IMPROVING DIGITAL ELEVATION MODELS THROUGH BETTER SAMPLING METHODS

Walker Gomes MSc
Head of Atmospheric Science Division
Centro Técnico Aeroespacial
12225 - São José dos Campos - São Paulo
Brazil

Commission IV

ABSTRACT:

This paper presents a new method for acquiring data points from topographic charts to be used on Digital Elevation Models (DEM's) generation. The DEM's accuracy are strongly influenced by the original data points position and grid regularization procedures (to obtain a rectangular grid). The first step was the understanding of the errors sources and their propagation. In order to put into practical use the new method, it was tested on a Geographical Information System which had a traditional data point acquisition method; based on nearest neighbors interpolation. The new method is based on fact that some algorithms first check the data input. If the original data are on unfavorable positions it automatically, through special procedures, chooses other data inputs. A linear interpolation is used on favorable circumstances, and Akima cubics interpolator on more difficult situations. It was used an IBM-PC-XT compatible computer, and the algorithms are fast enough for practical use. The examples have shown an excellent improvement over the traditional nearest neighbors interpolations.

KEY WORDS: MDE, GIS, 3-D, computer graphics.

1. INTRODUCTION

It is well known that there isn't a single perfect interpolation method, and that the required precision is a function of the model's application. While a given model can prove excellent for some usages, it can be inadequate or even useless for others.

In this sense, it's believed that there can be a large number of users interested in a DEM generation technique with the following characteristics:

- Low cost.
- Contour lines extracted from topographic charts as data source.
- IBM-PC-XT hardware or better.
- Very high resolution for a large number of applications.

The proposed method is designed to be a specialized system intended for use under the constraints mentioned above. Within this frame, a method was searched that could be both accurate and simple enough to be compatible with the data source and hardware used.

The quality of DEMs is closely dependent on sampling and interpolation processes, and it's possible to say that sampling is paramount because there is no way to obtain good results from any interpolation method if the samples are of poor quality.

So, after a series of experiments with linear, cubic and bicubic interpolators (Gomes,1990a), it was chosen to make use of (1) linear interpolation for the most part of model generation, (2) cubic interpolation in the cases where linear interpolation presents difficulties, and (3) bicubic interpolation for grid densification.

2. EXISTING METHODS

Literary review indicates three different approaches for modelling:

1) It's assumed that the relief data are randomly related, and a local sampling process with interpolation as a function of distance is used (e.g. Barnes,1964). Although this approach is largely used in other types of modelling, it is not adequate for modelling from contour lines.

2) Maybe the most popular of all is a linear interpolation along certain axes. The number of axes may be one, two or four, and interpolation is based on a weight average of the elevations of the supporting contour lines. When it's used one axis the sampling would be very poor on some regions, compromising the interpolator performance (Gomes,1990a). On two axes, sampling is generally good, but in some cases it causes serious errors as shown in Gomes,1990a. These problems are significantly reduced on four axes, but some errors still can be found in some few places, apart from the considerable increase in computing time.

3) The most recent works make use of four axes sampling, and use the steepest profile together with cubic interpolation. (Legates & Willmott,1986). The performance of these methods has shown to be better than the former ones, but they significantly increase the computing effort.

3. ERROR TYPES

In general terms, the errors associated to grid regularization can be divided in two large classes, which it will be called Type 1 and Type 2 errors.

3.1 Type 1 error

This error occurs when the algorithm for sample selection chooses non-representative data, even though they exist. The Figure 1 is an example of this kind of error when the nearest neighbors interpolator is used for each quadrant.

3.2 Type 2 error

In opposition to the type 1 error, this error isn't caused by the data choice algorithm, been due to the not existence of good samples in the regions where this error can be found. Figure 2 shows an example where equal samples chosen by the nearest neighbors method will be encountered both in area A (peak region) and area B (valley region), causing both peaks and valleys to become flatten.

