Stochastic heterogeneity modeling of braided river aquifers

A methodology based on multiple point statistics and analog data

thesis presented at the

Faculty of Sciences

Institute of Hydrogeology and Geothermics (CHYN)

University of Neuchâtel, Switzerland

for the degree of

Doctor of natural sciences

presented by

Guillaume Pirot

accepted on the recommendation of

Prof. Philippe Renard
Dr. Julien Straubhaar
Prof. Jaime Gomez-Hernandez
Prof. Niklas Linde
Prof. Ghislain de Marsily

Defended on January 14^{th}, 2015

UNIVERSITÉ DE NEUCHÂTEL

Faculté des sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel - Suisse
Tél: + 41 (0)32 718 2100
E-mail: secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur Guillaume PIROT

Titre:

"Stochastic modeling of braided river aquifer heterogeneity"

sur le rapport des membres du jury composé comme suit:

- Prof. Philippe Renard, Université de Neuchâtel, directeur de thèse
- Dr Julien Straubhaar, co-directeur de thèse, Université de Neuchâtel
- Prof. Jaime Gomez-Hernandez, Université Polytechnique de Valence, Espagne
- Prof. Niklas Linde, Université de Lausanne
- Prof. Ghislain de Marsily, Université Paris 6, France

Neuchâtel, le 29 janvier 2015

Le Doyen, Prof. B. Colbois

B. Colbois

Remerciements

Merci de tout coeur, Ghazal, ma femme, de m'avoir révélé ce trésor qu'est ton amour, source d'inspiration et d'énergie. Un grand merci à toi David, de m'avoir mis en contact avec un Directeur de thèse formidable lorsque j'étais mûr pour venir découvrir la recherche. Un très grand merci à Philippe mon Directeur de thèse et à Julien co-Directeur de ma thèse; entre vos conseils, vos idées, votre écoute et votre passion, travailler avec vous est un plaisir sans fin. Merci à toute l'équipe du projet Ensemble: Tobias, Emanuel, Laureline, Grégoire, Niklas, Ivan, Peter, Céline, Jef et Peter pour nos riches discussions, toutes les compétences que nous avons partagées, les divers conseils apportés, le terrain au Tagliamento... Merci Léa et Dimitri, pour vos travaux de Master en lien avec ma thèse, qui m'ont permis d'aller aussi loin dans le travail accompli. Merci Alessandro, Andrea, Martin et Christophe mes collègues de bureau pour votre bonne humeur, le partage de vos réflexions et tant d'autres choses... Merci à Fabio, Thomas, Lele, Paulo, Greg, Dylan... qui m'ont accompagné en musique lors de ces années, lors de nombreuses répétitions, jams et concerts! Merci à Lucien, Claire, Alice, Cécile, Lorienne, Axa, Julien, Eric pour les bons moments passés ensemble lors des pauses café, barbecues près du lac, apéros, week end au ski, sorties en peau de phoque. Merci spécial à Yoan pour les randonnées pédestres et à ski sans doute les plus improbables! Merci à tous les CHYNois pour ces moments partagés à l'Uni. Merci Luc, merci Céline de m'avoir accompagné à distance pendant cette thèse. Merci Michèle pour les souvenirs que j'ai de toi et tout ce que tu m'a transmis.

Acknowledments

The work presented in this thesis was financially supported by the Swiss National Science Foundation under the contract CRSI22 1222491.

Keywords

braided-river aquifer model, geological heterogeneity model, stochastic, pseudo-genetic, multiple-point statistics, LIDAR, topography model, groundwater flow and transport simulation

Mots-clés

modèle d'aquifère de rivière en tresses, modèle d'hétérogénéité géologique , stochastique, pseudo-génétique, statistiques multi-point, LIDAR, modèle de topographie, simulation d'écoulement et de transport des eaux souterraines

Abstract

In this thesis a new pseudo-genetic method to model the heterogeneity of sandy gravel braided-river aquifers is proposed. It is tested and compared with other modeling approaches on a case study of contaminant transport. Indeed, in Switzerland or in mountainous regions, braided-river aquifers represent an important water resource that need to be preserved. In order to manage this resource, a good understanding of groundwater flow and transport in braided-river aquifers is necessary. As the complex heterogeneity of such sedimentary deposits strongly influences the groundwater flow and transport, groundwater behavior predictions need to rely on a wide spectrum of geological model realizations.

To achieve realistic sedimentary deposits modeling of braided river aquifers, the proposed pseudo-genetic algorithm combines the use of analogue data with Multiple-Point Statistics and process-imitating methods. The integration of analogue data is a key feature to provide additional, complementary and necessary information in the modeling process. Assuredly, hydrogeologist are often subject to field data scarcity because of budget, time and field constraints. Multiple-Points Statistics recent algorithms, on one hand, allow to produce realistic stochastic realizations from training set with complex structures and at the same time allow to honor easily conditioning data. On the other hand, process-imitating methods allow to generate realistic patterns by mimicking physical processes.

The proposed pseudo-genetic algorithm consists of two main steps. The first step is to build main geological units by stacking successive topography realizations one above the other. So, it mimics the successive large flood events contributing to the formation of the sedimentary deposits. The successive topographies are Multiple-Point Statistics realizations from a training set composed of Digital Elevation Models of an analogue braided-river at different time steps. Each topography is generated conditionally to the previous one. The second step is to generate fine scale heterogeneity within the main geological units. This is performed for each geological unit by iterative deformations of the unit bottom surface, imitating so the process of scour filling. With three main parameters, the aggradation rate, the number of successive iterations and the intensity of the deformations, the algorithm allows to produce a wide range of realistic cross-stratified sedimentary deposits.

The method is tested in a contaminant transport example, using as reference Tritium tracer experiment concentration data from the MAcro Dispersion Experiment (MADE) site, Columbus, Mississippi, USA. In this test case, an assumption of data scarcity is made. Analogue data are integrated in the geological modeling process to determine the input parameters required – characteristic dimensions and conductivity statistical properties – for two variants of the proposed pseudo-genetic algorithm as well as for multi-gaussian simulation and object based methods. For each conceptual model, flow and transport simulations are run over 200 geological model realizations to cover a part of the uncertainty due to the input parameters. A comparison of the plume behavior prediction is performed between the different conceptual models.

The results show that geological structures strongly influence the plume behavior, therefore the choice or the restriction to specific conceptual models will impact the prediction uncertainty. Though little information are available for the modeler, it is possible to achieve reasonable predictions by using analogue data. Of course, with limited information, it is impossible to make an accurate prediction to match the reference, and none of each conceptual model produces better predictions but all are useful to cover the uncertainty range. The results also underline the need to consider a wide exploration of the input parameters for the various conceptual models in order to recover the uncertainty.

Résumé

Cette thèse propose une nouvelle méthode de modélisation de l'hétérogénéité des aquifères sablo-graveleux de rivières en tresse. la méthode est ensuite testée sur un cas de transport de contaminant et comparée d'autres méthodes de modélisation. De tels dépôts sédimentaires sont communs en Suisse ou dans les régions montagneuses et sont les réserves d'une part importante de l'eau consommée. Pour préserver et gérer au mieux la ressource, une bonne compréhension des phénomènes d'écoulement et de transport dans ces milieux est requise. De plus, ces écoulements étant fortement influencés par l'hétérogénéité et les structures complexes de ces dépôts, la prédiction de l'incertitude liée l'évolution d'un panache nécessite un éventail large et varié de modèles géologiques.

Afin d'obtenir des structures sédimentaires réalistes pour ces milieux générés par les rivières en tresses, l'algorithme pseudo-génétique proposé intègre l'utilisation de données analogue combinée avec des techniques basées sur les statistiques multi-point et l'imitation des processus. L'intégration de données analogues dans la modélisation est un élément clé pour palier au manque données de terrain auquel les hydrogéologues sont souvent confrontés. D'autre part, les récents algorithmes multi-point permettent de généré des réalisations stochastiques réalistes partir de données d'entrainement contenant des structures complexes. De plus, les méthodes basées sur l'imitation des processus permettent également de générer des motifs réalistes en tenant compte des phénomènes physiques observés.

L'algorithme pseudo-génétique est composé de deux étapes. La première consiste construire les unités géologiques importantes l'échelle du domaine modelisé. Cette étape est réalisée par l'empilement de topographies sucessives, ce qui revient imiter la sucession de grandes crues survenant dans les rivières en tresse et contribuant au dépôt des sédiments. Les topographies successives sont générées par simulation multi-point, partir de données d'entrainement composées de modèles numériques de terrain différentes dates d'une rivière en tresses analogue. La deuxième étape consiste générere l'hétérogénéité petite échelle, l'intérieur des unités géologiques principales. Pour cela, la surface de fond de chaque unité géologique est déformée itérativement, imitant ainsi le processus de remplissage des auges. Ainsi, avec trois paramètres essentiels, le taux d'aggradation, le nombre d'itération et l'intensité des déformation, il est possible de générer une grande varietés de structures sédimentaires entrecroisées réalistes.

Les modèles géologiques obtenus sont ensuite testés dans un exercice de prédiction d'évolution d'un plume contaminé, utilisant comme référence les données de concentration de l'expérience de traage au Tritium du site expérimental de Macro Dispersion (MADE), Columbus, Mississippi, USA. Pour rendre le contexte de modélisation réaliste l'hypothèse est faite que peu de données sont diponibles. Des données d'analogues sont intégérées aux processus de modélisation pour déterminer les paramètres d'entrée requis – comme les dimensions charactéristiques ou les propriétés statistiques des champs de conductivité – pour deux variantes de l'algorithme pseudo-génétique proposé et également pour une méthode basée sur les simulations de champs multi-gaussien et pour une méthode de modélisation par objets. Pour chaque concept, 200 modèles géologiques différents sont générés pour couvrir en partie l'incertitude liée aux paramètres des modèles. Les prédictions du comportement des panaches sont ensuite comparées entre les différents modèles.

Les résultats montrent que les différentes structures géologiques influencent fortement le comportement des panaches. Conséquemment, le choix des modèles conceptuels a donc une incidence sur l'incertitude des prédictions. Même si peu de données sont disponibles, il est tout fait possible d'obtenir des prédictions raisonables en utilisant des données d'analogues. Bien évidemment, avec des données limitées, il n'est pas étonnant de ne pas avoir une seule prédiction reproduisant la référence. Bien qu'aucune méthode en particulier ne pro-

duise de meilleures prédictions que les autres, toutes sont utiles pour couvrir la portée de l'incertitude. Cela souligne également la nécessité d'explorer l'espace des paramètres pour différents modèles conceptuels, afin de ne pas sous-estimer l'incertitude des prédictions.

Contents

Chapter 1

Introduction

1.1 Motivations

In alpine regions, braided river aquifers constitute a major resource tapped to produce drinking water and to support agricultural or industrial activities (FOEN, 2009). In a context of climate change (Middelkoop et al., 2001; Moeck, 2014), of intensive use of the resource and under urbanization constraints, the management of the resource becomes a challenge especially to preserve water quality, to protect groundwater dependent ecosystems and to contain flooding risks. This awareness is illustrated by river restoration projects all around the worlds (Sear, 1994; Caruso, 2006; Tockner et al., 2009; Glenz, 2013). But investigating surface groundwater interactions or quantifying the uncertainty of contaminant migration in these complex geological media (Renard, 2007; Brunner et al., 2009) require reliable aquifer heterogeneity models at the inundation plain scale.

Solute transport problems are essentially controlled by the geological heterogeneity and the resulting petrophysical properties (De Marsily et al., 2005; Eaton, 2006). Various modeling tools exist and propose different representations of geological heterogeneity. But none of them is known to be right or more appropriate to represent braided river deposits. Indeed, as far as we know, no full 3D detailed characterization of the internal heterogeneity of a braided aquifer exists at the inundation plain scale. One of the most detailed 3D characterizations of a braided river deposit was achieved by sedimentological observations on a series of seven cross-sections at the Herten site (Bayer et al., 2011), but on a rather small area ($16\,m \times 7\,m$ vertical sections along a $10\,m$ profile). Consequently, when considering the quantification and propagation of uncertainty, several conceptual models shall be considered.

Thanks to their mathematical properties and computing ease, multi-gaussian based geostatistical models (Felletti et al., 2006; Salamon et al., 2007; Glenz, 2013) have flourished. They have been tested on groundwater flow and transport problems. But by nature, multi-gaussian simulations favor the connectivity of the mean values and do not allow for reproduction of realistic geological patterns (Guardiano and Srivastava, 1993), therefore even if these models are well calibrated, accurate plume evolution prediction might be hard to achieve (Zinn and Harvey, 2003).

New object based algorithms have been developed recently to model braided river aquifers (Ramanathan et al., 2010; Huber et al., 2015). These approaches present a simplified representation of the expected geological structures. They allow to produce geological realizations presenting simplified structures compatible with field observations such as outcrops or Ground Penetrating Radar profiles. The inputs required are relatively simple and limited to length scale characteristics as well as hydrogeological property distributions for each kind of objects. An advantage of this method is that these algorithms are fast. However conditioning to field data can be difficult.

Introduced by Guardiano and Srivastava (1993), Multiple-Point Statistics (MPS) algorithms allow realistic simulations of complex patterns from a training data set. First practical implementations such as SNESIM (Strebelle, 2002) are apt for the simulation of categorical variables. More recent algorithms such as the direct sampling (DS) proposed by Mariethoz et al. (2010), bring more flexibility and can deal with joint simulations of multiple categorical and continuous variables. Using MPS (Straubhaar et al., 2011), Comunian et al. (2011) proposed an innovative way to model a three dimensional braided river deposit from the detailed characterization work at the Herten site (Bayer et al., 2011). This method proved the ability of MPS to reproduce complex fine scale geological structures and to honor conditioning data, but on a rather small area ($16\,m \times 10\,m \times 7\,m$). However, to our knowledge, no three dimensional training data set covers a large domain. More generally, the lack of training data set, especially in a three dimensional space is one of the difficulties in using MPS (Comunian et al., 2012).

Process-imitating methods also called pseudo-genetic algorithm might offer more realistic geological patterns in accordance with field and outcrop observations but are CPU expensive and not straightforward to condition to field data. So far, few such methods have been developed to model braided river aquifer heterogeneous deposits (Webb, 1994; Teles et al., 2004). The algorithm of Webb (1995) produces first successive topographies of the braided river by random walk simulations. Then, the topographies are stacked one over the other, which creates volumes that are assigned to a facies according to the local flow velocity given by an approximation of the Froude number. Note that aside from channels width and depth, all surfaces between channels are considered as flat. Furthermore, the facies assignment does not account for the geometry of the volumes.

An aspect that is important to account for when modeling braided river aquifer deposits at a specific site is that information quantity might be severely restricted. Indeed, the field data budget is always limited, and technical difficulties might prevent accurate and exhaustive data acquisition. In an active system, one might not be sure to retrieve piezometers or other equipments after a big flood event. In inactive systems, one is limited by vegetation or restrictions introduced to preserve the local ecosystems. In addition, quarries or outcrops might not be accessible close to the site of interest. Then the availability and the use of analogue site data may prove useful and necessary.

1.2 Objectives

This thesis pursues two main objectives.

The first aim is to propose a new modeling method, able to integrate analog data, to produce realistic geological models of gravel braided river deposits while combining the advantages of MPS and process-imitating methods. MPS tools allow to generate realistic structures, facilitate data conditioning and offer the possibility to integrate secondary information. By mimicking processes, pseudo-genetic methods produce realistic realizations that have a physical meaning in addition to their statistical meaning. The proposed pseudo-genetic algorithm revisits the method of Webb (1994) and uses MPS to generate realistic successive topographies from large scale and high resolution digital elevation models acquired by LIDAR (Lane et al., 2003).

The second goal is to assess in a contaminant transport example if analogue data integration allows to make 1) reasonable plume evolution predictions and 2) relevant uncertainty analysis. The proposed pseudo-genetic algorithm and methods based on other concepts of geological heterogeneity for braided river aquifer are then tested on a flow and transport problem. For this purpose, the very dense data set of the MADE II tracer experiment (Columbus, Mississippi, USA) has been used. It offers the possibility to compare predictions with actual solute concentrations on a large number of sample points.

1.3 Methodology

The thesis is organized in four main chapters following a common thread. Nevertheless, each chapter may be read independently.

Chapter 2 provides a detailed sensitivity analysis of the DS algorithm. The quality of two-dimensional and three-dimensional realizations is assessed as a function of the main parameters of the DS algorithm. It also provides an illustration of post-processing and conditioning options as well as multivariate simulations. This is an important preliminary step since this technique is used intensively in chapter 3 and 4 as an internal mechanism for the pseudo-genetic simulation of braided river aquifers.
This chapter was published in Computers & Geosciences (2013) as *"A practical guide to*

performing multiple-point statistical simulations with the Direct Sampling algorithm" by E. Meerschman, G. Pirot, G. Mariethoz, J. Straubhaar, M. Van Meirvenne and P. Renard.

Chapter 3 presents a method to simulate successive braided river topographies with MPS. It uses the DS algorithm on a large scale Digital Elevation Model dataset, derived from high resolution LIDAR data acquired at different time steps over the Waimakariri River, New Zealand (Lane et al., 2003). An assessment of the resulting simulations is performed through a connectivity analysis.
This chapter was published in Geomorphology (2014) as *"Simulation of braided river elevation model time series with multiple-point statistics"* by G. Pirot, J. Straubhaar and P. Renard.

Chapter 4 proposes a new process-imitating method to produce realistic realizations of geological heterogeneity for braided river aquifer deposits. The algorithm is based on the vertical stacking of successive braided river topographies, obtained by applying the method developed in chapter 3. It produces the main geological structures. Their inner structure heterogeneity is then assigned according to a scheme mimicking the trough filling process occurring during braided river floods.
This chapter will be submitted to Water Resources Research as *"A pseudo genetic model of coarse braided-river deposit"* by G. Pirot, J. Straubhaar and P. Renard.

Chapter 5 is a plume prediction exercise to test the resulting models proposed in chapter 4 with other conceptual models such as multi-gaussian simulations and object based models. It is performed by assuming a situation where limited field information is available and the use of analogue data is required. The MADE 2 Tritium tracer experiment is used as a reference in this test case.
This chapter will be submitted to a Special Issue of Journal of Hydrology as *"Influence of conceptual model choice on contaminant transport uncertainty forecasting in braided river aquifers"* by G. Pirot, P. Renard, E. Huber, J. Straubhaar and P. Huggenberger.

Chapter 6 summarizes the main results of the thesis and presents further perspectives.

Appendix A describes a method to explore model space parameters at reasonable cost, when approximate physical models are available to surrogate costly full physics models. This appendix was published in Advances in Water Resources as *"Distance-based kriging relying on proxy simulations for inverse conditioning"* by D. Ginsbourger, B. Rosspopoff, G. Pirot, N. Durrande and P. Renard.

Appendix B presents a methodology to condition MPS simulations using geophysical tomography data. This appendix was published in Mathematical Geosciences as *"Conditioning of Multiple-Point Statistics Facies Simulations to Tomographic Images"* by T. Lochbühler, G. Pirot, J. Straubhaar and N. Linde.

Chapter 2

A sensitivity analysis of the Direct Sampling algorithm

Remark

This chapter was written jointly with Eef Meershman, the first author of the corresponding paper. The contribution was equally shared between the first two authors. The first author, at the origin of the paper was a bit more involved in the writing while the second author was more involved in the generation of the simulations and in the design and computing of quality indicators. Both were strongly implied in the interpretation of their results.

Abstract

The Direct Sampling (DS) algorithm is a recently developed multiple-point statistical simulation technique. It directly scans the training image (TI) for a given data event instead of storing the training probability values in a catalog prior to simulation. By using distances between the given data events and the TI patterns, DS allows to simulate categorical, continuous and multivariate problems. Benefiting from the wide spectrum of potential applications of DS, requires understanding of the user-defined input parameters. Therefore, we list the most important parameters and assess their impact on the generated simulations. Real case TIs are used, including an image of ice-wedge polygons, a marble slice and snow crystals, all three as continuous and categorical images. We also use a 3D categorical TI representing a block of concrete to demonstrate the capacity of DS to generate 3D simulations. First, a quantitative sensitivity analysis is conducted on the three parameters balancing simulation quality and CPU time: the acceptance threshold t, the fraction of TI to scan f and the number of neighbors n. Next to a visual inspection of the generated simulations, the performance is analyzed in terms of speed of calculation and quality of pattern reproduction. Whereas decreasing the CPU time by influencing t and n is at the expense of simulation quality, reducing the scanned fraction of the TI allows substantial computational gains without degrading the quality as long as the TI contains enough reproducible patterns. We also illustrate the quality improvement resulting from post-processing and the potential of DS to simulate bi-variate problems and to honor conditioning data. We report a comprehensive guide to performing multiple-point statistical simulations with the DS algorithm and provide recommendations on how to set the input parameters appropriately.

2.1 Introduction

Multiple-point statistics (MPS) covers an ensemble of sequential simulation algorithms using a training image (TI) as input data for the spatial structure of a process instead of a two-point variogram (Guardiano and Srivastava, 1993; Strebelle and Journel, 2000). A TI is a conceptual image of the expected spatial structure and is often built based on prior information. Using a TI allows extracting multiple-point statistics and hence describing more complex patterns; this is especially important when spatial connectivity plays a key role in the model application (Bianchi et al., 2011; Gómez-Hernández and Wen, 1998; Renard et al., 2013; Zinn and Harvey, 2003).

As is characteristic for sequential simulations, the unknown locations $\mathbf{x}$ of the simulation grid are visited according to a predefined (random or regular) path. For each $\mathbf{x}$ the simulated value is drawn from a cumulative distribution function F conditioned to a local data event $\mathbf{d}_\mathrm{n}$:

$$F(z, \mathbf{x}, \mathbf{d}_\mathrm{n}) = \mathrm{Prob}\left\{ \mathrm{Z(x)} \leq \mathrm{z} | \, \mathbf{d}_\mathrm{n} \right\}. \tag{2.1}$$

This data event comprises the values of the known neighboring grid nodes $\mathbf{x}_\mathrm{i}$, i.e. the conditioning data and the already simulated grid nodes, and their relative positions. F is built based on the central nodes of TI patterns equal or similar to $\mathbf{d}_\mathrm{n}$.

The Direct Sampling (DS) algorithm is a recent MPS algorithm (Mariethoz et al., 2010)[1]. The particularities of DS consist in skipping the explicit modeling of F by directly sampling the TI during simulation, and in using dissimilarity distances between $\mathbf{d}_{\mathrm{n}}$ and the TI patterns. As soon as a TI pattern is found that matches $\mathbf{d}_{\mathrm{n}}$ exactly or as soon as the distance between the TI pattern and $\mathbf{d}_{\mathrm{n}}$ is lower than a given threshold, the value at the central node of the TI pattern is directly pasted to $\mathbf{x}$. Since the TI is scanned randomly, this strategy is equal to drawing a random value from F, but increases simulation speed. Other MPS algorithms, like the widely used *snesim* (Strebelle, 2002) and *IMPALA* (Straubhaar et al., 2011) algorithm, scan the TI beforehand for all possible $\mathbf{d}_{\mathrm{n}}$'s and store the TI probabilities in a catalogue. Therefore, they are restricted to the simulation of categorical variables and need to use a predefined template for $\mathbf{d}_{\mathrm{n}}$. Due to its unique strategy, DS allows to simulate both categorical and continuous variables, and to handle multivariate cases, only by selecting the appropriate distance measures.

Since DS is a promising simulation technique for a wide range of applications, it is important to understand precisely its capacities and its sensitivity to the user-defined input parameters. DS is implemented in the ANSI C language and all input and output files are in an ASCII SGeMS compatible format (Remy et al., 2009). Detailed algorithm steps and further implementation details of DS can be found in Mariethoz et al. (2010) and the user manual of DS.

Using DS requires the user to define some parameters: among them, the acceptance threshold t, the maximum fraction of TI to scan f and the maximum number of points in the neighborhood n are the most important since they are balancing simulation quality and CPU time (section 2.2.1). For these three parameters, a detailed sensitivity analysis was reported by generating non-conditional simulations for the entire 3D parameter space. Next to a visual inspection of the resulting simulations, we quantified the similarity between the simulations and the TI by means of simulation quality indicators (CASE 1). The same quality indicators were calculated for a 3D example (CASE 2). We also illustrated the potential of the post-processing option (CASE 3), the multivariate simulation option (CASE 4) and the data conditioning option (CASE 5) and discussed the corresponding user-defined input parameters. Tab. 2.1 summarizes the values of the parameter that we kept fixed and the range of values of the parameters that we varied.

Since Mariethoz et al. (2010) already showed good performance of DS with as much as 54 processors, the parallelization option is not discussed here. For more information about the option to use transform-invariant distances we refer to Mariethoz and Kelly (2011).

Many previous studies have used only one TI with sinuous channels. In contrast, we included a greater variety of patterns by performing sensitivity analyses on seven TIs: an image of ice-wedge polygons (Plug and Werner, 2002), a microscopic view of a thin marble slice, an image of snow crystals, all three as categorical and continuous images, and a categorical 3D image of concrete (Fig. 2.1). The continuous 2D TIs are grayscale photographs with pixel values between 0 and 255; the categorical 2D TIs were derived from these by classifying them into three categories. The 3D TI was generated by sequentially simulating 2D slices constrained by conditioning data computed at the previous simulation steps (Comunian et al., 2012). The figures shown in this paper are the results for the categorical ice-wedge TI, the continuous marble TI and the 3D concrete TI. They are presented with the same color scale as the TIs in Fig. 2.1. The results for the other TIs can be found as supplementary electronic material available online.

[1]It is the object of an international patent application (PCT/EP2008/009819). The code is available on demand for academic and research purposes. Requests should be sent to Philippe Renard, Grégoire Mariethoz or Julien Straubhaar.

Table 2.1: Fixed parameters with their default values chosen for this study (sorted according to their appearance in the parameter file) and parameters that were varied with their default values and range over which they were varied (sorted according to the case number in which they were studied).

Fixed parameters	
Name	**Default**
Simulation method	MPS
Number of realizations	10
Max search distance	125 125 0 ($\frac{1}{2}$ size simulation grid)
Anisotropy ratios in the search window (x, y, z)	1 1 1
Transformations	0 (no transformations)
Path type	0 (random path)
Type of variable	0 for categorical, 1 for continuous
Exponent of the distance function in the template	0
Syn-processing parameters (4)	0 0 0 0 (no syn-processing)
Initial seed	1350
Parameters reduction	1 (no parameters reduction)
Parallelization	1 (serial code, no parallelization)

Varied parameters			
Name	**Default**	**Range**	**Case**
Threshold position t	0.05	0.01-0.02-0.04-0.06-0.08-0.1-0.12 0.14-0.16-0.18-0.2-0.25-0.5-0.75-0.99	1
Max fraction of TI to scan f	0.5	0.05-0.1-0.15-0.2-0.3-0.4-0.5-0.6-0.75-1	1
Max number of points in neighborhood n	50	1-5-10-15-20-30-50-80	1
Number of post-processing steps (p)	0	0-1-2	2
Post-processing factor (pf)	0	0-1-3	2
Number of variables to simulate jointly	1	1-2	3
Relative weight of each variable	1	0.1 0.9-0.25 0.75-0.5 0.5-0.75 0.25-0.9	3
Weight of conditioning data (δ)	1	0-1-5	4
Data conditioning	No	No-yes	4

2.2 CASE 1: parameters balancing simulation quality and CPU time: t, f and n

2.2.1 Parameters t, f and n

An acceptance threshold t needs to be defined because a TI pattern matching $\mathbf{d}_n$ exactly is often not found, especially for continuous variables. When the distance between the TI pattern and $\mathbf{d}_n$ is smaller than t, the central node of the TI pattern is pasted at location $\mathbf{x}$. The default distances used in this paper are based on the fraction of non-matching nodes for categorical simulations and the mean squared errors for continuous simulations. The 'exponent to the distance function in the template' was set to 0, meaning that the distances are calculated without weighting the nodes in the TI pattern and the central node according to their proximity to the central node. All distances are normalized ensuring their minimum to be zero (exact match) and their maximum to be 1 (no match) (Mariethoz et al., 2010).

The maximum fraction of TI to scan f limits the number of TI patterns that are scanned for their similarity with $\mathbf{d}_n$: f ranges from 0 (no scan) to 1 (scan full TI if necessary). If the maximum fraction of the TI f is scanned and still no TI pattern with a distance smaller than t is found, the central node of the TI pattern with the lowest distance is pasted at location $\mathbf{x}$.

The neighborhood $\mathbf{d}_n$ is defined as the n grid nodes that are closest to $\mathbf{x}$ within the defined search area. This search area can be defined by setting the parameters 'maximum search distance', i.e. the radius in the x-, y- and z- direction of a rectangular search area. Generally, it is advised to use a large search area by setting the radii to half the size of the simulation grid, corresponding to the maximum neighborhood size, except when considering non-stationary variables (Mariethoz et al., 2010), or if patching occurs (discussed below).

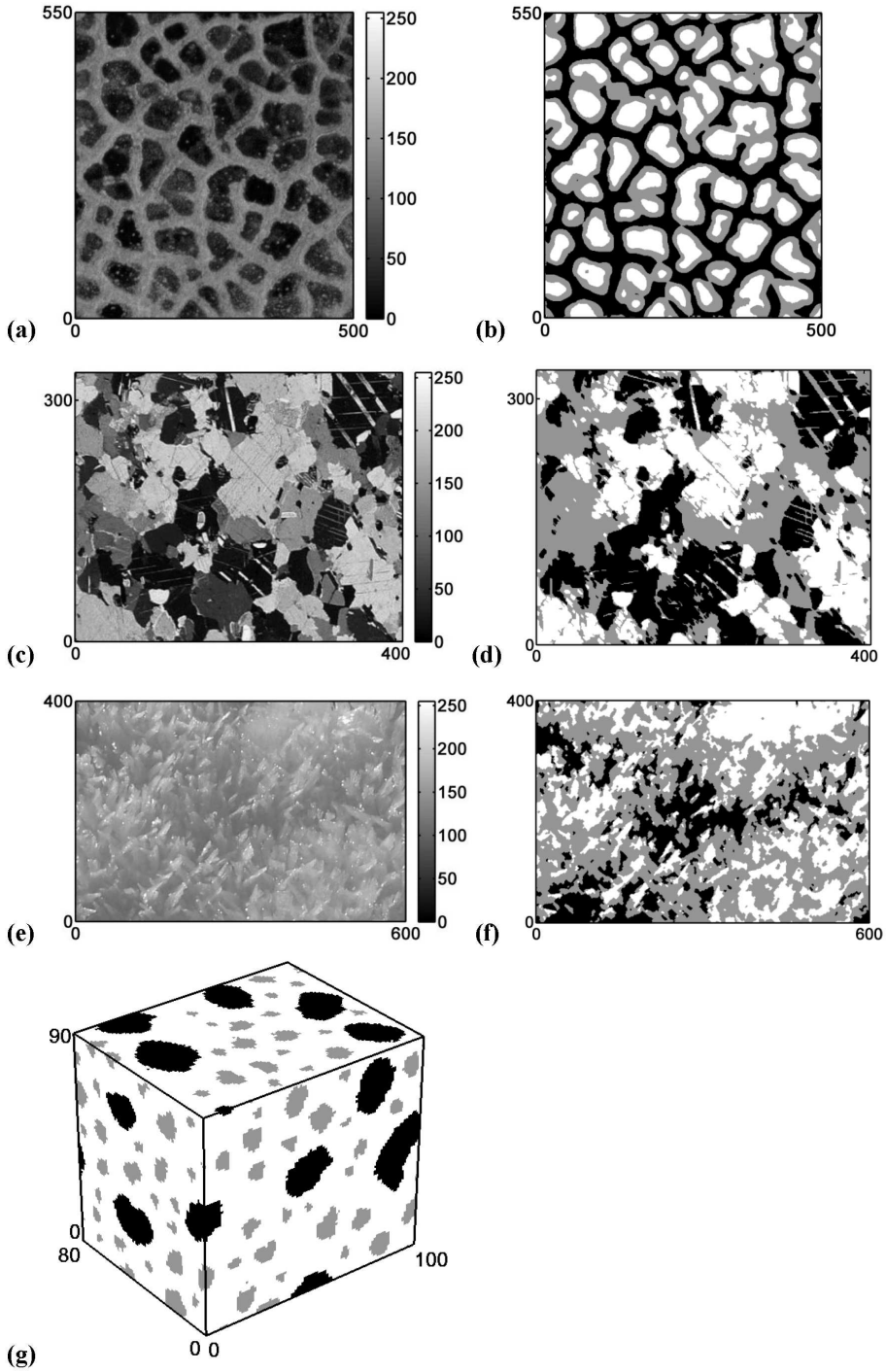

Figure 2.1: The six training images (TIs) that were used for the sensitivity analyses: (a) continuous (photograph: Plug and Werner (2002)) and (b) categorical ($k = 3$) TI of ice-wedge polygons; (c) continuous and (d) categorical ($k = 3$) TI of a thin marble slice; (d) continuous and (e) categorical ($k = 3$) TI of snow crystals. The x- and y-axes represent the number of pixels. The results of the sensitivity analyses for TIs (a) to (b) are illustrated in this paper; the results for TIs (c) to (f) can be found in the App. C, Fig. C.1 to C.4.

The definition of n results in $\mathbf{d}_\mathrm{n}$'s covering a large part of the search area when the first unknown grid nodes are simulated, and a progressive decrease of the size of the area covered

by $\mathbf{d_n}$ when the number of already simulated nodes increases. Consequently, DS ensures that structures of all sizes are present in the simulation, which is also the purpose of the multi-grid approach in *snesim* (Strebelle, 2002).

Fig. 2.2 gives an overview of the DS workflow and its three main input parameters.

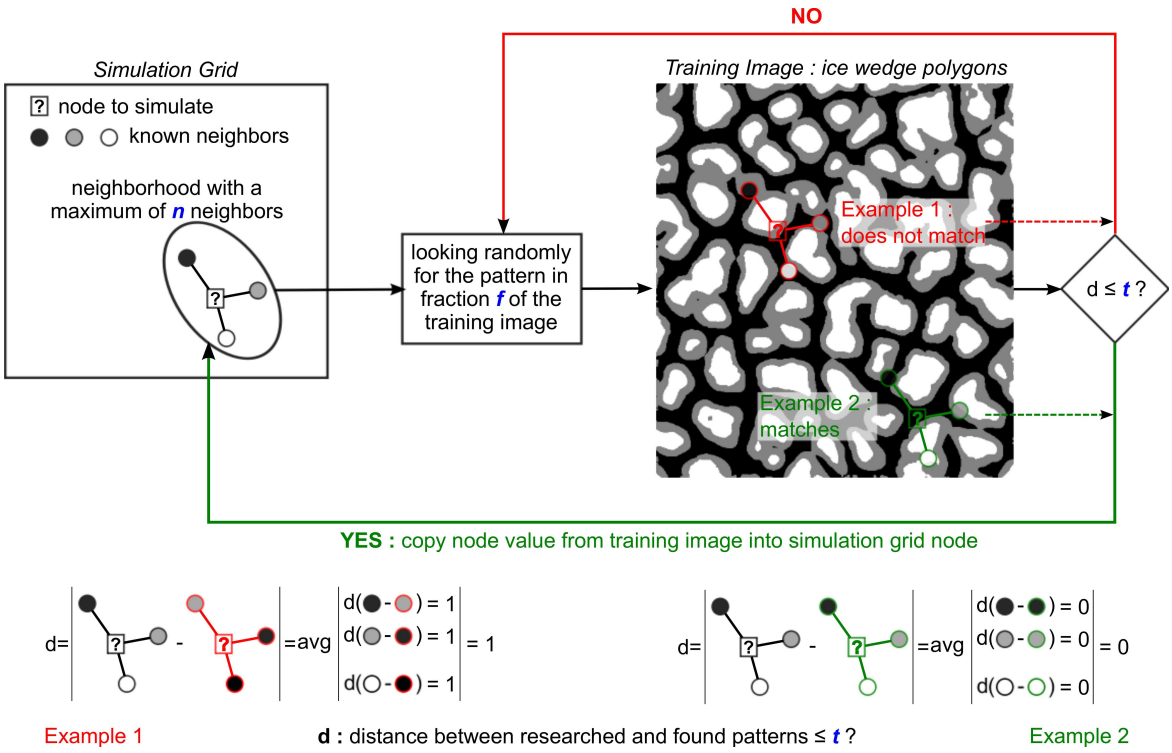

Figure 2.2: Workflow of DS explaining the three main input parameters: the acceptance threshold t, the maximum fraction of TI to scan f and the maximum of neighbors n.

It is clear that the larger n and the closer t to 0 and f to 1, the better the simulation quality will be. However, these settings will be very expensive in terms of CPU time. For all six TIs, we simulated 10 unconditional realizations for each parameter combination of 15 t values, 10 f values and 8 n values (Tab. 2.1), resulting in 12,000 realizations for each TI.

2.2.2 CPU time

Fig. 2.3 shows the CPU time needed to simulate one unconditional simulation for the categorical and the continuous case. First the influence of t and n is shown for $f = 0.5$ after which the influence of f is shown for different combinations of t and n.

Besides the fact that generating simulations based on the continuous TI generally takes longer, the results for the categorical and the continuous case show a similar behavior. Simulations with small t and large n require a long simulation time and decreasing f strongly reduces CPU time. Modifying one of the parameters t, f or n increases or decreases the CPU exponentially. The combined effect of relaxing all three parameters only slightly, can reduce CPU time significantly. For instance, generating one simulation for the categorical case with default parameters ($t = 0.05$, $f = 0.5$, $n = 50$) took 163 s. Relaxing t to 0.1 only took 44 s, relaxing all three parameters to $t = 0.1$, $f = 0.3$ and $n = 30$ only took 13 s.

This behavior is related to the scanning algorithm. When t is close to 0, f close to 1 and n high, the algorithm scans the entire TI for a very good match with complex data events (large neighborhoods). This takes a lot of time. In the opposite case, the algorithm finds very quickly a TI pattern that matches the criteria and the algorithm is fast. What is

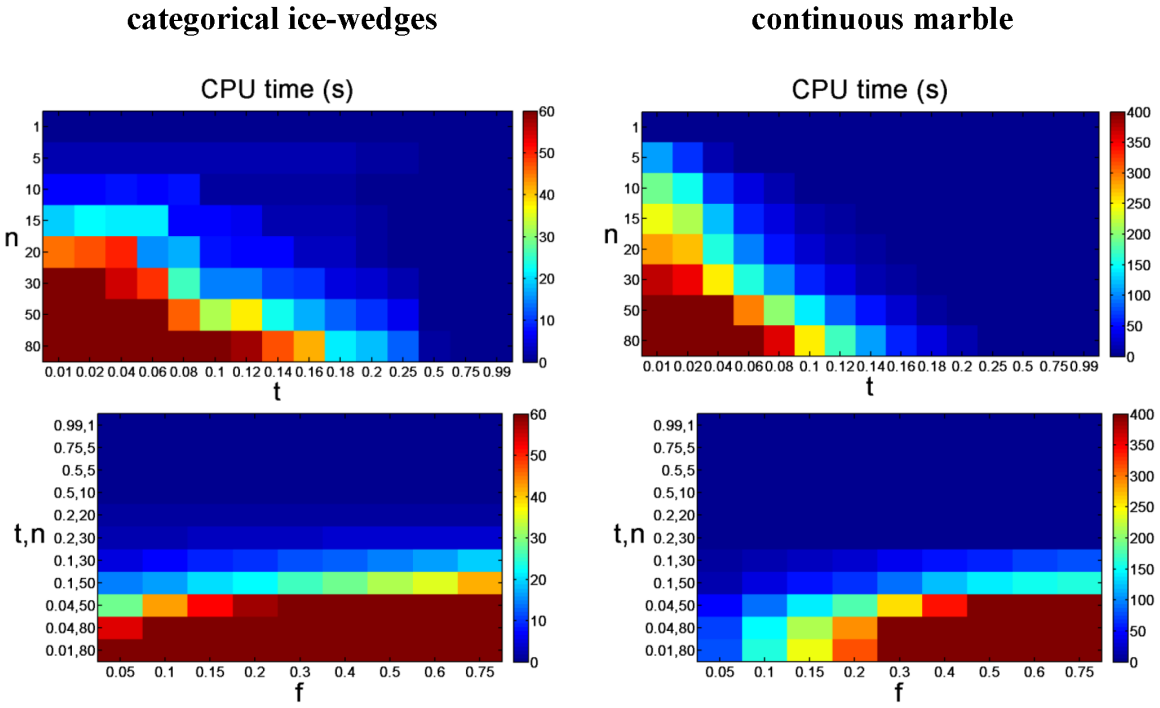

Figure 2.3: Influence of t and n (for $f = 0.5$) (top) and f (bottom) on the CPU time required to run one unconditional simulation.

striking in Fig. 2.3is that the evolution between these two cases is rather abrupt for some parameter values. When the parameter values are behind such an abrupt boundary, DS is very fast whatever the parameter values, below this boundary CPU time increases.

2.2.3 Visual quality inspection

Fig. 2.4 shows the first out of 10 simulations for some combinations of t, f and n using the categorical ice-wedge TI and Fig. 2.5 using the continuous marble TI. The results for the other TIs can be found as supplementary material. Similar as for Fig. 2.3, first a sensitivity analysis on t and n was performed (for $f = 0.5$), after which the effect of f was illustrated for some combinations of t and n. We selected simulations with different quality levels in order to illustrate the evolution of the simulation quality. As the quality steps depend on the TI, simulations with different t and n values were illustrated for each case.

For the categorical case, running DS with $t > 0.5$ or $n \leq 5$ resulted in noisy images. This is not surprising since then the sampling is not selective enough: many TI pattern can be accepted even if it is far away from $\mathbf{d}_n$. This corresponds to situations in which the algorithm is very fast. For $t \leq 0.5$ and $n > 5$, the ice-wedge pattern was reasonably well reconstructed. For $t \leq 0.2$ and $n \geq 30$, the simulation quality was very good. Not only the pattern reconstruction, but also the appearance of noise and the fuzziness of the edges between different categories were influenced by t and n (CASE 3). For the categorical marble TI (Fig. 2.5d) similar thresholds were found (App. C, Fig. C.2). For the categorical snow crystals TI (Fig. 2.1f) the simulation quality was good for $t \leq 0.1$ and $n \geq 50$ (App. C, Fig. C.3). In contrast to the effect of t and n, f had a much smaller effect on the simulation quality. Scanning a smaller part of the TI hardly resulted in a quality decrease (Fig. 2.4b). The same conclusion can be drawn from the simulations using the other categorical TIs .

Fig. 2.5 shows that generating continuous simulations generally requires stricter parameters (lower t, higher n and f). Running DS with $t \geq 0.2$ resulted in noisy images. The

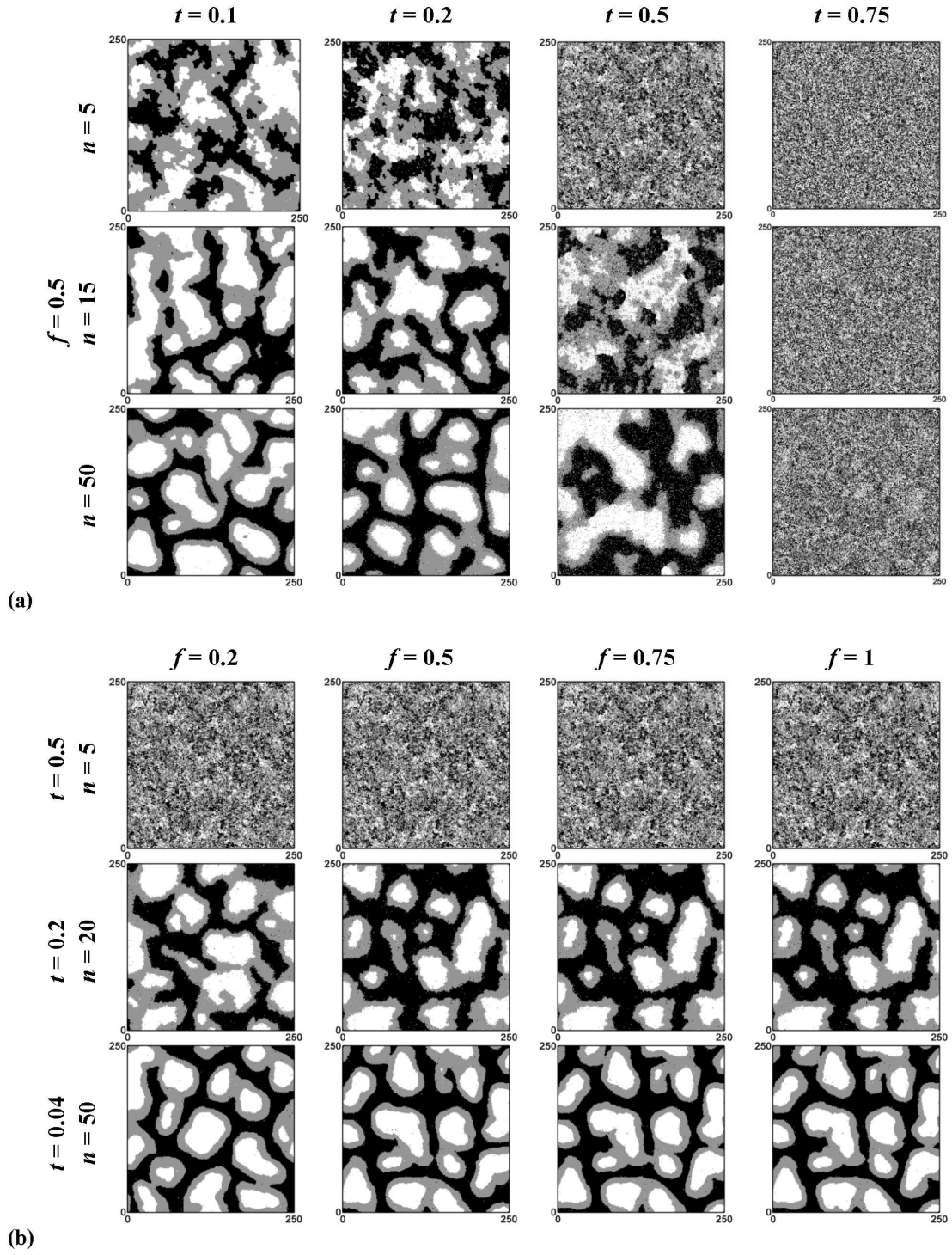

Figure 2.4: First out of 10 unconditional simulations illustrating the effect of parameters t and n with $f = 0.5$ and (b) first out of 10 unconditional simulations illustrating the effect of f for constant t and n based on the categorical ice-wedge TI (Fig. 2.1b).

simulation quality was good for $t \leq 0.1$ and $n \geq 30$. Simulations using the continuous ice-wedge TI (Fig. 2.1a) showed important changes in visual quality for the same values of t and n (App. C, Fig. C.1). The quality of the simulations using the continuous snow crystal

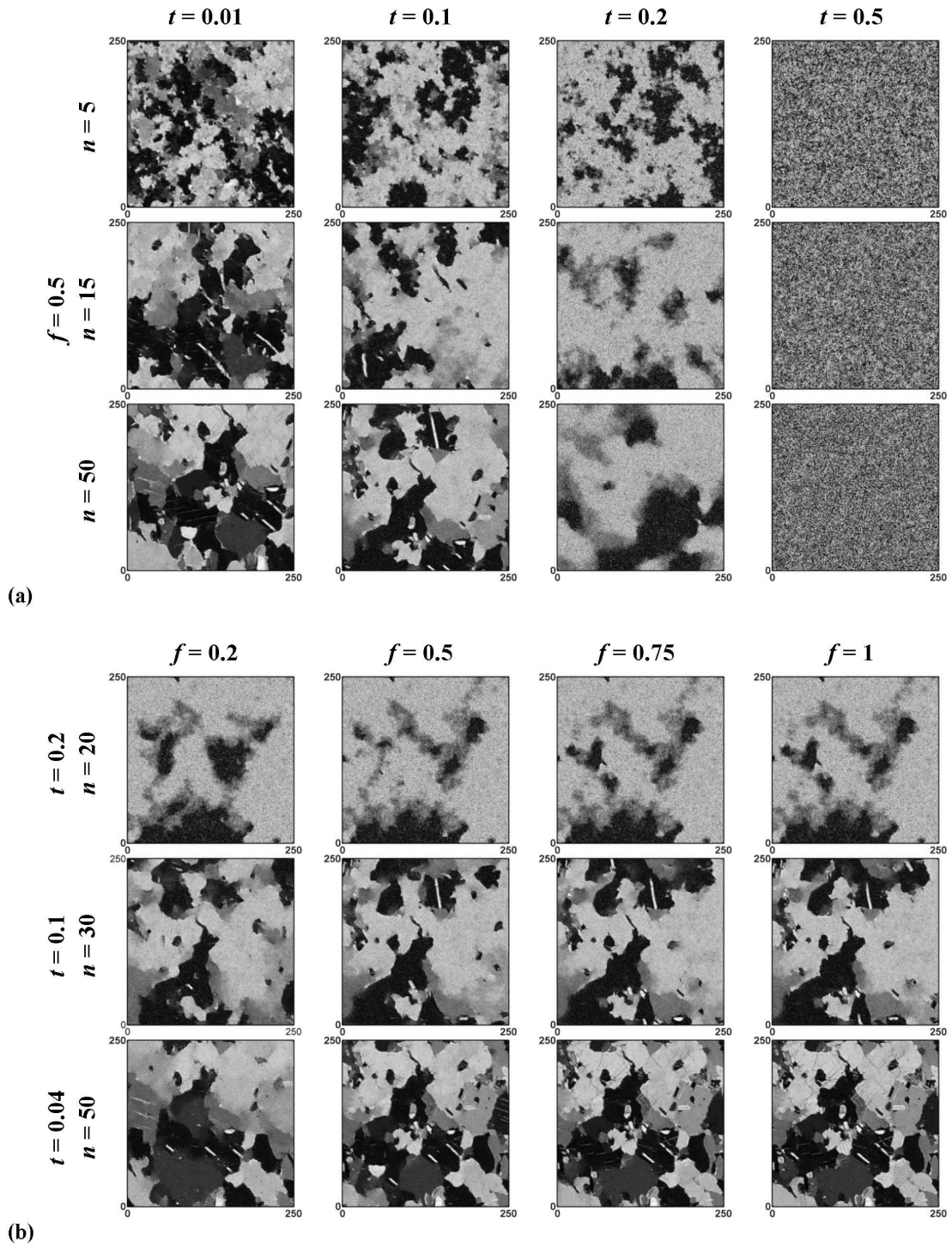

Figure 2.5: First out of 10 unconditional simulations illustrating the effect of parameters t and n with $f = 0.5$ and (b) first out of 10 unconditional simulations illustrating the effect of f for constant t and n based on the continuous marble TI (Fig. 2.1c).

TI (Fig. 2.1e) was generally less good: only for $t \leq 0.1$ and $n \geq 50$ the simulation quality was moderate (Supplementary material - Fig. d). For the continuous cases, it was observed that variations in f did not affect much on the simulation quality.

Especially for the continuous cases, it can be seen that some simulations are almost exact copies of parts of the TI. This phenomenon is called 'patching'. It is caused by copying each time the central node of the same best matching pattern. The issue of patching will be discussed further in sec. 2.2.5.

2.2.4 Simulation quality indicators

For each unconditional simulation we calculated several quality indicators by comparing the histogram, variogram and connectivity function of the TI and the simulations. The connectivity function $\tau(\mathbf{h})$ for a category s is defined as the probability that two points are connected by a continuous path of adjacent cells all belonging to s, conditioned to the fact that the two points belong to s (Boisvert et al., 2007; Emery and Ortiz, 2011; Renard et al., 2011, 2013):

$$\tau(\mathbf{h}) = \text{Prob}\left\{\mathbf{x} \leftrightarrow \mathbf{x}+\mathbf{h} \middle| s(\mathbf{x}) = s, s(\mathbf{x}+\mathbf{h}) = s\right\}. \tag{2.2}$$

Fig. 2.6 compares the histogram, variogram and connectivity function of the categorical ice-wedge TI (Fig. 2.1b) with these of a good simulation ($t = 0.01$, $f = 0.5$, $n = 80$) and a bad simulation ($t = 0.5$, $f = 0.5$, $n = 15$ for categorical and $t = 0.2$, $f = 0.5$, $n = 5$ for continuous). Both the indicator variogram values $\gamma^k(\mathbf{h})$ and the connectivity function $\tau^k(\mathbf{h})$ for each category k were calculated for 20 lag classes $\mathbf{h}$ with a lag width of 5. The histograms (proportions of the three categories) are represented for both simulations. The indicator variograms and the connectivity functions are only reproduced for the good simulation, except for the intermediate material (grey), where the bad simulation could partially reproduce the TI statistics.

Fig. 2.7 illustrates the same for the continuous case. Here the standard variogram $\gamma(\mathbf{h})$ was calculated instead of the indicator variogram. To calculate the connectivity functions, the TI and the simulations were first divided into three categories based on two thresholds representing connectivity jumps in the TI (Renard et al., 2013). Similarly to Fig. 2.6, the histograms (represented as the cdf) are represented for both the good and the bad simulation, whereas the variogram and the connectivity functions are only reproduced by the good simulation.

To quantify the dissimilarity between the simulations' statistics and those of the TI, we calculated three error indices for each simulation: a histogram error ε_{hist}, variogram error ε_{var} and connectivity error ε_{conn}. For the categorical simulations, ε^k_{hist} was defined as the absolute value of the difference between the proportion of k in the simulation grid f^k_{sim} and in the TI f^k_{TI}. For the continuous simulations, ε_{hist} was calculated as the Kullback–Leibler divergence D_{KL} (Kullback and Leibler, 1951)

$$\varepsilon_{hist} = D_{KL}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \tag{2.3}$$

with P defined as the probability distribution in the TI and Q the probability distribution in the simulations.

The variogram error ε_{var} for the continuous simulations was based on the weighted average difference between the variogram values of the simulations $\gamma_{sim}(\mathbf{h})$ and the TI $\gamma_{TI}(\mathbf{h})$ for 20 lag classes $\mathbf{h}_d$, and was calculated as

$$\varepsilon_{\text{var}} = \frac{\sum_{d=1}^{20} \frac{1}{\mathbf{h}_d} \frac{|\gamma_{sim}(\mathbf{h}_d) - \gamma_{TI}(\mathbf{h}_d)|}{\text{var}_{sim}}}{\sum_{d=1}^{20} \frac{1}{\mathbf{h}_d}} \tag{2.4}$$

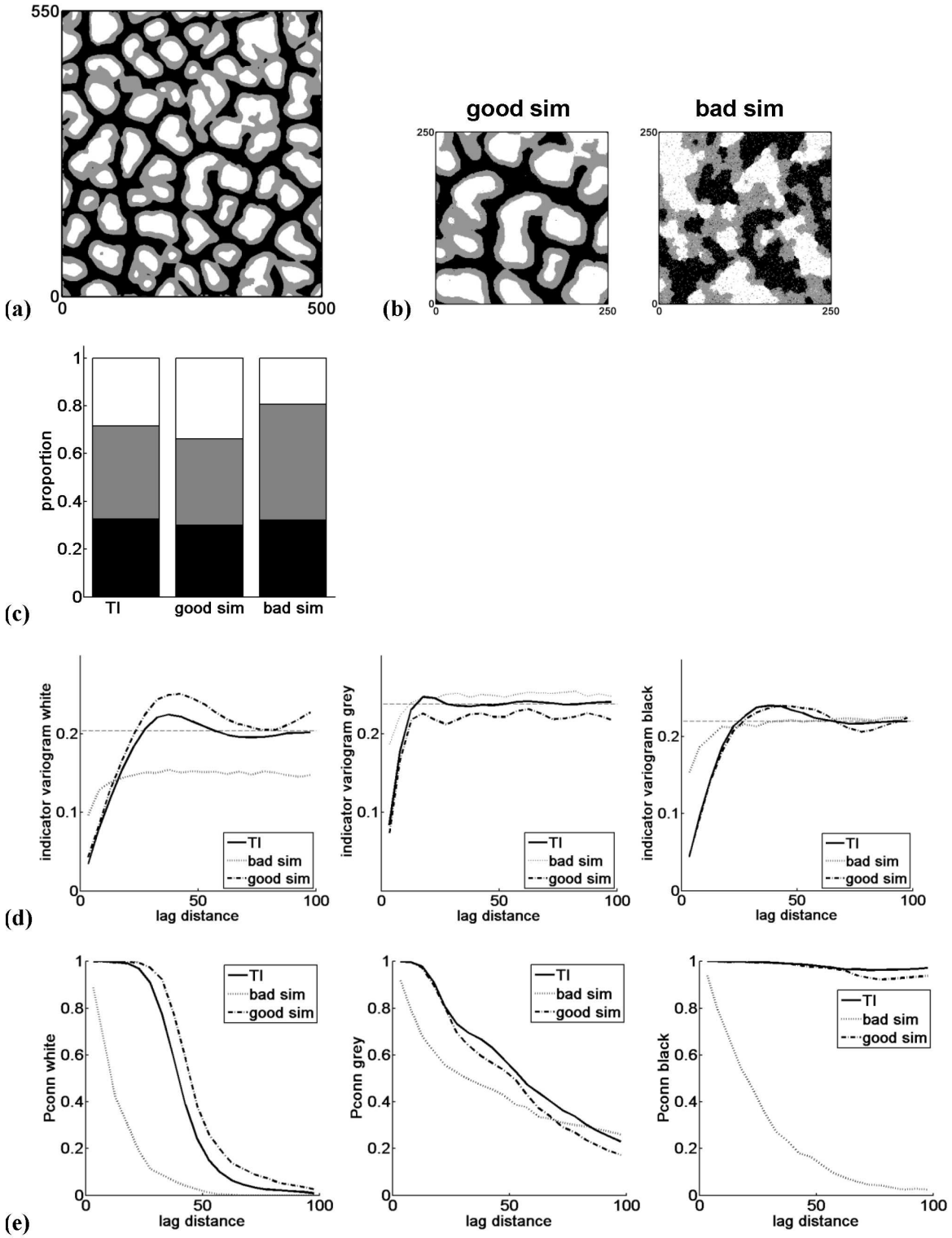

Figure 2.6: Reproduction of (a) the categorical ice-wedge TI statistics of (b) a good ($t = 0.01$, $f = 0.5$, $n = 80$) and a bad simulation ($t = 0.5$, $f = 0.5$, $n = 15$) with the reproduction of (c) the histogram, (d) the indicator variograms (the dotted lines represent the TI indicator variance) and (e) the connectivity functions.

with var_{sim} the simulation variance used to standardize the absolute errors, so they range between 0 and 1. The term $\frac{1}{\mathbf{h}_d}$ is included to give larger weights to errors corresponding to

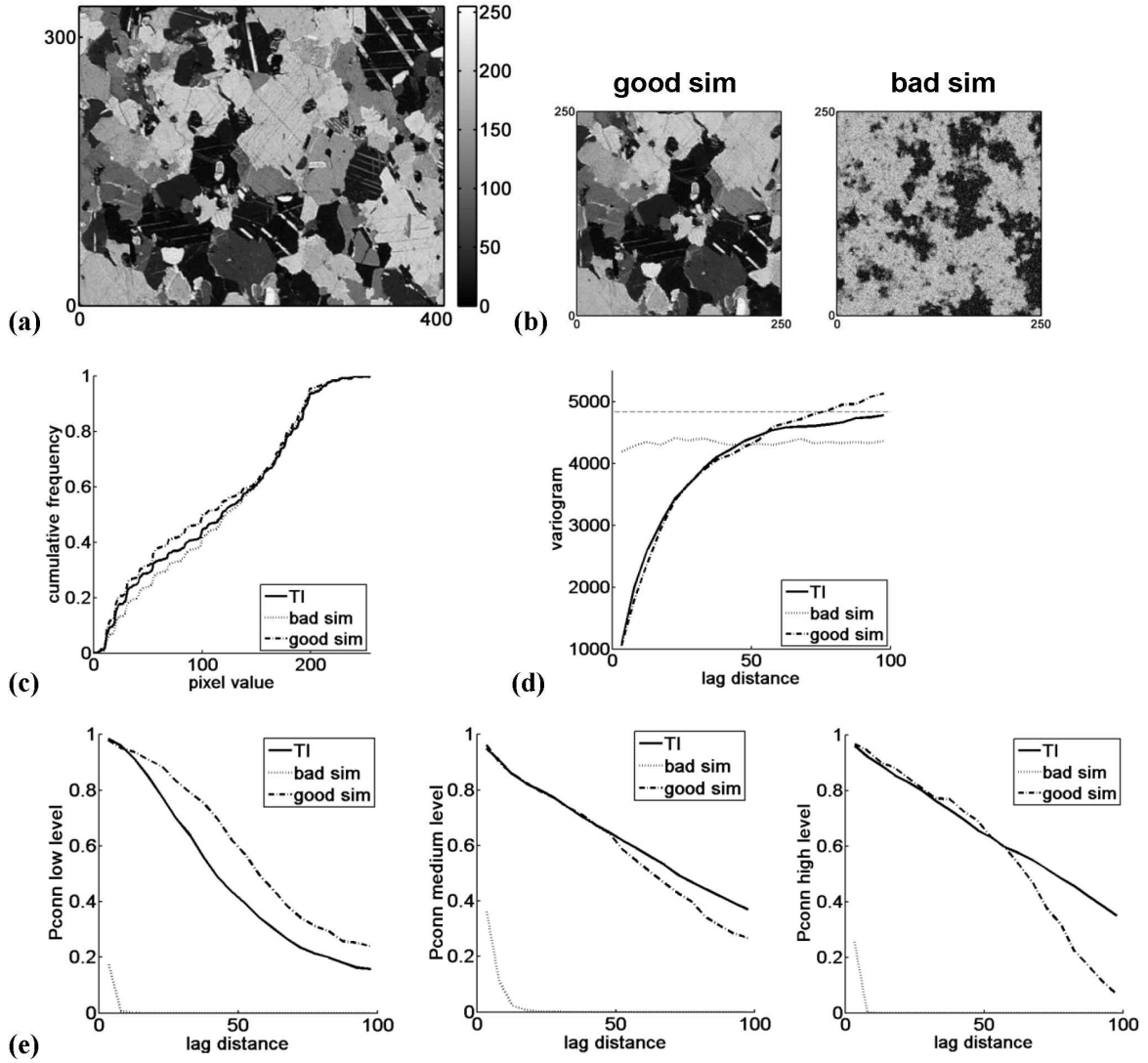

Figure 2.7: Reproduction of (a) the continuous marble TI statistics of (b) a good ($t = 0.01$, $f = 0.5$, $n = 80$) and a bad simulation ($t = 0.2$, $f = 0.5$, $n = 5$) with the reproduction of (c) the histogram, (d) the variogram (the dotted line represents the TI variance) and (e) the connectivity functions.

small variogram lags.

The variogram error ε_{var}^{k} for the categorical case was calculated similarly using the indicator variogram values.

The connectivity error ε_{conn} was calculated as

$$\varepsilon_{conn} = \frac{\sum_{d=1}^{20} \left| \tau_{sim}^{k}(\mathbf{h}_d) - \tau_{TI}^{k}(\mathbf{h}_d) \right|}{20} \tag{2.5}$$

and also ranges between 0 and 1.

2.2.5 Results and discussion

Fig. 2.8 and 2.9 show the results of the simulation quality indicators for the categorical and the continuous case. The first part of the figures illustrates the effect of t and n, the second

part the effect of f.

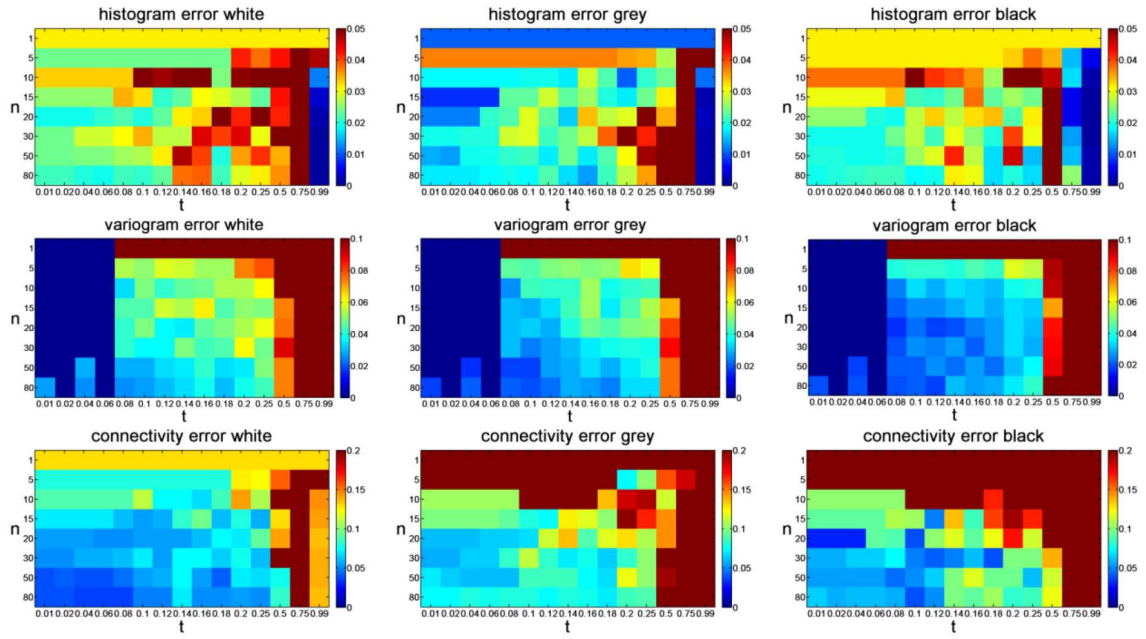

Figure 2.8: Influence of (a) t and n (for $f = 0.5$) and (b) f on the quality indicators based on the categorical ice-wedge TI (Fig. 2.1b).

As was already concluded from Fig. 2.6 and 2.7 ε_{hist} behaves differently than ε_{var} and ε_{conn}. The histogram was generally well reproduced for all simulations. Noisy images reproduced the histogram the best, which is especially clear for the continuous case. This can be explained by considering the extreme case of $t = 1$. With such a setting, the DS randomly samples values from the TI, resulting in a perfect reproduction of the marginal distribution ($\varepsilon_{hist} \approx 0$), but no reproduction of spatial continuity (very large ε_{var} and ε_{conn}).

