Math Geosci (2010) 42: 65–99
DOI 10.1007/s11004-009-9258-9

# High-order Statistics of Spatial Random Fields: Exploring Spatial Cumulants for Modeling Complex Non-Gaussian and Non-linear Phenomena

**Roussos Dimitrakopoulos · Hussein Mustapha · Erwan Gloaguen**

Received: 10 June 2008 / Accepted: 17 November 2009 / Published online: 11 December 2009


**Abstract** The spatial distributions of earth science and engineering phenomena under study are currently predicted from finite measurements and second-order geostatistical models. The latter models can be limiting, as geological systems are highly complex, non-Gaussian, and exhibit non-linear patterns of spatial connectivity. Non-linear and non-Gaussian high-order geostatistics based on spatial connectivity measures, namely spatial cumulants, are proposed as a new alternative modeling framework for spatial data. This framework has two parts. The first part is the definition, properties, and inference of spatial cumulants—including understanding the interrelation of cumulant characteristics with the in-situ behavior of geological entities or processes, as examined in this paper. The second part is the research on a random field model for simulation based on its high-order spatial cumulants.

Mathematical definitions of non-Gaussian spatial random functions and their high-order spatial statistics are presented herein, stressing the notion of spatial cumulants. The calculation of spatial cumulants with spatial templates follows, including anisotropic experimental cumulants. Several examples of two- and three-dimensional images, including a diamond bearing kimberlite pipe from the Ekati Mine in Canada, are analyzed to assess the relations between cumulants and the spatial behavior of geological processes. Spatial cumulants of orders three to five are shown to capture directional multiple-point periodicity, connectivity including connectivity of extreme values, and spatial architecture. In addition, they provide substantial information on geometric characteristics and anisotropy of geological patterns. It is further shown that effects of complex spatial patterns are seen even if only subsets of all cumulant templates are computed. Compared to second-order statistics, cumulant maps are found to include a wealth of additional information from underlying geological

R. Dimitrakopoulos (✉) · H. Mustapha · E. Gloaguen
COSMO–Stochastic Mine Planning Laboratory, Department of Mining and Materials Engineering, McGill University, Montreal, Qc, Canada H3A 2A7
e-mail: roussos.dimitrakopoulos@mcgill.ca

patterns. Further work seeks to integrate this information in the predictive capabilities of a random field model.

**Keywords** High-order statistics · Non-Gaussian spatial random functions · Spatial cumulants · Complex geology

## 1 Introduction

Random field models and stochastic data analysis, termed geostatistics, have long been established as the mainstream approach to modeling and predicting spatially distributed and location dependent natural phenomena from limited sets of measurements. Theoretical developments and applications exist in a variety of geoscience and engineering fields (e.g. Matheron 1971; David 1977, 1988; Journel and Huijbregts 1978; Cressie 1993; Kitanidis 1997; Goovaerts 1998; Chilès and Delfiner 1999; Caers 2005; Webster and Oliver 2007; Remy et al. 2009). Despite the considerable work and developments in the field over the past four decades, the mainstream modeling paradigm is founded upon both second-order spatial statistics and the geological information it contains. Although second-order statistics are adequate for the complete statistical description of Gaussian processes, they are inadequate for modeling geological phenomena which typically deviate from Gaussianity and exhibit complex non-linear spatial patterns. These concerns have been articulated since the 1990s (Guardiano and Srivastava 1993; Journel 1997; Tjelmeland 1998; and others).

If the effectiveness of geostatistical modeling, particularly in the presence of non-Gaussianity and non-linearity, is to be enhanced, more spatial information needs to be extracted from measurements and made available. This vital information enhances modeling applications such as the prediction and quantification of spatial uncertainty. Enhancing modeling and predictive capabilities have major applied implications, as demonstrated in mining operations, where the spatial grade distribution and connectivity of the high grade ore to other high grade or low grade ore drives the economics of a mining operation. Similarly, in oil reservoirs, complex spatial arrangements of permeable and impermeable units drive the production characteristics of the reservoir, and predictions from drilling and seismic data have major economic implications. One can certainly expand this list with examples from environmental modeling, ground water resources, $CO_2$ sequestration in geological formations, and so on.

Related to the efforts to enhance modeling is the early work by Guardiano and Srivastava (1993). They attempt to measure, and then use to quantify spatial uncertainty, the relations of multiple-point configurations in the context of connectivity. Connectivity is a notion defined earlier without any spatial reference in Cox (1972) for a set of jointly binomial random variables. The authors use template, an array of a limited number of spatial point indicator variables, to scan a known image. Then, they compute Bayesian conditional probabilities of 'success' of one categorical variable, given that a surrounding multinomial event is computed and subsequently used in an extension of the known sequential indicator simulation. All that is required in this multiple-point approach is the inference of the probabilities or indicator covariances.

Concerns about the above work include the physical meaning of point templates versus event covariances, and that the method ignores the role of high-order statistics such as cumulants, which in turn suggest that the choice of templates may be arbitrary in the context of spatial random fields. Furthermore, the approach may be limited in the information integrated into the modeling process, while the spatial templates used appear geologically un-interpretable, unlike the known second-order spatial connectivity measures (e.g. David 1988; Rendu and Ready 1982; Dreiss and Johnson 1989) and high-order spatial statistics, such as the spatial cumulants presented herein.

Since the earlier work, efforts have been made to further develop new techniques dealing with spatial complexity. These include the now established multiple-point approach described in Strebelle (2002), a multiple-point method based on filters (Zhang et al. 2006; Wu et al. 2008), new Markov random field based approaches (Daly 2004; Tjelmeland and Eidsvik 2004), computer graphic methods that reproduce multiple-point patterns (Arpat and Caers 2007) and other related developments (Boucher 2009; Chugunova and Hu 2008; Mirowski et al. 2008; Remy et al. 2009; Scheidt and Caers 2009; Gloaguen and Dimitrakopoulos 2009; and others). The above developments replace the two-point variogram with a training image (or analog) so as to account for high-order dependencies in geological processes. Although these are all notable advances, it can be argued that a well-defined spatial stochastic modelling framework can contribute further to the capacity to tackle the complex high-order geostatistical description of non-Gaussian and/or non-linear geological phenomena.

High-order spatial cumulants are introduced here as a theoretically grounded and general alternative that can capture the complex spatial geological characteristics, curvilinear features, geometric relations, complex spatial heterogeneity, and the connectivity of extreme values needed for the modeling of spatially dependent geological phenomena. Cumulants are combinations of statistical moments that allow the characterization of non-Gaussian random variables (Billinger and Rosenblatt 1966; Rosenblatt 1985). The cumulants of a random field are derived from its joint characteristic function, which is defined as the logarithm of the moment generating function. Cumulants are also critical contributors to non-Gaussian and non-linear modeling; related developments in the technical literature focus on cumulants for signal filtering and blind deconvolution, as discussed below. Spatial cumulants are a new concept introduced here because they characterize non-linear and non-Gaussian stationary and ergodic spatial random fields, and can provide a new, consistent framework in addressing the issues mentioned above. When considered in space, cumulants allow for large possible combinations of point random variables, suggesting a link to complex spatial connectivity patterns in a geological sense. Early work in the frequency domain is presented in Shiryaev (1960), Billinger and Rosenblatt (1966), and Mendel (1991). Nikias and Petropulu (1993) provide new definitions and terms, in a systematic way, for signal processing approaches that are widespread in the signal processing literature, including the use of high-order multivariate cumulants in non-linear signal processing (Zhang 2005). Another known area of applications of cumulants is astrophysics (e.g. Gaztanaga et al. 2000). In the field of geostatistics, Matérn (1960) refers to the notion of spatial higher-order moments without definitions. There is no documented attempt for defining spatial cumulants in the context of earth science and engineering problems.

The development of a high-order framework as an alternative to current models requires two key elements: (a) the definitions of spatial cumulants and the understanding of the interrelation of cumulant characteristics and in-situ behavior of geological entities or processes, and (b) the development of the related predictive aspects of random field models—the current framework is based on second-order statistics or multiple-point methods that cannot employ spatial cumulants.

This paper contributes to the first element mentioned above, while the second one is addressed in a subsequent publication (Mustapha and Dimitrakopoulos 2010). In the following sections, mathematical definitions, including random variables, their moments, and cumulants, are first outlined. Then, non-Gaussian spatial random functions and their high-order spatial statistics are described in detail. Definitions are followed by approaches to the calculation of spatial cumulants, including implementation of anisotropic experimental cumulants using spatial templates. Subsequently, examples with two- and three-dimensional images are presented. The three-dimensional data is from a kimberlite pipe at the Ekati Diamond Mine, Canada. The examples are used to calculate cumulants up to fifth-order from both complete training images and drillhole data. The examples provide interpretations of cumulant characteristics versus spatial characteristics of the images used in the context of the duality relations between cumulants and geological process. Conclusions and suggested further work follow.

## 2 High-order Statistics for Non-Gaussian Random Functions

Let $(\Omega, \Im, P)$ be a probability space. A function $Z : \Omega \to R$ is a real random variable, if for every $r \in R$ the subset $A_r = \{x \in \Omega / Z(x) \leq r\} \in \Im$. Using a more intuitive definition, a random variable may be seen as a numerical coding of the possible outcomes of a given experiment. If a real random variable is defined on the probability space $(\Omega, \Im, P)$, then it is fully characterized by its cumulative distribution function

$$F_Z(z) = P(Z \leq z), \tag{1}$$

or its probability density function, if it exists,

$$f_Z(z) = \frac{d}{dz} F_Z(z). \tag{2}$$

The probability distribution of a random variable is often parameterized by a small number of parameters, which have a practical interpretation. For example, it is often of interest to know the average value that is represented by the expected value of the random variable, $E[Z]$. Other typical parameters are the variance, skewness and kurtosis.

## 2.1 Moments and Cumulants of Random Variables

### *2.1.1 Definition*

Given a real-valued random variable, $Z$, its moment-generating function (Rosenblatt 1985) is defined by

$$M(w) = E\left[e^{wZ}\right] = \int_{-\infty}^{+\infty} e^{wz} f_Z(z)\,dz. \tag{3}$$

The cumulant-generating function of $Z$ is the Neperian logarithm of the moment-generating function $M$

$$K(w) = \ln\left(E\left[e^{wZ}\right]\right). \tag{4}$$

A random variable is fully determined by its probability density function, its cumulative distribution function, its first or second characteristic function. The $r$th ($r \geq 0$) moment of $Z$ is $\mathrm{Mom}[Z,\ldots,Z] = E[Z^r] = \int_{-\infty}^{+\infty} z^r f_Z(z)\,dz$. Provided that the moment-generating function $M$ has a Taylor expansion about the origin,

$$M(w) = E\left[e^{wZ}\right] = E\left[1 + wZ + \cdots + w^r Z^r/r! + \cdots\right] = \sum_{r=0}^{\infty} \frac{w^r\,\mathrm{Mom}[\overbrace{Z,\ldots,Z}^{r}]}{r!}, \tag{5}$$

then the $r$th moment of $Z$ is the $r$th derivative of $M$ at the origin. The cumulants of $Z$ are the coefficients in the Taylor expansion of the cumulant-generating function, $K$, about the origin

$$\begin{aligned} K(w) &= \ln\left(E\left[e^{wZ}\right]\right) = \ln\left(E\left[1 + wZ + \cdots + w^r Z^r/r! + \cdots\right]\right) \\ &= \sum_{r=0}^{\infty} \frac{w^r\,\mathrm{Cum}[\overbrace{Z,\ldots,Z}^{r}]}{r!}. \end{aligned} \tag{6}$$

Evidently, the moment of order zero is 1 and the cumulant of order zero is 0. The relationship between the first few moments and cumulants, obtained by extracting coefficients from the expansion, is as follows

$$\begin{aligned} \mathrm{Cum}[Z] &= \mathrm{Mom}[Z], \\ \mathrm{Cum}[Z,Z] &= \mathrm{Mom}[Z,Z] - \mathrm{Mom}[Z]^2, \\ \mathrm{Cum}[Z,Z,Z] &= \mathrm{Mom}[Z,Z,Z] - 3\,\mathrm{Mom}[Z,Z]\,\mathrm{Mom}[Z] + 2\,\mathrm{Mom}[Z]^3, \end{aligned}$$

and the expressions in the reverse direction are

$$\begin{aligned} \mathrm{Mom}[Z,Z] &= \mathrm{Cum}[Z,Z] + \mathrm{Cum}[Z]^2, \\ \mathrm{Mom}[Z,Z,Z] &= \mathrm{Cum}[Z,Z,Z] + 3\,\mathrm{Cum}[Z,Z]\,\mathrm{Mom}[Z] + \mathrm{Cum}[Z]^3. \end{aligned}$$

### 2.1.2 Multivariate Cumulants

The relation between moments and cumulants, somewhat surprisingly, is simpler and transparent in the multivariate case than in the univariate case. Let $Z = (Z_1, \ldots, Z_n)$ be the components of a random vector. From the above notation in the univariate case, one may write $E_r = E[Z_r]$ for the component of the mean vector, $E_{rs} = E[Z_r Z_s]$ for the component of the second moment matrix, $E_{rst} = E[Z_r Z_s Z_t]$ for the third moment, and so on. In addition, the cumulants are consistently introduced by $c_r = \mathrm{Cum}[Z_r]$ for the component of mean vector, $c_{r,s} = \mathrm{Cum}[Z_r, Z_s]$ and $c_{r,s,t} = \mathrm{Cum}[Z_r, Z_s, Z_t]$. It is convenient to use Einstein's summation convention, so $w_r Z_r$ denotes the linear combination $w_1 Z_1 + w_2 Z_2 + \cdots + w_n Z_n$, the square of the linear combination is $(w_r Z_r)^2 = w_r w_s Z_r Z_s$ a sum of $n^2$ terms, and so on for the higher powers. The Taylor expansion of the moment-generating function of $Z$ is now given by

$$M(w) = 1 + w_r E_r + w_r w_s E_{rs}/2! + w_r w_s w_t E_{rst}/3! + \cdots, \tag{7}$$

and the cumulants are defined as the coefficients in the Taylor series of the cumulants-generating function

$$K(w) = \ln\big(M(w)\big) = w_r c_r + w_r w_s c_{r,s}/2! + w_r w_s w_t c_{r,s,t}/3! + \cdots. \tag{8}$$