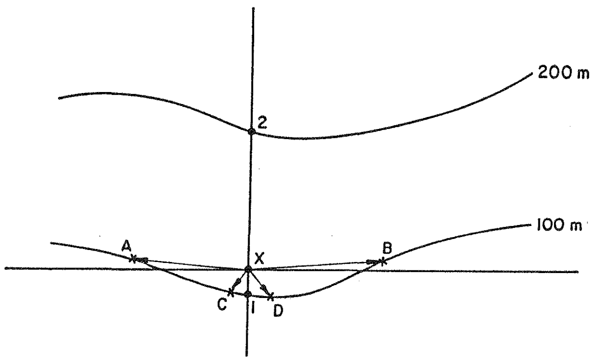

Figure 1 - Type 1 error.

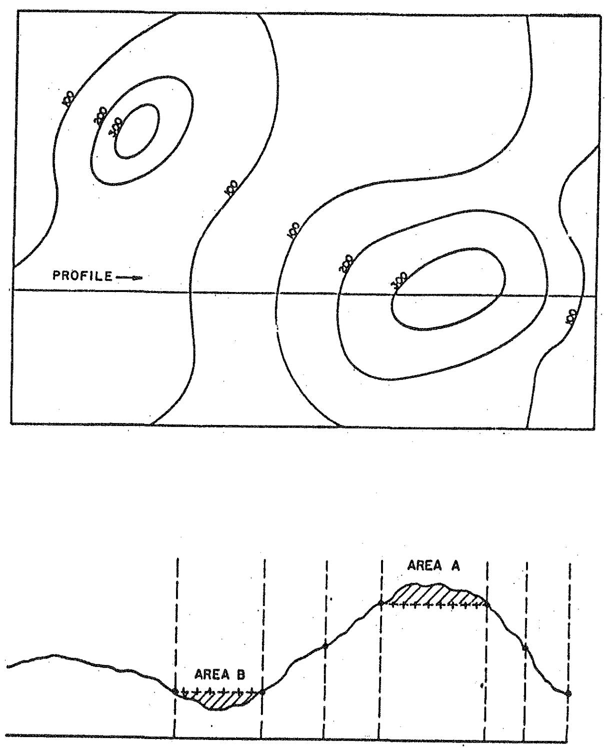

Figure 2 - Type 2 error

4. PROPOSED METHOD

The proposed method was divided in two modules for instructional reasons. The first one is the data acquisition module. The second one selects samples and interpolation methods as a result of the analysis of the previously sample points.

4.1 Data acquisition (Module 1)

Detailed information on this module can be found in Gomes & Silva,1991. The steps comprised by this module are summarized as follows:

1) At First one defines which is going to be the grid spacing in the "x" and "y" directions. The optimum grid is obtained by measuring the smaller distance between two isolines. both in the north/south and in the east/west directions. A detailed description can be found in Gomes, 1990a.

2) An overall sampling is performed on two axes, so that for each point in the grid, there is always a sample point to the left, to the right, above and below.

4.2 Selective sampling and interpolation (module 2)

The main feature of this method for data acquisition lies on the fact that the behavior of the samples in the model is transparent. It's therefore predictable, allowing manipulation during the selection of samples in order to select those that are specially representatives.

Another very important quality that comes out from predictability and ease of manipulation is the possibility of using different interpolators according to the region upon which the interpolation is performed. These characteristics are explored below.

4.2.1. Error Hypotheses and possible solutions

The basic idea is (1) to use linear interpolation in the regions where it gives good results, since this method requires minimum computing time; (2) in the type 1 error regions (the samples exist but aren't found) a search is made in order to find better samples because, as formerly stated, the problem within these areas isn't related to the interpolator, but with the sampling itself; and (3) in the regions where type 2 error occurs, the Akima's cubic interpolation (Akima,1970) is used because the use of linear interpolation in these areas would cause the flattening of peaks and valleys.

