

GAUSSIAN FIELD WITH UNKNOWN TREND CONDITIONED ON INEQUALITY DATA

P. Abrahamsen and F. E. Benth
Norwegian Computing Center
P.O. Box 114 Blindern
N-0314 Oslo, Norway

1 SUMMARY

A Gaussian random field with a linear trend for the mean is considered. Methods for obtaining the distribution for trend coefficients given exact and inequality data are established. Moreover, the conditional distribution for the random field at any location is calculated. The approach adapts the Data Augmentation Algorithm of Tanner and Wong [5, 6] which is a Monte Carlo technique for finding the fixed point of an integral operator.

Inequality data has previously been used for mapping random fields using indicator kriging [4], quadratic programming and splines [2], interval kriging [1], and Gaussian simulation [3]. However, these authors assume a known mean.

2 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, $\boldsymbol{\beta}$ is P coefficients, and $\epsilon(\mathbf{x})$ is a zero mean Gaussian random field with known covariance function. The P coefficients are assumed to have a prior multinormal distribution: $\boldsymbol{\beta} \sim \mathcal{N}_P(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$.

Let N exact observations of $Z(\mathbf{x})$ be given: $\mathbf{Z}^e = [Z(\mathbf{x}_1^e), \dots, Z(\mathbf{x}_N^e)]' = [z(\mathbf{x}_1^e), \dots, z(\mathbf{x}_N^e)]'$. Furthermore, assume that there are M inequality constraints on $Z(\mathbf{x})$: $\mathbf{Z}^i = [Z(\mathbf{x}_1^i), \dots, Z(\mathbf{x}_M^i)]' \in [B_1, \dots, B_M]'$, where B_1, \dots, B_M are 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 Borel set in \mathbb{R}^M and $\mathbf{Z}^i \in \mathbf{B}^i$.

In the following several probability density functions (pdf's) will be 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 $\boldsymbol{\beta}$ have a joint multinormal distribution, $\varphi(z(\mathbf{x}), \mathbf{z}^e, \mathbf{z}^i, \boldsymbol{\beta})$, and that e.g., the conditional distribution $\varphi(z(\mathbf{x})|\mathbf{z}^i, \mathbf{z}^e, \boldsymbol{\beta})$ is also multinormal. The pdf $\bar{\varphi}(z(\mathbf{x}), \mathbf{z}^i|\mathbf{z}^i \in \mathbf{B}^i) = \bar{\varphi}(z(\mathbf{x}), \mathbf{z}^i|\mathbf{B}^i)$ is truncated normal, where \mathbf{B}^i will be used as a short hand for $\mathbf{z}^i \in \mathbf{B}^i$. E.g. the density $f(z(\mathbf{x})|\mathbf{z}^e, \boldsymbol{\beta}, B^i)$ is not normal nor truncated normal.

Multinormal pdf's can be handled analytically and efficient simulation methods are available provided the dimensions are less than, say 500. Truncated normals are more difficult to handle analytically but simulation is straightforward using, for example, rejection sampling.

3 OBTAINING POSTERIOR DISTRIBUTIONS

Consider an arbitrary set of K locations, $\{\mathbf{x}_1, \dots, \mathbf{x}_K\}$, and the random field at these locations organised 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, $\mathbf{Z}^i \in \mathbf{B}^i$, denoted $f(\mathbf{z}|\mathbf{z}^e, \mathbf{B}^i)$. Given this pdf, predictors using expectation or the mode can be evaluated and uncertainty measures such as variance and quantiles can be calculated. In particular $z^*(\mathbf{x}) = \mathbb{E}\{Z(\mathbf{x})|\mathbf{z}^e, \mathbf{B}^i\}$ is the natural extension to the standard linear kriging predictor.

Also of interest is to find the posterior distribution of $\boldsymbol{\beta}$ denoted $f(\boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$ so it is convenient to consider the joint pdf $f(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$ from which marginals $f(\boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$ and $f(\mathbf{z}|\mathbf{z}^e, \mathbf{B}^i)$ are obtained.

