COMMISSION III Working Group 2

Presented Paper

GEOMETRIC REPRESENTATION METHODS FOR MULTI-TYPE SELF-DEFINING REMOTE SENSING DATA SETS

by

Paul E. Anuta

Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana 47907, U.S.A.

ABSTRACT

The use of digital remote sensing image data from aircraft and satellite-borne sensors is becoming widespread in land use, environment, and agricultural monitoring. In addition to sensor data, a wide variety of ancillary data types are being used in conjunction with remote sensing data to aid in its analysis and utilization. The requirement for creation of multidimensional digital data sets containing many types of sensor and ancillary data types has led to a number of approaches for transformation and merging of various data types. The basic concept presented here is that of the self-defining data set or SDDS. An SDDS contains in its accompanying support data all information needed for correct geometric and radiometric representation of all samples. The proposed structure of an SDDS will be defined and the majority of the paper will be devoted to the problem of geometric representation of a diverse set of data types. The paper closes with a description of a software system which accepts an SDDS and produces a registered high - dimensional data set. Examples of multi-data sets generated using these techniques are presented including a Landsat, earth surface gamma ray and magnetics combination for mineral exploration.

INTRODUCTION

Efficient and convenient representation of remote sensing image data and supporting ancillary data are necessary if effective use is to be made of the diversity of data types available today and which will be available in the future. Current approaches to the problem of merging different data types are to treat each case as an individual registration/rectification problem. Control points and distortion models are laboriously acquired for each data type for each area to be studied. This paper describes work being carried out to standardize the multi-data merging process to enable convenient user assembly of a selection of data types without being concerned with control points, distortion models resolution, projection. etc. A recent article in a publication on "smart" sensors discusses the data-flow problem and is relevant background for the ideas proposed here.[1]

THE MULTI-DATA MERGING PROBLEM

The requirement for merging and rectification of multiple data types has come about from the increased use of digital computer methods for resource monitoring and analysis. Land-use planning, geological investigations, surveys of agricultural, forestry and other vegetative cover resources, and many other applications require diverse data types to be brought together for correlative analysis. For example, in a mineral exploration activity, aircraft-gathered geophysical gamma ray and magnetics data were merged with Landsat satellite MSS data and ancillary data (geology map and digital topographic data) to form a high dimensionality data set.[2,3] The data were analyzed to identify areas having anomalous response in a number of variables which may contain mineralization. Many other examples exist in the fields mentioned. A strong need has arisen to formalize the approach to merging data to form user data sets.

SELF-DEFINING DATA SETS

The basic concept set forth in this paper is that of the self-defining data set (SDDS). When one considers the problem of merging a diversity of data types for many ground sites, it quickly becomes prohibitive to find control points to warp each data type into registration with all the others and to a reference. This is especially true for high-dimensionality data sets. We propose here the creation of a standard which would ideally be worldwide. This standard would be such that data which may be of interest in a large number of earth resources remote sensing applications would be in a format which allows convenient and automatic merging. This is possible only if each data set to be merged has complete information stored in its ancillary data regarding its geometric characteristics. Candidate data sets can be in two forms: uncorrected and corrected to some reference. The diagram in Figure 1 indicates some of the data types of interest and the data flow needed to create a self-defining data set for each. Clearly, geometric description of uncorrected data will require more complex functions than data rectified to some projection and coordinate system. For example, a full Landsat frame has been shown[4] to be adequately geometrically described by a two-dimensional fifth-degree polynomial. This requires 42 coefficients to represent geometric distortion in each frame. A fully corrected Landsat frame placed in a UTM or HOM projection requires only a minimum of ancillary parameters (scale, zone azimuth, center coordinates, etc.). Thus the fully corrected form is very desirable from the standpoint of geometric simplicity. On the other hand, correction of all data regardless of the likelihood of use is extremely wasteful of resources and requires resampling which modifies data values and is a costly operation. If all data were stored in uncorrected form with full geometric description, then only the pixels required for use need be extracted, geometrically transformed, resampled, and placed in the user's grid coordinate system. In either data form, the concept is the same and we submit the SDDS concept as a desirable goal. The current fully corrected Landsat CCT products (commonly called P tapes) are SDDS's, and this development is pursuant to the general concept.

