STATISTICAL ESTIMATION OF SOIL PROPERTIES

An Application

by

Daniel Robert Spikula

B.S., Carnegie-Mellon University
(1981)

Submitted to the Department of
Civil Engineering
in Partial Fulfillment of the
Degree of

MASTER OF SCIENCE IN CIVIL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1983

© Massachusetts Institute of Technology 1983

Signature of Author

Department of Civil Engineering
27 May 1983

Certified by

Gregory B. Baecher
Thesis Supervisor

Accepted by

Francois M.M. Morel, Chairman
Civil Engineering Department Committee

Archives
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUL 12 1983

LIBRARIES
STATISTICAL ESTIMATION OF SOIL PROPERTIES

An Application

by

Daniel Robert Spikula

Submitted to the Department of Civil Engineering on
27 May 1983 in partial fulfillment of the requirements
for the Degree of Master of Science in Civil Engineering

ABSTRACT

A best (minimum variance) linear unbiased estimator
(BLUE) is presented in the context of geotechnical
exploration. The estimation technique exploits spatial
correlation, manifest in the autocorrelation function, to
interpolate exploration data through a weighted sum of
observations.

The BLUE estimation technique is illustrated for data
from Standard Penetration Tests performed on sand at an
industrial site. Preliminary autocorrelation analyses are
performed, but large scatter and sparse data from borings
separated by less than the autocovariance distance prevented
precise determination of the autocorrelation function.
Imprecise autocorrelation functions are typical in
gеотехnісаl аррlісаtіоns, and because the function is
fundamental to the BLUE technique, the estimators are
parametrically studied. Estimators and estimation errors for
the depth average of SPT's are calculated at numerous
locations with various combinations of autocorrelation
parameters to

1. assess effects of parameter values on the BLUE
   estimators and estimation variances, and to

2. verify BLUE estimators and estimation variances by
   comparison to observed data.

Analysis of observed and predicted estimation errors provides
indirect estimates of autocorrelation distances and
measurement errors.

Thesis Supervisor: Dr. Gregory B. Baecher
Title: Associate Professor of Civil Engineering
Head, Constructed Facilities Division
ACKNOWLEDGEMENTS

The writer is grateful for the helpful guidance and supervision provided by Professor Gregory B. Baecher during the execution of this thesis. T. William Lambe, Inc., provided data for the study, and the M.I.T. Seagrant Program provided partial funding of the work.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>1</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>2</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>3</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>4</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>6</td>
</tr>
<tr>
<td>II. SITE AND DATA DESCRIPTIONS</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Site Description</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Soil Classifications</td>
<td>9</td>
</tr>
<tr>
<td>2.4 SPT Data</td>
<td>9</td>
</tr>
<tr>
<td>III. AUTOCORRELATION ANALYSIS</td>
<td>14</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>14</td>
</tr>
<tr>
<td>3.2 Application to Field Data</td>
<td>16</td>
</tr>
<tr>
<td>3.2.1 Review of Boring Logs</td>
<td>18</td>
</tr>
<tr>
<td>3.2.2 Regression Analyses</td>
<td>20</td>
</tr>
<tr>
<td>3.2.3 Autocorrelation Analyses</td>
<td>21</td>
</tr>
<tr>
<td>3.3 Summary</td>
<td>23</td>
</tr>
<tr>
<td>IV. BEST LINEAR UNBIASED ESTIMATORS</td>
<td>39</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>39</td>
</tr>
<tr>
<td>4.2 Derivation of BLUE Estimators</td>
<td>42</td>
</tr>
<tr>
<td>4.3 Matrix Representation of the Estimation System</td>
<td>43</td>
</tr>
<tr>
<td>4.4 Discussion of the Estimation System Solution</td>
<td>44</td>
</tr>
</tbody>
</table>
V. APPLICATION OF BLUE ESTIMATORS TO FIELD DATA ........ 48

5.1 Introduction .......................................... 48
5.1.1 Depth-average Blow Count. ....................... 48
5.1.2 Parameters of the BLUE Estimator ............... 49
5.1.3 Estimation of Parameters ......................... 51
5.1.4 Parametric Study .................................. 52

5.2 Effects of Autocorrelation Parameters on BLUE Estimators .............................................. 52
5.2.1 BLUE Estimators ..................................... 52
5.2.2 Estimation Variance ............................... 54
5.2.3 Summary .............................................. 55

5.3 Accuracy of Estimators ............................... 56
5.3.1 Observed Estimation Error ....................... 57
5.3.2 Error Ratio ........................................ 57
5.3.3 Summary .............................................. 60

VI. CONCLUSION ........................................... 84

VII. REFERENCES ........................................... 85
I. INTRODUCTION

Geotechnical engineers frequently design foundations and earth structures based on scattered and variable data from expensive exploration programs. Traditional evaluation of this data primarily depended upon "engineering judgement" without the guidance of analytical models. However, approximate statistical methods are now available to provide a rational framework for the selection of engineering parameters such as soil strength or compressibility. These methods can ultimately provide an estimate of the reliability of a structure, thereby leading to rational selection of safety factors (Whitman, 1983; Baecher, 1982; Vanmarcke, 1977) and more economical engineering designs, but these methods can also be used with deterministic analyses. For example, deterministic stability or settlement calculations can be performed for a site where exploration data are statistically interpreted.

In current practice, exploration data are typically interpreted in simplified and sometimes irrational ways. For example, soil strata changes are commonly interpolated linearly between boring locations without physical reason, and Standard Penetration Test (SPT) blow counts for settlement or bearing capacity analyses are usually estimated as the arithmetic mean of all SPT's at a site. The latter estimator is efficient only under restrictive conditions which are often not met. One such condition is that the SPT data be uncorrelated.
A simple statistical estimation technique founded upon the concept of spatial correlation is presented in this paper for the interpolation of soil properties between observed values. The technique assigns a linear weighting factor to each observation of a soil property, producing the best linear unbiased estimator (BLUE) of the property at a point in the soil mass.

In addition to presentation of the BLUE estimation technique, this paper also describes application of the technique to SPT data obtained from a Boston area consulting firm. The data are from an industrial site where two sandy soils, one natural and one hydraulic fill, are present.

A subset of the sand data was studied for spatial correlation via an autocorrelation analysis. Also, the depth-average blow count of the natural sand and fill is computed for each boring, and the results are interpolated with the BLUE technique to provide estimators of depth-average blow count and estimation variance over the entire site.

The BLUE technique is parametrically studied to examine the effect of the degree of spatial correlation on the estimators and to determine appropriate values of autocorrelation parameters for the SPT data in lieu of conclusive autocorrelation analyses.

The data and site description are presented in Chapter II. Autocorrelation is defined and applied to the SPT data in Chapter III. The BLUE estimation technique is developed in Chapter IV and is applied as part of the parametric study in Chapter V. Finally, conclusions are summarized in Chapter VI.
II. SITE AND DATA DESCRIPTIONS

2.1 INTRODUCTION

An extensive boring program was performed at the site, located on reclaimed land. The program comprised 48 borings with SPT's spaced at about one meter vertical intervals and was summarized by SPT blow count versus elevation plots for each boring. Unfortunately, the scale of the plots prevented statistical analysis of the vertical structure of the soil, because elevations could only be determined to the nearest 0.1m. Nevertheless, the plots were adequate for horizontal correlation analyses required for the proposed study.

