Archeologia e Calcolatori
8, 1997, 55-65

SPATIAL STATISTICAL ANALYSIS APPLIED TO MAGNETOMETRIC ARCHAEOLOGICAL DATA

1. Introduction

The application of magnetometric techniques in underground survey is, at the present time, an element of great importance in prospections that are carried out in different disciplines: geophysics, mining, archaeology, etc., although such characteristics of each of these research fields decide a specific methodology of use.

Normally, sciences that use data of magnetometric surveys are focused on macrospatial studies, attempting to locate big bodies of material that spread over large areas. In this way, the characteristics decide the methodology of application, the type of data obtained and the analitical methods. However, the study of underground archaeological structures is characterized by microspatial scale, complexity of structures, type of material and shallowness of the location.

We propose, in this paper, an analytical method of geomagnetic data focused on spatial statistics. We will compare this method with the normally used techniques in archaeological applications. This methodology has been applied to data obtained in the magnetic survey of the Roman site of Las Gabias (Andalucía, Spain), over an area of 2000 m^2. The field data are collected during two months by means of the gradient method using 1 m of grid separation.

2. Analytical treatment of data

The analysis of data obtained in the magnetic prospections has, normally, carried out by the application of two different methodologies: filtering (mainly linear filtering) and data reduction to the pole:

A) *Filtering techniques* are based on transforming spatial data in a spatial frequencies domain for further treatment, employing methods of Image Analysis (contour filtering, band-pass-filtering, etc.). One of the most used filtering techniques transforms data with a Fourier transform obtaining a two-dimensional array of complex numbers. This matrix is multiplied by an appropriate matrix and an inverse Fourier transform is applied to the result (Bhattacharyya, Navolio 1976; Bhattacharyya 1978). These methods are widely used because they are available in commercial software.

B) The method of *reduction to the pole* transforms data in order to remove the influence of the earth's magnetic field inclination. In this way, new data is obtained simulating prospection in the earth's magnetic pole.

This data transformation places the peaks of the anomalies above the center of the underground structure, decreasing the negatives anomalies and emphasizing the positive anomalies, so that the interpretation is easier (BREINER 1973; details in analytical methods appears in GIBERT, GALDEANO 1985). The method is used in those cases where it is important to relate, in an accurate form, the located anomaly to the underground material. However, archaelogists often prefer to obtain magnetometric maps without making the reduction to the pole (SCOLLAR 1990).

3. GRAPHIC TREATMENT OF DATA

Data can be analyzed in graphic form, since structures have complex shapes and pronounced geometric shapes. Generally, data analysis is made by means of methods obtained from other fields of research, so their application in archaeological material presents some inconveniencies. Among these methods the following should be pointed out:

Contour plots. This technique is often used in geological surveys, usually macrospatial studies to locate big body of material. The resultant plot shows magnetic isolines similar to topographic maps, and it is necessary to carry out a smoothing process (for example a low-pass filter, convolution filter averaging the value by means a moving window) with the removal of fine structures. According to SCOLLAR (1990), these techniques are useful for shapes that appear with high contrast and large size and, for archaelogical data, the results are not satisfactory because of the macrospatial character of the method, the smoothing of details and the complexity of archaeological structures. Furthermore the magnitud of magnetic susceptibility values, obtained in an archaeological survey, is lower than that obtained in geological studies (Tav. II, a).

Perspective profiles plot. This method uses the grid design in the prospection area for plotting graphics of the functions $z = f(x)$, where x is a spatial coordinate and z is magnetization value. The graphic representation is preferable in those cases where the signal from substratum is weak, anomalies are large and there is a big contrast with the enviroment (Tav. II, b). One way of making magnetic data stand out is to display the data as a shaded relief on a 3-D graphic. This gives good results in geological surveys (PATERSON, REEVES 1985).

Image displays. Because of the advances in the digital analysis of images, these techniques are applied in the study of data obtained in magnetic prospections. The method treats magnetic data as a bitmap (raster) image. The coordinates x and y of the prospected area are the coordinates of the image while the magnetic field value (or gradient or differential magnetism) is transformed in a gray scale or artificial colour.

