Math Geosci (2011) 43: 183–202
DOI 10.1007/s11004-010-9305-6

A Comparison of Random Field Models Beyond Bivariate Distributions

Xavier Emery · Julián M. Ortiz

Received: 27 June 2007 / Accepted: 25 May 2010 / Published online: 19 October 2010

Abstract In order to determine to what extent a spatial random field can be characterized by its low-order distributions, we consider four models (specifically, random spatial tessellations) with exactly the same univariate and bivariate distributions and we compare the statistics associated with various multiple-point configurations and the responses to specific transfer functions. The three- and four-point statistics are found to be the same or experimentally hardly distinguishable because of ergodic fluctuations, whereas change of support and flow simulation produce very different outcomes. This example indicates that low-order distributions may not discriminate between contending random field models, that simulation algorithms based on such distributions may not reproduce the spatial properties of a given model or training image, and that the inference of high-order distribution may require very large training images.

Keywords Multivariate distributions · Poisson hyperplane tessellation · Stable iterated tessellation · Dead leaves tessellation · Random mosaic · Multiple-point statistics

1 Introduction

Geostatistical simulation is widely used in the analysis of regionalized data in order to quantify spatial uncertainty and to perform spatial prediction. It relies on the definition of a random field model over a Euclidean space ($\mathbb{R}^d$, $d \geq 1$) and the con-

X. Emery (✉) · J.M. Ortiz
Department of Mining Engineering, University of Chile, Santiago, Chile
e-mail: xemery@ing.uchile.cl

J.M. Ortiz
e-mail: jortiz@ing.uchile.cl

X. Emery · J.M. Ortiz
ALGES Laboratory, Advanced Mining Technology Center, University of Chile, Santiago, Chile

struction of multiple realizations of this random field, conditioned to the available data. In applications, the Gaussian random field model is very popular because it is fully characterized by a mean value and a covariance function that measures bivariate dependencies. To add flexibility and to describe a wider class of regionalized phenomena, other simulation approaches have been proposed, based on the modeling of multiple-point statistics, that is, statistics that depend on more than two points and therefore measure multivariate dependencies (Guardiano and Srivastava 1993; Strebelle 2002; Ortiz and Deutsch 2004; Daly 2005; Chugunova and Hu 2008).

Multiple-point statistics are often inferred from a training image deemed representative of the random field to be reproduced. In practice, however, such statistics only give an insight into the low-order distributions of this random field and high-order distributions remain unspecified. Indeed, a training image provides a realization of a random vector of finite size n and, therefore, it cannot provide information on distributions of orders greater than n. Furthermore, it may contain little information on low-order distributions (for instance, on the mean value, variance, and covariance function) if the integral range or range of correlation is large with respect to the image size (Lantuéjoul 1991, 2002; Chilès and Delfiner 1999).

The previous statements raise the following questions:

(1) How much information do low-order distributions convey, and to what extent do they characterize a random field? In particular, does the knowledge of these distributions imply little uncertainty in the responses to transfer functions used in practice?
(2) Are there privileged multiple-point configurations (e.g., arrays of aligned points) for which statistics better discriminate contending random field models?

To partially answer these questions, it is of interest to define random fields with the same univariate and bivariate distributions and to compare their higher-order distributions. To achieve this goal, we will perform analytical calculations on conceptual random field models, rather than numerical calculations on training images, for the following three reasons:

(1) Image resolution: a training image is available only at a given resolution that may not be convenient for calculating given statistics (e.g., variograms at lag distances smaller than the image mesh). For instance, suppose that one avails assays of gold grade every 0.5 meter along drill holes in a massive gold deposit. Practitioners may be interested in having an image that reproduces the short-scale behavior (metric scale) observed on the data because of its importance on change of support and assessment of mining selectivity. This can lead to an unmanageably large image for the whole deposit, or even for the size of the local neighborhood one is working with.
(2) Ergodicity problem: the fluctuations of the experimental statistics around the model statistics may be considerable because the training image is never infinitely large. Experimental statistics may therefore not be identifiable with model statistics.
(3) Specification problem: the training image should be a faithful representation of the phenomenon under study and should have (up to fluctuations) the same statistics as the available data, for instance, the same histogram and variogram. If this is not the case, compensation techniques must be introduced (Ortiz et al. 2007).

In the following sections, we will examine examples of random fields for which analytical calculations are tractable and exact simulation is possible, and where the indetermination of high-order distributions can have a non-negligible impact on the response to transfer functions. Although it is easy to identify random field models with the same covariance function, finding models with exactly the same univariate and bivariate distributions is a much more arduous task, as bivariate distributions generally contain broader information than the sole covariance. Two exceptions are the random set indicators and random spatial tessellations, on which we will focus our analysis.

2 Basics About Random Tessellations

Following Lantuéjoul (2002), a random tessellation or random mosaic can be viewed in three different fashions:

(1) A partition of the space $\mathbb{R}^d$ into pairwise disjoint open random cells whose closures fill the whole space.
(2) A population of objects (random cells). This population is characterized by a typical cell, which is obtained by considering a very large domain of $\mathbb{R}^d$ and selecting one cell of the partition at random, each cell within the domain having the same probability of being selected.
(3) A random field, by assigning independent and identically distributed valuations to the cells of the partition. The univariate distribution of the random field so defined is the same as that of the cell valuations, while the bivariate distributions are characterized by the correlation function, which is itself proportional to the geometric covariogram of the typical cell (Rivoirard 1994; Lantuéjoul 2002). Recall that the covariogram for a given lag vector $\mathbf{h}$ is the expected measure of the intersection between a cell and the same cell shifted by $\mathbf{h}$. Accordingly, if two tessellations have the same distribution for the cell valuations and the same correlation function (or, equivalently, the same covariance function), then their univariate and bivariate distributions are the same.

Random tessellations are popular models in the fields of mathematical morphology and stochastic geometry (Serra 1982; Stoyan et al. 1996) and have found a wide range of applications, including agriculture (Guillot et al. 2006), landscape and forest ecology (Hasegawa and Tanemura 1976; Matérn 1979; Garrigues et al. 2007), geography (Furuyama 1976), telecommunication (Gloaguen et al. 2006; Fleischer 2007), biology (Fleischer 2007), image analysis (Lee et al. 2001; Bordenave et al. 2006), materials science and engineering (Fairclough and Davies 1990; Jeulin 1994, 2000; Coster and Chermant 2002), glaciology (Ukita and Moritz 2000), hydrology and geology, for example, for modeling crack patterns in porous media or fracture networks in rock masses (Andersson et al. 1984; Dershowiz and Einstein 1988; Chilès 1989).

In geostatistics, tessellations have been recognized as random fields with no destructuring of the extreme high or low values, as opposed to diffusion random fields (Rivoirard 1994). They can also be used to create other random field

models, such as the following examples: random set indicators, by assigning a valuation of 0 or 1 to each cell; random fields defined by combining independent tessellations (Emery 2005), or a tessellation and a multivariate Gaussian random field (Garrigues et al. 2007); piecewise stationary random fields (Kim et al. 2005); object-based random fields whose components are tessellation cells, such as Boolean and random token random fields (Serra 1982; Chilès and Delfiner 1999; Lantuéjoul 2002).

3 Random Tessellations with an Isotropic Exponential Correlation Function in $\mathbb{R}^3$

This section presents four tessellations with the same correlation function. For the sake of simplicity, we restrict ourselves to the stationary framework and to the three-dimensional space.

