Math Geosci (2016) 48:559–579
DOI 10.1007/s11004-015-9606-x

SPECIAL ISSUE

Inverse Modeling of Moving Average Isotropic Kernels for Non-parametric Three-Dimensional Gaussian Simulation

Oscar Peredo[1] · **Julián M. Ortiz**[1,2] · **Oy Leuangthong**[3]

Received: 6 October 2014 / Accepted: 3 July 2015 / Published online: 28 July 2015

Abstract Moving average simulation can be summarized as a convolution between a spatial kernel and a white noise random field. The kernel can be calculated once the variogram model is known. An inverse approach to moving average simulation is proposed, where the kernel is determined based on the experimental variogram map in a non-parametric way, thus no explicit variogram modeling is required. The omission of structural modeling in the simulation work-flow may be particularly attractive if spatial inference is challenging and/or practitioners lack confidence in this task. A non-linear inverse problem is formulated in order to solve the problem of discrete kernel weight estimation. The objective function is the squared euclidean distance between experimental variogram values and the convolution of a stationary random field with Dirac covariance and the simulated kernel. The isotropic property of the kernel weights is imposed as a linear constraint in the problem, together with lower and upper bounds for the weight values. Implementation details and examples are presented to demonstrate the performance and potential extensions of this method.

Keywords Moving average · Convolution · Gaussian simulation · Variogram · Inverse problems

1 Introduction

Geostatistical simulation is useful to quantify uncertainty of a variable in space and assess its performance to various processes. These processes may be related to flow,

✉ Oscar Peredo
operedo@gmail.com

1 ALGES Laboratory, Advanced Mining Technology Center, University of Chile, Santiago, Chile

2 Department of Mining Engineering, University of Chile, Santiago, Chile

3 SRK Consulting (Canada) Inc., Toronto, Canada

transport, thresholding, etc., and they require computing multiple numerical realizations that discretize the geographical space. The generation of multiple realizations can be cumbersome, since the spatial structure must be inferred and imposed a priori in the numerical models. Many methods exist for generating the simulations once a structural model (a variogram) has been determined (Chilès and Delfiner 2012). However, a constant challenge is to simplify this inference process as much as possible and minimize the degree of uncertainty involved.

In the context of parametric variogram modeling, novel attempts to semi-automatize or fully automatize this process can be found in Cressie (1985), Pardo-Igúzquiza (1999), Emery (2010), and Desassis and Renard (2013). In these works the variables to be fitted are parameters from structures defined a priori (spherical, gaussian, exponential and others) minimizing some form of least-squares metric. Even though this approach is fast and accurate, the predictions might be seriously biased if the true variogram is not closely approximated by any combination of members of the parametric families.

As an alternative, the non-parametric approach does not require any knowledge of the prior structures, typically paying a higher price in computational cost and convergence. Wackernagel (1994) uses smoothing in the experimental covariance values and eigenvalue decomposition/correction of the smoothed covariance matrix. Works of Shapiro and Botha (1991), Rehman (1995), and Yao and Journel (1998) are based in Bochner's theorem, described by Bochner (1949), to map the positive-definite constraints to a spectrum representation using the Fourier transform. In the spectral domain, a long series of density spectrum values must be fitted to experimental transformed data. A different approach based in moving averages modeling has been used to provide an alternative to the standard classes of parametric variogram/covariance functions. Calder and Cressie (2007) give applied examples of this approach not only in non-parametric variogram modeling, but also in dimensional reduction (Higdon 1998, 2002), non-stationary and anisotropic modeling (Higdon 1998; Higdon et al. 1999; Fuentes 2002a, b), and multivariate spatial modeling (Barry and Ver Hoef 1996; Calder 2008; Majumdar and Gelfand 2007). This approach is particularly attractive because it guarantees a valid non-negative definite spatial covariance function (Matérn 1986; Thièbaux and Pedder 1987).

The (forward) moving averages, discussed by Oliver (1995) and previously applied in one dimension by Journel and Huijbregts (1978), can be summarized as a convolution between a spatial kernel of weights and a white noise random field. The kernel can be calculated once the variogram model is known. As explained in the referenced articles, the stochastic integral

$$\mathbf{Y}(\mathbf{x}) = \int_U f(\mathbf{x}, \mathbf{x}')Z(\mathbf{x}')\mathrm{d}\mathbf{x}' \tag{1}$$

allows us to obtain correlated random fields $\mathbf{Y}$ with covariance matrix $\mathbf{C}$ and mean 0, convolving a kernel function f with a stationary random field $\mathbf{Z}$ with a Dirac covariance measure. The covariance $\mathbf{C}(\mathbf{x}, \mathbf{x}')$ is related to the kernel function f through the following equation

$$\mathbf{C}(\mathbf{x}, \mathbf{x}') = \int_U f(\mathbf{x}, \mathbf{x}'') f(\mathbf{x}', \mathbf{x}'') \mathrm{d}\mathbf{x}'' \tag{2}$$

Given a known covariance model $\mathbf{C}(r)$ (e.g., exponential, gaussian, spherical, or others) with $r = |\mathbf{x} - \mathbf{x}'|$, and denoting $\mathcal{F}$ the Fourier transform, Eq. (2) can be solved by considering the inverse Fourier transform of $\sqrt{\mathcal{F}(\mathbf{C}(r))}$ as

$$f(r) = \mathcal{F}^{-1}\left(\sqrt{\mathcal{F}(\mathbf{C}(r))}\right)$$

Using the moving average non-parametric approach, Barry and Ver Hoef (1996) proposed a black-box method to kriging where the non-parametric kernels are piecewise constant functions of the form

$$f(r) = \sum_{j=1}^{m} a_j \mathcal{I}\left\{\frac{(j-1)R}{m} < r \leq \frac{jR}{m}\right\}, \tag{3}$$

where $\mathcal{I}$ denotes the indicator function, m the number of radial bands, and R is the kernel radius of influence. The values of the constants a_j are estimated solving a derivative-free optimization problem which minimizes the distance to experimental variogram values. Posteriorly, Ver Hoef et al. (2004) improve the computational efficiency of the kernel fitting using a Fast Fourier Transform (FFT) to compute the convolutions between the kernel and the random field $\mathbf{Z}$. Gribov et al. (2006) reported advanced techniques and novel insights implementing a similar approach into geostatistical software.

The approach taken in this work follows ideas from Barry and Ver Hoef (1996), where the kernel is fully discretized based in the simulation grid resolution. The kernel values at each discrete location are the variables to be estimated. With this choice, three-dimensional gaussian simulations can be obtained directly, just convolving the fitted kernel with a random field of Dirac covariance measure, using the same simulation grid resolution as before. The problem of estimation of the kernel values at discrete locations is formulated in terms of a non-linear inverse problem with linear constraints. In this problem the cost function is the weighted least-squares (WLS) function between two vectors of variographic data, one calculated using experimental values, and the other simulated using the unknown weights of the kernel function. The constraints are associated to statistical and geometrical properties that must be fulfilled by the values of f at user-defined neighbor locations. The combined output of this problem is a distribution of fitted kernels and the associated convoluted simulations, which have a similar semi-variogram curve as the experimental data. The distribution is generated using a Monte Carlo method by sampling maximum likelihood a posteriori kernels using an initial kernel guess as a priori model.

As a starting point, only isotropic kernels are modeled. Its symmetrical properties allow us to test new concepts and modify the models in a flexible and easy way, focusing the efforts in studying and analyzing the relation between the variogram and the kernel weights. However, an anisotropic kernel can also be modeled using the proposed method by just relaxing some of the linear constraints imposed in the

optimization problem and extending the definition of the kernel in Eq. (3), as shown by Barry and Ver Hoef (1996), to two or three radial dimensions.

In the next section, theoretical and numerical aspects of the inverse approach are explained. A proposed implementation and practical aspects of its usage are presented in Sect. 3. In Sect. 4, synthetic and real scenarios are addressed, quantifying the effect of kernel smoothness and sparsity of the available dataset. Finally, conclusions and future work are included in Sect. 5.

