High-Order Data-Driven Spatial Simulation of Categorical Variables

Ilnur Minniakhmetov[1] · Roussos Dimitrakopoulos[1]

© The Author(s) 2021

Abstract   Modern approaches for the spatial simulation of categorical variables are largely based on multi-point statistical methods, where a training image is used to derive complex spatial relationships using relevant patterns. In these approaches, simulated realizations are driven by the training image utilized, while the spatial statistics of the actual sample data are ignored. This paper presents a data-driven, high-order simulation approach based on the approximation of high-order spatial indicator moments. The high-order spatial statistics are expressed as functions of spatial distances that are similar to variogram models for two-point methods, while higher-order statistics are connected with lower-orders via boundary conditions. Using an advanced recursive B-spline approximation algorithm, the high-order statistics are reconstructed from the available data and are subsequently used for the construction of conditional distributions using Bayes' rule. Random values are subsequently simulated for all unsampled grid nodes. The main advantages of the proposed technique are its ability to (a) simulate without a training image to reproduce the high-order statistics of the data, and (b) adapt the model's complexity to the information available in the data. The practical intricacies and effectiveness of the proposed approach are demonstrated through applications at two copper deposits.

Keywords   Stochastic simulation · Data-driven · Categorical variables · High-order spatial statistics · Spatial model

✉   Ilnur Minniakhmetov
       delnuro@gmail.com; ilnur.minniakhmetov@bhp.com

       Roussos Dimitrakopoulos
       roussos.dimitrakopoulos@mcgill.ca

[1]   COSMO—Stochastic Mine Planning Laboratory, McGill University, Montreal, QC H3A 0E8, Canada

⌀ Springer

1 Introduction

Geostatistical simulations are often required in reservoir modeling, as well as in the quantification of geological uncertainty, pollutants in contaminated areas, and other spatially dependent geological and environmental phenomena. During the past few decades, geostatistical simulations of categorical variables, such as geological units with complex spatial geometries of mineral deposits and petroleum reservoirs, have largely been modeled within the framework of multiple-point spatial simulation (MPS) methods that were introduced in the 1990s and have been further developed since then (Guardiano and Srivastava 1993; Journel 1993; Strebelle 2002, 2021; Journel 2003; Zhang et al. 2006; Chugunova and Hu 2008; Remy et al. 2009; Mariethoz and Renard 2010; Straubhaar 2011; Stien and Kolbjørnsen 2011; Toftaker and Tjelmeland 2013; Strebelle and Cavelius 2014; Zhang et al. 2017; Gómez-Hernández and Srivastava 2021, others). The MPS framework is based on the use of training images (TI) or analogues of the attributes of interest being modeled and contains additional information about the complex spatial relations of the attributes to be simulated; however, the TIs are not conditioned to the available data and their spatial statistics. To retrieve and use the pertinent information from a TI, the similarity between the local neighborhood of an unsampled location to be simulated and the TI is calculated in an explicit or implicit form. Based on this similarity measure, the value of a node from the TI with the most similar neighborhood is assigned to the unsampled location being simulated. Generally, most of the multi-point simulation techniques are a Monte Carlo sampling of values from the TI in some form. No spatial models are used and, importantly, no spatial information is retrieved from the available sample data. As a result, simulated realizations of attributes of interest reflect the TI and its spatial aspects. In cases where there are relatively large data sets, conflict between the available data and the TI statistics is observed, and the resulting simulated realizations do not reproduce the spatial statistics of the data (Goodfellow et al. 2012; Osterholt and Dimitrakopoulos 2007; Dimitrakopoulos et al. 2010; Pyrcz and Deutsch 2014). Several attempts have been made to incorporate more information from the available data. Some authors suggest using replicates from the data in addition to TI (Mariethoz and Renard 2010); however, in practice, it is difficult to find any replicates for three-point relations when data are sparse. Others (Mariethoz and Kelly 2011) apply affine transformations to better condition to the data; however, a TI remains the main source of information.

Dimitrakopoulos et al. (2010) and Mustapha and Dimitrakopoulos (2010a, b) propose the use of high-order spatial cumulants to capture complex multi-point relations during the simulation of non-Gaussian random fields. The proposed high-order simulation approach estimates the third- and fourth-order spatial statistics from data and complements them with higher-order statistics from the TI. Further developments in algorithmic performance (Yao et al. 2018, 2020), generalization using splines (Minniakhmetov et al. 2018), a high-order decorrelation method (Minniakhmetov and Dimitrakopoulos 2017a), and efficient block simulations (de Carvalho et al. 2019), and training-image-free simulations (Yao et al.

2021) have made the approach more practical. These approaches are based on the approximation of a conditional distribution using Legendre polynomials, which are smooth functions and are incapable of an adequate approximation of the discrete distribution of categorical variables.

The topic of describing complex multi-point relations of categorical variables is addressed in Vargas-Guzman (2011) and Vargas-Guzman and Qassab (2006), who use high-order indicator statistics to characterize spatially distributed rock units, while Minniakhmetov and Dimitrakopoulos (2017b) develop this further by introducing the connection between different orders to the related mathematical model for the two-dimensional case. For example, consider a third-order spatial indicator moment of a stationary random field, which is a function of two lags. When one of the lags is equal to zero, the third-order indicator moment becomes the second-order indicator moment. In addition, instead of exponential functions, the B-spline functions are used to estimate high-order spatial indicator moments. It is known that B-splines provide an optimal (in terms of accuracy) estimation of equi-continuous functions defined on compacts (Evans et al. 2009; Babenko 1986). Based on the above, a new recursive algorithm is proposed for the better approximation of high-order spatial statistics with nested boundary conditions of lower-level relations. Then, the conditional distribution for the given neighborhood is calculated from high-order indicator moments and the related category is simulated. In addition to extending previous developments mentioned above, the present study also explores practical aspects of the proposed method through applications at two major, real-world copper deposits: Olympic Dam, Australia, and Escondida, Chile (the world's largest copper mine). Furthermore, the paper highlights the importance of high-order spatial statistics as the useful tool for the analysis of spatial contact relations between categories. Contrary to indicator variograms (Journel and Alabert 1990; Goovaerts 1997) that provide information about pair-wise relationships between categories, the third- and fourth-order spatial indicator moments reflect three- and four-wise relations between multiple categories in space. It should be noted that, as shown in the subsequent sections, the proposed method works without a TI; however, additional information from a TI can be incorporated as a secondary condition, ensuring that the high-order spatial indicator moments are driven by the available data.

The paper is organized as follows. First, high-order spatial indicator moments are introduced as a function of distances between points for two-point and multi-point cases. Then, a mathematical model for recursive approximation of high-order spatial indicator moments is presented, followed by the proposed high-order, data-driven, categorical simulation method. Subsequently, the proposed simulation algorithm is applied in two case studies to simulate the geological units of copper deposits. Discussion and conclusions follow.

2 High-Order Spatial Indicator Simulation

Let (Ω, F, P) be a probability space. Consider a stationary ergodic random vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_N)^{\mathrm{T}}$, $\mathbf{Z} : \Omega \to S^N$, defined on a regular grid $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, $\mathbf{x} \in R^n, n = 2, 3$, where Ω is a space of all possible outcomes, F contains all combinations of Ω, S^N is a set of states represented by categories $S = \{s_1, s_2, \ldots, s_K\}$,

and P is the probability measure, or probability. For example, the probability of Z_1 being at a state s_k is defined as

$$P(Z_1 = s_k) \equiv P(\{\omega \in \Omega : \mathbf{Z}(\omega) \in s_k \otimes S^{N-1}\}) \tag{1}$$

Without loss of generality, assume that $s_k = k, k = 1 \ldots K$. Let $\mathbf{d_n} = \{z_\alpha, \alpha = 1 \ldots n\}$ be a given set of conditioning data, where lowercase z stands for outcomes of random variable Z. The focus of high-order simulation techniques is to simulate the realization of the random vector $\mathbf{Z}$ for all nodes of a grid D with a given set of conditioning data $\mathbf{d_n}$.

Similarly to Minniakhmetov and Dimitrakopoulos (2017b), the high-order categorical simulation method is based on the concept of sequential simulation (Journel and Alabert 1990; Journel 1993), where the joint probability distribution $P(Z_1 = k_1, Z_2 = k_2, \ldots, Z_N = k_N | \mathbf{d_n})$ of the random vector $\mathbf{Z}$ can be decomposed into the product of conditional univariate distributions

$$\begin{aligned} P(Z_1 = k_1, \ldots, Z_N = k_N | \mathbf{d_n}) &= P(Z_2 = k_2, \ldots, Z_N = k_N | Z_1 = k_1, \mathbf{d_n}) P(Z_1 = k_1 | \mathbf{d_n}) \\ &= \prod_{i=2}^{N} P(Z_i = k_i | Z_1 = k_1, \ldots, Z_{i-1} = k_{i-1}, \mathbf{d_n}) P(Z_1 = k_1 | \mathbf{d_n}) \end{aligned} \tag{2}$$

According to Eq. (2), simulation of categorical variables can be done sequentially by visiting a grid node at a time and calculating $P(Z_i = k_i | Z_1 = k_1, \ldots, Z_{i-1} = k_{i-1}, \mathbf{d_n})$. However, in practice (Dimitrakopoulos and Luo 2004), instead of considering all previously simulated nodes and data $\{Z_1, \ldots, Z_{i-1}, \mathbf{d_n}\}$, only those in local neighborhood of Z_i are considered, i.e.

$$P(Z_i = k_i | Z_1 = k_1, \ldots, Z_{i-1} = k_{i-1}, \mathbf{d_n}) \approx P(Z_i = k_i | \Lambda_i), \tag{3}$$

where Λ_i is the set of previously simulated nodes and data within the local neighborhood of Z_i.