Figure 2-1 shows the boring locations and identifies 33 locations where a surcharge (typically 7m high) was in place at the time of exploration. The remaining 15 boring locations were not surcharged.

2.2 SITE DESCRIPTION

The borings revealed typical subsurface conditions depicted in Figure 2-2. The site consists of dumped miscellaneous fill (termed "upper fill", \(U_f\)), over hydraulically-placed fill of variable thickness, underlain by thick alluvial deposits.

The hydraulic fill was dredged from a bay and a nearby canal and was placed in two phases: first, the northern and western sections, and then the remainder of the site. The hydraulic fill contains considerable fines (passing the number 200 sieve) and is sub-classified as sandy clay fill (\(F_C\)) or sand
fill (F_s) with silt and clay parts. The fill is primarily F_s with layers of F_c except for the south-central and south-eastern portions which are primarily F_c with F_s layers.

Below the hydraulic fill is a layer of alluvial sand (A_g), consisting of silty sand and sandy silt. The alluvial sand is underlain by a deep alluvial clay (A_c) deposit.

2.3 SOIL CLASSIFICATIONS

All soils were classified by engineers directly involved in the exploration program. The hydraulic fill soil classifications were sometimes inconsistent. The distinction between F_c and F_s soils was not clear from available data. Boring logs occasionally appeared in two or more figures of the data set with conflicting soil classification.

Available summary grain size curves could not be matched to specific sample locations, rendering the curves useless for classification of samples. Nevertheless, the curves did indicate that the fill was generally granular with about 20 or 30 percent of the soil passing the number 200 sieve, although fines contents up to 60 percent were encountered.

Penetration resistance was also useless for soil classification, because the F_c blow counts were within the range of the F_s blow counts, which were as low as zero blows per foot.

2.4 SPT DATA

Of the soils identified at the site, only the hydraulic fill and alluvial sand were considered for the present study.
SPT data on other soils were identified but not studied.

Relevant raw SPT data were corrected for overburden effects to provide a relatively unbiased estimate of the true penetration resistance of the soil. The correction from Teng (1962) was used for effective vertical stress less than 40 psi:

\[ N_C = \frac{50N}{\sigma_v + 10} \quad (2-1) \]

where

\[ N_C = \text{corrected blow count (bpf),} \]
\[ N = \text{raw SPT blow count (bpf), and} \]
\[ \sigma_v = \text{vertical effective stress at the location of the test (psi).} \]

The vertical effective stress was estimated with the following assumptions based upon available information:

- ground surface at the level existing at the time of exploration
- phreatic surface at elevation +5.2m
- bouyant unit weight = 0.9 TCM below elevation +5.2m
- total unit weight = 1.7 TCM above elevation +5.2m
  = weight of surcharge

The corrected blow counts are plotted versus elevation in Figure 2-3. The figure represents 355 data points identified as hydraulic fill or sandy alluvium (that is, Fc, Fs, and As soils).
SURCHARGE
-AT SELECTED BORINGS-

\[ \text{UPPER FILL, } U_f \]

\[ \text{EL } 5.2 \text{ m} \pm \]

\[ \text{FILL SAND, } F_s \text{ / FILL CLAY, } F_c \]

\[ \text{EL } 1.5 \pm 1.8 \text{ m} \]

HYDRAULIC FILL

-VARIED LAYERING-

\[ \text{ALLUVIAL SAND, } A_s \]

\[ \text{LOCATION OF BOUNDARY FLUCTUATES} \]

\[ \text{EL } -8 \pm 1.6 \text{ m} \]

\[ \text{ALLUVIAL CLAY, } A_c \]

**FIGURE 2-2** Typical Subsurface Conditions
FIGURE 2-3  Variation of Corrected Blow Count with Elevation
IIII. AUTOCORRELATION ANALYSIS

3.1 INTRODUCTION

Autocorrelation analysis is derived from random process theory but is applicable to non-random processes such as the variability of a soil property across a site (Baecher, 1982; Vanmarcke, 1977). The autocorrelation function, $c[r]$, can be defined in terms of means and variances:

$$
C[r] = \frac{E[(z_i - E[z_i])(z_{i+r} - E[z_{i+r}])]}{E[(z - E[z])^2]} 
$$

(3-1)

where $z_i$ is the value of a soil property at location $i$, $z_{i+r}$ is the value of the same property measured at a distance $r$ from location $i$, and $E[\ ]$ is the expectation operator (Benjamin and Cornell, 1970).

The denominator of Equation 3-1 is the site variance. The numerator is termed the autocovariance, $C[r]$, and is the product of residuals off the mean trend of property $z$. Autocovariance is a description of the correlation of a soil property measured at different locations in a soil mass. If a property is uncorrelated (as one would expect at large separations), then the product of residuals would sometimes be positive, sometimes negative, and would be zero on average. However at small separation, one would expect a strong correlation and a positive product of residuals. In fact, with no separation ($r$ equal zero), the points are perfectly correlated, the autocovariance is maximized, and $C[r]$ equals the site variance. The
autocorrelation function is obtained by normalizing the autocovariance with respect to the site variance, thereby limiting the range of autocorrelation to \([-1, +1]\).

The distance at which \(c[r]\) equals \(1/e\), where \(e\) is the base of natural logarithms, is termed the "autocorrelation distance," or, equivalently, the "autocovariance distance". The autocovariance distance is an index of the fluctuation of a soil property about its expected value. It is also the distance over which a soil property can be expected to exhibit significant correlations. The autocovariance distance, when combined with the mean and variance, provides a summary of the variability of a soil property across a site.

A typical autocorrelation function (Figure 3-1a) has a discontinuity at \(r\) equal zero. The discontinuity is caused by random errors (i.e., errors with zero mean and a non-zero variance). The autocorrelation function for a purely random process (Figure 3-1b) is a spike equal to one at \(r=0\) and zero elsewhere. The autocorrelation function of a property which exhibits only spatial variability about its mean value equals one at \(r\) equal zero, but the function monotonically decreases to zero at sufficiently large distances (Figure 3-1c).

The variance of measurements of a soil property can be modeled as the sum of the random and spatial components of variance (Baecher, 1982):

\[
\text{Var[measured property]} = \text{Var[random]} + \text{Var[spatial]} \quad (3-2)
\]

This allows superposition of the autocorrelation functions in
Figures 3-1a and 3-1b, and, more importantly, allows an estimation of the fraction of variability due to spatial variability about the mean and of variability due to random noise in measurements. For example, the function of Figure 3-1a reveals 20 percent variability due to random noise and 80 percent due to spatial variability. Hence, the autocorrelation function is an important tool for statistically describing the structure of a soil deposit.

It should be noted that a unique autocorrelation function does not exist for a site. Rather, the functions depend upon the selected mean trend, determined through standard regression techniques (Benjamin and Cornell, 1977; Draper and Smith, 1981; Koutsoyiannis, 1977). Care should be taken to provide a sound, unbiased estimate of the mean trend, because some geotechnical tests (for example, standard penetration tests or field vanes) provide a biased measurement of the true soil property. Raw data from such tests must be corrected before correlation analysis results are interpreted.

