Mathematical Geology, Vol. 17, No. 5, 1985

Fitting Variogram Models by Weighted Least Squares[1]

Noel Cressie[2]

The method of weighted least squares is shown to be an appropriate way of fitting variogram models. The weighting scheme automatically gives most weight to early lags and down-weights those lags with a small number of pairs. Although weights are derived assuming the data are Gaussian (normal), they are shown to be still appropriate in the setting where data are a (smooth) transform of the Gaussian case. The method of (iterated) generalized least squares, which takes into account correlation between variogram estimators at different lags, offer more statistical efficiency at the price of more complexity. Weighted least squares for the robust estimator, based on square root differences, is less of a compromise.

KEY WORDS: generalized least squares, kriging, median polish, robustness, stationarity.

1. INTRODUCTION

Any geostatistical study should ideally involve many different areas of expertise. In mining applications, the team should include at least a geologist, a mining engineer, a metallurgist, a financial manager, and a statistician. This article is written from a statistician's point of view highlighting that role in the study; the broader perspective can be gained by reading King, McMahon, and Bujtor (1982). The statistician can typically be expected to lead the team through the following stages:

1. Graphing and summarizing data.
2. Detecting and allowing for nonstationarity.
3. Estimating spatial relationships, usually by estimating the variogram or covariogram.
4. Estimating *in situ* resources, often by kriging.
5. Assessing recoverable reserves.

Cressie (1984) presents a resistant approach (i.e., using techniques not affected by a small proportion of outlying or aberrant values) to stages 1 and 2,

[1]Manuscript received 21 November 1983; accepted 27 December 1984.
[2]Department of Statistics, Iowa State University, Ames, Iowa 50011 U.S.A.

0020-5958/85/0700-0563$04.50/0

and shows techniques of exploratory data analysis (EDA) to be adaptable to spatial data. Robust estimation of the variogram in the presence of contaminated data is already discussed at some length by Cressie (1979), Cressie and Hawkins (1980), Armstrong (1984), Hawkins and Cressie (1984), and Switzer (1984). Here we mainly address the problem of fitting a model to various variogram estimators, both classical and robust. Until now, fitting procedures have either been "by eye," by ad hoc methods particular to the model being fitted, or by least squares. These approaches will be improved by using *statistical criteria to weight the influence* of various parts of the estimator.

In order to invoke a certain amount of statistical rigor, we proceed directly to assumptions of our model. We assume throughout that the intrinsic hypothesis holds: Suppose that the grade of an ore body D at a point t (in general in $\mathbb{R}^3$, but for our purposes here in $\mathbb{R}^2$) is the realization of a random function $\{Z_t; t \in D\}$ and that this is observed at certain points $\{t_i\}$ (often a regular grid) of the ore body. Then the so-called *intrinsic hypothesis* assumes that for h, a vector in $\mathbb{R}^2$

$$\begin{aligned} E(Z_{t+h} - Z_t) &= 0 \\ E(Z_{t+h} - Z_t)^2 &= 2\gamma(h) \end{aligned} \tag{1}$$

This is almost second-order stationarity of first differences. The quantity $2\gamma(h)$ is known as the variogram and is the crucial parameter of geostatistics; see Matheron (1963) and Journel and Huijbregts (1978). It is a more general model than that of second-order stationarity of $\{Z_t\}$, but when the latter is appropriate

$$\operatorname{cov}(Z_t, Z_{t+h}) = C(h) \tag{2}$$

and $\gamma(h) = C(0) - C(h)$.

When data are nonstationary in the drift $d(t) = E(Z_t)$, Starks and Fang (1982b) show how naive attempts to estimate the variogram yield a substantial bias. If one thinks drift can be expressed as a polynomial in t with known order, then the technique based on the generalized covariance function (Delfiner, 1976) does allow unbiased inference. However it is a complicated procedure to implement (Starks and Fang, 1982a), and if one guesses the order of the polynomial wrongly, one is faced with exactly the same bias problems as with the variogram. A straightforward approach to the problem of nonstationarity in the mean is taken by Cressie (1984); there, *resistant* techniques are used to estimate drift. These are shown to ameliorate the bias problem (Cressie and Glonek, 1984); hence residuals from this resistant fit are used to estimate the variogram. So, if data are nonstationary in the mean, now an easy-to-apply method exists to reduce the problem to one involving mean stationarity. Nonresistant fitting of low-order polynomials in disjoint regions (Buxton, 1982) is another simple technique often employed, but its drawback is that the residual bias problem is still present. The classical estimator of the variogram based on data $\{Z_{t_i}, i = 1, \cdots, n\}$ is (Matheron, 1963)

$$2\hat{\gamma}(h) = \frac{1}{N_h} \sum_{i=1}^{N_h} (Z_{t_i+h} - Z_{t_i})^2 \qquad h = h(1), h(2), \cdots \tag{3}$$

where N_h is the number of lag-h differences. In order to use available knowledge of robust *location* estimation, Cressie and Hawkins (1980) take fourth roots of squared differences, yielding robust (to contamination by outliers; see Hawkins and Cressie, 1984) estimators

$$2\bar{\gamma}(h) = \left(\frac{1}{N_h} \sum_{i=1}^{N_h} |Z_{t_i+h} - Z_{t_i}|^{1/2}\right)^4 \Big/ (0.457 + 0.494/N_h) \tag{4}$$

$$2\tilde{\gamma}(h) = (\text{med}\,\{|Z_{t_i+h} - Z_{t_i}|^{1/2}\})^4/B_h \tag{5}$$

where med $\{\cdot\}$ denotes median of the sequence $\{\cdot\}$, and B_h corrects for bias.

We could consider other robust estimators proposed by Armstrong and Delfiner (1980), but the following argument shows them to be asymptotically equivalent to (5). Armstrong and Delfiner followed up Cressie and Hawkins' idea that the variogram is simply var $(Z_{t+h} - Z_t)$ under the intrinsic hypothesis, so that a "robustification" could be made by using a robust estimator of scale on the differences $\{(Z_{t_i+h} - Z_{t_i}); i = 1, \cdots, N_h\}$. Provided stationarity holds, $(Z_{t+h} - Z_t)$ is a symmetric random variable with mean 0, which ameliorates the scale estimation problem considerably. They defined "Huberized" variograms (i.e., using the scale estimator of Huber, 1964, rather than the sample variance of $\{(Z_{t_i+h} - Z_{t_i})\}$), and quantile variograms. Huberized variograms are lengthy to compute, requiring iteration at every lag. The square of the interquartile range of differences, however, is a resistant, quick, and easy alternative; consider then

$$(\text{UQ}\{Z_{t_i+h} - Z_{t_i}\} - LQ\{Z_{t_i+h} - Z_{t_i}\})^2$$

where UQ stands for "upper quartile" and LQ for "lower quartile." Furthermore, the quantile variogram is based on a sample quantile of $\{(Z_{t_i+h} - Z_{t_i})^2\}$, the most popular choice being the median. The idea is that quantiles are more resistant to outliers than the mean; consider then

$$\text{med}\,\{(Z_{t_i+h} - Z_{t_i})^2\}$$

Both of the above approaches need some normalization to make them unbiased; however, leaving this aside, we show that both are *equivalent* to $2\tilde{\gamma}(h)$ in (5), the fourth root type estimator based on the median. Now med $\{(Z_{t_i+h} - Z_{t_i})^2\} = (\text{med}\,\{|Z_{t_i+h} - Z_{t_i}|^{1/2}\})^4$, because $f(x) = x^{1/4}$ is a monotonic function. Also, asymptotically, $\text{UQ}\{Z_{t_i+h} - Z_{t_i}\} - \text{LQ}\{Z_{t_i+h} - Z_{t_i}\} = 2\ \text{med}\,\{|Z_{t_i+h} - Z_{t_i}|\} = 2(\text{med}\,\{|Z_{t_i+h} - Z_{t_i}|^{1/2}\})^2$. Hence the estimator $2\tilde{\gamma}(h)$, based on fourth roots of squared differences, simultaneously captures the essence of a robust scale approach and a quantile approach.

The next section shows that present methods of variogram fitting are inadequate. Under suitable asymptotics, using the criterion of weighted least squares will improve the fit. The method automatically gives most weight to early lags, and downweights those lags with a small number of contributing pairs.

2. WEIGHTED LEAST-SQUARES FITTING

The variogram $\{2\gamma(h)\}$, defined in (1), is a function of h that is typically estimated at discrete lags: $h(1), h(2), \cdots, h(k)$; for example, for data on a rectangular grid, and for a fixed direction of the grid, $h(j) = j; j = 1, 2, \cdots$, in units of the grid spacing. Through these estimated values, a variogram model (such as spherical, exponential, Gaussian, de Wijsian, linear, etc.), which typically depends on several parameters, is fitted. It is the method of fitting that is the subject of this section.

