



Three-dimensional high resolution fluvio-glacial aquifer analog – Part 2: Geostatistical modeling

A. Comunian^{a,*}, P. Renard^a, J. Straubhaar^a, P. Bayer^{b,**}

^a Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland

^b ETH Zürich, Engineering Geology, Sonneggstrasse 5, 8092 Zürich, Switzerland

ARTICLE INFO

Article history:

Available online 1 April 2011

This manuscript was handled by P. Baveye Editor-in-Chief

Keywords:

Multiple-point statistics

Heterogeneity

Hierarchical framework

Aquifer analog

Hydrofacies

Fluvio-glacial

SUMMARY

The heterogeneity of sedimentary structures at the decimeter scale is crucial to the understanding of groundwater flow and transport. In a series of two papers, we provide a detailed analysis of a fluvio-glacial aquifer analog: the Herten site. The geological data along a series of 2D sections in a quarry, the corresponding GPR measurements, and their sedimentological interpretation are described in the companion paper. In this paper, we focus on the three-dimensional reconstruction of the heterogeneity. The resulting numerical model is provided as an [electronic supplementary material](#) for further studies. Furthermore, the geostatistical parameters derived from this analysis and the methodology described in the paper could be used in the future for the simulation of similar deposits where less data would be available. To build the 3D model, we propose a hierarchical simulation method which integrates various geostatistical techniques. First, we model the subdivision of the domain into regions corresponding to main sedimentological structures (e.g. a sedimentation event). Within these volumes, we use multiple-point statistics to describe the internal heterogeneity. What is unusual here is that we do not try to use a complex training image for the multiple-point algorithm accounting for all the non-stationarity and complexity, but instead use a simple conceptual model of heterogeneity (ellipsoidal shapes as a training image) and constrain the multiple point simulations within the regions by a detailed interpolation of orientation data derived from the 2D sections. This method produces realistic geological structures. The analysis of the flow and transport properties (hydraulic conductivity and tracer breakthrough curves) of the resulting model shows that it is closer to the properties estimated directly from the 2D geological observations rather than those estimated from a model of heterogeneity based on probability of transitions and not including the modeling of the large-scale structures.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

This paper is the second of a two-part series investigating the heterogeneity of a fluvio-glacial aquifer analog. The overall motivation of the work is that a realistic modeling of aquifer heterogeneity requires on the one hand accurate data on the expected sedimentological structures and on the other hand appropriate modeling techniques.

While the first paper Bayer et al. (2011) in this series describes the field work and proposes a sedimentological interpretation of the data, this second part focuses on the modeling aspects and follows two aims. The first is to build a high resolution three-dimensional model of the heterogeneity of the Herten fluvio-

glacial aquifer analog. This model constitutes an important preliminary step for future studies investigating the impact of heterogeneity on hydrogeological processes or on the significance of field experiments.

The second aim of the paper is to propose a hierarchical approach for the modeling of such environment types. Among the various existing methods to model heterogeneity reviewed by Koltermann and Gorelick (1996) and de Marsily et al. (2005), multiple-point (MP) statistics (Guardiano and Srivastava, 1993; Strebel, 2000; Hu and Chugunova, 2008) is appealing because it provides a simple mean to integrate a conceptual geological model in a stochastic simulation framework. The conceptual model is provided as a training image (TI) representing the heterogeneity patterns that the user expects at a given site. For example, the TI can represent meanders if the user knows that the geological environment that he wants to model is made of meanders. Then, when using MP statistics, one of the most important questions to answer is how to get the TI? A possibility often cited in the literature is to directly use the geological maps or sections obtained from analog studies as TI.

* Corresponding author.

** Principal corresponding author.

E-mail addresses: alessandro.comunian@unine.ch (A. Comunian), philippe.renard@unine.ch (P. Renard), julien.straubhaar@unine.ch (J. Straubhaar), bayer@erdw.ethz.ch (P. Bayer).

In practice, only a few attempts to use real geological analogs as training images have been made so far (Maharaja, 2005; Huysmans et al., 2008). These works have shown that analogs often display significant non stationarity (because of the physical nature of the geological processes that lead to the heterogeneity). Using them directly as training images produces often unacceptable simulations (e.g. Boucher, 2009); indeed MP statistics assume that there is a repetition of the spatial heterogeneity patterns in the TI to build the MP probability distributions (ergodicity). If the TI is not stationary, the different patterns are mixed in the probability distributions. Different techniques have been developed to overcome this limitation. The main idea is to provide not only the TI but also some additional information (i.e. orientation maps, results of geophysical surveys, probability maps) to avoid mixing the patterns when computing the probability distributions (Boucher, 2009; Chugunova and Hu, 2008; de Vries et al., 2009; Mariethoz et al., 2010; Straubhaar et al., 2011). Even if the problem of non stationarity is important, one has to remember that it is not specific to MP statistics. It affects, in a similar manner, all the standard geostatistical methods based on lower order statistics. However, when only two-point statistical moments are used, the simulated structures are not expected to resemble the original ones (because the two-point moments do not contain enough information to constrain the shape of the structures, see e.g. Strebel (2000)) and the non stationarity is less apparent.

Another issue when trying to use analog data to simulate 3D heterogeneity with MP statistics is that one needs a 3D training image to simulate the 3D domain. In general, analog observations

only provide 2D training images. It is therefore necessary to develop specific methods that allow using MP statistics in that situation. If we look at standard geostatistical procedures, the practice in those cases would be to assume some invariance of the probability laws by rotation (Chilès and Delfiner, 1999). The covariances or variograms would be inferred along the available data in the directions included in the geological section, and the variability in the perpendicular directions would be assumed to follow the same laws or to be simply rescaled to account for some anisotropy. Here similar assumptions could be made, but the current implementations (both commercial and open source) of MP statistics do not yet offer these possibilities. When no other technique is available, the user has to build a 3D training image that is, as far as possible, compatible with the 2D observations. One can do it for example by using object based techniques, as proposed by Caers (2005) and Maharaja (2008), or by using 3D geometrical modeling techniques. If the geological structures are highly complex and intricate, this can be very tedious.

To overcome the problem of non stationarity and the problem of insufficient access to a 3D training image, the approach proposed in this paper consists of subdividing the domain in smaller regions corresponding to a series of successive sedimentological units corresponding to the large scale heterogeneity. In these units (also named regions), the heterogeneity is simpler and one can use basic but reasonable 3D training images constructed with object based techniques. In other words, the non stationarity related to the presence of major sedimentological units is treated by modeling a set of regions. The non stationarity of orientations of the

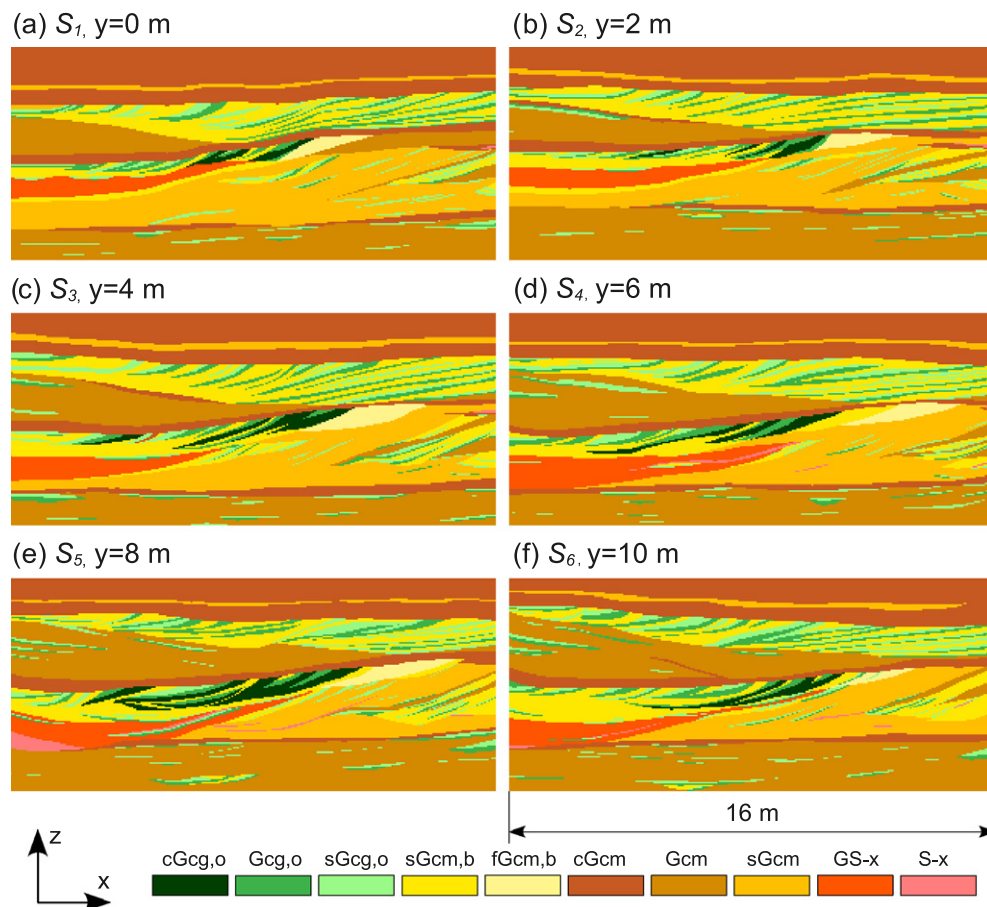


Fig. 1. Rasterized six sections S_j (16 by 7 m) of the Herten case study. Fig. 7a provides an ensemble view. The hydrofacies codes are explained in detail in the companion paper (Bayer et al., 2011).

