Should Geostatistics Be Model-Based?

Kanti V. Mardia
Department of Statistics, University of Leeds, Leeds
E-mail: k.v.mardia@leeds.ac.uk

1 Abstract

Often, there are two streams in statistical research - one developed by practitioners and other by main stream statisticians. Development of geostatistics is a very good example where pioneering work under realistic assumptions came from mining engineers whereas it is only now that statistical framework is getting more transparent. The subject with statistical emphasis has been evolving, as seen by various excellent books from statistical sides (Banerjee, S., Carlin, B.P. and Gelfand, A.E. 2004. Hierarchical Modelling and Analysis for Spatial Data, Chapman and Hall/CRC, New York; Cressie, N. 1992. Statistics for Spatial Data, Wiley, New York; Diggle, P.J. and Ribeiro Jr., P.J. 2007. Model-Based Geostatistics, Springer, New York; Ripley, B.D. 1981. Spatial Statistics, Wiley, New York; Stein, M.L. 1999. Interpolation of Spatial Data: Some Theory for Kriging, Springer, New York). We will mainly discuss here in detail maximum likelihood methods for spatial linear model (kriging). Some other key topics are briefly outlined, including rock fracture modelling and risk assessment for the safe storage of hazardous wastes in underground repositories.

2 Introduction

It is well known that the maximum likelihood (ML) method is a powerful statistical tool in estimation for parametric models. The method of maximum likelihood was introduced by R.A. Fisher as early as 1931 but its application in geostatistics for inference of variogram parameters has been slow and it was first introduced by Mardia (1980) - incidently, this paper was presented to the Geological Congress in Paris in 1980! Mardia and Marshall (1984) and Mardia (1990) have considered the problem in detail. In earth sciences applications, the method was studied by Kitandis (1983) and Kitandis and Lane (1985). Since then, the number of applications has been on increase. Additionally, the computer simulations have allowed numerical studies of problems that are difficult to handle theoretically: small sample theory, the effect of drift on bias of the covariance parameter estimates, robustness of the Gaussian assumption, effect of transformations of the data and transformation on the parameters.

There is no doubt that the availability of cheap and powerful computers has increased the interest in ML applications. With spatially correlated data, the maximization of the likelihood requires considerable computer power mainly for two numerical operations: matrix inversion and maximization in multidimensional parameter spaces.

The method of ML in geoestatistics can be used complementary with the well established method among practitioners of graphical variogram modelling. The variogram plots should be used as a tool for getting hints on selecting a model and the actual unknown parameters of the model could be estimated by ML. Furthermore, ML provides an uncertainty of the estimated parameters what may be used to decide if point estimates are sufficiently accurate or if interval estimation (giving a region of suitable parameters) is more adequate.

We discuss various practical and theoretical problems with MLE (The talk will have several illustrated examples). The conclusion is that the ML method is feasible for geostatistics; it can be implemented efficiently and provides a powerful tool for geostatistical inference. It provides a complete approach to variogram inference offering methodology for model selection, statistical inference and providing measures of uncertainty of the estimated parameters.

Historically, the following observation of Watson (1986) is a key in understanding the development of statistical geostatistics:
"In the mid 1970s the work of Georges Matheron and Jean Serra of the Center for Mathematical Morphology at the Paris School of Mines, attracted my attention. They seemed to be breathing new life into the application of statistics to geology and mining. As a result, I spent a lot of time persuading English-speaking geologists and statisticians that this was so, while trying to persuade the Fontainebleau School to integrate their writings with that of the anglophones! Our geologists receive very little mathematical

training (indeed they seem less mathematiicaly inclined than any scientific group I know) so French-style "geostatistics" was just too much for them."

The following comments on Matheron from Watson (1986) are rather intriguing:
"While I never persuaded Matheron to adopt the "Anglo-Saxon optique", I enjoyed his hospitality on many occasions, thereby sampling French family life (at its best, I suspect) which one can never know as a tourist."

The gap between the mainstream statistical geostatistics and practical geostatistics has narrowed since then!

