Mathematical Geology, Vol. 24, No. 4, 1992

Evaluation and Comparison of Spatial Interpolators[1]

Dennis Weber[2] and Evan Englund[3]

This study evaluates 15 different estimators to determine their relative merits in estimating block concentrations at contaminant waste sites. The evaluation was based on 54 subsets of data drawn from an exhaustive set of 19,800 data. For each subset, 198 block estimates were made with each estimator. The measurements of estimation quality were a linear loss function and a more standard statistic, the mean square error. The linear loss function showed that seven of the estimators produced scores close enough to be within the same statistical population. Results based on the mean square error were similar. The surprising results of this study were that inverse distance and inverse distance squared both produced better scores than kriging.

KEY WORDS: sampling, geostatistics, estimators, interpolation.

INTRODUCTION

Previous studies have examined the effects of various estimation parameters on the quality of estimates from spatially correlated data. Englund (1990) showed that the variance of estimates was high among 12 different statisticians who made local block estimates from two common sets of data. A second study by Englund et al. (in press) investigated different sampling design parameters, sample size, grid type, and noise level, and showed that the only statistically significant parameter was sample size. The present study evaluates the relative accuracy of 15 different spatial estimators.

EXPERIMENTAL PROCEDURE

Fifteen spatial estimation methods were used in this study. All methods produced estimates for the same 198 blocks, and had no missing values. The 198 block estimates for each of 54 sample sets and for each estimator were compared with the "true" values from the original 19,800 data. Evaluation

[1]Received 7 January 1991; accepted 22 September 1991.
[2]Environmental Research Center, University of Nevada, 4505 Maryland Parkway, Las Vegas, Nevada 89154.
[3]U.S. EPA EMSL-LV, P.O. Box 93478, Las Vegas, Nevada 89193.

statistics included a cost to society and the mean square error (MSE) averaged over the 54 sample sets.

WALKER LAKE DATABASE

A subset of the larger Walker Lake data set derived from a digital elevation model was used as a surrogate "site model." The local variance of elevation data was used as a surrogate for measured soil contamination data. A grid of 19,800 data in a 110 × 180 array (Fig. 1) was obtained by calculating the variance of 5 × 5 blocks of elevation data (Isaaks and Srivastava, 1989). The frequency distribution is in Figure 2. The site model was subdivided into 198 square blocks, each containing 100 variance values. These blocks, for which average "true values" were computed, represent units of a size assumed to be practical for remediation.

SAMPLING DESIGN

Fifty-four different sample data sets were drawn independently from the site model according to a 3 × 3 × 2 factorial design with three sample sizes, three sample patterns, and two levels of sample error: 18 different sample designs were produced, each combination of which was repeated three times. The sample sizes were 104, 198, and 308 data: the three sample patterns were simple

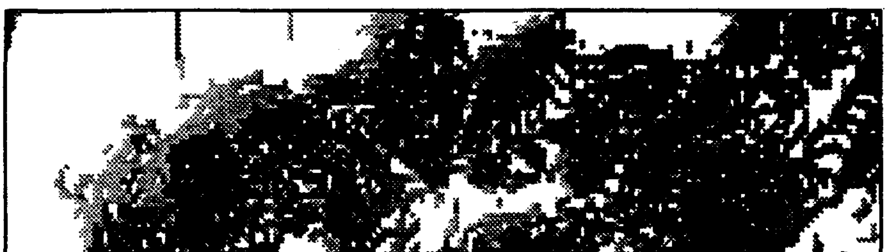

Fig. 1. Subset of the Walker Lake data set.

