MULTIDIMENSIONAL MULTIVARIABLE SMOOTHING

Donald E. Myers
Department of Mathematics
University of Arizona
Tucson, AZ 85721

ABSTRACT

Given a vector valued function defined in a region of R_k, $\overline{Z}(x) = [Z_1(x), \ldots, Z_m(x)]$ and values of $\overline{Z}(x)$ at a finite number of points $x_1, \ldots, x_n$, the objective is to estimate the values of certain linear functionals (with values in R_m). It is further assumed that there is some form of interdependence between the components. This problem occurs in mining (ore reserve estimation), hydrology, soil physics as well as in environmental monitoring. In the case of $m = 1$ three essentially equivalent methods are known; thin-plate splines, Radial Basis functions and the regression technique known as "kriging". The thin plate spline is defined by a specific smoothness condition. It is shown that this equivalence can be extended providing a simple extension of splines to the multivariate case and incorporating the "undersampled" case, e.g., wherein values for some components are not known at some sample locations. In each of the three formulations, the interpolator is written in terms of a kernel function which is positive definite in a generalized sense. The kernel in turn defines an operator on an appropriate function space which characterizes the smoothing property of the interpolator.

1. Introduction

In many applications in the Earth Sciences there are multiple variables of interest and which are correlated but for which state equations describing the inter-relationships are unknown. Moreover these variables are usually also associated with spatial coordinates. One objective in the analysis of such data may be interpolation or contouring. For example if one considers ore grades in a copper mine, copper may be of principal interest but other metals such as molybdenum, gold, silver, zinc may also be of critical interest. In the case of a coal deposit, the attributes of interest might include the thickness of the seam or depth of overburden, BTU content, percent ash or sulfur content. Ore grades are by no means uniform throughout a deposit and in the exploration stage, data (assayed sections of drill cores) are relatively sparse. Empirical evidence has long shown that samples are correlated approximately inversely with distance and that there is some degree of correlation between the different attributes not only at the same location but perhaps also at a distance. There are other applications wherein one attribute is of principal interest but it is sampled in different scales, for example rainfall measured by rain gauges vs radar measurements or insolation measured by heliographs vs satellite data. In such cases the objective may include calibration of one form of measurement against another in order to enhance contouring using the cheaper, quicker method of measurement. If a copper deposit is to be exploited by an open pit mine then it is necessary to estimate the average grade of each of the projected mining blocks, i.e., a spatial average. Although additional information will be available at the mining stage, at the exploration stage most blocks will not contain any samples. An overall average would not provide sufficient information since there will be a selection process based on a cutoff grade. The problem may be one of estimation of linear functionals of a vector valued function.

The spatial correlation of a single attribute might be characterized in several ways. One possibility is to assume that the interpolated surface must satisfy a smoothness condition which then implicitly imposes a spatial relationship. If the linear functionals to be estimated do not correspond to point evaluation then it is less obvious how to impose the smoothness condition. For example, is it imposed on the interpolated surface and then the linear functionals computed (which imposes further smoothing), or is it imposed on the interpolated linear functional surface? Is the degree of smoothness determined by the sample information or simply by mathematical convenience? The appropriate way of describing inter-variable correlation by a smoothness condition is even less obvious. The alternative that has been used in geostatistics is to consider the data as a (non-random) sample from one realization of a random function and describe the spatial correlations and inter-variable correlations by (generalized) covariances and cross-covariances. This has two disadvantages; the smoothness is characterized only indirectly, and secondly, the covariances and cross-covariances must be chosen somewhat arbitrarily or they must be estimated from the data. In the latter case distributional assumptions must be made or little will be known about the reliability of the estimators of the parameters since the sampling is non-random as well as generally not on a regular grid. Two classical methods have commonly been used that are similar to this alternative. The first, known as the polygonal method, considers each sample location as the barycenter of a polygon and then assumes that grades are constant on each such polygon (an extreme local smoothness condition). To estimate the grade at an unsampled location or the average grade over a block it is sufficient to find the nearest sample then the grade of this nearest sample is "extended". This method is highly dependent on the sampling pattern and would not reflect differences in the spatial correlation between different metals or

