INTEGRATION OF MULTI-SOURCE DATA IN MINERAL EXPLORATION[1]

Knut Conradsen
Bjarne Kjær Ersbøll
Allan Aasbjerg Nielsen

IMSOR – Institute of Mathematical Statistics and Operations Research
Technical University of Denmark, Building 321
DK–2800 Lyngby, Denmark

ABSTRACT

This paper describes several multivariate statistical analysis applications of geochemical, geophysical and spectral variables in mineral exploration. **Mahalanobis' distance** is described in some detail and based on four multi-source variables this measure is applied to produce a map that gives an expression of the statistical proximity of each point in the map to a mineralized area. The four multi-source variables chosen from a much larger set of variables have all been subject to extensive data processing: the geochemical variable is the noise MAF (**minimum/maximum autocorrelation factor**) of eleven **kriging** interpolated stream sediment variables; the geophysical variables are kriged aeromagnetic data iteratively moving average corrected to minimize the flight line striping and kriged Bouguer gravity anomaly data corrected for a quadratic trend; and the spectral variable is the density of automatically generated linear features based on Landsat TM data. The results indicate among other things a not previously recognized subsurface continuation of an already mapped lineament.

1.0 INTRODUCTION

In order to combine irregularly sampled spatial information such as geochemical and geophysical data with information given on a spatially regular grid such as the spectral data obtained from the Landsat MSS/TM or the SPOT HRV scanners the former can be brought onto a regular grid. One way of doing this is to use the geostatistical prediction method **kriging** (Journel and Huijbregts, 1978) which provides a best linear unbiased estimator (BLUE) based on a model of the empirical **semi-variogram** of the original data. The semi-variogram describes the **spatial autocorrelation** of the data. Kriging has two advantages over other estimation schemes, namely that it takes the spatial autocorrelation of the data into account, and that it provides you with an estimation variance for each estimated value.

When using several variables in each of the geochemical, geophysical and spectral data groups one might want to reduce the number of variables while still retaining most of the over-all information. Methods suited for this purpose include the generation of **orthogonalized variables** such as principal component scores, factor scores and minimum/maximum autocorrelation factor scores

[1] Presented at the Eighth Thematic Conference on Geological Remote Sensing, Denver, Colorado, USA, April 29 – May 2, 1991.

(MAFs – Switzer and Green, 1984). The MAFs have the advantage over most other orthogonalized variables that they take the spatial autocorrelation structure of the data into consideration. The MAFs separate the spatial variation in a multichannel image such that the highest order MAF describes the low frequency phenomena (signal) and the lowest order MAF describes the high frequency phenomena (noise).

Naturally, all data on a regular grid can be used directly in a geological interpretation. They can also be used as variables in multivariate statistical analyses of combined data sets, i.e. data sets including geochemical and geophysical data as well as spectral information from e.g. Landsat TM. The nature of the multivariate analysis will obviously depend on the nature of the problem and the questions posed. Below **Mahalanobis' distance** is used as an example of one way of performing multivariate statistical analysis on multi-source data. Mahalanobis' distance is a measure of the squared distance—that allows for any correlations that might exist between variables—for an observation to the center of the distribution of, say, a number of interesting variables in a mineralized training area. The work reported here constitutes part of a larger joint project between Minas de Almadén y Arrayanes S.A., Almadén, Spain, Department of Geology, Trinity College, Dublin, Ireland, and the Institute of Mathematical Statistics and Operations Research (IMSOR), Technical University of Denmark, Lyngby, Denmark. A full description of the results of this joint project is found in Conradsen et al. (1987).

2.0 MULTIVARIATE STATISTICAL TECHNIQUES

This section describes two multivariate statistical techniques used in the case study below. Mahalanobis' distance is described in some mathematical detail whereas the minimum/maximum autocorrelation factors are described verbally only.

2.1 MAHALANOBIS' DISTANCE

Mahalanobis' distance is a measure of the separability of the mean values of two Gaussian distributions. Consider the following two Gaussian distributions:

$$X \in N(\mu_1, \Sigma) \qquad Y \in N(\mu_2, \Sigma)$$

which means that X will follow a Gaussian distribution with mean vector μ_1 and variance-covariance structure Σ. Y on the other hand will follow a Gaussian distribution with mean vector μ_2 and variance-covariance structure Σ. Note that the variance-covariance structure Σ is the same for the two distributions. The elements of X and Y each describe measurements of a number of natural variables or derivatives hereof, e.g. content of zink, value of second principal component etc.

The mean vectors and variance-covariance structure are estimated in the usual way. Given the observations

$$X_1, X_2, \ldots, X_{n1} \qquad Y_1, Y_2, \ldots, Y_{n2}$$

which are known to come from the respective above distributions, we can then estimate $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\Sigma}$. We now define Mahalanobis' distance as:

$$D^2 = (\hat{\mu}_1 - \hat{\mu}_2)'\hat{\Sigma}^{-1}(\hat{\mu}_1 - \hat{\mu}_2)$$

We note that if $\Sigma = I$ (i.e. the identity matrix) then D^2 reduces to the squared Euclidian distance between $\hat{\mu}_1$ and $\hat{\mu}_2$. Mahalanobis' distance may be regarded as the squared norm of $\hat{\mu}_1 - \hat{\mu}_2$ with respect to $\hat{\Sigma}^{-1}$, written as

$$D^2 = ||\hat{\mu}_1 - \hat{\mu}_2||^2_{\hat{\Sigma}^{-1}}$$

or—put in simple terms—as the squared distance between $\hat{\mu}_1$ and $\hat{\mu}_2$ using $\hat{\Sigma}$ as a "measuring stick."

We now regard the special case where $n_2 = 1$ i.e. we have only one observation from the second Gaussian distribution. This case corresponds to the situation in this paper, where we are interested in determining the proximity of an arbitrary observation to some reference distribution. The reference distribution is known to represent some interesting class (e.g. observations known to come from mineralized areas).

As an illustration of the concept consider a simple example: we have already computed $\hat{\mu}_1$ and $\hat{\Sigma}$ from a training sample known to be rich in content of some natural variable, and we now wish to determine Mahalanobis' distance of other observations to this reference distribution

$$\hat{\mu}_1 = \begin{bmatrix} 124 \\ 82 \end{bmatrix} \qquad \hat{\Sigma} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

The distribution is depicted in Figure 1 with the mean vector shown as a bullet (•) and variance-covariance matrix represented as the ellipse drawn with the solid line. The Mahalanobis' distances $D^2 = 1, 4, 10, 20$ are also shown as ellipses, but now using dashed lines. The square (□) and the circle (○) both have the same Mahalanobis' distance (= 10) to the bullet (•). The Euclidean distances are 5.1 and 2.0 respectively. In directions where separability is good (e.g. towards the circle, ○) we use a small "measuring stick" according to the variability in that direction and vice versa. This feature also overcomes the difficulties Euclidean measures have with difference in scale of the variables.

