NASA

National Aeronautics and Space Administration

EARTH RESOURCES LABORATORY

WESTERN ENERGY RELATED OVERHEAD MONITORING PROJECT PHASE II SUMMARY

REPORT #176

JANUARY 1979

(E81-10141) WESTERN ENERGY RELATED OVERHEAD MONITORING PROJECT. PHASE 2: SUMMARY Progress Report, 1 Jan. 1977 - 30 Jun. 1978 (NASA) 113 p HC A06/MF A01 CSCL 08F
N81-23550
Unclas
G3/43 00141

NATIONAL SPACE TECHNOLOGY LABORATORIES

The EARTH RESOURCES LABORATORY is an
element of the NATIONAL SPACE TECHNOLOGY
LABORATORIES, NSTL STATION, MS 39529:
and is located at 1010 Gause Blvd.,
SLIDELL, LOUISIANA 70458. Telephone
504 255-6511 or FTS 685-6511.

WESTERN ENERGY RELATED OVERHEAD MONITORING PROJECT

PHASE II SUMMARY

(January 1, 1977 - June 30, 1978)

by

James E. Anderson

EARTH RESOURCES LABORATORY

NATIONAL SPACE TECHNOLOGY LABORATORY

NSTL STATION, MISSISSIPPI 39529

Report #176

January 1979

TABLE OF CONTENTS

Section

LIST OF FIGURES

LIST OF TABLES

Table		Page

I. BACKGROUND AND SUMMARY

During the summer of 1975, the National Aeronautics and Space Administration (NASA) entered into a five year project with the Environmental Protection Agency (EPA). The purpose of this interagency project, funded by EPA, was to transfer hardware and software techniques for processing remotely sensed digital data from NASA to EPA, so that EPA would be capable of establishing and maintaining a fully operational energy related overhead monitoring system. This system would not only incorporate maximum utilization of present aircraft capabilities within the EPA, but would allow for the integration of currently available (Landsat) and proposed satellite multispectral scanner data. Training of personnel in the skills necessary to implement such a highly technical system would also be provided during the project. Specific details of the project can be found in the five year project plan, approved September 3, 1975 (Energy Accomplishment Plan 77ABI, Memorandum of Understanding #D5-E771, Subagreement #77BEL).

Both agencies have designated laboratories through which the five year plan will be implemented. NASA/National Space Technology Laboratory-Earth Resources Laboratory (NSTL-ERL) located in Slidell, LA, and the EPA/Environmental Monitoring and Support Laboratory (EMSL) in Las Vegas, NV, are the two agency laboratories involved. Unless otherwise indicated, "NASA/NSTL" and "EPA/EMSL" will be used to refer to these two laboratories respectively.

The entire project has been divided into three phases. Phase I was an 18-month project segment, during which time existing NASA/NSTL remotely sensed data acquisition, analysis and information producing capabilities were applied to digital data acquired by both Landsat and aircraft remote sensing systems over coal strip mines in the Western United States. A data processing system was defined, assembled, and transferred to EPA in January 1977. Key EPA/EMSL personnel also received training in the use of a similar data processing system during November/December 1976 at NASA/NSTL. A summary of activities conducted during Phase I of the project can be found in Anderson, et.al., 1978.

In addition to the data analysis system assembled at NASA/NSTL and transferred to EPA/EMSL, an airborne multispectral scanner (MSS) system was also defined. Such an MSS system, delivered to EPA/EMSL and subsequently installed into an EPA/EMSL aircraft and integrated with data processing/analysis components, provides EPA/EMSL with the capability to collect and process digital data entirely on its own system, thus accomplishing the major goal of Phase I of the project.

Phase II of the project (reported on herein) began January 1, 1977, and lasted for 18 months. During this time, NASA/NSTL investigated problems specifically defined by EPA/EMSL as needing additional research. Such problems were associated with on-line coal strip mine related activities. New techniques developed to solve these problems were transferred to EPA/EMSL at the end of this portion of the project.

Phase III has been designated as a two-year project segment, during which time EPA/EMSL will test the system in a fully operational mode. During this portion of the project, NASA/NSTL will provide continued assistance in the use of the system and the addition of new capabilities.

Phase II activities consisted of analyzing remotely sensed MSS data dealing with active coal strip mines in the Western United States. Technical details of data processing and analysis procedures required during Phase II are contained in following sections of this report. Technical terms are defined in the body of the text where appropriate.

Accuracy and utility of the products will be evaluated by EPA/EMSL with respect to their needs. However, this report will address the degree to which the data processing/analysis results obtained agree with general patterns as observed in the field, on maps, and on aerial photography.

Three specific tasks were defined which were to be addressed by NASA/NSTL during Phase II of the project. These dealt with (1) regional analysis applications based on the use of Landsat acquired MSS data and auxiliary data, (2) developing a technique for using aircraft acquired MSS data to rapidly monitor site specific surface mine related activities, and (3) registration of aircraft acquired MSS data to a map base. Procedures and software developed by NASA/NSTL to solve these tasks were transferred to EPA/EMSL at the conclusion of Phase II activities.

II. AN INTRODUCTION TO MULTISPECTRAL SCANNERS

The effective application of the electronic data supplied by multispectral scanner (MSS) systems is enhanced by an understanding of the basic concepts associated with MSS systems. It is the purpose of this section to discuss these basic concepts, as well as to define several terms associated with MSS systems and the data that they produce.

Readers possessing an understanding of such details may elect to skip this section and continue with the following one. For those readers whose interest in this subject extends beyond the scope of the materials presented in this section, more detailed information dealing with MSS systems may be found in NASA 1976.

Most MSS systems that are currently in use are composed of electronic, mechanical, and optical components that collect, separate, and record information related to the electromagnetic energy reflected or emitted from objects located beneath the device. Since MSS systems located on aircraft or orbiting satellites are functionally similar, no distinctions will be made between them in this discussion, except where such differences significantly affect the reader's understanding.

The MSS system (Figure 1) is designed in such a way that the data collection optics are directed to a corridor or swath below the device, the swath oriented perpendicularly to the direction of travel of the MSS, be it a flight line (aircraft

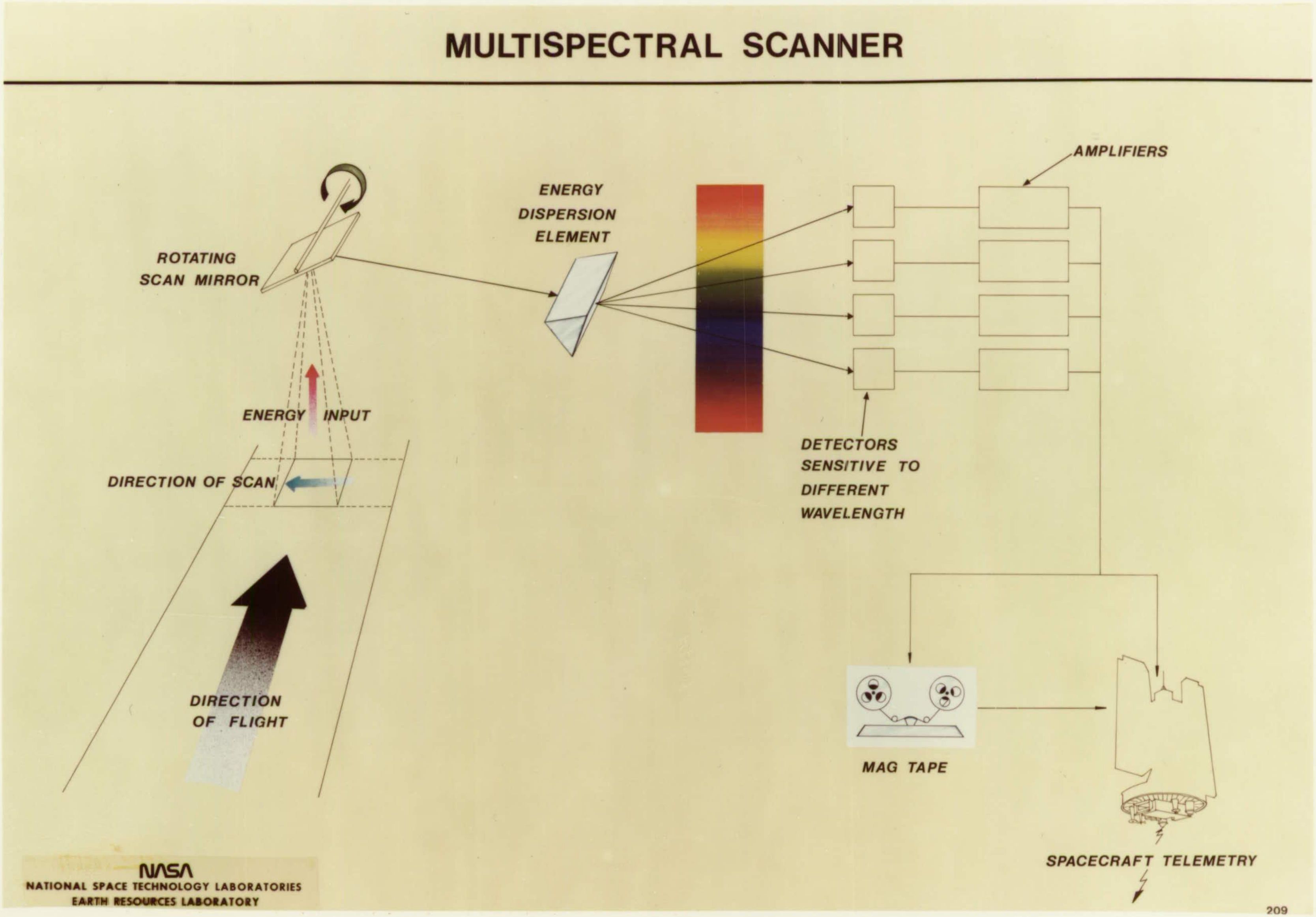

Figure 1 Schematic Representation of the Multispectral Scanner

systems) or an orbit track (satellite systems). This is accomplished through the use of a rotating mirror, which scans the swath from one side to the other. This feature of MSS systems has lead to the term "scan line" when describing the swath. As the mirror focuses the data collection optics across the scan line, electromagnetic radiation is recorded by the MSS system for discrete portions of the scan line. Each such portion of the scan line is referred to as an "element". There are numerous elements on each scan line, the actual number depending to a large degree on the optical field stop (measured in milliradians) of the MSS system and the effective scan angle of the optics involved. The scan angle is defined as twice the maximum angle, measured from the perpendicular, from which reflected electromagnetic radiation can be received by the MSS system.

Each data set thus collected is therefore composed of scan lines and elements, the number of scan lines depending on such factors as MSS altitude, scan rate of the mirror, speed of forward travel of the MSS, and length of the study area. Each scan line has associated with it a unique sequential number; similarly each element along a particular scan line has a sequential number assigned to it. The scan lines and elements can alternatively be envisioned as rows and columns of a grid, respectively. Thus, any geographic point below the MSS can be correlated with the unique scan line and element number pair of the MSS data.

Every MSS system has associated with it a calculable intrinsic limitation called "ground resolvable distance." This term refers to the actual dimensions of each element in a data set that is collected by a particular MSS, and is a function of (1) MSS altitude above a target, (2) MSS optical field stop, and (3) MSS scan angle. At nadir (point directly beneath the MSS system) the element size (ground resolvable distance) can be found as follows:

$$ES = A * mr$$

where: ES = Element Size on Each Side

A = MSS Altitude (in 1000's of feet) Above Target

mr = MSS Optical Field Stop (in milliradians).

Thus, if an MSS system were 30,000 feet above a target, with an optical field stop of 2 mr, the element size (at nadir) would be:

$$\begin{aligned} ES &= (30 \text{ thousand feet}) * 2 \text{ mr} \\ &= 60 \text{ feet on each side} \\ &= 3600 \text{ feet}^2 \text{ surface area} \end{aligned}$$

It can also be demonstrated that as the scan angle increases from the perpendicular (moves toward the ends of the scan line) element size increases, due to the geometry involved. This physical principle is most pronounced when examining MSS data collected at low altitudes (aircraft). Software designed to analyze MSS data usually takes this geometric problem into consideration.

As is portrayed in Figure 1, the electromagnetic radiation from each element is defracted into its constituent parts in the MSS. By positioning appropriate photo-detectors around the periphery of the defraction device at prescribed locations, a sample of the electromagnetic radiation pertaining to a particular element can be made. Such a sample is drawn from wavelength regions of the electromagnetic spectrum defined by the location and physical width of the individual photo-detectors used. These wavelength regions, to which the MSS system is sensitive, define the spectral selectivity or spectral resolution of the MSS, and have been referred to as "channels" or "bands" of the MSS. Specific spectral resolutions of the MSS systems employed for use in the study reported on herein are presented in Table 1.

For each element of any given scan line, several "readings" are recorded on magnetic tape onboard the aircraft or are transmitted from the satellite to ground receiving stations. These readings express, for any particular MSS channel, the relative amount of reflected and/or emitted electromagnetic energy received at the detector. Readings are taken for each of the channels incorporated into the MSS system. As an example, the 11-channel scanner outlined in Table 1 would produce 11 such readings per element.

The data thus collected and stored on magnetic tape are analyzed through the use of digital computers. These computers read the recorded data for each element from the magnetic tape and process it according to numerous software programs developed for this purpose. Detailed analysis procedure and explanations for software implementation can be found in the following section of this document.

Table 1 - SPECTRAL RESOLUTION OF THE AIRBORNE MULTISPECTRAL SCANNER SYSTEM* AND THE LANDSAT MULTISPECTRAL SYSTEMS USED IN THIS PROJECT WITH CORRESPONDING APPARENT COLORS

Aircraft Channel or Band	Wavelength Region in microns	Apparent Color
1	0.38 - 0.44	Violet
2	.44 - .47	Blue
3	.495 - .535	Cyan/Green
4	.54 - .58	Green/Yellow
5	.58 - .62	Yellow/Orange
6	.62 - .66	Orange/Red
7	.66 - .70	Red
8	.70 - .74	Far Red/Near Infrared (IR)
9	.76 - .86	Reflected IR
10	.97 - 1.06	Reflected IR
11	9.50 - 13.50	Thermal (Emitted) IR

Landsat Channel or Band	Wavelength Region in microns	Apparent Color
4	.5 - .6	Blue/Green/Yellow
5	.6 - .7	Orange/Red
6	.7 - .8	Far Red/Near Infrared
7	.8 - 1.1	Reflected Infrared
8**	10.4 - 12.6	Thermal (Emitted) IR

* Supplied by the manufacturer
** Landsat 3 only

The Landsat satellite system is composed of three sun-synchronous MSS bearing satellites. The satellite MSS system operates in a manner similar to the aircraft MSS used in this study except for the following:

(1) Element size is approximately 1.1 acres (56m x 79m element size)

(2) Spectral resolution is restricted to four channels for Landsat I and II, with the addition of a fifth channel in Landsat 3 (Table 1)

(3) Each satellite has a regular 18-day cycle period; i.e., each satellite passes over a particular ground spot every 18 days

(4) Data stored onboard the satellite is transmitted to a ground receiving station where it is processed and placed on magnetic tape for subsequent user requests

(5) Data is placed on magnetic tapes corresponding to 100 by 100 nautical mile areas (115 x 115 statute miles) after reception from the satellite

(6) Scan angle effects are minimized due to the relatively narrow scan angle involved (11.56 degrees).

Aircraft-acquired MSS data has been shown to be an effective method for dealing with site-specific, limited geographical area considerations (Anderson, et.al., 1978). However, other factors which might be under investigation include such things as the cumulative effect of several sites in a regional sense and non-site specific effects; e.g., drainage, etc. Aircraft coverage of such factors may be impractical, due to (1) the very large areas involved (requiring large amounts of data to be collected, stored, and analyzed), and (2) the possible need for frequent repetitive coverage to monitor change for which aircraft associated costs would be prohibitive. The use of Landsat satellite-acquired digital data facilitates assembling information over such extensive areas of interest at frequent intervals.

Several features associated with Landsat digital data serve to enhance its use for the "regional" type of analysis. First, costs associated with computer processed data covering large areas are relatively low when compared to the corresponding aircraft data-derived cost figures. This is not only related to the amount of data needed to cover a particular area, but other special considerations discussed in the following section. Second, since the satellite system has a regular, established coverage date (charts are available), repetitive coverage is possible.

In addition to currently orbiting earth resources oriented satellites, plans for future satellite MSS systems include the Landsat follow-on thematic mapper, with seven channels and 40m element size (planned 1981 launch).

III. DATA PROCESSING CONSIDERATIONS

This section contains information related to the more technical aspects of data processing which occurred during Phase II activities of the Western Energy Project, and has been included in an attempt to familiarize the reader with data processing methodology employed by NASA/NSTL. References, where made, can be found in the Literature Cited section situated at the end of the main body of this report.

In general, data processing and analysis activities which occur at NASA/NSTL follow a defined flow or pattern. This overall flow is graphically depicted as Figure 2, and is applicable for the most part, to either aircraft or satellite acquired MSS data. Where differences do occur for aircraft or satellite MSS data processing, they will be discussed. Throughout the discussion presented in this section, reference will be made to Figure 2 so that the reader will be better able to orient himself to the data processing flow at any given moment.

The mini computer that serves as the heart of the data analysis system (DAS) requires specific computer compatible magnetic tape (CCT) and card input formats for data processing activities. (See NASA/Earth Resources Laboratory, 1975, for specific formats required of CCT's.) This requirement is directly related to the programming method employed when the DAS was established. Landsat MSS data, which is stored on CCT's, is not in the required format when received from the supplier. In order to use such data

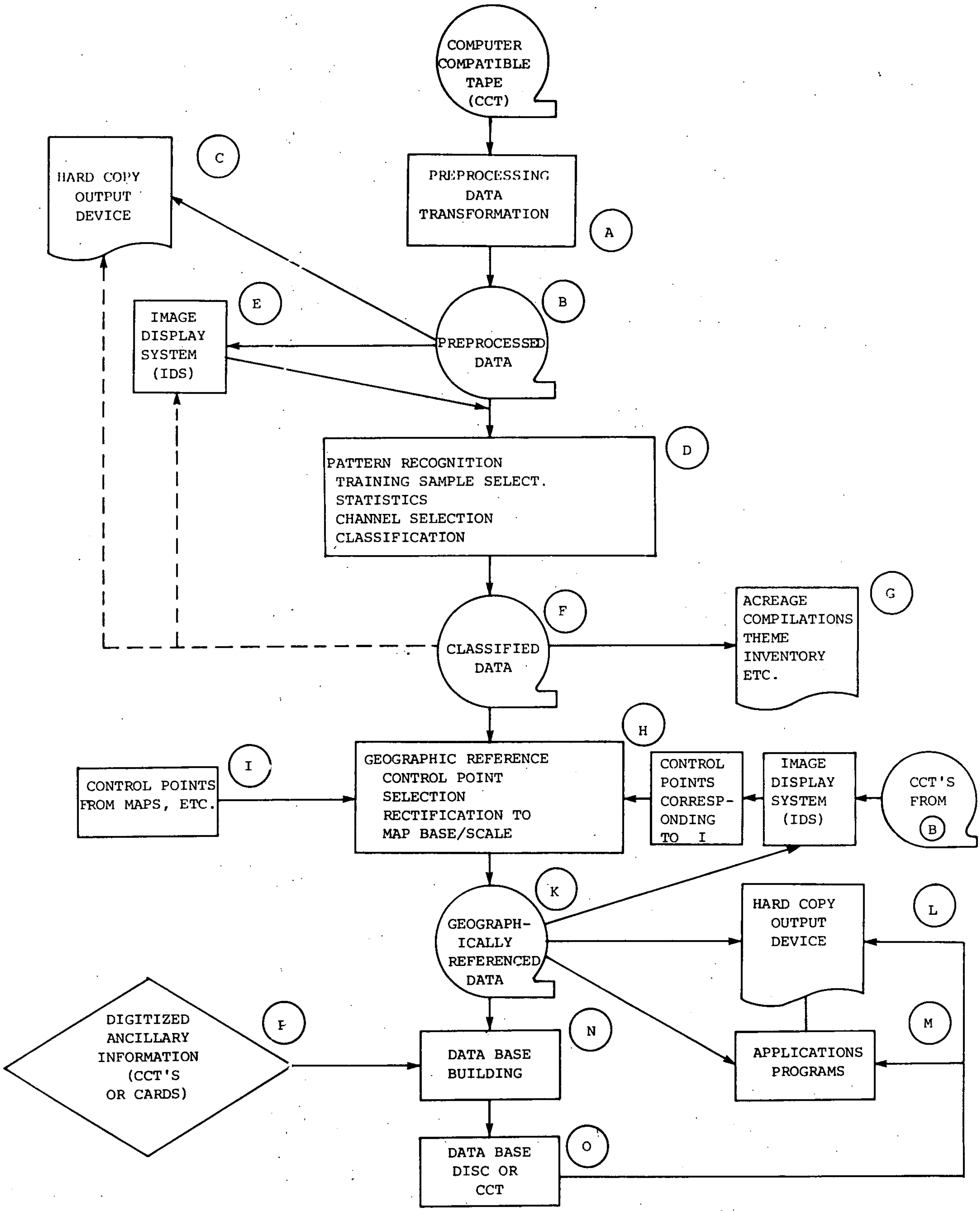

Figure 2. Flow Diagram of General Data Processing and Analysis Activities Conducted at NASA.

for analysis, the format must be altered to that acceptable to the DAS. This reformatting procedure is the initial step of the Preprocessing Data Transformation (PDT) Module (A in Figure 2), and is accomplished through the use of software specifically designed for this purpose.