The results obtained in the prospection are presented as a two-dimensional matrix (Tav. III, a). Thus, it is possible to transform data to obtain a new matrix with a large number of rows and columns (enlargement) or to reduce the z values of data (data range compression).

The enlargement consists of a digital expansion of the point coordinates (in the original range) to obtain a matrix fitting the raster format (generally 1024x1024 pixels). The z values, for points where there are no original data, are obtained by means of different methods:

– Pixel replication. Each input data value is displayed using a rectangle of pixels. These pixels have the same z value as that of the data. In this way an image with a block-like appearance is obtained.

– Interpolation. Between each point and the nearest neighbour points, new ones are originated by means of interpolating the original values. The results depend upon the interpolation technique used, for example:

a) bilinear interpolation: z value is obtained in each point as the average distance (weighted or not) of the nearest neighbour pixels in the input image; this method causes a loss of high frequencies and a poorly contrasted image (Tav. III, b).

b) biquadratic and bicubic interpolation: uses a fit to polinomical functions of second or third grade. They are very adapted for magnetometrics surveys.

c) separable cubic spline sinusoid function approximation: uses a convolution kernel of 4 x 4 points based in a truncated approximation of the sin (x)/x function. If adquate limits are established, this method avoids computational results and will give good results (Scollar 1990).

– Compression. It consists of reducing the dynamic range of data. It is carried out in cases where data show a large dynamic range with features of interest both small and large, simultaneously present. The general manner is based on transforming the z variable by means of specific functions or methods, e.g.:

a) Arctangent. Reduces the small anomalies around zero level and compresses the larger ones to a maximum of between ± 1.57.

b) Linear transformation. This method can be used if data are obtained by a gradiometer or the general trend has been removed.

c) Mean/standard deviation normalisation to mean=0 and standard deviation=0.5.

d) Median/interquantile difference normalisation.

On the other hand, when the data are in image form, the image processing techniques (gaussian filters, Wallis algorithm, band-pass filters, contrast filters, etc.) can be used. The newly obtained data can be treated with any method mentioned above.

4. Data analysis method: SPATIAL

Using magnetometric techniques for underground studies, the specific characteristics of prospected material have to be taken into account; the following aspects should be considerated to obtain good results:

1) *Limited areas*: the surface seldom has more than 1000 m^2 and the survey normally needs to be carried out to find material of small dimensions.
2) *Structure*: the material forms structures with an important geometric component (by example, linear or circular structures).
3) *Anomaly types*: they usually have a local character and sometimes are complex in shape.
4) *Shallow depth*: the material appears between 0.5 and 2 m depth currently, although there are exceptions that exceed this limit.

The data analysis methods used may be appropriate to these peculiarities, taking into account the characteristics of archaeological settlements. Statistical spatial analysis is a discipline focused on the study of univariant data defined from two spatial variables, x and y. The spatial component is the basic element, that is to say, x and y are the reference variables.

In accordance with the methodology of Statistical Spatial Data Analysis (Cressie 1991; Upton, Fingleton 1995), it is proposed to study data in the following way:

1) Exploratory data analysis by means of elemental descriptive statistics (histograms, frequency polygons and stem-and-leaf diagrams). In this way a primary approximation of data distribution and the necessity of data transforming (compression, expansion, etc.) are obtained.
2) Use of the data set can be: all data, positive data $z>0$ or negative data $z<0$. Though the anomalies are bipolar, the positive and negative anomalies separation, for the purpose of data treatment, is carried out by the following considerations: if the structure has more magnetic susceptibility than environment, it originates a bipolar anomalie with a positive component greater than the corresponding negative component; and conversely. The application of the analytical methods that we propose provides greater emphasys to the positive component, in the first case, or to the negative component, in the other. Thus, it is possible to superimpose the analysis of both cases to obtain a clearer result for interpretation. This method is currently used in magnetometric studies because the zero value constitutes a problem working with continuos positive and negative data simultaneously.
3) Original data transformation, if necessary, by means of the use of mathematical functions in order to obtain greater stability in the data variance. Generally, they are transformations obtained from Exploratory Data Analy-

sis (TUKEY 1977) or from usual statistical analysis:

