# TERRAIN SURFACE APPROXIMATION AND ON-LINE QUALITY ASSESSMENT

James S. Bethel and Edward M. Mikhail
Kern Instruments, Inc.
Geneva Road
Brewster, NY, 10509, USA

School of Civil Engineering
Purdue University
W. Lafayette, IN, 47907, USA

Commission III WG III-3

ABSTRACT

Least squares regression using bicubic spline functions is discussed. The computational aspects are detailed for a parameterization with tensor product B-splines. The special case of gridded elevation observations is analyzed and the application of array algebra for this case is presented. A derivation is given for the array normal equations which ties in closely with the corresponding derivation for the classical least squares normal equations, and demonstrates the equivalence of the two approaches.

On-line quality assessment is investigated using the spline regression and a post adjustment analysis of the residuals. The residual analysis is performed by finding a maximum test statistic of externally studentized quadratic forms and associated critical values. The critical factor of bicubic patch size (or basis function spacing) is shown to influence the effectiveness of gross error detection. The particulars of this influence are presented as results of a statistical experiment using Digital Terrain Model, DTM, data and artificially induced 'spike' blunders.

## 1. INTRODUCTION

The use of digital terrain models for collecting, archiving, and graphically displaying topographic information is becoming increasingly accepted. Consequently we are experiencing a distinct transition from pure theory and research into the practical aspects of application to photogrammetry and cartography.

Terrain elevation data can be gathered in a manual mode from analog stereo instruments. Increasingly, however, various forms of computer assistance are being rendered in this often tedious process. Such assistance can range from automatic stepping in an analytical stereo instrument to a complete automation of the stereo perception and height sensing operation. As automation plays an ever larger role, the data may be created in prodigious quantities.

On-line quality control checks of observation data would be viewed as beneficial by many production managers. By this means one could minimize the risk of corrupting a data base, and minimize the resulting effort of cleaning and restoring the data base which is found to contain significant errors. From a more abstract point of view, the fundemental tasks in digital terrain modelling are:

(a) To observe a continuous surface at a set of discrete points, and

(b) To reconstruct, within error tolerances, the surface from the sampled points.

Step (b) implies the existence of a mathematical model by which one 'fills in' the reconstructed surface between the samples. Step (a) does not explicitly require such a model. If, however, a quality control check is incorporated into the observation process, then the existence of a mathematical model becomes necessary as a criterion to check for consistency among the observations. Thus the formation of a model is an important component of the process of collecting and using digital terrain data.

The terrain surface, unfortunately, is not expressible as an exact mathematical function. This is contrasted with the more tractable problems of photogrammetry and geodesy wherein well-known functions relate the observations and the unknown parameters. From the foregoing description of the general DTM problem, two specific and important issues have been extracted and chosen for investigation:

(1) Can we obtain a mathematical model or models which will follow arbitrary terrain shapes, and yet remain reasonably concise for efficient computation ?

(2) Can we devise an automated algorithm to check observations for gross inconsistency with the model, and hence for inconsistency with the terrain surface ?

## 2. THE MATHEMATICAL MODEL

The fact that most DTM data is collected in a regularly gridded pattern suggests using a surface function which expoits this regularity. Bicubic, tesor product spline functions possess this capability and were thus chosen for this investigation.

A two-dimensional tensor product function function may be thought of as a combination of two one-dimensional fuctions. If we have $f(x)$ and $g(y)$, then the function:

(2-1) $$h(x,y) = f(x) * g(y)$$

is called the tensor product of f and g and may also be denoted by:

(2-2) $$h = f \otimes g$$

This is analogous to the way in which a bivariate density function is constructed from the univariate density functions of two independent random variables. If the two component functions happen to be B-splines, then the resulting 'hill' function is referred to as a tensor product B-spline. Two-dimensional spline regression may be performed using the two-dimensional B-spline basis functions rather than the usual polynomial parameters and constraints. The unknowns that we solve for in a least squares estimation are the coefficients of a 'matrix' of overlapping basis functions.