3 The spatial linear model

Let $\{X(t)\}$ be a stochastic process where t represents a point in d-dimensional space where we write $T = R^d$ for the Euclidean space and $T = Z^d$ for points on a regular lattice. Suppose the process is sampled at points $t_1, t_2, \ldots, t_n$ to give the sample vector $X = \{x(t_1), x(t_2), ..., x(t_n)\}^T \equiv \{x_1, x_2, \ldots, x_n\}^T$, say. Further, we assume that

$$X(t) = \mu(t) + \epsilon(t) \tag{1}$$

where $\{\epsilon(t)\}$ is Gaussian with zero means,

$$\mu(t) = E\{X(t)\} = f(t)^T\beta, \tag{2}$$

β is a q-by-1 parameter vector and $f(t)$ is a vector of known functions, possibly monomials, which may describe a trend. Suppose that $\epsilon(t)$ is second-order stationary with

$$\text{Cov}\{\epsilon(t), \epsilon(t+h)\} = \text{Cov}\{X(t), X(t+h)\} = \sigma(h;\theta), \tag{3}$$

where $\sigma(\cdot;\theta)$ is a positive definite function of h, assumed known apart for a p-by-1 vector of parameters θ. Thus

$$E(X) = F\beta, \tag{4}$$

where $F = \{f(t_1), f(t_2), \ldots, f(t_n)\}^T$. Let the covariance matrix of X be $\Sigma = \Sigma(\theta)$ with

$$\sigma_{ij} = (\Sigma)_{ij} = \{\sigma(t_i - t_j;\theta)\}. \tag{5}$$

Our model is of the form

$$\text{observation} = \text{deterministic trend} + \text{stochastic fluctuation}$$

where trend measures the long-term variation whereas the stochastic fluctuation measures the short-term or the local variation. We can specify $\Sigma(\theta)$ by modelling $\sigma(h;\theta)$ directly as done in Geostatistics, see Section 3.

4 ML estimation

<u>ML equations.</u> In this section, we follow Mardia (1980) and Mardia and Marshall (1984). From Section 2, since X is multivariate normal, the log-likelihood function of X with the parameters (β, θ) is

$$l = l(X;\beta,\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}(X - F\beta)^T[\Sigma(\theta)]^{-1}(X - F\beta). \tag{6}$$

Hence, the MLE of β is given by

$$\hat{\beta} = (F^T\hat{\Sigma}^{-1}F)^{-1}F^T\hat{\Sigma}^{-1}X \tag{7}$$

where $\hat{\Sigma} = \Sigma(\hat{\theta}), \hat{\theta} =$ MLE of θ. Next consider the scale parameter. Note that we could write $\theta^T = (\theta_1, \theta_2^T)$, where $\theta_1 = \sigma^2$ is such that $\Sigma(\theta) = \sigma^2 P(\theta_2)$ and $P(\theta_2)$ represents a correlation matrix. Then the ML equations for β and σ^2 are

$$\hat{\beta} = (F^T\hat{P}^{-1}F)^{-1}F^T\hat{P}^{-1}X, \tag{8}$$

$$\text{and} \quad \hat{\sigma}^2 = [(X - F\hat{\beta})^T\hat{P}^{-1}(X - F\hat{\beta})]/n. \tag{9}$$

Profile likelihood. One method to check that a solution to the ML equations for the global maximum is to plot the profile likelihood when the number of covariance parameters are very few. Substituting $\hat{\beta}$ and $\hat{\sigma}^2$ into the log-likelihood, we get the *profile likelihood*

$$l_p(X;\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|P(\theta)| - \frac{n}{2}\log\{[X - F\hat{\beta}(\theta)]^T P(\theta)^{-1}[X - F\hat{\beta}(\theta)]\},$$

where $\hat{\beta}(\theta) = [F^T\Sigma(\theta)^{-1}F]^{-1}F^T\Sigma(\theta)^{-1}X$. If $\hat{\theta}$ maximizes $l_p(X;\theta)$, then $\hat{\theta}, \hat{\beta}(\hat{\theta})$ and $\hat{\sigma}^2(\hat{\beta},\hat{\theta})$ also maximize this likelihood. Usually θ is a scalar, especially for the stationary case, so that it is relatively easy to plot $l_p(X;\theta)$ rather than $l(X;\theta,\sigma^2)$. Also, we then can obtain $\hat{\beta} = \hat{\beta}(\hat{\theta})$ and $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\theta})$ from (8) and (9).