Following Tanner and Wong [5, 6] the posterior density can be expressed as the fixed point of an integral equation using basic probabilistic manipulations:

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

where the transition kernel is

$$K(\mathbf{z}, \boldsymbol{\beta}; \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}) = \int_{\mathbb{R}^M} \varphi(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{z}^i) \bar{\varphi}(\mathbf{z}^i|\mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{B}^i) d\mathbf{z}^i. \quad (1b)$$

The fixed point can be computed by iterating

$$f^{(n+1)}(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i) = \int_{\mathbb{R}^P} \int_{\mathbb{R}^K} K(\mathbf{z}, \boldsymbol{\beta}; \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}) f^{(n)}(\tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}|\mathbf{z}^e, \mathbf{B}^i) d\tilde{\mathbf{z}} d\tilde{\boldsymbol{\beta}}. \quad (2)$$

Convergence is ensured since the kernel behaves nicely according to criteria given by Tanner and Wong [5].

Analytical integration of the integrals in the fixed point iterations are 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 (2):

Data Augmentation Algorithm.

Given the current approximation $f^{(n)}(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$ of $f(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$.

- (a) Draw S samples, $\mathbf{z}_{(1)}^i, \dots, \mathbf{z}_{(S)}^i$, from the current approximation to the predictive density $f(\mathbf{z}^i|\mathbf{z}^e, \mathbf{B}^i)$. This is done in two steps:
 - (a1) Draw $\tilde{\boldsymbol{\beta}}$ from $f^{(n)}(\boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$.
 - (a2) Draw $\mathbf{z}_{(s)}^i$ from $\bar{\varphi}(\mathbf{z}^i|\tilde{\boldsymbol{\beta}}, \mathbf{z}^e, \mathbf{B}^i)$, where we use $\tilde{\boldsymbol{\beta}}$ from Step (a1). This is done by drawing from the normal density $\varphi(\mathbf{z}^i|\tilde{\boldsymbol{\beta}}, \mathbf{z}^e)$ until $\mathbf{z}_{(s)}^i \in \mathbf{B}^i$.
- (b) Update the current approximation to $f(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$ to be the mixture of conditional densities of $(\mathbf{z}, \boldsymbol{\beta})$ given the augmented data patterns generated in Step (a), that is,

$$f^{(n+1)}(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i) = \frac{1}{S} \sum_{s=1}^S \varphi(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{z}_{(s)}^i), \quad (3)$$

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

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

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

The $\bar{\varphi}(\mathbf{z}^i|\mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{B}^i)$ part comes from (1b) and $f^{(n)}(\tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}|\mathbf{z}^e, \mathbf{B}^i)$ is found in (2). Thus, Step (a) essentially performs two of the integrals in the fixed point iteration. However, the $\tilde{\mathbf{z}}$ integral is removed from the integral since $f^{(n)}(\mathbf{z}^i|\mathbf{z}^e, \mathbf{B}^i) = \int_{\mathbb{R}^P} \bar{\varphi}(\mathbf{z}^i|\mathbf{z}^e, \tilde{\boldsymbol{\beta}}, \mathbf{B}^i) f^{(n)}(\tilde{\boldsymbol{\beta}}|\mathbf{z}^e, \mathbf{B}^i) d\tilde{\boldsymbol{\beta}}$. In Step (a1) sampling $\boldsymbol{\beta}$ from $f^{(n)}(\boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$ is done by sampling from one randomly selected distribution in the sum in (3); \mathbf{z} is only considered after the final iteration. Step (b) is an evaluation of the kernel (2) of the fixed point integral: $\int_{\mathbb{R}^M} \varphi(\mathbf{z}, \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{z}^i) \bar{\varphi}(\mathbf{z}^i|\mathbf{z}^e, \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{B}^i) d\mathbf{z}^i$, 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 constraints $\mathbf{z}^i \in \mathbf{B}^i$ 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 $f^{(n)}(\mathbf{z}^i|\mathbf{z}^e, \mathbf{B}^i)$ approach $f(\mathbf{z}^i|\mathbf{z}^e, \mathbf{B}^i)$. The second problem must be handled by implementing smart rejection sampling techniques.

