Published in Advances in Water Resources 31, issue 1, 147-159, 2008
which should be used for any reference to this work

Issues in characterizing heterogeneity and connectivity in non-multiGaussian media

Jaouher Kerrou [a,*], Philippe Renard [a], Harrie-Jan Hendricks Franssen [b], Ivan Lunati [c]

[a] *Centre for Hydrogeology, University of Neuchâtel, Rue Emile Argand 11 – CP158, CH-2009 Neuchâtel, Switzerland*
[b] *Institute of Environmental Engineering, ETH Zurich, CH-8093 Zurich, Switzerland*
[c] *Laboratory of Soil and Environmental Physics, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

Abstract

The performances of kriging, stochastic simulations and sequential self-calibration inversion are assessed when characterizing a non-multiGaussian synthetic 2D braided channel aquifer. The comparison is based on a series of criteria such as the reproduction of the original reference transmissivity or head fields, but also in terms of accuracy of flow and transport (capture zone) forecasts when the flow conditions are modified. We observe that the errors remain large even for a dense data network. In addition some unexpected behaviours are observed when large transmissivity datasets are used. In particular, we observe an increase of the bias with the number of transmissivity data and an increasing uncertainty with the number of head data. This is interpreted as a consequence of the use of an inadequate multiGaussian stochastic model that is not able to reproduce the connectivity of the original field.

Keywords: Aquifer characterization; Inverse; Kriging; Stochastic simulations; Uncertainty; Well-capture zones; Connectivity; Multiple-point statistics

1. Introduction

Groundwater flow and transport are controlled by physical properties that are characterized by a high degree of heterogeneity and by scales of variation that span several orders of magnitude. A major difficulty is that this heterogeneity, whose knowledge is fundamental for modelling relevant environmental problems (e.g. protection zone design, contaminant migration prediction, aquifer remediation, seawater intrusion), has to be inferred on the basis of sparse measurements. In the past decades, a large number of techniques has been developed with the aim of characterizing the spatial variability of aquifer parameters and their uncertainty [11,26,31]. Generally speaking, the characterization of the heterogeneity can be addressed based on direct observations of the *physical parameters* (direct methods) or on observations of the *state variables* of the system (inverse methods).

Direct methods infer the distribution of the physical parameters (transmissivity, porosity, etc.) from the local information about the parameters themselves. Note, however, that this information is often obtained by solving an inverse problem involving the state variables (e.g. pumping tests) but their interpretation is local and provides parameter values that become the input of the characterization methods. Additional direct information about the characteristics and the position of geological facies can be provided by geophysical observations. Among the most widely employed direct interpolation techniques are kriging and stochastic multiGaussian simulations. They both are two-point geostatistical methods that proved efficient in a wide variety of applications in hydrogeology but also in other fields such as mining, or petroleum engineering [4,13,25].

* Corresponding author. Tel.: +41 32 7182677; fax: +41 32 7182603.
E-mail addresses: jaouher.kerrou@unine.ch (J. Kerrou), philippe.renard@unine.ch (P. Renard), hendricks@ifu.baug.ethz.ch (H.-J. Hendricks Franssen), ivan.lunati@epfl.ch (I. Lunati).

In practice, however, direct methods are seldom used alone because they do not account for the global information on state variables (e.g. hydraulic head or concentration) and lead to groundwater models that do not reproduce the observed values of those state variables. Instead, this information is taken into account by inverse methods, which characterize the physical parameters from measurements of the state variables. Inverse modelling has been a topic of intense research and developments [1,5,12,19,21]. As argued in a recent review by Carrera et al. [9], most methods do not differ from each other in essence, but they may differ with respect to the computational details. Among the inverse techniques, the Monte-Carlo approach, in which multiple equally likely realisations of aquifer properties are conditioned to hydraulic head and concentration data, allows estimating uncertainty. Inverse methods have been applied successfully in a wide range of problems [7,27,28,41].

Since both direct and inverse methods rely on measurements acquired at few discrete locations, some hypotheses have to be made on the parameter statistics in order to infer the continuum distribution. Most two-point geostatistical techniques illustrated in the previous sections are based on the assumption that the physical parameters follow a multiGaussian distribution, which is analytically simple and fully characterized by a mean and covariance function. Numerical testing is usually performed by applying the characterization techniques to synthetic fields that also feature multiGaussian statistics. The results demonstrate the accuracy and the consistency of the methods. While remaining in the multiGaussian framework, few studies have shown how uncertainty can be reduced by increasing the number of transmissivity, head or concentration data [19,40].

MultiGaussian fields maximise entropy (disorder), and in return, minimize the spatial continuity of the extreme values, thus, a loss of connectivity [24]. This feature has a high impact on flow and transport as shown by a number of numerical investigations [17,33,39,42]. While these studies were limited to the context of direct techniques, they all showed that the selection of a multiGaussian model might be consequential on flow and transport simulations. Both direct and inverse techniques are available to handle non-multiGaussian media. Examples of direct techniques include the sequential indicator simulation method [16], truncated pluriGaussian simulations [2], or multiple-point statistics [36]. Examples of inverse methods that are able to handle non-multiGaussian media are the conditional probabilities method [8], a combination of truncated pluriGaussian simulation and the gradual deformation approach [21], the inverse modelling of multimodal hydraulic conductivity distributions with the representer method [22], or the probability perturbation method combined with multiple-point statistics [6] among others.

Although it is known already now for more than a decade that multiGaussian models have severe limitations, and although alternative methods exists, most groundwater hydrology studies adopt a multiGaussian model, often also because data are not available to infer a non-multiGaussian model. Therefore, this study explores what happens if a multiGaussian model is adopted for a non-multiGaussian medium, a situation that is most probably very common in practice. In particular, we do not only investigate the implications of a wrong random function model in direct studies, but also in inverse problems.

In summary, the goal of the present study is to investigate and compare the reliability of direct and inverse multiGaussian techniques when applied to characterize fields that are not multiGaussian and exhibit preferential flow paths. A main question of interest in this study was to what extend hydraulic head data, used in inverse modelling, are able to correct the consequences of the wrong assumption of a multiGaussian random function model. Two situations may occur: either the inverse conditioning is able to detect non-multiGaussian structures and would alleviate the problems associated with a wrong random function (multiGaussian), or the discrepancy between the model and the reality would still lead to inaccurate predictions. In the latter case, checking the multiGaussianity assumption for a particular case study would be extremely important; moreover, the groundwater community would be encouraged to further develop and adopt methods based on different statistics.

