Les Cahiers du GERAD ISSN: 0711–2440

High-Order Simulation with Clustered Datasets

D.F. Machuca-Mory
R. Dimitrakopoulos

G–2013–57

September 2013

High-Order Simulation with Clustered Datasets

David F. Machuca-Mory

Roussos Dimitrakopoulos*

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

david.machucamory@mcgill.ca

roussos.dimitrakopoulos@mcgill.ca

* *and GERAD*

September 2013

Les Cahiers du GERAD

G–2013–57

Abstract: Preferential sampling of high grade zones is a common practice in the drilling campaigns of ore deposits. This may lead to a bias in the statistics derived from these clustered datasets. The global mean and variance, particularly, tend to be overestimated and high-order statistics, such as the skewness and kurtosis are distorted. High-order simulation uses high-order spatial statistics, known as cumulants, to approximate non-Gaussian local distributions of grades conditioned by neighbouring samples. Data clustering affects the inference of the spatial cumulants, and consequently, the conditional distributions fitted using them. A weighted cumulant estimator is proposed to account for data clustering. This estimator is implemented in the program HOSC, used for calculating the experimental spatial cumulants of irregularly spaced datasets. Additionally, the fitting of non-Gaussian conditional distributions using Legendre polynomial series derived from weighted spatial cumulants is proposed and implemented in the algorithm for simulation with high-order statistics, HOSIM. The geological model and diamond drilling dataset corresponding to the Apensu gold deposit in Ghana is used to illustrate these implementations and to compare the new results with the realizations and statistics produced without accounting for data clustering. This example shows that as expected, the data cdf and its statistics are reproduced by simulation with declustered high-order statistics.

Key Words: Spatial cumulants, non-Gaussian distributions, training image.

1 Introduction

Spatial clustering of direct measurements, or hard data, is a common feature in mining and other geosciences or geoengineering datasets, which is a result of technical and logistical constraints, and also from the interest on obtaining more information in areas where the sampled rock attribute exhibits anomalous values. Sample clustering may cause biases in the inference of the attribute histogram and other statistics, thus, it is common practice to weight the samples inversely proportional to the local sample density to minimize the bias introduced by preferential or clustered sampling. When the declustered histogram is presented to geostatistical simulation methods, such as the sequential Gaussian simulation, the resulting realizations reproduce this histogram and the one and two point statistics within ergodic fluctuations (Goovaerts, 1999; Leuangthong et al., 2004).

Recently, simulation based on high-order statistics has been proposed to overcome the limitations imposed by the use of two-point spatial statistics and the multi-Gaussian model (Mustapha and Dimitrakopoulos, 2010). High-order spatial statistics, in particular cumulants, are able to capture more complex spatial relationships than two-point statistics such as the variogram and the covariance. High-order statistics are also used to fit non-Gaussian conditional distributions via polynomial series without any assumption on the data distribution. Following a random path over a grid, the fitted conditional probability distribution functions (*cpdf*s) are randomly sampled to produce realizations of a non-Gaussian random field. The low- and high-order statistics required to fit the *cpdf*s are primarily inferred directly from samples and completed by the information obtained from a training image. As higher orders of the statistics are considered, there is an increasing reliance on the training image. In order to avoid transferring the bias caused from clustered data to the fitted *cpdf*s, it is necessary to minimize this bias in the lower and high-order spatial statistics.

In this paper weighted high-order statistics are proposed to account for sample clustering in the fitting of non-Gaussian conditional distributions. First, weighted high-order spatial moments and cumulants are presented. Next, the fitting of a conditional distribution using Legendre polynomial series based on weighted high-order spatial moments and cumulants is developed. The impact of incorporating declustering weights in high-order simulation is illustrated with the help of a diamond drill-hole dataset from the Apensu gold deposit in Ghana. The declustered high-order statistics maps show the impact of incorporating declustering weights and the resulting realizations reproduce the declustered global histogram.

2 Spatial Simulation Based on Weighted High-Order Spatial Statistics

High-order spatial statistics correspond to statistics of n-tuples of samples arranged in a spatial template. Given a spatial random function $Z(\boldsymbol{u})$ defined over a domain D, the vectors $\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n$ define a template $T_{n+1}^{\boldsymbol{h}_1,\ldots,\boldsymbol{h}_n}$ of $n+1$ points $\boldsymbol{u}_0,\ \boldsymbol{u}_0+\boldsymbol{h}_1, \ldots,\ \boldsymbol{u}_0+\boldsymbol{h}_n \in D$. A high-order spatial moment for this arrangement is defined by the integration:

$$E\left[Z_0^{i_0} \cdot Z_1^{i_1} \cdots Z_n^{i_n}\right] = \int_{D^{n+1}} z_o^{i_0} z_1^{i_1} \cdots z_n^{i_n} f_Z(z_0 z_1 \cdots z_n)\, d\boldsymbol{z} \tag{1}$$

with $\boldsymbol{z} = (z_0, z_1, \cdots, z_n)$. For the sake of brevity, the random variables in the template are denoted as $Z(\boldsymbol{u}_o) = Z_0$, $Z(\boldsymbol{u}_o+\boldsymbol{h}_1) = Z_1, \cdots, Z(\boldsymbol{u}_o+\boldsymbol{h}_n) = Z_n$ and their corresponding realizations as $z_0, z_1, \cdots, z_n$. Normally, when inferring spatial high-order moments from scattered data, all samples are considered as equally weighted. The inference of weighted experimental high-order statistics and their use in high-order spatial simulation is presented next.

