VARIOGRAM ESTIMATION IN A BAYESIAN FRAMEWORK

To be presented at Fifth International Geostatistics Congress, 22-27 September 1996, Wollongong, Australia

P. F. MOSTAD, T. EGELAND, N. L. HJORT[2],
A. G. KRAGGERUD AND P. Y. BIVER[3]

Norwegian Computing Center
P.O. Box 114 Blindern, N-0314 Oslo, Norway

[2]*Department of Mathematics, University of Oslo*
P.O. Box 1053, Blindern, N-0316 Oslo, Norway

AND

[3]*Elf Aquitaine Production*
Avenue Larribau, 64018 Pau Cedex, France

Abstract. Variograms are fundamental in the application of random fields, but they are frequently difficult to estimate from data. This is particularily true when the assumption of isotropy is dropped. In such cases, we propose to obtain the estimate using Bayesian statistics, where prior knowledge may be used in a structured way.

In this article, several approaches are presented and tested on a simple model. We also discuss extensions to more complex models, as well as a case study.

1. Introduction

In most applications of random fields there is a need to estimate the variogram, i.e., the function that describes the spatial covariance between different locations. When few observations are available, as is frequently the case in petroleum applications, variogram estimation becomes quite difficult. In practice one may have to combine the observations with prior knowledge to obtain an estimate. In this article, we discuss several ways to make such a combination in a rigorous manner using a Bayesian framework.

It is particularily difficult to estimate variograms when the assumption of isotropy is dropped. To illustrate how our methods may help here, we will focus on a class of parametric variograms incorporating geometric anisotropy, i.e., anisotropy resulting from combining an isotropic covariance function with a linear transformation of the underlying space.

Section 2 discusses some traditional approaches, in particular, maximum likelihood estimation. A simple example is presented to illustrate the methods. Section 3 explains the Bayesian approaches and applies them to the example. Section 4 then extends the previous ideas to more complex and versatile models, and contains a discussion of some of the problems and possibilities involved. It also discusses an application to real data. Finally, Section 5 draws some conclusions and points to further work.

2. Traditional estimates

There is a classical non-parametric estimator for variograms based on measuring covariance between pairs of observations. It may be turned into a parametric estimator by fitting a parametric variogram to the nonparametric estimate. Such estimators may yield good answers in many cases, but may be suspect when there is little data, and when one is trying to estimate anisotropy. We will not consider these methods here, but for more information the reader is referred to Cressie (1991).

2.1. MAXIMUM LIKELIHOOD ESTIMATOR

A standard approach to estimation is to specify a parametric model and estimate its parameters using the maximum likelihood principle. For the pur-

Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo, Norway, Tel.: (+47) 22 85 25 00

pose of an example in this article, we use a model with a two-dimensional Gaussian field with expectation 0 and variance 1. If $Z_d = (z_1, \ldots, z_n)$ are the values at points $x_1, \ldots, x_n$, they follow a multinormal distribution $Z_d \sim \mathcal{N}(0, K)$ with covariance matrix K defined by $K_{ij} = [k(x_i - x_j)]$. To incorporate the possibility of geometric anisotropy, we introduce the following covariance function:

$$k(x) = \exp\left(-\frac{3}{\gamma_1}\left\|\begin{bmatrix} \gamma_2^{-\frac{1}{2}} & 0 \\ 0 & \gamma_2^{\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \cos\gamma_3 & \sin\gamma_3 \\ -\sin\gamma_3 & \cos\gamma_3 \end{bmatrix} x\right\|\right). \qquad (1)$$

This is a way to parametrize geometric anisotropy. Any exponential covariance function with geometric anisotropy may be brought to the form above by using a singular value decomposition of the linear transformation of the underlying space. Analogue parametrizations exist in higher dimensions. The advantage with the form above is that the parameters γ_1, γ_2, and γ_3 are easy to interprete. γ_1 corresponds to the geometric average of the range, γ_2 to the anisotropy degree, (the length of the major axis divided by the length of the minor axis), while γ_3 corresponds to the direction of the major axis. We may illustrate a particular choice of γs by plotting an ellipse outlining the set of points whose covariance with the center is $e^{-3} \approx 0.05$ or more (see Figure 1). In the parametrization above, we assume that $\gamma_1 > 0$, $\gamma_2 \geq 1$, and $0 \leq \gamma_3 < 180$ (degrees).