2 Inverse Moving Average Isotropic Kernels

Given a lag vector $\mathbf{h}$, the semi-variogram $\gamma(\mathbf{h})$ can be estimated from available N_d data samples, denoted $\{y(\mathbf{x}_i)\}_{i=1}^{N_d}$, located in a regular three-dimensional lattice using the formula

$$\gamma(\mathbf{h}) = \frac{1}{2N_h} \sum_{\{(p,q):\mathbf{x}_q=\mathbf{x}_p+\mathbf{h}\}} [y(\mathbf{x}_p) - y(\mathbf{x}_q)]^2 \tag{4}$$

Assuming that the $\mathbf{Y}$ random field was generated using (1), for some kernel f and some stationary random field $\mathbf{Z}$ with Dirac covariance measure, it holds

$$\begin{aligned}\gamma(\mathbf{h}) &= \frac{1}{2N_h} \sum_{\{(p,q):\mathbf{x}_q=\mathbf{x}_p+\mathbf{h}\}} [y(\mathbf{x}_p) - y(\mathbf{x}_q)]^2 \\ &\approx \frac{1}{2N_h} \sum_{\{(p,q):\mathbf{x}_q=\mathbf{x}_p+\mathbf{h}\}} \left[\sum_{i=-R}^{R} \sum_{j=-R}^{R} \sum_{k=-R}^{R} w_{i,j,k} (z_{p(i,j,k)} - z_{q(i,j,k)}) \right]^2 \end{aligned} \tag{5}$$

with R the radius of the kernel's neighborhood (in this particular discretization, a cube of side $2R+1$ is considered as neighborhood) and the variables $z_{p(i,j,k)}$ and $w_{i,j,k}$ defined as

$$z_{p(i,j,k)} =: z(\mathbf{x}_p + ib\hat{\mathbf{i}} + jb\hat{\mathbf{j}} + kb\hat{\mathbf{k}}) \tag{6}$$

$$w_{i,j,k} =: f(ib\hat{\mathbf{i}} + jb\hat{\mathbf{j}} + kb\hat{\mathbf{k}}) \tag{7}$$

being b the step length of the underlying lattice, $i, j, k \in \{-R, \ldots, R\}$, $(\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}})$ the canonical vectors in each dimension and $p(i,j,k), q(i,j,k) \in \{1, \ldots, N_h\}$ indices of the corresponding pair. Concatenating the indices $i, j, k \in \{-R, \ldots, R\}$ using a one-dimensional index $r \in \{1, \ldots, N_w\}$, with $N_w = (2R+1)^3$, expression (5) can be written as follows

$$\begin{aligned}\gamma(\mathbf{h}) &\approx \frac{1}{2N_h} \sum_{\{(p,q):\mathbf{x}_q=\mathbf{x}_p+\mathbf{h}\}} \left[\sum_{r=1}^{N_w} w_r (z_{p(r)} - z_{q(r)}) \right]^2 \\ &= \mathbf{w}^T \boldsymbol{\Delta}(\mathbf{h})^T \boldsymbol{\Delta}(\mathbf{h}) \mathbf{w} \\ &= \|\boldsymbol{\Delta}(\mathbf{h})\mathbf{w}\|^2 \end{aligned} \tag{8}$$

with $\mathbf{w} \in \mathbb{R}^{N_w}$ the kernel weights vector and $\boldsymbol{\Delta}(\mathbf{h}) \in \mathbb{R}^{N_h \times N_w}$ the differences matrix defined as

$$\boldsymbol{\Delta}(\mathbf{h}) = \frac{1}{\sqrt{2N_h}} \left[\begin{array}{c|c|c} z_{p_{1(1)}} - z_{q_{1(1)}} & \cdots & z_{p_{1(N_w)}} - z_{q_{1(N_w)}} \\ \vdots & \vdots & \vdots \\ z_{p_{N_h(1)}} - z_{q_{N_h(1)}} & \cdots & z_{p_{N_h(N_w)}} - z_{q_{N_h(N_w)}} \end{array}\right] \tag{9}$$

being $\{(p_1, q_1), \ldots, (p_{N_h}, q_{N_h})\}$ an enumeration of the set $\{(p, q) : \mathbf{x}_q = \mathbf{x}_p + \mathbf{h}\}$.

According to Eq. (8), the semi-variogram values for each lag distance $h = \|\mathbf{h}\|$ can be approximated by a convex quadratic function of the kernel weights $\mathbf{w}$, using the random field $\mathbf{Z}$ to build the differences matrix $\boldsymbol{\Delta}(\mathbf{h})$. In some cases, when large values of N_h and N_w are being used, it can be computationally expensive to calculate and store $\boldsymbol{\Delta}(\mathbf{h})$. Instead, its value can be obtained by executing a computer program that calculates $\mathbf{Y}$ according to a discretized version of Eq. (1) (convolving $\mathbf{Z}$ and $\mathbf{w}$), and after that it calculates the semi-variogram values of $\mathbf{Y}$ for different lag distances.

3 Numerical Solution

Using Eq. (8) an ordinary least-squares (OLS) type of cost function can be defined, measuring the squared euclidean distance between a vector of experimental semi-variogram values at different lag distances $\mathbf{h}_1, \ldots, \mathbf{h}_n$ and the simulated semi-variogram depending on the weight values $\mathbf{w}$. Since the fitting at the origin of the semi-variogram curve must be emphasized, a WLS type of cost function must be used. Each term is multiplied by the squared inverse of the observed experimental semi-variogram values, adding a small positive constant $\epsilon > 0$ to assure positiveness of the denominator

$$\text{cost}_{\text{WLS}}(\mathbf{w}) = \sum_{i=1}^{n} \left(\frac{\|\boldsymbol{\Delta}(\mathbf{h}_i)\mathbf{w}\|^2 - (\gamma^{\text{target}}(\mathbf{h}_i) + \epsilon)}{\gamma^{\text{target}}(\mathbf{h}_i) + \epsilon} \right)^2 \tag{10}$$

Several constraints can be added to the problem in order to reduce the size of the search space and get more accurate results. In this work the following constraints have been considered:

(i) *Upper and lower bounds* $w_i \in [0, B]$ for some large constant $B > 0$. With this constraint, the convolved values y_i will have mean 0 and variance $\sum_i w_i^2$, with an upper bound equal to $N_w \times B^2$. If the initial kernel is sufficiently close to local optima and small and controlled perturbations of weights are allowed in optimization process, the upper bound is never reached in practice. However, it is included for theoretical reasons (compactness of the feasible set).

(ii) *Piecewise radial symmetry* the number of variables can be reduced using the piecewise representation of the kernel presented in Eq. (3). This constraint can be assessed by imposing $w_p = w_q$ if $\exists r_p, r_q \geq 0$ and $\exists j \in \{1, \ldots, m\}$ such that $r_p, r_q \in \{r \geq 0 : \frac{(j-1)R}{m} < r \leq \frac{jR}{m}\}$, $w_p = f(r_p)$ and $w_q = f(r_q)$. With this constraint the total number of variables to be fitted is the number of radial bands

m, which must be less or equal to n, the number of experimental semi-variogram observations (number of lags).

The previous constraints are linear, so they can be written in matrix form as $\mathbf{Aw} = \mathbf{b}$, allowing an easy feasibility checking. Additional non-linear constraints can be added, for example imposing a fixed value for the semi-variogram sill, as $sill = \sum_i w_i^2$. This constraint is equivalent to $\mathbf{C}(\mathbf{0}) = \int_U f(r)^2 \mathrm{d}r$ and can replace the boundedness condition (i). However, as detailed in Bazaraa et al. (1993), checking the feasibility of non-linear constraints is not straightforward and can add considerable amount of computational work. For this reason the authors decided to include only linear constraints. Additionally, since the convolved data can be standardized using a normal score transformation with the same table as the standardized experimental data, the sill of the non-standardized semi-variogram is not relevant in the proposed framework.

Mathematically, minimizing function (10) subject to the previous constraints corresponds to a 4th-order polynomial minimization (non-linear least-squares on $\mathbf{w}$) linearly constrained. It can be seen as an ill-posed non-linear inverse problem, in the sense that it searches for optimal kernels (local optima) which completely model the structural property of the underlying phenomenon, based only in a few observed measurements. As explained by Mosegaard and Tarantola (1995) and Tarantola (2004), Monte Carlo methods are the best alternative to obtain samples of the a posteriori probability distribution of the model, in this case the kernel weights. This means that the solution of the inverse problem is not a model but a collection of models that are consistent with both the data (experimental semi-variogram values) and the a priori information (initial kernel weights, used as starting point). Even though the non-linearities are weak and highly efficient techniques can be used to solve the problem, such as the Gauss-Newton method for non-linear least-squares problems described in Nocedal and Wright (2006), the probabilistic approach was chosen. This decision is based in the potentially prohibitive amount of computational resources required to store the first-order derivatives of the cost function from Eq. (10), which involves the storage of matrix $\Delta(\mathbf{h}_i)$ for each i. Additional techniques, like using a regularization term as explained by Tikhonov and Arsenin (1977), are also considered in this work. As shown in several geophysics-based inverse problems, prior information can be introduced in the problem as a penalty term that regularizes the cost function. In this case, the regularization term is added to the cost function (10) as

$$\mathrm{cost}_{\mathrm{WLS}}^{\mathrm{reg}}(\mathbf{w}, \beta) = \mathrm{cost}_{\mathrm{WLS}}(\mathbf{w}) + \beta R(\mathbf{w}) \tag{11}$$

where $\beta \geq 0$ is the regularization parameter and $R(\mathbf{w})$ can be a measure of the smoothness of the kernel or prior information imposed over the feasible solutions. Common choices for $R(\mathbf{w})$ are linear combinations of $\|\mathbf{w} - \mathbf{w}^{\mathrm{prior}}\|_s$, $\|\mathbf{w}\|_s$, $\|\nabla \mathbf{w}\|_s$ and $\|\nabla^2 \mathbf{w}\|_s$ with $s \in \{0, 1, 2, \infty\}$. $\mathbf{w}^{\mathrm{prior}}$ is defined as the prior information vector. $\nabla \mathbf{w}$ and $\nabla^2 \mathbf{w}$ are defined as the discrete first- and second-order three-dimensional spatial derivatives of the kernel $f(r)$.

