[A version of this manuscript has been published in MATHEMATICAL GEOLOGY 24:3 (1992) 329-343]

TEE EFFECTS OF SAMPLING DESIGN PARAMETERS ON BLOCK SELECTION

E. Englund, D. Weber, N. Leviant

Cost-effective spatial sampling strategy requires balancing sampling costs with the expected benefits from improved information. A contaminated site model was used to test various sampling schemes, which were evaluated based on the quality of block selections from interpolated values. Different numbers of samples, different sampling patterns, and two levels of sampling error were used. The number of samples was the only one of these factors observed to be significant. Modest levels of bias (<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 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 analogous to that of grade control in mining operations.

The effects of three important parameters in sampling design have been investigated through a factorial experiment on a computerized site model. The parameters are sample size, sample pattern, and sample error. The model exhibits realistic characteristics such as high positive skewness, discontinuity, and a spatial correlation structure. The objective is 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, we used a surrogate "site model" data set. We chose a subset of the larger Walker Lake data set (Isaaks and Srivastava, 1989) which 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 110x180 array (Figure 1), and has been described in detail elsewhere (Englund, 1990).

The site model was subdivided into 198 square blocks, each containing 100 data values (Figure 2). The blocks, for which average true values were computed, represent units of a size assumed to be practical for remediation.

SAMPLING DESIGNS

The experimental approach taken to evaluate the different sampling design parameters is a 3x3x2 factorial design, with three sample sizes, three sample patterns, and two levels of sample error. Combinations of these lead to 18 different sample designs, each of which was repeated three times for a total of 54 samplings.

Sample size simply refers to the number of samples to be collected in a given sampling. Previous work with the site model suggested that sample 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 simple random, cellular stratified, and regular grid (Figure 3). Cellular stratified sampling involves selecting a randomly located sample within each grid cell.

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 at a relative standard deviation (RSD) of 32 percent. RSD is given by

$$error(x;\mu,\sigma^2) = \frac{\sigma}{\sqrt{2\pi}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}$$

where μ is the mean of the sample set, σ is 0.32μ, and x is the number from minus to plus infinity randomly selected to assign the error. Errors, therefore, were assumed to be normally distributed, with a mean equal to zero.

Bias was not included as a part of the factorial design. The effects of constant bias were evaluated later by adding the bias to the kriged estimates and recalculating the decision quality measures.

The sample value assigned to any selected sample location was the value of the nearest of the 19,800 values plus the randomly generated error term when required.

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

MEASURES OF QUALITY

Each kriging estimation from a sampling produced 198 block estimates which were compared to the corresponding true block values. To 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 (1989, 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. 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, we assumed their sum to be a linear function of concentration, while the remediation cost was assumed to be constant.

To balance these costs, we define an *action* level (a decision variable) for remediation as society's best estimate of the breakeven point, i.e., where 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" and normalize the linear loss function to the value "one" at the action level. The linear loss function is divided into four categories as shown in Table 1. When a block's *estimated* concentration is greater than the action level, it is assigned the loss 1.0; when less than the action level, it is assigned the loss "true value/action level". Note that the decision is made based on the *estimated* concentration, but the loss in the latter case is determined by the *true* concentration of the block. One can see from Table 1 that any incorrect decision will result in a greater cost to society than the true cost.

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 three and four. 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.

Table 1.
Linear Loss Function

Line	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 (<l)
3	Incorrect	> AL	< AL	1	TV/AL (<l)
4	Incorrect	< AL	> AL	TV/AL (>l)	1

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

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 (LLS) was obtained by averaging total loss over the nine action levels, then further averaging over the 54 data sets as follows.

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

This LLS was compared with the *ideal case* where the score was calculated by using the true block values.

To illustrate this evaluation, figure 4 presents a scatter plot of one set of estimates for data subset #l 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 1, line 1) and lower left (Table 1, 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 (<l).

Incorrect Decisions: The upper left quadrant represents the *false* positives (Table 1, 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 "l", which are greater than those obtained in the *ideal case* (<l).

