



High-Order Sequential Simulation via Statistical Learning in Reproducing Kernel Hilbert Space

Lingqing Yao^{1,2}  · Roussos Dimitrakopoulos² · Michel Gamache¹

Received: 2 April 2019 / Accepted: 20 November 2019 / Published online: 7 December 2019
© The Author(s) 2019

Abstract The present work proposes a new high-order simulation framework based on statistical learning. The training data consist of the sample data together with a training image, and the learning target is the underlying random field model of spatial attributes of interest. The learning process attempts to find a model with expected high-order spatial statistics that coincide with those observed in the available data, while the learning problem is approached within the statistical learning framework in a reproducing kernel Hilbert space (RKHS). More specifically, the required RKHS is constructed via a spatial Legendre moment (SLM) reproducing kernel that systematically incorporates the high-order spatial statistics. The target distributions of the random field are mapped into the SLM-RKHS to start the learning process, where solutions of the random field model amount to solving a quadratic programming problem. Case studies with a known data set in different initial settings show that sequential simulation under the new framework reproduces the high-order spatial statistics of the available data and resolves the potential conflicts between the training image and the sample data. This is due to the characteristics of the spatial Legendre moment kernel and the generalization capability of the proposed statistical learning framework. A three-dimensional case study at a gold deposit shows practical aspects of the proposed method in real-life applications.

Keywords Stochastic simulation · High-order spatial statistics · Statistical learning · Reproducing kernel · Multipoint simulation

✉ Lingqing Yao
yaolingqing@gmail.com

¹ Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montreal, QC H3T 1J4, Canada

² COSMO – Stochastic Mine Planning Laboratory, Department of Mining and Materials Engineering, McGill University, 3450 University Street, Montreal, QC H3A 2A7, Canada

1 Introduction

Stochastic simulations are used to quantify the spatial uncertainty in earth science or engineering applications. Since the early 1990s, the so-termed multipoint statistical simulation (MPS) methods (Guardiano and Srivastava 1993; Journel and Zhang 2006; Strebelle 2002) were first proposed to overcome the limitation of the second-order simulation approaches in reproducing the complex spatial patterns encountered in natural phenomena. Instead of using a theoretical variogram/covariance model, as is the case with conventional two-point geostatistical simulations, the MPS methods consider that the so-called training image (TI) contains the prior information of the spatial statistics or patterns of the attribute to be simulated. A spatial template is defined as a geometrical configuration of the relative locations among the multiple points, regardless of the coordinates. The known data within the spatial template at a certain location on the simulation grid acts as the conditioning data in the simulation and is termed a data event. Over the past decade, several state-of-the-art MPS algorithms have been proposed to improve the efficiency and reproduction of the curvilinear features (Mariethoz and Caers 2014; Remy et al. 2009).

An inherent limitation of the MPS algorithms is that the high-order spatial statistics of the available data are not systematically considered and are partly integrated in ad-hoc ways. This issue becomes more prominent when the spatial statistics of the TI and the sample data are different, leading to realizations conflicting with the spatial statistics of the sample data, especially when the latter data is relatively dense as is the case in mining applications (Goodfellow et al. 2012; Osterholt and Dimitrakopoulos 2007). As an alternative, high-order simulation methods are proposed to model a random field without any presumption of its probability distribution, and high-order spatial statistics are systematically incorporated in the model (Dimitrakopoulos et al. 2010; Mustapha and Dimitrakopoulos 2010a, b, 2011). The first algorithm of high-order simulation, HOSIM, approximates the probability density function (PDF) by the Legendre polynomial series through the so-called spatial cumulants (Dimitrakopoulos et al. 2010; Mustapha and Dimitrakopoulos 2010b, 2011). Further developments of the high-order simulation paradigm include the simulation of spatially correlated variables (Minniakhmetov and Dimitrakopoulos 2017) and the direct simulation at the block scale (de Carvalho et al. 2019). Most recently, Yao et al. (2018) proposed a new computational model of high-order simulation as a unified empirical function, which avoids CPU-demanding computations of expansion coefficients. Furthermore, a kernel function can be derived from this model and will be used in the present work.

A common issue that runs across all of the above-mentioned high-order simulation methods is that the approximation of the PDF by orthogonal polynomials cannot be guaranteed to be positive. The sensitivity of high-order polynomials to the rounding errors near the endpoints of the approximation weakens the convergence of polynomial series to a stable analytic function, as discussed in Minniakhmetov et al. (2018), who propose an approximation of the PDF using Legendre-like orthogonal splines as the basis functions, resulting in a significant improvement in numerical stability. As the deviation of the empirical statistics from the true expectation arises due to possible statistical conflicts between the sample data and the TI, the convergence of the approximation to the actual underlying PDF could be undermined. Under such a

circumstance, a postprocessing step has to be introduced to correct the approximation. For example, the correction procedure through interpolation around the points of negative densities is applied in Mustapha and Dimitrakopoulos (2010b).

The present work proposes a new high-order simulation framework based on statistical learning (Vapnik 1995, 1998), which deliberately mitigates the statistical conflicts between the sample data and the TI, and also overcomes the limitation of approximating the PDF with the orthogonal expansion series. Statistical learning theory (Vapnik 1998) develops a new learning paradigm to explore functional dependency from a given data set without relying on prior knowledge, which contrasts with the classical statistical methods that are based on parametric models. According to the learning paradigm, a target model needs to be learned from the available data set, which represents the training data. The so-called learning machine (Vapnik 1998) is frequently given as a set of functions, from which a specific learning model is selected to approximate the target model according to certain criteria.

To interpret high-order simulation in terms of statistical learning, the training data are regarded as the available data from the sample data and/or the TI. The target model is the probability distribution related to the random field of the spatial attributes. The learning model is the approximated PDF of the target probability distribution, from which the realizations can be generated. The learning process for high-order simulation is driven by matching the expected high-order spatial statistics of the target probability distribution to the high-order spatial statistics observed from the available data. The matching of the high-order spatial statistics is the most challenging part and is approached herein by a learning process in a reproducing kernel Hilbert space (RKHS) (Scholkopf and Smola 2001). A spatial Legendre moment (SLM) reproducing kernel is proposed to construct the specified RKHS (SLM-RKHS), such that the high-order spatial statistics are systematically incorporated in this Hilbert space for a certain probability distribution. The elements in the original data space are mapped into the SLM-RKHS, termed RKHS embedding (Muandet et al. 2016; Smola et al. 2007; Song et al. 2008, 2013). In addition, the high-order spatial statistics of the available data are carried over to the domain after this RKHS embedding. Eventually, the statistical learning regarding high-order simulation leads to a convex optimization in SLM-RKHS where the solutions amount to solving a quadratic programming problem.

In the following sections, the general theory of kernel methods, including the RKHS and RKHS embedding of probability distributions, are introduced. Section 3 describes the main workflow of high-order sequential simulation via statistical learning, and a spatial Legendre moment reproducing kernel is defined to construct the specific SLM-RKHS. Furthermore, this SLM-RKHS is decomposed to lower-dimensional subspaces, such that conditional probability density functions (CPDF) in the context of sequential simulation can be embedded into the corresponding subspaces. Subsequently, a high-order stochastic simulation method is presented as a learning process based on the embedding of the CPDF into the decomposed subspace of the SLM-RKHS. Next, the proposed simulation method is tested using a fully known data set. A case study at a gold deposit is then presented to show the practical aspects of the proposed method. Conclusions follow.

2 Methods

2.1 Overview of Kernel Space and Embedding a Probability Distribution

In the general setting of kernel methods, a kernel space needs to be set up and associated with a predefined kernel function, and a feature mapping is defined to map an arbitrary element from the original data space into the kernel space. The related general concepts and theory are formalized in the followed subsections.

2.1.1 Reproducing Kernel Hilbert Space

A Hilbert space \mathcal{H} is a vector space over a field endowed with an inner product (Stein and Shakarchi 2005). For simplicity, the Hilbert space \mathcal{H} over the set \mathbb{R} of real numbers is considered here, and the inner product is defined as

$$\langle f, g \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}, \quad \forall f, g \in \mathcal{H}.$$

The norm is defined as

$$\|f\|_{\mathcal{H}} = \langle f, f \rangle^{1/2}, \quad \forall f \in \mathcal{H}.$$

Other essential properties can be found in Stein and Shakarchi (2005). The concepts of reproducing kernel and positive definite function are from Berlinet and Thomas-Agnan (2004) with the modification of the range of kernel function to \mathbb{R} .

Reproducing Kernel Let \mathbb{E} be a non-empty set and \mathcal{H} be a Hilbert space of functions defined on \mathbb{E} . Then, a function $K: \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$ is a reproducing kernel of a Hilbert space \mathcal{H} if and only if

- (1) $\forall t \in \mathbb{E}, K(\cdot, t) \in \mathcal{H}$, and
- (2) $\forall t \in \mathbb{E}, \forall f \in \mathcal{H}, \langle f, K(\cdot, t) \rangle = f(t)$.

The last condition is called “the reproducing property,” because any function in \mathcal{H} can be reproduced by its inner product with the kernel K . In addition, as a direct derivation of the above conditions, the reproducing kernel can be written as the inner product

$$K(s, t) = \langle K(\cdot, s), K(\cdot, t) \rangle, \quad \forall s, t \in \mathbb{E}.$$

Naturally, a Hilbert space in possession of a reproducing kernel is called a reproducing kernel Hilbert space. The feature map associated with an RKHS \mathcal{H} with kernel K is defined as $\phi: \mathbb{E} \rightarrow \mathcal{H}$ such that $\langle \phi(s), \phi(t) \rangle = K(s, t)$. In fact, $\phi(t): \mathbb{E} \rightarrow \mathcal{H}, t \mapsto K(\cdot, t), \forall t \in \mathbb{E}$ satisfies such a definition as the feature map according to the reproducing property. This type of feature map is called a reproducing kernel map (Scholkopf and Smola 2001) or canonical feature map (Steinwart and Christmann 2008) and will be used in the present paper.

Positive Definite Function A real-valued function $K: \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$ is positive definite if $\forall n \geq 1, \forall (a_1, \dots, a_n) \in \mathbb{R}^n, \forall (x_1, \dots, x_n) \in \mathbb{E}^n$, there is

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j K(x_i, x_j) \geq 0$$

The reason for introducing the concept of the positive definite function is that a reproducing kernel is equivalently a positive definite function (Berlinet and Thomas-Agnan 2004). Thus, in practical terms, constructing an RKHS is equivalent to defining a positive definite function.

2.1.2 RKHS Embedding of a Probability Distribution

The range of the feature mapping spans RKHS \mathcal{H} by definition (Scholkopf and Smola 2001). Thus, the feature mapping ϕ is crucial in embedding a data element into the RKHS \mathcal{H} . Accordingly, two mappings are important to embed a probability distribution into the RKHS \mathcal{H} (Smola et al. 2007)

$$\mu [p] = E_{x \sim p} [\phi(x)], \tag{1}$$

and

$$\mu [X] = \frac{1}{M} \sum_{i=1}^M \phi(X_i), \tag{2}$$

where the first equation is the expectation kernel mean map regarding the density p and the second one is the empirical kernel mean map with the finite sample set $X = \{X_1, \dots, X_M\}$. The expectation kernel mean map $\mu[p]$ is an element in the RKHS \mathcal{H} as long as $E_{x \sim p} [K(x, x)] < \infty$ (Smola et al. 2007). Suppose that the samples from X are independently drawn from the same probability distribution with density p , then $\mu[p]$ can be approximated by $\mu[X]$ (Song et al. 2008), with the bound of the deviation $\|\mu[p] - \mu[X]\|_{\mathcal{H}}$ with the probability given by Altun and Smola (2006).

