Reprinted from

Tenth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Thematic Mapper Data and

Geographic Information Systems

June 12 - 14, 1984

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

GEOCODING ISARITHMIC MAPS WITH A MICROCOMPUTER

C. ASPIAZU, J.C. RIBEIRO,
L. MOTA ALVES, G. ASSUNCAO RIBEIRO

Federal University of Viçosa
Viçosa, Minas Gerais, Brazil

ABSTRACT

A uniformly-spaced grid system is a convenient and simple way to collect and store geographical data; these data to be used as input in an information system mostly oriented toward land planning and management. A BASIC program for coding isarithmic maps in vector format and filling with interpolated values on a uniformly-spaced grid was prepared. Interpolation methods used were: numerical approximation and surface fitting (polynomial and multiquadric functions). The program is run on a 55,536 bytes-RAM microcomputer.

That software has already been used to transform photogrammetrically obtained terrain elevation contours into space grid structures. With the coded and interpolated contour lines, three-dimensional surface views of a sparsely forested area were prepared. Similarly, numerical data gathered from several maps depicting individual climatological parameters over the entire State of Minas Gerais, Brazil, were organized into grids which were subjected to a factor analysis, by means of the same computer equipment, to propose a method for regional climatic classification.

I. INTRODUCTION

Efficient land use planning and management are basically dependent on the availability of reliable data organized as to facilitate their handling and interpretation. A uniformly-spaced grid placed on a source map is one of the most useful methods for coding and storing spatial data; these data to be used as input in a geographical information system (GIS). A grid presents data organized in matrix form, with the location of each individual cell being specified by its row and column numbers. Since geographical information may almost certainly be expressed as a set of numbers, a grid system allows the combination of data originating from such diverse sources as isarithmic and choropleth maps, as well as from duly rescaled, digitized aerial photographs and orbital imagery, with each set of registered cells of the combined matrices becoming a vector.

The size of the grid cells depends on the dimensions of the area and the scale of the details being pursued. Smaller cells result in increased processing time and relatively higher inaccuracies due to grid misalignment. Larger cells result in smoothing of details. MacDougall (1976, p. 60) gave a set of general rules to define the most suitable grid spacing for a particular map.

In geographic information systems, polygon boundaries are usually represented by an ordered sequence of x-y coordinate pairs of points representing the end points of very short, straight line segments. This is known as vector format (Marble and Peuquet, 1983, p. 925). Information can be coded automatically by means of an optical scanning device or, manually, through the use of a digitizing tablet. Interpolation of the coded data takes care of filling grid cells with pertaining values.

Any available hardware can be used to store and handle all generated data. A microcomputer, despite the usual drawbacks concerning execution time and memory requirements, allows for satisfactory machine-user interaction, enhanced by its graphical possibilities. Data handling for specific purposes requires the use of programs for retrieval and manipulation of informational files. At this step, too, a microcomputer may be satisfactorily employed.

II. METHODS

A computer program was written in BASIC language for coding isarithms and then to interpolate to fill the grid cells with data. The program runs on an HP 9845B microcomputer with 55,536 bytes RAM and the following peripherals: a digitizing tablet measuring 35x45 cm; a plotter; two mass storage units using 500 kbytes-flexible disks; a card reader; a thermal printer; and, a dot-matrix printer.

A. TWO-DIMENSIONAL INTERPOLATION

The program permits the user to choose one of three interpolating procedures: numerical approximation, polynomial and multiquadric.

Each cell central value, C, is estimated by the numerical approximation procedure based on the average of digitized points surrounding the cell, weighted by the reciprocal of their powered distances to C. In symbols:

$$C = \sum_{k=1}^{M} (Z_k/D_k^{\ r}) / \sum_{k=1}^{M} (1/D_k^{\ r}) \qquad (1)$$

where Z_k are the surrounding M data values and D_k are their distances to C; r is the power to which distances are raised (MacDougall, 1976, pp. 109-111).

