Journal of Hydrology 531 (2015) 124–141

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

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

Guillaume Pirot [a,*], Philippe Renard [a], Emanuel Huber [b], Julien Straubhaar [a], Peter Huggenberger [b]

[a] University of Neuchâtel, Centre for Hydrogeology and Geothermics, Rue Emile Argand 11, CH-2000 Neuchâtel, Switzerland
[b] University of Basel, Department of Environmental Sciences, Basel, Switzerland

ARTICLE INFO

Article history:
Available online 29 July 2015

Keywords:
Uncertainty
Contaminant transport
Modeling
Braided river
Aquifer
Groundwater

SUMMARY

Hydrogeologist are commonly confronted to field data scarcity. An interesting way to compensate this data paucity, is to use analog data. Then the questions of prediction accuracy and uncertainty assessment when using analog data shall be raised. These questions are investigated in the current paper in the case of contaminant transport forecasting in braided river aquifers. In using analog data from the literature, multiple unconditional geological realizations are produced following different geological conceptual models (Multi-Gaussian, Object-based, Pseudo-Genetic). These petrophysical realizations are tested in a contaminant transport problem based on the MADE-II tracer experiment dataset. The simulations show that reasonable contaminant transport predictions can be achieved using analog data. The initial concentration conditions and location regarding the conductivity heterogeneity field have a stronger influence on the plume behavior than the resulting equivalent permeability. The results also underline the necessity to include a wide variety of geological conceptual models and not to restrain parameter space exploration within each concept as long as no field data allows for conceptual model or parameter value falsification.

1. Introduction

The aims of this paper are to investigate if the use of analog data allows to make (1) reasonable contaminant transport predictions in braided river aquifers, (2) relevant uncertainty analysis, given different geological conceptual models and limited field data observations, and (3) to assess how the predictions are influenced by the different geological structures resulting from the different geological concepts.

Indeed braided river aquifers are frequent in alpine regions such as Switzerland and constitute an important part of the tapped water resources (FOEN, 2009). In an environment influenced by climate change (Middelkoop et al., 2001; Moeck, 2014) and under constraints related to human activity such as infrastructures or industries, preserving the water resources quality becomes a priority to pursue irrigation practice, drinking water supply and of course to protect the natural ecosystems relying on this resource. More generally, regardless of the sedimentary deposit nature, a good understanding of groundwater flow and transport in aquifers is necessary to manage the water resource (defining protection perimeters around pumping wells, planning restoration projects or planning remediation actions after a contamination).

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 aquifer models – not limited to the braided river type – have different characteristics and can be classified in process-imitating, structure-imitating, or descriptive methods (Koltermann and Gorelick, 1996).

Among the various types of aquifer models, some have been compared through numerical experiments. For example, Teles et al. (2004) compare the flow and transport properties in fluvio-glacial alluvium 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 of alluvial sediments are obtained when facies details are simulated within

* Corresponding author.
E-mail address: guillaume.pirot@unine.ch (G. Pirot).

http://dx.doi.org/10.1016/j.jhydrol.2015.07.036
0022-1694/

previously established structural elements. Many experiments have also been conducted for other types of environments (Zinn and Harvey, 2003; Falivene et al., 2006; Lee et al., 2007; Zhang et al., 2013; Siirila-Woodburn and Maxwell, 2015). 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 from these comparisons 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 model comparison. First of all, we assume that little data is available about a given site, as it is most often the case for hydrogeologists. Secondly, our interest is focused on braided river aquifers. To compensate the data paucity, we consider different geological conceptual models and we use prior information from analog sites elsewhere in the world to constrain the heterogeneity models. In a context where no conductivity measures are available, the issue of model conditioning or even inverse conditioning cannot be addressed. We then perform a set of concentration predictions and compare the results with a highly detailed reference data set. As stated at the beginning of the introduction, 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 question is whether one conceptual model is more robust in a sense that it is likely to provide more reliable uncertainty estimates than others. Here, three types of models, designed for braided river aquifers, are tested in this situation (a multi-gaussian model, an object based model, and a process-imitating model). 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. The concentration data measured at the MADE site are compared a posteriori with the concentrations simulated for the different models. These comparisons allow to compute various errors, such as error on the contaminant mass, error on the concentrations, error on the plume center of mass, error on the plume dispersion for instance. In addition, to assess the quality of the predictions uncertainty in a context of information paucity, the cumulative mass of the contaminant is predicted in different zones.

The paper is structured as follows. First, Section 2 presents what are the information used from the MADE site data-set to setup the problem and how analog sites described in the literature are compiled to complete these information. Then, the different geological conceptual models are described in Section 3. The transport model and its specific parameters and boundary conditions are presented in Section 4. Finally, the results of the plume simulations are presented and analyzed in Section 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 limited 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.

2.1. The MADE site

The MADE site is located on the Air Force Base of Colombus, Mississippi (see Fig. 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 4–5‰. The alluvial aquifer is shallow and its mean thickness is about 11 m. 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–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 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. 1 for an illustration of the piezometers and Fig. 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 20–30 sampling points space 0.38 m 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.

2.2. Analog site data

Facies conductivity mean and variance and porosity values as well as characteristic length scales can be retrieved or computed

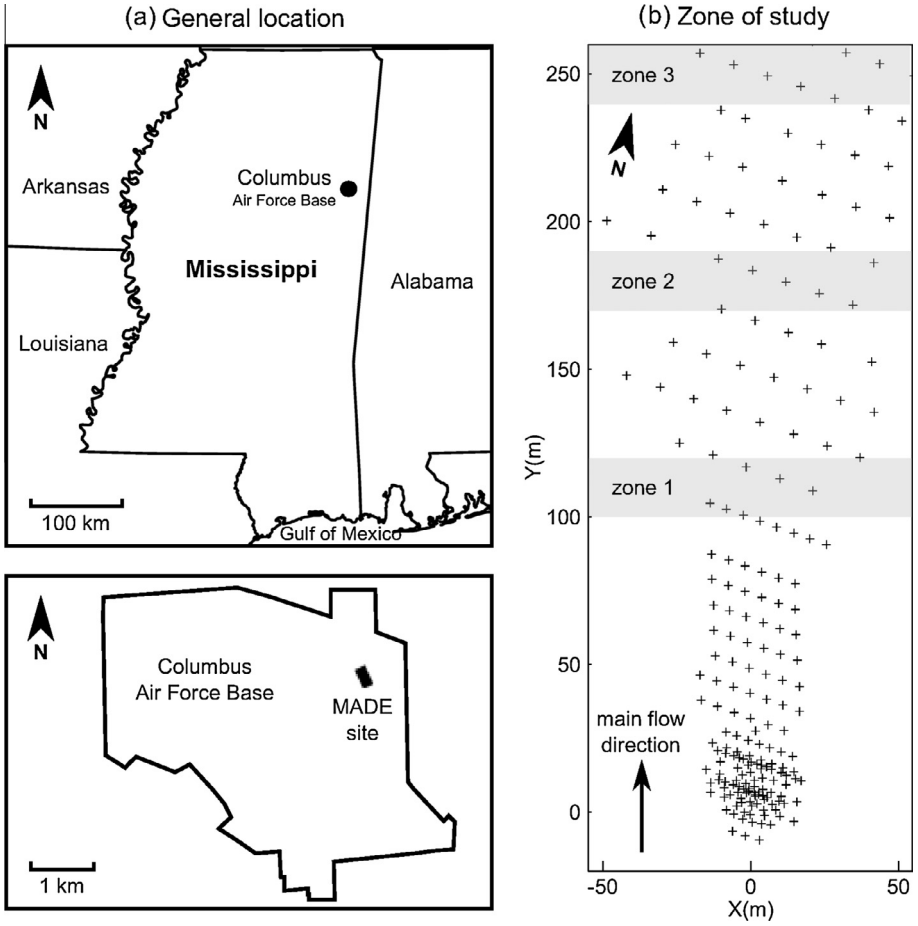

Fig. 1. The MADE site: location (a) and piezometer network (b) – injection location corresponds to coordinates (0, 0).

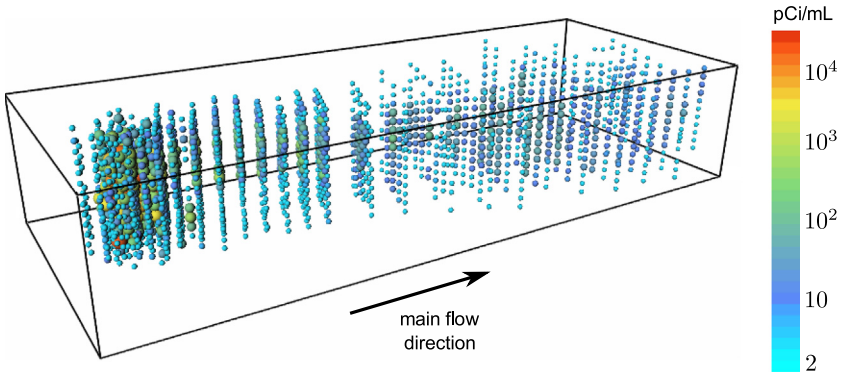

Fig. 2. Reference concentrations measured 328 days after injection.

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 Table 1.

One might note that the number of structural elements to consider might differ between two braided river aquifers. Indeed for numerical models, the number of structural elements relies on the retained geological concept, as illustrated in Section 3. This is why the main structural and depositional elements proposed by Jussel et al. (1994) are not described in details here.

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_{lnK}^2 of Y as:

$$\mu_{lnK} = \sum_i p_i \mu_i, \tag{1}$$

$$\sigma_{lnK}^2 = \sum_i p_i \mu_i^2 - \mu^2 + \sum_i p_i \sigma_i^2. \tag{2}$$

Table 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

Using the values of Jussel et al. (1994) summarized in Table 1 leads to a mean of the *log* conductivity and a standard deviation of:

$$\mu_{lnK} \simeq -9.14 \qquad \sigma^2_{lnK} \simeq 3.05 \tag{3}$$

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 $\mu_n = 0.20$. These values are of course different from a mean $ln(K)$ of -11.6 and variance of 6.6 obtained by direct-push injection logger measures (Bohling et al., 2012) and from a mean porosity of 0.31 as reported by Boggs and Adams (1992). These differences are the consequence of the data paucity assumption and they do not prevent pursuing the aims of the papers.