Three situations may arise when using the proposed method:

a) Case 1

The sampled points have the following characteristics: 2 points with elevation Z1 and 2 points with elevation Z2 (Fig. 3). It can be noticed that points 1 and 2 have the same elevation (200 m), while points 3 and 4 have another height (both with 300 m).

In this case, the distances between points 1 and 3 (dx) and between points 2 and 4 (dy) are measured. The interpolation uses only two points: points 1 and 3 are used if dx < dy, while points 2 and 4 are used if dx > dy.

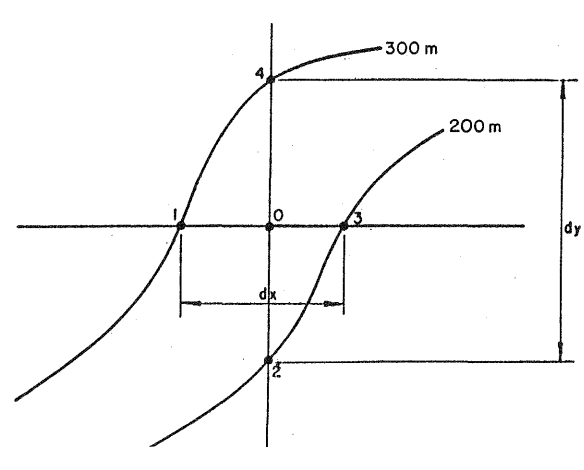

Figure 3 - Case 1 sampling

b) Case 2

The sampled points have the following characteristics: 3 points with elevation Z1 and one point with elevation Z2, or vice versa (Fig. 4).

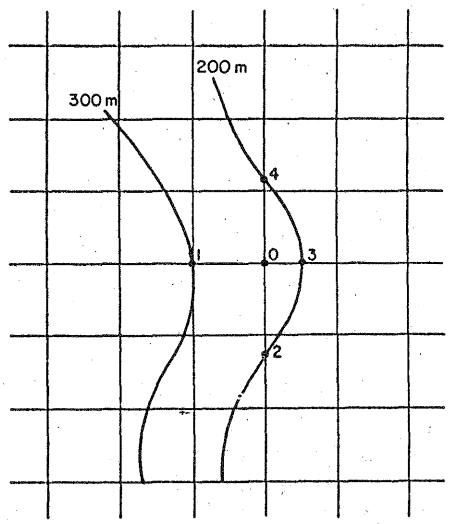

Figure 4 - Case 2 sampling

The example given in Fig. 4 shows points 2, 3 and 4 with an elevation of 200 m and point 1 with a different elevation (300 m). In order to avoid that the 200 m isoline have a greater weight than the 300m isoline, the method finds the point of different elevation, selecting this point and the one belonging to the other isoline facing the first point. In the given illustration, these points are point 1 and point 3, respectively.

c) Case 3

In this case all four points have the same value.

Two situations may be encountered: partially-closed areas, where errors of type 1 occur (grid points 1, 2 and 3 in Fig. 5), or closed area, where type 2 error is found (points A, B and C). The problem of partially-closed areas is discussed in 4.2.2, and that of the completely closed areas is discussed in 4.2.3.

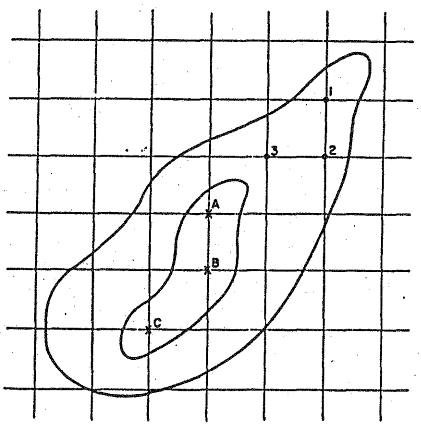

Figure 5 - Case 3 sampling

4.2.2 Possible solutions for type 1 errors

Sampling method proved to be inefficient in partially-closed areas, because the samples aren't representative (type 1 error). In the performed experiments these regions presented the largest errors and, therefore, this problem required special handling to be solved or attenuated.

