A new approach for geological pattern recognition using high-order spatial cumulants

Hussein Mustapha, Roussos Dimitrakopoulos

Abstract

Spatially distributed natural phenomena represent complex non-linear and non-Gaussian systems. Currently, their spatial distributions are typically studied using second-order spatial statistical models, which are limiting considering the spatial complexity of natural phenomena such as geological applications. High-order geostatistics is a new area of research based on higher-order spatial connectivity measures, especially spatial cumulants as suitable for non-Gaussian and non-linear phenomena. This paper presents HOSC or High-order spatial cumulants, an algorithm for calculating spatial cumulants, including anisotropic experimental cumulants based on spatial templates. High-order cumulants are calculated on two- and three-dimensional synthetic training images so as to elaborate on their characteristics. Spatial cumulants up to and including the fifth-order are found to be relatively insensitive to the number of data used, and relatively small data sets are sufficient to provide cumulant maps. HOSC has been coded in FORTRAN 90 and is easily integrated to the S-GeMS open source platform.

1. Introduction

High-order cumulants are combinations of moment statistical parameters that allow the characterization of non-Gaussian random variables (Billinger and Rosenblatt, 1966), and may be seen as an extension of the well-known covariance function. They are critical contributors to non-Gaussian and non-linear modelling, where related developments include cumulants for signal filtering and deconvolution (Al-Smadi, 2004; Nikias and Petropulu, 1993; Sadler et al., 1995; Delopoulos and Giannakis, 1996; Dembélé and Favier, 1998; Zhang, 2005), or for estimating the gravitational evolution of the cosmic distribution function (Gaztanaga et al., 2000) and conditional cumulants and high-order statistics in the so-called high-precision astronomy (Bernardeau et al., 2002). A key justification for the use of cumulants is the wealth of information they contain compared to second-order statistical measures (Pan and Szapudi, 2005). The recently introduced concept of spatial cumulants (Mustapha and Dimitrakopoulos, 2008; Dimitrakopoulos et al., 2010) is important because spatial cumulants completely characterize non-linear and non-Gaussian stationary and ergodic spatial random fields. From an applied point of view, when considered in space, cumulants allow for large possible combinations of random point variables, suggesting a link to complex spatial connectivity patterns in a geological sense. This combination of multiple locations in space generates configurations that are considered capable of capturing substantially complex spatial configurations and patterns reflecting geologic characteristics of rock formations. As a result, cumulants are more enriched in the information they contain than the combinations of pairs of locations traditionally used for the same reason in the form of variograms or covariance functions (David, 1988, Section 2.7; Rendu and Ready, 1982; Dreiss and Johnson, 1989). Systematic definitions of spatial cumulants, including random variables, their moments and cumulants, non-Gaussian spatial random fields and their high-order spatial statistics, are given by Dimitrakopoulos et al. (2010), who also show examples of cumulant maps, for two- and three-dimensional training images, as well as a case study from a diamond bearing kimberlite pipe. In this paper, an efficient algorithm called High-Order Spatial Cumulants (HOSC) is detailed and used to compute experimental high-order spatial cumulants up to order five, for two- and three-dimensional fields, defined on both regular and irregular grids. Calculations are based on the notion of spatial templates that express the order of cumulant being calculated, direction and lag distances. The efficiency of HOSC is seen in the context of its ability to describe complex geological configurations. In the context of implementation, HOSC is optimal...
in terms of information stored and, consequently, it is optimal in time because only the necessary data have been used. It is important to note that simulation algorithms such as SNESIM (Strebelle, 2002) and FILTERSIM (Zhang et al., 2006) can be divided into two main parts. The first part is a pre-processing phase which contains the search of trees in SNESIM or the application of linear filters with a classification procedure in FILTERSIM. Incorporation of these trees and filters comes in a second part related to the simulation algorithm used. The work presented herein, i.e., the calculation of high-order cumulants calculation, can be used in a pre-processing step of a future simulation method, an avenue that is different than the one adopted in the present work.

The well-know problem cumulants problem or moments problem (Kendall and Stuart, 1977; Stuart and Ord, 1991) has been studied extensively from a theoretical point of view.

---

**Fig. 1.** Examples of experimental third-order cumulant templates for 2D regular grids. \( h_i (i=1, 2) \equiv |z(x) - z(x_i)| \). Note that, number of points in each template is equal to cumulant’s order, while number of directions is equal to cumulant’s order. –1.

**Fig. 2.** Examples of experimental fourth- and fifth-order cumulant templates for 2D regular grids.

**Fig. 3.** Examples of experimental fourth-order cumulant templates for 3D regular grids.
This problem concerns the determination of a probability distribution given its cumulants. Examples of solving this type of problem can be found in Edgeworth (1905, 1907); Daniels (1954); Welling (1999); Gaztanaga et al. (2000). Generalization of these methods to the multivariate case, cumulants represent well-defined properties of a consistent spatial random field model. At the same time, cumulants maps can be calculated in the same context of the search trees procedure used in SNESIM (Strebelle, 2002). However, the difference here is that these cumulant based methods incorporate much more information from a training image than information used by SNESIM. In addition, cumulant based methods can be employed for continuous variables. Accordingly, the order and the template’s shape to be used to approximate a conditional density at a given point can be determined by the neighbours of this point. In other words, the order is identical the number of neighbours plus one, and the template is defined by joining the point, to be simulated, to each point in the neighbours. It should be noted that, spatial cumulants, as developed in this paper, can be used as non-linear filters in a pre-processing step of a simulation algorithm and/or for image processing. However, these ideas need further work to determine, for example, the accurate orders and templates to be used, and how the different cumulant maps will communicate with one another so as to utilize the information available.