Jussel et al. (1994) also provide characteristic length scales per facies, as presented in Table 1. There are no dimension characteristics defined for the predominant facies GG/BG, as it is considered as the background 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 Table 2).

To sum-up, only the BG structural element is used to define characteristic length scale scenarios. All structural elements proportions, conductivity mean and variance are used to compute a common conductivity mean and variance.

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): $110 \times 280 \times 10.5$ m with a $1 \times 1 \times 0.1$ 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. 1). Each conceptual geological model is decomposed in 5 scenarios to cover the uncertainty on characteristic dimensions (Table 2). For each scenario, 40 geological realizations are generated, which builds an ensemble of 200 geological realizations per conceptual model.

Table 2
Dimension characteristics per scenario.

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

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

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) developed by Huber et al. (submitted for publication) (As the paper is not yet published, the reader might refer to Huber (2015)). The third model uses a pseudo genetic algorithm developed by Pirot et al. (submitted for publication) (As the paper is not yet published, the reader might refer to Pirot (2015)) for two different parameter sets (PG1 and PG2). In the present paper, we are interested in contaminant transport predictions in braided-river aquifers. This is why we limited the modeling methods to the one based on geological description or genesis of braided-river aquifers plus a standard multi-gaussian model as comparison reference. Conductivity field realizations for the different conceptual models are illustrated in Figs. 3 and 4.

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 Section 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.

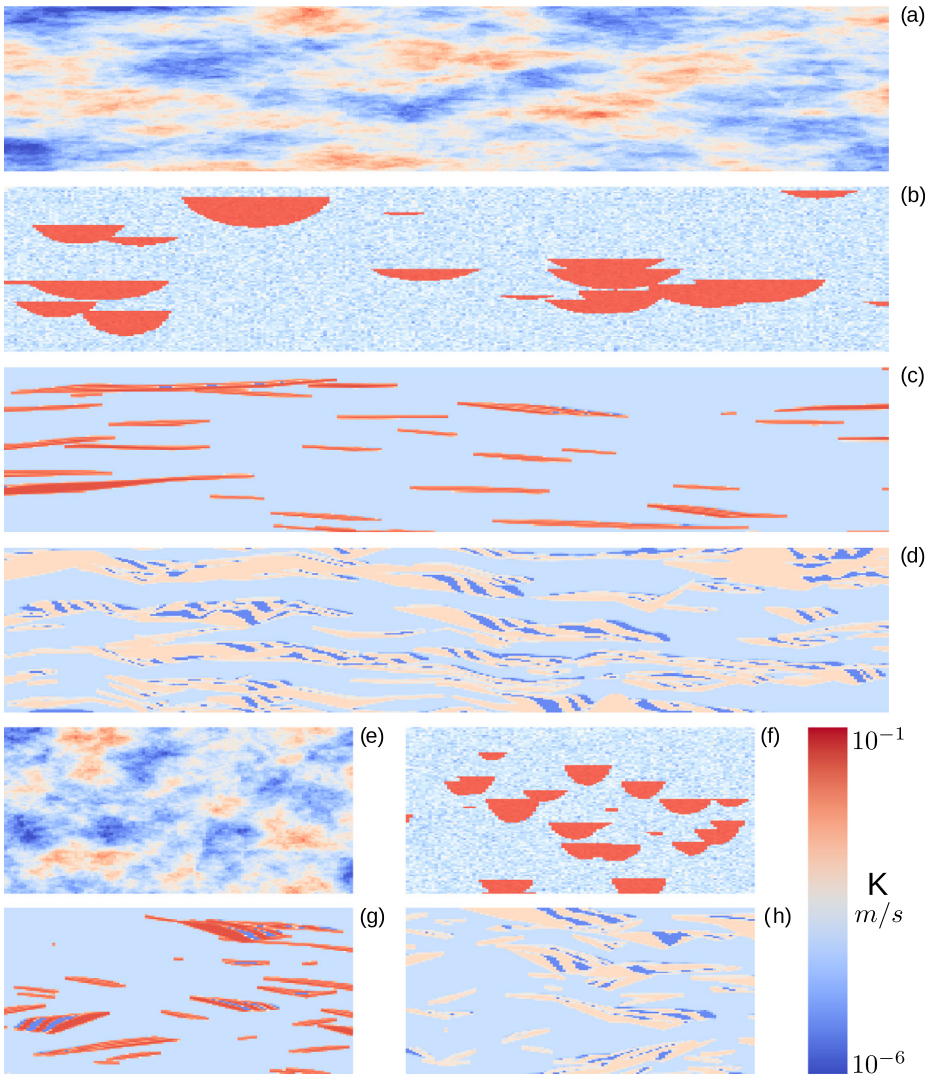

Fig. 3. Two dimensional longitudinal (a–d) and lateral (e–h) sections of scenario 3 conductivity fields for the different conceptual geological models (a) and (e): MGS, (b) and (f): OBJ, (c) and (g): PG1, (d) and (h): PG2.

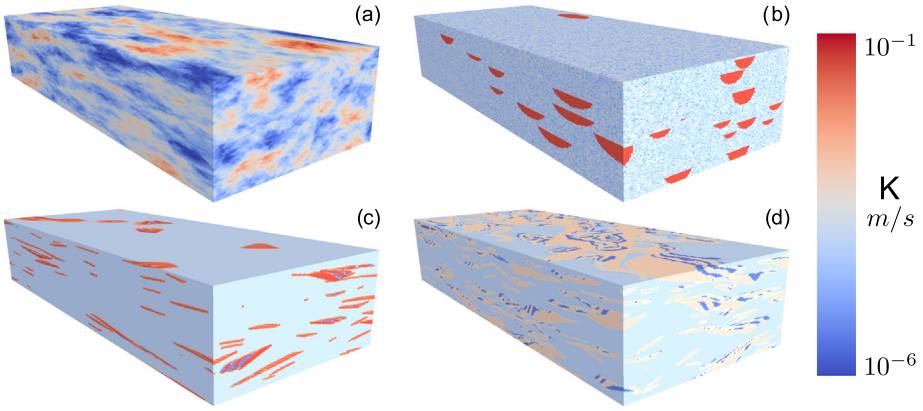

Fig. 4. Three dimensional representations of scenario 3 conductivity fields for the different conceptual geological models (a): MGS, (b): OBJ, (c): PG1, (d): PG2.

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 Table 3.

To allow a systematic model comparison later on, the mean and variance of the *log* conductivity fields are normalized (Eq. (4)) so that the resulting mean and variance are those given by Eq. (3):

Table 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 (m)	21.0	35.5	50.0	64.5	79.0
Lateral range (m)	8.4	14.2	20.0	25.8	31.6
Vertical range (m)	0.84	1.42	2.00	2.58	3.16
Kriging mean	$\mu_{lnK} = -9.14$				
Kriging variance	$\sigma^2_{lnK} = 3.05$				

$$lnK_{normalized} = (lnK_{sim} - \mu_{lnK_{sim}}) \times \frac{\sigma_{lnK}}{\sigma_{lnK_{sim}}} + \mu_{lnK}, \quad (4)$$

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 Figs. 3(a) and (e), 4(a) and 5(a).

3.2. Object and process based models

Field observations of gravel pit exposures of Pleistocene coarse deposits in Switzerland showed that the main types of depositional elements are horizontal layers and "cross-bedded sets [...] with trough-shaped, erosional concave upward lower bounding surfaces" (Huggenberger and Regli, 2006). The horizontal layers correspond to poorly sorted gravel (the grey gravel (GG) in Table 1). The erosional bounding surfaces of the trough-fills are generally spoon-shaped and oriented in the main flow direction (e.g. Siegenthaler and Huggenberger, 1993). Therefore, they were identified as being formed by scour-pool fills (e.g. Siegenthaler and Huggenberger, 1993; Jussel et al., 1994; Klingbeil et al., 1999; Heinz et al., 2003; Huggenberger and Regli, 2006; Bayer et al., 2011). The trough-fill mostly consists of open framework – bimodal gravel couplet cross-beds (OW/BM in Table 1, see also Siegenthaler and Huggenberger, 1993; Huggenberger and Regli, 2006). Due to the strong contrast in hydraulic conductivity between GG and OW/BM, the three dimensional spatial distribution of OW/BM in the subsurface is likely to act as a fast pathway for the subsurface flow (Huber and Huggenberger, 2015). The GG can be considered as background. A brief description of the algorithm is proposed in the next paragraph. All the details can be found in Huber et al. (submitted for publication).

The algorithm mimics the deposition and erosion of sediments on the braidplain for successive large flood events. The trough-fills are modeled with truncated ellipsoids aligned with the horizontal plane. At each event a horizontal layer of sediment is deposited on the previous braidplain elevation and truncated ellipsoids are randomly distributed on the horizontal layer. The top elevation of the truncated ellipsoids coincides with the top elevation of the deposited layer. The vertical distribution of the layer is simulated by a homogeneous Poisson process of intensity $\lambda = \frac{Z_{max} - Z_{min}}{ag}$, where ag (m/layer) is the expected aggradation rate and $Z_{max} - Z_{min}$ (m) is the desired aquifer thickness. The width, length and depth of the truncated ellipsoids as well as their orientation are supposed uniformly distributed. The center of the truncated ellipsoids is distributed according to a Strauss process defined by the parameters β, γ and r. 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 within each facies (Jussel et al., 1994) without accounting for spatial correlation. The cells associated to the matrix and the scours respectively are assigned conductivity statistical properties of facies GG and OW/BM respectively (see Table 1).