To find the maximum likelihood (ML) estimate for the γs in our simple example, we must determine numerically the γs maximizing $-\log|K| - Z_d K^{-1} Z_d'$, where K is the function of the γs indicated above.

2.2. SIMULATED EXAMPLE

To illustrate the methods discussed in this article, we have simulated some test data at 40 points located (by simulating from a homogeneous Poisson distribution) in a disc of radius 2500 m. They appear as points in Figure 1, and may represent wells in an oil field. Data values have been simulated (using Splus) at each point using a Gaussian field with expectation 0, variance 1, and an exponential variogram with average range $\gamma_1 = 1000$, anisotropy degree $\gamma_2 = 4$, and angle of anisotropy $\gamma_3 = 60$ degrees. That is, the major axis of anisotropy has an angle of 30 degrees with the y-axis and 60 degrees with the x-axis, and the correlation range is 2000 along this axis, while the correlation length along the perpendicular axis is 500 (Figure 1). The data could represent average porosities, or any other measurement in the wells, suitably transformed.

Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo, Norway, Tel.: (+47) 22 85 25 00

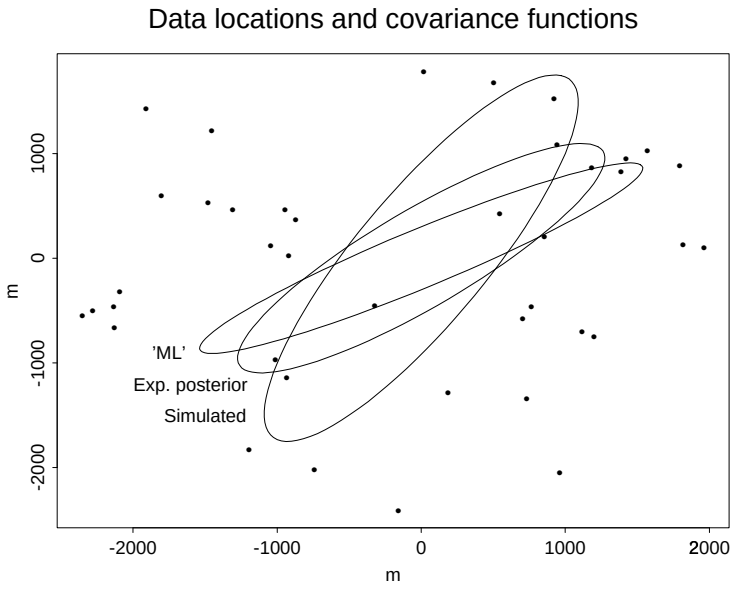

Figure 1. The figure shows 50 points simulated with a homogeneous Poisson distribution within the disc with radius 2500 around the origin. The ellipse marked "Simulated" illustrates the covariance function used for simulating values at these points. The other ellipses represent the estimates derived from the ML method, and by taking the expectation of the posterior, using prior number 2.

We applied the ML method to these data, using the "ms" optimization method of Splus (Venables & Ripley 1994) and varying the starting point. This numerical optimization found a maximum when $\gamma_1 \to \infty$, $\gamma_2 \to \infty$, and $\gamma_3 = 65$, whereas a more sensible local maximum was found at $(\gamma_1, \gamma_2, \gamma_3) = (680, 6.75, 29.9)$. These last numbers appear in Table 1, and are indicated in Figure 1 by the 'ML' ellipse.