After creating a synthetic reality, our methodology mimics the procedure that would be followed during a practical case study. We start constructing a synthetic transmissivity field such that it possibly represents a real aquifer characterized by long-correlation structures such as channels and lenses (Section 2.1). A reference head field, which mimics natural flow conditions, is obtained by simulating the flow on this transmissivity field (Section 2.2). The transmissivity and the head fields are sampled in order to obtain a series of datasets with an increasing number of data points, which represent synthetic experimental data to be used as input for the aquifer characterization (Section 3.1). At this point, we first analyse the datasets in order to infer the statistics required for the characterization step and compute histograms and variograms (Section 3.2). Then, for each dataset, we reconstruct three transmissivity fields by applying three multiGaussian characterization techniques, i.e. kriging, stochastic direct simulations and self-calibrated sequential simulations (Sections 3.3 and 3.4). The performance of the characterization techniques is evaluated both in terms of reproduction of the real transmissivity field and reproduction of the initial flow situation (Section 4.1). Most important, the inferred transmissivity fields are used to make predictions on different flow scenarios. In particular, we consider the response of the aquifer to the construction of a well. Forecasts in terms of total fluxes through the domain, head in the pumping well, and protection-zone extension around the pumping well are considered (Sections 4.2 and 4.3). The reliability of the techniques is estimated as a function of the number of transmissivity and head data used to condition the transmissivity fields.

2. Synthetic reality

2.1. Transmissivity field

The study is conducted on a synthetic transmissivity (T) field, which consists of channels and lenses displaying internal heterogeneity (Fig. 1a–c). It is built from an aerial photograph displaying braided channels in the Ohau River, New Zealand [30]. According to the classification of natural rivers from Rosgen [32], the architecture displayed on this photograph belongs to the *type D: braided channels*. This kind of sedimentary environment can be regarded as the ancestor of our synthetic aquifer.

The Ohau aerial photograph is digitized and used at its real scale so that the size of the channels and lenses is realistic. The image size is 1000 m by 400 m and discretized into 1 m by 1 m cells. Two multiGaussian unconditional simulations are separately generated to populate the channels and the lenses with natural logarithm of transmissivity (T) values. The first simulation describes the T distribution in the channels and has an exponential variogram with a short correlation range (3 m). The second simulation describes the lenses and has a nested variogram that includes one isotropic exponential model with a 3 m range, plus a cubic anisotropic model with a long range in the x direction (600 m) and a shorter one in the y direction (300 m). Note that the mean and the variance of the two distributions were chosen such that the values are realistic for such a geological environment. The long-range correlation is used to mimic a regional trend in the deposition of fine sediments. This leads to a bimodal, non-multiGaussian, anisotropic transmissivity filed with about 50% of the surface occupied by channels (highly permeable coarse material) and 50% occupied by lenses (poorly permeable fine material).

The decimal logarithm of T has a mean of -2.3 and a variance of 0.67, which corresponds to $\sigma^2_{\ln(T)} = 3.6$ (note that the variance is computed considering the T field as a whole, regardless to its bimodal nature). The overall integral scale was estimated to be 27 m in the x direction (approximately 2.7% of the length of the domain) and 8 m in the y direction (2% of the domain width) by integrating numerically the correlation function calculated from the reference T field.

2.2. Reference flow

A reference 2D flow field, which mimics natural flow conditions, is obtained by prescribing the hydraulic heads on the eastern and western boundaries, and imposing no-flow conditions on the northern and the southern boundaries (Fig. 1d). It is assumed that the aquifer is confined and there is no recharge (no source term). The flow problem on the reference transmissivity field is solved with the Feflow code [14] and a head distribution is obtained that reflects the flux variations generated by the permeability contrast between the channels and the lenses (Fig. 1d). The reference head field will be sampled to provide the input data for the inverse characterization technique.

3. Characterization procedure

The characterization procedure consists in applying independently or successively different techniques. The starting point is the sampling of the transmissivity field to obtain the data for computing the experimental histogram

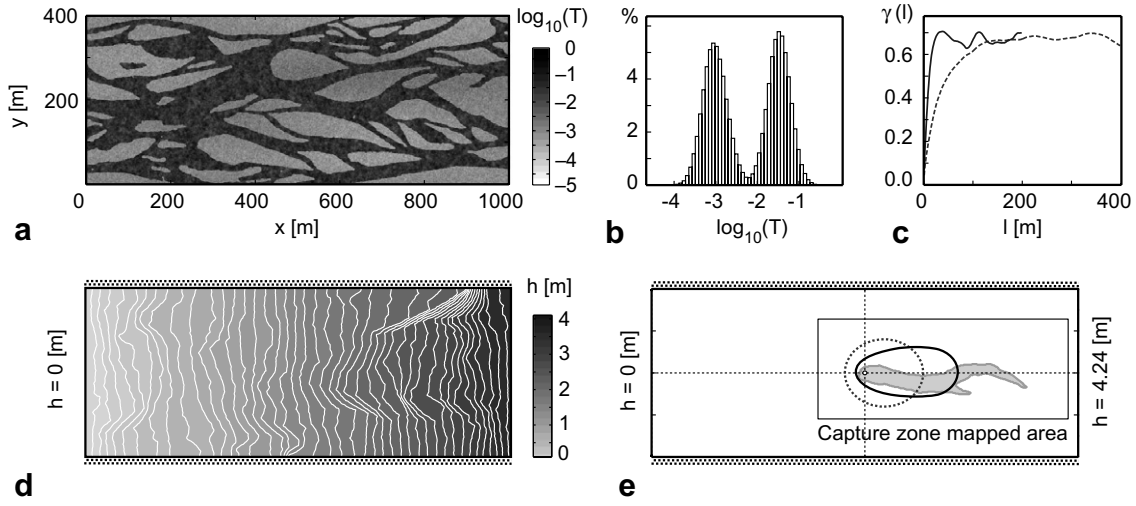

Fig. 1. (a) The synthetic transmissivity field, (b) histogram of the log decimal of the transmissivities, (c) x (dashed line) and y (solid line) directional variograms, (d) head field and boundary conditions for the uniform steady flow situation, (e) 10-days capture zone under the radial steady flow conditions (gray shape) and the mapped area; the 10-days capture zone calculated using the analytical solution of Bear and Jacobs [3] on the basis of the geometric mean (dashed line) and using the equivalent transmissivity (solid line).

and variogram. In addition, either the data are interpolated by kriging, or a set of stochastic simulations conditional to transmissivity measurements is generated. These stochastic simulations can be additionally conditioned to hydraulic-head data obtained by sampling the reference flow field. Finally, we consider the case in which an exhaustive knowledge on the geology is available (i.e. channel location and structure are exactly known).