The lower right quadrant represents the *false* negatives (Table 1, 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 "l". Since the loss based on the true values is never greater than "l", 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: The second quality measure is the mean squared error (MSE), averaged over all 198 blocks and all 54 sample sets, which is

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

where Z_{ij} and T_i 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 LLS because it does not depend on the action level (the decision variable). Furthermore, the correlation between the estimator and the two statistics might not be the same because the linear loss function 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.

RESULTS

Effects of Sample Size, Pattern, and Error

Figure 5 and table II show the results of the factorial design study according to the linear loss score. Each of the three groups, i.e., sample 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 representing plus and minus two standard errors. We make the following interesting observations:

By examining the mean values of the LLS and MSE, one sees that the number of samples is clearly the most important of the sampling design factors. The decreases in LLS as sample size increases are significant compared to the standard error in all cases.

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 better estimates than random samples, and that the "randomized grid" used in the cellular stratified sampling should be intermediate. The observed results are consistent with this theory, but the decreases in the means from random to regular grids are not *statistically* significant.

Table II
Average Values of Variable Groups

DESIGN FACTORS	LINEAR LOSS SCORE Mean	sd(Mean)	MEAN SQUARE ERROR Mean	sd(Mean)
Sample 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
33% RSD	151.2	1.66	10,007	649

sd(Mean) is the standard error (standard deviation of the mean)

Somewhat surprisingly, the results show no statistically significant difference between samples with no error and those with an error of 32% RSD. 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.

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 Figure 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 we see the mean loss for each factor plotted against the decile action levels. For reference, we also plot the losses obtained by selecting all of the blocks, none of the blocks, and for perfect selection. Note that for action levels near the tails of the distribution, block selection does not appear to have an advantage over the all-or-nothing approach, and in some cases, may be worse. In Figure 7, the sample number curves show that the incremental loss reduction due to increased sampling is significantly greater for action levels near the median.

Effects of Sampling Bias

If we 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. We can, therefore, evaluate the effect of a constant multiplicative bias by multiplying the kriged estimates by the constant and recomputing the quality measures. We used a computationally simpler equivalent: 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 (+lO% 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 9 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 we examine the curves for individual action levels, as illustrated in Figure 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.

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. Nevertheless, there are significant practical implications for sampling and decision-making in this type of situation.

The relative insensitivity to moderate amounts of bias and sampling error strongly supports the use of field screening and portable analytical methods, if they are significantly less expensive than conventional sample collection 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. At the other extreme, this bias would be strongly 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 mean 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, 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.

In a previous study (Englund, 1990) a single data set of 126 samples drawn from the site model was interpolated by 12 different investigators, 10 of whom used some form of kriging. Linear loss scores for the 10, 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.

REFERENCES

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

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

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.

Journel, A. G., 1989, Fundamentals of Geostatistics in Five Lessons, Short Course in Geology, v. 8: American Geophysical Union, Washington DC, 40 p.

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

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.

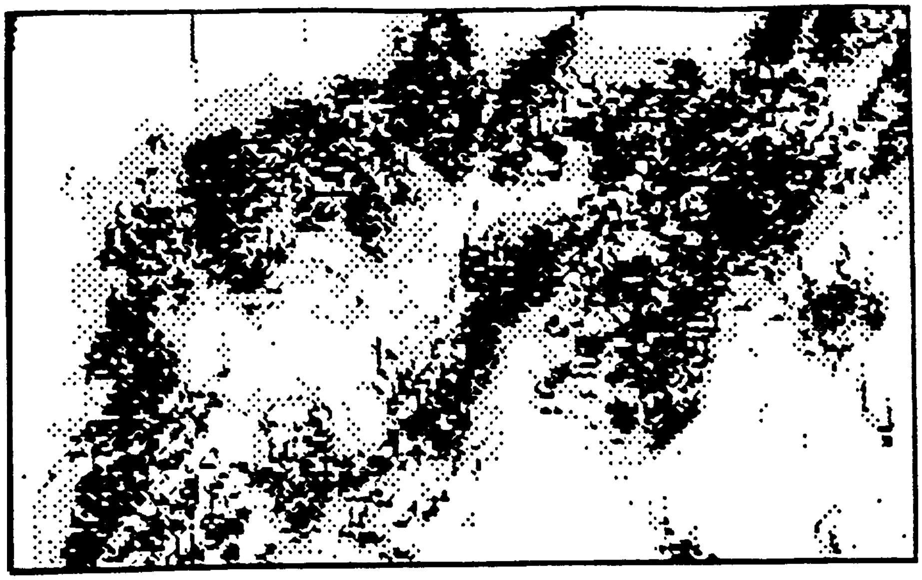

Figure 1. Shaded map of the site model showing 19,800 points.