3.2 APPLICATION TO FIELD DATA

Autocorrelation analyses were performed on the SPT data to illustrate the concept of autocorrelation and to illustrate a typical methodology for estimation of an autocorrelation function. The autocorrelation function was assumed to be similar to the function of Figure 3-1(a) with a spike equal to one at r=0, decaying exponentially for separations greater than zero:
\[ c[r] = b \exp(-r/r_0) \quad \text{for} \ r > 0 \]  

(3-3)

where

- \( b \) is the ratio of spatial to total variance,
- \( r \) is the separation distance, and
- \( r_0 \) is the autocovariance distance.

The values of \( r_0 \) and \( b \) were estimated from the field data.

The autocorrelation analysis of the field data focused on the horizontal correlation structure of the sandy alluvium and hydraulic fill SPT data. Vertical correlation was not investigated, because elevations of test locations were not known with sufficient accuracy.

The \( F_c \) fill was identified as well as possible (although identification was arbitrary in some cases) and was removed from the analysis, leaving only the \( F_s \) and \( A_s \) sands, which were not differentiated. Sixteen sand SPT data greater than 2.0 standard deviations (in an absolute sense) from the mean blow count were considered anomalies, possibly due to rocks or other obstructions, and were also removed from the correlation analysis, leaving 267 sand SPT data.

The consequences of removing so many "anomalous" data are discussed after presentation of the horizontal correlation analysis, which was performed in three steps:

1. review of boring logs,
2. regression analyses, and
3. autocorrelation analyses.
3.2.1 Review of Boring Logs

Before regression and autocorrelation analyses were begun, the general subsurface conditions were reviewed. The boring logs were aligned with a standard elevation datum at the site (the KP base), and the variation of corrected blow count with elevation was perused for the sands. Subsequently, the effect of referencing SPT data to a physical "elevation datum" (the top of the alluvial clay deposit), as opposed to the arbitrary KP base, was investigated.

Inspection of the 48 boring logs revealed an apparent dichotomy, possibly caused by the two-phase filling operation noted earlier in this paper. Fourteen borings, identified in Figure 3-2, were labeled as "Type 2" borings. The hydraulic fill of Type 2 borings generally exhibited considerable amounts of fine-grained soil with $F_C$ clay occurring in substantial thicknesses and $F_S$ sand showing low blow counts on average (a typical Type 2 boring is presented in Figure 3-3b). In contrast, the hydraulic fill of the other 34 borings (termed "Type 1" borings) contained fewer fines and produced higher blow counts on average (see Figure 3-3a for a typical Type 1 boring).

The average corrected blow counts for Type 1 and Type 2 borings were calculated for arbitrary 2.0m layers. The averages are shown in Figures 3-4a and 3-4b along with layer standard deviations and the number of data per interval. The Type 2 borings yielded an average blow count of 13 bpf over elevations -1 to -3m. Type 1 borings yielded an average of 19 bpf for
the same layer. Otherwise, both Type 1 and Type 2 borings were similar with layer averages on the order of 17 to 20 bpf regardless of boring type.

A composite of all borings (Figure 3-4c) masks the 13 bpf layer of the Type 2 borings with a layer average of 18 bpf over elevations -1 to -3m. The prevailing Type 1 borings dominate the layer means and standard deviations, as one would expect given the number of data per layer per boring type.

One should realize that very low blow count $F_s$ sand exists in Type 1 borings (evidenced by the large standard deviations of Figure 3-4) as well as Type 2 borings. Type 2 borings have lower average $F_s$ blow counts, because the additional fines of Type 2 borings limit the occurrence of high blow counts. (Blow count generally decreases with increasing percentage of fine-grained material.)

If one were to consider the blow count for a 2.0m layer of any boring at the site, the type of boring (that is, Type 1 or Type 2) would be indistinguishable. The "apparent" dichotomy would disappear and the Type 2 borings would be lost in the scatter of the Type 1 borings.

Hence, two treatments of the borings logs were chosen for correlation analyses:

1. Distinguish between Type 1 and Type 2 borings, and operate on only one type at a time.

2. Do not distinguish between borings, and operate on the composite of all borings.
Additionally, the effect of using the top of the alluvial clay as an elevation datum was investigated. The alignment of borings with respect to the AC datum is schematically illustrated in Figure 3-5. The AC alignment tends to reduce the number of $A_g$ sand data in the upper layers of the profile, thereby reducing the average blow count in the hydraulic fill (Figure 3-5). Both boring Types 1 and 2 show reduced average blow counts over a range of 4 to 8m above the top of the alluvial clay. Consequently, the AC alignment deserves consideration in the correlation analysis.

3.3.2 Regression Analyses

Numerous regression analyses (Koutsoyiannis, 1977) were performed in an attempt to find a simple trend in the corrected blow count data. Linear trends of $N_C$ with elevation were considered over various ranges of elevations for several elevation datums. Additionally, the Type 1/Type 2 division of the site led to investigation of a scheme with two coordinate systems. The basic scheme postulates that $N_C$ depends upon the vertical location of the point of interest and the distance from a low blow count soil zone (i.e., the Type 2 boring zones with substantial occurrences of $F_C$ clay).

The origins of the coordinate systems are shown in Figure 3-6. Three parameters are used: elevation, $r_1$, and $r_2$ where $r_1$ is the distance from origin $i$ to the point of interest with $i$ equal to 1 or 2.
Proposed regression models are presented in Tables 3-1 and 3-2. Table 3-1 contains general expressions, and Table 3-2 summarizes analyses performed. Results of all cases in Table 3-2 reveal no discernable trends, with r-squared values (the square of the sample correlation coefficient) consistently below 0.1. No reason was found for differentiating between Type 1 and Type 2 borings. Therefore, in the absence of trends, the mean and variance were considered to be constant over limited ranges of elevation.

3.2.3 Autocorrelation Analyses

Overview

Horizontal autocorrelation functions were calculated for corrected blow counts assuming statistical stationarity at each layer of soil. Layer means and variances were estimated by the mean and variance of the sample blow count data, and a moment estimator of the autocorrelation function (Equation 3-1) was found by averaging the product of residuals over 25m intervals. Few data exist at small spacings (less than 40m), so the autocorrelation results must be interpreted with caution. Additional data might reveal a smaller autocovariance distance for any or all of the analyses of this study.

Results and Discussion

Autocorrelation analyses were performed on selected layers of Type 1 borings, Type 2 borings, and a composite of all borings; both the KP and AC alignments were analyzed. Insufficient data precluded meaningful interpretation of analyses
performed on Type 2 borings, but results were obtained for Type 1 borings and for the composite. Four cases, numbered 1 through 4, are summarized in Table 3-3. The autocorrelation functions are plotted in Figures 3-7 through 3-10, respectively.

The use of Type 1 borings compared to the use of all borings had null effect on the autocorrelation curves. The AC alignment produced nearly identical results for Case 3 (Type 1 borings) and Case 4 (composite of borings). Cases 3 and 4 appear to have longer autocovariance distances and slightly smaller coefficients of variation (Table 3-3) than Cases 1 and 2 (the KP alignments); however, the lack of closely spaced data and the large data scatter of Figures 3-7 through 3-10 diminish the significance of the apparent differences in the autocorrelation functions of the cases.

Based upon the four cases, the horizontal autocovariance distance for the Aₕ and Fₕ sand SPT data is about 50m and is roughly constant with depth, although one layer at the site appears to have a smaller rₒ. The autocorrelation functions represented by the open squares of Figures 3-7 through 3-10 are poorly correlated, and the small degree of correlation is accentuated by cases 1 and 2.