Covariance functions. One of the most popular covariance functions with statisticians is what is called Matern scheme given by

$$\{2/\Gamma(\nu)\}(r/2c)^\nu K_\nu(r/c),\ r = |h| > 0, \nu > 0, c > 0,$$

where $K_\nu(.)$ is the Bessel function of the second kind and order ν, and c is a scale parameter. In fact, the parameter ν determines the differentiability of the underlying process; the special case $\nu = 0.5$ and $\nu = \infty$ correspond to the exponential scheme and the Gaussian scheme respectively. This scheme was introduced by Matern (1960) but Whittle (1954) proposed the use of this scheme with $\nu = 1$ in 2 dimensions. This scheme for $\nu = 0.5$ can be related approximately to the spherical scheme.

5 The nugget case

The Model. Consider

$$\mathrm{Cov}(X) = \sigma^2 P + \psi^2 I, \tag{10}$$

where P is a correlation matrix with $\rho_{ij} = \rho(t_i - t_j)$, $\mathrm{Cov}(x_i, x_j) = \sigma^2\rho_{ij} + \psi^2\delta_{ij}$, $\rho(0) = 1$, and $\delta_{ij} = 1$ if $i = j; = 0$ if $i \neq j$. We will call ψ^2 a nugget parameter. We could write the underlying process (with zero drift) as

$$X(t_i) = \epsilon(t_i) + \eta(t_i), \tag{11}$$

where $\mathrm{Cov}(\epsilon(t_i), \epsilon(t_j)) = \sigma^2\rho(t_i - t_j), \rho(0) = 1,\ \mathrm{Var}[\eta(t_i)] = \psi^2$, $\mathrm{Cov}\{\epsilon(t_i), \eta(t_i)\} = 0$.

Hence (11) is an errors-in-variable model. It is important to distinguish here whether the process is continuous or discrete.

A profile likelihood for the nugget case. Let us now take the vector θ with just three parameters, and write σ^2, $\delta = \psi^2/\sigma^2$, and θ, where we now assume that θ is a correlation parameter in $P(\theta)$. For a given θ, we can use the spectral decomposition of $P(\theta)$, so that $P(\theta) = C(\theta)^T\Lambda(\theta)C(\theta)$, where $C(\theta)$ is an orthogonal matrix, and $\Lambda(\theta) = \mathrm{diag}\{\lambda_1(\theta), \lambda_2(\theta), \ldots, \lambda_n(\theta)\}$. As before, suppose the "trend" is $F\beta$. Then the log-likelihood is simply

$$l = -(n/2)\log(\sigma^2) - (1/2)\sum_{i=1}^n \log[\delta + \lambda_i(\theta)] - \sum_{i=1}^n u_i^2(\theta,\beta)/\{(2\sigma^2)[\delta + \lambda_i(\theta)]\}, \tag{12}$$

where $u(\theta,\beta) = C(X - F\beta)$. It can be shown, as in Mardia (1980), that

$$\hat{\sigma}^2(\delta,\theta,\beta) = \{\sum_{i=1}^n u_i^2[\delta + \lambda_i(\theta)]^2\}/\sum_{i=1}^n[\delta + \lambda_i(\theta)]^{-2}. \tag{13}$$

The ML equations for δ and β do no depend on σ^2, and are

$$\hat{\beta}(\theta,\delta) = [F^T C(\Lambda + \delta I)^{-1}C^T F]^{-1}C^T(\Lambda + \delta I)^{-1}CX, \tag{14}$$

$$\text{and}\quad \left[\sum_{i=1}^n u_i(\theta,\hat{\beta})^2[\hat{\delta} + \lambda_i(\theta)]^{-1}\right]\left[\sum_{i=1}^n[\hat{\delta} + \lambda_i(\theta)]^{-2}\right] = n\sum_{i=1}^n u_i(\theta,\hat{\beta})^2[\hat{\delta} + \lambda_i(\theta)]^{-2}. \tag{15}$$

Now in substituting $\hat{\beta}$ from (14), for a given θ, a solution can be obtained in terms of $\delta(\theta)$. In turn, we substitute δ into (14) to obtain $\hat{\beta}(\theta,\hat{\delta})$, and then from (13) we get $\sigma^2 = \sigma^2(\hat{\delta},\theta,\hat{\beta})$. Thus the profile likelihood with respect to θ can be obtained from (12). Hence, the ML estimate of θ can be derived, which in turn gives the MLEs of δ, β, and σ^2. Mardia and Gill (1982) gave some early interpretations.