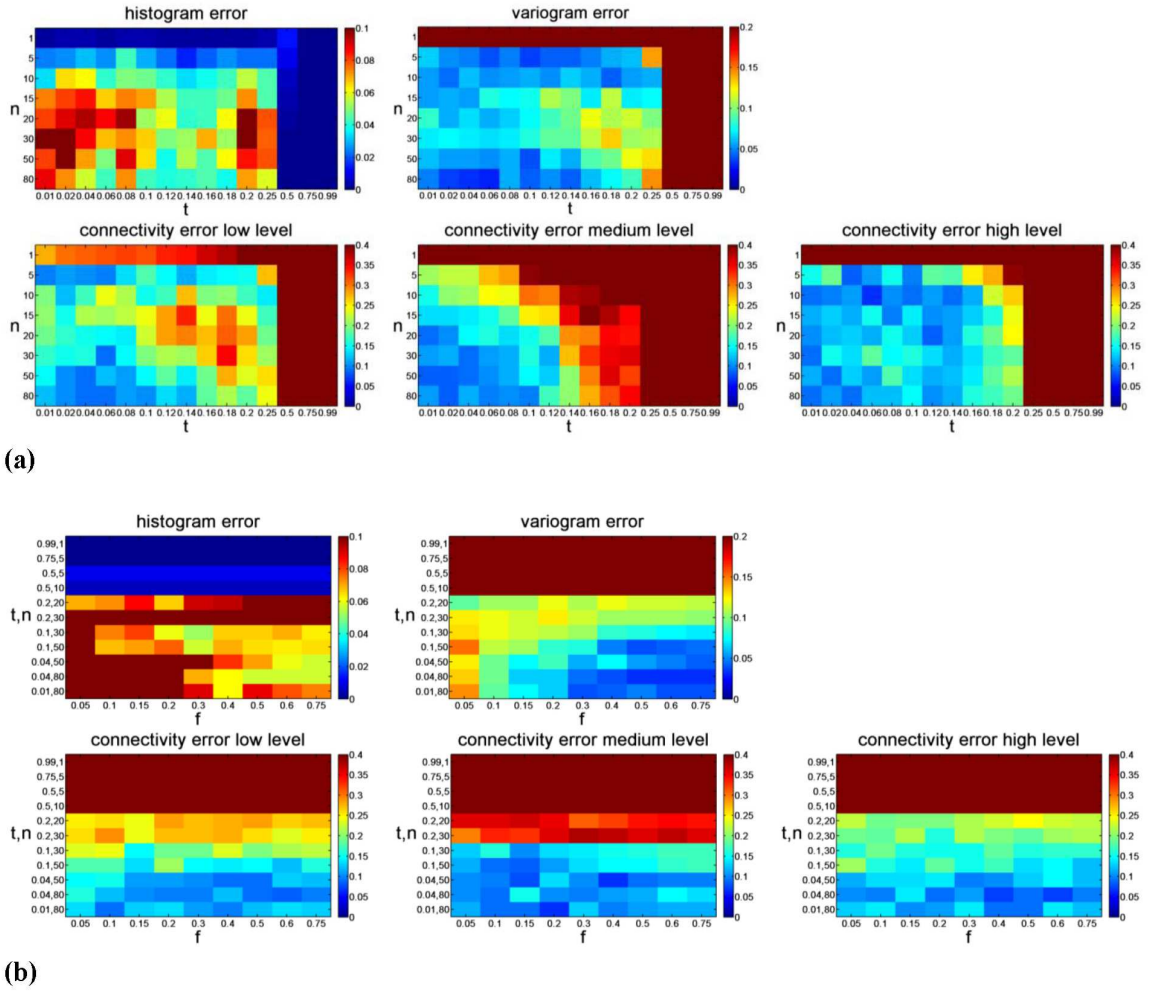

Figure 2.9: Influence of (a) t and n (for $f = 0.5$) and (b) f on the quality indicators based on the continuous marble TI (Fig. 2.1c).

For intermediate combinations of t and n, ε_{hist} is generally larger. In the areas with good simulation quality ($t \leq 0.2$ and $n \geq 30$) ε_{hist} behaves differently for the categorical and the continuous case. For the categorical case, low t and high n guarantee both low ε_{hist} and good simulation quality (Fig. 2.4, App. C, Fig. C.6 and C.7). For the continuous case, the high quality simulations ($t \leq 0.2$ and $n \geq 30$) have higher ε_{hist}. This counter-intuitive result can be explained as follows: with low t and high n, the simulation has to honor very strong spatial constraints. When the structures are made of objects that are large with respect to the domain size, respecting such spatial constraints means to respect the connectivity of facies and the objects size even if it contradicts the target pdf. Because of a slight non-stationarity in the TI, the simulation can then follow the pdf of one specific part of the TI that is different than the rest. This can result in significant variability in the pdfs of the simulations. This is the opposite as the case of $t = 0$, where the TI distribution is honored because there is no constraint on the spatial continuity.

For the continuous ice-wedge TI (Fig. 2.1a) and the snow crystal TI (Fig. 2.1e), the histogram was well reproduced in the high quality simulations (App. C, Fig. C.5 and C.8). Since certain applications require the histogram to be reproduced, this issue could be further addressed by the DS developers. In contrast ε_{var} and ε_{conn} increase for larger t and smaller n, which is a more intuitive behavior. Both errors show similar quality jumps as were derived

from the visual inspection and therefore behave as sTab. simulation quality indicators.

These results allow us to derive some rules of thumb. Running DS using a categorical TI should result in good simulations for $t \leq 0.2$ and $n \geq 30$. Selecting $t \geq 0.5$ and $n < 15$ should be avoided. For continuous TI, it is advised to use $t \leq 0.1$, $n \geq 30$ and to avoid $t \geq 0.2$ and $n \leq 15$. The quality of intermediate combinations is hard to predict. It is obvious that the simulation quality steps strongly depend on the TI. The greater the pattern repeatability in the TI, the better the simulation quality will be.

Fig. 2.8b and 2.9b lead again to the conclusion that f does not have a strong influence on the simulation quality. This was confirmed by the other TIs (App. C, Fig. C.5 to C.8). Only for certain situations, like for f<0.2 in the continuous case (see results for ε_{var}), the pattern reproduction degraded with lower f since the probability of finding a matching TI pattern is lower. Note also that using a small f value for TIs that contain insufficient diversity (Mirowski et al., 2009), might aggravate the statistical scarcity and lead to poor results. But generally decreasing f results in large computation gains without a substantial decrease in simulation quality, which is an important conclusion for an efficient use of DS.

It is interesting to juxtapose the CPU time (Fig. 2.3) with the corresponding quality indicators (Fig. 2.8 and 2.9). This reveals where the interesting boundaries are located in terms of quality over CPU time ratio. For instance, for the categorical case the quality was moderate from $n \geq 15$ and $t \leq 0.18$ ($f = 0.5$) (Fig. 2.8a), whereas the CPU time really increased from $n \geq 30$ and $t \leq 0.1$ (Fig. 2.3a). In between these boundaries, the simulation quality was good, as was confirmed by the visual inspection. In case CPU time is a limiting factor, users are recommended to investigate the quality over CPU time ratio for different parameter combinations running trial simulations on a small grid.

It is good practice to run DS initially with $f = 0.5$, t between 0.05 and 0.2 and n between 20 and 50. From this, the parameters need to be fine-tuned for every particular situation, knowing that decreasing t and increasing n and f should result in better simulation quality. However, one should keep in mind that using parameters which guarantee very good simulations has two drawbacks. First, these configurations will require large CPU times. Second, there is a risk of generating simulations which are all exact copies of (part of) the TI (patching effect or verbatim copy). This risk is higher when the TI does not show enough pattern repeatability (which is more often the case for continuous TIs) and when there are no conditioning data (CASE 5). Strategies to avoid patching are choosing f <1, thus avoiding to pick each time the same best matching mode, slightly relaxing t and n, or using a smaller 'maximum search distance'.

2.3 CASE 2: 3D simulation

Similarly to two dimensions (CASE 1), DS can generate 3D simulations. To demonstrate this, we performed a limited sensitivity analysis using the 3D concrete TI (Fig. 2.1g). We generated 10 unconditional simulations for each combination of eight t [0.01 − −0.05 − −0.1 − −0.15 − −0.2 − −0.25 − −0.3 − −0.5] and eight n [1 − −5 − −8 − −16 − −32 − −64 − −125 − 216] values, using a fixed value of 0.5 for f. The other parameters were set as indicated in Tab. 2.1, with exception of the maximum search distance that was defined as half of the search grid in three directions (x, y and z).

The CPU time as a function of t and n (not shown here) behaved very similar as for 2D (Fig. 2.3). For instance, generating one simulation with $t = 0.1$ and $n = 32$ took 194 s, with $t = 0.05$ and $n = 32$ 1998 s and with $t = 0.05$ and $n = 64$ 6804 s.

Fig. 2.10b shows the simulation quality indicators (see 2.2.4) as a function of t and n. Overall, the results were analogous to those of CASE 1. The simulation quality generally improved with increasing n and decreasing t with a quality jump for $t = 0.2$ and $n = 32$, as

can be seen from $\varepsilon_{conn}^{grey}$ and $\varepsilon_{conn}^{black}$, and the unconditional simulations shown in Fig. 2.10a. Again, ε_{hist} was the smallest for noisy simulations with very small n.

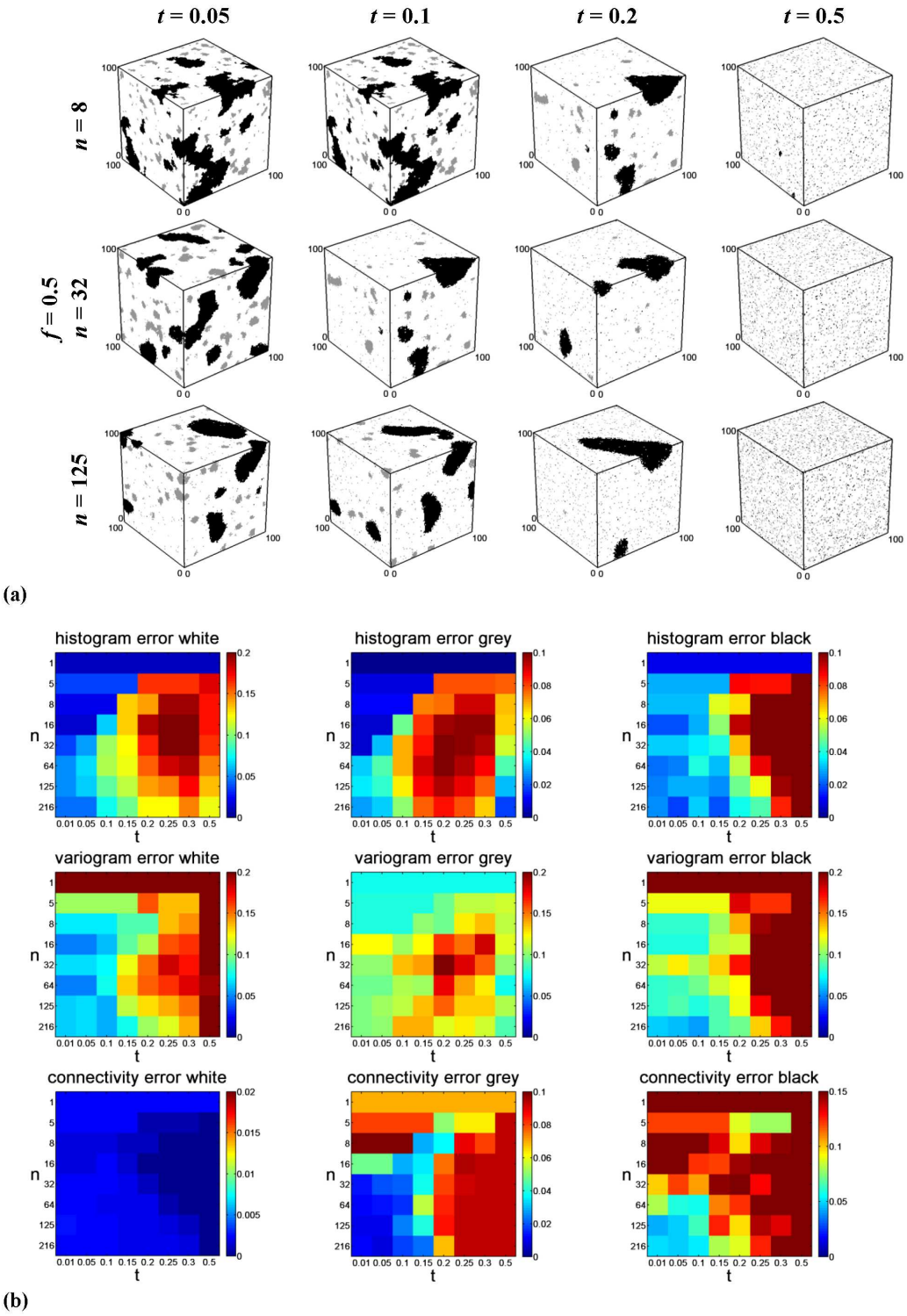

Figure 2.10: 3D example with (a) first out of 10 unconditional simulations illustrating the effect of parameters t and n with $f = 0.5$ and (b) influence of t and n (for $f = 0.5$) on the quality indicators based on the 3D concrete TI (Fig. 2.1g).

Since the white category represents the background volume, $\varepsilon_{conn}^{white}$ was very small for

all parameter combinations and hence not informative. With parameters producing noisy simulations ($t \geq 0.3$ and $n \leq 8$), $\varepsilon_{\text{var}}^{grey}$ was again lower. This could be explained by the small range of the grey indicator variogram, causing $\varepsilon_{\text{var}}^{grey}$ to be small for pure nugget variograms reproducing the sill correctly.

2.4 CASE 3: Post-processing for noise removal

To further improve the simulation quality and more specifically to remove noise, DS includes a post-processing option. After having simulated all the unknown grid nodes, it is possible to resimulate each node with an entirely informed neighborhood. Two post-processing parameters need to be defined: the number of post-processing steps p and the post-processing factor p_f. The latter is the factor by which f and n are divided aiming to save CPU time in the additional post-processing steps (Mariethoz, 2009). For example, if $p = 2$ and $p_f = 3$, all nodes are resimulated two times with parameters $f3$ and $n3$ times lower than their original values. Fig. 2.11 illustrates the noise removal effect of post-processing for increasing t and varying p and p_f for the categorical ice-wedge TI (Fig. 2.1b).

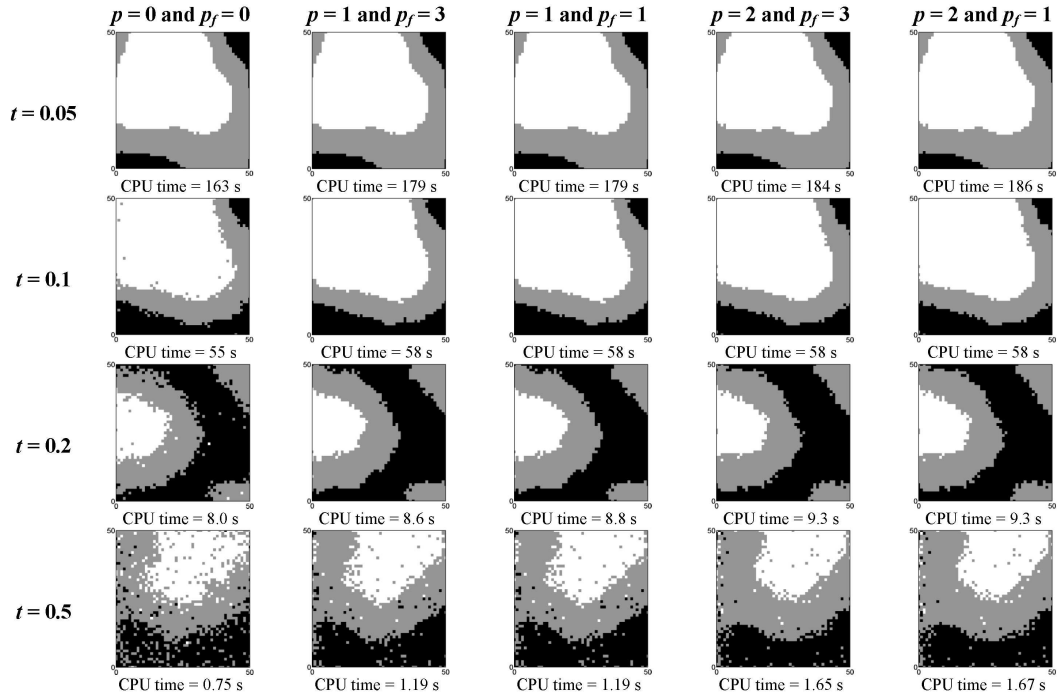

Figure 2.11: Illustration of the noise removal effect of post-processing using the categorical ice-wedge TI (Fig. 2.1b) for increasing t, and sensitivity analysis for the number of post-processing steps (p) and the post-processing factor (p_f) showing the lower left corner of the simulation grid.

The post-processing step proved to be valuable especially for intermediate t values (0.1 and 0.2), since the noise could be considered as entirely removed without substantially increasing CPU time. The simulations obtained with intermediate t after post-processing were similar to these obtained with small t, except for the boundaries which were less sharp. Furthermore, post-processing allowed for a significant reduction in CPU time. With $t = 0.1$ and one post-processing step, one obtained in 58 s realizations similar to when using $t = 0.05$ without post-processing, which took 163 s. For small t (0.05) the post-processing step was not necessary since the simulation quality was already good without it. For high t (0.5) it is insufficient since post-processing only removes noise and does not improve structures

at larger scale. Repeating the post-processing step did not result in significant quality improvements, and whether or not f and n were decreased in the post-processing step neither decreased the CPU time nor improved the simulation quality.

The effect of a post-processing step was less substantial for the continuous case than for the categorical case and the CPU cost was much higher (not shown here). Hence, for continuous cases, the quality loss of selecting a high t, cannot be corrected with one or more additional post-processing steps. Since p and p_f have to be chosen in advance, it can be considered as good practice to add one additional post-processing step when simulating categorical variables. When noise appears, it will be reduced and the extra CPU time needed is relatively low. For p_f a value of 1 can be selected, since adjusting p_f did not seem to have an effect. If the simulations still contain noise after post-processing, it is however advised to decrease t instead of adapting p and p_f.

2.5 CASE 4: Multivariate simulation

Among the MPS methods, only DS had demonstrated its potential to simulate m variables simultaneously based on m TIs. These variables can be continuous, categorical or a mixture of both since for each a different distance measure can be chosen (distance type parameter set to 0 for categorical and 1 for continuous). Several implementations have been tested. The one of Mariethoz et al. (2010) was used in this paper. First, a path is defined that goes randomly through all the unknown grid nodes for each variable $\mathbf{x}_m$. When one variable is simulated at one location, the other variable at the same location can be simulated later in the path. For each $\mathbf{x}_m$ a multivariate $\mathbf{d}_{\mathrm{n}}$ is built containing the neighboring data for the m variables, which do not have to be collocated. For each variable the maximum number of neighbors n_m can be defined separately. Based on a weighted average of the m selected distance measures, the multivariate TI pattern, centered at the same node for each TI, is chosen that is similar to the multivariate $\mathbf{d}_{\mathrm{n}}$. The weights used to define the multivariate distance measure w_m are user-defined. DS automatically normalizes their sum to one. The value at the central node of this multivariate TI pattern is then pasted at $\mathbf{x}_m$. If conditioning data are given for all or some of the m variables, they will be honored (Mariethoz et al., 2010) as shown in CASE 5.

A potential application is a situation where one variable is (partially) known and the other(s) are to be simulated (the collocated simulation paradigm). DS becomes especially interesting when the relationship between the variables is known via the training data set but not expressed as a simple mathematical relation. Applications can be found in Mariethoz et al. (2010), Meerschman and Van Meirvenne (2011) and Mariethoz et al. (2012). As an illustration we show five unconditional bivariate simulations using the categorical and continuous ice-wedge TI (Fig. 2.1a and 2.1b) as bivariate TI and performed a sensitivity analysis on the weights given to both variables (Fig. 2.12). For the other parameters we used the default values as given in Tab. 2.1: both n_{cat} and n_{cont} were 50.

Fig. 2.12 shows that not only the spatial texture of each TI was reproduced, but also the multiple-point dependence between the TIs. The weights given to each variable strongly influenced the continuous variable. The larger w_{cont}, the better the quality of the continuous variable. The quality of the continuous variable decreased for smaller w_{cont}. In such cases, the bivariate relationship between both TIs was well respected, but the spatial continuity of the continuous variables was not strongly imposed. The quality of the categorical simulations was less affected by the choice of the weights. Note that for large w_{cont} the continuous simulation was almost an exact copy of the continuous TI (Fig. 2.1a), which is again an example of the patching effect described in CASE 1.

These results and other numerical experiments (not shown here) suggest that it is often

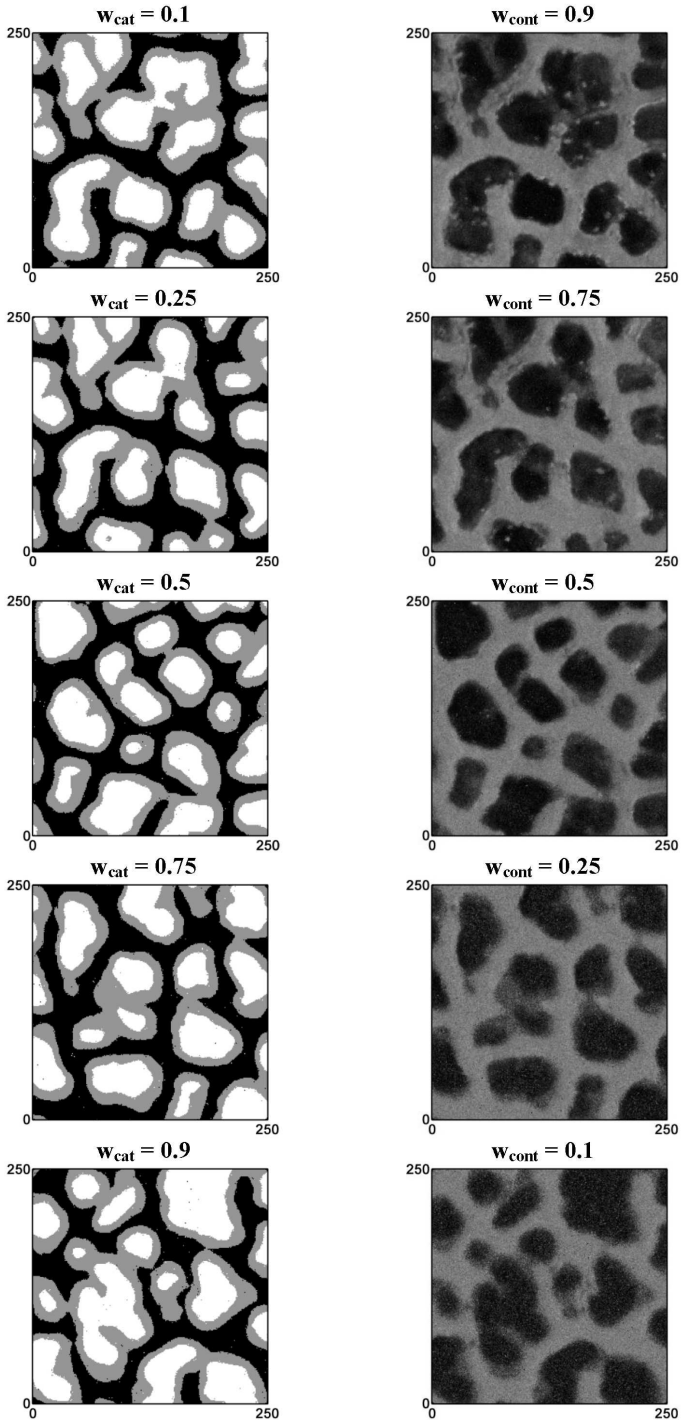

Figure 2.12: Illustration of the multivariate simulation option: five unconditional bivariate simulations using Fig. 2.1a and 2.1b as bivariate TI, and sensitivity analysis for the weights given to the two variables (w_{cat} and w_{cont}). The left column represents the categorical variable for each simulation, and the right column represents the corresponding continuous variable.

beneficial for the quality of the simulation of continuous variables to co-simulate a categorical variable that helps reproducing the continuity of the structures. This is a generally accepted technique in image processing in which the categorical variable is called 'feature map' (Lefebvre and Hoppe, 2006; Zhang et al., 2003).

2.6 CASE 5: Data conditioning

DS always honors conditioning data by assigning them to the closest grid node prior to simulation. Hence, local accuracy is guaranteed (the pixels at the data locations will have the correct values) but the simulated structures need to be coherent with the conditioning data. Therefore, one needs to check whether the fixed grid nodes are included in the spatial pattern or whether they appear as noise. The parameter that can be used to enforce pattern consistency in the neighborhood of the conditioning data is the data conditioning weight δ. This parameter is used in the distance computation to weight differently data event nodes that correspond to conditioning data. If $\delta = 1$, all the nodes in $\mathbf{d}_n$ have the same importance. For $\delta > 1$, higher weights are given to the data event nodes that are conditioning data, while for $\delta < 1$ they are given lower weights (Mariethoz et al., 2010; Zhang et al., 2006).

For both the categorical and continuous cases, one of the best unconditional simulations ($t = 0.01$, $f = 1$, $n = 80$) was used as reference image. To avoid using simulations that were copies of part of the TI due to patching, we first mirrored both simulations horizontally and vertically, before sampling 100 conditioning data from each according to a stratified random sampling scheme (Fig. 2.13 and 2.14). Using the default parameters for t, f and n (Tab. 2.1), we ran 50 simulations using these conditioning data and the corresponding TI. To remove noise, one post-processing step was performed with $p_f = 1$ (CASE 3). Conditioning data nodes are not resimulated during post-processing. For $\delta = 0$, $\delta = 1$ and $\delta = 5$, the first conditional simulation is shown together with the conditional probabilities for category k in the categorical case (Fig. 2.13), and the median over the 50 simulations for the continuous case (Fig. 2.14).

It can be concluded that δ is an important parameter when conditioning data are available. For $\delta = 0$ the 50 simulations could be considered as unconditional simulations, since the conditioning data grid nodes were ignored in $\mathbf{d}_n$. The simulation patterns were not at all consistent with the conditioning data and the large variation between the simulations resulted in non-informative summarizing images. For $\delta = 1$ the simulations showed patterns that were more or less consistent with the conditioning data. The remaining inconsistencies disappeared with $\delta = 5$, resulting in summarizing images that closely resembled the reference images. The better results for $\delta = 5$ are due to the high quality of the conditioning data, which perfectly represented the reference image without measurement errors. Generally, we advise to set δ to a value higher than or equal to 1. The lower the expected uncertainty of the conditioning data, the higher δ can be chosen.

Note that for this example the conditioning data were sampled from a field with a spatial pattern that was very similar to the TI. When one expects that the underlying spatial pattern of the conditioning data deviates more from the TI, the use of transform-invariant distances can be beneficial. This option of DS increases the number of TI patterns by randomly scaling or rotating the patterns found in the TI (Mariethoz and Kelly, 2011).

2.7 Conclusions

This paper has reported the first comprehensive sensitivity analysis for the DS algorithm, aiming to encourage users to benefit more efficiently from the potential of DS and its wide spectrum of applications. Given these results we provide the following general guidelines.

For categorical TIs, choosing $t \leq 0.2$ and $n \geq 30$ will generally result in high quality simulations. Smaller t and larger n result in better simulation quality a lower the level of noise. However, this choice will also depend on the available CPU time. Furthermore, for small t and large n, the user should check if there is still sufficient variability between the simulations. For continuous TIs, we advise to select $t \leq 0.1$ and $n \geq 30$. For continuous

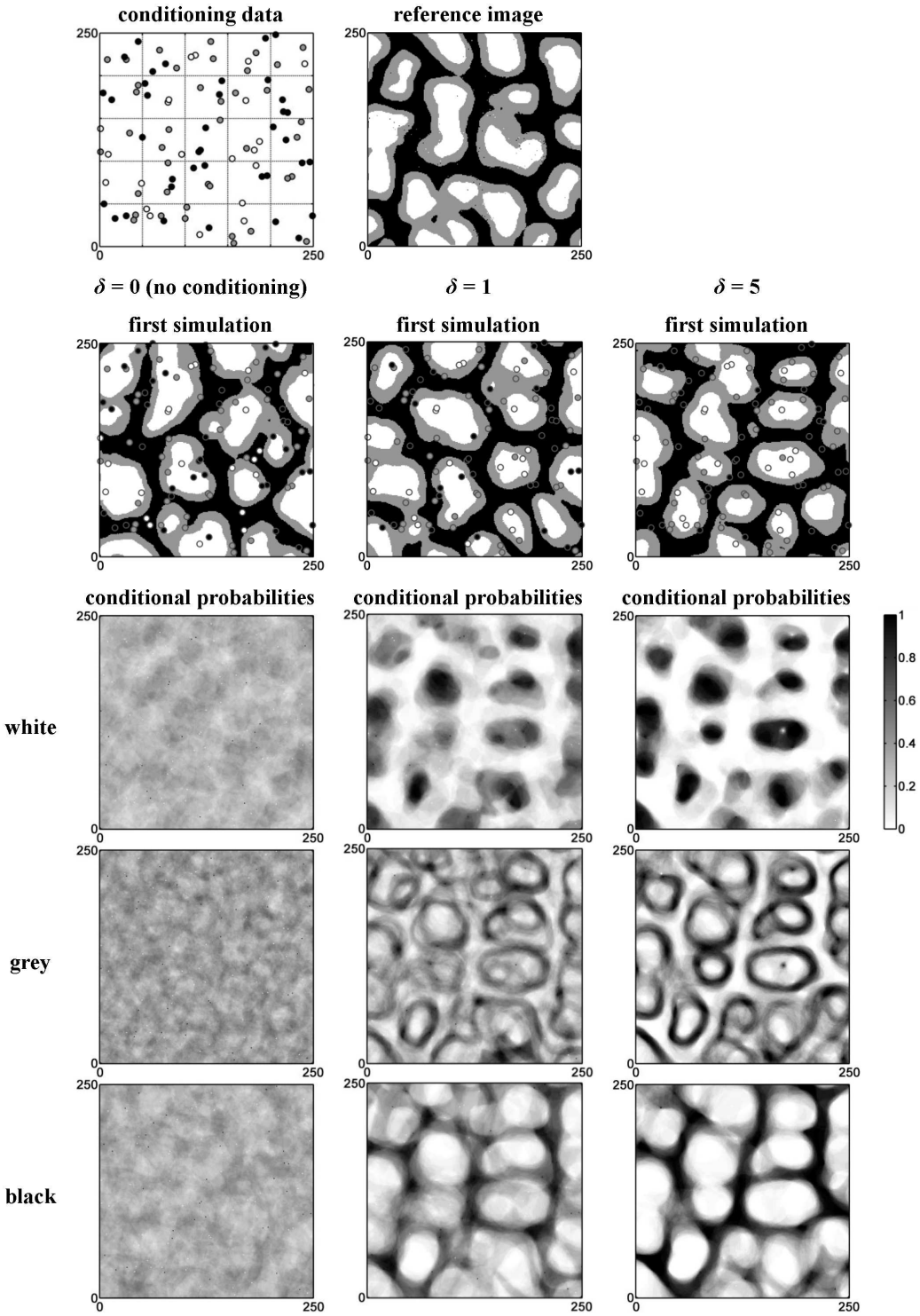

Figure 2.13: Illustration of data conditioning for the categorical ice-wedge TI (Fig. 2.1b) based on 100 conditioning data. For $\delta = 0$, $\delta = 1$ and $\delta = 5$ the first simulation is shown together with the conditional probabilities for each category summarizing 50 simulations.

cases, the selection of t and n is a delicate balance between ensuring good simulation quality and still guaranteeing sufficient variability between the simulations (avoiding patching). A good strategy to reduce both CPU time and the risk of patching is setting $f < 1$, and

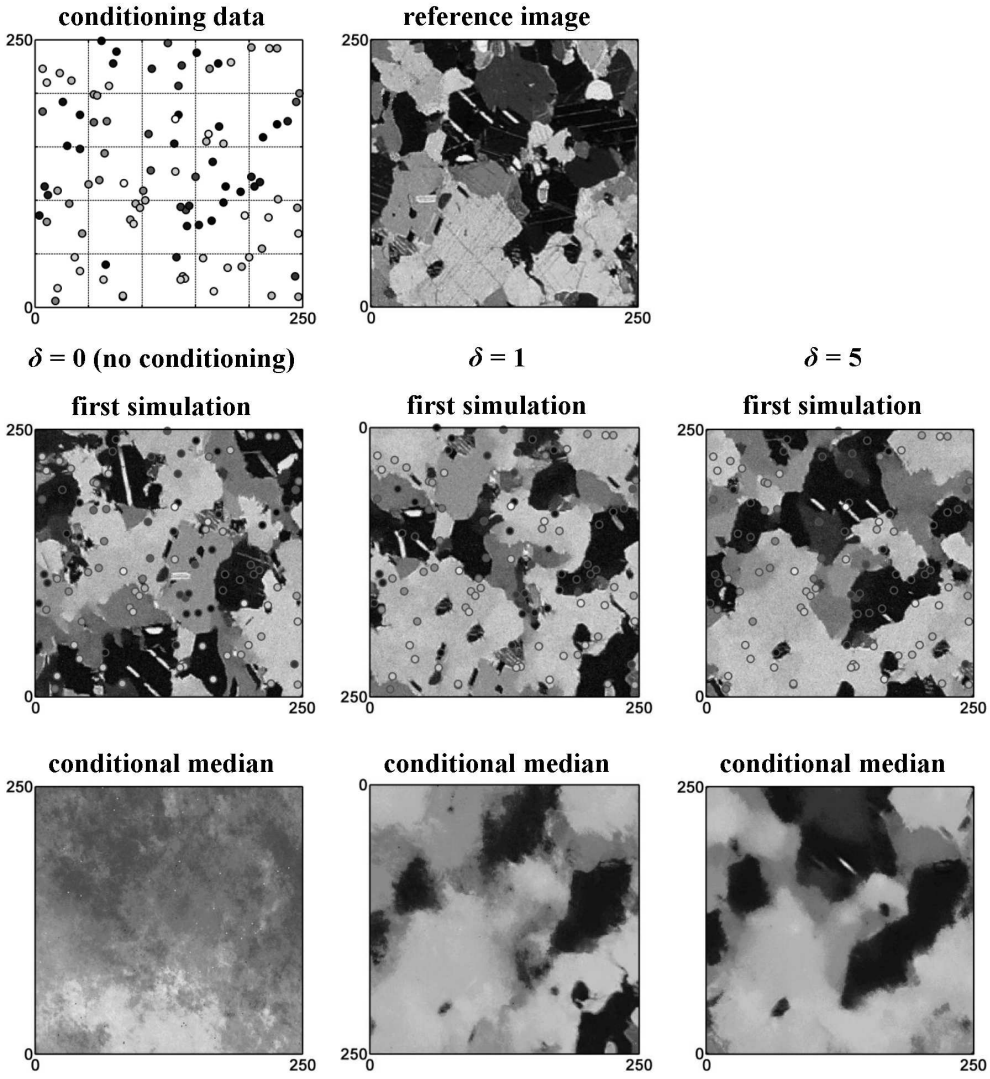

Figure 2.14: Illustration of data conditioning for the continuous marble TI (Fig. 2.1c) based on 100 conditioning data. For $\delta = 0$, $\delta = 1$ and $\delta = 5$ the first simulation is shown together with the conditional median for each category summarizing 50 simulations.

reducing the maximum search distance to a third the domain size or less, thus scanning a different fraction of the TI for each unknown grid node.

For categorical simulations in particular, it is advised to always add one post-processing step for noise removal. If the final simulations still contain (too much) noise, improvement should be sought by adapting t and n.

Simulating bivariate images is a very new and promising technique first offered by the DS algorithm. With the illustrative example in this paper we have shown that the weights given to each variable clearly affect simulation quality. In case of continuous variable simulation, it is beneficial to add an auxiliary categorical variable that is co-simulated with a relative small weight. This generally improves the simulation of the continuous variable.

When conditioning data are available, it is interesting to put the weights given to the conditioning data (parameter δ) higher than the weights given to the already simulated nodes. This results in simulated patterns more consistent with the conditioning data.

Acknowledgments

The work presented in this paper was financially supported by the Flemish Fund for Scientific Research (FWO-Vlaanderen), by the Swiss National Science Foundation under the contract CRSI22_122249/1, and by the National Centre for Groundwater Research and Training.

Chapter 3

Simulation of braided river elevation model time series with Multiple-Point Statistics

Abstract

A new method is proposed to generate successive topographies in a braided river system. Indeed, braided rivers morphology is a key factor influencing river-aquifer interactions and having repercussions in ecosystems, flood risk or water management. It is essentially based on multivariate multiple-point statistics simulations and digital elevation models as training data sets. On the one hand, airborne photography and LIDAR acquired at successive time steps have contributed to a better understanding of the geomorphological processes although the available data are sparse over time and river scales. On the other hand, geostatistics provide simulation tools for multiple and continuous variables, which allow for exploring the uncertainty of many assumption scenarios. Illustration of the approach demonstrates the ability of multiple-point statistics to produce realistic topographies from the information provided by digital elevation models at two time steps.

3.1 Introduction

Braided rivers constitute an important part of alluvial systems in alpine regions such as Switzerland. Many of these rivers were channelized in the past and are now targeted by restoration projects (Glenz, 2013) for flood prevention, water management purposes, biodiversity preservation, and leisure activities (FOEN, 2009; Peter, 2009), particularly in a context of climate change (Macklin and Rumsby, 2007). As a result of the erosion and deposition processes, the morphology of braided rivers is a signature of such active systems. Morphology is a key parameter, first toward the understanding of the dependant ecosystems (Amoros and Bornette, 2002; Richards et al., 2002; Clarke et al., 2003; van der Nat et al., 2003; Tockner et al., 2009), and also to better understand the main geological structures of the resulting aquifers in order to study groundwater flow and transport (Thomas and Nicholas, 2002; Käser et al., 2014), or surface and subsurface relationships. In a hydrogeological context, simulations of successive morphologies could also be used to produce three-dimensional heterogeneous geological models. These issues are not addressed in this paper but they justify the need of topography models. The purpose of this work is to present a new way of modeling braided river topography and its evolution.

Static models of braided river morphology can be achieved by LIDAR data acquisition followed by image processing (Westaway et al., 2003) and analyses can be derived from descriptive methods characterizing the length scale and the main topographic structures (Rust, 1972; Miall, 1977; Germanoski and Schumm, 1993; Goff and Ashmore, 1994; Warburton and Davies, 1994; Foufoula-Georgiou and Sapozhnikov, 2001; Hundey and Ashmore, 2009; Lane, 2009). But these approaches are often limited to a single time step (static aspect) and restrained to the area of acquired data.

Simulations based on process imitating methods such as cellular automata models (Murray and Paola, 1994; Nicholas et al., 2009) or such as event-based models (Pyrcz et al., 2009), which can be validated by comparisons to laboratory experiments (Ashmore, 1982), allow for models of the system over successive time steps. Nevertheless, the conditioning to field measurements such as borhole data is often very difficult. To overcome this drawback, an alternative is the use of MPS simulations. These techniques are non parametric and allow for reproduction of complex spatial features from a conceptual model called training image (TI), as well as to account for conditioning to field data. To our knowledge, multiple-point statistics (MPS) has not yet been used to simulate successive braided river morphologies. MPS has been introduced by Guardiano and Srivastava (1993), and first practical algorithms such as SNESIM (Strebelle, 2002) were designed for the simulation of categorical variables. The algorithm proposed by Mariethoz et al. (2010), the direct sampling (DS), is much more

flexible and can deal with joint simulations of multiple categorical and continuous variables. Because the DS method can reproduce spatial structures and complex correlations between several continuous variables (Mariethoz et al., 2012), this algorithm allows for simulations of successive digital elevation models (DEMs).

Following those ideas, in this paper, the principle of a new method to simulate braided river topographies at successive time steps is proposed. It combines the advantages of the DS algorithm, with the large-scale data available from LIDAR topography. The approach is illustrated with statistical simulations of topography time series. The training datasets (TIs) are based on successive DEMs of the Waimakariri River, New Zealand (Lane et al., 2003) aquired with LIDAR at four time steps.

The paper is structured as follows. The DS algorithm is briefly presented in section 3.2. The simulation of DEM time series with MPS algorithms is not straightforward, mainly because of large scale heterogeneities and trends in the TIs. Therefore 3.3 describes first a data analysis of the available TIs. That leads us to propose a methodology making use of auxiliary variables to enable realistic simulations of successive DEMs, with respect to the observed non stationarities present in sparse training data sets. The method is detailed and demonstrated within section 3.4. The paper ends with a statistical validation of the simulations in section 3.5.

3.2 The direct sampling, an MPS algorithm

Multiple-point statistics (MPS) algorithms allow us to simulate a random function Z on a domain called the simulation grid. The random function spatial statistics are retrieved from a conceptual model known as TI. In the TI, Z is known over all its domain (Fig. 3.1).

Each pixel of the simulation grid is simulated sequentially, one after another. A random path visiting every node in the simulation grid is defined, and each location $\mathbf{x}$ in the path is successively simulated as follows. The data event $\mathbf{d}(\mathbf{x})$ is the pattern constituted by the spatial ensemble of known values $Z(\mathbf{x}+\mathbf{h_i})$ in the neighbourhood of $\mathbf{x}$ ($\mathbf{h_i}$ being a lag vector), i.e., the conditioning data and the already simulated points. Then, the value $Z(\mathbf{x})$ to simulate at location $\mathbf{x}$ is drawn from the cumulative distribution function F conditionally to the local data event $\mathbf{d}(\mathbf{x})$: $F(z, \mathbf{x}, \mathbf{d}(\mathbf{x})) = Prob\{Z(\mathbf{x}) \leq z | \mathbf{d}(\mathbf{x})\}$. F is derived from similar local data events present in the TI. F can be dealt with in two ways.

The first efficient MPS approaches — implemented for instance by Strebelle (2002) in SNESIM or by Straubhaar et al. (2011, 2013) in IMPALA — rely on computing an histogram of the patterns present in the TI and compatible with the local data event $\mathbf{d}$. These approaches are limited to categorical variables and with fixed search neighbourhood templates. The use of successive multigrids simulation allows us, in that case, to capture and reproduce the structures from the TI at different scales while *using small search templates*, but may create some artefacts.

In a more recent approach, the DS algorithm (Mariethoz et al., 2010) skips the patterns histogram computation. Instead, for the simulation of a value at location $\mathbf{x}$ in the simulation grid, the TI is randomly scanned until a location $\mathbf{y}$ with a pattern $\mathbf{d}(\mathbf{y})$ compatible with the data event $\mathbf{d}(\mathbf{x})$ is found. Then, the value $Z(\mathbf{y})$ is simply pasted into the location $\mathbf{x}$, and the simulation continues with the next node $\mathbf{x}$ in the random path. The data event retrieved from the simulation grid is made up, at maximum, of the N closest informed nodes. Then, the scale covered by the data event is large in the beginning of the simulation, and becomes smaller at the end, allowing the capture of structures within the TI at different scales. Specifying a normalized distance $D(\mathbf{d}(\mathbf{x}), d(\mathbf{y}))$ between patterns, the compatibility of two patterns is defined as follows: $\mathbf{d}(\mathbf{x})$ and $\mathbf{d}(\mathbf{y})$ are compatible if and only if $D(\mathbf{d}(\mathbf{x}), \mathbf{d}(\mathbf{y})) < t$, where t is a tolerance (or threshold) chosen by the user. Moreover, in order to reduce the

computational time for the simulation of one node $\mathbf{x}$, a maximal fraction f (chosen by the user) of the TI is scanned and, if a compatible pattern is not found, the value $Z(\mathbf{y})$ at the visited location $\mathbf{y}$ corresponding to the best match between $\mathbf{d}(\mathbf{x})$ and $\mathbf{d}(\mathbf{y})$ is assigned to $\mathbf{x}$. For a categorical variable, the distance $D(\mathbf{d}(\mathbf{x}), \mathbf{d}(\mathbf{y}))$ is usually defined as the proportion of the nodes in the patterns having a different value. Normalized distances (of type L_1 or L_2 for instance) can be used to deal with a continuous variable. The algorithm is straightforwardly extended to the multivariate case: in a situation with m joint variables, the simulation proceeds by comparing m pair of data events. Fig. 3.1 illustrates the basic principles of DS. Further details about the DS algorithm can be found in Mariethoz et al. (2010) and

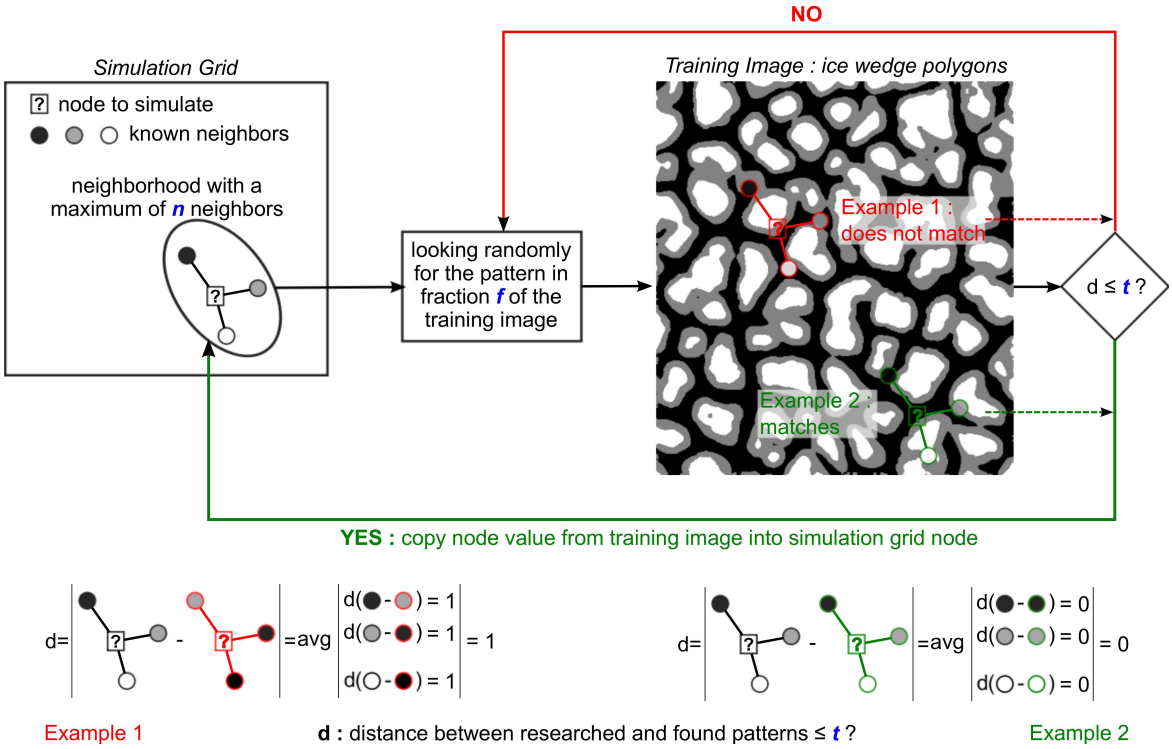

Figure 3.1: Workflow of DS, courtesy of Meerschman et al. (2013).

additional examples of DS simulations are available in the papers of Mariethoz and Kelly (2011) and Meerschman et al. (2013).

3.3 Topography data analysis and description with auxiliary variables

In this section, we focus on the LIDAR data analysis, showing why it cannot be used directly as a TI, and how to overcome this issue.

3.3.1 DEM as training image

The DEM data used in this study were provided by Lane et al. (2003). This data set from the Waimakariri River was acquired between February 1999 and May 2000 at four different time steps as shown in Fig. 3.2. Originally, the data had a resolution of 1 $m/pixel$, and the elevation was the absolute elevation above sea level. In order to work with simulation grids larger than the TI, and to give as much importance to the change of elevation in both transversal and longitudinal directions, the main trend of the DEMs, i.e., their main slope,

was removed. In other words, the average elevation computed over the longitudinal (x) axis is removed from the DEM. It means that for each x-coordinate (one column here), the average elevation is computed and subtracted to the measured elevation for all points of this x-coordinate. To speed up the MPS simulations, the resolution is coarsened by a factor of 20 ($\Delta X = 20 \times dx_{initial}$, $\Delta Y = 20 \times dy_{initial}$). The upscaling is done by averaging the local altitudes. It induces a smoothing of the data, but still respects the main topography structures from the original data (Lane et al., 2003).

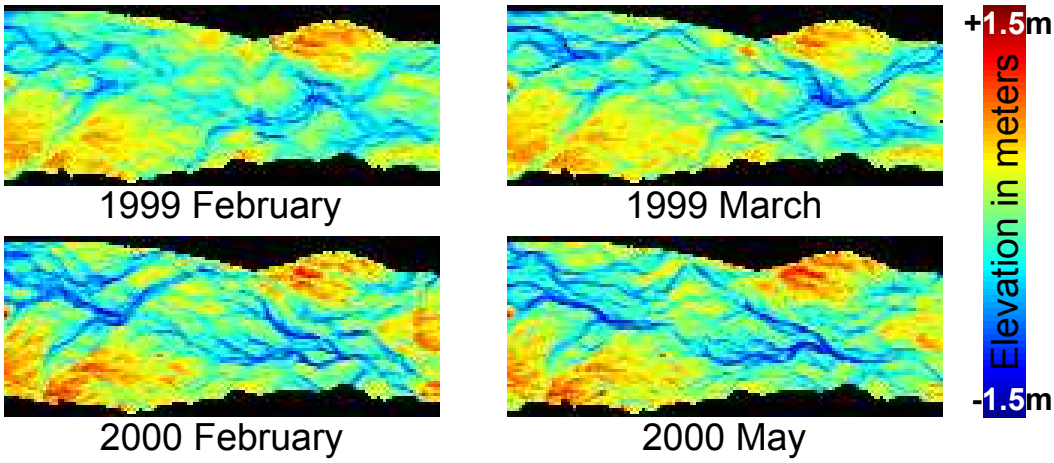

Figure 3.2: 2900 m × 1200 m (145 $pixels$ × 60 $pixels$ - x axis × y axis) DEMs of the Waimakariri River, New Zealand at four time steps. Red: highest elevations; blue: lowest elevations.

Some simulations of DEMs with the DS algorithm, using directly the Waimakariri River DEMs (Fig. 3.2) as TI, are presented in Fig. 3.3. We can observe that many structures of

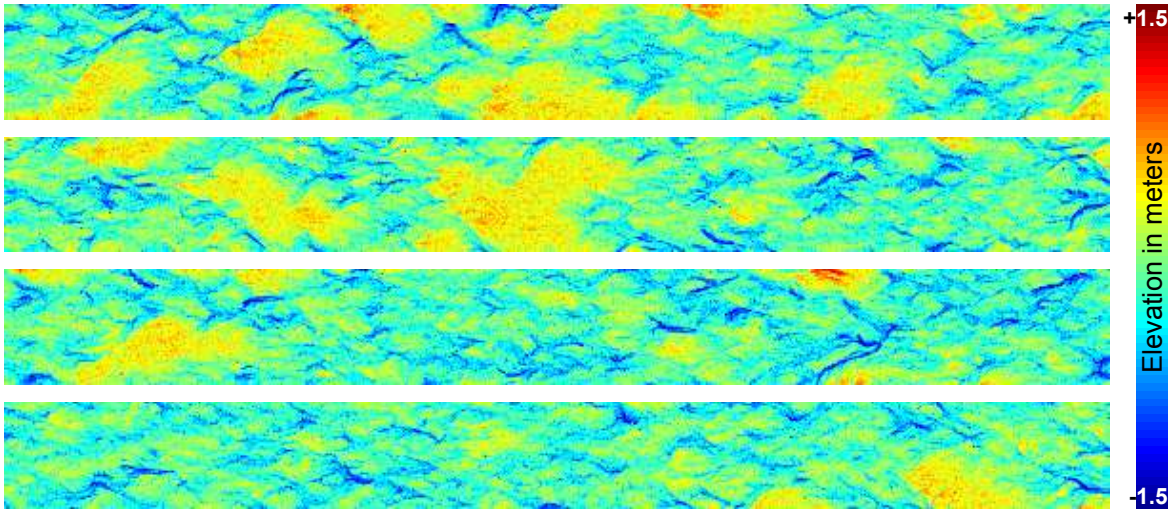

Figure 3.3: 11600 $m \times 1200\,m$ (580 $pixels \times 60\,pixels$) simulations of DEMs with DS, using a monovariate TI. Red: highest elevations; blue: lowest elevations.

the topography present in the TI can be reproduced in the simulations. But an important drawback shows up: higher elevation parts in yellow or red and deeper channels in dark blue seem to be randomly located in the simulation grid. However, the higher parts are expected to be located on the sides of the domain. It is an observation that the human brain can deduce from the TI, but that the algorithm can not infer automatically. This is a well known

problem — the MPS algorithm assumes that the TI is stationary, meaning that any pattern present in the TI has the same probability to occur at any location. If this is not the case, special techniques must be applied. Nonstationarity in the TI can be overcome by using probability maps or auxiliary variables (Chugunova and Hu, 2008; Boucher, 2009; de Vries et al., 2009) in classical MPS algorithms such as SNESIM (Strebelle, 2002) or IMPALA (Straubhaar et al., 2011). With the DS algorithm, auxiliary variables can be used as well to guide the spatial positioning of the topography structures (Mariethoz et al., 2010).

3.3.2 Defining auxiliary variables

Looking at the DEM variable in Fig. 3.2, two main features can be distinguished: the less active part of the river is characterized by larger and higher elevation zones on the sides of the river, and the most active part of the river called the *main channel*, is characterized by deeper channels and small to medium elevation zones.

The main channel can be represented by its centerline. This allows us to compute a distance between each location in the river and this centerline. This first auxiliary variable — distance to the centerline — shall prevent to generate large and high elevation zones within the main channel. In addition, a distance can be computed between each location in the river and the river edges. This second auxiliary variable — distance to the edges — shall ensure the absence of channels or pools cutting one of the large and high elevation zones present on the edges. As illustrated in section 3.4.2, these two distances will prove to be sufficient to help the algorithm localize the topography structures. Figure 3.4 gives an overview of the process to compute the centerline and the auxiliary variables. More details

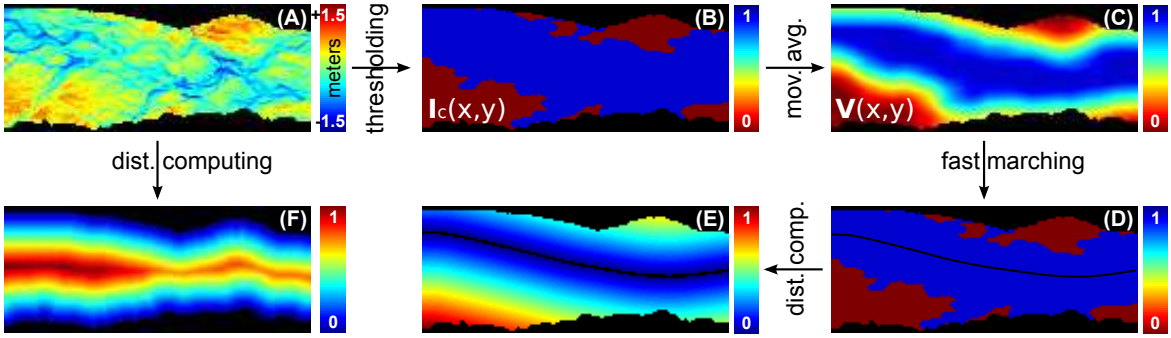

Figure 3.4: Auxiliary variables workflow. (A) initial DEM; (B) main channel; (C) velocity field; (D) main channel and its centerline; (E) normalized distance to the centerline; (F) normalized distance to the edges.

are available in the following paragraphs. The process is repeated for all available DEMs (Fig. 3.2).

Delineating a centerline in the main channel

The (x, y) coordinates system, used in what follows, refers to the x (longitudinal) and y (transversal) axes introduced in section 3.3.1.

First of all, the main channel is delineated by thresholding the elevation in the DEMs. In our example, a threshold of $0.3\,m$ is used. Zones above the threshold are post-processed to remove isolated groups or zones that are not adjacent to the edges. The result is an indicator function $I_c(x, y)$ locating the main channel when $I_c(x, y) = 1$, as shown in blue in Fig. 3.4B.

Then the centerline is obtained, by using a Matlab implementation of the fast-marching algorithm (Cao and Greenhalgh, 1994; Sethian, 1996; Peyré and Cohen, 2006). It consists

in finding a least effort path — i.e. shortest distance path — in a velocity field between two points. The velocity field $V(x, y)$ (Fig. 3.4C) is obtained by averaging the main channel indicator function $I_c(x, y)$ with a $20 pixels$ radius moving window, $V(x, y) = \frac{1}{S} \times \int_S I_c(x, y) dS$. The starting point or inlet is defined on the left edge as the point of coordinate $Inlet_x = 0$ and $Inlet_y$ as the average of y for which $max\{V(x, y), x = 0\}$ is reached. The ending point or outlet is chosen on the right edge as the point of coordinate $Outlet_x = 145$ and $Outlet_y$ as the average of y for which $max\{V(x, y), x = 145\}$ is reached. The resulting centerline is displayed in Fig. 3.4D.

Now that both the centerline and the edges of the river are known, the auxiliary variables can be computed.

Distance to the centerline

The distance to the centerline is defined for each point of the TI as the minimum euclidean distance between the considered point and all the points describing the centerline. It is then normalized by the maximum distance over the domain. The variable is displayed in Fig. 3.4E.

Distance to the edges

The distance to the edges is defined for each point of the TI as the minimum euclidean distance between the considered point and all the points describing the edges. It is then normalized by the maximum distance over the domain. The variable is displayed in Fig. 3.4F.

3.4 Simulation of successive topographies

We proceed in three main steps to simulate successive DEMs.

The first step (section 3.4.1) consists in defining the initial setup. It consists of :

- creating an initial centerline,
- simulating the river edges — these are then fixed across time and valid for the whole time-series and they determine the lateral size of the simulation domain for the topographies —
- generating a successive centerline time-series,
- computing the associated auxiliary variables.

The second step (section 3.4.2) is to simulate an initial topography. The third step (section 3.4.3) is a Markov chain transition kernel that consists of simulating the topography at time step $n + 1$, with respect to the previously simulated topography at time step n. All simulations are obtained with the DS algorithm.

3.4.1 Test case setup

For this example, 100 time steps are considered for a two-dimensional topography simulation grid of $580\, pixels \times 84\, pixels$ that corresponds to $11,600\, m \times 1680\, m$.

Edges and centerlines are two-dimensional features that can be described as one-dimensional variables as shown by Mariethoz et al. (2014). In our case, the centerline or the edges are described not as a succession of angles but as a succession of derivatives dy/dx for simplicity. This allows us to work on a fixed length of $580\, pixels$ along the x-axis (longitudinal). The

edge and centerline derivatives TI used for the following simulations are extracted from the four successive DEMs of the Waimakariri River (Fig. 3.2).

Creating an initial centerline and river edges

The first centerline is obtained by MPS simulation of its lateral derivative along the longitudinal axis. The cumulative sum of the derivatives gives the lateral coordinate for each point on the longitudinal axis of the grid.

The edges are similarly simulated with MPS but constrained loosely to some conditioning points. For the upper or the lower edges, we compute a straight line parallel to the linear approximation of the centerline but ensuring a minimum distance of 10 $pixels$ with the centerline. Then for each edge, the conditioning points are the starting point, ending point, and 15 points uniformly drawn from the corresponding straight line. Once the edges are simulated, we consider them fixed and valid for all time steps. The width (y-axis) of the simulation domain is then fixed to 84 $pixels$ in our example.

The conditioning of edge simulations to sampled points is obtained by using an iterative accept–reject algorithm. The sampled points are ordered as (P_i). Let us denote S_i the i^{th} segment (i.e., part of a river edge) between P_i and P_{i+1}. The segments are successively simulated. To generate a segment S_i, successive derivatives are simulated and cumulatively added to obtain a line starting from P_i; then if the distance d_{i+1} between its ending point is too far from P_{i+1}, for instance $d_{i+1} \geq 2pixels$, the simulation of S_i derivatives is rejected, and another one is launched, until it is accepted. When a segment S_i is accepted, the simulation continues with the simulation of the next segment, S_{i+1}, starting from the ending point of S_i. A resampling (of points P_i) is done if too many simulation attempts are rejected.

Generating a time-serie of successive centerlines

Knowing the n^{th} centerline, the lateral coordinates of both starting point and ending point are locally perturbed. Additional points are retained between the starting point and the ending point with a longitudinal interdistance uniformly drawn between 40 $pixels$ and 100 $pixels$. Both lateral and longitudinal coordinates of these points are locally perturbed. The lateral perturbations are uniformly drawn in $[-3\ pixels, 3\ pixels]$. The longitudinal perturbations are uniformly drawn in $[-5pixels, 5pixels]$. All the perturbed points are used as conditioning data. Only centerlines that exist between the edges and with a minimum distance of $10pixels$ to the edges are kept. If the simulation of the $n+1^{th}$ centerline is rejected, the n^{th} centerline is sampled again and the $n+1^{th}$ centerline is resimulated. Figure 3.5 illustrates the first four centerlines.

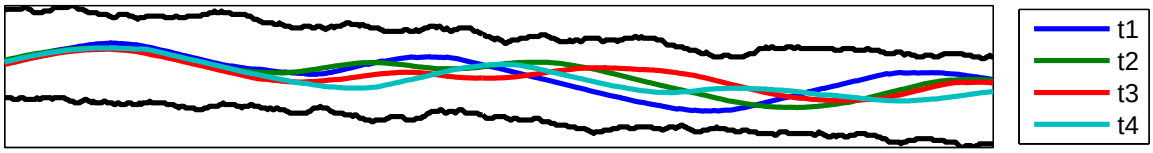

Figure 3.5: Simulated edges and centerlines at four successive time steps for 580 $pixels \times$ 84 $pixels$ simulations.

Computing auxiliary variables

Once edges and centerlines are defined, the fully informed auxiliary variables may be computed, as explained in section 3.3.2.

3.4.2 Initial topography

To generate an initial DEM, a slope-free topography is simulated with the DS algorithm. The distance to the centerline and the distance to the edges variables are used as auxiliary variables. They are always exhaustively informed and guide the DS algorithm to localize the different kind of structures of the variable of interest — the topography — in the realizations. The TI contains all DEMs for the available time steps (Fig. 3.2). As for each state, the topography is a two-dimensional data set; it results in a three-dimensional TI, the third dimension representing the time steps. Figure 3.6 illustrates a 2D slice of the TI and its auxiliary variables at time step 'March 1999'.

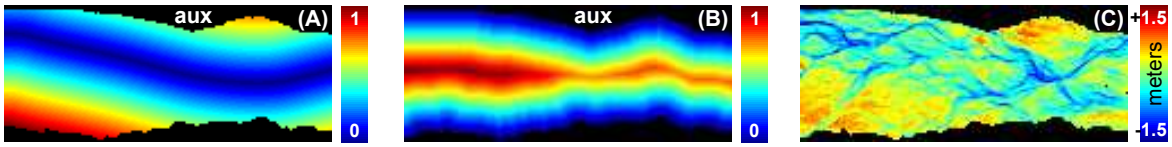

Figure 3.6: Auxiliary variables (A,B) and variable of interest (C) for the TI; (A) distance to centerline, (B) distance to edges, (C) topography.

The auxiliary variables — distances to the centerline and to the edges — are exhaustive for the simulations as shown in Fig. 3.7A and 3.7B. Four possible realizations of DEMs with the same values for the exhaustive auxiliary variables (Fig. 3.7A and 3.7B) are displayed in Fig. 3.7C, 3.7D, 3.7E, and 3.7F. The parameters employed for the simulations with the DS algorithm are listed in Table 3.1.

Table 3.1: DS algorithm parameters for initial topography simulation

Parameters	Value
Homothety	None
Rotation	None
Search neighbourhood	$30\,pixels \times 15\,pixels$
Maximum number of neighbouring nodes	20 for auxiliary variables 30 for the simulated variable
Maximum density of neighbouring nodes	0.1 for auxiliary variables 1.0 for the simulated variable
Type of acceptance distance	L_1
Distance acceptance threshold	0.2 for the distance to centerline variable 0.15 for the distance to edge variable 0.05 for the simulated variable
Maximum scanning fraction for each TI	0.7
Post-processing	None

As one can see, the different topography structures are well reproduced and well placed according to the edges and to the main active channel. Moreover, the exhaustive auxiliary variables do not constraint too much the simulation: with the same auxiliary variables, every realization is unique. So, the main active channel centerline and edges appear to be sufficient to simulate realistic DEMs. A quantitative criteria to test the quality of the DEM simulations is proposed in section 3.5.

3.4.3 Simulation of a topography conditional to a previous topography

The DS algorithm enables realistic simulations of DEMs using exhaustive auxiliary variables based on given centerline and edges. As a consequence, the same auxiliary variables — distance to the centerline and distance to the edges — shall be kept to simulate successive DEMs. In order to simulate a DEM at time step $n+1$ with respect to the previous DEM at time step n, the auxiliary variables are informed at two different time steps, n and $n+1$; and an additional auxiliary variable providing the elevation in meters at time step n is considered.

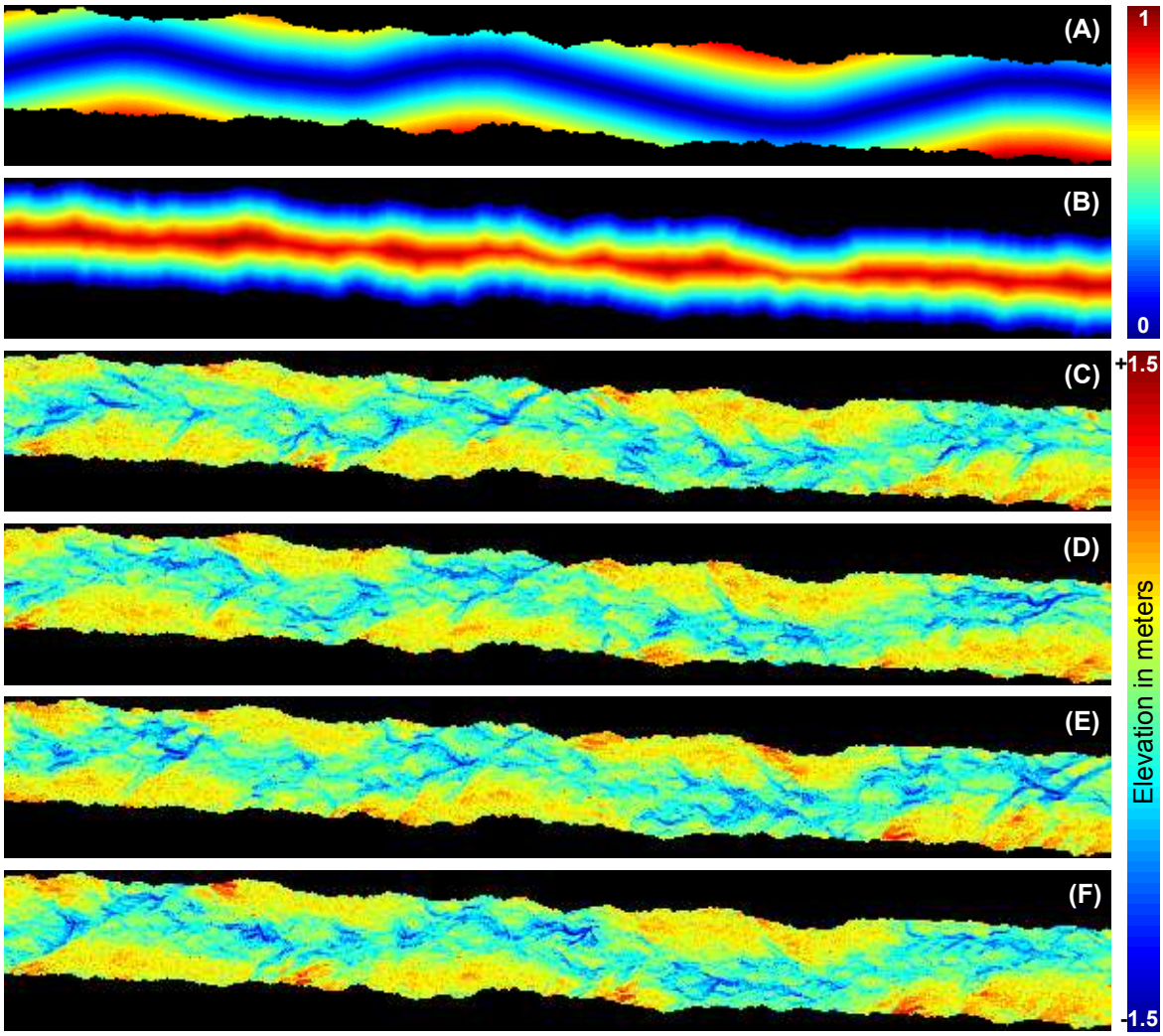

Figure 3.7: Conditioning data: (A) distance to the centerline, (B) distance to the edges, and four DEMs' simulations (C), (D), (E), and (F) with a three variables TI.

