# Random Functions and Geological Subsurfaces†

Petter Abrahamsen[1,3] Henning Omre[2,3]

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

[2]Norwegian Institute of Technology, Department of Mathematical Sciences, N-7034 Trondheim, Norway.

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

**Abstract**

The objective of the presentation is to show how the theory of Gaussian random functions (fields) can be used for describing geological structures. It will be demonstrated how Gaussian random functions can be used to obtain the most probable description and to model variability. Depth conversion of seismic travel time maps to depth maps will be used as an illustration. The ability for Gaussian random field models to integrate such diverse information as depth, dip and velocity information in wells, seismic travel time and velocity maps, and even subjective knowledge on velocity fields, will be outlined. Properties of Gaussian random functions will be presented. Some underlying theoretical properties will be given, but emphasis is made on the practical side. Especially the use of spatial prediction and spatial simulation will be considered in some detail.

## 1 Introduction

Modelling of natural phenomena requires the ability to quantify the uncertainty not accounted for by measurements and interpretation. This calls for stochastic models to enrich the deterministic description by adding random components describing variability. The mathematical complexity makes the toolbox of available models restrictive. The most thoroughly studied models are continuous random functions and in particular Gaussian random functions. These have been extensively studied in the literature, see e.g. Doob (1953), Cramér & Leadbetter (1967), Matheron (1973), Adler (1981, 1990),

†Invited lecture presented at ECMOR IV, 4th European Conference on the Mathematics of Oil Recovery, Røros, Norway, 7.–10. june 1994.

Vanmarcke (1983), Yaglom (1986*a*, 1986*b*), and Matérn (1986). Most of these books are theoretical and sparsely consider practical applications.

During the sixties random functions were put into practical work for predicting ore reserves with associated precision measures in the mining industry. This was the start of an activity usually referred to as geostatistics. The methods developed are commonly referred to as kriging. Kriging predictors were first described by the French mathematician George Matheron for use in mining applications (Cressie 1990). By now, kriging is a standard technique described in numerous textbooks. The most recent are Journel & Huijbregts (1978), Ripley (1981, 1988), Isaaks & Srivastava (1989), Cressie (1991), and Christakos (1992). The main applications of kriging are still within earth sciences such as mining, petroleum exploration, hydrology, and metrology, see e.g. the collection of articles in Soares (1993). Thus, the random fields considered are usually defined on $\mathbf{R}^d$ where $d = 2, 3$ or 4 in spatial-temporal settings. Recently, kriging techniques in a high dimensional parameter space has found its way into the exploration and utilization of experimental designs (Sacks, Welch, Mitchell & Wynn 1989).

The cost of computer resources has dropped considerably during the last decade, so geostatistical simulation, as presented in Journel (1974), has become more easily available. This has boosted the development of simulation methods having a much broader scope than the more restricted analytical kriging approaches. Thus, more complicated problems can be assessed but usually at the cost of extensive computer resources. In parallel with the application of traditional kriging techniques, new developments breaking out of the Gaussian framework has been made. The disjunctive kriging (Matheron 1976) and indicator kriging (Journel 1989) are examples of this. Moreover, models for discrete random functions like mosaic variables and event variables are to a lesser extent used in the earth sciences (Hjort & Omre 1993). Examples of these random functions are Markov random fields (Besag 1974) and marked point models (Stoyan, Kendall & Mecke 1987).

This paper focuses on Gaussian random functions and the intention is to shed light on the possibilities and obstacles of their use in modelling of geological subsurfaces.

## 2 Gaussian Random Functions

Consider a standardized Gaussian random function $\mathcal{E}(\mathbf{x})$ with $\mathbf{x} \in \mathbf{R}^d$, defined such that

$$
\begin{aligned}
\mathrm{E}\{\mathcal{E}(\mathbf{x})\} &= 0 \\
\mathrm{Var}\{\mathcal{E}(\mathbf{x})\} &= 1 \\
\mathrm{Cov}\{\mathcal{E}(\mathbf{x}'), \mathcal{E}(\mathbf{x}'')\} &= \rho(\mathbf{x}', \mathbf{x}'').
\end{aligned}
$$

The function $\rho(\mathbf{x}', \mathbf{x}'')$ is the spatial correlation function. Gaussianity entails that for any $n$, and for any configuration $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the random vector $\boldsymbol{\mathcal{E}}^T = [\mathcal{E}(\mathbf{x}_1), \ldots, \mathcal{E}(\mathbf{x}_n)]$ has a multi-Gaussian distribution $\boldsymbol{\mathcal{E}} \sim N_n(\mathbf{0}, \mathbf{C})$, where $\mathbf{C}_{ij} = \rho(\mathbf{x}_i, \mathbf{x}_j)$ is a correlation matrix.

Since linear transformations of Gaussian random functions are Gaussian random functions it is possible to construct a general Gaussian random function based on the standardized Gaussian random function as

$$Z(\mathbf{x}) = m(\mathbf{x}) + \sigma(\mathbf{x})\, \mathcal{E}(\mathbf{x}). \tag{2.1}$$

The expectation and covariances become

$$\begin{aligned} \mathrm{E}\{Z(\mathbf{x})\} &= m(\mathbf{x}) \\ \mathrm{Var}\{Z(\mathbf{x})\} &= \sigma^2(\mathbf{x}) \\ \mathrm{Cov}\{Z(\mathbf{x}'), Z(\mathbf{x}'')\} &= \sigma(\mathbf{x}')\, \sigma(\mathbf{x}'')\, \rho(\mathbf{x}', \mathbf{x}''). \end{aligned}$$

Virtually all properties of a Gaussian random function are determined by the expectation, the variance, and the correlation function:

(i) The spatially dependent expectation, $m(\mathbf{x})$, determines the most probable value of $Z(\mathbf{x})$ at location $\mathbf{x}$.

(ii) The spatially dependent standard deviation, $\sigma(\mathbf{x})$, determines the variance at $\mathbf{x}$.

(iii) The correlation function, $\rho(\mathbf{x}', \mathbf{x}'')$, determines the regularity of the residual random function, $\mathcal{E}(\mathbf{x})$. Assuming a smooth $m(\mathbf{x})$, this carry over to the random function, $Z(\mathbf{x})$, it self. The regularity or smoothness is determined by the behavior of $\rho(\mathbf{x}', \mathbf{x}'')$ as $\mathbf{x}'$ approach $\mathbf{x}''$. Usually $\rho(\mathbf{x}', \mathbf{x}'')$ approach zero for large separations, and hence independence is reached.

EXAMPLE 2.1 The choice of correlation function determines the roughness of $\mathcal{E}(\mathbf{x})$. Figure 2.1 shows four simulated realizations using the exponential correlation function class

$$\rho(\mathbf{x}', \mathbf{x}'') = \exp\left(-3\left(\|\mathbf{x}' - \mathbf{x}''\| / R\right)^{\nu}\right); \qquad (0 < \nu \le 2)$$

with $\nu = 0.5$, 1, 1.5, and 2. It is seen that the choice of the roughness parameter, $\nu$, has large influence on the appearance. The correlation range, $R$, defines the length scale; $\rho \approx 0.05$ for $\|\mathbf{x}' - \mathbf{x}''\| = R$. Choosing $\nu = 2$, as in Figure 2.1(d), gives analytical realizations, whereas letting $\nu \to 0$ give white noise. A wide variety of permissible correlation functions exist; see e.g. Matérn (1986), Yaglom (1986*a*), or Abrahamsen (1994*b*) for more examples. It is occasionally possible to estimate the correlation function, but usually the correlation function must be chosen based on knowledge of the phenomena. □

In many applications the expectation is conveniently modeled as a trend formed by a linear combination of known functions $f_p(\mathbf{x})$:

$$m(\mathbf{x}) = \sum_{p=1}^{P} \beta_p\, f_p(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}. \tag{2.2}$$

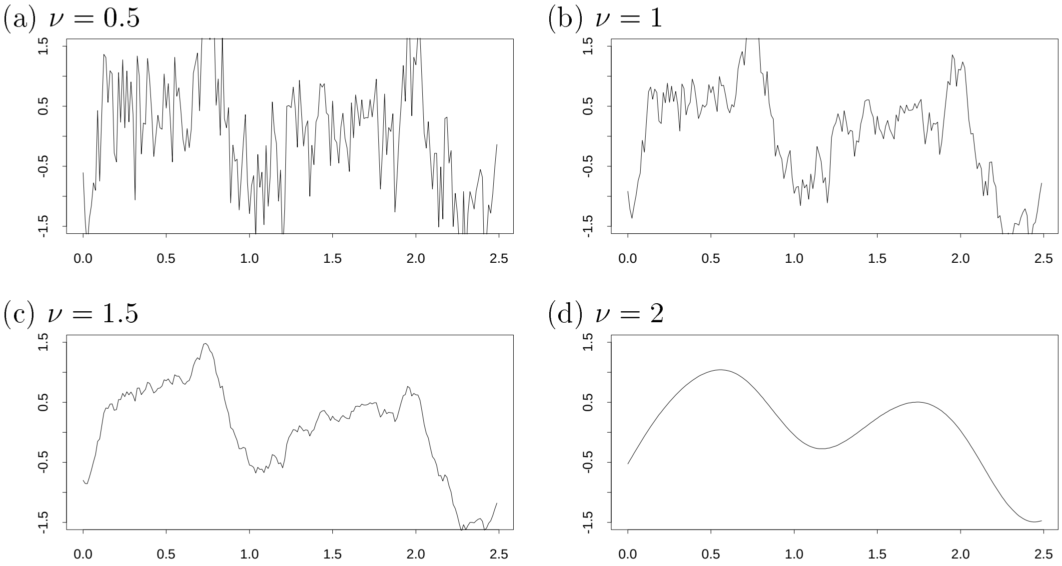


Figure 2.1: *Cross sections of simulated standardized random fields, $\mathcal{E}(\mathbf{x})$, using exponential correlation functions with different roughness parameter $\nu$. The correlation length equals one in all pictures. The same pseudo random numbers has been used in each simulation.*