6 Restricted Maximum Likelihood (REML)

Assume again the model $X = F\boldsymbol{\beta}$ + error, but now the error process may be only increment stationary, ie. we are dealing with intrinsic random function (IRF). The main principle of the REML is to project the data in a hyperplane perpendicular to F, and work on the marginal likelihood in that plane, namely on

$$X^* = PX, P = I - F(F^TF)^{-1}F^T \tag{16}$$

where $PF = 0$ is the hyperplane. The ML can be biased so REML is used for estimating θ as above. For implementation, we can take an ortho-normal basis in this hyperplane. Many distribution takes a form independent of basis but not in the intrinsic case (see, for example Mardia *et al.*, 2006). There are two main approaches: (1) Basis free method (ie. implicit basis only) and (2) Basis explicitly specified. In the first approach, we have the marginal distribution as singular normal distribution, so that the pdf is given by

$$|2\pi P\Sigma P|^{-\frac{1}{2}} \exp[-\frac{1}{2}X^T(P\Sigma P)^- X]. \tag{17}$$

In the second approach, we chose a basis and the work on co-ordinates. Thus we have various specific choices. We can work on X_2^* defined by $X_2^* = A_2X$ where $A_2^TF = 0$. Now we can write fully the distribution of X_2^* where $\text{Cov}(X_2^*)$ is of the full rank. Also $A = (A_1, A_2)$ is non-singular. In this transformation, it is clear that X_1^* does not give any information on θ. Whether the full likelihood is better or the REML, it depends except for a singular Σ (the intrinsic case) the choice is clear - you have to have REML! Harville (1977) simplified (17) when Σ^{-1} exists. Let

$$X_1^* = A_1^TX, \quad X_2^* = A_2^TX; \quad A_1 = F(F^TF)^{-\frac{1}{2}}, A_2^TF = 0 \tag{18}$$

so that A_1 is an orthogonal basis of F, and the columns of A_2 are orthonormal to F. Thus $A = (A_1, A_2)$, $\Sigma^* = A^T\Sigma A$. The REML principle says use the resulting likelihood which from (17) is given by

$$|\Sigma_{22}^*|^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}X_2^{*T}(\Sigma_{22}^*)^{-1}X_2^*\}$$

and if Σ^{-1} exists then Harville (1977) has shown that

$$|\Sigma_{22}^*| = |\Sigma|\ |F^T\Sigma^{-1}F|/|F^TF|, X_2^{*T}\Sigma_{22}^{*-1}X_2^* = (X - F\hat{\beta})^T\Sigma^{-1}(X - F\hat{\beta}).$$

Hence we are back to the original quantities X, β, Σ, etc. in the likelihood. Although REML is widely recommended for geostatistical models, it can be more sensitive than ML to the chosen model for the drift (see, for example, Diggle and Ribeiro Jr., 2007, p.117). Also it can have higher MSE.

7 General Issues with MLE

We only consider estimation of θ since the drift estimation is generally not a challenge.

Multimodality. The likelihood could have several modes for θ depending on the scheme but there seems to be always a global maximum. There has been a very close scrutiny of the behaviour of the MLE for the spatial linear model (see Christensen, 2004; Mardia and Watkins, 1989; Ripley, 1988; Smith, 2000; Stein, 1999; Warnes and Ripley, 1987).

For the multimodality, there could be several reasons, for examples, the likelihood maybe flat, e.g. due to a wrong model; the scheme could be non-differentiable, e.g.,the spherical scheme with respect to the range parameter; the nugget parameter can be on the boundary; the numerical method may not give enough accuracy.