3.1 Poisson Hyperplane Tessellation

Let S_3 be the unit sphere of $\mathbb{R}^3$ and $\{H_n : n \in \mathbb{N}\}$ a network of Poisson planes

$$H_n = \left\{\mathbf{x} \in \mathbb{R}^3 : \langle \mathbf{x}, \boldsymbol{\alpha}_n \rangle = p_n\right\}, \tag{1}$$

where $\{(\boldsymbol{\alpha}_n, p_n) : n \in \mathbb{N}\}$ is a homogeneous Poisson point process in $S_3 \times \mathbb{R}_+$ with intensity $\theta = \frac{1}{\pi}$, and $\langle , \rangle$ is the inner product in $\mathbb{R}^3$. The cells of the Poisson hyperplane tessellation (hereunder, "Poisson tessellation" for short) are defined by the polyhedra delimited by the planes (Matheron 1975; Chilès and Delfiner 1999) (Fig. 1(A)). The correlation function ρ of such a tessellation is an isotropic exponential function with a practical range equal to 3 (Appendix 1)

$$\forall \mathbf{h} \in \mathbb{R}^3, \quad \rho(\mathbf{h}) = \exp\bigl(-\|\mathbf{h}\|\bigr). \tag{2}$$

In practice, the tessellation has to be simulated over a bounded domain $\mathbb{D} \subset \mathbb{R}^3$. Therefore, it is not necessary to simulate $\{p_n : n \in \mathbb{N}\}$ in $\mathbb{R}_+$, but only in a bounded interval such that the Poisson planes have a non-zero probability of intersecting $\mathbb{D}$.

3.2 Iterated (Nested) Poisson Tessellation

Consider a Poisson tessellation with intensity $\frac{1}{2\pi}$ (practical range 6), and subdivide each cell with an independent Poisson tessellation with intensity $\frac{1}{2\pi}$ (Fig. 1(B)). The probability that two locations separated by vector $\mathbf{h}$ belong to a single cell of the nested tessellation is the product of the corresponding probabilities for each elementary tessellation, that is, $\exp(-\frac{\|\mathbf{h}\|}{2})^2$. The correlation function is therefore the same as in the previous tessellation (2). Furthermore, the typical cell is also the same (Poisson polyhedron), insofar as it is delimited by a network of Poisson hyperplanes.

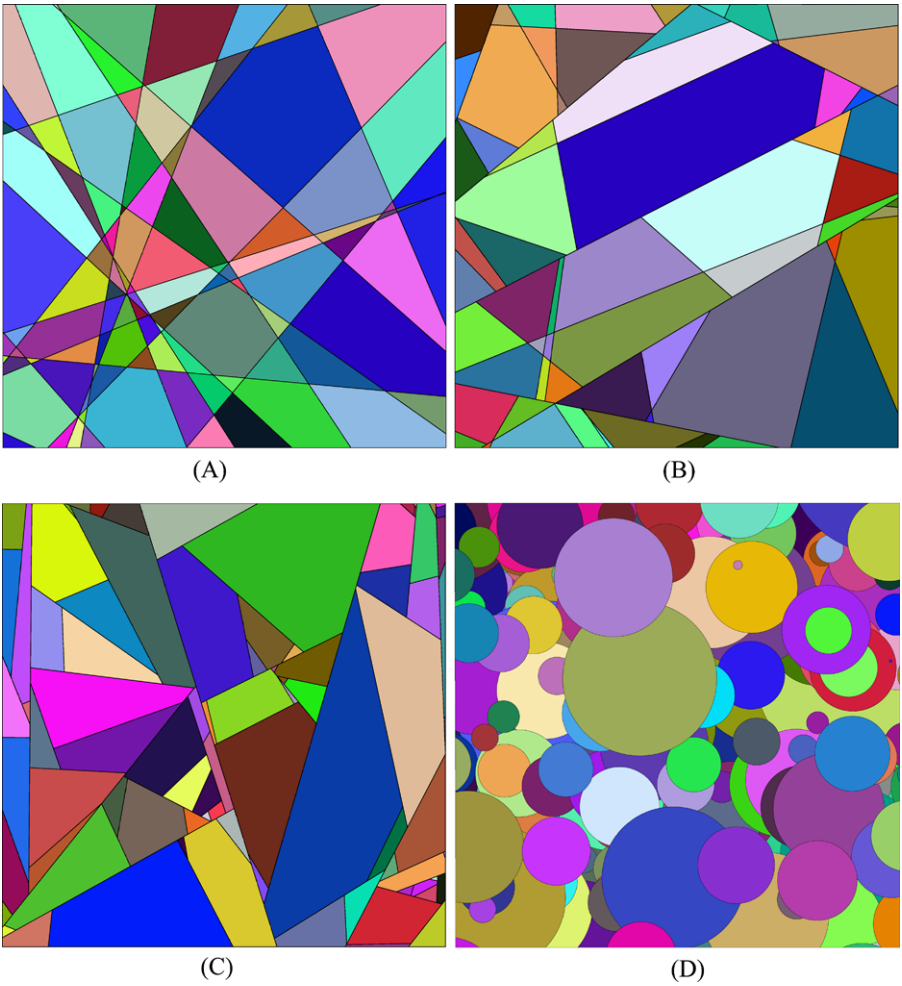

Fig. 1 Realizations of (**A**) Poisson, (**B**) iterated Poisson, (**C**) stable iterated, and (**D**) dead leaves tessellations with isotropic exponential correlation functions (representations in a plane)

3.3 Stable Iterated Tessellation

A stable iterated tessellation of a bounded domain $\mathbb{D} \subset \mathbb{R}^3$ is constructed in the following fashion (Fig. 1(C)) (Nagel and Weiss 2005, 2008):

(1) Consider an initial cell equal to $\mathbb{D}$ and assign to this cell a random lifetime drawn from an exponential distribution.
(2) For every cell whose lifetime is less than a given maximum time,
   (a) draw a Poisson plane intersecting $\mathbb{D}$. This plane may not intersect the cell under consideration, or may divide it into two sub-cells
   (b) increase the lifetime assigned to the cell or the two sub-cells by an exponentially distributed random variable.
(3) Go back to Step 2.

The tessellation model so obtained is invariant with respect to the operation of iteration (or nesting) and rescaling, hence its name. It can be shown (Nagel and Weiss 2008; Nagel et al. 2008) that the typical cell of the tessellation is a Poisson polyhedron. This implies that the correlation function (proportional to the geometric covariogram of the typical cell) is still an exponential function (2).

3.4 Dead Leaves Tessellation

Consider the following components:

(1) A homogeneous Poisson point process in $\mathbb{R}^3 \times \mathbb{R}_+$ (space–time) with intensity $\theta = 1$; this is a countable set of points, hereunder denoted as $P = \{(\mathbf{x}_n, t_n) : n \in \mathbb{N}\}$.
(2) A countable family of independent, nonempty compact random sets (called primary grains) $\{A_n : n \in \mathbb{N}\}$ in $\mathbb{R}^3$.

The components (Poisson process and primary grains) are assumed mutually independent. Let us associate the nth primary grain A_n with the nth point $(\mathbf{x}_n, t_n)$ of P, so that A_n is centered at $\mathbf{x}_n$ and dated on t_n. In $\mathbb{R}^3$, this grain partially covers the grains fallen at earlier dates. After an infinite time, the visible parts of the grains form a partition of $\mathbb{R}^3$ known as the dead leaves tessellation (Jeulin 1997).