Up to now, variogram fitting procedures have been either "by eye," ad hoc methods particular to the model being fitted, or by least squares through the points $\{(h(j), 2\hat{\gamma}(h(j))); j = 1, 2, \cdots, k\}$ (David, 1977, Sect. 6.1, 6.2; Journel and Huijbregts, 1978, Sect. III.C.6, Chap. IV; Clark, 1979, Chap. 2). What we would like to do here is present a general fitting procedure which gleans the correct features from current practice, discards incorrect features, and produces a statistical rationale for an overall approach. For example, some variograms have a sill parameter, which in turn appears as a multiplicative factor in prediction (kriging) variances. Currently (David, 1977, p. 122; Journel and Huijbregts, 1978, p. 231; Clark, 1979, p. 29), we are told that if $\{Z_t\}$ is stationary and mixing (i.e., weak dependence at large lags) and hence $\sigma^2 = \text{var}(Z_t) = \lim_{h \to \infty} \gamma(h)$,

$$\frac{1}{N-1} \sum_{i=1}^{N} (Z_{t_i} - \bar{Z})^2 = \hat{\sigma}^2$$

is a good estimator of σ^2. Yet in the same body of theory, we are told that should nonstationarity in the mean $E(Z_t) = d(t)$ exist, then the variogram estimator based on residuals $\{Z_{t_i} - \hat{d}(t_i)\}$ will be intolerably biased in estimating the variogram of the errors, $E[Z_{t+h} - d(t+h) - Z_t + d(t)]^2$ (Matheron, 1971, p. 196). But in exactly the same way as for the nonstationary case, residuals $\{Z_{t_i} - \bar{Z}\}$ in the stationary case produce a *biased* estimator of σ^2. Suppose the points $\{t_i\}$ are equally spaced on a transect (e.g., $\{Z_{t_i}\}$ is a time series); then

$$E(\hat{\sigma}^2) = \sigma^2 - \frac{2}{N-1} \sum_{h=1}^{N} [1 - (h/N)]\, C(h)$$

which *always* exhibits a negative bias when $C > 0$. Recent results of Cressie and Glonek (1984) indicate that this situation can be ameliorated by choosing a resistant quantity to estimate the constant mean, such as med $\{Z_{t_i}\}$ instead of $\bar{Z} \equiv \sum Z_{t_i}/N$.

Furthermore, various ad hoc ways of obtaining the slope of the variogram at the origin, the nugget effect, the range, and so forth, are unsatisfactory in that the statistical fluctuations of a variogram estimator $\{2\gamma^*(h(j)); j = 1, \cdots, k\}$ are not taken into account when fitting a model $\{2\gamma(h; \boldsymbol{\lambda})\}$. This could, but usually does not, lead to serious errors (in the first instance statistical, but eventually financial) because the geostatistician usually returns to look at the fit plotted against the estimate, to "eyeball" it and adjust maverick parts of the fit accordingly. Moreover, recent interesting results by Diamond and Armstrong (1983) show the prediction (kriging) stage of the analysis to be reasonably insensitive to the variogram chosen. This, however, should not stop us from trying to make the best of the data $\{Z_{t_i}\}$, to estimate the variogram (robustly), and to fit a model (efficiently), free of unconscious biases.

The method of least squares is not statistical; it is purely a numerical criterion used to find "the most appropriate" parameter values. In our context, suppose $\{2\gamma(h; \boldsymbol{\lambda})\}$ is a variogram model depending on parameters $\boldsymbol{\lambda}$. Then, the method of *least squares* says to choose the value of $\boldsymbol{\lambda}$ which minimizes

$$\sum_{j=1}^{k} \{2\gamma^*(h(j)) - 2\gamma(h(j); \boldsymbol{\lambda})\}^2 \tag{6}$$

call it $\boldsymbol{\lambda}_I^*$. However, when $2\boldsymbol{\gamma}^* = \{2\gamma^*(h(1)), \cdots, 2\gamma^*(h(k))\}$ is a vector of random variables with variance matrix var $(2\boldsymbol{\gamma}^*) = \Sigma$, then the method of *generalized least squares* says to choose the value of $\boldsymbol{\lambda}$ which minimizes

$$[2\boldsymbol{\gamma}^* - 2\boldsymbol{\gamma}(\boldsymbol{\lambda})]'\Sigma^{-1}[2\boldsymbol{\gamma}^* - 2\boldsymbol{\gamma}(\boldsymbol{\lambda})] \tag{7}$$

and call it $\boldsymbol{\lambda}_\Sigma^*$.

Between $\boldsymbol{\lambda}_I^*$ and $\boldsymbol{\lambda}_\Sigma^*$ is an intermediate stage, the method of *weighted least squares*, which says to choose the value of $\boldsymbol{\lambda}$ which minimizes

$$\sum_{j=1}^{k} \{\text{var}\,[2\gamma^*(h(j))]\}^{-1} \{2\gamma^*(h(j)) - 2\gamma(h(j); \boldsymbol{\lambda})\}^2 \tag{8}$$

and call it $\boldsymbol{\lambda}_V^*$, where $V = \text{diag}\,\{\text{var}\,[2\gamma^*(h(1))], \cdots, \text{var}\,[2\gamma^*(h(k))]\}$ is a diagonal matrix with zero's everywhere except for variances of $2\gamma^*(h(j))$ on the diagonal. Notice that we have *not* considered the maximum likelihood estimator of $\boldsymbol{\lambda}$, because it is strongly model dependent; also Carroll and Ruppert (1982) show the generalized least-squares estimator of $\boldsymbol{\lambda}$ to possess superior robustness to misspecification of error structure.

Under appropriate asymptotics, use of $\boldsymbol{\lambda}_V^*$, *weighted least-squares estimator*, is shown to be a statistically sensible estimator of $\boldsymbol{\lambda}$. Furthermore, it can be used as an initial value in iterative generalized least squares.

2.1 Classical Estimator

Suppose $\{Z_t\}$ is Gaussian; that is, any finite linear combination of Z_t's has a Gaussian (normal) distribution; this assumption is relaxed later. Then the intrinsic hypothesis (1) implies that in distribution

$$(Z_{t+h} - Z_t)^2 = 2\gamma(h) \cdot \chi_1^2$$

where χ_1^2 denotes a chi-square random variable on 1 df. Cressie and Hawkins (1980) base a robust location estimator of $2\gamma(h)$ on this fact. Now

$$E[(Z_{t+h} - Z_t)^2] = 2\gamma(h)$$
$$\text{var}\,[(Z_{t+h} - Z_t)^2] = 2[2\gamma(h)]^2$$

$$\begin{aligned} &\text{corr}\,[(Z_{t+h(1)} - Z_t)^2, (Z_{s+h(2)} - Z_s)^2] \\ &= \{\text{corr}\,[(Z_{t+h(1)} - Z_t), (Z_{s+h(2)} - Z_s)]\}^2 \\ &= \left\{\frac{\gamma(|t-s+h(1)|) + \gamma(|t-s-h(2)| - \gamma(|t-s|) - \gamma(|t-s+h(1)-h(2)|)}{\{2\gamma(h(1))\}^{1/2}\{2\gamma(h(2))\}^{1/2}}\right\}^2 \end{aligned} \tag{9}$$

where "corr" denotes correlation. The last expression of (9) comes from an easily proven fact that if X_1, X_2 jointly are normal with zero means and corr $(X_1, X_2) = \rho$, then corr $(X_1^2, X_2^2) = \rho^2$.

Recall from (3) the formula for $\{2\hat{\gamma}(h); h = h(1), h(2), \cdots\}$. The contents of this subsection are, by necessity, rather technical. We make the following asymptotic assumptions:

Assumption A1: k is fixed (see subsection 2.3 for practical guidelines on the choice of k)

Assumption A2: $N_{h(j)} \to \infty$ for each $j = 1, \cdots, k$, as $N \to \infty$, and $N \to \infty$ as $|D| \to \infty$ such that $N/|D|$, the sampling rate per unit area, is constant.

Furthermore assume:

Assumption A3: $\gamma(h) = \sigma^2$, for $h > a$ (i.e., beyond the range a, random variables Z_t, Z_{t+h} are uncorrelated); or $\gamma(h) = c_1 h + c_2$, for $h > a$.

This last assumption includes many models which either are covariance stationary or satisfy the intrinsic hypothesis. Some small modifications to the expressions to follow will also take care of the exponential, Gaussian, and de Wijsian models, not covered, strictly speaking, by A3.