GEOMETRIC DESCRIPTION OF MULTI-TYPE DATA SETS

The most widely available remote sensing data type in the world today is Landsat MSS spectral reflectance data. Prior to 1979, the CCT data were corrected for basic system distortions but not geometrically corrected and placed in a projection. New products from the NASA MDP system are accurately rectified to a

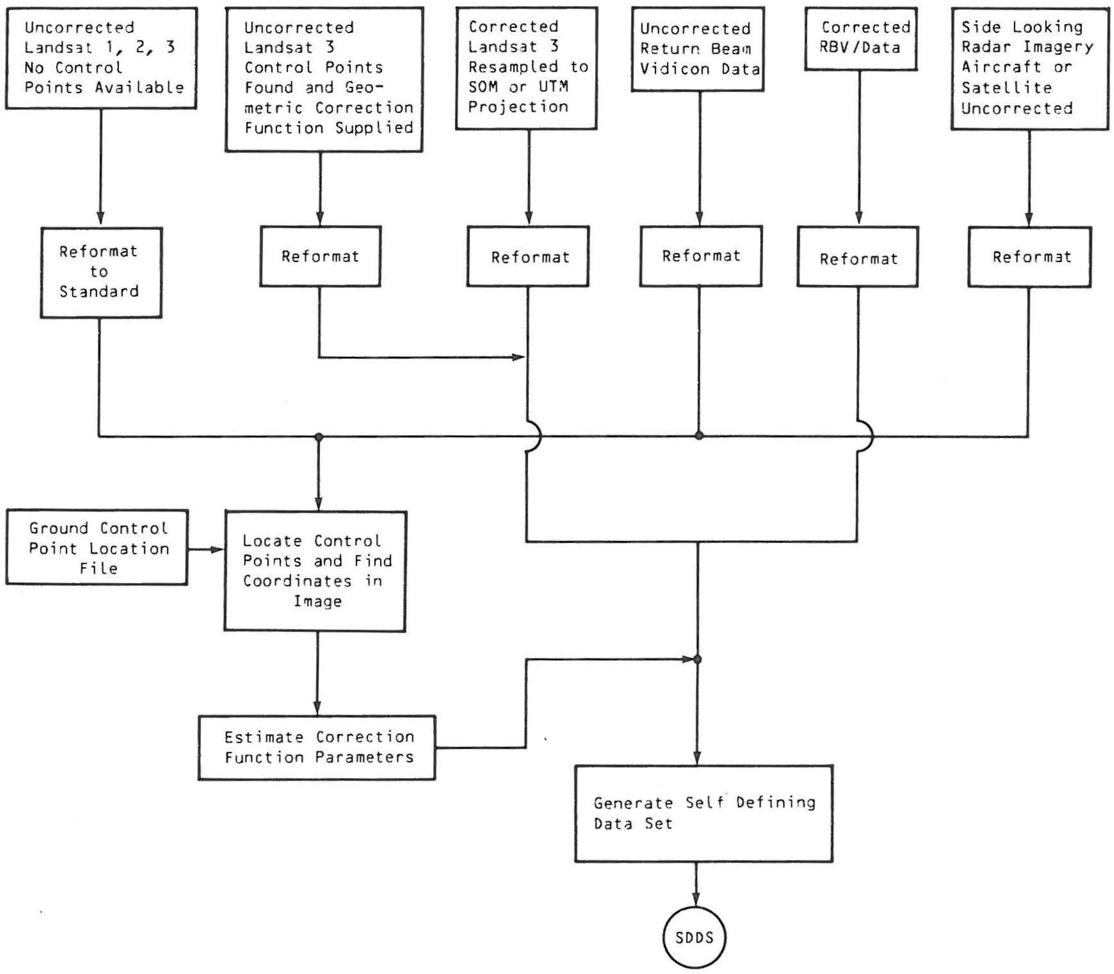

Fig. 1. Example data flow for self-defining data set generation.

UTM or HOM (Hotine Oblique Mercator) projection. The self-definition parameters required for fully corrected data are minimal:

1. Scale of pixels.
2. Azimuth of latitude zone.
3. Center latitude
4. Center longitude.
5. Frame center row.
6. Frame center column.

