

## Constraining distance-based multipoint simulations to proportions and trends



Gregoire Mariethoz<sup>a, b, c, \*</sup>, Julien Straubhaar<sup>a</sup>, Philippe Renard<sup>a</sup>, Tatiana Chugunova<sup>d</sup>, Pierre Biver<sup>d</sup>

<sup>a</sup> University of Neuchâtel, Centre of Hydrogeology and Geothermics, Neuchâtel, Switzerland

<sup>b</sup> University of Lausanne, Institute of Earth Surface Dynamics, Switzerland

<sup>c</sup> The University of New South Wales, School of Civil and Environmental Engineering, Sydney, Australia

<sup>d</sup> TOTAL SA, CTJS Avenue Larribau, 64000, Pau, France

### ARTICLE INFO

#### Article history:

Received 20 October 2014

Received in revised form

3 May 2015

Accepted 12 July 2015

Available online xxx

#### Keywords:

Training image

Probability

Non-stationarity

Spatial modeling

Geostatistics

### ABSTRACT

In the last years, the use of training images to represent spatial variability has emerged as a viable concept. Among the possible algorithms dealing with training images, those using distances between patterns have been successful for applications to subsurface modeling and earth surface observation. However, one limitation of these algorithms is that they do not provide a precise control on the local proportion of each category in the output simulations. We present a distance perturbation strategy that addresses this issue. During the simulation, the distance to a candidate value is penalized if it does not result in proportions that tend to a target given by the user. The method is illustrated on applications to remote sensing and pore-scale modeling. These examples show that the approach offers increased user control on the simulation by allowing to easily impose trends or proportions that differ from the proportions in the training image.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Multiple-point geostatistics (MPS) has emerged in the last years as a family of stochastic simulation tools that can be used in various areas of earth systems imaging. Although the original application domain focused on modeling the internal structures of subsurface deposits ranging from pore-scale (El Ouassini et al., 2008; Okabe and Blunt, 2007; Tahmasebi and Sahimi, 2013; Zhang et al., 2006a) to reservoir-scale (Huysmans and Dassargues, 2012; Ronayne et al., 2008; Yin, 2013), it has thereafter been extended to very different fields such as mining (Goodfellow et al., 2012; Rezaee et al., 2014), soil science (Meerschman et al., 2014), remote sensing (Boucher, 2009; Ge and Bai, 2011; Jha et al., 2013; Stisen et al., 2011; Vannamettee et al., 2014), for modeling the occurrence of rainfall (Oriani et al., 2014; Wojcik et al., 2009), and even in medical imaging (Pham, 2012).

Several multiple-point simulation methods have been developed over the last decade. MPS algorithms can be classified in two categories, based on their underlying statistical approach. The first

category comprises methods that are in the continuity of classical variogram-based geostatistics such as SGS (Deutsch and Journel, 1998) where a probability distribution is inferred for each grid node, which is subsequently sampled for a simulated value. This is the case of the first available MPS method, SNESIM (Strebelle, 2002), as well as its memory-efficient implementation, IMPALA (Straubhaar et al., 2011). One important advantage of computing local probability density functions is that it is possible to perturb them in order to influence the sampling. SNESIM and IMPALA allow using probability aggregation methods (Allard et al., 2012; Krishnan, 2008) to update, on-the-fly, the local probability density functions (pdfs) for each simulated node. In the context of the patchwork simulation (El Ouassini et al., 2008), a method based on the perturbation of the transition probabilities between patches has been proposed by Faucher et al. (2014). Approaches based on the perturbation of a probability allow the user to impose two specific constraints on the resulting simulations:

- 1) To control the global proportions of a categorical variable (also known as servo-system).
- 2) To increase or decrease locally the proportion of a given facies, therefore allowing to generate non-stationary models.

\* Corresponding author. University of Lausanne, Institute of Earth Surface Dynamics, UNIL-Mouline, Geopolis, 1015 Lausanne, Switzerland  
E-mail address: [gregoire.mariethoz@unil.ch](mailto:gregoire.mariethoz@unil.ch) (G. Mariethoz).

This flexibility has been used to impose spatial trends (Koch et al., 2013; Liu, 2006), proportions of land cover categories (Tang et al., 2013) as well as to perturb models in the context of inverse problems (Caers and Hoffman, 2006). In the same vein, perturbing the local probability distributions can be used to impose a specific relationship with covariates by aggregating the local conditional probability with a conditional non-parametric joint distribution. Importantly, when data indicate different proportions than the ones of the training image (for example geophysics or remote sensing measurements), it is often easier to adjust the proportions in the MPS simulation than to obtain a new training image with modified proportions (although modifying the training image has been proposed, see for example (although modifying the training image has been proposed, see for example de Vries et al., 2009).

The second category of MPS algorithms is based on the computation of distances between patterns. It comprises some patch-based methods (Arpat and Caers, 2007; Honarkhah and Caers, 2010; Mahmud et al., 2014; Tahmasebi et al., 2012; Zhang et al., 2006b) as well as the pixel-based method of direct sampling (DS) (Mariethoz et al., 2010). A distance between patterns is used in these algorithms to compare the neighborhood of a node or patch currently simulated with a data event in the training image (or a prototype of data event that represents a group of patterns). One of the main advantages of this approach is that several types of distances can be considered, allowing to simulate continuous variables, or even multiple variables simultaneously (multivariate simulation). Therefore these methods have been preferred for a range of applications. However, using a distance also means that no local probability distribution is computed, and therefore it cannot be perturbed, for example with a technique like the tau-model to influence local trends or global averages.

To this day, the only way of controlling local trends in distance-based MPS is inspired from the method proposed by Chugunova and Hu (2008), which introduces an auxiliary variable, or control map, that determines which parts of the training image correspond to which parts of the simulation. This method is general and offers the possibility to impose complex non-stationarities (for example a trend in the spatial patterns and not only the proportions of categories). Its essential concept is to restrict the sampling of the training image to the locations that are likely to present the desired proportions. For a binary categorical case, controlling proportions with the approach of Chugunova and Hu (2008) is accomplished by computing on the training image  $Z(\mathbf{y})$  a moving average  $Z'(\mathbf{y})$ , and assembling both variables together in a single bivariate training image, which empirically describes the relationship between  $Z$  and its local average  $Z'$ . Target local probabilities can then be imposed in the simulated domain as conditioning data on  $Z'(\mathbf{x})$ . The simulation of the values  $Z(\mathbf{x})$  will then be influenced by those conditioning data through the selection of multivariate patterns in the training image (Mariethoz et al., 2010). The main drawback of this auxiliary variable procedure is that it does not explicitly formulate a target proportion, and therefore only provides an indirect control on the local probability of occurrence of a given category. In essence, it can only reproduce local proportions that are represented at some location in the training image. Since it cannot depart from the proportions of the training image, it is not possible for the user to impose arbitrary proportions constraints. Furthermore, even in the range of proportions represented in the training image it is not possible to precisely match a target local proportion, nor to define global proportion targets.

In this paper we pursue a different route which, in contrast to the auxiliary variable method, is simpler for the user because it does not necessitate to build for each category an auxiliary variable map in the training image (such as for example by using a moving average), and at the same time allows for a precise control of the

proportions in the simulated domain. The approach we propose allows imposing such constraints by a perturbation of the distance function used to compare patterns. It is found that our distance perturbation scheme allows adjusting local and global proportions with a given precision, and also offers control parameters to choose between reproduction of the fine structures and honoring the desired proportions, when these two constraints are not compatible. For example, local proportion beyond the range of what is present in the training image can be imposed on the simulation grid.

This paper is organized as follows: we start giving a brief background on the direct sampling method which is used to illustrate the approach presented. Next, we describe our distance perturbation method. We then analyze the effect and the sensitivity of the main parameters of the method on a simple 2D example, and demonstrate its applicability on two additional synthetic case studies: the first one simulates future remote sensing images of land use in Africa and the second examples simulates the precipitation of deposits in the interstices of a 3D porous sample.

## 2. Using a distance perturbation

### 2.1. Background on direct sampling

The direct sampling algorithm (DS) (Mariethoz et al., 2010) works as follows: for each simulated node a local data event is extracted. The training image is then scanned with this data event (typically in a random or pseudo-random fashion). As soon as the distance  $d\{\cdot\}$  between the data event searched for and the one currently scanned is below a given threshold  $t$ , the corresponding training image value is assigned to the simulated location. A simplified pseudocode of the algorithm is given below:

The key principle is that as soon as the condition mentioned at point 5 is not honored, the corresponding loop is interrupted, drastically accelerating the sampling procedure compared to an exhaustive scan. A maximal scanned fraction of the TI  $f$  can be defined to control the computational time. The algorithm is essentially driven by the three key parameters  $n$  (data event size),  $t$  (threshold) and  $f$ . Some practical recommendations on how to choose the parameters are provided in (Meerschman et al., 2013). The formulation of the distance  $d\{\mathbf{N}_x, \mathbf{N}_y\}$  is important and several possibilities are proposed in (Mariethoz et al., 2010), but it is generally defined such that its values are bounded in the interval  $[0, 1]$ . Acting on the distance is the main lever that allows having control over the algorithm's behavior. Distances are used in a similar fashion in other distance-based MPS algorithms. The method we propose in this paper is based on modifying this distance computation to include a term that takes into account local or global proportion constraints.