The space of all probability distributions forms a convex set \mathcal{P} ; thus, the image of the expectation kernel mean map $\mathcal{M} := \{\mu[p], \forall p \in \mathcal{P}\}$ is also convex and is called the marginal polytope (Smola et al. 2007). In terms of the RKHS embedding, the goal of the density estimation is to find an optimal probability density $\hat{p} \in \mathcal{P}$ such that the deviation $\|\mu[X] - \mu[\hat{p}]\|_{\mathcal{H}}$ is minimized. In practice, the density estimator \hat{p} is assumed as a mixture of a set of candidate densities or prototypes p_i (Smola et al. 2007; Song et al. 2008) as

$$\hat{p} = \sum_{i=1}^n \alpha_i p_i, \tag{3}$$

where $\sum_i^n \alpha_i = 1$ and $\alpha_i \geq 0, \forall 1 \leq i \leq n$.

Let us define the subset \mathcal{P}_0 of \mathcal{P} as

$$\mathcal{P}_0 := \left\{ \hat{p} = \sum_i^n \alpha_i p_i \mid \sum_{i=1}^n \alpha_i = 1 \text{ and } \alpha_i \geq 0, \quad \forall 1 \leq i \leq n \right\}.$$

It can be seen that \mathcal{P}_0 is a convex hull of the prototypes since \hat{p} is a convex combination of the candidate densities. The density estimation amounts to solving the minimization problem restricted to a convex set \mathcal{P}_0 as

$$\min_{\hat{p} \in \mathcal{P}_0} \|\mu[X] - \mu[\hat{p}]\|_{\mathcal{H}}^2. \tag{4}$$

Explicit expansion of Eq. (4) leads to solving a quadratic program for $\alpha = (\alpha_1, \dots, \alpha_n)$ as the following (Song et al. 2008)

$$\begin{aligned} & \min_{\alpha} \frac{1}{2} \alpha^T (\mathbf{Q} + \lambda \mathbf{I}) \alpha - \mathbf{q}^T \alpha \\ \text{s.t. } & \sum_{i=1}^n \alpha_i = 1, \\ & \alpha_i \geq 0, \quad \forall 1 \leq i \leq n \end{aligned} \tag{5}$$

where λ is a regularization constant to prevent overfitting, and \mathbf{I} is the identity matrix. $\mathbf{Q} = [Q_{ij}]_{n \times n}$ is a matrix, and $\mathbf{q} = (q_1, \dots, q_n)$ is a vector of length n , both of which are entries that depend on the kernel function. The matrix \mathbf{Q} is positive definite; hence the above quadratic program (5) is a convex optimization problem.

2.2 High-Order Simulation Method in Spatial Legendre Moment Kernel Space

2.2.1 SLM Reproducing Kernel

The motivation for applying statistical learning to the high-order simulation is to match the high-order spatial statistics of the output realizations to the training data through the learning process. This goal is achieved by the learning procedure in a newly defined kernel space, while the kernel is defined as

$$K(X, Y) = \prod_{i=0}^N \left[\sum_{w=0}^W \left(w + \frac{1}{2}\right) P_w(x_i) P_w(y_i) \right], \tag{6}$$

and is called a spatial Legendre moment kernel (SLM-kernel for short) of order W , where $X, Y \in [-1, 1]^{N+1}$, $X = (x_0, x_1, \dots, x_N)$, $Y = (y_0, y_1, \dots, y_N)$, and $P_w(\cdot)$ is the Legendre polynomial of order w defined on the interval $[-1, 1]$.

As the name of the kernel suggests, one reason to define the SLM-kernel in the form of Eq. (6) is that past studies of high-order simulations based on Legendre-polynomial series have shown the capacity for capturing complex spatial patterns with spatial cumulants or spatial Legendre moments (Dimitrakopoulos et al. 2010; Mustapha and Dimitrakopoulos 2010b; Yao et al. 2018). In other words, the SLM-kernel is constructed in a way that the distance between two distributions embedded

into the kernel space actually represent the deviation of spatial Legendre moments from each other. The other reason stems from the fact that the computational model from Yao et al. (2018) leads to a kernel-like expression of approximating the CPDF [cf. Eq. (14) in Yao et al. (2018)].

To prove that $K(\mathbf{X}, \mathbf{Y})$ is positive definite, one can first define a simpler function $k(s, t) = P_w(s) P_w(t), \forall s, t \in [-1, 1]$ and show that it is positive definite. In fact,

$$\forall n \geq 1, \quad \forall a_i, a_j \in \mathbb{R}, \quad \forall t_i \in [-1, 1], \quad 1 \leq i, j \leq n,$$

it is easy to see that

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j P_w(t_i) P_w(t_j) = \left[\sum_{i=1}^n a_i P_w(t_i) \right]^2 \geq 0$$

Therefore, $k(s, t)$ is positive definite. Now, we denote

$$\begin{aligned} K'(\mathbf{X}, \mathbf{Y}) &= \sum_{w=0}^W \left(w + \frac{1}{2} \right) P_w(x_i) P_w(y_i) \\ &= \sum_{w=0}^W \left(w + \frac{1}{2} \right) k(x_i, y_i) \end{aligned}$$

$K'(\mathbf{X}, \mathbf{Y})$ is positive definite because the weighted sum of positive definite functions with non-negative coefficients is also positive definite. Finally, $K(\mathbf{X}, \mathbf{Y})$ can be written as $K(\mathbf{X}, \mathbf{Y}) = \prod_{i=0}^N K'(\mathbf{X}, \mathbf{Y})$. Given that the finite product of positive definite functions is also positive definite (Steinwart and Christmann 2008), it is proven that the function $K(\mathbf{X}, \mathbf{Y})$ is positive definite, and thus, it defines a reproducing kernel.

2.2.2 Sequential Simulation via Statistical Learning in SLM-Kernel Space

The implementation of a high-order stochastic simulation is under the framework of a sequential simulation (Journel 1994). By means of decomposing the multivariate probability distribution into a consecutive set of univariate distributions, the simulation is carried out sequentially to generate random values from conditional distributions per a random path visiting the simulation grid. Specifically, let us denote the random field to be simulated as $\mathbf{Z}(\mathbf{u})$, which composes a multivariate distribution regarding the variable locations \mathbf{u} at a discrete simulation grid. Suppose an arbitrary node Z_0 to be simulated within a random path is located at \mathbf{u}_0 with a neighborhood Λ of N conditioning data that contains either the sample data or the previously simulated nodes along the random path. Without loss of generality, the key problem in the stochastic simulation is to find an estimation of the CPDF $f(Z_0|\Lambda)$, given the center node Z_0 and the N conditioning data. From the spatial configuration of the neighborhood, a spatial template can be extracted as $\mathbf{T} = (\mathbf{u}_0, \mathbf{u}_0 + h_1, \dots, \mathbf{u}_0 + h_N)$, where h_1, \dots, h_N are distance vectors of the location of each conditioning data from the center node \mathbf{u}_0 .

Clearly, statistical learning for the simulation aims to learn a target probability distribution from the available training data, and this turns out to be minimizing the distance of the empirical distribution and the target distribution after embedding them into the SLM-kernel space. By the definition of the Dirac delta function, one can define an empirical probability density function (EPDF) (Scott 2015) corresponding to a sample set X of size M as

$$f_{\text{emp}}(x) = \frac{1}{M} \sum_{i=1}^M \delta(x - X_i). \quad (7)$$

Then, the empirical kernel mean map $\mu[X]$ can be rewritten as a convolution with the kernel K as

$$\mu_K[f_{\text{emp}}] := \mu[X] = \int f_{\text{emp}}(x) K(x, \cdot) dx. \quad (8)$$

Similarly, the expectation kernel mean map $\mu[p]$ can also be written as

$$\mu_K[p] = \int p(x) K(x, \cdot) dx. \quad (9)$$

In this way, both the empirical kernel mean map $\mu[X]$ and the expectation kernel mean map $\mu[p]$ can be regarded as an integral operator μ_K determined by the kernel K acting on the EPDF or the PDF. The convolution of the density function with kernels can be analogous to the regularization of the integral operator to solve the ill-posed problem of density estimation (Vapnik 1998; Vapnik and Mukherjee 1999).

Given the above-mentioned template $\mathbf{T} = (\mathbf{u}_0, \mathbf{u}_0 + h_1, \dots, \mathbf{u}_0 + h_N)$ and the replicate encountered in the TI as $\boldsymbol{\zeta}_t = (\zeta_{t,0}, \zeta_{t,1}, \dots, \zeta_{t,N})$ corresponding to \mathbf{T} , the EPDF f_{emp} embedded in the SLM-RKHS is identical to the density estimator in Yao et al. (2018) in the kernel form as

$$\mu_K[f_{\text{emp}}] = \frac{1}{M} \sum_{t=1}^M K(\boldsymbol{\zeta}_t, \cdot). \quad (10)$$

Furthermore, under the sequential simulation framework, the CPDF $f(Z_0|\Lambda)$ can be mapped into a lower-dimensional kernel space through decomposition of the kernel space, so that the high-order simulation can be reduced to a one-dimensional optimization problem.

Note that the kernel K in Eq. (6) can be decomposed as a product of lower-dimensional kernels K_0 and K_N as

$$K_0(x_0, y_0) = \sum_{w=0}^W \left(w + \frac{1}{2}\right) P_w(x_0) P_w(y_0), \quad (11)$$

and

$$K_N(\mathbf{X}', \mathbf{Y}') = \prod_{i=1}^N \left[\sum_{w=0}^W \left(w + \frac{1}{2}\right) P_w(x'_i) P_w(y'_i) \right], \quad (12)$$

where K_0 is one-dimensional and K_N is N -dimensional with $\mathbf{X}' = (x_1, \dots, x_N)$, $\mathbf{Y}' = (y_1, \dots, y_N)$. Through marginalization of Eq. (10), the approximation of the CPDF $f_W(z_0|\Lambda)$ can be rewritten in terms of the kernels as

$$\tilde{f}_W(z_0|\Lambda) = \frac{\sum_{t=1}^M K_0(\zeta_{t,0}, z_0) \cdot K_N(\boldsymbol{\zeta}'_t, \Lambda)}{\sum_{t=1}^M K_N(\boldsymbol{\zeta}'_t, \Lambda)}, \tag{13}$$

where $\boldsymbol{\zeta}'_t = (\zeta_{t,1}, \dots, \zeta_{t,N})$. By letting

$$\beta_t = \frac{K_N(\boldsymbol{\zeta}'_t, \Lambda)}{\sum_{t=1}^M K_N(\boldsymbol{\zeta}'_t, \Lambda)}, \tag{14}$$

the approximation of the CPDF $\tilde{f}_W(z_0|\Lambda)$ can be expressed as

$$\tilde{f}_W(z_0|\Lambda) = \sum_{t=1}^M \beta_t \cdot K_0(\zeta_{t,0}, z_0). \tag{15}$$

From Eq. (15), it turns out that the approximated CPDF $\tilde{f}_W(z_0|\Lambda)$ is a linear combination of kernel bases, and therefore, it lies in the SLM-RKHS with the kernel K_0 . Furthermore, it can be regarded as the embedding of the empirical CPDF into the SLM-RKHS. In other words, the kernel mean map μ_{K_0} for the conditional distributions can be defined as

$$\mu_{K_0}[f_{\text{emp}}(z_0|\Lambda)] = \sum_{t=1}^M \beta_t \cdot K_0(\zeta_{t,0}, \cdot), \tag{16}$$

and

$$\mu_{K_0}[f(z_0|\Lambda)] = \int f(z_0|\Lambda) K_0(z_0, \cdot) dz_0 = E[K_0(z_0, \cdot)], \tag{17}$$

where Eqs. (16) and (17) correspond to the SLM-RKHS embedding of the empirical CPDF and the target CPDF, respectively.