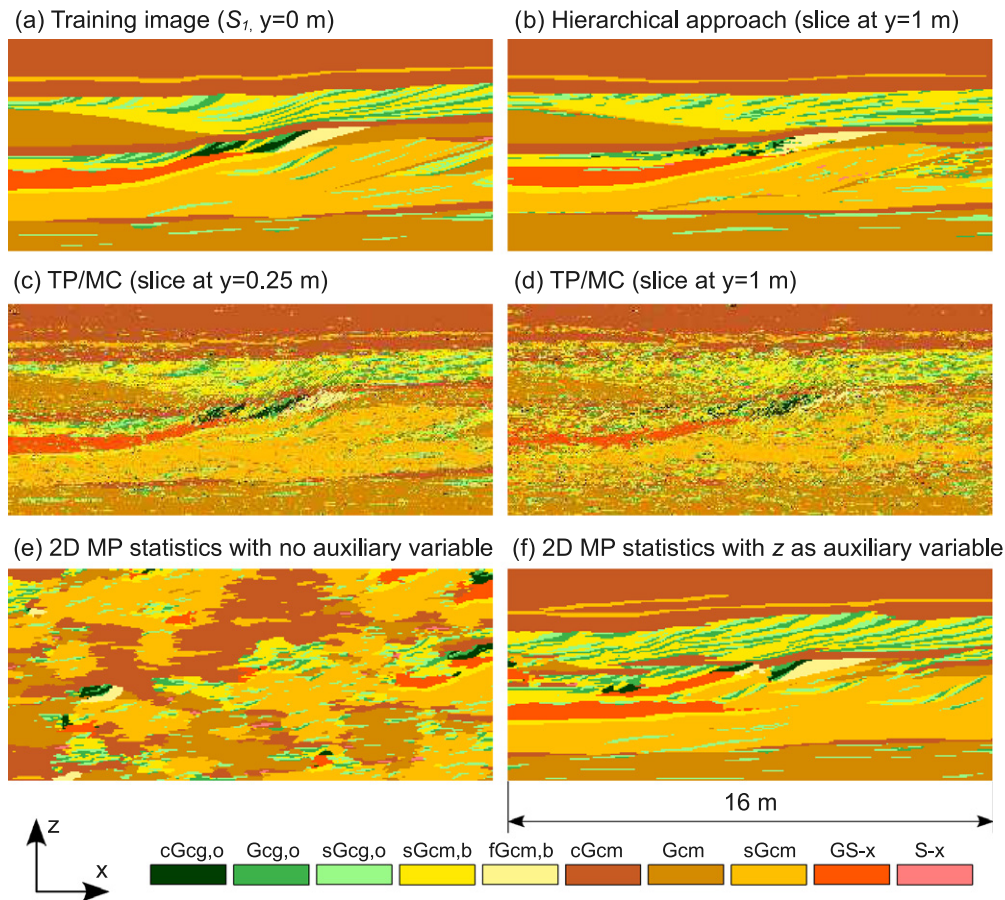


Fig. 2. A comparison between the simulations obtained with TP/MC, MP statistics with and without auxiliary variables and the proposed hierarchical approach. Figures b–d are slices parallel to the plane xz of 3D simulations obtained using the six S_j as conditioning data. Figures e and f are obtained with the training image (a) and no conditioning data. The hydrofacies codes are explained in detail in the companion paper (Bayer et al., 2011).

smaller scale heterogeneity within the regions is modeled separately by interpolating their orientations. In deltaic reservoirs, a similar approach has already proven to be very efficient (Khan et al., 2008).

The paper is structured as follows. First we present the Herten analog data set. Then we illustrate the limits of current modeling approaches on this data set in order to justify the need for the proposed approach. The following sections are devoted to the step by step description of the approach. The paper ends with a presentation of the results and a discussion. All the MP statistic simulations are obtained with the algorithm improved multiple-point algorithm using a list approach (IMPALA, Straubhaar et al. (2011)). The resulting 3D models are provided as [electronic supplementary material](#) downloadable from the web site of the journal or available upon request to the authors.

2. The Herten dataset

The Herten site is located in the Rhine river valley in southern Germany, close to the city of Basel (Switzerland). The site was selected because it is one of the best characterized gravel bodies of this size, and at the same time the heterogeneity and the non-stationarity of the data set are a challenging benchmark for a simulation technique. Moreover, other studies were performed on the same dataset, allowing a direct comparison of the results. The data set contains six parallel geological sections having a dimension of

16 by 7 m (Fig. 1). In the following, we will refer to the six sections as $S_1, \dots, S_6$. These data were obtained by mapping the geological facies in successive outcrop sections exposed during the pit exploitation. The distance between each section is 2 m. In the reference system adopted in this manuscript these sections are parallel to the plane xz .

Ten different hydrofacies types are mapped within the sections with a discretization of 5 cm along both x and z direction. For each hydrofacies, hydraulic conductivity and effective porosity were measured on a number of samples ranging from 3 to 15. For an exhaustive description of the mapping procedure, the hydrofacies coding conventions, as well as the available sedimentary, hydrogeological and geophysical data sets, the reader is referred to part 1 of the companion paper (Bayer et al., 2011).

In the present paper, we use the hydrofacies data and build a 3D stochastic model within a hierarchical MP statistics framework. The hydrofacies observations from the outcrops (Fig. 1) are used both as hard conditioning data (268,000 points) to constrain the simulation and more generally as a source of information to develop the whole methodology. The discretization is identical to that used by Maji et al. (2006). The simulation domain Ω is divided into 9 million ($320 \times 201 \times 140$) cubic cells having a side of 5 cm.

This discretization is selected for three reasons: (1) to reflect the very high resolution of the data; (2) to allow direct comparison with previous work on the same site; and (3) to provide a high resolution 3D model of the heterogeneity of fluvio-glacial deposits

that could then be used in the future to investigate how heterogeneity can control physical processes in these systems.

3. Motivations for a hierarchical approach

Before entering into the description of the proposed hierarchical approach, it is important to visualize and understand the impact of some limits of available approaches for modeling the 3D heterogeneity of the Herten site.

Several studies have used the Herten dataset (see companion paper, Bayer et al., 2011). Among them, the most interesting for our purposes are the works of Maji (2005) and Maji et al. (2006). These authors applied a transition-probability-based Markov chain approach (TP/MC, Carle, 1996), using the software TProgs, to reconstruct the complete 3D volume and to use it for analyzing the impact of heterogeneity on DNAPL transport. The TP/MC approach accounts for the probability of transition from one hydrofacies to another with the use of a transition probability matrix. This matrix is composed of transition probability curves which are adjusted from the data contained in the available sections; the 3D simulation is performed using all the sections as conditioning data.

Two slices within a 3D simulation obtained with the TP/MC approach are displayed in Fig. 2c and d. Fig. 2c shows a slice parallel

and close (0.25 m) to the conditioning data of the first section S_1 (Fig. 2a); Fig. 2d is placed in between S_1 and S_2 , at a distance of one meter from both. These two slices show that the simulation becomes very noisy when the simulated nodes get further away from the conditioning data. More precisely, thin continuous layers, especially in the upper part of the domain, become clearly disconnected, and may have substantial influence on the hydraulic behavior of the whole heterogeneous domain. Even if the transition probability curves shown by Maji (2005) agree rather well with the experimental data, it is possible that better results could be obtained with an improved parametrization within TProgs. Actually, we believe that the main reasons for the problems with the TP/MC simulations are (1) the high non stationarity of the data set and (2) the poor information available for the computation of the probability transition matrix along the direction normal to the outcrops.

As a remedy, the TP/MC approach could be applied, for example, in a sequence stratigraphic framework (Weissmann and Fogg, 1999). However, this would be still constrained by the use of statistical models fitted on experimental curves that are computed considering only two points at a time. To overcome this last limitation, we investigate MP statistics.

The analog sections are only 2D, thus we cannot use them directly as training images to simulate a 3D domain with MP

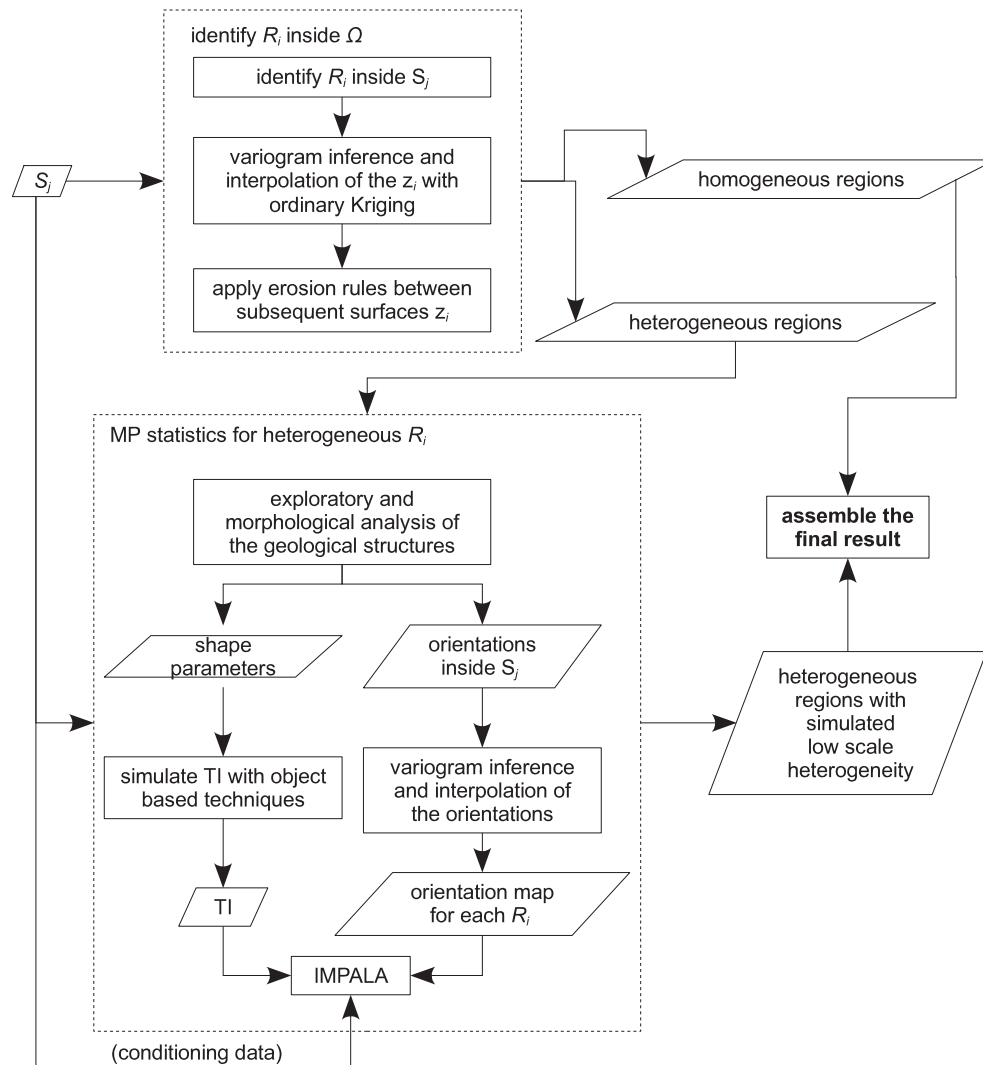


Fig. 3. A schematic representation of the proposed hierarchical framework.

