Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

ESTIMATING FUNCTIONS FOR SEMIVARIOGRAM ESTIMATION

Subhash Lele
The Johns Hopkins University
and
Montana State University

Abstract

This paper proposes the use of estimating functions based on composite likelihood for the estimation of semivariogram parameters. These estimating functions eliminate the need for the subjective specification of the range and bin width parameter at the same time retain the model robustness of the classical procedures based on the method of moments. Improvement in the efficiency can be anywhere from 50 to 100%. Further extensions and uses are discussed. They attest to the power and flexibility of the composite likelihood approach.

Key words: Composite likelihood, geostatistics, kriging, restricted maximum likelihood, variogram.

1 Introduction

Various scientific disciplines require the collection and prediction of data over space. For example, in mining where the goal is to predict ore concentration over the entire study area, samples are collected at various locations. To predict concentrations at locations where the samples are not collected, geostatistics uses a technique known as kriging. Kriging produces a map of ore concentrations for the entire site which can be used for planning and operating mining activities. This same technique has applications in environmental data collection where the goal is to predict environmental degradation or clean-up based on data collected at a discrete number of monitoring locations at a site. As in mining, a useful tool for site assessment and clean-up of a contaminant site is a contour map of contaminant concentrations over the area of interest. Environmental decision makers then could use this map to identify those areas which should be excavated to protect public health, those which pose little or no risk, and those where the uncertainty is large enough to warrant additional sampling.

The attraction of the kriging procedure in these applications is twofold. First, it offers a statistical justification for the way it takes point data (data from locations that have been sampled) and generates a smooth, interpolated map (i.e., a contour plot) of contaminant concentrations. Second, the kriging procedure generates explicit uncertainty measures (e.g., prediction intervals) for the interpolated and smoothed estimates-both for estimates of concentrations at particular locations, and for estimates of averages within a defined area. These uncertainty measures can be used for building precise margins of safety into a decision rule.

We refer the reader to Cressie (1991) or Journel and Huijbregt (1978) for the details of kriging. But the basic idea behind kriging is simple to explain. We observe the phenomenon under study only at a finitely many sampled locations. We want to predict the values of the phenomenon at the unobserved locations. Obviously we need to model the relationship between the data that are observed and the data that need to be predicted. This model generally has two components: the trend component and the spatial association component. Once such a model is specified, it can be used to make the prediction for the observations at the unsampled locations such that the prediction error is minimized.

In practice, although a parametric form of a model can be specified with reasonable confidence, one has to estimate the parameters based on the observed data. The purpose of this paper is to introduce the framework of estimating functions for the estimation of the spatial association, modeled by semivariogram, and use it to compare and improve existing methods of estimation. The main results of the paper are:

1. A new class of estimation procedures based on composite likelihood (Lindsay, 1988) is introduced. It is shown that in the case of ordinary kriging, this method leads to an estimating function similar to the popular method of weighted least squares (Cressie, 1991, page 96).

2. A composite likelihood for simple kriging leads to estimating functions, described by Godambe and Thomson (1989), that combine the information in conditional mean and conditional variance structure. This is shown to lead to a substantial improvement in the efficiency.

3. Composite likelihood estimating functions can be interpreted as minimizing prediction error. Kriging is used for prediction. It makes sense that the estimation of the semivariogram should also be based on minimization of prediction error.

4. The use of the composite likelihood, similar to maximum likelihood and restricted maximum likelihood methods, eliminates the need for subjective choices of range and bin width parameters needed for the classical methods; at the same time, it retains the model robustness properties of the classical methods.

5. Godambe's optimality criterion is used for comparing finite sample performance of different methods. It is shown that the use of composite likelihood can improve efficiency anywhere from 50 to 100% over the classical least squares method.

6. Extensions of these ideas to the case of universal kriging as well as robustness issues are indicated.

2 Classical semivariogram estimation

We assume that the reader is familiar with the basic ideas about kriging and semivariogram. Cressie (1991) or Journel and Huijbregt(1978) are excellent references. The following is a very brief introduction to the required concepts.

We use the following notation.

s denotes the location coordinates.

$U(s)$ denotes the value of the process at location s.

