Reconstruction of Incomplete Data Sets or Images Using Direct Sampling

Gregoire Mariethoz · Philippe Renard

Abstract With increasingly sophisticated acquisition methods, the amount of data available for mapping physical parameters in the geosciences is becoming enormous. If the density of measurements is sufficient, significant non-parametric spatial statistics can be derived from the data. In this context, we propose to use and adapt the Direct Sampling multiple-points simulation method (DS) for the reconstruction of partially informed images. The advantage of the proposed method is that it can accommodate any data disposition and that it can indifferently deal with continuous and categorical variables. The spatial patterns found in the data are mimicked without model inference. Therefore, very few assumptions are required to define the spatial structure of the reconstructed fields, and very limited parameterization is needed to make the proposed approach extremely simple from a user perspective. The different examples shown in this paper give appealing results for the reconstruction of complex 3D geometries from relatively small data sets.

Keywords Geostatistics · Simulation · Multiple-point · Multipoint · Reconstruction · Completion · Partial images · Direct sampling · DS

1 Introduction

Missing or partial information on parameter fields is a ubiquitous problem in Earth Sciences and is the major cause of (hydro)geological uncertainty. This is not likely to change since there will always be an unsampled volume between wells or outcrops where no direct information is available. Geological heterogeneity only makes

G. Mariethoz (✉) · P. Renard
Centre for Hydrogeology, University of Neuchâtel, 11 Rue Emile Argand, CP 158, CH-2000 Neuchâtel, Switzerland
e-mail: gregoire.mariethoz@unine.ch

Present address:
G. Mariethoz
ERE Department, Stanford University, 367 Panama st, rm 65, Stanford, CA 94305, USA

things worse because characterizing it calls for more data (Renard et al. 2005). A new method for completion (reconstruction) of such sparsely informed images is presented in this paper. The focus is on its application to the reconstruction of geological structures. Kriging has often been used for reconstruction problems, like filling the missing zones of geophysical surveys (Cornacchiulo and Bagtzoglou 2004), reconstructing time series of aquifer piezometric levels (Kumar and Ahmed 2008) or for rain field reconstruction (Fiorucci et al. 2001). Several studies (Guadagnini et al. 2004; dell'Arciprete et al. 2008) use geostatistical simulations to reconstruct geological facies. Other applications include the reconstruction of rock fracture surface (Marache et al. 2002) and the generation of 3D permeability blocks from 2D thin sections in order to evaluate the permeability of samples (Youngseuk et al. 2004).

The statistical properties of natural phenomena are often best reproduced by nonparametric approaches (Mariethoz et al. 2009b). Since they use non-parametric high-order statistics, multiple-points (MP) simulation methods (Guardiano and Srivastava 1993; Strebelle 2002; Zhang et al. 2006; Arpat and Caers 2007) have a high potential for reproducing complex spatial structures present in the data. Reconstructions of geological structures using multiple-points simulations have been performed by Wu et al. (2008) using the FILTERSIM algorithm (Zhang et al. 2006), Hurley and Zhang (2009) for borehole imagery, and Okabe and Blunt (2004) to generate small scale 3D images of porous media. An important practical limitation of MP techniques is that finding an appropriate training image (TI) is not straightforward. Therefore it is common practice to build a TI using other geostatistical methods, such as Boolean simulations (Lantuéjoul 2002), sequential indicator simulations (Journel and Isaaks 1984; Goovaerts 1997), or a mix of different techniques (Journel and Zhang 2006). This practice, although justified, reduces the use of multiple-points methods to a mere way of conditioning simulations coming from other methods. Hence, most of the possibilities offered by MP to integrate a non-parametric spatial model are hindered because the high-order multiple-points model is fed by parametric, low-order statistics.

We present a reconstruction approach that uses MP statistics but does not require a TI when the amount of data is sufficiently large. The idea is to extract MP statistics directly from the data set, which becomes a training data set (TD). We show that the inference of consistent MP statistics is possible even when the TD amounts to less than 1% of a 3D volume to reconstruct. Moreover, examples show that hypothesis of stationarity can be somewhat relaxed if enough information on non-stationarity is contained in the data. While traditional MP simulation techniques are well suited for inferring multiple-points statistics from a densely and regularly informed domain, they cannot fully exploit the information contained in sparse TDs due to the existence of fixed neighborhoods (search templates) and to the necessity of multiple-grids. Instead, we propose to use the technique of multiple-points simulations by direct sampling (DS). DS is ideal to fully exploit the information present in sparse TDs because it accommodates any data disposition and does not need multiple-grids. In addition, it is able to handle both categorical and continuous variables, thus allowing to deal with a wide range of different applications. Note that FILTERSIM can also handle both categorical and continuous variables, but like other MP techniques it uses multiple-grids and relies on the storage of patterns of a fixed size. The DS method, as well as its performances and the sensitivity to its various parameters, is described in detail in Mariethoz (2009) and Mariethoz et al. (2010).

As indicated by Ortiz and Deutsch (2004), the model information contained in the data is based on the true field, often more reliable than a training image derived from interpretation (provided that there are no large errors in the data). With a data-driven approach, the model choice depends on the available data only and there is little room for model uncertainty (Kitanidis 1986), even if uncertainty remains on the data measurements and their statistical significance. Although we investigate cases where MP statistics are inferred from relatively small data sets, we do not systematically explore the border between problems where the TD is sufficient and problems where an external TI should be used. The first part of the paper presents a summary of the DS algorithm, which is the main tool for solving the reconstruction problem, and how it is used for the reconstruction of partial data sets. The rest of the paper is devoted to the presentation of examples, because we think that it is the best way to grasp the principle, the practical advantages and limitations of this reconstruction method. Five reconstruction examples are presented. The first one uses a simple 2D case to illustrate the algorithm behavior with different TDs. The second example reconstructs a satellite photograph in both categorical and continuous variables. The third example validates the method by performing a reconstruction based on a TD extracted from a 3D synthetic reservoir, therefore allowing comparison between the synthetic reference and the reconstruction. The fourth reconstruction example is based on a real 3D data set made of a series of profiles digitized on a quarry front in Italy. The last example shows the application of the reconstruction method to borehole imaging data.

2 The Reconstruction of Partial Images Using Direct Sampling

Geostatistical simulation algorithms are aimed at producing realizations of a spatially correlated variable Z at all N locations $\mathbf{x}_i$ of a regular grid, with $i = [1, \ldots, N]$. Each such realization is a sample of the N-dimensional joint cumulative distribution function (cdf)

$$F(z, \mathbf{x}) = \text{Prob}\{Z(\mathbf{x}_1) \leq z, Z(\mathbf{x}_2) \leq z, \ldots, Z(\mathbf{x}_N) \leq z\}, \tag{1}$$

for the case of a continuous variable. Sequential simulation algorithms (Deutsch and Journel 1992) are a practical way of sampling (1) by performing the following decomposition

$$F(z, \mathbf{x}) = \text{Prob}\{Z(\mathbf{x}_1) \leq z\} \cdot \text{Prob}\{Z(\mathbf{x}_2) \leq z | z(\mathbf{x}_1)\}$$
$$\cdot \cdots \cdot \text{Prob}\{Z(\mathbf{x}_N) \leq z | z(\mathbf{x}_1), \ldots, z(\mathbf{x}_{N-1})\}. \tag{2}$$

The conditional cumulative distribution function (ccdf) of Z at each location depends of the ensemble of all previously determined Z locations. Note that a similar decomposition can be accomplished for discrete variables. Sequential simulations usually proceed by considering only a limited neighborhood (i.e. a subset of locations) of size n for determining local ccdfs, with $n \ll N$ to limit computational burden. At each location $\mathbf{x}$, the lag vectors $\mathbf{L} = \{\mathbf{h}_1, \ldots, \mathbf{h}_n\} = \{\mathbf{x}_1 - \mathbf{x}, \ldots, \mathbf{x}_n - \mathbf{x}\}$ define the position of the informed neighboring locations. The data event at location $\mathbf{x}$ is defined by $\mathbf{d}_n(\mathbf{x}, \mathbf{L}) = \{Z(\mathbf{x} + \mathbf{h}_1), \ldots, Z(\mathbf{x} + \mathbf{h}_n)\}$, the combination of the lag

vectors and the value at the neighboring locations. Given this data event, sequential simulation techniques estimate at each successive location the ccdf for the variable of interest Z

$$F(z, \mathbf{x}, \mathbf{d}_n) = \mathrm{Prob}\{Z(\mathbf{x}) \leq z | \mathbf{d}_n(\mathbf{x}, \mathbf{L})\}. \tag{3}$$

Once the ccdf of (3) is determined, a value for $Z(\mathbf{x})$ is drawn from it, and is thereafter considered as conditioning data when simulating the remaining nodes. The DS algorithm (Mariethoz et al. 2010) is a geostatistical sequential simulation algorithm using MP statistics, i.e. the ccdf of (3) is derived from a training image (TI) representing the desired spatial structure of the variable of interest. The particularity of DS lies in the fact that it does not require to estimate the ccdf (3). Consequently, there is no need to store the probabilities associated with all pixels configurations found in the TI. Other algorithms have to store them either in a tree structure (Strebelle 2002) or in lists (Straubhaar et al. 2008). Instead, at each simulated node $\mathbf{x}$, the TI is directly and randomly sampled until the appropriate pattern is found. For each of the successive samples, the distance or mismatch, $d\{\mathbf{d}_n(\mathbf{x}, \mathbf{L}), \mathbf{d}_n(\mathbf{y}, \mathbf{L})\}$, between the data event observed in the simulation $\mathbf{d}_n(\mathbf{x}, \mathbf{L})$ and the one sampled from the training image $\mathbf{d}_n(\mathbf{y}, \mathbf{L})$, is calculated ($\mathbf{y}$ denotes the nodes of the TI). If there is no mismatch or if the mismatch is lower than a given threshold, the sampling process is stopped and the value at the central node of the data event found in the TI $Z(\mathbf{y})$ is directly pasted in the simulation grid (SG) at the location $\mathbf{x}$. It is known (Shannon 1948) that pasting a randomly sampled matching value is strictly equivalent to computing the ccdf based on the entire image and then drawing a value from the ccdf. The first matching value is a sample of (3). It is not the most likely value (which is used in the SIMPAT algorithm of Arpat and Caers 2007).