Mahalanobis' distance is described in most textbooks on multivariate statistical analysis cf. e.g. Anderson (1984).

2.2 MINIMUM/MAXIMUM AUTOCORRELATION FACTORS

An often applied technique for information extraction and compression in multivariate analysis of (image) data is the principal components (PRC) transform. The PRC transform is described in most textbooks on multivariate statistical analysis cf. e.g. Anderson (1984). The application of this transform requires knowledge of or an estimate of the sample covariance matrix. **The PRCs maximize the variance represented by each component.** PRC one is the linear combination of the original bands that explains the maximum amount of variance in the original data. A higher order PRC is the linear combination of the original bands that explains the maximum amount of variance subject to the constraint that it is orthogonal to lower order PRCs. The most notable objection to PRC analysis is that it performs a pixel-wise operation that does not take the spatial structure of an image into account.

As opposed to the principal components transform the minimum/maximum autocorrelation

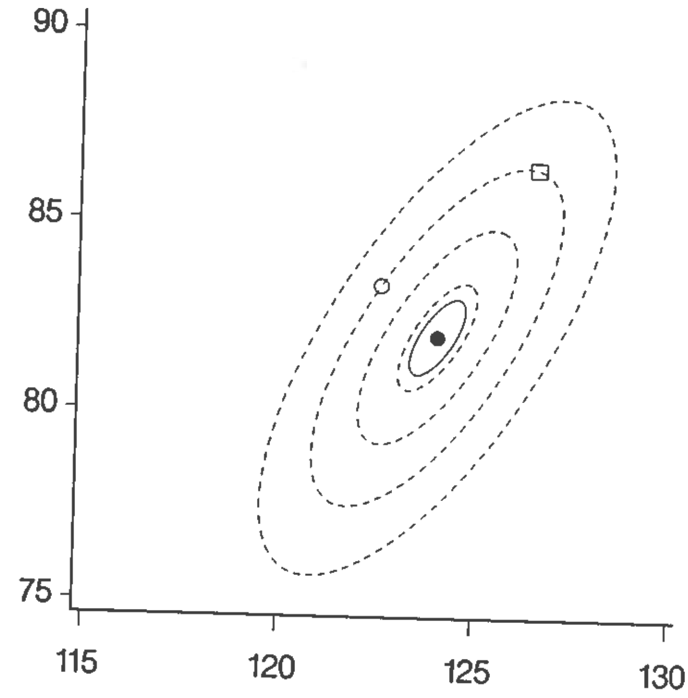

Figure 1: Illustration of Mahalanobis' Distance.

factors (MAF) transform allows for the spatial structure of the image data. The MAF transform was introduced by Switzer and Green (1984). The application of this transform requires knowledge of or an estimate of the sample covariance matrix and the covariance matrix of the spatially shifted image—hereby giving the neighbourhood information. The MAF transform minimizes the autocorrelation rather than maximizing the data variance (PRC). **In reverse order the MAFs maximize the autocorrelation represented by each component.** MAF one is the linear combination of the original bands that contains minimum autocorrelation between neighbouring pixels. A higher order MAF is the linear combination of the original bands that contains minimum autocorrelation subject to the constraint that it is orthogonal to lower order MAFs. The MAF procedure thus constitutes a (conceptually) more satisfactory way of orthogonalizing image data than PRC analysis. The MAF transform is equivalent to a transformation of the data to a coordinate system in which the covariance matrix of the spatially shifted image data is the identity matrix followed by a principal components transformation. Another important property of the MAF procedure is its invariance to linear transforms, a property not shared by ordinary PRC analysis. This means that it doesn't matter whether the data have been scaled e.g. to unit variance before the analysis is performed.

The PRCs, the MAFs and other orthogonal transforms are described in Conradsen and Nielsen (1986).

3.0 CASE STUDY – CENTRAL SPAIN

The data used to illustrate the above techniques comes from a 30×20 km^2 test area in central Spain. Mercury has been mined in this area since Roman times. Four variables representing geochemical, geophysical and spectral data chosen from a larger data set consisting of 126 variables are used.

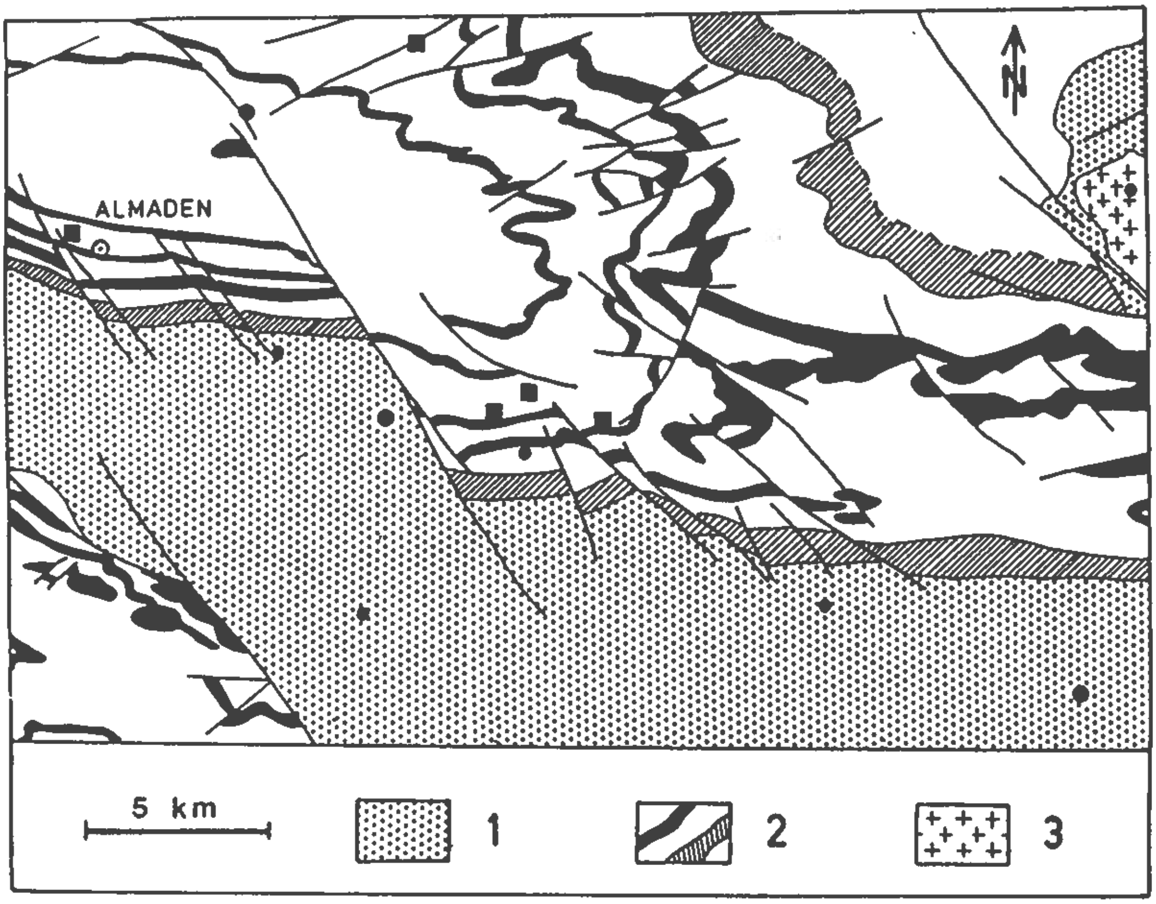

Figure 2: Simplified Geological Map of Test Area.

