Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Locality-based 3-D multiple-point statistics reconstruction using 2-D geological cross-sections

Qiyu Chen [1,2,3], Gregoire Mariethoz [2], Gang Liu [1,3,*], Alessandro Comunian [4], Xiaogang Ma [5]

[1] School of Computer Science, China University of Geosciences, Wuhan 430074, China
[2] Institute of Earth Surface Dynamics, University of Lausanne, 1015 Lausanne, Switzerland
[3] Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China
[4] Dipartimento di Scienze della Terra "A.Desio", Università degli Studi di Milano, Milan, Italy
[5] Department of Computer Science, University of Idaho, 875 Perimeter Drive MS 1010, Moscow, ID 83844-1010, USA

* *Correspondence to:* liugang67@163.com (G. Liu)

Abstract: Multiple-point statistics (MPS) has shown promise in representing complicated subsurface structures. For a practical three-dimensional (3-D) application, however, one of the critical issues the difficulty to obtain a credible 3-D training image. However, bidimensional (2-D) training images are often available because established workflows exist to derive 2-D sections from scattered boreholes and/or other samples. In this work, we propose a locality-based MPS approach to reconstruct 3-D geological models on the basis of such 2-D cross-sections, making 3-D training images unnecessary. Only several local training sub-sections closer to the central uninformed node are used in the MPS simulation. The main advantages of this partitioned search strategy are the high computational efficiency and a relaxation of the stationarity assumption. We embed this strategy into a standard MPS framework. Two probability aggregation formulas and their combinations are used to assemble the probability density functions (pdfs) from different sub-sections. Moreover, a novel strategy is adopted to capture more stable pdfs, where the distances between patterns and flexible neighborhoods are integrated on several multiple grids. A series of sensitivity analyses demonstrate the stability of the proposed approach. Several hydrogeological 3-D application examples illustrate the applicability of our approach in reproducing complex geological features. The results, in comparison with previous MPS methods, show better performance in portraying anisotropy characteristics and in CPU cost.

Keywords: Three-dimensional reconstruction, Multiple-point statistics, Locality, Cross-sections, Non-stationarity, Probability aggregation

1. Introduction

3-D characterization of geological architectures plays a crucial role in the quantification of subsurface water, oil and ore resources (*Chen et al.*, 2017; *Foged et al.*, 2014; *Hoffman and Caers*, 2007; *Jackson et al.*, 2015; *Kessler et al.*, 2013; *Raiber et al.*, 2012; *Wambeke and Benndorf*, 2016). Heterogeneity and connectivity of sedimentary reservoirs exert controls on

underground fluid transport (*Gaud et al.*, 2004; *Renard and Allard*, 2013; *Weissmann et al.*, 1999) which is vital to quantify and forecast the formation and distribution of subsurface resources. For a practical 3-D application, however, these attributes are notoriously difficult to characterize and model since the informed data we can acquire are very sparse. A variety of two-point geostatistical approaches (*Goovaerts*, 1998; *Journel*, 1993; *Pyrcz and Deutsch*, 2014; *Ritzi*, 2000) have been employed to reproduce 3-D models of subsurface sedimentary structures and facies distribution. Unfortunately, two-point geostatistics cannot capture high-order statistics and hence it cannot properly reproduce anisotropic features and connectivity patterns (*Heinz et al.*, 2003; *Klise et al.*, 2009; *Knudby and Carrera*, 2005; *Lee et al.*, 2007; *Phelps and Boucher*, 2009; *Vassena et al.*, 2010). Object-based methods allow a realistic reconstruction of the heterogeneous lithofacies (*Deutsch and Tran*, 2002; *Maharaja*, 2008; *Pyrcz et al.*, 2009), but it is not always possible to parameterize complicated geological phenomena by a few geometric parameters from irregular data (*Comunian et al.*, 2012). To overcome the abovementioned limitations, multiple-point statistics (MPS) was developed over the recent years and has shown excellent prospects in modeling subsurface anisotropic structures, such as porous media, hydrofacies, reservoir, and other sedimentary structures (*Comunian et al.*, 2011; *Dell Arciprete et al.*, 2012; *Hajizadeh et al.*, 2011; *Hu and Chugunova*, 2008; *Oriani et al.*, 2014; *Pirot et al.*, 2015; *Tahmasebi et al.*, 2014; *Wu et al.*, 2006).

A first MPS approach was suggested by *Guardiano and Srivastava* (1993) which is designed to reproduce heterogeneous geometries by extracting spatial patterns from training images directly rather than through variograms. A training image is a conceptual model derived from observations, and it bears a crucial role in MPS-based stochastic simulation. The first efficient implementation of MPS was developed by *Strebelle* (2002) on the basis of a tree structure. Later, the memory-efficient implementation IMPALA based on lists of data events was developed (*Straubhaar et al.*, 2011). Several attempts have thereafter focused on improving MPS algorithms (*Arpat and Caers*, 2007; *Caers*, 2001; *Mariethoz et al.*, 2010; *Tahmasebi et al.*, 2012; *Wu et al.*, 2008; *Yang et al.*, 2016; *Zhang et al.*, 2006). With these methods, training images are scanned with a fixed search template and the MPS patterns are stored in a tree or a list data structure. For the currently simulated node, the conditional probability density function (cpdf) is calculated according to the current data event. An important difficulty lies in choosing the size of data template to best reproduce large-scale structures (*Strebelle*, 2002). The larger the size of the data event, the fewer replicates of this data event will be found over the training images for inferring the corresponding cpdf. However, when the size of data template is too small, large scale structures of the training image cannot be reproduced (*Mariethoz et al.*, 2010). In addition, a search template including too many nodes can lead to storing a large number of pdfs in the tree or list structure, increasing CPU cost and memory consumption. The multiple-grids concept (*Tran*, 1994; *Strebelle*, 2002) mitigates the above-mentioned limitations, but they still present due to the rigidity of data templates and multiple grids. A more straightforward MPS method, Direct Sampling (DS), was proposed by *Mariethoz et al*. (2010) where the high order statistics are sampled directly from the training image without storing patterns and without the need of multiple grids. One of the main advantages of this approach is that several types of distances between patterns can be considered, making it possible to simulate continuous variables, or even multivariate simulation. As an approximation, pattern distance was used to express the matching degree between the neighborhood of a node and a data event in the

training image (*Chugunova and Hu*, 2008; *Mariethoz et al.*, 2010, 2015). For the pdf-based MPS methods, using the distances between patterns greatly decreases the amount of stored patterns. Some patch-based methods (*Arpat and Caers*, 2007; *Honarkhah and Caers*, 2010; *Tahmasebi et al.*, 2012; *Zhang et al.*, 2006) were proposed on the basis of this concept. By means of computer graphics, two very efficient MPS algorithms (*Li et al.*, 2016; *Mahmud et al.*, 2014) were developed to decrease the computational burden of traditional methods.

No matter which MPS algorithm is used, a suitable training image is the fundamental requirement. Although such algorithms are gaining popularity in hydrogeological applications (*Hermans et al.*, 2015; *He et al.*, 2014; *Høyer et al.*, 2017; *Hu and Chugunova*, 2008; *Huysmans et al.*, 2014; *Jha et al.*, 2014; *Mahmud et al.*, 2015), they still suffer from one vital limitation: the lack of training images, especially for 3-D situations. Object-based or process-based techniques are one possibility to build 3-D training images (*de Marsily et al.*, 2005; *de Vries et al.*, 2009; *Feyen and Caers*, 2004; *Maharaja*, 2008; *Pyrcz et al.*, 2009). Besides inherent limitations in the parameterization of these algorithms, it is also challenging to reproduce the various aspects of geological geometries from a high-resolution outcrop map, or even from insufficient borehole data (*Comunian et al.*, 2014; *Pirot et al.*, 2015). To overcome this difficulty, some assumptions have been implemented to reconstruct 3-D models using low-dimensional data (e.g. boreholes, cross-sections, outcrop and remote sensing images) directly instead of a training image in the entire 3-D domain (*Bayer et al.*, 2011; *Comunian et al.*, 2011; *Hu et al.*, 2011; *Weissmann et al.*, 2015). A promising reconstruction method of partially informed data sets was proposed by *Mariethoz and Renard* (2010) by using and adapting the DS algorithm. However, large-scale 3-D models contain millions of nodes, thus a very large number of scan attempts will be carried out for each simulated node, especially in early stages of a simulation due to the sparse known data. Therefore, this method still suffers from a severe computational burden for fine 3-D applications. Moreover, it assumes stationarity of the modeled domain, which is not often the case in practice. *Comunian et al*. (2012) proposed an approach to tackle the lack of a full 3-D training image using sequential 2-D simulations with conditioning data (s2Dcd): a 3-D domain is filled by preserving an overall coherence due to that a series of 2-D simulations performed using 2-D training images along orthogonal directions. To combine the cpdfs from different directions, several probability aggregation methods were tested and discussed (*Allard et al.*, 2012; *Bordley*, 1982; *Genest and Zidek*, 1986; *Journel*, 2002; *Krishnan*, 2008; *Mariethoz et al.*, 2009; Stone, 1961). Other 3-D applications to represent geological structures using MPS and partial data include filling in the shadow zone of a 3-D seismic cube (*Wu et al.*, 2008), generating small scale 3-D models of porous media (*Okabe and Blunt*, 2007) and building a 3-D training image with digital outcrop data (*Pickel et al.*, 2015).

From another perspective, using very common workflows, geologists can obtain 2-D geological maps or sections from scattered boreholes and/or other samples by studying analogs (*Caumon et al.*, 2009). With increasingly sophisticated data acquisition methods, 2-D high-resolution images can be acquired. For example, large-scale outcrop maps can be captured by using terrestrial lidar (*Dai et al.*, 2005; *Heinz et al.*, 2003; *Nichols et al.*, 2011; *Pickel et al.*, 2015; *Zappa et al.*, 2006), and fine-scale pore images can be derived from progressive imaging techniques (*Okabe and Blunt*, 2005; *Zhang et al.*, 2010). Therefore, there are many ways to acquire low-dimensional data for reconstructing a full 3-D model. In practice, however,

these works using real geological analogs or sections as training images still face significant non-stationarity due to the heterogeneity of geological phenomena and processes (*Comunian*, 2011; *de Vries et al.*, 2009).

To address the insufficient access to a 3-D training image and the challenge of non-stationarity, we present a new strategy to reconstruct 3-D geological heterogeneities using 2-D cross-sections instead of an entire training image. Compared to previous MPS implementations relying on partial data, our proposal is to use only several local sub-sections closer to the uninformed node as training images, rather than full planes perpendicular to the *x*, *y* and *z* directions (*Comunian et al.*, 2012; *Okabe and Blunt*, 2007) or searching in the entire 3-D domain (*Mariethoz and Renard*, 2010). The local sub-sections are able to offer more coherent and reliable statistics since they are spatially closer to the simulated node which is going to be simulated. Moreover, the original cross-sections are divided into many sub-sections according to their spatial relationships, thus non-stationarity is reduced since it is restricted into a local cube consisting of six or fewer sub-sections. In principle, our proposal can be applied into any multiple-point stochastic simulation method. In this work, we embed this strategy into a standard MPS framework called ENESIM (*Guardiano and Srivastava*, 1993). The blocking strategy proposed in this work can significantly reduce the search space of training images, which makes it possible to get a 3-D reconstruction using ENESIM for a reasonable CPU cost. As with DS, in our method MP statistics are not stored and the neighborhood is flexible. To integrate the patterns from different sub-sections, two probability aggregation formulas and their combinations are used. As an approximation of the matching degree between neighborhoods and data events, pattern distances are used to enhance the stability of cpdfs. Furthermore, we adapt multiple-grids into our approach, where the geometries of data templates are not fixed for grids of different scales. Besides cross-sections, any other scattered samples can also be involved into the proposal as conditional data (hard data).

The remainder of this paper is organized as follows. Section 2 presents the main concepts of the locality-based 3-D MPS reconstruction using 2-D cross-sections and the detailed steps of the proposed approach. Section 3 shows a parameter sensitivity analysis and the performance comparison with other MPS algorithms. Section 4 gives a comprehensive application case to illustrate the effectiveness of our approach when facing the real geological field data. The final section discusses some concluding remarks and ideas for future work.

2. Methodology

2.1. Local Search Strategy of 3-D MPS Reconstruction

Comunian et al. (2012) presented a commonly used 3-D MP simulation approach using 2-D training images. The key idea is to capture 3-D MP statistics by using two/three orthogonal training images and then merging the MP statistics from different directions into an integrated pdf. However, any locations of the training images are scanned even they are far away from the simulated node so that one spatial pattern will be carried to a distant position. Therefore, the use of this method is restricted to stationary training images, which are in practice seldom available. In this work, we propose a local search

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Hydrology and Earth System Sciences
Discussions

strategy that allows to palliate this problem, by taking into account the spatial relationships of the real geological cross-sections in a given 3-D domain.