HOSC is a public domain algorithm easily incorporated in the SGEMS platform (Remy et al., 2009). HOSC may assist validating simulated realizations obtained by other simulation algorithms beyond the validation steps currently in practice. In the following sections, the definitions of cumulants are first reviewed. Then, the calculation of experimental high-order cumulants and description of a general algorithm follow, together with the details of program HOSC. Subsequently, HOSC is used to calculate cumulants up to fifth order for two- and three-dimensional images. Interpretations of cumulant characteristics versus spatial characteristics of the images are then provided. Finally, the availability of the program on the SGEMS platform (Remy et al., 2009) is discussed and conclusions follow.

2. Mathematical definitions of spatial cumulants

Let \((\Omega, \mathcal{A}, P)\) be a probability space and let \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) be a measurable space. A spatial random field \(Z(x), x \in \mathbb{R}^n\) is a family of random variables \(\{Z(x_1), Z(x_2), \ldots\}\) at locations \(x_1, x_2, \ldots\), where each random variable is defined on \((\Omega, \mathcal{A}, P)\) and takes values in \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\). The moments of order \(r\) of \(Z_i=Z(x_i)\) at the origin are defined by

\[
\text{Mom}[Z_i^r] = E[Z_i^r] = \frac{1}{\mathcal{F}_0} \frac{d^r}{d\Phi^r}(\Phi(\omega))_{\omega=0}.
\]

where \(E[\cdot]\) denotes the expectation operation, \(j^2=-1\) and \(\Phi(\omega)\) is the first characteristic function (Rosenblatt, 1985) given by \(\Phi(\omega) = E[e^{i\omega Z}] = \int_{-\infty}^{\infty} e^{i\omega u} dF_Z(u)\). The cumulants of order \(r\) are given by the derivative of order \(r\) of the second characteristic function \(\Psi(\omega) = \ln(\Phi(\omega))\), at \(\omega=0\):

\[
\text{Cum}[Z_i, \ldots, Z_j] = \frac{1}{\mathcal{F}_0} \frac{d^r}{d\Phi^r}(\Psi(\omega))_{\omega=0} = 0
\]

For random variables with known distribution, cumulants can be calculated analytically.

Denote by \(E_{j_1 \cdots j_n} = E(X_1, \ldots, X_n)\) and \(c_{j_1 \cdots j_n} = \text{cum}(X_1, \ldots, X_n)\) the \(n\)th-order moment and cumulant, respectively. Then, the translation of moments to cumulants, and vice versa, can be obtained recursively (Smith, 1995) as

\[
m_{j_1 \cdots j_n} = \sum_{j_1 = 0}^{l_1} \cdots \sum_{j_n = 0}^{l_n} \left( \prod_{i=1}^{n} \left( \frac{j_i - 1}{j_i} \right) \left( \frac{j_i - 2}{j_i - 1} \right) \cdots \left( \frac{j_i - 1}{j_i - 1} \right) \right) \times C_{j_1, j_1 \cdots j_n} m_{j_1 \cdots j_n}.
\]
Mom is defined as indexed in similarly, the \( r \) and \( c_i \) and, \( \text{Algorithm 1 for third-order (ND=2) cumulants calculation} \)

\[
\begin{align*}
1: & \quad \text{DO } n = 1 \text{, NN} :: \quad \text{Loop on the total number of nodes NN.} \\
2: & \quad \text{DO } i = 1 \text{, NLags_Dir[1]} :: \quad \text{Loop on the acceptable lags for } n \text{ in the direction 1.} \\
3: & \quad \text{DO } j = 1 \text{, NLags_Dir[2]} :: \quad \text{Loop on the acceptable lags for } n \text{ in the direction 2.} \\
   & \quad a. \quad \text{lag}_i = \text{NDL}[i,1] :: \quad \text{Read the lag number } i \text{ acceptable for } n \text{ in the direction 1.} \\
   & \quad b. \quad \text{lag}_j = \text{NDL}[j,2] :: \quad \text{Read the lag number } j \text{ acceptable for } n \text{ in the direction 2.} \\
   & \quad c. \quad m_i = \text{NDL}[i,3] :: \quad \text{Read the number of the second node acceptable for } n \text{ in the direction 1.} \\
   & \quad d. \quad m_j = \text{NDL}[j,3] :: \quad \text{Read the number of the second node acceptable for } n \text{ in the direction 2.} \\
   & \quad e. \quad N_{lag_i,lag_j} = N_{lag_i,lag_j}+1 :: \quad \text{Increase the number of triplets acceptable for lags } \text{lag}_i \text{ and } \text{lag}_j. \\
   & \quad f. \quad C^{lag_i,lag_j} = C^{lag_i,lag_j} + V[n] :: \quad \text{Update } C^{lag_i,lag_j} \text{ from (Eq.2); } V[n] \text{ is the value at the node } i. \\
4: & \quad \text{END DO } (n, i, j) \\
5: & \quad \text{DO } i = 1 \text{, NLag[1]} :: \quad \text{Loop on the lags of direction 1} \\
6: & \quad \text{DO } j = 1 \text{, NLag[2]} :: \quad \text{Loop on the lags of direction 2} \\
   & \quad g. \quad C^{T_i,j} = C^{T_i,j} / N_{ij} :: \quad \text{Average } C^{T_i,j} \\
7: & \quad \text{END DO } (i, j)
\end{align*}
\]

and,

\[
c_i_{i_1...i_n} = \sum_{j_1 = 0}^{i_{1}} \sum_{j_2 = 0}^{i_{2}} \sum_{j_{n-1} = 0}^{i_{n-1}} \sum_{j_n = 0}^{i_{n}} \binom{i_1}{j_1} \binom{i_2}{j_2} \cdots \binom{i_{n-1}}{j_{n-1}} \binom{i_n}{j_n} \\
x_m(j_1...j_{n-1}...j_n) c_{j_1...j_n},
\]