attributes in the same deposit. The second method is Inverse Distance Weighting (IWD) and is well-known in many contexts, it implicitly incorporates spatial correlation but in a crude kind of way and does not directly determine a smoothness condition. While these latter two methods are relatively simple to apply, the results were frequently found to be less than satisfactory. Neither of these latter two methods extends easily to the vector case. In the following section three methods; splines, radial basis functions and kriging, are reviewed for the case of $m = 1$. These three seemingly different methods are in fact essentially equivalent but kriging is more easily extended in a natural way to the case of $m > 1$, in addition the thin plate spline is somewhat more complicated in the case of $k > 1$ in contrast to the other two methods. The extension of kriging to $m > 1$ and the implications for extensions of the other two are discussed in subsequent sections as well as difficulties, special cases and open problems.

2. A brief review: three methods

A. Splines. The interpolation of a function in R_1 (and to a lesser extent in R_2) by the use of a thin plate spline is well known. Given the values of an unknown function $f(x_1), \ldots, f(x_n)$ at a finite number of locations, one utilizes a function $g(x)$ with continuous second derivatives such that the L^2 norm is minimal and $g(x_i) = f(x_i)$; $i = 1, \ldots, n$. These conditions might be re-stated in more general form as follows: Let H_1, H_2 be two Hilbert spaces of functions, B a bounded linear operator from H_1 onto H_2. Let $L_1, \ldots, L_n$ be continuous linear functionals in the dual of H_1. The "spline" is the element, g in H_1 such that B_g has minimal norm in H_2 and the $L_i(g)$; $i = 1, \ldots, n$ have prescribed values. In the simplest case the linear functionals are point evaluations.

B. Kriging. Suppose instead that $Z(x)$ is a (real-valued) random function defined in R_k such that $Z(x) = Y(x) + m(x)$ with

(1a) $E[Z(x)] = m(x) = \sum a_k f_k(x)$,

(1b) $\gamma(h) = 0.5 \operatorname{Var}[Z(x+h) - Z(x)] = 0.5 \operatorname{Var}[Y(x+h) - Y(x)]$

where the f_k's are known functions (often taken as monomials in the position coordinates) and $\gamma(h)$ is assumed dependent only on h. Usually one of the f_k is taken to be identically one. The data is taken to be a (non-random) sample from one realization of $Z(x)$. Under an independence assumption the estimator might be taken to be the mean and hence estimable by the sample mean. Instead a more general linear combination is used with two conditions imposed; namely unbiasedness and minimal mean square error. That is,

$$Z^*(x) = \sum \phi_i(x) Z(x_i) \tag{2}$$

where

$$E[Z^*(x) - Z(x)] = 0 \quad \text{and} \quad \operatorname{Var}[Z^*(x) - Z(x)] \text{ is minimized.} \tag{3}$$

A sufficient condition for the first part of (3) is that

$$\sum \phi_i(x) f_k(x_i) = f_k(x);\ k = 0, \ldots, p. \tag{4}$$

Minimizing the variance of the error of estimation subject to the conditions given in (4) produces the following additional equations

$$\sum \phi_i(x) \gamma(x_i - x_j) + \sum \mu_k(x) f_k(x_j) = \gamma(x - x_j);\ \ j = 1, \ldots, n \tag{5}$$

This stochastic/regression formulation was described independently by Matern (1960), Matheron (1965), and in slightly different form by Goldberger (1962). It is contained at least implicitly in the work of others and has been re-discovered, sometimes in lesser generality, a number of times since. Matern assumed second order stationarity for $Z(x)$, and used (auto) covariances. Matheron placed more emphasis on the use of the variogram which only requires second order stationarity of the first order differences and called the method Universal Kriging (D. Krige is a South African mining engineer whose ideas led Matheron to this formulation). The stationarity is necessary for the estimation of the covariance or variogram (Originally Matheron called 2γ the variogram). In the case where $Z(x)$ is second order stationary with covariance $\sigma(h)$, it is easily seen that $\gamma(h) = \sigma(0) - \sigma(h)$. The covariance is always bounded but variograms need not be. Covariances must be positive definite and this implies the boundedness whereas variograms need only be conditionally negative definite.