A simplified geological map of the test area is shown in Figure 2, legend: (1) late Precambrian sediments, (2) lower Palaeozoic volcanics and sediments with outlined quartzite horizons, (3) Hercynian granite, (black squares) mercury occurrences, and (black bullets) lead-zinc-silver occurrences.

First, 2,970 stream sediment samples were collected in the test area. The samples were analyzed by Inductively Coupled Plasma emission spectrometry (ICP) for the content of sixteen elements: Pb, Zn, Cd, Hg, Cu, Ba, Mn, Ni, Co, Cr, Sn, W, Mo, V, Sb and Ag. The distributions of Cu, Sn, Mo, Sb and Ag are strongly skewed and these elements are not included in the multivariate analyses. The remaining eleven elements also show skewed distributions but transforming the concentration values by the natural logarithm gave satisfactory histograms.

Experimental semi-variograms were estimated for the eleven elements and parameters **nugget effect** and **range of influence** were estimated based on non-linear least squares fits to (double) spherical semi-variogram models. The semi-variogram describes the spatial autocorrelation of the data. The eleven elements were kriged to a 200×200 m^2 grid. Kriging provides a best linear unbiased estimator (BLUE) based on the semi-variogram model and the original data. For a thorough description of these geostatistical concepts cf. Journel and Huijbregts (1978). MAF analysis was performed on the eleven kriged variables. The "noise" MAF (MAF1)—which in the case of kriged (and thus smoothed) data is not really "noise"—reveals a lot of geochemical information and it is therefore chosen as our geochemical variable.

Below this variable is called ICPMAF1.

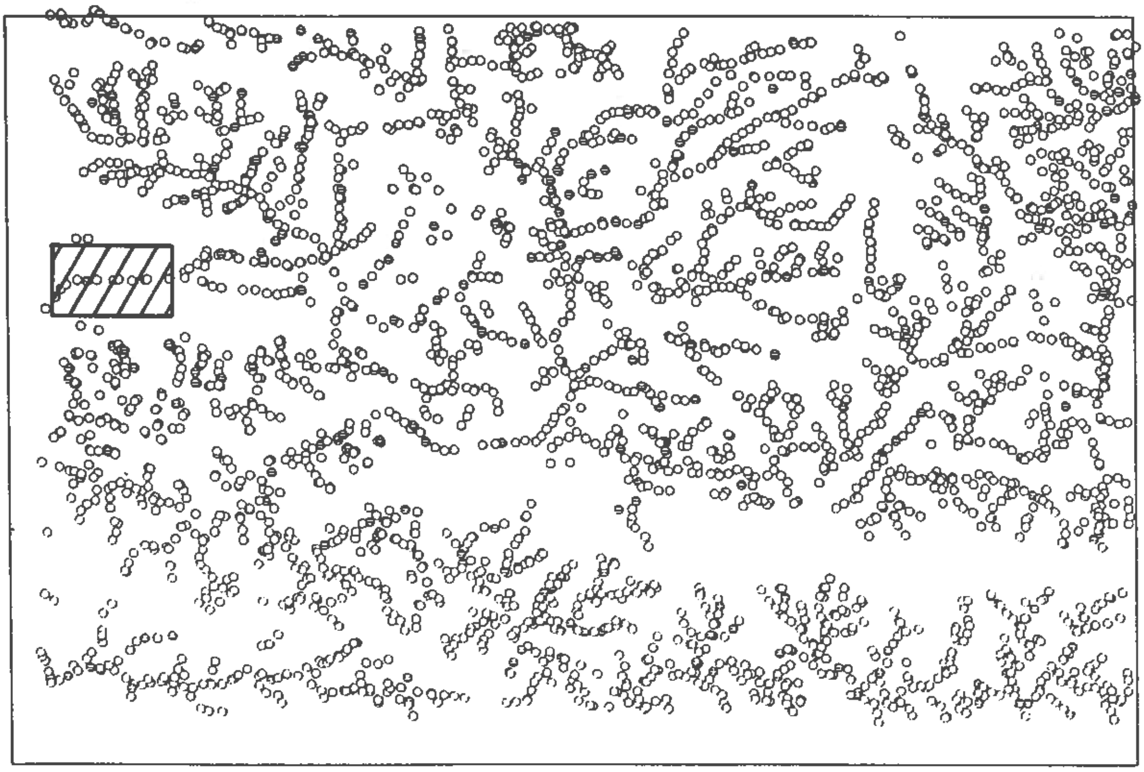

Figure 3: Sample Sites for the Geochemical Data, Training Site Marked.

Figure 3 shows the sample sites for the geochemical data. The training area used for computation of the reference distribution consisting of 112 observations positively known to contain mercury is marked. Figure 4 shows the experimental semi-variogram and a double spherical model with nugget effect fitted by means of a non-linear least squares routine (range of influence is approx. 9 km and nugget effect is approx. 50 %). Figure 5 shows kriged stream sediment ICP analyzed data, Hg. Figures 6 and 7 show MAF1 and MAF11 for eleven kriged geochemical variables.

