Stochastic Models and Radial Basis Function Interpolators

Donald E. Myers [(1)]

1 Introduction

There are many examples of stochastic formulations or solutions for problems that otherwise are treated as deterministic. Bras and Rodríguez-Iturbe,(1985) and de Marsily (1986) are good examples of the treatment of stochastic equations in lieu of deterministic ones in order to adequately incorporate heterogeneity in the subsurface and hence in the hydrogeologic parameters appearing in the flow and transport equations. Stochastic methods are widely used in petroleum, Yarus and Chambers (1994). There are examples more specific to Radial Basis Function interpolators, see Matheron (1973, 1980-81), Wabha and Kimeldorf (1970), Liu et al (2002), Weller et al (2002).
Likewise there are multiple reasons for considering a stochastic formulation. These include (1) additional or different insights into the problem arising from the stochastic model, (2) new results not easily obtainable from the deterministic model, (3) simpler derivations of some results when the stochastic formulation is used. But another important reason might be the difference between "error" and "uncertainty". Madych and Nelson (1988) give a bound on the point approximation error but it is given in terms of the norm of the unknown function. In general the function being interpolated will be unknown and hence the "errors" will be unknown, hence there is 'uncertainty" which is often best described in statistical terms.

2 Positive Definiteness

Recall that the strict (conditional) positive definiteness is essential in the derivation and application of the RBF interpolator for two reasons; (1) to determine the norm (or semi-norm) for the interpolation space and (2) to ensure that the system of equations determining the coefficients in the RBF interpolator has a unique solution. In the stochastic context, a positive definite function is a covariance function for a second order stationary random function (a conditionally positive definite function is a generalized covariance function for an intrinsic stationary random function). Matheron (1973) gives a representation theorem for conditionally positive definite functions (generalizing Bochner's Theorem) based on results in Gelfand and Vilenkin (1964). In a stochastic model rather than minimizing the norm, one minimizes the estimation variance (which is computed in terms of the generalized covariance function).

[(1)] Department of Mathematics, University of Arizona, Tucson, AZ 85721-0089 USA.

3 The Equivalence

In the notation of Madych and Nelson (1988) the RBF interpolator is of the form

$$s(x) = \sum_{[j=1,\dots,n]} c_j\, h(x-x_j) + \sum_{|\alpha| < m} k_\alpha\, x^\alpha \tag{1}$$

where

$$\sum_{[j=1,\dots,n]} c_j\, h(x-x_j) + \sum_{|\alpha| \le m} k_\alpha\, x^\alpha = v_i\ ; \quad i = 1, \dots, n \tag{2a}$$

$$\sum_{[i=1,\dots,n]} c_i\, x_i^{\alpha} = 0\ ; \quad |\alpha| < m \tag{2b}$$

By simple linear algebra this is transformed into

$$s(x) = \sum_{[j=1,\dots,n]} C_j\,(x)\, v_j \tag{1'}$$

where

$$\sum_{[j=1,\dots,n]} C_j\,(x)\, h(x_i - x_j) + \sum_{|\alpha| \le m} K_\alpha\, x_i^{\alpha} = h(x - x_i)\ ; \quad i = 1, \dots, n \tag{2a'}$$

and

$$\sum_{[j=1,\dots,n]} C_j\,(x)\, x_j^{\alpha} = x^{\alpha}\ ; \quad |\alpha| < m \tag{2b'}$$

It is common to write $C_j\,(x)$ as simply C_j. Note that the coefficient matrix is exactly the same in both systems. The difference is that in the RBF formulation one solves the system once and obtains a functional form that must be evaluated at each point, in the stochastic form the coefficient matrix need be inverted only once but the right hand side changes with each point. In the stochastic formulation the data, i.e., $vi;\ i= 1,\dots,n$, is a non-random sample from one realization of a random function V(x) with generalized covariance h(u). The system (2a', 2b') is obtained by minimizing the variance of the error of estimation under the constraint of unbiasedness assuming that the estimator is of the form given in (1'). This form is motivated by the multivariate gaussian case where the conditional expectation is linear in the data. It is well known that in general the conditional expectation is the minimum variance unbiased estimator. Thus (1') is a linear approximation to the conditional expectation. The minimized estimation variance is given by

$$\sigma^2(x) = \sum_{[j=1,\dots,n]} C_j\, h(x - x_j) + \sum_{|\alpha| \le m} K_\alpha\, x^\alpha \tag{3}$$

Note that the coefficients are not the same as in (1). This variance is not data dependent hence one must be careful about interpreting it in the usual way, e.g., to compute confidence intervals.

