

Kriging with Inequality Constraints¹

Petter Abrahamsen^{2,3} and Fred Espen Benth²

A Gaussian random field with an unknown linear trend for the mean is considered. Methods for obtaining the distribution of the trend coefficients given exact data and inequality constraints are established. Moreover, the conditional distribution for the random field at any location is calculated so that predictions using e.g. the expectation, the mode, or the median can be evaluated and prediction error estimates using quantiles or variance can be obtained. Conditional simulation techniques are also provided.

KEY WORDS: Bayesian kriging, Data Augmentation Algorithm, Gaussian random field, fixed point iterations.

INTRODUCTION

Consider the problem of mapping a random field $\{Z(\mathbf{x}); \mathbf{x} \in \mathbb{R}^d\}$ where the mean is unknown. Linear kriging techniques are widely used for this purpose but they do not allow the use of inequality constraints. The objective of this paper is therefore to suggest a method for predicting and simulating a Gaussian random field given inequality constraints where the value of $Z(\mathbf{x})$ is known to belong to a set of values $B_{\mathbf{x}}$, i.e., $Z(\mathbf{x}) \in B_{\mathbf{x}}$. The set $B_{\mathbf{x}}$ can be an interval, a one sided constraint, or even any Borel set.

Methods for predictions with associated prediction errors and conditional simulations will be given. The approach—first presented in Abrahamsen and Benth (1998)—adapts the Data Augmentation Algorithm (Tanner and Wong, 1987; Tanner, 1993), which is a Monte Carlo technique for finding the fixed point of an integral operator. The Data Augmentation Algorithm is efficient for calculating the distributions of trends and $Z(\mathbf{x})$ at a small number of locations. However, calculating the distributions of $Z(\mathbf{x})$ at every node in a large grid will require too much computer resources. Therefore, an approximation for the expectation and

¹Received 29 February 2000; accepted 6 September 2000.

²Norwegian Computing Center, P. O. Box 114 Blindern, N-0314 Oslo, Norway.

³Present address: Roxar Software Solution AS, P. O. Box 165 Skøyen, N-0212 Oslo, Norway, e-mail: Petter.Abrahamsen@roxar.com

variance of $Z(\mathbf{x})$ is also provided. The approach is illustrated by a small synthetic one-dimensional (1D) example and a two-dimensional (2D) example for mapping a geological horizon using observations from vertical and horizontal bore-holes.

Several approaches to include interval constraints have been proposed in the literature. Indicator kriging approaches include inequality constraints in a natural way (Journel, 1986). Dubrule and Kostov (1986) and Kostov and Dubrule (1986) proposed a solution to fit a surface to both exact data and inequality constraints using the dual kriging formalism and quadratic programming. The approach finds the minimum number of violated constraints and use their bounds as exact data. This method has been adapted for larger data sets by Langlais (1989). However, citing Journel (1986, p. 273): "Of course, one may question a solution consisting of replacing inequality data by hard data identified to the bounds." The dual kriging formalism does not provide prediction errors. Diamond (1988) develops interval kriging where interval valued data are the natural observations. Freulon and de Fouquet (1993) use a rejection sampling technique for simulating a Gaussian field conditioned on a small number of inequality constraints and a Markov chain Monte Carlo approach when the number of constraints is large. Stein (1992) also use a Monte Carlo technique to evaluate high dimensional integrals to obtain posterior distributions. Choi, Christakos, and Serre (1998) have explored the effects of inequality constraints using the so-called Bayesian maximum entropy analysis and Militino and Ugarte (1999) have used the Expectation-maximization (EM) algorithm for predicting uranium grade at censored locations.

NOTATION

We consider a random field Z on \mathbb{R}^d with a linear trend:

$$Z(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\boldsymbol{\beta} + \epsilon(\mathbf{x}); \quad \mathbf{x} \in \mathbb{R}^d,$$

where $\mathbf{f}'(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_P(\mathbf{x})]$ are P known functions and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_P]'$ are P coefficients. The $\epsilon(\mathbf{x})$ is a zero mean Gaussian random field with known covariance function. Moreover, the P coefficients are assumed to have a prior multinormal distribution:

$$\boldsymbol{\beta} \sim \mathcal{N}_P(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{1}$$

Thus, Z is a Gaussian random field with expectation and covariance:

$$\begin{aligned} E\{Z(\mathbf{x})\} &= \mathbf{f}'(\mathbf{x})\boldsymbol{\mu}_0 \\ \text{Cov}\{Z(\mathbf{x}), Z(\mathbf{x}')\} &= \mathbf{f}'(\mathbf{x})\boldsymbol{\Sigma}_0\mathbf{f}(\mathbf{x}') + \text{Cov}\{\epsilon(\mathbf{x}), \epsilon(\mathbf{x}')\} \end{aligned}$$

Data

Let N be exact observations of $Z(\mathbf{x})$ given by

$$\mathbf{Z}^e = \begin{bmatrix} Z(\mathbf{x}_1^e) \\ \vdots \\ Z(\mathbf{x}_N^e) \end{bmatrix} = \begin{bmatrix} z(\mathbf{x}_1^e) \\ \vdots \\ z(\mathbf{x}_N^e) \end{bmatrix}$$

Furthermore, we assume that there are M inequality constraints on $Z(\mathbf{x})$:

$$\mathbf{Z}^i = \begin{bmatrix} Z(\mathbf{x}_1^i) \\ \vdots \\ Z(\mathbf{x}_M^i) \end{bmatrix} \in \begin{bmatrix} B_1 \\ \vdots \\ B_M \end{bmatrix}$$

where B_1, \dots, B_M are intervals on the real line, or more general, Borel subsets of \mathbb{R} . The whole set of inequality constraints are denoted $\mathbf{B}^i = B_1 \times \dots \times B_M$, so that \mathbf{B}^i is a Cartesian product of M intervals, or more general, \mathbf{B}^i could be any Borel set in \mathbb{R}^M . The inequality constraint on \mathbf{Z}^i is written $\mathbf{Z}^i \in \mathbf{B}^i$.

In the following, several probability density functions (pdf's) are considered. For a general (non-normal) pdf the symbol $f(\cdot)$ is used, for normal densities $\varphi(\cdot)$ is used, and truncated normal densities are denoted $\bar{\varphi}(\cdot)$. Note that $Z(\mathbf{x}), \mathbf{Z}^e, \mathbf{Z}^i$, and β have a joint multinormal distribution, i.e., $\varphi(z(\mathbf{x}), \mathbf{z}^e, \mathbf{z}^i, \beta)$, and that the conditional distribution $\varphi(z(\mathbf{x}) \mid \mathbf{z}^i, \mathbf{z}^e, \beta)$ is also normal. The pdf $\bar{\varphi}(z(\mathbf{x}), \mathbf{z}^i \mid \mathbf{B}^i) = \bar{\varphi}(z(\mathbf{x}), \mathbf{z}^i \mid \mathbf{B}^i)$ is truncated normal, where \mathbf{B}^i is used as shorthand for $\mathbf{z}^i \in \mathbf{B}^i$. The density $f(z(\mathbf{x}) \mid \mathbf{z}^e, \beta, \mathbf{B}^i)$ is neither normal nor truncated normal.

POSTERIOR DISTRIBUTIONS

Let us consider an arbitrary set of K locations $\{\mathbf{x}_1, \dots, \mathbf{x}_K\}$, and the random field at these locations organized as a vector:

$$\mathbf{Z} = [Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_K)]'$$

The objective is to find the posterior distribution of \mathbf{Z} given the exact data, $\mathbf{Z}^e = \mathbf{z}^e$, and the inequality constraints, $\mathbf{Z}^i \in \mathbf{B}^i$. Denote this density as $f(\mathbf{z} \mid \mathbf{z}^e, \mathbf{B}^i)$. Given this pdf, predictors using expectation, median, or mode can be evaluated and uncertainty measures such as variance and quantiles can be calculated. In particular, $z^*(\mathbf{x}) = E\{Z(\mathbf{x}) \mid \mathbf{z}^e, \mathbf{B}^i\}$ is the natural extension of the standard linear kriging predictor. Also of interest is to find the posterior distribution of β , denoted $f(\beta \mid \mathbf{z}^e, \mathbf{B}^i)$, so it is convenient to consider the joint pdf $f(\mathbf{z}, \beta \mid \mathbf{z}^e, \mathbf{B}^i)$.

Following Tanner and Wong (1987) and Tanner (1993), the posterior density can be expressed as the fixed point of an integral equation (see Appendix):

$$f(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{B}^i) = \int_{\mathbb{R}^P} \int_{\mathbb{R}^K} K(\mathbf{z}, \beta; \tilde{\mathbf{z}}, \tilde{\beta}) f(\tilde{\mathbf{z}}, \tilde{\beta} | \mathbf{z}^e, \mathbf{B}^i) d\tilde{\mathbf{z}} d\tilde{\beta} \quad (2a)$$

where the transition kernel is

$$K(\mathbf{z}, \beta; \tilde{\mathbf{z}}, \tilde{\beta}) = \int_{\mathbb{R}^M} \varphi(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{z}^i) \bar{\varphi}(\mathbf{z}^i | \mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\beta}, \mathbf{B}^i) d\mathbf{z}^i. \quad (2b)$$

The fixed point can be computed by iterating

$$f^{(n+1)}(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{B}^i) = \int_{\mathbb{R}^P} \int_{\mathbb{R}^K} K(\mathbf{z}, \beta; \tilde{\mathbf{z}}, \tilde{\beta}) f^{(n)}(\tilde{\mathbf{z}}, \tilde{\beta} | \mathbf{z}^e, \mathbf{B}^i) d\tilde{\mathbf{z}} d\tilde{\beta} \quad (3)$$

Convergence follows since the kernel satisfies the criteria in Tanner and Wong (1987).