As illustrated in Figure 1, a 3-D domain is segmented into nine small blocks by six cross-sections from three orthogonal directions where there are two sections in each direction. Every local block is surrounded by n local sub-sections ($1 \le n \le 6$). It should be noted that, sometimes, local blocks are not closed (i.e. the surrounding sub-sections are less than six) (Figure 1b); and it is also allowed sections along some planes are missing; however, at least one section should be provided. For each unknown node in the local block (e.g. the gray cubes in Figure 1c), the MP statistics are captured from the surrounding sub-sections rather than from the entire sections. Namely, there are n corresponding training images for each simulated node. These local sub-sections are the parts of the global cross-sections which are closer to the uninformed nodes in the local block, thus they are more likely to be regarded as statistically representative.

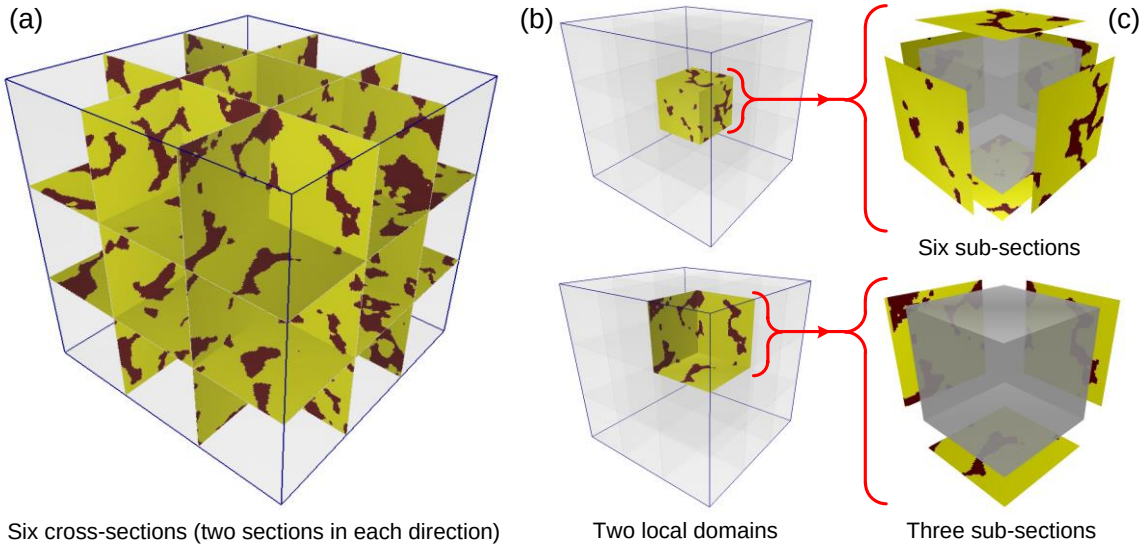

Figure 1. Local sub-sections divided by their spatial relationships and the corresponding training images. (a) Six cross-sections in a 3-D domain: two sections along each direction; (b) two local domains: a central cube and a corner cube; (c) corresponding sub-sections (training images).

Another important point is related to handling of the search window when scanning a sub-section. Here, we allow all locations of a sub-section to be visited by the central node of a data event. The neighbor nodes of the data event can be placed in other adjacent sub-sections when matching with the training images. As shown in Figure 2, the area inside the blue line is the search window. If only the nodes of the data event are from the sub-section itself (case 1 on the figure), the training patterns are seriously reduced. We adopt a search strategy where neighbor nodes can be searched in the neighboring sub-sections (case 2 on the figure). Its main advantages are the coherence of the spatial patterns in a realization and the larger number of training patterns available. In addition, the size of the data events is constrained by the boundary of the global section, as illustrated in *Mariethoz et al.* (2010).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Hydrology and Earth System Sciences
Discussions

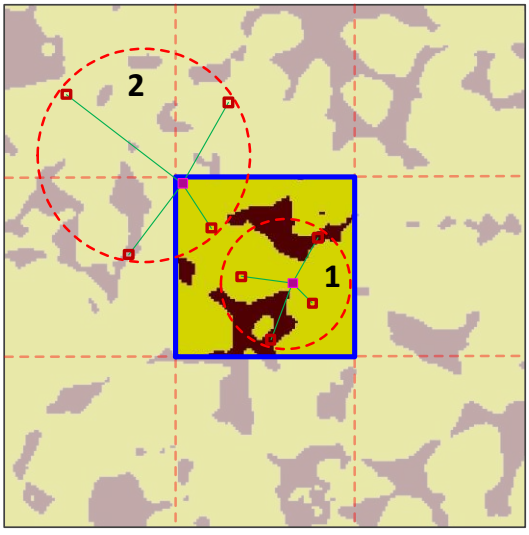

Figure 2. Search window in sub-sections.

If more cross-sections are available, a finer spatial subdivision can be used. In this case, the size of each sub-section is smaller and the computational cost is reduced significantly. However, extremely small training images cannot offer enough spatial patterns, thus a minimal sub-section size has to be considered. In practice, if there are many sections in each direction, a feasible solution is to select several ones as the references and others are used as conditioning data only.

2.2. Probability Aggregation

Allard et al. (2012) presented a comprehensive literature review for aggregating probability distributions. These can be divided into additive methods and multiplicative methods according to their mathematical properties. Linear Pooling formula (*Stone*, 1961) is a widely used method (for example, it was used by *Okabe and Blunt*, 2007) based on the addition of probabilities. It is appealing because of its flexibility and simplicity. Multiplicative methods include Bordley/Tau models and log-linear pooling (based on odd ratios) (*Bordley*, 1982; *Journel*, 2002; *Genest and Zidek*, 1986).

2.2.1. Linear Pooling Formula

The linear pooling formula, proposed by *Stone* (1961), probably is the most intuitive way of aggregating the probabilities $P_1,...,P_n$ of an event A.

$$P_G(A) = \sum_{i=1}^{n} w_i P_i(A) \quad \text{with } w_1,...,w_n \in \mathbf{R}^+. \tag{1}$$

In this formula, w_i are positive weights and their sum must equal one to obtain a global probability $P_G \in [0,1]$.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

2.2.2. Log-Linear Pooling Formula

A log-linear pooling operator is a linear operator of the logarithms of the probabilities (*Genest and Zidek*, 1986). If a prior probability $P_0(A)$ must be included, the log-linear pooling formula is written:

$$P_G(A) \propto P_0(A)^{1-\sum_{i=1}^{n} w_i} \prod_{i=1}^{n} P_i(A)^{w_i}. \quad (2)$$

$\sum_{i=0}^{n} w_i = 1$ is needed to verify external Bayesianity. There are no other constraints whatsoever on the weights w_i, $i = 0,...,n$. The sum $S = \sum_{i=1}^{n} w_i$ plays an important role in this formula. If $S = 1$, the prior probability P_0 is filtered out because $w_0 = 0$. Otherwise, if $S > 1$, the prior probability has a negative weight and P_G is further away from P_0 than other probabilities. Conversely, if $S < 1$, P_G is always closer to P_0. Therefore, we can adjust the influence of the prior probability P_0 on the aggregated result P_G by changing the value of S.

2.2.3. Strategy for Aggregating the pdfs from Local Sub-Sections

As an additive aggregation method, the linear pooling formula corresponds to a mixture model, which is related to the union of events and to the logical operator OR (*Allard et al.*, 2012). This method is thus used to unite several independent probabilities into a global term P_G. The log-linear pooling formula, based on the multiplication of probabilities, is related to the intersection of events and to the logical operator AND. Therefore, we usually use such a method to aggregate the probabilities with significant correlation to acquire a conjunction probability.

In this study, n pdfs ($1 \le n \le 6$) are computed from the surrounding local sub-sections (Figure 1). For the illustrative case proposed here, a local 3-D domain is surrounded by six sub-sections, and six pdfs are being aggregated. There are two parallel sub-sections (training images) in each direction. Intuitively, an additive aggregation operator seems more appropriate to combine such two disjunctive probability distributions, since we just expect a larger number of samples and thus and more robust pdf by uniting both. Then, three orthogonal pdfs are obtained. We then join these pdfs containing the statistics from different directions. In summary, an optimal probability aggregation strategy is proposed by the procedure described below:

1. Aggregate the pdfs collected along the same direction for parallel sub-sections using the linear pooling formula.
2. Aggregate the orthogonal pdfs from the above step by using the log-linear pooling formula.

Of course, the probability aggregation step is not required when for step 1 there is only one sub-section along a given plane, and for step 2 the pdf that along some direction are missing are simply not included in the aggregation process. For the step 1, the weights w_1 and w_2 are related to the distances between the current location and the two parallel sub-sections d_1 and d_2 s, and computed as:

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

$$w_1 = \frac{1/d_1}{1/d_1 + 1/d_2},\ w_2 = \frac{1/d_2}{1/d_1 + 1/d_2}. \tag{3}$$

Such parameterization ensures that within-block trends are accounted for.

For the step 2, an influence of the prior probability is desired to tune the other orthogonal pdfs. Thus, we usually use $0 < w_0 < 1$, and set $w_i (i = 1, ..., n)$ equal, i.e. $w_i = (1 - w_0)/n$, where n is the number of pdfs to be aggregated. However, the weights $w_i (i = 1, ..., n)$ can also change, for example, they can vary at each simulation step as described in *Comunian et al.* (2012), according to the contributions of the different training images, while sum still respects the condition $\sum_{i=0}^{n} w_i = 1$.

2.3. Pattern Distance

A pattern distance $d\{\mathbf{N}_X, \mathbf{N}_Y\}$ is an approximation of the dissimilarity between patterns, which is used to compare the neighborhood of a node currently simulated with a data event in the training image (*Mariethoz et al.*, 2010). Approximate matches are accepted by using a distance threshold t. Namely for a data event $\mathbf{N}_X$ from the simulation grid, when the condition $d\{\mathbf{N}_X, \mathbf{N}_Y\} \le t\ (t \ge 0)$ is met, the pattern $\mathbf{N}_Y$ from the training image will be used to update the current cpdf. For a categorical variable, the distance can be formulated as:

$$d\{\mathbf{N}_X, \mathbf{N}_Y\} = \frac{1}{n}\sum_{i=1}^{n} a_i \in [0,1], \quad where\ \ a_i = \begin{cases} 0 & \text{if } Z(x_i) = Z(y_i), \\ 1 & \text{if } Z(x_i) \ne Z(y_i). \end{cases} \tag{4}$$

For a non-stationary training image from an actual geological phenomenon, repeatability of spatial patterns could be weak so that it is hard to acquire a stable cpdf. Therefore, we adopt a patterns distance with a threshold as an approximation to sample more patterns and get a more stable cpdf.

2.4. Flexible Search Template on Multiple Grids

When large neighborhoods are considered, it is more difficult to find matching data events in the training image and thus a larger distance threshold t is required to obtain a sufficient number of samples for an acceptable cpdf. This can lead to degrading small-scale features or the removal of categories that have a low proportion. To address this issue, we propose a novel implementation of multiple grids where the search template is flexible and the distance threshold t varies according to the radius of the neighborhood.

As illustrated in Figure 3, an example of multiple grids with three levels is used to show the relationship between neighborhoods, search radius R and distance threshold t on different grids. A neighborhood is identified by the informed and/or simulated nodes located in the circle with a radius R and the current node (the gray nodes in Figure 3) as a central.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Hydrology and Earth System Sciences
Discussions

The initial radius R_0 and distance threshold t_0 for the first grid are assigned as the input parameters. The radius R linearly reduces to 1 from the first to the last grid, and the threshold t similarly varies from 1 to 0. A large data event is divided into several small parts placed on the different grids which results in smaller neighborhoods on each grid. An acceptable threshold t is thus assigned to each neighborhood. For the last grid, the radius is reduced to 1 and at most there are eight nodes in a neighborhood (Figure 3 is a 2-D sketch of multiple grids, but we mainly focus on 3-D simulations in this paper). This strategy considers that small data events located on the last grid are much more repetitive (thus easier to find) than the large data events of the first grid.