In the alternative form, (1'), the estimated (interpolated) value is a weighted average of the data values. It is easy to see that the degree to which this weighted average differs from a simple unweighted average depends on the spatial "correlation". That is, the extent to which the values at close locations are more similar than the values at distant locations. The covariance function explicitly quantifies this similarity (or lack of it). Eq(2') incorporates the similarity of the values at the data locations as well as the similarity of the value at the location to be estimated with respect to the separate data locations.

4 The Data

In the context of RBF the data are "numbers", i.e., scalars and each is the value of the unknown *function* at a single point. In the stochastic formulation the data may have other characteristics. In particular the data may in fact be the values of linear functionals applied to the unknown function. Two in particular are of interest, the derivative and a spatial integral. Likewise rather than "interpolating" values of the unknown function one may wish to estimate values of linear functionals applied to the unknown function. The derivation of the RBF interpolator does not provide an easy way to deal with either of these variations, in contrast it is rather easy in the stochastic formulation. There are other possibilities for the data as well, e.g., the data might be given in terms of inequalities or as probability distributions. The data might also be vector valued, e.g., barometric pressure and wind speed. Again the stochastic formulation provides a logical extension. This particular example is considered in Chauvet et al (1976)

5 Interpolation vs Smoothing

The RBF interpolator is "exact" (also called "perfect"), i.e. $s(x_i) = v_i\,;\ i = 1,\ldots,n$. This seems to imply that the there is no error in the data. If the data represent measurements or the results of some form of analysis (e.g., laboratory analysis) then this may be unreasonable. The thin plate spline is a particular RBF interpolator but the smoothing spline is not. However incorporating the variance of the error is easily accomplished in the stochastic formulation, this requires only a minor change in the equations in (2a'). The smoothing spline is then a special case, see Cressie (1989, 1990)

6 Non-Uniqueness

If the RBF $h(u)$ in (2) is strictly conditionally positive then the system (2), (3) will always have a unique solution irrespective of the choice of the $h(u)$ but in general the interpolating function given in (1) will be different. That is, at non-data points the interpolated values need not be the same. The error bound given in Madych and Nelson (1988) depends on the choice of $h(u)$. However the data is not really used in the derivation. There are many empirical studies that compare choices of $h(u)$ for particular data sets generated from known functions but do not provide any particular guidance as to how to choose the basis function *a priori* and when the function is unknown. There are various ways to "fit" the generalized covariance to the data and hence attempt to minimize the nonuniqueness. These arise rather naturally in the context of the stochastic formulation.

Cross-validation. At a non-data point, the error is given by $s(x)-V(x)$ or in standardized form, $[s(x)-V(x)]/\sigma(x)$. If the generalized covariance is a "good fit" to the data then all of these errors should be small. More specifically

$$(1/n)\ \sum_{[j=1,\ldots,n]} |s(x_j)-v_j|$$

should be close to zero because the expected value of each term in the sum is zero. If the entire data set is used to generate the estimated values this statistic would not be useful. Therefore it is necessary to "jackknife" the data. That is, the values at data locations are systematically estimated one at a time using only *other* data locations. In considering how large a discrepancy (from zero) is significant it is important to note the scale of the data values. A second statistic is given in terms of the normalized errors

$$(1/n)\ \sum_{[j=1,\ldots,n]} |s(x_j)-v_j|^2 / \sigma^2(x_j)$$

The expected value of this statistic is one. Finally one might consider the distribution of the normalized errors. Using Chebyshev's Inequality (which does not depend on any distributional assumptions)

$$P\{|s(x)-V(x)|/\sigma(x) > k\} \leq 1- 1/k^2$$

Thus ~ 90% or more of the normalized errors should be ≤ 3. Under the stronger assumption of a multigaussian distribution, more than 95% of the normalized errors should be ≤ 2.5. Particularly when the data is generated by some phenomenon rather than by a known function it can be useful to plot the errors vs location, to construct a histogram of the normalized errors and to plot estimate vs true value. For a discussion of this in the context of RBF's, see Myers (1992b). While there are some similarities with cross-validation as it is sometimes used with the smoothing spline there are also significant differences. In this case the objective is to evaluate the fit of the covariance/generalized covariance. With the smoothing spline one is only determining the degree of difference between an exact interpolator and a smoothed one.

