Selected Aspects of Multiple-Point Statistics

Julián M. Ortiz (jmo1@ualberta.ca)
Department of Civil & Environmental Engineering
University of Alberta

Abstract

This paper discusses several aspects of the inference and use of multiple-point statistics in geostatistical simulation. We briefly review the problems encountered when integrating this information with other sources. The issues of stationarity and representativeness of the high-order statistics are discussed.

Several methods are proposed to integrate multiple-point statistics extracted from a training image or dense production data to the conventional sequential indicator simulation. Some exploratory examples are presented and we discuss the problems encountered during the implementation of these methods. The property of an estimate to be non-convex is explained and some attempts at comparing the performance of the methods are shown.

Introduction

Integration of multiple-point (MP) statistics into geostatistical models is difficult because the inference of these statistics is often unreliable and their use in a kriging-like framework requires the positive semi-definite modelling of covariances between MP events and single-point events. This modelling would call for the knowledge of the multivariate distribution of the variable.

One way to overcome these problems is to use a training image deemed representative of the phenomenon being studied [1, 3, 8, 16, 17]. This raises different problems: the representativeness of this training image, the scale of the training image, the grid definition of the simulated model, and the univariate distribution of the training image. One concern is the amount of information deemed general to the phenomenon and the amount considered particular to the training image. The goal is not to reproduce the training image, but to simulate a model that shares some of the multivariate characteristics of the true distribution.

The paradigm of the extended normal equations, really a single normal equation, solves most of these problems [8, 16]. A conditional distribution function is calculated that considers the MP configuration in a neighborhood. In general, this is done by discretizing the conditional distribution in the indicator framework. The probability of that particular location to be below a threshold is given by the frequency of occurrence of that MP configuration in the training image. Therefore, there is no kriging system to solve. The multivariate distribution is being approximated by the frequencies extracted from the training image. If

a MP configuration is not found enough times to reliably estimate its frequency, the search is reduced and a different MP configuration is considered.

The question of the representativeness of the training image remains. The only apparent way to avoid this problem is to use MP statistics extracted from the actual data [10, 14]. Borrowing the MP information from a training image is equivalent to using the variogram from a different area, deposit, or reservoir to build a numerical model that reproduces the spatial continuity of the phenomenon. Expert judgement is used.

In most geostatistical techniques the multivariate distribution is commonly assumed multivariate Gaussian. Consequently, long range and non-linear connectivity is not reproduced in the numerical model. Responses after a transfer function are not well reproduced [7]. Transfer functions can be sensitive to this high order correlation or connectivity, that is, continuity of high and low values in the model. In mining, the mine design, mine plan, and grade control could change as the high order correlation is better reproduced. Better reproduction of multiple-point statistics at point support may allow block averaged values to follow more closely the true block distribution, improving grade control and recoverable reserves. Common practice is to assume that the simulation method provides a realistic representation of the connectivity.

The commonly used multi-Gaussian approach freezes the connectivity at different thresholds. A novel approach to avoid this assumption and also avoid the use of multiple-point statistics has been proposed recently. A sequential simulation can be performed considering calibration parameters that allow choosing the isofactorial family that better approximates the bivariate law of the data. The connectivity of extremes is then controlled by the isofactorial family considered [6]. Although this method appears as an improvement in controlling the connectivity of extremes, it can be considered as a parametric approach to indicator kriging, where the estimation of the values of the conditional distribution at several thresholds is replaced by a distributional assumption (not necessarily multi-Gaussian) that imparts correlation at extreme thresholds in a controlled manner. However, there is no explicit control over higher-order statistics.

If the data present a multivariate distribution that departs from multi-Gaussian, multiple-point statistics have to be explicitly imposed in the simulation algorithm.

A short discussion on the representativeness of the training images follows. Then, the issue of the scale of the model and statistical inference are presented. Some techniques to update the conditional probability in an indicator framework are proposed and some examples are shown. Some attempts are made to compare the performance of the proposed methods. Finally, the important concept of non-convexity of the estimator of the conditional probability is discussed.

About Training Images

Representativeness

Training images are deemed representative of the multivariate spatial characteristics of the phenomenon under study. This assertion is required to use the statistics extracted from the training image into the geostatistical model. There is, however, little chance to check this assumption, and the decision of using or not a training image is left to the expert's

judgement.

In general, the training image shares *some* aspects with the variable of interest, although the final goal is to extract from it, through statistical inference, some of the *key* features. The goal of using a training image is not to reproduce exactly its features, but to extract its *essence*. Which features are *key* and which are not? There is not a clear answer to this question and the answer is left again to the expert judgement of the modeler. In general, the objective of using a training image is to obtain models that lead to better decision making. This cannot be directly expressed as a set of mathematical constraints during the process of inferring multiple-point statistics from the training image, so heuristic methods are used.

In practice, not all the features of the training image are captured mainly because of the method used to define the conditional probabilities. For example, in the *single normal equation simulation* (`snesim`) [16] locations are visited in random order and at every point, the conditioning data (sample data and previously simulated nodes) are defined within a search neighborhood. The configuration of the conditioning data is used to scan the training image. If not enough replications of the geometric arrangement of the conditioning points are available for reliable statistical inference, then the farthest point is dropped and the training image is scanned for a reduced multiple-point configuration. Because the simulation path is random and the conditioning information is changed for inference, there is no specific set of multiple-point statistics explicitly honored at the end of the simulation. Each realization will have used a different set of statistics to condition the simulated points. For an infinitely large (ergodic) domain, the simulated models will tend to honor the same statistics, but these are not necessarily the same as the training image. Discrepancies may come from the fact that some configurations are absent in the training image.

Another approach to extracting the essential features of the training image has been proposed in a neural network context [2], where a validation is performed that permits to define an error function. The training image is split into a training and a validation set. The coefficients calculated to train the neural network are also checked against the validation set. If the neural network starts over fitting the training data, the error in the validation set increases and the training process is stopped. This procedure still seems questionable, since the validation is performed with more training data. One possible solution to this problem would be to validate the training image considering the sample data (if available) that will condition the simulated models. A cross-validation could be devised where each conditioning data is taken out of the data set in turn and the conditional distribution at that data location is estimated from the training image. This would require a fair amount of data to reliably compute the accuracy in the frequency of the conditional probabilities from the training image as compared with the data. Training images representing the same type of deposition environment could be tested to find the one that better fits the data. The construction of a MP statistics library is seen as a step towards improving the understanding of these statistics and to allow further research in this field [15].

Extracting multiple-point information from production data in a mining context is possible [10, 14]. This permits the construction of a data-driven model, where all the information is extracted from the data available and the use of a training image is avoided. Abundance of data is required and, realistically, petroleum applications will not benefit from this approach.

MP Information and Modelling Scale

When multiple-point information is available and used to infer high-order statistics, the scale at which the models are built must correspond to the resolution of this available information. If the simulated model required is denser that the resolution of the data or the training image used to infer the multiple-point statistics, then some assumption about the behavior of these statistics at distances smaller than their spacing must be considered. This poses another problem: positive semi-definiteness of the statistics is required and this implies some type of modelling of the multiple-point statistics. To clarify this idea, the problem is similar to inferring the variogram near the origin and its nugget effect in a mathematically consistent manner, in order to have variogram values for short distances (actually, the variogram values are needed only up to the resolution of the simulated grid). In the case of the variogram, there are known models that are licit in the sense that their positive semi-definiteness has been already proven. Linear combinations of licit variogram models are also licit, and this provides the flexibility to model almost any experimental variogram. There is no similar set of functions in the case of multiple-point statistics.

Statistical Inference and Stationarity

Statistical inference is a problem when dealing with multiple-point configurations, since the estimation of any statistic that simultaneously considers several points in space requires a large number of replications of that spatial configuration of points and values.

If enough data are available, frequencies are used for inference. The expected value of a given combination of values and locations is estimated by the frequency of that arrangement of values and locations found in some training set. This training set can be a training image or a data set. However, this does not ensure positive semi-definiteness.

As the size of the data sets become larger to allow inference, the issue of stationarity becomes more critical. We must assume all data belong to the same population and have a consistent behavior throughout the field, to perform statistical inference. This problem is also found when dealing with training images.

Updating Techniques to Account for Multiple-Point Information

Bayes' Law

If the information is to be integrated in cokriging, cross-covariances are required between the single-point data and the multiple-point information. Furthermore, direct covariances between different multiple-point configurations are also required. This makes the use of MP statistics unappealing. Some simplifications are required to allow easier integration.

The indicator framework works in terms of probabilities to be below a threshold. Consider the case where the probability of a variable Z to be below a threshold z_k at location $\mathbf{u}$ is of interest. We denote this event $\mathbf{A}$. Let all single point information be event $\mathbf{B}$, which includes the sample data and previously simulated nodes used to calculate the conditional probability at $\mathbf{u}$ (event $\mathbf{A}$) as in indicator kriging. If multiple-point information is available

(event $\mathbf{C}$), we seek a method to integrate these diverse sources of information to infer the conditional probability of $\mathbf{A}$ given $\mathbf{B}$ and $\mathbf{C}$.

Of course, this approach can be generalized to any number of events of different type (support, dimension, precision) [10].

Bayes' law gives a formalism to calculate this conditional probability, but it requires the knowledge of the joint probability of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{P(\mathbf{A},\mathbf{B},\mathbf{C})}{P(\mathbf{B},\mathbf{C})} \tag{1}$$

The multivariate distributions $P(\mathbf{A},\mathbf{B},\mathbf{C})$ and $P(\mathbf{B},\mathbf{C})$ are difficult to infer, thus their use is avoided in practice. Three simplifying assumptions can be considered:

- Independence between $\mathbf{B}$ and $\mathbf{C}$.
- Conditional independence between $\mathbf{B}$ and $\mathbf{C}$ given $\mathbf{A}$, which leads to the permanence of ratios or Naive-Bayes assumption [11, 14].
- A multi-Gaussian approximation to find the conditional probability $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ as a linear combination of the two independent conditionals $P(\mathbf{A}|\mathbf{B})$ and $P(\mathbf{A}|\mathbf{C})$ [10].

Assumption of Independence Between Multiple Events

The assumption of independence between $\mathbf{B}$ and $\mathbf{C}$ permits the calculation of the joint probability $P(\mathbf{B},\mathbf{C})$ as the product of the independent probabilities:

$$P(\mathbf{B},\mathbf{C}) = P(\mathbf{B}) \cdot P(\mathbf{C})$$

The expression for the conditional probability of $\mathbf{A}$ given $\mathbf{B}$ and $\mathbf{C}$ is simplified to (see **Appendix** for derivation):

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{P(\mathbf{A}|\mathbf{B})}{P(\mathbf{A})} \cdot \frac{P(\mathbf{A}|\mathbf{C})}{P(\mathbf{A})} \cdot P(\mathbf{A}) \tag{2}$$

The assumption of independence between the events $\mathbf{B}$ and $\mathbf{C}$ can be relaxed and conditional independence can be assumed instead.

Permanence of Ratios Assumption or Naive-Bayes Model

Assuming conditional independence between $\mathbf{B}$ and $\mathbf{C}$, we can write the following expressions:

$$P(\mathbf{B}|\mathbf{A},\mathbf{C}) = P(\mathbf{B}|\mathbf{A})$$
$$P(\mathbf{C}|\mathbf{A},\mathbf{B}) = P(\mathbf{C}|\mathbf{A})$$

The expression for the conditional probability of $\mathbf{A}$ given $\mathbf{B}$ and $\mathbf{C}$ is simplified:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{P(\mathbf{A},\mathbf{B},\mathbf{C})}{P(\mathbf{B},\mathbf{C})} = \frac{P(\mathbf{A}) \cdot P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{C}|\mathbf{A},\mathbf{B})}{P(\mathbf{B},\mathbf{C})} = \frac{P(\mathbf{A}) \cdot P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{C}|\mathbf{A})}{P(\mathbf{B},\mathbf{C})}$$

To get rid of the term for the joint probability of **B** and **C** without assuming independence between them, we can write the same expression, but now for the conditional probability of **A** not occurring, that is, of $\bar{\mathbf{A}}$, the complement of **A**:

$$P(\bar{\mathbf{A}}|\mathbf{B},\mathbf{C}) = \frac{P(\bar{\mathbf{A}}) \cdot P(\mathbf{B}|\bar{\mathbf{A}}) \cdot P(\mathbf{C}|\bar{\mathbf{A}})}{P(\mathbf{B},\mathbf{C})}$$

Now, taking the ratio between these two expressions and after some algebra (see **Appendix**), we obtain the permanence of ratios assumption [9]:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{\frac{1-P(\mathbf{A})}{P(\mathbf{A})}}{\frac{1-P(\mathbf{A})}{P(\mathbf{A})} + \frac{1-P(\mathbf{A}|\mathbf{B})}{P(\mathbf{A}|\mathbf{B})} \cdot \frac{1-P(\mathbf{A}|\mathbf{C})}{P(\mathbf{A}|\mathbf{C})}} \tag{3}$$

This expression does not require a prior knowledge of the relationships between **B** and **C**, that is, all conditional relationships are built based on the assumption that the incremental information provided by the event **B** regarding the event **A** is constant regardless of the conditioning event **C**.

