Mathematical Geology, Vol. 12, No. 2, 1980

Robust Estimation of the Variogram: I[1]

Noel Cressie[2] and Douglas M. Hawkins[3]

It is a matter of common experience that ore values often do not follow the normal (or lognormal) distributions assumed for them, but, instead, follow some other heavier-tailed distribution. In this paper we discuss the robust estimation of the variogram when the distribution is normal-like in the central region but heavier than normal in the tails. It is shown that the use of a fourth-root transformation with or without the use of M-estimation yields stable robust estimates of the variogram.

KEY WORDS: Geostatistics, kriging, robust estimation, variogram.

INTRODUCTION

The variogram is a cornerstone of geostatistics. From it one deduces the model form that is applicable to the ore body, the kriging weights, and the consequent standard errors of estimation by kriging. Published work on geostatistics, while stressing the importance of the variogram, usually underplays the problems in estimating it. This is because most discussion assumes that the underlying data are either normal to an adequate degree of approximation or (if, for example, they are lognormal) have already been transformed to normality.

In actual practice, by contrast, one discovers all too often that the data are not normal, and in particular tend to be contaminated by occasional outliers. These problems are mentioned inter alia by David (1977, 1978) who considers that the estimation of the variogram in the face of nonnormal data is an important and unsolved problem. This paper addresses this problem and provides an introductory evaluation of several possible robust methods of estimating the variogram.

---

[1]Manuscript received 18 June 1979; revised 2 November 1979.

[2]Visiting Scientist, NRIMS, during the period in which this work was carried out. School of Mathematical Sciences, The Flinders University of South Australia, Australia.

[3]National Research Institute for Mathematical Sciences, CSIR, P.O. Box 395, Pretoria 0001, South Africa.

OVERVIEW OF ROBUST STATISTICS

In view of the comparative novelty of the statistical theory of robust estimation, a brief introduction may be in order. As a simple illustrative example, suppose one has a sample $X_1, \ldots, X_n$ from a distribution F, symmetric about an unknown location parameter μ that we wish to estimate. If F is the normal (Gaussian) distribution, then the best estimator is the sample mean $\overline{X}$. This is sufficient (i.e., contains all the sample information about μ) and has a minimum variance among all unbiased estimators of μ. The efficiency of an asymptotically unbiased estimator may be measured by the ratio to its asymptotic variance of the minimum asymptotic variance among all such estimators. For the normal distribution the mean thus has efficiency 100%, while the sample median, which is also unbiased, has an efficiency of 64%.

Both the mean and the median are particular members of the class of *trimmed means* defined as follows. Let $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ represent the ordered sample and let $\alpha \in [0, \frac{1}{2})$ be such that $n\alpha$ is an integer. Then the $100\alpha\%$ trimmed mean is defined by

$$T(\alpha) = \sum_{\alpha n+1}^{n-\alpha n} X_{(i)}/n(1 - 2\alpha)$$

that is, the $100\alpha\%$ largest and the $100\alpha\%$ smallest observations are deleted from the sample, and the remaining observations averaged. If $\alpha = 0$, $T(\alpha)$ is the mean of all the observations while $\alpha = \frac{1}{2} - \epsilon$ corresponds to the median. Clearly the trimmed means with $\alpha > 0$ are insensitive to the presence of up to $n\alpha$ outliers on either side of the sample; hence they are robust against outliers, in contrast to the sample mean, which is strongly affected by outliers.

The common assumption underlying much of data analysis is that the normal distribution represents a good point of departure, and when the robustness of estimators to the nonnormality of the data is studied, it is quite common to assume that the data are from a contaminated normal distribution. Letting Φ denote the cumulative distribution function (cdf) of the standard normal $N(0, 1)$ distribution, the cdf of a contaminated normal is defined by

$$F_\epsilon(x) = (1 - \epsilon)\,\Phi\,[(x - \mu)/\sigma] + \epsilon\Phi\,[(x - \mu)/g\sigma]$$

For $\epsilon = 0$ or $\epsilon = 1$, this reduces to the normal distribution, while if $g > 1$ and $\epsilon > 0$, it defines a distribution that resembles $N(\mu, \sigma^2)$ in the central regions but which has considerably heavier tails.

Taking the family of contaminated normal distributions as a paradigm for the class of normal-like but heavy-tailed distributions, one then seeks estimators whose efficiency is high when the data are normal ($\epsilon = 0$) and which are reasonably insensitive to departures from normality ($\epsilon > 0$).

