SOME ASPECTS OF ROBUSTNESS

Donald E. MYERS*

CONTENTS

ABSTRACT 64

A - INTRODUCTION 65

B - ROBUSTNESS 66

C - NUMERICAL EXAMPLES 70

D - SAMPLING CONFIGURATIONS 71

E - ROBUST OR NON-ROBUST 72

RESUME 64

F - CIRCULAR PATTERNS 78

G - SUMMARY 78

H - ACKNOWLEDGEMENT 78

REFERENCES 79

ILLUSTRATIONS

Figure 1 - 73

Figure 2 - 73

----

*Department of Mathematics - University of Arizona - TUCSON, AZ 85721 - USA.

"ETUDES GEOSTATISTIQUES", Séminaire CFSG - 17-18 Juin 1985 - Fontainebleau, France, in Sci. de la Terre, Sér. Inf., n° 24

ABSTRACT

There are several parameters in a kriging system : the variogram, the sampling configuration, the size and shape of the block and its relationship to the sample pattern, etc. The robustness of kriging with respect to these parameters is considered. Armstrong and Diamond have considered a particular form of neighborhood for variograms. Other definitions are considered as well as the distinction between the weights vector with and without the Lagrange multiplier. The problem is most easily seen in the context of continuity. Numerical examples are given.

RESUME

Quelques aspects de la Robustesse

Un système de krigeage fait intervenir plusieurs paramètres : le variogramme, l'implantation spatiale des sondages, la taille et la géométrie du bloc à estimer ainsi que sa position par rapport au plan d'échantillonnage, etc. Dans cet article, la robustesse d'un système de krigeage est étudiée par rapport à ces paramètres. Armstrong et Diamond l'ont utilisé dans le cas d'une forme de voisinage d'estimation particulier. D'autres types de voisinage d'estimation, de plan d'échantillonnage et formes de blocs sont étudiés ici, dans le cas de pondérateurs tenant compte ou non des paramètres de Lagrange. Il est plus facile de poser ce problème dans le cas continu. Des exemples numériques sont donnés à titre d'illustration.

A - INTRODUCTION

Robustness is a desirable property of estimators which has multiple meanings. In the context of classical statistics, it refers to the insensitivity of the estimator to changes in the distribution of the population sampled and/or to the presence of outliers see Hampfel (1971). Huber (1981) has discussed various modifications of estimators for the mean that make the estimator more robust in the classical sense.

In the practice of geostatistics, estimation is a multiple stage process and if the estimation is non-robust or instable at any one step the cumulative process may be constable. Consider the standard problem of utilizing data in 1, 2 or 3-space to estimate spatial averages over intervals, areas or volumes. Whether Simple, Ordinary, Universal, Disjunctive Indicator Kriging or Co-Kriging is used, one must first estimate or model the variogram (respectively generalized covariance). It is "well-known" that Kriging is relatively robust with respect to the variogram model. It is also well-known that the standard estimator for the variogram is not robust. Armstrong and Delfiner (1980) utilized quartile estimatos to minimize the effect of outliers. Cressie and Hawkins (1980) have shown that under normality assumptions that other estimators are more robust. However there is a distinction between estimating the values of the variogram at each of a finite number of lags and modelling the variogram which must satisfy positive definiteness conditions. BLUEPACK (1983) incorporates an algorithm for estimating the coefficients in a polynomial generalized covariance. Kitanidis (1983) has proposed a maximum likelihood estimator for these coefficients.

Whichever variogram or generalized covariance estimator is used, it is common practice to use cross-validation to verify the appropriateness of the model selected. The efficacy of cross-validation in turn is dependent on the robustness of Kriging. Unfortunately cross-validation does not result in a single characteristic to be maximized or minimized. Even if it did, there would remain the question of whether cross-validation is sufficiently sharp to distinguish between variogram models that are close. Hawkins and Cressie (1984) have proposed an iterative process as has Neumann (1984). Armstrong (1984), Diamond and Armstrong (1984) have examined the more fundamental question of the robustness of Kriging. One may examine the Kriging estimator in at least two ways (1) the set of weights (2) the estimated value. Since the weights vector is itself data free it would seem preferable to consider only the weights but ultimately it is the estimated value that is of interest. In either case it is possible to view robustness in the sense of continuity.

