

Fitting Matrix-Valued Variogram Models by Simultaneous Diagonalization (Part II: Application)¹

Tailiang Xie², Donald E. Myers³, and Andrew E. Long⁴

As an application, we demonstrate a proposed variogram modeling scheme using a spatial data set. Because the scheme relies on a procedure for simultaneously diagonalizing several matrices, we briefly describe the FG and least-squares algorithms. The model obtained by our scheme is used to cokriging the data. In addition, the proposed scheme is compared to more traditional methods.

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

SIMULTANEOUS DIAGONALIZATION ALGORITHMS

Given k $p \times p$ symmetric matrices A_1, \dots, A_k , there exist k orthonormal matrices B_1, \dots, B_k such that $B_i^T A_i B_i = \Lambda_i = \text{diag}(\lambda_1^{(i)}, \dots, \lambda_p^{(i)})$, for $i = 1, \dots, k$, where $\lambda_j^{(i)}$ are eigenvalues of A_i . It is well known that A_1, \dots, A_k can be diagonalized simultaneously if and only if A_1, \dots, A_k have a common eigenvector space: in other words, if A_1, \dots, A_k are mutually commutative. When the commutativity condition is not satisfied, we seek an orthonormal matrix B such that the A_1, \dots, A_k are *nearly* diagonalized by B : that is, such that the squares of off-diagonal elements of $B^T A_i B$ are relatively small (in a sense to be defined).

Two quantities may be used to measure simultaneous diagonalizability:

$$\Phi(B|A_i, n_i; i = 1, \dots, k) = \prod_{i=1}^n (\det(\text{diag}(B^T A_i B)))^{n_i} / \prod_{i=1}^n (\det(A_i))^{n_i} \quad (1)$$

$$\Psi(B|A_i, n_i; i = 1, \dots, k) = \sum_{i=1}^k \sum_{j \neq l}^p n_i (b_j^T A_i b_l)^2 \quad (2)$$

¹Received 6 December 1994; accepted 3 March 1995.

²Biometry Section, Arizona Cancer Center, University of Arizona, Tucson, Arizona 85724.

³Department of Mathematics, University of Arizona, Tucson, Arizona 85721.

⁴Department of Mathematics, University of Wisconsin-Oshkosh, Oshkosh, Wisconsin 54901.

where $B = (b_1, \dots, b_p)$ is any $p \times p$ orthonormal matrix and n_i ($i = 1, 2, \dots, k$) are weights. Note that (a) Φ measures the relative deviation from diagonality, whereas Ψ measures the absolute deviation from diagonality, (b) Φ requires A_1, \dots, A_k to be nonsingular, but Ψ does not, and (c) $\Phi \geq 1$ and $\Psi \geq 0$, with exact equality if and only if B simultaneously diagonalizes the A_1, \dots, A_k . The goal is to determine an orthonormal matrix B_0 such that Φ or Ψ is minimized.

Flury and others (Flury and Constantine, 1985; Flury and Gautschi, 1986) developed a simultaneous diagonalization algorithm, termed the FG-algorithm, by using (1) as the criterion for measuring the simultaneous diagonalizability. De Leeuw and Pruzansky (1978) developed another simultaneous diagonalization algorithm, termed the least-squares algorithm, by using (2) as the criterion for simultaneous diagonalizability. The idea of the least-squares algorithm is to apply Jacobi rotations (Press and others, 1992, p. 94), at each stage minimizing the sum of squares of the k off-diagonal elements selected by a rotation pair (Flury, 1988).

Because sample variograms may be singular, (2) will be used in this paper for variogram modeling. We briefly introduce this algorithm.