There are some effective solutions. It is found that a modest increase in additional number of the model parameters can resolve some of the issues. But it is always wise to check various points: the underlying geometry, long range correlations versus trend; stability of the method for finding the global maximum, and so on. Thus except for these precautions, one can handle even extreme situations!

Model Indeterminancy. We can assess the competing models by the Akaike Information Criterion. However, it is worthwhile to check whether to fit a higher order drift and white noise vs simply coloured error with long range correlation: expert opinion can help but the drift can be poor in extrapolation.

Computational Procedures. With increase in the computational power, the ML method is becoming easier to implement for even large data sets. Now inverting even a 500×500 matrix is feasible. Various methods are available - methods using derivatives of likelihood versus methods not using the derivatives but using the likelihood itself. For small numbers of parameters, grid method seems to be adequate. For very large data set Vecchia (1988) type selection of neighbourhood (a pseudo-likelihood type approach) is found to be useful even when $n >> 500$!

8 Generalized Linear Geostatistical Models (GLGM)

We now describe a versatile model of Diggle *et al.* (1998). Let $S(t)$ be a stationary Gaussian process with mean zero and variance σ^2, and correlation function $\rho(h)$. Suppose that the observations $x_i|S(.), \quad i = 1, \ldots, n$ are conditionally independent with

$$E\{x_i|S(.)\} = g(\alpha + S_i) = \mu_i \text{ say}, \quad \text{Var}\{x_i|S(.)\} = v(\mu_i).$$

The function of $g(.)$ is the analytic inverse of the link function of Generalized Linear Model. We can now obtain the semi-variogram of the process S as (Diggle and Ribeiro Jr., 2007; Diggle *et al.*, 1998):

$$\gamma_s(h) = \frac{1}{2}[E_S\{g(\alpha + S_i) - g(\alpha + S_j)\} + 2E_S\{v(g(\mu + S_i))].$$

All this looks daunting but the model can be implemented even for non-Gaussian cases such as Poisson, Binomial, etc. Indeed, there is a supporting library in R called GeoR. For the Poisson case, we have $g(y) = \exp\{\alpha + y\} = v(y)$. Further, we have

$$\text{cov}(x_i, x_j) = e^{\alpha + \frac{1}{2}\sigma^2} + e^{2\alpha + \sigma^2}(e^{\sigma^2} - e^{\sigma^2 \rho(h)}), h = \|t_i - t_j\|.$$

The aim is to explore the latent process $S(.)$ as for dynamic linear model. In general, the expressions are not straightforward so one has to resort to numerical work.

9 Discussion

Of course, Matheron was aware of the ML method (see Matheron, 1971, pp.155-156). But he was more concerned with the drift component rather than the covariance parameters and his comments on p.156 are worth noting (the equation (4-2-7) below refers to the universal kriging equations and S to the Domain): "Our optimal estimator is more general than the maximum likelihood estimator, for it is not related to a Gaussian hypothesis and is generalized to the case where S is infinite. In applications, the systems (4-2-7) is easier to solve, as it requires only one matrix inversion."

Within Geostatistical models, there are many other key ideas: for non-Gaussian cases the use of the Box and Cox transformation (for example, Mardia and Goodall, 1993), link between kriging and thin plate spline (for example, Kent and Mardia, 1984), use of spectral representation for covariances (for example, Kent and Mardia, 1996), kriging with derivative (for example, Mardia *et al.*, 1996), CAR and SAR models (for example, Mardia, 1990). There are other estimation methods , e.g. minimum norm quadratic estimators (for example, Mardia and Marshall, 1985). Directional statistics is another area of great importance (for examples, Mardia, 1981a,b, 1989, 2002; Jupp and Mardia, 2000). Incidentally ,the directional statistics paper ,Mardia (2002), was presented as a plenary talk to the IAMG Conference in Berlin!

We have not covered here various other key topics. Bayesian methods in geostatistics are becoming effective (see, for example, Banerjee *et al.*, 2004; Diggle and Ribeiro Jr., 2007). Another area is of spatial temporal modelling through Kriged Kalman filter with kriging providing principal fields (Mardia *et al.*, 1998; Sahu and Mardia, 2005). Beyond the spatial linear models, there are many other type of statistical models ,eg. for rock fractures (Mardia *et al.*, 2007a; Xu *et al.*, 2006) to assess risk for the safe storage of hazardous wastes in underground repositories.

