Les Cahiers du GERAD ISSN: 0711–2440

Multiple-point simulation of Olympic Dam Copper Deposit, Australia, using wavelet analysis

M. Teixeira
R. Dimitrakopoulos

G–2015–91

September 2015

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Multiple-point simulation of Olympic Dam Copper Deposit, Australia, using wavelet analysis

Murilo Teixeira [a]

Roussos Dimitrakopoulos [a,b]

[a] *COSMO – Stochastic Mine Planning Laboratory, Department of Mining and Materials Engineering, McGill University, Montreal (Quebec) Canada, H3A 2A7*

[b] *and GERAD*

murilo.teixeira@mail.mcgill.ca
roussos.dimitrakopoulos@mcgill.ca

September 2015

Les Cahiers du GERAD

G–2015–91

Abstract: Traditional geostatistical simulation methods assume that the first two order statistics are sufficient to model mineral deposits. However, these methods are not able to model some complex pattern present in geological environments and the spatial connectivity of high values. So, multiple-point simulation (MPS) methods were developed as an attempt to overcome these limitations. In this work a MPS approach based on wavelet analysis is applied to Olympic Dam deposit, to simulate both material types and grades. The method, termed *wavesim*, work as follows: first, it scans a training image with a template to generate a pattern database; then this database has its dimension reduced by applying wavelet analysis; after that, the patterns are divided into classes using k-means clustering algorithm, considering the approximate sub-band image; and finally the grid is simulated by comparing the conditioning data event in each node with the classes prototypes and choosing a pattern from that class. Olympic Dam's simulation results show that *wavesim* can be applied successfully to a large instance. The resulting simulated realizations are analysed and validated in terms of histograms, variograms and high-order statistics, the latter being performed by using high-order spatial cumulants (HOSC).

1 Introduction

Geostatistical simulations provide a technical framework to spatially model attributes of interest in mineral deposits, and provide means to quantify geological uncertainty both in terms of grades and material types. Sets of simulated models of these geological attributes can then be considered as inputs for stochastic optimization in mining planning using stochastic optimization methods, hence enhancing the decision making process.

Traditionally, geostatistical simulations are generated through two-point simulation algorithms, such as sequential Gaussian simulation and sequential in indicator simulation (Goovaerts, 1997). Two-point statistics or variogram based methods present an important limitation: they are unable to model complex spatial patterns, such as curvilinear patterns and connectivity of extreme values. Multiple point simulations (MPS) methods, which rely on a training image instead of variograms, are developed to overcome these limitations. Early efforts include the *snesim* algorithm (Strebelle, 2002) which is based on extended normal equation simulation (Guardiano and Srivastava, 1993). In *snesim*, the conditional probability distribution is retrieving from a search tree, which is generated according to a training image or analogue. More recent improvements to *snesim* algorithm include removal of stationary assumptions (Strebelle and Zhang, 2005), parallelization and use of lists instead of search trees (Straubhaar et al., 2011), adjustments in order to decrease memory demand and increase speed (Strebelle and Cavelius, 2014), and GPU programming (Li et al., 2013). Modelling geological units in mineral deposits using *snesim* are shown in Jones et al. (2013), Boucher et al. (2014), Goodfellow et al. (2012) and Osterholt and Dimitrakopoulos (2007). Simulation of continuous attributes using MPS is introduced in *filtersim* algorithm (Zhang, Switzer and Journel, 2006) where the idea of grouping similar patterns into classes is presented, so that the pattern database is scanned just once, before simulation takes place. The grouping is performed based on some filter scores of the patterns. This reduces drastically the computational time of the algorithm and makes possible the application for relatively large size simulations. Other well known MPS methods include *simpat* (Arpat and Caers, 2007), *distpat* (Honarkhah and Caers, 2010), *cdfsim* (Mustapha et al., 2104) and direct sampling (Mariethoz et al., 2010).

Extension of the two-point statistics to high-order spatial statistics in geostatistical simulations is presented in Dimitrakopoulos et al. (2010). The proposed method tries to reproduce the high-order spatial cumulants (HOSC) from the data and training image in the simulated realizations. The main advantages of this methodology over the MPS ones showed before are being data-driven and to explicitly reproduce the high-order statistics of samples in the simulated realizations.

Since the simulation method used herein is based on wavelet analysis, it is important to define what it is. Wavelet analysis is widely used, and some examples of its application are signal processing and data compression. (Bénéteau and Fleet, 2011) present and intuitive introduction to this subject, and Mallat (1998) shows it in more detail. The idea in wavelet analysis applied to image compression is to decompose an image in two types of information: 1) average type information of the nearby pixels, called approximate sub-band of the image; and 2) how these pixels depart from local average. Both of them together keep all the information about the image, but the first one keeps most of its variability. Hence, the approximate sub-band is usually a sufficient representation of the complete image for several applications, and can be used in instead of the original image. The advantage of using the sub-band image instead of the original one is related to less memory requirements and efficiency, in computational terms. Gloaguen and Dimitrakopoulos (2009) apply wavelet transformation to geostatistical simulation, exploring the inter-scale dependency in the wavelet domain. Although the direct conditioning to hard data is simple in this technique, it is difficult to fitting the conditioning data in wavelet domain. Chatterjee et al. (2011) and (2012) also use wavelet analysis to simulate both continuous and categorical attributes from mineral deposits, termed *wavesim*. This method is described in detail in the next section.

In this paper, the *wavesim* algorithm is used to simulate both geological units and copper grades at the Olympic Dam, a copper deposit located in southern Australia. It has been exploited since 1988 and it is the fourth largest producer of copper in the world nowadays. Olympic Dam is a huge breccia complex, hosted by deformed and highly brecciated granite, which is slightly older than the mineralization. There are two

types mineralization: the strata-bound bornite-chalcopyrite-pyrite one; and chalcocite-bornite in lenses and cross-cutting veins. Four geological domains are defined according to the content of copper minerals.

In the following sections, the wavelet simulation method used here is outlined. Then, the application of such method to Olympic Dam deposit is presented and discussed. Finally, conclusions follow.

2 Multiple-point simulation using wavelet analysis

The wavelet transformation based simulation method of Chatterjee et al. (2012) is outlined in this section, pointing at the differences between categorical and continuous simulations, where appropriate.

2.1 Pattern database generation

The first step of the method used herein is to scan the training image with a given spatial template $ti_T(u)$, which is a multiple-point vector centered in u. Its definition is presented in (1).

$$ti_T\left(u\right) = \left\{ti\left(u\right),\ ti\left(u+h_2\right), \ldots, ti\left(u+h_\alpha\right), \ldots, ti\left(u+h_{nT}\right)\right\} \tag{1}$$

Here, h_α are lag vectors that define the geometry of the template containing nT nodes, and $\alpha = \{1, 2, \ldots, nT\}$. It is noteworthy that the pattern database built is independent of the pattern locations.

For the continuous simulation cases, the patterns are stored exactly as they appear in the training image. However, for a categorical training image with M categories, the training image is first transformed into M sets of binary values, according to:

$$I_m\left(u\right) = \begin{cases} 1, & \text{if } u \text{ belongs to category } m \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Accordingly, each location is represented by a vector of binary values, where the m^{th} element is 1 if that node belongs to category m, and 0 otherwise. Thus, for each node, there is exactly one element equal to 1.

2.2 Pattern database decomposition

Before splitting the pattern database into classes, the dimension of each of the patterns is reduced through discrete wavelet transformation (DWT) (Mallat 1998), so the classification can be done more efficiently. The dimension of the patterns is reduced by DWT by selecting the scale of wavelet decomposition, defined as the resolution or support size on which a given wavelength of a training image or analog is defined.

A pattern can be decomposed into a series of orthogonal basis function through DWT Mallat (1998), which provides both frequency and spatial information. The wavelet function provides a series of orthonormal base functions by scaling and shifting the original basis function, known as the mother wavelet function. Three-dimension data sets are decomposed into eight components, where one is the scaling and the other seven are the high-frequency components. A pattern ti_T with dimensions $N \times N$ can be decomposed as:

$$ti_T = \sum_{i,l=0}^{N_j-1} a_{J,i,l}\phi^{LL}_{J,i,l} + \sum_{B\in D}\sum_{j=1}^{J}\sum_{i,l=0}^{N_j-1} \omega^B_{j,i,l}\psi^B_{j,i,l} \tag{3}$$

where $D = \{LH, HL, HH\}$, L and H are low-pass and high-pass filters, $N_j = N/2^j$, J is the number of scales, $N = p$ when p is even and $N = (p+1)$ when p is odd, ϕ_j is the scaling function and ψ^B_j are the wavelet functions. The coefficients a_{j-1} and ω_{j-1} are calculated by taking the inner products between a pattern (ti_T) and scaling (ϕ_j) and wavelet functions $\left(\psi^B_j\right)$, respectively.

$$\begin{aligned} a_{j-1} &= \left\langle ti_T, \phi_j \right\rangle \\ \omega^B_{j-1} &= \left\langle ti_T, \psi^B_j \right\rangle \end{aligned} \tag{4}$$

The Haar wavelet is use as basis function, and is defined as follows:

$$\psi^H(x) = \begin{cases} 1 & 0 \le x < 1/2 \\ -1 & 1/2 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \phi^L(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The size of the scaling image for a M categories data set is:

$$LN = \left(\frac{N}{2^j}\right)^d xM \tag{6}$$

where d is the dimension of the image, and j is the number of scales. As the size of the original image is $\left(N^d\ xM\right)$, the reduction factor is 2^{jd}. Therefore, there is a substantial gain in computer efficiency if only the scaling component is used for the classification step.

2.3 Classification of the pattern database

After the patterns had their dimensionality reduced, they are divided into classes, using the k-means clustering algorithm (Hartigan and Wong, 1979; Lloyd, 1982). First, the number (k) of clusters needs to be provided. Then, k patterns are randomly chosen from the pattern database to be the initial centroids of the classes. Subsequently, the database is visited and each pattern is compared to the initial centroids; the class corresponding to the most similar centroid is the chosen one for that pattern. Then, the k new centroids are calculated by averaging all elements within each of the classes. This is performed iteratively, until the position of the centroids does not change any longer, which is when the final configuration of the clusters is reached. At each iteration, the objective is to minimize the following function:

$$Z = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| t_i^{(j)} - c_j \right\|^2 \tag{7}$$

where Z is the sum over all distances between the patterns and their respective class centroid, $t_i^{(j)}$ represents patterns i classified in cluster j, c_j is the centroid of class j and $\|t_i^{(j)} - c_j\|^2$ is the squared Euclidean distance between $t_i^{(j)}$ and c_j. When this process is completed, each class is labeled by its prototype, which is the average over all patterns in that class (same as centroid).

2.4 Simulation

Once the patterns are classified, spatial simulation is carried out based on the sequential simulation approach (Goovaerts, 1997). At each node to be simulated, the conditioning data event is retrieved by placing the same spatial template used for the generation of the pattern database on the realization grid. The data event comprises samples and previously simulated nodes. A distance function (Eq. 8 below) is used to compare the similarity between the conditioning data event and each of the classes' prototypes. The closest prototype, associated to a given class of patterns, is retrieved. In this application, L2-norm is used for the dissimilarity calculation between data events and classes prototypes:

$$d(x,y) = \sum_{i=1}^{3} \left\{ \omega_i * \left[\frac{1}{n_{type}} \sum_{j=1}^{n_{type}} \left(x(j) - y(j)\right)^2 \right] \right\} \tag{8}$$

$$\sum_{i=1}^{3} \omega_i = 1 \tag{9}$$

Here, $x(j)$ represents each node of the conditioning data event, $y(j)$ represents each node of the class prototype, n_{type} is the amount of data of a given data type and ω_i is the weight associate of data type i. Three data types are considered here: hard data, previously simulated node inside inner patch and previously simulated

node outside inner patch. The weight associated with hard data is the highest, and the one associate with previously simulated node outside inner patch is the lowest.