Assuming that the CPDF can be expressed as the convex combination of some candidate distributions p_i as in Eq. (3), such that $f(z_0|\Lambda) \in \mathcal{P}_0$, then the density estimation for the CPDF can be solved by a similar minimization problem as Eq. (4) with the kernel mean map changing to μ_{K_0} . Explicit expansion of the minimization problem leads to a quadratic program similar to Eq. (5), whereas the matrix \mathbf{Q} and the vector \mathbf{q} are expressed as

$$Q_{ij} = E_{z_0 \sim p_i, z'_0 \sim p_j} [K_0(z_0, z'_0)], \tag{18}$$

$$q_j = \sum_{t=1}^M \beta_t \cdot E_{z_0 \sim p_j} [K_0(\zeta_{t,0}, z_0)]. \tag{19}$$

Therefore, combining Eqs. (5), (11), (18) and (19), the RKHS embedding of the CPDF leads to a quadratic program expressed by the one-dimensional kernel K_0 . The solution

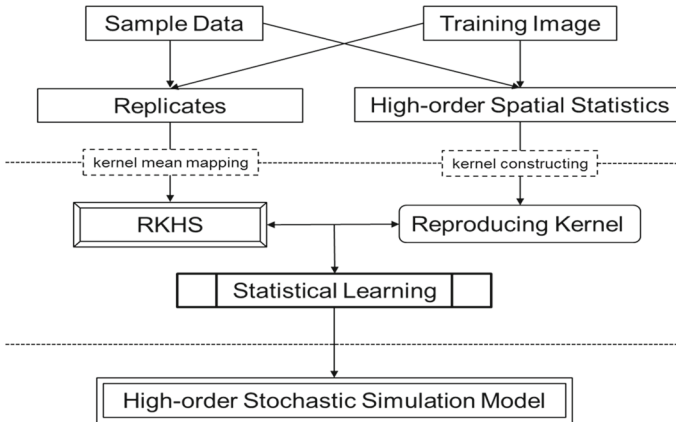


Fig. 1 Workflow of high-order simulation via statistical learning

to the optimization problem will give the weights α_i of each candidate distribution p_i , which leads to a target distribution matching to the high-order spatial statistics of the available data.

A general high-order stochastic simulation workflow via statistical learning is shown in Fig. 1. The main difference between the new high-order simulation workflow and the other geostatistical simulation methods is that the key element in the proposed workflow becomes the kernelization, including the kernel construction and the kernel mean mapping. A detailed implementation of the algorithm is given in Sect. 4.

3 Sequential Simulation Algorithm Based on Statistical Learning in SLM-RKHS

The SLM-RKHS embedding of the CPDF projects the density estimation in high-order stochastic simulation into a quadratic program in the feature space with SLM-kernel K_0 defined in the interval $[-1, 1]$. Hence, the sample data and the TI are first transformed into the interval $[-1, 1]$. The truncated normal densities on the interval $[-1, 1]$ are used as the prototypes. Let us denote the normal density with mean m_i and standard deviation σ as g_{σ, m_i} and its corresponding cumulative distribution function as G_{σ, m_i} . Then, the density functions of the prototypes are $p_i = g_{\sigma, m_i}/c_i$, with $c_i = G_{\sigma, m_i}(1) - G_{\sigma, m_i}(-1)$. Thus, the approximated CPDF can be expressed as

$$\hat{f}(z_0|\Lambda) = \sum_{i=1}^n \alpha_i g_{\sigma, m_i}(z_0)/c_i, \quad (20)$$

where n is the number of the prototypes. The computations of the matrix \mathbf{Q} and the vector \mathbf{q} are essential to build the quadratic program for solving the weights α_i . Further expansions of Q_{ij} and q_j in Eqs. (18) and (19) give

$$Q_{ij} = \sum_{w=0}^W \left(w + \frac{1}{2}\right) E_{z_0 \sim p_i} [P_w(z_0)] \cdot E_{z'_0 \sim p_j} [P_w(z'_0)], \tag{21}$$

$$q_j = \sum_{t=1}^M \beta_t \cdot \left(\sum_{w=0}^W \left(w + \frac{1}{2}\right) P_w(\zeta_{t,0}) E_{z_0 \sim p_j} [P_w(z_0)]\right). \tag{22}$$

As the computations of the coefficients β_t and the Legendre polynomial $P_w(\zeta_{t,0})$ are straightforward according to their definitions, the Legendre polynomial moment with the truncated normal density $E_{z_0 \sim p_i} [P_w(z_0)]$ remains the only term of more consideration. Here, a recursive algorithm to compute the Legendre polynomial moment $E_{z_0 \sim p_i} [P_w(z_0)]$ is developed.

Let us denote $A_{w,i} = E_{z_0 \sim p_i} [P_w(z_0)]$ and $B_{w,i} = E_{z_0 \sim p_i} [z_0 P_w(z_0)]$. Note that $P_0(z_0) = 1$, and $P_1(z_0) = z_0, \forall z_0 \in [-1, 1]$. There are

$$A_{0,i} = 1, \tag{23}$$

and

$$A_{1,i} = B_{0,i} = m_i + \sigma^2 [g_{\sigma,m_i}(-1) - g_{\sigma,m_i}(1)] / c_i. \tag{24}$$

The recursive relations of Legendre polynomials (Lebedev and Silverman 1965) are

$$(w + 1) P_{w+1}(z_0) = (2w + 1) z_0 P_w(z_0) - w P_{w-1}(z_0), \tag{25}$$

and

$$(2w + 1) P_w(z_0) = \frac{d}{dz_0} [P_{w+1}(z_0) - P_{w-1}(z_0)]. \tag{26}$$

By Eqs. (25) and (26) and through integration by parts, one can derive the following recursive equations

$$(w + 1) A_{w+1,i} = (2w + 1) B_{w,i} - w A_{w-1,i}, \tag{27}$$

and

$$B_{w,i} = m_i A_{w,i} + \sigma^2 [(-1)^w g_{\sigma,m_i}(-1) - g_{\sigma,m_i}(1)] / c_i + \sigma^2 [2(w - 1) + 1] A_{w-1,i} + \sigma^2 [2(w - 3) + 1] A_{w-3,i} + \dots \tag{28}$$

Combining with the initial conditions in Eqs. (23) and (24), Eqs. (27) and (28) form a complete recursive procedure to compute $E_{z_0 \sim p_i} [P_w(z_0)]$. The computations in turn build the quadratic program for density estimation of the conditional probability distribution in the simulation.

In a situation with high-dimensional space, the location parameters m_i of the prototypes can be determined by clustering the available data. Here, since the density estimation problem is cast to the one-dimensional space by kernel decomposition, the locations of the prototypes are given by a set of peak points of the function from

Eq. (15). Specifically, the interval $[-1, 1]$ is divided evenly into 100 subintervals, and the prototypes are selected from the subintervals which contain the peak points of the function Eq. (15). This heuristic approach to selecting prototypes further simplifies the quadratic program and makes the simulation feasible for implementation. The scale parameter σ can be chosen by the method of stochastic gradient descent where the gradients can be derived from the recursive equations in Eqs. (27) and (28).

In summary, the high-order stochastic simulation algorithm based on RKHS embedding (KERNELSIM hereafter for simplification) can be described as follows:

- (1) Scale the property values of the samples and the TI to the interval $[-1, 1]$.
- (2) Generate a random path to visit the simulation grid.
- (3) Pick one node from the random path to simulate, with the conditioning data taken from the neighborhood containing both the sample data and the previously simulated nodes.
- (4) Replicates are scanned from the TI according to the template defined by the spatial configuration of the conditioning data.
- (5) Compute the SLM-kernel moments to build the quadratic program.
- (6) Solve the quadratic program to estimate of the CPDF with regard to the center node. Draw a random value from the CPDF as the data value of the center node.
- (7) Repeat from step (3) until the simulation is completed.
- (8) Back transform the property values of the simulation from $[-1, 1]$ to the original data space.

In a practical implementation, step (5) can be simplified to precompute the Legendre polynomial moments for each prototype distribution, as well as the Legendre polynomial values of the replicates, and therefore the computations can be greatly reduced at the cost of more memory usage. The solver for the quadratic program in step (6) applied to the present paper is based on the algorithm from Goldfarb and Idnani (1983).

The time complexity of the proposed algorithm is of polynomial time overall. Suppose that the size of the simulation grid is S and the size of the training data is M , the maximum order of the Legendre moments is W , the maximum number of conditioning data is N , and the number of the prototype distributions is n_p . Searching the replicates of the conditioning data from a regular grid takes $O(M \cdot N)$ operations. Computing the kernel moments and building the quadratic program takes $O(M \cdot n_p (W^3 + W^2N))$ arithmetic operations. Solving the quadratic program problem also takes polynomial time of $O(n_p^4 \cdot L)$, where L is the size of the problem encoding in binary (Vavasis 2001). Hence, the overall time complexity is a polynomial of $O(S \cdot (M \cdot n_p (W^3 + W^2N) + n_p^4 \cdot L))$.

4 Case Studies

4.1 Case Study at a Fully Known Reservoir

The porosity attributes from the Stanford V reservoir data set (Mao and Journel 1999) are considered for simulation. Two horizontal sections at different depths are extracted

Fig. 2 Exhaustive image: a horizontal section from a fully known reservoir

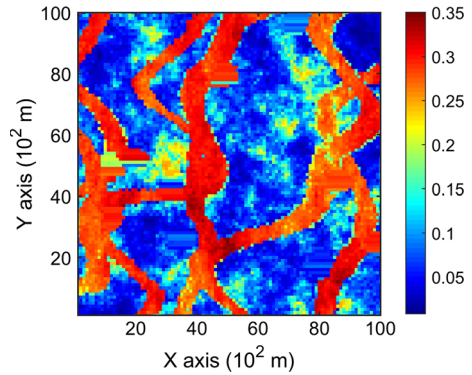
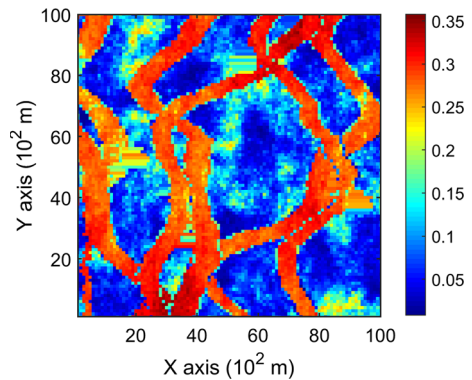


Fig. 3 TI-1: another horizontal section from a fully known reservoir



from the reservoir, acting as the exhaustive image and the TI, respectively. For comparison, the two horizontal sections shown in Figs. 2 and 3 are selected to be the same ones used in a previous study (Yao et al. 2018). Firstly, the TI extracted from the original reservoir data set is rotated 45° clockwise to generate a new TI with seemingly different spatial structures, which are noted as TI-1 and TI-2 (Figs. 3, 4), respectively. Furthermore, two different sets of sample data as DS-1 and DS-2 are drawn from the exhaustive image and are shown in Figs. 5 and 6, corresponding to the relatively sparse and dense samples, respectively. The main purposes of performing a simulation on these different cases are: (1) testing the sensitivity of KERNELSIM to the statistical conflicts between the sample data and the TI; (2) testing the impact of the number of sample data on the realization of KERNELSIM.

4.1.1 Example 1

This example consists of simulation results generated by KERNELSIM with the TI-1 as the training image and DS-1 and DS-2 as the sample data sets. This example generally represents the situation where the sample data and the TI are of different origin but are sharing some similarity in spatial patterns. For instance, the channels in both the exhaustive image and TI-1 are preferential in the vertical directions.