Such behaviour seems to be fairly typical in the above model, although of course some examples seem to yield sensible ML estimates, especially when a lot of data points are used. The problem with multimodality is well known (Mardia & Watkins 1989), in particular when a spherical variogram is used instead of the exponential variogram above. The additional problems of finding a maximum at all indicate that ML methods may not be very well suited for this model. In fact, the theory that supports the ML method in most non-spatial models does not go through for the present type of spatial statistics models, so there is no a priori guarantee that the ML should produce better results than competing methods. The theory of maximum quasi-likelihood (Hjort & Omre 1994) may have better theoretical properties. Results in Kraggerud (1996), and Mostad, Egeland & Hjort

Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo, Norway, Tel.: (+47) 22 85 25 00

	γ_1	γ_2	γ_3	Max. axis	Min. axis
True values	1000	4	60	2000	500
Maximum likelihood	(680)	(6.75)	(29.9)	(1768)	(262)
Max. posterior (prior 2)	719	3.99	30.2	1435	360
Max. posterior (prior 3)	627	3.33	30.8	1145	343
Exp. posterior (prior 1)	787	6.98	37.3	2079	298
Exp. posterior (prior 2)	827	3.87	40.0	1627	420
Exp. posterior (prior 3)	772	3.90	39.8	1525	391

TABLE 1. The results for the different methods. The two rightmost columns show the maximal and minimal ranges, computed from γ_1 and γ_2 in each case; they correspond to the major and minor axes of the ellipses that describe the covariance function.

(1994) indicate that the estimates obtained from this method are similar to those from the ML method, while maximum quasi-likelihood is much faster to compute.

3. Bayesian methods

When the ML method does not give reliable answers, we may try to combine it with prior information using a Bayesian framework. In our simple example above, we can choose a prior distribution $f(\gamma_1, \gamma_2, \gamma_3)$, and then compute the posterior distribution using Bayes theorem. We have chosen three different priors, to illustrate how the choice of prior influences the results. The first prior contains only the information that γ_1 and γ_2 are not too extreme; i.e., we assume we know that $100 \leq \gamma_1 \leq 5000$ and $\gamma_2 < 20$. The second prior assumes we have quite good prior information, setting $\gamma_1 \sim \mathcal{N}(1000, 500)$ truncated at 0 (i.e., γ_1 has a distribution that is proportional to the positive part of the normal distribution with expectation 1000 and standard deviation 500), and setting γ_2 to follow a Gamma distribution with expectation 4 and standard deviation 2, truncated at 1. See Figure 2 for an illustration of the priors. The third prior is based on vaguer, and slightly wrong information: We set $\gamma_1 \sim \mathcal{N}(1500, 1500)$ (truncated at 0), and we let γ_2 have a Gamma distribution with expectation 3 and standard deviation 6, truncated at 1. Note that we have not included any prior information about the direction of anisotropy γ_3, and that we have chosen a particular form of the priors where γ_1 and γ_2 are independent. Obviously,

one may instead choose to have independent priors on the maximum and minimum axes of the anisotropy, or indeed choose any other form of prior best reflecting actual prior knowledge.