Aircraft acquired MSS data must also be converted to a suitable format. However, unlike Landsat MSS data which is in a digital mode when it is received, aircraft MSS data is recorded in an analog mode, which cannot be decoded by the digital mini-computer of the DAS. It therefore must be decommutated, which is a process whereby the original analog mode aircraft MSS data is converted to digital mode data, in a format suitable for use on the DAS. The result of this procedure is a tape which is quite similar to a reformatted Landsat tape. Decommutation is also considered as being a component of the PDT module (A in Figure 2).

Upon completion of reformatting or decommutation and while still in the PDT module, the data is analyzed in an effort to ascertain whether or not correctable data related problems exist which might interfere with or reduce the overall effectiveness of subsequent data processing and analysis. One such problem, associated with Landsat MSS data, is termed "scan line banding."

Scan line banding manifests itself as periodically occurring "stripes" or "bands" in the Landsat MSS data image, usually only one scan line in width, which appear either lighter or darker in tone than the scan lines which preceed and follow. This problem is directly related to the fact that the MSS on the Landsat satellite actually collects data for six scan lines simultaneously with each oscillation of the scanning mirror. This means that there are six detectors for each of the channels on the satellite, and scan line banding occurs when one (or more) of the detectors is radiometrically out of calibration with the remaining detectors for a particular channel. Software has been designed at NASA which permits the investigator to minimize the effects of scan line banding, while at the same time maintaining the data integrity of the affected scan lines.

Additional problems encountered when processing Landsat or aircraft acquired MSS data are missing or radiometrically "garbled" scan lines. These problems are easily detected when the data is viewed in picture form. Missing data might result in physiographic or topographic features making abrupt and unexplainable changes in course or direction. "Garbled" data is seen as a "salt and pepper" scan line, with no detectable correlations to scan lines surrounding it, and no continuity of data within the scan line. If not occurring in large amounts, these problems can be dealt with (reconstructed) through the use of additional software programs developed at NASA/NSTL. Such software are included in the PDT module (A in Figure 2).

Aircraft MSS data, collected at relatively low altitudes when compared with the satellite data, exhibits a somewhat unique data problem, termed "sun angle-scan angle" effect. This problem results in a change of radiometric data values across a scan line for a particular land cover type. For instance, a coal seam that extends across the MSS scan angle may have one data value on the left extreme of scan, a different value at nadir, and yet a different value at the extreme right of scan, even though the coal seam does not change in physical makeup. Confounding the problem is the fact that the lower wave lengths (blue end of the spectrum) are affected to a greater degree than the high or red end. These problems are related to scan angle and azmith of the sun. In order to compensate for this situation, software has been developed which will remove such conditions from the aircraft acquired MSS data, thereby improving the data for subsequent processing and analysis.

The PDT step, whether applied to Landsat or aircraft acquired MSS data, results in a CCT containing MSS digital data that have been selected by the investigator for subsequent analysis, and hence has most problems associated with data quality removed (B in Figure 2). Such a CCT serves as the base data set for all ensuing data processing and analysis activities and is in a format that permits hard copy reproduction of the raw MSS data (to any suitable scale) by either an electrostatic printer/plotter device, or through the use of a color continuous strip film recorder (C in Figure 2). (Both of these devices are described in Whitley, 1976). While hard copy reproduction of the raw MSS data is not mandatory for further data processing, it produces a product which can serve as a source of quick reference for the entire data set, and a "map" which can be used to record data-oriented information.

Actual data processing and analysis begin when the software programs contained in the <u>Pat</u>tern <u>Rec</u>ognition (PATREC) module are implemented (D in Figure 2). It is within this block that the MSS data are actually classified and therefore the component steps in this module will be presented separately.

The initial step in the PATREC module is training sample selection. (A training sample is defined as an area in the MSS data corresponding to an area on the ground which has been previously selected since it contains a uniform, homogeneous land cover. Such sites are, in a simplistic sense, used to "train" the computer to recognize the same land cover elsewhere.) The procedure of training sample selection deals with the task of locating predefined training sample boundaries in the digital data, matching the boundaries as closely as possible to the way in which they occur on the earth's surface. This task is accomplished through the use of an Image Display System (IDS)(See E in Figure 2). This device, linked to the mini-computer and described in detail in Whitley, 1976, permits a visual display of the digital MSS data to be made. Such a display can be in either a black and white or color enhanced mode. Using such a display, training sample boundaries can be located through the use of a movable cursor, which is shaped in the form of a "+" sign. Each corner of the polygon containing the sample is located, and the corresponding scan line-element coordinates are read/recorded by the mini-computer. This process is repeated until all training samples have been located and their boundary coordinates have been stored on disc memory.

The next process that occurs is the computation of statistics associated with the MSS digital data contained in each training sample (D in Figure 2). This step can actually occur immediately after each individual training sample is located, or all training samples may be chosen and the statistics can then be computed for all training samples at the same time. However, regardless of the technique involved, the resultant statistics (including a mean vector and covariance matrix) define the distribution of training sample data MSS radiometric values in a multi-dimensional space (which in turn is defined by the number of channels being used). This characteristic of the training sample statistics has led to the usage of the term "spectral signature" to refer to training sample multispectral statistics. Training sample selection by which spectral signatures are derived has been referred to in the literature as "signature development".

As the statistics are computed for each training sample, the investigator must make several decisions. First, it must be determined whether the training sample population is statistically homogeneous. This means that the histograms or frequency counts of the data (for all channels) appear to be fairly normally distributed, with no "outliers" (data points located at some unusually large distance from the mean) or multimodal tendencies. Second, the actual number of data points (elements) in each training sample must be of a sufficient number to constitute a statistically "valid" population (generally thought to be greater than 30 elements). By selecting the largest possible training sample (and hence including the greatest number of elements), statistical tests conducted at a later time can be made with greater levels of confidence. This helps to assure a much more reliable output.

The third decision that must be made relative to the statistics is to determine that the coefficient of variation (computed as $\frac{\text{standard deviation}}{\text{mean}}$ X 100%) for any and all channels is of a sufficiently low magnitude. Experience with various data sets indicates that acceptable maximum values for the coefficient of variation (CV) lie in the range of 10-15% for aircraft acquired MSS data, and 5-10% for Landsat MSS data. However, these values are by no means considered absolute; the investigator must maintain enough flexibility to alter the acceptable maximum values of the CV in response to the quality and type of data being processed. Thus for radiometrically "clean" data, CV's of 5% for Landsat MSS data might be upper limit values; for low quality data 15% might be the maximum. Selection of an acceptable maximum value becomes somewhat easier as experience is gained in processing MSS data.

The final decision that must be made is one dealing with the overall radiometric "purity" of the data. As previously mentioned, outliers or multimodal tendencies would indicate potential problems with training sample location or composition. Other problems include (but are not restricted to) data value skipping (observable on the histograms as skips or gaps in the histograms at regular intervals), saturation (data values consistently occurring at extreme high limit of signal for a particular channel), and pure samples (no variance--all data have the same radiometric value--primarily a problem associated with water).

If none of the problems mentioned above are observed, the statistics for the training sample under consideration can be stored for later use. If, however, one or more problems are encountered, corrective actions should be undertaken. These actions may be taken in such a manner that the training sample is simply dropped and is no longer considered in the analysis. Alternately, the training sample can be relocated in the digital data and a new set of statistics computed. This latter action is highly desirable in situations where the particular training sample in question is the only one representing a particular land cover type. If this land cover type is significant, the loss of the training sample might prove detrimental to the overall utility of the resultant product. If, however, the training sample is but one of several representing a particular land cover category, the loss might not be critical. (All of the above activities are represented by the block labeled D in Figure 2).

It is for these reasons that it is highly desirable to utilize at least three training samples per land cover category. By doing so, the possibility of "losing" an important land cover type is minimized. In addition, reselection of a training sample would not be necessary. This is quite important, particularly if ground truth information[1] was obtained by

[1] Ground truth information refers to data which describe the land cover in existence on a training sample site (in the field), and deals with vegetative cover, non-vegetative factors such as rock outcroppings, sand bars, etc., and topographic considerations, all of which will effect the spectral characteristics of the site. Detailed procedures for gathering ground truth information can be found in Joyce, 1978.

actual onsite field visitation by an earth scientist (e.g., forester, geologist, etc.). Reselection of samples would require revisitation and additional time and financial costs would be incurred. If, however, ground truth information is in the form of interpretations taken from aerial photographs, land use maps, soil maps, etc., reselection of training samples would not impose severe setbacks in terms of time or costs.

Training sample selection and subsequent computation of statistics are activities conducted on the interactive IDS (E in Figure 2). This system is composed of an operator's console, which allows for communication between the computer (not considered part of the IDS) and the investigator, and an image display device. This device is composed of a color cathode ray tube (CRT) similar to a standard color television set, and in addition, a disc memory that stores the information derived by the mini-computer from the CCT containing the MSS data and places it in picture form on the CRT. Descriptions and photographs of the IDS can be found in Whitley, 1976. (The IDS, mini-computer, peripherals, etc., form the DAS.)

The complete set of training sample statistics which are to be used in subsequent data processing steps are stored on a separate, addressable computer disc memory file. However, before proceeding to the actual classification of all elements in the data set, one more statistical analysis is necessary. This step, referred to as divergence analysis, permits the investigator to determine whether or not any particular pair of training samples are spectrally separable from one another, and if not, how meaningful groupings of similar training samples can be made. (This step is also encompassed by the set of activities labeled D in Figure 2).

The vehicle for determining the relative spectral separability which exists between two training samples is a computed value called divergence. (See Jones, 1976, for a detailed discussion of divergence.) This value is calculated by the interactive program IASTAT (Interactive Statistics). Training sample statistics are used to estimate the capability of the computer to spectrally discriminate two training samples from each other. The divergence between the two samples, calculated from the individual training sample statistics, is printed out in matrix form for up to 64 training samples at a time. As the relative magnitude of the divergence between any two training samples increases, the spectral separability and hence, the computers discriminatory capability increases. The converse is also true.

Perhaps the only problem associated with the use of divergence is that of deciding a "critical" divergence value. This value is defined as being that divergence magnitude, above which training samples are considered spectrally separable and below which training samples are considered inseparable and hence should be grouped. For experience, it has been found that this value fluctuates in a manner influenced by (1) the land cover categories to be classified, (2) the overall radiometric characteristics of the digital data, and (3) the overall range in values for a partial set of training samples. It therefore becomes necessary to establish the critical divergence value only after analyzing the initial IASTAT run (no groupings made previous to this initial run).

Should two training samples (for example "coal seam" and "water" which are totally dissimilar as land cover types) appear to be spectrally inseparable, several checks should be made before reaching the conclusion that they are not spectrally separable (and hence the land cover types they represent are not spectrally separable). First, the location of the training samples in question should be rechecked on the CRT, and second, the descriptive name associated with it should be verified, to assure that a "water" training sample was not mislabeled as being "coal seam" or vice versa.

Training samples which are not spectrally separable may be combined; i.e., the data points (elements) for two or more training samples may be grouped into larger populations and the statistics for this larger group are then computed and stored. Thus, training samples representing the same or similar land cover types, and exhibiting appropriate divergence values, could be grouped into one larger "training sample."

After computation of statistics for the grouped sample populations has been completed, the statistics for the grouped populations and all training samples not grouped are compared again, with inseparable ones being combined. This process is iterated as many times as necessary to produce training sample populations which are spectrally separable. From this point on, such spectrally separable populations are referred to as "classes," each class corresponding (in a spectral sense) to a particular land cover category defined by the investigator. Classes may be composed of several land cover types (e.g., natural vegetation class is composed of sagebrush, rabbitbrush, etc.) in the case of classes composed of several individual training samples. It can be readily seen that as more and more groupings are made, the detail of the classification may decrease.

Before proceeding on to the classification technique, a special characteristic associated with aircraft acquired MSS data should be mentioned. Most airborne MSS systems contain more than four channels (which do not in general relate to the Landsat MSS channels). However, software developed by NASA/NSTL restricts the number of channels used as input to the classifier to a maximum of four. This restriction to four channels does not seem to affect the adequacy of spectral separation of major land cover types of interest to EPA/EMSL. (Anderson, et.al., 1978)

Several techniques can be used to reduce the number of channels to four; one such technique utilizes the program SEPARATION. This program has been designed to select, from an input set of "n" channels, the four channels that will lead to the best overall spectral separation of the training samples used as inputs. Only one set of four channels will be selected for use in classification. The statistics generated for the individual training samples are actually utilized in a manner similar to that described in Jones, 1976. The statistics are then recomputed for the four chosen channels for each training sample, divergence analysis is conducted, and groupings, again based on only four channels, are made. (PATREC module - D in Figure 2).

Computer implemented classification of MSS data is dependent on the use of multivariate statistics describing spectrally separable land cover types. Classical "supervised" techniques previously described rely on the investigator selecting training samples from the MSS data, analyzing the resultant statistics, and grouping such statistics if it is determined that they represent spectrally non-separable land cover types. This procedure requires that ground truth information be available for use for all such training samples used, and that the boundaries of the training samples be recognizable in the MSS data.

SEARCH is a program which can be used to develop statistics for spectrally separable land cover types without the need for ground truth information. In this sense, it can be called an "unsupervised" technique, since no manual manipulation of the data is required by the investigator.

SEARCH evaluates every contiguous six scan line by six element block of data for spectral "purity". This is done by computing the statistics for the block of data under evaluation, and comparing them to threshold values input by the investigator. A block of data is retained only if the standard deviations for all channels lies in the range specified by the investigator, and if the computed coefficient of variation ((standard deviation/mean)*

100%) is less than or equal to that input as a threshold. Statistics for those blocks deemed acceptable for use are stored in memory.

The process of evaluating six scan line by six element blocks is continued until 50 sets of statistics have been stored. At this point, pairwise divergences are computed, and that pair with the smallest pairwise divergence is combined. This merger permits the use of the fiftieth slot, and an alternating statistics evaluation-statistics merger continues until all input data has been evaluated. If at any time during execution of SEARCH, the smallest pairwise divergence is larger than a user input maximum separability for merger, all statistics derived from a single six by six block with a divergence of less than 4.5 times the maximum separability for merger, when paired with statistics derived from two or more six by six blocks, are eliminated. If in this case no statistics are removed from memory, the maximum separability for merger is incremented and the process is repeated.

Statistics retained in memory after a final deletion has been made are written into a computer disc memory file. These statistics can then be used as inputs to the classifier, which in the case of this study was MAXL4 (Maximum Likelihood-4 channels of data).

At this point in the data processing flow, data classification occurs (D in Figure 2). Software developed by NASA/NSTL and referred to as MAXL4 accomplishes this task.

MAXL4 is a four channel maximum likelihood data classification program which uses as inputs inverses of class covarience matrices, probabilities at input quadratic thresholds, class constants, and area of the input data to be classified. Unlike straight forward maximum likelihood classifiers currently in use however, MAXL4 contains three "speed-up" execution sections which greatly cut down the processing time in most cases (with respect to simple maximum likelihood classifiers).

The first time-efficient processing section is the unpack/compare subroutine. This subroutine compares the element under consideration with the one previously classified. If this element is identical to the last element classified, the result of the classification of the previous element is duplicated. If an identical match is not found, the program continues on with a maximum likelihood type approach. This first "element comparison" technique speed-up achieved is somewhat data dependent, but has been found to be very helpful when processing data covering areas of relatively homogeneous land cover. Water bodies result in a substantial speed-up, as might be expected.

The second manner in which a speed-up in the classification can be achieved is the manner in which elements not "classified" in the "element comparison" are subsequently classified. During this procedure, an initial probability (PROB) is established as a negative number with a large absolute value, and an index (M) is initialized to one. This point is subsequently referred to as the class re-entry point.

The mean vector of class M is subtracted from the vector of the element under consideration, and the quadratic threshold of the difference is computed. This computed quadratic threshold is subtracted from the "class constant" (which is simply the log of the probability of the occurrence of class M given the mean vector). The resultant difference represents the log probability of the occurrence of class M. This value is compared with PROB. If the log probability of class M is greater than PROB, then it is stored in PROB and M is stored in an index parameter "IND". M is subsequently indexed by one and the process is repeated. Should PROB be less than the class constant of class M, the program is directed to the class re-entry point (mentioned previously), since otherwise the element most likely is represented by class number "IND". If, on the other hand, PROB is greater than the class constant of class M, then none of the remaining classes could possibly be most likely to represent the element under consideration, since the class constants have previously been sorted and stored accordingly. The classification (with or without a priori values) can now be made.

The log probability/class constant comparison speed-up is most influences by input a priori values when processing satellite data, since the determinants of the covarience matrices do not tend to vary to any great degree. This technique has been shown to greatly speed-up the processing of aircraft data, where the trend is for the determinants to vary much more.

The output from MAXL4 is a CCT containing classified data (F in Figure 2), and line printer statistics indicating the number of elements (and percent of data) classified into each class. The tape serves two purposes. First, it can be used to furnish acreage statistics, calculated by an acreage program called POLYAC (Polygon Acreage). In addition, other software packages permit individual theme (class) inventories to be made. Such software is contained in module G of Figure 2.

The second use for the classified CCT is in the production of hard copy products (C in Figure 2), such as color or black and white film recordings (to desired scale), electrostatic printer/plotter output for use in the generation of CROMALIN products, or simply viewing on the IDS. The latter is generally performed first, to verify, in a broad sense, that the classification fairly reliably represents the surface land cover patters in the area of interest.

Prior to the time work had been conducted for one task of Phase II of this Project, only Landsat MSS data would have been subjected to additonal processing, since no technique had been developed for referencing aircraft acquired MSS data to a map base. However, aircraft MSS data can now be treated in a manner similar to that of Landsat MSS data in as much as the capability exists to geometrically correct aircraft MSS data and reference it, like Landsat MSS data, to the Universal Transverse Mercator (UTM) map projection system. This process occurs at H of Figure 2, and is completed for both aircraft and Landsat acquired MSS data in two steps. (The procedure for registering aircraft acquired MSS data is described in a later section of this report.)