$\gamma_u(s_i, s_j)$ denotes the semivariogram for the U process. It is given by:

$$\begin{aligned} \gamma_u(s_i, s_j) &= \frac{1}{2} var(U(s_i) - U(s_j)) \\ var(U(s)) &= \sigma^2 \end{aligned}$$

Simple kriging corresponds to $E(U(s)) = 0$. Ordinary kriging corresponds to $E(U(s)) = \mu$.

Isotropic semivariograms:

Suppose that

$$\gamma_u(s_i, s_j) = \gamma_u(d(s_i, s_j))$$

that is, the semivariogram depends only on the distance between the locations. Such semivariograms are called isotropic semivariograms. The covariance, when it exists, is given by

$$C_u(d(s_i, s_j)) = \gamma_u(+\infty) - \gamma_u(d(s_i, s_j))$$

It is obvious that in practice one has to estimate the parameters involved in the semivariogram or covariogram models using the available data. There are two statistical questions that become relevant here. One is related to the methods of estimation and the other relates to the properties of the estimators obtained from various methods. This section will describe various methods of estimation. They all can be looked upon as particular cases of the method based on estimating functions.

Let us begin the discussion with two simple but important cases of simple and ordinary kriging. For the sake of simplicity, let us assume that $\sigma^2 = 1$.

The classical estimation procedure for the estimation of the variogram can be described in an algorithmic form as follows.

1. Calculate all pairwise distances between locations.
$$d(s_i, s_j) = |s_i - s_j|$$
2. Calculate the squared deviations for the observations (empirical semi-variogram).
$$\nu(s_i, s_j) = \frac{1}{2}(u(s_i) - u(s_j))^2$$
3. Plot $\nu(s_i, s_j)$ (on the y-axis) versus $d(s_i, s_j)$ (on the x-axis).
4. If the parametric model to be fitted is
$$\gamma_u(s_i, s_j) = 1 - \theta^{d(s_i,s_j)}$$
then the parameter θ is estimated by minimizing the following
$$\sum_{i \le j}(\nu(s_i, s_j) - 1 + \theta^{d(s_i,s_j)})^2$$

One can also use some robust criterion such as absolute deviations instead of squared deviations; or one can use the weighted least squares criterion.

Usually the values of $\nu(s_i, s_j)$ (being chisquare random variables) are quite scattered. Some smoothing of these values through the use of local averages is suggested before attempting the fitting. The suggested values of the span in the local averaging are such that there are at least 30 points within a window. This is the *bin width*. One also has to throw away those pairs which are 'too far' apart, that is are outside the *range* parameter. This method of estimation is sensitive to the bin width and range parameters.

Maximum likelihood (ML) and restricted maximum likelihood (REML) approach:

If one assumes a Gaussian distribution, one can also use the method of maximum likelihood to estimate the parameter in the covariogram. In this case, the value of θ is obtained by maximizing the following.

$$L(\theta,\mu) = \frac{-1}{2}\log|C(\theta)| - \frac{1}{2}(U-\mu)^T C(\theta)^{-1}(U-\mu) \tag{1}$$

where U is the vector of observations, μ is the vector of means and $C(\theta)$ is the covariance matrix with $C_{ij}(\theta) = \theta^{d(s_i,s_j)}$.

It is advisable to eliminate the nuisance parameter μ before estimating θ. This is achieved by considering the likelihood of the contrasts. Consider the vector of contrasts $u_c = \{U(s_i) - U(s_1), i = 2,3,..,n\}$. It is easy to see that this vector corresponds to multiplying the original data vector U by a matrix A such that its first column consists of -1 and the i-th column consists of zeros except in the i-th place. Thus:

$$V = AU \sim N(0, AC(\theta)A^T)$$

The likelihood corresponding to V is quite complicated. It is given by

$$L(\theta,u_c) = \frac{1}{(2\pi)^{n/2}\,|\Psi(\theta)|^{1/2}}\exp\{-\frac{1}{2}u_c^T\Psi^{-1}(\theta)u_c\} \tag{2}$$

where $\Psi_{ii}(\theta) = 2\gamma_u(s_i,s_1;\theta)$ and $\Psi_{ij}(\theta) = \gamma_u(s_i,s_1;\theta) + \gamma_u(s_j,s_1;\theta)$ $- \gamma_u(s_i,s_j;\theta)$. Notice that this is a function of θ only. Maximizing this with respect to θ yields the REML estimator.