Analytical integration of the integrals in the fixed point iterations is not possible and numerical integration will at best be inaccurate for higher dimensions. The following algorithm is a Monte Carlo evaluation of the integrals in (3) where approximate densities are represented by (large) samples.

Data Augmentation Algorithm. Given the approximation $f^{(n)}(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{B}^i)$:

- (a) Draw S samples, $\mathbf{z}_{(1)}^i, \dots, \mathbf{z}_{(S)}^i$ from $f^{(n)}(\mathbf{z}^i, \beta | \mathbf{z}^e, \mathbf{B}^i)$. This is done in two steps:
 - (a1) Draw $\tilde{\beta}_{(s)}$ from $f^{(n)}(\beta | \mathbf{z}^e, \mathbf{B}^i)$.
 - (a2) Draw $\mathbf{z}_{(s)}^i$ from $\bar{\varphi}(\mathbf{z}^i | \tilde{\beta}_{(s)}, \mathbf{z}^e, \mathbf{B}^i)$ by sampling from $\varphi(\mathbf{z}^i | \tilde{\beta}_{(s)}, \mathbf{z}^e)$ until $\mathbf{z}_{(s)}^i \in \mathbf{B}^i$.
- (b) Update the approximation to be

$$f^{(n+1)}(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{B}^i) = \frac{1}{S} \sum_{s=1}^S \varphi(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{z}_{(s)}^i) \quad (4)$$

where $\varphi(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{z}_{(s)}^i)$ are normal densities.

Step (a) generate “latent” inequality constraints given the exact data by sampling from the density

$$f^{(n)}(\mathbf{z}^i | \mathbf{z}^e, \mathbf{B}^i) = \int_{\mathbb{R}^P} \int_{\mathbb{R}^K} \bar{\varphi}(\mathbf{z}^i | \mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\beta}, \mathbf{B}^i) f^{(n)}(\tilde{\mathbf{z}}, \tilde{\beta} | \mathbf{z}^e, \mathbf{B}^i) d\tilde{\mathbf{z}} d\tilde{\beta}$$

The $\bar{\varphi}(\mathbf{z}^i | \mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\beta}, \mathbf{B}^i)$ part comes from (2b) and $f^{(n)}(\tilde{\mathbf{z}}, \tilde{\beta} | \mathbf{z}^e, \mathbf{B}^i)$ comes from (3).

Thus, Step (a) essentially performs two of the integrals in the fixed point iteration. However, $\tilde{\mathbf{z}}$ are dummy variables and can be ignored since

$$f^{(n)}(\mathbf{z}^i | \mathbf{z}^e, \mathbf{B}^i) = \int_{\mathbb{R}^p} \bar{\varphi}(\mathbf{z}^i | \mathbf{z}^e, \tilde{\beta}, \mathbf{B}^i) f^{(n)}(\tilde{\beta} | \mathbf{z}^e, \mathbf{B}^i) d\tilde{\beta}$$

In Step (a1) drawing β from $f^{(n)}(\beta | \mathbf{z}^e, \mathbf{B}^i)$ is done by drawing from one randomly selected distribution in the sum in (4); \mathbf{z} is only considered in the final iteration. Step (b) is an evaluation of the kernel (3) of the fixed point integral:

$$\int_{\mathbb{R}^M} \varphi(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{z}^i) \bar{\varphi}(\mathbf{z}^i | \mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\beta}, \mathbf{B}^i) d\mathbf{z}^i$$

where the integral is replaced by the sum over \mathbf{z}^i 's drawn from $f^{(n)}(\mathbf{z}^i | \mathbf{z}^e, \mathbf{B}^i)$ in Step (a).

The efficiency of the algorithm depends on the rate of rejections in Step (a2). The rejection rate is large when either $f^{(n)}(\mathbf{z}^i | \mathbf{z}^e, \mathbf{B}^i)$ is a poor approximation to $f(\mathbf{z}^i | \mathbf{z}^e, \mathbf{B}^i)$ or when the constraint $\mathbf{z}^i \in \mathbf{B}^i$ is very restrictive, that is, when $\bar{\varphi}(\mathbf{z}^i | \mathbf{z}^e, \mathbf{B}^i)$ is very different from $\varphi(\mathbf{z}^i | \mathbf{z}^e)$. The first problem is solved by starting the algorithm using a small S in the initial iterations and then increasing the number as $f^{(n)}(\mathbf{z}^i | \mathbf{z}^e, \mathbf{B}^i)$ approaches $f(\mathbf{z}^i | \mathbf{z}^e, \mathbf{B}^i)$. Our experience is that starting with $S = 2$ at the first iteration and then doubling S for the 10–13 first iterations is efficient. The problem of very restrictive constraints must be solved by implementing smart rejection sampling techniques.

The Data Augmentation Algorithm needs an initial distribution of β . In the examples below we have used $f^{(0)}(\beta | \mathbf{z}^e, \mathbf{B}^i) = \varphi(\beta | \mathbf{z}^e)$ but more sophisticated choices including some inequality constraint information are possible.

Posterior Distribution of β

If the objective is limited to obtain moments for the distribution of β , the Data Augmentation Algorithm simplifies slightly. Step (a) is exactly the same but in Step (b) any reference to \mathbf{z} can be removed so (4) is replaced by

$$f^{(n+1)}(\beta | \mathbf{z}^e, \mathbf{B}^i) = \frac{1}{S} \sum_{s=1}^S \varphi(\beta | \mathbf{z}^e, \mathbf{z}_{(s)}^i)$$

Note that this is the algorithm we would obtain if we removed all references to \mathbf{z} and $\tilde{\mathbf{z}}$ in (2a) to (3) and in the Data Augmentation Algorithm.

Posterior Distribution for Z

Notice that \mathbf{z} -samples from previous iterations are never used in later iterations in the Data Augmentation Algorithm. Therefore, to obtain the posterior distribution $f(\mathbf{z} | \mathbf{z}^e, \mathbf{B}^i)$, only β is sampled in all but the last iteration. After $f^{(n)}(\beta | \mathbf{z}^e, \mathbf{B}^i)$ has converged to the necessary precision, \mathbf{z} and β can be drawn in the last iteration of the Data Augmentation Algorithm.

Usually we are interested in prediction of Z at a single location, \mathbf{x} . So replacing \mathbf{z} by $z(\mathbf{x})$ in (4) and simply ignoring the β samples in the final iteration gives

$$f^{(n)}(z(\mathbf{x}) | \mathbf{z}^e, \mathbf{B}^i) = \frac{1}{S} \sum_{s=1}^S \varphi(z(\mathbf{x}) | \mathbf{z}^e, \mathbf{z}_{(s)}^i)$$

An approximation to the conditional expectation at \mathbf{x} is therefore

$$z^*(\mathbf{x}) = E\{Z(\mathbf{x}) | \mathbf{z}^e, \mathbf{B}^i\} \approx \frac{1}{S} \sum_{s=1}^S E\{Z(\mathbf{x}) | \mathbf{z}^e, \mathbf{z}_{(s)}^i\}$$

Multimodal distributions might give $z^*(\mathbf{x}_n^i) \notin B_n^i$ so the mode or some other statistic should be used for the predictor. This is illustrated in Example B in the following section.

Ignoring β in the Data Augmentation Algorithm

To obtain $f(\beta | \mathbf{z}^e, \mathbf{B}^i)$, all references to \mathbf{z} can be ignored in (2a) to (3) and in the Data Augmentation Algorithm. Similarly, it is possible to ignore β in (2a) to (3) and obtain an alternative Data Augmentation Algorithm for calculating $f(\mathbf{z} | \mathbf{z}^e, \mathbf{B}^i)$. Step (a2) is then removed, together with all references to β and $\tilde{\beta}_{(s)}$. Our experience, however, has shown that it is advantageous to include β in the algorithm, since the rejection rate in Step (a2) is decreased significantly. The samples $\tilde{\beta}_{(s)}$ contain information about the inequality constraints which makes $\phi(\mathbf{z}^i | \tilde{\beta}_{(s)}, \mathbf{z}^e)$ a better approximation to $\bar{\phi}(\mathbf{z}^i | \tilde{\beta}_{(s)}, \mathbf{z}^e, \mathbf{B}^i)$ than $\phi(\mathbf{z}^i | \mathbf{z}^e)$. This argument holds only if the cost of drawing $\tilde{\beta}_{(s)}$ does not out-balance the increased acceptance rate in Step (a2).