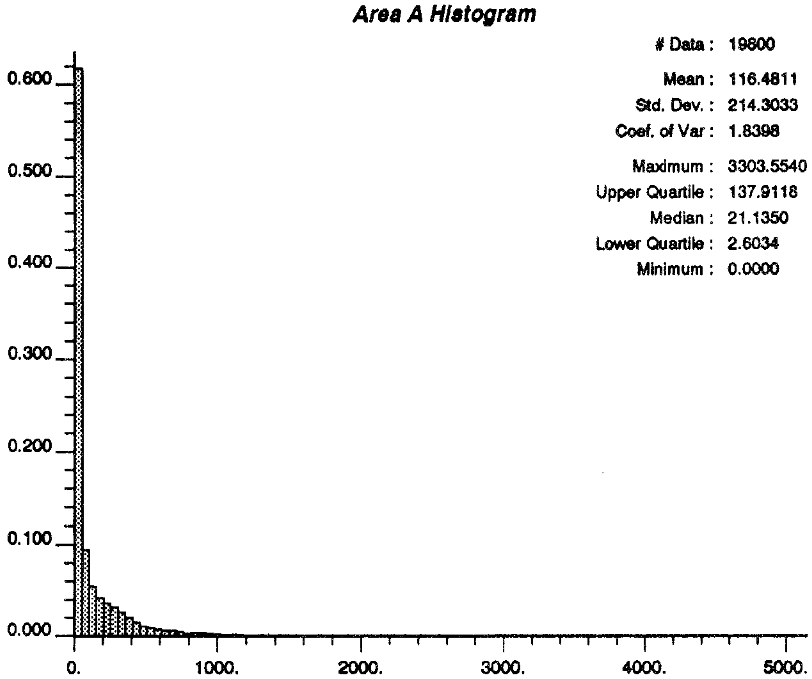

Fig. 2. Frequency distribution of 19,800 data from the subset of the Walker Lake database.

random, cellular stratified, and regular grid. The two levels of sample error were a base level of zero error, and a high level at a relative standard deviation of 32%. Errors were normally distributed, with zero mean.

EVALUATION STATISTICS

Linear Loss Score. The primary measure, the linear loss score (LLS), is calculated from an asymmetric linear loss function. The underlying assumption is that society pays a cost for all contaminated areas, either as a remediation cost for each block cleaned, or as a less easily defined group of costs (health effects, ecological damage, etc.) for each block which remains contaminated. We assumed their sum to be a linear function of concentration, while the unit remediation cost was assumed to be constant.

An "action level" for remediation is assumed to be society's best estimate of the breakeven point, i.e., the contamination level at which the cost of cleaning a block is exactly equal to the cost of not cleaning it. We define loss in units of "block remediation cost," which we normalize to "one" at the "action level.'' The loss assigned to a block can fall in one of the four categories shown in Table I.

For a block of any concentration, the cost associated with a correct remediation decision is found from lines 1 and 2; the loss of an incorrect decision is found from lines three and four. The sum of the 198 block scores is the total loss for the site, excluding sampling costs.

To minimize the effect of action level on linear loss score, we computed the scores (excluding sampling costs) at nine action levels. Thus, the linear loss score (LLS) derived from the linear loss function is expressed as

$$\text{Linear Loss Score} = \frac{1}{54}\sum_{i=1}^{54}\left[\frac{1}{9}\sum_{j=1}^{9}\left(\sum_{k=1}^{198}\text{Loss}_{ijk}\right)\right]$$

where the summations i, j, and k are over the 54 data sets, the nine action levels, and the 198 blocks, respectively. An example and additional detail regarding the calculation of the LLS are given in Englund et al., in press.

Mean Square Error. A second quality measure is the mean square error (MSE), averaged over all 198 blocks and all 54 sample sets, which is

$$\text{MSE} = \frac{1}{54}\sum_{j=1}^{54}\left[\frac{1}{198}\sum_{i=1}^{198}(Z_{ij}^{\text{estimate}} - Z_{i}^{\text{true}})^2\right]$$

where $\mathbf{Z}^{\text{estimate}}$ and $\mathbf{Z}^{\text{true}}$ are the estimates and true values for the blocks, and i and j represent the blocks and data sets, respectively. MSE does not depend on the action level.