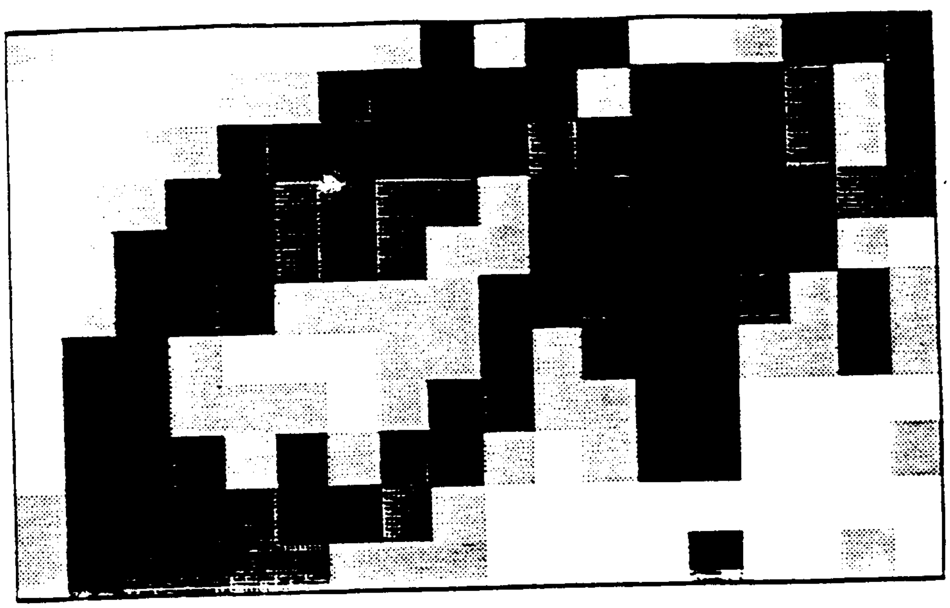

Figure 2. Shaded map of site model, showing 198 true block means.

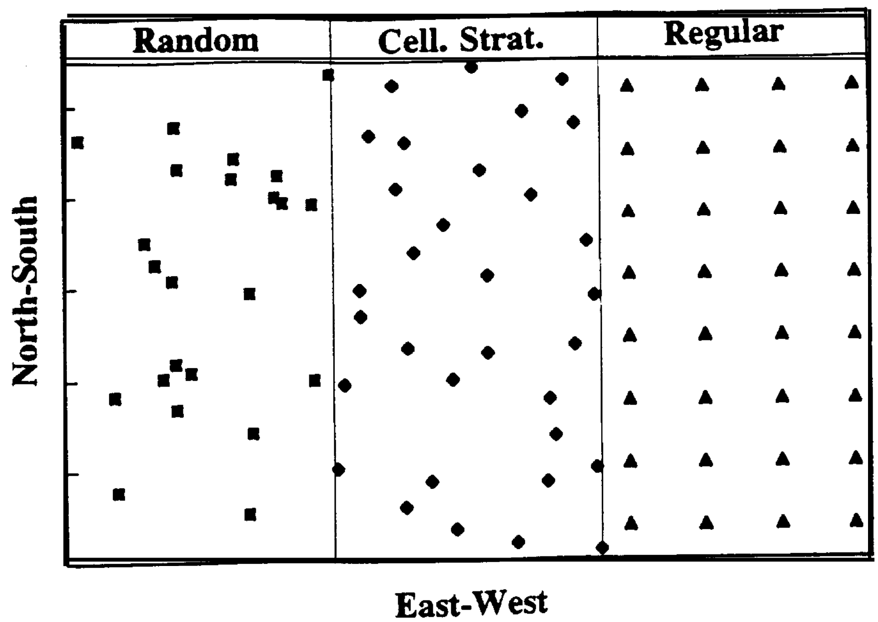

Figure 3. Example of random (left), cellular stratified (center), and regular grid (right) sample patterns.