It is convenient to write the interpolator given in (2), as well as the system of equations, in matrix form since it de-emphasizes the dependence on both m, k as well as makes the connection with the other two formulations more apparent. (2) may be re-written as

$$Z^*(x) = [\phi_1(x), \ldots, \phi_n(x), \mu_0, \ldots, \mu_p][Z(x_1), \ldots, Z(x_n), 0, \ldots, 0]^T \qquad (2')$$

or

$$Z^*(x) = [\gamma(x - x_1), \ldots, \gamma(x - x_n), f_0(x), \ldots, f_p(x)][b_1, \ldots, b_n, a_0, \ldots, a_p]^T \qquad (2'')$$

and the combined system of equations in (4), (5) can be written as

$$\begin{bmatrix} K & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} K_0 \\ F_0 \end{bmatrix} \begin{bmatrix} K & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} Z \\ 0 \end{bmatrix}. \qquad (4')$$

C. Radial basis functions. Hardy (1971) proposed the used of an interpolator given by (2") except that the terms corresponding to the f's and the a's were omitted. For the kernel function γ he used a bi-harmonic function since he was considering the interpolation of gravity potential. Subsequently Micchelli (1986) showed that conditional positive definiteness of the kernel with respect to the functions $f_0, \ldots, f_p$ is sufficient for a unique solution to the system given in (4') which is obtained from requiring that the interpolator be exact. Micchelli only considered isotropic kernel functions and assumed that the functions $f_0, \ldots, f_p$ were polynomials. As noted in Myers (1988c), this positive definiteness condition is easily generalized to anisotropic kernels and to a more general definition. This generalization is implicit in the work of Matheron (1973).

3. The general problem

A vector valued function $\overline{Z}(x) = [Z_1(x), \ldots, Z_m(x)]$ defined in R_k is observed at a finite number of locations, the objective is to estimate the value of a linear functional of some or all of components of $\overline{Z}$. In practice the linear functionals of interest are point evaluation (at an unsampled location) or average over a volumes. It is also assumed that there is some form of interdependence between the components of $\overline{Z}$ otherwise the problem may be reduced to that discussed above. The data may be incomplete in the sense that at some locations data is available on only some of the components (this

is frequently referred to as the Undersampled case). In general the only knowledge of $\overline{Z}$ is given by the data and the interpolation method must incorporate some form of model. If $\overline{Z}$ is considered as a random function and the multivariate density is known or assumed then conditional expectation would provide the best interpolator. Inferring the density from a non-random sample from one realization of the random function would not be possible. One alternative is to assume at least a weak form of stationarity and utilize an interpolator which only requires second moment properties.

A. Co-kriging

Assume that $\overline{Z}(x)$ has the following form

$$\overline{Z}(x) = \overline{Y}(x) + \overline{M}(x), \quad E\{\overline{Z}(x)\} = \overline{M}(x) = \overline{F}(x)\overline{M} \tag{5a}$$

where

$$\overline{F}(x) = [f_0(x), \ldots, f_p(x)] \tag{5b}$$

$$\overline{\tau}(h) = 0.5E\{\overline{Y}(x+h) - \overline{Y}(x)\}^T\{\overline{Y}(x+h) - \overline{Y}(x)\} \text{ exists and depends only on } h. \tag{5c}$$

