IMPROVEMENT OF GBF/DIME FILE COORDINATES IN A GEOBASED INFORMATION SYSTEM BY VARIOUS TRANSFORMATION METHODS AND "RUBBERSHEETING" BASED ON TRIANGULATION

Dr. George L. Fagan and Henry F. Soehngen
Bowne Management Systems Inc.
235 East Jericho Turnpike
P.O. Box 109
Mineola, New York 11501
(516) 248-6840

ABSTRACT

The Geobased Information System for the Town of Oyster Bay, Long Island, NY stores a street network represented by a GBF/DIME file and tax parcel centroids. When layers of coordinate information from the various sources are displayed on a screen or drawn on maps, obvious misfits of parcel centroids within logical blocks delineated by GBF/DIME segments have been encountered, leading to the need to "clean-up" the views for aesthetic and planning purposes.

The large amount of coordinate data for the 114 square mile area with over 96,000 parcels, made automation of "correcting" the coordinates to a common base essential. It was decided to use the Parcel Centroid layer as the base and transform the GBF/DIME coordinates to this base. Control points were established on the Parcel Centroid layer as apparent street intersections and digitized. The corresponding GBF/DIME coordinates were then transformed by least squares using a number of transformation procedures including linear affine, bilinear affine, orthogonal, and 2nd order polynomial.

The results while satisfactory in some cases, still left residuals of the control points which could only be resolved by arbitrary rules. Therefore, for the final production, a transformation method based on triangulation of the control point network by Delaunay triangles followed by use of linear affine transformation parameters for each triangle was employed. The results were highly satisfactory.

INTRODUCTION

The Town of Oyster Bay Geobased Information System (GIS) serves to provide highly useful graphical and non-graphical data for planning, including pavement management and the routing of solid waste collection vehicles. The GIS stores a street network represented by a GBF/DIME file, parcel centroids and associated files containing other features and information.

Created during a project to develop computerized collection routes for the Town's solid waste collection vehicles (Fagan 1986), the Geobased Information System can accomodate municipal information such as maps, assessment records, census information, and land use information as though it were a single continuous map. The database is extremely flexible in its design and permits Town departments to access geographic and related attribute information from a single source. This centralization provides efficiency from reduced record maintenance and the availability of consistent information.

Database Development

The Geobased Information System presently consists of the following base map and overlay features:

- Street Network
- Non-Street Features
- Municipal Boundaries
- Parcel Centroids
- Pavement Management Information
- Solid Waste Collection Information
- Annotation

Various data sources were employed to gather the above information and create the GIS. Several of these sources were available on computer media and included the GBF/DIME file for the Nassau-Suffolk County Standard Metropolitan Statistical Area (SMSA) from the U.S. Bureau of the Census, the Nassau County Parcel Coordinate/Area File, and the Town of Oyster Bay Assessment File.

The street segments from the GBF/DIME file were used to define a graphic network of all streets within the Town and a non-graphic database of attribute information for each segment. This network was checked for missing or incorrect streets, improper street geometry and topology and incorrect street names or other attribute information. The database was updated to reflect any necessary changes.

The Town's Assessment File contained useful information but lacked coordinate information which prevented it from being integrated into the GIS. Coordinate information was available for each parcel in Nassau County's Parcel Coordinate/-Area File. A computer program was written to merge these two files and bulk load the information as the Parcel Centroid layer within the GIS.

Misfit of the GIS Overlays

When the layers of coordinate information from the various sources used to create the GIS were displayed on a screen or drawn on maps, obvious misfits of parcel centroids within the logical blocks delineated by GBF/DIME segments were encountered (see Figure 1). These positional discrepancies, which often result when map data from different sources are combined or overlaid onto a single map, are usually the result of differing source, projection and accuracy requirements of the original documents.

These misfits lead to the need to "clean-up" the views for aesthetic and planning purposes. The large amount of coordinate data for an area of 114 square miles and over 96,000 parcels made automation of "correcting" the coordinates to a common base very essential.

Establishment of Control Points

In order to improve the positional relation between map layers in the Geobased Information System, it was decided to use the Parcel Centroid layer as the base and transform the GBF/DIME coordinates to this base. To accomplish such a transformation, definition of control points was necessary. These control points are used to force registration of selected GBF/DIME intersections over their apparent counterparts in the Parcel Centroid layer and in turn bring other intersections into near coincidence with their counterparts.