Like in the first implementation of multiple-points simulations by Guardiano and Srivastava (1993), the DS scans the TI with a data event that can have any shape. However, the original algorithm is aimed at computing the ccdf of (3) for each configuration and therefore it requires to scan the entire TI for each simulated node. Here, the aim is only to scan a small part of the TI to find one single acceptable event. This seemingly slight difference between algorithms has large implications. First, DS does not need an explicit formulation of the probability distribution of (3), and therefore it does not need to store it. Values are drawn from the TI, whereas all existing sequential simulation methods draw values from a distribution. Although equivalent, the concept is radically different. Second, the DS algorithm is orders of magnitude faster than the method of Guardiano and Srivastava (1993) because a matching data event is often found after scanning a small fraction of the TI. As far as it is possible to benchmark algorithms with different input parameters, our experience is that DS can be as fast as tree-based multiple-points simulation codes (e.g. Remy et al. 2009) (i.e. it takes similar time to obtain simulations of similar quality). Moreover, DS is easy to parallelize (Mariethoz 2010; Mariethoz et al. 2010). Because the multiple-points statistics of the TI are not stored, there is no limit on the size of the neighborhood that can be used. Moreover, the neighboring nodes do not need to be located at fixed positions relatively to the central node, whereas this set of positions, usually referred to as template, is necessary for traditional MP methods. The data event can have any geometry and can change

for each simulated node. Therefore, we define the neighborhood of a simulated node as the n closest previously simulated or conditioning nodes within a search radius l, where n and l are neighborhood parameters. This flexibility is a major advantage when neighborhoods are sparse or incomplete, which is precisely the case in the problem of reconstructing partial images. Moreover, this definition of neighborhoods is very convenient as it allows the mean radius of the data events to decrease linearly as the density of informed grid nodes becomes higher. This natural variation of the data events' size has the same effect as multiple-grids (Strebelle 2002), and makes their use unnecessary.

Using different measures of distance also offers a high degree of flexibility. It is therefore possible to accommodate categorical as well as continuous variables. For categorical variables, the distance between a data event found in the simulation and another one found in the TI, $d\{\mathbf{d}_n(\mathbf{x}, \mathbf{L}), \mathbf{d}_n(\mathbf{y}, \mathbf{L})\}$, is given by the proportion of non-matching nodes. It is calculated by using the indicator variable a that equals 0 if two nodes have identical values and 1 otherwise

$$d\{\mathbf{d}_n(\mathbf{x}), \mathbf{d}_n(\mathbf{y})\} = \frac{1}{n} \sum_{i=1}^{n} a_i \in [0, 1], \quad \text{where } a_i = \begin{cases} 0 & \text{if } Z(\mathbf{x}_i) = Z(\mathbf{y}_i), \\ 1 & \text{if } Z(\mathbf{x}_i) \neq Z(\mathbf{y}_i). \end{cases} \quad (4)$$

A measure of distance allowing to accommodate continuous variables is the normalized pair-wise Manhattan distance

$$d\{\mathbf{d}_n(\mathbf{x}), \mathbf{d}_n(\mathbf{y})\} = \frac{1}{n} \sum_{i=1}^{n} \frac{|Z(\mathbf{x}_i) - Z(\mathbf{y}_i)|}{\text{argmax}(Z^{ti}) - \text{argmin}(Z^{ti})} \in [0, 1]. \quad (5)$$

The normalizing factor is the largest interval within the Z values found in the TI, and it is used to ensure that the distance values are comprised in the interval [0,1]. Without this normalization, the choice of a threshold value would be more difficult. For example, numerical tests have shown that 0.05 is a low threshold and 0.5 a high threshold. Such guidelines can be more difficult to establish without normalization. Mariethoz et al. (2010) propose several alternative types of distances, some of them including a term to weight the nodes of the data event according to their position relative to the central node. In the applications presented in this paper, we only use the distances (4) and (5), and we do not include the weighting term.

Simulations resulting from DS reproduce the spatial properties of the scanned image. Even complex properties are reproduced because the method allows exploiting high-order statistics from a data set. Since DS does not need multiple-grids for reproducing structures at different scales, it does not encounter problems for the relocation of conditioning data on different simulation grids (Saripally 2008), making it easy to integrate large data sets. As the reconstruction problem relies on large data sets, this feature is a major advantage. The core of the DS algorithm remains unchanged if a data set has to be scanned instead of a TI. This allows simulating unknown nodes by scanning the informed ones only. The goal is for the unknown zones of the SG to be filled with patterns coming from the TD. If a large diversity of patterns is present in the TD, this diversity can be reproduced in the unknown zones of the SG. The DS algorithm described in Mariethoz et al. (2010) has to be slightly modified in order to

scan the TD instead of a TI. The main change is that **y** does not designate any more a location in the TI, but a location in the TD, once it has been migrated on the SG (for more clarity, the ensemble of the SG nodes where TD points have been migrated from here on is designated TD). The resulting algorithm is summarized in the following steps

(1) Assign all points of the TD to the closest grid nodes in the SG. When several conditioning data should be assigned to the same grid node, we assign the closest one to the grid node. The other alternatives often used in geostatistics are to assign the most frequent one, or to sample randomly one of these values for each simulation. If the support size of the data does not correspond to the size of the SG nodes, upscaling methods may also be used here.
(2) Define a path through the remaining nodes of the SG.
(3) For each successive location **x** in the path:
  (a) Find the neighbors of **x**. They consist of a maximum of the n closest grid nodes $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ that were already assigned or simulated in the SG and located within a search radius l. Compute the lag vectors $\mathbf{L} = \{\mathbf{h}_1, \ldots, \mathbf{h}_n\} = \{\mathbf{x}_1 - \mathbf{x}, .., \mathbf{x}_n - \mathbf{x}\}$ defining the neighborhood of **x**.
  (b) Define the data event $\mathbf{d}_n(\mathbf{x}, \mathbf{L}) = \{Z(\mathbf{x} + \mathbf{h}_1), \ldots, Z(\mathbf{x} + \mathbf{h}_n)\}$. It is a vector containing the values of the variable of interest at all the nodes of the neighborhood.
  (c) Randomly draw a location **y** in the TD and from that location scan systematically the TD. For each location **y**:
    (i) Find the data event $\mathbf{d}_n(\mathbf{y}, \mathbf{L})$.
    (ii) Compute the distance $d\{\mathbf{d}_n(\mathbf{x}, \mathbf{L}), \mathbf{d}_n(\mathbf{y}, \mathbf{L})\}$ between both data events. If the lag vectors point to nodes of the SG that are not informed by the TD or that are outside the SG, these nodes are counted as false ($a_i = 1$) if the variable to reconstruct is categorical (distance 4). In case of continuous variable (distance 5), they are given the maximum possible distance, which is the largest interval within the Z values found in the TD.
    (iii) If it is the lowest distance obtained so far in the scanning of the TD, store **y**, $Z(\mathbf{y})$ and $d\{\mathbf{d}_n(\mathbf{x}, \mathbf{L}), \mathbf{d}_n(\mathbf{y}, \mathbf{L})\}$.
    (iv) If $d\{\mathbf{d}_n(\mathbf{x}, \mathbf{L}), \mathbf{d}_n(\mathbf{y}, \mathbf{L})\}$ is smaller than the acceptance threshold t, the value $Z(\mathbf{y})$ is assigned to $Z(\mathbf{x})$. The algorithm proceeds then to Step 3 in order to simulate another node of the SG.
    (v) If no satisfying distance under the threshold has been found in the TD, the node **y** with the lowest distance is accepted and its value $Z(\mathbf{y})$ is assigned to $Z(\mathbf{x})$.

The data event $\mathbf{d}_n(\mathbf{x}, \mathbf{L})$, including its shape given by the lag vectors, may not correspond exactly to any informed pattern in the TD. In those cases, the distance measure favors the selection of a data event in the TD that maximizes the number of correctly informed nodes. This is done by attributing the maximum distance to the non-informed nodes in the data event (point 3.c.ii). This is equivalent to reducing the order of the statistics if no suitable data event is found in the TD. In the worst case, there is no location in the TD corresponding to the searched lags vectors, and the marginal probabilities are then sampled. There are only a limited number of parameters governing the algorithm. The main ones are the maximum number of neighbors

n and the threshold t. Setting n to a large value (between 30 and 50 nodes) and t to a very low value (typically 0) guarantees to yield the best possible results DS can provide, sometimes at a high cost in terms of CPU time. Therefore, we used $t = 0$ for all the examples presented below, except the continuous-variable case (because a perfect match can never be found between continuous-variable data events and it makes no sense to set $t = 0$ in this case). It should be noted that $t = 0$ does not mean an absence of variability between realizations. Rather, it is a strong constraint on the reproduction of the statistics of the TD.