Multi-Gaussian Assumption

The redundancy between the sources of information can be calculated if the multivariate spatial distribution of the variable is known. This is not the case in general; however, the full multivariate distribution is known in the multi-Gaussian case.

Under the multi-Gaussian assumption, a multiple-point covariance can be fully described as combinations of second-order covariances (see for example, [13]). Simple indicator kriging can be used to estimate the conditional expectation of the indicator variable. A multiple-point event can be expressed as the set of all the single-point events that constitute it. There is no need to work with multiple-point events in the multi-Gaussian case. Single points can be used and the conditional expectation can be estimated by simple kriging.

Although variables are not multi-Gaussian, this model can be used to approximate the redundancy term. In the context of integrating MP information, building an estimator that combines the estimate of the conditional probability at an unsampled location given different events could be considered. These events could be the indicator kriging (IK) estimate of the probability at **u**, and MP probabilities for different configurations. These conditioning events are noted as **B** and **C**, respectively. Notice that because the assumption of multi-Gaussianity is used, the events must be disjoint, in the sense that no one single point should belong to more than one event. Otherwise, the simple indicator kriging system of equations will be singular.

Assume $P(\mathbf{A}|\mathbf{B})$ is known from indicator kriging and $P(\mathbf{A}|\mathbf{C})$ can be obtained from a training image or from production data. We are interested in finding the weights ω_B and ω_C such that, *if the variable were multi-Gaussian*, the linear combination of $P(\mathbf{A}|\mathbf{B})$ and $P(\mathbf{A}|\mathbf{C})$ would yield the conditional probability $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ (known under multivariate Gaussianity).

The conditional probability of **A** can be written as a linear combination of the conditional probabilities calculated with each one of the conditioning events separately (see **Appendix** for further explanation):

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) = \omega_B \cdot (P(\mathbf{A}|\mathbf{B}) - P(\mathbf{A})) + \omega_C \cdot (P(\mathbf{A}|\mathbf{C}) - P(\mathbf{A})) \tag{4}$$

If event $\mathbf{B}$ is made of n_B single points and event $\mathbf{C}$ is a multiple-point configuration of n_C points, then, under the multi-Gaussian assumption, the joint distribution of both events can be calculated. The distribution is multivariate Gaussian of order $1 + n_B + n_C$ with mean and variance-covariance matrix:

$$\mu_{(1+n_B+n_C)\times 1} = \begin{pmatrix} P(\mathbf{A}) \\ \vdots \\ P(\mathbf{A}) \end{pmatrix}$$

$$\Sigma_{(1+n_B+n_C)\times(1+n_B+n_C)} = \begin{pmatrix} \Sigma_{A,A} & \Sigma_{A,B} & \Sigma_{A,C} \\ \Sigma_{B,A} & \Sigma_{B,B} & \Sigma_{B,C} \\ \Sigma_{C,A} & \Sigma_{C,B} & \Sigma_{C,C} \end{pmatrix}$$

Each sub-matrix corresponds to the covariance matrix between the single point events that constitute each event, for example:

$$\Sigma_{B,C} = \begin{pmatrix} Cov(\mathbf{u}_1^B, \mathbf{u}_1^C) & Cov(\mathbf{u}_1^B, \mathbf{u}_2^C) & \cdots & Cov(\mathbf{u}_1^B, \mathbf{u}_{n_C}^C) \\ Cov(\mathbf{u}_2^B, \mathbf{u}_1^C) & Cov(\mathbf{u}_2^B, \mathbf{u}_2^C) & \cdots & Cov(\mathbf{u}_2^B, \mathbf{u}_{n_C}^C) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\mathbf{u}_{n_B}^B, \mathbf{u}_1^C) & Cov(\mathbf{u}_{n_B}^B, \mathbf{u}_2^C) & \cdots & Cov(\mathbf{u}_{n_B}^B, \mathbf{u}_{n_C}^C) \end{pmatrix}$$

where $\mathbf{u}_i^B, i =, ..., n_B$ are the locations of the n_B single-point events used to obtain the indicator kriging estimate $P(\mathbf{A}|\mathbf{B})$, and $\mathbf{u}_j^C, j = 1, ..., n_C$ are the locations of the n_C points that constitute the multiple-point event $\mathbf{C}$ and that allow inferring the conditional probability $P(\mathbf{A}|\mathbf{C})$.

When considering the event $\mathbf{C}$, the conditional probability of $\mathbf{A}$ given $\mathbf{C}$ is calculated as the simple indicator kriging estimate at location $\mathbf{u}$ given the n_C samples that constitute the multiple-point event $\mathbf{C}$.

The weights ω_B and ω_C in **Equation 4** can be seen as a measure of closeness and redundancy between the events $\mathbf{B}$ and $\mathbf{C}$.

Notice that the goal is to use the conditional probabilities $P(\mathbf{A}|\mathbf{B})$ and $P(\mathbf{A}|\mathbf{C})$ obtained from different sources. The use of simple indicator kriging is meant to solve for the weights that yield the right conditional probability, *if the variable was multivariate Gaussian* and all the conditional probabilities were calculated by simple indicator kriging.

We calculate the conditional probability of $\mathbf{A}$ given a single event $\mathbf{B}$. This conditional probability can be seen as the indicator kriging estimate of $\mathbf{u}$, given the vector $\mathbf{i^B}$ of indicator values $i^B(\mathbf{u}_\alpha^B), \alpha = 1, ..., n_B$, or equivalently, as the linear regression estimate of the probability of $\mathbf{A}$ given $\mathbf{B}$:

$$P(\mathbf{A}|\mathbf{B}) - P(\mathbf{A}) = \Sigma_{A,B} \cdot \Sigma_{B,B}^{-1} \cdot (\mathbf{i^B} - P(\mathbf{A})) \tag{5}$$

Now, expressing the conditional probability of $\mathbf{A}$ given both events, that is, accounting for the redundancy between them, under the multi-Gaussian assumption:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) = \begin{pmatrix} \Sigma_{A,B} & \Sigma_{A,C} \end{pmatrix} \cdot \begin{pmatrix} \Sigma_{B,B} & \Sigma_{B,C} \\ \Sigma_{C,B} & \Sigma_{C,C} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{i^B} - P(\mathbf{A}) \\ \mathbf{i^C} - P(\mathbf{A}) \end{pmatrix}$$

The inverse of the covariance matrix can be written as:

$$\begin{pmatrix} \Sigma_{B,B} & \Sigma_{B,C} \\ \Sigma_{C,B} & \Sigma_{C,C} \end{pmatrix}^{-1} = \begin{pmatrix} D_1 & D_0 \\ D_0^T & D_2 \end{pmatrix}$$

with

$$\begin{aligned} D_1 &= (\Sigma_{B,B} - \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \cdot \Sigma_{C,B})^{-1} \\ D_0 &= -(\Sigma_{B,B} - \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \cdot \Sigma_{C,B})^{-1} \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \\ D_2 &= \Sigma_{C,C}^{-1} + \left((\Sigma_{B,B} - \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \cdot \Sigma_{C,B})^{-1} \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1}\right)^T \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \end{aligned}$$

That is,

$$\begin{aligned} P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) &= \left(\Sigma_{A,B} \cdot D_1 + \Sigma_{A,C} \cdot D_0^T\right) \cdot (\mathbf{i^B} - P(\mathbf{A})) + \\ &\quad (\Sigma_{A,B} \cdot D_0 + \Sigma_{A,C} \cdot D_2) \cdot (\mathbf{i^C} - P(\mathbf{A})) \end{aligned} \tag{6}$$

In order to determine the weights assigned to the conditional probabilities $P(\mathbf{A}|\mathbf{B})$ and $P(\mathbf{A}|\mathbf{C})$, we can rewrite the conditional probability on **Equation 4** using **Equation 5**:

$$\begin{aligned} P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) &= \omega_B \cdot \left(\Sigma_{A,B} \cdot \Sigma_{B,B}^{-1} \cdot (\mathbf{i^B} - P(\mathbf{A}))\right) + \\ &\quad \omega_C \cdot \left(\Sigma_{A,C} \cdot \Sigma_{C,C}^{-1} \cdot (\mathbf{i^C} - P(\mathbf{A}))\right) \end{aligned} \tag{7}$$

To identify the conditional probabilities of **Equations 6** and **7**, the approximate values for ω_B and ω_C can be defined:

$$\omega_B = \frac{\|\Sigma_{A,B} \cdot D_1 + \Sigma_{A,C} \cdot D_0^T\|}{\|\Sigma_{A,B} \cdot \Sigma_{B,B}^{-1}\|} \qquad \omega_C = \frac{\|\Sigma_{A,B} \cdot D_0 + \Sigma_{A,C} \cdot D_2\|}{\|\Sigma_{A,C} \cdot \Sigma_{C,C}^{-1}\|} \tag{8}$$

Because the goal is to combine two estimates, the weights must be numbers, not vectors. The ideal case to apply this method is when the vector $\Sigma_{A,B} \cdot D_1 + \Sigma_{A,C} \cdot D_0^T$ is proportional to $\Sigma_{A,B} \cdot \Sigma_{B,B}^{-1}$, and the vector $\Sigma_{A,B} \cdot D_0 + \Sigma_{A,C} \cdot D_2$ is proportional to $\Sigma_{A,C} \cdot \Sigma_{C,C}^{-1}$. In general, these pairs of vectors are not proportional to each other. The calculation of ω_B and ω_C under these circumstances is an approximation of the proportionality coefficient if they were proportional.

We can consider some extreme cases of the linear combination of estimates presented above. A first extreme case is to assume independence between the events. This amounts to setting the cross-terms $\Sigma_{B,C}$ and $\Sigma_{C,B}$ as 0. D_0 becomes zero, D_1 is reduced to $(\Sigma_{B,B})^{-1}$, and D_2 is simplified to $(\Sigma_{C,C})^{-1}$. Thus, $\omega_B = \omega_C = 1$. The estimate built as a linear combination is then just the sum of the two (independent) estimates of the conditional probability.

Independence of **B** and **C** is, in the context presented here, not a good assumption, since **C** is very close to **u** and **B** is correlated with **A**, therefore it should also be correlated with **C**.

Notice that this procedure to estimate weights that measure the redundancy between the sources of information can be extended to any arbitrary number of conditioning events [10].

In a general case, where the indicator variable does not come from a multivariate Gaussian distribution but where quantifying the redundancy between the two sources of information **B** and **C** appears impossible, the weighting proposed under the multi-Gaussian assumption could be used to approximate the solution.

Implementation

In the implementation of these methods a few problems may arise. First, the property of Bayesian updating under the independence assumption between the sources of information can lead to severe order relation deviations, since under this assumption, the updated probability can easily take values below zero and above one. This problem is not severe under the assumption of permanence of ratios and under the multi-Gaussian assumption to assess the relationship between the sources of information. These two methods show order relation deviation of the same magnitude as the estimation of the conditional probability by indicator kriging. Permanence of ratios gives an updated probability always within the interval [0,1], provided each component falls within this interval. In the case of approximating the redundancy through a multi-Gaussian assumption, order relations should be minor, since the weighting will prevent the estimated probability from departing largely from the interval [0,1]. Deviations should occur with approximately the same frequency as when indicator kriging is used.

A second problem is the inconsistency between the univariate distributions of the different sources of information. The methods assume stationarity: the proportions below the thresholds are the same no matter what source of information is considered. This is not always the case. Corrections are needed to solve this problem. The goal is to make the estimator unbiased. The simplest solution is to standardize the conditional probability from the source whose marginal distribution differs from the target distribution. Consider that the source from where $P(\mathbf{A}|\mathbf{B})$ is obtained has a marginal distribution such that $P(\mathbf{A})_{\mathbf{B}} \neq P(\mathbf{A})$, then the following standardized conditional probability is used:

$$P(\mathbf{A}|\mathbf{B})^* = P(\mathbf{A}|\mathbf{B}) \cdot \frac{P(\mathbf{A})}{P(\mathbf{A})_{\mathbf{B}}}$$

The methods for integrating different events have been implemented. Consider the following definitions for the events (**Figure 1**):

A : The event of having $Z(\mathbf{u}) \leq z_k$, where z_k corresponds to a threshold value.

B : The n_B single points used to inform the location **u**. These points are located in a neighborhood of **u**. They correspond to the data used to solve the simple indicator kriging system to estimate the conditional probability at **u**.

C : The multiple-point event formed by a set of n_C points, usually very close or adjacent to the location of interest **u**.