The Object model parameters are given in Table 4. The semi-axes a, b and c as well as the vertical truncation ratio rH were defined such that the objects dimensions do not exceed the characteristic dimensions defined for each scenario in Table 2. According to

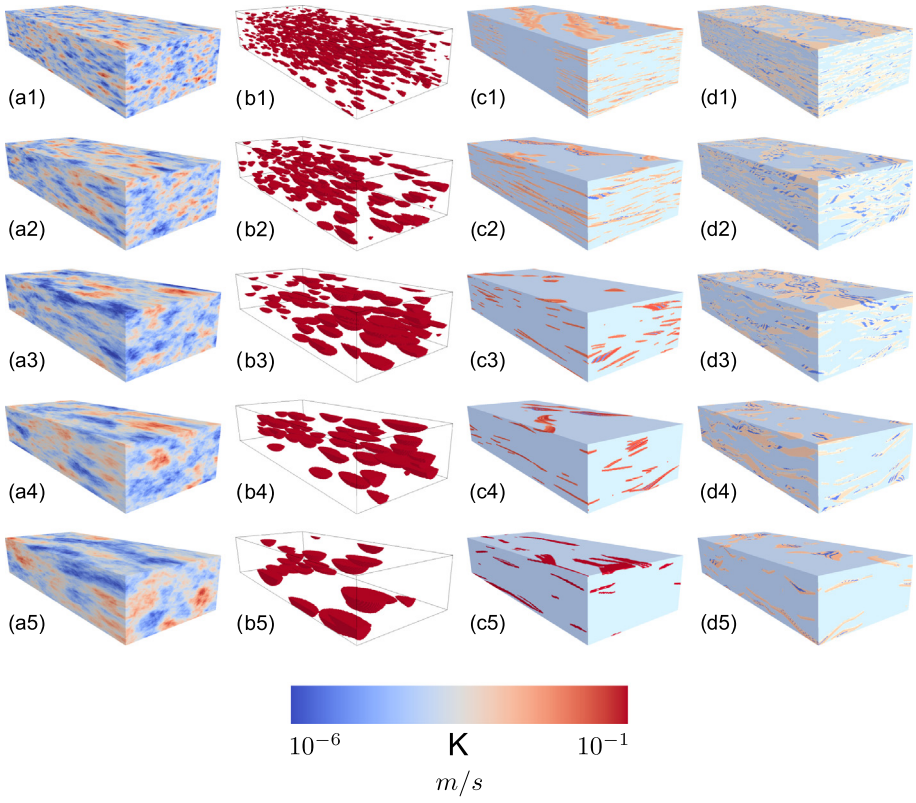

Fig. 5. Three dimensional representations of conductivity fields for the five scenarios (1–5) and the different conceptual geological models (a): MGS, (b): OBJ, (c): PG1, (d): PG2.

Table 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	$\left[\frac{0.2}{0.3}, \frac{0.84}{0.8}\right]$	$\left[\frac{0.4}{0.3}, \frac{1.42}{0.8}\right]$	$\left[\frac{0.4}{0.3}, \frac{2}{0.8}\right]$	$\left[\frac{0.55}{0.3}, \frac{2.58}{0.8}\right]$	$\left[\frac{1}{0.3}, \frac{3.16}{0.8}\right]$
Ellipsoid orientation α	$\left[\frac{-40\pi}{180}, \frac{40\pi}{180}\right]$	$\left[\frac{-40\pi}{180}, \frac{40\pi}{180}\right]$	$\left[\frac{-40\pi}{180}, \frac{40\pi}{180}\right]$	$\left[\frac{-40\pi}{180}, \frac{40\pi}{180}\right]$	$\left[\frac{-40\pi}{180}, \frac{40\pi}{180}\right]$
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	20	45	60	75	90

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 were adjusted by trial and error to keep approximatively the same facies proportions when the size of the objects changes. It allows to keep focused on the impact of geological structures, as stated in the objectives.

Then, to ensure the same *log* conductivity mean and variance for all realizations, the *log* conductivity fields simulated are normalized according to Eq. (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 Figs. 3(b) and (f), 4(b) and 5(b).

3.3. Pseudo genetic aglorithm

The algorithm produces facies models. The pseudo-genetic method (Pirot et al., submitted for publication) 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 produces 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. (submitted for publication).

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 presented in Pirot et al. (2014) are used here. 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 contributes (1) to erode partly the underlying geological layers and (2) to deposit sediments, forming so a new geological layer. A progressive and iterative deformation scheme of the geological layers base, taking into account approximate local flow and local topography, allows to generate the cross stratified deposits by repeating the sequence of successive facies 1 and 3, as illustrated in Fig. 6.

In the facies sequence (Tables 6 and 7), the matrix is defined as facies 2. It corresponds to the GG/BG horizontal structural element

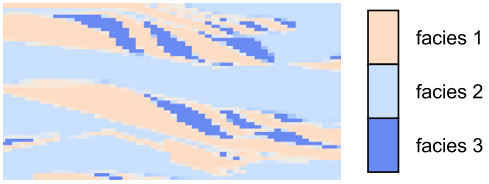

Fig. 6. Repetition of the facies sequence forming cross-stratified deposits for the pseudo genetic models – extracted from a vertical section orthogonal to the flow direction.

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, 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),

$$\mu_{lnK} = p_1 \times ln(K_1) + p_2 \times ln(K_2) + p_3 \times ln(K_3) \tag{5}$$

where $\{p_1, p_2, p_3\}$ and $\{K_1, K_2, K_3\}$ are respectively the proportions and conductivity values of facies $\{1, 2, 3\}$. Note that the proportions are not input parameters in the algorithm, but are resulting parameters. The variance is known and given by Eq. (6),

$$\sigma^2_{lnK} = E(ln(K_{sim})^2) - E(ln(K_{sim}))^2 \tag{6}$$

where K_{sim} is the conductivity field for the simulation. Developing it with facies proportions and conductivities, it becomes Eq. (7):

$$\sigma^2_{lnK} = p_1 \times ln(K_1)^2 + p_2 \times ln(K_2)^2 + p_3 \times ln(K_3)^2 - (p_1 \times ln(K_1) + p_2 \times ln(K_2) + p_3 \times ln(K_3))^2. \tag{7}$$

The values of μ_{lnK} and σ^2_{lnK} are set to -9.14 and 3.05 according to Eq. (3) and the matrix (facies 2) conductivity value is fixed to 8×10^{-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: $K_1 > K_2 > K_3$.

Scaling parameters give a freedom degree to reproduce 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 5–7. 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. Let us denote d_l, d_w, d_d the length, width and depth dimensions of the resulting objects. PG1 version produced objects of dimensions $d_l \times d_w \times d_d = 37 \times 22 \times 1.5$ m and of lower density than PG2 version that produced

Table 5
Dimension, resolution and general parameters.

	Parameter	Value
Zone of interest aquifer model parameters	Length (flow direction), m	280
	Width (orthogonal to the flow direction), m	110
	Minimum thickness – depth, m	10.5
	Cell length, m	1
	Cell width, m	1
	Cell height, m	0.1
Braided river topography dimensions	Length, m	11,600
	Width, m	1200
	Cell length, m	20
	Cell width, m	20
Interpolation parameters	Margin length, m	5
	Margin width, m	5
	Margin depth, m	0
facies parameters	Coarse grain size sediment facies value	1
	Unsorted mixed grain size sediment facies value	2
	Fine grain size sediment facies value	3

objects of dimensions $d_l \times d_w \times d_d = 66 \times 25 \times 1$ m. For each scenario j, the scaling factors $s_l^j,\ s_w^j,\ s_d^j$ along the length, width and depth axis are computed as $s_i^j = I_i^j \div s_i^j,\ i \in \{l, w, d\}$ where $I_l^j,\ I_w^j,\ I_d^j$ are the dimension characteristics for scenario j, as defined in Table 2.

As the resulting model is defined on a pillar grid (Pirot et al., submitted for publication) – 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. 7).

To limit the loss of information such as sharp conductivity contrasts or the presence of thin facies layers during this vertical upscaling 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 $K_{zz} = \frac{1}{\sum_i \frac{p_i}{K_i}}$.

3.4. Integral scale cross comparison by scenario and geological concept

One could wonder if the conductivity models generated by different algorithm present comparable length scale characteristics within each scenario, as the different model parameters were design to facilitate it. In order to provide a quantitative comparison of the resulting length scale characteristics, the integral scales of the $ln\ K$ realizations were computed along the three directions of the orthonormal basis. The lateral direction represents the horizontal direction orthogonal to the main flow direction. The longitudinal direction corresponds to the main flow direction. Given a specific direction, the integral scale I_s is computed as $I_s = \int_0^\infty \rho(h)\ dh$ where $\rho(h) = \frac{\sigma^2 - \gamma(h)}{\sigma^2}$ is the two-point auto-correlation as a function of the distance h along the retained direction, $\gamma(h)$ being the semi-variogram. The box-plots of the integral scales per scenario and per concept are presented in Fig. 8.

It shows that scenario characteristic length scales can be retrieved for the MGS and OBJ models and in a less pronounced way for the PG1 models, as the integral scale is an increasing function of the scenario length scale. However for the PG2 models, it is not possible to differentiate the scenarios by the integral scale, as its mean looks almost constant, only the variance seems to increase. A most probable reason is the presence of the sharp contrasts due to the cross stratification deposits which shortens the characteristic length scale of the conductivity. In spite of our efforts to customize model parameters in order to have comparable characteristic length scales per scenario, the integral scales are different from a concept to another. The integral scales of the OBJ models seems limited to the integral scales of scenarios 1–4 for

Table 6
Scenario dependent structural parameters and 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 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*				

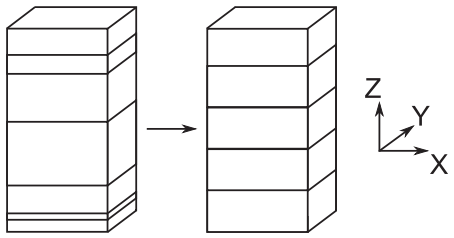

Fig. 7. Vertical grid regularization.

MGS models while those of the PG1 models reflect those of scenarios 3 and 4 for the MGS models, and those of the PG2 models are comparable to those of scenario 1 for MGS and OBJ models.

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 $1 \times 1 \times 0.1$ m). Iso-surfaces of the transport simulations are illustrated for the different conceptual models and their variant at various concentrations threshold in Section 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. 9). 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. 9).

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 Section 2.2. Though it is superfluous due to the heterogeneity models, the longitudinal dispersivity is a parameter required by *Groundwater*, the flow and transport simulator. 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}.

4.2. Interpolations of the concentration data

The concentrations sampled on the field at day 27 (Fig. 10(a)) are interpolated on the whole domain of the model to obtain the initial state for the transport simulations (Fig. 10(b) and (c)).

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 Table 8.