Two ways were thought to deal with this situation:

a) Generic approach

This line-of-action assumes that the sampling method isn't efficient and that it's necessary to enhance the process of sampling. Under this point-of-view one alternative could use the additional sampling of points placed on the lines diagonal to the original axes. This process will most certainly improve the results in these regions, but it showed to be very expensive, because it requires a significantly larger computational effort. It should also be noticed that, although it considerably reduces the probability of occurrence of type 1 error, this error still can happen anyway, as shown in Fig. 6. Due to these considerations this approach was abandoned.

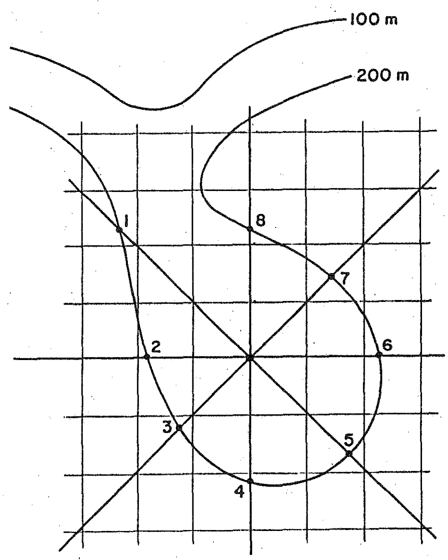

Figure 6 - Four axes sampling.

b) Local approach

The idea of this approach is to make use of the characteristics of the regions that are subjected to errors (samples with the same elevation) in order to separate them from the well-behaved regions. In this way, the additional effort is applied only to the troublesome areas. It was noticed from the experiments that the probability of having equal samples (case 3) is less than 10%. Therefore, more than 90% of the grid points to be interpolated can be solved in a simple and fast way, as described in items 4.2.1a and 4.2.1b.

This approach assumes that the sampling is poor only in those areas and that just a local search for better samples can solve the problem. In the case that a more representative sample is found, this sample is used. If not so, it's assumed that this point is located within a closed area and, therefore, it's subjected to the type 2 error.

Two ways are considered for searching the more representative sampling points

1) Manual search

When a set of four points of same elevation is encountered, the algorithm interrupts the process, indicating on the screen the point to be interpolated and asking the operator to indicate if it is a closed or partially-closed area (type 2 or type 1 error, respectively). If the operator informs that it's a type 2 error, the algorithm proceeds to solve the problem as it will be discussed in 4.2.3. If the problem is said to be of type 1, the operator will be asked to select two samples through a mouse. Fig. 7 shows that points A and B are more representative than points 1, 2, 3 and 4. After that, the algorithm uses the points selected by the operator for the interpolation, proceeding until it finds another ambiguity, when it will ask for the operator to take decision again.

This technique presents the following advantages:

- Less computational effort, and in particular, lesser effort during the sampling process, in comparison with the generic approach.

- Manual sampling for specific points is very representative, because the operator helps the computer in the selection process.

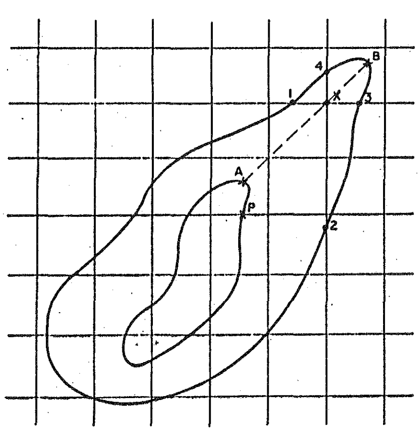

Figure 7 - Manual sample selection

This method is inconvenient because it isn't completely automatic, requiring the presence of a trained operator to interfere in the process whenever this case is encountered.