The non-linear optimization problem that must be solved to infer the a posteriori kernel weight values is as follows

$$\begin{aligned} \underset{\mathbf{w}\in\mathbb{R}^{N_w}}{\text{minimize}} \quad & \text{cost}^{\text{reg}}_{\text{WLS}}(\mathbf{w}, \beta) \\ \text{subject to} \quad & w_i \leq B, \quad \forall i \\ & w_i \geq 0, \quad \forall i \\ & w_i = w_j, \ i, j \text{ are piecewise radially symmetric} \end{aligned} \tag{12}$$

4 Implementation

Simulated annealing (Kirkpatrick et al. 1983; Geman and Geman 1984; Deutsch 1992; Peredo and Ortiz 2011) is used to obtain approximate solutions of problem (12). The calculation of the current simulated semi-variogram values $\{\|\mathbf{\Delta}(\mathbf{h}_i)\mathbf{w}\|^2\}_{i=1}^n$ must be done with the same parameters used to obtain the target experimental semi-variogram $\{\gamma^{\text{target}}(\mathbf{h}_i)\}_{i=1}^n$. If the convolved values y_i are also required to follow the same distribution (histogram), the user must standardize the experimental data before obtaining the target semi-variogram values, and back transform the convolved values using the transformation table from experimental data.

Algorithm 1 summarizes the methodology. Important aspects to consider in the proposed algorithm include the relation between the spacing of the regular lattice and the semi-variogram spacing parameters, and the feasibility enforcement policy. The regular lattice Ω_h can be defined in any way with the only requirement that it contains enough data to reproduce the semi-variogram using all lags defined in the parameters. For instance, if a variogram uses a lag distance of 2.0 and Ω_h has a step length of 4.0, the resulting variogram will have holes in each step non-divisible by 4.0 (assuming that the lag tolerance is sufficiently small). Feasibility enforcement consists in modifying each potential new state of the weights vector, denoted $\tilde{\mathbf{w}}^k$, in order to belong to the interval $[0, B]$ and satisfy the banded radial-symmetric constraint. Only controlled perturbations are allowed, defined as

$$\tilde{\mathbf{w}}_i^k = \begin{cases} \mathbf{w}_i^k(1 \pm \delta), & \text{if } f(r^*) = \mathbf{w}_i^k \\ \mathbf{w}_i^k, & \text{if not} \end{cases} \tag{13}$$

with $\pm\delta \in (-1, 1)$ a percentage of increment or decrement and $r^* \in [0, R]$ a random radial distance. If $\mathbf{w}_i^k(1 \pm \delta) \notin [0, B]$, a new percentage $\pm\delta$ is generated until the constraint is fulfilled.

5 Practitioner's Usage

From a practitioner's point of view, the main goal is to obtain several realizations of a random field where each one preserves the same spatial structure of the input dataset. In order to run a conventional geostatistical simulation algorithm, like sequential gaussian simulation (Alabert 1987; Deutsch 2002; Deutsch and Journel 1998; Isaaks 1990), the first task is to model the variogram parameters, typically requiring expert knowledge from the practitioner in order to fit the modeled variogram to the experimental one. If the practitioner has not enough experience, this task can be cumbersome and error

Input: Initial kernel weights $\mathbf{w}^0 := f^0(r)$ and radius R as in Eq. (7). Dataset values $\mathbf{V}$ in a domain Ω (can be irregular and non-gaussian). Number of lags n, azimuth φ, dip θ and variogram tolerances tol. Initial temperature T_0, cooling scheme λ, number of iterations N_{iters}, convergence tolerance ϵ_{tol}. Regularization parameter $\beta \geq 0$.

1: $\widehat{\mathbf{V}} \leftarrow$ Apply a normal score transformation over $(\mathbf{V}, \Omega)$
2: $\{\gamma^{target}(\mathbf{h}_i)\}_{i=1}^n \leftarrow$ Compute semi-variogram over $(\widehat{\mathbf{V}}, \Omega)$ using $(n, \varphi, \theta, tol)$
3: $\Omega_h \leftarrow$ Generate a regular lattice Ω_h, with similar dimensions of Ω
4: $\mathbf{Z} \leftarrow$ Generate random field with Dirac covariance defined in Ω_h
5: $\mathbf{Y}^0 \leftarrow$ Convolute $\mathbf{w}^0 := f^0(r)$ and $\mathbf{Z}$ in domain Ω_h using formula (1)
6: $\{\|\mathbf{\Delta}(\mathbf{h}_i)\mathbf{w}^0\|^2\}_{i=1}^n \leftarrow$ Compute semi-variogram over $(\mathbf{Y}^0, \Omega_h)$ using $(n, \varphi, \theta, tol)$
7: $cost^0 \leftarrow$ Calculate cost function of formula (11) using $\{\|\mathbf{\Delta}(\mathbf{h}_i)\mathbf{w}^0\|^2\}_{i=1}^n$ and $\beta \geq 0$
8: $k \leftarrow 0,\ T \leftarrow T_0$
9: **while** $\frac{cost^k}{cost^0} < \epsilon_{tol}$ **or** $k > N_{iters}$ **do**
10: $\quad \tilde{\mathbf{w}}^k \leftarrow$ Modify $\mathbf{w}^k$
11: $\quad$ Force feasibility of $\tilde{\mathbf{w}}^k$ using constraints of problem (12)
12: $\quad \tilde{\mathbf{Y}}^k \leftarrow$ Convolute $\tilde{\mathbf{w}}^k := \tilde{f}^k(r)$ and $\mathbf{Z}$ in domain Ω_h using formula (1)
13: $\quad \{\|\mathbf{\Delta}(\mathbf{h}_i)\tilde{\mathbf{w}}^k\|^2\}_{i=1}^n \leftarrow$ Compute semi-variogram over $(\tilde{\mathbf{Y}}^k, \Omega_h)$ using $(n, \varphi, \theta, tol)$
14: $\quad \widetilde{cost}^k \leftarrow$ Calculate cost function of formula (11) using $\{\|\mathbf{\Delta}(\mathbf{h}_i)\tilde{\mathbf{w}}^k\|^2\}_{i=1}^n$ and $\beta \geq 0$
15: $\quad \tau \leftarrow 1$ if $\widetilde{cost}^k < cost^k$; $e^{\frac{cost^k - \widetilde{cost}^k}{T \cdot cost^k}}$ otherwise
16: $\quad (\mathbf{w}^{k+1}, cost^{k+1}) \leftarrow (\tilde{\mathbf{w}}^k, \widetilde{cost}^k)$ if `rand(0,1)` $< \tau$; $(\mathbf{w}^k, cost^k)$ otherwise
17: $\quad k \leftarrow k + 1$
18: $\quad$ Apply cooling scheme λ to the temperature T
19: **end while**
20: $\mathbf{Y}^*_{back} \leftarrow$ Back transform values of best convolved image $\mathbf{Y}^*$

Output: Optimal kernel weights $\mathbf{w}^*$ and best convolved and back transformed images $\mathbf{Y}^*$ and $\mathbf{Y}^*_{back}$

Algorithm 1: General steps of simulated annealing technique to obtain approximate solutions of problem (12) with optional regularization term

prone. The long-term objective of this contribution is to provide a tool for the practitioner to automatically generate one or more realizations of the random field without requiring a fitted parametric variogram model. A possible work-flow using this tool can be viewed in Algorithm 2.

1: $\gamma^{target} \leftarrow$ Obtain experimental variogram values.
2: $\mathbf{w}^0 \leftarrow$ Set initial kernel weights.
3: $\Omega \leftarrow$ Set a simulation lattice.
4: $\Omega_{\min} \leftarrow$ Calculate the minimal domain size needed to obtain statistically valid variogram values.
5: $\{\mathbf{w}^{aposteriori}\} \leftarrow$ Estimate the a posteriori model distribution by approximately solving problem (12) with inputs γ^{target}, $\mathbf{w}^0$ and several realizations $z_1, \ldots, z_n$ with Dirac covariance in $\Omega_{\min}$.
6: $y \leftarrow$ Generate one realization of $\mathbf{Y}$ convolving a sample from the a posteriori model distribution and a realization z with Dirac covariance in Ω.
7: Store y in a file.
8: **while** More realizations of $\mathbf{Y}$ in Ω are needed **do**
9: $z \leftarrow$ Obtain realization of $\mathbf{Z}$ (Dirac covariance) in Ω.
10: $y \leftarrow$ Convolve z with a sample from the a posteriori model distribution in Ω (forward moving averages).
11: Store y in a file.
12: **end while**

Algorithm 2: Practitioner's work-flow using the proposed method to generate realizations without modeling the variogram

6 Examples

Two sets of examples are presented: using synthetic datasets and real datasets. The synthetic examples are designed to show scenarios where the initial semi-variogram values are different from the target ones, using well-known analytic kernels.