The polynomial procedure calculates C through the equation:

$$C_i = a_1+a_2X_i+a_3Y_i+a_4X_iY_i+a_5X_i^{\ 2}+ a_6Y_i^{\ 2}+a_7Y_iX_i^{\ 2}+a_8X_iY_i^{\ 2}+ a_9X_i^{\ 3}+a_{10}Y_i^{\ 3} \qquad (2)$$

$$i = 1, 2, \ldots, n$$

where a_1, a_2, ..., a_{10} are coefficients and X_i, Y_i coordinates of cell i (Gama and Pereira Jr., 1980, p. 15). Thus, a polynomial equation is determined that best fits the pattern of the surficial data.

The remaining procedure is also a surface fitting one and uses the multiquadric function:

$$C_i = \sum_{j=1}^{n} C_j((X_i-X_j)^2+(Y_i-Y_j)^2)^{1/2} \qquad (3)$$

$$i = 1, 2, \ldots, n$$

where n is the number of cells surrounding data values; C_j a coefficient; X_i, Y_i the coordinates of cell i; and X_j, Y_j the coordinates of the cell j (Hardy, 1977, p. 489).

B. CODING AND INTERPOLATING CONTOURS

The program has already been used to fill a grid structure with data from photogrammetrically-obtained terrain elevation contours (scale 1:10,000) of a sparsely forested natural area. The area of 194.36 ha (479.9 acres), called Paraiso's Basin, is located near the University Campus at Viçosa, State of Minas Gerais, Brazil. The contour lines were digitized in vector format. The interpolated grid consisted of 2,048 cells (64 lines x 32 columns), each cell being a .4 x .4-cm square.

The cell-filled grid was later retrieved and processed by means of software designed to represent graphically the ground relief from a central perspective. Coordinates for each observed point on the projection plane were calculated as:

$$X_p = (Y_o.X-X_o.Y)/(Y_o-Y) \quad (4)$$

$$Y_p = (Y_o.Z-Z_o.Y)/(Y_o-Y) \quad (5)$$

where X_o, Y_o, and Y_o are the coordinates of the observer in the three-dimensional space; X, Y, and Z the coordinates of the observed point in three-dimensional space.

C. CODING AND INTERPOLATING CLIMATOLOGICAL MAPS

Grid structures were also obtained from climatological maps depicting 15 parameters for the entire State of Minas Gerais, an area of 587,000 km^2 (about 1.45×10^9 acres). A map of the State (scale 1:3,000,000) was subdivided into 1 cm-side cells. Cells that fell outside of the State were disregarded and, thus, each grid consisted of 508 cells. Values for each cell were obtained by numerical approximation. At the end of the interpolating procedure, the cells resembled a network of regularly-spaced meteorological stations. Computer subroutines from the System/360 Scientific Subroutine Package (IBM, 1970, pp. 55-56) were translated to BASIC language and run on the previously mentioned hardware to process the data and to propose a methodology for performing regional climatic classifications.

First, a matrix of simple correlations between the input variables is obtained by the program. Next, a factor analysis is carried out, retaining for further utilization only those factors with eigenvalues equal to or greater than one. At the end, a Varimax orthogonal rotation is applied. The final result of the factor analysis consists of sets of indices that summarize the original climatic variables through a smaller number of composite variables. Each set of indices originates values with which to fill grid cells, with the result that there obtain as many grids as indices calculated. These grids are identical to the ones filled with the original climatic variables. Performing a cluster analysis of those grids, all information is finally combined, and the region can be subdivided in seemingly homogeneous areas (clusters). The number of areas to be shown in the climatic classification is calculated using W. D. Fisher's algorithm (Hartigan, 1975, pp. 141-142). Further details concerning climatic classifications built using factor analysis can be found in the works of McBoyle (1973, pp. 110-118) and Powell and MacIver (1977, pp. 1-37). The cluster analysis performed therein was based on the work by McCammon (1968, pp. 1663-1670).