Second, aeromagnetic data from the test area were collected. The data is strongly influenced by the northeast-southwest flight direction chosen in carrying out the measurements. The striping observed in the raw data reflects calibration problems from flight line to flight line. As no transverse

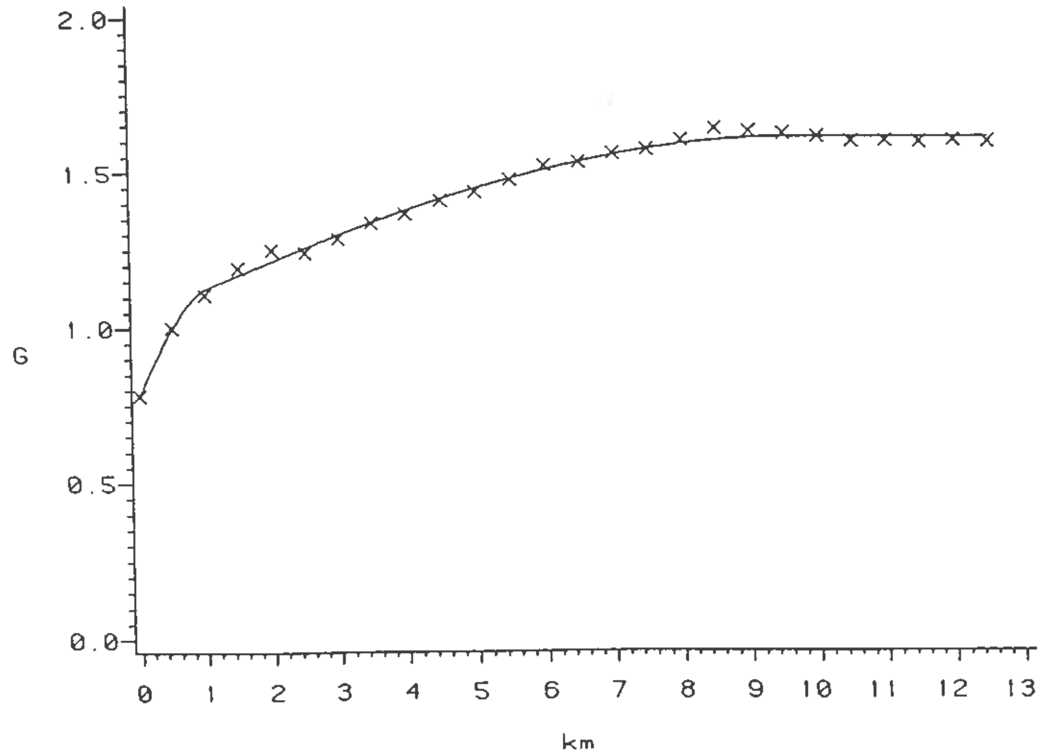

Figure 4: Experimental and Modelled Semi-variogram for Hg.

line was flown iterative moving average operations on flight lines and kriging to the same 200×200 m^2 grid as for the geochemical variables was used to minimize the striping effect.

Below this variable is called Aeromag.

Figures 8 and 9 show kriged raw aeromagnetic data and kriged aeromagnetic data iteratively moving average corrected (to minimize flight line striping).

Third, a Bouguer gravity anomaly survey was performed in the test area. Residual data from 539 stations were kriged to the 200×200 m^2 grid after subtraction of a quadratic trend surface. While this trend may reflect some large scale phenomena the residual values most likely reflect local, shallow geological features.

Below this variable is called Bouguer.

Figures 10 and 11 show kriged raw Bouguer gravity anomaly data and kriged Bouguer gravity anomaly data after subtraction of a quadratic trend surface.

Fourth, based on Landsat TM band 5 summer data a procedure for automated detection of linear features was applied. The basic idea behind the procedure is this: a linear feature in an image represents some kind of irregularity with respect to the surrounding pixels. A measure of the

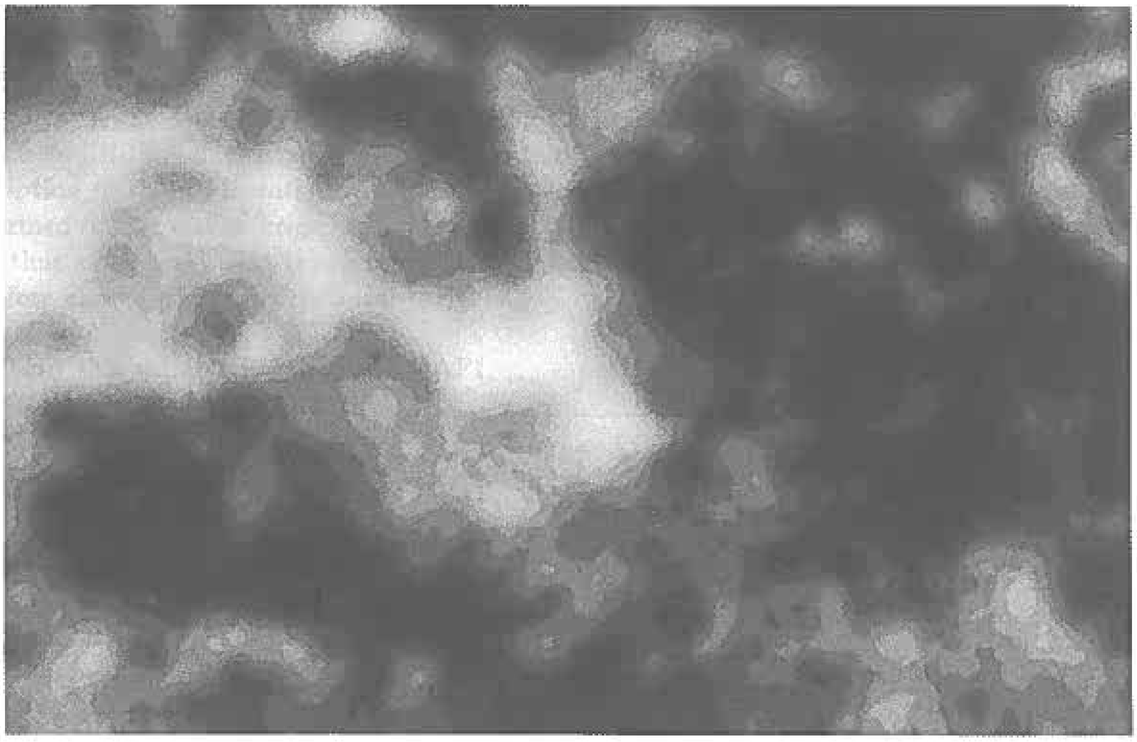

Figure 5: Kriged Stream Sediment Geochemistry, Hg.

magnitude of this irregularity is the difference between the actual pixel value at a given point and a prediction of that pixel value based on the remaining pixel values. The procedure filters the image by a data dependent filter. The filtered image is binarized to enhance the largest residuals and skeletonized to reduce the width of the linear structures to one pixel. Based on this skeletonized image a local orientation in each pixel on the skeleton is estimated by the use of a moving window. A directional sector chosen for production of lineament density maps can be histogram based or it can be chosen as a division of the compass rose into sectors; here 15° sectors are chosen. The sector used below is northeast-southwest (45° to 59° in mathematical notation).

Below this variable is called A045059.

Figures 12 and 13 show Landsat TM band 5 summer data and an automatically derived density map for linear features oriented northeast-southwest (45° to 59° in mathematical notation).

Figure 14 shows the Mahalanobis' distance map based on the features described above. It is noteworthy that the indicated lineament extends further south than depicted in the geological map in Figure 2.