ONE-DIMENSIONAL EXAMPLES

To illustrate how inequality constraints influence the prediction of random fields, we used the Data Augmentation Algorithm on two simple one-dimensional examples. The examples will show that inequality data may produce highly

non-Gaussian posterior distributions, unlike kriging with equality data. In fact, distributions may become bimodal (or more generally, multimodal). We shall also through these examples see that the standard choice of the expectation as a predictor may be questioned in the context of inequality kriging, since it may violate the constraints. Because of the simplicity of the examples, we can discuss the convergence properties of the Data Augmentation Algorithm.

Let us consider the linear regression model

$$Z(x) = \beta_0 + \beta_1 x + \epsilon(x); \quad x \in \mathbb{R}$$

where $\epsilon(x)$ is a Gaussian random field with zero mean, unit variance, and a spherical correlation function with range 4. The prior distribution for the coefficients is $\beta \sim \mathcal{N}_2(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10^2 & 0 \\ 0 & 2 \end{bmatrix})$. The prior variances are two orders of magnitude larger than the posterior variances obtained below and is almost “noninformative.” We assume three equality constraints at the locations $x = 2.5, 5.5,$ and 8 are given, with values $Z(2.5) = 0.1, Z(5.5) = 0.9,$ and $Z(8) = 0.6$. At the locations $x = 0$ and $x = 10,$ we suppose in Example A that we have the inequality data $Z(0) < 0$ and $Z(10) < 1$. Example B substitutes the upper bound on the field $Z(x)$ at location $x = 10$ with the constraint $Z(10) \notin [1, 3.5]$, that is, the field $Z(x)$ is not allowed to pass through the interval $[1, 3.5]$ at $x = 10$. We now discuss the results from Examples A and B in more detail.

Predicting β

Figure 1 shows the regression lines obtained in Examples A. The marginal posterior probability densities for the intercept β_0 and the slope β_1 are shown in Figure 2. The pdf’s are obtained using kernel smoothing of $8192 (=2^{13})$ samples. The figures illustrate that both inequality constraints alter the expectations and reduce the variances. It is also seen that the conditional distributions for the intercept and the slope are very close to being Gaussian. In all examples we have tested using closed or one-sided intervals as constraints, the conditional distributions for β looks Gaussian. Example B, however, gives a highly non-Gaussian result.

Figure 3 shows the regression lines and data when the inequality constraint at $x = 10$ is changed so that $Z(10) \notin [1, 3.5]$. Figure 4 shows the corresponding posterior marginal probability distributions for intercept and slope. Observe that the regression lines pass through the “illegal” interval and the distribution for the slope becomes bimodal.

Predicting $Z(x)$

Figure 5 illustrates the marginal distributions of $Z(x) | \mathbf{z}^e, \mathbf{B}^i$ by showing quantiles and expectations as a function of x for Examples A and B. The statistics

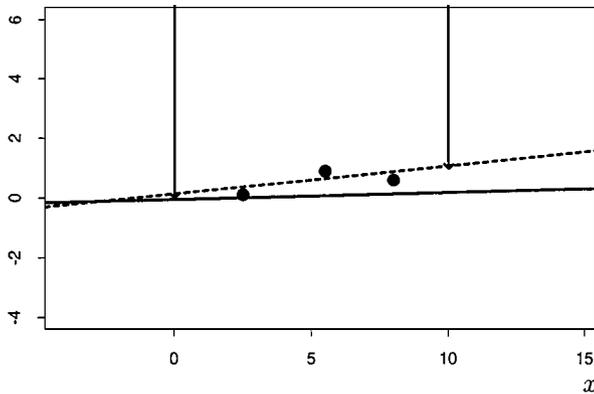


Figure 1. Trends using coefficients $E\{\beta | z^c, \mathbf{B}^1\}$ (solid line) and $E\{\beta | z^c\}$ (dashed line) for Example A. Exact data are plotted as dots and interval constraints as arrows. The allowed region is outside the arrows. Vertical axis is z .

are based on 8,192 samples of $Z(x)$. Note that the predictor $E\{Z(x) | z^c, \mathbf{B}^1\}$ passes through the “illegal” interval at $x = 10$ in Example B. The median however, honors the interval constraint. In Example A the conditional expectation behaves nicely, but the median is slightly above since the distributions are skew. The skewness is caused by the inequality constraint and is not a random effect coming from the Monte Carlo sampling.

Convergence Rates

There are two sources of errors in the algorithm: The number of fixed point iterations are limited and the number of samples S of β and $Z(x)$ in the final iterations are also limited. The two examples above were obtained by using $S = 2$ in the initial fixed point iteration and then doubling S at each iteration until $S = 8192$ at iteration 13. From then on S was kept unchanged. Figure 6 shows how the distributions of the slope evolves as the number of fixed point iterations increase. The convergence for Example A is rapid and after 10 iterations ($S = 1024$) the levels seem to stabilize. Increasing S after this is mainly to reduce Monte Carlo noise. In Example B, however, convergence is very slow. Forty iterations are shown on the plot Figure 6 and the levels does not stabilise until at least 35 fixed point iterations have been run. The problem seems to be that the number of β_1 samples from the highest mode is underestimated during the initial iterations.

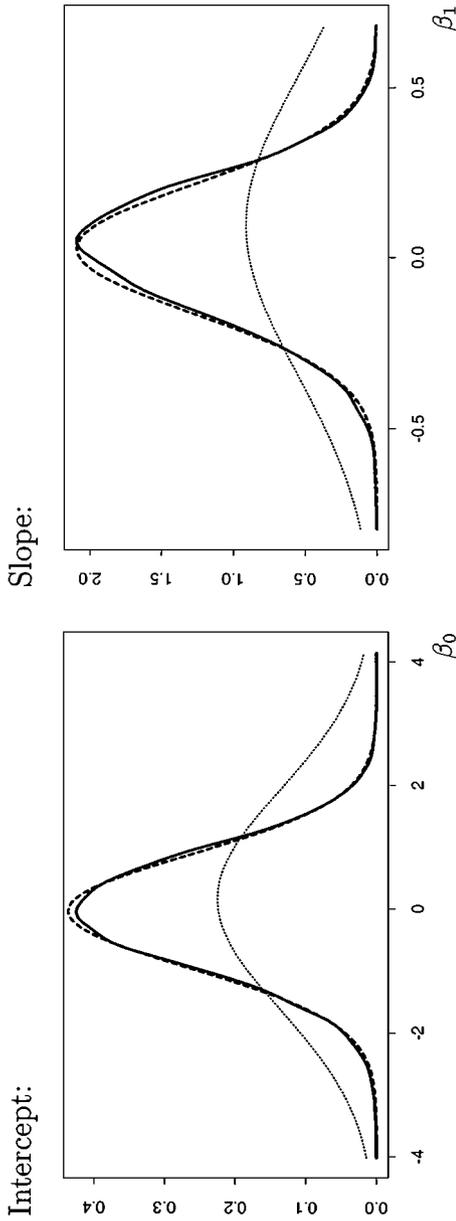


Figure 2. Probability densities for intercept and slope for Example A. Solid lines are posterior pdf's while dashed lines are normal distributions with identical expectations and variances. The thin dotted curves are the posterior normal distributions given only exact data.

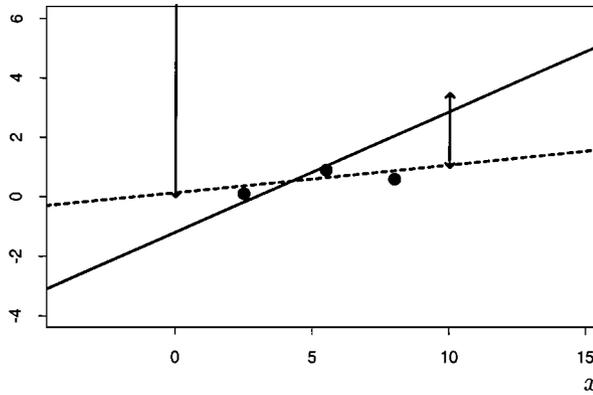


Figure 3. Trends using coefficients $E\{\beta | z^e, \mathbf{B}^i\}$ (solid line) and $E\{\beta | z^e\}$ (dashed line) for Example B. Exact data are plotted as dots and interval constraints as arrows. The allowed region is outside the arrows. Vertical axis is z .

A FAST APPROXIMATION

Evaluating the conditional expectation using the Data Augmentation Algorithm for every node in a large grid is very time consuming so an approximation is called for.

Let us assume that the inequality constraints B_1, \dots, B_M are such that the posterior pdf's for $Z(\mathbf{x})$ are close to being normal. Then the conditional expectation, $z^*(\mathbf{x}) = E\{Z(\mathbf{x}) | \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}^i \in \mathbf{B}^i\}$, can be used as a predictor. The idea is to replace the constraints $\mathbf{Z}^i \in \mathbf{B}^i$ by artificial “exact” data with multinormal “measurement errors” at the locations of the inequality constraints, that is,

$$\mathbf{Z}^i \in \mathbf{B}^i \rightarrow \mathbf{Z}_d^i = \mathbf{Z}^i + \epsilon_{\text{err}}^i = \mathbf{z}_d^i$$

with a multinormal zero mean “measurement error”: $\epsilon_{\text{err}}^i \sim \mathcal{N}_M(0, \Sigma_{\text{err}}^i)$. The “measurement error” is independent of β and $\epsilon(\mathbf{x})$. The posterior distribution, $\varphi(z(\mathbf{x}) | \mathbf{z}^e, \mathbf{z}^i)$, is normal since the “measurement error” is multinormal and efficient (linear) kriging methods work. In particular $E\{Z(\mathbf{x}) | \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\}$ and $\text{Var}\{Z(\mathbf{x}) | \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\}$ are given by standard formulas from multinormal analysis.

