



## American Society for Quality

---

A Bayesian Analysis of Kriging

Author(s): Mark S. Handcock and Michael L. Stein

Source: *Technometrics*, Vol. 35, No. 4 (Nov., 1993), pp. 403-410

Published by: American Statistical Association and American Society for Quality

Stable URL: <http://www.jstor.org/stable/1270273>

Accessed: 03/10/2009 01:48

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=astata>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



American Statistical Association and American Society for Quality are collaborating with JSTOR to digitize, preserve and extend access to *Technometrics*.

<http://www.jstor.org>

# A Bayesian Analysis of Kriging

**Mark S. HANDCOCK**

Department of Statistics and Operations Research  
Leonard N. Stern School of Business  
New York University  
New York, NY 10012-1126

**Michael L. STEIN**

Department of Statistics  
University of Chicago  
Chicago, IL 60637

This article 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 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. These ideas are implemented using topographical data from Davis.

KEY WORDS: Bayesian statistics; Interpolation; Robustness; Spatial statistics.

## 1. INTRODUCTION

In this article 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 Section 2. 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 important from our perspective, an estimate of the uncertainty of that prediction can be derived from the model.

The statistical model is seldom exactly known beforehand and is usually estimated from the very same data from which the predictions are made. The objective of this article 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 that we wish to predict the elevation at which the streams join at the base of the hill. We actually have an observation there, indicat-

ing 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 expected when the estimated uncertainties are based on alternative point estimates of the model.

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 in which 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 Halvorsen (1989), and Woodbury (1989), there appears to be no work from within the geostatistical community using the Bayesian perspective. Omre and Halvorsen (1989) described a Bayesian approach to predicting the depth of geologic horizons based on seismic reflection times. They noted the Bayesian interpretation of ordinary kriging and used prior information about the mean function only, not accounting for uncertainty in the covariance structure. Of course, the situation is a direct extension of

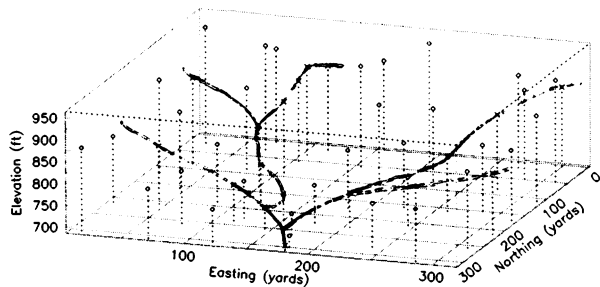


Figure 1. Elevations of the North Face of a Hill From Davis (1973), Table 5.11. The aspect ratio in the figure is spatially correct. The solid lines indicate the small streams running down the hill. Elevations on the streams are indicated by  $\times$  and those not on the streams are marked by  $o$ .

standard Bayesian work in linear models, on which, for example, Box and Tiao (1973, sec. 2.7) and Zellner (1971, sec. 7) are textbook references. Both sketched results using natural vague prior distributions for the parameters. Much of the work in Bayesian time series focuses on the estimation of the parameters of particular autoregressive moving average (ARMA) models. Zellner (1971, sec. 5) derived the predictive distribution of a future observation from a not necessarily stationary AR(1) process. Broemeling (1985) gave a discussion of standard regression and mixed models. His section 5 extended Zellner's (1971) work by using the proper conjugate prior distributions for autoregressive series. He did not extend the work on prediction. All of these analyses assume that the ARMA orders are known. The work of Harrison and Stevens (1976) and West and Harrison (1986) on forecasting using dynamic fully Bayesian models is also of interest. The last few years have seen an explosion of work in Bayesian time series, especially on state-space approaches using the Kalman filter. For a summary, consider the essay by Broemeling and Shaarawy (1986). Here we will focus on a parametric representation of the covariance structure as its direct interpretation is of interest. Nonparametric approaches were developed by Le and Zidek (1992) and Pilz (1991). An alternative non-Bayesian approach to covariance parameter uncertainty was given by Switzer (1984). What is novel about this article 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. The assumption that the covariance structure is known is relaxed and the Bayesian formulation is developed in Section 3. The focus is the evaluation of the performance of the traditional *plug-in* kriging procedure. This evaluation is illustrated in Section 4 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 that  $Z(x)$  is a real-valued stationary Gaussian random field on  $R$  with mean  $E\{Z(x)\} = f(x)'\beta$ , where  $f(x) = \{f_1(x), \dots, f_q(x)\}'$  is a known vector-valued function and  $\beta$  is a vector of unknown regression coefficients. Furthermore, the covariance function is represented by  $\text{cov}\{Z(x), Z(y)\} = \alpha K_\theta(x, y)$  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  $\Theta$  is an open set in  $\mathbb{R}^p$ . The division is purely formal because  $\theta$  may also determine aspects of scale. In the general case, we observe  $\{Z(x_1), \dots, Z(x_n)\}' = Z$  and will focus on the prediction of  $Z(x_0)$ . The kriging predictor is the best linear unbiased predictor of the form  $\hat{Z}_\theta(x_0) = \lambda(\theta)'Z$ —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  $\hat{Z}_\theta(x_0)$  is given by