We want to find var $\{2\hat{\gamma}(h(j))\}$, and cov $\{2\hat{\gamma}(h(i)), 2\hat{\gamma}(h(j))\}$. From (9), and Cressie and Hawkins (1980)

$$\operatorname{var}\{2\hat{\gamma}(h(j))\} = \frac{2\{2\gamma(h(j))\}^2}{N_{h(j)}^2}\left\{N_{h(j)} + \sum_{l \neq m=1}^{N_{h(j)}} \sum_{m=1}^{N_{h(j)}} \cdot \left[\frac{\gamma(t_m - t_l - h(j)) + \gamma(t_m - t_l + h(j)) - 2\gamma(t_m - t_l)}{2\gamma(h(j))}\right]^2\right\} \tag{10}$$

where we adopt the convention that $\gamma(-h) = \gamma(h); h \geqslant 0$, and of course, by definition, $\gamma(0) = 0$. Equation (10) gives the *exact* expression for diagonal elements of Σ, but it is to say the least, unwieldy; off-diagonal elements are equally so

$$\operatorname{cov}\{2\hat{\gamma}(h(i)), 2\hat{\gamma}(h(j))\} = \frac{2\{2\gamma(h(i))\}\{2\gamma(h(j))\}}{N_{h(i)}N_{h(j)}}\left\{\sum_{l=1}^{N_{h(i)}} \sum_{m=1}^{N_{h(j)}} \cdot \left[\frac{\gamma(t_m - t_l + h(j)) + \gamma(t_m - t_l - h(i)) - \gamma(t_m - t_l) - \gamma(t_m - t_l + h(j) - h(i))}{\{2\gamma(h(i))\}^{1/2}\{2\gamma(h(j))\}^{1/2}}\right]^2\right\} \tag{11}$$

We need to make further assumptions to obtain some guidance from (10) and (11) namely

Assumption A4: $\{Z_{t_i}; i = 1, \cdots, N\}$ occur on a transect, and furthermore at equally spaced points. Write $t_i = i$, in units of the spacing.

Under A1 to A4, (10) becomes

$$\operatorname{var}[2\hat{\gamma}(j)] = \frac{2[2\gamma(j)]^2}{N_j}\left\{1 + 2\sum_{m=1}^{j+a}\left[\frac{\gamma(m+j) + \gamma(m-j) - 2\gamma(m)}{2\gamma(j)}\right]^2\right\} + 0\left(\frac{1}{N_j^2}\right) \tag{12}$$

and (11) becomes, for $i = j - 1; 2 \leqslant j \leqslant k$

$$\operatorname{cov}[2\hat{\gamma}(j-1), 2\hat{\gamma}(j)] = \frac{2[2\gamma(j-1)][2\gamma(j)]}{N_{j-1}} \cdot \left\{2\sum_{m=1}^{j+a}\left[\frac{\gamma(m+j-1) + \gamma(m-j) - \gamma(m) - \gamma(m-1)}{(2\gamma(j-1))^{1/2}(2\gamma(j))^{1/2}}\right]^2\right\} + 0\left(\frac{1}{N_{j-1}^2}\right)$$

for $i = j - 2$; $3 \leqslant j \leqslant k$

$$\operatorname{cov}[2\hat{\gamma}(j-2), 2\hat{\gamma}(j)] = \frac{2[2\gamma(j-2)][2\gamma(j)]}{N_{j-2}} \left\{ -\left[\frac{2\gamma(j-1) - 2\gamma(1)}{[2\gamma(j-2)]^{1/2}[2\gamma(j)]^{1/2}}\right]^2 + 2\sum_{m=1}^{j+a} \left[\frac{\gamma(m+j-2) + \gamma(m-j) - \gamma(m) - \gamma(m-2)}{[2\gamma(j-2)]^{1/2}[2\gamma(j)]^{1/2}}\right]^2 \right\} + 0\left(\frac{1}{N_{j-2}^2}\right)$$

and in general, for $l + 1 \leqslant j \leqslant k$

$$\operatorname{cov}[2\hat{\gamma}(j-l), 2\hat{\gamma}(j)] = \frac{2[2\gamma(j-l)][2\gamma(j)]}{N_{j-l}} \left\{ c(j,l) + 2\sum_{m=1}^{j+a} \left[\frac{\gamma(m+j-l) + \gamma(m-j) - \gamma(m) - \gamma(m-l)}{[2\gamma(j-l)]^{1/2}[2\gamma(j)]^{1/2}}\right]^2 \right\} + 0\left(\frac{1}{N_{j-l}^2}\right) \tag{13}$$

where (for $l + 1 \leqslant j \leqslant k$)

$$c(j,l) = \begin{cases} -\left[\dfrac{2\gamma(j-l/2) - 2\gamma(l/2)}{[2\gamma(j-l)]^{1/2}[2\gamma(j)]^{1/2}}\right]^2 & l = 2, 4, 6, \cdots, \\ 0 & l = 1, 3, 5, \cdots \end{cases}$$

In order to interpret these results, take the simple model

$$\gamma_f(h) = \begin{cases} 0 & h = 0 \\ f\sigma^2 & h = 1 \\ \sigma^2 & h = 2, 3, \cdots, \end{cases}$$

where $\frac{1}{4} < f \leqslant 1$, to ensure positive definiteness of $C(h) = \sigma^2 - \gamma(h)$. Note that $f = 1$ gives uncorrelated (independent in the Gaussian case) $\{Z_{t_i}\}$. Let us start with diagonal elements of $\Sigma = (\sigma_{ij})$; to leading order

$$\sigma_{11} = \frac{2[2\gamma_f(1)]^2}{N_1}\left[1 + \frac{(5f^2 - 6f + 2)}{2f^2}\right]$$

$$\sigma_{22} = \frac{2[2\gamma_f(2)]^2}{N_2}\left[1 + \frac{(2f^2 - 4f + 3)}{2}\right]$$

$$\sigma_{jj} = \frac{2[2\gamma_f(j)]^2}{N_j}\left[1 + \frac{(6f^2 - 12f + 7)}{2}\right] \qquad j = 3, \cdots, k$$

Now every pairwise ratio of the terms in large brackets belongs to $[\frac{1}{4}, 4]$ and, hence, for this model we retain enough statistical efficiency (Cressie, 1980) to work with

$$\mathrm{var}\,[2\hat{\gamma}(j)] \simeq 2[2\gamma(j;\boldsymbol{\lambda})]^2/N_j \tag{14}$$

in (8), giving the approximate weighted least-squares estimate $\hat{\boldsymbol{\lambda}}_V$, obtained by minimizing

$$\sum_{j=1}^{k} N_{h(j)} \left\{ \frac{\hat{\gamma}(h(j))}{\gamma(h(j);\boldsymbol{\lambda})} - 1 \right\}^2 \tag{15}$$

Minimizing (15) is a vast improvement over least squares, although more efficient estimators yet can be obtained by iterating (see subsection 2.3).

One might hope for off-diagonal elements of Σ to be negligible, but this is not the case, as the simple model illustrates; to leading order

$$\sigma_{j-1,j} = \frac{2[2\gamma_f(j-1)]\,[2\gamma_f(j)]}{N_{j-1}} \left(\frac{4f^2 - 10f + 8}{2}\right) \qquad j = 4, \cdots, k$$

$$\sigma_{j-2,j} = \frac{2[2\gamma_f(j-2)]\,[2\gamma_f(j)]}{N_{j-2}} \left(\frac{5f^2 - 10f + 7}{2}\right) \qquad j = 5, \cdots, k$$

Therefore, even when data are uncorrelated, which corresponds to $f = 1$ in the simple model, Σ has typical diagonal term $\sigma_{jj} = 2[2\gamma_1(j)]^2\ \{\frac{3}{2}\}/N_j$, and typical off-diagonal term $\sigma_{j-l,j} = 2[2\gamma_1(j-l)]\ \ [2\gamma_1(j)]\ \ \{1\}/N_{j-l}$; thus $\sigma_{j-l,j}/\sigma_{jj} = \{\frac{2}{3}\}\ [\gamma_1(j-l)/\gamma_1(j)]\ (N_j/N_{j-l})$, nonnegligible for $l = 1, 2, 3$.

We will see that this situation is ameliorated considerably by considering the robust estimator based on square root differences (Cressie and Hawkins, 1980).

2.2 Robust Estimator

Recall from (4) the formula for $2\bar{\gamma}(h)$; $h = h(1), h(2), \cdots$, based on $\{|Z_{t_i+h} - Z_{t_i}|^{1/2}\}$. Consider for the moment the quantity

$$\bar{A}(h) \equiv \frac{1}{N_h} \sum_{i=1}^{N_h} |Z_{t_i+h} - Z_{t_i}|^{1/2} \qquad h = h(1), h(2), \cdots,$$

whose variance matrix we wish to find.

