# Fitting Matrix-Valued Variogram Models by Simultaneous Diagonalization (Part I: Theory)<sup>1</sup>

Tailiang Xie<sup>2</sup> and Donald E. Myers<sup>3</sup>

Suppose that  $\overline{Z}(x_1), \ldots, \overline{Z}(x_n)$  are observations of vector-valued random function  $\overline{Z}(x)$ . In the isotropic situation, the sample variogram  $\gamma^*(h)$  for a given lag h is

$$\overline{\gamma}^{*}(h) = \frac{1}{2N(h)} \sum_{\mathbf{x}(h)} (\overline{Z}(x_{i}) - \overline{Z}(x_{i})) (\overline{Z}(x_{i}) - \overline{Z}(x_{i}))^{T}$$

where s(h) is a set of paired points with distance h and N(h) is the number of pairs in s(h).

For a selection of lags  $h_1, h_2, \ldots, h_k$  such that  $N(h_i) > 0$ , we obtain a k-tuple of (semi) positive definite matrices  $\overline{\gamma}^*(h_1), \ldots, \overline{\gamma}^*(h_k)$ . We want to determine an orthonormal matrix B which simultaneously diagonalizes the  $\overline{\gamma}^*(h_1), \ldots, \overline{\gamma}^*(h_k)$  or nearly diagonalizes them in the sense that the sum of squares of off-diagonal elements is small compared to the sum of squares of diagonal elements is small compared to the sum of squares of diagonal elements. If such a B exists, we linearly transform  $\overline{Z}(x)$  by  $\overline{Y}(x) = B\overline{Z}(x)$ . Then, the resulting vector function  $\overline{Y}(x)$  has less spatial correlation among its components than  $\overline{Z}(x)$  does. The components of  $\overline{Y}(x)$  with little contribution to the variogram structure may be dropped, and small cross-variograms fitted by straightlines. Variogram models obtained by this scheme preserve the negative definiteness property of variograms (in the matrix-valued function sense). A simplified analysis and computation in cokriging can be carried out. The principles of this scheme are presented in this paper.

KEY WORDS: variogram modeling, positive definite function, matrix diagonalization, algorithm.

### INTRODUCTION

If Z(x) is an intrinsic spatial random function, then

$$\gamma(h) = \frac{1}{2} E [Z(x + h) - Z(x)]^2$$

exists and does not depend on x. The  $\gamma(h)$  quantifies the spatial correlation of Z(x). The sample variogram for a given lag h is

$$\gamma^{*}(h) = \frac{1}{2N(h)} \sum_{s(h)} (Z(x_{i}) - Z(x_{j}))^{2}$$

867

<sup>&</sup>lt;sup>1</sup>Received 6 December 1994; accepted 3 March 1995.

<sup>&</sup>lt;sup>2</sup>Biometry Section, Arizona Cancer Center, University of Arizona, Tucson, Arizona, 85724.

<sup>&</sup>lt;sup>3</sup>Department of Mathematics, University of Arizona, Tucson, Arizona 85721.

where

$$s(h) = \{(i, j) : dis(x_i, x_i) = h\}$$
 and  $N(h) = |s(h)|$ 

Of course, the *h* should be selected so that  $s(h) \neq \emptyset$ , i.e., N(h) > 0. Note that points with a given distance are clumped together and the estimator is unbiased.

If Z(x) is an intrinsic vector spatial random function, then the matrix sample variogram would be

$$\overline{\gamma}^*(h) = \frac{1}{2N(h)} \sum_{s(h)} \left( \overline{Z}(x_i) - \overline{Z}(x_j) \right) \left( \overline{Z}(x_i) - \overline{Z}(x_j) \right)^T$$

Although there are various methods to obtain a sample variogram, the sample variogram is not sufficient for spatial prediction schemes, such as kriging and interpolation. A functional form that best represents this sample variogram could be used. However, not every smooth function passing through or close to the sample variogram is a valid variogram model. Recall that the variogram should be a conditionally negative definite function of a certain order. The lack of this property can result in a negative mean-squared error of prediction (Cressie, 1991). Therefore, we have to fit a variogram using a class of functions with the appropriate negative definiteness property.