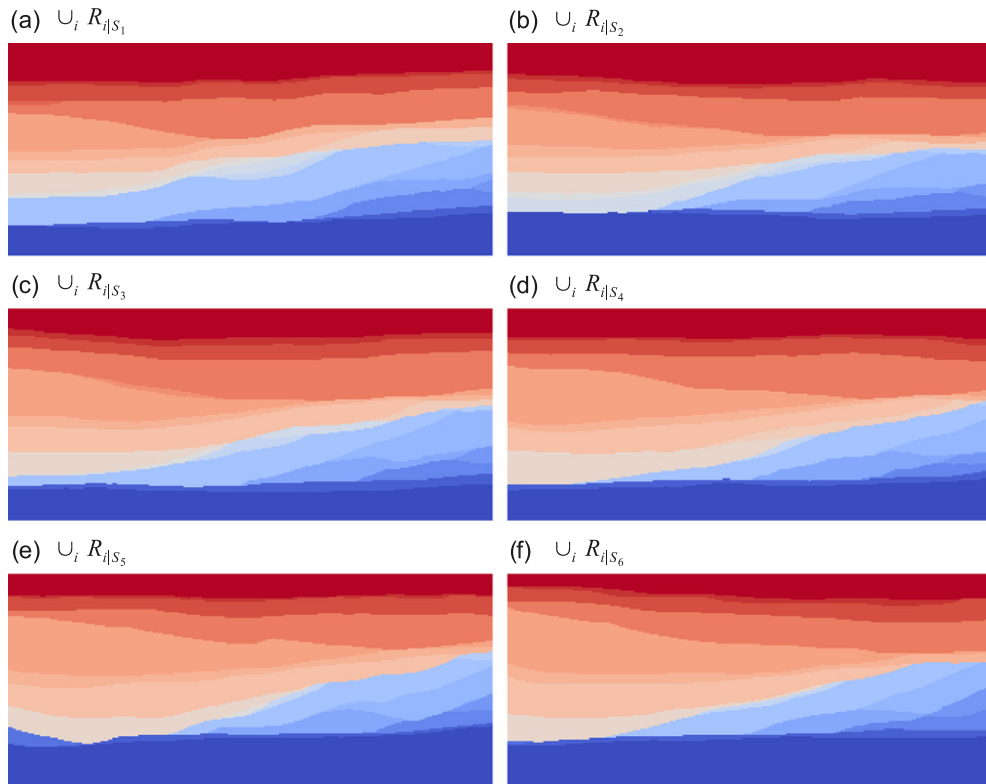


Fig. 4. The different colors represent the regions R_i identified inside each section S_j from the geological sections displayed in Fig. 1. For a color-region correspondence see Fig. 7c.

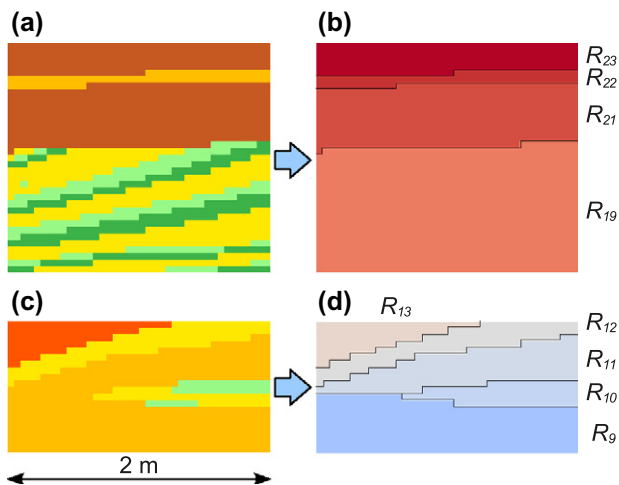


Fig. 5. Procedure adopted to identify the regions R_i inside subsets of the section S_1 . Lower part of (a) and (b): R_{19} is created aggregating neighboring cells which contain small-scale architectural elements. Upper part of (a) and (b): R_{21} , R_{22} , and R_{23} are created aggregating homogeneous layers. (c) and (d) Two regions (in this case R_9 and R_{13}) are created in order to facilitate the interpolation of their upper limit z .

statistics. However, one of these sections (i.e. S_1) can be used as training image to perform 2D unconditional MP statistics simulations. Fig. 2e is a simulation obtained without taking into account the non-stationarity of the training image. As shown by Chugunova and Hu (2008) and Boucher (2009), all the patterns present in the training image are mixed up in the simulation and the architectural elements contained in the section are not reproduced.

To overcome this, one can use auxiliary variables to describe the trends in the training image and in the simulation domain

(Chugunova and Hu, 2008; Straubhaar et al., 2011). Briefly, an auxiliary variable can be represented by a continuous field superimposed to the training image and which is known in the simulation domain. This field is used to constrain MP statistics: the patterns which are observed in the training image, only for a given value of the auxiliary variable field, are reproduced in the simulation domain only for the corresponding values of the auxiliary variables (with a given tolerance). Here, for example, the training image (Fig. 2a) contains an evident vertical trend along the z . The results obtained with the MP statistics using z as auxiliary variable (Fig. 2f) are remarkably improved if compared with the ones obtained without the use of the auxiliary variable (Fig. 2e). Probably, one could go a step further by using a second auxiliary variable along the x coordinate (horizontal) to improve the results. In addition, note that the two 2D MP statistics simulations are obtained without using conditioning data. Even if the simulation of Fig. 2f is rather good, we still note that the thin inter-bedded layer (with hydrofacies code *sGcm*, see companion paper, Bayer et al., 2011) on top of the section is duplicated. This observation reflects the fact that, in general, MP statistics is suitable for reproducing repetitive patterns, but not suited to modeling single discrete features that should rather be modeled with other methods.

Finally, the application of MP statistics on the whole 3D domain remains a problem because there is still a need for a 3D training image or for an improved MP statistics methods allowing 3D volumes to be simulated from 2D training images. An appropriate procedure to overcome this problem is presented subsequently.

Therefore we propose to dichotomize the problem and to model the main layers and the small-scale heterogeneity separately. A first encouraging result of this method is shown in Fig. 2b precisely at the same location as the one shown in Fig. 2d for the TP/MC based approach.

Table 1

The orientation angles used to define the five rotation zones (quantiles 20%, 40%, 60% and 80%) and the five corresponding rotations (quantiles 10%, 30%, 50%, 70% and 90%) associated to each region.

Region	[	Zone 0	)]	Zone 1	)]	Zone 2	)]	Zone 3	)]	Zone 4	]
	min (°)	Q10 (°)	Q20 (°)	Q30 (°)	Q40 (°)	Q50 (°)	Q60 (°)	Q70 (°)	Q80 (°)	Q90 (°)	max (°)
R_1	-17.37	-1.27	-0.25	0.33	0.79	1.20	1.57	2.11	3.00	4.50	24.18
R_4	-15.93	-9.27	-3.12	0.00	2.19	5.48	6.53	7.59	8.38	9.35	14.94
R_5	-28.99	-16.91	-15.27	-12.19	-8.85	-4.89	0.00	0.00	5.07	10.08	16.51
R_6	-2.55	0.00	0.00	0.00	0.00	2.55	5.36	8.71	10.64	11.80	13.23
R_7	-36.87	7.88	9.37	10.20	12.56	14.23	15.62	17.18	19.41	22.28	31.88
R_8	-10.78	4.51	7.29	10.37	13.42	16.19	17.35	19.06	20.48	22.78	30.51
R_9	-6.26	-2.47	-2.33	-2.11	-0.91	0.00	0.00	1.10	3.78	7.53	18.05
R_{10}	-0.95	10.82	13.89	15.11	18.50	19.21	20.64	22.87	24.71	26.79	32.75
R_{13}	-18.55	-1.70	0.00	1.53	3.38	5.18	7.44	10.07	13.45	17.23	23.04
R_{15}	-44.54	-0.20	2.43	4.31	5.98	7.81	9.53	11.48	13.81	17.24	44.48
R_{17}	-17.69	-11.26	-9.21	-7.69	-6.50	-6.04	-5.66	-5.22	-4.33	-2.44	4.26
R_{19}	-12.56	-1.96	0.58	2.31	3.69	4.97	6.29	7.72	9.77	13.87	38.00

Table 2

The arithmetic mean of the z_i of the regions R_i for each section S_j (in meters; here the values of z_2 are not shifted). The last column contains the shift between the second section and the mean of the first and the third section. The bold value at the bottom of the last column is the mean of the column values. The values of z_{23} coincide with the upper part of the simulation domain (7 m).

