Evaluation and Comparison of Spatial Interpolators II[1,2]

Dennis D. Weber[3] and Evan J. Englund[4]

DRAFT: August 5, 1993 - for submission to Mathematical Geology

The performance of several variations on ordinary kriginq and inverse distance estimators is evaluated. Mean squared errors (MSE) were calculated for estimates made on multiple resamplings from five exhaustive data bases representing two distinctly different types of estimation problem. Ordinary kriging, when performed with variogram estimated from the sample data, ***was more robust than inverse-distance methods to the type of estimation problem, and to the choice of estimation parameters such as number of neighbors.***

Keywords: kriging, geostatistics, spatial estimation, inverse-distance estimation

INTRODUCTION

Spatial interpolation is important in many environmental studies. The U.S. EPA Environmental Monitoring Systems Laboratory -Las Vegas has been investigating the performance of various interpolators, especially as they apply to sampling, estimation, and remediation of contaminated soils and sediments. Previous studies by the authors have investigated the effects of various estimation parameters on the quality of spatial estimates. Englund (1990) showed that the variability was high among estimates (obtained primarily by kriging) by 12 different statisticians working with two common sets of data. Englund, et al., (1992) investigated the effects of sample size, grid type, and sampling error on estimation accuracy, by using 54 sample data sets drawn from a single large exhaustive data base. The most important parameter proved to be sample size, where the estimation accuracy improved with increasing sample number.

Weber and Englund (1992) evaluated the relative accuracy of 15 different spatial estimators by using the same 54 sample data sets, showing that inverse-distance and inverse-distance-squared interpolators performed slightly better than ordinary and simple kriging. The authors concluded that these results, while provocative, did not mean that inverse-distance methods are necessarily superior to kriging estimators

---

[1] Received.!..;accepted....

[2] Notice: Although the research described in this article has been funded in part by the United States Environmental Protection Agency through Cooperative Agreement CR818526 to the Harry Reid Center for Environmental Studies, University of Nevada-Las Vegas, it has not been subjected to Agency review. Therefore it does not necessarily reflect the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

[3] University of Nevada Las Vegas, Harry Reed Center for Environmental Studies, 4505 S. Maryland Parkway, Las Vegas, Nevada 89154.

[4] U. S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory - Las Vegas, P.O.Box 93478, Las Vegas, Nevada 89193

in all cases; reasonable variations of that experiment could be imagined wherein kriging would be expected to have a distinct advantage over the particular inverse-distance algorithms used. The nature of the data base may have fortuitously favored inverse-distance. Both the kriging methods and variogram modeling were relatively simplistic: changing either might have significantly altered the results. Finally, strong anisotropy and clustering of samples, which favors kriging, were not present in the data.

In this paper, we begin to address these issues with a more extensive comparison of several inverse-distance and kriging interpolators. We evaluate their relative performance on five exhaustive data bases that represent distinctly different types of physical phenomena.

DATA BASE DESCRIPTIONS

The five data bases, like the Walker Lake data base used in the earlier studies, were taken from digital elevation models obtained from the National Cartographic Information Center. The goal was to select several data bases representing surfaces that have different characteristics. Three of the data bases use the original elevation data, and two of them are transformations obtained by calculating the variance of the original elevation data. All of the data bases contain 21,600 data on a common grid of 120 rows by 180 columns. Corresponding 20 by 30 arrays of block averages were generated from the arithmetic means of 6 by 6 arrays of points in each of the data bases. Statistical parameters are given below in table I. Figures la-le show three-dimensional perspective views of the five data bases, and Figures 2a-2e show the corresponding histograms. The elevation data bases are relatively unskewed, with fairly smooth, continuous surfaces. We believe that they can fairly represent other phenomena sharing these characteristics: geological structural surfaces: thickness of lithologic units, hydraulic head, surface water temperature, barometric pressure, etc. The variance data bases are highly skewed, and present discontinuous, noisy surfaces. These distributions are similar to those of contaminants or trace elements in soils, sediments, and rocks; and to phenomena such as porosity and hydraulic conductivity.

1. **Bend Elevation and Variance-of-Elevation:** The original data base contained 240 rows by 360 columns of elevation data. Two final data bases (120 by 180) were obtained by calculating the mean or variance of each 2 by 2 array of the original data. The most prominent features of the elevation base (Fig. la) are the mountains on the east side. The terrain slopes down toward the west side where canyons are found. The most prominent features of the variance base (Fig. lb) are produced by the valleys on the west side and the mountains on the east side.

2. **Black Butte Elevation and Variance-of-Elevation:** Elevation and

variance data base were calculated as for Bend. The elevation base (Fig. 1c) includes Black Butte in the far northeast corner and several prominent mountains in the southwest corner. The area surrounding Black Butte is very flat, as opposed to the landscape around the mountains in the Bend elevation area. The prominent features in the original elevation model produce corresponding prominent variance features (Fig. 1d).

3. **Steamboat Falls Elevation:** An original data base of 180 columns by 120 rows was used **as** the final data base. The mutual confluence of two creeks with the North Umpqua river produces an interesting surface divided into four sections (Fig. 1e).

Table 1
Statistics for Data Bases and Block Averages

Data Base Statistics

Data Base Data Type	Bend Elevation	Bend Variance	Butte Elevation	Butte Variance	Steam Elevation
Mean	1,682	142	1,347	48	704
St. Dev.	411	255	253	112	209
Skewness	.16	4.24	.28	5.43	.45
Kurtosis	2.89	35.97	2.22	46.19	2.67
CV	.24	1.80	.18	2.34	.30

Block Average Statistics

Data Base Data Type	Bend Elevation	Bend Variance	Butte Elevation	Butte Variance	Steam Elevation
Mean	1,682	142	1,347	48	705
Std. Dev.	408	158	252	78	204

SAMPLE SETS

Sixty sample sets were drawn by random sampling from each data base, including 20 sets each of 25, 100, and 400 samples. The total number of sample sets was 300. Figure 3 shows typical sample configurations.

ESTIMATORS

For each of the 300 sample sets, a 20 by 30 grid of block estimates was produced by each estimator. In all cases, blocks were numerically approximated by a discrete 2x2 array of point estimates. Estimators included four variations on ordinary kriging, four variations on inverse-distance, and one spline, as described below. For two of the estimators, inverse-distance squared and ordinary kriging with traditional variogram estimation, the estimates were rerun with different numbers of neighboring samples included.

Geo-EAS - Ordinary Kriging: Ordinary kriging was performed by using Geo-EAS software (Englund and Sparks, 1991). The ordinary kriging equations have been described elsewhere (e.g. , Journel and Huijbregts, 1978; Isaaks and Srivastava 1989). A circular, single sector search was used, with a large enough search radius to ensure the maximum desired number of neighboring sample locations. The four variations on ordinary kriging differ only in the choice of variogram model used. In one case, a "black box" approach was used - a zero-nugget linear model was assigned regardless of the experimental variogram. The other three ordinary kriging estimators used variogram models fitted to three alternate estimators of the experimental variogram, namely, the traditional variogram, the general, or lag-wise, relative variogram (David, 1977), and Cressie's (1985) robust variogram. These are described in detail below. In these three cases, isotropic double-nested exponential models were fitted by minimizing the squared residual function (SR) given by

$$SR = \sum_{i=1}^{N} [F_{kk}(\vec{h}_i) - \gamma_X(\vec{h}_i)]^2 \cdot \omega_Y(\vec{h}_i)$$

where the functions $F_{kk}(\vec{h}_i)$, $\gamma_X(\vec{h}_i)$, and $\omega_Y(\vec{h}_i)$ are given below.