THE LEAST-SQUARES ALGORITHM

We denote by $O(p)$ the group of all $p \times p$ orthonormal matrices; " \leftarrow " as assignment.

step L_0 . Select an initial approximation $B = (b_1, \dots, b_p) \in O(p)$ to the orthonormal matrix minimizing Ψ , e.g., $B \leftarrow I_p$ where I_p is identity matrix; and set ϵ_L , a convergence tolerance.

$$l \leftarrow 0; \quad A_i \leftarrow B^T A_i B$$

step L_1 . $B^{(l)} \leftarrow B; \quad l \leftarrow l + 1$.

step L_2 . For $j = 1$ to p and $l = j + 1$ to p , do steps L_{21} and L_{22} :

step L_{21} . Define $\bar{Q}(j, l, \theta) = (\bar{q}_{st})$ as the identity except for $\bar{q}_{jj} = \cos \theta = \bar{q}_{ll}, \bar{q}_{jl} = -\bar{q}_{lj} = \sin \theta$. The angle θ is selected such that the sum of off-diagonal elements of $\bar{Q}(j, l, \theta)^T A_1 \bar{Q}(j, l, \theta), \dots, \bar{Q}(j, l, \theta)^T A_k \bar{Q}(j, l, \theta)$ is minimized: that is,

$$\Psi(\bar{Q}(j, l, \theta) | A_i, n_i; i = 1, \dots, k)$$

is minimized.

step L_{22} . $A_i \leftarrow \bar{Q}(j, l, \theta)^T A_i \bar{Q}(j, l, \theta); \quad B \leftarrow B \bar{Q}(j, l, \theta)$.

step L_3 . If $|\Psi(B) - \Psi(B^{(l)})| < \epsilon_L$, then step; else go to L_1 .

DIAGONALIZATION EFFICIENCY

If A_1, \dots, A_k are not all diagonal and B is an orthonormal matrix used to diagonalize simultaneously the A_1, \dots, A_k , we employ

$$\kappa_B = 1 - \frac{\Psi(B|A_i, n_i; i = 1, \dots, k)}{\Psi(I_p|A_i, n_i; i = 1, \dots, k)}$$

to measure the diagonalization efficiency of B . We may say that B simultaneously diagonalizes A_1, \dots, A_k with efficiency κ_B . From a practical point of view we think of $B^T A_1 B, \dots, B^T A_k B$ as being in nearly simultaneously diagonal form if $\kappa_B \geq 90\%$.

Both Flury (Flury, 1988) and Clarkson (Clarkson, 1988) pointed out in their work that, at the time, convergence of the least-squares algorithm had not been demonstrated. Recently, this convergence problem was proven by Xie (1994).

APPLICATION

In this section, we demonstrate the variogram modeling scheme proposed in Part I of the paper (Xie and Myers, this issue) using a real spatial data set from a recent study of nitrate pollution of water wells in an area around Phoenix, Arizona. This data set contains 171 spatial locations (wells) which were sampled for the three variables bicarbonate, calcium, and magnesium, circa 1977. In the data set, easting and northing are the spatial coordinates and bicarbonate, calcium, and magnesium values are given as log-transformed and scaled chemical concentrations at corresponding locations.

We treated bicarbonate, calcium, and magnesium as components of a (spatial) random vector function. Sample variogram matrices were computed at 50 lags. Because of symmetry, each sample lag matrix is determined fully as an array of six components, composed of the sample variograms and cross-variograms of the three variables:

$$(\gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{22}, \gamma_{23}, \gamma_{33})$$

The least-squares algorithm was employed to perform simultaneous diagonalization for the set of 50 sample variogram matrices. An orthonormal matrix B was obtained:

$$B = \begin{pmatrix} 0.4013 & -0.9145 & -0.0502 \\ 0.6194 & 0.3114 & -0.7207 \\ 0.6747 & 0.2581 & 0.6915 \end{pmatrix} \tag{3}$$

Table 1. Simultaneous Diagonalization (Before and After Diagonalization)^a

	ssd	ss0	ssa
Before	124.265	83.873	208.138
After	206.640	1.498	208.138

^assd: sum of squares of diagonal entries, ss0: sum of squares of off-diagonal entries, ssa: sum of squares of all entries.

having a diagonalization efficiency of 98.2%. We omit the sample variogram matrices (before and after diagonalization) because of space limitations (these are available from the authors upon request, or see Xie, 1994).

Table 1 summarizes results from the simultaneous diagonalization. Descriptive statistics for the components of the sample variogram matrices before and after diagonalization are given in the Table 2.

Figure 1 shows scatter plots of all six components (diagonal and off-diagonal) of the sample variogram matrices before and after diagonalization. Before diagonalization the six components are similar in shape and magnitude.