With these six coefficients, the geographic coordinates of any point in the frame can be determined to sub-pixel accuracy. In the case of uncorrected frames, 42 coefficients plus some housekeeping parameters are necessary. The geometric description of the MSS data is well studied and considerable experience has been accumulated in correction of this data type. In effect, SDDS format is currently available from U.S. Landsat data sources.

For the Return Beam Vidicon (RBV) sensor on board Landsat, both fully corrected and uncorrected data are also available. The fully corrected data are in a map projection and the six parameters discussed will define the geometry. The uncorrected data will have a similar pair of two-dimensional polynomials as ancillary data which will describe the corrections as with the

MSS data. We assume the polynomials are fifth degree or less since the area covered is less and distortions less than with the MSS.

A multitude of other sensor types are of interest in earth resources remote sensing, including thermal scanners, side-looking radar, geophysical, and geochemical sensors. Of particular current interest is side-looking synthetic aperture radar which promises all-weather observation of the earth surfaces from aircraft and satellites. Registration of aircraft SAR with Landsat[3] was studied and an analysis of the geometric distortions in the SAR was carried out. The data set studied covered a 16-kilometer wide by 18km long area in eastern Maryland, U.S.A. First through fifth degree polynomials were estimated using least squares from control points selected from the SAR imagery and from corresponding fully corrected Landsat imagery. The residual errors were tabulated and are presented in Table 1.

Table 1

Residual Errors for Geometric Functions
to Match Aircraft SAR Imagery
to Landsat MSS Imagery

Distortion Function	Residual Errors (Pixels)	
	Along-Track	Across-Track
Affine	5.00	4.04
Bi-Quadratic	4.13	4.16
Bi-Cubic	4.28	3.48
Fourth Degree	3.55	3.67
Fifth Degree	3.82	3.38

From these results it can be seen that higher degree representations tend to reduce error but not to a sub-pixel level, which would be desirable. The degree of function chosen for geometric representation is constrained by storage and evaluation complexity. For practical reasons, the fifth-degree case is assumed to be a limit and the 42 terms needed to represent this function are taken as the maximum needed to be stored in an SDDS. This limit also coincides with the degree of Landsat MSS distortion representation polynomials.

In the case of ancillary data types, these are generally taken from map sources which are already in some projection and coordinate system. One example is the conversion of geologic units from map form to a digital image grid format. The geology polygons were traced with a digitizing cursor and coordinates recorded. A gridded image file was generated to conform to the map geometry which in effect places the data in the map projection, which was Lambert Conformal Conic. Over small areas, a linear relationship between data set coordinates and geographic coordinates can be used; thus, nominally six terms suffice for map-based ancillary data. Another example is topographic data which have been obtained from USGS 1:250,000 scale maps. The data are available from the Geological Survey in blocks covering one-half of a quadrangle and are in a Transverse Mercator Projection. Again, a linear relationship can be used to relate pixels to other coordinate systems over areas of one quadrangle. Thus it is concluded that fifth-degree polynomials or less can be used to represent most data that would be input to a combined multi-type data set. In the ancillary record, up to 42 coefficients would be stored in addition to general housekeeping quantities. In many cases, however, far less than 42 would be required.

IMAGE RECONSTRUCTION FROM TRACK-TYPE DATA

Many remote sensing sensor types are non-imaging and collect data only along lines or tracks corresponding to the path of an aircraft, satellite, helicopter, or ground survey. Before a geometric characteristic is defined, a gridded image representation of the data is generated. Numerous methods exist for processing such data. One method[5] has the goal of isotropic reconstruction by requiring the along-track and across-track frequency characteristics to be the same. An along-track filter is designed to have a cutoff at the Nyquist frequency determined by the across-track spacing and an across-track interpolator is designed to match along-track filter in frequency response. The resulting two-dimensional filtering interpolating algorithm was applied to airborne geophysical survey data (uranium, thorium, potassium and magnetics) to produce an image data set. The data flow for the algorithm is shown in Figure 2. The geometry of the result was determined by a uniform grid of cells linearly related to latitude and longitude over a small area. The parameters necessary to define this data set are the scale of the cells, coordinates of the four corners of the data set, or a total of nine. Since this data set was north oriented and rectangular, only five coefficients are required.

