Conditioning geological surfaces to horizontal wells

Ariel Almendral Vázquez1 - Pål Dahle1 - Petter Abrahamsen1 - Audun Sektnan1

Abstract
Kriging is a standard method for conditioning surfaces to observations. Kriging works for vertical wells, but may produce surfaces that cross horizontal wells between surface observations. We establish an approach that also works for horizontal wells, where surfaces are modeled as a set of correlated Gaussian random fields. The constraints imposed by the horizontal wells makes the conditional surfaces non-Gaussian. We present a method for exact conditional simulation and an approximation for prediction and prediction uncertainty. Thousands of constraints can be handled efficiently without numerical instabilities. The approach is illustrated with synthetic and real examples that show how the constraints influence the surfaces and reduce uncertainty.

Keywords  Surface prediction · Kriging · Horizontal wells · Inequality constraints

1 Introduction
Conditioning a set of stratigraphic surfaces to surface observations was considered by [1]. The basic idea is to describe each surface as the top of a stratigraphic layer. Layer thicknesses are assumed to be modeled as Gaussian random fields, making surfaces a set of correlated Gaussian random fields. Standard kriging and conditional simulation techniques provide surface maps conditioned to the surface observations.

As horizontal wells became the standard drilling technique, a new and important source of data became available: zone-log information. The zone-logs constraint the position of the surfaces along the well path, giving a large number of surface (inequality) constraints, when compared with direct surface observations. In [3], the authors proposed a method to condition one surface to surface constraints. Their main idea was to represent the surface constraints by synthetic data with a particular covariance structure, and thereafter use standard kriging to generate the surface map.

In this paper we generalize the inequality constrain method in [3] to multiple surfaces, while respecting the formalism in [1]. One important challenge is to handle large amounts of surface constraints that appear on real applications with tens of surfaces and hundreds of wells. The Monte Carlo method in [3] does not scale to real applications and requires modifications. One of the main improvements in this work is the use of an efficient version of the Gibbs sampler that uses the geometrical, pair-wise distribution of the constraints to efficiently draw accurate samples. The finite number of samples necessarily implies that there is an inherent Monte Carlo sampling error in the estimated positions of the surfaces. We have now provided an upper estimate of this mismatch. The method was briefly explained in [4] using two surfaces, but we now provide a rigorous mathematical presentation, several examples and details on the algorithms.

Conditioning to inequality constraints have been studied by a number of authors, going back to [16] with indicator kriging and [12], [18] and [19] that use the dual kriging formalism combined with quadratic programming. Others have used rejection sampling Monte Carlo simulations [14] to evaluate high dimensional integrals to obtain posterior distributions [29] and the Expectation-Maximization algorithm for predicting uranium grade at censored locations [24]. A kriging formalism for interval-valued data was developed in [11], and [8] worked with constraints in...
the area of vectorial and multi-point Bayesian maximum entropy analysis. An MCMC Bayesian model for modeling the spatial distribution of a parameter with censored observations, using a Metropolis-Hastings algorithm was presented by [25]. Versions of the Gibbs sampler have also been proposed to handle large number of constraints [13]. The paper [23] gives a rigorous presentation and application of the Gibbs method for truncated distributions from physically constrained problems (non-negativity of concentrations, censored data, etc.) in environmental applications. The author illustrates how the procedure is used to estimate uncertainty bands in both geostatistical interpolation and inverse modeling. The Gibbs sampler is also an important step when simulating conditioned pluri-Gaussian random fields, in the context of facies modeling [7].

The Data Augmentation Algorithm [31] in our application is also used in [15] to detect the presence of a pollutant below detection threshold. Gaussian process emulators that satisfy constraints (inequalities, convexity, etc.) with respect to some input variables are studied by [20]. The method in this paper is classified as 2.5D by [33] and cannot handle multi-valued surfaces like overturned folds and salt domes.

The next section describes the basic ideas without any technical details. Section 3 describes the stochastic model for surfaces and layers, and the formal representation of well data is given in Section 4. Section 5 describes how to calculate the conditional probability distribution of surface depth given surface observations and constraints from horizontal wells. This section also presents efficient ways to calculate conditional predictions, prediction uncertainties and an algorithm to draw simulated realizations from the conditional probability distribution. Some technical details are elaborated on in Appendix A. Section 6 provides two examples that illustrate the methodology, while Section 7 shows results from a real case.

2 Basic idea

Horizontal wells impose restrictions on the surfaces in the form of inequality constraints. Well observations (logs, cores) along a well are typically measured every foot. Assuming surfaces are somewhat smooth indicates that restricting a surface laterally at every foot is unnecessary. Also, the finite grid resolution (typically 10–100 meters) limits the need for lateral precision. It is therefore convenient to resample the well to reduce the number of constraints. It is sufficient to keep samples separated by approximately the grid resolution since this is the resolution of the predictions. Proper sampling of the well must take into account location of well points (observation of surfaces), well inclination, and interaction with other wells.

Standard kriging and conditional simulation techniques do not include the possibility to condition on inequality constraints so a few initial steps are added to the standard approaches. The following steps outlines the proposed method and they are illustrated in Fig. 1:

(a) **Initial state:** Consider the vertical cross section through a well and two surfaces. The upper surface has been observed at one well point (green disc) where the well enter the layer between the two surfaces. The task is to predict or simulate the two surfaces given this information.

(b) **Condition on the well point:** Calculate the depth prediction (solid red lines) and the prediction uncertainty envelope (dashed lines) along the well using kriging. Since surfaces are correlated, the surface observation of the upper surface influences the lower surface.

(c) **Select constraining sample points:** These are the sample points that are close to a surface, that is, the vertical distance between the well path and surface predictions are small relative to the prediction uncertainty. Blue discs has been identified as constraining sample points.

(d) **Draw surface values at the constraints:** Draw 256 sets of correlated surface depth values at the constraining sample points. This is done by using the Data Augmentation Algorithm described in Section 5.1. These 256 sets of depth values represent the probability distribution for the surface depth at the constraints. They are illustrated as small point clouds in Fig. 1(d). Note that this is not a multivariate Gaussian distribution since the combination of high correlations and truncations makes any marginal distribution non-Gaussian. The number of sets could be larger than 256, but experience has shown that this is a proper balance between speed and precision.

(e) **Prediction:** Estimate the expectation and the covariance of the 256 sets of surface depth values from Step (d). The estimated expectations and covariances are used to calculate a set of data points that are used in the kriging equations. The expectations and variances of these data points are chosen such that the prediction and prediction uncertainty exactly reproduce the estimated expectations and covariances of the 256 sets of surface depth values (Section 5.5). Figure 1(e) shows the resulting prediction and prediction uncertainty.

(f) **Conditional simulation:** Select randomly one of the 256 sets of correlated surface depth values drawn in Step (d). The selected set of depth values are used as additional well points when conditioning the simulated realization.
Fig. 1 Cross section along a well. The well path is shown in black, and the well point as a green disc. White and blue discs are well samples. The blue discs are constraining well samples. Surfaces are red and surface uncertainty envelopes (±2 SD) are indicated by red dashed lines. The red discs are constrained surface points.