The problem is to find good values for artificial data \mathbf{z}_d^i , and the “measurement error” covariance matrix, Σ_{err}^i . We propose to select these such that the posterior expectations and covariances at the constrained locations are correct:

$$E\{\mathbf{Z}^i | \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\} \stackrel{!}{=} E\{\mathbf{Z}^i | \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}^i \in \mathbf{B}^i\} \tag{5}$$

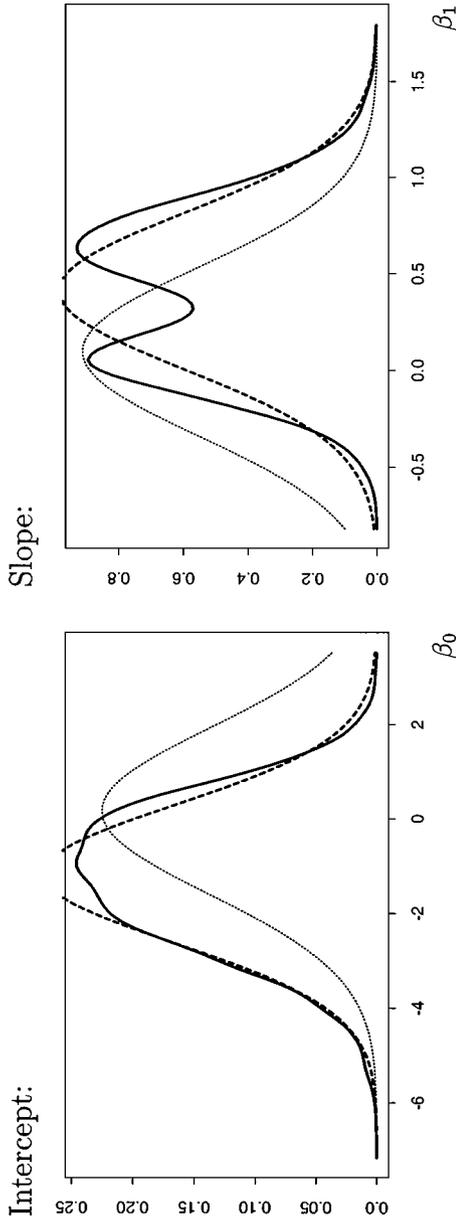


Figure 4. Probability densities for intercept and slope for Example B. Solid lines are posterior pdf's while dashed lines are normal distributions with identical expectations and variances. The thin dotted curves are the posterior normal distributions given only exact data.

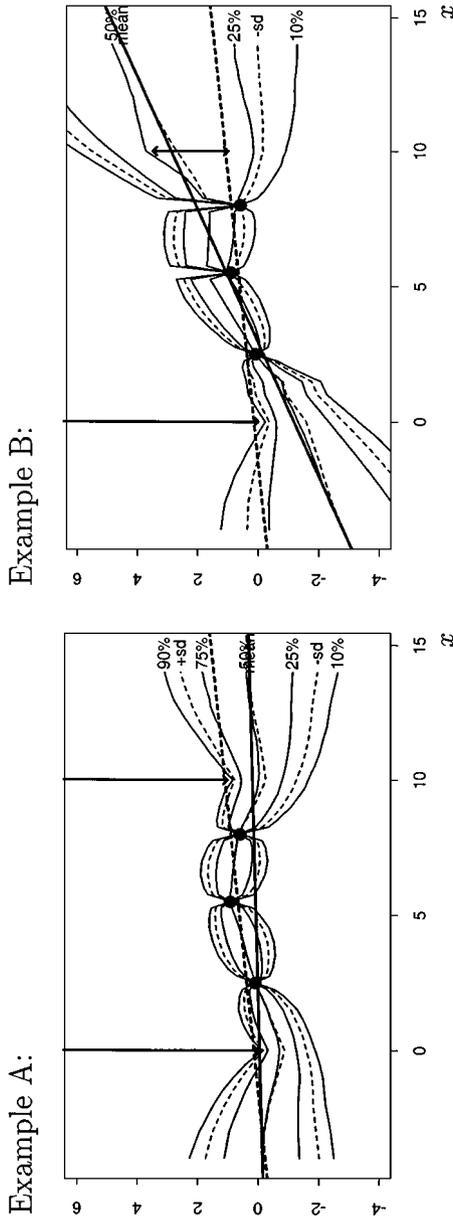


Figure 5. Quantiles of $Z(x) | z^c, B^i$ shown as solid lines. Expectation plus/minus standard error shown as dashed lines.

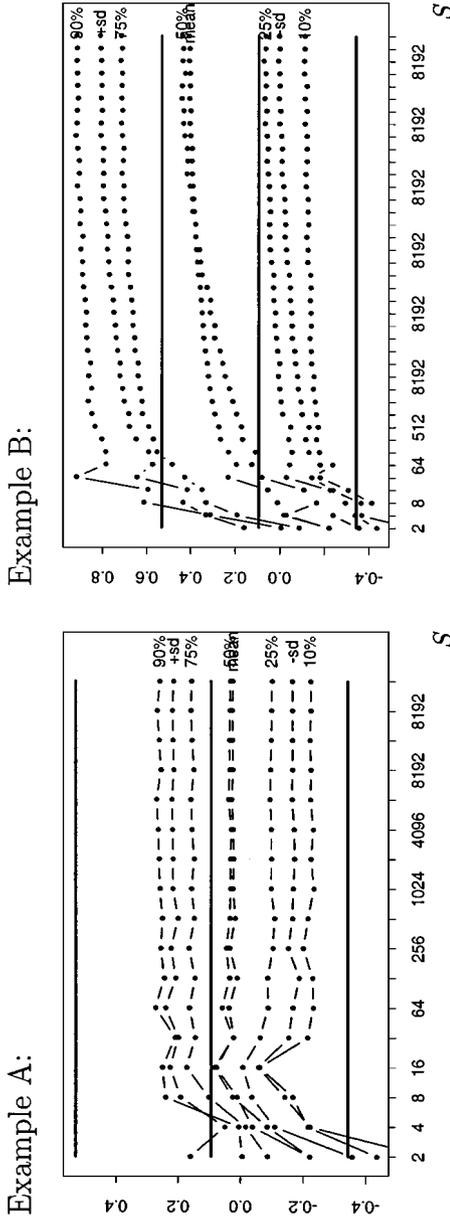


Figure 6. Quantiles and moments for $\beta_1 \mid z^c, B^i$ as a function of the number of fixed point iterations. The horizontal lines are expectation \pm standard deviation for $\beta_1 \mid z^c$.

$$\text{Var}\{\mathbf{Z}^i \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\} \stackrel{!}{=} \text{Var}\{\mathbf{Z}^i \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}^i \in \mathbf{B}^i\}. \tag{6}$$

The correct moments given interval constraints—that is, the right-hand sides—must be obtained using the Data Augmentation Algorithm. Hopefully, approximations

$$\begin{aligned} E\{Z(\mathbf{x}) \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\} &\approx E\{Z(\mathbf{x}) \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}^i \in \mathbf{B}^i\} \\ \text{Var}\{Z(\mathbf{x}) \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\} &\approx \text{Var}\{Z(\mathbf{x}) \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}^i \in \mathbf{B}^i\} \end{aligned}$$

for arbitrary locations \mathbf{x} , will have an acceptable small error since they are correct at the locations of exact data and inequality constraints.

It is possible to find closed forms for the artificial data vector \mathbf{z}_d^i and the error covariance matrix Σ_{err}^i satisfying requirements (5) and (6). Before expressions for these are given, some additional notation is introduced for convenience. We will need covariances between two arbitrary vectors containing values from the random field $Z(\mathbf{x})$:

$$\mathbf{C}^{uv} = \text{Cov}\{\mathbf{Z}^u, \mathbf{Z}^v\} = \mathbf{K}^{uv} + \mathbf{F}^u \Sigma_0 \mathbf{F}^{v'}$$

where $\mathbf{K}^{uv} = \text{Cov}\{\epsilon^u, \epsilon^v\}$, $F_{np}^u = f_p(\mathbf{x}_n^u)$, and Σ_0 is the prior covariance matrix for β ; see (1).

