Mathematical Geology, Vol. 30, No. 2, 1998

Highly Robust Variogram Estimation[1]

Marc G. Genton[2]

The classical variogram estimator proposed by Matheron is not robust against outliers in the data, nor is it enough to make simple modifications such as the ones proposed by Cressie and Hawkins in order to achieve robustness. This paper proposes and studies a variogram estimator based on a highly robust estimator of scale. The robustness properties of these three estimators are analyzed and compared. Simulations with various amounts of outliers in the data are carried out. The results show that the highly robust variogram estimator improves the estimation significantly.

KEY WORDS: spatial statistics, robust variogram, scale estimation, M-estimator, influence function, breakdown point.

INTRODUCTION

Variogram estimation is a crucial stage of spatial prediction, because it determines the kriging weights. It is important to have a variogram estimator which remains close to the true underlying variogram, even if outliers (faulty observations) are present in data. Otherwise kriging can produce noninformative maps. Experience from a broad spectrum of applied sciences shows that measured data contains as a rule between 10–15% of outlying values (Hampel, 1973) due to gross errors, measurement mistakes, faulty recording, etc. This proportion can even go up to 30% (Huber, 1977). One might argue that any reasonable exploratory data analysis would identify outliers in the data, for example by examining **h**-scatterplots. However, this approach is subjective and informal. Furthermore, existence of exploratory techniques does not supersede the utility of robust techniques. In this paper, we advocate the use of estimators which take account of all the available information in data.

Let us consider a spatial stochastic process $\{Z(\mathbf{x}): \mathbf{x} \in D\}$, where D is a fixed subset of $\mathbb{R}^d$, $d \geq 1$. Assume that this process is ergodic and satisfies the

[1] Received 6 February 1997; revised 30 May 1997.

[2] Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland. e-mail: genton@dma.epfl.ch

0882-8121/98/0200-0213$15.00

hypothesis of intrinsic stationarity given by

$$\text{(a) } E(Z(\mathbf{x})) = \mu = \text{constant}, \qquad \forall \mathbf{x} \in D$$

$$\text{(b) } \operatorname{Var}(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) = 2\gamma(\mathbf{h}), \qquad \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in D$$

where $2\gamma(\mathbf{h})$ is the variogram. This is a very simple model which can be used in practice after detrending data (Cressie, 1991) or in some cases even directly. Let $\{Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)\}$ be a sample of such a spatial process. The classical variogram estimator proposed by Matheron (1962), based on the method of moments, is

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{N_{\mathbf{h}}} \sum_{N(\mathbf{h})} (Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2, \qquad \mathbf{h} \in \mathbb{R}^d \tag{1}$$

where $N(\mathbf{h}) = \{(\mathbf{x}_i, \mathbf{x}_j): \mathbf{x}_i - \mathbf{x}_j = \mathbf{h}\}$ and $N_{\mathbf{h}}$ is the cardinality of $N(\mathbf{h})$. This estimator is unbiased, but behaves poorly if there are outliers in the data. One single outlier can destroy this estimator completely. For that reason, Cressie and Hawkins (1980) proposed a more robust estimator for gaussian independent data:

$$2\hat{\gamma}(\mathbf{h}) = \left[\frac{1}{N_{\mathbf{h}}} \sum_{N(\mathbf{h})} |Z(\mathbf{x}_i) - Z(\mathbf{x}_j)|^{1/2}\right]^4 \Big/ \left(0.457 + \frac{0.494}{N_{\mathbf{h}}}\right), \qquad \mathbf{h} \in \mathbb{R}^d \tag{2}$$

where the denominator corrects for bias under gaussianity. However, this estimator can also be destroyed by a single outlier in the data and is, therefore, not really a solution to the problem.

To view variogram estimation as a problem of identifying the scale at various lags (Cressie, 1991) is intuitively appealing and opens up new perspectives. By a scale estimator of a sample $\{V_1, \ldots, V_n\}$ we mean any positive function $S_n(V_1, \ldots, V_n)$ which satisfies

$$S_n(\alpha V_1 + \beta, \ldots, \alpha V_n + \beta) = |\alpha| S_n(V_1, \ldots, V_n) \tag{3}$$