Assuming $\boldsymbol{\beta}$ is a $P$-dimensional Gaussian random vector means that $Z(\mathbf{x})$ is still a Gaussian random function since the $\beta_p$'s enter linearly. The expectation and covariances become

$$
\begin{aligned}
\mathrm{E}\{Z(\mathbf{x})\} &= \mathbf{f}^T(\mathbf{x})\,\mathrm{E}\{\boldsymbol{\beta}\} \\
\mathrm{Var}\{Z(\mathbf{x})\} &= \mathbf{f}(\mathbf{x})^T \boldsymbol{\Sigma}\,\mathbf{f}(\mathbf{x}) + \sigma^2(\mathbf{x}) \\
\mathrm{Cov}\{Z(\mathbf{x}'), Z(\mathbf{x}'')\} &= \mathbf{f}(\mathbf{x}')^T \boldsymbol{\Sigma}\,\mathbf{f}(\mathbf{x}'') + \sigma(\mathbf{x}')\,\sigma(\mathbf{x}'')\,\rho(\mathbf{x}', \mathbf{x}''),
\end{aligned}
$$

where $\boldsymbol{\Sigma} = \mathrm{Var}\{\boldsymbol{\beta}\}$. The variability increases due to the uncertainty in the random vector $\boldsymbol{\beta}$. Also note that the random variables, $\beta_p$, add infinite range correlations since $\mathbf{f}(\mathbf{x}')^T \boldsymbol{\Sigma}\,\mathbf{f}(\mathbf{x}'')$ is non-zero even for large separation distances.

The model (2.1) with the linear trend (2.2) is a fairly flexible model for a wide variety of natural phenomenon. It can be considered as a generalized regression model in functions $f_p(\mathbf{x})$ with unknown coefficients $\beta_p$.

Example 2.2 As an illustration, consider depth conversion of interpreted seismic travel times to a particular subsurface. The travel times, $\{t(\mathbf{x}); \mathbf{x} \in D \subset \mathbf{R}^2\}$, are assumed known on a dense grid, $\mathfrak{L}_D$, covering $D$. The travel time at $\mathbf{x}$ constitutes spatial averages over an area round $\mathbf{x}$. Moreover, assume that the interval velocity field can be described as a function of the travel times, say

$$v(\mathbf{x}) = v_0 + v_1\, t(\mathbf{x}).$$

This entails a possible increase in interval velocity at larger depths for positive $v_1$. Let $Z(\mathbf{x})$ denote the depth to the subsurface:

$$Z(\mathbf{x}) = v(\mathbf{x})\, t(\mathbf{x}) + \sigma(\mathbf{x})\, \mathcal{E}(\mathbf{x}).$$

(a) $\nu = 1.5$ (b) $\nu = 2$

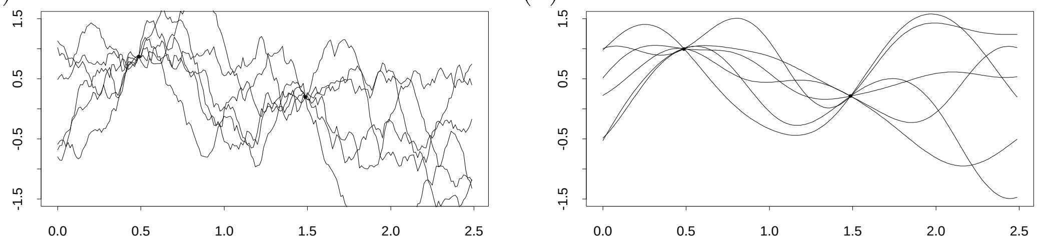


Figure 2.2: *Cross sections of simulated standardized random fields conditioned on data at 0.5 and 1.5. Exponential correlation functions with roughness parameter $\nu$ equal 1.5 and 2 has been used. The correlation length is one in both pictures. Each picture contains 6 different realizations.*

This obviously corresponds to (2.1) with the linear trend

$$m(\mathbf{x}) = v(\mathbf{x})\, t(\mathbf{x}) = v_0\, t(\mathbf{x}) + v_1\, t^2(\mathbf{x}).$$

The travel times give the expected depth, $\sigma(\mathbf{x})$ is related to the precision of the seismic measurements and interpretation of travel times, and $\mathcal{E}(\mathbf{x})$ represents the averaging process. □

## Observations

The modeling of random functions is normally supported by observations in some locations, say $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$:

$$\mathbf{Z}_{\text{obs}}^T = [Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)]; \qquad \mathbf{x}_i \in D,$$

with realizations $\mathbf{z}_{\text{obs}} = [z(\mathbf{x}_1), \ldots, z(\mathbf{x}_n)]$. Hence, the stochastic model of interest is the conditional Gaussian random function

$$[Z(\mathbf{x}) | \mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}]; \qquad \mathbf{x} \in D.$$

The conditional random function inherit the stochastic properties of $Z(\mathbf{x})$ except that it is known at the locations of the observations. Consequently, the variance at the observed location is zero and the conditional expectation must interpolate the observations. The correlation function however, remains unchanged.

Example 2.3 Realizations of conditional Gaussian random functions can be obtained by first simulating unconditional Gaussian random functions and secondly tie them to the observations using simple kriging as discussed below. Figure 2.2 show cross-sections of two sets of standardized Gaussian random functions, $[\mathcal{E}(\mathbf{x}) | \mathcal{E}(0.5) = 0.87, \mathcal{E}(1.5) = 0.20]$, having exponential correlation functions with roughness parameter, $\nu$, equal to 1.5 and 2 respectively. Notice that the spread of the realizations increase rapidly away from the observations; for $\nu = 2$ the increase is linear whereas for $\nu = 1.5$ the increase starts vertical! □

Often properties of a function of the random function is studied:

$$H = h\Big([Z(\mathbf{x})|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]\Big); \qquad \mathbf{x} \in D.$$

The function $h(\cdot)$ may represent the bulk volume between a subsurface and a given oil-water contact, the highest spill point, or the depth at a particular location. The objective is then to determine the probability distribution of $H$. This is most generally done through a sampling approach. This entails generating a set of independent realizations from the stochastic model for $[Z(\mathbf{x})|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]$, denoted by

$$\{[Z(\mathbf{x})|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]_i; \quad i = 1, \ldots, s\}.$$

Inserting each realization into $h(\cdot)$ give a set

$$\{H_i = h\left([Z(\mathbf{x})|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]_i\right); \quad i = 1, \ldots, s\}$$

with empirical distribution approaching the sought probability distribution as $s$ increase. Thus, from this set, the stochastic properties of $H$ is easily inferred. The number of realizations needed, depends on the objective; to obtain a reliable estimate for the expectation or median requires a modest $s$ while reliable estimates for variance or large quantiles require a much larger $s$. Note however, that the empirical distribution is unbiased for any $s$, it is the precision that suffers under small $s$.

The sampling approach outlined above requires that it is possible to simulate realizations of the conditional Gaussian random function $[Z(\mathbf{x})|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]$. This means that one must be able to obtain realizations $\{[Z(\mathbf{x}_l)|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]; l \in \mathfrak{L}_D\}$ where $\mathfrak{L}_D$ is a grid covering $D$. In most applications the number of nodes in $\mathfrak{L}_D$ exceed $10^4$ which means that traditional stochastic simulation techniques for multi-Gaussian random variables fail. Thus, approximate algorithms must be used.

## Analytical Prediction — Kriging

The sampling approach outlined above applies to any function $h(\cdot)$. For certain classes of functions however, analytical solutions are possible. The simplest cases appear when $h(\cdot)$ is linear; for instance

$$\begin{aligned}
H &= [Z(\mathbf{x}_0)|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]; && \text{point value,}\\
H &= a + b[Z(\mathbf{x}_0)|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]; && \text{linear combination,}\\
H &= \int_D w(\mathbf{u})[Z(\mathbf{u})|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]\,\mathrm{d}\mathbf{u}; && \text{spatial average,}\\
H &= \frac{\partial[Z(\mathbf{u})|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}]}{\partial u_i}; && \text{derivative.}
\end{aligned}$$

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

All these $H$'s are Gaussian random variables with expectation and variances deduced from the expectation and covariance function of $[Z(\mathbf{x})|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}]$.

The first case is of particular interest; it provides the probability distribution of $Z(\mathbf{x})$ at a location $\mathbf{x}_0$ given the observations. For the depth conversion example this is the probability distribution for the depth to the subsurface at an arbitrary location $\mathbf{x}_0 \in D$. The distribution is Gaussian so it suffices to obtain the expectation and the variance for a full account of the probability distribution.