If X is the solution vector of a Kriging system for the variogram model γ and ΔX is the incremental change in X resulting from the change $\Delta\gamma$ then Diamond and Armstrong (1984) (hereafter referred to as D/A) have obtained bounds for $\|\Delta X\|$ and $\|\Delta X\|/\|X\|$ for a particular choice of a neighborhod of γ and for certain matrix norms. This work extends that of D/A in several respects. The solution considered by D/A includes the Lagrange multiplier(s) whereas the estimated values do not directly depend on the multiplier(s). The neighborhood definition utilized by D/A is non-symmetric and is too restrictive, their definition can be extended and other neighborhood definitions will be considered. The question of robustness will first be placed in a more general context.

B - ROBUSTNESS

In examining the Kriging equation for the block estimator, there are at least four "parameters" that determine the Kriging weights.

1 - THEORETICAL

a. the variogram or generalized covariance model

b. the geometrical configuration and the number of sample locations used in the estimation

c. the block shape and the geometrical relationship of the block to the sample location configuration

d. the order of the drift (and the choice of drift functions)

While not present in the Kriging equations in the same sense as the theoretical parameters the solution vector is also affected by at least two computational "parameters".

a. the mesh and pattern of the grid of points in the block used to numerically integrate $\overline{\gamma}$

b. the method used to solve the system of equations, e.g. Gaussian Reduction, Pivoting, partitioning etc.

Finally there is the question of the support of the data, because the Kriging equations utilize point support values, i.e. the weights are associated with point values, that is, there is the question of what changes should be made in the weight vector to reflect non-punctual support.

Let $x_1,\ldots,x_n$ denote the sample locations and $Z(x_1),\ldots,Z(x_n)$ the (punctual) data. Let $\gamma(h)$ denote the variogram model (or generalized covariance) then the Kriging equations may be written in the form

$$\Gamma\overline{\lambda} + F\overline{\mu} = \Gamma_0 \qquad (1)$$

$$F^T\overline{\lambda} = F_0 \qquad (2)$$

where

$$\Gamma = \begin{bmatrix} \gamma_{11},\ldots\gamma_{1n} \\ \vdots \qquad \vdots \\ \gamma_{n1},\ldots\gamma_{nn} \end{bmatrix}, \quad \gamma_{ij} = \gamma(x_i - x_j) \qquad (3)$$

$$\Gamma_0 = \begin{bmatrix} \overline{\gamma}_1 \\ \vdots \\ \overline{\gamma}_n \end{bmatrix}, \quad \overline{\gamma}_i = \frac{1}{v}\int_v \gamma(x_i - x)^{dx} \qquad (4)$$

$$\overline{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}, \quad \overline{\mu} = \begin{bmatrix} \mu_0 \\ \vdots \\ \mu_k \end{bmatrix} \qquad (5)$$

$$F = \begin{bmatrix} 1 & \\ 1 & \\ \vdots & ? \\ 1 & \end{bmatrix} \qquad (6)$$

With respect to the effects of the four theoretical parameters, changes in Γ, Γ_0, $\overline{\lambda}$, $\overline{\mu}$, F, F_0 could be written in the form

$$\Gamma \rightarrow \Gamma + \Delta\Gamma$$

$$\Gamma_0 \rightarrow \Gamma_0 + \Delta\Gamma_0$$

$$\overline{\lambda} \rightarrow \overline{\lambda} + \Delta\lambda \qquad (7)$$

$$\overline{\mu} \rightarrow \overline{\mu} + \Delta\overline{\mu}$$

$$F \rightarrow F + \Delta F$$

$$F_0 \rightarrow F_0 + \Delta F_0$$

For each of these components of the Kriging equations, the change may include the effective deletion of a row and/or column, i.e. the replacement by a row/column of zeros. For example, if $Z(x)$ is an Intrinsic Random Function order zero then F would consist of 1's in column one and zeros in the remaining columns whereas for an order one IRF in three space there would be three more non-zero columns. In a similar fashion if moving neighborhoods are used, Γ would have one or more rows and columns with all zeros.