ESTIMATORS

For each estimator described below, 198 block estimates were produced for each of the 54 data sets. Blocks were numerically approximated by discrete 2×2 arrays of point estimates. All kriging estimates were made by using Geo-EAS software (Englund and Sparks, 1988).

Table 1. Linear Loss Function[a]

	Decision	Estimate	True value	Assigned linear loss	True linear loss
1	Correct	>AL	>AL	1	1
2	Correct	<AL	<AL	TV/AL (<1)	TV/AL (<1)
3	Incorrect	>AL	<AL	1	TV/AL (<1)
4	Incorrect	<AL	>AL	TV/AL (>1)	1

[a] AL and TV represent action level and true value, respectively.

Ordinary Kriging and Simple Kriging. Variograms using spherical models for each of the 54 data sets were estimated by a single investigator who modeled the variograms visually according to a prescribed set of instructions. The kriging neighborhood was defined as the 20 closest samples. For each of the 54 data sets, the mean value of the data samples was provided to the simple kriging program.

Log Kriging. The natural logarithms of the data sets were used to calculate the variograms for the 54 data sets. They were calculated according to the same procedures used above in ordinary kriging. Then ordinary kriging was performed as above to obtain the estimates in log space. To recover block estimates in units of concentration, six methods of backtransform (Rendu, 1979; David, 1988; Rivoirard, 1990) were used:

$$Z_{A_{ij}} = e^{Y_{ij}}, \quad Z_{B_{ij}} = \exp\,[Y_{ij} + \tfrac{1}{2}V_{ij}], \quad Z_{C_{ij}} = \exp\,[Y_{ij} + \tfrac{1}{2}V_{ij} - \lambda_{ij}],$$

$$Z_{D_{ij}} = \frac{\mu_{dj}}{\mu_{D_j}}\, e^{Y_{ij}},$$

and

$$Z_{E_{ij}} = \frac{\mu_{dj}}{\mu_{E_j}} \exp\,[Y_{ij} + \tfrac{1}{2}V_{ij}], \quad Z_{F_{ij}} = \frac{\mu_{dj}}{\mu_{F_j}} \exp\,[Y_{ij} + \tfrac{1}{2}V_{ij} - \lambda_{ij}]$$

where $\mathbf{Z}_{\mathbf{K}ij}$ is the **K**-backtransform (**K** $= A, B, C, D, E,$ and F) for block i in data set j. Y_{ij} is the log-kriged estimate, V_{ij} is the Kriging variance, and λ_{ij} is the Lagrange multiplier. Backtransform bias was accounted for by μ_{dj} which is the mean value of the sample data values for set j, divided by μ_{Kj} which is the mean of the 198 **K**-backtransformed estimates for the same set.

Rank Kriging. For a data set of n measured values (Z_i), the (Z_i) were assigned ranks (R_i) according to their magnitudes (Singh and Sparks, in preparation). Rank values were treated as measured values in ordinary kriging. The backtransform from the kriged rank estimate was obtained by linear interpolation

$$Z_o = Z_j + \left[\frac{(R_o - R_j)}{(R_{j+1} - R_j)}\right] \cdot [Z_{j+1} - Z_j]$$

where R_j and R_{j+1} are the ranks below and above the estimated rank, respectively, and Z_j and Z_{j+1} are the measured values that correspond to R_j and R_{j+1}. A bias correction was made as in the log kriging backtransforms $(\mu_{dj}/\mu_{\text{rank}j})$, where j represents the jth data set.

Mean. For each of the 54 data sets, the mean value of the data samples was assigned to each of the 198 blocks.

Radian CPS/PC. The following four estimators use the CPS/PC software package (Radian, 1990). Available software options did not permit the same search scheme used for the kriging estimators. The four CPS/PC estimators all used a search scheme defined as a circle, divided into octants, with a search radius (r_s) of 800 feet. Estimates were made using a maximum of three samples per octant, with a minimum total of eight.