As demonstrated in Appendix 2, the dead leaves tessellation has an isotropic exponential correlation function (2) if the primary grains are balls with random diameters with the following probability density function (Fig. 1(D))

$$\forall r \in \mathbb{R}_+, \quad f(r) = \frac{8\exp(r)[r + r\exp(2r) - 4r\exp(r) + \exp(2r) - 1]}{r^2[1 + \exp(r)]^4}. \tag{3}$$

4 Comparison of Multivariate Distributions

For each of the previous tessellations, one can define a random field by assigning each cell with a random variable (valuation). Provided that these variables are drawn from the same distribution, the random fields so defined have exactly the same univariate and bivariate distributions, and can be discriminated only on the basis of their higher-order distributions.

4.1 Trivariate Distributions

Let us consider three spatial locations $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ and denote by a, b and c the sides of the triangle formed by these locations. The distribution of the random variables at $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ depends on the following:

(1) The univariate distribution (distribution of cell valuations).
(2) The probabilities that any two of these locations belong to the same cell, which depend only on the correlation function (2).
(3) The probability that the three locations belong to the same cell: $p_{123} = \text{Prob}\{X \subseteq 1 \text{ cell}\}$.

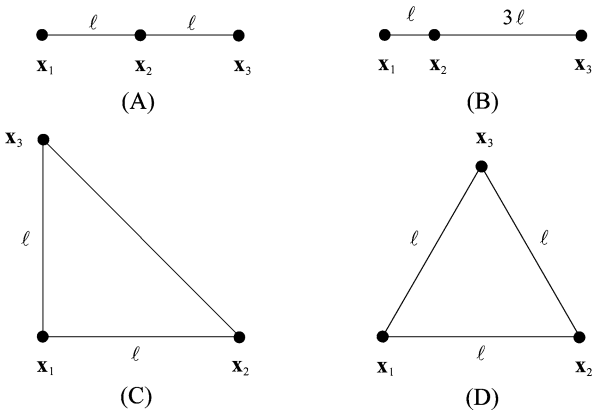

Fig. 2 (**A**)–(**D**): Examples of three-point configurations in the plane

Because the four tessellations under consideration have identical univariate distribution and correlation functions, their trivariate distributions are characterized by the probability p_{123}.

Now, this probability is the same in the Poisson, iterated Poisson, and stable iterated tessellations, insofar as in each case the cells are Poisson polyhedra. To calculate p_{123}, note that Poisson polyhedra are convex, so that X belongs to a single cell of the tessellation if the entire triangle with vertices $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ is contained in one cell, which yields, according to (17) in Appendix 1 (Santaló 2004)

$$p_{123}^{\text{Poisson}} = \exp\left\{-\frac{a+b+c}{2}\right\}. \tag{4}$$

As for the dead leaves tessellation, as shown in (19) in Appendix 2, one has

$$p_{123}^{\text{Dead leaves}} = \frac{E\{V_I(a,b,c;R)\}}{E\{V_U(a,b,c;R)\}}, \tag{5}$$

where $V_I(a,b,c;R)$ and $V_U(a,b,c;R)$ are the expected volumes of the intersection and union of three balls centered at $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, with the same diameter R drawn randomly from the distribution given in (3). By introducing the geometric covariogram K of the primary grains, as shown in (21) in Appendix 2, one has

$$E\{V_U(a,b,c,R)\} = 3K(0) - K(a) - K(b) - K(c) + E\{V_I(a,b,c;R)\}. \tag{6}$$

The volume $V_I(a,b,c;r)$ of the intersection of three balls with equal diameter r can be expressed analytically (Helte 1994; Powell 1964). The expected volume is obtained by a numerical integration over $\mathbb{R}_+$

$$E\{V_I(a,b,c,R)\} = \int_0^{+\infty} V_I(a,b,c;r)\,f(r)\,dr. \tag{7}$$

Four configurations are examined: aligned points (Fig. 2(A) and (B)), isosceles right triangle (Fig. 2(C)) and equilateral triangle (Fig. 2(D)). In each case, the probability

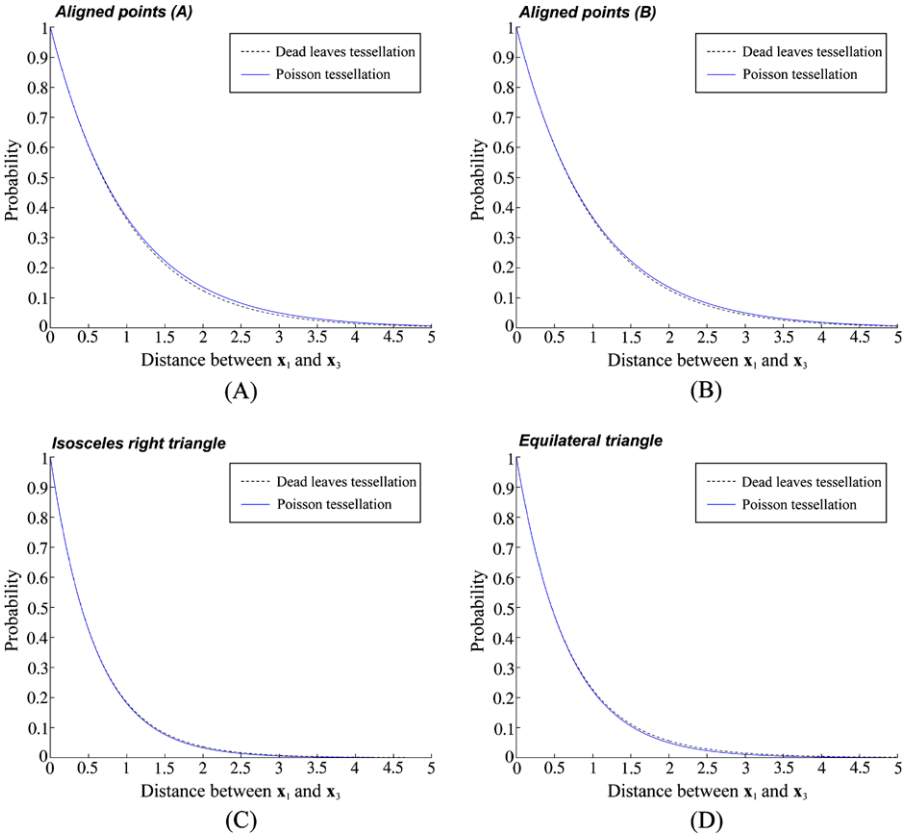

Fig. 3 (**A**)–(**D**): Probability that three points $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ belong to a single cell as a function of the distance between $\mathbf{x}_1$ and $\mathbf{x}_3$ (Poisson and dead leaves tessellations). The images correspond to the configurations shown in Fig. 2

p_{123} that the three points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ of the configuration belong to the same cell of the tessellation is calculated and plotted as a function of the distance between $\mathbf{x}_1$ and $\mathbf{x}_3$ (Fig. 3). Note that the contrast between the Poisson and dead leaves tessellations is quite small, especially when the distance between consecutive locations is small or large (in which cases p_{123} tends to 1 or 0). However, this contrast is greater for the first configuration (aligned and regularly spaced points) than for the triangle configurations, which is because aligned point configurations give an insight into the Markov property (screening effect) of the random field model: this property holds in the Poisson, iterated Poisson, and stable iterated models, but not in the dead leaves model because the cells in this model are not necessarily convex or connected sets.

