MEASURING THE UNCERTAINTY IN KRIGING

MARK S. HANDCOCK *
Department of Statistics & Operations Research
Leonard N. Stern School of Business
New York University
New York, NY 10012-1126 U. S. A.

Spatial phenomena is commonly modelled as a realization from a stochastic process. Even when the reality is unique such models can usefully represent the uncertainty the modeler has about the phenomena. This paper is concerned with predicting for Gaussian random fields in a way that appropriately deals with uncertainty in the covariance function. To this end, we analyze the best linear unbiased prediction procedure (kriging) within a Bayesian framework. Particular attention is paid to the treatment of parameters in the covariance structure and their effect on the quality, both real and perceived, of the prediction. We show how this model can be improved by accounting for the uncertainty in the model parameters.
Key words: Bayesian statistics; Interpolation; Robustness; Spatial Statistics.

1. INTRODUCTION

In this paper we consider a modeling approach for spatially distributed data. As an illustration of the types of problems addressed consider the spatially distributed data in Figure 1 . The data are 52 topological elevations over a small area on the northern side of a hill. The data were measured by a surveying class, using a plane table and alidade. Davis (1973) was interested in the analysis of maps and used the survey to produce contours of the region. An important feature is the small streams running northward down the hill and joining together at the base of the region.

How should we analyze the data if our objective is to predict the elevations within the region surveyed? The perspective taken is that the actual elevations at each possible survey location (i.e. northing and easting from a reference point) taken together are a 'realization' from a particular stochastic process. The general process is described in the next section. Based on the observed data at the 52 locations and this statistical model a prediction of the elevation at unobserved locations can be made. Just as importantly from our perspective, an estimate of the uncertainty of that prediction can be derived from the model.

The statistical model is usually estimated from the very same data from which the predictions are made. The objective of this paper is to assess the effect of the fact that the model is estimated, rather than known, on the prediction and the associated prediction uncertainty. We describe a method for achieving this objective.

For example, suppose we wish to predict the elevation where the streams join at the base of the hill. We actually have an observation there, indicating that the elevation is 705 feet, but we will ignore it at this point except to use it to check our prediction. The commonly used method of maximum likelihood for estimating the best model suggests that the true elevation has a 95% chance of being in the interval (699, 707) feet. This interval only represents the uncertainty in the prediction given this particular estimated model and does not represent the uncertainty in estimating the model itself. When this is accounted for, using the method described in the next sections, the Bayesian 95% prediction interval is (694, 713) feet. The posterior probability content, incorporating the model uncertainty, of the maximum likelihood interval is 73%. Similar inaccuracies are to be excepted when the estimated uncertainties are based on alternative point estimates of the model.
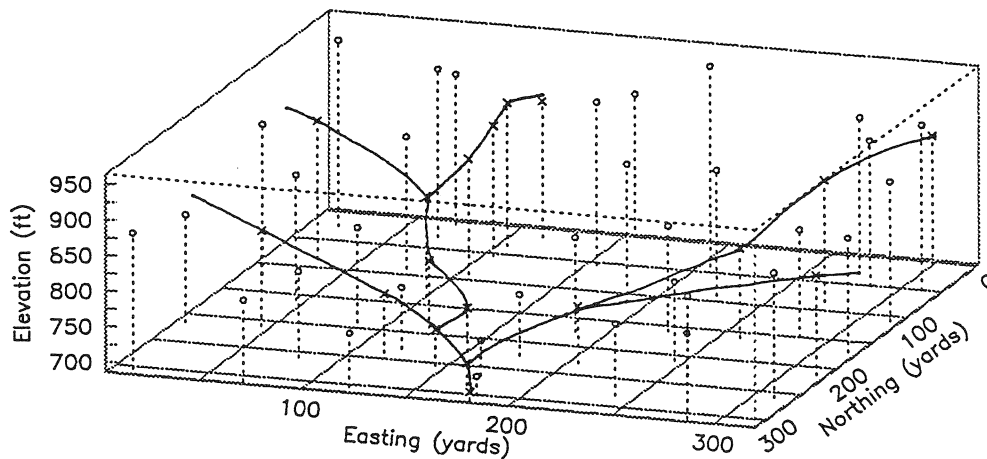

FIGURE 1: Elevations of the north face of a hill from Davis (1973), Table 5.11. The aspect ratio in the figure is spatially correct. The solid lines indicate the small streams running down the hill. Elevations on the streams are indicated by × and those not on the streams are marked by o.

We conclude that in many practical situations this uncertainty has a large impact on the estimated uncertainty of the prediction and a lesser effect on the predicted value itself. If there is little information about the model in the data the approach guards against gross error. In situations where substantial previous knowledge of the phenomena exists, the approach allows the information to be incorporated easily.