The class of trimmed means is one that provides such estimators of location, and is in fact a subclass of the larger class of L-estimators, formed by taking linear combinations of order statistics of the sample.

Another large class is that of M-estimators T which minimize

$$\sum_{1}^{n} \rho(X_i - T)$$

or equivalently, which solve

$$\sum_{1}^{n} \psi(X_i - T) = 0, \qquad \text{where} \quad \psi = \rho'$$

Any ψ which is antisymmetric about the origin yields a potential estimator from this class. The choices $\psi(x) = x$ and $\psi(x) = \operatorname{sgn}(x)$ correspond, respectively, to the sample mean and median. If F is known to have a density f, then the choice $\psi(x) = -d \log f(x)/dx$ yields T as the maximum-likelihood estimator, a property reflected in the mnemonic name M-estimator.

These brief comments do not and cannot do full justice to the growing area of robust inference, and the interested reader should consult the very readable overviews by Huber (1978) and by Jaeckel (1971).

METHODS OF ESTIMATING THE VARIOGRAM

Like the proverbial cat which may be skinned in more than one way, the variogram has a number of properties that allow it to be estimated in different ways. We distinguish four that seem to us to be potentially useful. Suppose $\{Z_t\}$ is a collection of regionalized variables (for example, ore grade at the point t), whose differences $Z_{t+h} - Z_t$, have a zero mean and a variance depending only on h. Matheron (1963) has coined the term variogram for this variance $2\gamma(h)$. Ultimately one would like to have $\gamma(h)$ as some explicit mathematical function of h; in practice however this is best and most easily done by estimating $\gamma(h)$ for various discrete values of h, and then carrying out a modeling exercise on these values to fit a suitable function or to test possible hypothesized models (Davis, 1978). A satisfactory conclusion of the latter exercise thus depends on good estimates in the former, and it is this that we consider here. The four approaches are

A Define the quantities $Y_t = (Z_{t+h} - Z_t)^2$. Then $2\gamma(h) = E(Y)$, and the estimation of the variogram becomes a problem of estimating the expectation of the random variables Y_t which, under the normality of the Z_t, follow scaled χ_1^2 distributions.

B If the sequence Z_t may be assumed to be stationary (or to have been transformed initially to stationarity) then $E(Z_{t+h}) = E(Z_t)$. Thus $2\gamma(h) = \text{var}$

$(Z_{t+h} - Z_t)$ and so the estimation of the variogram may be approached by estimating the variance of the *symmetric* random variables $Z_{t+h} - Z_t$.

C For many purposes (including the estimation of kriging weights), it is enough to determine the set $\gamma(h)$ for all h up to an unknown constant multiple. Now, provided var (Z_t) is finite:

$$2\gamma(h) = 2 \text{ var } (Z_t) \; [1 - \rho(h)]$$

where $\rho(h)$ here refers to the autocorrelation at lag h. Thus an estimate of the set $\rho(h)$ also determines the set $\gamma(h)$ up to the unknown variance.

D Assuming for the moment that the data form a traverse, there exists an autoregressive-moving average (ARMA) model describing them

$$\varphi(B)Z_t = \psi(B)a_t$$

where φ and ψ are power series in the backshift operator B and the series a_t is white noise. From φ, ψ, and the variance of the a_t, one may deduce $\gamma(h)$, and any method of estimating φ and ψ can yield $\gamma(h)$ up to an unknown constant multiple of the variance of the a_t.

All four methods could provide the basis for robust (or for that matter conventional) methods of estimating the variogram; this paper, however, deals only with approach A.

PRELIMINARY TRANSFORMATION

The problem for which we have by far the most powerful tools, algorithms, and theoretical results is that of the robust estimation of a center of symmetry (Huber 1964, 1972; Hogg 1974). By contrast, the estimation of some characteristic of a nonsymmetric distribution or of a scale parameter is not well understood at all. Now we are interested in estimating the variogram