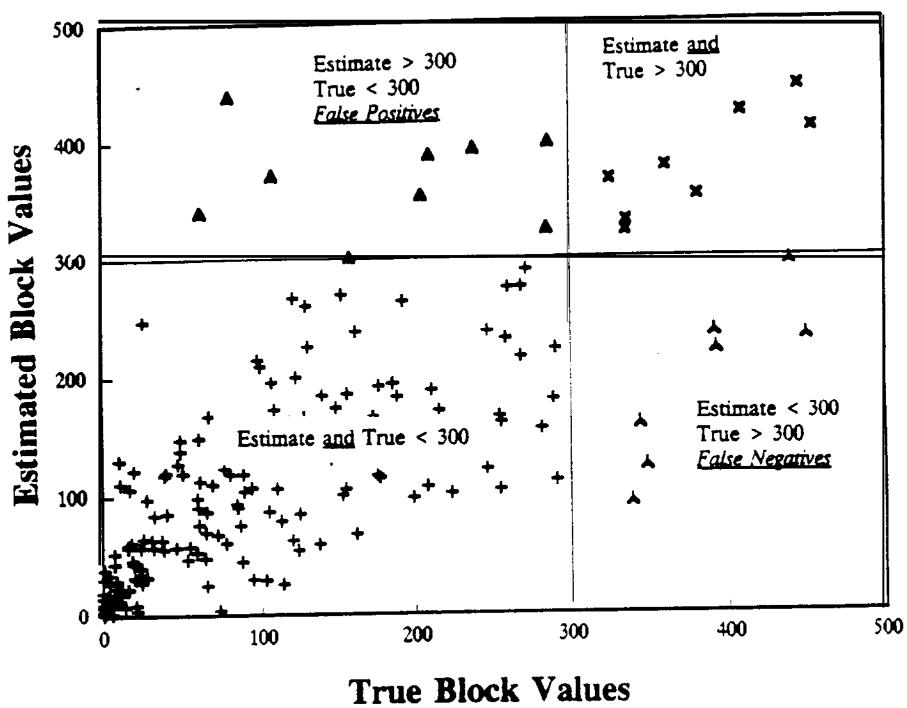

Figure 4. Scatter plot of true versus estimated concentration values for data subset number 1.

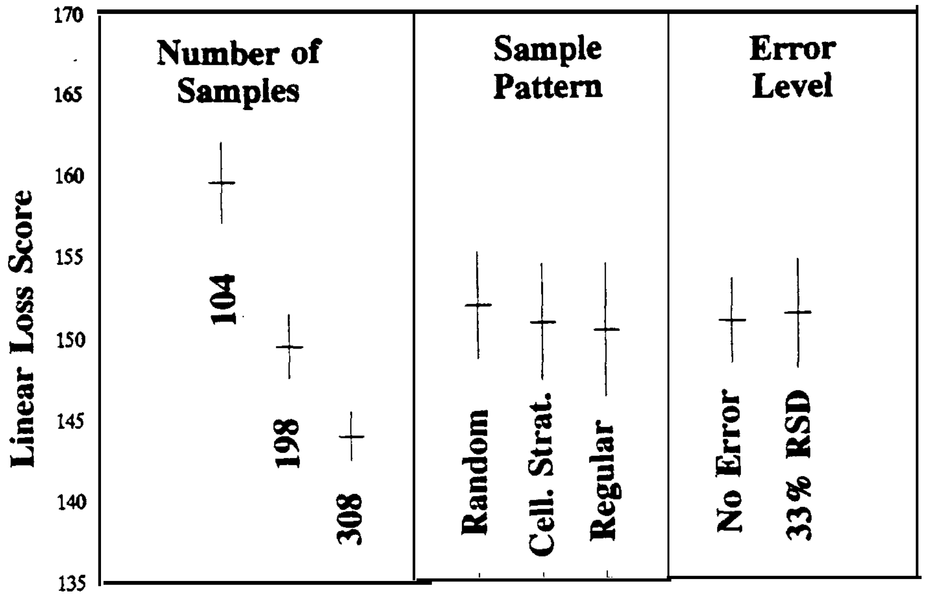

Figure 5. Effects of sample size, pattern, and error as measured by the Linear Loss Score. The horizontal and vertical bars represent the means and ranges including plus and minus two standard errors, respectively.