2.1 Weighted spatial cumulants

Declustering weights can be obtained from several declustering methods, such as cell declustering (Deutsch, 1989), polygonal declustering (Isaaks and Srivastava, 1989), global kriging or random pick. A common feature of these methods is that they assign to each sample at a location $\boldsymbol{u}_0$ a weight $w_0 = w(\boldsymbol{u}_0)$ that is inversely proportional to the sample density in its neighbourhood. For an arrangement of $n+1$ sample locations

$\boldsymbol{u}_0, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$, a $n+1$-point weight is defined as the geometric average of the individual declustering weights:

$$w_{0,1,\ldots,n} = w(\boldsymbol{u}_0, \boldsymbol{u}_1, \cdots, \boldsymbol{u}_n) = \sqrt[n+1]{w_0 w_1 \cdots w_n}$$

For a template $T_{n+1}^{\boldsymbol{h}_1,\ldots,\boldsymbol{h}_n}$ that has $N_{h_1 \cdots h_n}$ replicates within the domain, the corresponding declustered experimental high-order moment can be obtained by:

$$\hat{E}^{(w)}\left[Z_0^{i_0} Z_1^{i_1} \cdots Z_n^{i_n}\right] = \frac{1}{\sum_k^{N_{h_1 \cdots h_n}} w_{\boldsymbol{u}_k,1,\ldots,n}} \sum_{j=1}^{N_{h_1 \cdots h_n}} w_{\boldsymbol{u}_j,1,\ldots,n} z_j^{i_0} z_1^{i_1} \cdots z_n^{i_n} \tag{2}$$

As combinations of high-order spatial moments, spatial cumulants capture more information than single moments. For a zero-mean random function, the spatial cumulants from order one to four are given by (Dimitrakopoulos et al., 2010):

$$\begin{aligned}
C_1^{(w)}(Z_0) &= E^{(w)}[Z_0] = 0 \\
C_2^{(w)}(Z_0, Z_1) &= C_2^{(w)}(\boldsymbol{h}_1) = E^{(w)}[Z_0 Z_1] \\
C_3^{(w)}(Z_0, Z_1, Z_2) &= C_3^{(w)}(\boldsymbol{h}_1, \boldsymbol{h}_2) = E^{(w)}[Z_0 Z_1 Z_2] \\
C_4^{(w)}(Z_0, Z_1, Z_2, Z_3) &= C_4^{(w)}(\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3) \\
&= E^{(w)}[Z_0 Z_1 Z_2 Z_3] - E^{(w)}[Z_0 Z_1] E[Z_2 Z_3] \\
&\quad - E^{(w)}[Z_0 Z_2] E^{(w)}[Z_1 Z_3] \\
&\quad - E^{(w)}[Z_0 Z_3] E^{(w)}[Z_1 Z_2]
\end{aligned} \tag{3}$$

If the spatial high-order moments incorporate declustering weights, the spatial cumulants obtained from them are also declustered. It is important to notice the relation between cumulants and classical descriptive statistics. The second declustered cumulant for a zero lag is the declustered variance:

$$\sigma_{(w)}^2 = C_2^{(w)}(0) = E^{(w)}\left[Z_0^2\right] - \left(E^{(w)}[Z_0]\right)^2 \tag{4}$$

The declustered skewness and the declustered 3$^{\text{rd}}$-order cumulant for lags zero are related by:

$$\gamma_1^{(w)} = \frac{C_3^{(w)}(0,0)}{\left(C_2^{(w)}(0)\right)^{3/2}} \tag{5}$$

And the declustered kurtosis is related to 4$^{\text{th}}$-order by:

$$\beta_2^{(w)} = \frac{C_4^{(w)}(0,0,0)}{\left(C_2^{(w)}(0)\right)^2} \tag{6}$$

The unbiased inference of these statistics is important for the approximation of the univariate and multivariate *cpdf* by Legendre polynomial series.

2.2 Simulation with declustered spatial high-order statistics

Conditional sequential simulation requires the conditioning of local distributions by neighbouring sample values and other sources of information about the attribute being modelled. These conditional distributions are randomly sampled to generate realizations that are used for the conditioning of the local distribution where no realization yet exists. In the case that the initial conditioning dataset, $\Delta_0 = \{d_n\}$, is limited to hard data values at n locations, such as $d_n = \{z(\boldsymbol{u}_\alpha) = z_\alpha,\ \alpha = 1, \ldots, n\}$, the conditional *pdf* at the first simulated location $\boldsymbol{u}_0$ is given by:

$$f_{Z_0}(z_o|\Delta_0) = \frac{f_Z(z_0, z_1, \ldots, z_n)}{\int_D f_Z(z_0, z_1, \ldots, z_n)\, dz_0} \tag{7}$$

The realization z'_0 is obtained from the Monte Carlo drawing of this conditional distribution and added to the dataset. Thus, the new conditioning data consists of $\Delta_1 = \{d_n\} \cup \{z'_0\}$ and it is used to condition the *pdf* at a second node. This process continues by following a random path until all nodes of the predefined grid are simulated (Gómez-Hernández, 1993; Journel, 1994).