Traditionally, one would model the sample variograms and cross-variograms of the original data (left, Fig. 1), and use these models in a cokriging program. Although variogram modeling is relatively straightforward, cross-variogram modeling is not; and it is for this reason especially that we seek rather to diagonalize the sample variogram matrices (right, Fig. 1). Nearly simultaneous diagonalization results in an increase in the spatial information carried by the variograms (diagonal components), and a consequent reduction in the spatial information carried by the cross-variograms (off-diagonal components).

Table 2. Descriptive Statistics for Off-Diagonal Components (Before and After Diagonalization)^a

	n	mean	std dev	min	max
γ_{12}^h	50	0.2679	0.1127	0.055	0.519
γ_{12}^a	50	0.0049	0.1054	-0.401	0.208
γ_{13}^h	50	0.3347	0.1366	0.080	0.784
γ_{13}^a	50	0.0031	0.0542	-0.146	0.076
γ_{23}^h	50	0.7721	0.1692	0.274	1.234
γ_{23}^a	50	0.0005	0.0344	-0.133	0.050

^a γ^h : before diagonalization, γ^a : after diagonalization.

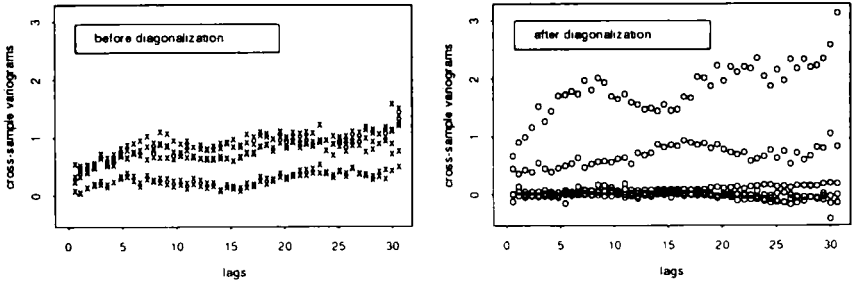


Figure 1. Scatter plots of components of sample variogram matrix, and components of nearly diagonalized variogram matrix.

As shown in Figures 2–7, diagonalization results in a concentration of spatial information in the first two diagonal components, and a reduction of all off-diagonal and the third diagonal components. Based on this result, the original data vectors \bar{Z} were transformed linearly by the orthonormal matrix B , to give the transformed data \bar{Y} :

$$\bar{Y}(x) = B^T \bar{Z}(x) = (y_1(x), \dots, y_k(x))^T \tag{4}$$

We assumed that the off-diagonal elements were negligible, modeling them as zeros, and modeled each of the three diagonal components using least-squares methods and standard models. Assuming no spatial cross-correlation (zero cross-variograms) indicates that cokriging is equivalent to separate kriging, which is less computationally intensive than cokriging, and more stable. The diagonalized data $\{\bar{Y}(x)\}$ thus were kriged, and the kriging estimates were retransformed by B to obtain estimators of the original data.

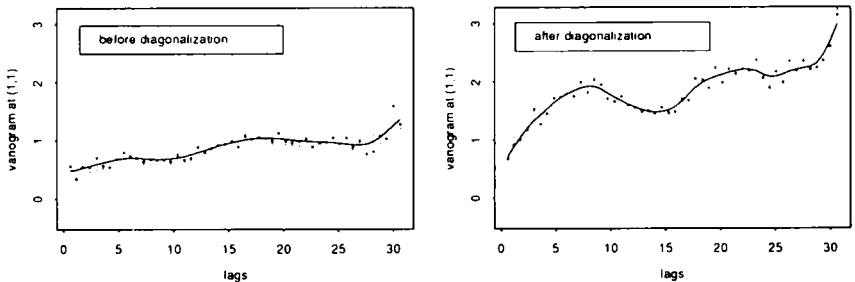


Figure 2. Sample variograms of γ_{11}^h and γ_{11}^l .