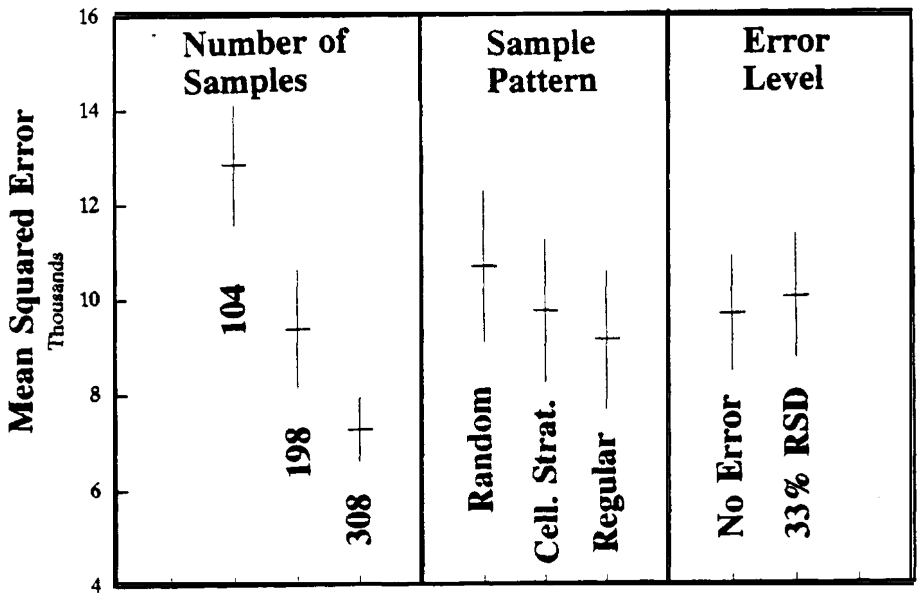

Figure 6. Effects of sample size, pattern, and error as measured by the Means Square Error. The horizontal and vertical bars represent the means and ranges including plus and minus two standard errors, respectively.

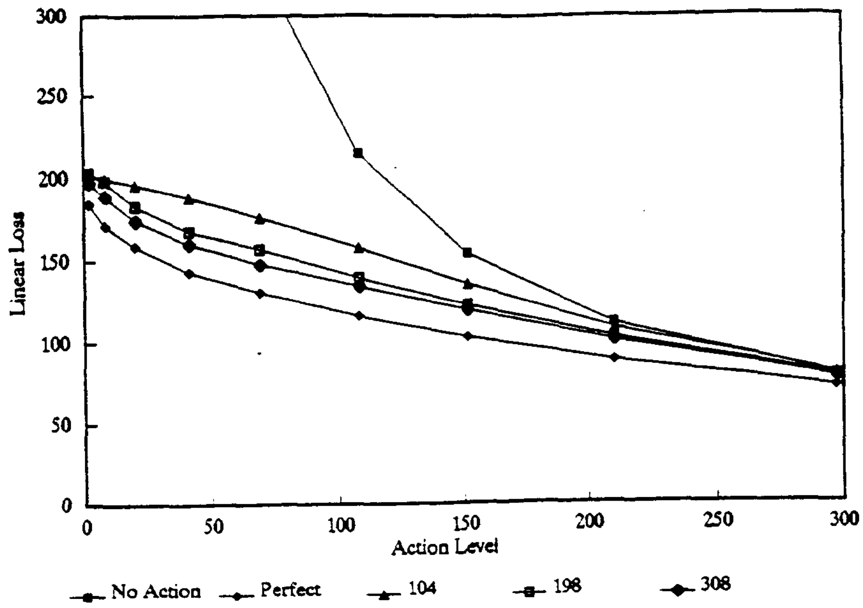

Figure 7. Mean loss for three sample sizes, plotted vs. action level.