### 2.2. Methodology defining errors on pdfs

The problem addressed in this paper is to directly constrain the proportion of simulated categories in Direct Sampling, which has to date not been possible. Such a constraint can be given by a user-provided map that prescribes for each grid node  $\mathbf{x}$  the target proportion for facies  $k$ , and is denoted  $P_t(\mathbf{x}, k)$ , with  $k = 1 \dots K$ . The local proportions constraint can be formulated as a category appearance frequency in the defined area/volume around and including a simulated value  $Z(\mathbf{x})$ . Therefore, the definition of a local proportion as we envision in this paper necessarily involves a neighborhood in a limited volume for computing the local proportions. Under the hypothesis of local ergodicity we may treat these local proportions as univariate probabilities at point  $Z(\mathbf{x})$ . Thus, this volume should be

Input: Simulation grid SG with values denoted  $Z(\mathbf{x})$ , training image TI with values denoted  $Z(\mathbf{y})$ , distance function  $d(\cdot)$ , distance threshold  $t$  (distance under which two neighborhoods are considered sufficiently similar).

1. Define a random simulation path through the SG.
2. **while** # non-informed locations in SG > 0 **do**
3. Choose a non-informed location  $\mathbf{x}$  along the random path in SG and identify the neighborhood  $\mathbf{N}_x$  consisting of the  $n$  closest informed nodes around  $\mathbf{x}$ , denoted  $\mathbf{x}_i, i=1 \dots n$ .
4. Initialize  $d\{\mathbf{N}_x, \mathbf{N}_y\} = \infty$
5. **while**  $d\{\mathbf{N}_x, \mathbf{N}_y\} \geq t$  **do** \*
6. Sample a location  $\mathbf{y}$  in TI.
7. Define  $\mathbf{N}_y$  as the neighborhood centered on  $\mathbf{y}$  having identical lag vectors as  $\mathbf{N}_x$ .
8. Compute the distance  $d\{\mathbf{N}_x, \mathbf{N}_y\}$ .
9. **end while**
10. Assign  $Z(\mathbf{x}) = Z(\mathbf{y})$
11. **end while**

Output: Completed simulation SG

\* Actual implementations may vary. In particular, if no distance under the threshold is found, the scan is interrupted and the best candidate so far is used.

large enough to yield univariate statistics that are representative of a local frequency of a facies at  $Z(\mathbf{x})$ . This neighborhood can be viewed as the spatial support of the imposed local proportions. In the case of a target global proportion, the proportion of a category  $k$  is given by a target  $P_t(k)$  which is considered on the entire simulation domain.

We define the current proportion of facies  $k$  at any simulated node  $\mathbf{x}$ ,  $P_c(\mathbf{x}, k)$ , based on the informed values in the neighborhood  $\mathbf{x}_i$  as well as a candidate value  $Z(\mathbf{y})$  from the TI. Our goal is that this current proportion is similar to the desired target proportion in the simulation,  $P_t(\mathbf{x}, k)$ . The concept is that the distance  $d\{\mathbf{N}_x, \mathbf{N}_y\}$  used in the DS can be perturbed on-the-fly in order to penalize data events that decrease the match between the local proportion in the simulation and the user-defined target proportion. It uses a similar philosophy as the servo-system procedure implemented in the SNESIM algorithm (Strebelle, 2002), but on a local scale and applied to the computation of distances instead of probabilities.

During the simulation, with  $Z(\mathbf{x})$  that is not yet simulated, the local proportion at  $\mathbf{x}$  can be estimated as

$$P_c(\mathbf{x}, k) = \frac{1}{n} \sum_{i=1}^n [Z(\mathbf{x}_i) = k], \quad (1)$$

the expression  $[Z(\mathbf{x}_i) = k]$  denoting an indicator variable that is a test returning a value 0 if false and 1 if true. This local proportion

will be updated by accounting for the candidate value. We can define the local proportion corresponding to the proposed value  $Z^*(\mathbf{x})$  for which the DS algorithm is currently computing the distance  $d\{\mathbf{N}_x, \mathbf{N}_y\}$ :

$$P_c^*(\mathbf{x}, k) = \frac{1}{n+1} \left\{ \sum_{i=1}^n [Z(\mathbf{x}_i) = k] + [Z^*(\mathbf{x}) = k] \right\}. \quad (2)$$

It corresponds to the estimated local proportion if the candidate value  $Z^*(\mathbf{x})$  was accepted by the sampling procedure. To achieve our goal, which is that the current local proportions estimated from the simulation are the same as the target proportions, we want to accept values in the simulation such that the following equations are honored as much as possible:

$$P_c^*(\mathbf{x}, k) = P_t(\mathbf{x}, k), \quad k = 1 \dots K \quad (3)$$

In order to enforce (6), the proposed methodology preferentially accepts values for location  $\mathbf{x}$  that tend to minimize the difference between target and current proportions, through a modification of the distance measure. For each candidate value  $Z^*(\mathbf{x})$ , we consider the root mean squared error of the frequency for all facies:

$$E_p = \sqrt{\frac{1}{K} \sum_{k=1}^K [ \{ P_c^*(\mathbf{x}, k) - P_t(\mathbf{x}, k) \}^2 ]} \quad (4)$$

Equation (4) has the advantage to be bounded in the interval  $[0,1]$ , therefore comparable to the distance values that have the same bounds. The distance used in DS is then perturbed, the perturbation being a function of  $E_p$ . We use an additive perturbation where a quantity proportional to  $E_p$  is added to the original distance:

$$d^* = d\{\mathbf{N}_x, \mathbf{N}_y\} + wE_p \quad (5)$$

$w$  is a user-defined weight that can strengthen or weaken the constraint on local proportions. It results from equation (5) that  $d^*$  is now bounded in the interval  $[0, 1 + w]$ . All data events that would usually be unacceptable (i.e. when  $d\{\mathbf{N}_x, \mathbf{N}_y\} \geq t$ ) still cannot be used in the simulation. However, among the pool of acceptable data events (for which  $d\{\mathbf{N}_x, \mathbf{N}_y\} < t$ ) a penalty proportional to the local proportion mismatch is applied. The additive perturbation of equation (5) therefore ensures that the distance on accepted patterns is under the threshold for all simulated values, resulting in a good reproduction of the training image patterns.

### 2.3. Modeling choices and alternative implementations

For local proportion constraints, the neighborhood  $\mathbf{N}_x$  used for comparing patterns in the DS and the neighborhood used for defining local proportions can be different. For all cases illustrated in this paper we have tested the use of different neighborhoods and found that the results are not significantly different from those obtained with a single neighborhood definition for both uses. In this study we therefore only illustrate the case where these neighborhoods are the same.

This said, we cannot exclude that there are cases where the neighborhood of the local proportion can play a role. This neighborhood corresponds to the area or volume (in 3D, a sphere with user-defined radius  $R$ ) that plays the role of a support for local univariate statistics. In practice, this radius could be derived from the origin of the local proportions constrain. For example, if the local proportions are constructed by a kernel smoothing applied to point measurements, the kernel bandwidth can be seen as the support of the local statistics. If the local proportions are derived from an auxiliary attribute, the resolution of this attribute is an indication of the support of the local probability.

Another modeling choice we did is to use a root mean squared error in equation (4). Other error measures between the current and target local proportions could be used, such as for example outlined in Bennett et al. (2013). Moreover, instead of penalizing the distance as proposed in equation (5), another possible way of implementing our method can be to consider two distances simultaneously. The acceptance criterion in step 5 of the algorithm is then replaced by a criterion imposing that both the distance  $d\{\mathbf{N}_x, \mathbf{N}_y\}$  is below the threshold  $t$  and the error on the proportion  $E_p$  is also lower than a fixed error norm. If one of these two criteria is not respected then the proposed candidate value is rejected. This technique has been implemented and tested as well, but results are again not significantly different from the one shown with the penalization approach and we therefore only consider the presented approach in this paper for the sake of simplicity.

### 2.4. Distance perturbation for global proportion constraints

An identical on-the-fly approach can be adopted to impose global proportion constraints. The only difference being that  $P_c$  and  $P_t$  are not localized, hence the current proportion is simply defined as the proportion of each facies:

$$P_c(k) = \frac{\#[Z(\mathbf{x}) = k]}{\#[Z(\mathbf{x})]}, \quad (6)$$

$\#[Z(\mathbf{x})]$  denoting the number of nodes simulated so far, and  $\#[Z(\mathbf{x}) = k]$  denoting the number of simulated nodes that have been attributed a facies value of  $k$ . The proposed global proportion is

$$P_c^*(k) = \frac{\#[Z(\mathbf{x}) = k] + [Z^*(\mathbf{x}) = k]}{\#[Z(\mathbf{x})] + 1} \quad (7)$$

and the error on the proportions is a non-localized form of equation (4):

$$E_p = \sqrt{\frac{1}{K} \sum_{k=1}^K \{P_c^*(k) - P_t(k)\}^2}. \quad (8)$$

### 2.5. Activation radius for the perturbation

The above distance perturbation strategy should lead to facies proportions that are closer to the target. However, a drawback is that imposing such systematic perturbations can degrade the structures of the training image. This is especially the case at the end of the simulation, when the local proportions are computed on a very small spatial area around the simulated node  $\mathbf{x}$ . In the context of sequential simulation, the perturbations applied to the first simulated nodes will influence the outcome of all nodes simulated later. This cumulative effect means that it is preferable to impose perturbations on the nodes values that are determined at the beginning of the simulation path rather than on the last ones. Distance perturbation applied at the final stage of the simulation may not be effective because most of the spatial structures are already in place and the final proportions can no longer be significantly altered. Forcing a perturbation for the last simulated values may then result in the addition of noise and simulated patterns that do not honor the structures of the training image.

Another problem occurs if the spatial structure of the local proportion map is not entirely compatible with the training image patterns, or if the proportion map is wildly different from the proportions found in the training image. The target proportions can then be matched, but the spatial structure necessarily has to be degraded. In order to keep the correct spatial structure, and at the same time to honor the local proportions, a trade-off has to be found, which would ideally consist in using the target proportion when simulating the large scale structures, and leaving it to the training image only to complete the small-scale structures.