Fig. 4 TI-2: rotation of TI-1 45° clockwise

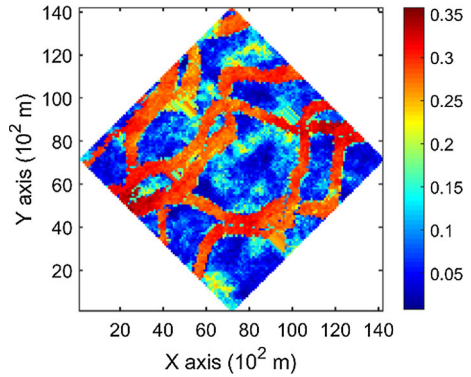


Fig. 5 DS-1: data samples of 200 points drawn from the exhaustive image

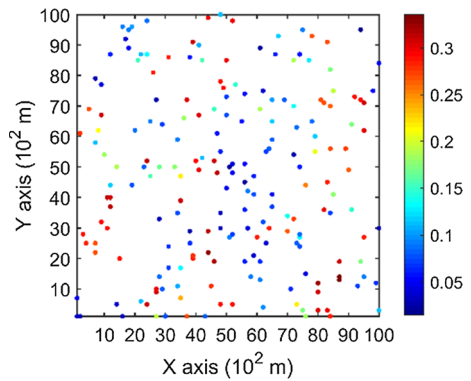


Fig. 6 DS-2: data samples of 400 points drawn from the exhaustive image

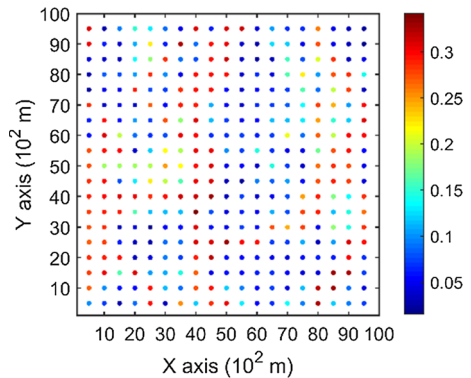


Figure 7 shows one realization of KERNELSIM using TI-1 as the training image and with DS-1 and DS-2 as the sample data, respectively. For comparison, both realizations are generated by the same random path to visit the nodes on the grid. It is clear that both realizations reproduce the main spatial structures of the exhaustive image along the vertical channels from the visualization (Fig. 7). The realization shown in Fig. 7a is comparable to the case study in Yao et al. (2018), and it shows that

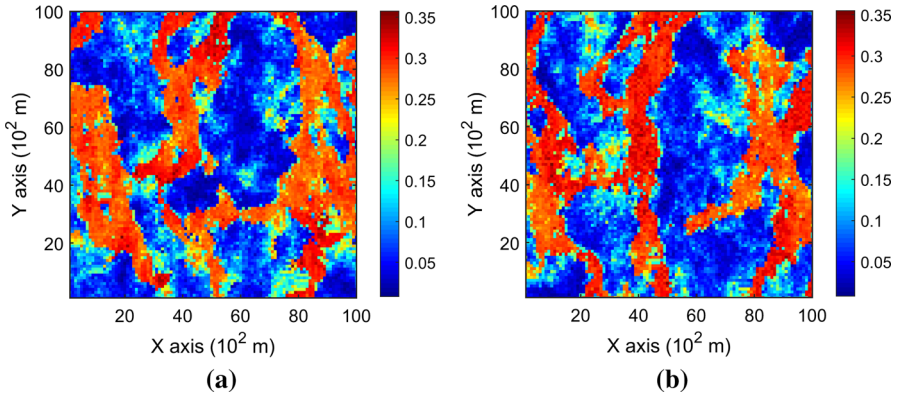


Fig. 7 One realization from KERNELSIM using TI-1. **a** DS-1 as the sample data, **b** DS-2 as the sample data

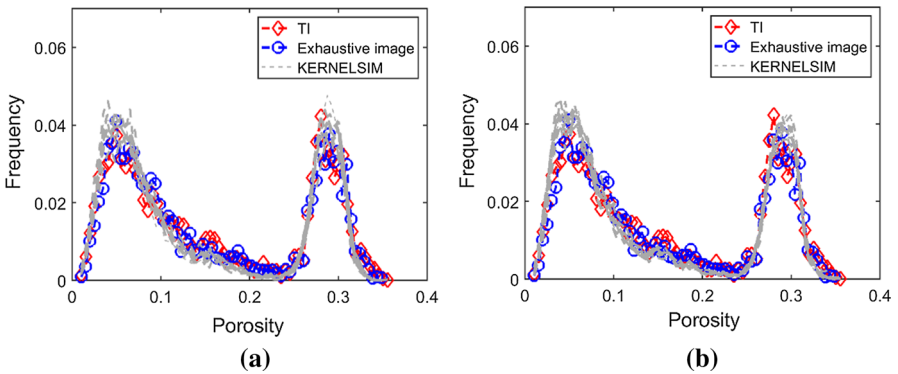


Fig. 8 Histograms of 10 realizations of KERNELSIM using TI-1. **a** DS-1 as the sample data, **b** DS-2 as the sample data

the present method reproduces channel connectivity better and eliminates the noisy points that appeared in the realizations generated using past approaches, which were caused by the impact of statistical conflicts between the sample data and the TI. Comparisons of the histograms and variograms of 10 realizations of KERNELSIM using either DS-1 or DS-2 as the sample data are illustrated in Figs. 8 and 9, respectively. The third-order cumulant maps of the sample sets DS-1 (smoothed for visualization) and DS-2 are shown in Fig. 10a, b. The cumulant maps of the exhaustive image and the TI are shown in Fig. 10c, d. For comparison, the third-order cumulant maps of the realizations of KERNELSIM using either the DS-1 or DS-2 as the sample data are shown in Fig. 10e, f. Figure 10g, h shows the average third-order cumulant maps of 10 realizations using the DS-1 and DS-2 as the sample data, respectively. Similarly, a further comparison of fourth-order cumulant maps is displayed in Fig. 11. The spatial template for computing the fourth-order cumulant maps included directions along the X-axis, Y-axis and the diagonal direction. The fourth-order cumulant maps are scaled by their deviations for clearer visualization of the patterns. Both the third-order and

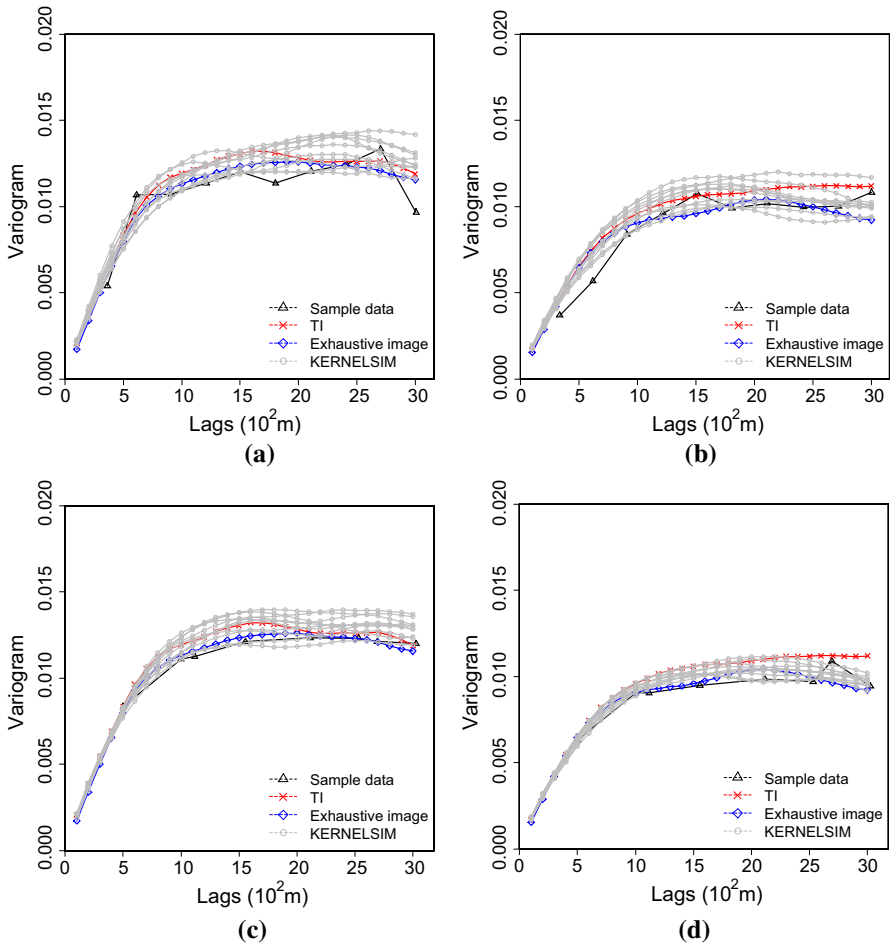


Fig. 9 Variograms of 10 realizations of KERNELSIM using TI-1. **a, b** Along the X and Y axes with DS-1 as the sample data; **c, d**, along the X and Y axes with DS-2 as the sample data

the fourth-order cumulant maps clearly show that the KERNELSIM realization tends to have similar spatial patterns to the sample data and the exhaustive image. The above results show that the KERNELSIM method reproduces both the lower and higher spatial statistics of the underlying random field given that the TI and the sample data share some similarity in their spatial distributions. Specifically, regardless the number of sample data being used, the main spatial features of the reservoir are retained in the realizations of KERNELSIM, as supported from the visual appearance of the vertical channels and the variograms, as well as from the cumulant maps.

4.1.2 Example 2

By rotating the TI-1 45° clockwise and creating a new training image as TI-2, shown in Fig. 4, there is seemingly a difference in the channel orientations between the TI-2

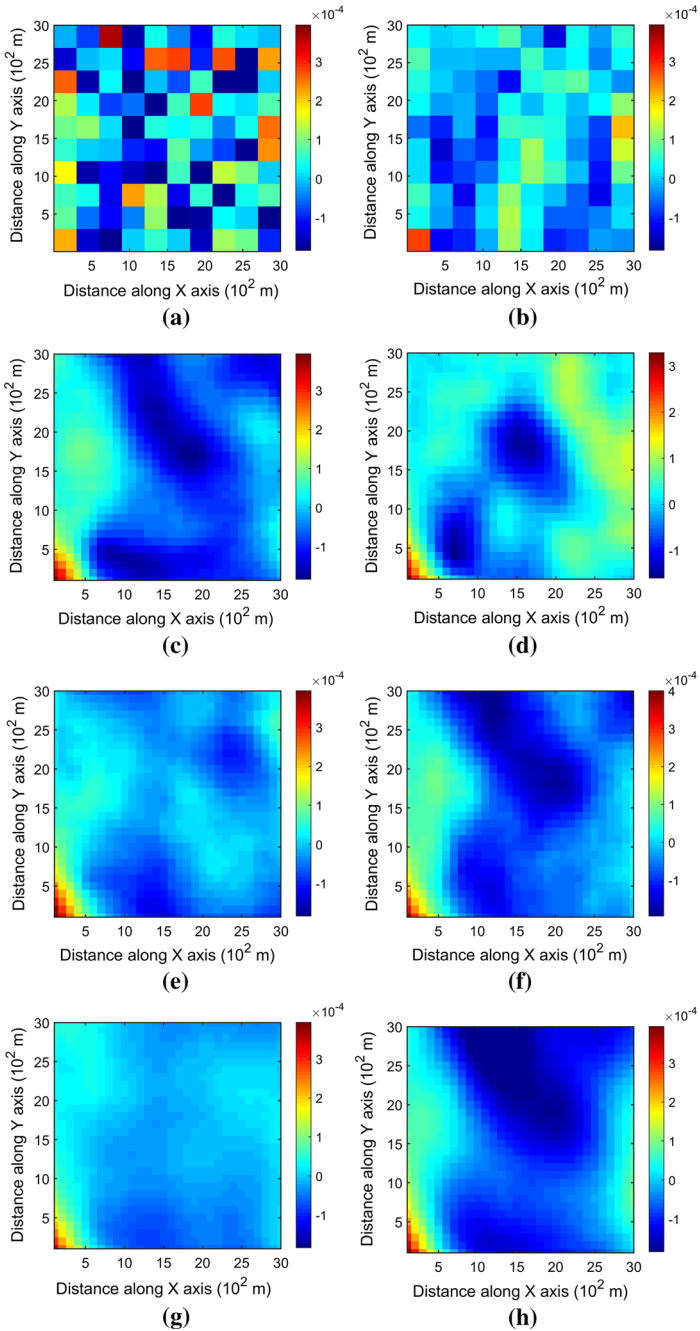


Fig. 10 Third-order cumulant maps of **a** DS-1, **b** DS-2, **c** exhaustive image, **d** TI-1, **e** realization in Fig. 7a with DS-1 as the sample data, **f** realization in Fig. 7b with DS-2 as the sample data, **g** 10 realizations in average with DS-1 as the sample data, and **h** 10 realizations in average with DS-2 as the sample data

