Les Cahiers du GERAD ISSN: 0711–2440

High-order simulation at block support scale

D.F. Machuca-Mory,
R. Dimitrakopoulos, T. Pham

G–2014–57

August 2014

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

High-order simulation at block support scale

David F. Machuca-Mory[a]

Roussos Dimitrakopoulos[a,b]

Truong Pham[a]

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

[b] *and GERAD*

david.machucamory@mcgill.ca
roussos.dimitrakopoulos@mcgill.ca

August 2014

Les Cahiers du GERAD

G–2014–57

Abstract: The use of spatial high-order statistics has been previously proposed as an alternative to introduce richer information about complex spatial patterns in the simulation of continuous attributes. These statistics are normally inferred from exhaustive quasi -support training images. Spatial high-order statistics values are combined within series of orthogonal polynomials to approximate local conditional distributions that can be used for the drawing of simulated point-support values. This paper extends this formalism to direct block-support simulation. This is achieved by inferring block-point high-order statistics from up-scaled training images and incorporating these statistics in the orthogonal polynomials approximation of the conditional distributions. This methodology is computationally expensive, so a reasonable option is to approximate all the required local conditional distributions only once. These can be subsequently sampled by different fields of correlated probabilities to produce multiple realizations of the attribute. The resulting simulated maps reproduce the high-order statistics of the up-scaled training image and they also match the up-scaled global distribution of the attribute.

Key Words: Geostatistical simulation, high-order spatial statistics, spatial uncertainty, orthogonal polynomials, conditional probabilities, block support.

1 Introduction

In mining, environmental sciences, and other contexts, numerical models are usually needed in a resolution that is much coarser than the sampling resolution. The geographical space where an ore deposit or a contaminated area occurs is represented by a 2D or 3D mesh with cell, or block, sizes that correspond to the minimum volume over which real life decisions can be taken effectively and efficiently. The problem of modelling block-support attributes from practically point volume data has been an important aim since the early years of Geostatistics (David 1977).

Spatial simulation techniques produce numerical models that reproduce the probability distribution (pdf) and spatial variability of the input dataset (Journel and Huijbregts 1978, Journel 1974). These techniques are conditioned by available hard and soft information about the spatial attribute. An ensemble of simulated numerical models, or realizations, all of them with similar statistics but locally different, can be used to quantify the uncertainty about the spatial distribution of ore grades and other rock attributes (Journel 1994). In mining, pit optimization and production scheduling algorithms that incorporate the uncertainty space provided by the simulated models can result in production scheduling that maximizes the economic return while minimizing risks related to production fluctuations (Godoy 2003, Godoy and Dimitrakopoulos 2004). Underlying most conditional simulation methods for continuous attributes is the inference of a local conditional cumulative distribution function (ccdf) given hard-data and secondary soft-information, if this last is available (Christakos 1992, Chilès and Delfiner 1999). These ccdfs are sampled randomly to produce simulated values (Isaaks 1990). Under a sequential approach, previously simulated values are used for the conditioning of local cdf at further locations along a random path. Changing the random path leads to different full field realizations of the numerical model of the attribute. This require the re-estimation of all the local ccdfs for each new realization.

The sequential approach for conditional simulation is theoretically sound but can be computationally demanding. An efficient alternative is the probability field (p-field) methodology (Srivastava 1992, Froidevaux 1993) although its statistical properties are not well established for conditional simulation (Chilès and Delfiner 2012). Contrarily to the sequential simulation approach, p-field simulation requires the inference of the local ccdfs only once. To produce different realizations of the attribute, the local ccdfs are sampled by different fields of simulated spatially correlated probabilities. P-field realizations, however, present artifacts nearby conditioning samples and do not reproduce the input model of spatial correlation satisfactorily (Pyrcz and Deutsch 2001, Srivastava and Froidevaux 2005)

Geostatistical simulation is commonly performed over small volumes of size comparable to the individual samples dimensions. The simulated block support values are obtained by averaging the sample scale simulated values within each block. This can be efficiently achieved by applying the LU methodology for obtaining realizations at the discretization points within each block (Glacken 1996). The block-support simulated values can then be used for conditioning other blocks (Godoy 2003). Alternatively, instead of discretizing the blocks, a change of support model can be applied to directly obtain the block-support realizations from point-support data (Emery 2009). The ensemble of block scale simulated values, obtained either by averaging or by a change of support model, over multiple realizations describes the local block scale uncertainty. These methods are limited by their reliance on the multiGaussian model and 2-point statistics, such as variograms and covariances. Gaussian based methods result in maximum entropy realizations given a distribution and variogram model (Journel and Deutsch 1993). This, coupled with the limited capability of 2-point statistics for describing complex non-linear patterns (Journel 2005), difficult the reproduction of geologically realistic numerical models by using the traditional simulation methods.

During the last decade, multiple-point simulation methods have been developed in response to the limitations of traditional methods based on 2-point statistics. Multiple-point simulation methods for continuous variables, such as those based on the filtering of patterns (Zhang, Switzer and Journel 2006, Wu, Boucher and Zhang 2008) or direct sampling (Mariethoz, Renard and Straubhaar 2010) work at the same scale of the samples. Up-scaling of the fine scale realizations is then needed to produce the block scale realizations. More important, multiple-point methods do not guarantee the reproduction of all input low and high-order statistics by the resulting realizations (Boucher 2009, Osterholt 2006).

Spatial high-order moments and cumulants are able to describe complex non-linear patterns (Mustapha and Dimitrakopoulos 2010c, Dimitrakopoulos, Mustapha and Gloaguen 2010). They have been proposed for informing the fitting of non-Gaussian conditional distributions by incorporating them in either Legendre or Laguerre polynomial series (Mustapha and Dimitrakopoulos 2010b, Mustapha and Dimitrakopoulos 2010a). An initial algorithm for sequential simulation using high-order statistics was presented later with encouraging results (Mustapha and Dimitrakopoulos 2011).

This paper explains the theoretical and practical details of the approximation of local cdfs conditioned by hard data using high-order moments obtained from a training image and Legendre polynomial series. This approach is extended to the conditioning of block-support values by point-support samples. Spatial stochastic simulations based on multiple-point and high-order statistics are considerably more computationally demanding than those based on 2-point statistics. Extending high-order simulation to direct block scale simulation requires obtaining a very large number of point-point and cross point-block high-order moments. This, together with the calculation-intensive ccdf fitting by Legendre polynomials, precludes the implementation of a sequential approach for high-order simulation at block scale, at least until smarter algorithms are devised. Instead, a p-field simulation approach is adopted for high-order simulation. This paper presents an algorithm that implements point-point and point-block conditioning of local cdfs using high-order moments. The obtained ccdfs are sampled by simulated fields of correlated probabilities to obtain non-Gaussian realizations of the attribute. Generating the probability fields is considerably much less expensive than approximating the local ccdfs by series of high-order statistics. Thus, despite the recognized drawbacks of the p-field method, this seems to be the only viable alternative for high-order simulation at block scale by now.

2 Point-block high-order simulation

Normally, in geostatistical modelling, a 2D or 3D domain D is discretized by a mesh of $N_{\mathbf{v}}$ blocks $\mathbf{v}_\beta,\ \beta = 1,\ldots,N_v$ of support $\mathbf{v}$. The unknown values of the attribute at each location are modelled by different Random Variables (RVs) that together form a spatially correlated Random Field (RF) (Matheron 1970, Christakos 1992). Given a quasi-point support dataset $z(\mathbf{u}_\alpha),\ \alpha = 1,\ldots,n$, which may include previously simulated values, and under the stationarity decision (Myers 1989) over D, the ccdf of the RV Z at a point $\mathbf{u}_k$ within block $\mathbf{v}$ is expressed by (Isaaks 1990)

$$\begin{aligned} F_Z(\mathbf{u}_k \in \mathbf{v}; t|z(\mathbf{u}_1) &= z_1,\ldots,z(\mathbf{u}_n) = z_n) \\ &= Prob\{Z(\mathbf{v}) \le t|z(\mathbf{u}_1) = z_1,\ldots,z(\mathbf{u}_n) = z_n\} \\ &= \frac{Prob\{Z(\mathbf{u}_k) \le t, z_1,\ldots,z_n\}}{Prob\{z_1,\ldots,z_n\}}. \end{aligned} \tag{1}$$

A simulated value at block support $z^l(\mathbf{v})$ can be obtained by averaging the simulated values drawn from multiple ccdfs evaluated at K points inside $\mathbf{v}$ such as $\mathbf{u}_k \in \mathbf{v}, \forall k = 1,\ldots,K$. The block support ccdf, $F(\mathbf{v}; z|z(\mathbf{u}_1),\ldots,z(\mathbf{u}_n))$, is then approximated by the distribution of multiple $z^l(\mathbf{v})$ values obtained from different realizations of the simulation algorithm.

The construction of the joint distribution in the equation above is well defined and computationally efficient if they are assumed multiGaussian (Ripley 1987, Alabert 1987a, Chilès and Delfiner 1999, Alabert and Massonnat 1990). In this case, as well as for sequential indicator simulation (Alabert 1987b, Journel and Isaaks 1984), the knowledge of the univariate marginal distribution and the spatial 2-point statistics is enough for building the required joint multivariate distributions.

In high-order simulation, the estimation of the non-Gaussian ccdfs requires the inference of high-order moments from sampling data and/or training images. Moments are quantities that, depending of their order, describe different features of univariate or multivariate distributions. Thus, the first order moment, the mean, gives the central tendency of the distribution; the second order moment yields to the variance and the covariance, which measure the spread and correlation of univariate and bivariate distributions, respectively; the third and fourth order moments inform of the asymmetry or "peakedness" of the probability distribution. In a spatial context, the combinations of high-order moments, known as cumulants, are able to describe

complex spatial relationships that 2-point statistics, such as the variogram and the covariance, cannot resolve. This section presents the point-block high-order moments and how they can be used for approximating the block-support local ccdf. These ccdfs can subsequently be used as the input for p-field simulation.

2.1 Block-point high-order statistics

Let us consider $Z_0 = Z(\mathbf{u}_0)$, a Random Variable (RV) anchored at point $\mathbf{u}_0$, and $Z_1, \dots \cdot Z_n$ as the RVs anchored at the end-points of vectors $\mathbf{h}_1, \dots, \mathbf{h}_n$ that radiate from $\mathbf{u}_0$. A $(n+1)$-point moment of order ω of these RVs is given by the expectation (Kendall et al. 1994)

$$\mu_{w_0 w_1 \dots w_n} = E\left[Z_0^{w_0} \cdot Z_1^{w_1} \cdot \dots \cdot Z_n^{w_n}\right], \tag{2}$$

where $w_0 + w_1 + \dots + w_n = \omega$, and $w_0, w_1, \dots, w_n \in \mathbb{N}^0$. The set of points $\mathbf{u}_0$, $\mathbf{u}_0+\mathbf{h}_1, \dots, \mathbf{u}_0+\mathbf{h}_n$ is termed as the $n+1$-point template of the moment $\mu_{w_0 w_1 \dots w_n}$, and it is represented as $\tau_{\mathbf{h}_1,\dots,\mathbf{h}_n}$. Finding enough data events that replicate this template in order to obtain a robust estimate of the high-order moment may not be feasible with scattered samples, thus, an exhaustive image is usually needed for such purpose. This image, known as training image (TI), is deemed to contain the same spatial patterns as the geological setting from where the samples were taken from.