Note that the kriging means are set to the low boundary values of the NST concentration values, to ensure that the interpolated

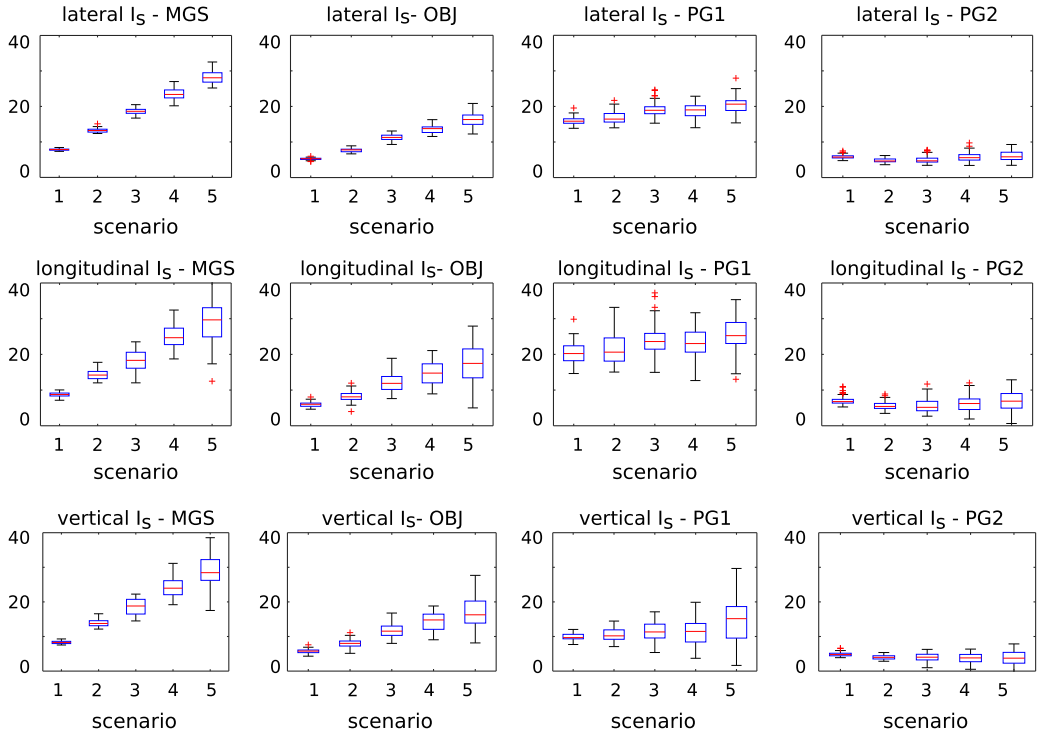

Fig. 8. Integral scales of the $\ln K$ realizations along the orthonormal basis directions presented for all scenarios and grouped by conceptual model.

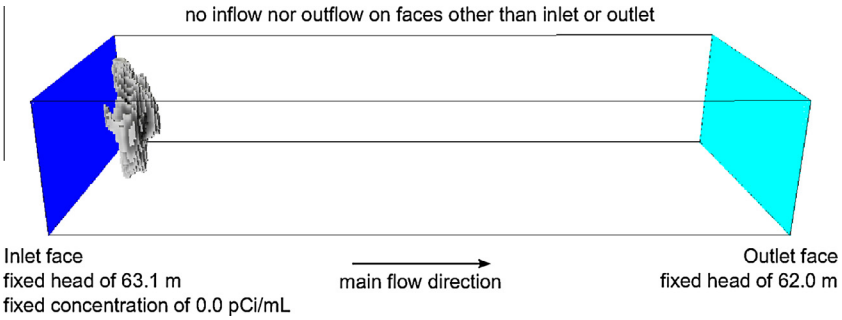

Fig. 9. Initial and boundary conditions.

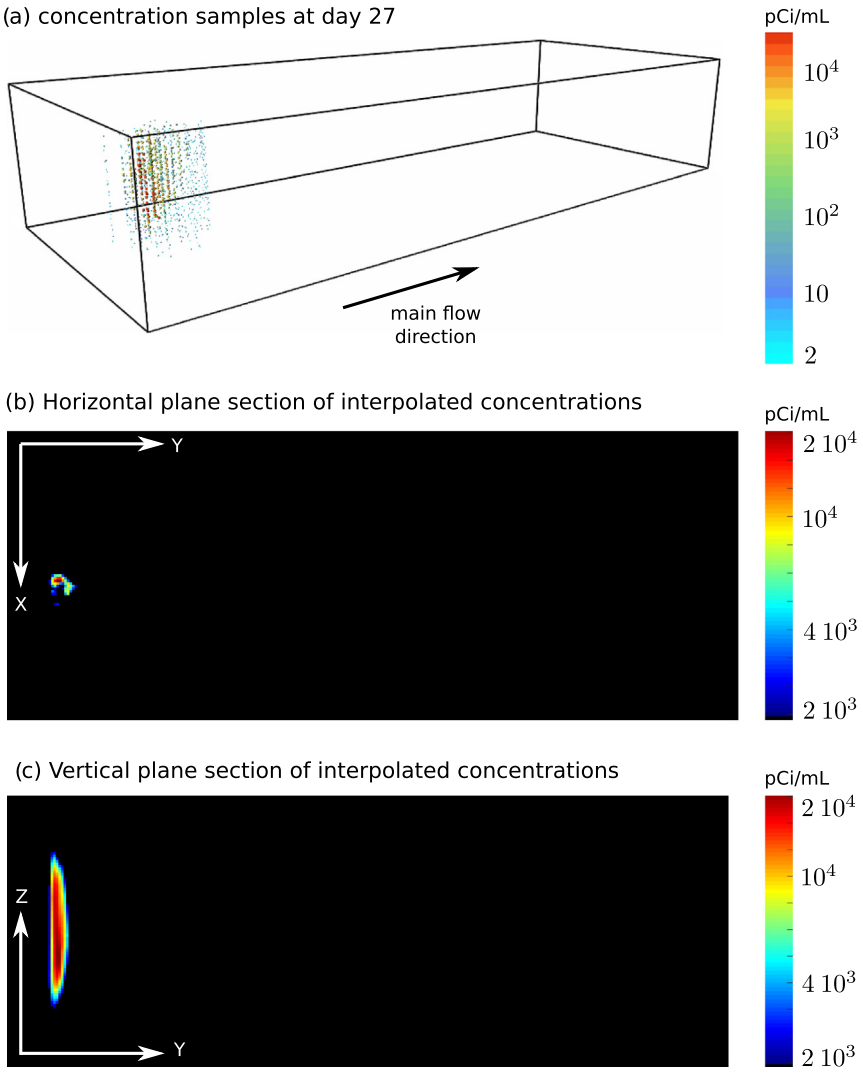

Fig. 10. Concentration samples (a) and thresholded interpolation of the highest concentrations representing 80% of the mass contaminant (b) and (c) at day 27.

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. Comparison of simulated versus measured concentrations

In this section, the simulated plumes are analyzed qualitatively and quantitatively for each conceptual model. The analysis compares the different conceptual models but not the different scenarios as these are used to cover the aquifer characteristic dimensions uncertainty. Most of the quantitative errors are computed directly from the sample data without intermediate interpolation, to avoid errors due to (1) non-linear NST of the reference data at day 328

Table 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 + 7.2 m	16 m
Simple kriging mean	−3.2	−3.4
Moving neighborhood ellipsoid	21 × 21 × 3 m	21 × 21 × 3 m

before interpolation and (2) back-transformation of the interpolated data. However, it is reasonable to use such smoothing interpolation method when computing a smooth representation of the reference plume such as a cumulative mass. The section ends with a prediction uncertainty analysis of the cumulative mass in various target zones.

5.1. Qualitative comparison of the plumes

Qualitatively, from the examples displayed in Fig. 11, one can observe that 328 days after injection, the plume is getting out of the domain for MGS and OBJ simulations; even high level of concentration are observed close to the exit of the domain.

Regarding PG1 simulations, the plume has spread with low concentrations toward the exit of the model, high concentrations staying closer to the injection zone. For the PG2 simulation, the plume is really slower and its furthest expansion is limited within the first half of the domain, whatever the level of concentration observed. Though these comments are valid and limited to the observation of a single realization per conceptual model, they are quite representative of the situation as confirmed by the indicators that follow.

5.2. Quantitative errors

Fig. 12 presents the boxplots of errors between the simulation and the reference. These quantitative indicators are grouped by type of conceptual model. The first boxplot represents the error on the total contaminant mass conservation ε_T (Eq. (8)) in the model throughout the transient transport simulation duration. As explained in Section 4.2, the contaminant total mass should be preserved between day 27 and day 328.

$$\varepsilon_T = 1 - \frac{\int_V C^{sim}(x,y,z,t=328)\ \mu_n\ dv}{\int_V C^{ref}(x,y,z,t=27)\ \mu_n\ dv} \tag{8}$$

Fig. 12(a) shows that for MGS and OBJ realizations, the contaminant leave the model for most of the realizations, while for PG1 and PG2 realizations, the plume stays within the model boundaries for more than 50% of all realizations.