The three experimental variograms $\gamma_X(\vec{h}_i)$ are defined as

$$\gamma(\vec{h}) = \frac{1}{2N(\vec{h})} \sum_{i=1}^{N(\vec{h})} \left[z_{t_i+\vec{h}} - z_{t_i}\right]^2 \qquad (Traditional)$$

$$\gamma_{GR}(\vec{h}) = \frac{\gamma(\vec{h})}{\left(\frac{m_{-\vec{h}} + m_{+\vec{h}}}{2}\right)^2} \qquad (General\ relative)$$

The maximum distance over which the experimental variogram were computed and fitted was determined by finding, in each case, the greatest distance needed to include the selected number of nearest neighbors. This distance

$$2\gamma_{NC}(\vec{h}) = \frac{\left(\frac{1}{N_{\vec{h}}}\sum_{i=1}^{N_{\vec{h}}}\sqrt{\left|z_{t_i+\vec{h}} - z_{t_i}\right|}\right)^4}{(0.457+0.497/N_{\vec{h}})} \qquad \textit{(Cressie's Robust)}$$

was divided into n equal-length lag intervals. The numbers of pairs in each lag were used to weight the least squares. The three variogram models are:

$$G_{exponential\ 1} = Nugget + Sill_1 \cdot \left(1-e^{-\frac{x}{1.732R_1}}\right) + Sill_2 \cdot \left[1-e^{-\frac{x}{1.732R_2}}\right]$$

$$G_{Gauss} = Nugget + Sill_1 \cdot \left[1-e^{-\frac{x^2}{3R_1^2}}\right] + Sill_2 \cdot \left[1-e^{-\frac{x^2}{3R_2^2}}\right]$$

$$G_{spherical} = Nugget + Sill_1 \cdot \left[\frac{3x}{2R_1} - \left(\frac{x}{2R_1}\right)^3\right] + Sill_2 \cdot \left[\frac{3x}{2R_2} - \left(\frac{x}{2R_2}\right)^3\right]$$

where R is the "practical" range (Journel and Huijbregts, 1978).

CPS/PC[5] - Inverse-distance Squared: This estimator, from the CPS/PC software package, is defined as:

$$Z_o = \sum_{i=1}^{n} w_i Z(x_i)$$

where Z_o represents the estimated value, w_i are weights, $Z(x_i)$ are sample values at locations xi, and the summation is over the n samples included in the estimate. The weights for inverse-distance squared (w^{IDS}) are defined as

$$w_i^{IDS} = \frac{\left(\frac{r_s - r_i}{r_i}\right)^2}{\sum_{i=1}^{n}\left(\frac{r_s - r_i}{r_i}\right)^2}$$

where r_i is the distance between estimate and the i[th] sample location, and r_s is the search radius. Note that these weighting equations are not "simple" inverse-distance schemes because sample weights are forced to equal zero at the search radius.
An octant search with a maximum of 16 points was used.

---

[5] CPS/PC is a commercial software product of Radian Corporation, Austin, Texas.

Surfer[6] - Inverse-distance, Inverse-distance Squared, and Inverse-distance Cubed: This estimator used the Surfer Version 4 software package. It is defined as:

$$Z_o = \sum_{i=1}^{n} w_i Z(x_i)$$

where Z_o represents the estimated value, w_i are weights, $Z(x_i)$ are sample values at locations x_i, and the summation is over the n samples included in the estimate. The weights for inverse-distance are defined as

$$w_i^p = \frac{\left(\frac{1}{r_i}\right)^p}{\sum_{i=1}^{n}\left(\frac{1}{r_i}\right)^p}$$

where r_i is the distance between estimate and the i^{th} sample location, and p is the inverse-distance exponent. A single sector search with a large search radius was used.

BIHASH - Spline: Spline estimates were performed by using a program developed by Foley (1987). This multi-stage method has many options controlled by an integer array of six elements that enable it to yield several different functions that either interpolate or approximate the given scattered data. As applied in this study, the first stage generated gridded estimates on a default uniform grid by using a polynomial least-squares approximation of degree two on eight neighboring sample values. For the variance sample sets, these estimates were constrained by the minimum and maximum sample values used. This was not done for elevation data. The second stage computes a piecewise bicubic Hermite interpolant with partial derivatives calculated from first stage grid. This interpolant does not necessarily pass through the sample values. The third stage adds a correction term to force the estimate to honor the data. Foley does not recommend this option for rapidly varying data; therefore, it was used elevation data, but not for variance data.

Mean Square Error: Evaluation was performed by calculating the mean square error (MSE) of the block estimates with respect to the true block values. MSE is defined here as
where $Z^{estimate}$ and Z^{true} are the estimates and true block values for the blocks, and i and j represent the blocks and sample data sets,

[6] Surfer is a commercial software product of Golden Software, Inc., Golden, Colorado.

$$MSE = \frac{1}{20}\sum_{j=1}^{20}\left[\frac{1}{600}\sum_{i=1}^{600}\left(Z_{ij}^{estimate}-Z_{i}^{true}\right)^2\right]$$

respectively. The quantity used here is the average MSE taken over the 20 sample data sets of one size from one data base. Each estimator thus receives 15 average MSE scores.

To aid in comparing relative performance, each of these average MSE scores was also normalized to a Z-score, $[MSE_i(k)-MSE_{GM}(k)]/SD_{GM}(k))$ where MSE_{GM} and SD_{GM} are the "grand" means and standard deviations, defined as

$$MSE_{GM}(k) = \frac{1}{N_{est}(k)}\sum_{i=1}^{N_{est}(k)} MSE_i(k)$$

$$SD_{GM}(k) = \frac{1}{N_{est}(k)}\sum_{i=1}^{N_{est}(k)} (MSE_{GM}(k) - MSE_i(k))^2$$

where $MSE_i(k)$ is the ith estimator of the kth group. The grand means and standard deviations for the 15 groups (Table 2) were computed from a larger group of 34 estimators, only 12 of which are presented in this paper. The Z-score representation allows a quick comparison of results because an entry below zero is better than the average estimator performance for a particular sample size and data base.

Table 2
Grand Means and Grand Standard Deviations

Samples/	BEA		BUA		STA		BEV		BUV	
data set	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
25	33,644	3,716	15,260	1,652	31,104	2,645	35,959	6,520	7,164	596
100	13,531	3,127	6,264	1,108	10,705	2,575	26,145	2,463	5,341	310
400	3,947	1,371	1,635	708	2,837	1,072	17,394	1,876	2,566	159

RESULTS

The results for all estimators are shown together as raw MSE scores in Figures 4a-4e, and as Z-scores in Figure 5a-5e. The designation numbers refer to estimators and corresponding parameters listed in Tables 3 and 4. Each graph shows the results for all estimators at 25, 100, and 400 samples for one data base. Figure 6a-6e shows z-scores for inverse-distance vs. kriging for different numbers of neighbors.

Table 3
Group I: Inverse-distance Estimators

Inv. Dist. Program	Exponent	Search Pattern Sectors	Samples/Sector
Surfer	2	1	04
Surfer	2	1	12
Surfer	2	1	20
Surfer	1	1	16
Surfer	2	1	16
Surfer	3	1	16
CPS/PC	2	8	02

Exponent is the power of the inverse-distance

Table 4
Group II: Kriging & Spline

Estimator	Experimental	Variogram Model	Fitting	Sectors	Search Pattern Samples/Sector
Spline	None	None	None	N/A	N/A
Kriging	None	Linear	None	1	16
Kriging	Traditional	Exponential	NP	1	20
Kriging	Traditional	Exponential	NP	1	12
Kriging	Traditional	Exponential	NP	1	4
Kriging	Gen. Rel.	Exponential	NP	1	16
Kriging	Cressie Robust	Exponential	NP	1	16

DISCUSSION

Inverse-distance estimators are very sensitive to the type of data base, to the number of neighbors used in the estimate, and to the power of distance used in weighting. For the elevation data bases, estimation quality generally increases as the number of neighbors decreases and as the power of distance increases. For variance data bases, the opposite occurs. . . there are dramatic increases in quality with increasing numbers of neighbors and decreasing powers of distance. The modified weighting scheme in the CPS/PC algorithm appears to have an effect roughly equivalent to a slight increase in the power.

By contrast, ordinary kriging using fitted variogram models is relatively robust to the type of data base and the method of estimating the experimental variogram. The kriging estimates consistently improve with increasing number of neighbors, regardless of data type. Four nearest neighbors appear to be generally inadequate for kriging, especially for the variance data bases; however, the results appear to stabilize with 12 or more neighbors. The results from ordinary kriging with the linear variogram model illustrate that it is not robust to an arbitrary choice of variogram model. This choice of linear model with zero nugget worked very well for elevation data bases, but extremely poorly for variance.

The relative performance of the estimators, with the exception of the spline, does not appear to be greatly affected by sample density.

From these results, it is apparent that the superior performance of inverse-distance over kriging reported earlier (Weber and Englund, 1992) was mostly due to a fortuitous (or judicious) selection of options with respect to the data base being used.