Usually variogram modeling consist of three stages:

- Compute the sample variogram at different lags.
- Correctly identify parametric functions that are conditionally negative definite. For a scalar random function, we can plot the sample variograms and obtain graphical information on the parametric form of the variogram. Because the family of conditionally negative definite functions is large, correct identification is not easy. In geostatistics, several standard models are used, such as the Gaussian model, Exponential model, or Spherical model. More complicated models can be obtained by nesting these standard models.
- Fit the sample variograms to the parametric form. Different methods have been developed for variogram modeling of a scalar spatial random function. The most widely used technique in geostatistics is the subjective method, by which (with aid of some graphic device) a fit is made to the sample variograms, with adjustments to the model parameters (such as sill, nugget, and range). Obviously, this method can be inaccurate and inefficient. Several goodness-of-fit methods for fitting a best variogram model have been proposed. Weighted least-squares method may be used (Cressie, 1985).

For the vector spatial random function situation, the subjective method is no longer adequate, because of the lack of suitable graphic aids for plotting

868

### Matrix-Valued Variogram Models. I

sample variogram matrices. Correctly identifying a parametric form of matrixvalued variogram thus is more difficult. The major difficulty comes from insufficient knowledge of matrix-valued (conditionally) negative definite functions. Even when we select a parametric form of matrix-valued variogram, it is a complicated procedure to fit a model, because there are many parameters to be estimated.

It is usual to reduce the matrix-valued variogram modeling problem to a scalar variogram modeling problem. Let  $\overline{Z}(x) = (z_1(x), \ldots, z_m(x))^T$ . Then, the matrix-valued variogram is

$$\overline{\gamma}(h) = \frac{1}{2} E[(\overline{Z}(x+h) - \overline{Z}(x)) (\overline{Z}(x+h) - \overline{Z}(x))^T]$$
$$= (\gamma_{rs}(h)), \qquad r, s = 1, \dots, m$$

where

$$\gamma_{rs}(h) = \frac{1}{2} E[(z_r(x+h) - z_r(x)) (z_s(x+h) - z_s(x))]$$

When r = s,  $\gamma_{rr}(h)$  is the variogram of  $z_r(x)$  and when  $r \neq s$ ,  $\gamma_{rs}(h)$  is the cross-variogram of  $z_r(x)$  and  $z_s(x)$ . It is well-known that  $\gamma_{rr}(h)$  (r = 1, ..., m) are conditionally negative definite functions, and the  $\gamma_{rs}(h)$  must satisfy the Cauchy-Schwartz condition

$$|\gamma_{rs}(h)| \leq \sqrt{\gamma_{rr}(h)\gamma_{ss}(h)} \tag{1}$$

Unfortunately, this inequality is only a necessary condition. At this time there is no sufficient condition for  $\gamma_{rs}(h)$   $(r \neq s)$  to be valid. Modeling cross-variograms is even more difficult, because of their incomplete characterization.

In an attempt to avoid this problem, Myers (1982, 1988) proposed instead to model variograms of the sum and difference of  $z_r(x)$ ,  $z_s(x)$ .

Define

$$\gamma_{rs}^{+}(h) = \frac{1}{2} E(z_r(x+h) + z_s(x+h) - z_r(x) - z_s(x))^2$$
(2)

$$\gamma_{rs}^{-}(h) = \frac{1}{2} E (z_r(x+h) - z_s(x+h) - z_r(x) + z_s(x))^2$$
(3)

$$r, s = 1, 2, \ldots, m$$

Note that  $\gamma_{rs}^+(h)$  and  $\gamma_{rs}^-(h)$  are the variograms of  $z_r(x) + z_s(x)$  and  $z_r(x) - z_s(x)$ , respectively. It is easy to see that