Bayesian analyses of kriging procedures are relatively new. Except for the work of Omre (1987), Omre and Halvarsen (1989) and Woodbury (1989) there appears to be no work from within the geostatistical community using the Bayesian perspective. Omre and Halvarsen (1989) describe a Bayesian approach to predicting the depth of geologic horizons based on seismic reflection times. They note the Bayesian interpretation of ordinary kriging and utilize prior information about the mean function only, not accounting for uncertainty in the covariance structure. Of course, the situation is a direct extension of standard Bayesian work in linear models where, for example, Box and Tiao (1973) §2.7 and Zellner (1971) §7 are textbook references. Here we will focus on a parametric representation of the covariance structure as its direct interpretation is of interest. Non-parametric approaches have been developed by Le and

Zidek (1992) and Pilz (1991). An alternative non-bayesian approach to covariance parameter uncertainty is given by Switzer (1984). What is novel about this paper is the spatial setting with irregularly observed locations and the general treatment of parameters in the covariance structure other than location and scale.

The framework of prediction is developed in Section 2. In Section 3 the assumption that the covariance structure is known is relaxed and the Bayesian formulation is developed. The focus is the evaluation of the performance of the traditional 'plug-in' kriging procedure. Section 4 illustrates this evaluation using topographical data from Davis (1973).

2. METHODOLOGY

2·1 Prediction using kriging

In this section we present the traditional kriging procedure as the basis for the later developments. Suppose $Z(x)$ is a real-valued stationary Gaussian random field on R with mean

$$E\{Z(x)\} = f(x)'\beta,$$

where $f(x) = \{f_1(x), \ldots, f_q(x)\}'$ is a known vector-valued function and β is a vector of unknown regression coefficients. Furthermore, the covariance function is represented by

$$\text{cov}\{Z(x), Z(y)\} = \alpha K_\theta(x, y) \qquad \text{for } x, y \in R$$

where $\alpha > 0$ is a scale parameter, $\theta \in \Theta$ is a $p \times 1$ vector of structural parameters and Θ is an open set in $\mathbb{R}^p$. The division is purely formal as θ may also determine aspects of scale. In the general case, we observe $\{Z(x_1), \ldots, Z(x_n)\}' = Z$ and will focus on the prediction of $Z(x_0)$. The kriging predictor is the best linear unbiased predictor of the form $\widehat{Z}_\theta(x_0) = \lambda(\theta)'Z$; that is, the unbiased linear combination of the observations that minimizes the variance of the prediction error. It is straightforward to show that the corresponding weight vector $\lambda(\theta)$ defining $\widehat{Z}_\theta(x_0)$ is given by

$$\lambda(\theta)' = b_\theta'(F'K_\theta^{-1}F)^{-1}F'K_\theta^{-1} + k_\theta'K_\theta^{-1}, \qquad (2.1)$$

where

$$\begin{aligned}
F &= \{f_j(x_i)\}_{n \times q}, \\
k_\theta &= \{K_\theta(x_0, x_i)\}_{n \times 1}, \\
K_\theta &= \{K_\theta(x_i, x_j)\}_{n \times n}, \\
b_\theta &= f(x_0) - F'K_\theta^{-1}k_\theta.
\end{aligned}$$

In the example $x = (x^1, x^2)$ and we can take $f_1(x) = x^1$ and $f_2(x) = x^2$, the northing and easting of the survey locations, respectively. A third component of the mean will be added in Section 4. The covariance function represents the covariance between the elevation at the survey locations $x = (x^1, x^2)$ and $y = (y^1, y^2)$.

2·2 Assessing Uncertainty in kriging

The quality of the prediction is determined by the distribution of the prediction error, $e_\theta(x_0) = Z(x_0) - \hat{Z}_\theta(x_0)$. Note that the prediction weights $\lambda(\theta)'$ do not depend on α or β. Under our Gaussian model, for fixed $\alpha, \beta,$ and θ, the conditional distributions of $Z(x_0)$ and $e_\theta(x_0)$ given Z are

$$Z(x_0) \mid Z \sim N\left(k_\theta' K_\theta^{-1} Z + b_\theta' \beta, \ \alpha\{ K_\theta(x_0, x_0) - k_\theta' K_\theta^{-1} k_\theta \} \right)$$

$$e_\theta(x_0) \mid Z \sim N\left(b_\theta'(\beta - \hat{\beta}(\theta)), \ \alpha\{ K_\theta(x_0, x_0) - k_\theta' K_\theta^{-1} k_\theta \} \right)$$

where $\hat{\beta}(\theta) = (F' K_\theta^{-1} F)^{-1} F' K_\theta^{-1} Z$ and $N(\cdot, \cdot)$ denotes the Gaussian distribution. The sampling (or unconditional) distribution for $e_\theta(x_0)$ is

$$e_\theta(x_0) \sim N\left(0, \ \alpha V_\theta \right) \tag{2.2}$$

where $V_\theta = K_\theta(x_0, x_0) - k_\theta' K_\theta^{-1} k_\theta + b_\theta'(F' K_\theta^{-1} F)^{-1} b_\theta$ and αV_θ is the usual prediction error variance as given in Ripley (1981).