III. RESULTS AND DISCUSSION

By applying the coding and interpolating procedures to two different situations, greater insight was gained about them and the electronic equipment used. Once gathered, the data has to be stored in ways apt for further retrieval and management for specific purposes. It can be said that the results obtained here were most encouraging.

A. THREE-DIMENSIONAL DISPLAY

Fig. 1 shows the Paraiso's Basin area and its contours. In Fig. 2 is the three-dimensional view of the Basin, obtained by numerical approximation, taking r=2 on equation (1). As expected making r=4, the resulting projection, not included here, showed local factors having higher, undesirable, influences denoted by a more "noisy" surface.

Other projections, also not shown here, were obtained by using the polynomial interpolation and the multiquadric function. The former produced good results but the required processing time was considered excessive; a faster processing was feasible by reducing the number of coded data, but the results were degraded in quality. The latter gave undue weight to local factors and, at times, discrepancies with reality were evident.

MacDougall (1976, p. 109) considered that compared with surface fitting, numerical approximation allows representation of more complex surfaces, restricting the spatial influence of any errors, and the interpolated surface will pass through the original data values. Surface fitting, on the contrary, would tend to smooth the raw data, which in some cases, is desirable.

Six surrounding points were used to obtain the interpolated values with which Fig. 2 was built. It is fairly obvious that quite different maps can be produced by numerical interpolation by simply varying the number of surrounding coded data points; the number of these to be used will depend on the spacing of the points and the spatial variation within the map. MacDougall (1976, p. 110) says that the appropriate number seems to be in the range of 6 to 9. For interpolation purposes, the use of a number higher than the given range upper limit will produce smoothed maps; whereas, a number lower than 6 will result in abrupt changes over the surface being mapped.

B. CLIMATIC CLASSIFICATION

Three indices (factors) were obtained after performing the Varimax rotation. The first index (Factor I) explained 47.3% of the total variance. This was called the "Thermal Index" ince it was made up exclusively of air temperature data. The second index (Factor II) explained 17.8% and was designated as the "Winter Air Humidity Index". The third (Factor III), known as the "Summer Air Humidity Index" accounted for 18.4% of the total variance. These three factors explained 83.5% of the total variation. The results above were in complete agreement with those obtained by processing the data with the Statistical Package for the Social Sciences (SPSS) on a University IBM 360 computer.

Three grid systems, with 508 cells each, were filled with the values (factor scores) as calculated by the mentioned indices. Performing a cluster analysis with these data was not possible since the microcomputer does not have sufficient memory for processing a 508x3-matrix. Thus, the number of cells was reduced to 121 by arbitrarily averaging values in 4-cell subsets. After applying the Fisher's algorithm, ten clusters were mapped (Fig. 3). As would be expected, the classification obtained here, using 15 variables, differentiated the State's climate to a greater extent than does Köppen's classification, which recognizes 4 climatic types by the consideration of only local temperature and precipitation data.

The main purpose of the present work was to propose a method for regional climatic classification based on micro-computer processed meteorological data. Considering specific climatic, edaphic and topographical constrains for a given plant species, it would also be possible, through the same procedures presented here, to subdivide a region into zones according to the ecological requirements of the species.

Isarithmic maps were digitized when point data would have been used instead. Obtaining data from lists of spatial distributions (coordinates and values) rather than by taking samples from maps, avoids measurement errors and the addition of interpretation and bias introduced by the map compiler. Regrettably, the original data lists were unavailable for this work.

The density of the isarithms to be coded and interpolated were very different for each climatological variable -a few lines for air relative humidity, many of them for precipitation data. Following the guides given by

MacDougall regarding grid size would result in the use of a different grid spacing for each map, even though these were of the same area. Selection of a common grid was made based on the best grid for the most important maps, while being appropriate for the least detailed map and also in consideration of the accuracy limits of the maps.