3.1. Sampling the transmissivity and head fields

To simulate the aquifer characterization procedure in a real case study, the reference transmissivity field and the head field are sampled at random locations in order to mimic field measurements. For each variable three datasets consisting of 21, 250, and 1000 measurements are obtained. (Note that transmissivity and heads are sampled at identical locations.) Fig. 2c and 2f shows the sample locations (circles) of the 21 and 250 datasets, respectively (1000-dataset locations are not shown because they are too dense). The transmissivity and the head datasets represent error-free local data. The mean distance between the samples is 89 m, 26 m, and 12 m, respectively. In order to avoid unit-dependent indicators and include the integral scale in our reasoning, the mean distance is normalized by the integral scale in the x direction. The corresponding dimensionless mean distances, d, are 3.5, 1, and 0.47 for 21, 250, and 1000 samples, respectively. In other words, the three datasets represent measurements whose mean spacing ranges from three times to half the integral scale. Note that the reference image is anisotropic and that the integral scale in the y direction is much smaller, even smaller than the mean distance between the samples in the 1000-measurement case. The statistics of the decimal log-transmissivity of the three datasets used are presented in Table 1.

3.2. Experimental variograms

The three transmissivity datasets (25, 250 and 1000 T samples) are analysed separately in a geostatistical framework. The 250 and 1000 T samples clearly exhibit a bimodal histogram (Fig. 2d and g). Before analysing the spatial correlation, the $\log_{10} T$ values are transformed into a normal variable N via a Gaussian transformation since the simulation algorithms require to work with normal distributed variables. Note that, after comparison with the normal score transform, Hermite polynomials decomposition was selected to compute the Gaussian transform as it provided a slightly better histogram reproduction for the direct and back transform. Then, the assumption of bi-Gaussianity of the normal variable N is tested and cannot be rejected, even if this may be

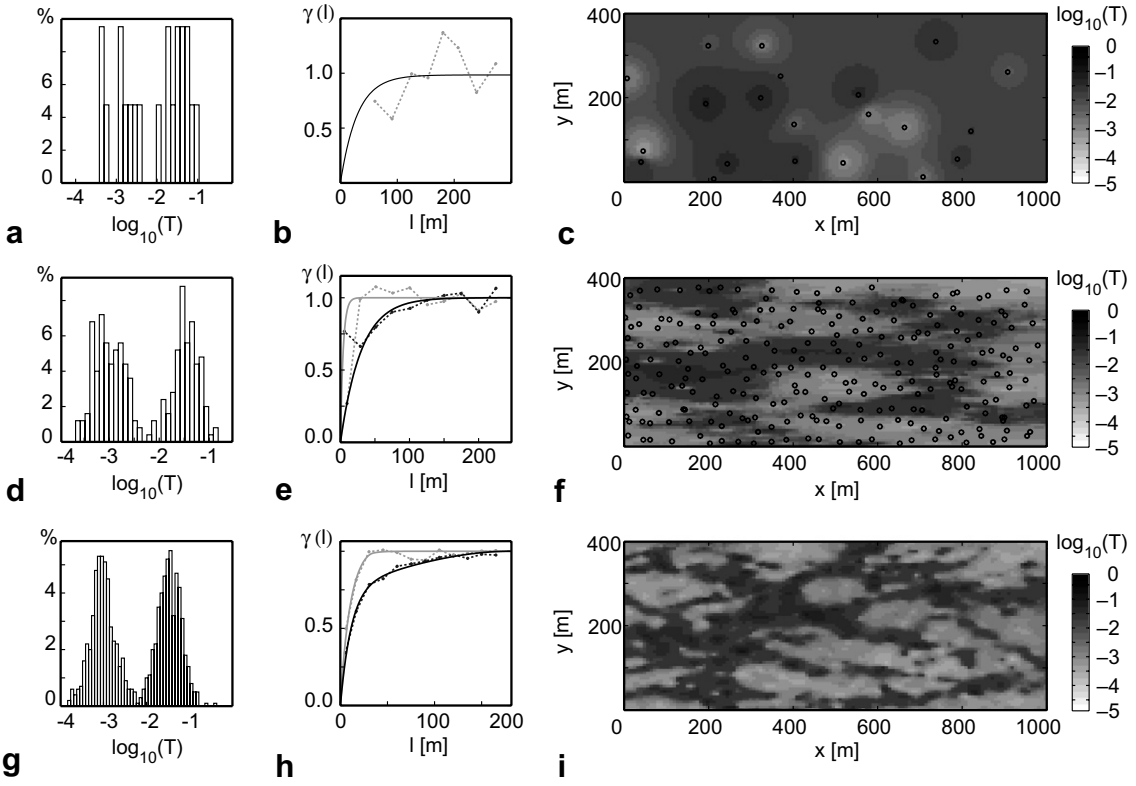

Fig. 2. T fields estimation steps for 21 T (a–c), 250 T (d–f), and 1000 T data (g–i): (a, d, and g) histogram of the T data, (b, e and h) experimental (dashed line) and model (solid line) variograms of the N Gaussian transform of the decimal log of the T data in both x (black line) and y (gray line) directions except for 21 T which was isotropic, (c, f and i) kriged maps and T samples locations except for 1000 T because of its high density.