The transformation routines, which are performed interactively at a graphics workstation, were implemented so as to allow an iterative, piecewise solution (White and Griffen 1985) over successive user defined polygons until a suitable solution is obtained. These routines include a utility that assists in the insertion of control points.

The interactive routine to place control points allows the user to select an apparent intersection location in the Parcel Centroid layer, digitize a control point at this location, and assign the corresponding GBF/DIME intersection (see Figure 1). All required attribute information regarding the control points is automatically collected from the GIS and the control points are stored in a data layer created for this purpose.

The rules used for control point placement were subjective and based primarily on the orientation of misfit between the layers for the defined polygon. General guidelines included:

- Include control points near the corners of the polygon.
- Place control points strategically around the periphery of the polygon.
- Place several control points near the center of the polygon.
- Do not place control points too close together.
- Control should be placed in locations of extreme offset.
- For more precise coincidence, add more control points. A minimum of 8 control points was recommended per polygon.

EXPERIMENTATION WITH GLOBAL MAP TRANSFORMATIONS

The literature in photogrammetry and remote sensing abounds with the use of various mathematical transformation functions to eliminate systematic and random errors in observational data to improve results (Kratky 1972, Bahr 1976). Experience with these methods seemed to indicate that one or more of the standard transformation concepts would produce satisfactory improvement in the final composite map product and database.

Transformation Models Utilized

2 Dimensional Affine Transformation. [6 Parameters]

$$X = a_1 + a_2x + a_3y \quad (1)$$

$$Y = b_1 + b_2x + b_3y \quad (2)$$

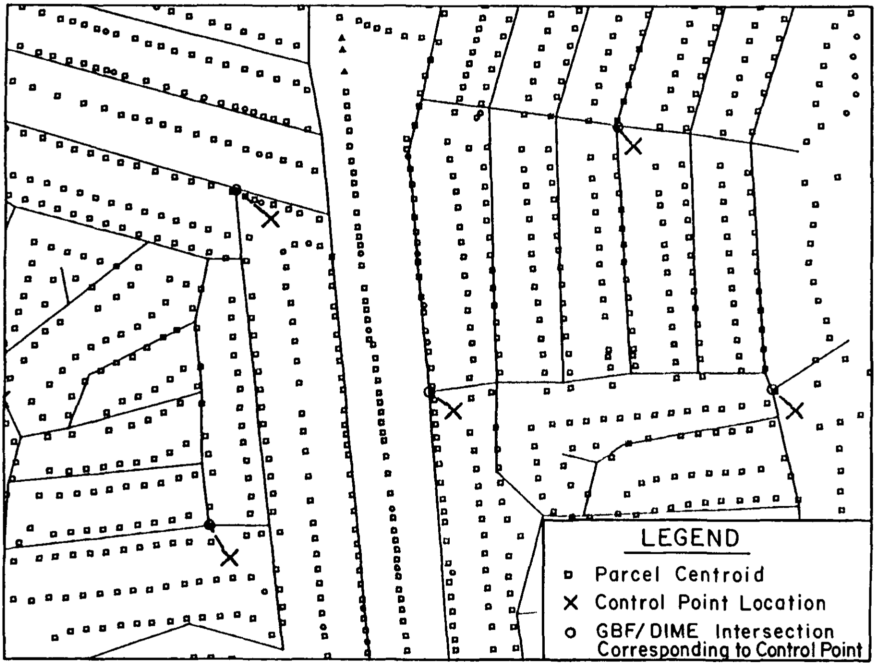

FIGURE 1
Misfit of Parcel Centroids within Logical Blocks
Delineated by GBF/DIME Segments

Bilinear Polynomial. [8 Parameters]

$$X = a_1x + a_2y + a_3xy + a_4 \quad (3)$$
$$Y = b_1x + b_2y + b_3xy + b_4 \quad (4)$$

Linear Conformal (Similarity Transformation). [4 Parameters]

$$X = a_1x - a_2y + a_3 \quad (5)$$
$$Y = a_1y + a_2x + a_4 \quad (6)$$

2nd Order Polynomial. [12 Parameters]

$$X = a_1 + a_2x + a_3y + a_4x^2 + a_5y^2 + a_6xy \quad (7)$$
$$Y = b_1 + b_2x + b_3y + b_4x^2 + b_5y^2 + b_6xy \quad (8)$$

Projective Transformation. [8 Parameters]

$$X = (a_1x + b_1y + c_1)/(d_1x + e_1y + 1) \quad (9)$$
$$Y = (f_1x + g_1y + h_1)/(d_1x + e_1y + 1) \quad (10)$$

In the above equations:
- X,Y are the control point coordinates in the parcel centroid layer.
- x,y are the corresponding original GBF/DIME coordinates for the control points.
- $a_1, a_2, \ldots, a_6$; $b_1, b_2, \ldots, b_6$; and $a_1, b_1, c_1, \ldots, h_1$ are required transformation parameters.