3 Reconstruction Examples

3.1 Spatial Repartition of the TD

The first example illustrates how the spatial scattering of the TD influences the reconstruction results. Figure 1a shows a categorical image, representing sand channels in a clay matrix (Strebelle 2002). The central zone (36% of the entire image) is considered unknown and it is reconstructed by scanning the edges of the image with a neighborhood constituted of the 40 closest nodes. The meaning of this parameter is that for any simulated pixel, the pattern formed by the 40 closest neighbors is considered. Starting from a random location, the training image is scanned until encountering a pixel whose neighborhood matches the desired pattern. The value of this pixel is then assigned to the simulated location. The method reproduces the statistics of the training image up to the 40th order (Shannon 1948). Three realizations of reconstruction are shown in Figs. 1b to d, generated using the algorithm described above. Distance (4) is used in the DS simulation because the variable of interest is categorical. The resulting realizations have the features one would expect when looking at the original training image. This reconstruction is possible because the training data set contains information at all scales, such as many small portions of channels on the East and West sides of the image and large connected features in the North and South. This example is based on a data set that is spatially incomplete, but informatively complete in terms of scale. The presence of data on large and continuous zones results in the presence of small and large scale patterns. This diversity of patterns present in the TD extensively feeds the MP model. Nevertheless, in real cases adjacent grid nodes may not be informed. With dispersed data, small scale patterns have less chance to occur in the TD, tending to produce a degenerated MP model.

Figure 2 shows the results of reconstruction using three different sparse TDs coming from the image of Fig. 1a. The first TD consists of 5,000 nodes taken at random locations (8% of a total of 62,500 nodes in the image grid). This is displayed in Fig. 2a, with the corresponding reconstruction in Fig. 2b. Figures 2c and e show larger TDs of respectively 10,000 and 20,000 informed nodes (16% and 31%), and their corresponding reconstructions in Figs. 2d and f. The random position of the informed nodes explains the relatively low quality of the reconstructed images for the TD of 5,000 informed nodes (Fig. 2b), where small scale patterns are virtually absent (see zoomed part of Fig. 2a). Gómez-Hernández and Wen (1998) show that using point data only, it is impossible to know whether the high values of hydraulic conductivity
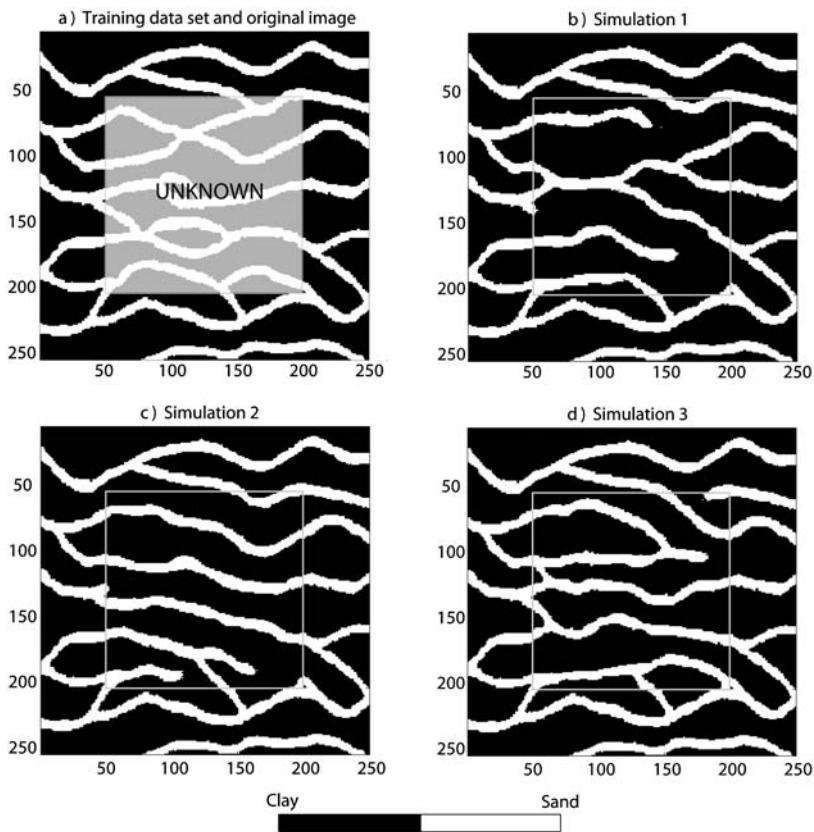
Fig. 1 The reconstruction of a simple categorical image with a hole (36% of the entire image). (**a**) The original complete image (Strebelle 2002) and the unknown zone to reconstruct. (**b**) to (**d**) Three realizations of reconstruction of the central zone

are connected or not. Here the reconstruction method does not make any assumption on the spatial continuity of the variable to simulate. If this spatial continuity is not represented in the TD, it will be absent in the reconstruction. Hence, the resulting reconstructions are very affected by the lack of information at small scale. If the TDs are large and contain information at many scales, DS reconstructions will perform better. In order to capture all scales, an optimum sampling strategy would consist of densely measured regions and sparse measurements in other areas (this is valid for all simulation/estimation methods). Figures 2d and f show that an important fraction of informed nodes contains enough information to accurately reconstruct the entire image, because such a dense data set also includes small scale patterns (zoomed parts of Figs. 2c and e). For geological applications, outcrops and quarry fronts can provide dense and multi-scale information on 3-dimensional geological structures (this is illustrated further in the paper). To a lesser extent, well logs also provide small scale patterns in the vertical direction. Remote sensing applications are also numerous, often involving problems similar to the setting of Fig. 1a, where one well-defined zone is unknown in an otherwise complete image.
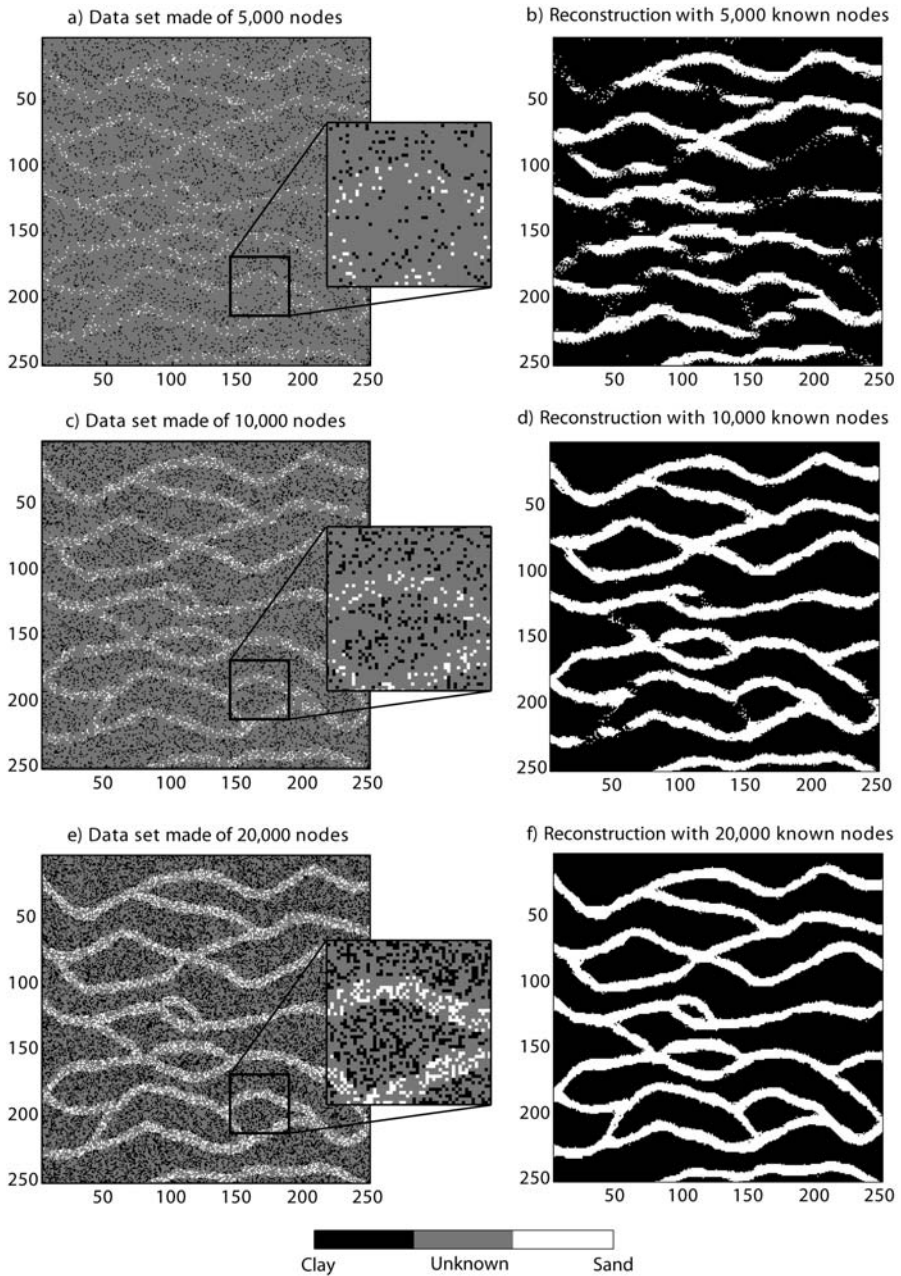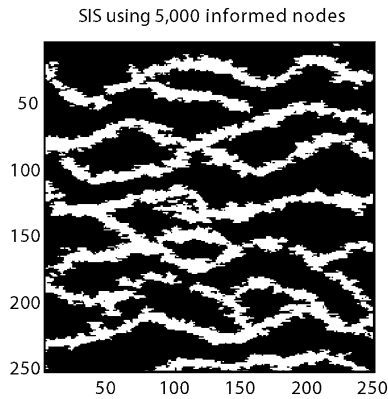
Fig. 2 Reconstruction using scattered data sets. (**a**) Data set consisting of 5,000 nodes. (**b**) One realization of reconstruction with a data set of 5,000 nodes (8% of the entire image). (**c**) and (**d**) Same with 10,000 informed nodes (16%). (**e**) and (**f**) Same with 20,000 informed nodes (31%)

