[A version of this manuscript has been published *Mathematical Geology 24:4 (1992) 381-391]*

EVALUATION AND COMPARISON OF SPATIAL INTERPOLATORS

Dennis Weber and Evan Englund

Abstract

This study evaluates fifteen 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 been made to determine the effects of various estimation parameters on the quality of estimates of spatially correlated data. A study by Englund (1990) showed that the variance of estimates was high among 12 different statisticians

who primarily used Kriging to obtain estimates on two common sets of data. That study, however, was preliminary because it used only two data sets for two different areas. A second study by Englund, et al., (1991) investigated the effects of different sampling design parameters on estimation accuracy. The second study used a 3x3x2 factorial design of 54 sample subsets drawn from an exhaustive data base of 19,800 original samples where the parameters were sample size, grid type, and noise level. That study showed that the only statistically significant parameter was the sample size where the estimation accuracy improved with increasing sample number. The present study evaluates the relative accuracy of 15 different spatial estimators by using the same 54 sample subsets.

EXPERIMENTAL PROCEDURE

Fifteen spatial estimation methods were used in this study. The only requirements were that all methods produce estimates for the same 198 blocks and have no missing values. The 198 block estimates for each of the 54 sample sets and for each estimator were processed in a program that compared them with the "true" values as calculated from the original 19,800 data. For each data set, the program produced evaluation statistics that included a cost to society and the mean square error (MSE) . These values, averaged over the 54 sample sets, were used to compare the relative merit of each estimator.

Walker Lake Data Base

For the evaluation, a surrogate "site model" data set was used. The site, a subset of a larger Walker Lake data set, was derived from digital elevation data. The variance of these elevation data was used as a surrogate for measured soil contamination data. The subset used in this study contains 19,800 data in a 11OX18O array (figure 1). The details have been described elsewhere (Englund, 1990, Isaacs, 1989). The site model has been subdivided into 198 square blocks, each containing 100 data values. The blocks, for which average "true Values" were computed, represent units of a size assumed to be practical for remediation.

Figure 1. Walker Lake Data Set

Sampling Design

The 3x3x2 factorial design, with 3 sample sizes, 3 sample patterns, and 2 levels of sample error produced 18 different sample designs, each combination of which was repeated 3 times for a total of 54 sample sets. This design was used by Englund (1990) to evaluate the effects of different sampling design parameters on the final estimates. The sample sizes were 104, 198, and 308 data. The 3 sample patterns were simple random, cellular stratified, and regular grid. Sample error represents the cumulative total of all possible error components included in the collection, handling, preparation, and analysis of a sample. The 2 levels of sample error were a base level of zero error, and a high level at a relative standard deviation of 32 percent. Errors were assumed to be normally distributed, with a mean equal to zero.

Evaluation Statistics

Linear Loss Score: The primary measure, the *linear loss score* (LLS), is calculated from a simple and economically based 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. In the absence of good models for the latter costs, we assumed their sum to be a linear function of concentration, while the remediation cost was assumed to be constant.

We make the additional assumption that when an "action level" for remediation is specified, it is 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 below.

Table I

Linear Loss Function

	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

AL and TV represent Action Level and True Value, respectively.

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. The optimal sampling design, of course, would be the one which minimizes total loss, including the sampling costs .

In order to minimize the effect of the choice of action level on the final linear loss score, we have computed the scores (excluding sampling costs) for each set of kriged estimates at nine

action levels. The action levels correspond to the decile class bounds on the true block values. In effect, the lowest action level treats the site model as if it were relatively highly contaminated; that is, 90% of the blocks are actually above the action level. Conversely, with the highest action level, only 10% of the blocks should be selected for remediation.

Finally, the *linear loss score* (LLS) derived from the linear loss function is expressed as

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

where the summations $i,j,$ and k are over the 54 data sets, the 9 action levels, and the 198 blocks, respectively. An example and additional detail regarding the calculation of the LLS is given in Englund, et. al., 1991.

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

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

where $\mathbf{Z}^{estimate}$ and $\mathbf{Z}^{true}$ are the estimates and true values for the blocks, and i and j represent the blocks and data sets, respectively. MSE is a purer statistic than the LLF because it does not depend on the action level. Furthermore, the correlation between the estimator and the two statistics might not show the

same correlation because the LLF assigns the same value to all blocks that are selected for removal, i.e., those for which the block estimate is greater than the action level.