Because the number of available control points per map area always exceeded the minimum number required to obtain the transformation parameters for a particular transformation model (by design), a Least Squares solution for the determination of the particular parameters was always utilized.

Least Squares Principles

A solution based on a unit matrix to represent the weights assigned to the control points was considered to be most logical. Thus, the system of observation equations for the Least Squares method can be expressed in matrix form as:

$$AX - L = V \qquad (11)$$

where:

X is the vector of required transformation parameters.
A is the coefficient matrix of observation equations.
L is the vector of observed quantities.
V is the vector of residuals.

Then,

$$A^TAX = A^TL \qquad (12)$$

is a system of normal equations (number of equations equals number of unknowns) which can be solved for the unknown parameters of the particular transformation model chosen.

The solution for the elements of the X vector can be obtained by a number of elimination methods or by matrix inversion operations. In the computer program system to be described, the solution vector elements were obtained by use of the Gauss-Jordan algorithm in subroutine form using an augmented matrix concept. Subroutines to perform matrix transposition, matrix multiplication and matrix times vector multiplication were also utilized.

Preliminary Preparations and Concepts

Digitizing of the control points in each map area utilizing the graphics workstation resulted in a file of coordinate data for each control point. One set of x,y coordinate data, expressed in State Plane Rectangular Coordinate values related to the digitization of apparent street intersections in the Parcel Centroid layer and the other represented the intersection node coordinates as provided in the GBF/DIME file for the relevent SMSA. Table 1 illustrates a short list of the structure of this file.

TABLE 1
CONTROL POINT COORDINATE FILE

Point No.	Census Basic	Tract Suffix	Census Node	Base Map Coordinates X	Base Map Coordinates Y	GBF/DIME Coordinates X	GBF/DIME Coordinates Y
1	5218	2	57	2155573.	163166.	2155596.	163088.
2	5218	2	48	2156517.	163041.	2156651.	162949.
3	5218	2	73	2156129.	162660.	2156099.	162508.
4	5218	2	81	2155836.	161957.	2155798.	161483.
5	5218	2	1103	2155289.	161537.	2155273.	161379.

The computer system developed to produce transformed composite maps was conceived to operate in an on-line, interactive mode and to permit as many iterations as desired based on selection of the number and distribution of the control points in the map area and on the choice of the transformation method. The production of parameters and consequent residuals on the control points is very fast on a Digital Equipment Corporation VAX 11/750.

Tests and Conclusions

As an illustration of the potential of utilizing global map transformation concepts, various tests were performed on a particular map area where the overlay of original GBF/DIME data on the base map was poor. Quick numerically oriented comparisons based on number of control points and transformation type produced results illustrated in Table 2. The results of this type of analysis (even before plotting) seem to indicate that a 2nd order polynomial would fit the particular map area best. A plot (Figure 2) of the area using the transformed GBF/DIME node data definitely shows great improvement over the original view.

Table 2 illustrates the tendency for the x,y residuals to increase as the number of control points chosen also increases. This is to be expected. In general, considering the randomness of the digitizing errors in the GBF/DIME system and also in the base system, it seems advisable to utilize a large number rather than a sparse number of control points. Taking out control points that have large residuals is also counterproductive as far as the final map results are concerned. The orthogonal similarity transformation produced the poorest solution as was expected considering the nature of the problem.

TABLE 2
CONTROL POINT RESIDUALS (M.S.E)
(in feet)

Transformation Type	Number of Control Points: 8 Residual X	8 Residual Y	10 Residual X	10 Residual Y	20 Residual X	20 Residual Y
Linear Affine	24.5	18.0	25.8	20.1	31.6	26.9
Orthog. Similarity	27.4	19.6	26.6	20.7	36.1	36.5
Bilinear Polynom.	16.4	13.5	25.0	19.6	29.0	25.5
2nd Order Polynom.	15.7	12.4	23.9	14.4	24.3	23.6

One final point, the residuals must be made to vanish to avoid gaps or gores in the final product. The simple and effective solution utilized in this application was to force the GBF/DIME node at the control point to have the same x,y values as the digitized base point x,y values.