Similarly to Mustapha and Dimitrakopoulos (2010a, b), the conditional distribution in Eq. (3) can be calculated from the joint distribution. Without loss of generality, consider conditional distribution $P(Z_0 = k_0 | Z_1 = k_1, \ldots, Z_n = k_n)$; it can then be calculated using Bayes' rule (Ripley 1987)

$$P(Z_0 = k_0 | Z_1 = k_1, \ldots, Z_n = k_n) = P(Z_0 = k_0, Z_1 = k_1, \ldots, Z_n = k_n) / P(Z_1 = k_1, \ldots, Z_n = k_n), \tag{4}$$

where $P(Z_1 = k_1, \ldots, Z_n = k_n)$ can be considered as a normalization coefficient due to the relations

$$P(Z_1 = k_1, \ldots, Z_n = k_n) = \sum_{k_0=1}^{K} P(Z_0 = k_0, Z_1 = k_1, \ldots, Z_n = k_n), \tag{5}$$

It can be shown that the probability is equivalent to spatial indicator moment (Vargas-Guzman 2011)

$$P(Z_{i_0} = k_0, Z_{i_1} = k_1, \ldots) = E\left[I_{k_0}(Z_{i_0})I_{k_1}(Z_{i_1})\ldots\right], \quad \forall i_0, i_1 \ldots = 1 \ldots N, \quad \forall k_0, k_1, \ldots = 1 \ldots K \tag{6}$$

where E is the expected value operator and $I_k(Z_i)$ is an indicator function

$$I_k(Z_i) = \begin{cases} 1, & Z_i = k \\ 0, & Z_i \neq k_0 \end{cases}. \tag{7}$$

From here on, indicator moments are denoted as

$$M_{\mathbf{k}}(Z_{i_0}, Z_{i_1}, \ldots) = M_{k_0 k_1, \ldots}(Z_{i_0}, Z_{i_1}, \ldots) = E\left[I_{k_0}(Z_{i_0})I_{k_1}(Z_{i_1})\ldots\right]. \tag{8}$$

Finally, Eqs. (2–5) define simulation algorithm of categorical random vector **Z**.

Algorithm A.1

1. Define a random path visiting all the unsampled nodes.
2. For each node $\mathbf{x}_{i_0}$ in the path:

   a. Find the closest data samples $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots \mathbf{x}_{i_n}$. The categories at these nodes are denoted by $k_1, \ldots k_n$.
   b. For all $k_0 = 1 \ldots K$, calculate the high-order spatial indicator moments $M_{\mathbf{k}}(Z_{i_0}, Z_{i_1}, \ldots, Z_{i_n})$
   c. Calculate the conditional distribution from joint distribution

   $$P\left(Z_{i_0} = k_0 | Z_{i_1} = k_1 \ldots, Z_{i_n} = k_n\right) = AM_{\mathbf{k}}(Z_{i_0}, Z_{i_1}, \ldots, Z_{i_n}), \tag{9}$$

   where coefficient A is the normalization coefficient as in Eq. (4)

   $$A = 1 / \sum_{k_0=1}^{K} M_{\mathbf{k}}(Z_{i_0}, Z_{i_1}, \ldots, Z_{i_n}) \tag{10}$$

   d. Draw a random value z_{i_0} from this conditional distribution (10) and assign it to the unsampled location $\mathbf{x}_{i_0}$.
   e. Add z_{i_0} to the set of sample hard data and the previously simulated values.
   f. Repeat Steps 2a–e for all the points along the random path defined in Step 1.

The next section presents a new method to calculate high-order spatial indicator moments for Eqs. (9–10).

3 High-Order Spatial Indicator Moments

Similarly to second-order statistics such as variograms, high-order spatial moments are calculated by discretizing distances between data samples into lags and finding all pairs, triplets, or multiplets separated by the same lags. To define lags in the two-dimensional case, the local neighborhood of each data sample is divided into

eight sectors (oct $= 1 \ldots 8$) and concentric circles with radius ($r = \{r_1, r_2, \ldots r_{\max}\}$) increasing by logarithmic law (Fig. 1). The choice of logarithmically increasing lags is driven mainly by computational resource limits. For example, to cover extents of 400 m (with 200 m typical continuity range for the deposits under consideration), the constant lag division with resolution at first lags of about 20 m requires 20 lags, which correspond to $20^4 = 160{,}000$ bins in a fourth-order map for each possible combination of categories in four points, whereas logarithmically increasing lags $\{20, 30, 40, 70, 160, 400\}$ require $6^4 = 1296$ bins, i.e. 123 times fewer calculations. In Fig. 1, data samples are denoted by x_0 (central point), x_1, x_2, and x_3. Data sample x_1 is located in octant $o = 7$ and lag $h = r_3$, x_2 is in octant $o = 1$ and lag $h = r_2$, and x_3 is in octant $o = 5$ and lag $h = r_2$. Any point located in the central circle belongs to all octant and lag 0. Thus, high-order moments can be expressed as functions of octant index and lags, e.g. $M_{k_0 k_1 k_2}(Z_0, Z_1, Z_2) \approx M_{k_0 k_1 k_2}(o_1 = 7, o_2 = 1; h_1 = r_3, h_2 = r_2)$, where Z_0, Z_1, Z_2 are random variables at locations x_0, x_1, x_2. From here on, calculated indicator moments are denoted as

$$M_{\mathbf{k}}(\mathbf{o}; \mathbf{h}) = M_{k_0 k_1, \ldots, k_n}(o_1, \ldots o_n; h_1, \ldots h_n) = M_{\mathbf{k}}(Z_{i_0}, Z_{i_1}, \ldots Z_{i_n}), \qquad (11)$$

and are calculated using sampling average

$$\hat{M}_{\mathbf{k}}(\mathbf{o}; \mathbf{h}) = \frac{1}{N_{\mathbf{o}, \mathbf{h}}} \sum_{j=1}^{N_{\mathbf{o}, \mathbf{h}}} I_{k_0}(z_{i_0}^j) \ldots I_{k_n}(z_{i_n}^j), \qquad (12)$$

where $\wedge$ denotes statistics calculated from data samples, i.e. sampling statistics, and $N_{\mathbf{o}, \mathbf{h}}$ is the number of all data samples $z_{i_0}^j \ldots z_{i_n}^j, j = 1 \ldots N_{\mathbf{o}, \mathbf{h}}$ falling in the octants

$\mathbf{o} = \{o_1, \ldots, o_n\}$ and lags $\mathbf{h} = \{h_1, \ldots, h_n\}$.

The data samples in Fig. 1 contribute to experimental high-order spatial statistics from order 2 to 4: second-order indicator moments $\hat{M}_{k_0 k_1}(o = 7; h = r_3)$, $\hat{M}_{k_0 k_2}(o = 1; h = r_2)$, $\hat{M}_{k_0 k_3}(o = 5; h = r_2)$, third-order indicator moments $\hat{M}_{k_0 k_1 k_2}(o_1 = 7, o_2 = 1; h_1 = r_3, h_2 = r_2)$, $\hat{M}_{k_0 k_1 k_3}(o_1 = 7, o_2 = 5; h_1 = r_3, h_2 = r_2)$, $\hat{M}_{k_0 k_2 k_3}(o_1 = 7, o_2 = 5; h_1 = r_3, h_2 = r_2)$, and fourth-order indicator moment $\hat{M}_{k_0 k_1 k_2 k_3}(o_1 = 7, o_2 = 1, o_3 = 5; h_1 = r_3, h_2 = r_2, h_3 = r_2)$. It should be noted that during simulation Step 2a, only one data sample per octant is used as conditional data to avoid calculation of high-order moments with repetitive random variables,

Fig. 1 Octant division of local neighborhood with logarithmically increasing lags. x_0 (central point), x_1, x_2, and x_3 are data samples
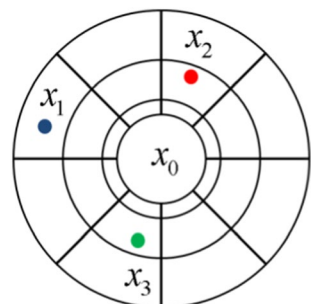
e.g. $\hat{M}_{k_0 k_1 k_1}(Z_0, Z_1, Z_1)$, $\hat{M}_{k_0 k_1 k_3 k_3}(Z_0, Z_1, Z_3, Z_3)$. This is quite similar to the octant search approach for second-order statistics methods (Remy et al. 2009). If several data samples fall in the same octant, the choice is made randomly.

In the three-dimensional case, instead of octants, the quadraginta octant division (Biswas and Bhowmick 2017) with 48 sectors (6 sectors per quadrant) is used (Fig. 2).

Following the logic of fitting theoretical variograms to variogram models (Journel and Huijbregts 1978), the sampling statistics (12) are not used directly to calculate joint distributions, but they are used to model high-order indicator moments.