ESTIMATORS

For each estimator, a regular grid of 198 block estimates at specified spatial locations was produced for each of the 54 data sets. In all cases, blocks were numerically approximated by a discrete 2x2 array of point estimates. The following are brief descriptions of the estimators used in the comparison study. All of the kriging estimates were made by using the Geo-EAS geostatistical software (Englund and Sparks, 1988).

Ordinary Kriging: All semivariograms were estimated by a single investigator who fit the models visually with the aid of Geo-EAS graphics according to a prescribed set of instructions. The investigator was instructed to fit a spherical model to the data. The kriging neighborhood was defined as the 20 closest samples.

Simple Kriging: Simple Kriging requires the data mean value to be provided by the investigator. For each of the 54 data sets, the mean value of the data samples was calculated and provided to the Geo-EAS program. The same search parameters were used as in the Ordinary Kriging above.

Log Kriging: The natural logarithms of the data sets were used to

calculate the semivariograms for the 54 data sets. They were calculated subjectively by one person according to the same procedures used above in Ordinary Kriging. Then Ordinary Kriging was performed also as above to obtain the estimates in Log space. To recover the block estimates in units of concentration, the following six methods of backtransform were used.

$$Z_{A_{ij}} = e^{LogY_{ij}}, \quad Z_{B_{ij}} = e^{\left[LogY_{ij}+\frac{1}{2}V_{ij}\right]}, \quad Z_{C_{ij}} = e^{\left[LogY_{ij}+\frac{1}{2}V_{ij}-\lambda_{ij}\right]},$$

$$Z_{D_{ij}} = \frac{\mu_{dj}}{\mu_{D_j}} e^{LogY_{ij}}, \quad Z_{E_{ij}} = \frac{\mu_{dj}}{\mu_{E_j}} e^{\left[LogY_{ij}+\frac{1}{2}V_{ij}\right]}, \quad \text{and } Z_{F_{ij}} = \frac{\mu_{dj}}{\mu_{F_j}} e^{\left[LogY_{ij}+\frac{1}{2}V_{ij}-\lambda_{ij}\right]}$$

where Z_{Kij} 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. For the second set of three backtransforms, the backtransform bias was accounted for by the factor 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: Here, the measured Z(s) values were assigned a rank according to their magnitudes. The rank values were then treated as the measured value in an ordinary kriging procedure, that is, a semivariogram was modeled, and the rank values were kriged. The backtransform from the kriged rank value at an unsampled location to the variable value was obtained by a linear interpolation between measured Z values as follows:

where Z_o is the variable estimate, s is the spatial coordinate, j

$$Z_o - Z(s_j) + \left[\frac{(R_o - R_j)}{(R_{j+1} - R_j)}\right] \cdot [Z(s_{j+1}) - Z(s_j)]$$

and j+l are the subscripts corresponding to their ranks below and above the estimated rank, and the subscript o corresponds to the estimate. A bias correction was made as in the log kriging backtransforms (μ_{dj}/μ_{rankj}), where here j represents the j^{th} 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 by Radian. Block estimation is approximated by using four point estimations, whereby the four points lie on the diagonals half-way between the midpoint and corners of each of the 198 blocks. All used the same search parameters which is defined as a circle, divided into 8 equal octants, with a fixed search radius (r_s) of 800 feet. If a sample location lies within .005 feet of the estimate location, the estimate (Z) was assigned the value of that sample; however, no such case existed. The search is defined as follows:

Check for sample within 0.005 feet. (none existed)

Yes: Assign Z_o equal to that sample value

No: Check for at least 3 samples in each octant.

Yes: Estimate Z_o by using the 24 samples

No: Check for at least 8 samples within neighborhood regardless of octant.

Yes: Estimate, but still using a maximum of 3 samples per octant

No: Assign Z_o as missing value.

The search parameters, however, were set to obtain estimates for all estimate locations; there was no missing value or sample value assigned to an estimate.

Inverse Distance and Inverse Distance Squared: Here the estimator 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 and inverse distance squared, respectively, are defined as

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

where r_i is the distance between estimate and the i^{th} sample location, and r_s is the search radius. This function will not become indeterminate because $r_i < .003$ feet is excluded.