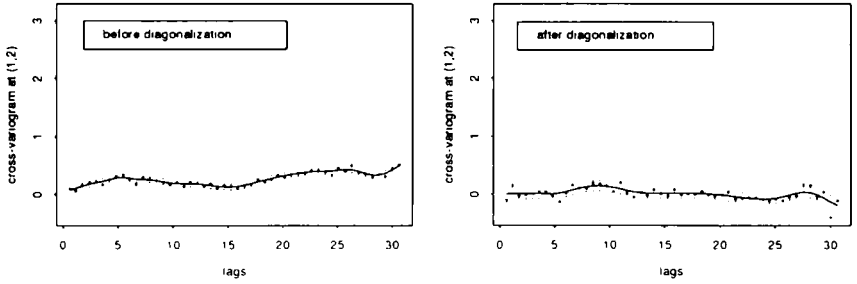


Figure 3. Sample cross-variograms of γ_{12}^b and γ_{12}^u .

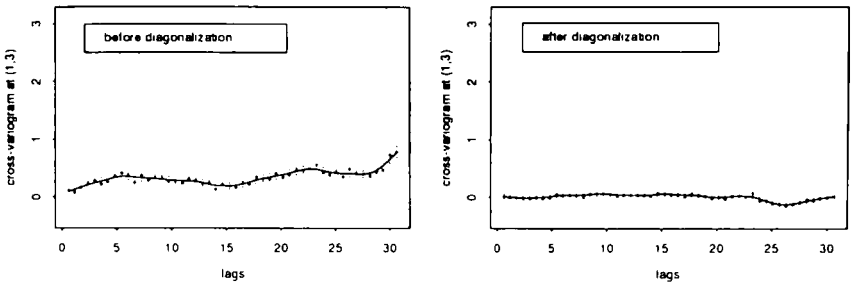


Figure 4. Sample cross-variograms of γ_{13}^b and γ_{13}^u .

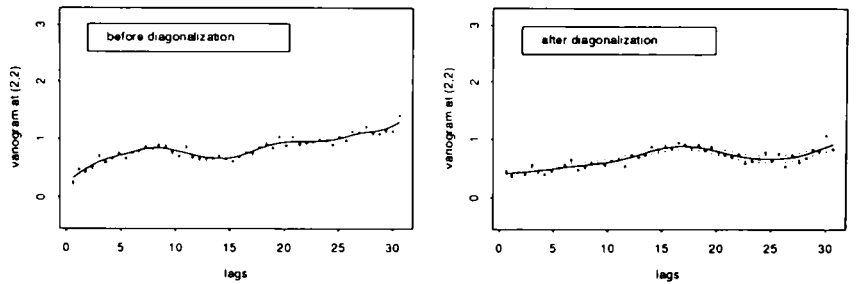


Figure 5. Sample variograms of γ_{22}^b and γ_{22}^u .

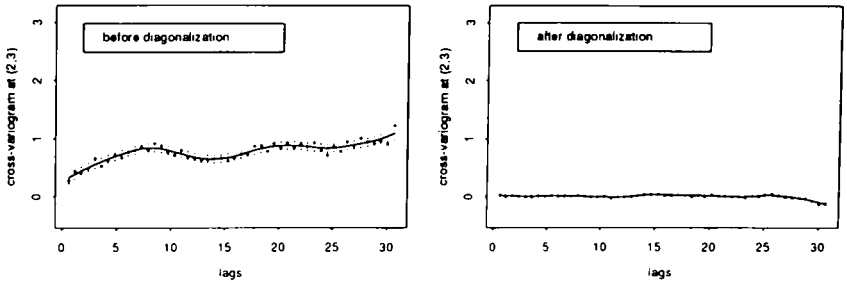


Figure 6. Sample cross-variograms of γ_{23}^h and γ_{23}^d .

DISCUSSION

Table 3 summarizes cross-validation results of kriging the original data directly, cokriging the original data directly, and cokriging indirectly (using the diagonalized data). For all variables, the cross-validation statistics of the kriged variable and the estimates obtained by transformation followed by kriging and retransformation were extremely similar (note that the kriging variance of the diagonalized data could not be retransformed). In this example, we cannot claim improvement using this method over separate kriging of the variables, only that we did about as well using the diagonalized variables in place of the original variables.

One advantage of the diagonalized variables is that the B matrix introduces the potential for some interpretation of the composition of the diagonalized variables: for example, the first transformed variable is "composed of" 0.4013 parts bicarbonate, 0.6194 parts calcium, and 0.6747 parts magnesium.⁵ This

⁵From the equation $Y = B^T Z$ (4): the proportions are from the first column of B (3).