Table 1
Samples statistics and adjusted variograms, N represents the Gaussian transform of $\log_{10}(T)$

Number of T samples [m]	21	250	1000
Average distance between samples [m]	89	26	12
Dimensionless distance d [–]	3.5	1	0.47
Mean $\log_{10}(T)$ [m^2/s]	−2.06	−2.28	−2.23
Minimum $\log_{10}(T)$ [m^2/s]	−3.31	−3.86	−3.82
Maximum $\log_{10}(T)$ [m^2/s]	−1.03	−0.82	−0.37
Std. $\log_{10}(T)$ [m^2/s]	0.77	0.82	0.82
N variogram type	Exponential	Exponential	Exponential + Spherical
Nugget effect	0	0	0
Variogram sill	0.98	0.99	0.77 + 0.23
Variogram range along x [m]	89	85	35 & 185
Variogram range along y [m]	89	15	25 & 35

surprising since the transmissivity data are clearly not bi-Gaussian. After analysing the anisotropy of the variogram maps of the transformed data, directional experimental variograms are calculated and modelled for each dataset in both x and y directions (Fig. 2b, e and h). Note that the experimental variogram of the 21 T samples did not show any anisotropy, thus an isotropic exponential model is used in this case. The variogram of 1000 T showed a nested structure, which is modelled by one exponential and one spherical model. Cross validation is performed to test whether the fitted variogram models are acceptable. Note that no nugget effect is considered since data are known to be error free. Furthermore, the absence of nugget effects allows a higher degree of control on the T field during the inverse calibration.

3.3. Kriging and conditional simulations

First, ordinary kriging of the N normally distributed values is applied to each transmissivity dataset (21, 250, 1000 T samples) with the corresponding modelled variogram. The T fields (Fig. 2c, f and i) are obtained by back-transforming the kriged N field in two steps. First, N is back transformed into $\log_{10}(T)$ with the inverse Gaussian transform (Hermite polynomials) based on the histogram of the data, then we take the decimal power of this field to obtain the transmissivity. Note that the last part of the back transform is often corrected to avoid the bias that it induces in the value of the arithmetic mean of T, but for 2D aquifers we should be more concerned with the bias in the geometric mean of T rather than in the arithmetic mean because the later controls the mean flow through the system. Taking the decimal power preserves the geometric mean, therefore this is what we did. Then, 100 conditional stochastic simulations of the normally distributed N values are generated for each dataset using the Turning Band Method and back transformed into T fields. The upper row of Fig. 3 shows examples of the resulting T-field simulations (the lower row shows how these fields are improved by inverse modelling). In the end, the geometric means of the simulations, of the kriging and of each dataset have been compared and are very similar.

3.4. Inverse modelling

The three sets of stochastic simulations of log-transmissivity can, in addition, be conditioned to the hydraulic-head data by means of the sequential self-calibration method as implemented in the INVERTO code [20]. Six pairs of data are considered with an increasing number of transmissivity (T) and head (h) samples, i.e. (21 T, 21 h), (21 T, 250 h), (21 T, 1000 h), (250 T, 250 h), (250 T, 1000 h), (1000 T, 1000 h). Note that the number of conditioning head data is always at least as large as the number of transmissivity data. For each pair, 100 equally likely inverse realisations are generated that are conditioned both to the transmissivity and hydraulic-head data.

During the inversion, the variogram estimated from the data (Section 3.2), is used as a model to interpolate the perturbations optimised at the master blocks. Two master blocks are laid out per correlation length, and the position of the master blocks is modified during the inverse conditioning. The lower row of Fig. 3 shows examples of the resulting transmissivity fields.

3.5. Characterization with exhaustive geological conditioning

In addition to the multiGaussian characterization presented in the previous sections, a test was conducted to estimate the efficiency of the characterization methods when the location of the channels would be exhaustively known. This is clearly unrealistic but it constitutes an end member that allows investigating what is the best estimate that can be obtained if one would know the position of the channels. We considered only the dataset with 21 T samples, but in addition we have a binary map that indicates if a pixel is within a channel or not. The 21 T samples are then divided into two groups, according to their locations in a channel or in a lens, and they are analysed and modelled separately. Note that, due to the lack of data, the two variograms are very weakly constrained. The resulting models are spherical with a range of 85 m for the channels data and 190 m for the lenses. Following the procedure presented above, Turning Band simulations are used to separately populate the channels and the lenses and generate 100 T fields.

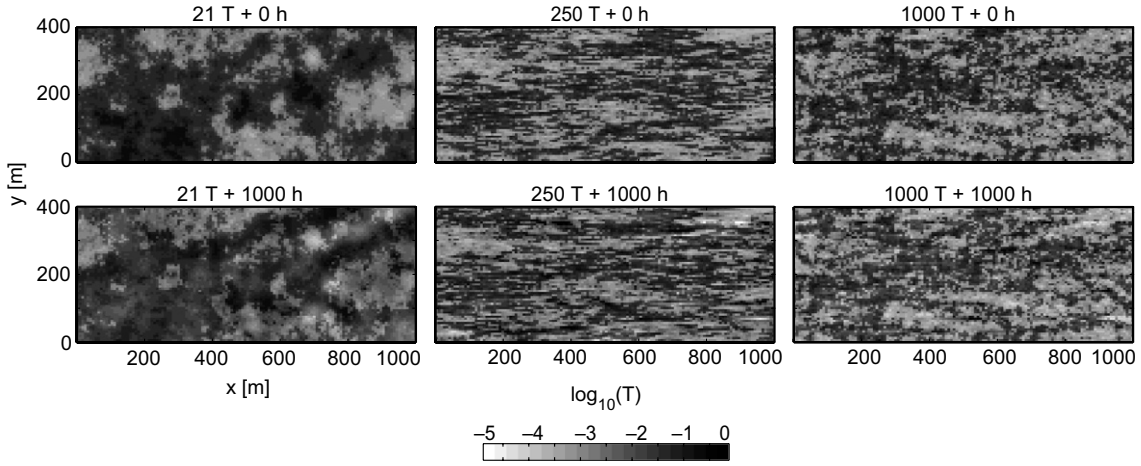

Fig. 3. Simulations conditioned only to the T values (upper maps), and the same simulations conditioned to T and h measurements according to the map title.

Finally, these fields are processed with INVERTO to produce 100 equally likely realisations conditional to 21 transmissivity data, 250 hydraulic-head data, and to the exhaustive geological information. In the inverse conditioning, each geological unit is separately perturbed. This means that perturbations in lenses do not affect the hydraulic conductivities in channels and vice versa. Both in the lenses and the channels 100 master blocks each are placed. In that case, INVERTO uses the geological knowledge to optimize the local values of the transmissivity by having different variograms for the different facies.

4. Numerical results

In order to evaluate the reliability of the characterization techniques, the estimated T fields and the simulated h fields are compared with the corresponding reference fields. Note that the two techniques provide different kind of results: the estimated field is unique and aims at providing a locally accurate map, the simulations aim at reproducing the variogram. Consequently, their comparison may be theoretically questionable, but we argue that it has practical relevance and this is why we perform it. Moreover, since we want to assess the ability of predicting the behaviour of the aquifer under different flow scenarios, the estimated transmissivity fields are used to forecast the system response when a pumping well is located in the middle of the domain.