### 2.1.3 Relation Between Moments and Cumulants

Comparison of coefficients reveals that each moment $E_{rs}, E_{rst}, \ldots$ is a sum of cumulant products over the partitions of the subscripts. For example,

$$\begin{aligned} E_{rs} &= c_{rs} + c_r c_s, \\ E_{rst} &= c_{rst} + c_{rs} c_t + c_{rt} c_s + c_{st} c_r + c_r c_s c_t. \end{aligned}$$

Moreover, the translation of moments to cumulants, and vice versa, can be obtained recursively as

$$\begin{aligned} E_{i_1 \ldots i_n} = \sum_{j_1=0}^{i_1} \cdots \sum_{j_{n-1}=0}^{i_{n-1}} \sum_{j_n=0}^{i_n - 1} \binom{i_1}{j_1} \cdots \binom{i_{n-1}}{j_{n-1}} \binom{i_n - 1}{j_n} \\ \times c_{i_1 - j_1, \ldots, i_{n-1} - j_{n-1}, i_n - j_n} E_{j_1 \ldots j_n}, \end{aligned} \tag{9}$$

and

$$\begin{aligned} c_{i_1, \ldots, i_n} = \sum_{j_1=0}^{i_1} \cdots \sum_{j_{n-1}=0}^{i_{n-1}} \sum_{j_n=0}^{i_n - 1} \binom{i_1}{j_1} \cdots \binom{i_{n-1}}{j_{n-1}} \binom{i_n - 1}{j_n} \\ \times E_{(i_1 - j_1) \ldots (i_{n-1} - j_{n-1})(i_n - j_n)} c_{j_1, \ldots, j_n}. \end{aligned} \tag{10}$$

Using the relation between moments and cumulants, it is shown that

$$c_{1,2,3} = E_{123} - E_1 E_{23} - E_2 E_{13} - E_3 E_{12} + 2 E_1 E_2 E_3.$$

Applying the same relation for $\{Z_1, Z_2, Z_3, Z_4\}$ gives

$$\begin{aligned} c_{1,2,3,4} = {} & E_{1234} \\ & - E_{12}E_{34} - E_{13}E_{24} - E_{14}E_{23} - E_1E_{234} - E_2E_{134} - E_3E_{124} - E_4E_{123} \\ & + 2E_{12}E_3E_4 + 2E_{13}E_2E_4 + 2E_{14}E_2E_3 + 2E_{24}E_1E_3 + 2E_{34}E_1E_2 \\ & + 2E_{23}E_1E_4 - 6E_1E_2E_3E_4. \end{aligned}$$

If $Z$ is a zero-mean random vector, then the cumulant reduces to

$$c_{1,2,3,4} = E_{1234} - E_{12}E_{34} - E_{13}E_{24} - E_{14}E_{23}.$$

## 2.2 Non-Gaussian Stationary Spatial Random Functions

### *2.2.1 Spatial Moments and Cumulants*

Let $(\Omega, \Im, P)$ be a probability space and let $(R^n, \beta(R^n))$ be a measurable space. A spatial random field $Z(x)$, $x \in R^n$, is a family of $N$ random variables $\{Z(x_1), Z(x_2), \ldots, Z(x_N)\}$ at locations $x_1, x_2, \ldots, x_N$, where each random variable is defined on $(\Omega, \Im, P)$ and takes values in $(R, \beta(R))$. Assuming $Z(x)$ is a zero-mean ergodic stationary random field indexed in $R^n$, the moments and cumulants can be modeled by distance functions. For example, the third-order cumulants, $c_{1,2,3}$, can be written as

$$c_{1,2,3} = \mathrm{Cum}\big[Z(x_1), Z(x_2), Z(x_3)\big] = c_3^z(\mathbf{h}_1, \mathbf{h}_2), \tag{11}$$

where the subscript in $c_3^z$ denotes the order of cumulant. The terms $\mathbf{h}_1 = h_1\mathbf{e}_1$ and $\mathbf{h}_2 = h_2\mathbf{e}_2$ represent, respectively, two vectors oriented from the point $x_1$ and the points $x_2$ and $x_3$; $h_1$ and $h_2$ are the distances along these vectors, and $\mathbf{e}_1$, $\mathbf{e}_2$ describe the orientations of $\mathbf{h}_1$ and $\mathbf{h}_2$. For simplicity, $\mathbf{h}_i$ is replaced by $h_i$ in the following.

Similarly, the $r$th-order cumulants of $Z(x)$ can be denoted as

$$c_3^z(h_1, h_2, \ldots, h_{r-1}) = \mathrm{Cum}\big[Z(x), Z(x+h_1), \ldots, Z(x+h_r)\big]. \tag{12}$$

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

$$c_2^z(h) = E\big[Z(x)Z(x+h)\big] - E\big[Z(x)\big]^2. \tag{13}$$

Its third-order cumulant is given by

$$\begin{aligned} c_3^z(h_1, h_2) = {} & E\big[Z(x)Z(x+h_1)Z(x+h_2)\big] - E\big[Z(x)\big]E\big[Z(x+h_1)Z(x+h_2)\big] \\ & - E\big[Z(x)\big]E\big[Z(x+h_1)Z(x+h_3)\big] \\ & - E\big[Z(x)\big]E\big[Z(x+h_2)Z(x+h_3)\big] + 2E\big[Z(x)\big]^3. \end{aligned} \tag{14}$$

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

$$c_2^z(h) = E\big[Z(x)Z(x+h)\big], \tag{15}$$

the third-order cumulant is defined as

$$c_3^z(h_1,h_2) = E\big[Z(x)Z(x+h_1)Z(x+h_2)\big], \tag{16}$$

the fourth-order cumulant is

$$\begin{aligned} c_4^z(h_1,h_2,h_3) = {} & E\big[Z(x)Z(x+h_1)Z(x+h_2)Z(x+h_3)\big] - c_2^z(h_1)c_2^z(h_2-h_3) \\ & - c_2^z(h_2)c_2^z(h_3-h_1) - c_2^z(h_3)c_2^z(h_1-h_2), \end{aligned} \tag{17}$$

and the fifth-order cumulant is

$$\begin{aligned} c_5^z(h_1,h_2,h_3,h_4) = {} & E\big[Z(x)Z(x+h_1)Z(x+h_2)Z(x+h_3)Z(x+h_4)\big] \\ & - c_2^z(h_1)c_3^z(h_3-h_2,h_4-h_2) - c_2^z(h_2)c_3^z(h_3-h_1,h_4-h_1) \\ & - c_2^z(h_3)c_3^z(h_2-h_1,h_4-h_1) - c_2^z(h_4)c_3^z(h_2-h_1,h_3-h_1) \\ & - c_2^z(h_2-h_1)c_3^z(h_3,h_4) - c_2^z(h_3-h_1)c_3^z(h_2,h_4) \\ & - c_2^z(h_4-h_1)c_3^z(h_2,h_3) - c_2^z(h_3-h_2)c_3^z(h_1,h_4) \\ & - c_2^z(h_4-h_2)c_3^z(h_1,h_3) - c_2^z(h_4-h_3)c_3^z(h_1,h_2). \end{aligned} \tag{18}$$

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

Spectral representation of regularly sampled spatial cumulants is of interest, both in terms of computational efficiency and physical properties. For more details, the reader is referred to the Appendix.

### 2.2.2 *Summary of Main Properties of Moments and Cumulants*

- $\mathrm{Mom}[w_1Z_1,\ldots,w_rZ_r] = w_1\ldots w_r\,\mathrm{Mom}[Z_1,\ldots,Z_r]$ and $\mathrm{Cum}[w_1Z_1,\ldots,w_rZ_r] = w_1\ldots w_r\,\mathrm{Cum}[Z_1,\ldots,Z_r]$, where $(w_1,\ldots,w_r)$ are constants.
- Moments and cumulants are symmetric functions in their arguments, e.g. $\mathrm{Mom}[Z_1,Z_2,Z_3] = \mathrm{Mom}[Z_2,Z_1,Z_3] = \mathrm{Mom}[Z_1,Z_3,Z_2]$, and so on.
- If the random variables $\{Z_1,Z_2,\ldots,Z_r\}$ can be divided into any two or more groups which are statistically independent, their $r$th-order cumulant is identical to zero, i.e. $\mathrm{Cum}[Z_1,Z_2,\ldots,Z_r] = 0$ whereas, in general, $\mathrm{Mom}[Z_1,Z_2,\ldots,Z_r] \neq 0$.
- If the sets of random variables $\{Z_1,Z_2,\ldots,Z_r\}$ and $\{Y_1,Y_2,\ldots,Y_r\}$ are independent, then $\mathrm{Cum}[Z_1+Y_2,Z_2+Y_2,\ldots,Z_r+Y_r] = \mathrm{Cum}[Z_1,Z_2,\ldots,Z_r] + \mathrm{Cum}[Y_2,Y_2,\ldots,Y_r]$, whereas, in general, $\mathrm{Mom}[Z_1+Y_2,Z_2+Y_2,\ldots,Z_r+Y_r] \neq \mathrm{Mom}[Z_1,Z_2,\ldots,Z_r] + \mathrm{Mom}[Y_2,Y_2,\ldots,Y_r]$.

- If the set of variables $\{z_1, z_2, \ldots, z_r\}$ are jointly Gaussian then all joint cumulants of order $r > 2$ are identical to zero.
- The $r$th-order cumulant function of a non-Gaussian stationary random process $Z(x)$ can be written (Nikias and Petropulu 1993), for $r = 3, 4$ only, as

$$c_3^z(h_1, h_2, \ldots, h_{r-1}) = \mathrm{Mom}\big[Z(x), Z(x+h_1), \ldots, Z(x+h_r)\big] - \mathrm{Mom}^{Gz}\big[Z(x), Z(x+h_1), \ldots, Z(x+h_r)\big],$$

where $\mathrm{Mom}^{Gz}[Z(x), Z(x+h_1), \ldots, Z(x+h_r)]$ is the $r$th-order moment function of an equivalent Gaussian process that has the same mean and autocorrelation function as $Z(x)$. In addition, if $Z(x)$ is Gaussian, then

$$\mathrm{Mom}\big[Z(x), Z(x+h_1), \ldots, Z(x+h_r)\big] = \mathrm{Mom}^{Gz}\big[Z(x), Z(x+h_1), \ldots, Z(x+h_r)\big], \quad \text{and}$$
$$c_3^z(h_1, h_2, \ldots, h_{r-1}) = 0.$$

## 2.3 Calculating Experimental Anisotropic Spatial Cumulants

In this section, the definitions and implementation details of the calculations of experimental cumulants from exhaustive and sparse data sets are described.

### *2.3.1 Definitions*

Spatial cumulants are defined in terms of distances in space. Existing cumulant calculations assume regularly sampled data and/or a regularly sampled training data set. In general, however, geological data is available only on irregularly spaced borehole locations. Similarly to anisotropic experimental variograms, it is possible to restrict the calculation of cumulants to a given direction. For this purpose, the concept of a spatial template for calculating cumulants is introduced. A spatial template $T$ is defined as a particular geometry of points in space; more formally, given a set of directional vectors $\{h_1, \ldots, h_n\}$, the associated spatial template of order $(n+1)$ is defined (considering a spatial location $x$ as a reference) as

$$T_{n+1}^{e_1, e_2, \ldots, e_n}(h_1, h_2, \ldots, h_n) = \big\{(x, x+h_1, x+h_2, \ldots, x+h_n) / \{x, x+h_i, i = 1, \ldots, n\} \subset \text{input image}\big\}. \quad (19)$$

For example, the third-order cumulant of a zero-mean random function $Z(x)$ with the given template $T_3^{e_1, e_2}$ is computed from

$$C^{T_3^{e_1,e_2}} = \frac{1}{N_{h_1,h_2}} \sum_{k=1}^{N_{h_1,h_2}} Z(x_k) Z(x_k + h_1) Z(x_k + h_2),$$
$$\{x_k; x_k + h_1; x_k + h_2\} \in T_3^{e_1,e_2}, \quad (20)$$

and its fourth-order cumulant with the given template $T_4^{e_1,e_2,e_3}$ is calculated as

$$
\begin{aligned}
C^{T_4^{e_1,e_2,e_3}} = {} & \frac{1}{N_{h_1,h_2,h_3}} \sum_{k=1}^{N_{h_1,h_2,h_3}} Z(x_k)Z(x_k+h_1)Z(x_k+h_2)Z(x_k+h_3) \\
& - \frac{1}{(N_{h_1,h_2,h_3})^2}\left[\left(\sum_{k=1}^{N_{h_1,h_2,h_3}} Z(x_k)Z(x_k+h_1)\right)\right. \\
& \left.*\left(\sum_{k=1}^{N_{h_1,h_2,h_3}} Z(x_k+h_2)Z(x_k+h_3)\right)\right] \\
& - \frac{1}{(N_{h_1,h_2,h_3})^2}\left[\left(\sum_{k=1}^{N_{h_1,h_2,h_3}} Z(x_k)Z(x_k+h_2)\right)\right. \\
& \left.*\left(\sum_{k=1}^{N_{h_1,h_2,h_3}} Z(x_k+h_1)Z(x_k+h_3)\right)\right] \\
& - \frac{1}{(N_{h_1,h_2,h_3})^2}\left[\left(\sum_{k=1}^{N_{h_1,h_2,h_3}} Z(x_k)Z(x_k+h_3)\right)\right. \\
& \left.*\left(\sum_{k=1}^{N_{h_1,h_2,h_3}} Z(x_k+h_1)Z(x_k+h_2)\right)\right], \\
& \{x_k; x_k+h_1; x_k+h_2; x_k+h_3\} \in T_4^{e_1,e_2,e_3}, \qquad (21)
\end{aligned}
$$

where $N_{h_1,h_2}$ and $N_{h_1,h_2,h_3}$ are the number of elements of $T_3^{e_1,e_2}$ and $T_4^{e_1,e_2,e_3}$, respectively. The high-order cumulants can recursively be calculated using (10). Note that (21) shows the difference between fourth-order cumulants and fourth-order moments expression.