It is also of interest to determine the size of the realization required for the estimate of p_{123} (empirical proportion of three-point configurations that belong to a single cell) to be accurate enough in order to discriminate between the Poisson and dead leaves models. As an illustration, Fig. 4 presents the estimates obtained for the configuration in Fig. 2(A) by considering five realizations of a Poisson tessellation on a cubic domain with an edge length of $L = 15$ and a discretization mesh of one tenth

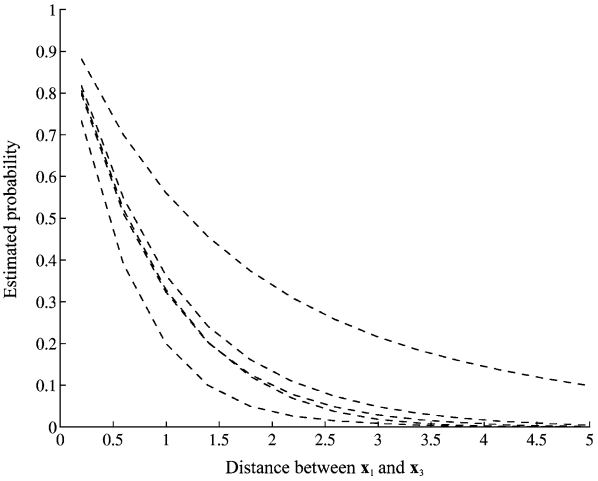

Fig. 4 Estimates of the probability that three points belong to a single cell of a Poisson tessellation. Estimates are calculated from realizations of the tessellation in a cubic domain with edge length $L = 15$ for the configuration shown in Fig. 2(**A**)

(that is, the realizations contain ten pixels per unit along each axis). One sees that the fluctuations of the experimental curves are much greater than the difference between the theoretical curves associated with the Poisson and dead leaves models (Fig. 3(A)). Numerical tests indicate that L should be greater than 3000 for the fluctuations to be less than half this difference in at least 95% of the realizations. In other words, using the trivariate distributions to discriminate between the two tessellations requires the availability of data in a domain with an edge length 1000 times larger than the range of correlation, or of a training image with this size. In practice, such a requirement is extremely difficult (if not impossible) to fulfill and makes evident the ergodicity problem mentioned in the introductory section.

4.2 Quadrivariate Distributions: Square Configuration

In this subsection, we restrict ourselves to the three tessellations (Poisson, iterated Poisson, and stable iterated) that are undistinguishable on the basis of their trivariate distributions, and we consider four locations $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ in the same plane, at the vertices of a square with edge length ℓ (Fig. 5(A)). To simplify the presentation, we will denote the four-point probabilities by the lowercase p followed by the location indices in subscript, separating by a comma the indices of the locations that do not belong to the same cell. For instance, $p_{14,2,3}$ is the probability that $\mathbf{x}_1$ and $\mathbf{x}_4$ are in the same cell, while $\mathbf{x}_2$ and $\mathbf{x}_3$ are in two separate cells.

For each model, the distribution of the tessellation at X is characterized by a set of 15 probabilities (Table 1). Furthermore, the cells are Poisson polyhedra, therefore $p_{1234} = \exp\{-2\ell\}$ ((17) and (18)); the cells are convex sets, therefore $p_{13,24} = 0$; the tessellation is isotropic, therefore $p_{12,34} = p_{14,23}$; and the 15 probabilities add to 1.

Accordingly, in order to compare the quadrivariate distributions of the tessellations under consideration, it suffices to examine only one probability (say, $p_{12,34}$). After some simple but long calculations based on the Poisson planes parameterization given in (1), one finds

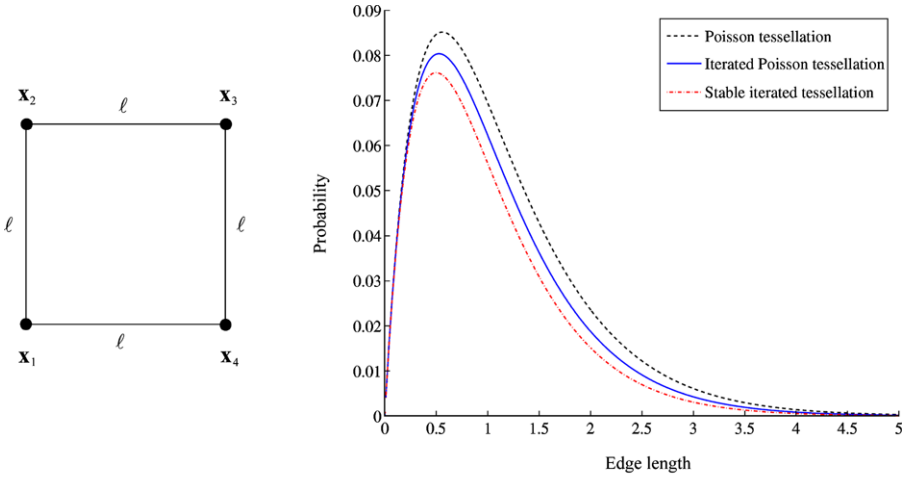

Fig. 5 (**A**) Square configuration in the plane and (**B**), probability $p_{12,34}$ as a function of the edge length (Poisson, iterated Poisson, and stable iterated tessellations)

Table 1 Probability system characterizing the distribution of a random tessellation at $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$

Probability	Relation with other probabilities
p_{1234}	
$p_{123,4}$	$p_{123} - p_{1234}$
$p_{124,3}$	$p_{124} - p_{1234}$
$p_{134,2}$	$p_{134} - p_{1234}$
$p_{234,1}$	$p_{234} - p_{1234}$
$p_{12,34}$	
$p_{13,24}$	
$p_{14,23}$	
$p_{12,3,4}$	$p_{12} - p_{1234} - p_{123,4} - p_{124,3} - p_{12,34}$
$p_{13,2,4}$	$p_{13} - p_{1234} - p_{123,4} - p_{134,2} - p_{13,24}$
$p_{14,2,3}$	$p_{14} - p_{1234} - p_{134,2} - p_{124,3} - p_{14,23}$
$p_{23,1,4}$	$p_{23} - p_{1234} - p_{123,4} - p_{234,1} - p_{14,23}$
$p_{24,1,3}$	$p_{24} - p_{1234} - p_{124,3} - p_{234,1} - p_{13,24}$
$p_{34,1,2}$	$p_{34} - p_{1234} - p_{134,2} - p_{234,1} - p_{12,34}$
$p_{1,2,3,4}$	

- Poisson tessellation

$$p_{12,34}^{\text{Poisson}} = \exp\left(-(3-\sqrt{2})\ell\right) - \exp(-2\ell); \tag{8}$$

- iterated Poisson tessellation

$$p_{12,34}^{\text{Iterated Poisson}} = 2\exp(-\ell)\left\{\exp\left(-\frac{3-\sqrt{2}}{2}\ell\right) - \exp(-\ell)\right\}; \tag{9}$$

- stable iterated tessellation

$$p_{12,34}^{\text{Stable Iterated}} = \exp(-2\ell)\ell(\sqrt{2}-1). \tag{10}$$

Although all the three-point statistics are identical, the four-point statistics based on a square configuration allow a distinction between the models (Fig. 5(B)). By repeating the same exercise as in Sect. 4.1, it is found that a training image with size $120 \times 120 \times 120$ (that is, an edge length 40 times larger than the range of correlation, which is still quite large in practice) would be required to experimentally differentiate the three tessellation models on the basis of their quadrivariate distributions.