The quadraginta octant division is the critical part of step 2b, which entails the calculation of high-order spatial indicator moment $M_{\mathbf{k}}(Z_{i_0}, Z_{i_1}, \ldots, Z_{i_n})$ of Algorithm A1 above. For the sake of simplicity, consider three possible categories $k \in \{0, 1, 2\}$ and three points: central point random value Z_0 to be simulated at location $\mathbf{x}_0$ and two neighborhood data samples z_1 and z_2 in arbitrary directions $\mathbf{x}_1$ and $\mathbf{x}_2$. First, the octants (o_1, o_2) and lag distances (h_1, h_2) are calculated from lags $\mathbf{h}_1 = \mathbf{x}_1 - \mathbf{x}_0$ and $\mathbf{h}_2 = \mathbf{x}_2 - \mathbf{x}_0$ centered at point$\mathbf{x}_0$. Next, the two-dimensional surfaces $\hat{M}_{k_0=0, k_1, k_2}(o_1, o_2; u, v)$, $\hat{M}_{k_0=1, k_1, k_2}(o_1, o_2; u, v)$, $\hat{M}_{k_0=2, k_1, k_2}(o_1, o_2; u, v)$ are approximated using all available data and the nested algorithm presented in Sect. 3.1. Note that (k_1, k_2) and (o_1, o_2) are fixed and known from neighborhood data at $\mathbf{x}_1$ and $\mathbf{x}_2$; (u, v) are distances along directions (o_1, o_2)—the only variable part of $\hat{M}_{k_0=0, k_1, k_2}(o_1, o_2; u, v)$.

It should be noted that using quadraginta octant search in Algorithm A1 is quite different from the search in the classical MPS methods, such as SNESIM (Strebelle and Cavelius 2014). Firstly, MPS methods reduce the number of neighborhood data when no exact replicates are found for a particular spatial configuration, whereas octant search provides all possible replicates at different lags with the fixed angles; for example, for the third-order moments (3-point statistics) for the fixed octants (o_1, o_2) defined by neighborhood data, all replicates at different lag distances (h_1, h_2) are found and used in the simulation of a value in a node. Secondly, octant search

Fig. 2 Quadraginta octant division for the tree-dimensional case (Biswas and Bhowmick 2017)
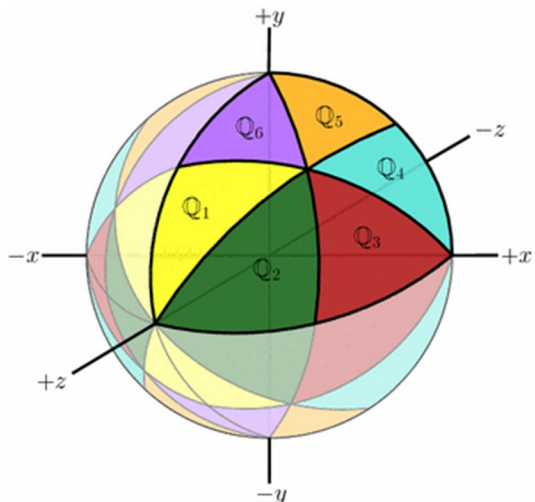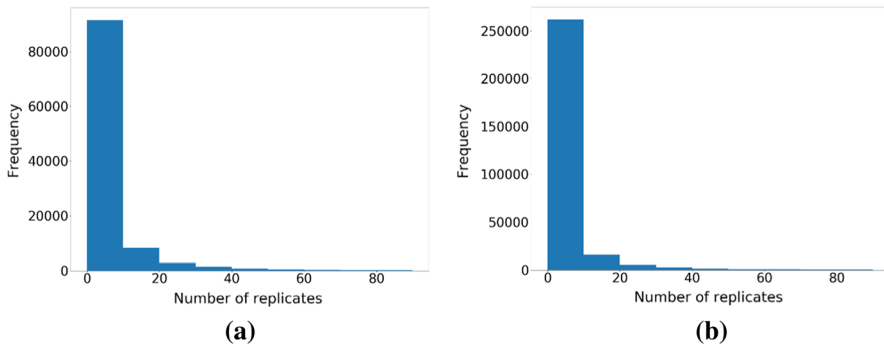
Fig. 3 Histogram of number of replicates for **a** fourth- and **b** fifth-order spatial indicator moments found in the Olympic Dam data set

Table 1 Percentiles of number of replicates for (a) fourth-order and (b) fifth-order spatial indicator moments found in the Olympic Dam data set

	P50	P60	P70	P80	P90
Number of replicates in fourth-order moments	2	3	4	8	18
Number of replicates in fifth-order moments	2	2	3	6	12

does not look for exact spatial replicates, but replicates within tolerances of lags and angles, which dramatically increases the number of replicates used in the simulation process. Lastly, there is no restriction on the "regularity" of the data sample locations or TI (if used), as the octant search works on points rather than on a grid.

3.1 Approximating High-Order Indicator Moments

Data available in some applications can be dense in space and give the impression that the high-order spatial statistics in Eq. (10) can be directly calculated from samples. However, the dimension of space in high-order spatial statistics grows exponentially as the order increases, such that even dense drilling is insufficient for the direct calculation of high-order statistics from the data. For example, in the fourth-order indicator moments, there are 17,296 possible combinations of directions (all possible three neighbor directions from 48 quadraginta octant division) and 2^4 possible combinations of indicators, for the case with two categories only, which results in $2^4 \times 17{,}296 = 138{,}368$ fourth-order spatial indicator moments. Figure 3 and Table 1 show the histogram and percentiles, respectively, of the number of replicates for fourth- and fifth-order spatial indicator moments found in the Olympic Dam data set described in Sect. 4.1. Replicates have been calculated using quadraginta octant division and 20 m lag tolerance. Overall, 80% of all possible spatial configurations from the drill-hole data had fewer than eight samples to calculate the fourth-order

indicator moment and fewer than six samples for the fifth-order indicator moments. Such a small number of replicates is not sufficient to provide a robust calculation of high-order moments for Eq. (10) of Algorithm A.1.

To increase the amount of information and, consequently, the quality of approximation of high-order moments, Minniakhmetov and Dimitrakopoulos (2017a, b) show that high-order indicator moments are bound by low-order moments

$$M_{\mathbf{k}}(\mathbf{o}; h_1, \ldots, h_p = 0, \ldots, h_n) = M_{\mathbf{k}\backslash k_p}(\mathbf{o}; \mathbf{h}\backslash h_p)\delta_{k_0, k_p}, \forall p \in 1 \ldots n, \quad (13)$$

where $\mathbf{h}\backslash h_p$ denotes all the lags $\mathbf{h}$ excluding the lag h_p, and $\mathbf{k}\backslash k_p$ denotes all the categories $\mathbf{k}$ excluding the category k_p. If the directions are close to orthogonal, then additional boundary conditions are valid

$$M_{\mathbf{k}}(\mathbf{o}; h_1, \ldots, h_p \to \infty, \ldots, h_n) = M_{\mathbf{k}\backslash k_p}(\mathbf{o}; \mathbf{h}\backslash h_p)M_{k_p}, \forall p \in 1 \ldots n. \quad (14)$$

where M_k is first-order statistics, i.e. proportion of category k in the data. Using the boundary conditions Eqs. (13–14) and sampling statistics (12), high-order indicator moments are approximated by

$$M_{\mathbf{k}}(\mathbf{o}; \mathbf{h}) = M_{\mathbf{k}}^0(\mathbf{o}; \mathbf{h}) + \delta M_{\mathbf{k}}(\mathbf{o}; \mathbf{h}), \quad (15)$$

where $M_{\mathbf{k}}^0(\mathbf{o}, \mathbf{h})$ is a trend defined just by boundary conditions, and $\delta M_{\mathbf{k}}(\mathbf{o}; \mathbf{h})$ is the B-spline regression of the mismatch between sampling statistics and trend $M_{\mathbf{k}}^0(\mathbf{o}; \mathbf{h})$.

The trend $M_{\mathbf{k}}^0(\mathbf{o}; \mathbf{h})$ connects lower-order moments with high-order moment by recursive formula

$$M_{\mathbf{k}}^0(\mathbf{o}; \mathbf{h}) = \frac{1}{\sum_{p=1}^n e^{-ah_p} + e^{-a(1-h_p)}} \left[\sum_{p=1}^n M_{k_0 \ldots k_n}(\mathbf{o}; h_1, \ldots, h_p = 0 \ldots, h_n)e^{-ah_p} \right.$$
$$\left. + \sum_{p=1}^n M_{k_0 \ldots k_n}(\mathbf{o}; h_1, \ldots, h_p \to \infty \ldots, h_n)e^{-a(1-h_p)} \right], \ \forall p = 1 \ldots n, \quad (16)$$

where $a = c/r_{\max}$, $r_{\max}$ is the radius that defines the local neighborhood, and c is the user-defined parameter that controls the influence of boundary conditions, i.e. small values of c force the approximation to use lower-order statistics close to boundaries ($h_p = 0, r_{\max}$) and sampling statistics in the area far from boundaries, whereas large values of c increase the influence of low-order statistics on high-order statistics. For data-rich environments, such as mining of mineral deposits, the use of smaller values of c is suggested, and for sparse data, large values.

As the term $M_{\mathbf{k}}^0(\mathbf{o}; \mathbf{h})$ incorporates the connection between low and high orders, the $\delta M_{\mathbf{k}}(\mathbf{o}; \mathbf{h})$ is only responsible for approximation of sampling statistics with zero boundary conditions.