For a thorough description of the automated lineament detection procedure cf. Conradsen and Nilsson (1987) and Conradsen et al. (1987).

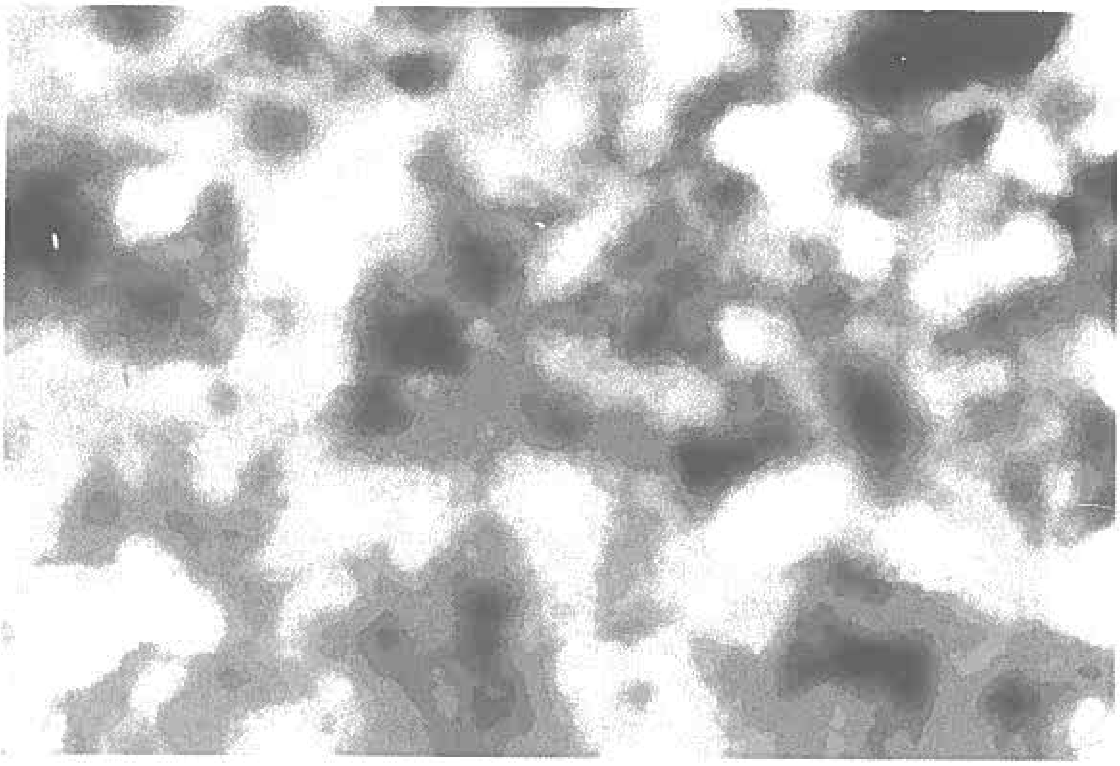

Figure 6: MAF1 of Kriged Stream Sediment Geochemistry.

Figure 7: MAF11 of Kriged Stream Sediment Geochemistry.

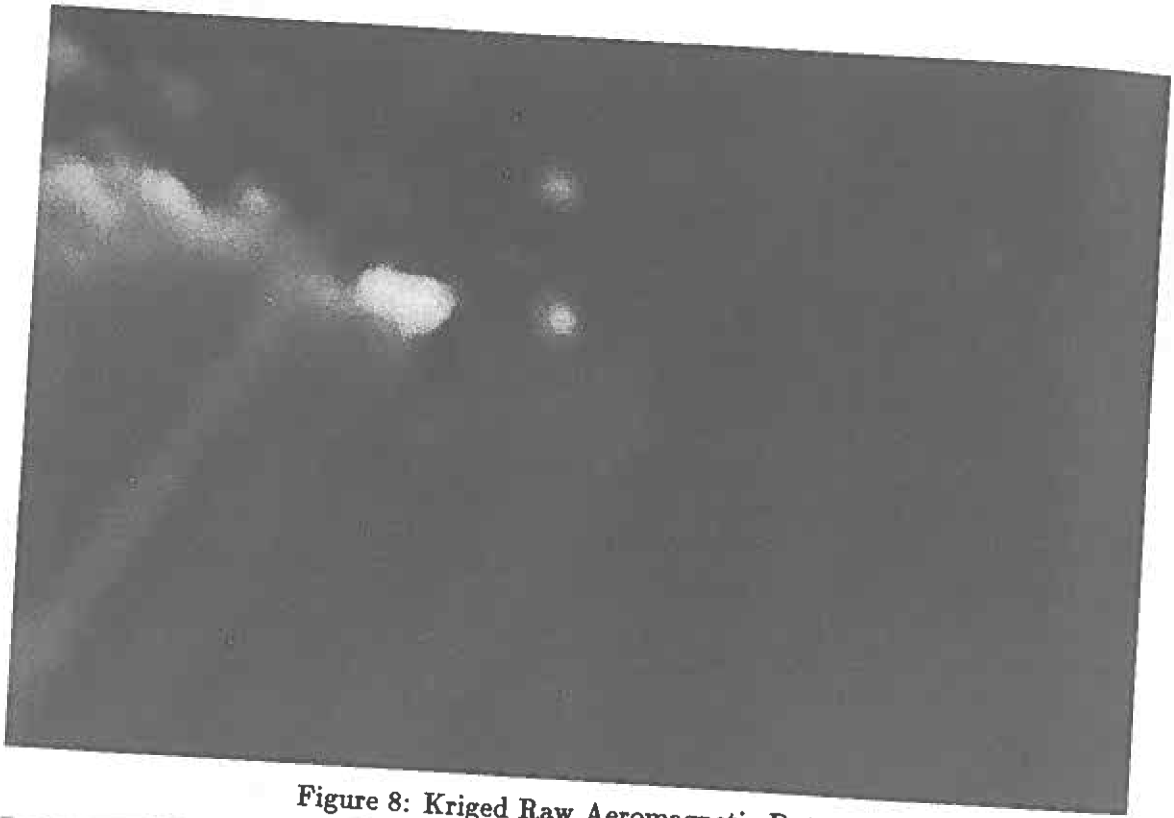

Figure 8: Kriged Raw Aeromagnetic Data.