Note that the underlying kriging procedure is motivated by sampling considerations, producing point predictions and associated measures of uncertainty for those predictions both based on sampling distributions unconditional on the observed Z. However kriging, when the mean is of known regression form, can be given a Bayesian interpretation. Traditionally, it is assumed that the covariance function is known exactly and the investigator has little knowledge about β prior to analyzing the data. The underlying kriging approach usually presumes ignorance about β and the unrelatedness of β to the behavior of the covariance function. This latter philosophy will be followed throughout the paper. Under these assumptions, an appropriate prior distribution has $\mathrm{pr}(\beta \mid \alpha, \theta)$ locally uniform. The posterior distribution of β is then

$$\beta \mid \alpha, \theta, Z \sim N_q\left(\hat{\beta}(\theta), \ \alpha(F' K_\theta^{-1} F)^{-1} \right)$$

The posterior distribution of the prediction error is then

$$\mathrm{pr}\big(e_\theta(x_0) \mid \alpha, \theta, Z\big) \ \propto \ \int_\beta \mathrm{pr}(e_\theta(x_0) \mid \alpha, \beta, \theta, Z)\mathrm{pr}(\beta \mid \alpha, \theta, Z)d\beta,$$

which is, by direct calculation,

$$e_\theta(x_0) \mid \alpha, \theta, Z \sim N\left(0, \ \alpha V_\theta \right) \tag{2.3}$$

the same as the sampling distribution (2.2). Similarly we have

$$Z(x_0) \mid \alpha, \theta, Z \sim N\left(\hat{Z}_\theta(x_0), \ \alpha V_\theta \right) \tag{2.4}$$

where $\hat{Z}_\theta(x_0) = k_\theta' K_\theta^{-1} Z + b_\theta' \hat{\beta}(\theta)$ is the usual kriging point predictor. These distributions form the basis for all inferential statements about the prediction and prediction error. Hence, except for the usual differences in interpretation, we end up with the same analysis as the traditional approach. This comparison may be loosely stated as: ordinary kriging is 'Bayesian' with the non-informative prior for the mean parameter.

3. KRIGING WITH UNKNOWN COVARIANCE

In this section, the assumption that the covariance function is known exactly is relaxed to allow the covariance function to be unknown, but still a member of the parametric class Θ.

In traditional kriging, one estimates α and θ by either likelihood methods or various *ad hoc* approaches. The likelihood approach to the estimation of the covariance structure was first applied in the hydrological and geological fields following Kitanidis and Lane (1985) and Hoeksema and Kitanidis (1985). Mardia and Marshall (1984) is a standard reference in the statistical literature. Usually the predictor and the behavior of the prediction error are themselves estimated by 'plugging-in' the estimates into (2.1) and (2.2). If θ is known so that only the location parameter β and the scale parameter α are uncertain then we are in a standard generalized least-squares setting. The distinction between the generalized least squares setting and the random field setting is the uncertainty in the structural parameter θ. While the restriction to a parametric class is a significant assumption, it still allows great latitude.

As β is a location parameter we expect that our prior opinions about β bear no relationship to those about α and *a priori* might expect α and β to be independent, leading to the use of Jeffreys's prior. Partly for convenience, the form of the prior used here will be

$$\mathrm{pr}(\alpha,\ \beta,\theta) \propto \mathrm{pr}(\theta)/\alpha$$

It easily follows from Zellner (1971) that the predictive distribution of $Z(x_0)$ conditional on θ and Z is

$$Z(x_0) \mid \theta, Z \sim t_{n-q}\left(\widehat{Z_\theta}(x_0),\ \frac{n}{n-q}\widehat{\alpha}(\theta)V_\theta \right), \tag{3.1}$$

a shifted t distribution on $n-q$ degrees of freedom.

The marginal posterior distribution of θ can be shown to be

$$\mathrm{pr}(\theta \mid Z) \propto \mathrm{pr}(\theta) \cdot |K_\theta|^{-1/2}|F'K_\theta^{-1}F|^{-1/2}\widehat{\alpha}(\theta)^{-(n-q)/2} \tag{3.2}$$

The Bayesian predictive distribution for $Z(x_0)$ is

$$\mathrm{pr}(Z(x_0) \mid Z) \propto \int_\Theta \mathrm{pr}(Z(x_0) \mid \theta, Z) \cdot \mathrm{pr}(\theta \mid Z)d\theta$$

where the integrand is given by (3.1) and (3.2). As the dependence of K_θ on θ is not specified this expression can not be simplified and further exploration will in general require numerical computation. If prior information is available it may be directly incorporated into (3.2), although additional numerical integration may be necessary if prior dependencies among (α, β, θ) are envisaged.