"RUBBERSHEETING" BASED ON TRIANGULATION

Development and Implementation of Delaunay Triangles

The preceeding approach of using global transformation parameters to fit one map into the framework of another (in this case GBF/DIME transformed to a Parcel Centroid base map) while yielding an improvement over the original untransformed composite, can still be improved by using a network of triangles connecting the control points in the map area. From the special analytic principles relating to the geometry of triangles, unique transformation parameters applicable to points only in each triangle, can be expected to result in better local correction of distortions instead of using a fixed set of transformation parameters considered applicable over the whole map area. The development of a computer solution for this methodology is more complex

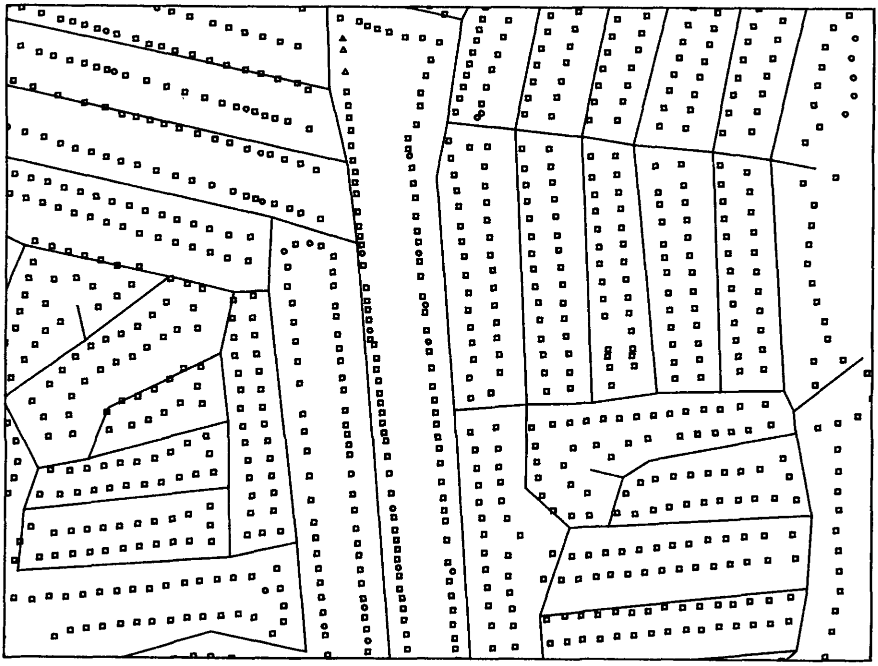

Figure 2
Illustration of Improvement by 2nd Order Polynomial Transformation

than the previously discussed approach and will be described below.

Background

For some time now, Delaunay triangles (Delaunay 1934) and triangulation facets covering areas or surfaces have been found useful for contour mapping and in finite element analysis (Gold, Charters and Ramsden 1977; Elfick 1979; McCullagh and Ross 1980; Lewis and Robinson 1977; Brassel and Reif 1979). More recently the U.S. Bureau of the Census has focused attention on the advantages of using Delaunay triangulation to transform detail from one map into the domain of another map base (Gilman 1985, White and Griffen 1985; Saalfeld 1985).

Concepts and Algorithms Required

Many triangle arrangements are possible to connect a series of control points in a map area. Only a Delaunay triangulation is unique and independent of the starting point. Delaunay triangles are characterized by the fact that they contain no other control points. Thus in Figure 3, if points 1 and 2 determine a possible base of a triangle (point 2 is the closest neighbor to 1) then a new neighbor to the right of line 1-2 must be found such that the circumscribing circle passing through points 1 and 2 and the other new point contains no other data point to form a Delaunay triangle.

Obviously, the proper selection of the new point to form a triangle requires checking into possibly numerous candidates in the proximity of point 2. It is necessary to have an efficient proximity selection algorithm and many of these have been proposed in the literature (Friedman, Baskett and Shustek 1975). In the case of this program development, elements of the X and Y coordinates of the control points were encoded into a character string or integer word using the features of FORTRAN 77 available on the VAX. This scheme develops a one dimensional representation when sorted of the two dimensional map space of discrete points and emphasizes proximity relationships. Using this concept, relatively few points had to be tried to properly form a Delaunay triangle.