In all cases, initial and final/best images are presented. Comparisons between the initial, final and targeted semi-variogram and the kernel weights are also shown. To draw the a posteriori distribution, 30 runs of the algorithm were executed in order to collect samples in each case. The average convergence of the algorithm is also provided via the minimization of the cost function over 6000 iterations using a cooling scheme $T \leftarrow 0.5T$ every 150 iterations (restarting to the initial temperature every 10 cooling

steps). Initial temperature is set individually in each example, keeping a reasonable value after some running trials.

6.1 Synthetic Data Examples

In these examples the semi-variogram of the convolved field $\mathbf{Y}$ is calculated over three-dimensional images where each node is located in a regular lattice of step length 1.0. The semi-variogram is calculated using the omni-directional lag vector with lags of separation $h = 1.0$ with a small tolerance, using a normal score transformation to produce the target data. All these examples use random fields $\mathbf{Z}$ with dimension $\left(n_x + \lceil \frac{R}{h} \rceil\right) \times \left(n_y + \lceil \frac{R}{h} \rceil\right) \times \left(n_z + \lceil \frac{R}{h} \rceil\right)$, where $n_x \times n_y \times n_z$ is the original dimension, and keeping a buffer zone of $\lceil \frac{R}{2h} \rceil$ nodes at the boundaries. The regularization parameter used in these examples is $\beta = 0$, letting the semi-variogram residual of Eq. (11) be the only active term in the cost function.

The first example (Fig. 1) illustrates convergence of the algorithm when it is initialized with a very smooth semi-variogram, while the targeted semi-variogram is less continuous. Specifically, it starts with gaussian values for each kernel weight

$$f(r) = \sqrt{\frac{4}{a^2\pi}} e^{\frac{-2r^2}{a^2}}, \quad a = 9.0 \tag{14}$$

with the target specified as three-dimensional spherical kernel weights

$$f(r) = \begin{cases} \frac{2}{a\sqrt{\pi}} & \text{for } r \leq \frac{a}{2} \\ 0 & \text{for } r > \frac{a}{2} \end{cases}, \quad a = 9.0 \tag{15}$$

The a posteriori models, depicted in Fig. 1f, are sampled with the proposed algorithm using a small three-dimensional domain $\Omega_{\min}$ with dimension $14 \times 14 \times 14$. Convolving the initial kernel Eq. (14), target kernel Eq. (15), and one of the sampled a posteriori kernels with a domain Ω of size $64 \times 64 \times 1$, standardized simulated images can be obtained as shown in Fig. 1a–c.

In contrast, the second example, depicted in Fig. 2, begins with short range continuity [spherical kernel from Eq. (15)], while the targeted semi-variogram is much more continuous [gaussian kernel from Eq. (14)]. As for the first example, the a posteriori models, depicted in Fig. 2f, are sampled with the proposed algorithm using a small three-dimensional domain $\Omega_{\min}$ with dimension $14 \times 14 \times 14$. Convolving the initial kernel, target kernel, and one of the sampled a posteriori kernels with a domain Ω of size $64 \times 64 \times 1$, standardized simulated images can be obtained as shown in Fig. 2a–c.

In both instances, the final/best image matches the target image remarkably well. This is further confirmed when the best and target semi-variograms are compared.

These examples can be considered as validity tests, in the sense that the obtained results are in concordance with the author's expectations. The target images correspond to the standardized convolution $\mathbf{Z} * \mathbf{w}^{\text{target}}$, with $\mathbf{Z}$ a standard normal random field and $\mathbf{w}^{\text{target}}$ acting as the target weight values (spherical in the first example and gaussian in

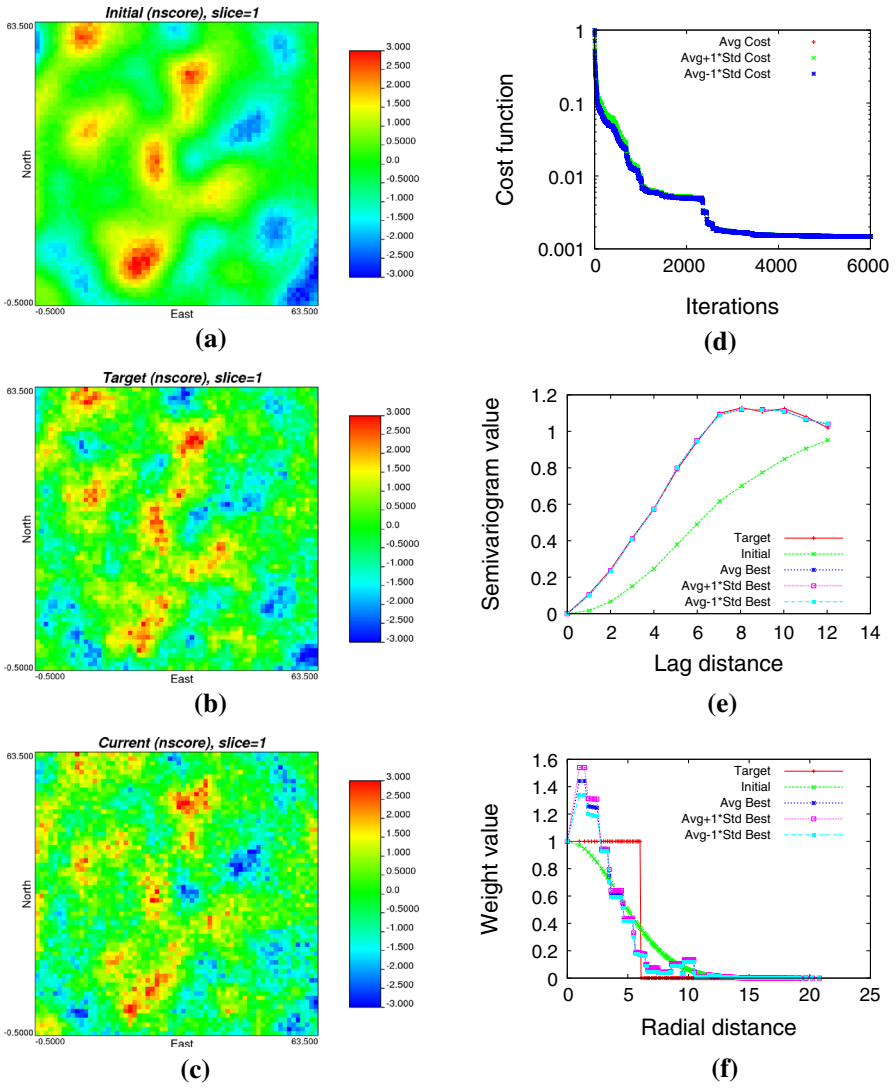

Fig. 1 Example 1: synthetic scenario with $64 \times 64 \times 1$ nodes in the simulation domain. **a** Initial, **b** best and **c** target simulated images, **d** cost function, **e** semi-variograms, and **f** kernel weights distributions. Initial semi-variogram is generated by a gaussian kernel and the target is generated by a spherical kernel

the second). The a posteriori distribution of models obtained solving the optimization problem generates semi-variogram curves that matches as much as possible the target experimental values. Even though the semi-variogram curves fit, the sampled kernels may not fit exactly with the target kernel weights. As shown in Figs. 1f and 2f, the shape of the a posteriori samples, $\mathbf{w}^{\text{best}}$, is similar to $\mathbf{w}^{\text{target}}$, that is $\mathbf{w}^{\text{best}} \approx \alpha \mathbf{w}^{\text{target}}$ for some $\alpha > 0$. In the second example, the distribution shows a dissimilarity at the tip of the kernel's discontinuity ($r = a/2$). This phenomenon is mitigated on average