Suppose we use an estimation procedure to select the parameters $(\tilde{\alpha}, \tilde{\theta})$ of a covariance structure. These may be arrived at by any procedure, although the usual methods are maximum likelihood, weighted least squares or derived from empirical correlation functions. The distribution that an investigator would use as a basis for inference about $Z(x_0)$ would be

$$Z(x_0) \mid \tilde{\alpha}, \tilde{\theta}, Z \sim N\left(\widehat{Z_{\tilde{\theta}}}(x_0),\ \tilde{\alpha}V_{\tilde{\theta}} \right),$$

plugging in $(\tilde{\alpha}, \tilde{\theta})$ for (α, θ) in (2.4).

Depending on the influence of θ on the spread and location of $\mathrm{pr}(Z(x_0) \mid \theta, Z)$, the Bayesian predictive distribution might be wider or narrower than the plug-in predictive distribution. The location of the plug-in predictive distribution may also be quite different from the Bayesian predictive distribution. Typically the Bayesian predictive distribution will have no simple analytic form and must be determined numerically. The difference between the plug-in and Bayesian predictive distributions represents the difference in inference between the traditional kriging approach and the full Bayesian approach.

Note the plug-in prediction error, $\tilde{e}_\theta(x_0) = Z(x_0) - \hat{Z}_{\hat{\theta}}(x_0)$, is just a shifted version of $Z(x_0)$, so that comparisons of performance of the plug-in estimates will be the same whether we consider $Z(x_0)$ or $\tilde{e}_\theta(x_0)$. We could interpret this as a comparison between the plug-in distribution for $\tilde{e}_\theta(x_0)$ and the actual distribution for $\tilde{e}_\theta(x_0)$ under the full Bayesian model, although the latter distribution would not be used for inference.

3·1   The Matérn Class of Covariance Functions

In this section we describe a general class of covariance functions that we feel provides a sound foundation for the parametric modeling of Gaussian random fields. The class is motivated by the smooth nature of the spectral density, the wide range of behaviors covered and the interpretability of the parameters. It will be used throughout the later sections. The properties of the covariance function directly determine the properties of the random field model. The Matérn class is characterized by the parameter $\theta = (\theta_1, \theta_2)$. $\theta_1 > 0$ is a scale parameter controlling the range of correlation. The smoothness parameter $\theta_2 > 0$ directly controls the smoothness of the random field. The Exponential class corresponds to the sub-class with smoothness parameter $\theta_2 = 1/2$, that is

$$K_E(x) = \exp(-x/\theta_1').$$

The sub-class defined by $\theta_2 = 1$ was introduced by Whittle (1954) as a model for two dimensional fields. It is commonly used in hydrology (see e.g. Jones (1989)). As $\theta_2 \to \infty$, $K_\theta(x) \to \exp(-x^2/\theta_1^2)$, often called the 'Gaussian' covariance function. We shall refer to it as the Squared Exponential model. This model forms the upper limit of smoothness in the class and will rarely represent natural phenomena as realizations from it are infinitely differentiable.

The isotropic correlation functions have the general form

$$K_\theta(x) = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)} \cdot \left(\frac{x}{\theta_1'}\right)^{\theta_2} \mathcal{K}_{\theta_2}\left(\frac{x}{\theta_1'}\right)$$

where $\theta_1' = \theta_1/(2\sqrt{\theta_2})$ and $\mathcal{K}_{\theta_2}$ is the modified Bessel function of order θ_2 discussed in Abramowitz and Stegun (1964), §9.

A field with this covariance function is $\lceil\theta_2 - 1$ times (mean-square) differentiable where $\lceil$ is the integer ceiling function. The realizations will have continuous $\lceil\theta_2 - 1$ derivatives if $\theta_2 > \lceil\theta_2 - \frac{1}{2}$. If the field is Gaussian the realizations will have continuous $\lceil\theta_2 - 1$ derivatives (almost certainly) (See Cramèr and Leadbetter (1967), §4.2, §7.3, and §9.2-9.5).

All calculations of $\mathcal{K}_{\theta_2}$ in this work use the RKBESL algorithm from the SPECFUN library available free from NETLIB. A general treatment is given in the seminal work by Matérn (1986).