First assume that $m(\mathbf{x})$ is known, that is, $\boldsymbol{\beta}$ are known numbers. The centered observations are $\mathbf{z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}$, where $F_{ip} = f_p(\mathbf{x}_i)$. Since $Z(\mathbf{x}_0)$ and the observation vector $\mathbf{Z}_{\text{obs}}$ are $n+1$ multi Gaussian variables, well known formulas for the conditional expectation and variance applies. The result is

$$\mathrm{E}\{Z(\mathbf{x}_0)|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\} = \mathbf{f}^T(\mathbf{x}_0)\,\boldsymbol{\beta} + \mathbf{k}^T(\mathbf{x}_0)\mathbf{K}^{-1}(\mathbf{z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}) \tag{2.3}$$

$$\mathrm{Var}\{Z(\mathbf{x}_0)|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\} = \sigma^2(\mathbf{x}_0) - \mathbf{k}^T(\mathbf{x}_0)\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}_0), \tag{2.4}$$

where the vector $\mathbf{k}(\mathbf{x}_0)$ and the matrix $\mathbf{K}$ are

$$\begin{aligned} \mathbf{k}(\mathbf{x}_0) &= \mathrm{Cov}\{Z(\mathbf{x}_0), \mathbf{Z}_{\text{obs}}\} = \mathrm{Cov}\{Z(\mathbf{x}_0), \mathbf{Z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}\} \\ \mathbf{K} &= \mathrm{Var}\{\mathbf{Z}_{\text{obs}}\} = \mathrm{Var}\{\mathbf{Z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}\}\,. \end{aligned}$$

The predictor (2.3) is commonly called the simple kriging predictor (Journel & Huijbregts 1978, Cressie 1991) whereas the covariance matrix $\mathbf{K}$ is called the kriging matrix.

Consider now the case where the expectation is given as a linear trend (2.2) where the random vector $\boldsymbol{\beta}$ has a prior $P$-dimensional Gaussian distribution: $\boldsymbol{\beta} \sim N_P(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$. The corresponding result can be written as

$$\mathrm{E}\{Z(\mathbf{x}_0)|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\} = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}_0 + \mathbf{k}_B^T(\mathbf{x}_0)\mathbf{K}_B^{-1}(\mathbf{z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}_0) \tag{2.5}$$

$$\begin{aligned} &\mathrm{Var}\{Z(\mathbf{x}_0)|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\} \\ &\qquad = \sigma^2(\mathbf{x}_0) + \mathbf{f}^T(\mathbf{x}_0)\boldsymbol{\Sigma}_0\mathbf{f}(\mathbf{x}_0) - \mathbf{k}_{\mathrm{B}}^T(\mathbf{x}_0)\mathbf{K}_{\mathrm{B}}^{-1}\mathbf{k}_{\mathrm{B}}(\mathbf{x}_0), \end{aligned} \tag{2.6}$$

where the Bayesian covariance vector and kriging matrix are

$$\begin{aligned} \mathbf{k}_{\mathrm{B}}(\mathbf{x}_0) &= \mathrm{Cov}\{Z(\mathbf{x}_0), \mathbf{Z}_{\text{obs}}\} = \mathbf{k}(\mathbf{x}_0) + \mathbf{f}^T(\mathbf{x}_0)\boldsymbol{\Sigma}_0\mathbf{F}^T \\ \mathbf{K}_{\mathrm{B}} &= \mathrm{Var}\{\mathbf{Z}_{\text{obs}}\} = \mathbf{K} + \mathbf{F}\boldsymbol{\Sigma}_0\mathbf{F}^T. \end{aligned}$$

The predictor (2.5) is called the Bayesian kriging predictor (Omre & Halvorsen 1989). It allows prior believes on the coefficients $\boldsymbol{\beta}$ to be incorporated through the choice of the prior expectation, $\boldsymbol{\beta}_0$, and the prior covariances, $\boldsymbol{\Sigma}_0$.

In the limit $\boldsymbol{\Sigma}_0 \to \mathbf{0}$, that is, the prior information is exact, the simple kriging predictor appears. The opposite limit: $\boldsymbol{\Sigma}_0 \to \boldsymbol{\infty}$ (i.e. $\boldsymbol{\Sigma}_0^{-1} \to \mathbf{0}$), the ignorant prior, provides

$$\mathrm{E}\{Z(\mathbf{x}_0)|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\} = \mathbf{f}^T(\mathbf{x}_0)\,\widehat{\boldsymbol{\beta}} + \mathbf{k}^T(\mathbf{x}_0)\,\mathbf{K}^{-1}(\mathbf{z}_{\text{obs}} - \mathbf{F}\widehat{\boldsymbol{\beta}}) \tag{2.7}$$

$$\text{(2.8)}\quad \operatorname{Var}\{Z(\mathbf{x}_0)|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\} = \sigma^2(\mathbf{x}_0) + \mathbf{f}^{*T}(\mathbf{x}_0)\widehat{\boldsymbol{\Sigma}}\mathbf{f}^*(\mathbf{x}_0) - \mathbf{k}^T(\mathbf{x}_0)\,\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}_0),$$

where $\mathbf{f}^*(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0) - \mathbf{k}^T(\mathbf{x}_0)\mathbf{K}^{-1}\mathbf{F}$, and the generalized least squares estimate and estimation covariances for $\boldsymbol{\beta}$ are

$$\text{(2.9)}\qquad \widehat{\boldsymbol{\beta}} = (\mathbf{F}^T\mathbf{K}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{K}^{-1}\mathbf{z}_{\text{obs}}, \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}} = (\mathbf{F}^T\mathbf{K}^{-1}\mathbf{F})^{-1}.$$

The predictor (2.7) is called the universal kriging predictor (Journel & Huijbregts 1978, Ripley 1981, Cressie 1991).

The conditional expectations (2.3), (2.5), or (2.7) respectively, are the best prediction of $Z(\mathbf{x}_0)$ at location $\mathbf{x}_0$. By varying $\mathbf{x}_0$ over $D$ in the depth conversion example, the best pointwise predicted subsurface is obtained:

$$\{\mathrm{E}\{Z(\mathbf{x})|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\}\,;\mathbf{x} \in D\}$$

The corresponding prediction variances given by (2.4), (2.6), or (2.8) respectively provides corresponding error maps given as

$$\{\operatorname{Var}\{Z(\mathbf{x})|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\}^{1/2}\,;\mathbf{x} \in D\}.$$

Whether to use simple, Bayesian, or universal kriging depends on the general experience with the variable under study, on the number of parameters in the trend, and on the number of data. If the trend is known simple kriging provides the obvious method. If the number of data is small and experience recommends the use of a trend model including several parameters, Bayesian kriging provides a robust and flexible method. If the number of data is large compared to the number of parameters, universal kriging gives good predictions without requiring any prior believes to be specified.

EXAMPLE 2.4 As an illustration a 5.5 kilometer cross-section of a geological dome observed in three wells are considered. The trend model is based on observed seismic travel times and is similar to the model in Example 2.2. Figure 2.3(a) shows the prior guess on the trend. Figures 2.3(b), (c), and (d) show predictions using simple, Bayesian, and universal kriging respectively. A spherical correlation function with 3000 meter correlation length has been used. For simple kriging the trend is assumed known and is taken equal to the expectation of the prior guess. For the Bayesian approach the prior guess is a suggestion on the trend of the subsurface with precision given by the prior variance. For the universal kriging approach, the prior guess is ignored. The three well data are hardly enough for estimating the acceleration parameter, $v_1$, so the universal kriging prediction is unreliable away from the wells. On the other hand the simple kriging prediction assumes a known trend model which is hardly realistic, so the associated prediction errors are to small. The Bayesian approach is a reasonable compromise; the acceleration parameter is mainly defined by the prior guess while the $v_0$ parameter is properly estimated from the data. Since the model parameters, $(v_0, v_1)$, have physical interpretations, prior believes on their values are usually present. This justify the use of a Bayesian approach. □

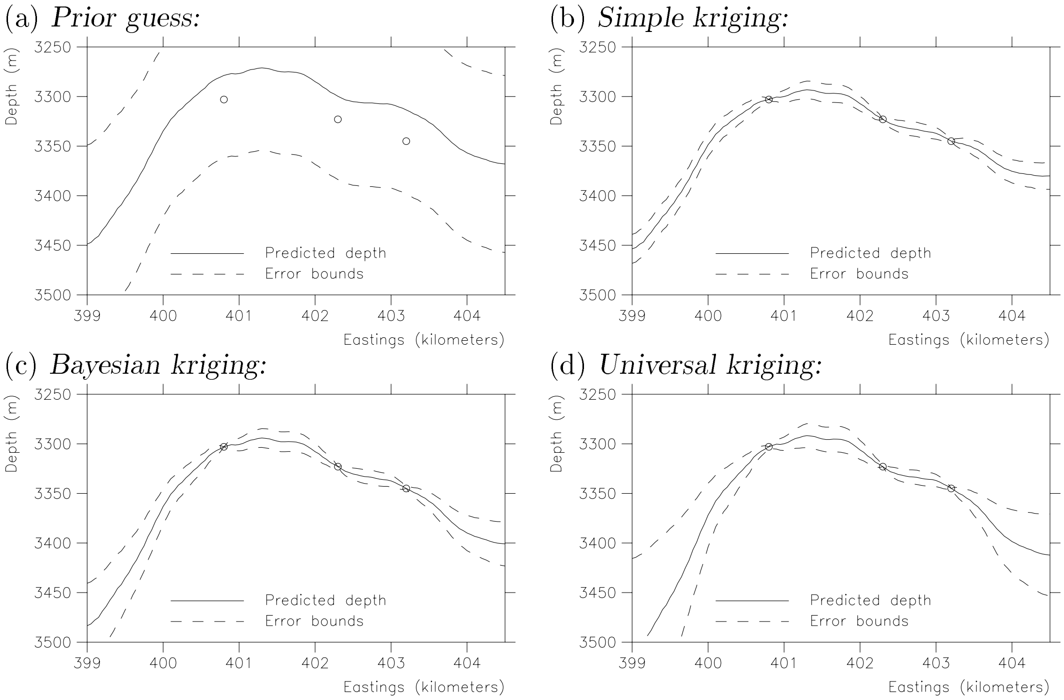


Figure 2.3: *Cross sections of predicted subsurfaces. Figure* (a) *shows the prior guess with uncertainty bounds. Figures* (b), (c), *and* (d) *are predictions obtained using simple, Bayesian, and universal kriging respectively. Error bounds are one standard deviation.*

# 3 Model Extensions

In the previous section depth conversion of seismic travel times was used as an illustration. In this section it will be shown how the stochastic model for a subsurface can be extended to:

- include observation errors,
- model the velocity field as a Gaussian random function and use observations of the velocity field for depth prediction,
- use gradient data from dip meters to improve depth predictions,
- model multiple subsurfaces consistently.