If the attribute's values average linearly, the block support RV is obtained by the integration of point-support values within the block volume (Matheron 1963):

$$Z(\mathbf{v}) = \frac{1}{|\mathbf{v}|} \int\limits_{\mathbf{u}\in\mathbf{v}} Z(\mathbf{u})\, d\mathbf{u}. \tag{3}$$

If the support of the centre of the template is a volume, instead of a point, the corresponding block-point high-order moment is

$$\mu_{\mathbf{v}, w_0 w_1 \dots w_n} = E\left[Z_{\mathbf{v}}^{w_0} \cdot Z_1^{w_1} \cdot \dots \cdot Z_n^{w_n}\right]. \tag{4}$$

When $w_0 = 1$ the high order moment becomes,

$$\begin{aligned} E\left[Z_{\mathbf{v}} \cdot Z_1^{w_1} \cdot \dots \cdot Z_n^{w_n}\right] &= E\left[\left(\frac{1}{|\mathbf{v}|} \int\limits_{\mathbf{x}\in\mathbf{v}} Z(\mathbf{x})\, d\mathbf{x}\right) \cdot Z_1^{w_1} \cdot \dots \cdot Z_n^{w_n}\right] \\ &= \frac{1}{|\mathbf{v}|} \int\limits_{\mathbf{x}\in\mathbf{v}} E\left[Z(\mathbf{x}) \cdot Z_1^{w_1} \cdot \dots \cdot Z_n^{w_n}\right] d\mathbf{x}. \end{aligned}$$

This can be regarded as a multiple-point generalization of the block-point non centred covariance. For any integer power $w \geq 1$ we can generalize the high-order block-point moment as

$$\begin{aligned} E\left[Z_{\mathbf{v}}^{w} \cdot Z_1^{w_1} \cdot \dots \cdot Z_n^{w_n}\right] &= E\left[\left(\frac{1}{|\mathbf{v}|} \int\limits_{\mathbf{x}\in\mathbf{v}} Z(\mathbf{x})\, d\mathbf{x}\right)^{w} \cdot Z_1^{w_1} \cdot \dots \cdot Z_n^{w_n}\right] \\ &= \frac{1}{|\mathbf{v}|^{w}} \int\limits_{\mathbf{x_1}\in\mathbf{v}} \cdots \int\limits_{\mathbf{x_w}\in\mathbf{v}} E\left[\left(Z(\mathbf{x}_1) \cdots Z(\mathbf{x}_w)\right) \cdot Z_1^{w_1} \cdot \dots \cdot Z_n^{w_n}\right] d\mathbf{x}_1 \cdots d\mathbf{x}_w. \end{aligned} \tag{5}$$

These high-order multiple-point statistics can be computationally very demanding to infer from a high resolution TI by using expression (5). An approximate, but more efficient alternative is get the $Z_{\mathbf{v}}$ values by up-scaling the TI and using these up-scaled values directly, as in Expression (4). This alternative is suggested to obtain the block-point high-order moments required for the non-Gaussian cdf conditioning explained next.

2.2 Block cdf conditioning by point scale data using high-order statistics

In image processing literature, it is common to find the bivariate pdf approximation by a truncated series of Legendre cumulants L_{k_0,k_1} and orthogonal Legendre polynomials $P_k(z)$ (Yap and Paramesran 2005, Liao

and Pawlak 1996, Teh and Chin 1988, Teague 1980, Hosny 2007). These and other authors use Legendre and other orthogonal polynomials for the compression and reconstruction of 2D images and 3D volumes. In geostatistics we are not trying to rebuild an image, although we use a training image as a support for the inference of the high-order moments. But this training image is analogous to the attribute we intend to model only in the sense that we assume that both share the same spatial patterns and the same stationary low and high-order statistics. The available samples of the attribute are not scattered information pieces of the training image we want to rebuild, nor does the training image provide locally accurate information about the spatial distribution of the attribute. Instead, we try to use the information about complex spatial structures provided by a training image and, at a less extent, by the dataset to model the space of uncertainty due to the lack of exhaustive information about the attribute. The local ccdfs model this uncertainty.

Applying the Legendre series to the approximation of local ccdfs is a considerably more challenging task than their application to image reconstruction since it requires moving from a 2D or 3D context to a multidimensional context. But let us consider by now three continuous RVs, one at block support $Z_{\mathbf{v}}$, and the other two, Z_1 and Z_2 at point support, all of them with values within the interval $[-1,1] \in \mathbb{R}$. The point support RVs are separated from the centre of the block $\mathbf{v}$ by vectors separation $\mathbf{h}_1$ and $\mathbf{h}_2$. The corresponding 3-variate pdf can be approximated by the Legendre series of maximum order ω in the following way (Teh and Chin 1988),

$$f_Z(z_{\mathbf{v}}, z_1, z_2) \approx \sum_{i_{\mathbf{v}}=0}^{\omega} \sum_{i_1=0}^{i_{\mathbf{v}}} \sum_{i_2}^{i_1} P_{k_{\mathbf{v}}}(z_{\mathbf{v}})\, P_{k_1}(z_1)\, P_{k_2}(z_2) L_{k_{\mathbf{v}},k_1 k_2}, \tag{6}$$

with $k_{\mathbf{v}} = i_{\mathbf{v}} - i_1$, $k_1 = i_1 - i_2$ and $k_2 = i_2$. The Legendre polynomials $P_k(z)$ can be obtained recursively (Abramowitz and Stegun 1964) or, alternatively, from its explicit formulation as (Weisstein):

$$P_k(z) = \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j 2^{-k} \frac{(2k-2j)!}{j!\,(k-1)!\,(k-2j)!} z^{k-2j} = \sum_{j=0}^{\lfloor k/2 \rfloor} a_{j,k} z^{k-2j}, \tag{7}$$

with $floor(k/2)$ as the highest integer smaller than $k/2$. Whereas, the 3-variate block-point Legendre cumulants are defined by (Lebedev and Silverman 1965)

$$L_{k_{\mathbf{v}},k_1,k_2} = \frac{(2k_{\mathbf{v}}+1)(2k_1+1)(2k_2+1)}{2^{n+1}} \times \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} P_{k_{\mathbf{v}}}(z_{\mathbf{v}})\, P_{k_1}(z_1)\, P_{k_2}(z_2)\, f(z_{\mathbf{v}}, z_1, z_2)\, dz_{\mathbf{v}} dz_1 dz_2. \tag{8}$$

Since the Legendre polynomials are defined in the interval $[-1,1] \in \mathbb{R}$, the original point and block support values must be transformed within that interval. The norm $(2k_{\mathbf{v}}+1)\cdots(2k_n+1)/2^{n+1}$ depends of the number of variables in the multivariate distribution. In this case, since there are two conditioning samples, $n = 2$. Replacing the Legendre polynomials in (8) by their equivalences from (7), and applying the definition of non-centred multivariate moments, the Legendre cumulant becomes (Teague 1980)

$$L_{k_{\mathbf{v}},k_1,k_2} = \frac{(2k_{\mathbf{v}}+1)(2k_1+1)(2k_2+1)}{2^{n+1}} \times \sum_{j_{\mathbf{v}}=0}^{\lfloor k_{\mathbf{v}}/2 \rfloor} \sum_{j_1=0}^{\lfloor k_1/2 \rfloor} \sum_{j_2}^{\lfloor k_2/2 \rfloor} a_{j_{\mathbf{v}},k_{\mathbf{v}}} a_{j_1,k_1} a_{j_2,k_2} E\left[Z_{\mathbf{v}}^{k_{\mathbf{v}}-2j_{\mathbf{v}}} \cdot Z_1^{k_1-2j_1} \cdot Z_2^{k_2-2j_2}\right], \tag{9}$$

where $E\left[Z_{\mathbf{v}}^{k_{\mathbf{v}}-2j_{\mathbf{v}}} \cdot Z_1^{k_1-2j_1} \cdot Z_2^{k_2-2j_2}\right]$ is a block-point high-order moment. The Legendre cumulants above can also be expressed in terms of spatial cumulants (Mustapha and Dimitrakopoulos 2010b), since moments can be expressed in terms of cumulants, and vice versa (Smith 1995). However, there is no practical advantage of doing so, because it is easier to obtain the spatial high-order moments than the spatial cumulants, and the former are the building blocks of both cumulants and Legendre cumulants. Being combinations of moments, spatial cumulants carry more information than single moments, but the same can be said of Legendre cumulants.

The block support pdf conditioned by the two point support values can be obtained by the relation:

$$f_Z\left(Z_{\mathbf{v}}|z_1,z_2\right)=\frac{f_Z\left(z_{\mathbf{v}},z_1,z_2\right)}{f_Z\left(z_1,z_2\right)}=\frac{f_Z\left(z_{\mathbf{v}},z_1,z_2\right)}{\int_{-1}^{1}f_Z\left(z_{\mathbf{v}},z_1,z_2\right)dz_{\mathbf{v}}}, \tag{10}$$

And the conditional probability of the block value being smaller or equal to a threshold t can be obtained by,

$$F_Z\left(Z_{\mathbf{v}}\leq t|z_1,z_2\right)=\frac{\int_{-1}^{t}f_Z\left(z_{\mathbf{v}},z_1,z_2\right)dz_{\mathbf{v}}}{\int_{-1}^{1}f_Z\left(z_{\mathbf{v}},z_1,z_2\right)dz_{\mathbf{v}}}. \tag{11}$$

By integrating the multivariate pdf (6) according to the expression above, the approximated block-support ccdf is given by

$$F_Z\left(Z_{\mathbf{v}}\leq t|z_1,z_2\right)\approx \frac{\sum_{i_{\mathbf{v}}=0}^{\omega}\sum_{i_1=0}^{i_{\mathbf{v}}}\sum_{i_2}^{i_1}P_{k_1}\left(z_1\right)P_{k_2}\left(z_2\right)L_{k_{\mathbf{v}},k_1k_2}\sum_{j_{\mathbf{v}}=0}^{\lfloor k_{\mathbf{v}}/2\rfloor}a_{j_{\mathbf{v}},k_{\mathbf{v}}}\left(\frac{t^{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}-(-1)^{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}}{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}\right)}{\sum_{i_{\mathbf{v}}=0}^{\omega}\sum_{i_1=0}^{i_{\mathbf{v}}}\sum_{i_2}^{i_1}P_{k_1}\left(z_1\right)P_{k_2}\left(z_2\right)L_{k_{\mathbf{v}},k_1k_2}\sum_{j_{\mathbf{v}}=0}^{\lfloor k_{\mathbf{v}}/2\rfloor}a_{j_{\mathbf{v}},k_{\mathbf{v}}}\left(\frac{1^{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}-(-1)^{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}}{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}\right)}, \tag{12}$$

where, as previously, $k_{\mathbf{v}}=i_{\mathbf{v}}-i_1$, $k_1=i_1-i_2$ and $k_2=i_2$.