This is accomplished by using the perturbed distance  $d^*$  for the initial part of the simulation, which would correspond to the coarsest multiple-grid levers in the context of other implementations such as SNESIM. The distance  $d$  is then used to simulate the remaining grid nodes. To this end, we introduce an activation radius  $R_{min}$ . If the spatial extension of the data event  $R_d$ , computed as the average distance to the neighbor nodes  $\mathbf{x}_i$ , is larger than  $R_{min}$ , the perturbations are used. Note that  $R_d$  varies because the DS algorithm considers a neighborhood made of a fixed number of the  $n$  closest nodes whose position is not predefined; the neighborhood covers a large area at the beginning of the simulation and reduces while the simulation grid gets more and more informed. For smaller data events, perturbations are deactivated. The functioning of the activation radius is illustrated in Fig. 1. For both local and global proportion constraints, the use of the activation radius ensures that the small-scale structures are not overly affected by the proportion constraints. Moreover, in cases where the alteration of

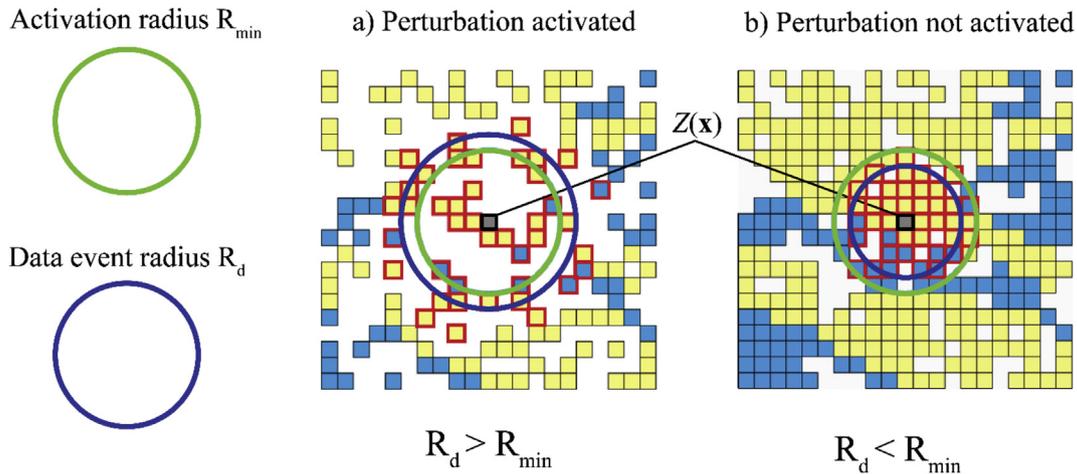


Fig. 1. The principle of the activation radius. a) Distance perturbation activated. b) Distance perturbation not activated.

the structures is inevitable (for example when using a stationary training image and using local proportions to impose a strong non-stationarity),  $R_{min}$  offers a lever for the user to find a trade-off between patterns reproduction and honoring of the proportion constraints.

### 3. Parameterization

#### 3.1. Optimal parameters in the case of global proportions constraints

In our approach, the parameters controlling the outputs are the target proportion weight  $w$  and the activation radius  $R_{min}$ . Intuitively,  $w$   $[0 \infty]$  controls the degree to which the training image patterns can be degraded, and  $R_{min}$   $[0 \infty]$  allows deciding up to which spatial scale this degradation is allowed to take place. While the general principle of these parameters can be qualitatively understood from equation (5) and Fig. 1, a quantitative investigation is needed to understand the interplay of these parameters.

If the target proportions are different from those in the training image, they can only be reached by altering the training image statistics. It is therefore expected that some sort of trade-off needs to be found, and the parameters  $w$  and  $R_{min}$  are the levers at the disposal of the user to allow defining how this trade-off takes place. In essence, we want to degrade the patterns up to the point where they allow reasonably reaching the target proportions (within a tolerance), but we do not want to degrade them more than necessary. As a consequence, a first modeling decision is to determine how close the target proportions should be matched. For example, reaching target proportions that are very different than those of the TI will only be possible by significantly altering the patterns. Then, once an acceptable error on the proportions is defined (this choice will be application-specific), it is possible to find the parameters that yield the highest possible quality realizations (in terms of patterns reproduction) within this constraint. This 2-step parameterization approach was tested on a simple case, based on a training image which was obtained from an aerial photograph of sand dunes in the Gobi desert (Allard et al., 2011). The image, displayed in Fig. 2a, has dimensions of 114 by 114 cells and consists of three facies with global proportions of 0.5149 (facies 0), 0.2311 (facies 1) and 0.2539 (facies 2).

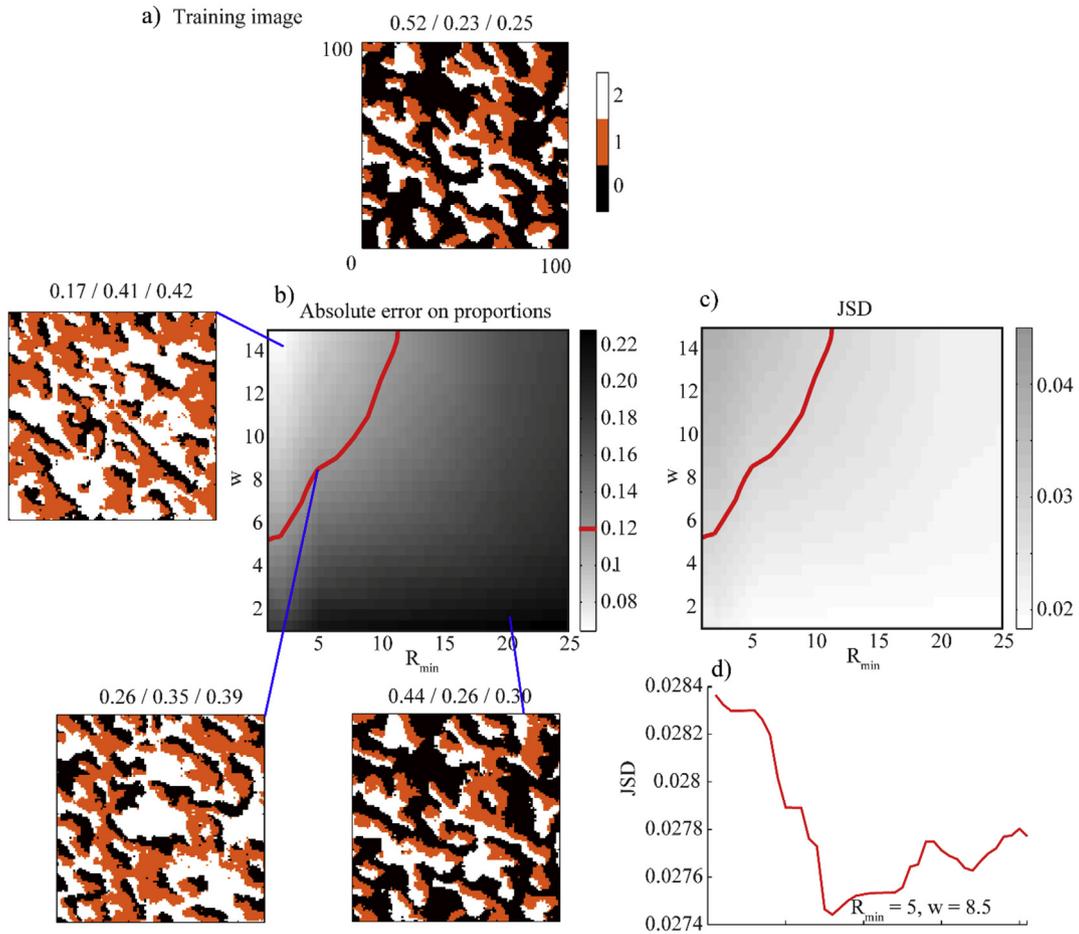
Using this training image, we determine the optimal  $w$  and  $R_{min}$  parameters for three scenarios of target global proportions. In the first scenario (shown in Fig. 2), the target proportions for the three

facies are 0.10 (facies 0), 0.45 (facies 1) and 0.45 (facies 2), denoted thereafter as 0.10/0.45/0.45. These proportions are very different from those of the training image, and therefore significant patterns alteration is expected. It can be seen that in the range of parameters considered, the target proportions cannot be reached exactly in this scenario because they are too far from what is observed in the training image.

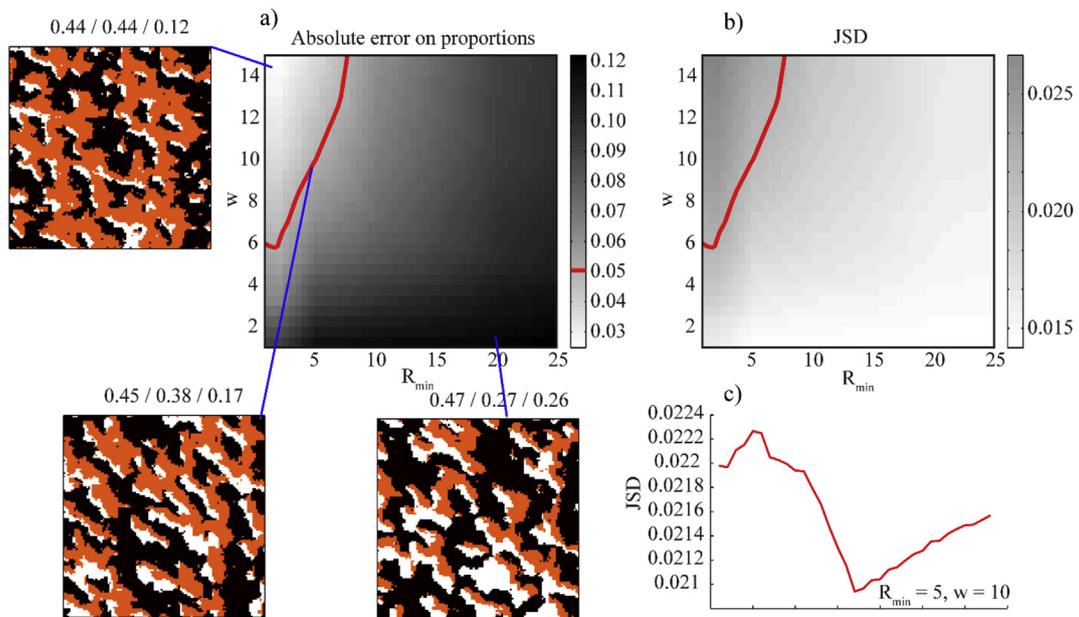
In the second scenario the target global proportions are 0.45/0.45/0.10 (Fig. 3), which is closer to what is found in the training image. Here the target proportions can be approached, but it can be difficult to match them exactly because the constraint is deactivated for patterns smaller than the activation radius.