After that, a pattern from this best-matched class has to be drawn, and then pasted onto the simulation grid (an inner patch is used to identify the nodes of the pasted pattern that are frozen and no longer visited during the sequential simulation). For categorical simulation cases, a cumulative distribution function (cdf) relative to the central node is built for that class. Then, a Monte-Carlo sampling is done in that cdf in order to choose the category of the central node only. After that, a pattern is randomly chosen among the ones which have the central nodes belonging to the same category drawn in the Monte-Carlo sampling. For continuous simulations, a pattern is randomly drawn from the best matched class; no cdf is generated. This procedure is repeated until all nodes of the realization grid are simulated.

In summary, the steps *wavesim* algorithm follows are:

1. Generate the pattern database by scanning the training image with a given template.
2. Decompose the patterns using wavelet analysis.
3. Group these patterns into classes using k-means algorithm.
4. Calculate the prototypes of each class.
5. Define random path to visit all nodes to be simulated.
6. Compare data event to prototypes.
7. Choose a pattern from best matched.
8. Past it back onto simulation grid.
9. Repeat steps 6 to 8 until all nodes are simulated.
10. Repeat steps 5 to 9 to generate multiple realizations.

2.5 Training image generation

In this section, the methods used to generate both categorical and continuous training images to be used in the case studies are presented.

2.5.1 Categorical training image

The generation of categorical training images may be based on different information, according to its availability. Some examples are: outcrops, photographs of present day deposits or depositional systems, drawings from experts and geological and geophysical interpretation (Caers and Zhang, 2002). Boucher (2008) discusses the importance of training images for capturing various features of the area to be modeled, as well as some intricacies in its generation.

In mining, categorical training images are usually generated through a geological interpretation using drill hole data and geological background information (Jones et al., 2013; Boucher et al., 2014; and Goodfellow et al. 2012), as was the case of this work. Exhaustive information may also be used as in Osterholt and Dimitrakopoulos (2007), in which case information from a previously mined area from the same mine was available.

2.5.2 Continuous training image

The continuous training image used in the case study was generated using low rank tensor completion (LRTC) (Yahya, 2012 and Liu et al., 2013). The goal of tensor completion methods is to determine values for missing elements, considering all the information available.

Low-rank tensor completion (LRTC) can be formulated as an optimization problem as follows:

$$\begin{aligned} min \ & \sum_{i=1}^{n} \alpha_i \left\| X_{(i)} \right\|_* \\ s.t. \ & X_\Omega = T_\Omega \end{aligned} \tag{10}$$

where $\alpha_i \left\| X_{(i)} \right\|_*$ is the trace norm of tensor X, α are constants satisfying $\sum_{i=1}^{n} \alpha_i = 1$ and $\alpha_i \geq 0$ and $X_{(i)}$ represents the unfolded tensor along each mode. In this problem, the matrices $X_{(i)}$ share the same entries and as a result, they cannot be optimized directly. Therefore, additional matrices $M_1, \ldots, M_n$ are introduced and the problem can be relaxed:

$$\begin{aligned} min \ & \sum_{i=1}^{n} \alpha_i \left\| M_{(i)} \right\|_* + \frac{\beta_i}{2} \left\| X_{(i)} - M_i \right\|_F^2 \\ & s.t. \ \ X_\Omega = T_\Omega \end{aligned} \tag{11}$$

where $\beta_i \geq 0$ and $\|.\|_F$ is the Frobenius norm operator of a matrix. As this problem is convex, but non-differentiable, the block coordinate descent (BCD) can be used for its optimization. The basic idea of BCD is to optimize a group (block) of variable while the rest are fixed. The variables are divided in $n+1$ blocks: $X, \ M_1, \ M_2, \ldots, M_n$.

The optimal solution of X with all the other variables fixes is given by:

$$\begin{aligned} min \ & \sum_{i=1}^{n} \frac{\beta_i}{2} \left\| M_i - X_{(i)} \right\|_F^2 \\ & s.t. \ \ X_\Omega = T_\Omega \end{aligned} \tag{12}$$

And the solution is:

$$X_{i1,\ldots,in} = \begin{cases} \left(\frac{\sum_i \beta_i fold_i(M_i)}{\sum_i \beta_i} \right)_{i1,\ldots,in} & (i_1, \ldots, i_n) \notin \Omega \\ T_{i1,\ldots,in} & (i_1, \ldots, i_n) \in \Omega \end{cases}$$

Finally, M_i is given by solving the following:

$$\begin{aligned} min \ & \sum_{i=1}^{n} \alpha_i \left\| M_{(i)} \right\|_* + \frac{\beta_i}{2} \left\| X_{(i)} - M_i \right\|_F^2 \equiv \\ & \sum_{i=1}^{n} \frac{\alpha_i}{\beta_i} \left\| M_{(i)} \right\|_* + \frac{1}{2} \left\| X_{(i)} - M_i \right\|_F^2 \end{aligned} \tag{13}$$

hence, the optimal M_i is given by $D_\tau \left(X_{(i)} \right)$, where $D_\tau (.)$ is the shrinkage operator and $\tau = \alpha_i / \beta_i$.

3 Simulation of the Olympic Dam Copper Deposit, South Australia

3.1 The deposit

Olympic Dam deposit is part of Gawler Craton, and it is located in Australia, in the center of the province South Australia, approximately 520 km NNW of Adelaide. It is a very large (6 km × 3 km × 800 m) polymetallic orebody, containing Cu, U, Au and Ag. Nowadays, it is the fourth largest producer of copper and the largest producer or uranium ("Olympic Dam Mine", InfoMine Inc.). In this case study, only copper will be considered in the simulation. Olympic Dam is a huge breccia complex, hosted by deformed and highly brecciated granite, which is slightly older than the mineralization. It is covered by 300 meter layer of flat-lying sedimentary rocks. This deposit is a copper-gold type of mineralization: it presents a complex copper mineral zoning pattern, centered on a structurally controlled barren quartz-hematite breccia. There are two types mineralization: 1) the strata-bound bornite-chalcopyrite-pyrite one, confined to the Olympic Dam formation, and 2) chalcocite-bornite in lenses and cross-cutting veins in both Olympic Dam and neighboring formations. Moving outward/downward, the following copper minerals are more common: chalcocite-bornite, bornite, chalcopyrite-bornite, chalcopyrite and chalcopyrite-pyrite, where the highest grades are usually associated with bornite ± chalcopyrite. Sulfide mineral assemblages in the Olympic Dam deposit are demonstrably in equilibrium with ubiquitous hematite (Fe_2O_3) alteration of the granite host rock that is thought to be older than the sulfide deposition. The disappearance of chalcocite in favor of chalcopyrite and subsequently bornite for pyrite mark the locations where these reactions proceed to completeness (Roberts and Hudson, 1983, Skirrow et al., 2007, Belperio and Freeman, 2004, and Hahn, 2008).

3.2 Simulation of material types

3.2.1 Data and training image

The drill hole data set consists of 5-m composites and the total of sulfides is calculated as the sum of the following minerals: chalcocite (Cu_2S), bornite (Cu_5FeS_4), chalcopyrite ($CuFeS_2$) and pyrite (FeS_2). As mention before, the sulfides are zoned from inner bornite ± chalcocite through bornite ± chalcopyrite to outer chalcopyrite ± pyrite. The mineral content is calculated based on copper and sulfide sulfur assays, and they are believed to be at the same level of accuracy as the underlying assays. Sulfide-bearing intervals are defined as greater than or equal to 0.05 % of total sulfides, which is a somewhat arbitrary threshold value. Table 1 shows how the material types to be simulated are defined. In this table, BN, CC, PY and CPY mean, respectively, bornite, chalcocite pyrite and chalcopyrite. Domains 1 and 2 are the most important. Domain 3 also contains important amount of copper and, finally, domain 0 is mostly waste.

Table 1: Definition of material types used to classify data into categories. BN, CC, PY and CPY mean, respectively, bornite, chalcocite pyrite and chalcopyrite.

Geological Domain	Definition
0	Total Sulfides $< 0.10\%$
1	(BN+CC) ≥ 0.05 and (PY+CPY) < 0.05
2	(BN+CC) ≥ 0.05 and (PY+CPY) ≥ 0.05
3	(PY+CPY) ≥ 0.05 and (BN+CC) < 0.05

Figure 1 shows the samples within the study area. The colors correspond to the definitions in Table 1.

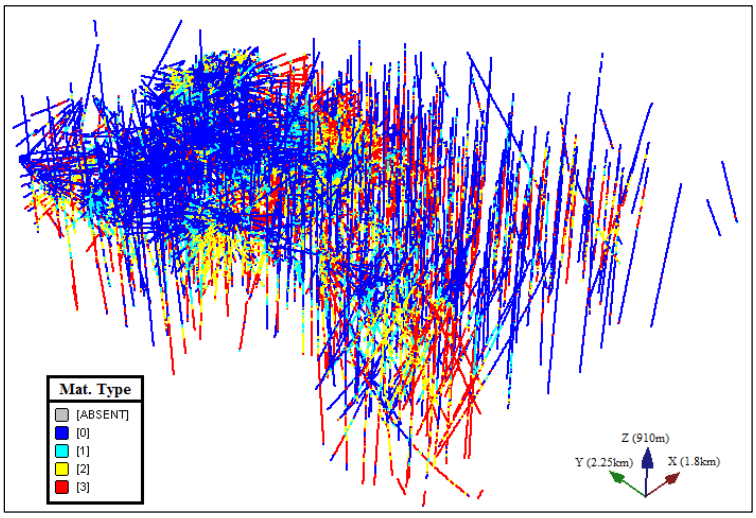

Figure 1: Drill hole samples of Olympic Dam copper deposit, colored according to material types defined in Table 1.

The training image is generated through a geological interpretation, as mentioned in Section 3.1, using the data in Figure 1. Figure 2 shows three sections of the training image generated, whereas Figure 3 displays the spatial configuration of material types 1 and 2, so that it is possible to see the spatial complexity of these units.

The training image is discretized in a grid of 15 m × 15 m × 5 m, resulting in 120 × 150 × 170 nodes in X, Y and Z directions, respectively and a total of 2,813,670 nodes. The deposit to be simulated is discretized the same way as the training image.

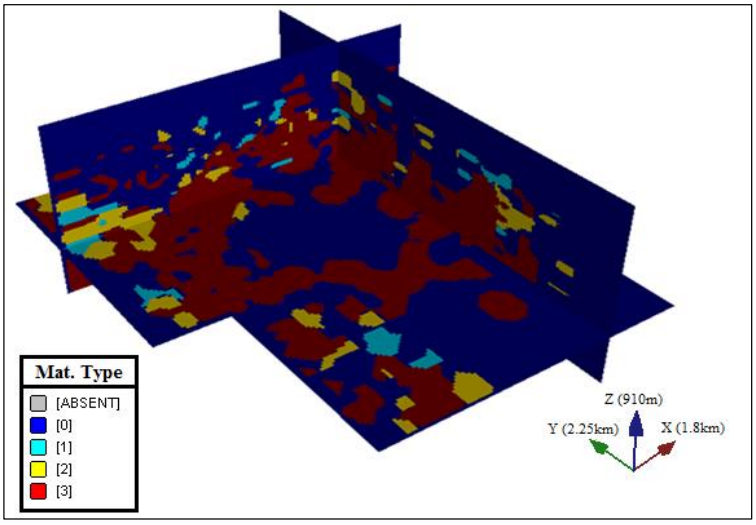

Figure 2: Cross-sections of the training image ($X = 96$, $Y = 118$ and $Z = 51$).