Alternatively, the block-support 1-point w-order moment conditioned by two point-support data is given by the relation

$$\begin{aligned}E\left[Z_{\mathbf{v}}^{w}|z_1,z_2\right]&=\int_{-1}^{1}z_{\mathbf{v}}^{w}f_Z\left(Z_{\mathbf{v}}|z_1,z_2\right)dz_{\mathbf{v}}\\&=\frac{\int_{-1}^{-1}z_{\mathbf{v}}^{w}f_Z\left(z_{\mathbf{v}},z_1,z_2\right)dz_{\mathbf{v}}}{\int_{-1}^{1}f_Z\left(z_{\mathbf{v}},z_1,z_2\right)dz_{\mathbf{v}}}.\end{aligned} \tag{13}$$

Therefore, similar to expression (12), the conditional high-order moment can be approximated by

$$E\left[Z_{\mathbf{v}}^{w}|z_1,z_2\right]\approx \frac{\sum_{i_{\mathbf{v}}=0}^{\omega}\sum_{i_1=0}^{i_{\mathbf{v}}}\sum_{i_2}^{i_1}P_{k_1}\left(z_1\right)P_{k_2}\left(z_2\right)L_{k_{\mathbf{v}},k_1k_2}\sum_{j_{\mathbf{v}}=0}^{\lfloor k_{\mathbf{v}}/2\rfloor}a_{j_{\mathbf{v}},k_{\mathbf{v}}}\left(\frac{1^{w+k_{\mathbf{v}}-2j_{\mathbf{v}}+1}-(-1)^{w+k_{\mathbf{v}}-2j_{\mathbf{v}}+1}}{w+k_{\mathbf{v}}-2j_{\mathbf{v}}+1}\right)}{\sum_{i_{\mathbf{v}}=0}^{\omega}\sum_{i_1=0}^{i_{\mathbf{v}}}\sum_{i_2}^{i_1}P_{k_1}\left(z_1\right)P_{k_2}\left(z_2\right)L_{k_{\mathbf{v}},k_1k_2}\sum_{j_{\mathbf{v}}=0}^{\lfloor k_{\mathbf{v}}/2\rfloor}a_{j_{\mathbf{v}},k_{\mathbf{v}}}\left(\frac{1^{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}-(-1)^{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}}{k_{\mathbf{v}}-2j_{\mathbf{v}}+1}\right)} \tag{14}$$

The first conditional moments, from order 1 to 4, usually suffice to provide an acceptable approximation of the conditional cdf. This is a more efficient alternative to fitting the ccdf for 9 or more thresholds using expression (12). The general expressions for Legendre cumulants (9), the approximate block-support ccdf (12) and the conditional moments (14) for any number of point and block-support conditioning data are presented in the Appendix A.

Given a maximum approximation order ω and n conditioning samples, the number of coefficients required for approximating the ccdf or the conditional moments using either expression (12) or (14) is calculated by the formula:

$$N_{coeff}=\frac{(\omega+1)(\omega+2)\cdots(\omega+n+1)}{(n+1)!}. \tag{15}$$

This quickly leads to a very high number of required coefficients as ω and n increases, and, consequently, a rapidly exploding computational effort. Mustapha and Dimitrakopoulos (2011) suggest that, in order to reduce the computational complexity of the ccdf approximation, the moments $E\left[Z_0^{w_0}\cdot Z_1^{w_1}\cdot\ldots\cdot Z_n^{w_n}\right]$ with $w_0,\ w_1\cdots w_n>1$ can be neglected since they tend to zero. This is true for very high orders, i.e. $w_0,\ w_1\cdots w_n>10$, and when the number of conditioning samples is large. Nevertheless, for few conditioning samples and $\omega=w_0+,\ w_1+\cdots+w_n\leq 6$, the contribution of those moments is still considerable. Also, they state that moment of the form $E\left[Z_0^{w_0}\cdot Z_1^{w_1}\cdot Z_n^{1}\right]$ will tend to $E\left[Z_0^{1}\cdot Z_1^{1}\cdot Z_n^{1}\right]$ if w_0 and w_1 are close

to 1, but the difference between moments $E\left[Z_0^2 \cdot Z_1^2 \cdot Z_n^1\right]$ and $E\left[Z_0^1 \cdot Z_1^1 \cdot Z_n^1\right]$ can be also considerable. Another suggestion for the efficient approximation of the ccdf is to consider only the Legendre cumulants $L_{k_{\mathbf{v}}=0,k_1=0,k_2=0}$, $L_{k_{\mathbf{v}}=1,k_1=0,k_2=0}$, $L_{k_{\mathbf{v}}=1,k_1=1,k_2=0}$, and $L_{k_{\mathbf{v}}=1,k_1=1,k_2=1}$, but to disregard $L_{k_{\mathbf{v}}=1,k_1=0,k_2=1}$, $L_{k_{\mathbf{v}}=0,k_1=1,k_2=1}$ since they are implicitly included in $L_{k_{\mathbf{v}}=1,k_1=1,k_2=1}$ (Mustapha and Dimitrakopoulos 2011). Nonetheless, the information that $L_{k_{\mathbf{v}}=1,k_1=1,k_2=0}$ conveys is different from the information provided by $L_{k_{\mathbf{v}}=1,k_1=0,k_2=1}$ and $L_{k_{\mathbf{v}}=0,k_1=1,k_2=1}$, thus, including the first Legendre cumulant but skipping the other two may introduce unwarranted fluctuations in the ccdf approximation by expression (12). These observations are supported by the results of the test example in the following section. In view of these issues, a provisional rule of thumb for keeping the computational effort within the limits of available computational resources is to limit the approximation order $\omega \leq 5$.

The methodology that implements the equations presented in this section is described next.

3 An algorithm for ccdf approximation using high-order statistics

As mentioned before, p-field simulation is an efficient alternative to a sequential approach that would require the inference of point-block cross high-order moments and much more complex Legendre cumulants and ccdf approximation. P-field simulation requires a field of previously estimated ccdfs to sample it by multiple realizations of a field of correlated probabilities. This section describes the methodology for inferring non-Gaussian ccdfs using the theory explained in the previous section.

3.1 Description of the algorithm

The following is the general overview of a proposed algorithm for the inference of local ccdfs based on high-order statistics:

1. Transform the original data, training image and reference distributions to the interval $[-1, 1]$
2. Rasterize the search ellipsoid
3. On basis of the block dimensions, create an up-scaling template for averaging training image values
4. Initialize the containers for storing the high-order moments and Legendre cumulants
5. Start the iteration over all the nodes of the output grid
6. At each block $\mathbf{v}$ find the informed neighbours $z(\mathbf{u}_1), \ldots, z(\mathbf{u}_n)$ at $\mathbf{v}+\mathbf{h}_1, \ldots, \mathbf{v}+\mathbf{h}_n$ and build a custom search template $\tau_{\mathbf{h}_1,\ldots,\mathbf{h}_n}$. Go to point 11 if the number of neighbours is less than the chosen minimum
7. Scan the training image with the template $\tau_{\mathbf{h}_1,\ldots,\mathbf{h}_n}$ and using the up-scaling template to get the block-support values from the high-resolution training image.
8. Convert the high-order moments to Legendre cumulants and prompt the approximation of the ccdf for the current node.
9. Correct order relations in the fitted ccdf,
10. Write in the output grid the approximate ccdf values for all user-defined thresholds,
11. Visit another block and repeat from points 7 to 11 until all nodes have been visited.

The key components of this algorithm are the methods related to the inference of block-point high-order moments, the storage and retrieval of high-order moments and Legendre cumulants, and the approximation of the local ccdfs via Legendre polynomial series. These methods are detailed next.

3.1.1 Inference of point-block high-order moments

The definition of the block dimensions as multiples of the cell dimensions of the training image permits building an up-scaling template for obtaining block-support training values values (see Figure 1(b)). Let us represent the up-scaling template by $\nu_{\mathbf{h}_0,\mathbf{h}_1,\ldots,\mathbf{h}_{n_d}}$, with n_d as the number of training image cells that can be accommodated in individual output blocks, and $\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{n_d}$ as the vectors that radiate from the centre of the block to the training image cells, including the one located at $\mathbf{h}_0 = (0, 0, 0)$.

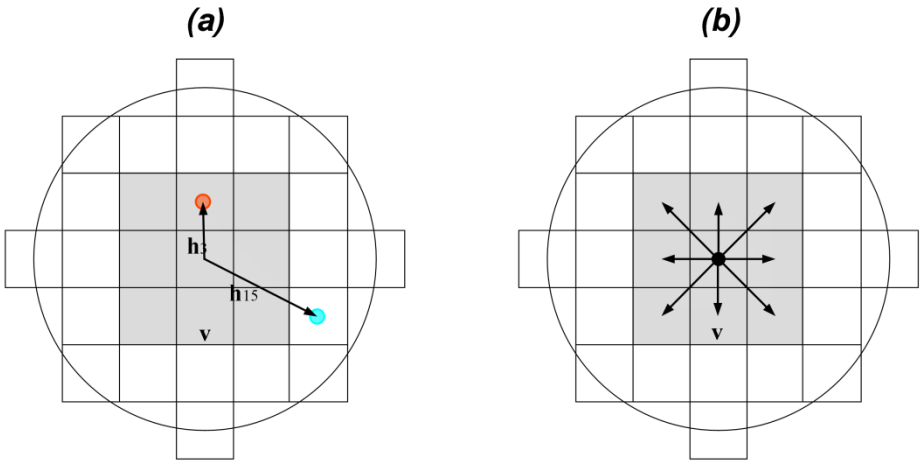

Figure 1: Moment *(a)* and up-scaling *(b)* templates for the inference of high-order moments. The grey area represents a block, the small squares the training image cells, the small circles the samples, the big circle a neighbourhood, and the arrows the vectors that form the templates.

The iteration path along the output grid is ordered, just like in classic estimation algorithms. At each point or block-support node of the output grid, the rasterized search neighbourhood located over the node collects the samples within its limits. The spatial configuration of these samples in relation to the centre of the node forms a moment template $\tau_{\mathbf{h}_1,\ldots,\mathbf{h}_n}$ (see Figure 1(a)).

These two templates, $\nu_{\mathbf{h}_0,\mathbf{h}_1,\ldots,\mathbf{h}_{n_d}}$ and $\tau_{\mathbf{h}_1,\ldots,\mathbf{h}_n}$, travel together along the training image. At each cell, the up-scaling template $\nu_{\mathbf{h}_0,\mathbf{h}_1,\ldots,\mathbf{h}_{n_d}}$ collects the values that are averaged to produce a value of the block-support value of $Z_{\mathbf{v}}$, while the template $\tau_{\mathbf{h}_1,\ldots,\mathbf{h}_n}$ collects the point-support values of $Z_1 \cdot \ldots \cdot Z_n$. The replicates thus obtained are used to estimate the point-block high-order moments (equation (4)) for the $\tau_{\mathbf{h}_1,\ldots,\mathbf{h}_n}$ template and all its sub-templates.