4.1. Reproduction of the reference T and h fields

The kriged T fields and some examples of simulations conditional to transmissivity and head data are shown in Figs. 2 and 3. By visual comparison with the reference transmissivity field (Fig. 1) it is evident that a large number of samples is required to start recovering the channel structure. Even in these cases, the simulated images are quite different from the reference.

To quantify this discrepancy and assess the accuracy both in terms of reproduction of the log-transmissivity and hydraulic-head fields, we introduce the average absolute error,

$$\varepsilon(X) = \frac{1}{n}\sum_{i=1}^{n}\left|\overline{X}_{s,i} - X_{\mathrm{ref},i}\right|, \tag{1}$$

and the average ensemble standard deviation,

$$\varepsilon_\sigma(X) = \frac{1}{n}\sum_{i=1}^{n}\sigma_{X_i}, \tag{2}$$

where n is the number of grid cells, i a grid cell index, X the variable considered (decimal log-transmissivity or steady-state hydraulic head), the overbar indicates ensemble average, $\sigma_{X_i} = \left[\overline{\left(X_{s,i}^2\right)} - \left(\overline{X}_{s,i}\right)^2\right]^{1/2}$ is the ensemble standard deviation of X at a given node, the subscript s refers to the realisations, and the subscript ref to the reference (synthetic) values.

Table 2 presents calculated average absolute error ε and average ensemble standard deviation ε_σ for transmissivity and head fields of all the combinations of T and h data. From the results presented in Table 2, one can observe

Table 2
Performance measures for the generated ensembles, conditioned to different amounts of data. The absolute average error ε and average ensemble standard deviation ε_σ are expressed in percent of the values calculated for 21 T

Conditioning data	$\varepsilon(Y)$	$\varepsilon_\sigma(Y)$	$\varepsilon(h)$	$\varepsilon_\sigma(h)$
$T = 21$	100.0	100.0	100.0	100.0
$T = 21$, $H = 21$	98.9	90.1	41.9	19.5
$T = 21$, $H = 250$	94.3	80.9	13.2	8.1
$T = 21$, $H = 250$ Known geology	48.0	54.4	7.2	4.5
$T = 21$, $H = 1000$	92.9	74.9	9.2	6.1
$T = 250$, $H = 250$	79.5	106.7	11.0	7.4
$T = 250$, $H = 1000$	79.4	106.0	8.7	6.5
$T = 1000$, $H = 1000$	64.9	87.0	10.0	5.7

that, in general, conditioning improves the characterisation of the log-transmissivity and hydraulic-head fields, and reduces the ensemble variance as expected. However, the reduction is smaller than observed in similar multiGaussian studies [20]. It is noteworthy to indicate that the uncertainty reduction for the $\log T$ and h estimation ($\varepsilon_\sigma(Y)$ and $\varepsilon_\sigma(h)$) are in most cases smaller than the improvement in the characterisation of the log-transmissivity field ($\varepsilon(Y)$) and the hydraulic head field ($\varepsilon(h)$), unlike the studies with a multiGaussian model for log-transmissivity [20].

Note that, in spite of using a wrong $\log T$ model in the inverse conditioning and in spite of a somewhat limited improvement in the characterisation of the $\log T$ field, the characterisation of the hydraulic head field improves spectacularly. Overall, the inverse modelling improves the characterisation of hydraulic-head and log-transmissivity fields and reduces uncertainty, but compared with multiGaussian cases using the correct random function model the improvements tend to be smaller. Note that in case of 1000 head samples no clear trend can be observed when the number of T samples is increased from 21 to 250 and 1000.

As expected, knowing the geology (the position of channels and lenses) considerably improves both the characterisation of the log-transmissivity field and the reproduction of the heads.

4.2. Forecasting the flow: fluxes and heads

In this section, the transmissivity fields obtained from the characterization procedures are used to predict the system behaviour under a different flow scenario, i.e. when a pumping well is added in the middle of the domain. The boundary conditions are identical as for the reference flow, but an additional pumping well with a constant flow rate of 700 m^3/d is added in the middle of the domain ($x = 500$ m, $y = 200$ m).

Two criteria are used to evaluate the performance of the T estimated fields: the accuracy in estimating the outflux at western boundary (Q_{out}) and the well-bore head (H_w). Fig. 4 shows the histograms of Q_{out} and H_w for some of the considered cases. The vertical lines represent the reference value calculated with the reference T field, the ensemble average from the simulations, and the value calculated with the kriged field. Note that heads and fluxes estimated with the kriged field are different from those obtained by the ensemble averages of the simulations. This is due to the fact that even if the flow equations are linear for heads, they are not linear in terms of hydraulic conductivities. However, none of these two estimations can be said more accurate than the other.

Without conditioning to head data, we observe a reduction of the uncertainty on the forecast heads or fluxes when the number of T data increases (compare Fig. 4a, i and m or Fig. 4b, j and n). It is instructive to compare the results provided by an exhaustive geological knowledge with those provided by conditioning to 1000 transmissivity data (Fig. 4e–f and m–n, respectively). As expected, the information on the position of channels and lenses truly improves the knowledge of the system, as proved by the reduction of both errors and uncertainty. In the case of 1000 T data, instead, the uncertainty reduces, but the error does not, as can be observed in Fig. 4n, where the histogram of the forecast outfluxes does not contain the reference value. The uncertainty has reduced but the forecasts are incorrect. This clearly shows the risk of conditioning to many T data with a wrong statistical model. The systematically smaller fluxes indicate that the connectivity is lower than in the reference field.

Adding head data generally has a positive impact but does not necessarily reduce the uncertainty. When the number of T data is small, conditioning with head data reduces uncertainty (compare Fig. 4a and c for example). It was surprising to observe that when the number of T data is very large, conditioning on heads increases the uncertainty (Fig. 4n and p). However, it has to be observed that this effect is such that the forecast histograms always include the reference values. In other words, it seems that if too many points are conditioned to transmissivity, information on the hydraulic heads is not able to substantially affect the flux ensemble average, but it improves the correctness of the prediction by increasing the uncertainty. This result will be addressed in more detail in the discussion.