Under the assumption that $\{Z_t\}$ is Gaussian, which is relaxed later, the intrinsic hypothesis (1) implies that, in distribution,

$$|Z_{t+h} - Z_t|^{1/2} = [2\gamma(h)]^{1/4} \cdot (\chi_1^2)^{1/4}$$

Then,

$$
\begin{aligned}
E[|Z_{t+h} - Z_t|^{1/2}] &= [2^{1/4}\Gamma(\tfrac{3}{4})/\pi^{1/2}]\,[2\gamma(h)]^{1/4} \\
\operatorname{var}[|Z_{t+h} - Z_t|^{1/2}] &= 2^{1/2}[\pi^{-1/2} - \Gamma^2(\tfrac{3}{4})/\pi]\,[2\gamma(h)]^{1/2} \\
\operatorname{corr}[|Z_{t+h(1)} - Z_t|^{1/2}, &|Z_{s+h(2)} - Z_s|^{1/2}] \\
&= \phi(\operatorname{corr}[(Z_{t+h(1)} - Z_t), (Z_{s+h(2)} - Z_s)])
\end{aligned}
\tag{16}
$$

These results parallel those of (9); mean and variance come from Cressie and Hawkins (1980) whereas the correlation is not so straightforward. If X_1, X_2 are jointly normal with zero means and corr $(X_1, X_2) = \rho$, then tedious algebra yields

$$
\operatorname{corr}(|X_1|^{1/2}, |X_2|^{1/2}) = \phi(\rho) \equiv \frac{\Gamma^2(\frac{3}{4})}{\pi^{1/2} - \Gamma^2(\frac{3}{4})}[(1-\rho^2)\,{}_2F_1(\tfrac{3}{4}, \tfrac{3}{4}; \tfrac{1}{2}; \rho^2) - 1]
\tag{17}
$$

where

$$
{}_2F_1(a, b; c; z) = 1 + \frac{ab}{c}z + \frac{a(a+1)\,b(b+1)}{c(c+1)}\frac{z^2}{2!} + \cdots +
$$

is the hypergeometric function. Thus for ρ small, $\phi(\rho) \simeq (\frac{5}{8})\,\rho^2$, which should be compared with corr $(X_1^2, X_2^2) = \rho^2$; this correlation attenuation is an added bonus to those who estimate the variogram with the robust estimator (4). We would like at this point, to acknowledge D. M. Hawkins with whom we collaborated to obtain this result.

We want to find $\tau(h(i), h(j)) \equiv \operatorname{cov}\{\bar{A}(h(i)), \bar{A}(h(j))\}$. From (16) and Cressie and Hawkins (1980)

$$
\tau(h(j), h(j)) = \frac{2^{1/2}[\pi^{-1/2} - \Gamma^2(\frac{3}{4})/\pi]\{2\gamma(h(j))\}^{1/2}}{N_{h(j)}^2} \cdot \left\{N_{h(j)} + \sum_{l \neq m=1}^{N_{h(j)}} \sum_{m=1}^{N_{h(j)}} \phi\left[\frac{\gamma(t_m - t_l - h(j)) + \gamma(t_m - t_l + h(j)) - 2\gamma(t_m - t_l)}{2\gamma(h(j))}\right]\right\}
$$

$$
\tau(h(i), h(j)) = \frac{2^{1/2}[\pi^{-1/2} - \Gamma^2(\frac{3}{4})/\pi]\;\{2\gamma(h(i))\}^{1/4}\{2\gamma(h(j))\}^{1/4}}{N_{h(i)}N_{h(j)}} \left\{\sum_{l=1}^{N_{h(i)}} \sum_{m=1}^{N_{h(j)}} \phi\left[\frac{\gamma(t_m - t_l + h(j)) + \gamma(t_m - t_l - h(i)) - \gamma(t_m - t_l) - \gamma(t_m - t_l + h(j) - h(i))}{\{2\gamma(h(i))\}^{1/2}\{2\gamma(h(j))\}^{1/2}}\right]\right\}
$$

Now it is a simple matter to show that for a (continuously differentiable) smooth function g, $\operatorname{cov}[g(X), g(Y)] \simeq g'[E(X)]\ g'[E(Y)]\ \operatorname{cov}(X, Y)$, and because $2\bar{\gamma}(h) = [\bar{A}(h)]^4/(0.457 + 0.494/N_h)$

$$\Omega \equiv \operatorname{cov}\{2\bar{\gamma}(h(i)), 2\bar{\gamma}(h(j))\}$$

$$\simeq 4^2[2^{1/4}\Gamma(\tfrac{3}{4})/\pi^{1/2}]^6 \left\{ \frac{\{2\gamma(h(i))\}^{3/4}}{0.457 + 0.494/N_{h(i)}} \cdot \frac{\{2\gamma(h(j))\}^{3/4}}{0.457 + 0.494/N_{h(j)}} \cdot \tau(h(i), h(j)) \right\} \tag{18}$$

This expression for variances and covariances of the robust estimator is the analogue of (10) and (11) (variances and covariances for the classical estimator). Analogous simplifications under assumptions A1 to A4 can be made, and we find, to the leading order of magnitude retained in (12) and (13)

$$\operatorname{var}[2\bar{\gamma}(j)] = \frac{2.885[2\gamma(j)]^2}{N_j} \left\{ 1 + 2\sum_{m=1}^{j+a} \phi\left[\frac{\gamma(m+j) + \gamma(m-j) - 2\gamma(m)}{2\gamma(j)}\right]\right\} \tag{19}$$

and for $l + 1 \leqslant j \leqslant k$

$$\operatorname{cov}[2\bar{\gamma}(j-l), 2\bar{\gamma}(j)]$$

$$= \frac{2.885[2\gamma(j-l)][2\gamma(j)]}{N_{j-l}} \left\{ d(j,l) + 2\sum_{m=1}^{j+a} \phi\left[\frac{\gamma(m+j-l) + \gamma(m-j) - \gamma(m) - \gamma(m-l)}{[2\gamma(j-l)]^{1/2}[2\gamma(j)]^{1/2}}\right]\right\} \tag{20}$$

where (for $l + 1 \leqslant j \leqslant k$)

$$d(j,l) = \begin{cases} -\phi\left[\dfrac{2\gamma(j-l/2) - 2\gamma(l/2)}{[2\gamma(j-l)]^{1/2}[2\gamma(j)]^{1/2}}\right] & l = 2, 4, 6, \cdots, \\ 0 & l = 1, 3, 5, \cdots, \end{cases}$$

Notice that (19) and (20), variances and covariances of the robust estimator, differ from those of the Matheron estimator (12 and 13), most importantly through correlation terms involving, respectively, $\phi(\cdot)$ (given by 17) and $(\cdot)^2$.

Take the simple model $\gamma_f(h)$ of the previous subsection, and use the approximation $\phi(\rho) \simeq (\frac{5}{8})\rho^2$. Elements of $\Omega = (\omega_{ij})$ are, to leading order

$$\omega_{11} = \frac{2.885[2\gamma_f(1)]^2}{N_1}\left[1 + \frac{(\frac{5}{8})(5f^2 - 6f + 2)}{2f^2}\right]$$

$$\omega_{22} = \frac{2.885[2\gamma_f(2)]^2}{N_2}\left[1 + \frac{(\frac{5}{8})(2f^2 - 4f + 3)}{2}\right]$$

$$\omega_{jj} = \frac{2.885[2\gamma_f(j)]^2}{N_j}\left[1 + \frac{(\frac{5}{8})(6f^2 - 12f + 7)}{2}\right] \qquad j = 3, \cdots, k$$

$$\omega_{j-1,j} = \frac{2.885[2\gamma_f(j-1)][2\gamma_f(j)]}{N_{j-1}}\left[\frac{(\frac{5}{8})(4f^2 - 10f + 8)}{2}\right] \qquad j = 4, \cdots, k$$

and so forth. When $f = 1$ (data uncorrelated), Ω has typical diagonal term $\omega_{jj} = 2.885[2\gamma_1(j)]^2\{\frac{21}{16}\}/N_j$, and typical off-diagonal term $\omega_{j-l,j} = 2.885[2\gamma_1(j-l)]\,[2\gamma_1(j)]\,\{\frac{5}{8}\}/N_{j-l}$. Then $\omega_{j-l,j}/\omega_{jj} = \{\frac{10}{21}\}\,[\gamma_1(j-l)/\gamma_1(j)]\,(N_j/N_{j-l})$, which is *30% smaller* than the corresponding expression $\sigma_{j-l,j}/\sigma_{jj}$ for the Matheron estimator.

The correlation attenuation of the robust estimator (4) means that the approximate weighted least-squares estimate $\bar{\boldsymbol{\lambda}}_V$, obtained by minimizing

$$\sum_{j=1}^{k} N_{h(j)}\left\{\frac{\bar{\gamma}(h(j))}{\gamma(h(j);\boldsymbol{\lambda})} - 1\right\}^2 \tag{21}$$

is statistically more efficient than $\hat{\boldsymbol{\lambda}}_V$, obtained by minimizing (15). More efficient estimators yet can be obtained by iterating (see subsection 2.3).