With experience, good judgement, and knowledge of the type of phenomenon being estimated, it is possible to obtain good estimates with inverse-distance methods. With the same experience, judgement and knowledge, however, it is also possible to select a good ordinary kriging approach with lower sensitivity to errors in judgement.

Although ordinary kriging is relatively robust, there are nevertheless significant differences in performance among the approaches used here. For example, the traditional variogram appears superior for elevation data bases, while the general relative variogram is better for the variance cases. Cressie’s robust variogram does in fact perform robustly, but not more so than the traditional variogram. The geostatistician is still faced with the problem of selecting the best tool for the job at hand. It is worth asking whether, in the absence of a *priori* knowledge, it is possible to use some characteristic of the sample data set itself to assist in this choice. In the examples presented here, skewness and kurtosis might provide an effective classifier.

The different approaches to variogram estimation presented here only begin to address the problem of variography. We have compared alternative variogram estimators, models, and fitting techniques in considerable detail (Englund and Weber, in preparation) .

Finally, we would like to emphasize that while MSE provides a simple and useful basis for comparison, minimization of MSE is not necessarily always what is needed. We have shown (Weber and Englund, 1992) that the relative ranking of estimator with respect to decision quality as measured by a loss function can be quite different from the ranking with respect to MSE.

REFERENCES

Cressie, N., 1985, Fitting Variogram Models by Weighted Least Squares, Math. Geol. v.17, n.5.

David, M., 1977, Geostatistical Ore Reserve Estimation, Elsevier Scientific Publishing Co., 364 pp.

Englund, E. J., 1990, A Variance of Geostatisticians: Math. Geol., v. 22, p. 417-456.

Englund, E. J. and Sparks, A. R., 1991, Geo-EAS 1.2.1 User's Guide, EPA/600/8-91/008: U.S. EPA, Las Vegas, 132 pp.

Englund. E.J. and Weber, D.D., and Leviant, N. 1992, The Effects of Sampling Design Parameters on Block Selection: Math. Geol. v.22, n.3.

Foley, T. A., 1987, Interpolation and Approximation of 3-D and 4-D Scattered Data, Comput. Math. Applic. v.12, n.8, 711-740.

Isaaks, E. H., and Srivastava, R. M., 1989, An Introduction to Applied Geostatistics: Oxford University Press, New York, 561 p.

Journel, A. G. and Huijbregts, Ch. J., 1978, Mining Geostatistics: Academic Press, London, 600 pp.

Weber, D. D. and Englund, E. J., 1992, Evaluation and Comparison of Spatial Interpolators, Mathematical Geology, Vol. 24, No. 4.

FIGURE 1

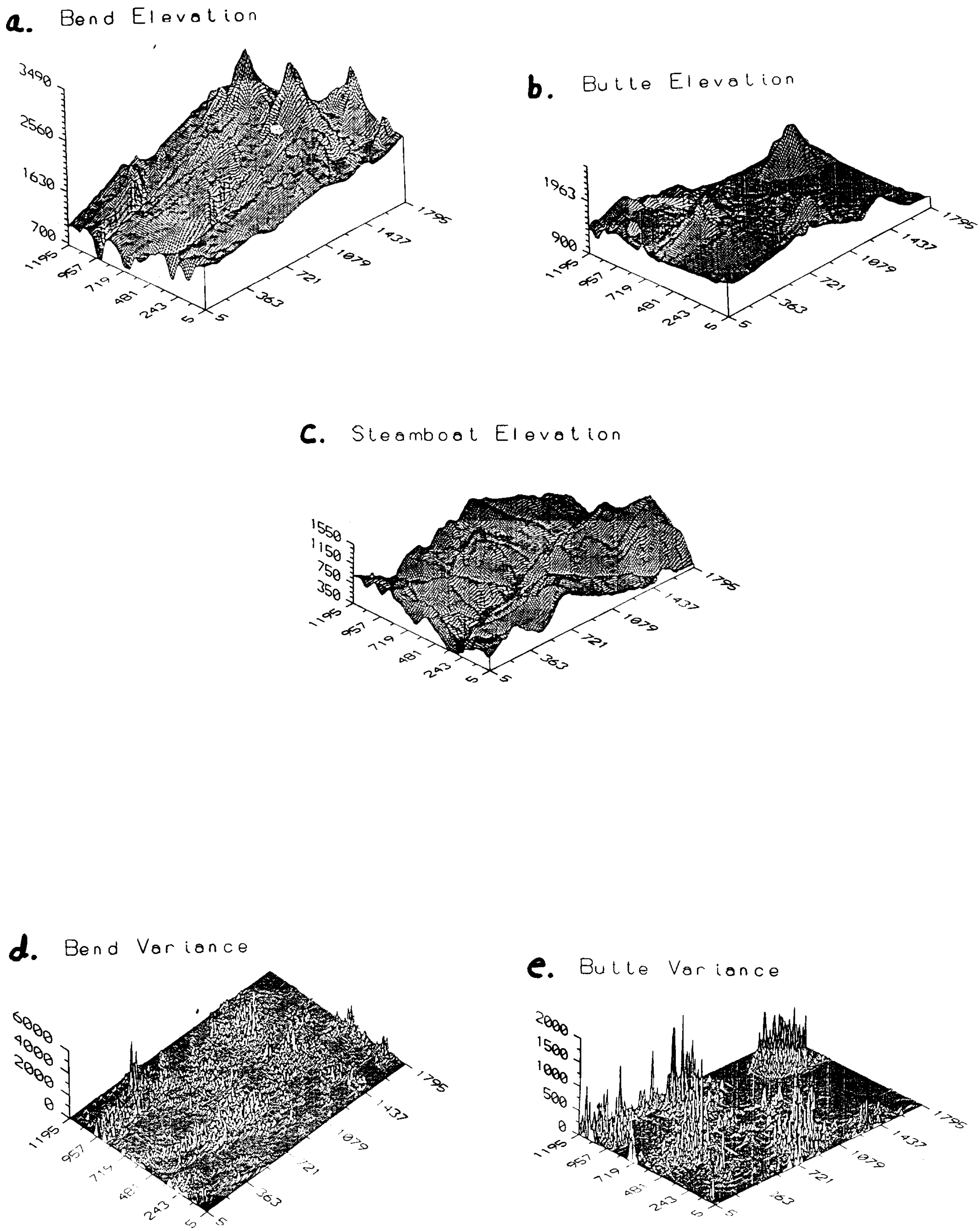

a. Bend Elevation
3490
2560
1630
700
1195
957
719
481
243
5
5
363
721
1079
1437
1795
b. Butte Elevation
1963
900
c. Steamboat Elevation
1550
1150
750
350
d. Bend Variance
6000
4000
2000
0
e. Butte Variance
2000
1500
1000
500
0