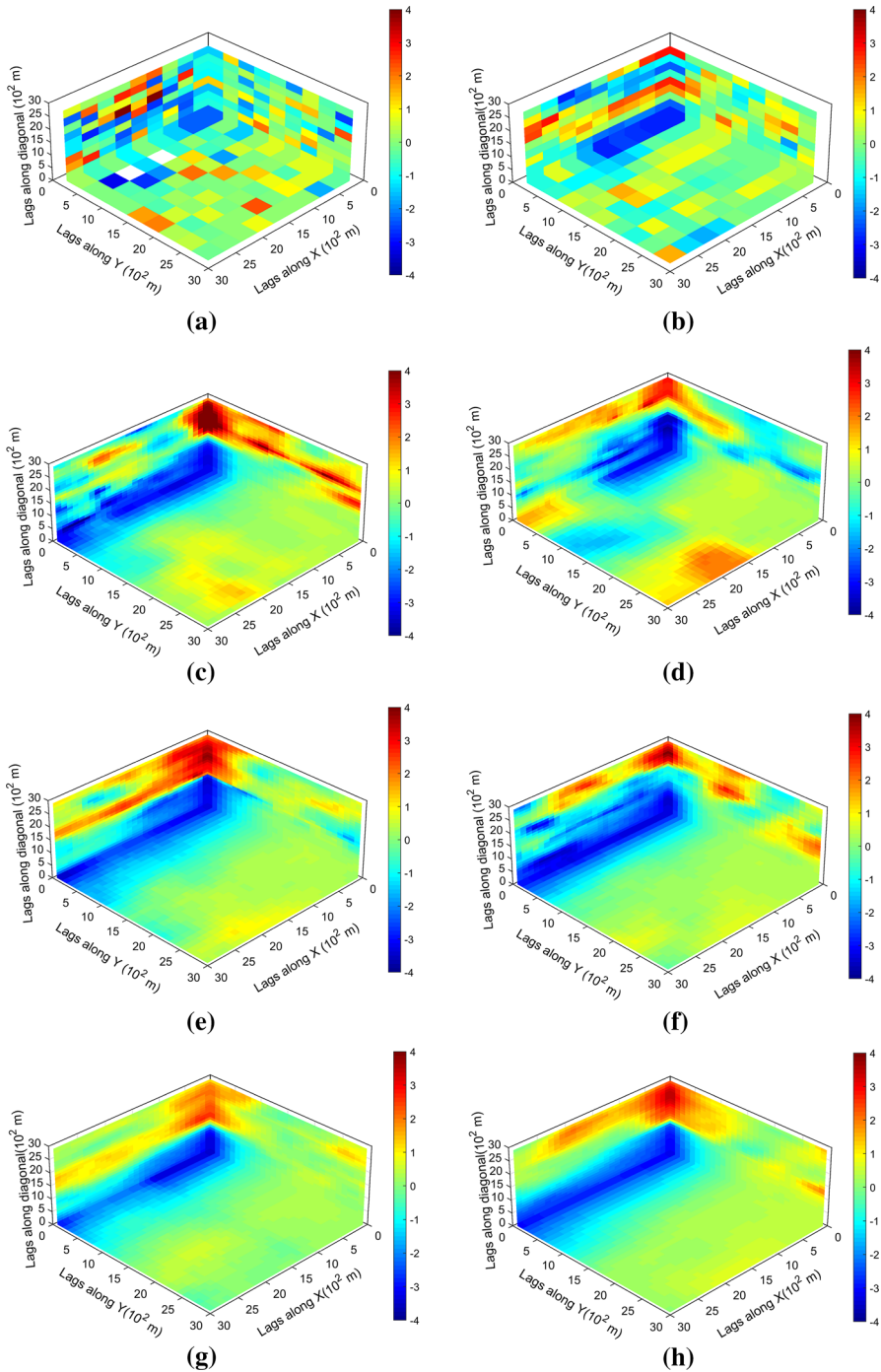


Fig. 11 Fourth-order cumulant maps of **a** DS-1, **b** DS-2, **c** exhaustive image, **d** TI-1, **e** realization in Fig. 7a with DS-1 as the sample data, **f** realization in Fig. 7b with DS-2 as the sample data, **g** 10 realizations in average with DS-1 as the sample data, and **h** 10 realizations in average with DS-2 as the sample data

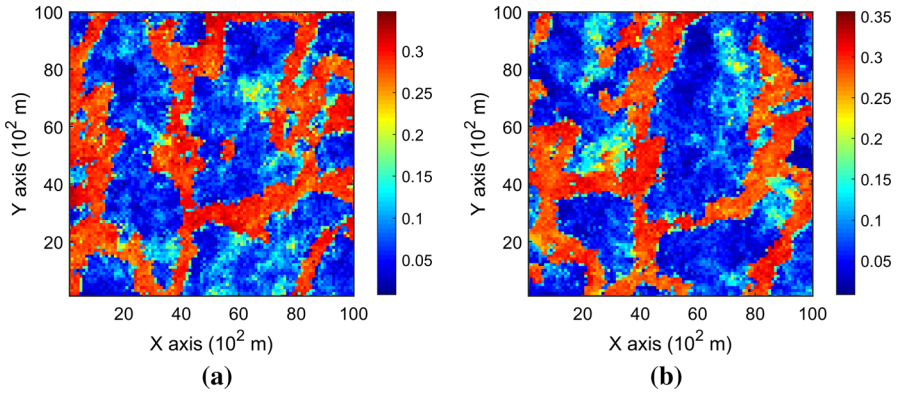


Fig. 12 One realization from KERNELSIM using TI-2. **a** DS-1 as the sample data, **b** DS-2 as the sample data

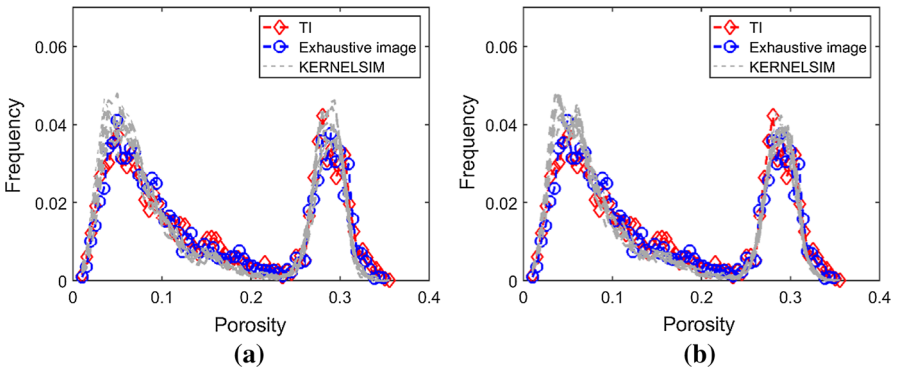


Fig. 13 Histograms of 10 realizations of KERNELSIM using TI-2. **a** DS-1 as the sample data, **b** DS-2 as the sample data

and the exhaustive image. Thus, this specific example aims to test the sensitivity of the KERNELSIM method to the more apparent statistical conflicts between the TI and the sample data. Figure 12 shows one realization of KERNELSIM using TI-2 as the training image, along with DS-1 and DS-2 as the sample data, respectively. Interestingly, even with relatively sparse sample data DS-1, the realization of KERNELSIM still reflects the vertical channels well. The same phenomena can also be observed in the realization using the denser sample data DS-2. Comparisons of the histograms and the variograms are shown in Figs. 13 and 14, respectively. Further, a comparison of high-order spatial statistics is shown in Figs. 15 and 16 in a similar way as in Example 1. While the third-order and the fourth-order cumulant maps of the TI and the exhaustive image are very different, the cumulant maps of the realizations still maintain the main spatial features of the one from the exhaustive image. This specific example shows that the KERNELSIM method is capable of generalizing the learning model to adapt to situations in the presence of statistical conflicts between the sample data and the TI. Of note, even with relatively sparse sample data, the proposed method can

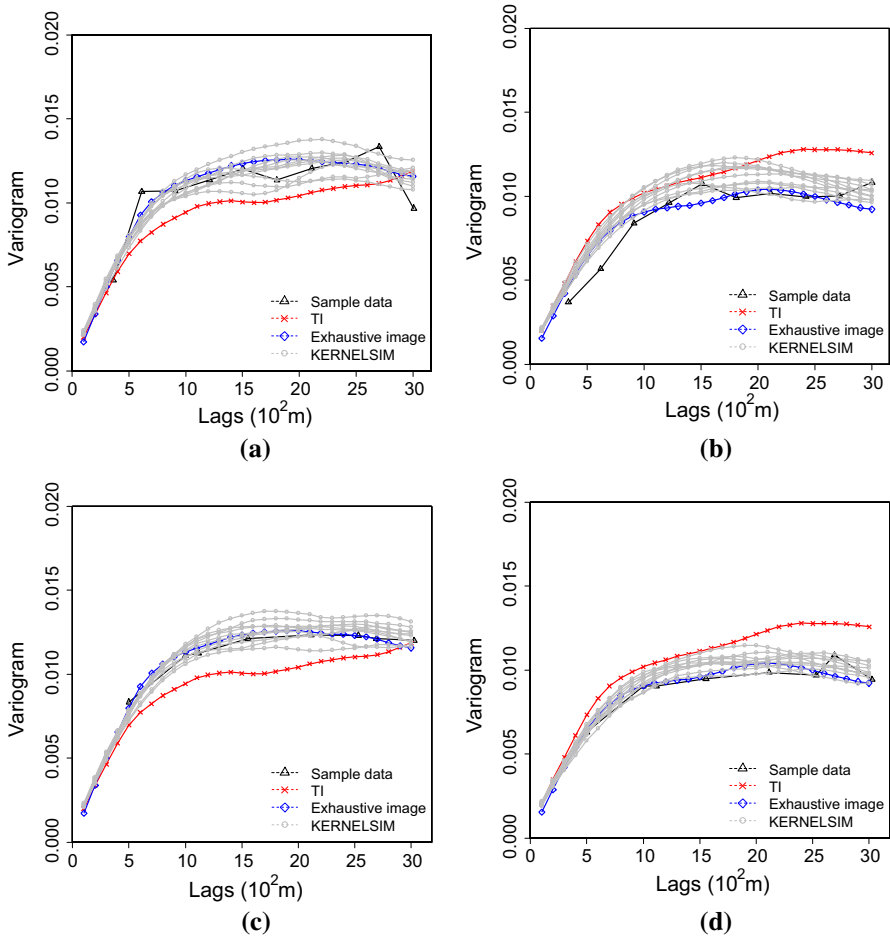


Fig. 14 Variograms of 10 realizations of KERNELSIM using TI-2. **a, b** Along the X and Y axes with DS-1 as the sample data; **c, d**, along the X and Y axes with DS-2 as the sample data

generate realizations with a reasonable reproduction of spatial statistics of the sample data from the lower to the higher orders.