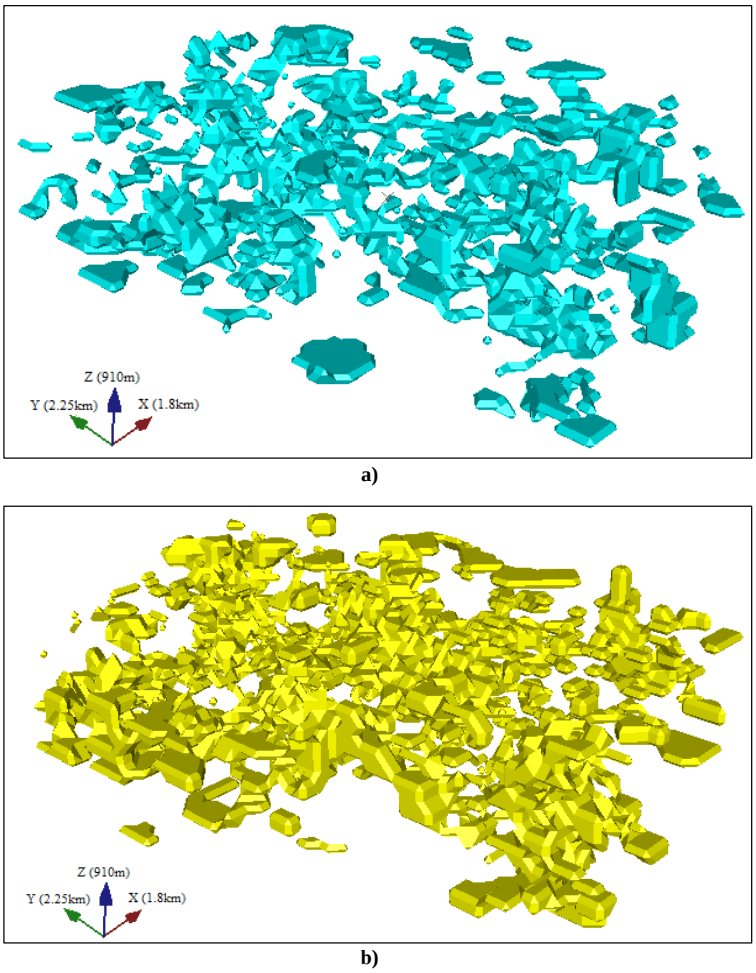

Figure 3: Spatial visualization of wireframes representing domains 1 and 1, in the training image: a) domain 1; and b) domain 2.

3.2.2 Simulation results

The number of cluster and template size are defined through trial and error. It is important to note that, the larger the template and the number of cluster, the better the results tend to be, however the run time of the algorithm becomes longer. Hence, the values to be used for these parameters are the ones showing the best trade-off between quality of results and computational time. Table 2 shows the parameters that were chosen through testing to simulate Olympic Dam deposit's material types.

Table 2: Parameters used in simulation.

Parameter	Value
Template	$11 \times 11 \times 5$
Inner Patch	$5 \times 5 \times 3$
Number of Clusters	500
Number of Realizations	20

Template, inner patch and number of clusters are defined in Sections 2.1, 2.4 and 2.3, respectively. Figure 4 displays 3 sections of two simulated realization. These sections are the same ones shown for the training image in Figure 2.

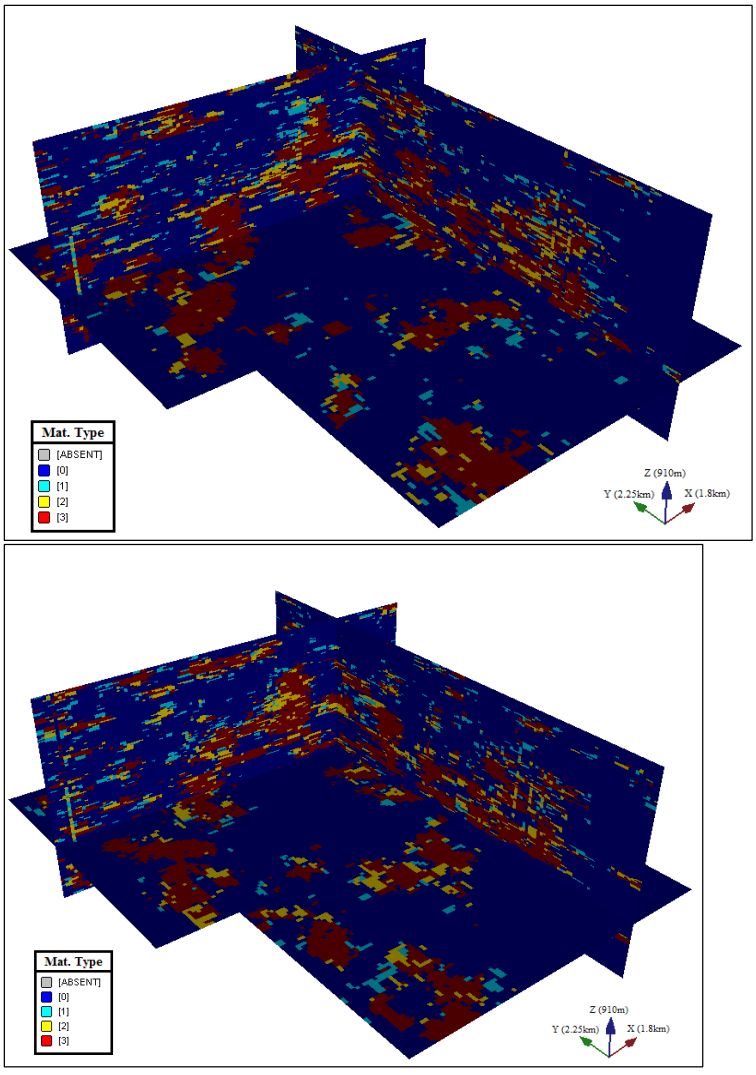

Figure 4: Two simulations of material types. The sections are the same than in Figure 2.

Comparison between training image and simulations section shows that both present the same regional configuration: same categories tend to appear in the same regions. However, as expected, the simulated realizations are smooth, presenting more variable patterns. Figure 5 shows the histogram of the 3 mineralized categories in the 20 simulations, compared to training image and declustered samples.

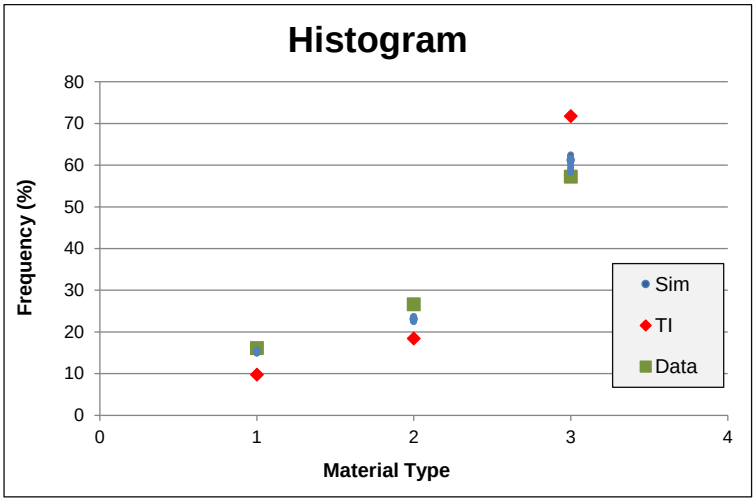

Figure 5: Histogram of 20 material types' simulations compared to data and training image.

According to Figure 5, simulations reproduce well the proportions of the 3 material types. It also shows the effect of the training image on the simulations' histograms. In this case study, the simulations reproduced data's histogram as opposed to training image's one due to the large amount of samples available. Figures 6, 7 and 8 show the direct variogram for the 3 material types, in 3 directions: East-West (EW), North-South

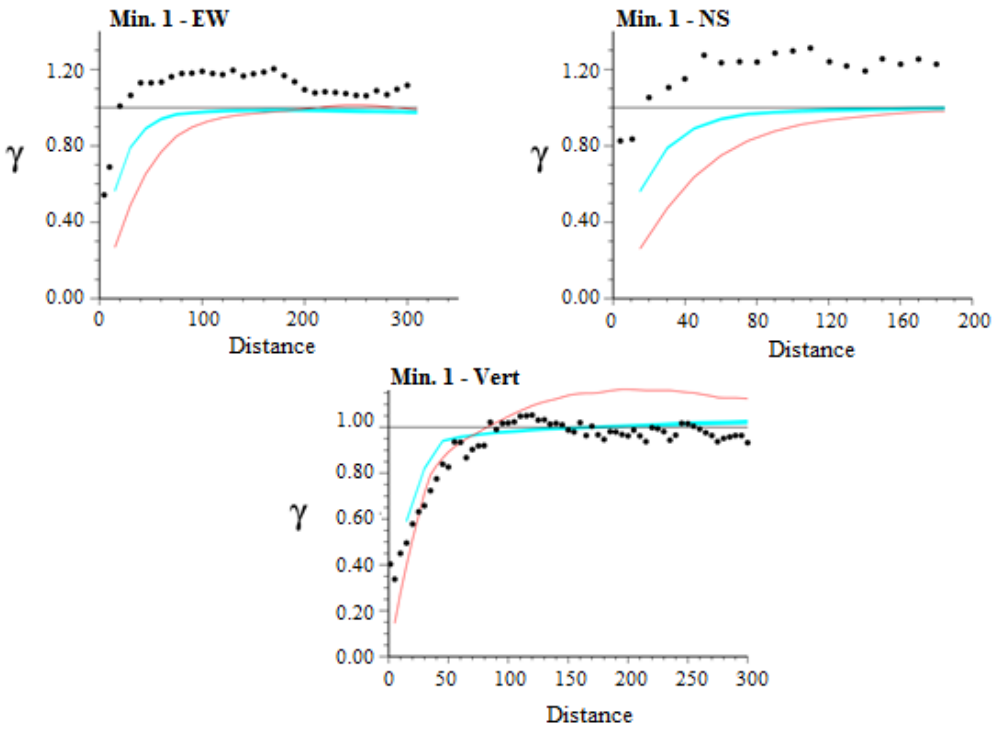

Figure 6: Variograms of 20 simulations of material types (light blue line), samples (black dot) and training image (red line), for material type 1.

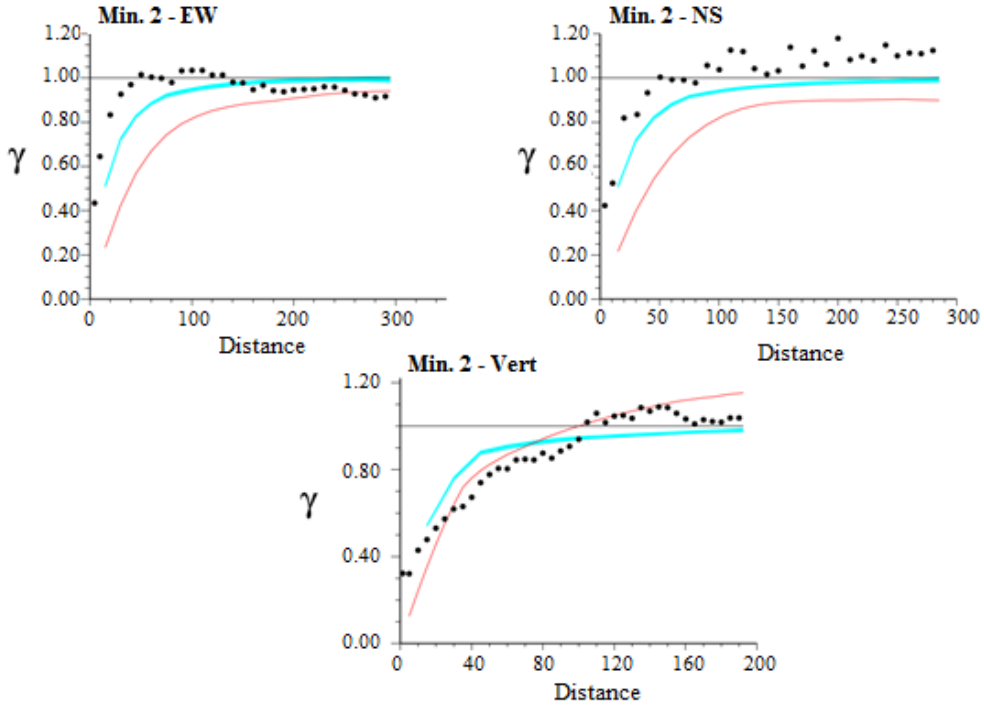

Figure 7: Variograms of 20 simulations of material types (light blue line), samples (black dot) and training image (red line), for material type 2.