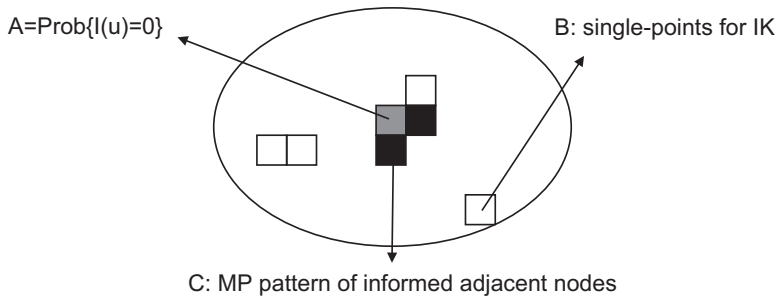

Figure 1: Example of three events. **A** corresponds to the probability of the grey node not to exceed a threshold value. Six single-point events are found in a search neighborhood of the node being estimated. They are illustrated as white and black nodes. **B** is formed by all the single point events that are not part of the multiple-point configuration that constitutes **C**. From the six single-points found in the search neighborhood, only the white nodes ($n_B = 4$ points) will be used to get the IK estimate, to avoid singularity due to full redundancy with the nodes in **C**. **C** is the multiple-point event formed by $n_C = 2$ points adjacent to the node whose conditional probability is being estimated. They appear as black nodes. The probability of **A** given the information in **C** is obtained either from a training image or from data.

B will help reproduce the long range structure of the variable, while **C** will locally modify the estimation of the probabilities to consider more complex multiple-point information. It will integrate some of the MP information to the indicator kriging estimator.

In the following examples, a set of very simple two-dimensional patterns are used and the probabilities are inferred from training images built to investigate the performance of each technique. The idea is to use only the four adjacent nodes to the one being estimated (**Figure 2**). The probabilities are associated with the frequencies of having those configurations in the training image. Since every node is coded as an indicator, for each p-point configuration there are 2^p combinations of zeros and ones possible. The total number of MP events from which the probabilities of occurrence must be obtained, is 81:

$$\sum_{i=0}^{4} \binom{4}{i} \times 2^i = 1 \times 1 + 4 \times 2 + 6 \times 4 + 4 \times 8 + 1 \times 16 = 81$$

If a combination of the indicators is not found in the training image, no updating will take place and the updated probability will correspond exactly to the one obtained by indicator kriging.

Examples

Several reference images have been built to have a training image for MP statistics, and for a visual reference. Four methods are used to generate unconditional realizations: sequential indicator simulation (`SIS`), updating under the independence assumption between **B** and **C**, updating under the permanence of ratios assumption, and combining estimates under

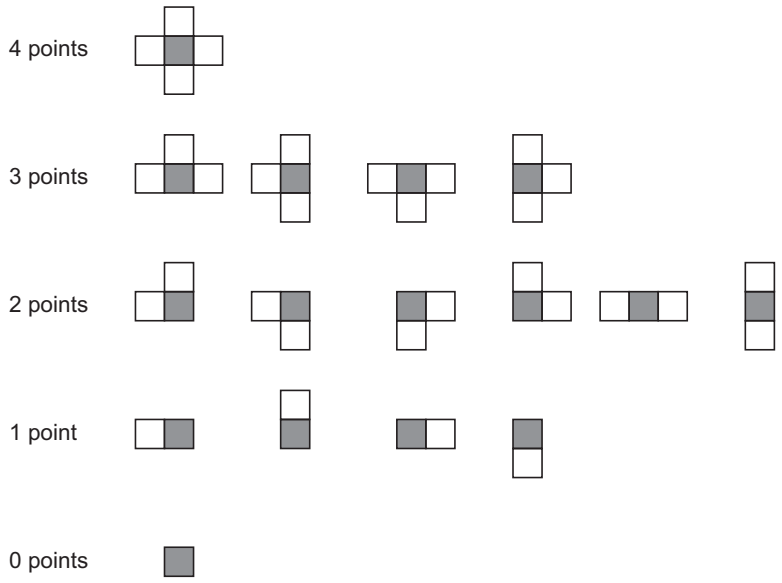

Figure 2: Multiple-point patterns with adjacent grid nodes. The gray node is the one being estimated. The patterns correspond to the four adjacent nodes to the node of interest. The probabilities are extracted from the training image even when some of the nodes are not informed, generating the three-, two-, one-, and zero-point patterns.

the multi-Gaussian assumption. All methods are simulated unconditionally and with the same random path for comparison.

Binary images are built with the object based algorithm `ellipsim` in `GSLIB` [5]. Different proportions are tested (p= 10, 50, and 90%) and cases with and without anisotropy.

This section also presents a binary example using training images of continuous variables and finally an example with multiple thresholds is shown.

Small Isotropic Objects Objects of radius equal to 3 units on a two-dimensional domain of 100 by 100 nodes, with a spacing of one unit are simulated as a reference. The four methods proposed are used to generate one realization of the phenomenon. **Figures 3**, **4**, and **5** show the results.

Large Isotropic Objects Larger objects were simulated, with a radius of 9 units, under the same condition as before. The results are shown on **Figures 6**, **7**, and **8**.

Small Anisotropic Objects Ellipses with an anisotropy at 30 degrees and major radius of 6 units and minor radius of 3 units were generated, again using different proportions (10, 50, and 90 %). The results are presented in **Figures 9**, **10**, and **11**.

Large Anisotropic Objects Finally, larger anisotropic ellipses were generated to extract the MP statistics and then simulated with the four methods. The major radius is 18 and the minor radius is 9 units. **Figures 12**, **13**, and **14** show the resulting maps.

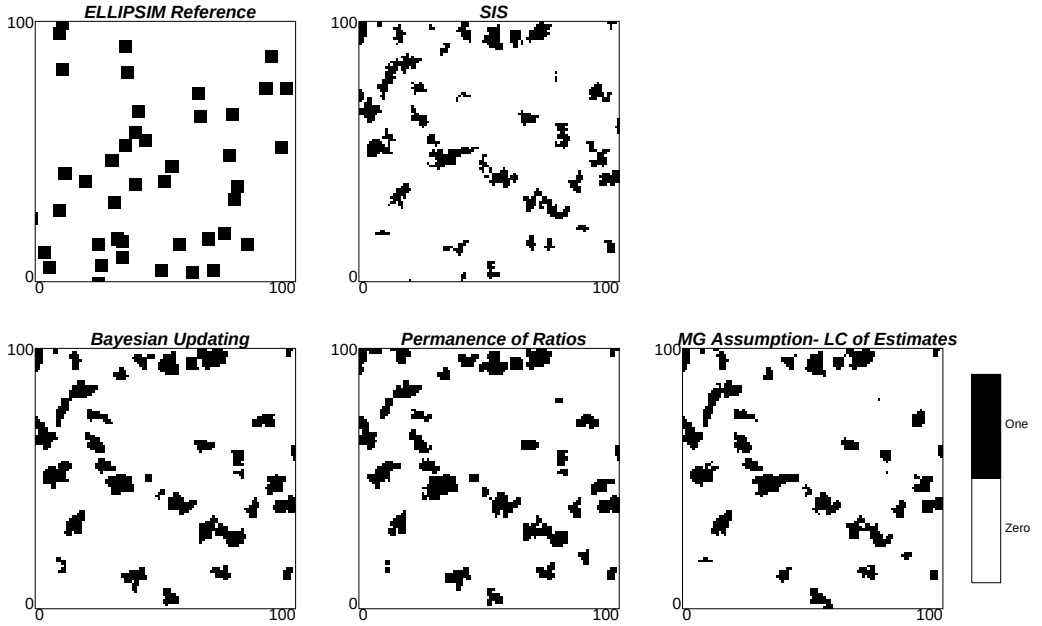

Figure 3: Maps of simulated values for small isotropic objects. Proportion above the threshold is 10 %

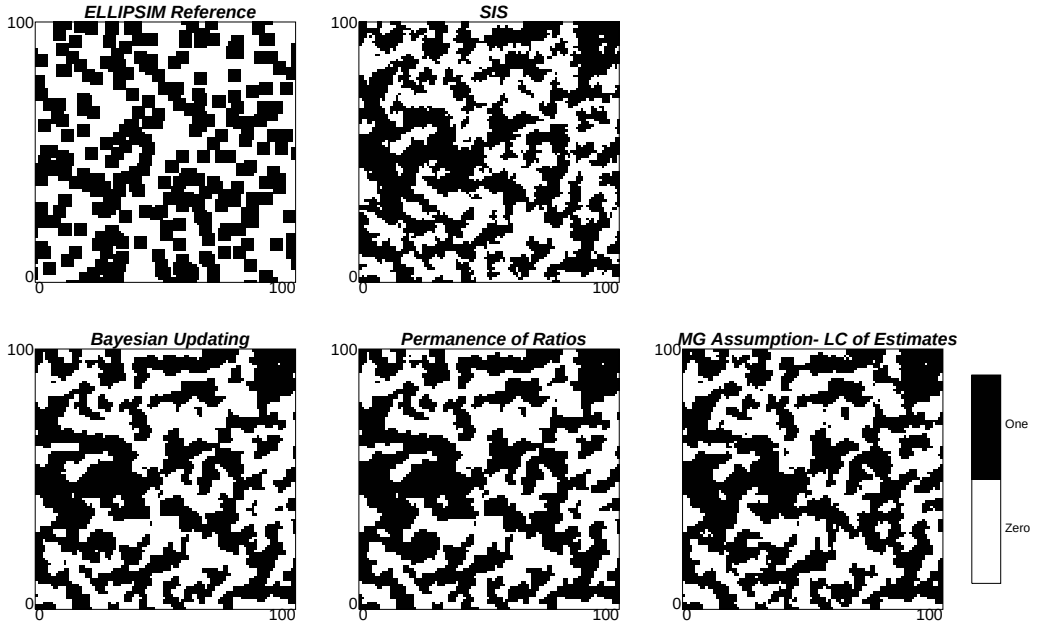

Figure 4: Maps of simulated values for small isotropic objects. Proportion above the threshold is 50 %

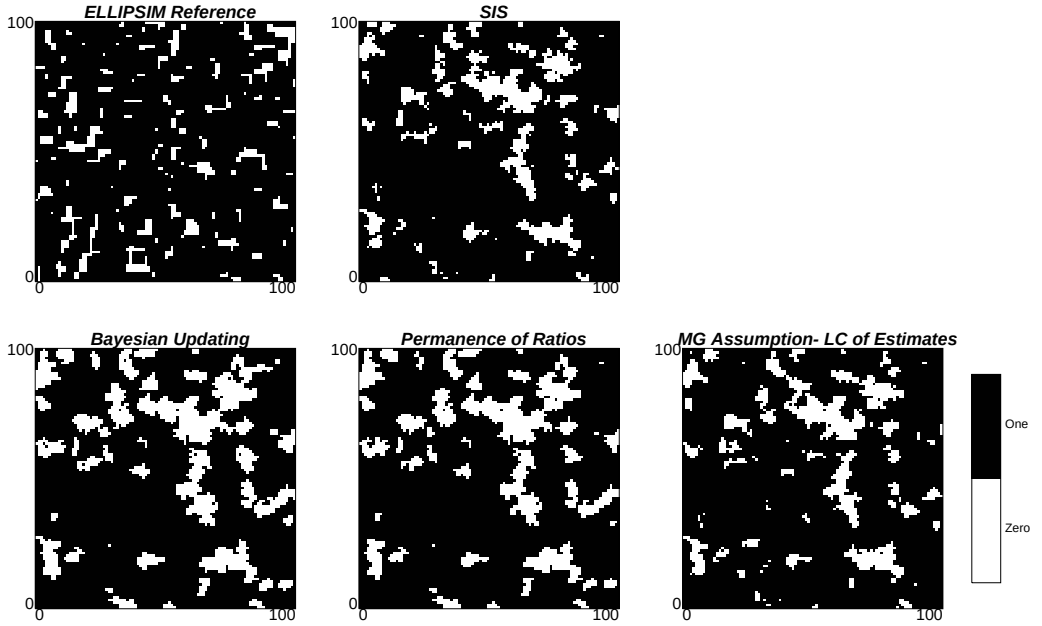

Figure 5: Maps of simulated values for small isotropic objects. Proportion above the threshold is 90 %

