

1-1-1976

# Analysis of Geophysical Remote Sensing Data Using Multivariate Pattern Recognition Techniques

Paul E. Anuta

Hans Hauska

Donald W. Levandowski

Follow this and additional works at: [http://docs.lib.purdue.edu/lars\\_symp](http://docs.lib.purdue.edu/lars_symp)

---

Anuta, Paul E.; Hauska, Hans; and Levandowski, Donald W., "Analysis of Geophysical Remote Sensing Data Using Multivariate Pattern Recognition Techniques" (1976). *LARS Symposia*. Paper 109.  
[http://docs.lib.purdue.edu/lars\\_symp/109](http://docs.lib.purdue.edu/lars_symp/109)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

Reprinted from

**Symposium on  
Machine Processing of  
Remotely Sensed Data**

**June 29 - July 1, 1976**

The Laboratory for Applications of  
Remote Sensing

Purdue University  
West Lafayette  
Indiana

IEEE Catalog No.  
76CH1103-1 MPRSD

Copyright © 1976 IEEE  
The Institute of Electrical and Electronics Engineers, Inc.

Copyright © 2004 IEEE. This material is provided with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the products or services of the Purdue Research Foundation/University. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Analysis of Geophysical Remote Sensing Data  
Using Multivariate Pattern Recognition Techniques

Paul E. Anuta, Hans Hauska\*  
Donald W. Levandowski

Laboratory for Applications of Remote Sensing  
Purdue University  
W. Lafayette, Indiana

I. ABSTRACT

Multivariate statistical pattern recognition techniques have been widely used in the analysis of multispectral scanner remote sensing data for crop surveys, forest mapping, land use surveys and in many other applications. These applications are restricted basically to surface cover reflectance and emissivity phenomena. In the study described in this paper multivariate analysis techniques were applied to geophysical remote sensing data which measures phenomena occurring beneath the surface of the earth. Three types of geophysical data: magnetic anomaly, induced pulse transient, and gamma ray data were digitized, registered and analyzed to observe relationships to known geology. In addition several types of surficial remote sensing data including LANDSAT multispectral scanner, side looking airborne radar (SLAR) and thermal infrared scanner data were included in the multivariate data set to enable evaluation of all the available remote sensing variables. The preprocessing and analysis techniques are discussed and results showing correlations between variables and relationships to geology is presented.

II. INTRODUCTION

LANDSAT and aircraft multispectral scanner (MSS) data provides measurements relating to surface reflectance and emissive properties of scene objects. MSS data has been used widely in the geological field for mapping surface features which often relate to subsurface conditions but only indirectly. Geophysical remote sensing on the other hand measures the effects of processes originating beneath the earth's surface and data from such measurements is widely used for mineral and petroleum

exploration purposes. Manual interpretation methods are generally used to process this data and a great need exists for quantitative and automated methods for processing multivariable geophysical remote sensing data. In the work reported here geophysical data was digitized, registered, and analyzed using the multivariate techniques which have been widely applied to multispectral scanner data. In addition to the geophysical data, topographic, geological map, side looking radar, thermal infrared data, and LANDSAT MSS data were also digitally registered together to produce a large dimensionality measurement vectors for expanded flexibility in the analysis. This data was analyzed using a variety of the statistical analysis and pattern recognition techniques to explore the relationship of groups in the resulting hyperspace to known geological features.

The study was conducted in these phases: 1) data preprocessing, 2) data analysis, and 3) results interpretation. Since the bulk of the geophysical data was in contour map and film format manual digitizing, gridding and registration was required and this process required a large amount of resources and technique development. This work is described in Section III. Once all the geophysical and surface remote sensing data was digitized and registered to a uniform grid and stored as a multi-channel tape file, computer analysis could then be conducted. These activities are described in Section IV. Since the goal of the study was to observe the relationship of multivariate analysis results to the geology and known mineralization characteristics the computer processing results were subjected to interpretation by geologically and geophysically trained personnel. An example of these interpretations is presented in Section V.