### 2.3.2 *Implementation on Irregular Grids*

The algorithm developed in this paper is conceptualized for irregular grids. In addition, the general algorithm adapts its algorithm for regular grids and treats them as particular cases of irregular grids. The template presented in Fig. 1 is convenient for regular grids. For irregular grids, tolerances in distances, angles, and bands are incorporated as shown for the third-order template in Fig. 2. In Fig. 2, $h_1$ and $h_2$ are the lags distances with $Th_1$ and $Th_2$ tolerance distances, respectively; $a_1$ and $a_2$ are the angles with tolerances $Ta_1$ and $Ta_2$, 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 $(n+3)$rd-order cumulant calculation. For the two-dimensional case, the rotation angle, measured in degrees clockwise, which rotates the original $Y$ axis (principle direction oriented to

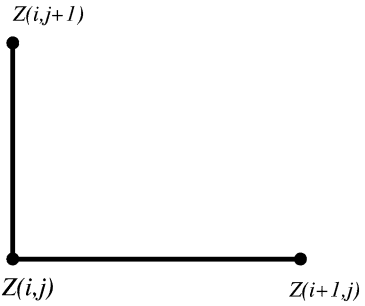



**Fig. 1** L-shape template for cumulant calculation on regular grid data

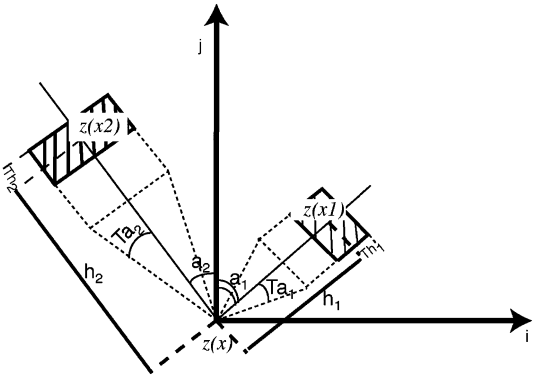



**Fig. 2** Irregular template for 3rd order cumulant calculation. $h_2$ and $h_1$ are the distances with $Th_1$ and $Th_2$, tolerance distances, respectively. $a_1$ and $a_2$ are the angles with tolerances $Ta_1$ and $Ta_2$, respectively. $(i, j)$ is the basis of the Cartesian coordinate system

the north) in the horizontal plane is used; however, in 3D, in addition to the rotation angle in the horizontal plane, a second rotation angle is also employed. The second angle is measured in negative degrees down from the horizontal, which rotates the principal direction from the horizontal.

## 3 Numerical Examples and Interpretations

This section presents several examples of third- to fifth-order cumulants calculated on two- and three-dimensional data sets. Results are interpreted so as to understand the pattern recognition and description capabilities of spatial cumulants, thus assisting the understanding of the interrelation of cumulant characteristics and in-situ behavior of geological entities. If this is understood, then spatial cumulants can be used to enhance predictive modeling capabilities. Most data sets utilized herein represent complete images, and the regular grid approach described in a previous section is followed. In addition, an incomplete and irregularly spaced (drillhole) data set is used to elaborate on the calculation of cumulants up to order five. Calculations are as discussed in a previous section and shown in Fig. 2. It should be noted that the covariance is a measure of the periodicity between pairs of points separated by given distances; similarly, the spatial cumulants of orders higher than two are also a measure of periodicity, but in the direction of the symmetry axis of the template used, that is, the multiple point symmetry. In the examples that follow, cumulants are computed on zero mean data sets.

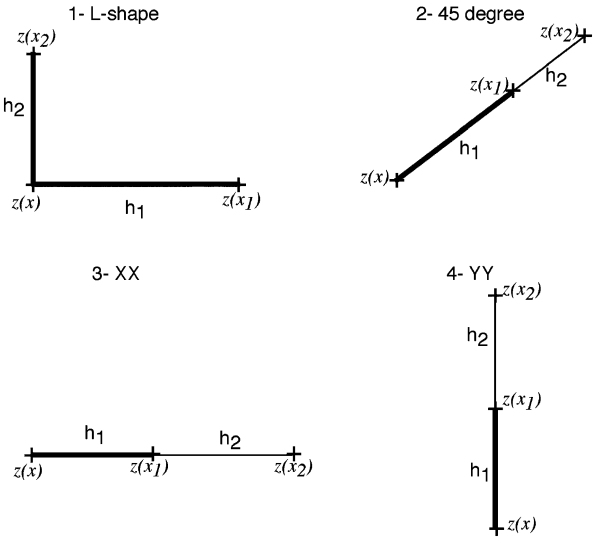



**Fig. 3** The four experimental third-order cumulant templates used in the following examples. (**1**) L-shape, (**2**) 45°, (**3**) $XX$ and (**4**) $YY$. $h_1$ is the distance between $x$ and $x_1$ and $h_2$ is the distance between $x$ and $x_2$

## 3.1 Spatial Templates

In the following example, the covariance map and four directional experimental cumulants are presented, unless otherwise specified. Cumulants are computed with the templates L-shape, 45 degree direction, $xx$ axis direction and $yy$ axis direction, as shown in Fig. 3, and fourth- and fifth-order cumulants are, respectively, computed with the $xy(-x)$ and $xy(-x)(-y)$ templates in 2D and with $xyz$ and $xyz(-z)$ templates in 3D, as shown in Fig. 4. The number of data pairs and/or replicates underlying the estimates varies from 220 to the total number of points in the image under study. Note that only the largest template will have a relatively small number of replicates.

## 3.2 Two-dimensional Images

### *3.2.1 Simple Binary Images*

The first example in Fig. 5(1) consists of five squares of high positive value (0.8) compared to the background of negative value (−0.07). The overall mean value of the original image is zero. Note that the top left, the bottom left and bottom right squares have L-shape symmetry. This example shows that similarly to the second-order cumulant (covariance), all of the third-order cumulant maps identify the size of the largest object in the image. This is represented by the high positive value around the origin in all the cumulant maps. The extent of this high value is equal to the size of the largest squares. In addition, this example shows that the L-shape cumulant in Fig. 5(3) detects the redundancy of features in the direction of the template axis. For example, the high intensity anomaly on the top right corner in Fig. 5(3) corresponds to the L-shape interaction between the top left squares, the bottom left square and the bottom right square of the original image in Fig. 5(1). When the branches of the

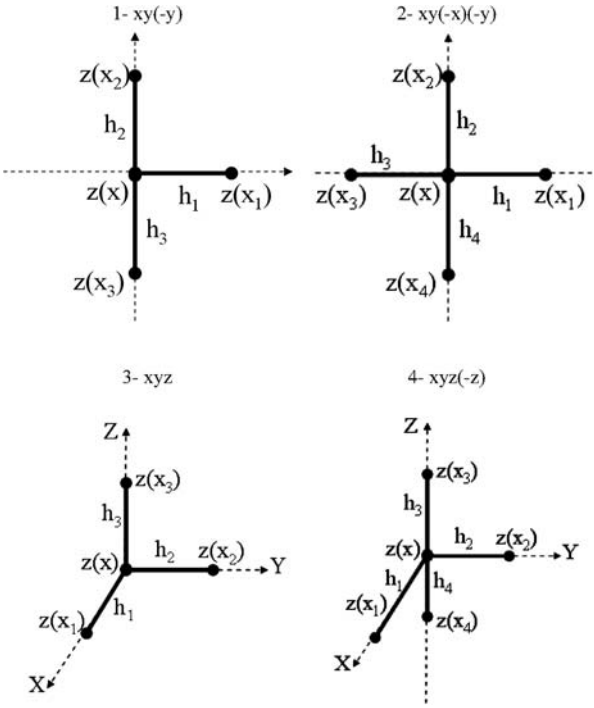



**Fig. 4** Examples of fourth-order cumulant templates (**1**) and fifth-order cumulant templates (**2**)

L-shape template are inside the squares, the cumulant then involves the product of three high positive values, leading to a high intensity third-order cumulant for the given distances and cumulant template.

It is interesting to observe what happens when the L-shape symmetry in the original image is broken, as shown in Fig. 6. Figure 6(1) shows the same image as Fig. 5(1), but without the top left square. This removal breaks the L-shape symmetry of the previous image and allows for observation of the impact on the L-shape cumulant. As expected, the high value at the top right corner is replaced by a strong negative value. The reason is that when removing the mean, white corresponds to $-0.1$, whereas black corresponds to 0.8. Thus, for the lags equal to the separation between squares, the multiplication of three high 0.8 values is replaced with two times 0.8 multiplying $-0.1$, leading to a strong negative value. Note that the other cumulant maps do not show many changes, except that the intensity of the anomalies decreased.

The next example shows the importance of the choice of the orientation of the template. Figure 7(1) shows five squares showing no L symmetry at large lags in the configuration of the squares in the image Cartesian basis. Hence, the corresponding L-shape cumulant is very similar to the one corresponding to a single square, but with higher magnitude around the origin. On the other hand, at distances 50 units along $X$ and along $Y$, it can be seen that there is a negative anomaly. This anomaly is due to the redundancy of the 3 squares on the diagonal of the image. In this L-shape configuration, there is non-negligible contribution of triplets where the two ends are in the positive (black) squares and the corner is in the negative (white) zone. This leads

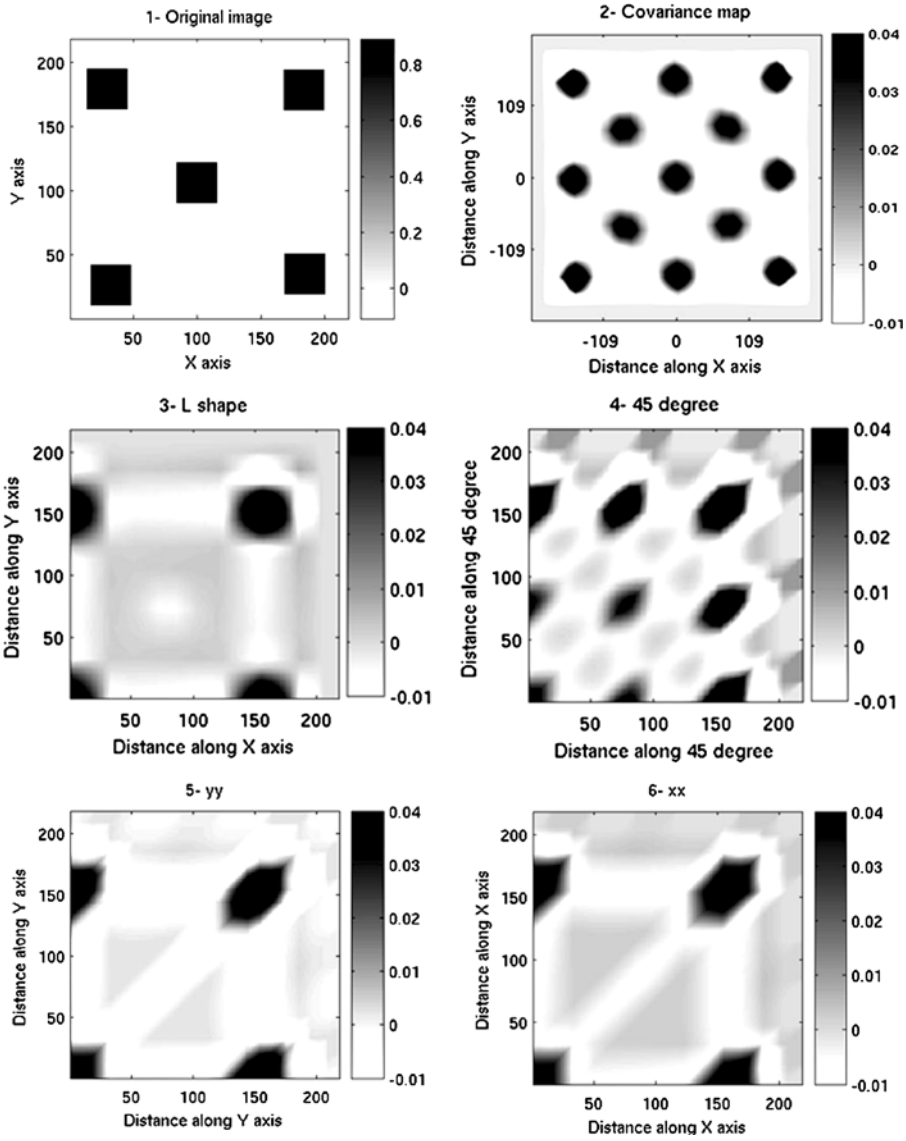



**Fig. 5** From *top left* to *bottom right*: (**1**) original image, (**2**) covariance map, and (**3**) to (**6**) cumulant maps using templates shown in Fig. 3

to a strong negative value. The same appears for distances of 100 units, where the L-shape takes into consideration the first and the last positive square on the diagonal. Figure 7(7) also shows the cumulant map of the L-shape template rotated by 45 degrees, so as to be in the same symmetry axis as the squares in Fig. 7(1). The latter cumulant map shows high positive values at 75 units along $X$ axis and 75 units along $Y$ axis, corresponding to the interaction between the square located in the center, the bottom right, and the top right squares of the image in Fig. 7(1).