FIGURE 2

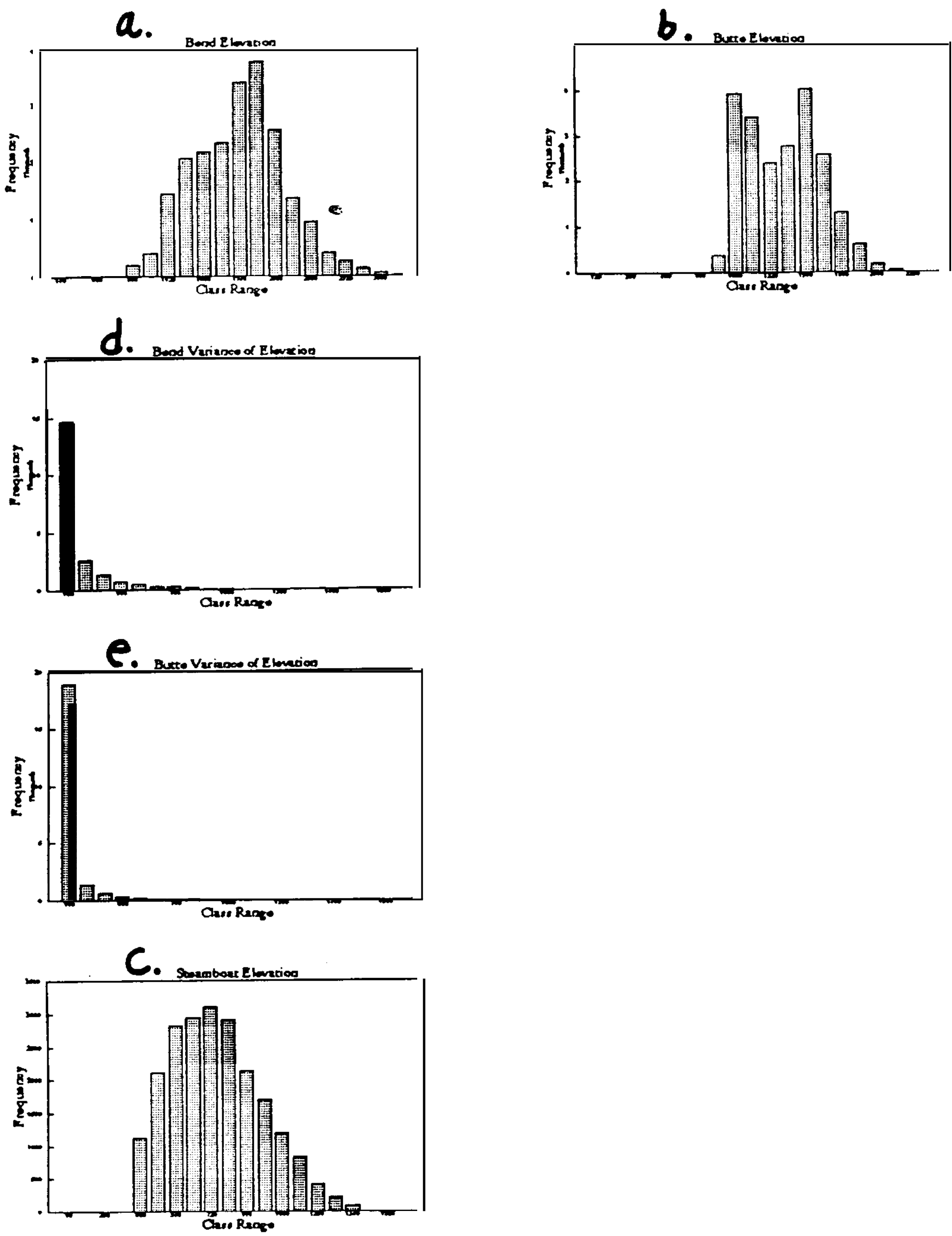

a.
Bend Elevation
Frequency
Class Range
b.
Butte Elevation
Frequency
Class Range
d.
Bend Variance of Elevation
Frequency
Class Range
e.
Butte Variance of Elevation
Frequency
Class Range
c.
Steamboat Elevation
Frequency
Class Range

FIGURE 3

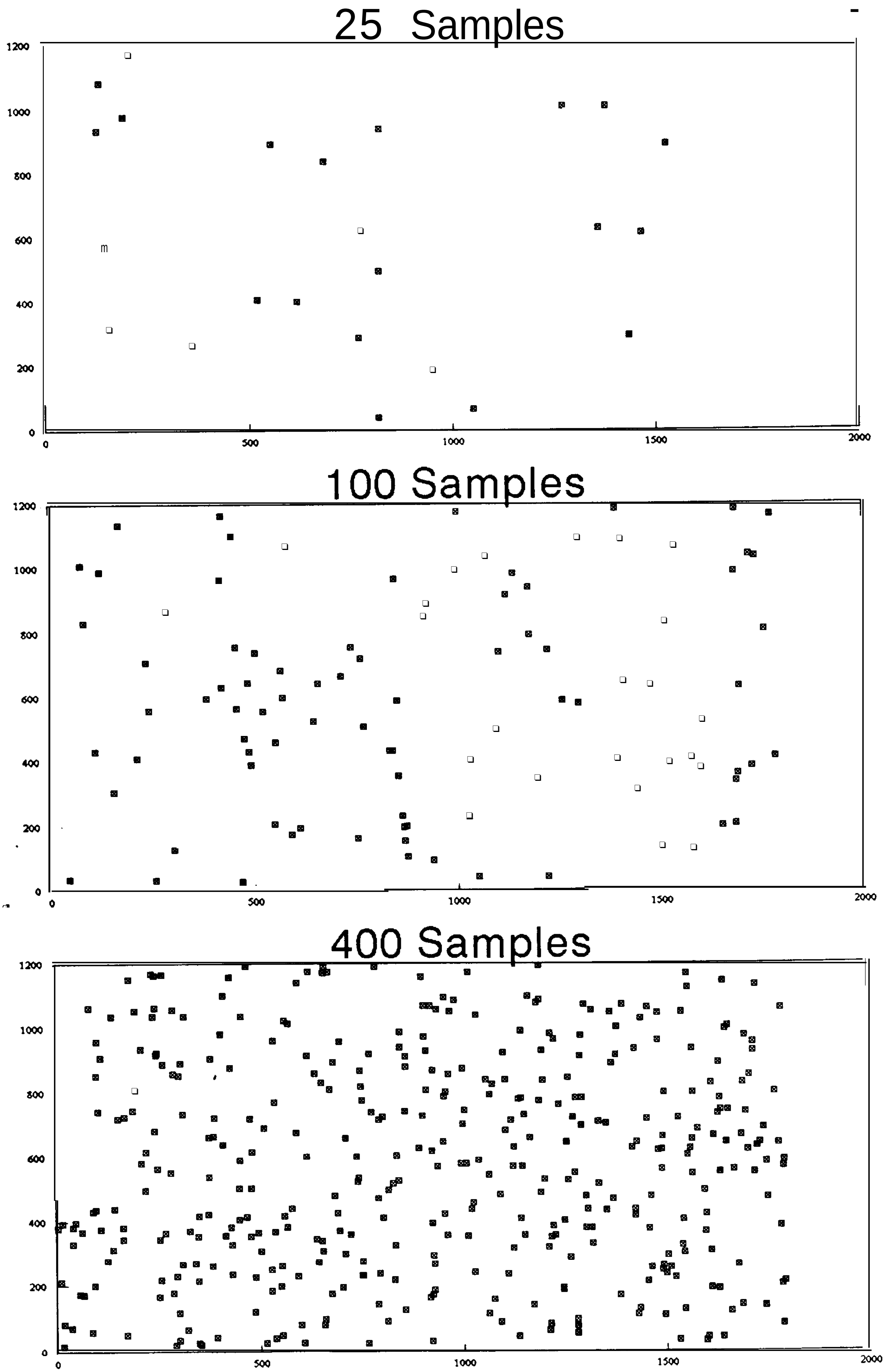

25 Samples
1200
1000
800
600
400
200
0
0
500
1000
1500
2000
100 Samples
1200
1000
800
600
400
200
0
0
500
1000
1500
2000
400 Samples
1200
1000
800
600
400
200
0
0
500
1000
1500
2000