Following is a summary of the merits and demerits of these classical methods of estimation of variogram parameters.

1. The method of moments estimator does not require specification of the Gaussian or any particular distribution. On the other hand, it requires that the bin width and range be specified.

2. The methods of ML or REML theoretically yield an optimal estimator but require a full specification of the probabilistic model. Moreover, they involve inversion of large matrices. This can be computationally prohibitive. Uniqueness of the maximum is also not always guaranteed.

3. Although the variogram is used for the purpose of prediction, this purpose does not seem to enter the classical estimation procedures. Notice also that classical MLE implicitly includes a term that derives from prediction error.

3 Estimating functions based on composite likelihood

The idea of composite likelihood, although discussed in various disguises such as Pseudolikelihood (Besag, 1975) or Partial Likelihood (Cox, 1975), was developed in its own right by Lindsay (1988). There are two motivations for constructing the composite likelihoods: first, they provide a substitute method of estimation when maximum likelihood is very difficult to calculate; secondly, they sometimes represent that portion of the model we are most comfortable with modelling and the resultant estimators can be consistent even when full maximum likelihood estimators are not, a form of consistency robustness. Let us concentrate on the first aspect for now.

It is quite clear that the likelihood function in equation (1) is extremely difficult to deal with computationally. The same holds true for the Restricted Likelihood based on the contrasts. A natural question to ask would be: can we approximate the likelihood function by something that behaves almost like a likelihood but is easy to deal with, both computationally and mathematically?

We will start with the simplest case of 'simple kriging' and then generalize it to the case of 'ordinary kriging'. Again we assume that $\sigma^2 = 1$.

3.1 Simple kriging:

In this case, we assume that

$$\begin{aligned} E(U(s_i)) &= 0 \\ var(U(s_i)) &= 1 \\ cov(U(s_i), U(s_j)) &= (1 - \gamma_u(s_i, s_j; \theta)) \end{aligned}$$

Then, under the Gaussian assumption, it is clear that $U \sim N(0, C(\theta))$ and the likelihood can be written as $L(\theta, u) = f(u(s_1), u(s_2), ..., u(s_n); \theta)$. Now, following Lindsay (1988), suppose we approximate this likelihood by the product of two dimensional marginal densities, namely:

$$CL_0(\theta, u) = \prod_{i=1}^{n-1} \prod_{j>i} f(u(s_i), u(s_j); \theta)$$

This is what is called a 'composite likelihood' because it is a composition of two dimensional marginal likelihoods. Consider the estimating function generated by this 'composite likelihood'.

$$\sum_{i=1}^{n-1} \sum_{j>i} \frac{d}{d\theta} \log f(u(s_i), u(s_j); \theta)$$

This is a zero unbiased estimating function. Notice that the composite likelihood involves only two dimensional densities and hence is computationally substantially simpler than the total likelihood $L(\theta, u)$.

These estimating functions have a very intuitive appeal. For the sake of illustration, consider the case corresponding to the isotropic exponential variogram model. The negative log-composite likelihood is, then, given by

$$\sum_{i=1}^{n}\sum_{j>i}\frac{1}{2(1-\theta^{2d(s_i,s_j)})}(u(s_i)-\theta^{d(s_i,s_j)}u(s_j))^2+\frac{1}{2}\log(1-\theta^{2d(s_i,s_j)})$$

This is minimized with respect to θ. Notice that the first term of this expression is just the weighted prediction error, where $u(s_i)$ is being predicted by $\theta^{d(s_i,s_j)}u(s_j)$. The second term can be interpreted as a smoothing factor or the factor that makes the estimating function unbiased. This factor, surprisingly does not depend on the assumption of Gaussianity of the underlying process but only needs existence of the variance. (In fact, a similar justification of minimizing the prediction error can be given to the full likelihood as well.) Kriging is a tool for prediction. It makes sense that we estimate the parameters of the semivariogram based on their prediction performance (Marcotte, 1995).