3.1) Standardization to mean 0 and standard deviation 1 of data defined by means of a quantitative variable. In this way the distortions caused by the existence of extreme values are reduced:

$$Z_i = \frac{z_i - \bar{z}}{n}$$

3.2) Freeman-Tukey transformation for qualitative data:

$$Z_i = (1000 \cdot z_i / n_i)^{1/2} + (1000 \cdot (z_i + 1) / n_i)^{1/2}$$

defined from the frequencies of z_i (CRESSIE 1991). This transformation gives more stability than (FREEMAN, TUKEY 1950):

$$Z_i = (1000 \cdot (z_i + 1) / n_i)^{1/2}$$

3.3) Calculation of data logarithm, obtaining new values better adapted to a Gaussian (normal) distribution:

$$Z_i = \log z_i$$

4) Stem-and-leaf plot: established by TUKEY (1977) as an element of previous study to discern if data show uniformity (that is absence of anomalies), if there are elements or a group of elements with high values showing the presence of anomalies, etc. This study can be carried out in other ways, by means of plotting histograms or frequency polygons. In this way, z value intervals, where anomalies exist, and the interval showing background noise can be found.

The previous points come from nonspatial analysis, that is to say, the location of variable values are not taken into account. From the point of view of spatial analysis, each point is considered as the center of a rectangle where the width equals minimum horizontal distance between points and where the hight equals minimum vertical distance between the same points; although, the software that carries these ideas out allows the selection of the rectangle size.

The spatial analysis includes the following parts:

5) Thematic scheme: where data are represented by means of rectangles of variable sizes. Data distribution is divided into classes of width that can be selected and these classes are represented by means of different shaded tones or colours. A value for the z variable is selected from the options detailed below:

– Z variable is assigned to each rectangle as the sum of the z values of the points located in the rectangle. If rectangles have the minimum size, Z will equal the original z.

- Z variable is assigned to each rectangle with a value equal to the minimum z values of points located in the rectangle.
- Z variable is assigned to each rectangle with a value equal to the maximum z values of points located in the rectangle.
- Z variable is assigned to each rectangle with a value that equals the mean z values.
- Z values can be expressed as relative frequencies with regard to the total data.

Tavv. IV, a-b show the thematic schemes for positive and negative data (the Z variable is the sum of the original variables to locate the areas with most important anomalies).

6) Thematic map shows original data but selecting a class distribution by means of an algorithm designed by the authors in order to detect significant differences between consecutive z values. This algorithm is not spatial and takes into account the total relations in the data. In schematic form, it is as follows:
   a) z values are put in order from the lowest to the highest values and the differences $z_i - z_{i-1}$ are calculated.
   b) A threshold value (can be assigned previously) and a maximum value of differences are selected to begin the iteration. The variable $a_0 = z_1$ is assigned (a_i values show interval limits for distribution of data).
   c) If $(z_i - z_{i-1})$/ maximum > threshold, a new interval $[a_i - a_{i-1}]$ is marked. Its beginning is the end of the previous interval. A new maximum value is obtained.
   d) z_i is assigned to the interval $[a_i - a_{i-1}]$ (if the condition has been fulfilled, the interval will be a new one, otherwise the interval will be the last one).
   e) Steps c and d are repeated until all data is classified in intervals.

The fundamental reason for the algorithm is to find the contrasts existing in data, since archaeologic prospection requires locating the maximum contrast between anomalies and the environment. The value of this contrast is quantified by the threshold variable and the choice of this value allows one to obtain different structure levels. The algorithm detects structures more contrasted with the environment (large anomalies) by applying a high threshold value, although detail is lost. If the threshold value is low, the algorithm detects a great number of structures although some of them can be computational artefacts. The values $1 <$ threshold < 1.5 are the ones that give a high contrast and optimize the results. Tavv. IV, c and V, a show the original data, and Tav. V, b shows the residuals obtained by means of the median-polish algorithm (in this form the local anomalies are located).

Representation of values in each point (x,y) is carried out by assigning the same shade or colour to all points belonging to the same interval.