The components in $\overline{F}(x)$ are usually taken to be linearly independent monomials in the position coordinates and in particular $f_0(x)$ is taken to be identically one. The matrix of coefficients $\overline{M}$ is unknown however. One of the simplest ways to incorporate the data for all components at all sample locations is to use a linear estimator of the form:

$$\overline{Z}^*(x) = \sum \overline{Z}(x_i)\Gamma_i(x). \tag{6}$$

Theorem 1. *(Myers, 1982). If $\overline{Z}(x)$ satisfies the conditions (5a), (5b) and (5c) above and the interpolator in (6) is to be unbiased and with minimum variance in the following sense:*

$$\sum a_i^2 \operatorname{Var}\{Z_1^*(x) - Z_i(x)\} \text{ is minimal for any choice of the } a_i\text{'s} \tag{7}$$

then the weight matrices in (6) are obtained as the solution to the system of matrix equations

$$\sum \overline{\tau}(x_i - x_j)\Gamma_j(x) + F_k(x_i)\mu_k(x) = \overline{\gamma}(x_i - x); \quad i = 1, \ldots, n \tag{8a}$$

$$\sum F_k(x_i)\Gamma_i(x) = F_k(x); \quad k = 0, \ldots, p \tag{8b}$$

$$F_k(x_i) = f_k(x_i)I. \tag{8c}$$

It was subsequently shown (Myers, 1984) that a simple modification of the system given in (8a), (8b), (8c) would incorporate the undersampled case and this was incorporated in a program (Carr, Myers and Glass, 1985). It was also shown that the solution is independent of the weights in (7) and in particular the case of one weight equal to one and all others equal to zero corresponds to the case of estimation of only one component. Early applications considered only this case and it was not noted that when the system given in (8a, b, c) is written in matrix form analogous to that of (4') then the coefficient matrix is the same whether estimating one component or all (Myers, 1988b,

c). This reformulation (sometimes called the Dual form) is given in Myers (1988a) and shows that the Radial Basis formulation is easily generalized to the vector case and to the undersampled case as well as allowing for incorporation of anisotropies. If linear functionals other than point valuation are to be estimated then it is only necessary to modify the right hand side of (8a).

B. Inference

Application of co-kriging assumes knowledge of the functions in $\overline{F}(x)$ (or at least the maximum order if monomials are used) as well as the variogram matrix function $\overline{\gamma}(h)$. In the case of $m = 1$, necessary and sufficient conditions are given as well as an integral representation theorem which generalizes the Bochner theorem. A form of positive definiteness is required in order that the variance in (3) be positive, this same form of positive definiteness is the sufficient condition for the coefficient matrix in (4') to be invertible as noted in Micchelli (1986) and Myers (1988c). In practice scalar variograms are estimated by sample variograms using the data and then fitting to a positive linear combination of valid models. This procedure is not as appropriate in the vector case. In particular the positive definiteness condition is more complicated. The following section presents three equivalent forms of positive definiteness arising out of three different formulations of the estimation problem. In turn this provides a practical method for modeling matrix variogram functions.

4. Positive definiteness

A. General co-kriging

In the general formulation of co-kriging, with all the weights in the sum of the variances equal to one, the "variance" is given by a matrix quadratic form

$$-\text{trace} \sum\sum \Gamma_i^T \overline{\gamma}(x_i - x_j)\Gamma_j \,. \tag{9}$$

This quadratic form should be positive for all points $x_1, \ldots, x_n$ and all weight matrices $\Gamma_i, \ldots, \Gamma_n$ which satisfy the following conditions

$$\sum F_k(x_i)\Gamma_i = 0; \quad k = 0, \ldots, p \,. \tag{10}$$

It is easy to see that this definition of positive definiteness is the same as positive definiteness of the matrix whose sub-blocks are the values of the matrix valued function $\overline{\gamma}(x_i - x_j)$. In non-matrix form this becomes

$$-\sum\sum\sum\sum C_s^i \gamma_{st}(x_i - x_j) C_t^j > 0 \,. \tag{11}$$

where $\gamma_{st}(h)$ is the general entry in $\overline{\gamma}(h)$.