$$\lambda(\theta)' = b'_\theta(F'K_\theta^{-1}F)^{-1}F'K_\theta^{-1} + k'_\theta K_\theta^{-1}, \quad (2.1)$$

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

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  $\alpha$  or  $\beta$ . Under our Gaussian model, for fixed  $\alpha, \beta$ , and  $\theta$ , the conditional distribution of  $Z(x_0)$  and  $e_\theta(x_0)$  given  $Z$  are

$$Z(x_0)|Z \sim N(k'_\theta K_\theta^{-1}Z + b'_\theta\beta, \alpha\{K_\theta(x_0, x_0) - k'_\theta K_\theta^{-1}k_\theta\})$$

and

$$e_\theta(x_0)|Z \sim N(b'_\theta(\beta - \hat{\beta}(\theta)), \alpha\{K_\theta(x_0, x_0) - k'_\theta K_\theta^{-1}k_\theta\}),$$

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

$$e_\theta(x_0) \sim N(0, \alpha V_\theta), \quad (2.2)$$

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  $\alpha V_\theta$  is the usual prediction error variance as given by 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  $Z$ . Kriging, however, 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  $\beta$  prior to analyzing the data. The underlying kriging approach usually presumes ignorance about  $\beta$  and the unrelatedness of  $\beta$  to the behavior of the covariance function. This latter philosophy will be followed throughout the article. Under these assumptions, an appropriate prior distribution has  $\text{pr}(\beta|\alpha, \theta)$  locally uniform. The posterior distribution of  $\beta$  is then

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

The posterior distribution of the prediction error is then

$$\begin{aligned} \text{pr}(e_\theta(x_0)|\alpha, \theta, Z) \\ \propto \int_\beta \text{pr}(e_\theta(x_0)|\alpha, \beta, \theta, Z) \text{pr}(\beta|\alpha, \theta, Z) d\beta, \end{aligned}$$

which is, by direct calculation,

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

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

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

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 follows: Ordinary kriging is "Bayesian" with the noninformative prior for the mean parameter.

### 3. KRIGING WITH UNKNOWN COVARIANCE PARAMETERS

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  $\Theta$ .

In traditional kriging, one estimates  $\alpha$  and  $\theta$  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 (1983), Kitanidis and Lane (1985), and Hoeksema and Kitanidis (1985). An article by 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 to (2.1) and (2.2). If  $\theta$  is known so that only the location parameter  $\beta$  and the scale parameter  $\alpha$  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  $\theta$ . Although the restriction to a parametric class is a significant assumption, it still allows great latitude.

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

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

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

$$Z(x_0)|\theta, Z \sim t_{n-q} \left( \hat{Z}_\theta(x_0), \frac{n}{n-q} \hat{\alpha}(\theta) V_\theta \right), \tag{3.1}$$

a shifted  $t$  distribution on  $n - q$  df.

The marginal posterior distribution of  $\theta$  can be shown to be

$$\begin{aligned} \text{pr}(\theta|Z) \propto \text{pr}(\theta) \\ \cdot |K_\theta|^{-1/2} |F' K_\theta^{-1} F|^{-1/2} \hat{\alpha}(\theta)^{-(n-q)/2}. \end{aligned} \tag{3.2}$$

The Bayesian predictive distribution for  $Z(x_0)$  is  $\text{pr}(Z(x_0)|Z) \propto \int_\Theta \text{pr}(Z(x_0)|\theta, Z) \cdot \text{pr}(\theta|Z) d\theta$ , where the integrand is given by (3.1) and (3.2). Because the dependence of  $K_\theta$  on  $\theta$  is not specified, this expression cannot 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  $(\alpha, \beta, \theta)$  are envisaged.