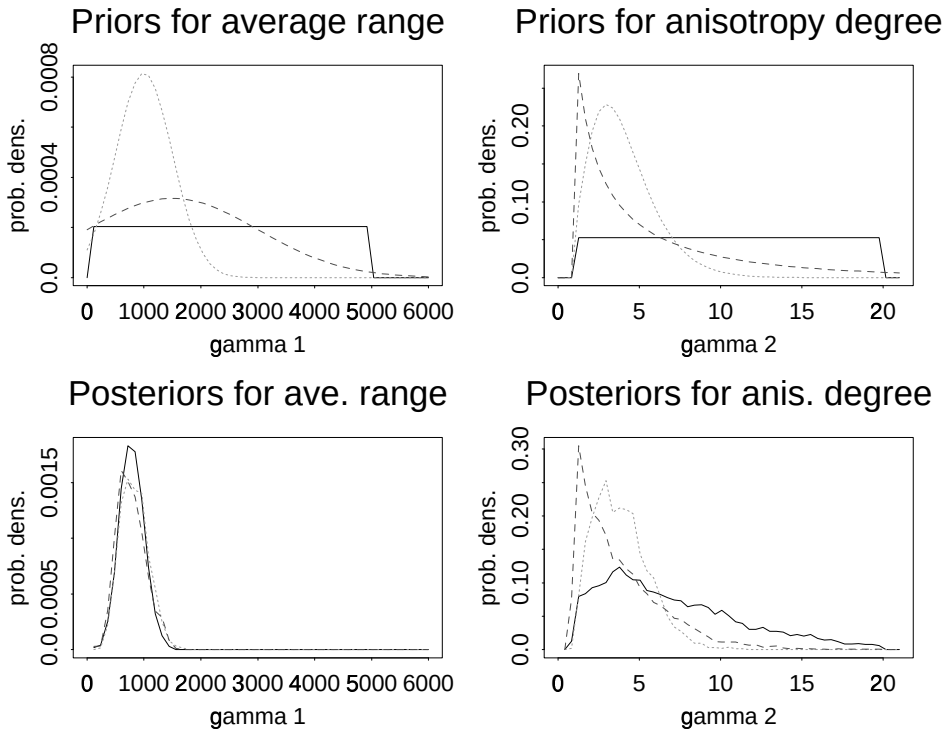

Figure 2. Marginal probability distributions for the average range (γ_1) and the anisotropy degree (γ_2). The fully drawn curves correspond to prior 1, the curves with smallest dots to prior 2, and the curves with the longest dots to prior 3. The posterior distributions in the lower figure are obtained using the Metropolis simulation algorithm.

Using the Bayesian paradigm, our estimates for the γs should now be based on the posterior distribution. In our simple example, this distribution is given, up to a constant, by

$$|K|^{-\frac{1}{2}} \exp(-\frac{1}{2} Z_d K^{-1} Z_d') f(\gamma_1, \gamma_2, \gamma_3). \tag{2}$$

As before, K is the covariance matrix defined by $K_{ij} = [k(x_i - x_j)]$ (see Equation 1), and $Z_d = (z_1, \ldots, z_n)$ is the data vector.

We have studied two alternatives for the actual estimate of the γs: The maximum of the posterior distribution and its expectation. The maximum must be found by numerical maximization of the function above. We used

the same optimization procedure as for the maximum likelihood. As we see, the only difference compared to the ML method is the factor $f(\gamma_1, \gamma_2, \gamma_3)$, and so this approach suffers from some of the same problems. With prior 1, the maximum seems to be reached when $\gamma_2 = 20$, but of course the same local maximum as the ML case appears. For prior 2 and prior 3, the results appear in Table 1. We see that the results are reasonable. In fact, this method generally gives reasonable estimates in simulated cases when the priors exclude extremely wrong answers. In practical use, the quality of the result depends on the quality of the prior, which again depends on how much the user of the method knows about for example the geology being described.

As the variance of the ML estimate seems to be quite large in this model, the maximum posterior estimate inherits this problem (unless the chosen priors are narrow). It can also inherit the multimodality problem from the ML method.

The second approach is to use the expectation of the posterior distribution as the estimate. We have studied two ways of computing this expectation: Numerical integration and simulation with the Metropolis algorithm. Only the latter approach is presented here.

The Metropolis algorithm (Ripley 1987) is based on computing the quotient of the values of the probability density function at different points. The algorithm simulates from the distribution by using a transition function Q specifying the probability of choosing a new vector $(\gamma_1', \gamma_2', \gamma_3')$, given a current vector $(\gamma_1, \gamma_2, \gamma_3)$. The choice is then kept or rejected based on the quotient above. The Q function must be symmetric with respect to switching $(\gamma_1, \gamma_2, \gamma_3)$ and $(\gamma_1', \gamma_2', \gamma_3')$, and should be chosen to optimize the convergence speed and acceptance rate of the algorithm. We have used a transition function Q which independently updates the lengths of the major and minor axes and the anisotropy angle, using normal distributions. (A standard deviation of 50 was used to update the lengths of the axes, while a standard deviation of 0.1 radians was used for the angle.) We should mention how we have chosen to compute the "mean" of the periodic variable γ_3: If $\gamma_{31}, \gamma_{32}, \ldots, \gamma_{3m} \in [0, 180)$ is a sample, then we define the mean as