B. Linear combinations

In the case of ore grades, as well as in some other applications, it makes sense to form linear combinations called equivalent grades obtained by weighting by relative prices. It is then possible to avoid the necessity of joint estimation of the components of the vector function by converting all data to equivalent grades and then the interpolation

is applied to the linear combination. As noted in Myers (1983) this is suboptimal but it is instructive to consider the translation of the positive definiteness condition. Let $W(x) = \overline{Z}(x)B$ be a linear combination then the variogram for W in terms of the variogram for $\overline{Z}$ is given by

$$\overline{\gamma}_W(h) = B^T \overline{\gamma}_Z(h) B. \tag{12}$$

When the positive definiteness condition is applied to $\gamma_W(h)$ the result is equivalent to (11). This approach is completely analogous to that used by Grenander (1957) for covariances.

C. Single component estimation

In it's earliest form, co-kriging was used only for estimation of one component but using data from the other components. This can be written in the form

$$Z_s^*(x) = \sum \sum Z_t(x_i) C_{ts}^i \tag{13}$$

and the error variance is of the form

$$-\sum \sum \sum \sum C_s^i \gamma_{st}(x_i - x_j) C_t^j \tag{14}$$

i.e., exactly the same as (11).

D. A generalization

The extension of Micchelli's theorem given in Myers (1988c) can be extended to the vector case. Let $f_0, \ldots, f_p$ be linearly independent functions and as above let F_i be $f_i I$ (I an $m \times m$ identity matrix) then it is easily seen that

$$\sum F_i(x) \Gamma_i = 0 \tag{15}$$

implies that all the Γ_i are zero matrices, i.e., the linear independence of the scalar matrix functions $f_0, \ldots, f_p$ is equivalent to the linear independence of the scalar matrix functions $F_0, \ldots, F_p$.

Definition. Let $F_0, \ldots, F_p$ be linearly independent scalar matrix functions defined on R_k and $\overline{g}(x, y)$ be a function from $R_k \times R_k$ into the ring of $m \times m$ real symmetric matrices. Without loss of generality assume that the identity matrix is one of the linearly independent scalar matrix functions. Then $\overline{g}$ is said to be positive definite with respect to the $F_0, \ldots, F_p$ if for all sets of points $x_1, \ldots, x_n$ in R_k

$$\text{Tr} \sum \sum \Gamma_i^T \overline{g}(x_1, x_j) \Gamma_j > 0 \tag{16}$$

for all $\Gamma_1, \ldots, \Gamma_n$ such that

$$\sum F_i(x_j) \Gamma_j = 0 \quad \text{for} \quad i = 0, \ldots, p. \tag{17}$$

Consider now a vector estimator/interpolator of the form

$$\overline{Z}^*(x) = \sum B_i \overline{g}(x_i, x) + \sum A_j F_j(x) \tag{18}$$

where $B_1, \ldots, B_n$ and $A_0, \ldots, A_p$ are $m \times m$ matrices. Given data $\overline{Z}(x_1), \ldots, \overline{Z}(x_n)$ and requiring that the interpolator be exact, i.e., $\overline{Z}^*(x_i) = \overline{Z}(x_i)$ for $i = 1, \ldots, n$ a linear system of equations is obtained. Unless the kernel function is positive definite in the strong sense then the coefficient matrix may not be invertible, however by imposing additional conditions on the $B_1, \ldots, B_n$ a unique solution is obtained. The term, $\sum A_j F_j(x)$, will be seen to determine the behavior of the estimator outside of the convex hull of the sample/data locations. The form of the estimator given by (18) is also seen to be a slight generalization of the dual form of the co-kriging estimator as given in Myers (1988a).

Theorem 2. *Let $F_0, \ldots, F_p$, $\overline{g}(x, y)$ be as above; $B_1, \ldots, B_n$ and $A_0, \ldots, A_p$ as in (18) then the latter are obtained as the unique solution to the system*

$$\begin{bmatrix} \overline{g}(x_1,x_1) & \ldots & \overline{g}(x_1,x_n) & F_0(x_1) & \ldots & F_p(x_1) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \overline{g}(x_n,x_1) & \ldots & \overline{g}(x_n,x_n) & F_0(x_n) & \ldots & F_p(x_n) \\ F_0(x_1) & \ldots & F_0(x_n) & & & \\ \ldots & & \ldots & & 0 & \\ F_0(x_1) & \ldots & F_p(x_n) & & & \end{bmatrix} \begin{bmatrix} B_1^T \\ \vdots \\ B_n^T \\ A_0^T \\ \vdots \\ A_P^T \end{bmatrix} = \begin{bmatrix} Z(x_1)^T \\ \vdots \\ Z(x_n)^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} . \quad (19)$$