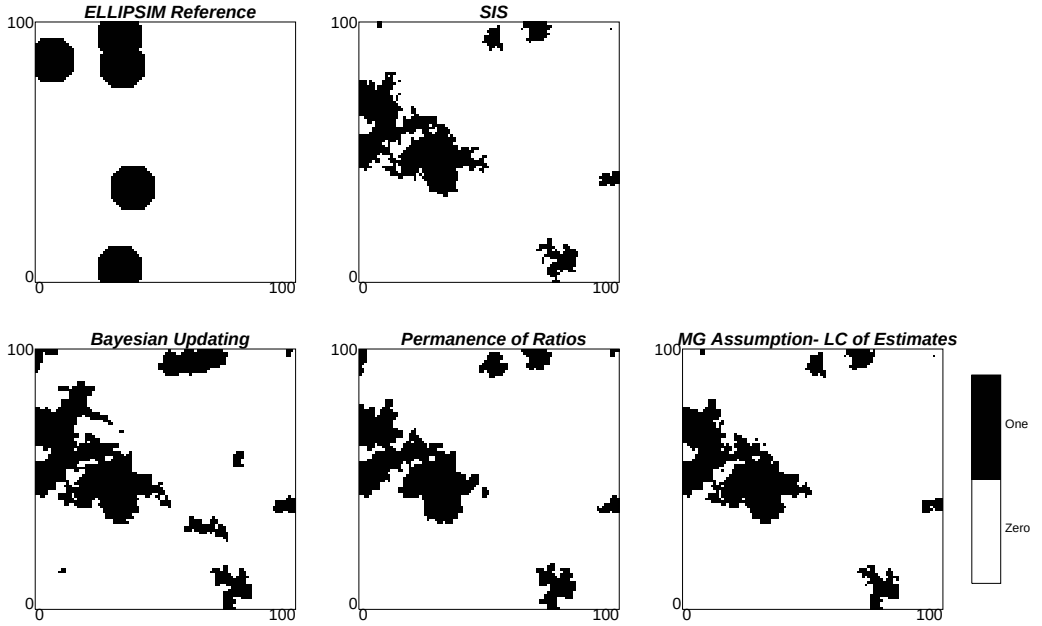

Figure 6: Maps of simulated values for large isotropic objects. Proportion above the threshold is 10 %

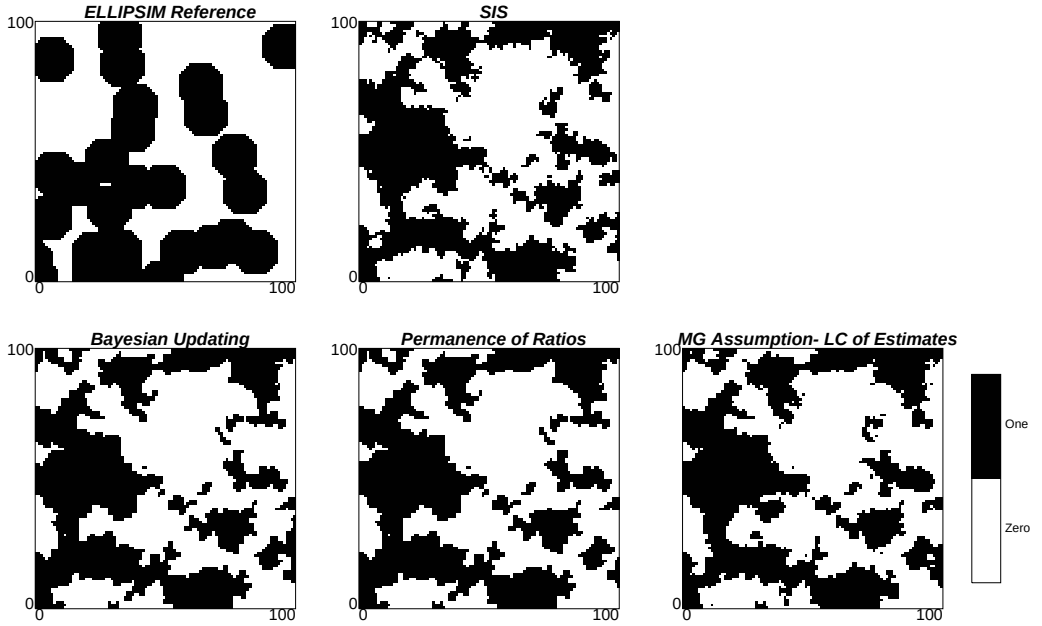

Figure 7: Maps of simulated values for large isotropic objects. Proportion above the threshold is 50 %

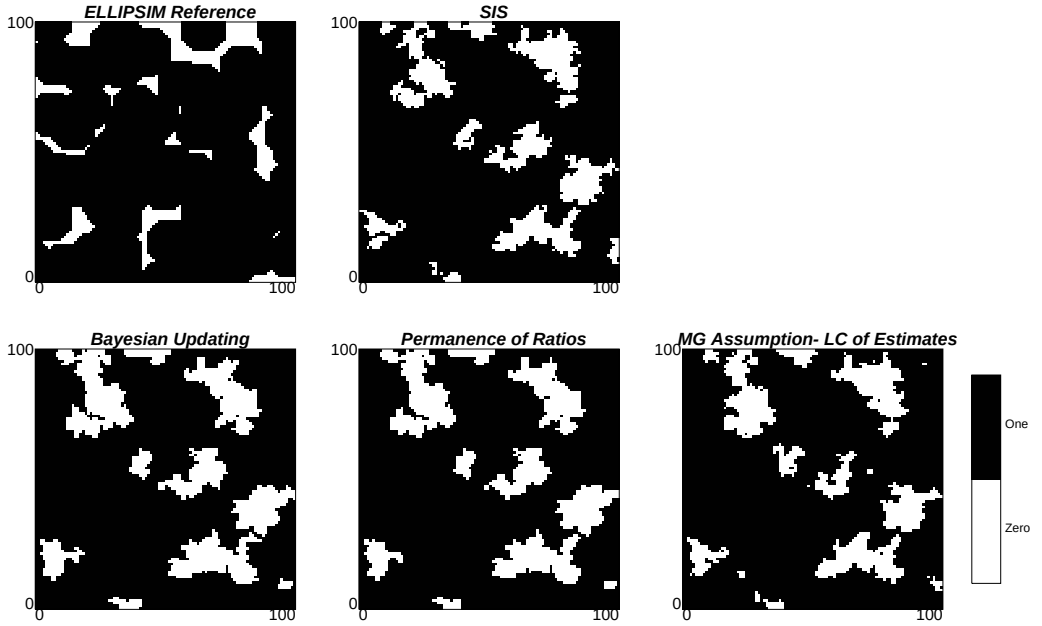

Figure 8: Maps of simulated values for large isotropic objects. Proportion above the threshold is 90 %

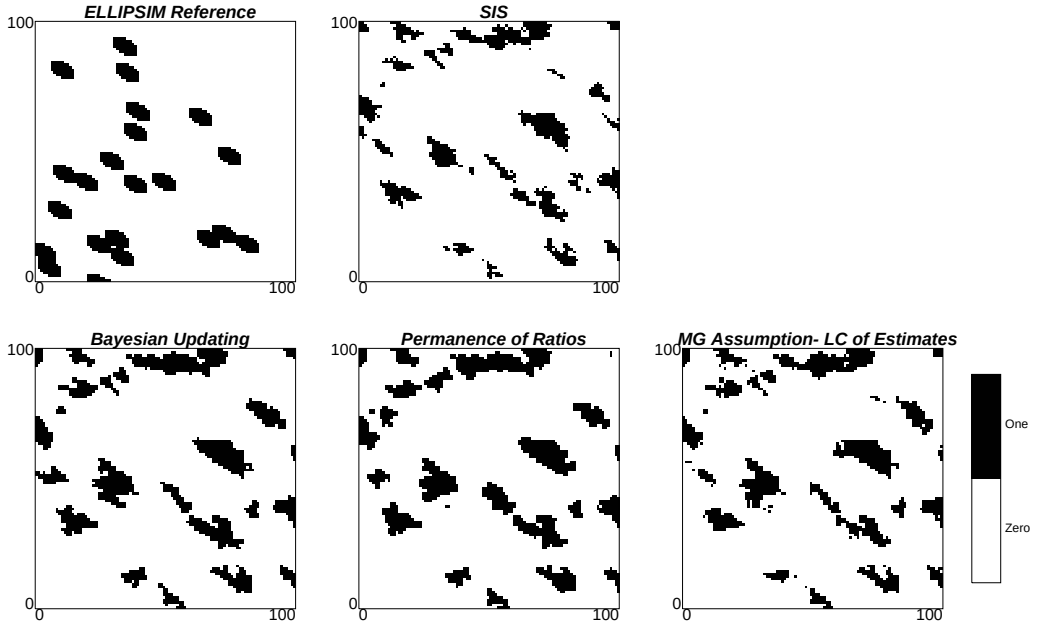

Figure 9: Maps of simulated values for small anisotropic objects. Proportion above the threshold is 10 %

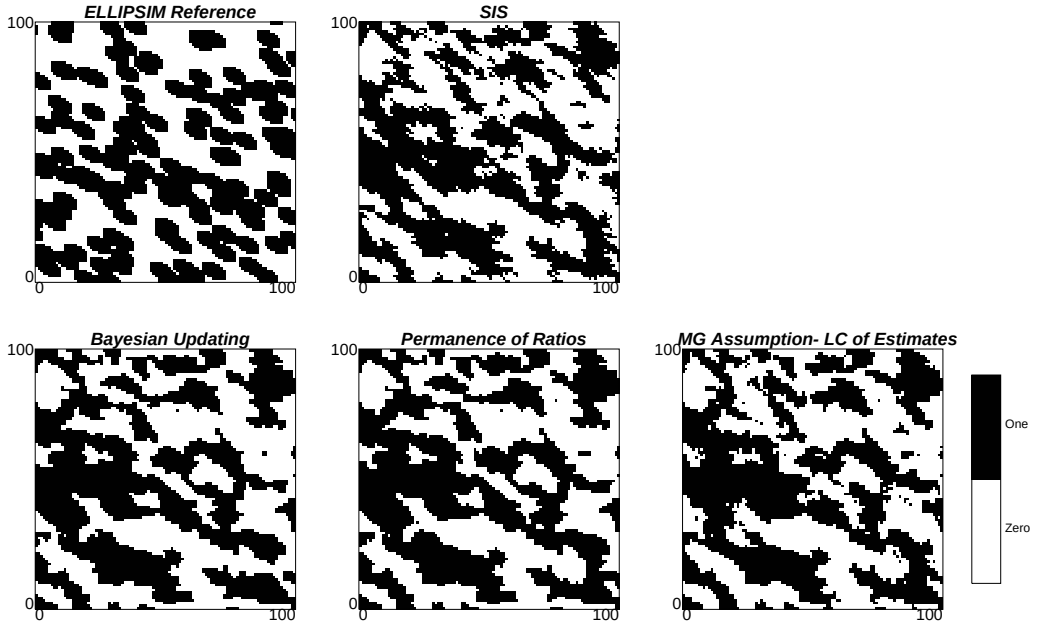

Figure 10: Maps of simulated values for small anisotropic objects. Proportion above the threshold is 50 %

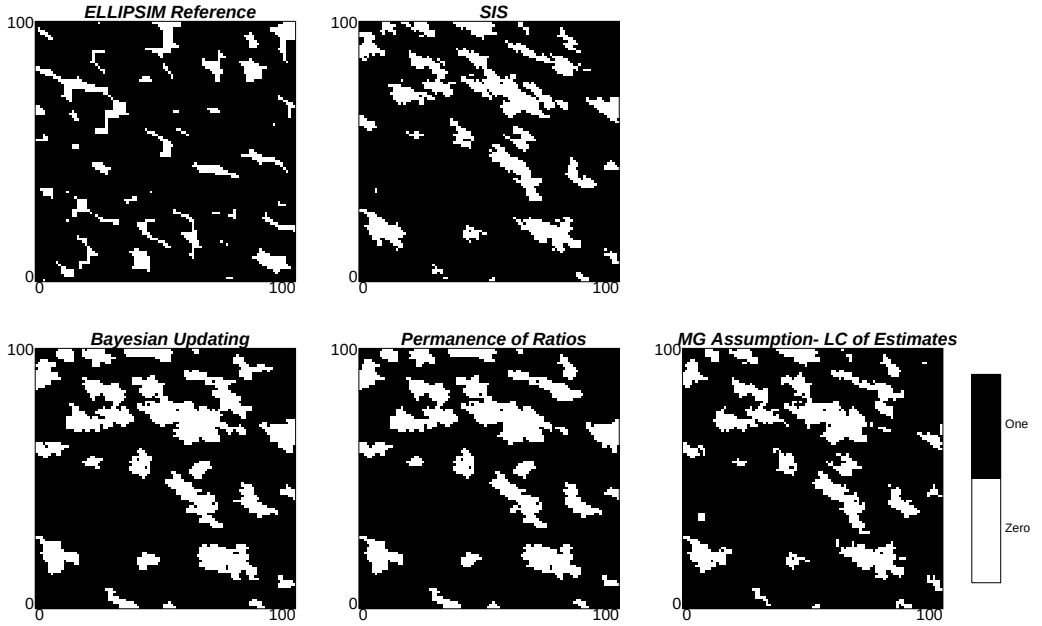

Figure 11: Maps of simulated values for small anisotropic objects. Proportion above the threshold is 90 %