All in all, the interplay between "statisticians" and "geostatisticians" has really become stronger, and we end with a few quotations on modelling (Speed, 2007):

George Box: "All models are wrong, some models are useful".

John Tukey: "Our focus should be on questions, not models . . . Models can - and will - get us in deep trouble if we expect them to tell us what the unique proper questions are".

Basil Rennie: "All thought and all communication is modelling, and most misunderstandings arise by someone confusing either a model with reality or one model with another. Every model embodies a half-truth, and as one of our wiser politicians once remarked, half-truths are like half-bricks, they are better because they carry further".

Answer to all these questions lies in Holistic Statistics (see, Mardia and Gilks, 2005)!

10 Acknowledgements

I wish to thank Eulogio Pardo-Iguzquiza for his help in preparing this paper; our joint paper with full details will appear elsewhere.

References

Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004). Hierarchical Modelling and Analysis for Spatial Data. Chapman and Hall/CRC, New York.

Christensen, O.F. (2004). Monte Carlo maximum likelihood in model-based geostatistics. Journal of Computational and Graphical Statistics, 13, pp.702-718.

Cressie, N. (1993). Statistics for Spatial Data, Wiley, New York.

Diggle, P.J. and Ribeiro Jr., P.J. (2007). Model-Based Geostatistics. Springer, New York.

Diggle, P.J., Tawn, J.A. and Moyeed, R.A. (1998). Model based geostatistics (with discussion). Applied Statistics, 47, pp.299-350.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72, pp.320-340.

Jupp, P.E. and Mardia, K.V. (2000). Directional Statistics. J. Wiley & Sons, Chichester.

Kent, J.T. and Mardia, K.V. (1994). Link between kriging and thin plate splines. In: Kelly F.P. (Ed.) Festschrift Volume to P. Whittle. Probability, Statistics and Optimization. Wiley, pp.325-339.

Kent, J.T. and Mardia, K.V. (1996). Spectral and circulant approximations to the likelihood for stationary Gaussian random fields. Journal of Statistical Inference and Planning, 50, pp.379-394.

Kitandis, P.K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. Water Resources Research, 19, pp.909-921.

Kitandis, P.K. and Lane, R.W. (1985). Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton methods. Journal of Hydrology, 79, pp.53-71.

Mardia, K.V. (1980). Some statistical inference problems in Kriging II: Theory. In: Proceedings 26th International Geology Congress Sciences de la Terre: Advances in Automatic Processing and Mathematical Models in Geology, Series "Informatique Geologie", 15, pp.113-131.

Mardia, K.V. (1981a). Are statisticians helping earth sciences enough?. In: Mardia, K.V. (Ed.) Communications in Statistics, Special Issue on Statistics in Earth Sciences, A10, 15, pp.1-2.

Mardia, K.V. (1981b). The evolution of directional models in geosciences since Fisher. In: Merriam, D.F. Down-To-Earth Statistics: Solution Targeting Geological Problems, Syracuse Univ. Geology Contributions 8, pp.39-46.

Mardia, K.V. (1988). Multi-dimensional multivariate Gaussian Markov random fields. Journal of Multivariate Analysis, 24, pp265-284.

Mardia, K.V. (1989). Statistics in Earth Sciences. In: Mardia, K.V. (Ed.) Special Issue of Communications in Statistics. Marcel Decker, A10, 15.

Mardia, K.V. (1990). Maximum likelihood estimation for spatial models. In: Griffith, D.A. (Ed.) Proceedings Spatial Statistics: Past, Present and Future, Institute of Mathematical Geology, Michigan Document Service. pp.203-25.

Mardia, K.V. (2002). Why is directional statistics pivotal to geosciences?. Opening Address at the 8th Annual Conference IAMG Proceedings, 1, pp.41-47.

Mardia, K.V., Bookstein, F.L., Kent, J.T. and Meyer, C.R. (2006). Intrinsic random fields and image deformations. Journal of Mathematical Imaging and Vision, 26, pp.59-71.