Let us look at the estimating function generated by the composite likelihood for simple kriging. It is simple to check that it has the following form.

$$\sum_{i=1}^{n}\sum_{j>i}\frac{d(s_i,s_j)\theta^{d(s_i,s_j)-1}}{(1-\theta^{2d(s_i,s_j)})}(u(s_i)-\theta^{d(s_i,s_j)}u(s_j))$$
$$+\sum_{i=1}^{n}\sum_{j>i}\frac{d(s_i,s_j)\theta^{2d(s_i,s_j)-1}}{(1-\theta^{2d(s_i,s_j)})}\left\{\frac{(u(s_i)-\theta^{d(s_i,s_j)}u(s_j))^2}{(1-\theta^{2d(s_i,s_j)})}-1\right\}$$

Notice that this is a linear combination of two estimating functions; the first one uses only the first moment and the second one uses the second moment of the process. This estimating function is zero unbiased when the conditional mean is linear and the marginal variances exist. This holds for probability structures more general than the Gaussian probability structure and thus retains the model robustness of the classical estimators. This, intuitively, is also a better use of the available information than using only the second moment information as done by the classical approaches. See Godambe and Thompson (1989) for a discussion of such estimating functions.

3.2 Ordinary kriging:

In this case, we assume that

$$E(U(s_i)) \;=\; \mu$$

$$\begin{aligned} var(U(s_i)) &= 1 \\ cov(U(s_i), U(s_j)) &= (1 - \gamma_u(s_i, s_j; \theta)) \end{aligned}$$

We can write down a composite likelihood quite simply. Consider the product of the marginal densities of the contrasts $v_{ij} = U(s_i) - U(s_j)$, namely,

$$CL_1(\theta, v) = \prod_{i=1}^{n-1} \prod_{j>i} f(v_{ij}; \theta)$$

Let us look at this particular case in detail. In the Gaussian case, notice that

$$f(v_{ij}; \theta) = \frac{1}{\sqrt{2\pi}\sqrt{2\gamma_u(s_i, s_j; \theta)}} \exp\{-\frac{1}{4\gamma_u(s_i, s_j; \theta)}(U(s_i) - U(s_j))^2\}$$

Hence, ignoring constant terms, negative log-composite likelihood upto a constant can be written as

$$\sum_{i=1}^{n-1} \sum_{j>i} \frac{1}{2\gamma_u(s_i, s_j; \theta)}(U(s_i) - U(s_j))^2 + \log \gamma_u(s_i, s_j; \theta)$$

The estimating function corresponding to this is given by:

$$\sum_{i=1}^{n-1} \sum_{j>i} \frac{\frac{d}{d\theta}\gamma_u(s_i, s_j; \theta)}{\gamma_u(s_i, s_j; \theta)} \left\{ \frac{(U(s_i) - U(s_j))^2}{2\gamma_u(s_i, s_j; \theta)} - 1 \right\} \tag{3}$$

Notice that this estimating function corresponds very closely to the weighted least squares method (Cressie 1991, page 96, equation 2.6.12).

Continuing with the theme of composite likelihoods, observe that there are many other composite likelihoods that can be considered also. For example, we may consider two contrasts at a time to get

$$CL_2(\theta, v) = \prod_{i,j} \prod_{l,m} f(v_{ij}, v_{lm}; \theta)$$

How this affects statistical efficiency is a question of interest. In practice, one will have to achieve a balance between statistical efficiency and computational efficiency. We will discuss comparison of statistical efficiency of estimating functions in the next section.

4 Efficiency comparisons

Having defined various estimating functions, the next natural question that arises is: which estimating function should be used when? In the following, we will define the optimality criterion used by Godambe (1960), now known as Godambe's criterion. We then use it to choose an estimating function for semivariogram estimation. We will illustrate its use in a single parameter case.

Godambe's optimality criterion

We will not provide all the regularity conditions explicitly here. The details can be found in Godambe (1960).