2) Automatic search

In this configuration, when samples of equal elevation are found, the algorithm looks at the neighboring points in search of a sample with different height. If such point is found, it's selected and sent to be interpolated together with the opposite point. A number of ways to perform this search was figured out, and the method illustrated on Fig. 8 was developed for this work:

- In order to find a sample of different elevation, two grid points belonging to the same Y coordinate are placed immediately to the left and to the right of the point being evaluated (points A and B, on Fig. 8a, respectively to the left and to the right of point X under evaluation).

- These two points are used as two different and constant X coordinates from which a search is done respectively above and below the Y coordinate where the two points are placed (Fig. 8a: points A2 and B2 above the points A and B respectively, and points A1 and B1 below these two points).

- If a sample of different elevation is found (point A1 in the example) this routine is finished, and this point is selected and used for interpolation together with the diametrically opposed point (point 2 in Fig. 8a), as in case 2 discussed in 4.2.1b.

- If no point of different height is found, the same procedure is used starting by selecting two points above and below the point under evaluation (i.e. belonging to the same X coordinate), and then find the points to the left and to the right of these two points. In Fig. 8b, it can be seen that starting the search by first selecting two points to the left and to the right of point X (points A and B), the four points derived (points A1, A2, B1 and B2) have the same height. On the other hand, if the search starts by selecting points below and above point X (in this case, points A' and B'), one of the four derived points (A'1, A'2, B'1 and B'2) have different elevation (point B'1 in the example). Therefore, this routine would select points 3 and B'1 to be used in the interpolation.

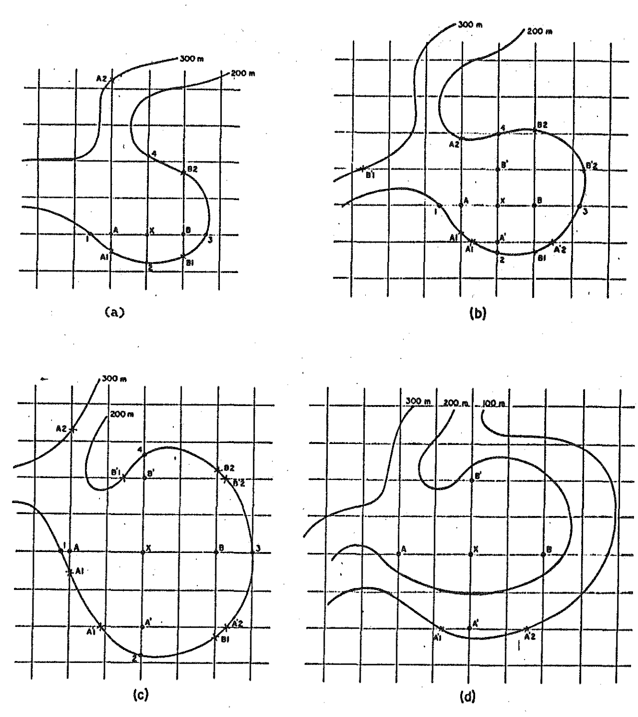

Figure 8 - Search algorithm:
(a) Displacement keeping Y constant.
(b) Displacement keeping X constant.
(c) Second iteraction.
(d) Protection mechanism.

- If no point of different elevation is found in this first iteraction, the former routine is repeated but displacing the points A, B, A' and B' two grid units distant from point X, as shown in Fig. 8c. In this illustration, the algorithm will find point A2 as the point of different elevation.

- If no point is found with this second iteraction, a new one is started with points A, B, A' and B' placing three grid points away from point X. The process repeats the same steps, enlarging the search area as necessary until it finds one point with different height, or until the protection mechanism described in the next step is used in all search directions.

- Care must be taken to avoid crossing the isoline already found when searching for points A, B, A' and B'. In Fig. 8d, the path from X to A' crosses an isoline and, therefore, the derived points A'1 and A'2 are discarded.