The approximation of the multidimensional function $\delta M_{\mathbf{k}}(\mathbf{o}; \mathbf{h})$ is the classical linear regression problem with constraints. In the present work, the multidimensional cardinal B-spline regression is used (Friedman et al. 2001), where $\delta M_{\mathbf{k}}(\mathbf{o}; \mathbf{h})$ is approximated using a linear combination of B-splines defined on uniform intervals.

$$\delta M_{\mathbf{k}}(\mathbf{o}; \mathbf{h}) = \sum_{i_1=1}^{\omega} \cdots \sum_{i_n=1}^{\omega} \alpha_{i_1,\dots,i_n} B_{i_1,r}(h_1) \dots B_{i_n,r}(h_n), \tag{17}$$

where $\alpha_{i_1,\dots,i_n}$ are coefficients of B-spline approximation, and $B_{i,r}(t)$ is the ith B-splines of order r on uniformly divided knot sequence $\{0, dr, 2dr \dots r_{max}\}$, which are separated by step dr that increases with the order of moment $\delta M_{\mathbf{k}}(\mathbf{o}; \mathbf{h})$, thus providing more regularized approximation for higher orders and detailed approximation for lower orders. In practice, only orders up to 5 can be adequately calculated from the data; therefore, $dr = r_{max}/6$, $dr = r_{max}/4$, and $dr = r_{max}/2$ are used for moments of order 3, 4, 5, respectively. For second-order moments, standard variogram modeling is utilized (David 1988).

The coefficients $\alpha_{i_1,\dots,i_n}$ are found using a least-squares algorithm to fit points, which are the residual of high-order moments calculated from the data samples and trend $M_{\mathbf{k}}^0(\mathbf{o}; \mathbf{h})$ from Eq. (16)

$$\delta M_{\mathbf{k}}(\mathbf{o}; \mathbf{h}^d) = \hat{M}_{\mathbf{k}}(\mathbf{o}; \mathbf{h}^d) - M_{\mathbf{k}}^0(\mathbf{o}; \mathbf{h}^d), \tag{18}$$

under zero boundary constraints

$$\delta M_{\mathbf{k}}(\mathbf{o}; h_1, \dots, h_p = 0, \dots, h_n) = 0,$$
$$\delta M_{\mathbf{k}}(\mathbf{o}; h_1, \dots, h_p \to \infty, \dots, h_n) = 0, \forall p = 1 \dots n, \tag{19}$$

where $\mathbf{h}^d$ are centers of lags used to calculate high-order statistics from the data (12).

Using all the above, the high-order moments are recursively constructed by starting from the second-order indicator moments. Figure 4 illustrates the approximation process. First, second-order indicator moments $M_{k_0 k_p}(o_p; h_p), p = 1 \dots n$, depicted by red solid lines in Fig. 4(a), are calculated from the basic variogram model. Then, the trend $M_{k_0,k_1,k_2}^0(\mathbf{o}; \mathbf{h})$, which is the surface in Fig. 4(a), is calculated using these boundary conditions and Eq. (16). Next, the residuals $\delta M_{k_0,k_1,k_2}(\mathbf{o}; \mathbf{h}^d)$, which are the black points in Fig. 4(b), are estimated by subtracting the trend $M_{k_0,k_1,k_2}^0(\mathbf{o}; \mathbf{h})$ from the sampling statistics $\hat{M}_{k_0,k_1,k_2}(\mathbf{o}; \mathbf{h}^d)$ using Eq. (18). Then, residual spatial moments $\delta M_{k_0,k_1,k_2}(\mathbf{o}; \mathbf{h})$, depicted as the surface in Fig. 4(c), are approximated from residuals $\delta M_{k_0,k_1,k_2}(\mathbf{o}; \mathbf{h}^d)$ and zero boundary conditions (red lines in Fig. 4c), using the B-spline regression in Eq. (17). Finally, the third-order spatial indicator moments, shown as the surface in Fig. 4(d), are retrieved by adding residual spatial moments $\delta M_{k_0,k_1,k_2}(\mathbf{o}; \mathbf{h})$ to the trend $M_{k_0,k_1,k_2}^0(\mathbf{o}; \mathbf{h})$ using Eq. (15).

The calculated third-order spatial indicator moments $M_{k_0,k_1,k_2}(\mathbf{o}; \mathbf{h})$ are used as the boundary conditions for the fourth-order spatial indicator moments $M_{k_0,k_1,k_2,k_3}(\mathbf{o}; \mathbf{h})$ (Fig. 5). Note that the fourth-order moment is a three-dimensional function that requires eight boundary conditions. The same procedure is recursively repeated for the fourth and fifth orders.
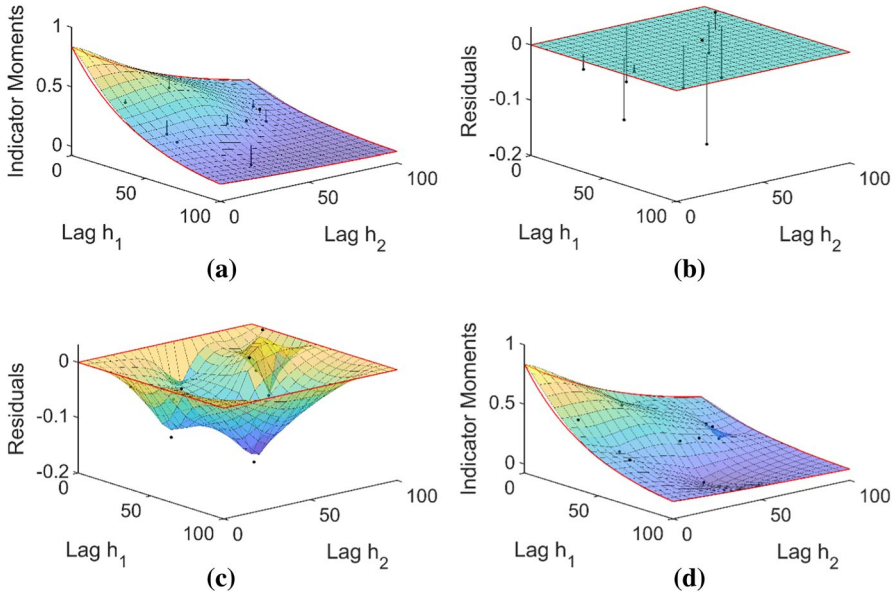
Fig. 4 Approximation of the third-order spatial indicator moments $M_{k_0,k_1,k_2}(\mathbf{o};\mathbf{h})$: **a** calculating the trend $M^0_{k_0,k_1,k_2}(\mathbf{o};\mathbf{h})$ from boundary conditions shown by red lines, **b** residuals $\delta M_{k_0,k_1,k_2}(\mathbf{o};\mathbf{h}^d)$ with sampling statistics and zero boundary conditions, **c** B-spline regression of the residual third-order map $\delta M_{k_0,k_1,k_2}(\mathbf{o};\mathbf{h})$, **d** the final reconstructed third-order spatial indicator moments $M_{k_0,k_1,k_2}(\mathbf{o};\mathbf{h})$

4 Applications

4.1 Olympic Dam Copper Deposit, Australia

Olympic Dam, located in South Australia, is the fourth largest copper deposit in the world. A part of the Olympic Dam deposit covering an area 1 km by 1 km is used here to demonstrate the application of the proposed simulation method in a
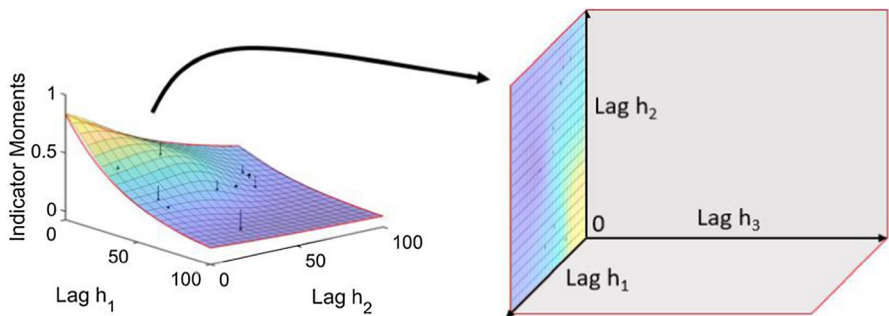
Fig. 5 The third-order indicator moments (left) $M_{k_0,k_1,k_2}(\mathbf{o};\mathbf{h})$ as the boundary condition for the fourth-order indicator moments $M_{k_0,k_1,k_2,k_3}(\mathbf{o};\mathbf{h})$

Fig. 6 Drill-holes from Olympic Dam copper deposit. Red represents grades above 1% Cu and blue below
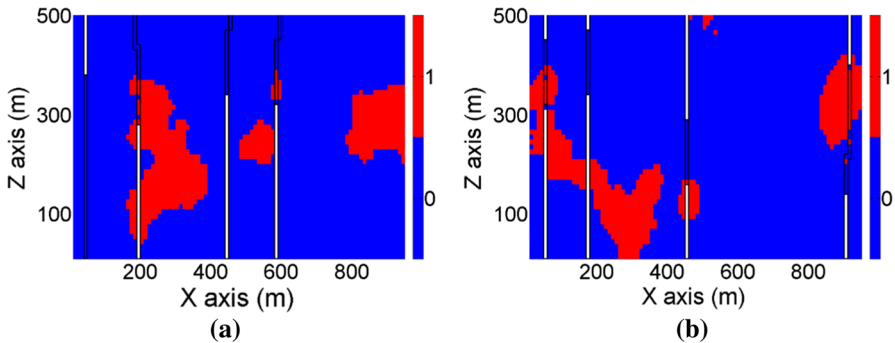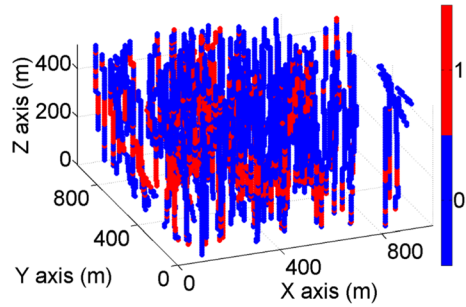
Fig. 7 Vertical sections of a simulated realization. Red represents grades above 1% Cu and blue below. Drill-holes are traced by black outlines, and white sections represent missing values

case study. The grade cut-off of 1% Cu is applied to the available drillhole data to define two categories to be simulated; then results are validated using high-order spatial indicator moments. Data are available in 515 exploration drill-holes shown in Fig. 6. Three-dimensional simulated realizations are generated on a regular grid with $100 \times 100 \times 50$ grid nodes and a block of $10 \times 10 \times 10$ m^3. Figure 7 depicts vertical sections from a simulated realization of the deposit. The figure shows that the proposed method generates spatially complex geometric structures that honor the drill-hole data. The spatial statistics and contacts are validated using high-order spatial indicator moments.