The first step in the registration process is to locate, on a suitable map base, several geographic or cultural features that are recognizable on the CCT's from which training samples were selected. The coordinates for these points (in the UTM system) are recorded and serve as part of the input to the Geographic Referencing (GEOREF) program (I in Figure 2). These same points are located in the digital data, and their scan line/element coordinates are recorded (J in Figure 2). These coordinates make up the other input to GEOREF (Graham, 1977).

Using both sets of control point coordinates, GEOREF calculates correction parameters for the MSS data. Should the output indicate it, one or more control points can be deleted, due to excessively large errors encountered in the calculation of the corrected location for that point. For this reason, it is recommended that numerous control points be selected for each data set, distributed as uniformly as possible over the entire geographic area contained in the data.

Geographic correction parameters, when applied to the classified data, accomplish two things. First, the data are geographically positioned to within 100 meters rms (at 66% level of confidence) for Landsat MSS data and to within a user specified rms error for aircraft acquired MSS data, based on an analysis of the "fit" of the control points. Second, the data are converted from the 59 m by 79 m element of the original MSS data to a 50 m x 50 m GEOREF element in the case of Landsat MSS data, and to any desired output cell size in the case of aircraft MSS data. This conversion is accomplished by comparing the GEOREF cell to its "nearest neighbors" of the original MSS data.

GEOREF CCT's (K in Figure 2) can be utilized in numerous ways, including:

1. Map making (Item L in Figure 2) through filming (to a desired scale) the digital classification on color or black and white continuous strip film, or through density plot/CROMALIN process (Whitley, 1976)

2. Application programs (Item M in Figure 2) including change detection, acreage compilation (by class), individual class inventory, etc

3. Data Base (DATBAS) construction (Item N in Figure 2). One difference between a GEOREF CCT and a DATBAS (Item 0 in Figure 2) is that, in the DATBAS, each GEOREF element has associated with it up to 26 types of information, acquired from MSS data and from other sources (such as elevation from topographic maps, soil survey, etc.). This information (Item P in Figure 2) is useful in either mapmaking (based on more detailed and accurate sources of information) which is computer-oriented, or in special purpose applications (Item M in Figure 2). This capability to utilize auxiliary information adds a new dimension to the processing/utilization capabilities of remotely sensed digital data.

It should be noted, however, that if the need for using other map projection system exists, software modifications must be made which would allow for the use of such systems.

IV. REGIONAL ANALYSIS APPLICATIONS BASED ON THE USE OF LANDSAT MULTISPECTRAL SCANNER DATA AND AUXILIARY DATA FOR REHABILITATION POTENTIAL ESTIMATION

A. Description of the Task

The objective of this task of the project was to develop an automated procedure which utilizes Landsat acquired MSS data and auxiliary data in such a manner that a classification or estimation of rehabilitation potentials for user selected areas could be produced. Such a procedure would be developed for use on a "mini-computer" data processing system, of a configuration similar to that assembled for and transferred to EPA during Phase I of the project. (The reader is referred to Anderson, 1976 (Appendix C) and Whitley, 1976, for technical information dealing with the configuration of the EPA (and similar) data processing system.) Software required to implement the concept was developed by NASA/NSTL and subsequently transferred to EPA at the conclusion of Phase II of the project.

Auxiliary data, as used in the context of this section of this report, refers to information not directly associated with Landsat MSS data, but nonetheless considered as being of a significant importance to the overall effectiveness of the rehabilitation potential estimation procedure devised. Auxiliary data is quite commonly (although not always) found in map form, of varying detail and map scale, and is available from numerous sources. Examples of such data include soil information (from soil maps developed by the Soil Conservation Service or other agencies), rainfall (from maps or data compiled by the National Weather Service), and elevation (derived from topographic map sheets of various scales produced by the U.S. Geological Survey).

B. Scope of the Task

Demonstration of the procedure was to be conducted in a one degree latitude by one degree longitude study area situated in such a manner that the northern two-thirds of Campbell County, Wyoming was included (Figure 3). This area was chosen for the following reasons:

1. The area is surrounded by the Powder River Basin, an area of vast coal reserves and of interest to EPA/EMSL

2. The Powder River Basin contains numerous active and proposed coal strip mine sites, thereby affording the opportunity to check results obtained from the implementation of the procedure

3. The area typifies a semi-arid environment which exists over a large portion of the western U.S.

4. Landsat MSS data, of good quality and of a date of collection corresponding to "peak of green" conditions in the study area, was available for use

5. Auxiliary data, in a format suitable for use, was obtainable.

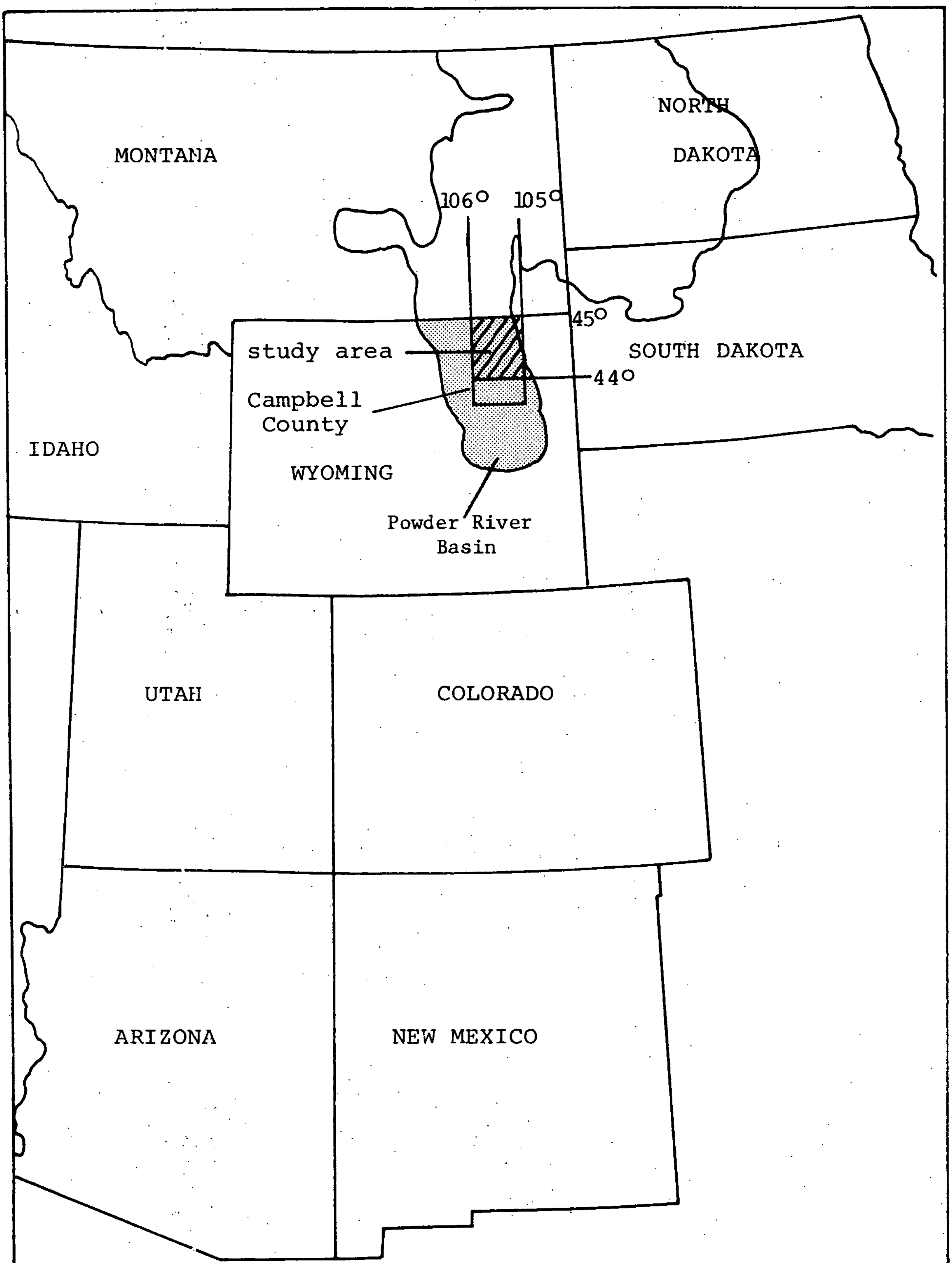

Figure 3. Location of the Study Area for the Investigation Dealing with the Regional Application of Landsat Data

Implementation of the procedure would depend on the use of Landsat MSS data to furnish land cover information, and the incorporation of auxiliary parameters which, when examined together, could be used as a basis for estimating the rehabilitation potential of a user selected area.

C. Technical Approach

The solution to the problem of developing an automated rehabilitation potential estimation procedure lies in the use of a geographically referenced data base (DATBAS). The specifications provided by EPA/EMSL for the development of the DATBAS were as follows:

1. It was to be based on the use of an interactive processing system

2. It was to accept both polygon digitized and grid digitized input

3. It was to be capable of containing data for a one degree latitude by one degree longitude area (at the equator)

4. It was to be capable of storing a minimum of 6 parameters of environmental information in addition to locational parameters

5. It was to provide for a choice of cell size in multiples of 50 meters from 50 meters up to 400 meters.

The DATBAS designed for use in this task is capable of storing information for up to 26 parameters of interest over a one degree of latitude by one degree of longitude geographic area. Not all 26 parameters are available for use in storing environmental or other such data, however, as several must be reserved for use as locational aids (State, County, etc.).

One additional item, which adds a great deal to the utility of the DATBAS structure developed in this task, warrants mention at this time. The investigator has the opportunity to interactively edit the data contained in the DATBAS. Such an editing capability could possibly be related to the addition of newly acquired data, the insertion of data previously thought to be unavailable for use, the modification of data found to be incorrect, etc.

The investigator can also make use of the interactive nature of the DATBAS in yet another manner. Several data processing subroutines have been incorporated into the DATBAS interactive software which will permit the investigator to perform tasks which are more or less data dependent, e.g. compute acreages for specific areas, perform data overlays (described later in this section of the report), and produce annotated products (examples of which are contained in this report) all of which are accomplished in a real time mode.

Operational implementation of the procedure involves the steps outlined below, assuming that all pertinent data have been stored in the appropriate DATBAS location:

a. The user selects the area that is of interest, and isolates that area from all else surrounding it. In rehabilitation applications, this is envisioned as being a sub area of the original 1° by 1° DATBAS. This step is most commonly carried out using the land cover information contained in the DATBAS, as it is quite easy to relate land cover patterns to maps, etc. However, the DATBAS is structured in such a manner that the data contained within it is referenced to the Universal Transverse Mercator (UTM) map projection system, and hence areas of interest can be located through the use of UTM northing and easting coordinates defining the boundaries of such areas.

b. The user next examines the appropriate parameters of the DATBAS, either singly or as they co-exist, and compares each or all with a set of values that he has developed to define various levels of rehabilitation potential. The actual method by which this examination/comparison step is accomplished is discussed in a later section of this report.

c. The user then assigns or estimates the rehabilitation potential for the area under consideration based on his set of values developed for his area of interest. In this manner, any area located in the DATBAS can be examined and a rehabilitation potential estimate can be produced. Should an area of interest be located in two (or more) DATBAS's, the procedure is repeated for each fraction of the study area on each DATBAS containing a portion of the entire study area. The results are simply pooled together to arrive at the desired estimate.

D. Procedures and Methodology

1. Processing Landsat MSS Data

Work commenced with the acquisition of Landsat MSS data from the EROS Data Center, located in Sious Falls, South Dakota. The EROS system is set up in such a manner that users of Landsat data can obtain computer data searches for data covering a particular geographic area during a specific time frame. The output obtained from these computer searches is a tabularization of the available Landsat data which covers a desired geographic area within dates specified by the user. Additional information contained in the computer printout indicates such things as Landsat acquisition (scene) I.D., date of collection, relative quality of the MSS channels (on a scale of one (low) to eight (high)), approximate percent cloud cover (in increments of 10 percent) and latitude/longitude coordinates of the corners of each Landsat scene. In most cases, this is sufficiently detailed to permit the user to select and order MSS data. However, regional centers have microfilm cassettes available for viewing which contain black and white format transparencies of all Landsat MSS data available, and should the user desire to preview the data before ordering it, these microfilms can be used. In addition, if the nearest EROS office is at some distance from the users location, black and white prints of the data can be ordered from the Sioux Falls office which can serve the same preview function. This, of course, adds to the overall data receipt time, but in many cases can cut down the expense of receiving data that can not be used.

EROS charges a flat fee of $200.00 for CCT's containing Landsat four channel MSS data covering a 100 N. mile by 100 N. mile Landsat scene. (Prices included in this report reflect charges in effect at the time of this writing. For current prices, consult with the respective data supplier.) In general, the time lag between ordering and receipt of Landsat MSS data is between 2 weeks to one month.

Data for this project, specifically Landsat scene 2889-16452 (29 June 1977 collection date), was received in January 1978. This Landsat scene was selected for use since: (1) the entire study area was located within the geographic boundaries of the data set, (2) all of the MSS channels were rated as being excellent, (3) cloud manifested problems were very minimal, (4) data collection date (29 June) is compatible with other data processed in the same area, and (5) the data was the most recent available. A black and white photographic print of Landsat channel 5 (.6-.7m) is included as Figure 4, and will serve to orient the reader to the study area as it is situated in the Landsat data scene. Upon receipt, the CCT's containing the data for the study area were reformatted and corrected for scan line banding.[2]

In September, 1977, ground truth teams visited the study area and collected detailed information for training sample sites

[2] The reader is referred to the section of this report entitled "Data Processing Considerations" for details of computer processing steps mentioned herein.

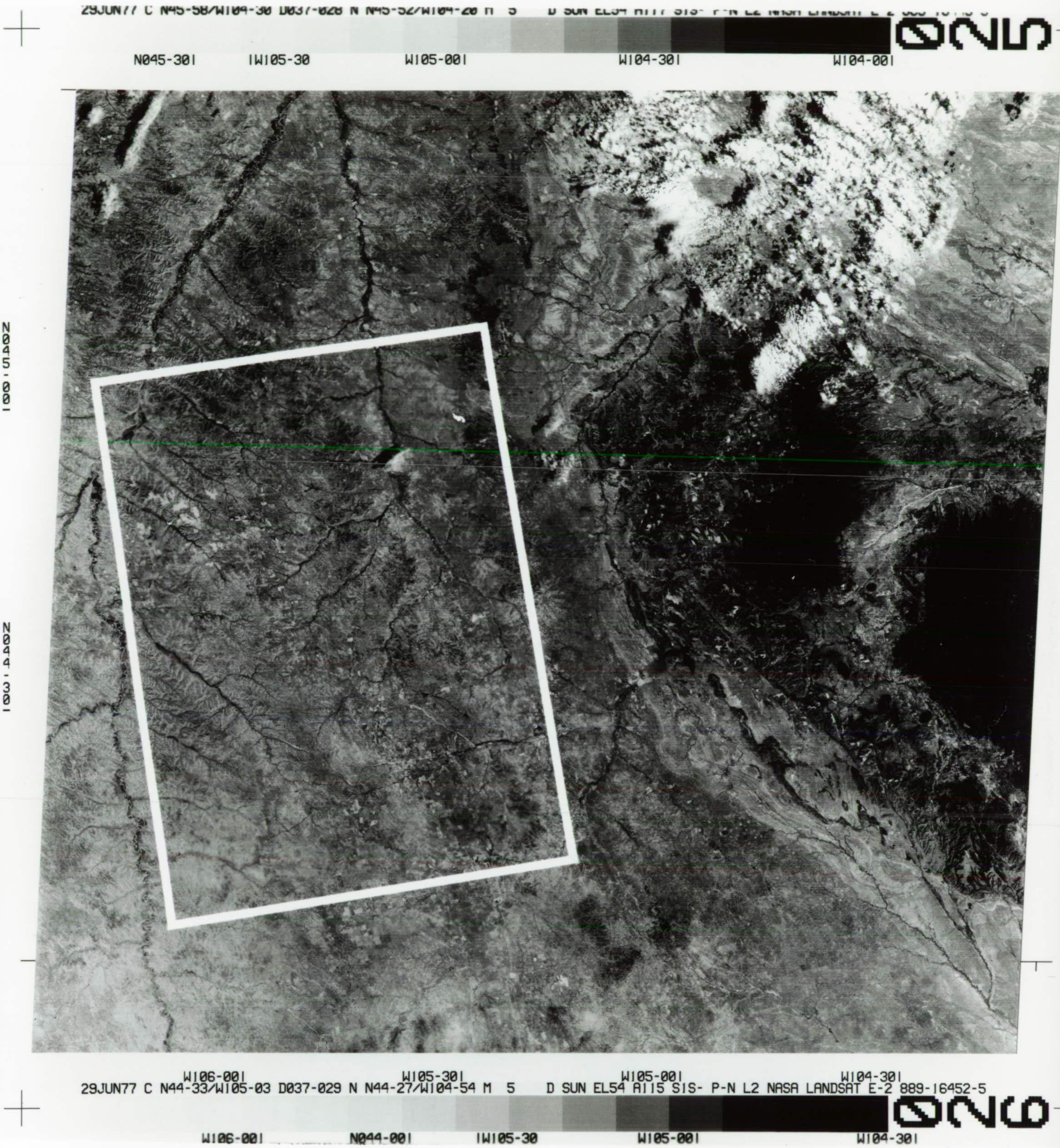

Figure 4 1:1,000,000 Scale Black & White Print of Landsat Channel 5 for Scene 2889-16452 (29 June 1977), with the Study Area Outlined selected for use in the study. The training sample sites were plotted on 1:24,000 scale topographic map sheets, and this information was used in the training sample selection procedure conducted on the IDS. Resulting training sample statistics were analyzed, and a classification was conducted based on 47 spectrally separable classes. These 47 classes can be consolidated into 8 land cover classes representing broad cover types of: (1) water, (2) coal, (3) exposed/piled overburden, (4) residential areas, (5) cultivated fields (including growing plants, stubble and bare soil), (6) range land, (7) reseeded range land, and (8) forested land. These broad categories reflect major land cover types in the study area.

In an attempt to determine the accuracy of the classification, in terms of some standardized scheme, it was decided to compare the classified Landsat data to the U.S. Geological Survey's "Land Use and Land Cover Classification System for Use with Remote Sensor Data", (Anderson, J. R., et.al., 1976).

The classifications produced by this system are in the form of transparent map overlays (at various scales), on which land use and land cover classifications have been delineated. Each such land use or land cover delineation has been assigned a unique land use classification number (e.g. 75 represents strip mines, gravel pits, and quarries). The USGS is currently digitizing these overlays, but at the time of this study, the Gillette 1:250,000 scale topographic map overlay (containing the study area) had not been digitized.

Comparison in this study were made with the USGS Level II classifications, since, as Anderson points out, "...Level II will probably be most appropriate for state-wide and interstate regional land use and land cover compilation and mapping..." (Anderson, J. R., et. al., 1976) and the classification descriptions match those land cover types of interest in this study.

The overlay sheet containing the USGS classification delineations was then positioned on an X-Y digitizer and initialization of the digitizer was then conducted. This procedure permits the operator to digitize points by their true UTM coordinates, thus facilitating the placement of such data into the UTM referenced DATBAS.

Sample points were located on the map overlay on a ½" grid, with the grid orientation determined in a random fashion. This systematic sample with random start produced 816 points within the sample area, which were then labeled with their corresponding USGS Level II classification number. These points were tabulated by USGS classification, and several classes (primarily residential (USGS Class 11), Commercial and Services (USGS Class 12), Industrial (USGS Class 13), Transportation, Communication, and Utilities (USGS Class 14), other Urban or Built up land (USGS Class 17), and Reservoirs (USGS Class 53)) were not represented in the sample. This situation was corrected by resampling the map overlay in a stratified fashion, sampling only those classes which were not represented in the initial systematic sample.