This results in the use of five exhaustive auxiliary variables along the variable of interest for the simulations, taking into account the evolution of the centerline and the evolution of the edges in addition to the evolution of the elevation. As illustrated in Fig. 3.8, the simulated variable (elevation at time step $n+1$) in the training data set (Fig. 3.8) is the February 2000 DEM (C2). The five exhaustive auxiliary variables are the February 1999 DEM (C1) and the distance to the edges (B1 and B2) and the distance to the centerline (A1 and A2) both at time steps February 1999 and February 2000.

The parameters employed for the simulations with DeeSse are listed in table 3.2.

Some examples of successive topography simulations are presented in Fig. 3.9(B,C,D,E). The first DEM (Fig. 3.9A) is obtained with the simulation method described in section 3.4.2. The following simulations allow us to model the evolution of the channels and of the side bars, grouping, diverging, or moving through time. Compared to the TI, the results are visually satisfactory. A statistical test of the successive DEMs simulations is proposed in section 3.5.

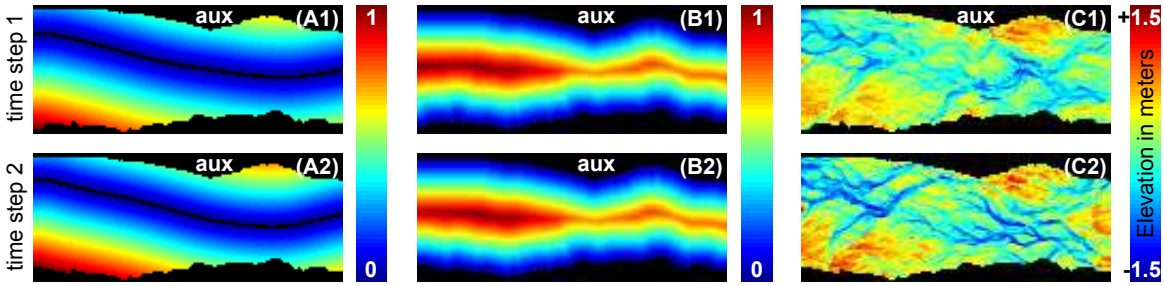

Figure 3.8: TI and its five auxiliary variables to simulate successive topographies: distance to centerline A1 at time step 1 and A2 at time step 2 ; distance to edges B1 at time step 1 and B2 at time step 2; elevation in meters C1 at time step 1 and C2 at time step 2.

Table 3.2: DS algorithm parameters for successive topography simulations

Parameters	Value
Homothety	None
Rotation	None
Search neighbourhood	$30\,pixels \times 15\,pixels$
Maximum number of neighbouring nodes	20 for distance variables
	30 for elevation variables
Maximum density of neighbouring nodes	0.1 for auxiliary variables
	1.0 for the simulated variable
Type of acceptance distance	L_1
Distance acceptance threshold	0.2 for distance auxiliary variables
	0.15 for the elevation auxiliary variable
	0.05 for the simulated variable
Maximum scanning fraction for each TI	0.7
Post-processing	None

3.5 Statistical validation of topography simulations

This section is focused on the variable of interest: the elevation in meters, as the auxiliary variables are only used for guiding the simulation of successive DEMs. To test the quality of the simulated DEMs, they are compared to the TIs (Fig. 3.2) through (i) the empirical cumulative distribution functions (ECDF) regardless of the location (one-point statistics), and (ii) the gamma connectivity function (Renard et al., 2013) for testing the spatial organization of the structures.

Two-point statistics like variograms or multiple-moint statistics such as spatial cumulants (Dimitrakopoulos et al., 2010) or statistics on patterns could also be used. However two-point statistics are not sufficient to characterize complex structures (Guardiano and Srivastava, 1993; Gómez-Hernández and Wen, 1998). Multiple-point statistics are limited by the dimensions and the geometries of the retained neighbourhood as well as by the necessity of discretizing continuous variables into categorical variables if the assessment is based on MPS histograms (Boisvert et al., 2010). The main reason why we use the gamma connectivity function is that the DEM simulations are intended to be used for solute transport and groundwater flow simulations. In this case, the reproduction of the connectivity patterns is crucial (Renard, 2007). Moreover, the connectivity function provides a global measure for the entire domain, and it can be computed over a continuous field such as the elevation in meters for instance.

The validation is performed for the two following cases. The initial topography simulations presented in section 3.4.2 will be referred to as case *(A)*. In this case, 100 realizations obtained with the same fully informed auxiliary variables are considered; the centerline and the edges are fixed for all the realizations. The successive topography simulations presented in section 3.4.3 will be referred to as case *(B)*. In that case, 100 realizations represent 100

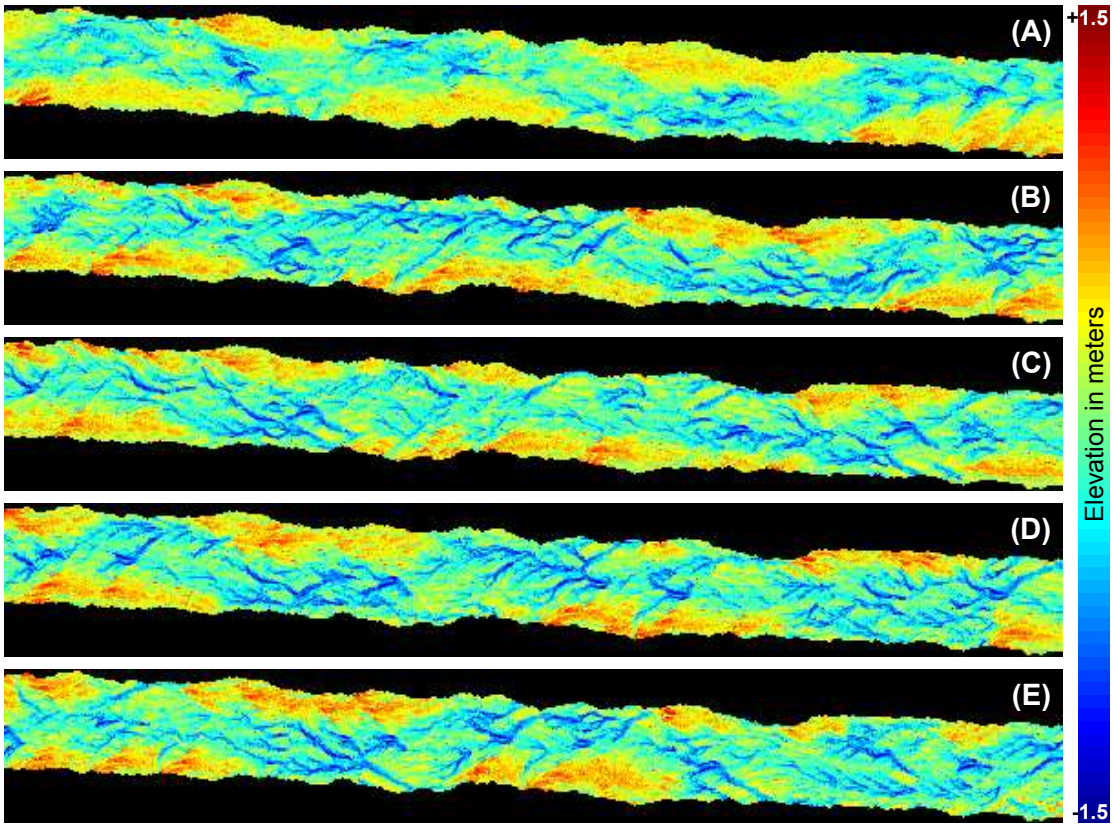

Figure 3.9: Realization of five successive topographies: (A) initial DEM at time step 1; (B), (C), (D), and (E) successive DEMs at time steps 2, 3, 4, and 5, respectively.

successive time steps, and therefore have different fully informed auxiliary variables as the centerline evolves at each time step.

3.5.1 Empirical cumulative distribution function (ECDF)

To assess the quality of simulations from a one-point statistics point of view, we compare the elevation variable ECDF of the TIs in black with those of the realizations in grey in Fig. 3.10.

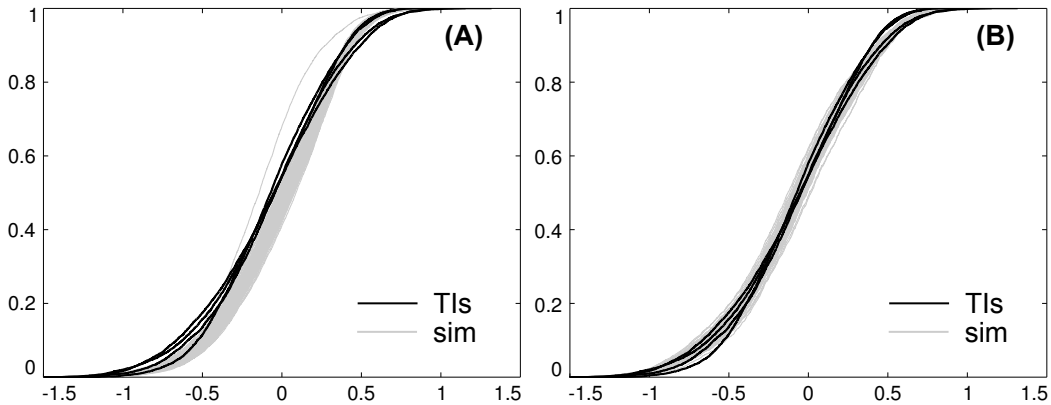

Figure 3.10: ECDF of the elevation variable; comparison between the TIs (black) and the simulations (grey), for cases *(A)*and *(B)*.

The first thing to notice is that in the TIs, the ECDFs are close but not identical for the four available time steps. This differences can be explained because the centerlines are slightly different, which induces different distributions of the nonstationarities. Another reason is that these time steps are part of some cycles in the evolution of the braided river: the topography is sligthly but constantly eroded between two floods, but each flood event is a way of resetting the topography distribution (Lane et al., 2003).

Seemingly the ECDFs are better reproduced in case *(B)* than in case *(A)*. The main reason for it, again, is a difference in the localization of the nonstationarities. As in case *(A)*, the centerline is the same for the 100 realizations, it does not offer as much exploration possibilities in the TIs as in case *(B)* where the centerlines are different.

3.5.2 Gamma connectivity indicators

As previously discussed, the gamma connectivity function defined by Renard et al. (2013) is used to check the quality of the pattern reproduction. Given an indicator variable I, its gamma-connectivity measure, Γ_I, is defined as the probability that two points belonging to the medium $I = 1$ are connected. In our situation, this connectivity measure is computed on indicator variables obtained by thresholding the simulated altitude field at several levels. Each threshold τ defines two complementary zones: a lower level (zone 1) below the threshold and an upper level (zone 2) above the threshold. The gamma connectivity, $\Gamma_{\tau,z}$, is then defined for each threshold level τ and zone $z \in \{1, 2\}$ as the proportion of connected pairs of points belonging to z regarding the number of pairs of points in z. Results are displayed in Fig. 3.11.

For both cases *(A)* and *(B)* as well as for both zones, the simulations have a gamma connectivity function quite close to the reference gamma connectivity functions of the TIs: the jumps in the connection proportions take place at the same range of thresholded values and have the same amplitude. The quality of the realizations in terms of connectivity may then be considered as fulfilled.

3.6 Discussion and conclusion

The models presented above reproduce the spatial statistics of braided river elevation at successive time steps. The proposed method allows to generate models showing a realistic evolution of the bars and channels, while it is not based on direct physical processes. We have shown that with a relatively small training data set, the DS algorithm successfully simulates the successive time-related DEMs, using exhaustive auxiliary variables that help localize structural nonstationarities. Variable transformation allows reducing the complexity of the problem and gives satisfactory results within really short computing times, especially when transforming a two-dimensional variable such as a centerline or edges into a one-dimensional variable.

The statistical validation of the models is based on the comparison of connectivity functions as it is of main concern for flow and transport modeling. A comparison of computing requirements and similar or other statistics with cellular automata and event-based models could also be performed for further performance testing and validation.

The successive DEMs are simulated under the assumption that the simulated edges and the simulated centerlines are valid themselves. The initial centerline simulation can be considered valid as it is simulated with a TI derived from braided river centerline interpretation. The way of constructing the edges might seem artificial, but in real case studies, edges can be easily delimited through aerial photography. Furthermore, their evolution is usually much slower than the topography or the centerlines. Similarly, the way centerline perturbations

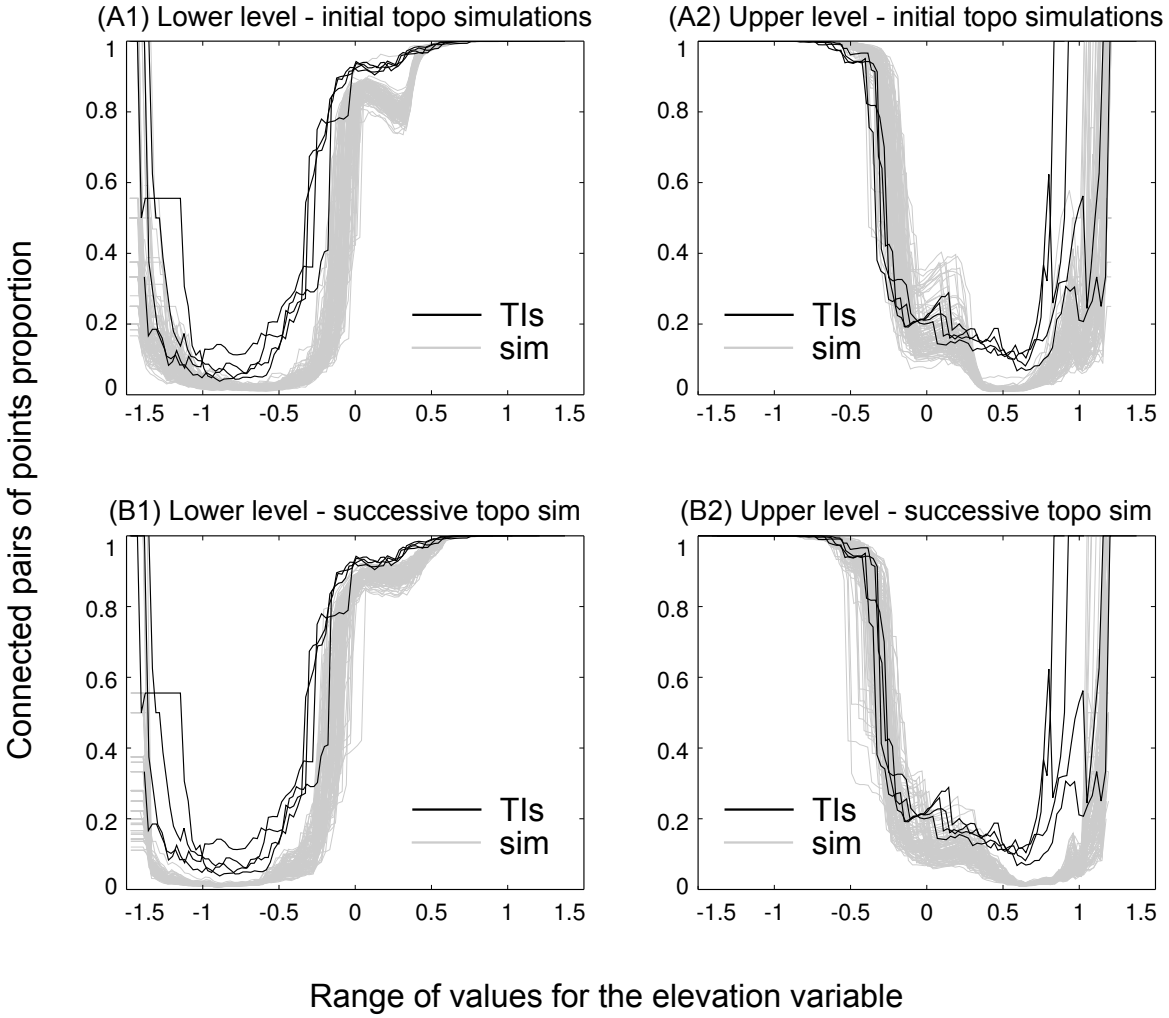

Figure 3.11: Comparison of the gamma connectivities, function of the inundation threshold, betweens the TIs (black) and the simulations (grey) for cases (A) and (B), lower (1) and upper (2) levels.

are created seems artificial too. Here, as no additional information was available from the field, random perturbations were used to simulate their evolution. Acquiring more field data about centerline evolution over decades would be possible through low resolution aerial photography time series, which are not as expensive as high resolution LIDAR data acquisition. Having these data, it would be straightforward to model their evolution using the same statistical principle used here for the topography.

Finally, the simulation of DEM time series opens a wide range of perspectives in the fields of water-related risk management and braided river aquifer modeling. Indeed such aquifers are built by successive erosion and deposition processes affecting geological records and surface morphology. Stacking up successive topographies could enable us to mimic the erosion and deposition steps and allow simulating the main internal structures of a braided river aquifer. Of course it would require some scale adjustment as the scale characteristics of the site to model might differ from the scale characteristics of the successive DEMs used as TIs for the simulations. This work is going to be pursued in that direction.

Acknowledgments

The work presented in this paper is part of the ENSEMBLE project, financially supported by the Swiss National Science Foundation under the contract CRSI22 1222491. The authors would like to thank Stuart Lane for having provided the Waimakariri River data set, Gregoire Mariethoz for his constructive and motivating discussions, Michael Pyrcz for his positive review and advice, the reviewers and the editor for their will of clarifying and publishing the manuscript.

Chapter 4

A pseudo genetic model of coarse braided river deposits

Abstract

A new method is proposed to produce facies models of braided river aquifers, from analog data. It is based on stacking Multiple-Point Statistics simulations of successive topographies to reproduce the main inner structures of the aquifer. The smaller scale structures are generated by mimicking the trough filling process occurring in braided rivers. It produces realistic cross-stratified sediments. The three main input parameters of the algorithm offer a control over the proportions, the continuity and the dimensions of the deposits.

4.1 Introduction

In alpine regions such as Switzerland, water for drinking, irrigation or industrial purposes is frequently tapped from gravel braided river aquifers (FOEN, 2009). To understand, manage and protect the groundwater resource in this type of systems, or to better model aquifer-river interaction at multiple scales, models of their internal geological heterogeneity are required. These models allow for example to assess the uncertainty of a pollution plume trajectory towards a pumping well. In a previous paper (Pirot et al., 2014), we proposed to model the evolution of braided river topographies using LIDAR data from analog sites and multiple-point statistics. In this paper, we pursue this research and consider the internal geological architecture of the braided system.

To constrain our model, we rely on a vast literature that describes the geology of braided systems. It includes numerous sedimentological descriptions (Williams and Rust, 1969; Miall, 1977; Huggenberger and Regli, 2006), geometrical or topological characterizations (Leopold et al., 1957; Howard et al., 1970; Sapozhnikov and Foufoula-Georgiou, 1996; Lane, 2009) conducted on specific sites. These observations as well as the classification of braided river components (Allen, 1983; Miall, 1985; Labourdette and Jones, 2007) provided an essential nomenclature for the description and analysis of such systems. Length scale characterization studies (Foufoula-Georgiou and Sapozhnikov, 2001; Hundey and Ashmore, 2009), detailed outcrop analysis (Klingbeil et al., 1999; Labourdette and Jones, 2007; Bayer et al., 2011) and ground penetrating radar measurements and interpretations (Bridge et al., 1995; Lunt and Bridge, 2004; Huber and Huggenberger, 2015) reinforced the knowledge of sedimentary structure and heterogeneity in braided river aquifers.

The understanding of the dynamic processes occurring in braided rivers was enforced by on site process description (Rust, 1972; Ashworth et al., 1992; Jones and Schumm, 2009), planform and morphological evolution analyses (Ashworth, 1996; Lane et al., 2003; Brasington et al., 2012), granulometry and bed-load transport studies (Ashmore, 1988; Dawson, 1988; Surian, 2002) and completed by many flume experiments (Ashmore, 1982; Kleinhans and Brinke, 2001; Van De Lageweg et al., 2013). Derived from physical laws, flow and sediment transport models have been developed (Fredsøe, 1978; Ashworth et al., 1994; De Serres et al., 1999; Dargahi, 2004; Millar, 2005; Davy and Lague, 2009). These physically based models, as well as cellular automata models (Murray and Paola, 1994; Coulthard et al., 2002; Thomas and Nicholas, 2002), take first into account the transport of sediments on the surface topography and eventually some other processes such as avulsion (Jerolmack and Paola, 2007) or the interaction with the vegetation (Edwards et al., 1999; Murray and Paola, 2003; Thomas et al., 2007). They allow analyzing the evolution of the surface morphology of braided rivers but they do not provide the resulting internal structure and heterogeneity of the underground.

All this conceptual and process-based knowledge allowed researchers to build various kinds of structure imitating models of the internal heterogeneity of braided river aquifers, with the aim to better understand the control of heterogeneity on solute or contaminant

transport in such aquifers (Boggs et al., 1993; Rodgers et al., 2004). The simplest models are based on traditional multi-gaussian geostatistics (Felletti et al., 2006; Salamon et al., 2007; Glenz, 2013) or indicator simulations (Klise et al., 2009). More realistic three dimensional models of geological heterogeneity can be generated by pseudo-genetic, object-based and some recent geostatistical methods. Webb (1994) was one of the first to propose a process imitating method to produce models of geological heterogeneity for braided river aquifers. His approach is based on the vertical stacking of successive topographies of the braided river simulated by random walk (Webb, 1995). One limitation of this method is that aside from channel width and depth, all surfaces between channels are considered as flat. The method has been used by Anderson et al. (1999) to model hydrogeological properties of a braided stream deposit and has been tested through a groundwater flow and transport problem. Teles et al. (2001) proposed an agent based model allowing to describe a fluvial system at a rather large scale. The processes of erosion and deposition are simulated using a set of simple rules and a multi-agent system. Using this framework, and providing paleo information about the flow and sediment load in the fluvial plain, over time scales of some thousands of years, the model allows reconstructing possible heterogeneity structures in 3D. Recently, stochastic object based models have been developed for braided rivers (Ramanathan et al., 2010; Huber and Huggenberger, 2015). These algorithms are computationally efficient when no conditioning to field data is required. Comunian et al. (2011) proposed an innovative way – based on multiple point statistics (Straubhaar et al., 2011) – to model a three dimensional braided river deposit from sedimentological observations on a series of seven cross-sections collected at the Herten site, Germany, by Bayer et al. (2011). This method proved the ability of MPS to reproduce complex fine scale geological structures, but the available data in this case allowed only to model a rather small area ($16\,m \times 10\,m \times 7\,m$).

In this paper, a new pseudo genetic algorithm is proposed. The method revisits the principles established by Webb (1994). It is based on the stacking of successive topography simulations to create geological units, within which geological facies or geophysical properties can be assigned. The novelties rely on the way to produce the successive topography simulations and on the way to assign the geological facies. Realistic successive topographies are simulated with the method developed by Pirot et al. (2014), which is based on the Direct Sampling multiple-point statistics (MPS) algorithm (Mariethoz et al., 2010). It provides not only a realistic topography of the channels, but also of bars and islands. Furthermore, the temporal evolution of the topography is modeled as well, using multivariate MPS. Then the assignment of the geological facies is not based on an estimation of the Froude number as done by Webb (1994), but on a deformation process that mimic the deposit of a repeated facies sequence, producing cross-bedded deposits. The approach is illustrated with a training data-set composed of successive DEMs of the Waimakariri River, New Zealand (Lane et al., 2003) acquired with LIDAR at four time steps.

The paper is organized in four parts. An overview of the pseudo genetic algorithm and its global parameters is given in section 4.2. The method to produce successive MPS simulations of topographies conditionally to the previous one and the way of stacking the successive topographies to build a geological layer are detailed in section 4.3. Then, the generation of facies heterogeneity within the geological layers is presented in section 4.4. Finally, the resulting models are presented as well as a sensitivity analysis of the algorithm to its main input parameters in section 4.5.

4.2 Algorithm and main parameters

The general algorithm proposed in this paper is outlined in Alg. 1 and its general principles are described thereafter. The initialization consists in producing an initial topography that

will constitute the bottom of the aquifer, using the Direct Sampling MPS algorithm of Mariethoz et al. (2010). Then successive iterations follow until the aquifer model reaches the desired thickness. Knowing a given topography, the next one is simulated by MPS conditionally to the previous one, as proposed by Pirot et al. (2014). The next topography is then stacked over the previous one with an aggradation rate fixed for all iterations (see Table. 4.1). The new topography erodes partly the underlying geological layers and deposits sediments that constitute a new geological layer. Facies heterogeneity can then be generated within the new geological layer. If the resulting aquifer model does not reach the desired thickness, the algorithm's loop continues, else it stops.

Algorithm 1 pseudo genetic algorithm for three-dimensional facies heterogeneity models of braided river aquifers

procedure BRAHMS(*parameter file*)
  simulate initial (bottom) *topography*
  initialize *current model thickness* $\leftarrow 0$
  while *current model thickness* $<$ *desired thickness* **do**
    simulate *next topography* conditionally to the previous *topography*
    stack the *next topography* over the previous *topography* to build a *geological layer*
    subtract the induced erosion to the underlying layers
    generate geological heterogeneity within the new *geological layer*
    compute the *current model thickness*
    set the previous *topography* $\leftarrow$ *next topography*
  end while
  return *3D heterogeneous facies model*
end procedure

This approach assumes that aggradation takes place during large flood events that remodel the topography of the river. On a field site, it is not possible to assess what part of the aquifer deposit has been mobilized and transported or remodeled during a flood. So, at a specific location, an observed aggradation does not mean that there is no erosion below the previous topography. However, in order to produce the main geological structures as objects and not large continuous layers, the method assumes that successive deposits have structures largely influenced by the observed evolution of the surface topography of braided rivers. In other words, the erosion surfaces defining facies interfaces in the recorded geology are supposed to be at least similar to the surface topography evolution.

Field observations on actual active braided river systems often do not show significant aggradations. Some detailed studies (Lane et al., 2003) show only some cyclic evolution of braided rivers geomorphology. Though gauging stations provide informations about flood frequency and magnitude, they do not allow to link these clearly with the aggradation rate (Sambrook Smith et al., 2010). In addition, braided river system outcrops do not show significant changes or trends in the dimensional characteristics of the deposits at one location, as if there were some kind of medium to large scale stationarity of the deposit dimensions. Therefore in absence of more specific information, it has been decided to keep the aggradation rate fixed in the algorithm for the moment. Note however, that a more sophisticated distribution of the aggradation quantities over time can easily be handled by the proposed modeling framework.

More generally, the algorithm uses two types of parameters. The first type, described in Table 4.1, are the parameters defining the dimensions of the model and are required to build the main structures – the geological layers – of the model. The second type, listed in Table 4.2, are used to generate heterogeneity within the geological layers. The parameters have to be adjusted and inferred from field observations. Outcrop analysis, GPR section interpretations, borehole data or analog sites shall provide information about the thickness of the main geological layers and therefore about the aggradation rate, as well as about

encountered structures of the deposits and facies proportions and properties.

4.3 Building the main depositional structures

The main depositional structures or geological layers are obtained sequentially by stacking together successive simulations of digital elevation models (DEMs). In order to produce coherent successive topographies, it is recommended to generate DEMs at the river scale, which might be really larger than the aquifer model scale depending on the case study. The aquifer dimensions and location are referred to as zone of interest. Interpolation parameters (see Table 4.1) allow to re-scale the extracted zone of the simulated topographies to the zone of interest, especially if an analog with other length scale characteristics is used to produce the topographies.

4.3.1 Simulation of successive topographies

The first step consists in producing sequentially successive topographies of a braided river according to the method described in details by Pirot et al. (2014). Here it is applied on a grid of size $11,600\,m \times 1,200\,m$, as indicated in the braided river topography dimension parameters of Table 4.1. To illustrate the method, successive DEMs of the Waimakariri river

Table 4.1: Main structural parameters

	Parameter	Value
zone of interest aquifer model parameters	length (flow direction)	$200\,m$
	width (orthogonal to the flow direction)	$100\,m$
	minimum thickness - depth	$12\,m$
	cell length	$2\,m$
	cell width	$2\,m$
	cell height	$0.1\,m$
braided river topography dimensions	length	$11,600\,m$
	width	$1,200\,m$
	cell length	$20\,m$
	cell width	$20\,m$
interpolation parameters	margin length	$5\,m$
	margin width	$5\,m$
	margin depth	$0\,m$
	scaling factor along length axis	1
	scaling factor along width axis	1
	scaling factor along depth axis	1
facies parameters	coarse grain size sediment facies value	1
	medium grain size sediment facies value	2
	fine grain size sediment facies value	3
aggradation rate α	default value	$0.3\,m/geological\,layer$

at four different time steps (Lane, 2000) are chosen as training data-set, though other analogs of braided river might have been used. At each time step t, the topography is simulated conditionally to the previous topography at time step $t-1$. The other parameters to run the MPS algorithm and produce the topography time series are identical to the ones used by Pirot et al. (2014). The result of four successive topography simulations are displayed in Fig. 4.1.

Then the zone of interest is randomly retrieved so that all pixels within it are in the domain of the large scale topography simulation grid. This extraction steps takes into account the different scaling factors (see Table 4.1) that allow to stretch or shrink the topographies along the model axis. Due to the possible different grid resolutions of the braided river topography simulations (coarser scale) and the aquifer geological model (finer scale), margin parameters are used to extend the zone of interest, which allow for a better interpolation of the topography at the aquifer model scale. The extracted topography is interpolated

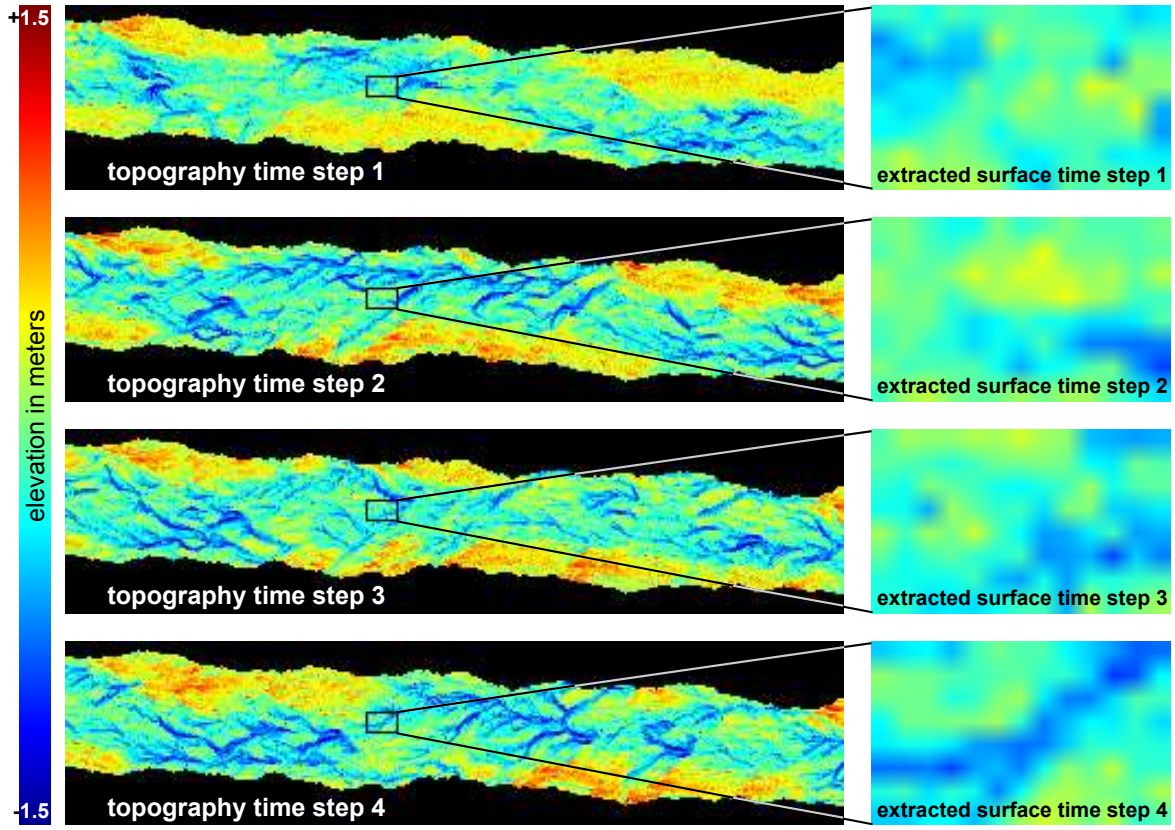

Figure 4.1: Simulations of four successive topographies, followed by extraction and interpolation

linearly on the aquifer model grid. Though smoother interpolations technique like splines could have been used, the resulting models do not suffer from this rough linear interpolation (see figs. 4.5 to 4.7).

4.3.2 Stacking topographies to create erosion and deposit volumes

Now that a series of successive topographies is available, it is possible to stack them with a vertical increment, i.e. the aggradation rate α. As illustrated on a two-dimensional fictive section in Fig. 4.2, this process produces some erosion volumes and some deposition volumes. A positive aggradation rate allows then to build a succession of main geological layers and to reach a minimum thickness or depth for the aquifer.

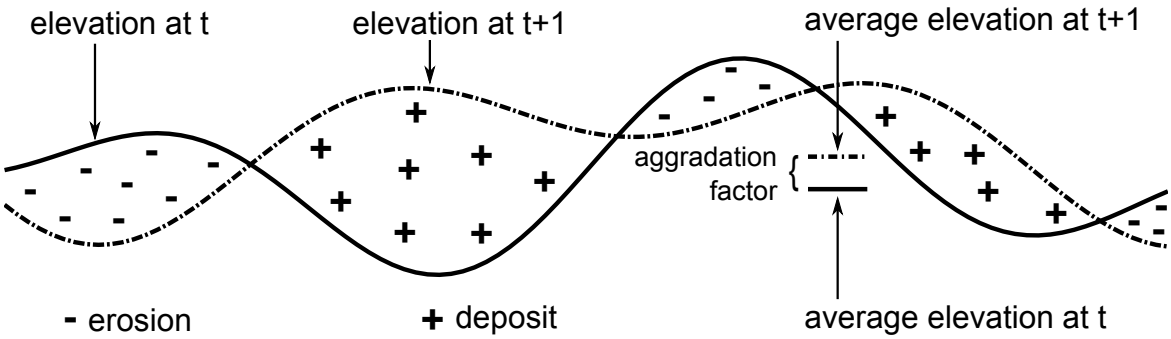

Figure 4.2: Pseudo processes of erosion and deposit

More precisely, let us denote $(E_i)_{i\in\mathbb{N}}$ the ensemble of simulated elevations. The ensemble of centered elevations $(C_i)_{i\in\mathbb{N}}$ are then defined as $C_i = E_i - \mu(E_i), \forall i \in \mathbb{N}$, where $\mu(E_i)$ is the mean of E_i. Let us denote $(Z_i)_{i\in\mathbb{N}}$ the elevation of the layers. The bottom of the aquifer model Z_0 is initialized with C_0. Then, sequentially, for each iteration i, the elevation of the next main layer is computed as a function of the previous elevation and the aggradation rate: $Z_i = Z_{i-1} + \alpha + C_i$. Additionally, the resulting erosions may impact all the previous elevations that need to be updated: $Z_j = min(Z_j, Z_i), 0 \leq j \leq i-1$. The resulting elevations $(Z_i)_{i\in\mathbb{N}}$ are stored as a pillar grid. Once the main geological layers are defined, it is possible to define heterogeneous physical properties or facies within each layer.

4.4 Generate geological heterogeneity within the deposit layers

The method presented hereafter to generate heterogeneous facies within the main geological layers is inspired by field observations of outcrops, processes, and studies linking surface topography with ground penetrating radar (GPR) measurements (Huber and Huggenberger, 2015). During a flood event and looking at the intra-layer scale, sandy gravel sheets are moving forward, filling progressively the scours formed by erosion at channel confluences. When a gravel sheet collapses while moving forward over a scour, a granulometric sorting occurs. By gravity, coarse grain size materials are deposited first, followed by finer grain size sediments. With the combined action of the flow, this process produces cross-stratified deposits that can be observed in gravel-pits (Heinz et al., 2003).

To mimic this process and obtain stratified deposits, the bottom topography of a given layer is iteratively shifted and deformed locally, in accordance with local flow and topography constraints. Each iteration defines a deposition volume for a distinct facies. The sequence of the facies are defined in the parameter Table 4.2. Within the framework of sand and gravel

Table 4.2: Parameters for the flow and gradient deformation scheme to generate fine scale heterogeneity

Parameter	Value
Number of iterations n – default value	6
Facies sequence – default sequence	[3 ; 1 ; 3 ; 1 ; 3 ; 1 ; 2]
flow power f_p – default value	5
smoothing radius r	3

braided river, it is assumed that three main facies types can be retained. The three facies are differentiated according to their granulometry and sorting: facies 1 represents fine grain size and well sorted sediments, facies 2 represents unsorted mixed size sediments, and facies 3 represents coarse grain size and well sorted sediments. Facies 2 can be interpreted as a low permeable medium, representing gravel sheets heterogeneity. Facies 1 and 3 represent the sorted sediments whose iterative successions form cross-stratifications.

To shift and deform the topography, we first evaluate the direction in which the sediments should move. This is given by a rough estimation of the possible river flow direction that will be described in section 4.4.1. Then we also consider the orientation of the local slope since scour filling occurs at the edges of gravel sheets where they can collapse, and are oriented along the local flow direction. The same local flow conditions are used for all iterations within a geological layer. However the downward slope is updated at each iteration. The deformation process is detailed in section 4.4.2.

Furthermore, note that due to the nature of the algorithm, mimicking successive physical erosion and deposit, each geological layer filled with heterogeneous facies can be affected by

the following erosions. This is why the heterogeneity of the previous layers are updated at each time step as mentioned in Alg. 1.

4.4.1 Surface flow approximation

The flow approximation is computed on a larger scale than the zone of interest to avoid boundary effects and to take into account the global flow scale which provide continuity at the local scale. So the flow is approximated on a grid of the same dimensions $(11,600m \times 1,200m)$ as the one used to generate the successive topographies.

First, an against current fast-marching (Cao and Greenhalgh, 1994; Sethian, 1996) is performed on the topography, to compute a travel time T_t minimizing the effort from the downstream outlet of the river topography (right edge on Fig. 4.3) to its upstream inlet (left edge on Fig. 4.3), using a Matlab implementation developed by Peyré and Cohen (2006). The input data is a grid of local front propagation velocity which is set proportional to the opposite of the topography, raised by a constant so that the velocity field contains strictly positive values.

Then the approximated flow field $\widehat{\vec{F}}$ is computed as the opposite of the gradient of the travel time T_t divided by its squared norm (see Eq. 4.1).

$$\widehat{\vec{F}} = -\frac{\vec{grad}(T_t)}{\|\vec{grad}(T_t)\|^2} \tag{4.1}$$

This method produces a flow map following the channel orientations and whose norm is slightly bigger in the channels, as displayed in Fig. 4.3. It is important to note that this flow

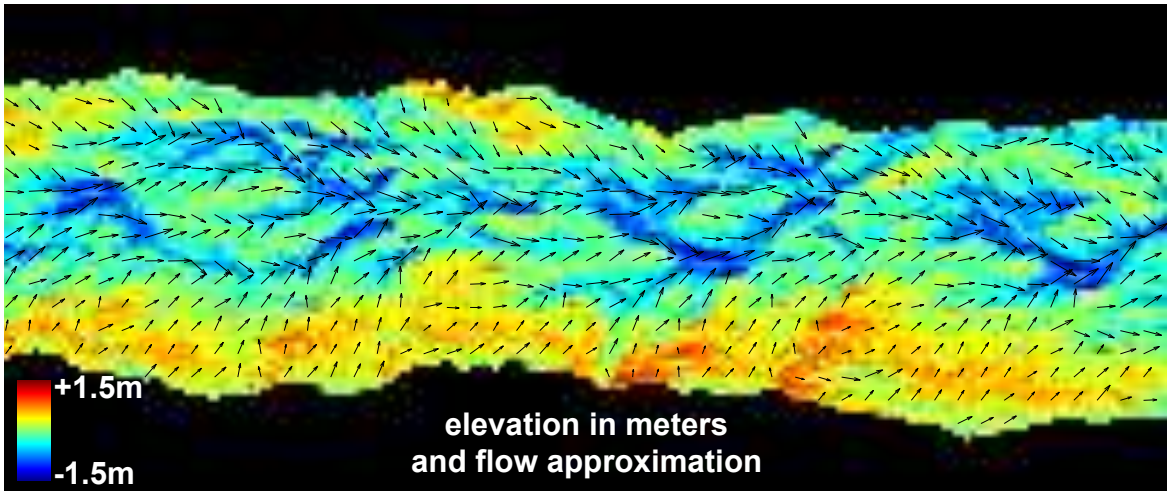

Figure 4.3: Approximation of the global flow over the topography at the river scale

map does not pretend to be accurate. It is used only to get a rough approximation of the flow directions that can be computed very rapidly. More sophisticated calculations could be used as well but this was not deemed necessary at this point of the research.

4.4.2 Iterative deformation scheme

The number of iterations n as well as the facies sequence are defined in the parameter Table 4.2. Within one geological layer, all iterations use the same flow map $\widehat{\vec{F}} = (v_x, v_y)^t$. It is assumed constant within each layer as the deformations are progressive and smooth to produce cross-stratifications. The intensity of the flow might be adjusted with the flow

parameter f_p. At each iteration i, the gradient $\vec{grad}(T_i) = (-u_x, -u_y)^t$ of the current topography T_i is computed. T_0 denotes the bottom of the geological layer. The deformation and shifting $\vec{d_i}$ is computed as a vector of scalar products (see Eq. 4.2) constrained by co-directionality: opposite flow and downward slope direction cannot generate a sorted deposit and therefore cannot shift the topography.

$$\vec{d_i} = f_p \cdot \begin{pmatrix} -min(u_x,0) \cdot min(v_x,0) + max(u_x,0) \cdot max(v_x,0) \\ -min(u_y,0) \cdot min(v_y,0) + max(u_y,0) \cdot max(v_y,0) \end{pmatrix} \tag{4.2}$$

The deformation $\vec{d_i}$ is applied to the coordinates $(X_i, Y_i)^t$ of the current topography T_i. The new coordinates $(X_{i+1}, Y_{i+1})^t = (X_i, Y_i)^t + \vec{d_i}$ are used to compute the next smoothed and shifted topography TS_i (within the current layer) through a nearest neighbor interpolation followed by a moving average whose parameters are defined in Table 4.2: $TS_i = smooothInterpolation(X_{i+1}, Y_{i+1}, T_i)$. The topography T_{i+1} is constrained to be below the top topography T_{n+1} of the current layer, and as no erosion is desired, it is also limited by the value of the previous topography T_i: $T_{i+1} = min(T_{n+1}, max(T_i, TS_i))$. Fig. 4.4 provides an illustration of successive deformations over three iterations from an initial topography and flow map. The facies assigned to the volume comprised between T_{i-1} and T_i is the i^{th} facies

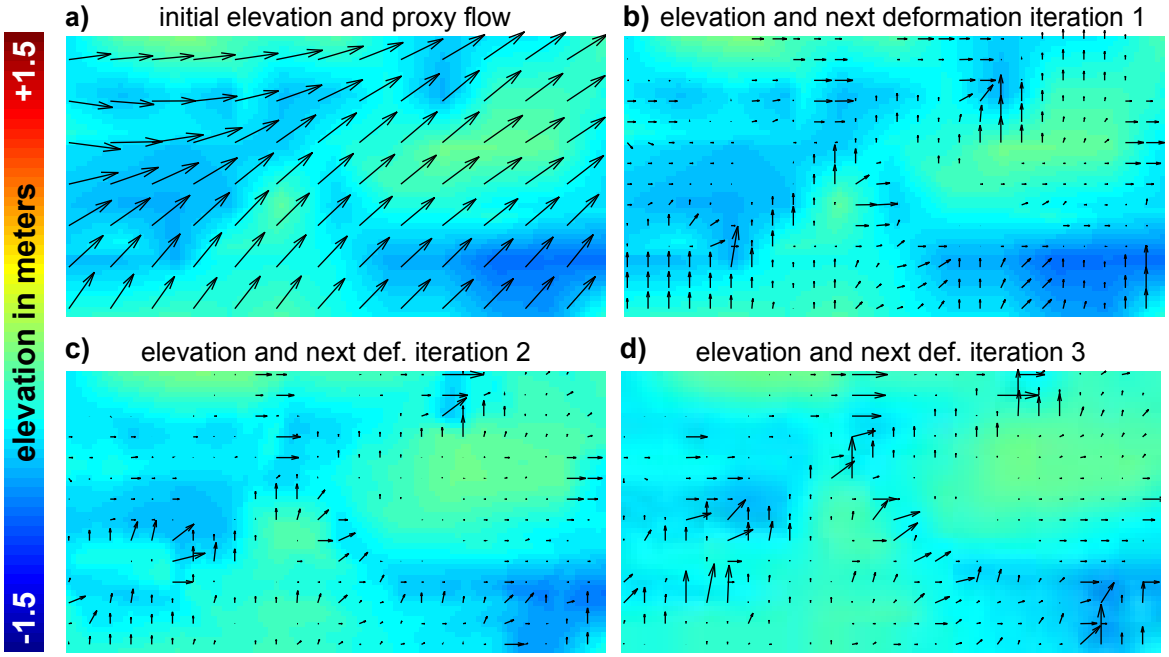

Figure 4.4: Iterative topography deformations at the aquifer scale

of the sequence (see Tab. 4.2). The resting volume in the geological layer after n iterations, located between T_n and T_{n+1} – denoting the roof of the geological layer, is assigned the $(n+1)^{th}$ facies of the sequence.

4.4.3 Resulting heterogeneity

As all topographies within each geological layer are generated over a regular grid, the resulting 3D model can be stored as a regular pillar grid. Each layer of the model provides thickness and facies information at all locations of the regular horizontal grid.

Three examples of the resulting models using different aggradation rate α, flow power f_p and number of iterations n are presented in figs. 4.5 to 4.7. As one can see, cross-stratifications are present in the 3D model. The facies 2, which fills the resting volume in the

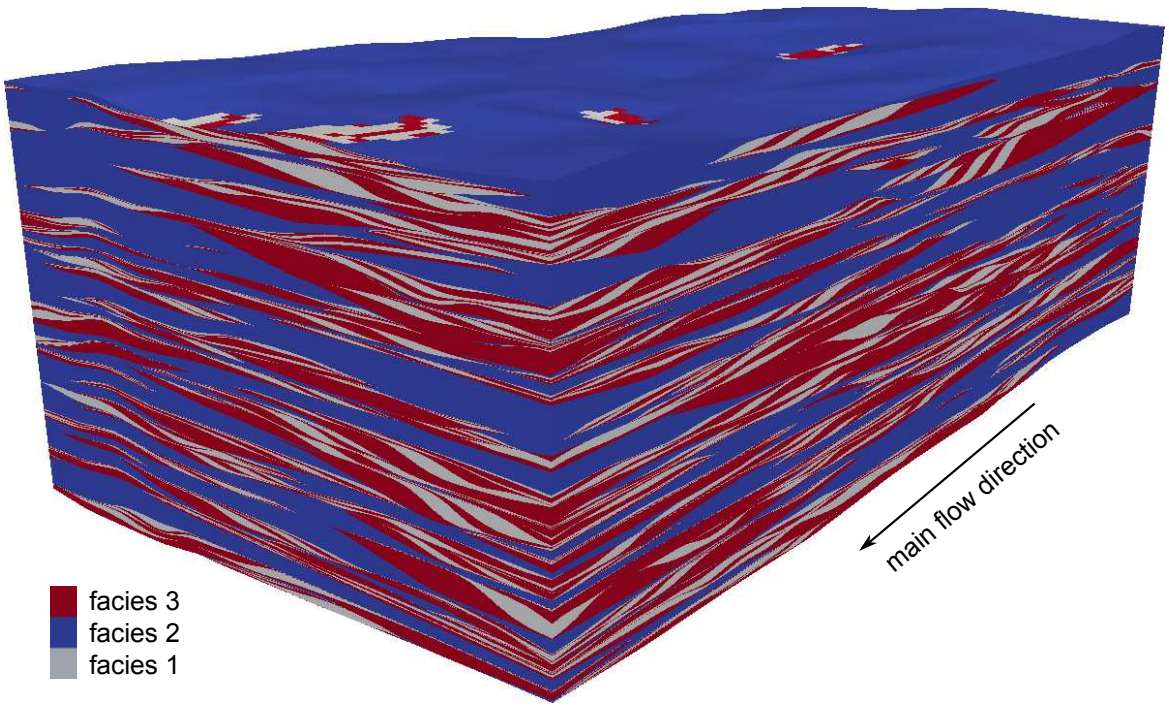

Figure 4.5: 3D heterogenous facies model – example 1 using reference parameters defined in tables 4.1 and 4.2 – obtained with $n = 6$ inner-layer iterative deformations, a flow power $f_p = 5$ and an aggradation $\alpha = 0.3\,m$

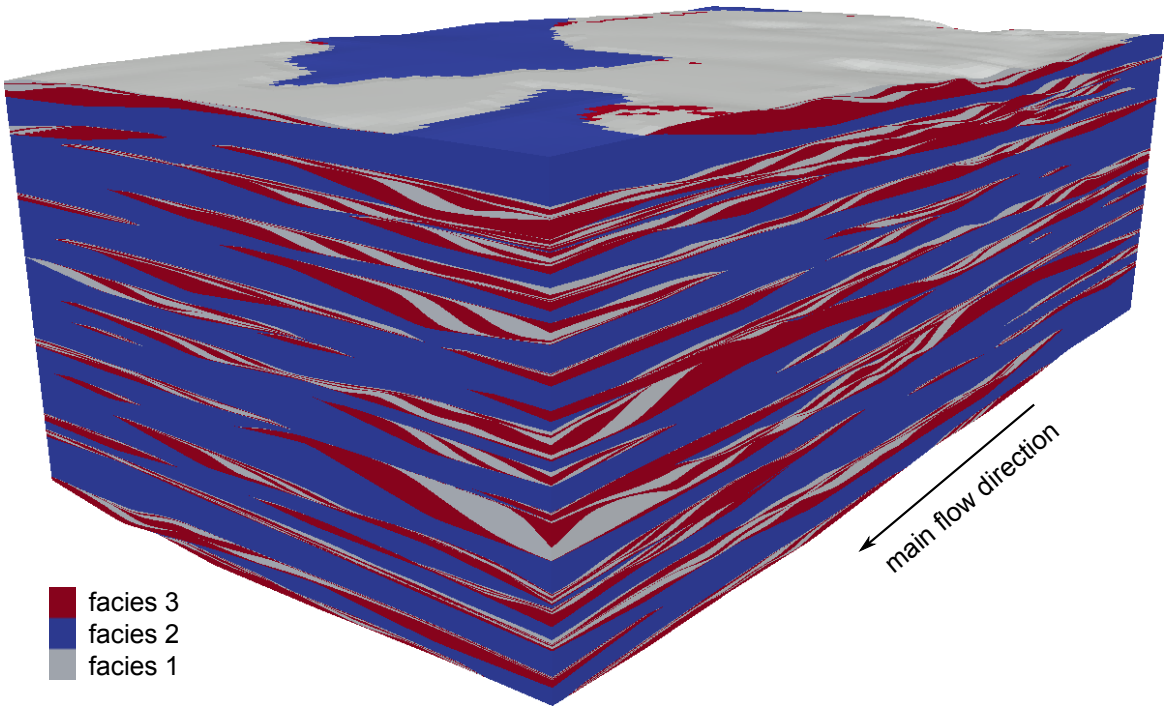

Figure 4.6: 3D heterogenous facies model – example 2 – obtained with $n = 6$ inner-layer iterative deformations, a flow power $f_p = 5$ and an aggradation $\alpha = 0.9\,m$

geological layers after the iterations composing facies 1 and 2 stratified deposits, represents a kind of geological matrix. By modifying the input parameters one can update the geometry

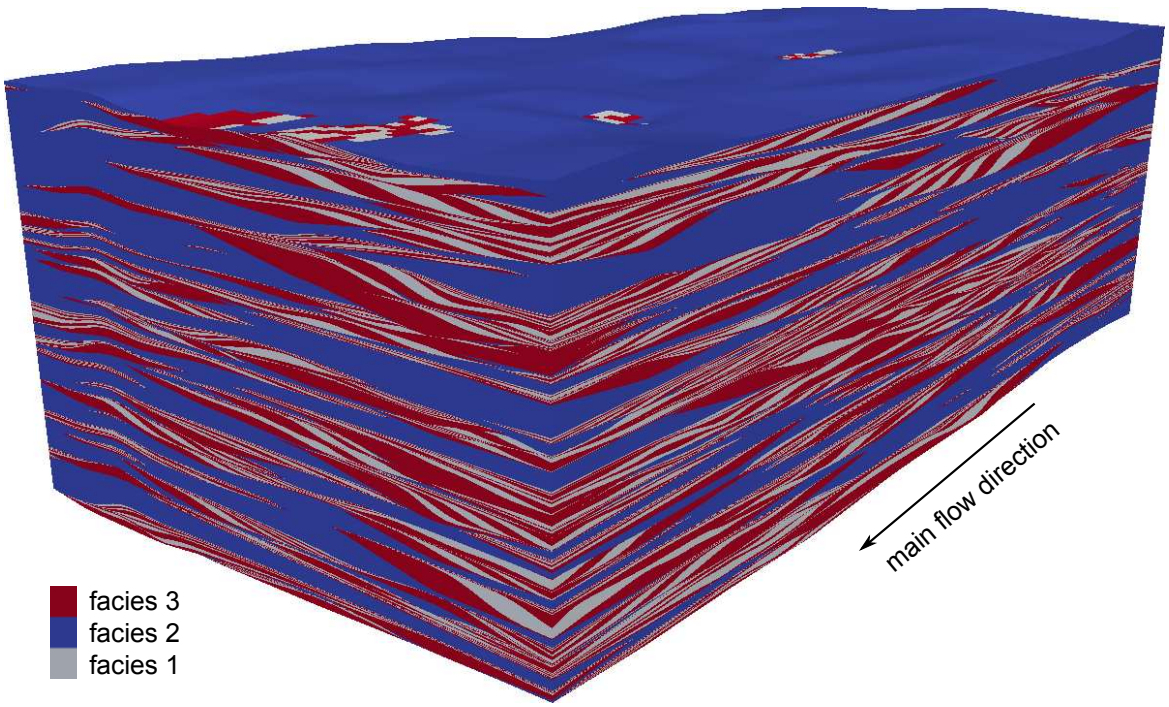

Figure 4.7: 3D heterogenous facies model – example 3 – obtained with $n = 10$ inner-layer iterative deformations, a flow power $f_p = 5$ and an aggradation $\alpha = 0.3\,m$

of the sedimentary structures (dimensions, degree of continuity, etc.) as well as some global properties such as the proportions of the different facies or their connectivity. In the next section, a sensitivity analysis shows how the three main parameters of the algorithm control the properties of the resulting heterogeneity models.

4.5 Parameters and sensitivity analysis

The aim of this section is to give some recommendations on how to calibrate the three main parameters of the algorithm – the aggradation rate α, the flow power f_p and the number of iterations n – regarding the constraints of the user. To achieve this, a sensitivity analysis of these parameters on the resulting facies proportions and facies geobody connectivities is conducted hereafter. Proportions and connectivity are retained as they are important properties for underground flow and transport modeling.

As connectivity indicator we propose to use the gamma connectivity measure (Renard et al., 2013). Given an indicator variable I, its gamma-connectivity measure, Γ_I, is defined as the probability that two points belonging to the medium ($I = 1$) are connected. Each facies allows to compute an indicator variable. It is also possible to consider specific directions as in directional variogram computing. In what follows, we are interested in the connectivity measure along the vertical direction and in the horizontal plane.

4.5.1 Influence of the aggradation rate parameter

The sensitivity analysis is performed in five different configurations. The range of values used for the aggradation parameter α as well as the values of the flow power f_p and the number of iterations n for each scenario are detailed in Table 4.3. The results for the aggradation parameter are illustrated in Fig. 4.8. The black curve shows the evolution of the indicators for the flow power f_p and the number of iterations n of reference as a function

of the aggradation parameter. The dark gray curves show the joint effect of the flow power parameter. The light gray curve illustrates the joint effect of the number of iterations.

Table 4.3: Parameters used to study the influence of the aggradation rate

Parameters	Range of values	
aggradation in meters	[0.05; 0.9]	
Curve name	f_p value	n value
reference	5	6
low flow power	0.2	6
high flow power	9	6
few iterations	5	3
many iterations	5	11

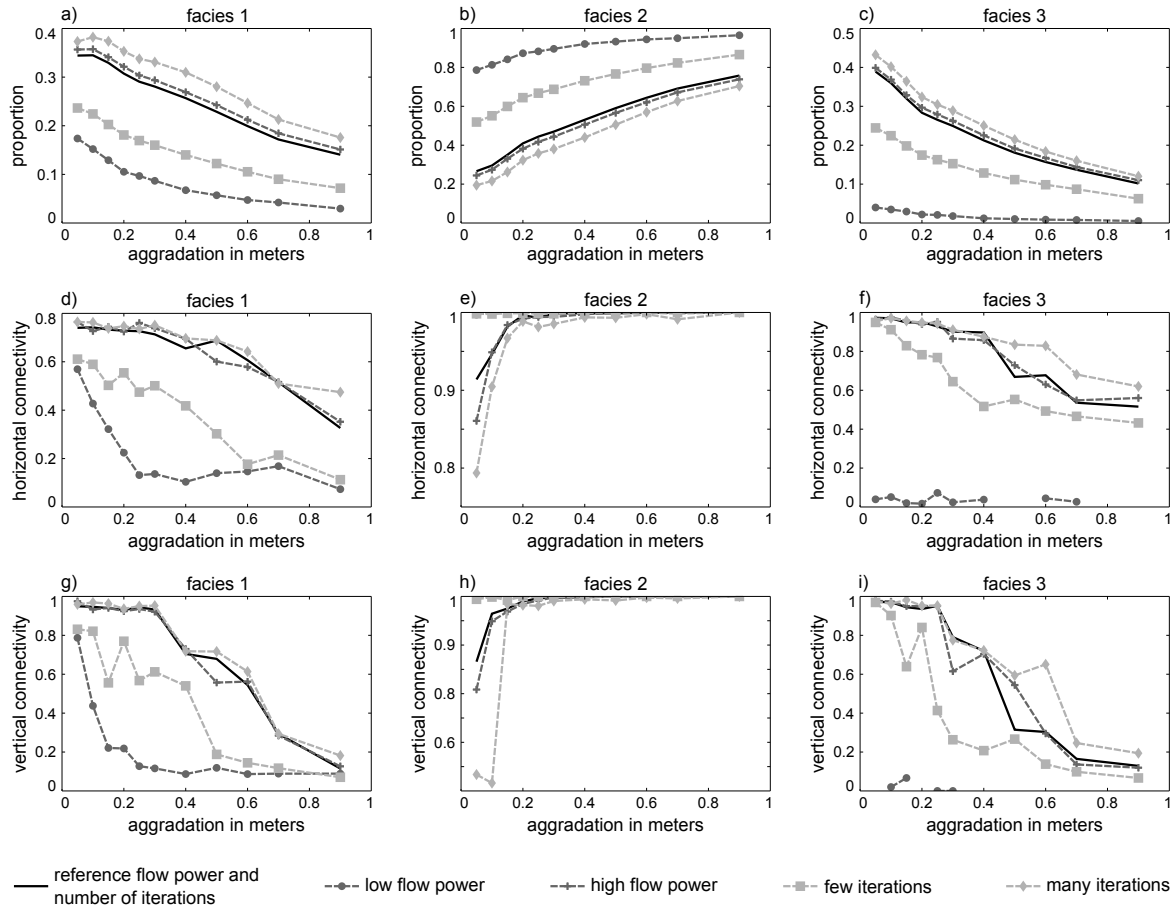

Figure 4.8: Influence of the aggradation on facies proportions and connectivity

The proportions of facies 1 and 3 decrease to the benefit of facies 2 proportion when the aggradation rate increases. It makes sense as an increase of aggradation creates thicker geological layers and thererfore more volume for the matrix facies after the iterative deformations of the bottom layer. Low values for the flow power and number of iterations reinforce this phenomena. The increasing proportion of facies 2 implies an increase of the connectivity indicator (Fig. 4.8e and 4.8h), which reaches a value almost equal to 1 (all pixels are connected) when the aggradation parameter is greater than 0.2. The decrease of connectivity and therefore of the corresponding geobodies length scale characteristics is however progressive for facies 1 and 3 while α increases. Again, low values for the flow power and the number of iterations parameters reinforce this behavior.

4.5.2 Influence of the flow power parameter

In this section, the influence of the flow parameter is analyzed. The range of parameter values are given in Table 4.4. The results are illustrated in Fig. 4.9. Again, the black curve shows the variations of the indicators for the reference values of the aggradation parameter α and the number of iterations n. The dark gray curves show the effect of the aggradation parameter. The light gray curves illustrate the effect of the number of iterations.

Table 4.4: Parameters used to study the influence of flow power parameter

Parameters	Range of values	
flow power	$[0.2; 9]$	
Curve name	α value	n value
reference	0.3	6
low aggradation rate	0.05	6
high aggradation rate	0.9	6
few iterations	0.3	3
many iterations	0.3	11

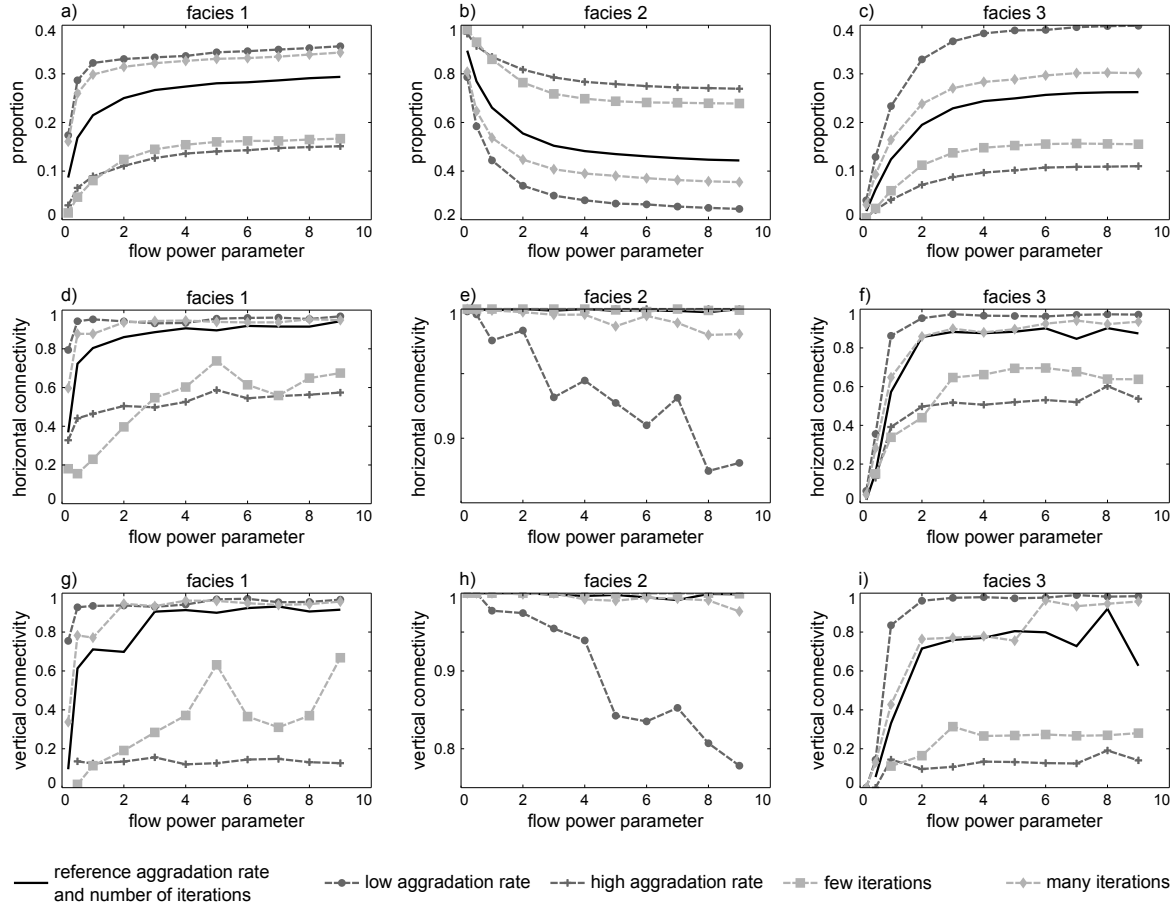

Figure 4.9: Influence of flow power parameter on facies proportions and connectivity

The proportions of facies 1 and 3 increase quite rapidly at the expense of facies 2 when the flow power parameter f_p increases. This is logical since the amplitude of the iterative deformations is proportional to f_p. More deformation implies more cross-beds and more facies 1 and 3. Furthermore, we observe that the proportions reach a sill. It makes sense as once scours or pools are filled in the layer, the resting volume is controlled by the aggradation rate. For facies 1 and 3, the sill of the proportion curves is lower and reached for higher

f_p values with high aggradation rate value or few iterations. It is the opposite for facies 2. The connectivity of facies 1 and 3 increases with f_p. Full connectivity for facies 2 is almost constant ; it decreases slightly with increasing f_p only for a low aggradation rate. The sill and range of the connectivity curves follow the same pattern as the proportion curves regarding high or low values of the aggradation or number of iterations parameters.