As new locations are visited along the iteration of the output grid, the moment template needs to be redefined according to the spatial configuration of the neighbouring samples. A new scanning of the training image is required if the new moment template contains at least one vector that is different from the templates used before. But only those moments that do not exist yet in the moment's container are calculated.

If block support data is not available, the point-block high-order moments are inferred only from a training image. In such case, the dataset can be used only for the inference point-point moments. In initial applications of high-order simulation, the data was combined with the training image (Mustapha and Dimitrakopoulos 2010b, Mustapha, Dimitrakopoulos and Chatterjee 2011, Machuca-Mory and Dimitrakopoulos 2012). From the practical point of view, however, the samples could distort the patterns informed by the training image, and, consequently, the inferred high-order moments. Moreover, doing so requires a training image that covers all the sampled area. Furthermore, from the conceptual point of view, the hard data should not be treated as pieces of the training image, as it was explained before (Section 2.2). Instead, two containers of moments can be allocated, one for those inferred from the training image and the other for those that can be inferred from the dataset. A container of combined high-order moments is then obtained by the weighted average of the moments that exist in both containers and the inclusion of the moments that exist only in the first container.

3.1.2 Storage of high-order moments and Legendre cumulants

The containers for the storage and retrieval of moments and Legendre cumulants were designed as vectors of maps. Each element of the vector corresponds to a map that contains only moments or Legendre cumulants of the same order ω, including $\omega = 0$. For example, if the approximation is performed with up to the 5th order of Legendre series, the containers for both, the moments and the Legendre cumulants will contain 6 maps and a moment such as $E\left[Z^2_{\mathbf{v}} \cdot Z^2_3 \cdot Z^1_{15}\right]$ will be stored in the 6th map. This data structure design, rather

than using a single map, is intended to improve the efficiency of storage and retrieval of the moments and Legendre cumulants.

Being spatial statistics, high-order moments are dependent of the spatial template used for their inference. Thus, it is important to keep the information about the spatial template $\tau_{\mathbf{h}_1,\ldots,\mathbf{h}_n}$ linked to the moment value. This is achieved by indexing the moments by integer arrays of size $\eta+1$, with η as the number of nodes of a rasterized circular or elliptical neighbourhood. In these map keys, the position of each element in the array corresponds to the position of a template point within a neighbourhood, and its value, to the power $w_\alpha \in \mathbb{N}^0, \alpha = 0, 1, \cdots, \eta$ with $w_0+w_1+\cdots+w_\eta = \omega$. For instance, Figure 2 shows a small circular neighbourhood of 3-units radius centred in a block v, the neighbouring rasterized nodes coded by their proximity to the centre, and two conditioning samples that coincide with nodes 3 and 15. The key for moment $E\left[Z_{\mathbf{v}}^2 \cdot Z_3^2 \cdot Z_{15}^1\right]$ in this neighbourhood is $[2,0,0,2,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0]$. An equivalent indexing method is used for the vector of Legendre cumulant maps.

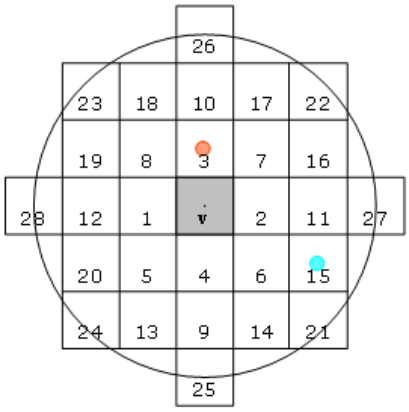

Figure 2: Rasterized neighbourhood with node ordering and two neighbouring samples.

3.1.3 Approximation of the local ccdfs and conditional moments

In expressions (12) and (14), the Legendre cumulants $L_{k_{\mathbf{v}},k_1,\ldots,k_n}$ provide information about the stationary multi-point spatial structure. The local conditioning is achieved by the inclusion of the n neighbouring sample values $z_1,\cdots,z_1$ in the Legendre polynomials $P_{k_1}(z_1),\cdots,P_{k_n}(z_n)$. At each node of the output grid, the high-order moments are recombined into Legendre cumulants according to (9) if the current moment template has not been yet found at previous nodes. Otherwise, the Legendre cumulant is just retrieved from the corresponding container. A recursive method implements the transformation of high-order moments into Legendre cumulants. Expressions (12) and (14) are also calculated using recursive algorithms.

The ccdfs fitted using Legendre or other polynomials will inevitably suffer of order relation problems. These are fixed by applying the same corrections applied to the ccdfs build using indicator methods (Goovaerts 1997, Deutsch and Journel 1998). However, in some cases, invalid ccdfs that cannot be fixed by those corrections may be produced. A suggested solution is to remove the invalid ccdfs from the output grid and to rebuild them by interpolating the thresholds of neighbouring valid ccdfs.

By selecting higher orders of approximation ω it is possible to fit more complex ccdfs and approximate conditional univariate moments of higher order. This also reduces the occurrence of order relation problems and invalid ccdfs. However as the order of the approximation increases, the computational costs explode.

The overall computational cost of this algorithm depends not only on the order of approximation and the number of neighbouring samples, but also on the number of dimensions (2D or 3D), the size of the search ellipsoid, the size of the training image, and the output grid and block sizes. Some ideas for improving the computational performance include the parallelization of processes, the use of graphical processing units, smaller but pattern-rich local training images, the implementation of multigrid search neighbouhoods and profiting of approximated symmetries of the type $E\left[Z_0^a \cdot Z_1^b\right] \approx E\left[Z_0^b \cdot Z_1^a\right]$.

3.2 Test example

Before attempting to run a first full case study, it is crucial to verify if the algorithm is actually fitting the local ccdfs accordingly to the theory presented in Section 2. Failing to do so would raise justified concerns about the validity of the proposed algorithm. The test example presented next is designed for such aim. Figure 3(a) shows a small grid formed by 4 blocks of 3×3 pixels cell size. Two samples are available in the grid. The 2D training image for extracting the point-point and point-block high-order moments appears in the right side of Figure 3. A search circle of 3 pixel radius, such as the one shown in Figure 2, was selected. By using this search neighbourhood, the cdf at block 1 (B1) is conditioned only by the high value sample, at B2, it is conditioned by both samples, no conditioning data is available for B3, and the cdf at B4 is conditioned only by the low value sample.

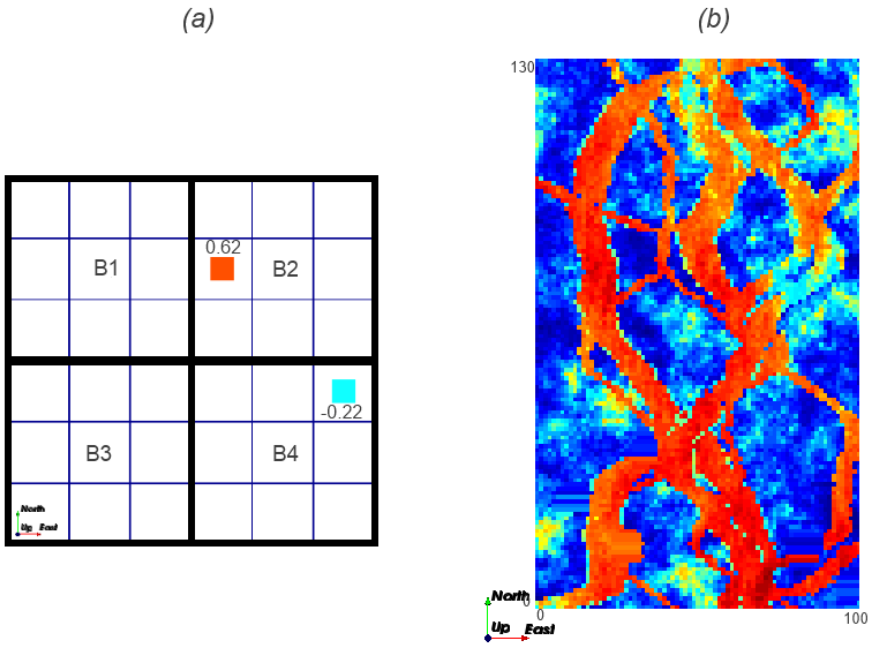

Figure 3: *(a)* Four 3×3 pixel blocks with two neighbouring samples and *(b)* a 2D training image.

Figure 4 shows the point and block scale fitted ccdfs at the centres of the 4 blocks. These results correspond to a maximum order of 4 for the ccdf approximation. All the required low and high-order moments were calculated from the training image and used in the Legendre approximation of the ccdfs. For B1 (Figure 4(a)), the presence of a high value neighbour bends the point and block-support ccdf towards higher probabilities for high values. In B2 (Figure 4(b)), the influence of both values results in a bimodal point-support ccdf, whereas the block-support ccdf shows a more uniform shape between the thresholds defined by the two sample values. This shape of the block-support ccdf is coherent with the averaging effect of up-scaled values. B3 (Figure 4(c)) has no conditioning neighbours, thus, the corresponding point and block-support ccdfs approach the respective global cdfs. Notice that the block-support global cdf is less bimodal than the point-support global cdf. B4 (Figure 4(d)) is conditioned only by the low value sample, thus, the point-support ccdf is shifted towards the lower values giving no probability of occurrence to high values. Contrastingly, the block-support ccdf, as it would be expected, is more resilient to the influence of a single low value. These results show that multi-point high-order moments can be used effectively to approximate conditional cdfs in a spatial context and for different supports.