$$\gamma_{rs}(h) = \frac{1}{2} (\gamma_{rr}(h) + \gamma_{ss}(h) - \gamma_{rs}^{-}(h)) = \frac{1}{2} (\gamma_{rs}^{+}(h) - \gamma_{rr}(h) - \gamma_{ss}(h))$$
  
$$\gamma_{rs}(h) = \frac{1}{4} (\gamma_{rs}^{+}(h) - \gamma_{rs}^{-}(h))$$

Thus we can use  $\gamma_{rs}^+$  and  $\gamma_{rs}^-$  to estimate each entry  $\gamma_{rs}$  of  $\overline{\gamma}$  individually. Because  $-\overline{\gamma}(h)$  must satisfy the conditional positive definiteness condition and the Cauchy–Schwartz condition is only a necessary condition for positive definiteness, estimating  $\gamma_{rs}(t)$  by the given method may not guarantee consistently positive definiteness (Goovaerts, 1994).

In this paper, we propose an alternate method for variogram modeling by simultaneously diagonalizing the sample variogram matrices. We will see that this method guarantees the negative definiteness of the matrix-valued variogram.

## MOTIVATION

Let  $\overline{Z}(x)$  be an intrinsic random vector function. We only consider the isotropic situation. Then its variogram is a matrix-valued function of *Euclidian* distance:

$$\overline{\gamma}(h) = \frac{1}{2} E[(\overline{Z}(x+h) - \overline{Z}(x)) (\overline{Z}(x+h) - \overline{Z}(x))^T]$$

As a (matrix-valued) function, it is conditional negative definite of order 1. As a matrix (for each given h), it is (*Hermitian*) semipositive definite.

Suppose that  $\overline{Z}(x_1), \ldots, \overline{Z}(x_n)$  are observations of  $\overline{Z}(x)$ . The sample variogram  $\overline{\gamma}^*(h)$  for a given lag h is

$$\overline{\gamma}^*(h) = \frac{1}{2N(h)} \sum_{x(h)} (\overline{Z}(x_i) - \overline{Z}(x_j)) (\overline{Z}(x_i) - \overline{Z}(x_j))^T$$

For a selection of lags  $h_1, h_2, \ldots, h_k$  such that  $N(h_i) > 0$ , we obtain a k-tuple of (semi) positive definite matrices  $\overline{\gamma}^*(h_1), \ldots, \overline{\gamma}^*(h_k)$ . We want to determine an orthonormal matrix B which simultaneously diagonalizes the  $\overline{\gamma}^*(h_1), \ldots, \overline{\gamma}^*(h_k)$ , or nearly diagonalizes them, in the sense that the sum of squares of off-diagonal elements is small compared to the sum of squares of diagonal elements. If such a B exists, we linearly transform  $\overline{Z}(x)$  by  $\overline{Y}(x) = B\overline{Z}(x)$ . Then, the resulting vector function  $\overline{Y}(x)$  has less spatial correlation among its components than  $\overline{Z}(x)$  does. Components with little contribution to the variogram structure may be dropped. The analysis now can focus on  $\overline{Y}(x)$ . The sample variogram matrices  $\overline{\gamma}^*_{\overline{Y}}(h_1), \ldots, \overline{\gamma}^*_{\overline{Y}}(h_k)$ , will be in nearly diagonal form. Therefore, a simplified analysis and computation can be carried out.