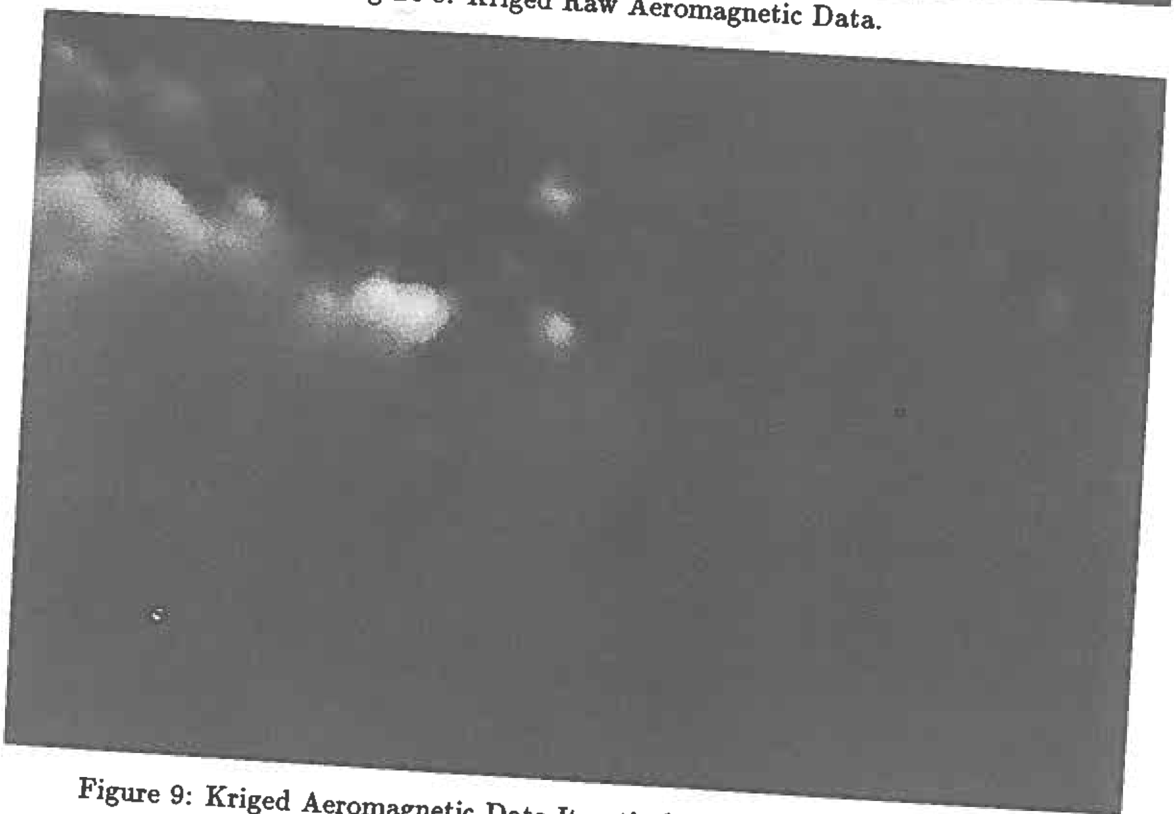

Figure 9: Kriged Aeromagnetic Data Iteratively Moving Average Corrected.

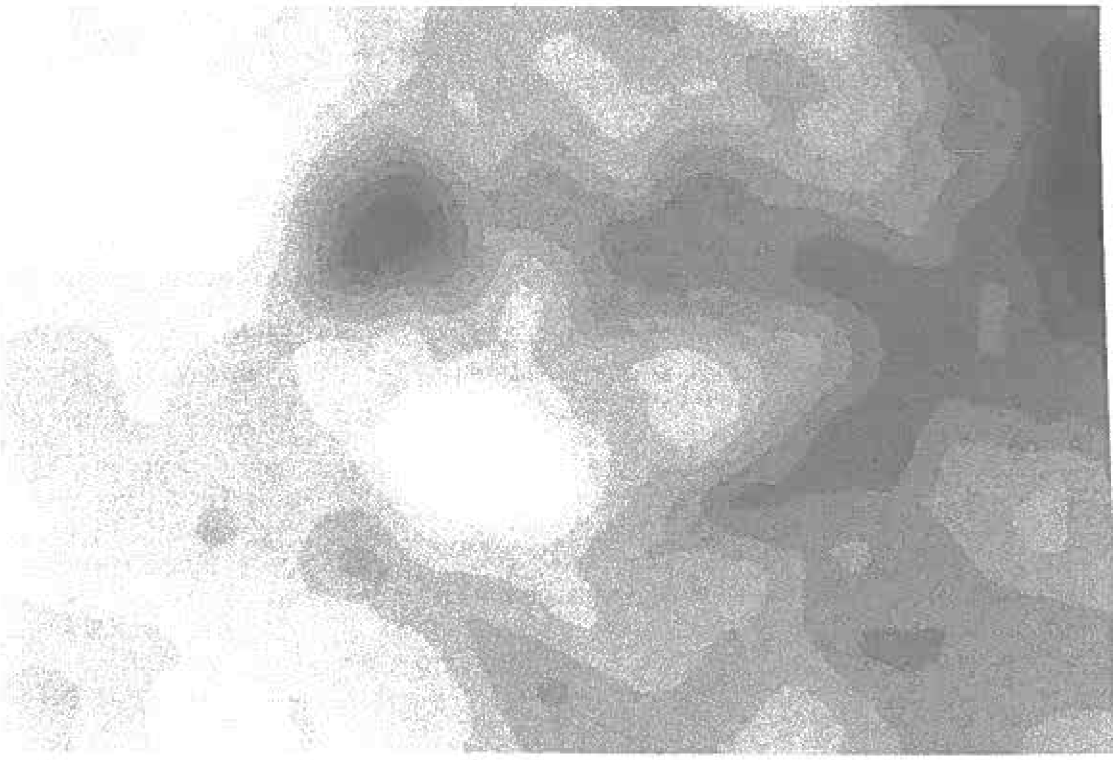

Figure 10: Kriged Raw Bouguer Gravity Anomaly Data.