7) Use of the unweighted median-polish krigging method (Tukey 1977), that gives (details in Kemperman 1984):

$$z(x_i,y_j) = \tilde{a}+\tilde{r}_i+\tilde{c}_j+R(x_i,y_j)$$

where a= total effect (regional trend), r_i= row effect, c_j= column effect, $R(x_i,y_j)$ is the local effect (local trend) and ~ is for the estimations of parameters (Cressie 1991). The study of each of these effects allows one to find the different trends: NS and EW directional, general and local effects, having as a base the coordinates of each rectangle. This algorithm is a robust krigging method using the median for descompose the data into large-scale and small-scale variation.

5. SPATIAL software

The previous mentioned ideas have been realized in our own program, SPATIAL, designed to analyze univariant spatial data. The spatial component is the analysis base and determines the methods used (Cressie 1991). In this way, data given to SPATIAL must include for each element the x and y coordinates (in any coordinate system) and the z variable (Fig. 1).

The program is structured in units that give the following possibilities:

1) *Data transformation*. This unit allows one to work with total data, only positive data or only negative values, so total anomalies or z>0 and z<0 can be studied independently. In magnetometric studies the software works very well with z>0 and z<0 anomalies because the neighborhoods of zero value correspond to noise.

In addition, this part incorporates some of the more used statistical techniques: standardization to mean 0 and standard deviation 1 (Chatfield 1988; Krzanowski 1988), calculation of the logarithm of data and Freeman-Tukey transformation (Freeman, Tukey 1950; Cressie 1991).

2) *Nonspatial elemental descriptive statistics*. The program plots histograms, frequency polygons and stem-and-leaf diagrams (Snedecor 1980; Tukey 1977), with total flexibility to select the number of intervals. In this way, the nonspatial results are obtained to determine the distribution type of z variable and to discern if a previous structure in the data exists, or if the data indicate the nonexistence of structures, etc.

3) *Thematic schemes*. This unit builds thematic maps of the area of study from the size of the spatial grid selected by the researcher; in this way it is possible to obtain a thematic map by means of squares of optional size. This map is superimposed over the spatial area and the data are fitted to this distribution. Each new rectangle shows the information of all the points