Note that Step (e) is an approximation where non-Gaussian marginals are replaced by a symmetric (multi-Gaussian) distribution calculated using kriging. Step (f) however, is an exact simulation algorithm that provides perfect samples from the conditional distribution given well points and the inequality constraints at the constraining sample points.

3 Stochastic model for surfaces and layers

Consider $L$ surfaces. The depth to surface $l$ is obtained by adding the thicknesses of the stratigraphic layers above the surface

$$z^l(x) = z^0(x) + \sum_{k=0}^{l-1} \Delta z^k(x), \quad x \in \mathbb{R}^2, \quad 1 \leq l \leq L.$$ 

In the following we will assume that the reference surface $z^0(x) = 0$ to simplify the notation.

The thickness of layer $l$ (below surface $l$) is modeled as

$$\Delta z^l(x) = \Delta m^l(x) + \xi^l(x),$$

where $\Delta m^l(x)$ is the layer thickness trend and $\xi^l(x)$ is the layer thickness residual. The layer thickness residual is modeled as a zero mean Gaussian random field [2] specified by the correlation function $\rho_l$ and the standard deviation $\sigma_l$ so that

$$E[\xi^l(x)] = 0,$$

$$\text{Var}(\xi^l(x)) = \sigma^2_l(x),$$

$$\text{Cov}(\xi^k(x), \xi^l(y)) = \delta_{kl} \sigma_k(x) \sigma_l(y) \rho_l(||x - y||),$$

where residuals from different layers have been assumed independent. Here $\delta_{kl}$ is 1 if $k$ is equal to $l$, and 0 otherwise.

The thickness trend $\Delta m^l(x)$ for layer $l$ is a linear combinations of $P^l$ known trend maps $g_p^l(x)$

$$\Delta m^l(x) = \sum_{p=1}^{P^l} \beta_p^l g_p^l(x),$$

where $\beta_p^l$ are the regression or trend coefficients. These trend maps can be constants, isochore maps, or any map useful to model the thickness of layer $l$. Isochore maps can be constructed from interpretation of seismic data or a depositional geological concept. The role of these trend maps is to capture the large scale shape of the surface.

To simplify notation we introduce

$$g^l(x) = [g_1^l(x), \ldots, g_{P^l}^l(x)],$$

and $\beta^l = [\beta_1^l, \ldots, \beta_{P^l}^l]$ so that Eq. 1 can be written

$$\Delta m^l(x) = g^l(x) \cdot \beta^l.$$

Using this notation we can write the depth to a surface as

$$z^l(x) = m^l(x) + \epsilon^l(x),$$
where the depth trend $m_i^l(x)$ and depth residual $e_i^l(x)$ are given by
\[
m_i^l(x) = \sum_{k=0}^{l-1} g_i^k(x) \cdot b_i^k,
\]
\[
e_i^l(x) = \sum_{k=0}^{l-1} \xi_i^k(x).
\]

The depth residual is a sum of Gaussian random fields and is therefore also a Gaussian random field.

To simplify notation, all trend coefficients are organized in a vector
\[
\beta' = [\beta_i^1, \ldots, \beta_i^{L_i}]
\]
of length $P = \sum_{l=1}^{L_i} P_l$. A corresponding trend map vector $f_i^l(x)$ for each surface is formed by replacing trailing non-contributing trend maps by zeros
\[
f_i^l(x) = [g_i^1(x), \ldots, g_i^l(x), 0, \ldots, 0].
\]

We can now write Eq. 2 as
\[
z_i^l(x) = f_i^l(x) \cdot \beta_i^l + e_i^l(x).
\]

This is a linear multiple regression model with correlated Gaussian errors.

The trend coefficients are assumed to have a prior multi-Gaussian distribution
\[
\beta_i \sim N_p(\mu_0, \Sigma_0).
\]

The trend coefficients are assumed to be a priori independent of the residuals.

Different surfaces are coupled in two ways; they share some common trend maps and trend coefficients, and they have correlated residuals since they share some common layer thickness residuals. Covariances between two arbitrary surfaces $a$ and $b$ at two different lateral locations are
\[
\text{Cov}(z_i^a(x), z_i^b(y)) = f_i^a(x) \Sigma_0 f_i^b(y)
\]
\[
+ \sum_{k=0}^{\min(a,b)-1} \text{Cov}(\xi_i^k(x), \xi_i^k(y)),
\]

since $\xi_i^k$ and $\xi_i^l$ are assumed uncorrelated for $k \neq l$.

The framework above is sufficient to specify the multi-Gaussian distribution of any collection of surface depth values. We will use the symbol $\varphi(\cdot)$ for any (multi) Gaussian probability density function in the following.

\section*{4 Well observations}

Consider all $N$ locations in space along the wells where the stratigraphic layer changes. These are observations of the surfaces separating the stratigraphic layers and are referred to as well points. They are denoted
\[
w = \{[x_i, z_i, l_i]; \ i = 1, \ldots, N\},
\]

where $l_i$ is the surface index. For all points $[x_i, z_i, l_i] \in w$ we have that
\[
z_i^l(x_i) = z_i.
\]

Consider also a discrete sampling of the well path $w$ that excludes all well points. These are the white and blue discs in Fig. 1(c). These sample points are denoted
\[
w^l = \{[x_i^l, z_i^l, l_i^l]; \ i = 1, \ldots, N^l\},
\]

where $l_i^l$ is the layer index. Sample points $[x_i^l, z_i^l, l_i^l] \in w^l$ impose the inequalities constraints
\[
z_i^l(x_i^l) < z_i^l \quad \text{and} \quad z_i^l < z_i^{l+1}(x_i^l).
\]

Note that every sample point imposes constraints on all $L$ surfaces; all surfaces above layer $l_i^l$ must be shallower than $z_i^l$ and all surfaces below layer $l_i^l$ must be deeper than $z_i^l$. It is however sufficient to consider only the constraints on the two surfaces confining a layer since all surfaces are correlated so a constraint on one surface will influence all other surfaces. Therefore, we will only consider the two constraints in Eq. 5. This is illustrated in Fig. 2.

Sample points are chosen as a subset of the original well samples. To control surfaces, a sampling density equal to the grid resolution is sufficient. Since surface control points are mostly needed in horizontal or near horizontal sections of the well, the sampling is made as a function of the horizontal component of measured depth rather than as a function of measured depth. This way, vertical wells get no sample points except at the end of the well, while horizontal well sections are sampled at grid resolution. Sample points are not chosen close to well points unless they are at the very end of the well. For multilateral wells, samples should be

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig2.png}
\caption{Cross section along a well path that intersects surface $l$ at $z_i^l(x_i) = z_i$ (green disc). Well sample points, which are indicated by blue discs, constrain surface points above and below the path. These constrained surface points are marked as red discs.}
\end{figure}
identical up to the branching point since only one of these sets are needed.