4. AN ANALYSIS OF DAVIS'S TOPOGRAPHICAL DATA

In this section we analyze the data introduced in Section 1 and originally from Davis (1973). It has been studied by Ripley (1981, pp. 58-72), and subsequently by Warnes (1986), Ripley (1988, pp. 15-21) and Mardia and Watkins (1989). The original data were scaled so that 50 yards in location corresponds to one map unit. We will use the more natural units of yards, although the later references continued the original scaling. The survey locations are recorded to two significant figures and the elevations to three significant figures.

The major assumptions implicit in the model are stationarity of the Gaussian random field, isotropy of the correlations and the correct specification of the mean. These are interdependent so that checking them individually is usually not the best approach. There are available methods to test if the marginal distribution of the observations is Gaussian. However it is difficult to determine if the joint distribution of the observations is Gaussian in the presence of an unknown correlation structure. In particular the marginal distribution of the observations is little guide to the joint distribution. The realizations of the random field can be assumed to be smooth, at least continuous and maybe even differentiable. Given the nature of the data and the measurement procedure it will be assumed that the measurement error is small so that the (observed) field is continuous.

The Exponential model, while providing a reasonable initial covariance class, does not allow the field to have differentiable realizations. Given that a priori the form of the covariance is unknown it is unreasonable to exclude the possibility of smoother random fields.

As indicated in the Section 2, the mean function should clearly include the Northing and Easting of the survey locations. In addition, there is information in the locations of the streams that should be taken into account. One crude way is to include, as $f_3(x)$, the horizontal distance of the survey point to the closest stream.

4·1 Posterior Knowledge Based on a Flat Mean

Initially we will entertain the model with a constant mean. The marginal posterior for the smoothness parameter based on the uniform prior for the smoothness and range parameters is given in Figure 2.

The mass of the distribution is between $\theta_2 = 0.5$ and $\theta_2 = 1.5$. The mode is slightly below $\theta_2 = 1$, which corresponds to Whittle's covariance function (Whittle (1954, 1962)). Interestingly, Whittle regarded this model as the natural extension of the Exponential model, $\theta_2 = 1/2$, from one to two dimensions. It corresponds to a random field with continuous realizations that are on the margin of mean-square differentiability. For $\theta_2 > 1$ the field is mean-square differentiable. It is interesting to note that the ratio of the density at the mode to the density at the Exponential model is about 5 : 1, so that the Exponential appears too rough for this field. Such posterior densities are a useful tool for describing and understanding the behavior of the phenomena underlying the data.

Of course, alternative prior distributions can easily be used. One could express prior knowledge about (β, α) by taking the marginal prior of (β, α) to be the usual Gaussian-Gamma conjugate prior from generalized least squares. An informative prior for θ_2 could deemphasize smoothnesses less than a half or much greater than two, the rationale being that we do not expect the realizations to be discontinuous or much smoother than twice differentiable. Such a prior distribution would have little effect on the predictive distribution as the likelihood places little weight on smoothness values in that range.
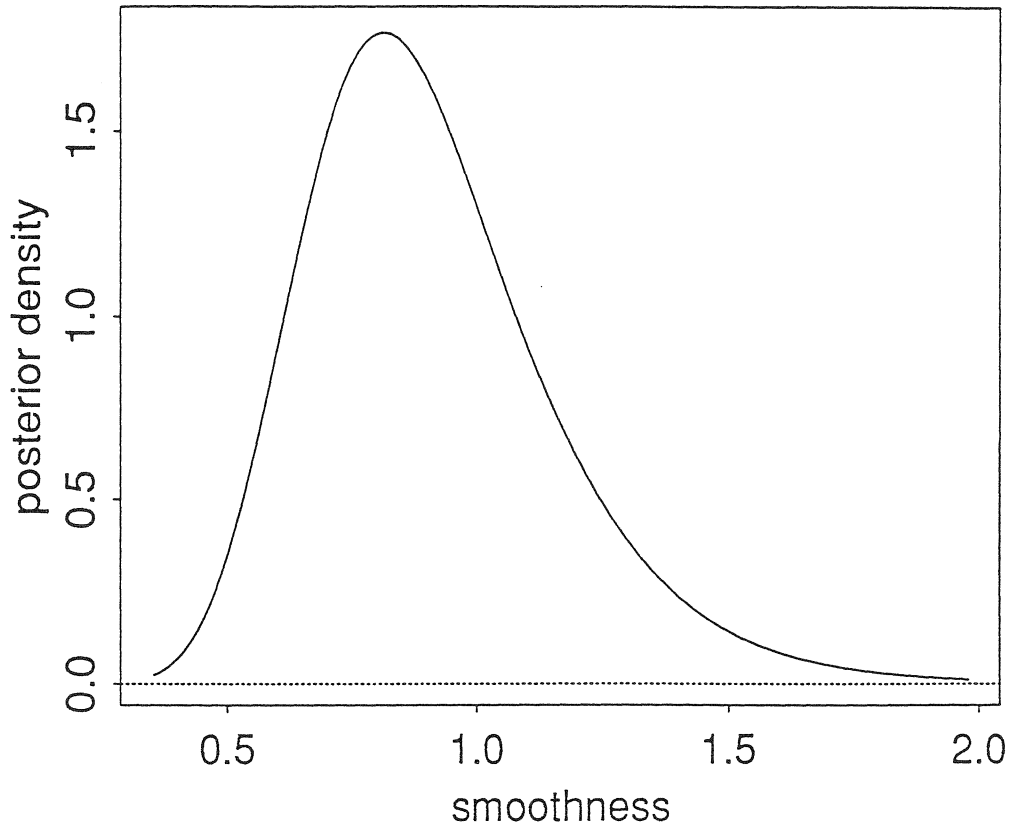
FIGURE 2: Posterior distribution for the smoothness parameter based on the Matérn model with constant mean.