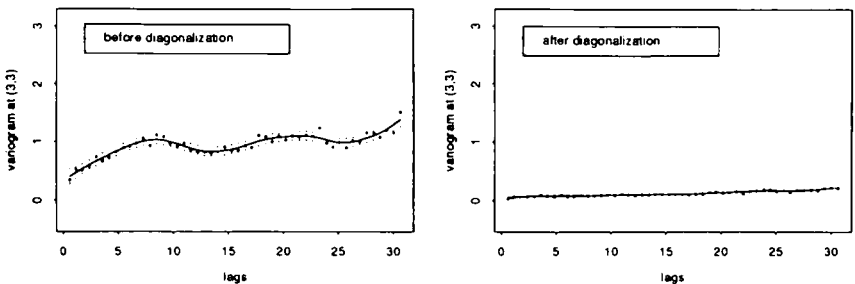


Figure 7. Sample variograms of γ_{33}^h and γ_{33}^d .

Table 3. Cross-Validation Statistics for Interpolation Results (Long, 1994) from Kriging, Cokriging All Three Variables, and All Subcokrigings^a

Statistic	$\overline{ z^* - z }$	$\overline{(z^* - z)^2}$	$\overline{\left(\frac{z^*(h) - z(h)}{\sigma(h)}\right)^2}$	$\rho(z^*, z)$	$\rho\left(z^*, \frac{z^*(h) - z(h)}{\sigma(h)}\right)$	$\overline{\frac{z^*(h) - z(h)}{\sigma(h)^2}}$
Ideal:	0	0	1	1	0	0
Bicarbonate	0.0113	0.5325	0.9238	0.6760	0.0523	0.5827
trans	0.0087	0.5325	N/A	0.6750	N/A	N/A
bi + ca*	0.0058	0.5224	1.0060	0.6824	0.0065	0.5200
bi + mg	0.0060	0.5585	1.1004	0.6572	-0.0730	0.5034
bi + ca + mg	0.0050	0.7323	1.4915	0.5595	-0.2877	0.4948
Magnesium	0.0104	0.4128	0.7636	0.7603	0.0585	0.5544
trans	0.0076	0.4152	N/A	0.7582	N/A	N/A
bi + ca*	0.0033	0.4116	0.8346	0.7599	0.0112	0.4986
ca + mg	0.0585	2.0629	5.0059	0.4292	-0.6783	0.1619
bi + ca + mg	invalid	invalid	invalid	invalid	invalid	invalid
Calcium*	0.0018	0.4466	0.7627	0.7411	0.0640	0.6037
trans	0.0080	0.4610	N/A	0.7309	N/A	N/A
bi + mg	0.0031	0.5037	0.9395	0.7009	-0.0527	0.5251
ca + mg	0.0324	0.6248	2.8418	0.7828	-0.5721	0.1528
bi + ca + mg	0.0018	0.5229	3.0511	0.7828	-0.4790	0.1510

^a“Sums” of variables represent cokrigings. Starred results were judged best for that variable. “trans” results are obtained by kriging diagonalized (transformed) variables, then linearly retransforming to obtain estimates for original variables. (N/A—“Not Applicable”—occurs in table for these estimates because variance-related cross-validation statistics could not be retransformed. “Invalid” occurs in table because of negative cokriging variances.

may or may not indicate anything to a researcher, but at least the information is available for study.

Note that cokriging the original variables actually gave poorer results in certain situations, using cross-validation statistics as the judge: the models may have been invalid, a danger which increases as the cross-variograms do.

Figures 8, 9, and 10 show the contour maps for cokriging the original data and kriging of the transformed data. For bicarbonate (Fig. 8), it appears that the transformed, kriged, and retransformed map is closer to the cokriged map, whereas in the situation of calcium (Fig. 9) the kriged map of calcium itself looks slightly more similar.