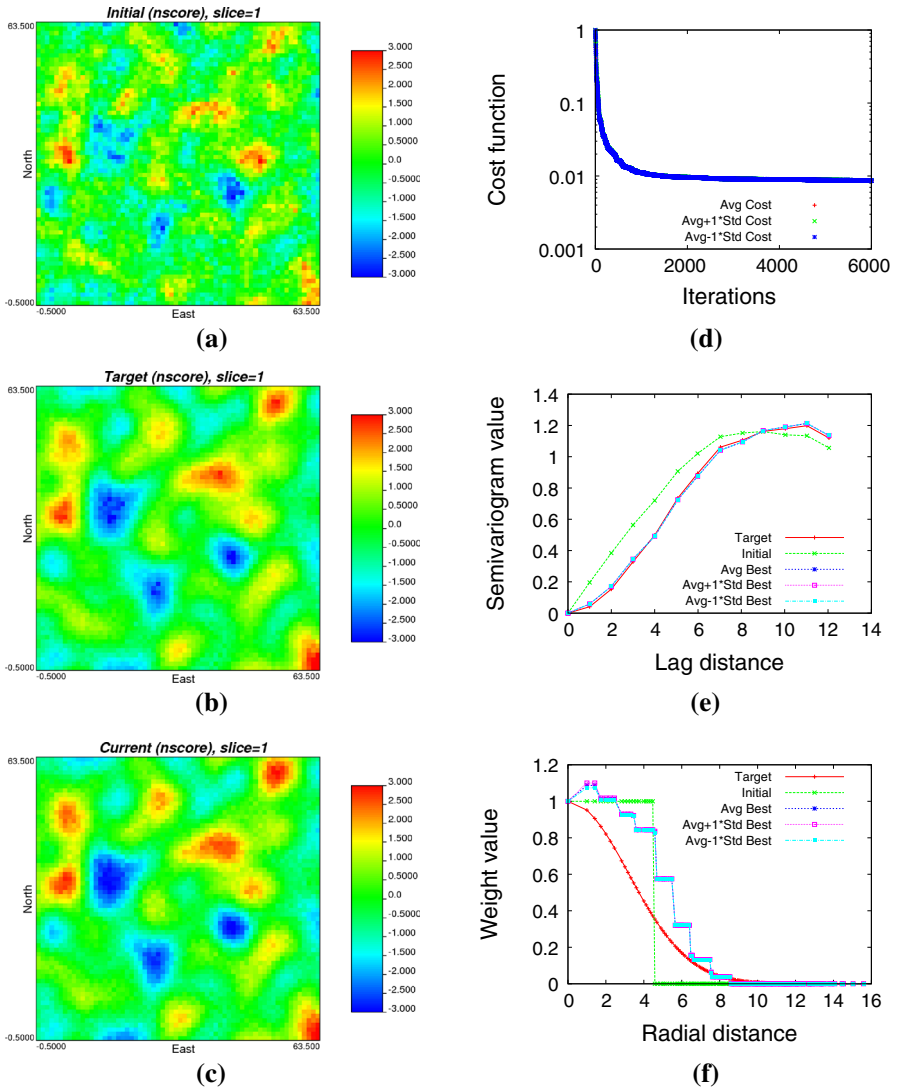

Fig. 2 Example 2: synthetic scenario with $64 \times 64 \times 1$ nodes in the simulation domain. **a** Initial, **b** best and **c** target images, **d** cost function, **e** semi-variograms and **f** kernel weights distributions. Initial semi-variogram is generated by a spherical kernel and the target is generated by a gaussian kernel

by all distribution samples, with higher weight values in the first half ($r \in [0, a/4]$) and lower values in the second half ($r \in [a/4, a/2]$) of the non-zero contribution of the kernel ($r < a/2$).

The quality of the convolved images depends on the distribution of the initial values (which depends of $\mathbf{w}^{\text{initial}}$) and convergence quality of the optimization solver, so in order to refine the approximation, new samples can be generated using different parameters in the optimization algorithm.

In both examples, initial and target kernels f^{initial} and f^{target} were chosen according to Eqs. (14) and (15) scaled to fit $f^{\text{initial}}(0) = f^{\text{target}}(0) = 1.0$, and a weight upper bound $B = 2.0$. These values allow us to visualize in the same scale, $f(r) \in [0.0, 2.0]$, all the a posteriori samples. It is worth to mention that the selection of the upper bound B controls the convergence rate and defines the size of the feasible set of solutions. However, the previous scale and upper bound selection fit the author's purposes, since the weight perturbations are performed in a smooth and controlled fashion, as explained in Sect. 4.

6.2 Real Data Examples

Two datasets of copper grades are used as base for real data examples. The first dataset is composed by 2376 diamond drill hole samples with information of copper grade composites (Serrano et al. 1996), and the second dataset, also with copper grade information, is composed by 20893 blast hole samples (Emery and Ortiz 2005). Both datasets were collected from well-known Chilean porphyry copper deposits.

The semi-variogram of the convolved field $\mathbf{Y}$ is calculated over three-dimensional regular domains with step length 10.0 for each example. Omni-directional vectors with lag separation $h = 20$, lag tolerance $\text{tol}_{\text{lag}} = 0.2$, and bandwidth tolerance $\text{tol}_{\text{band}} = 200$ for the first example, and $h = 10$, $\text{tol}_{\text{lag}} = 0.2$ and $\text{tol}_{\text{lag}} = 0.001$ for the second example are used as parameters. A normal score transformation is applied to each base dataset to produce the target data. As in the synthetic case, all these examples use random fields $\mathbf{Z}$ with similar dimension, keeping a buffer zone of $\lceil \frac{R}{2h} \rceil$ nodes at the boundaries. Unlike the synthetic examples, no reference or target image is available, so only initial and best images are included in this section.

In the first example (drill holes), the initial weights are generated using a gaussian kernel function with $a = 100.0$, as shown in Eq. (14). The number of lags in the target semi-variogram is $n = 10$. The a posteriori models are sampled with the proposed algorithm using a small three-dimensional domain Ω_{min} with dimension $21 \times 21 \times 21$ and step length 10. The corresponding initial semi-variogram curve, depicted in Fig. 3d, does not match the experimental values, but is sufficiently close to start the optimization process. After 6000 iterations with no regularization term active ($\beta = 0$), the cost function has been reduced on average to approximately 0.25 % of its initial value, obtaining the distribution of kernel weights of Fig. 3e. By looking at the standard deviation bands in this figure, the weight distribution is wider near the origin and narrow far from it. This behavior was not observed in the both synthetic examples, where the standard deviation band was narrow for all radial distances. A slice image generated via convolution of a larger domain Ω with dimension $128 \times 128 \times 128$ and a sample a posteriori kernel, is depicted in Fig. 3b. It shows similar semi-variogram values compared with the experimental ones, and the histogram of the back transformed values has similar shape to the base dataset (Fig. 4).

In the second example (blast holes), depicted in Fig. 5, the initial image is also generated using a gaussian kernel with $a = 100.0$. The number of lags in the target semi-variogram is $n = 15$. The a posteriori models are sampled with the proposed algorithm using a small three-dimensional domain Ω_{min} with dimension $21 \times 21 \times 21$

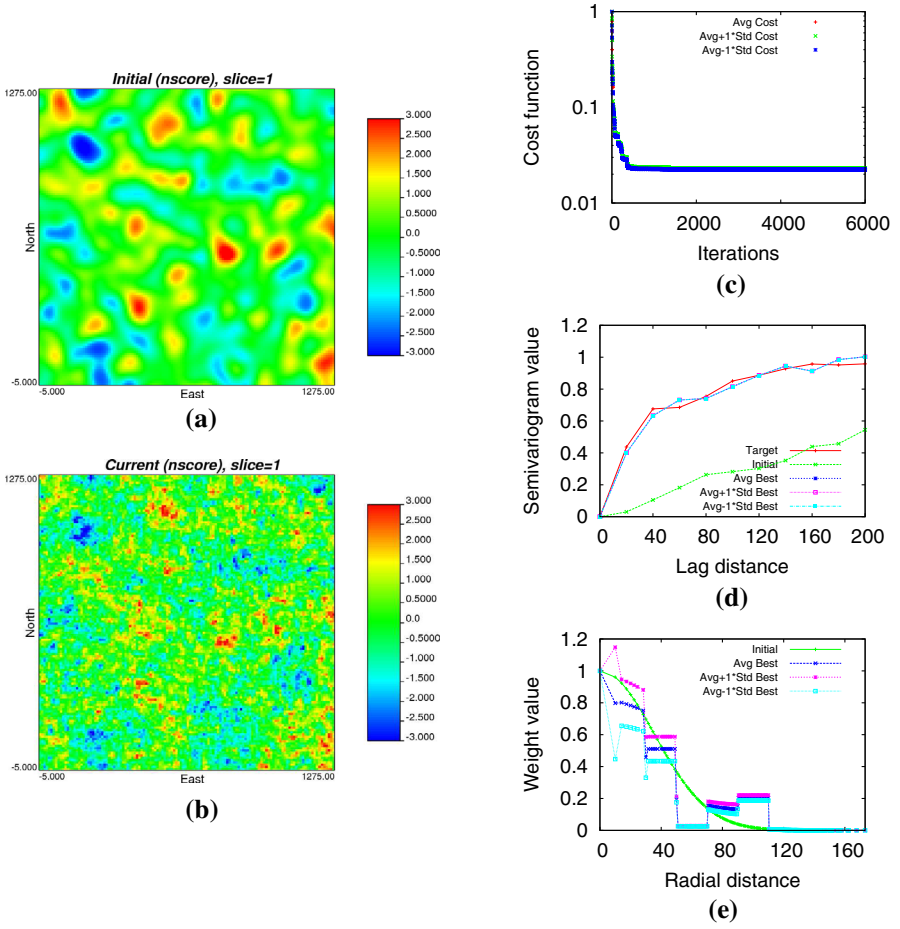

Fig. 3 Example 3: Three-dimensional dataset of 2376 drill hole copper grade composites and a simulation domain of $128 \times 128 \times 128$ nodes. **a** Initial and **b** best images of slice 1; **c** cost function, **d** semi-variograms and **e** kernel weights distributions