4.5.3 Influence of the number of iterations parameter

Here, the impact of the number of iterations parameter is assessed. The range of values used are specified in Table 4.5. The results are illustrated in Fig. 4.10. Once more, the black

Table 4.5: Parameters used to study the influence of the number of shifting iterations

Parameters	Range of values	
number of iterations	[3; 11]	
Curve name	f_p value	α value
reference	5	0.3
low flow power	0.2	0.3
high flow power	9	0.3
low aggradation rate	5	0.05
high aggradation rate	5	0.9

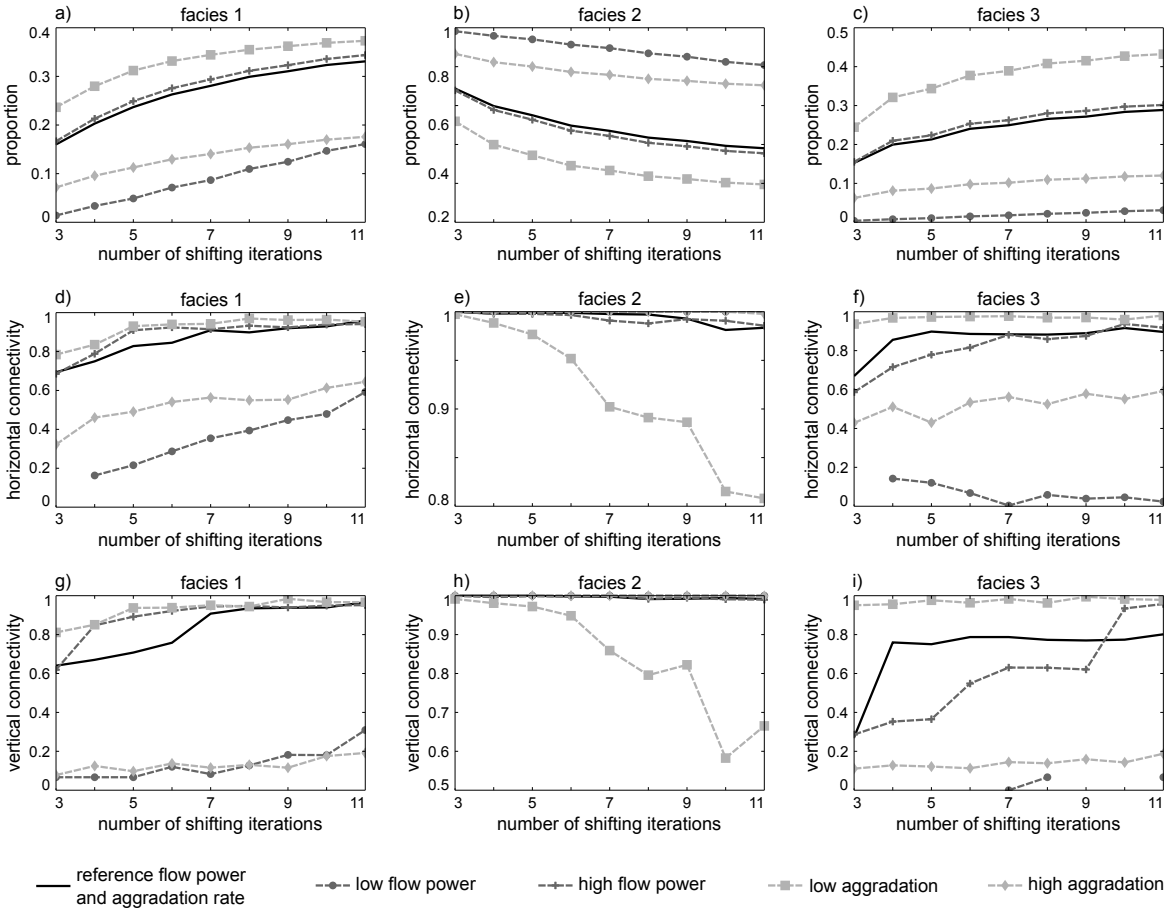

Figure 4.10: Influence of the number of shifting iterations on facies proportions and connectivity

curve shows the evolution of the indicators for reference values of the flow power f_p and the aggradation rate α as a function of the number of iterations n. The dark gray curves shows

the combined effect of the flow power parameter. The Light gray curve illustrates the joint effect of the aggradation parameter.

The proportions of facies 1 and 3 increase slowly to the detriment of facies 2 when the number of shifting iterations n increases. It is not surprising as by construction, the proportions of alternative facies 1 and 3 shall increase as a function of their occurrence in the facies sequence. For facies 1 and 3, the proportion curves are lower with high aggradation rates α or with low flow power f_p values. It is the opposite for facies 2. One could expect similar sills as observed in the indicator plotted as a function of the flow power f_p. However, these sill are not reached with the range of values, because at each iteration the shifting and the deformation of the topography gets smaller and smaller. Indeed, even though the local flow is the same, the topography gradient absolute value decreases at each iteration. The connectivity of facies 1 and 3 increases slowly with n. Full connectivity for facies 2 is almost constant; it decreases slightly with increasing n only for low aggradation rates. The connectivity curves follow the same pattern as the proportion curves regarding high or low values of the aggradation rate α or flow power f_p parameters.

4.5.4 Recommendations

In the proposed algorithm, parameters are required at two levels. In order to facilitate an intuitive understanding of the impact of the parameters on the structures, it is highly recommended to adjust them in two steps. The first parameters to be considered shall be the so called "main structural parameters" (Table 4.1), as they will influence the structural morphology of the deposit. Only after this step should the "geological heterogeneity parameters" (Table 4.2) be adjusted.

Among the limited number of parameters used to build the main geological layers, a few of them require some calibration to either fix them or define their range. Indeed, given a specific morphological context defined by the topography training images, the scaling factors as well as the aggradation rate parameters will strongly influence the size of the resulting elements such as scour width and length for instance. A simple way to help calibrating these parameters is to use interpretations of ground penetrating radar (GPR) data to provide a length scale for these dimensional characteristics. If no GPR data is available, a possibility is to rely on analog data, and define scaling factor to link observable length characteristics of the site to model with length characteristics of the analog.

The geological heterogeneity parameters such as the flow power f_p or the number of shifting iterations n might unfortunately be difficult to calibrate. A suggestion is to use facies proportions, that could be inferred from borehole sample, outcrop analysis, gravel pit analogues or sediment grain size distribution along the river, or a combination of these field observations and analysis. Another possibility is to measure facies thickness from boreholes or outcrop – possibly from analog sites – and compare them with boreholes or section in the simulated depositional model to adjust the parameters.

4.6 Discussion and conclusion

The pseudo genetic algorithm presented here allows to simulate braided river heterogeneous deposits. Three-dimensional representations (see figs. 4.5 to 4.7) of some stochastic realizations obtained with this model reveal realistic sedimentary structures such as cross-stratifications as one can observe on gravel outcrops. The main advantages of the method are the small number (three) of influential parameters (α, f_p, n), and the possibility to use analog data, in particular if no DEM is available for the modeling site. Another advantage is that the approach does not require extensive field investigations as only length scale char-

acteristics are necessary to calibrate the model, and these can be obtained by non-invasive methods such as GPR measurements or accessible outcrops. At last, it is clearly possible to control the connectivity of the different facies by adjusting the input parameters.

One limitation of this approach is however that it might not be straightforward to calibrate the three main parameters to fit characteristic length scales of the modeling site. And though the resulting models seem satisfactory compared to outcrop or field observations, it does not guarantee that compared with other modeling approaches, the resulting depositional models will be a better support for solute and contaminant migration prediction. Therefore, further work shall be conducted to investigate this issue.

Another limitation of the method is that it presently does not allow for data conditioning. Two conditioning aspects shall be considered in the future. The first aspect is to account for large geological structures inferred from GPR section interpretation for instance. It means simultaneously constraining the MPS topography simulations with the aggradation rate and there might be a high degree of uncertainty related to the interpretations. The second aspect is to deal with borehole data. Such conditioning could be achieved by identifying the successive interfaces between the main geological layers and then using these altitudes to constrain simultaneously the MPS topography simulations with the aggradation rate, as while dealing with GPR section interpretations. Another alternative could be to compute a distance function between the properties observed and the properties generated by the model, and then add a post-processing step which would consist in deforming the simulated model to fit the field data while minimizing a deformation cost function.

An even simpler way to use the model described in this paper would be to use it for constructing 3D training images. Then these images can be used in a standard multiple-point statistics (MPS) framework to model the 3D distribution of the facies. This is straightforward and would ensure the conditioning to borehole data directly. The GPR conditioning could then be envisioned in an inversion framework based on summary statistics (Lochbühler et al., 2015).

Acknowledgments

The work presented in this paper is part of the ENSEMBLE project, funded by the Swiss National Science Foundation under the contract CRSI22 1222491. The authors would like to thank Stuart Lane for having provided the Waimakariri River data set, Grégoire Mariethoz and Nicolas Flipo for their constructive and motivating discussions.

Chapter 5

Influence of conceptual model choice on contaminant transport uncertainty forecasting in braided river aquifers

5.1 Introduction

In their review article, Sanchez-Vila et al. (2006) state that "Heterogeneity is the single most salient feature of hydrogeology". Not surprisingly, how to deal with heterogeneity has therefore been the subject of intense research (e.g. De Marsily et al., 2005) and a broad range of models and techniques have been developed to represent heterogeneity in groundwater models. Most of these approaches allow not only building models of the spatial variability but also to quantify the corresponding uncertainty and its influence on flow and transport processes. All those models have different characteristics and can be classified in process-imitating, structure-imitating, or descriptive methods (Koltermann and Gorelick, 1996).

Among them some have been compared through numerical experiments. For example, in the case of braided systems, Teles et al. (2004) compare the flow and transport properties for genesis facies models, indicator simulations and an equivalent uniform medium. dell'Arciprete et al. (2012) show through a comparison between sequential indicator, transition probability and Multiple-Point Statistics (MPS) simulations, that the most realistic geological realizations are obtained when facies details are simulated within previously established structural elements. Zinn and Harvey (2003), using different two dimensional Gaussian based simulations, underline how the mass transfer behavior might be influenced by the conductivity of the most connected components and by the global conductivity variance. Siirila-Woodburn and Maxwell (2015) compare Gaussian, Truncated Gaussian and Facies models through mean arrival time and plume spreading indicators. Lee et al. (2007) compare sequential Gaussian and transition probability based geostatistical simulations and found that pumping test behaviors are strongly influenced by the spatial pattern and connectivity of the facies or hydraulic properties. Zhang et al. (2013) perform a sensitivity analysis of input parameters for a sedimentary depositional facies model, showing that the layer thicknesses distribution controls the non-Fickian tailing of tracer transport at late times. Falivene et al. (2006) compares the resulting connectivity and saturation indicators of two dimensional truncated Gaussian, sequential indicator, MPS and object based simulations. In general, the main objective of all these studies was to compare the flow and transport properties (equivalent conductivity, macrodispersivity, etc.) resulting from various heterogeneity models. An important conclusion is that the choice of a given conceptual heterogeneity or conceptual geological model has in general a larger impact on the transport properties than the variability resulting only from the random variability within a given conceptual model.

In the present paper, we carry out a different type of comparison. We assume that little data is available about a given site (as it is most often the case) and we use prior information from analog sites elsewhere in the world to constrain the heterogeneity models. We then perform a set of predictions and compare the results with a highly detailed data set. The general idea is to test if it is possible to use analog data to make truly relevant uncertainty analysis. In this framework, one important issue is if one conceptual model is more robust and more likely to provide reliable uncertainty estimates than others. Therefore, the 3 main types of models, existing for braided river aquifers, are tested in this situation (a multi-gaussian model, an object based model, and a process-imitating model).

As we assume that limited data is available from the site itself, we do not consider the issue of model conditioning or even inverse conditioning. The comparison is rather difficult since the three types of models have different requirements in terms of input. Nevertheless, this type of exercise is crucial to provide a better understanding of the consequences of the choice of a given modeling approach to quantify uncertainty in cases of data sparsity.

The reference data set comes from the MAcro Dispersion Experiment (MADE) site which is an extremely well studied example of complex geological heterogeneity resulting from braided river deposit. Over the years, a large number of experiments have been carried

out on the site (Zheng et al., 2011), resulting in an impressive and high quality data set. In particular, several experiments resulted in very dense tracer data sets. This data is used here to test the quality of the predictions uncertainty in a context of low level of information.

The paper is structured as follows. First, section 5.2 presents what are the few information used from the MADE site data-set to setup the problem and how analog sites described in the literature may complete these information. Then, the different geological conceptual models are described in section 5.3. The transport model and its specific parameters and boundary conditions are presented in section 5.4. Finally, the results of the plume simulations are presented and analyzed in section 5.5.

5.2 Data and setup

Though the MADE experiments provide a lot of information from conductivity measurements, piezometer levels and concentrations at different locations and time-steps, hydrogeologists in general do not have access to so much data because of budget and time constraints. Therefore, we assume that field data information is limited to the sandy gravel nature of the braided river aquifer and to a few measurements allowing to set the boundary conditions for the transport simulations: piezometric level upstream and downstream of the domain, as well as concentrations of the contaminant at day 27. Here, the injection at day 0 is not simulated to avoid numerical issues due to a strong gradient of the concentrations. This is what we define as our low level of information.

Additionally, to study the impact of the model structures on the uncertainty of contaminant transport prediction, all conductivity fields are normalized to ensure that they have the same mean and the same variance. In order to compare concentration predictions with a reference, petrophysical properties shall be calibrated somehow. This is why the conductivity mean and variance as well as data like characteristic length scales, required to generate realizations of hydraulic property fields with the different geological conceptual models, shall be retrieved from analog data.

5.2.1 The MADE site

The MADE site is located on the Air Force Base of Colombus, Mississippi (see Fig. 5.1), about $2\,km$ South of the Buttahatchee River and about $6\,km$ East of the Tombigbee River. Its surface of about $0.25\,km^2$ is quite flat. It comprises elevations between $64.6\,MSL$ and $66.5 MSL$. The topography presents a slope of 4 to 5‰. The alluvial aquifer is shallow and its mean thickness is about $11m$. The aquifer is constituted of a sandy gravel Pleistocene alluvial terrace associated to the Buttahatchee River over an aquitard composed of Cretaceous marine sediments.

Due to its highly heterogeneous hydraulic properties, it has been actively studied (Zheng et al., 2011) from the mid nineteen eighties, to investigate transport problems in complex porous medium. From 1986 to 2007, three main tracer experiments under natural hydraulic gradient have been conducted at the whole site scale (MADE I to III). It has been followed by two tracer experiments under forced hydraulic gradient at smaller scales (MADE IV and V). More recently, geophysical campaigns including Ground Penetrating Radar (GPR), Direct Push (DP) and Direct Current (DC) resistivity measurements have been conducted to complete the characterization of the site (Bowling et al., 2005; Dogan et al., 2011; Bohling et al., 2012).

In this paper, we choose to focus on the MADE II Tritium contaminant experiment because it offers a high quality data set with a perfect conservative tracer allowing to investigate the effect of heterogeneity on uncertainty without having to consider additional process such

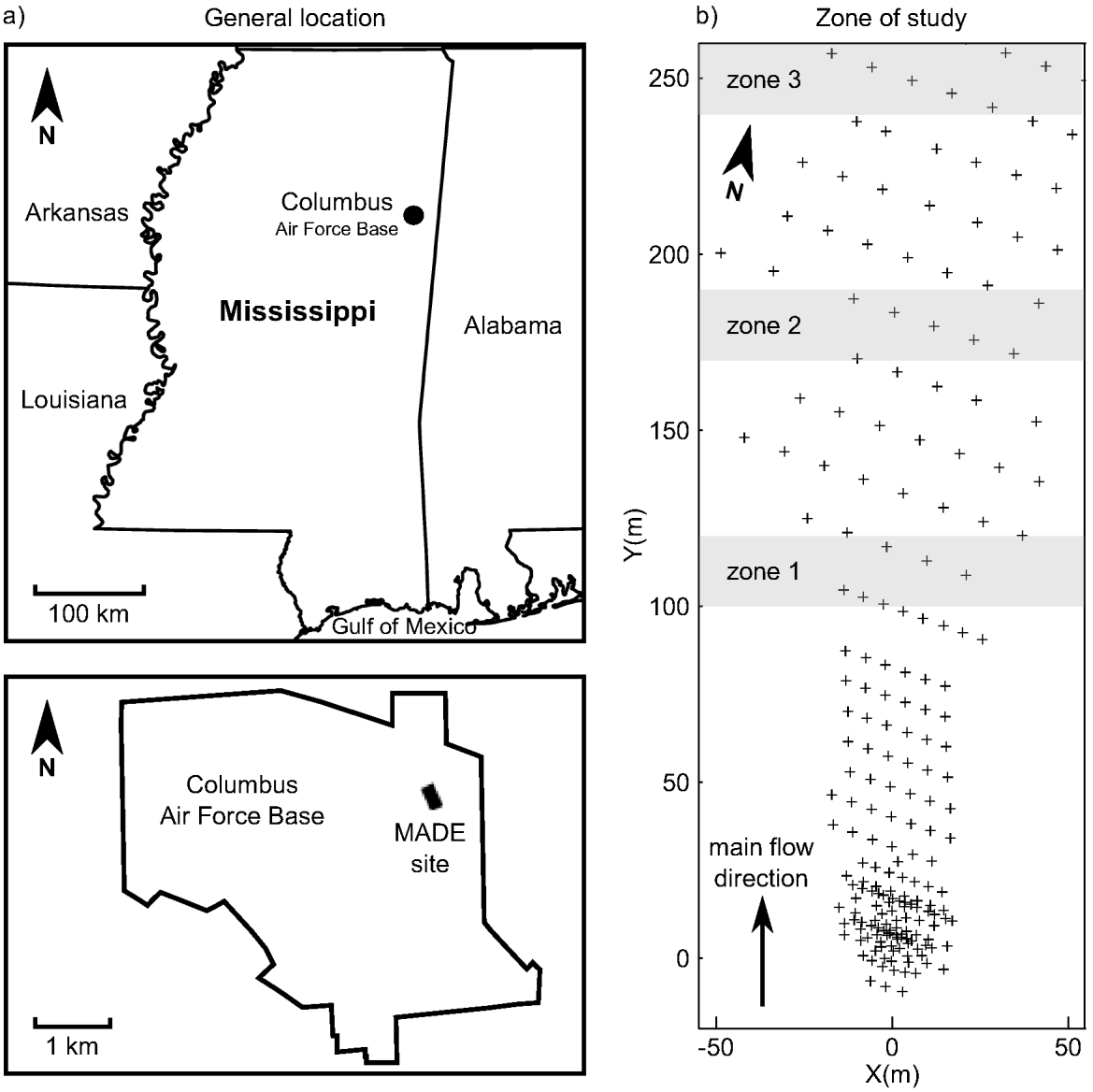

Figure 5.1: The MADE site: location (a) and piezometer network (b) – injection location corresponds to coordinates (0,0)

as density contrasts, retardation or chemical reactions. Furthermore, this experiment has been modeled by numerous teams (Feehley et al., 2000; Barlebo et al., 2004; Salamon et al., 2007; Guan et al., 2008; Llopis-Albert and Capilla, 2009), which will facilitate the comparison and the interpretation of the resulting plumes.

The MADE II experiment (Boggs et al., 1993) lasted 15 months. Five snapshots of the concentrations and the hydraulic heads were taken at time steps 27, 132, 224, 328 and 440 days after injection (see Fig. 5.1 for an illustration of the piezometers and Fig. 5.2 for an illustration of the measured concentrations at day 328). The data were measured thanks to 328 multilevel samplers, each of them being equipped with 20 to 30 sampling points space $0.38m$ apart vertically. To avoid high numerical gradient due to the injection in this exercise, day 27 will be considered as the starting date for initial conditions. As the Tritium plume was not completely sampled during the fifth snapshot, the ending date of the exercise is fixed to day 328, corresponding to the fourth snapshot and providing thus the reference points set.

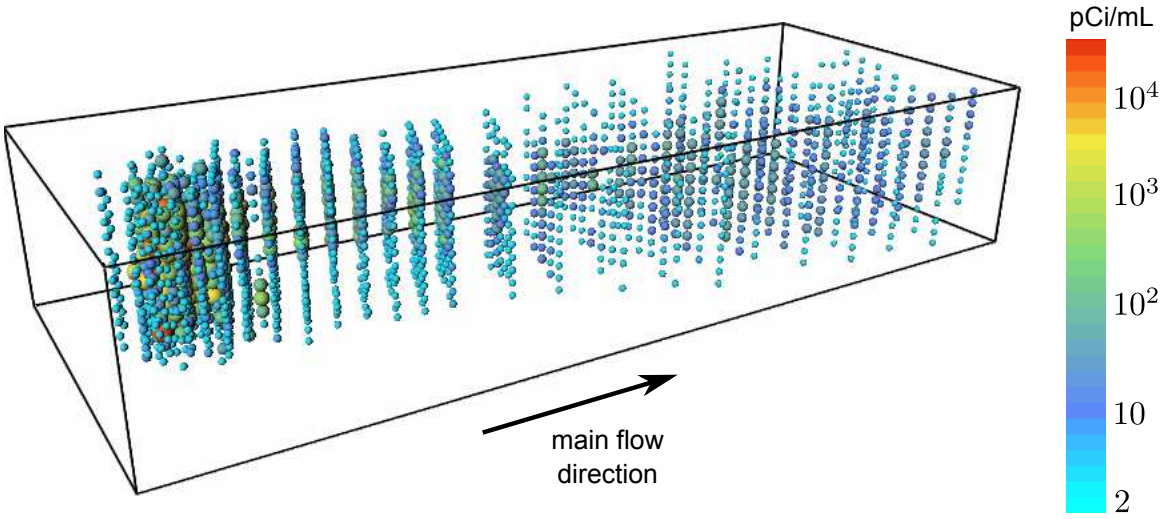

Figure 5.2: Reference concentrations measured 328 days after injection

5.2.2 Analog site data

Facies conductivity mean and variance and porosity values as well as characteristic length scales can be retrieved or computed from analog site studies (Brierley, 1989; Anderson et al., 1999; Bayer et al., 2011) or from shared database like *wwhypda* (Comunian and Renard, 2009). In this paper, all the necessary analog site data comes from the work of Jussel et al. (1994) where the main structural elements of a braided river aquifer are described precisely in terms of dimensions, hydraulic conductivity distribution and porosity. The essential information is resumed in Tab. 5.1.

Table 5.1: Geometrical and Hydraulic Parameters from Jussel et al. (1994) – GG: gray gravel, BG: brown gravel, OW: open framework, BM: bi-modal gravel

Parameters	Structural Element Type			
	GG/BG Horizontal	GG	BG	GG/BG inclined
Volumetric fraction p_i, %	57.8	9.4	15.8	4.4
Real lens length $L_i \pm \sigma_{L_i}$, m	...	27 ± 16	50 ± 29	10 ± 6
Length/Width $L/B_i \pm \sigma_{L/B_i}$	...	2.0 ± 0.5	2.0 ± 0.5	2.0 ± 0.5
Max. lens height H_i	...	$0.06L_i$	$0.04L_i$	$0.15+ 0.03L_i$
Porosity, %	17.0	20.1	14.1	17.0
Conductivity K_i, mm/s	0.08	0.15	0.02	0.1
σ_{lnK_i}	0.8	0.5	0.6	0.8
	OW/BM	Sand	Silt	OW Single
Volumetric fraction p_i, %	5.3	5.0	0.4	1.9
Real lens length $L_i \pm \sigma_{L_i}$, m	22 ± 14	9.8 ± 5.7	8.6 ± 5.8	2.6 ± 1.5
Length/Width $L/B_i \pm \sigma_{L/B_i}$	2.1 ± 0.6	2.4 ± 1.1	2.0 ± 0.5	2.0 ± 0.5
Max. lens height H_i	$0.45+ 0.033L_i$	$0.25+ 0.021L_i$	$0.14+ 0.027L_i$	$0.072+ 0.008L_i$
Porosity, %	30.0	42.6	40.0	34.9
Conductivity K_i, mm/s	10	0.26	0.005	100
σ_{lnK_i}	0	0.4	0	0

In order to compute, from the analog data, a common mean and variance of the *log* conductivity and porosity for the different geological conceptual models, we proceed as follow. Let us denote X a discrete variable representing the facies $X = i$, $i = 1, \ldots, n$, characterized by a known distribution $p_i = P(X = i)$ and $Y = ln(K)$ a continuous variable representing the *log* conductivities, noting $Y_i = (Y|X = i)$. Knowing p_i, $\mu_i = E(Y|X = i)$, $\sigma_i^2 =$

$VAR(Y|X=i)$ we can compute the mean μ_{lnK} and variance σ^2_{lnK} of Y as:

$$\mu_{lnK} = \sum_i p_i \mu_i \quad , \tag{5.1}$$

$$\sigma^2_{lnK} = \sum_i p_i \mu_i^2 - \mu^2 + \sum_i p_i \sigma_i^2 \quad . \tag{5.2}$$

Using the values of Jussel et al. (1994) summarized in Tab. 5.1 leads to a mean of the *log* conductivity and a standard deviation of:

$$\begin{aligned} \mu_{lnK} &\simeq -9.14 \\ \sigma^2_{lnK} &\simeq 3.05 \end{aligned} \tag{5.3}$$

or in base 10 $\mu_{log_{10}K} \simeq -3.97$ and $\sigma^2_{log_{10}K} \simeq 0.58$, K being expressed in m/s. The mean is porosity $\mu_n = 0.20$.

Jussel et al. (1994) also provide characteristic length scales per facies, as presented in Tab. 5.1. There are no dimension characteristics defined for the predominant facies GG/BG, as it is considered as the matrix. We assume that the connectivity through the aquifer will be mainly influenced by the dimensions of the largest and most represented structural element BG. As one can see, its dimensions might vary over a large range. To take into account this uncertainty on the prior, we define 5 scenarios of characteristic dimensions (see Tab 5.2).

Table 5.2: Dimension characteristics per scenario

Parameters	Scen. 1	Scen. 2	Scen. 3	Scen. 4	Scen. 5
Length	21.0 m	35.5 m	50.0 m	64.5 m	79.0 m
Width	8.4 m	14.2 m	20.0 m	25.8 m	31.6 m
Depth	0.84 m	1.42 m	2.00 m	2.58 m	3.16 m

5.2.3 Experimental setup

The domain dimensions retained here are the same than the one used by Salamon et al. (2007) and Llopis-Albert and Capilla (2009): $110\ m \times 280\ m \times 10.5\ m$ with a $1\ m \times 1\ m \times 0.1\ m$ resolution (3.2 millions cells). The coordinate system is designed so that the injection zone is centered on $(0, 0)$ in the horizontal plane (see Fig. 5.1). Each conceptual geological model is decomposed in 5 scenarios to cover the uncertainty on characteristic dimensions (Tab. 5.2). For each scenario, 40 geological realizations are generated, which builds an ensemble of 200 geological realizations per conceptual model.

To test the quality of the predictions obtained using different conceptual models, we define downstream of the injection site three target zones (zone 1: $100\,m < Y \leq 120\,m$, zone 2: $170\ m < Y \leq 190\ m$, zone 3: $240\ m < Y \leq 260\ m$) where simulated iso-concentrations or cumulative mass can be compared with the MADE II reference.

5.3 Geological Conceptual Models

Three different models are described hereafter to represent the geological heterogeneity of a braided river aquifer. The first one is the multi-gaussian model. Multi-Gaussian Simulations (MGS) have been generated using the turning bands technique (Matheron, 1973; Journel, 1974; Emery and Lantuéjoul, 2006). This type of model has been frequently used to represent the heterogeneity of the MADE site (Barlebo et al., 2004; Salamon et al., 2007; Llopis-Albert and Capilla, 2009), and is repeated here to facilitate interpretation and comparison with

previous work. The second method is an object based model (OBJ) developped by Huber et al. (2015). The third model uses a pseudo genetic algorithm developed by Pirot et al. (2015) for two different parameter sets (PG1 & PG2). Conductivity field realizations for the different conceptual models are illustrated in figs. 5.3 and 5.4.

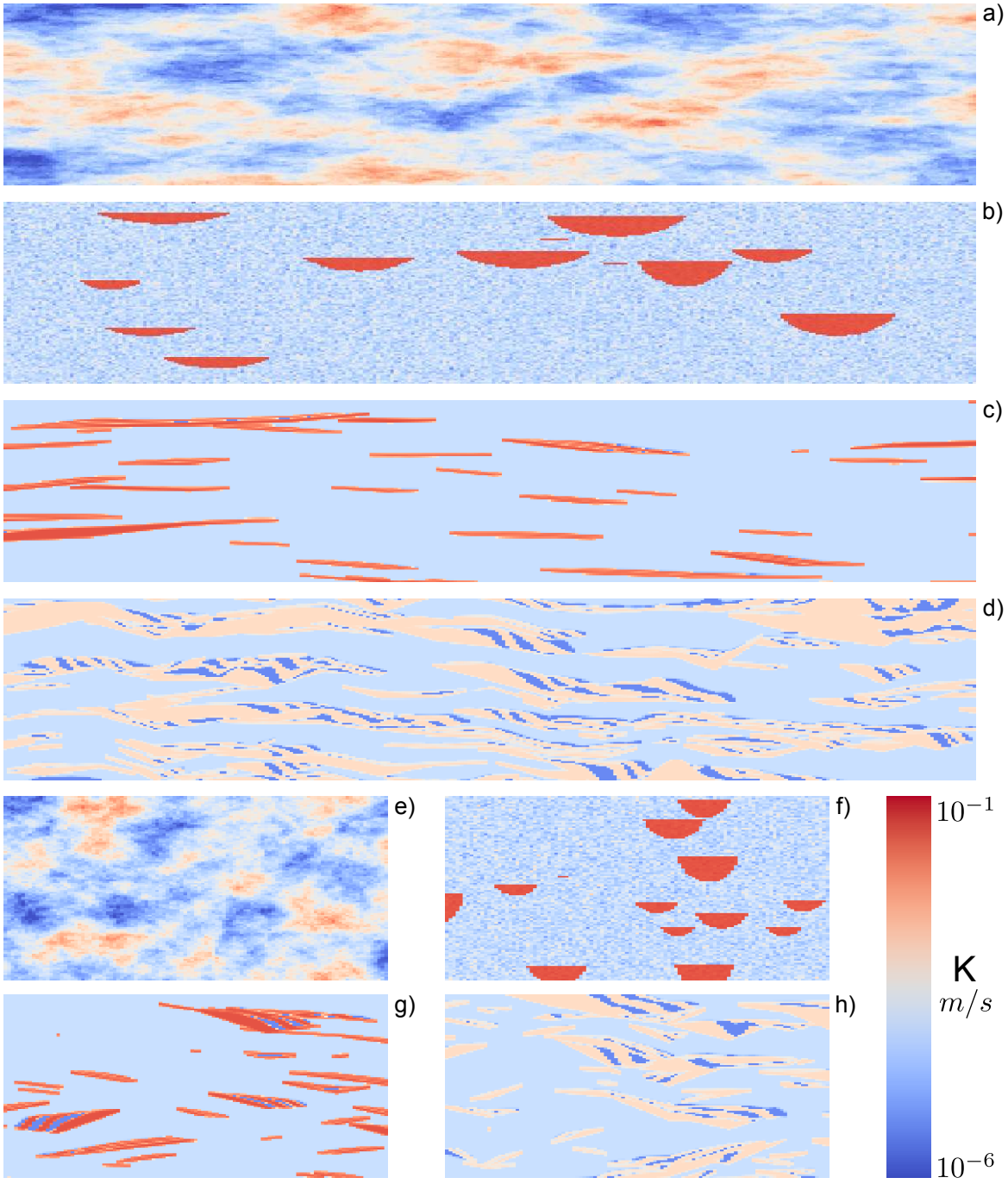

Figure 5.3: Two dimensional longitudinal (a,b,c,d) and lateral (e,f,g,h) sections of conductivity fields for the different conceptual geological models (a) & e): MGS, b) & f): OBJ, c) & g): PG1, d) & h): PG2)

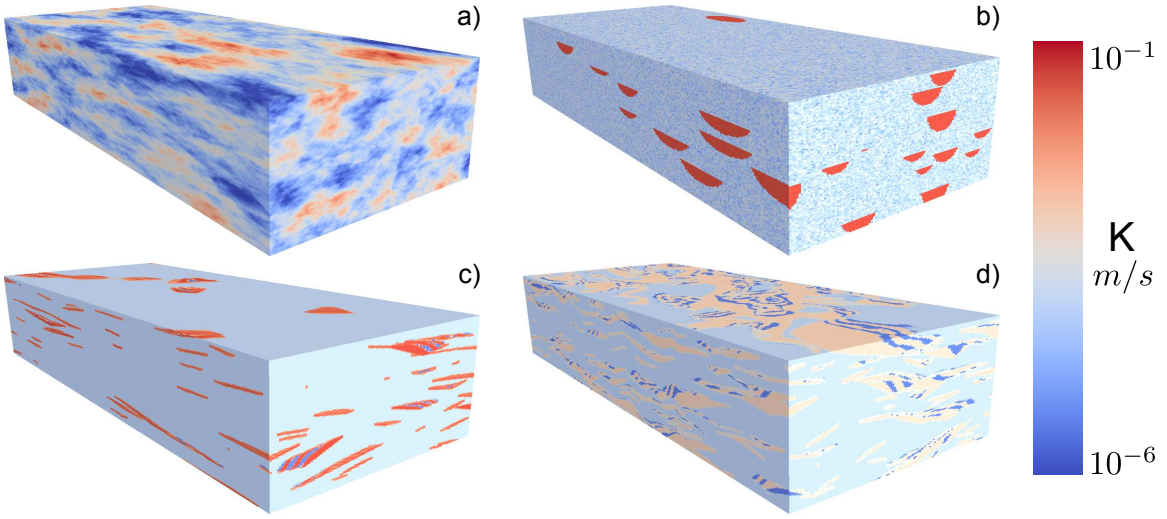

Figure 5.4: Three dimensional representations of scenario 3 conductivity fields for the different conceptual geological models (a): MGS, b): OBJ, c): PG1, d): PG2)

5.3.1 Multi Gaussian Simulations

In a data sparse context, the theoretical variogram models cannot be inferred from a geostatistical analysis of 'non-existing' conductivity samples. However *log* conductivity mean and variance as well as structural element characteristic dimensions for the same kind of geological environment are available in the literature (e.g. Jussel et al., 1994), as explained in sec. 5.2.2. Here we make the assumption that one can use the characteristic dimensions provided by Jussel et al. (1994) to estimate the variogram ranges. Without specific information, no nugget effect is considered. Here we propose to use a spherical variogram model. The ensemble of parameters to define the variogram models for each scenario are summed up in Tab. 5.3. To allow a systematic model comparison later on, the mean and variance of

Table 5.3: Variogram models per scenario

Parameters	Scen. 1	Scen. 2	Scen. 3	Scen. 4	Scen. 5
Type of variogram	spherical				
Sill	$\sigma^2_{lnK} = 3.05$				
Longitudinal range	21.0 m	35.5 m	50.0 m	64.5 m	79.0 m
Lateral range	8.4 m	14.2 m	20.0 m	25.8 m	31.6 m
Vertical range	0.84 m	1.42 m	2.00 m	2.58 m	3.16 m
Kriging mean	$\mu_{lnK} = -9.14$				
Kriging variance	$\sigma^2_{lnK} = 3.05$				

the *log* conductivity fields are normalized (Eq. 5.4) so that the resulting mean and variance are those given by Eq. 5.3:

$$lnK_{normalized} = (lnK_{sim} - \mu_{lnK_{sim}}) \times \frac{\sigma_{lnK}}{\sigma_{lnK_{sim}}} + \mu_{lnK}, \tag{5.4}$$

with $\mu_{lnK_{sim}}$ and $\sigma_{lnK_{sim}}$ the *log* conductivity mean and standard deviation of the simulated field. Illustrations of MGS simulations are given in Fig. 5.3a), 5.3e) and 5.4a).

5.3.2 Object and Process based models

From field observations and GPR profiles gathered at the Tagliamento River site (NE Italy), Huber and Huggenberger (2015) suggest that braided river aquifers heterogeneity is dominated by two main depositional elements: gravel sheets and scours. Gravel sheets are

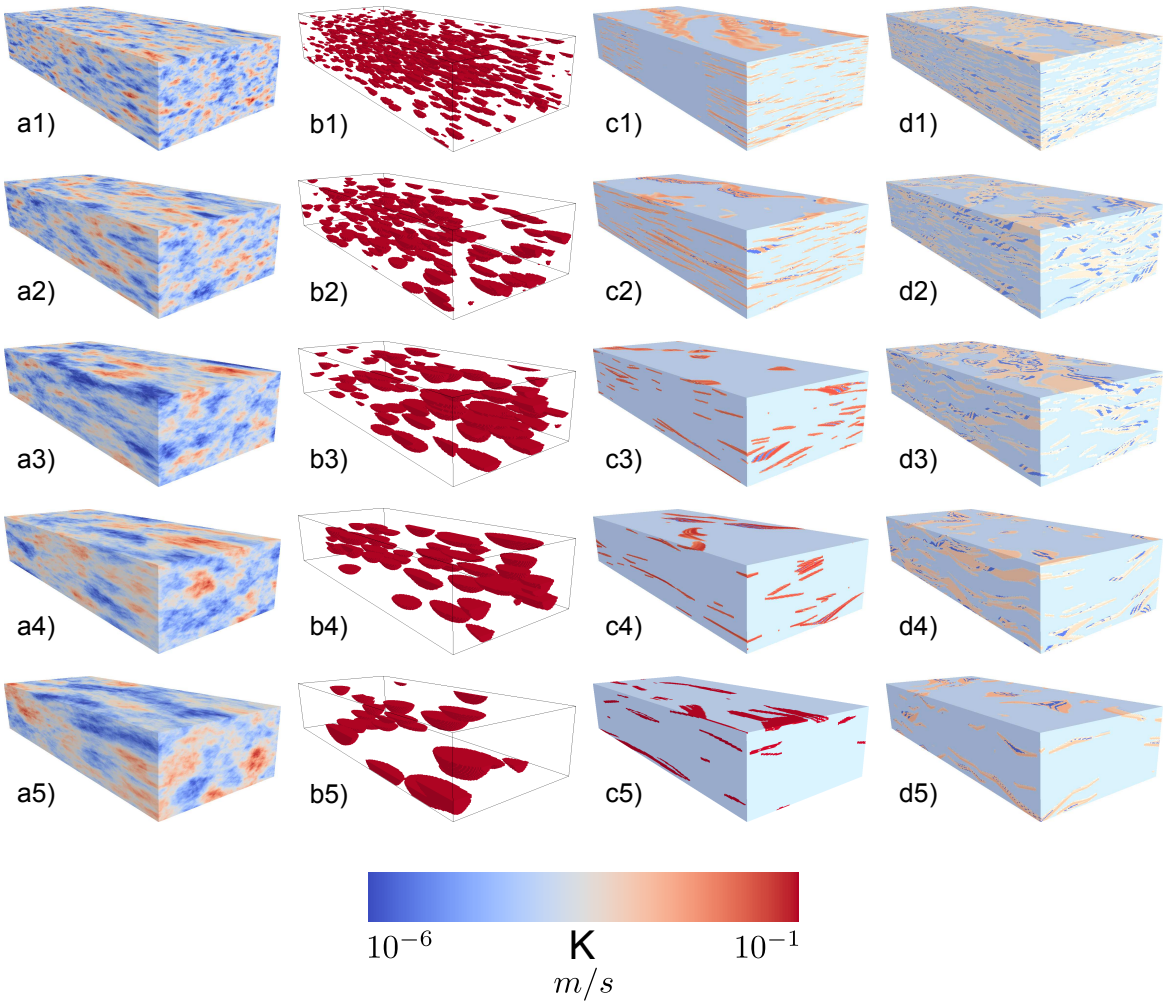

Figure 5.5: Three dimensional representations of conductivity fields for the five scenarios (1 to 5) and the different conceptual geological models (a): MGS, b): OBJ, c): PG1, d): PG2)

widespread structures composed of unsorted gravel and sand and therefore have a comparatively low conductivity while scours formed at channel confluences are filled by sorted sediment and consequently have a greater conductivity. Those observations lead the authors to develop an object based model for braided river aquifers. A brief description of the algorithm is proposed in the next paragraph. All the details are available in Huber et al. (2015).

The algorithm mimics the successive occurrence of large flood events. It produces a sequence of horizontal layers, one per large event, and each layer is composed of a gravel sheet and vertically truncated semi-ellipsoids for filled scours. The number of layers is modeled from a Poisson distribution whose parameter λ (expectation) is defined here as the ratio of the mean net sediment deposition, also called the mean aggradation ag (m), divided by the aquifer thickness. The position of each layer is taken randomly in a uniform distribution over the aquifer thickness. Then within each layer, the number of objects and their position are retrieved from a Strauss process using parameters β, γ and r. Finally, width, length, depth, orientation α and truncation parameters of the semi-ellipsoids are uniformly retrieved from bounded intervals. The resulting 2 facies models are then meshed to form a regular grid and the cells are assigned conductivities retrieved by assuming log normal distributions (Jussel

et al., 1994) without accounting for spatial correlation. The parameters and their bounding intervals used for the different scenarios are detailed in Tab. 5.4.

Table 5.4: Object model parameters per scenario

Parameters	Scen. 1	Scen. 2	Scen. 3	Scen. 4	Scen. 5
Ellipsoid semi-length a, m	$[8.0, 10.5]$	$[10.0, 17.75]$	$[15.0, 25.0]$	$[15.0, 32.25]$	$[20.0, 39.0]$
Ellipsoid semi-width b, m	$[2.0, 4.2]$	$[4.0, 7.1]$	$[8.0, 10.0]$	$[8.0, 12.9]$	$[10.0, 15.8]$
Ellipsoid semi-depth c, m	$[\frac{0.2}{0.3}, \frac{0.84}{0.8}]$	$[\frac{0.4}{0.3}, \frac{1.42}{0.8}]$	$[\frac{0.4}{0.3}, \frac{2}{0.8}]$	$[\frac{0.55}{0.3}, \frac{2.58}{0.8}]$	$[\frac{1}{0.3}, \frac{3.16}{0.8}]$
Ellipsoid orientation α	$[\frac{-40\pi}{180}, \frac{40\pi}{180}]$	$[\frac{-40\pi}{180}, \frac{40\pi}{180}]$	$[\frac{-40\pi}{180}, \frac{40\pi}{180}]$	$[\frac{-40\pi}{180}, \frac{40\pi}{180}]$	$[\frac{-40\pi}{180}, \frac{40\pi}{180}]$
Ellipsoid truncation rH, m	$[0.3, 0.8]$	$[0.3, 0.8]$	$[0.3, 0.8]$	$[0.3, 0.8]$	$[0.3, 0.8]$
Aggradation ag, m	0.05	0.1	0.16	0.2	0.25
Strauss Beta β	0.00015	0.00006	0.00004	0.000025	0.000013
Strauss Gamma γ	1	1	1	1	1
Strauss r r	20	45	60	75	90

The semi-length a, semi-width b, semi-depth c and Ellipsoid truncation rH upper boundaries are defined for each scenario such that $2 \times a$, $2 \times b$ and $c \times rH$ do not exceed the characteristic dimensions defined in Tab. 5.2. The lower range, as well as the orientation α are set accordingly to field observations by Huber and Huggenberger (2015). According to Heinz et al. (2003), the aggradation rate and the spatial density of the scours are approximately inversely proportional to the dimensions of the scours. For each scenario, the aggradation rate ag and the Strauss point process parameters β, γ and r where adjusted by trial and error to follow this relationship.

To ensure the same *log* conductivity mean and variance for all realizations, the *log* conductivity fields simulated are normalized according to Eq. 5.4. Note that because of this normalization, the mean conductivity values defined for the distribution laws do not matter and it only influences the variances to a low extent. Illustrations of OBJ simulations are given in Fig. 5.3b), 5.3f) and 5.4b).

5.3.3 Pseudo Genetic Aglorithm

This method (Pirot et al., 2015) assumes first that the main structures of a braided river aquifer, obtained after many erosion and deposit events present some internal interfaces that are similar to some extent to the observable surface topography. It supposes secondly, that the sediments are deposited in erosion scours under local flow and topography constraints which produce cross-stratifications. Inspired by principles first established by Webb (1994), the proposed method is based on stacking successive topography simulations which creates erosion surfaces and deposit volumes called geological units. These geological units are then decomposed in different geological facies according to local geomorphological conditions. Here we provide a brief overview of the method and the parameter values used for this test case. Full details are available in Pirot et al. (2015).

The successive topographies are simulated conditionally to the previous state with the Direct Sampling (DS) Multiple-Point Statistics (MPS) algorithm (Mariethoz et al., 2010) following the method developed by Pirot et al. (2014). The same DS parameters are given in Pirot et al. (2014). The training data set is composed of successive Digital Elevation Models of an analog braided river, the Waimakariri River, New Zealand, acquired by LIDAR at different time steps (Lane et al., 2003). Note that data from another analog braided river could be used. Stacking the topographies successively over the previous ones with a fixed aggradation rate erodes partly the underlying geological layers and deposits sediments that form a new geological layer. A progressive deformation scheme of the geological layers base, taking into account approximate local flow and local topography, allows for generating the cross stratified deposits.

Scaling parameters allow producing the different characteristic length scales desired for

each scenario. The algorithm is here executed with two different parameter sets PG1 and PG2 to illustrate the potential of various aggradation rates and local flow conditions. The parameters used to produce the geological units and to generate intra-unit heterogeneity are described in tables 5.5 to 5.7. For each version, the aggradation and deformation scheme parameters have been set so that the resulting sediment deposited in scours formed object with identifiable dimensions when setting the scaling factors to 1 ; PG1 version produced objects of dimensions $37m \times 22m \times 1.5m$ and of lower density than PG2 version that produced objects of dimensions $66m \times 25m \times 1$.

Table 5.5: Dimension, resolution and general parameters

	Parameter	Value
zone of interest aquifer model parameters	length (flow direction)	$280\,m$
	width (orthogonal to the flow direction)	$110\,m$
	minimum thickness - depth	$10.5\,m$
	cell length	$1\,m$
	cell width	$1\,m$
	cell height	$0.1\,m$
braided river topography dimensions	length	$11,600\,m$
	width	$1,200\,m$
	cell length	$20\,m$
	cell width	$20\,m$
interpolation parameters	margin length	$5\,m$
	margin width	$5\,m$
	margin depth	$0\,m$
facies parameters	coarse grain size sediment facies value	1
	unsorted mixed grain size sediment facies value	2
	fine grain size sediment facies value	3

Table 5.6: Scenario dependent structural parameters & iterative deformation scheme parameters for the PG1 variant

	Parameter	PG1 Value				
		Scen. 1	Scen. 2	Scen. 3	Scen. 4	Scen. 5
structural parameters	scaling factor along length axis	0.57	0.96	1.35	1.74	2.14
	scaling factor along width axis	0.38	0.65	0.91	1.17	1.44
	scaling factor along depth axis	0.56	0.95	1.33	1.72	2.11
	aggradation rate α	$0.4\,m/geological\ layer$				
deformation scheme parameters	number of iterations n	6				
	facies sequence	[1 ; 3 ; 1 ; 3 ; 1 ; 3 ; 2]				
	flow power f_p	2				
	smoothing radius r	$3\,cells$				

Table 5.7: Scenario dependent structural parameters & iterative deformation scheme parameters for the PG2 variant

	Parameter	PG2 Value				
		Scen. 1	Scen. 2	Scen. 3	Scen. 4	Scen. 5
structural parameters	scaling factor along length axis	0.32	0.55	0.77	0.99	1.22
	scaling factor along width axis	0.34	0.57	0.80	1.03	1.26
	scaling factor along depth axis	0.84	1.42	2.00	2.58	3.16
	aggradation rate α	$0.3\,m/geological\ layer$				
deformation scheme parameters	number of iterations n	6				
	facies sequence	[1 ; 3 ; 1 ; 3 ; 1 ; 3 ; 2]				
	flow power f_p	5				
	smoothing radius r	$3\,cells$				

The facies numbering is illustrated in fig. 5.6. In the facies sequence (tables 5.6 and 5.7), the matrix is defined as facies 2. It corresponds to the GG/BG horizontal structural element defined by Jussel et al. (1994). Facies 1 represents sorted coarse grain sediments as in the natural sorting occurring during the scour filling process, coarse sediments will deposit first,

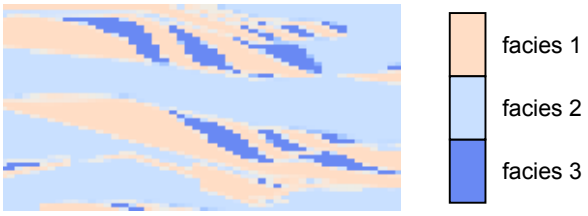

Figure 5.6: Facies illustration for the pseudo genetic models

followed by finer sediments. This is also why facies 3 represents fine grain sorted sediments. The *log* conductivity mean is known and given by Eq. 5.5,

$$\mu_{lnK} = p_1 \times ln(K_1) + p_2 \times ln(K_2) + p_3 \times ln(K_3) \tag{5.5}$$

where $\{p_1, p_2, p_3\}$ and $\{K_1, K_2, K_3\}$ are respectively the proportions and conductivity values of facies $\{1, 2, 3\}$. The variance is known and given by Eq. 5.6,

$$\sigma^2_{lnK} = E(ln(K_{sim})^2) - E(ln(K_{sim}))^2 \tag{5.6}$$

where K_{sim} is the conductivity field for the simulation. Developing it with facies proportions and conductivities, it becomes Eq. 5.7:

$$\begin{aligned} \sigma^2_{lnK} = & p_1 \times ln(K_1)^2 + p_2 \times ln(K_2)^2 + p_3 \times ln(K_3)^2 \\ & - (p_1 \times ln(K_1) + p_2 \times ln(K_2) + p_3 \times ln(K_3))^2 \quad . \end{aligned} \tag{5.7}$$

The values of μ_{lnK} and σ^2_{lnK} are set to -9.14 and 3.05 according to Eq. 5.3 and the matrix (facies 2) conductivity value is fixed to $8 \times 10^{-5} m/s$, the value estimated by Jussel et al. (1994) for the GG/BG Horizontal structural element. It allows to provide then a unique solution for K_1 and K_3, respecting the constraints edicted by the description of the facies: $K_1 > K_2 > K_3$.

As the resulting model is defined on a pillar grid (Pirot et al., 2015) – the layers of the geological model have different thickness defined on a regular grid (X, Y) – the conductivity values have to be transferred on a vertical regular grid (see Fig. 5.7). To limit the loss

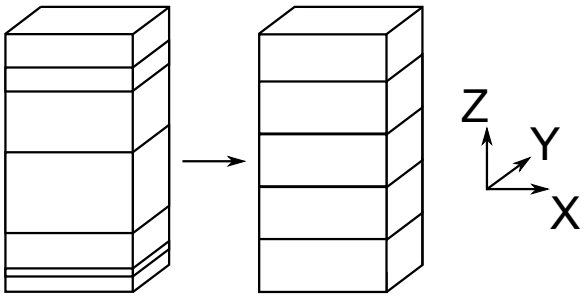

Figure 5.7: Vertical grid regularization

of information in this process, we derive vertical and horizontal equivalent conductivities (Renard and De Marsily, 1997). The horizontal component of the conductivity is computed by taking the arithmetic mean $K_{xx} = K_{yy} = \sum_i p_i K_i$ where i denotes the facies index, p_i the proportion of facies i with conductivity K_i in the concerned regular grid cell on which we want to compute de equivalent conductivity. Similarly, the vertical component of the conductivity tensor is computed as the harmonic mean of the local values $K_{zz} = \frac{1}{\sum_i \frac{p_i}{K_i}}$.

5.4 Transport model

All the heterogeneity models described in the previous section have been used as input to simulate groundwater flow and transport using the finite element code Groundwater (Cornaton, 2007). Piezometric head measurements show that the hydraulic gradient across the area has not changed significantly during the experiment. In addition the injection phase is not modeled here and the flow state is assumed to be stabilized 27 days after the injection. Therefore the flow is modeled in steady state. Solute transport is modeled by solving the advection dispersion equations in transient regime. For all the models, the mesh resolution has been kept identical to the geological models and includes about 3.2 million cells ($110 \times 280 \times 105$ cells of dimension $1m \times 1m \times 0.1m$). Iso-surfaces of the transport simulations are illustrated for the different conceptual models and their variant at various concentrations threshold in Fig. 5.9.

5.4.1 GW parameters

Following Salamon et al. (2007) and to avoid numerical instabilities due to the initial injection conditions, the tritium plume at day 27 is considered as the initial concentrations state (see Fig. 5.8). Prescribed heads of 63.1 m and 62.0 m are set respectively on the inlet and outlet faces of the model as well as a fixed concentration of 0.0 pCi/mL on the inlet face of the model; all these conditions are kept constant for the whole duration of the transient transport simulation. Inflow and outflow are limited to the inlet and outlet faces (Fig. 5.8).

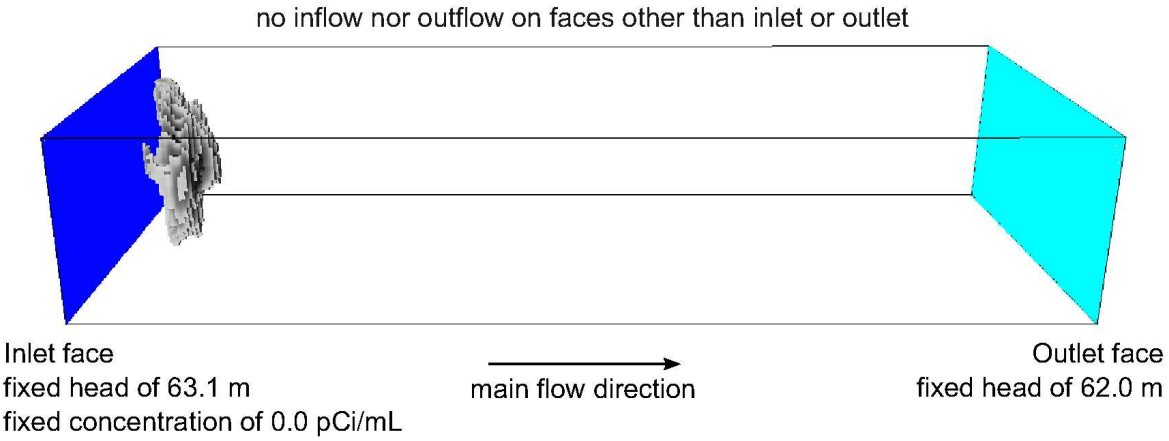

Figure 5.8: Initial and boundary conditions

The transient transport simulations are performed over 301 days, until day 328, using an automatic time-stepping strategy. The main transport parameters are kept identical for all the different subsurface models. The porosity fields are assumed equal to the mean porosity $\mu_n = 0.20$ computed in sec. 5.2.2. The longitudinal dispersivity α_L is set to 1.0 m, the transverse horizontal dispersivity α_{Th} to 0.1 m the molecular diffusion D_m to $10^{-9}\, m^2/s$ and the storage coefficient S_S to $10^{-5}\, m^{-1}$.

5.4.2 Interpolations of the concentration data

The field measurements of concentration are interpolated on the whole domain of the model at day 27 to obtain the initial state for the transport simulations. The same procedure is applied on the data corresponding to day 328 to estimate the cumulative mass along the longitudinal axis, the main flow direction. The raw data are transformed using a Normal Scored Transformation (NST) before being interpolated by Simple Kriging. The kriging results are

back-transformed into interpolated concentrations. As the natural Tritium concentration at the MADE site is around $2\,pCi/mL$ (Boggs et al., 1993), interpolated concentration values below or equal to this threshold are set to 0.

The interpolations are performed using Isatis, and the kriging parameters are described in Tab. 5.8. Note that the kriging means are set to the low boundary values of the NST

Table 5.8: Normal Scored Transformed concentrations simple kriging parameters

Parameter	Value for concentrations	
	at day 27	at day 328
kriging model	exponential + linear	exponential
Nugget	0	0.32
Sill	$0.32 + 0.45$	0.74
Range	$2.0\,m + 7.2\,m$	$16\,m$
Simple Kriging mean	-3.2	-3.4
Moving Neighborhood ellipsoid	$21\,m \times 21\,m \times 3\,m$	$21\,m \times 21\,m \times 3\,m$

concentration values, to ensure that the interpolated values located far away from the plume are equal to the background natural concentration after back transformation. Based on the field observations and on previous interpolations of the Tritium plume at day 328 (Feehley et al., 2000), it can be asserted that the entirety of the plume stays within the model boundaries. Then to keep the mass consistency between the simulations and the reference at day 328, the back-transformed interpolated concentration field is normalized so that the observed mass at day 328 is the same as at day 27.

5.5 Results

In this section, the simulated plumes (Fig. 5.9) are analyzed for each conceptual model using both quantitative and qualitative criteria. On the displayed examples, one can observe that the plume has moved close to the exit for MGS and OBJ simulations while the plume has spread with low concentrations toward the exit of the model for the PG1 simulation, high concentrations staying close to the injection zone. For the PG2 simulation, the plume stays stuck in the first half of the domain.

As the method used to interpolate the reference Tritium plume provides a smooth plume, it is reasonable to use it for computing the cumulative mass along the longitudinal axis (main flow direction), which as a sum is a smoothed representation of the plume, but it could induce some interpretation errors to rely on the result of the interpolation at day 328 as a reference over the whole domain without considering the uncertainty on the interpolated concentration, after a non linear NST. This is why quantitative indicators shall rather be computed from the sample locations.

Fig. 5.10 presents the boxplots of quantitative indicators grouped by type of conceptual model. The first boxplot represents the error on the total contaminant mass conservation ε_T (Eq. 5.8) in the model throughout the transient transport simulation duration. As explained in 5.4.2, the contaminant total mass should be preserved between day 27 and day 328.

$$\varepsilon_T = 1 - \frac{\int_V C^{sim}(x, y, z, t = 328)\,\mu_n\,dv}{\int_V C^{ref}(x, y, z, t = 27)\,\mu_n\,dv} \tag{5.8}$$

Fig. 5.10a) shows that for MGS and OBJ realizations, most of the realizations let the contaminant leave the model, while for PG1 and PG2 realizations, more than 50% of all the realizations maintain the plume within the model boundaries.

The other quantitative indicators are based on the existing sample locations. Let us denote N_S the number of samples, s the sample cell index, $\vec{x_s}$ their coordinates and C_s the

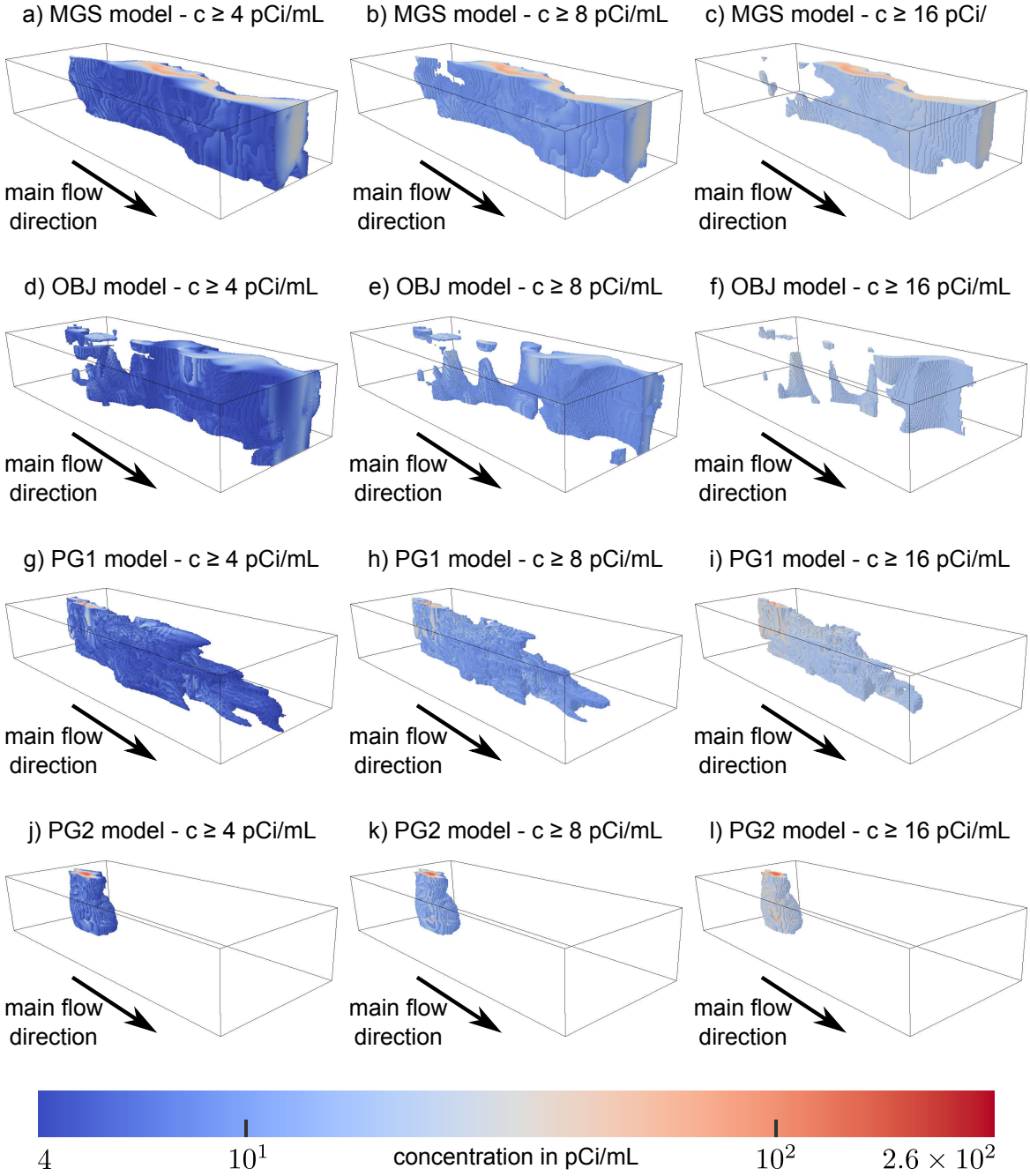

Figure 5.9: Plume examples at day 328 for different conceptual models and for scenario 3 at different concentration thresholds – vertical scale exaggerated by 5

concentrations at these locations. A simple way to define an error between a simulated plume and the reference plume might be to compute the L^1 norm error ε_C between the simulated concentrations and the reference concentrations at the sample locations (Eq 5.9).

$$\varepsilon_C = \sum_{s=1}^{N_S} \frac{|C_s^{sim} - C_s^{ref}|}{C_s^{ref}} \tag{5.9}$$

Fig. 5.10b) shows that globally, the simulated concentrations are closest to the reference for OBJ realizations than for MGS realizations and even furthest for PG1 and PG2 realizations.

Another way to look at the plume characteristics is to analyze the position of the plume

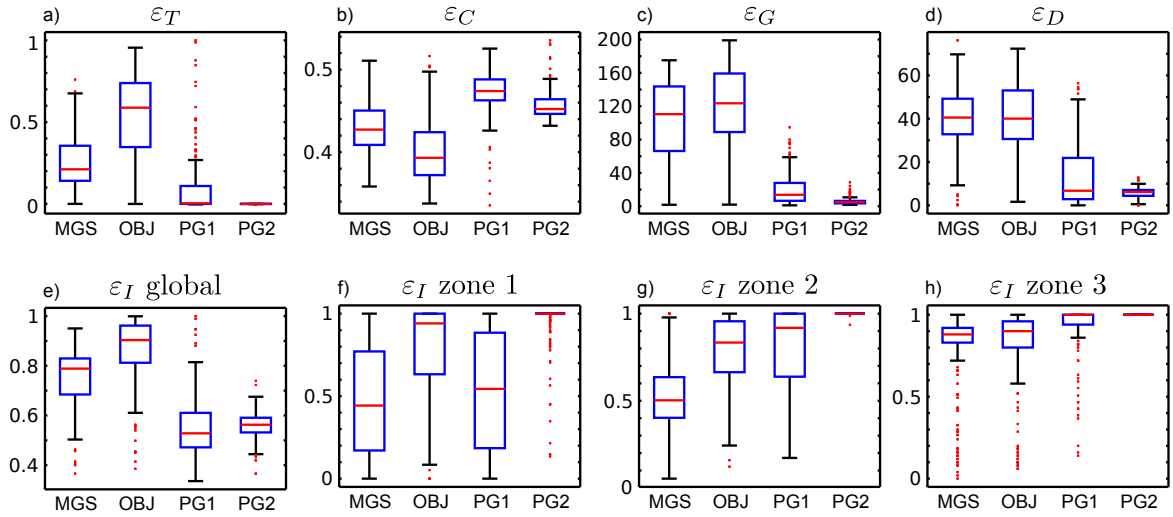

Figure 5.10: Error boxplots; a) error on the total mass; b) L_1 error on concentrations at sample location; c) error on the center of mass longitudinal position; d) error on the center of mass dispersion; e),f),g) & h) errors on iso-concentration classification by zone

center of mass and the plume dispersion around it. An error ε_G on the center of mass location is defined as the euclidean distance between the center of mass of the simulated plume $\vec{g}^{sim}$ and of the reference plume $\vec{g}^{ref}$ (Eq 5.10).

$$\varepsilon_G = \left\| \vec{g}^{sim} - \vec{g}^{ref} \right\| \text{ where } \vec{g}^{\cdots} = \frac{\sum_{s=1}^{N_S} C_s^{\cdots} \vec{x_s}}{\sum_{s=1}^{N_S} C_s^{\cdots}} \tag{5.10}$$

Fig. 5.10c) shows that the center of mass distance is small for PG1 and PG2 realizations while it is far greater for MGS and OBJ realizations. It is not surprising that these results are close to those shown by the error on the total contaminant mass conservation ε_T because a plume moving faster out of the model boundaries implies that its center of mass will also move faster towards the outlet. The plume dispersion around the plume center of mass is computed as the average distance between the center of mass and the sample location weighted by the concentration. It allows defining an error on the plume dispersion ε_D as the difference between the simulated plume dispersion and the reference plume dispersion (Eq. 5.11).

$$\varepsilon_D = \left| \frac{\sum_{s=1}^{N_S} C_s^{sim} \|\vec{x_s} - \vec{g}^{sim}\|}{\sum_{s=1}^{N_S} C_s^{sim}} - \frac{\sum_{s=1}^{N_S} C_s^{ref} \|\vec{x_s} - \vec{g}^{ref}\|}{\sum_{s=1}^{N_S} C_s^{ref}} \right| \tag{5.11}$$

Fig. 5.10d) shows that the plume dispersion is better reproduced for PG1 and PG2 models than for MGS or OBJ models.

Often, in practical applications, the key prediction is to estimate if the concentration downstream of a contaminant plume may get higher than a regulatory level. To quantify a contamination risk above determined thresholds C_k^{th} for $k \in \{1, 2, 3\}$, one might use iso-concentration classes on the simulated plumes, and to build an error ε_I based on the miss-classification at the sample locations in the simulations versus the reference classes (Eq. 5.12).

$$\varepsilon_I = \frac{1}{N_S \times N_K} \sum_{s=1}^{N_S} \sum_{k=1}^{N_K} \mathbb{1}_{C_s^{sim} \geq C_k^{th}} \times \mathbb{1}_{C_s^{ref} \geq C_k^{th}} \tag{5.12}$$

This error is computed for the whole domain (ε_I), and for three delimited zones: ε_I^1 for zone 1, ε_I^2 for zone 2 and ε_I^3 for zone 3. Indeed, it is interesting to know how the errors may vary depending on the position of the zone of interest. Zones 1, 2 and 3 are restrained on the longitudinal axis respectively between Y coordinates $]100, 120]$, $]170, 190]$ and $]240, 260]$ as illustrated in Fig.5.1. Fig. 5.10e), f), g) & h) shows that MGS simulations give better iso-concentration predictions in zones 1, 2 and 3 corresponding to the downstream part of the model while PG1 and PG2 simulations offer a better prediction on the whole model domain and by deduction in the upstream part of the model.

One might compute three-dimensional exceeding concentration probability maps of the simulated plumes $P_{C_{th}}(x_i, y_i, z_i)$, where i denotes the grid cell index in the model and C_{th} represents the concentration threshold have been computed as $P_{C_{th}}(x_i, y_i, z_i) = \frac{1}{N_R} \sum_{j=1}^{N_R} \mathbb{1}_{C_i^j \geq C_{th}}$, C_i^j being the simulated concentration for realization j in cell i. These probabilities are computed here for two concentration thresholds $C_{th} \in \{4, 16\}$, and are illustrated as probability density maps through vertical and lateral integration (Eq. 5.13 & 5.14) in Fig. 5.11 and 5.12.

$$P^Z_{C_{th}}(x, y) \propto \frac{\sum_k P_{C_{th}}(x, y, z_k)}{\sum_i P_{C_{th}}(x_i, y_i, z_i)}, \quad k \text{ layer index on the vertical } z \text{ axis} \tag{5.13}$$

$$P^X_{C_{th}}(y, z) \propto \frac{\sum_l P_{C_{th}}(x_l, y, z)}{\sum_i P_{C_{th}}(x_i, y_i, z_i)}, \quad l \text{ layer index on the transverse } x \text{ axis} \tag{5.14}$$

Vertical projections offer a top view of the iso-concentration probability densities, while the lateral projections give a representation of the plume iso-concentration densities along the vertical and longitudinal axes. One might expect to observe smaller probabilities for higher concentration thresholds, but as $P^Z_{C_{th}}(x, y)$ and $P^X_{C_{th}}(y, z)$ are re-normalized, some density maps present higher values for the highest concentration threshold in Fig. 5.12. Globally, Fig. 5.11 and 5.12 confirm the trends observed with other indicators. MGS and OBJ plumes go faster through the model. OBJ plumes seems globally a little bit faster for concentrations $< 16\, pCi/mL$ while for MGS plumes, the highest concentrations goes faster through. For MGS and OBJ plumes, the plumes seems equally dispatched on the vertical dimension, while it stays more concentrated along the longitudinal axis in the lateral–longitudinal plane. PG1 and PG2 plumes are much slower. PG2 plumes stay really close to the injection location.