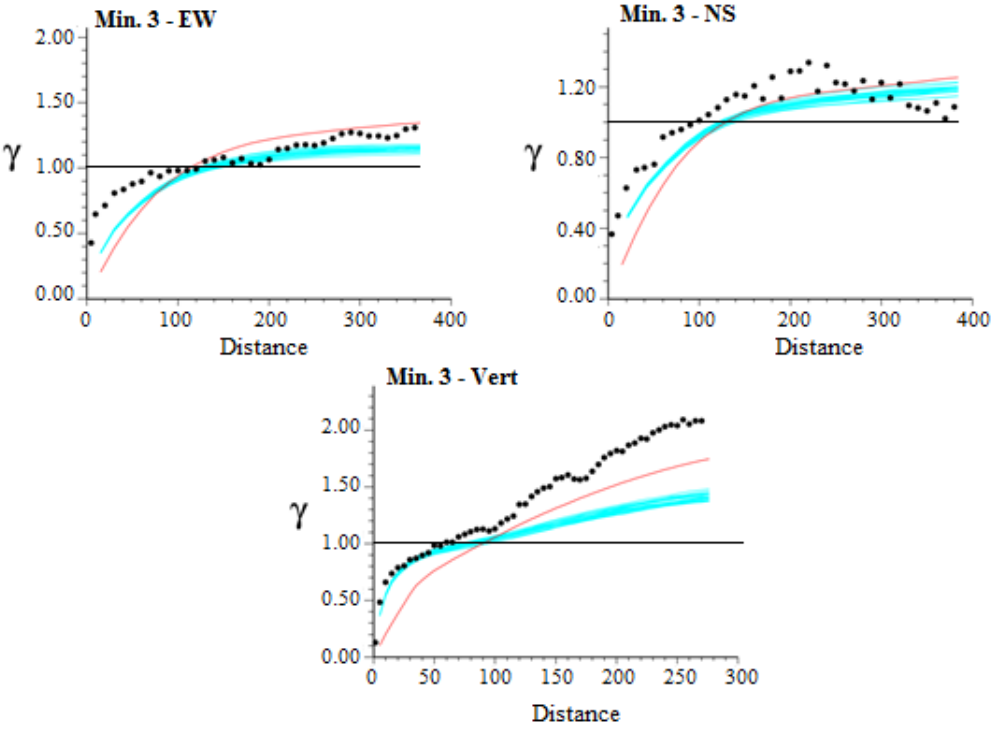

Figure 8: Variograms of 20 simulations of material types (light blue line), samples (black dot) and training image (red line), for material type 3.

(NS) and vertical. The cross-correlograms between these 3 material types for EW, NS and vertical directions are displayed in Figures 9, 10 and 11.

All results suggest a reasonable reproduction of training image's and samples' variograms and cross-variograms by the simulated realizations of the 3 material types at Olympic Dam. As the simulation was performed with a training image driven method, it was expected that simulations' variograms were closer to training image's ones. However, as there are a large amount of samples in this case study, they are more similar to data's variograms. Validations are also performed in terms of high-order statistics, which are analyzed through cumulant maps, as shown in De Iaco and Maggio, 2011. In order for the cumulant maps to be calculated, spatial templates have to be defined. In this case, an L-shape template was used to calculate 3th order cumulant {(1,0,0); (0,1,0)} and for the 4th one, the template {(1,0,0); (0, 1, 0); (0,0,1)} was used. Figures 12 and 13 show cumulant maps calculated considering material types 1, 2 and 3. As it may be noted, the simulations' cumulant maps can be seen as being "in between" samples' and training image's ones: they show a similar general pattern to latter, but with lesser continuity, due to influence of the former.

3.3 Simulation of copper grades

3.3.1 Data set and training image

Having defined material types of Olympic Dam deposit through categorical simulation, copper grades are simulated within these boundaries, using the same wavelet based method. Figure 14 and Table 3 show, respectively, the histograms and statistics of copper grade for each material type, separately. Material types 1 and 2 are the richest ones; type 3 also contains high copper grade samples; and finally type 0 is mostly waste and, as result, it is not simulated.

The training image used in this case study is generated through low rank tensor completion (Liu et al., 2013), as described in Section 3.2. The algorithm works better for higher density of samples, hence the

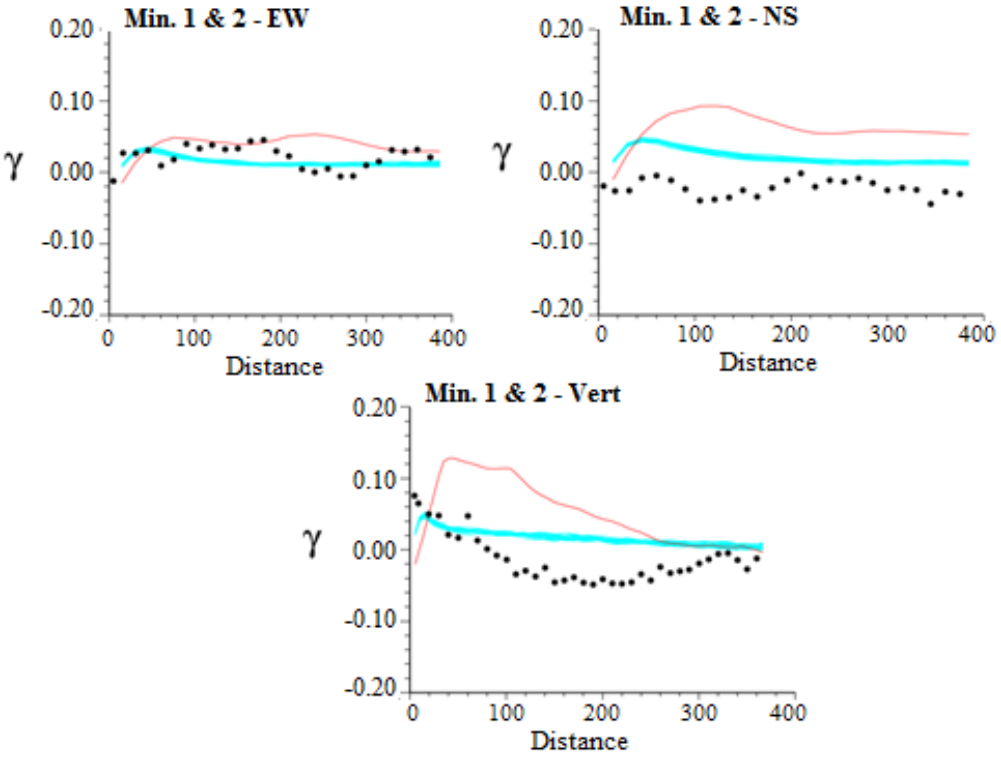

Figure 9: Cross-correlograms of 20 simulations of material types (light blue line), samples (black dot) and training image (red line), for material types 1 and 2.

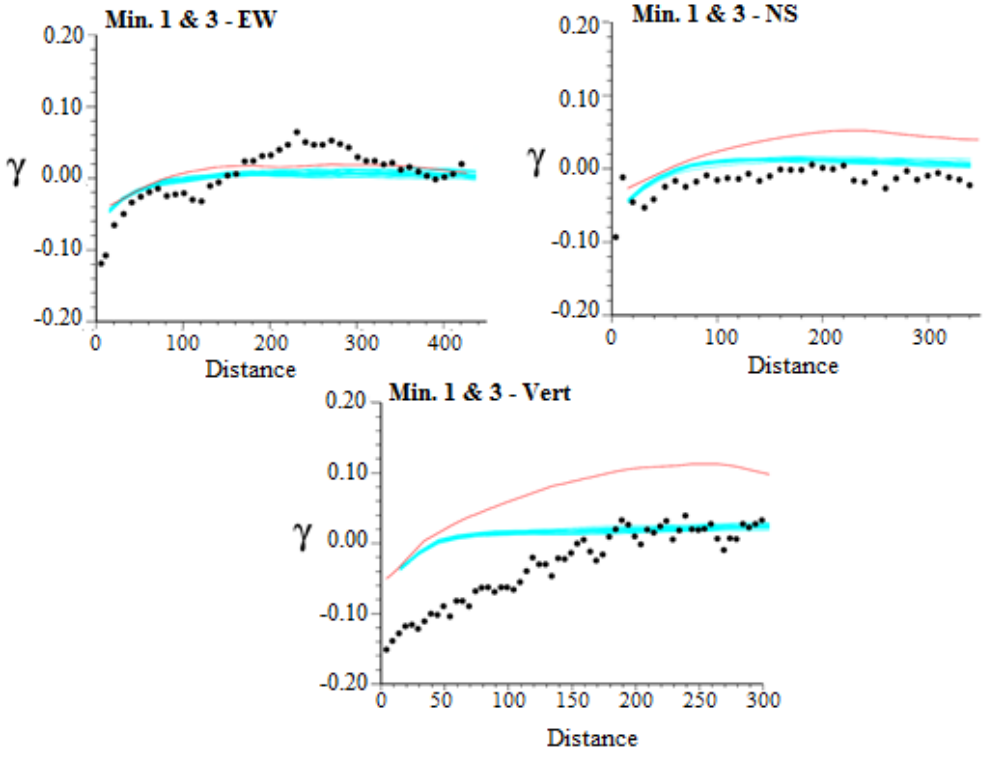

Figure 10: Cross-correlograms of 20 simulations of material types (light blue line), samples (black dot) and training image (red line), for material types 1 and 3.

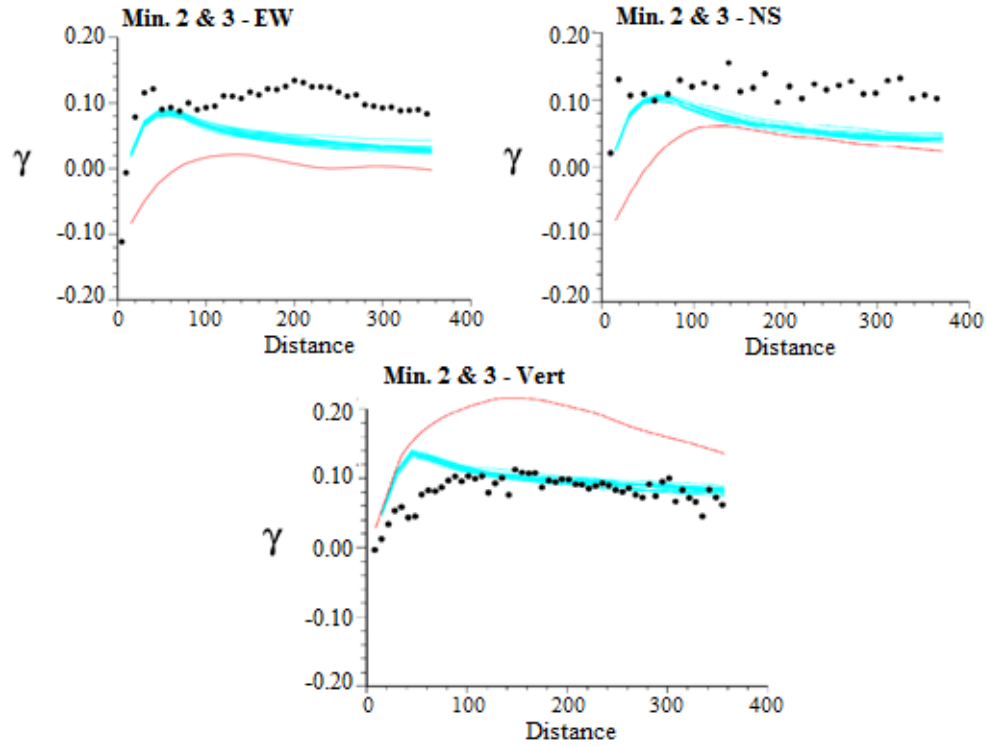

Figure 11: Cross-correlograms of 20 simulations of material types (light blue line), samples (black dot) and training image (red line), for material types 2 and 3.