We have demonstrated modeling matrix-valued variogram by using the simultaneous diagonalization technique described in Part I of this paper. The advantages of this method are: (1) a simplified analysis and computation in preparation for cokriging, and (2) a guarantee of the negative definiteness of resulting variogram model. Our experience shows that the scheme works well

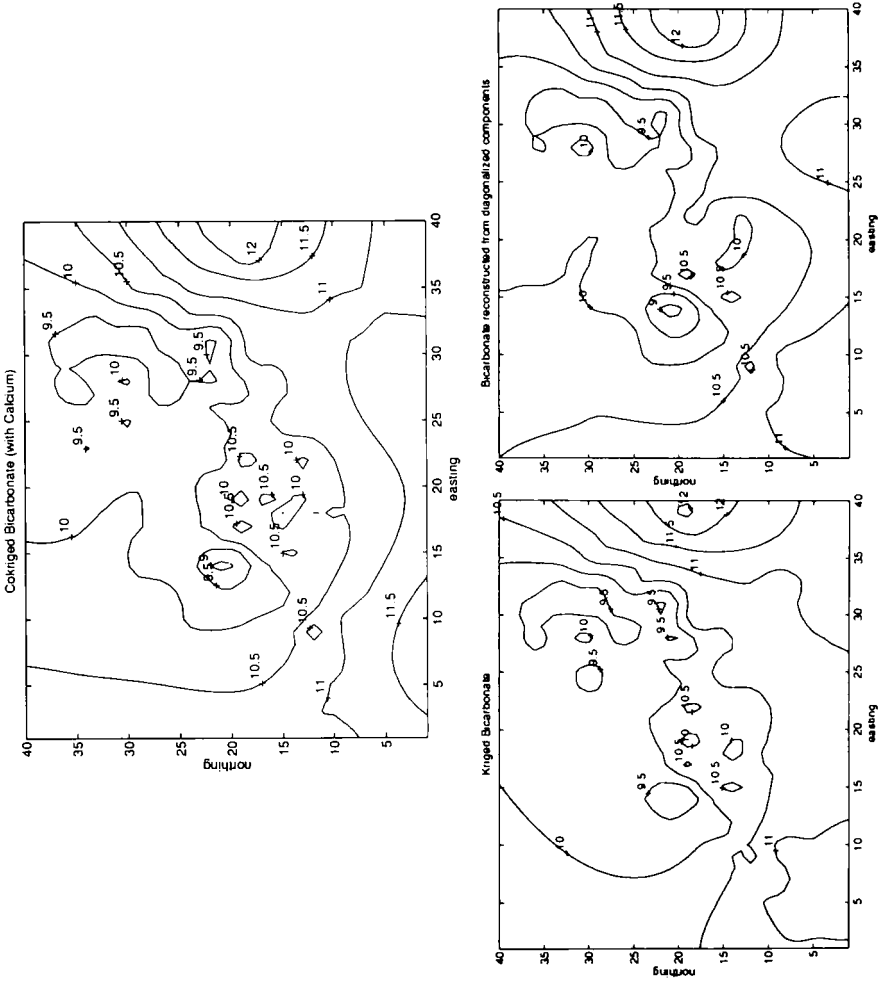


Figure 8. Bicarbonate contours, for cokriging and kriging of raw data, and kriging of transformed data, retrans-