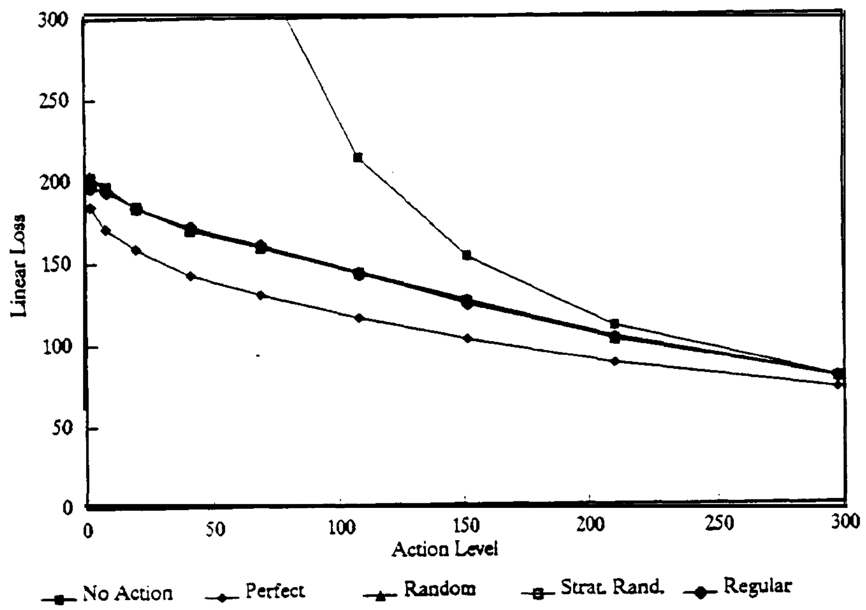

Figure 8. Mean loss for three sample patterns, plotted vs. action level.

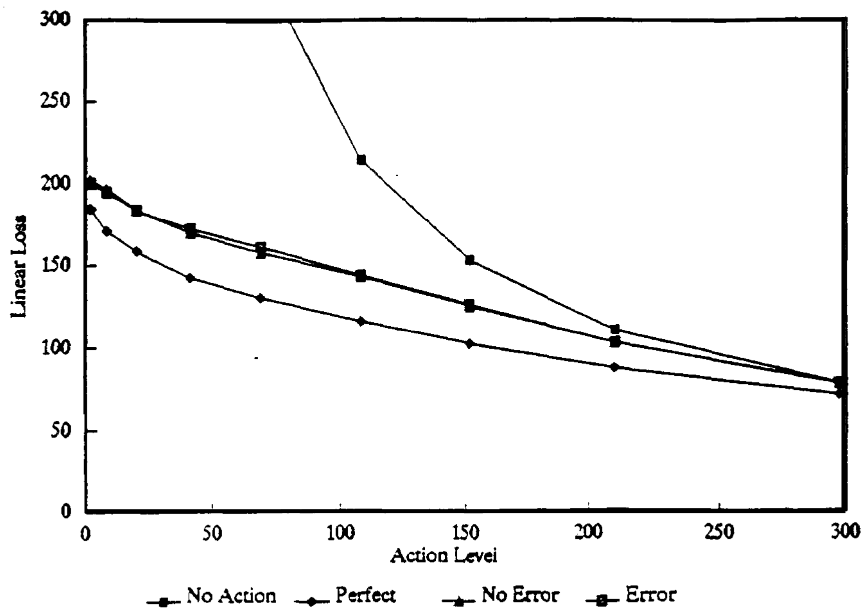

Figure 9. Linear loss for two error levels plotted vs. action level.