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

Considering the problem as one in estimating the expectation of $(Z_{t+h} - Z_t)^2$, we find at once that under the normal model which forms our point of departure, $(Z_{t+h} - Z_t)^2$ follows a $2\gamma(h)\chi_1^2$ distribution, which is highly skewed. Therefore, in order to use the well-known results on robust estimation, we looked for a method of transforming the problem to one of estimating a center of symmetry. The class of power transformations $Y_t = \{(Z_{t+h} - Z_t)^2\}^\lambda$ was chosen since it is very broad and includes several that one might believe a priori to be suitable: $\lambda \to 0$ gives the logarithmic, $\lambda = \frac{1}{3}$ the Wilson-Hilferty, and $\lambda = 1$ the identity transformation, respectively. A theoretical study of the cumulants of Y_t as a function of λ showed that the choice $\lambda = 0.25$ leads to a Y_t that is very close to normal if the Z_t are normal, having cumulant coefficients of skewness and kurtosis of 0.08 and -0.52, respectively. In the subsequent study use was thus made of this transformation (to $Y_t = |Z_{t+h} - Z_t|^{1/2}$) and the problem regarded as one of measuring the center of symmetry of the Y_t. Huber (1972) in fact has suggested estimating the scale by first taking logarithms and then estimating location.

Undoing the effect of this transformation is a relatively straightforward operation. $Y_t/\{2\gamma(h)\}^{1/4}$ is the fourth root of a χ_1^2 variate and so can be shown quite easily to have

$$\text{mean} = 2^{1/4}\,\Gamma\left(\tfrac{3}{4}\right)/(\pi)^{1/2} = 0.82216$$

$$\text{and variance} = 2^{1/2}\left[\pi^{-1/2} - \Gamma^2\left(\tfrac{3}{4}\right)/\pi\right] = 0.12192.$$

Let $\bar{Y}$ be the arithmetic mean of n independent Y_t. Assuming that $\bar{Y}$ is normally distributed (which is true for n large by the central limit theorem and is, as noted above, nearly true for n as small as 1) some simple but unilluminating distribution theory then shows that

$$E\left[\bar{Y}^4/2\gamma(h)\right] = 0.457 + (0.494/n) + (0.045/n^2)$$

and so an almost completely unbiased estimator for $2\gamma(h)$ is given by $\bar{Y}^4/(0.457 + 0.494n^{-1} + 0.045n^{-2})$.

Later we use this correction factor, not only with $\bar{Y}$ but with other, more robust estimators. Provided that these estimators are also asymptotically normal (as they are) and have high asymptotic efficiency (as most of them do) then the same correction procedure will turn these estimators, too, into almost unbiased estimators of the variogram.

The above reasoning applies to independent Y_t. In fact, our Y_t are not independent, since they are functions of the Z_t which are mutually correlated. However, this correlation is not of great practical import, as may be shown by the following reasoning

$$\operatorname{cov}(Z_{t+h} - Z_t, Z_{s+h} - Z_s) = E(Z_{t+h} - Z_t)(Z_{s+h} - Z_s)$$
$$= \operatorname{var}(Z)\,\{2\rho(|s-t|) - \rho(|s-t+h|) - \rho(|s-t-h|)\}, \qquad \text{while}$$
$$\operatorname{var}(Z_{t+h} - Z_t) = \operatorname{var}(Z_{s+h} - Z_s) = \operatorname{var}(Z)\,2[1 - \rho(h)]$$

Thus the correlation between $Z_{t+h} - Z_t$ and $Z_{s+h} - Z_s$ is

$$\{\rho(|s-t|) - \tfrac{1}{2}[\rho(|s-t+h|) + \rho(|s-t-h|)]\}/[1 - \rho(h)]$$

Now in a real-life problem

(i) If $|s-t|$ is large, then $\rho(|s-t|)$ and $\rho(|s-t \pm h|)$ are small.
(ii) The autocorrelation function $\rho(h)$ varies smoothly, and so if $|s-t|$ and h are small, it may be approximated locally by a linear function for which $2\rho(i-j) = \rho(i-j+h) + \rho(i-j-h)$.

(iii) For h around 1, there may be nonnegligible correlations between two differences; however, their effect in estimation is merely to increase slightly the width of confidence intervals and hence create a mild degree of conservatism (Cressie, 1980).

Thus we see that for problems of practical interest the interdependence between Y_t and Y_s will be negligible except for a negligibly small proportion of the Y_t, Y_s pairs. We therefore feel justified in treating the $\{Y_t\}$ as if they were an independent random sample.

ESTIMATORS USED

Altogether 10 estimators of the type A were studied. Those of the class of M-estimators solved for T an equation of the form

$$\sum_{1}^{n} \psi\,\{(Y_t - T)/cS\} = 0$$

where S is a measure of scale for which we used the median absolute deviation of the Y_t about their median (Hogg 1974, p. 910).