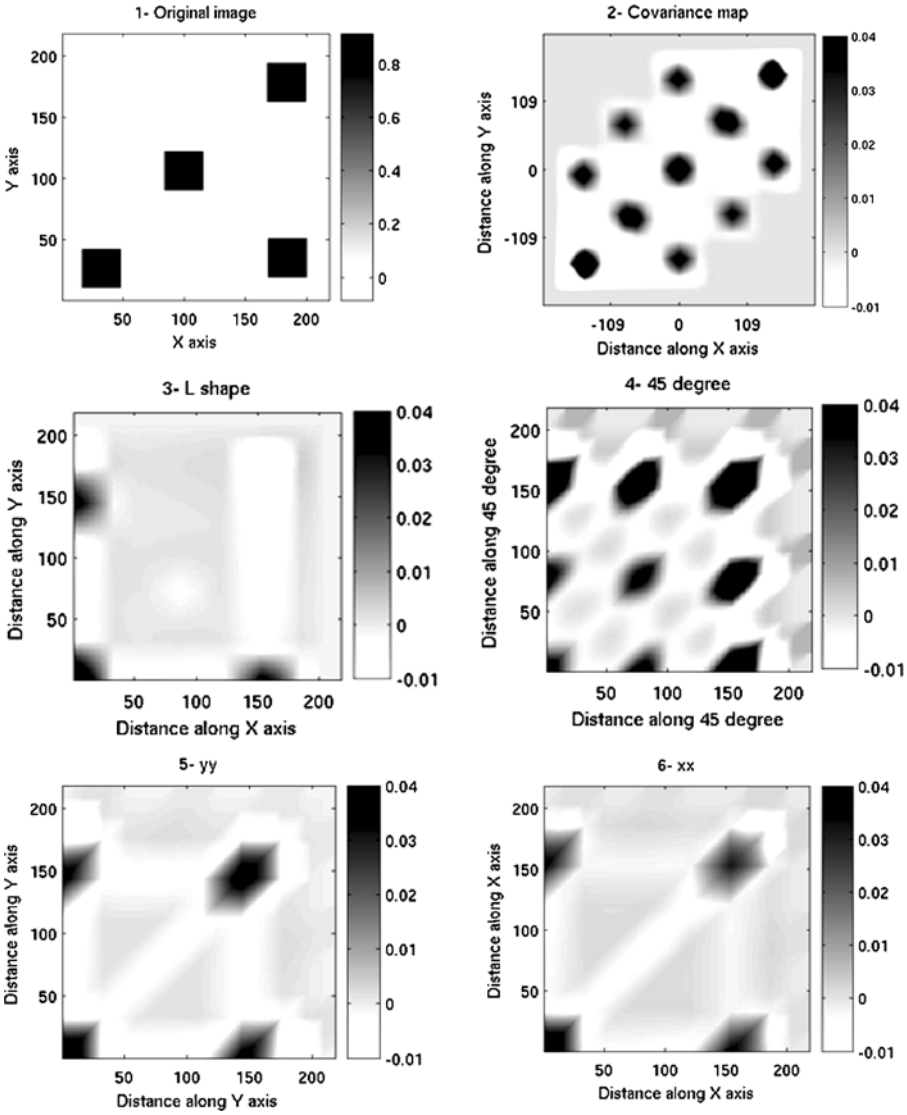



**Fig. 6** From *top left* to *bottom right*: (**1**) original image, (**2**) covariance map, and (**3**) to (**6**) cumulant maps using templates shown in Fig. 3

### *3.2.2 Images Sharing the Same Histogram and Variogram*

Two images sharing the same histogram and variogram are examined here. The examples were used in the past to demonstrate the limits of second-order spatial statistics in capturing geologic information (Krishnan and Journel 2003). The first image and the corresponding cumulant maps in Example 1 are shown in Fig. 8; and the second image, also with the corresponding cumulant maps in Example 2, is shown in Fig. 9. Example 2 corresponds to the set of elongated ellipses of heights of 15 units

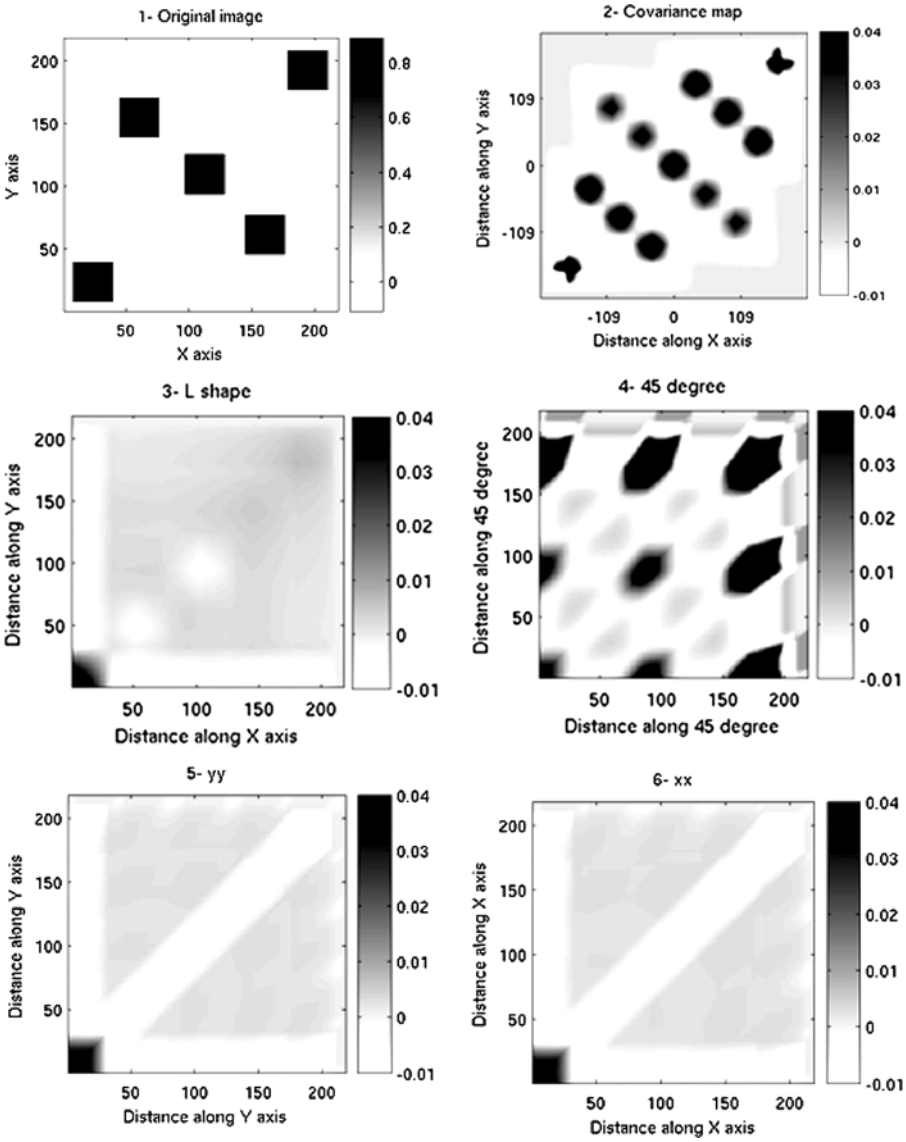
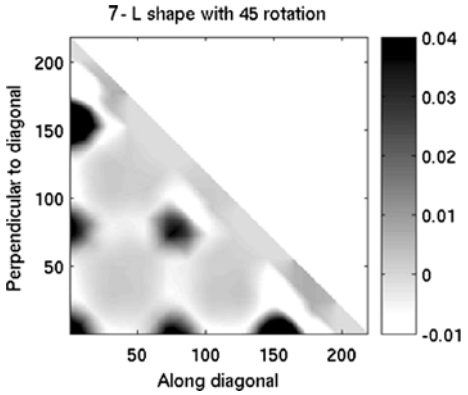



**Fig. 7** From *top left* to *bottom right*: (**1**) original image, (**2**) covariance map, and (**3**) to (**6**) cumulant maps using templates shown in Fig. 3

and width of 25 units. Comparing Example 1 and Example 2 provides several observations. Cumulant maps of images in Example 2 are quite different to those of Example 1. In particular, the $YY$ direction cumulant in Fig. 9(5) show a strong difference compared to $YY$ cumulant in Fig. 8(5). These are features not shown in the cumulant maps of Example 1.

**Fig. 7** (*Continued*)




### *3.2.3 Continuous Examples*

Two two-dimensional continuous images, shown in Figs. 10(1) and 11(1), are considered in this section. These images show porosity variation in $100 \times 100$ m$^2$ sections of the field. The data originates from the Stanford V Reservoir Data Set (Mao and Journel 1999). Covariance maps in Figs. 10(2) and 10(3) illustrate the existence of high continuity in the vertical direction corresponding to the direction of the continuity. However, several more details can be found in the third-order cumulant maps, as shown by the following observations made from Figs. 10 and 11.

The high values between $0 < x < 5$ and along $y$, in the L-shape cumulant map in Fig. 10(3), reflect the existence of at least one vertical and continuous channel with width equal to about 5 m. The other high values around $x = 40$ m and vertically along $y$ come from interactions between different vertical and continuous channels separated by a distance equal to about 40 m in the original image. In addition, the region of high values between $80 < x < 100$ and along $y$ indicates that two channels exist in the original image and are separated by 80 m. Given that the width of this region is about 20 m (in the bottom of the figure), the two or three close channels in the original image are 80 m from theses channels, and show that other vertical channels exist. The 45 degree cumulant map in Fig. 10(4) shows important information related to the location of the channels. The intersections of the diagonal of Fig. 10(4) and the different channels are highlighted at the borders of the map, i.e. $x = 100$ or $y = 100$. The high values along the diagonal in Fig. 10(5) between about $0 < x < 80$ indicate that the channels are irregular and at most they are about 80 m apart. Finally, indications about the channels' separation along $x$ axis can be found in Fig. 10(6). The diagonal shows that between $60 < x < 80$ there are no channels; at least two channels are within a range of 20 m along $x$.

The same type of interpretation is valid for the interpretation of the cumulant maps in Fig. 11. For example, at the borders of the 45 degree map in Fig. 11(4), four clear and separated sets of high values are shown. These high values reflect the intersections of the channels in the original image and the diagonal as explained earlier; there is at least one channel in the region $0 < x < 5$ and along the vertical direction. The L-shape map in Fig. 11(3) shows around $x = 20$ and up to $y = 70$ a region of high

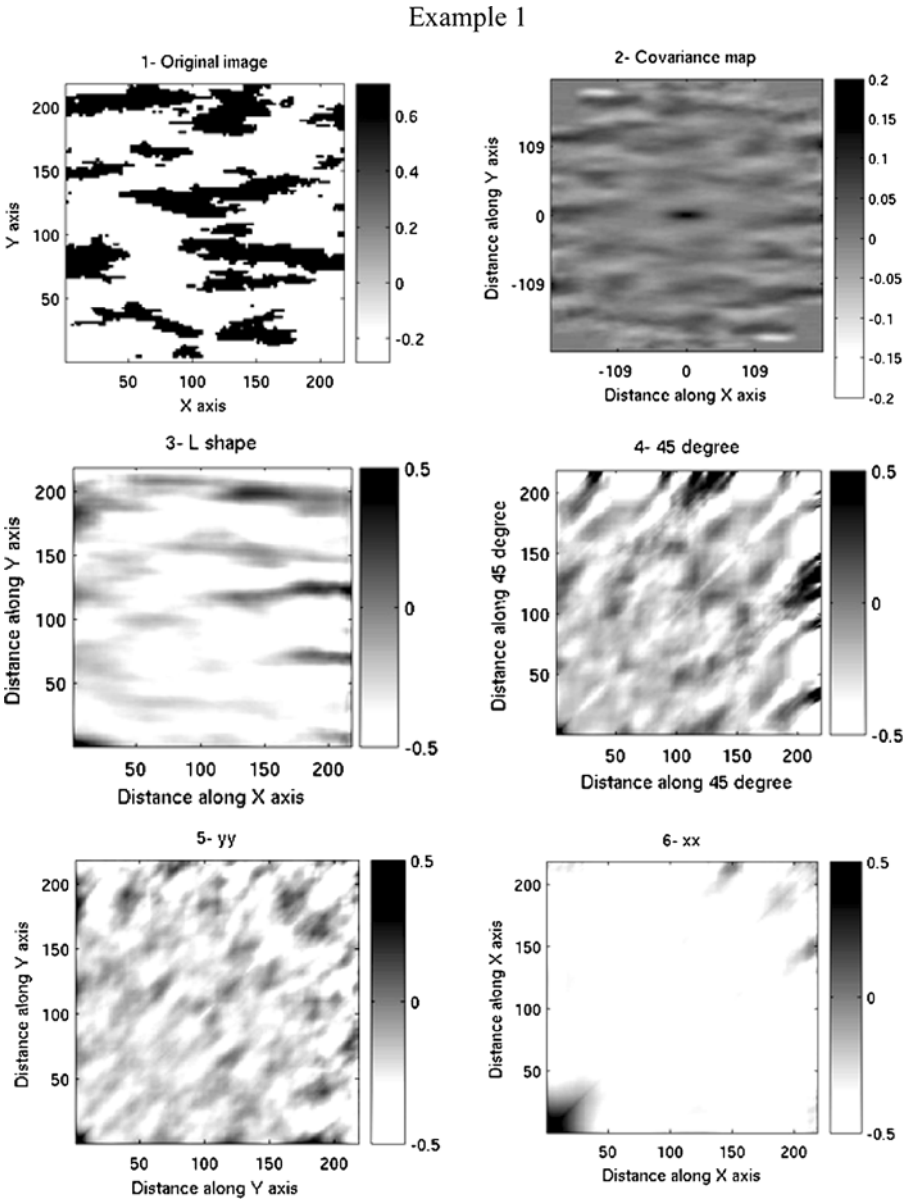



**Fig. 8** Example 1, from *top left* to *bottom right*: (**1**) original image, (**2**) covariance map, and (**3**) to (**6**) cumulant maps using templates shown in Fig. 3

values; this reflects the existence of two sub-channels parallel and separated by a distance equal to about 20 m. This is consistent with the information provided by the 45 degree map. The length of the larger regular part of a channel is equal to about 70 m, as shown in the $yy$ cumulant map in Fig. 11(5). This is exactly the part of the channel