2.3 Weighted Least Squares and Generalized Least Squares

The generalized least-squares estimator obtained by minimizing (7), is statistically more efficient than the weighted least-squares estimator from (8). Under the asympototics A1, A2 above, both yield consistent estimates.

We have shown (subsections 2.1, 2.2) that minimizing (15) or (21) yields an approximate weighted least-squares estimate, although the latter, based on the robust estimator (4), is more efficient. Tractability of (15) and (21) make them attractive to work with, whereas minimizing (7), with Σ given by (10) and (11) or Ω given by (18), is forbidding; besides, Σ or Ω themselves depend on variogram parameters $\boldsymbol{\lambda}$. We suggest iteration as a way to resolve this impasse.

For example, suppose that we use the Matheron estimator (3), and that model parameters $\hat{\boldsymbol{\lambda}}_V$ are obtained by minimizing (15). Then substitute $2\gamma(h(j);\hat{\boldsymbol{\lambda}}_V)$ into (12) and (13) (or, more exactly, into 10 and 11) to obtain $\hat{\Sigma}$. The next stage of the iteration is to minimize (7) using $\Sigma = \hat{\Sigma}$. This new set of estimates of $\boldsymbol{\lambda}$ obtained can be used in the same way in (12) and (13), to obtain an updated estimate of Σ, which is in turn used in the minimization of (7), and so forth.

Under asymptotics A1, A2, a trivial generalization of Davis and Borgman (1982) shows that $\{2\hat{\gamma}(h(j));\ j = 1, \cdots, k\}$ and $\{2\bar{\gamma}(h(j);\ j = 1, \cdots, k\}$ are jointly normal. This justifies the iterated generalized least-squares approach, or its less efficient cousin, weighted least squares, as being sensible procedures.

When the data are not regularly spaced, some of the $h(j)$'s may be close together, and hence it will make a big difference to the estimators whether (iterated) generalized least squares or weighted least squares is used. This is *not* the case for the examples treated in Section 3.

Although all of the above was derived for the stationary Gaussian distribution, in fact all that was needed was $(Z_{t+h} - Z_t)^2 = 2\gamma(h) \cdot W$, where W is a unit mean random variable whose variance does not depend on h. That this happens on many scales, even those which are not normal, is witnessed by the following approximation (g is a continuous function differentiable in a neighborhood of μ)

$$\begin{aligned}[g(Z_{t+h}) - g(Z_t)]^2 &= \{[\mu + (Z_{t+h} - \mu) g'(\mu) + \cdots] \\ &\qquad - [\mu + (Z_t - \mu) g'(\mu) + \cdots]\}^2 \\ &\simeq [g'(\mu)]^2 (Z_{t+h} - Z_t)^2\end{aligned}$$

Hence these fitting procedures possess a robustness to a change of scale.

An important practical consideration is the choice of k. Let $H = \max\{h: N_h > 0\}$ denote the largest *possible* lag to be considered in the fit. Then Journel and Huijbregts (1978, p. 194) have the following "practical rule"

$$\text{Fit only up to lags } h \text{ for which } N_h > 30, \text{ and } 0 < k \leqslant H/2 \tag{22}$$

This guide is useful, although at times other considerations, such as when it is known the kriging equations will not make use of the variogram beyond a certain lag, need be taken into account. Currently, variogram estimates at large lags tend to over-influence the various ad hoc fitting procedures being used.

In summary then, for a fixed k and variogram estimator $2\gamma^*(\cdot)$, the (approximate) weighted least-squares procedure is to minimize with respect to $\boldsymbol{\lambda}$

$$\sum_{j=1}^{k} N_{h(j)} \left\{ \frac{\gamma^*(h(j))}{\gamma(h(j);\boldsymbol{\lambda})} - 1 \right\}^2 \tag{23}$$

The estimator could be either used as an improvement over least squares, or could itself be the starting value of an iterative generalized least-squares approach.

3. VARIOGRAM MODEL FITTING

In this section we estimate variogram parameters (e.g., sill, nugget effect, range, etc.) by minimizing (23). We now give several variogram models that will be applied to coal ash and iron ore data.

3.1 Spherical model

$$\gamma(h;\boldsymbol{\lambda}) = \begin{cases} c_0 + c_s[(\tfrac{3}{2})(h/a_s) - (\tfrac{1}{2})(h/a_s)^3] & 0 < h \leqslant a_s \\ c_0 + c_s & h \geqslant a_s \end{cases} \tag{24}$$

where $\boldsymbol{\lambda} = (c_0, c_s, a_s)$ is the vector of parameters to be estimated; c_0 is the nugget effect, $c_0 + c_s$ is the sill, and a_s is the range.

3.2 Exponential Model

$$\gamma(h; \boldsymbol{\lambda}) = c_0 + c_e[1 - \exp(-h/a_e)] \qquad h > 0 \tag{25}$$

where $\boldsymbol{\lambda} = (c_0, c_e, a_e)$ is the vector of parameters.

3.3 Linear Variogram

$$\gamma(h; \boldsymbol{\lambda}) = c_0 + b_l h \qquad h > 0 \tag{26}$$

where $\boldsymbol{\lambda} = (c_0, b_l)$ is the vector of parameters. Other models are found in Journel and Huijbregts (1978, p. 61ff).

The sperical model is not linear in its parameters; that is, we cannot write $\boldsymbol{\gamma}(\boldsymbol{\lambda}) = X\boldsymbol{\lambda}$ for some matrix X, and it is not differentiable in its parameters. It increases toward a sill, and so the covariogram given by (2) exists. The exponential model is not linear in its parameters, but it is differentiable everywhere. It increases toward a sill, and so the covariogram exists. The linear model is linear and differentiable in its parameters. It increases without bound, and so no covariogram exists. The combination of peculiarities of each model provides a good cross-section of the type of problems that arise when fitting.

Probably the most difficult model to fit using the method of weighted least-squares (i.e., by minimizing 23) is the spherical model of (24). We give some computational details for this case; let

$$f(h; c_0, c_s, a_s) = \sum_{h=1}^{[a_s]} N_h \left\{ \frac{\bar{\gamma}(h)}{c_0 + c_s[(\frac{3}{2})(h/a_s) - \frac{1}{2}(h/a_s)^3]} - 1 \right\}^2 + \sum_{h=[a_s]+1}^{k} N_h \left[\frac{\bar{\gamma}(h)}{c_0 + c_s} - 1 \right]^2$$

For a_s fixed, in particular $a_s = l$, l integer, a minimum with respect to c_0, c_s can be found by setting $\partial f/\partial c_0 = 0$, and $\partial f/\partial c_s = 0$. And for $a_s \in (l, l+1)$, f is differentiable with respect to a_s. Therefore the minimization can be done progressively. At the lth "node point," minimizing values of c_0 and c_s can be obtained and the appropriate f evaluated. Then by differentiation, we see if a stationary point of f occurs when $a_s \in (l, l+1)$. If not, proceed to the $(l+1)$st node point and repeat. Minimizing (23) for exponential and linear models is relatively straightforward.

Two data sets are used to illustrate the *weighted least-squares fitting procedure.* The first set is coal-ash measurements obtained from Gomez and Hazen (1970, Tables 19 and 20) for the Robena Mine Property in Greene County, Pennsylvania; a mostly conventional geostatistical analysis can be found in Buxton (1982). Resistant techniques for graphing and summarizing these data, and detecting and allowing for nonstationary, are developed in Cressie (1984). This should also be used as a source for the spatial locations of the original and residual coal ash values, where "residual values" here are residuals from a drift estimated by median polish; details are in Cressie (1984). The various analyses have shown the E-W direction to possess trend, but the N-S direction to be relatively stationary. We fit a model to robust variogram estimator (4).

Variogram estimators for coal ash (Table 1), together with the number of pairs used in estimating them, is given. Estimated values plotted with the best fit superimposed are shown in (Fig. 1). The "practical rule" (22) can be applied to coal ash originals (Fig. 1a), which means a fit up to and including lag 10. Little parameter change occurs when the maximum lag is extended to 16 (Fig. 1b). The same configuration can be applied to the residual data (Fig. 1c).