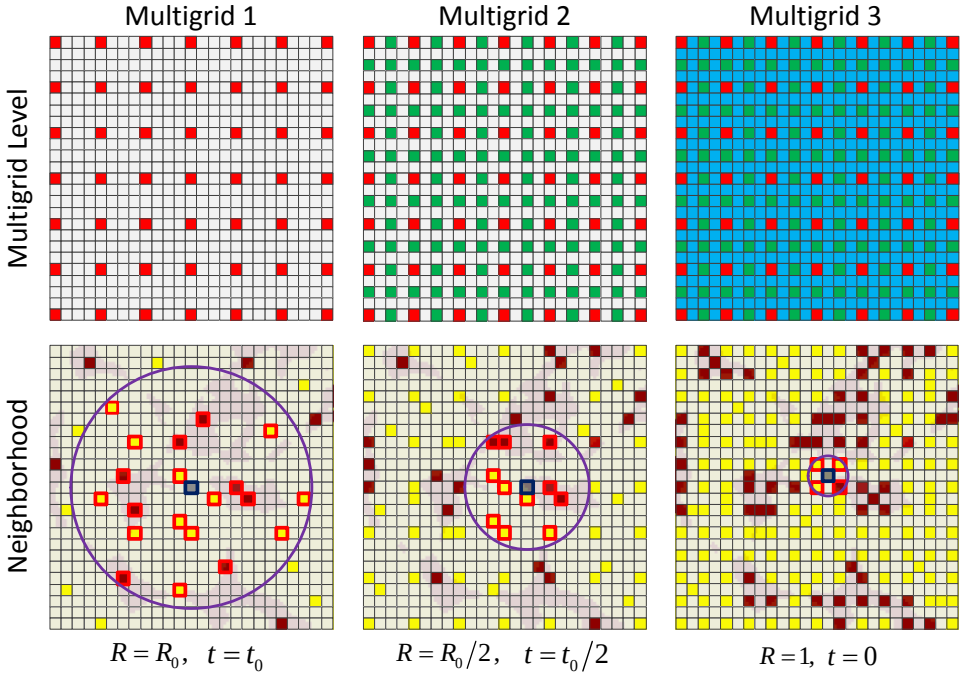

Figure 3: An example of multiple grids and the corresponding neighborhoods, search radius R and distance threshold t.

2.5. Step-by-Step Algorithm Using the Local Search Strategy

Based on the strategies proposed in the above sections, the detailed steps of our simulation algorithm proceed as illustrated in Algorithm 1.

Algorithm 1: Reconstruct 3-D geological structures using 2-D cross-sections

1 Load data files, assign all points of the training data (cross-sections and other samples) into the SG.
2 Record the indexes of the sections in *x*, *y*, *z* directions and compute the prior proportions $\mathbf{P}_p$ of the local domains.
3 **For** each multiple grid g :
4 Define a random simulation path for grid g according to the remaining nodes.
5 **Do** until all uninformed nodes in g have been visited:

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

6 Get the index of current node $\mathbf{x}$, and identify its neighborhood $\mathbf{N}_x$.
7 Obtain the indexes of the closest sections around $\mathbf{x}$: $\{x_0, x_1\}$, $\{y_0, y_1\}$, $\{z_0, z_1\}$.
8 Randomly scan the sub-sections (TIs) and get the corresponding cpdfs (see algorithm 2).
9 Get the prior proportion $\mathbf{p}$ of the local domain according to the location of the node $\mathbf{x}$.
10 Combine the cpdfs and $\mathbf{p}$ into a joint pdf using the strategy presented in section 2.2.
11 Randomly draw a value from the final pdf, and assign it to location $\mathbf{x}$.
12 **End**
13 **End**

As mentioned above, we capture the MP statistics from several sub-sections of a local domain. Thus, the corresponding prior proportion should also be computed on the basis of these surrounding sub-sections (step 2). Obviously, step 8 is the most important procedure in our simulation algorithm, and the idea is inspired from ENESIM (*Guardiano and Srivastava*, 1993) and DS (*Mariethoz et al.*, 2010). The main procedure is demonstrated in Algorithm 2.

Algorithm 2: Scan a local sub-section (training image) in one certain direction

Input: $\mathbf{x}$**:** current simulation location; id : index of the training image that will be scanned;
$\chi_0, \chi_1, \gamma_0, \gamma_1$: indexes of the closest training image s in the other two directions.
Output: cpdf: conditional probability density function from the current training image.

1 **Function** ScanTI($\mathbf{N}_x, id, \chi_0, \chi_1, \gamma_0, \gamma_1$, &cpdf)
2 Get the sub-section sub_S (training image) according to the id and $\chi_0, \chi_1, \gamma_0, \gamma_1$;
3 Set a random path p, and initialize the counter of matched patterns $sum = 0$;
4 **for** $i := 0 \rightarrow p.\text{size}()$ **such that** $i < p.\text{size}() \times f$ **do**
5 Sample a location in the training image and get the neighborhood $\mathbf{N_Y}$;
6 Compute the distance $d\{\mathbf{N_X}, \mathbf{N_Y}\}$ using equation (4);
7 **if** $d\{\mathbf{N_X}, \mathbf{N_Y}\} \le t$ **then**
8 update the cpdf according to the facies of the central point in the training image;
9 sum++;
10 **end if**
11 **if** $sum > N_{\max}$ **then**
12 **break;**
13 **end if**
14 **end for**
15 **end Function**

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

The fraction of the scanned training image f and the distance threshold t are borrowed from DS and they play the same roles. $\chi_0, \chi_1, \gamma_0, \gamma_1$ are the indexes of the closest training images in the other two directions and they are used to determine the current sub-section (training image). A new parameter, the maximum number of matched patterns from the training image $N_{\max}$ is adopted to avoid unnecessary searches. For some small neighborhoods, especially in the last multiple grid, the cpdf will rapidly stabilize with the increasing number of matched patterns.

3. Parameterization and Performance Analysis

In this section, we apply our method on several synthetic cases where the cross-sections are extracted from existing 3D references. Using these examples, we perform a parameter sensitivity analysis and compare it with two widely used methods, DS-based 3-D reconstruction (*Mariethoz* and *Renard*, 2010) and s2Dcd (*Comunian et al.*, 2012). The workflows and algorithms proposed in this work are developed in the C++ programming language. All experiments presented in this paper are implemented on a laptop computer with Intel 4-Cores i5-62000U Quad-core CPU 2.30 GHz, 8 GB RAM and 64 bit Windows 10.

3.1. Parameter Sensitivity

The majority of parameters of our approach are similar to DS. Therefore, only the sensitivity of three parameters specific to our approach are tested against the 3D reference shown in Figure 4, considering CPU cost and statistical and geometrical features of the realizations obtained. All cross-sections used in the following tests in the section 3.1 are extracted from this 3D model.

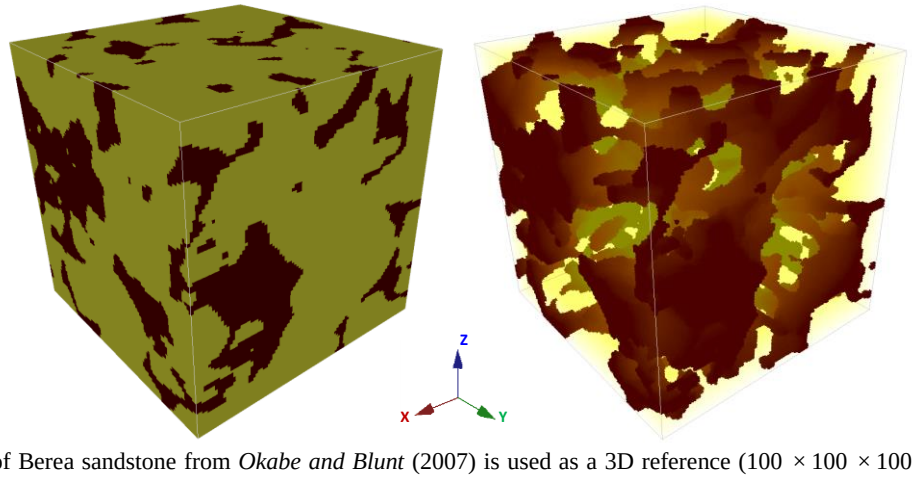

Figure 4. A sample of Berea sandstone from *Okabe and Blunt* (2007) is used as a 3D reference (100 × 100 × 100 voxels). The crimson color represents pores and the yellow color represents matrix. The porosity of this model is 20.33%.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

3.1.1. Number of Cross-Sections

The number of cross-sections is a critical parameter in our approach. They are not only regarded as the training images and conditioning data, but also control the computing speed and the quality of the reconstructions. Figure 5 and Table 1 show different reconstructions and their statistical properties by increasing the sections in every direction. In this test, the number of cross-sections in each direction increases from one to six, and other parameters are fixed: maximum search distance = 50, maximum number of points in a neighborhood = 35, distance threshold $t_0 = 0.2$, fraction of training image to scan $f_0 = 0.8$, maximum of matched patterns from each training image = 100, number of multiple grids = 3, weights of the probability aggregation $w_0 = w_1 = w_2 = w_3 = 0.25$. We obtain 20 realizations for each set of cross-sections. The main difference between the different settings is the improvement of computational efficiency with the increase in cross-sections. The proportion of pores (porosity) is reproduced at a similar level for each group, Also, when increasing the number of cross-sections, the number of geobodies gets closer to the reference, and the variability is decreased and the connectivity is more and more stable, which are caused by the increase of conditioning data (i.e. informed cross-sections). On the other hand, using too many cross-sections will lead to a number of artifacts since the training sub-sections for each sub-block are very small, resulting in insufficient number of samples. As a consequence, we recommend that several sections can be chosen if there are abundant candidates in one direction, which must ensure that the features of selected ones are diverse and contain enough spatial patterns, but not incurring artifacts. In this test, 3 to 5 sections in each direction are recommended, but it related to the size of simulation grid in other 3-D application. When there are very few or no sections in a direction, a feasible solution has been suggested by *Gueting et al.* (2017) where sequential 2D simulations are performed to obtained some sections first, and then both the original informed data and the obtained sections are used to reconstruct the model of the entire 3D domain.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

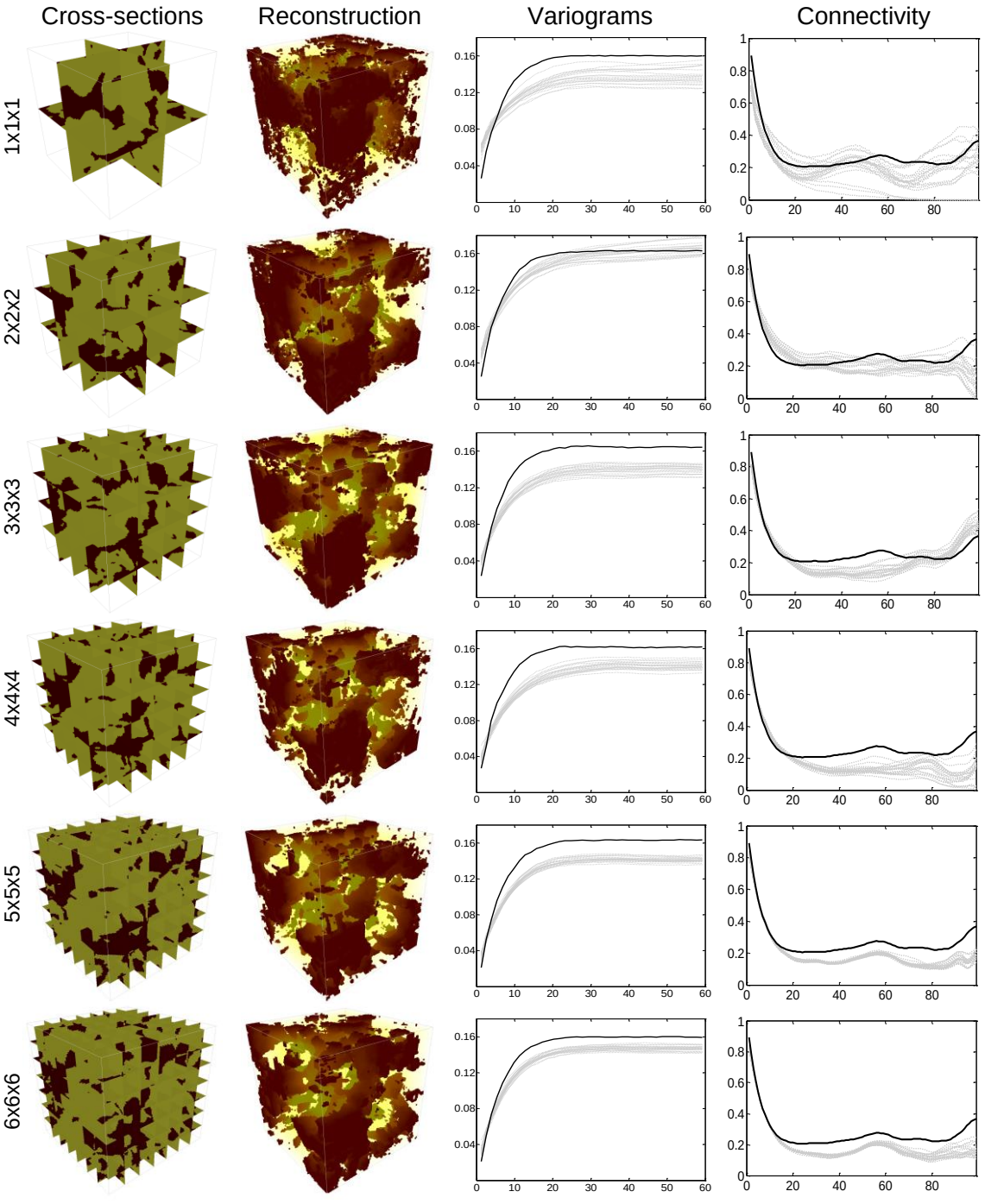

Figure 5. Reconstructions and their statistical properties with increasing the number of sections in each direction. We only present the connectivity functions computed along the coordinate Y since their features are similar in three directions.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Hydrology and Earth System Sciences
Discussions

Table 1. Comparison of the performance of the tests in Figure 5. All the statistics are the averages of 20 realizations.