4.3 Multivariate Distributions: Aligned Points

Suppose now that X consists of a set of points distributed along a line Δ. In such a case, the distributions of the Poisson, iterated Poisson, and stable iterated tessellations at X are exactly the same. Indeed, for each tessellation, the typical cells are Poisson polyhedra and the intersection of Δ with the cell boundaries form a Poisson point process: the tessellation restricted to Δ is a partition of the line into exponentially distributed intervals. In summary, along a line crossing the space, the above-mentioned tessellations cannot be discriminated.

4.4 Multivariate Distributions: Line Segment and Ball

To complete the analysis, two spatial configurations containing an infinite non-countable number of points are examined: a line segment and a ball. In both cases, we are interested in the probability that the configuration is contained in a single cell of the tessellation. This probability is the same in the Poisson, iterated Poisson, and stable iterated tessellations because the typical cells are the same; as such, only the Poisson and dead leaves tessellations will be compared.

First, let us consider the case of a line segment X with length ℓ. The probability that X is contained in a single cell of the Poisson tessellation, as shown in (17) in Appendix 1, is

$$\text{Prob}\{X \subseteq 1 \text{ cell}\}_{\text{Poisson}} = \exp\{-\ell\}. \tag{11}$$

Concerning the dead leaves tessellation, as shown in (19) in Appendix 2, one has

$$\text{Prob}\{X \subseteq 1 \text{ cell}\}_{\text{Dead leaves}} = \frac{E\{V_I(\ell; R)\}}{E\{V_U(\ell; R)\}}, \tag{12}$$

where R is a random diameter drawn from the distribution given in (3), $V_I(\ell; R)$ is the volume of the intersection of two balls centered at the endpoints of X and with diameter R, and $V_U(\ell; R)$ is the volume of the union of these two balls and the joining cylinder. This leads to the following expression

$$\text{Prob}\{X \subseteq 1 \text{ cell}\}_{\text{Dead leaves}} = \frac{K(\ell)}{K(0) + \frac{\pi\ell}{4}\int_0^{+\infty} r^2 f(r)\, dr}. \tag{13}$$

The difference between the Poisson and dead leaves tessellations (Fig. 6(A)) is slightly greater than in the case of three aligned points (Fig. 3(A)). In practice, the line

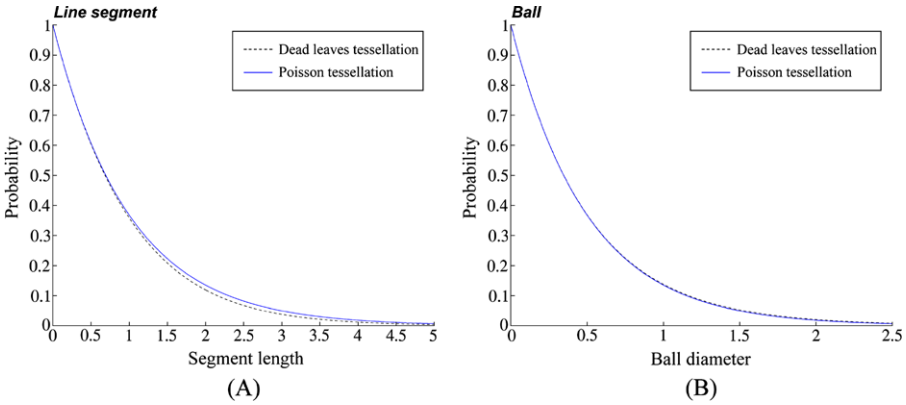

Fig. 6 Probability that (**A**) a line segment, and (**B**) a ball contained in a single cell, as a function of the segment length or ball diameter (Poisson and dead leaves tessellations)

segment configuration can be investigated when working with data collected along drill holes, but it is not well suited when working with training images because of the discretization inherent to the image, which contrasts with the continuous nature of the segment.

We now turn to the case when X is a ball with diameter δ. From (17) and (19) one has

$$\begin{aligned} \text{Prob}\{X \subseteq 1 \text{ cell}\}_{\text{Poisson}} &= \exp\{-2\delta\} \\ \text{Prob}\{X \subseteq 1 \text{ cell}\}_{\text{Dead leaves}} &= \frac{\int_{\delta}^{+\infty} (r-\delta)^3 f(r)\, dr}{\int_{0}^{+\infty} (r+\delta)^3 f(r)\, dr}. \end{aligned} \tag{14}$$

As shown in Fig. 6(B), the differences between the Poisson and dead leaves tessellations are hardly perceptible. The use of statistics from line segments or arrays of aligned points therefore appears as the best option for experimentally distinguishing both tessellation models, although it requires a very large domain to attenuate ergodic fluctuations (Sect. 4.1).

5 Responses to Transfer Functions

Although the high-order statistics of the random field models under consideration may be hard to differentiate (as shown in the previous section), the responses to specific transfer functions can differ significantly. To support this assertion, two examples are presented next.

5.1 Change of Support

Consider a Poisson and a stable iterated tessellation in $\mathbb{R}^3$ with cell valuations equal to 1 or 0 (with probabilities p and $1-p$, respectively) and an isotropic correlation function with practical range 3 (2).

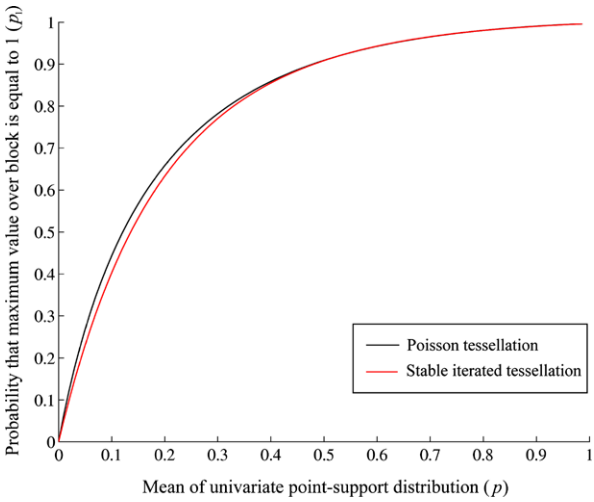

Fig. 7 Probability that the maximum point-support value over a cube is equal to 1 as a function of the mean value of the point-support distribution

Assume that one is interested in determining the maximum value of the tessellation over a cubic domain X with edge length ℓ. In order to differentiate the two tessellation models, the maximum has been considered rather than the average over X, insofar as the latter has the same expected value and variance in both models. The distribution of this maximum value is given by

$$p_1 = \text{Prob}\{\text{maximum value over } X = 1\} = E\left\{1 - (1-p)^N\right\}, \tag{15}$$

where N is the number of tessellation cells intersecting X. In each case (Poisson and stable iterated tessellations), the distribution of N can be assessed by constructing a large number of realizations of the tessellation on X, which yields an estimate of p_1.

The estimates of p_1 as a function of p, obtained by assuming a unit edge length (i.e., $\ell = 1$) and by using 10,000 realizations of both tessellations, are displayed in Fig. 7. The differences between the plotted curves are far from being negligible: for instance, for $p = 0.1$ one obtains $p_1 = 0.444$ for the Poisson tessellation and $p_1 = 0.403$ for the stable iterated tessellation.