Suppose that 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 one derived from empirical correlation functions. The distribution that an investigator would use as a basis for inference about  $Z(x_0)$  would be  $Z(x_0)|\tilde{\alpha}, \tilde{\theta}, Z \sim N(\hat{Z}_{\tilde{\theta}}(x_0), \tilde{\alpha} V_{\tilde{\theta}})$ , plugging in  $(\tilde{\alpha}, \tilde{\theta})$  for  $(\alpha, \theta)$  in (2.4).

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

#### 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 subclass with smoothness parameter  $\theta_2 = \frac{1}{2}$ ; that is  $K_E(x) = \exp(-x/\theta_1)$ . The subclass defined by  $\theta_2 = 1$  was introduced by Whittle (1954) as a model for two-dimensional fields. It is commonly used in hydrology (Creutin and Obled 1982; Jones 1989; Mejía and Rodríguez-Iturbe 1974). As  $\theta_2 \rightarrow \infty$ ,  $K_\theta(x) \rightarrow \exp(-x^2/\theta_1^2)$ , 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 because realizations from it are infinitely differentiable.

The isotropic correlation functions have the general form

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

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

A field with this covariance function is  $\lceil \theta_2 - 1$  times (mean-squared) 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 Cramér and Leadbetter 1967, secs. 4.2, 7.3, and 9.2–9.5).

All calculations of  $\mathcal{H}_{\theta_2}$  in this work use the RKBESL algorithm from the SPECFUN (Cody 1987) library available from NETLIB (Dongarra and Du Croz 1985). A general treatment was 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, originally from Davis (1973). It was studied by Ripley (1981, pp. 58–72) and subsequently by Warnes (1986), Warnes and Ripley (1987), Ripley (1988, pp. 15–21), and Mardia and Watkins (1989). The original data were scaled so that 50 yards in location correspond 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. It is difficult to determine, however, 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 give little guidance 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.

As indicated in 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.

One of the most common methods for fitting a covariance model to data is to match by eye a theoretical curve to the empirical correlation plot of the detrended observations. This guide to intuition is very dubious for four reasons. First, the values in

the plot are very highly correlated so that the additional information in the latter points is very small. Second, each point is based on the average of greatly differing numbers of pairs of points. Third, misspecification of the mean function will have a big effect on the points at medium to large lags. Fourth, fitting by eye is particularly inappropriate for smooth processes. If the field is differentiable,  $\theta_2 > 1$ , then the slope at the origin of the correlation function will be zero. Thus the smoothness of an underlying differentiable field will be difficult to determine based on the empirical correlation plot.

The Exponential model, although 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. Handcock (1989) undertook an analysis based on the much richer Matérn class. It was shown that the corresponding likelihood surfaces were unimodal, and the class appears to be an appropriate model for this topographical data when the mean function included the northing and the distance to the closest stream. As the model for the mean function became more sophisticated, the estimated covariance structures became shorter ranged and smoother. In addition, the uncertainty in the parameters increases.

#### 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.

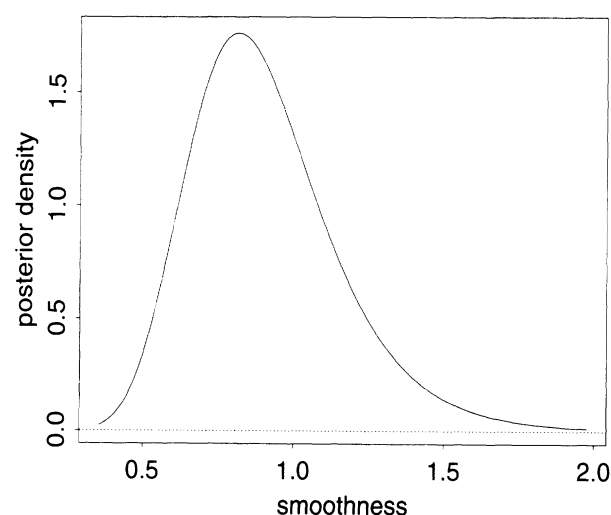


Figure 2. Posterior Distribution for the Smoothness Parameter Based on the Matérn Model With Constant Mean.