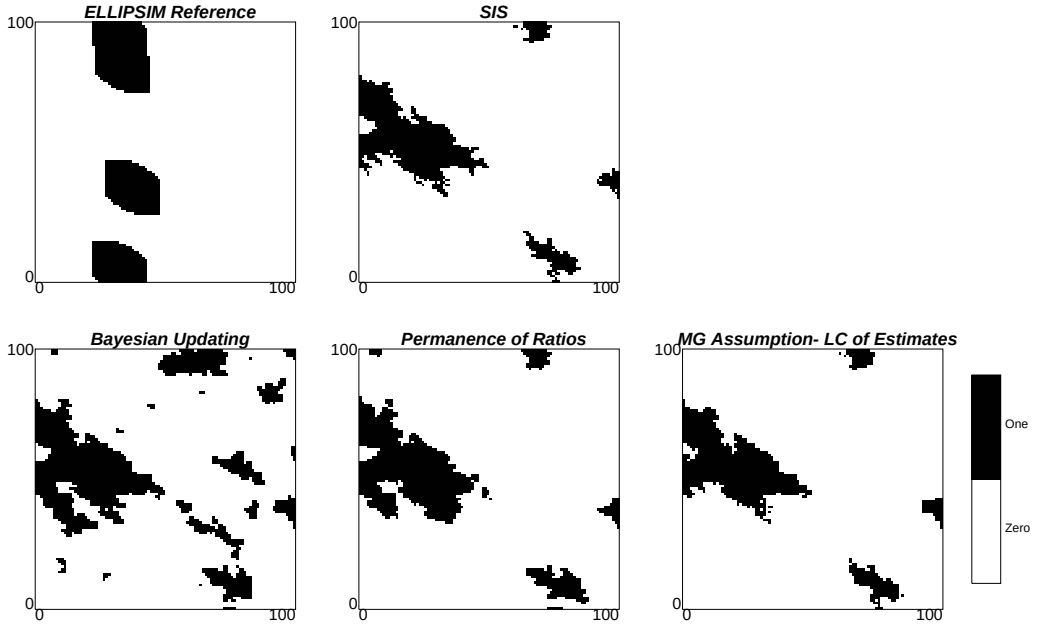

Figure 12: Maps of simulated values for large anisotropic objects. Proportion above the threshold is 10 %

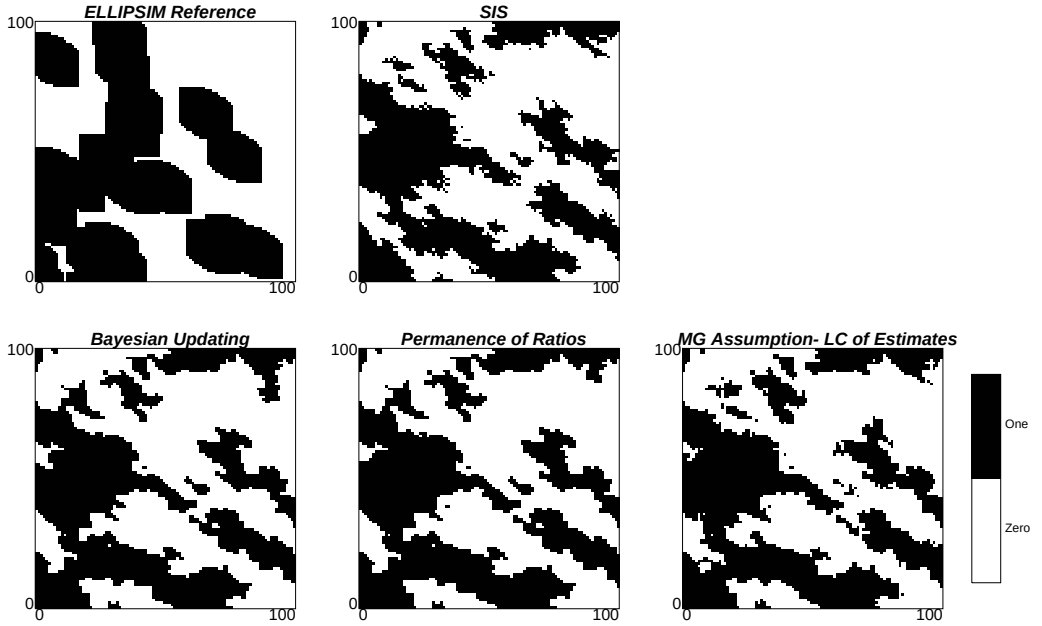

Figure 13: Maps of simulated values for large anisotropic objects. Proportion above the threshold is 50 %

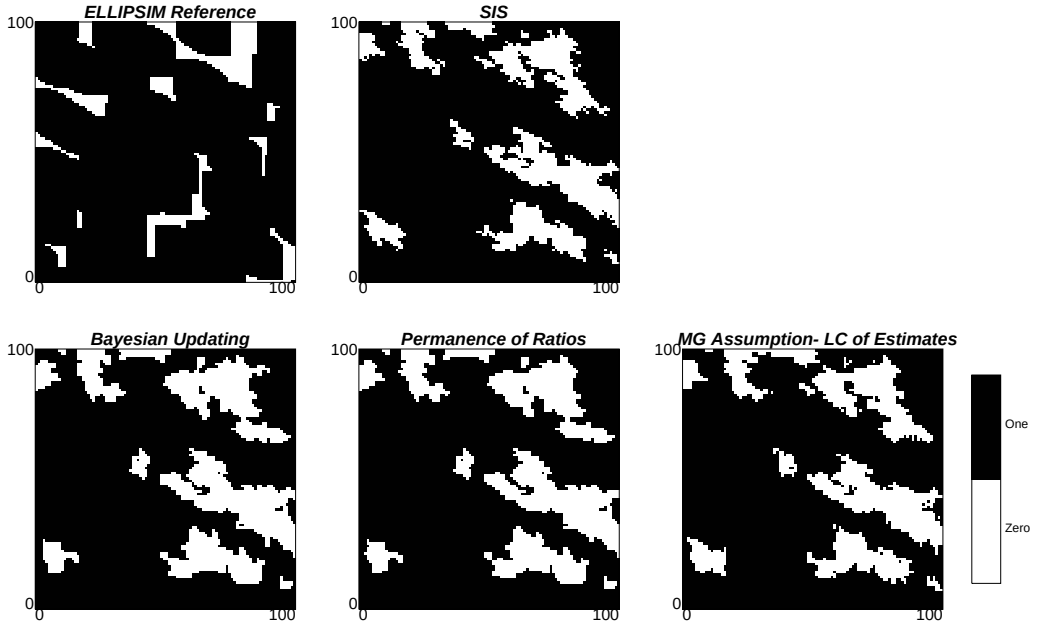

Figure 14: Maps of simulated values for large anisotropic objects. Proportion above the threshold is 90 %

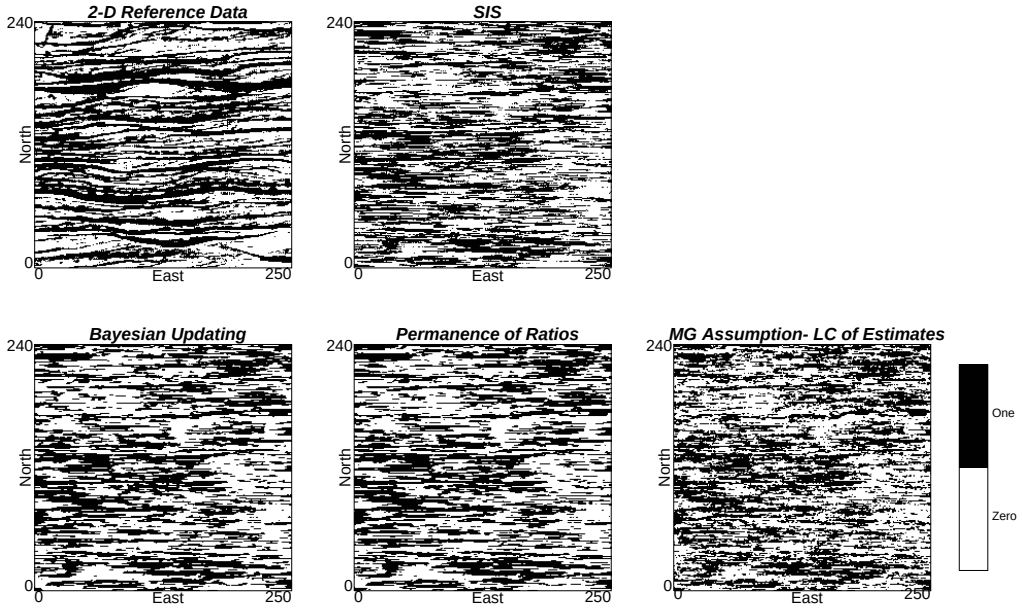

Figure 15: Maps of simulated values for a binary image taken from a continuous variable. Proportion above the threshold is 50 %

MP Statistics Extracted From a Continuous Training Image A training image showing a continuous variable is used in this example. The median threshold has been chosen to create a binary image. The probabilities of multiple-point events as shown on **Figure 2** are extracted to update the `SIS` algorithm. The results are shown on **Figure 15**. Although the curvilinearity is not captured by any of the methods, the layering that can be seen on the training image is also seen on the simulated models. The models obtained with Bayesian updating assuming data independence and permanence of ratios look too clean. These two methods erase all the noise that can be seen in the reference. On the contrary, the updating by combining the estimates of the conditional probability under the multi-Gaussian assumption seems to add too much noise.

Continuous Variable Example The same training image of a continuous variable used in the last example is now utilized to simulate using multiple thresholds. A variable with a positively skewed distribution has been characterized with 10 thresholds corresponding to each one of the 9 deciles, in addition to the quantile 0.95. The indicator variograms for each thresholds were calculated to perform sequential indicator simulation. The frequency of occurrence of the MP-patterns that have at least one adjacent node informed are extracted from the reference image. These statistics are then used to update the `SIS` simulated model. Again, all four methods were applied and the results are shown on **Figure 16**.

Discussion

There are several comments:

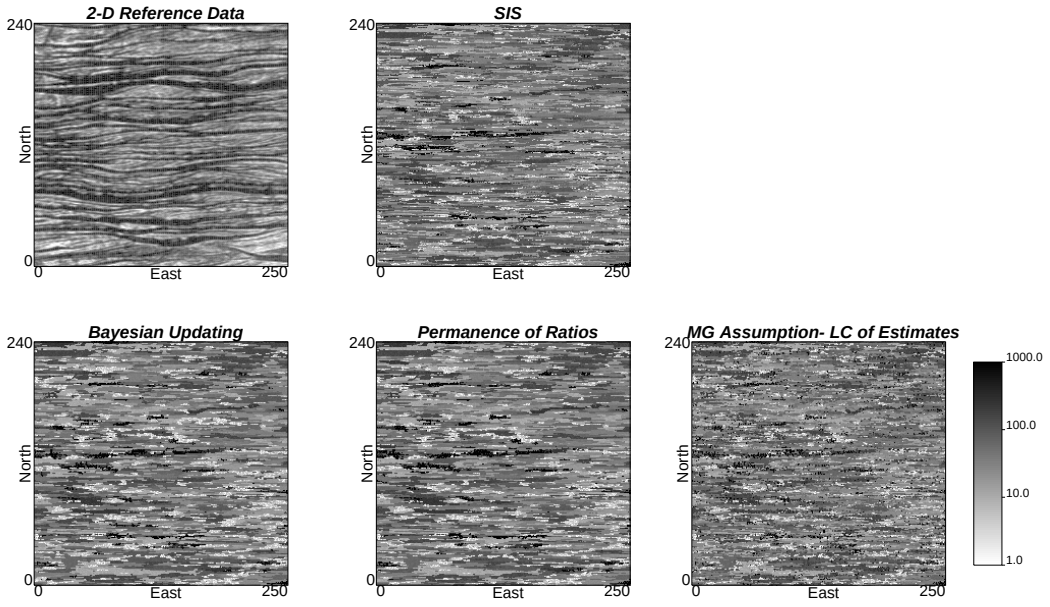

Figure 16: Maps of the continuous simulated values using all five methods for a continuous variable. Ten thresholds were used to characterize the spatial continuity of the variable.

- `SIS` generates a map with the correct long range continuity, but there seems to be too much noise in the short range: multiple-point statistics are clearly not reproduced. This problem is partially solved via image cleaning. Algorithms such as MAPS [4] reduce the excessive randomness, but without direct control of the multivariate distribution.
- Updating under the data independence assumption cleans the image. It looks like a cleaned `SIS` output [4]. The short-range anisotropic features of the `SIS` output look more isotropic than the training image, maybe because of the small size of the pattern used to extract the MP statistics, which does not reflects anisotropies very clearly. In general, the result is more similar to the training image.
- Updating under the permanence of ratios assumption also cleans the image. It is hard to judge which method gives the better result.
- The idea of combining both estimates of the conditional probability, the one calculated by indicator kriging of the data not belonging to the pattern used to infer the MP statistics, and the probability obtained from the MP pattern, performed well. It compares similarly to the algorithm where the updating is made under the data independence and permanence of ratios assumptions. However, it approximates the effect of redundancy between both sources of information, so this assumption should be more realistic.