5.2 Effective Permeability

Let us now consider four random set models, represented by indicator variables taking on the values 0 and 1 with equal probability. The first three models correspond to Poisson, stable iterated and dead leaves tessellations whose cells are valued independently (Fig. 8(A), (C), and (D)), while the fourth model is a Poisson tessellation in which the cells are valued alternatively (Fig. 8(B)). The random set indicators so defined possess the same univariate distribution (0/1), same bivariate distributions (characterized by an exponential correlation function), and also same trivariate distributions because of the autoduality property, that is, the distributions are unchanged when permuting the 0s and the 1s (Lantuéjoul 2002).

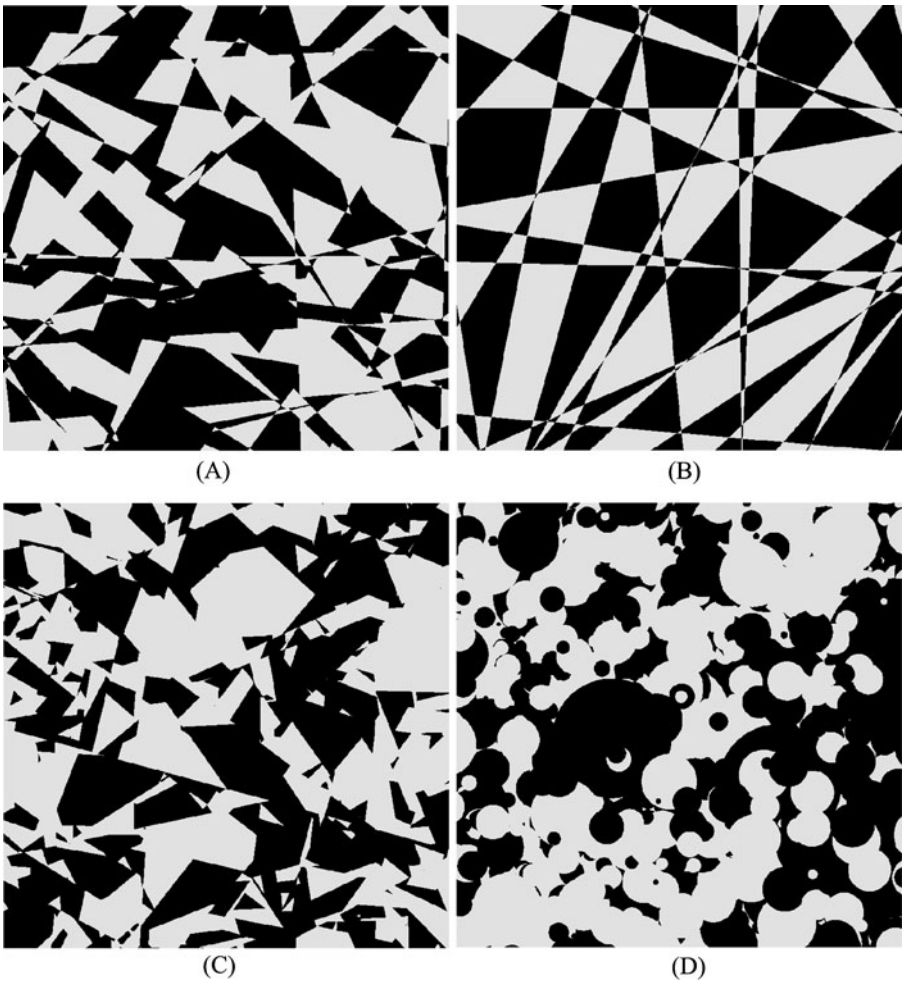

Fig. 8 Realizations of random set indicators with same univariate, bivariate and trivariate distributions: (**A**) Poisson tessellation, (**B**) alternating Poisson tessellation, (**C**) stable iterated tessellation, and (**D**) dead leaves tessellation

Suppose now that the indicators represent two rock types in a reservoir, for instance the 0s may stand for poor-quality shale with a permeability of 0.01 mD and the 1s for good-quality sandstone with a permeability of 100 mD. It is of interest to compare the four rock type models with respect to flow transport by calculating an equivalent permeability at a block support.

We constructed 500 realizations of each indicator variable over a two-dimensional regular grid with 200×200 nodes (the practical range of the correlation function has been set to 20 units). In each realization, single-phase flows have then been simulated by using the FLOWSIM program (Deutsch 1989) and the effective permeability along the abscissa direction has been calculated.

Table 2 Basic statistics on equivalent permeability along the abscissa direction, calculated from five hundred realizations over a square domain

Random set model	Average of simulated equivalent permeability (mD)	Median of simulated equivalent permeability (mD)
Poisson tessellation	1.83	0.78
Alternating Poisson tessellation	1.06	0.66
Stable iterated tessellation	2.85	1.04
Dead leaves tessellation	3.36	1.38

It is seen (Table 2) that the resulting permeability values greatly depend on the model under consideration. In the alternating Poisson tessellation (Fig. 8(B)), due to the sandstone/shale alternations, most of the flow paths go through shale, which provokes a considerable decrease in the effective permeability with respect to the other three models.

The cells of the Poisson (Fig. 8(A)) and stable iterated (Fig. 8(C)) tessellations are Poisson polygons, but their spatial arrangements are not the same: on average, the typical cell of the stable iterated tessellation is surrounded by more cells than that of the Poisson tessellation (the cells of the stable iterated tessellation are not face-to-face). This implies a smaller probability that sandstone be surrounded by shale in the stable iterated model, therefore having a higher effective permeability.

Concerning the dead leaves tessellation (Fig. 8(D)), the effective permeability is the highest one (on average, more than three times that of the alternating Poisson tessellation), as the curved shape of the tessellation cell boundaries yields a greater spatial connectivity of sandstone.

In summary, although they have the same univariate, bivariate, and trivariate distributions, the random set models under study are easily discernible on the basis of flow transport properties. Such aggregated statistics therefore appear as an interesting alternative to multiple-point statistics (in practice, limited to a few points) for discriminating between contending models.

6 Conclusions

One question that motivated this work was to know whether low-order distributions could characterize a random field or leave little uncertainty in higher-order distributions and in the responses to transfer functions. The answer to this question greatly depends on the type of random field under consideration. For instance, for diffusion-type random fields, bivariate distributions probably contain much information, if not all. In particular, it can be shown (Appendix 3) that if a random field has bivariate Gaussian distributions and a stationary cosine covariance function, then it is necessarily a multivariate Gaussian random field. This leaves open the question of whether or not there exist stationary random fields with bivariate Gaussian distributions other than multivariate Gaussian random fields.

In contrast, for random tessellations and random sets, low-order distributions convey relatively little information, since we have presented several such models that cannot be distinguished on the basis of their bivariate or trivariate distributions and that produce very different outcomes concerning change of support or flow simulation. In other words, having specified the distributions up to a finite order, there may still be a large indetermination about the underlying random field and about its response to specific transfer functions.