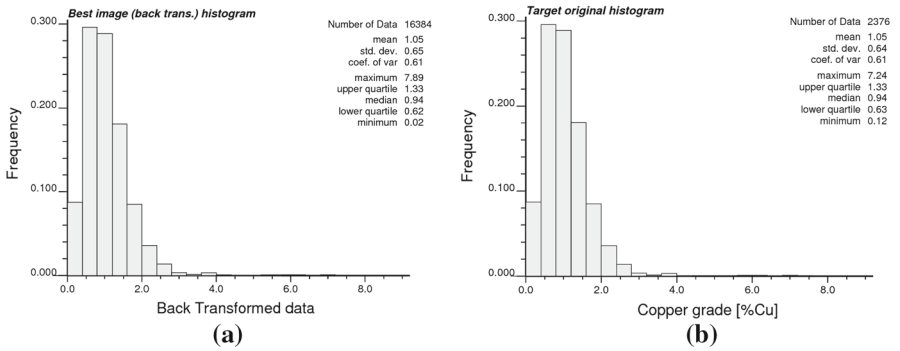

Fig. 4 Example 3: **a** best back transformed and **b** target original data histogram values (drill hole dataset)

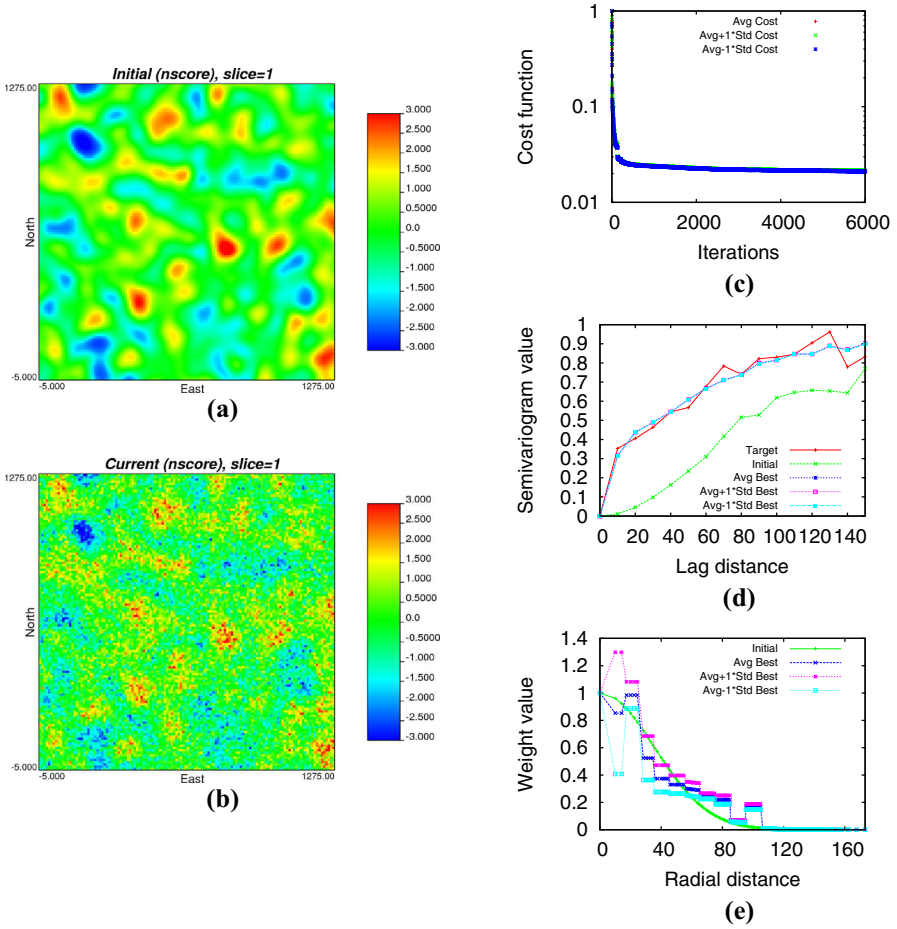

Fig. 5 Example 4: Three-dimensional dataset of 20893 blast hole samples of copper grade composites and a simulation domain of 128 × 128 × 128 nodes. **a** Initial and **b** best images of slice 1; **c** cost function, **d** semi-variograms and **e** kernel weights distributions

and step length 10. After the optimization ends, without using regularization terms ($\beta = 0$), the cost function has been reduced on average to 0.1 % of its initial value, according to Fig. 5c. The a posteriori kernel distribution can be viewed in Fig. 5e. As in the previous example, the weight distribution is wider near the origin and narrow far from it. A slice image generated via convolution of a larger domain Ω with dimension $128 \times 128 \times 128$ and a sample a posteriori kernel, is depicted in Fig. 5b. It shows similar semi-variogram values compared with the experimental ones, and the histogram of the back transformed values has similar shape to the base dataset (Fig. 6).

Compared with the previous example, in this case the a posteriori kernel distribution shows a smoother behavior as a function of the radial distance to the origin. This behavior is due to the granularity used by the semi-variogram parameters, where the

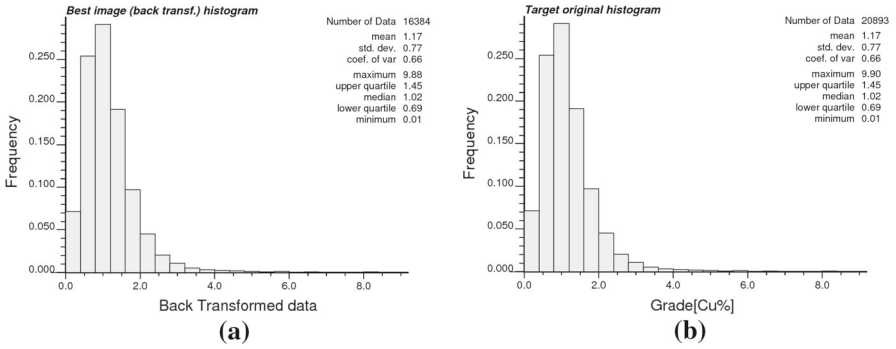

Fig. 6 Example 4: **a** best back transformed and **b** target original data histogram values (blast hole dataset)

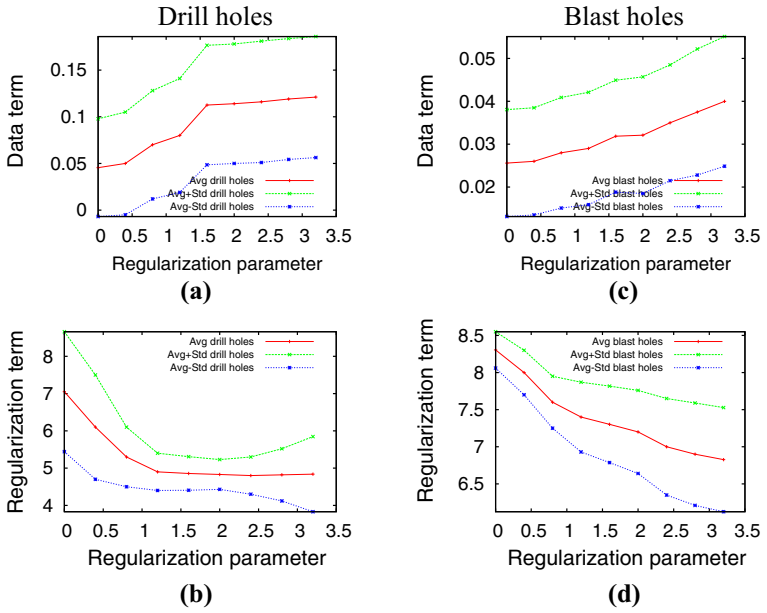

Fig. 7 Regularization term effect in real data examples. **a** Drill holes and **c** blast holes data term values. **b** Drill holes and **d** blast holes regularization term values

first example uses 10 lags with a high tolerance (in lag separation and bandwidth), and the second uses 15 lags with relatively smaller tolerance values.