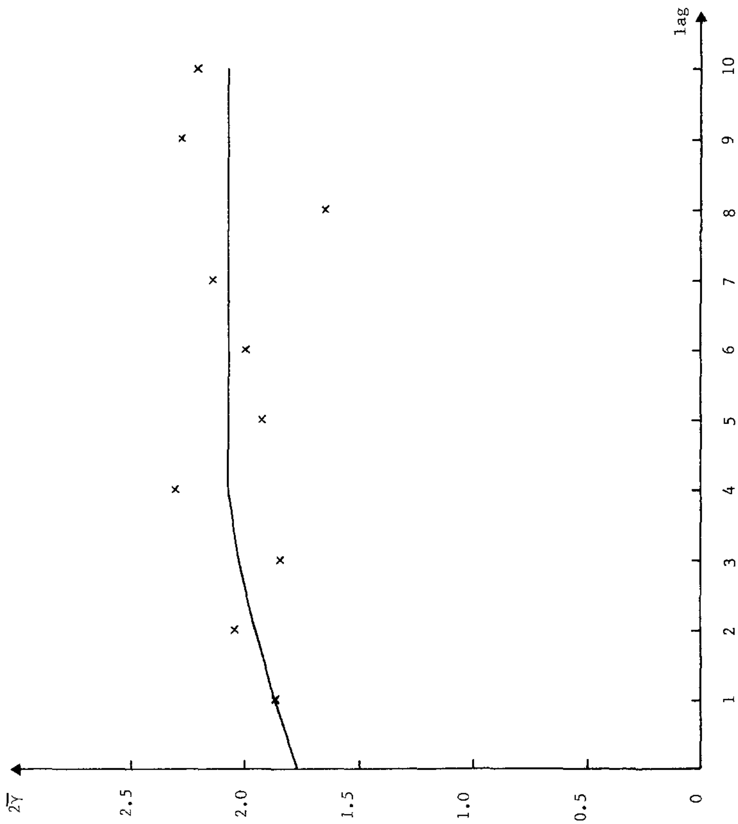
The second data set is an iron ore deposit in Australia. An analysis similar to that described in Cressie (1984) showed approximate stationarity in the E-W

Table 1. Coal Ash Data[a]

Originals				Residuals		
h	$2\bar{\gamma}(h)$	$(2\hat{\gamma}(h))$	N_h	$2\bar{\gamma}(h)$	$(2\hat{\gamma}(h))$	h
1	1.86	(2.40)	186	1.50	(2.07)	1
2	2.04	(2.53)	171	1.62	(2.20)	2
3	1.84	(2.70)	155	1.59	(2.31)	3
4	2.30	(3.00)	145	1.71	(2.49)	4
5	1.92	(2.62)	134	1.65	(2.16)	5
6	1.99	(2.43)	123	1.54	(1.84)	6
7	2.14	(2.42)	111	1.35	(2.07)	7
8	1.65	(2.32)	102	1.46	(1.90)	8
9	2.27	(2.80)	94	1.63	(2.33)	9
10	2.20	(2.87)	87	1.57	(2.35)	10
11	2.20	(2.88)	77	1.51	(2.32)	11
12	2.12	(2.65)	67	1.62	(2.11)	12
13	1.82	(2.19)	57	1.20	(1.83)	13
14	1.59	(3.08)	48	0.90	(1.96)	14
15	2.44	(3.65)	40	1.38	(3.14)	15
16	1.70	(1.83)	32	0.89	(1.42)	16
17	1.24	(1.56)	25	0.91	(0.88)	17
18	1.79	(1.84)	17	0.85	(1.48)	18
19	3.45	(1.97)	10	0.40	(0.90)	19
20	3.72	(5.17)	4	1.38	(1.50)	20

[a]N-S direction; originals and residuals (from median polish). Entries show robust variogram estimator (4) (classical variogram estimator 3 is in parentheses), and number of pairs N_h involved in the lag h estimation.

(a)



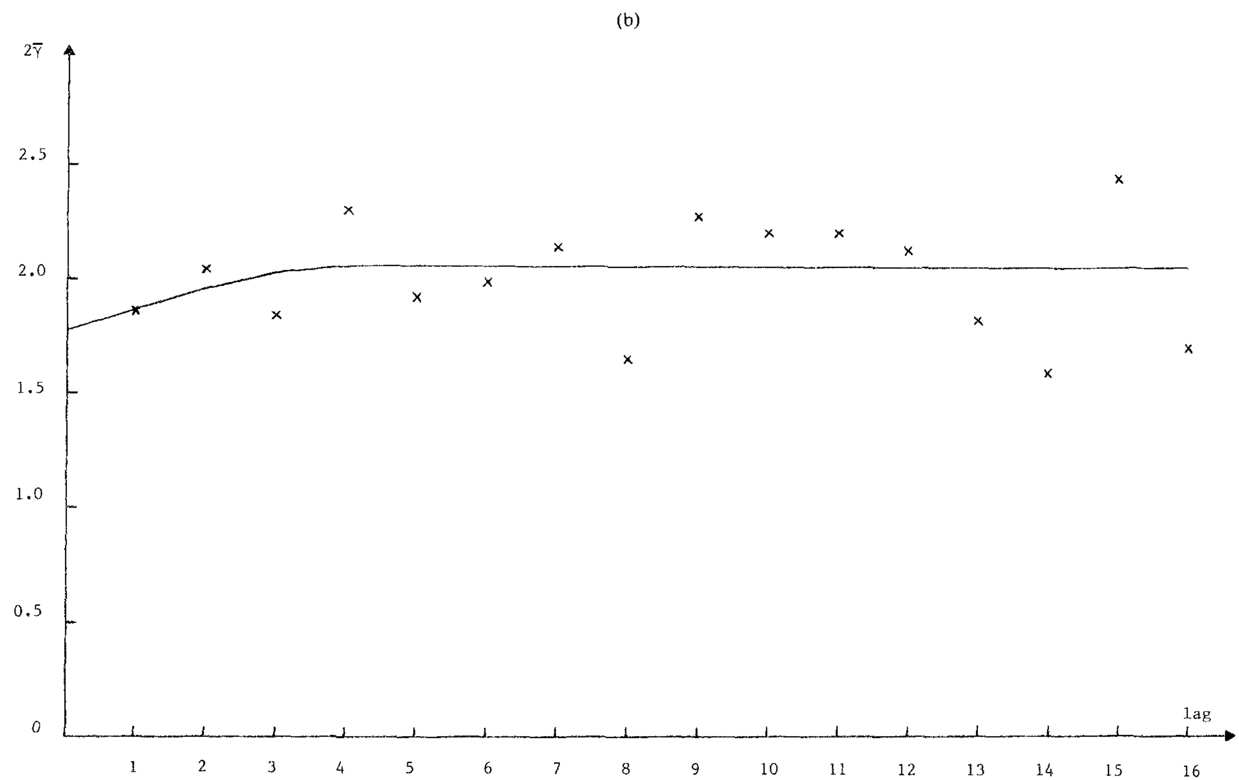


Fig. 1. (a) Coal ash data; N–S direction; originals. Weighted least-squares fit of spherical variogram to estimated variogram (marked with ×) up to lag 10. (b) Coal ash data; N–S direction; originals. Weighted least-squares fit of spherical variogram to estimated variogram (marked with ×), up to lag 16. (c) Coal ash data; N–S direction; residuals from median polish. Weighted least-squares fit of pure nugget effect to estimated variogram (marked with ×), up to lag 16.