Piecewise Least Squares: Here a polynomial equation (a surface) is fitted simultaneously to all of the selected samples within the search radius. This equation relates the value of the variable of interest to the spatial coordinates of locations within the sampling domain as follows:

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

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 weight of the i^{th} sample was w_i 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. Fitting is constrained by a least squares criterion applied to the weighted residuals of the *remainder* of the selected samples in the neighborhood. In

this step, the distance weighting is calculated as w_i above where r_i are the distances between the *selected sample location* and the *remaining sample locations*. This procedure results in the construction of as many surfaces as sample locations, each having radius of 800 feet. The estimate is calculated from the surface values at the intersections of the vertical projection of the estimate location with the surfaces that overlap that location. The values at the intersections are weighted again as w_i above, but, here, the distances r_i are the distances between the *estimate* location and the *sample* locations.

RESULTS

The LL score and MSE for the fifteen estimators are summarized in table II and figures 2 and 3. Figures 2 and 3 give the means and standard error of the means for the linear loss score and mean square error, respectively. Table II also gives the percent increase in LL score over the values obtained from using the true values; the standard errors are also given. The range of loss increase was from 14 to 47 percent, with seven of the estimators scoring between 14 and 20 percent. In figure 2, the estimators are grouped according to the their LLS means. The seven estimators having the lowest values represent one group, while the other two groups have four estimators each. Referring to figure 3, it is seen that the same correspondence does not apply for the MSE except for the group having the highest scores. This is to be expected

since the LL function assigned a value of "one" to any block having a concentration value above the action level. Therefore, a block estimate that counts very high in the MSE calculation would still be assigned only the value "one" for the LL score.

Table II shows the surprising result that both inverse distance squared 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.

The rank kriging procedure without bias correction did the poorest with respect to both LL score and MSE; however, by applying the 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. Simple kriging requires the mean value of the data set to be provided, whereas ordinary kriging calculates a mean for each individual block, based on the samples included in estimate. The local mean appears to be more meaningful in a situation where the global mean is not constant, i.e. , where the intrinsic hypothesis is not valid.

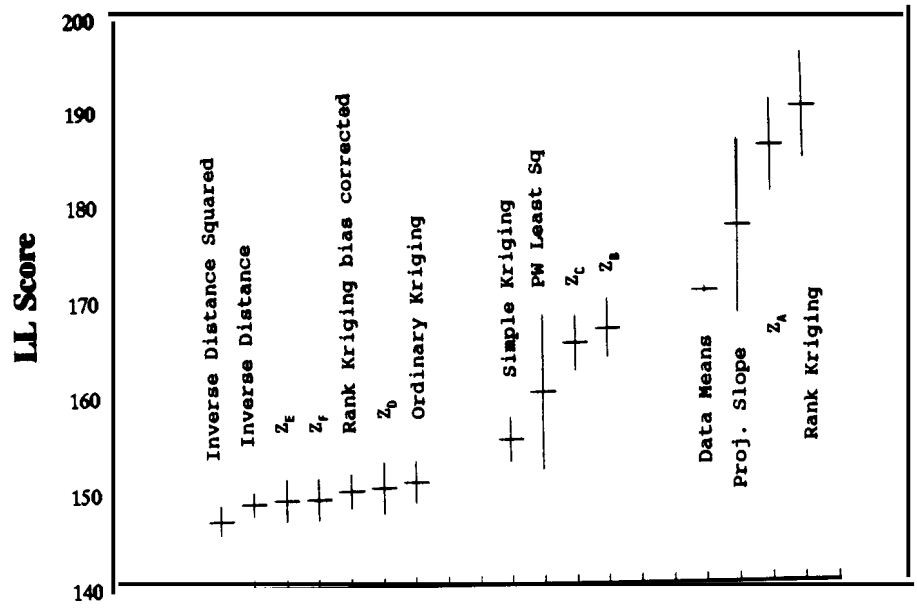

Figure 2. LL Score means and standard errors for the 15 estimators. The horizontal and vertical bars represent the means and ranges including plus and minus two standard errors, respectively. Log kriging backtransforms are labeled according to their definitions given under **Log Kriging**.