The examples presented correspond to unconditional simulations. It was found in these examples that all the methods generated a bias on the probabilities below each threshold.

Algorithm	SIS	Full Indep.	Perm. of Ratios	MG Assumption
MSE	3.18E-05	1.98E-05	2.48E-05	4.78E-06
MAS	3.36E-02	3.85E-02	4.14E-02	2.71E-02

Table 1: Mean squared error and mean absolute error in the MP probability for the different algorithms.

This bias was always towards the median. Investigation of the bias showed that it is due to the correction of order relation deviations as discussed in [12].

The bias is greatly reduced if the following conditions are met [10]: (1) enough conditioning information is available; (2) the size of the simulated domain is large with respect to the range of correlation of the variogram; and (3) multiple grid search is used to simulate.

Assessing Performance

Comparing the methods is not an easy task. One straightforward approach would be to look at the histograms of MP probabilities obtained from each simulated model, compared with the reference probabilities. Classical measures of mismatch could be used, such as a mean squared error or mean absolute error. Unfortunately, not all the MP configurations have the same importance, so a small mismatch in an important configuration can pass unnoticed with a summary statistic of this kind.

This analysis was done to the binary example generated with a training image from a continuous variable; otherwise, we would need to determine the probabilities for each threshold and joint-probabilities between thresholds.

The probabilities of each one of the 81 MP configurations was calculated for the reference and for each one of the four maps generated with the different algorithms. For completeness, this should be done over multiple realizations with each algorithm, to avoid problems of ergodicity. The mismatch between the MP probability for each configuration from the training image and from each simulated map was calculated and plotted (**Figure 17**). The absolute value of this error is also presented in **Figure 18**. In this plot, the multi-Gaussian assumption appears as the closest to the target probabilities.

Interestingly, the graphs show quite clearly that the estimation of the conditional probability combining the IK estimate (from the data) and the MP probability estimate (from the training image) generates smaller errors than the other algorithms. This is also seen when looking at summary statistics, such as the mean squared error and the mean absolute error (**Table 1**). Cases where a given configuration of the multiple-points was not found are not easily comparable, since the impact of not having a configuration that is present in the target statistics cannot easily be quantified.

Implementation with real data should provide more insight regarding which method performs the best. The exploratory examples developed in this section only help to anticipate the improvement in performance of each one of the methodologies proposed.

Non-Convexity of Estimators

One of the good properties of the estimators, when integrating information from various sources is the possibility to be non-convex, that is, to generate probabilities outside the range of the input information. However in the kriging context, this is sometimes deemed inappropriate since the probabilities may go beyond the allowed range [0,1]. The idea is that the new method performs better than all the individual sources of information when estimating the conditional probability at a location of interest.

Bayesian updating under the independence and permanence of ratios assumptions gives a unique map of the updated probability given the marginal probability of the event of interest. That is, given $P(\mathbf{A})$, the map of $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ as a function of $P(\mathbf{A}|\mathbf{B})$ and $P(\mathbf{A}|\mathbf{C})$ is fixed (for example, see **Figure 19**).

The non-convexity can easily be obtained for these methods, by coding with a different color all the area of this map where the resulting probability $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ is outside the range defined by the two sources of information $P(\mathbf{A}|\mathbf{B})$ and $P(\mathbf{A}|\mathbf{C})$. These maps are presented in **Figure 20**. The non-convexity of all the methods can also be quantified during the simulation procedure. At every node, the three probabilities $P(\mathbf{A}|\mathbf{B})$, $P(\mathbf{A}|\mathbf{C})$, and $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ are known for all the updating methods. The proportion of the time in which the corresponding algorithm generates a result outside the range of the two input probabilities can be used as a measure of non-convexity.

Table 2 shows the result for the examples previously presented. In general, the linear combination of estimates under the multi-Gaussian assumption generated more estimates outside the range of the probabilities inferred from the two sources **B** and **C**. The high fraction of estimates outside the range of the input probabilities for the case of assuming independence of the sources of information is mostly due to the large order relations that this method generates. The non-convexity of this method is undermined by the corrections required to keep the probabilities in the allowable interval.

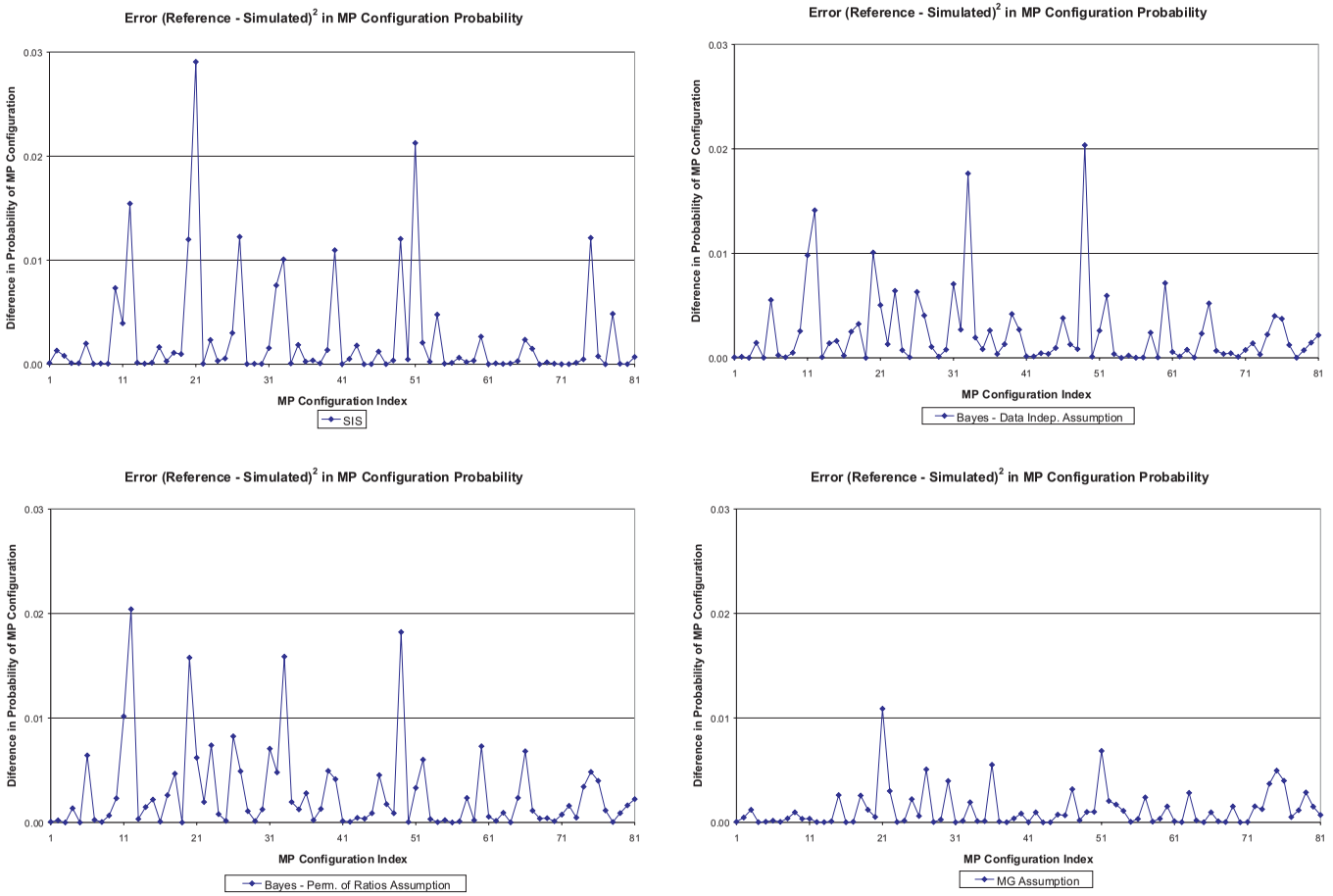

Figure 17: Mismatch in MP probability for all 81 MP configurations for four of the methods.

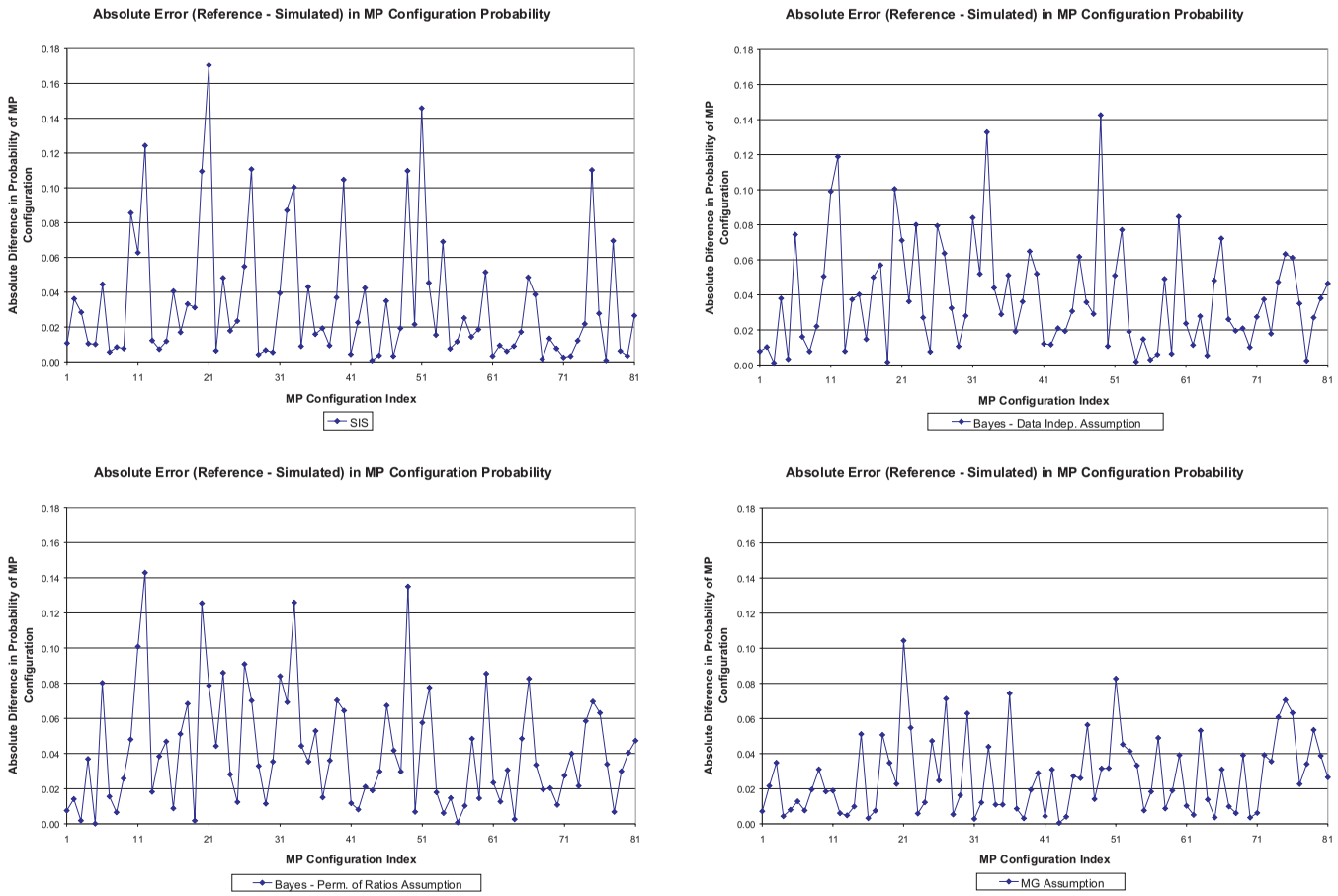

Figure 18: Absolute value of the mismatch in MP probability for all 81 MP configurations for four of the methods.

Example	Full Indep.	Perm. of Ratios	MG Assumption
Small isotropic objects p = 0.10	0.76	0.25	0.65
Small isotropic objects p = 0.50	0.66	0.45	0.55
Small isotropic objects p = 0.90	0.40	0.34	0.52
Large isotropic objects p = 0.10	0.72	0.24	0.72
Large isotropic objects p = 0.50	0.59	0.31	0.63
Large isotropic objects p = 0.90	0.26	0.17	0.63
Small anisotropic objects p = 0.10	0.75	0.25	0.66
Small anisotropic objects p = 0.50	0.60	0.37	0.59
Small anisotropic objects p = 0.90	0.33	0.26	0.58
Large anisotropic objects p = 0.10	0.68	0.22	0.74
Large anisotropic objects p = 0.50	0.59	0.29	0.65
Large anisotropic objects p = 0.90	0.25	0.14	0.64
Real example - One threshold p = 0.50	0.85	0.85	0.35
Real example - Ten thresholds	0.82	0.78	0.29

Table 2: Fraction of the nodes updated where $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ was outside the range of $P(\mathbf{A}|\mathbf{B})$ and $P(\mathbf{A}|\mathbf{C})$.