Besides the need for a proximity analysis, building a Delaunay triangle requires a point of view in terms of clockwise or counter-clockwise rotation about the focal point 1 in Figure 3. The direction concept chosen in this application was clockwise. Therefore, the new vertex point 3 must be to the right of the line 1-2 as well as a close neighbor to point 2.

The process used to check the direction of the selected point is based on the fact that the equation of a straight line in a plane

$$Ay + Bx + C = 0 \qquad (13)$$

divides the plane into two half planes such that points in one half plane satisfy

$$Ay + Bx + C > 0 \qquad (14)$$

and those in the other half plane satisfy

$$Ay + Bx + C < 0 \qquad (15)$$

For the clockwise point of view in Figure 3, a point will lie to the right of a line if the parameters are computed as:

$$A = (x_2 - x_1)/2(-1) \qquad (16)$$

$$B = (y_1 - y_2)/2(-1) \qquad (17)$$

$$C = (y_2x_1 - y_1x_2)/2(-1) \qquad (18)$$

and if Equation (14) is satisfied.

Substituting the x and y coordinates of the proposed point into the equation Ay + Bx + C, the point can be quickly rejected or accepted in terms of direction. The final review of the new vertex is done by a subroutine which incorporates a distance check with near neighbors.

When a Delaunay triangle is formed it must be entered into *a database structure* that allows easy retrieval of triangle elements and adjacency relationships. The simple system involved array storage of:

Triangle Number, 3 Vertex Point Numbers, Adjacent Triangles (up to 3), and 6 Triangle Transformation Parameters.

The method elected to move the GBF/DIME file triangle vertex coordinates into exact correspondance with the base map triangle coordinates (see Figure 4) involved the use of the linear affine equations discussed earlier in this paper:

$$X_1 = a_1x_1 + a_2y_1 + a_3 \quad (19)$$
$$X_2 = a_1x_2 + a_2y_2 + a_3 \quad (20)$$
$$X_3 = a_1x_3 + a_2y_3 + a_3 \quad (21)$$
$$Y_1 = b_1x_1 + b_2y_1 + b_3 \quad (22)$$
$$Y_2 = b_1x_2 + b_2y_2 + b_3 \quad (23)$$
$$Y_3 = b_1x_3 + b_2y_3 + b_3 \quad (24)$$

where:
X_1, X_2, X_3 are the final base map coordinates and
x_1, x_2, x_3 are the original GBF/DIME coordinates.

This set of six equations based on 3 points is just sufficient to derive the 6 transformation parameters for an individual triangle. The solution of the equations was done with the same Gauss-Jordan algorithm used in the earlier investigations. It is to be noted that this type of solution using the global coordinates contrasts with the use of "local" coordinate systems with their own concepts which has been favored by other investigators. It produces the same transformed results.

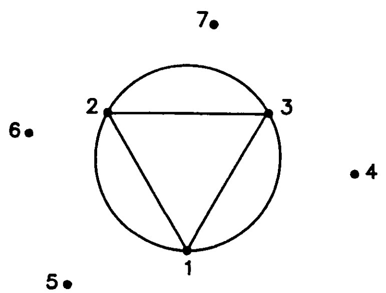

Figure 3. Creation of Delaunay Triangles

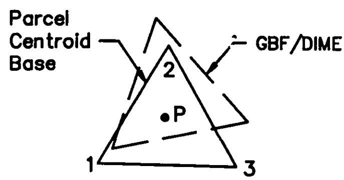

Figure 4. Generalization of Triangle Warpage

For the final transformation of the GBF/DIME nodes over the whole map area, each node point had to be identified as a point in a particular Delaunay triangle. To accomplish this the GBF/DIME nodes were placed in a file with an indexed structure which had the x,y coordinates encoded in the key similar to the earlier use in the point proximity algorithm described earlier.

By building an enclosing rectangle around each triangle and using the key structure to find a potential point, each point could be verified as being in a particular triangle by using the right of line testing method described previously. The triangle point transformations using the derived x,y transformation parameters was very quickly done. In this way each triangle, with its quota of points inside, was treated to provide the final total map area presentation in the base map system.

Summary and Conclusions

Figure 5 illustrates the type of solution achieved by using the triangulation approach. The scheme is particularly successful in handling layers of information including annotation in terms of preserving relationships for a pleasing composite map.