Consider the conditional expectation for $\mathbf{Z}' = [Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_K)]$ given exact data, $\mathbf{Z}^e = \mathbf{z}^e$, and artificial data, $\mathbf{Z}_d^i = \mathbf{z}_d^i$. From standard multinormal theory, conditional expectation and covariances are

$$E\{\mathbf{Z} \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\} = \mathbf{F}\boldsymbol{\mu}_0 + \begin{bmatrix} \mathbf{C}^e & \mathbf{C}^i \end{bmatrix} \begin{bmatrix} \mathbf{C}^{ee} & \mathbf{C}^{ei} \\ \mathbf{C}^{ie} & \mathbf{C}^{ii} + \Sigma_{\text{err}}^i \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{z}^e - \mathbf{F}^e \boldsymbol{\mu}_0 \\ \mathbf{z}_d^i - \mathbf{F}^i \boldsymbol{\mu}_0 \end{bmatrix} \tag{7}$$

$$\text{Var}\{\mathbf{Z} \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\} = \mathbf{C} - \begin{bmatrix} \mathbf{C}^e & \mathbf{C}^i \end{bmatrix} \begin{bmatrix} \mathbf{C}^{ee} & \mathbf{C}^{ie} \\ \mathbf{C}^{ei} & \mathbf{C}^{ii} + \Sigma_{\text{err}}^i \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C}^{e'} \\ \mathbf{C}^{i'} \end{bmatrix} \tag{8}$$

where $\mathbf{C} = \text{Var}\{\mathbf{Z}\}$ and $\mathbf{C}^u = \text{Cov}\{\mathbf{Z}, \mathbf{Z}^u\}$. Expression (7) is the Bayesian kriging predictor and (8) is the associated prediction error (Kitanidis, 1986, Omre, 1987).

By letting $\mathbf{Z} = \mathbf{Z}^i$ in (8) and solving for Σ_{err}^i gives

$$\Sigma_{\text{err}}^i = \left[\text{Var}\{\mathbf{Z}^i \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}_d^i = \mathbf{z}_d^i\}^{-1} - (\mathbf{C}^{ii} - \mathbf{C}^{ie} \mathbf{C}^{ee-1} \mathbf{C}^{ie'})^{-1} \right]^{-1}$$

This result can be confirmed using matrix algebra found in, e.g., Mardia, Kent and Bibby (1979, pp. 458–459). Introducing $\Sigma^i = \text{Var}\{\mathbf{Z}^i \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}^i \in \mathbf{B}^i\}$, and

$$\mathbf{C}^{ile} = \text{Var}\{\mathbf{Z}^i \mid \mathbf{Z}^e\} = \mathbf{C}^{ii} - \mathbf{C}^{ie} \mathbf{C}^{ee-1} \mathbf{C}^{ie'}$$

and imposing (6) gives

$$\Sigma_{\text{err}}^i = [\Sigma^{i-1} - \mathbf{C}^{ile-1}]^{-1} = \Sigma^i + \Sigma^i [\mathbf{C}^{ile} - \Sigma^i]^{-1} \Sigma^i \tag{9}$$

It can be seen that $\mathbf{C}^{ile} > \Sigma^i$ in the sense that $(\mathbf{C}^{ile} - \Sigma^i)$ must be positive definite for (9) to give a valid covariance matrix. This is reasonable since the constraints should add information and therefore reduce the uncertainty at the constrained locations. In the following, two examples of this problem is encountered and a remedy is prescribed in the section containing the case-study. Note however that the algebra leading to (9) does not require that $\mathbf{C}^{ile} - \Sigma^i$ is positive definite. It only requires that the inverse of $\mathbf{C}^{ile} - \Sigma^i$ exist. Moreover, if Σ_{err}^i fails to become a valid covariance matrix, (7) and (8) can still be used and will give meaningful results. However, the interpretation of \mathbf{z}_d^i as artificial data with error covariance Σ_{err}^i is no longer valid.

The next step is to solve (7) for \mathbf{z}_d^i and impose (5). Again for notational simplicity, $\mu^i = E\{\mathbf{Z}^i \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}^i \in \mathbf{B}^i\}$ is introduced, so that the solution can be written

$$\mathbf{z}_d^i = \mathbf{F}^i \mu_0 - \Sigma_{\text{err}}^i \mathbf{C}^{ile-1} \mathbf{C}^{ie} \mathbf{C}^{ee-1} (\mathbf{z}^e - \mathbf{F}^e \mu_0) + \Sigma_{\text{err}}^i \Sigma^{i-1} (\mu^i - \mathbf{F}^i \mu_0) \tag{10}$$

Inserting \mathbf{z}_d^i and Σ_{err}^i from (10) and (9) respectively in the predictor (7) and prediction variance (8) give the required properties. In most practical applications, \mathbf{Z} in (7) and (8) is replaced by a single variable $Z(\mathbf{x})$.

In some cases the condition $\mathbf{C}^{ile} > \Sigma^i$ fails so that Σ_{err}^i is not positive definite. Then, the interpretation of ϵ_{err}^i being a “measurement error” is not valid. However, using expressions (7) and (8) as approximations to (5) and (6) is still possible as long as $[\begin{smallmatrix} \mathbf{C}^{ce} & \mathbf{C}^{ie} \\ \mathbf{C}^{ci} & \mathbf{C}^{ii} + \Sigma_{\text{err}}^i \end{smallmatrix}]$ is positive definite.

Example

Consider the simple one-dimensional example illustrated in Figure 1. In Figure 7 the exact expectation, $E\{Z(\mathbf{x}) \mid \mathbf{Z}^e = \mathbf{z}^e, \mathbf{Z}^i \in \mathbf{B}^i\}$, is compared to the

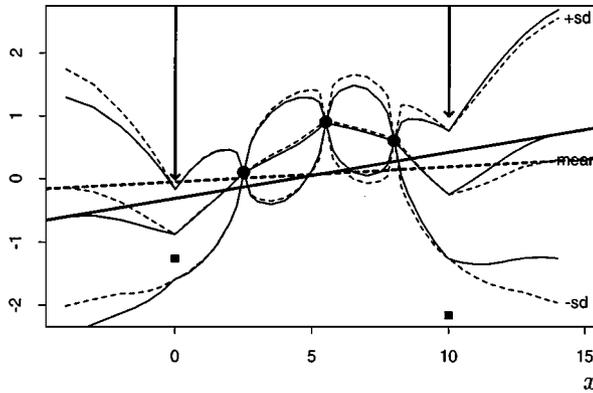


Figure 7. The approximated expectation \pm standard error are shown as solid lines. Expectation \pm standard error for $Z(x) \mid \mathbf{Z}^c = \mathbf{z}^c, \mathbf{Z}^i \in \mathbf{B}^i$ are shown as dashed lines. Corresponding linear trend lines are shown as straight dashed and solid lines respectively. The two artificial data \mathbf{z}_d^i are shown as squares.

approximation $E\{Z(\mathbf{x}) \mid \mathbf{Z}^c = \mathbf{z}^c, \mathbf{Z}_d^i = \mathbf{z}_d^i\}$. This figure shows that the approximate method does not obtain the correct trend since the expectation curves depart away from data locations.

Getting the Trends to Fit

To adjust the P trend coefficients, P additional artificial data are added in the following way

$$Z_{dM+p}^i = \beta_p + \epsilon_{errM+p}^i = z_{dM+p}^i; \quad p = 1, \dots, P$$

This means that $Z_{dM+p}^i; p = 1, \dots, P$ links to $Z(\mathbf{x})$ through β alone; all $M + P$ elements of ϵ_{err}^i are assumed independent of β and $\epsilon(\mathbf{x})$.

Additional requirements in (5) and (6) must be added to incorporate the trend parameters. This is done by replacing \mathbf{Z}^i by $[\mathbf{Z}^i \beta']$ in (5) and (6).

The arguments leading to the determination of the artificial data vector \mathbf{z}_d^i and the observation error covariance matrix Σ_{err}^i holds, but some care in expanding the different \mathbf{F} and \mathbf{C} matrices is necessary. For instance \mathbf{F}^i is replaced by $[\mathbf{F}^i \mathbf{I}_P]$, where \mathbf{I}_P is a P dimensional identity matrix. Similarly, covariances involving \mathbf{Z}_d^i must be replaced by enlarged matrices including contributions from the P new

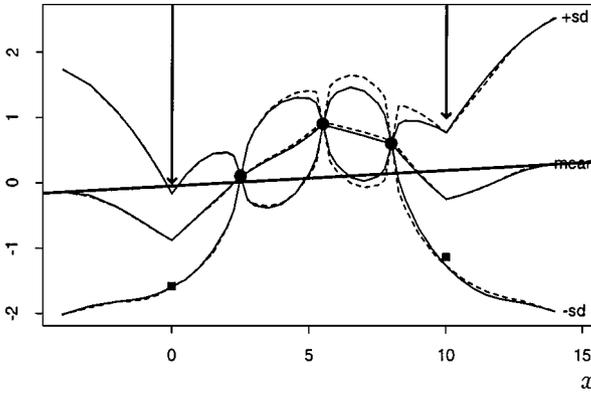


Figure 8. The approximated expectation \pm standard error using additional artificial data to fit the trend are shown as solid lines. Expectation \pm standard error for $Z(x) | Z^e = z^e, Z^i \in B^i$ are shown as dashed lines. Corresponding linear trend lines are shown as straight dashed and solid lines respectively. The two artificial data z_d^i are shown as squares.

elements:

$$C^i \rightarrow \begin{bmatrix} C^i \\ F \Sigma_0 \end{bmatrix}, \quad C^{ei} \rightarrow \begin{bmatrix} C^{ei} \\ F^e \Sigma_0 \end{bmatrix}, \quad \text{and} \quad C^{ii} \rightarrow \begin{bmatrix} C^{ii} & F^i \Sigma_0 \\ \Sigma_0 F^{i'} & \Sigma_0 \end{bmatrix}$$

Example with Fitted Trend

Once again consider the one-dimensional example illustrated in Figure 1. In Figure 8 the exact expectation $E\{Z(\mathbf{x}) | Z^e = z^e, Z^i \in B^i\}$ is compared to the approximation $E\{Z(\mathbf{x}) | Z^e = z^e, Z_d^i = z_d^i\}$, where two additional data are added to fit the trend. The figure shows that the trends are correct in this second attempt.