Assuming \( Z(x) \) is a zero-mean ergodic stationary random field indexed in \( \mathbb{R}^n \), then the rth-order moment of the random field \( Z(x) \) is defined as

\[
\text{Mom}[Z(x), Z(x+h_1), \ldots, Z(x+h_{r-1})] = E[Z(x)Z(x+h_1)\ldots Z(x+h_{r-1})].
\]

The moments depend only on the distances \( h_1, h_2, \ldots, h_{r-1} \). Similarly, the rth-order cumulants of \( Z(x) \) can be denoted as

\[
c_r(h_1, h_2, \ldots, h_{r-1}) = \text{Cum}[Z(x), Z(x+h_1), \ldots, Z(x+h_{r-1})].
\]

For example, the second-order cumulant of a non-centered random function \( Z(x) \) known as the covariance is given using Eq. (4) by

\[
c_2(h) = E[Z(x)Z(x+h)] - E[Z(x)]^2.
\]

Its third-order cumulant is given by

\[
c_3(h_1, h_2) = E[Z(x)Z(x+h_1)Z(x+h_2)].
\]

\[
- E[Z(x)]E[Z(x+h_1)Z(x+h_2)] \\
- E[Z(x)]E[Z(x+h_1)Z(x+h_3)] \\
- E[Z(x)]E[Z(x+h_2)Z(x+h_3)] + 2E[Z(x)]^3.
\]

The moments depend only on the distances \( h_1, h_2, \ldots, h_{r-1} \) as

\[
E[Z(x)] = \int_{\mathbb{R}^n} Z(x) f(x) dx,
\]

where \( f(x) \) is the probability density function of \( x \).
The cumulants are invariant to additive constants; thus, if a given process $Z(x)$ is not zero-mean, its cumulants can be computed as the cumulants of $Z(x) - E(Z(x))$ (Nikias and Petropulu, 1993). It can be computationally convenient to consider zero-mean random functions as some of the terms vanish. For example, the second-order cumulant of a zero mean random function $Z(x)$, known as centered covariance, is given by

$$c_2^z(h) = E[Z(x)Z(x+h)],$$

the third-order cumulant is defined as

$$c_3^z(h_1, h_2) = E[Z(x)Z(x+h_1)Z(x+h_2)],$$

the fourth-order cumulant is

$$c_4^z(h_1, h_2, h_3) = E[Z(x)Z(x+h_1)Z(x+h_2)Z(x+h_3)],$$

and the fifth-order cumulant is

$$c_5^z(h_1, h_2, h_3, h_4) = E[Z(x)Z(x+h_1)Z(x+h_2)Z(x+h_3)Z(x+h_4)].$$

The cumulants of order higher than three of a zero mean random functions are related to their moments of lower orders and a combination of their moments of order two.

### 3. Experimental calculation of high-order spatial cumulants

The spatial cumulants expressions in Eqs. (9)–(12) are experimentally evaluated by specifying lags (i.e., distances) and their directions. For example, in Eq. (12), lags $h_1$, $h_2$, and $h_4$ denote four lags supported by four directions and the point $x+h_i$ (in nD-space, $n=1$, 2, or 3) is implicitly calculated from $x+1d_i$ where the unit vector $d_i$ determines the direction of $h_i$. In this context, we first select a set of $n$ directions for a $(n+1)$-order cumulants calculation. Then, different sets of $n$-lags are chosen. In each of these sets, the lag $i$ is in the direction $i$. Consider a set of $n$ directions $(d_i, i=1, n)$ that are supported by the direction angles $(a_1, a_2, ..., a_n)$. For a set of $n$ lags $(h_i, i=1, n)$, the associated spatial template of order $(n+1)$ is defined (considering a spatial location $x$ as a reference) as

$$T_{n+1}^{h_1, h_2, ..., h_n} = \{x, x+h_1, x+h_2, ..., x+h_n\},$$

such that the points $\{x, x+h_i, i=1, n\}$ are a set of the original points distribution.

For example, the third-order cumulant Eq. (10) with the given template $T_3^{h_1, h_2}$ is computed from

$$C_{h_1, h_2}^{T_3^{h_1, h_2}} = \frac{1}{N_{h_1, h_2}} \sum_{i=1}^{N_{h_1, h_2}} Z(x_i)Z(x_i+h_1)Z(x_i+h_2), \quad \{x_i, x_i+h_1, x_i+h_2\} \in T_3^{h_1, h_2},$$

and the fourth-order cumulant Eq. (11) with the given template $T_4^{h_1, h_2, h_3}$ is calculated as

$$C_{h_1, h_2, h_3}^{T_4^{h_1, h_2, h_3}} = \frac{1}{N_{h_1, h_2, h_3}} \sum_{i=1}^{N_{h_1, h_2, h_3}} Z(x_i)Z(x_i+h_1)Z(x_i+h_2)Z(x_i+h_3),$$

$$- \frac{1}{(N_{h_1, h_2, h_3})^2} \left[ \sum_{i=1}^{N_{h_1, h_2, h_3}} Z(x_i)Z(x_i+h_1)Z(x_i+h_2) \right] \left[ \sum_{i=1}^{N_{h_1, h_2, h_3}} Z(x_i)Z(x_i+h_2)Z(x_i+h_3) \right]$$