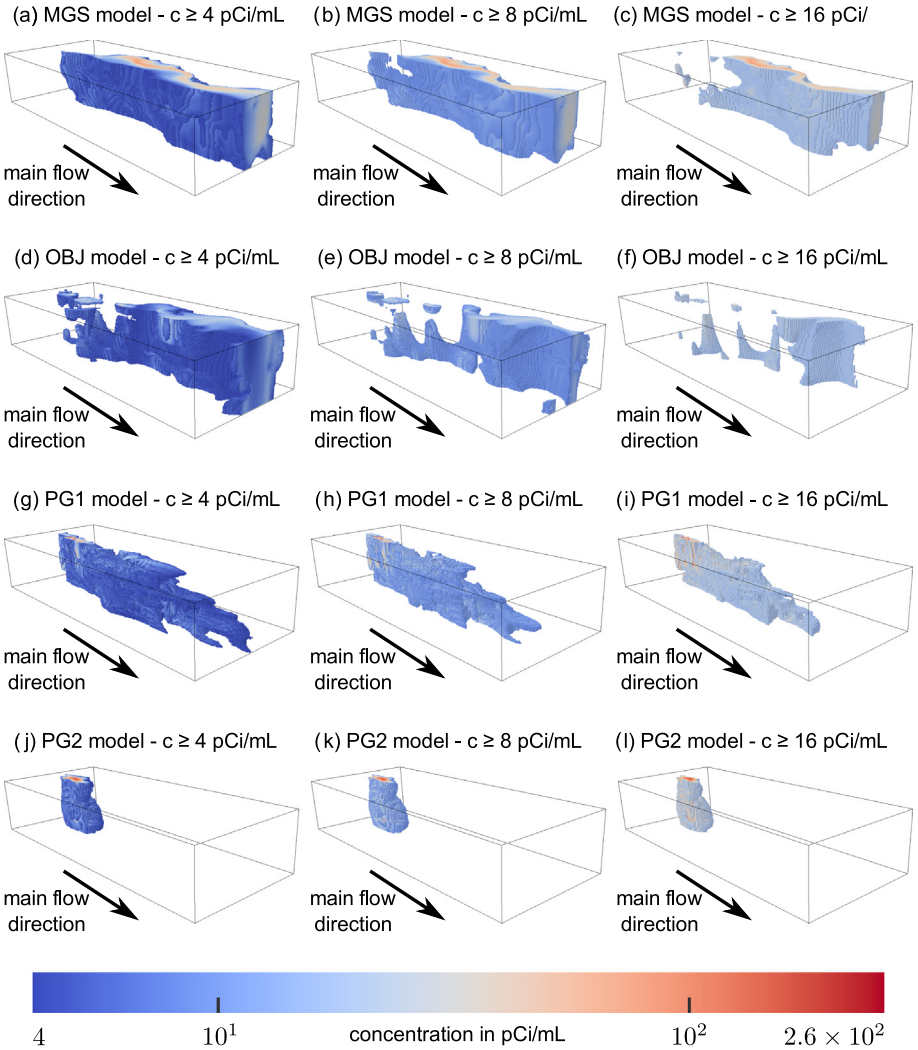

Fig. 11. Plume examples at day 328 for different conceptual models and for scenario 3 at different concentration thresholds – vertical scale exaggerated by 5.

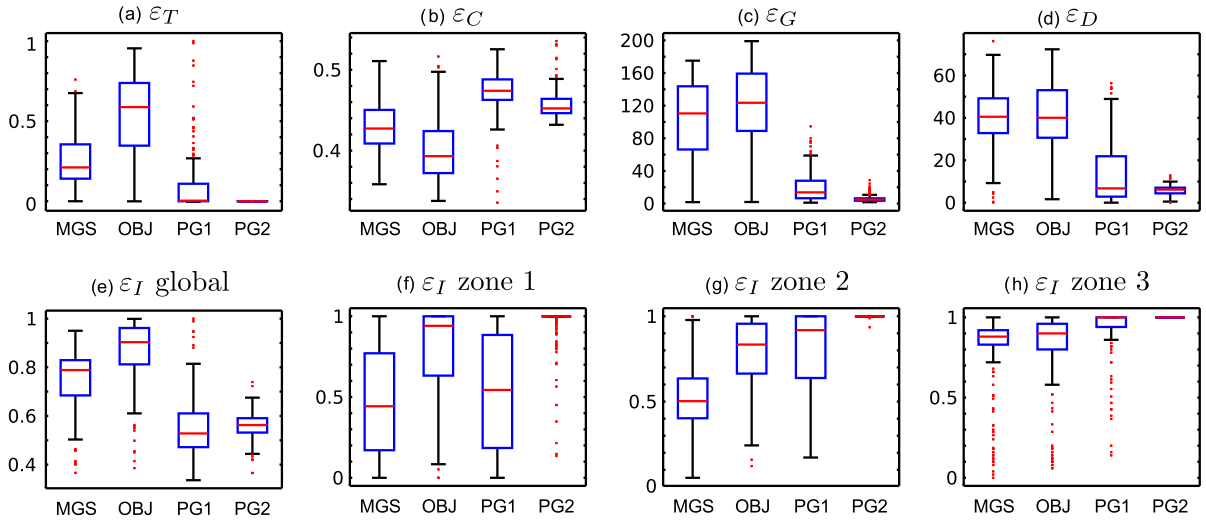

Fig. 12. Boxplots of errors at day 328; (a) error on the total mass; (b) L_1 error on concentrations at sample location; (c) error on the center of mass longitudinal position; (d) error on the center of mass dispersion; and (e)–(h) errors on iso-concentration classification by zone.

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 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. (9)).

$$\varepsilon_C = \sum_{s=1}^{N_S} \frac{|C_s^{sim} - C_s^{ref}|}{C_s^{ref}} \tag{9}$$

Fig. 12(b) shows that globally, the simulated concentrations are closer to the sampled 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 center of mass and the plume dispersion

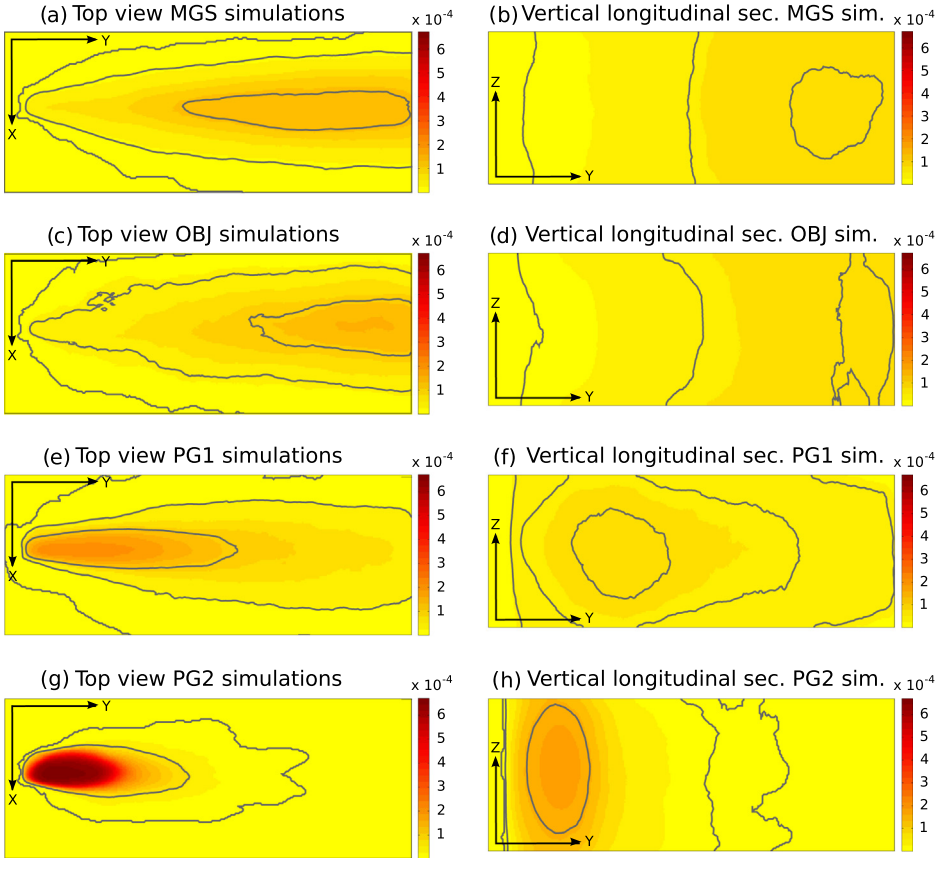

Fig. 13. Renormalized probability densities per model for $C \geqslant 4$ pCi/mL at day 328.

around its center of mass. An error ε_G on the center of mass location is defined as the euclidean distance between the center of mass $\vec{g}^{sim}$ of the simulated plume and the center of mass $\vec{g}^{ref}$ of the reference plume (Eq. (10)).

$$\varepsilon_G = \left\|\vec{g}^{sim} - \vec{g}^{ref}\right\| \text{ where } \vec{g}^{\cdots} = \frac{\sum_{s=1}^{N_S} C_s^{\cdots} \vec{x}_s}{\sum_{s=1}^{N_S} C_s^{\cdots}} \tag{10}$$

Fig. 12(c) shows that the error on the center of mass longitudinal position is small for PG1 and PG2 realizations while it is much 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 to define an error on the plume dispersion ε_D as the difference between the simulated plume dispersion and the reference plume dispersion (Eq. (11)).

$$\varepsilon_D = \left| \frac{\sum_{s=1}^{N_S} C_s^{sim} \|\vec{x}_s - \vec{g}^{sim}\|}{\sum_{s=1}^{N_S} C_s^{sim}} - \frac{\sum_{s=1}^{N_S} C_s^{ref} \|\vec{x}_s - \vec{g}^{ref}\|}{\sum_{s=1}^{N_S} C_s^{ref}} \right| \tag{11}$$

Fig. 12(d) shows that the plume dispersion is better reproduced for PG1 and PG2 models than for MGS or OBJ models.