If $\overline{\lambda}$ is written in the form

$$\overline{\lambda} = G(\Gamma,\Gamma_0,F,F_0) \qquad (8)$$

then

$$\Delta\overline{\lambda} = G(\Gamma+\Delta\Gamma,\ \Gamma_0+\Delta\Gamma_0,\ F+\Delta F,\ F_0+\Delta F\ ) - G(\Gamma,\Gamma_0,F,F_0) \qquad (9)$$

For a data vector $\overline{Z} = [Z(x_1),\ldots,Z(x_n)]$

$$Z^*(x_0) = \overline{Z}\ \overline{\lambda} \qquad (10)$$

and

$$\Delta Z^*(x_0) = \overline{Z}\ \Delta\overline{\lambda} \qquad (11)$$

Note that in the case of $\Delta F = 0$, $\overline{Z}\ \Delta\lambda$ is an authorized linear combination, that is, $F^T\Delta\lambda = 0$.

This suggests a possible test statistic for comparing one variogram model against another. Since $\overline{Z}\ \Delta\lambda$ is an authorized linear combination, the variance of $\overline{Z}\Delta\lambda$ is $\Delta\lambda^T\Gamma\Delta\lambda$ or $\Delta\lambda^T(\Gamma+\Delta\Gamma)\Delta\lambda$. If $\Delta\Gamma$ is small in an appropriate sense then the quotient of these two variances should be close to one.

The distinction between X and $\overline{\lambda}$ is diminished when considering the Kriging variance and we shall limit consideration to some numerical examples.

3 - NEIGHBORHOODS FOR VARIOGRAMS

Let **A** be the family of valid variograms in R^k.

Definition 1. (D/A) For $\gamma \in A$ and $0 < \delta < 1$

$$N_\delta(\gamma) = \{g \,|\, g\varepsilon\mathbf{A}, \sup_{|h|} \left|\frac{g(h)}{\gamma(h)} - 1\right| < \delta\} \tag{12}$$

Diamond and Armstrong have shown that this is a very useful neighborhood definition and by using the condition number of the variogram matrix obtain bounds for $\|\Delta X\|$, $\|\Delta X\|/\|X\|$ where $X^T = [\overline{\lambda}\ \overline{\mu}]^T$.

Several comments are in order

It is not necessary to assume that γ, g are isotropic

The definition is not quite symmetric in γ, g, $g \in N_\delta(\gamma)$ implies $\gamma \in N_{\delta'}(g)$ where $\delta' = \delta/(1+\delta)$

The definition requires that g, γ coincide for $h = 0$ and that the values be close for large $|h|$. For spherical models this means the nuggets must be the same and the sills the same or very close

Since moving or local neighborhoods are commonly used the definition can be weakened by using

$$N_{\delta,r}(\gamma) = \{g \,|\, g\varepsilon\mathbf{A}, \sup_{|h|\leq r} \left|\frac{g(h)}{\gamma(h)} - 1\right| < \delta\} \tag{13}$$

This would permit a linear model to be in a neighborhood of a Spherical whereas the original definition would not.

Definition 2 Let $\gamma \varepsilon \mathbf{A}$ and $\varepsilon > 0$.

$$M_{\varepsilon,r}(\gamma) = \{g \mid g \varepsilon \mathbf{A}, \sup_{|h| \leq r} |\gamma(h) - g(h)| < \varepsilon\} \qquad (14)$$

Yakowitz and Szidarovsky (1984) have given a bound on $\|\Delta\lambda\|$ using this neighborhood definition. Moreover it would seem a more natural definition except that in general it is the values of the variogram for short lags that are most critical.