Recalling the gridded pattern of the height observations suggests the use of what has become known in the photogrammetric/geodetic community as array algebra (Rauhala, 1978). Under certain conditions of the mathematical model and of the observation pattern, this method allows a greatly reduced computational effort to solve what appears to be a much more massive problem. The conditions under which it may be used are:

(1) The parameter coefficient matrix or design matrix of the mathematical model must be decomposable into a kronecker product
(2) The observations must occur in a full, gridded pattern
(3) The observation weight matrix must be decomposable as in condition (1)

The spline model, the gridded observations, and the assumption of uncorrelated, equal precision observations together fulfill these three conditions.

The following represents an outline of the derivation of the normal equations for the case of a 'seperable' model and gridded observations. The condition equations may be written:

(2-3) $$V + B_1 \Delta B_2^t = F$$

in which V, , and F are rectangular matrices. It is equivalent to the classical formulation:

(2-4) $$v + B\Delta = f$$

in which B is decomposed as:

(2-5) $$B = B_1 \otimes^R B_2$$

and correspondingly the weight matrix:

(2-6) $$W = W_1 \otimes^R W_2$$

where $x^R$ is the reverse kronecker product. In the classical approach we minimize the quadratic form of the residuals:

(2-7) $$v^t W v \rightarrow \text{minimum}$$

Under our chosen assumptions the quadratic form may be written equivalently as:

(2-8) $$v^t W v = tr(W_2 V^t W_1 V)$$

where tr() indicates the trace of the matrix brackets. This equality is proven in Neudecke this new expression for the quantity to be minimized, we solve Equation (2-3) for the parameter matrix, $\Delta$ , subject to the condition that:

(2-9) $$tr(W_2 V^t W_1 V) \rightarrow \text{minimum}$$

Expanding this and using the derivative of a trace with respect to a matrix (See Turnbull, 1930), we obtain after a number of algebraic steps:

(2-10) $$(B_1^t W_1 B_1)\Delta(B_2^t W_2 B_2) = B_1^t W_1 F W_2 B_2$$

Using auxiliary matrices, Equation (2-10) becomes:

(2-11) $$N_1 \Delta N_2^t = T$$

This represents the normal equations for the array or multilinear model. It may be solved by triangular decomposition or by inversion, depending on the explicit need for the inverse(s).

## 3. GROSS ERROR DETECTION

Having fit a bicubic surface to a set of height observations, we now wish to examine the residuals for the existence of possible outliers. The full quadratic form for the entire set of residuals has the form:

(3-1) $$q^2 = v^tWv$$

We partition the observation vector, and likewise the residual vector into two parts:

(3-2) $$l = \begin{bmatrix} l_1 \\ l_2 \end{bmatrix} \qquad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

such that $l_1$ is sufficient to define the model, and $l_2$ is a (usually small) subgroup to be tested for the presence of gross errors. If the adjustment were carried out using only the $l_1$ set of observations, we would obtain the reduced quadratic form:

(3-3) $$\dot{q}^2 = v_1{}^tW_1v_1$$

in which $W_1$ corresponds to the $l_1$ set. This reduced quadratic form, $\dot{q}^2$, will always be of smaller value than the full quadratic form, $q^2$. It has been shown (Stefanovic, 1978) that the difference:

(3-4) $$d = q^2 - \dot{q}^2$$

can be obtained from the results of the original full adjustment by:

(3-5) $$d = v_2{}^tQ_{v_2v_2}{}^{-1}v_2$$

in which $Q_{v_2v_2}$ is the submatrix of the full $Q_{vv}$ which corresponds to the subvector $v_2$. Thus, $\dot{q}^2$ itself may be obtained without recomputation of the adjustment by rearranging Equation (3-4). The redundancy, or degrees of freedom in the adjustment is denoted by r. The number of residuals in the test group $v_2$ is p.