FIGURE 4

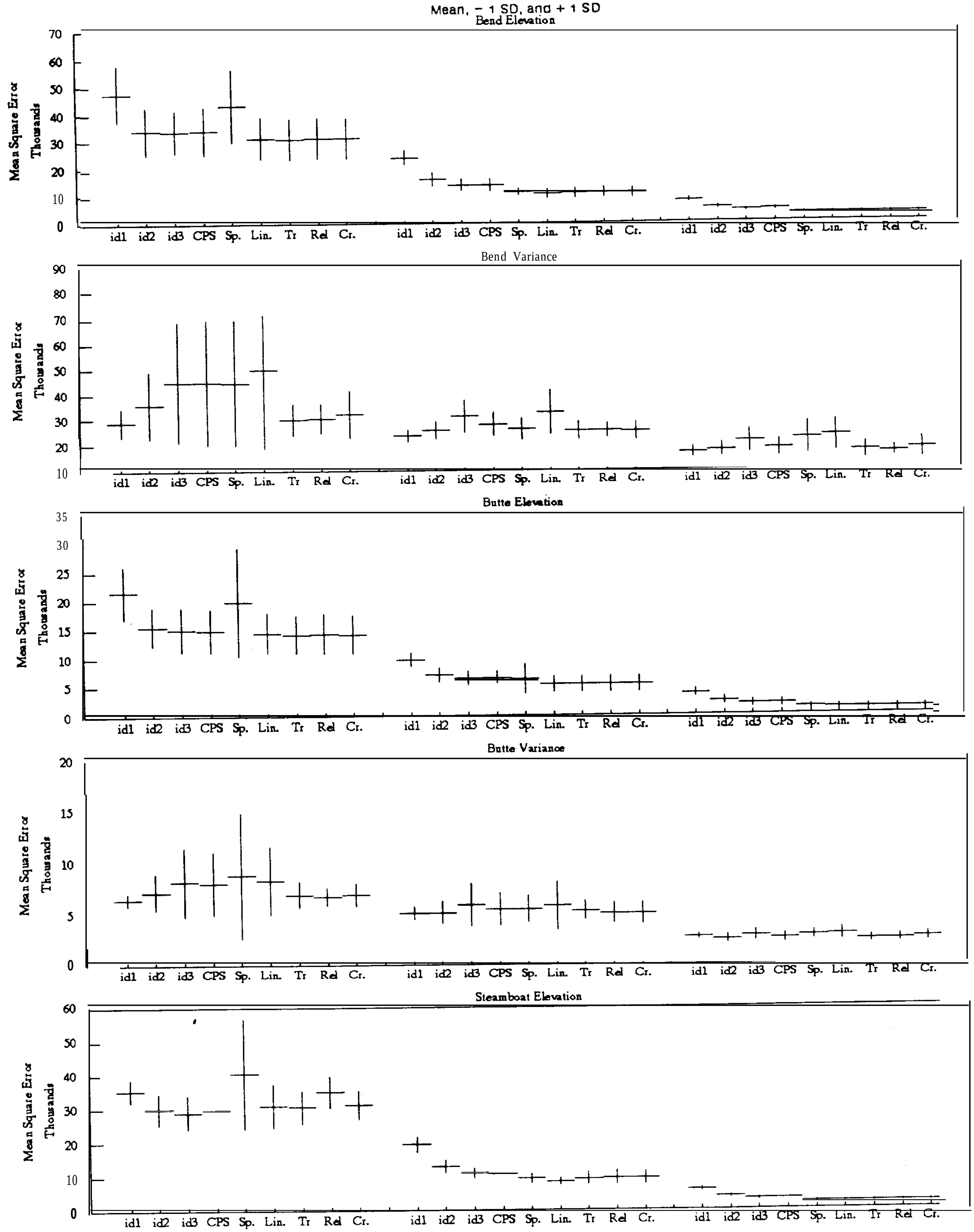
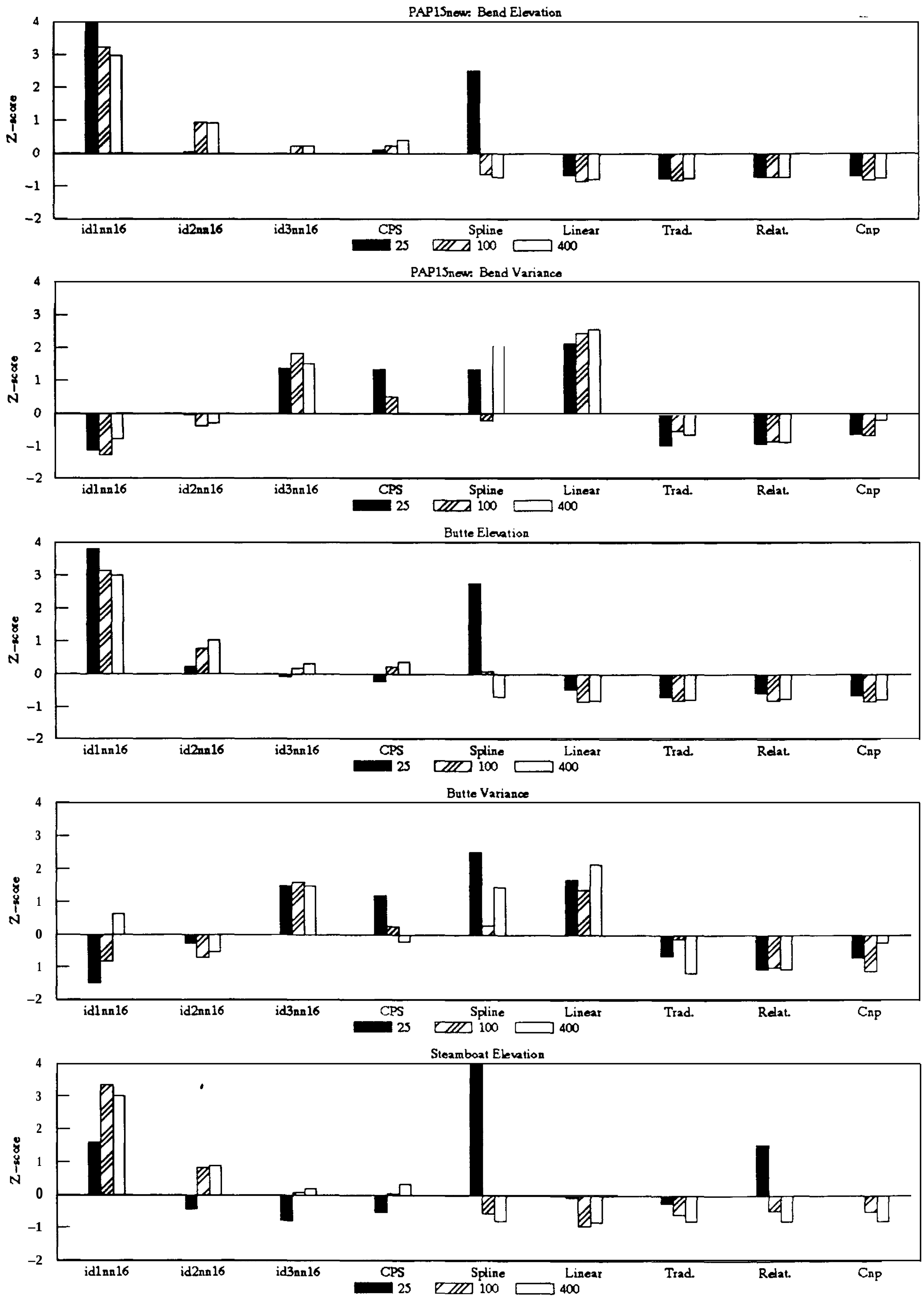
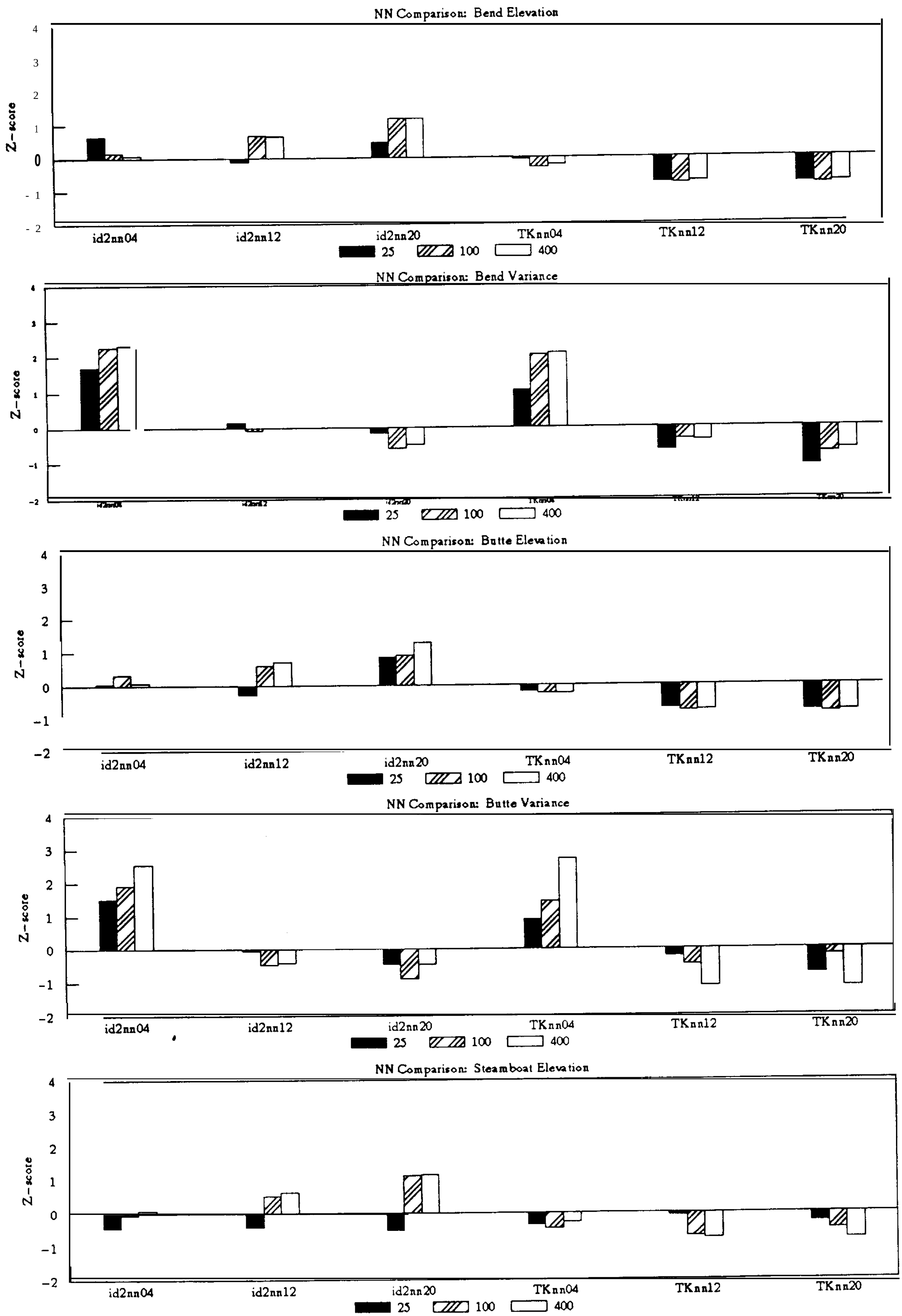

