Mathematical Geology, Vol. 24, No. 3, 1992

The Effects of Sampling Design Parameters on Block Selection[1,2]

E. Englund,[3] D. Weber,[4] and N. Leviant[5]

Cost-effective spatial sampling strategy requires balancing sampling costs with the expected benefits from improved information. A contaminated site numerical model was used to test various single-phase sampling schemes, which were evaluated based on the quality of block selections from interpolated values. Different sample set sizes, different sampling patterns, and two levels of sampling precision were used. The sample set size was the only one of these factors observed to be significant. Bias was also examined. Modest levels (<20%) had minimal impact; the effects of higher levels of bias varied with the selection level concentration.

KEY WORDS: sampling, geostatistics.

INTRODUCTION

The problem of designing a single-phase spatial soil sampling plan at a contaminated site is of considerable economic interest. The specific question addressed in this paper is that of block selection, that is, the identification of sub-areas of a site which require remedial action. Although the economic factors differ, the problem is similar to that of grade control in mining operations.

The problem of spatial sampling network design has been addressed by workers in many different fields. A brief overview of the topic is provided by Barnes (1989). The most common geostatistical approach, exemplified by Bur-

[1]Received December 12, 1990; accepted March 28, 1991.

[2]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.

[3]U.S. Environmental Protection Agency, P.O. Box 93478, Las Vegas, Nevada 89193.

[4]Environmental Research Center, University of Nevada, 4505 S. Maryland Parkway, Las Vegas, Nevada 89154.

[5]Computer Sciences Corporation, P.O. Box 93478, Las Vegas, Nevada 89193.

0882-8121/92/0400-0329$06.50/1

gess et al. (1981), is to look for the lowest cost design which satisfies a specified upper limit on the maximum (or mean) kriging variance.

In this paper, a strictly empirical approach was taken. The effects of three sampling design parameters were examined by using kriged estimates of sample sets obtained by repeatedly resampling a numerical site model: variogram models were inferred from the sample sets. The parameters were sample set size, sample pattern, and sample error. Here, sample set size simply refers to the number of individual samples to be collected in a given sample set. The model exhibits realistic characteristics such as high positive skewness, discontinuity, and a spatial correlation structure. The objective was to obtain information on the relative importance of the design parameters under realistic conditions, in order to prepare practical guidelines for cost-effective sampling programs.

THE SITE MODEL

To test the effects of different sampling parameters, a surrogate "site model" data set was used which is a subset of the larger Walker Lake data set (Isaaks and Srivastava, 1989). It was derived from a digital elevation data, with elevation variance used to simulate soil contamination. The subset of the Walker Lake data set used in this study contains 19,800 data in a 110 × 180 array (Fig. 1), and has been described in detail elsewhere (Englund, 1990).

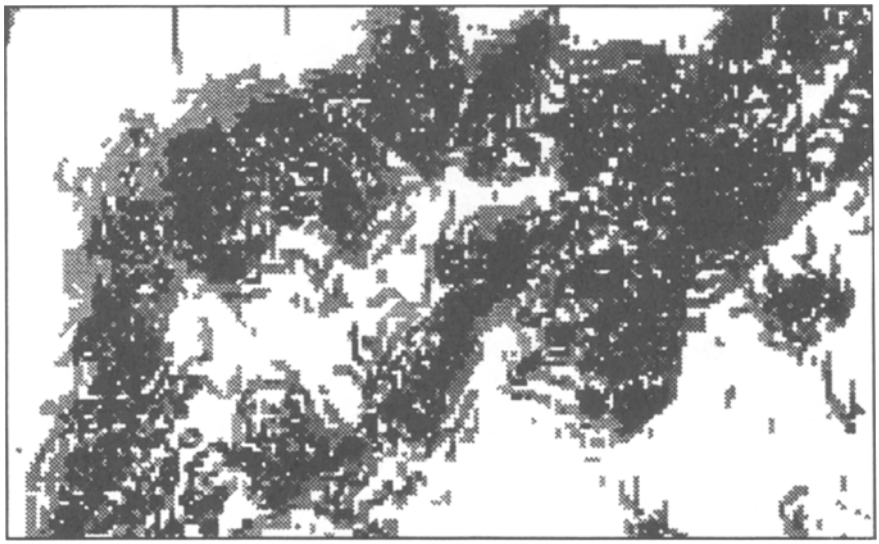

Fig. 1. Shaded map of the site model showing 19,800 points. Shading is based approximately on the quartiles of the data values. Darker shading represents higher values.