We are, in general, looking for a situation in which d is large in relation to its degrees of freedom, and $\dot{q}^2$ is small. This suggests forming the following ratio:

(3-6) $$\frac{\dfrac{d/\sigma_0^2}{p}}{\dfrac{\dot{q}^2/\sigma_0^2}{r-p}} \approx \frac{\dfrac{\chi_p^2}{p}}{\dfrac{\chi_{r-p}^2}{r-p}}$$

and look for that partition of v which maximizes this test statistic with respect to its distribution under a null hypothesis of no blunders. This maximum chi-squared ratio (MCSR) is an example of 'external studentization' and has been suggested by several statisticians (Cook, 1982). In order to obtain a preliminary idea of what the density function of the MCSR would look like it was decided to simulate this random variable using monte-carlo methods. A program was developed to obtain n N(0,1) deviates using a library pseudo-random number generator. For each of p=1,2,3,4,5 the maximum of:

(3-7) $$\frac{\chi_p^{2*}/p}{\chi_{n-p}^{2*}/n-p}$$

was found and tabulated as an entry in a histogram for each of the p's. As a check, a numerical approximation of the distribution has also been achieved by starting with the basic random variables in the MCSR and proceding via the mathematical propagation of distributions. This has been implemented in a series of computer programs, and runs have been made to duplicate the cases estimated by the histogram simulations. Figure 1 shows one final MCSR density (the shaded curve) overlayed with the histogram from the simulation. They should be representing the same function and it is seen that the agreement is reasonably good. This