$$\frac{1}{2}\left(\arctan\frac{\sum_{i=1}^{m}\sin 2\gamma_{3i}}{\sum_{i=1}^{m}\cos 2\gamma_{3i}} + 180^\circ \cdot I\left[\sum_{i=1}^{m}\cos 2\gamma_{3i} > 0\right]\right), \qquad (3)$$

where I denotes the indicator function. (As $2\gamma_3$ has period 2π, we may construct unit vectors with angles $2\gamma_{3i}$ with the x-axis and let the angle of their vector sum represent the "mean".)

The Metropolis method was applied to our example, with 10,000 triplets of γs simulated for each prior, and the first 1000 discarded. The resulting averages are shown in Table 1 and Figure 1. The estimates are reasonably good, although there are still sizable errors. In general, the method seems to give reasonable estimates. It avoids the perils of optimizing a possibly multimodal function, but it is time consuming to compute to a reasonable degree of accuracy. It also yields a sample from the entire posterior distribution, which may be useful. The marginal distributions of this sample for our case-study example are plotted in Figures 2 and 3. Note, in particular, the information one obtains about the anisotropy direction (γ_3) without using any prior information about direction.

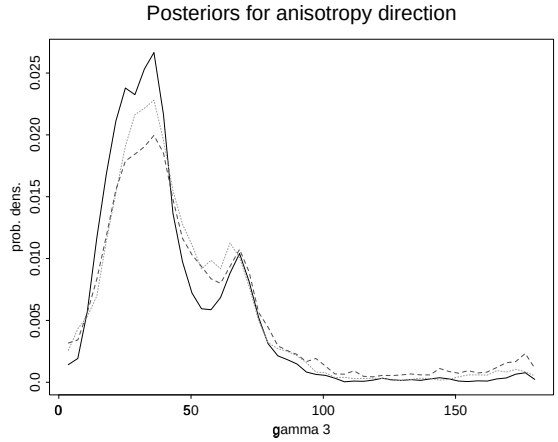

Figure 3. The marginal posterior distributions for γ_3, the direction of anisotropy, simulated with the Metropolis algorithm. The curve with the highest maximum corresponds to prior 1, the second highest to prior 2, and the lowest to prior 3. Note the local maximum around 60 degrees.

4. Extensions to more complex models

Much of the above may be fairly straightforwardly extended to more complex models. First of all, the mean of the Gaussian field, and the variance σ^2, may be unknown instead of fixed. We may also use trend functions $f_j(x)$ $(j = 1, \ldots, s)$ for the field, so that the total trend for the field at a point x may be expressed as a vector product $[f_1(x), \ldots, f_s(x)]\beta$, where β is a vector of weights. The vector of observations Z_d now has distribution $Z_d \sim \mathcal{N}(F\beta, \sigma^2 K)$, where $F_{ij} = [f_j(x_i)]$, and K is defined as before.

There are well-known formulas for the maximum likelihood estimates for β and σ^2 in this case, and from them, we get a formula in the γ parameters which must be numerically optimized in terms of these to find their ML estimates (see Mostad & Egeland (1995)). In the Bayesian framework, it is natural to use priors for β and σ^2; we use priors of the form $\beta \mid \sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 T)$ and $\sigma^2 \sim \mathcal{IG}(a, b)$ to facilitate computations. (Here, $\mathcal{IG}$ denotes the inverse Gamma distribution.) The special form of the prior is of course a limitation, but in practice, prior information will probably often fit into this form. If we use this framework together with a fixed variogram, we get Bayesian Kriging (Hjort & Omre 1994). One may instead use the full model to obtain posterior predictions for values at unsampled locations directly (Gaudard, Karson, Linder & Sinha 1995). Comparing this approach to one where we first estimate a variogram and then use this variogram to make a prediction by e.g. Kriging, we see that we implicitly use the entire posterior distribution for the variogram, instead of just a fixed estimate.