i	$z_{i S_1}$	$z_{i S_2}$	$z_{i S_3}$	$z_{i S_4}$	$z_{i S_5}$	$z_{i S_6}$	$(z_{i S_1} + z_{i S_3})/2 - z_{i S_2}$
1	1.10	1.64	1.03	1.17	1.58	1.54	-0.58
2	1.23	1.76	1.26	1.35	1.69	1.70	-0.51
3	1.23	1.76	1.26	1.35	1.74	1.71	-0.51
4	1.41	1.92	1.37	1.42	1.83	1.82	-0.53
5	1.42	1.94	1.40	1.45	1.86	1.89	-0.53
6	1.72	2.10	1.57	1.66	2.03	2.04	-0.46
7	1.85	2.35	1.85	1.79	2.25	2.23	-0.50
8	2.76	2.69	2.43	2.28	2.64	2.71	-0.10
9	2.78	2.71	2.44	2.29	2.66	2.71	-0.10
10	2.83	2.75	2.50	2.31	2.69	2.75	-0.09
11	2.90	3.02	2.56	2.31	2.69	2.75	-0.28
12	2.99	3.10	2.60	2.31	2.70	2.75	-0.31
13	3.22	3.36	2.83	2.74	3.03	3.02	-0.34
14	3.38	3.44	2.89	2.88	3.03	3.02	-0.30
15	3.68	3.83	3.49	3.45	3.76	3.67	-0.24
16	3.94	4.05	3.70	3.66	4.03	3.93	-0.23
17	4.29	4.51	4.26	4.44	4.92	4.85	-0.23
18	4.29	4.55	4.30	4.44	4.92	4.85	-0.25
19	5.22	5.54	5.44	5.43	5.70	5.69	-0.21
20	5.23	5.54	5.44	5.43	5.70	5.69	-0.20
21	5.74	5.93	5.90	5.95	6.20	6.18	-0.11
22	5.86	6.07	6.10	6.06	6.28	6.32	-0.09
							-0.30

4. The hierarchical simulation framework

Fig. 1 exhibits large-scale structures composed of sub-horizontal layers with internal small-scale heterogeneities. It becomes natural to adopt a hierarchical approach in order to simulate the two structures separately.

Fig. 3 describes the major steps of the proposed framework, which are explained in the following.

The first major step is to recognize the main layered large-scale structures inside the sections, and extend the large-scale structures to the 3D simulation domain with this procedure:

1. define the hierarchy among structures and identify in each section separate regions R_i
2. infer variogram and interpolate with ordinary Kriging, in the whole 3D domain, the separation surfaces between two adjacent regions
3. create the large-scale layered structure, applying given erosion rules on the interpolated surfaces.

At the end of this step, the simulation domain is divided into a number of homogeneous and heterogeneous regions.

The second major step concerns the heterogeneous regions only. It consists in applying the MP statistics method to simulate, at small scale, the fine heterogeneous structures:

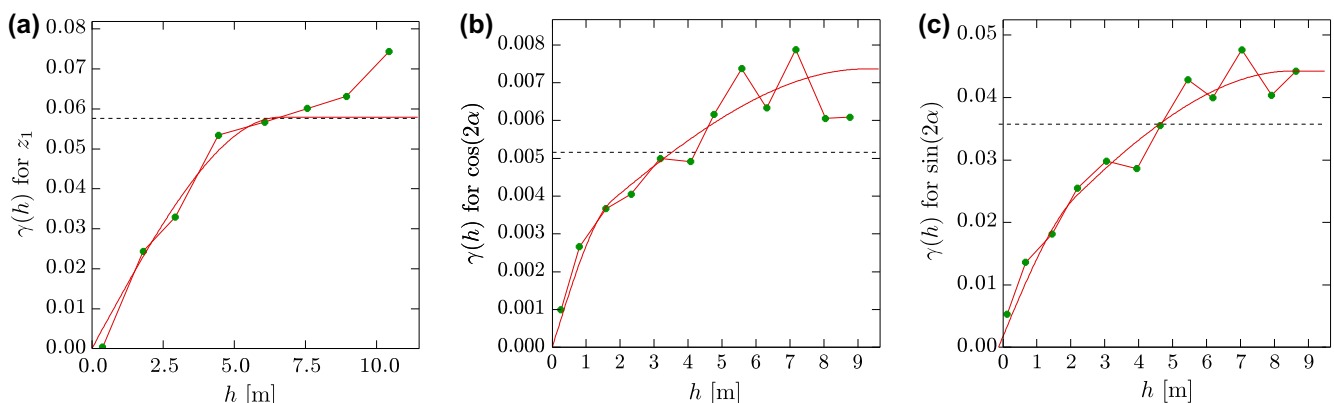


Fig. 6. The variograms adopted to interpolate (a) the top of region R_1 ; (b) $\cos(2\alpha)$ and (c) $\sin(2\alpha)$ for R_{19} . Here α is the orientation of the objects in the plane xz ($\alpha = 0$ on the x axis and positive counterclockwise).

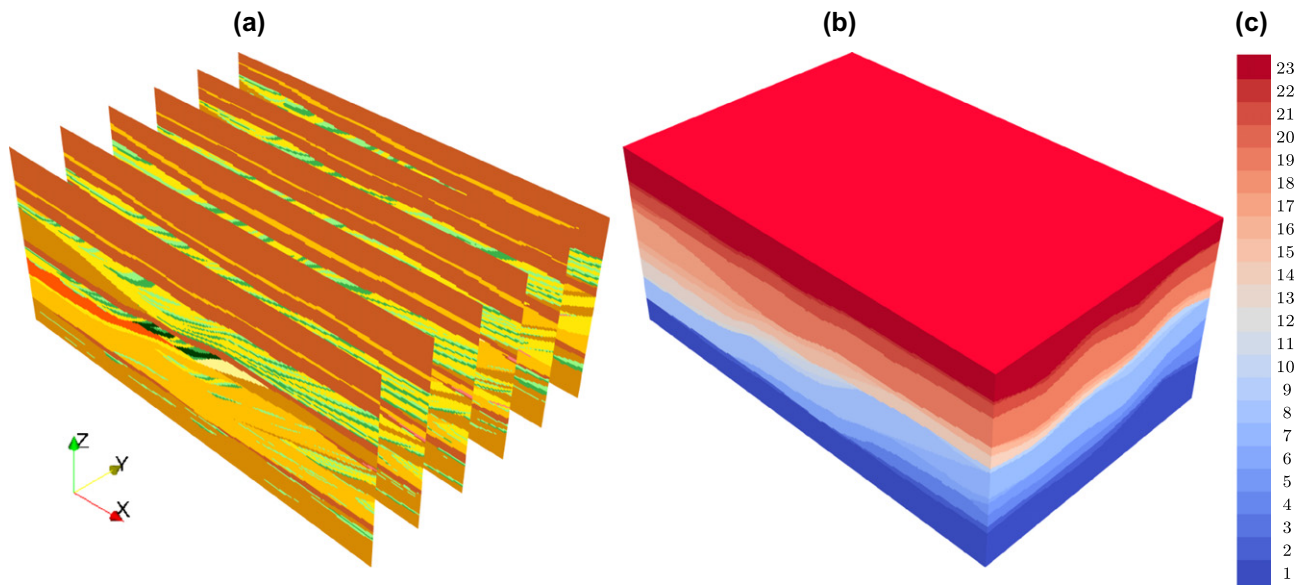


Fig. 7. (a) The six sections S_j of the Herten data set and (b) the layered structure obtained after the interpolation of the upper limit of the 23 regions R_i .

Table 3
The geometrical parameters used to simulate with an object based method the training images required to apply the MP simulation method in the heterogeneous regions R_i . For all the objects $\lambda = 0.5$.

Region	r_e min (m)	r_e max (m)	r_p min (m)	r_p max (m)
R_1	0.55	3.50	0.05	0.35
R_4	0.75	3.50	0.05	0.20
R_5	0.75	3.50	0.05	0.20
R_6	0.75	1.50	0.05	0.10
R_7	2.00	5.00	0.05	0.30
R_8	1.00	4.00	0.05	0.20
R_9	1.50	3.50	0.05	0.15
R_{10}	1.50	3.50	0.05	0.20
R_{13}	1.50	3.50	0.05	0.20
R_{15}	1.50	3.50	0.20	0.75
R_{17}	1.50	5.00	0.10	0.20
R_{19}	3.00	7.50	0.10	0.30

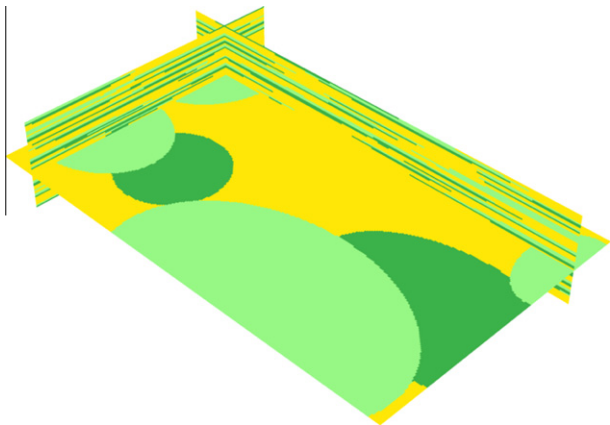


Fig. 9. The 3D training image used to simulate with MP statistics the region R_{19} .

1. exploratory and morphological analysis of the shapes contained in each heterogeneous region, in order to get the geometrical parameters suited for the application of object based simulation methods
2. simulate suitable 3D training images with an object based technique
3. interpolate in all the R_i the orientation of the small-scale heterogeneities, in order to create a map to be used as input for the MP statistics simulation technique
4. simulate internal heterogeneity within each region with the MP statistic using the conditioning data provided by the outcrops and the orientation fields
5. assemble the results.

The heterogeneous regions simulated in this second step are then included in the 3D domain.

4.1. Subdivide the domain into large scale regions

The first step of the hierarchical framework consists of partitioning the whole domain Ω in a number of regions R_i such that $\Omega = \cup_i R_i$ and $R_i \cap R_j = \emptyset$ when $i \neq j$.