Fig. 3 Reconstruction using
SIS with the same data set of
5,000 nodes

SIS using 5,000 informed nodes

To emphasize the poor performance of DS for reconstruction based on scattered data sets, it has been compared to sequential indicator simulation (SIS, Deutsch and Journel 1992) using the same 5,000 conditioning data as in Fig. 2b. The variogram model used for SIS was adjusted on the 5,000 data TD. It is a spherical model, with ranges of 27 m in the EW direction and 10 m in the NS direction. The same 5,000 data were used for conditioning. The results of this reconstruction using SIS are shown in Fig. 3. It displays better connectivity of sand bodies than DS reconstruction, even if artifacts appear due to the anisotropic variogram structure. Since DS reconstruction is an entirely data-driven approach, it cannot generate patterns that are absent from the data and fails when data are insufficient. Conversely, approaches based on a model (such as SIS) allow creating patterns that are not present in the data by using external information such as a variogram model or a training image, and such external information can be invaluable when data are scarce.

3.2 Continuous Variable Example

The second example illustrates the same problem using the synthetic (but realistic) transmissivity field presented by Kerrou et al. (2008). The reference image was built from an aerial photograph of braided channels in the Ohau River, New Zealand (Mosley 1982). The image size is 1,000 m by 400 m and it has been discretized on a grid of 440 by 176 nodes. It is divided in two facies codes: 0 for gravelly channels and 1 for silty lenses. Facies were converted to a continuous transmissivity field as described by Kerrou et al. (2008). Two multi-Gaussian unconditional simulations were separately generated to populate the channels and the lenses with logarithm of transmissivity values. The first simulation (channels) uses an exponential variogram with correlation range of 3 m. The second simulation (lenses) has a nested variogram that includes one isotropic exponential model with a 3 m range, plus a cubic anisotropic model with a range of 600 m in the EW direction and 300 m in the NS direction. This leads to a bimodal, non-multi-Gaussian, anisotropic transmissivity field.

Similarly to the previous example, Fig. 4a shows the extension of the unknown zone, which represents 20% of the domain. Figure 4b displays one realization of
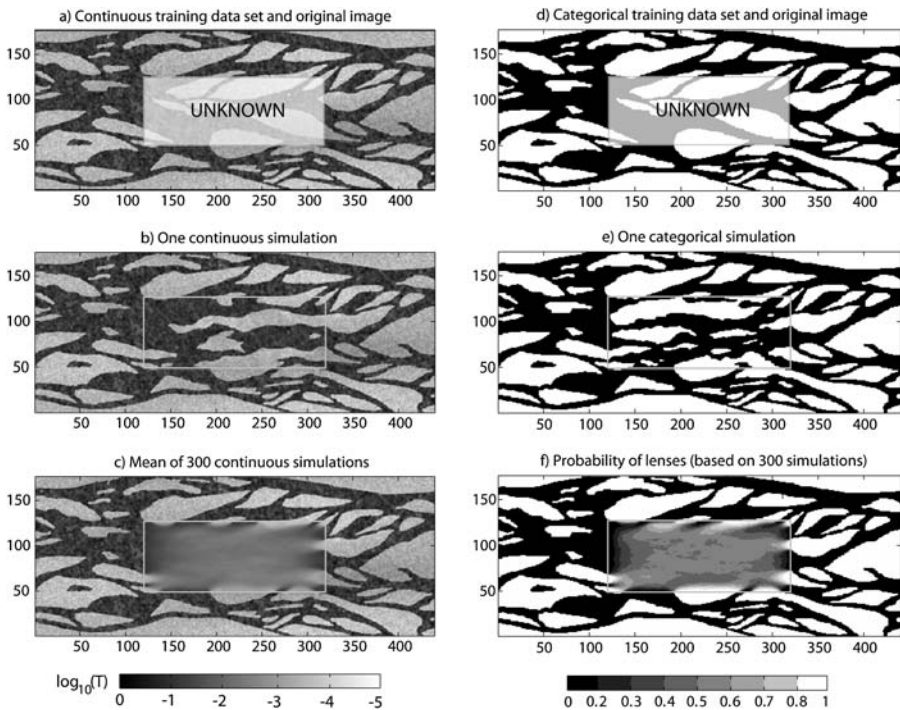
Fig. 4 Reconstruction of an image depicting an aerial view of a braided system. The size of the image is 440 by 176 nodes. (**a**) The original complete image (Kerrou et al. 2008) and the unknown zone to reconstruct. (**b**) One realization of reconstruction of the central zone. (**c**) Mean $\log_{10} T$, based on 300 realizations. (**d**) to (**f**) Same reconstruction for categorical variable. Here, (**f**) is the probability of occurrence of lenses

reconstruction. The distance measure (5) is used since the variable is continuous. In this case, a neighborhood of only 20 nodes is enough to give satisfying results, with a threshold set at $t = 0.025$. Visually, the reconstructed central zone has a similar spatial structure as the rest of the image. Figure 4c displays the mean of 300 reconstruction simulations, which come close to a conditional estimation. It shows that there is a zone of low uncertainty on the edges of the unknown zone. The features in this zone of low uncertainty are coherent with the position of the channels and lenses in the reference field. Figures 4d f illustrate the same reconstruction for the corresponding categorical variable, using the measure of distance (4) and the same neighborhood. Structures are visible in the central part of the estimation, showing that uncertainty is reduced even far from the edges.

Two sorts of constraints are imposed on the reconstruction of the unknown zone. The first one is given by the patterns present in the TD and that DS reproduces in the reconstruction. This is the structural constraint. The second constraint is imposed by the conditioning nodes that are close to the unknown zone. This is the local constraint. Ideally, these two constraining factors should be compatible as they both come from a single TD. In practice, non-stationarity can be present and the local condition-
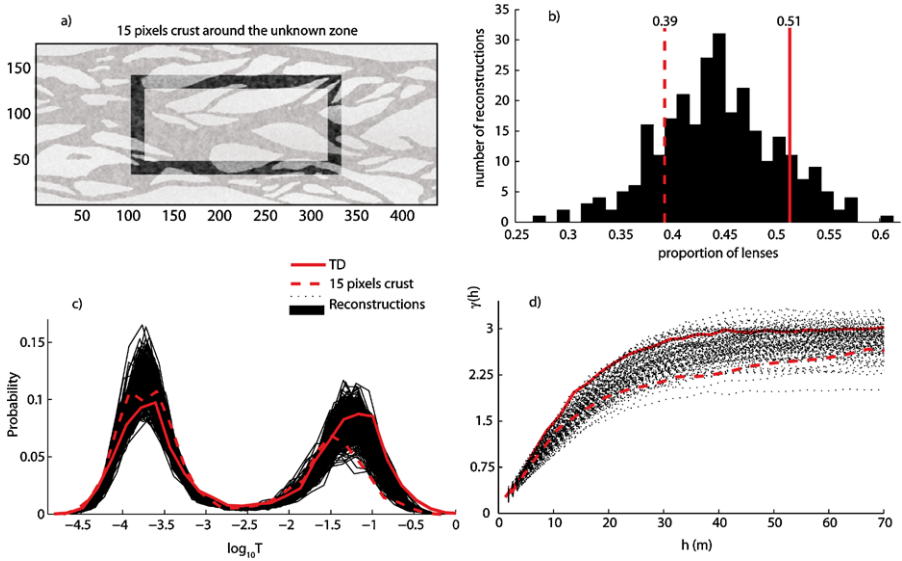
Fig. 5 Illustration of local and structural constraints. (**a**) 15 pixels' crust around the unknown zone. (**b**) Proportion of lenses in 300 continuous-variable reconstructions' simulations and in the different areas of the reference image. (**c**) Histograms of $\log_{10} T$ for 300 continuous-variable reconstructions, for the 15 pixels' crust and for the entire TD. (**d**) Experimental omnidirectional variograms for 300 continuous-variable reconstructions, for the 15 pixels' crust and for the entire TD

ing may become antagonistic with structural constraints. This situation occurs in the present example. The reference image has a global proportion of lenses of 0.50, but it is imbalanced between the different zones, with a proportion of lenses of 0.44 in the central (unknown) part and 0.51 on the outside. Figure 5a highlights the area of the TD within 15 pixels from the unknown zone, where the proportion of lenses is only 0.39, much lower than the marginal proportion. The result of this antagonism is a reconstruction that has statistical properties lying in between the ones of the entire TD and the area of the TD close to the unknown zone. Figure 5b shows the histogram of the proportion of lenses in 300 reconstructions. It is centered in between the proportions found in the TD and the proportions found in the 15 pixels' crust around the unknown zone.