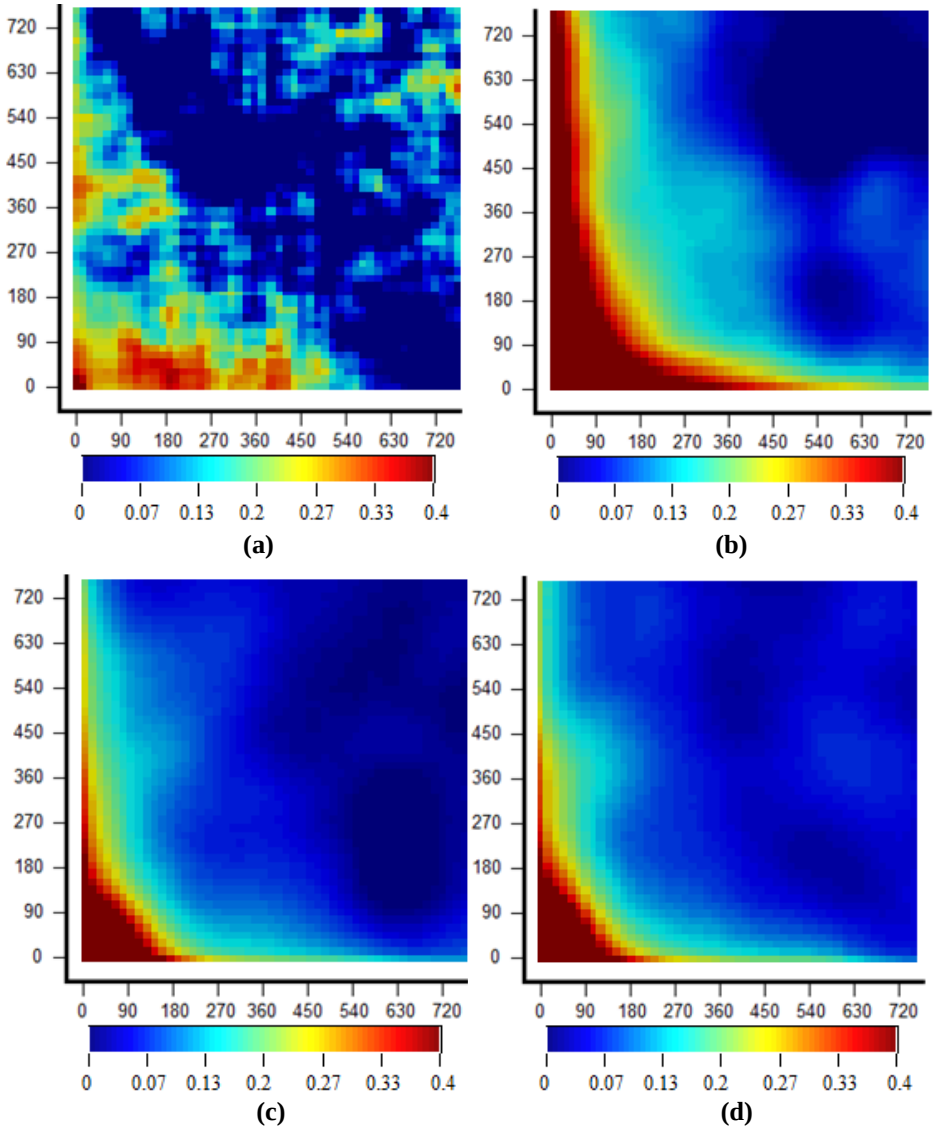

Figure 12: Third-order cumulant maps for a) samples; b) training image; c) and d) two realizations. Direction of cumulant: {(1,0,0); (0,1,0)}.

Table 3: Statistics of copper grade per domain.

Statistics	Cat0	Cat1	Cat2	Cat3
Mean	0.104	0.733	0.741	0.453
Stand. Dev.	0.372	1.137	1.012	0.614
Variance	0.139	1.292	1.023	0.377
Kurtosis	67.301	7.318	5.854	5.250
Skewness	7.655	2.567	2.249	2.139
Minimum	0.005	0.005	0.005	0.005
Maximum	4.58	7.66	6.866	4.08
10th Perc.	0.007	0.020	0.024	0.010
25th Perc.	0.014	0.109	0.123	0.071
50th Perc.	0.030	0.200	0.256	0.180
75th Perc.	0.059	0.858	0.993	0.621
90th Perc.	0.114	2.184	2.125	1.259

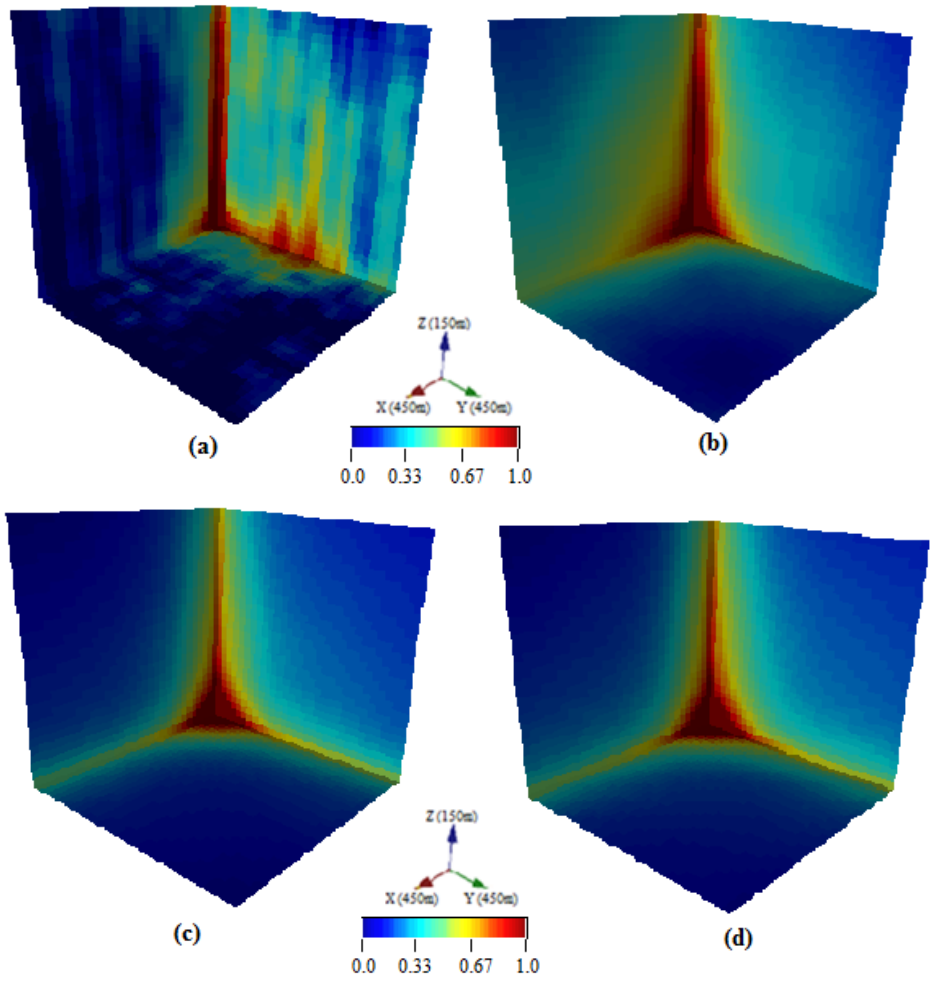

Figure 13: Fourth-order cumulant maps for a) samples; b) training image; c) and d) two realizations. Direction of cumulant: {(1,0,0); (0, 1, 0); (0,0,1)}.

training image was generated based on the densest sampled part of the deposit, as shown in Figure 15. Figure 16 displays 3 cross-sections of the training image. Its dimensions in X, Y and Z direction are 80, 80 and 111, respectively.

The continuous simulation is performed as follows. To simulate material type 1, for example, only patterns which lie inside the wireframe related to this material, as defined in Figure 3(a), are taken from the training image. Then, these patterns are pasted on the simulation grid, but only on the nodes which lie inside the simulated wireframe related to this material type, according to the simulation of material types in the previous section. The same procedure is then repeated to simulate copper grades for categories 2 and 3.

3.3.2 Simulation results

The parameters used to simulate the copper grades within each material type are displayed in Table 4. As in the simulation of material types, the simulation grid is discretized by 15 m $\times$ 15 m $\times$ 5 m, resulting in 120 $\times$ 150 $\times$ 170 nodes in X, Y and Z directions respectively, and a total of 2,813,670 nodes.

The template size and the number of clusters to be used are defined based on trial and error. Several values are tested, their results are analyzed and the values of these parameters which present the best trade-off between good quality results and computational time are chosen. The tendency is that, the bigger the template and number of clusters, the better the results are and the longer it takes to run. Figures 17, 18

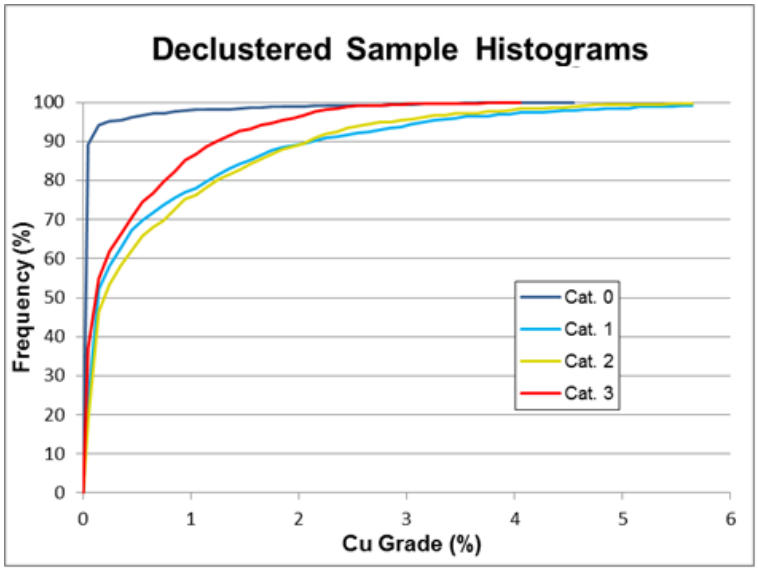

Figure 14: Declustered copper grade histograms for each geological domain.

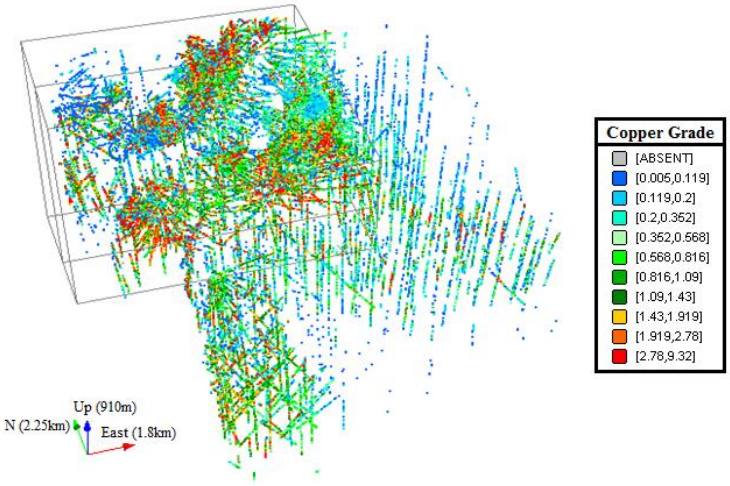

Figure 15: Data set colored regarding copper grade. Samples used to build training image is highlighted.

Table 4: Parameters used for the copper simulations.