$$- \frac{1}{(N_{h_1, h_2, h_3})^2} \left[ \sum_{i=1}^{N_{h_1, h_2, h_3}} Z(x_i)Z(x_i+h_1)Z(x_i+h_3) \right] \left[ \sum_{i=1}^{N_{h_1, h_2, h_3}} Z(x_i)Z(x_i+h_1)Z(x_i+h_2) \right]$$

$$x_{i}(x_i, x_i+h_1, x_i+h_2, x_i+h_3) \in T_4^{h_1, h_2, h_3},$$

where $N_{h_1, h_2}$ and $N_{h_1, h_2, h_3}$ are the number of elements of $T_3^{h_1, h_2}$ and $T_4^{h_1, h_2, h_3}$, respectively. For more details, the reader is referred.
to Mustapha and Dimitrakopoulos (2008). Looking back at Eqs. (14) and (15), expressions for fifth-order cumulants can be obtained in a similar fashion. The cumulants maps are obtained by evaluating the experimental expressions for some sets of lags. Each set of lags (for example, three lags in fourth-order or four lags for fifth-order) provides one cumulant value.

For two-dimensional training images, four directional experimental cumulants are used for the third-order cumulants. As detailed above, the template directions are first needed.

Table 3
Parameters for HOSC program.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># DIMENSION OF THE PROBLEM</td>
<td>1=1D, 2=2D and 3=3D</td>
</tr>
<tr>
<td># REGULAR GRID OR NO</td>
<td>1 = YES, 0 = NO.</td>
</tr>
<tr>
<td># IF REGULAR GRID PROVIDE:NX,NY,NZ,DIMX,DIMY,DIMZ</td>
<td>0</td>
</tr>
<tr>
<td># ORDER OF CUMULANTS</td>
<td>1, 2, 3, 5</td>
</tr>
<tr>
<td># THE NUMBER OF DIRECTIONS</td>
<td>1</td>
</tr>
<tr>
<td># CONSIDRED IS EQUAL TO (ORDER OF CUMULANTS - 1)</td>
<td>0</td>
</tr>
<tr>
<td># NUMBER OF LAGS IN EACH DIRECTION:</td>
<td>35 41 53 510 600 780</td>
</tr>
<tr>
<td># NLAG_DIR1, NLAG_DIR2, ..</td>
<td>35 41 53 53</td>
</tr>
<tr>
<td># LAG SEPARATION IN EACH DIRECTION</td>
<td>15 15 15 15</td>
</tr>
<tr>
<td># LAG TOLERANCE: A FRACTION OF THE LAG SEPARATION</td>
<td>0.1 0.1 0.1 0.1</td>
</tr>
<tr>
<td>AZIMUTH,AZIMUTH_TOL,BANDH,DIP,DIP_TOL,BANDV</td>
<td>90 5 0 0 0 0</td>
</tr>
<tr>
<td>0 5 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>180 5 1 90 5 1</td>
<td></td>
</tr>
<tr>
<td>180 5 1 -90 5 1</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 7. Fourth-order cumulant map in (1) and (2) are two-dimensional cross-sections along z. Light blue color in upper part of (2) is a non-value region, i.e., no replicates in training image correspond to considered templates.
Figs 1(1–4) show four possible templates directions that are: \( \{\vec{x}, \vec{x}\} \), \( \{\vec{y}, \vec{y}\} \), \( \{45^\circ\text{-axis}, 45^\circ\text{-axis}\} \) or simply \( \{45,45\} \), and \( \{\vec{x}, \vec{y}\} \). The 45\(^\circ\)-axis is obtained from \( \vec{x} + \vec{y} \). Any other combination or transformation of a basis vector \( \{\vec{x} \text{ or } \vec{y}\} \) provides a possible direction. For example, the 45\(^\circ\) rotations of \( \{\vec{x}, \vec{y}\} \) provide the template directions in Fig. 1(5). The fourth- and fifth-order cumulants consist of the use of three and four directions, respectively. As for the third-order, directions are linear combinations or transformations of the basis vectors \( \{\vec{x}, \vec{y}\} \). See for example Fig. 2. For three-dimensional images the basis vectors are \( \{\vec{x}, \vec{y}, \vec{z}\} \). Similarly, directions fourth- and fifth-order are selected from combinations or transformations of these basis vectors. Some examples are presented in Figs. 3 and 4.

The algorithms developed in HOSC are conceptualized for irregular grids. In addition, HOSC adapts its algorithm for regular grids and treats them as particular cases of irregular grids. Templates presented above are convenient for regular grids. For irregular grids, tolerances in distances, angles and bands are incorporated as shown in Fig. 5. In Fig. 5, \( h_1 \) and \( h_2 \) are the lags distances with \( T_{h_1} \) and \( T_{h_2} \), tolerance distances respectively; \( a_1 \) and \( a_2 \) are the angles with tolerances \( T_{a_1} \) and \( T_{a_2} \), respectively and \( (i,j) \) is the basis of the Cartesian coordinate system. Other \( n \) \((n \geq 1)\) directions can be added by specifying \( n \) supplementary angles with the corresponding tolerances for \((n+3)\)-order cumulant calculation.

3.1. The general algorithm

The main subroutines in HOSC are developed for searching pairs and calculating high-order cumulants. The program is written with the standard Fortran 90 language, and is organized by using MODULES (classes). The data structures are described next.

HOSC reduces the computational costs by using the following idea. Fig. 6(1) shows an example of a data distribution. The third-order cumulant calculation using the two-directional templates in Fig. 6(2) starts with searching pairs as shown in Fig. 6(3). This figure shows the node \( A \) which is the intersection between two lags. As mentioned before, the calculation of \( T_{h_1;h_2}^{3} \), in Eq. (14), uses only the triplet \( \{A, A+h_1, A+h_2\} \) and then, all the other remaining pairs are inactive. A pair of lags is considered inactive if there is no common extremity between them. The HOSC algorithm stores and manipulates only the active pairs. Needs link—what is the staff that follows.