The same phenomenon occurs for the reproduction of the histograms (Fig. 5c) of the simulated $\log_{10} T$ values. Here again, most histograms of simulated values are halfway between the TD and the conditioning 15 pixels' crust. The difference in lenses proportion that was noted for Fig. 5b is visible in the second mode of the distributions. Similarly, one can observe quite different omnidirectional variograms for the TD and the 15 pixels' crust (Fig. 5d). The reconstructions should have a variogram similar to the one of the TD, because they are made of patterns of nodes coming directly from the TD. Nevertheless, coherence with the existing bodies of the external area generates a structure that tends to the one of the areas of the TD neighboring the unknown zone. The resulting variogram lies in between these two models, which is necessary in order to have a coherent reconstruction.
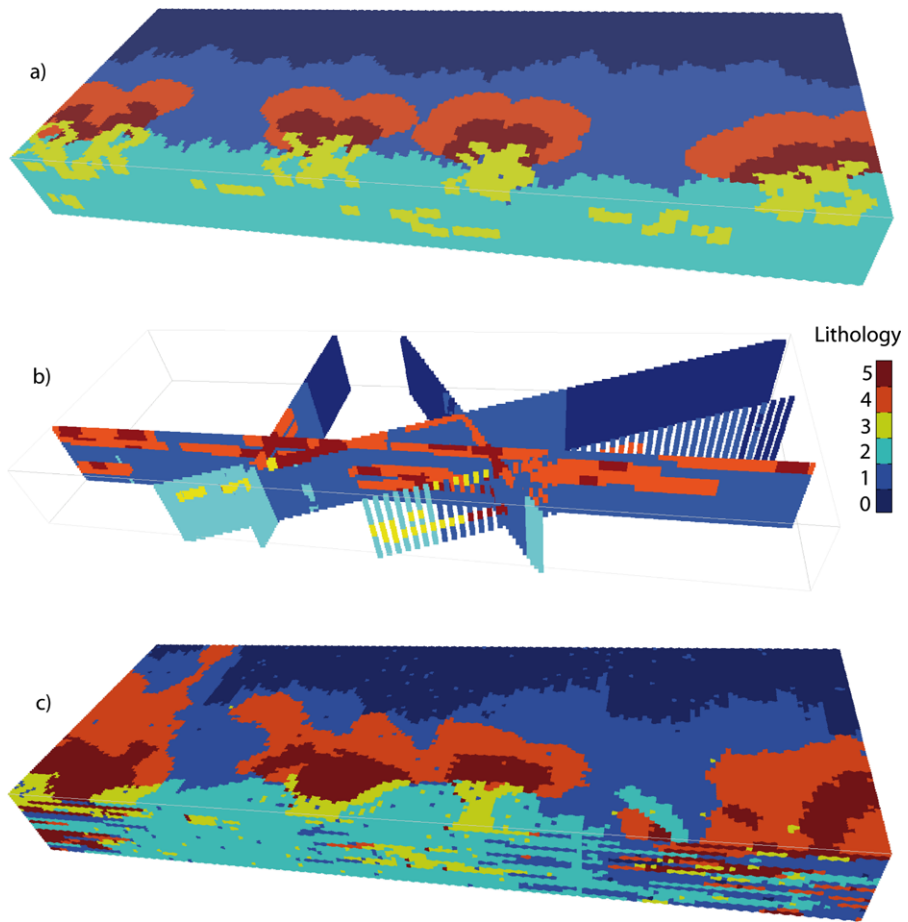
Fig. 6 Reconstruction of turbidite lobes. (**a**) Original reference image, from which the data set consisting of 5 cross sections (**b**) has been extracted. (**c**) One realization of reconstruction

3.3 3D Synthetic Example

The third example shows how much information can be extracted from a TD composed of continuous zones, such as 2D slices in a 3D bloc. The reference image is an object-based simulation depicting a geological setting of turbidites on a continental margin (Strebelle 2002). This reference image (Fig. 6a) is non-stationary in both horizontal and vertical directions because its scale is very large, ranging from proximal turbiditic sediments to distal pelagic deposits in the abyssal plain. It is made of 6 geological facies and discretized in $70 \times 183 \times 20$ grid nodes. From facies code 0 to 5, the different lithologies correspond respectively to abyssopelagic sediments, bathypelagic sediments, epipelagic sediments, coarse turbiditic sediments, medium turbiditic sediments, and fine turbiditic sediments.

The TD (Fig. 6b) is created by extracting points along 5 vertical cross sections in the reference image. In a real setting, these slices would typically correspond to

interpreted geophysical surveys, such as seismic profiles. The reconstruction grid has the same discretization as the reference image, with a total of 256,200 nodes, 11,380 of which (4.4%) are informed by the extracted slices. For the reconstruction, a neighborhood of 20 nodes is again used. In order to ensure that the first simulated nodes are close to the informed locations (close to the cross sections), a custom path in the SG is used instead of a random path. It consists of an altered random path ensuring that each simulated node has at least one of its direct neighbors informed (direct neighbors are only one grid block away from the simulated node). Therefore, the nodes close to the informed cross sections are simulated first, and the simulation proceeds away from there on. If a random path were used, it would start simulating nodes far away from the informed cross sections. Because large patterns have less chance to be found in the TD, it is likely that low-order (or even marginal) probabilities would be sampled at the beginning of the simulation. This would cause inconsistencies in the large patterns, which may deteriorate the resulting images.

Twenty realizations of reconstruction have been performed, one of them shown in Fig. 6c. At poorly informed locations, such as in the right and left edges of the image, the reconstruction produces degraded patterns compared to the reference. Although the reconstructions show a certain degree of complexity in these areas, some important features are missing. On the right side, facies 2 is noticeably absent. On the left, facies 1 and 4 extend horizontally until the abyssal zone; and in general, vertical alternations of facies are not geologically consistent. This illustrates the limits of the data-driven approach that cannot go beyond the information contained in the data. Again, if structures are not present in the TD, they will also not be present in the reconstruction. Conversely, at locations where sufficient information is present, complex shapes can be inferred from the data. In the central part of the image where three cross sections intersect, shapes of facies 4 and 5 resemble the lobes and elongated bodies of facies 3 are similar to what is found in the reference. Such features cannot be easily deduced from the TD.

In general, non-stationary trends are adequately reproduced in the reconstruction procedure, since the cross sections are large enough to capture it. The hierarchy of the turbiditic structures is respected in the reconstructions, with the channels coming first, followed by the proximal and distal parts of the lobes. Figures 7a and b show the probability maps of facies 4 and 5 (areas with a probability higher than 0.7 are highlighted). Facies 4 is generally more distal than facies 5 and surrounds it, which is a characteristic of lobes. Figure 7c displays the probability map of facies 0 (probability higher than 0.7). This facies is mostly present in the abyssal zone, as it should be according to the reference image. Similarly, turbiditic structures only appear on the proximal part of the image, along with facies 1 and 2.

In Fig. 7c, a sharp transition in the probability map of facies 0 is visible at the equidistance of two parallel TD cross sections (right side of the image). We believe such artifacts are related to the custom path used. The simulation grid is filled starting with the nodes close to the data. When simulated nodes originating from distant slices meet, incompatibilities can occur. One way to deal with these artifacts could be to use larger neighborhoods. This would increase the probability that the simulated nodes coming from another slice belong to the neighborhood of the node to simulate. A more proper solution would be to use syn-processing to remove these artifacts
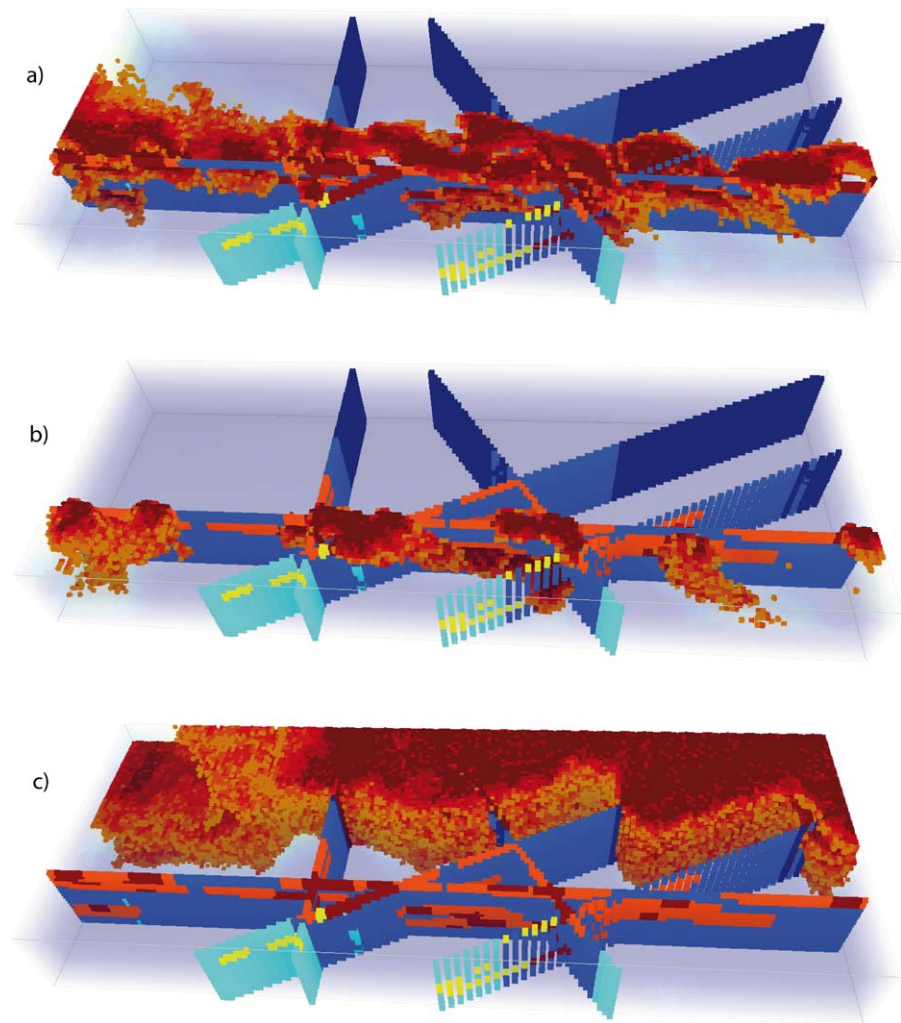
Fig. 7 Probability maps of facies, computed from 20 realizations. (**a**) Area where the probability of facies 4 exceeds 0.7. (**b**) Area where the probability of facies 5 exceeds 0.7. (**c**) Area where the probability of facies 0 exceeds 0.7