The resulting 842 total sample points were digitized and placed into the DATBAS. The classified Landsat data was then geographically referenced and was placed into the DATBAS in such a manner that the resulting DATBAS "cell" was approximately 40 acres in size (400 m to a side). This cell size is thought to be small enough to be of utility to EPA/EMSL for purposes of a regional applications demonstration.

A print of the contents of the DATBAS was then used to match each of the sample points (with known USGS classification) to the corresponding Landsat classification DATBAS cell. Since the Landsat classes were developed from training samples of known land cover composition, each Landsat class could be related to a particular USGS class. It was therefore a simple task to tabulate those Landsat cells that were either in agreement or disagreement with the USGS classification for the sample points.

Based on the results of this comparison, it was decided to field check a portion of those points that were found to be in disagreement, since the USGS map overlay was prepared by photo interpreting aerial photography (dated 1974-76) of a small scale, and since errors in either system could produce "false" disagreements to occur. Areas found to be in agreement were not field checked, as it was felt that when the two systems were in agreement, the chance that both were in error was quite remote.

The purpose of this sample was to obtain a reliable estimate of the accuracy of the classified Landsat data. This would be accomplished by using the results of the field check to estimate the actual number of points thought to be in error but in fact correctly classified, and combining this figure with that of the points found to be in agreement in the original sample.

A ten percent sample of those points in disagreement was taken. This sample was structured in such a way that allocation of field check sites was in proportion to the total number of Landsat points found to be in disagreement with a particular USGS classification. Points selected were plotted on 1:24,000 scale topographic map sheets, and in July, 1978, a team was sent to conduct field checks for those sites.

2. Integration of Auxiliary Data

The six types of auxiliary data to be used in the rehabilitation potential estimation procedure were:

a. Aspect (slope orientation)
b. Elevation
c. Rainfall
d. Slope
e. Soil type or series
f. Temperature

These factors, when considered jointly for an area of the surface of the earth, serve to define environmental niches in which plant species will be expected to successfully establish themselves and maintain growth after rehabilitation activities are discontinued. For the purposes of this demonstration, the above listed factors were considered of an adequate number, and it is recognized by the author that these six factors by no means encompass all of the environmental factors which could have been included in this demonstration. The rehabilitation potential estimation procedure has been developed to allow for the inclusion of additional (up to 18 total) factors should an investigator wish.

Not only do environmental factors greatly affect the final success or failure of rehabilitation, but they also define, to a large degree, the type of vegetation that might be found growing (in a natural state) in a specific area. Thus, by knowing the type of vegetation that is presently growing in an area, an insight is gained as to the vegetation type that might prove to be the best suited for rehabilitation efforts for a specific combination of environmental factors found in the area, since that vegetation type has already adapted to the environment. This is the basic assumption used in the rehabilitation potential estimation procedure developed.

Of course, disturbances to the land might alter water tables, soil types, surface temperature, etc. to such an extent that this assumption no longer holds, but at the present time, an insufficient amount of literature exists that presents research results dealing with plant species suitable for rehabilitation of disturbed lands in the study area, and therefore concrete conclusions along these lines cannot be drawn. Should research produce evidence that there are plant species more suitable for use in rehabilitation than those found growing naturally in the area, this can be incorporated into the scheme.

Auxiliary data was obtained from the sources listed in Table 2. Soils and rainfall data were digitized from map sheets and placed directly into the DATBAS (Figures 5 and 6). The soils data was digitized on a 400 m grid, with the predominant soil type serving to identify the entire 400 m cell. Original scale of the soils maps was 1:63,360 (1" = 1 mile), and six such sheets were required to completely cover the 1^{o} by 1^{o} study area. Total time required to digitize these sheets was approximately 260 manhours.

The rainfall data was alternately digitized from a 1:1,000,000 scale map using a polygon approach, which required two manhours to complete. In the polygon approach, the boundaries for a particular delineation (in this case, areas of equal annual rainfall) are digitized one at a time, and an entire polygon is placed into the data base at a desired scale. This method of digitizing has the advantage that once the data has been digitized and placed on a CCT, it can be read into a DATBAS of any cell size. Thus, once the data is on a CCT, it is not necessary to go back and redigitize an area (as in the case of the cell-by-cell approach) to compensate for a change in DATBAS cell size. This saves a considerable amount of time, should additional DATBAS applications requiring different cell sizes be required.

Table 2 - SOURCES FOR THE AUXILIARY DATA USED IN THE INVESTIGATION OF REGIONAL APPLICATION OF LANDSAT DATA

Data	Format	Source
Soil Type	1:63,360 map	Soil Conservation Service
Rainfall	1:1,000,000 map	Packer, Paul E. 1974 Rehabilitation Potentials & Limitations of Surface-Mined Lands in the Northern Great Plains, USDA Forest Service Report, INT-14 Ogden, Utah (Figure 1)
Elevation	CCT	National Cartographic Information Center
Aspect		Derived from Elevation Data
Slope		Derived from Elevation Data

Figure 5 Soils and Rainfall Data as Contained in the Data Base

Figure 6 Color Code for the Soils Data Portion of Figure 5

The terrain related topographic factors of aspect, slope, and elevation, were derived from data purchased from the National Cartographic Information Center Topographic Division (USDI Geological Survey) in Reston, Virginia (22092). The data itself is located on 9 track computer compatible magnetic tapes in either 800 or 1600 byte-per-inch density and is composed of point elevations digitized from a 1:250,000 scale topographic map sheet on a .01" grid (which represents approximately 200 ft. on the ground). Each 1:250,000 scale map (1° latitude by 2° longitude) is divided into two 1° by 1° areas, such that two data blocks are required for one 1:250,000 scale map.

Costs at the time of this writing are $15 for the tape itself plus $6 for each 1° by 1° area required. Specific instructions and additional technical information can be found in NCIC Digital Terrain Tapes Users Guide, available from NCIC USGS Reston, Virginia.

In order to obtain elevation, slope, and aspect parameters from the original elevation data on the NCIC CCT's several computer programs developed at NASA were implemented. These programs reformat the original data and compute the aspect and slope for the desired output (DATBAS) cell size from the elevation data.

Output options included in the software permit the user to specify: (1) output cell size, (2) output elevation ranges (up to 256 mutually exclusive ranges may be specified), (3) aspect ranges (up to 8 mutually exclusive ranges), and (4) slope ranges (up to 20 mutually exclusive ranges). Computer run time required to produce results which can be integrated into the DATBAS depends on the output options selected, as well as the number of ranges for each output parameter used. For the data used in this study, 12 elevation ranges (200' increments), 8 aspect ranges, and 8 slope ranges were used. (Figures 7, 8, and 9, respectively).

Temperature data, of the type needed, was not available from a sufficient number of weather stations in the study area, and thus meaningful temperature delineations could not be derived. This data type has been dropped from further considerations in this demonstration project, but it could be added to the procedure in future projects if available in other areas. It should be noted that temperature generally correlates with elevation.

E. Results and Analysis

1. Accuracy of the Data

Accuracy figures calculated for the computer implemented supervised land cover classification of the Landsat MSS data (Figure 10 and 11) are presented in Table 3. As can be seen, the composite value is 86%, with values of individual land cover classes ranging from 50 to 100%. The low values were not unexpected, as training samples were not specifically selected to represent these land cover types at the outset of this task. A few training samples did match the description of these land cover classes, but this can be attributed to coincidence rather than a conscious effort. This fact, coupled with the low

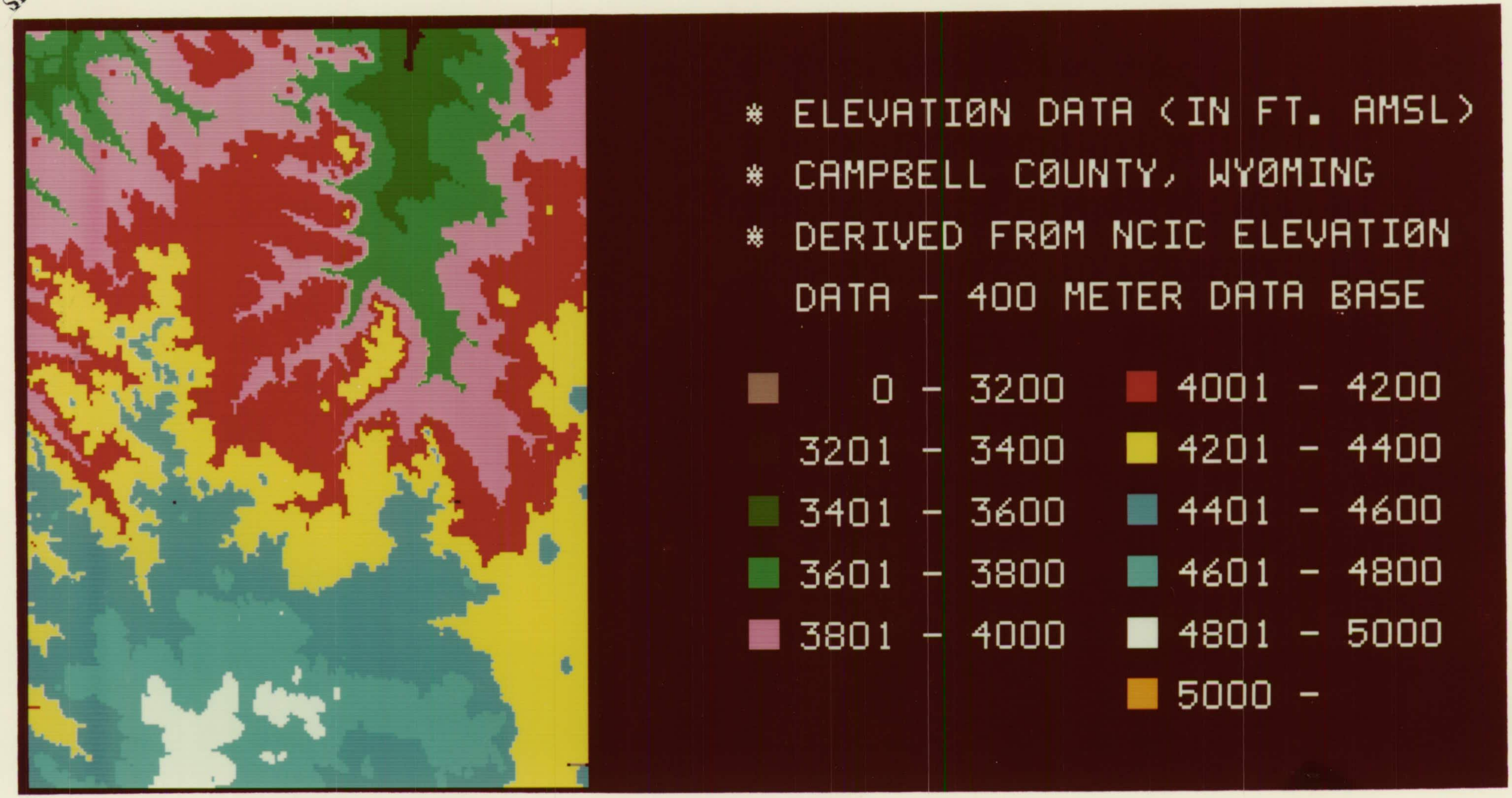

Figure 7 Elevation Data (derived from NCIC elevation tapes) as Contained in the Data Base

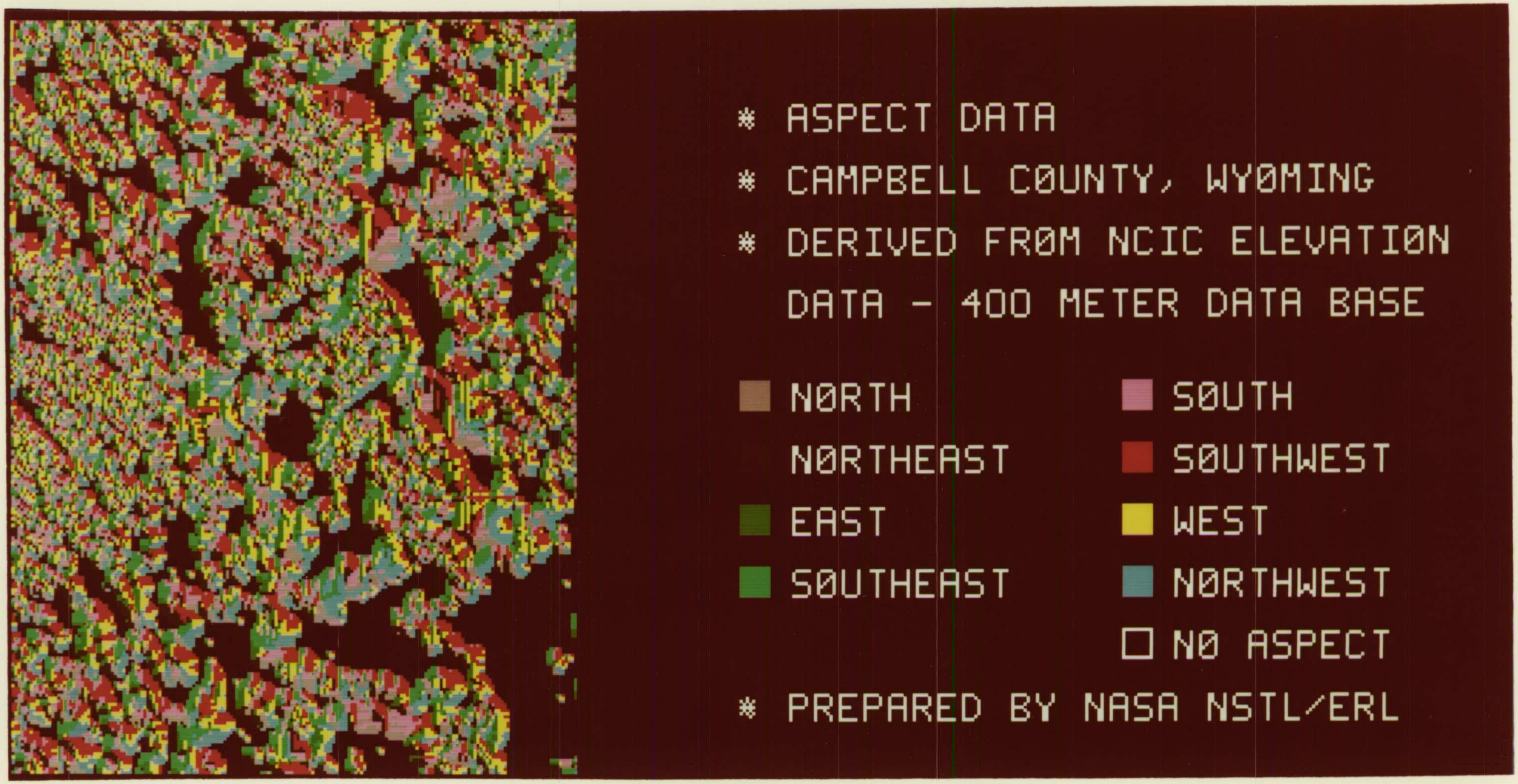

Figure 8 Aspect Data (derived from NCIC elevation tapes) as Contained in the Data Base

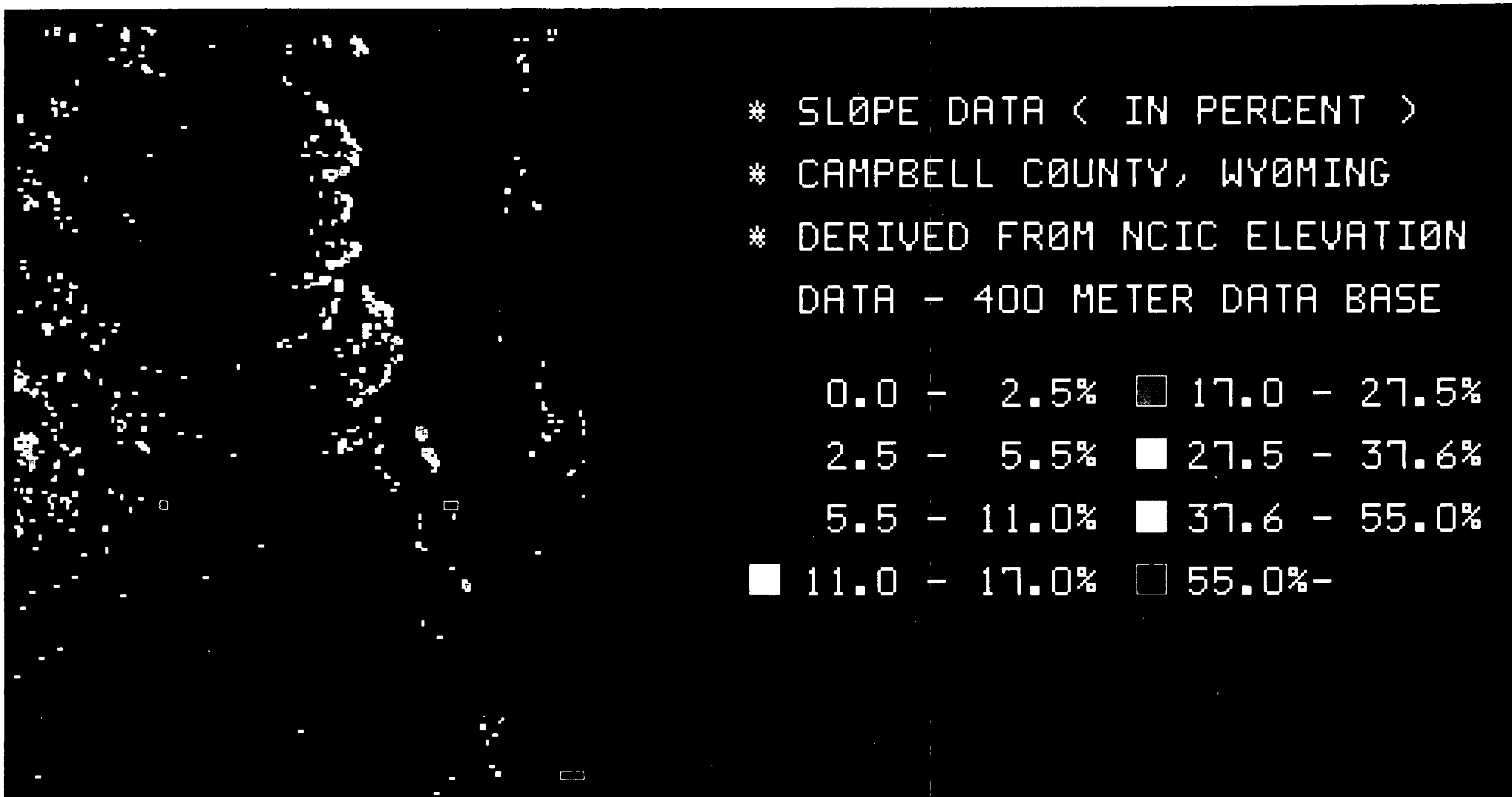

Figure 9 Slope Data (derived from NCIC elevation tapes) as Contained in the Data Base

Table 3 - ACCURACY FOR THE COMPUTER IMPLEMENTED SUPERVISED LAND COVER CLASSIFICATION OF THE LANDSAT MSS DATA USED IN THIS STUDY

USGS Land Cover Classification Number	USGS Level II Land Cover Category	Accuracy of Classification of Landsat MSS Data
11	Residential	67%
12	Commercial & Services	100%
13	Industrial	100%
14	Transportation, Communication, and Utilities	100%
17	Other Urban or Built Up Land	50%
21	Cropland & Pasture	78%
24	Other Agricultural Land	100%
31	Herbaceous Rangeland	77%
32	Shrub & Brush Rangeland	80%
33	Mixed Rangeland	75%
41	Deciduous Forest Land	100%
42	Evergreen Forest Land	79%
53	Reservoirs	100%
62	Unforested Wetlands	50%
71	Dry Salt Flats	100%
75	Strip Mines, Quarries, Gravel Pits	100%
76	Transitional Areas	100%
	Overall	86%