The proof is completely analogous to the case when $m = 1$ as given in Myers (1988b). If as was assumed above, the identity matrix is one of the linearly independent matrix functions then the sum of the B's is the zero matrix (i.e., this is one of the equations in (19)) hence if $\overline{g}(x, y)$ is a constant matrix for the distance between x, y sufficiently large then $\overline{Z}^*(x)$ is determined only by the second part of (18) when the minimum of the distances $x - x_j$ is large enough. The upper portion of (19) is obtained by imposing the exactness condition and the lower part corresponds to the unbiasedness conditions if formulated in the context of co-kriging. However the estimator can be obtained without the stochastic formulation and provides a natural generalization of a spline. The choice of $\overline{g}$ is determined by a smoothness condition. From the geostatistical/co-kriging perspective $\overline{g}$ is assumed to be "uniquely" determined by the data and the principal problem is one of adequately estimating/modeling the kernel function as well as determining the appropriate order of the polynomial functions whereas from the perspective of splines or radial basis functions the choice of the linearly independent functions and the kernel function is more arbitrary and determined by external conditions imposed on the estimator.

4. Applications to image analysis

Remotely sensed data is usually available on a dense grid but may also be available only with low resolution. One way to enhance the use of such data is to calibrate it against ground-based high resolution data. Co-kriging provides a method for carrying out this calibration. Examples of such data include the measurement of daily insolation by satellite vs measurement by heliograph or rainfall as measured by radar vs that measured by rain gauges. Co-kriging incorporates spatial variability for each attribute as well as inter-attribute correlation and it also incorporates the scale of the measurements.

Another direction is suggested by the work of Switzer and Green (1984). Principal Components Analysis has frequently been used to distinguish noise from signal in images, Switzer and Green applied PCA to the covariance matrix of the first order differences. This is analogous to a diagonalization of the variogram matrix function as suggested in Myers (1988a). Ma and Royer (1988) have shown that the smoothing property of the kriging estimator can be re-formulated as filtering when applied to images. By using co-kriging the same results can be extended to multi-spectral images.

5. Final remarks

In practice not all of the data set is used for each estimation, instead a moving neighborhood is used. The vector of weight matrices is a function of the variogram matrix, the order of the functions in $F(x)$, the moving neighborhood (that is, the search method). To characterize the continuity it is necessary to define a neighborhood. In the univariate case three different neighborhood definitions are given in Myers (1985a, 1986) but none will easily generalize to the vector case.

One advantage of the stochastic formulation for the interpolator is that it lends itself to simulating new realizations but preserving first and second moment properties as well as the marginal distribution. The algorithm/program given in Carr and Myers (1985) partially extends this to the vector case, the algorithm given in Myers (1989) provides a full extension.

Notice

Although the research described in this article has been funded wholly or in part by the U.S. Environmental Protection Agency through a Cooperative Research Agreement with the University of Arizona, it has not been subjected to Agency review: therefore, neither does it reflect the views of the Agency nor should any official endorsement be inferred.

Bibliography

Carr, J., & Myers, D. E. (1984). *Application of the Theory of Regionalized Variables to the Spatial Analysis of LANDSAT data.* in the Proceedings of the Ninth Pecora Symposium on Remote Sensing, 2–4 Oct., IEEE Computer Society Press, 55–61.

Carr, J., Myers, D. E., & Glass, C. (1985). Co-kriging: A Computer Program. *Computers and Geosciences* **11**, 111-127.

Carr, J., & Myers, D. E. (1985). COSIM: A Fortran IV program for co-conditional simulation. *Computers and Geosciences* **11**, 675–705.