Although the autocovariance distance can be roughly inferred from the analyses, accurate estimation of the ratio, b, of spatial variance to total variance is very difficult. The value of b is estimated by extrapolating the autocorrelation function back to zero separation distance. However, many data are needed at small spacings to differentiate between random
noise in measurements and small scale (but real) spatial variations. The widely spaced data of Figures 3-7 through 3-10 suggest a $b$-value on the order of 0.6, but this value is little more than a reasonable guess. The exponential autocorrelation function is plotted in Figure 3-11 for $r_o = 50m$ and $b = 0.6$.

Effect of Anomalous Data Removal

In light of the estimation problems caused by bore-hole spacing, and scatter in the autocorrelation functions, the removal of 16 "anomalous" data (or 5 percent of the sand data) probably had little effect on the estimation of the autocorrelation function parameters, $r_o$ and $b$. Inclusion of the data would slightly increase the scatter in the functions, but would not effect the bore-hole spacing problem.

Hence, removal of 5 percent of the data was inconsequential for this analysis; however, identification of anomalous data should generally receive more attention than was given for the preceding autocorrelation analyses. Care must be taken to ensure that the data are truly anomalous or else important features of the correlation structure may be overlooked.

3.3 SUMMARY

The autocorrelation analyses emphasized a very important point: ample data must be available at spacings in the neighborhood of, and smaller than, the autocovariance distance. This data serves two purposes:
1. It improves estimation of the autocovariance (the numerator of Equation 3-1), thereby reducing scatter in the autocorrelation function and providing improved estimates of $r_o$.

2. It facilitates extrapolation of the autocorrelation function back to $r_o = 0$, providing improved estimates of $b$.

Of course, extrapolation of the curve to $r=0$ will always involve some uncertainty, because very small scale, but real, variability may not be identified. Nevertheless, this uncertainty can usually be reduced to a negligible level with judicious spacing of explorations.

Unfortunately, the data for this study were spaced greater than the autocovariance distance. A shortage of SPT data at horizontal distances less than 40m hindered determination of the key autocorrelation parameters, $r_o$ and $b$, of the exponential autocorrelation function.

The autocovariance distance appeared to be about 50m for most layers of sand, although one layer exhibited a lower degree of autocorrelation with $r_o$ apparently less than 25m. The true value of $r_o$ may be much smaller than the values suggested by the analyses, but the estimate of 50m is reasonable.

The value of $b$ was much more difficult to determine than $r_o$. A value of 0.6 can be reasonably, but unconfidently, assumed.
**TABLE 3-1: Regression Analyses - Proposed Models**

<table>
<thead>
<tr>
<th>Model Number</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$N_c = \beta_0 + \beta_1d$</td>
</tr>
<tr>
<td>2</td>
<td>$N_c = \beta_0 + \beta_1d + \beta_2\left(\frac{r_1}{r_1 + r_2}r_2 + \frac{r_2}{r_1 + r_2}r_1\right)$</td>
</tr>
<tr>
<td>3</td>
<td>$N_c = \beta_0 + \beta_1d + \beta_2\left(\frac{1}{r_1}\right) + \beta_3\left(\frac{1}{r_2}\right)$</td>
</tr>
<tr>
<td>4</td>
<td>$N_c = \beta_0 + \beta_1d + \beta_2\left(\frac{1}{r_1^2}\right) + \beta_3\left(\frac{1}{r_2^2}\right)$</td>
</tr>
</tbody>
</table>

$N_c$ = corrected blow count  
$d$ = elevation  
$r_1$ = distance from coordinate origin 1  
$r_2$ = distance from coordinate origin 2  
$\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$ = regression coefficients
TABLE 3-2: Regression Analyses - Summary

<table>
<thead>
<tr>
<th>Elevation datum</th>
<th>Regression Model</th>
<th>Range of Elevations considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP</td>
<td>1</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>-1.0 to -3.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-3.0 to -5.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-5.0 to -7.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>all</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>all</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>all</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>1</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>+1.0 to +4.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+4.0 to +10.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>all</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+2.0 to +4.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+4.0 to +6.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+6.0 to +8.0</td>
<td></td>
</tr>
</tbody>
</table>

Notes:
- Model numbers are defined in Table 3-1.
- All sample correlation coefficients ($r^2$ values) are less than 0.1.
### TABLE 3-3: Summary of Blow Count Autocorrelation Analyses

<table>
<thead>
<tr>
<th>Case No.</th>
<th>Borings Type</th>
<th>Elevation Datum</th>
<th>Range of Elevation</th>
<th>Layer Mean (bpf)</th>
<th>Layer Coefficient of Variation</th>
<th>Estimated Autocorrelation Distance (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34 Typel</td>
<td>KP</td>
<td>-1.0 to -3.0</td>
<td>19.4</td>
<td>0.53</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-3.0 to -5.0</td>
<td>19.6</td>
<td>0.60</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-5.0 to -7.0</td>
<td>22.0</td>
<td>0.45</td>
<td>&lt;25</td>
</tr>
<tr>
<td>2</td>
<td>48 Typel &amp;</td>
<td>KP</td>
<td>+1.0 to -3.0</td>
<td>17.8</td>
<td>0.57</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Type2</td>
<td></td>
<td>-3.0 to -5.0</td>
<td>19.0</td>
<td>0.60</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-5.0 to -7.5</td>
<td>21.6</td>
<td>0.51</td>
<td>&lt;25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-7.5 to -9.0</td>
<td>16.9</td>
<td>0.61</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>34 Typel</td>
<td>AC</td>
<td>8.0 to 10.5</td>
<td>20.8</td>
<td>0.43</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.0 to 8.0</td>
<td>16.6</td>
<td>0.61</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0 to 4.0</td>
<td>22.8</td>
<td>0.47</td>
<td>&lt;35</td>
</tr>
<tr>
<td>4</td>
<td>48 Typel &amp;</td>
<td>AC</td>
<td>8.0 to 10.5</td>
<td>20.4</td>
<td>0.46</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Type2</td>
<td></td>
<td>4.0 to 8.0</td>
<td>15.7</td>
<td>0.64</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0 to 4.0</td>
<td>22.5</td>
<td>0.49</td>
<td>&lt;35</td>
</tr>
</tbody>
</table>
a. Typical Autocorrelation Function

b. Autocorrelation Function for a Random Process

c. Autocorrelation Function Exhibiting Pure Spatial Variability

FIGURE 3-1 Autocorrelation Functions
a. Typical Type 1 Boring  

b. Typical Type 2 Boring  

FIGURE 3-3 Typical Boring Types
FIGURE 3-4 Summary of Boring Logs - KP Alignment. Average corrected sand blow counts for arbitrary 2.0m intervals. Numbers in parentheses denote the number of data per interval.
FIGURE 3-5 Summary of Borings Logs - AC Alignment. Average corrected sand blow counts for arbitrary 2.0m intervals. Numbers in parentheses denote the number of data per interval.
FIGURE 3-6  Coordinate System for Regression Analysis
FIGURE 3-7 Autocorrelation Functions for Case 1: Type I Borings, KP Alignment
FIGURE 3-8  Autocorrelation Functions for Case 2.  All Boringa, KP Alignment
FIGURE 3-9  Autocorrelation Functions for Case 3, Type 1 Borings, AC Alignment
FIGURE 3-10 Autocorrelation Functions for Case 4: All Borings, AC Alignment
FIGURE 3-11 Suggested Autocorrelation Function
IV. BEST LINEAR UNBIASED ESTIMATORS

4.1 INTRODUCTION

Simple statistical techniques can be used to combine a set of soil property measurements at locations throughout a soil mass to yield rational estimates of the property at a location where measurements are unavailable. The rational estimator should account for spatial correlation of soil properties, should be calculable from the data set, and should be easy to use. For simplicity, a linear estimator, \( \hat{z} \), is chosen:

\[
\hat{z} = \sum_{i=1}^{n} w_i z_i
\]  
(4-1)

such that

\[
\frac{1}{n} \sum_{i=1}^{n} w_i = 1.0
\]

where

- \( n \) = number of observations of soil property \( z \),
- \( z_i \) = \( i \)th observation of \( z \) at location \( x_i \) in the soil mass, and
- \( w_i \) = weighting factor assigned to the \( i \)th observation.