These statements and the results of the previous sections call for the following comments:

(1) The inference of high-order distributions may require the availability of very large training images or data sets, which is generally out of reach. As seen with the Poisson and dead leaves tessellations, ergodic fluctuations are too important to identify the experimental high-order statistics with the model statistics and to discriminate between contending models, unless the size of the training image is greater than one thousand times the range of correlation (a condition that is generally not met in practice).
(2) Simulation algorithms based only on low-order distributions (e.g., sequential indicator simulation) or on slightly misspecified high-order distributions may fail at reproducing the desired random field and accurately predicting its response to a transfer function. This situation has been highlighted with the tessellation models presented in this paper and may therefore arise in other practical cases.
(3) Because the inference from data or from training images is restricted to low-order distributions, the choice of a particular random field model remains a decision of the practitioner to a great extent. Following the examples presented in this study, in order to reduce the indetermination on the model as much as possible, our advice is to avoid examining only directional statistics (that is, statistics on aligned points) and restricting multiple-point spatial configurations to small distances (that is, points close to each other, compared with the range of correlation), insofar as contending models may not be distinguishable on the basis of such configurations (Figs. 3, 5, and 6).
(4) It is necessary to investigate whether other kinds of statistics (not only multiple-point) could improve the statistical differentiation between models, in particular aggregated statistics such as change-of-support statistics or flow properties. Also, instead of considering the distribution on a limited number of points, one may examine the distribution over a volume containing an infinite non-countable number of points (this requires the availability of observations over non-point supports). For example, the tessellations under consideration can be simply discriminated within any compact volume by examining the boundary edges: in the Poisson tessellation, each edge is the crossing of two planes and therefore always meets four faces (Fig. 1(A)); in the stable iterated tessellation, each edge only meets three faces (Fig. 1(C)); in the iterated Poisson tessellation, the edges can meet either three or four faces (Fig. 1(B)); as for the dead leaves tessellation, it is the only one with curved boundaries (Fig. 1(D)).
(5) The alternative to training images is the recourse to process oriented models, so that the realizations have the generic properties of the desired process without depending on finite-order statistics that are insufficient to specify these properties.

Acknowledgements This research was funded by the Chilean Commission for Scientific and Technological Research through FONDECYT projects 1061260 and 1090056. The authors acknowledge the support of the Advanced Laboratory for Geostatistical Supercomputing (ALGES) at the University of Chile.

Appendix 1: Poisson Tessellation

Let X be a convex polyhedron in $\mathbb{R}^3$ and consider a Poisson tessellation with intensity $\theta = \frac{1}{\pi}$. By using the parameterization of the Poisson planes given in (1), one can establish that the number N_X of planes crossing X is a Poisson random variable with mean value (Lantuéjoul 2002)

$$E(N_X) = 2b_3(X), \tag{16}$$

where $b_3(X)$ is the mean breadth of X. Accordingly, the probability that X is contained in a single cell of the tessellation, that is, the probability that N_X is equal to zero, is

$$\text{Prob}\{X \subseteq 1 \text{ cell}\}_{\text{Poisson}} = \exp\bigl\{-2b_3(X)\bigr\}. \tag{17}$$

If X is a convex polygon, then

$$b_3(X) = \frac{1}{4}|\partial X|, \tag{18}$$

where $|\partial X|$ stands for the length of the boundary of X; this expression can be derived from that of the mean breadth of X in the plane (Lantuéjoul 2002). Equations (4) and (11) are particular cases of (18) for a triangle and for a line segment.

The correlation function at lag vector $\mathbf{h}$ is the probability that two locations separated by $\mathbf{h}$ belong to a single cell of the tessellation (Rivoirard 1994). Because the cells are convex and isotropic, this is the same as the probability that a segment of length $\|\mathbf{h}\|$ is contained in a single cell, which yields (2).

Appendix 2: Dead Leaves Tessellation

Let X be a compact set in $\mathbb{R}^3$ and A be a primary grain of the dead leaves tessellation. The probability that X is contained in a single cell of the tessellation is (Jeulin 1997)

$$\text{Prob}\{X \subseteq 1 \text{ cell}\} = \frac{E\{\text{measure}(A \text{ eroded by } X)\}}{E\{\text{measure}(A \text{ dilated by } X)\}}, \tag{19}$$

where $(A \text{ eroded by } X) = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} - \mathbf{y} \in A, \mathbf{y} \in X\}$, $(A \text{ dilated by } X) = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in X, \mathbf{y} \in A\}$.

In particular, if $X = \{\mathbf{x}, \mathbf{x}+\mathbf{h}\}$ consists of two locations separated by vector $\mathbf{h}$, one finds the correlation function of the tessellation random field at lag $\mathbf{h}$,

$$\forall \mathbf{h} \in \mathbb{R}^3, \quad \rho(\mathbf{h}) = \frac{K(\mathbf{h})}{2K(\mathbf{0}) - K(\mathbf{h})}, \tag{20}$$

where $K(\mathbf{h})$ is the geometric covariogram of the primary grain at lag $\mathbf{h}$. Hereunder, as only the isotropic case is examined, the geometric covariogram will be denoted as a function of the vector length $r = \|\mathbf{h}\|$ instead of $\mathbf{h}$.

According to (20), the dead leaves tessellation has an isotropic exponential correlation function if the geometric covariogram of the primary grains is of the form

$$\forall \mathbf{h} \in \mathbb{R}^3, \quad K(r) = \frac{2K(0)}{1+\exp(r)} \quad \text{with } r = \|\mathbf{h}\|. \tag{21}$$

Let us assume that the primary grain is a ball with a random diameter in $\mathbb{R}^3$. In this case, its geometric covariogram is a mixture of isotropic spherical covariograms

$$K(r) = \int_r^{+\infty} \frac{\pi u^3}{6}\left(1 - \frac{3}{2}\frac{r}{u} + \frac{1}{2}\frac{r^3}{u^3}\right) f(u)\, du, \tag{22}$$

where f is a probability density function to be determined. Let us denote by F the associated cumulative density function. By differentiating twice with respect to r and by identifying with (21), one has

$$1 - F(r) = \frac{2K''(r)}{\pi r} = -\frac{4K(0)}{\pi r}\exp(r)\frac{1-\exp(r)}{[1+\exp(r)]^3}. \tag{23}$$

Because $F(0) = 0$, one finally obtains

$$F(r) = 1 + \frac{8}{r}\exp(r)\frac{1-\exp(r)}{[1+\exp(r)]^3} = 1 - \frac{2\sinh(r/2)}{r\cosh^3(r/2)}. \tag{24}$$

The probability density function f is given in (3). To simulate a random variable with such a density, an acceptance–rejection algorithm can be used (Lantuéjoul 2002) by noting the following inequality

$$\forall r \geq 0, \quad f(r) \leq 3\exp(-r). \tag{25}$$

Appendix 3: Random Fields with Bi-Gaussian Distributions and Cosine Covariance

In this appendix, we establish that a stationary random field $\{Y(\mathbf{x}) : \mathbf{x} \in \mathbb{R}\}$ with bivariate Gaussian distributions and cosine covariance function $\rho(\mathbf{h}) = \cos(\mathbf{h})$ is necessarily a multivariate Gaussian random field. Let us consider three locations, $\mathbf{x}$, $\mathbf{x}+\mathbf{h}$ and $\mathbf{x}+\mathbf{h}'$. The regression of $Y(\mathbf{x}+\mathbf{h})$ upon $Y(\mathbf{x})$ gives

$$Y(\mathbf{x}+\mathbf{h}) = \cos(\mathbf{h})Y(\mathbf{x}) + \sin(\mathbf{h})V(\mathbf{x}, \mathbf{x}+\mathbf{h}), \tag{26}$$

with $V(\mathbf{x}, \mathbf{x}+\mathbf{h})$ a standard Gaussian random variable independent of $Y(\mathbf{x})$. Similarly,

$$Y(\mathbf{x}+\mathbf{h}') = \cos(\mathbf{h}')Y(\mathbf{x}) + \sin(\mathbf{h}')V(\mathbf{x}, \mathbf{x}+\mathbf{h}'). \tag{27}$$