The objective is then to infer the multivariate distribution in the numerator of Eq. (7), whereas the denominator can be obtained by numerical integration of the multivariate distribution. In sequential Gaussian simulation, for instance, the multivariate *pdf* is modelled by a multi-Gaussian distribution, which is fully defined by a covariance matrix and a vector of local means. The local Gaussian *cpdf*s are efficiently inferred by simple kriging (Chilès and Delfiner, 1999), however this comes at the price of maximum entropy and lack of high-order structure and connectivity (Journel and Deutsch, 1993). Alternatively, high-order simulation uses high-order statistics and Legendre polynomials to approximate non-Gaussian multivariate distributions. In case of an unsampled location $\boldsymbol{u}_0$ and two conditioning data, z_1 and z_2, the corresponding declustered multivariate *pdf* is given by (Jain and Nanda, 1995; Navarro et al., 2006):

$$f_Z^{(w)}(z_o, z_1, z_2) = \frac{w(z_o, z_1, z_2)\, f_Z(z_o, z_1, z_2)}{E[w(z_o, z_1, z_2)]} \tag{8}$$

If the marginal *pdfs* are defined in the interval $[-1, 1]$, which usually requires rescaling the original values, the multivariate *pdf* can be expressed as a series of Legendre polynomials (Mustapha and Dimitrakopoulos, 2010):

$$f_Z^{(w)}(z_o, z_1, z_2) = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} L_{l,m,n}^{(w)} \overline{P}_l(z_0)\overline{P}_m(z_1)\overline{P}_n(z_2) \tag{9}$$

$L_{l,m,n}^{(w)}$ are the weighted Legendre coefficients and $\overline{P}_\alpha(z)$, $\alpha = l, m, n$, are the α-order normalized Legendre polynomials, such as:

$$\overline{P}_\alpha(z) = \sqrt{\frac{2\alpha+1}{2}} P_\alpha(z) \tag{10}$$

The first two Legendre polynomials are $P_0(z) = 1$ and $P_1(z) = z$. The following derivation allows obtaining Legendre polynomials of any order $\alpha \geq 0$ (Abramowitz and Stegun, 1964):

$$P_\alpha(z) = \frac{1}{2^\alpha \alpha!}\frac{d^\alpha}{dz^\alpha}\left[\left(z^2-1\right)^\alpha\right] = \sum_{j=1}^{\alpha} a_{j,\alpha} z^\alpha, \text{ for } z \in [-1, 1] \tag{11}$$

The Legendre polynomials fulfill the orthogonally property; if they are normalized this property is defined as:

$$\int_A \overline{P}_m(z)\,\overline{P}_n(z)\,dz = \begin{cases} 0, & \text{if } m \neq n \\ 1, & \text{if } m = n \end{cases} \tag{12}$$

After applying this property on Equation (9), the Legendre coefficients are derived as:

$$L_{l,m,n}^{(w)} = \int_{A^3} \overline{P}_l(z_0)\overline{P}_m(z_1)\overline{P}_n(z_2) f_Z^{(w)}(z_o, z_1, z_2)\, dz_0 dz_1 dz_2 \tag{13}$$

Doing $\overline{a}_{i,\alpha} = a_{i,\alpha}\sqrt{(2\alpha+1)/2}$ and substituting into Equation (11), the Legendre coefficients above become:

$$L_{l,m,n}^{(w)} = \sum_{i=0}^{l}\sum_{j=0}^{m}\sum_{k=0}^{n} \overline{a}_{i,l}\overline{a}_{j,m}\overline{a}_{k,n} \int_{A^3} z_0^i z_1^j z_2^k f_Z^{(w)}(z_o, z_1, z_2)\, dz_0 dz_1 dz_2 \tag{14}$$

Therefore, these coefficients can be expressed as combinations of the weighted high-order moments:

$$L_{l,m,n}^{(w)} = \sum_{i=0}^{l}\sum_{j=0}^{m}\sum_{k=0}^{n} \overline{a}_{i,l}\overline{a}_{j,m}\overline{a}_{k,n} E^{(w)}\left[Z_0^i Z_1^j Z_2^k\right] \tag{15}$$

Fitting a multivariate non-Gaussian distribution in this way can be computationally demanding. For this reason, the expansion of the Legendre polynomial series in Eq. (9) is truncated to an order ω, which, as a rule of thumb, is not larger than 12. Therefore, approximation of the multivariate *pdf* becomes:

$$f_Z^{(w)}(z_o, z_1, z_2) = \sum_{l=0}^{\omega} \sum_{m=0}^{l} \sum_{n=0}^{m} L_{l,m,n}^{(w)} \overline{P}_l(z_0) \overline{P}_m(z_1) \overline{P}_n(z_2) \tag{16}$$

To further elucidate the distribution fitting using declustered high-order statistics, consider a univariate example. Figure 1 shows the Walker Lake clustered dataset (Isaaks and Srivastava, 1989). The sizes of the sample marks in this figure are proportional to the declustering weight assigned to each sample. It is clear in this figure that high-grade areas are oversampled, while sample density in poor areas is lower. The fitting of the univariate distribution is first performed with cumulants of order 1 to 10 obtained from equally weighted samples. In this case, the Legendre fitting approximates the distribution of clustered data, as it can be observed in Figure 1(b). The fitting of the univariate distribution using declustered Legendre cumulants (red curve in Figure 1(b)) fits better the declustered histogram (black bars in Figure 1(b)).