$\forall \alpha \in \mathbb{R}$, $\forall \beta \in \mathbb{R}$. In effect, the stochastic process of differences at lag $\mathbf{h}$, $V(\mathbf{h}) = Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$, has zero expectation and a variance of $2\gamma(\mathbf{h})$. Thus, if $\{V_1(\mathbf{h}), \ldots, V_{N_{\mathbf{h}}}(\mathbf{h})\}$ is the sample of $V(\mathbf{h})$ corresponding to the sample $\{Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)\}$ of Z, Matheron's classical variogram estimator takes the form

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{N_{\mathbf{h}}} \sum_{i=1}^{N_{\mathbf{h}}} V_i(\mathbf{h})^2, \qquad \mathbf{h} \in \mathbb{R}^d \tag{4}$$

i.e., it is simply the classical estimator of the sample variance of $\{V_1(\mathbf{h}), \ldots, V_{N_{\mathbf{h}}}(\mathbf{h})\}$. We are now going to use the theory of M-estimators of scale to derive robustness properties. In the third section, I propose a new variogram estimator based on a highly robust sample scale.

M-ESTIMATORS OF SCALE

Recalling the definition of an M-estimator of scale (Hampel and others, 1986), suppose we have one-dimensional observations $V_1(\mathbf{h}), \ldots, V_{N_\mathbf{h}}(\mathbf{h})$ which are identically distributed according to a distribution from the parametric model $\{F_\sigma;\ \sigma > 0\}$, where $F_\sigma(v) = F(v/\sigma)$. An M-estimator $S_{N_\mathbf{h}}(V_1(\mathbf{h}), \ldots, V_{N_\mathbf{h}}(\mathbf{h}))$ of σ is defined by the implicit equation

$$\sum_{i=1}^{N_\mathbf{h}} \chi(V_i(\mathbf{h})/S_{N_\mathbf{h}}) = 0 \tag{5}$$

and corresponds asymptotically to the statistical functional S defined by

$$\int \chi(v/S(F))dF(v) = 0 \tag{6}$$

where χ is a real, symmetric (even), and sufficiently regular function (Hampel and others, 1986). The influence function of an M-estimator of scale S at a distribution F is well known (Hampel and others, 1986)

$$IF(v, S, F) = \frac{\chi(v/S(F))S^2(F)}{\int v\chi'(v/S(F))dF(v)} \tag{7}$$

The importance of the influence function lies in its heuristic interpretation: it describes the effect on the estimator of an infinitesimal contamination at point v. An important summary value of the influence function is gross-error sensitivity of S at F, defined by

$$\gamma^* = \sup_v |IF(v, S, F)| \tag{8}$$

This quantity measures the worst influence that a small amount of contamination can have on the estimator. It is desirable that γ^* be finite, in which case S is B-robust (bias-robust) at F. Another important robustness property is the breakdown point ε^* of a scale estimator. This indicates how many data points need to be replaced to make the estimator explode (tend to infinity) or implode (tend to zero). In the case of M-estimators of scale, it has been shown (Huber, 1981) that

$$\varepsilon^* = \min\left(\frac{-\chi(0)}{\chi(+\infty) - \chi(0)}, \frac{\chi(+\infty)}{\chi(+\infty) - \chi(0)}\right) \leq \frac{1}{2} \tag{9}$$

The special choice $\chi(v) = |v|^q - \int |v|^q\, dF(v)$, $q > 0$, leads to the so-called L^q M-estimators of scale (Genton and Rousseeuw, 1995), which are shown to be never B-robust, for every value of $q > 0$, that is to say, $\gamma^* = \infty$. Moreover, it is easily seen that $\varepsilon^* = 0\%$, for any value of $q > 0$. A closer look at Equation (4) and the corresponding equation for the Cressie and Hawkins estimator shows that they correspond to the L^2 and $L^{1/2}$ estimators, respectively. Thus, these two estimators are not robust in the sense of the influence function and breakdown point.

A HIGHLY ROBUST VARIOGRAM ESTIMATOR

In the context of scale estimation, Rousseeuw and Croux (1992, 1993) proposed a simple, explicit and highly robust estimator, called $Q_{N_\mathbf{h}}$, defined by

$$Q_{N_\mathbf{h}} = 2.2191\{|V_i(\mathbf{h}) - V_j(\mathbf{h})|;\ i < j\}_{(k)} \tag{10}$$