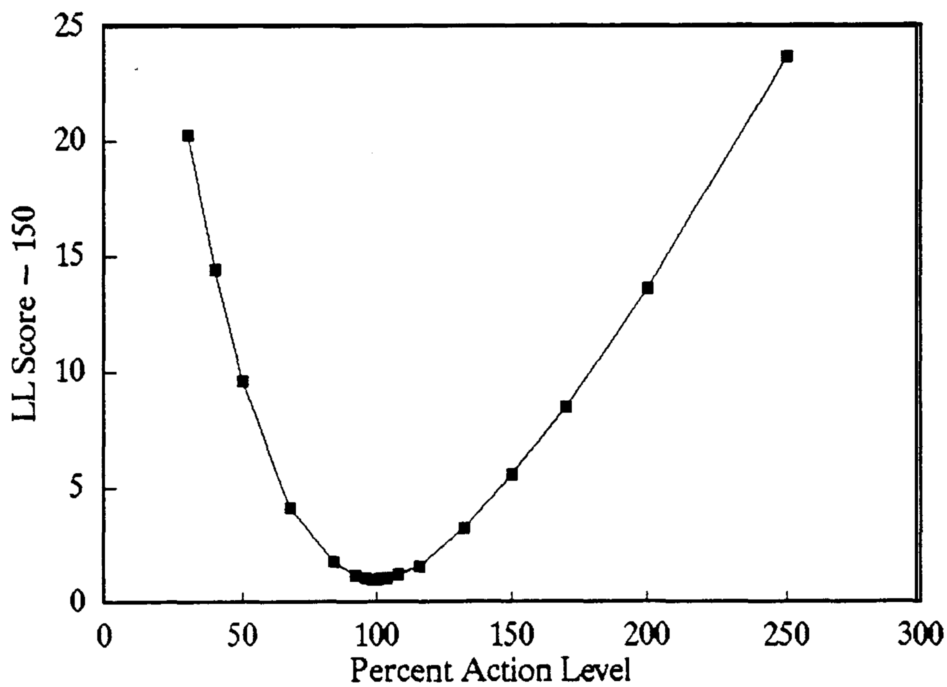

Figure 10. Effects of sample bias on block selection quality; linear loss score (averaged over all action levels).

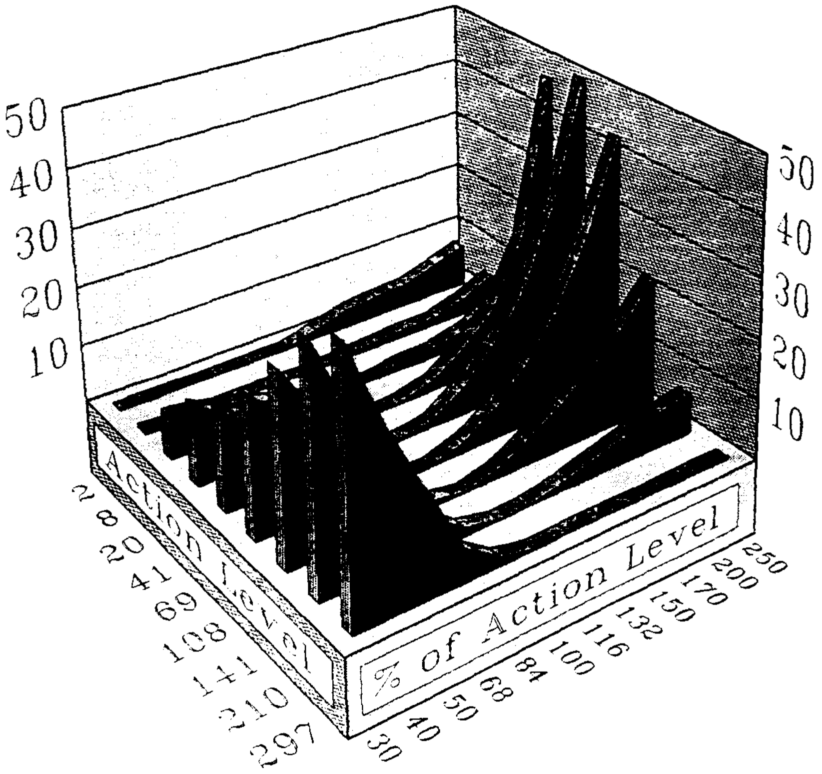

Figure 11. Effects of sample bias on block selection quality; linear loss for each action level. Percents less than 100 simulate positive bias and percents more than 100 simulate negative bias.

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 at the University of Nevada in Las Vegas and through Contract #68017365-154 to Computer Sciences Corporation, it has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.