The extended models will consist of a collection of correlated Gaussian random functions and the kriging techniques described in the previous section will be modified to cope with this. Simple kriging conditioned on data from additional random functions is commonly called cokriging. Bayesian and universal kriging can also be modified to include observations from a collection of random functions, so appropriate names could be Bayesian and universal cokriging.

The contents of this section is specific to the application of seismic depth conversion but the basic ideas are useful in many contexts.

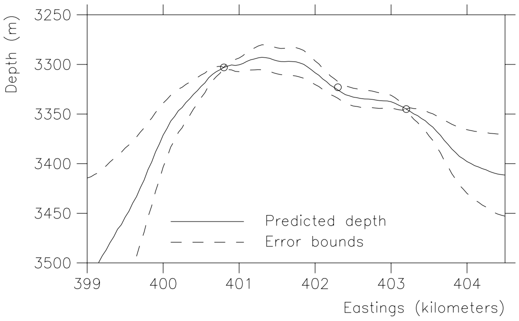


Figure 3.1: *Similar to Figure* 2.3(d) *except that the middle observation is assumed inaccurate.*

### Observation Errors

Observations of random functions are occasionally corrupted by errors. Consider a set of $n$ observations organized in a vector

$$\mathbf{Z}_{\text{obs}}^{\epsilon T} = \mathbf{Z}^T + \boldsymbol{\epsilon}^T = [Z(\mathbf{x}_1) + \epsilon(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n) + \epsilon(\mathbf{x}_n)]; \qquad \mathbf{x}_i \in D,$$

where $\epsilon(\mathbf{x}_i)$ are multi Gaussian measurement errors independent of $Z(\mathbf{x})$ with vanishing expectation. Observing that

$$\begin{aligned} \text{Cov}\{Z(\mathbf{x}_0), \mathbf{Z}_{\text{obs}}^{\epsilon}\} &= \text{Cov}\{Z(\mathbf{x}_0), \mathbf{Z}_{\text{obs}}\} = \mathbf{k}(\mathbf{x}_0) \\ \text{Var}\{\mathbf{Z}_{\text{obs}}^{\epsilon}\} &= \text{Var}\{\mathbf{Z}_{\text{obs}}\} + \text{Var}\{\boldsymbol{\epsilon}\} = \mathbf{K} + \mathbf{K}^{\epsilon} \end{aligned}$$

shows that the expectations and variances conditioned on the noise corrupted data are obtained by substituting $\mathbf{K}$ by $\mathbf{K}+\mathbf{K}^{\epsilon}$ in the predictors and prediction variances, (2.3) through (2.8). The simplest example occurs for independent measuring errors where $\mathbf{K}^{\epsilon} = \text{diag}(\sigma_{\text{err}}^2(\mathbf{x}_1), \ldots, \sigma_{\text{err}}^2(\mathbf{x}_n))$, with $\sigma_{\text{err}}^2(\mathbf{x}_i)$ being the variance of the measuring error at $\mathbf{x}_i$. Thus, the measuring error is simply introduced by adding the corresponding variances to the diagonal of the kriging matrix. The consequence of introducing measuring errors is that the kriging predictors no longer interpolate the observations, and moreover, that the prediction errors are nonzero at these locations.

Example 3.1 Assume that the middle observation in Example 2.4 for some reason is inaccurate. Figure 3.1 show a prediction using universal kriging. This figure should be compared to Figure 2.3(d). It is seen that the prediction fails to honour the middle well observation and moreover, the prediction error is nonzero. □

### Velocity Observations

In the previous section the interval velocity was modelled using a trend formed by a linear sum of known functions multiplied by unknown or partially known coefficients such as the model in Example 2.2. Such a trend is a rough approximation to the real interval velocity field. A more realistic model is to add a stochastic component describing local variations in velocity

$$V(\mathbf{x}) = m_V(\mathbf{x}) + \sigma_V(\mathbf{x})\, \varepsilon_V(\mathbf{x}), \tag{3.1}$$

where $m_V(\mathbf{x}) = \mathbf{g}^T(\mathbf{x})\boldsymbol{\beta}$ is, as previously, a linear trend. Moreover, $\mathcal{E}_V(\mathbf{x})$ is a standardized Gaussian random function accounting for variations not properly modelled by the trend, and $\sigma_V^2(\mathbf{x})$ is the associated laterally dependent variance. To obtain a model for the depth to a subsurface an additional model for the travel times is needed:

$$T(\mathbf{x}) = t(\mathbf{x}) + \sigma_T(\mathbf{x})\,\mathcal{E}_T(\mathbf{x}),$$

where the interpreted travel times, $t(\mathbf{x})$, are assumed to be a proper trend for the 'true' travel time, $T(\mathbf{x})$. Now the residual, $\sigma_T(\mathbf{x})\,\mathcal{E}_T(\mathbf{x})$, is supposed to account for interpretation errors and the smoothing in the seismic signals. The depth model becomes:

$$Z(\mathbf{x}) = V(\mathbf{x})\,T(\mathbf{x}) = \Big(m_V(\mathbf{x}) + \sigma_V(\mathbf{x})\,\mathcal{E}_V(\mathbf{x})\Big)\Big(t(\mathbf{x}) + \sigma_T(\mathbf{x})\,\mathcal{E}_T(\mathbf{x})\Big).$$

To make this expression more tractable two modifications will be made. The product involving $\mathcal{E}_V(\mathbf{x})\,\mathcal{E}_T(\mathbf{x})$ is ignored. This is justified provided the trends are significantly larger than the residuals. Secondly, $m_V(\mathbf{x})\,\sigma_T(\mathbf{x})\,\mathcal{E}_T(\mathbf{x})$ is replaced by a residual $\sigma_Z(\mathbf{x})\,\mathcal{E}_Z(\mathbf{x})$ which is assumed to consist of a known variance, $\sigma_Z^2(\mathbf{x})$, and a standardized Gaussian random function, $\mathcal{E}_Z(\mathbf{x})$ independent of $\mathcal{E}_V(\mathbf{x})$. The resulting expression for the depth reads

$$Z(\mathbf{x}) = m_V(\mathbf{x})\,t(\mathbf{x}) + [\sigma_V(\mathbf{x})\,t(\mathbf{x})]\,\mathcal{E}_V(\mathbf{x}) + \sigma_Z(\mathbf{x})\,\mathcal{E}_Z(\mathbf{x}). \tag{3.2}$$

This expression is in principal equivalent to (2.1). The depth model includes three parts: the velocity trend multiplied by the interpreted travel times, a residual caused by anomalies in the velocity field, and a residual caused by inaccuracies in the interpreted travel time map.

At this point a model for the velocity (3.1) and a model for the depth (3.2) have been established. Both models are Gaussian random functions with properly defined expectations and covariance functions and fits into the kriging framework outlined in the previous section. The new feature is that the models are correlated through the common velocity trend, $m_V(\mathbf{x})$, and the common residual function, $\mathcal{E}_V(\mathbf{x})$. The dependency is given by the cross covariances:

$$\mathrm{Cov}\{Z(\mathbf{x}), V(\mathbf{x}')\} = [\sigma_V(\mathbf{x})\,t(\mathbf{x})]\sigma_V(\mathbf{x}')\,\rho_V(\mathbf{x},\mathbf{x}'),$$

or assuming $\boldsymbol{\beta}$ are multi Gaussian random variables, the cross covariance are

$$\mathrm{Cov}\{Z(\mathbf{x}), V(\mathbf{x}')\} = t(\mathbf{x})\mathbf{g}^T(\mathbf{x})\boldsymbol{\Sigma}_0\,\mathbf{g}(\mathbf{x}') + [\sigma_V(\mathbf{x})\,t(\mathbf{x})]\sigma_V(\mathbf{x}')\,\rho_V(\mathbf{x},\mathbf{x}'),$$

where $\boldsymbol{\Sigma}_0$ is the specified prior covariance matrix of $\boldsymbol{\beta}$.