SPATIAL SOFTWARE

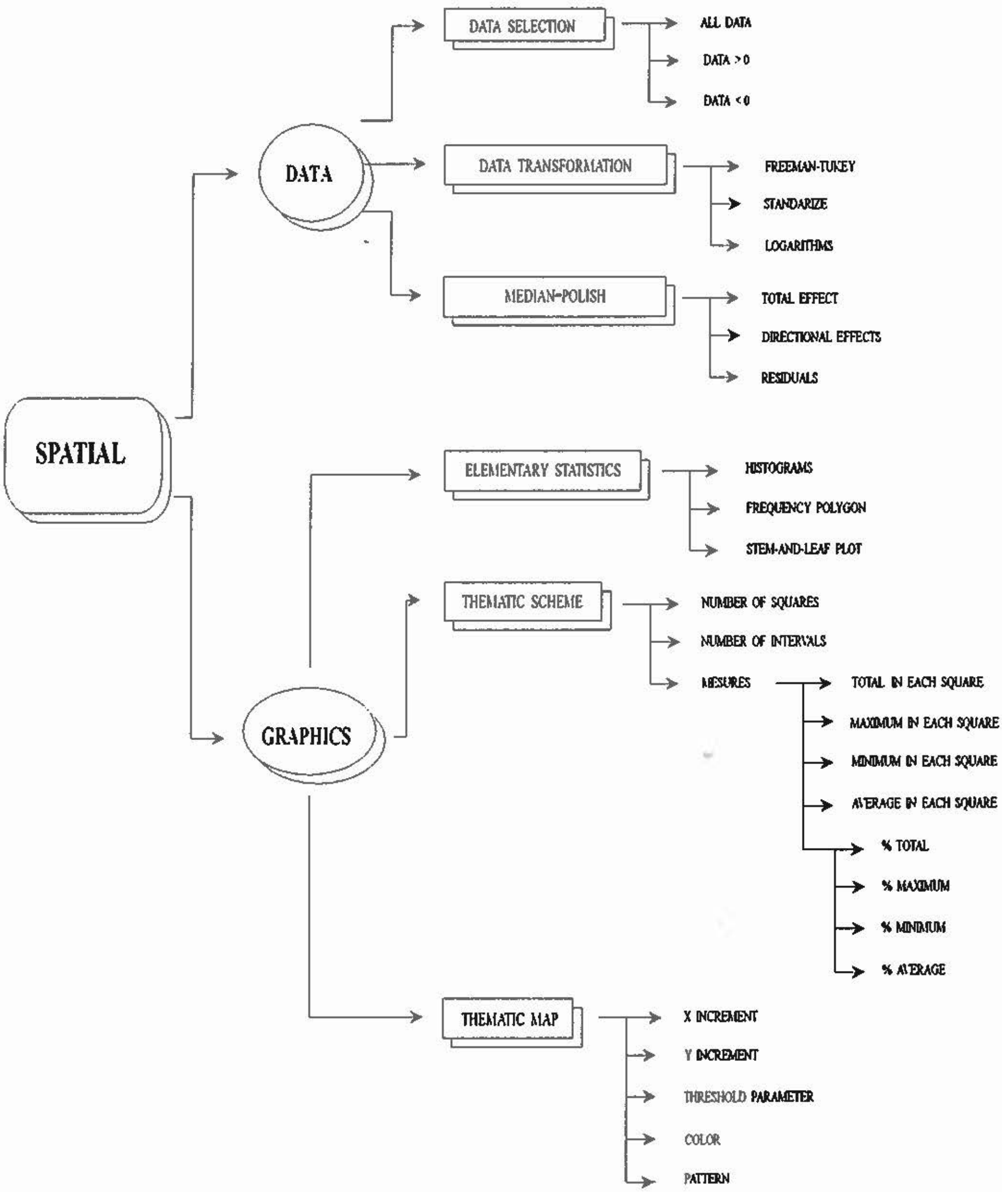

Fig. 1

contained within it. Center coordinates are assigned to each square and the values that a new variable Z adquires in this square as we say previously.

The squares are coloured or shaded depending upon Z values (Upton, Fingleton 1989), dividing the variable range in intervals of the same length. The interval number is selected by the researcher (with maximum of 8).

4) *Thematic maps*. This procedure divides the area of study in rectangles in the following way: from x and y increments given by the researcher or by default (in this case the software analyzes the data and computes the minimum values for x e y, so each rectangle includes as maximum a point of the distribution only or none) (Cressie 1991).

Later, the program makes a division of the range of Z between the intervals selected by means of the algorithm, locating the highest increments of Z and establishing the intervals to maximizing these increments, with a maximum of 8 intervals (if the algorithm obtains more than 8 intervals, an optimization process is made to limit this number). In this way, areas with a higher contrast with regard to adjacent areas are displayed, thus avoiding the masking of the less contrasting anomalies, due to the averages and approximations between adjacent areas. The contrast quality is determined by the threshold parameter p. The algorithm fixes this parameter at 2 by default, but the researcher may elect to carry out a complete study in order to obtain the low contrast anomalies, or the high contrast anomalies, etc. (if p is high, the highest contrast anomalies stand out).

The plotting of the thematic map can be made in colour or in black and white tones.

5) *Spatial trends*. The use of the median-polish algorithm, developed by Tukey (Tukey 1977; Cressie 1991), is focused on breaking down the data in different trends to determinate spatial influences such as: global trend, NS and EW directional trends and local trends. In this way it is possible to study data distribution in terms of these points of view.

6) *Display characteristics*. The program fits the display size to the spatial limits shown by the data and adjusts the scale to obtain the maximum window for display on the monitor screen. The display appears like a grid, with a scale adapted to the x and y. Each rectangle is coloured or shaded in terms of the interval where Z is assigned.

6. Conclusions

1) In this paper we establish a general methodology that is applicable to a data set defined by means of spatial 2D variables and a quantitative variable.

2) The proposed algorithm is focused on the whole data set in nonspatial way, though the results have spatial character.

3) The algorithm eliminates the little meaningful anomalies and provides a notable anomalies map that emphasizes the most important areas for subsequent intervention.
4) The division of the quantitative variable in intervals accomplishes a good contrast of the anomalies, giving the structural direction and the structures intersection.
5) The application in The Gabias settlement has permitted to locate the areas with meaningful anomalies, the direction of the structures and low isolated anomalies. The excavation of some prospected areas and the contrast with several profiles of subsurface radar has permitted to confirm the validity of the method.