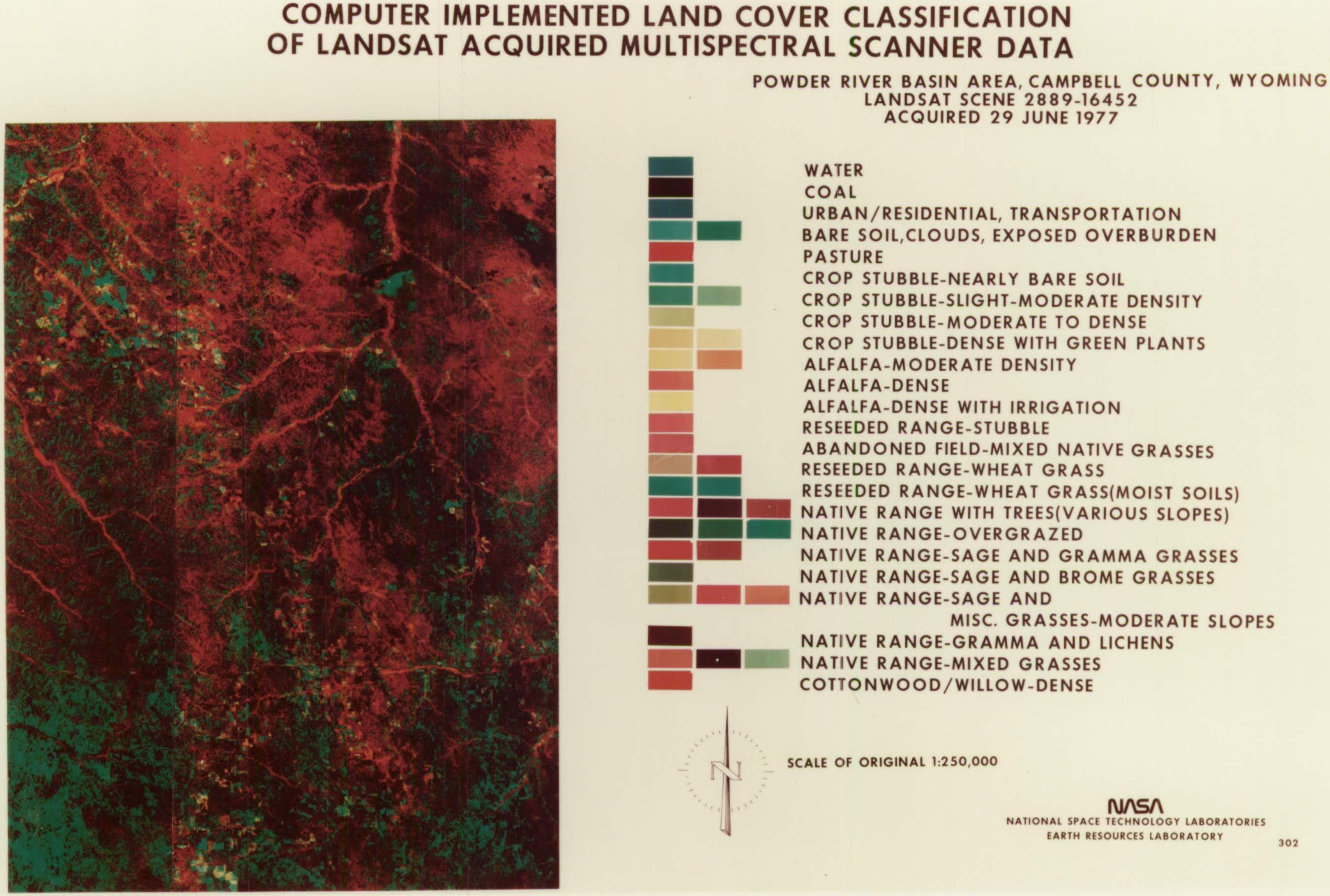

Figure 10 Computer Implemented Land Cover Classifications of Landsat Acquired Multispectral Scanner Data

Figure 11 A Portion of the Computer Implemented Land Cover Classifications of Landsat Acquired Multispectral Scanner Data Registered to a 1:250,000 Scale Topographic Map Sheet Overlay

accuracy figures associated with the two land cover types in question, serve to emphasize the need for complete representation by training samples of all land cover types of interest in the study area. Had training samples been developed for "other urban and built up land" and "non forested wetlands", accuracy figures associated with these two classes would most probably have increased. Figure 11 demonstrates the geometric fidelity of the classified Landsat data contained in the DATBAS to be the original 1:250,000 scale topographic sheet. The class colors are the same as those in Figure 10. Acreage figures for the classes represented in Figure 10 are given in Table 4.

No effort was made to verify the accuracy of the ancillary data used in this study. The sources of the data have quoted accuracies for their respective products, and these accuracies effectively hold for the data used in the procedure demonstration reported on herein.

2. <u>Operational Implementation of the DATBAS</u>

There exist numerous ways that the DATBAS can be used to deal with specific applications requiring investigator-data interaction. For the purpose of demonstrating the utility of the DATBAS in this report, the following hypothetical example has been constructed. It should be noted that this is not the only way in which a solution could have been reached, and as an investigator becomes familiar with the DATBAS, other options would become apparent.

Table 4 - ACREAGE COMPUTATIONS FOR THE LANDSAT CLASSIFICATION PRESENTED AS FIGURE 10

CLASS	ACREAGE
Water	79
Coal	2728
Urban/Residential, Transportation	13838
Bare Soil, Clouds, Exposed Overburden	25936
Pasture	3914
Crop Stubble - Nearly Bare Soil	14273
Crop Stubble - Slight-Moderate Density	16526
Crop Stubble - Moderate to Dense	23682
Crop Stubble - Dense with Green Plants	19373
Alfalfa - Moderate Density	20164
Alfalfa - Dense	633
Alfalfa - Dense with Irrigation	120
Reseeded Range - Stubble	15498
Abandoned Field - Mixed Native Grasses	149290
Reseeded Range - Wheat Grass	139248
Reseeded Range - Wheat Grass (moist soils)	25976
Native Range with Trees (various slopes)	344958
Native Range - Overgrazed	646742
Native Range - Sage and Gramma Grasses	283320
Native Range - Sage and Brome Grasses	161863
Native Range - Sage and Misc. Grasses (moderate slopes)	163998
Native Range - Gramma and Lichens	50686
Native Range - Mixed Grasses	43214
Cottonwood/Willow - Dense	32301
TOTAL	2198360

Consider the following problem. A coal company has submitted a request for a permit to surface mine coal in the Campbell County area of Northeastern Wyoming. In the request, it is specified that rehabilitation will be conducted, using plant species A, B, C, and D to establish a vegetative ground cover. No irrigation will be conducted on the rehabilitated areas. In addition, due to limitations associated with overburden depth, slopes, etc., mining will be restricted to Ulm loam soils occuring on a SE aspect. The entire proposed mine site is to be situated in an elevation range of 4001-4200 ft. above mean sea level. The question is whether or not the plant species mentioned in the request are suitable for use in rehabilitating the area mentioned in the permit, given all of the ecological restrictions mentioned in the request.

Following the outline for operational implementation of the rehabilitation potential estimation procedure mentioned earlier in this section, the first step that is necessary in order to answer the question at hand is to isolate the area in question from all other areas. For the purposes of this demonstration, the boundaries were arbitrarily selected for a somewhat large area contained in the DATBAS. This was done for ease of presentation and reader comprehension, and should not be misconstrued as a limitation of the applications software. Actual boundaries can be located (computed) through the implementation of a specific series of command instructions, or the boundaries can be digitized and placed into a parameter in the DATBAS.

After the boundaries have been located, the next step involves examination of the appropriate DATBAS parameters and a comparison of each parameter, either singly or as they co-exist, with a set of values. For this example, the comparison was conducted as follows.

In the original request, it was specified that no irrigation was to be conducted. Past experience (hypothetical) indicates that an annual minimum of 16 inches of rainfall is required to support vegetation in the area. Thus, the 16 inch annual rainfall minimum becomes a "limiting factor", in as much as irrigation would be required in areas with less than that amount if revegetation efforts are to be successful. And since the hypothetical permit request makes no allowance for irrigation, mining would have to be restricted to those areas where the vegetation would have a high probability of receiving sufficient moisture.

The upper left portion of Figure 12 depicts the annual rainfall for the hypothetical proposed lease area. Yellow in this portion of the figure represents an annual rainfall of 14-16 inches, blue represents 16 inches or greater per year. So, taking into consideration the 16 inch minimum annual rainfall constraint mentioned above, the blue area in the rainfall data would be the only area suitable for mining from the standpoint of availability of sufficient moisture.

In addition to the moisture constraint, the mining was to be limited to ulm loam soils occuring on SE aspects. In order to locate these areas, the soils data was examined first for the presence of the required rainfall. In this respect, the rainfall data was used as a "mask" to selectively restrict subsequent analysis to soil types meeting the rainfall constraint. The upper right portion of Figure 12 shows all soils (irrespective of aspect) which occur in the proposed lease areas which do, in fact, have 16 inches or greater annual rainfall associated with them. Ulm loam soils appear orange in this portion of the figure.

Aspect is included in the analysis by using the ulm loam soil type as a "mask"; the result being a display of the aspect of each 400 m DATBAS cell which is of the ulm loam soil type, and which meets the moisture constraint (center right portion of Figure 12). SE aspect slopes have been colored olive green in this display.

Elevation considerations mentioned in the permit request (4001-4200 ft. a.m.s.l.) can now be included, by simply examining all elevations of each DATBAS cell meeting the previously mentioned constraints. The SE aspect is used as a mask to display elevation data. The elevation range of 4001-4200 ft. a.m.s.l. appears as a dark green in the center portion of Figure 12.

With this last "masking", all ecological constraints of concern have been examined. It is now a simple task to use the 4001-4200 ft. a.m.s.l. elevation range as a "mask", and display the land cover types (derived from Landsat MSS data) that can be found occurring on areas which meet all of the constraints with which this example has dealt. Such is the nature of the lower left portion of Figure 12.

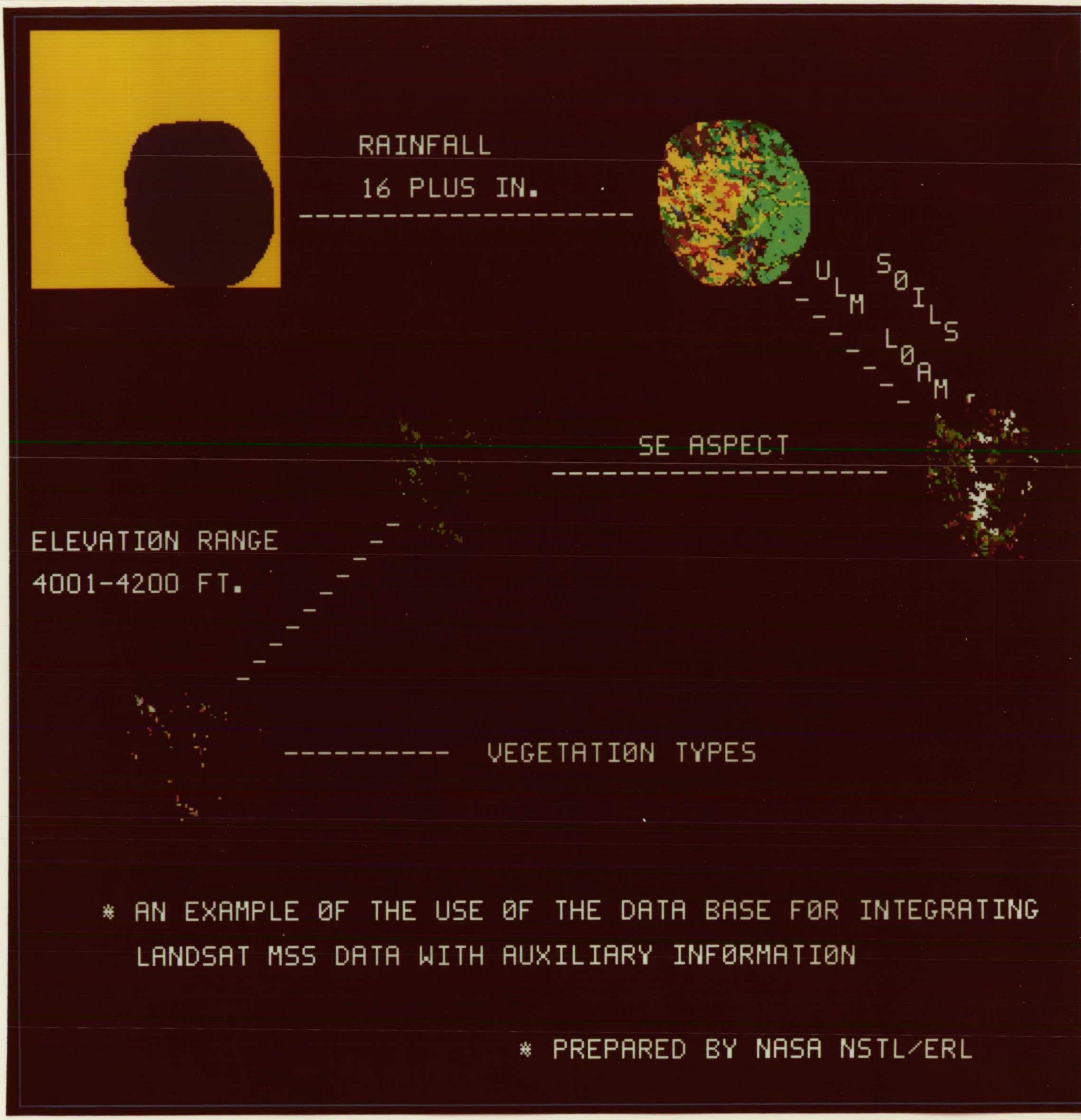

Figure 12 An Example of the Use of the Data Base for Integrating Landsat MSS Data with Auxiliary Information

The vegetation types associated with each color on this final display can then be compared to those mentioned in the permit request. Should those species proposed for use in rehabilitation efforts be included in the species determined to be growing on the sites, no problems exist. However, should one or more proposed species be excluded from the list of species found "naturally" growing in the area, a revision of the permit request may be in order.

As each successive "masking" is made, the resulting display is augmented by an acreage tabulation by class. As an example, when the rainfall data is used as a mask, the resulting soils data display is accompanied by an acreage tabulation, indicating the acreages of each soil type with 16 inches or greater rainfall (annual) contained in the study area. This automatic acreage calculation feature has numerous uses, particularly if acreage figures weigh heavily in the problem at hand.

F. Conclusions

This hypothetical example served to illustrate but one way in which the procedure could have been executed. Other factors (such as slope length, slope steepness, mean annual temperatures, monthly rainfall, newly released research results, etc.) could have been included, different problems could have been dealt with, etc. The major point to be emphasized here is that the procedure is highly flexible and thus adaptable to numerous applications not restricted to rehabilitation oriented problems.

V. SITE SPECIFIC APPLICATIONS BASED ON THE USE OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA FOR MONITORING SURFACE MINED LANDS

A. Description of the Task

The objective of this task was to develop an automated data processing procedure which would permit EPA/EMSL to classify aircraft acquired MSS data without the need for ground truth information gathering. Such a procedure would permit the "real time" processing of the multitude of aircraft MSS data which would be required to monitor the many surface mining operations currently in production, or planned for future development. On site visitation would be restricted to post data classification trips deemed necessary through analysis of the classification produced. Software needed in order to implement such a procedure was developed, tested, and refined by NASA/NSTL, and transferred to EPA/EMSL at the conclusion of Phase II activities.

B. Scope of the Task

Demonstration of the task was conducted using three on-line coal strip mines, for which aircraft acquired MSS data had been collected in 1975 (Phase I activity). Table 5 presents details of these mine sites.

Table 5 - ON-LINE COAL STRIP MINES USED IN THE SITE SPECIFIC TASK DEALING WITH AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA

Coal Strip Mine	Location	1974 Production (Tons)
Belle Ayr	Campbell Co, WY	3,301,472
Decker No. 1	Big Horn Co, MT	7,000,000
Navajo	San Juan Co, NM	6,955,000

These MSS data sets for these three mines were chosen for the following reasons:

(1) Aircraft acquired MSS data of the mines, collected in 1975, was on-hand
(2) Detailed ground truth information (1975) existed for these mines, facilitating checkout of the procedure developed
(3) A diversity of environmental conditions is represented, providing a means by which the procedure can be evaluated for sensitivity to changing site conditions
(4) Aircraft acquired MSS data was scheduled to be collected for these mine sites in 1977 (by EPA/EMSL), which would function as a "test case" of the procedure developed.

C. Technical Approach

In order to permit the classification of aircraft acquired MSS data without the need for pre-classification ground truth information gathering, software was developed and refined which would search the data for areas of spectrally uniformity, group those areas which were not spectrally separable, and output the statistics for the resulting spectrally separable classes. The program thus developed was termed "SEARCH".[1]

Operational implementation of the procedure involves the following steps:

1. The investigator specifies the data set to be used, along with several statistical parameters necessary to implement the software
2. Execution of the software SEARCH
3. Use of the resulting statistics as input to a classification program
4. Analysis of the classification produced.

D. Procedures and Methodology

Past experience in dealing with aircraft acquired MSS data of on-line coal strip mines has shown that results can be obtained which meet EPA/EMSL requirements for products, based entirely on the

[1]The reader is referred to the section of this report entitled "Data Processing Considerations" for details of computer processing steps mentioned herein.

use of a four channel subset of the original aircraft MSS data set (Anderson, et al., 1978). The use of such a subset has several important considerations, among which are:

(1) Data processing costs are reduced on a per unit area basis

(2) Software developed for four channel Landsat MSS data can be applied to the four channel aircraft MSS data subset (e.g. SEARCH).

1. Aircraft Data Processing

The manner in which the four channel subset is chosen varies from one investigator to another, but for this study, channel selection was based on the output from the sun angle-scan angle correction program. The output consists of a graph, displayed on the IDS, which plots average spectral response by element for each channel used as input. (This average spectral response is calculated from the spectral responses of each element over all scan lines.)

The graphic display tends to indicate the correlation that exists between channels for a particular data set, in as much as channels which are highly correlated will have more or less identical graphs, perhaps the only significant difference being an almost constant offset distance.

Using such a display, the correlated channels were systematically eliminated, in all cases favoring the channel with the highest (most near to red) wavelength. If this procedure did not reduce the total number of channels to four, channels were eliminated which had fairly "flat" graphs, indicating no particularly high sensitivity to spectral change in the data. This in all cases reduced the number of channels to the desired four.

The data for the three mine sites previously mentioned were analyzed independently, and four channels were subsequently selected for each data set. It is of interest to note that the channels selected were identical for all mine sites (channels 2, 7, 8, and 9). The four channels thus selected were used as inputs to SEARCH.

Since this was the first instance in which SEARCH was to be applied to aircraft acquired MSS data, it was necessary to establish threshold values which would maximize the utility of the resulting classification, while at the same time kept processing turn around time as minimal as possible. In order to accomplish this task, three SEARCH/MAXL4 runs were made for each of the three 1975 data sets. These runs were structured in such a manner that the only influence on the classification would result from altering the user specified statistical input parameters of SEARCH.

The first SEARCH/MAXL4 run for each data set was established in a manner identical to that commonly used for processing Landsat data in Southeastern Louisiana (Table 6). This run served as a base from which alterations to the values of the input parameters were made.