Let  $\mathfrak{C}_1$  be a set of matrix-valued functions in which each function is conditionally negative definite of order 1. For given sample variogram (matrix form) of  $\overline{Z}(x)$ ,  $\overline{\gamma}_{\overline{Z}}^*(h_1)$ , ...,  $\overline{\gamma}_{\overline{Z}}^*(h_k)$  and for an admissible candidate  $\overline{\gamma}(h) \in \mathfrak{C}_1$ , the goodness-of-fit can be measured by

$$\Phi(\overline{\gamma}) = \sum_{i=1}^{k} \omega_i Tr(\overline{\gamma}(h_i) - \overline{\gamma} \frac{*}{Z}(h_i))^2$$

where  $\omega_i \ge 0$  (i = 1, ..., k) with  $\Sigma \omega_i = 1$ . Because a matrix-valued variogram is symmetric, we use the square of  $\overline{\gamma}(h_i) - \overline{\gamma}^*(h_i)$  instead of  $(\overline{\gamma}(h_i) - \overline{\gamma}^*(h_i))$  $(\overline{\gamma}(h_i) - \overline{\gamma}^*(h_i))^T$ . In order to determine a best candidate, we minimize  $\Phi(\overline{\gamma})$ . The minimization may occur in two senses: one is to minimize  $\Phi(\overline{\gamma})$  for the parameter space of  $\overline{\gamma}$ , if  $\overline{\gamma}$  is specified (for instance, altering nugget, sill, and range if  $\overline{\gamma}$  is a spherical model); the other is to minimize  $\Phi(\overline{\gamma})$  for the set of valid models, C<sub>1</sub>. The latter requires some topological properties of the space  $C_1$ , such as compactness, which are not clear. From a practical point of view, we may assume that the interesting candidates are in some compact set of functions in  $\mathcal{C}_1$ . We also assume that modeling scalar variograms is workable and efficient. Our idea is to rotate  $\overline{\gamma}_{\overline{z}}^{*}(h_{i})$  (i = 1, ..., k) by B so that the main diagonal elements carry as much information as possible; in other words, the off-diagonal elements carry as little information as possible (are minimized); then the diagonal elements are modeled as scalar variograms. The off-diagonal elements can be dropped (if they are nearly zero), or at most be modeled by linear functions. We will show how this idea works and how much loss occurs if we use the resulting model for cokriging.

# THE PRINCIPLE

Suppose that there exists an orthonormal  $m \times m$  matrix B such that

$$B\overline{\gamma} \stackrel{*}{\underline{Z}}(h_1)B', \ldots, B\overline{\gamma} \stackrel{*}{\underline{Z}}(h_k)B'$$

are nearly diagonal. Denote  $B\overline{\gamma}_{\overline{Z}}^*(h_i)B' = D(h_i)$  and let  $\overline{Y}(x) = B\overline{Z}(x)$ . Then  $\overline{\gamma}_{\overline{Y}}(h) = B\overline{\gamma}_{\overline{Z}}(h)B'$  and sample variograms of  $\overline{Y}(x)$  at lags  $h_1, \ldots, h_k$  are  $\overline{\gamma}_{\overline{Y}}^*(h_i) = B\overline{\gamma}_{\overline{Z}}^*(h_i)B' = D(h_i)$ ,  $(i = 1, \ldots, k)$ . Because *B* is orthonormal,

$$\Phi(\overline{\gamma}) = \sum_{i=1}^{k} \omega_{i} Tr(\overline{\gamma}_{\overline{Z}}(h_{i}) - \gamma_{\overline{Z}}^{*}(h_{i}))^{2} = \sum_{i=1}^{k} \omega_{i} Tr(B\overline{\gamma}_{\overline{Z}}(h_{i})B' - B\overline{\gamma}_{\overline{Z}}^{*}(h_{i})B')^{2}$$
$$= \sum_{i=1}^{k} \omega_{i} Tr(\overline{\gamma}_{\overline{Y}}(h_{i}) - \overline{\gamma}_{\overline{Y}}^{*}(h_{i}))^{2} = \sum_{i=1}^{k} \omega_{i} Tr(\overline{\gamma}_{\overline{Y}}(h_{i}) - D(h_{i}))^{2}$$
(4)

Let  $\overline{\gamma}_{\overline{Y}}(h_i) = (\gamma_{rs}^{\overline{Y}}(h))$  and  $D(h) = (d_{rs}(h))$ . Then

1.