The site model was subdivided into 198 square blocks, each containing 100 data values (Fig. 2). The blocks represent units of a size assumed to be practical for remediation. A "true value" for each block was calculated by taking the average of the 100 data values within the respective block.

SAMPLING DESIGNS

The experimental approach taken to evaluate the different sampling design parameters is a 3 × 3 × 2 factorial design, with three sample set sizes, three sample patterns, and two levels of sample error. Combinations of these lead to

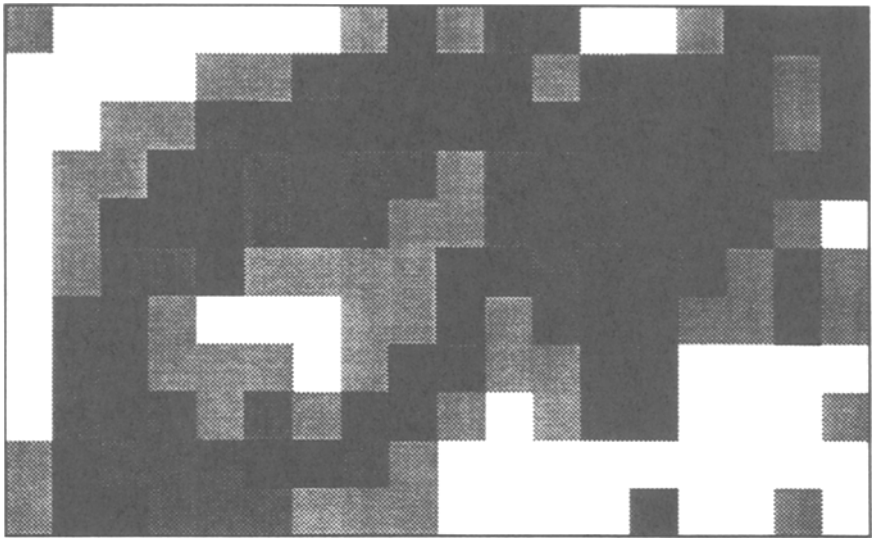

Fig. 2. Shaded map of site model, showing 198 true block means. Shading is based approximately on the quartiles of the block means.

Table I. 3 × 2 × 2 Factorial Sample Design Showing Number of Measurements per Sample Set[a]

	Error	No error
Random	104, 198, 308	104, 198, 308
Cellular stratified	104, 198, 308	104, 198, 308
Regular	104, 198, 308	104, 198, 308

[a] Each entry represents three samplings to obtain a total of 54 sample sets.

18 different sample set designs as shown in Table I, each of which was repeated three times for a total of 54 sample sets.

Previous work (Englund, 1990) with the site model suggested that sample set sizes of 100, 200, and 300 would be reasonable for this study; the actual sizes of 104, 198, and 308 reflect adjustments required to accommodate the regular grid pattern.

The three sample patterns used were random, cellular stratified, and regular grid (Fig. 3). Cellular stratified sampling involves selecting a randomly located sample within each grid cell.

The sample value assigned to any selected sample location was the value of the nearest of the 19,800 values plus a randomly generated error term when required. Sample error represents the cumulative total of all possible error components included in the collection, handling, preparation, and analysis of a sample. Two levels of sample error were considered—a base level at zero error, and a high level normally distributed error with a relative standard deviation of 32% of the true value.

Bias was not included as a factor in the experimental design. The effects of bias were evaluated later by multiplying the kriged estimates by a bias factor and recalculating the decision quality measures. The details are discussed later in this paper.