For the node number \( n \), if there exists a node number \( m \) such that the distance between nodes \( n \) and \( m \) is equal to some given
lag $l$ in a given direction $D$, then $m$, $l$ and $D$ are acceptable for the node $n$. Using this property, we fill for each node $n$ the class Node_Dir_Lag. This class contains

1. $\text{NLags}_\text{Dir}[ND]$: the number of the acceptable lags for the node $n$ in each direction. $ND$ is total number of directions considered.

For each direction $D$:

2. $\text{NLD}[1]$: a vector of the acceptable lags for the node $n$. This vector is defined as:
   a. $\text{NLD}[1]: \text{NLags}_\text{Dir}[D], 1 \!= l$
   b. $\text{NLD}[1]: \text{NLags}_\text{Dir}[D], 2 \!= n$
   c. $\text{NLD}[1]: \text{NLags}_\text{Dir}[D], 3 \!= m$

The definition of “active node” is introduced to only use the necessary information. Consider for example a point $p$ and the $x!y!g$-template third-order cumulant. If for any lags $lag_x$ and $lag_y$, Fig. 9. Interaction between three blocks in (1) using third-, fourth- and fifth-order cumulants in (2), (3) and (4), respectively.
Fig. 10. Interaction between three blocks using third-order cumulants (1) and using fourth- and fifth-order cumulants, respectively (2).

Fig. 11. (1) original image, (2) covariance map, (3) \((\bar{x}, \bar{y})\) third-order cumulant map, (4) 2D cross-section of \((\bar{x}, \bar{y})\) fourth-order cumulant map at \(-x=0\) and (4) 2D cross-section from \((\bar{x}, \bar{y}, -\bar{x}, -\bar{y})\) templates cumulant map at \(-x=0\) and \(y=0\).
there exist points \( p_x \) and \( p_y \) such that \( \text{lag}_x - \text{tolerance}_x < \text{dist}(p_x, p) < \text{lag}_x + \text{tolerance}_x \) and \( \text{lag}_y - \text{tolerance}_y < \text{dist}(p_y, p) < \text{lag}_y + \text{tolerance}_y \), where tolerances \( x \) and \( y \) are the lag tolerances, then the point \( p \) is considered as active. This approach facilitates the calculation of \( \text{cum}(\text{lag}_x, \text{lag}_y) \) using Eq. (13). It is possible that a given point \( p \) does not satisfy this property for all the lags considered and then this node will not be stored with the related information, and consequently it will not be considered for the evaluation of the cumulant maps.

The calculation algorithm of Eq. (9) can be generalized to compute the first term in the right-hand side of Eqs. (10)–(12). This algorithm computes the average of two-point (one direction) multiplications in Eq. (9). Then, this is extended to three points (two directions in Eq. (10)), four points (three directions in Eq. (11)) and five points (four directions in Eq. (12)) multiplications average. The remaining terms in Eqs. (10)–(12) are obtained from multiplications of second-order cumulant terms. These terms can be evaluated using the similar calculation algorithm of Eq. (9). Then, third-, fourth- and fifth-order cumulants are calculated based on generalizations of the second-order cumulants calculation algorithm. See for example, the third-order cumulants calculation algorithm in Table 1.

### 3.2. Program parameters

The HOSC program is suitable for experimenting with high-order cumulants to characterize complex patterns. The parameters required for the program are listed below and parameters files are shown in Tables 2 and 3:

**Grid file description (Table 2)**

This file has the same structure for regular and irregular grids and is described as

- Line 2: the total number of nodes (\( NN \)) used in the geometry description.
- Line 4: the number of variables (\( NVAR \)). If for example \( NVAR=4 \), then we have \( X, Y, Z \) and value at the node \( (X, Y, Z) \).
- Lines 7–22: the nodes information. For each node \( p (p=1, ..., NN) \), we provide its coordinates \( X_p, Y_p, Z_p \) and, \( V[p] \), the value at \( p \). For 2D problem \( Z_p \) is zero.

**Parameters file description (Table 3)**

- Line 2: the problem dimension \( Dim \). \( Dim=k \) for \( k \)-dimensional problem.
- Line 4: the data is on a regular or irregular grid. If the value is 1, then we have a regular grid, otherwise it is an irregular grid.
- Lines 7–11: the order of cumulant \( (Norder) \). \( Norder=1, 2, 3, ... \)

---

Fig. 12. Interpretations of high positive anomalies in third-order map (Fig. 11 (2)). (1.a)–(8.a) explains, respectively, high anomalies in (1.b)–(8.b) from interactions between black zones in red rectangles.
Line 14: the number of lags in each direction.
Line 16: the lag distance separation in each direction.
Line 18: the lag tolerances.
Line 21–24: the definition of the directions. The number of directions considered is \( ND = N_{\text{order}} - 1 \). For each direction \( D \), we provide: \( \text{azimuth}_D \), \( \text{tol}_{\text{azimuth}}_D \), \( \text{bwh}_D \), \( \text{dip}_D \), \( \text{tol}_{\text{dip}}_D \), \( \text{bwd}_D \), where
- \( \text{azimuth}_D \): is the azimuth angle for the direction \( D \).
- \( \text{tol}_{\text{azimuth}}_D \): is the half-window azimuth tolerance for the direction \( D \).
- \( \text{bwh}_D \): is the azimuth bandwidth for the direction \( D \).
- \( \text{dip}_D \): is the dip angle for direction \( D \).
- \( \text{tol}_{\text{dip}}_D \): is the half-window dip tolerance for the direction \( D \).
- \( \text{bwd}_D \): is the dip bandwidth for direction \( D \).