The third scenario imposes equal proportions for all three facies (0.33/0.33/0.33, Fig. 4), which is even more similar to the proportions in the training image and should be possible to accomplish without significant alteration of the training image patterns. In the case of  $w = 20/R_{min} = 1$ , the proportions are exactly matched although with some addition of noise and degradation of the patterns, the edges of the dunes presenting unrealistic small-scale angular features.

We consider in details the first scenario. Different  $w$  and  $R_{min}$  parameters are used and 100 realizations are generated for each parameters combination. The parameters used for the DS simulations are a neighborhood made of the 25 closest neighbors to the simulated node, and a distance threshold of 0.01, meaning that the algorithm is looking for very close pattern matches. These realizations are used to measure 1) how well the target proportions are honored and 2) how faithfully the training image statistics are reproduced. The error to target proportions is easily computed using a mean absolute error over all categories (Fig. 2b). Quantifying the patterns errors is however more complex, and many different metrics could be adopted, such as variograms (Croft et al., 2013; de Iaco and Maggio, 2011), cumulants (Dimitrakopoulos et al., 2010), connectivity functions (Renard and Allard, 2013) or comparison of patterns (Pérez et al., 2014). Here we adopt multiple-point histograms (Boisvert et al., 2010), which inventory the frequency of each patterns in an image. We compute the multiple-point histogram of each realization and compare it to the multiple-point histogram using a Jensen-Shannon divergence or JSD (Fig. 2c). The higher the JSD, the larger patterns degradation occurs in the realization. In order to compare both large and small patterns, we consider the JSD for patterns of size 3 by 3 nodes averaged over different offsets (1,2,3), in the same fashion as the offsets used in multiple-grids (Tan et al., 2014).



**Fig. 2.** Determination of the optimal parameters in the case where the target proportions (0.10/0.45/0.45) are very different from the training image proportions (0.52/0.23/0.25). The red line represents the isoline of an absolute error on proportions of 0.12 (in b, superimposed in grid c), and the JSD values for the corresponding parameters combinations (in d). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** Determination of the optimal parameters in the case where the target proportions (0.45/0.45/0.10) are moderately different than the training image proportions (0.52/0.23/0.25).

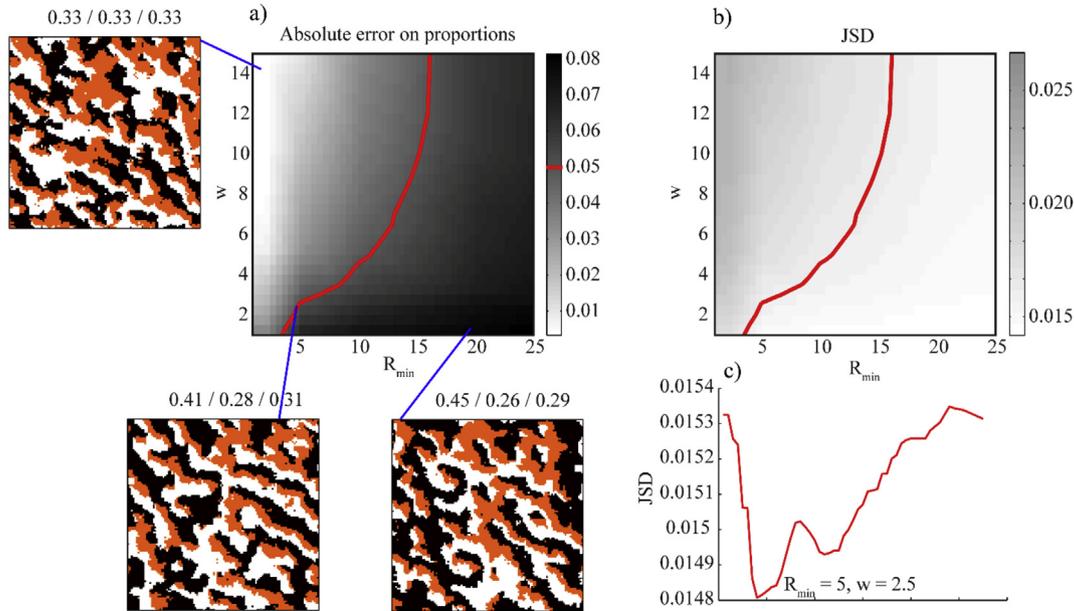


Fig. 4. Determination of the optimal parameters in the case where the target proportions (0.33/0.33/0.33) are relatively similar to the training image proportions (0.52/0.23/0.25).

On Fig. 2 it is clear that the parameters combinations yielding low JSDs, result in poor proportion control, whereas reaching the target proportions involves significant degradation of the patterns (i.e. high JSD). The proportion error and the JSD are therefore inversely proportional. This trade-off is also present when other target proportions are considered (see the other scenarios in Figs. 3 and 4), indicating that the trade-off between proportions constraints and patterns constraints is a general characteristic.

To select the parameter values, we adopt the point of view that in practice, the target proportions are based on some available data, which also have uncertainty bounds. It therefore makes sense to start deciding to what level the target proportions have to be matched, and within that constraint it is then possible to optimize the patterns reproduction (quantified here using the JSD). This is shown in figure Fig. 2b, where the red line represents the isoline of an absolute error on proportions of 0.12 (a relatively large error, but which should be considered in the context of target proportions being very different from the ones in the training image). All realizations obtained with parameters along this line will have equal proportion error, however they may reproduce the training image patterns to different degrees. This is shown in figure Fig. 2c, where the same 0.12 proportion error isoline is superimposed on the JSD values grid (note that in figure Fig. 2c the line is not an isoline in the JSD, but the same isoline as in figure Fig. 2b that is superimposed). While the JSD values seem relatively uniform along this isoline, they do present small variations, as shown in figure Fig. 2d that displays the JSD values along this red line. A point of minimum JSD can be identified, corresponding to the optimal parameters  $R_{min} = 5$  and  $w = 8.5$ . It is however observed that the variations of JSD along the isoline are very small, indicating that equivalent trade-offs between proportions constraints and patterns constraints can be reached with different sets of parameters.

Similar trade-offs can be established to define the optimal parameters for different target proportions, such as 0.45/0.45/0.10 (Fig 3) and 0.33/0.33/0.33 (Fig. 4). For these two cases, the target proportions are not radically different from those in the training image, and therefore we have set the proportion error to 0.05. It is found that in all cases, the optimal  $R_{min}$  value is 5, which is proportional to the size of the structures to reproduce. This is a

general observation that is not particular to this case study and has also been observed with the examples presented in the next sections, therefore providing an easy rule of thumb for an approximate  $R_{min}$  value when a complete sensitivity analysis is not practical.

### 3.2. Parameterization when using local proportions constraints

The use of local proportion constraints is illustrated using the same training image as in Fig. 2. The same parameters for the DS are used as previously (neighborhood made of 25 nodes, distance threshold of 0.01). We use this stationary training image to generate non-stationary unconditional realizations on a larger domain of 100 by 600 cells where the proportions of the facies vary from one side of the domain to the other. Fig. 5a–c shows the model setting: we impose a target proportion for facies 0 varying horizontally between 0 and 1, the complementary being the two other facies having equal target proportion.

When local proportions constraints are used, similar patterns degradations will take place as in the case of global constraints, except that the degree of the degradation will vary spatially. It is however impractical to define spatially-variable  $w$  and  $R_{min}$  parameters. We therefore base ourselves on the optimal  $R_{min} = 5$  parameter inferred for the case of global proportions constraints, and perform a sensitivity study on  $w$  to investigate the trade-off between proportions constraints and patterns constraints. In our example this trade-off will be problematic near the right and left edges of the image where the target proportion for facies 0 is respectively 1 and 0, whereas the training image is stationary with a global proportion of 0.5149. Such a drastic change in proportions will necessarily require either a significant alteration in the training image patterns, or a non-respect of the local proportion constraints.