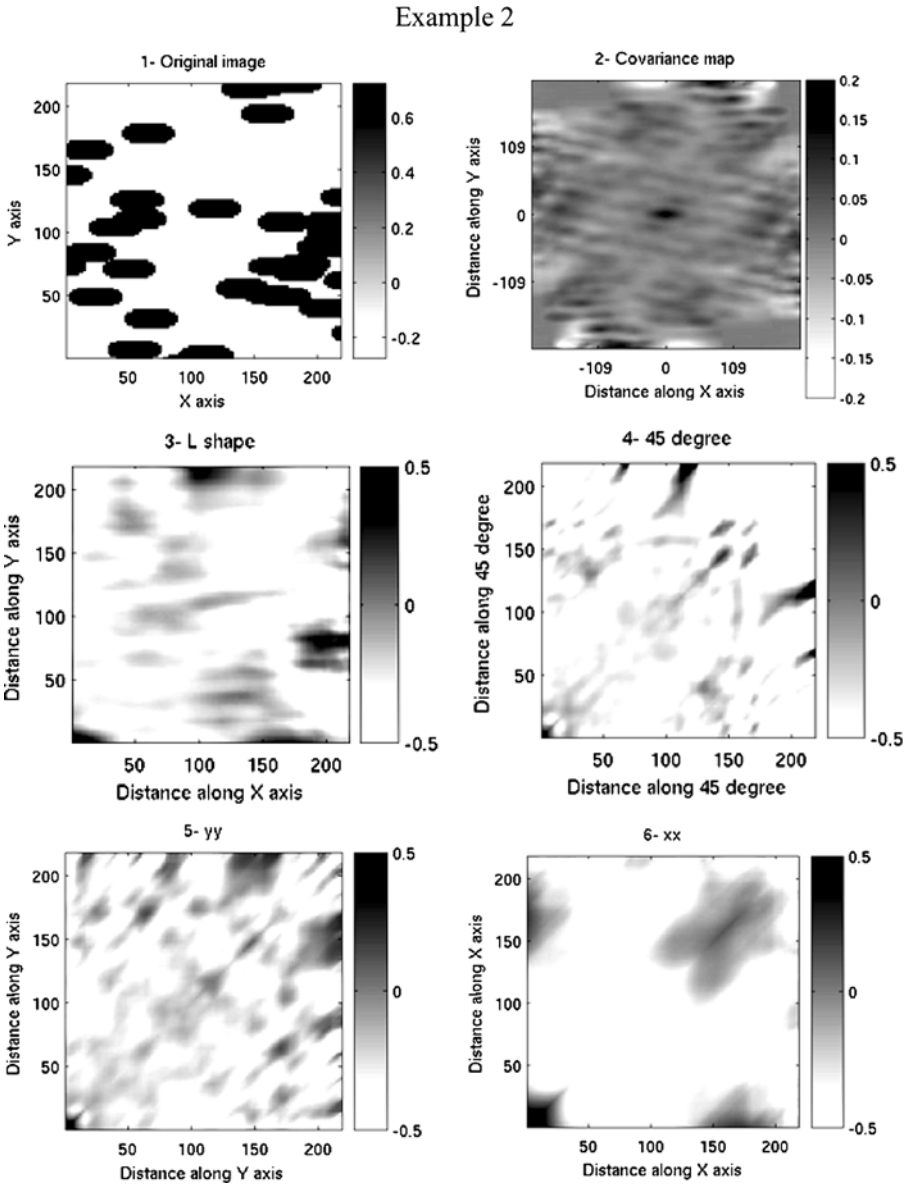



**Fig. 9** Example 2, from top left to bottom right: (**1**) original image, (**2**) covariance map, and (**3**) to (**6**) cumulant maps using templates shown in Fig. 3

that is located at $x = 20$ in the original image. The separation of channels along $x$ axis can be shown in Fig. 11(6).

Note that the values of the different maps may not be in the same range, depending on the order of calculation. The interactions between objects, help in describing interesting properties of the main patterns. For example, the 45 degree third order

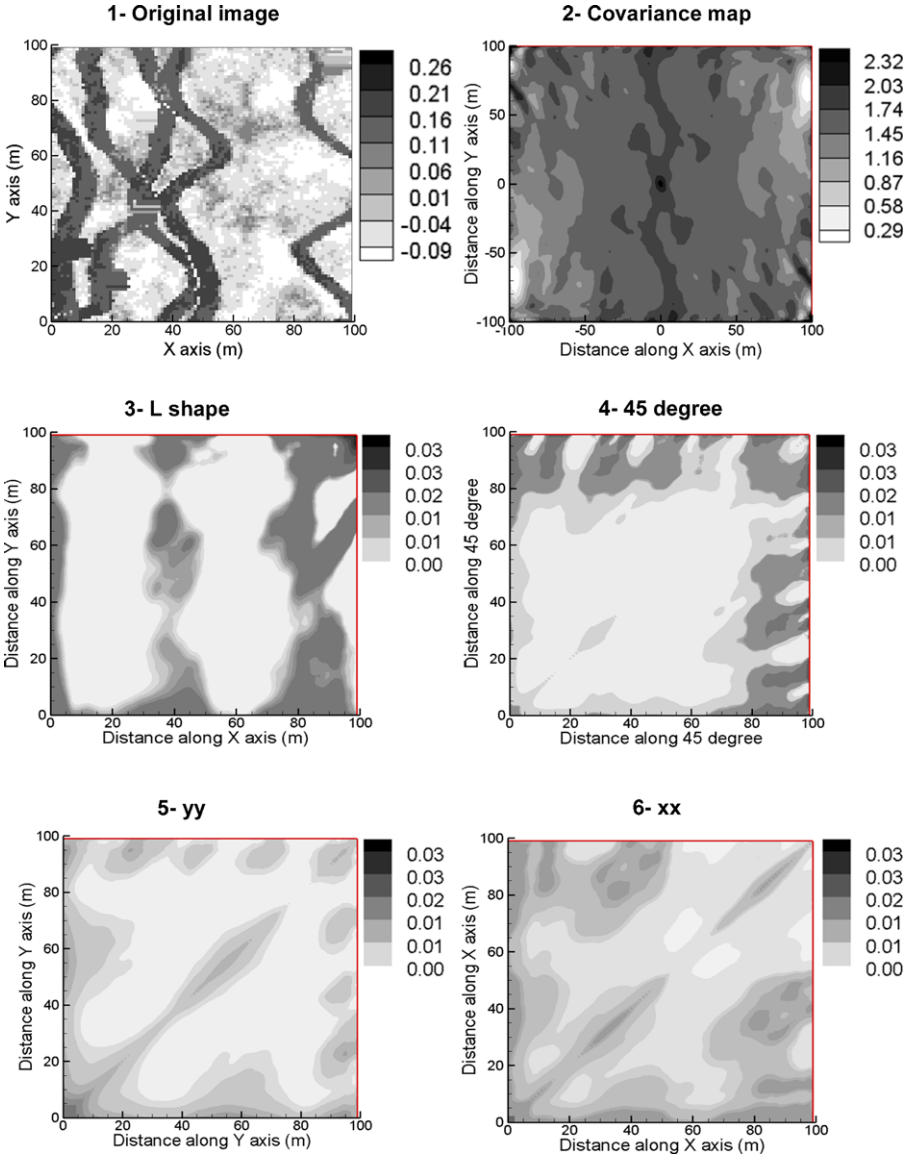



**Fig. 10** From *top left* to *bottom right*: (**1**) original image, (**2**) covariance map, and (**3**) to (**6**) cumulant maps using templates shown in Fig. 3

cumulants in Figs. 8(4) and 9(4), provide the intersection between the objects, i.e. the channels, in the original image and the diagonal as shown at the right hand borders of these figures. This property, combined with what the other orders provide, will better characterize the images. Then, the goal from this discussion is to show that in the cumulant maps there are clear indicators of object locations; at $x = 80$ and along $y$

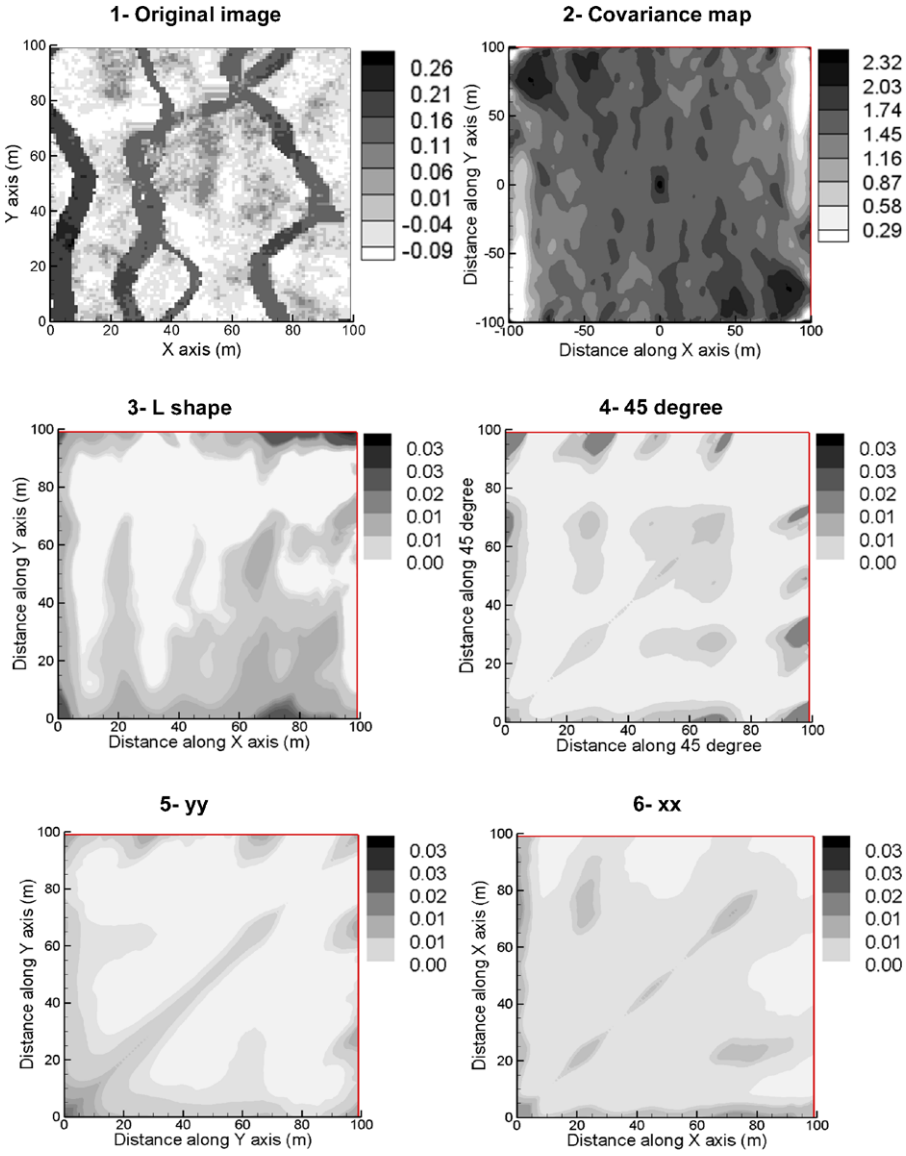



**Fig. 11** From *top left* to *bottom right*: (**1**) original image, (**2**) covariance map, and (**3**) to (**6**) cumulant maps using templates shown in Fig. 3

there are a reasonable number of replicates for the L-shape template. For example, for $h_1 = 80$ and $h_2 = 10$, the number of replicates estimated is about 1350.

### *3.2.4 Information Gain Through Higher-orders: A Simple Example*

In this example, three different block shapes are considered in a binary image as shown in Fig. 12(1). Cumulants of high-order provide information on the interac-

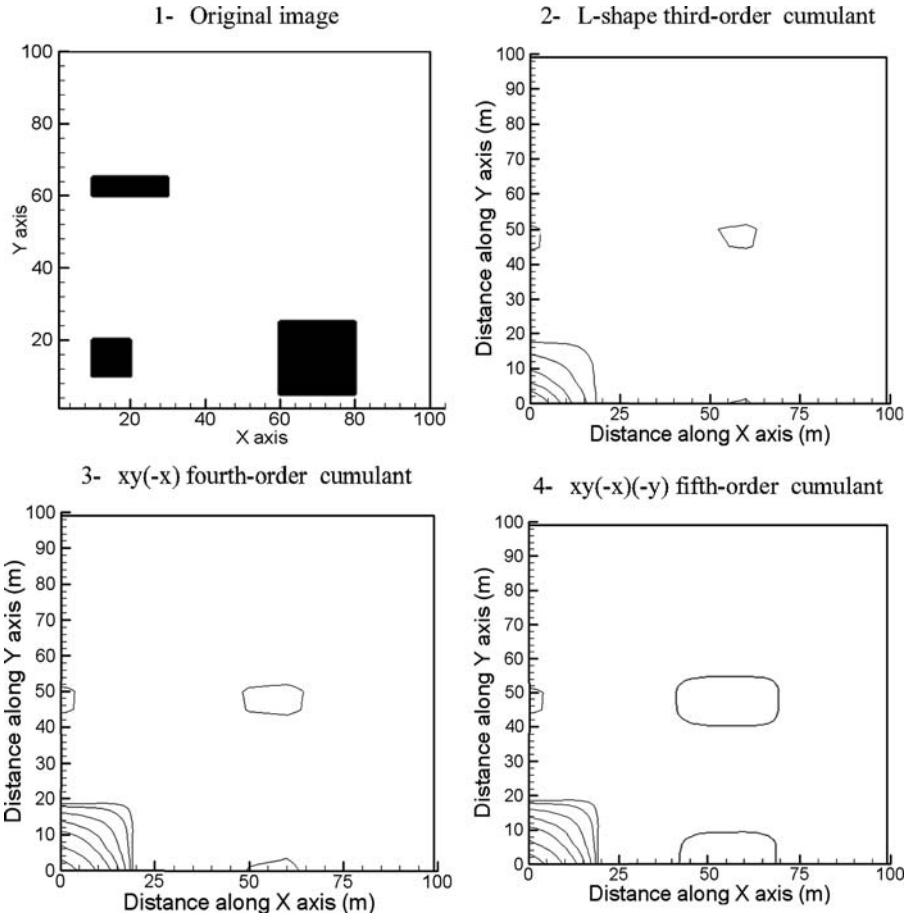



**Fig. 12** The interaction between three blocks in (**1**) using third-, fourth- and fifth-order cumulants in (**2**), (**3**) and (**4**), respectively

tion between these blocks. Figure (2) presents the L-shape (third-order) cumulant map. This map reflects the size of the 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 interaction between blocks as shown in Fig. 12(2). Both the length and width of this block are approximately the minimal length and the minimal width of interacted blocks. Thus, the third-order cumulant maps provide the approximations of the intersections of the different objects in the directions of the L-shape used.