(Mariethoz et al. 2010). It is a recursive procedure that un-simulates and re-simulates nodes with the aim of removing inconsistencies in the patterns of the simulation. Figure 8 displays the facies proportions in the reference, in the simulations and in the TD. In the TD, there is a large bias for most of the facies because the irregular disposition of the slices and the global non-stationarity of the reference image result in a sampling that is not representative. Facies 0, 2, and 3 are under-represented and facies 1, 4, and 5 are over-represented. In the reconstructions, the largest biases (facies 0, 1, 4, and 5) have been reduced. Structural constraints are not well defined because the TD is not fully representative and because of non-stationarity. On the other hand, local constraints take very well into account the non-stationarity inherent to the image.
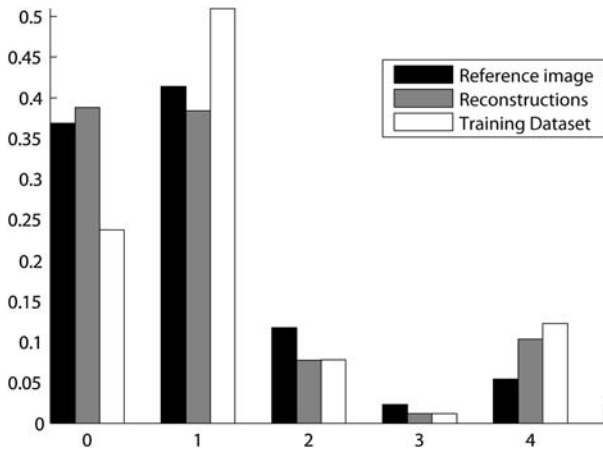
Fig. 8 Facies proportions in the reference image, the 20 reconstructions and the TD. Proportions of reconstructions have been computed using 20 realizations

As observed in the previous example, even if the structural constraints are biased, the combination with the local constraints corrects the structure to a certain extent. The result is that the facies proportions in the reconstructions are closer to the reference proportions than what is observed in the TD. Here again, accounting for multiple-point statistics allows exploiting the available data with maximum coherence, even in the presence of non-stationarity.

3.4 Real-Case 3D Application

The next example is the application of the direct sampling reconstruction algorithm to a real 3D data set. The study site is a quarry located in the sediments of the Lambro River, in northern Italy. The geological environment is an extremely heterogeneous setting of meandering sediments with point-bar deposits. The geometries of such sedimentological formations are widely analyzed and described in the literature (Miall 1996; Nichols 1999; Kostic and Aigner 2007) and are a classical challenge for geostatistical simulations (Deutsch and Tran 2002; Zappa et al. 2005; Rubin et al. 2006; Mariethoz et al. 2009a). The available data are based on a series of measurements taken on 5 different exposure stages of the quarry front. On each of the 5 documented fronts, centimeter-scale logging, textural and petrophysical analysis have been performed. Hierarchical classification and interpretation of the sediments have led to the definition of 4 hydrofacies, represented in a data set of 483,472 point values located along the 5 quarry fronts (details can be found in Bersezio et al. 2007). Migrated on the SG, these points constitute the training data set used for the reconstructions (Fig. 9a). A previous study (dell'Arciprete et al. 2008) performed reconstructions based on the same data set, using well-established methods of SIS (Deutsch and Journel 1992) and transition probabilities (Carle and Fogg 1997). Contrarily to dell'Arciprete et al. (2008), the present application does not intend to benchmark methods, but only to apply DS reconstruction to a real data set.
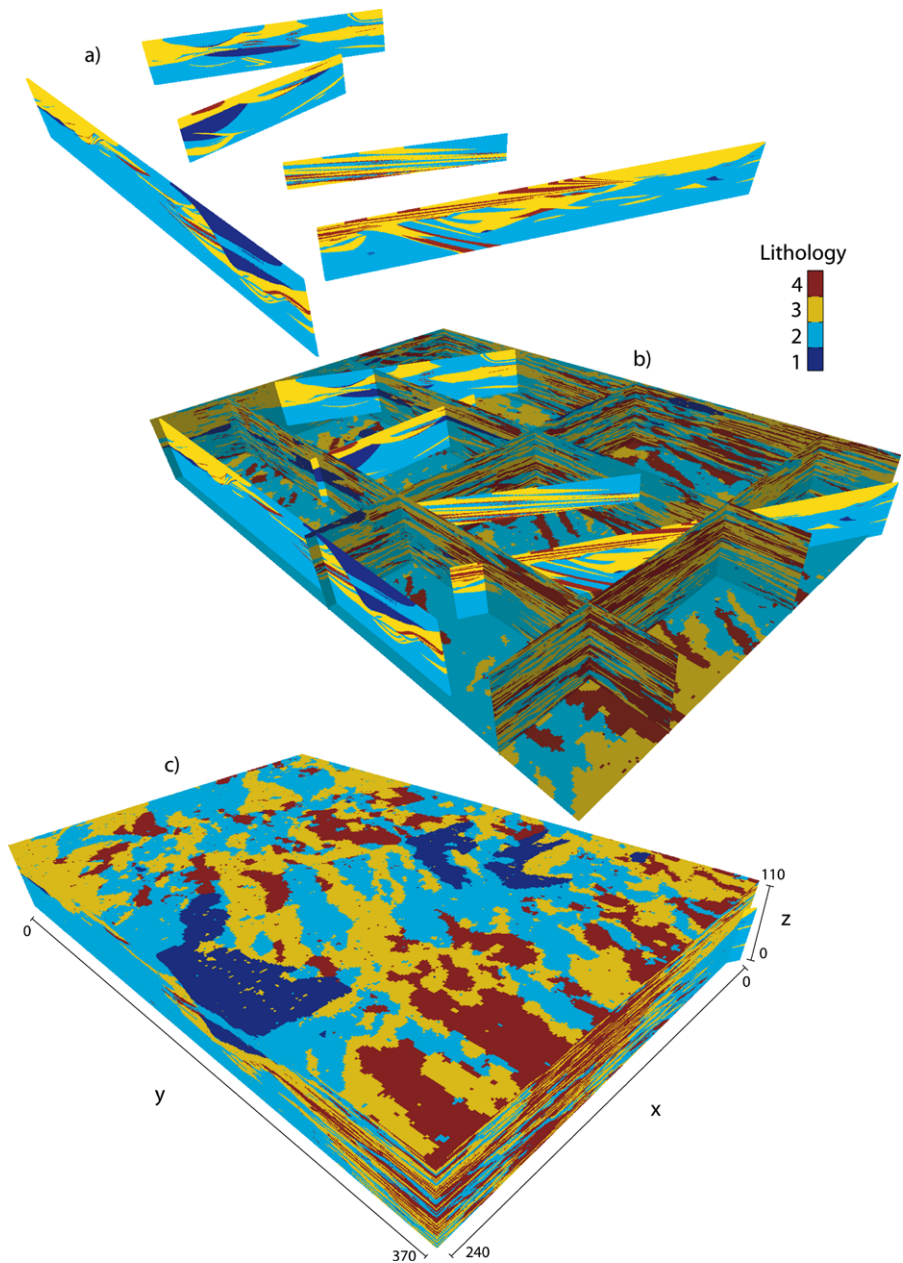
Fig. 9 (**a**) Original data set consisting of 5 interpreted cross sections. (**b**) One high-resolution simulation of reconstruction, slices view (10 million cells). (**c**) Same simulation of reconstruction, block view

Two different grids are used for reconstruction. The first one, coarse, is constituted of 2,764,800 grid blocks. The second one, finer, has a total of 9,768,000 smaller blocks. Because of the very high resolution of measurements, each facies point data

Table 1 Summary of the properties of both reconstruction grids

	Size of grid x, y, z			Total no. of nodes	Blocks size x, y, z (cm)			No. of informed nodes
Coarse grid	160	240	72	2,764,800	31.25	31.25	12.5	36,291 (1.31%)
Fine grid	240	370	110	9,768,000	20	20	8	85,056 (0.87%)

cannot be migrated on a separate grid block, even on the fine grid. The number of informed grid nodes is different on the two grids, but in both cases it remains a very low proportion of the entire grid (1.31% for the coarse grid and 0.87% for the finer grid). A summary of the grid properties is presented in Table 1. On the finer grid, a single reconstruction realization was made to keep CPU time reasonable (one fine-grid reconstruction takes two days on a 7-CPU parallel machine, but only a few hours for the coarse grid). Figures 9b and c show two different views of the reconstructed lithologies. Figure 10 displays 6 cross sections of the reconstruction on fine grid. Twenty other realizations were performed on the coarse grid. These simulations are used to compute statistics on the probability of occurrence of facies (Fig. 11). All reconstructions use a neighborhood of 35 nodes and the same custom path as in the previous example.