The weighting factors are selected to reflect the degree of spatial correlation. If the measurements, \( z_i \), are independent, then an appropriate estimator is the sample mean, \( \bar{z} \), given by Equation 4-1 with \( w_i = 1/n \) for all \( i \). If the measurements are spatially correlated, as indicated by the phenomenon of autocovariance, then the weights must satisfy several spatial relations. These relations are illustrated by
the data of Figure 4-1. A plan view of six data, labeled $z_1$ to $z_6$, is shown. The data are to be weighted and summed to estimate $z$ at point A.

The location of datum $z_1$ is a mirror-image of the location of datum $z_6$ about line $y-y$. Similarly, location 2 is a mirror-image of location 5. These symmetries should be manifested by the weights, as should a lack of symmetry. That is, $w_1$ should equal $w_6$, and $w_2$ should equal $w_5$, but $w_1$ should, generally, differ from $w_2$, because locations 1 and 2 are different distances from point A.

Increasing the separation distance between a datum and the point of interest should reduce the influence of the datum, and because $w_1$ is much further from point A than location 2, $w_1$ should be smaller than $w_2$. Additionally, $w_3$ should be larger than $w_4$, because datum $z_3$ is closer to point A than datum $z_4$. Data are weighted by geometric considerations, independent of the value of individual data; therefore, excessively high or low data will have small effects on the estimator if the data are located at sufficiently large distances from the point of interest. Conversely, if an extreme datum exists in the proximity of a point of interest, this datum will dominate the estimator. Nearby data tend to control the estimator and reduce the effect of distant data.

The geometrical relation between the data and the point of interest should not be the only geometrical consideration;
relative locations of the borings are also significant. If data are located in a cluster, as are data \( z_1 \) and \( z_6 \), the net effect of the cluster should be smaller than the sum of the effects of individual data in the cluster, because the data in the cluster should exhibit relatively strong correlations with one another (that is, they contain redundant information). For example, knowledge of \( z_1 \) gives little additional information for the estimate at point A than datum \( z_6 \); and although data \( z_1, z_6, \) and \( z_3 \) are roughly equidistant from point A, \( z_3 \) should receive a larger weight than \( z_1 \) or \( z_6 \), because the correlation between \( z_3 \) with \( z_1 \) or \( z_6 \) should be small compared to the correlation between \( z_1 \) and \( z_6 \).

One final property of the estimator is desirable. Suppose a seventh datum existed at point A. Because the value of \( z_7 \) is known, the estimator should be equal to \( z_7 \) at point A; that is, \( w_7 \) should equal one and all other weights should equal zero.

The above properties can be ensured if two conditions are imposed on the estimator of Equation 4-1:

1. Unbiasedness
2. Minimum variance

The "best linear unbiased estimator" (BLUE) is that having minimum estimation variance of any linear unbiased estimator.

BLUE estimators are common for many problems of prediction (Whittle, 1963) and are the basis of "Kriging" in ore reserve estimation (Journel and Huijbregts, 1978).
The development of kriging presented by Journel and Huijbregts is closely related to the BLUE estimation technique for a stationary soil property described below.

4.2 DERIVATION OF BLUE ESTIMATORS

The estimator \( z \) of Equation 4-1 becomes a BLUE estimator if the weights, \( w_i \), are determined by a three-step procedure:

1. Apply the unbiased condition, thereby restricting the \( w_i \) such that:

\[
\sum_{i=1}^{n} w_i = 1
\]  
(4-2)

that is, \( E[\hat{z}] = E[\Sigma w_i z_i] \)

\[ = \Sigma w_i E[z_i] \]

\[ = E[z_i] \text{ iff } \Sigma w_i = 1 \]

2. Determine the estimation variance, \( \sigma_e^2 \), for the estimator \( \hat{z} \) of the true soil property \( z^* \) at location \( x^* \):

\[
\sigma_e^2 = E[(\hat{z} - z^*)^2]
\]  
(4-3)

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j C(z_i, z_j) - 2 \sum_{i=1}^{n} w_i C(z_i, z^*) + C(z^*, z^*)
\]  
(4-4)

where \( C(z_i, z_j) \) is the autocovariance of \( z \) between locations \( x_i \) and \( x_j \).

3. Minimize Equation 4-4 over all \( w_i \) subject to Equation 4-2.

This yields optimal values of \( w_i \) and the BLUE estimator.
Minimization is conveniently accomplished by Lagrangian techniques resulting in a system of \( n+1 \) equations with \( n+1 \) unknowns:

\[
\sum_{i=1}^{n} w_i C(z_i, z_j) + \lambda = C(z_j, z^*) \quad \text{for } j=1 \text{ to } n \quad (4-5)
\]

\[
\sum_{i=1}^{n} w_i = 1 \quad (4-2)
\]

where \( \lambda \) is a Lagrange multiplier. For convenience, Equations 4-5 and 4-2 are termed the "estimation system."

The autocovariance terms depend only on the autocovariance function, the geometry of the observations, and the point to be estimated. Therefore, only the \( w_i \) and \( \lambda \) are unknown. Solution of the estimation system for all \( w_i \) and \( \lambda \) results in optimal values and the optimal estimator. Substitution of the optimal \( w_i \) into Equation 4-4 produces the minimum estimation variance, \( \sigma_e^2 \); or the optimal weights and \( \lambda \) may be substituted into an equivalent equation (David, 1977):

\[
\sigma_e^2 = C(z^*, z^*) - \lambda - \sum_{i=1}^{n} w_i C(z_i, z^*) \quad (4-6)
\]

4.3 MATRIX REPRESENTATION OF THE ESTIMATION SYSTEM

The estimation system can be represented as partitioned matrices:
where

\[
\begin{bmatrix}
C & \mathbf{I} \\
\mathbf{1}^T & \lambda
\end{bmatrix}
\begin{bmatrix}
\mathbf{W} \\
\lambda
\end{bmatrix} =
\begin{bmatrix}
G \\
\mathbf{1}
\end{bmatrix}
\] (4-7)

\( C \) = the covariance matrix (n×n)
\( \mathbf{1} \) = column vector with all elements equal to one (n×1)
\( \mathbf{0} \) = null matrix (1×1)
\( \lambda \) = Lagrange multiplier matrix (1×1)
\( \mathbf{W} \) = column vector of \( w_i \) terms (n×1)
\( \mathbf{G} \) = column vector of \( C(z_i, z^*) \) terms (n×1)
\( \mathbf{1} \) = matrix with element equal to one (1×1)