DATA SET GENERATION SYSTEM

A prototype software system has been developed at LARS for generating and utilizing self-defining data sets. The basic elements of the system are outlined in Figure 3. The first element is a control point location function which estimates the location of known ground control points in uncorrected imagery and generates visual displays for precise location of control in image coordinates. Line printer, graphics dot printer, CRT display, or CRT alphanumeric terminal images can be generated for user interaction. Next, an image distortion modeling function generates parameters for up to fifth-degree polynomials to describe the image geometry. The GEOD function determines other parameters, such as center point geographic coordinates and heading for complete geometric description. The CNVLRS function generates the SDDS format from standard LARSYS 3.1 data tape format. The REGTAP routine is a tape-buffered image transformation routine which accepts a definition of the desired user grid and transforms one or more SDDS input to fill cells in the desired user output grid. The user grid specification includes such parameters as center point latitude and longitude, grid cell size, azimuth, and vertical and horizontal size of data set.

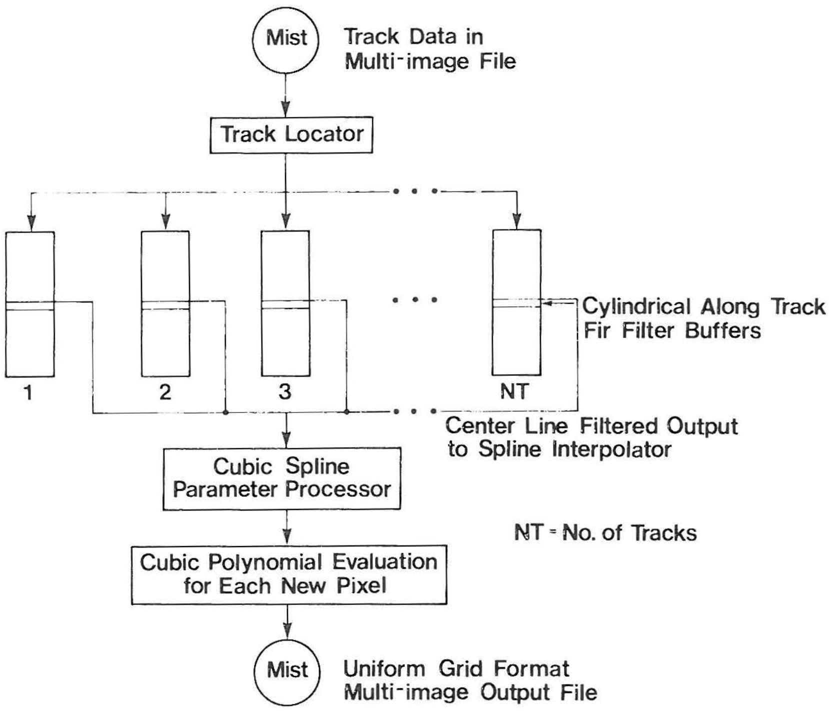

Figure 2. System structure for track-type data interpolation and filtering algorithm.

This system is experimental and exists only in developmental software for generation of research data sets. Further development is needed for it to be user-level software.

EXAMPLE MULTI-TYPE DATA SET

A high-dimensionality multi-type data set was required for a research project on remote sensing techniques for mineral exploration. For one test site in southeast Arizona, U.S.A., a variety of geophysical and other data types were available. The geophysical data consisted of airborne gamma ray and magnetic data gathered along tracks spaced 4.8km apart at an altitude of 137 meters. The track-type interpolation algorithm was used to generate an image data set with a grid cell size of 152 meters for an area approximately 254km square centered near San Manuel, Arizona. Landsat MSS data was transformed and placed in the user grid from frames which closely relate to the geophysical data in time of collection. Ancillary data in the form of geology maps was digitized and placed in SDDS format and transformed to the user grid. The topographic elevation data was obtained from the U.S. Geological Survey in digital form for 1:250,000 scale map quadrangles. This data is in effect in SDDS format since it matches the area and projection of the source map and requires only a linear transformation to place it into the user grid. The data set has the channels listed in Table 2.