4. Examples and interpretation of high-order cumulants of two- and three-dimensional images

This section presents examples of calculating third-, fourth-, and fifth-order experimental spatial cumulants using HOSC. The cumulants are calculated for both 2D and 3D images. Results are interpreted so as to demonstrate the use of spatial cumulants for geological pattern recognition. The \( n \)-order (here, \( n = 2, 3, 4, 5 \)) cumulant maps are plotted in \( (n - 1) \)-dimensional space. For the presentation of three- and four-dimensional cumulant maps, the notion of cross-section is first introduced. In terms of geometry, a cross-section is the intersection of a body in 2-dimensional space with a line, or of a body in 3-dimensional space with a plane, etc. Similarly for the \( n \)-dimensional objects (\( n > 3 \)), the cross-sections reduce the dimensionality to \( (n - 1) \). This cross-section principle can be applied \( m (m < n) \) times successively to reduce the dimensionality to \( (n - m) \). For example, a 4D object, A, is reduced to 3D object, B, for the first 3D cross-section; and a 2D object, C, is obtained using another 2D cross-section of B. Here, the dimension is reduced by 2 using 2 successive cross-sections.

The fourth-order cumulant maps are 3D images. Cross-sections of these maps provide 2D images and help to clarify the results. Similarly, 2D images are obtained by 2 successive cross-sections from...
the fifth-order cumulant map. For example, a fifth-order \( \mathcal{C}_{5}(X, Y, Z, -Z) \) cumulant map is in four-dimensional space. A 3D cross-section, \( S \), can be obtained (\( S \) is in \( (X, Y, Z) \) space) by \( z=0' \) (a cut at the fourth direction). From \( S \), a 2D cross-section, \( M \), (\( M \) is in \( (X, Y) \) plane) can be searched by \( z=0' \) (a cut at the third direction). For example, Fig. 7(1) shows the \( (X, Y, Z) \) fourth-order map and cross-sections along \( z \) in Fig. 7(2). These sections can be used separately as shown in Fig. 8. Mathematically, these 2D cross-sections are the restrictions of the cumulant functions Eqs. (11) and (12) to a two-dimensional vector space. For example, the restricted fourth- and fifth-order cumulant functions to the \( (X, Y) \) vector space are given by

\[
c_4^{2}(X, Y) = c_4^{2}(h_1, h_2, 0) = E[Z(x)Z(x)Z(x+h_1)Z(x+h_2)] - 2c_2^{2}(h_1)c_2^{2}(h_2) - c_2^{2}(0)c_2^{2}(h_1-h_2)
\]

and,

\[
c_5^{3}(X, Y) = c_5^{3}(h_1, h_2, 0, 0) = E[Z(x)Z(x)Z(x+h_1)Z(x+h_2)] - c_3^{3}(h_2-h_1, -h_1) + c_3^{3}(h_2-h_1, h_1) + c_3^{3}(h_1, h_2) - 2c_3^{3}(h_1, h_2, 0) - c_2^{2}(h_2)c_3^{3}(h_1, h_1) + 2c_3^{3}(h_1, 0) - c_2^{2}(h_2-h_1)c_3^{3}(0, 0)
\]

Eqs. (16) and (17) can be updated, depending on the position of the cross-section. Note that the restriction of an \( n \)-order cumulant function to a \( (n-1) \)-dimensional space is different from the \( (n-1) \)-order cumulant function.

For 3D images, the third-order cumulant map is created based on two directions. This map is obtained by moving and averaging over the pipe, using different templates which are oriented in two specific directions. For each couple of lags \( (h_1, h_2) \), the pipe is scanned to search the template set \( T_n \) in Eq. (13), and evaluating the cumulant at the point \( (h_1, h_2) \) as described in Eq. (14). The same approach is used for fourth- (three directions) and fifth-order (four directions) cumulant maps. The covariance is a measure of the periodicity between pairs of points separated by given distances (Rani and Mitra, 1994). Similarly, the higher-order cumulant is also a measure of periodicity, but in the direction of the symmetry axis of template used, that is, the multiple point symmetry.

4.1. Results and interpretations

This subsection provides an intuitive approach to the interpreting the cumulant maps. For binary images, the high values of the third-order cumulant Eq. (9) reflect the frequency that the...
three points of the template used are in the black zones (i.e., they have a value of 1). Then, the interpretation of the high intensity anomalies is based on the interactions between any three black points, or for more general three black zones. For a zero lag in one of the two directions, the third-order provides the interactions between two black zones distributed in the second direction. Locally and for one black zone, the third-order cumulant map reflects its dimensions using lags that are smaller than its size. In addition to the third-order properties and for non-zero lags, the fourth- and fifth-order cumulants consider not only the interactions between any four or five black zones, but also the cross-relations between these zones as shown in Eqs. (16) and (17). In the following parts of this paragraphs, a simple example is presented to illustrate the interpretation procedure and the information gained through higher-order cumulants. In this example, a binary image with three different blocks shape is considered as shown in Fig. 9(1). Fig. 9(2) presents the \{x,y\} third-order cumulant map. This map reflects the size of the 

![Image](original_image)

**Fig. 13.** An interpretation of a high positive anomaly in fourth-order map (Figs. 11(3)). (1a) explains high anomaly in (1b) from interactions between black zones in red rectangles.