Sample points near the surfaces will impose strong restrictions on the surface depth whereas sample points far away from surfaces will not impose any restriction. In practice, therefore, only a subset of all the sample points will suffice to guarantee that surfaces are consistent with the well paths.

5 Conditional probability density function

Let \( z = \{ z^{I_1}(x_1), \ldots, z^{I_N}(x_N) \} \) be any collection of surface values. The moments of the distribution \( \varphi(z \mid w) \) are given by the standard kriging equations in Eqs. 26 and 27.

Our goal is to obtain the conditional probability density function

\[
f(z \mid w, w^I).
\]

Conditioning on the inequality constraints \( w^I \) makes this a non-Gaussian density. This implies that efficient approaches such as kriging for calculating predictions and prediction uncertainties cannot be applied. In the following, we will make a representation of this function that allows standard techniques such as linear prediction (kriging) and conditional simulation to be exploited. Let

\[
z^I = \{ z^{I_1}(x^I_1), z^{I_1+1}(x^I_1); i = 1, \ldots, N^I \}
\]

be the collection of all surface points located above or below the sample points. These are the surface points illustrated as red discs in Fig. 2. The probability density function Eq. 6 can be reformulated as

\[
\begin{align*}
f(z \mid w, w^I) &= \int \int f(z, \beta, z^I \mid w, w^I) \, dz \, d\beta \\
&= \int \int f(z \mid w, \beta, z^I) f(\beta, z^I \mid w, w^I) \, dz \, d\beta \\
&= \int \int \varphi(z \mid w, \beta, z^I) f(\beta, z^I \mid w, w^I) \, dz \, d\beta.
\end{align*}
\]

where \( \varphi(\cdot) \) denotes a (multi-)Gaussian probability density function. The second line follows from the definition of a conditional distribution and the probability density function \( \varphi(z \mid w, \beta, z^I) \) is Gaussian since \( z \) is independent of \( w^I \) given \( \beta \) and \( z^I \).

Notice that \( \varphi(z \mid w, \beta, z^I) \) is a multivariate Gaussian probability density function fully specified by the stochastic model described in Section 3. Explicit calculation of \( f(\beta, z^I \mid w, w^I) \) is difficult. Instead, we approximate this function using \( S \) simulated realizations \( \{ z^I_{1(s)}, \ldots, z^I_{S(s)} \} \) and \( \{ \beta_{1(s)}, \ldots, \beta_{S(s)} \} \) drawn from this distribution. The next sections explain an efficient way to draw these samples.

5.1 A Monte Carlo approximation to \( f(\beta, z^I \mid w, w^I) \)

A naive approach to draw samples from \( f(\beta, z^I \mid w, w^I) \) is to draw samples from \( \varphi(\beta, z^I \mid w) \) and use rejection sampling to satisfy the constraints imposed by \( w^I \). However, this will not work in practice since the number of constraints is large \((2N^I)\) and even a small rejection rate for each constraint would lead to a vanishing acceptance rate. This is illustrated in Table 1 for a few examples. A second problem is that \( \varphi(\beta \mid w) \) can be very different from \( f(\beta \mid w, w^I) \) reducing the acceptance rate even further. The efficiency problem caused by \( \beta \) is solved by using an expectation-maximization-like algorithm called the Data Augmentation Algorithm [31] and by replacing the rejection sampling at the core of this algorithm by a Gibbs sampler. This sampler is explained in detail in Appendix A.

The Data Augmentation Algorithm uses the following approximation that converges as the number of simulated samples \( S \) increases

\[
f(\beta \mid w, w^I) \approx \frac{1}{S} \sum_{s=1}^{S} \varphi(\beta \mid w, z^I_{(s)}),
\]

where the conditional expectation and covariances of \( \varphi(\beta \mid w, z^I_{(s)}) \) is given by Eqs. 24 and 25. The algorithm is a fixed point iteration where the probability densities, including \( \beta \), becomes more accurate at every iteration.

Data Augmentation Algorithm

1: \( f^{(0)}(\beta \mid w, w^I) = \varphi(\beta \mid w) \) \hspace{1cm} Initialize \( \beta \) distribution
2: \( \text{for } n = 1 \text{ to } N_f \text{ do } \) \hspace{0.5cm} \( N_f \) fixed point iterations
3: \( \hspace{1cm} \text{for } s = 1 \text{ to } S \text{ do } \) \hspace{1cm} Draw \( S \) samples
4: \( \hspace{2.5cm} \text{Draw } \beta_{(s)} \approx f^{(n-1)}(\beta \mid w, w^I) \)
5: \( \hspace{2.5cm} \text{Draw } z^I_{(s)} \sim \varphi(z^I \mid \beta_{(s)}) \) until \( z^I_{(s)} \) satisfy \( w^I \)
6: \( \hspace{1cm} \text{end for} \)
7: \( \hspace{1cm} f^{(n)}(\beta \mid w, w^I) \approx \frac{1}{S} \sum_{s=1}^{S} \varphi(\beta \mid w, z^I_{(s)}) \)
8: \( \text{end for} \)
9: \( \text{Return } \{ z^I_{1(s)}, \ldots, z^I_{S(s)} \} \) from line 5 and \( \{ \beta_{1(s)}, \ldots, \beta_{S(s)} \} \) from line 4

The Data Augmentation Algorithm generates samples from \( f(\beta, z^I \mid w, w^I) \). It is initialized by drawing \( \beta \) independently of the constraints. The algorithm then alternates between drawing \( z^I \) dependent on \( \beta \) (line 5) and drawing \( \beta \) dependent on \( z^I \) (line 4). The constraints \( w^I \) are only included explicitly in the rejection sampling step when drawing \( z^I \).

If the number of constraints \( w \) is large, the impact of \( w^I \) is expected to be small and we can use \( f(\beta \mid w, w^I) \approx \varphi(\beta \mid w) \). In this case, line 4 in the algorithm can be replaced by \( \beta_{(s)} \sim \varphi(\beta \mid w) \) and the fixed point iterations become obsolete. This approach is numerically more stable and saves calculation time.
Experience shows that \( N_f = 20 \) fixed point iterations is sufficient for most practical applications, see Fig. 5. Increasing this number also increases the calculation time.

### 5.2 Conditional distribution

The Data Augmentation Algorithm generates \( S \) samples of \( \mathbf{z}' \) and \( \beta \) that can be used to represent the density \( f(\beta, \mathbf{z}' | \mathbf{w}, \mathbf{w}') \). This allows the double integral of Eq. 7 to be replaced by a sum of Gaussian probability density functions

\[
f(\mathbf{z} | \mathbf{w}, \mathbf{w}') = \int \int \varphi(\mathbf{z}|\mathbf{w}, \beta, \mathbf{z}') f(\beta, \mathbf{z}' | \mathbf{w}, \mathbf{w}') d\mathbf{z}' d\beta \\
\approx \frac{1}{S} \sum_{s=1}^{S} \varphi(\mathbf{z} | \mathbf{w}, \beta(s), \mathbf{z}'(s)).
\]

This approximation converges to the correct distribution as \( S \to \infty \).