4.3. Capture zone forecast and performance analysis

As a last comparison between the characterisation methods, we assess the accuracy of the different T fields in forecasting the 10-days capture zone of the pumping well, which is delineated by simulating advective–dispersive transport. We assume a constant porosity equal to 0.3, whereas the longitudinal and transversal dispersivities are small and equal to 2 and 0.2 m, respectively. This yields an advection-dominated transport. The capture zones are calculated by solving the Kolmogorov backward equation [29,37]. The mean-life expectancy is calculated with the GroundWater finite element code [10] and the 10-days capture zone is defined as the region around the well where the mean-life expectancies are less than 10 days (Fig. 1e). Note that the shape of the capture zone is strongly controlled by the local transmissivity distribution in the vicinity of the well. The reference capture zone is depicted in Fig. 1e, together with the capture zones calculated by means of the analytical solution of Bear and Jacobs [3] for a purely advective transport of an inert solute. Shown are both results obtained by using the geometric mean of the T reference field and the equivalent homogenous transmissivity obtained from the Darcy flux. Note that the 10-days capture zone resulting from the geometric mean reflects the assumed isotropy of the field.

In the direct conditional simulation approach and in the inverse approach, several simulations have been constructed for a given dataset. In these two cases, the

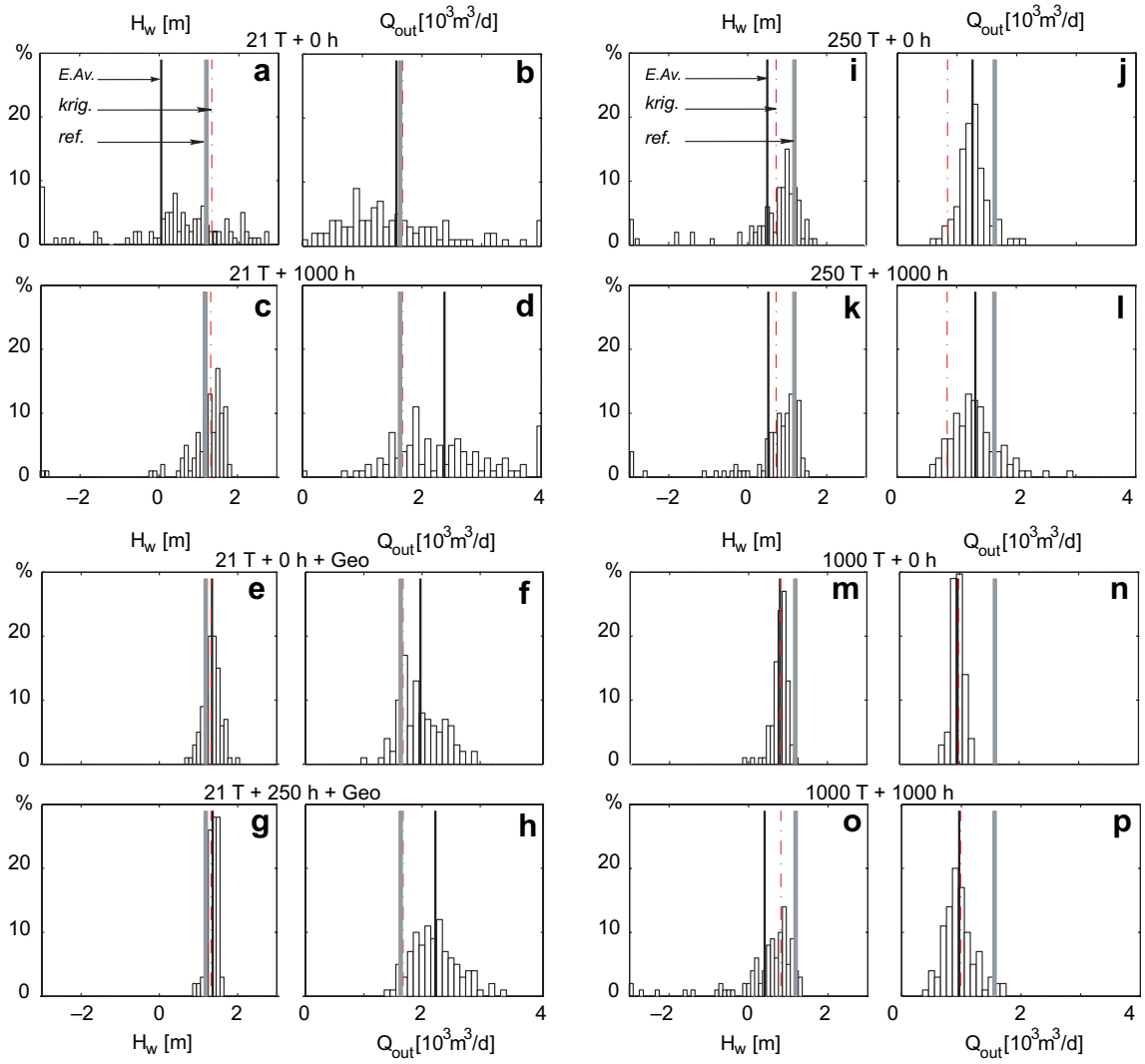

Fig. 4. Histograms of the final head in the well (a, e, i and m) for the simulations conditioned only with T; (c, g, k and o) the same plus 1000 head data according to the title. The integral out flux (b, f, j and n) for the simulations conditioned only with T; (d, h, l and p) the same plus 1000 head data according to the title. For each set of the 100 simulations: the thick gray line represents the reference field values (ref.); the solid thin line represents the ensemble average of the simulations (E.Av.), the dash-points line represents the kriged values (krig.).

uncertainty is estimated by constructing the probability maps for a point to belong to the capture zone. These maps are shown in Fig. 5. Visual inspection of the results (Fig. 5) reveals that the forecasts can significantly differ from the reference. We note that the reference is not always completely included in the forecasted capture zones. Indeed, the eastern finger of the reference capture zone is predicted neither by kriging, nor by the 0.5 isoprobability contour of the simulations, nor by the geometric mean. The latter leads to a prediction very similar to that obtained with the analytical isotropic solution of Bear and Jacobs [3]. As observed before, the a priori knowledge of the geology significantly improves the characterization of the T fields. If the position of the channels is known, the predicted capture zone is already accurate with a small number of T data (21) and even better when additional head data (250) are used for conditioning the transmissivity field.