Let Θ denote the parameter space. Let G be the class of all zero unbiased estimating functions, that is, if $g \in G$, then $E_\theta(g(U,\theta)) = 0$ for all $\theta \in \Theta$. Let us also assume that $g \in G$ are differentiable and all the relevant expectations exist. Then information content in g regarding the parameter θ is given by

$$Infn(g;\theta) = \frac{\{E_\theta(\frac{d}{d\theta}g(U,\theta))\}^2}{E_\theta(g^2(U,\theta))}$$

Given two zero unbiased estimating functions g_1 and g_2, it is now easy to compare their performance. One can plot $Infn(g_1;\theta)$ and $Infn(g_2;\theta)$ as a function of θ and choose that estimating function which is uniformly better. However, in most practical situations, there may be a certain part of the parameter space where g_1 may be better than g_2; and on the other part g_2 may be better than g_1. In this situation, researcher will have to consider his prior opinion about the most likely parameter value for the data at hand and choose the relevant estimating function. One can also use other approaches such as minimax or non-informative priors or proper priors to calculate 'average information' to select an estimating function in these situations. Appropriateness of these approaches is a foundational issue and will not be discussed here.

Comparison of composite likelihood and the classical method

In the following we discuss the information comparison for the composite likelihood based estimating function and the classical method of estimation. The details of the model are as follows.

We consider the ordinary kriging with exponential semivariogram. That is:

$$E(U(s)) = \mu$$

$$Var(U(s)) = 1$$

$$Cov(U(s_i), U(s_j)) = \theta^{d(s_i,s_j)}$$

We assume the underlying probabilistic model to be Gaussian.

Tables 1, 2 show the information comparisons at various values of θ for 4 x 4 and 8 x 8 regular grids (increasing domain asymptotics) and table 3 shows the information comparison for the 8 x 8 grid nested inside a 4 x 4 grid (infill asymptotics). It is quite clear that the composite likelihood estimating functions are substantially better, an increase in efficiency of about 50%, than the least squares approach.

In table 4 we compare the composite likelihood for the simple kriging with the least squares approach. Recall that, in the case of simple kriging, composite likelihood uses both the conditional mean structure and the conditional variance to obtain estimating functions. From table 4 it is clear that the efficiency gains are substantial, to the order of 75%. Multiparameter extensions of this are straightforward and are not considered here.

Recently Lele and Curriero (1997) have shown that the predictive performance of composite likelihood based estimation of variogram is comparable with the traditional approach.

5 Further extensions

In this section we will discuss various extensions of the use of composite likelihood.

Geometric anisotropy and non-euclidean distances

Curriero (1996) introduces the use of noneuclidean distances in variogram modelling. Lele and Curriero (1997) extend the use of composite likelihood approach to automatically estimate geometric anisotropy in the data.

Universal kriging and Intrinsic Random Function kriging

In practice, it is seldom the case that the mean is known. If the mean is constant, we saw how composite likelihood uses the marginal distribution of contrasts $U(s_i) - U(s_j)$, which is independent of μ, to obtain the semivariogram parameters. This was shown to be related to the weighted least squares approach. Now suppose that the mean structure is given by

$$E(U(s)) = X(s)\beta$$

where $X(s)$ is a vector of known covariates such as elevation, rock type etc. The vector parameter β is a nuisance parameter. It is well known (Cressie, 1991; page 153) that knowledge of β is inessential for kriging predictor as long as the semivariogram is known completely. Unfortunately that is seldom the case. The semivariogram is estimated based on the residual obtained by first estimating β using least squares approach. This estimate is also known to be biased (Cressie, 1991; page 167). One approach that overcomes this bias is REML. We consider contrasts of the observations such that each contrast has mean 0.

Suppose

$$U \sim N(X\beta, C(\theta))$$

Construct a matrix A such that $AA^T = I - X(X^TX)^{-1}X^T$ and $A^TA = I$. Then:

$$V = AU \sim N(0, AC(\theta)A^T)$$

Thus the distribution of V is independent of β. The likelihood for V can be written as:

$$L(\theta, v) = f(v_1, v_2, ..., v_{n-p}; \theta)$$

where p is the number of covariates X. Utilizing ideas of composite likelihood, we can approximate the above likelihood in several ways. Since the marginal distribution of v_i depends on θ, the simplest possibility is:

$$CL_2(\theta, v) = \prod_{i=1}^{n-p} f(v_i; \theta)$$

Following the discussion of the corresponding composite likelihood for ordinary kriging, it can be seen that this essentially generalizes the weighted least squares approach for semivariogram estimation from ordinary kriging to universal kriging but without the need of estimation of the trend parameters and hence avoiding the bias considerations. Our initial studies show that $CL_2(\theta, v)$ depends weakly on θ and hence is not very informative. The bivariate composite likelihood described below, however, seems to be fairly informative.