5.3. Accuracy of downstream forecast

Often, in practical applications, the objective 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 to build an error ε_I based on the miss-classification of the simulated sample versus the reference sample classification (Eq. (12)).

$$\varepsilon_I = \frac{1}{N_S \times N_K} \sum_{s=1}^{N_S} \sum_{k=1}^{N_K} \mathbb{1}_{C_s^{sim} \geqslant C_k^{th}} \times \mathbb{1}_{C_s^{ref} \geqslant C_k^{th}} \tag{12}$$

This error is computed for the whole domain (ε_I), and for three delimited zones: zone 1, zone 2 and 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. 1. Fig. 12(e)–(h) shows that MGS simulations give better iso-concentration predictions in zones 1–3 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. These probabilities have been computed as $P_{C_{th}}(x_i, y_i, z_i) = \frac{1}{N_R} \sum_{j=1}^{N_R} \mathbb{1}_{C_i^j \geqslant C_{th}}$ for two concentration thresholds $C_{th} \in \{4, 16\}$ and C_i^j is the simulated concentration for realization j in cell i. These probabilities are illustrated as probability density maps after vertical and lateral integration (Eqs. (13) and (14)) in Figs. 13 and 14.

$$P_{C_{th}}^{Z}(x,y) \propto \frac{\sum_k P_{C_{th}}(x,y,z_k)}{\sum_i P_{C_{th}}(x_i,y_i,z_i)}, \quad k \text{ layer index on the vertical } z \text{ axis} \tag{13}$$

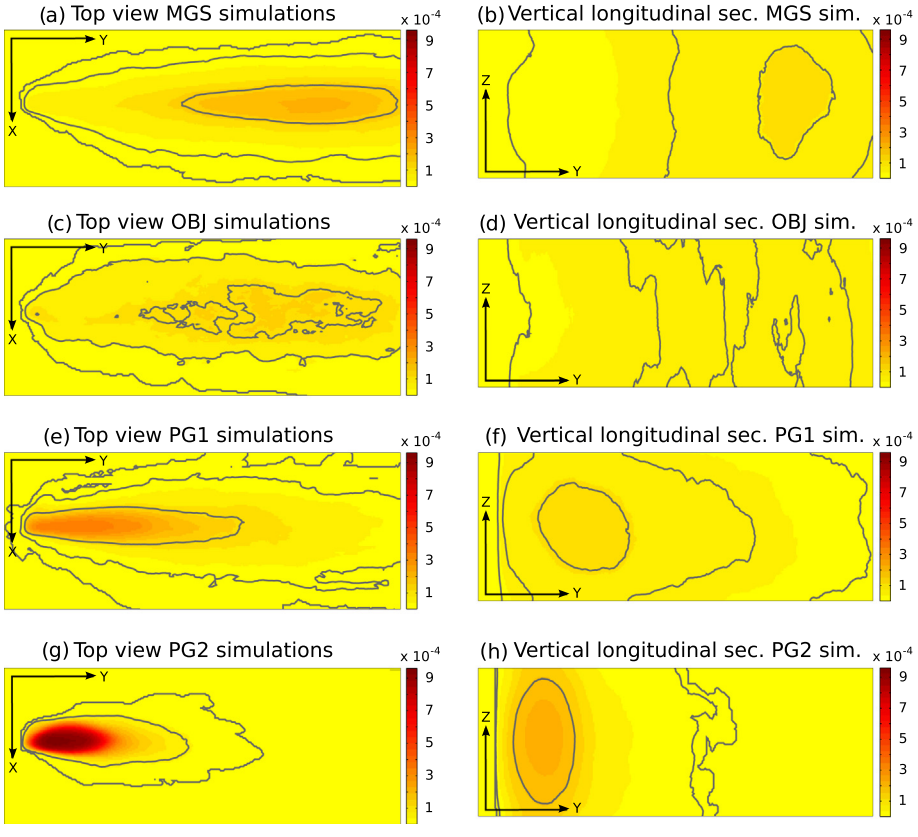

Fig. 14. Probability densities per model for $C \geqslant 16$ pCi/mL at day 328.

$$P^{X}_{C_{th}}(y,z) \propto \frac{\sum_l P_{C_{th}}(x_l,y,z)}{\sum_i P_{C_{th}}(x_i,y_i,z_i)}, \quad l \text{ layer index on the transverse } x \text{ axis} \tag{14}$$

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. 14. Globally, Figs. 13 and 14 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 smaller than 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.

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 cumulative 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). For each longitudinal coordinate along the flow direction, the mass is cumulated over the corresponding transversal section of the modeled domain. The ensemble of mass distribution curves and some typical profiles are illustrated separately for each conceptual model in Fig. 15.

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. 15a realization 106 and Fig. 15c realization 182 with a strong attenuation due to mass loss).

The 10, 50 and 90 percentile of the mass distribution curves are shown in Fig. 16.

They confirm the global trends observed with the quantitative indicators (Fig. 12), that MGS and OBJ simulations (Fig. 16(a) and (b)) 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. 16(c) and (d)). 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 histograms of the cumulative mass in the volumes delimited by the three control zones (Fig. 17) for the simulations versus the reference. Indeed, in practical applications, it is common to define a control zone to observe if a contaminant exceeds a concentration. In this exercise, the cumulative mass might be interpreted as a mean concentration over the control zone, as the porosity is assumed constant over the model.

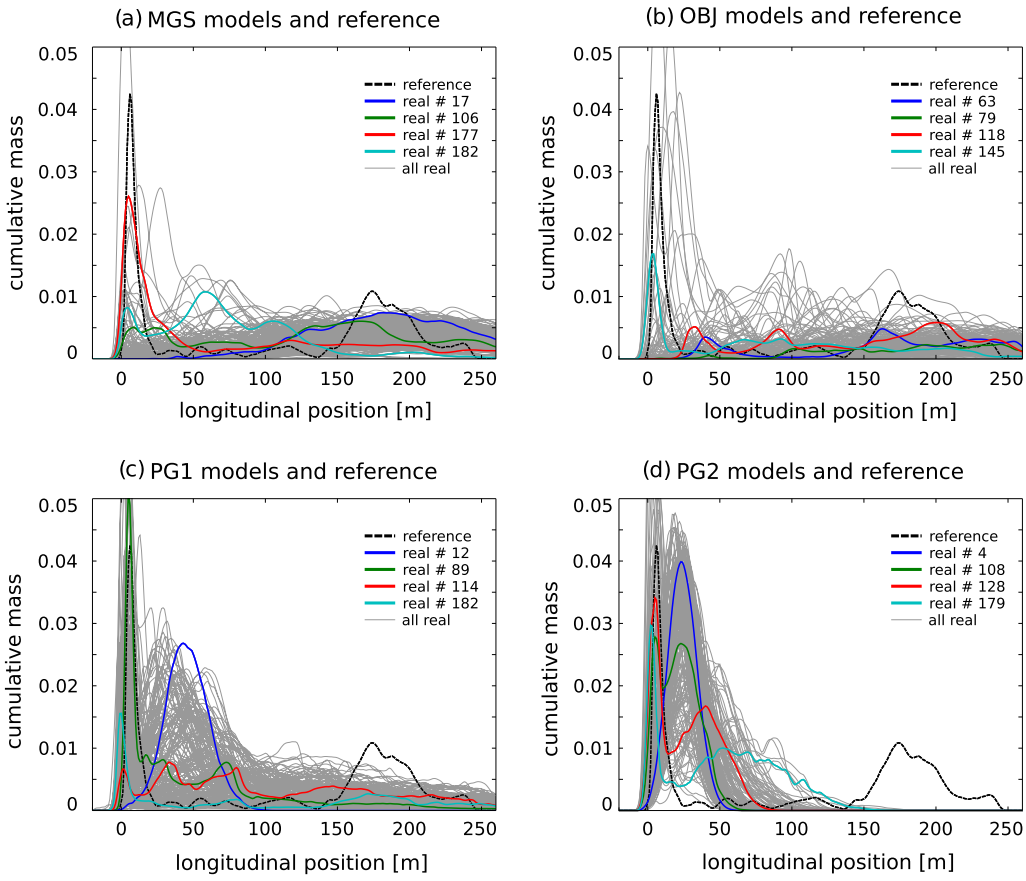

Fig. 15. Cumulative mass along main flow direction at day 328: typical breakthrough curves.

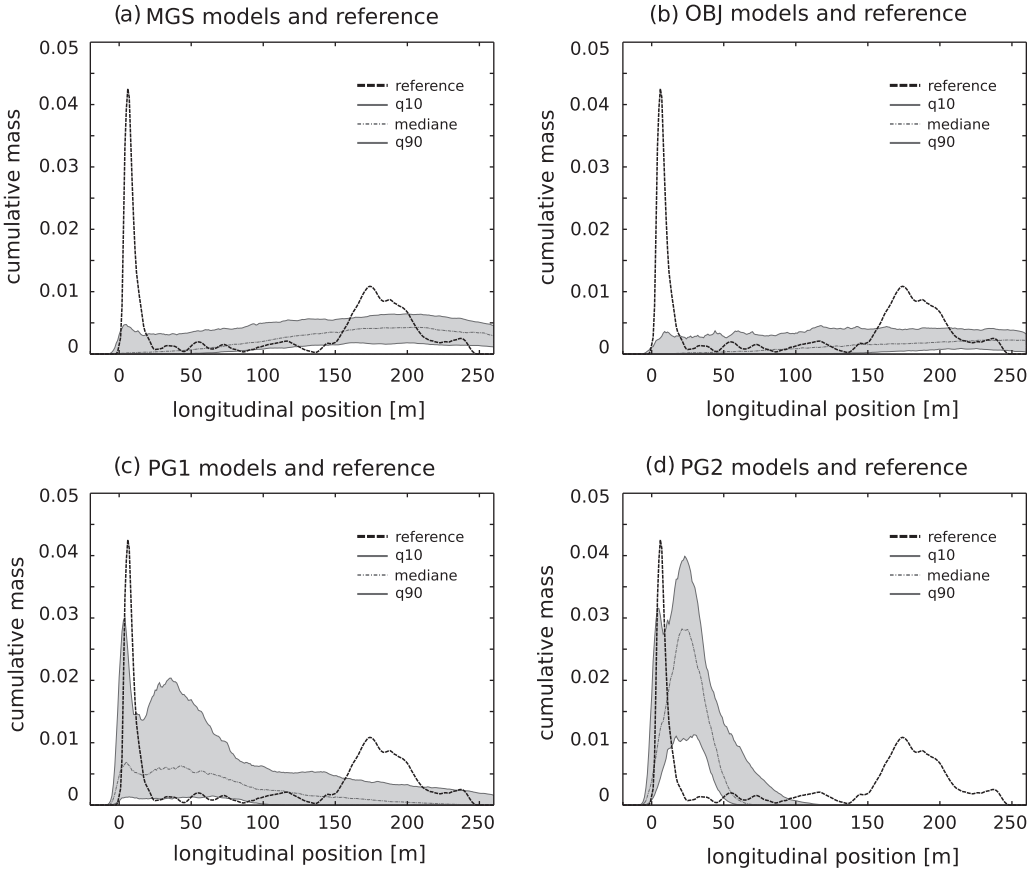

Fig. 16. Cumulative mass along main flow direction at day 328: quantiles on breakthrough curves.

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. In an inversion framework, the PG2 parameterization could have been disregarded. However, would have a zone been further downstream, PG2 realizations 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.