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

3.1 Posterior distribution for $\boldsymbol{\beta}$ and $Z(\mathbf{x})$

If the objective is limited to obtaining moments for the distribution for $\boldsymbol{\beta}$ 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 (3) is replaced by $f^{(n+1)}(\boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i) = \frac{1}{S} \sum_{s=1}^S \varphi(\boldsymbol{\beta}|\mathbf{z}^e, \mathbf{z}_{(s)}^i)$.

To obtain the posterior distribution $f(z(\mathbf{x})|\mathbf{z}^e, \mathbf{B}^i)$ the Data Augmentation Algorithm must be iterated until $f^{(n)}(z(\mathbf{x}), \boldsymbol{\beta}|\mathbf{z}^e, \mathbf{B}^i)$ has converged to the necessary precision. Replacing \mathbf{z} in (3) by $z(\mathbf{x})$ and simply ignoring the $\boldsymbol{\beta}$'s give $f^{(n)}(z(\mathbf{x})|\mathbf{z}^e, \mathbf{B}^i) = \frac{1}{S} \sum_{s=1}^S \varphi(z(\mathbf{x})|\mathbf{z}^e, \mathbf{z}_{(s)}^i)$.

4 EXAMPLES

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 $\boldsymbol{\beta} \sim \mathcal{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10^2 & 0 \\ 0 & 2^2 \end{bmatrix}\right)$. The prior variances are two order of magnitudes larger than the posterior variances obtained below and is almost ‘non informative’.

4.1 Predicting $\boldsymbol{\beta}$

Figure 1 shows regression lines obtained using three exact data and two one sided inequality constraints. The marginal posterior probability densities for the

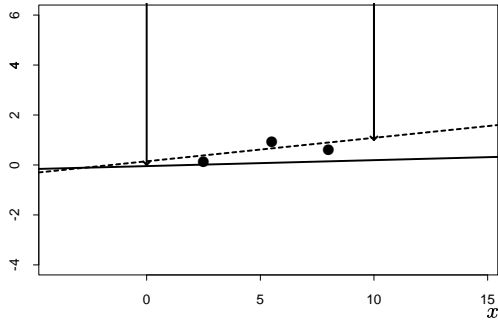


Figure 1: Trends using coefficients $E\{\beta|z^e, \mathbf{B}^i\}$ (solid line) and $E\{\beta|z^e\}$ (dashed line). Exact data are plotted as dots and interval constraints as arrows. Vertical axis is z .

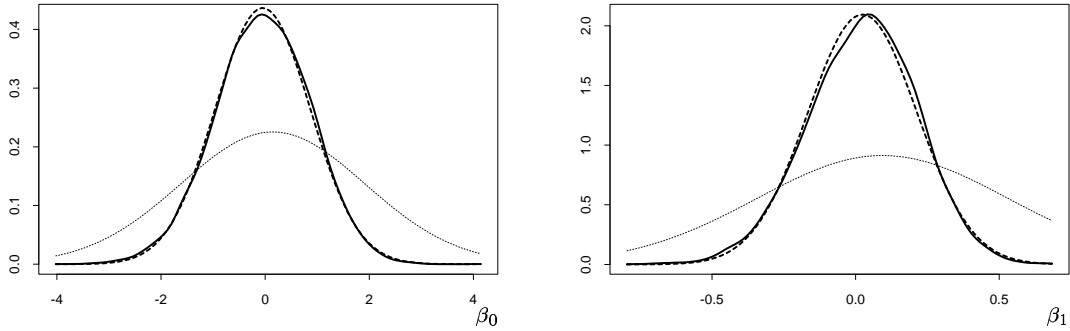


Figure 2: Probability densities. Solid black lines are obtained using a kernel smoother of 8192 samples while dashed lines are normal distributions with identical expectations and variances. Dotted curves are the posterior normal distributions given only exact data.

intercept, β_0 , and the slope, β_1 , are shown in Figure 2. The figures illustrate that the two inequality constraints alter expectations and reduce the variances. It is also seen that the conditional distributions for intercept and 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. The next example however gives a highly non-Gaussian result.