Parameter	Value
Template	$11 \times 11 \times 9$
Inner Patch	$5 \times 5 \times 5$
Number of Clusters	500
Numnber of Realizations	3

and 19 show cross-sections in X, Y and Z directions respectively for 2 simulations, at point support. The 3 material types are being displayed together.

The gray color corresponds to the category 0, which was not simulated. It is possible to see that, despite of presenting different shapes, due to be related to different categorical simulations, the copper simulations above tend to present the same spatial pattern; i.e. coincident rich and poor areas. Figures 20, 21 and 22 display the histograms of the 3 simulated realizations compared to training image and hard data, for material types 1, 2 and 3, respectively.

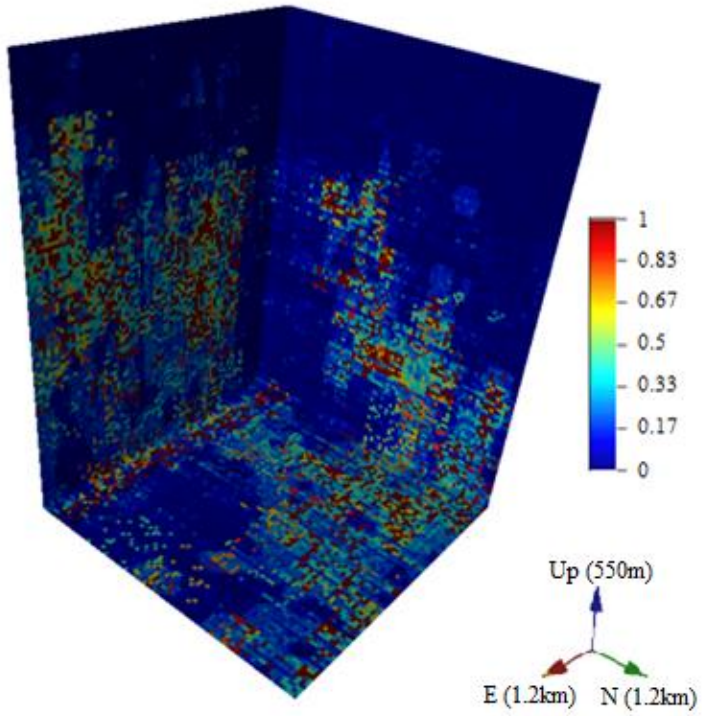

Figure 16: Cross-sections of the continuous training image ($X = 0$, $Y = 0$ and $Z = 0$).

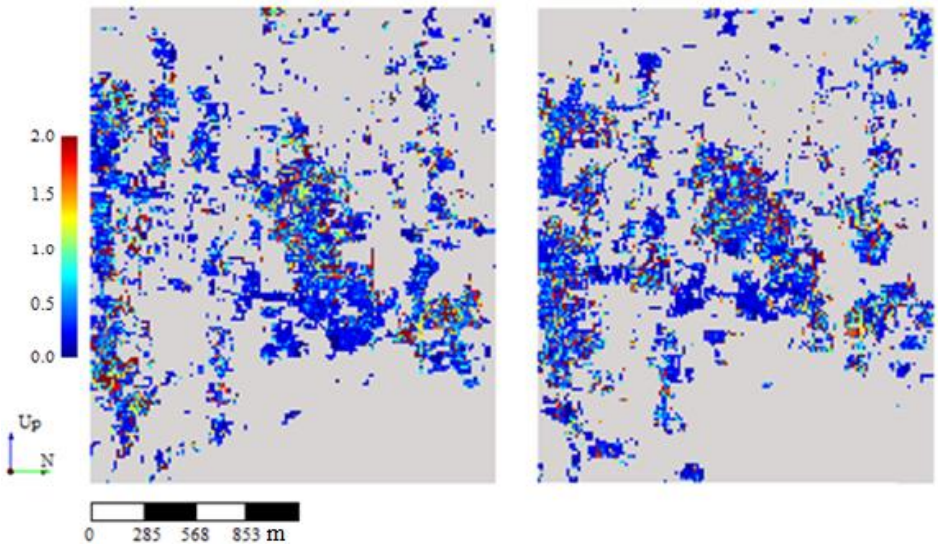

Figure 17: Cross-section ($X = 40$) showing realizations of copper grades (%), for 2 simulations.

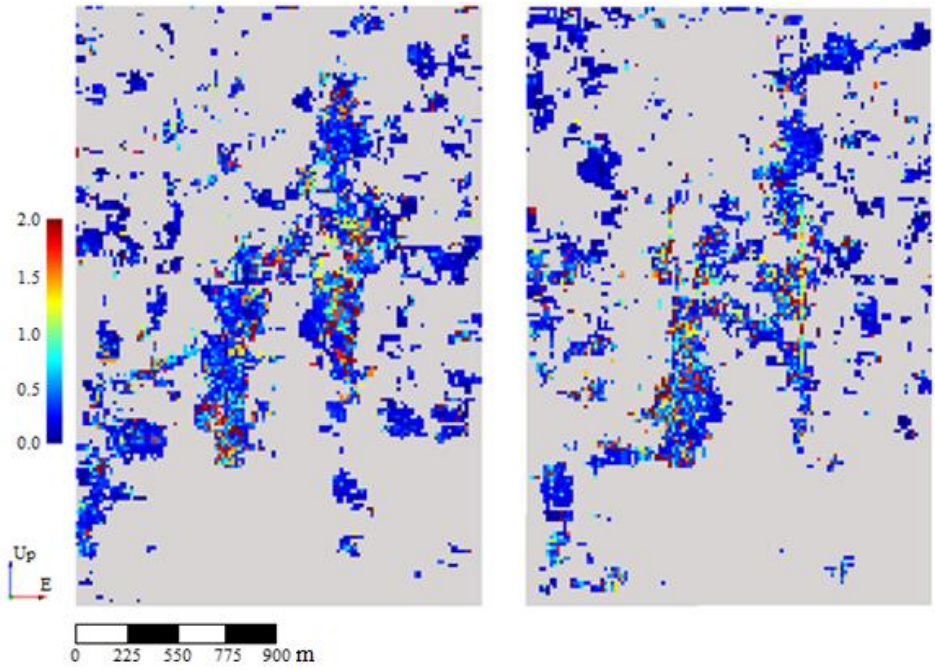

Figure 18: Cross-section ($Y = 120$) showing realizations of copper grades (%), for 2 simulations.

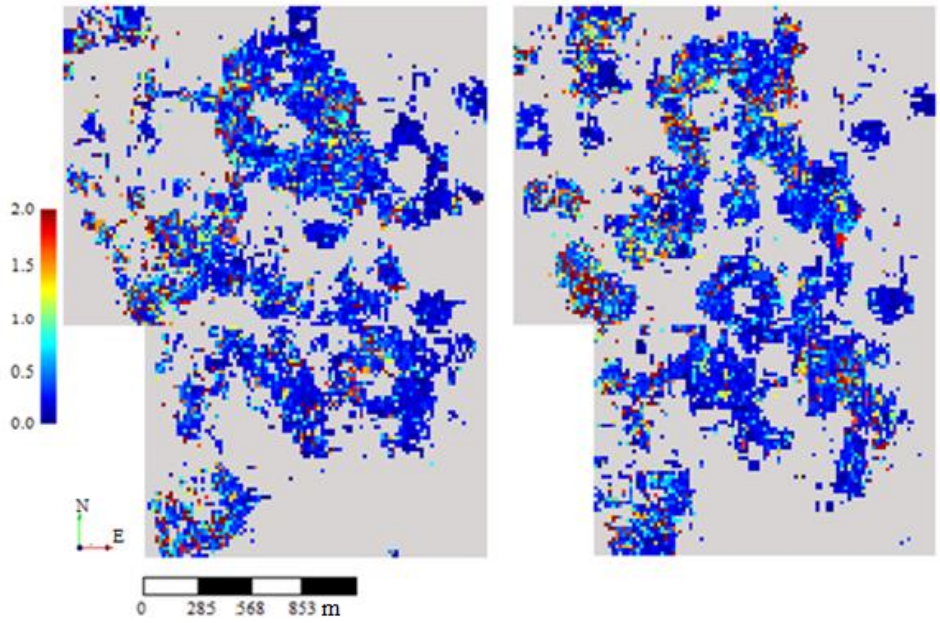

Figure 19: Horizontal view ($Z = 80$) showing realizations of copper grades (%), for 2 simulations.

Figures 20, 21 and 22 show that the simulated copper realizations present very similar histograms, compared to both samples and training image. It is noteworthy that the histograms of samples and training image are very similar: it is important to multiple-point simulation techniques, including the method used herein, that the training image presents similar statistics to the hard data. Otherwise, the simulated realizations may show conflicting features. Figures 23, 24 and 25 show the copper grades variograms in the east/west, north/south and vertical directions for material types 1 and 2 and 3, respectively.

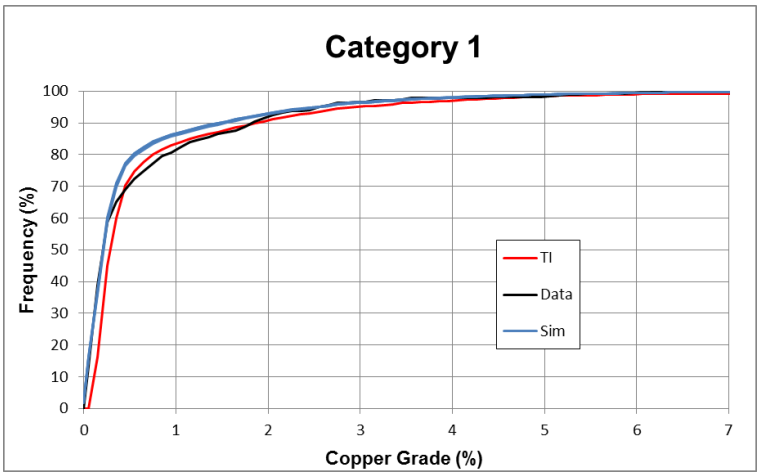

Figure 20: Copper histogram of material type 1. Comparison between training image, hard data and simulations.

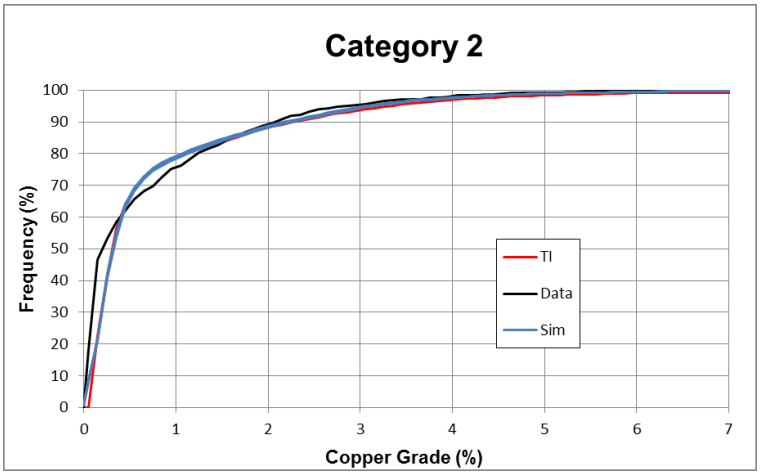

Figure 21: Copper histogram of material type 2. Comparison between training image, hard data and simulations.

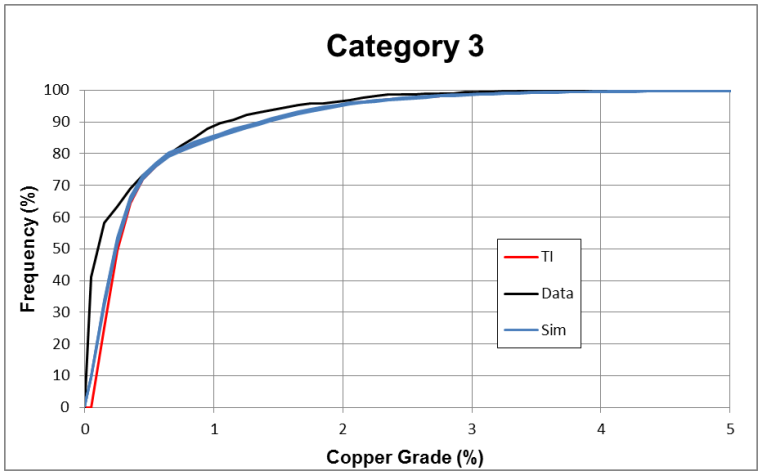

Figure 22: Copper histogram of material type 3. Comparison between training image, hard data and simulations.