Test	Num. of sections	Sub-blocks	Porosity (%)		No. of geobodies	Time (s)
			Training sections	Results		
1×1×1	3	8	19.07	16.36	1781	1382
2×2×2	6	27	21.35	19.95	908	718
3×3×3	9	64	18.70	16.22	572	396
4×4×4	12	125	19.62	16.21	471	271
5×5×5	15	216	19.81	16.80	340	183
6×6×6	18	343	19.74	17.32	326	127
3D Ref.			**20.33**		**144**	

3.1.2. Maximum number of Matched Patterns from Each Training Image

Table 2 shows the statistics of 20 realizations obtained by varying the maximum of matched patterns from each training image N_{max}, which is a novel parameter adopted in this work to avoid the unnecessary searches during obtaining a cpdf from training images. Other parameters are the same as in the former test presented in Figure 5, except for the sections in each direction which are fixed to 3. We find that the computational cost increases sharply when $N_{max}>160$ and then stabilizes. Concerning the compared statistical properties, low values of N_{max} result in variabilities are closer to the reference with bigger variances which is due to it is almost like sampling the result directly from training images resulting in the role of cpdfs is lost. For the remaining cases, the statistics are similar except for a decrease of variances with increasing N_{max} (Table 2). In order to better grasp the effect of N_{max}, three cases are selected (N_{max} = 5, 40, 320) and the corresponding realizations are shown in Figure 6a-c. The connectivity functions vary in a large range for small N_{max} values. Conversely, they become more stable when increasing N_{max} (Figure 6d). The variance of variables bears the same tendency by increasing N_{max} (Figure 6e). Consequently, N_{max} = 40 to 120 is recommended resulting in a balance between a stable cpdf and computational cost.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Table 2. Comparison of the performance for 20 realizations with three sections in each direction, and varying the maximum of matched patterns from each training image N_{max}. Other parameters are fixed and are same with the test of Figure 5. All the statistical values are the mean of 20 realizations. ∞ represents no constraint for N_{max}.

N_{max}	Porosity (%)	Variance	No. of geobodies	Time (s)
5	18.39	0.150	365	132
10	17.22	0.143	440	161
20	16.69	0.139	486	200
40	16.47	0.138	505	251
80	16.48	0.138	510	417
160	16.38	0.137	519	495
320	16.50	0.138	503	549
640	16.66	0.139	508	587
∞	16.89	0.138	497	589
Ref.	**18.70**	**0.152**	**144**	

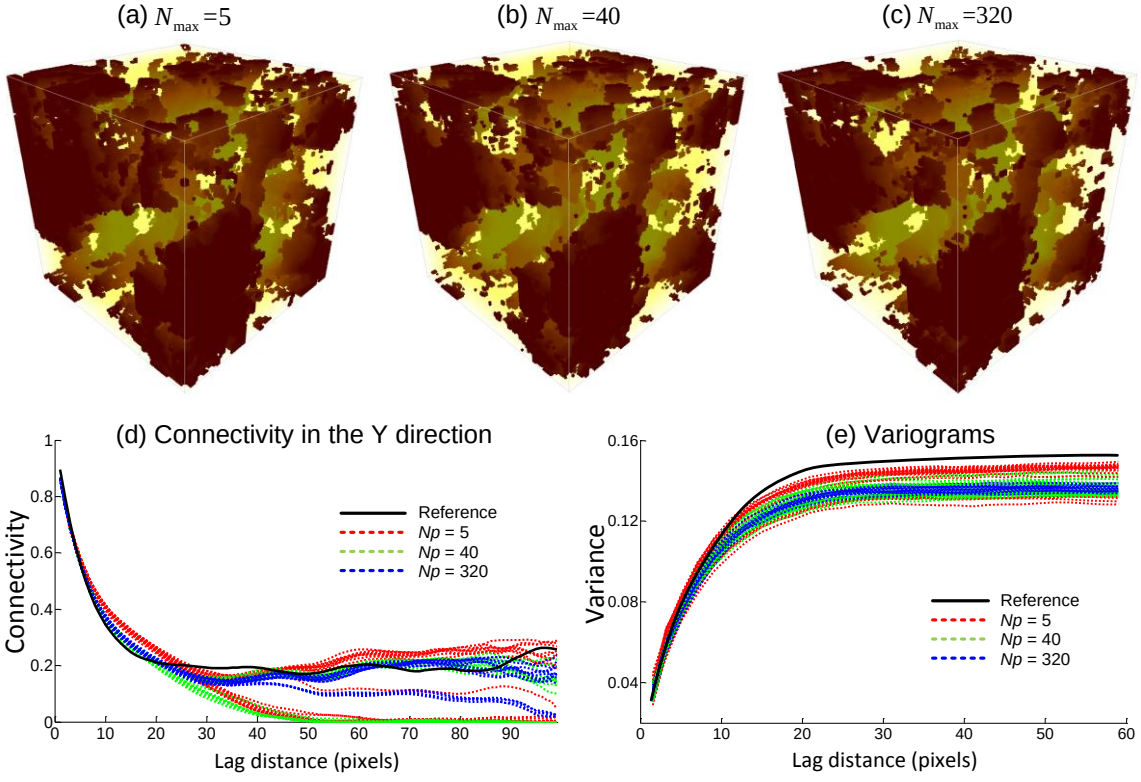

Figure 6. Reconstructions and their statistical properties with N_{max} = 5, 40, 320 selected from Table 2.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Hydrology and Earth System Sciences
Discussions

3.1.3. Weights of the Probability Aggregation Formulas

In this work, the strategy for aggregating the pdfs from local sub-sections includes two steps. In the first step the weights of Linear Pooling Formula for two parallel sub-sections are selected depending on the distances between the current location and the two sub-sections in the first step. Therefore, the weights are automatically set and do not need to be set. In the second step, the appropriate weights for the prior probability distribution and three orthogonal cpdfs are to be selected by the user. Figure 7 shows different realizations obtained by varying the four weights w_0, w_1, w_2, w_3. Here we increase the weight of the prior probability distribution w_0 and let the other three weights equal, since the cpdfs from three orthogonal directions have the same contribution. Of course, if users think the cpdf of one direction is more important than others, they can be changed, under the constraint that $\sum_{i=0}^{3} w_i = 1$. It can be observed that when $w_0 = 0$, the spatial structures are well reproduced, but with larger variance (Figure 7a) since all spatial patterns are inferred from the MP statistics of the surrounding sub-sections rather than using prior information. When increasing w_0, the continuity of the spatial structures is degraded, but the facies proportions are closer to the reference (Figure 7b). Finally, in the extreme case of (Figure 7c) the continuity of spatial structures is lost. Therefore, $0 \le w_0 \le 0.25$ is a recommended range and other three weights can be determined by the importance (e.g. complexity or variety of patterns) of the sections in each direction.

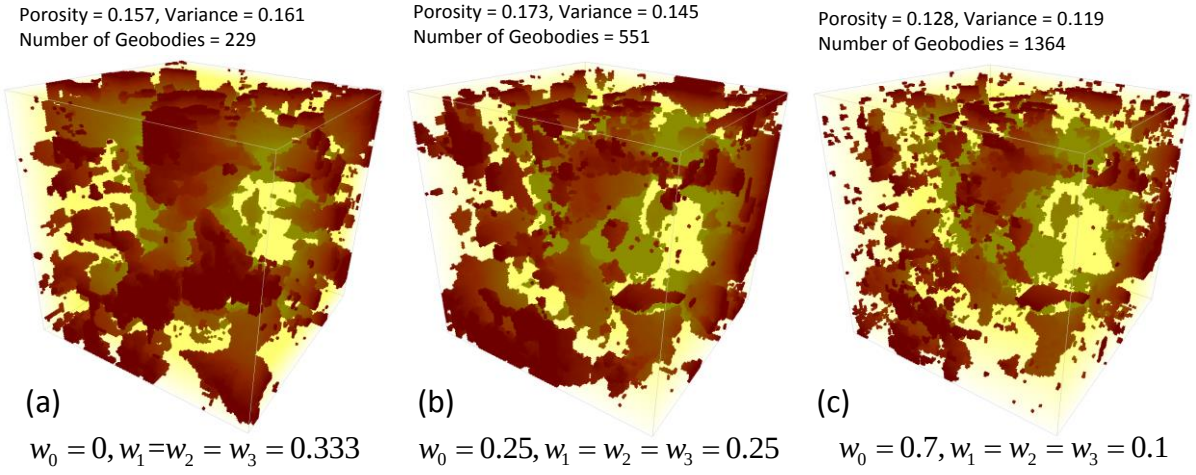

Figure 7. Three realizations obtained by varying the weights of the probability aggregation formulas. Three sections in each direction are used and other parameters are same with the test of Figure 5.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

For the other parameters involved in our algorithm, most of them are similar to the parameterization of DS which have been tested thoroughly in *Meerschman et al*. (2013). However, our method allows larger initial values for the neighborhood size and the distance threshold because multiple grids are used so that these initial values are decreased with increasing the level of multiple grids.

3.2. Comparison of Reproducing Heterogeneities with Existing Methods

To verify the validity of our approach for reproducing heterogeneous structures, we compare it with two MPS implementations that use partial data: DS (*Mariethoz and Renard*, 2010) and s2Dcd (*Comunian et al.*, 2012). As shown in Figure 8, six cross-sections extracted from a 3-D model of folds (180 ×150 × 120 voxels) (*Mariethoz and Kelly*, 2011) are utilized in this test. s2Dcd is a wrapper library that requires an external MPS engine. In order to ensure comparability, here DS is employed as the engine of s2Dcd. The detailed parameters are as follows: maximum search distance = 40, maximum number of points in a neighborhood = 40, distance threshold $t_0 = 0.2$, fraction of training image to scan $f_0 = 0.8$, maximum of matched patterns from each training image = 100, level of multiple grids = 3, weights of the probability aggregation $w_0 = w_1 = w_2 = w_3 = 0.25$. In other two methods, a smaller distance threshold $t = 0.05$ is considered and other essential parameters are same with our method. Because the implementation of DS is parallel, we use 4 processors to carry out this test in DS and s2Dcd. Only one processor is used in our method because our implementation is not parallel. In Figure 8, one selected realization for each method is presented. From their visual appearance, it looks that s2Dcd and our method have the similar performance for reproducing the patterns shown in 3-D reference and informed cross-sections. Therefore, histograms, variograms, and connectivity functions are used to further analyze performance. Figure 9 shows the comparison of proportions of the facies for the realizations by using three MPS methods. 20 realizations are performed for each method. It can be seen that the facies proportions with our methods are closer to the proportion of the informed cross-sections. The variograms and the connectivity functions on three directions for the 3-D reference and the generated 20 realizations of each method are shown in Figures 10 and 11, indicating that all three methods are able to reproduce the basic statistics of the 3-D reference, but the proposed method is generally closer to the reference.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