Assume there exists observations of depth and observations of interval velocities:

$$\mathbf{Z}_{\text{obs}}^T = [Z(\mathbf{x}_1^z), \ldots, Z(\mathbf{x}_n^z)] \quad \text{and} \quad \mathbf{V}_{\text{obs}}^T = [V(\mathbf{x}_1^v), \ldots, V(\mathbf{x}_m^v)].$$

The position of the observations are arbitrary but usually velocity observations are available at a subset of the velocity observations. The objective is to assess the probability distribution of the conditional Gaussian random functions

$$[Z(\mathbf{x})|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}, \mathbf{V}_{\text{obs}} = \mathbf{v}_{\text{obs}}] \quad \text{and} \quad [V(\mathbf{x})|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}, \mathbf{V}_{\text{obs}} = \mathbf{v}_{\text{obs}}].$$

Prediction of $Z(\mathbf{x}_0)$ given the available data are of particular interest. Assume the trend, $m_V(\mathbf{x})$, is known and consider the centered observations $\mathbf{z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}$, where $F_{ip} = g_p(\mathbf{x}_i^z)\,t(\mathbf{x}_i^z)$ and $\mathbf{v}_{\text{obs}} - \mathbf{G}\boldsymbol{\beta}$, where $G_{ip} = g_p(\mathbf{x}_i^v)$. Since $Z(\mathbf{x}_0)$, $\mathbf{Z}_{\text{obs}}$, and $\mathbf{V}_{\text{obs}}$ are a set of $n+m+1$ multi Gaussian random variables, the formulas for conditional expectation and variances give

$$\begin{aligned}(3.3)\quad & \mathrm{E}\{Z(\mathbf{x}_0)|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}, \mathbf{V}_{\text{obs}} = \mathbf{v}_{\text{obs}}\} \\ &\qquad = m_V(\mathbf{x}_0)\,t(\mathbf{x}_0) + \mathbf{k}^T(\mathbf{x}_0)\mathbf{K}^{-1}\begin{bmatrix}\mathbf{z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta} \\ \mathbf{v}_{\text{obs}} - \mathbf{G}\boldsymbol{\beta}\end{bmatrix} \\ (3.4)\quad & \mathrm{Var}\{Z(\mathbf{x}_0)|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}, \mathbf{V}_{\text{obs}} = \mathbf{v}_{\text{obs}}\} \\ &\qquad = [\sigma_V(\mathbf{x}_0)\,t(\mathbf{x}_0)]^2 + \sigma_Z^2(\mathbf{x}_0) - \mathbf{k}^T(\mathbf{x}_0)\mathbf{K}^{-1}\mathbf{k}^T(\mathbf{x}_0),\end{aligned}$$

where the covariance vector and kriging matrix now reads

$$\mathbf{k}(\mathbf{x}_0) = \mathrm{Cov}\left\{Z(\mathbf{x}_0), \begin{bmatrix}\mathbf{Z}_{\text{obs}} \\ \mathbf{V}_{\text{obs}}\end{bmatrix}\right\} = \begin{bmatrix}\mathrm{Cov}\{Z(\mathbf{x}_0), \mathbf{Z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}\} \\ \mathrm{Cov}\{Z(\mathbf{x}_0), \mathbf{V}_{\text{obs}} - \mathbf{G}\boldsymbol{\beta}\}\end{bmatrix}$$

$$\mathbf{K} = \mathrm{Var}\left\{\begin{bmatrix}\mathbf{Z}_{\text{obs}} \\ \mathbf{V}_{\text{obs}}\end{bmatrix}\right\} = \begin{bmatrix}\mathrm{Var}\{\mathbf{Z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}\} & \mathrm{Cov}\{\mathbf{Z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}, \mathbf{V}_{\text{obs}} - \mathbf{G}\boldsymbol{\beta}\} \\ \mathrm{Cov}\{\mathbf{V}_{\text{obs}} - \mathbf{G}\boldsymbol{\beta}, \mathbf{Z}_{\text{obs}} - \mathbf{F}\boldsymbol{\beta}\} & \mathrm{Var}\{\mathbf{V}_{\text{obs}} - \mathbf{G}\boldsymbol{\beta}\}\end{bmatrix}.$$

Similar expressions for the velocities are obtained by using the trend $m_V(\mathbf{x})$ in (3.3) and replacing $Z(\mathbf{x}_0)$ by $V(\mathbf{x}_0)$ in the covariance vector $\mathbf{k}(\mathbf{x}_0)$. Comparing (3.3) and (3.4) to (2.3) and (2.4) reveals that the equations for simple kriging are very similar; the difference is that the kriging matrix and covariance matrix must be expanded to accommodate the cross covariances.

It is straight forward to extend the above formalism to Bayesian or universal kriging. Just as $\mathbf{k}(\mathbf{x}_0)$ and $\mathbf{K}$ were expanded to accommodate the additional cross covariances, $\mathbf{f}(\mathbf{x}_0)$ and $\mathbf{F}$ must be expanded to accommodate the trend functions at all the locations of the observations. For more details see Abrahamsen (1993) from where the following example is taken.

Example 3.2 To illustrate the properties of cokriging consider a velocity function with a constant unknown trend, $m_V(\mathbf{x}) = v_0$. Three velocity and four depth observations are available. Figures 3.2(a), (b), and (c) show cross sections of predictions of the velocity function conditioned on velocity observations alone, depth observations alone and all seven observations. The velocity residual has been chosen to have a exponential correlation function with $\nu = 2$, giving smooth bell-shaped forms around the observations. Bayesian cokriging were used. The 'hump' on the right side of Figures 3.2(b) and (c) are caused by the depth observation with no accompanying velocity observation. From Figure 3.2(b), where velocity observations are ignored, it is seen that the depth data carry valuable information on the shape of

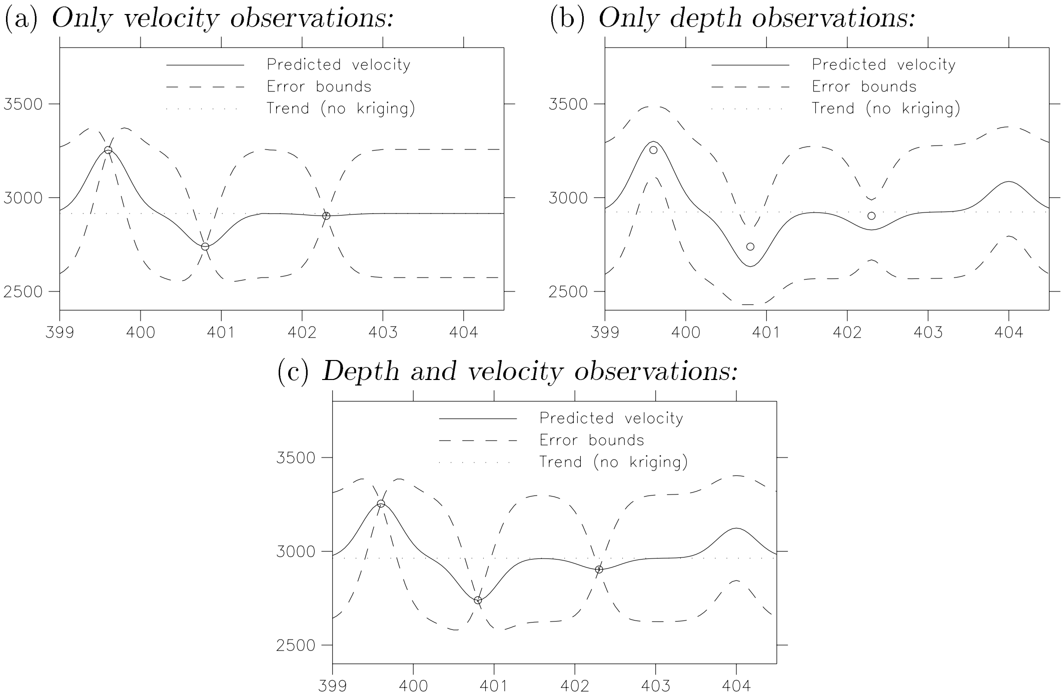


Figure 3.2: *Cross sections of Bayesian velocity predictions conditioned on velocity* (a)*, depth* (b)*, or all observations* (c)*. Velocity along the vertical axis and lateral position in kilometers along the horizontal axis.*

the velocity function. Comparing Figure 3.2(a) to 3.2(c) reveals that using depth data has minor influence in regions where velocity data are already present but significant influence in regions where velocity observations are missing. This is due to high correlations between depth and velocity observation in the same location. This is a sign of redundancy in the data. □

## Dip Observations

Occasionally reliable dip meter observations are available. They carry information on the gradient of the subsurface, that is, the slope and the direction of the slope. When considering the gradient of a Gaussian random function it is tacitly assumed that the gradient actually exist. Assuming a smooth expectation, $m(\mathbf{x})$, and variance, $\sigma^2(\mathbf{x})$, the existence of the gradient is related to the behaviour of the correlation function, $\rho(\mathbf{x}, \mathbf{x}')$, as $\mathbf{x} \to \mathbf{x}'$. A sufficient condition is that [Cramér & Leadbetter (1967, p. 84) and Abrahamsen (1994*b*, p. 25)]

$$\lim_{\mathbf{x} \to \mathbf{x}'} \frac{\partial^2 \rho(\mathbf{x}, \mathbf{x}')}{\partial x_i \, \partial x'_j} = \text{finite} \quad \text{for all } i, j = 1, \ldots, d.$$

($d = 2$ for subsurfaces.) This puts strong restrictions on the correlation function. For instance the exponential correlation functions in Example 2.1 do not compile to this requirement unless $\nu = 2$. Thus, only display (d) in Figure 2.1

possess differentiable curves; this should not come as a surprise considering the appearance of the figures. Display (a–c) are too rough even though they are continuous everywhere. Several other correlation functions give differentiable Gaussian random functions; see Abrahamsen (1994*a*) for some examples.