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

4.2 Predicting $Z(x)$

Figure 5 illustrates the marginal distributions of $Z(x)|z^e, \mathbf{B}^i$ by showing quantiles and expectations as a function of x . The statistics are based on 8192 samples of $Z(x)$. The right hand plot in Figure 5 shows that using $E\{Z(x)|z^e, \mathbf{B}^i\}$ as the predictor for $Z(x)$ may lead to poor results since it is in the “illegal” interval at $x = 10$. The median however, passes above this interval. For the example using one sided constraints the conditional expectation behaves nicely but notice that the median is slightly above since the distributions are skew.

4.3 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 limited. The examples above were obtained by using $S = 2$ in the initial fixed point iteration and then doubling S at each iteration until $S = 8192$

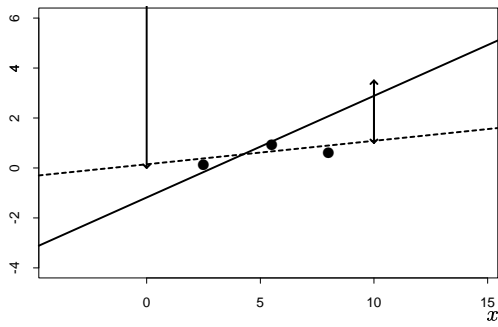


Figure 3: Trends using coefficients $E\{\beta|z^e, \mathbf{B}^i\}$ (solid line) and $E\{\beta|z^e\}$ (dashed line). Exact data are plotted as dots and interval constraints as arrows. Vertical axis is z .

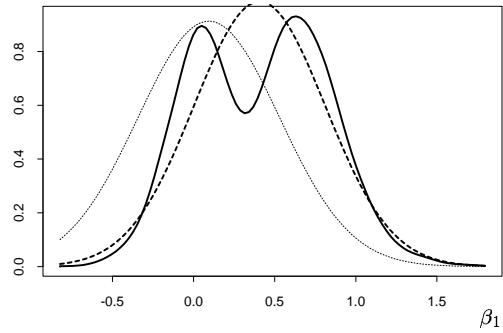
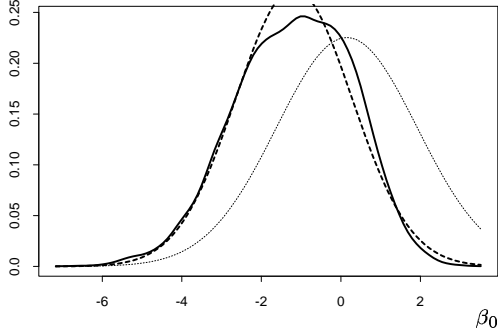


Figure 4: Probability densities. Solid black lines are obtained using a kernel smoother of 8192 samples while dashed lines are normal distributions with identical expectations and variances. Dotted curves are the posterior normal distributions given only exact data.

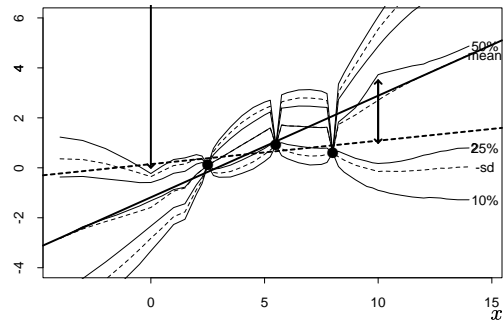
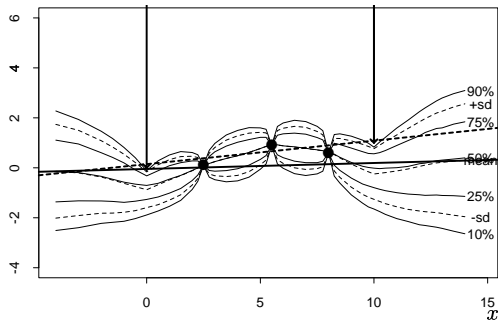