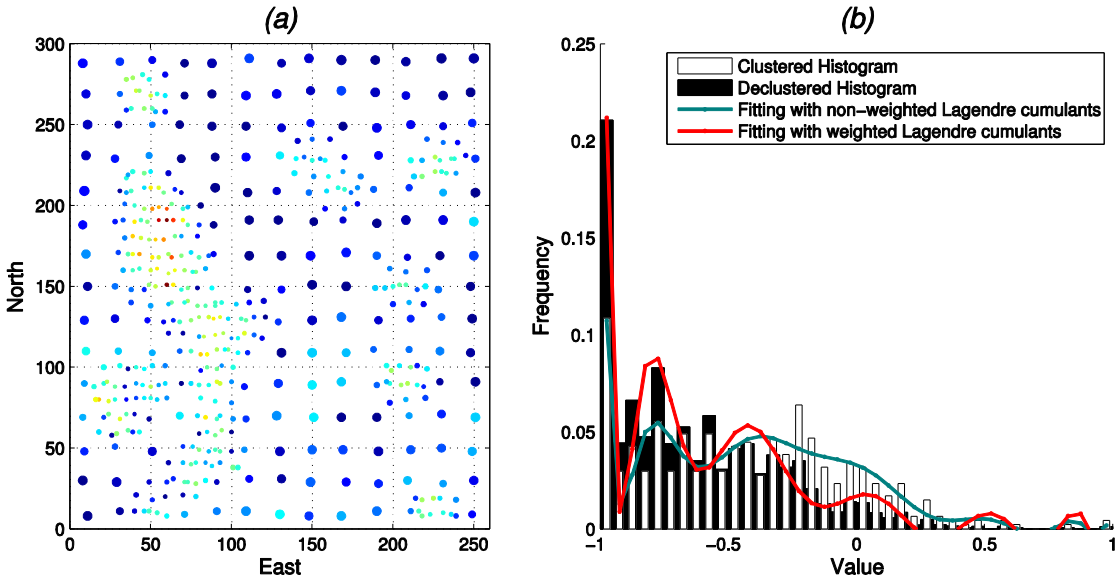

Figure 1: (a) Clustered dataset, (b) fitting of the clustered (white bars) and declustered (black bars) histogram using clustered (cyan curve) and declustered (red curve) Legendre cumulants.

Figure 2 compares the 3$^{\text{rd}}$-order clustered (Figure 2(b)) and declustered (Figure 2(c)) cumulant maps with a similar map obtained from the exhaustive Walker Lake dataset (Figure 2(a)). Both maps obtained from scattered data are les continuous than the one obtained from the exhaustive image, however, this is explained by the incompleteness of the information provided by the sampling dataset. The declustered 3$^{\text{rd}}$-order cumulant map shows some features that are comparable to the corresponding map for the exhaustive dataset.

The fitting of non-Gaussian multivariate distributions by means of series of Legendre cumulants and polynomials was originally implemented in the algorithm HOSIM (Mustapha and Dimitrakopoulos, 2011) for sequential simulation based on high-order spatial statistics. The modified version of HOSIM incorporates data declustering weights in the inference of spatial high-order statistics required for fitting declustered conditional distributions.

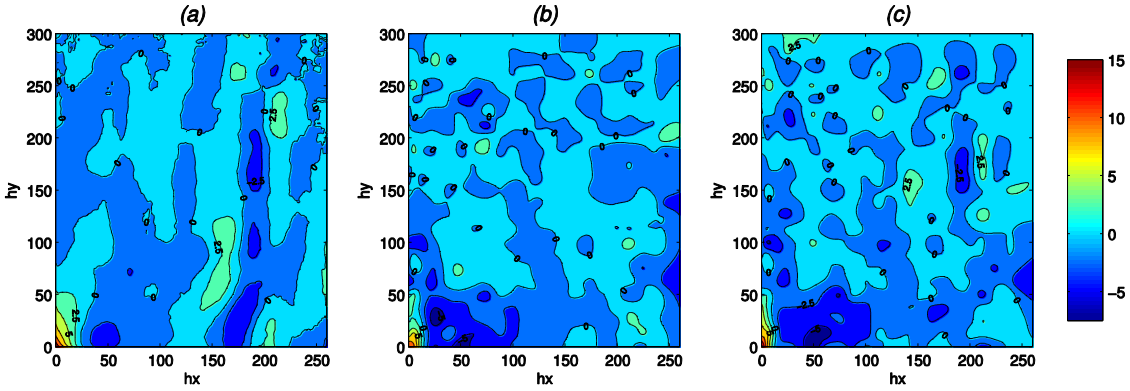

Figure 2: 3rd order cumulant maps for (a) Exhaustive dataset, (b) clustered samples with equal weight, (c) clustered samples with declustering weights.

3 Case Study: The Apensu Gold Deposit, Ghana

The dataset used for this case study is a subset of a larger drilling campaign dataset from the Apensu deposit, located in the Ahafo South district of Ghana and owned by Newmont Ghana Gold Ltd. The gold mineralization in this deposit is controlled by the dominant thrust fault and its subsidiary structures. The richer gold grades are found mainly in the hanging wall of the thrust fault (Jones et al., 2011). Newmont provided the drilling dataset containing the gold assays, as well as a tridimensional geological block model with blocks identified as the thrust fault, the two different families of subsidiary structures and the host rock. Taking into account the host rock, this model defines four mineralization domains.

3.1 Original data and training image

The selected region for this study comprises a volume of 300m × 600m × 150m, which is rotated to make the fault thrust strike parallel to the North South direction; this region is located at the center of the deposit, which is the most densely sampled area. Figure 3(a) shows a 3D view of the block model containing the mineralized domains in the selected region. The block size in the geological block model is 5m × 5m × 5m. Figure 3(b) shows the drill-hole traces in the selected region. Drillholes are, for the most part, perpendicular or sub-perpendicular to the main fault thrust.

The samples within the selected region are composited in 5-meter intervals and the composited gold grades rescaled between 0 and 10. This last was done for confidentiality reasons. The cell declustering method is used to assign declustering weights to the composites.

Table 1 shows the basic statistics for the clustered and declustered 4106 composited in the selected region. As this table shows, the impact of data clustering in the Au *cdf* and its statistics is significant.