It should be noted that the simulations are performed using all possible directions from the quadraginta octant division. For the third-order indicator moments, there are 1,128 possible combinations of directions, and each of the combinations has 2^3 possible combinations of categories 0 and 1. Thus, the total number of indicator maps is 9,024. For the fourth-order indicator moments, there are 17,296 possible combinations of directions, which results in 138,368 forth-order indicator maps. Without a doubt, the information in these indicator maps is quite similar, and the complex high-order relations can be expressed using a smaller number of moments. Therefore, only the orthogonal directions, so-called L-shape templates, are used for validation purposes.

The third-order indicator moment maps $\hat{M}_{111}(\mathbf{h}_1, \mathbf{h}_2)$ are calculated using the L-shape template (Mustapha and Dimitrakopoulos 2010a) with lags $\mathbf{h}_1 = (idx, 0)$ $\mathbf{h}_2 = (0, jdy)$, indexed by $i = 0 \ldots 8$, $j = 0 \ldots 8$, where dx and dy are 50 m $\times$ 50 m. The moment map $\hat{M}_{111}(\mathbf{h}_1, \mathbf{h}_2)$ shows the probability of having category 1 in three points separated from each-other by lags along X and Y (L-shape). Note that validation of high-order moments is done with an approach with regular steps and two directions as in Mustapha and Dimitrakopoulos (2010a, b), in contrast to the octant approach with logarithmically increasing lags in Sect. 3. The logarithmically increasing lags with small steps close to the origin are helpful to better inform the approximation of high-order moments close to the central node, whereas for validation it is important to visualize both short-range and long-range connectivity. The moment map shows the probability of having copper grades above 1% at three points separated by lags along X and Y, that is, the continuity of high grades. As can be seen in Fig. 8, the simulated realization generated by the proposed method (Fig. 8b) reproduces red and yellow areas in the moment map of the hard data (Fig. 8a), that is, continuity ranges. As shown above, values along axes are indicator moments of order 2, that is the conventional indicator covariances. The third-order indicator moment maps $\hat{M}_{101}(\mathbf{h}_1, \mathbf{h}_2)$ are shown in Fig. 9. The moment map shows the probability of having copper grades above 1% at two points in the Y direction and copper grade below 1% at one point in the X direction. $\hat{M}_{101}(\mathbf{h}_1, \mathbf{h}_2)$ reflects the contact plane between the copper grades above and below 1%, respectively, in the east–west direction. As can be seen in Fig. 9, the simulation using the proposed method (Fig. 9b) reproduces the spatial characteristics of the data between copper grades above and below 1% (Fig. 9a), represented by the yellow–red cone in the lower right part of the moment maps in Fig. 9.

The third-order indicator moment maps $\hat{\mathbf{M}}_{110}(h_1, h_2)$ are shown in Fig. 10. Similarly to the above, the moment map shows the probability of having copper grades above 1% at two points in the X direction and copper grade below 1% at one point in the Y direction. $\hat{M}_{101}(\mathbf{h}_1, \mathbf{h}_2)$ reflects the contact plane between copper grades above and below 1%, respectively, in the north–south direction. As can be seen in Fig. 10, the realization generated using the proposed method (Fig. 10b) reproduces the
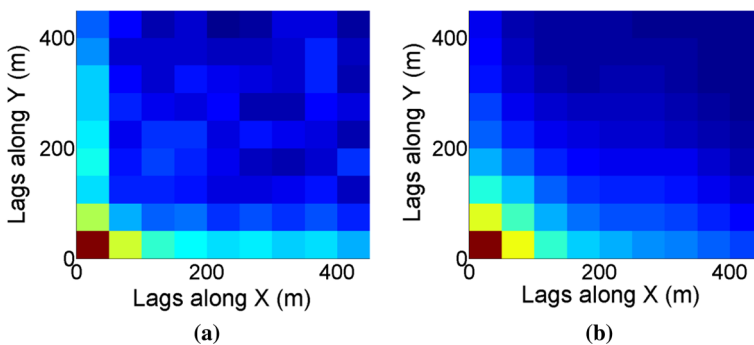

Fig. 8 The third-order indicator moment maps $\hat{M}_{111}(\mathbf{h}_1, \mathbf{h}_2)$ of **a** the available data, and **b** a simulated realization generated with the proposed method
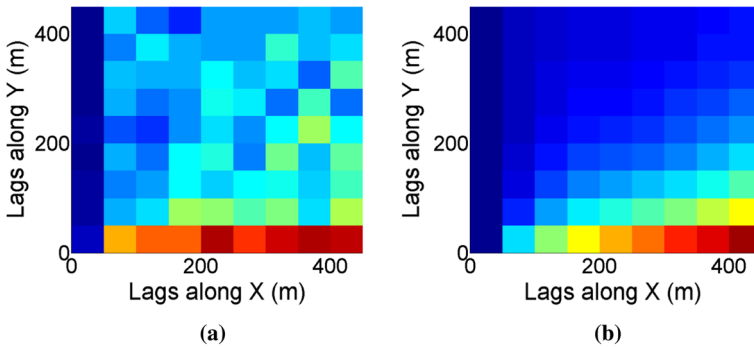
Fig. 9 The third-order indicator moment maps $\hat{M}_{101}(\mathbf{h}_1, \mathbf{h}_2)$ of **a** the data, and **b** a simulation using the proposed method

contact between copper grades above and below 1% as found in the data (Fig. 10a), represented by the yellow–red cone in the upper left part of the moment maps in Fig. 10.

The fourth-order spatial indicator moment $\hat{M}_{1000}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ (Fig. 11) is calculated with lags $\mathbf{h}_1 = (i\mathrm{d}x, 0)$, $\mathbf{h}_2 = (0, j\mathrm{d}y)$, and $\mathbf{h}_3 = (0, k\mathrm{d}z)$ indexed by $i, j, k = 1 \ldots 8$, where $\mathrm{d}x$, $\mathrm{d}y$, and $\mathrm{d}z$ are 50 m × 50 m × 50 m. The moment $\hat{M}_{1000}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ reflects the complex contact between copper grades above and below 1%, where grades with values above 1% are surrounded by grades below 1% in the X, Y, and Z directions. The fourth-order moment map of the simulation using the proposed method (Fig. 11b) reproduces the yellow–red area in the moment map of the data (Fig. 11a). All other spatial indicator moments, that is, $\hat{M}_{k_1 k_2 k_3}(\mathbf{h}_1, \mathbf{h}_2)$, $\hat{M}_{k_1 k_2 k_3 k_4}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ $, \forall k_1, k_2, k_3, k_4 = 0, 1$, were analyzed and found consistent with the spatial indicator moments from the drill-hole data.

To highlight the importance of high-order calculations, a realization from the sequential indicator simulation method (SISIM) is analyzed in terms of third-order spatial indicator moments $\hat{M}_{k_1 k_2 k_3}(\mathbf{h}_1, \mathbf{h}_2)$. A section of the realization, shown
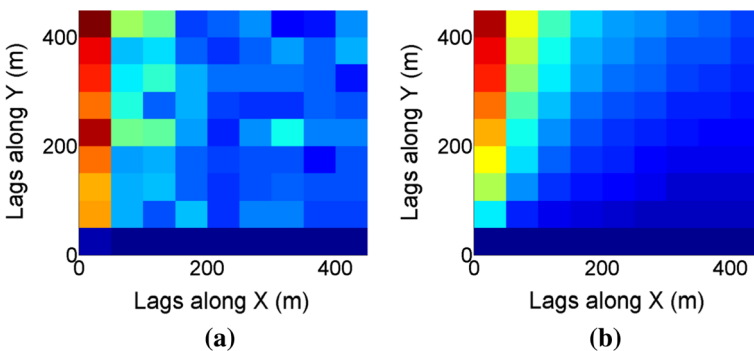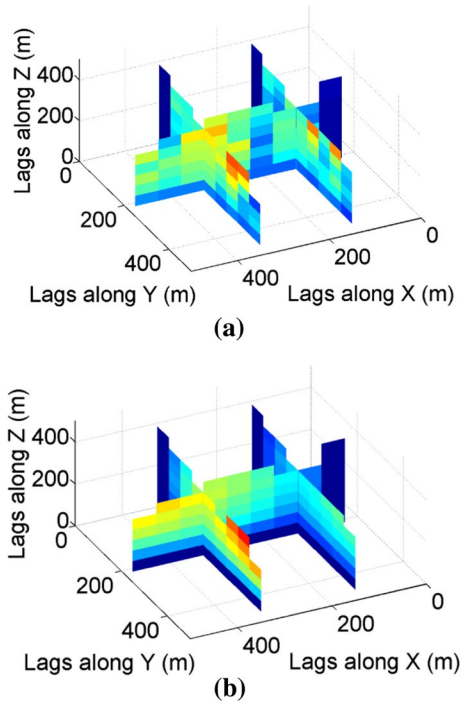