José Antonio Esquivel
Beatriz Benjumea
José Antonio Peña
Department of Prehistory and Archaeology
University of Granada
Andalusian Institute of Geophysics
University of Granada

ACKNOWLEDGEMENTS

This study is included in the project "Development of Methods for the Diagnosis and Conservationist Research of the Archaeological Heritage", financed by the Interministerial Commission of Science and Technology of Spain (I+D project).

BIBLIOGRAPHY

Bhattacharyya B.K., Navolio M.E. 1976, *A fast Fourier transform method for rapid computation of gravity and magnetic anomalies due to arbitrary bodies*, «Geophysical Prospecting», 24, 633-649.

Bhattacharyya B.K. 1978, *Computer modeling in gravity and magnetics interpretation*, «Geophysics», 43, 912-929.

Breiner S. 1973, *Applications Manual for Portable Magnetometers*, Sunnyvale, Geometrics.

Chatfield C. 1988, *Problem Solving*, London, Chapman & Hall.

Cressie N. 1991, *Statistics for Spatial Data*, Chichester, John Wiley & Sons.

Freeman M.F., Tukey J.W. 1950, *Transformations related to the angular and the square root*, «Annals of Mathematical Statistics», 21, 607-611.

Gibert D., Galdeano A. 1985, *A computer program to perform transformations of gravimetric and aeromagnetic surveys*, «Computer & Geoscience», 11, 5, 553-588.

Hasek V., Vencalak R. 1989, *Methods for processing digital geophysical data in archaeology*, in S. Rahtz, J. Richards (eds.), *Computer Applications and Quantitative Methods in Archaeology*, BAR International Series, 548, Oxford, 179-192.

Kemperman J.H.B. 1984, *Least absolute value and median polish*, in Y.L. Tong (ed.), *Inequalities in Statistics and Probability*, Institute of Mathematical Statistics, Monograph Series, 5, Hayward, 84-103.

Krzanowski W.J. 1988, *Principles of Multivariate Data Analysis*, Oxford, Clarendon Press.

PATERSON N.R., REEVES C.V. 1985, *Applications of gravity and magnetic surveys: The state of the art in 1985*, «Geophysics», 50, 2558-2594.

SCOLLAR I. 1990, *Topics in Remote Sensing 2: Archaeological Prospecting and Remote Sensing*, Cambridge, Cambridge University Press.

SHAFFER G. 1993, *An archaeomagnetic study of a wattle and daub building collapse*, «Journal of Field Archaeology», 20, 1, 59-75.

SNEDECOR G.W., COCHRAN W.G. 1980, *Statistical Methods*, Iowa, Iowa State University Press.

TUKEY J.W. 1977, *Exploratory Data Analysis*, Reading, Addison-Wesley.

UPTON G.J.G., FINGLETON B. 1985, *Spatial Data Analysis by Example, vol. 1: Point Pattern and Quantitative Data*, Chichester, John Wiley & Sons.

UPTON G.J.G., FINGLETON B. 1989, *Spatial Data Analysis by Example, vol. 2: Categorical and Directional Data*, Chichester, John Wiley & Sons.

WEYMOUTH J. 1986, *Geophysical methods of archaeological site surveying*, in M.B. SCHIFFER (ed.), *Advances in Archaeological Method and Theory*, vol. 9, Orlando, Academic Press, 311-395.

ABSTRACT

The geophysical techniques applied to the archaeological exploration involve specific problems of data interpretation, due to the spatial characteristics that they possess and to the particularities of material remains in archaeological sites. This work shows an application of statistical spatial analysis to the study of data in 2D coordinates together with the value of a variable z, that quantifies a spatial soil characteristic (in the magnetometric survey these values correspond to the magnetic intensity or gradient values). The analysis includes exploratory statistical no spatial methods and other methods extracted from the statistical spatial analysis, and performs thematic maps by means of a new algorithm that finds the greater contrasts in the z values. This methodology and the algorithm have been applied to the microspatial study of magnetometric data in a Roman settlement in Las Gabias (Granada, Spain) and checked with the excavation carried out in this site.