### 5.3 Conditional simulation

Surfaces are normally represented on a grid with thousands or even millions of grid nodes. If \( \mathbf{z} \) represent the surface values at all grid nodes, the representation in Eq. 8 provides a straightforward simulation algorithm for \( f(\mathbf{z} | \mathbf{w}, \mathbf{w}') \)

\[
z \sim f(\mathbf{z}|\mathbf{w}, \mathbf{w}') \\
\sim \varphi(\mathbf{z}|\mathbf{w}, \beta(s), \mathbf{z}'(s)),
\]

where \( \beta(s), \mathbf{z}'(s) \) is one sample drawn from the \( S \) samples provided by the Data Augmentation Algorithm. This corresponds to one point from each of the point clouds in Fig. 1(d). We can draw from the conditional multi-Gaussian distribution Eq. 9 using standard methods of drawing samples from Gaussian random fields conditioned to point observations.

### 5.4 Prediction and prediction uncertainty

The conditional expectation is a commonly used predictor. The conditional expectation of Eq. 8 is

\[
z^* = \mathbb{E}[z | \mathbf{w}, \mathbf{w}'] \\
\approx \frac{1}{S} \sum_{s=1}^{S} \mathbb{E}[z|\mathbf{w}, \beta(s), \mathbf{z}'(s)].
\]

Since the density \( \varphi(\mathbf{z}|\mathbf{w}, \beta(s), \mathbf{z}'(s)) \) is multi-Gaussian, Eq. 10 is essentially a simple kriging equation system evaluated for every sample \( s = 1, \ldots, S \). The conditional variance is given by a similar Monte Carlo approximation

\[
\Sigma = \text{Var}(\mathbf{z} | \mathbf{w}, \mathbf{w}') \\
\approx \text{Var}(\mathbf{z}|\mathbf{w}, \beta(s), \mathbf{z}'(s)) \\
+ \frac{1}{S} \sum_{s=1}^{S} \left( \mu(s) - \bar{\mu} \right)^2, \tag{11}
\]

where \( \mu(s) = \mathbb{E}[z|\mathbf{w}, \beta(s), \mathbf{z}'(s)] \) and \( \bar{\mu} = \frac{1}{S} \sum_{s=1}^{S} \mu(s) \).

Note that \( \text{Var}(\mathbf{z}|\mathbf{w}, \beta(s), \mathbf{z}'(s)) \) is equal for all \( S \) samples. The conditional variance is a measure of the prediction uncertainty and becomes the standard kriging error if \( \mathbf{w}' \) contains no information.

In principle, \( \mathbf{z} \) can include all grid nodes, but since the expected surface depths only contain marginals, it is more efficient to repeat Eq. 10 for every grid node

\[
z^*(x) \approx \frac{1}{S} \sum_{s=1}^{S} \mathbb{E}[z(x)|\mathbf{w}, \beta(s), \mathbf{z}'(s)]; \quad \forall x \in \mathcal{G}, \tag{12}
\]

where \( \mathcal{G} \) is the collection of all grid nodes.

Similar to Eq. 12, the prediction uncertainty at every grid node is

\[
\sigma^2(x) \approx \text{Var}(z(x)|\mathbf{w}, \beta(s), \mathbf{z}'(s)) \\
+ \frac{1}{S} \sum_{s=1}^{S} \left( \mu(x) - \bar{\mu}(x) \right)^2, \tag{13}
\]

where \( \mu(s)(x) = \mathbb{E}[z(x)|\mathbf{w}, \beta(s), \mathbf{z}'(s)] \) and \( \bar{\mu}(x) = \frac{1}{S} \sum_{s=1}^{S} \mu(s)(x) \).

The Monte Carlo uncertainty in Eq. 12 is

\[
\text{Var}(\frac{1}{S} \sum_{s=1}^{S} \mathbb{E}[z(x)|\mathbf{w}, \beta(s), \mathbf{z}'(s)]) \\
= \frac{1}{S^2} \sum_{s=1}^{S} \text{Var}(\mathbb{E}[z(x)|\mathbf{w}, \beta(s), \mathbf{z}'(s)]) \\
= \frac{1}{S} \left( \text{Var}(z(x)|\mathbf{w}) - \text{Var}(z(x)|\mathbf{w}, \beta(s), \mathbf{z}'(s)) \right), \tag{14}
\]

which follows from independent samples and the identity \( \text{Var}(A) = \mathbb{E}[\text{Var}(A|B)] + \text{Var}(\mathbb{E}[A|B]) \).
Equation 14 shows that the Monte Carlo variance is inverse proportional to the sample size and proportional to the difference between the kriging variance without constraints and the kriging variance with the constraints. This difference is zero if there are no inequality constraints and reduces to the kriging variance at the constraints. This implies that an upper boundary for the Monte Carlo uncertainty is the standard kriging variance divided by the sample size, $\text{Var}(z(x)|w)/S$. In the examples below we have chosen $S = 256$, giving a Monte Carlo uncertainty of at most 6.25 percent of the standard kriging uncertainty.

### 5.5 An efficient approximation for prediction and prediction uncertainties

The predictions Eq. 12 and prediction uncertainties Eq. 13 are time consuming to calculate for every grid node. This can be solved by replacing the non-Gaussian density conditioned to constraints $w^I$ by a Gaussian density conditioned to synthetic help points $z_{\text{syn}}^I$ located at the constrained surface points

$$f(z|w, w^I) \approx \varphi(z|w, z_{\text{syn}}^I).$$

[3] proposed to select synthetic well points such that

$$E[z^I|w, z_{\text{syn}}^I] = E[z^I|w, w^I],$$

$$\text{Var}(z^I|w, z_{\text{syn}}^I) \approx \text{Var}(z^I|w, w^I).$$

The right hand sides of these equations are approximated using Eqs. 10 and 11 that collapses into simple sums since e.g. $E[z^I|w, \beta(s), z_{\text{syn}}^I(s)] = z_{\text{syn}}^I(s)$. The first requirement is trivially ensured by selecting $z_{\text{syn}}^I = z^I$ obtained from Eq. 10. This will however give zero variances at $z^I$ so the synthetic well points must be assigned a multi-Gaussian measurement error specified by a $2 \cdot N^I$-dimensional covariance matrix $\Sigma_{\text{crt}}$ [3] give explicit formulas for the expectation $E[z_{\text{syn}}^I]$ and the measurement error $\Sigma_{\text{crt}}$ that fulfills requirements Eqs 16 and 17.

The multi-Gaussian approximation Eq. 15 gives the following alternative approximations to Eqs. 12 and 13

$$z^*(x) = E[z(x)|w, w^I] \approx E[z(x)|w, z_{\text{syn}}^I],$$

$$\sigma^2(x) = \text{Var}(z(x)|w, w^I) \approx \text{Var}(z(x)|w, z_{\text{syn}}^I).$$

This is the conditional expectation and variance of a multi-Gaussian distribution given well points and the additional synthetic well points. Calculating the expectations and variances are done by traditional techniques such as kriging.