Both examples show the application of the algorithm generating the desired distribution of kernels, and their corresponding usage to generate a simulation convolving a sample kernel with a larger domain. The structural properties of the convolved images retain the original variogram and histogram shapes calculated in the base dataset. No regularization term was used in the presented examples; however, it is insightful to show the effect of a regularization term in the cost function. In Fig. 7 it can be observed the relative values of the data and regularization terms from cost function (11), $\text{cost}_{\text{WLS}}(\mathbf{w})$, and $R(\mathbf{w})$, respectively. The function $R(\cdot)$ was chosen to control the

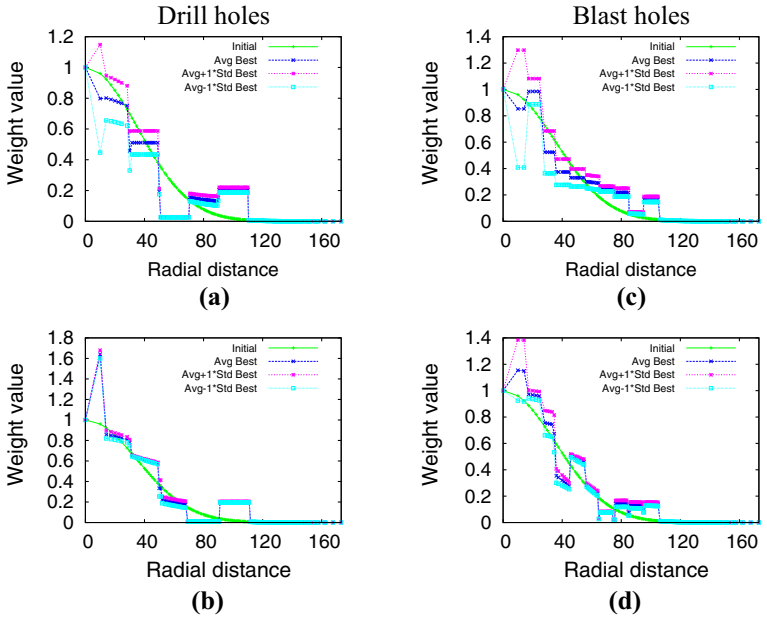

Fig. 8 Examples of **a**, **c** non-regularized and **b**, **d** regularized with $\beta = 1.6$ and $R(\mathbf{w}) = \|\nabla^2\mathbf{w}\|_2$ in each real data example

smoothness of the kernel as a function of the radial distance, that is $R(\mathbf{w}) = \|\nabla^2\mathbf{w}\|_2$. As mentioned in Sect. 3, $\nabla^2\mathbf{w}$ is the discrete second- order spatial derivative of the kernel function. Data and regularization term values associated to the first example (drill holes) are depicted in Fig. 7a, b, and associated to the second example (blast holes) are depicted in Fig. 7c, d. In those figures, it is plotted the average value of the data residual term, defined as $\frac{\text{cost}_{\text{WLS}}(\mathbf{w}^{\text{best}})}{\text{cost}_{\text{WLS}}(\mathbf{w}^{\text{initial}})}$, and the average value of the regularization residual term, defined as $\frac{R(\mathbf{w}^{\text{best}})}{R(\mathbf{w}^{\text{initial}})}$. In both examples the data residual term increases steadily as the regularization parameter β increases from 0.0 to 3.2 in step of size 0.4. On the contrary, the regularization residual term decreases steadily as β increases. In the drill holes example, the increase/decrease in the data/regularization term starts a slowdown since the value $\beta = 1.6$. No slowdown can be observed in the blast holes example. Independently of the slowdown in the increase/decrease of each term, it can be observed that the standard deviation of the regularization term increases with larger values of β. Examples of non-regularized and regularized kernel distributions are shown in Fig. 8a–d, respectively.

Another important aspect of the algorithm is the number of available data used to infer the kernel distribution. The blast holes dataset, with 20893 samples, will be used as test to obtain a measure of accuracy if less samples are available to infer the kernels keeping the same semi-variogram parameters. In Fig. 9 are shown the results using 25 % (5225 samples) and 6 % (1260 samples) selected uniformly at random. As in the blast holes original example, the a posteriori models are sampled using a small three-dimensional domain Ω_{min} of $21 \times 21 \times 21$ nodes and step length 10. Although

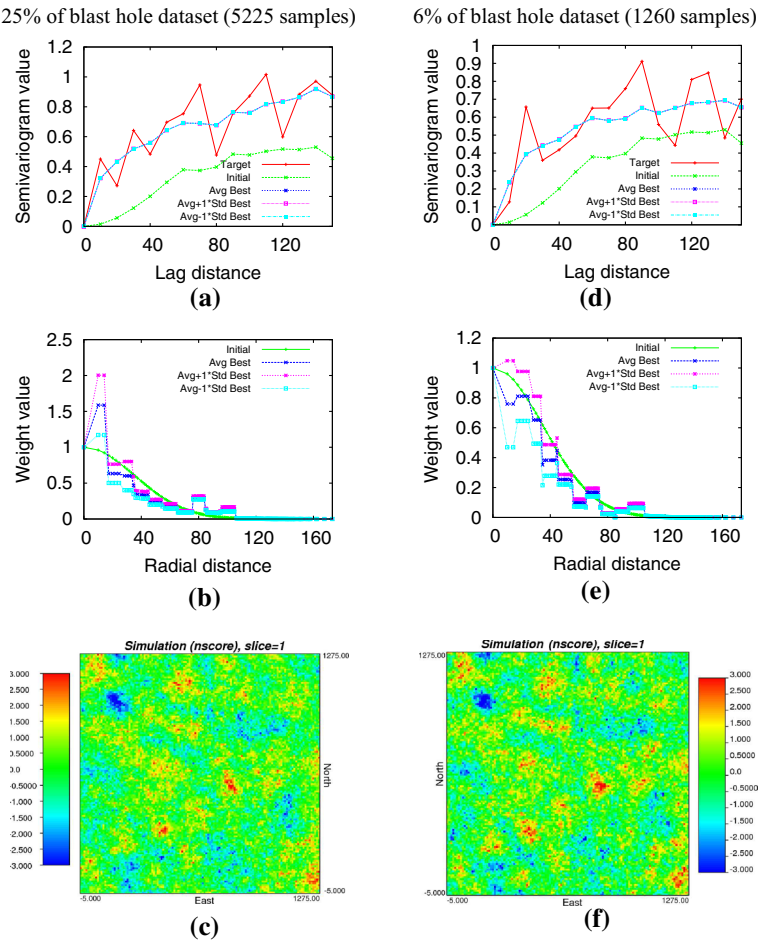

Fig. 9 Effect of uniform sub-sampling the blast hole dataset (20893 samples). **a**, **d** Semi-variograms, **b**, **e** kernel weights distributions and a simulation with $128 \times 128 \times 128$ nodes, using **c** 25 % and **f** 6 % of the dataset

the level of convergence of the cost function in these tests is not as good as using 100 % of the dataset (avg. cost reduction is 14 and 27 %, respectively, and using 100 % the avg. cost reduction is 2 %), the semi-variogram curves are fitted resembling the original scenario with 100 % of the dataset. Two sample standardized simulations using a larger domain Ω of $128 \times 128 \times 128$ nodes are depicted in Fig. 9c, f, each one generated with kernel weights obtained using 25 and 6 % of the dataset, respectively. Both images look similar, but further analysis shows slight differences. In Fig. 10 are depicted the residual error histograms between both standardized images and a standardized simulation of the same size using a kernel weight inferred with 100 % of the blast hole dataset. The three simulations use the same random seed to generate the underlying random field $\mathbf{Z}$ which is convolved with the kernel weights. Both residual errors have mean equal to zero. Using 25 % of the dataset, the standard deviation of

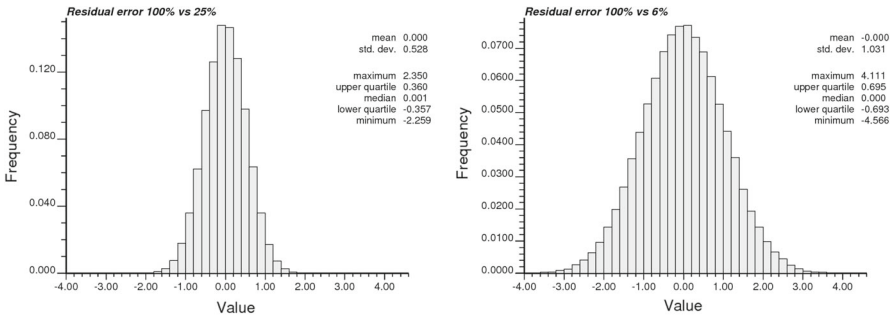

Fig. 10 Residual error histograms between standardized simulations of size 128 × 128 × 128 nodes. **a** Using kernel weights inferred with 100 % of the blast hole dataset versus 25 % of it and **b** 100 % versus 6 % of it

the residual error is $\sigma = 0.528$, while using 6 % its value is $\sigma = 1.031$. It can be observed that the results remain remarkably similar using 25 % of the dataset, on the contrary, using 6 % of the dataset the deviation of the error is considerably larger. This difference in the residual errors can be consequence of the number of sample pairs used to compute $\gamma^{\text{target}}(\mathbf{h}_i)$. While the average number of pairs using 100 and 25 % of the dataset is 656 and 140, respectively, this number is reduced to 33 using 6 % of the dataset, which is at the limit of a practical value. As Journel and Huijbregts (1978) suggests the number of sample pairs must be greater than 30–50 in order to be useful for an experimental semi-variogram.