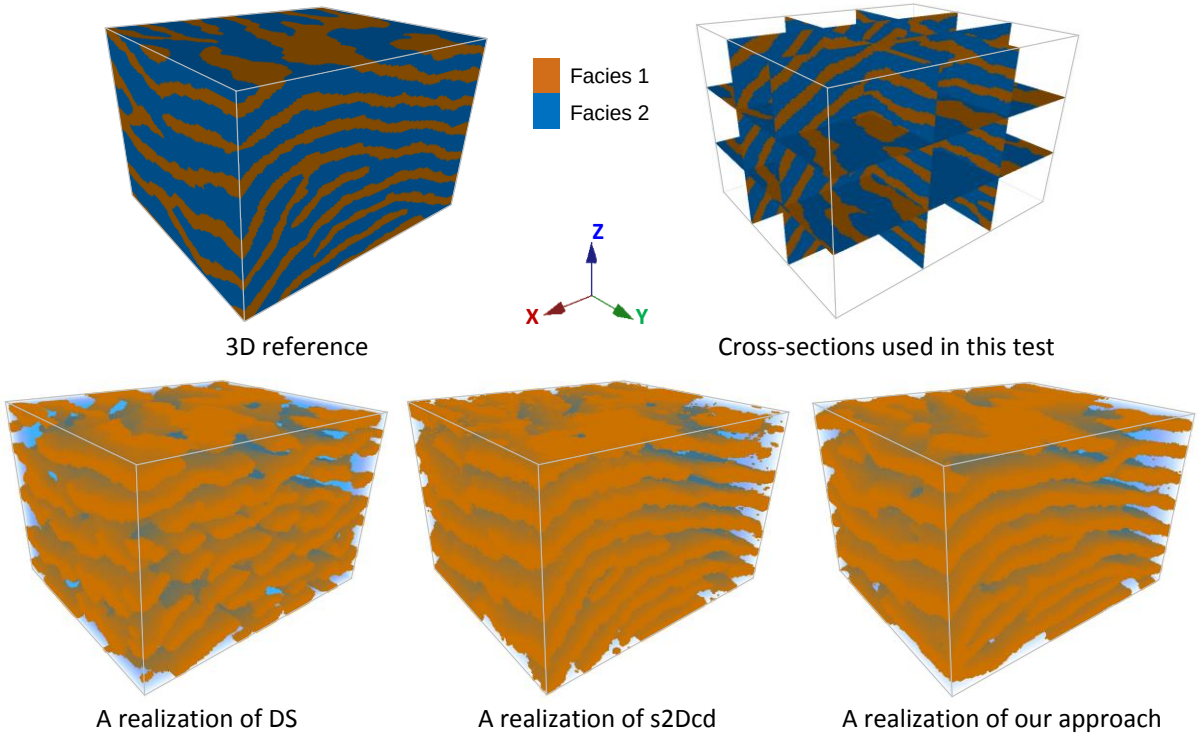

Figure 8. Realizations of three different MPS reconstruction methods. In our method, the parameters.

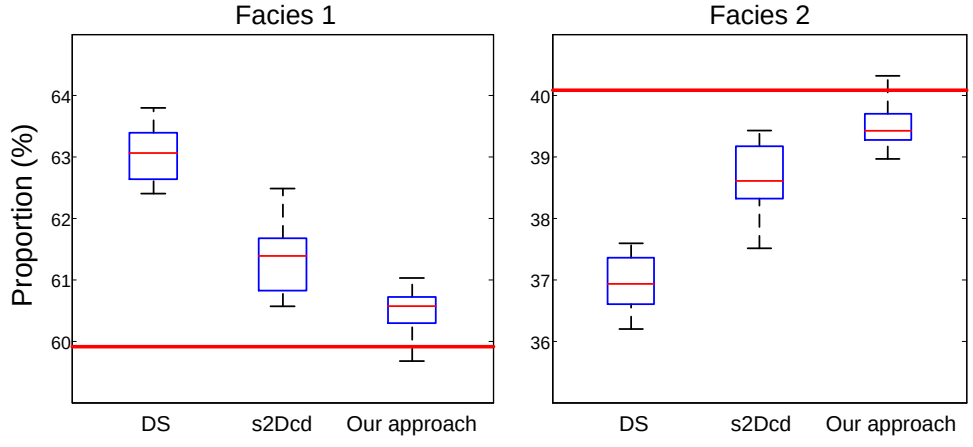

Figure 9. Proportions of the facies for 20 reconstructions by using three MPS methods. The red horizontal line represents the proportions of facies in the cross-sections used as training images.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

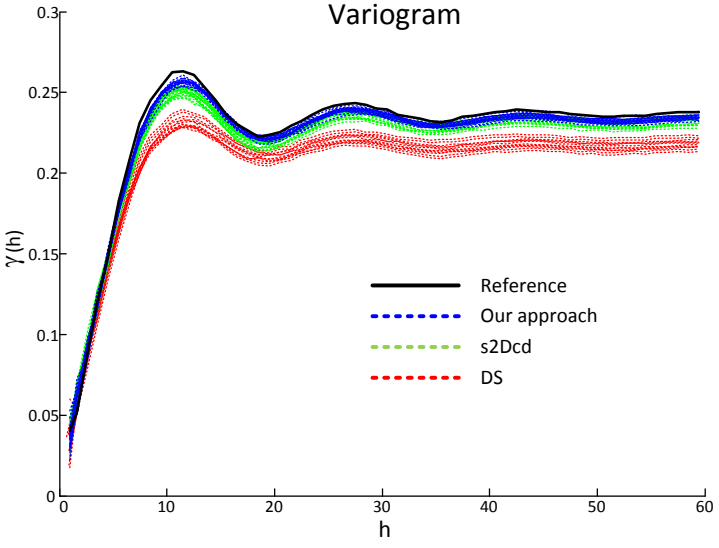

Figure 10. Comparison of the variograms between DS, s2Dcd and our approach.

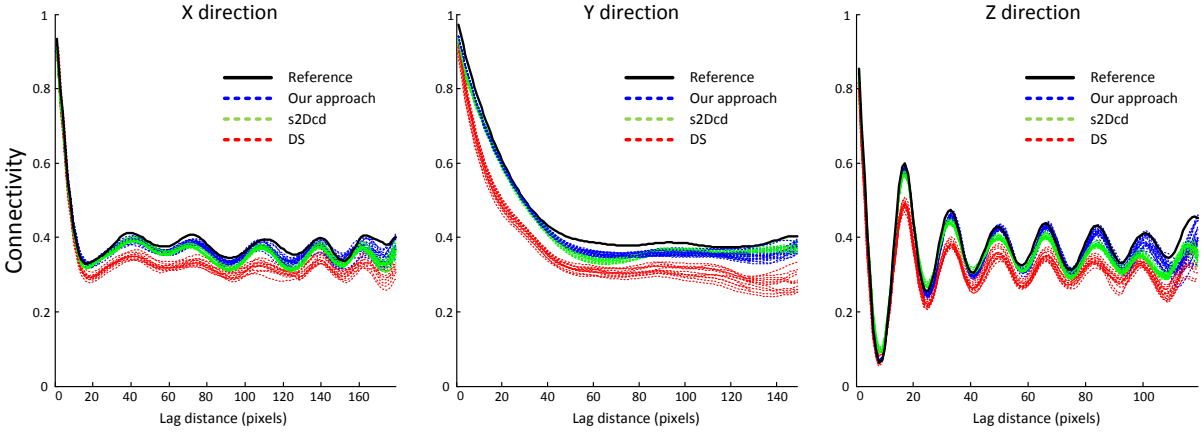

Figure 11. Comparison of the connectivity functions in three directions with three MPS methods.

Tan et al. (2014) proposed a distance-based approach to evaluate the quality of MP simulation outcomes where the Jensen–Shannon (JS) divergence is used to depict the dissimilarity of MP histograms as a quantitative metric. The information in the dissimilarity of MP histograms can be visualized using multidimensional scaling (MDS) (*Caers*, 2011). MDS approximates these distances by a lower-dimensional Euclidean distance in Cartesian space, which facilitates the visualization of the dissimilarity of MP histograms. To further compare the models obtained using the three different MPS approaches, MDS plots are constructed by calculating the distance of MP histograms between all the realizations of three approaches and a 3-D reference. The resulting MDS map is shown in Figure 12 and it can be observed that the realizations of our method are closer to the reference in the MDS map than the results obtained by the other two approaches.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

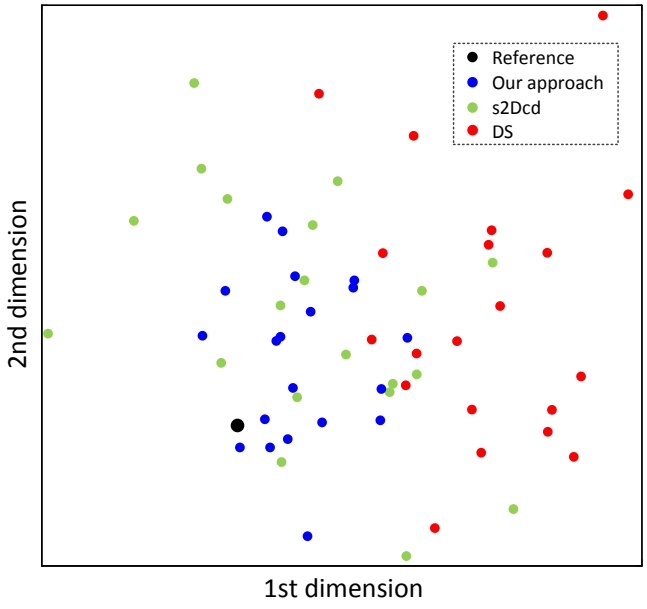

Figure 12. MDS representation for 20 realizations of each MPS methods.

In practice, there is no fully informed 3-D reference and we only have several informed cross-sections. Thus, the statistical features of the reconstructions (e.g. variograms, connectivity functions and MDS plots) are close to the reference but no one can surpass it in the above test. However, these comparisons are still able to validate the reproduction of spatial patterns for the different MPS approaches.

3.3. Computational Performance

In this section we focus on the analysis of computational performance. For the test illustrated in section 3.1.1, the average time per realization for different number of cross-sections, shown in Figure 13a, decreases rapidly with increasing the number of cross-sections from one to six in each direction. Although this factor seriously affects the computational efficiency, we should strike a balance between the computational efficiency and variability of spatial patterns which is described in detail in section 3.1.2. For the maximum of matched patterns from each training image $N_{\max}$ depicted in section 3.1.2, the computational cost increases sharply in the first half part by changing $N_{\max}$, and then approximately tends to be stable (Figure 13b). Thus, a good choice for this factor is close to the turning point before the curve tends to be stable. The weights of the probability aggregation formulas do not affect CPU time and the computational performance of other parameters has been assessed clearly by *Meerschman et al.* (2013).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Hydrology and Earth System Sciences
Discussions

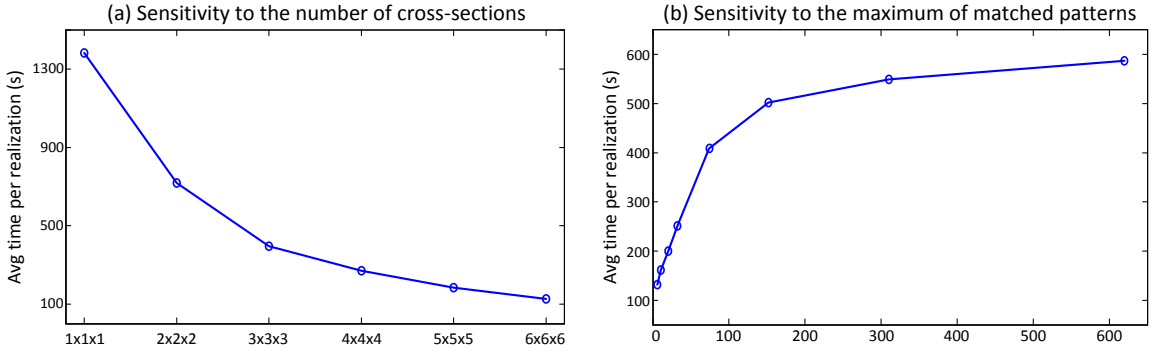

Figure 13. Effect of (a) the number of cross-sections in each direction and (b) the maximum of matched patterns from each training image. The times are averaged on 20 realizations.

A comparison of computational performance between DS, s2Dcd and our approach is presented in Figure 14. Because our method is sensitive to the number of input cross-sections, we offer two and four sections in each direction respectively, and the computational efficiencies are shown in Figures 14a and 14b with increasing the total number of grid cells. Other parameters are the same as the test in section 3.2. Note that a different time axis is used for DS-based reconstruction because it uses much more CPU time than the other two methods, even though four processors are used for DS-based reconstruction and only one for the other two methods. As shown in Figures 14a, s2Dcd and our approach present similar computational performance but much better than DS-based 3-D reconstruction since the MP statistics are captured from a smaller domain composed of several 2-D sections in s2Dcd and our approach. When increasing the number of cross-sections, the search space is divided into more subdomains in our approach so as to achieve a better performance than s2Dcd (see Figure 14b).