### 6 Synthetic examples

We shall investigate a synthetic example to illustrates the properties of the proposed method. The example uses the general exponential correlation function $\rho(h) = e^{-(h/R)^\alpha}$, with an isotropic range $R = 1500$ m and power $\alpha = 1.5$. The grid dimension is 5800 m by 3800 m, the grid resolution is 25 m and well paths are sampled every 100 m of lateral deviation. The sampling is coarser than recommended to make the pictures below easier to read.

For many well path sample points there are no corresponding constrained surface points. The reason for this is two-fold: First, constrained surface points are only added for well path sample points where the surface is at risk of crossing the well when doing a stochastic realization. Experience shows that such crossings are unlikely if $t = |z(x) - m(x)|/\sigma(x) < 2.5$, that is, when the distance between the surface and the well is less than 2.5 standard deviations of the surface uncertainty. For an individual point, this value compares to a probability of 98.76 percent of crossing the well. Secondly, many constraints are highly correlated and carry the same information, and it is sufficient to keep the constrained surface points where the inequality constraints are likely to be violated. The closer the well is to the surface the fewer constraints are removed.

#### 6.1 Two surfaces

Consider an anticline with two domes and two drilled wells, as shown in Fig. 3. Well w1 enters a geological layer from the top of the main dome and runs horizontally just below the saddle and leaves the second dome on its flank. Figure 4 shows a vertical cross section along well w1. The upper cross section shows two predicted surfaces, s1 and s5, conditioned on well points from both wells. Uncertainty bands of plus/minus one standard deviation are shown as grey. In the center cross section, surfaces have been conditioned to well points and constraining sample points. We can see that adding the constraints has pushed surface s1 correctly above the well and the uncertainty
bands have been reduced. Also note that the deeper surface s5 is strongly affected. It is pushed down by the leftmost constraining sample point and the mid section is pushed upwards because it is correlated to surface s1. The lower cross section of Fig. 4 shows one simulated realization.

According to Eq. 5, constrained surface points come in pairs, but this is not necessary in many cases. A simple criterion to evaluate a constrained surface point is to compute its distance to the well. If this distance is large compared to the surface uncertainty, then we remove it from the calculations. In the center cross section of Fig. 4, several of the pairs have been reduced to a single constraint; one red disc is lacking at locations of constraining sample points.

The convergence of one of the trend coefficients is shown in Fig. 5. Other trend coefficients show similar convergence. A grey uncertainty band is plotted as plus-minus one standard deviation of the trend coefficient. In this simple case we have convergence after a few fixed point iterations. The remaining fluctuations are due to Monte Carlo noise, and are significantly less than the standard deviation.

6.2 Five surfaces

The reservoir is the same as in the previous example, but three intermediate surfaces (s2, s3, s4) have been inserted between s1 and s5. The number of well points increased from two to eight for each well. The upper cross section of Fig. 6 shows the five surfaces conditioned to these well points. There are several segments along the well path where the well incorrectly crosses surface s3 and s4. The surfaces are correctly placed when they are also conditioned to constraining well samples (center cross section). As discussed in Section 4, constraints are only applied to the surfaces above and below the well path. The other surfaces adapt to these through the correlation structure Eq. 4. The lower cross section shows one simulated realization.

Figure 7 shows the depth uncertainty map of s3 when conditioned to well points (top) and additional constraining
Fig. 6 Cross sections through well w1 (double black line), showing surfaces s1, s2, s3, s4 and s5 (red lines), well points (green discs), constraining sample points (blue discs) and constrained surfaces points (red discs). Top figure: The surfaces have been conditioned to well points only. Center figure: The surfaces are conditioned to well points and constraining sample points. Bottom figure: One simulation of the surfaces, conditioned to well points and constraining sample points.

Fig. 7 Top: Depth uncertainty (SD in meters) of surface s3 when conditioned to well points. Bottom: Depth uncertainty of surface s3 when conditioned to well points and constraining sample points. The map is 5.8 km by 3.8 km.

sample points (bottom). The constraints clearly reduce the uncertainty along the well path.

7 Real case

We now turn to a real world example where data have been anonymized. The example has seventeen surfaces and fourteen wells, most of which are strongly deviated or horizontal. The grid dimensions are 15 000 m by 6 000 m and the resolution is 25 m giving a total of 145 000 grid cells for each surface. The total length of logged well sections is 74 km, and wells are sampled at grid resolution. All residuals are specified by a generalized exponential correlation function with isotropic range $R = 2000$ m and power $\alpha = 1.4$.

The upper part of Fig. 8 shows the predicted depths in a vertical cross section of a particular well. The surfaces are conditioned to the 215 available well points, eleven of which belongs to the well shown. The surfaces are seen to be inconsistent with the well path for large parts of the well, starting at the heel of the well and continuing for nearly two kilometers. Since the well runs inside a layer that is less than 5 m, it is impossible to get the layering correct without more information.

In the center part of Fig. 8, constraining sample points have been added to the calculation, and the predicted surfaces are now consistent with the well path. Notice that all surfaces are affected since they are highly correlated. The length of inconsistent layering is reduced from 3 700 m to 22 m for the fourteen wells, and all remaining errors are
The bottom part of Fig. 8 shows a stochastic simulation, see [5]. For all wells, in this particular realization, the total length of inconsistent layering is 68 m out of which 2 m has an error larger than one meter. One hundred stochastic simulations were run in 50 minutes. The average length of inconsistent layering in these simulations was 101 m. The worst simulation had 418 m of inconsistent layering, whereas the best simulation had only 18 m.

7.1 Comparing the Gibbs sampler to rejection sampling

A Gibbs sampler is used in line 5 of the Data Augmentation Algorithm. For comparison, the Gibbs sampler has been replaced by a naive rejection sampler that propose samples from a multivariate Gaussian distribution until all the constraints are satisfied. Table 1 compares the CPU times spent by the Gibbs and a rejection sampler for the examples in Sections 6 and 7. The rejection sampler failed to produce a single sample after three days of computation on the real example while the Gibbs sampler provided a sample in 7 seconds. So naive rejection sampling can be used for simple examples but fails for realistic cases with many constraints.

8 Discussion and closing remarks

A method for predicting and simulating a set of correlated geological surfaces conditioned to well points and inequality constraints has been presented. At the core of the approach is a resampling of the well to a finite set of constraints and the Data Augmentation Algorithm that finds a representation of the multivariate distribution of the surface depths at the constrained sample points. The multivariate distribution is a truncated version of a high-dimensional, highly correlated multi-Gaussian distribution. Drawing samples from this truncated distribution efficiently is challenging. Common trend coefficients impose a global dependency between all observations that is resolved by the iterative estimation scheme in the Data Augmentation Algorithm. An important step in the Data Augmentation Algorithm is rejection sampling from a multi-Gaussian distribution. Straightforward rejection sampling will not work since the acceptance rate is practically zero. Therefore, a very efficient Gibbs sampler that exploits the structure of the problem is used for drawing samples from the truncated distribution.