4·2 Incorporating Additional Information in the Mean

The model with a constant mean may be inadequate as compared to the models including the survey locations and distance to streams as regressors because of non-stationarity in the mean (Mardia and Watkins (1989)). The location chosen to be predicted at is the surveyed location closest to the most northern junction of the stream (at (180, 300) on Figure 1). It was chosen to be reasonably close to the other survey locations. The models will be developed without this location and the elevation there will be used as a check on the predictions.

Figure 3 is the profile log-likelihood surface under this model. Figure 4 presents the predictive densities based on the Matérn model with this more sophisticated mean function and a uniform prior on the smoothness and range parameters. The plug-in predictive distribution based on the maximum likelihood estimate, $(\hat{\alpha}, \hat{\theta}) = (955, 68, 7.8)$, is Gaussian centered at 703 feet with a standard deviation of about 2.1 feet. The effect of the additional regressors is to substantially reduce the variability of the predictive distributions.
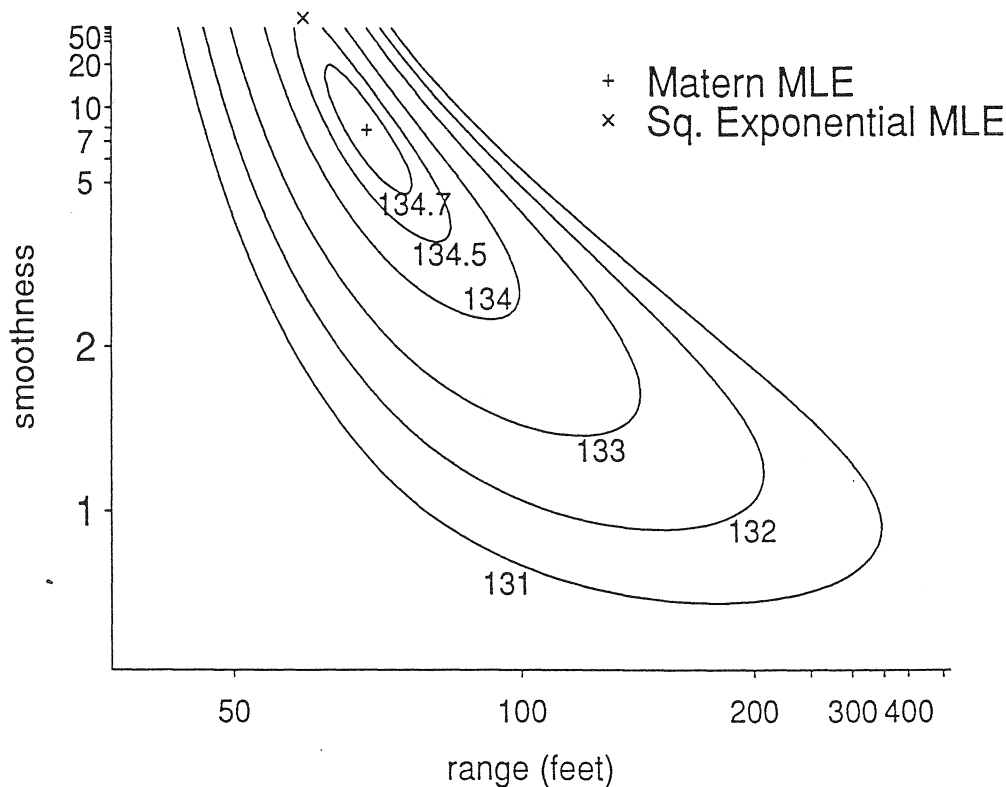
FIGURE 3: Profile log-likelihood for the Matérn model with mean based on the northing and the distance to closest stream.