For a subsurface $Z(\mathbf{x})$ defined on $\mathbf{R}^2$ the associated gradient field, $\dot{\mathbf{Z}}(\mathbf{x})$, is a two-dimensional vector in $\mathbf{R}^2$ defined by its components in a Cartesian coordinate system:

$$\dot{Z}_i(\mathbf{x}) = \frac{\partial Z(\mathbf{x})}{\partial x_i}; \qquad i = 1, 2.$$

Assuming $\mathrm{E}\{Z(\mathbf{x})\} = m(\mathbf{x})$, then

$$\dot{m}_i(\mathbf{x}) = \mathrm{E}\Big\{\dot{Z}_i(\mathbf{x})\Big\} = \frac{\partial m(\mathbf{x})}{\partial x_i}; \qquad i = 1, 2.$$

Further, assume that the covariance function, $C(\mathbf{x}, \mathbf{x}') = \mathrm{Cov}\{Z(\mathbf{x}), Z(\mathbf{x}')\} = \sigma(\mathbf{x})\,\sigma(\mathbf{x}')\,\rho(\mathbf{x}, \mathbf{x}')$, is simultaneously differentiable in $\mathbf{x}$ and $\mathbf{x}'$, that is, $Z(\mathbf{x})$ is mean square differentiable. Then, the cross-covariance function between $Z(\mathbf{x})$ and a component of $\dot{\mathbf{Z}}(\mathbf{x})$ is a vector defined by the components

$$\dot{C}_i(\mathbf{x}, \mathbf{x}') = \mathrm{Cov}\Big\{Z(\mathbf{x}), \dot{Z}_i(\mathbf{x}')\Big\} = \frac{\partial}{\partial x_i'} C(\mathbf{x}, \mathbf{x}'); \qquad i = 1, 2.$$

The covariance functions between components of $\dot{\mathbf{Z}}(\mathbf{x})$ are

$$\ddot{C}_{ij}(\mathbf{x}, \mathbf{x}') = \mathrm{Cov}\Big\{\dot{Z}_i(\mathbf{x}), \dot{Z}_j(\mathbf{x}')\Big\} = \frac{\partial^2}{\partial x_i \partial x_j'} C(\mathbf{x}, \mathbf{x}'); \quad i, j = 1, 2.$$

The matrix $\ddot{C}_{ij}(\mathbf{x}, \mathbf{x}')$ is a second rank tensor called the covariance tensor of the gradient field, whereas the components of the vector, $\dot{C}_i(\mathbf{x}, \mathbf{x}')$, form a first rank tensor.

Assume there exist observations of depth and dip in a number of wells:

$$\mathbf{Z}_{\mathrm{obs}}^T = [Z(\mathbf{x}_1^z), \ldots, Z(\mathbf{x}_n^z)] \quad \text{and} \quad \dot{\mathbf{Z}}_{\mathrm{obs}}^T = [\dot{Z}_1(\mathbf{x}_1^{\dot{\mathbf{z}}}), \dot{Z}_2(\mathbf{x}_1^{\dot{\mathbf{z}}}), \ldots, \dot{Z}_1(\mathbf{x}_m^{\dot{\mathbf{z}}}), \dot{Z}_2(\mathbf{x}_m^{\dot{\mathbf{z}}})].$$

The position of the observations are arbitrary but usually dip observations are available in a subset of the wells. Since $Z(\mathbf{x}_0)$, $\mathbf{Z}_{\mathrm{obs}}$, and $\dot{\mathbf{Z}}_{\mathrm{obs}}$ are a set of $n + 2m + 1$ multi Gaussian random variables, the formulas for conditional expectation and variances apply once more to give the predictor:

$$\mathrm{E}\Big\{Z(\mathbf{x}_0)|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}, \dot{\mathbf{Z}}_{\mathrm{obs}} = \dot{\mathbf{z}}_{\mathrm{obs}}\Big\} = m(\mathbf{x}_0) + \mathbf{k}^T(\mathbf{x}_0)\mathbf{K}^{-1}\begin{bmatrix}\mathbf{z}_{\mathrm{obs}} - \mathbf{F}\boldsymbol{\beta} \\ \dot{\mathbf{z}}_{\mathrm{obs}} - \dot{\mathbf{F}}\boldsymbol{\beta}\end{bmatrix}$$

$$\mathrm{Var}\Big\{Z(\mathbf{x}_0)|\mathbf{Z}_{\mathrm{obs}} = \mathbf{z}_{\mathrm{obs}}, \dot{\mathbf{Z}}_{\mathrm{obs}} = \dot{\mathbf{z}}_{\mathrm{obs}}\Big\} = \sigma^2(\mathbf{x}_0) - \mathbf{k}^T(\mathbf{x}_0)\mathbf{K}^{-1}\mathbf{k}^T(\mathbf{x}_0),$$

where the covariance vector and kriging matrix now reads:

$$\mathbf{k}(\mathbf{x}_0) = \mathrm{Cov}\left\{Z(\mathbf{x}_0), \begin{bmatrix}\mathbf{Z}_{\mathrm{obs}} \\ \dot{\mathbf{Z}}_{\mathrm{obs}}\end{bmatrix}\right\} = \begin{bmatrix}\mathrm{Cov}\{Z(\mathbf{x}_0), \mathbf{Z}_{\mathrm{obs}} - \mathbf{F}\boldsymbol{\beta}\} \\ \mathrm{Cov}\Big\{Z(\mathbf{x}_0), \dot{\mathbf{Z}}_{\mathrm{obs}} - \dot{\mathbf{F}}\boldsymbol{\beta}\Big\}\end{bmatrix}$$

$$\mathbf{K} = \mathrm{Var}\left\{\begin{bmatrix}\mathbf{Z}_{\mathrm{obs}} \\ \dot{\mathbf{Z}}_{\mathrm{obs}}\end{bmatrix}\right\} = \begin{bmatrix}\mathrm{Var}\{\mathbf{Z}_{\mathrm{obs}} - \mathbf{F}\boldsymbol{\beta}\} & \mathrm{Cov}\Big\{\mathbf{Z}_{\mathrm{obs}} - \mathbf{F}\boldsymbol{\beta}, \dot{\mathbf{Z}}_{\mathrm{obs}} - \dot{\mathbf{F}}\boldsymbol{\beta}\Big\} \\ \mathrm{Cov}\Big\{\dot{\mathbf{Z}}_{\mathrm{obs}} - \dot{\mathbf{F}}\boldsymbol{\beta}, \mathbf{Z}_{\mathrm{obs}} - \mathbf{F}\boldsymbol{\beta}\Big\} & \mathrm{Var}\Big\{\dot{\mathbf{Z}}_{\mathrm{obs}} - \dot{\mathbf{F}}\boldsymbol{\beta}\Big\}\end{bmatrix}.$$

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

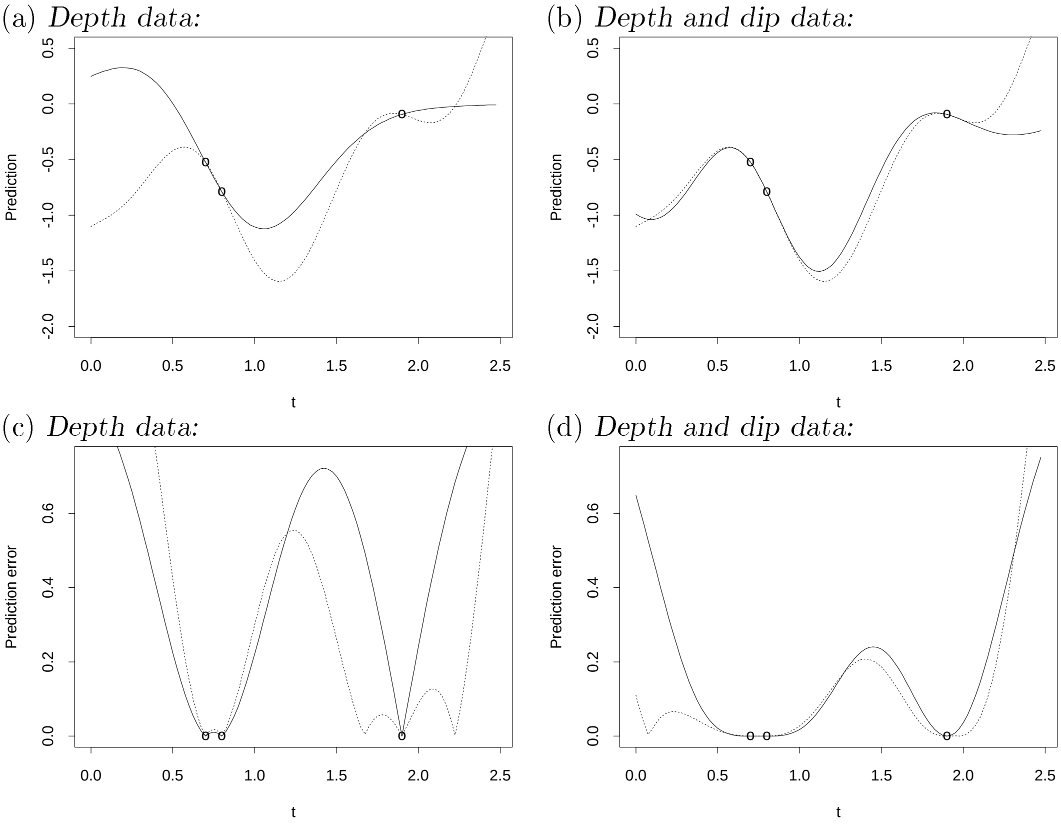


Figure 3.3: *Cross sections of predictions conditioned on depth and dip data. Figures* (a) *and* (b) *show predictions as solid lines and the dashed lines are the 'true' function. Figures* (c) *and* (d) *show the corresponding prediction errors as solid lines whereas the dashed lines are the difference between the prediction and the true function (absolute value).*

All these covariances are given by the covariance function, $C(\mathbf{x}, \mathbf{x}')$, the cross-covariance function, $\dot{C}_i(\mathbf{x}, \mathbf{x}')$, and the covariance tensor, $\ddot{C}_{ij}(\mathbf{x}, \mathbf{x}')$.