As before, we could consider pairs of v_i's instead of singletons, to obtain

$$CL_3(\theta, v) = \prod_{i=1}^{n-p} \prod_{j>i} f(v_i, v_j; \theta)$$

One would expect a substantial gain in efficiency parallel to the gains reported in table 4.

Intrinsic random function kriging is very similar to universal kriging in spirit. See Cressie (1991, pages 299-309) for detailed description. The usual method of estimation of the generalized covariance function is based on REML. It is clear that composite likelihood should be applicable to the problem of estimation of the generalized covariance functions.

Nonparametric semivariogram estimation

Practitioners are reluctant to specify a particular model for semivariogram. Recently there have been several papers (Shapiro and Botha, 1991; Cherry, 1995; Lele, 1995) proposing methods for nonparametric semivariogram estimation. It is known (Schoenberg, 1938) that the class of all semivariograms corresponds to a mixture of some kernel semivariogram. The methodology of composite likelihood is easily applicable in such a situation. Moreover it is possible to fit such models using composite likelihood even in the case of universal kriging or intrinsic random function kriging.

Robustness issues

A legitimate concern of the practitioners is that these estimating procedures may not be robust against outliers. Lindsay (1994) discusses the issue of robustness versus efficiency in terms of estimating functions. The weighted estimating functions considered by Lindsay (1994) have a form very similar to the estimating functions obtained from composite likelihoods. Consider equation 3. Let

$$\delta(u(s_i), u(s_j)) = \frac{(U(s_i) - U(s_j))^2}{2\gamma_u(s_i, s_j; \theta)} - 1$$

Then a modified version of the above estimating function may be written as

$$\sum_{i=1}^{n} \sum_{j>i} \frac{\frac{d}{d\theta}\gamma_u(s_i, s_j; \theta)}{\gamma_u(s_i, s_j; \theta)} A(\delta(u(s_i), u(s_j)) = 0$$

If we take $A(\delta) = \delta$,we recover the original estimating function. If we take $A(\delta) = \frac{(1+\delta)^{\lambda+1}-1}{\lambda+1}$, we recover power weighted divergence family described by Cressie and Read(1984). Different values of λ lead to different robustness properties. These robust equations extend easily to the universal kriging case.

6 Summary

This paper proposes the method of composite likelihood for the estimation of semivariogram parameters. Several advantages of this method are outlined.

1. This method eliminates the need for the specification of the bin width and range parameters, making it automatic and objective, at the same time retains the model robustness of the classical approach.

2. Composite likelihood estimating functions have a very intuitive justification of minimizing the prediction error. The ultimate use of variograms is for prediction. It makes sense to estimate the variograms in such a manner that the prediction error is minimized.

3. The efficiency gains obtained by this method are shown to be substantial.

4. The flexibility of this method can lead to better estimation in the case of universal and intrinsic random function kriging. This flexibility also allows for estimation of mixtures of semivariograms in the presence of trend. Robustness to outliers also can be achieved fairly easily.

7 Acknowledgments

This work was partially supported by a grant from DOE (DE-FC07-94, Professor Daniel Goodman PI). I would like to thank Professor Goodman for his generous support, encouragement and many insightful comments.

References

Besag, J. E. (1975) Statistical analysis of non-lattice data. *The Statistician*, 24, 179-195.

Cherry, S. (1995) *Nonparametric estimation of variogram.* Doctoral dissertation submitted to the Department of Mathematical Sciences, Montana State University, Bozeman, Montana.

Cox, D. R. (1975) Partial likelihood. *Biometrika*,62, 269-276.

Cressie, N.A.C. (1991) *Statistics for spatial data.* John Wiley and Sons, NY.

Cressie, N.A.C. and T.R.C. Read (1984) Multinomial goodness of fit tests. *J. Roy. Statist. Soc. Ser. B* 46, 440-464.

Curriero, F. (1996) *The use of non-euclidean distances in geostatistics.* Doctoral dissertation, Department of Statistics, Kansas State University, U.S.A.