The efficiency of the methodology has been demonstrated on a number of real cases, see [30] and [21] for two case studies from the North Sea. In those cases, large surface grids are divided into subgrids that are handled semi-independently for better efficiency [32]. Further applications and practical details can be found in [6].
The methods proposed in this paper have been extended to model well path uncertainty by [10].

### Appendix A: Gibbs sampler for truncated multi-Gaussian distribution

The time consuming part of the Data Augmentation Algorithm is the rejection sampling in line 5 since acceptance rates are minute in realistic examples. The number of sample points and constraints can be hundreds or more. We aim thus at an efficient rejection sampler for a multivariate truncated Gaussian distribution, where the goal is to obtain a valid sample \( z' \). One possibility is to use the Gibbs sampler in [27]. The essential idea is to sequentially update each single variable conditioned on all other variables and loop through all variables many times.

Let \( z = z' \), \( z = \{z_1, \ldots, z_n\} \) be one sample \( s \) of the constrained surface points \( z' \) with size \( n = 2 \cdot N^I \). We define next an iteration \( z^k \) that approximates the sample \( z \), for every iteration in the Gibbs sampler. Starting with an initial sample \( z^0 = \{z_1^0, \ldots, z_n^0\} \), we obtain the sample \( k \) from the sample \( k-1 \) by sequentially drawing from the univariate distribution

\[
z_i^k \sim N_{z_i \in R_i} \left( E \left[ z_i^k \mid z_{\neq i}^k \right], \text{Var} \left( z_i^k \mid z_{\neq i}^k \right) \right)
\]

for \( i = 1, \ldots, n \). Here \( N_{z_i \in R_i} \) is the univariate truncated Gaussian distribution, truncated on some interval \( R_i = (-\infty, z_i^f) \) or \( R_i = (z_i^l, \infty) \) and the symbol \( i \) denotes all the indices except \( i \). The acceptance rate in each step will normally be quite low but can be improved by using the optimal accept-reject algorithm in [27]. The sequence of iteration updates could be arbitrary so we may update any permutation \( z_{j_1}, \ldots, z_{j_n} \) of the components \( z_1, \ldots, z_n \). A convenient permutation that takes into account the hardest constraints first improves the convergence of this sampler.

This univariate Gibbs sampler will in principal converge to the correct distribution but it suffers from extremely slow convergence since the constrained surface points \( z' \) are close laterally and are therefore highly correlated. The change in each iteration in Eq. 18 is therefore small and convergence is extremely slow. To improve on this we need to utilize the geometry of the problem.

#### A.1 Block Gibbs sampler

Define next the distance pairs

\[
d_i = \{d_{i}^+, d_{i}^-\}; \quad i = 1, \ldots, N^I,
\]

and collect them into the set of all pairs \( d = \{d_1, \ldots, d_{N^I}\} \).

The constraints Eq. 5 take on a simple form using the new variates, i.e., \( d_i \in R^+ = R^+ \times R^+ \), for all \( i \). We may thus equivalently draw from the distribution in terms of the distance variates

\[
f(d \mid w, d \in (R^+)^{N^I}).
\]

A bivariate Gaussian distribution with mean \( \mu \) and covariance \( \Sigma \) truncated on the set \( R^+ = R^+ \times R^+ \) will be denoted by \( N_{R^+}(\mu, \Sigma) \). The conditional mean and expectations of a pair \( d_i \) given well points \( w \) are denoted by \( \mu_i \) and \( \Sigma_i \).

We now present a block version of the Gibbs sampler Eq. 18 where variates are simultaneously updated in pairs. The pair strategy has proven to be way more efficient than the univariate strategy explained in Section A. Starting with initial realization pairs \( d^0 = \{d_1^0, \ldots, d_{N^I}^0\} \), the block Gibbs algorithm computes the sample \( k \) from the sample \( k-1 \) by sequentially drawing from the bivariate distribution

\[
d_i^k \sim N_{R^+} \left( E \left[ d_i \mid d_{\neq i}^k \right], \text{Var} \left( d_i \mid d_{\neq i}^k \right) \right).
\]

Again, the symbol \( i \) denotes all the indices except \( i \). The two-dimensional mean \( \hat{\mu}_i = E \left[ d_i \mid d_{\neq i}^k \right] \) and the \( 2 \times 2 \) covariance matrix covariance \( \hat{\Sigma}_i = \text{Var} \left( d_i \mid d_{\neq i}^k \right) \) are given by the well-known expressions

\[
\hat{\mu}_i = \mu_i + \Sigma_{i,j}^{-1} \Sigma_{i}^{-1} (\mu_j - \mu_j), \quad i,j = 1, \ldots, N^I, \quad \Sigma_i = \Sigma_{i,j}^{-1} \Sigma_{i}^{-1} \Sigma_{i,j}.
\]

Here \( \mu_i = E \left[ d_i \right] \) denotes the \( 2 \times 1 \) mean of the pair \( i \) and \( \Sigma_{i,j} = \text{Cov} \left( d_i, d_j \right) \) is the \( 2 \times 2 \) covariance matrix. The vector \( \mu_i \) excludes the index \( i \) from the full mean vector \( \mu = E \left[ \bar{d} \right] \). The matrix \( \Sigma_{u,v} \) is the part of the full covariance matrix \( \Sigma = \text{Var} (d) \) with indices \( u \in \mathbf{u} \) and \( v \in \mathbf{v} \).

The computation of the matrix \( \Sigma_{i,j}^{-1} \) of size \( (n - 2) \times (n - 2) \) can be expedited by pre-computing and storing the full precision matrix \( V = \Sigma^{-1}. \) We re-utilize parts of this matrix with the help of the following identity

\[
\Sigma_{i,j}^{-1} = V_{i,j} - V_{i,\ell} V_{\ell,j}^{-1} V_{\ell,i}.
\]

Paper [9] presents a satisfactory fast rejection sampler to draw from this type of distributions. The basic idea behind the fast truncated bivariate sampler is to divide the variable space into cases for which a conveniently chosen proposal distribution gives theoretical acceptance rates of 0.5.
A.2 Iterative approximation of the multivariate mean

The large number of dimensions and highly correlated points poses a threat to the convergence of the Gibbs sampler in Eq. 19 since the large number of iterations needed for convergence makes it prohibitively expensive. The Gibbs sampler requires therefore an initial state. We obtain that state by a (non-linear) iterative scheme whose objective is to approximate the mean of the truncated multivariate distribution of all distance pairs. The main idea is that the bivariate mean $\mathbf{\delta}$ of a truncated pair given the rest of the pairs obtained in a previous iteration is used as hard points that pose a threat to the convergence of the Gibbs sampler.