4.1.3 Conditional Probability on Different Spatial Patterns

Three configurations of the conditioning data are intentionally picked at different locations to represent the typical spatial patterns that are possibly encountered in the data event. The KERNELSIM method is applied to generate the conditional probability distributions on these different spatial patterns to compare the behaviors of the CPDF at different locations (Fig. 17). Since the attribute values are transformed to the domain $[-1, 1]$ of Legendre polynomials, both the conditioning data and the CPDFs are also in this domain. Figure 17a shows the pattern of transition between lower values and higher

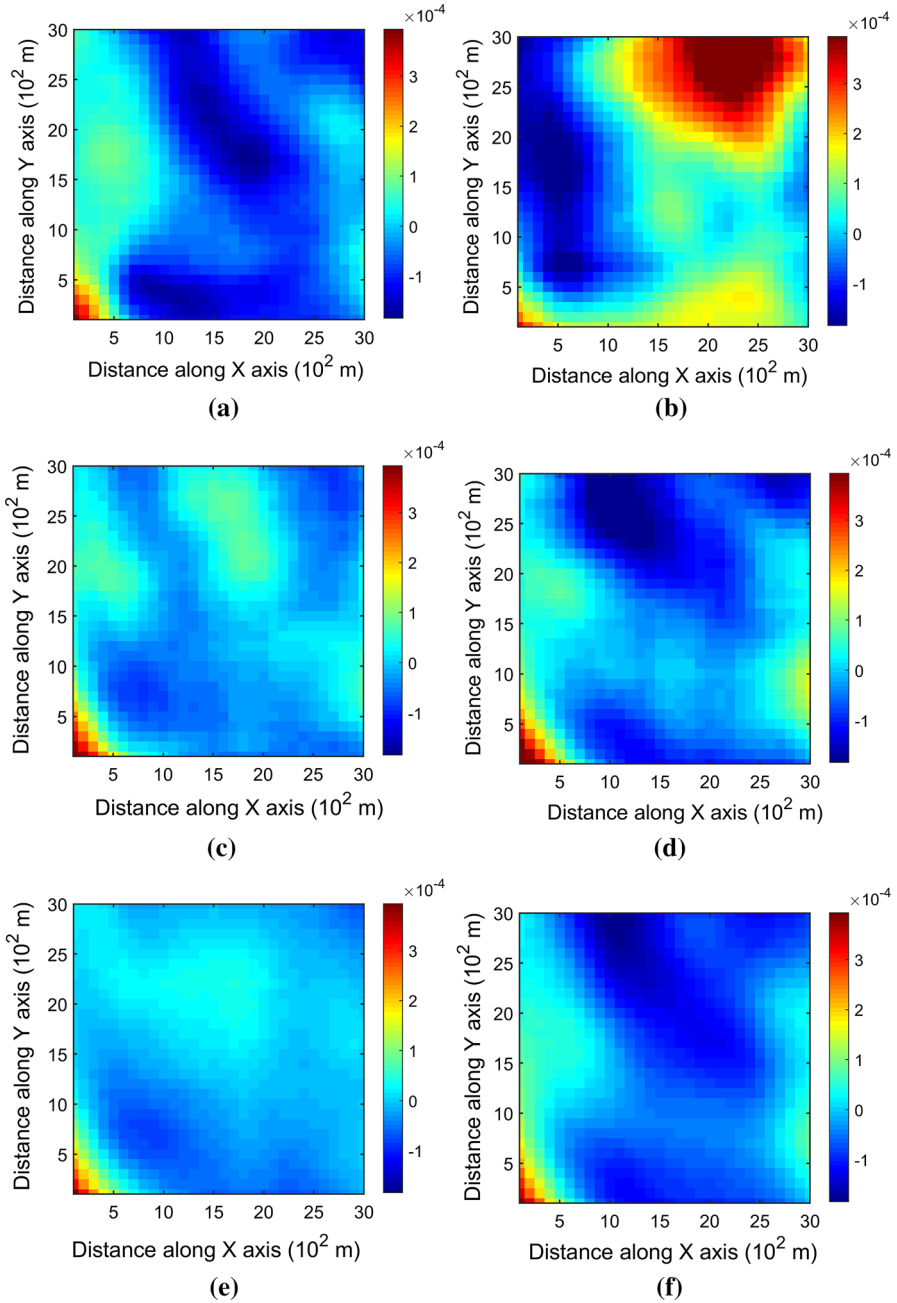


Fig. 15 Third-order cumulant maps of **a** exhaustive image, **b** TI-2, **c** realization in Fig. 12a with DS-1 as the sample data, **d** realization in Fig. 12b with DS-2 as the sample data, **e** 10 realizations in average with DS-1 as the sample data, and **f** 10 realizations in average with DS-2 as the sample data

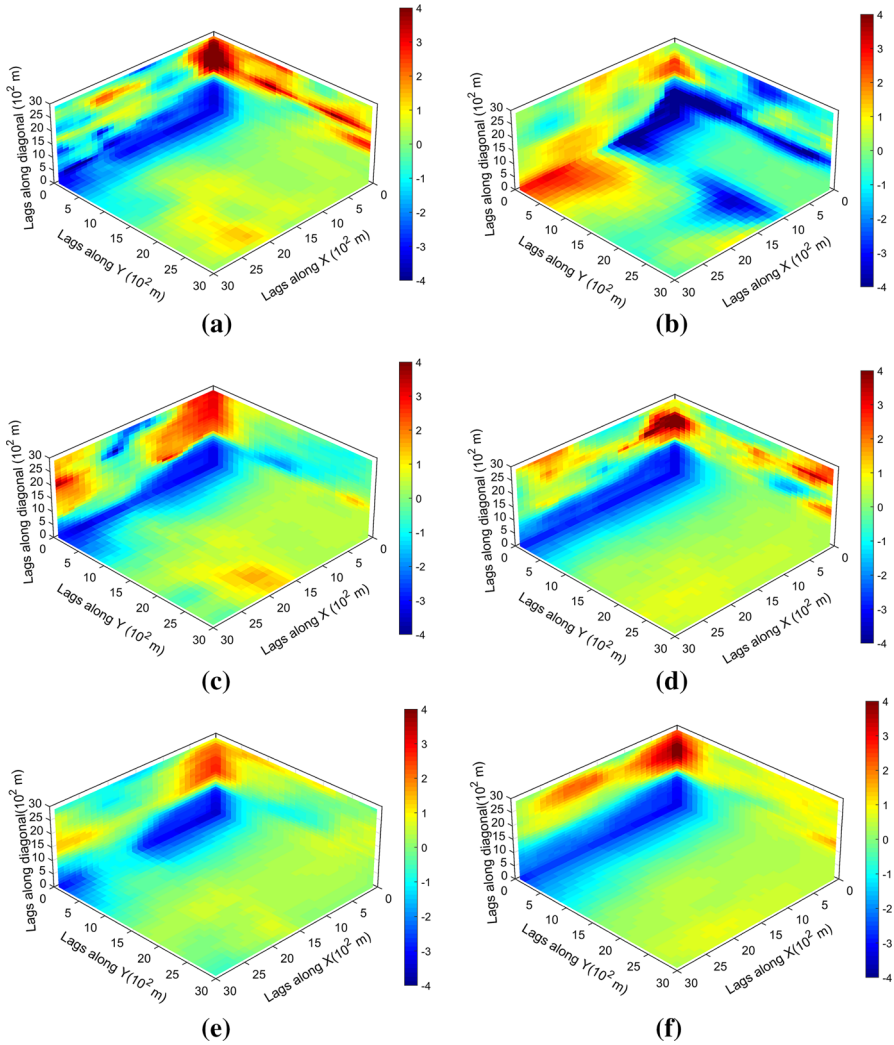


Fig. 16 Fourth-order cumulant maps of **a** exhaustive image, **b** TI-2, **c** realization in Fig. 12a with DS-1 as the sample data, **d** realization in Fig. 12b with DS-2 as the sample data, **e** 10 realizations in average with DS-1 as the sample data, and **f** 10 realizations in average with DS-2 as the sample data

values, which usually happens near the boundary of the channels in the exhaustive image, while Fig. 17b shows its corresponding CPDF at the center node. In this case, the CPDF has two different modes at the values of -0.41 and 0.74 , which interestingly implies that the possible prediction could either be a lower value or a higher value, while the higher value has a higher likelihood. It turns out that the true value at this location after transformation is 0.745 . However, it should be noted here that this double-modal behavior is reasonable near the boundary of transitioning between lower and higher values. This kind of probability distribution cannot be characterized by the second-

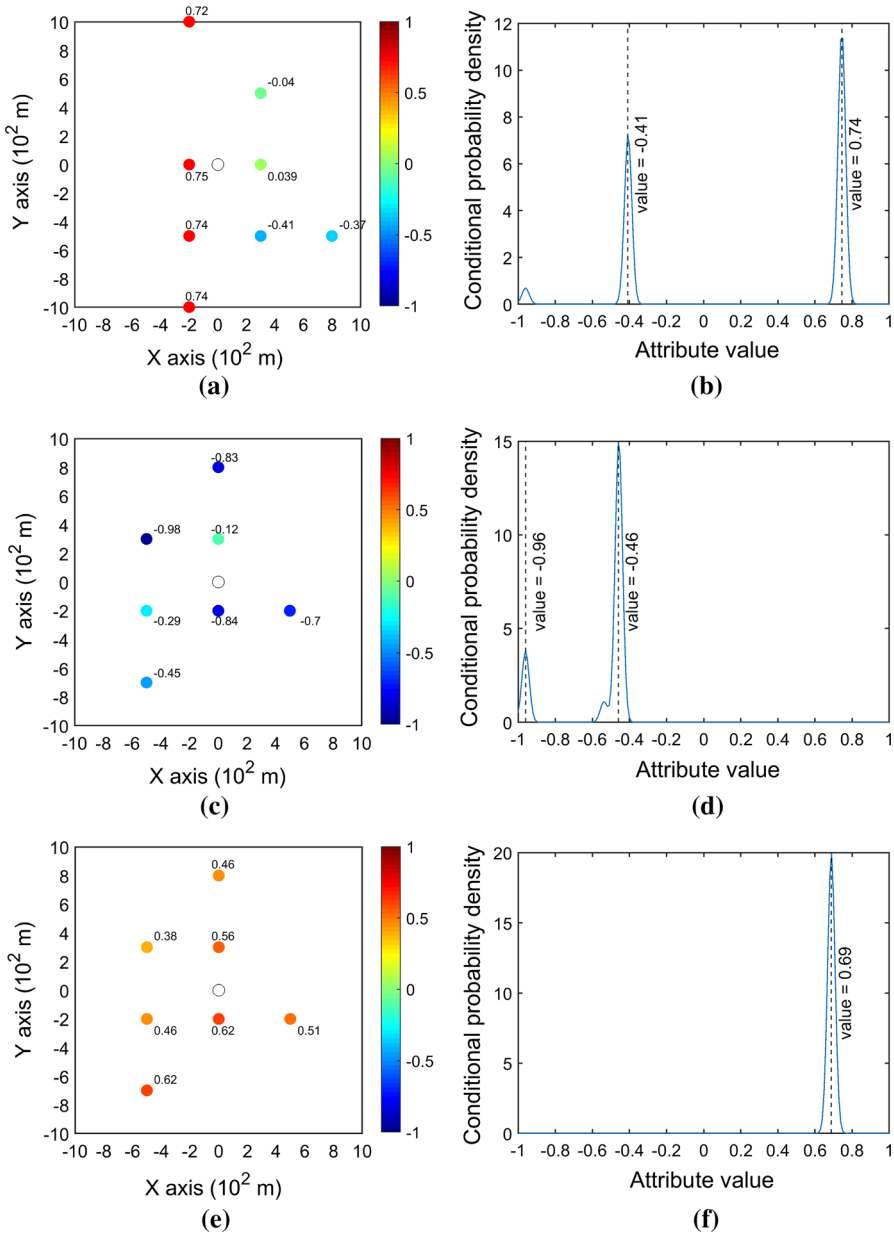


Fig. 17 Behaviors of conditional probability distributions corresponding to conditioning data with different spatial patterns. The central circle represents the center node to be simulated, and the colored nodes are the conditioning data in the neighborhood

order geostatistical simulation methods based on Gaussian assumption. Figure 17c, d shows the simulation behavior at a location where the center node is surrounded by nodes with relatively lower values. Again, the CPDF also shows a bimodal shape