where the factor 2.2191 is for consistency at the gaussian distribution,

$$k = \binom{[N_\mathbf{h}/2] + 1}{2},$$

and $[N_\mathbf{h}/2]$ denotes the integer part of $N_\mathbf{h}/2$. This means that we sort the set of all absolute differences $|V_i(\mathbf{h}) - V_j(\mathbf{h})|$ for $i < j$ and then compute its kth quantile ($k \approx \frac{1}{4}$ for large $N_\mathbf{h}$). This value is multiplied by the factor 2.2191, thus yielding $Q_{N_\mathbf{h}}$. Note that this estimator computes the kth order statistic of the $\binom{N_h}{2}$ interpoint distances. It is of interest to remark that $Q_{N_\mathbf{h}}$ does not rely on any location knowledge and is therefore said to be location-free. This is in contrast to Matheron's estimator which implicitly makes use of the zero expectation of $V(\mathbf{h})$. Estimator $Q_{N_\mathbf{h}}$ has an $\varepsilon^* = 50\%$ breakdown point, the highest possible value, and a bounded influence function with $\gamma^* = 2.069$ at the standard gaussian distribution. Gaussian asymptotic efficiency (Hampel and others, 1986) of $Q_{N_\mathbf{h}}$ attains 82%, which is close to the 100% of L^2, whereas $L^{1/2}$ reaches only 69.3%. At first sight, estimator $Q_{N_\mathbf{h}}$ appears to need $O(N_\mathbf{h}^2)$ computation time, which would be a disadvantage. However, it can be computed using no more than $O(N_\mathbf{h} \log N_\mathbf{h})$ time and $O(N_\mathbf{h})$ storage, by means of the fast algorithm described in Croux and Rousseeuw (1992).

Using the previous results and definitions of scale estimator $Q_{N_\mathbf{h}}$, define the highly robust variogram estimator to be

$$2\hat{\gamma}(\mathbf{h}) = (Q_{N_\mathbf{h}})^2, \qquad \mathbf{h} \in \mathbb{R}^d \tag{11}$$

Of course, this estimator has the robustness properties of $Q_{N_\mathbf{h}}$.

SIMULATIONS

In order to analyze performances of $Q_{N_\mathbf{h}}$, we carried out a limited simulation study of spatial data in $\mathbb{R}^1$. If data is clean, that is to say without outliers, each of the three previous variogram estimators behaves correctly. In fact, the interesting situations are those with outliers in data, which are more likely to arise in real geostatistical data sets. For that reason, we simulate an equally-spaced gaussian sample of size $n = 200$ from a spherical variogram

$$\gamma(h, a, b, c) = \begin{cases} 0 & \text{if } h = 0 \\ a + b\left(\frac{3}{2}\left(\frac{h}{c}\right) - \frac{1}{2}\left(\frac{h}{c}\right)^3\right) & \text{if } 0 < h \leq c \\ a + b & \text{if } h > c \end{cases} \tag{12}$$

with parameters $a = 1$, $b = 2$, and $c = 15$. Then, we perturb this sample by simply randomly replacing ϵ percent of the data by new values, independently and identically distributed according to a gaussian distribution $N(0, \sigma^2)$. We choose the following situations:

[1] $\epsilon = 0\%$ [2] $\epsilon = 10\%, \sigma = 5$ [3] $\epsilon = 20\%, \sigma = 5$

[4] $\epsilon = 30\%, \sigma = 5$ [5] $\epsilon = 10\%, \sigma = 10$ [6] $\epsilon = 10\%, \sigma = 20$

Situations [1] to [4] correspond to an increase in the percent of outliers from a given distribution, whereas situations [5] and [6] correspond to a fixed amount of outliers from a distribution with increasing variance. Note that the choice of gaussian distributions is not restrictive. One could use other—more heavy tailed and skewed—distributions and would get the similar behavior for the variogram estimators. Next, on each sample, the variogram is estimated by Matheron's classical estimator, that of Cressie and Hawkins and our highly robust one. Results of the variogram estimations for each situation are shown in Figures 1, 2, and 3, up to lag $h = 100 = n/2$, following the empirical rule of Journel and Huijbregts (1978). On each graph, the true underlying variogram is represented by a solid line.

Effect of outliers in the data is shown by a greater vertical variability of variogram estimates. For Matheron's classical estimator, as well as for that of Cressie and Hawkins, a horizontal deformation is added, which leads to an increase in the range, expressed through parameter c. This phenomena hardly occurs, if at all, for the highly robust variogram estimator. In fact, parameter c, characterizing shape and range of the spherical variogram, is of primary importance for ordinary kriging because of the invariance of weights under linear transformation of the variogram (Genton, 1995). The biggest problem with the two nonrobust estimators is not only that the parameters of the model are estimated incorrectly, but that the shape of the variogram is wrong. Even for small perturbations, these variograms suggest a pure nugget effect model rather than a spherical one. In all cases, the highly robust estimator performs well and estimates the shape of the underlying variogram with great accuracy.