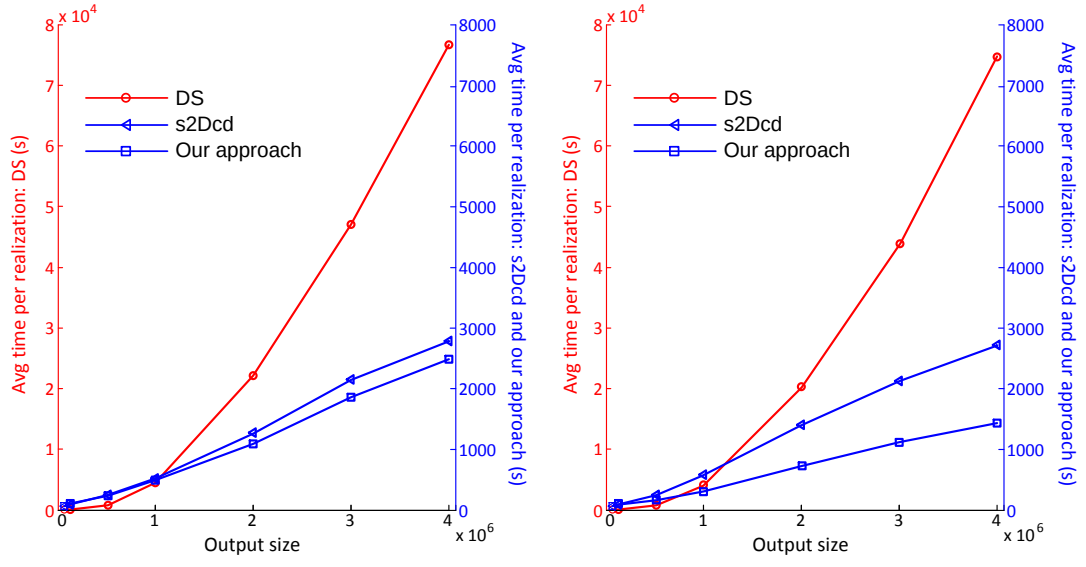

Figure 14. Comparison of computational performance between DS, s2Dcd and our approach with increasing the size of output grid: (left) two cross-sections and (right) four cross-sections in each direction. Note that different time axises are used in the two subplots, and four processors are used for DS-based reconstruction and only one for the other two methods.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Hydrology and Earth System Sciences
Discussions

4. Application Example

To further demonstrate the applicability of our algorithm, an example from a real geological application is presented in this section. The Descalvado aquifer analog dataset (Figure 15) depicts the complex hydrofacies of a small area (28m × 7m × 5.8m) in Brazil (*Bayer et al.*, 2015). In the original dataset, there are five cross-sections derived from outcrops, which are marked by black lines in Figure 15a. They are referenced in a 3-D domain consisting of 280 × 70 × 58 voxels. These sections allow creating only two parts so subdomains, which is insufficient for an application of our method. Therefore, we borrow the strategy of *Gueting et al.* (2017) to insert three additional sections in yz direction using sequential 2-D MP simulation approach (s2Dcd) firstly which are marked by blue lines in Figure 15a. Then, all the tests are implemented on the basis of eight cross-sections (three in xz direction and five in yz direction) which are shown in Figure 15b.

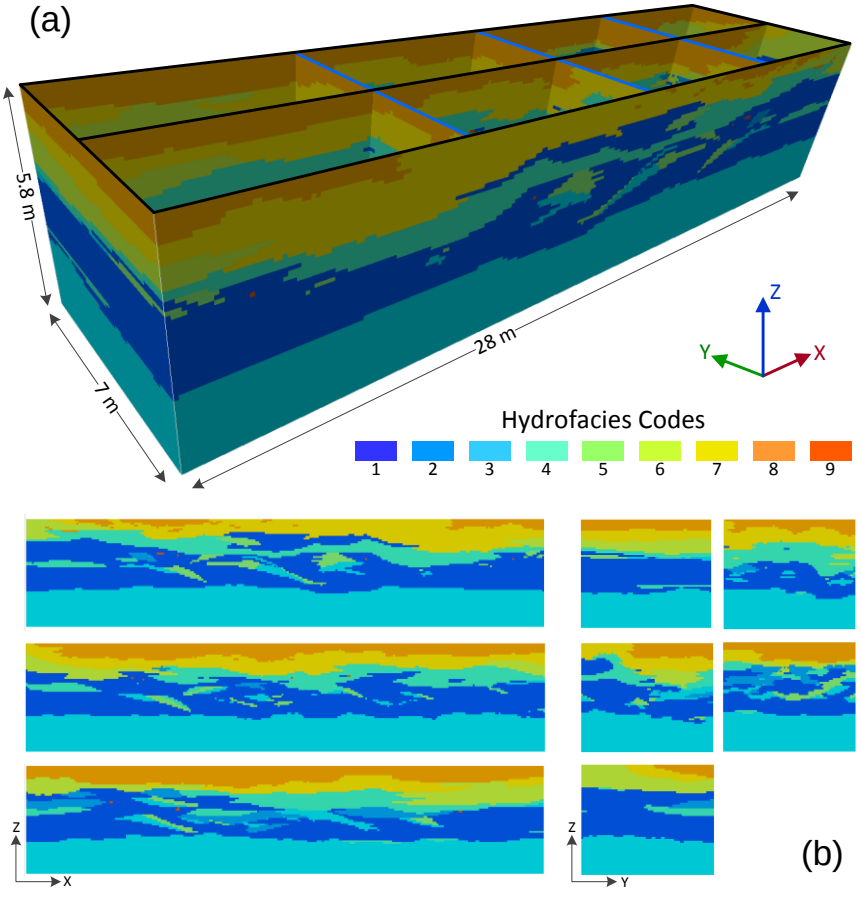

Figure 15. Descalvado aquifer analog dataset (*Bayer et al.*, 2015). (a) 3-D presentation of the informed cross-sections: three sections in xz direction, and five sections in yz direction; (b) 2-D presentation of these informed sections.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

Figure 16 shows realizations obtained by using three different MPS approaches on the basis of the above-mentioned dataset. The white lines indicate the locations of informed sections in each realization. Note that an auxiliary variable along the *z* coordinate is used in s2Dcd and our approach, and the detailed description is given by *Comunian et al.* (2012). To further reveal the performance of the different approaches, we use MDS maps to visualize the dissimilarity of MP histograms (Figure 17), similarly as in section 3.2. However here we use it to reveal the dissimilarity between the sections exacted from the realizations and the informed sections, rather than different 3-D realizations. MDS is very appropriate to present the dissimilarity for this kind of applications shown in this paper because we only have partial cross-sections instead of an entire 3-D training image. Therefore, it is necessary to assess the dissimilarity between the reconstructed sections and informed sections. As shown in Figure 15, the sections from *xz* and *yz* directions are very different, such as the correlation lengths and the complexity of structures. Thus, we draw different MDS maps respectively for the *xz* and *yz* directions (Figures 16a and 16b). Individual sections from the realizations are compared in Figures 16c and 16d. Overall, it can be observed both visually and in the MDS maps that the sections obtained from our approach are closest to the informed sections.

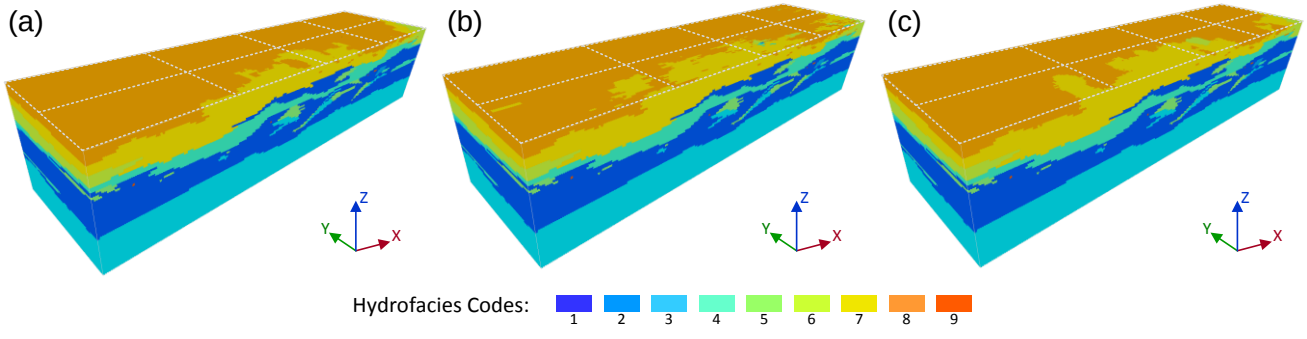

Figure 16. Three realizations using three different MPS approaches: (a) DS, (b) s2Dcd with the coordinate z as auxiliary variable; and (c) our approach with the coordinate z as auxiliary variable.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

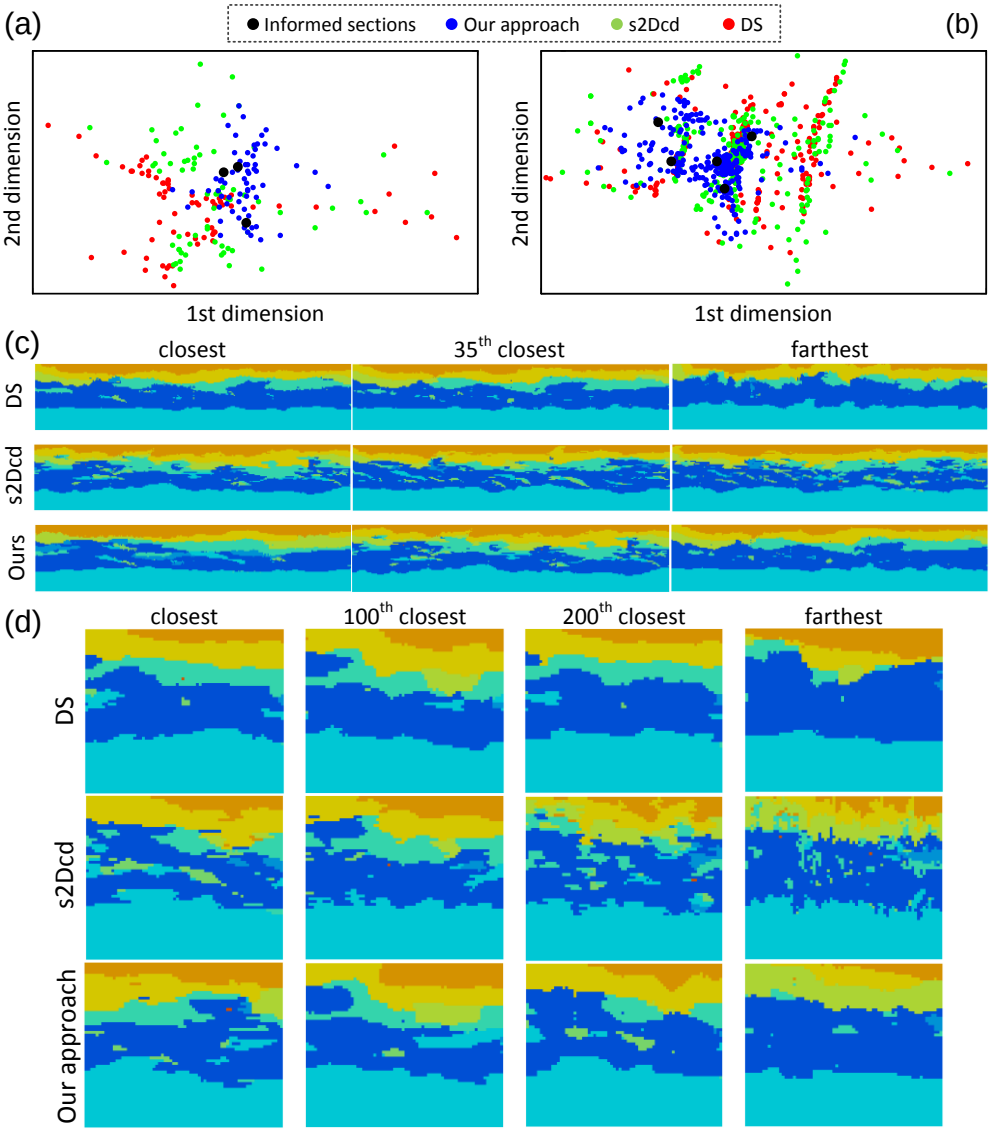

Figure 17. MDS maps of sections extracted from realizations using three different MPS approach. (a) MDS map of 70 sections for each realization along *xz* direction; (b) MDS map of 280 sections for each realization along *yz* direction; (c) selected sections for each method according to the JS divergence in *xz* direction; (d) selected sections in *yz* direction.