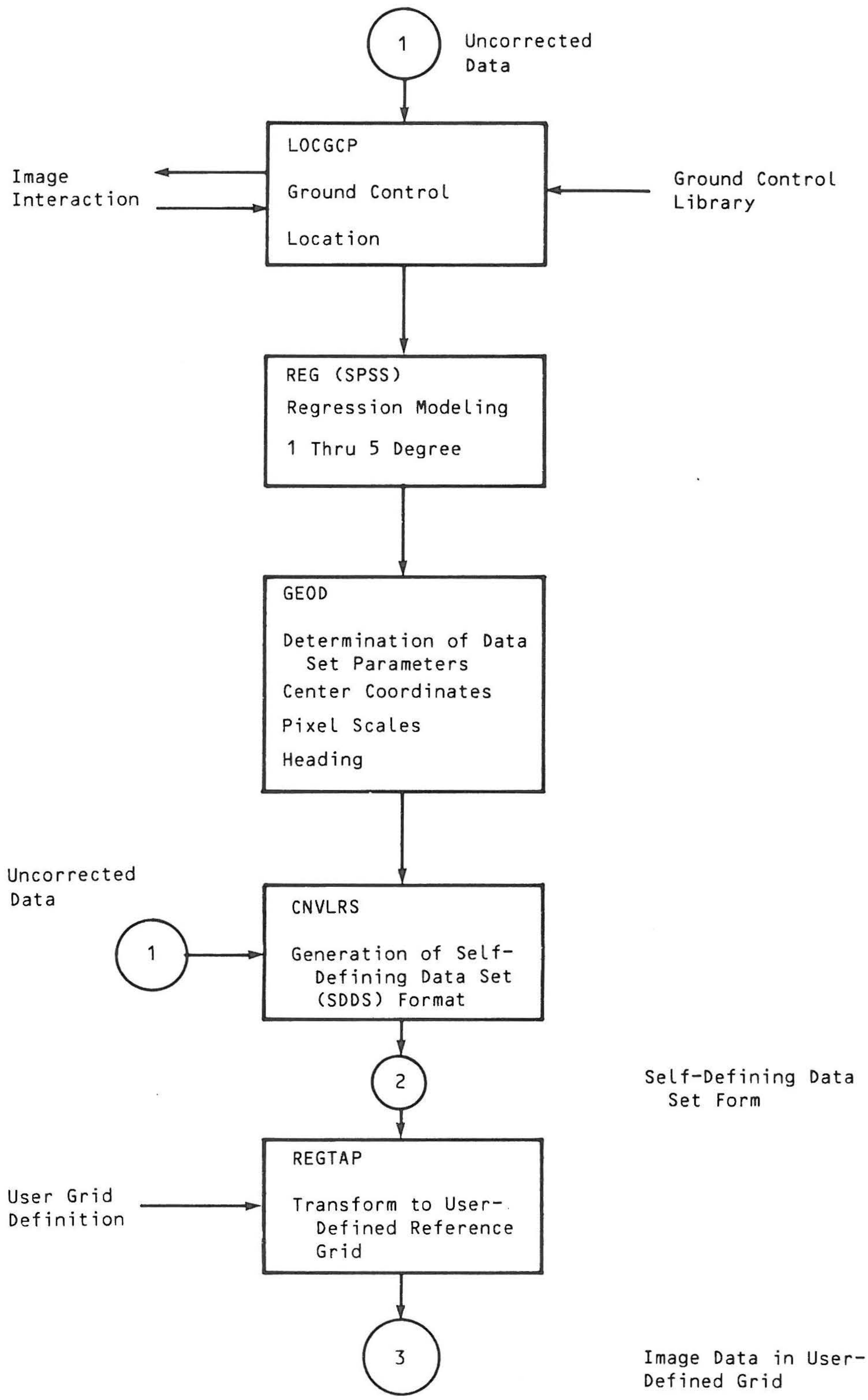

Figure 3. Prototype software system for generating and merging self-defining data sets.