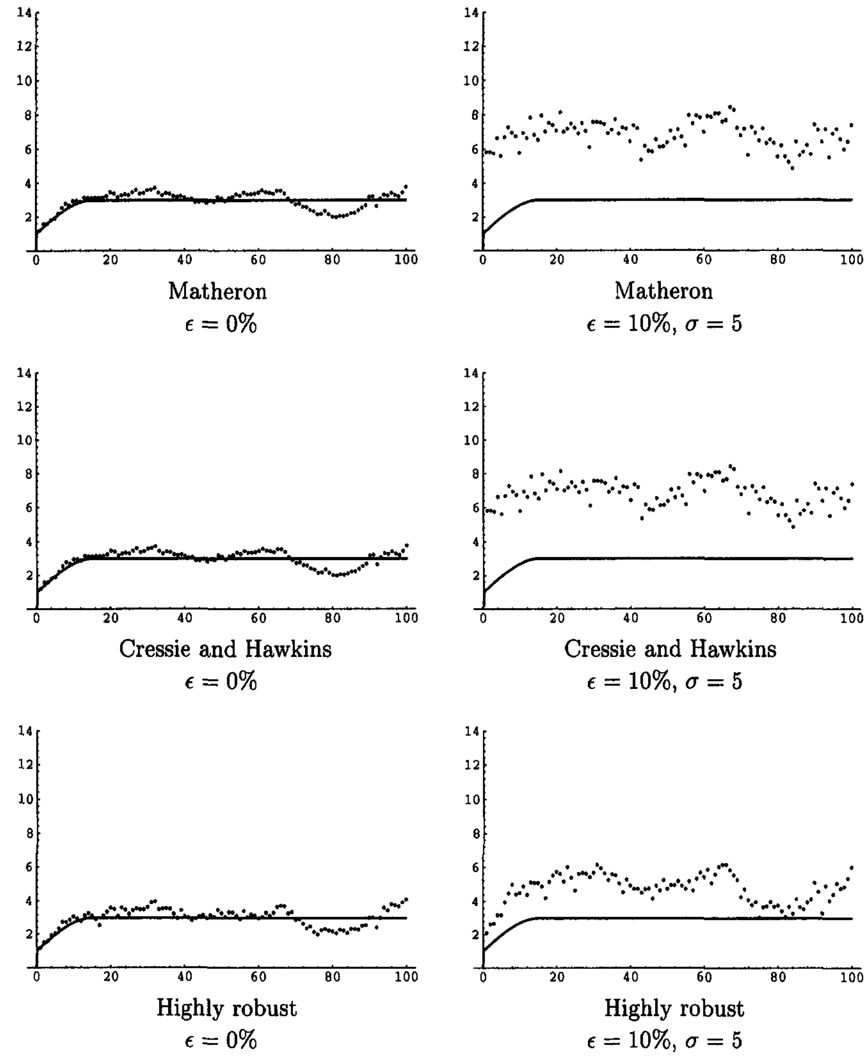

Figure 1. The three rows of this figure show on the left-hand side the estimation from unperturbed data and on the right-hand side estimation from perturbed data with a small amount of contamination (ϵ = 10% of $N(0, 25)$). The three estimators considered are Matheron, Cressie and Hawkins, and the highly robust one. The underlying spherical variogram is indicated by the solid line.