One could wonder if the plume speed differences are related to the global equivalent permeability. However, a plot of the plume center of mass longitudinal coordinates as a function of the equivalent permeability (Fig. 5.13) shows that even though the greatest equivalent permeabilities are obtained for the PG1 models, the longitudinal position of the center of mass for these models are not the farthest from the injection location. Therefore the equivalent conductivity does not explain all the effects on the plume speed differences between the various conceptual models in this example. Several reasons might be considered. One of them could be the sensitivity of the initial conditions such as the injection location on the plume dispersion. Another one is probably the way to set the conductivity property for the matrix or to fix the minimum and maximum conductivity values for the facies models.

The previous errors or probability density maps are rather global and aggregated indicators. One might want less aggregated indicators to assess the individual quality of the predictions. A more qualitative way to assess the quality of the simulated plumes is to compare the mass distribution along the longitudinal axis as performed in previous studies of the MADE-II experiment (Salamon et al., 2007; Llopis-Albert and Capilla, 2009). The

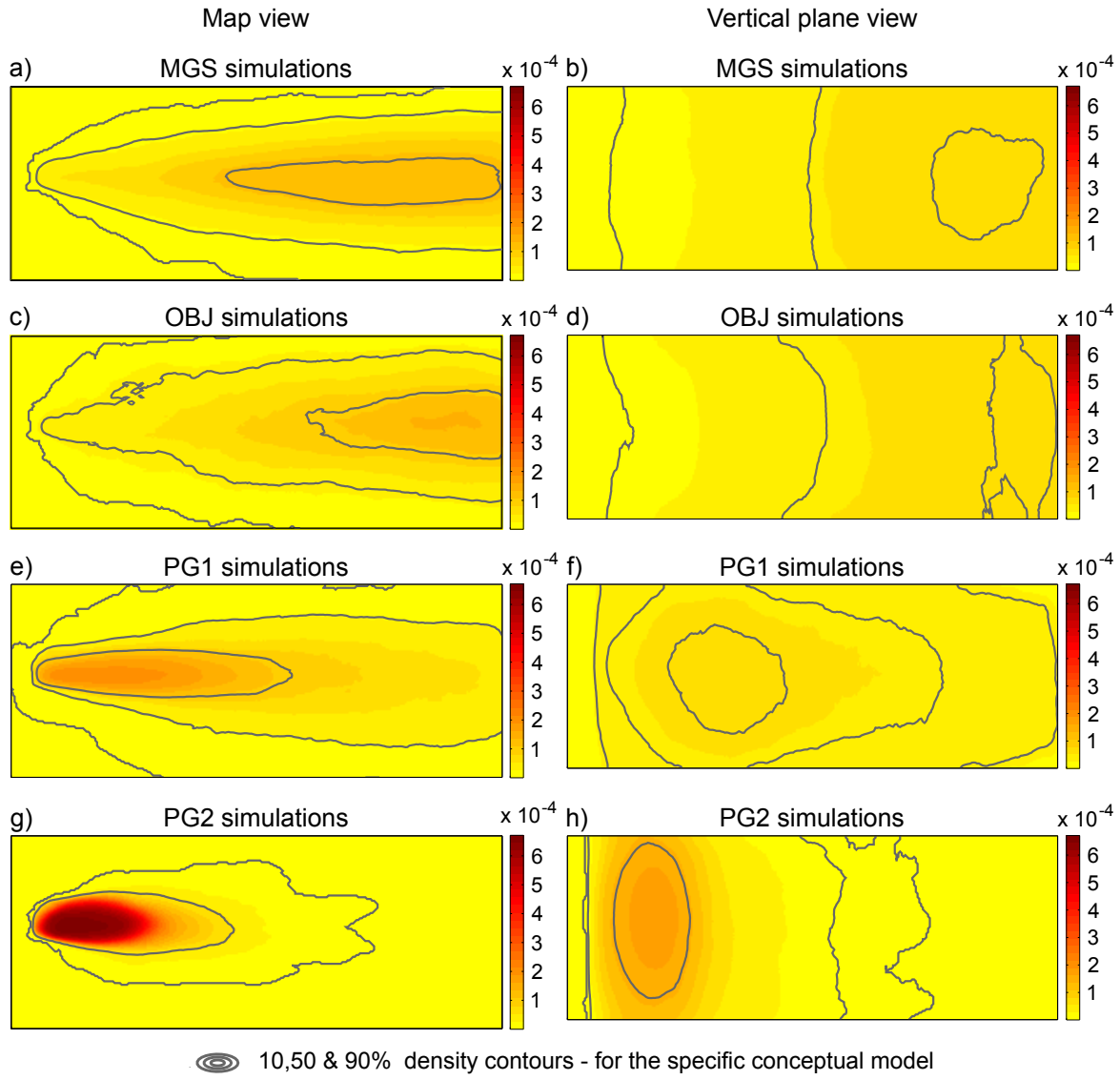

Figure 5.11: Renormalized probability densities per model for $C \geq 4\,pCi/mL$

ensemble of mass distribution curves and some typical profiles are illustrated separately for each conceptual model in Fig. 5.14. The mass distribution curves present different profiles from gaussian bell to multimodal distributions, depending on the length scale characteristics scenario and on the type of conceptual model. Multimodal profiles are produced by all types of conceptual models, but profiles showing a high peak close to the coordinate 10 and an accumulation zone around coordinate 170 are sparse (Fig. 5.14a realization 106 and Fig. 5.14c realization 182 with a strong attenuation due to mass loss). The 10, 50 and 90 percentile of the mass ditribution curves are shown in Fig. 5.15. They confirm the global trends observed with the quantitative indicators (Fig. 5.10), that MGS and OBJ simulations (Fig. 5.15a & 5.15b) produce faster plumes, do not include the first peak in the mass distribution curve, but are better at predicting the 'accumulation' zone, in opposition to PG1 and PG2 simulations (Fig. 5.15c & 5.15d). However, the global shape of the cumulative mass distribution along the flow direction looks better and is more promising for PG1 and PG2 simulations: the peak around $Y = 0\,m$ and the accumulation around $Y = 170\,m$ for the reference show a bi-modal behavior that seems more often obtained by the PG1 and PG2 simulations.

To assess the quality and the uncertainty of the simulated plume, one can look at the

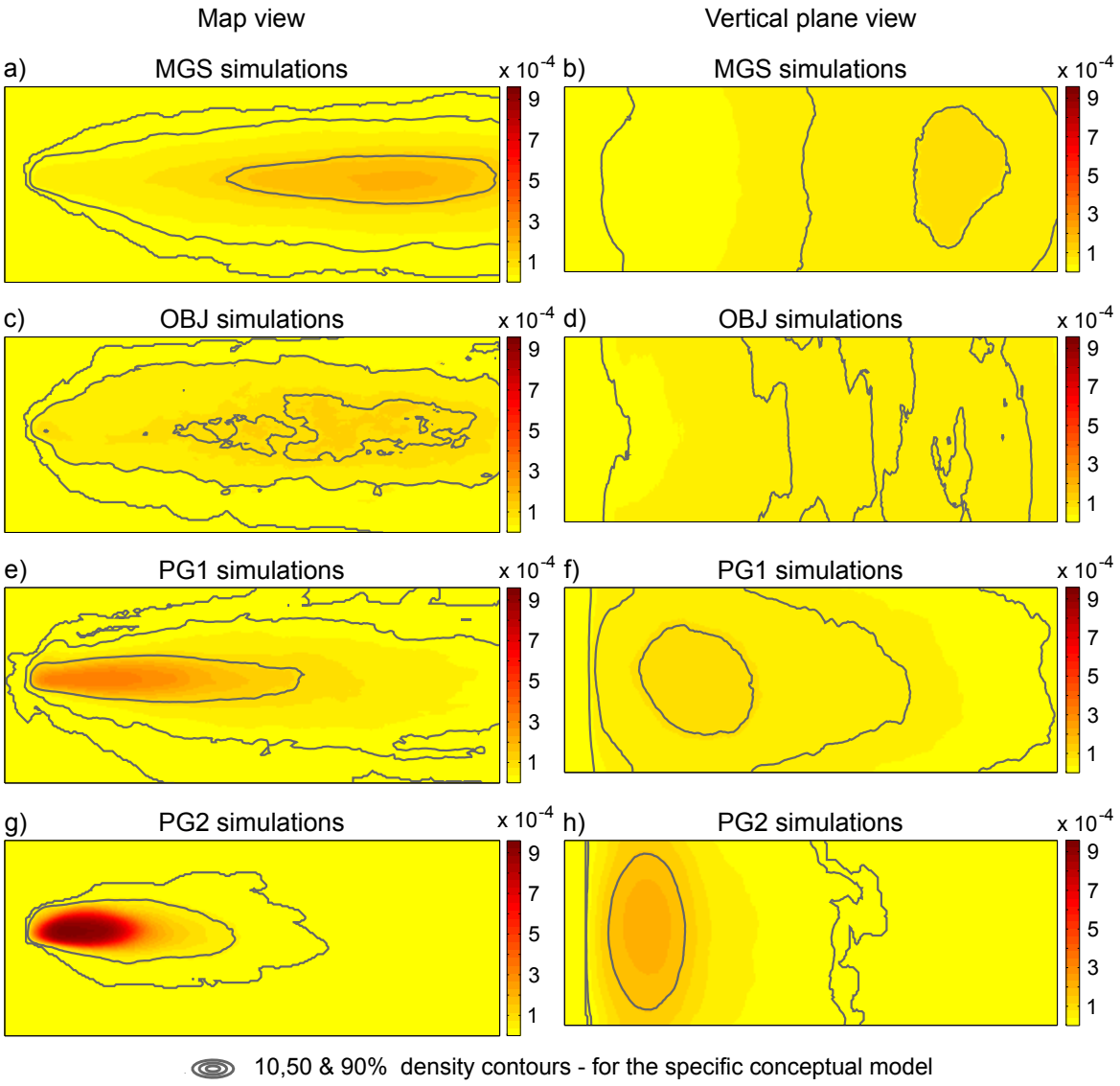

Figure 5.12: Probability densities per model for $C \geq 16\, pCi/mL$

histograms of the cumulative mass in the volumes delimited by the three target zones (Fig. 5.16) for the simulations versus the cumulative mass for the reference. For each zone, the different distributions around the reference value show that the uncertainty completely depends on the observation zone. Of course it also depends a lot on the kind of model. The best predictions for zone 1 are produced by OBJ simulations. The best predictions for zone 2 are given by MGS simulations. The best predictions for zone 3 come from PG1 simulations. PG2 simulations are outperformed in the target zones. However, would have a zone been further downstream, it would have outperformed the other simulations. The asymmetric uncertainty distribution around the reference values confirm the previous observations about the global plume speed and spreading.

A global statement is that the variability of these indicators is not uniform and strongly depends on the type of indicator, on the observation zone, and on the type of conceptual model.

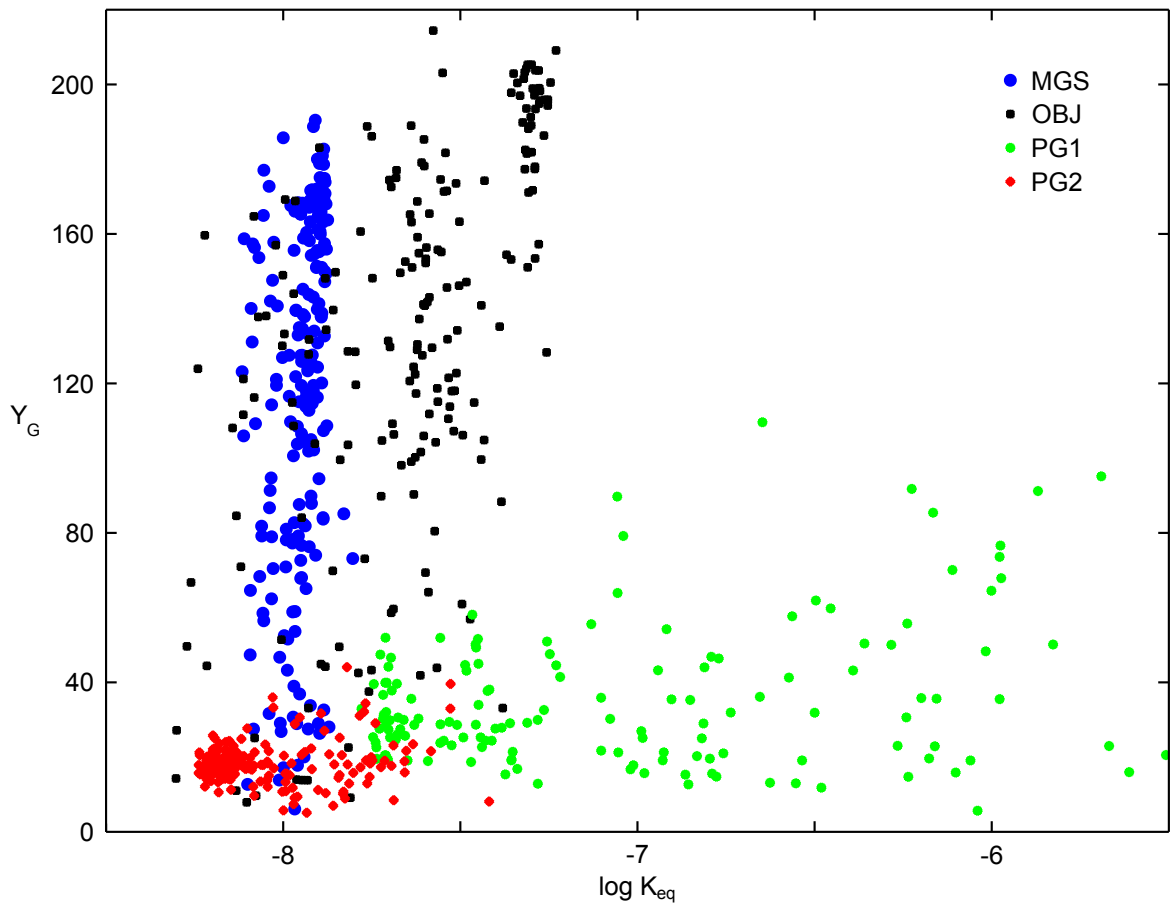

Figure 5.13: Plume center of mass longitudinal position as a function of the equivalent permeability

5.6 Discussion and conclusion

It has been shown that with limited site-specific information, it is possible to propose geological realizations based on different types of conceptual models and analog data, and then to run transient transport simulations afterwards. Globally, the predicted plumes are quite acceptable with respect to the assumption of field information scarcity. It is probably due to the closeness of the conductivity mean $\mu_{log_{10}K} \simeq -3.97$ and variance $\sigma^2_{log_{10}K} \simeq 0.58$ values computed from analogue data (Jussel et al., 1994) with the conductivity mean $\mu_{log_{10}K} \simeq -4.32$ and variance $\sigma^2_{log_{10}K} \simeq 0.80$ measured on the MADE site, as reported by Salamon et al. (2007). None of the conceptual models was able to predict the plume behavior and uncertainty in a completely satisfying manner. However, considering multiple conceptual models together might help to compensate the conceptual model imperfections. In the demonstration example, MGS or OBJ plume realizations can help localize the accumulation zone while PG1 or PG2 models would help considering the slow release of the contaminant. In a sense, all conceptual models can be useful.

An important aspect of this case study is that the variability of the OBJ and PG geological models was not fully explored, inducing then an underestimation of the simulated plume uncertainty. Indeed the two parameterizations proposed for the PG algorithm show that a wide variety of structures, densities and connectivities might be generated, by playing with different input parameters. Also, our strategy to reduce the degree of freedom to compute the conductivity values so that the mean and variance of the *log* conductivity field is the same

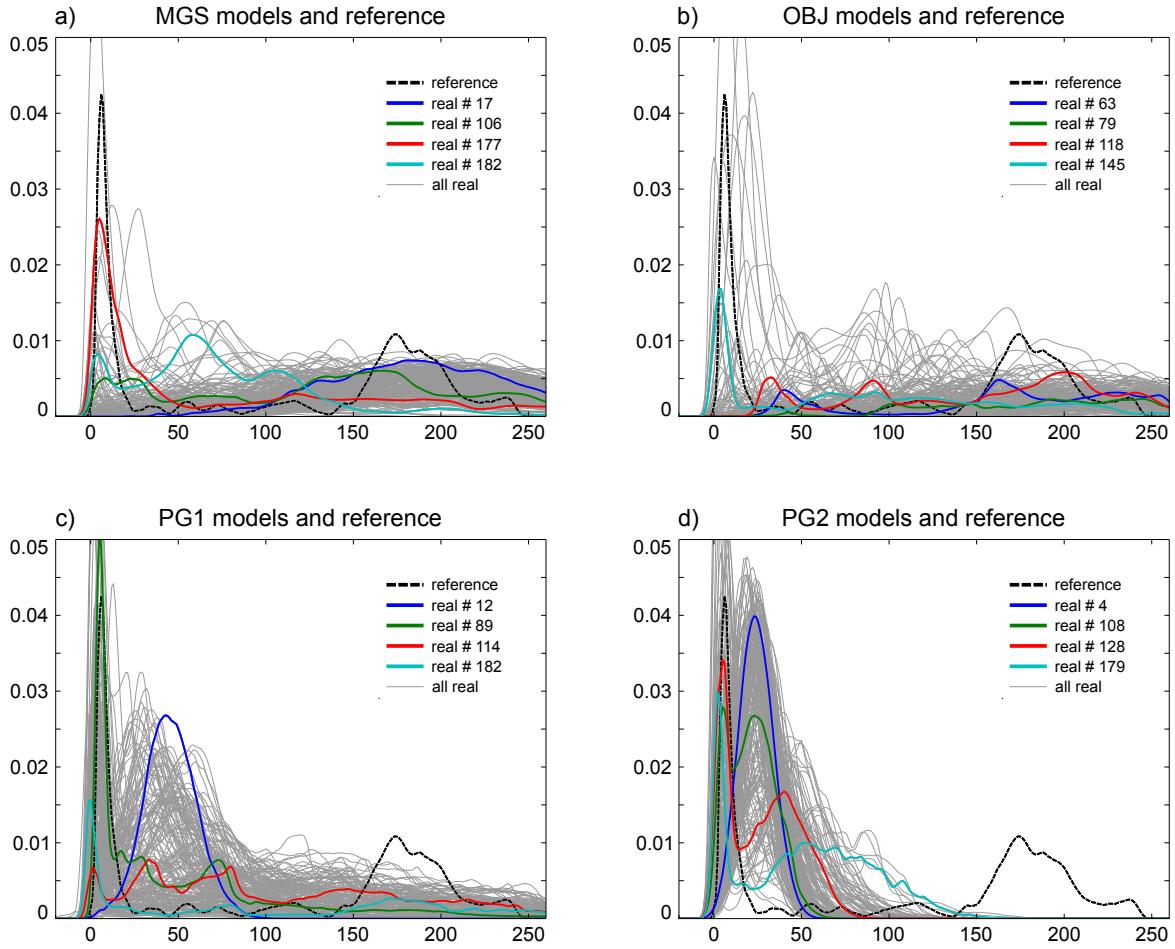

Figure 5.14: Cumulative mass along main flow direction: typical breakthrough curves

for all realizations is also a limiting factor. Fixing the minimum or maximum value instead of the matrix conductivity value would have changed the simulated plumes. Regarding the OBJ model, varying the density of objects per layer or the aggradation rate would also have changed the structures and connectivities of the geological models.

Further parameter space exploration should be performed, and as it would increase the contaminant plume uncertainty, the question of adding information during the modeling process to reduce the prediction uncertainty should be raised. As it would be unrealistic in many cases to add as much conductivity measurements as available for the MADE site, a suggestion is to refine length scale characteristics by measuring Ground Penetrating Radar profiles on the site to model. These profiles could be transformed into summary statistics, offering a global selection criteria, or also used locally as conditioning data for geological layer interfaces. This is work the authors intend to perform soon. More generally, the expected impact of data conditioning is a reduction of the variability within and between the conceptual models. This could also be verified.

Some limitations of this study offer further perspectives. Indeed, the study has been conducted on a single reference site and it is not possible to draw a general conclusion. A similar modeling effort shall be conducted on other braided river aquifer study sites, to check if the resulting prediction uncertainty confirms the results described here. At last, as the plume behavior might not be explained by the equivalent conductivity, the other factors controlling the plume transport shall be analyzed in more details. In particular, the initial location of the plume might have a considerable impact on the contaminant transport,

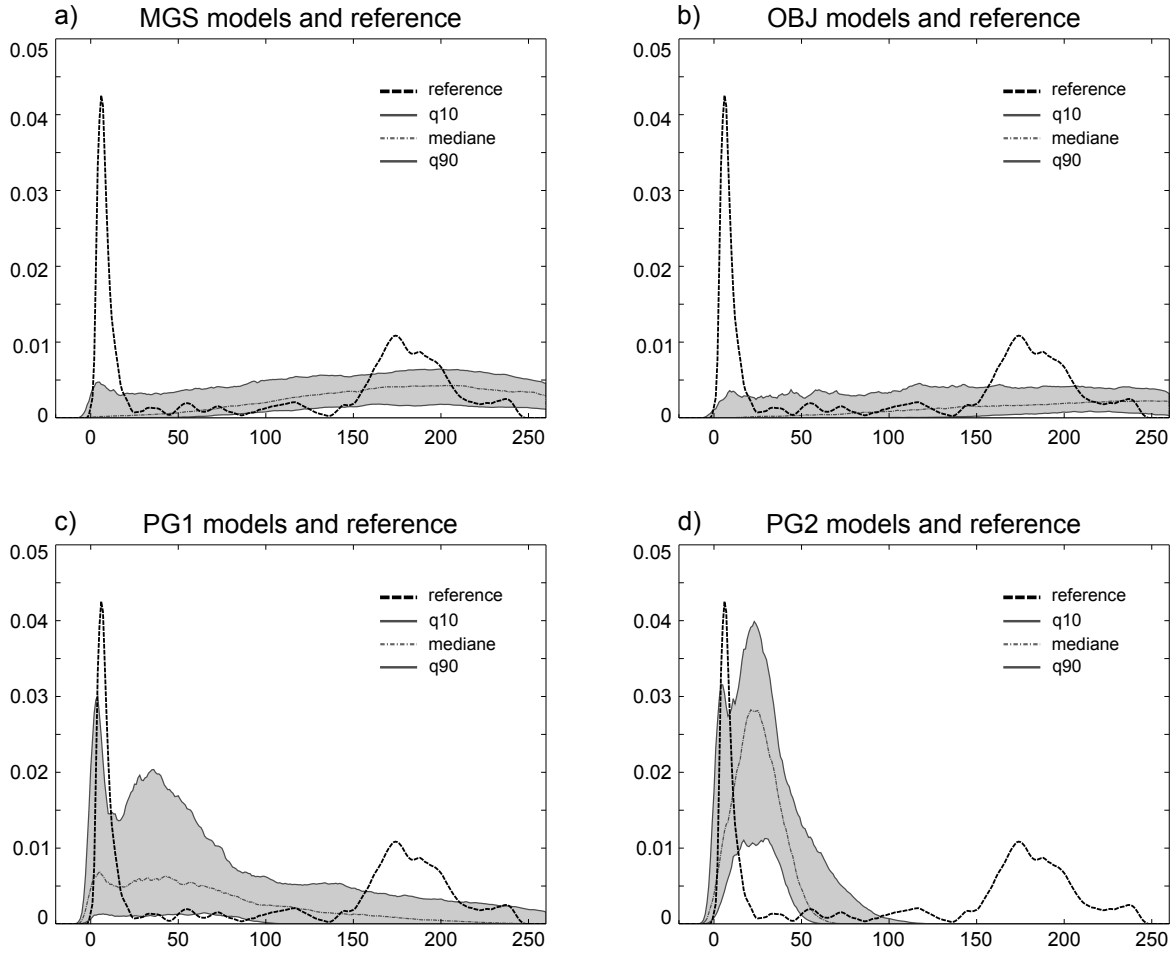

Figure 5.15: Cumulative mass along main flow direction: quantiles on breakthrough curves

especially when the high concentrations are very localized.

Acknowledgments

The work presented in this paper is part of the ENSEMBLE project, funded by the Swiss National Science Foundation under the contract CRSI22 1222491. The authors would like to thank Jim Butler for its enthusiasm in sharing the data from the MADE site with us and Fabien Cornaton for his great advice about the parameterization and use of Groundwater .

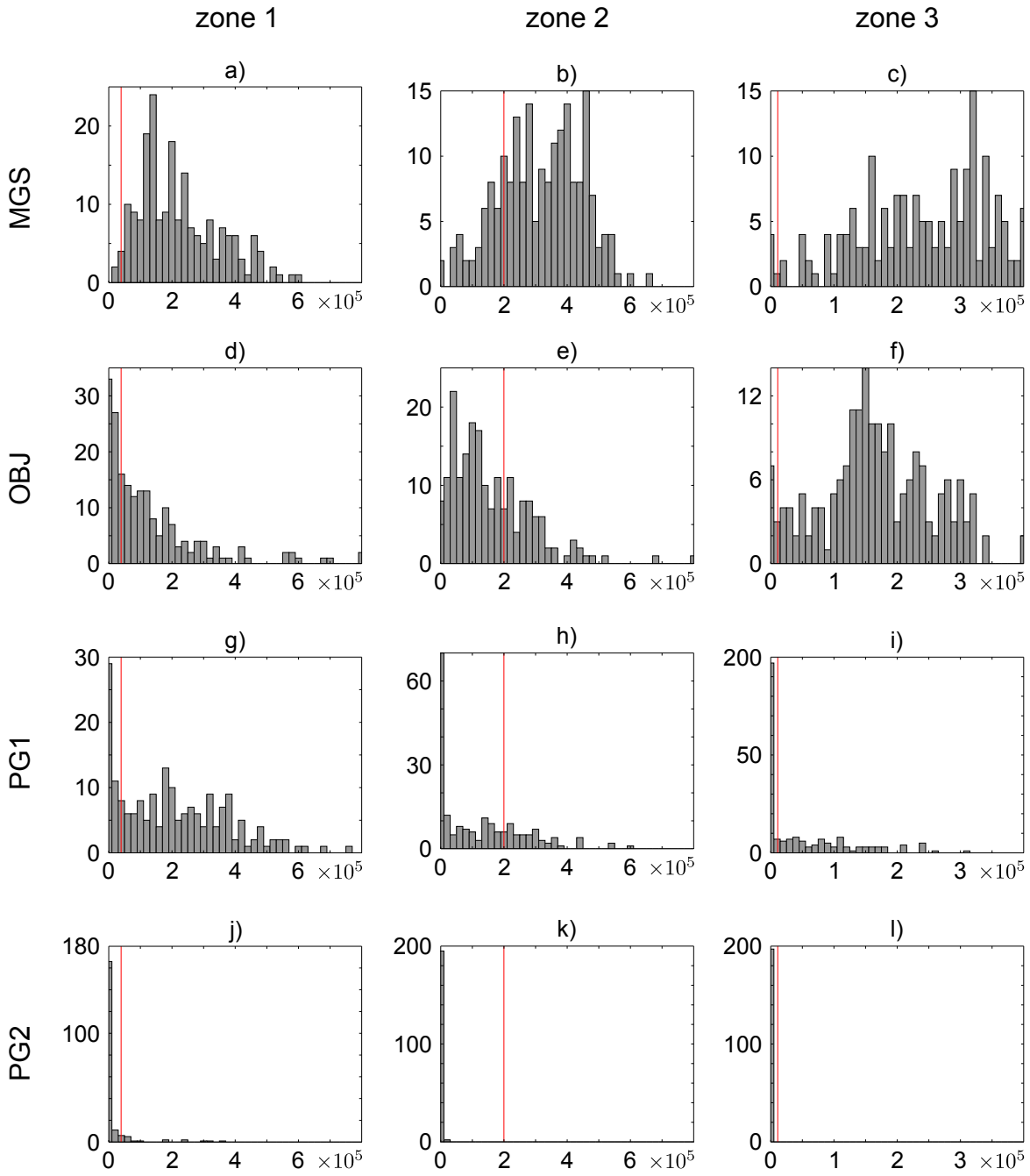

Figure 5.16: Histograms of cumulative mass (pCi) predictions in the three target zones; reference value: vertical red line

Chapter 6

Conclusion

6.1 Summary of the results

The sensitivity analysis of the DS algorithm (chapter 2) provides recommendations on the three main input parameters, and shows examples of MPS simulations for a wide range of new training sets. An acceptance threshold t smaller than 0.2 and a maximum number of neighbors n larger than 30 provide rather good simulations for almost any training set, even if the quality of the realizations also relies on the quality of the training sets. Reducing t and increasing n improves the quality of the simulations at the expense of the computational time and decreases the variability between realizations. The quality of the realizations seems to converge to an asymptote as soon as the computational time increases faster. A possible way to decrease the computational time is to reduce f, the proportion of the training set used, as long as the training set contains enough similar patterns. Another possibility is to work on an upscaled grid.

The simulation of topographies at successive time steps with MPS (chapter 3) is another demonstration of the possibilities offered by the DS algorithm. It shows that combining the variable of interest with auxiliary data that help describing the non-stationarity of the variable to simulate is a very efficient way to generate complex structures with MPS techniques, confirming the results of Chugunova and Hu (2008) and Mariethoz and Kelly (2011). It also underlines the potential of variable transformation to reduce complexity and to gain computational time, as also performed by Mariethoz et al. (2012). A noticeable result is the possibility to use time series of maps for training sets. It complements the ability of MPS to model local time series (Oriani et al., 2014).

These results allowed us to propose a new process imitating method to model braided river aquifers (chapter 4) using analogue data. Successive topographies, obtained by MPS simulations using a training set coming from an analogue site, are stacked iteratively to produce geological layers whose mean thickness is a function of the aggradation parameter. The inner layer heterogeneity produced by successive deformation leads to realistic cross stratified sedimentary structures. Though limited to the aggradation rate, the number of iterative deformation and the power of the deformations, the main parameters might not be easy to calibrate but they offer a control on the facies proportion and connectivity.

The test of the proposed process imitating method on a real contaminant transport data set and the comparison with a multi-gaussian technique and an object based algorithm (chapter 5) was not a success nor a failure for any model. It demonstrates the potential of building geological heterogeneity models using analogue data when few field data are available, as it is often the case in practice for hydrogeologists. An important result is that the uncertainty of the plume predictions depends on both the kind of geological model and the position of the target zone for the prediction. It also points out that when little information is available, even if none of the models is right, each of them should be considered in order to not underestimate the uncertainty of the predictions. A drawback of the conducted test is the restriction applied to the model parameters. A wider exploration of the parameter space should be performed for both object based and pseudo genetic models.

6.2 Directions for further research

The limitations encountered and the questions raised throughout this thesis open further perspectives. MPS topography simulations are potentially useful to study river aquifer interactions (Käser et al., 2014) and potentially constitute a new tool for flood risk management. Indeed, this new way of modeling topographies is rather computationally efficient in a probabilistic framework, and probably faster than process or physics based models that require a minimum number of time steps before stabilizing to the desired state.

The sensitivity analysis performed on the DS algorithm allows to consider an auto-calibrated method to optimize the DS algorithm parameters. As a sudden increase of the computation time is always observed when quality indicators converge to their optimum, one might consider performing a parameter space exploration on a small size grid to determine the best parameter set. The best parameters could be determined around an elbow of the multi-dimensional surface representing the computational time as a function of the parameters. Of course, the grid size used for testing should be large enough to account for the inner ergodicity of the training set.

This study also points out the need for new tools to assess quantitatively the quality of MPS simulations and of training sets, in addition to already existing indicators (Boisvert et al., 2010; Dimitrakopoulos et al., 2010; Emery and Ortiz, 2011). Meanwhile, Rongier et al. (2013) developed some control indicators to check properties of connected geobodies, based on the facies proportions and connectivities and on the shape, geometrical properties and skeleton of the geobodies. Tan et al. (2014) proposed a distance based on multiple-point histogram comparison between two data-sets (training-set or realization). Pérez et al. (2014) suggested a multiple and high order consistency check, based on the retrieving in the training set of randomly sampled pattern in realizations. Another aspect in the quality of MPS realizations is the tendency to patch (copy paste) from the training set. This is observed when too much constraints are applied on the number of neighbors and on the acceptance threshold parameters. Under such constraints, the algorithm chooses the best pattern and performs an optimization instead of sampling among acceptable patterns. This could be avoided by adopting some accept-reject step as in classical Markov Chain Monte Carlo algorithms.

Further developments of the pseudo-genetic model could be proposed to allow field data to be honored. One aspect relates to the conditioning to erosion surfaces corresponding for instance to major reflectors in Ground Penetrating Radar profiles. It involves simultaneous constraints on both the topography simulations and the aggradation rate. Another aspect is to take borehole data into account. An identification of the successive interfaces between the main geological layers as well as a correlation between borehole stratification (Olea, 2004) would serve for conditioning both the topography simulations and the aggregation rate. Another possibility would be to deal with conditioning as a minimizing cost function, balancing a deformation cost versus a non conditioning cost.

The pseudo-genetic model developed in this thesis is an opportunity to generate 3D training sets for MPS simulations. Indeed, though MPS are able to reproduce many complex patterns, in particular for heterogeneity models of braided river aquifer (Comunian et al., 2011), there is a lack of available training sets (Comunian et al., 2012). As a matter of fact, using MPS on a training set generated with the proposed approach would be another solution to achieve data conditioning, either by re-simulating with around the conditioning data locations or by simulating the entire simulation grid by MPS.

Other aspects of the pseudo-genetic model such as the parameterization of the number of facies could be investigated further. In this thesis, three facies are retained to describe a connected matrix and the sediments deposited in scours. The matrix, that represents gravel sheets, is composed of unsorted mixed sediments and therefore has a rather low conductivity. The two facies of the scour sediments correspond to well sorted coarse grain size sediments and well sorted fine grain size sediments. Their alternation allows to create contrasts in the sedimentary deposits, which produces cross-stratifications. One could wonder if two facies as conceived in the object-based model of Huber and Huggenberger (2015) would be enough or if more facies would offer better results regarding the uncertainty of the plume predictions.

Another possible variant in terms of access to analog data would be to consider topographies issued from physics based models such as developed by Schuurman et al. (2013). It

could be used either as training set or to provide the successive topographies with the corresponding aggradation rate. A great advantage to use such a physics model is the possibility to link the aggradation with the two successive topographies used as training set.

Testing the proposed pseudo genetic algorithm on the MADE site Tritium experiment indicates that a wider parameter space exploration is required. As it could dramatically increase the uncertainty of the predictions, it shall be combined with some parameter selection or calibration tools. For instance, it could be combined with summary statistics comparisons (Lochbühler et al., 2015) between field geophysical measures such as Ground Penetrating Radar profiles and simulated geophysics. This requires the development of indicators to compare forward geophysical methods with field measurements.

The plume simulation exercise performed on the MADE site with three kinds of conceptual models of geological heterogeneity should be repeated on other investigated sites in order to draw some general conclusions. One might indeed wonder if the same form of plume prediction uncertainty would be obtained for various sites. However, such experiments conducted at a large scale for braided river aquifers are scarce in the literature.

Finally, it would be interesting to account for the different sedimentary environments at the MADE site in the modeling process and assess their impact on contaminant transport. The braided river alluvial aquifers stand on an aquitard formed by Cretaceous marine sediments. The two meter thick surface of the aquifer is composed by a mixture of fine particles, sand and gravels. Underneath, the main part of the aquifer is essentially constituted of sand and gravel. It would require an additional modeling effort to characterize each zone in accordance with its composition and also an additional computational effort as the aquifer thickness to model would be greater.

Bibliography

Bibliography

Alcolea, A. and P. Renard (2010). Blocking Moving Window algorithm: Conditioning multiple-point simulations to hydrogeological data. *Water Resources Research 46*(8), W08511.

Alharthi, A. and J. Lange (1987). Soil water saturation: Dielectric determination. *Water Resources Research 23*(4), 591–595.

Allen, J. (1983). Studies in fluviatile sedimentation: bars, bar-complexes and sandstone sheets (low-sinuosity braided streams) in the Brownstones (L. Devonian), Welsh Borders. *Sedimentary Geology 33*(4), 237–293.

Alumbaugh, D., P. Y. Chang, L. Paprocki, J. R. Brainard, R. J. Glass, and C. A. Rautman (2002). Estimating moisture contents in the vadose zone using cross-borehole ground penetrating radar: A study of accuracy and repeatability. *Water Resources Research 38*(12), 45–1.

Amoros, C. and G. Bornette (2002). Connectivity and biocomplexity in waterbodies of riverine floodplains. *Freshwater Biology 47*(4), 761–776.

Anderson, M., J. Aiken, E. Webb, and D. Mickelson (1999). Sedimentology and hydrogeology of two braided stream deposits. *Sedimentary Geology 129*(3), 187–199.

Ashmore, P. E. (1982). Laboratory modelling of gravel braided stream morphology. *Earth Surface Processes and Landforms 7*(3), 201–225.

Ashmore, P. E. (1988). Bed load transport in braided gravel-bed stream models. *Earth Surface Processes and Landforms 13*(8), 677–695.

Ashworth, P., J. L. Best, J. O. Leddy, and G. W. Geehan (1994). *The Physical Modelling of Braided Rivers and Deposition of Fine-grained Sediment*. John Wiley and Sons.

Ashworth, P., R. Ferguson, P. Ashmore, C. Paola, D. Powell, and K. Prestegaards (1992). Measurements in a braided river chute and lobe: 2. Sorting of bed load during entrainment, transport, and deposition. *Water Resources Research 28*(7), 1887–1896.

Ashworth, P. J. (1996). Mid-channel bar growth and its relationship to local flow strength and direction. *Earth Surface Processes and Landforms 21*(2), 103–123.

Barlebo, H. C., M. C. Hill, and D. Rosbjerg (2004). Investigating the Macrodispersion Experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model. *Water Resources Research 40*(4), W04211.

Bayer, P. (2000). Aquifer-analog-studie in grobklastischen braided riverablagerungen: Sedimentäre/hydrogeologische wandkartierung und kalibrierung von georadarmessungen. *Diplomkartierung, Universitäet Tübingen*.

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

Bect, J., D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing 22*(3), 773–793.

Bianchi, M., C. Zheng, C. Wilson, G. R. Tick, G. Liu, and S. M. Gorelick (2011). Spatial connectivity in a highly heterogeneous aquifer: From cores to preferential flow paths. *Water Resources Research 47*(5), W05524.

Bliznyuk, N., D. Ruppert, C. Shoemaker, R. Regis, S. Wild, and P. Mugunthan (2008). Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation. *Journal of Computational and Graphical Statistics 17*(2), 270–294.

Boggs, J. M., L. M. Beard, W. R. Waldrop, T. B. Stauffer, W. G. MacIntyre, and C. P. Antworth (1993). Transport of tritium and four organic compounds during a natural-gradient experiment (MADE-2). Technical report, Electric Power Research Inst., Palo Alto, CA (United States).

Bohling, G. C., G. Liu, S. J. Knobbe, E. C. Reboulet, D. W. Hyndman, P. Dietrich, and J. J. Butler (2012). Geostatistical analysis of centimeter-scale hydraulic conductivity variations at the MADE site. *Water Resources Research 48*(2), W02525.

Boisvert, J. B., M. J. Pyrcz, and C. V. Deutsch (2007). Multiple-point statistics for training image selection. *Natural Resources Research 16*(4), 313–321.

Boisvert, J. B., M. J. Pyrcz, and C. V. Deutsch (2010). Multiple point metrics to assess categorical variable models. *Natural Resources Research 19*(3), 165–175.

Boucher, A. (2009). Considering complex training images with search tree partitioning. *Computers & Geosciences 35*(6), 1151–1158.

Bowling, J. C., A. B. Rodriguez, D. L. Harry, and C. Zheng (2005). Delineating alluvial aquifer heterogeneity using resistivity and GPR data. *Groundwater 43*(6), 890–903.

Brasington, J., D. Vericat, and I. Rychkov (2012). Modeling river bed morphology, roughness, and surface sedimentology using high resolution terrestrial laser scanning. *Water Resources Research 48*(11), W11519.

Bridge, J. S., J. Alexander, R. E. Collier, R. L. Gawthorpe, and J. Jarvis (1995). Ground-penetrating radar and coring used to study the large-scale structure of point-bar deposits in three dimensions. *Sedimentology 42*(6), 839–852.

Brierley, G. J. (1989). River planform facies models: the sedimentology of braided, wandering and meandering reaches of the Squamish River, British Columbia. *Sedimentary Geology 61*(1), 17–35.

Brunner, P., P. G. Cook, and C. T. Simmons (2009). Hydrogeologic controls on disconnection between surface water and groundwater. *Water Resources Research 45*(1), W01422.

Caers, J. (2011). *Modeling Uncertainty in the Earth Sciences*. John Wiley & Sons.

Caers, J. and X. Ma (2002). Modeling conditional distributions of facies from seismic using neural nets. *Mathematical Geology 34*(2), 143–167.

Caers, J., K. Park, and C. Scheidt (2010). Modeling uncertainty of complex earth systems in metric space. In *Handbook of Geomathematics*, pp. 865–889. Springer.

Cao, S. and S. Greenhalgh (1994). Finite-difference solution of the eikonal equation using an efficient, first-arrival, wavefront tracking scheme. *Geophysics 59*(4), 632–643.

Cardiff, M. and P. Kitanidis (2009). Bayesian inversion for facies detection: An extensible level set framework. *Water Resources Research 45*(10), W10416.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten (2005). Inverse problem in hydrogeology. *Hydrogeology Journal 13*(1), 206–222.

Caruso, B. S. (2006). Project river recovery: restoration of braided gravel-bed river habitat in New Zealand's high country. *Environmental Management 37*(6), 840–861.

Christie, M., A. Cliffe, P. Dawid, and S. S. Senn (2011). *Simplicity, complexity and modelling*, Volume 114. John Wiley & Sons.

Chugunova, T. and L. Hu (2008). Multiple-point simulations constrained by continuous auxiliary data. *Mathematical Geosciences 40*(2), 133–146.

Clarke, S. J., L. Bruce-Burgess, and G. Wharton (2003). Linking form and function: towards an eco-hydromorphic approach to sustainable river restoration. *Aquatic Conservation: Marine and Freshwater Ecosystems 13*(5), 439–450.

Coléou, T., M. Poupon, and K. Azbel (2003). Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *The Leading Edge 22*(10), 942–953.

Comunian, A. and P. Renard (2009). Introducing wwhypda: a world-wide collaborative hydrogeological parameters database. *Hydrogeology Journal 17*(2), 481–489.

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

Comunian, A., P. Renard, J. Straubhaar, and P. Bayer (2011). Three-dimensional high resolution fluvio-glacial aquifer analog–Part 2: Geostatistical modeling. *Journal of Hydrology 405*(1), 10–23.

Constable, S. C., R. L. Parker, and C. G. Constable (1987). Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics 52*(3), 289–300.

Cordua, K. S., T. M. Hansen, and K. Mosegaard (2012). Monte carlo full-waveform inversion of crosshole gpr data using multiple-point geostatistical a priori information. *Geophysics 77*(2), H19–H31.

Cornaton, F. J. (2007). Ground water: a 3-D ground water and surface water flow, mass transport and heat transfer finite element simulator, reference manual. *University of Neuchâtel, Neuchâtel, Switzerland*.

Coulthard, T. J., M. G. Macklin, and M. J. Kirkby (2002). A cellular model of Holocene upland river basin and alluvial fan evolution. *Earth Surface Processes and Landforms 27*(3), 269–288.

Cui, T., C. Fox, and M. O'Sullivan (2011). Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. *Water Resources Research 47*(10), W10521.

Dafflon, B. and W. Barrash (2012). Three-dimensional stochastic estimation of porosity distribution: Benefits of using ground-penetrating radar velocity tomograms in simulated-annealing-based or Bayesian sequential simulation approaches. *Water Resources Research 48*(5), W05553.

Dargahi, B. (2004). Three-dimensional flow modelling and sediment transport in the River Klarälven. *Earth Surface Processes and Landforms 29*(7), 821–852.

Davis, J. and A. Annan (1989). Ground-penetrating radar for high-resolution mapping of soil and rock stratigraphy. *Geophysical Prospecting 37*(5), 531–551.

Davy, P. and D. Lague (2009). Fluvial erosion/transport equation of landscape evolution models revisited. *Journal of Geophysical Research: Earth Surface (2003–2012) 114*(F3).

Dawson, M. (1988). Sediment size variation in a braided reach of the Sunwapta River, Alberta, Canada. *Earth Surface Processes and Landforms 13*(7), 599–618.

Day-Lewis, F. and J. Lane (2004). Assessing the resolution-dependent utility of tomograms for geostatistics. *Geophysical Research Letters 31*(7).

Day-Lewis, F. D., K. Singha, and A. M. Binley (2005). Applying petrophysical models to radar travel time and electrical resistivity tomograms: Resolution-dependent limitations. *Journal of Geophysical Research: Solid Earth (1978–2012) 110*(B8).

De Marsily, G., A. Buoro, et al. (1999). 40 years of inverse problems in hydrogeology. *Comptes Rendus de l'Academie des Sciences Series IIA Earth and Planetary Science 329*(2), 73–87.

De Marsily, G., F. Delay, J. Gonçalvès, P. Renard, V. Teles, and S. Violette (2005). Dealing with spatial heterogeneity. *Hydrogeology Journal 13*(1), 161–183.

de Matos, M. C., P. L. Osorio, and P. R. Johann (2006). Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. *Geophysics 72*(1), P9–P21.

De Serres, B., A. G. Roy, P. M. Biron, and J. L. Best (1999). Three-dimensional structure of flow at a confluence of river channels with discordant beds. *Geomorphology 26*(4), 313–335.

de Vries, L. M., J. Carrera, O. Falivene, O. Gratacós, and L. Slooten (2009). Application of multiple point geostatistics to non-stationary images. *Mathematical Geosciences 41*(1), 29–42.

dell'Arciprete, D., R. Bersezio, F. Felletti, M. Giudici, A. Comunian, and P. Renard (2012). Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments. *Hydrogeology Journal 20*(2), 299–311.

Dimitrakopoulos, R., H. Mustapha, and E. Gloaguen (2010). High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena. *Mathematical Geosciences 42*(1), 65–99.

Dogan, M., R. L. Van Dam, G. C. Bohling, J. J. Butler, and D. W. Hyndman (2011). Hydrostratigraphic analysis of the MADE site with full-resolution GPR and direct-push hydraulic profiling. *Geophysical Research Letters 38*(6), L06405.

Doherty, J. and S. Christensen (2011). Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. *Water Resources Research 47*(12), W12534.

Doser, D. I., K. D. Crain, M. R. Baker, V. Kreinovich, and M. C. Gerstenberger (1998). Estimating uncertainties for geophysical tomography. *Reliable Computing 4*(3), 241–268.

Eaton, T. T. (2006). On the importance of geological heterogeneity for flow simulation. *Sedimentary Geology 184*(3), 187–201.

Edwards, P., J. Kollmann, A. Gurnell, G. Petts, K. Tockner, and J. Ward (1999). A conceptual model of vegetation dynamics on gravel bars of a large Alpine river. *Wetlands Ecology and Management 7*(3), 141–153.

Emery, X. and C. Lantuéjoul (2006). Tbsim: A computer program for conditional simulation of three-dimensional gaussian random fields via the turning bands method. *Computers & Geosciences 32*(10), 1615–1628.

Emery, X. and J. M. Ortiz (2011). A comparison of random field models beyond bivariate distributions. *Mathematical Geosciences 43*(2), 183–202.

Eppstein, M. J. and D. E. Dougherty (1998). Efficient three-dimensional data inversion: Soil characterization and moisture Monitoring from cross-well ground-penetrating radar at a Vermont Test Site. *Water Resources Research 34*(8), 1889–1900.

Falivene, O., P. Arbus, A. Gardiner, G. Pickup, J. A. Muoz, and L. Cabrera (2006). Best practice stochastic facies modeling from a channel-fill turbidite sandstone analog (the Quarry outcrop, Eocene Ainsa basin, northeast Spain). *AAPG bulletin 90*(7), 1003–1029.

Feehley, C. E., C. Zheng, and F. J. Molz (2000). A dual-domain mass transfer approach for modeling solute transport in heterogeneous aquifers: Application to the Macrodispersion Experiment (MADE) site. *Water Resources Research 36*(9), 2501–2515.

Felletti, F., R. Bersezio, and M. Giudici (2006). Geostatistical simulation and numerical upscaling, to model ground-water flow in a sandy-gravel, braided river, aquifer analogue. *Journal of Sedimentary Research 76*(11), 1215–1229.

FOEN (2009, December). Groundwater in switzerland.

Foufoula-Georgiou, E. and V. Sapozhnikov (2001). Scale invariances in the morphology and evolution of braided rivers. *Mathematical Geology 33*(3), 273–291.

Fredsøe, J. (1978). Meandering and braiding of rivers. *Journal of Fluid Mechanics 84*(04), 609–624.

Fu, J. and J. J. Gómez-Hernández (2009). A blocking markov chain monte carlo method for inverse stochastic hydrogeological modeling. *Mathematical Geosciences 41*(2), 105–128.

Germanoski, D. and S. Schumm (1993). Changes in braided river morphology resulting from aggradation and degradation. *The Journal of Geology 101*(4), 451–466.

Ginsbourger, D., R. Le Riche, and L. Carraro (2010). *Computational Intelligence in Expensive Optimization Problems*, Chapter "Kriging is well-suited to parallelize optimization", pp. 1867–4534. Studies in Evolutionary Learning and Optimization. Springer-Verlag.

Glenz, D. (2013). *Inverse modeling of groundwater flow in the Rhône alluvial aquifer - Impact of the Third Rhône correction*. Ph. D. thesis, University of Neuchâtel, Switzerland.

Goff, J. R. and P. Ashmore (1994). Gravel transport and morphological change in braided Sunwapta River, Alberta, Canada. *Earth Surface Processes and Landforms 19*(3), 195–212.

Gómez-Hernández, J. and X. Wen (1998). To be or not to be multi-gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources 21*(1), 47–61.

Guan, J., F. J. Molz, Q. Zhou, H. H. Liu, and C. Zheng (2008). Behavior of the mass transfer coefficient during the MADE-2 experiment: New insights. *Water Resources Research 44*(2), W02423.

Guardiano, F. and R. Srivastava (1993). Multivariate geostatistics: beyond bivariate moments. In *Geostatistics Troia 1992*, pp. 133–144. Springer Netherlands.

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

Hendricks Franssen, H., A. Alcolea, M. Riva, M. Bakr, N. Van der Wiel, F. Stauffer, and A. Guadagnini (2009). A comparison of seven methods for the inverse modelling of groundwater flow. application to the characterisation of well catchments. *Advances in Water Resources 32*(6), 851–872.

Howard, A. D., M. E. Keetch, and C. L. Vincent (1970). Topological and geometrical properties of braided streams. *Water Resources Research 6*(6), 1674–1688.

Hu, L. and T. Chugunova (2008). Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. *Water Resources Research 44*(11), W11413.

Hubbard, S. S., J. Chen, J. Peterson, E. L. Majer, K. H. Williams, D. J. Swift, B. Mailloux, and Y. Rubin (2001). Hydrogeological characterization of the South Oyster Bacterial Transport Site using geophysical data. *Water Resources Research 37*(10), 2431–2456.

Hubbard, S. S. and Y. Rubin (2000). Hydrogeological parameter estimation using geophysical data: a review of selected techniques. *Journal of Contaminant Hydrology 45*(1), 3–34.

Huber, E. and P. Huggenberger (2015). Linking geomorphology and subsurface characterization of a gravel-braided river reach: Tagliamento River (NE Italy). *Sedimentary Geology* (Submitted).

Huber, E., P. Huggenberger, and J. Caers (2015). Geophysical stereology: Quantifying the uncertainty on 3D subsurface structure from 2D GPR data – Application to the Tagliamento River. *Journal of Applied Geophysics* (In preparation).

Huggenberger, P. and C. Regli (2006). A sedimentological model to characterize braided river deposits for hydrogeological applications. *Braided rivers: Process, deposits, ecology, and management. Blackwell Publ., Oxford, UK. A sedimentological model to characterize braided river deposits for hydrogeological applications*, 51–74.

Hundey, E. and P. Ashmore (2009). Length scale of braided river morphology. *Water Resources Research 45*(8), W08409.

Jerolmack, D. J. and C. Paola (2007). Complexity in a cellular model of river avulsion. *Geomorphology 91*(3), 259–270.

Jones, L. and S. Schumm (2009). Causes of avulsion: an overview. *Fluvial Sedimentology VI*, 171–178.

Journel, A. G. (1974). Geostatistics for conditional simulation of ore bodies. *Economic Geology 69*(5), 673–687.

Journel, A. G. and D. Posa (1990). Characteristic behavior and order relations for indicator variograms. *Mathematical Geology 22*(8), 1011–1025.

Jussel, P., F. Stauffer, and T. Dracos (1994). Transport modeling in heterogeneous aquifers: 1. statistical description and numerical generation of gravel deposits. *Water Resources Research 30*(6), 1803–1817.

Kaipio, J. and E. Somersalo (2005). *Statistical and computational inverse problems*, Volume 160. Springer.

Kaipio, J. and E. Somersalo (2007). Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics 198*(2), 493–504.

Käser, D., P. Brunner, T. Graf, F. Cochand, R. McLaren, and R. Therrien (2014). Channel Representation in Physically Based Models Coupling Groundwater and Surface Water: Pitfalls and How to Avoid Them. *Groundwater 52*, 827–836.

Keating, E. H., J. Doherty, J. A. Vrugt, and Q. Kang (2010). Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. *Water Resources Research 46*(10), W10517.

Kennedy, M. C. and A. O'Hagan (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika 87*(1), 1–13.

Kitanidis, P. (1997). *Introduction to geostatistics: applications in hydrogeology*. Cambridge University Press.

Kleinhans, M. G. and W. B. T. Brinke (2001). Accuracy of cross-channel sampled sediment transport in large sand-gravel-bed rivers. *Journal of Hydraulic Engineering 127*(4), 258–269.

Klingbeil, R., S. Kleineidam, U. Asprion, T. Aigner, and G. Teutsch (1999). Relating lithofacies to hydrofacies: outcrop-based hydrogeological characterisation of Quaternary gravel deposits. *Sedimentary Geology 129*(3), 299–310.

Klise, K. A., G. S. Weissmann, S. A. McKenna, E. M. Nichols, J. D. Frechette, T. F. Wawrzyniec, and V. C. Tidwell (2009). Exploring solute transport and streamline connectivity using lidar-based outcrop images and geostatistical representations of heterogeneity. *Water Resources Research 45*(5), W05413.

Klotzsche, A., J. van der Kruk, G. Meles, and H. Vereecken (2012). Crosshole gpr full-waveform inversion of waveguides acting as preferential flow paths within aquifer systems. *Geophysics 77*(4), H57–H62.

Koltermann, C. E. and S. M. Gorelick (1996). Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research 32*(9), 2617–2658.

Kostic, B., A. Becht, and T. Aigner (2005). 3-d sedimentary architecture of a quaternary gravel delta (sw-germany): Implications for hydrostratigraphy. *Sedimentary Geology 181*(3), 147–171.

Krishnan, S. and A. Journel (2003). Spatial connectivity: from variograms to multiple-point measures. *Mathematical Geology 35*(8), 915–925.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 79–86.

Künze, R. and I. Lunati (2011). A matlab toolbox to simulate flow through porous media. Technical report, University of Lausanne, Switzerland.

Künze, R. and I. Lunati (2012). An adaptive multiscale method for density-driven instabilities. *Journal of Computational Physics 231*(17), 5557–5570.

Labourdette, R. and R. R. Jones (2007). Characterization of fluvial architectural elements using a three-dimensional outcrop data set: Escanilla braided system, South-Central Pyrenees, Spain. *Geosphere 3*(6), 422–434.

Lane, S. (2000). The measurement of river channel morphology using digital photogrammetry. *The Photogrammetric Record 16*(96), 937–961.

Lane, S. (2009). Approaching the system-scale understanding of braided river behaviour. In I. A. of Sedimentologists Special Publication (Ed.), *Braided Rivers: Process, Deposits, Ecology and Management*, Volume 36, pp. 107–135. Blackwell Publishing Ltd.

Lane, S., R. Westaway, and D. Hicks (2003, March). Estimation of erosion and deposition volumes in a large, gravel-bed, braided river using synoptic remote sensing. *Earth Surface Processes and Landforms 28*(3), 249–271.

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

Lefebvre, S. and H. Hoppe (2006). Appearance-space texture synthesis. In *ACM Transactions on Graphics (TOG)*, Volume 25, pp. 541–548. ACM.

Leopold, L. B., M. G. Wolman, M. G. Wolman, and M. G. Wolman (1957). *River channel patterns: braided, meandering, and straight*. US Government Printing Office Washington (DC).

Linde, N., A. Binley, A. Tryggvason, L. B. Pedersen, and A. Revil (2006). Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltime data. *Water Resources Research 42*(12), W12404.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer.

Llopis-Albert, C. and J. E. Capilla (2009). Gradual conditioning of non-Gaussian transmissivity fields to flow and mass transport data: 3. Application to the Macrodispersion Experiment (MADE-2) site, on Columbus Air Force Base in Mississippi (USA). *Journal of Hydrology 371*(1), 75–84.

Lochbühler, T., J. Vrugt, and N. Linde (2015). Summary statistics from training images as model constraints in probabilistic inversion. *Geophysical Journal International* (Accepted).

Lødøen, O. P. and H. Tjelmeland (2010). Bayesian calibration of hydrocarbon reservoir models using an approximate reservoir simulator in the prior specification. *Statistical Modelling 10*(1), 89–111.

Lunt, I. and J. Bridge (2004). Evolution and deposits of a gravelly braid bar, Sagavanirktok River, Alaska. *Sedimentology 51*(3), 415–432.

Macklin, M. and B. Rumsby (2007). Changing climate and extreme floods in the British uplands. *Transactions of the Institute of British Geographers 32*(2), 168–186.

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

Mariethoz, G. (2010). A general parallelization strategy for random path based geostatistical simulation methods. *Computers & Geosciences 36*(7), 953–958.

Mariethoz, G., A. Comunian, I. Irarrazaval, and P. Renard (2012). A new angle: meandering river modeling using Direct Sampling. In *34th International Geological Congress, Brisbane, Australia 5-10 August 2012*.

Mariethoz, G., A. Comunian, I. Irarrazaval, and P. Renard (2014). Analog-based meandering channel simulation. *Water Resources Research 50*, 836–854.

Mariethoz, G. and B. Kelly (2011). Modeling complex geological structures with elementary training images and transform-invariant distances. *Water Resources Research 47*(7), W07527.

Mariethoz, G., M. McCabe, and P. Renard (2012). Spatiotemporal reconstruction of gaps in multivariate fields using the direct sampling approach. *Water Resources Research 48*(10), W10507.

Mariethoz, G., P. Renard, and J. Straubhaar (2010). The Direct Sampling method to perform multiple-point geostatistical simulations. *Water Resources Research 46*, W11536.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology 58*(8), 1246–1266.

Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability 5*(3), 439–468.

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

Meerschman, E. and M. Van Meirvenne (2011). Using bivariate multiple-point geostatistics and proximal soil sensor data to map fossil ice-wedge polygons. In *Pedometrics 2011: Innovations in pedometrics*, pp. 51–51. Czech University of Life Sciences Prague.

Meles, G., S. Greenhalgh, J. v. der Kruk, A. Green, and H. Maurer (2012). Taming the non-linearity problem in gpr full-waveform inversion for high contrast media. *Journal of Applied Geophysics 78*, 31–43.

Menke, W. (1989). *Geophysical Data Analysis: Discrete Inverse Theory*, Volume 45 of *International Geophysics*. Academic Press.

Miall, A. (1977). A review of the braided-river depositional environment. *Earth-Science Reviews 13*(1), 1 – 62.

Miall, A. D. (1985). Architectural-element analysis: a new method of facies analysis applied to fluvial deposits. *Earth-Science Reviews 22*(4), 261–308.

Middelkoop, H., K. Daamen, D. Gellens, W. Grabs, J. C. Kwadijk, H. Lang, B. W. Parmet, B. Schädler, J. Schulla, and K. Wilke (2001). Impact of climate change on hydrological regimes and water resources management in the Rhine basin. *Climatic Change 49*(1-2), 105–128.

Millar, R. G. (2005). Theoretical regime equations for mobile gravel-bed rivers with stable banks. *Geomorphology 64*(3), 207–220.

Mirowski, P. W., D. M. Tetzlaff, R. C. Davies, D. S. McCormick, N. Williams, and C. Signer (2009). Stationarity scores on training images for multipoint geostatistics. *Mathematical Geosciences 41*(4), 447–474.

Mockus, J. (1989). *Bayesian Approach to Gobal Optimization*, Volume 37. Kluwer Academic Pub.

Moeck, C. (2014). *Evaluating the effect of climate change on groundwater resources: from local to catchment scale*. Ph. D. thesis, University of Neuchâtel, Switzerland.

Mondal, A., Y. Efendiev, B. Mallick, and A. Datta-Gupta (2010). Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump Markov chain Monte Carlo methods. *Advances in Water Resources 33*(3), 241–256.

Mosegaard, K. and A. Tarantola (1995). Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth (1978–2012) 100*(B7), 12431–12447.

Moysey, S., J. Caers, R. Knight, and R. Allen-King (2003). Stochastic estimation of facies using ground penetrating radar data. *Stochastic Environmental Research and Risk Assessment 17*(5), 306–318.

Moysey, S., R. J. Knight, and H. M. Jol (2006). Texture-based classification of ground-penetrating radar images. *Geophysics 71*(6), K111–K118.

Moysey, S., K. Singha, and R. Knight (2005). A framework for inferring field-scale rock physics relationships through numerical simulation. *Geophysical Research Letters 32*(8).

Mugunthan, P. and C. A. Shoemaker (2006). Assessing the impacts of parameter uncertainty for computationally expensive groundwater models. *Water Resources Research 42*(10), W10428.

Murray, A. and C. Paola (1994, September). A cellular-model of braided rivers. *Nature 371*(6492), 54–57.

Murray, A. and C. Paola (2003). Modelling the effect of vegetation on channel pattern in bedload rivers. *Earth Surface Processes and Landforms 28*(2), 131–143.

Nicholas, A. P., R. Thomas, and T. A. Quine (2009). Cellular modelling of braided river form and process. In I. A. of Sedimentologists Special Publication (Ed.), *Braided Rivers: Process, Deposits, Ecology and Management*, Volume 36, pp. 137–151. Blackwell Publishing Ltd.

Olea, R. A. (2004). CORRELATOR 5.2a program for interactive lithostratigraphic correlation of wireline logs. *Computers & Geosciences 30*(6), 561–567.

Oliver, D. S. and Y. Chen (2011). Recent progress on reservoir history matching: a review. *Computational Geosciences 15*(1), 185–221.

Oliver, D. S., L. B. Cunha, and A. C. Reynolds (1997). Markov chain monte carlo methods for conditioning a permeability field to pressure data. *Mathematical Geology 29*(1), 61–91.

Oriani, F., J. Straubhaar, P. Renard, and G. Mariethoz (2014). Simulation of rainfall time-series from different climatic regions using the Direct Sampling technique. *Hydrology and Earth System Sciences Discussions 11*(3), 3213–3247.

Ory, J. and R. Pratt (1995). Are our parameter estimators biased? the significance of finite-difference regularization operators. *Inverse Problems 11*(2), 397.

Paciorek, C. J. (2003). *Nonstationary Gaussian processes for regression and spatial modelling*. Ph. D. thesis, Citeseer.

Paige, C. C. and M. A. Saunders (1982). Lsqr: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS) 8*(1), 43–71.

Pérez, C., G. Mariethoz, and J. M. Ortiz (2014). Verifying the high-order consistency of training images with data for multiple-point geostatistics. *Computers & Geosciences 70*, 190–205.

Peter, A. (2009). The restoration of Swiss watercourses is urgently necessary.

Peterson, Jr, J. E. (2001). Pre-inversion corrections and analysis of radar tomographic data. *Journal of Environmental & Engineering Geophysics 6*(1), 1–18.

Peyré, G. and L. D. Cohen (2006). Geodesic remeshing using front propagation. *International Journal of Computer Vision 69*(1), 145–156.

Pirot, G., J. Straubhaar, and P. Renard (2014). Simulation of braided river elevation model time series with multiple-point statistics. *Geomorphology 214*, 148–156.

Pirot, G., J. Straubhaar, and P. Renard (2015). A pseudo genetic model of coarse braided-river deposit. *Water Resources Research* (in preparation).

Plug, L. J. and B. Werner (2002). Nonlinear dynamics of ice-wedge networks and resulting sensitivity to severe cooling events. *Nature 417*(6892), 929–933.

Podvin, P. and I. Lecomte (1991). Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools. *Geophysical Journal International 105*(1), 271–284.

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

Ramanathan, R., A. Guin, R. W. Ritzi, D. F. Dominic, V. L. Freedman, T. D. Scheibe, and I. A. Lunt (2010). Simulating the heterogeneity in braided channel belt deposits: 1. a geometric-based methodology and code. *Water Resources Research 46*(4), W04515.

Rankey, E. and J. Mitchell (2003). That's why it's called interpretation: Impact of horizon uncertainty on seismic attribute analysis. *The Leading Edge 22*(9), 820–828.

Rasmussen, C. E. (2006). Gaussian Processes for Machine Learning.

Rauber, M., F. Stauffer, P. Huggenberger, and T. Dracos (1998). A numerical three-dimensional conditioned/unconditioned stochastic facies type model applied to a remediation well system. *Water Resources Research 34*(9), 2225–2233.

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

Renard, P. (2007). Stochastic hydrogeology: what professionals really need? *Groundwater 45*(5), 531–541.

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

Renard, P. and G. De Marsily (1997). Calculating equivalent permeability: a review. *Advances in Water Resources 20*(5), 253–278.

Renard, P., J. Straubhaar, J. Caers, and G. Mariethoz (2011). Conditioning facies simulations with connectivity data. *Mathematical Geosciences 43*(8), 879–903.

Richards, K., J. Brasington, and F. Hughes (2002). Geomorphic dynamics of floodplains: ecological implications and a potential modelling strategy. *Freshwater Biology 47*(4), 559–579.

Richardson, J., J. Sangree, R. Sneider, et al. (1987). Applications of geophysics to geologic models and to reservoir description. *Journal of petroleum technology 39*(07), 753–755.

Rodgers, P., C. Soulsby, J. Petry, I. Malcolm, C. Gibbins, and S. Dunn (2004). Groundwater–surface-water interactions in a braided river: a tracer-based assessment. *Hydrological Processes 18*(7), 1315–1332.

Rongier, G., P. Collon-Drouaillet, P. Renard, J. Straubhaar, and J. Sausse (2013). Reproduction assessment of connected geobodies in multiple-point simulation. In *GOCAD Meeting Abstracts*, Volume 1, pp. 0561.

Roth, K., R. Schulin, H. Flühler, and W. Attinger (1990). Calibration of time domain reflectometry for water content measurement using a composite dielectric approach. *Water Resources Research 26*(10), 2267–2273.

Rust, B. R. (1972). Structure and process in a braided river. *Sedimentology 18*(3-4), 221–245.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989). Design and analysis of computer experiments. *Statistical science*, 409–423.

Salamon, P., D. Fernàndez-Garcia, and J. Gómez-Hernández (2007). Modeling tracer transport at the MADE site: the importance of heterogeneity. *Water Resources Research 43*(8), W08404.