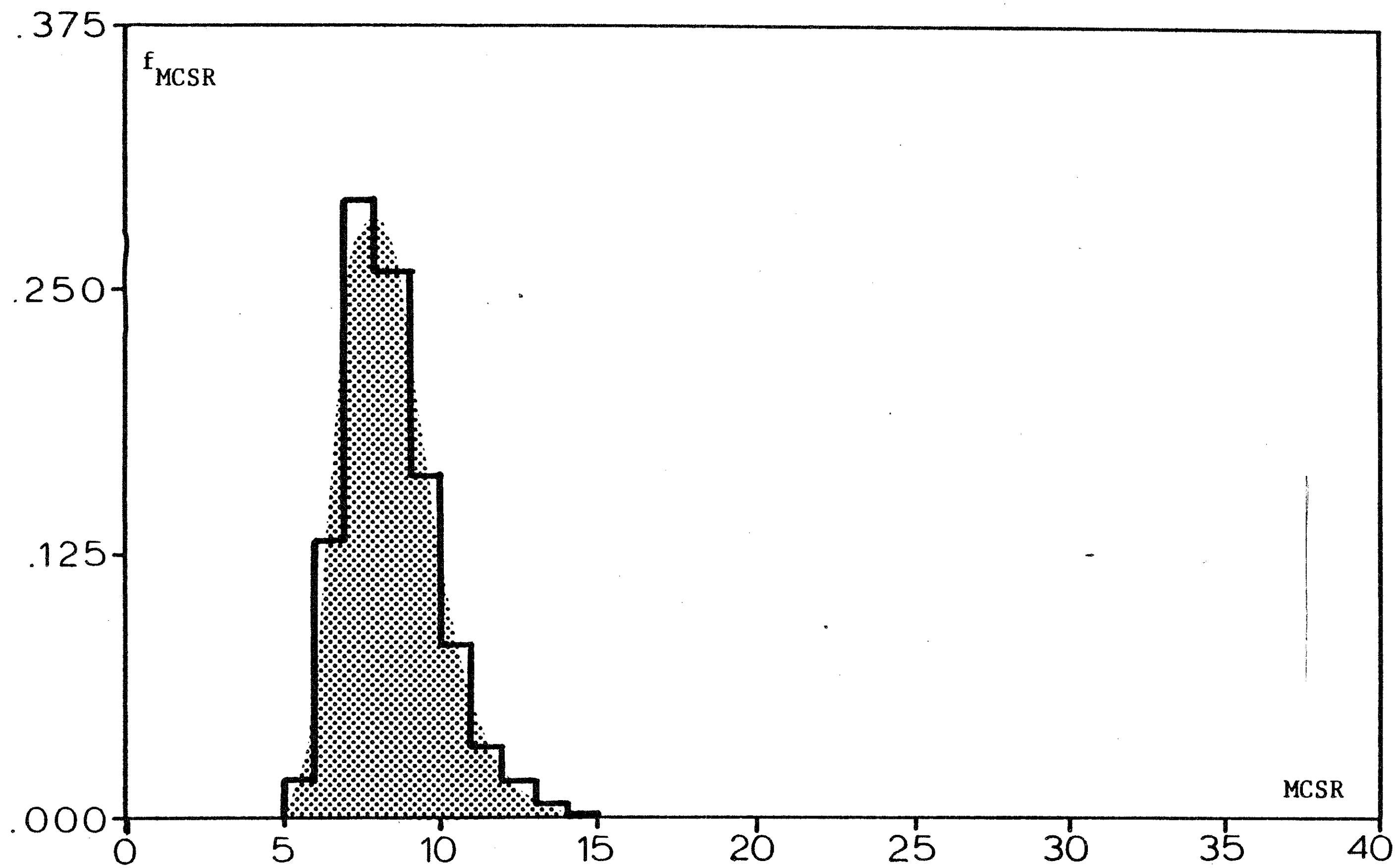


Figure 1. MCSR Density Function for d.f.=(3,228)

provides some confirmation that the analysis and its numerical realization are valid. A flow chart of the strategy for analyzing the observations for gross error detection is shown Figure 2.

## 4. EXPERIMENTAL ANALYSIS AND RESULTS

The approach investigated here for quality control of on-line DTM acquisition involves checking small areas of observed elevation 'posts' and insuring that these observations are free of gross errors. This checking may take place immediately following observation, i.e. on-line, or at a later time in a batch mode. In either case, the small check regions would successively cover the entire data set so that all observations would be subject to the test for blunders. In the experiment described here, the check region is a square area having 16x16=256 observations.

An inherent characteristic of terrain topography is that the 'order' of the surface representing it changes from one geographic area to the next. Significant variation in this surface order may often occur within a single photogrammetric model, or certainly within a quadrangle or map sheet cartographic unit. This presents an apparent difficulty to the quality assurance analyst, or to an automated equivalent, in trying to insure that the raw observations define an approximation to a valid topographic surface. The naturally occurring variation in terrain heights will be confounded with any observation errors. Thus an algorithm for detecting large observation errors should address the question of this possible confounding and how it may vary with terrain character. An intuitive approach to surface fitting for different terrain types would involve densifying the mesh of bicubic patches of the spline function for 'higher order' or more rugged terrain areas. Ideally, the residuals after a surface should be comparable in magnitude to the uncertainty in the observation process. The purpose of the experimental work described here is to look at the effects of terrain type and mesh density (model type) on the effectiveness of the gross error detection algorithm.

The degree to which a patchwise polynomial function can conform to an irregular surface is directly related to the mesh size of the patches. If the mesh size diminishes to the point where the number of parameters is equal to the number of sample points, then the fit will be perfect as there is no redundancy. In this experimental work, four mesh sizes have been chosen for modelling the 16x16 point sample regions: model one: 1 patch, model two: 2x2 patches, model three: 4x4 patches, model four: 8x8 patches. Two of these are shown in Figure 3. The experiments involved selecting examples of different terrain types (classes) from a DTM data base, and artificially injecting 'spike' blunders. The blunders were introduced singly and in multiples with magnitudes of five and ten feet, where the sample spacing was 32 feet. The surface functions at four mesh densities were fit in each case and the 'success' (percentage of valid detections) was assigned a