Mean, - 1 SD, and + 1 SD
Bend Elevation
Mean Square Error
Thousands
id1 id2 id3 CPS Sp. Lin. Tr Rel Cr.
Bend Variance
Butte Elevation
Butte Variance
Steamboat Elevation

FIGURE 5

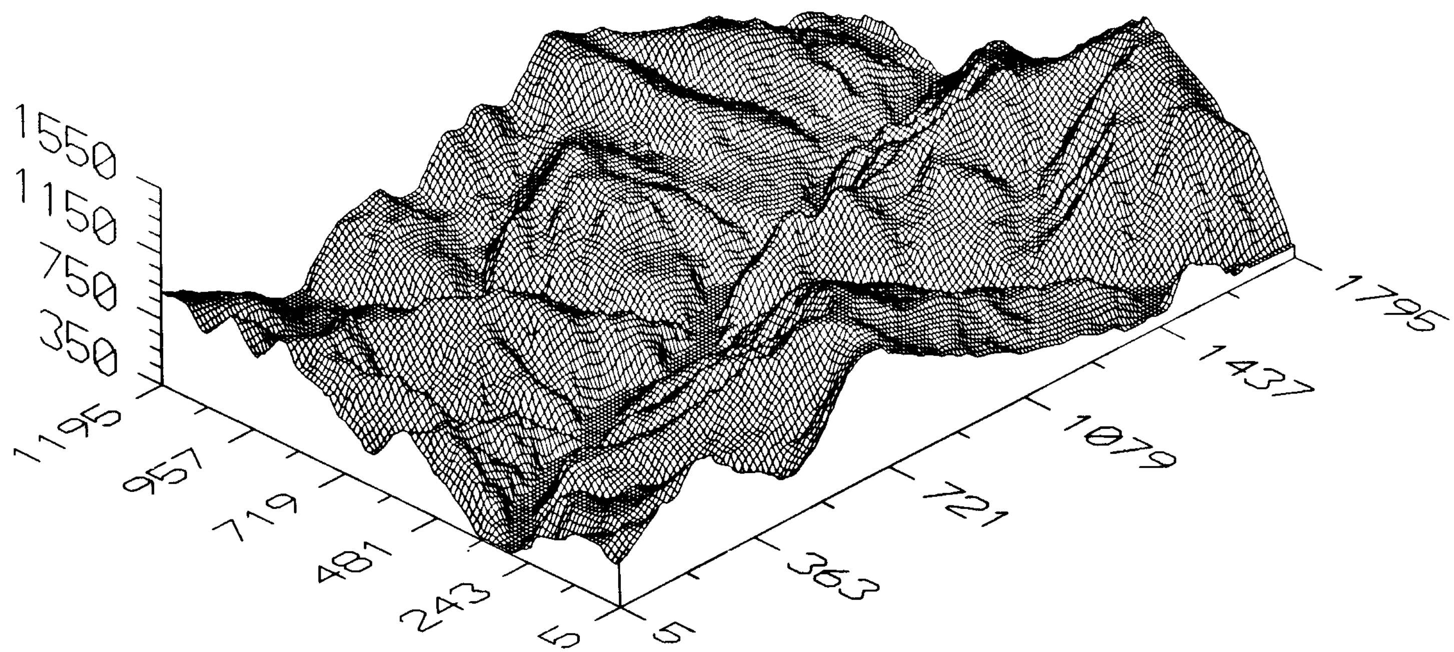


PAP15new: Bend Elevation
Z-score
id1nn16
id2nn16
id3nn16
CPS
Spline
Linear
Trad.
Relat.
Cnp
25
100
400
PAP15new: Bend Variance
Z-score
id1nn16
id2nn16
id3nn16
CPS
Spline
Linear
Trad.
Relat.
Cnp
25
100
400
Butte Elevation
Z-score
id1nn16
id2nn16
id3nn16
CPS
Spline
Linear
Trad.
Relat.
Cnp
25
100
400
Butte Variance
Z-score
id1nn16
id2nn16
id3nn16
CPS
Spline
Linear
Trad.
Relat.
Cnp
25
100
400
Steamboat Elevation
Z-score
id1nn16
id2nn16
id3nn16
CPS
Spline
Linear
Trad.
Relat.
Cnp
25
100
400