Fig. 10 The third-order indicator moment maps $\hat{\mathbf{M}}_{110}(h_1, h_2)$ of **a** the data, and **b** a simulation using the proposed method

Fig. 11 The fourth-order indicator moment maps $\hat{M}_{1000}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ of **a** the data, and **b** a simulation using the proposed method

(a)

(b)

in Fig. 12(a), exhibits low nonlinear connectivity and a high number of small disconnected shapes. This is confirmed by third-order indicator moments $\hat{M}_{111}(\mathbf{h}_1, \mathbf{h}_2)$, $\hat{M}_{101}(\mathbf{h}_1, \mathbf{h}_2)$, $\hat{M}_{110}(\mathbf{h}_1, \mathbf{h}_2)$ in Fig. 12b–d, respectively. In contrast to Figs. 8, 9, and 10, the third-order indicator moments from the SISIM method have low values for non-zero lags $(\mathbf{h}_1, \mathbf{h}_2)$ and a high contrast between values along axes, i.e. second-order statistics and values away from the axes. This indicates that the realization from the SISIM method does not reproduce complex relations of data and exhibits lower nonlinear connectivity of related categories.

4.2 Escondida Norte Copper Deposit, Chile

Escondida is a large porphyry copper deposit in Chile consisting of two open-pit mines, Escondida and Escondida Norte. A part of Escondida Norte, 2.5 km by 2.5 km by 0.5 km, is used in this section to present a case study. Four mineralization zones are simulated using the proposed approach: oxides, sulfides, mix of oxides and sulfides, and waste. Complex geometrical shapes of mineralization zones and geological contacts are validated here using high-order spatial indicator moments. High-order spatial indicator moments allow us to analyze cross-categorical relations and take into account geological aspects of mineral deposits, such as which category is always embedded within another, which categories cannot be in contact, and so on.
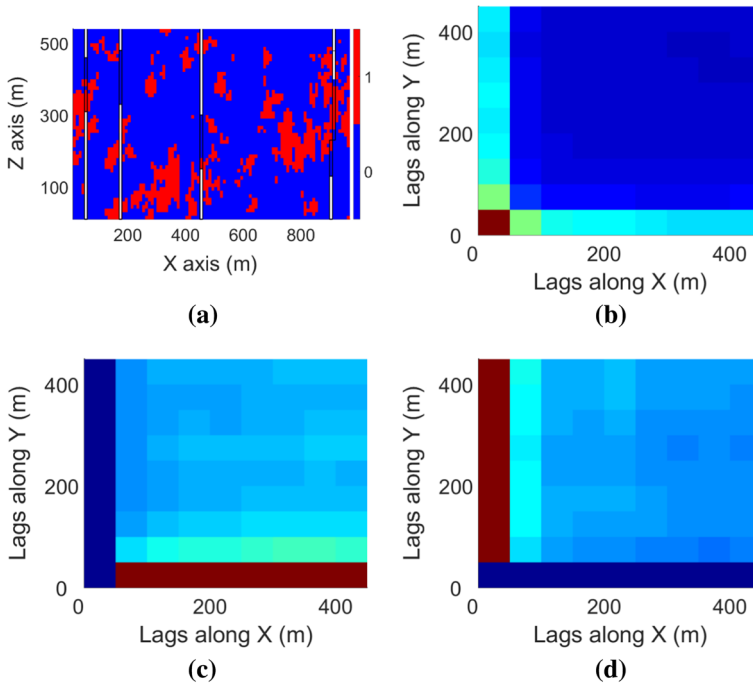
Fig. 12 Section of a realization from the SISIM method **a** and corresponding third-order indicator moment maps. **b–d** $\hat{M}_{111}(\mathbf{h}_1, \mathbf{h}_2)$, $\hat{M}_{101}(\mathbf{h}_1, \mathbf{h}_2)$, $\hat{M}_{110}(\mathbf{h}_1, \mathbf{h}_2)$

The drill-holes available are shown in Fig. 13. In general, mineralization zones are quite variable, and the uncertainty of related contacts need to be quantified. The three-dimensional simulated realizations generated are defined on $115 \times 107 \times 55$ grid of blocks of $25 \times 25 \times 15$ m^3 size. Sulfides are predominantly located in the bottom part and are covered by layers of mix and oxide zones. The upper part of the deposit consists mostly of waste materials. Vertical and horizontal sections of two simulated realizations using the proposed method are shown in Figs. 14 and 15, respectively. The simulations honor the layered structure of the mineralization zones and demonstrate higher variability of oxides and mix mineralization zones.

Fig. 13 Drill-holes from Escondida Norte copper mine; colors represent the mineralization zones
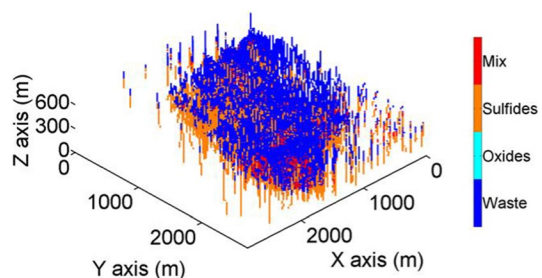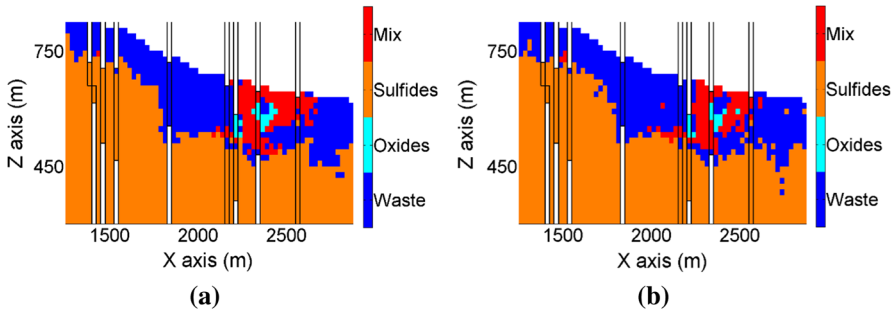
Fig. 14 Vertical section from two simulated realizations; colors represent mineralization, and drill-hole locations are represented by black outlines
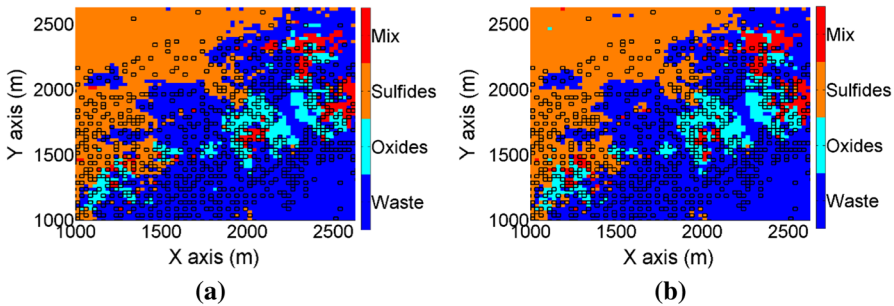

Fig. 15 Horizontal section from two simulated realizations; colors represent mineralization, and drill-hole locations are represented by black outlines

The third-order indicator moment maps $\hat{M}_{OSS}(\mathbf{h}_1, \mathbf{h}_2)$ (O stands for oxides, S for sulfides, M for mix, and W for waste) are calculated using lags $\mathbf{h}_1 = (idx, 0)$ $\mathbf{h}_2 = (0, jdy)$ indexed by $i = 0 \ldots 8$, $j = 0 \ldots 8$, where dx and dy are 50 m × 50 m. The moment map shows the probability of having oxide separated from sulfides by lags along X and Y, that is, a complex contact between oxides and sulfides. As can be seen in Fig. 16, the simulated realization generated with the proposed method (Fig. 16b) reproduces the red and yellow areas in the moment map of the data (Fig. 16a). The third-order indicator moment maps $\hat{M}_{SSW}(\mathbf{h}_1, \mathbf{h}_2)$ are shown in Fig. 17. The moment map shows the probability of having sulfides at two points in the X direction and waste in the Y direction. $\hat{M}_{SSW}(\mathbf{h}_1, \mathbf{h}_2)$ reflects the contact plane between sulfides and waste with a north to north–south direction. As can be seen in Fig. 17, the simulation using the proposed method (Fig. 17b) reproduces the corresponding spatial relations found in the data (Fig. 17a).