We investigate the effect of different combinations of  $R_{min}$  and  $w$ . Fig. 6 shows individual realizations, confirming that the trend in the target proportion map is globally respected, although less so for low  $w$  values (Fig. 6a) where the patterns cannot depart significantly from those of the training image. In Fig. 6d, matching the target proportions results in structures composed of facies 1 and 2 of reduced size on the right side of the domain. This is caused by the target proportions that vary with a constant gradient. As a result

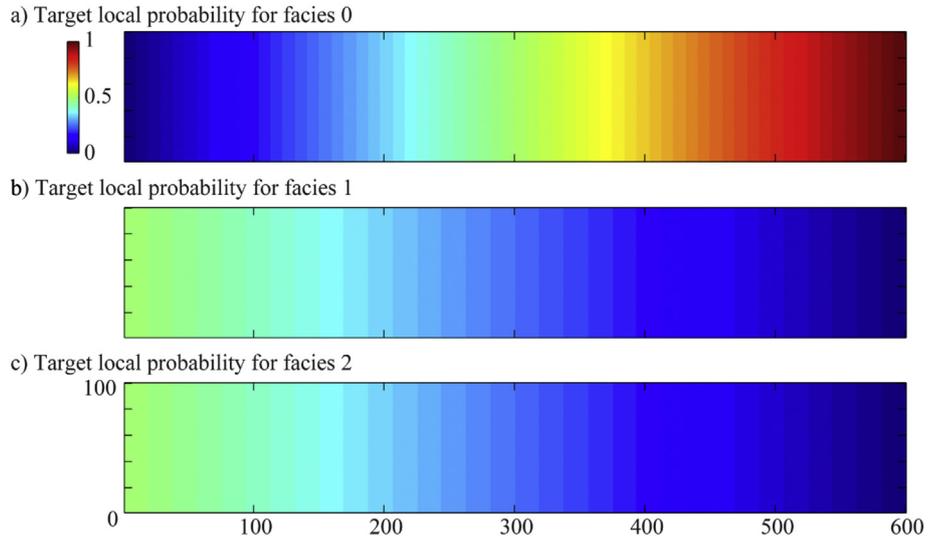


Fig. 5. Target local proportion maps.

smaller structures are produced, allowing the mean proportion to vary more smoothly. Similarly, large structures of increased connectivity appear on the left side of the domain, showing that modifying the proportions is associated with morphological changes. Fig. 6e is an extreme case illustrating that it is possible to closely follow the target proportions, but with the necessary consequence that the patterns are globally degraded. The result is small-scale patterns degradation that correspond to proportions corrections carried out at the end of the random path.

In the context of the DS algorithm, altering the patterns means that more scanning of the training has to be done. This translates in an increase in CPU cost. In the examples of Fig. 6, 11.70 s in average are needed for Fig. 6a, to be compared with 9.86 s if no probability constraints are considered. When the proportions constraints are stronger, the computational times increase, with 16.85 s for Fig. 6c and 30.93 s for Fig. 6e.

The overall respect of the proportions constraints for the same 5 parameters sets is shown in Fig. 7, where the average proportion along the X axis is displayed. The shaded areas correspond to  $\pm 2$  standard deviations around the mean proportion, computed on 100 realizations for each case, and the dashed lines represent the target proportions. Fig. 7a shows that the target proportions cannot be honored near the edges of the domain. On the other hand the cases in Fig. 7c to e fit the target proportions to a large degree, at the cost of degraded patterns.

#### 4. Application examples

##### 4.1. Example with land use categories

While geological modeling is a traditional application domain of multiple-point geostatistics, the possibility to impose variations of

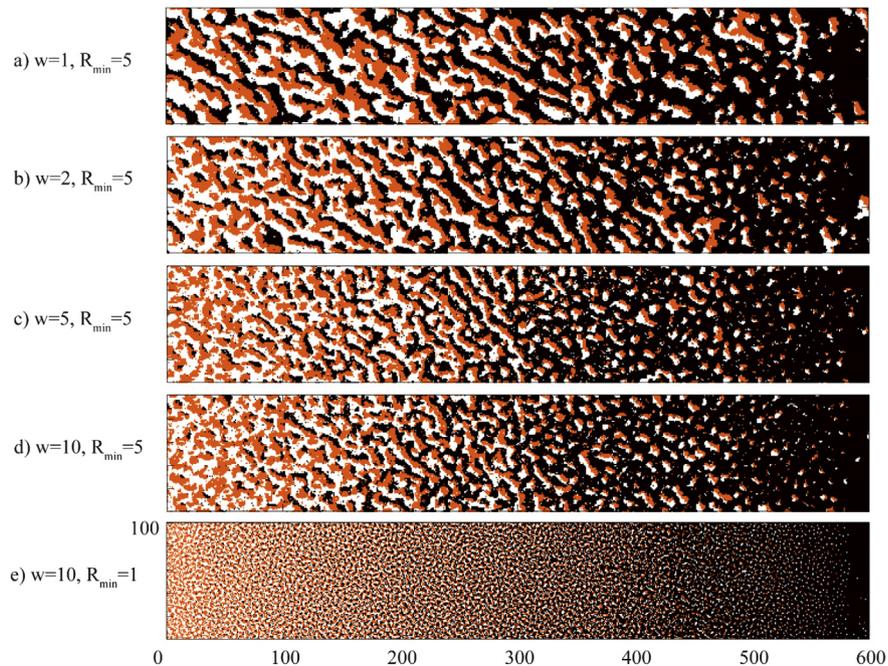
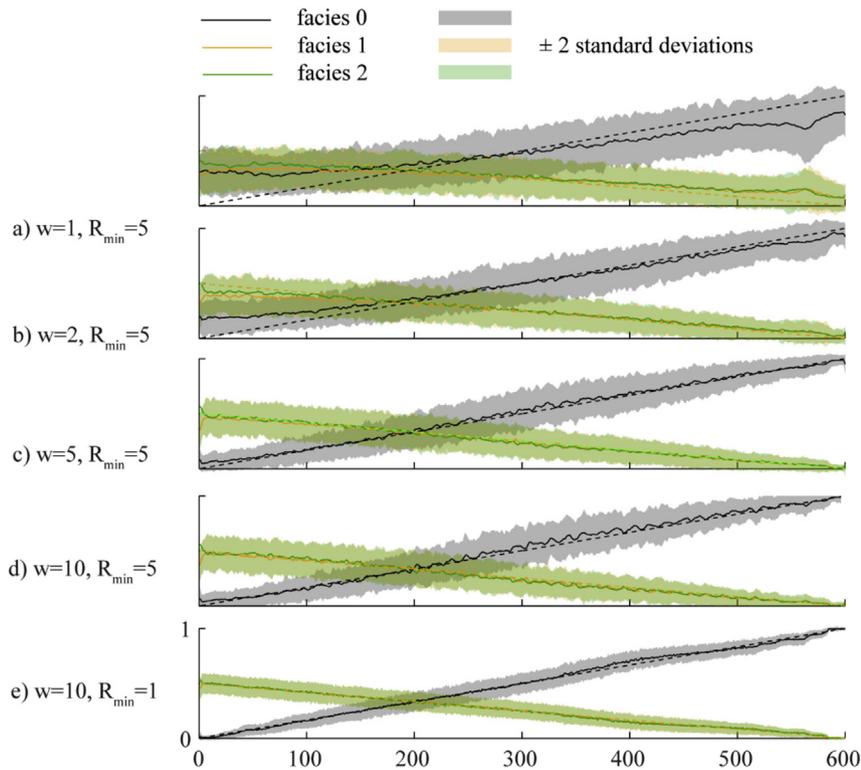


Fig. 6. Sensitivity analysis of parameters  $w$  and  $R_{min}$  for local probability constraints.



**Fig. 7.** Comparison of target and simulated proportions. Solid lines represent mean facies proportions. Dashed lines denote target proportions. Shaded areas represent the area corresponding to  $\pm 2$  standard deviations around the mean proportion, computed on 100 realizations. Note that the shaded area of facies 1 is mostly overlapping the shaded area of facies 2.

categories proportions opens possibilities in different fields of study where it is important to account for strong non-stationarity. To illustrate the use of local proportions control in a realistic case that is not a classical application domain of MPS, we consider the distribution of land use classes throughout West Africa and how it could be altered by the end of the 21st century. We base ourselves on land use data derived from remote sensing (Global land cover GLCNMO). The present-day distribution of land use classes is displayed in Fig. 8a, with a simplified classification consisting of only 6 land use classes.

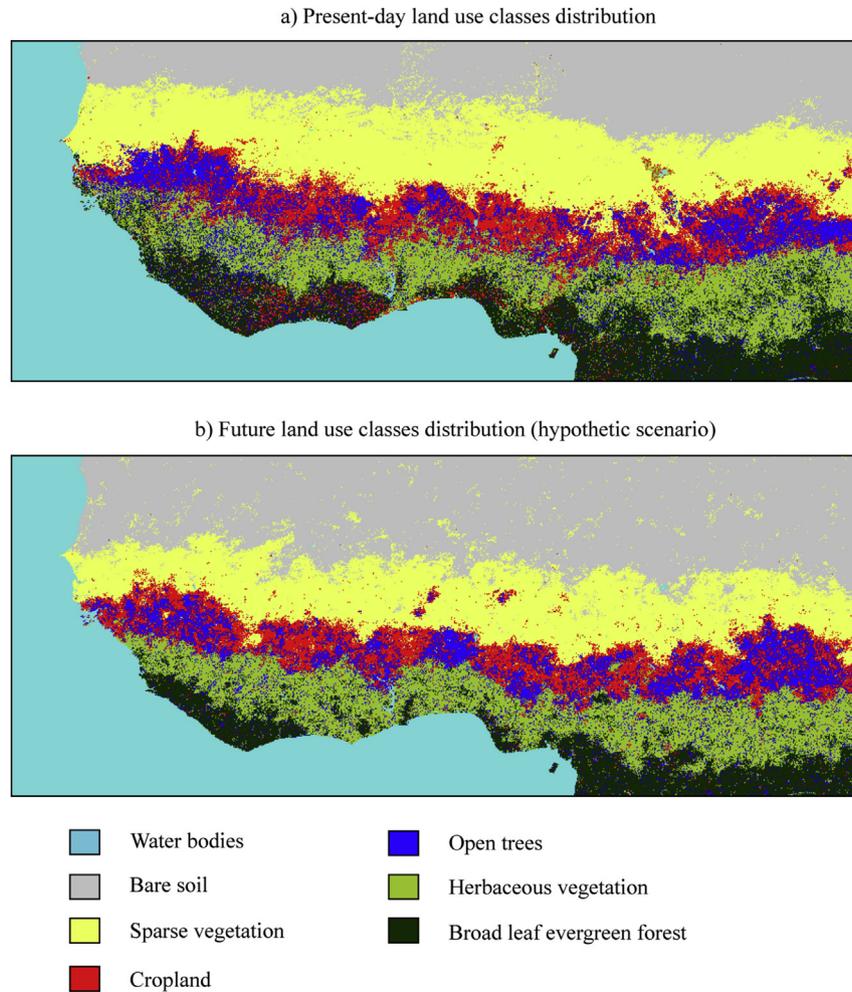
In the literature, shifts in climatic zones of approximately 500–600 m/year for the second half of the 20th century have been reported in the Sahel area (Gonzalez, 2001). We place ourselves in a hypothetical scenario where these shifts are accelerated in the 21st century, with faster shifts in the arid zone (north of the studied area) than in the equatorial zone. Based on this, we assume different shifts for each land use class resulting, by the end of the 21st century, in southwards translations between 40 and 200 km depending on the areas. Note that these values are arbitrarily chosen to test our method and do not reflect an actual prediction of future land use distribution. Our goal in this example is to use our local proportion approach to depict what the future land use would look like in such a hypothetical scenario. The domain size is 901 by 361 pixels, each pixel representing an area of 4 km by 4 km.