Estimation and Fitting of the Generalized Covariance. There are natural estimators for both the covariance function and the generalized covariance of order zero (usually called the variogram). Then the appropriate model should be "close" to the estimated covariance/generalized covariance. A form of weighted least squares can be used to both estimate parameters and also to quantify the goodness of fit. See the R packages *gstat* and *geoR*.

Maximum Likelihood. Under an assumption of a multigaussian distribution and a covariance function of general type (but with unknown parameters, e.g., Matérn) the problem reduces to one of estimating the parameters in the density function including the mean. This method is discussed in Diggle et al (2003) and implemented in the *geoR* package for R.

A Non-consequential Non-uniqueness. By definition covariance functions and generalized covariance functions satisfy $h(0) = 0$. Although positive definite functions satisfy this, conditionally positive definite functions need not satisfy this condition. However *if* $h(u)$ conditionally positive definite $a + h(u)$ produces the same solution in the systems given by (2a), (2b) and by (2a')., (2b'). For example, $h(u) = -[\delta^2 + |u|^2]^{0.5}$ vs $|\delta| - [\delta^2 + |u|^2]^{0.5}$.

7 Linear Functionals

Obviously "point values" are linear functionals but there are at least two other linear functionals of interest.

Integrals. Let A be a measurable set in $\boldsymbol{R}^k$, then the objective is to estimate

$$(1/|A|) \int_A f(x)\, dx$$

where $f(x)$ is the unknown (assumed integrable) function. Obviously one solution is to simply integrate (analytically or numerically) the interpolating function. The problem is that there is no really good error bound, note that the error bound given by Madych and Nelson (1988) is only local. In the stochastic formulation the problem appears slightly different. The estimator in (1') could be used to generate estimates at each point on a fine grid superimposed on A and then use numerical integration.

While there will be a minimized estimation variance at each grid point these do not directly produce an estimation variance for the integral. How ever it is possible to show that the integral can be estimated directly by using an estimator exactly like eq (1') but eqs (2a') will be modified, namely the term on the right hand side of the equations is replaced by the integral of that function, i.e.

$$(1/|A|) \int_A h(x-x_i)\, dx$$

These integrals might be obtained numerically in the software. It can be shown that in the limit the numerical integration using estimates on a grid will converge to the direct estimate, Myers (1999). The minimized estimation variance can be computed using a slight variation of eq(3).

Derivatives. Essentially the same questions arise with respect to estimation(or interpolation) of the derivative of the unknown function as arise for integrals. If the basis function is differentiable then the derivative of the unknown function might be estimated/approximated by differentiating the interpolating function. This does not result in error bounds for the derivative however. As a linear functional, the values of the derivative can be estimated using eq(1') by only slightly modifying the equations in (2a'), (2b'). Namely it necessary to change the right hand side of the equations in (2a') by using the derivative. It can be shown that this is equivalent to differentiating the interpolating function, this is of course dependent on the differentiability of the basis function. In some cases data for the derivative is an auxiliary variable as discussed in the next section, Chilès and Delfiner (1999).

8 The Vector Valued Case

There are many examples where data is available for a second related variable. The relationship may not be a functional one but rather one of correlation, i.e., statistical dependence. The stochastic model provides a natural way to incorporate the information contained in the data of auxiliary variables. As shown in Myers (1992a) each data "value" is a vector, the generalized covariance becomes a matrix function and the coefficients in (1), (2a), (2b) are matrices. The diagonal entries in the matrix function are covariances, the off diagonal entries are cross-covariances and this matrix function must satisfy a matrix version of positive definiteness. Sometimes the auxiliary variable represents data that is easier or cheaper to obtain and often data will not be available for all variables at all locations. The software can be written to adapt to this. The practical problem is a lack of known positive definite (conditionally positive definite) matrix functions to choose from. For that reason it is common to use what is known as a Linear Coregionalization Model, which is a generalization of a positive linear combination of known positive definite functions, Wackernagel (2003).