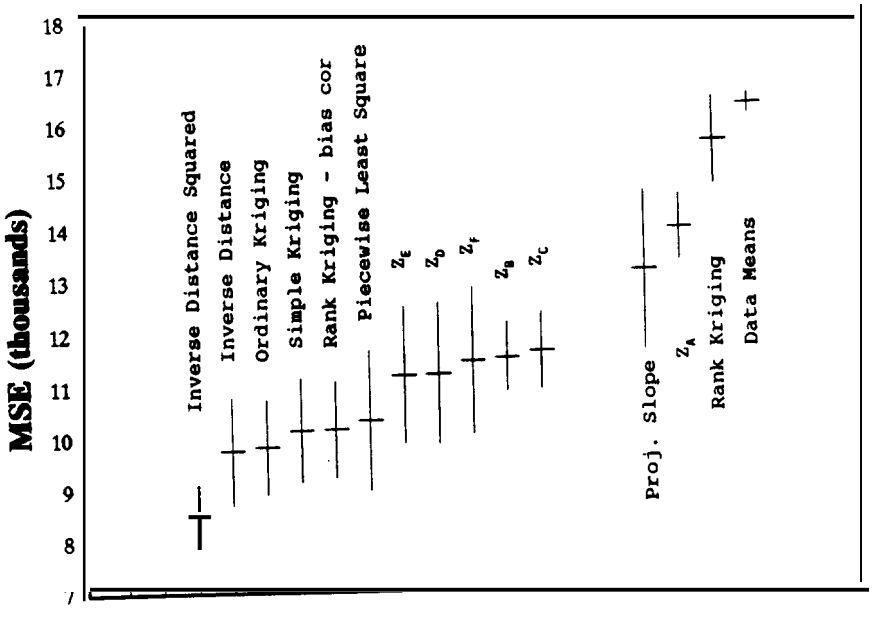

Figure 3. MSE means and standard errors for the 15 estimators. The horizontal and vertical bars represent the means and ranges including plus and minus two standard errors, respectively.

Table II

Summary of Evaluation Results

Estimator	LL Score/st.err.	%inc.	MSE/st.err.
True Values	128.98	0	0
Inverse Distance Squared	147.1/0.87	14.06	9763/513
Inverse Distance	148.9/0.62	15.44	8498/305
LK - Z_E (Y+½V & bias)	149.3/1.10	15.72	11167/654
LK - Z_F (Y+½V-λ & bias)	149.3/1.06	15.78	11425/696
Rank Kriging (bias cor.)	150.2/0.88	16.44	10151/459
LK - Z_D (Y & bias)	150.4/1.35	16.61	11190/675
Ordinary Kriging	151.0/1.07	17.06	9830/449
Simple Kriging	155.3/1.16	20.39	10142/497
Piecewise Least Squares	160.1/4.03	24.14	10312/671
LK - Z_C (Y+½V-λ)	165.2/1.43	28.07	11592/365
LK - Z_B (Y+½V)	166.7/1.52	29.23	11486/326
Data Means	170.5/0.16	32.15	16215/100
Projected Slope	177.1/4.49	37.31	13081/754
LK - Z_A (Y)	185.5/2.43	42.74	13875/308
Rank Kriging	189.7/2.82	47.07	15518/418

where LK is Log Kriging and the LK designations correspond to the definitions given under Log Kriging

The eight poorest performing estimators scored percentage increases 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.

Because of the closeness of the LL scores of the best few estimators, a study was made where the action levels were changed slightly to determine how the choice of action level affects the LL score. First, all nine action levels were changed by a constant percent of their original values, then only one level was changed to note the effect. The results for the Inverse Square Distance estimator are shown in table III. They show a linear relationship

between the percent change in action levels and the resulting percent change in linear loss score when all nine action levels are increased by the same percent. Where only one action level is changed, a percent change of action level normalized by dividing by 9 causes a similar percent change of LL score. Because the LLF is strongly correlated to the action levels, it is safe to assume that the differences seen between estimators is not due to the choice of action levels.

Table III
Inverse Square Distance Estimator

Change in all Action Levels		
% change	LL score	% change of LL score
0.0	147.1259	
0.1	147.0926	-0.023
0.2	147.0593	-0.045
0.5	146.9600	-0.112
1.0	146.8167	-0.210
5.0	145.5389	-1.079
10.0	143.9204	-2.179

Change in single Action Level		
% change	LL score	% change of LL score
.74	147.4370	0.21
.74	146.7852	-0.23

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to Allen Sparks and Ashok Singh for their contributions to this work.

********** NOTICE **********

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

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

Englund, Evan J., "A Performance Evaluation of Interpolation Methods" EPA Report EPA/600/X-89/049, March 1989.

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., Weber, D. D., and Leviant, N., 1991, The Effects of Sampling Design Parameters on Block Selection: Math. Geol., v. ??, p. ???-???.

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