Conditioning on gradient data is just another variety of cokriging. The special feature is that all covariance functions are uniquely determined by the basic covariance function $C(\mathbf{x}, \mathbf{x}')$. Extensions to Bayesian and universal kriging predictors are straight forward. For more details consult Abrahamsen (1994*a*). An example of predicting a subsurface is found in Renard & Ruffo (1993), but they do not use the travel times nor do they compute the prediction error.

Example 3.3 Consider a standardized Gaussian random function with the exponential correlation function having roughness parameter $\nu = 2$. This function has been observed at three locations. Figure 3.3(a) show a prediction using these three data and Figure 3.3(c) show the associated prediction error. Assume further that the derivatives or dip has been observed at the same three locations. Figure 3.3(b) show a prediction using the three depth and three dip data. Figure 3.3(d) show the corresponding prediction errors. The figures mainly speaks for them selves but note in particular the effect of two near observations. In Figure 3.3(a) they mimic a combination of a depth and a dip data, whereas in Figure 3.3(b) they mimic a combination of a depth, a dip, and a second order derivative data. □

It is seen that observations of dip potentially have strong predictive power. It should however be mentioned that the correlation function used in this example describe the smoothest possible Gaussian random function. For real subsurfaces this is unrealistic so other correlation functions must be used. These will in comparison reduce the importance of dip information.

### Multiple Subsurface Model

In most applications of depth conversion several subsurfaces are considered. Thus, a consistent simultaneous description of several subsurfaces are called for. Some of these could be seismic reflectors whereas others are invisible on the seismics and must be modelled by using geological knowledge. More details are found in Abrahamsen (1993).

The main idea is to use a stochastic description of each interval above the deepest subsurface. The stochastic description of the thickness of, say interval $l$, will contain a term accounting for the trend and a residual including a standardized Gaussian random field:

$$\Delta Z_l(\mathbf{x}) = m_l(\mathbf{x}) + \sigma_l(\mathbf{x})\mathcal{E}_l(\mathbf{x}).$$

The model for the depth to subsurface $l$ is $Z_l(\mathbf{x}) = \sum_{i=1}^{l} \Delta Z_i(\mathbf{x})$. Now consider a subsurface $k$ and a deeper subsurface $l$, that is, $l > k$. These subsurfaces are correlated since $Z_l(\mathbf{x}) = Z_k(\mathbf{x}) + \sum_{i=k+1}^{l} \Delta Z_i(\mathbf{x})$, i.e., they have the trends and residuals describing $Z_k(\mathbf{x})$ in common. Thus, a prediction of subsurface $l$ must be conditioned on all well observations, including depth observations of subsurfaces above and below. A major advantage of this approach, compared to the traditional, where each interval is treated separately, is the ability to include deviating wells consistently, and moreover, the ability to compute realistic prediction errors for all subsurfaces.

Simple, Bayesian, and universal kriging predictors using all available data are established in the same way as before. An example illustrates the practical consequences:

EXAMPLE 3.4 Consider two seismic reflectors. Depth observations from three vertical wells and one deviating well are available. Figure 3.4(a) show predictions performed in the traditional way; The upper subsurface is conditioned on the four observations from this subsurface. Then the interval is predicted given the prediction of the upper subsurface and the observations of the interval thickness. Figure 3.4(b) show predictions of the two subsurfaces conditioned on all the data. Note in particular the ability to assess the uncertainty around the deviating well correctly. It is also seen that the upper subsurface is modified by the rightmost observation of the lower subsurface. □

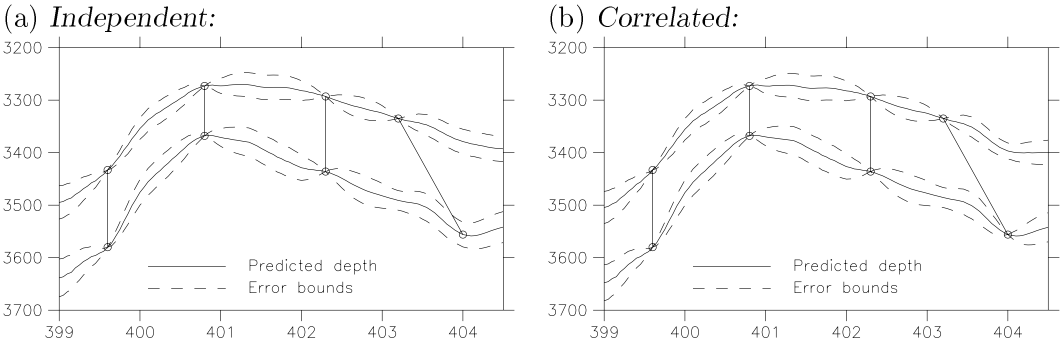


Figure 3.4: *Cross-section of prediction of two correlated subsurfaces. Figure* (a) *shows the traditional approach where the interval is predicted independently of the upper subsurface. Figure* (b) *shows predictions conditioned on all observations. Depth along the vertical axis and lateral position in kilometers along the horizontal axis.*

When considering seismic reflectors, the models should be simplified by ignoring the errors in the travel times above subsurface $l$. This approximation is excellent provided the velocity contrasts from layer to layer are moderate.

# 4 Parameter Inference

For Gaussian random functions three classes of model parameters are defined: (i) trend parameters, e.g. $\boldsymbol{\beta}$, (ii) scale of variance, $\sigma(\mathbf{x})$, and (iii) parameters in the correlation function, e.g. $\nu$ and $R$. Moreover, the general form of correlation function has to be specified. In this section only a brief outline of the problem of estimating these parameters from data will be outlined. Some important references will be given.

Inference about the parameters are complicated since the number of observations usually is small, the observations are occasionally inaccurate, they are usually correlated, and they are frequently preferentially sampled. Thus, the assumptions on which traditional inference is based are normally violated. Large parts of the inference theory of spatial models remains to be developed.

The geostatistical tradition has developed many rules of thumb for assessing parameter values. Journel & Huijbregts (1978), Isaaks & Srivastava (1989), and Cressie (1991) provides a good overview. Attempts on applying more classical statistical criteria in the estimation has also been made. Mardia & Marshall (1984) used a maximum likelihood estimator and reported bias problems although multi modality mixed with numerical instability could have been the cause. The sensibility of using maximum likelihood was challenged by Warnes & Ripley (1987). In Hjort & Omre (1993) a maximum pseudo-likelihood approach like Besag (1974), was suggested, and a maximum quasi-likelihood procedure was developed in some detail. An elegant procedure for testing the significance of parameters in the spatial correlation function was proposed by Switzer (1984).

The Bayesian formulation of the problem seems to simplify the inference of parameters. Kitanidis (1986), Omre (1987) Omre & Halvorsen (1989), Hjort & Omre (1993), and Handcock & Stein (1993) all developed posterior distributions for the model parameters. To reach the parameters in the spatial correlation function, numerical integration is occasionally required.

# 5 Simulation of Random Functions

The challenge is to generate a realization of $\{[Z(\mathbf{x})|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}]; \quad \mathbf{x} \in D\}$ according to the specified model. In practice the Gaussian random function has to be represented on a grid, $\mathfrak{L}_D$, covering $D$. Hence, the problem is to simulate a Gaussian random vector. This may appear simple since all Gaussian random vectors can be expressed as linear combinations of independent Gaussian random variables. The coefficients in the linear combination are easily obtainable from the cholesky decomposition of the covariance matrix of the random vector. However, the problem is the dimensionality since the grid, $\mathfrak{L}_D$, may contain $10^7$ elements. So in practice, algorithms based on approximations must be used.

Traditionally, conditional simulation has been based on the following decomposition:

$$\begin{aligned}&[Z(\mathbf{x})|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}]_i \\ &\qquad = \mathrm{E}\{Z(\mathbf{x})|\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\} + Y_i(\mathbf{x}) - \mathrm{E}\{Y(\mathbf{x})|\mathbf{Y}_{\text{obs}i}\}\,; \quad i = 1, \dots, s,\end{aligned}$$

where $Y_i(\mathbf{x})$ is a realization of a Gaussian random function $Y(\mathbf{x})$ having the same variance and correlation function as $Z(\mathbf{x})$, and $\mathbf{Y}_{\text{obs}i}$ being a vector of observations from $Y_i(\mathbf{x})$ taken at the locations of $\mathbf{Z}_{\text{obs}}$. For a known expectation, the simple kriging predictor and the associated prediction error are orthogonal. From this fact it is easily proven that the above expression give the correct distribution.