The samples obtained with this method are not always the best, but they considerably improve the final model. This is illustrated by Fig. 9. It shows that, although points Z and W are the ideal interpolation samples, very similar results can be obtained by using the points 3 and B'1 selected by the proposed method.

In a few partially-closed areas, this algorithm won't be able to find the different-height point (e.g. if the 200 m isoline "neck" in Fig. 9 were too narrow or too long). In such case, the area is considered as being completely closed and is treated as a type 2 error region, and another algorithm is used to solve it, as discussed in 4.2.3.

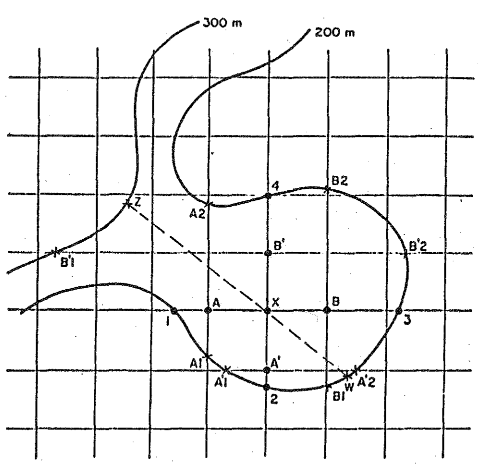

Fig. 9 - Samples obtained by the search algorithm.

4.2.3 Type 2 error solution

As previously defined, type 2 errors occur in completely-closed areas and, in this case, no better samples will be available. Whatever is the selected sample, the result will be the same, i.e. four identical elevations. In these regions, it is necessary to have an interpolator that uses slope information, as the Akima's interpolation method (Akima, 1970).

Most fortunately, the areas that cannot be solved by linear interpolation offer very good samples for cubic interpolation methods. It can be noticed in Fig. 10 that whatever is the direction of interpolation (keeping X or Y constant), points 2 and 3 will necessarily be different from each other. The same happens with points 4 and 5. In this situation the cubic interpolation deliver good results. Akima's interpolation method was chosen because it's less unstable as well as very efficient for computing (Gomes, 1990).

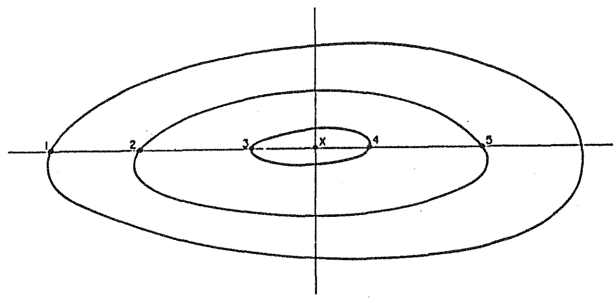

Figure 10 - Sampling for cubic interpolation in closed areas

4.4 Grid densification

Grid densification was performed with the Akima's bicubic interpolation (Akima, 1974a/b), which presented good results.

5. EVALUATION OF THE PROPOSED METHOD

The mathematical performance of this method was compared with the DEM generation method used in the Geographic Information System (GIS) developed by the Instituto Nacional de Pesquisas Espaciais (INPE, 1987) of Brazil. This GIS samples the nearest neighbor method in each quadrant.

5.1 Evaluation method

The two methods were compared as follows:

1) A 1:50000 topographic chart was used as a data source.

2) A regular grid prepared by a Brazilian cartographic authority (DSG - Diretoria de Serviço Geográfico do Exército Brasileiro) was used as ground truth. This grid was generated by photogrammetric restitution using the stereopairs employed in the preparation of the chart mentioned above.

3) A grid regularization was performed by the method used in INPE's GIS, and another using the proposed method.

4) Data generated with both methods were statistically compared with the ground truth data.