Figure 5 shows the impact of the maximum order of approximation in the fitted point and block-supports ccdfs, in this case for location B2 of the example. At order 0, the Legendre cumulants only allow fitting a uniform cdf. At order 1 the influence of conditioning data on the fitted ccdfs is minimal. At orders 2 and 3 the influence of conditioning data is noticeable, but the fitted ccdfs lack of detail and may be biased. Only for

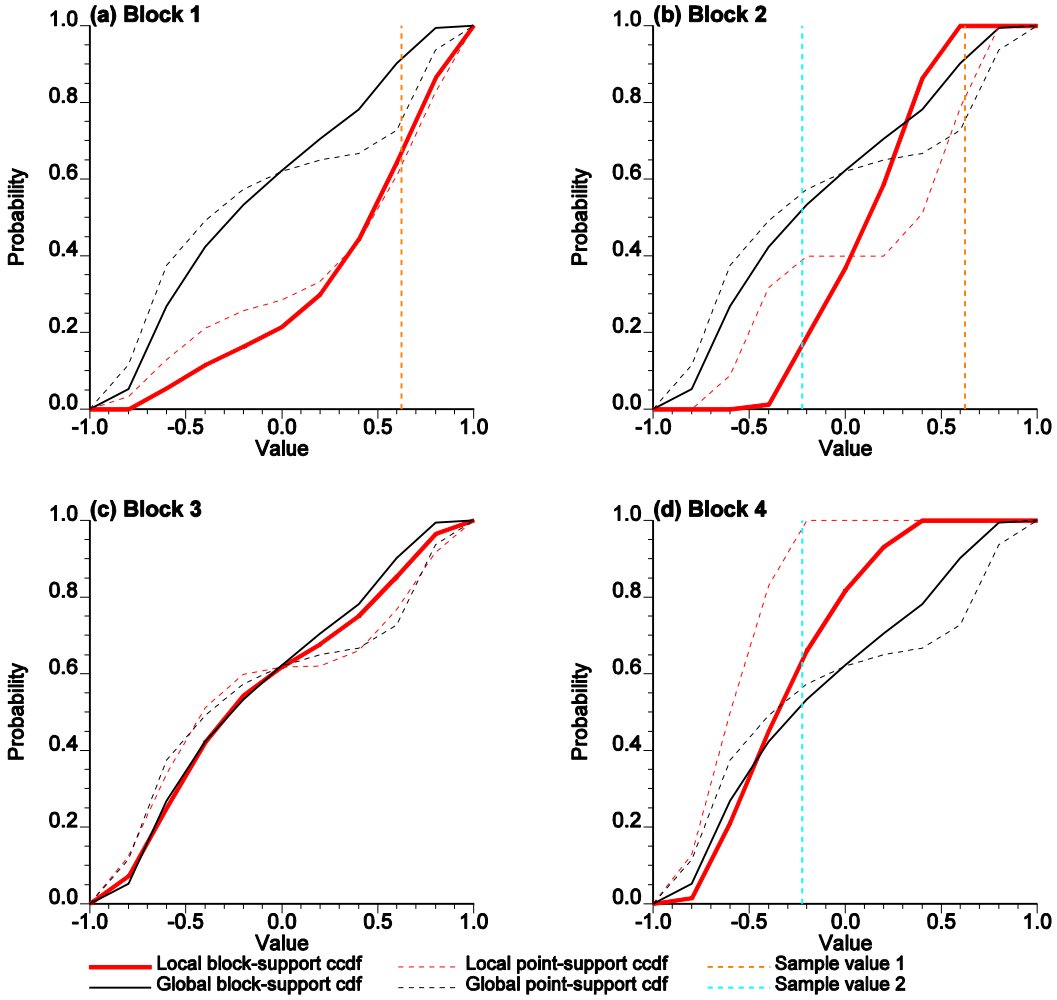

Figure 4: Point and block-support 4th-order Legendre ccdf approximation at 4 different locations and with 0 to 2 conditioning samples. The vertical dashed lines mark the values of the conditioning samples.

orders greater or equal to 4, the ccdf approximation by Legendre cumulants is able to fit complex asymmetric and multimodal ccdfs.

Initial works of simulation based on high-order statistics in a geostatistical context suggest that high-order moments and Legendre cumulants with exponents higher than 1 should be disregarded or approached by lower order moments of similar number of points (Mustapha and Dimitrakopoulos 2011). Doing so would reduce dramatically the computational cost of high-order simulation since only $n+2$ moments and Legendre cumulants would suffice, with n as number of conditioning samples (2 in this example).

335 high-order moments and the same number of Legendre cumulants were needed to fit all the point-support ccdfs in Figure 4. This number was 55 for the block-support ccdfs. Appendix C shows some of the high order moments (Table 1) and their corresponding Legendre cumulants (Table 2) required for the 4th order approximation of the point-support ccdf at B2. As these tables show, the moments of the form $E\left[Z_0^{w_0} \cdot Z_1^1\right]$, with w_0 close to 1, cannot be approximated by moments of the form $E\left[Z_0^1 \cdot Z_1^1\right]$. Neither can they be dismissed for $2 \leq w_0 \leq 4$. The same can be said for the Legendre cumulants.

These observations preclude the previously suggested approach for reducing the computational complexity of high-order simulation algorithms. Adopting it would result in a poor fitting of the ccdf, in the best case, or, in increased unwarranted fluctuations that yield to invalid ccdfs, in the worst.

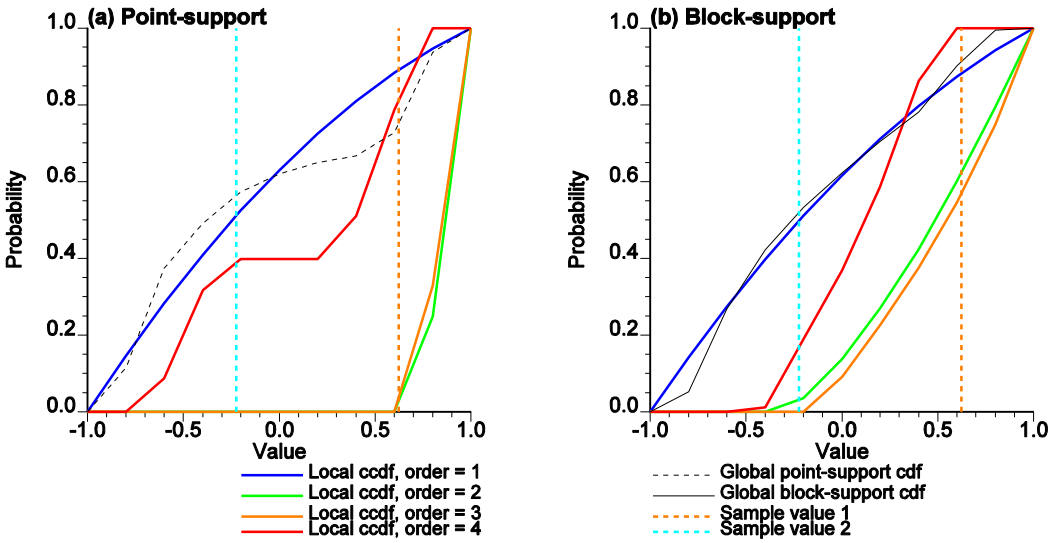

Figure 5:

4 Case study

The exhaustive 2D dataset used as the training image for this case study is the horizontal slice of the Stanford V reservoir (Mao and Journel 1999) that appears in Figure 3(b). 251 scattered samples were taken from the training image at random locations within a 100×100 units area. Sampling directly from an exhaustive TI is not feasible in real life applications. The TI is used only for obtaining the high-order moments, but it does not provide locally specific information. More than reconstructing the training image, this case study aims to model the space of uncertainty due to lack of complete locally specific information but counting with knowledge about the complex spatial structure of the attribute. Figure 6(a) shows the clearly bimodal cdf of the scattered dataset, while Figure 6(b) presents the sample locations. Note that the original values were transformed to the $[-1, 1]$ range to comply with the definition range of Legendre polynomials.

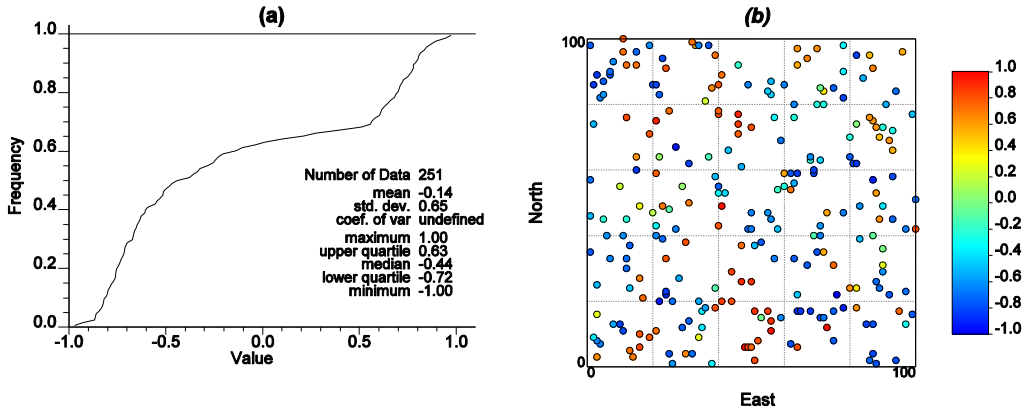

Figure 6: *(a)* Cumulative distribution of the scattered dataset, and *(b)* sample locations.

The output grid block size is 4×4 units. The local block-support ccdfs were approximated using the locally specific information provided by the dataset and the spatial structure information provided by the high-order statistics obtained from the training image. Three orders of approximation, from 3 to 6 were tried. Figure 7 shows the cumulative probability maps for three thresholds applied on the ccdfs obtained using different orders of approximation. The conditional cumulative probabilities in these maps are lower in regions of high values and reproduce the curvilinear features observed in the training image. The blank

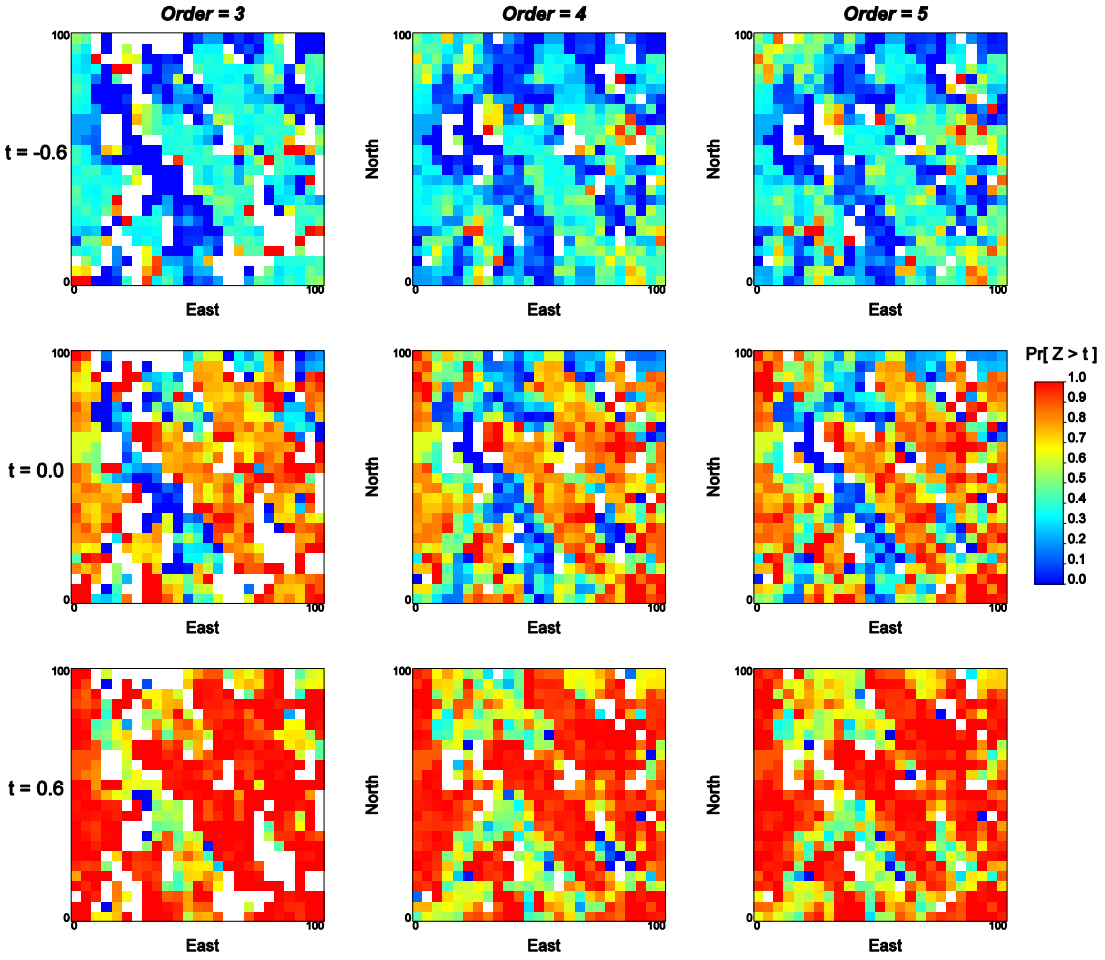

Figure 7:

blocks in these maps correspond to locations were the Legendre polynomials approximation yielded invalid ccdfs. At order 3 the number of invalid ccdfs is 119, at order 4 it is reduced to 46, and at order 5, it remains stable with 46 invalid ccdfs. However, increasing the order of the approximation results in a considerable increase of the computational time. At order 3, it took 8.6 minutes to complete the ccdfs approximation for all blocks in the grid; whereas at order 4 it took 27.4 minutes, and 76.1 minutes at order 5. The blank ccdfs can be fixed by interpolating the percentiles of the valid neighbouring ccdfs.