Problems with Σ_{err}^i

In the first approximation (see Fig. 7) the “measurement error” covariance matrix is

$$\Sigma_{err}^i = \begin{bmatrix} 2.37 & -0.01 \\ & 0.64 \end{bmatrix}$$

which is a covariance matrix. When introducing two additional data to adjust the trend, the error matrix becomes

$$\Sigma_{\text{err}}^i = \begin{bmatrix} 1.36 & 0.11 & -0.29 & -0.02 \\ & 0.68 & -0.07 & 0.02 \\ & & 3.35 & -0.32 \\ & & & -0.003 \end{bmatrix}$$

which is *not* a covariance matrix since the last “variance,” -0.003 , is negative. Despite this, the approximate predictor works in the sense that the algebra is valid and the result shown in Figure 8 is meaningful. However, this is a sign of problems that have to be handled. We present more on this in the full-scale case study in the next section.

A TWO-DIMENSIONAL CASE STUDY

The example in this section is an application of the approximation outlined in the previous section. A geological horizon is known exactly at seven locations (Fig. 9) where bore-holes have penetrated the horizon. Two of these bore-holes (26B and 27) proceed almost horizontally below the horizon toward the south. They provide the information that the horizon must be above these two well trajectories. The two well trajectories have been sampled at approximately 100-m intervals, which makes a total of 33 samples. This gives 7 “exact” data and 33 “inequality” constraints. The regular grid used to represent the surface consists of $61 \times 91 = 5,551$ grid nodes.

The model for the depth to the horizon is

$$Z(\mathbf{x}) = \beta_1 t(\mathbf{x}) + \beta_2 t(\mathbf{x})(t(\mathbf{x}) - 1.17 \text{ sec}) + \epsilon(\mathbf{x}); \quad \mathbf{x} \in \mathbb{R}^2,$$

where $t(\mathbf{x})$ is the one-way seismic travel times to the horizon. A map of the travel times can be seen in Figure 9. The linear regression model for the trend is constructed such that the seismic sound velocity model is

$$v(\mathbf{x}) = \beta_1 + \beta_2(t(\mathbf{x}) - 1.17 \text{ sec})$$

The approximate average of the $t(\mathbf{x})$ -map is 1.17 sec, and is subtracted to reduce

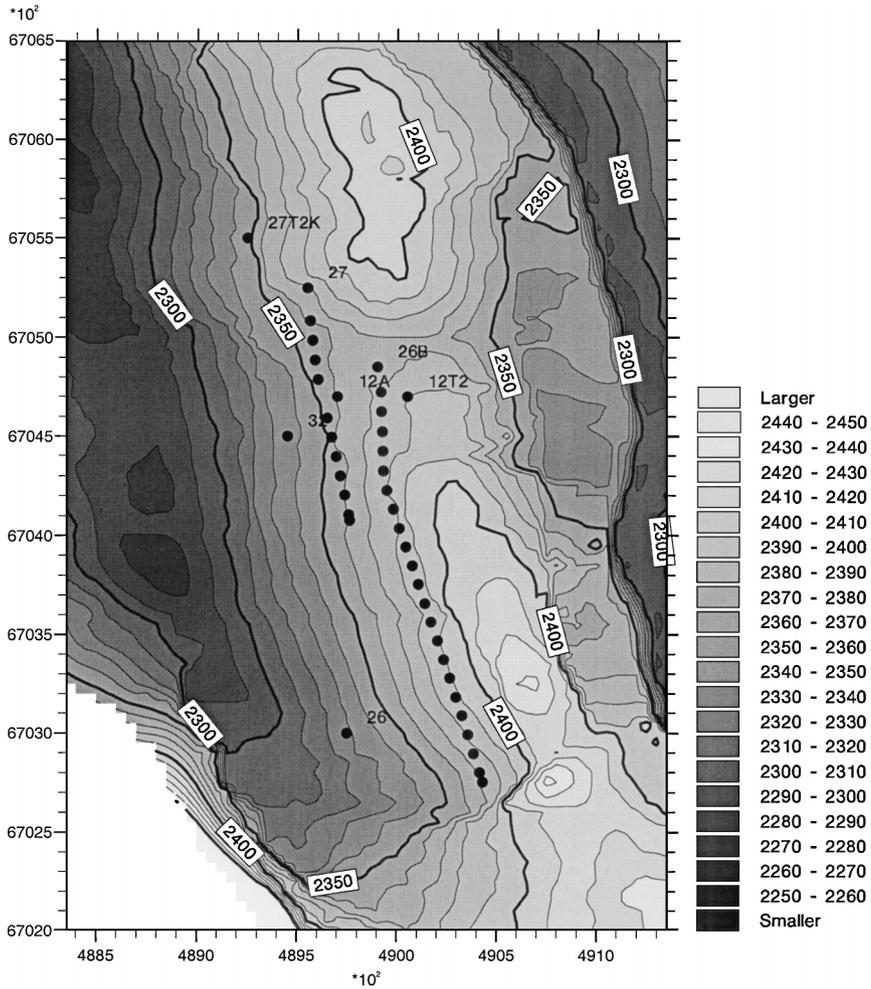


Figure 9. Contoured map of two-way travel times to horizon in milliseconds. The location of vertical well observations are marked by dots and the name of the well. The two horizontal wells are seen as two rows of dots.

the otherwise high collinearity in the model. This has no influence on the final result but makes β estimates less correlated.

We have selected a spherical variogram with range 3000 m for the error $\epsilon(\mathbf{x})$. This has in our experience proven to be a reasonable choice in several similar case studies. The standard error of the residual was chosen to be 40 m.

Table 1. Expectation and Standard Errors in the Prior and Posterior Distributions of β

	Prior guess		Posterior given exact data		Posterior given all data	
	Expec.	Stand. err.	Expec.	Stand. err.	Expec.	Stand. err.
β_1 (m/sec)	2.000	1.000	2.160	24	2.157	23
β_2 (m/sec ²)	0	10.000	247	929	-772	723

Estimation of Trend Parameters

The prior guess on β and the result from conditioning on data is summarized in Table 1. The posterior distribution given all data was obtained using the Data Augmentation Algorithm. A systematic reduction in the standard errors is verified; moreover, we can see that inequality constraints change the posterior distribution significantly.

Surface Prediction

Figures 10 and 11 show maps of predicted depth using only exact data and all data respectively. The map in Figure 11 was obtained using the approximation presented in the previous section. Figure 12 shows the difference between the predictions. The figures show that constraints from the horizontal wells “push” the predicted depths up.

Note that there is hardly any local effect from the inequality constraints; they mainly influence the trend. So each individual constraint has hardly any influence but as a group they have considerable impact. This caused some severe problems discussed in the next section.

Removal of Noninformative Constraints

When trying to compute the prediction using all data we immediately ran into problems. The positive definite condition on the matrix $(\mathbf{C}^{ile} - \Sigma^i)$ in (9) was not met. The simple reason was that the impact of some of the inequality constraints i was negligible so $[\mathbf{C}^{ile}]_{ii} \approx [\Sigma^i]_{ii}$. The Monte Carlo uncertainty in Σ^i makes this even worse so negative differences occurred. This will be a frequent practical problem because it is very difficult in advance to know whether an inequality constraint actually carry information or not. For densely sampled constraints such as observations along a horizontal well we must also expect complications since the constraints carry almost identical information.