Discussion

The use of multiple-point statistics in geostatistical simulation is cumbersome. Many issues make their use difficult. Stationarity, inference, and the scale of the model are some of the issues we discussed in this paper. The integration of MP statistics requires simplifications to the correct expression obtained by Bayes' law. We discussed several approximations to facilitate the calculation of the conditional probability considering the multiple-point information in an indicator framework.

The assumption of permanence of ratios has several nice properties that suggest it should perform better than the full independence assumption for updating a conditional probability. The examples presented here showed a similar performance of these updating techniques.

Linearly combining the estimates of indicator kriging with the farthest data and the probability of the MP event constituted by the closest (adjacent) nodes to the location being simulated calls for an assumption of dependence. The multi-Gaussian framework is used because of its convenient mathematical properties. The estimation of MP covariances is done by combining two-point covariances. This greatly simplifies the integration of information and allows the calculation of factors (or weights) to be assigned to each estimate. The resulting probability is often outside the range of the input probabilities, which is a nice property, since it means that the updated probability is better informed than the two input probabilities. The examples presented showed a reasonably good performance of this method.

Further research is required to better assess the performance of these methods and to explore some of the other issues exposed in this paper.

References

[1] B. G. Arpat, J. Caers, and A. Haas. Characterization of West-Africa submarine channel reservoirs: A neural network based approach to integration of seismic data. In *2001 SPE Annual*

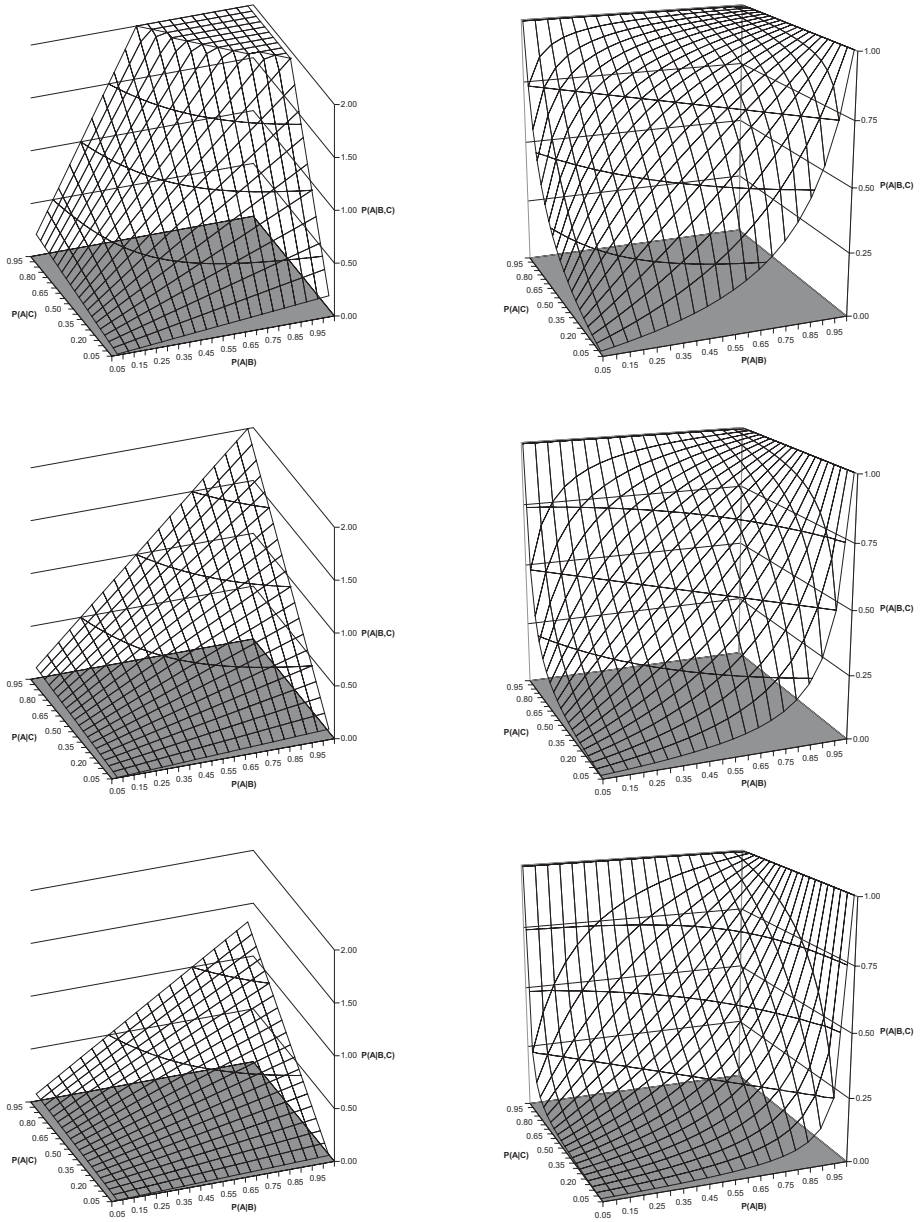

Figure 19: Graphs of $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ given $P(\mathbf{A}) = 0.25$ (top), $P(\mathbf{A}) = 0.50$ (middle), and $P(\mathbf{A}) = 0.75$ (bottom). The plots on the left show the maps under the assumption of full independence between $\mathbf{B}$ and $\mathbf{C}$. The plots on the right show the maps under the assumption of permanence of ratios.

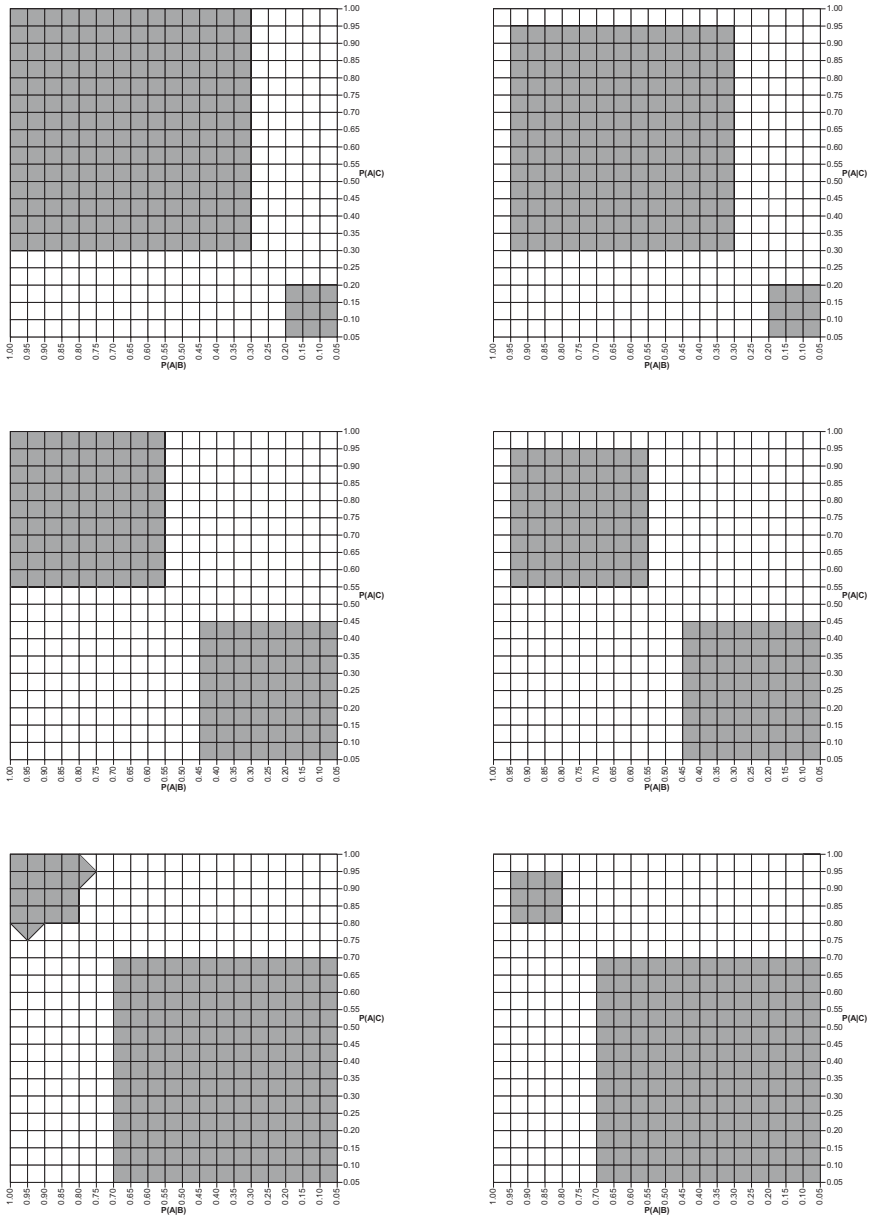

Figure 20: Graphs showing in grey the area where the estimated probability $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ is outside the range defined by $P(\mathbf{A}|\mathbf{B})$ and $P(\mathbf{A}|\mathbf{C})$ given $P(\mathbf{A}) = 0.25$ (top), $P(\mathbf{A}) = 0.50$ (middle), and $P(\mathbf{A}) = 0.75$ (bottom). The plots on the left show the maps under the assumption of full independence between $\mathbf{B}$ and $\mathbf{C}$. The plots on the right show the maps under the assumption of permanence of ratios.

Technical Conference and Exhibition, New Orleans, LA, September 2001. Society of Petroleum Engineers. SPE paper # 71345.

[2] J. Caers. Towards a pattern recognition-based geostatistics. In *Report 12, Stanford Center for Reservoir Forecasting*, Stanford, CA, May 1999.

[3] J. Caers and A. G. Journel. Stochastic reservoir simulation using neural networks trained on outcrop data. In *1998 SPE Annual Technical Conference and Exhibition*, pages 321–336, New Orleans, LA, September 1998. Society of Petroleum Engineers. SPE paper # 49026.

[4] C. V. Deutsch. Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection. *Computers & Geosciences*, 24(6):551–562, 1998.

[5] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 2nd edition, 1998.

[6] X. Emery. Conditional simulation of non-Gaussian random functions. *Mathematical Geology*, 34(1):79–100, 2002.

[7] J. Gómez-Hernández. Issues on environmental risk assessment. In E. Y. Baafi and N. A. Schofield, editors, *Geostatistics Wollongong '96*, volume 1, pages 15–26. Kluwer, 1997.

[8] F. Guardiano and M. Srivastava. Multivariate geostatistics: Beyond bivariate moments. In A. Soares, editor, *Geostatistics Tróia '92*, volume 1, pages 133–144. Kluwer, 1993.

[9] A. G. Journel. Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses. *Mathematical Geology*, 34(5):573–596, July 2002.

[10] J. Ortiz C. *Characterization of High Order Correlation for Enhanced Indicator Simulation*. PhD thesis, University of Alberta, Edmonton, AB, Canada, 2003.

[11] J. Ortiz C. Naive Bayes model and permanence of ratios. In *Centre For Computational Geostatistics*, volume 5, Edmonton, AB, 2003.

[12] J. Ortiz C. Short note: Bias in SIS due to order relation corrections and a quick fix. In *Centre For Computational Geostatistics*, volume 5, Edmonton, AB, 2003.

[13] J. Ortiz C. and C. V. Deutsch. Calculation of uncertainty in the variogram. *Mathematical Geology*, 34(2):169–183, 2002.

[14] J. Ortiz C. and C. V. Deutsch. Indicator simulation accounting for multiple-point statistics. In *Centre For Computational Geostatistics*, volume 5, Edmonton, AB, 2003.

[15] M. J. Pyrcz and C. V. Deutsch. A library of training images and associated code. In *Centre For Computational Geostatistics*, volume 5, Edmonton, AB, 2003.

[16] S. Strebelle and A. G. Journel. Sequential simulation drawing structures from training images. In *6th International Geostatistics Congress*, Cape Town, South Africa, April 2000. Geostatistical Association of Southern Africa.

[17] S. Strebelle and A. G. Journel. Reservoir modeling using multiple-point statistics. In *2001 SPE Annual Technical Conference and Exhibition*, New Orleans, LA, September 2001. Society of Petroleum Engineers. SPE paper # 71324.