9 Simulation

The thin plate spline is obtained by imposing a smoothness condition on the interpolating function, more generally RBF interpolators are obtained by imposing less obvious conditions on the interpolating function. In the stochastic formulation, the estimator is essentially an approximation to the conditional expectation. From the perspective of a random function, there may be multiple realizations that will satisfy the data. As realizations of that random function they all exhibit the spatial correlation implied by the covariance/generalized function. Essentially all interpolation methods "smooth" the data, some more than others. This may not be desirable and in many applications is realistic, think of interpolating elevation for a mountain range. In hydrology it is common to interpolate hydraulic conductivity and then use that parameter in a flow model (stochastic differential equation). It is often important to see how much variation might occur in the ultimate prediction, e.g. best case vs worst case. Thus one wants to generate multiple "equally likely" realizations subject to certain constraints; (1) the spatial correlation structure should be preserved, (2) the marginal distribution should be preserved, i.e., the distribution implicit in the data. This is a form of Monte Carlo but it is more complex because of the constraints. Monte Carlo usually is only constrained by the distribution type. A number of algorithms are in common use, e.g., Sequential Gaussian, Simulated Annealing (which uses the simulated annealing optimization algorithm), Cholesky decomposition (of the covariance matrix) are examples. The Sequential Gaussian algorithm is implemented in *geoR* package for R.

References

- R. Bras and I. Rodríguez-Iturbe,(1985) *Random Functions and Hydrology*, Dover Paperback
- P. Chauvet, J. Pailleux and J.P. Chilès, (1976) Analyse objective des champs météorologiques par cokrigeage. *La Météorologie, Sciences et Techniques, 6ème Série* 4, 37-54
- J.P.Chilès and P.Delfiner, (1999) *Geostatistics: Modeling Spatial Uncertainty*. Wiley & Sons
- N. Cressie, (1989) Geostatistics. *The American Statistician* 43, 197-202
- N. Cressie, (1990) Reply to Letter to the Editor, *The American Statistician* 44, 256-258
- G. de Marsily, (1986) *Quantitative Hydrogeology*., Academic Press, New York
- P.J. Diggle, P.J. Rubeiro Jr and O.F. Christensen, (2003), An Introduction to Model Based Geostatistics. In *Spatial Statistics and Computational Methods*, J. Möller (ed), Springer-Verlag, 43-80
- I.M. Gelfand and N. Ya Vilenkin, (1964) *Generalized Functions*. Vol. 4, Academic Press
- G.R. Liu, K.Y. Dai, Y.T. Gu and K.M. Lim, (2002) A comparison between radial point interpolation method (RIPM) and kriging based meshfree method. In *Advances in Meshless and X-FEM Methods*, G.R. Liu (ed), World Scientific, Singapore
- K.V. Mardia, J.T. Kent, C.R.Goodall, J.A. Little (1996) Kriging and splines with derivative information. *Biometrika* 83, 207-221
- W. R. Madych and S.A. Nelson, (1988), Multivariate interpolation and conditionally positive definite functions. *Approx. Theory & its Appl.* 4, 77-89
- G. Matheron, (1973) The Intrinsic Random Functions and their applications. *Adv. Appl. Prob.* 5, 439-468
- G. Matheron, (1981) Splines and Kriging: their formal equivalence. In: D.F. Merriam. (Ed.), *Downto-Earth Statistics: Solutions Looking for Geological Problems*, Syracuse Univ. of geology contributions Edition. pp. 77–95. (Also see "Splines and kriging: Their formal equivalence," Internal Report, Centre de Geostatistique, Ecole des Mines de Paris, Fountainebleau, 1980)
- D.E. Myers, (1988) Interpolation with Positive Definite Functions. *Sciences de la Terre*, 28, 251-265
- D.E. Myers, (1992a) Cokriging, Radial Basis Functions and the role of Positive Definiteness. *Computers Math. Applications* 24, 139-148
- D.E. Myers, (1992b) Selection of a radial basis function for data interpolation. in Advances in Computer Methods for Partial Differential Eq. VII, R. Vichnevetsky, D. Knight and G.Richter (eds), IMACS, 553-558
- D.E. Myers, (1999) Smoothing and Interpolation with radial basis functions. in *Boundary Element Technolgy XIII*, C.S. Chen, C.A. Brebbia and D.W. Pepper (eds), 365, 376, WIT Press, Southampton
- H. Wackernagel, (2003) *Multivariate Geostatistics* (3rd edition) Springer-Verlag, 3rd edition, Berlin,
- G. Wahba, G.,Kimeldorf, (1970). Spline functions and stochastic processes. *Sankhyä: the Indian Journal of Statistics*: Series A 32 (2), 173–180.
- U. Weller, W. zu Castell, M. Zipprich, M. Sommer and M. Wehrhan, (2002), Kriging and interpolation with radial base functions - a case study. http://ibb.gsf.de/preprints/2003/pp03-14.pdf
- J. Yarus and R. Chambers, (eds), (1994) *Stochastic Modeling and Geostatistics*. AAPG Computer Applications in Geology, No. 3.