Our method is able to reduce the non-stationarity effect of real geological data to a certain extent due to the local search strategy. As shown in the above analysis, the patterns in the informed cross-sections are very complicated where the distribution of hydrofacies is anisotropic and non-stationary, especially for the facies with a lower proportion. As illustrated in Figure 18a, a local domain is surrounded by four segments from the informed cross-sections. It should be noted that there is no facies 2 in all the four segments. We extract the local parts from three realizations by using different MPS approaches. Then we check all the segments of the three local models, and we find that facies 2 is reproduced in this local area in the

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-256
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 14 June 2018

realizations of DS and s2Dcd. Three segments are selected from the three local models, and they are shown in Figure 18b where the boundaries of facies 2 are marked by red lines. Figure 18c shows the histograms of the four informed segments and the three local models. It can be observed that, although there is no facies 2 in the closest four segments, it is still reproduced in this local area by DS and s2Dcd. Conversely, our approach can maintain the distribution of facies well since all the MP statistics are captured from the surrounded sub-sections. This indicates that our approach allows involving the non-stationary geological analogs in the 3-D real applications, and spatial patterns are restricted into a local domain so that they are not carried to faraway locations.

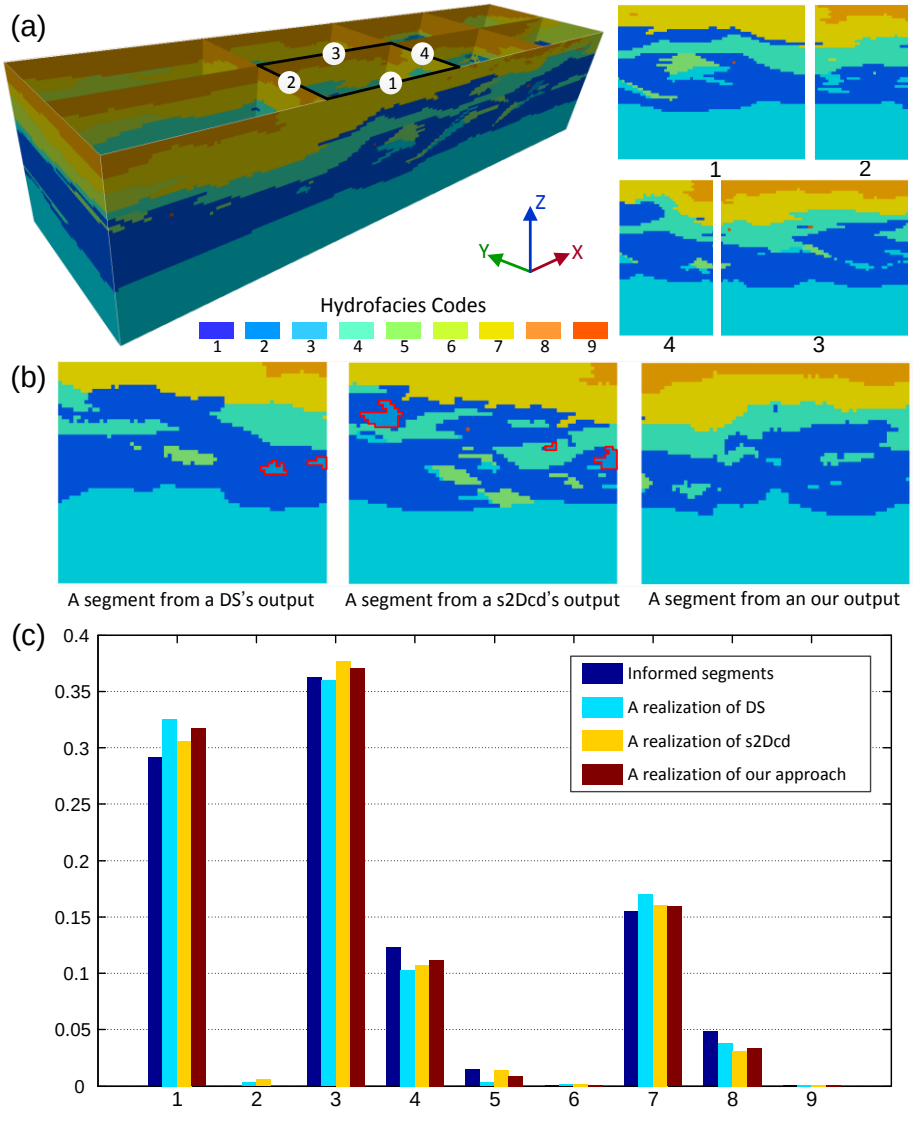

Figure 18. Comparison of reproduction of non-stationary patterns. (a) a local domain and the four corresponding segments; (b) three selected segments from the realizations obtained using different approaches in the local area; (c) histograms of the four informed segments and the three local models exacted from three realizations.

5. Discussion and Conclusion

In this paper, we presented a novel method for reconstructing 3-D complex heterogeneous structures by using partial lower dimensional data. Indeed, this is a very general issue since inferring high-dimensional patterns from low-dimensional data (e.g. boreholes, outcrops and other analogs) is a very common workflow for geologists. In practice, reliable 3-D models of complex geological structures are still difficult to construct due to the heterogeneity of geological phenomena and processes, even though there are many real geological analogs or sections that can be used. Our method makes it possible to reconstruct 3-D structures with MPS when no 3-D training image is available. The synthetic experiments and practical applications presented in this paper demonstrate the capacity to reconstruct such heterogeneous structures.

As compared to the previous MPS implementations that use partial data, the proposed method requires several local training sub-sections surrounding a simulated node, rather than a full section (*Comunian et al.*, 2012) or points in a 3-D domain (*Mariethoz and Renard*, 2010). The local search strategy proposed in this paper allows to compute more reliable MP statistics because it avoids that spatial patterns from faraway locations are considered in the simulation of the current node. In this strategy, the original cross-sections are divided into many sub-sections according to their spatial relationships. Therefore, the non-stationarity of real geological analogs is reduced to a certain extent because the training patterns cannot be borrowed from further than a local subdomain. Of course, besides cross-sections, other scattered samples also can be included as hard data.

Moreover, our approach increases the computational efficiency compared with existing MPS methods. The local search strategy allows acquiring MP statistics from the local sub-sections so that the searches are significantly reduced. Its good computational performance makes it potentially applicable to real 3-D modeling problems such as porous media, hydrofacies, reservoir, and other complex sedimentary structures. In addition, a new parameter, the maximum of matched patterns from each training image is adopted to avoid the unnecessary searches. The experimental results demonstrated that a reasonable choice for this parameter can not only ensure to capture a stable cpdf, but also gain a further performance speed-up.

The method presented here retains many advantages of DS (*Mariethoz et al*., 2010), such as unnecessary storing for MP statistics, pattern distances, flexible neighborhood. Nevertheless, we propose an adaptive and flexible implementation of the search template on multiple grids where the radius of the neighborhood, the distance threshold and the size of data events decrease linearly with the rising of levels of multiple grids. As a result, a big data event is divided into several small parts placed on the different grids, which results in a smaller neighborhood on each grid. An acceptable distance threshold is assigned to the first grid to make it easier to obtain a stable cpdf and to capture the large-scale features from the original sparse samples. For the last grid, the radius of neighborhood is reduced to one and the highest criterion is carried out for the threshold (i.e. $t = 0$) which avoids the small-scale features or lower proportion facies are filtered out. Hence, the simulation of each multi-grid is simulated with different parameters, allowing for flexibility in simulating different structures at different scales.

Another important advantage of our approach is the probability aggregation strategy where the combinations of two different formulas are used to combine the cpdfs from different sub-sections. First, an additive aggregation method, linear pooling formula is used to combine two disjunctive probability distributions from each pair of parallel sub-sections to obtain a more stable pdf. The weights of this step are related to the distances between the current location and the two parallel sub-sections. Such parameterization is able to ensure the pattern trend changing from one sub-section to another one. And then, we aggregate the orthogonal pdfs and prior probability distribution by using a multiplicative method, log-linear pooling formula. This step can enhance the capability for reconstructing connectivity of spatial patterns in comparison with the method using a series of 2-D MPS simulations to fill a 3-D domain along given orthogonal directions (*Comunian et al.*, 2012).

The limitations of our method come from that it is not always possible to obtain abundant sections in each direction, and extremely small local blocks cannot offer enough spatial patterns, thus a minimal sub-section size has to be considered. In addition, our method is not able to perform the simulation of continuous variables. The proposed method can be further improved to overcome these limitations. Another possible direction is to parallelize the proposed MPS implementation and further enhance its computational performance.

*Competing interests***.** The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (U1711267, 41172300) and the National High-tech R&D Program of China (863 Program) (2012AA121401). The authors wish to thank Philippe Renard and Julien Straubhaar for providing the MPS algorithm DeeSse; Moctar Dembele, Min Zeng and Luiz Gustavo Rasera for the fruitful discussions. An executable program of the proposed algorithm is available on the website of the first author (http://www.escience.cn/people/chenqiyu/index.html), and the source code developed by using C++ is available on request from the first author (qiyu.chen@cug.edu.cn).

References

Allard, D., Comunian, A., and Renard, P.: Probability aggregation methods in geoscience, *Mathematical Geosciences*, 44, 545-581, 2012.

Arpat, G. B., and Caers, J.: Conditional simulation with patterns. *Mathematical Geology*, 39(2), 177-203, 2007.

Bayer, P., Comunian, A., Höyng, D., and Mariethoz, G.: High resolution multi-facies realizations of sedimentary reservoir and aquifer analogs, *Scientific Data*, 2, 150033, 2015.

Bayer, P., Huggenberger, P., Renard, P., and Comunian, A.: Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study, *Journal of Hydrology*, 405, 1-9, 2011.

Bordley, R. F.: A multiplicative formula for aggregating probability assessments, *Management Science*, 28, 1137-1148, 1982.

Caers, J.: Geostatistical reservoir modelling using statistical pattern recognition, *Journal of Petroleum Science and Engineering*, 29, 177-188, 2001.

Caers, J.: Modeling uncertainty in the earth sciences. Wiley, Hoboken, 2011.

Chen, Q., Liu, G., Li, X., Zhang, Z., and Li, Y.: A corner-point-grid-based voxelization method for the complex geological structure model with folds, *Journal of Visualization*, 20, 875-888, 2017.

Chugunova, T. L., and Hu, L. Y.: Multiple-point simulations constrained by continuous auxiliary data, *Mathematical Geosciences*, 40, 133-146, 2008.

Comunian, A., Renard, P., and Straubhaar, J.: 3-D multiple-point statistics simulation using 2-D training images, *Computers & Geosciences*, 40, 49-65, 2012.

Comunian, A., Renard, P., Straubhaar, J., and Bayer, P.: Three-dimensional high resolution fluvio-glacial aquifer analog – Part 2: Geostatistical modeling, *Journal of Hydrology*, 405, 10-23, 2011.

Comunian, A., Jha, S. K., Giambastiani, B. M. S., Mariethoz, G., and Kelly, B. F. J.: Training Images from Process-Imitating Methods, *Mathematical Geosciences*, 46, 241-260, 2014.

Caumon, G., Collon-Drouaillet, P., De Veslud, C. L. C., Viseur, S., and Sausse, J.: Surface-based 3-D modeling of geological structures, *Mathematical Geosciences*, 41(8), 927-945, 2009.

Dai, Z., Ritzi, R. W., and Dominic, D. F.: Improving permeability semivariograms with transition probability models of hierarchical sedimentary architecture derived from outcrop analog studies, *Water Resources Research*, 41, 2005.

Deutsch, C. V., and Tran, T. T.: FLUVSIM: a program for object-based stochastic modeling of fluvial depositional systems, *Computers & Geosciences*, 28(4), 525-535, 2002.

de Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., and Violette, S.: Dealing with spatial heterogeneity, *Hydrogeology Journal*, 13, 161-183, 2005.

de Vries, L. M., Carrera, J., Falivene, O., Gratacós, O., and Slooten, L. J.: Application of multiple point geostatistics to non-stationary images, *Mathematical Geosciences*, 41, 29-42, 2009.

Dell Arciprete, D., Bersezio, R., Felletti, F., Giudici, M., Comunian, A., and Renard, P.: Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments, *Hydrogeology Journal*, 20, 299-311, 2012.

Feyen, L., and Caers, J.: Multiple-point geostatistics: a powerful tool to improve groundwater flow and transport predictions in multi-modal formations. In: Renard, P., Demougeot-Renard, H., Froidevaux, R. (Eds.), *GeoENV V: Geostatistics for Environmental Applications*. Springer, Berlin Heidelberg, pp. 197–208, 2004.

Foged, N., Marker, P. A., Christansen, A. V., Bauer-Gottwein, P., Jørgensen, F., Høyer, A.-S., and Auken, E.: Large-scale 3-D modeling by integration of resistivity models and borehole data through inversion, *Hydrology and Earth System Sciences*, 18, 4349-4362, 2014.