For each iteration, $k$, we compute $\mathbf{\delta}^k = \{\delta_0^k, \ldots, \delta_{N_L}^k\}$ as the mean of the positively truncated bivariate Gaussian distribution with untruncated mean. I.e., the bivariate Gaussian

$$\mathbf{X} = \mathbf{d}_1 | \mathbf{d}_2 = \delta_2^{k-1} \sim N(\mu_{\delta}, \Sigma_{\delta})$$

has the moments computed with formulas Eqs. 20–21 and

$$\delta_i^k = E[\mathbf{X} | \mathbf{X} \in \mathcal{R}^+] . \quad (22)$$

The mean of this bivariate truncated Gaussian variable is computed analytically using formulas in [28].

Our experience on applications indicates that 1000 fixed-point iterations is sufficient for a good approximation. In [17] each variable is updated using univariate conditional truncated variables. The authors show that a sufficient condition for this iterative scheme to converge is that the full precision matrix is diagonally dominant. This condition is not satisfied in our application. The authors show numerically that the iteration gives good approximations for the multivariate mean, even if diagonal dominance is not satisfied.

This exploratory analysis of our multivariate truncated distribution needs to be handled with care. Using a Gibbs sampler in Eq. 19 with the initial vector resulting from Eq. 22 underestimates the uncertainty. A remedy that produces better spread is to start the Gibbs sampler with higher conditional variances which accounts to adding random noise to the initial sample.

Appendix B: Estimation and prediction

The linear regression model Eq. 3 can be written in matrix form as

$$\mathbf{z} = \mathbf{F} \cdot \mathbf{\beta} + \mathbf{\epsilon} .$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}^1(\mathbf{x}_{1,1}) \\ \vdots \\ \mathbf{z}^1(\mathbf{x}_{1,N_1}) \\ \mathbf{z}^2(\mathbf{x}_{2,1}) \\ \vdots \\ \mathbf{z}^2(\mathbf{x}_{2,N_1}) \\ \vdots \\ \mathbf{z}^{L}(\mathbf{x}_{L,1}) \\ \vdots \\ \mathbf{z}^{L}(\mathbf{x}_{L,N_1}) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{f}^1(\mathbf{x}_{1,1}) \\ \vdots \\ \mathbf{f}^1(\mathbf{x}_{1,N_1}) \\ \mathbf{f}^2(\mathbf{x}_{2,1}) \\ \vdots \\ \mathbf{f}^2(\mathbf{x}_{2,N_1}) \\ \vdots \\ \mathbf{f}^{L}(\mathbf{x}_{L,1}) \\ \vdots \\ \mathbf{f}^{L}(\mathbf{x}_{L,N_1}) \end{bmatrix}, \quad \mathbf{\epsilon} = \begin{bmatrix} e^1(\mathbf{x}_{1,1}) \\ \vdots \\ e^1(\mathbf{x}_{1,N_1}) \\ \vdots \\ e^2(\mathbf{x}_{2,1}) \\ \vdots \\ e^2(\mathbf{x}_{2,N_1}) \\ \vdots \\ e^{L}(\mathbf{x}_{L,1}) \\ \vdots \\ e^{L}(\mathbf{x}_{L,N_1}) \end{bmatrix},$$

$$\mathbf{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_P \end{bmatrix} .$$

B.1 Estimating the trend coefficients

The best linear unbiased estimator (BLUE) for the coefficients is the generalized least squares (GLS) estimator [22, p. 172]

$$\mathbf{\beta}_{\text{GLS}} = (\mathbf{F}' \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{F}' \mathbf{K}^{-1} \mathbf{z} , \quad (23)$$

$$\hat{\Sigma}_{\text{GLS}} = (\mathbf{F}' \mathbf{K}^{-1} \mathbf{F})^{-1} ,$$

where the kriging matrix $\mathbf{K}$ is the covariance matrix

$$\mathbf{K} = \text{Var} (\mathbf{z}) = \text{Var} (\mathbf{\epsilon}) ,$$

with elements given by Eq. 4.

A Bayesian estimate is obtained by specifying the prior means and the prior covariances in the prior $P$-dimensional Gaussian distribution for the coefficient values

$$\mu_0 \sim N_P (\mu_0, \Sigma_0) .$$

The Bayesian estimate for the posterior expectations and covariances are

$$\hat{\mathbf{\beta}}_{\text{Bayes}} = \mathbf{\mu}_0 + \Sigma_0 \mathbf{F}' (\mathbf{K} + \mathbf{F} \Sigma_0 \mathbf{F}')^{-1} (\mathbf{z} - \mathbf{F} \mu_0) , \quad (24)$$

$$\hat{\Sigma}_{\text{Bayes}} = \Sigma_0 - \Sigma_0 \mathbf{F}' (\mathbf{K} + \mathbf{F} \Sigma_0 \mathbf{F}')^{-1} \mathbf{F} \Sigma_0 . \quad (25)$$

This estimate is robust for any number of surface observations, $N$, including zero. If the prior uncertainty vanishes, $\Sigma_0 \to 0$, we recover the prior mean $\mu_0$. It can also be shown, with reasonable assumptions, that if $\Sigma_0 \to \infty$, we obtain the GLS estimate [26].

B.2 Conditional distribution given surface observations

The kriging predictor for a surface $l$ is

$$\hat{\mathbf{z}}_l(\mathbf{x}) = \hat{\mathbf{m}}_l(\mathbf{x}) + \mathbf{k}_l(\mathbf{x}) \mathbf{K}^{-1} (\mathbf{z} - \hat{\mathbf{m}}) , \quad (26)$$

where the estimated means are obtained from the coefficient estimates as $\hat{\mathbf{m}}_l(\mathbf{x}) = \mathbf{f}_l(\mathbf{x}) \cdot \hat{\mathbf{\beta}}$ and $\hat{\mathbf{m}} = \mathbf{F} \cdot \hat{\mathbf{\beta}}$. For universal kriging, with unknown trend coefficients, the GLS estimates Eq. 23 are used for the coefficients, whereas
for Bayesian kriging, the Bayesian estimates Eqs. 24–25 are used for the coefficients. The covariance vector $\mathbf{k}(x)$ is the covariance between the location of interest and all the surface observations: $\mathbf{k}(x) = \text{Cov}(Z'(x), z)$, where covariances are given by Eq. 4.

The Bayesian prediction error is given by

$$
\sigma^2(x) = \sigma^2(x) + \sum f_i(x) \Sigma_0 f_i(x) - \left( \mathbf{k}'(x) + \sum f_i(x) \Sigma_0 \mathbf{F} \right) \left( \mathbf{K} + \mathbf{F} \Sigma_0 \mathbf{F} \right)^{-1} \left( \mathbf{k}'(x) + \sum f_i(x) \Sigma_0 \mathbf{F} \right) \cdot f^i(x).
$$

(27)