C - NUMERICAL EXAMPLES

D/A have given several examples wherein a 1×1 block is to be estimated using the sample locations at (-0.4,0), (0.4,0) and (0.39,0.1)

1-γ, g ARE GAUSSIAN

γ: a = 1, c = 1

g: a = 1.1, c = 1

Points	λ: γ	λ: g
(-0.4, 0)	.4982	.4985
(0.4, 0)	.3970	.4112
(0.39, 0.1)	.1048	.0903
μ	-.0190	-.0140

Table 1

$$\frac{\|\Delta X\|}{\|X\|} = .0324, \quad \frac{\|\Delta\lambda\|}{\|\lambda\|} = .0314$$

where X is the solution of $BX = B_0$ as given in D/A. $\overline{\lambda}$ the solution of the system (1), (2).

The bound given in D/A for $\|\Delta X\|/\|X\|$ is

$$\frac{\|\Delta X\|}{\|X\|} \leq \frac{2\delta\lambda(B)}{1-\delta\lambda(B)} \tag{15}$$

but $\delta = .1$ and hence $k(B) \approx 100$ and the bound is not applicable.

2-γ, g ARE GAUSSIAN

γ, a = 1, sill = 1, nugget = 0

g, a = 1, sill = .99, nugget = .1

	λ	
points	γ	g
(-0.4, 0)	.4982	.4954
(0.4, 0)	.3970	.3253
(0.39, 0.1)	.1048	.1793
μ	-.0190	-.0230

Table 2

$$\frac{\|\Delta X\|}{\|X\|} = .1602, \quad \frac{\|\Delta\lambda\|}{\|\lambda\|} = .1602$$

D – SAMPLING CONFIGURATIONS

One may consider moving the block to be estimated. The latter is computationally simpler and is used in the following examples

1-γ IS GAUSSIAN

a = 1, c = 1

	λ				
Block Center	(.4,0)	(-.39,.1)	(-.4,0)	μ	σ^2
(0,0)	.498	.108	.393	-.019	.017
(.1,.1)	.595	.688	-.283	-.028	.0205
(-.1,-.1)	.403	-.550	1.147	-.032	.0232
(.1,-.1)	.612	-.474	.862	-.032	.0247
(-.1,.1)	.384	.776	-.160	-.027	.0178

Table 3

Note that only small pertubations of the block result in sign reversals in the weights.

4 γ IS GAUSSIAN,

$a = 1.1,\ c = 1.0$

	λ				
Block Center	(.4,0)	(-.39,.1)	(-.4,0)	μ	σ^2
(0,0)	.499	.094	.408	-.014	.0128
(.1,.1)	.598	.728	-.326	-.022	.0149
(-.1,-.1)	.400	-.612	1.212	-.026	.0169
(.1,-.1)	.615	-.541	.906	-.026	.0185
(-.1,.1)	.381	.808	-.189	-.022	.0115

Table 4

The Gaussian model is particularly sensitive to small changes in the variogram as well as in the sample configuration.

E - ROBUST OR NON-ROBUST

When collecting soil samples for monitoring for the presence of pollutants, it has often been the practice of the EPA (Environmental Protection Agency, USA) to sample each block to be estimated by using locations regularly spaced on the circumference of circle centered in the block. For simplicity consider a 1×1 block with a circle of radius r centered in the