The purpose of the reconstruction on a fine grid is to check the reproduction of the very fine stratified structures present in the TD, and a single realization suffices for this. Figure 9b displays the TD together with cross sections in the SG. The intersections of all interfaces between the different transects are coherent. The stratifications observed in the TD (alternating beddings of lithologies 2, 3 and 4) are clearly visible in the reconstruction.

Figure 10 shows that the non-stationarity of the data is preserved. For example, there is more stratification of lithofacies 3 and 4 with increasing y coordinate (compare Figs. 10a, c and f). Moreover, one can observe large bodies of lithofacies 3 in a matrix of facies 2 in Fig. 10a, whereas smaller and less continuous bodies of facies 2 are found at the other end of the domain. Similarly, facies 1 is only present in certain areas of the TD. The reconstruction respects this spatial distribution, since this facies is essentially found in Figs. 10a and b, and is almost absent in the other cross sections.

Structures showing different orientations are present in the TD. This is especially visible with facies 4 that is present in continuous sub-horizontal layers in the upper part of the domain, but also dipping structures having an angle of about 45° in the lower part of the domain (Fig. 9a). Modeling such structural non-stationarity would normally call for dividing the domain into separate structural units that would be simulated separately. This has not been done here. Nevertheless, Figs. 10c, d and e show that this dual orientation of structures is present in the reconstruction. Figure 11 shows that the lithofacies proportions in the TD and in the reconstructions are very close, but there is no exact match. As the proportions of the real field are unknown, it is not possible to know if the reconstructed proportions are closer to the reality than the proportions of the TD.
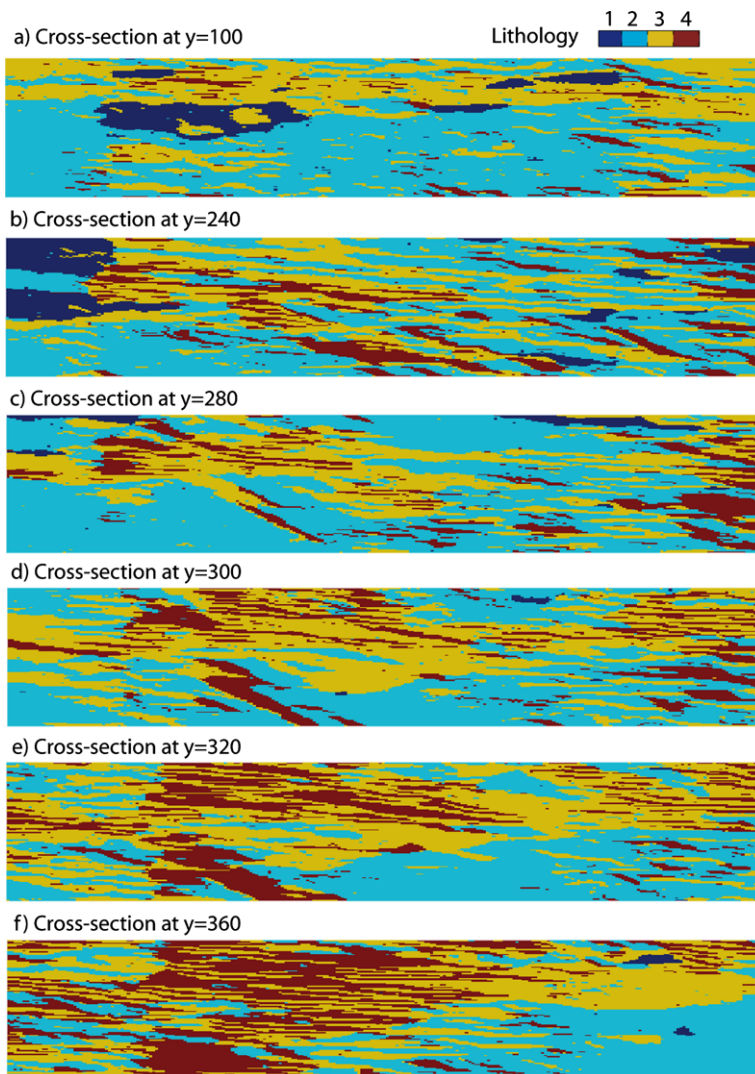
Fig. 10 (**a**) to (**f**) Cross sections in the reconstruction on fine grid. All sections are parallel to the x-axis, at a constant y coordinate indicated on the figure

3.5 Real-Case Borehole Imagery Example

The last reconstruction example illustrates an application of the DS reconstruction method to a borehole microresistivity image. Borehole images are a powerful way of obtaining data for the orientation of bedding planes and fractures (Wu and Pollard 2002; Khoshbakht et al. 2009) to identify cycles in sedimentary sequences (Lefranc et al. 2008), or for visual inspection of boreholes. Borehole imagery is a common example of unknown zones in an otherwise known image. Borehole images can be obtained by inserting tools such as an FMI (Fullbore Formation MicroImager, Schlumberger)

Fig. 11 Facies proportions in the TD and in 20 reconstruction realizations on the coarse grid
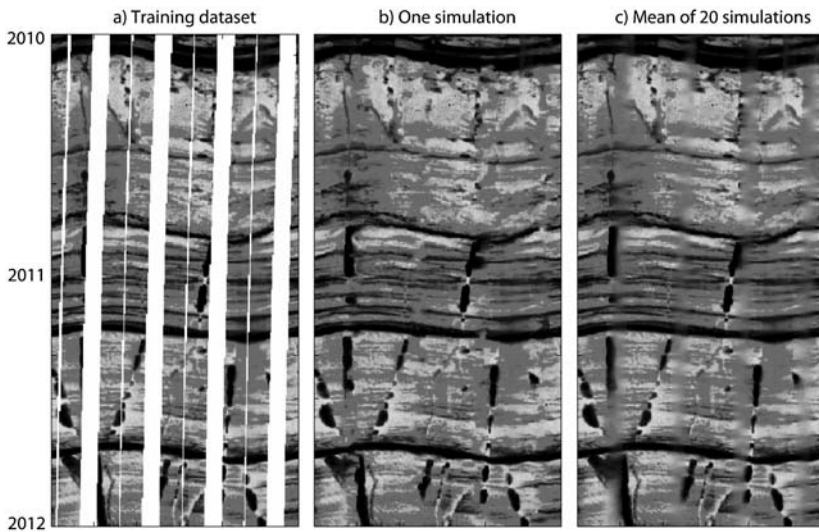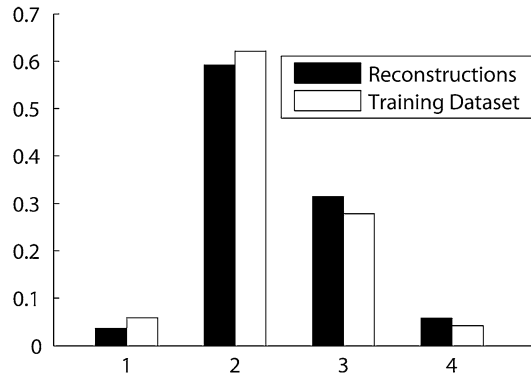

Fig. 12 (**a**) Training data set consisting of microresistivity measurements taken by the 8 sensors of a FMI device (courtesy of HEF Petrophysical Consulting). The vertical axis represents the borehole depth. (**b**) One reconstruction realization. (**c**) Mean of 20 reconstruction realizations

device into the borehole. The device is composed of 4 pads carrying 2 sensors each. The sensors induce a current into the formation that is modulated in amplitude, providing rich information about the petrophysical properties. The resulting images are exhaustively informed along the pads' trajectory, but not informed at all between the pads. Therefore, accurate estimation of bedding planes and fractures' orientation calls for interpolation and reconstruction techniques. Reconstruction could also be a precious aid for the visual inspection of the inside of a borehole. It is also much easier to carry on geological interpretation based on complete images. Figure 12a displays a data set obtained with an FMI device. The unknown zones caused by the spacing between pads are shown in white. In this case, 70% of the image is informed, constituting the TD available for reconstruction.

Twenty realizations of reconstruction have been generated using the distance in (5) with a neighborhood of $n = 20$ nodes and a threshold of $t = 0.01$. One of them is displayed in Fig. 12b. The fine-scale heterogeneity and the continuity of horizons and fractures are preserved. At locations where they are present in the TD, bedding planes are connected (e.g. dark lines in the central part of the image). The internal texture within the sedimentary units (gray and white) is also ensured. The TD is non-stationary because it is made of different zones comprising thick and thin horizons, fractured and non-fractured areas, massive or finely laminated structures. Here again, the method is able to accommodate non-stationarity because the different zones of the TD are not mixed in the reconstructed areas. In Fig. 12c, the mean of the entire stack shows that the continuity of the major features is well reproduced on all simulations, with more variability on the small scale heterogeneity resulting in blurry areas along the unknown zones.