Alternatively, as Expression (14) shows, the Legendre cumulants approximation can also yield conditional 1-point low and high-order moments. Figure 8 shows the conditional moments maps from order 1 to 4. The blank areas in these maps correspond to blocks for which the conditional moments exceeded the valid range.

The probability fields were generated using sequential Gaussian conditional simulation. The conditioning data was created by replacing the original data values with the uniform distribution median value (0.5) (Goovaerts 2002). This dataset was transformed to Gaussian space where different realizations were generated. The realizations were back-transformed into uniform distributions defined within the interval $[0, 1]$. The variogram model for simulating the probability fields was obtained from the uniform transforms of the scattered dataset. The multiple probability fields were used to sample the local ccdfs obtained with the 5th order Legendre approximation. Figure 9 shows some of the resulting realizations. Notice the similitude of the patterns in the ccdf maps of Figure 7 with the patterns of simulated values in Figure 9.

The use of sequential Gaussian simulation to generate the probability fields may raise concerns about the preservation of the high-order spatial structure of the ccdfs in the realizations. Nonetheless, as the 3rd order

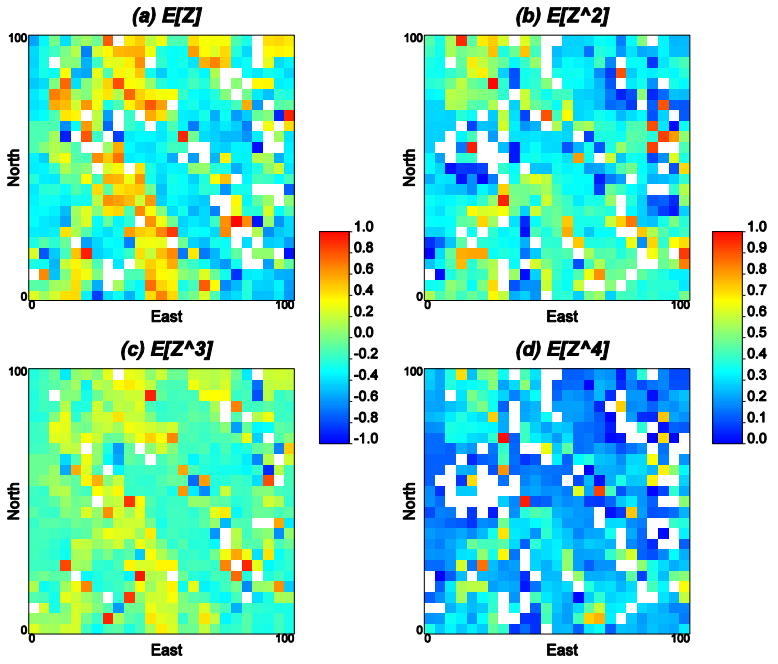

Figure 8: Approximated conditional 1st *(a)*, 2nd *(b)*, 3rd *(c)* and 4th *(d)* moments using Legendre cumulants.

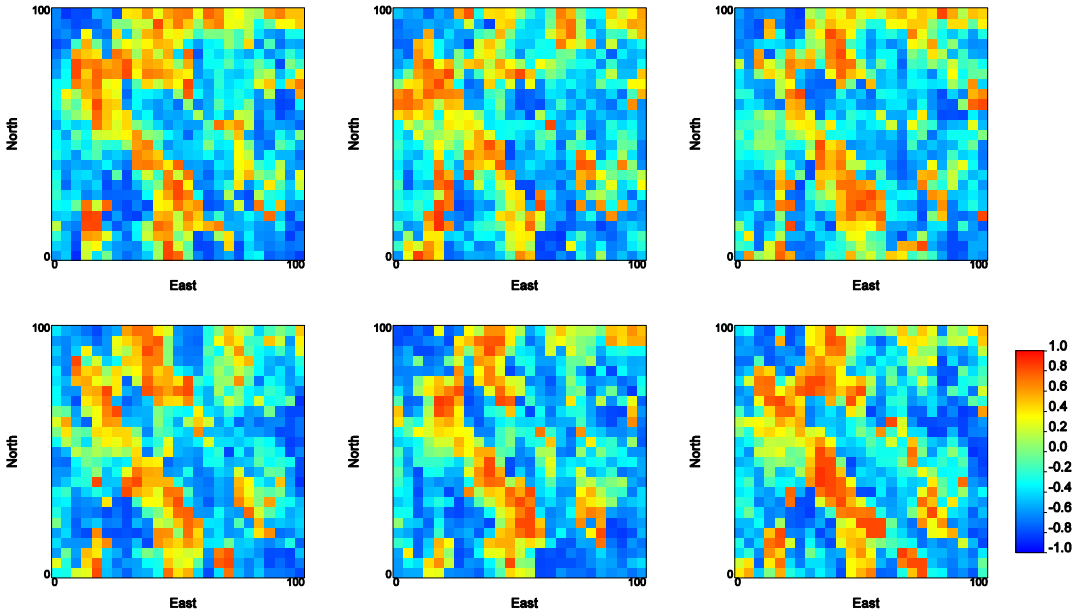

Figure 9: Block-support realizations obtained by probability sampling of the ccdfs.

cumulant maps of Figure 10 show, the individual realizations still contain complex spatial relations that are comparable to those found in the 3rd order cumulant maps of the dataset (Figure 11(a)) and, specially, of the Training Image (Figure 11(b)).

High-order block-support simulation not only allows for a reasonable reproduction of the block-support high-order spatial statistics, but also of the block-support low-order statistics and histogram. As Figure 12 shows, the global cdfs of the block-support realizations match the global cdf of the up-scaled Training Image within ergodic fluctuations.

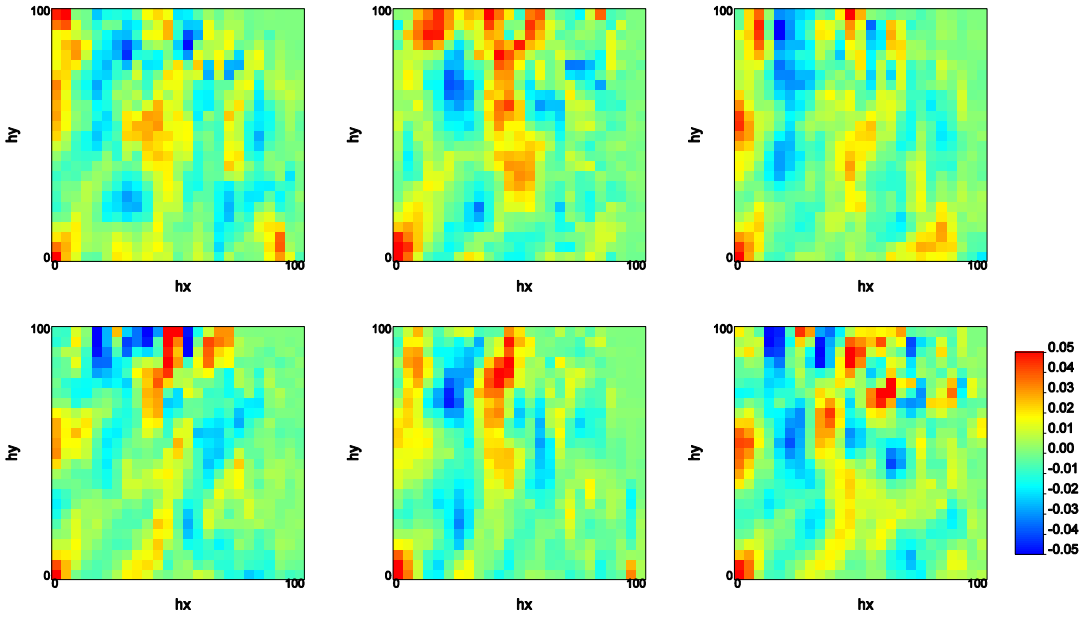

Figure 10: 3rd order cumulant maps for the realizations presented above.

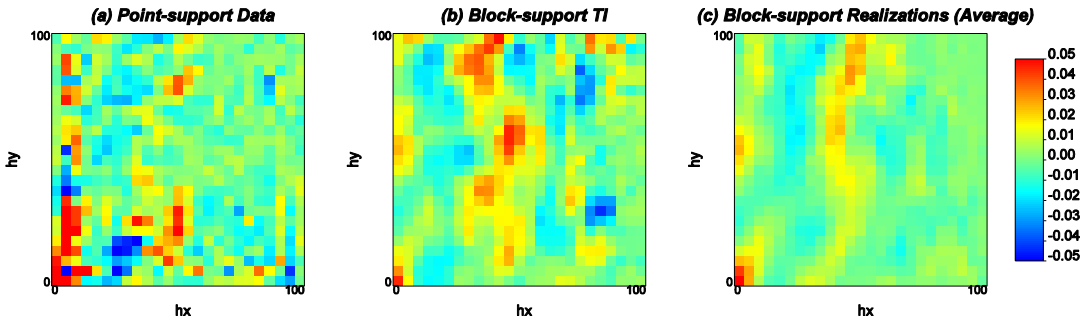

Figure 11: 3rd order cumulants maps obtained from the scattered samples *(a)*, the block-support training image *(b)* and by averaging the 3rd order cumulants of 10 block-support realizations *(c)*.

5 Discussion and conclusions

A methodology based on high-order spatial statistics for the local conditioning of the cumulative distributions functions of continuous variables at point and block support has been presented. The high-order spatial statistics carry information about non-linear spatial correlations into the orthogonal polynomial series that are used to approximate the conditional cdfs. Legendre polynomial series were used in the case study to approximate local conditional cdfs from a bimodal dataset. Other types of polynomials, such as Laguerre or Jacobi could be more appropriate if the data distribution is much skewed.