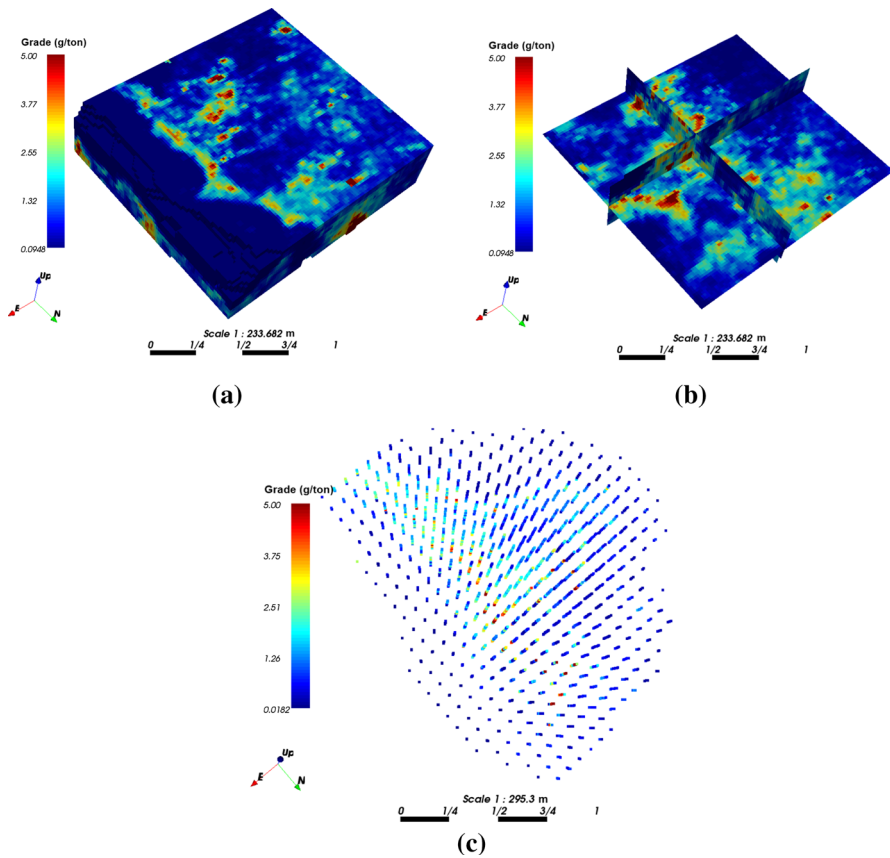


Fig. 18 **a** TI, **b** a cross section of the TI, and **c** the sample data of the Au grades

due to the large variation in the spatial patterns. Figure 17e, f shows the behavior of simulation at a location where the center node is surrounded by nodes with relatively higher values. The CPDF exhibits a unimodal distribution as the variation in the spatial pattern is small. Although the behaviors of CPDF could be case-dependent due to different spatial distributions of attributes of interest, these experiments show that the CPDFs generated by KERNELSIM are driven by the training data instead of a fixed covariance function, and thus can reflect the characteristics of different spatial patterns. In fact, several past studies have also shown the advantage of high-order simulation methods in reproducing the complex spatial patterns over the traditional second-order simulation methods, such as sequential Gaussian simulation (de Carvalho et al. 2019; Minniakhmetov et al. 2018; Mustapha and Dimitrakopoulos 2010b).

4.2 Case Study at a Gold Deposit

The case study at a gold deposit is presented here to demonstrate the practical aspects and the performance of KERNELSIM in its application to a real-life example. The

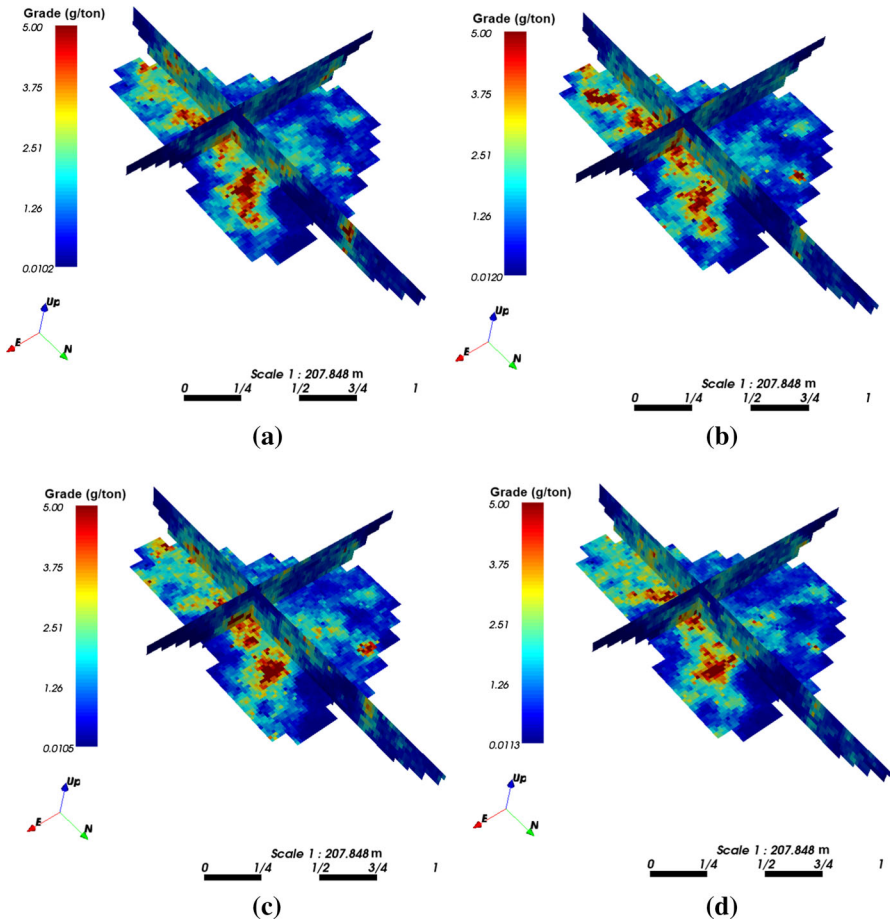


Fig. 19 Cross sections of four different realizations of KERNELSIM of the Au grades

sample data are from 407 exploration drill holes and are composited to 10 m in length. The TI comes from the blast hole data located at a mined-out area of the ore body. Figure 18 shows the TI, a cross section of the TI and the sample data in a three-dimensional view. The TI is generated from the blast hole data assuming that the geological settings of the studied area are similar to the mined-out area, where conflicts would be mitigated by the statistical learning process dominated by the sample data. Figure 19 shows cross sections of four different realizations of KERNELSIM for the gold deposit in a three-dimensional view. The histogram of the gold grades resembles the histogram of the sample data, as can be seen from Fig. 20. The variograms of the sample data and the TI are plotted for comparison with the variograms of 10 realizations of KERNELSIM from the gold deposit in Fig. 21. Figure 22 shows the third-order cumulant maps of the samples, the TI and the realization of KERNELSIM, respectively, along with the L-shape spatial template in the X–Y plane. Furthermore, the fourth-order cumulant maps of the samples, the TI and the realization of KERNEL-

Fig. 20 Histograms of 10 realizations of KERNELSIM for the Au grades of the gold deposit in comparison to the TI and the samples

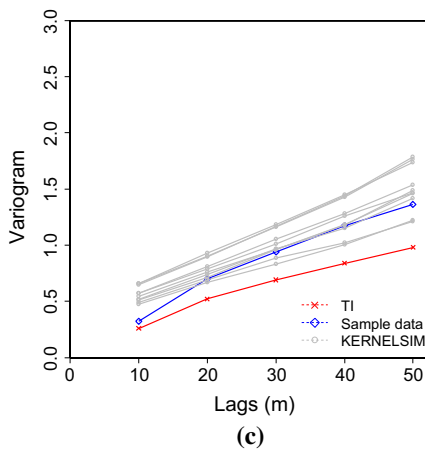
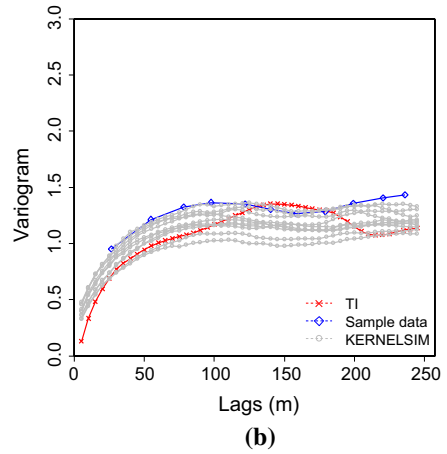
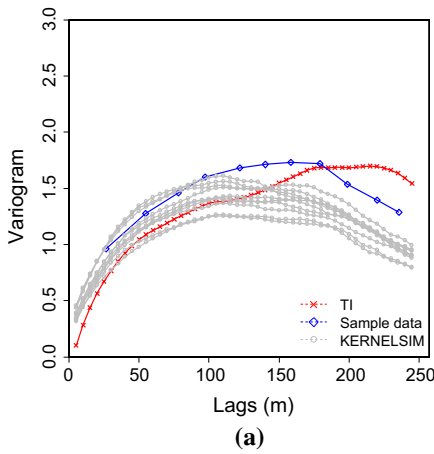
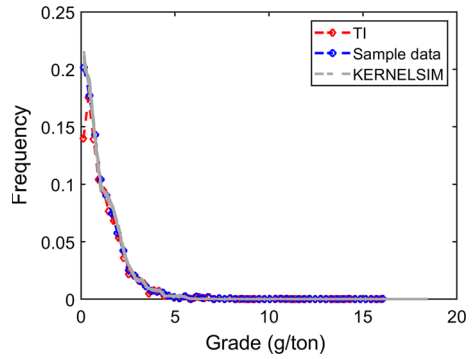


Fig. 21 Variograms of 10 realizations of KERNELSIM for Au grades at the gold deposit along **a** E–W, **b** N–S, and **c** down drill holes, in comparison to the sample data and the TI

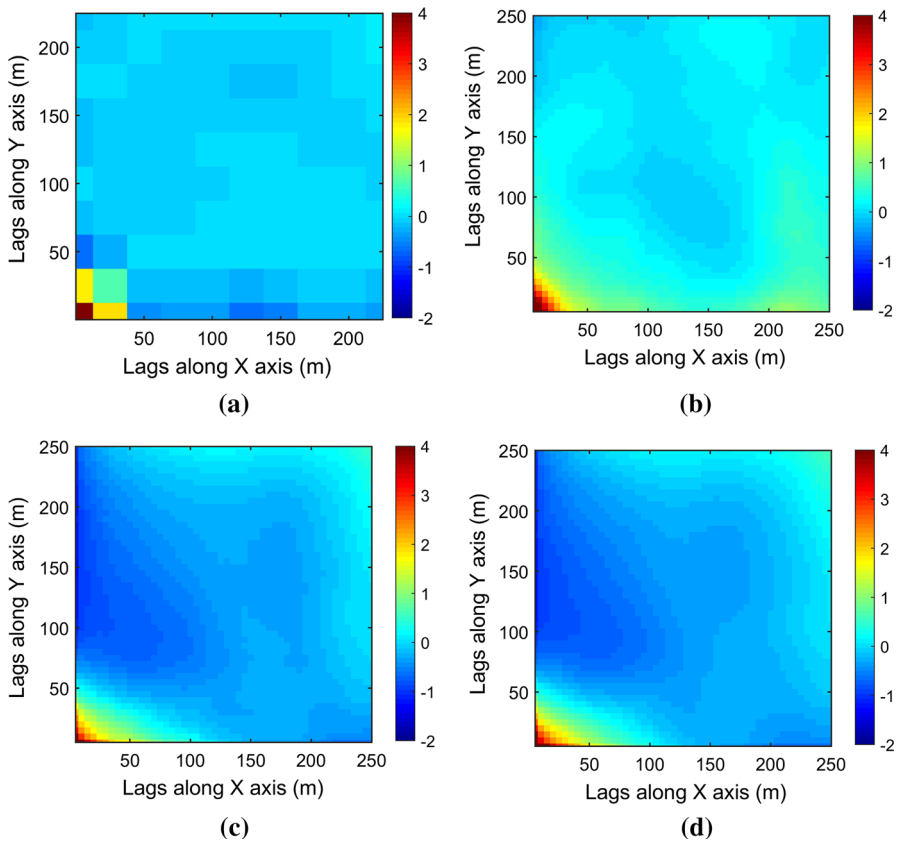


Fig. 22 Third-order cumulant maps of **a** the sample data, **b** the TI, **c** the realization of KERNELSIM and **d** the 10 realizations of KERNELSIM in average

SIM are respectively displayed in Fig. 23. The results of the comparison in Figs. 22 and 23 show that the KERNELSIM reproduces the high-order spatial statistics of the sample data in addition to the lower-order statistics, even though the spatial patterns of the third-order and fourth-order cumulant maps of the TI are different to those of the sample data.