To quantify the accuracy of the characterization methods, we introduce two error norms for comparing the forecast capture zone, Z, with the reference capture zone, Z_{ref}: the missed area,

$$e_{\mathrm{ma}} = \frac{N_{\mathrm{p}}}{B_{\mathrm{p}} + N_{\mathrm{p}}}, \tag{3}$$

and the unnecessarily protected area,

$$e_{\mathrm{uc}} = \frac{P_l}{B_{\mathrm{p}} + P_l}, \tag{4}$$

where N_{p} [m^2] is the area of the reference protection zone that is not correctly identified by the forecast, i.e. the area of $Z_{\mathrm{ref}} \cap Z$, B_{p} [m^2] is the area of the reference protection zone that is correctly forecasted, i.e. the area of $Z \backslash Z_{\mathrm{ref}}$, P_l [m^2] is the area wrongly forecasted as belonging to the protection zone, i.e. the area of $Z_{\mathrm{ref}} \backslash Z$. In other words, e_{ma} is the percentage of the reference that has not been

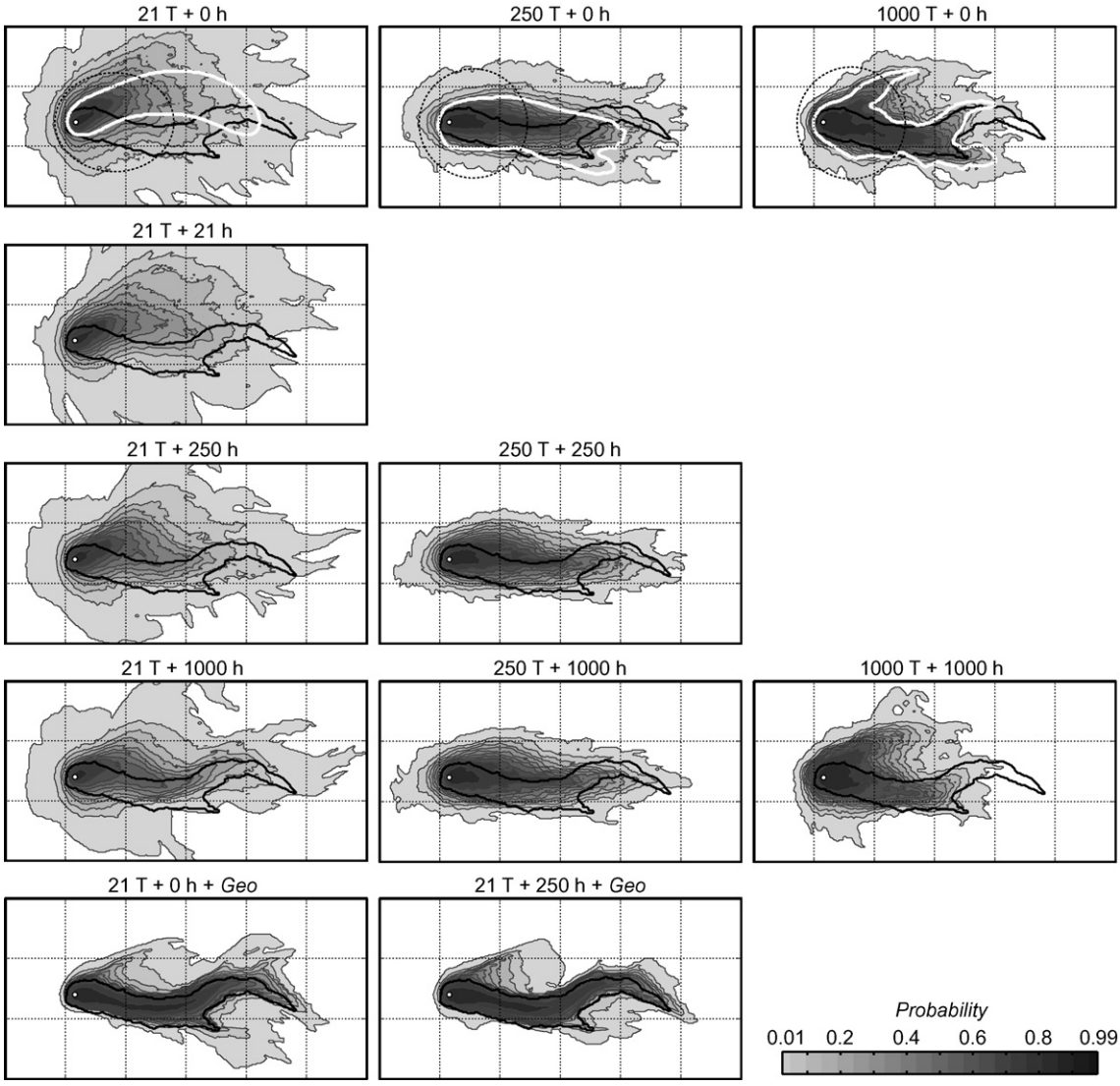

Fig. 5. 10-Days capture zone probability maps. The black line represents the reference 10-days capture zone. The white line (in the upper maps) represents the zone forecasted by kriging. The gray levels represent the isoprobability contours for all the combinations of T and h datasets (according to the map title). The dashed line on the upper maps represents the 10-days capture zone for an homogenous field in which T is equal to the geometric mean of the T samples.

identified, whereas e_{uc} is the percentage of the forecast that is unnecessarily protected.

In addition, an evaluation of the total error is provided by the average squared residual between the forecasted (Z) and the reference well-capture zone (Z_{ref}),

$$\varepsilon_s = \frac{1}{100}\sum_{s=1}^{100}\left(\frac{1}{n}\sum_{i=1}^{n}(Z_{s,i} - Z_{ref,i})^2\right), \quad (5)$$

where s indicates the simulation, i is the grid node and n the total number of nodes in the grid. s, resp. ref, is an indicator variable, equal to 1 if the node belongs to Z, resp. Z_{ref}, and 0 if not. Finally, the uncertainty is quantified by the dimensionless ratio between the uncertain area and the reference protection zone,

$$u = \frac{I}{N_p + B_p}, \quad (6)$$

where $N_p + B_p$ [m^2] is the area of the reference protection zone, Z_{ref}, and I [m^2] is the area located between the 0.9 and 0.1 isoprobability contours.