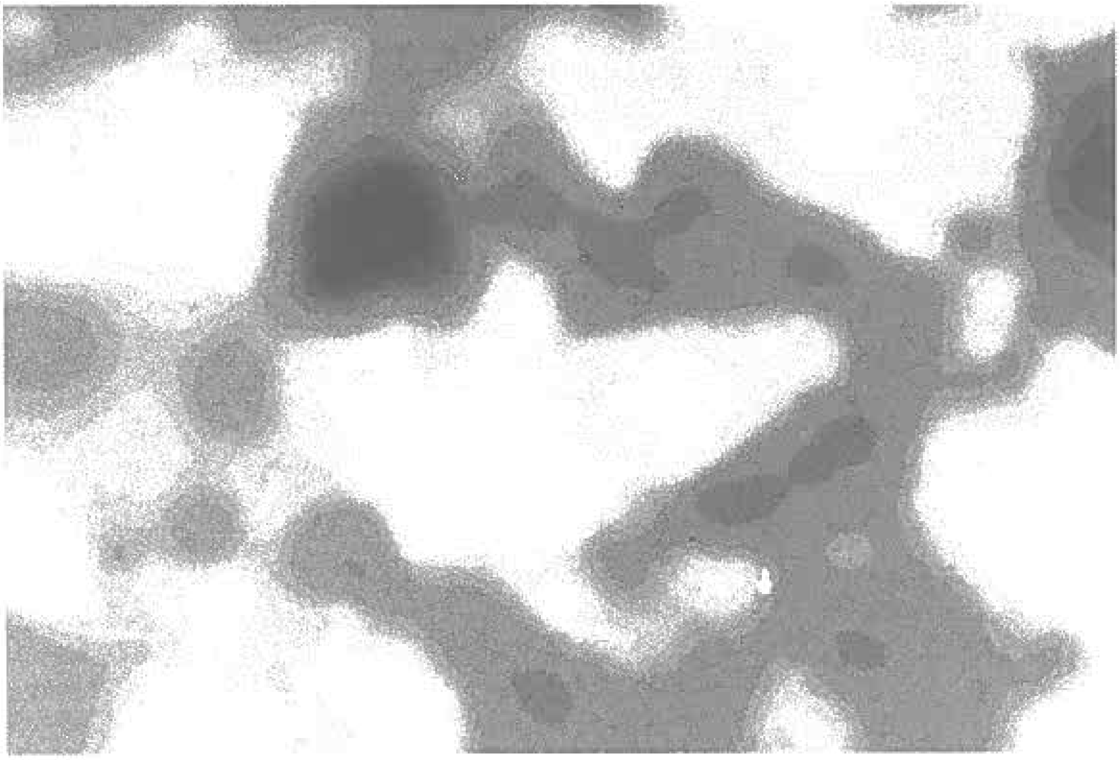

Figure 11: Kriged Bouguer Gravity Anomaly Data, Quadratic Trend Subtracted.

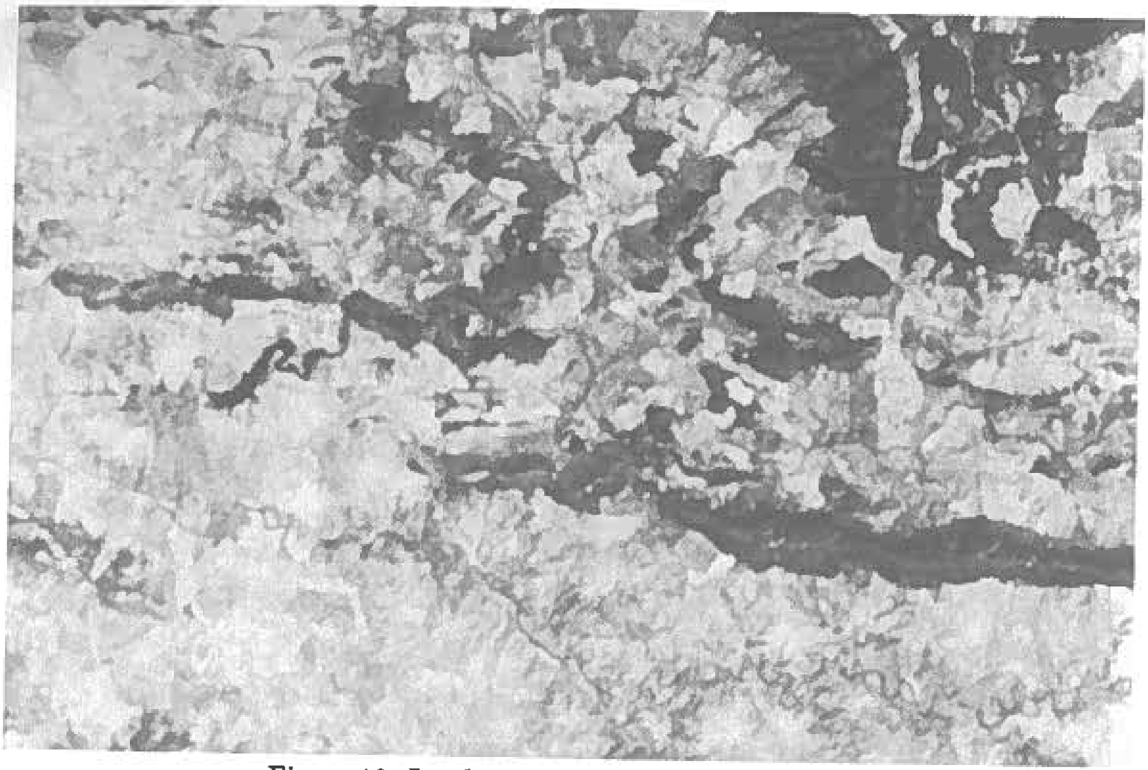

Figure 12: Landsat TM Band 5, Summer Data.

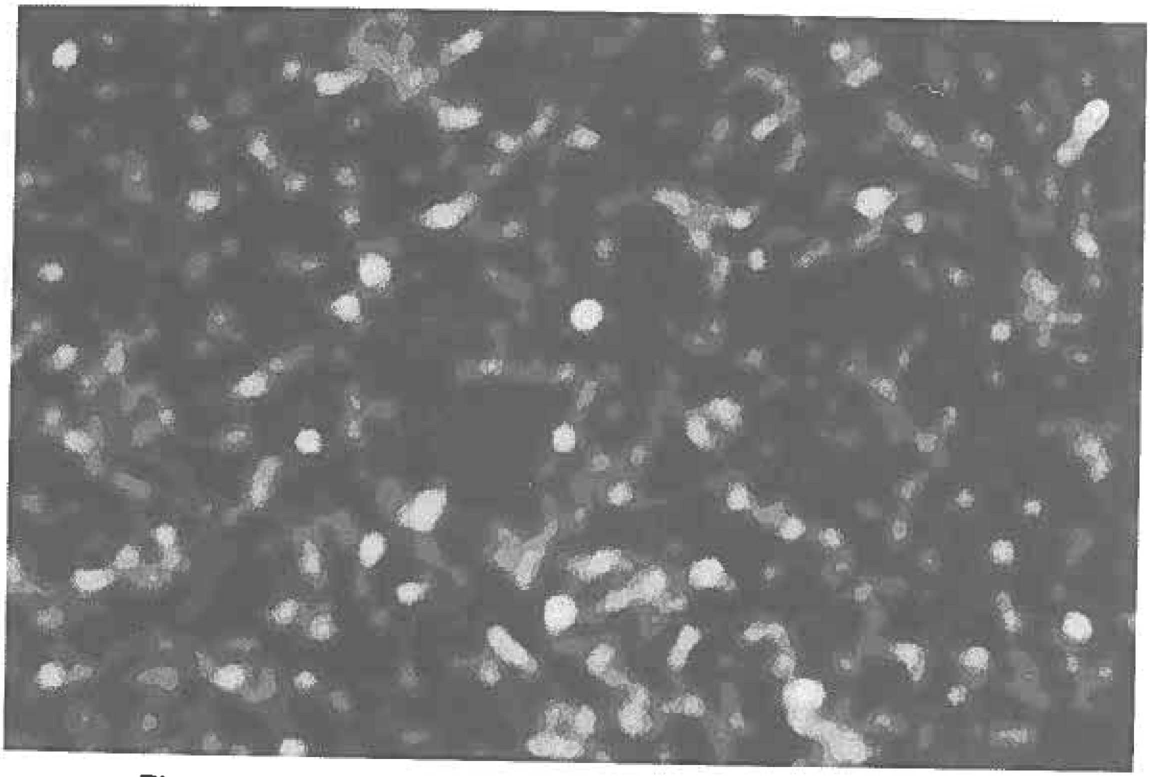

Figure 13: Automatically Derived Density Map for Linear Features.