5 Conclusions

The paper presents a new high-order stochastic simulation framework based on statistical learning. Within this statistical learning workflow, the density estimation in the sequential simulation is kernelized, which renders it equivalent to solving a quadratic programming problem. The kernelization is approached by embedding the original data space into a kernel Hilbert space. A spatial Legendre moment reproducing kernel is proposed to construct an RKHS that can incorporate the high-order spatial statistics of the original data. In addition, a kernel decomposition technique is proposed

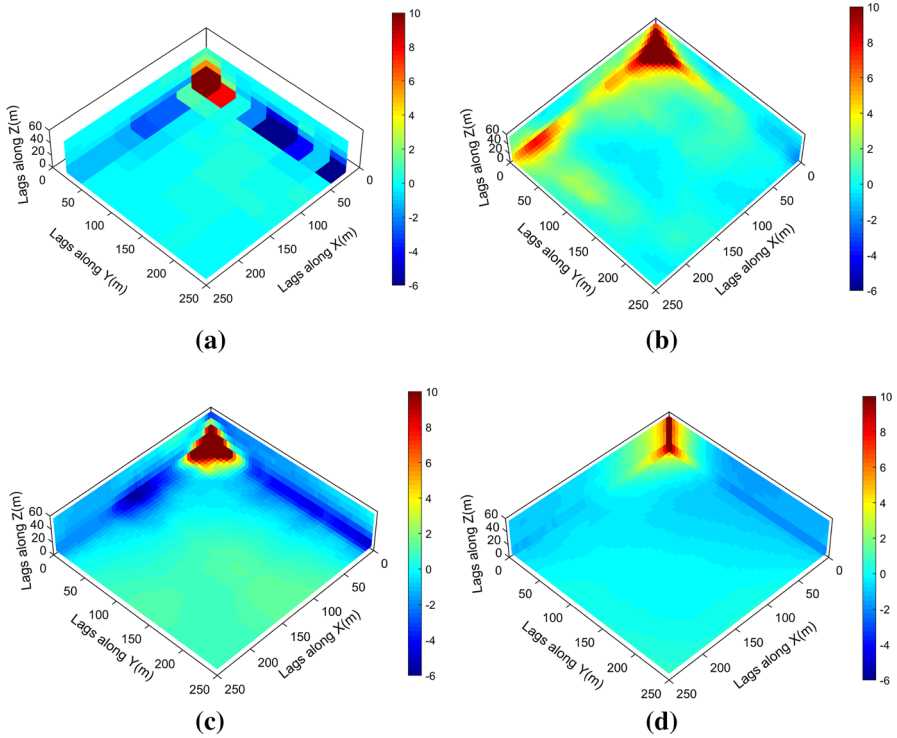


Fig. 23 Fourth-order cumulant maps of **a** the sample data, **b** the TI, **c** the realization of KERNELSIM and **d** the 10 realizations of KERNELSIM in average

to project the kernelization into a one-dimensional kernel Hilbert space to approach the sequential simulation procedure and to reduce computational complexity. The proposed statistical learning framework is general and can cope with the possible statistical conflicts between the sample data and the TI. The implementation of the method presented, termed KERNELSIM, is tested in different case studies. The examples, which use a fully known reservoir, show that KERNELSIM can reproduce the main spatial patterns of the sample data. Notably, the generalization capacity of the proposed method mitigates the statistical conflicts between the sample data and the TI and retains high-order statistical features from the sample data. The two examples in the first case study also provide some insights on how the number of the sample data and the relation of the sample data to the TI affect the simulation results. It should be noted that the simulation results only use the replicates from the TI to infer a conditional probability distribution. Hence, the proposed statistical framework provides an approach to condition the local probabilistic models learning from the TI to the existing configuration of the sample data based on the generalization capacity of the learning framework. However, the assumption made is that the TI shares some similarities in the local spatial structures with the sample data, even though their global structures could be different. The impact of the TI can also be reduced by only using replicates from the sample data, if the number of the replicates reaches a cer-

tain threshold of statistical significance, similarly to the approach adopted in previous publications (Mustapha and Dimitrakopoulos 2010b; Yao et al. 2018). A case study at a gold deposit demonstrates the performance of KERNELSIM in a three-dimensional example. The results show that the KERNELSIM method reproduces the high-order spatial statistics of the drill hole samples well. Thus, the method provides an effective approach to simulate the ore body using the drill hole samples with the TI originating from a suitable mined-out part of the same deposit.

It is important to note that in the context of the general learning process, the compromise between minimization of the training error (prediction error on the training data) and the test error (prediction error with a new input other than the one from the training data) leads to the well-known bias-and-variance tradeoff (Hastie et al. 2009). More specifically, the complexity of the learning machine can be balanced to avoid overfitting and to increase the possibility of generalizing the learning model, which stabilizes the prediction output of the learning model. On the one hand, the learning process targeting to match the high-order spatial statistics aims to minimize the bias (i.e., the deviation of the high-order statistics of the estimated model from that of the available data). On the other hand, by applying simpler and more relevant models to the solution space of the target distributions, such as using the convex space of prototype distributions in the present method, the solutions tend to be less sensitive to the noisy fluctuation and have a greater capacity for generalization. In fact, this balancing of the bias-and-variance tradeoff demonstrates the flexibility of the new statistical learning framework of high-order simulation, which other simulation methods lack.

Acknowledgments The authors would like to thank the two reviewers for their critical comments and their help to improve the quality of the manuscript. This work was funded by Fonds de recherche du Québec-Nature et technologies, FQRNT Grant “Développement d’une stratégie globale d’optimisation de sites miniers avec incertitude: Amélioration de la viabilité économique et de la gestion environnementale des résidus miniers d’une mine de fer dans le nord,” with New Millennium Iron Corporation being the industry collaborator, the Natural Sciences and Engineering Research Council (NSERC) of Canada, CRD Grant CRDPJ 500414-16, the COSMO Mining Industry Consortium (AngloGold Ashanti, Barrick Gold, BHP, De Beers, IAMGOLD, Kinross, Newmont Mining and Vale), and NSERC Discovery Grant 239019.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Altun Y, Smola A (2006) Unifying divergence minimization and statistical inference via convex duality. In: Proceedings of the 19th annual conference on learning theory, Pittsburgh, PA. Springer, Berlin, pp 139–153. https://doi.org/10.1007/11776420_13
- Berlinet A, Thomas-Agnan C (2004) Reproducing kernel Hilbert spaces in probability and statistics. Kluwer, Boston
- de Carvalho JP, Dimitrakopoulos R, Minniakhmetov I (2019) High-order block support spatial simulation method and its application at a gold deposit. *Math Geosci* 51:793–810. <https://doi.org/10.1007/s11004-019-09784-x>
- Dimitrakopoulos R, Mustapha H, Gloaguen E (2010) High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena. *Math Geosci* 42:65–99. <https://doi.org/10.1007/s11004-009-9258-9>

- Goldfarb D, Idnani A (1983) A numerically stable dual method for solving strictly convex quadratic programs. *Math Program* 27:1–33. <https://doi.org/10.1007/bf02591962>
- Goodfellow R, Albor Consuegra F, Dimitrakopoulos R, Lloyd T (2012) Quantifying multi-element and volumetric uncertainty, Coleman McCreeedy deposit, Ontario, Canada. *Comput Geosci* 42:71–78. <https://doi.org/10.1016/j.cageo.2012.02.018>
- Guardiano F, Srivastava RM (1993) Multivariate geostatistics: beyond bivariate moments. In: Soares A (ed) *Geostatistics Tróia '92. Quantitative Geology and Geostatistics*, vol 5. Kluwer, Dordrecht, pp 133–144. https://doi.org/10.1007/978-94-011-1739-5_12
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York
- Journel A (1994) Modeling uncertainty: Some conceptual thoughts. In: Dimitrakopoulos R (ed) *Geostatistics for the next century. Quantitative Geology and Geostatistics*, vol 6. Springer, Dordrecht, pp 30–43. https://doi.org/10.1007/978-94-011-0824-9_5
- Journel AG, Zhang T (2006) The necessity of a multiple-point prior model. *Math Geol* 38:591–610. <https://doi.org/10.1007/s11004-006-9031-2>
- Lebedev NN, Silverman RA (1965) *Special functions and their applications*. Prentice-Hall, Englewood Cliffs
- Mao S, Journel A (1999) Generation of a reference petrophysical/seismic data set: the Stanford V reservoir. Stanford
- Mariethoz G, Caers J (2014) *Multiple-point geostatistics: stochastic modeling with training images*. Wiley, Hoboken
- Minniakhmetov I, Dimitrakopoulos R (2017) Joint high-order simulation of spatially correlated variables using high-order spatial statistics. *Math Geosci* 49:39–66. <https://doi.org/10.1007/s11004-016-9662-x>
- Minniakhmetov I, Dimitrakopoulos R, Godoy M (2018) High-order spatial simulation using Legendre-like orthogonal splines. *Math Geosci* 50:753–780. <https://doi.org/10.1007/s11004-018-9741-2>
- Muandet K, Fukumizu K, Sriperumbudur B, Schölkopf B (2016) Kernel mean embedding of distributions: a review and beyonds. arXiv preprint [arXiv:1605.09522](https://arxiv.org/abs/1605.09522)
- Mustapha H, Dimitrakopoulos R (2010a) Generalized Laguerre expansions of multivariate probability densities with moments. *Comput Math Appl* 60:2178–2189. <https://doi.org/10.1016/j.camwa.2010.08.008>
- Mustapha H, Dimitrakopoulos R (2010b) High-order stochastic simulation of complex spatially distributed natural phenomena. *Math Geosci* 42:457–485. <https://doi.org/10.1007/s11004-010-9291-8>
- Mustapha H, Dimitrakopoulos R (2011) HOSIM: a high-order stochastic simulation algorithm for generating three-dimensional complex geological patterns. *Comput Geosci* 37:1242–1253. <https://doi.org/10.1016/j.cageo.2010.09.007>
- Osterholt V, Dimitrakopoulos R (2007) Simulation of wireframes and geometric features with multiple-point techniques: application at Yandi iron ore deposit, Australia. In: *Orebody modelling and strategic mine planning*, vol 14, 2 edn. AusIMM Spectrum Series, pp 51–60
- Remy N, Boucher A, Wu J (2009) *Applied geostatistics with SGeMS: a user's guide*. Cambridge University Press, Cambridge
- Schölkopf B, Smola A (2001) *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge
- Scott DW (2015) *Multivariate density estimation: theory, practice, and visualization*. Wiley, New York
- Smola A, Gretton A, Song L, Schölkopf B (2007) A Hilbert Space Embedding for Distributions. In: Hutter M, Servadio RA, Takimoto E (eds) *Algorithmic Learning Theory*. Springer, Berlin, Heidelberg, pp 13–31
- Song L, Fukumizu K, Gretton A (2013) Kernel embeddings of conditional distributions: a unified kernel framework for nonparametric inference in graphical models. *IEEE Signal Process Mag* 30:98–111. <https://doi.org/10.1109/MSP.2013.2252713>
- Song L, Zhang X, Smola A, Gretton A, Schölkopf B (2008) Tailoring density estimation via reproducing kernel moment matching. In: *Proceedings of the 25th international conference on machine learning*. ACM, New York, pp 992–999
- Stein EM, Shakarchi R (2005) *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton University Press, Princeton
- Steinwart I, Christmann A (2008) *Support vector machines*. Springer, New York

- Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. *Math Geol* 34:1–21. <https://doi.org/10.1023/A:1014009426274>
- Vapnik VN (1995) *The nature of statistical learning theory*. Springer, New York
- Vapnik VN (1998) *Statistical learning theory*. Wiley, New York
- Vapnik VN, Mukherjee S (1999) Support vector method for multivariate density estimation. In: *Proceedings of the 12th international conference on neural information processing systems*, Denver, CO. MIT Press, Cambridge, pp 659–665
- Vavasis SA (2001) Complexity theory: quadratic programming. In: Floudas CA, Pardalos PM (eds) *Encyclopedia of optimization*. Springer, Boston, pp 304–307. https://doi.org/10.1007/0-306-48332-7_65.
- Yao L, Dimitrakopoulos R, Gamache M (2018) A new computational model of high-order stochastic simulation based on spatial Legendre moments. *Math Geosci* 50:929–960. <https://doi.org/10.1007/s11004-018-9744-z>