BLOCK ESTIMATES

Mean concentration values were estimated for each of the 198 blocks by the method of ordinary kriging with Geo-EAS software (Englund and Sparks, 1988). The kriging neighborhood was defined as the 20 closest samples. Variogram model functions required for kriging were estimated subjectively from

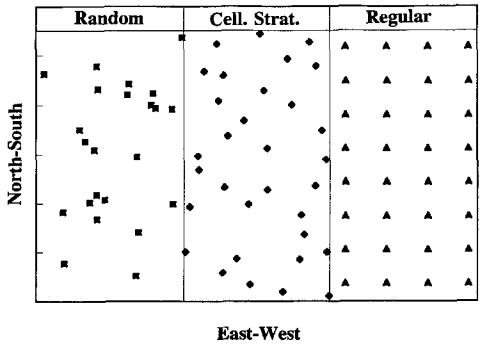

Fig. 3. Example of random (left), cellular stratified (center), and regular grid (right) sample patterns.

the sample set data; to minimize this as a source of variability in the study, all 54 models were estimated by one person according to a given set of instructions.

MEASURES OF QUALITY

Each interpolation of any one sample set produced 198 kriged block estimates which were compared to the corresponding true block values. To unambiguously compare one set of estimates with another, it was necessary to reduce the set of 198 block estimation errors to a single quality statistic. A variety of such measures were described by Englund (1990). They include statistical measures such as mean and standard deviation of the errors, decision quality measures such as the numbers of false positives and negatives, and loss functions which quantify the economic consequences of selection decisions. The most appropriate quality measure depends on the nature of the decision to be made. In this paper, two measures of quality, a linear loss score and the mean square error were used.

Linear Loss Score

In this study, the primary evaluation statistic is the linear loss score which is calculated from a linear loss function (Journel, 1984). A linear loss function was used because it is simple and economically based. 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, their sum was assumed to be directly proportional to concentration, while the remediation cost was assumed to be constant.

To balance these costs, an action level (a decision variable) for remediation was defined as society's best estimate of the breakeven point, i.e., where, on the average, the cost of cleaning a block was equal to the cost of not cleaning it. Loss was defined in units of "block remediation cost" and the linear loss function was normalized to the value "one" at the action level.

The linear loss function can be divided into four categories as shown in Table II. When a block's estimated concentration was greater than the action level, the evaluation scheme used in this paper assigned the loss the value 1.0; when it was less than the action level, the loss was assigned the value "(true block value)/(action level)" or TV/AC as shown in Table II. The latter represents the proportional part of the loss function curve. Note that the decision was made based on the estimated concentration, but the loss in the latter case was determined by the true concentration of the block. One can see from Table II that any incorrect decision will result in a greater cost to society than will the

Table II. Linear Loss Function[a]

Line	decision	Estimated value	True value	Assigned linear loss[a]	True linear loss
1	Correct	> AL	> AL	1	1
2	Correct	< AL	< AL	TV/AL (<1)	TV/AL (<1)1
3	Incorrect	> AL	< AL	1	TV/AL (<1)1
4	Incorrect	< AL	> AL	TV/AL (>1)	1

[a] AL and TV represent action level and true value, respectively.
[b] Assigned loss is based on the estimated block value and the action level.

correct decision. For a block of any concentration, the loss associated with a correct remediation decision is found from lines 1 and 2: the cost of an incorrect decision is found from lines 3 and 4. For a given action level and data set, the sum of the 198 block costs, excluding sampling costs, would be the total cost for the site. The optimal sampling design would be the one which minimizes total cost, including the sampling costs.

It should be noted that if the block estimate were considered to be the expected value of a conditional probability distribution, then the expected loss from non-remediation could be computed by integrating the loss function over the distribution. The optimal decision then, would be to remediate when this expected loss exceeded the cost of remediation. With the linear loss model, expected loss is a function only of the expected value of the conditional probability distribution, and not its shape; thus, the simple decision rule used here is optimal. This would not necessarily be the case for other loss models.

In order to minimize the effect of the choice of action level on the total loss score, we have computed the total cost (excluding sampling costs) for each set of 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. The final Linear Loss Score was obtained by averaging the total loss over the nine action levels, then further averaging over the 54 data sets as follows.

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

This Linear Loss Score was compared with the ideal case where the score was calculated by using the true block values.