The results show that the resulting cdfs are effectively affected by the values and spatial configuration of the conditioning samples and the spatial pattern informed by the high-order moments extracted from a training image. The approximation block-support conditional cdfs from block support data and training images is achieved by the incorporation of cross block-point high-order statistics with up-scaled block-support values. This methodology is computationally expensive, thus, a reasonable option is to approximate the conditional cdfs only once and use different fields of correlated probabilities to sample them. The use of conditional sequential Gaussian simulation to generate the probability fields does not affect considerably the high-order spatial structure of the realizations produced by the proposed methodology, whose cumulant maps

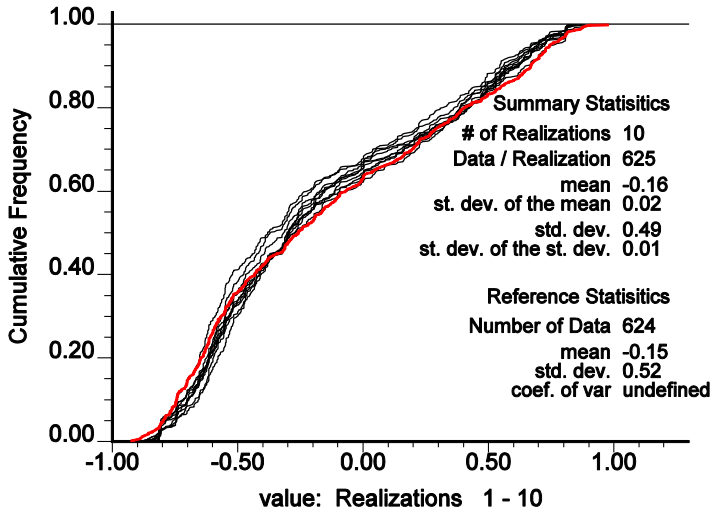

Figure 12: Global cdfs of the block-support realizations (black curves) and the up-scaled training image (red curve).

show reasonable agreement with the cumulant map obtained from the up-scaled training image. The global cdfs of the realizations also match closely the up-scaled training image cdf within ergodic fluctuations.

Besides the high demand of computational resources, the proposed methodology suffers of cases where the Legendre approximation yields invalid conditional cdfs. So far this problem is tackled by order-relation corrections or, for the extreme cases, by removing those cdfs and rebuilding them by interpolating the probability values of surrounding valid conditional cdfs. More detailed studies are needed to identify the conditions that difficult the approximation of the local conditional cdfs using series of orthogonal polynomials. Future work also include the development of leaner and faster algorithms for the implementation of conditional cdf approximation using orthogonal polynomials and high-order statistics at block and point-support scale. These algorithms are needed to implement high-order sequential simulation.

Appendix A – General form of the conditional cdf approximation using Legendre polynomials and high-order statistics

If the local cdf is conditioned by N point-support and M block-support data, it can be approximated by

$$F_Z\left(Z_0 \leq t \mid z_1, \ldots, z_N, z_{N+1}, \ldots z_M\right) \approx \frac{\left(\sum_{i_0}^{\omega} \cdots \sum_{i_N}^{i_{N-1}} \sum_{i_{N+1}}^{i_N} \cdots \sum_{i_M}^{i_{N+M+1}} L_{i_0-i_1,\ldots,i_N-i_{N+1},\ldots i_{N+M}} \prod_{k=1}^{N+M} P_{i_k-i_{k+1}}\left(z_k\right) \sum_{j=0}^{\left\lfloor \frac{i_0-i_1}{2} \right\rfloor} a_{j,i_0-i_1} \frac{\left(t^{i_0-i_1-2j+1}-(-1)^{i_0-i_1-2j+1}\right)}{i_0-i_1-2j+1}\right)}{\left(\sum_{i_0}^{\omega} \cdots \sum_{i_N}^{i_{N-1}} \sum_{i_{N+1}}^{i_N} \cdots \sum_{i_M}^{i_{N+M+1}} L_{i_0-i_1,\ldots,i_N-i_{N+1},\ldots i_{N+M}} \prod_{k=1}^{N+M} P_{i_k-i_{k+1}}\left(z_k\right) \sum_{j=0}^{\left\lfloor \frac{i_0-i_1}{2} \right\rfloor} a_{j,i_0-i_1} \frac{\left(1-(-1)^{i_0-i_1-2j+1}\right)}{i_0-i_1-2j+1}\right)}.$$

Where Z_0 can be a point or block-support random variable, z_1 to z_N are point-support data values, z_{N+1} to z_{N+M} are block support data values, the Legendre polynomials $P_{i_k-i_{k+1}}\left(z_k\right)$ are obtained by Expression (7),

and the point-block Legendre cumulants $L_{i_0-i_1,\ldots,i_N-i_{N+1},\ldots i_{N+M}}$ are given by

$$L_{i_0-i_1,\ldots,i_N-i_{N+1},\ldots i_{N+M}} = \frac{\prod_{k=0}^{M+N+1}\left(2\left(i_k-i_{k+1}\right)+1\right)}{2^{M+N+1}} \times \sum_{j_0=0}^{\lfloor (i_0-i_1)/2\rfloor} \cdots \sum_{J_{N+M}=0}^{\lfloor i_{N+M}/2\rfloor} E\left[Z_0^{i_0-i_1-2j_{\mathbf{0}}} \cdot \ldots \cdot Z_{M+N}^{M+N-2j_{\mathbf{0}}}\right] \prod_{l=0}^{N+M} a_{j_l,i_l-i_{l+1}}.$$

The coefficients $a_{j_l,i_l-i_{l+1}}$ are the same as those in Expression (7).

Appendix B – Some high-order moment and Legendre cumulant values

Table 1: Some point-point and point-block high-order moments calculated for the test example in Section 3.

High-order Moment	Point-point	Block-point
$E\left[Z_0^1 \cdot Z_1^1\right]$	0.334954	0.312679
$E\left[Z_0^2 \cdot Z_1^1\right]$	−0.0578584	−0.0576177
$E\left[Z_0^3 \cdot Z_1^1\right]$	0.185085	0.148872
$E\left[Z_0^1 \cdot Z_{14}^1\right]$	0.342108	0.295744
$E\left[Z_0^2 \cdot Z_{14}^1\right]$	−0.055852	−0.0571939
$E\left[Z_0^3 \cdot Z_{14}^1\right]$	0.188256	0.124784
$E\left[Z_0^1 \cdot Z_1^1 \cdot Z_{14}^1\right]$	−0.055968	−0.0545197
$E\left[Z_0^2 \cdot Z_1^1 \cdot Z_{14}^1\right]$	0.123183	0.123839

Table 2: Legendre cumulant values for the same orders and points as the high-order moments presented in the table above.

Legendre cumulant	Point-point	Block-point
$L_{1,1,0}$	0.37682325	0.351763875
$L_{2,1,0}$	−0.040103813	−0.039699938
$L_{3,1,0}$	−0.10426605	−0.254199225
$L_{1,0,1}$	0.3848715	0.2834145
$L_{2,0,1}$	−0.04125825	−0.039133313
$L_{3,0,1}$	−0.111621825	−0.1730568
$L_{1,1,1}$	−0.094446	−0.092001994
$L_{2,1,1}$	0.04864725	0.082740375

Appendix C – The HOEst program

HOEst stands for high-order estimation, and its objective is to estimate local non-Gaussian point and block-support ccdfs and conditional high-order moments over a regular grid and given a set of conditioning hard data. The resulting ccdfs can be used as an input to p-field simulation, or, alternatively, the more compact conditional moments output can be used to approximate the ccdfs.

The HOEst program was written in C++ and implemented as a SGeMS plugin (Remy 2009). It takes advantage of diverse classes and functions of SGeMS and the Geostatistical Template Library, GsTL (Remy 2001). HOEst is formed by three main classes: The HOEst class, the Legendre_ccdf class and the HO_moment_map class. HOEst class links the algorithm with SGeMS and the user interface with the algorithm. This class also executes the algorithm and manages the other two classes. The HO_moment_map class calculates the high-order moments and stores them in a vector of maps. The Legendre_ccdf class reads the

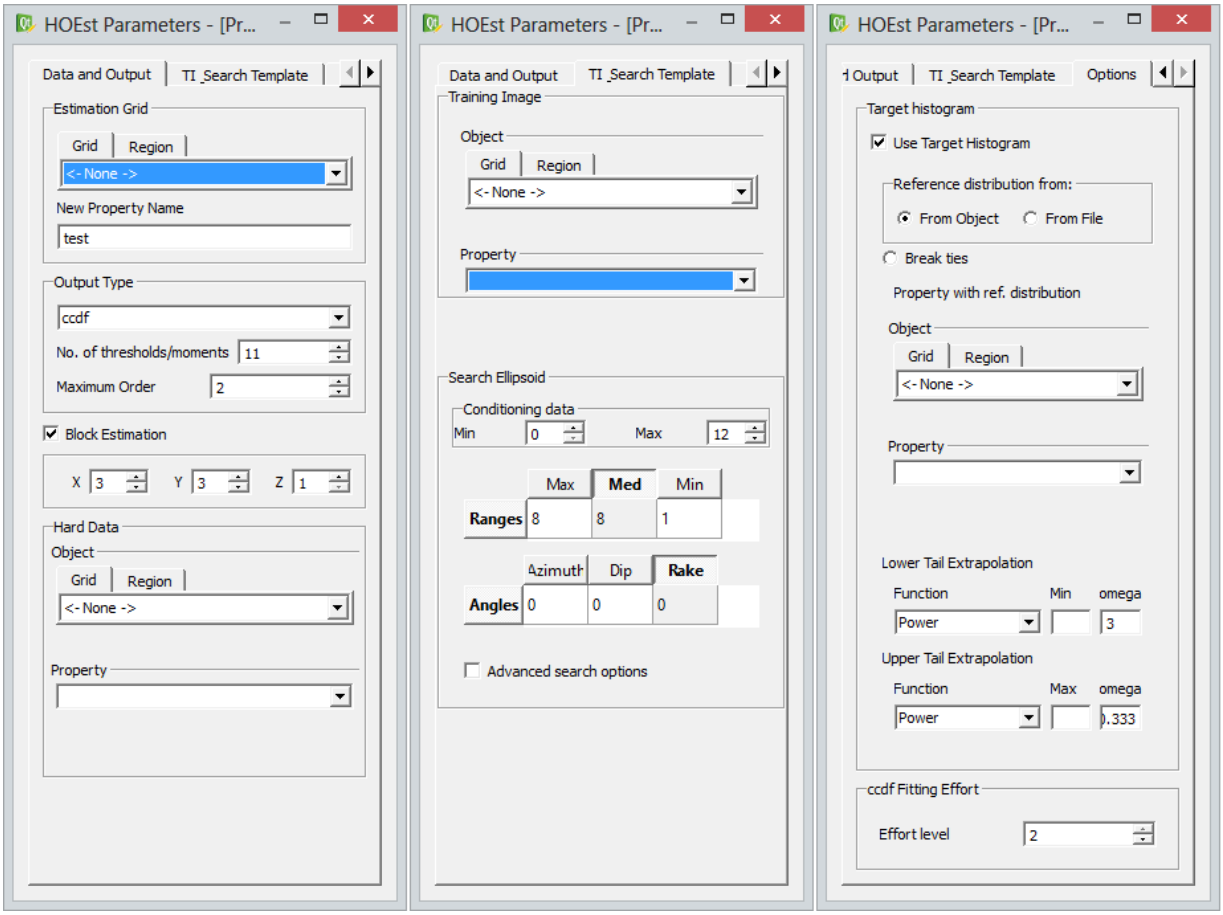

Figure 13: HOEst user interface.

high-order moments maps and uses them to, first, obtain and store the Legendre cumulants, and, second, to approximate the ccdf or the conditional moments. It returns the ccdf for a user-defined number of thresholds or, alternatively, the local conditional moment up to a user-defined order.