\* Postdoctoral Research fellow, sponsored by the European Space Agency (ESA), on leave from Luleå University of Technology, Luleå, Sweden.

### III. DATA HANDLING

The greatest challenge in the study was to digitize, transform to a uniform grid and register a wide variety of remote sensing, and map data types. Three basic forms of data were handled: 1) Digital, 2) Graphic Contour and Polygon Maps, and 3) Film. Data for five test sites were acquired and processed with up to twenty variables or "channels" generated for each site. The variables by data type are listed in Table 1. LANDSAT data was obtained in digital form on CCT's thus no preprocessing was required prior to registration. A 250 foot square grid cell size was chosen as the reference grid and all other forms of data were registered to this grid.

The map data contains information in two forms: 1) contours and, 2) polygons. All the geophysical data plus the topographic data were in the form of manually or machine generated contours - the geologic maps are of the polygon type. The six geophysical variable maps and the geology map for each site were manually digitized using a coordinate digitizing table. The contours were punched on cards for processing by gridding software.

Contour gridding is a particularly difficult problem and many methods are in use for interpolating contoured data to a uniform grid. The method used in this study (obtained from a report by Turner<sup>1</sup>) forms a linear combination of the six nearest points according to the formula:

$$z'_{jk} = \frac{z_n}{2} + \frac{\sum_{i=1}^N z_i / \sqrt{D_i}}{2 \sum_{i=1}^N 1/\sqrt{D_i}}$$

where:  $z'_{jk}$  is the interpolated value for the grid cell at column j, row k.

$z_n$  is the nearest contour data point.

$z_i$  are the N nearest contour points.

N = 6 for this case

$D_i$  are the corresponding Euclidean distances from  $z_i$  to  $z'_{jk}$ .

A uniform grid of points for the chosen 250 ft. interval was created for the six geophysical variables for each site. The polygons from the geologic map require a different and much simpler data handling process. All that is required is to fill the region inside each polygon with a code

number for the geologic unit surrounded by each polygon. Careful records must be kept defining the contents of the area to the left and right of the boundary but the process is straight-forward. The result of the digitizing and gridding process is thus a LANDSAT compatible image like data set for each of the seven map variables.

The film sources presented different and more difficult data handling challenges. The thermal infrared aircraft scanner data was recorded on film in strips nominally 3000 feet wide with a dozen or more strips (flight lines) covering the test sites with nominally 50% overlap. The strips had severe panoramic distortion due to the 45° maximum look angle and were elongated in scale by a 5 to 1 factor. A further problem was with sensor calibration which prevented accurate strip to strip matching of the data. This data was preprocessed by first digitizing the strip film laydowns on a microdensitometer then breaking the data set into individual strips which were then panoramically and scale corrected. Control points were manually located in each strip and used to register the strip to the reference grid. Finally the registered strips were joined into one data set covering the test site.

The side looking radar film data was received in one mosaiced block for the test site; however, control point location proved to be a difficult task. Matching points on the SLAR image and topographic maps were found using a Zoom Transfer Scope which registered the two sources and allowed visual location of control in the SLAR. Color and color infrared aerial photography was reduced and digitized through blue, green and red separation filters and the data was recorded on tape.

The sixteen data variables were registered with the four LANDSAT bands using the LARS image registration system<sup>2</sup>. Manually derived control points were used to define a bi-quadratic warp function to geometrically transform each variable to match the 250 ft. reference grid. Control points were derived from map data (variables 5 thru 12 in Table 1) via the coordinate digitizing table. Control points in the digitized film (variables 13 thru 20) were obtained by displaying the data on a Digital Image Display system<sup>3</sup> and visually identifying the points and recording them via a light pen. The registration process produced a twenty channel data file which forms the basic input to the multivariate geophysical data study.

### IV. DATA ANALYSIS PROCEDURES

The main objective of this study was to evaluate the different data sources in order to select the best possible subset

to be used in further surveys. A large number of processing functions for multivariate analysis are available in LARSYS<sup>4</sup>, a system which was originally developed for the analyses of multispectral data only. The wide variety of data available in this study lead to the belief that the common approach to multispectral analysis including maximum likelihood classification based on a number of training classes would not be feasible. Therefore, it was decided to use the LARSYS nonparametric clustering routine (called \*CLUSTER), which is in principle based on the ISOCLS-algorithm (Ball and Hall<sup>5</sup>) to detect clumps of data with similar properties.