To illustrate this evaluation, Fig. 4 presents a scatterplot of one set of

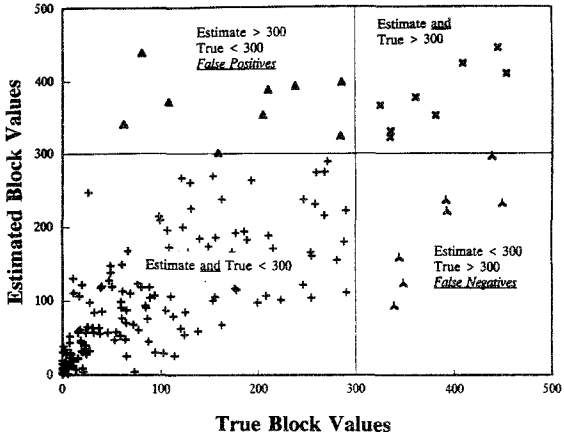

Fig. 4. Scatterplot of true v. estimated concentration values for data subset number 1.

estimates for data subset 1 in which the 198 true and estimated block values are plotted on the x and y axes, respectively, and the action level is 300 units.

Correct Decisions

The blocks falling in the upper right (Table II, line 1) and lower left (Table II, line 2) quadrants represent correct decisions, i.e., the decision (and hence, the cost) would be the same based on either the estimate or the true values. All blocks in the upper right quadrant receive scores of "1" and those in the lower left quadrant receive scores equal to their true values divided by 300 (< 1).

Incorrect Decisions

The upper left quadrant represents the false positives (Table II, line 3) where the estimates are greater than the action level, but the true values are less than the action level. These blocks receive scores of "1," which are greater than those obtained in the ideal case (< 1). The lower right quadrant represents the false negatives (Table II, line 4) where the estimates are less than the action level, but the true values are greater. These blocks receive scores equal to their true values divided by 300, but since they are greater than 300, their scores are greater than "1." Since the loss based on the true values is never greater than "1," these linear loss scores also will be greater than in the ideal case.

Therefore, for both false negatives and false positives, the losses are greater than those based on the true values. The desired objective for an estimator is to achieve a score equal to that obtained in the ideal case.

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.

RESULTS

Effects of Sample Set Size, Pattern, and Error

Figure 5 and Table III show the results of the factorial design study according to the linear loss score. Each of the three groups, i.e., sample set size, pattern, and error, contains all 54 results. Both presentations give the means and standard error of the means for each group. Figure 5 shows the means and the range identifying plus and minus two standard errors. The following observations were made.

The mean values of the Linear Loss Score and Mean Squared Error show that the sample set size is the most important of the sampling design factors. The decreases in Linear Loss Score as sample set size increases are significant compared to the standard error in all cases.

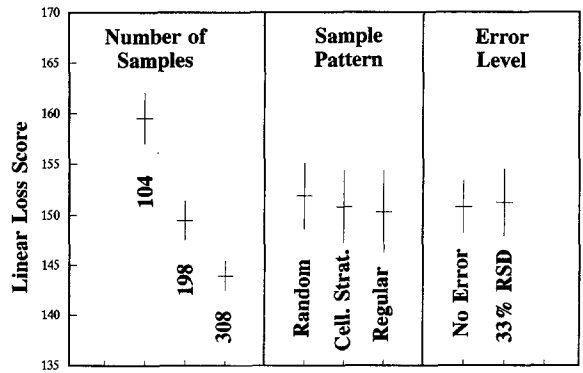

Fig. 5. Effects of sample set size, pattern, and error as measured by the linear loss score. The vertical and horizontal bars represent the means and ranges including plus and minus two standard errors, respectively.

Table III. Average Values of Variable Groups[a]

	Linear loss score		Mean squared error	
Design factors	Mean	SD (mean)	Mean	SD (mean)
Sample set size				
104	159.6	1.26	12,839	636
198	149.5	0.97	9,389	623
308	143.9	0.77	7,264	334
Pattern				
Random	151.9	1.63	10,661	799
Cell. strat.	150.8	1.81	9,718	748
Regular	150.3	2.02	9,113	719
Error				
No error	150.8	1.31	9,654	607
32% RSD	151.2	1.66	10,007	649

[a]SD (mean) is the standard error (standard deviation of the mean); RSD is relative standard deviation.