Sambridge, M. (1999). Geophysical inversion with a neighbourhood algorithmi. searching a parameter space. *Geophysical Journal International 138*(2), 479–494.

Sambrook Smith, G., J. Best, P. Ashworth, S. Lane, N. Parker, I. Lunt, R. Thomas, and C. Simpson (2010). Can we distinguish flood frequency and magnitude in the sedimentological record of rivers? *Geology 38*(7), 579–582.

Sampson, P. D. and P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association 87*(417), 108–119.

Sanchez-Vila, X., A. Guadagnini, and J. Carrera (2006). Representative hydraulic conductivities in saturated groundwater flow. *Reviews of Geophysics 44*(3), RG3002.

Santner, T. J., B. J. Williams, and W. Notz (2003). *The design and analysis of computer experiments*. Springer.

Sapozhnikov, V. and E. Foufoula-Georgiou (1996). Self-affinity in braided rivers. *Water Resources Research 32*(5), 1429–1439.

Sasena, M. J., P. Papalambros, and P. Goovaerts (2002). Exploration of metamodeling sampling criteria for constrained global optimization. *Engineering optimization 34*(3), 263–278.

Scheidt, C. and J. Caers (2009). Representing spatial uncertainty using distances and kernels. *Mathematical Geosciences 41*(4), 397–419.

Schuurman, F., W. A. Marra, and M. G. Kleinhans (2013). Physics-based modeling of large braided sand-bed rivers: Bar pattern formation, dynamics, and sensitivity. *Journal of Geophysical Research: Earth Surface 118*(4), 2509–2527.

Sear, D. (1994). River restoration and geomorphology. *Aquatic Conservation: Marine and Freshwater Ecosystems 4*(2), 169–177.

Sen, P., C. Scala, and M. Cohen (1981). A self-similar model for sedimentary rocks with application to the dielectric constant of fused glass beads. *Geophysics 46*(5), 781–795.

Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences 93*(4), 1591–1595.

Siirila-Woodburn, E. R. and R. M. Maxwell (2015). A heterogeneity model comparison of highly resolved statistically anisotropic aquifers. *Advances in Water Resources 75*, 53–66.

Singha, K., F. D. Day-Lewis, and S. Moysey (2007). Accounting for tomographic resolution in estimating hydrologic properties from geophysical data. *Subsurface hydrology: Data integration for properties and processes*, 227–241.

Singha, K. and S. Moysey (2006). Accounting for spatially variable resolution in electrical resistivity tomography through field-scale rock-physics relations. *Geophysics 71*(4), A25–A28.

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

Straubhaar, J., A. Walgenwitz, and P. Renard (2013). Parallel multiple-point statistics algorithm based on list and tree structures. *Mathematical Geosciences 45*(2), 131–147.

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

Strebelle, S. and A. Journel (2000). Sequential simulation drawing structures from training images.

Surian, N. (2002). Downstream variation in grain size along an Alpine river: analysis of controls and processes. *Geomorphology 43*(1), 137–149.

Suzuki, S. and J. Caers (2008). A distance-based prior model parameterization for constraining solutions of spatial inverse problems. *Mathematical Geosciences 40*(4), 445–469.

Suzuki, S., G. Caumon, and J. Caers (2008). Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. *Computational Geosciences 12*(1), 105–119.

Tan, X., P. Tahmasebi, and J. Caers (2014). Comparing Training-Image Based Algorithms Using an Analysis of Distance. *Mathematical Geosciences 46*(2), 149–169.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Siam.

Teles, V., J. P. Bravard, G. De Marsily, and E. Perrier (2001). Modelling of the construction of the Rhone alluvial plain since 15 000 years BP. *Sedimentology 48*(6), 1209–1224.

Teles, V., F. Delay, and G. De Marsily (2004). Comparison of genesis and geostatistical methods for characterizing the heterogeneity of alluvial media: Groundwater flow and transport simulations. *Journal of Hydrology 294*(1), 103–121.

Thomas, R. and A. P. Nicholas (2002). Simulation of braided river flow using a new cellular routing scheme. *Geomorphology 43*(3), 179–195.

Thomas, R., A. P. Nicholas, and T. A. Quine (2007). Cellular modelling as a tool for interpreting historic braided river evolution. *Geomorphology 90*(3), 302–317.

Tinga, W. R., W. Voss, and D. Blossey (1973). Generalized approach to multiphase dielectric mixture theory. *Journal of Applied Physics 44*(9), 3897–3902.

Tockner, K., A. Paetzold, U. Karaus, C. Claret, and J. Zettel (2009). Ecology of braided rivers. In I. A. of Sedimentologists Special Publication (Ed.), *Braided Rivers: Process, Deposits, Ecology and Management*, Volume 36, pp. 339–359. Blackwell Publishing Ltd.

Tossavainen, O.-P., J. Percelay, M. Stacey, J. P. Kaipio, and A. Bayen (2011). State estimation and modeling error approach for 2-D shallow water equations and Lagrangian measurements. *Water Resources Research 47*(10), W10510.

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

Van De Lageweg, W., W. Van Dijk, and M. Kleinhans (2013). Morphological and stratigraphical signature of floods in a braided gravel-bed river revealed from flume experiments. *Journal of Sedimentary Research 83*(11), 1032–1045.

van der Nat, D., K. Tockner, P. Edwards, J. Ward, and A. Gurnell (2003). Habitat change in braided flood plains (Tagliamento, NE-Italy). *Freshwater Biology 48*(10), 1799–1812.

Vidale, J. (1988). Finite-difference calculation of travel times. *Bulletin of the Seismological Society of America 78*(6), 2062–2076.

Vrugt, J. A. and C. J. Ter Braak (2011). Dream (d): an adaptive markov chain monte carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences 15*(12), 3701–3713.

Warburton, J. and T. Davies (1994). Variability of bedload transport and channel morphology in a braided river hydraulic model. *Earth Surface Processes and Landforms 19*(5), 403–421.

Webb, E. K. (1994). Simulating the three-dimensional distribution of sediment units in braided-stream deposits. *Journal of Sedimentary Research 64*(2b), 219–231.

Webb, E. K. (1995). Simulation of braided channel topology and topography. *Water Resources Research 31*(10), 2603–2611.

Westaway, R., S. Lane, and D. Hicks (2003). Remote survey of large-scale braided, gravel-bed rivers using digital photogrammetry and image analysis. *International Journal of Remote Sensing 24*(4), 795–815.

Western, A. W., G. Blöschl, and R. B. Grayson (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research 37*(1), 83–97.

Williams, P. F. and B. R. Rust (1969). The sedimentology of a braided river. *Journal of Sedimentary Research 39*(2), 649–679.

Williamson, P. (1991). A guide to the limits of resolution imposed by scattering in ray tomography. *Geophysics 56*(2), 202–207.

Zhang, J., K. Zhou, L. Velho, B. Guo, and H.-Y. Shum (2003). Synthesis of progressively-variant textures on arbitrary surfaces. In *ACM Transactions on Graphics (TOG)*, Volume 22, pp. 295–302. ACM.

Zhang, T., P. Switzer, and A. Journel (2006). Filter-based classification of training image patterns for spatial simulation. *Mathematical Geology 38*(1), 63–80.

Zhang, Y., C. T. Green, and G. E. Fogg (2013). The impact of medium architecture of alluvial settings on non-Fickian transport. *Advances in Water Resources 54*, 78–99.

Zheng, C., M. Bianchi, and S. M. Gorelick (2011). Lessons learned from 25 years of research at the MADE site. *Groundwater 49*(5), 649–662.

Zhou, H., J. J. Gómez-Hernández, and L. Li (2012). A pattern-search-based inverse method. *Water Resources Research 48*(3), W03505.

Zinn, B. and C. F. Harvey (2003). When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. *Water Resources Research 39*(3), 1051.

Appendix A

Distance-based kriging relying on proxy simulations for inverse conditioning

Abstract

Let us consider a large set of candidate parameter fields, such as hydraulic conductivity maps, on which we can run an accurate forward flow and transport simulation. We address the issue of rapidly identifying a subset of candidates whose response best match a reference response curve. In order to keep the number of calls to the accurate flow simulator computationally tractable, a recent distance-based approach relying on fast proxy simulations is revisited, and turned into a non-stationary kriging method where the covariance kernel is obtained by combining a classical kernel with the proxy. Once the accurate simulator has been run for an initial subset of parameter fields and a kriging metamodel has been inferred, the predictive distributions of misfits for the remaining parameter fields can be used as a guide to select candidate parameter fields in a sequential way. The proposed algorithm, *Proxy-based Kriging for Sequential Inversion* (ProKSI), relies on a variant of the *Expected Improvement*, a popular criterion for kriging-based global optimization. A statistical benchmark of ProKSI's performances illustrates the efficiency and the robustness of the approach when using different kinds of proxies.

A.1 Introduction

Inverse techniques are one of the corner stones of groundwater modeling. Their aim is to identify model structure and model parameter values from observed state variables. In practice, a wide range of approaches exist and have been compared extensively (De Marsily et al., 1999; Kaipio and Somersalo, 2005; Carrera et al., 2005; Tarantola, 2005; Hendricks Franssen et al., 2009; Oliver and Chen, 2011). Often, the inverse problem is formulated in a least-square or maximum likelihood manner. A data misfit quantifies the difference between measured and calculated state variables. The aim is then to find a parameter field minimizing the misfit.

Less frequently in practice, the problem is solved in the Bayesian framework with the aim to recover an ensemble of representative samples (parameter fields) from the posterior probability distribution. This is particularly important when prior geological knowledge is available and can be expressed using geological models describing the parameter fields. Techniques such as multiple point statistics, object- or process-based geological simulations (De Marsily et al., 2005) are often used to express this prior knowledge, but then solving the inverse problem becomes very challenging because it is usually not possible to provide an explicit analytical expression of the posterior distribution. In such situations, one needs to rely on computational resources and statistical sampling techniques (Mosegaard and Tarantola, 1995; Tarantola, 2005; Caers, 2011; Vrugt and Ter Braak, 2011) such as Markov Chain Monte Carlo (MCMC) (Oliver et al., 1997; Liu, 2008; Fu and Gómez-Hernández, 2009; Alcolea and Renard, 2010; Mariethoz, 2010; Mondal et al., 2010; Cui et al., 2011). A practical difficulty in that approach is that evaluating the likelihood function, involving in itself a calculation of the misfit, is often computationally very demanding. This inhibits the user to let an MCMC procedure run for a sufficiently large number of iterations to enable convergence (Keating et al., 2010; Mariethoz, 2010). Similar computational issues arise in optimization problems related to groundwater management: if each evaluation of the objective function that has to be minimized requires a significant amount of computational resources, it becomes infeasible to reach the optimum in reasonable time.

To reduce the computational demand, one can use the concept of *metamodel* (or *response surface*). The response (e.g. the misfit) of the flow simulator is computed for a small set of inputs and can then be predicted by the metamodel for any other input. Various interpolation techniques can be employed such as radial basis functions, splines, or kriging (Bliznyuk et al., 2008; Mugunthan and Shoemaker, 2006; Sacks et al., 1989; Matheron, 1963; Rasmussen,

2006; Paciorek, 2003). An advantage of kriging is its ability to provide both a prediction of the possible response (kriging mean m) and a corresponding prediction uncertainty (kriging variance s^2). The prediction uncertainty drops to zero where the response has actually been computed with the simulator and increases when moving away from those input points. In the global optimization problem consisting in finding inputs minimizing the objective function, one can use m and s^2 to express a trade-off between the exploitation of the mean prediction (finding locations where m is low) and exploration of the design space (finding locations where the prediction is the most uncertain). This idea gave birth to the *Expected Improvement* (EI) criterion (Mockus, 1989): for every location within the input space, the kriging metamodel is used to derive a predictive distribution for the improvement that might be obtained by evaluating the objective function at that location. Here the term improvement refers to the difference between the best (i.e. the lowest) response observed so far and the response at the new location if this difference is positive, and 0 otherwise.

The input point with the highest EI is then chosen to run the numerical model again and update the metamodel. Such approaches based on kriging metamodels have been very successfully used for sequential design of computer experiments since the development of the *Efficient Global Optimization* algorithm (Sasena et al., 2002) in the late 1990's. Several other criteria were later proposed for neighbouring problems (Bect et al., 2012).

Another approach to reduce the computational demand is to use a concept of distance between parameter fields (Suzuki and Caers, 2008; Scheidt and Caers, 2009; Caers, 2011). Several types of distances can be defined, but the important point is that the distance should be chosen such that it can be computed rapidly and help predicting if two parameter fields will lead to similar or different responses. For example, Suzuki et al. (2008) used the Hausdorff distance to quantify the differences in the geometry of complex 3D models (having different fault systems, horizon geometries, etc.), coupled with the neighborhood algorithm (Sambridge, 1999) to search efficiently, within the prior ensemble, the models that match field observations of oil production. Scheidt and Caers (2009) propose a general framework based on the concept of distance to quantify uncertainty. In their example, the problem consists in estimating oil recovery in a production well. The models all have the same geometry, but very different parameter fields (obtained using multiple-point statistics with different training images). The prior ensemble is large and the aim is to rapidly obtain a good estimation of the uncertainty on the forecast. For that purpose, Scheidt and Caers (2009) define the square distance between two parameter fields as the integrated square difference between the responses computed for the two parameter fields with a fast streamline solver. The distances between every pair of parameter fields is computed and used as a base for mapping all the parameter fields in an abstract metric space in which it is possible to select a small number of them covering comprehensively the variability of the complete ensemble. Running the forward two-phase flow numerical simulator only on these selected geological models allows a fast and rather accurate uncertainty assessment. Going a step further, Caers et al. (2010) use the same framework to formulate the inverse problem.

A promising direction for reducing the computational demand is the joint use of a pair of complex and simple models (Kaipio and Somersalo, 2007; Kennedy and O'Hagan, 2000; Lødøen and Tjelmeland, 2010; Doherty and Christensen, 2011; Tossavainen et al., 2011). The distinction between the complex and simple models is not straightforward, but to remain general we can state that the complex model tends to account for all important and relevant physical processes as well as all the necessary geometrical complexity of the reservoir. On the opposite, the simple model neglects some aspects of this complexity with the aim of being much more computationally efficient. The simplification may be based on neglecting some physical processes, on reducing the problem dimension (2D instead of 3D), or on a coarse spatial or temporal resolution. In the remaining of this paper, we will use the terminology

accurate model for the complex one, and *proxy* for the simple one. To use a combination of accurate and proxy models in practice, one needs to establish a link between the two. Several approaches can be devised. For example, Doherty and Christensen (2011) identify some parameters of the proxy model by solving an inverse problem where the results of the accurate model have to be reproduced.

In this paper, we propose to link an accurate and a proxy model using a distance-based kriging metamodel. It allows to forecast outputs of the accurate model as it is done with traditional kriging metamodels. However, those methods are usually limited to parameter spaces of small dimensions. This makes their application for the identification of complete parameter fields impossible. The novelty of the proposed approach lies therefore in the way we define the covariance kernel at the core of the kriging metamodel. The concept is simple, we assume that the same parameter fields can be used as input data for the proxy and the accurate model. As suggested by Caers and his collaborators (Suzuki et al., 2008; Scheidt and Caers, 2009; Caers et al., 2010; Caers, 2011) we use a distance based on proxy responses, but we include that distance into the covariance kernel of the kriging equations. The consequence is a drastic reduction of the problem dimension, allowing to infer covariance parameters. Once the statistical relation between the proxy and the accurate model is established, it can be used to predict the misfit for any parameter field whose proxy response is known. It can also be updated when new runs of the accurate model become available. This general idea can be applied to a very wide range of problems.

One of the main aims of this paper is therefore to describe the concept of the distance-based kriging technique. We also illustrate how this technique can be used in a sequential algorithm aiming at quickly identifying a set of parameter fields whose responses computed with an accurate model match some reference data. In an inverse problem, the purpose is often not only be to find the global minimizer(s) but more to sample from a posterior distribution, and so we propose a variant of the EI criterion meant to spend more time exploring the possible various minima of the misfit function than EI. For illustration purpose, we consider a simple flow and solute transport problem. The geological heterogeneity is modeled using a multiple-point statistics technique (Mariethoz et al., 2010) allowing to account for prior geological knowledge typical for a fluvio glacial environment. Numerous experiments with a randomization procedure are conducted to test the robustness of the method.

The paper is organized as follows. In section A.2 we give an overview of the sequential algorithm used to solve the inverse problem. In section A.3, we describe in detail the proposed kriging metamodel. The equations of ordinary kriging are recalled, with a focus on the role of the covariance kernel. The original kernel underlying our work is introduced, followed by a discussion on its interpretation and mathematical foundations. Some practical details follow on the estimation of covariance parameters. We end the presentation of the method in section A.4 by describing how the sequential search is driven. Sections A.5 and A.6 are dedicated to results and discussion. We first introduce a case study to illustrate the methodology. Then we present the results of a randomized experiment and statistically assess the method's performances based on a benchmark of 100 reference curves. We conclude and propose a few theoretical and practical perspectives in section A.7.

A.2 Overview of the sequential algorithm

The proposed sequential algorithm is named *Proxy-based kriging for Sequential Inversion* (ProKSI). Its aim is to identify rapidly, within a large ensemble of parameter fields, the ones whose responses computed with the accurate model fit a given reference curve. In practice, the algorithm consists in sequentially selecting among all the available parameter fields which one will be used as input for the accurate numerical model at the next iteration (Fig. A.1

to A.2). Before sketching the key phases of the algorithm, let us set a few notations.

Each candidate parameter field is denoted $\mathbf{x}_i \in E$ $(1 \leq i \leq N)$, where E is a vector space, typically of dimension 10^4 to 10^6 when representing a discretization of the subsurface. In the following examples, $\mathbf{x}_i$ represents a categorical field obtained from multiple-point statistics simulation. But the proposed methodology is more general and can be applied without much modifications to models having various geometries or even based on different conceptual assumptions. The only requirement is that it is possible to compute the accurate and proxy responses for any of those input parameter fields.

The accurate numerical simulator is considered as a function f returning a vector of values. In the example, we assume more specifically that the simulator f returns for any parameter field $\mathbf{x} \in E$ a breakthrough curve $f_{\mathbf{x}}$ (concentration versus time):

$$f_{\mathbf{x}} : t \in [0, T] \rightarrow f_{\mathbf{x}}(t) \in [0, +\infty) \tag{A.1}$$

where t represents the time. The space of such curves is denoted by F.

Now, given a reference curve $f_{\text{ref}} \in F$, the goal is to recover in a limited time which $\mathbf{x}_i$'s $(1 \leq i \leq N)$ minimize the misfit $g^{\circ}(\mathbf{x}) := d(f_{\text{ref}}, f_{\mathbf{x}})$, where d is some metric on F. For example, if we use the L^2 norm, the misfit will be expressed as:

$$g^{\circ}(\mathbf{x}) = \int_0^T \left(f_{\text{ref}}(t) - f_{\mathbf{x}}(t)\right)^2 dt \tag{A.2}$$

Ideally, one wishes to describe the subset of input fields leading to a good fit, relying on a fixed number of evaluations $k < N$ dictated by computation time constraints. In addition to f, we assume that a "*proxy*" $p : E \rightarrow F$ is available, providing an approximate solution to the flow and transport equations significantly faster than f. p may stem for instance from an auxiliary simulator solving similar equations with simplified physics, or from degrading the accurate simulator f by reducing the time or spatial resolution.

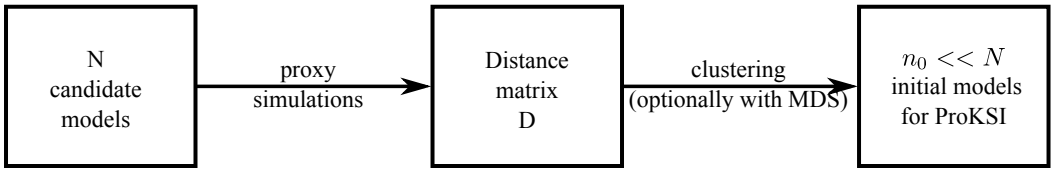

Figure A.1: Initialization steps of the ProKSI algorithm.

The ProKSI algorithm starts with a series of initialization steps (Fig. A.1):

1. A sample of parameter fields $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ is drawn from a chosen prior distribution (e.g., by multiple-points statistics simulation).

2. The proxy responses $p(\mathbf{x}_i, t)$ are computed for all $\mathbf{x}_i$'s $(1 \leq i \leq N)$. The distances $d_{i,j}$ between the proxy responses of any pair of parameter fields are then computed:

$$d_{i,j}^2 = \int_0^T \left(p(\mathbf{x}_i, t) - p(\mathbf{x}_j, t)\right)^2 dt \tag{A.3}$$

   This allows assembling a distance matrix D between all proxy responses.

3. A clustering technique (k-means) is used to group the parameter fields in n_0 classes. For each class, the parameter fields that are the closest to the centroïd are selected to get a subset $\mathbf{X}_{n_0} = \{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{n_0}}\}$ of n_0 initial models (See Fig. A.6). Multidimensional scaling (MDS) is optionally used to map all the input parameter fields in a small-dimensional Euclidean space (Fig. A.6).

For each of those n_0 selected parameter fields, the accurate response $f_{\mathbf{x}_{i_j}}$ is computed with the accurate numerical solver. We obtain a vector $\mathbf{g}^\circ = \{g^\circ_{i_1}, \ldots, g^\circ_{i_{n_0}}\}$ (where $g^\circ_{i_j} := g^\circ(\mathbf{x}_{i_j})$, $1 \leq j \leq n_0$) containing the misfits for the n_0 parameter fields.

The values of $\mathbf{g}^\circ$ are transformed to obtain a sample $\mathbf{g}$ with a close-to-Gaussian distribution. Such a Gaussian transformation is important later when the kriging variance is used to represent the prediction uncertainty for the misfit. Different techniques can be used for the transformation such as normal score transform, or Gaussian anamorphosis. Here, a power-law transform $g_{i_j} = \left[g^\circ_{i_j}\right]^\gamma$ (Box-Cox type) is used as it is simple to implement and robust even when a small number of samples is available, and the value of γ is obtained by minimizing the skewness of the sample of transformed values $\{g_{i_j}, 1 \leq j \leq n_0\}$.

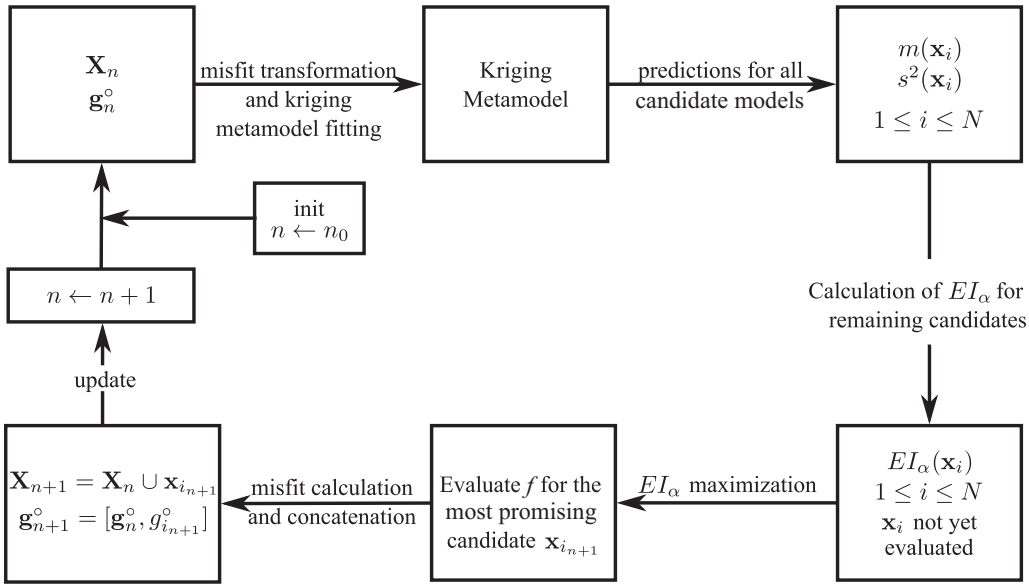

Figure A.2: Sequential loop of the ProKSI algorithm.

A sequential loop (Fig. A.2) then allows selecting a new parameter field at each iteration on which to run the accurate solver. This enables building progressively a set of parameter fields with low misfit values. The steps in that loop are the following (n is first set to n_0):

1. If not already done, apply a normalizing transform to the sample of misfits (See detail above). Estimate the covariance parameters τ, θ, and σ^2 as described in section A.3. Compute the kriging mean $m(\mathbf{x}_i)$ and the variance $s^2(\mathbf{x}_i)$ for all inputs $\mathbf{x}_i \notin \mathbf{X}_n$.

2. After having computed the value of the modified expected improvement criterion $EI_\alpha(\mathbf{x}_i)$ (see section A.4 for its definition) for all the remaining candidate models, Select a model with maximal EI_α value as next candidate, called $\mathbf{x}_{i_{n+1}}$.

3. Set $\mathbf{X}_{n+1} = \mathbf{X}_n \cup \{\mathbf{x}_{i_{n+1}}\}$. Compute $f_{\mathbf{x}_{i_{n+1}}}$ with the accurate numerical solver. Calculate the new corresponding misfit and append it to the vector of misfits: $\mathbf{g}^\circ_{n+1} = \{\mathbf{g}^\circ_n, g^\circ_{i_{n+1}}\}$. Go to step 1 and resume the search until a convergence criterion is met.

The algorithm stops when the EI_α reaches a prescribed lower threshold, or a desired number of evaluations has been done, for instance because the allocated search time is elapsed.

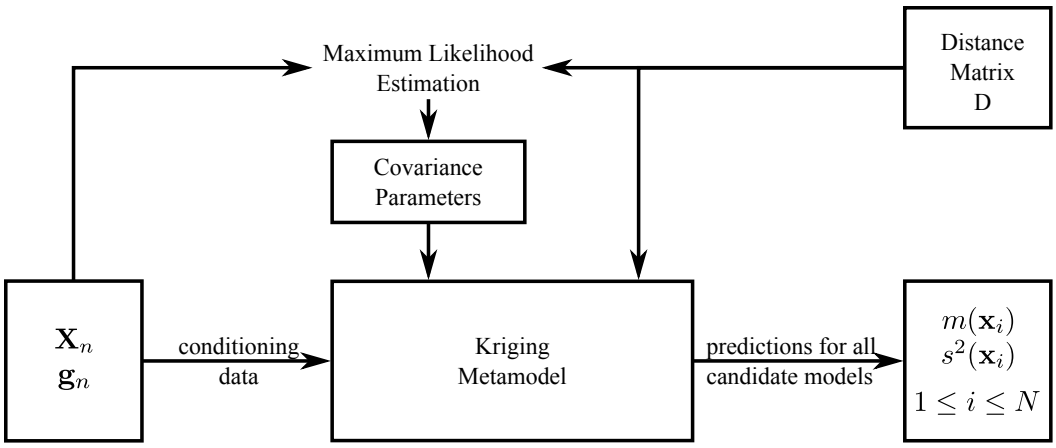

Figure A.3: Overview of the proxy-based kriging prediction workflow (after misfit transformation).

A.3 High-dimensional kriging with a proxy-based kernel

The most important difference between the existing methods and what we propose here is the distance-based kriging approach. It lies at the heart of sequential algorithm described earlier in Figure A.2. In this section, we will describe in detail how this step is performed. The main idea is to integrate the distance between proxy responses within the covariance kernel of the kriging metamodel (Fig. A.3).

A.3.1 kriging for Computer Experiments

We adopt the framework of Gaussian processes Rasmussen (2006) to model the transformed misfit between f_{ref} and the response of the accurate numerical simulator. The transformed misfit g is assumed to be one realization of a Gaussian process with high-dimensional index space $(G_{\mathbf{x}})_{\mathbf{x}\in E}$, with mean function μ and non-stationary covariance kernel k. We assume that μ is an unknown constant, as in the case of ordinary kriging. We denote by $\mathbf{g}$ the vector of known misfit values at the current design of experiments $\mathbf{X}_n := \{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_n}\}$ $(n \geq n_0)$. The kriging mean $m(\mathbf{x}) = \mathbb{E}\left[G_{\mathbf{x}} | G_{\mathbf{x}_{i_1}} = g(\mathbf{x}_{i_1}), \ldots, G_{\mathbf{x}_{i_n}} = g(\mathbf{x}_{i_n})\right]$ and kriging variance s^2 at any arbitrary point $\mathbf{x} \in E$ are written:

$$m(\mathbf{x}) = \widehat{\mu} + \mathbf{k}(\mathbf{x})^T K^{-1}(\mathbf{g} - \widehat{\mu}\mathbf{1}) \tag{A.4a}$$

$$s^2(\mathbf{x}) = k(\mathbf{x},\mathbf{x}) - \mathbf{k}(\mathbf{x})^T K^{-1}\mathbf{k}(\mathbf{x}) + \frac{(1 - \mathbf{k}(\mathbf{x})^T K^{-1}\mathbf{1})^2}{\mathbf{1}^T K^{-1}\mathbf{1}} \tag{A.4b}$$

where K is a $n \times n$ matrix with entries $K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$, referred to as the *covariance matrix of observations*, $\mathbf{k}(\mathbf{x}) := (k(\mathbf{x},\mathbf{x}_1), \ldots, k(\mathbf{x},\mathbf{x}_n))'$ is a $n \times 1$ *covariance vector*, and $\widehat{\mu} = \frac{\mathbf{1}^T K^{-1}\mathbf{g}}{\mathbf{1}^T K^{-1}\mathbf{1}}$ is the best linear unbiased estimator of μ.

One of the attracting features of kriging is that m interpolates the observations (i.e. $\forall j \in \{1, \ldots, n\}$, $m(\mathbf{x}_{i_j}) = g(\mathbf{x}_{i_j})$). Furthermore, s^2 vanishes at the design points ($s^2(\mathbf{x}_{i_j}) = 0$), and gives a quantification of the prediction uncertainty at unobserved points. A very important feature is that both properties remain valid whatever the chosen covariance kernel k. Hence, equations (A.4a) and (A.4b) give a potentially infinite set of interpolating metamodels, and selecting k appropriately for the studied phenomenon appears to be a crucial issue in practice.

A.3.2 A new kernel for high-dimensional kriging based on fast proxies

Designing a suiTab. covariance kernel over $E \times E$ is very challenging because E is a space of parameter fields of typical dimensions ranging between 10^4 to 10^6. Hence, taking kernels usually employed in d-dimensional ($d \approx 10$) cases, e.g., an anisotropic power exponential kernel, will a priori not make sense in the present framework. Alternatively, uncovering features of the parameter fields $\mathbf{x} \in E$ leading to similar response curves would be ideal.

Here, we take advantage of the proxy responses in order to define a relevant measure of similarity. More precisely, we propose to use a covariance kernel of the following form:

$$k(\mathbf{x}, \mathbf{y}) := \sigma^2 \exp\left(-\frac{1}{\theta^2}\int_0^T (p(\mathbf{x}, t) - p(\mathbf{y}, t))^2 dt\right) + \tau^2 \mathbf{1}_{\mathbf{x}=\mathbf{y}} \tag{A.5}$$

In words, the closer two proxy curves associated with two parameter fields $\mathbf{x}, \mathbf{y}$ are, the closer the fits to the reference are expected to be when running the accurate simulator with those inputs. In addition to this transformed Gaussian kernel, the term $\tau^2 \mathbf{1}_{\mathbf{x}=\mathbf{y}}$ stands for the nugget effect, and allows to model a possible dissimilarity between the accurate responses of the inputs $\mathbf{x}, \mathbf{y}$, even if their associated proxy responses are close or even identical.

In fact, the proposed covariance kernel k can be seen as a standard stationary Gaussian kernel over $F \times F$, chained with the "proxy operator", that is with the function p:

$$k(\mathbf{x}, \mathbf{y}) := \sigma^2 \exp\left(-\frac{1}{\theta^2}||p(\mathbf{x}) - p(\mathbf{y})||_F^2\right) + \tau^2 \mathbf{1}_{\mathbf{x}=\mathbf{y}} \tag{A.6}$$

where $||f||_F := \sqrt{\int_0^T f(t)^2 dt}$ ($f \in F$) stands for the L^2 norm over F (the functions of F being further assumed continuous). This basic fact ensures that the proposed kernel is an admissible covariance. k is indeed positive definite over $E \times E$ (but not necessarily strictly) in virtue of the following property, for which a proof is proposed in appendix:

Property Let E and F be two arbitrary spaces. Given a positive definite kernel k_F over $F \times F$, the kernel k_E defined by

$$k_E(\mathbf{x}, \mathbf{y}) := k_F(p(\mathbf{x}), p(\mathbf{y})) \tag{A.7}$$

is positive definite over $E \times E$ for any function $p : E \longrightarrow F$.

Note that in different contexts, similar methods relying on a change of variables within a positive definite kernel were already proposed, for example in Sampson and Guttorp (1992) and subsequent works. Coming back to Eq. (A.5), the basis kernel k_F corresponding to Prop. A.7 is none other than an isotropic Gaussian kernel $k_F(\mathbf{u}, \mathbf{v}) = \sigma^2 \exp\left(-\frac{1}{\theta^2}||\mathbf{u} - \mathbf{v}||_F^2\right)$, parametrized by a sill σ^2 and a range parameter $\theta > 0$. Note also that chaining the proxy operator with other kernels being positive definite in any dimension (e.g. kernels of the Matérn class) would be admissible too; a Matérn kernel may be preferred to a Gaussian one depending on the settings.

The next subsection focuses in detail on the chosen methodology for estimating the three parameters σ^2, θ, τ^2 from available data.

A.3.3 Parameter fitting for the proposed kriging model

Several methods can be considered for estimating σ^2, θ, and τ^2 based on available data. In particular, following the logic of Eq. (A.6), the problem boils down to a low-dimensional one thanks to the isotropy assumption relatively to the space F, and so usual variographic tools may well be applied in theory. However, most implemented variographic methods take

low-dimensional vectors as inputs, while only the norm of the increments is available here. Alternative automatic estimation methods include cross-validation error minimization, and Maximum Likelihood Estimation (MLE).

Here we chose to base covariance parameter estimation on MLE, while keeping an eye on variographic tools for assisting MLE with initial values or bounds for the parameters. In ordinary kriging settings, MLE consists in maximizing the likelihood of σ^2, θ, τ^2 given $\mathbf{g}$ under the assumption that $G_{\mathbf{X_n}} \sim \mathcal{N}(\widehat{\mu}\mathbf{1}, K)$, or equivalently (See, e.g., Santner et al. (2003)) in minimizing:

$$l(\sigma^2, \theta, \tau^2; \mathbf{g}) := \log(\det(K)) + (\mathbf{g} - \widehat{\mu}\mathbf{1})^T K^{-1} (\mathbf{g} - \widehat{\mu}\mathbf{1}), \tag{A.8}$$

where K and $\widehat{\mu}$ are functions of $(\sigma^2, \theta, \tau^2)$. Various global optimization procedures can be used for solving this non-convex optimization algorithm, ranging from Nelder and Mead's simplex to genetic algorithms using derivatives. Here we adopt a pragmatic one-at-a-time approach involving a concentration step on the variance parameter (described in appendix), for which convincing experimental results could be obtained as shown in Sections 5 and 6.

A.4 Sequential search driven by proxy-based kriging

For any candidate parameter field $\mathbf{x}_i$, the kriging metamodel of the previous section allows predicting the (transformed) misfit $g(\mathbf{x}_i)$ by $m(\mathbf{x}_i)$ with prediction variance $s^2(\mathbf{x}_i)$. Now, in a sequential procedure aiming at identifying the parameter fields with the lowest misfits such as considered here, m and s^2 can be used at any given iteration for selecting on which candidate parameter field to run the accurate numerical model next. For that purpose, we propose to use a variant of the *Expected Improvement* (EI) criterion, meant to spend more time exploring the basins of optima than the genuine EI.

By definition, EI is intended to point towards promising points, but also to foster space exploration. Hence, in EI algorithms like *EGO* Sasena et al. (2002), a typical behavior when evaluating the objective function at a good point (i.e. at a point becoming the current best) is to spend some additional iterations in its neighborhood, and then to get attracted by unexplored regions with higher kriging variances. This can be explained by coming back to EI's formal definition. Let us denote by $g(\mathbf{X}_n)$ the vector of misfit values after n accurate evaluations of f, and by $\min(g(\mathbf{X}_n))$ the minimum misfit value found so far. The aim is now to find a parameter field $\mathbf{x}$ such that the magnitude of the improvement of $g(\mathbf{x})$ with respect to $\min(g(\mathbf{X}_n))$ be the highest in expectation. Let us remind the reader that the (transformed) misfit is modelled here as a Gaussian Process $(G_{\mathbf{x}})_{\mathbf{x}\in E}$. For any $\mathbf{x} \in E$, the difference between the current minimum and the unknown value of the misfit, $\min(G_{\mathbf{X}_n}) - G_{\mathbf{x}}$, is then a random variable. Only positive values are usually taken into account when one is not interested in regions with worse misfit, and the improvement is therefore defined as $(\min(G_{\mathbf{X}_n}) - G_{\mathbf{x}})^+ := \max(\min(G_{\mathbf{X}_n}) - G_{\mathbf{x}}, 0)$. The EI criterion for a candidate parameter field $\mathbf{x}$ is then defined as the expectation of this improvement conditional on $G_{\mathbf{X}_n} = g(\mathbf{X}_n)$:

$$EI(\mathbf{x}) := \mathbb{E}\left[(\min(G_{\mathbf{X}_n}) - G_{\mathbf{x}})^+ \,|G_{\mathbf{X}_n} = g(\mathbf{X}_n)\right] \tag{A.9}$$

where conditioning on the event $G_{\mathbf{X}_n} = g(\mathbf{X}_n)$ turns $\min(G_{\mathbf{X}_n})$ into $\min(g(\mathbf{X}_n))$, and leads to the well-known Gaussian conditional distribution for $G_{\mathbf{x}}$:

$$\mathcal{L}(G_{\mathbf{x}}|G_{\mathbf{X}_n} = g(\mathbf{X}_n)) = \mathcal{N}(m(\mathbf{x}), s^2(\mathbf{x})) \tag{A.10}$$

Owing to this convenient property, the EI criterion offers the advantage of being analytically tracTab. (Sasena et al., 2002). Noting $T = \min(g(\mathbf{X}_n))$ and $f_{\mathcal{N}(m(\mathbf{x}),s^2(\mathbf{x}))}(\cdot)$ for the density

of the $\mathcal{N}(m(\mathbf{x}), s^2(\mathbf{x}))$ distribution, we have :

$$\begin{aligned} EI(\mathbf{x}) &= \int_{-\infty}^{T} (T-u) f_{\mathcal{N}(m(\mathbf{x}),s^2(\mathbf{x}))}(u) du \\ &= (T-m(\mathbf{x}))\Phi\left(\frac{T-m(\mathbf{x})}{s(\mathbf{x})}\right) + s(\mathbf{x})\phi\left(\frac{T-m(\mathbf{x})}{s(\mathbf{x})}\right), \end{aligned} \tag{A.11}$$

where Φ and ϕ stand for the cumulative distribution function and the probability distribution function of the standard Gaussian distribution, respectively. Here we propose a variant of EI meant to put more emphasis on the exploration of basins of minimum while remaining tractable. Indeed, the aim in our motivating applications is not only to find the global minimizer(s) of g as quickly as possible, but also to find *a representative subset* of inputs leading to a response curve close to the reference, i.e. to a small misfit. The proposed trick to lower the repulsion effect of current best points is to replace $\min(g(\mathbf{X}_n))$ by a quantile of $g(\mathbf{X}_n)$ in the definition of EI. Calling α the level of this quantile, we define

$$EI_\alpha(\mathbf{x}) = (q_\alpha - m(\mathbf{x}))\Phi\left(\frac{q_\alpha - m(\mathbf{x})}{s(\mathbf{x})}\right) + s(\mathbf{x})\phi\left(\frac{q_\alpha - m(\mathbf{x})}{s(\mathbf{x})}\right) \tag{A.12}$$

where $q_\alpha = q_\alpha(\mathbf{X}_n)$ is the empirical α%-quantile of the sample of misfits $g(\mathbf{X}_n)$. Varying α allows tuning the criterion from normally explorative to very local. Indeed, when $\alpha = 0$, $q_{\alpha,n}$ coincides with the minimum of $g(\mathbf{X}_n)$, so that $EI_0 \equiv EI$. However, when tuning α to a strictly positive value (obviously smaller than 1), the tendency of EI to vanish near the observation points disappears. To prevent the algorithm from resampling at already explored points, we exclude them from the search. However, we are interested in points very close to the already explored points in terms of the proposed kernel, since they have similar proxy responses but may be very different in terms of inputs. Different values of α will be investigated in the application section, where the benefit of taking $\alpha > 0$ will be illustrated.

A.5 Illustration of the method through a case study

To illustrate the proposed approach, we consider a synthetic example. The input parameter fields $\mathbf{x}_i$ are categorical and describe a 2D vertical geological section (example Fig. A.4(a)). The ensemble of fields $\mathbf{x}_i$ is generated using multiple-point statistics. The forward problem consists in computing a breakthrough curve $f_\mathbf{x} : t \in [0, T] \to f_\mathbf{x}(t) \in \mathbb{R}_+$ using an accurate numerical solver f. An example of a breakthrough curve is illustrated in Fig. A.4(e).

The general aim, in this example is to test whether the proxy based kriging approach proposed in the previous section is able to help identifying efficiently a subset of fields $\mathbf{x}_j$ with relatively small misfit values $g^\circ(\mathbf{x}) := d(f_{\text{ref}}, f_{\mathbf{x}_j})$.

To test the reliability of the method, the procedure is repeated using different reference curves and configurations. The details of the generation of the geological models (i.e. the parameter fields), the forward numerical solver, and two proxy simulators are described in the following sections before presenting and discussing the results.

A.5.1 Geological facies simulations

The prior geological model for this case study is based on an aquifer analogue: the Herten site (Bayer et al., 2011), typical of a glacio-fluvial environment, and on a multiple-point statistics geological model. Thousand realizations (the $\mathbf{x}_i$'s) of the geological medium are generated using the Direct Sampling (DS) multiple-point statistics method (Mariethoz et al., 2010) using a geological section mapped from the Herten site as a training image. Fig. A.4(a)

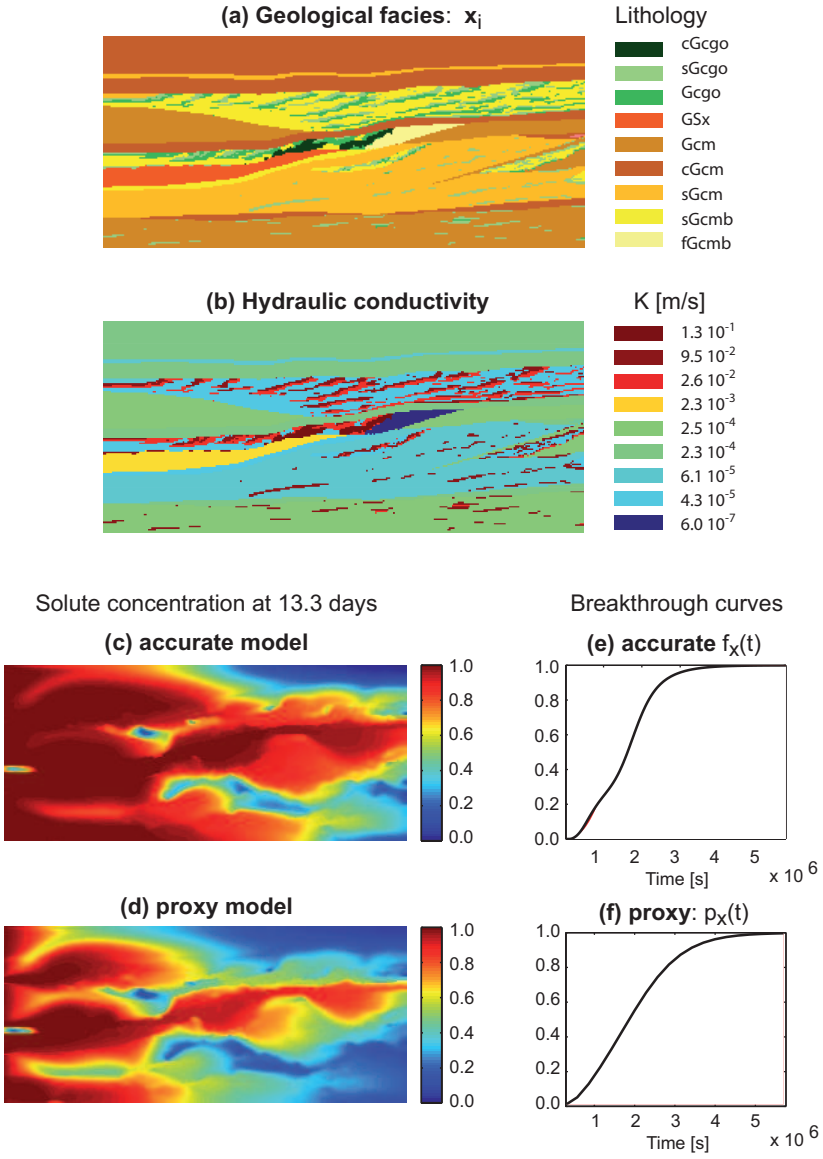

Figure A.4: Illustration of the main components of the illustrative example: (a) one simulation of a categorical field representing the geological facies. Each color corresponds to a type of sediment; (b) hydraulic conductivity field derived from the geological facies; (c,f) computation of the transient distribution of the concentration fields and breakthrough curve using an accurate numerical simulator; (d,f) approximation using a proxy simulator.

shows one example. The grid has a size of 320 by 140 pixels covering an area of 16m by 7m. Each color corresponds to a geological facies. The parameters of the DS method are: a search neighborhood of 20 cells on each axis, a maximal number of neighboring nodes of 15, a distance threshold of 0.01, and a maximal scan fraction of 0.5. All the stochastic realizations are constrained by a secondary variable (describing the large scale sedimentary structures) in the training image and in the simulations, following the approach used by Comunian et al. (2011). Fig. A.5(a) displays 9 of those realizations. The variability between them is present only at small scales within the main sedimentary bodies. The large scale structures are identical in all realizations. The ensemble of those geological models constitutes a sample of our prior distribution on parameter fields.

A.5.2 Flow and transport simulations

The breakthrough curves are obtained by solving the classical advection-dispersion equation in transient state using a finite volume technique (Künze and Lunati, 2011, 2012). The spatial discretization is kept identical to the one used for the geological simulations. The boundary conditions and parameters are summarized in Tab. A.1. A constant value of the hydraulic conductivity is assigned to each facies (Fig. A.4(b)) according to the mean values obtained from laboratory experiments and described by Bayer et al. (2011). For the sake of simplicity, the porosity is considered homogeneous over all facies. A constant head is prescribed on the left (0.1m) and right boundaries (0m) and remain constant while the upper and lower boundaries are no flow boundaries. Those boundary conditions lead to a uniform steady-state flow from left to right.

The initial distribution of the solute concentration is set to zero everywhere in the domain. A fixed concentration of 1 is prescribed on the left boundary. The transport problem is solved in transient state. Figure A.4(c) shows the map of the solute concentration for the realization shown in Fig. A.4(a) after 13.3 days of simulations. On the right boundary, the solute fluxes are integrated to compute the breakthrough curve $f_{\mathbf{x}}(t)$ representing the mean concentration at the outlet versus time (Fig. A.4(e)).

Table A.1: Parameter values for the solute transport model

Parameter	Value
Porosity	0.35
Molecular diffusion	$4.0 \times 10^{-9}\ m/s$
Longitudinal dispersivity (along x axis)	$0.1\ m$
Transversely dispersivity (along z axis)	$0.01\ m$
Total simulation time	$1.44 \times 10^{7}\ s$
Time steps length	$1.44 \times 10^{4}\ s$

Despite the apparent limited variability in the geological structure discribed above, a rather wide range of tracer breakthrough responses are obtained on the prior ensemble (Fig. A.5(b)). This illustrates the importance of the internal heterogeneity of the high permeability features within the main sedimentary layers.

A.5.3 Two different proxies

A good proxy is faster than the accurate numerical model and allows to distinguish parameter fields that have similar or different responses in terms of tracer breakthrough. Such a proxy is generally not expected to provide an accurate simulation of the breakthrough or of solute concentration states. It should simply be a fast approximation allowing to discriminate parameter fields.

For this case study, we consider two different proxies and check their performances and reliability. The first one, $p^1_{\mathbf{x}}(t)$, is based on simplified physics. We use the same solver (Künze and Lunati, 2011, 2012) and the same spatial and temporal resolution as for the accurate model based on the full physics, but we disregard diffusion and dispersion effects. The numerical simulation thereby only accounts for advection and numerical dispersion phenomena. The second proxy, $p^2_{\mathbf{x}}(t)$, is based on simply coarsening the time discretization of the accurate model. The number of time steps is reduced; their duration is increased to $2.88 \times 10^5\ s$ (i.e. a division by 20 of the number of time steps).

The breakthrough curves computed with the two proxies are displayed in Figs. A.5(c) and A.5(d). The first proxy gives breakthrough curves whose general shape resemble more the accurate model than the second proxy: some of the curves display a sigmoidal shape

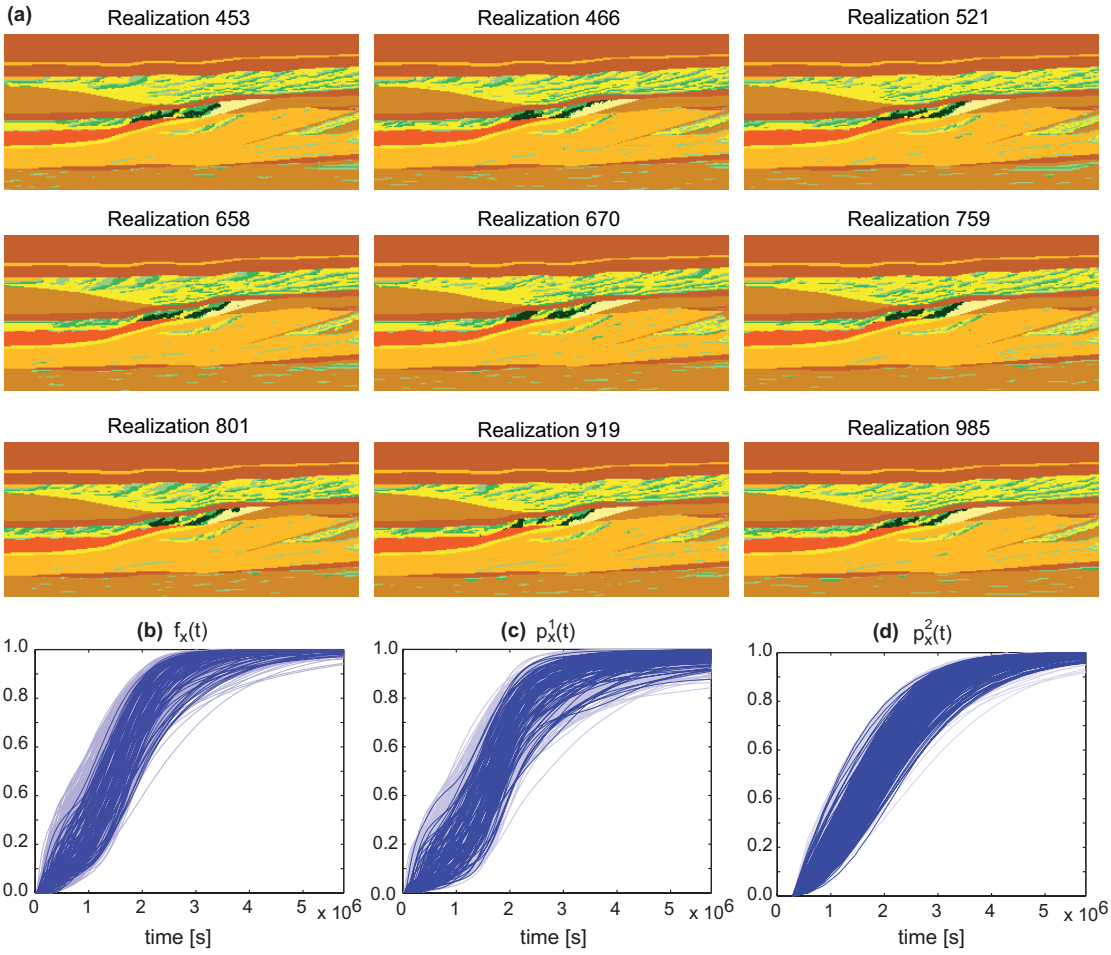

Figure A.5: (a) 9 realizations of the lithofacies. Because all the simulations are constrained by the large scale structure data, only the internal architecture within the main layers is displaying some variability between the simulations. (b) Ensemble of the breakthrough curves obtained with the accurate numerical model and the two proxies (c and d) for the 1000 parameter fields. To make the Fig. more readable, some breakthrough curves are represented in light gray color.

like the fine scale solution. The second proxy results in breakthrough curves that are more regular. For this proxy, the first arrivals of the tracer are almost identical for all geological models because of the coarse temporal resolution. The responses for $p^2_{\mathbf{x}}(t)$ present some variability, but less than $f_x(t)$ and the first proxy. For both proxies, the computational time is reduced by a factor of about 20. The accurate numerical solution takes about 7.5 minutes on a PC, while the two proxies run in about 20 seconds each.

A.5.4 Results

Let us now apply our kriging model to the problem of predicting the transformed misfit between the breakthrough curves of a given reference and the responses associated with the 1000 candidate geological media (i.e. parameter fields). The proxy used here is $p^1_x(t)$, the one with simplified physics. For now we arbitrarily choose one of the actual response curves (the realization with index 800) as a reference for illustration purposes. Note that more general results will be presented in section A.6, where statistics will be derived based on 100 randomly chosen reference curves.

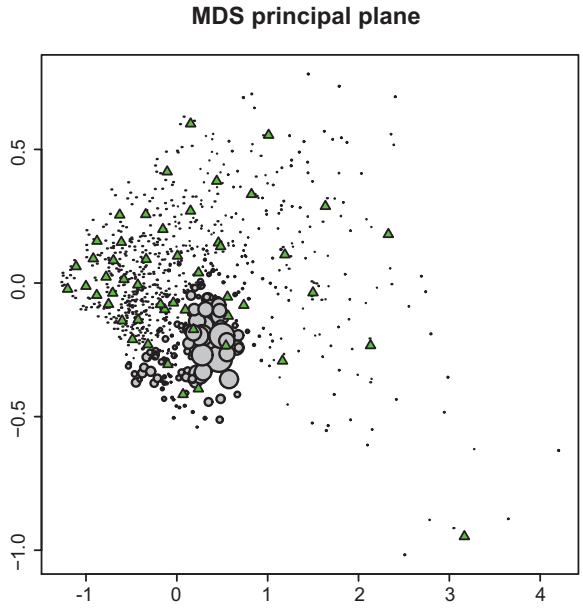

Figure A.6: Every point in the MDS space represents a parameter field. The triangles indicate the models that were selected by the K-means algorithm for the initial design of experiments, and the radius of the circles are proportional to the EI_α criterion.

Among the 1000 considered inputs, 50 are chosen based on a clustering technique using proxy-induced distance (Fig. A.6), in the flavor of Scheidt and Caers (2009)'s approach. The actual response curves are calculated by using the accurate numerical model with the latter inputs, and the 50 corresponding values of misfit to the reference curve are calculated and stored in a vector, denoted by $g^\circ(X_{50})$ or $\mathbf{g}^\circ$, as in section A.3.

As shown on Fig. A.7, a transformation is used to make the data misfits closer to Gaussian. For simplicity, we restrict the transformation to be a power transform, $g = (g^\circ)^\gamma$. The ad hoc approach proposed here to determine the coefficient of this transform is to set the skewness of the transformed sample equal to zero. As will be presented in more detail in section A.6 (performance assessment), such transform significantly improves the predictivity of the kriging model, as well as the performances of the inversion algorithm proposed in the next section.

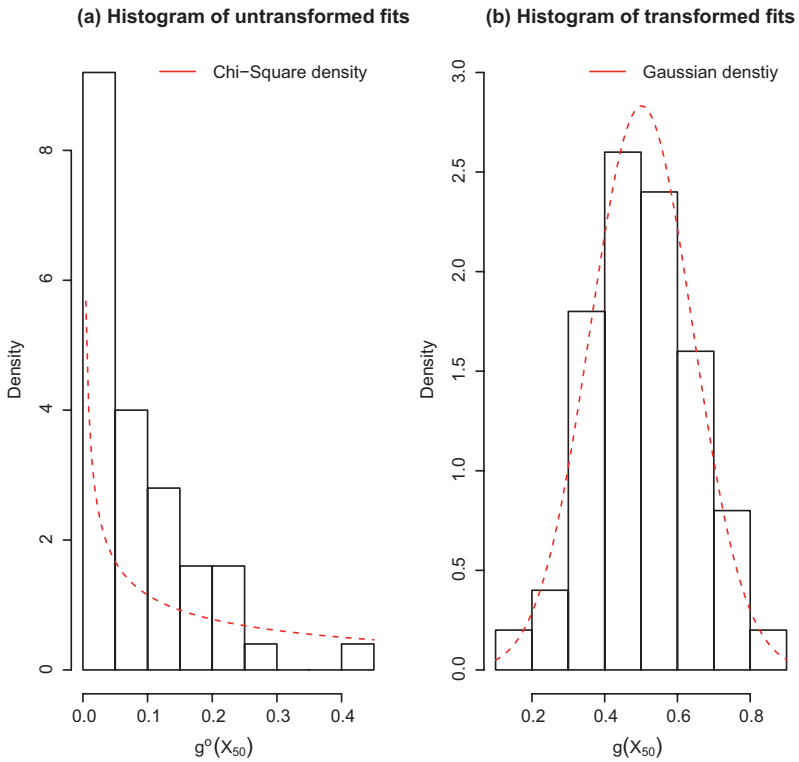

Figure A.7: Samples of untransformed (left) and transformed (right) misfit values obtained at a 50-point initial design of experiments in the case of a proxy with simplified physics. The histogram of the untransformed sample is closer to a chi-square distribution, whereas the one obtained by a power transformation, although remaining positive, is much more similar to the one of a Gaussian sample. The exponent used in the power transformation ($\gamma \approx 0.24$ here) is obtained by setting the skewness of the transformed sample to 0.

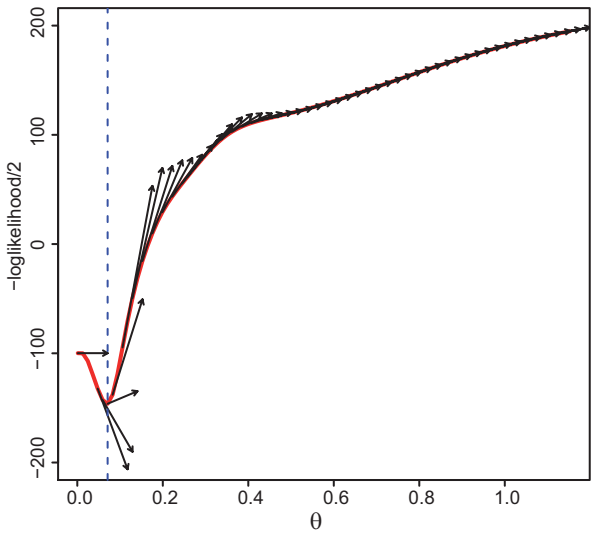

Figure A.8: Identification of θ by maximizing the concentrated log-likelihood function. The large curvature at the minimum indicates a well-identified parameter value. The arrows stand for the gradients of the log-likelihood function, which have been calculated analytically, and are used within the optimization procedure.

In a second step, we estimate the kernel parameters by maximum likelihood (MLE) based on the transformed sample of fits. We can see in Fig. A.8 that the optimal value of θ is very clearly defined since the log-likelihood curve has a large curvature at its minimum value.

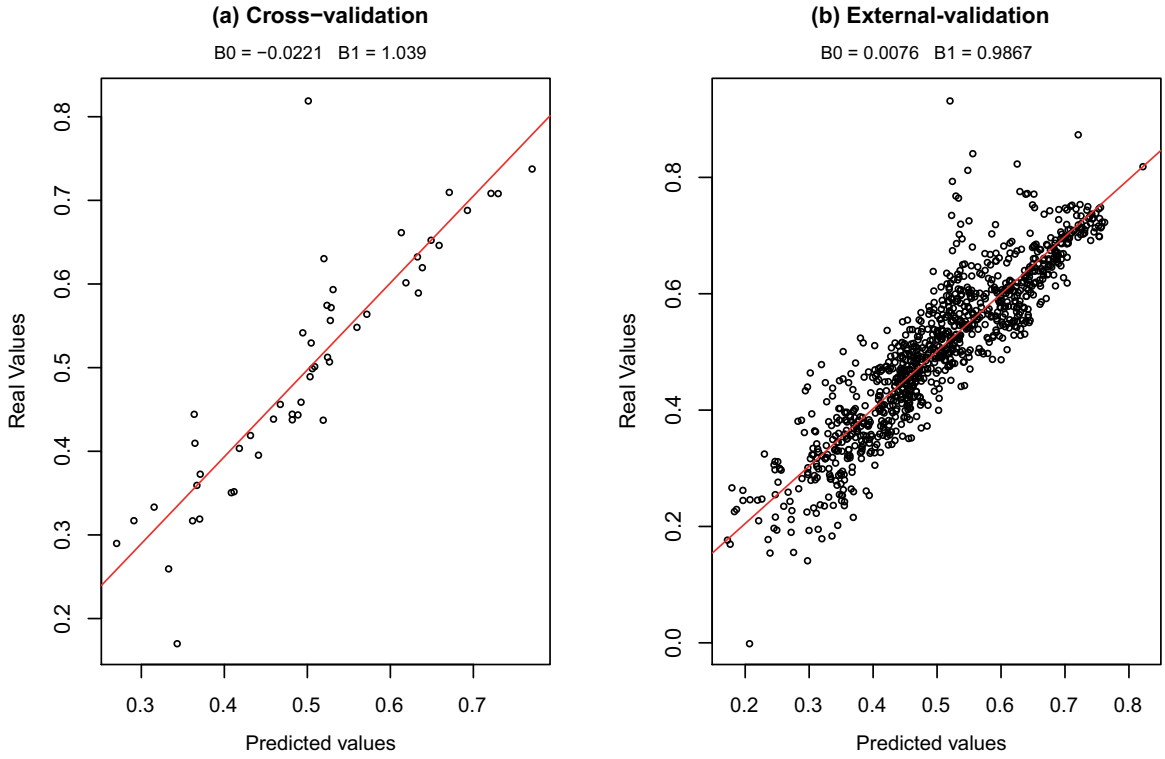

Figure A.9: (a) Cross validation and (b) External validation

The quality of the resulting kriging predictions is then evaluated: we first use a standard cross validation technique on the 50 samples used to build the kriging model (Fig. A.9(a)) and then extend the comparison to an external validation on the complete ensemble of 1000 values (Fig. A.9(b)). In both cases, the predicted values obtained by kriging are in good agreement with the true values; the regression line of predicted versus actual values has in intercept $B0$ close to zero and a slope $B1$ close to 1 (Fig. A.9), indicating that the kriging predictions are not highly biased. Furthermore, one can see that the leave-on-out errors of (a) give a reasonable estimate of the prediction errors observed a posteriori on the exhaustive validation set.

The 50 iterative selections operated by the ProKSI algorithm provide candidate parameter fields that are assessed through the accurate numerical flow simulator. The misfit with the reference is plotted on Fig. A.10.

A.6 Performance assessment

The good results obtained in the leading example (See Figures A.9 and A.10) are of course conditioned by the chosen reference breakthrough curve f_{ref} (arbitrarily chosen as the one with index 800) and do not constitute a sufficient basis to appraise the ProKSI algorithm. Furthermore, the method is proxy-dependent, and it would make sense to test the sensitivity of the performances to both an improvement or a degradation in the proxy. In this section, we propose a more systematic benchmarking of the algorithm's performance by analysing the results obtained with 100 different response curves, and for three different proxies, with

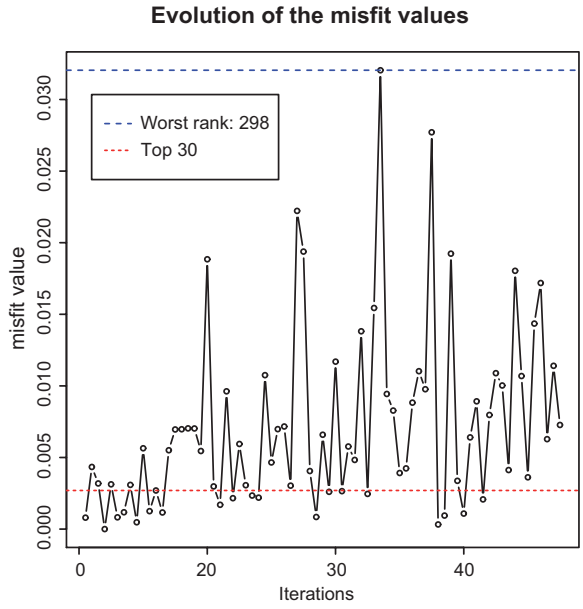

Figure A.10: Monitoring of the misfit values obtained for the parameter fields sequentially chosen by the ProKSI algorithm.

a comparison to Monte Carlo random search in the case of the worse proxy. In that last situation, we will use a completely inadequate proxy model to test the robustness of the method. Furthermore, the effect of the power transform applied to the misfit function, as well as the effect of the replacement of the minimum by a quantile in the *EI* criterion are investigated. Before giving more details about the benchmark and the obtained results, let us first present the main performance evaluation metrics.

A.6.1 Performance evaluation metrics

EM1: current best model's rank. One of the most natural way of evaluating an optimization method consists in monitoring the evolution of the misfit as a function of the number of iterations (Fig. A.10). One can also plot the smallest misfit value achieved so far as function of the number of iterations. However, the curve obtained for such a metric would have a scale (on the y-axis) depending on the considered f_{ref}, which would prevent us from making comparisons between different tests. As a consequence, we choose to focus on the evolution of the rank of the current best model among the 1000 candidates. This rank would normally be unknown but here we can compute it because we evaluate the true misfit for all the candidate models (even those which are not selected by the ProKSI algorithm) in order to be able to test the efficiency of the method. Repeating the numerical experiment with multiple references, we can then plot some statistics of the rank as a function of the number of iterations (Fig. A.11(a)).

EM2: number of evaluated models from the top 30. The first metric (EM1) focuses on the capacity of the method to find at least one parameter field with a low misfit value, but not on its ability to explore the set of parameter fields with low misfit values. EM2 is meant to be a complement to EM1, by measuring the number of models of the top 30 (i.e. the 3% best models in terms of misfit value) evaluated along the algorithm. Though this proportion might seem rather arbitrary, EM2 gives a good picture of the algorithm's tendency to explore the possible multiple peaks of the posterior distribution of models. Again, the statistics of EM2 are plotted as a function

of the number of iterations (Fig. A.11(b)). Note that in the best configuration tested here (proxy 2 with $\alpha = 0.6$), 26 parameter fields among the top 30 were found in median after 75 iterations, so exploring all fields of top 30 out of 75 iterations appears as a kind of reachable best case (contrarily to exploring 75 among 75).

EM3: probability that random search outperforms the proposed algorithm. It is expected that an elaborated algorithm like ProKSI (relying on a metamodel) performs better than random search, and at least not much worse in cases where the proxy is misspecified. The metric EM2 is well-adapted to base a comparison of ProKSI to a naive Monte Carlo (MC) algorithm, since the probability distribution of the number of points visited in the top 30 can be analytically derived for the case of a random search (this number then follows a hyper-geometric distribution). EM3 gives at each iteration of ProKSI the probability that MC sampling finds more points in the top 30.