4 Discussion and Conclusion

In this paper, we presented a method for reconstructing partial images. The method can be used for a wide range of applications. Indeed, this problem is very general even though we mainly presented 2D and 3D Earth Science applications. We have shown several possible applications of the method, but many others can be envisioned. For example, time series observations are frequently left incomplete because an instrument has been damaged. When this occurs, one usually has to build a model to reconstruct the missing data. With the proposed method, this can be done rapidly, even if the patterns and statistical distributions of the measurements are complex. The proposed algorithm propagates the properties of a known data set to locations where the values of the variable at stake are unknown. This method is based on the technique of multiple-points simulations by direct sampling that presents several advantages over traditional methods. First, it uses multiple-points statistics, and therefore can propagate whole patterns rather than two-point statistics. Second, because the neighborhoods are flexible, it is possible to use patterns that are combinations of irregularly positioned data, such as a set of randomly positioned points or different cross sections. Third, it can accommodate either categorical or continuous variables. And fourth, there are no problems of data relocation as multiple-grids are not necessary.

The presented reconstruction method has the major advantage that it needs very little parameterization. Since the spatial model is non-parametric and is given in the training data, no model inference is needed. Similarly, a certain amount of non-stationarity is directly accounted for if it is present in the training data, and trends do not need to be known in advance. A clear gain is the reduction of the uncertainty associated with the model selection and inference. Another major advantage of using this easy parameterization is that it minimizes the time spent adjusting the parameters for a geological model. This can be of great appeal for real-case applications where time and money are key factors. In the different examples presented, the method successfully reconstructed partial data sets made of continuous unknown zones, such as holes in an otherwise complete image or completely informed cross sections in a 3D

domain. However, reconstruction based on very scattered data sets remains a challenge due to the lack of information at small scale. The available data set must also be sufficiently rich to contain ample diversity of patterns, and to represent any non-stationarity present in the domain. Moreover, because the reconstruction algorithm is mainly data-driven, it is not possible to validate the spatial model. If measurement or upscaling errors are present in the training data set, the reconstructed images can be seriously biased. For example, if permeability data are migrated on the SG using a poor upscaling scheme, the entire reconstruction will reflect this choice. Similarly, realizations can be adversely affected if non-stationarity is not captured in the data or if the sampling of the data is preferential. Since the model is defined by the data, any approximation in the handling or in the interpretation of these data will be incorporated in the results, and there are no means of correcting such bias by adding external structural information.

A possible direction for further research could be to use an external complete training image when the training data set is insufficient. For each simulated grid node, the TD would first be scanned. The training image would be scanned only if no matching data event can be found in the TD. It could allow the modeler to introduce extreme values and specific connectivity patterns not present in the TD. As for classic multiple-points simulation, the training image may be derived from expert interpretation and qualitative knowledge. The TI would then be a prior model, as opposed to the implicit spatial model based on the TD.

References

Arpat B, Caers J (2007) Conditional simulations with patterns. Math Geol 39(2):177–203

Bersezio R, Giudici M, Mele M (2007) Combining sedimentological and geophysical data for high resolution 3-d mapping of fluvial architectural elements in the quaternary Po Plain (Italy). Sediment Geol 202:230–247

Carle SF, Fogg GE (1997) Modeling spatial variability with one and multi-dimensional continuous Markov chains. Math Geol 7(29):891–918

Cornacchiulo D, Bagtzoglou C (2004) Geostatistical reconstruction of gaps in near-surface electrical resistivity data. Vadose Zone J 3:1215–1229

dell'Arciprete D, Felletti F, Bersezio R (2008) Simulation of fine-scale heterogeneity of meandering river aquifer analogues: comparing different approaches, paper presented at geoENV 2008. Southampton, 8–10 September, 2008

Deutsch C, Journel A (1992) GSLIB: geostatistical software library. Oxford University Press, Oxford

Deutsch C, Tran T (2002) FLUVSIM: A program for object-based stochastic modeling of fluvial depositional systems. Comput Geosci 2002(28):525–535

Fiorucci P, La Barbera P, Lanza L, Minciardi R (2001) A geostatistical approach to multisensor rain field reconstruction and downscaling. Hydrol Earth Syst Sci 5(2):201–213

Gómez-Hernández JJ, Wen X-H (1998) To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. Adv Water Resour 21(1):47–61

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, Oxford

Guadagnini L, Guadagnini A, Tartakovski D (2004) Probabilistic reconstruction of geologic facies. J Hydrol 294:57–67

Guardiano F, Srivastava M (1993) Multivariate geostatistics: beyond bivariate moments. In: Geostatistics-Troia. Kluwer Academic, Amsterdam, pp 133–144

Hurley N, Zhang T (2009) Method to generate fullbore images using borehole images and multi-point statistics. In: SPE middle east oil and gas show and conference, 15–18 March 2009, Bahrain, Bahrain

Journel A, Isaaks E (1984) Conditional indicator simulation: application to a Saskatchewan deposit. Math Geol 16(7):685–718

Journel A, Zhang T (2006) The necessity of a multiple-point prior model. Math Geol 38(5):591–610

Kerrou J, Renard P, Hendricks-Franssen H-J, Lunati I (2008) Issues in characterizing heterogeneity and connectivity in non-multi-Gaussian media. Adv Water Resour 31(1):147–159

Khoshbakht F, Memarian H, Mohammadnia M (2009) Comparison of Asmari, Pabdeh and Gurpi formation's fractures, derived from image log. J Petroleum Sci Eng 67(1–2):65–74

Kitanidis P (1986) Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resour Res 22(4):499–507

Kostic B, Aigner T (2007) Sedimentary architecture and 3-D ground penetrating radar analysis of gravelly meandering river deposits (Neckar valley, SW Germany). Sedimentology 54:789–808

Kumar D, Ahmed S (2008) Reconstruction of water level time series in an aquifer using geostatistical technique. In: Groundwater dynamics in hard rock aquifers. Springer, Berlin, pp 191–200

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

Lefranc M, Beaudoin B, Chiles J-P, Guillemot D, Ravenne C, Trouiller A (2008) Geostatistical characterization of Callovo–Oxfordian clay variability from high-resolution log data. Phys Chem Earth, Parts A/B/C 33(Supplement 1):S2–S13

Marache A, Riss J, Gentier S, Chiles J-P (2002) Characterization and reconstruction of a rock fracture surface by geostatistics. Int J Numer Anal Methods Geomech 26(9):873–896

Mariethoz G (2009) Geological stochastic imaging for aquifer characterization. University of Neuchâtel, Neuchâtel, Switzerland

Mariethoz G, Renard P, Cornaton F, Jaquet O (2009a) Truncated pluri-Gaussian simulations to characterize aquifer heterogeneity. Ground Water 47(1):13–24

Mariethoz G, Renard P, Froidevaux R (2009b) Integrating collocated auxiliary parameters in geostatistical simulations using joint probability distributions and probability aggregation. Water Resour Res 45:W08421

Mariethoz G (2010, in press) A general parallelization strategy for random path based geostatistical simulation methods. Comput Geosci

Mariethoz G, Renard P, Straubhaar J (2010, submitted) The direct sampling method to perform multiple-points simulations. Water Resour Res

Miall A (1996) The geology of fluvial deposits. Springer, New York

Mosley P (1982) Analysis of the effect of changing discharge on channel morphology and instream uses in a Braided river, Ohau river, New Zealand. Water Resour Res 18(4):800–812

Nichols G (1999) Sedimentology and stratigraphy. Blackwell, London

Okabe H, Blunt M (2004) Multiple-point statistics to generate pore space images. In: Geostatistics Banff. Springer, Berlin, pp 763–768

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

Remy N, Boucher A, Wu J (2009) Applied geostatistics with SGeMS: A user's guide. Cambridge University Press, Cambridge

Renard P, Gómez-Hernández J, Ezzedine S (2005) Characterization of porous and fractured media. In: Encyclopedia of hydrological sciences. Wiley, New York

Rubin Y, Lunt I, Bridge J (2006) Spatial variability in river sediments and its link with river channel geometry. Water Resour Res 42:W06D16

Saripally I (2008) Evaluating data conditioning accuracy of MPS algorithms and the impact on flow modeling. In: 21th SCRF Meeting, Stanford University, May 8–9, 2008

Shannon CE (1948) A mathematical theory of communication. Bell Syst Techn J 27:379–423

Straubhaar J, Walgenwitz A, Renard P, Froidevaux R (2008) Optimization issues in 3D multipoint statistics simulation, paper presented. Geostats 2008, 1–5 Dec 2008, Santiago, Chile

Strebelle S (2002) Conditional simulation of complex geological structures using multiple point statistics. Math Geol 34(1):1–22

Strebelle S, Payrazyan KJC (2002) Modeling of a deepwater turbidite reservoir conditional to seismic data using multiple-point geostatistics. In: The 2002 SPE annual technical conference and exhibition. San Antonio, September 29–October 2, SPE paper 77425

Wu H, Pollard D (2002) Imaging 3-D fracture networks around boreholes. AAPG Bull 86(4):593–604

Wu J, Boucher A, Zhang T (2008) A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. Comput Geosci 34(12):1863–1876

Youngseuk K, Mukerji T, Amos N (2004) Permeability prediction from thin sections: 3D reconstruction and Lattice–Boltzmann flow simulation. Geophys Res Lett 31:(L04606)

Zappa G, Bersezio R, Felletti F, Giudici M (2005) Modeling heterogeneity of gravel-sand, braided stream, alluvial aquifers at the facies scale. J Hydrol 325(1–4):134–153

Zhang T, Switzer P, Journel A (2006) Filter-based classification of training image patterns for spatial simulation. Math Geol 38(1):63–80