Also represented are three alternate plug-in predictive distributions based on the maximum likelihood estimators under the Exponential and Squared Exponential classes and the maximum *a posteriori* Matérn value. Probability intervals based on this plug-in predictive distributions will markedly differ from those based on the Bayesian predictive distribution. The latter is a better reflection of the uncertainty in the covariance structure and should be regarded as a superior reference for inference. Smoother estimated models correspond to less perceived uncertainty in the prediction. For example, the Bayesian 95% prediction interval has nominally 99.99% probability under the plug-in predictive distribution for the maximum likelihood Matérn model. Alternatively the nominally 95% interval for this plug-in predictive distribution actually has 73% probability.

4·3 Sensitivity to Prior Specification

How sensitive is our inference to the choice of prior distributions? In these examples a prior distribution uniform on the positive values the smoothness parameter is used. Alternatively one could use the prior

$$\mathrm{pr}(\theta_2) \quad = \quad \frac{1}{(1 + \theta_2)^2}$$

reflecting the belief that larger smoothness values are a *priori* less likely than smaller values. Physically the belief is that the field is more likely to be one or two times differentiable, rather than, say, 101 times. This prior is uniform for $\theta_2/(1 + \theta_2)$ on [0, 1].
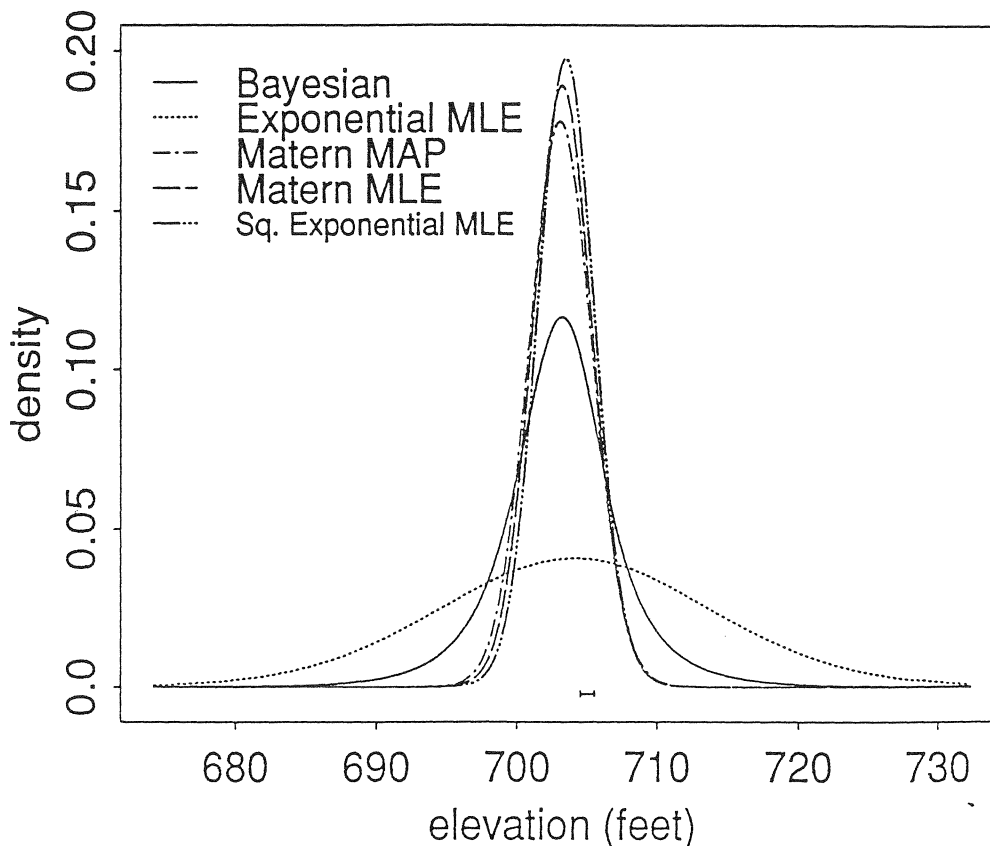
FIGURE 4: Predictive distributions based on the Matérn model with mean using the northing and the distance to closest stream. The observed elevation at this location is represented by the small horizontal bar reflecting the recording accuracy.

The effect of using this prior to deemphasize larger smoothness values relative to the uniform prior is to increase the uncertainty in the prediction, in line with the less smooth models. The inference appears to be insensitive to moderate changes in the prior for θ_2. Based on figures not presented here, we find that using a uniform prior for α instead of the usual $1/\alpha$ results in a predictive distribution with slightly thinner tails and that the predictive distribution is insensitive to changes in the prior for β. While the maximum likelihood estimate is a good representative value, the overall flatness of the likelihoods would suggest against choosing any particular member as the 'truth'. Clearly we need additional information before we can choose between members of the same class. The same comments apply to the choice of mean model. It is tempting to base the decisions on the changes in log-likelihood. It is still an

open question as to the validity of this procedure in the face of the interdependence of the mean and covariance structures.