Inverse Distance and Inverse Distance Squared. These estimators are 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 (w^{ID}) and inverse distance squared (w^{IDS}) are defined as

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

where r_i is the distance between estimate and the ith sample location, and r_s is the search radius.

Piecewise Least Squares. A polynomial equation (a surface) is fitted to the selected samples

$$Z_o = a_{0,0} + a_{1,0} + a_{0,1}y + a_{1,1}xy$$

where Z_o is the estimated value at coordinates x and y, and $a_{i,j}$ are the fitting coefficients. Coefficients are obtained by regressing sample values against their x, y coordinates while constraining the fit to minimize the weighted residual sum-of-squares. The residual weights were w^{IDS} as calculated above.

Projected Slope. This procedure individually fits a first order polynomial function as described by Z_o above through each selected sample location. The distance weighting is calculated as w^{IDS} above where r_i are the distances between the selected sample location and the remaining sample locations. This procedure results in as many surfaces as sample locations. The estimate is calculated from the surface values at their intersections with the estimate location, weighted again as w^{IDS} above, but where the distances r_i are the distances between the estimate location and the sample locations.

RESULTS

Results for the 15 estimators are summarized in Fig. 3 and 4 where the means and standard error of the means are given for the linear loss score and mean square error, respectively. In Fig. 3, the estimators are grouped by LLS means.

Figures 3 and 4 show the surprising result that both inverse distance squared

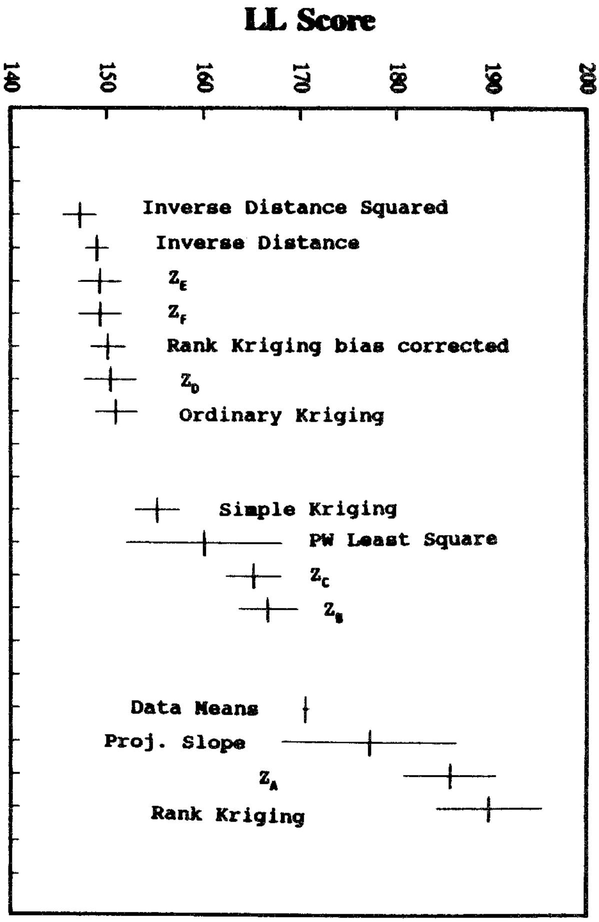

Fig. 3. LL Score means and standard errors for the 15 estimators. Horizontal and vertical bars represent means and ranges including plus and minus two standard errors, respectively. Log kriging back-transforms (Z_A, Z_B, etc.) are labeled according to their definitions given under *Log Kriging* in the section on Estimators.