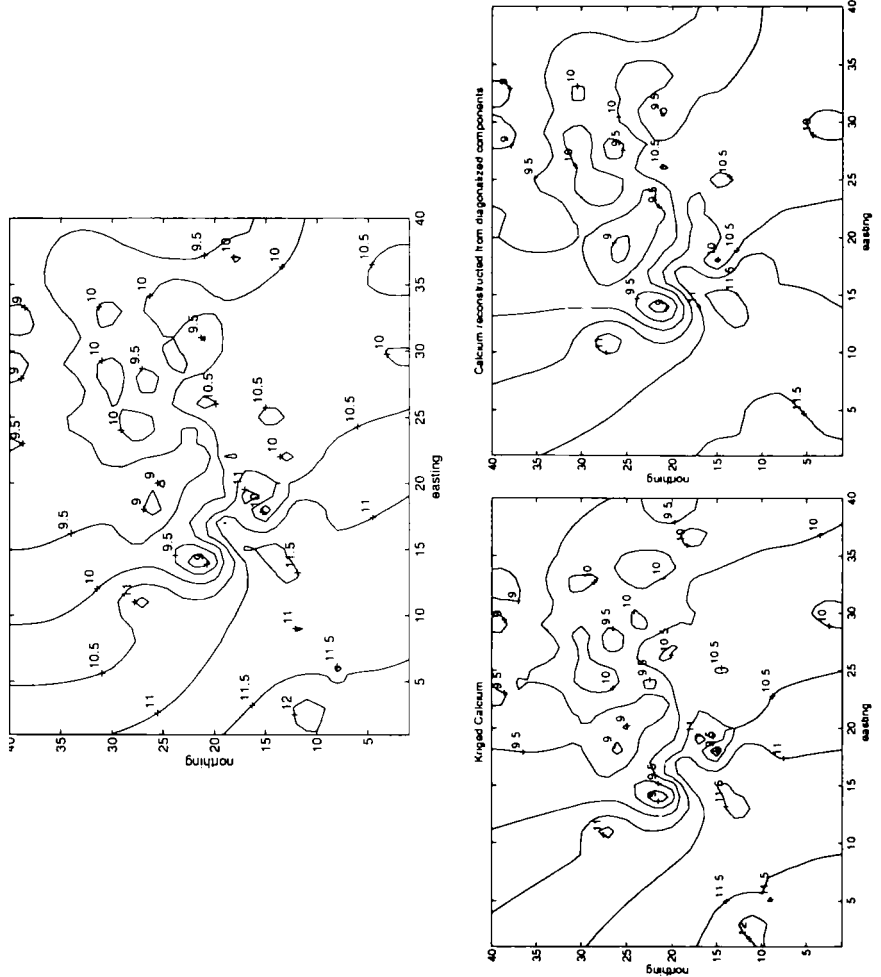


Figure 9. Calcium contours, for cokriging and kriging of raw data, and kriging of transformed data, retransformed to original. Results were contoured to same intervals.

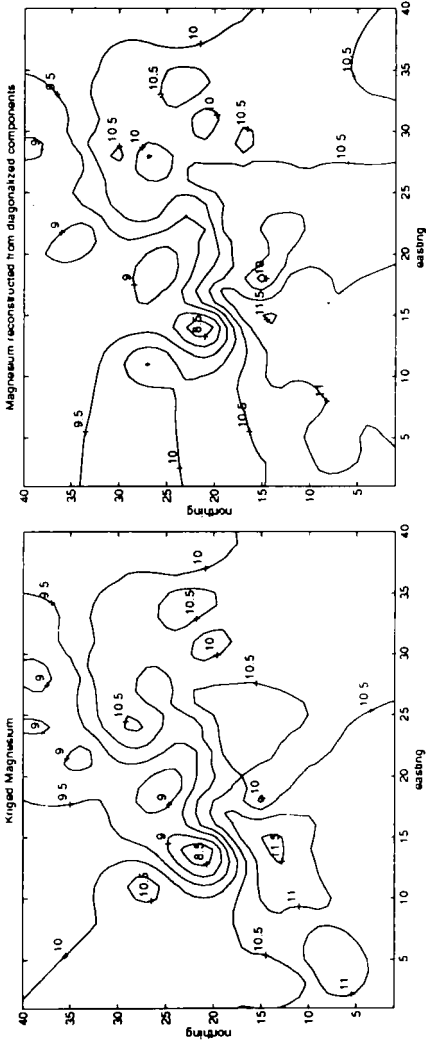


Figure 10. Magnesium contours, for kriging of raw data, and kriging of transformed data, retransformed to original. Results were contoured to same intervals. Kriging was better than cokriging, and raw kriging did better than did kriging transformed data.

in some situations of coregionalization, especially if there are not many components in the vector random function (because efficiency of diagonalization relies on the size of matrices).

A FORTRAN program for the simultaneous diagonalization scheme is available upon request. A public domain UNIX version GeoEas package, developed by the third author, which used to perform the major computations can be obtained anonymously through math.arizona.edu.

REFERENCES

- Clarkson, D. B., 1988, A least squares version of algorithm AS 211, the F-G diagonalization algorithm: *Applied Statistics*, v. 37, no. 2, p. 317-321.
- 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., 1988, *Common principal components and related multivariate models*: John Wiley & Sons, New York, 258 p.
- 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.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 1992, *Numerical recipes in FORTRAN (2nd ed.)*: Cambridge Univ. Press, Cambridge, 963 p.
- Xie, T., 1994, *Positive definite matrix-valued functions and variogram modeling*: unpubl. doctoral dissertation, Univ. Arizona, 144 p.