IV. CONCLUDING REMARKS

No attempt was made, at this time, to evaluate statistically the accuracy of the interpolated data used in the situations presented. This, and further refinements, will follow with the aim to integrate a system for gathering data from any available source (isarithmic and choropleth maps), and for data management and interpretation. It is felt that such a system, based on a microcomputer, might become a nice and relatively inexpensive tool for land use planning and management for forest and agricultural purposes.

Geographical data were not yet combined by these authors with orbital or other kind of imagery. Doing this, hopefully, will not be a very difficult task once a somewhat crude, locally prepared, remote sensing software package already available, is improved.

V. REFERENCES

Gama, C. D. and Pereira Jr. G. 1980. Métodos otimizados de mapeamento automático. In: Simpósio sobre Aplicações Gráficas por Computador e Sistemas Gráficos Interativos, 2º., São Paulo. Anais, 2º. vol., p. 11-29.

Hardy, R. L. 1977. Least squares prediction. Photogrammetric Engineering and Remote Sensing. 43(4):475-492.

Hartigan, J. A. 1975. Clustering algorithms. John Wiley & Sons, Inc., New York.

IBM. 1970. IBM application program: system/360 Scientific Subroutine Package. Version III. Programmer's manual. 5th. ed. New York.

McBoyle, G. R. 1973. Climatic classification of Australia by computer. In: McBoyle, G. R., ed. Climate in Review. Houghton Mifflin Co., Boston, Mass.

McCammon, R. B. 1968. The dendrograph: a new tool for correlation. Geological Soc. of America Bull. 79:1663-1670.

MacDougall, E. B. 1976. Computer Programming for Spatial Problems. Edward Arnold Publ. Ltd., London.

Marble, D. F. and Peuquet D. J. 1983. Geographic Information Systems and Remote Sensing. In: Simonett, D. S. and Ulaby, F. T., eds. Manual of Remote Sensing, Vol. I, 2nd. edition. American Society of Photogrammetry, Falls Church, Va.

Powell, J. M. and MacIver, D. C. 1977. A summer climate classification for the forested area of the Prairie Provinces using factor analysis. Northern Forest Research Centre, Edmonton, Alberta, Canada. Information Report Nor-X-177.

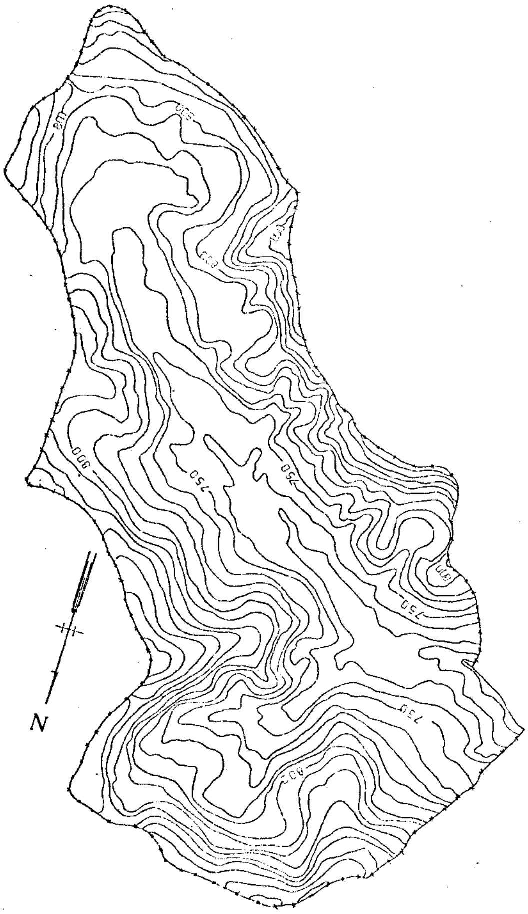

Figure 1. Contour lines of the Paraiso's Basin.