In a first step, we compute the local proportions of each category for present-day land use by using a moving average on a window of size 30 pixels, corresponding to 480 km<sup>2</sup> (Fig. 9a). The moving average is applied separately on the indicator variable that corresponds to the occurrence of each category, resulting in 6 separate local proportion maps. In order to represent the proportions of categories in the future, we shift the proportion map for each category southwards by a certain distance (see Fig. 9 for the amount

of shift for each category). On each proportion map, the southwards shift causes a horizontal zone to be unknown. The proportion values are extrapolated in those areas using nearest neighbor interpolation, which is adequate in this case since the proportion of each category is expected to be relatively uniform in this upper horizontal zone. Since a different shift is applied for each category, the proportions need to be renormalized to sum up to one at each location. The resulting proportion maps are displayed in Fig. 9b. These are used as target proportions for simulating the future land use. Fig. 8a is used as training image for the application of our distance perturbation method, under the assumption that the spatial patterns will remain similar in the future, and in particular that the spatial relationship between the different categories will not change drastically.

The window moving average used to obtain the present-day proportions is based on a 30 pixels window. This ensures that the local proportions are relatively constant over such a radius, reflecting the gradient of the local proportion map. In consequence, we use a value of  $R_{min} = 30$  pixels to avoid patterns degradations in the spatial structures. We use a relatively low value of  $w = 1$  to preserve the spatial patterns of the training image. The parameters of the DS algorithm are the same as in the previous synthetic example (neighborhood made of 25 nodes, distance threshold of 0.01). Fig. 8b shows one resulting realization. The specific spatial patterns for each category are globally preserved in the simulation. One can notice in particular the following spatial features:

- Bare soil and sparse vegetation are relatively uniform within each category, although they are intertwined at their interface.
- Cropland and Open trees are intermingled, with fringes of cropland generally surrounding relatively compact open trees areas.



**Fig. 8.** Case study of land use changes in Western Africa. a) present-day land use (Global land cover GLCNMO, <http://www.iscgm.org>). b) One realization of future land use according to a hypothetical scenario.

- Open trees pixels are speckled in the zones dominated by herbaceous vegetation and broad leaf forest.
- Numerous “Islands” of broad leaf forest occur in the herbaceous vegetation zone.
- Almost no “Islands” of herbaceous vegetation occur in the zone covered by broad leaf forest.

Fig. 9c shows the probability of occurrence of each category calculated over 50 realizations. It was computed by applying to each realization the same procedure as was used to determine the local proportions in the training image of Fig. 8a: a moving average on a window of size 30 pixels for the indicator variable corresponding to each category. Fig. 9c represents the mean local proportion over all 50 realizations, and corresponds well to the target proportions of Fig. 9b.

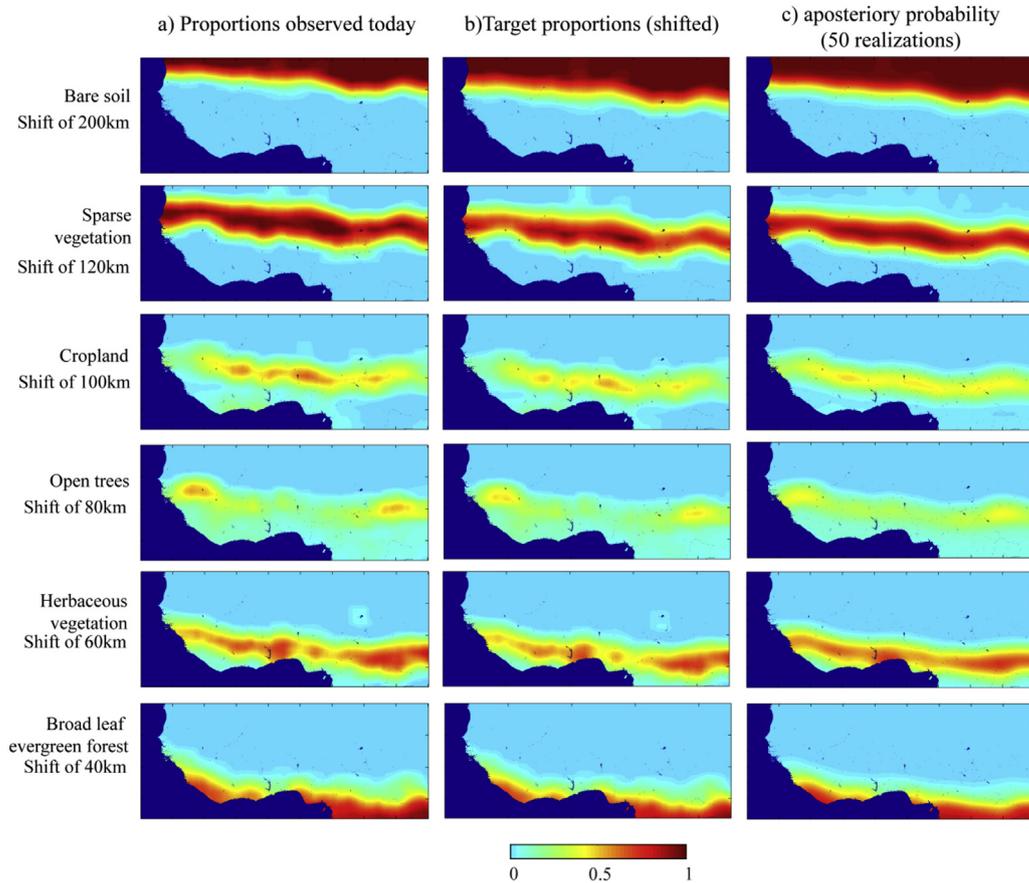
#### 4.2. 3D example

The applicability of our method in 3D is shown on a pore-scale modeling example. The training image represents a synthetic sandstone sample where partial clogging occurs by precipitation in a part of the pore volume. Here the training image is itself a construct based on a simulation using the method of elementary training images and transform-invariant distances (Mariethoz and Kelly, 2011). The elementary training image is shown in Fig. 10a

( $50 \times 50 \times 50$  voxels). It consists of a regular array of cubes (in blue) that will be used to represent the sand grains, interspaced with green separations (the space between pores). Along certain planes the pore space is taken up by a third facies shown in red, which will represent the precipitation clogging. Using this elementary training image and transform-invariant distances with random rotations in the range  $[+90^\circ -90^\circ]$ , we obtain the image displayed in Fig. 10b ( $100 \times 100 \times 100$  voxels). The random rotations cause the cubic grains in Fig. 10a to produce spheroid-type shapes in Fig. 10b, the pore space and clogging phase taking up the remaining volume. For the purpose of this test, we assume Fig. 10b to be representative of a sandstone where the pores have been partially clogged. For graphical reasons on Fig. 10b the sand grains are represented in transparent (empty pore space and clogging are left in green and red colors respectively). The proportions in Fig. 10b for the categories grains/empty pore/clogging are respectively 71%, 21% and 8%.

Modeling this type of medium is important for applications ranging from reactive transport in porous media where the precipitation alters the hydraulic properties of a rock formation, to the modeling of hydrothermal ore deposits where the precipitated deposits have a commercial value and are mined.

We set up a case where reactive transport caused differentiated amounts of precipitation in a portion of sandstone. The setting consists in a domain of  $100 \times 100 \times 200$  voxels, where the target

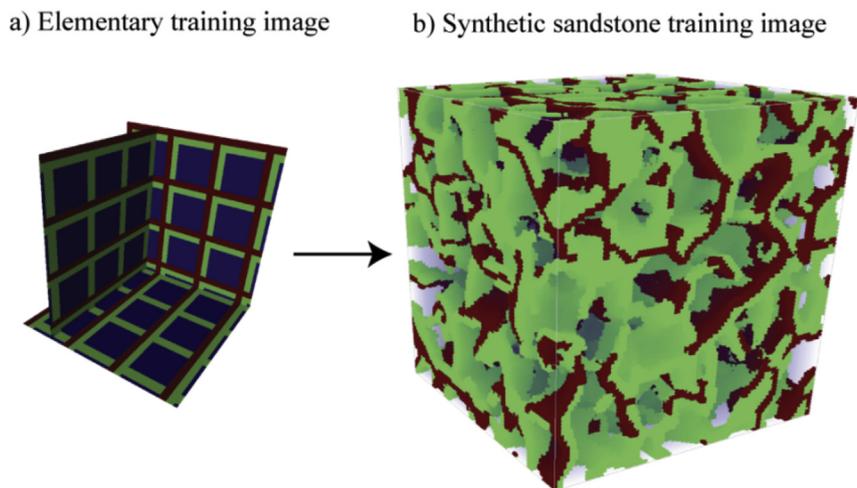


**Fig. 9.** Local proportion maps for Western Africa land use. a) Present-day distribution of land use categories. b) Target proportions with different shifts for each category. c) Local proportions based on 50 realizations.

proportion of sand grains is constant and stationary at a value of 61%, the remaining volume being shared between pore space and clogging that has precipitated in this pore space. The bottom part of the volume considered has pores that are mostly clogged by precipitation, while the top part has pores with almost no clogging. This situation is represented with the target proportions

represented in Fig. 11: the target proportion for the sand grains is uniform and the remaining fraction varies between 5% pores/34% clogging and 34% pores/5% clogging.