The regions R_i are first identified manually inside the sections S_j ($R_{i|S_1, \dots, S_6}$, see Fig. 4). The genetic units defined in the companion paper (Bayer et al., 2011) are used as guidelines for the definition of the regions, which are identified following one of the two following rules. The first rule is used when the distribution of the hydrofacies presents decimeter-scale heterogeneities (see lower part of Fig. 5a). With this rule, the cells are clustered in order to form a unique region; this is the case of R_{19} in Fig. 5b. The regions defined with this rule are heterogeneous.

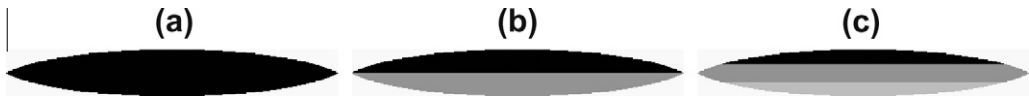


Fig. 8. The slices of the three kinds of flattened ellipsoids, used for the boolean simulation of the TIs, composed by (a) one, (b) two and (c) three facies.



Fig. 10. The hydrofacies proportions computed for each heterogeneous region R_i . For each region, the first bar on the left represents the proportions computed on the conditioning data, the second the proportions on the training image I_i used for the simulation, the third and the fourth the proportions of two realizations of a conditional MP simulation.

The second rule is used when homogeneous large-scale layers or architectural elements are observed. This is for example the case of the genetic unit IV (see companion paper, Bayer et al., 2011), where three homogeneous regions R_{21} , R_{22} and R_{23} are identified (Fig. 5a and b).

A third rule was added in order to facilitate the interpolation procedure: the identified regions R_i must not overlap. This is illustrated in Fig. 5c and d, where following only the two rules presented before the regions R_9 and R_{11} would have been considered as a whole; the upper limit z of this last region would not have been uniquely defined by its x and y coordinates, creating a problem for the subsequent interpolation procedure.

Adding this last rule, the relation among the upper limits z of the identified regions is the following:

$$z_{i|S_1, \dots, S_6} \geq z_{i-1|S_1, \dots, S_6} \quad \text{for } i = 2, \dots, n_R - 1. \quad (1)$$

In the Eq. (1) n_R stands for the number of identified regions R_i ; the vertical z axis is pointing upwards and the index increases from lower z to higher z .

Eleven homogeneous regions and twelve heterogeneous regions $R_{i|S_1, \dots, S_6}$ are identified (in total, $n_R = 23$) following the rules illustrated before. The results of this identification procedure are depicted in Fig. 4.

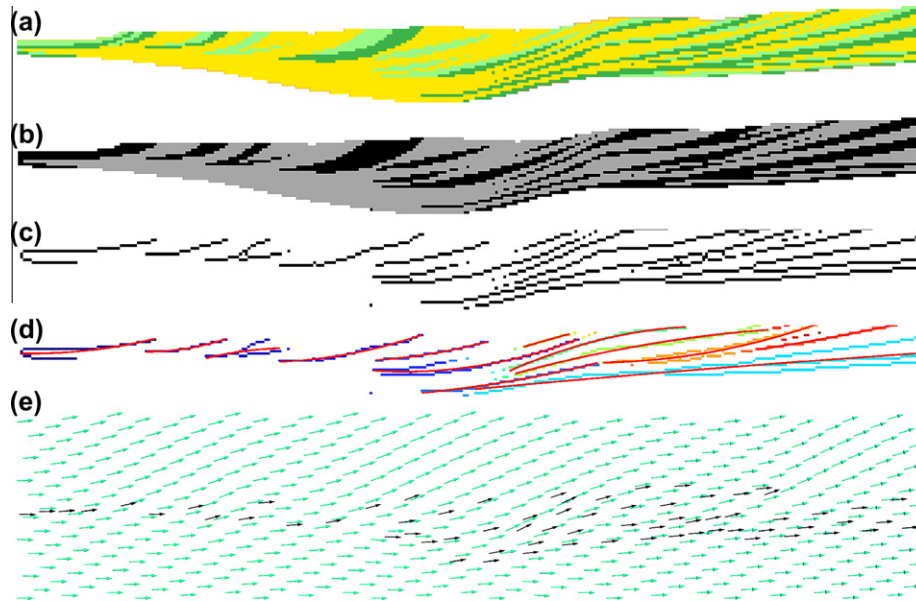


Fig. 11. Part of the steps performed to compute an orientation map for region R_{19} and the section S_1 : (a) $R_{19} \cap S_1$ (b) the objects inside the matrix are identified; (c) objects are reduced into their skeleton; (d) the skeletons of distinct objects are labeled (with distinct colors), and polynomials of order two are fitted on the skeletons (red lines); (e) the orientations are deduced from the derivatives of the polynomials (dark arrows). The bright arrows in (e) are a slice of the 3D interpolation of the orientations.

A systematic shift of the heights $z_{i|S_2}$, especially for the lower regions (R_i with $i \leq 7$) was identified by preliminary interpolations (see Table 2). Due to the impossibility of investigating the reason for this systematic shift in depth, we decided to subtract its mean value (0.3 m) from all the $z_{i|S_2}$ (for more details see Table 2). Another possibility would be to ignore the z values provided by the second section during the interpolation process, for example for all $z_{i|S_2}$ with $i \leq 7$. However, we preferred to apply the shift to all the $z_{i|S_2}$ in order to keep all the data for the interpolation.

Structural analysis is performed on the corrected heights, for each region, in order to fit a variogram model and interpolate the z_i with ordinary Kriging on Ω . The variogram model adopted for z_1 is depicted in Fig. 6a.

To avoid overlap among the interpolated surfaces, some erosion rules are applied according to Eq. (1). The z_i interpolated are corrected taking as reference the lower surface, starting from $i = 1$, with the following rule:

$$z_i = \begin{cases} z_i & \text{if } z_i \geq z_{i-1} \\ z_{i-1} & \text{if } z_i < z_{i-1} \end{cases} \quad \text{for } i = 2, \dots, n_R - 1, \quad (2)$$

Fig. 7b represents the layered structure obtained in this first reconstruction step.

4.2. Small scale: MP statistics

While eleven regions of the layered structure obtained by interpolation can be considered homogeneous, the other twelve regions R_i with $i = 1, 4, 5, 6, 7, 8, 9, 10, 13, 15, 17, 19$ have a heterogeneous internal structure. Their structure is simulated using MP statistics with an individual training image I_i for each region.

The following sections describe the procedure adopted to generate the training images required to apply MP statistics. Also, the steps required to take into account the non stationarity of

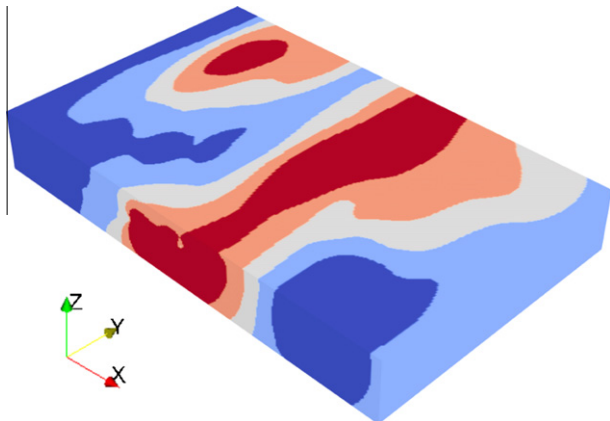


Fig. 12. The five orientation zones defined to take into account for the non stationarity in the orientation of the geological structures for region R_{15} .

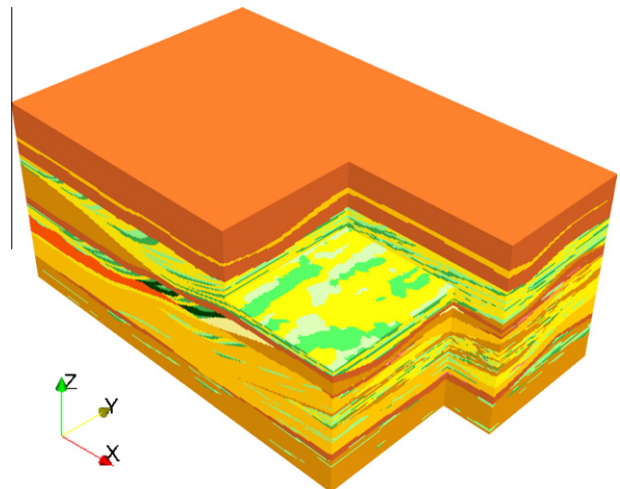


Fig. 13. One realization obtained with the hierarchical simulation framework taking into account the conditioning data.

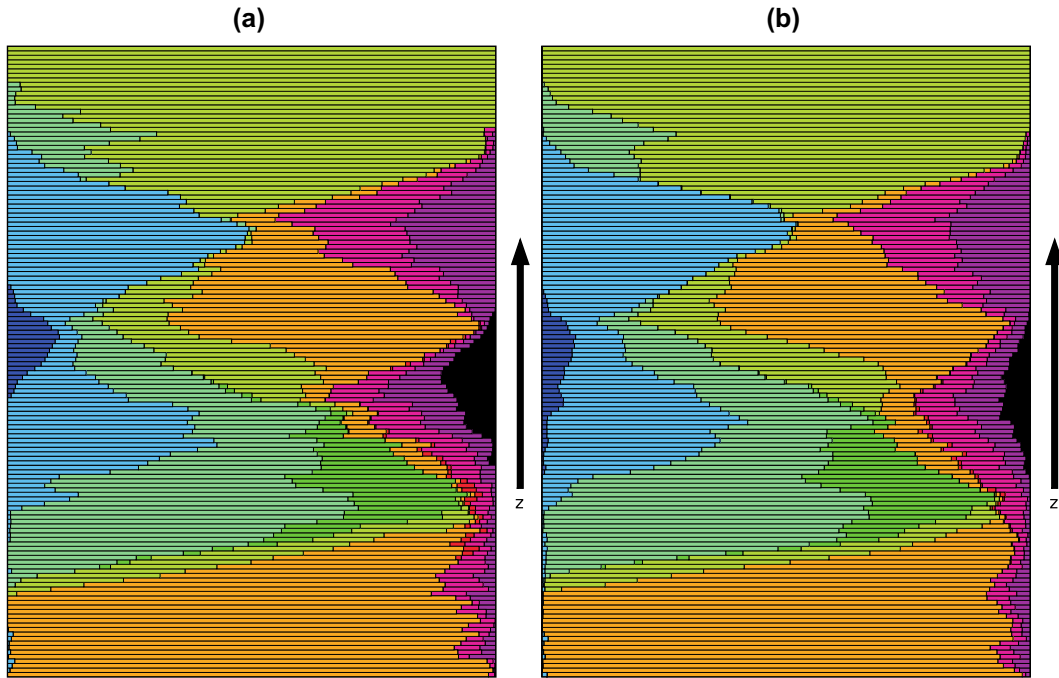


Fig. 14. The hydrofacies proportions computed on horizontal slices (parallel to xy plane) of 3D simulations. Slices are computed for each discretization step along the axis z (140 slices in total). Proportions on slices are computed (a) on the conditioning data and (b) on the 3D simulation of Fig. 13.

the orientations of the geological structures in MP simulations are reported. The following procedure is applied to each heterogeneous region.