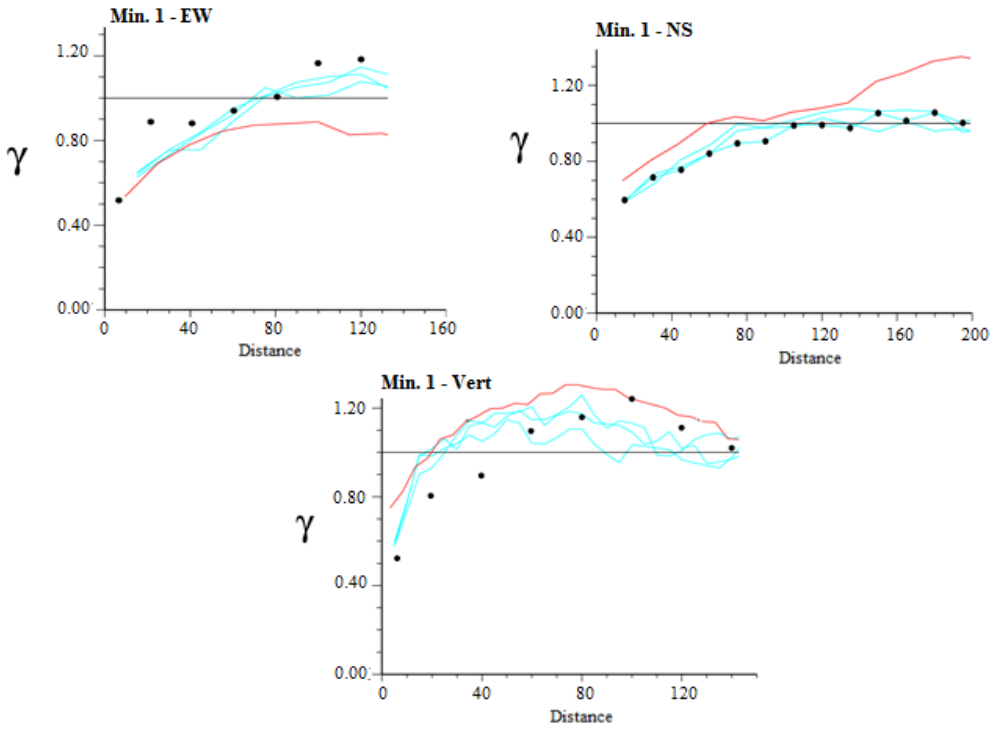

Figure 23: Copper grade variograms of 3 simulations (light blue line), samples (black dot) and training image (red line) for material type 1.

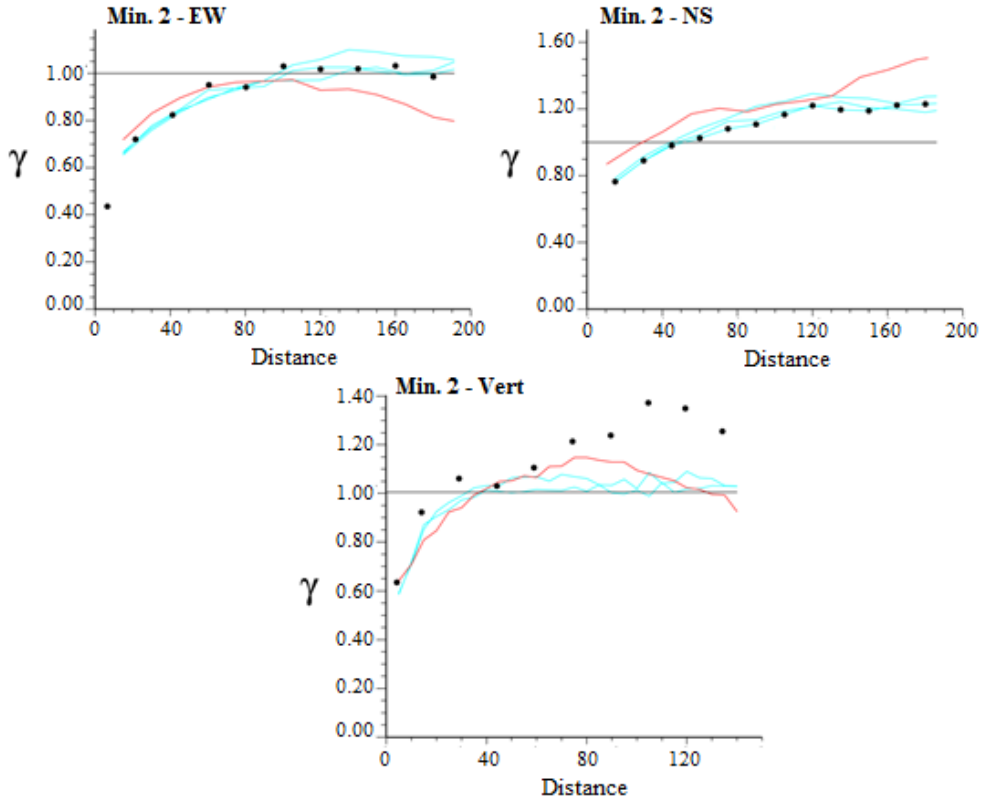

Figure 24: Copper grade variograms of 3 simulations (light blue line), samples (black dot) and training image (red line) for material type 2.

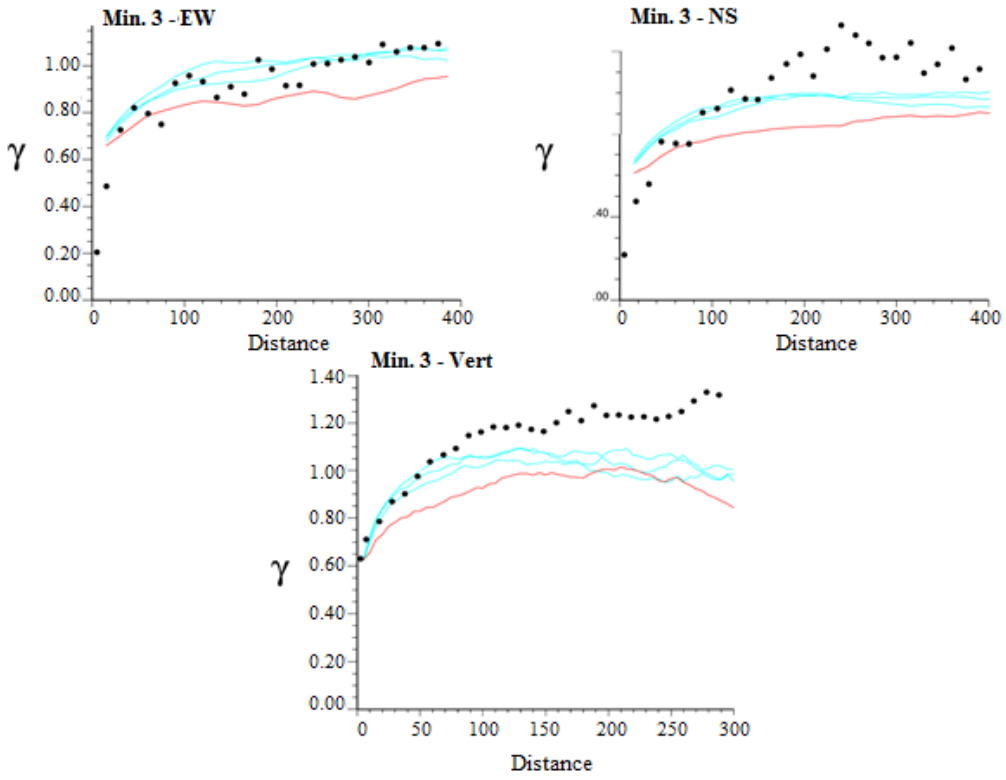

Figure 25: Copper grade variograms of 3 simulations (light blue line), samples (black dot) and training image (red line) for material type 3.

Simulated realizations show good reproduction of samples' and training image's variograms. Here, it is possible to note, again, the tendency of realizations to present their variograms in between the ones of samples and training image. Besides, the variograms of training image and samples are very similar to each other. Figures 26 and 27 show, cumulant maps of copper grade regarding each material type for samples, training image and two simulated realizations.

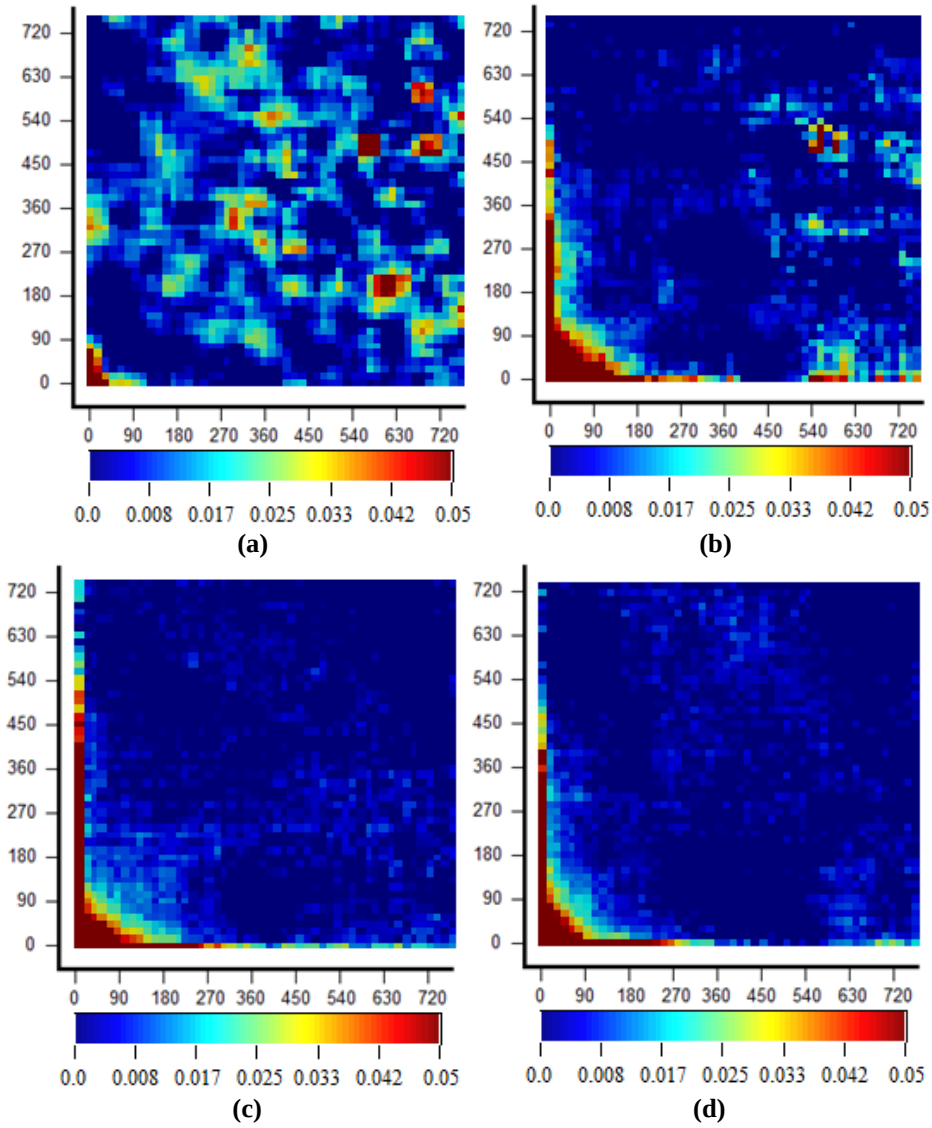

Figure 26: Copper grade third-order cumulant maps of a) training image; b) and c) two realizations. Direction of cumulant: $\{(1,0,0);(0,1,0)\}$. Distance in meters.