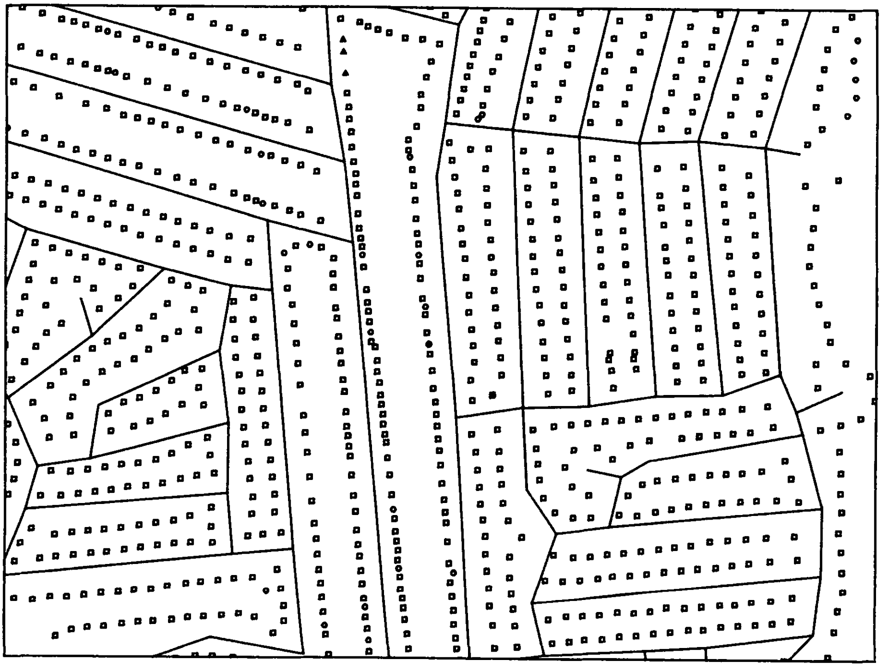

Figure 5.
Illustration of Improvement by Triangulation Using Delaunay Triangles

REFERENCES

Bahr, H.P. 1976, Geometric Models for Satellite Scanner Imagery: ISP Commission III, Helsinki.

Brassel, K. and Reif, D. 1979, A Procedure to Generate Thiessen Polygons: Geographical Analysis, Vol. 11, No. 3, 289-303.

Delaunay, B. 1934, Sur la sphere vide: Bulletin of the Academy of Sciences of the USSR, Classe Sci. Mat. Nat., 793-800.

Elfick, M. H. 1979, Contouring by Use of a Triangular Mesh: Cartographic Journal, Vol. 16, June, 24-29.

Fagan, G. L. 1986, Routing of Municipal Service Vehicles, Town of Oyster Bay, Long Island, New York: Proceedings of the Annual Convention of the American Society of Photogrammetry and the American Congress on Surveying and Mapping, Volume 3, Washington, D.C.

Friedman, J., Baskett, F. and Shustek, L. 1975, An Algorithm for Finding Nearest Neighbors: IEEE Transactions on Computers, October, 1000-1006.

Gilman, D. 1985, Triangulations for Rubber Sheeting: Proceedings. AutoCarto 7, Digital Representations of Spatial Knowledge, American Society of Photogrammetry and the American Congress on Surveying and Mapping, Washington, D.C., 191-197.

Gold, C., Charles, T. and Ramsden, J. 1977, Automated Contour Mapping Using Triangular Element Data: Computer Graphics, 11, 170-175.

Kratky, V. 1972, Image Transformations: Photogrammetric Engineering, May, 463-471.

Lee, D. T. and Schachter, B. J. 1980, Two Algorithms for Constructing a Delaunay Triangulation: International Journal of Computer and Information Sciences, Vol. 9, No. 3, 219-242.

Lewis, B. A. and Robinson, J. S. 1977, Triangulation of Planar Regions with Applications: The Computer Journal, Vol. 21, No. 4, 324-332.

McCullagh, M. and Ross, C. G. 1980, Delaunay Triangulation of a Random Data Set for Isorithmic Mapping: The Cartographic Journal, Vol. 17, No. 2, 93-99.

Mikhail, E. M. 1976, Observations and Least Squares: Harper & Row, New York.

Saalfeld, A. 1985, A Fast Rubber-Sheeting Transformation Using Simplicial Coordinates: The American Cartographer, Vol. 12, No. 2, 169-173.

White, M. and Griffen, P. 1985, Piecewise Linear Rubber-Sheet Map Transformation: The American Cartographer, Vol. 12, No. 2, 123-131.