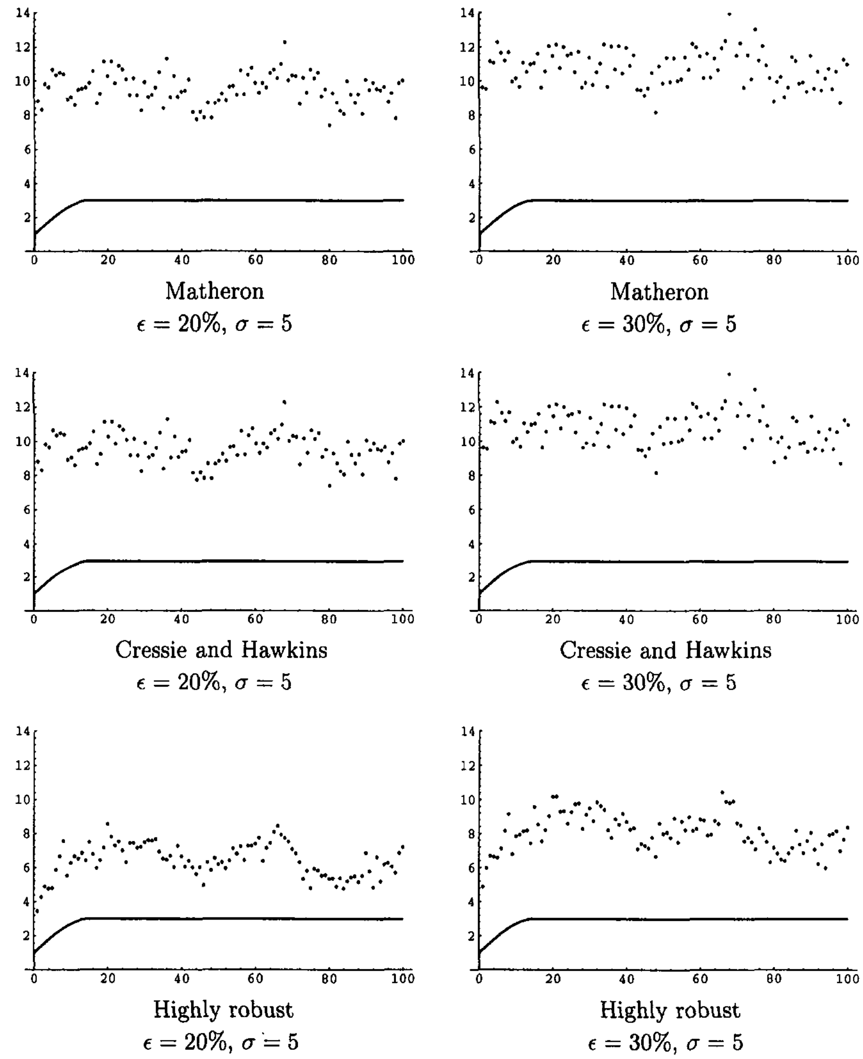

Figure 2. The three rows of this figure show the estimation from perturbed data, on the left-hand side with $\epsilon = 20\%$ and on the right-hand side with $\epsilon = 30\%$ from a $N(0, 25)$ distribution. The three estimators considered are Matheron, Cressie, and Hawkins, and the highly robust one. The underlying spherical variogram is indicated by the solid line.