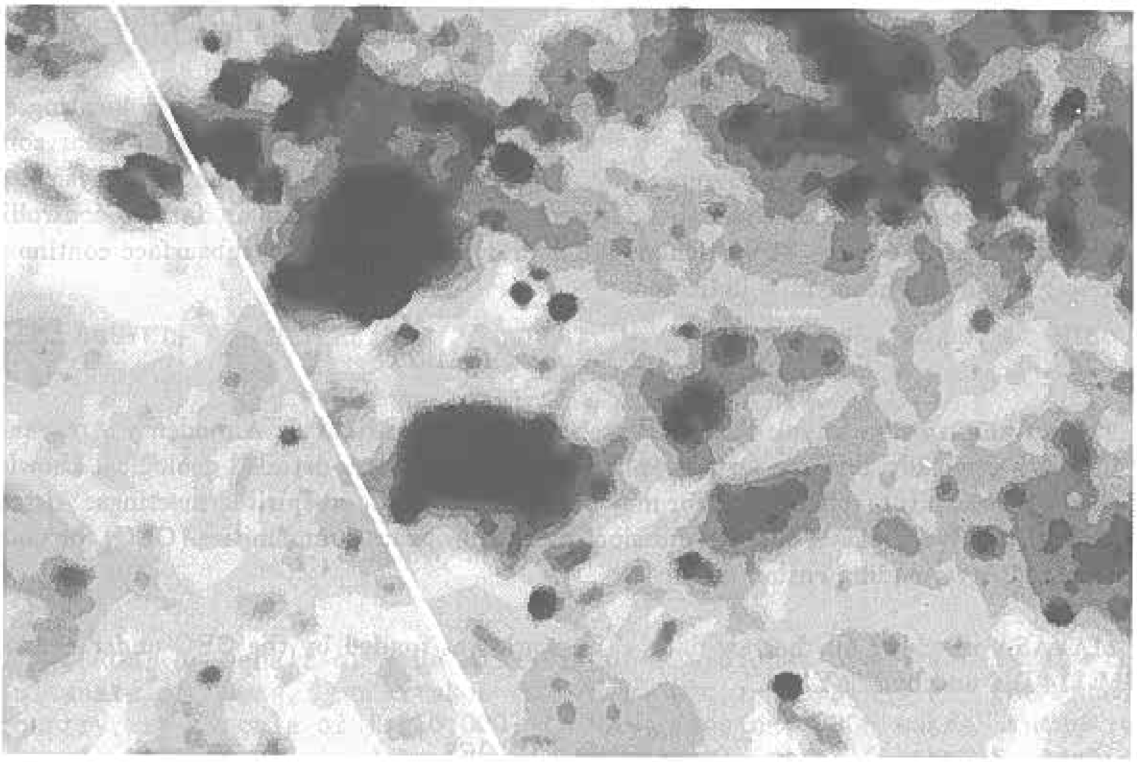

Figure 14: Mahalanobis' Distance Map with Lineament Indicated.

More data including radiometric information on content of radioactive material (U, Th and K) is available in this area but not used in the analysis reported here.

Table 1 shows mean vector and covariance/correlation matrices for the four variables chosen estimated in a small (112 pixels) area centered around the town of Almadén situated in the top left corner of the test area and hosting the oldest mercury mine in the world. The covariance/correlation matrices are given as upper/lower triangle matrix respectively. The actual numbers are scaled so that all variables are stretched linearly between 0 and 100.

Table 1: Mean Vector and Covariance/Correlation Matrices

	A045059	Bouguer	ICPMAF1	Aeromag
Mean	7.049	58.910	24.610	23.830
A045059	41.260	−2.365	5.903	7.429
Bouguer	*−0.065*	32.200	−10.990	13.730
ICPMAF1	*0.075*	*−0.158*	150.500	−11.180
Aeromag	*0.305*	*0.638*	*−0.240*	14.400

4.0 CONCLUSIONS

The multivariate statistical techniques illustrated hold a strong potential in locating areas of interest in mineral exploration. In the test area used mineralizations linked to shear zones and related to granitic bodies are recognized. The application of directional and geophysical data in the Mahalanobis' distance analysis therefore ensures a link to the main factors controlling the mineralizations in the area. In particular, a not previously recognized subsurface continuation of an already mapped lineament is indicated.

ACKNOWLEDGEMENTS

The authors wish to thank the Geology Department of Minas de Almadén y Arrayanes S.A. (MAYASA) especially Enrique Ortega and Jesús Artieda whose detailed geological knowledge of the test area was invaluable; thanks for many interesting and good-spirited meetings. Also, thanks to Leopold van Wambeke of the Commission of the European Communities (CEC) for many suggestions and never-ending enthusiasm.

MAYASA provided all the raw data. The work was funded by the CEC under contract No. MSM-114-DK and by MAYASA.

REFERENCES

T.W. Anderson (1984): *An Introduction to Multivariate Statistical Analysis.* 2nd Edition. John Wiley and Sons, 675 pp.

K. Conradsen and B.K. Nielsen (1986): *The Use of Some Data-dependent Orthogonal Transformations in the Analysis of Multichannel Imagery.* Technical Report, IMSOR, 68 pp.

K. Conradsen and G. Nilsson (1987): Data Dependent Filters for Edge Enhancement of Landsat Images. *Computer Vision, Graphics and Image Processing,* Vol. 38, No. 2, pp. 101-121.

K. Conradsen et al. (1987): *The Application of Remote Sensing and Data Integration as an Aid to Mineral-Exploration in the Almadén Region.* Minas de Almadén y Arrayanes S.A., 354 pp.

A.G. Journel and Ch.J. Huijbregts (1978): *Mining Geostatistics.* Academic Press, London, 600 pp.

P. Switzer and A.A. Green (1984): *Min/Max Autocorrelation Factors for Multivariate Spatial Imagery.* Technical Report No. 6, Department of Statistics, Stanford University, 10 pp.