The 10 estimators were

(i) The mean $\bar{Y}$ of the Y_t values.

(ii) The median of the Y_t values.

(iii), (iv), and (v) The trimmed mean of the Y_t values trimming, respectively, 5, 10, and 25% of the extreme order statistics on each side. (Gastwirth and Rubin, 1969).

(vi) The Huber M-estimator, where

$$\psi(x) = \begin{cases} x & |x| \leqslant 1 \\ \operatorname{sgn} x & |x| > 1, \quad \text{and} \quad c = 2.2. \end{cases}$$

(vii) The Tukey bisquare M-estimator using

$$\psi(x) = \begin{cases} x(1 - x^2)^2 & |x| \leqslant 1 \\ 0 & |x| > 1, \quad \text{where} \quad c = 6 \text{ (Gross 1977, p. 342)} \end{cases}$$

(viii) The Hampel M-estimator using

$$\psi(x) = \begin{cases} x & |x| \leqslant 3 \\ 3\operatorname{sgn}(x)\,[(14 - |x|)/11], & 3 < |x| < 14 \\ 0 & |x| \geqslant 14, \end{cases}$$

where $c = 1$ (Shorack 1976, p. 124)

(ix) The Andrews M-estimator using

$$\psi(x) = \begin{cases} \sin x & |x| \leqslant \pi \\ 0 & |x| > \pi, \quad \text{where} \quad c = 3.11 \end{cases}$$

(x) The conventional estimator (Matheron, 1963)

$$\sum_{1}^{N} (Z_{t+h} - Z_t)^2/N, \quad \text{where} \quad N = n - h$$

The study consisted of simulating 500 traverses of length 50 according to the model

$$Z_t = 0.6Z_{t-1} + U_t$$

and estimating $\gamma(1)$ using each estimator. The U_t were independent identically distributed random variables from the following distributions

A: $N(0, 1)$
B: Standard Laplace: $f(x) = \frac{1}{2}\exp(-|x|)$
C: $N(0, 1)$ 5% contaminated with $N(0, 9)$
D: $N(0, 1)$ 10% contaminated with $N(0, 9)$
E: $N(0, 1)$ 20% contaminated with $N(0, 9)$
F: $N(0, 1)$ 5% contaminated with $N(0, 100)$

In addition, as case G, we analyzed a set of real data from the Hartebeestfontein gold mine, kindly supplied by Dr. D. G. Krige. These latter data consisted of 83 parallel traverses each containing up to 128 values. Since these values were expected a priori to be lognormally distributed, they were log transformed for analysis and the following results refer to the transformed values.

RESULTS

The overall means and standard deviations across traverses of the estimates of $\gamma(1)$ are given in Table 1. Turning first to the standard deviations, we see that the conventional estimator is by far the most stable for the normal distribution, and by far the least stable in almost all of the heavy-tailed runs (including that on the real data). The estimators having consistently the smallest standard deviations in the nonnormal cases are the M-estimators. Except for the grossly (and probably unrealistically) contaminated case F, the median performs poorly, and perhaps surprisingly, the mean of the Y_t performs very well. All of the trimmed means were dominated by M-estimators, the best of which seems to be Tukey's bisquare.

Table 1. Means and Standard Deviations of 10 Variogram Estimates, for Seven Different Data Sets.

	Data set						
	A	B	C	D	E	F	G
Estimator	Standard deviations						
1	0.35	0.65	0.46	0.57	0.78	1.32	1.88
2	0.47	0.82	0.56	0.66	0.87	0.86	2.19
3	0.41	0.73	0.50	0.62	0.87	1.29	2.10
4	0.41	0.73	0.50	0.60	0.85	1.06	2.14
5	0.45	0.76	0.52	0.61	0.83	0.88	2.20
6	0.37	0.65	0.45	0.54	0.76	0.91	1.92
7	0.39	0.66	0.46	0.54	0.75	0.71	1.97
8	0.36	0.65	0.44	0.54	0.76	0.96	1.89
9	0.35	0.65	0.46	0.56	0.78	1.28	1.88
10	0.28	0.78	0.65	0.84	1.15	6.65	2.68
	A	B	C	D	E	F	G
	Means						
1	1.31	2.03	1.57	1.85	2.47	2.61	0.70
2	1.43	2.00	1.62	1.83	2.35	2.04	0.70
3	1.51	2.29	1.77	2.07	2.75	2.69	0.74
4	1.52	2.23	1.75	2.02	2.65	2.43	0.74
5	1.54	2.19	1.76	2.00	2.57	2.26	0.73
6	1.34	1.97	1.55	1.80	2.35	2.13	0.69
7	1.36	1.95	1.55	1.77	2.29	1.87	0.68
8	1.32	1.99	1.54	1.80	2.39	2.14	0.69
9	1.31	2.03	1.57	1.85	2.46	2.54	0.70
10	1.27	2.46	1.76	2.21	3.15	7.38	0.84