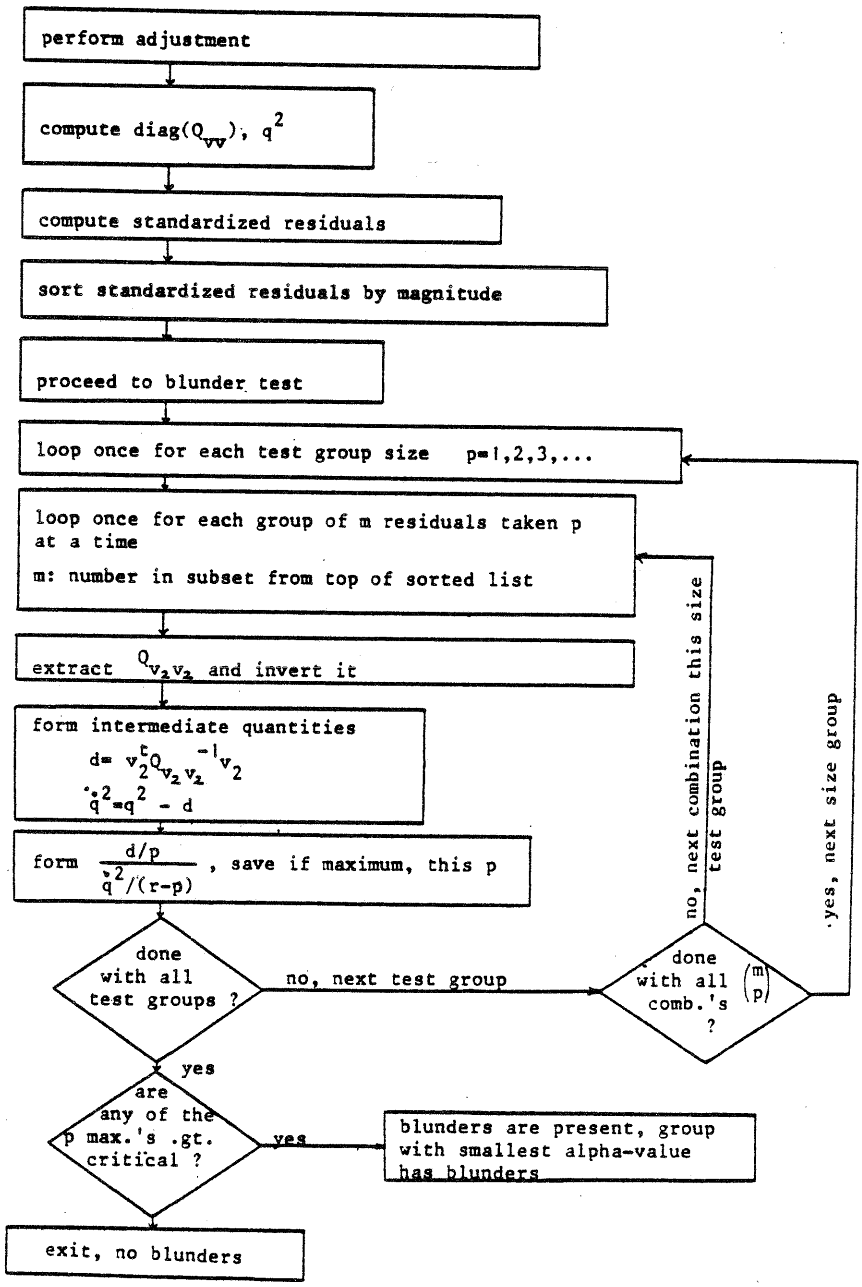


Figure 2. Flowchart of Strategy for Blunder Detection

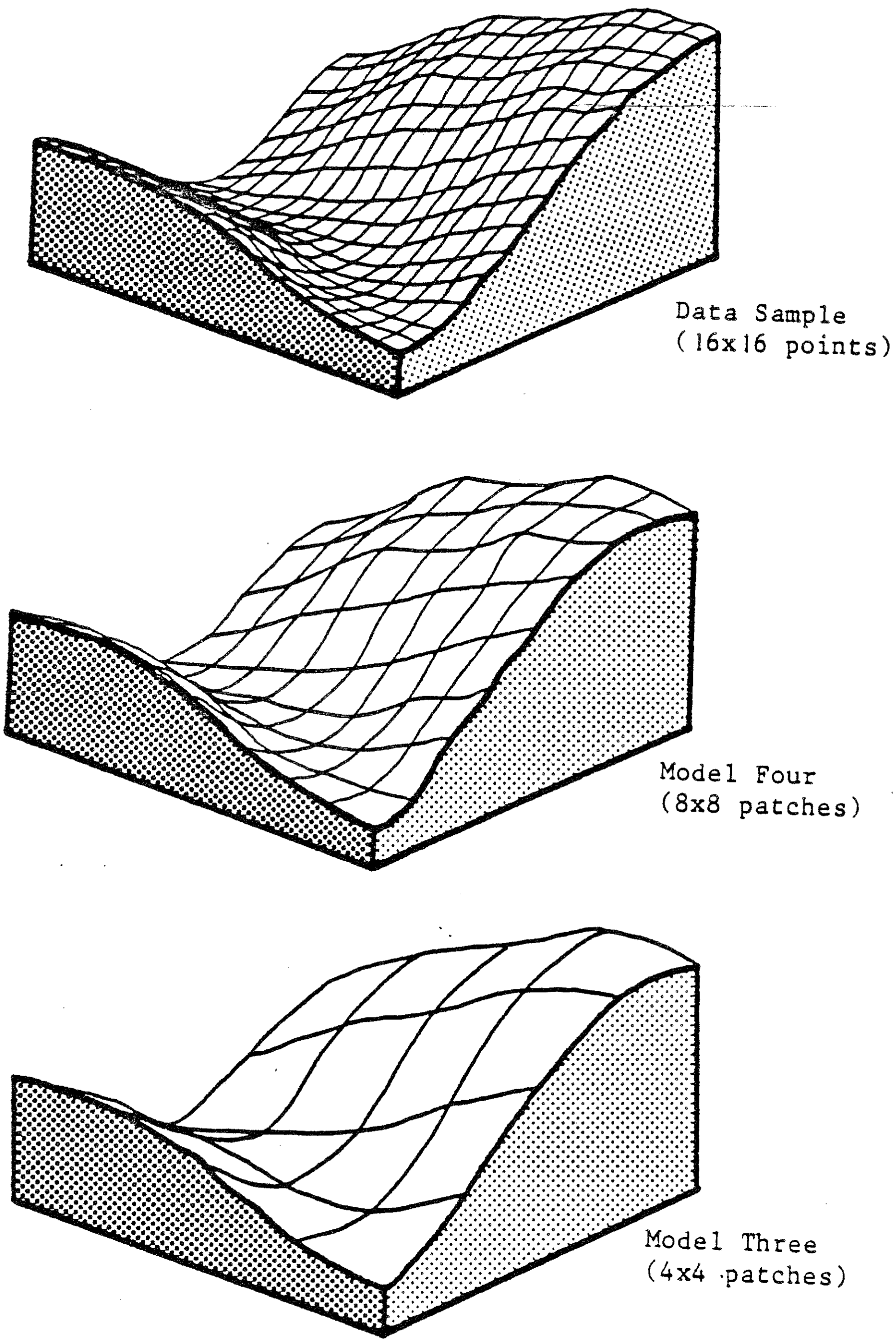


Figure 3. Data Sample and Two Spline Models

score between 0 and 1. An important consideration was the relation between the terrain class and the spline mesh density (model). A summary of the scores are given for model 1 in Table 1.