Gaud, M. N., Smith, G. A., and McKenna, S. A.: Relating small-scale permeability heterogeneity to lithofacies distribution, In: Bridge, J., and D. W. Hyndman (Eds.), *Aquifer Characterization*, vol. 80. SEPM, Special Publication, pp. 55-66, 2004.

Genest, C., and Zidek, J. V.: Combining probability distributions: A critique and an annotated bibliography, *Statistical Science*, 114-135, 1986.

Guardiano, F. B., and Srivastava, R. M.: Multivariate geostatistics: beyond bivariate moments. In *Geostatistics Troia'92*. Springer Netherlands, pp. 133-144, 1993.

Gueting, N., Caers, J., Comunian, A., Vanderborght, J., and Englert, A.: Reconstruction of three-dimensional aquifer heterogeneity from two-dimensional geophysical data. *Mathematical Geosciences*, 50(1), 53-75, 2017.

Hajizadeh, A., Safekordi, A., and Farhadpour, F. A.: A multiple-point statistics algorithm for 3-D pore space reconstruction from 2-D images, *Advances in Water Resources*, 34, 1256-1267, 2011.

He, X. L., Sonnenborg, T. O., Jørgensen, F., and Jensen, K. H.: The effect of training image and secondary data integration with multiple-point geostatistics in groundwater modelling, *Hydrology and Earth System Sciences*, 18, 2943-2954, 2014.

Heinz, J., Kleineidam, S., Teutsch, G., and Aigner, T.: Heterogeneity patterns of Quaternary glaciofluvial gravel bodies (SW-Germany): application to hydrogeology. *Sedimentary geology*, 158(1), 1-23, 2003.

Hermans, T., Nguyen, F., and Caers, J.: Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: Workflow and case study, *Water Resources Research*, 51, 5332-5352, 2015.

Hoffman, B. T., and Caers, J.: History matching by jointly perturbing local facies proportions and their spatial distribution: Application to a North Sea reservoir, *Journal of Petroleum Science and Engineering*, 57(3), 257-272, 2007.

Honarkhah, M., and Caers, J.: Stochastic simulation of patterns using distance-based pattern modeling, *Mathematical Geosciences*, 42, 487-517, 2010.

Høyer, A.-S., Vignoli, G., Hansen, T. M., Vu, L. T., Keefer, D. A., and Jørgensen, F.: Multiple-point statistical simulation for hydrogeological models: 3-D training image development and conditioning strategies, *Hydrology and Earth System Sciences*, 21, 6069-6089, 2017.

Hu, L. Y., and Chugunova, T.: Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review, *Water Resources Research*, 44, 2008.

Hu, R., Brauchler, R., Herold, M., and Bayer, P.: Hydraulic tomography analog outcrop study: Combining travel time and steady shape inversion, *Journal of Hydrology*, 409, 350-362, 2011.

Huysmans, M., Orban, P., Cochet, E., Possemiers, M., Ronchi, B., Lauriks, K., Batelaan, O., and Dassargues, A.: Using multiple-point geostatistics for tracer test modeling in a clay-drape environment with spatially variable conductivity and sorption coefficient, *Mathematical Geosciences*, 46, 519-537, 2014.

Jackson, M. D., Percival, J. R. , Mostaghiml, P., Tollit, B. S., Pavlidis, D., Pain, C. C., Gomes, J. L. M. A., El-Sheikh, A. H., Salinas, P., Muggeridge, A. H., and Blunt, M. J.: Reservoir modeling for flow simulation by use of surfaces, adaptive unstructured meshes, and an overlapping-control-volume finite-element method, *SPE Reservoir Evaluation & Engineering*, 18, 115-132, 2015.

Jha, S. K., Comunian, A., Mariethoz, G., and Kelly, B. F. J.: Parameterization of training images for aquifer 3-D facies modeling integrating geological interpretations and statistical inference, *Water Resources Research*, 50, 7731-7749, 2014.

Journel, A. G.: Geostatistics: roadblocks and challenges, In *Geostatistics Troia'92*. Springer Netherlands, pp. 213-224, 1993.

Journel, A. G.: Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses, *Mathematical Geology*, 34, 573-596, 2002.

Kessler, T. C., Comunian, A., Oriani, F., Renard, P., Nilsson, B., Klint, K. E., and Bjerg, P. L.: Modeling Fine-Scale Geological Heterogeneity-Examples of Sand Lenses in Tills, *Ground Water*, 51, 692-705, 2013.

Klise, K. A., Weissmann, G. S., McKenna, S. A., Nichols, E. M., Frechette, J. D., Wawrzyniec, T. F., and Tidwell, V. C.: Exploring solute transport and streamline connectivity using lidar‐based outcrop images and geostatistical representations of heterogeneity, *Water Resources Research*, 45(5), 2009.

Knudby, C., and Carrera, J.: On the relationship between indicators of geostatistical, flow and transport connectivity. *Advances in Water Resources*, 28(4), 405-421, 2005.

Krishnan, S.: The Tau Model for Data Redundancy and Information Combination in Earth Sciences: Theory and Application, *Mathematical Geosciences*, 40, 705-727, 2008.

Lee, S. Y., Carle, S. F., and Fogg, G. E.: Geologic heterogeneity and a comparison of two geostatistical models: Sequential Gaussian and transition probability-based geostatistical simulation, *Advances in water resources*, 30(9), 1914-1932, 2007.

Li, X., Mariethoz, G., Lu, D., and Linde, N.: Patch-based iterative conditional geostatistical simulation using graph cuts, *Water Resources Research*, 52, 6297-6320, 2016.

Maharaja, A.: TiGenerator: Object-based training image generator, *Computers & Geosciences*, 34, 1753-1761, 2008.

Mahmud, K., Mariethoz, G., Baker, A., and Sharma, A.: Integrating multiple scales of hydraulic conductivity measurements in training image-based stochastic models, *Water Resources Research*, 51, 465-480, 2015.

Mahmud, K., Mariethoz, G., Caers, J., Tahmasebi, P., and Baker, A.: Simulation of Earth textures by conditional image quilting, *Water Resources Research*, 50, 3088-3107, 2014.

Mariethoz, G., and Kelly, B. F. J.: Modeling complex geological structures with elementary training images and transform-invariant distances, *Water Resources Research*, 47(7), 2011.

Mariethoz, G., and Renard, P.: Reconstruction of incomplete data sets or images using direct sampling, *Mathematical Geosciences*, 42, 245-268, 2010.

Mariethoz, G., Straubhaar, J., Renard, P., Chugunova, T., and Biver, P.: Constraining distance-based multipoint simulations to proportions and trends, *Environmental Modelling & Software*, 72, 184-197, 2015.

Mariethoz, G., Renard, P., and Straubhaar, J.: The Direct Sampling method to perform multiple-point geostatistical simulations, *Water Resources Research*, 46(11), 2010.

Mariethoz, G., Renard, P., and Froidevaux, R.: Integrating collocated auxiliary parameters in geostatistical simulations using joint probability distributions and probability aggregation, *Water Resources Research*, 45(8), 2009.

Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., and Renard, P.: A practical guide to performing multiple-point statistical simulations with the Direct Sampling algorithm. *Computers & Geosciences*, 52, 307-324, 2013.

Nichols, E. M., Weissmann, G. S., Wawrzyniec, T. F., Frechette, J. D., and Klise, K. A.: Processing of outcrop-based lidar imagery to characterize heterogeneity for groundwater models, *SEPM concepts in sedimentology and paleontology*, 10, 239-47, 2011.

Okabe, H., and Blunt, M.J.: Pore space reconstruction using multiple-point statistics, *Journal of Petroleum Science and Engineering*, 46, 121-137, 2005.

Okabe, H., and Blunt, M. J.: Pore space reconstruction of vuggy carbonates using microtomography and multiple-point statistics, *Water Resources Research*, 43(12), 2007.

Oriani, F., Straubhaar, J., Renard, P., and Mariethoz, G.: Simulation of rainfall time series from different climatic regions using the direct sampling technique, *Hydrology and Earth System Sciences*, 18, 3015-3031, 2014.

Phelps, G., and Boucher, A.: Mapping locally complex geologic units in three dimensions: the multi-point geostatistical approach, *Three-Dimensional Geological Mapping*, 36-39, 2009.

Pickel, A., Frechette, J. D., Comunian, A., and Weissmann, G. S.: Building a training image with Digital Outcrop Models, *Journal of Hydrology*, 531, 53-61, 2015.

Pirot, G., Straubhaar, J., and Renard, P.: A pseudo genetic model of coarse braided-river deposits, *Water Resources Research*, 51, 9595-9611, 2015.

Pyrcz, M. J., Boisvert, J. B., and Deutsch, C. V.: ALLUVSIM: A program for event-based stochastic modeling of fluvial depositional systems, *Computers & Geosciences*, 35(8), 1671-1685, 2009.

Pyrcz, M. J., and Deutsch, C. V.: Geostatistical reservoir modeling. Oxford university press, 2014.

Raiber, M., White, P. A., Daughney, C. J., Tschritter, C., Davidson, P., and Bainbridge, S. E.: Three-dimensional geological modelling and multivariate statistical analysis of water chemistry data to analyse and visualise aquifer structure and groundwater composition in the Wairau Plain, Marlborough District, New Zealand, *Journal of Hydrology*, 436-437, 13-34, 2012.

Renard, P., and Allard, D.: Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*, 51, 168-196, 2013.

Ritzi, R. W.: Behavior of indicator variograms and transition probabilities in relation to the variance in lengths of hydrofacies, *Water resources research*, 36(11), 3375-3381, 2000.

Stone, M.: The opinion pool. *The Annals of Mathematical Statistics*, 32(4), 1339-1342, 1961.

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O.: An improved parallel multiple-point algorithm using a list approach, *Mathematical Geosciences*, 43(3), 305-328, 2011.

Strebelle, S.: Conditional simulation of complex geological structures using multiple-point statistics, *Mathematical Geology*, 34, 1-21, 2002.

Tahmasebi, P., Hezarkhani, A., and Sahimi, M.: Multiple-point geostatistical modeling based on the cross-correlation functions, *Computational Geosciences*, 16, 779-797, 2012.

Tahmasebi, P., Sahimi, M., and Caers, J.: MS-CCSIM: Accelerating pattern-based geostatistical simulation of categorical variables using a multi-scale search in Fourier space, *Computers & Geosciences*, 67, 75-88, 2014.

Tran, T. T.: Improving variogram reproduction on dense simulation grids. *Computers & Geosciences*, 20(7-8), 1161-1168, 1994.

Vassena, C., Cattaneo, L., and Giudici, M.: Assessment of the role of facies heterogeneity at the fine scale by numerical transport experiments and connectivity indicators, *Hydrogeology Journal*, 18(3), 651-668, 2010.

Wambeke, T., and Benndorf, J.: An integrated approach to simulate and validate orebody realizations with complex trends: A case study in heavy mineral sands, *Mathematical Geosciences*, 48, 767-789, 2016.

Weissmann, G. S., Carle, S. F., and Fogg, G. E.: Three-dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics, *Water Resources Research*, 35(6), 1761-1770, 1999.

Weissmann, G. S., Pickel, A., McNamara, K. C., Frechette, J. D., Kalinovich, I., Allen-King, R. M., and Jankovic, I.: Characterization and quantification of aquifer heterogeneity using outcrop analogs at the Canadian Forces Base Borden, Ontario, Canada, *Geological Society of America Bulletin*, 127(7-8), 1021-1035, 2015.

Wu, J., Boucher, A., and Zhang, T.: A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM, *Computers & Geosciences*, 34, 1863-1876, 2008.

Wu, K., Van Dijke, M. I. J., Couples, G. D., Jiang, Z., Ma, J., Sorbie, K. S., Crawford, J., Young, I., and Zhang, X.: 3-D stochastic modelling of heterogeneous porous media – Applications to reservoir rocks, *Transport in Porous Media*, 65, 443-467, 2006.

Yang, L., Hou, W., Cui, C., and Cui, J.: GOSIM: A multi-scale iterative multiple-point statistics algorithm with global optimization, *Computers & Geosciences*, 89, 57-70, 2016.

Zappa, G., Bersezio, R., Felletti, F., and Giudici, M. Modeling heterogeneity of gravel-sand, braided stream, alluvial aquifers at the facies scale, *Journal of Hydrology*, 325, 134-153, 2006.

Zhang, T., Li, D., Lu, D., and Yang, J.: Research on the reconstruction method of porous media using multiple-point geostatistics, *Science China Physics, Mechanics and Astronomy*, 53, 122-134, 2010.

Zhang, T., Switzer, P., and Journel, A.: Filter-based classification of training image patterns for spatial simulation, *Mathematical Geology*, 38, 63-80, 2006.