Since the target proportions are very different than in the training image we use a relatively high value of  $w = 20$ , and  $R_{min} = 5$  which is comparable to the size of the structures. Regarding the DS



**Fig. 10.** Clogged sandstone example. a) Elementary training image used for the construction of b) (grains in blue, pores in green, clogging in red). b) Training image representing sandstone with partial clogging occurring within the pore space (grains represented in transparent). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

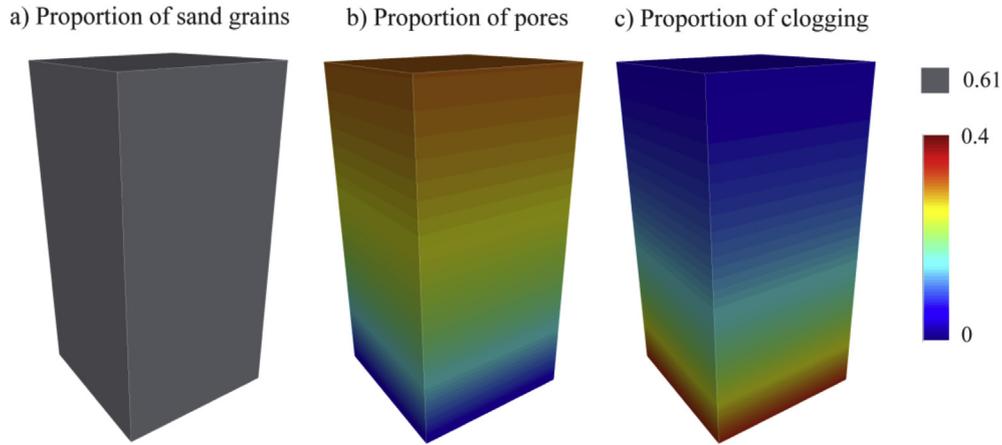


Fig. 11. Target proportions for the three facies considered.

parameters, the distance threshold is 0.01 as in the previous examples, but this time the number of neighbors is set to 50 to be able to represent the complex 3D patterns. Fig. 12a shows one resulting realization and Fig. 12b–d shows the variation in the proportion of the different categories, computed from 75 realizations. The proportion of sand grains is relatively uniform, despite local variations due to the limited number of realizations used. The proportions of open pores and clogging are inversely proportional to each other, as in the target of Fig. 11.

However when looked at closely, the target proportions are not well respected, as shown in Fig. 13. This is especially the case towards the lower part of the modeled area, where it seems very difficult to increase the proportion of clogging. We explain this by the particular structure of the training image, where the clogging facies (red) has the same spatial structure as the pore scale facies (green) consisting of sections, but for clogging these sections only occur in two directions of space (see Fig. 10a). It is therefore very difficult to have more clogging than pores, which is what is observed on Fig. 13. Moreover, in this case the target proportions impose to inflate the proportion of a facies that is very infrequent in the training image (8%), which can only be done by completely altering the spatial patterns. This could be achieved with lower values of  $R_{min}$  and high  $w$ , but then the patterns alterations would be so large that it would be preferable to use a different training image to have an explicit control on the generated patterns.

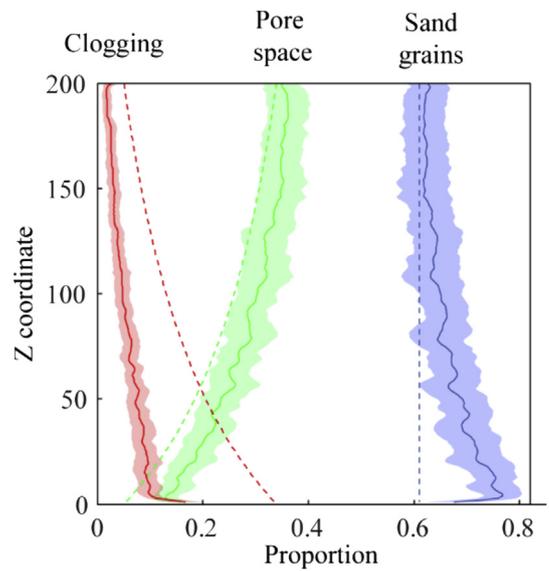


Fig. 13. Comparison of target and simulated proportions. Solid lines represent mean facies proportions. Dashed lines denote target proportions. Shaded areas represent the area corresponding to  $\pm 2$  standard deviations around the mean proportion, computed on 75 realizations.

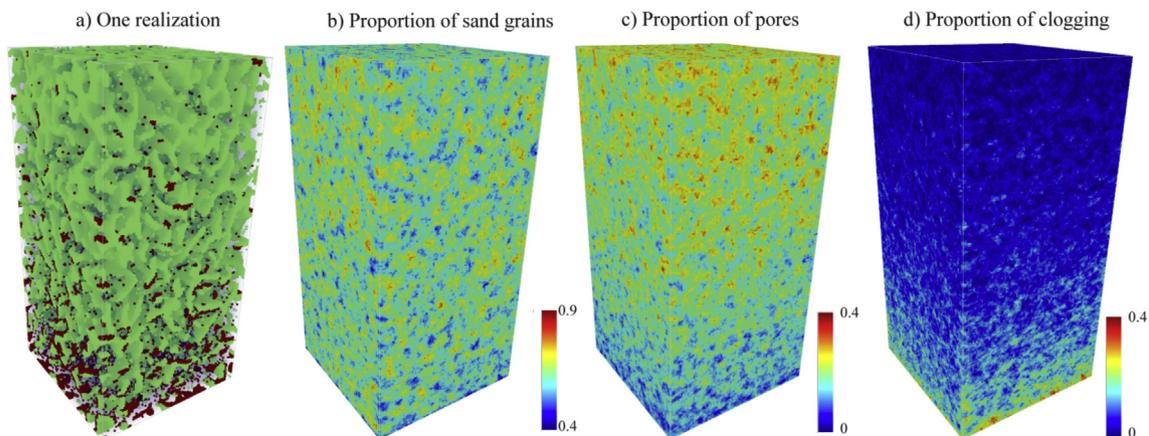


Fig. 12. a) One realization of sandstone with differential clogging. b–d) proportion of each facies, computed over 75 realizations.

## 5. Discussion and conclusion

A distance perturbation method is proposed to impose local proportions of categories in the context of Direct Sampling multiple-point simulation. Thus far, such a control on proportions was only possible for methods that explicitly formulate a probability distribution for the simulated values such as for example SNESIM (Strebelle, 2002). Distance-based methods, which are often more computationally efficient and more flexible, do not offer for the moment proportion control, and this is their main limitation. In this paper we presented an approach that lift this limitation. It is illustrated with the DS algorithm. The rationale of the approach is to penalize, for each simulated node, the locations in the training image that would result in a proportion further away from the target. For each candidate value in the training image, the current proportions in the simulation are computed accounting for the already informed nodes in the neighborhood of the simulated node and for this candidate value. If the new candidate value influences the proportions in a direction that is contrary to the target, a penalty proportional to a given weight is attributed to the candidate.

The method is tested in a variety of cases borrowed from domains ranging from remote sensing to pore-scale modeling, demonstrating that our approach is applicable beyond the traditional field of geological modeling of lithofacies. It is generally able to produce realizations that are constrained either to a local proportion map or to given global proportions.

In case of non-stationarity in the proportions, there is only a very small part of the simulation domain where the target proportions correspond to what is observed in the training image. At all other locations, patterns will necessarily depart from those in the training image in order to match the target proportions. It is a general observation that when the local proportions significantly differ from the proportions in the training image, honoring the target proportions comes at the price of altering the spatial structures. The parameters  $w$  and  $R_{min}$  offer a way for the user to settle a trade-off between reproduction of the training image patterns and respect of the target proportions.  $w$  determines the strength of the distance perturbation relative to the distance on patterns and  $R_{min}$  provides a minimum spatial scale under which no distance perturbation is used, ensuring that the small-scale structures are preserved.

Related to this discussion, an important aspect is that the distance perturbation method allows generating patterns that were not initially present in the training image, and to control these new patterns through the parameters  $w$  and  $R_{min}$ . This can be interesting in applications of uncertainty modeling where one wants to explore models that depart from the strict pool of patterns given in the training image.

In the approach presented, we use the same neighborhood to compute local proportions and for determining a patterns distance in the training image. This choice is motivated by the simplified parameterization (only a single neighborhood needs to be chosen) and the computational efficiency. However it is straightforward to have a different neighborhood for the computation of local proportions. Tests done on a separate implementation have showed that it is possible to compute proportions on a neighborhood of a fixed radius, therefore defining a constant support for the local proportions. This could be useful in applications where there is a physical relationship between the magnitude of the simulated values and the volume on which these values are averaged, such as for example in the case of simulating permeability values. Specific proportions neighborhoods could also be used if there are sharp contrasts in the proportion map, which would require to be captured by smaller neighborhoods.