Sample pattern did not have a significant effect on the quality of selection decisions. Geostatistical theory (Olea, 1984; Yfantis et al., 1987) predicts that regular grids should provide lower variance estimates than random samples, and that the "randomized grid" used in the cellular stratified sampling should be intermediate. The results are not inconsistent with this theory, but support the view of Switzer (1979) that in the absence of clustering, estimation errors are insensitive to the data configuration.

Somewhat surprisingly, the results show no statistically significant difference between sample sets with no error and those with the error added. A possible explanation lies in the fact that even with the high relative errors, the variance of the distribution of absolute errors is less than 10% of the total population variance. This is consistent with common rules-of-thumb for good sampling. In addition, the variogram of the exhaustive site model (by using all 19,800 samples) indicates that approximately one-half the total population variance is already present at the scale of adjacent data points. This "spatial noise" is only increased about 20% by the additional sampling error. Furthermore, the error added here is strictly random and independent of the true values, which may be unrealistic.

It is also interesting, and perhaps somewhat sobering, to note that there is overlap in the results obtained with 104 and 308 samples. This results from variance unexplained by the sampling design parameters. The probable source is simply luck-of-the-draw in the sampling process. This illustrates the point

that using an optimal sampling design will not guarantee the best (or even a good) result in any specific case.

Figure 6 illustrates that observations similar to those made from Fig. 5 can also be made when quality is measured by the more traditional mean square errors.

Figures 7–9 provide an alternate view of the results. Here the mean loss for each factor was plotted against the decile action levels. For reference, the losses obtained by selecting none of the blocks (no action) and by perfect selection were also plotted. Note that for action levels near 300, block selection does not appear to have an advantage over the all-or-nothing approach, and in some cases, may be worse. These results are summarized in Table IV. They show that most of the variation occurs in the mid-range action levels, and that the statistically significant differences reflected in the LL Score are also shown to be consistent for the individual levels.

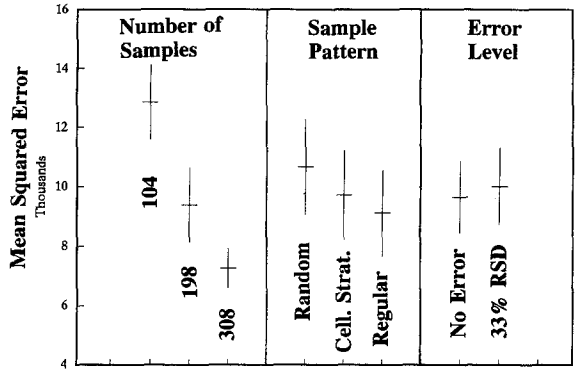

Fig. 6. Effects of sample set size, pattern, and error as measured by the mean squared error. The vertical and horizontal bars represent the means and ranges including plus and minus two standard errors, respectively.

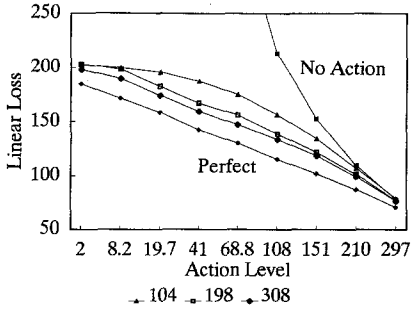

Fig. 7. Mean loss for three sample set sizes, plotted vs. action level.

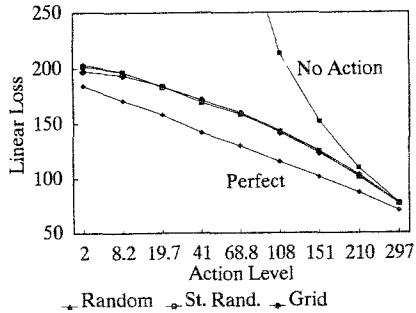

Fig. 8. Mean loss for three sample patterns, plotted vs. action level.