$$\Phi(\bar{\gamma}) = \sum_{i=1}^{k} \omega_{i} \sum_{r=1}^{m} (\gamma_{rr}^{\bar{\gamma}}(h_{i}) - d_{rr}(h_{i}))^{2} + 2 \sum_{i=1}^{k} \omega_{i} \sum_{1 \le r < s \le m}^{m} (\gamma_{rs}^{\bar{\gamma}}(h_{i}) - d_{rs}(h_{i}))^{2}$$

Let

$$\Phi_{rr} = \sum_{i=1}^{k} \omega_i (\gamma_{rr}^{\overline{Y}}(h_i) - d_{rr}(h_i))^2, \quad r = 1, \dots, m$$
  
$$\Phi_{rs} = \sum_{i=1}^{k} \omega_i (\gamma_{rs}^{\overline{Y}}(h_i) - d_{rs}(h_i))^2, \quad 1 \le r < s \le m$$

In order to minimize  $\Phi(\overline{\gamma})$ , it is sufficient to minimize the  $\Phi_{rs}$ ,  $1 \le r \le s \le m$ . Note that  $\gamma_{rr}^{\overline{\gamma}}(h)$  is the variogram of the *r*th component  $y_r(x)$  of Y(x) and the  $d_{rr}(h_i)$  are the sample variograms. Minimizing  $\Phi_{rr}$   $(r = 1, \ldots, m)$  is the same as modeling the variogram of  $y_r(x)$  based on its sample variograms  $d_{rr}(h_i)$ ,  $(i = 1, \ldots, k)$ . This can be done in the same way as for a scalar random function.

Similarly, minimizing  $\Phi_{rs}$   $(1 \le r < s \le m)$  is the same as modeling the cross-variogram of  $y_r(x)$  and  $y_s(x)$ . We may require some smoothness property for  $\gamma_{rs}^{\tilde{Y}}(h)$ . Hence the  $\gamma_{rs}^{\tilde{Y}}(h)$  could be any smooth functions such that  $\Phi_{rs}$  is minimized and  $\gamma_{rs}(h)$  along with  $\gamma_{rr}(h)$  (r = 1, ..., m) (obtained by modeling  $y_r(x)$ ) constitute a conditionally negative definite matrix-valued function of order 1.

Before trying to model the cross-variograms  $\gamma_{rs}^{\overline{\gamma}}(h)$   $(1 \le r < s \le m)$ , look at the data  $\{d_{rs}(h_i)\}$  (that is, the sample variogram values) more closely.

Because the  $D(h_i)$  (i = 1, ..., k) are simultaneously nearly diagonal,

$$\sum_{i=1}^{k} \sum_{1 \le r < s \le m} (d_{rs}(h_i))^2 = d$$

is small compared to the sum of the squares of main diagonal elements of  $D(h_i)$ .

$$\bar{d}_{rs} = \frac{1}{k} \sum_{i=1}^{k} d_{rs}(h_i), \qquad \hat{d}_{rs} \approx \sum_{i=1}^{k} \omega_i d_{rs}(h_i)$$
(5)

When  $\omega_i = 1/k$ , then  $\overline{d}_{rs} = \hat{d}_{rs}$ . We call  $\overline{d}_{rs}$  the sample mean and  $\hat{d}_{rs}$  the weighted sample mean. Similarly we obtain sample variances and weighted sample variances

$$\bar{\sigma}_{rs}^{2}(d) = \frac{1}{k-1} \sum_{i=1}^{k} (d_{rs}(h_{i}) - \bar{d}_{rs})^{2}, \ \hat{\sigma}_{rs}^{2}(d) = \sum_{i=1}^{k} \omega_{i}(d_{rs}(h_{i}) - \hat{d}_{rs})^{2}$$
$$\bar{\sigma}_{h}^{2} = \frac{1}{k-1} \sum_{i=1}^{k} (h_{i} - \bar{h})^{2}, \ \hat{\sigma}_{h}^{2} = \sum_{i=1}^{k} \omega_{i}(h_{i} - \hat{h})^{2}$$