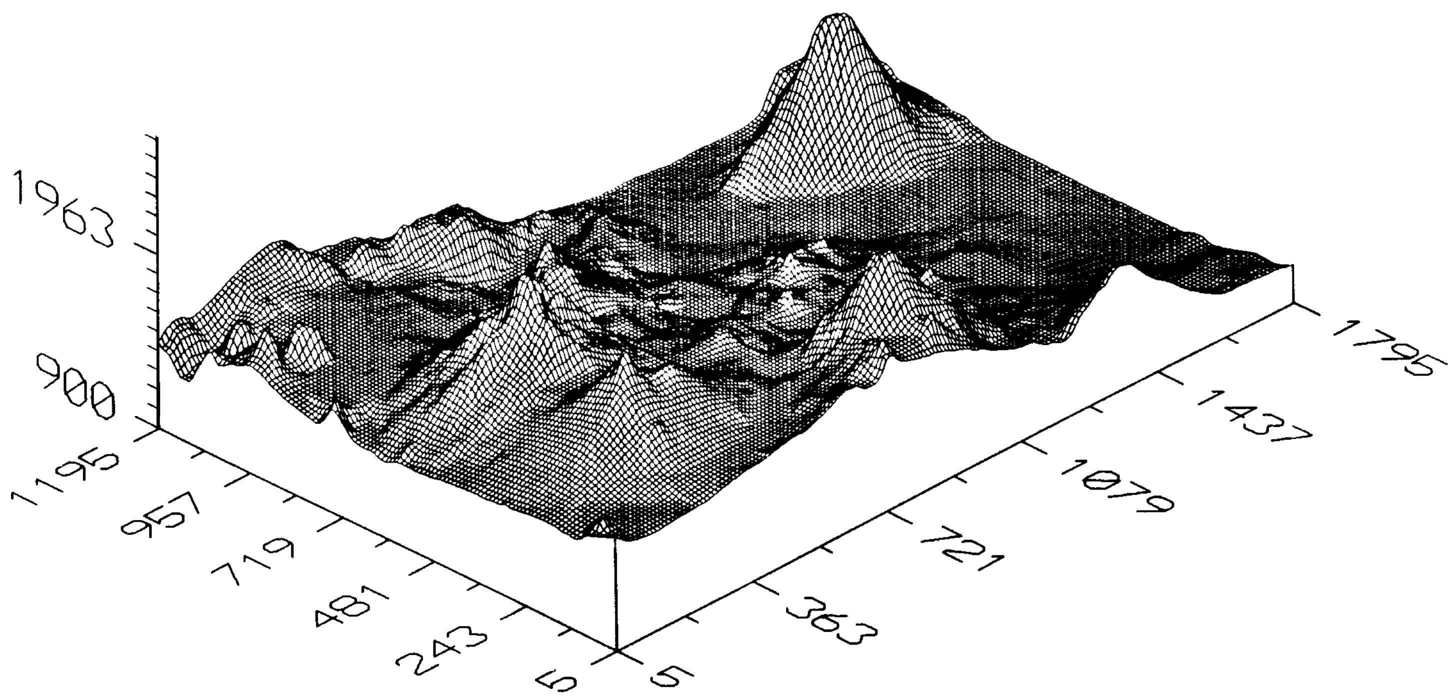
FIGURE 6


NN Comparison: Bend Elevation
Z-score
id2nn04
id2nn12
id2nn20
TKnn04
TKnn12
TKnn20
25
100
400
NN Comparison: Bend Variance
Z-score
25
100
400
NN Comparison: Butte Elevation
Z-score
id2nn04
id2nn12
id2nn20
TKnn04
TKnn12
TKnn20
25
100
400
NN Comparison: Butte Variance
Z-score
id2nn04
id2nn12
id2nn20
TKnn04
TKnn12
TKnn20
25
100
400
NN Comparison: Steamboat Elevation
Z-score
id2nn04
id2nn12
id2nn20
TKnn04
TKnn12
TKnn20
25
100
400