Table 6 - INPUT PARAMETERS USED FOR SEARCH/MAXL4 PROCESSING OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA

Parameter		Run 1	Run 2	Run 3
Standard Deviation	LO	0.5	0.5	*0.1
	HI	1.0	1.0	*1.5
Coefficient of Variation (in%)		5	5	5
Maximum Separability (Divergence) for Merger		5	*1	*1

* Indicates change from Run 1

The second and third runs were set up in an attempt to progressively increase the number of six by six blocks of data found acceptable, and hence force the software to evaluate more data. It was anticipated that the result of this action would be an increase in the accuracy of the classification made, since statistics used would have been developed from a significantly larger population of sample points and hence would be more representative of the data being classified.

The EPA/EMSL photo-interpreted aerial photography (collected concurrently with the MSS data) was used to define the minimum level of detail acceptable, since it was felt that the photo-interpretation was produced to currently acceptable EPA/EMSL levels of detail, and that the outcome of the classification of MSS data must contain at least as much detail as was in the photo-interpretation to be of real use to EPA/EMSL. It was therefore decided to use the photo-interpretation as a base, and to compare all classifications produced to that base.

The EPA/EMSL photo-interpreted aerial photography was thus used to establish randomly located check plots; approximately five such plots were used for each of the land cover types delineated on the EPA/EMSL photo-interpretation. The check plots were then located in the raw MSS data, each plot being three scan lines by three elements in size. The scan line-element coordinates for the corners of each plot were stored on mini-computer disc memory for later use.

After a classification was produced, a program was executed which tabulates the occurrence and acreages of classes within each of the check plots. The check plot tabulations were then arranged so that the program tabulations for check plots representing a singular land cover type were situated together. Analysis continued by examining the results of the acreage tabulations for the check plots, and matching the SEARCH developed spectrally separable classes to the EPA land use types found on the photo-interpretation. In certain instances (e.g. coal), this was an easy task. In other cases (e.g. vegetation) several SEARCH classes corresponded to a single EPA land cover type, and separation of SEARCH classes was somewhat involved.

Two factors weighted heavily in the determination of which classification run was the most acceptable. First, the detail of the classification was considered. This was necessary since the classifications were being compared to an acceptable base, and any detail over and above that contained in the base set would be unnecessary. Accuracy (agreement of the classification to actual land cover conditions) was also considered in this factor.

The second factor which was considered was that of overall turn around time. This factor is related to the first (i.e. classification detail), since longer turn around times are normally required for more detailed classifications. In terms of the objectives of this investigation, the most acceptable classification run would be the one which met the detail requirement, while at the same time requiring the least turn around time.

Analysis of the classifications produced for each mine site was carried out with the objectives of acceptable detail and turn around time in mind. In this respect, a check plot which was classified <u>correctly</u> as being 100% of one SEARCH class was the most desirable situation which could exist, since the plots were established in singular EPA land cover types. The greater the number of classes occurring in a check plot, the more undesirable it became.

The check plots were then grouped by using the percent composition of SEARCH classes found in each check plot. For instance, check plots composed entirely of one class were placed in one group, those composed of 90 percent of one class, and ten percent of another were placed into a second group, etc. The results were grouped for later use.

Computer turn around time for each computer run was also determined, and included such items as card deck preparation, and SEARCH and MAXL4 run times. These times were added together for each computer run made for each mine site and were also stored for later use.

Upon completion of analysis and the selection of the most desirable method for processing the data, the 1977 aircraft acquired MSS data was processed. This was done in an attempt to determine the overall effectiveness of the technique developed, since no aerial photography was used in conjunction with data processing of the 1977 MSS data. Processing was conducted in a manner identical to that described herein for the 1975 aircraft data, with the exception that only the most desirable method of executing SEARCH/MAXL4 was used.

E. Analysis and Results

Tabulations for each computer pass for the MSS data for each mine included in this study are presented in Table 7. Selection of the most desirable Run for Belle Ayr and Decker #1 mines is simplified somewhat by the fact that Run #1 (for both mines) resulted in the largest percentage of check plots classified entirely by one class. In addition for Belle Ayr, the total computer run time for Run #1 was less than that for either of the other two runs. Run #1 for Decker #1 was just one minute more than Run #2, and this fact is balanced by a two percent difference in percent of check plots classified entirely by one class. In this particular case,

Table 7 - TABULATED RESULTS FOR THE DATA ANALYZED FROM THREE COMPUTER RUNS OF AIRCRAFT ACQUIRED MSS DATA FOR THE MINES INCLUDED IN THE STUDY.

Mine	Run	Percent of Check Plots of "One Class"	# Classes	SEARCH Run Time (min)	MAXL4 Run Time (min)	Total Computer Run Time (min)	Total Data Processed (in Pixels)
Belle Ayr	1	35	28	5	40	45	44,640
	2	31	43	5	56	61	
	3	26	38	6	45	51	
Decker #1	1	41	25	5	44	49	156,000
	2	39	23	5	43	48	
	3	36	36	12	51	63	
Navajo	1	27	35	74	230	304	830,250
	2	24	35	71	224	295	
	3	27	36	86	192	278	

the "percentage of one class" figure was given more weight, and the result was that Run #1 was determined to be the most desirable for both Belle Ayr and Decker #1 mines.

Total computer run times for the Navajo data processing are somewhat misleading, in as much as the classification software is executed through card input (background). This type of job is given low priority with respect to processing done on the IDS (foreground). Thus, if computer time must be allocated between a background job and a foreground job, the foreground job will be given first priority to the time available. This means that should a foreground job be in progress at the same time that a background job is being executed, the background job must "wait" until computer processing time is available for use, and hence the total computer time required will be inflated by the "wait" time. There is no way to isolate this "wait" time at present, and thus take into account background-foreground interface problems.

One additional fact affects the total computer run time required, that being the amount of data to be processed. As the amount of data to be processed increased, the total amount of background "wait" time that could be expected also increases. This is not a linear relationship however, but is dependent on foreground activity.

The greatest difference in total run time for the three runs used for Navajo was 26 minutes. This occurred between Runs #1 and #3, which were the Runs with the greatest (and equal) "percent one class" figures. As can be seen in Table 7, the SEARCH run time for Run #1 was 12 minutes less than that for Run #3. This difference was expected, since the statistical parameters were relaxed for Run #3 and quite restrictive for Run #1.

It was therefore decided to use the SEARCH run time as an indicator of total run time, thus decreasing the influence of "wait" time encountered during the long classification runs. With this decision made, Run #1 was subsequently chosen as the most desirable for Navajo.

Even though the data were analyzed independently for the MSS data classifications for the three mine sites, Run #1 was chosen as the most desirable run for all mine sites. This is rather significant since the mines are situated in diverse geographic regions, flight line orientation was different (North-South for Navajo and Decker #1, East-West for Belle Ayr), and volume of data analyzed varied considerably (Navajo contained approximately 18.6 times the data of Belle Ayr). This outcome is very desirable, since it demonstrates the relative consistency of the software over a wide range of conditions.

Classified results obtained from Run #1 for Belle Ayr, Decker #1, and Navajo, along with their associated color tables, are included as Figures 13, 14, and 15 respectively. These figures correspond to the areas around each mine site that have been photo-interpreted by the EPA and for which aircraft acquired MSS data was processed. Results obtained agree very closely with

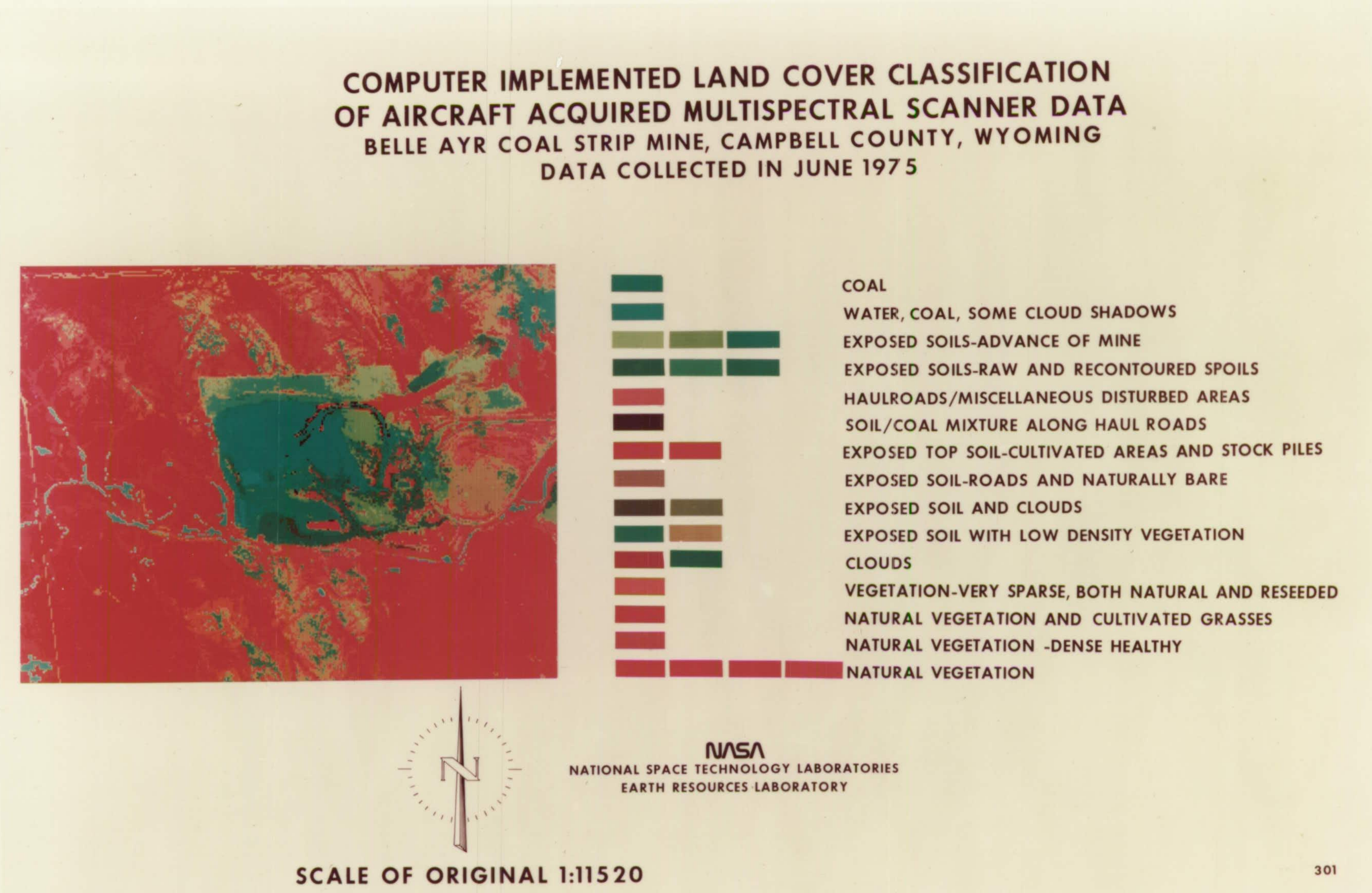

Figure 13 Computer Implemented Land Cover Classification of Aircraft Acquired Multispectral Scanner Data, Belle Ayr Coal Strip Mine, Campbell Co., WY (1975 data)

Table 8 - ACREAGES FOR THE COMPUTER IMPLEMENTED LAND COVER CLASSIFICATION OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA (1975) OF THE BELLE AYR COAL STRIP MINE

Class	Acreage
Coal	36
Water, Coal, Some Cloud Shadows	25
Exposed Soils-Advance of Mine	65
Exposed Soils-Raw & Recontoured Spoils	107
Haulroads/Miscellaneous Disturbed Areas	9
Soil/Coal Mixture along Haulroads	3
Exposed Top Soil-Cultivated Areas & Stock Piles	43
Exposed Soil-Roads & Naturally Bare	12
Exposed Soil & Clouds	12
Exposed Soil with Low Density Vegetation	125
Clouds	5
Vegetation-Very Sparse, both Natural & Reseeded	91
Natural Vegetation & Cultivated Grasses	47
Natural Vegetation-Dense Healthy	188
Natural Vegetation	375
TOTAL	1143 Acres ($1.89\ mi^2$)

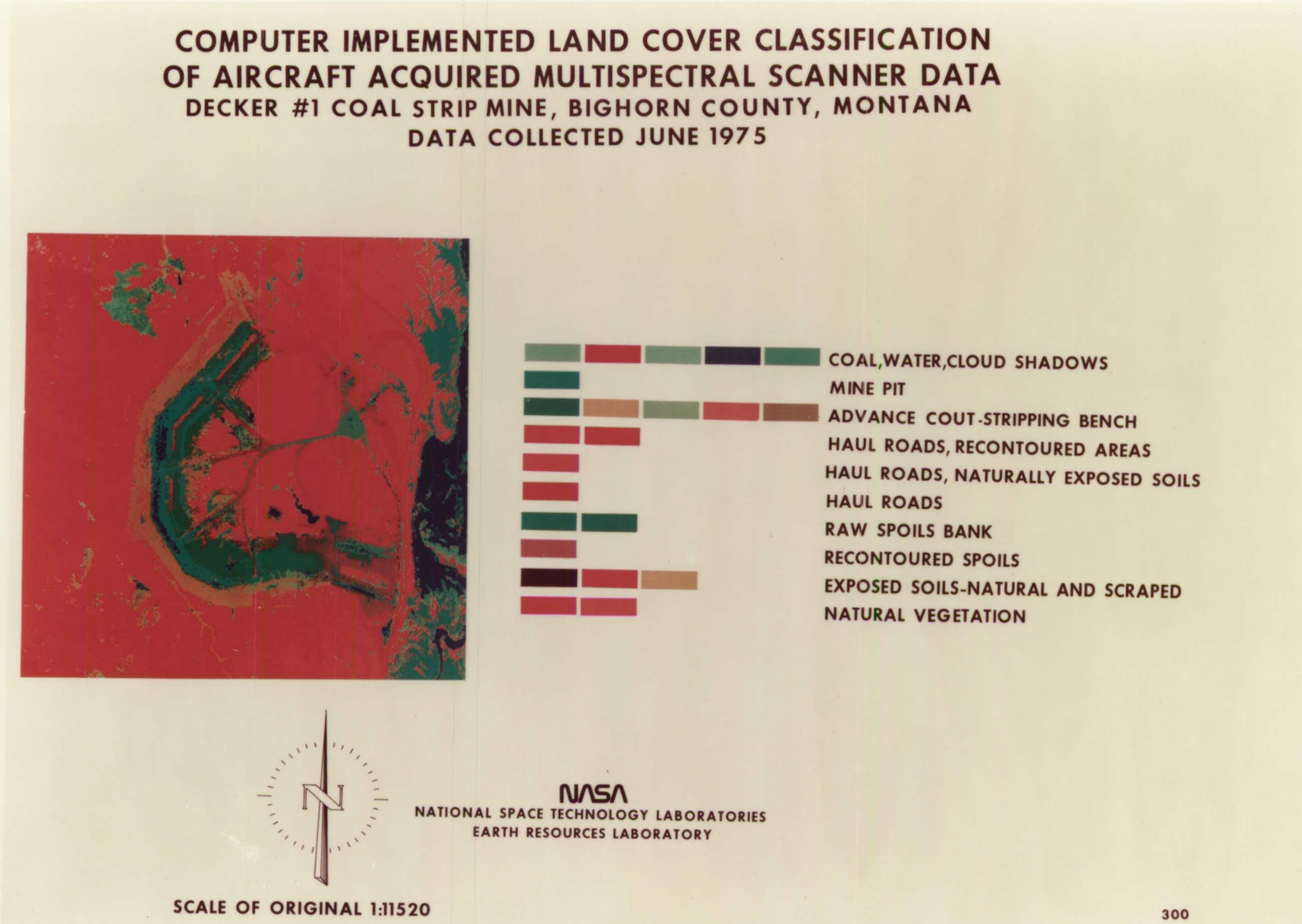

Figure 14 Computer Implemented Land Cover Classification of Aircraft Acquired Multispectral Scanner Data, Decker Coal Strip Mine, Bighorn Co., MT (1975 data)

Table 9 - ACREAGES FOR THE COMPUTER IMPLEMENTED LAND COVER CLASSIFICATION OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA (1975) FOR THE DECKER #1 COAL STRIP MINE

Class	Acreage
Coal, Water, Cloud Shadows	213
Mine Pit	185
Advance Cot-Stripping Bench	146
Haul Roads, Recontoured Areas	145
Haul Roads, Naturally Exposed Soils	271
Haul Roads	7
Raw Spoils Bank	117
Recontoured Spoils	102
Exposed Soils-Natural & Scraped	410
Natural Vegetation	1836
TOTAL	3432 Acres
	(5.36 mi^2)

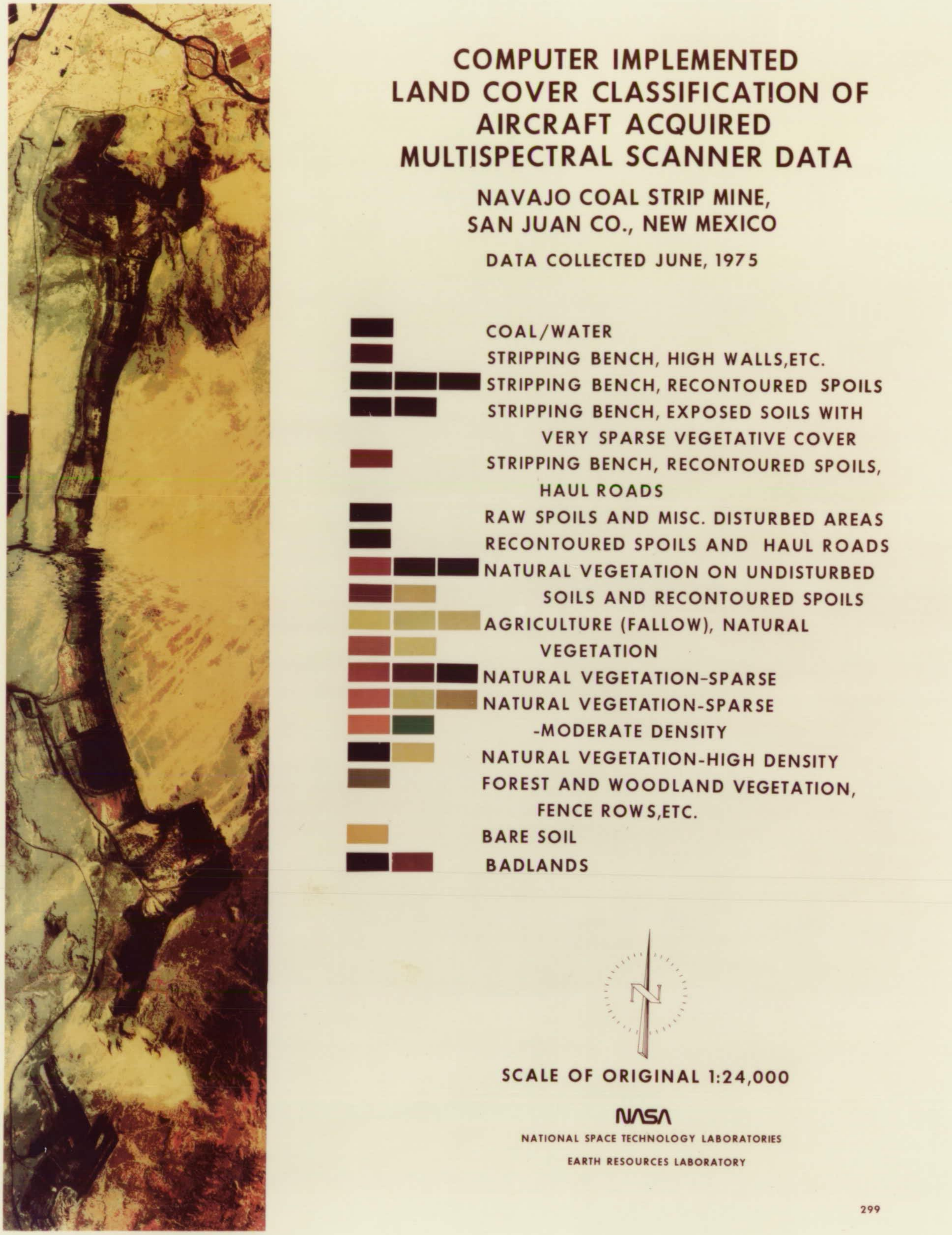

Figure 15 Computer Implemented Land Cover Classification of Aircraft Acquired Multispectral Scanner Data, Navajo Coal Strip Mine, San Juan Co., NM (1975 data)

Table 10 - ACREAGES FOR THE COMPUTER IMPLEMENTED LAND COVER CLASSIFICATION OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA (1975) FOR THE NAVAJO COAL STRIP MINE

Class	Acreage
Coal/Water	270
Stripping Bench, High Walls, Etc.	157
Stripping Bench, Recontoured Spoils	98
Stripping Bench, Exposed Soils with Very Sparse Vegetative Cover	455
Stripping Bench, Recontoured Spoils, Haul Roads	341
Raw Spoils & Misc. Disturbed Areas	45
Recontoured Spoils & Haul Roads	140
Natural Vegetation on Undisturbed Soils & Recontoured Spoils	1377
Agriculture (Fallow), Natural Vegetation	1251
Natural Vegetation-Sparse	5
Natural Vegetation-Spares-Moderate Density	3799
Natural Vegetation-High Density	369
Forest & Woodland Vegetation, Fence Rows, Etc.	319
Bare Soil	56
Badlands	899
TOTAL	9581 Acres ($14.97\ mi^2$)

EPA photo-interpretations. Acreages associated with the classes for each mine site can be found in Tables 8, 9, and 10.