However, one may prefer to fix a variogram before continuing on with other computations. In Kraggerud (1996), the above model is studied as the basis for Bayesian variogram estimation. The added unknown parameters seem to increase the variance of the estimates, thus making the inclusion of prior information even more pertinent. Among other things, there seems to be a bias in the ML estimate for γ_2, the degree of anisotropy. This should also be clear from theoretical arguments. However, the amount of bias is hard to compute. One possibility is to fix the positions of data points (the "well locations"), and simulate many data sets within the used model. By performing ML estimates of the parameters for each simulated data set, it should be possible to estimate the bias function numerically. However, such computations are likely to be very time consuming.

The above model may of course also be extended further, for example, into more dimensions, or to other types of covariance functions. But adding more parameters is likely to increase the difficulty of estimation. It is also possible to apply the ideas above to data which are not normally distributed in the marginal by applying an appropriate transformation.

4.1. APPLICATION TO POROSITY DATA

The case study was based on porosity measurements from 43 wells ranging from 0.031 to 0.140 with a mean 0.078 and standard deviation 0.028. For each well, we used the average measured porosity. The data set has been

used on several previous occasions for various purposes, and Mostad & Egeland (1995) provide details beyond what space permits here. Presently, the aim is to interpolate the porosity surface as accurately as possible. In addition, we would like to estimate a model reflecting the uncertainty appropriately in some sense. Different models are compared based on leave–one–out predictions. An initial model based on a spherical variogram with ranges 1000 and 3000 in the x and y directions performs poorly – the predictions and the observations are slightly negatively correlated! The situation is improved by estimating the parameters and changing to an exponential variogram. Ranges 1000 and 3000 were retained as prior guesses while prior standard deviations were set to 750 and 1500. The ranges were estimated as 291 and 3494, indicating that there is less continuity in the x direction and more in the y–direction than assumed *a priori*. Visual inspection revealed a better model. The mentioned correlation increases to 0.26, confirming that an improved model results from variogram estimation.

5. Concluding remarks

A Bayesian framework seems well suited to combine prior knowledge with data observations into a variogram estimate. Both the maximum posterior method and the posterior expectation method give reasonable estimates, although the variances of the estimates could still be large in many applications. The maximum posterior method is faster to compute, but may inherit the multimodality problem from the ML estimate, and it may have larger variance. There seems to be a bias (overestimation) in the estimation of the degree of anisotropy, and it should be interesting to try to measure this bias. Finally, the method of quasi-likelihood represents a different approach to the problems of the ML method, and should be investigated further.

References

Cressie, N. (1991), *Statistics for Spatial Data*, John Wiley & Sons, New York.

Gaudard, M., Karson, M., Linder, E. & Sinha, D. (1995), Bayesian spatilal prediction, Preprint.

Hjort, N. L. & Omre, H. (1994), 'Topics in spatial statistics', *Scand. J. Statist.* **21**(4), 289–357. With discussion.

Kraggerud, A. G. (1996), Parametric variogram estimation in a Bayesian framework, Master thesis (diplomoppgave), University of Trondheim.

Mardia, K. V. & Watkins, A. J. (1989), 'On multimodality of the likelihood in the spatial linear model', *Biometrika* **76**(2), 289–295.

Mostad, P. & Egeland, T. (1995), A program for Bayesian Kriging and variogram estimation, NR-note SAND/14/95, Norwegian Computing Center, P.O.Box 114 Blindern,

N-0314 Oslo, Norway.

Mostad, P., Egeland, T. & Hjort, N. L. (1994), Geostatistical modeling. Bayesian Kriging of facies proportions, NR-note SAND/24/94, Norwegian Computing Center, P.O.Box 114 Blindern, N-0314 Oslo, Norway.

Ripley, B. D. (1987), *Stochastic Simulation*, John Wiley & Sons, New York.

Venables, W. N. & Ripley, B. D. (1994), *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.