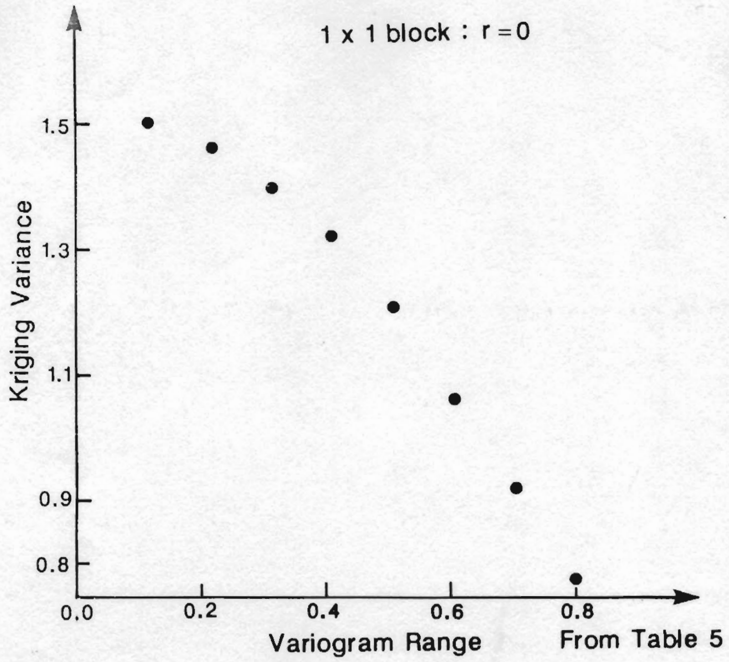

Figure 1

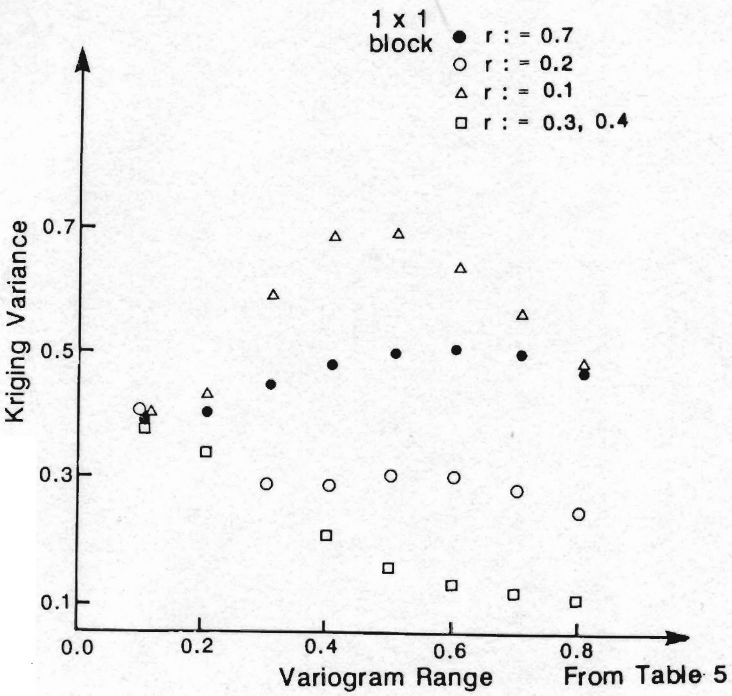

Figure 2

Table 5 - Kriging Variances

Spherical model, sill = 1.5
8 × 8 grid for numerical integration
4 sample points on circle, radius r, center at center of block

radius r	variogram range	block size 1.0 × 1.0	1.0 × 1.2	
0.0	.10	1.520	1.522	1.523
	.20	1.468	1.471	1.473
	.30	1.408	1.418	1.427
	.40	1.324	1.343	1.358
	.50	1.212	1.241	1.265
	.60	1.067	1.111	1.148
	.80	0.775	0.842	0.888
0.1	0.10	0.389	0.390	0.390
	0.20	0.429	0.432	0.436
	0.30	0.597	0.607	0.615
	0.40	0.685	0.703	0.718
	0.50	0.690	0.718	0.742
	0.60	0.638	0.676	0.710
	0.70	0.561	0.604	0.646
	0.80	0.485	0.528	0.571
0.2	0.10	0.382	0.384	0.388
	0.20	0.341	0.344	0.348
	0.30	0.286	0.296	0.305
	0.40	0.287	0.304	0.319
	0.50	0.302	0.326	0.347
	0.60	0.301	0.350	0.358
	0.70	0.281	0.312	0.344
	0.80	0.251	0.283	0.316
0.3	0.10	0.382	0.389	0.394
	0.20	0.340	0.344	0.348
	0.30	0.284	0.293	0.303
	0.40	0.217	0.229	0.240
	0.50	0.165	0.170	0.191
	0.60	0.138	0.154	0.171
	0.70	0.124	0.140	0.159
	0.80	0.116	0.133	0.151