Note that

$$\bar{\sigma}_{rs}^{2}(d) = \frac{1}{k-1} \sum_{i=1}^{k} \left( d_{rs}(h_{i}) \right)^{2} - \frac{k}{k-1} \left( \bar{d}_{rs} \right)^{2} < \frac{d}{k-1} - \frac{k}{k-1} \left( \bar{d}_{rs} \right)^{2} \quad (6)$$

which implies that  $|\overline{d}_{rs}| < \sqrt{d/k}$ . Similarly

$$\hat{\sigma}_{rs}^{2}(d) = \sum_{i=1}^{k} \omega_{i}(d_{rs}(h_{i}))^{2} - (\hat{d}_{rs})^{2} < \sum_{i=1}^{k} \omega_{i}(d_{rs}(h_{i}))^{2}$$

and

$$|\hat{d}_{rs}^2| \leq \sum_{i=1}^k \omega_i (d_{rs}(h_i))^2$$

### Matrix-Valued Variogram Models. I

These imply that

$$\sum_{r < s} \hat{\sigma}_{rs}^2(d) < \sum_{i=1}^k \omega_i \sum_{r < s} (d_{rs}(h_i))^2 < d$$
<sup>(7)</sup>

and

$$\sum_{r < s} |\hat{d}_{rs}^{2}| \leq \sum_{i=1}^{k} \omega_{i} \sum_{r < s}^{k} (d_{rs}(h_{i}))^{2} < d$$

Therefore, for  $1 \le r < s \le m$ , what we are going to fit is a set of data for which the data values are small and their variations are even smaller. This type of data may have a horizontally linear tendency. To see this, suppose  $\alpha_{rs} + \beta_{rs}h$  is the linear function (regression line) used for fitting  $\{d_{rs}(h_i)\}$   $(1 \le r < s \le m)$ . We build a weighted least-squares framework by minimizing

$$\sum_{i=1}^{k} \omega_i (d_{rs}(h_i) - \alpha_{rs} - \beta_{rs}h_i)^2$$

By a simple calculation, we obtain the estimated intercepts and slopes

$$\tilde{\alpha}_{rs} = \hat{d}_{rs}$$

$$\tilde{\beta}_{rs} = \frac{\sum_{i=1}^{k} \omega_i (d_{rs}(h_i) - \overline{d}_{rs}) (h_i - \overline{h})}{\sum_{i=1}^{k} \omega_i (h_i - \overline{h})^2}$$

By Holder inequality,

$$|\tilde{\beta}_{rs}| \leq \frac{(\sum_{i=1}^{k} \omega_i (d_{rs}(h_i) - \overline{d}_{rs})^2)^{1/2} (\sum_{i=1}^{k} \omega_i (h_i - \overline{h})^2)^{1/2}}{\sum_{i=1}^{k} \omega_i (h_i - \overline{h})^2} \leq \frac{\hat{\sigma}_{rs}(d)}{\hat{\sigma}_h}$$
(8)

By (7), we have

$$\sum_{r < s} \hat{\beta}_{rs}^2 \le \frac{\sum_{r < s} \hat{\sigma}_{rs}^2(d)}{\hat{\sigma}_h^2} < \frac{d}{\hat{\sigma}_h^2}$$

It is not easy to perform statistical testing of the hypothesis  $\beta_{rs} = 0$ , because of dependences in  $\{d_{rs}(h_i)\}$ . It should be noted that (1) *d* is small by our assumption, (2) we can achieve a large variation of  $\{h_i\}$  easily. Therefore, the quantity  $d/\hat{\sigma}_h^2$  would be small enough so that  $|\tilde{\beta}_{rs}|$  is negligible.