![Image](fourth_order)

**Fig. 14.** An interpretation of a high positive anomaly in fifth-order map (Figs. 11(4)). (1a) explains high anomaly in (1b) from interactions between black zones in red rectangles.
biggest object and translates it at the origin. In addition, this map shows an object between $50 < x < 60$ and $45 < y < 50$ which results from the interaction of blocks as shown in Fig. 10(1). The length and width of this block are, approximately, the minimal length and the minimal width of interacted blocks, respectively. Then, the third-order cumulant maps provide the approximations of the intersections of the different objects in the directions of the third-order template used. Figs. 9(3, 4) show, respectively, two 2D cross-sections from the $xy(-x)$ fourth-order cumulant map (at $-x=0$ or for $h_3=0$) and $xy(-x)(-y)$ fifth-order cumulant map (at $-x=0$ and $-y=0$ or for $h_3=0$ and $h_4=0$). Eqs. (16) and (17), for the fourth- and fifth-order, express not only the interaction between blocks at the extremities of the template used, but also the cross-relations between these blocks as shown.

Fig. 15. Original images and third-order ($\vec{x}, \vec{y}$) cumulant maps.
in Fig. 10(2). These cross-relations between objects provide more information about the size of the anomaly as shown in Figs. 9(3, 4).

**Remarque 1.** Two-dimensional third-, fourth-, and fifth-order cumulant maps are interpreted through the interaction between the values at the extremities of the template used. For example, for binary image, the high values result from the interactions between the black points as explained above.

**Remarque 2.** Three-dimensional third, fourth-, and fifth-order cumulant maps are interpreted through the interaction between the different sections of the image. This is a generalization of Remarque 1. For example, each 2D-section at $h_3=d$ of the fourth-order template $(h_1,h_2,h_3)$ cumulant map is obtained by the interactions of every two-sections, $S_1$ and $S_2$, from the original image such that the distance between $S_1$ and $S_2$ is equal to $d$.

### 4.2. Cumulant maps for two-dimensional images

#### 4.2.1. Binary images

This example considers sub-horizontal lenses with average width of 15 units and average length of 100 units. Fig. 11(2) shows the third-order $(X, Y)$ cumulant map of Fig. 11(1). This figure shows a horizontal high positive anomaly and reflects certain horizontal distribution of the black points in the original image.
image. The interaction between these points provides a good interpretation of the high anomalies (Fig. 12). For example, the high anomaly between $175< x < 220$ and at $y=75$ (Fig. 12(1.b)) results from the interaction between three black zones as shown in Fig. 12(1.a). Figs. 12(2.a–8.a) provide the interpretations of the anomalies in Figs. 12(2.b–8.b), respectively. The size of the anomaly is reflected by the minimal length and width of the three black zones considered. For example, in Fig. 12(2.b), the length of the anomaly is about 50 m and its width is about 15 m. This corresponds to the minimal length (left-bottom zone) and width (left-top zone) of the zones considered in Fig. 12(2.a).

The fourth- and fifth-order maps (Figs. 11(3, 4) and are obtained using cross-sections, or from Eqs. (16) and (17) as explained above. These maps show first an increase in the intensity of the high positive anomalies and provide a better description of their sizes. In addition to the third-order cumulants, the fourth- and fifth-order cumulants include the interaction between the zones at the diagonal through the term second order cumulant $c_{2}(h_{1}-h_{2})$. This term provides the interaction between the zones at the diagonal of the template. An example is shown in Figs. 13 and 14. The fourth-order map in Fig. 13(1.b) shows that the size of the anomaly is close to top-left black zone in Fig. 13(1.a), while the fifth-order map in Fig. 14(1.b) approximates the bottom-right black zone in Fig. 14(1.a).

4.2.2. Continuous images
In this section, the Walker Lake dataset (Srivastava and Isaaks, 1989) shown in Fig. 15(1.a) is considered. It represents a digital elevation model of the National Cartographic Information Center, part of the United States Geological Survey. The considered image is $260 \times 300$ m$^2$ and contains 78 000 nodes. The $(\bar{x}, \bar{y})$ third order cumulant in Fig. 15(1.b) shows the main anomaly direction until 125 m along $y$ axis and then an inflexion. This inflexion is related to the inflexion of the main anomaly itself, indicating that third-order cumulants are a measure of the connectivity. Fig. 15(1.a) is sampled using a stratified random sampling. Stratified random sampling involves dividing the domain into homogeneous subgroups and then taking a simple random sample in each subgroup. The number of samples is decreased successively until 460 nodes in Fig. 15(3.a). Results in Figs. 15(2.b–3.b) show the ability of the third-order cumulant to provide good results with sparse data compared to the finest one with 78000 nodes. This conclusion reflects the efficiency of high-order spatial cumulants to detect the main characteristics of high-order spatial variability on sparsely and irregularly sampled images. Figs. 15(4.b, 5.b) present, respectively, the fourth- and the fifth-order cumulants maps using only 460 points as shown in Figs. 15(4.a, 5.a). These maps provide, approximately, similar results to the third-order, and then well approximate the third-order map of Fig. 15(1.a). Thus, the main patterns are inferred from the third-, fourth- and fifth-order with 460 points as well as using 78 000 points. The cumulant maps reflect the main configurations formed by the initial data. Moreover, significantly fewer data points are found enough to reproduce the complex structures in the images. The sensitivity of the cumulant maps to the positions of the data points will be studied in the future work.
4.3. Cumulant maps for three-dimensional images

In this example, a three-dimensional image of channels of different sizes (Fig. 16(1)) is considered and it is used here to provide an insight to cumulant calculations and interpretations in three dimensions. This image shows the porosity variation in a 100 \times 130 \times 30 \text{m}^3 field. The data set used (syn_poro.out) are available in the Stanford V Reservoir Data Set (Mao and Journel, 1999). Fig. 14 presents two-dimensional cross-sections along \(x\), \(y\) and \(z\), respectively. Sections along \(x\) (resp., \(y\)) show some similarity in the channel configurations in each section as presented in Fig. 16(3) (resp., Fig. 16(4)). These channels deviate from \(y\) directions to \(x\) directions in the cross-sections along \(z\) as shown in Figs. 17(1–10). In this context, the different order of cumulants will be compared to describe this variation along \(z\).