4.2.1. Simulate the missing training images

At the scale of the regions R_i one can recognize geological structures inside the six 2D sections $R_{i|S_{1...6}}$ and deduce some 2D training images. However, 3D training images are required in order to apply MP statistics methods.

The missing 3D training images are generated using object-based techniques, relying on the information provided by the 2D sections. We suppose that the 3D shapes can be represented as lenses, as in the case study described by Jussel et al. (1994). The objects are generated at random locations with a shape defined by:

$$\frac{(x - x_0)^2 + (y - y_0)^2}{r_e^2} + \frac{(z - z_0 - [z - z_0]\lambda r_p)^2}{r_p^2} = 1, \quad \lambda \in [0, 1]. \quad (3)$$

Expression (3) represents an ellipsoid flattened by a factor λ , centered in the point (x_0, y_0, z_0) with equatorial radii r_e and polar radius r_p . The parameters r_e and r_p are varied with uniform laws $\mathcal{U}([r_{e \min}, r_{e \max}])$ and $\mathcal{U}([r_{p \min}, r_{p \max}])$ for each simulated object; the simulation is stopped when a target proportion is reached.

The values of the parameters $r_{e \min}$, $r_{e \max}$, $r_{p \min}$, $r_{p \max}$ and λ adopted for the boolean simulations are reported in Table 3.

These values are estimated manually by trial and error from the outcrops instead of using an automated procedure which would have been required to account for the limited size of the outcrops, for object truncation due to erosion processes and for the non stationarity in orientation of the objects. The facies proportions used as threshold to stop the object simulation procedure are extracted from the data contained in the outcrops. However, inferred 3D object properties from 2D outcrops introduces a bias: for this reason, even if for each heterogeneous region R_i we try to simulate a training image with proportions close to those of the outcrops $R_{i|S_{1...6}}$, we also accept training images simulated with proportions having

the same order of magnitude of the proportions extracted from the sections. For each heterogeneous region R_i , the proportions extracted from the outcrops $R_{i|S_{1...6}}$, the proportions obtained for the training image I_i , and the proportions obtained from two realizations of conditional MP simulations are in agreement (Fig. 10).

The boolean simulation code allows the generation of three different types of flattened ellipsoids having a composition of one, two or three facies (Fig. 8); in fact, we suppose that the shape of the small-scale geological structures which can be elicited from the outcrops can be approximated with these types of ellipsoids. Consequently, the training images for the 12 heterogeneous regions are generated with a combination of the three types, in order to mirror the compositions observed in the sections S_j (see Fig. 9).

4.2.2. Objects orientation

The non stationarity inherent to the layered structure of the site at large scale can be taken into account by subdividing the domain Ω into smaller regions R_i . However, at the scale of the heterogeneous regions R_i , there is another source of non stationarity: the one related to the orientation of the heterogeneities within the region. With the MP statistics simulation algorithm IMPALA (Straubhaar et al., 2011) it is possible to account as well for this type of non stationarity; inside a simulation domain, different zones can be defined where individual simulation parameters can be applied, like a dilatation or a rotation of the training image. In order to perform MP simulation in each heterogeneous region R_i , IMPALA requires a suitable training image and an orientation map for the objects. This section is devoted to this last map.

In the following examples, and for testing the sets of control variables of the MP simulation, we focus on region R_{19} ; in fact this region groups a great number of features whose reproduction are a challenge in the 3D simulations, like, for example, extended continuous layers.

Fig. 11 depicts the main steps performed in order to detect the orientations of the objects. The steps are implemented in Matlab and the figure reports as an example the case of region R_{19} inside

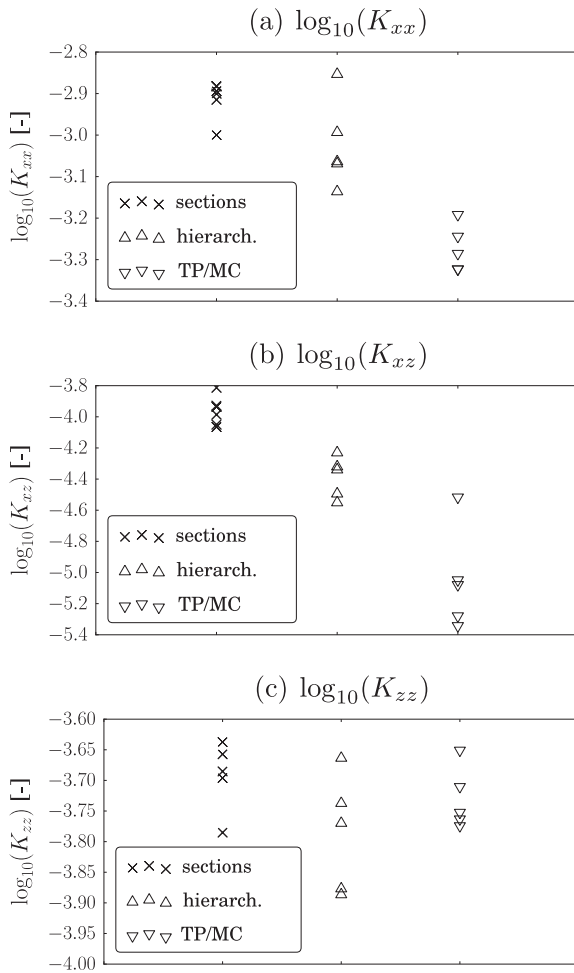


Fig. 15. The $\log_{10}$ of the components of the equivalent conductivity tensor $\mathbf{K}$. (a) K_{xx} , (b) K_{xz} and (c) K_{zz} are computed on the six sections of the data set, on five 2D slices in the 3D simulation obtained with the hierarchical approach and on five 2D slices obtained with the TP/MC approach.

section S_1 . Because we only have access to sections along the outcrop plane xz , we can only determine orientations on this plane.

For each heterogeneous region $R_i | S_1, \dots, S_6$ the orientations are computed with the following procedure:

- Distinguish the matrix and the objects. For example the hydrofacies $G_{cg,o}$ and $sG_{cg,o}$ can be considered as forming unique objects inside a matrix of $sG_{cm,b}$ (Fig. 11a and b).
- Apply a morphological operation in order to reduce the objects into thin shapes (Fig. 11c).
- Label each reduced object using the connectivity rule described by Haralick and Shapiro (1992) and considering eight neighboring points (Fig. 11d).
- Fit a polynomial (red lines,¹ Fig. 11d) for each labeled object with a least squares method. Use the derivative of the polynomial to get the orientation value α in each point of the object skeleton (dark arrows, Fig. 11e).

4.2.3. Orientation interpolation

To define the 3D orientation maps required by the MP statistics simulations the punctual values of orientations computed along

the skeleton of the small-scale architectural elements have to be interpolated.

Interpolating local directional data is a challenge (Lajaunie et al., 1997; Gumiaux et al., 2003). The main problem is that orientation data are circular when they are represented as angles. They only vary between 0° and 360° , and the value 0° is closer to 350° than the value 20° , for example. A change of variable must be used to ensure that the interpolation is made properly. In geostatistics, several approaches have been designed for the interpolation of orientation data. Lajaunie et al. (1997) and Chilès and Delfiner (1999), chapter 5.5 observe that orientation data can be represented as the gradient of a scalar field; starting from this consideration, they propose using general co-Kriging equations to interpolate the scalar field and map the orientations. Gumiaux et al. (2003) provide a review of the main methods and also propose a simple solution to the problem. Instead of directly interpolating the orientation angle α , they recommend interpolating $\cos(2\alpha)$ and $\sin(2\alpha)$; then the values of α on the whole interpolation domain are computed with $\arctan(\sin(2\alpha)/\cos(2\alpha))/2$. In this work, due to the small range of variation of the orientation of the object, we apply this last approach.

Structural analysis is performed on $\cos(2\alpha)$ and $\sin(2\alpha)$ for each heterogeneous region R_i . Here we report as examples the variogram models adopted to interpolate with ordinary Kriging $\cos(2\alpha)$ and $\sin(2\alpha)$ for R_{19} (Fig. 6b and d). The bright arrows of Fig. 11e represent a slice of the orientations interpolated for R_{19} .

For each heterogeneous region R_i , the interpolated orientations are grouped in five orientation classes, in order to define five zones in which to apply the MP statistic method with the same training image but with different orientations. The classification of the orientations is done according to the quantiles: a quantile step of 10% is computed, and while the 0%, 20%, 40%, 60%, 80% and 100% quantiles are adopted as bounds for the five classification intervals, the 10%, 30%, 50%, 70%, and 90% quantiles are used as the orientations to characterize each zone. Table 1 contains all the values adopted for the previous classifications. Fig. 12 represents the five orientation zones used to apply MP statistics to region R_{15} .

A test performed on R_{19} using ten orientation classes (quantile step of 5%) did not improve the simulation result. Therefore, the classification using five classes is considered sufficient.