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,  $\theta_2 = \frac{1}{2}$ , from one to two dimensions. It corresponds to a random field with continuous realizations that are on the margin of mean-squared differentiability. For  $\theta_2 > 1$  the field is mean squared 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 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  $(\beta, \alpha)$  by taking the marginal prior of  $(\beta, \alpha)$  to be the usual Gaussian-gamma conjugate prior from generalized least squares. An informative prior for  $\theta_2$  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.

#### 4.2 Bayesian Prediction Based on the Matérn Class

The location chosen to be predicted is the center of the region [at (150, 150) on Fig. 1]. It was chosen to be reasonably distant from the survey locations.

Figure 3 presents the predictive densities based on the Matérn model with constant mean function. The plug-in predictive distribution based on the maximum likelihood estimator (MLE)  $(\hat{\alpha}, \hat{\theta}) = (3900, 192, 0.97)$ , is Gaussian centered at 817 feet with a standard deviation of about 20 feet. It is quite close in shape to the Bayesian predictive distribution under the model. Hence probability regions based on this plug-in predictive distribution will be similar to those under the Bayesian predictive distribution.

#### 4.3 Assessing Particular Covariance Functions

Ripley (1981) investigated covariance functions based on fitting by eye the empirical correlation function. The model suggested by Warnes and Ripley (1987) and Ripley (1988), again based on empirical correlation plots, is Exponential with standard deviation 65 feet and range 2 units [ $(\hat{\alpha}, \hat{\theta}) = (4225, 141, 0.5)$  in our notation] and constant mean. The plug-in predictive distribution based on this estimate is Gaussian centered at 820 feet with a standard deviation of about 39 feet (Fig. 3). Probability intervals

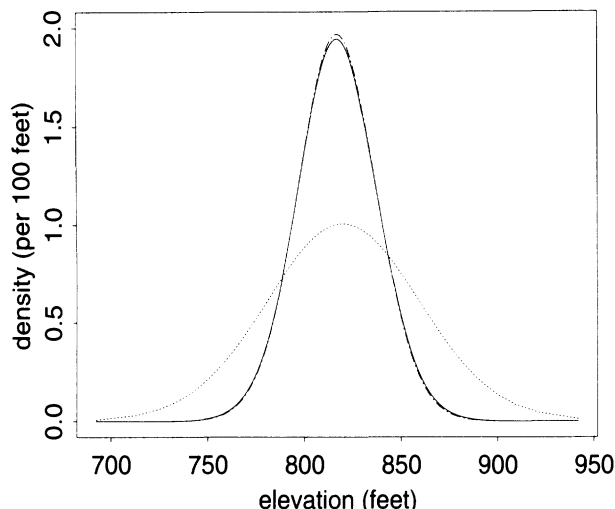


Figure 3. Predictive Distributions Based on the Matérn Model With Constant Mean: —, Bayesian, ···, Based on a Fit by Eye; ---, Based on the MLE. The plug-in distributions are based on the MLE and the value suggested by Warnes and Ripley (1987).

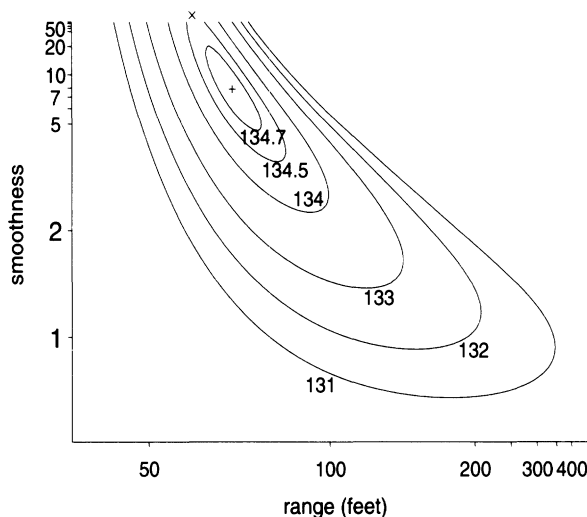


Figure 4. Profile Log-likelihood for the Matérn Model With Mean Based on the Northing and the Distance to Closest Stream: +, Matérn MLE; x, Squared Exponential MLE. The observed elevation at this location is represented by the small horizontal bar reflecting the recording accuracy.

based on this plug-in predictive distribution 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. For example, the Bayesian 95% prediction interval has nominally 71% probability under the plug-in predictive distribution. Alternatively, the nominally 95% interval for the plug-in predictive distribution actually has 99.96% probability.