Due to the large differences in variability and range in the data, e.g., aeromagnetics range from 4000 - 6000 gamma and INPUT channel ranges from 0-30 arbitrary units, clustering could not be performed on the original data values. The data was clustered by using the raw value as stored on the magnetic tapes, where each point is represented by a series of integers in the range between 0 and 255. Still, the clustering was not successful when using channel 8 (the ratio). This can be attributed to the fact that the ratio between the two radiometric channels is usually represented by a smooth surface with very low frequency of spatial variation. Also, the ratio channel exhibited in the histograms the smallest number of modes. It therefore dominated any combination of channels in which it was used.

In order to assign initial cluster centers in a correct fashion a version of the cluster program was used which assigns cluster centers along the largest eigenvalue of the data hyperellipsoid. The number of clusters originally asked for was fifteen in all cases. This is an arbitrary number which appeared to be quite effective for practical purposes.

The cluster output consists of a map on which the different symbols characterize the cluster to which each point has been assigned. In order to display this map in a proper fashion a processor was developed which performs a classification based on the same principle as \*CLUSTER, namely by assigning each point to the nearest cluster center. The output from this processor is written onto magnetic tape for further processing.

In order to systematically evaluate the results, another program was developed which counts the number of points in each cluster which are found in each geological unit. The result is written out in the form of a Table. From this table one is able to draw conclusions as to which variables best represent the given geology.

Furthermore, in order to investigate the interdependence of the different variables the LARSYS \*STATISTICS processor was used to compute the means and covariances for each area for all the channels, geology excluded. This was done with the purpose of investigating the dependence of the geophysical variables on topography. Some correlation was found, but not significant - at least not in statistical terms. To go a step further, a multiple linear regression was computed using all the variables, deleting the geology and also deleting the four LANDSAT channels.

## V. INTERPRETATION OF ANALYSIS RESULTS

Initially, the analysis procedure was applied to an area that included a known, porphyry copper deposit surrounded by an alteration halo. The copper mineralization is confined to a porphyritic quartz monzonite intrusive which is one to one and a half miles in size. Alteration consists of intense quartz-sericite ranging outward from the copper area to a zone of weak propylitic and clay alteration.

A cluster map of the area was prepared using the following types of data: two LANDSAT channels (0.5-0.6 $\mu$ m and 0.8-1.1 $\mu$ m), uranium, potassium, and magnetics (Figure 1). The original geophysical contour maps show that the area incorporating the quartz monzonite, altered rocks, and mineralization is associated with broad anomaly highs on the uranium and potassium maps and a steep gradient on the magnetic map. The cluster map, on the other hand, depicts the quartz monzonite (dash symbol), the area of intense quartz-sericite alteration (equal symbol), and the weak clay and propylitic alteration zone (I symbol).

Thus the clustering of the multivariate data allows the examination of the several types of data as well as their intercorrelations in a single image which in this particular case enables one to extract more information than from any one given data type alone.

## VI. CONCLUSIONS

This study was an effort to apply existing multivariate remote sensing data analysis techniques to geophysical data to aid in the interpretation of this type of data for mineral exploration. The majority of the achievements were in the area of data handling and digital analysis algorithm development. There were many difficulties in digitizing, gridding, and registering very different film and map data types and development of these methods alone is thought to be a significant contribution to the state of the art.

Interpretation of the products of multi-variate digital processing of the combined geophysical and surficial remote sensing data has only begun and much work will be required to evaluate the benefits of the multivariate digital approach. Only a brief interpretation example is shown since the data used is proprietary in nature; however, the general concept is thought to have great potential and for this reason this limited discussion is presented.