4.2.4. MP statistics simulation

Selecting a suitable set of control variables of the MP statistics simulation is a crucial point. Here we test different sets with a “trial and error” procedure, following the guidelines provided by Liu (2006) and guided by our experience. A large part of the tests are performed only on region R_{19} .

An important question is how to evaluate and compare the results obtained with different sets of variables. In this context the results of the simulations are evaluated for their ability to reproduce, first, the observed hydrofacies proportions, and, second, the continuity of the structures, which can be deduced from visual inspection of the sections S_j .

For a detailed explanation of the variables tested and their influence on the simulated results, refer to Liu (2006) and Straubhaar et al. (2011); here after we report only a brief overview of the variables considered in the tests. The first variables inspected are the number of multi-grid levels and the size of the search neighborhood. With a big search neighborhood it is possible to capture large-scale features of the training image, but an increased neighborhood requires more RAM and more CPU time. Increasing the multi-grid level allows the reproduction of large-scale features with a smaller search neighborhood. Here we test neighborhood of size $3 \times 3 \times 3$, $5 \times 5 \times 5$ and $7 \times 7 \times 7$ and multi-grid levels from 1 to 6.

¹ For interpretation of color in Figs. 1, 2, 4–7, 9–14, 16 and 17, the reader is referred to the web version of this article.

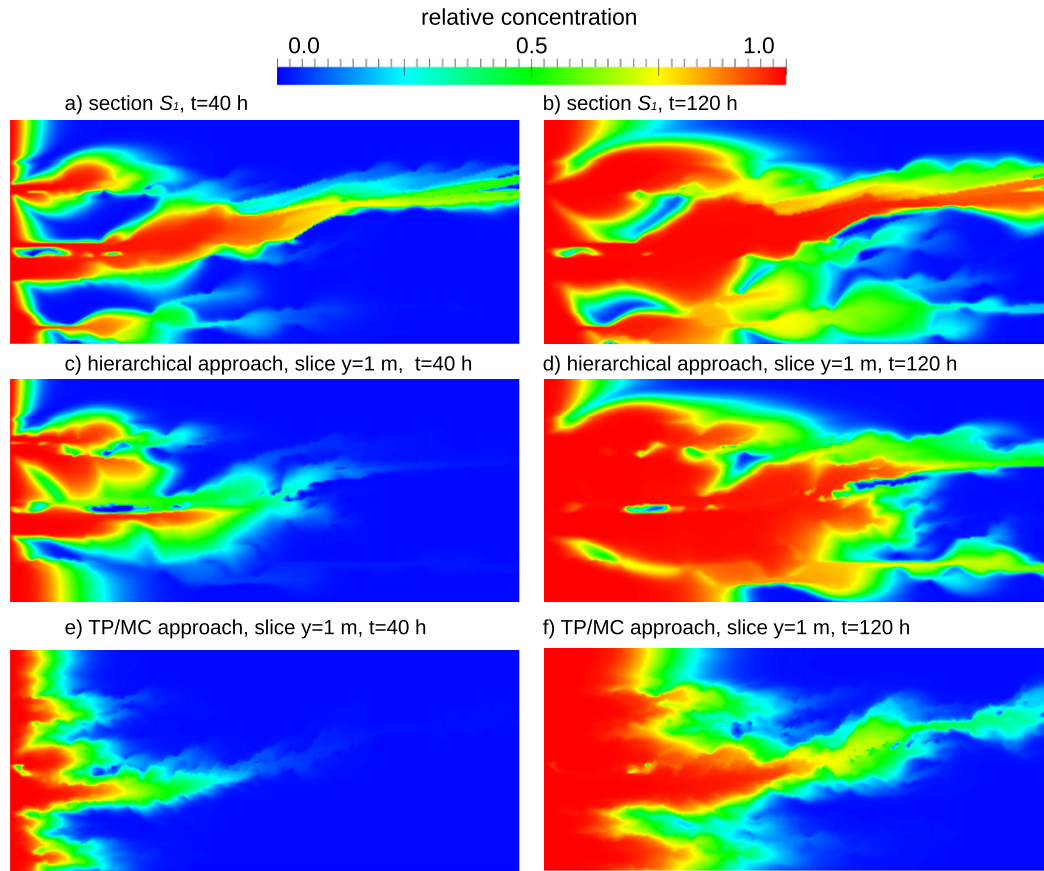


Fig. 16. The relative concentration after 40 h (left column) and 120 h (right column) for the section S_1 (a,b), a slice parallel to S_1 and located between S_1 and S_2 ($y = 1$) within the 3D simulation obtained with the hierarchical approach (c,d), or obtained with the TP/MC approach (e,f).

Another control variable of interest for the MP statistics is the simulation path. Depending on the disposition of the conditioning data and on the shapes which have to be reproduced, this variable can influence the simulation result. The simulation paths tested in this case study are the following: a random path; a path performed varying the coordinates of the simulation grid in the order x , y and z (“unilateral” path); a path in which the grid nodes closer to the conditioning data are simulated first.

5. Results and discussion

Fig. 13 illustrates one realization obtained with the hierarchical simulation framework. Different realizations are obtained with the same set of control variables, while using a different seed for the random MP statistics simulations. The control variables used are 5 multi-grid levels along the directions x and y and 1 multi-grid level along z , a random simulation path and a search neighborhood of size $5 \times 5 \times 5$. Using this set of variables the realizations obtained reproduce reasonably well the complex structure of the medium given by the six vertical sections of Fig. 1. Thin layers are now continuous over most of the domain. Internal heterogeneity within a region is constrained within the region. Moreover, the variation of the hydrofacies proportion computed along the vertical axis z for the simulated domain and the conditional data are in good agreement (Fig. 14).

These results are closer to the observation and much less noisy than previous simulations made using transition probability techniques (Fig. 2c and d).

Two numerical flow simulation tests are performed in order to corroborate the results of the visual comparisons and of the comparisons on the proportions.

The first tests are performed under steady-state flow conditions. The values of hydraulic conductivity provided in the companion paper (Bayer et al., 2011) for each hydrofacies are used. Two sets of linearly varying head boundary conditions are employed. The first is used to prescribe a hydraulic gradient along the horizontal direction and the second one to prescribe a hydraulic gradient along the vertical direction. The flow is computed using the Groundwater finite element code (Cornaton, 2007) and post-processed to compute the equivalent hydraulic conductivity tensor $\mathbf{K}$ (Rubin and Gómez-Hernández, 1990; Renard et al., 2001) of the heterogeneous medium. $\mathbf{K}$ is computed for the six sections of the data set, for five 2D slices in the 3D simulation obtained with the TP/MC approach, and for five 2D slices in the 3D simulation obtained with the hierarchical approach. These slices are parallel to the planes of the sections and located in between them, that is for $y = 1, 3, 5$ and 7 m. K_{xx} and K_{xz} are the components of $\mathbf{K}$ which provide indications about the continuity of the lithological layers observed in the sections along the direction x . We observe that the values of K_{xx} and K_{xz} computed from the hierarchical approach are closer to the reference than the values computed from the TP/MC approach (Fig. 15a and b). For the component K_{zz} , taking into account the limited range of variations of the estimated values, we consider that the results from the two simulation techniques are not significantly different (Fig. 15c).

The second tests are based on transport simulations performed with the Groundwater code on the same sections. The values of

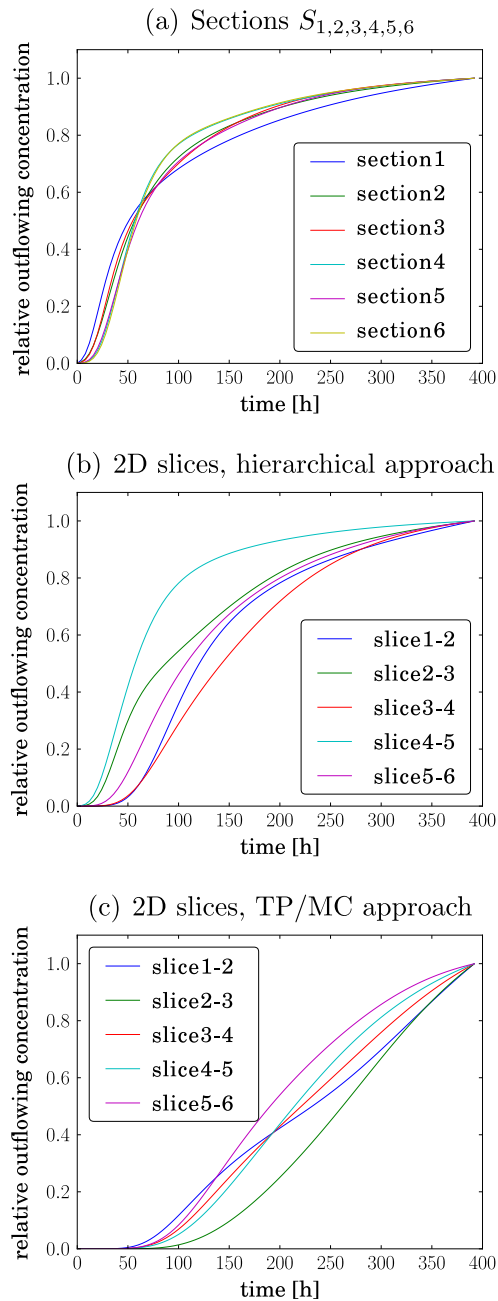


Fig. 17. Contaminant breakthrough curves for (a) the section S_1 , (b) a slice on the 3D simulation obtained with the hierarchical approach, parallel to S_1 and in between of S_1 and S_2 , (c) a slice on the 3D simulation obtained with the TP/MC approach in the same position as the slice (b).

hydraulic conductivity and porosity provided in the companion paper (Bayer et al., 2011) for each hydrofacies are used, together with coefficients of longitudinal and vertical dispersivity of respectively 10 cm and 1 cm. The solute transport problem considers a horizontal flux. It is computed by prescribing a fixed head on the two opposite faces of the section and no flow boundaries on top and bottom, a concentration of 1 is prescribed at the inlet of the system at $x = 0$ m at $t = 0$ (Fig. 16). The evolution of the average concentration is recorded at the outflow boundary (Fig. 17). Fig. 17 shows that the contaminant breakthrough curves for the slices obtained with the hierarchical approach (Fig. 17b) are closer to the reference data set (Fig. 17a) than those obtained with the TP/MC approach (Fig. 17c). The same considerations can be applied to the results

shown in Fig. 16, where the relative concentration at two different time steps in the section S_1 and in one 2D slice parallel to S_1 at $y = 1$ m is extracted from the 3D simulations obtained with the hierarchical and the TP/MC approaches.