Now the covariance between $Y(\mathbf{x}+\mathbf{h})$ and $Y(\mathbf{x}+\mathbf{h}')$ is equal to $\cos(\mathbf{h}-\mathbf{h}')$. From (26) and (27) it ensues

$$\operatorname{cov}\left\{V(\mathbf{x},\mathbf{x}+\mathbf{h}),V(\mathbf{x},\mathbf{x}+\mathbf{h}')\right\}=1. \tag{28}$$

Hence $V(\mathbf{x},\mathbf{x}+\mathbf{h})=V(\mathbf{x},\mathbf{x}+\mathbf{h}')$ for any $\mathbf{h}$ and $\mathbf{h}'$ in $\mathbb{R}$. Denoting this random variable by $V(\mathbf{x})$, and using (26) with $\mathbf{x}=\mathbf{0}$, one has

$$\forall \mathbf{h}\in\mathbb{R},\quad Y(\mathbf{h})=\cos(\mathbf{h})Y(\mathbf{0})+\sin(\mathbf{h})V(\mathbf{0}), \tag{29}$$

with $Y(\mathbf{0})$ and $V(\mathbf{0})$ being two independent standard Gaussian random variables. Under this condition, it is seen that any weighted average of variables of $\{Y(\mathbf{x}):\mathbf{x}\in\mathbb{R}\}$ is normally distributed, indicating that this random field has multivariate Gaussian distributions. In this particular example, the bivariate distributions convey all the information about the random field.

References

Andersson J, Shapiro AM, Bear J (1984) A stochastic model of a fractured rock conditioned by measured information. Water Resour Res 20(1):79–88

Bordenave C, Gousseau Y, Roueff F (2006) The dead leaves model: a general tessellation modeling occlusion. Adv Appl Probab 38(1):31–46

Chilès JP (1989) Three-dimensional geometric modelling of a fracture network. In: Buxton BE (ed) Geostatistical, sensitivity, and uncertainty methods for ground-water flow and radionuclide transport modeling. Battelle Press, Columbus, pp 361–385

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(2):133–146

Coster M, Chermant JL (2002) On a way to material models for ceramics. J Eur Ceram Soc 22(8):1191–1203

Daly C (2005) Higher order models using entropy, Markov random fields and sequential simulation. In: Leuangthong O, Deutsch CV (eds) Geostatistics Banff 2004. Springer, Dordrecht, pp 215–224

Dershowiz WS, Einstein HH (1988) Characterizing rock joint geometry with joint system model. Rock Mech, Rock Eng 21(1):21–51

Deutsch CV (1989) Calculating effective absolute permeability in sandstone/shale sequences. SPE Form Eval 4(3):343–348

Emery X (2005) Geostatistical simulation of random fields with bivariate isofactorial distributions by adding mosaic models. Stoch Environ Res Risk Assess 19(5):348–360

Fairclough ARN, Davies GA (1990) Poisson line processes in 2-space to simulate the structure of porous media: methods of generation, statistics and applications. Chem Eng Commun 92(1):23–48

Fleischer F (2007) Analysis and fitting of random tessellation models: applications in telecommunication and cell biology. PhD Dissertation, Universität Ulm, Fakultät für Mathematik und Wirtschaftswissenschaften

Furuyama M (1976) Applications of L-mosaic map to spatial analysis of urban map. Mem Fac Ind Arts, Kyoto Tech Univ, Sci Technol 25:113–134

Garrigues S, Allard D, Baret F (2007) Using first- and second-order variograms for characterizing landscape spatial structures from remote sensing imagery. IEEE Trans Geosci Remote Sens 45(6):1823–1834

Gloaguen C, Fleischer F, Schmidt H, Schmidt V (2006) Fitting of stochastic telecommunication network models, via distance measures and Monte Carlo tests. Telecommun Syst 31(4):353–377

Guardiano F, Srivastava M (1993) Multivariate geostatistics: Beyond bivariate moments. In: Soares A (ed) Geostatistics Tróia'92. Kluwer Academic, Dordrecht, pp 133–144

Guillot G, Kan-King-Yu D, Michelin J, Huet P (2006) Inference of a hidden spatial tessellation from multivariate data: application to the delineation of homogeneous regions in an agricultural field. J R Stat Soc, Ser C, Appl Stat 55(3):407–430

Hasegawa M, Tanemura M (1976) On the pattern of space division by territories. Ann Inst Stat Math 28(1):509–519

Helte A (1994) Fourth-order bounds on the effective conductivity for a system of fully penetrable spheres. Proc R Soc Lond, Ser A, Math Phys Eng Sci 445(1923):247–256

Jeulin D (1994) Fracture statistics models and crack propagation in random media. Appl Mech Rev 47(1):141–150

Jeulin D (1997) Dead leaves models: from space tessellation to random functions. In: Jeulin D (ed) Advances in theory and applications of random sets. World Scientific, Singapore, pp 137–156

Jeulin D (2000) Random texture models for material structures. Stat Comput 10:121–132

Kim HM, Mallick BK, Holes CC (2005) Analyzing non-stationary spatial data using piecewise Gaussian processes. J Am Stat Assoc 100(470):653–668

Lantuéjoul C (1991) Ergodicity and integral range. J Microsc 161(3):387–403

Lantuéjoul C (2002) Geostatistical simulation: models and algorithms. Springer, Berlin

Lee AB, Mumford D, Huang J (2001) Occlusion models for natural images: a statistical study of a scale-invariant dead leaves model. Int J Comput Vis 41(1–2):35–59

Matérn B (1979) The analysis of ecological maps as mosaics. In: Cormack RM, Ord JK (eds) Spatial and temporal analysis in ecology. Fairland International Cooperative Publishing House, Maryland, pp 271–288

Matheron G (1975) Random sets and integral geometry. Wiley, New York, p 261

Nagel W, Mecke J, Ohser J, Weiss V (2008) A tessellation model for crack patterns on surfaces. Image Anal Stereol 27:73–78

Nagel W, Weiss V (2005) Crack STIT tessellations: characterization of stationary random tessellations stable with respect to iteration. Adv Appl Probab 37(4):859–883

Nagel W, Weiss V (2008) Mean values for homogeneous STIT tessellations in 3D. Image Anal Stereol 27:29–37

Ortiz JM, Deutsch CV (2004) Indicator simulation accounting for multiple-point statistics. Math Geol 36(5):545–565

Ortiz JM, Lyster S, Deutsch CV (2007) Scaling multiple-point statistics to different univariate proportions. Comput Geosci 33(2):191–201

Powell MJD (1964) The volume internal to three intersecting hard spheres. Mol Phys 7(6):591–592

Rivoirard J (1994) Introduction to disjunctive kriging and nonlinear geostatistics. Oxford University Press, Oxford

Santaló LA (2004) Integral geometry and geometric probability. Cambridge University Press, Cambridge

Serra J (1982) Image analysis and mathematical morphology. Academic Press, London

Stoyan D, Kendall WS, Mecke J (1996) Stochastic geometry and its applications, 2nd edn. Wiley, New York

Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. Math Geol 34(1):1–21

Ukita J, Moritz RE (2000) Geometry and the deformation of pack ice: II. Simulation with a random isotropic model and implication in sea-ice rheology. Ann Glaciol 31:323–326