Turning next to the means given in Table 1, it becomes necessary for us to consider very carefully a fundamental question that until now we have ignored, viz what exactly we are trying to achieve with robust estimation of the variogram. By the usual definition, $2\gamma(h) = E(Z_{t+h} - Z_t)^2$, and the conventional estimator is an unbiased estimator of $2\gamma(h)$. Thus any other estimator whose mean does not coincide with that of the conventional estimator is biased, and so may seem to be undesirable. We do not believe this to be the case, and put forward the following justification.

First suppose the following Kalman filter type model holds, giving outliers that are qualitatively the same as the first order a.v. above (Meditch 1969)

$$W_t = \int_h \alpha(h)\, W_{t+h}\, dh + U_t$$

$$Z_t = W_t + e_t$$

Here the W_t represents a continuous unobservable underlying process defined by the autoregression coefficient function $\alpha(h)$ and the white noise U_t. We may observe Z_t, which is W_t plus a white noise error e_t, at any point t. The objective is to estimate the set W_t from the set Z_t. Note where this model differs from the usual geostatistical one: instead of using Z_t as the basic entity and saying that we are only interested in values of Z_t averaged over a block, we introduce the abstract quantity W_t, which is defined at a point but varies smoothly, thus being similar in practice to the average of Z_t over a block, but conceptually different. The variogram $2\gamma(h) = E(Z_{t+h} - Z_t)^2$

$$
\begin{aligned}
2\gamma(h) &= E(Z_{t+h} - Z_t)^2 \\
&= E(W_{t+h} - W_t + e_{t+h} - e_t)^2 \\
&= 2\,[\gamma_W(h) + \sigma_e^2]
\end{aligned}
$$

where $2\gamma_W(h)$ is the variogram of W_t and σ_e^2 the variance of e. By definition, the W-process is continuous, and so as $h \to 0$, $\gamma_W(h) \to 0$. Thus the nugget effect (which is the limit of $\gamma(h)$ as $h \to 0$) is just σ_e^2. If the distribution of e_t is altered, but not that of U_t or $\alpha(h)$, the variogram is merely moved up or down parallel to its starting profile. The information about $\alpha(h)$ and the variance of the U_t is contained in $\gamma_W(h)$, and so these parallel movements of the variogram do not affect it. Thus the interpretation of the variogram may be reduced to (i) estimating the nugget effect σ_e^2, (ii) subtracting this from $\gamma(h)$, and (iii) inferring the properties of the W_t process. From these separate pieces of information one may infer an optimal scheme for regressing W_t on the set Z_t.

Let us interpret these remarks in the context of robust estimation. If one assumes both U_t and e_t to be normally distributed, then a normal distribution for Z_t and a "classical" geostatistical estimation and kriging scheme, may be deduced. If, however, the U_t process remains normal but the e_t process is assumed to have some heavier-tailed distribution, then the Z_t are not normal but exhibit occasional outliers. It is here that the use of good robust estimators of the variogram is desirable. These estimators, by reducing the sampling variability of estimates of $\gamma(h)$, can provide a more stable estimation procedure for $\gamma_W(h)$, and any downward bias will merely imply that they underestimate σ_e^2 while leaving the structure of $\gamma_W(h)$ unaffected.

Since the important parameters defining the deposition process relate to the $\alpha(h)$, we are thus able to estimate these with greater accuracy from such a robust variogram. The determination of an optimum set of weights for kriging or robust kriging (as set out in the following paragraph) requires estimates of both $\alpha(h)$ and various parameters, including perhaps σ_e^2, of the distribution of the e_t. This determination in the nonnormal case may however have to be done in an iterative way by successively computing estimates $\widehat{W}_t$ and residuals $Z_t - \widehat{W}_t$ and using these residuals to refine our estimates of the distribution of the e_t, and, hence, improve our estimates $\widehat{W}_t$.