The fourth-order spatial indicator moments $\hat{M}_{MOWM}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ (Fig. 18) are calculated with lags $\mathbf{h}_1 = (idx, 0)$, $\mathbf{h}_2 = (0, jdy)$, and $\mathbf{h}_3 = (0, kdz)$ indexed by $i, j = 1 \ldots 8$, $k = 1 \ldots 5$, where dx, dy, and dz are 50 m × 50 m × 50 m. The moment $\hat{M}_{MOWM}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ reflects the spatial aspects of the contact between mixed oxides and waste zones. The fourth-order moment maps of the simulation
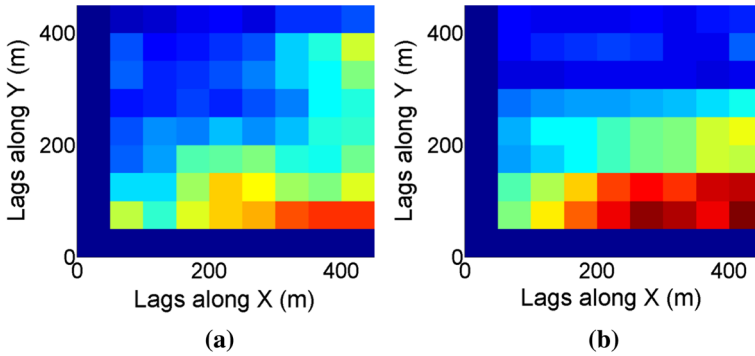
Fig. 16 The third-order indicator moment maps $\hat{M}_{OSS}(\mathbf{h}_1, \mathbf{h}_2)$ of (a) the data, and (b) a simulated realization generated with the proposed method
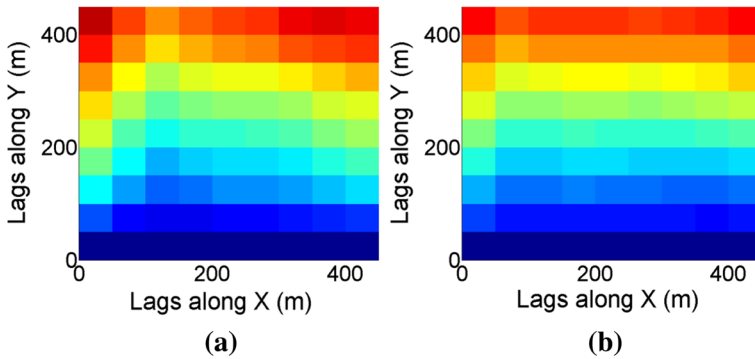

Fig. 17 The third-order indicator moment maps $\hat{M}_{SSW}(\mathbf{h}_1, \mathbf{h}_2)$ of **a** the hard data, and **b** the simulation using the proposed method

using the proposed method (Fig. 18b) reproduce the yellow–red area in the moment map of the data (Fig. 18a).

Similarly to the above, all other spatial indicator moments, that is, $\hat{M}_{k_1 k_2 k_3}(\mathbf{h}_1, \mathbf{h}_2)$, $\hat{M}_{k_1 k_2 k_3 k_4}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3), \forall k_1, k_2, k_3, k_4 = \{S, O, M, W\}$, of the simulated realizations were analyzed and found consistent with the spatial indicator moments from the drill-hole data.

Fig. 18 The fourth-order indicator moment maps $\hat{M}_{MOWM}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ of **a** the hard data and **b** the simulation using the proposed method
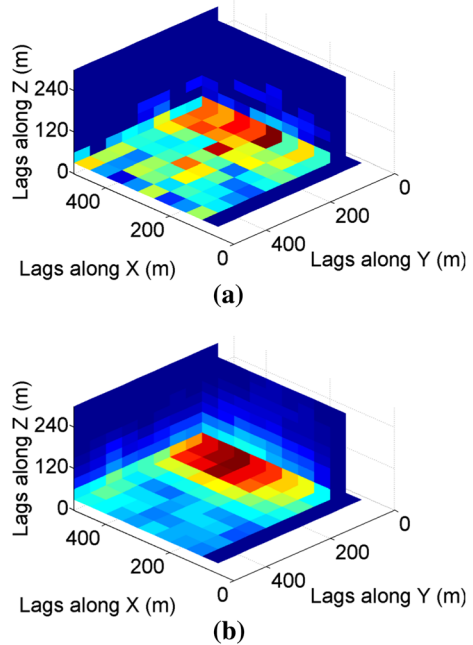
5 Conclusions

This paper presents a new data-driven, high-order sequential method for the simulation of categorical random fields. The sequential algorithm is based on the B-spline approximation of high-order spatial indicator moments that are consistent with each other. The main distinction from commonly used MPS methods is that in the proposed approach, conditional distributions are constructed using high-order spatial indicator moments as functions of distances based on hard data. Thus, simulated realizations can be generated without a TI. Note that in applications with relatively large data sets, such as the simulation of mineral deposits, the higher-order statistics are deduced from hard data. However, the option of adding a TI to a data set is available only if sparse data sets are available, as is the case with petroleum reservoirs.

The basic concept of the method presented is to use recursive approximation models with enclosed boundary conditions, which are derived from the nested nature of high-order spatial indicator moments, as presented herein. To provide a robust estimation, the regularized B-splines are used. An additional critical aspect of the proposed approach is that different amounts of information can be retrieved for different levels of relations. Each order of spatial statistics is approximated using the appropriate number of B-splines to provide robustness to the approach and to avoid overfitting. Thus, lower-order statistics are estimated with a higher resolution than the higher-order statistics.

The simulation algorithm presented was tested at two real copper deposits, without using TIs. The results of the applications demonstrate that the proposed method reproduces complex spatial patterns and preserves the high-order spatial statistics in

the data. While the proposed technique is fully data-driven, information from a TI can be incorporated with the proposed model as a trend to capture high-frequency features when available in the TI. Further research may consider improving the approximation methods.

Appendix

Consider an arbitrary spatial configuration of three points in space. Using Eqs. (7) and (8), the third-order spatial indicator moment can be expressed as

$$M_{k_0,k_1,k_2}(Z_0, Z_1, Z_2) = E\big[I_{k_0}(Z_0)I_{k_1}(Z_1)I_{k_2}(Z_2)\big]. \tag{20}$$

Let us consider that Z_0 has the same location as Z_1, i.e. $h_1 = x_1 - x_0 = 0$. Then, Z_0 has the same value as Z_1

$$M_{k_0,k_1,k_2}(Z_0, Z_1, Z_2) = E\big[I_{k_0}(Z_0)I_{k_1}(Z_1)I_{k_2}(Z_2)\big] = E\big[I_{k_0}(Z_0)I_{k_1}(Z_0)I_{k_2}(Z_2)\big], \text{ when } x_0 = x_1. \tag{21}$$

It is not hard to see that the term $I_{k_0}(Z_0)I_{k_1}(Z_0)$ is zero when $k_1 \neq k_0$, and otherwise equal to $\big[I_{k_0}(Z_0)\big]^2 = I_{k_0}(Z_0)$. Therefore,

$$M_{k_0,k_1,k_2}(Z_0, Z_1, Z_2) = E\big[I_{k_0}(Z_0)I_{k_2}(Z_2)\big]\delta_{k_0,k_1} = M_{k_0,k_2}(Z_0, Z_2)\delta_{k_0,k_1}, \text{ when } x_0 = x_1. \tag{22}$$
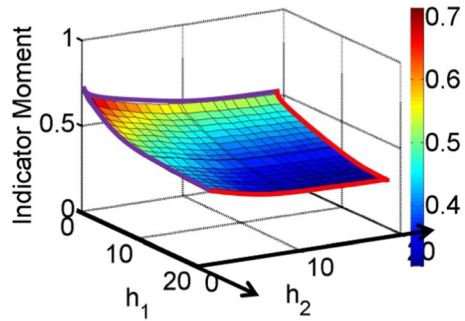
Thus, Eq. (22) connects third-order spatial statistics with second-order statistics and can be generalized for high-order spatial statistics in the form of Eq. (13)

$$M_{\mathbf{k}}(\mathbf{o}; h_1, \dots, h_p = 0, \dots, h_n) = M_{\mathbf{k}\setminus k_p}(\mathbf{o}; \mathbf{h}\setminus h_p)\delta_{k_0,k_p}, \forall p \in 1 \dots n,$$

In Eq. (20), let us consider the other extreme $h_1 \to \infty$; then, under the assumption of finite spatial correlation length of the random field $\mathbf{Z}$, the random variable Z_1 can be considered as independent from random variables Z_0 and Z_2. Then, using Eq. (3),

$$M_{k_0,k_1,k_2}(Z_0, Z_1, Z_2) = P(Z_0 = k_0, Z_1 = k_1, Z_2 = k_2) = P(Z_0 = k_0, Z_2 = k_2)P(Z_1 = k_1)$$

Fig. 19 The three-point spatial indicator moment map $M_{000}(\mathbf{o}; h_1, h_2)$. The solid violet lines represent second-order spatial statistics at $h_1 = 0$ or $h_2 = 0$, and the solid red lines represent second-order spatial statistics when either $h_1 \to \infty$ or $h_2 \to \infty$

$$M_{k_0,k_1,k_2}(Z_0, Z_1, Z_2) = M_{k_0,k_2}(Z_0, Z_2) M_{k_1}(Z_1). \tag{23}$$

Thus, Eq. (23) connects third-order spatial statistics with second-order statistics and can be generalized for high-order spatial statistics in the form of Eq. (14)

$$M_{\mathbf{k}}(\mathbf{o}; h_1, \ldots, h_p \to \infty, \ldots, h_n) = M_{\mathbf{k}\backslash k_p}(\mathbf{o}; \mathbf{h}\backslash h_p) M_{k_p}, \forall p \in 1 \ldots n$$

.Graphically, for the third-order spatial indicator moments, the boundary conditions, i.e. second-order spatial indicator moments, are show on Fig. 19. The three-point spatial indicator moment $M_{000}(\mathbf{o}; h_1, h_2)$ with $k_0 = k_1 = k_2 = 0$ is considered. The solid violet lines represent second-order spatial statistics at $h_1 = 0$ or $h_2 = 0$, i.e. $M_{000}(\mathbf{o}; h_1 = 0, h_2) = M_{00}(o_2; h_2)$ and $M_{000}(\mathbf{o}; h_1, h_2 = 0) = M_{00}(o_1; h_1)$, respectively. The red solid lines represent second-order spatial statistics when either $h_1 \to \infty$ or $h_2 \to \infty$, i.e. $M_{000}(\mathbf{o}; h_1 \to \infty, h_2) = M_{00}(o_2; h_2) M_0(Z_1)$ and $M_{000}(\mathbf{o}; h_1, h_2 \to \infty) = M_{00}(o_1; h_1) M_0(Z_2)$, respectively.