The cumulant maps presented next are obtained using a rotated template of lag vectors $\boldsymbol{h}_{X'}$ (Az. 270°, Dip −55°), parallel to the main down-the-hole direction, $\boldsymbol{h}_{Y'}$ (Az. 0°, Dip 0°) and $\boldsymbol{h}_{Z'}$ (Az. 270°, Dip 35°), perpendicular to the first two lag vectors. These directions correspond to the configuration that allows maximizing the collection of replicates in three directions for the inference of robust high-order statistics.

Table 1: Basic statistics for clustered and declustered gold grades.

Statistics for (Au g/t):	Mean	Std. Dev.	Skewness	Kurtosis
Clustered composites:	0.24	0.52	6.10	52.5
Declustered composites:	0.19	0.44	7.04	78.8

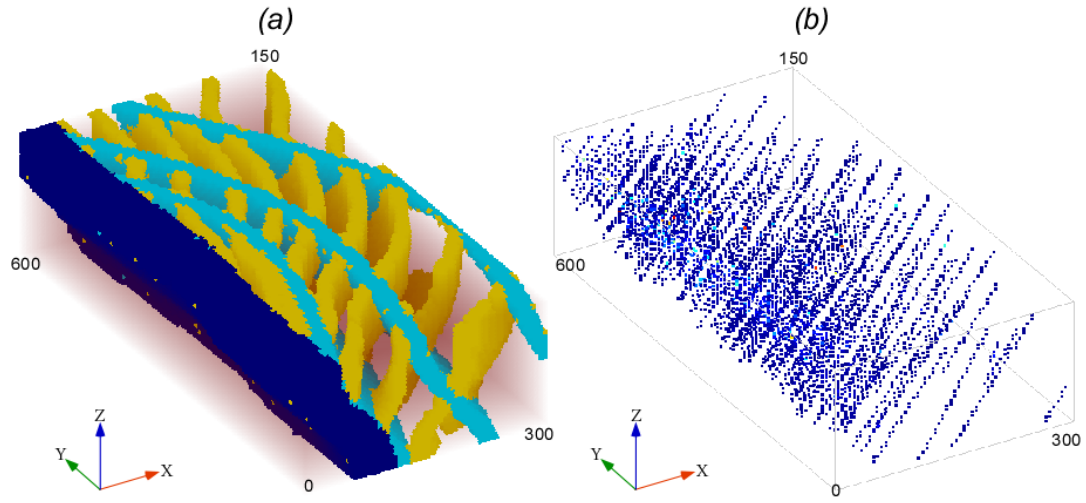

Figure 3: (a) Geological model view showing the main fault zone (dark blue) and the two families of subsidiary structures (pale blue and golden). (b) A similar view showing the drill-hole traces.

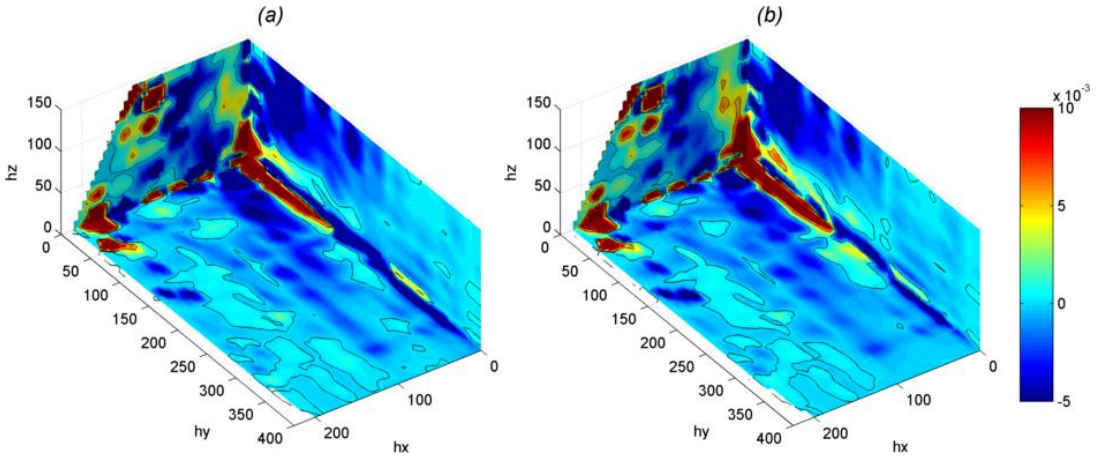

Figure 4: (a) clustered and (b) declustered third order cumulant maps obtained from drill hole data.

Figure 4 shows the clustered and declustered 3^{rd}-order cumulant maps obtained from the drill hole dataset. The differences between both maps are not very clear, although the 3^{rd}-order cumulant value at zero separation distance, which is related to the skewness, falls from 0.83 to 0.59.

Figure 5 presents three slices of the declustered (a) and clustered (b) 4^{th}-order cumulant volumes obtained from the drill hole samples. As for the 3^{rd}-order cumulants, the differences between clustered and declustered 4^{th}-order cumulants are not visually clear, however, the zero-lag 4^{th}-order cumulant falls from 3.69 to 2.89 after applying declustering weights.