The next question is whether a model generated in this way preserves the conditional negative definiteness property. This will be answered by the following Theorem (Xie, 1994, theorem 5.3.1, p. 70):

### Theorem.

Let h = ||x - y|| for  $x, y \in \mathbb{R}^k$ , where the norm is Euclidian 2-norm. Let

$$\gamma(h) = g(h) + \sum_{u=0}^{p-1} A_u h^{2u}$$

where  $A_{\mu}$  are any real symmetric matrices. If g(h) is conditionally negative definite of order p, then  $\gamma(h)$  also is conditionally negative definite of order p.

This Theorem suggests that we can construct a (matrix-valued) variogram model in the way that g(h) takes care of all diagonal parts and  $\sum_{u=0}^{p-1} A_u h^{2u}$  takes care of all off-diagonal parts. Because the sample variograms of Y(x) are nearly diagonal, we simply set  $g(h) = \text{diag}(\gamma_{11}(h), \ldots, \gamma_{mm}(h))$ , and  $A_u = (a_{rs}^u)$  with  $a_{rr}^u = 0$  ( $u = 0, \ldots, p - 1$  and  $r = 1, \ldots, m$ ). Note that  $\overline{\gamma_{\overline{Y}}}(h)$  is conditionally negative definite of order 1. Then  $\sum_{u=0}^{p-1} A_u h^{2u} = A_0 = (a_{rs}^0)$  with  $a_{rr}^0 = 0$  ( $r = 1, \ldots, m$ ). The model finally becomes

$$\overline{\gamma}_{\overline{Y}}(h) = \operatorname{diag}(\gamma_{11}(h), \ldots, \gamma_{mm}(h)) + A_0$$

From preceding discussion, it is natural to set  $a_{rs}^0 = \tilde{\sigma}_{rs} = \hat{d}_{rs}$   $(1 \le r < s \le m)$ .

Let us now return to  $\Phi(\overline{\gamma})$ 

$$\Phi(\widetilde{\gamma}) = \sum_{r=1}^{m} \Phi_{rr} + \sum_{1 \le r < s \le m}^{m} \Phi_{rs}$$

$$= \sum_{i=1}^{k} \omega_{i} \sum_{r=1}^{m} (\gamma_{rr}^{\widetilde{Y}}(h_{i}) - d_{rr}(h_{i}))^{2} + 2 \sum_{i=1}^{k} \omega_{i} \sum_{1 \le r < s \le m}^{m} (\gamma_{rs}^{\widetilde{Y}}(h_{i}) - d_{rs}(h_{i}))^{2}$$

$$= \sum_{i=1}^{k} \omega_{i} \sum_{r=1}^{m} (\gamma_{rr}^{\widetilde{Y}}(h_{i}) - d_{rr}(h_{i}))^{2} + 2 \sum_{i=1}^{k} \omega_{i} \sum_{1 \le r < s \le m}^{m} (d_{rs} - d_{rs}(h_{i}))^{2}$$

$$< \sum_{i=1}^{k} \omega_{i} \sum_{r=1}^{m} (\gamma_{rr}^{\widetilde{Y}}(h_{i}) - d_{rr}(h_{i}))^{2} + 2 \sum_{i=1}^{k} \omega_{i} \sum_{1 \le r < s \le m}^{m} (d_{rs}(h_{i}))^{2}$$

$$< \sum_{i=1}^{k} \omega_{i} \sum_{r=1}^{m} (\gamma_{rr}^{\widetilde{Y}}(h_{i}) - d_{rr}(h_{i}))^{2} + 2d \qquad (9)$$

The first part of the right-hand side of (9) is the error caused by fitting the main diagonal elements. The second part of (9) is an upper bound on the error caused by fitting the off-diagonal elements (which are small by our assumption).

We remark that this scheme can be extended easily to modeling conditionally negative definite matrix-valued functions of order p > 1 in the light of the Theorem, fitting the off-diagonal elements  $\{d_{rs}(h_i)\}$  by even polynomials with degree less than or equal to 2(p - 1). If  $\{d_{rs}(h_i)\}$  has small variation, the coefficients for the higher degree terms can be neglected.