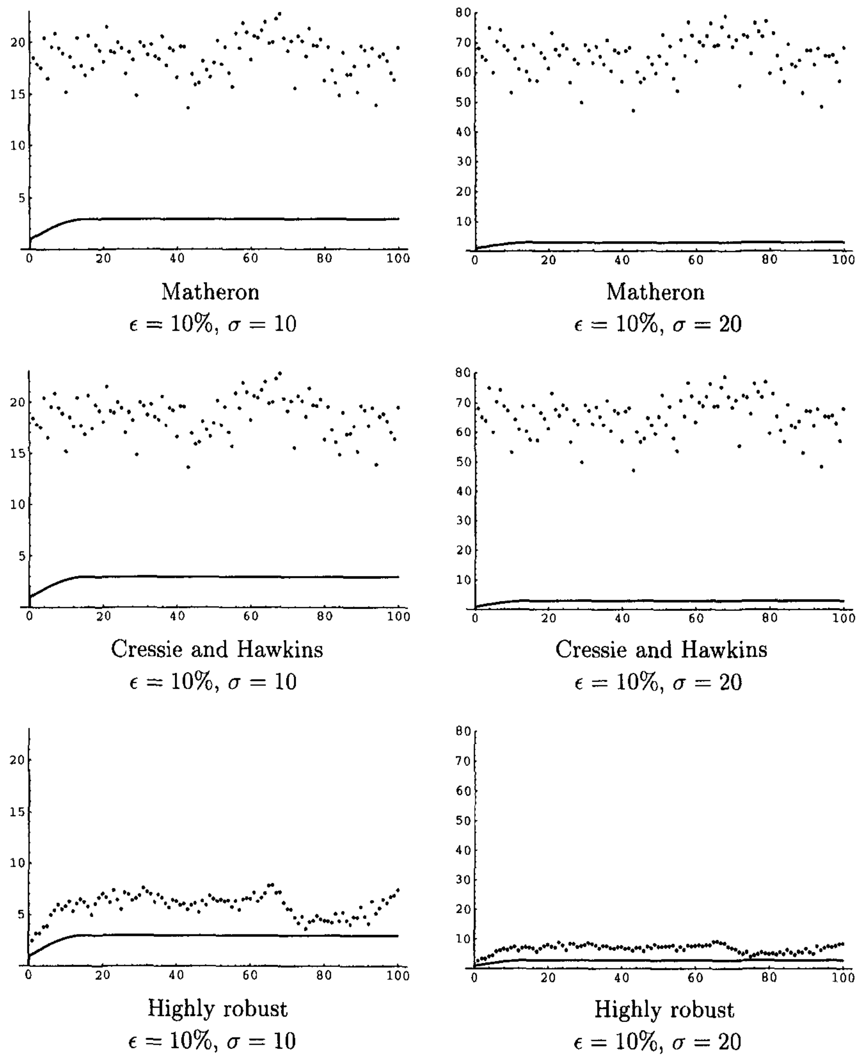

Figure 3. The three rows of this figure show the estimation from perturbed data with a small amount of contamination ($\epsilon = 10\%$), on the left-hand side from a $N(0, 100)$ distribution and on the right-hand side from a $N(0, 400)$ distribution. The three estimators considered are Matheron, Cressie, and Hawkins, and the highly robust one. The underlying spherical variogram, is indicated by the solid line.

CONCLUSIONS

In this paper, variogram estimation has been approached via scale estimation. The theory of M-estimators of scale demonstrates that neither Matheron's classical variogram estimator, nor that of Cressie and Hawkins are robust in the sense of influence function and breakdown point. For that reason, I propose a highly robust variogram estimator by applying a highly robust scale estimator to the problem. A small simulation study of spatial data containing outliers illustrated the behavior of these three estimators and confirmed theoretical results. Instead of using only the highly robust variogram estimator in practice, I rather suggest computing it along with Matheron's estimator. If they are very close to each other, one can assume that outliers had negligible effect. If they are significantly different, one has to think and act with care.

ACKNOWLEDGMENTS

This work contains parts of my PhD dissertation, which was written under the generous guidance of Prof. Stephan Morgenthaler. I wish to thank the Swiss National Science Foundation for its financial support.

REFERENCES

Cressie, N., 1991, Statistics for spatial data: John Wiley & Sons, New York, 900 p.

Cressie, N., and Hawkins, D. M., 1980, Robust estimation of the variogram, I: Math. Geology, v. 12, no. 2, p. 115–125.

Croux, C., and Rousseeuw, P. J., 1992, Time-efficient algorithms for two highly robust estimators of scale: Computational Statist., v. 1, p. 411–428.

Genton, M. G., 1995, Robustesse dans l'estimation du variogramme: Bulletin de l'Institut International de Statistique, Beijing, China, v. 1, p. 400–401.

Genton, M. G., and Rousseeuw, P. J., 1995, The change-of-variance function of M-estimators of scale under general contamination: Jour. Comp. Appl. Math., v. 64, p. 69–80.

Hampel, F. R., 1973, Robust estimation, a condensed partial survey: Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, v. 27, p. 87–104.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A., 1986, Robust statistics, the approach based on influence functions: John Wiley & Sons, New York, 502 p.

Huber, P. J., 1977, Robust statistical procedures: Soc. Industrial and Applied Mathematics, Philadelphia, 56 p.

Huber, P. J., 1981, Robust statistics: John Wiley & Sons, New York, 308 p.

Journel, A. G., and Huijbregts, Ch. J., 1978, Mining geostatistics: Academic Press, London, 600 p.

Matheron, G., 1962, Traité de géostatistique appliquée, Tome I: Mémoires du Bureau de Recherches Géologiques et Minières, no. 14, Editions Technip, Paris, 333 p.

Rousseeuw, P. J., and Croux, C., 1992, Explicit scale estimators with high breakdown point, *in* Dodge, Y., ed., L_1 *Statistical analyses and related methods*: North-Holland, Amsterdam, p. 77–92.

Rousseeuw, P. J., and Croux, C., 1993, Alternatives to the median absolute deviation: Jour. Am. Stat. Assoc., v. 88, no. 424, p. 1273–1283.