The required training image was built using sequential Gaussian simulation (SGS) within the domains. The domains are defined by hard boundaries, and only composites within the boundaries of a domain were used for simulating the grades in it. The dimensions of the training image are the same as the dimensions of the zone of study. It is important to remark that the use of SGS for the construction of the training image raises concerns about its ability to reproduce the high-order statistics of data, given that the high-order cumulants of Gaussian distributions tend to zero (Sadler et al., 1994). However, there are few practical alternatives to SGS within domains for the construction of training images from scattered samples. Figure 6(a) and (b) show the 3^{rd} and 4^{th} order cumulant maps, respectively, obtained from the training image. The lag vectors

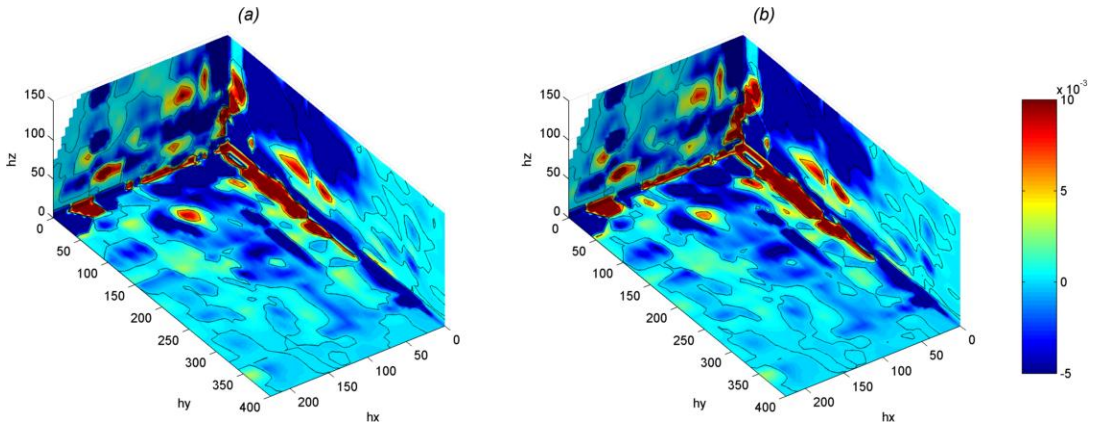

Figure 5: (a) unweighted and (b) weighted fourth order cumulant maps.

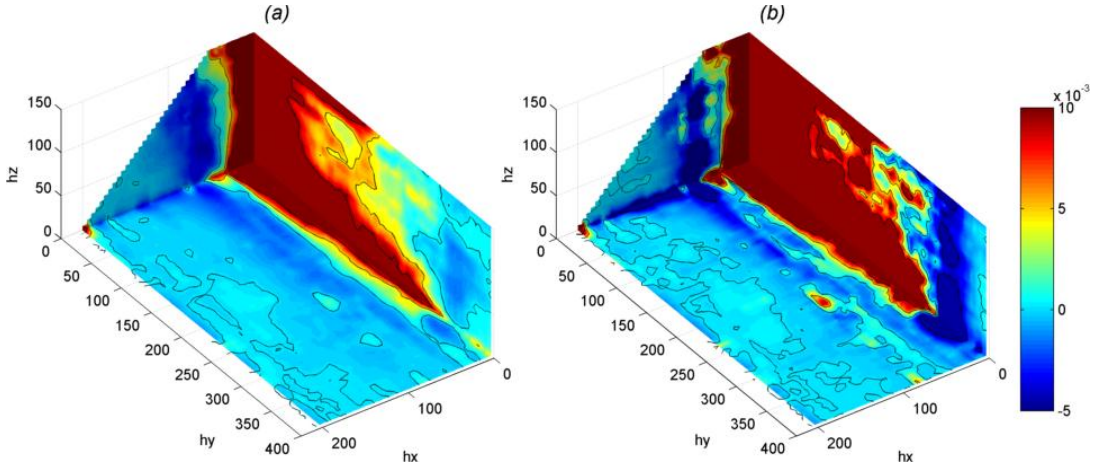

Figure 6: 3^{rd}-order (a) and 4^{th}-order (b) cumulant maps obtained from the training image.

that form the corresponding cumulant template layout are $\boldsymbol{h}_{X'}$ (Az. 270°, Dip −45°), $\boldsymbol{h}_{Y'}$ (Az. 0°, Dip 0°) and $\boldsymbol{h}_{Z'}$ (Az. 270°, Dip 45°). The main feature in these cumulant maps is the high value region in the plane of vectors $\boldsymbol{h}_{Y'}$ and $\boldsymbol{h}_{Z'}$. This is produced by the high-grade population simulated inside the fault domain. In the direction of $\boldsymbol{h}_{X'}$, that is sub-parallel to the main down-the-hole direction, the cumulant values fall quickly and stabilizes around zero. Besides some fluctuations in the $\boldsymbol{h}_{X'}$ and $\boldsymbol{h}_{Y'}$ plane, which are related to the presence of high grades in the subsidiary structures, this lack of spatial high-order features is expected for a Gaussian simulated image (Sadler et al., 1994).

3.2 High-order simulation

The composite dataset and the training image described in the previous section are used as inputs for the HOSIM algorithm. HOSIM superimposes the data to the training image and rescales the merged values between −1 and 1, which is necessary in order to respect the range of definition for the Legendre polynomials. Before simulating the grades, HOSIM scans the merged training to obtain and store the high-order statistics for all of the spatial templates that can be accommodated in a search template. In this case, the dimensions of the search template are 75m × 135m × 55m, with the longer side parallel to the y axis. The declustering sample weights are also superimposed to the equal weights of the training image nodes. The maximum number of neighboring samples to condition the simulation of a node is set as 24, and the order of the polynomial series is 7. A larger number of samples or higher orders for the Legendre polynomials increase considerably the computational effort with modest improvement of the realizations. Twenty HOSIM realizations have been

generated for both, the clustered and declustered, datasets. Figure 7 shows the average of these realizations, or E-type estimates, corresponding to (a) the clustered dataset and (b) the declustered dataset. Both groups of realizations reproduce the general features of mineralization, such as the high grade population in the fault zones and moderate grades in the subsidiary structures. However, as Table 2 shows, the averages of low and high-order statistics of the twenty realizations obtained using declustered cumulants match closely the equivalent statistics of the declustered dataset.