The computational cost of the proportion control is not prohibitive, with only a minor increase of the order of 15% in cases where the target proportions do not require too much alteration of the training patterns. However the computational cost can be much higher if the proportions constraints require generating new patterns, with up to a tripling of the computational time compared to the absence of proportions control.

The present paper only deals with categorical variables, where target local or global proportions are specified for every category. Following the same approach, one can define classes gathering several categories and target proportions for each class. The concept could be extended to continuous variables as well by defining classes and the corresponding proportions. In this case the error between two distributions  $E_p$  (used in equation (4)) could be defined as an error between the current and the target histograms. Alternatively, the computation of the error term could be simplified by only considering the error between moments of the current and target histograms, such as for example the mean and the standard deviation.

## Acknowledgments

Funding from the Australian Research Council, the National Water Commission, and the Groundwater Education Investment Fund is acknowledged. This work was also supported by Total.

## References

- Allard, D., Comunian, A., Renard, P., 2012. Probability aggregation methods in geoscience. *Math. Geosci.* 44 (5), 545–581.
- Allard, D., D'Or, D., Froidevaux, R., 2011. An efficient maximum entropy approach for categorical variable prediction. *Eur. J. Soil Sci.* 62 (3), 381–393.
- Arpat, B., Caers, J., 2007. Conditional simulations with patterns. *Math. Geol.* 39 (2), 177–203.
- Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. *Environ. Model. Softw.* 40 (0), 1–20.
- Boisvert, J.B., Pyrcz, M.J., Deutsch, C.V., 2010. Multiple point metrics to assess categorical variable models. *Nat. Resour. Res.* 19 (3), 165–175.
- Boucher, A., 2009. Sub-pixel mapping of coarse satellite remote sensing images with stochastic simulations from training images. *Math. Geosci.* 41 (3), 265–290.
- Caers, J., Hoffman, T., 2006. The probability perturbation method: a new look at Bayesian inverse modeling. *Math. Geol.* 38 (1), 81–100.
- Chugunova, T., Hu, L., 2008. Multiple-Point simulations constrained by continuous auxiliary data. *Math. Geosci.* 40 (2), 133–146.
- Croft, H., Anderson, K., Brazier, R.E., Kuhn, N.J., 2013. Modeling fine-scale soil surface structure using geostatistics. *Water Resour. Res.* 49 (4), 1858–1870.
- de Iaco, S., Maggio, S., 2011. Validation techniques for geological patterns simulations based on variogram and multiple-point statistics. *Math. Geosci.* 43 (4), 483–500.
- de Vries, L., Carrera, J., Falivene, O., Gratacos, O., Slooten, L., 2009. Application of multiple point geostatistics to non-stationary images. *Math. Geosci.* 41 (1), 29–42.
- Deutsch, C.V., Journel, A.G., 1998. *GSlib: Geostatistical Software Library and User's Guide*, second ed.
- 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 (1), 65–99.
- El Ouassini, A., Saucier, A., Marcotte, D., Favis, B.D., 2008. A patchwork approach to stochastic simulation: a route towards the analysis of morphology in multi-phase systems. *Chaos Solit. Fractals* 36 (2), 418–436.
- Faucher, C., Saucier, A., Marcotte, D., 2014. Corrective pattern-matching simulation with controlled local-mean histogram. *Stoch. Environ. Res. Risk Assess.* 1–24.
- Ge, Y., Bai, H., 2011. Multiple-point simulation-based method for extraction of objects with spatial structure from remotely sensed imagery. *Int. J. Remote Sens.* 32 (8), 2311–2335.
- Gonzalez, P., 2001. Desertification and a shift of forest species in the West African Sahel. *Clim. Res.* 17 (2 Special 8), 217–228.
- Goodfellow, R., Albor Consuegra, F., Dimitrakopoulos, R., Lloyd, T., 2012. Quantifying Multi-element and Volumetric Uncertainty, Coleman McCreeley Deposit. In: *Computers and Geosciences*, vol. 42, pp. 71–78. Ontario, Canada.
- Honarkhah, M., Caers, J., 2010. Stochastic simulation of patterns using distance-based pattern modeling. *Math. Geosci.* 42 (5), 487–517.

- Huysmans, M., Dassargues, A., 2012. Modeling the effect of clay drapes on pumping test response in a cross-bedded aquifer using multiple-point geostatistics. *J. Hydrol.* 450–451, 159–167.
- Jha, S.K., Mariethoz, G., Evans, J.P., McCabe, M.F., 2013. Demonstration of a geostatistical approach to physically consistent downscaling of climate modeling simulations. *Water Resour. Res.* 49 (1), 245–259.
- Koch, J., He, X., Jensen, K.H., Refsgaard, J.C., 2013. Challenges in conditioning a stochastic geological model of a heterogeneous glacial aquifer to a comprehensive soft dataset. *Hydrol. Earth Syst. Sci. Discuss.* 10 (12), 15219–15262.
- Krishnan, S., 2008. The tau model for data redundancy and information combination in earth sciences: theory and application. *Math. Geosci.* 40 (6), 705–727.
- Liu, Y., 2006. Using the Snesim program for multiple-point statistical simulation. *Comput. Geosci.* 23, 1544–1563.
- Mahmud, K., Mariethoz, G., Caers, J., Tahmasebi, P., Baker, A., 2014. Simulation of earth textures by conditional image quilting. *Water Resour. Res.* 50 (4), 3088–3107.
- Mariethoz, G., Kelly, B.F.J., 2011. Modeling complex geological structures with elementary training images and transform-invariant distances. *Water Resour. Res.* 47 (7).
- Mariethoz, G., Renard, P., Straubhaar, J., 2010. The direct sampling method to perform multiple-point geostatistical simulations. *Water Resour. Res.* 46 (11).
- Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., Renard, P., 2013. A practical guide to performing multiple-point statistical simulations with the direct Sampling algorithm. *Comput. Geosci.* 52, 307–324.
- Meerschman, E., Van Meirvenne, M., Mariethoz, G., Islam, M.M., De Smedt, P., Van De Vijver, E., Saey, T., 2014. Using bivariate multiple-point statistics and proximal soil sensor data to map fossil ice-wedge polygons. *Geoderma* 213, 571–577.
- Okabe, H., Blunt, M., 2007. Pore space reconstruction of vuggy carbonates using microtomography and multiple-point statistics. *Water Resour. Res.* 43, W12S02.
- Oriani, F., Straubhaar, J., Renard, P., Mariethoz, G., 2014. Simulation of rainfall time-series from different climatic regions using the direct sampling technique. *Hydrol. Earth Syst. Sci. Discuss.* 11 (3), 3213–3247.
- Pérez, C., Mariethoz, G., Ortiz, J.M., 2014. Verifying the high-order consistency of training images with data for multiple-point geostatistics. *Comput. Geosci.* 70 (0), 190–205.
- Pham, T.D., 2012. Supervised restoration of degraded medical images using multiple-point geostatistics. *Comput. Methods Programs Biomed.* 106 (3), 201–209.
- Renard, P., Allard, D., 2013. Connectivity metrics for subsurface flow and transport. *Adv. Water Resour.* 51, 168–196.
- Rezaee, H., Asghari, O., Koneshloo, M., Ortiz, J., 2014. Multiple-point geostatistical simulation of dykes: application at Sungun porphyry copper system, Iran. *Stoch. Environ. Res. Risk Assess.* 1–15.
- Ronayne, M., Gorelick, S., Caers, J., 2008. Identifying discrete geologic structures that produce anomalous hydraulic response: an inverse modeling approach. *Water Resour. Res.* 44, W08426.
- Stisen, S., McCabe, M.F., Refsgaard, J.C., Lerer, S., Butts, M.B., 2011. Model parameter analysis using remotely sensed pattern information in a multi-constraint framework. *J. Hydrol.* 409 (1–2), 337–349.
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., Besson, O., 2011. An improved parallel multiple-point algorithm using a list approach. *Math. Geosci.* 43 (3), 305–328.
- Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* 34 (1), 1–22.
- Tahmasebi, P., Hezarkhani, A., Sahimi, M., 2012. Multiple-point geostatistical modeling based on the cross-correlation functions. *Comput. Geosci.* (42), 18–27.
- Tahmasebi, P., Sahimi, M., 2013. Cross-correlation function for accurate reconstruction of heterogeneous media. *Phys. Rev. Lett.* 110, 078002.
- Tan, X., Tahmasebi, P., Caers, J., 2014. Comparing training-image based algorithms using an analysis of distance. *Math. Geosci.* 46 (2), 149–169.
- Tang, Y., Atkinson, P.M., Wardrop, N.A., Zhang, J., 2013. Multiple-point geostatistical simulation for post-processing a remotely sensed land cover classification. *Spat. Stat.* (5), 69–84.
- Vannameteer, E., Babel, L.V., Hendriks, M.R., Schuur, J., de Jong, S.M., Bierkens, M.F.P., Karssenberg, D., 2014. Semi-automated mapping of landforms using multiple point geostatistics. *Geomorphology* 221, 298–319.
- Wojcik, R., McLaughlin, D., Konings, A., Entekhabi, D., 2009. Conditioning stochastic rainfall replicates on remote sensing data. *IEEE Trans. Geosci. Remote Sens.* 47 (8), 2436–2449.
- Yin, Y., 2013. A new stochastic modeling of 3-D mud drapes inside point bar sands in meandering river deposits. *Nat. Resour. Res.* 22 (4), 311–320.
- Zhang, T., Bombarde, S., Strebelle, S., Oatney, E., 2006a. 3D porosity modeling of a carbonate reservoir using continuous multiple-point statistics simulation. *SPE J.* 11 (3), 375–379.
- Zhang, T., Switzer, P., Journel, A., 2006b. Filter-based classification of training image patterns for spatial simulation. *Math. Geol.* 38 (1), 63–80.