Godambe, V. P. (1960) An optimum property of regular maximum likelihood estimation. *Ann. Math.Stat.*, 31, 1208-1212.

Godambe, V. P. (ed.) (1991) *Estimating Functions.* Oxford University Press, NY.

Godambe V. P. and Kale, B. K. (1991) Estimating functions: An overview. In *Estimating Functions*, Godambe V. P. eds. Oxford University Press, NY.

Godambe V. P. and M. Thompson (1989) An extension of quasi-likelihood estimation (with discussion). *J. Stat. Plan. and Inf.* 22, 137-172.

Journel A. G. and C. J. Huijbregt (1978) *Mining Geostatistics.* Academic Press, London.

Lele, S. (1995) Inner product matrices, kriging and nonparametric estimation of variogram. *Math. Geology* 27(5), 673-692.

Lele, S. and F. Curriero(1997) A composite likelihood method for semivariogram estimation. (submitted manuscript)

Lindsay, B. G. (1988) Composite likelihood methods. *Contemporary Mathematics*, 80, 221-239.

Lindsay, B. G. (1994) Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Statist.* 22, 1081-1114.

Marcotte, Denis (1995) Generalized Cross-Validation for Covariance Model Selection. *Math. Geology* 27(5), 659-672.

Schoenberg, I.J. (1938) Metric spaces and completely monotone functions. *Ann. Mathematics* 39(4), 811-841

Shapiro A. and J.D. Botha (1991) Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis* 11, 87-96.

Table 1: Efficiency comparison between the classical least squares method and the composite likelihood based method for a 4 x 4 grid. The second two columns are Monte -Carlo estimates of the information in the estimating functions and the third column is the ratio of the informations or the efficiency gain.

Parameter	Least squares	Composite likelihood	Efficiency gain
0.1	4.44	4.49	1.01
0.2	4.55	5.39	1.18
0.3	5.75	7.73	1.34
0.4	6.65	9.20	1.38
0.5	7.79	11.83	1.52
0.6	12.70	18.99	1.50
0.7	24.03	37.41	1.56
0.8	53.74	81.73	1.52
0.9	223.265	342.31	1.53

Table 2: Efficiency comparison between the classical least squares method and the composite likelihood based method for an 8 x 8 grid (increasing domain). The second two columns are Monte-Carlo estimates of the information in the estimating functions and the third column is the ratio of the informations or the efficiency gain.

Parameter	Least squares	Composite likelihood	Efficiency gain
0.1	14.63	15.81	1.08
0.2	12.75	16.11	1.26
0.3	11.42	16.85	1.48
0.4	10.94	18.37	1.68
0.5	14.12	25.95	1.84
0.6	15.10	30.14	2.00
0.7	24.43	48.80	2.00
0.8	50.76	101.32	1.99
0.9	212.052	386.67	1.82

Table 3: Efficiency comparison between the classical least squares method and the composite likelihood based method for an 8 x 8 grid at a distance 0.5 (infill asymptotics). The second two columns are Monte-Carlo estimates of the information in the estimating functions and the third column is the ratio of the informations or the efficiency gain.

Parameter	Least squares	Composite likelihood	Efficiency gain
0.1	29.83	43.79	1.47
0.2	15.94	27.80	1.74
0.3	11.36	21.33	1.88
0.4	10.42	20.94	2.00
0.5	12.35	25.23	2.04
0.6	15.59	28.48	1.83
0.7	28.14	54.22	1.93
0.8	42.91	83.69	1.95
0.9	253.018	412.54	1.63

Table 4: Efficiency comparison for a 4 x 4 grid in the case of simple kriging. Here composite likelihood corresponds to a combination of linear and quadratic estimating functions. The second two columns are Monte-Carlo estimates of the information in the estimating functions and the third column is the ratio of the informations or the efficiency gain.

Parameter	Least squares	Composite likelihood	Efficiency gain
0.1	4.40	23.84	5.42
0.2	4.34	15.64	3.60
0.3	5.46	16.29	2.98
0.4	6.12	12.60	2.06
0.5	8.49	15.01	1.77
0.6	11.71	22.61	1.93
0.7	21.61	34.73	1.61
0.8	53.364	79.22	1.48
0.9	220.301	322.51	1.46