Results obtained from processing the 1977 aircraft acquired MSS data are presented as Figures 16 (Belle Ayr), 17 (Decker #1), and 18 (Navajo). Acreages associated with these classifications are summarized in Tables 11, 12, and 13.

It is very interesting to note the change that has occurred in the two year period between data collection. For instance, the Northern Boundary of the Belle Ayr mine was situated in a more or less east-west direction in 1975 (Figure 13), but has progressed to the convaluated nature as shown for the 1977 data set (Figure 16). Similar change can be found in the other two mine sites examined in this study.

Analysis of the data stopped at this point, and as pointed out earlier in this report, EPA/EMSL will evaluate the resultant products and methodology with respect to their applicability to requirements as defined by EPA/EMSL. It should be noted however, that results obtained agree very well with land cover patterns delineated on aerial photography.

One final consideration warrents mention at this time. Applications software could be developed which would permit map comparison to be performed in an automated fashion using as inputs the two classifications produced for a particular mine site. This capability would permit the rapid screening of the data and pin point location of areas which might be in need of

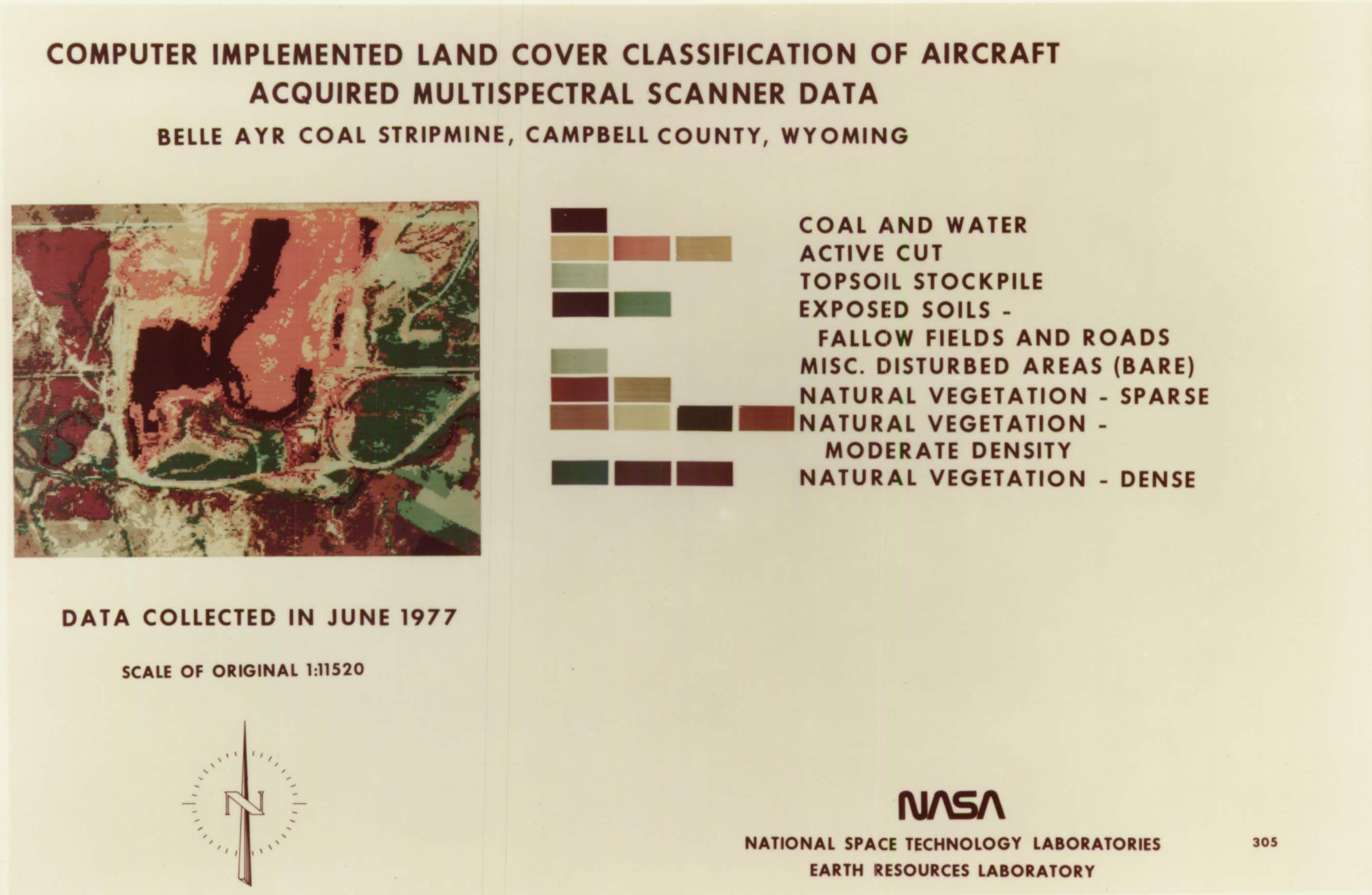

Figure 16 Computer Implemented Land Cover Classification of Aircraft Acquired Multispectral Scanner Data, Belle Ayr Coal Strip Mine, Campbell Co., WY (1977 data)

Table 11 - ACREAGES FOR COMPUTER IMPLEMENTED LAND COVER CLASSIFICATION OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA (1977) OF THE BELLE AYR COAL STRIP MINE, CAMPBELL COUNTY, WY

Class	Acreage
Coal and Water	38.6
Active Cut	137.5
Topsoil Stockpile	12.1
Exposed Soils - Fallow Fields & Roads	16.5
Misc. Disturbed Areas (Bare)	15.8
Natural Vegetation - Sparse	42.5
Natural Vegetation - Dense	106.2
TOTAL	445.9 Acres ($.07\ mi^2$)

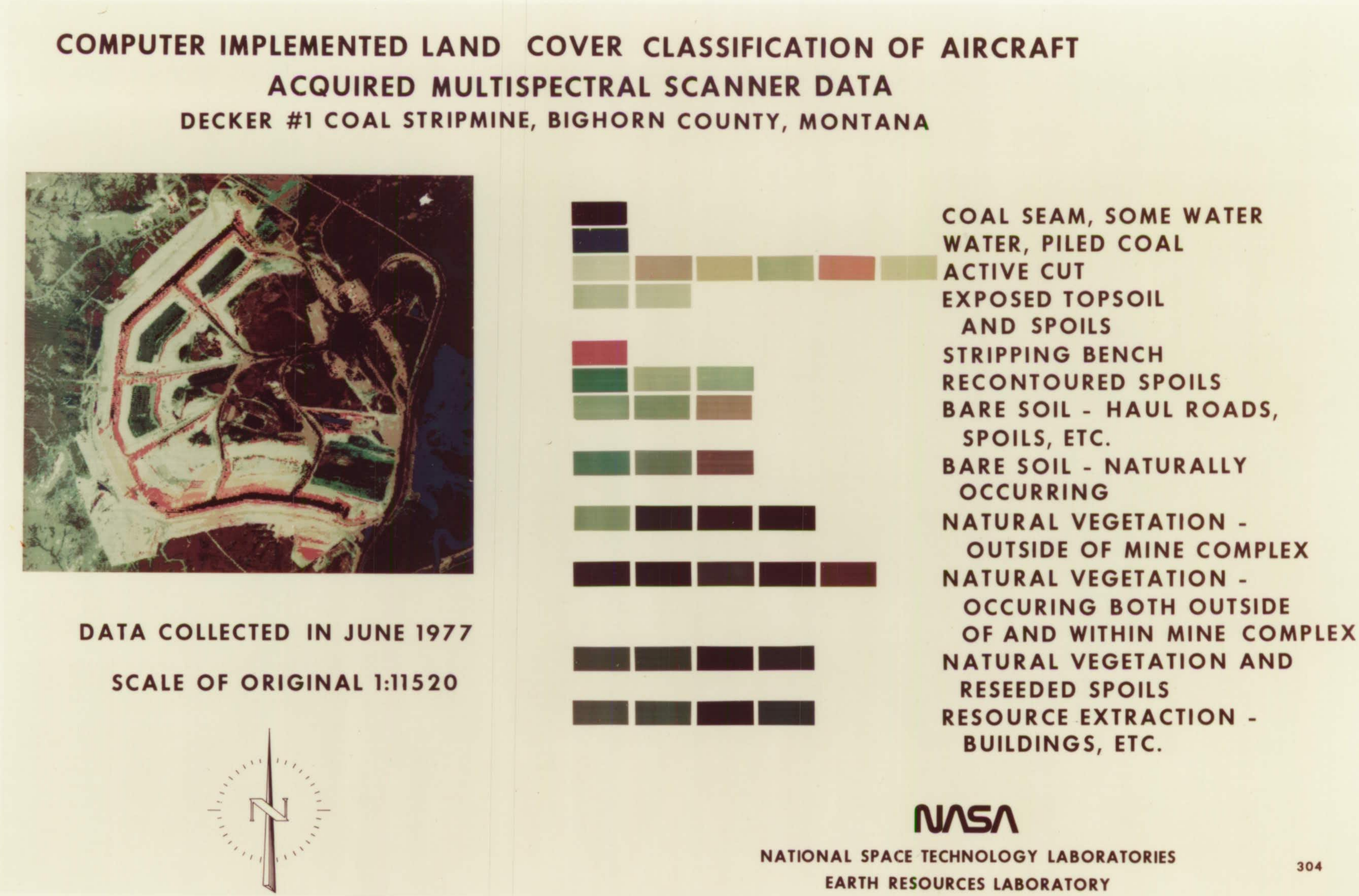

Figure 17 Computer Implemented Land Cover Classification of Aircraft Acquired Multispectral Scanner Data, Decker Coal Strip Mine, Bighorn Co., MT (1977 data)

Table 12 - ACREAGES FOR THE COMPUTER IMPLEMENTED LAND COVER CLASSIFICATION OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA (1977) OF DECKER #1 COAL STRIP MINE, BIGHORN COUNTY, MT

Class	Acreage
Coal Seam, Some Water	328.2
Water, Piled Coal	231.7
Active Cut	183
Exposed Topsoil and Spoils	187.6
Stripping Bench	72.3
Recontoured Spoils	127
Bare Soil - Haul Roads, Spoils, Etc.	954.8
Bare Soil - Naturally Occurring	529.2
Natural Vegetation - Outside of Mine Complex	218.9
Natural Vegetation - Occurring Both Outside of and Within Mine Complex	1164.4
Natural Vegetation and Reseeded Spoils	639.1
Resource Extraction - Buildings, Etc.	36.4
TOTAL	4636.2 Acres ($7.24\ mi^2$)

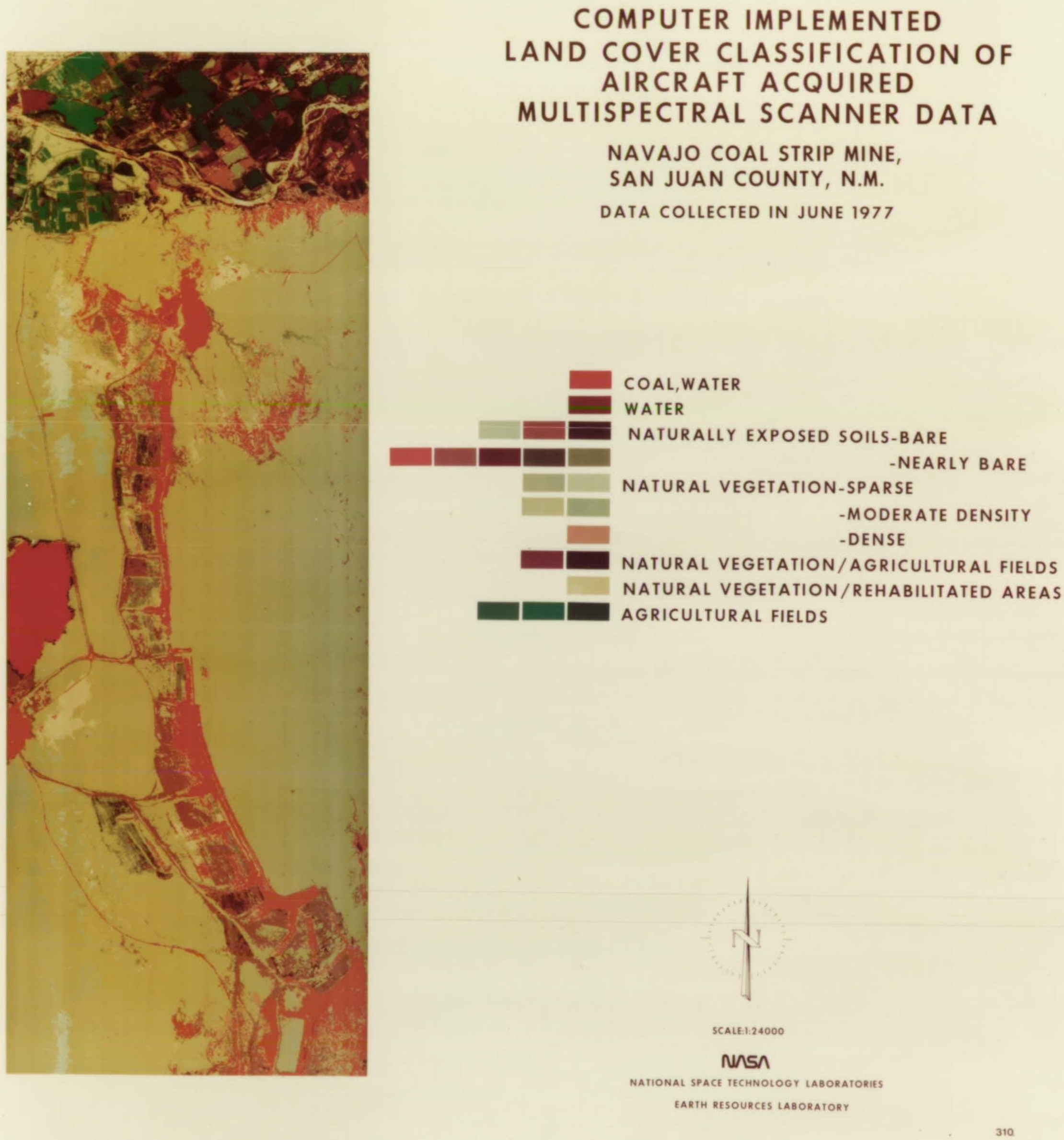

Figure 18 Computer Implemented Land Cover Classification of Aircraft Acquired Multispectral Scanner Data, Navajo Coal Strip Mine, San Juan Co., NM (1977 data)

Table 13 - ACREAGES FOR THE COMPUTER IMPLEMENTED LAND COVER CLASSIFICATION OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA (1977) OF THE NAVAJO COAL STRIP MINE

Class	Acreage
Coal, Water	10
Water	294.2
Naturally Exposed Soils - Bare	106.6
Naturally Exposed Soils - Nearly Bare	1444.7
Natural Vegetation - Sparse	415.1
Natural Vegetation - Moderately Dense	4374.0
Natural Vegetation - Dense	54
Natural Vegetation/Agricultural Fields	159.8
Natural Vegetation/Rehabilitated Areas	105.2
Agricultural Fields	412.1
TOTAL	7375.7 Acrea ($11.5\ mi^2$)

subsequent ground visitation by an earth scientist. Add to this the capability to rectify aircraft acquired MSS data to a map base (discussed in the following section of the report), and the utility of the procedure developed increases substantially.

F. Conclusions

Based on the results obtained from the work conducted on this task of the project, the following conclusions can be made:

1. The procedure developed provides investigators with a method of rapidly processing aircraft acquired MSS data, which produces results of a similar nature to those obtained using conventional aerial photo-interpretation techniques

2. The diversity of environmental conditions over which the procedure was tested did not adversely affect the classifications obtained, indicating that the procedure established is relatively unaffected by geographic location

3. Development of software which would permit automated map comparison to occur would greatly enhance the utility of this data processing methodology.

C-2

VI. REGISTRATION OF AIRCRAFT ACQUIRED MULTISPECTRAL SCANNER DATA TO A MAP BASE

A. Description of the Task

The objective of this task of the project was to develop a computer implemented procedure which would permit the registration of aircraft acquired MSS data to a map base. Such a procedure would enhance the utility of aircraft acquired MSS data since:

1. It would permit the investigator to remove scale related distortions caused by variations in aircraft speed or direction, and would compensate for geometric distortions caused by MSS angle of look

2. It could be used to register raw or classified data

3. It would permit the investigator to define the output data cell size desired and to place the registered data into a DATBAS

4. Two aircraft missions flown over the same area could be compared through use of DATBAS applications software.

Software required to implement this registration procedure was developed, tested, and refined by NASA/NSTL and documentation was subsequently transferred to EPA/EMSL at the conclusion of Phase II activities.

B. Scope of the Task

Demonstration of the effectiveness of the software developed was conducted on the aircraft acquired MSS data for Belle Ayr and Navajo coal strip mines dealt with in the preceeding section of this report. The 1975 data set for each mine was used.

The software was developed in such a manner so as to permit the registration to be made to the UTM map projection system. This would facilitate placement of the registered data into the DATBAS, where especially developed applications software could be used to conduct analysis in a manner similar to that mentioned earlier in this report.

C. Technical Approach

Three interrelated software programs were developed which can be used to register aircraft acquired MSS data to a UTM map base. These programs can handle aircraft data containing up to 11 channels, and permit the user to select the registration method desired.

The first program in the series, BGRID (Build GRID), constructs a grid of scan line and element numbers on a disc memory file. These numbers correspond to a rectangular grid of UTM northing and easting coordinates. Input to this program consists of such items as UTM northing and easting defined boundaries for the study area, number of input channels, nadir element of the scanner, desired output cell size, scan line-element boundaries of interest (if any, if none are used, the entire data set is used), and the registration method to be used. Data can be registered using an overall cubic fit with either a bilinear or nearest neighbor interpolation, or a bilinear interpolation between grid nodes with either a bilinear or nearest neighbor interpolation.