The decomposition above requires only a non-conditional Gaussian random function to be simulated, i.e. $Y_i(\mathbf{x})$. Many approaches for this have been proposed; turning bands (Matheron 1973, Christakos 1992), frequency-domain (Borgman, Taheri & Hagan 1984), and sequential algorithms (Omre, Sølna & Tjelmeland 1993). Another procedure for simulating conditional Gaussian random functions on large lattices follows from the factorization of the multi Gaussian pdf into univariate conditional Gaussian pdf's. A conditional simulation algorithm using this is described by Gómez-Hernández & Journel (1993).

# 6 Closing Remarks

The methods described have been successfully applied to several case studies. They are mainly extensions to methods widely used within the petroleum industry.

There are several advantages of using a formalized stochastic model for describing geological subsurfaces:

- Model assumptions are fully specified so results can be reproduced and consequences of the model can be checked.
- Different types of data are included consistently adhering to varying precision.
- Prediction errors take into account the complicated relationships between different subsurfaces and interval velocity fields. Thus, prediction errors are correctly calculated and they are realistic provided the specified model is realistic.
- Introducing new observations requires no adjustments of the models, — the conditional model will automatically honour new observations.
- Observations from deviating wells and interval velocity observations are handled correctly according to the model.

Moreover, using Gaussian random functions and linear trend models give efficient calculations since the conditioning is done analytically. Using non-linear trend models requires numerical optimization whereas using a non-Gaussian residual could imply an optimization criteria different from the least squares criteria implicitly used in the previous sections. The conditional expectations and variances would be replaced by numerical integrals that could be evaluated at the cost of extensive computer resources.

In the search for a more accurate description of natural phenomena, improved understanding of the physics governing the relationships between measurements and the phenomena itself is the key. Stochastic models should ideally describe a small gap between the state of the art deterministic description and the phenomena itself. Thus, a future challenge is to include more precise deterministic descriptions of the phenomena to minimize the stochastic parts of the model. The role of the stochastic part of the model is twofold: it reduces uncertainty by prescribing optimal ways of estimating model properties, and it quantifies the uncertainties so that financial risks can be studied.

# Acknowledgements

The work of Petter Abrahamsen was supported by a research fellowship from The Research Council of Norway. The authors benefitted significantly from facilities and computer programs at Norwegian Computing Center.

# References

Abrahamsen, P. (1993), Bayesian kriging for seismic depth conversion of a multi-layer reservoir, *in* Soares (1993), p. 1088.

Abrahamsen, P. (1994*a*), Kriging derivatives, 25 pp., Preliminary, planned for submission to *Biometrika*.

Abrahamsen, P. (1994*b*), A review of Gaussian random fields and correlation functions, Report 878, Norwegian Computing Center, P.O.Box 114 Blindern, N-0314 Oslo, Norway, 57 pp.

Adler, R. J. (1981), *The Geometry of Random Fields*, John Wiley & Sons, New York, 280 pp.

Adler, R. J. (1990), An introduction to continuity, extrema, and related topics for general Gaussian processes, Lecture notes—monograph series 12, Institute of Mathematical Statistics, Hayward, California, 160 pp.

Besag, J. (1974), 'Spatial interaction and the statistical analysis of lattice systems [with discussion]', *J. Roy. Statist. Soc. Ser. B* **36**, 192–236.

Borgman, L., Taheri, M. & Hagan, R. (1984), Three-dimensional, frequency-domain simulation of geological variables, *in* Verly, David, Journel & Marechal (1984), pp. 517–541. 2 volumes.

Christakos, G. (1992), *Random Field Models in Earth Sciences*, Academic Press Inc., San Diego, 474 pp.

Cramér, H. & Leadbetter, M. R. (1967), *Stationary and Related Stochastic Processes*, John Wiley & Sons, New York, 348 pp.

Cressie, N. (1990), 'The origins of kriging', *Math. Geol.* **22**(3), 239–252.

Cressie, N. (1991), *Statistics for Spatial Data*, John Wiley & Sons, New York, 900 pp.

Dimitrakopoulos, R., ed. (1994), *Geostatistics for the Next Century*, Vol. 6 of *Quantitative Geology and Geostatistics*, proceedings from 'An International Forum in Honour of Michel David's Contribution to Geostatistics', Montreal 1993, Kluwer Academic Publishers, Dordrecht, 497 pp.

Doob, J. L. (1953), *Stochastic Processes*, John Wiley & Sons, New York, 654 pp.

Gómez-Hernández, J. J. & Journel, A. G. (1993), Joint sequential simulation of multiGaussian fields, *in* Soares (1993), pp. 85–94.

Handcock, M. S. & Stein, M. L. (1993), 'A Bayesian analysis of kriging', *Technometrics* **35**(4), 403–410. A preliminary version found in Dimitrakopoulos (1994).

Hjort, N. L. & Omre, H. (1993), Topics in spatial statistics, Report 871, Norwegian Computing Center, P.O.Box 114 Blindern, N-0314 Oslo, Norway, 45 pp. Invited lecture at Nordic Conf. Math. Statist., Odense, June 1990. To appear (with discussion contributions and rejoinder) in 1995 in Scand. J. Statist.

Isaaks, E. H. & Srivastava, R. M. (1989), *Applied Geostatistics*, Oxford University Press Inc., New York, 561 pp.

Journel, A. G. (1974), 'Geostatistics for conditional simulation of ore bodies', *Economic Geol.* **69**, 673–687.

Journel, A. G. (1989), *Fundamentals of Geostatistics in Five Lessons*, Vol. 8 of *Short Course in Geology*, American Geophysical Union, Wahington DC, 40 pp. Presented at the 28th International Geological Congress, Washington DC.

Journel, A. G. & Huijbregts, C. J. (1978), *Mining Geostatistics*, Academic Press Inc., London, 600 pp.

Kitanidis, P. K. (1986), 'Parameter uncertainty in estimation of spatial functions: Bayesian analysis', *Water Resour. Res.* **22**(4), 499–507.

Mardia, K. V. & Marshall, R. J. (1984), 'Maximum likelihood for models for residual covariance in spatial regression', *Biometrika* **71**(1), 135–146.

Matérn, B. (1986), *Spatial Variation*, 2nd edn, Springer-Verlag Inc., Berlin, 151 pp. Fist edition published in 1960 by the Swedish Forest Research Institute.

Matheron, G. (1973), 'The intrinsic random functions and their applications', *Advances in Appl. Probab.* **5**, 439–468.

Matheron, G. (1976), A simple substitute for conditional expectation: disjunctive kriging, *in* M. Guarascio, C. J. Huijbregts & M. David, eds, 'Advanced Geostatistics in the Mining Industry', proceedings from '1st International Geostatistics Congress', Rome Italy, 1975, D. Reidel Publishing Company, Dordrecht, Netherlands, pp. 221–236.

Omre, H. (1987), 'Bayesian kriging—merging observations and qualified guesses in kriging', *Math. Geol.* **19**(1), 25–39.

Omre, H. & Halvorsen, K. B. (1989), 'The Bayesian bridge between simple and universal kriging', *Math. Geol.* **21**(7), 767–786.

Omre, H., Sølna, K. & Tjelmeland, H. (1993), Simulation of random functions on large lattices, *in* Soares (1993), pp. 179–199.

Renard, D. & Ruffo, P. (1993), Depth, dip and gradient, *in* Soares (1993), pp. 167–178.

Ripley, B. D. (1981), *Spatial Statistics*, John Wiley & Sons, New York, 252 pp.

Ripley, B. D. (1988), *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge, 148 pp.

Sacks, J., Welch, W. J., Mitchell, T. J. & Wynn, H. P. (1989), 'Computer experiments', *Statistical Science* **4**(4), 409–423.

Soares, A., ed. (1993), *Geostatistics Tróia '92*, Vol. 5 of *Quantitative Geology and Geostatistics*, proceedings from '4th International Geostatistics Congress', Tróia Portugal, 1992, Kluwer Academic Publishers, Dordrecht, 1088 pp. 2 volumes.

Stoyan, D., Kendall, W. S. & Mecke, J. (1987), *Stochastic Geometry and its Applications*, John Wiley & Sons, New York.

Switzer, P. (1984), Inference for spatial autocorrelation functions, *in* Verly et al. (1984), pp. 127–140. 2 volumes.

Vanmarcke, E. (1983), *Random Fields: Analysis and Synthesis*, The MIT Press, Cambridge, Massachusetts, 382 pp.

Verly, G., David, M., Journel, A. G. & Marechal, A., eds (1984), *Geostatistics for Natural Resource Characterization — Part 1/2*, Vol. 122 of *NATO ASI Series C*, proceedings from 'NATO ASI on Geostatistics for Natural Resource Characterization', Stanford Sierra Lodge, Lake Tahoe, California 1983, D. Reidel Publishing Company, Dordrecht, 1092 pp. 2 volumes.

Warnes, J. J. & Ripley, B. D. (1987), 'Problems with likelihood estimation of covariance functions of spatial Gaussian processes', *Biometrika* **74**(3), 640–642.

Yaglom, A. M. (1986*a*), *Correlation Theory of Stationary and Related Random Functions I: Basic Results*, Springer Series in Statistics, Springer-Verlag Inc., New York, 526 pp.

Yaglom, A. M. (1986*b*), *Correlation Theory of Stationary and Related Random Functions II: Supplementary Notes and References*, Springer Series in Statistics, Springer-Verlag Inc., New York, 258 pp.