Figures 12(3) and 12(4) show two 2D cross-sections from the $xy(-x)$ fourth-order cumulant map (at $-x = 0$) and $xy(-x)(-y)$ fifth-order cumulant map (at $-x = 0$ and $-y = 0$). Equations (3) and (4), 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 in Fig. 13. These cross-relations be-

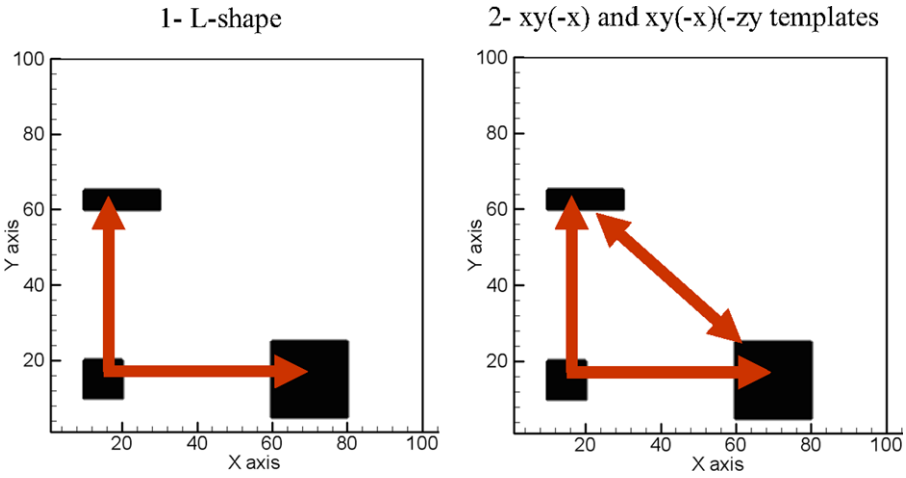



**Fig. 13** The interaction between three blocks using L-shape third-order cumulants (**1**) and $xy(-x)$ and $xy(-x)(-y)$ using fourth- and fifth-order cumulants, respectively

tween objects provide more information about the size of the anomaly as shown in Figs. 12(3) and 12(4).

It is important to comment on the difference between cumulants and moments, which can be demonstrated and stressed with the example in Fig. 12. More specifically, if the same templates employed in this example are used to calculate the fourth- and fifth-order moments, no difference will be present when compared to the third-order moment. This is because moments provide only, and are limited to, the intersection between the objects located at the extremities of the template used.

## 3.3 Three-dimensional Case Study

### *3.3.1 Training Image*

In Figs. 14(2) and 14(3), we consider an interpretation of a diamond bearing kimberlite pipe of the Ekati Diamond Mine, NWT, Canada (Nowicki et al. 2004), and its translation to a 3D binary training image (76,055 nodes). The geological interpretation of the pipe geometry suggests two parts: one on the top and another one on the bottom, as shown in Fig. 15. The part on the bottom is close to a three-dimensional geometric cone shape. The base of this cone, or the intersection between the two parts of the pipe, is a nearly horizontal section 150 m along $x$ and 200 m along $y$ (Fig. 15).

The third-order cumulant maps provide approximations of the pipe shape as shown in Fig. 16(1) to (3). The fourth-order $xyz$ cumulants average the objects and translate them to the origin. For example, the pipe, considered as the only object in Fig. 14(3), is translated to the origin as shown in Fig. 17(1). From this figure, 2D cross-sections are shown in Figs. 17(2) and 17(3). These cross-sections provide, approximately, the results obtained by the third-order cumulant maps in Fig. 16. This conclusion is justified by the fact that the pipe shape is approximately reflected, in a specific plane, by using orthogonal shapes ($xy$ or L-shape, $xz$ or $yz$) cumulants, while it is fully

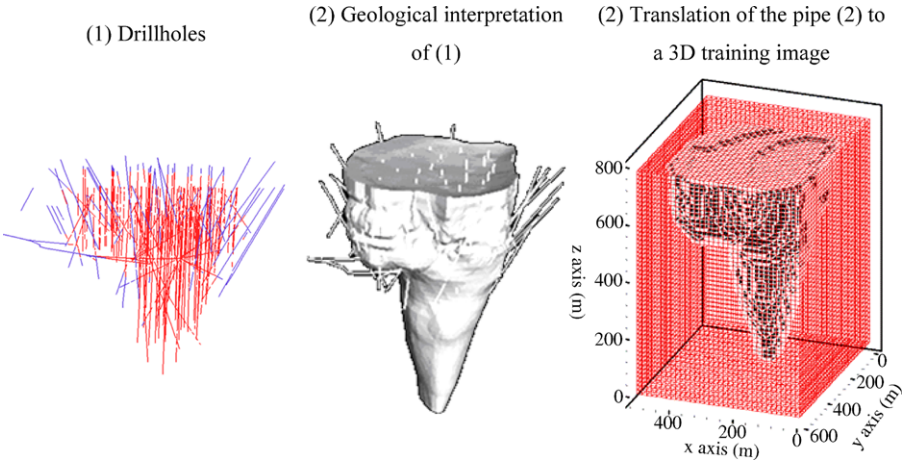



**Fig. 14** A geological interpretation of a kimberlite pipe, (**1**) drillholes; (**2**) a geological interpretation of (**1**); (**3**) regular block approximation of the pipe surrounding rock

characterized using $xyz$ cumulants. It is not surprising that these observations show the ability of the higher order cumulants to include key relations seen in lower orders. More generally, relations between cumulants can be extended for an order higher than four and, in particularly, order five. The fifth-order cumulant maps are based on four directions, and placed in four-dimensional space. Then, cross-sections are used as detailed in a previous section. For example, Fig. 18 shows a 3D cross-section of the $xyz(-z)$ fifth-order cumulant map. This figure translates the pipe to the origin and reflects the results of the fourth-order and, consequently, the third-order results are reflected too.

The pipe shape varies strongly between the bottom and top along the vertical axis ($z$), while the variations are less along the horizontal axis ($x$ and $y$), as shown in Fig. 15. Figure 19 shows results of the third-, fourth- and fifth-order with a different way. Figure 19(1.a) to (3.a) present third-order cumulant maps using line contours from Fig. 16. Several 2D cross-sections along $x$, $y$ and $z$ directions are selected from the fourth-order cumulant maps in Fig. 17. The fronts from 2D cross-sections of Fig. 17 are selected as shown in Fig. 19(1.b) to (3.b). The third-order cumulant map in Fig. 19(1.a) shows a regular shape of the horizontal sections of the pipe, while the fourth- and fifth-order cumulant maps, in Fig. 19(1.b) to (1.c), reflect some horizontal irregularity of the pipe. The main reason comes from the variation of the pipe size along the vertical axis.

This variation is better described by the fourth- and the fifth-order cumulants because they manipulate points in more than two directions. For example, four points are considered for the fourth order-cumulant and one of the points varies along the $z$ direction. The variations along the $x$ and $y$ axes are less than the variation along $z$ axis. The third-order cumulant maps in Fig. 19 (2.a) to (3.a) show, approximately, results close to those obtained from fourth- and fifth-order maps as shown in Figs. 17(2) and 18(3). The fourth- and fifth-order maps detect with more precision

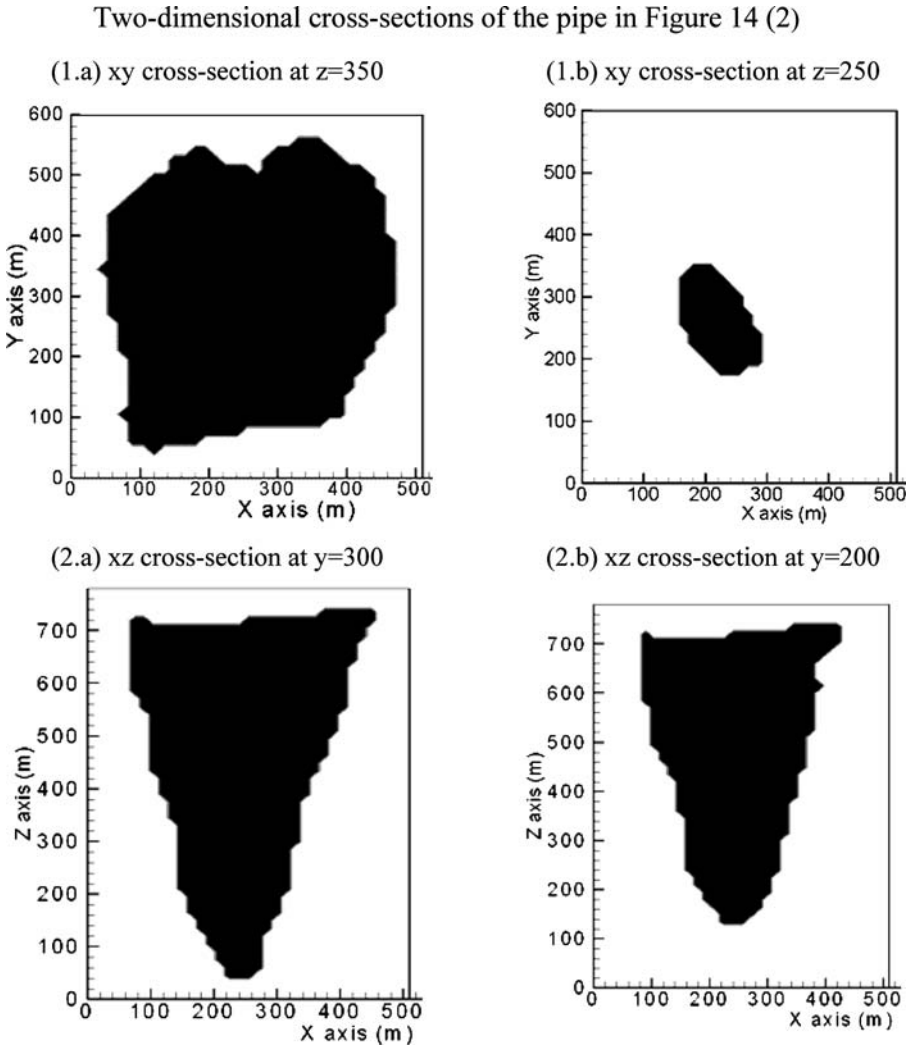



**Fig. 15** Different horizontal and vertical 2D cross-sections of the pipe in Fig. 14

the size along $x$ (150 m) and $y$ (200 m) of the intersection between the two parts of the pipe.

### *3.3.2 Drillhole Data*

The cumulant maps calculated on the 3D training image provide good interpretation for the pipe's shape. In the following, cumulants are calculated on the original data obtained from the pipe drillholes in Fig. 14(1). Figures 20 to 22 show the results obtained. In these figures, the red lines represent the borders of the set of high values in the cumulant maps. These borders reflect the shape of the pipe along $x$

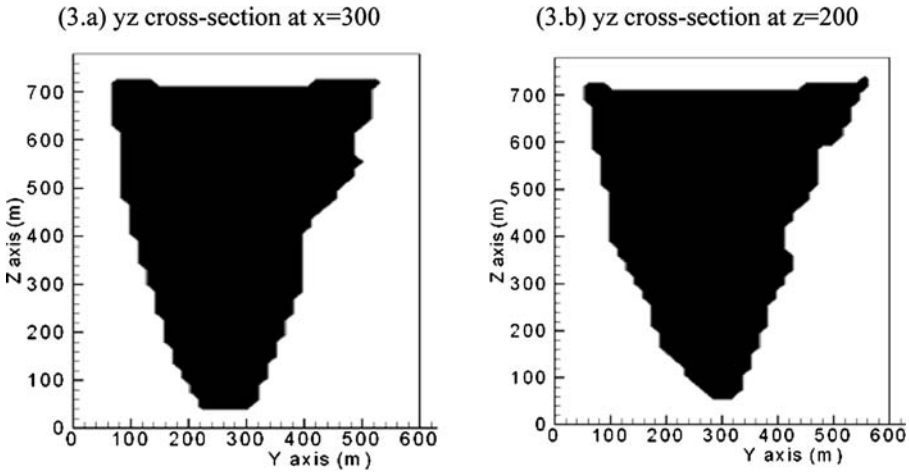



**Fig. 15** (*Continued*)

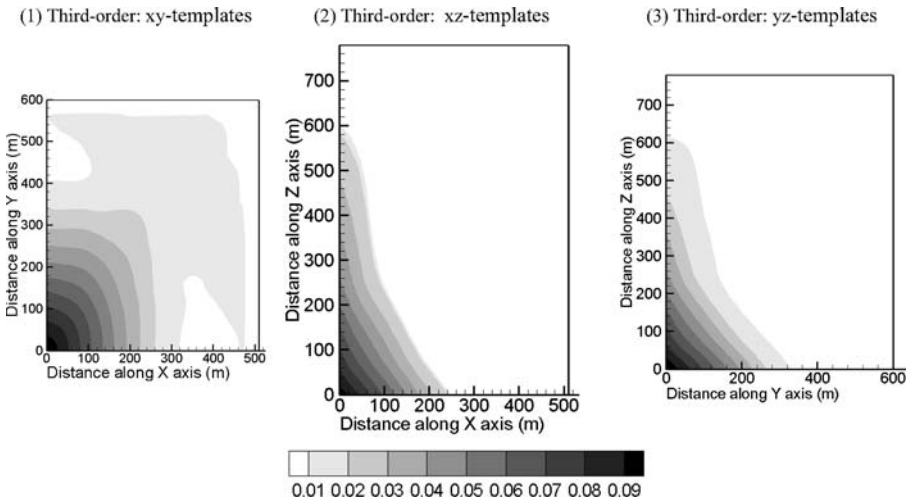



**Fig. 16** Third-order cumulants. (**1**) to (**3**) are third-order cumulants for the Fox kimberlite pipe, NT in Fig. 14(3), using $xy$, $xz$ and $yz$ templates

and $y$ directions, which is approximately similar to the results obtained on the 3D training image (Fig. 19). The top part of the pipe is more easily detected than the bottom part because most of the data points are in the top part until a 300 m depth is reached, as the pipe drillholes show in Fig. 14(1). In Figs. 21 and 22, the fourth- and fifth-order cumulant maps calculated provide better description of the horizontal sections than the third-order map. They show irregularity between $400 < x < 500$ and along $y$, as detected with the fourth- and fifth-order maps from the 3D training image.