Table 6 - Kriging Variances

Spherical model, sill = 1.5
8 × 8 grid for numerical integration
8 sample points on circle, radius r, center at center of block

radius r	variogram range	block size 1.0 × 1.0	1.0 × 1.1	1.0 × 1.2
0.1	.10	0.215	0.214	0.215
	.20	0.367	0.370	0.374
	.30	0.557	0.567	0.574
	.40	0.655	0.673	0.687
	.50	0.665	0.693	0.717
0.2	0.10	0.199	0.202	0.204
	0.20	0.182	0.186	0.189
	0.30	0.210	0.220	0.229
	0.40	0.229	0.247	0.262
	0.50	0.254	0.278	0.300
0.3	0.10	0.191	0.191	0.195
	0.20	0.152	0.155	0.159
	0.30	0.124	0.132	0.141
	0.40	0.109	0.128	0.135
	0.50	0.079	0.103	0.119
0.4	0.10	0.197	0.204	0.203
	0.20	0.154	0.158	0.161
	0.30	0.108	0.113	0.119
	0.40	0.086	0.090	0.096
	0.50	0.079	0.080	0.084
0.5	0.10	0.204	0.197	0.199
	0.20	0.172	0.168	0.168
	0.30	0.146	0.139	0.138
	0.40	0.124	0.112	0.107
	0.50	0.124	0.109	0.100

Table 7 - Kriging Variances

Spherical model, sill = 1.5
4 × 4 grid for numerical integration
5 sample points on circle, radius r, center at center of block

radius r	variogram range	block size 1.0 × 1.0	1.0 × 1.1	1.0 × 1.2
0	.2	1.579	1.588	1.593
	.3	1.443	1.458	1.478
	.4	1.364	1.375	1.387
	.5	1.246	1.279	1.384
	.6	1.090	1.114	1.178
	.7	0.935	0.985	1.034
	.8	0.801	0.850	0.399
.1	.2	0.453	0.451	0.476
	.3	0.610	0.622	0.636
	.4	0.696	0.715	0.731
	.5	0.703	0.734	0.758
	.6	0.641	0.683	0.723
	.7	0.561	0.607	0.651
	.8	0.480	0.525	0.570
.2	.2	0.303	0.298	0.294
	.3	0.268	0.275	0.280
	.4	0.278	0.294	0.306
	.5	0.287	0.313	0.334
	.6	0.289	0.317	0.344
	.7	0.270	0.301	0.331
	.8	0.238	0.270	0.303
.3	.2	0.336	0.346	0.349
	.3	0.240	0.253	0.265
	.4	0.168	0.185	0.195
	.5	0.145	0.161	0.178
	.6	0.116	0.133	0.153
	.7	0.110	0.126	0.144
	.8	0.103	0.121	0.140
.4	.2	0.284	0.298	0.317
	.3	0.228	0.238	0.253
	.4	0.175	0.181	0.192
	.5	0.125	0.125	0.137
	.6	0.102	0.104	0.112
	.7	0.087	0.089	0.095
	.8	0.073	0.073	0.078

Table 8 - Kriging Variances

Spherical model, sill = 1.5
8 × 8 grid for numerical integration
5 sample points on circle, radius r, center at center of block

radius r	variogram range	block size 1.0 × 1.0	1.1 × 1.0	1.2 × 1.0
	.1	1.520	1.523	1.523
	.2	1.478	1.471	1.473
.1	.1	0.301	0.300	0.299
	.2	0.398	0.401	0.404
.2	.1	0.297	0.300	0.305
	.2	0.262	0.266	0.271
	.3	0.244	0.254	0.263
	.4	0.257	0.274	0.289
	.5	0.275	0.300	0.323
	.6	0.280	0.309	0.337
	.7	0.264	0.296	0.324
	.8	0.235	0.269	0.302
.3	.1	0.305	0.304	0.300
	.2	0.265	0.268	0.270
	.3	0.207	0.216	0.223
	.4	0.147	0.160	0.172
	.5	0.125	0.139	0.154
	.6	0.104	0.114	0.135
	.7	0.097	0.113	0.131
	.8	0.093	0.110	0.128
.4	.1	0.313	0.315	0.316
	.2	0.267	0.270	0.273
	.3	0.221	0.225	0.232
	.4	0.172	0.174	0.180
	.5	0.121	0.122	0.127
	.6	0.100	0.099	0.103
	.7	0.086	0.085	0.088
	.8	0.069	0.068	0.071
.5	.1	0.309	0.313	0.310
	.2	0.283	0.282	0.281
	.3	0.262	0.254	0.252
	.4	0.238	0.227	0.220
	.5	0.211	0.195	0.186
	.6	0.176	0.159	0.148
	.7	0.158	0.140	0.128
	.8	0.149	0.131	0.118