A.6.2 Benchmark: design and implementation

Design of the benchmark

The aim of the benchmark was to assess the global performances of the ProKSI algorithm on the considered case study with the following specific questions in mind. **How sensitive are the performances to: (Q1) the chosen proxy, (Q2) the value of the quantile α, and (Q3) the normalizing transform of the misfit values?**

Consequently, we ran replications of the algorithm (by varying the reference curve) with different proxies, with or without power transform of the misfit function, and with different values of α. In order to obtain results based on solid statistical analysis, rather than on an arbitrary set of examples with a potentially low generalizability, we ran the ProKSI algorithm 100 times for each configuration (i.e. for each considered (proxy, transform, α) combination). For each considered proxy (p^1, p^2, and a third mismatched one described below), 50 parameter fields were chosen by Scheidt and Caers clustering technique, and 100 f_{ref} were randomly chosen among the 950 remaining candidates. Then, for any given configuration (in terms of transform and/or α value), 75 iterations of the ProKSI algorithm were run for the 100 chosen f_{ref}. The results are visualized in terms of box-plot sequences representing the statistical distributions of 100 values for the considered evaluation metric, evolving over the 75 iterations. Finally, for EM3, one sequence of 75 probabilities that a Monte Carlo algorithm would lead to more points in the top 30 than the proposed approach (one probability per iteration) can be produced for each replicate. We chose to summarize these results by representing sequences of box-plots for the selected configurations.

Implementation of the benchmark

All the benchmark algorithm runs and the performance evaluation calculations were done using the open source statistical software R, based on the numerical simulation results obtained for the 1000 multiple-statistics simulations (see implementation details in section A.5). The R code, gathered in form of a package (*ProKSI*, forthcoming on the *Comprehensive R Archive Network*), was called for each task of the following loop, forming the basic brick of the benchmark for any fixed configuration:

A.6.3 Results

The first benchmark results, displayed on Fig. A.11, deal with the performances on the ProKSI algorithm when applied to our test-case with proxy 1, and default settings concerning the normalizing transform and the *EI* variant (power transformation done, and

Algorithm 2 Testing procedure for a proxy with a given algorithm configuration

1: **Choose** the initial design of experiment (**50** points using Scheidt et Caers approach).
2: **Choose 100** different simulations among the 950 remaining points.
3: **for** $i = 1$ to $i = 100$ **do**
4: **Run 75** iterations of the algorithm on the i^{th} reference.
5: **Evaluate** the 3 EM's for each iteration of the i^{th} run.
6: **end for**

$\alpha = 0.15$). Fig. A.11(a) represents the evolution of the statistics (box-plot) of EM1 over

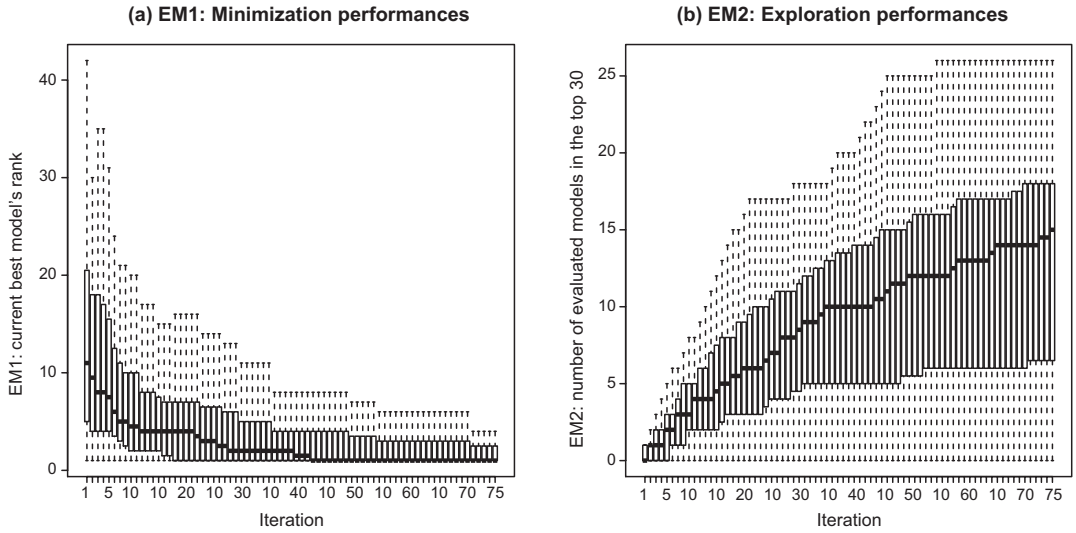

Figure A.11: Performances of the ProKSI algorithm (based on proxy 1) with a transformed misfit. (a) box-plot of the EM1 criterion over the 100 numerical experiment replicates. (b) box-plot of the EM2 criterion.

the 100 replicates, along the 75 iterations of the algorithm. We can see here that in 42 iterations, the actual best parameter field (out of 1000) has been found for more than 50% of the replications. In Fig. A.11(b), the exploration performances are investigated in terms of EM2; it is found here that 15 parameter fields among the 30 best ones have been evaluated in median after 75 iterations of ProKSI. In total, these results show both how the proposed kriging metamodel helps reaching a fast convergence, and that ProKSI achieves a rather satisfying exploration of the set of best parameter fields in a limited number of iterations.

Effect of the misfit transformation on the algorithm performances

Fig. A.12 represents the performances (in terms of EM1 and EM2) obtained by applying the ProKSI algorithm to our case study with default settings concerning the *EI* criterion ($\alpha = 0.15$) but without normalizing power transform for the misfit function.

The results appear to be clearly inferior to the ones obtained with the transformation: even after the 75 iterations, the median rank of the best evaluated parameter field is strictly above 1, which expresses a significantly slower convergence of ProKSI as with the transformed misfits. Similarly, the number of models forming the top 30 evaluated along the algorithm stagnates around 8 in median after the 75 iterations. The normalizing transform has thus clearly a positive effect on the efficiency of the algorithm, both in terms of fast convergence, and in terms of global exploration of the nearly optimal parameter fields.

However, as illustrated on Fig. A.13, the results in terms of EM2 are good enough to

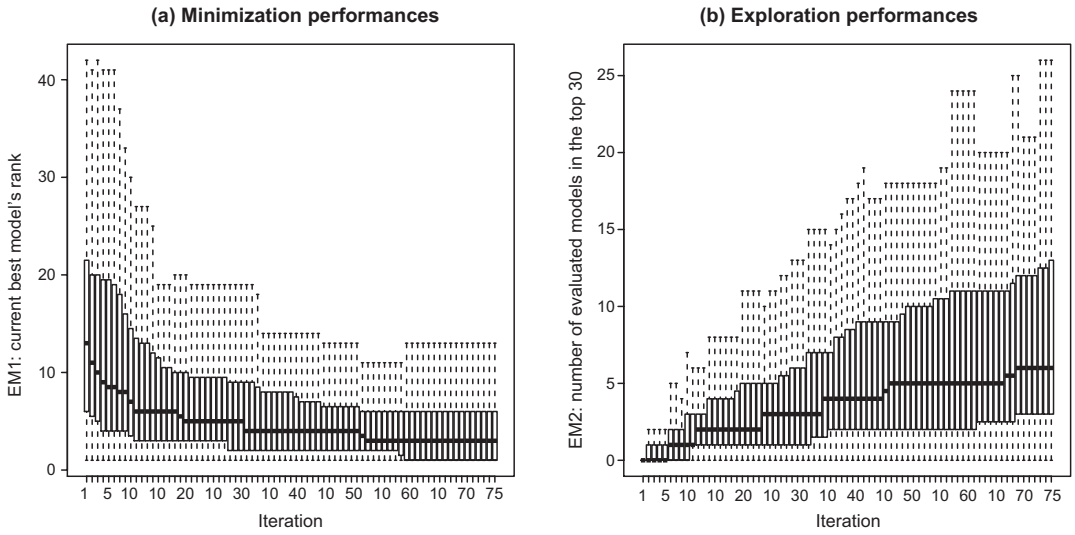

Figure A.12: Performances of the ProKSI algorithm (based on proxy 1) without power transform of the misfit. (a) box-plot of EM1 over the 100 replicates of the numerical experiment. (b) box-plot of EM2.

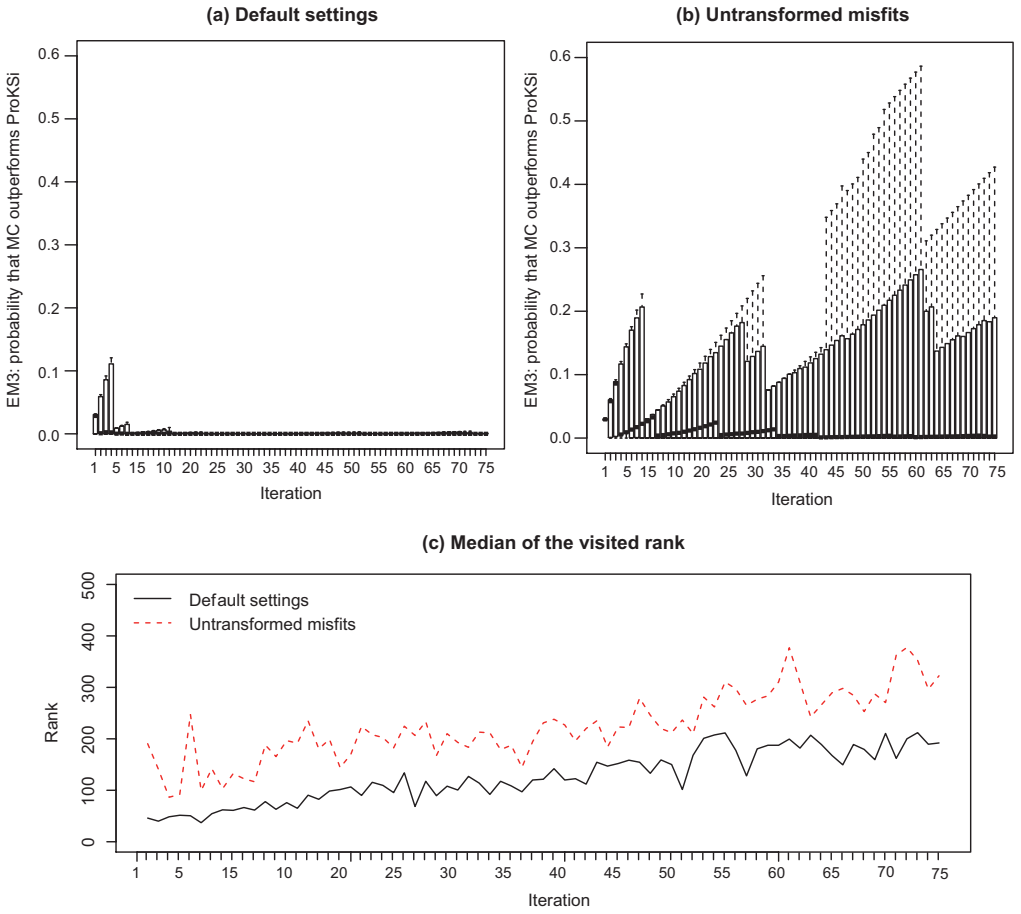

Figure A.13: Effect of the misfit transformation on the performances of the ProKSI algorithm in terms of its superiority with respect to a Monte-Carlo search and median rank of the evaluated models.

outperform a pure random search (upper right graphic). On the lower graphic, the evolution of the median rank for the parameter fields evaluated by ProKSI with or without transform illustrate that the algorithm with transform spends more time in low misfit regions.

Effect of an improved proxy on the algorithm performances

Let us now present the results obtained when using the second proxy with default settings. The most striking result when looking at Fig. A.14 is the impressively fast convergence of the

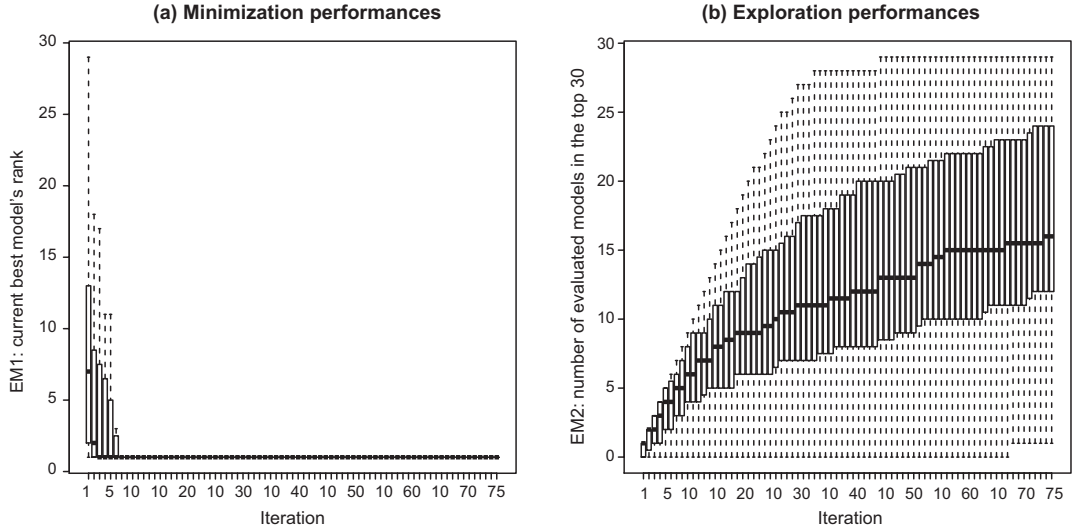

Figure A.14: Performances of the ProKSI algorithm (based on proxy 2) with default settings.

algorithm in terms of EM1 criterion. Indeed, in 7 iterations, the minimizer has been found in all considered cases (100 replicates). ProKSI successfully relies here on the information given by proxy 2 for uncovering the best point, only based on slightly more than the misfit values for the set of 50 initial models. What seems really outstanding in that case is that such a result is uniformly obtained for the 100 reference curves. To milden this success a bit, let us remark that the performances in terms of exploration are comparable to the first proxy, that is one half of the top 30 models were evaluated in median after termination.

Effect of the α parameter (from EI_α) on the algorithm performances

We investigate here the effect of the parameter α, tuning the quantile level in the proposed generalization of EI, on the performances of the algorithm. We obtained very different results for the two proxys. Indeed, the performances of ProKSI were not very sensitive to α when using the first proxy, so that we do not discuss this case here, and refer the interested reader to the appendix for more detail. However, α was found to be strongly influencing the algorithm's performances when using the second proxy, as illustrated on Fig. A.15.

It is indeed observed on Fig. A.15 (a and b) that using ProKSI with the standard EI criterion ($\alpha = 0$) is less efficient compared to the considered default value $\alpha = 0.15$ (See Fig. A.14): even though the algorithm convergence to the minimum is always comparably fast, the exploration performances are strongly affected by this change of criterion (median number of points in the top 30 after termination decreased from 15 to 10). On the other hand, increasing *alpha* to 0.6 was found to greatly improve the results in terms of exploration (again, without affecting the minimization performances, see Fig. A.15 (c)) since the median

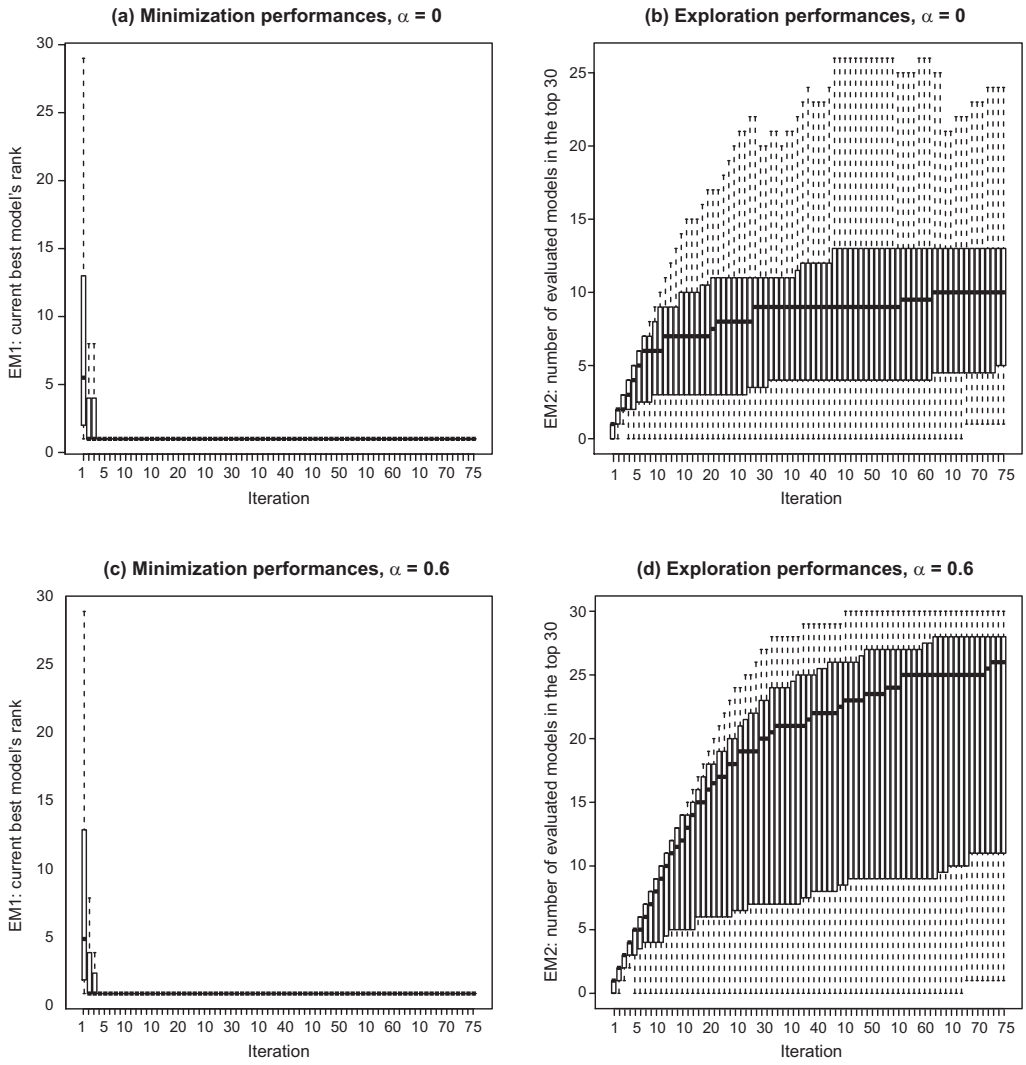

Figure A.15: Effect of the α parameter on the performances when using the second proxy

number of points in the top 30 jumped to 25, as can be seen on Fig. A.15 (d). In summary, introducing this parameter α was found beneficial for forcing the algorithm to spend more iterations (out of the 75 allocated ones) exploring the top candidates. Its optimal tuning is of course problem-dependent. The rather arbitrary default value $\alpha = 0.15$ chosen here gave improved results in both considered cases, even though better performances were reached by using a larger α value in the case of the second proxy.

Effect of a non-informative proxy on the algorithm performances

Finally, we propose to test the performances of ProKSI when using a completely inadequate proxy model. The idea is to see if the algorithm remains consistently applicable when the simplified model is poorly (or not at all) informative, and how using ProKSI in such degraded conditions would perform compared to a naive Monte Carlo search.

In order to emulate a non-informative proxy, we started from proxy 1, and randomly permuted the 1000 indices. We then ran the ProKSI algorithm with this "mismatched" proxy, and compared them to trajectories obtained by Monte Carlo (the whole replicated for the 100 reference curves).

As illustrated on Fig. A.16, the performances of ProKSI with "mismatched" proxy are

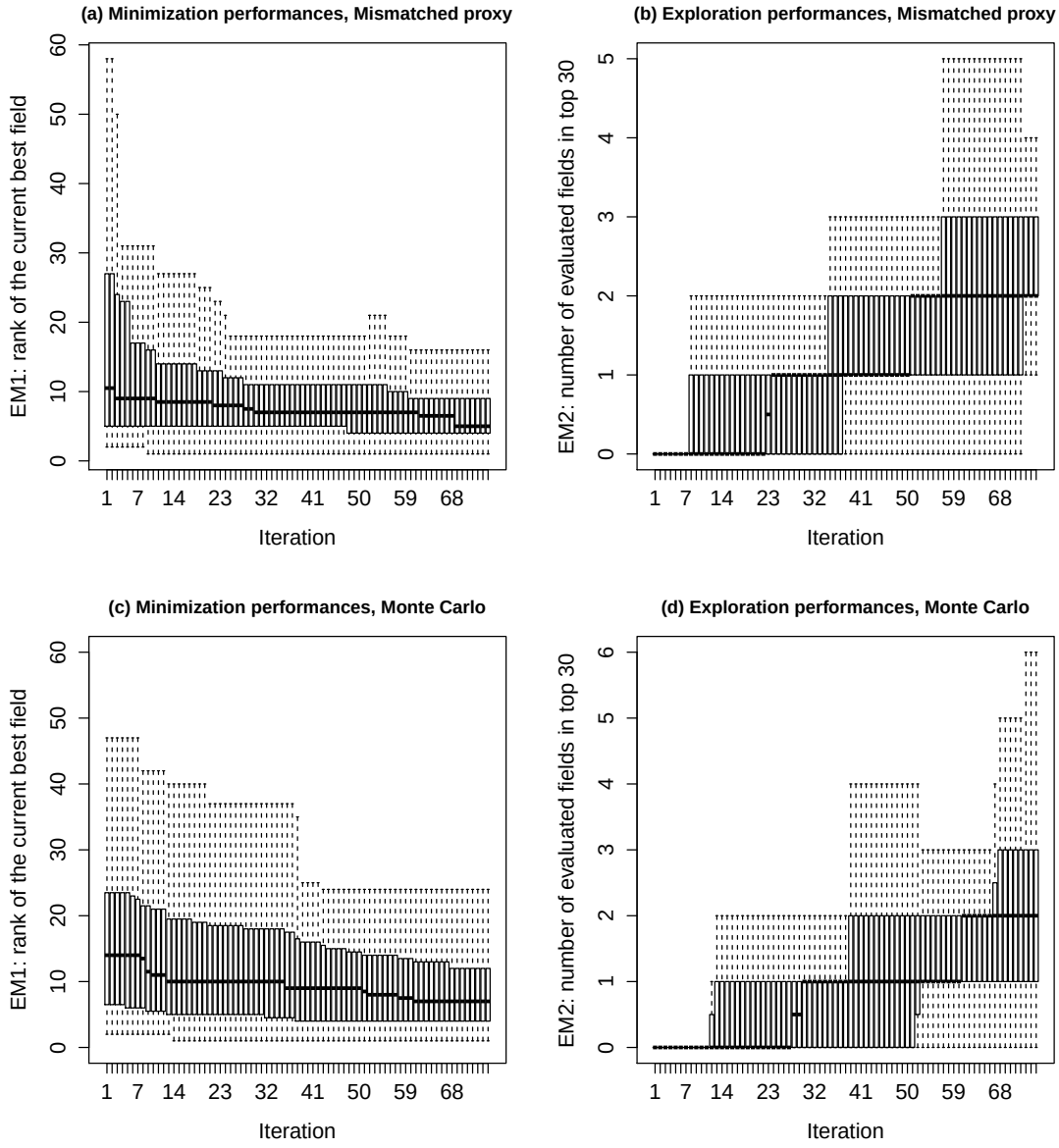

Figure A.16: Effect of a non-informative proxy on the performances.

comparable to those of Monte Carlo. The algorithm hence appears reasonably robust to a proxy misspecification, while being potentially very efficient for well-chosen proxies.

A.7 Conclusion

Handling high resolution geological models in combination with complex physics solvers requiring heavy computational load to provide an accurate representation of a system while representing uncertainty is often mutually exclusive. Accurate complex models are often too computationally demanding to be used in the general framework of a Monte Carlo approach and analytical propagation of uncertainty is intractable. Resolving this issue is an important research topic both from a theoretical perspective and for a wide range of applications (Christie et al., 2011), including hydrogeology.

In this paper, we propose a contribution which consists in coupling an accurate model,

a simple model (the proxy), and a statistical metamodel. The statistical metamodel is used to link the results of the proxy with those of the accurate model. More precisely, this is achieved by developing a specific covariance kernel accounting for the difference in responses from the proxy models and allowing to predict by kriging the misfit between a given reference curve and the response of accurate model. One of the strengths of this idea is that the use of the distance between proxy responses permits to drastically reduce the dimension of the kriging problem and allows an efficient inference of the covariance parameters. The quality of the relation between the accurate and the proxy models is also directly taken into account via the covariance kernel. In addition, the chosen covariance kernel can be tailored to the practical problem that has to be solved (through the proxy, the kernel k_F, and more), which makes the approach quite flexible.

In the case study, we showed how such an approach can help in the case of an inverse conditioning problem where the prediction refers to the misfit between observations and the accurate model responses. As a first step, we propose here an iterative search algorithm. This example extends previous work done by Caers and colleagues (Suzuki et al., 2008; Scheidt and Caers, 2009; Caers, 2011) in which we add a step based on the kriging model described above to orient the search. We propose to guide the selection of a model during the search by defining a modified Expected Improvement criterion EI_α such that the algorithm will potentially explore multiple basins of minimum, if they exist.

The systematic analysis of the case study showed the following results.

- When the proxy is informative, the method is extremely efficient in finding the parameter fields that minimize the misfit.
- When the proxy is less informative, the method efficiency decreases remains competitive with respect to a random search.
- The proposed modified expected improvement criteria allows both identifying the global minimizer and exploring the existing basins of minimum.
- The method is more efficient if the misfit is properly transformed so as to get a close-to-Gaussian sample.
- The parameter α –defining the quantile of the misfit distribution below which a parameter field is considered as an interesting candidate– allows to control the degree of exploration of the method. For low values of α, the algorithm tends to rapidly leave regions of parameter fields with proxy responses similar to those of already evaluated fields. For higher values of α, the algorithm spends more iterations in regions of fields with proxy responses close to the ones of evaluated fields with the lowest misfit values, hence performing a more extensive exploration of top candidates.

The results obtained so far are very encouraging and show that the use of a kriging technique to couple a complex and a simple model will open a broad range of new perspectives. Note that in the presented example, selecting the misfit minimizer among parameter fields generated by multiple-points simulation provides a useful parameter reconstruction without any regularization. If one wants to obtain not only the best solution but an ensemble of models, then the selection criterion and the iterative search procedure will have to be modified in order to ensure that the final ensemble will be a representative sample of the posterior distribution. The method can also be extended in a relatively straightforward manner to allow generating new candidate models by coupling it, for example, with the Iterative Spatial Resampling method (Mariethoz, 2010). It is also very clear that this type of approach can be parallelized to improve the numerical performances (Ginsbourger et al., 2010).

As a final note, the model selection approach illustrated in the case study is applicable when detailed geological knowledge is available (type of geological environment, sufficient data to build a good geological model, and known properties for the various facies) and when this information can be used to generate a prior ensemble of geological models. An obvious limitation of that approach is that the algorithm is only searching within a predefined ensemble of parameter fields. If all the proposed candidates are very unrealistic, it may happen that none of the candidates will lead to a reasonable fit to the observations. However, the approach is very flexible: users can apply it with many different geological priors, and explore their relevance in an efficient manner with the proposed method. There is no limit in the type of prior on the parameter field used to generate the initial ensemble (e.g. boolean model, truncated plurigaussian, pseudo-genetic model, etc.). Coping with such a variety of priors is usually not possible with more traditional inversion methods.

Acknowledgments

The work presented in this paper is part of the *Integrated methods for stochastic ensemble aquifer modeling (ENSEMBLE)* project supported by the Swiss National Science Foundation under the contract CRSI22_122249/1. The authors want to thank I. Lunati, who provided the finite volume matlab code to solve the flow and transport equations and helped to setup the flow problem and proxy simulations, as well as N. Linde, P. Brunner and five anonymous reviewers for their constructive comments.

Proof that a p.d. kernel chained with a proxy is p.d.

Property Let E and F be two arbitrary spaces. Given a positive-definite kernel k_F over $F \times F$, the kernel k_E defined by

$$k_E(\mathbf{x}, \mathbf{y}) := k_F(p(\mathbf{x}), p(\mathbf{y})) \tag{A.13}$$

is positive-definite over $E \times E$ for any function $p : E \longrightarrow F$.

Proof. Let $n \in \mathbb{N}$, $\mathbf{x}_1, \ldots, \mathbf{x}_n \in E$, and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Then

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j k_E(\mathbf{x}_i, \mathbf{x}_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j k_F(p(\mathbf{x}_i), p(\mathbf{x}_j)) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j k_F(\mathbf{y}_i, \mathbf{y}_j) \geq 0\end{aligned}$$

by using the definition of positive-definiteness applied to k_F with the points $\mathbf{y}_i := p(\mathbf{x}_i) \in F$ $(1 \leq i \leq n)$ and the coefficients $\alpha_1, \ldots, \alpha_n$ as above.

On the approximate MLE used here for the covariance parameters

When $\tau^2 = 0$, it is known (Santner et al., 2003) that $\widehat{\mu} = \frac{\mathbf{1}^T R(\theta)^{-1} \mathbf{g}}{\mathbf{1}^T R(\theta)^{-1} \mathbf{1}}$, and the optimal value of σ^2 can be expressed as a function of θ only:

$$\sigma^{2*}(\theta) := \frac{1}{N}(\mathbf{g} - \widehat{\mu}\mathbf{1})^T R(\theta)^{-1}(\mathbf{g} - \widehat{\mu}\mathbf{1}), \tag{A.14}$$

where $R(\theta)$ is the correlation matrix of $G_{\mathbf{X}_n}$. Minimizing l is then equivalent to the one-dimensional minimization over θ of the so-called *concentrated*(or *profile*) log-likelihood:

$$l_c(\theta; \mathbf{g}) := l(\sigma^{2*}(\theta), \theta, 0; \mathbf{g}). \tag{A.15}$$

When $\tau^2 > 0$, it is not possible to concentrate on σ^2 as in Eq. (A.14) because the correlation matrix itself then depends on σ^2 too.

Here we approach the problem sequentially, and preserve the concentration step at the price of a minor approximation. First, an estimate of τ^2 is derived based on variographic considerations (variance of differences corresponding to increments with smaller norm values). Then, a first guess of σ^2, say σ_0^2, is made using a similar approach. This guess may also stem from a previous iteration in the case of a sequential design of experiments. Based on τ^2 and σ_0^2, an approximate formula –analogue to Eq. (A.14)– is then proposed for the optimal variance as a function of the range:

$$\widetilde{\sigma^{2*}}(\theta) := \frac{1}{N}(\mathbf{g} - \widehat{\mu}(\theta)\mathbf{1})^T \left(R(\theta) + \frac{\tau^2}{\sigma_0^2} I \right)^{-1} (\mathbf{g} - \widehat{\mu}(\theta)\mathbf{1}), \tag{A.16}$$

θ is then tuned by optimizing the following approximate concentrated likelihood:

$$\widetilde{l_c}(\theta; \mathbf{g}) := l(\widetilde{\sigma^{2*}}(\theta), \theta, \tau^2; \mathbf{g}) \tag{A.17}$$

Appendix B

Conditioning of Multiple-Point Statistics Facies Simulations to Tomographic Images

Abstract

Geophysical tomography captures the spatial distribution of the underlying geophysical property at a relatively high resolution, but the tomographic images tend to be blurred representations of reality and generally fail to reproduce sharp interfaces. Such models may cause significant bias when taken as a basis for predictive flow and transport modeling and are unsuitable for uncertainty assessment. We present a methodology in which tomograms are used to condition multiple-point statistics (MPS) simulations. A large set of geologically reasonable facies realizations and their corresponding synthetically calculated crosshole radar tomograms are used as a training image. The training image is scanned with a direct sampling algorithm for patterns in the conditioning tomogram, while accounting for the spatially varying resolution of the tomograms. In a post-processing step, only those conditional simulations that predicted the radar traveltimes within the expected data error levels are accepted. The methodology is demonstrated on a two-facies example featuring channels and an aquifer analog of alluvial sedimentary structures with five facies. For both cases, MPS simulations exhibit the sharp interfaces and the geological patterns found in the training image. Compared to unconditioned MPS simulations, the uncertainty in transport predictions is markedly decreased for simulations conditioned to tomograms. As an improvement to other approaches relying on classical smoothness-constrained geophysical tomography, the proposed method allows for: (1) reproduction of sharp interfaces, (2) incorporation of realistic geological constraints and (3) generation of multiple realizations that enables uncertainty assessment.

B.1 Introduction

Predictive modeling of subsurface flow and solute transport requires detailed models of the spatial distribution of hydraulic properties. A lot of recent research has focused on finding ways to use geophysical data for hydrological parameter estimation (e.g., Hubbard and Rubin, 2000; Linde et al., 2006; Eppstein and Dougherty, 1998; Dafflon and Barrash, 2012). The benefit of geophysical techniques is that a high number of sensors can be used at rather low costs and with little invasive impact. Data sets of thousands of data points of high spatial density are easily acquired and when these data are inverted, tomographic images of high resolution can be obtained. Unfortunately, inference of the geophysical property distribution from measured data, that is, the process of inversion or tomography, is generally non-unique due to limited data coverage, noisy data and non-linear physics. To overcome this non-uniqueness, the inverse problem is regularized where the most common approach is to explicitly search least-structure models by applying a smoothness constraint. Smoothness-constrained inversions are optimal in the sense of reducing regularization bias (Ory and Pratt, 1995) and turned out to be a forceful tool to prevent inversion artifacts. Their main disadvantage is that they produce tomograms that tend to image the subsurface property distributions as blurred or patchy structures and usually fail to recover sharp interfaces or small-scale structures. For accurate flow and transport modeling, we often need models that capture such features since they may control the subsurface connectivity.

The classical geostatistical approach aims at defining a model such that the two-point statistical relations observed in the field are matched (Kitanidis, 1997). These methods are invariant towards connectivity patterns (e.g., Gómez-Hernández and Wen, 1998; Krishnan and Journel, 2003), which means that parameter fields of very different connectivity characteristics can be identical in terms of their probability density function (pdf) or variogram. For example, Zinn and Harvey (2003) showed for parameter fields of near-identical lognormal univariate conductivity distributions, how flow and transport behavior changed dramatically

for different connectivity patterns. Various methods have been proposed to fully characterize parameter fields by including the connectivity patterns. Journel and Posa (1990) proposed to describe connectivity by indicator variograms that provide the probability that two distant points share the same indicator value. This approach has further been advanced by modified connectivity functions that account for points within a certain value range that are connected by an arbitrary continuous path (Western et al., 2001). A recent development are multiple-point statistics (MPS) algorithms that simulate parameter fields by accounting for the statistical relations between large sets of points, thereby allowing the reproduction of complex geological structures and connectivity patterns. The algorithms take the statistics from training images that feature the representative patterns that are subject to simulation. MPS approaches make use of the fact that certain geological settings are often built up by a limited number of repeating structural elements.

At the same time, the need to recover geological patterns and facies structures has been recognized and caught up by the geophysics community. The application of pattern-based approaches for the characterization of subsurface structures is a subject of intense research, mainly led by the petroleum industry (e.g., Coléou et al., 2003; de Matos et al., 2006). Caers and Ma (2002) extracted probabilistic facies distributions from seismic data using neural networks on colocated windows of seismic information. Similarly, Moysey et al. (2003) trained neural networks on geological analogs or well data to retrieve facies information from surface ground-penetrating radar (GPR) images. Based on this, Moysey et al. (2006) compared different transform techniques to translate the radargram into a facies classification. They found that textual measures that preserve the spatial structure of the GPR image often outperform measures based on univariate statistics (variance). A rather novel approach was presented by Cardiff and Kitanidis (2009), who developed an inversion scheme in which facies boundaries as well as parameter values within facies are updated by a Bayesian level set methodology. An example of how a GPR cross-section is used to establish a facies model, which in turn is used for flow and transport simulations is given in Rauber et al. (1998). Dafflon and Barrash (2012) use radar tomograms together with neutron borehole logs to condition stochastic simulations of the 3D porosity distribution in a well field. They show that conditioning to tomograms improves both simulated-annealing-based and Bayesian sequential simulation results.

In this study, we present a novel approach of combining geophysical models and conceptual geological knowledge to produce realistic high-definition subsurface facies models. A categorical training image (TI) containing the geological facies information available about the site of interest is converted into a tomogram by geophysical forward and inverse modeling. The tomogram can be seen as a filtered version of the facies TI that is dependent on the spatial distribution of the underlying geophysical property. Information about resolution of the tomographic setup, and about regularization and data error effects are included in the tomogram since all these factors change the 'tomographic filter' and consequently influence the tomography results. The discrete facies TI and its continuous tomogram are combined to form a bivariate TI. A conditioning tomogram is obtained by applying the same tomography procedure to geophysical data that are measured in the field. We then generate MPS simulations by scanning the bivariate TI for patterns found in the conditioning tomogram, in this case, we use tomograms from crosshole GPR experiments. Compared to conditioning to individual data points placed somewhere in the model domain (e.g., Zhou et al., 2012), we exploit the spatial information in the tomograms which is distributed over the entire domain.

The proposed method circumvents a problem generally inherent in the interpretation of geophysical models: Inferring the true geology from tomographic images requires correcting the models for all inversion effects, that is, one needs to deconvolve the tomogram and the

'tomographic filter' (Singha et al., 2007). Approaches to tackle this problem are linearized resolution analysis (Menke, 1989; Day-Lewis et al., 2005) or the use of numerical analogs to field studies (Moysey et al., 2005; Singha and Moysey, 2006). Since we are mimicking the geophysical tomography applied to the measured data when creating the bivariate TI, the 'tomographic filter' is implicitly taken into account. This and the fact that conceptual geological information is incorporated at an early stage of the inverse modeling process reduces the risk of misinterpretation due to limited knowledge about the 'tomographic filter' (e.g., Richardson et al., 1987; Rankey and Mitchell, 2003). Compared to classical inversion studies, we do not leave the interpreter with a geophysical image on which to base a geological interpretation, but we present a tool to condition a conceptual geological model to the distribution of the inferred geophysical parameters.

B.2 Methodology

B.2.1 General workflow

We developed a procedure to condition multiple-point geostatistical simulations to tomographic images. The method builds on the work of Mariethoz et al. (2010), who present a direct sampling algorithm for multiple-point geostatistical simulations. This multiple-point direct sampling (MPDS) algorithm allows to co-sample from a TI that contains both a distribution of a categorical and a continuous variable. Here, we use a categorical TI featuring characteristic geological structures and its corresponding geophysical tomogram. We then use a tomogram obtained at the field site of interest as a known continuous image to condition the simulation. In this way, we produce a categorical simulation, that contains the structures from the categorical training image and whose tomographic image is in agreement with the geophysical tomogram given as input variable. The general workflow is (see also Fig. B.1):

1. Invert the available geophysical field data by smoothness-constrained deterministic least-squares inversion. The resulting image is referred to as 'original tomogram' in the following.

2. Choose or create a training image that represents conceptually the expected geological structures/patterns. This is referred to as 'original TI'. This is a crucial step, a poorly chosen training image will severely bias the modeling outcome.

3. Create an ensemble of realizations of the chosen original TI. Here, this is done using MPDS.

4. For each realization of the original TI, generate a synthetic tomogram:

   (a) Assign realistic values of the underlying geophysical parameters (e.g., seismic or radar slowness, electrical resistivity) to the different facies in the realizations. This step can be based on established petrophysical models, for example the CRIM model (complex resistive index method, Tinga et al. (1973); Alharthi and Lange (1987); Roth et al. (1990)) or petrophysical relationships inferred from previous investigations (e.g., Hubbard et al., 2001).

   (b) Calculate synthetic geophysical data by simulating field experiments. Model the experiments using a setup similar to that used to acquire the actual data (Moysey et al., 2005). Contaminate the synthetic data with an expected amount of noise.

(c) Invert the synthetic data. Apply the same modeling parameters (grid discretization, initial model, regularization type and weights) as used to produce the original tomogram (Moysey et al., 2005).

5. Combine the realizations of the categorical original TI and their continuous tomograms to one bivariate 'working TI' that contains the facies and the geophysical parameters as two separate variables.

6. Simulate facies models based on direct sampling:

   (a) Pick a location in the simulation grid whose facies is not yet defined. Retrieve the data events (i.e., certain patterns formed by a pre-defined number of grid cells) centered at this location in the original tomogram and for the facies variable. Then, scan through the ensemble of synthetic tomograms and their corresponding facies models until similar data events are found. 'Similar' here means that the distances between the actual events and the found ones are below some predefined threshold. Building the working TI as stacks of two-dimensional models (as will be described in detail later) allows for choosing a scanning procedure so that the spatially varying resolution within a tomogram is accounted for.

   (b) Paste the facies value (at the central node of the data event) found in the TI to the selected location in the simulation grid.

   (c) Randomly pick a new location in the simulation grid and repeat until it is entirely filled.

7. For each simulated facies model, calculate geophysical forward data and compare to the true data. Accept the simulation as a possible model, if the weighted root-mean square error (WRMSE) of the predicted data is below a threshold $WRMSE_{target}$.

The resulting ensemble of accepted models forms a possible basis for geological interpretation, flow and transport modeling or uncertainty assessment.

B.2.2 Deterministic inversion

Both the original data (i.e., the 'true data' measured in the field) and the predicted data for all facies realizations undergo smoothness-constrained deterministic inversion (steps 1 and 4c in previous section). This inversion type is commonly referred to as Occam's inversion (Constable et al., 1987). In the smoothness-constrained inversion scheme, a model $\mathbf{m}$ is sought that features the least structure possible and still predicts the data within the assumed data error levels. In a general sense, we seek to minimize an objective function $\boldsymbol{\phi}$ that contains both a measure of the data misfit and the model roughness (e.g., Menke, 1989):

$$\boldsymbol{\phi} = \left|\left|\mathbf{C}_d^{-0.5}(\mathbf{d} - \mathbf{F}(\mathbf{m}))\right|\right|_2^2 + \alpha \left|\left|\mathbf{C}_m^{-0.5}(\mathbf{m} - \mathbf{m}_{ref})\right|\right|_2^2, \tag{B.1}$$

where the first term is the error-weighted misfit between original data ($\mathbf{d}$) and data predicted by the forward model ($\mathbf{F}(\mathbf{m})$) in a least-squares sense. The second term is the model misfit or regularization. In Eq. B.1,
$\mathbf{C}_d$ is the data error covariance matrix;
$\mathbf{C}_m$ is the model covariance matrix;
α is a trade-off parameter that weights the model regularization with respect to the data misfit term;
$\mathbf{m}_{ref}$ is a (uniform) reference model.

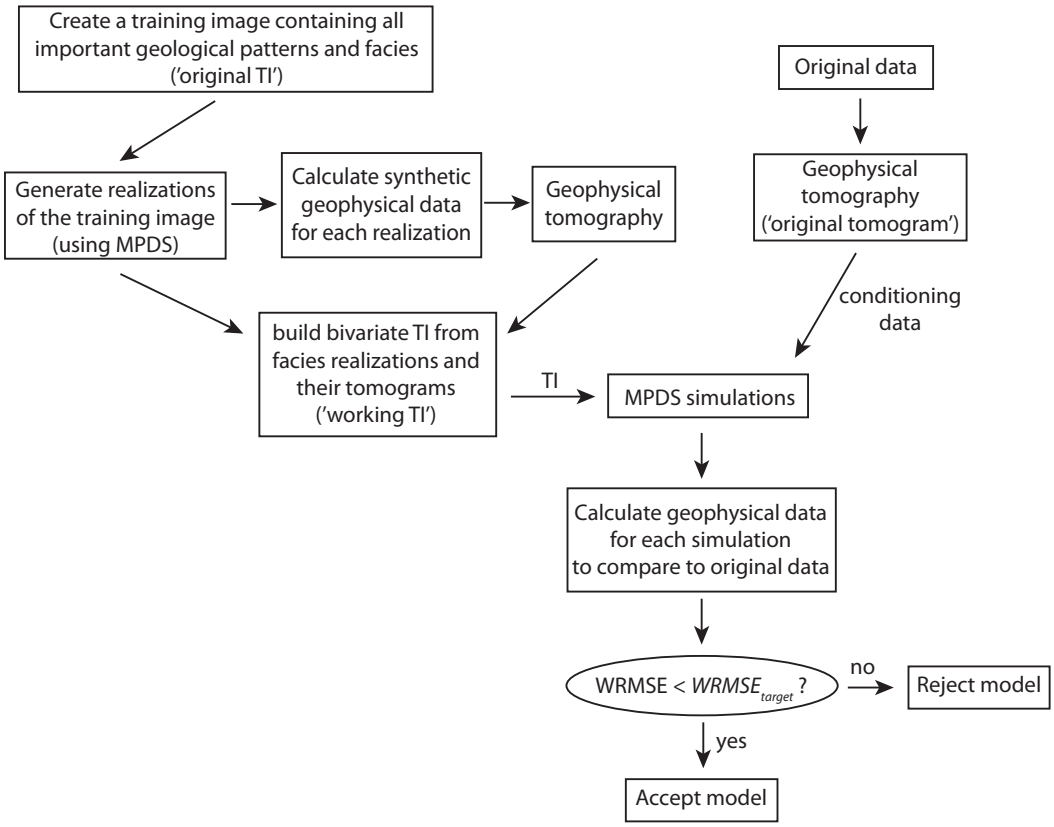

Figure B.1: Schematic workflow of the proposed conditioning method.

In our smoothness-constrained inversion, $\mathbf{C}_m$ is an anisotropic first-order differences operator, used to penalize abrupt changes between adjacent model parameters. In this study, horizontal-to-vertical anisotropy ratios of 3:1 and 10:1 were used for the first and the second application. Since the forward problem is non-linear, it is linearized about a reference model (for the first iteration) or the model of the previous iteration (for subsequent iterations), resulting in the following system of equations, that is solved with a conjugate gradient solver (Paige and Saunders, 1982)

$$\begin{pmatrix} \mathbf{C}_d^{-0.5}\mathbf{J}_k \\ \alpha\mathbf{C}_m^{-0.5} \end{pmatrix} (\Delta\mathbf{m}_{k+1}) = \begin{pmatrix} \mathbf{C}_d^{-0.5}\left[\mathbf{d} - \mathbf{F}(\mathbf{m}_k) + \mathbf{J}_k\Delta\mathbf{m}_k\right] \\ 0 \end{pmatrix}, \tag{B.2}$$

where k indicates the iteration index and $\mathbf{J}$ is the Jacobian matrix containing the elements $J_{ij} = \partial d_i/\partial m_j$. The model update we solve for, $\Delta\mathbf{m}_{k+1}$, is then used to construct a new model $\mathbf{m}_{k+1} = \mathbf{m}_{ref} + \Delta\mathbf{m}_{k+1}$. During the inversion, we decrease α consecutively from one iteration to the other in order to start with a smooth model where structure is gradually added to decrease the data misfit. The inversion is stopped when the weighted root-mean squared error $WRMSE_{target}$ is reached. The weighting by the data errors implies that for $WRMSE = 1$, on average all data are fit to their error levels. For a more rigorous description of the inversion scheme, we refer to Linde et al. (2006).

B.2.3 Multiple-point statistics simulation

Multiple-point statistical simulations describe the procedure to draw fields of facies or geological properties based on the spatial dependencies between a set of points larger than 2. Unlike simulations based on two-point statistics, such as Gaussian sequential simulation, it is insufficient to know only the pdf. Rather, patterns (so-called data events) of a certain predefined

size and density are extracted from a training image and the pixel of interest is simulated according to its statistics with respect to previously assigned pixels in the surrounding. The main advantage of MPS over two-point statistics is the ability to deal with non-stationarity and sharp interfaces and consequently allowing simulation of categorical facies models with curvilinear features, which can be crucial to adequately represent connectivity (e.g., Hu and Chugunova, 2008).

Classical MPS simulation codes (e.g., Strebelle, 2002) scan the training image for patterns and store the multi-point statistics of these in a structured tree. To overcome the limitations due to the large storage demand required by the trees, a list approach was developed by Straubhaar et al. (2011, 2013). In these classical implementations of MPS, the multiple grid approach is used to capture patterns of different scales within the TI (Tran, 1994). In this work, we use the MPDS algorithm *DeeSse*, which is an enhanced and commercialized version of the algorithm described by Mariethoz et al. (2010). Note that in the following, we use 'MPDS' to describe the general direct sampling procedure developed by Mariethoz et al. (2010), and '*DeeSse*' when we talk about the particular algorithm we used. In *DeeSse* (as in all direct sampling algorithms), the training image is scanned for data events in the simulation grid and directly simulates the pixel of interest, thus avoiding storage limitations, computation of pdf and the use of multiple grids. We give a brief summary of this sampling procedure here; for a complete description we refer to Mariethoz et al. (2010). If a simulation grid with nodes $\mathbf{x}$ is to be filled using the training image with nodes denoted $\mathbf{y}$, all locations $\mathbf{x}$ are visited successively. At each location $\mathbf{x}$, the n closest neighboring nodes that already have been assigned a value are found and the pattern of their locations and values defines one data event. Starting from a random location in the training image, the training image is systematically scanned for the particular data event. At every location $\mathbf{y}$, the distance between the data event and the event found in the TI is calculated. The data event found in the TI is kept and the node $\mathbf{x}$ of interest is simulated by copying the value from $\mathbf{y}$, if the calculated distance is below a threshold t. If a certain fraction f of the TI has been scanned without success, the event with the smallest distance is accepted and used for simulation. Choosing an appropriate distance measure is crucial for the simulation (e.g., Zhou et al., 2012). We applied an ℓ_1-norm for continuous variables and the sum of non-matching nodes for categorical variables. In addition, *DeeSse* allows to decrease the node density in the group of neighboring nodes to also consider nodes in the pattern recognition that are not directly clustered around the node of interest.

The resolution of a geophysical tomogram is not evenly distributed, but often varies markedly in space. This is due to (1) limitations in signal coverage, and (2) the fact that in non-linear problems, the resolution depends on the property distribution (e.g., Day-Lewis et al., 2005; Singha and Moysey, 2006; Singha et al., 2007). Therefore, scanning a training image horizontally, that is moving through a single 2-D image until a certain data event is scanned, can produce large bias since it treats the tomograms as if resolution is constant in space. To overcome this, we stacked a large number of realizations of a training image and the corresponding tomograms, similar to Zhou et al. (2012).
Technically speaking, a 3-D training image with four variables is set up. In our case, this constitutes two stacks of the catalogue of the categorical training images and the corresponding tomograms. The other two variables are the x- and y-coordinates of all nodes. By doing so, we treat the individual layers of the stacks as projections of a single plane and sample for 2-D patterns within these bodies. The x- and y-coordinates are dummy variables needed to keep the MPDS algorithm from seeking data events in 3-D. The 'working TI' is thus scanned until a data event is found for which

- the distance between the pattern in the $facies$ variable and the pattern formed by the

already simulated pixels in the simulations grid is below the threshold t_{facies},

- the distance between the pattern in the *tomogram* variable and the pattern in the conditioning tomogram is below the threshold $t_{tomogram}$ and
- the distances between the x- and y-coordinates of the pixel of interest and the undefined pixel in the data event are below thresholds t_x and t_y.

Once this data event is found in the *tomogram* variable of the 'working TI', the value for the pixel of interest is pasted from the $facies$ variable in the 'working TI' and copied to the simulation grid. Fig. B.2 depicts the general scanning procedure. The possibility to deal with multiple variables, whether categorical or continuous, is an advantage of MPDS over conventional MPS algorithms.

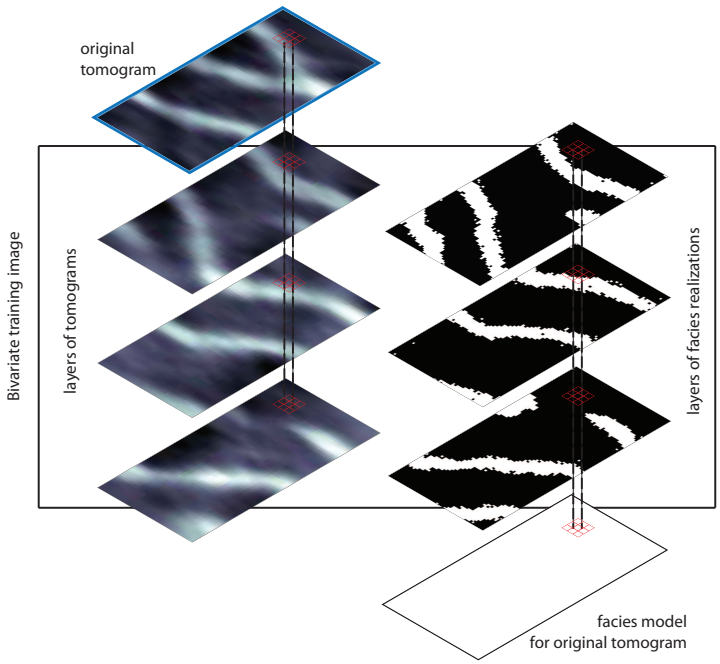

Figure B.2: The sampling principle used to account for the spatially variable resolution of geophysical tomograms.

B.2.4 Forward modeling

Though the framework of our approach is general in terms of the type of tomograms used, we will now focus on applications to crosshole GPR tomograms. In crosshole GPR experiments, transmitter and receiver antennas are placed in adjacent boreholes and the subsurface electromagnetic properties are sensed by measuring the transmitted signal for different transmitter-receiver configurations. Here, we focus on the first arrivals of the radar wave, which bear information about the radar wavespeed in the interwell region. In our first application, we assume that the relation between the facies type and the radar wavespeed is known (in practice, this could be known from laboratory experiments). In our second application, detailed information about the porosity and the hydraulic conductivity of each facies

type are available. The porosity information for the individual facies was translated into values of radar wavespeed v using (Sen et al., 1981)

$$\epsilon = \epsilon_w \phi^m \left(\frac{1 - \frac{\epsilon_m}{\epsilon_w}}{1 - \frac{\epsilon_m}{\epsilon}} \right)^m \tag{B.3}$$

and (e.g., Davis and Annan, 1989)

$$v = \frac{c}{\sqrt{\epsilon}}, \tag{B.4}$$

where ϵ_w, ϵ_m and ϵ [-] are the relative electrical permittivity of water, of the grains and of the bulk material, respectively. We here used $\epsilon_w = 81$ and $\epsilon_m = 3$. The porosity of each facies is ϕ, the so-called cementation factor m is typically taken to be 1.5 in unconsolidated media and the speed of light $c = 3 \times 10^8$ m/s.
Based on the velocity model, we determined radar traveltimes for a crosshole GPR experiment using

$$t_{GPR} = \int_{\mathbf{x}_1}^{\mathbf{x}_2} u(s) ds, \tag{B.5}$$

where $u(s)$ [s/m] is the radar slowness (i.e., the reciprocal of the radar wavespeed) along the trajectory s, starting at point $\mathbf{x}_1$ and ending at $\mathbf{x}_2$. The trajectories of the radar signals are dependent on the spatial distribution of the radar wavespeed. We solve the eikonal equation to get the spatial distribution of first-arrival traveltimes for each source location using the finite-difference algorithm of Podvin and Lecomte (1991) and perform ray-tracing by back-propagation for each receiver location (Vidale, 1988).
In the two applications considered below, transmitters and receivers were placed at the left and right boundary of the model domain at a spacing of 0.25 m and experiments were repeated for reciprocal transmitter and receiver boreholes so that transmitter positions become receiver positions and vice versa. As is common practice, only rays for which the transmitter-receiver path is inclined less than 50 to the horizon are taken into account (e.g., Peterson, 2001). This resulted in data sets of 2430 and 1388 traveltimes for the first and second application. All synthetic radar data were contaminated with Gaussian noise corresponding to a relative error of 1% of the measurement value.

B.3 Application to a channel scenario

We first applied the method to a two-facies case featuring channel-like structures (Strebelle, 2002). Radar wavespeeds of 70 and 80 m/μs were assigned to facies 0 and 1, respectively. A total of 1000 realizations of the original TI (generated with *DeeSse*, with a resolution of 0.1×0.1 m) and the corresponding tomograms build the bivariate working TI from which we sample. The number of neighboring nodes n in the simulation defines the order of the statistics that are honored during the simulation. For the facies variable, the number of neighboring nodes is set to $n = 25$ and the distance threshold to $t_{facies} = 0.04$. This means that if no more than one of the 25 nearest cells ($25 \times 0.04 = 1$) around the cell to be simulated differs from the cell values in the data event, the pattern for this variable is considered acceptable (Mariethoz et al., 2010). For the tomogram variable, the same parameters are used ($n = 25$ and $t_{tomogram} = 0.04$), whereas for the dummy-variables (i.e., the x- and y-coordinates), we chose $n = 1$ and the distance thresholds t_x and t_y so that the data events used for simulation are in an area that differs maximally three grid cells in x- and y-direction from the simulation cell (Tab. B.1).

Results are shown in Fig. B.3. The tomograms clearly highlight the smearing effect inherent in smoothness-constrained tomography: Even if the true image is built up only

Table B.1: Overview of MPDS parameters. The summation signs denote that the simple sum of non-matching nodes is taken as a distance measure. The scan fraction f is a parameter for the multivariate TI and is not chosen for each individual variables.

Variable	n	t	f	n-density	distance type
Case 1: Channels					
facies	25	0.04	0.1	1.0	$\sum$
tomogram	25	0.04	0.1	1.0	ℓ_1-norm
x-coordinate	1	0.06	0.1	1.0	ℓ_1-norm
y-coordinate	1	0.03	0.1	1.0	ℓ_1-norm
Case 2: Herten					
facies	25	0.04	0.1	0.5	$\sum$
tomogram	25	0.04	0.1	0.5	ℓ_1-norm
x-coordinate	1	0.03	0.1	1.0	ℓ_1-norm
y-coordinate	1	0.02	0.1	1.0	ℓ_1-norm

by two facies (black and white), classical tomography always produces shades of grey. A total of 100 MPDS simulations, conditioned to the original tomogram, were generated with *DeeSse*. For each simulation, we calculated the radar traveltimes and compared them to the true radar data. Keeping only simulations for which the WRMSE is below 1.0, left us with 68 simulations (Fig. B.4a). For illustration of the MPDS simulations with different conditioning (Fig. B.5), we adapted the so-called movie strategy for the visualization of stochastic inversion results (Tarantola, 2005). As expected, the conditioning to a tomogram brings the simulations closer to the true facies model. For simulated facies models that fail to predict the true radar traveltimes, the main structures of the true model are reproduced, but they often fail to represent the connectivity of the channels. The accepted facies models are visually very similar to the true facies model and the connectivity of the channels is recovered.

B.4 Application to a multi-facies aquifer analog

As a more realistic example, we applied our methodology to a facies model based on a detailed gravel aquifer mapping study (Bayer et al., 2011). The underlying model includes features that are typical for aquifers built up by alluvial gravel deposits. These types of aquifer bodies are of high importance in many parts of the world due to their generally high storage capacity, high permeability and accessibility.

B.4.1 The Herten aquifer analog

The Herten aquifer model is based on detailed mapping of sedimentary deposits in a gravel pit in southwest Germany. Bayer et al. (2011) set up a geological map of the deposits to establish a realistic aquifer analog. The analog is a 3-D representation of the gravel body, which was made possible by mapping during ongoing excavation. The deposits are mainly formed by unconsolidated fluvio-glacial and fluvial sediments in the Rhine basin (Bayer et al., 2011). The dominant structural elements, that are, erosional surfaces, scours, gravel sheets, are typical for sediments deposited in braided river environments. Ten different subunitary lithofacies are defined to describe the gravel deposits. For every lithofacies, porosity and hydraulic conductivity values and ranges are provided by experimental studies of Bayer (2000), Heinz et al. (2003) and Kostic et al. (2005).

The detailed aquifer analog was digitized to use it as a base for geostatistical simulations (Comunian et al., 2011). We used a 2-D slice of the Herten analog as a conceptual model, that is, as an original training image. The dimensions of one slice of the analog are 7×16 m. We generated 1000 realizations of the original training image with *DeeSse*. All realizations

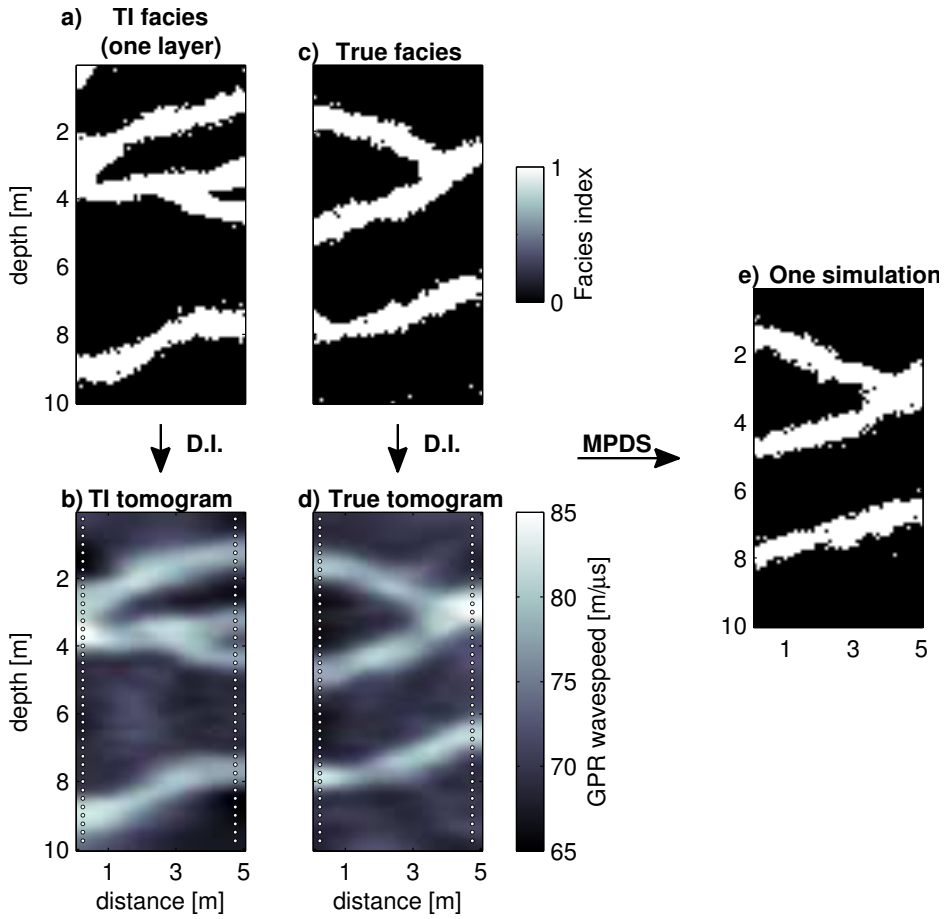

Figure B.3: (a) and (b): One layer of the bivariate training image; (a) facies and (b) corresponding tomogram. (c) The true facies we seek to recover and (d) its tomogram, which is used as conditioning data in the MPDS simulations. D.I. stands for deterministic inversion. White dots indicate transmitter and receiver positions for the radar traveltime calculations. (e) One simulated facies model.

show some of the geological patterns and of the lithofacies found in the analog, but the direction and dimension of these patterns varies between the realizations. The ensemble of realizations consequently represents a rather wide variety of geological scenarios.

For this study, the available facies realizations were simplified in that (1) only a $7\times4\,\mathrm{m}$ part of the images were considered for simulation (keeping the original cell size of $0.05\times0.05\,\mathrm{m}$) and (2) the number of facies was reduced from 10 to 5 by clustering similar facies together. The clustering is based on both similarities in geological texture (Bayer et al., 2011), as well as in hydrogeological properties (Fig. B.6). The modified facies and their properties are summarized in Tab. B.2.

B.4.2 Results

We chose slightly different parameters in the MPDS algorithm to account for the higher degree of complexity in this multi-facies case. To define a data event for the facies and radar wavespeed ('tomogram') parameters, the maximum density of neighboring nodes within the search environment was reduced to 0.5 to capture more complex patterns and to prevent just reproducing overly fine small-scale variations. This means that we consider n out of the $2n$ closest cells for pattern definition. Again, the distance thresholds for the x- and y-

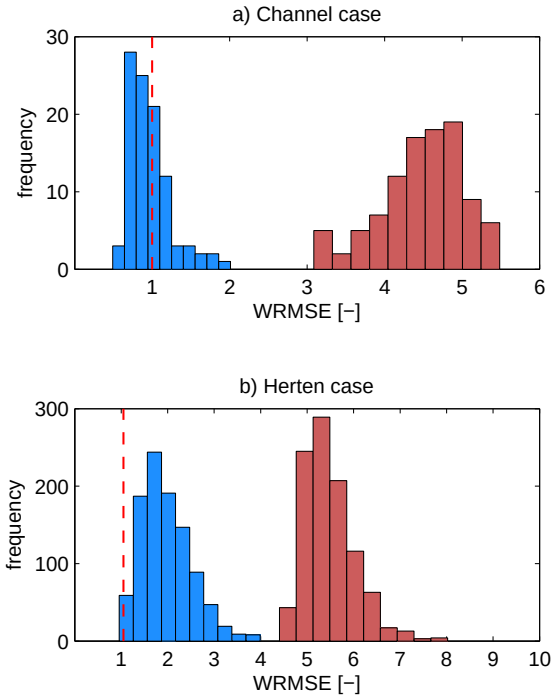

Figure B.4: (a) Distribution of weighted root-mean squared errors (WRMSE) of radar traveltimes for 100 unconditioned (red) and conditioned (blue) MPDS simulations for the channel scenario. The acceptance threshold (dashed vertical line) is the same as for the deterministic inversion. 68% of the conditioned simulations predict traveltimes with an WRMSE below the indicated threshold of 1.0 (red line). (b) As for a, but for the Herten case and for 1000 unconditioned and conditioned simulations. Ten conditioned simulations predict the traveltimes with an WRMSE below the threshold of 1.05.

coordinates, t_x and t_y, were chosen such that the node to be simulated is constrained solely by data events occurring in the same region of the grid (Tab. B.1).

Unlike the tomograms, the simulations feature the desired sharp interfaces between facies (Fig. B.7). Typical facies-in-facies patterns (e.g., facies 1 and 5 are frequently connected in tight layering) are reproduced rather well. We simulated 1000 facies models, of which 10 predicted the radar traveltimes within the error bounds (Fig. B.4b). Different simulations that predict the radar data within the error bounds are displayed in Fig. B.8.