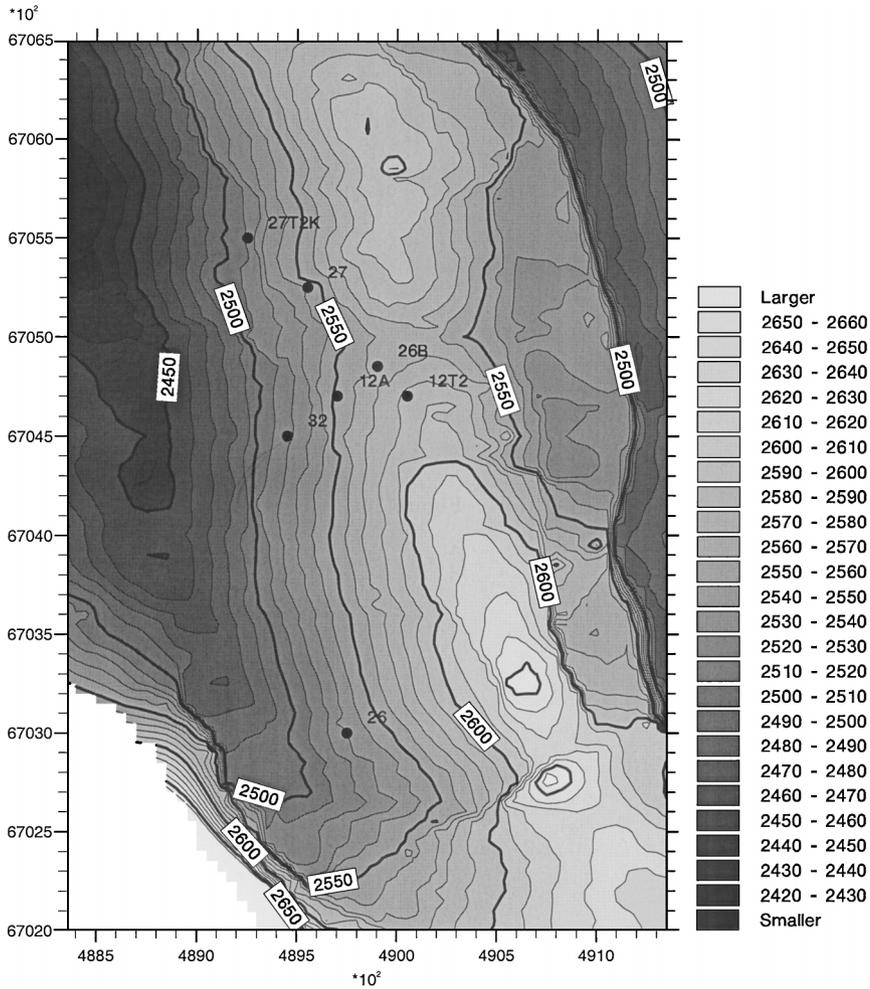


Figure 10. Contoured map of predicted depth using the 7 exact data. The location of vertical well observations are marked by dots and the name of the well.

To cure these problems, we tried to identify and remove the troublesome constraints. We suggest the following steps for removing them:

1. Remove constraint i if

$$\sqrt{C_{ii}^{i|e}} - \sqrt{\Sigma_{ii}^i} < 0$$

This means that constraint i must reduce the prediction error.

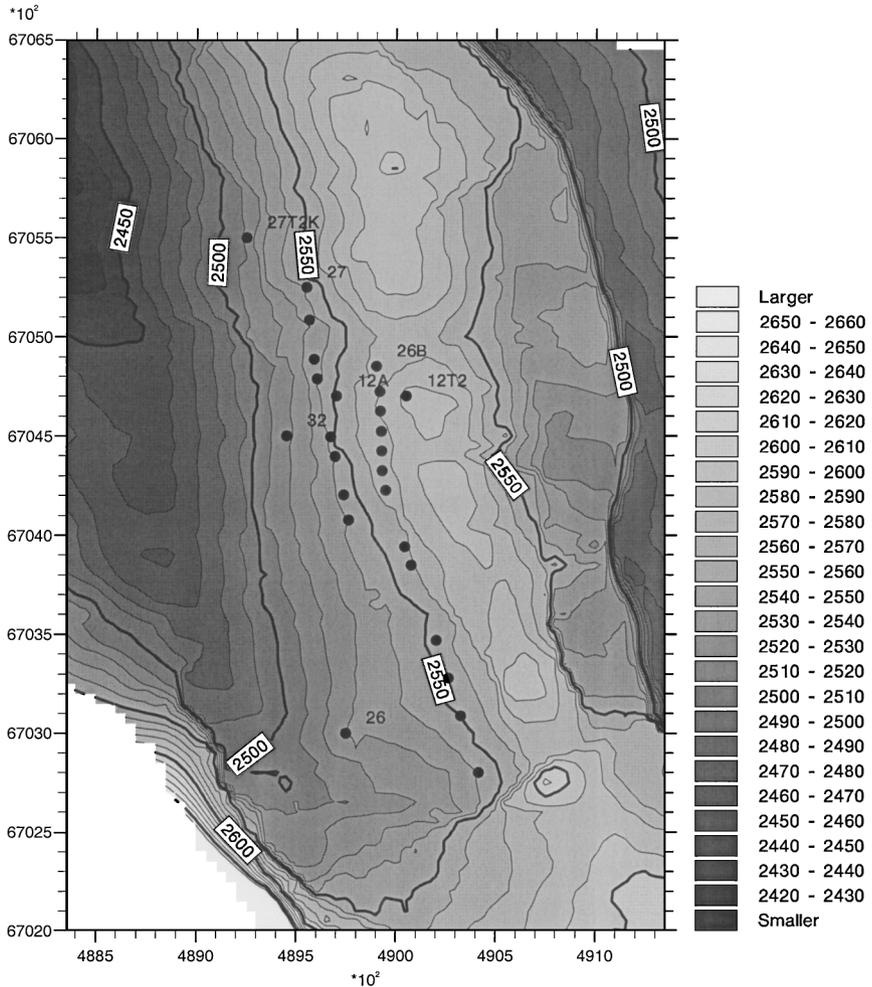


Figure 11. Contoured map of predicted depth using 7 exact and 19 inequality constraints. The location of vertical well observations are marked by dots and the name of the well. The two horizontal wells are seen as two rows of dots. Note that some of the observations seen in Figure 9 are absent.

2. The matrix $(\mathbf{C}^{ile} - \Sigma^i)$ must have the properties of a covariance matrix with a corresponding correlation matrix. Off-diagonal elements of this correlation matrix must be in the range $[-1, 1]$. The minimum necessary number of constraints was removed to make all correlations within the range $[-0.99, 0.99]$.

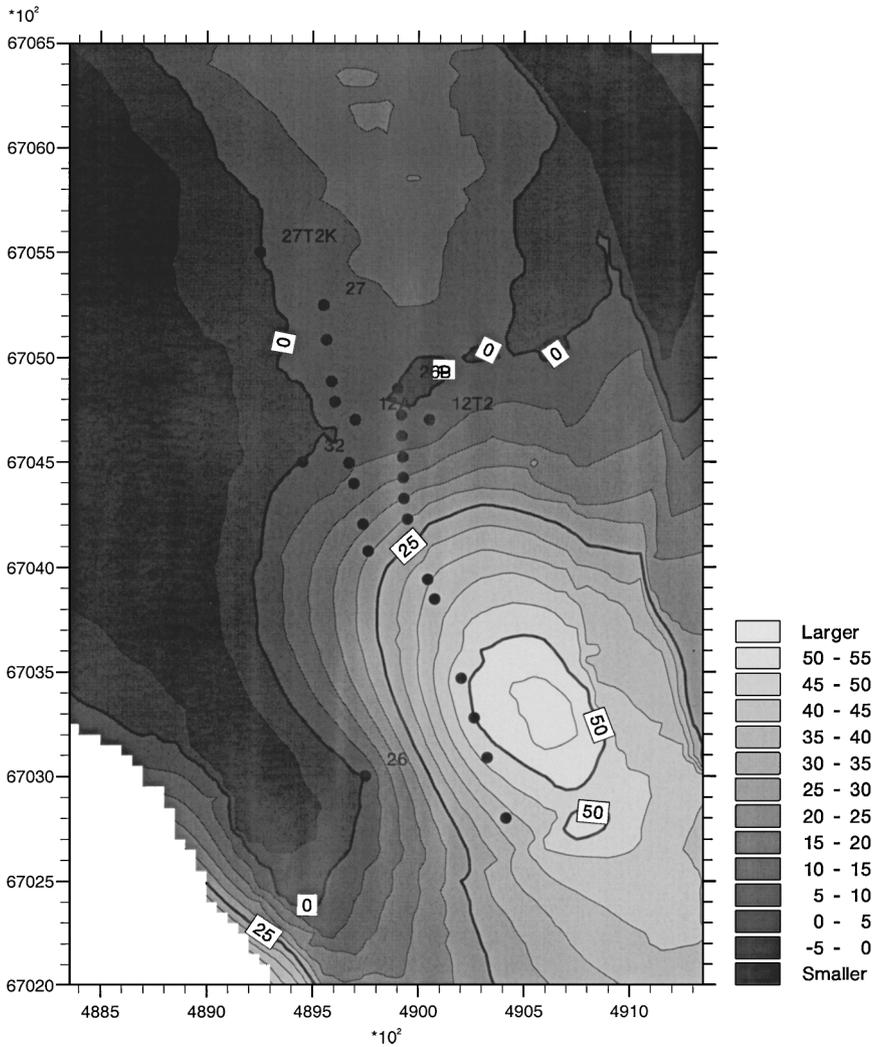


Figure 12. Contoured map of the difference between the predicted depth maps in Figure 10 and Figure 11.

3. Remove remaining constraints i if none of the following criteria are met:

$$\left(\sqrt{\mathbf{C}_{ii}^{\text{ile}}} - \sqrt{\Sigma^i}\right) / \sqrt{\mathbf{C}_{ii}^{\text{ile}}} > 0.05 \quad \text{or} \quad (|\mathbb{E}\{\mathbf{Z}_i^i | \mathbf{Z}^e\} - \mu_i^i|) / \sqrt{\mathbf{C}_{ii}^{\text{ile}}} > 0.1$$

Thus, either the constraint must reduce the prediction error more than 5% or it must influence the expectation by more than 10% relative to the prediction error at the location.