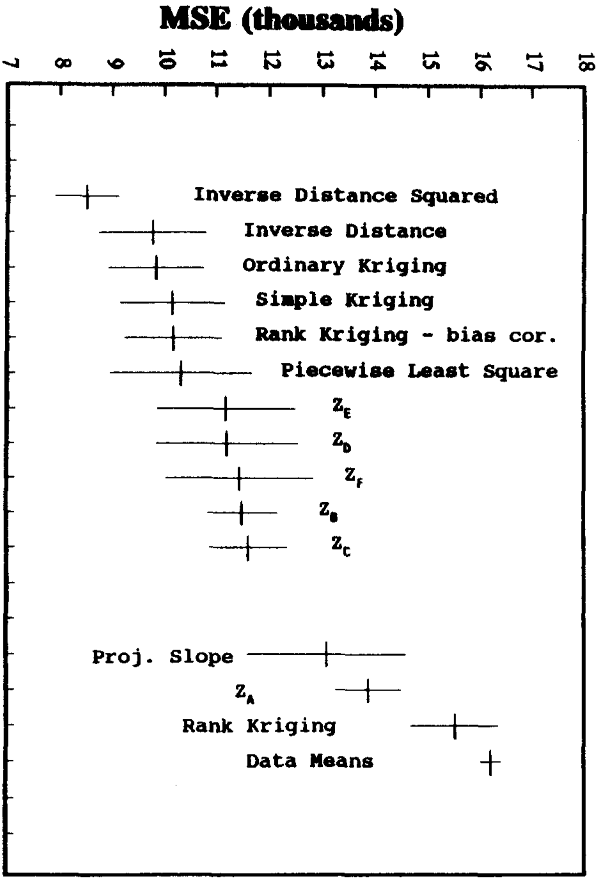

Fig. 4. MSE means and standard errors for the 15 estimators. Horizontal and vertical bars represent means and ranges including plus and minus two standard errors, respectively.

and inverse distance estimators scored better than any other methods according to both quality measures. The next best estimates were obtained from the log krigings and rank kriging, all with bias corrections. The best log kriging was obtained by using the back transform which included the log value and its kriging variance. All three log kriging backtransforms with bias correction outperformed ordinary kriging, although the difference is not statistically significant.

Rank kriging without bias correction did the poorest with respect to both LL score and MSE; however, with bias correction, the LL score was better than ordinary kriging and only slightly worse with respect to MSE.

Ordinary Kriging produced better estimates than simple kriging because of the non-stationarity of the data. The original data set had large areas where the values were low and large areas where the values were high.

The eight poorest performing estimators scored percentage increases in LLS from 20.39 to 47.07 over the true values. These increases are significantly greater than the seven better performing estimators. They will, therefore, not be considered as competing methods in further studies.

Table II shows the linear loss for 11 of the interpolators vs. the action level. For action levels below 19.7 contamination units, the linear losses for the Piecewise Least Squares and Projected Slope interpolators are significantly higher than the other 13 interpolators. This is caused by underestimating blocks by these two interpolators.

DISCUSSION

Based on the results of this experiment, should geostatisticians abandon the use of kriging and adopt the simpler inverse distance algorithms? In the authors' opinion, the answer is no, for two reasons. First, an argument can be made that the particular results in this experiment are fortuitous, and second, reasonable variations of this experiment can be imagined wherein kriging would be expected to have a distinct advantage over the particular inverse distance algorithms used here.

In theory, kriging is an optimal interpolator in the sense that it minimizes estimation variance when the variogram is known and the expected values of the mean and variance are constant over the area of interest. In practice, these

Table 2. Linear Loss vs. Action Level

	Action level								
Estimator	2.0	8.2	19.7	41.0	68.8	108	151	210	297
Inverse distance squared	206	192	177	164	152	137	120	100	77
Inverse distance	203	196	182	167	154	139	122	101	77
Log kriging − Z_E[a]	204	196	186	170	155	136	119	99	78
Log kriging − Z_F[a]	202	196	185	171	155	137	119	99	78
Rank kriging[a]	216	196	181	168	154	138	121	101	78
Log kriging − Z_D[a]	212	199	186	171	155	136	118	99	77
Ordinary kriging	201	196	184	171	159	143	125	102	77
Simple kriging	202	199	195	182	165	148	126	102	77
Piecewise least squares	294	211	184	167	152	137	120	99	77
Data means	198	198	198	198	198	202	154	110	78
Projected slope	389	238	199	175	156	138	121	100	78

[a] With bias correction.

conditions are never met; kriging has become a popular interpolator in large part because it is robust—it generally produces good sub-optimal estimates even when reality departs substantially from the ideal.