Representing the n+1 by n+1 square matrix of Equation 4-7 as \( \mathbf{K} \) allows the estimation system's solution to be written as:

\[
\begin{bmatrix}
\mathbf{W} \\
\lambda
\end{bmatrix} = \mathbf{K}^{-1}
\begin{bmatrix}
\mathbf{G} \\
\mathbf{1}
\end{bmatrix}
\] (4-8)

4.4 DISCUSSION OF ESTIMATION SYSTEM SOLUTION

The solution can be obtained by various linear equation solvers or numerical inversion schemes (Hornbeck, 1975). The covariance matrix is a full symmetric matrix of positive elements with maximum elements on the main diagonal and is positive definite. Hence the estimation system is very well conditioned and tractable.
Another interesting feature of the estimation system is the relationship between the optimal weights and the autocovariance function. If the autocovariance function is of the form

\[ C(r) = C(0) \exp(-r/r_o) \]  \hspace{1cm} (4-9)

where \( C(0) \) is the autocovariance at separation distance \( r=0 \) (that is, the spatial variance), then the optimal weights will be independent of the magnitude of \( C(0) \). Nevertheless, a change in \( C(0) \) results in a proportional change in \( \lambda \) and, consequently, a proportional change in the estimation variance.

For example, reducing \( C(0) \) by a factor \( b \) reduces all covariance terms in Equation 4-5 by the factor \( b \), has no effect on \( W \), and reduces \( \lambda \) by the factor \( b \). Therefore, the autocovariance distance, \( r_o \), is the key parameter for obtaining BLUE estimators, but both \( r_o \) and \( C(0) \) influence the estimation variance.

One final aspect of the estimation system is the effect of random measurement noise, which causes a spike in the autocovariance curve at zero separation distance. Because the measurement noise is random and has equal effect (in a statistical sense) on all observations, the estimation system can be solved considering only spatial variance. This results in the optimal weights; but the total estimation variance \( \sigma_z^2 \), is the sum of \( \sigma_e^2 \), the spatial estimation variance of Equation 4-6, and the random noise, \( \sigma_r^2 \):
\[ \sigma_z^2 = \sigma_e^2 + \sigma_r^2 \]  \hspace{1cm} (4-10)

Hence the existence of measurement noise does not complicate the mathematics of the estimation system solution.
FIGURE 4-1 Plan Location of Six Data
V. APPLICATION OF BLUE ESTIMATORS TO FIELD DATA

5.1 INTRODUCTION

SPT data from the hydraulic fill and sandy alluvium (Chapter II) provide an opportunity for application of BLUE estimators to a typical engineering problem: estimation of depth-average blow counts at various locations across a site. The depth average may be required for settlement calculations, for example, and knowledge of the spatial variation of the depth-average is particularly useful for differential settlement problems.

The BLUE technique can determine important parameters for settlement analysis, such as the thickness of the sandy layer, and can also be used to directly estimate settlement (Baecher, 1981) or other quantities of interest to the geotechnical engineer. However, the technique is herein limited to the estimation of depth-average blow count. Settlement calculations are not performed.

5.1.1 Depth-average Blow Count

The observed depth-average blow count, $X_i$, is defined for the $i^{th}$ boring as the arithmetic mean of the corrected blow counts for the sandy alluvium ($A_s$ soil) and the hydraulic fill (both $F_s$ and $F_c$ soil). That is,

$$X_i = \frac{1}{K_i} \sum_{j=1}^{k_i} (N_{c,j})$$

(5-1)
where

\[ i = \text{boring number}; \]

\[ k_i = \text{total number of } A_g, F_g, \text{ and } F_C \text{ data from boring } i; \]

\[ (N_C)_j = \text{corrected blow count from the } j^{\text{th}} \text{ SPT in the boring on } A_g, F_g, \text{ or } F_C \text{ soil.} \]

The \( F_C \) soil is included with the more granular \( A_g \) and \( F_g \) soil for several reasons. The \( F_C \) soil generally occurs as lenses or thin layers within the hydraulic fill, and distinct layer boundaries are unapparent. Additionally, both the \( F_g \) and \( F_C \) soil are fairly granular, and in practice, settlement analyses based upon the available data would not differentiate between the \( F_g \) and \( F_C \) data, hence the inclusion of \( F_C \) data in depth-average blow counts.

The observed depth-average blow counts are shown for their respective boring locations in Figure 5-1. The figure also defines a cartesian coordinate system for referencing locations of borings and estimators.

5.1.2 Parameters of the BLUE Estimator

At any location, the BLUE estimator of depth-average blow count depends upon the covariance terms of Equation 4-5. The covariance terms are mandated by the autocorrelation function, or more appropriately, the autocovariance function, which is the product of the total variance and the autocorrelation function. The autocovariance function and, therefore, the estimator depend upon three fundamental parameters:
1. total variance, $\sigma^2$;
2. spatial component of variance, $\sigma^2$; and
3. autocovariance distance, $r_0$.

The random component of variance, $\sigma^2$, is simply the difference of $\sigma^2$ from $\sigma_s^2$; therefore, $\sigma^2$ is fixed, given $\sigma_s^2$ and $\sigma_r^2$.

The value of $\sigma^2$ is estimated as $45.4 \text{ (bpf)}^2$ by the sample variance,

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - m)^2$$  \hspace{1cm} (5-2)

where

$n = \text{number of data (i.e., borings)},$

$X_i = \text{depth-average blow count for the } i^{th} \text{ boring},$ and

$m = \text{mean of the depth-average blow counts, calculated as}$

$\frac{1}{n} \sum_{i=1}^{n} X_i = 18.4 \text{ bpf.}$

The values of $\sigma_s^2$ and $r_0$ can be directly obtained from an autocovariance function if the function is clearly defined, but in the present case, a well-defined function is not available. The large spacing between bore-holes, which prevented precise estimation of an autocorrelation function for the individual SPT's on the $A_s$ and $F_s$ soil, prevents precise determination of the autocorrelation parameters for the depth-averages of the $A_s$, $F_s$ and $F_c$ blow counts. Therefore, formal autocorrelation analyses are not performed to determine $\sigma_s^2$ or $r_0$. 
Instead, the parameters of the autocorrelation function are assumed to be in the neighborhood of the values suggested by Chapter II, and the effects of varying the values of the parameters are studied. The parametric study yields refined estimates of $\sigma_s^2$ and $r_o$ for the depth-average of the $A_s$, $F_s$, and $F_c$ blow counts.

5.1.3 Estimation of Parameters

Parameter values are determined by an iterative process. Values of $r_o$ and $\sigma_s^2$ are assumed and used to estimate the depth-average at all boring locations while temporarily ignoring the observed value of the depth-average. Next, the estimator is compared to the observed value, and the observed estimation error is compared to the estimation error predicted by the BLUE technique. The errors are analyzed for all estimators produced by the pair of parameters (that is, for all boring locations). The parameter values are then revised, and the estimators and errors are recalculated for several other values of the parameters. The values producing the minimum errors are the best estimators of $\sigma_s^2$ and $r_o$.