Figures 26 and 27 show how similar realizations' and training image's maps are, but with less continuity. This is due to the influence of more discontinuous samples' maps over their cumulant maps. When comparing samples' cumulant maps to simulation's and training image's ones, it is important to recall that the samples contains much less replicates for each lag. This may cause it to have some local artifacts, such as the discontinuities seen in the Figures 26 and 27, which probably would not happen if more data were available.

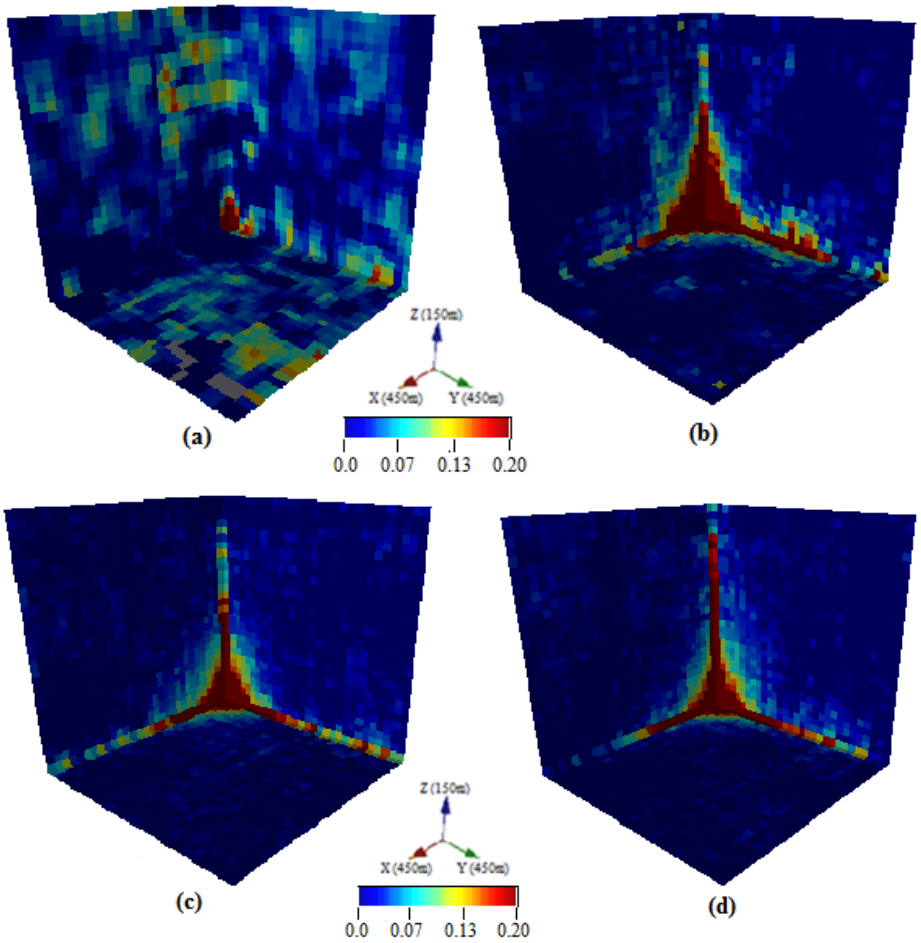

Figure 27: Copper grade fourth-order cumulant maps of a) training image; b) and c) two realizations. Direction of cumulant: $\{(1,0,0);(0,1,0);(0,0,1)\}$.

4 Conclusions

In this work, a multiple-point simulation method based on wavelet analysis is applied to Olympic Dam copper deposit, located in southern Australia, in order to evaluate both volumetric and grades uncertainty. Olympic Dam presents 4 different material types and the training image used to simulate the orebody model is obtained through geological interpretation. Then, 3 of those categorical realizations are retained for the simulation of copper grades, which is performed considering one material type at a time. The continuous training image was generated through low rank tensor completion. The result is a set of equiprobable orebody models which accounts for volumetric and grade uncertainty. The validation of the simulations' results is performed for low order statistics (histograms and variograms) and also for high-order statistics, through third and fourth-order spatial cumulant maps. This validation suggested that method tested herein can be successfully applied to real sized deposits, since Olympic Dam was discretized in 2,813,670 nodes. In all the cases, the statistics of the realizations were compared to both training image's and samples' statistics. This comparison showed an interesting point, which is common to all multiple point simulation methods, since they are training image driven: the statistics values of the realization tend to be in between training image's and data's ones. Because of that, conflicting information between training image and data samples has a negative impact on the simulation's results. Therefore, the training image has to be representative of the deposit to be simulated.

References

Arpat, G., and Caers, J. (2007). Conditional simulation with patterns. Mathematical Geology, 39(2), 177–203.

Arpat, G.B. (2005). Sequential simulation with patterns. PhD thesis, Stanford University.

Bastrakov, E.N., Skirrow, R.G., Davidson, G.J. (2007). Fluid evolution and origins of iron oxide Cu-Au prospects in the Olympic Dam district, Gawler Craton, South Australia. Economic Geology, 102(8), 1415–1440.

Belperio, A., and Freeman, H. (2004). Common geological characteristics of Prominent Hill and Olympic Dam-Implications for iron oxide coppergold exploration models. Australian Institute of Mining Bulletin, Nov-Dec. Issue, 67–75.

Bénéteau, C., Van Fleet, P.J. (2011). Discrete Wavelet Transformation and Undergraduate Education. Notices of AMS, 58(5), 656–666.

Boucher, A. (2009). Considering complex training images with search tree partitioning. Computers and Geosciences, 35(6), 1151–1158.

Boucher, A., Costa, J.F., Rasera, L.G., Motta, E. (2014). Simulation of Geological Contacts from Interpreted Geological Model Using Multiple-Point Statistics. Mathematical Geosciences, 46, 561–572.

Caers, J. (2011), Modeling Uncertainty in the Earth Science, Willey-Blackwell, Chichester, UK.

Caers, J., and Zhang, T. (2004). Multiple-point geostatistics: A quantitative vehicle for integrating geologic analogs into multiple reservoir models. AAPG Special Volumes, 383–394.

Chatterjee, S., Mustapha, H., Dimitrakopoulos, R. (2011). Geologic heterogeneity representation using high-order spatial cumulants for subsurface flow and transport simulations. Water Resources Research, 47(8).

Chatterjee, S., Dimitrakopoulos, R, and Mustapha, H. (2012). Dimensional Reduction of Pattern-Based Simulation Using Wavelet Analysis. Mathematical Geosciences, 44, 343–374.

De Iaco, S., and Maggio, S. (2011). Validation techniques for geological patterns simulations based on variogram and multiple-point statistics. Mathematical Geosciences, 43(4), 483–500.

Dimitrakopoulos, R., Mustapha, H., and Gloaguen, E. (2010). High-order statistics of spatial random fields: Exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena. Mathematical Geosciences, 42(1), 65–99.

Ding, C., and He, X. (2004). K-means Clustering via Principal Component Analysis. Proc. of Int'l Conf. Machine Learning (ICML 2004), 225–232.

Gloaguen, E., Dimitrakopoulos, R. (2009). Two-dimensional conditional simulations based on the wavelet decomposition of training images. Mathematical Geosciences, 41(6), 679–701.

Goodfellow, R., Albor Consuegra, F., Dimitrakopoulos, R., Lloyd, T. (2012). Quantifying multi-element and volumetric uncertainty, Coleman McCreedy deposit, Ontario, Canada. Computers and Geosciences, 42, 71–78.

Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation (Applied Geostatistics Series). Oxford University Press, Oxford.

Guardiano, F., Srivastava, R. (1993). Multivariate geostatistics: beyond bivariate moments. In A. Soares (Ed.), Geostatistics Troia. Vol. 1, 133–144. Kluwer Academic Publications.

Hahn, T. (2008). The Olympic Dam Cu-U-Au-REE Deposit, Australia. Retrieved from: http://www.geo.tu-freiberg.de/oberseminar/os07_08/Thomas%20Hahn.pdf.

Hartigan, J.A., Wong, M.A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society, Series C (Applied Statistics) 28(1), 100–108.

Honarkhah, M., and Caers, J. (2010). Stochastic simulation of patterns using distance-based pattern modeling. Mathematical Geosciences, 42(5), 487–517.

Huang, T., Lu, D.T., Li, X., Wang, L. (2013). GPU-based SNESIM implementation for multiple-point statistical simulation. Computers and Geosciences, 54, 75–87.

InfoMine Inc. Retrieved from http://www.infomine.com/minesite/minesite.asp?site=olympicdam.

Jones, P., Douglas, I., Jewbali, A. (2013). Modeling Combined Geological and Grade Uncertainty: Application of Multiple-Point Simulation at the Apensu Gold Deposit, Ghana. Mathematical Geosciences, 45(8), 949–965.

Liu, J., Musialski, P., Wonka, P., and Ye, J. (2013). Tensor completion for estimating missing values in visual data. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(1), 208–220.

Lloyd, S.P. (1982). Least Square Quantization in PCM. IEE Transactions on Information Theory, 2(2), 129–137.

MacQueen, J.B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1, 281–297.

Mallat, S. (1998). A Wavelet Tour of Signal Processing. Academic Press, San Diego, CA.

Mariethoz, G., Renard, P., Straubhaar, J. (2010). The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resources Research, 46(11).

Mustapha, H., Chatterjee, S., Dimitrakopoulos, R. (2014). CDFSIM: Efficient Stochastic Simulation Through Decomposition of Cumulative Distribution Functions of Transformed Spatial Patterns. Mathematical Geosciences, 46(1), 95–123.

Mustapha, H., Dimitrakopoulos, R. (2010). High-order stochastic simulations for complex non-Gaussian and non-linear geological patterns. Mathematical Geosciences, 42(5).

Osterholt, V., Dimitrakopoulos, R. (2007). Simulation of wireframes and geometric features with multiple-point techniques: Application at Yandi iron ore deposit. Orebody modelling and strategic mine planning, AusIMM, Spectrum Series, 14, 95–124.

Roberts, D.E., Hudson, G.R.T. (1983). The Olympic Dam copper-uranium-gold deposit, Roxby Downs, South Australia. Economic Geology, 78(5), 799–822.

Skirrow, R.G., Bastrakov, E.N., Barovich, K., Fraser, G.L., Creaser, R.A., Fanning, C.M., Raymond, O.L., Davidson, G.J. (2007). Timing of iron oxide Cu-Au-(U) hydrothermal activity and Nd isotope constraints on metal sources in the Gawler Craton, South Australia. Economic Geology, 102(8), 1441–1470.

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. Mathematical Geosciences, 43(3), 305–328.

Strebelle, S. (2000). Sequential Simulation Drawing Structures from Training Images. PhD thesis, Stanford University.

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple point statistics. Mathematical Geology, 34(1), 1–21.

Strebelle, S., Cavelius, C. (2014). Solving Speed and Memory Issues in Multiple-Point Statistics Simulation Program SNESIM. Mathematical Geosciences, 46(2), 171–186.

Strebelle, S., and Zhang, T. (2005) Non-stationary multiple-point geostatistical models. Geostatistics Banff 2004, 235–244. Springer Netherlands.

Wu, J., Zhang, T., and Journel, A. (2008). Fast FILTERSIM simulation with score-based distance. Mathematical Geosciences, 40(7), 773–788.

Yahya, W.J. (2011). Image reconstruction from a limited number of samples: A matrix-completion-based approach. McGill University Libraries, Montreal, Canada.

Zhang, T., Switzer, P., and Journel, A. (2006). Filter-based classification of training image patterns for spatial simulation. Mathematical Geology, 38(1), 63–80.