block, n points uniformly spaced on the circle and γ a spherical model with no nugget, sill 1.5 and range a. It is easily seen that because of the symmetry all weights are $1/n$ and hence as r, a change the weight vector changes not at all.

While the sampling pattern changes with r, it does so in an isotropic way. As the range on the variogram changes, the weights do not. In contrast the Kriging variance does change. Judged in this context, the Kriging estimator may be too robust or not robust at all since clearly the estimated value changes unless the random function is a constant.

F – CIRCULAR PATTERNS

The following tables and figures illustrate the behavior of the Kriging variance when a circular pattern is used. Note especially the behavior for $r = 0$, where the variance is montone as contrasted with the case of $r > 0$. This suggests that the ability of cross-validation to distinguish the "better" of several variogram models is effected by the local neighborhood search process in Kriging.

G – SUMMARY

The robustness of the Kriging estimator and the Kriging variance as a function of the variogram model, sampling pattern configuration, block size and shape, order of drift is not uniform. For some variogram types and sampling configurations the estimator is very non-robust. The robustness of the estimator may be gauged either in terms of the weight vector or the estimated value. The efficacy of cross-validation is dependent on the robustness and the non-robustness of the Kriging estimator.

ACKNOWLEDGEMENT

Portions of this work were completed while visiting at the Centre de Géostatistique. Appreciation is expressed to M. Armstrong for the use of the data in Tables 1-4.

REFERENCES

Armstrong, M. (1984) "Improving the estimation and modelling of the variogram", in Geostatistics for Natural Resource Characterization, (ed.) G. Verly, et. al., D. Reidel Publishing Co., Dordrecht

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

BLUEPACK, (1983) Centre de Geostatistique, Fontainebleau, France

Cressie, N. and Hawkins, D. H. (1980) "Robust Estimators of the variogram: I: MATHEMATICAL GEOLOGY 12, 115-126.

Diamond, Phil and Armstrong, Margaret, (1984) "Robustness of Variograms and Conditioning of Kriging Matrices", MATHEMATICAL GEOLOGY, 16, No. 8, 809-822.

Hampfel, F. R. (1971) "A more qualitative definition of robustness" ANN. MATH. STAT. 42, 1887-1896.

Hawkins, Douglas M. and Cressie (1984) "Robust Kriging-A proposal", MATHEMATICAL GEOLOGY 16, No. 1, 3-18.

Huber, P. J. (1981) "Robust Statistics", John Wiley & Sons, New York

Kitanidis, P. K. (1983) "Statistical estimation of polynomial generalized covariance functions and hydrologic applications", WATER RESOURCE RES. 19(4), 909-921

Neuman, Shlomo P. (1984) "Role of Geostatistics in Subsurface Hydrology" in Geostatistics for Natural Resource Characterized (ed) G. Verly et. al., D. Reider Publishing Co.

Myers, Donald E., and Armstrong, Margaret (1985) "Robustness of Kriging; Variogram Modelling and Applications", in preparation

Myers, Donald E. (1984) "Determination of Sample Size and Sampling patterns for Soil Sampling", Final Report LV84-89 ESML-LV, EPA

Yakowitz, S. J. and Szidarovzky F. (1984) "A comparison of Kriging with Nonparametric Regression Methods", J. MULTIVARIATE ANALYSIS, 15