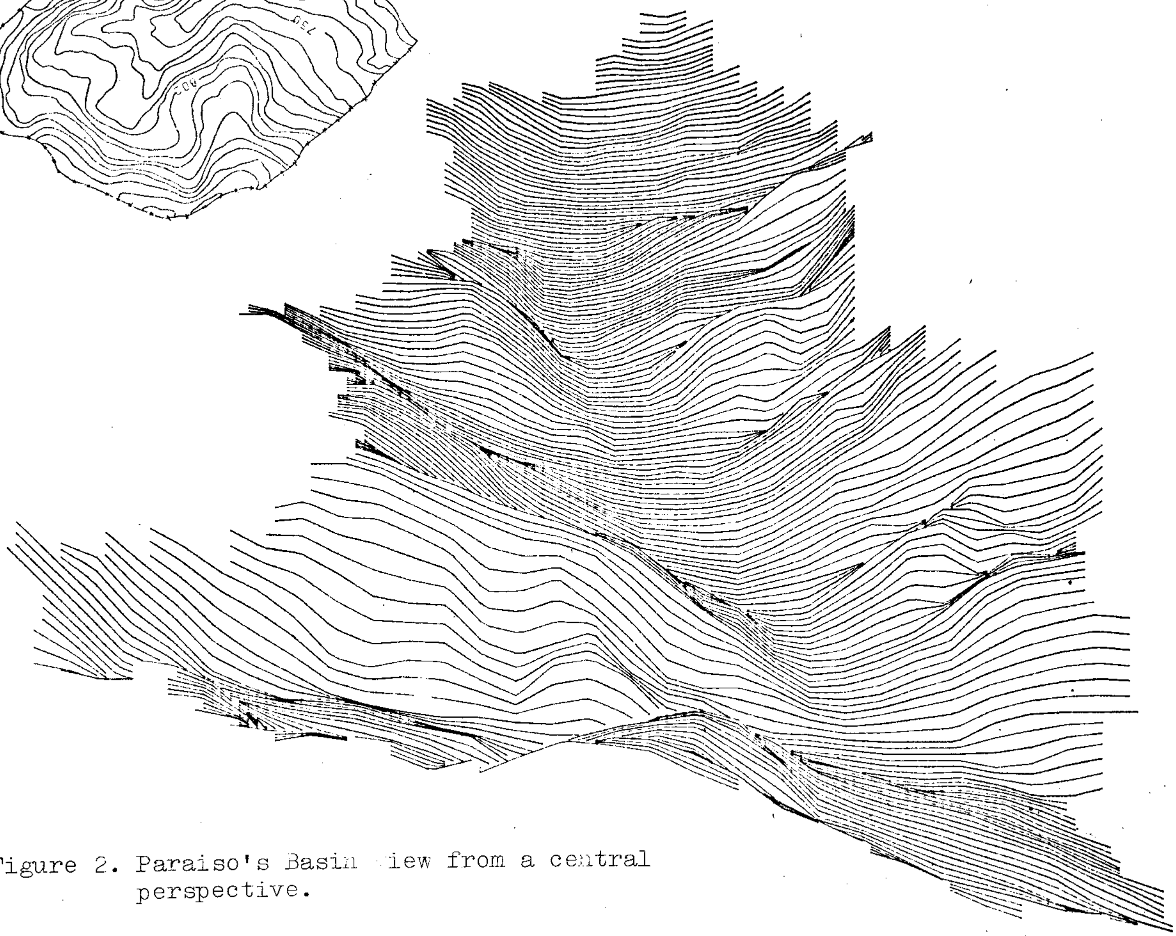

Figure 2. Paraiso's Basin view from a central perspective.