In general these simulations show that the hierarchical approach better reproduces the flow and transport properties of the heterogeneity observed in the outcrops than the TP/MC method. This is certainly related to the fact that the proposed method produces more continuous layers in the horizontal direction which better ensures the continuity of the sedimentary structures.

More generally, the hierarchical approach described in this paper allows the use of MP simulation tools with simple training images generated using well known boolean simulation techniques. This is an important advantage because the accessibility of a suitable training image, especially in 3D cases, is often a problem. At the same time, while the small-scale intricacy is reproduced using MP methods, the large-scale structures are modeled using standard geostatistical methods such as ordinary Kriging.

With the simplification introduced by the hierarchical framework, it is possible to take into account the different aspects of the non stationarity which characterizes complex sites (like i.e. the Herten case study). One aspect of the non stationarity is tackled by splitting the domain in sub-horizontal regions R_i . Another aspect of the non stationarity is related to the orientation of the geological structures contained in the heterogeneous regions R_i . The hierarchical approach allows one to apply MP statistics algorithms which can handle this last aspect. Again, within this hierarchical framework a simple procedure is developed to detect and interpolate spatial orientation of the geological structures.

Another problem faced in this study is the choice of the best MP statistics simulation parameters. A number of tests were performed with different sets of parameters for the region R_{19} . They show that selecting a big search neighborhood and a high number of multi-grid levels does not always coincide with a better simulation result: often it is possible to obtain good results at moderate computational effort burden with a search neighborhood with a radius of three grid cells and five multi-grid levels. Moreover, since the continuity of the geological structures is mainly horizontal, along the direction z one multi-grid level provided satisfactory results.

In regards to the simulation path, this study, in which conditioning data are not sparse, shows that a simulation path based on the distance from the simulated point and the conditioning data is not ideal. The main structures are simulated starting from the vertical conditional sections, resulting in an artificial lack of continuity between the two planes. The best results are obtained with a random simulation path.

6. Conclusions

With this work, we have shown that hybrid or hierarchical models allow their users to circumvent most difficulties (non stationarity and lack of a 3D training image) encountered when dealing with multi-scale data such as the Herten site. With the advent of MP statistics, there was a temptation to think that all types of geological structures could be modeled with a single tool provided that an adequate training image is available. It is clear that progress has been made in that direction, but we also see that it is often a better strategy to combine different techniques.

One should note that splitting the domain into regions is something that is done routinely when building 3D geological models. Here the difference is that we apply the same principle at a small scale and within a single geological formation to describe different scales of internal structures within the formation.

The 3D model of the Herten site is provided as [electronic supplementary material](#) that can be used in the future in several ways.

Indeed, the 3D heterogeneity of the model is certainly close to the one which can be observed in this type of fluvio-glacial aquifers and at that scale. Therefore, the model could be used to investigate physical processes that are controlled by this heterogeneity (see e.g. Maji et al., 2006). A whole range of numerical experiments could be done.

Another possibility is to use some statistical information that is derived from the sections (for example the number of regions per unit thickness, the variograms of the boundaries between the regions, the type of orientation structures, etc.) to extend the dimension of the 3D model. Indeed, it is possible to simulate successively an ensemble of random surfaces following the statistical parameters and the erosion rules inferred by the dataset to build an unconditional partition of a large domain. If borehole data are available, one could also simulate these surfaces conditional to the borehole data. The surfaces can be used to define separate regions where one could simulate orientation fields based on the orientation observed in the data set. The orientation fields can then be used to simulate pertinent geological structures with MP statistics (with the training images derived as depicted before).

The last point that we want to stress is that analog data sets almost always provide only 2D information. This fact clearly emphasizes the need to extend the MP statistics framework in order to allow simulating 3D domains from 2D training images and a set of reasonable assumptions in the future. The presented procedure demonstrates a successful first step in this direction.

Acknowledgments

Funding for this research was provided by Swiss National Science foundation (Grants PP002-106557 and PP002-124979) and the Swiss Confederation's Innovation Promotion Agency (CTI Project No. 8836.1 PFES-ES). The authors are also grateful to G. de Marsily, G. Caumon, D. Allard, M. Giudici and two anonym reviewers who provided useful suggestions for the improvement of the manuscript. Special thanks to F. Cornaton who made the finite element code Groundwater available for the study.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jhydrol.2011.03.037](https://doi.org/10.1016/j.jhydrol.2011.03.037).

References

- Bayer, P., Huggenberger, P., Renard, P., Comunian, A., 2011. Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study. *Journal of Hydrology* 405 (1–2), 1–9.
- Boucher, A., 2009. Considering complex training images with search tree partitioning. *Computers & Geosciences* 35 (6), 1151–1158.
- Caers, J., 2005. *Petroleum Geostatistics*. Society of Petroleum Engineers.
- Carle, S., 1996. A transition probability-based approach to geostatistical characterization of hydrostratigraphic architecture. Ph.D. thesis, Department of Land, Air and Water Resources, University of California Davis.
- Chilès, J., Delfiner, P., 1999. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.
- Chugunova, T.L., Hu, L.Y., 2008. Multiple-point simulations constrained by continuous auxiliary data. *Mathematical Geosciences* 40 (2), 133–146.
- Cornaton, F., 2007. *Ground Water: A 3-D Ground Water Flow and Transport Finite Element Simulator*. Reference manual. Tech. rep., University of Neuchâtel, Switzerland.
- de Marsily, G., Delay, F., Goncalves, J., Renard, P., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. *Hydrogeology Journal* 13 (1), 161–183.
- de Vries, L.M., Carrera, J., Falivene, O., Gratacos, O., Luit, J.S., 2009. Application of multiple point geostatistics to non-stationary images. *Mathematical Geosciences* 41 (1), 29–42.
- Guardiano, F.B., Srivastava, R.M., 1993. Multivariate geostatistics: beyond bivariate moments. In: Soares, A. (Ed.), *Geostatistics: Troia '92*, vol. 1. Kluwer, Dordrecht, The Netherlands, pp. 133–144.
- Gumiaux, C., Gapais, D., Brun, J., 2003. Geostatistics applied to best-fit interpolation of orientation data. *Tectonophysics* 376 (3–4), 241–259.
- Haralick, R.M., Shapiro, L.G., 1992. *Computer and Robot Vision*, vol. 1. Addison-Wesley.
- Hu, L.Y., Chugunova, T.L., 2008. Multiple-point geostatistics for modeling subsurface heterogeneity: a comprehensive review. *Water Resources Research* 44 (W11413), 14.
- Huysmans, M., Peeters, L., Moermans, G., Dassargues, A., 2008. Relating small-scale sedimentary structures and permeability in a cross-bedded aquifer. *Journal of Hydrology* 361 (1–2), 41–51.
- Jussel, P., Stauffer, F., Dracos, T., 1994. Transport modeling in heterogeneous aquifers: 1. Statistical description and numerical generation of gravel deposits. *Water Resources Research* 30 (6), 1803–1817.
- Khan, D., Strebelle, S., Hanggoro, D., Willis, B., Pyrcz, M., 2008. Non-stationary multiple point simulation to model complex deltaic deposits for flow simulation. In: Ortiz, J., Emery, X. (Eds.), *Geostatistics Santiago 2008*. Springer, Netherlands, pp. 217–226.
- Koltermann, C.E., Gorelick, S.M., 1996. Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research* 32 (9), 2617–2658.
- Lajaunie, C., Courrioux, G., Manuel, L., 1997. Foliation fields and 3D cartography in geology: principles of a method based on potential interpolation. *Mathematical Geology* 29 (4), 571–584.
- Liu, Y., 2006. Using the Snesim program for multiple-point statistical simulation. *Computers & Geosciences* 32 (10), 1544–1563.
- Maharaja, A., 2005. Stochastic modeling of the Rhine–Meuse delta using multiple-point geostatistics. In: Leuangthong, O., Deutsch, C.V. (Eds.), *Geostatistics Banff 2004*, Mathematics and Statistics, vol. 14. Springer, Netherlands, pp. 653–659.
- Maharaja, A., 2008. Tgenerator: object-based training image generator. *Computers & Geosciences* 34 (12), 1753–1761.
- Maji, R., 2005. Conditional stochastic modelling of DNAPL migration and dissolution in a high-resolution aquifer analog. Ph.D. thesis, University of Waterloo.
- Maji, R., Sudicky, E., Panday, S., Teutsch, G., 2006. Transition probability/markov chain analyses of DNAPL source zones and plumes. *Ground Water* 44 (6), 853–863.
- Mariethoz, G., Renard, P., Straubhaar, J., 2010. The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research* 46 (11), W11536.
- Renard, P., Genty, A., Stauffer, F., 2001. Laboratory determination of the full permeability tensor. *Journal of Geophysical Research* 106 (B11), 26443–26452.
- Rubin, Y., Gómez-Hernández, J.J., 1990. A stochastic approach to the problem of upscaling of conductivity in disordered media: theory and unconditional numerical simulations. *Water Resources Research* 26, 691–701.
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., Besson, O., 2011. An improved parallel multiple-point algorithm using a list approach. *Mathematical Geosciences* 43 (3), 305–328.
- Strebelle, S., 2000. Sequential simulation drawing structure from training images. Ph.D. thesis, Stanford University, Stanford, CA, USA.
- Weissmann, G., Fogg, G., 1999. Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. *Journal of Hydrology* 226 (1–2), 48–65.