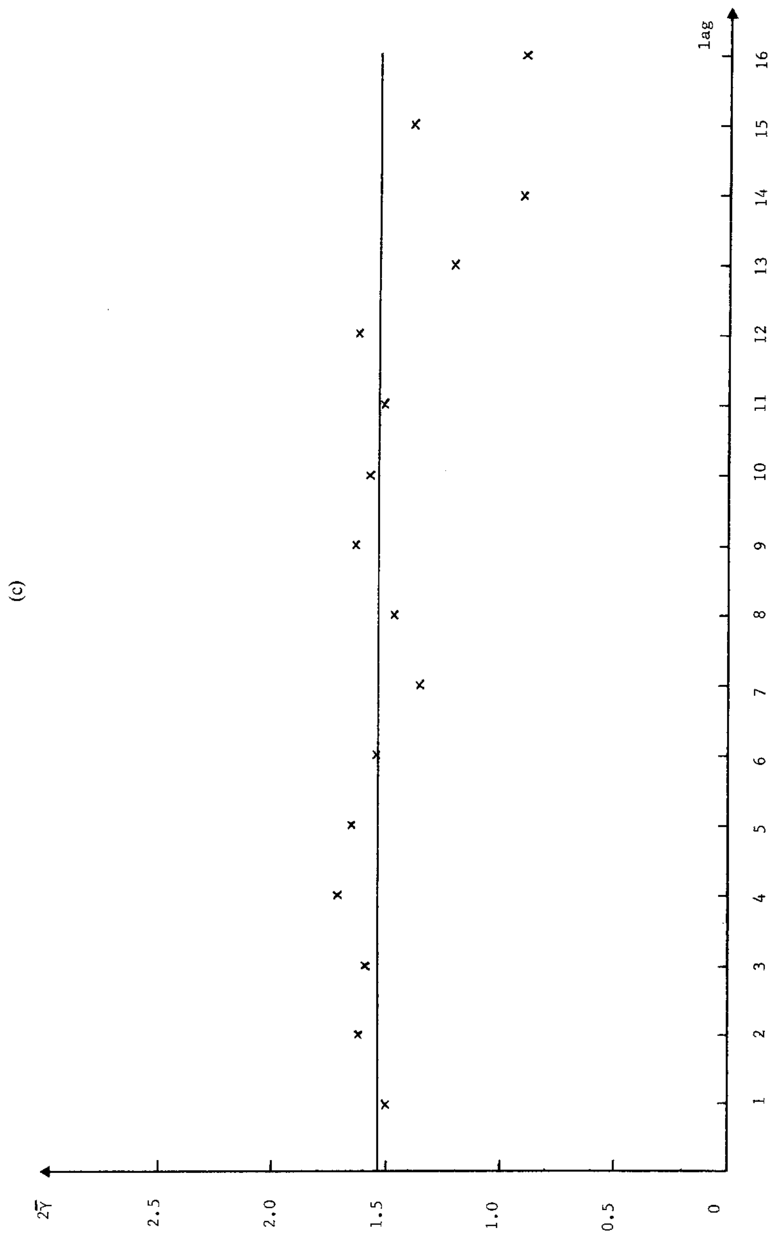


Fig. 1. Continued

direction, but a definite trend in the N-S direction. Also, geometric anisotropy was evident (Journel and Huijbregts, 1978, p. 177), but we will not be concerned here with this because we only fit variograms in the E-W direction.

Variogram estimators for iron ore (Table 2), together with number of pairs used in estimating them, are given. Estimated values using (4) are plotted with the best fit superimposed (Fig. 2). A straight line fit on data only up to lag 7 (Fig. 2a) was considered all that was necessary as kriging would not involve points at greater distances (screen effect), and weighted least-squares fitting in the straight line case would automatically mean that greater lags would contribute almost nothing to the weighted sum of squares. The residual variogram estimate, however, shows spherical structure beyond lag 7, and so estimates up to lag 11 were used in the weighted least-squares fit (Fig. 2b). For confidentiality reasons, the iron ore data set cannot be shown in its entirety, but a plan of the spatial locations of the sample (and of course, of the median polish residuals) (Fig. 3) is given.

In summary then, using the method of *weighted least-squares*, we have been able to estimate variogram values in the following situations:

Table 2. Iron Ore Data

Originals				Residuals		
h	$2\bar{\gamma}(h)$	$[2\hat{\gamma}(h)]$	N_h	$2\bar{\gamma}(h)$	$[2\hat{\gamma}(h)]$	h
1	13.10	(14.39)	103	11.50	(11.96)	1
2	14.81	(18.43)	94	11.89	(13.69)	2
3	14.66	(16.93)	85	11.29	(13.46)	3
4	17.31	(18.67)	77	13.08	(14.05)	4
5	26.50	(23.67)	69	17.38	(16.35)	5
6	19.14	(18.54)	61	14.74	(13.39)	6
7	27.31	(24.66)	53	18.49	(21.96)	7
8	41.65	(33.33)	45	17.86	(23.13)	8
9	36.00	(34.26)	38	16.97	(19.88)	9
10	27.16	(30.15)	31	12.90	(16.64)	10
11	37.37	(39.67)	25	16.31	(19.60)	11
12	39.89	(37.05)	19	15.32	(17.02)	12
13	48.08	(39.03)	13	22.18	(25.49)	13
14	53.18	(48.78)	7	37.90	(25.66)	14
15	165.29	(96.22)	4	19.02	(15.81)	15
16	203.52	(145.44)	2	9.40	(8.80)	16

[a]E-W direction; originals and residuals (from median polish). Entries show robust variogram estimator (4) (classical variogram estimator 3 is in parentheses), and number of pairs N_h involved in the lag h estimator.

(a)

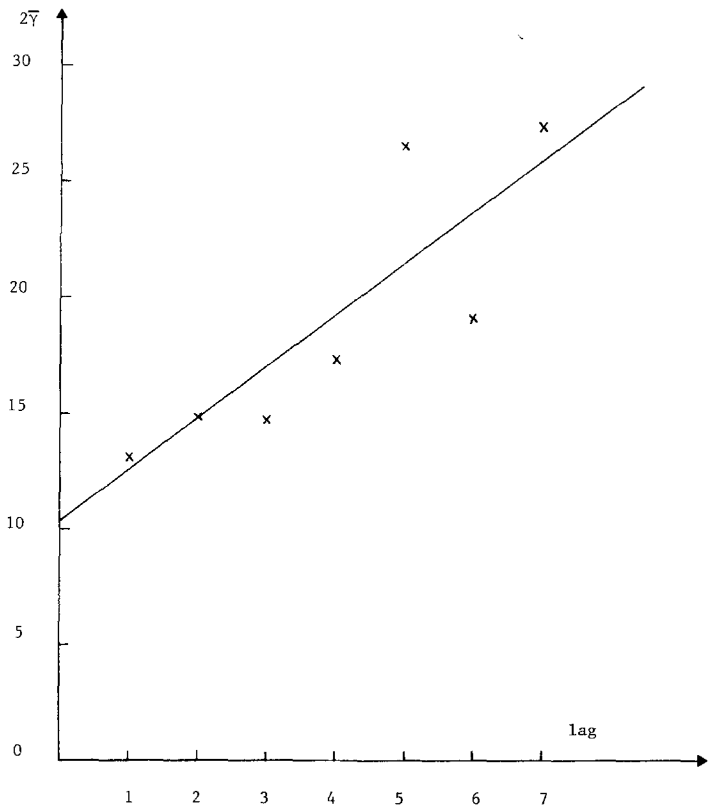


Fig. 2. (a) Iron ore data; E–W direction; originals. Weighted least-squares fit of straight line variogram to estimated variogram (marked with X), up to lag 7. (b) Iron ore data; E–W direction; residuals from median polish. Weighted least-squares fit of spherical variogram to estimated variogram (marked with X), up to lag 11.

Coal ash originals, N–S: Spherical model (24)

$$c_0 = 0.89, \qquad c_s = 0.14, \qquad a_s = 4.31$$

Coal ash residuals, N–S: Spherical model (24)

$$c_0 + c_s = 0.77, \qquad a_s = 0 \quad \text{(pure nugget effect)}$$

Iron ore originals, E–W: Linear model (26)

$$c_0 = 5.17, \qquad b_l = 1.11$$

Iron ore residuals, E–W: Spherical model (24)

$$c_0 = 4.83, \qquad c_s = 3.59, \qquad a_s = 8.73.$$

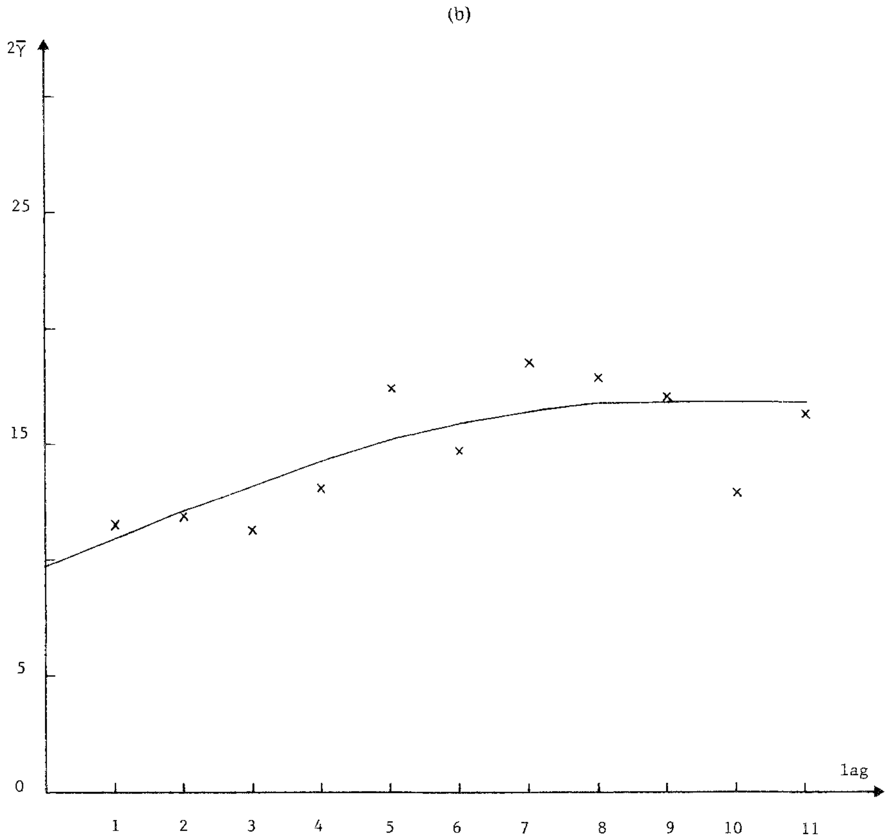


Fig. 2. Continued

We also tried exponential (see 25) fitting, and in each case the spherical fit improved the weighted sums of squares by a few percent.