5.2 RESULTS

The comparison between ground truth data and INPE's GIS data gave the following results:

- Mean square error:	595.01m^2
- Mean error:	19.65m
- Standard deviation:	15.22m
- Variance:	231.53m^2
- Maximum error:	39.84m

The comparison between ground truth data and the grid generated by the proposed method showed what follows:

- Mean square error:	218.85m^2
- Mean error:	11.87m
- Standard deviation:	8.83m
- Variance:	77.92m^2
- Maximum error:	29.29m

It's clear that the proposed method showed a statistically better performance in all parameters analyzed, as compared to the method used in the INPE's GIS.

6. CONCLUSION

Most of the works dealing with this subject present a generic approach when defining a method for sampling and interpolation that can give good results in general. This causes a dichotomy. When the more simple methods are used (requiring less computing time), the mean error tends to be large and the visual quality is poor in many cases. When more sophisticated methods are used two serious problems are faced:

1) The more powerful interpolators require a larger number of samples. This can be good on one hand, as the weight of a poor sample is reduced. On the other hand, however, it increases the probability of having poor samples.

2) The computation effort is also increased due to the more complex sampling method, the increased number of calculations required for the interpolation, and to avoid the possible occurrence of instability in the model.

The main contribution of this work' seems to be its ability to foresee the possible errors that can be found during a grid generation. This possibility, which at first glance appeared to be impracticable, allowed for the development of a method which eliminates or attenuates such errors by using local mechanisms in the evaluation and selection of samples, as well as in the selection of the interpolation method that best fits the requirements of each particular region of the study area. Therefore, the proposed method presents very good potential for the generation of high precision grids at low cost.

6. ACKNOWLEDGMENTS

The author would like to thank all those who contributed to the accomplishment of this by any means, and in particular to Dr. Luis Alberto Vieira Dias, who gave the necessary orientation at the beginning of this research; to computer Analyst João Camilo da Silva, for the implementation and adaptation of the computer programs which made possible many of the experiments performed to test the formulated hypotheses and to generate new concepts; and to M Sc student and researcher Jurandyr S. Fonseca for the translation of this text and suggestions for its improvement.

7. BIBLIOGRAPHY

+ References from JOURNALS:

AKIMA,H.,1970 A new method of interpolation and smooth curve fitting based on local procedures. Journal of the ACM, 17(4):589-602, Oct.

AKIMA,H.,1974a Method of bivariate interpolation and smooth surface fitting based on local procedures. Communications of the ACM, 17(1):18-20, Jan.

AKIMA,H.,1974b Algorithm 474 bivariate interpolation and smooth surface fitting based on local procedures. Communications of the ACM, 17(1):26-31, Jan.

BARNES, S.L.,1964 A Technique for maximizing details in numerical weather map analysis. Journal of applied meteorology, vol.3:396-409

LEGATES, D.R. ; WILLMOTT, C.J.,1986 Interpolation of point values from isoline maps. The American Cartographer, 13(4):308-323.

+ References from GREY LITERATURE:

GOMES, W.,1990a Amostragem e interpolações em modelos digitais de terreno. (Dissertation for M Sc graduation on Remote Sensing). Instituto de Pesquisas Espaciais, São José dos Campos, Brazil (INPE-5041-TDL/404).

GOMES, W.,1990b Proposta de um novo método de amostragem para modelos digitais de elevação. VI Simpósio de Sensoriamento Remoto. Manaus, Brazil, 24-29 jun. Anais. Instituto nacional de Pesquisas Espeaciais (INPE), São José dos Campos, Brazil, v.3. p.724-731.

GOMES, W.,1991 Metodologia para obtenção, organização e armazenamento de dados de terreno. (Presented in II Simpósio de Cartografia no EMFA, Nov. 1991, in the press).

INSTITUTO DE PESQUISAS ESPACIAIS (INPE).,1987 Sistema de Informações Geográficas. Versão 1.0. Manual do Usuário. São José dos Campos, Brazil.