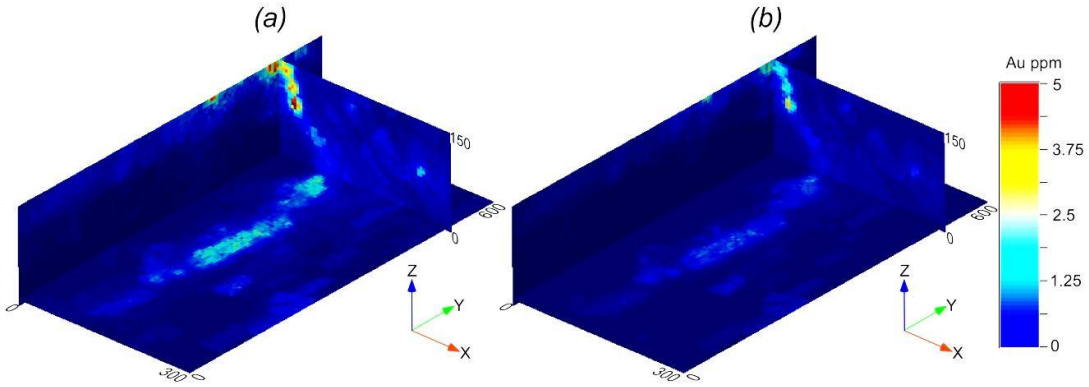

Figure 7: E-type estimates of 20 realizations obtained using (a) the clustered and (b) the declustered datasets.

Table 2: Statistics for clustered and declustered data, the training image and realizations with clustered and declustered high-order statistics.

Statistic	*Clustered data*	*Declustered data*	*Training image*	*Clustered HOSIM (average of 20 realizations)*	*Declustered HOSIM (average of 20 realizations*
Mean	0.24	0.19	0.19	0.24	0.18
Variance	0.27	0.19	0.20	0.26	0.19
Skewness	6.10	7.04	6.74	6.17	7.09
Kurtosis	52.5	78.8	66.6	52.8	71.0

Figure 8 shows the average 3$^{\text{rd}}$-order cumulant maps obtained from the twenty realizations using clustered (a) and declustered Legendre cumulants (b). Both maps show similar features, although the averages in the second map tend to be lower than in the clustered case. These maps combine features from both the data and the training image cumulant maps. For instance, the continuity of high values in the direction of the $\boldsymbol{h}_{Y'}$ vector is also present in the training image 3$^{\text{rd}}$-order cumulant map, whereas the high values at medium and long distances of vector $\boldsymbol{h}_{X'}$ are related to similar features in the data cumulant map. Similar observations can be made on the 4$^{\text{th}}$-order cumulant maps in Figure 9.

4 Discussion

The incorporation of declustering weights in the inference of the cumulants required for fitting the non-Gaussian conditional distributions in high-order simulation allows minimizing the bias introduced by clustered sampling. The resulting realizations reproduce the declustered data *cdf* and its low- and high-order single point statistics. This is important in resource evaluation, since otherwise, the recoverable reserves could be overestimated if samples are more abundant in richer zones. The training image has an important impact in the high-order multiple-point statistics reproduced by the realizations. This is expected given that as the points involved in the inference of the high-order statistics increases, more information needs to be borrowed

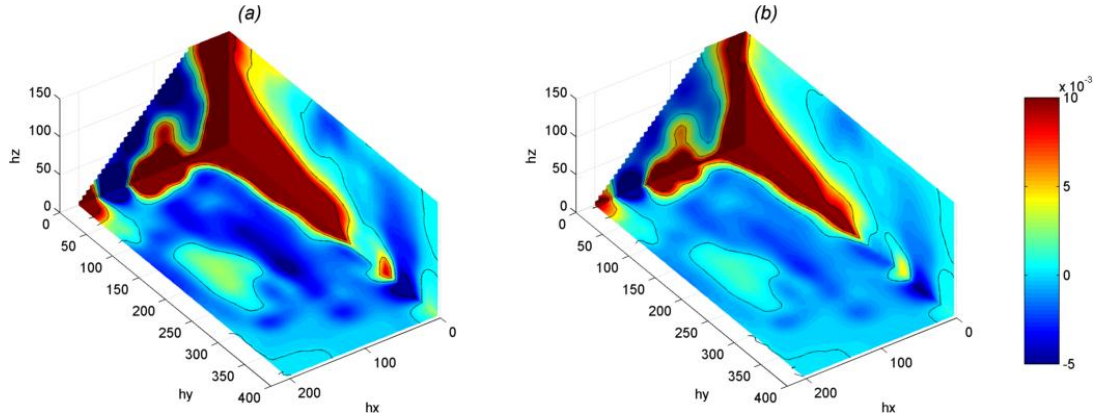

Figure 8: Average 3rd-order cumulant maps corresponding to the realizations generated using clustered (a) and declustered (b) Legendre cumulants.