In summary, all-order moments are connected via boundary conditions:

1. Marginal distribution $P(Z_0 = k_0) = M_{k_0}(Z_0)$, i.e. the first-order moment is the boundary for the second-order moment $M_{k_0,k_1}(Z_0, Z_1)$

$$M_{k_0,k_1}(Z_0, Z_1) = E[I_{k_0}(Z_0) I_{k_1}(Z_1)] = E[I_{k_0}(Z_0)] \delta_{k_0,k_1} = M_{k_0}(Z_0) \delta_{k_0,k_1}, \text{ when } x_0 = x_1$$

2. The second-order moment is the boundary for the third-order moment. For example,

$$M_{k_0,k_1,k_2}(Z_0, Z_1, Z_2) = E[I_{k_0}(Z_0) I_{k_2}(Z_2)] \delta_{k_0,k_1} = M_{k_0,k_2}(Z_0, Z_2) \delta_{k_0,k_1}, \text{ when } x_0 = x_1.$$

3. The third-order moment is the boundary for the fourth-order moment. For example,

$$\begin{aligned} M_{k_0,k_1,k_2,k_3}(Z_0, Z_1, Z_2, Z_3) &= E[I_{k_0}(Z_0) I_{k_1}(Z_1) I_{k_2}(Z_2) I_{k_3}(Z_3)] \\ &= E[I_{k_0}(Z_0) I_{k_1}(Z_1) I_{k_2}(Z_2)] \delta_{k_0,k_3} \\ &= M_{k_0,k_1,k_2}(Z_0, Z_1, Z_2) \delta_{k_0,k_3}, \text{ when } x_0 = x_3. \end{aligned}$$

References

Babenko K (1986) Fundamentals of numerical analysis. Nauka, Moscow

Biswas R, Bhowmick P (2017) On the functionality and usefulness of quadraginta octants of naive sphere. J Math Imaging Vis 59(1):69–83

David M (1988) Handbook of applied advanced geostatistical ore reserve estimation. Elsevier, Amsterdam

de Carvalho JP, Dimitrakopoulos R, Minniakhmetov I (2019) High-order block support spatial simulation method and its application at a gold deposit. Math Geosci 51(6):793–810. https://doi.org/10.1007/s11004-019-09784-x

Dimitrakopoulos R, Luo X (2004) Generalized sequential Gaussian simulation on group size ν and screen-effect approximations for large field simulations. Math Geol 36(5):567–590. https://doi.org/10.1023/B:MATG.0000037737.11615.df

Dimitrakopoulos R, Mustapha H, Gloaguen E (2010) High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena. Math Geosci 42(1):65–99. https://doi.org/10.1007/s11004-009-9258-9

Evans J, Bazilevs Y, Babuska I, Hughes T (2009) N-widths, sup–infs, and optimality ratios k-version of the isogeometric finite element method. Comput Methods Appl Mech Engrg 198:1726–1741

Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning. Springer, New York

Gómez-Hernández JJ, Srivastava RM (2021) One step at a time: the origins of sequential simulation and beyond. Math Geosci 53(2):193–209. https://doi.org/10.1007/s11004-021-09926-0

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York

Guardiano J, Srivastava RM (1993) Multivariate geostatistics: beyond bivariate moments. In: Soares A (ed) Geostatistics Tróia'92, vol 1. Springer, Dordrecht, pp 133–144

Journel A G (1993) Geostatistics: roadblocks and challenges. In: 7th Annual Report, Stanford Center for Reservoir Forecasting, Stanford University, Stanford, CA

Journel A G (2003) Multiple-point geostatistics: a state of the art. In: 16th Annual Report, Stanford Center for Reservoir Forecasting, Stanford University, Stanford, CA

Journel AG, Alabert F (1990) New method for reservoir mapping. Petrol Technol 42(2):212–218

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London

Mariethoz G, Kelly B (2011) Modeling complex geological structures with elementary training images and transform-invariant distances. Water Resour Res 47:1–2

Mariethoz G, Renard P (2010) Reconstruction of incomplete data sets or images using direct sampling. Math Geosci 42(3):245–268

Minniakhmetov I, Dimitrakopoulos R (2017a) Joint high-order simulation of spatially correlated variables using high-order spatial statistics. Math Geosci 49(1):39–66. https://doi.org/10.1007/s11004-016-9662-x

Minniakhmetov I, Dimitrakopoulos R (2017b) A High-order, data-driven framework for joint simulation of categorical variables. In: Gómez-Hernández J, Rodrigo-Ilarri J, Rodrigo-Clavero M, Cassiraga E, Vargas-Guzmán J (eds) Geostatistics Valencia 2016. Springer, Cham, pp 287–301

Minniakhmetov I, Dimitrakopoulos R, Godoy M (2018) High-order spatial simulation using Legendre-like orthogonal splines. Math Geosci 50(7):753–780. https://doi.org/10.1007/s11004-018-9741-2

Mustapha H, Dimitrakopoulos R (2010a) A new approach for geological pattern recognition using high-order spatial cumulants. Comput Geosci 36(3):313–334

Mustapha H, Dimitrakopoulos R (2010b) High-order stochastic simulation of complex spatially distributed natural phenomena. Math Geosci 42(5):457–485. https://doi.org/10.1007/s11004-010-9291-8

Pyrcz M, Deutsch C (2014) Geostatistical reservoir modeling, 2nd edn. Oxford University Press, Oxford

Remy N, Boucher A, Wu J (2009) Applied geostatistics with SGeMS: a user's guide. Cambridge University Press, Cambridge

Ripley BD (1987) Stochastic simulation. Wiley, New York

Stien M, Kolbjørnsen O (2011) Facies modeling using a Markov mesh model specification. Math Geosci 43:611

Straubhaar JRP (2011) An improved parallel multiple-point algorithm using a list approach. Math Geosci 43(3):305–328

Strebelle S (2002) Conditional simulation of complex geological structures using multiple point statistics. Math Geol 34(1):1–21. https://doi.org/10.1023/A:1014009426274

Strebelle S (2021) Multiple-point statistics simulation models: pretty pictures or decision-making tools? Math Geosci 53(2):267–278. https://doi.org/10.1007/s11004-020-09908-8

Strebelle S, Cavelius C (2014) Solving speed and memory issues in multiple-point statistics simulation program SNESIM. Math Geosci 46(6):171–186. https://doi.org/10.1007/s11004-013-9489-7

Toftaker H, Tjelmeland H (2013) Construction of binary multi-grid Markov random field prior models from training images. Math Geosci 45(4):383–409

Vargas-Guzman J (2011) The Kappa model of probability and higher-order rock sequences. Comput Geosci 15:661–671

Vargas-Guzman J, Qassab H (2006) Spatial conditional simulation of facies objects for modeling complex clastic reservoirs. J Petrol Sci Eng 54:1–9

Yao L, Dimitrakopoulos R, Gamache M (2018) A new computational model of high-order stochastic simulation based on spatial Legendre moments. Math Geosci 50(8):929–960

Yao L, Dimitrakopoulos R, Gamache M (2020) High-order sequential simulation via statistical learning in reproducing kernel Hilbert space. Math Geosci 52:693–723. https://doi.org/10.1007/s11004-019-09843-3

Yao L, Dimitrakopoulos R, Gamache M (2021) Training image free high-order stochastic simulation based on aggregated kernel statistics. Math Geosci. https://doi.org/10.1007/s11004-021-09923-3

Zhang T, Switzer P, Journel A (2006) Filter-based classification of training image patterns for spatial simulation. Math Geol 38(1):63–80. https://doi.org/10.1007/s11004-005-9004-x

Zhang T, Gelman A, Laronga R (2017) Structure- and texture-based fullbore image reconstruction. Math Geosci 49(2):195–215. https://doi.org/10.1007/s11004-016-9649-7

Osterholt V, Dimitrakopoulos R (2007) Simulation of wireframes and geometric features with multiple-point techniques: application at Yandi iron ore deposit. The Australasian Institute of Mining and Metallurgy (AusIMM) Spectrum Series, vol 14, pp 51–60

Chugunova TL, Hu LY (2008) Multiple-point simulations constrained by continuous auxiliary data. Math Geosci 40:133–146

Goodfellow R, Consuegra FA, Dimitrakopoulos R, Lloyd T (2012) Quantifying multi-element and volumetric uncertainty, Coleman McCreedy deposit, Ontario, Canada Comput Geosci 42:71–78