The process can now be summarized in the following steps.

(1) Compute the matrix sample variogram  $\gamma_{\overline{z}}^*(h_1), \ldots, \gamma_{\overline{z}}^*(h_k)$  for suitably selected lags  $h_1, \ldots, h_k$ .

(2) Determine an orthonormal  $m \times m$  matrix B such that

 $B\gamma_{\overline{Z}}^{*}(h_{1})B' = D(h_{1}), \ldots, B\gamma_{\overline{Z}}^{*}(h_{k})B' = D(h_{k})$ 

are nearly diagonal.

#### Matrix-Valued Variogram Models. I

(3) Linear transform Y(x) = BZ(x).

(4) Model each component  $y_r(x)$  separately. This can be done in the same way as for modeling scalar variograms.

(5) Select  $\gamma_{rs}(h) = \hat{d}_{rs}$   $(1 \le r < s \le m)$ . Then, the estimated variogram model of  $\overline{Y}(x)$  is

$$\gamma_{\overline{Y}}(h) = \operatorname{diag}(\gamma_{11}(h), \ldots, \gamma_{mm}(h)) + A_0$$

where  $A_0 = (a_{rs}^0)$  is symmetric,  $a_{rs}^0 = \hat{d}_{rs} \mid s \leq r < s \leq m$  and  $a_{rr}^0 = 0$ .

If the sum of squares of off-diagonal elements  $\sum_{i=1}^{k} \sum_{1 \le r < s \le m} (d_{rs}(h_i))^2$  is close to zero, we simply use the diagonal matrix-valued function

$$\gamma_{\overline{Y}}(h) = \operatorname{diag}(\gamma_{11}(h), \ldots, \gamma_{mm}(h))$$

as an estimate of  $\gamma_{\overline{Y}}(h)$ .

This scheme relies on a procedure for simultaneously diagonalizing several symmetric matrices. We can use a modified FG-algorithm (Flury and Constantine, 1985; Flury and Gautschi, 1986; Clarkson, 1988a; Xie, 1994) or a least-squares algorithm (De Leeuw and Pruzansky, 1978; Clarkson, 1988b; Xie, 1994) to complete the simultaneous diagonalization procedure.

### REFERENCES

- Clarkson, D. B., 1988a, A remark on algorithm AS 211, the F-G diagonalization algorithm: Applied Statistics, v. 37, no. 1, p. 147–151.
- Clarkson, D. B., 1988b, A least squares version of Algorithm AS 211, the F-G diagonalization algorithm: Applied Statistics, v. 37, no. 2, p. 317-321.
- Cressie, N., 1985, Fitting variogram models by weighted least squares: Math. Geology, v. 17, no. 5, p. 563–586.
- Cressie, N., 1991, Statistics for spatial data: Wiley-Interscience, New York, 900 p.
- De Leeuw, J., and Pruzansky, S., 1978, A new computation method to fit the weighted Euclidean distance model: Psychometrika, v. 43, no. 4, p. 479–490.
- Flury, B. N., and Constantine, G., 1985, The FG-diagonalization algorithm: Appl. Statist., v. 35, no. 2, p. 177-183.
- Flury, B. N., and Gautschi, W., 1986, An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form: SIAM Jour. Sci. Stat. Comput., v. 7, no. 1, p. 169–184.
- Goovaerts, P., 1994, On a controversial method for modeling a coregionalization: Math. Geology, v. 26, no. 2, p. 197-204.
- Myers, D. E., 1982, Cokriging: the matrix form: Math. Geology, v. 14, no. 3, p. 249-257.
- Myers, D. E., 1988, Multivariate geostatistics for environmental monitoring: Sciences de la Terre, v. 27, p. 411–428.
- Xie, T., 1994, Positive definite matrix-valued functions and variogram modeling: unpubl. doctoral dissertation, Univ. Arizona, 144 p.