The \((\bar{x}, \bar{y})\) template in the \(x\)-\(y\) plane presented in Fig. 18(1) shows a high positive anomaly oriented in the \(x\) direction. The size of the anomaly is 10–20 units along the \(y\) axis and about 40 along the \(x\) axis. The third-order cumulant averages the cross-sections along \(z\). Because the channels oriented in the \(x\) directions are bigger than those oriented in the \(y\) direction, then the total average is dominated by the \(x\)-direction channels. The fourth-order \((\bar{x}, \bar{y}, \bar{z})\) and the fifth-order \((\bar{x}, \bar{y}, \bar{z}, -\bar{z})\) cumulants are, respectively, three-dimensional and four-dimensional maps as explained before. Fig. 19 shows similarity between the fourth- and the fifth-order cumulants maps and Figs. 18(2, 3) present two cross-sections of the maps and confirm the similarity. The fourth-order cumulant expression includes the cross-relations between the points of the template used. Fig. 18(2) shows a small deviation from the \(x\)-direction; this deviation is presented in the original image as shown in Figs. 17(3–10).

A two-dimensional section at \(z=a\) from the fourth-order \((\bar{x}, \bar{y}, \bar{z})\) map is calculated based on the interactions between all the two sections along \(z\), from Fig. 16(1), which are separated by the distance \(a\). For the fifth-order maps, the two-dimensional

![Fig. 17. Two-dimensional cross-sections along z from Fig. 16(1).](image-url)
Fig. 17. (Continued)

(7) $z = 21$

(8) $z = 22$

(9) $z = 25$

(10) $z = 30$

Fig. 18. Third-, fourth- and fifth-order cumulant maps for Fig. 16(1).
Fig. 19. Fourth- and fifth-order cumulant maps for Fig. 16(1).

Fig. 20. Fourth-order cumulant map of Fig. 16(1). Distance between sections in (3) and (4) is 4 m.
maps are computed from the interaction of three sections along \( z \) and \(-z\). In this example, the main channels are oriented in the \( x \) and \( y \) directions while the discontinuity along \( z \) or any third direction is much lesser. Also, the \( x \)-direction channels dominate the \( y \)-direction channels. Thus, the interaction between two sections parallel to the \( x-y \) plane will be approximately the same obtained from the interaction between three sections. This is why fourth- and fifth-order maps are similar. As explained above, the fourth-section provides the interaction between the different sections. If these sections are approximately similar, then the third-order will provide results close to those obtained by the higher-order.

Fig. 20(2) shows a cross-section at \( z = 4 \) from the fourth-order map. Figs. 20(3, 4) are two sections, from the original image, separated by 4 m and they are used for the calculation of the map in Fig. 20(2). These sections present large channels oriented in the \( x \)-direction and will dominate all the interactions between the other sections. Fig. 21(2) presents another cross-section at \( z = 30 \). This results from the interaction between the sections in Figs. 21(3, 4) which are separated by the distance of 30 m. These sections are the only available in the original image. Here, the fourth-order map shows some high anomaly along \( y \), and then provides more information than the third-order in describing the different complex patterns presented in the training image.

5. Software availability

HOSC is a stand alone, public domain software program. However, it can be inserted into other software, such as the Stanford Geostatistical Modeling Software (SGeMS) platform (Remy et al., 2009), for specific applications. For example, the SGeMS’s graphical interface can be updated for user parameter input of HOSC using the SGeMS plug-in mechanism as shown in Fig. 22. The HOSC source code will be available to the public on the IAMG (International Association for Mathematical Geosciences) and the COSMO—Stochastic Mine Planning Laboratory websites.
6. Conclusions

This paper presented an algorithm for experimental calculations of high-order spatial cumulants. The calculation of spatial cumulants, including anisotropic experimental cumulant calculations using spatial templates, was introduced, and several examples for two- and three-dimensional data sets fields defined on regular and irregular grids were presented, and their characteristics were analyzed to assess the relations between cumulants and geological patterns. In this paper, the cumulant maps are obtained by using particular templates \( f(x, y); g \) and \( f(x, y); z \). However, HOSC manipulates more general and anisotropic templates composed by any set of two, three or four directions.

The fourth- and fifth-order cumulants were more accurate in describing complex patterns than the third-order and the observations showed, not surprisingly, the ability of the higher-order cumulants to include key relations seen in lower orders. Mathematically, fourth- and fifth-order cumulant functions include not only the interactions between the elements on the extremities of the template used, but also the cross-relations between these elements. This has an effect on the determination of the anomaly’s size as shown in Example 1. Also, the fourth- and fifth-order described well the complex patterns with sparse and dense data as shown in Examples 3 and 4. The sensitivity of the third-order cumulants to the number of nodes used is studied in Example 2. The results suggest that a relatively small data set (460 data out of 78 000 nodes) is sufficient to map the key patterns of the third order cumulant.

Acknowledgements

The work in this paper was funded from NSERC CDR Grant 335696 and BHP Billiton, as well NSERC Discovery Grant 239019 and McGill’s COSMO Lab. Thanks are in order to Brian Baird, Peter Stone, and Gavin Yates, as well as BHP Billiton Diamonds and, in particular, Darren Dyck, for their support, collaboration, data from Ekati Mine, and technical comments.

References