#### 4.4 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 nonstationarity 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 Fig. 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 4 is the profile log-likelihood surface under this model. Figure 5 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 MLE,  $(\hat{\alpha}, \hat{\theta}) = (955, 68, 7.8)$ , 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.

Three alternate plug-in predictive distributions based on the MLE's under the Exponential and Squared Exponential classes and the maximum a posteriori (MAP) Matérn value are also represented. The plug-in distributions differ substantially from the Bayesian predictive distribution. Smoother estimated models correspond to less perceived uncertainty in the prediction. For example, the nominally

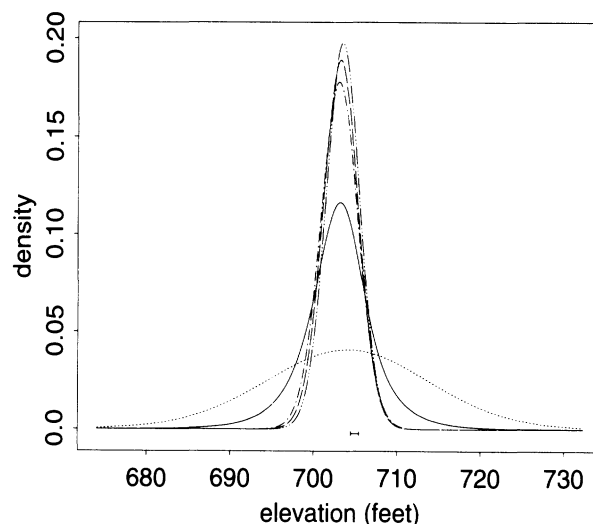


Figure 5. Predictive Distributions Based on the Matérn Model With Mean Using the Northing and the Distance to Closest Stream: —, Bayesian; ···, Exponential MLE; ---, Matérn MAP; - · -, Matérn MLE; — · —, Squared Exponential MLE.

95% confidence interval under the maximum likelihood Matérn model has 73% posterior probability content.

#### 4.5 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 of the smoothness parameter is used. Alternatively one could use the prior

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

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 6 compares the Bayesian predictive distribution using this prior deemphasizing larger smoothness values to the uniform prior used in Figure 5. The effect is to increase the uncertainty in the prediction, in line with the less smooth models. The influence appears to be insensitive to moderate changes in the prior for  $\theta_2$ . Based on figures not presented here, we find that using a uniform prior for  $\alpha$  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  $\beta$ .

Although the MLE is a good representative value, the overall flatness of the likelihoods would suggest against choosing any particular member as the "truth."

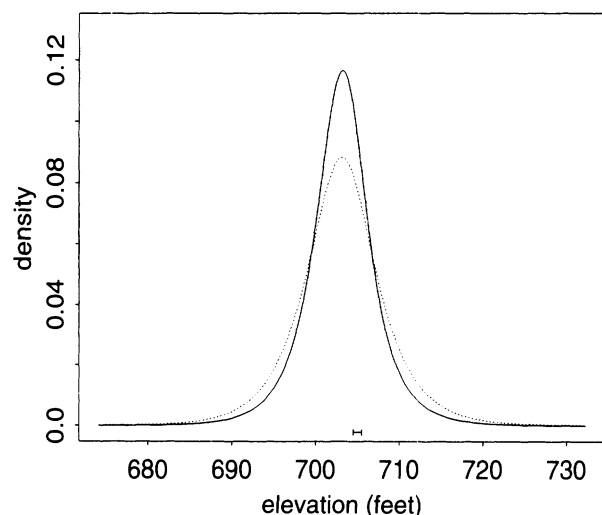


Figure 6. Predictive Distributions Based on the Matérn Model Under a Uniform Prior and a Prior Deemphasizing Larger Smoothness Values:  $\cdots$ , Prior Deemphasizing Smoothness; —, Based on Uniform Prior. The observed elevation at this location is represented by the small horizontal bar reflecting the recording accuracy.