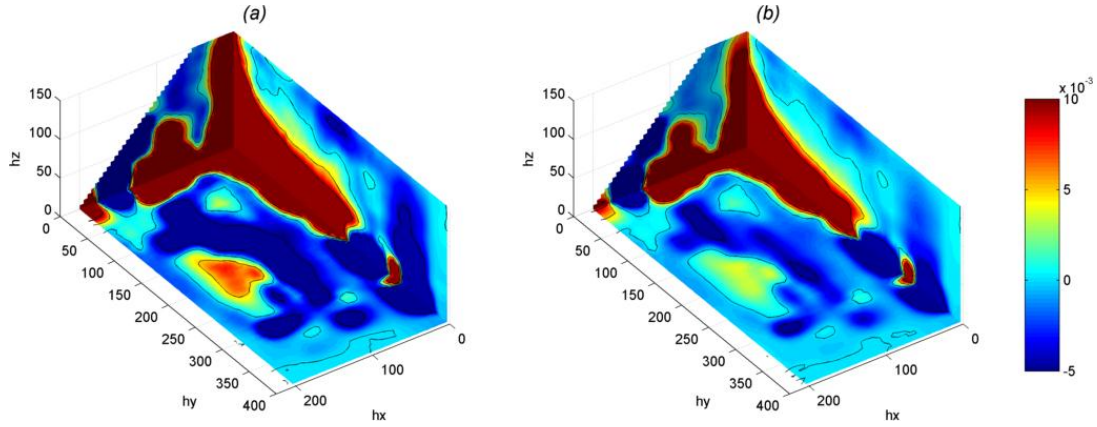

Figure 9: Average 4th-order cumulant maps corresponding to the realizations generated using clustered (a) and declustered (b) Legendre cumulants.

from the training image. In the previous study, since the training image is a composite of four different Gaussian images, one per each geological unit, it tends to override the high-order structures informed by the data cumulants. This poses the problem of generating training images that contain high-order statistics that are pertinent to the spatial features of the mineralization.

In this application to a structurally complex gold deposit, HOSIM succeeds in generating simulated models that respect the spatial features of gold mineralization. This is accomplished without the need to merge separated realizations of the geology and the grades. However, the mineralization in the fault zone, which is an abundantly sampled domain, is better reproduced in the simulated models than the mineralization in the smaller and less sampled subsidiary structures. The difficulty in modelling the spatial continuity of grades in the smaller structures is aggravated in this case by the statistical similarity between the gold population in these structures and in the background rock that host them. An option for ongoing research to deal with background noise is the use of weighted training images where features that are deemed more important are weighted heavier than others. Proceeding in this way could allow the spatial cumulants to capture more of the spatial structure of the important features.

5 Conclusions

Simulation using high-order statistics does not need to adopt a parametric model. Any shape of conditional distribution can be approximated using series of Legendre cumulants, where Legendre cumulants are functions of spatial cumulants. The incorporation of declustering weights in the inference of the spatial cumulants allows minimizing the biases caused by clustered sampling in the Legendre approximation of non-Gaussian distributions. High-order simulation allows modelling important spatial features of mineralization, even when they are complex. The resulting realizations reproduce the declustered one-point low- and high-order statistics of original data, but also tend to reproduce the multiple-point high-order statistics of training images. This is explained by the increased reliance on the information provided by the training image as the required spatial cumulants become more complex. An important issue is that Gaussian-based training images tend to override the information provided by samples in the inference of high-order statistics. The construction of non-Gaussian training images containing high-order statistics that are representative of the mineralization is a subject of future work.

References

Abramowitz, M., Stegun, I. A. (1964) *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Dover Publications.

Chilès, J.-P., Delfiner, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*. New York: Wiley.

Deutsch, C. (1989) DECLUS: A fortran 77 program for determining optimum spatial declustering weights. *Comput. Geosci.*, 15, 325–332.

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. *Mathematical Geosciences*, 42, 65–99.

Gómez-Hernández, J. J., Cassiraga, E. F. (1993) Theory and practice of sequential simulation. In *Geostatistical Simulations: Proceedings of the Geostatistical Simulation Workshop*, M. Armstrong & P. A. Dowd (eds.), Fontainebleau, France: Kluwer Academic Publishers.

Goovaerts, P. (1999) Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties. *Stochastic Environmental Research and Risk Assessment*, 13, 161–182.

Isaaks, E., Srivastava, R. M. (1989) *Applied Geostatistics*. New York: Oxford University Press.

Jain, K., Nanda, A. K. (1995) On multivariate weighted distributions. *Communications in Statistics – Theory and Methods*, 24, 2517–2539.

Jones, P., Douglas, I., Jewbali, A. (2011) Modelling geological uncertainty in mining using multiple-point statistics. In *World Gold 2011, 3rd International Conference*, Montreal, Canada.

Journel, A., Deutsch, C. (1993) Entropy and spatial disorder. *Mathematical Geology*, 25, 329–355.

Journel, A. G. (1994) Modeling uncertainty: Ssome conceptual thoughts. In *Geostatistics for the Next Century – an International Forum in Honour of Michel Davids Contribution to Geostatistics*, Montreal, 1993, R. Dimitrakopoulos (ed.), 30–43. Dordrecht: Kluwer Academic Publ.

Leuangthong, O., McLennan, J. A., Deutsch, C. V. (2004) Minimum acceptance criteria for geostatistical realizations. *Natural Resources Research*, 13, 131–141.

Mustapha, H., Dimitrakopoulos, R. (2010) High-order stochastic simulation of complex spatially distributed natural phenomena. *Mathematical Geosciences*, 42, 457–485.

Navarro, J., Ruiz J. M., Del Aguila, Y. (2006) Multivariate weighted distributions: A review and some extensions. *Statistics*, 40, 51–64.

Sadler, B. M., Giannakis, G. B., Keh-Shin, L. (1994) Estimation and detection in non-Gaussian noise using higher order statistics. *Signal Processing, IEEE Transactions on*, 42, 2729–2741.