It is appropriate to observe at this point that the adoption of this model has far-reaching consequences, for if the e_t are not normally distributed, then it is not correct to use the linear regressions implicit in kriging. Thus, parallel to the development of robust estimators of the geostatistical model, one needs to develop robust equivalents to kriging. For example, a robust alternative to the kriging equation $T = \Sigma_i b_i Z_i$ might be the solution of $\Sigma_i b_i \psi(Z_i - T) = 0$, which we recognize as an M-estimator. More generally, a regression equation merely provides one with an estimator of a location parameter which, in the case of normal data, is a center of symmetry.

As already noted, this is the best studied problem in robust estimation, and solutions abound. Unfortunately most of these, including those involving the use of M-estimators, cannot be applied in most practical cases because of the enormous volume of computation they would entail, and so effort in this area should concentrate on noniterative estimators based possibly on Winsorization, outlier rejection, or trimming of the apparently outlying Z_t.

Returning now to Table 1, we see that in the heavy-tailed cases all the robust estimators are biased relative to the conventional estimator. Again in all cases except case F, there is little to choose between the mean of the Y_t and the M-estimators; in case F however the M-estimators have a consistently greater downward bias.

Thus the M-estimators are the estimators of choice, having quite good efficiency for normal data coupled with stability for all the heavy-tailed distributions studied. Another very simple estimator, $\bar{Y}$, the mean of the Y_t performs equally well for the distributions contaminated with $N(0, 9)$ data and, which is more significant, for the real data. In a technical sense $\bar{Y}$ is not robust in that it goes to infinity if any Y_t does, however it does seem to be robust enough to handle data which, while not normal, have outliers that deviate by not more than six or seven standard deviations. This description is likely to cover the large majority of real data sets.

CONCLUSION AND OVERVIEW

The paper uncovers a number of very interesting conclusions about the robust estimation of the variogram, among which is the surprising finding that the arithmetic mean of the fourth root of $(Z_{t+h} - Z_t)^2$ gives a reasonably robust, stable estimate of the variogram. Furthermore, the orthodox M-estimators are equally stable, and theoretically more robust, while trimmed means and the median do not perform well.

Much work remains to be done. There is no pretence that all reasonable robust estimators were tried, nor that working via the fourth-root transformation is the only or necessarily the best approach to robust estimation of the variogram. Furthermore the whole area of kriging in the face of nonnormal disturbances warrants extensive attention.

ACKNOWLEDGMENTS

The authors are grateful to T. McCully for programming many of the computer routines used, and to Dr. D. G. Krige and J. E. Magri of Anglovaal for helpful discussions and for providing the data used as set G.

REFERENCES

Cressie, N., 1980, Relaxing assumptions in the one sample t-test: Australian J. Stat., v. 22, to appear.

David, M., 1977, Geostatistical ore reserve estimation: Elsevier, Amsterdam, 364 p.

David, M., 1978, Sampling and estimation problems for three-dimensional spatial stationary and non-stationary stochastic processes as encountered in the mineral industry: J. Stat. Planning Inference, v. 2, p. 211-244.

Davis, B. M., 1978, Some tests of hypothesis concerning variograms. Ph.D. Thesis, University of Wyoming, Laramie, p. 147.

Gastwirth, J. and Rubin, M., 1969, On robust linear estimators: Ann. Math. Stat., v. 40, p. 24-39.

Gross, A., 1977, Confidence intervals for bisquare regression estimates: J. Amer. Stat. Assoc., v. 72, p. 341-354.

Hogg, R. V., 1974, Adaptive robust procedures: A partial review and some suggestions for future applications and theory: J. Amer. Stat. Assoc., v. 69, p. 909-923.

Huber, P. J., 1964, Robust estimation of a location parameter. Ann. Math. Stat., v. 35, p. 73-101.

Huber, P. J., 1972, Robust statistics–review: Ann. Math. Stat., v. 43, p. 1041-1067.

Huber, P. J., 1978, Robust statistical procedures: SIAM, Philadelphia.

Jaeckel, L., 1971, Robust estimates of location: Symmetry and asymmetric contamination. Ann. Math. Stat., v. 42, p. 1020-1034.

Matheron, G., 1963, Principles of geostatistics: Econ. Geol., v. 58, p. 1246-1266.

Meditch, J. S., 1969, Stochastic optimal linear estimation and control: McGraw-Hill, San Francisco, 269 p.

Shorack, G., 1976, Robust studentization of location estimates: Statistica Neerlandica, v. 30, p. 119-141.