Mardia, K.V. and Gilks, W. (2005). Meeting the statistical needs of 21st-century science. Significance, 2, pp.162-165.

Mardia, K.V. and Gill, C.A. (1982). Some statistical inference problems in kriging I: numerical applications. Journal of the Geological Society, Computer Applications in Geology. pp.56-72.

Mardia, K.V. and Goodall, C.R. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In: Patil, G.P. and Rao, C.R. (Eds.) Multivariate Environmental Statistics, 6, pp.347-385.

Mardia, K.V., Goodall, C., Redfern, E.J. and Alonso, F.J. (1998). The Kriged Kalman filter. Discussion paper. Presented to the Spanish Statistical Society. TEST, 7, pp.217-276.

Mardia, K.V., Kent, J.T., Goodall, C.R. and Little, J.A. (1996). Kriging and splines with derivative information. Biometrika. 83, pp207-221.

Mardia, K.V. and Marshall, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika, 71, pp.135-46.

Mardia, K.V. and Marshall, R.J. (1985). Some minimum norm quadratic estimators of the components of spatial covariance. Journal of the International Association of Mathematical Geology, 17, pp.517-525.

Mardia, K.V., Nyirongo, V.B., Walder, A.N., Xu, C., Dowd, P.A., Fowell, R.J. and Kent, J.T. (2007). A line finding assignment problem and rock fracture modelling. In: Upadhaya, S.K., Singh, U. and Dey, D.K. (Eds.) Bayesian Statistics and its Applications, Anamaya Publisher, New Delhi, pp.319-330.

Mardia, K.V., Walder, A.N., Xu, C., Dowd, P.A., Fowell, R.J., Nyirongo, V.B. and Kent, J.T. (in press). MCMC implementation of rock fracture modelling. Mathematical Geology.

Mardia, K.V. and Watkins, A.J. (1989). On multimodality of the likelihood in the spatial linear model. Biometrika, 76, pp289-295.

Matern, B. (1960). Spatial Variation, 2nd edn., Springer, Berlin.

Matheron, G. (1971). The Theory of Regionalzed Variables and its Applications.École Nationale Superiéure des Mines de Paris.

Ripley, B.D. (1981). Spatial Statistics, Wiley, New York.

Ripley, B.D. (1988). Statistical Inference for Spatial Processes, Wiley, New York.

Sahu, S.K. and Mardia, K.V. (2005). A Bayesian kriged Kalman model for short-term forecasting of air pollution levels. Applied Statistics, 54, pp.223-244.

Smith, R.L. (2000). Spatial statistics in environmental science. In: Fitzgerald, W.J., Smith, R.L., Walder, A.T. and Young, P.C. (Eds.) Non-Linear and Non-Stationary Signal Pprocessing, Cambridge University Press, pp.152-183.

Speed, T. (2007). Terence's stuff: model skeptics. IMS Bulletin, 36, p.11.

Stein, M.L. (1999). Interpolation of Spatial Data: Some Theory for Kriging, Springer, New York.

Vecchia, A.V. (1988). Estimation and model identification for continuous spatial processes. Journal of the Royal Statistical Society, 50B, pp.297-312.

Warnes, J.J. and Ripley, B.D. (1987). Problems with likelihood estimation of covariance functions of spatial Gaussian processes. Biometrika, 74, pp.640-642.

Watson, G.S. (1986). A boy from the bush. In: Gani, J.M. (Ed.), The Craft of Statistical Modelling, Springer, New York, pp.43-60. Also reprinted in: Mardia, K.V. (Ed.) The Art of Statistical Science, Wiley, New York, pp.xxix-lii.

Whittle, P. (1954). On stationary processes in the plane. Biometrika, 41, pp.434-449.

Xu, C., Dowd, P.A., Mardia, K.V. and Fowell, R.J. (2006). A connectivity index for discrete fracture networks. Mathematical Geology, 38, pp.611-634.

Xu, C., Dowd, R.A., Mardia, K.V., and Fowell, R.J. (in press).A comparison of methods for the stochastic simulation of rock fractures. Mathematical Geology.