There are significant differences between classes for several of the size/number combinations for models one and two. In addition, for both model one and two the most significant effects occur for the small error magnitude. Also interesting is the fact that the class effects are apparent but not significant for models three and four. It appears that there is a significant degradation in the blunder detection ability with a low order model in high order terrain, but the reverse does not hold. To some extent, this confirms what one would expect from an intuitive point of view. The factor, size of blunder, was not tested but, as expected, the effectiveness of the detection algorithm is in every case higher for the large size blunder. Similarly, no testing is done for the factor of number of blunders, and no consistent pattern appears here. This itself is useful information, however, in that the algorithm (as intended) appears to function in the presence of multiple blunders (up to three) without serious degradation.

## 5. CONCLUSIONS

In summary, it has been demonstrated that it is possible to detect and isolate a small number of gross errors in gridded DTM observations in a manner suitable for on-line implementation. This has been accomplished by fitting, via least squares, a simultaneous patchwise polynomial (bicubic spline) to a small set of observations and analyzing the residuals for the presence of outliers. The statistical testing of the residuals is done through an externally studentized quadratic form, for which the maximum occurrence is found (maximum chi-squared ratio, MCSR). The critical values of this statistic are derived and this derivation is believed to provide more accurate critical values than other approximate methods. The bicubic spline modelling approach, while used as an intermediate step in the blunder detection, has obvious utility in the more extended application of interpolation over large data sets.

## 6. REFERENCES


1. Cook, R., and Weisberg, S., 1982, Residuals and Influence in Regression, Chapman and Hall, NY
2. Neudecker, H., 1969, 'Some Theorems on Matrix Differentiation with Special Reference to Kronecker Products', Journal of American Statistical Association, vol. 64, p.953
3. Rauhala, U., 1978, 'Array Algebra DTM' Proceedings of the DTM Symposium, ASP, St. Louis
4. Stefanovic, P., 1978, 'Blunders and Least Squares', ITC Journal, 1, p. 122
5. Turnbull, H. 1930, 'A Matrix Form of Taylor's Theorem', Proceedings of the Edinburgh Mathematical Society, II

## 7. ACKNOWLEDGEMENTS

This research was partially supported by the Defense Mapping Agency Aerospace Center (DMAAC) under contract number DMA 700-82-C0013, and by Teledyne Geotronics (TG). The authors would like to thank Mr. David Alspaugh of DMAAC and Mr. David Baker of TG for their cooperation and counsel.


Table 1. Summary of Experimental Results for Model 1

MODEL 1

1-BLUNDERS

| | CLASS1 | CLASS2 | CLASS3 | CLASS4 |
|---|---|---|---|---|
| SMALL | .800<br>5 | .800<br>5 | .000<br>1 | .333<br>9 |
| LARGE | 1.000<br>2 | 1.000<br>5 | .857<br>7 | .667<br>6 |

2-BLUNDERS

| | CLASS1 | CLASS2 | CLASS3 | CLASS4 |
|---|---|---|---|---|
| SMALL | .875<br>8 | .750<br>4 | .100<br>5 | .000<br>3 |
| LARGE | 1.000<br>4 | 1.000<br>7 | .750<br>6 | .000<br>3 |

3-BLUNDERS

| | CLASS1 | CLASS2 | CLASS3 | CLASS4 |
|---|---|---|---|---|
| SMALL | .944<br>6 | .750<br>4 | .467<br>5 | .200<br>5 |
| LARGE | 1.000<br>4 | 1.000<br>7 | 1.000<br>4 | .800<br>5 |

In each cell: upper value = cell mean; lower value = number of observations in the cell.

| | | |
|---|---|---|
| CLASS1 | .920 | N= 29 |
| CLASS2 | .906 | N= 32 |
| CLASS3 | .619 | N= 28 |
| CLASS4 | .387 | N= 31 |
| SMALL | .558 | N= 60 |
| LARGE | .858 | N= 60 |
| 1-BLUNDERS | .700 | N= 40 |
| 2-BLUNDERS | .650 | N= 40 |
| 3-BLUNDERS | .775 | N= 40 |