5. CONCLUSION

The kriging procedure is often described as optimal (Matheron (1965)) because it produces optimal predictions when the covariance structure of the random field is known. If the covariance structure needs to be estimated, then this primary motivation for kriging is in question. It is then necessary to assess the effect of the fact that the model is estimated, rather than known, on the prediction and the associated prediction uncertainty. In this paper we have seen that the Bayesian paradigm provides a framework in which to analyze the performance of the estimated kriging predictor. In conclusion, a better approach would be to base inference on the Bayesian predictive distribution. This approach takes into account the uncertainty about the covariance function expressed in the likelihood surface and ignored by point estimates of the covariance function. It also allows the performance of the usual plug-in predictive distribution based on an estimated covariance structure to be critiqued within a wider framework. The results also suggest that fitting the empirical correlation function 'by eye' may lead to plug-in predictive distributions that differ markedly from the Bayesian predictive distribution. The maximum likelihood estimate may be the best single representative available, but this reduction itself can be detrimental to the inference (See Figure 4).

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1965), Handbook of Mathematical Functions, New York: Dover.

BOX, G. E. P. and TIAO, G. C. (1973), Bayesian Inference in Statistical Analysis, Reading, Mass.: Addison-Wesley.

CRAMÉR, H. and LEADBETTER, M. R. (1967), Stationary and Related Stochastic Processes, New York: Wiley.

DAVIS, J. C. (1973), Statistics and Data Analysis in Geology, New York: Wiley.

HANDCOCK, M. S. (1989), Inference for Spatial Gaussian Random Fields when the Objective is Prediction. Ph.D. thesis, Department of Statistics, University of Chicago, Chicago, Illinois.

HOEKSEMA, R. J. and KITANIDIS, P. K. (1985), Analysis of the Spatial Structure of Properties of Selected Aquifers, Water Resources Research 21, 563-572.

JONES, R. H. (1989), Fitting a Stochastic Partial Differential Equation to Aquifer Head Data, Stochastic Hydrology and Hydraulics 3, 100-105

KITANIDIS, P. K. and LANE, R. W. (1985), Maximum Likelihood Parameter Estimation of Hydrologic Spatial Processes by the Gauss-Newton Method, Journal of Hydrology 17, 31-56.

LE, N. D. and ZIDEK, J. V. (1992), Interpolation with Uncertain Spatial Covariances: A Bayesian Alternative to Kriging, J. Multivariate Analysis 43, 2, 351-374.

MARDIA, K. V. and MARSHALL, R. J. (1984), Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika 71, 135-146.

MARDIA, K. V. and WATKINS, A. J. (1989), On Multimodality of the Likelihood in the Spatial Linear Model, Biometrika 76, 289-95.

MATÉRN, B. (1986), Spatial Variation Second Ed, Lecture Notes in Statistics, 36, Berlin: Springer-Verlag.

MATHERON, G. (1965), Les Variables Régionalisées et leur Estimation Paris: Masson.

OMRE, H. (1987), Bayesian Kriging - Merging Observations Qualified Guesses in Kriging, Mathematical Geology 19, 25-39.

OMRE, G. M. and HALVARSEN, D. F. (1989), A Bayesian Approach to Kriging, In M. Armstrong (ed.), Proceedings of the Third International Geostatistics Congress I, pp. 49-68, Dordrecht: Academic Publishers.

PILZ, J. (1991), Bayesian Estimation and Experimental Design in Linear Regression Models, New York: Wiley.

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

RIPLEY, B. D. (1988), Statistical Inference for Spatial Processes, Cambridge: Cambridge University Press.

SWITZER, P. (1984), Inference for Spatial Autocorrelation Functions, in G. Verly et al (eds.), Geostatistics for Natural Resources Characterization, NATO ASI Series C Vol. 122, pp. 127-140. Reidel Publishing Co., Boston, MA.

WARNES, J. J. (1986), A Sensitivity Analysis for Universal Kriging, Mathematical Geology 18, 653-676.

WHITTLE, P. (1954), On Stationary Processes in the Plane, Biometrika 41, 434-449.

WHITTLE, P. (1962), Topographic Correlation, Power-Law Covariance Functions, and Diffusion, Biometrika 49, 305-314.

WOODBURY, A. D. (1989), Bayesian Updating Revisited, Mathematical Geology 21, 285-308.

ZELLNER, A. (1971), An Introduction to Bayesian Inference in Econometrics, Reprinted 1987, New York: Wiley.