We face two different problems when simulating facies models based on the Herten TI. (1) The complexity of the realizations leads to higher computational costs since it takes a longer time until a suitable data event is found in the TI. Also, the choice of the simulation parameters becomes more crucial than in the simpler channel case. (2) Many conditioned simulations feature small-scale structures that are not present in the true facies model. Together with the wide range of radar wavespeeds (much larger than for the channel scenario), this leads to a high number of conditioned simulations that fail to predict the radar data within the defined error bounds. Thus, more simulations are required to get a decent num-

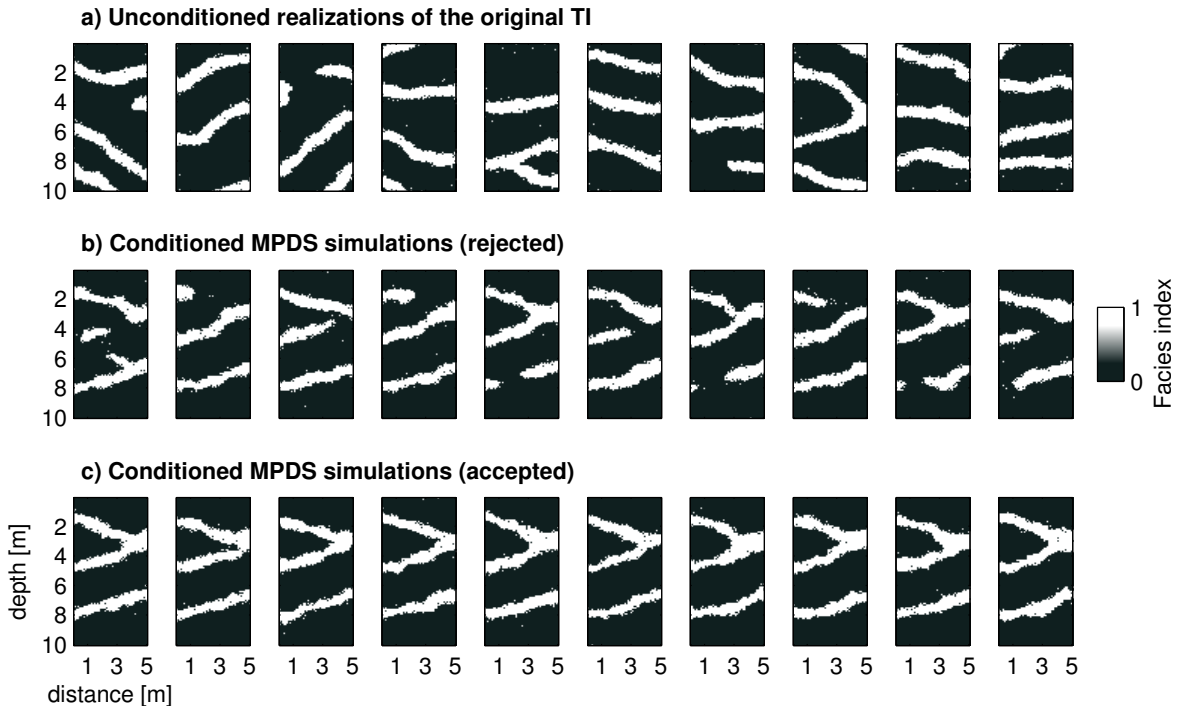

Figure B.5: Different MPDS simulations of the channel scenario in movie illustration. (a) Unconditioned simulations as they are used as layers of the 'working training image'. (b) MPDS simulations conditioned to a geophysical tomogram. These simulations failed to predict the radar data satisfactorily and were rejected. (c) As for (b) but with radar traveltimes for which the WRMSE is below the threshold, these simulation were accepted.

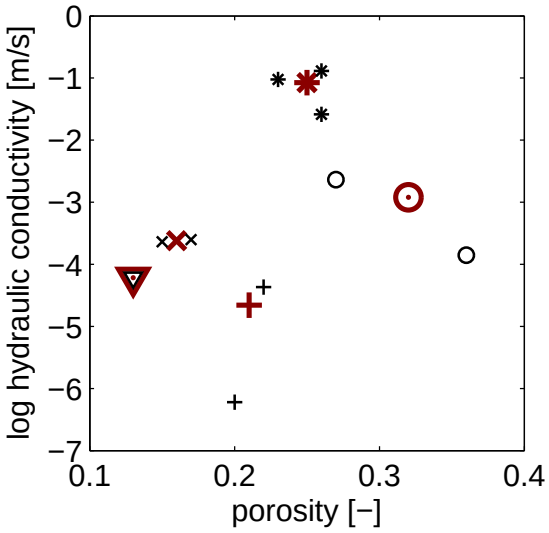

Figure B.6: Clustering of the original 10 facies to the reduced facies distribution. Vertical crosses (denoted facies 1 in Tab. B.2), circles (facies 2), triangles (facies 3), diagonal crosses (facies 4) and asterisks (facies 5) refer to the simplified facies model. In red the new facies.

ber of conditioned simulations that honor the geophysical data and not only patterns in the tomograms.

Table B.2: Description of the different facies types

Facies index	Description	Hydraulic conductivity K [m/s]	Porosity [-]
1 (+)	Alternating gravel, bimodal basal subunit (sand, silt, clay)	2.2×10^{-5}	0.21
2 (○)	Well-sorted gravel and sand	1.2×10^{-3}	0.32
3 (▽)	Sand-rich, poorly sorted, matrix supported gravel	6.1×10^{-5}	0.13
4 (×)	Poorly sorted, matrix supported gravel	2.4×10^{-4}	0.16
5 (∗)	Alternating gravel, open framework	8.4×10^{-2}	0.25

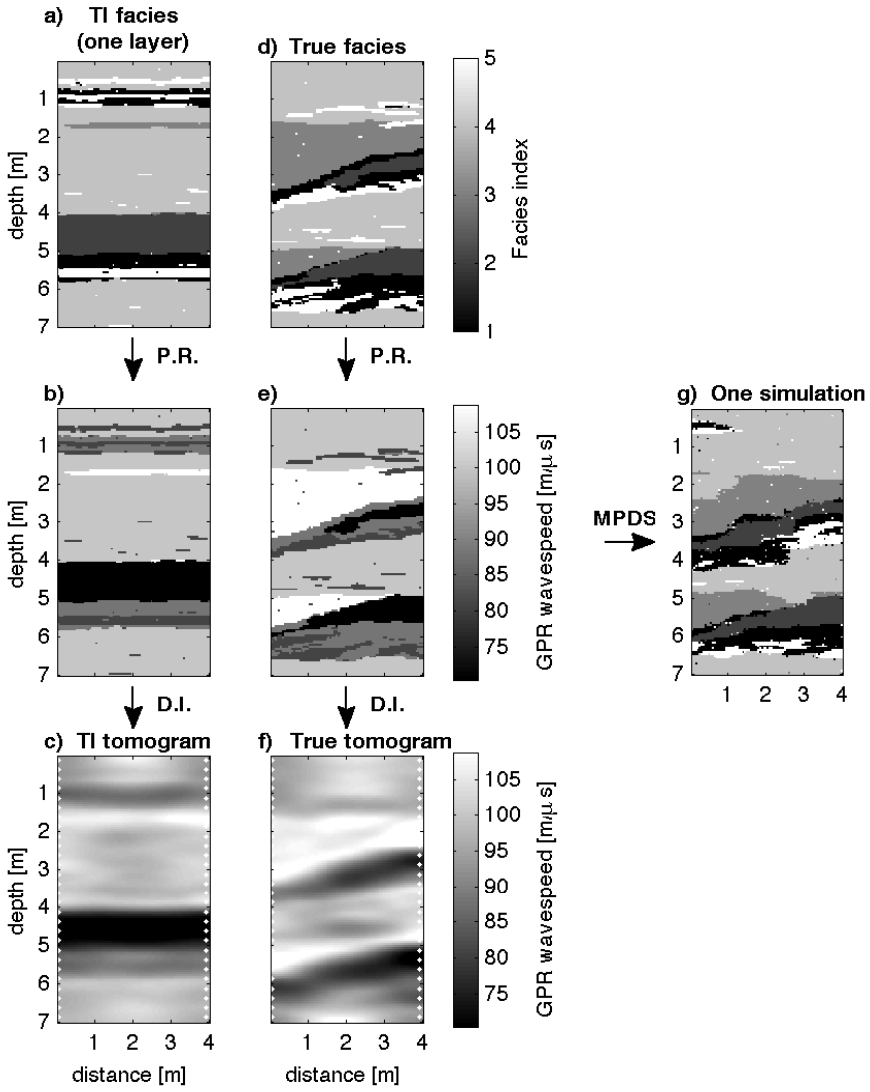

Figure B.7: (a) One layer of the facies training image and (b) corresponding radar wavespeed distribution, obtained by applying a petrophysical relation (P.R.). (c) The corresponding tomogram results from deterministic inversion (D.I.) of (b). (d) The true facies we seek to recover, (e) its wavespeed distribution and (f) its tomogram, which is used as conditional data in the MPDS simulations. White dots indicate transmitter and receiver positions for the radar traveltime calculations. (g) One simulated facies model.

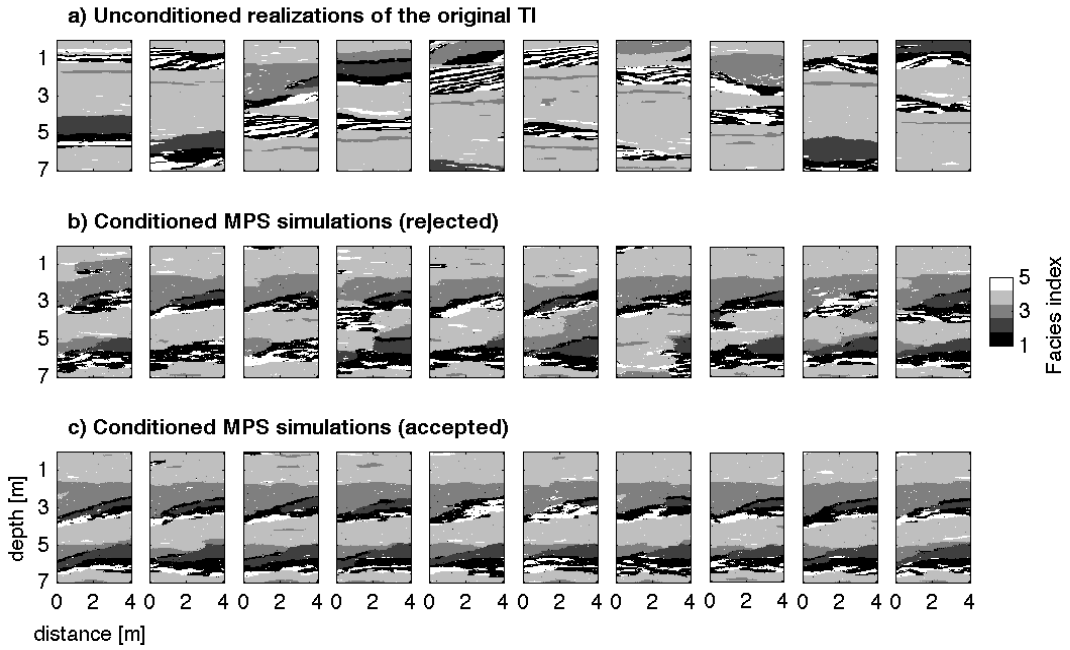

Figure B.8: Different MPDS simulations of the Herten training image in movie illustration. (a) Unconditioned simulations as they are used as layers of the training image. (b) MPDS simulations conditioned to a geophysical tomogram. These simulations failed to predict the radar data satisfactorily and were rejected. (c) As for (b) but with radar traveltimes for which the WRMSE is below the threshold, these simulation were accepted.

B.5 Impact on flow predictions

The main purpose for determining the spatial distribution of lithofacies and their hydrological parameters with high resolution is to use these models for prediction of flow or solute transport. To demonstrate the improvement of the models resulting from bivariate simulation, we conducted synthetic tracer experiments and compared unconditioned (i.e., the layers of the facies training image) and conditioned MPDS simulations with respect to their transport predictions. Tracer breakthrough at certain locations on the right boundary is calculated for an injection of a conservative tracer at the left boundary of the domain that is driven by a horizontal head gradient. We used the finite difference code *MaFloT* (Matlab Flow and Transport (maflot.com), Künze and Lunati, 2012) to solve the advection-dispersion equation and calculate the tracer concentration evolution during continuous tracer injection. For dispersivities in longitudinal and transversal direction, we used $\alpha_l = 10^{-2}\,\mathrm{m}$ and $\alpha_t = 10^{-3}\,\mathrm{m}$ and for molecular diffusion $D = 10^{-9}\,\mathrm{m}^2/\mathrm{s}$.

For the channel scenario, hydraulic conductivity values of $K_0 = 10^{-5}\,\mathrm{m/s}$ and $K_1 = 10^{-2}\,\mathrm{m/s}$ were assigned to facies 0 and 1, respectively, to model highly conductive channels in a low-conductivity matrix. The porosity was kept constant at $\phi = 0.3$. For the aquifer analog (Herten), hydrogeological facies information, that is, the porosity and the hydraulic conductivity values were directly available from an outcrop data set (Tab. B.2). In both cases, the uncertainty in the predictions for the tracer breakthrough is clearly reduced for the simulations that are conditioned to a tomogram (Fig. B.9). For simulations that predict the radar traveltimes within the error bounds, the uncertainty is further reduced. There is no indication of modeling bias, the predicted breakthroughs are consistent with the breakthrough curves of the true models.

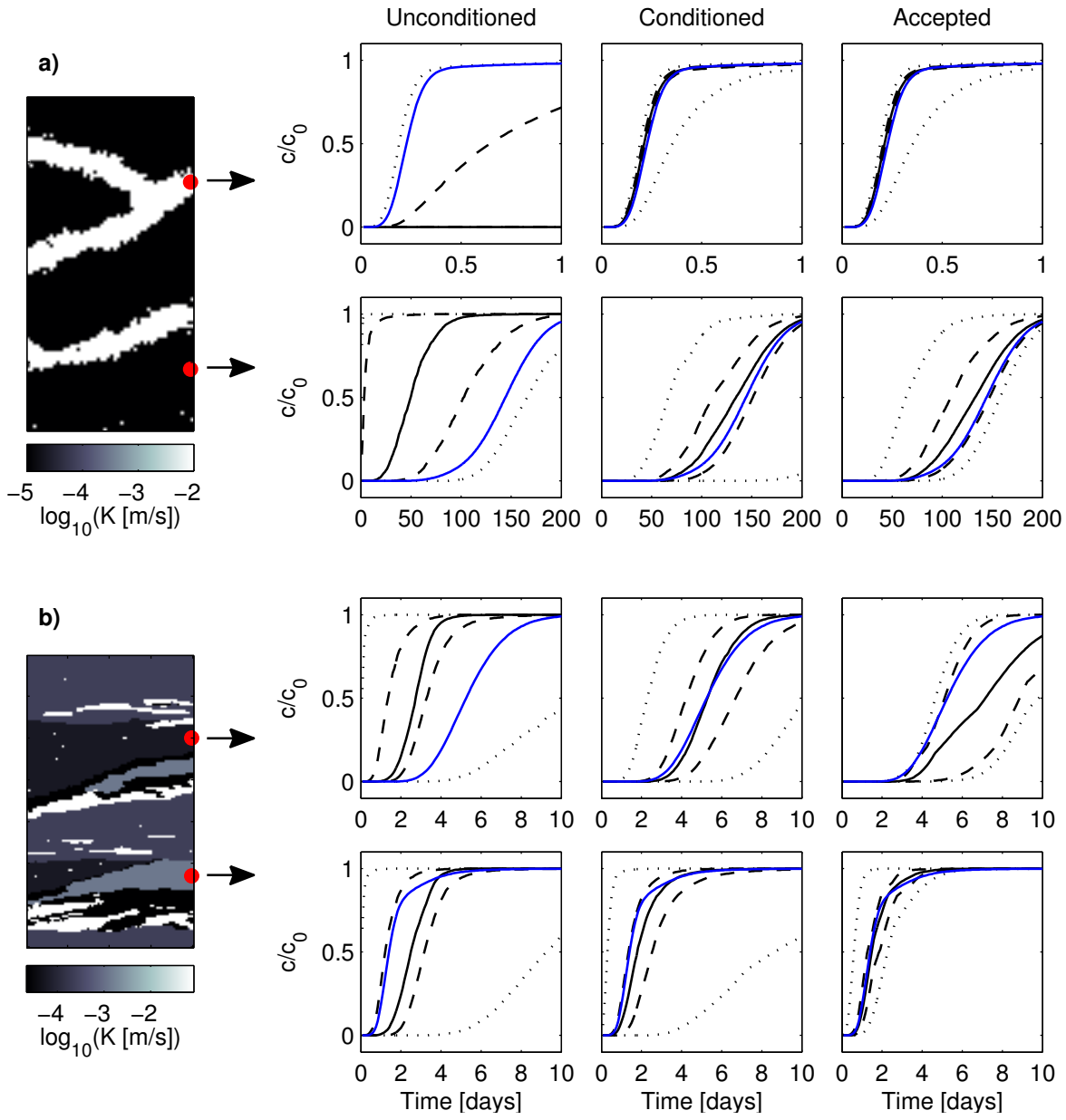

Figure B.9: Statistics of calculated breakthrough curves for (a) channel scenario and (b) Herten scenario. Solid, dashed and dotted lines depict mean, 50 % quantiles and 95 % quantiles. Blue line: breakthrough of the true model.

B.6 Discussion

We show that combining MPS simulations and classical geophysical tomography can produce models that overcome some of the shortcomings of both purely tomogram-based inverse models and unconditioned MPS simulations. Unlike classical geophysical tomography, the presented method allows to recover models that feature expected geological patterns and sharp interfaces between facies elements. At the same time, the inverse models are constrained by geophysical data whose information is spread over large parts of the model domain, which is not the case for unconditioned simulations or simulations conditioned to point data only. When applied in practice, the following issues must be taken into account:

- As for all methods that include MPS simulations, the choice of an adequate training image is of utmost importance. If no auxiliary geological information is available, MPS simulations carry a substantial risk of creating illusive results. This was observed in tests performed by applying the method with a poorly chosen TI: We tried to represent an object-based model featuring lenses with the same channel TI as used in the first application. The resulting models reproduce the overall distribution of the two facies but feature channels instead of lenses and fail to predict the data well. Problems in finding models with satisfactory data fit can thus be taken as an indicator for an incomplete or biased TI. A detailed assessment about how a poorly chosen TI affects the quality of MPS simulations is beyond the scope of this study and is left as a recommendation for further research.

- In the current formulation, the GPR signals are approximated by curved rays rather than accounting for the full waveform signal. This approximation has been proven robust in many applications (e.g., Davis and Annan, 1989; Alumbaugh et al., 2002; Dafflon and Barrash, 2012). New developments in full waveform inversion techniques (e.g., Meles et al., 2012; Klotzsche et al., 2012) promise improved inversion results for certain settings if further attributes of the waveform than just the first arrival are considered.

- In the two scenarios presented, we assumed the petrophysics (i.e., the relationship between the lithological properties and the geophysical parameters) to be perfectly known and exact, which is rarely the case in practice. We postulate that the uncertainty of the petrophysics can be accounted for by using a much larger TI that contains several tomograms for one facies model, each calculated with different petrophysical parameters. We leave this for further research.

- We applied the method to crosshole GPR tomograms, but the method is general with regard to the measured geophysical signal. Applications to other geophysical or hydrological tomographic methods, such as electrical resistivity tomography (ERT), seismic tomography or hydraulic tomography are straight-forward.

The 'weight' of the facies TI versus the conditioning data can be changed by varying the MPDS parameters. For a smaller distance threshold t, for example, patterns in the TI are reproduced more accurately. This requires larger parts of the TI to be scanned in order to find a matching data event. Thus in general, the choice of the sampling parameters is a compromise between reproduction accuracy and computational effort and to large parts controls the CPU burden in our workflow (Meerschman et al., 2013). Another computationally expensive step is creating the working TI, since each generation of a MPDS realization of the original TI is followed by a forward calculation and a geophysical inversion. However, the inversion parameters are taken as in the inversion of the field data, which implies that constructing the working TI demands no extensive parameter study and needs to be done only once.

Like purely stochastic inversions based on Markov chain Monte Carlo, the proposed method provides the user with an ensemble of possible models (e.g., Cordua et al., 2012). Since the models are not extracted from a Markov chain and are intrinsically uncorrelated, our method is significantly faster than stochastic inversions. Applying the method to a 3D case is straightforward, but CPU restrictions are non-negligible.

The quality of a geophysical inverse model strongly depends on the choice of the inversion parameters. This problem is largely circumvented by our approach, since we do not use the tomogram for interpretation but for conditioning of a geological conceptual model. We implicitly account for the resolution limitations of tomograms, thus there is no need for

an explicit resolution analysis. Also, we create several possible models, which are used to assess the uncertainty in the modeling results. This highlights the improvement over interpretation of individual tomograms, where accurate estimation of both resolution and uncertainty generally remains difficult (e.g., Williamson, 1991; Day-Lewis and Lane, 2004; Doser et al., 1998).

The resulting models reproduce the large- and small-scale structures of the true models rather well. They can be used for geological interpretation or as starting points for further inversion or conditioning. Additional reduction of the uncertainty in the parameter estimations might be achieved by including other types of geophysical or hydrological data in the presented workflow.

B.7 Conclusions

A method is presented to condition MPS simulations to geophysical tomograms. We create facies simulations by scanning a bivariate training image, constructed from geological facies realizations and their tomograms, for patterns found in a conditioning tomogram. This allows to exploit the advantages of both tomography and multiple-point geostatistical simulation. MPS simulations produce detailed facies models with sharp interfaces, where realistic geological constraints are honored by sampling from a TI. Conditioning the simulations to a tomogram, that resolves large-scale features over large parts of the model domain, allows to adequately recover realistic facies models. We found running a geophysical forward calculation on the proposed simulations and comparing the response to the measured data to be a valid acceptance criterion. Unlike the classical geophysical approach where the interpreter is left with one tomographic image, the method provides an ensemble of possible facies models that allows for uncertainty assessment. Testing the conditioned simulations with respect to their transport predictions reveals that they correctly predict the transport characteristics of the true model and that the uncertainty is reduced compared to predictions of unconditioned MPS simulations.

Acknowledgements

This research was funded by the Swiss National Science Foundation (SNF) and is a contribution to the ENSEMBLE project (grant no. CRSI22_132249). We would like to thank Philippe Renard for many helpful suggestions during the course of this study. We thank the three anonymous reviewers for their constructive comments that helped to improve the manuscript.

Appendix C

Supplementary figures for the chapter: A sensitivity analysis of the Direct Sampling algorithm

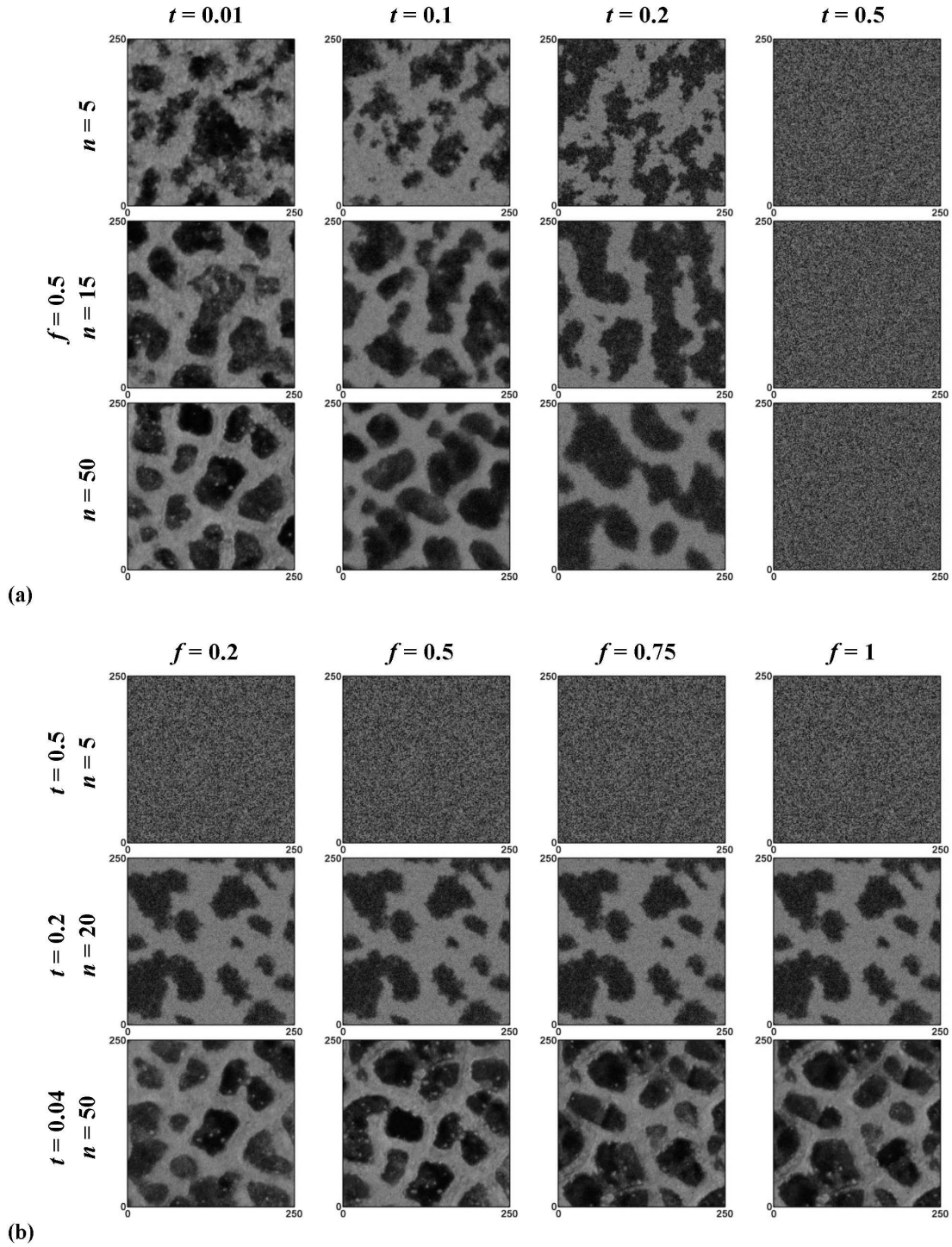

Figure C.1: (a) First out of 10 unconditional simulations illustrating the effect of parameters t and n with f=0.5 and (b) first out of 10 unconditional simulations illustrating the effect of f for constant t and n based on the continuous ice-wedges TI (Fig. 2.1a)

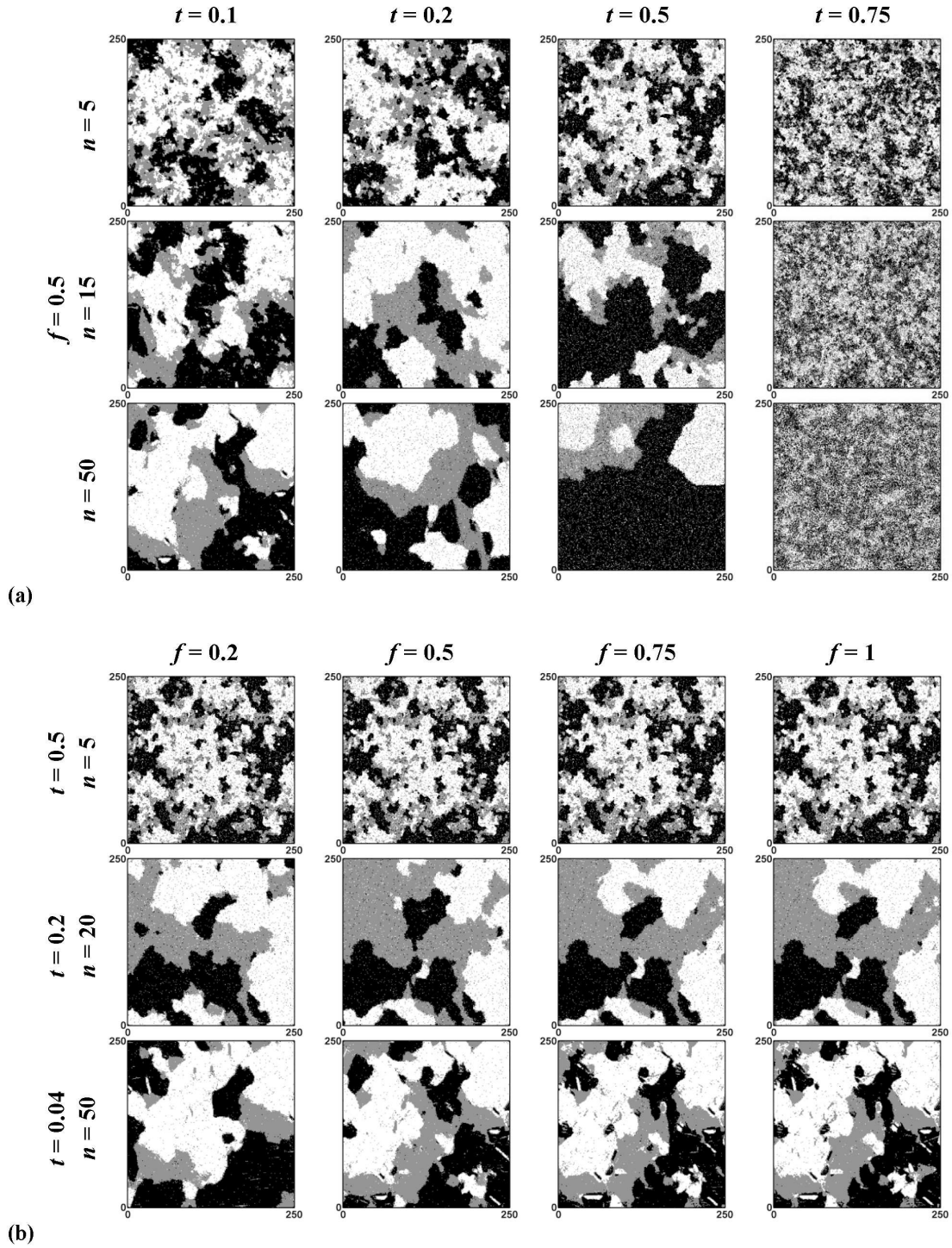

Figure C.2: (a) First out of 10 unconditional simulations illustrating the effect of parameters t and n with f=0.5 and (b) first out of 10 unconditional simulations illustrating the effect of f for constant t and n based on the categorical marble TI (Fig. 2.1d)

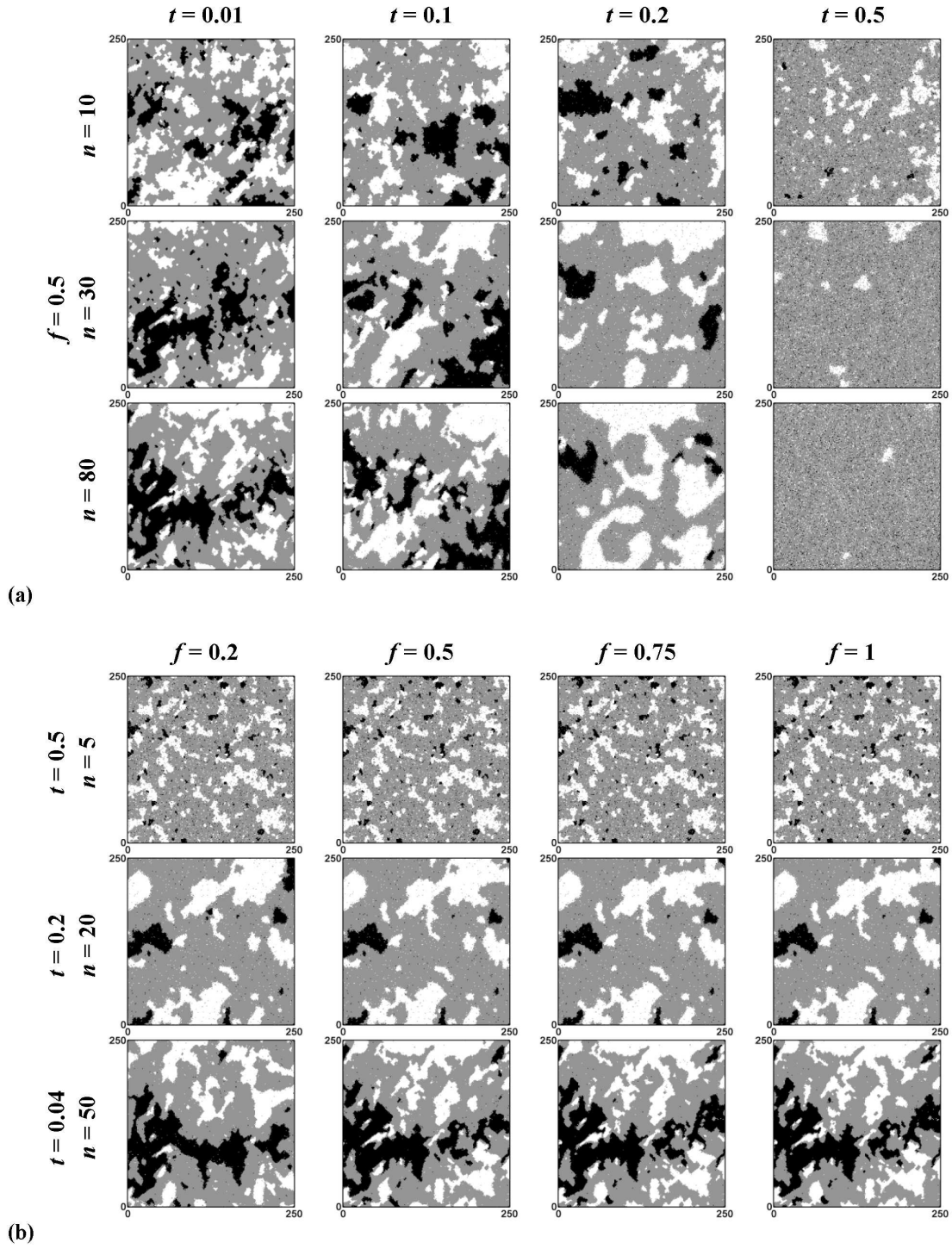

Figure C.3: (a) First out of 10 unconditional simulations illustrating the effect of parameters t and n with f=0.5 and (b) first out of 10 unconditional simulations illustrating the effect of f for constant t and n based on the categorical snow crystals TI (Fig. 2.1f)

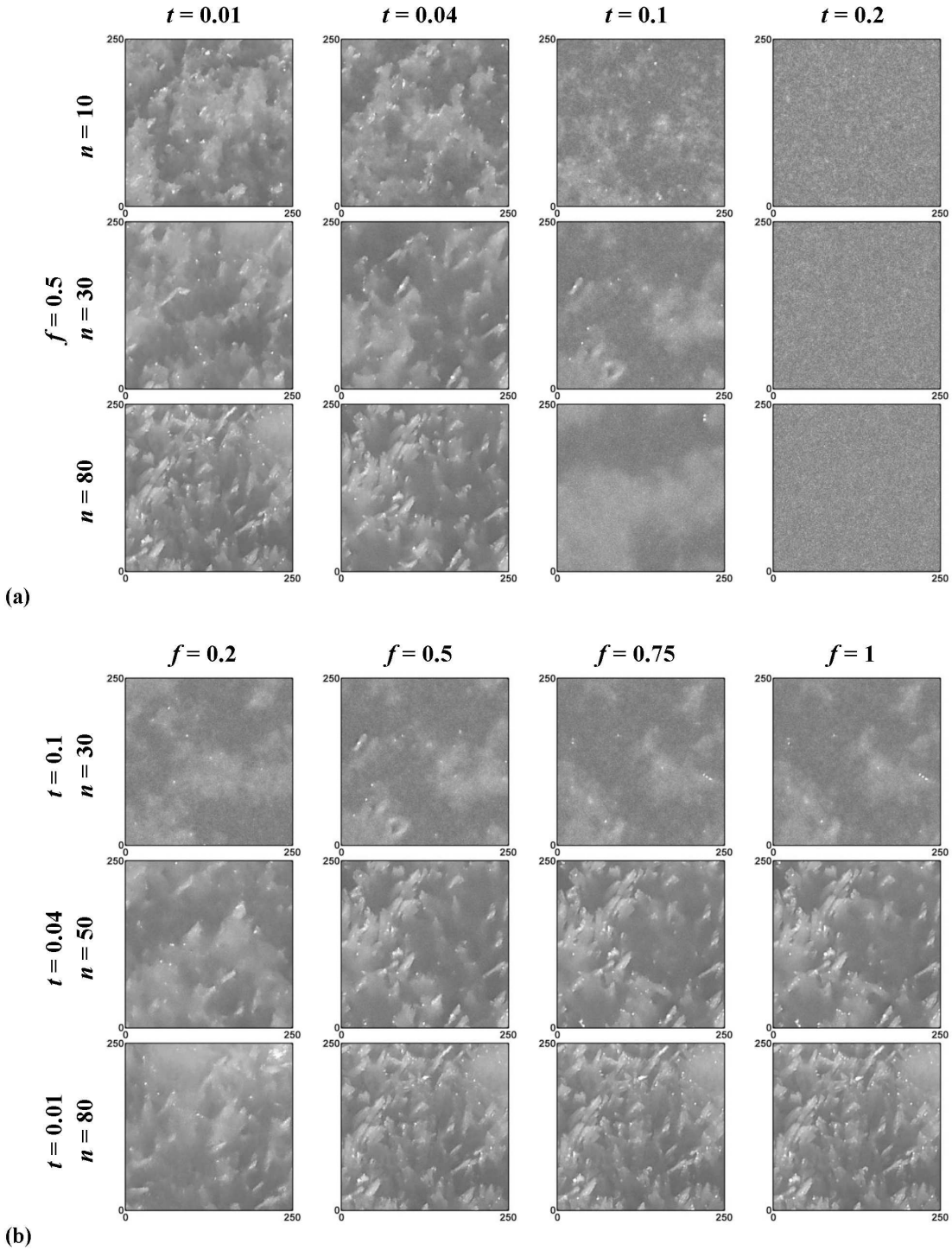

Figure C.4: (a) First out of 10 unconditional simulations illustrating the effect of parameters t and n with f=0.5 and (b) first out of 10 unconditional simulations illustrating the effect of f for constant t and n based on the continuous snow crystals TI (Fig. 2.1e)

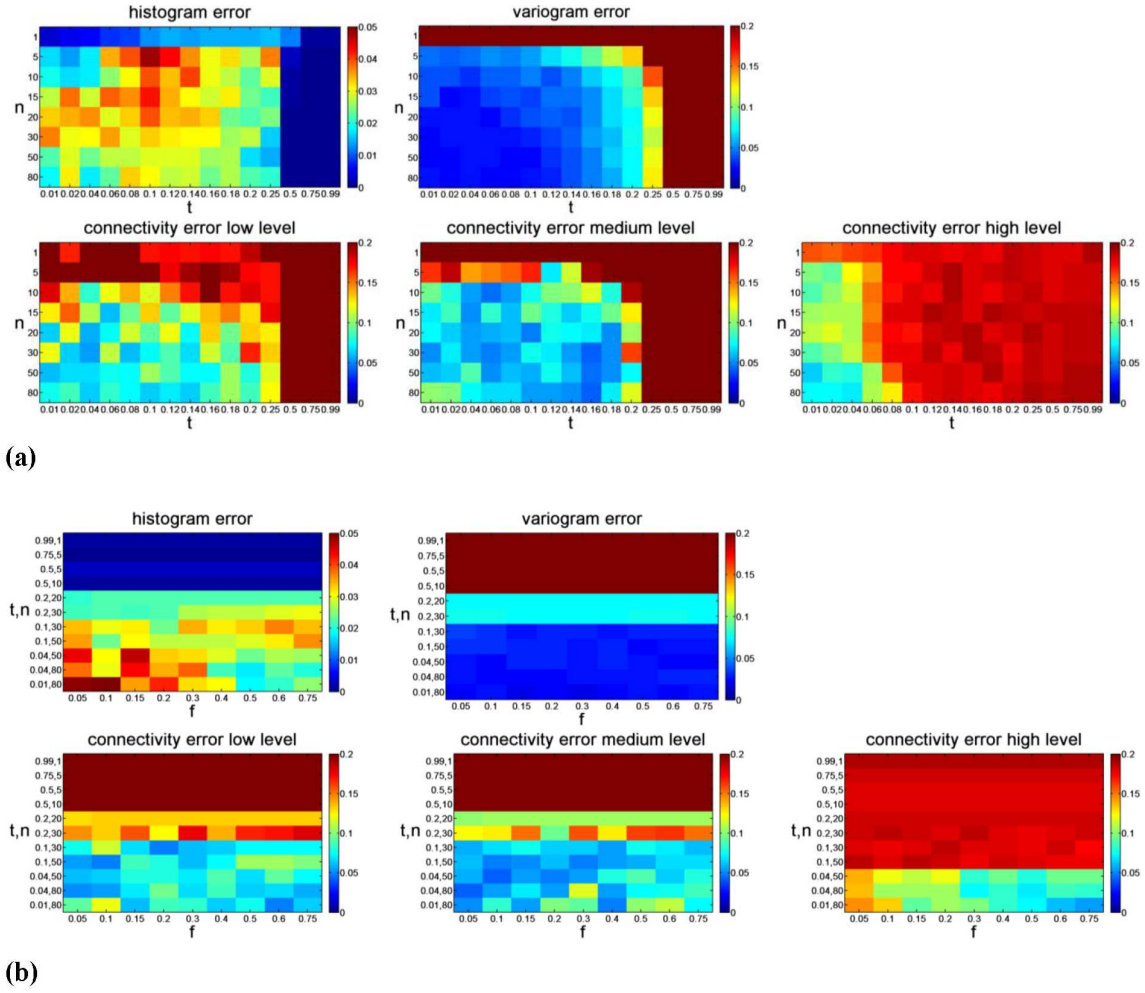

Figure C.5: Influence of (a) t and n (for f=0.5) and (b) f on the quality indicators based on the continuous ice-wedges TI (Fig. 2.1a)

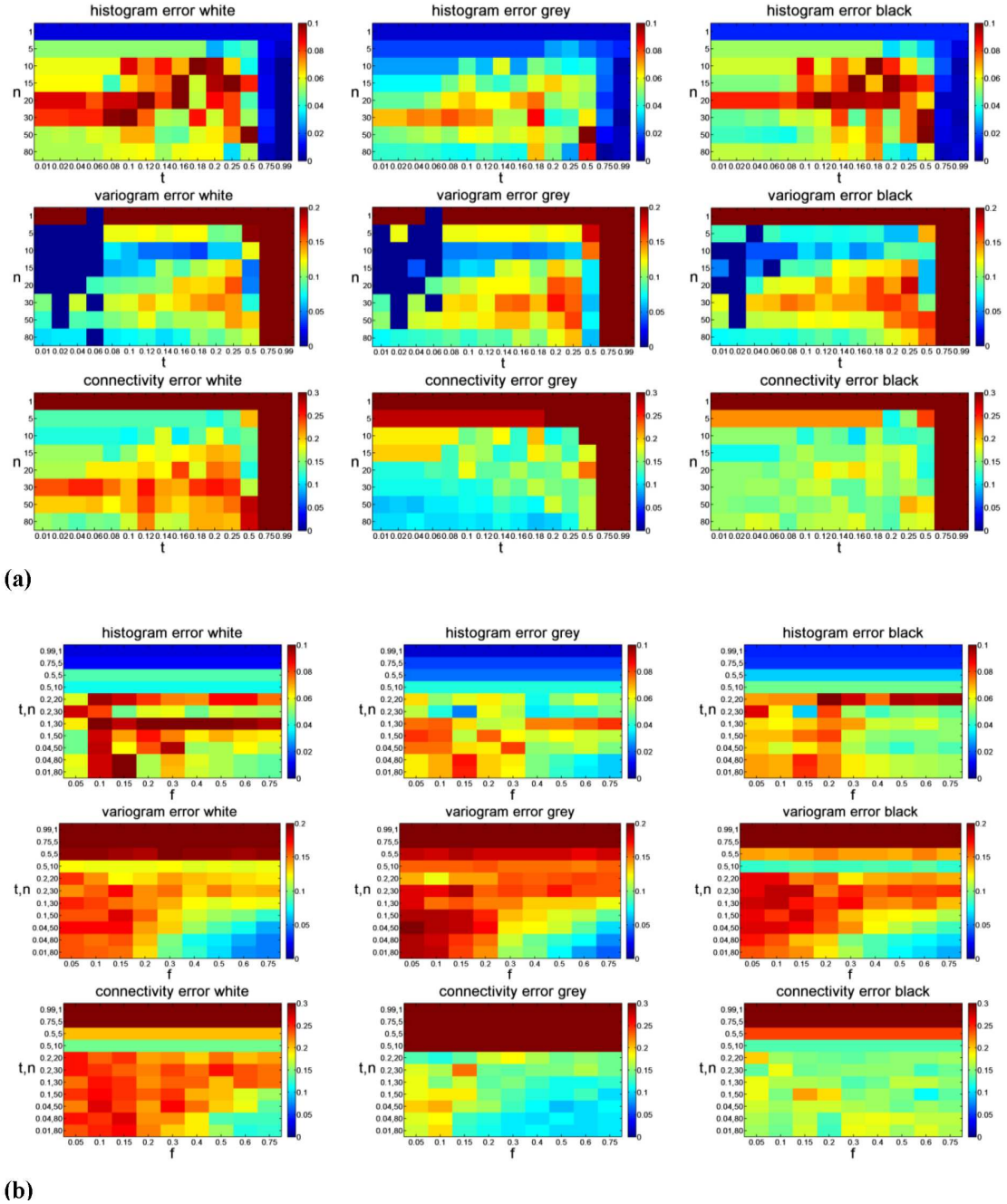

Figure C.6: Influence of (a) t and n (for f=0.5) and (b) f on the quality indicators based on the categorical marble TI (Fig. 2.1d)

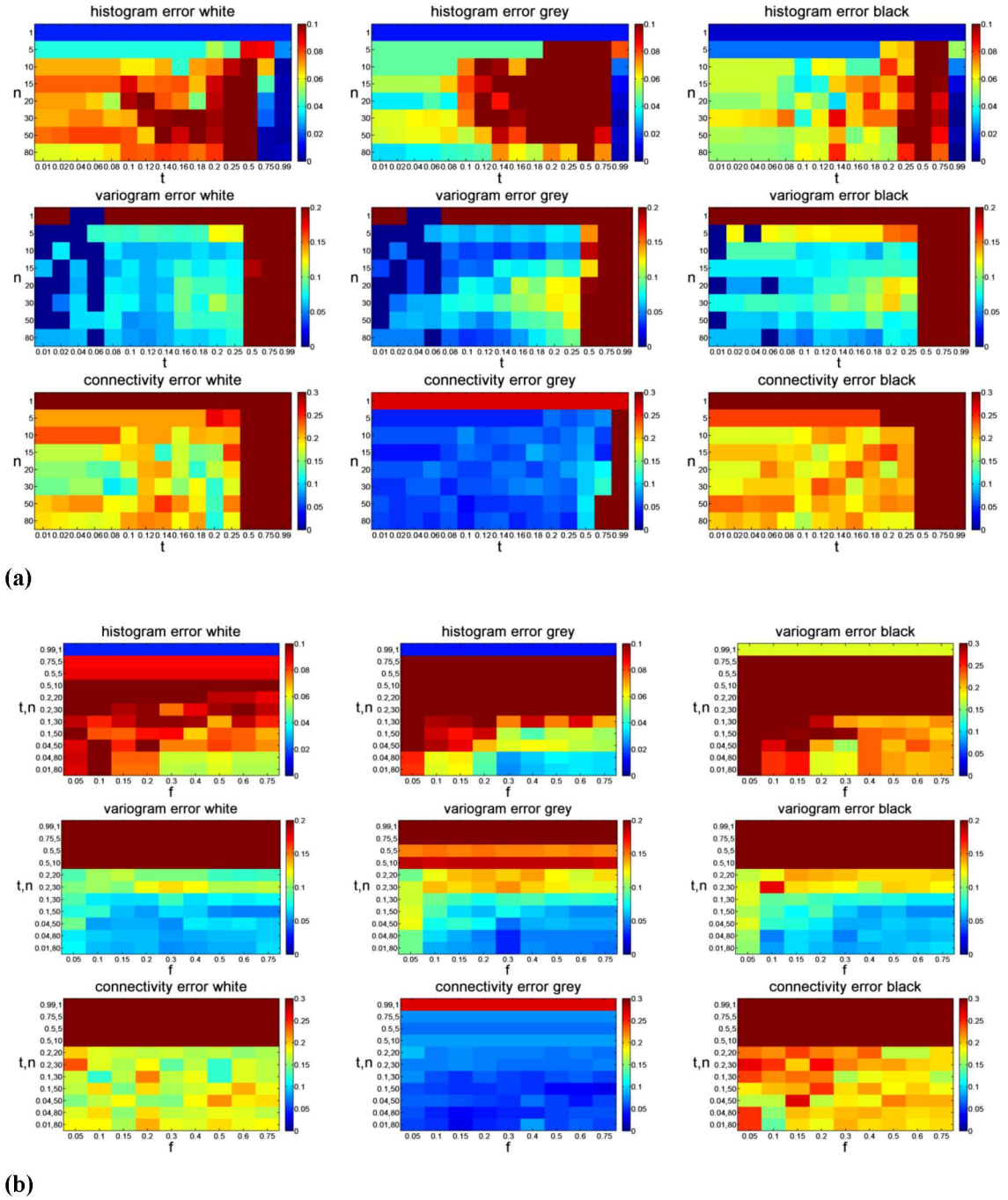

Figure C.7: Influence of (a) t and n (for f=0.5) and (b) f on the quality indicators based on the categorical snow crystals TI (Fig. 2.1f)

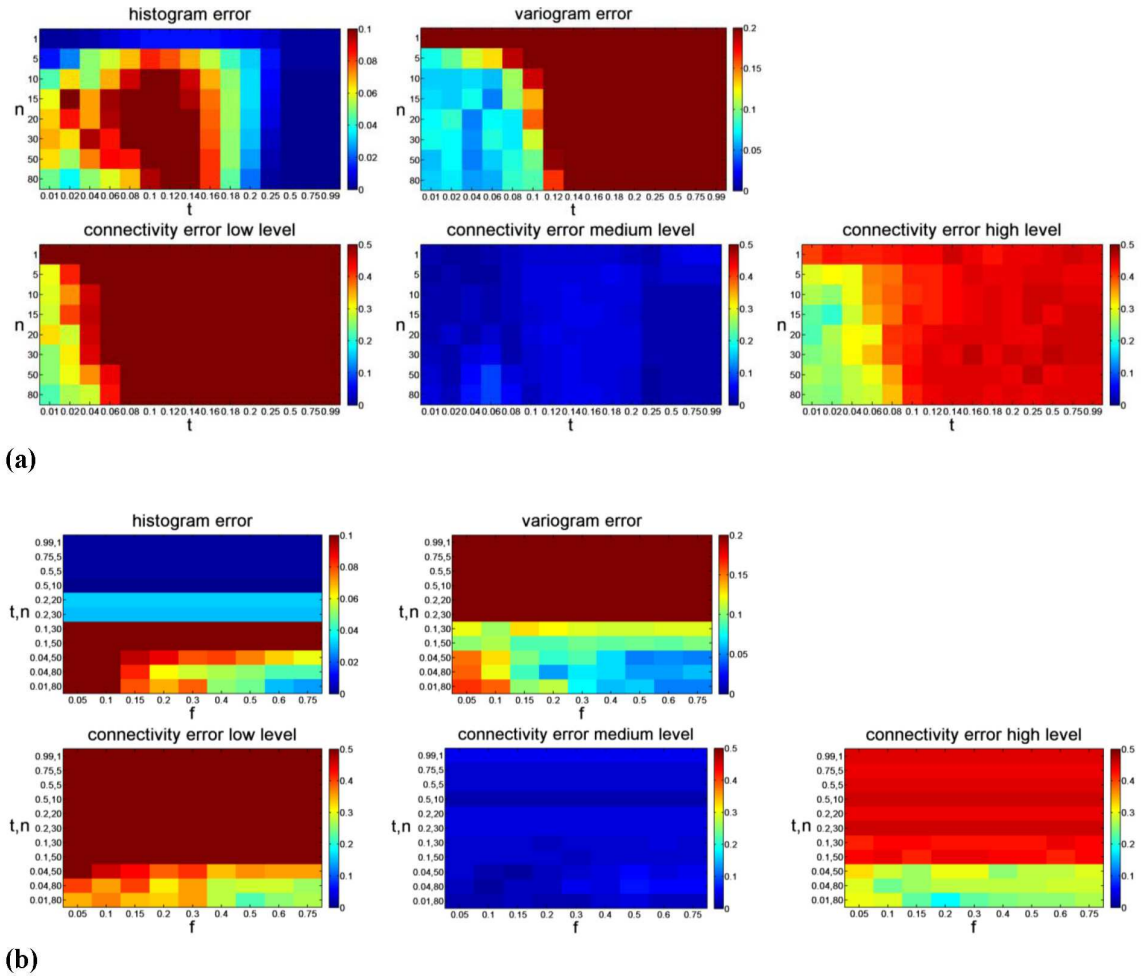

Figure C.8: Influence of (a) t and n (for f=0.5) and (b) f on the quality indicators based on the continuous snow crystals TI (Fig. 2.1e)

Appendix D

Supplementary figures for the chapter: Influence of conceptual model choice on contaminant transport uncertainty forecasting in braided river aquifers

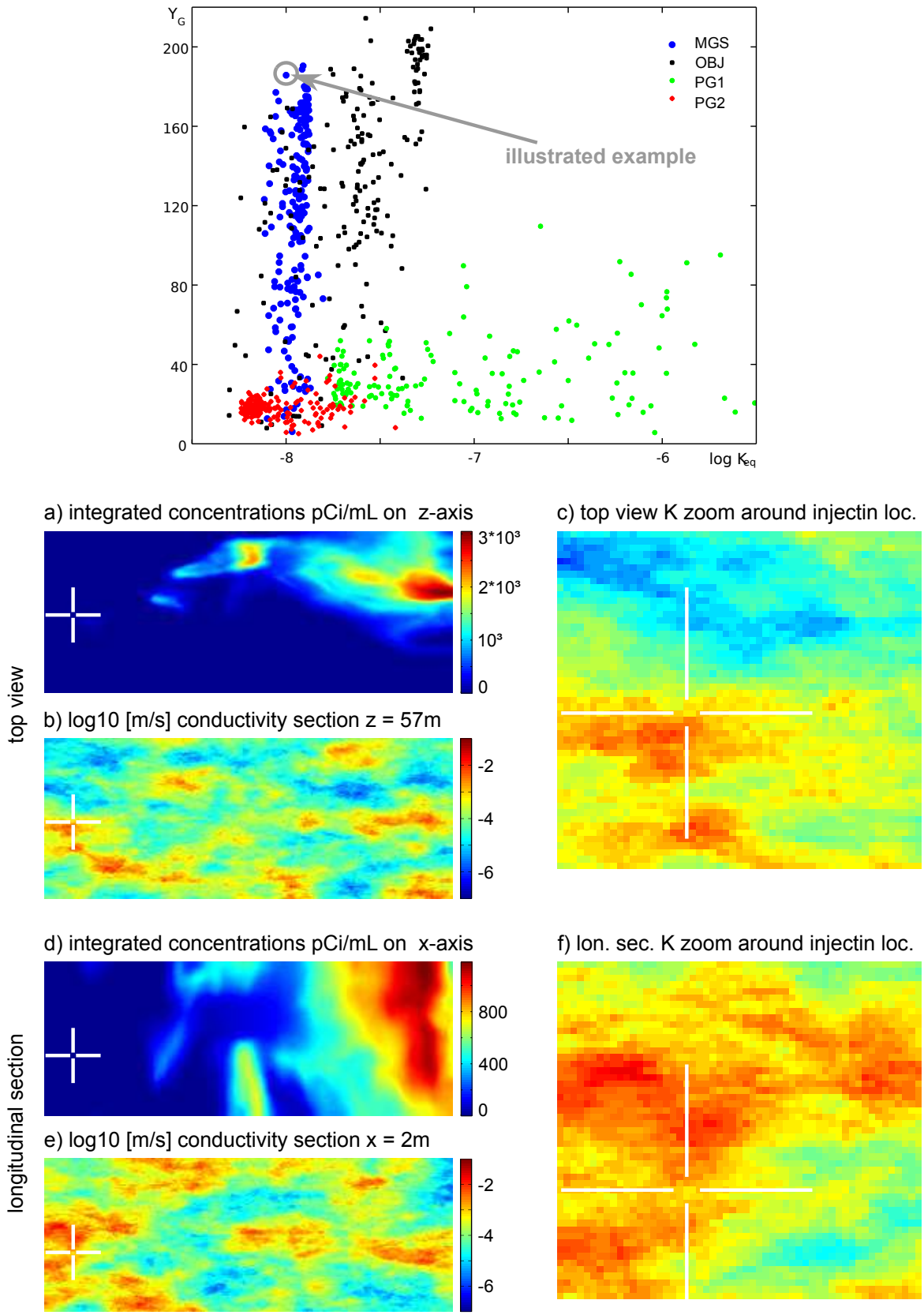

Figure D.1: Illustration of cumulative concentrations 328 days after injection and conductivity sections for a MGS K-field generating a fast plume; the white cross locates the highest concentration 27 days after injection.

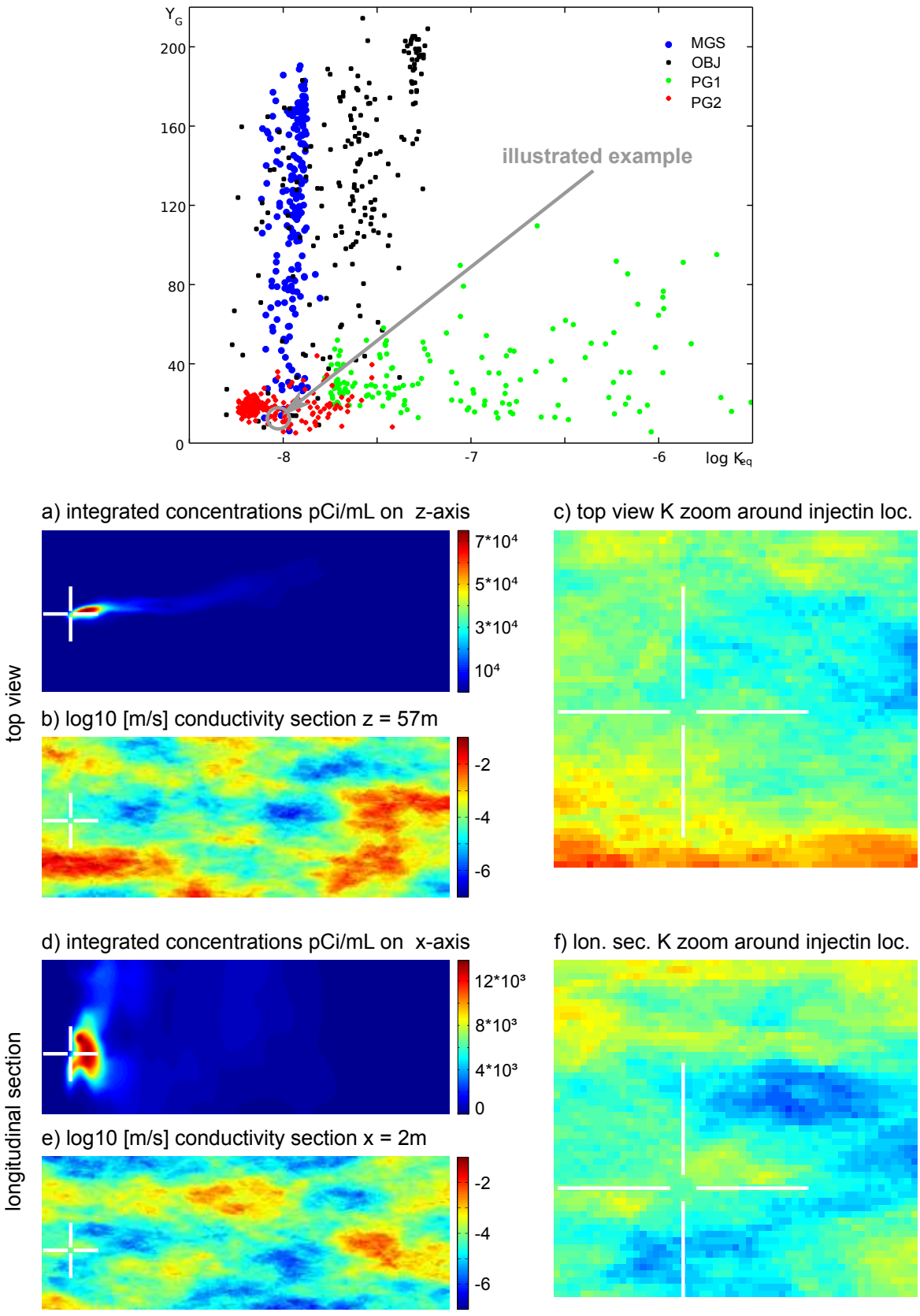

Figure D.2: Illustration of cumulative concentrations 328 days after injection and conductivity sections for a MGS K-field generating a slow plume; the white cross locates the highest concentration 27 days after injection.

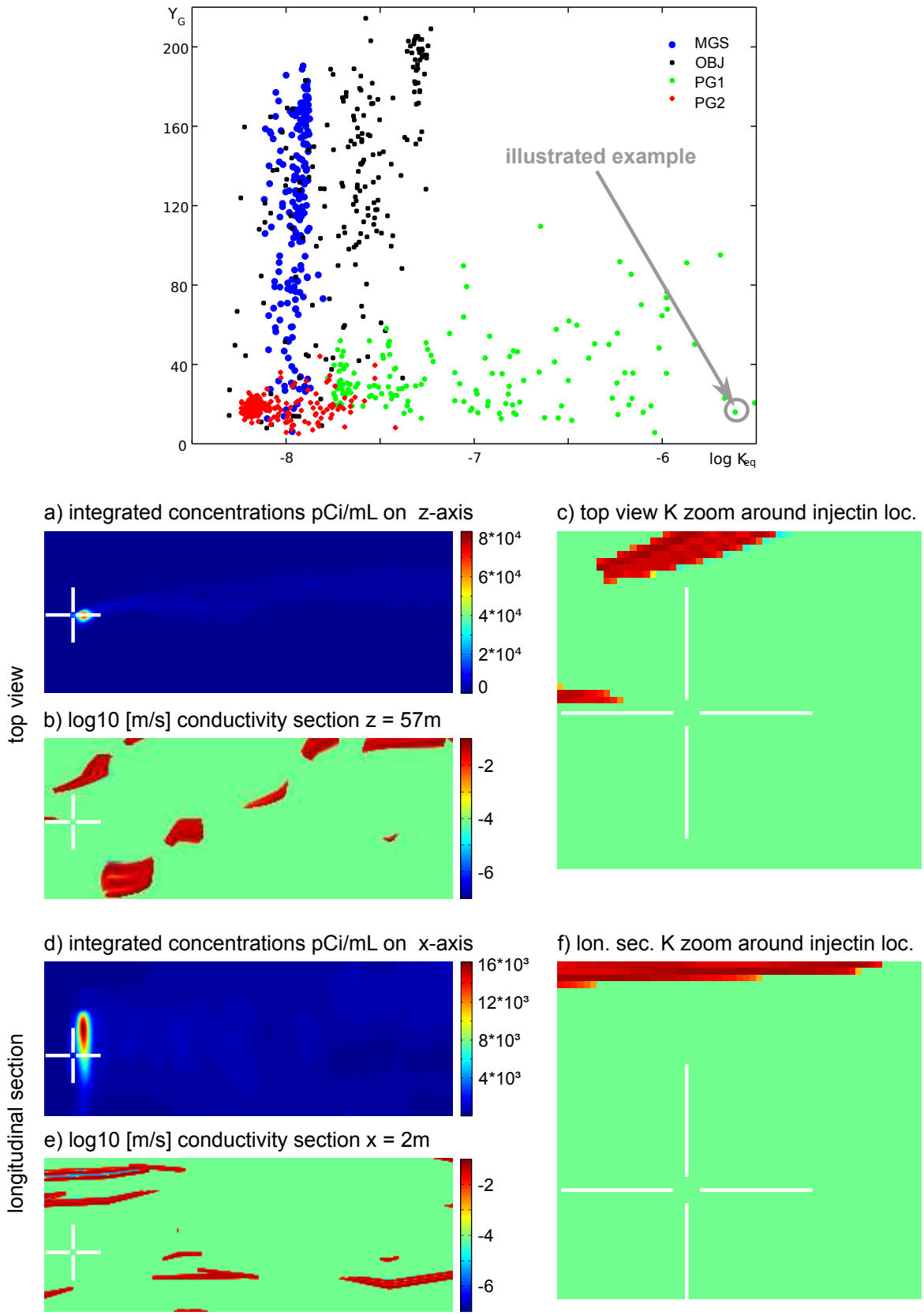

Figure D.3: Illustration of cumulative concentrations 328 days after injection and conductivity sections for a PG1 K-field – characterized by a low equivalent conductivity – generating a fast plume ; the white cross locates the highest concentration 27 days after injection.

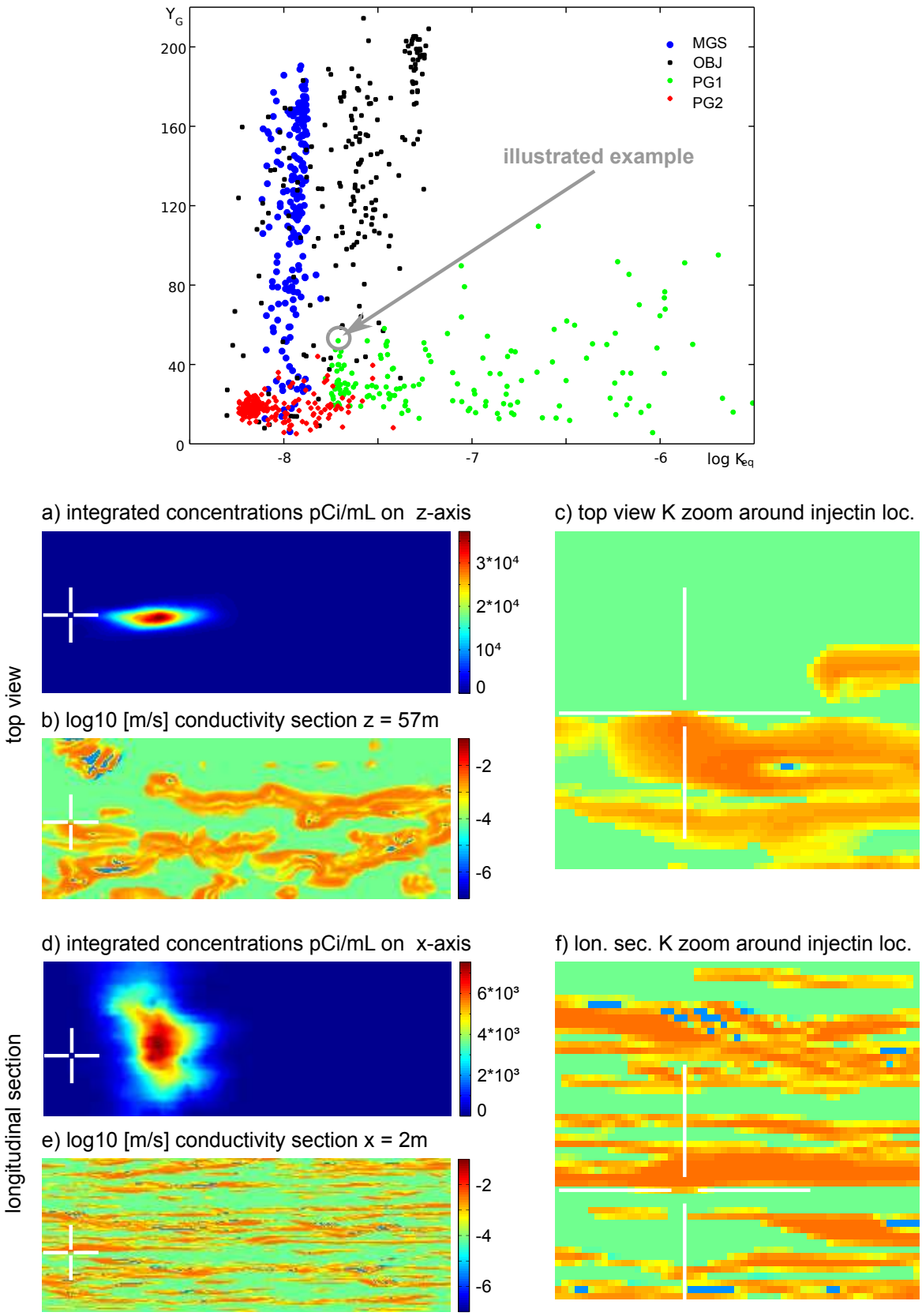

Figure D.4: Illustration of cumulative concentrations 328 days after injection and conductivity sections for a PG1 K-field – characterized by a high equivalent conductivity – generating a slow plume; the white cross locates the highest concentration 27 days after injection.