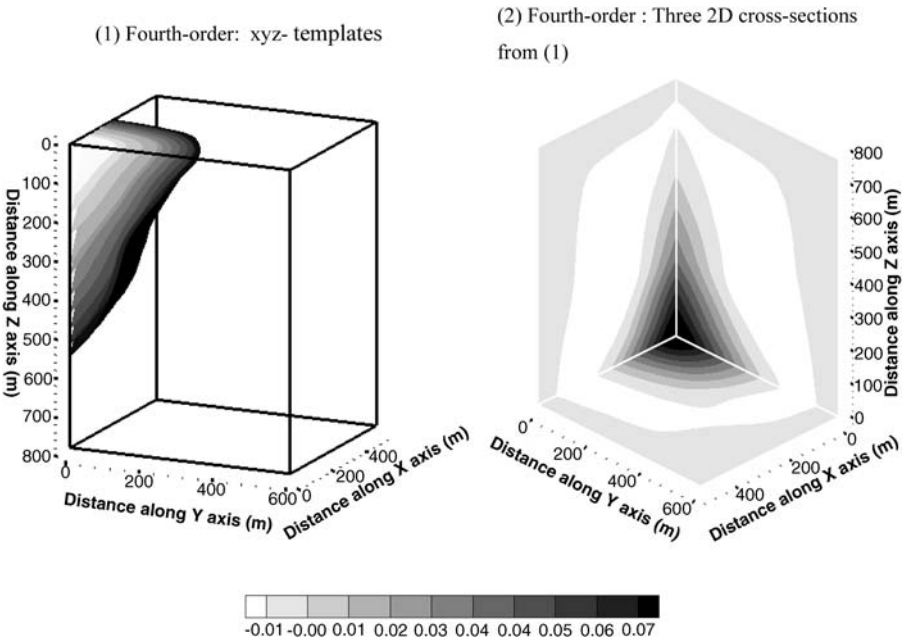



**Fig. 17** Fourth-order cumulants. (**1**) is a fourth-order cumulant map for the Fox kimberlite pipe, NT in Fig. 14(3), using $xyz$ templates; (**2**) are three 2D cross-sections from (**1**)

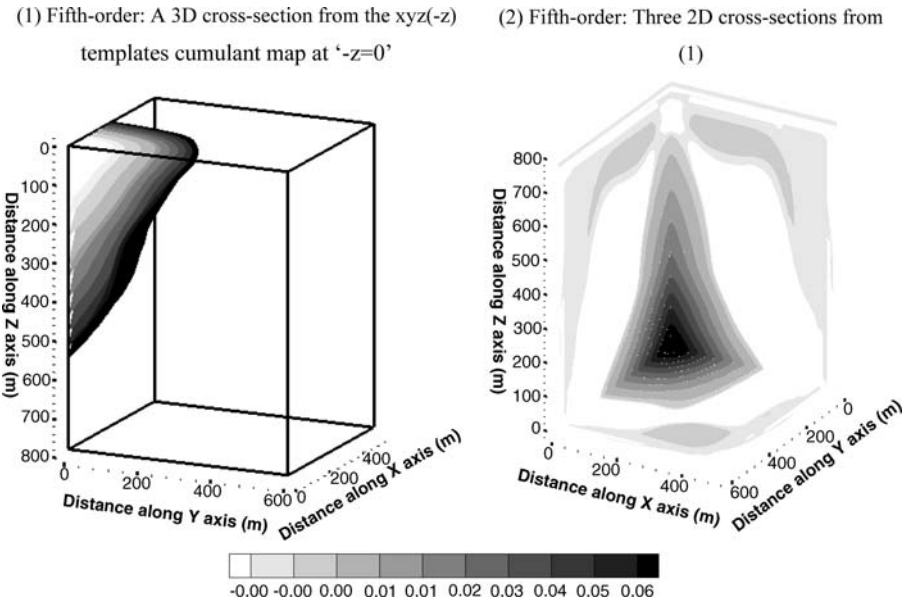



**Fig. 18** Fifth-order cumulants. (**1**) is a 3D cross-section at $-z = 0$ of the five-dimensional $xyz(-z)$ template cumulant map; (**2**) are three 2D cross-sections from (**1**)

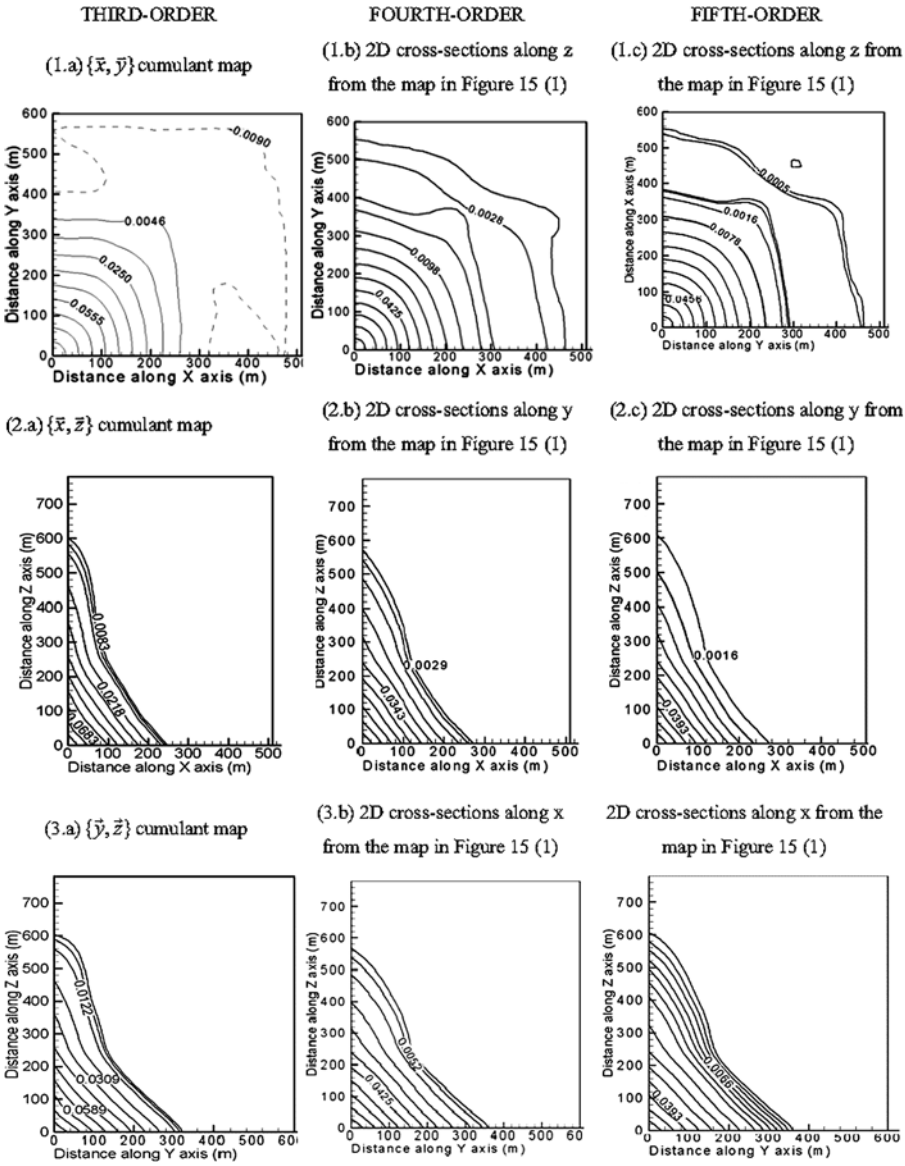



**Fig. 19** Third-, fourth- and fifth-order cumulant maps for the pipe (3D training image) in Fig. 14

## 4 Comments on Pattern Recognition and Duality Relations

The examples presented above provide an insight to the ability of the higher order spatial cumulants to reflect spatial characteristics of various geological patterns. Furthermore, the examples contribute to our understanding of the interrelation between characteristics of cumulants as mathematical entities, and the in-situ behavior of geological characteristics and patterns in the context of the so-called duality relations. As

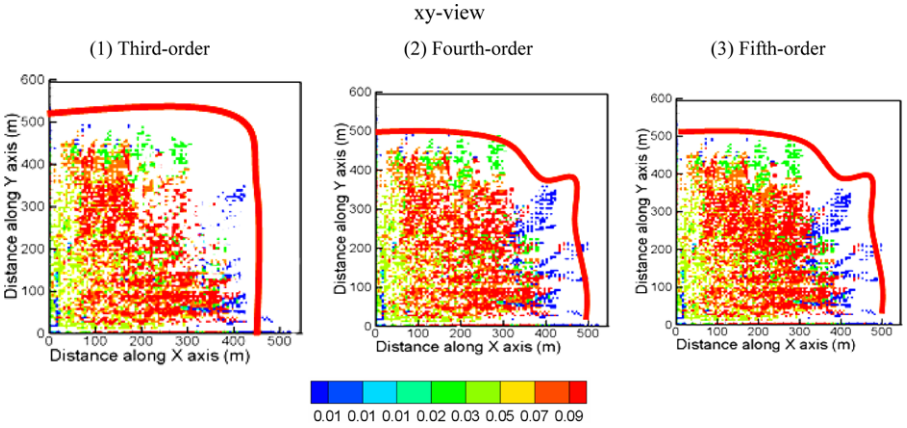



**Fig. 20** $xy$ views of the third-, fourth- and fifth-order cumulant maps calculated using the data from drillholes in Fig. 14(1)

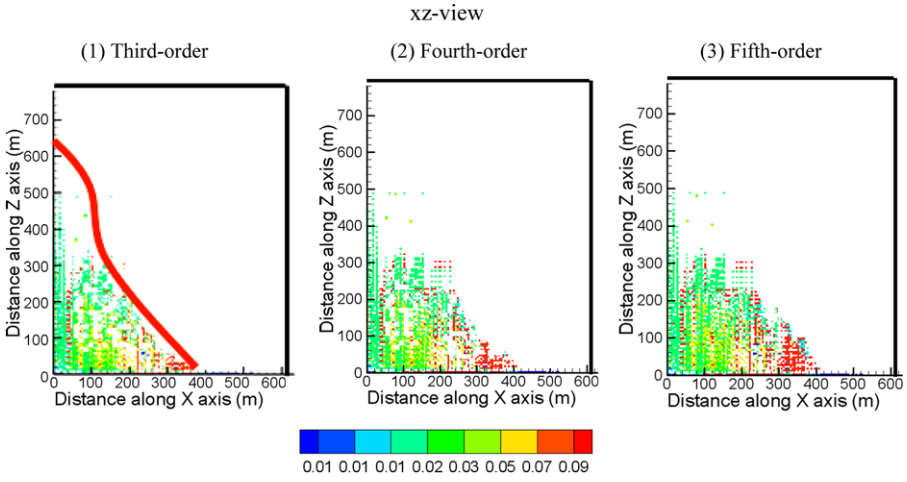



**Fig. 21** $xz$ views of the third-, fourth- and fifth-order cumulant maps calculated using the data from drillholes in Fig. 14(1)

noted earlier, understanding this type of relation is particularly important for the future use of high-order spatial cumulants in spatial prediction and simulation models. Main observations from the previous examples refer mostly to the cumulants calculated using orthogonal shaped templates (i.e. L-shape, $xyz$ and $xyz(-z)$ templates) and may be summarized as follows.

- L-shaped cumulants identify the size of the biggest object in the image considered. This depends on the homogeneity of the objects present in an image. If the geology contains some large objects, and some clusters of separated small objects, we may not be able to distinguish between the different small objects if we use a small

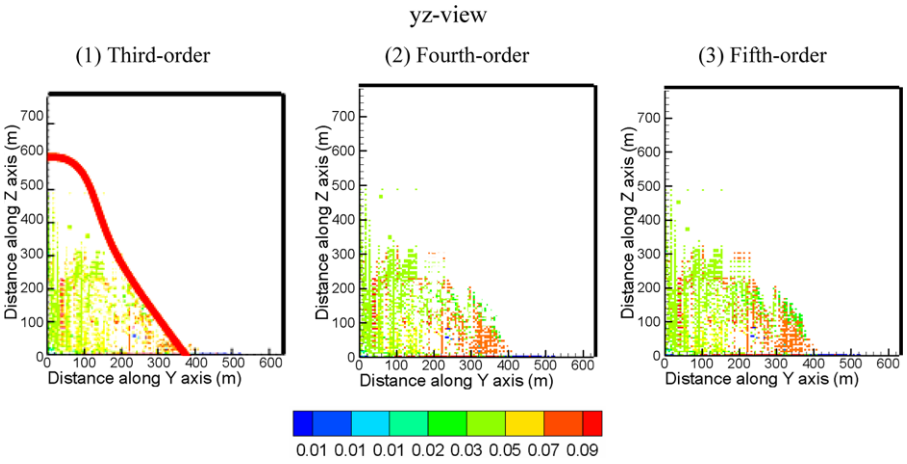



**Fig. 22** $yz$ views of the third-, fourth- and fifth-order cumulant maps calculated using the data from drillholes in Fig. 14(1)

number of lags $(h_1, h_2)$; however, if we increase the number of lags, the resolution would be better and the details will be clearer.