6. Impact of conceptual geological model on plume behavior

One could wonder if the plume speed differences are related to the global equivalent permeability K_{eq}. Let us denote K_{eq}^j the global equivalent permeability for realization j. K_{eq}^j (ms^{-1}) is computed from Darcy's law as $K_{eq}^j = -\frac{Q_j}{A} \times \frac{L}{\Delta h}$ where Q_j ($\mathrm{m}^3\,\mathrm{s}^{-1}$) is the total discharge computed by Groundwater in realization j, A (m^2) is the area of the mean section orthogonal to the flow direction, L (m) is the length of the modeled domain and Δh (m) is the water gradient between the outlet and the inlet of the model. One could expect to observe an increasing longitudinal coordinate of the center of mass at day 328 as a function of the global equivalent permeability.

However, a plot of the plume center of mass longitudinal coordinates as a function of the equivalent permeability (Fig. 18) does not reveal an increasing function.

It shows on one hand that each conceptual model produces a specific behavior, as a result of the geological structures that are proper to each concept. On the other hand, a strong noise exists within each conceptual model and many pairs of simulations (j_1, j_2) present characteristics – $K_{eq}^{j_1} < K_{eq}^{j_2}$ and $Y_G^{j_1} > Y_G^{j_2}$, where Y_G denotes the longitudinal position of the plume center of mass at day 328 – in opposition to what could be expected at a first glance. This intra-variability is probably due to the sensitivity of the initial conditions such as the injection location on the plume dispersion. Some of this noise might also be due to the computation of the plume center of mass at day 328. Indeed if the plume speed is highly heterogeneous, the mass staying within the model boundaries might be trapped in specific locations close to the injection while most of the mass would have made its way out of the model boundaries. A curious aspect is also that MGS simulations present a wide variance of the plume center of mass longitudinal traveling distance for a really thin distribution of equivalent permeabilities. In opposition, PG1 realizations have a huge variability in the equivalent permeability but produce plume traveling distances with a smaller variance. From these general observations arise two main questions. Firstly, what can explain the variance of the equivalent permeability peculiar to each concept? Secondly, which factors influence the variance of the longitudinal traveling distance of the plume center of mass for the different concepts?

A possible influential factor for the equivalent permeability variance might be the size of the heterogeneous structures regarding the size of the modeled domain. Indeed, for MGS models, the dimensions of the structures are really smaller than the model dimensions (see Fig. 5). It allows for statistical stationarity of the conductivity field within each realization. In addition, the realizations share the same conductivity mean and variance and therefore

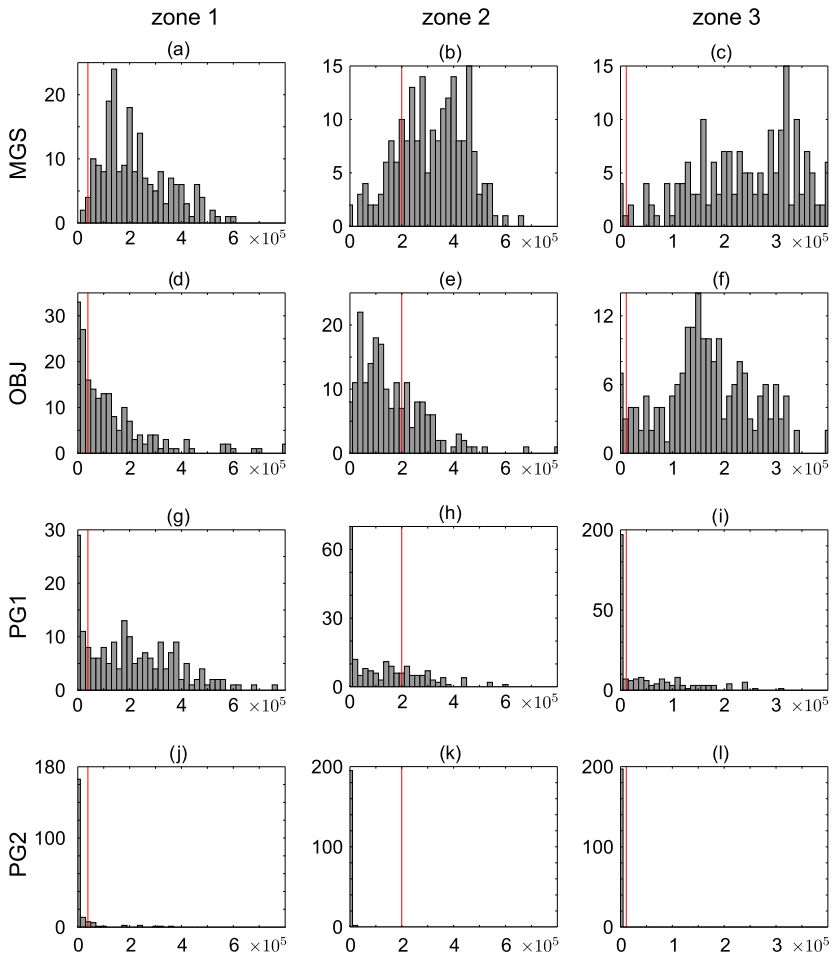

Fig. 17. Histograms of cumulative mass (pCi) predictions for day 328 in the three target zones; reference value: vertical red line.

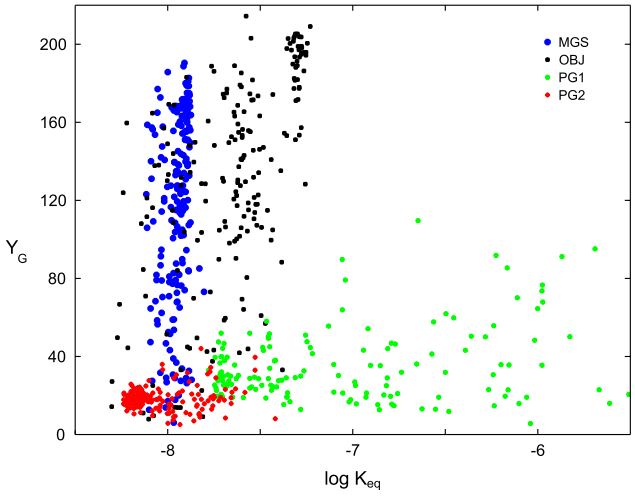

Fig. 18. Plume center of mass longitudinal position at day 328 as a function of the equivalent permeability.

almost the same histogram for MGS realizations due to the field re-normalization step. In that case, the flow paths might have a similar mean behavior and mean power, resulting in a thin equivalent permeability bandwidth. The extreme opposite conditions happens for the PG1 simulations: some objects with high conductivity values have a scale of the same order of magnitude than the domain dimensions and the extreme conductivity values may vary between the realizations (see Figs. A.21 and A.22 in the supplementary material). As the high conductivity values are greater than in the MGS case, it produces extremely strong flow paths crossing the whole domain, which results in higher equivalent permeabilities. Meanwhile, the connectivity of such strong flow path is not systematic, and as the matrix is characterized by a low conductivity, the disconnection length between preferential flow paths reduces the total discharge through the domain and consequently cuts back the equivalent permeability.

As illustrated in the supplementary material (Figs. A.19, A.20, A.21 and A.22), the traveling distance of the plume center of mass depends on the conductivity values at the plume center of mass location 27 days after injection. Indeed, the distance traveled by the plume is related to the proportion and probability of the contaminant mass present in preferential flow path at the initial conditions. In a sense, it is linked to the initial plume dispersion, the size and the density of the preferential flow paths. Though the envelope of the initial plume seems wide (see Fig. 9), its dispersion is rather small and concentrated on a column, as one can see in Fig. 10. For MGS realizations, as the probability for initial plume center of mass to be in a preferential flow path or in a low conductivity zone is the same and that the conductivity fields are continuous, it allows for a wide variance in the center of mass longitudinal displacement. For PG1 realizations, one might notice from conductivity sections (see Figs. 3–5 and A.21) that the probability of the initial plume to be located in a preferential flow path is small, as the high conductivity facies have a smaller proportion than the low conductivity matrix. As the number of realization is limited, here it might not be sufficient to create an example where the initial plume center of mass

would be within a preferential flow path and therefore, we only observe examples where the center of mass does not travel far from the injection location. To confirm this interpretation, a transport simulation was performed with an initial plume shifted so that its center of mass would be located in a preferential flow path. This case is illustrated by an animation in the supplementary material. This animation illustrates the first 13 days of transport simulation on 13 successive frames. In that case, the mass present in the preferential flow path is taken out of the model boundaries in a few days. Then, the mass initially present in the matrix slowly makes its way towards the preferential flow path, but when it reaches it, it is rapidly diluted by the strong flow and nothing is then visible in the preferential flow path.

7. Conclusion

It has been shown that with limited site-specific information, it is possible to propose geological realizations that are coherent with characteristic dimensions of braided-river deposits, based on different types of conceptual models and analog data. Globally, considering the assumption of field information scarcity, the predicted plumes are within reasonable orders of magnitude in terms of concentration and travel time. In this specific case, though the conductivity mean $\mu_{ln(K)} \simeq -9.1$ and variance $\sigma^2_{ln(K)} \simeq 3.1$ values computed from analogue data (Jussel et al., 1994) are different from the conductivity mean $\mu_{ln(K)} \simeq -11.6$ and variance $\sigma^2_{ln(K)} \simeq 6.6$ measured on the MADE site, as reported by Bohling et al. (2012), the K distribution with a smaller variance and higher conductivity mean is still overlapping with the DPIL measured K distribution. Thus combined with various different geological structure might allow reproducing comparable plume speed or dispersion. However, none of the conceptual models was able to predict the plume behavior and uncertainty in a completely satisfying manner with the selected parameters. Considering multiple conceptual models together help to compensate the conceptual model imperfections. In the test case, MGS or OBJ plume realizations could locate the accumulation zone while PG1 or PG2 models could describe the slow release of the contaminant. In a sense, all conceptual models were useful.

A limitation of this 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 can be generated, by varying the different input parameters. Also, the strategy adopted to assign the conductivity values so that the mean and variance of the *log* conductivity field are the same for all realizations reduces the degrees of freedom and is therefore a limiting factor. Considering different values within a given range for the matrix conductivity would have changed the flow and transport conditions and consequently would have lead to other plume shapes and behaviors. 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.

Globally, the variability of the computed indicators is not uniform and strongly depends on the type of indicator, on the observation zone, and on the type of conceptual model. In addition, the fact that errors are sometime (ε_I zone 2, ε_I zone 3) quite far from 0 for all or for most of the realizations and the fact that the reference is not within the quantiles for the cumulative mass distribution along the longitudinal axis suggest that there might not be enough realizations, not enough parameter exploration within each conceptual model (over constrained by the common conductivity mean and variance and conductivity assignment rules) or that the conceptual models used are not adapted.