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 is not known and 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 article, 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 MLE may be the best single representative available, but this reduction itself can be detrimental to the inference (see Fig. 5).

[Received March 1990. Revised December 1992.]

#### 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, MA: Addison-Wesley.
- Broemeling, L. D. (1985), *Bayesian Analysis of Linear Models*, New York: Marcel Dekker.
- Broemeling, L. D., and Shaarawy, S. (1986), "A Bayesian Analysis of Time Series," in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, eds. P. K. Goel and A. Zellner, Amsterdam: Elsevier, pp. 337–354.
- Cody, W. J. (1987), "SPECFUN—A Portable Special Function Package," in *New Computing Environments: Microcomputers in Large-Scale Scientific Computing*, ed. A. Wouk, Philadelphia: Society of Industrial and Applied Mathematics, pp. 1–12.
- Cramér, H., and Leadbetter, M. R. (1967), *Stationary and Related Stochastic Processes*, New York: John Wiley.



- Creutin, J. D., and Obled, C. (1982), "Objective Analyses and Mapping Techniques for Rainfall Fields: An Objective Comparison," *Water Resources Research*, 18, 413–431.
- Davis, J. C. (1973), *Statistics and Data Analysis in Geology*, New York: John Wiley.
- Dongarra, J. J., and Du Croz, J. (1985), "Distribution of Mathematical Software via Electronic Mail," Technical Report MCS-TM-48, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60440.
- Handcock, M. S. (1989), "Inference for Spatial Gaussian Random Fields When the Objective is Prediction," unpublished Ph.D. thesis, University of Chicago, Dept. of Statistics.
- Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 38, 205–247.
- 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. (1983), "Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrologic Applications," *Water Resources Research*, 19, 909–921.
- 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 Covariance: A Bayesian Alternative to Kriging," *Journal of Multivariate Analysis*, 43, 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* (2nd ed.). Berlin: Springer-Verlag.
- Matheron, G. (1965), *Les Variables Régionalisées et Leur Estimation*, Paris: Masson.
- Mejía, J. M., and Rodríguez-Iturbe, I. (1974), "On the Synthesis of Random Field Sampling From the Spectrum: An Application to the Generation of Hydrologic Spatial Processes," *Water Resources Research*, 10, 705–711.
- Omre, H. (1987), "Bayesian Kriging—Merging Observations and Qualified Guesses in Kriging," *Mathematical Geology*, 19, 25–39.
- Omre, G. M., and Halvorsen, D. F. (1989), "A Bayesian Approach to Kriging," in *Proceedings of the Third International Geostatistics Congress I*, ed. M. Armstrong, Dordrecht: Academic Publishers, pp. 49–68.
- Pilz, J. (1991), *Bayesian Estimation and Experimental Design in Linear Regression Models*, New York: John Wiley.
- Ripley, B. D. (1981), *Spatial Statistics*, New York: John Wiley.
- (1988), *Statistical Inference for Spatial Processes*, Cambridge, MA: Cambridge University Press.
- Switzer, P. (1984), "Inference for Spatial Autocorrelation Functions," (NATO ASI Ser. C, Vol. 122), in *Geostatistics for Natural Resources Characterization*, ed. G. Verly, M. David, A. G. Journel, and A. Marechal, Boston: Reidel, pp. 127–140.
- Warnes, J. J. (1986), "A Sensitivity Analysis for Universal Kriging," *Mathematical Geology*, 18, 653–676.
- Warnes, J. J., and Ripley, B. D. (1987), "Problems With Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes," *Biometrika*, 74, 3, 640–642.
- West, M., and Harrison, P. J. (1986), "Monitoring and Adaptation in Bayesian Forecasting Models," *Journal of the American Statistical Association*, 81, 741–750.
- Whittle, P. (1954), "On Stationary Processes in the Plane," *Biometrika*, 41, 434–449.
- (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*, New York: John Wiley. (Reprinted 1987.)