Before the final model is chosen, one has to take into account variograms in other directions, possible anisotropies, and so forth. These adjustments are usually made however, in light of a first sensible fitting of a model to estimated variogram values. This is why a statistically based approach to variogram estimation must be of interest to the practicing geostatistican.

4. CONCLUDING REMARKS

The variable $(Z_{t+h} - Z_t)^2$ is *very* skewed, and $\Sigma\,(Z_{t_i} - Z_{t_i+h})^2/N_h$ remains skewed (although less so), whereas $\Sigma\,|Z_{t_i+h} - Z_{t_i}|^{1/2}/N_h$ has less of a skewness problem. Weighted least squares, which is Gaussian-based, may not be all that appropriate for small N_h. Some other criterion which takes into account the inherent positiveness and skewness of the estimator would be needed when, say, $\max\{N_h; h \geqslant 1\} < 30$.

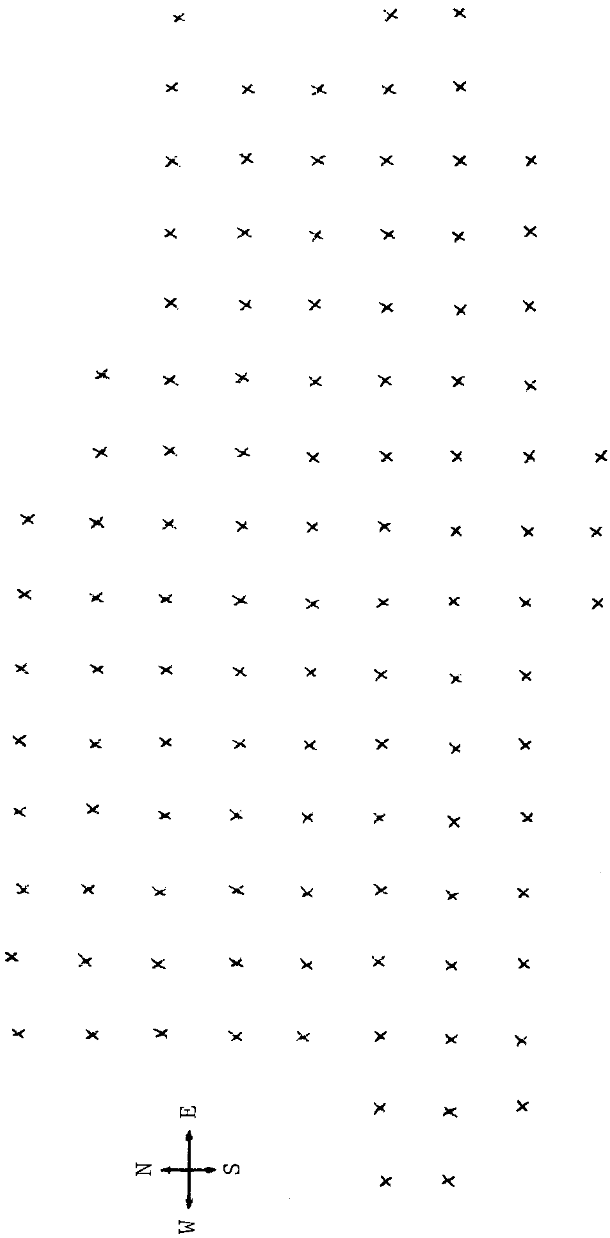


Fig. 3. Spatial locations of iron ore data set; samples are on an approximate 50-m square grid.

In conclusion then, we have tried to formalize the method of variogram model fitting by setting out three possible approaches.

(i) Least squares (see 6). This, and "eyeballing," are probably the methods most used today. David (1977, p. 522) does talk about weighting each variogram estimator according to the number of points involved in estimation, but does not say how.
(ii) Weighted least squares (see 23). This has the desirable feature that weighting is directly proportional to the number of observations N_h, and is larger for smaller lags.
(iii) Generalized least squares (see 7). A way to handle the forbidding expressions for Σ, through iteration, is given (Subsection 2.3).

Weighted least squares is the true compromise between simplicity and statistical efficiency. A fit based on the robust estimator (4) needs less compromise than the Matheron estimator (3). We *do not* recommend blind use of this method, but rather recommend it as a way of homing in on a satisfactory model that also takes into account such things as anisotropies and geological considerations.

ACKNOWLEDGMENT

Comments on a first draft, from Doug Hawkins, André Journel, and Geoff Laslett, are gratefully acknowledged.

REFERENCES

Armstrong, M., 1984, Improving the estimation and modelling of the variogram, *in* G. Verly et al. (Eds.) Geostatistics for natural resources characterization: Dordrecht, Reidel, p. 1-19.

Armstrong, M. and Delfiner, P., 1980, Towards a more robust variogram: A case study on coal, Internal note N-671: Centre de Geostatistique, Fontainebleau, France.

Buxton, B., 1982, A geostatistical case study, unpublished masters thesis: Stanford University, California, 84 p.

Carroll, R. J. and Ruppert, D., 1982, A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model: *Jour. Amer. Stat. Assoc.*, v. 77, p. 878-882.

Clark, I., 1979, Practical geostatistics: Applied Science Publishers, Essex, England, 129 p.

Cressie, N. A. C., 1979, Straight line fitting and variogram estimation (with discussion): *Bull. Inter. Stat. Inst.*, v. 48, Book 3, p. 573-582.

Cressie, N. A. C., 1980, M-estimation in the presence of unequal scale. *Stat. Neerlandica*, v. 34, p. 19-32.

Cressie, N. A. C., 1984, Towards resistant geostatistics, in G. Verley et al. (Eds.) Geostatistics for natural resources characterization: Dordrecht, Reidel, p. 21-44.

Cressie, N. A. C. and Glonek, G., 1984, Median based covariogram estimators reduce bias: *Stat. Prob. Lett.*, v. 2, p. 299-304.

Cressie, N. A. C. and Hawkins, D. M., 1980, Robust estimation of the variogram: *Jour. Inter. Assoc. Math. Geol.*, v. 12, p. 115-125.

David, M., 1977, Geostatistical ore reserve estimation: Amsterdam, Elsevier, 364 p.
Davis, B. M. and Borgman, L. E., 1982, A note on the asymptotic distribution of the sample variogram: *Jour. Inter. Assoc. Math. Geol.*, v. 14, p. 189–193.
Delfiner, P., 1976, Linear estimation of nonstationary spatial phenomena, in M. Guarascio et al. (Eds.) Advanced geostatics in the mining industry: Dordrecht, Reidel, p. 49–68.
Diamond, P. and Armstrong, M., 1983, Robustness of variograms and conditioning of kriging matrices: Internal note N-804, Centre de Geostatistique, Fontainebleau, France.
Gomez, M. and Hazen, K., 1970, Evaluating sulfur and ash distribution in coal seams by statistical response surface regression analysis: U.S. Bureau of Mines Report, RI 7377.
Hawkins, D. M. and Cressie, N. A. C., 1984, Robust kriging–a proposal: Jour. Inter. Assoc. Math. Geol., v. 16, p. 3–18.
Huber, P. J., 1964, Robust estimation of a location parameter: Ann. Math. Stat., v. 35, p. 73–101.
Journel, A. and Huijbregts, C., 1978, Mining geostatistics: London, Academic Press, 600 p.
King, H. F., McMahon, D. W., and Bujtor, G. J., 1982, A guide to the understanding of ore reserve estimation, *in* Proceedings of the Australasian Institute of Mining and Metallurgy, Supplement: v. 281, p. 1–21.
Matheron, G., 1963, Principles of geostatistics: Econ. Geol., v. 58, p. 1246–1266.
Matheron, G., 1971, The theory of regionalized variables and its applications: Cahiers du Centre de Morphologie Mathematique, No. 5, Fontainebleau, France.
Sharp, W. E., 1982, Estimation of semi-variograms by the maximum entropy method: *Jour. Inter. Assoc. Math. Geol.*, v. 14, p. 456–474.
Starks, T. and Fang, J., 1982a, On the estimation of the generalized covariance function: Jour. Inter. Assoc. Math. Geol., v. 14, p. 57–64.
Starks, T. and Fang, J., 1982b, The effect of drift on the experimental semi-variogram: Jour. Inter. Assoc. Math. Geol., v. 14, p. 309–320.
Switzer, P., 1984, Inference for spatial autocorrelation functions in G. Verly et al. (Eds.) Geostatistics for natural resources characterization: Dordrecht, Reidel, p. 127–140.