Appendix

Derivation of Equation 2

Recall **Equation 1**:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{P(\mathbf{A},\mathbf{B},\mathbf{C})}{P(\mathbf{B},\mathbf{C})} \tag{9}$$

Considering $\mathbf{B}$ and $\mathbf{C}$ are conditionally independent given $\mathbf{A}$ and also considering them independent from each other, we have the following expressions:

$$P(\mathbf{B}|\mathbf{A},\mathbf{C}) = P(\mathbf{B}|\mathbf{A})$$
$$P(\mathbf{C}|\mathbf{A},\mathbf{B}) = P(\mathbf{C}|\mathbf{A})$$

$$P(\mathbf{B},\mathbf{C}) = P(\mathbf{B}) \cdot P(\mathbf{C})$$

Now,

$$P(\mathbf{A},\mathbf{B},\mathbf{C}) = P(\mathbf{C}|\mathbf{A},\mathbf{B}) \cdot P(\mathbf{A},\mathbf{B}) = P(\mathbf{C}|\mathbf{A},\mathbf{B}) \cdot P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{A})$$

which can be simplified to

$$P(\mathbf{A},\mathbf{B},\mathbf{C}) = P(\mathbf{C}|\mathbf{A}) \cdot P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{A})$$

We also know that:

$$P(\mathbf{C}|\mathbf{A}) = \frac{P(\mathbf{A},\mathbf{C})}{P(\mathbf{A})} = \frac{P(\mathbf{A}|\mathbf{C}) \cdot P(\mathbf{C})}{P(\mathbf{A})} \tag{10}$$

Similarly,

$$P(\mathbf{B}|\mathbf{A}) = \frac{P(\mathbf{A},\mathbf{B})}{P(\mathbf{A})} = \frac{P(\mathbf{A}|\mathbf{B}) \cdot P(\mathbf{B})}{P(\mathbf{A})} \tag{11}$$

Equation 9 can be rewritten:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{P(\mathbf{A},\mathbf{B},\mathbf{C})}{P(\mathbf{B},\mathbf{C})} = \frac{P(\mathbf{C}|\mathbf{A}) \cdot P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{A})}{P(\mathbf{B}) \cdot P(\mathbf{C})}$$

which entails **Equation 2**:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{P(\mathbf{A}|\mathbf{B})}{P(\mathbf{A})} \cdot \frac{P(\mathbf{A}|\mathbf{C})}{P(\mathbf{A})} \cdot P(\mathbf{A})$$

Derivation of Equation 3

Consider the expression for $P(\mathbf{A}|\mathbf{B},\mathbf{C})$ given the assumption of conditional independence between $\mathbf{B}$ and $\mathbf{C}$ given $\mathbf{A}$:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{P(\mathbf{A}) \cdot P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{C}|\mathbf{A})}{P(\mathbf{B},\mathbf{C})}$$

A similar expression for the complement of $\mathbf{A}$, noted $\bar{\mathbf{A}}$ can be written:

$$P(\bar{\mathbf{A}}|\mathbf{B},\mathbf{C}) = \frac{P(\bar{\mathbf{A}}) \cdot P(\mathbf{B}|\bar{\mathbf{A}}) \cdot P(\mathbf{C}|\bar{\mathbf{A}})}{P(\mathbf{B},\mathbf{C})}$$

The ratio between these two expressions entails:

$$\frac{P(\bar{\mathbf{A}}|\mathbf{B},\mathbf{C})}{P(\mathbf{A}|\mathbf{B},\mathbf{C})} = \frac{P(\bar{\mathbf{A}}) \cdot P(\mathbf{B}|\bar{\mathbf{A}}) \cdot P(\mathbf{C}|\bar{\mathbf{A}})}{P(\mathbf{A}) \cdot P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{C}|\mathbf{A})}$$

Using **Equations 10** and **11**, and the fact that $P(\bar{\mathbf{A}}|\mathbf{B},\mathbf{C}) = 1-P(\mathbf{A}|\mathbf{B},\mathbf{C})$, $P(\mathbf{B}|\bar{\mathbf{A}}) = 1-P(\mathbf{B}|\mathbf{A})$, and $P(\mathbf{C}|\bar{\mathbf{A}}) = 1-P(\mathbf{C}|\mathbf{A})$, we can deduce **Equation 3**:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) = \frac{\frac{1-P(\mathbf{A})}{P(\mathbf{A})}}{\frac{1-P(\mathbf{A})}{P(\mathbf{A})} + \frac{1-P(\mathbf{A}|\mathbf{B})}{P(\mathbf{A}|\mathbf{B})} \cdot \frac{1-P(\mathbf{A}|\mathbf{C})}{P(\mathbf{A}|\mathbf{C})}}$$

Derivation of Equation 4

We now show how to derive the weights in **Equation 4**:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) = \omega_B \cdot (P(\mathbf{A}|\mathbf{B}) - P(\mathbf{A})) + \omega_C \cdot (P(\mathbf{A}|\mathbf{C}) - P(\mathbf{A}))$$

$P(\mathbf{A}|\mathbf{B},\mathbf{C})$ can be calculated by linearly combining the estimates from different sources. To find the weights, we consider that the phenomenon is multi-Gaussian, hence, we can solve for the conditional probability with conventional kriging that uses only two point statistics.

The procedure is:

- Use all $n_B + n_C$ single points to solve the simple indicator kriging system of equations and find an estimate of $P(\mathbf{A}|\mathbf{B},\mathbf{C})$. This estimate is the linear combination of the weights that solve the kriging system and the indicator kriging values of the $n_B + n_C$ values (samples and previously simulated nodes):

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) = \begin{pmatrix} \Sigma_{A,B} & \Sigma_{A,C} \end{pmatrix} \cdot \begin{pmatrix} \Sigma_{B,B} & \Sigma_{B,C} \\ \Sigma_{C,B} & \Sigma_{C,C} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{i^B} - P(\mathbf{A}) \\ \mathbf{i^C} - P(\mathbf{A}) \end{pmatrix}$$

  Considering the expression for the inverse of the covariance matrix:

$$\begin{pmatrix} \Sigma_{B,B} & \Sigma_{B,C} \\ \Sigma_{C,B} & \Sigma_{C,C} \end{pmatrix}^{-1} = \begin{pmatrix} D_1 & D_0 \\ D_0^T & D_2 \end{pmatrix}$$

  We can solve for the sub-matrices by multiplying the covariance matrix with its inverse and setting this product to be equal to the identity:

$$\begin{pmatrix} D_1 & D_0 \\ D_0^T & D_2 \end{pmatrix} \cdot \begin{pmatrix} \Sigma_{B,B} & \Sigma_{B,C} \\ \Sigma_{C,B} & \Sigma_{C,C} \end{pmatrix} = \begin{pmatrix} I_{n_B} & 0 \\ 0 & I_{n_C} \end{pmatrix}$$

  We have four equations to find three unknowns:

$$\begin{aligned} &D_1 \cdot \Sigma_{B,B} + D_0 \cdot \Sigma_{C,B} = I_{n_B} \\ &D_1 \cdot \Sigma_{B,C} + D_0 \cdot \Sigma_{C,C} = 0 \\ &D_0^T \cdot \Sigma_{B,B} + D_2 \cdot \Sigma_{C,B} = 0 \\ &D_0^T \cdot \Sigma_{B,C} + D_2 \cdot \Sigma_{C,C} = I_{n_C} \end{aligned}$$

  The first two equations allow the calculation of D_1:

$$\begin{aligned} &D_1 \cdot \Sigma_{B,B} + D_0 \cdot \Sigma_{C,B} = I_{n_B} \\ &D_0 = -D_1 \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \end{aligned}$$

  hence,

$$\begin{aligned} &D_1 \cdot \Sigma_{B,B} + -D_1 \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \cdot \Sigma_{C,B} = I_{n_B} \\ \Rightarrow \quad &D_1 = (\Sigma_{B,B} - \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \cdot \Sigma_{C,B})^{-1} \end{aligned}$$

And now, we can calculate D_0:

$$\begin{array}{rl} & D_0 = -D_1 \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \\ \Rightarrow & D_0 = -(\Sigma_{B,B} - \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \cdot \Sigma_{C,B})^{-1} \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \end{array}$$

Finally, we can obtain the expression for D_2:

$$\begin{array}{rl} & D_0^T \cdot \Sigma_{B,C} + D_2 \cdot \Sigma_{C,C} = I_{n_C} \\ \Rightarrow & D_2 = \Sigma_{C,C}^{-1} - D_0^T \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \\ \Rightarrow & D_2 = \Sigma_{C,C}^{-1} - (-(\Sigma_{B,B} - \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \cdot \Sigma_{C,B})^{-1} \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1})^T \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \\ \Rightarrow & D_2 = \Sigma_{C,C}^{-1} + \left((\Sigma_{B,B} - \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \cdot \Sigma_{C,B})^{-1} \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1}\right)^T \cdot \Sigma_{B,C} \cdot \Sigma_{C,C}^{-1} \end{array}$$

We can now re-write the expression for the conditional probability:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) = \begin{pmatrix} \Sigma_{A,B} & \Sigma_{A,C} \end{pmatrix} \cdot \begin{pmatrix} D_1 & D_0 \\ D_0^T & D_2 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{i^B} - P(\mathbf{A}) \\ \mathbf{i^C} - P(\mathbf{A}) \end{pmatrix}$$

$$\begin{array}{rcl} P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) & = & \left(\Sigma_{A,B} \cdot D_1 + \Sigma_{A,C} \cdot D_0^T\right) \cdot (\mathbf{i^B} - P(\mathbf{A})) + \\ & & (\Sigma_{A,B} \cdot D_0 + \Sigma_{A,C} \cdot D_2) \cdot (\mathbf{i^C} - P(\mathbf{A})) \end{array}$$

- Use only the n_B points of the farthest single points to solve the simple indicator kriging system and find an estimate of $P(\mathbf{A}|\mathbf{B})$.

$$P(\mathbf{A}|\mathbf{B}) - P(\mathbf{A}) = \Sigma_{A,B} \cdot \Sigma_{B,B}^{-1} \cdot (\mathbf{i^B} - P(\mathbf{A}))$$

- Repeat with the n_C single points (that constitute the multiple-point event) and solve a simple indicator kriging system to estimate the conditional probability $P(\mathbf{A}|\mathbf{C})$. Notice that this estimate does not have to be equal to the value obtained from a secondary source (either a training image or production quasi-exhaustive data).

$$P(\mathbf{A}|\mathbf{C}) - P(\mathbf{A}) = \Sigma_{A,C} \cdot \Sigma_{C,C}^{-1} \cdot (\mathbf{i^C} - P(\mathbf{A}))$$

- Since we are interested in obtaining the weights to linearly combine the estimates as if they were multi-Gaussian, we have the expression:

$$P(\mathbf{A}|\mathbf{B},\mathbf{C}) - P(\mathbf{A}) = \omega_B \cdot (P(\mathbf{A}|\mathbf{B}) - P(\mathbf{A})) + \omega_C \cdot (P(\mathbf{A}|\mathbf{C}) - P(\mathbf{A}))$$

$$\begin{array}{c} \left(\Sigma_{A,B} \cdot D_1 + \Sigma_{A,C} \cdot D_0^T\right) \cdot (\mathbf{i^B} - P(\mathbf{A})) + (\Sigma_{A,B} \cdot D_0 + \Sigma_{A,C} \cdot D_2) \cdot (\mathbf{i^C} - P(\mathbf{A})) = \\ \omega_B \cdot \Sigma_{A,B} \cdot \Sigma_{B,B}^{-1} \cdot (\mathbf{i^B} - P(\mathbf{A})) + \omega_C \cdot \Sigma_{A,C} \cdot \Sigma_{C,C}^{-1} \cdot (\mathbf{i^C} - P(\mathbf{A})) \end{array}$$

Identifying the terms common to $(\mathbf{i^B} - P(\mathbf{A}))$ and $(\mathbf{i^C} - P(\mathbf{A}))$, we can set the following identities:

$$\begin{array}{c} \left(\Sigma_{A,B} \cdot D_1 + \Sigma_{A,C} \cdot D_0^T\right) = \omega_B \cdot \Sigma_{A,B} \cdot \Sigma_{B,B}^{-1} \\ (\Sigma_{A,B} \cdot D_0 + \Sigma_{A,C} \cdot D_2) = \omega_C \cdot \Sigma_{A,C} \cdot \Sigma_{C,C}^{-1} \end{array}$$

And a possible solution (see **Equation 8**) is to take the ratio between the norms of the vectors on each side of the equations:

$$\omega_B = \frac{\|\Sigma_{A,B} \cdot D_1 + \Sigma_{A,C} \cdot D_0^T\|}{\|\Sigma_{A,B} \cdot \Sigma_{B,B}^{-1}\|} \qquad \omega_C = \frac{\|\Sigma_{A,B} \cdot D_0 + \Sigma_{A,C} \cdot D_2\|}{\|\Sigma_{A,C} \cdot \Sigma_{C,C}^{-1}\|}$$