In principal, the preceding method precisely estimates $\sigma_s^2$ and $r_o$. These estimators, combined with knowledge of the total variance, $\sigma^2$, define the autocorrelation function. Hence, the function can be indirectly determined without formal autocorrelation analyses. The indirect parametric method reduces the scatter typical of most autocorrelation analyses, but the method does not ameliorate the problem of sparse data at small separation distance. Therefore, as in
estimators are independent of \( b \); therefore, all values of \( b \) produce identical estimators at a given location for a particular \( r_0 \).

For each \( r_0 \) value, the average blow count was estimated at each grid-line intersection of Figure 5-2. The grid spacing of 50m is arbitrary, as the BLUE technique may be used to estimate any number of locations. The grid of estimators, along with observed average blow counts at boring locations are used to generate topographic maps of depth-average blow count (Figures 5-3).

The maps illustrate the smoothing effect of large autocovariance distances. This effect is emphasized by Figure 5-4. The figure shows the variation of the estimator along the \( x=200m \) line, which is a typical section through the site (Section 200) for \( r_0 \) of 25, 50, and 100m and \( b \) equal to 0.50. The \( r_0 = 25m \) line undulates about the regional average blow count. Observed blow counts rise and fall as individual data mandates, but the effect of individual data are rapidly reduced as the estimator's location moves away from data.

The maps and the section imply little difference (from a practical engineering standpoint) between estimators based on a range of reasonable \( r_0 \) values, from approximately 30 to 50m. The BLUE estimators are moderately robust with respect to \( r_0 \), and as previously stated, the estimators are independent of the magnitudes of the spatial and random components of variance. However, the estimation variance is dependent upon both the autocovariance distance and the variance components.
5.2.2 Estimation Variance

The spatial estimation variance, $\sigma^2_e$, associated with Figures 5-3 are mapped in Figures 5-5 for $b=0.50$. The total estimation variance, $\sigma^2$, is found by adding the random component of variance, $\sigma^2_r$, to the values indicated on the maps. The total estimation variance along Section 200 is plotted in Figure 5-6 for $r_o$ values of 25, 50, and 100m and $b=0.50$. This figure and the maps illustrate the rapid increase of the estimation variance as the distance from observations increases, although the scale of the maps obscures the rate of change in $\sigma^2$. To illuminate the effect of $r_o$ on $\sigma^2$, a 100m interval of the section is enlarged in Figure 5-7. The figure represents the typical behavior of the estimation variance in the vicinity of boring locations.

For distances within the autocovariance distance from the nearest boring, the estimation variance exhibits a wide range of values, but at distances exceeding $r_o$ from the nearest boring, the estimation variance approaches $\sigma^2$.

The estimation variance also approaches $\sigma^2$ for all values of $b$ at distance from data locations, but the estimation variance is significantly affected by $b$ in the vicinity of data locations. The estimation variance along section 200 is plotted (Figure 5-8) for $r_o$ equal to 50m and $b$ equal to 0.25, 0.50, and 1.00.

Figure 5-8 shows that $b$ appears to be the most important parameter for determining estimation variance in an
area near data. If there is no measurement noise, then the estimated soil property equals the measured soil property; however, if measurement noise exists, then no estimation technique will, on average, provide an estimator with a smaller variance than the measurement noise.

In general, the effect of b at any location is easy to visualize since spatial estimation variance is proportional to b and measurement noise is proportional to (1-b).

5.2.3 Summary

Producing BLUE estimates and associated estimation variances of the depth-average blow count for several values of the autocovariance distance, $r_o$, and the fraction of variance due to spatial variance, b, illustrates several important properties of the estimation technique:

- The value of b does not affect values of BLUE estimators, assuming the total variance is constant.

- The value of b strongly influences the estimation error at locations near data.

- Estimation variance decreases as $r/r_o$ increases, where r is the distance to a datum.

- The estimator equals the value of the observed datum at data locations and tends toward the regional data average as the distance from individual data increases.

- Large $r_o$ values produce a smooth variation in estimators across a site, but small $r_o$ values produce estimators which reflect small fluctuations in data, resulting in greater spatial variation in the estimators.
5.3 ACCURACY OF ESTIMATORS

The best BLUE estimators are searched for by selecting several combinations of \( r_0 \) and \( b \). For each combination, each of the 48 data (depth-average blow counts) are sequentially disregarded and the other 47 data are used to provide an estimator of depth-average blow count at the location of the disregarded datum. This procedure is used to calculate 48 occurrences for each of two statistics:

1. The observed estimation error, \( \varepsilon \), defined as the difference between the estimated value, \( \hat{X} \), and the observed value of depth-average blow count, \( X \):

\[
\varepsilon = \hat{X} - X
\]

2. The error ratio, \( E \), defined as the observed estimation error divided by the predicted estimation error:

\[
E = \frac{\hat{X} - X}{\sqrt{\frac{\sigma_e^2}{\sigma_r^2}}}
\]

The mean and variance of \( \varepsilon \) and \( E \) are tabulated for various combinations of \( r_0 \) and \( b \) (Table 5-1). The mean value of \( \varepsilon \) is approximately zero for all combinations at \( r_0 \) and \( b \) because of the unbiased condition imposed by the estimation system. Minimizing the variance of \( \varepsilon \) over all values of \( r_0 \) discloses the appropriate value of \( r_0 \), assuming a linear estimator and an exponential autocovariance function.
Minimizing the variance of $E$ discloses $b$, the fraction of variance as spatial variance.

5.3.1 Observed Estimation Error

The variance of the estimation error is plotted against autocovariance distance in Figure 5-9 along with the 90 percent confidence interval for the variance. The figure shows a minimum at about $r_o = 35\text{m}$, where the variance is equal to $40.9 \, \text{(bpf)}^2$. This is a 10 percent improvement over the variance associated with the assumption of spatially independent properties (that is, 10 percent of the site variance of 45.4). The large confidence intervals imply that the variance of observed estimation error may be greater or less than the values of Table 5-1, and the exact numerical values are insignificant. Nevertheless, the relative values of variance are significant, because the estimators and, therefore, the estimation errors, depend on $r_o$, as Chapter IV has shown. Therefore, the $r_o$ which produces the minimum variance should be reproducible, on average, given another set of boring logs, although the value of the minimum variance of the observed estimation error may vary.

5.3.2 Error Ratio

The error ratio, $E$, is useful for determining the relative magnitudes of the spatial and random components of variance, described by $b$. The error ratio usually resembles the standard normal variable (Journel and Huijbregts) with a mean value of zero and a variance of one.
Cumulative frequency distributions of $E$ are represented by normal probability plots in Figures 5-10 for various $r_0$ and $b$ values. The symbols indicate the cumulative frequencies based on frequency cell widths of 0.70, and the solid lines represent a standard normal random variable. For comparison, the probability plot of the error ratio generated by an estimator which assumes independent soil properties is presented in Figure 5-11. The figure was generated by letting

$$E = \frac{\bar{X} - X}{\sigma}$$

where $\bar{X}$, the sample mean, is an estimator of $X$; and $\sigma$, the square root of the site variance, is the estimation error associated with the estimator $\bar{X}$.

The probability plots illustrate the approximate normality of $E$ and the effects of $r_0$ and $b$ on the error ratio. The value of $b$ has a minor effect on the distribution of $E$ for the data of this study, although sparse data at small separations prevent strong statements on the effect of $b$.