Regarding the limited data available in the modeling context, one has to remind that no inverse conditioning has been performed, and therefore it would be pure luck to have really small errors on all concentration sample locations simultaneously (impact on ε_C, ε_I and cumulative mass distribution curves). Then, through the observation of the cumulative mass distributions for instance, the predicted plumes might be considered as acceptable and the different conceptual models shall not be simply ousted. Additionally, 200 realizations per conceptual model might not seem enough to quantify the prediction uncertainty but it already requires a large amount of computational time for such multi-million cell models. Moreover some freedom degrees have been restrained in the object-based and pseudo genetic conceptual geological models to fulfill the constraints imposed by a common conductivity mean and variance.

Consequently, further parameter space exploration should be performed. 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 the subsurface structure length scale characteristics with Ground Penetrating Radar surveys. The interpreted GPR data could be transformed into summary statistics, offering a global selection criterion, or it could also be used locally as conditioning data for geological layer interfaces. 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.

Other 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, especially when the high concentrations are very localized as illustrated in the supplementary material.

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 his enthusiasm in sharing the data from the MADE site with us and Fabien Cornaton for his support in the parameterization and use of Groundwater, Dimitri Migraine for his implication in the data preparation and the preliminary work, Ghislain De Marsily, Jaime Gomez-Hernandez and Niklas Linde for their constructive comments, the reviewers and the editor for their will of clarifying and publishing the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhydrol.2015.07.036.These data include MOL files and InChiKeys of the most important compounds described in this article.

References

Anderson, M., Aiken, J., Webb, E., Mickelson, D., 1999. Sedimentology and hydrogeology of two braided stream deposits. Sed. Geol. 129, 187–199.

Barlebo, H.C., Hill, M.C., Rosbjerg, D., 2004. Investigating the Macrodispersion Experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model. Water Resour. Res. 40, W04211.
Bayer, P., Huggenberger, P., Renard, P., Comunian, A., 2011. Three-dimensional high resolution fluvio-glacial aquifer analog – Part 1: Field study. J. Hydrol. 405, 1–9.
Boggs, J.M., Adams, E.E., 1992. Field study of dispersion in a heterogeneous aquifer – Part 4: Investigation of adsorption and sampling bias. Water Resour. Res. 28, 3325–3336.
Boggs, J.M., Beard, L.M., Waldrop, W.R., Stauffer, T.B., MacIntyre, W.G., Antworth, C.P., 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., Liu, G., Knobbe, S.J., Reboulet, E.C., Hyndman, D.W., Dietrich, P., Butler, J.J., 2012. Geostatistical analysis of centimeter-scale hydraulic conductivity variations at the MADE site. Water Resour. Res. 48, W02525.
Bowling, J.C., Rodriguez, A.B., Harry, D.L., Zheng, C., 2005. Delineating alluvial aquifer heterogeneity using resistivity and GPR data. Groundwater 43, 890–903.
Brierley, G.J., 1989. River planform facies models: the sedimentology of braided, wandering and meandering reaches of the Squamish River, British Columbia. Sed. Geol. 61, 17–35.
Comunian, A., Renard, P., 2009. Introducing wwhypda: a world-wide collaborative hydrogeological parameters database. Hydrogeol. J. 17, 481–489.
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.
De Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. Hydrogeol. J. 13, 161–183.
dell'Arciprete, D., Bersezio, R., Felletti, F., Giudici, M., Comunian, A., Renard, P., 2012. Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments. Hydrogeol. J. 20, 299–311.
Dogan, M., Van Dam, R.L., Bohling, G.C., Butler, J.J., Hyndman, D.W., 2011. Hydrostratigraphic analysis of the MADE site with full-resolution GPR and direct-push hydraulic profiling. Geophys. Res. Lett. 38, L06405.
Emery, X., Lantuéjoul, C., 2006. Tbsim: A computer program for conditional simulation of three-dimensional gaussian random fields via the turning bands method. Comput. Geosci. 32, 1615–1628.
Falivene, O., Arbus, P., Gardiner, A., Pickup, G., Muoz, J.A., Cabrera, L., 2006. Best practice stochastic facies modeling from a channel-fill turbidite sandstone analog (the Quarry outcrop, Eocene Ainsa basin, northeast Spain). AAPG Bull. 90, 1003–1029.
Feehley, C.E., Zheng, C., Molz, F.J., 2000. A dual-domain mass transfer approach for modeling solute transport in heterogeneous aquifers: application to the macrodispersion experiment (MADE) site. Water Resour. Res. 36, 2501–2515.
FOEN, 2009. Groundwater in switzerland. <http://www.bafu.admin.ch/grundwasser/07496/07516/index.html?lang=en>.
Guan, J., Molz, F.J., Zhou, Q., Liu, H.H., Zheng, C., 2008. Behavior of the mass transfer coefficient during the MADE-2 experiment: new insights. Water Resour. Res. 44, W02423.
Heinz, J., Kleineidam, S., Teutsch, G., Aigner, T., 2003. Heterogeneity patterns of quaternary glaciofluvial gravel bodies (SW-Germany): application to hydrogeology. Sed. Geol. 158, 1–23.
Huber, E., 2015. Incorporating Sedimentological Observations, Hydrogeophysics and Conceptual Knowledge to Constrain 3D Numerical Heterogeneity Models of Alluvial Systems. Ph.D. Thesis. University of Basel.
Huber, E., Huggenberger, P., 2015. Morphological perspective on the sedimentary characteristics of a coarse, braided reach: Tagliamento River (NE Italy). Geomorphology 248, 111–124.
Huber, E., Huggenberger, P., Caers, J., submitted for publication. Geophysical stereology: Quantifying the uncertainty on 3D subsurface structure from 2D GPR data – application to the Tagliamento River. J. Appl. Geophys. (2015).
Huggenberger, P., Regli, C., 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.
Journel, A.G., 1974. Geostatistics for conditional simulation of ore bodies. Econ. Geol. 69, 673–687.
Jussel, P., Stauffer, F., Dracos, T., 1994. Transport modeling in heterogeneous aquifers: 1. Statistical description and numerical generation of gravel deposits. Water Resour. Res. 30, 1803–1817.
Klingbeil, R., Kleineidam, S., Asprion, U., Aigner, T., Teutsch, G., 1999. Relating lithofacies to hydrofacies: outcrop-based hydrogeological characterisation of quaternary gravel deposits. Sed. Geol. 129, 299–310.
Koltermann, C.E., Gorelick, S.M., 1996. Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. Water Resour. Res. 32, 2617–2658.
Lane, S., Westaway, R., Hicks, D., 2003. Estimation of erosion and deposition volumes in a large, gravel-bed, braided river using synoptic remote sensing. Earth Surf. Proc. Land. 28, 249–271.
Lee, S.Y., Carle, S.F., Fogg, G.E., 2007. Geologic heterogeneity and a comparison of two geostatistical models: sequential Gaussian and transition probability-based geostatistical simulation. Adv. Water Resour. 30, 1914–1932.
Llopis-Albert, C., Capilla, J.E., 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). J. Hydrol. 371, 75–84.
Mariethoz, G., Renard, P., Straubhaar, J., 2010. The direct sampling method to perform multiple-point geostatistical simulations. Water Resour. Res. 46, W11536.
Matheron, G., 1973. The intrinsic random functions and their applications. Adv. Appl. Prob. 5, 439–468.
Middelkoop, H., Daamen, K., Gellens, D., Grabs, W., Kwadijk, J.C., Lang, H., Parmet, B.W., Schädler, B., Schulla, J., Wilke, K., 2001. Impact of climate change on hydrological regimes and water resources management in the Rhine basin. Climatic Change 49, 105–128.
Moeck, C., 2014. Evaluating the Effect of Climate Change on Groundwater Resources: From Local to Catchment Scale. Ph.D. Thesis. University of Neuchâtel, Switzerland.
Pirot, G., 2015. Stochastic Heterogeneity Modeling of Braided River Aquifers – A Methodology Based on Multiple Point Statistics and Analog Data. Ph.D. Thesis. University of Neuchâtel. <http://doc.rero.ch/record/234460/files/00002442.pdf>.
Pirot, G., Straubhaar, J., Renard, P., 2014. Simulation of braided river elevation model time series with multiple-point statistics. Geomorphology 214, 148–156.
Pirot, G., Straubhaar, J., Renard, P., submitted for publication. A pseudo genetic model of coarse braided-river deposit. Water Resour. Res. (2015).
Renard, P., De Marsily, G., 1997. Calculating equivalent permeability: a review. Adv. Water Resour. 20, 253–278.
Salamon, P., Fernàndez-Garcia, D., Gómez-Hernández, J., 2007. Modeling tracer transport at the MADE site: the importance of heterogeneity. Water Resour. Res. 43, W08404.
Sanchez-Vila, X., Guadagnini, A., Carrera, J., 2006. Representative hydraulic conductivities in saturated groundwater flow. Rev. Geophys. 44, RG3002.
Siegenthaler, C., Huggenberger, P., 1993. Pleistocene rhine gravel: deposits of a braided river system with dominant pool preservation. Geol. Soc. 75, 147–162 (London, Special Publications).
Siirila-Woodburn, E.R., Maxwell, R.M., 2015. A heterogeneity model comparison of highly resolved statistically anisotropic aquifers. Adv. Water Resour. 75, 53–66.
Teles, V., Delay, F., De Marsily, G., 2004. Comparison of genesis and geostatistical methods for characterizing the heterogeneity of alluvial media: groundwater flow and transport simulations. J. Hydrol. 294, 103–121.
Webb, E.K., 1994. Simulating the three-dimensional distribution of sediment units in braided-stream deposits. J. Sediment. Res. 64, 219–231.
Zhang, Y., Green, C.T., Fogg, G.E., 2013. The impact of medium architecture of alluvial settings on non-Fickian transport. Adv. Water Resour. 54, 78–99.
Zheng, C., Bianchi, M., Gorelick, S.M., 2011. Lessons learned from 25 years of research at the MADE site. Groundwater 49, 649–662.
Zinn, B., Harvey, C.F., 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 Resour. Res. 39, 1051.