Table 2. Contents of a Multi-Type Remote Sensing Data Set

Channel	Data Type
1	Total Field Magnetics
2	Uranium Gamma Ray Count (Bi^{214})
3	Thorium Gamma Ray Count (Th^{208})
4	Potassium Gamma Ray Count (K^{40})
5	Uranium/Thorium Ratio
6	Uranium/Potassium Ratio
7	Thorium/Potassium Ratio
8	Landsat Band 4 (.5-.6 μm)
9	Landsat Band 5 (.6-.7 μm)
10	Landsat Band 6 (.7-.8 μm)
11	Landsat Band 7 (.8-1.1 μm)
12	Digitized Geology Units
13	Topographic Elevations
14	Slope Derived From Elevation
15	Aspect Derived From Elevation

Detailed statistical and classification analysis has been carried out on this data set for mineralization potential. Various discriminant functions were developed which isolated anomalous areas in the study site. Further discussion of the applications deriving from the multi-type data set is beyond the scope of this paper and further information can be obtained in reference 6.

CONCLUSIONS

The generation and utilization of multi-type self-defining data sets has been briefly discussed. The paper describes a development effort and a final system has not been created. The main point of the paper is that creation of a standard SDDS format for remote sensing data will enable users to conveniently and relatively automatically assemble a variety of data types into an arbitrary grid system defined by the user at assembly time. The SDDS form is currently being created via the Landsat fully corrected "P" tape generated by the NASA MDP system in the U.S.A. We here extend this requirement to aircraft data, other satellite data, ancillary data and other data types which the user might want to assemble in his grid quickly and at low cost.

An ultimate system can be envisioned for the example presented here in which Landsat data would all be stored at the EROS Data Center, Sioux Falls, South Dakota, the geophysical data all at another center -- say, at Grand Junction, Colorado, and all topographic data at the USGS Center in Reston, Virginia. For example, a user at any suitably equipped center desires to assemble the data set of Table 2. He types in the data types, dates, and user-grid definition on a computer terminal and the appropriate data is retrieved from tape libraries and transmitted via satellite to the user facility. The user system assembles the data and writes the multichannel data set on tape and flashes a message indicating the process is complete in a matter of minutes. This is in essence the concept presented here. We hope progress can be made in this direction since the cost and time involved in currently assembling such data sets are prohibitive and this acts to defeat utilization of the data. The SDDS concept should be part of such a development.

ACKNOWLEDGEMENTS

The work discussed in this paper was sponsored by the National Science Foundation under Grants ENG7614400 and ENG7820466 and by the National Aeronautics and Space Administration under Contract NAS9-15466. Appreciation is extended to Mr. Timothy Grogan who did the programming for the developmental system.

REFERENCES

1. Price, R. D., "NASA End-to-End Data System Program," in Progress in Astronautics and Aeronautics, Vol. 67, Remote Sensing of Earth From Space: Role of "Smart Sensors," pp. 155-165, published by AIAA, 1979.

2. Anuta, P. E., "Computer-Assisted Analysis Techniques for Remote Sensing Data Interpretation," Geophysics, Vol. 42, No. 3, April 1977, pp. 468-481; also Laboratory for Applications of Remote Sensing, LARS Information Note 100675, Purdue University, West Lafayette, Indiana, U.S.A., 47907.

3. Anuta, P. E., "Generation of Multitype Remote Sensing Data Sets for Geophysical Exploration," Poster paper presented at 6th Annual Pecora Symposium, Sioux Falls, South Dakota, April 16, 1980.

4. NASA Reference Publication 1039, "Synthetic Aperture Radar/Landsat MSS Image Registration," National Aeronautics and Space Administration Report RP-1039, June 1979. Available from NTIS, Springfield, Virginia.

5. Anuta, P. E. and McGillem, C. D., "A Two-Dimensional Filter Design for Isotropic Reconstruction of Track-Type Airborne Geophysical Surveys," Proceedings of 1979 IEEE Acoustics, Speech and Signal Processing Conference, April 2-4, 1979, Washington, D.C.

6. L'Heureux, D. M.; Levandowski, D. W.; and Anuta, P. E., "Exploration for Porphyry Copper Deposits Through Multivariate Analysis of Computer-Registered Landsat and Geophysical Data Sets," presented at 6th Pecora Symposium, Sioux Falls, South Dakota, April 16, 1980; also to be published as LARS Technical Report by Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907.