Output from the program is in the form of a statement which directs the user to construct a file of a specified size which will eventually be used to store the grid constructed. This consideration which was built into the program is quite useful, as these files can now be constructed large enough to contain the data, and unnecessarily large files are avoided, thus saving on unused disc space.

The remaining two programs, GEOBL (Georef-Bilinear) and GEONN (Georef-Nearest Neighbor), use the grids constructed by BGRID, along with control points selected by the investigator, and the unregistered MSS data, to construct an output file containing the registered MSS data. The choice of GEOBL or GEONN depends on whether bilinear or nearest neighbor interpolation was used in BGRID. Registered data is oriented to north, and is in a format compatible with DATBAS requirements.

D. Procedures and Methodology

To demonstrate the utility of the software developed, the classified results obtained in the proceeding section of this report were subjected to registration through the use of the software developed. An overall cubic fit with bilinear interpolation was used to produce the output.

To demonstrate the procedure by which aircraft data is registered to a map base, reproductions of the printouts obtained through the use of an interactive remote job terminal (foreground job) have been included as Figures 19, 20, and 21. Reference will be made to these figures throughout the following discussion.

Figure 19 presents the interactive run stream for the initial execution of BGRID. Execution of BGRID actually occurs twice; the first execution is used to calculate the size of the output disc file required, and the second execution is used to actually construct the grid.

After the program is loaded (LOAD, BGRID, CONPTS.) the software directs the investigator to indicate whether this run is the first run (for calculation of disc file size) or the second run (for grid construction). Since Figure 19 is an example of the first run, a "1" was input.

The next eight inputs deal with parameters which are used to define the data that will be processed, as well as the geographic area of interest. These inputs are self explanatory, except perhaps for the fact that output pixel dimensions are given in meters.

The investigator is now given the choice of the method by which the data is to be processed. As was mentioned, the method chosen for the data used in this study was an overall cubic fit with bilinear resampling (interpolation). The relative merits of each option can be found in Graham, 1978.

Figure 19 - Example Runstream for First Execution of Program BGRID

```
LOAD, BGRID, CONPTS.

Input 1.  If initial parameters are to be input-otherwise, key
          return:  1.

Input scan angle in radians:  .7854

Input nadir element:  314.5

Input max northing, min northing:  4884000, 4882000.

Input min easting, max easting:  468500, 472000.

Input pixel dimensions for output-horizontal, vertical:  9.146,
                                                         9.146.

Input min element, max element:  192, 400.

Input min scan, max scan:  1, 250.

Input number of channels on tape.  If classified tape, input 1.

1.
Select one of the following types of grids.
BL-bilinear interpolation, NN-nearest neighbor
1. - overall cubic fit - BL resampling
2. - overall cubic fit - NN resampling
3. - BL Grid - BL resampling
4. - BL Grid - NN resampling
1.

Input upper bounds for scan error, elt. error: 3, 3.

Create grid file of 2400 words and 221 records.
```

Figure 20. Example Runstream for Second Execution of Program BGRID

```
LOAD, BGRID, CONPTS, FILE.

Input 1.  If initial parameters are to be input-otherwise, key return.

Deleted EAS= 469375.  NOR= 4881440.  SCAN= 227.  ELE= 431.
Deleted EAS= 469185.  NOR= 4879830.  SCAN= 223.  ELE= 584.
Deleted EAS= 468920.  NOR= 4883850.  SCAN= 307.  ELE= 230.
Deleted EAS= 468975.  NOR= 4884186.  SCAN= 313.  ELE= 199.
Deleted EAS= 470815.  NOR= 4882666.  SCAN=  82.  ELE= 325.
Deleted EAS= 470885.  NOR= 4883036.  SCAN=  66.  ELE= 294.
Deleted EAS= 469385.  NOR= 4880336.  SCAN= 220.  ELE= 536.

                    RESIDUAL ERRORS-CUBIC FIT
Points with SCAN errors >3.0 and ELT. errors >3.0 were deleted.

    C.P.                          CORRECTED
    NUMBER    SCAN    RESIDUAL     ELEMENT     RESIDUAL
      1        98.   -.220E-01      227.       .149E 01
      2       294.    .105E 01      286.      -.356E 00
      3       291.   -.326E 00      298.       .465E 00
      4       287.    .849E 00      332.      -.514E-02
      5       305.   -.148E 01      332.       .627E 00
      6       280.   -.439E 00      369.      -.205E 01
      7       260.    .233E 01      367.      -.190E 01
      8       259.   -.227E 00      535.       .126E 01
      9       258.   -.122E 01      543.       .190E 01
     10        24.    .628E 00      329.      -.233E 01
     11       138.   -.196E 01      227.      -.785E 00
     12       246.    .129E 01      525.      -.764E 00
     13        78.    .247E 00      250.       .618E 00
     14       114.    .556E 00      592.      -.122E 01
     15        75.    .147E 00      226.       .171E 01
     16        84.   -.142E 01      546.       .134E 01
    ---------------------------------------------------------
    SIGMA ERRORS      .11E 01                  .135E 01
                                CONTROL POINTS USED
    NUMBER    SCAN    C.ELT    EASTING    NORTHING
      1        98.    227.     470790.    4883856.
      2       294.    286.     468965.    4883156.
      3       291.    298.     468965.    4883030.
      4       287.    332.     468965.    4882630.
      5       305.    332.     468785.    4882630.
      6       280.    369.     468965.    4882190.
      7       260.    367.     469170.    4882216.
      8       259.    535.     468940.    4880356.
      9       258.    543.     468930.    4880270.
     10        24.    329.     471320.    4882606.
     11       138.    227.     470410.    4883840.
     12       146.    525.     469970.    4880440.
     13        78.    250.     470940.    4883570.
     14       144.    592.     470160.    4879680.
     15        75.    226.     471000.    4883860.
     16        84.    546.     470475.    4880210.
```

Figure 21 - Example Runstream for Execution of Program GEONN

```
LOAD, GEONN, TAPE, FILE

To copy what's on disk to tape, input 1. - otherwise key return.

File Name:  Tape,    Error:  11,    Record:  2,    File:  1,
Parities:  0
```

The final input defines the maximum allowable errors for registration, in meters, and any control points with an error greater than this will be discarded.

And finally, the program outputs the size of the disc file required for construction of the grid.

Figure 20 represents the interactive job stream for the second run of BGRID. The "LOAD, BGRID, CONPTS, FILE". Command loads the program, along with the control point file ("CONPTS") and the grid file ("FILE") created at the direction of the initial run of BGRIDS. In this case, no input is made on the initial program generated directive, and the machine proceeds to examine points, delete those with errors larger than the user input maximum, and finally constructs the grid.

Figure 21 contains the run stream for GEONN. This program requires the tape containing the MSS data ("TAPE") and the grid file constructed by the second run of BGRID ("FILE") as inputs, and outputs a disc file containing the registered MSS data. This disc file can then be copied to magnetic tape if desired, or used as input to the DATBAS.

E. Analysis and Results

Figure 22 demonstrates the results obtained when a particular set of data have been rectified to a map base, in this case raw single channel MSS data (1975) for the Belle Ayr coal strip mine. The original flight line ran east-west (top half of the figure),

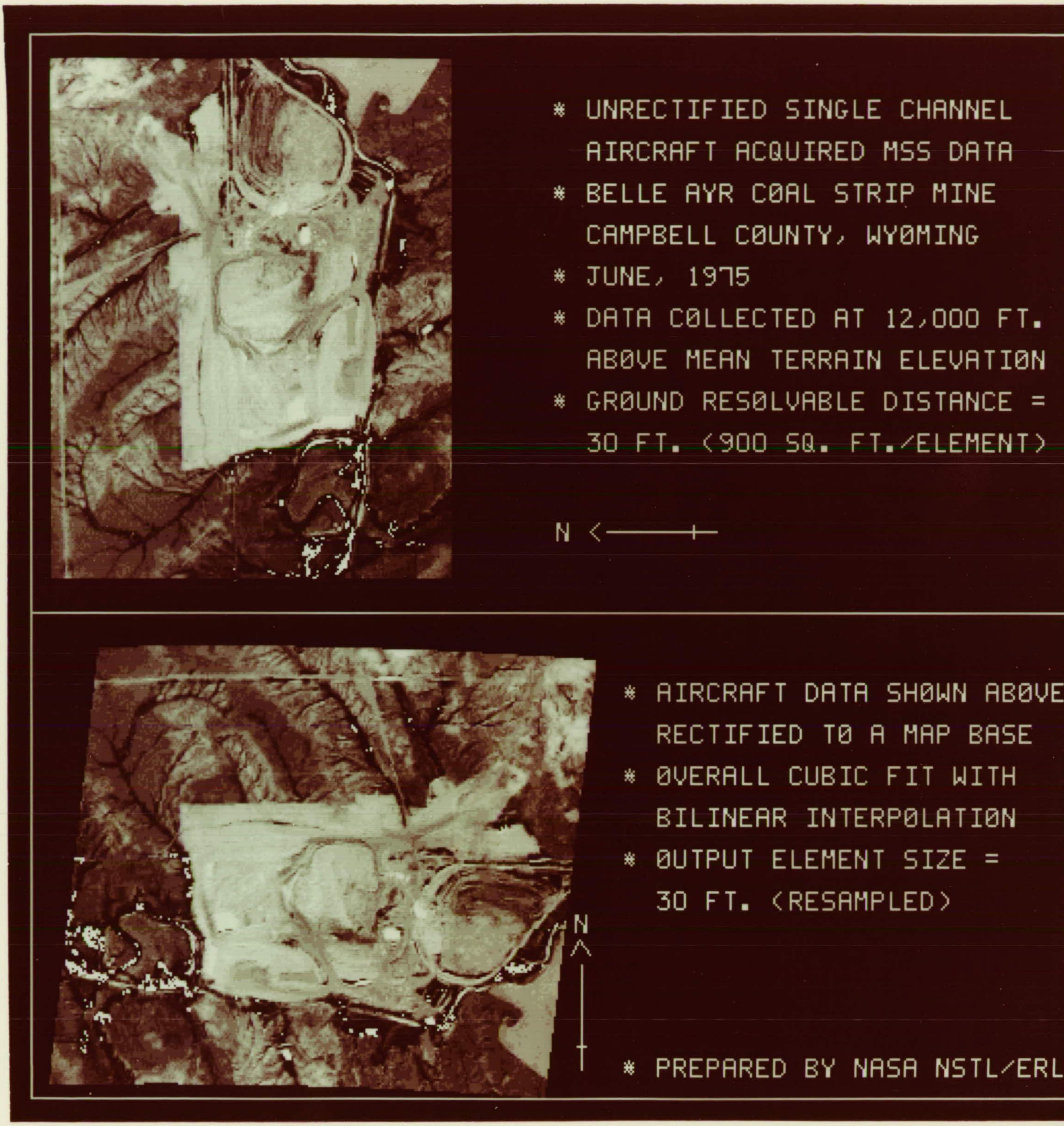

Figure 22 An Example of the Registration of Aircraft Acquired Multispectral Scanner Data to a Map Base

and the resulting rectified data (bottom half of the figure) has been rotated so that north is to the top. The rectification software can handle up to 11 channels of raw data, should the investigator desire to rectify raw data for map making or for use in other application oriented tasks.

The lower half of Figure 22 also represents data that is in a format compatible with the DATBAS, and such data could be used as input to the DATBAS software. Once the data has been stored in the DATBAS, software could be developed which would permit automated map comparision to occur as was mentioned in the preceeding section of this report.

Since the data is now referenced to the UTM map projection system, each element in the data has associated with it a specific UTM northing and easting coordinate. This fact would permit the investigator to obtain geographic boundaries for various surface mining related activities, and use these boundaries to compute acreages or to monitor progress of mining activities.

It should be noted that this is a significant achievement, because, to date, the use of aircraft acquired MSS data has been hindered by the fact that no such geographic reference has been available other than optical correction of hard copy reproductions of the data.

Classified results obtained from the analysis conducted in the proceeding section of this report for 1975 aircraft acquired

MSS data for the Navajo Coal Strip Mine which have been registered to a map base are presented as facing page Figures 23 and 24.

Figures 23 shows the classified data for the entire data set after rectification has been performed. (The reader is referred to Figure 15 for a presentation of unrectified 1975 MSS data for this mine.) It is of interest to note the curvature of the east and west edges of the data, resulting from a very slight turn by the aircraft while data collection was being made. Through the selection of control points and subsequent data processing, this was discovered. Unrectified data displays (e.g. Figure 15) do not readily bring this fact to light.

Figure 24 is included to demonstrate the degree to which the rectified data fits a map base. As can be seen, major features agree quite well. Also note the advance of the mine in the time period between mapping (1966) and MSS data collection (1975). The colors correspond to those in the key of Figure 23.

F. Conclusions

Several conclusions can be made as a result of registering aircraft acquired MSS data to a map base.

(1) Software designed for the registration can compensate for certain fluctuations in flight direction encountered when data is being collected, provided that these aircraft heading changes are not excessively abrupt

Figure 23 Computer Implemented Land Cover Classification of Aircraft Acquired Multispectral Scanner Data Registered to Map Base

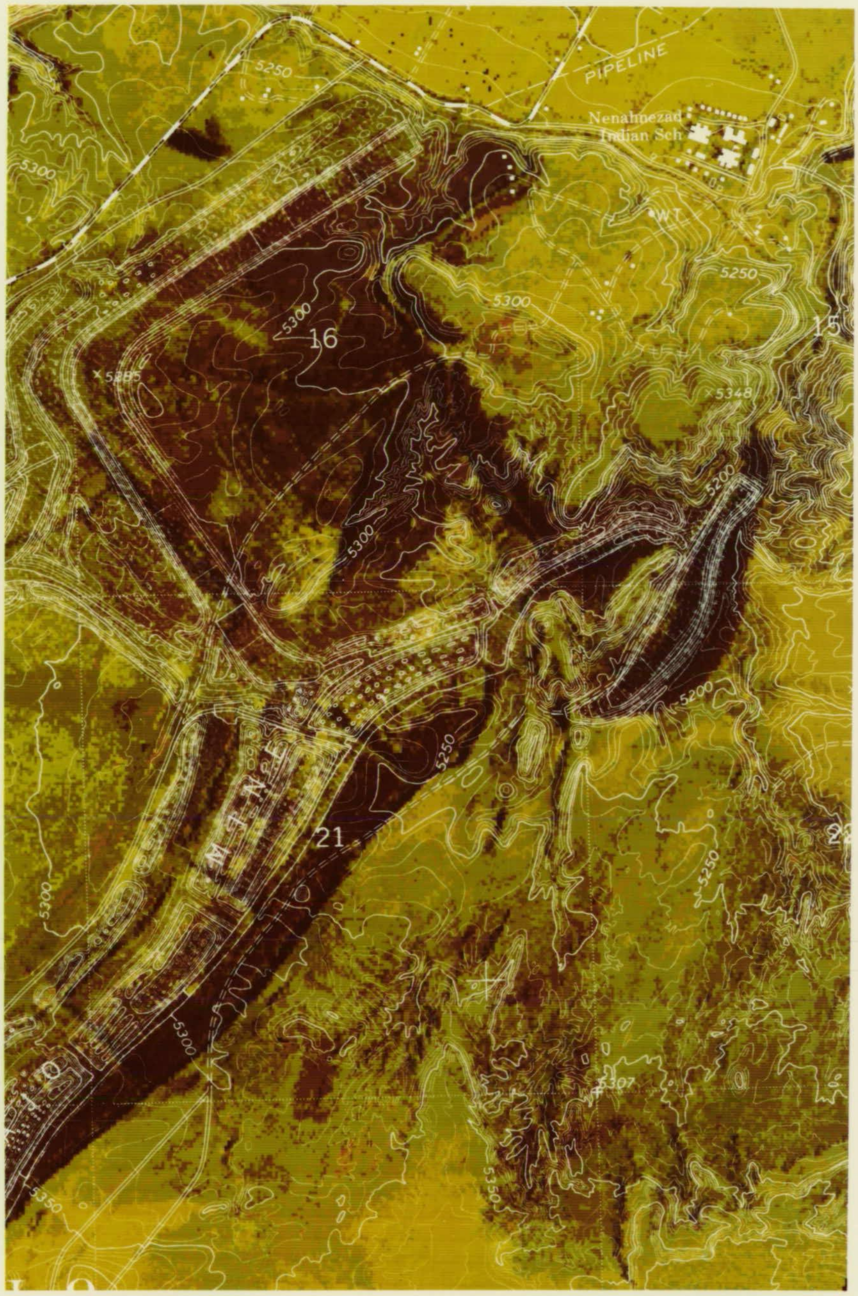

Figure 24 An Enlargement of a Section of Figure 23 with an Overlay of Features from 1:24,000 Scale Topographic Base Map Superimposed

(2) The overall fit to a map base is greatly enhanced through the selection of control points selected over the entire geographic area contained in the MSS data being processed. In addition, the more numerous the number of control points, the better the overall fit

(3) Better fits are obtained if control points are selected outside of the geographic area of interest. This implies that data collection activities need to be planned so as to include such areas in the MSS data. Including such data points results in a minimization of extrapolation (as opposed to interpolation) by the software, and a reduction in the "edge effect" (where the fit to a map is markedly poor at the edges of the data)

(4) The capability to register aircraft acquired MSS data to a map base permits direct comparisons of several flights

(5) Auxiliary data can be added (in a manner similar to that described in an earlier section of this report) and investigations can now incorporate such data where it would materially enhance the results obtained. In addition, registered aircraft data could conceivably be integrated with Landsat data and auxiliary data in a DATBAS format, and used in a systems approach to problems encountered by investigators.

LITERATURE CITED

NASA 1976. Landsat Data Users Handbook. NASA Goddard Space Flight Center Doc. 76SDS4258 (Revised April 1, 1977) Greenbelt, MD 20721

Whitley, S. W. 1976. Low Cost Data Analysis Systems for Processing Multispectral Scanner Data. NASA TR R-467. Johnson Space Center, Houston, TX 77058

Joyce, A. T. 1978. Procedures for Gathering Ground Truch Information for a Supervised Approach to a Computer Implemented Land Cover Classification of Landsat Acquired Multispectral Scanner Data. NASA Reference Publication 1015. NASA Johnson Space Center, Houston, TX 77058

Jones, W. C. 1976. CHOICE: 36 Band Feature Selection Software with Applications to Multispectral Pattern Recognition, NASA/ERL Report #59. NASA NSTL-Earth Resources Laboratory, 1010 Gause Blvd., Slidell, LA 70458

Anderson, J. E. and C. E. Tanner 1978. Remote Monitoring of Coal Strip Mine Rehabilitation EPA Document EPA-600/7-78-149. EPA Environmental Monitoring and Support Laboratory, Las Vegas, NV 89114

Graham, M. H. 1977. Digital Overlaying of the Universal Transverse Merca for Grid with Landsat Digital Data Derived Products. NASA TM 58200, Johnson Space Center, Houston, TX 77058

Anderson, J. E. 1976. Western Energy Related Overhead Monitoring Project Phase I Report. NASA NSTL-Earth Resources Laboraotry Report #162, NASA NSTL-ERL, 1010 Gause Blvd., Slidell, LA 70448

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. USGS Professional Paper 964, 28 pps.

NASA/NSTL Earth Resources Laboratory 1975. ERL DAS Tape Format Manual. NASA/NSTL Earth Resources Laboratory, 1010 Gause Blvd., Slidell, LA 70458

Graham, M. H. 1978. Aircraft Geographical Reference Module Program Documentation, NASA NSTL-Earth Resources Laboratory, 1010 Gause Blvd., Slidell, LA 70458