Carr, J., Glass, C., Yang, H., & Myers, D. E., (1987). Application of Spatial Statistics to Analyzing Multiple Remote Sensing Data sets. Johnson, A. I. & Peterson, C. B. (1987, eds.), *Geotechnical Applications of Remote Sensing and Remote Data Transmission, ASTM STP 967*, 138–150.

Franke, R. (1982). Scattered Data Interpolation: Tests of Some Methods. *Math. Comp.* **38**, 181–200.

Goldberger, A. S. (1962). Best Linear Unbiased Prediction in the Generalized Linear Regression Model. *Journal of American Statistical Association*, 369–375.

Grenander, U. (1957). *Statistical Analysis of Stationary Time Series.* J. Wiley and Sons, New York.

Hardy, R. H. (1971). Multiquadratic Equations of Topography and other Irregular Surfaces. *J. Geophysical Res.* **76**, 1905–1915.

Kimmeldorf, G. S., & Wahba, G. (1970). A Correspondence between Bayesian estimation on Stochastic Processes and Smoothing by Splines. *Annals. Math. Stat.* **41**, 495–502.

Ma, Y. Z., & Royer, J. J. (1988). Local Geostatistical Filtering: Application to Remote Sensing. *Sciences de la Terre* **27**, 17–36.

Matern, B. (1986). *Spatial Variation.* 2nd Edition, Lecture Notes in Statistics #36, Springer Verlag.

Matheron, G. (1973). The Theory of Intrinsic Random Functions and their Applications. *Advances in Applied Probability* **5**, 439–468.

Matheron, G. (1980). *Splines et Krigeage: Leur Equivalence Formelle.* Internal Report N–667, Centre de Geostatistique, Fontainebleau, 26p.

Micchelli, C. (1986). Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions. *Constructive Approximation* **2**, 11–22.

Myers, D. E. (1982). Matrix Formulation of Cokriging. *Math. Geology* **14**, 249–257.

Myers, D. E. (1983). Estimation of Linear Combinations and Cokriging. *Math. Geology* **15**, 633–637.

Myers, D. E. (1984). Cokriging: New Developments. Verly, G. (1984, ed.), *Geostatistics for Natural Resource Characterization*, D. Reidel Pub. Co., Dordrecht, 295–305.

Myers, D. E. (1985a). On Some Aspects of Robustness. *Sciences de la Terre* **24**, 63–79.

Myers, D. E. (1985b). Co-kriging: Methods and Alternatives. Glaeser, P. (1985, ed.), *The Role of Data in Scientific Progress*, Elsevier Scientific Pub, 425–428.

Meyers, D. E. (1986). *The Robustness and continuity of Kriging.* Joint Statistics Meeting, Chicago, Aug. 20–21.

Myers, D. E. (1988a). Some Aspects of Multivariate Analysis. Chung, C. F. (1988, ed.), *Quantitative Analysis of Mineral and Energy Resources*, D. Riedel Publishing Co., Dordrecht, 669–687.

Myers, D. E. (1988b). Multivariate Geostatistics for Environmental Monitoring. *Sciences de la Terre* **27**, 411–428.

Myers, D. E. (1988c). Interpolation with Positive Definite Functions. *Sciences de la Terre* **28**, 251–265.

Myers, D. E. (1989). Vector Conditional Simulation. Armstrong, M. (1989, ed.), *Geostatistics*, Volume I, 283–293.

Myers, D. E., & Carr, J. (1985). Cokriging and Principal Component Analysis: Bentonite Data revisited. *Sciences de la Terre* **21**, 65–79.

Salkkauskas, K. (1982). Some Relationships between Surface Splines and Kriging. Schempp, W. & Zeller, K. (1982, eds.), *Multivariate Approximation Theory II*, Birhauser, Basel, 311–325.

Switzer, P., & Green, A. A. (1984). *Min/Max Autocorrelation Factors for Multivariate Spatial Imagery.* Technical Report, Dept. Statistics, Stanford Univ., 14p.