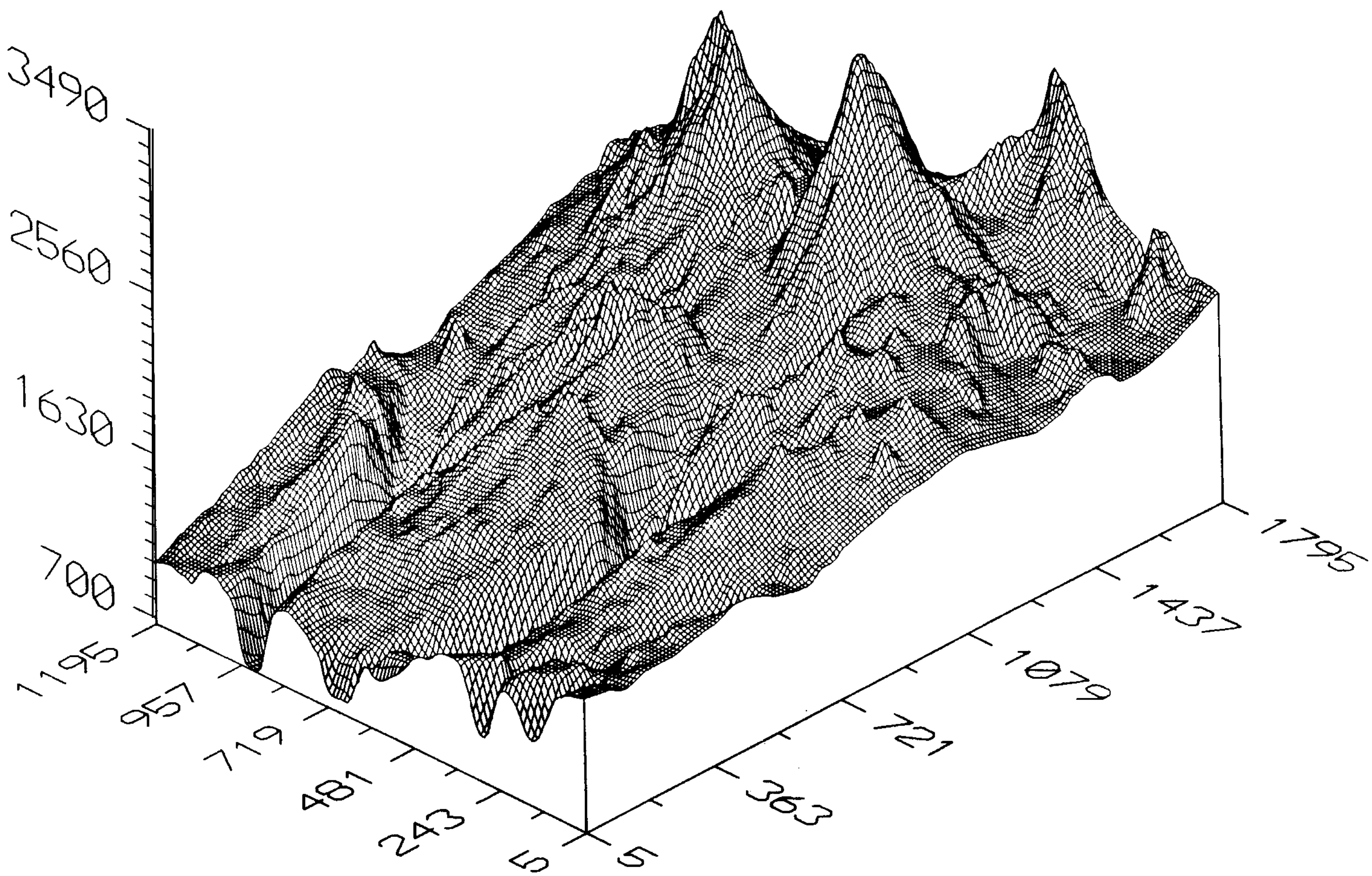
Steamboat Elevation



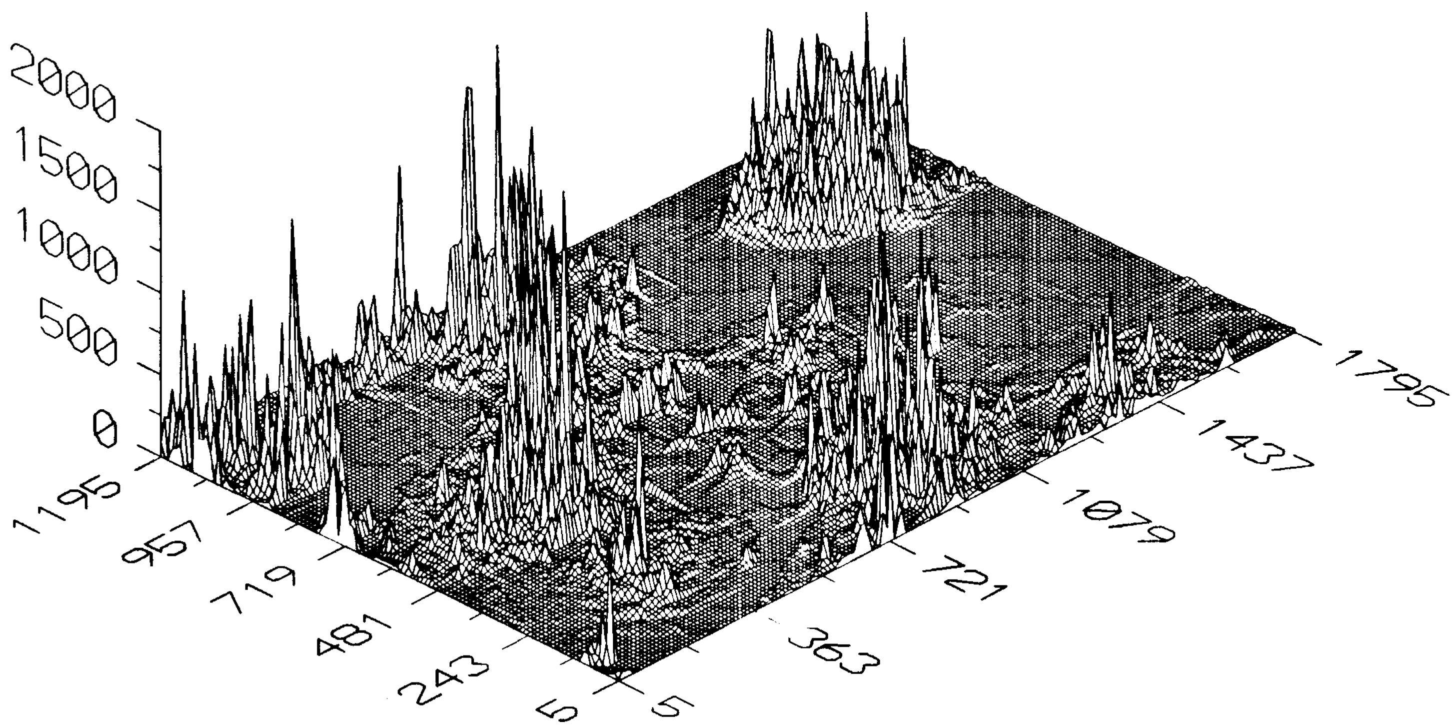
Butte E evation

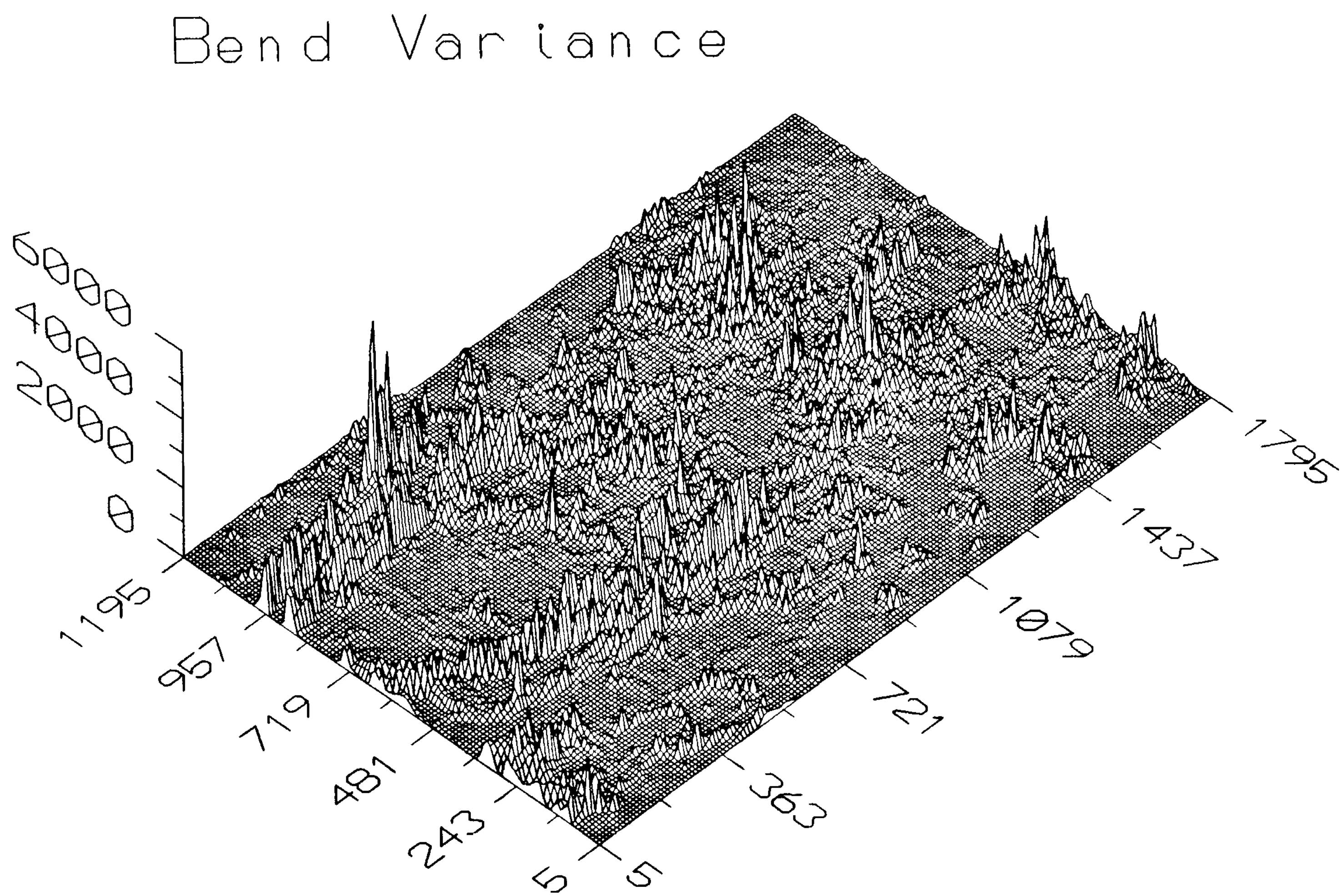


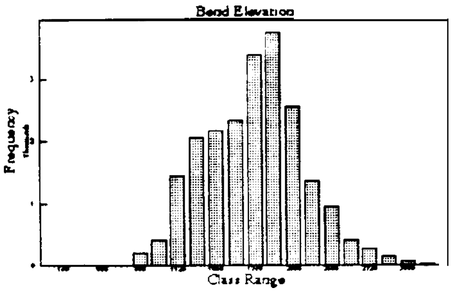
Bend E evation

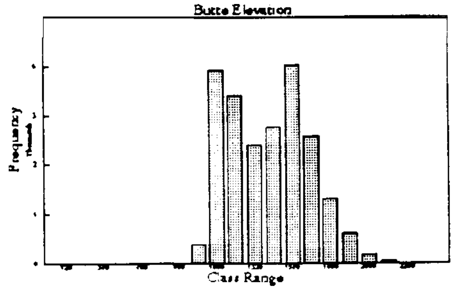


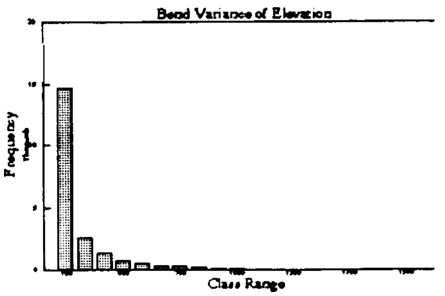
Butte Variance

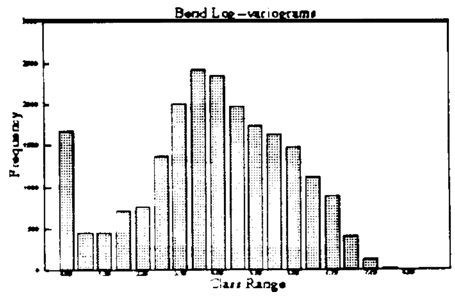


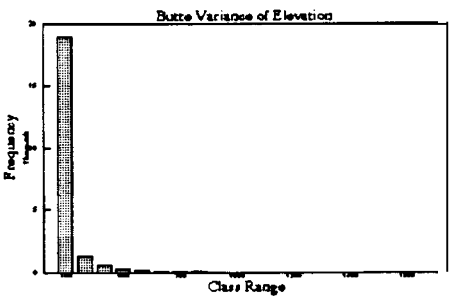

Bend Variance
6000
4000
2000
0
1195
957
719
481
243
5
5
363
721
1079
1437
1795

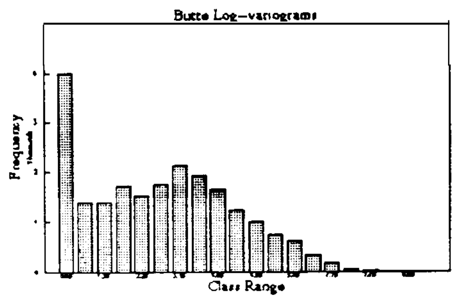

Bend Elevation
Frequency
Thousands
Class Range

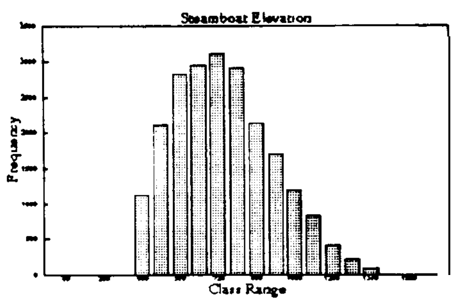

Butte Elevation
Frequency
Thousands
Class Range


Bend Variance of Elevation
Frequency
Thousands
Class Range


Bend Log-variograms
Frequency
Class Range


Butte Variance of Elevation
Frequency
Thousands
Class Range


Butte Log-variograms
Frequency
Thousands
Class Range


Steamboat Elevation
Frequency
Class Range