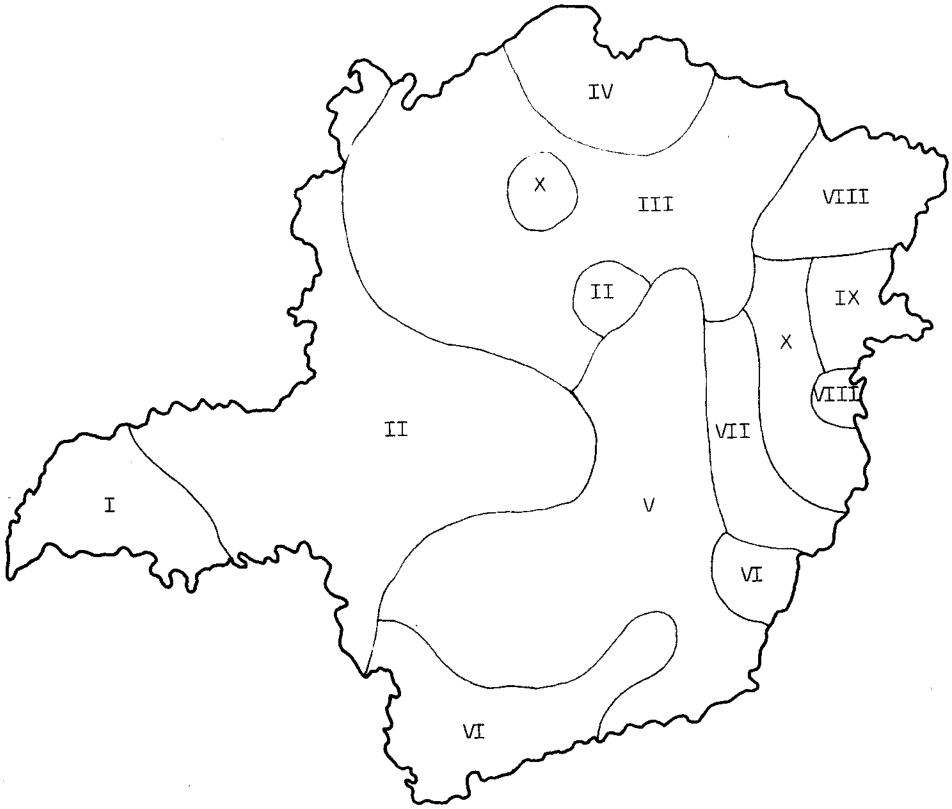

Figure 3. Climatic classification of the State of Minas Gerais, Brazil.

AUTHORS BIOGRAPHICAL DATA

CELESTINO ASPIAZÚ is Associate Professor of Forestry and Agronomy at the Federal University of Viçosa, Viçosa, Minas Gerais, Brazil. He is a former Assistant Professor of Agronomy at the University of Buenos Aires, Buenos Aires, Argentina. He is an Agricultural Engineer from the University of Buenos Aires. He received the M.S. and Ph.D. degrees in Agricultural Climatology from Iowa State University, Ames, Iowa. Dr. Aspiazú is a member of Gamma Sigma Delta, The Honor Society of Agriculture, and also of the American Society of Photogrammetry, the American Society of Agronomy and the American Meteorological Society. His areas of specialization include agricultural climatology and remote sensing (yield estimates, land use, and digital image processing).

JOSÉ CARLOS RIBEIRO is Assistant Professor of Forestry at the Federal University of Viçosa, Viçosa, Minas Gerais, Brazil. He is a Forest Engineer from the Federal University of Viçosa. His areas of specialization include photogrammetry, photointerpretation, and automatic data processing.

LACI MOTA ALVES is a former Research Assistant of Forestry at the Federal University of Viçosa, Viçosa, Minas Gerais, Brazil. Presently, he is teaching topography and photogrammetry at the Superior School of Agronomy, Paraguaçu Paulista, São Paulo, Brazil; he also does private consulting. He is a Forest Engineer from the Federal University of Viçosa, and holds the M.S. from the same University. His areas of specialization include topography, photogrammetry, and photointerpretation.

GUIDO ASSUNÇÃO RIBEIRO is Research Assistant of Forestry at the Federal University of Viçosa, Viçosa, Minas Gerais, Brazil. He is a Forest Engineer from the Federal University of Viçosa, and holds the M.S. from the same University. His areas of specialization include remote sensing, forest climatology and forest fires.