- L-shaped cumulants correctly identify dominant anisotropies in the training image.
- L-shaped cumulants average objects, translating them to origin; this implies that the L-shaped cumulant becomes increasingly insensitive to anisotropies when more objects with different anisotropies are present in an image.
- L-shaped cumulants outside a neighborhood of the origin measure the redundancy of binary objects in horizontal or vertical directions.
- 45-degree cumulant borders (opposite to origin) specify the number of times the 45 degree line intersects objects in the image examined. This provides an idea of the absolute position of some objects in the scene.
- Cumulants of order three are able to determine periodicity in their symmetry axis.
- $xyz$ fourth-order and $xyz(-z)$ fifth-order cumulants show, not surprisingly, the ability of the higher order cumulants to include key relations seen in the L-shape cumulant map.
- The L-shape properties, cited above, can be generalized for higher-order cumulants; for example, $xyz$ and $xyz(-z)$ templates identify the size of the biggest object in the image considered. In addition, they also average objects, translating them to origin.
- $xyz$ and $xyz(-z)$ are three-dimensional templates and they account for the interactions between anisotropies distributed in 3D, while the L-shape third-order cumulant is based on the variations along two main directions. This property is the main difference between the cumulants and moments. The fourth-order cumulant in (21) shows the extra terms added with respect to the fourth-order moment; these terms provide much more interactions between objects at the extremities of the template, in contrary to the fourth-order moment which only focuses on the intersection between these objects.

- For three-dimensional images, $xyz$ fourth- and $xyz(-z)$ fifth-order cumulants are more accurate in describing the main anisotropies. In particular, they provide more geometric information about the shape of the anomalies.
- $xyz$ fourth- and $xyz(-z)$ 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.

Duality relations between the mathematically related properties of cumulants and the behavior of geological processes may be expressed as follows.

- The variation of the $n$th-order spatial statistics values with several spatial templates provides information on the anisotropy of the $n$-point connectivity. For example, if the process under study consists of horizontal layers, the L-shape experimental third-order cumulant will show strong variations, whereas the third-order cumulant along the horizontal axis will be zero.
- The behavior of the cumulants around the origin determines the zone of influence of the average size of a possible existing anomaly in the field under study. For second-order cumulants, this behavior corresponds to the well-known range of influence. The high-order range is related to the rate of decrease of the absolute value of the cumulant. It shows that the samples of natural processes that are closer to each other are subject to stronger interactions than those that are far from each other. In addition, non-collinear templates quantify the directional connectivity of the process under study.
- The behavior of cumulants at large distance determines the degree of homogeneity and the connectivity of the process. As shown on the examples with the squares in the images, the third-order cumulant of such processes tends to constant values for large distances, confirming that for those distances the process is homogeneous. On the contrary, high anomalies at large distances would correspond to the interactions and pseudo-periodicity between the beginning and the end of the geological shapes present. The fourth- and the fifth-order cumulants show homogeneity for large distances in the example of the diamond pipe.
- The behavior of the L-shape third-order cumulant is a representation of the connectivity between perpendicular points. It is a tool to represent processes with main anisotropy in the L directions. In addition, this property is valid for a higher-order to describe anisotropies in the directions of the templates used.
- The relative energy between a set of the directional templates reflects the relative anisotropy in terms of their $n$-points interactions. For example, the continuous image presented earlier shows strong variation in the 45 degree cumulant compared to the $XX$ and $YY$ ones. This reflects the fact that the main anisotropy is in the 45 degree direction.

## 5 Conclusions

This paper presented developments towards a new approach to modeling complex, non-linear, non-Gaussian earth sciences and engineering data, as required in most applications. The new alternative framework is founded upon concepts from high-order statistics that are introduced here in a spatial context. Mathematical definitions

of non-Gaussian spatial random functions and their high-order spatial statistics were described in detail, stressing the notion of spatial cumulants. The calculation of spatial cumulants was introduced, including anisotropic experimental cumulant calculations using spatial templates. Several examples of two- and three-dimensional images were presented, and their characteristics analyzed, to understand and document the interrelations between cumulants and geological patterns.

The present work shows that high-order cumulants characterize spatial pattern redundancy in cumulants and are correlated to the orientation of the spatial template branches in the main axis of the original image. As a result, the choice of the cumulant appears to depend on the geological process, anisotropy and pattern redundancy. Spatial cumulants, up to and including fifth-order, are found to be efficient in characterizing spatial patterns on both binary and continuous images, in two and three dimensions. At the same time, and besides being part of a consistent mathematical model, spatial cumulants were shown to have specific relations between the higher and the lower order moments or cumulants, making the representation of spatial information consistent over a series of orders, globally as well as within a given neighborhood. This consistency, or relation, does not exist in spatial templates as used in currently defined multiple-point statistics. Further research will address the spatial predictive aspects of non-Gaussian, non-linear random field models, so as to facilitate the use of spatial cumulants and related information that current frameworks cannot accommodate.

**Acknowledgements** The work in this paper was funded from NSERC CDR Grant 335696 and BHP Billiton, as well NSERC Discovery Grant 239019. 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 the Ekati Mine, as well as technical comments.

## Appendix: Spectral Representation

### A.1 Spectral Representation of Spatial Cumulants

As the power spectrum is the spectral domain representation of the auto-covariance function, the polyspectrum is the spectral representation of cumulants of order three or higher. Assuming that the cumulant sequence satisfies the following conditions (Nikias and Petropulu 1993)

$$\sum_{h_1=-\infty}^{\infty} \cdots \sum_{h_{r-1}=-\infty}^{\infty} \left|c_r^z(h_1,\ldots,h_{r-1})\right| < \infty$$

or

$$\sum_{h_1=-\infty}^{\infty} \cdots \sum_{h_{r-1}=-\infty}^{\infty} \left(1+|h_j|\right)\left|c_r^z(h_1,\ldots,h_{r-1})\right| < \infty \quad \forall j \in \{1,\ldots,r-1\},$$

the $r$th-order polyspectrum $C_n^z(\omega_1,\ldots,\omega_{n-1})$ of $Z(x)$ exists, is continuous, and is defined as the $(r-1)$-dimensional Fourier transform of the $r$th-order cumulant se-

quence. The $r$th-order cumulant spectrum is thus defined as

$$C_r^z(\omega_1,\ldots,\omega_{r-1}) = \sum_{h_1=-\infty}^{\infty}\cdots\sum_{h_{r-1}=-\infty}^{\infty} c_r^z(h_1,\ldots,h_{r-1})\exp\bigl\{-j(\omega_1 h_1+\cdots+\omega_{r-1}h_{r-1})\bigr\}$$

with $|\omega_i|\le\pi\ \forall i\in\{1,\ldots,r-1\}$ and $|\sum_{i=1}^{r-1}\omega_i|\le\pi$.

In general, $C_r^z(\omega_1,\ldots,\omega_{r-1})$ is complex, carrying information on both magnitude and phase of the phenomena under study and can be represented as

$$C_r^z(\omega_1,\ldots,\omega_{r-1}) = \bigl|C_r^z(\omega_1,\ldots,\omega_{r-1})\bigr|\exp\bigl\{j\Psi_r^z(\omega_1,\ldots,\omega_{r-1})\bigr\},$$

where $|C_r^z(\omega_1,\ldots,\omega_{r-1})|$ is the amplitude and $\Psi_r^z(\omega_1,\ldots,\omega_{r-1})$ is the phase. It is also periodic with period $2\pi$, i.e.

$$C_r^z(\omega_1,\ldots,\omega_{r-1}) = C_r^z(\omega_1+2\pi,\ldots,\omega_{r-1}+2\pi).$$

The power spectrum, bispectrum, and trispectrum are special cases of the $r$th-order cumulant spectrum.

*Power Spectrum*: $r=2$

$$C_2^z(\omega) = \sum_{h=-\infty}^{\infty} c_2^z(h)\exp\bigl\{-j(\omega h)\bigr\},\quad |\omega|\le\pi.$$

*Bispectrum*: $r=3$

$$C_2^z(\omega_1,\omega_2) = \sum_{h_1=-\infty}^{\infty}\sum_{h_2=-\infty}^{\infty} c_3^z(h_1,h_2)\exp\bigl\{-j(\omega_1 h_1+\omega_2 h_2)\bigr\},\quad |\omega_1|\le\pi,\ |\omega_2|\le\pi,\ |\omega_1+\omega_2|\le\pi.$$

*Trispectrum*: $r=4$

$$C_4^z(\omega_1,\omega_2,\omega_3) = \sum_{h_1=-\infty}^{\infty}\sum_{h_2=-\infty}^{\infty}\sum_{h_3=-\infty}^{\infty} c_4^z(h_1,h_2,h_3)\exp\bigl\{-j(\omega_1 h_1+\omega_2 h_2+\omega_3 h_3)\bigr\},\quad |\omega_1|\le\pi,\ |\omega_2|\le\pi,\ |\omega_3|\le\pi,\ |\omega_1+\omega_2+\omega_3|\le\pi.$$

### A.2 The Link Between the Spatial and Spectral Domains

Similarly to the power spectrum and covariance case, the polyspectrum is related to cumulants through the inverse Fourier transform. The inverse Fourier transform

yields

$$c_r^z(h_1,\ldots,h_{r-1}) = \frac{1}{(2\pi)^{r-1}} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} \cdots \int_{-\pi}^{+\pi} C_r^z(\omega_1,\ldots,\omega_{r-1}) \times \exp\bigl\{ j(\omega_1 h_1 + \cdots + \omega_{r-1} h_{r-1}) \bigr\}\, d\omega_1 \cdots d\omega_{r-1}.$$

By choosing $r = 2, 3, 4$ and setting $(h_i) = 0$, $\forall i \in \{1,\ldots,r-1\}$, at $h_r = 0$, the following well known statistical parameters can be derived

$$c_2^z(0) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} C_2^z(\omega)\, d\omega \quad \bigl(\text{spatial variance } \gamma_2^z\bigr),$$

$$c_3^z(0,0) = \frac{1}{(2\pi)^2} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} C_3^z(\omega_1,\omega_2)\, d\omega_1\, d\omega_2 \quad \bigl(\text{spatial skewness } \gamma_3^z\bigr),$$

$$c_3^z(0,0,0) = \frac{1}{(2\pi)^3} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} C_4^z(\omega_1,\omega_2,\omega_3)\, d\omega_1\, d\omega_2\, d\omega_3$$

$\bigl(\text{spatial kurtosis } \gamma_4^z\bigr)$.

## References


Arpat GB, Caers J (2007) Conditional simulation with patterns. Math Geosci 39(2):177–203

Billinger DR, Rosenblatt M (1966) Asymptotic theory of *k*th-order spectra. In: Harris B (ed) Spectral analysis of time series. Wiley, New York, pp 189–232

Boucher A (2009) Considering complex training images with search tree partitioning. Comput Geosci 35:1151–1158

Caers J (2005) Petroleum geostatistics. SPE–Pennwell Books, Houston

Chilès JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

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

Cox DR (1972) Regression models and life-tables. J R Stat Soc B 34:45–58

Cressie NA (1993) Statistics for spatial data. Wiley, New York

Daly C (2004) Higher order models using entropy, Markov random fields and sequential simulation. In: Geostatistics Banff 2004. Springer, Berlin, pp 215–225

David M (1977) Geostatistical ore reserve estimation. Elsevier, Amsterdam

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

Dreiss N, Johnson S (1989) Hydrostratigraphic interpretation using indicator geostatistics. Water Resour Res 25(12):2501–2510

Gaztanaga EP, Fosalba P, Elizalde E (2000) Gravitational evolution of the large-scale probability density distribution. Astrophys J 539:522–531

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

Goovaerts P (1998) Geostatistics for natural resources evaluation. Cambridge University Press, Cambridge

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

Journel AG (1997) Deterministic geostatistics: A new visit. In: Baafy E, Shofield N (eds) Geostatistics Woolongong 1996. Kluwer, Dordrecht, pp 213–224

Journel AG, Huijbregts ChJ (1978) Mining geostatistics. Academic Press, San Diego

Kitanidis PK (1997) Introduction to geostatistics—Applications in hydrogeology. Cambridge Univ Press, Cambridge

Krishnan S, Journel AG (2003) Spatial connectivity: From variograms to multiple-point measures. Math Geol 35:915–925

Mao S, Journel A (1999) Generation of a reference petrophysical/seismic data set: the Stanford v reservoir. Report 12, Stanford Center for Reservoir Forecasting, Stanford, CA

Matérn B (1960) Spatial variation—Stochastic models and their application to some problems in forest surveys and other sampling investigations. Meddelanden fran Statens Skogsforskningsinstitud 49(5) Almaenna, Stockholm

Matheron G (1971) The theory of regionalized variables and its applications. Cahier du Centre de Morphologie Mathematique, No 5, Fontainebleau

Mendel JM (1991) Use of high-order statistics (spectra) in signal processing and systems theory: Theoretical results and some applications. IEEE Proc 79:279–305

Mirowski PW, Trtzlaff DM, Davies RC, McCormick DS, Williams N, Signer C (2008) Stationary scores on training images for multipoint geostatistics. Math Geosci 41:447–474

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

Nikias CL, Petropulu AP (1993) Higher–order spectra analysis: A nonlinear signal processing framework. PTR Prentice Hall, Upper Saddle River

Nowicki T, Crawford B, Dyck D, Carlson J, McElroy R, Oshust P, Helmstaedt H (2004) The geology of the kimberlite pipes of the Ekati property, Northwest Territories, Canada. Lithos 76:1–27

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

Rendu JM, Ready L (1982) Geology and the semivariogram—A critical relationship. In: APCOM'82, pp 771–783

Rosenblatt M (1985) Stationary sequences and random fields. Birkhäuser, Boston

Scheidt C, Caers J (2009) Representing spatial uncertainty using distances and kernels. Math Geosci 41:397–419

Shiryaev AN (1960) Some problems in the spectral theory of higher order moments I. Theory Probab Its Appl 5(3):265–284

Strebelle S (2002) Conditional simulation of complex geological structures using multiple point statistics. Math Geol 34:1–22

Tjelmeland H (1998) Markov random fields with higher order interactions. Scand J Statist 25:415–433

Tjelmeland H, Eidsvik J (2004) Directional Metropolis: Hastings updates for posteriors with nonlinear likelihoods. In: Geostatistics Banff 2004. Springer, Berlin, pp 95–104

Webster R, Oliver MA (2007) Geostatistics for environmental scientists. Wiley, New York

Wu J, Boucher A, Zhang T (2008) SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. Comput Geosci 34:1863–1876

Zhang F (2005) A high order cumulants based multivariate nonlinear blind source separation method source. Mach Learn 61(1–3):105–127

Zhang T, Switzer P, Journel AG (2006) Filter-based classification of training image patterns for spatial simulation. Math Geol 38(1):63–80