The value of $b$ is important within the autocovariance distance from a datum but has little effect at large distances. Because borings in the present case are typically greater than 40m apart, attempts to verify the value of $b$ by calculating BLUE estimators at data locations ignore the effect of $b$ at the small distances where it is most important.

Nevertheless, if borings did exist at small spacings, the value of $b$ could be determined from a parametric study. This method of determining $b$ is illustrated with the SPT data, but
the data underemphasize the effect of $b$.

Figure 5-12 shows the variance of $E$ as a function of $r_o$ and $b$. The minimum point of the surface representing variance has ordinates corresponding to the most appropriate parameters of the autocovariance function, because the point represents the minimum variance over all values of $r_o$ and $b$. For additional clarity, the projection of lines of constant $b$ are plotted on the $r_o$ - Variance of $E$ space (Figure 5-13), along with the 90 percent confidence interval for the variance of $E$, whose value should be 1.00. The confidence intervals reveal that the predicted estimation error is very robust, at least for estimators at large distances from data. With the exception of excessive values of $r_o$ or $b$, any combination of $r_o$ and $b$ yield reasonable results for the estimation variance.

Assuming an $r_o$ of 35m (Figure 5-9), Figure 5-13 suggests a $b$-value on the order of 0.50, although the exact value is not clear. Additional borings at spacings smaller than the autocovariance distance would probably accentuate the effects of varying $b$, resulting in a more pronounced curvature of the lines of Figure 5-13 and, therefore, a sharper evaluation of the true value of $b$.

The concave curvature of the lines in Figure 5-13 is intuitively satisfying. For $r_o$ too large, the spatial estimation variance will be too small because strong correlations are assumed. For $r_o$ too small, the spatial estimation variance will increase, causing the variance of $E$ to decrease; however, the observed estimation error, $\hat{X} - X$
will increase, because the estimator will not manifest the information provided by the true spatial correlation.

The occurrence of the variance of E below 1.00 is probably caused by statistical errors inherent in finite sample sizes. The confidence interval on the variance suggests that for repeated sampling, the variance can be expected to exhibit large scatter about the theoretical value of 1.00.

5.3.3 Summary

The most appropriate values of the autocovariance distances, \( r_0 \), and the ratio of spatial to total variance, \( b \), can be determined from a parametric study if data exist over a wide range of separation distances. The absence of data separated by less than the autocovariance distance prevents definitive estimation of \( b \), although an appropriate value of \( r_0 \) can be determined. The \( r_0 \) value which produces the minimum variance between the observed and estimated soil properties can be estimated independently of \( b \) and will be a reasonable estimator of the true autocovariance distance of the soil property.

The data of this study suggest an \( r_0 \) of 35m and that 50 percent of the variance in observed average blow count is caused by random measurement error. Therefore, Figure 5-3b is the appropriate map of estimated depth-average blow count for the site, and Figure 5-5b is representative of the spatial estimation variance. A measurement variance of \( 23(bpf)^2 \) must be added to the values of the figure to obtain the total estimation variance.
<table>
<thead>
<tr>
<th>( r_0 ) (m)</th>
<th>Mean[( \varepsilon )] (bpf)</th>
<th>Var[( \varepsilon )] (bpf)</th>
<th>( b = \sigma_s^2 / \sigma^2 )</th>
<th>Mean[E]</th>
<th>Var[E]</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.00</td>
<td>46.6</td>
<td>0.25</td>
<td>0.01</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.50</td>
<td>0.01</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.01</td>
<td>1.03</td>
</tr>
<tr>
<td>25</td>
<td>-0.06</td>
<td>42.2</td>
<td>0.50</td>
<td>-0.01</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
<td>-0.01</td>
<td>0.94</td>
</tr>
<tr>
<td>35</td>
<td>-0.12</td>
<td>40.9</td>
<td>0.50</td>
<td>-0.01</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
<td>-0.01</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>-0.01</td>
<td>0.98</td>
</tr>
<tr>
<td>50</td>
<td>-0.02</td>
<td>43.6</td>
<td>0.25</td>
<td>-0.03</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.50</td>
<td>-0.02</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
<td>-0.02</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.75</td>
<td>-0.02</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>-0.02</td>
<td>1.11</td>
</tr>
<tr>
<td>100</td>
<td>-0.28</td>
<td>45.3</td>
<td>0.50</td>
<td>-0.04</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
<td>-0.04</td>
<td>1.41</td>
</tr>
</tbody>
</table>
FIGURE 5-1 Plan Location of Depth-average Blow Count Observations
FIGURE 5-3a. Contour Map of BLUE Estimators for $r_0 = 25\text{m}$
FIGURE 5-3a. Contour Map of BLUE Estimators for $\sigma = 35m$
FIGURE 5-3c. Contour Map of BLUE Estimators for $c = 50m$
FIGURE 5-3d. Contour Map of BLUE Estimators for 6 = 100m
FIGURE 5-4  The Effect of $r_0$ on BLUE Estimators along Section 200
FIGURE 5-5a. Contour Map of Spatial Estimation Variance for $c_0 = 25$ m and $b = 0.50$
FIGURE 5-5b. Contour Map of Spatial Estimation Variance for $r_0 = 3.5m$ and $b = 0.5$
FIGURE 5-5c. Contour Map of Spatial Estimation Variance for $a = 50 \text{ m}$ and $b = 0.50$
FIGURE 5-6 Effect of $\gamma_0$ on the Total Estimation Variance along Section 200 for $b=0.50$
FIGURE 5-7 Spatial Estimation Variance Along Section 200 for 59 m < y < 159 m
FIGURE 5-8  Effect of $b = \sigma_s^2 / \sigma^2$ on the Total Estimation Variance along Section 200 for $r_0 = 50$ m
FIGURE 5-9  Variance of Observed Estimation Error versus Autocovariance Distance
Figure 5-102. Normal Probability Plot of the Error Ratio

$g_c = 25m$; $b = 0.50$ and 0.60
**FIGURE 5-10b. Normal Probability Plot of the Error Ratio**

$r_0 = 35m; \ b = 0.50, 0.65 \text{ and } 1.00$
FIGURE 5-10c. Normal Probability Plot of the Error Ratio
\( r_o = 50 \text{m}; b = 0.50, 0.65 \text{ and } 1.00 \)
FIGURE 5-10d. Normal Probability Plot of the Error Ratio
$R_0 = 100m$; $b = 0.50$ and $0.65$
**FIGURE 5-11** Normal Probability Plot of the Error Ratio on Independent Observations.
FIGURE 5-12 Variance of the Error Ratio versus $r_0$ and $b$
FIGURE 5-13 Variance of the Error Ratio versus Autocovariance Distance
VI. CONCLUSION

A minimum-variance linear unbiased estimation technique, founded upon statistical correlation among data, has been presented. The technique is useful for site characterization but is no panacea. The technique provides improved estimators, but it requires data from an extensive exploration program. The program should be preferably designed with consideration of the technique so that parameters of the autocorrelation function may be determined, either by autocorrelation analyses or by back-figuring from the BLUE estimators. Nevertheless, even rough approximations of the autocorrelation parameters are apparently adequate for calculation of the estimators. When data is abundant, the statistical estimation technique provides a rational and simple method of interpolating geotechnical data.
VII. REFERENCES