The three steps above were used to remove 14 inequality constraints, leaving 19 (Fig. 11). These 19 should carry almost all constraining information in the full set.

However, removing the 14 inequality constraints was not enough to ensure that the now 19-dimensional matrix $(\mathbf{C}^{ie} - \Sigma^i)$ was positive definite. To ensure this we changed the correlation matrix of $(\mathbf{C}^{ie} - \Sigma^i)$ toward the correlation matrix of \mathbf{C}^{ie} . This is an indirect way of changing the off-diagonal elements of Σ^i without affecting the variances. The idea is that maintaining the correct variances is far more important than getting the correlations exactly reproduced. The resulting modified Σ^i was positive definite although this procedure does not guarantee that.

The effect of the approximations and the *ad hoc* data reduction procedures on the maps is difficult to evaluate since calculating the correct distribution using the Data Augmentation Algorithm would require enormous amounts of computation resources.

CONDITIONAL SIMULATION

The Data Augmentation Algorithm can be used for sampling \mathbf{Z} from $f^{(n)}(\mathbf{z}, \beta | \mathbf{z}^e, \mathbf{B}^i)$ in (4). However, \mathbf{Z} might have 10^4 or more elements when a dense grid is used for representing the random field. The dimension of \mathbf{Z} will then limit the efficiency of the Data Augmentation Algorithm dramatically so an alternative is necessary.

The following is a sequential simulation algorithm motivated by the identity

$$f(\mathbf{z} | \mathbf{z}^e, \mathbf{B}^i) = \int_{\mathbb{R}^P} \int_{\mathbb{R}^m} \varphi(\mathbf{z} | \mathbf{z}^e, \mathbf{z}^i, \beta) \bar{\varphi}(\mathbf{z}^i | \mathbf{z}^e, \beta, \mathbf{B}^i) f(\beta | \mathbf{z}^e, \mathbf{B}^i) d\mathbf{z}^i d\beta \quad (11)$$

Simulation Algorithm

- (a) Draw $\tilde{\beta}$ from $f(\beta | \mathbf{z}^e, \mathbf{B}^i)$ using the Data Augmentation Algorithm.
- (b) Draw $\tilde{\mathbf{z}}^i$ from $\bar{\varphi}(\mathbf{z}^i | \mathbf{z}^e, \tilde{\beta}, \mathbf{B}^i)$ by drawing from $\varphi(\mathbf{z}^i | \mathbf{z}^e, \beta)$ until $\tilde{\mathbf{z}}^i \in \mathbf{B}^i$.
- (c) Draw \mathbf{z} from $\varphi(\mathbf{z} | \tilde{\beta}, \mathbf{z}^e, \tilde{\mathbf{z}}^i)$.

Drawing \mathbf{z} in Step (c) is the standard conditional simulation problem of conditioning a Gaussian random field with known trend, $\mathbf{f}'(\mathbf{x})\beta$, to “hard” data \mathbf{z}^e and $\tilde{\mathbf{z}}^i$. Thus, the Data Augmentation Algorithm is used as a preprocessor to established

a set of simulated values for β and \mathbf{z}^i . These, are simply used as hard data in any established conditional simulation procedure.

DISCUSSION

A method for conditioning a “universal kriging” model on inequality constraints has been presented. The Data Augmentation Algorithm provides a method to find the posterior distributions of Gaussian random fields. In our experience, it seems that the posterior trend has an almost Gaussian posterior density in many natural cases. However, if we use constraints forcing the field to be outside of a forbidden interval, the trend may become bimodal and the conditional expectation may violate the constraints.

Predicting values in a large grid requires approximations. We suggest a method where the inequality constraints are replaced by “equivalent” point data with error bounds. This method has been tested on a large-scale case study and showed that it needs refinement.

Conditional simulation can be performed by letting the Data Augmentation Algorithm provide samples of β and \mathbf{z}^i , and then conditioning the random field on these.

ACKNOWLEDGMENTS

The work has been supported by the Research Council of Norway, Norsk Hydro, Saga Petroleum, and Statoil. We are grateful to Norsk Hydro for providing data for the case study. We are also grateful to Arne Skorstad who has done some intricate modifications to old software.

REFERENCES

- Abrahamsen, P., and Benth, F. E., 1998, Gaussian field with unknown trend conditioned on inequality data, *in* Buccianti, A., Nardi, G., and Potenza, R., eds., Proceedings of IAMG '98: De Frede Editore, Naples, Italy, p. 333–338.
- Choi, M.-K., Christakos, G., and Serre, M. L., 1998, Recent developments in vectorial and multi-point BME analysis, *in* Buccianti, A., Nardi, G., and Potenza, R., eds., Proceedings of IAMG '98: De Frede Editore, Naples, Italy, p. 91–96.
- Diamond, P., 1988, Interval-valued random functions and the kriging of intervals: *Math. Geology*, v. 20, no. 3, p. 145–165.
- Dubrule, O., and Kostov, C., 1986, An interpolation method taking into account inequality constraints: I. Methodology: *Math. Geology*, v. 18, no. 1, p. 33–51.
- Freulon, X., and de Fouquet, C., 1993, Conditioning a Gaussian model with inequalities, *in* Soares, A., ed., *Geostatistics Tróia '92*, vol. 1: Kluwer Academic, Dordrecht, The Netherlands, p. 201–212.

- Journal, A. G., 1986, Constrained interpolation and qualitative information—The soft kriging approach: *Math. Geology*, v. 18, no. 3, p. 269–286.
- Kitanidis, P. K., 1986, Parameter uncertainty in estimation of spatial functions: Bayesian analysis: *Water Resources Res.*, v. 22, no. 4, p. 499–507.
- Kostov, C., and Dubrule, O., 1986, An interpolation method taking into account inequality constraints: II. Practical approach: *Math. Geology*, v. 18, no. 1, p. 53–73.
- Langlais, V., 1989, On the neighborhood search procedure to be used when kriging with constraints, *in* Armstrong, M., ed., *Geostatistics*, Vol. 2: Kluwer Academic, Dordrecht, The Netherlands, p. 603–614.
- Mardia, K. V., Kent, J. T., and Bibby, J. M., 1979, *Multivariate analysis*: Academic Press, London, 518 p.
- Militino, A. F., and Ugarte, M. D., 1999, Analyzing censored spatial data: *Math. Geology*, v. 31, no. 5, p. 551–562.
- Omre, H., 1987, Bayesian kriging—Merging observations and qualified guesses in kriging: *Math. Geology*, v. 19, no. 1, p. 25–39.
- Stein, M. L., 1992, Prediction and inference for truncated spatial data: *Jour. Comp. Graphical Statist.* v. 1, no. 1, p. 354–372.
- Tanner, M. A., 1993, *Tools for statistical inference*, 2. ed.n: Springer-Verlag, New York, 156 p.
- Tanner, M. A., and Wong, W. H., 1987, The calculation of posterior distributions by data augmentation (with discussion): *Jour. American Statist. Assoc.*, v. 82, no. 398, p. 528–550.

APPENDIX

Evaluation of (2a) and (2b)

$$\begin{aligned}
 f(\mathbf{z}, \boldsymbol{\beta} \mid \mathbf{z}^e, \mathbf{B}^i) &= \int_{\mathbb{R}^m} \varphi(\mathbf{z}, \boldsymbol{\beta} \mid \mathbf{z}^e, \mathbf{z}^i) f(\mathbf{z}^i \mid \mathbf{z}^e, \mathbf{B}^i) d\mathbf{z}^i \\
 &= \int_{\mathbb{R}^m} \varphi(\mathbf{z}, \boldsymbol{\beta} \mid \mathbf{z}^e, \mathbf{z}^i) \int_{\mathbb{R}^p \times \mathbb{R}^k} \bar{\varphi}(\mathbf{z}^i \mid \mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{B}^i) \\
 &\quad \times f(\tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}} \mid \mathbf{z}^e, \mathbf{B}^i) d\tilde{\mathbf{z}} d\tilde{\boldsymbol{\beta}} d\mathbf{z}^i \\
 &= \int_{\mathbb{R}^p \times \mathbb{R}^k} \left(\int_{\mathbb{R}^m} \varphi(\mathbf{z}, \boldsymbol{\beta} \mid \mathbf{z}^e, \mathbf{z}^i) \bar{\varphi}(\mathbf{z}^i \mid \mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{B}^i) d\mathbf{z}^i \right) \\
 &\quad \times f(\tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}} \mid \mathbf{z}^e, \mathbf{B}^i) d\tilde{\mathbf{z}} d\tilde{\boldsymbol{\beta}} \\
 &= \int_{\mathbb{R}^p \times \mathbb{R}^k} K(\mathbf{z}, \boldsymbol{\beta}; \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}) f(\tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}} \mid \mathbf{z}^e, \mathbf{B}^i) d\tilde{\mathbf{z}} d\tilde{\boldsymbol{\beta}}
 \end{aligned}$$