VII. REFERENCES

1. Turner, A. K., "The GCARS System," Fortran IV Programmer Manual, Part C - Programs for Contour Mapping, Joint Highway Research Project Report, Project C-36-72A, File 1-6-1, Purdue University, September 1969.
2. Anuta, P. E., "Computer-Aided Analysis Techniques for Remote Sensing Data Interpretation," LARS Information Note 100675, Purdue University, October 1975.

3. Wilson, L. L., "Purdue/LARS Digital Display Users' Guide," LARS Information Note 022675, Purdue University, Lafayette, Indiana, February 1975.
4. Phillips, T. L., "LARSYS Version 3 Users Manual, Volume 2, 1973," Laboratory for applications of Remote Sensing Purdue University, Lafayette, Indiana.
5. Ball, G. H., Hall, D. J., "ISODATA, A Novel Method of Data Analysis and Pattern Classification," Stanford Research Institute Report AD699616, Menlo Park, California, pp. 1-16, 1965.

Table 1. Data Types and Variables Digitized and Registered.

Channel	Type	Source	Description
1	Digital	LANDSAT MSS	.5-.6um Surface Reflectance
2			.6-.7um Surface Reflectance
3			.7-.8um Surface Reflectance
4			.8-1.1um Surface Reflectance
5	Geophysical Contour & Polygon Maps	Gamma Ray Radiation	Thorium (Tl <sup>208</sup> Isotope)
6			Uranium (Bi <sup>214</sup> Isotope)
7			Potassium (K <sup>40</sup> Isotope)
8			Potassium/Thorium Ratio
9		500 Ft. Magnetics	Magnetic Anomaly Measurement
10		INPUT (Induced Pulse Transient)	Conductivity of Subsurface
11		Topography	Topographic Elevation
12		Geologic Map	Mapped Geologic Units
13	Film	Side Looking Radar	Radar Reflectance
14			Thermal Infrared
15		Color Aerial Photography	Visible Blue Wavelengths
			Visible Green Wavelengths
			Visible Red Wavelengths
18		Color Infrared Aerial Photography	Visible Green Wavelengths
	Visible Red Wavelengths		
19		Reflective Infrared Wavelengths	
20			

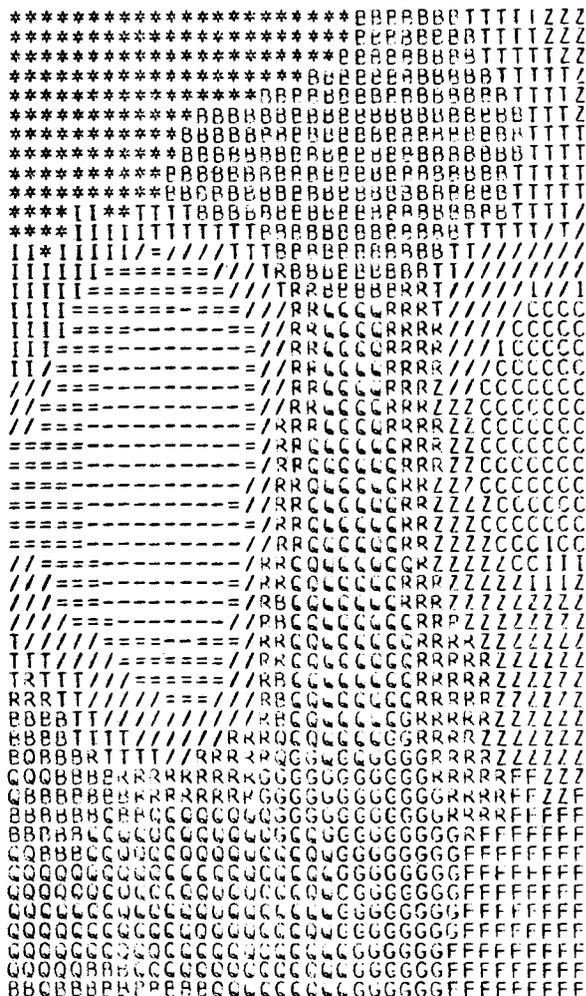


Figure 1. Clustering Results for LANDSAT, Gamma Ray, and Magnetic Field Data.