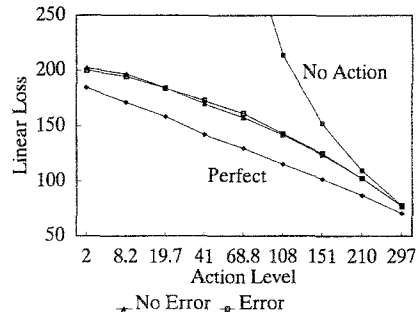

Fig. 9. Linear loss for two error levels plotted vs. action level.

Table IV. Linear Loss vs. Action Level for the Different Categories

	Action level								
Category	2	8.2	19.7	41	68.8	108	151	210	297
104	202	200	195	187	175	156	134	107	79
198	203	198	182	167	156	139	122	101	77
308	198	189	174	159	147	133	119	99	76
Random	204	197	184	172	160	144	126	104	78
Cel. strat.	202	196	184	170	159	143	125	101	77
Regular grid	197	194	184	172	160	142	123	102	77
No error	202	197	184	170	158	142	124	102	77
Added error	200	195	184	173	161	143	125	103	78

In Figure 7, the sample set size curves show that the incremental loss reduction due to increased sampling was significantly greater for action levels near the median value of 69. This indicates that the quality of the estimate is less important at the two extremes of action level. The main reason is that the

fraction of blocks impacted by poor estimates was small at the extremes. Therefore, assuming no bias, the effects of poorer estimates would be smaller. For very low action levels, most blocks were selected for remediation; therefore, the loss approached the value 198. However, false negatives could be costly because the assigned loss was "(true value)/(action level)" where action level was a small number. For high action levels, the loss assigned to false negatives is usually small because the action level is a large number. Therefore, assuming no bias, the effect of poorer estimates will again be minimized.

Effects of Sampling Bias

If one were to multiply a variable in a data set by a constant k, and then compute variograms and kriged estimates from the modified variable, all of the kriged estimates would be multiplied by k. One can, therefore, evaluate the effect of a constant multiplicative bias by multiplying the kriged estimates by the constant and recomputing the quality measures. Here a computationally simpler equivalent was used: biasing the selection level relative to the nominal action level. For example, given an action level of 100, selecting all blocks greater than 90.91 gives the same loss function score as multiplying all of the kriged estimates by 1.1 (+10% bias).

Figure 10 shows linear loss as a function of bias expressed in percent. Each point is the mean loss for all 54 cases averaged over the nine decile selection levels, where each selection level was multiplied by the bias factor. Note that the minimum of this curve occurs at zero bias, and that it is relatively flat near the minimum.

The bias relationship is much more complex when one examines the curves for individual action levels, as illustrated in Fig. 11. The average curve is only representative of the mid-range action level curves. Action levels near the tails become highly asymmetrical; at the extremes, the minimum loss may occur at significant levels of bias. The reasons for this effect can be seen by examining

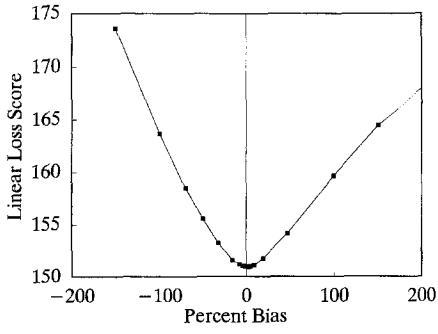

Fig. 10. Effects of sample bias on block selection quality; linear loss score (averaged over all action levels).

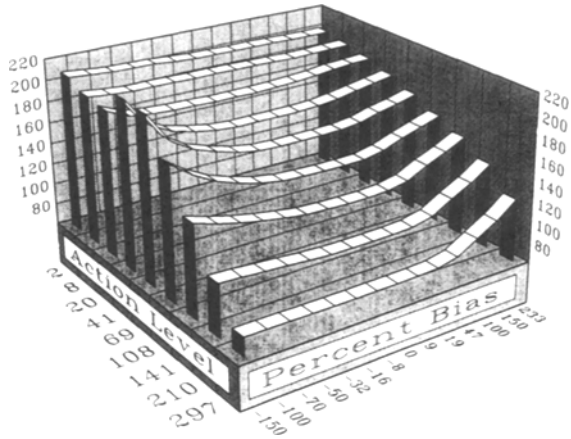

Fig. 11. Effects of sample bias on block selection quality; linear loss score for each action level. Vertical scale is the linear loss score.