The example used in this experiment is not ideal. Kriging at best should provide near-optimal estimates, and we would expect that if enough alternatives were examined, some would eventually be found which provide better estimates. That this occurred among the small number of interpolators tested in this study is, as was stated earlier, surprising, but not totally unexpected. The inverse distance methods are linear-weighted-average estimators in the same general category as kriging, and as such, can be expected to also be quite robust.

From a theoretical standpoint, only the MSE score should be used to compare interpolators, as this is the estimation variance which kriging attempts to minimize. Of the 15 interpolators tested, only ordinary and simple kriging use ordinary variograms and thus actually attempt to minimize the MSE as computed here. The log krigings, for example, use log variograms and attempt to minimize the mean squared log-error, which was not computed in this study. Thus, whereas ordinary and simple kriging might be expected to have the two lowest MSE scores under ideal conditions, they actually ranked a quite respectable third and fourth.

The kriging approaches used in this study can be described as relatively simplistic. Kriging varieties such as universal kriging, indicator kriging, and disjunctive kriging might be better suited to the case at hand.

The inverse distance methods used here, like kriging, involve some judgment on the part of the user. The search scheme described earlier has a significant effect by limiting the number of samples used for each estimate. The choice of search radius also affects the weights assigned. A different set of search parameters for the ID and IDS might produce worse (or even better) results. In addition to selection of a search strategy, the kriging estimators use variogram models which must be estimated from the data. In this study, single spherical models were fitted manually by visual inspection according to a simple set of instructions. This was done to minimize the learning effect of fitting multiple variograms from different samples of the same population, but it may have negatively affected the results.

Two frequently-encountered circumstances which would tend to favor kriging over inverse distance estimators were not present in this study. These are strong anisotropy and biased clustering of samples. The portion of the Walker Lake data set used here is only weakly anisotropic, and significant data clustering could occur only by chance in the 18 data sets generated by simple random sampling.

In conclusion, the results of this study, while provocative, should not be interpreted to mean that the inverse distance methods tested here are superior to kriging estimators in all cases. The search for the ideal interpolator is far from over.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Tom Starks, Allen Sparks, Nancy Leviant, Robert Enwall, and Ashok Singh for their contributions to this work. Although the research described in this article has been supported by the United States Environmental Protection Agency through cooperative agreement CR814701 to the Environmental Research Center of the University of Nevada, Las Vegas, it has not been subjected to Agency review and, therefore, does not necessarily reflect the view of the Agency and no official endorsement should be inferred.

REFERENCES

David, M., 1988, Handbook of Applied Advanced Geostatistical Ore Reserve Estimation: Dev. Geomathemat., v. 6, p. 116–118.

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

Englund, E. J., and Sparks, A. R., 1988, Geo-EAS (Geostatistical Environmental Assessment Software) User's Guide, EPA/600/4-88/033: U.S. EPA, Las Vegas, 174 p.

Englund, E. J., et al., The Effects of Sampling Design Parameters on Block Selection: Math. Geol. In press.

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 p.

Radian Corporation, 1990, Commercial software available from Radian Corporation for use on mainframe and PC computers.

Rendu, J. M., 1979, Normal and Lognormal Estimation: Math. Geol., v. 11, n. 4, p. 407–422.

Rivoirard, J., 1990, Teacher's Aid: A review of Lognormal Estimators for *in situ* Reserves: Math. Geol., v. 22, n. 2, p. 213–221.

Singh, A., and Sparks, A., Rank Kriging: An Alternative Nonparametric Geostatistical Method. In preparation.