Figure 5: Quantiles of $Z(x)|z^e, \mathbf{B}^i$ shown as solid lines. Expectation plus/minus standard error shown as dashed lines.

at iteration 13. From then on S was kept unchanged. Figure 6 shows how the distributions for the slope evolves as the number of fixed point iterations increase. The convergence for the first example is rapid and after 10 iterations ($S = 1024$) the levels seems to stabilise. Increasing S after this is mainly to reduce Monte Carlo noise. For the second example however, convergence is very slow. Forty iterations are shown in the right hand plot of Figure 6 and the levels does not stabilise until at least 35 fixed point iterations has been run. The problem seems to be that the number of β_1 samples from the highest mode is underestimated during the initial iterations.

5 DISCUSSION

A method for conditioning a ‘universal kriging’ model on inequality constraints has been presented. The Data Augmentation Algorithm provides us with an iterative method to find the posterior distributions in question. We have tried out

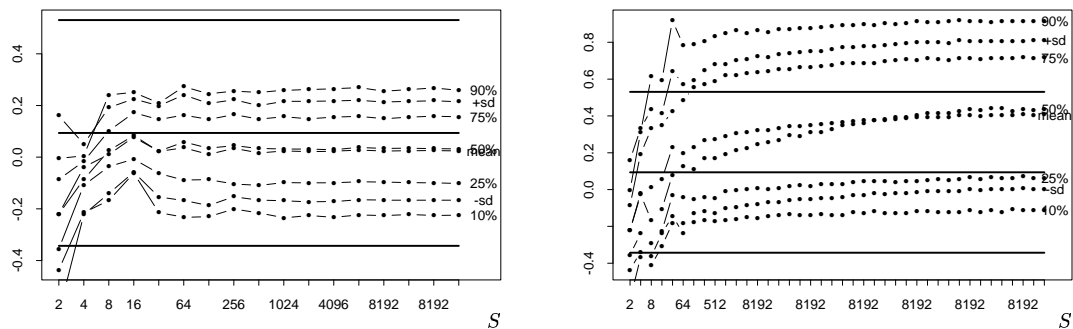


Figure 6: Quantiles and moments for $\beta_1 | z^e, B^i$ as a function of the number of fixed point iterations. The horizontal lines are expectation \pm standard deviation for $\beta_1 | z^e$.

the algorithm for several examples, showing that the posterior trend has an almost Gaussian posterior density for many natural cases. If we use constraints forcing the field to be outside a forbidden interval the trend gets bimodal. Moreover, conditional expectation is no longer a reasonable predictor, since it may violate the constraints.

Predicting values in a large grid requires approximations. An idea we are working on is to replace inequality constraints by ‘equivalent’ point data with error bounds.

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

ACKNOWLEDGEMENT

This work was sponsored by the Research Council of Norway.

REFERENCES

- [1] P. Diamond. Interval-valued random functions and the kriging of intervals. *Math. Geol.*, **20**, 145–165, 1988.
- [2] O. Dubrule and C. Kostov. An interpolation method taking into account inequality constraints: I. Methodology. *Math. Geol.*, **18**, 33–51, 1986.
- [3] X. Freulon and C. de Fouquet. Conditioning a Gaussian model with inequalities. In A. Soares, Ed., *Geostatistics Tróia '92*, Dordrecht, Kluwer Academic Publ., 1993, 201–212
- [4] A. G. Journel. Constrained interpolation and qualitative information—the soft kriging approach. *Math. Geol.*, **18**, 269–286, 1986.
- [5] M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.*, **82**, 528–550, 1987.
- [6] M. A. Tanner. *“Tools for Statistical Inference”*. Springer-Verlag Inc., New York, 2. edition, 1993.