The HOEst plugin appears in the Estimation group of the SGeMS algorithms. Figure 13 shows the three pages of the user interface for HOEst. The first page contains the data and output parameters. The widgets for the output grid and the hard data input are similar to those found in the user interfaces of classic SGeMS algorithms. Two output types are permitted: ccdf and moments. In case the first output type is selected, the "N. of thresholds/moments" parameter specify the number the number of cut-offs for discretizing the output local ccdfs. In case the second output type is selected, each block in the output grid will contain the univariate conditional moments up to an order given by the "N. of thresholds/moments" parameter. The "Maximum Order" parameter specifies the order ω for the approximation of the local ccdfs and conditional moments. It is recommended not to select a maximum order larger than the "N. of thresholds/moments" parameter if the output type is conditional moments. If the "Block Estimation" box is unchecked, the output ccdfs or moments will be at point-support. Otherwise, the dimensions of the output blocks must be provided. These dimensions are multiples of the training image cell sizes.

The second page of the HOEst user interface is devoted to the training image and search ellipsoid. The widget for selecting the training image and the training property are similar to those for selecting the hard data grid and property. The search ellipsoid definition parameters follow the same convention as other SGeMS plugins. A detailed explanation of the SGeMS search ellipsoid definition and the advanced search options can be found in Chapter 2 of the book Applied Geostatistics with SGeMS (Remy, Boucher and Wu 2009).

Chapter 6 of the same book explains the entries of the widget for a non-parametric distribution, as the one that appears in the third page of the HOEst user interface. The target histogram is used to transform the original data and TI cdfs to a declustered cdf. If a declustered histogram is not provided, the hard data histogram will be used as the target histogram.

The last parameter in the third page of the HOEst user interface is the ccdf fitting effort level. The three effort levels from 0 to 2 correspond to:

0. Only high-order moments and Legendre cumulants of the form $E\left[Z_0^{0\vee1}\cdot Z_1^{0\vee1}\cdot\ldots\cdot Z_n^{0\vee1}\right]$ and $L_{k_{\mathbf{v}}=0\vee1,k_1=0\vee1,\cdots k_2=0\vee1}$, respectively, are inferred and used for approximating the ccdfs,
1. only high-order moments of the form $E\left[Z_0^{0\vee1}\cdot Z_1^{0\vee1}\cdot\ldots\cdot Z_n^{0\vee1}\right]$ are inferred. Moments of the form $E\left[Z_0^{>1}\cdot Z_1^{>1}\cdot\ldots\cdot Z_n^{>1}\right]$ are approximated by moments $E\left[Z_0^{1}\cdot Z_1^{1}\cdot\ldots\cdot Z_n^{1}\right]$. These approximated moments are used for obtaining the Legendre cumulants of any order and attempting to fit the ccdfs.

All high-order moments and Legendre cumulants required by expressions (9), (12) and (14) are obtained up to the user selected order ω and used for approximating the local ccdfs or conditional moments. This is the option that yields best results but also the most computationally demanding one.

It is important to remark that while only the hard data that falls within the boundaries of the output grid is used for the local cdf and moments conditioning, the training image grid does not need to be coincident with the output grid. This allows greater flexibility for building and choosing a training image. The only restriction is that the block dimensions of the output grid must be entire multiples of the training image's cell dimensions.

After the iteration of the entire output grid is finished, the program prints the maps of high-order moments and Legendre cumulants in two debugging text files.

References

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

Alabert, F. 1987a. The practice of fast conditional simulations through the LU decomposition of the covariance matrix. Mathematical Geology, 19, 369–386.

Alabert, F. & G. Massonnat. 1990. Heterogeneity in a complex turbiditic reservoir: Stochastic modeling of facies and petrophysical variability: SPE paper 20604. In 65th Annual Technical Conference and Exhibition, 775–790. New Orleans, Louisiana: Society of Petroleum Engineers.

Alabert, F.G. 1987b. Stochastic imaging of spatial distributions using hard and soft information. In Department of Applied Earth Sciences, Stanford University.

Boucher, A. 2009. Considering complex training images with search tree partitioning. Computers & Geosciences, 35, 1151–1158.

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

Chilès, J.-P. & P. Delfiner. 2012. Geostatistics: Modeling Spatial Uncertainty. New York: Wiley.

Christakos, G. 1992. Random Field Models in Earth Sciences. Academic Press.

David, M. 1977. Geostatistical Ore Reserve Estimation. Elsevier Scientific Pub. Co.

Deutsch, C. & A. Journel. 1998. GSLIB: Geostatistical Software Library and User's Guide. New York: Oxford University Press.

Dimitrakopoulos, R., H. Mustapha & E. Gloaguen. 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.

Emery, X. 2009. Change-of-support models and computer programs for direct block-support simulation. Computers & Geosciences, 35, 2047–2056.

Froidevaux, R. 1993. Probability field simulation. In Geostatistics Tróia'92, 73–83. Springer.

Glacken, M.I. 1996. Change of support by direct conditional block simulation. In Department of Geological and Environmental Sciences, Stanford University.

Godoy, M. 2003. The effective management of geological risk in long-term production scheduling of open pit mines. In W.H. Bryan Mining Geology Research Centre. The University of Queensland.

Godoy, M. & R. Dimitrakopoulos. 2004. Managing risk and waste mining in long-term production scheduling of open-pit mines. Society for Mining, Metallurgy, and Exploration Transactions, 316, 43–50.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. New York: Oxford University Press.

Goovaerts, P. 2002. Geostatistical modelling of spatial uncertainty using p-field simulation with conditional probability fields. International Journal of Geographical Information Science, 16, 167–178.

Hosny, K.M. 2007. Exact Legendre moment computation for gray level images. Pattern Recognition, 40, 3597–3605.

Isaaks, E.H. 1990. The application of Monte Carlo methods to the analysis of spatially correlated data. In Department of Applied Earth Sciences, Stanford University.

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

Journel, A.G. & C. Huijbregts. 1978. *Mining Geostatistics*. London, New York: Academic Press.

Journel, A.G. & E. Isaaks. 1984. Conditional indicator simulation: Application to a Saskatchewan uranium deposit. Mathematical Geology, 16, 685–718.

Journel, A.G. 1974. Geostatistics for conditional simulation of ore bodies. Econ Geol, 69, 673–687.

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

Journel, A.G. 2005. Beyond covariance: The advent of multiple-point geostatistics. Geostatistics Banff 2004. eds. O. Leuangthong & C.V. Deutsch, 225–233. Springer Netherlands.

Kendall, M.G., A. Stuart, J.K. Ord, S.F. Arnold & A. O'Hagan. 1994. Kendall's Advanced Theory of Statistics. London, New York: Edward Arnold; Halsted Press.

Lebedev, N.N. & R.A. Silverman. 1965. Special Functions and their Applications. Englewood Cliffs, N.J.: Prentice-Hall.

Liao, S.X. & M. Pawlak. 1996. On image analysis by moments. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 18, 254–266.

Machuca-Mory, D.F. & R. Dimitrakopoulos. 2012. Simulation of a structurally-controlled gold deposit using high-order statistics. In Ninth International Geostatistics Congress, eds. P. Abrahamsen, R. Hauge & O. Kolbjornsen. Oslo, Norway.

Mao, S. & A. Journel. 1999. Generation of a reference petrophysical and seismic 3D data set: The Stanford V reservoir. In Annual Meeting Report. Stanford, CA, USA: Stanford Center for Reservoir Forecasting.

Mariethoz, G., P. Renard & J. Straubhaar. 2010. The direct sampling method to perform multiple-point geostatistical simulations. Water Resources Research, 46, W11536.

Matheron, G. 1963. Les principes de la géostatistique. In CG, École des Mines de Paris. Fontainebleau

Matheron, G. 1970. La théorie des variables régionalisées et ses applications. In Cahiers du Centre de Morphologie Mathématique de Fontainebleau. École des Mines de Paris.

Mustapha, H. & R. Dimitrakopoulos. 2010a. Generalized Laguerre expansions of multivariate probability densities with moments. Computers & Mathematics with Applications, 60, 2178–2189.

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

Mustapha, H. & R. Dimitrakopoulos. 2010c. A new approach for geological pattern recognition using high-order spatial cumulants. Computers & Geosciences, 36, 313–334.

Mustapha, H. & R. Dimitrakopoulos. 2011. HOSIM: A high-order stochastic simulation algorithm for generating three-dimensional complex geological patterns. Computers & Geosciences, 37, 1242–1253.

Mustapha, H., R. Dimitrakopoulos & S. Chatterjee. 2011. Geologic heterogeneity representation using high-order spatial cumulants for subsurface flow and transport simulations. Water Resources Research, 47, W08536.

Myers, D. 1989. To be or not to be... stationary? That is the question. Mathematical Geology, 21, 347–362.

Osterholt, V. 2006. Simulation of ore deposit geology and an application at the Yandicoogina iron ore deposit. In School of Engineering. University of Queensland.

Pyrcz, M.J. & C.V. Deutsch. 2001. Two artifacts of probability field simulation. Mathematical Geology, 33, 775–799.

Remy, N. 2001. GsTl: The geostatistical template library in C++. Master of Science these, Department of Petroleum Engineering, Stanford University.

Remy, N., A. Boucher & J. Wu. 2009. Applied geostatistics with SGeMS a user's guide. Cambridge, UK; New York: Cambridge University Press.

Ripley, B.D. 1987. Stochastic Simulation. New York: Wiley.

Smith, P.J. 1995. A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. The American Statistician, 49, 217–218.

Srivastava, R. 1992. Reservoir characterization with probability field simulation. In SPE Annual Technical Conference and Exhibition.

Srivastava, R.M. & R. Froidevaux. 2005. Probability field simulation: A retrospective. In Geostatistics Banff 2004, 55–64. Springer.

Teague, M.R. 1980. Image analysis via the general theory of moments. J. Opt. Soc. Am, 70, 920–930.

Teh, C.H. & R.T. Chin. 1988. On image analysis by the methods of moments. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 10, 496–513.

Weisstein, E.W. 2006. Legendre Polynomial. http://mathworld.wolfram.com/LegendrePolynomial.html (last accessed June 2013).

Wu, J., A. Boucher & T. Zhang. 2008. A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. Computers & Geosciences, 34, 1863–1876.

Yap, P.-T. & R. Paramesran. 2005. An efficient method for the computation of Legendre moments. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27, 1996–2002.

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