7 Conclusions

A novel approach to non-parametric Gaussian simulation that does not require explicit variogram modeling by the practitioner is presented. The method is based on a non-linear inverse moving average simulation, where kernel estimations from experimental semi-variogram values are sampled solving a constrained optimization problem. Simulated annealing is used to sample approximate solutions.

As mentioned in Sect. 5, once the posterior distribution of kernels is obtained, several simulations can be generated convolving kernel samples with different random images (Dirac covariance). The method ensures that all those simulations will have a semi-variogram curve that matches as much as possible the experimental curve, and the histograms will follow the same distribution as the target original data.

Early application of this methodology to synthetic and real data examples shows reasonable convergence and kernel weight values. The usage of smoothing regularization terms in the cost function allows the user to shape the profile of the kernel distribution, with the trade-off of losing data residual accuracy. The real data examples presented in this work confirm in part that the effect of the regularization is not negligible and must be studied using more complex datasets. Additionally, the number of available samples in the dataset must guarantee, together with the variogram tolerance parameters, that the number of sample pairs must be greater than a practical lower bound.

These results are encouraging; however, much remains to be done in the development of this method. Future tasks include anisotropy availability, hard data conditioning, and acceleration of the current implementation. Allowing anisotropy in the kernel weights can be modeled by expanding the definition of the piecewise-linear representation of the kernel. As mentioned by Barry and Ver Hoef (1996), the radial bands can be extended to two or three dimensions in a simple way. Conditioning of simulated results could be incorporated a posteriori into the convoluted image. A possible strategy is to use kriging with the modeled semi-variogram values in order to impose the variographic structure defined by the algorithm. A simple code modification of covariance calculation using these values can be implemented. Regarding the computational efficiency, the current implementation was designed to support small scale isotropic scenarios. Even though the usage of a small test domain $\Omega_{\min}$ accelerates the sampling of the posterior distribution of kernels, complex variogram shapes and future anisotropic features make the current implementation unpractical. Using high performance techniques and programming models, like OpenMP (OpenMP Architecture Review Board 2011) and CUDA (NVIDIA Corporation 2010), the current version of this application can be accelerated, making it usable in large-scale real scenarios.

Acknowledgments The authors thankfully acknowledge the computer resources, technical expertise, and assistance provided by the Barcelona Supercomputing Center—Centro Nacional de Supercomputación (Spain) which supports the Marenostrum supercomputer, and the National Laboratory for High Performance Computing (Chile), which supports the Leftraru supercomputer. Additional thanks are owed to industrial supporters of ALGES laboratory, in particular Yamana Gold, as well as the Advanced Mining Technology Center (AMTC) and the whole ALGES team. The authors would also thank two anonymous reviewers for the suggested insightful ideas.

References

Alabert FG (1987) Stochastic imaging of spatial distributions using hard and soft information (Master thesis), Stanford University, Stanford

Barry RP, Ver Hoef JM (1996) Blackbox kriging: Spatial prediction without specifying variogram models. J Agric Biol Environ Stat 1(3):297–322

Bazaraa MS, Sherali HD, Shetty CM (1993) Nonlinear programming : theory and algorithms. Wiley, New York

Bochner S (1949) Fourier transform. Princeton University Press, London

Calder C, Cressie N (2007) Some topics in convolution-based spatial modeling. In: Proceedings of the 56th session of the International Statistical Institute, Lisbon

Calder CA (2008) A dynamic process convolution approach to modeling ambient particulate matter concentrations. Environmetrics 19(1):39–48

Chilès JP, Delfiner P (2012) Geostatistics: modeling spatial uncertainty. Wiley, New York

Cressie N (1985) Fitting variogram models by weighted least squares. J Int Assoc Math Geol 17(5):563–586

Desassis N, Renard D (2013) Automatic variogram modeling by iterative least squares: univariate and multivariate cases. Math Geosci 45(4):453–470

Deutsch CV (1992) Annealing techniques applied to reservoir modeling and the integration of geological and engineering (Well test) data. PhD thesis, Stanford University, Stanford

Deutsch CV (2002) Geostatistical reservoir modeling. Oxford University Press Inc, New York

Deutsch CV, Journel AG (1998) GSLIB: geostatistical software library and user's guide. Oxford University Press, Oxford

Emery X (2010) Iterative algorithms for fitting a linear model of coregionalization. Comput Geosci 36(9):1150–1160

Emery X, Ortiz JM (2005) Estimation of mineral resources using grade domains: critical analysis and a suggested methodology. J S Afr Inst Min Metall 105(4):247–255

Fuentes M (2002) Interpolation of nonstationary air pollution processes: a spatial spectral approach. Stat Model 2(4):281–298
Fuentes M (2002) Spectral methods for nonstationary spatial processes. Biometrika 89(1):197–210
Geman S, Geman D (1984) Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6(6):721–741
Gribov A, Krivoruchko K, Ver Hoef JM (2006) Modeling the semivariogram; new approach, methods comparison, and simulation study. In: Coburn TC, Yarus JM, Chambers RL (eds) Stochastic modeling and geostatistics; principles, methods and case studies. American Association of Petroleum Geologists, Tulsa, pp 45–57
Higdon D (1998) A process-convolution approach to modelling temperatures in the north atlantic ocean. Environ Ecol Stat 5(2):173–190
Higdon D (2002) Space and space-time modeling using process convolutions. In: Anderson C, Barnett V, Chatwin PC, El-Shaarawi AH (eds) Quantitative methods for current environmental issues. Springer, London, pp 37–56
Higdon D, Swall J, Kern J (1999) Non-stationary spatial modeling. In: Bernardo JM et al (eds) Bayesian Statistics 6. Oxford University Press, Oxford, pp 761–768
Isaaks EH (1990) The application of monte carlo methods to the analysis of spatially correlated data. PhD thesis, Stanford University, Stanford
Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London
Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220(4598):671–680
Majumdar A, Gelfand A (2007) Multivariate spatial process modeling using convolved covariance functions. Math Geol 79:225–245
Matérn B (1986) Spatial variation, 2nd edn. Springer, New York
Mosegaard K, Tarantola A (1995) Monte Carlo sampling of solutions to inverse problems. J Geophys Res 100(B7):12431–12447
Nocedal J, Wright SJ (2006) Numerical optimization. Springer, New York
NVIDIA Corporation (2010) NVIDIA CUDA C programming guide, version 3:2. NVIDIA Corporation, Santa Clara
Oliver DS (1995) Moving averages for gaussian simulation in two and three dimensions. Math Geol 27(8):939–960
OpenMP Architecture Review Board (2011) Openmp application program interface. Tech rep. URL:http://www.openmp.org/mp-documents/OpenMP3.1.pdf
Pardo-Igúzquiza E (1999) Varfit: a fortran-77 program for fitting variogram models by weighted least squares. Comput Geosci 25(3):251–261
Peredo O, Ortiz JM (2011) Parallel implementation of simulated annealing to reproduce multiple-point statistics. Comput Geosci 37(8):1110–1121
Rehman S (1995) Semiparametric modeling of cross-semiovariograms. PhD thesis, Georgia Institute of Technolog
Serrano L, Vargas R, Stambuk V, Aguilar C, Galeb M, Holmgren C, Contreras A, Godoy S, Vela I, Skewes M, Stern C (1996) The Late Miocene to Early Pliocene Ró Blanco—Los Bronces copper deposit, central Chilean Andes. In: Camus SR F, Petersen R (eds) Andean copper deposits: new discoveries, mineralizations, Styles and Metallogeny, Society of Economic Geologists, Special Publication no. 5, Littleton, Colorado, pp 119–130
Shapiro A, Botha JD (1991) Variogram fitting with a general class of conditionally nonnegative definite functions. Comput Stat Data Anal 11(1):87–96
Tarantola A (2004) Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics, Philadelphia
Thièbaux HJ, Pedder MA (1987) Spatial objective analysis with applications in atmospheric science. Academic Press, London
Tikhonov AN, Arsenin VY (1977) Solutions of ill-posed problems. Scripta series in mathematics, Winston
Ver Hoef JM, Cressie N, Barry RP (2004) Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT). J Comput Graph Stat 13(2):265–282
Wackernagel H (1994) Cokriging versus kriging in regionalized multivariate data analysis. Geoderma 62(1–3):83–92
Yao T, Journel AG (1998) Automatic modeling of (cross) covariance tables using fast fourier transform. Math Geol 30(6):589–615