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Abrahamsen, P.: Bayesian kriging for seismic depth conversion of a multi-layer reservoir. In: Soares, A. (ed.) Geostatistics Tróia ’92, pp. 385–398. Kluwer Academic Publ., Dordrecht (1993). https://doi.org/10.1007/978-94-011-1739-5_31
2. Abrahamsen, P.: A review of Gaussian random fields and correlation functions. Report 917, Norwegian Computing Center, PO.Box 114 Blindern, N-0314 oslo norway (1997)
3. Abrahamsen, P., Benh, F.E.: Kriging with inequality constraints. Math. Geol. 33(6), 719–744 (2001). https://doi.org/10.1023/A:1011078716252
4. Abrahamsen, P., Dahle, P., Hauge, V.L., Almendral-Vazquez, A., Vignes, M.: Surface prediction using rejection sampling to handle non-linear constraints. Bulletin of Canadian Petroleum Geology 63(4), 304–317, 12 (2015). ISSN 0007-4802. https://doi.org/10.2113/gscpgbull.63.4.304
5. Abrahamsen, P., Kvernerv, V., Barken, D.: Simulation of gaussian random fields using the fast fourier transform (fft). In: Proceeding of ECMOR XVI - 16th European Conference on the Mathematics of Oil Recovery, pp. 1–14, EAGE (2018). https://doi.org/10.3997/2214-4609.201802134
6. Abrahamsen, P., Dahle, P., Kvernerv, V., Sektman, A., Vázquez, A.A., Aarnes, I.: Cohiba User Manual, 2020. https://nr.no/en/industries/natural-resources/cohiba/
7. Armstrong, M., Galli, A., Beucher, H., Loc’h, G., Renard, D., Doligez, B., Rémi Eschard, F.G.: Plurigaussian Simulations in Geosciences. Springer-Verlag Inc., Berlin (2011)
8. Choi, K.-M., Christakos, G., Serre, M.L.: Recent developments in vectorial and multi-point bme. In: Buccioni, A., Nardi, G., Potenza, R. (eds.) Proceeding of IAMG’98, pp. 91–96. Isola d’Ischia, Naples (1998)
9. Chopin, N.: Fast simulation of truncated gaussian distributions. Statistics and Computing 21(2), 275–288 (2011). https://doi.org/10.1007/s11222-009-9168-1
10. Dahle, P., Almendral-Vazquez, A., Abrahamsen, P.: Simultaneous prediction of geological surfaces and well paths. In: Petroleum Geostatistics 2015, EAGE, pp 5 (2015). https://doi.org/10.3997/2214-4609.20141358
11. Diamond, P.: Interval-values random functions and the kriging of intervals. Math. Geol. 20(3), 145–165 (1988). https://doi.org/10.1007/BF00890251
12. Dubrule, O., Kostov, C.: An interpolation method taking into account inequality constraints. 1. Methodology. Math. Geol. 18(1), 33–51 (1986). https://doi.org/10.1007/BF00897654
13. Emery, X., Arroyo, D., Pelaez, M.: Simulating Large Gaussian Random Vectors Subject to Inequality Constraints by Gibbs Sampling. Math. Geol. 46(3), 265–283 (2014). https://doi.org/10.1007/s11004-013-9495-9
14. Freulon, X., de Fouquet, C.: Conditioning a gaussian model with inequalities. In: Soares, A. (ed.) Geostatistics Tróia ’92, pp. 201–212 (1993). https://doi.org/10.1007/978-94-011-1739-5_17
15. Fridley, B.L., Dixon, P.: Data augmentation for a Bayesian spatial model involving censored observations. Environmetrics 18(2), 107–123 (2007). https://doi.org/10.31274/RTD-180813-12076
16. Journel, A.: Constrained interpolation and qualitative information - the soft Kriging approach. Math. Geol. 18(3), 269–305 (1986). https://doi.org/10.1007/BF00898032
17. Konstantinos, D.K., Konstantinos, E.T., Athanasios, A.R.: On the efficient estimation of the mean of multivariate truncated normal distributions. arXiv:1307.0680 (2014)
18. Kostov, C., Dubrule, O.: An interpolation method taking into account inequality constraints. 2.Practical approach. Math. Geol. 18(1), 53–73 (1986). https://doi.org/10.1007/BF00897655
19. Langlais, V.: On the neighborhood search procedure to be used when kriging with constraints. In: Armstrong, M. (ed.) Geostatistics, pp. 603–614. Springer Netherlands, Dordrecht (1989). https://doi.org/10.1007/978-94-015-6844-9_47
20. Maatouk, H., Bay, X.: Gaussian Process Emulators for Computer Experiments with Inequality Constraints. Math. Geol. 49(5), 557–582 (2017). https://doi.org/10.1007/s11004-017-9673-2
21. Mannini, A., Pearse, S.: How big an elephant can be. In: 76th EAGE Conference & Exhibition 2014 Amsterdam RAI, The Netherlands, 16-19 June 2014, pp 1–5 (2014). https://doi.org/10.3997/2214-4609.20141503
22. Mardia, K.V., Kent, J.T., Bibby, J.M.: Multivariate Analysis. Academic Press Inc., London (1979)
23. Michalak, A.M.: A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling. Water Resources Res 44(9). https://doi.org/10.1029/2007WR006645 (2008)
24. Militino, A.F., Ugarte, M.D.: Analyzing censored spatial data. Math. Geol. 31(5), 551–561 (1999). https://doi.org/10.1023/A:1007516023962
25. Oliveira, V.D.: Bayesian inference and prediction of gaussian random fields based on censored data. Journal of Computational and Graphical Statistics 14(1), 95–115 (2005). https://doi.org/10.1198/106186005X27518
26. Omre, H., Halvorsen, K.B.: The Bayesian bridge between simple and universal kriging. Math. Geol. 21(7), 767–786 (1989). https://doi.org/10.1007/BF00893321
27. Robert, C.P.: Simulation of truncated normal variables. Statistics and Computing 5, 121–125 (1995). https://doi.org/10.1007/BF00143942
28. Rosenbaum, S.: Moments of a truncated bivariate normal distribution. Journal of the Royal Statistical Society. Series B (Methodological) 23, 405–408 (1961). http://www.jstor.org/stable/2984029
29. Stein, M.L.: Prediction and inference for truncated spatial data. Journal of Computational and Graphical Statistics 1(1), 91–110 (1992). https://doi.org/10.2307/1390602
30. Stenerud, V.R., Kallekleiv, H., Abrahamsen, P., Dahle, P., Skorstad, A., Viken, M.H.A.: Added value by fast and robust conditioning of structural surfaces to horizontal wells for real-world reservoir models. In: Proceedings of the SPE ATCE, page 8 (2012). https://doi.org/10.2118/159746-MS

31. Tanner, M.A., Wong, W.H.: The calculation of posterior distributions by data augmentation (with discussion). J. Amer. Statist. Assoc. 82(398), 528–550 (1987). https://doi.org/10.2307/2289457

32. Vigsnes, M., Kolbjørnsen, O., Hauge, V.L., Dahle, P., Abrahamsen, P.: Fast and accurate approximation to kriging using common data neighborhoods. Math. Geol. 49(5), 619–634 (2017). https://doi.org/10.1007/s11004-016-9665-7

33. Wellmann, J.F., Caumon, G.: 3-D Structural geological models: Concepts, methods, and uncertainties, volume 59 of Advances in Geophysics, pp. 1–121. Elsevier, Amsterdam (2018). https://publications.rwth-aachen.de/record/750586

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.