Fig. 4. For the relatively high action level shown, there are 17 positives, nine of which are false. The total remediation cost for these blocks equals 17. One can see by inspection, however, that the mean true concentration of these blocks is less than the action level. Thus, if the estimates were sufficiently negatively biased such that they were not remediated, their mean loss would be less than one and their total loss would be less than 17. A negative bias, therefore, would reduce total loss. At low action levels, a comparable effect occurs for positive bias.

DISCUSSION

These results should be interpreted with caution, as they can be generalized to only one class of sampling problem, namely highly skewed (approximately log-normal) populations with well-defined spatial correlation and a high degree of random variability over short distances. The model represents only sites which have been almost entirely contaminated to some degree, as opposed to sites which have discrete, localized "hot spots" surrounded by clean areas. Such sites could not be modeled with Gaussian-related distributions, but would call for mixtures of distributions. Nevertheless, there are practical implications for sampling and decision-making in this type of situation.

The relative insensitivity to moderate amounts of linear multiplicative bias and sampling error supports the use of field screening and portable analytical methods, if they are significantly less expensive than conventional sample col-

lection and laboratory analysis. In addition, the relatively broad zone of acceptable data quality provides considerable flexibility in combining data from different sources of varying quality.

It is current practice at some sites to compute confidence limits around block concentration estimates, and to select for remediation all blocks whose upper 95% bound exceeds the action level. This is equivalent to positively biased sampling, and is not optimal except when the action level is near the low tail of the distribution. Near the high tail, however, this bias would increase the losses and, thus, would be counterproductive.

The potential benefit from sampling and block selection, as opposed to making an all-or-nothing decision about the entire site, is greatest when the action level is near the median of the distribution. As the action level approaches either end of the distribution, the benefit approaches zero.

The relatively small effect of sample pattern on the results suggests that for practical purposes, in the absence of additional information, the particular sample pattern selected should be a matter of convenience. Usually, it is easier to sample on a regular grid; fortunately, this provides results at least as good as the other patterns. It should be emphasized that the sampling schemes evaluated here are all single-phase designs. The results are not applicable to multi-phase, adaptive designs.

In a previous study (Englund, 1990), a single data set of 126 samples drawn from the site model was interpolated by 12 different investigators, ten of whom used some form of kriging. Linear loss scores for the ten, computed as in the current study, showed a 12-point range, from 144 to 156. This is the same order of magnitude as the difference between the means of the 104-sample and 308-sample cases, suggesting that optimization of sampling and optimization of interpolation are economic problems of comparable importance.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Tom Starks, Allen Sparks, Robert Enwall, and Ashok Singh for their contributions to this work.

REFERENCES

Barnes, R. J., 1989, A Partial History of Spatial Sampling Design: Geostatistics (Newsletter of the North American Council on Geostatistics), v. 3, n. 2, p. 10–13.

Burgess, T. M., Webster, R., and McBratney, A. B. 1981, Optimal Interpolation and Isarithmic Mapping of Soil Properties: IV. Sampling Strategy: J. Soil Sci., v. 32, n. 4.

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.

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

Journel, A. G., 1984, in G. Verly et al. (Eds.), Geostatistics for Natural Resources Characterization, Part I: D. Reidel Publishing Company, p. 261–270.

Olea, R. A., 1984, Systematic Sampling of Spatial Functions: Kansas Geological Survey, Lawrence, 50 p.

Switzer, P., 1979, Statistical Considerations in Network Design: Water Res. Res., v. 15, n. 6.

Yfantis, E. A., Flatman, G. T., and Behar, J. V., 1987, Efficiency of Kriging Estimates for Square, Triangular, and Hexagonal Grids: Math. Geol., v. 19, p. 183–205.