The missed area e_{ma} and unnecessarily protected area e_{uc} errors are calculated for different levels of probability threshold (0.2, 0.5 and 0.8) and are summarized in Fig. 6. It can be observed that the average squared residual (ε_s) is almost identical for kriging and for the simulations conditioned only on T (Fig. 6a) and that ε_s decreases regularly with decreasing d. Adding head data clearly reduces ε_s only when the mean distance d between the sample is large with respect to the integral scale, i.e. when the simulations are conditioned to few T data. The best results are obtained with the geological knowledge.

When kriging is employed, ε_s, the missed area (e_{ma}), and the unnecessarily protected area (e_{uc}) decrease when d decreases (Fig. 6a, and c–f). When d is lower than 1, the

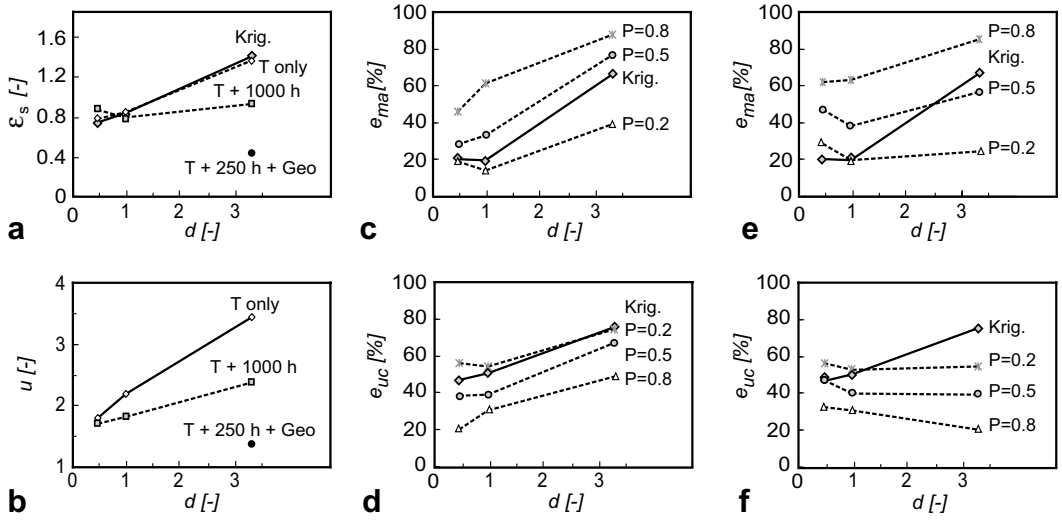

Fig. 6. Results: (a) the averaged squared error between forecasted and reference capture zone vs. mean distance between the samples for the three methods, (b) uncertainty vs. mean distance between the samples, (c) e_{ma} (missed area) for 0.2, 0.5, and 0.8 probability threshold vs. distance between the samples with transmissivity data only, (d) e_{uc} (unnecessarily protected area) for 0.2, 0.5, and 0.8 probability threshold vs. distance between the samples with transmissivity data only, (e) and (f): same as (c) and (d) but 1000 additional head measurements used to condition the T fields.

errors do not reduce significantly (or stabilize) when new samples are added. In the best cases, the missed area represents around 20% of the reference protection zone and the area unnecessarily protected represents around 45% of the forecasted zone.

When stochastic simulations are employed, the uncertainty area u can be estimated. We find that it varies between 1.5 and 3.5 times the real area of the reference capture zone. Fig. 6b shows how u decreases regularly when d decreases and that adding head data reduces the uncertainty very significantly, especially when the distance d between the transmissivity data is large.

The probability maps obtained from stochastic simulations (Fig. 5) are tools to help making decisions. The level of risk can be quantified by the probability threshold P that is employed to define the capture zone. A risk-prone decision maker may for example take a probability threshold of 0.5 while a conservative decision maker may take a probability threshold of 0.8. Fig. 6c and d shows the evolution of the missed area and unnecessarily protected area as a function of d, and of the probability threshold P in the case of simulations conditioned only to T data, while Fig. 6e and 6f shows the equivalent when 1000 h data are used in addition. These results suggest that the error decreases with d only if the latter is larger than 1. When conditioning to head data, it does not appear clear whether conditioning to more T data improves the characterization. On the contrary, in case 1000 T data are used for conditioning, the unnecessarily protected area is larger than in case only 21 T data are used for conditioning (using in both cases 1000 head data for inverse conditioning). There are also some counterintuitive results that show that when the distance between the sample is small and when a large number of head and transmissivity data are provided, the errors can be larger than with a smaller number of data (see for example the curves of the unnecessarily protected area e_{uc} in Fig. 6f and for a probability threshold of 0.2; e_{uc} decreases when the distance between the samples decreases). This is in contrast with what was observed for the fluxes, where the inverse simulations yield a better histogram than the simulations conditioned only to T. In all cases, the most remarkable fact is that errors remain large even with a considerable amount of data.

5. Discussion

The numerical simulations presented in this paper consider a 2D aquifer that displays channels and lenses and follows non-multiGaussian statistics. Our approach has been to consider this field, which supposed to reflect the properties of a realistic aquifer, as "the reality" and try to characterize it on the basis of simulated local measurements of transmissivity and heads. It has to be remarked that we have used a single T field as reference (as in a real test case there is only one real aquifer) and the results are therefore influenced by the specific transmissivity arrangement and location of the well. However, we believe that the difficulties encountered might be faced anytime that the problem of characterizing a channelized aquifer is addressed. Having located the well in the most permeable region reflects the fact that real wells are placed in transmissive regions in order to optimize efficiency. Our assumption is that the T-field considered qualitatively illustrates the difficulties encountered with media that exhibit a highly connected network of permeable channels. In practice, vertical averaging of the 3D hydraulic conductivity might filter the complexity of the porous media by yielding a model characterized by smoother contrasts between permeable

and less-permeable regions. However, if we accept the picture of the 3D aquifer as having a hydraulic conductivity that displays an interconnected network of permeable channels and isolated poorly permeable lenses, we can expect difficulties similar to those described in this paper. Even more problematic, the smoother 2D data of the transmissivity would tend to additionally hide the extreme values of the hydraulic conductivity, which are expected to play an important role, in particular for transport (e.g. for capture-zone estimation or remediation problems).

Therefore, despite the limitations, we argue that the numerical results presented in the previous section enable some general considerations. Some conclusions are expected, such as the fact that exhaustive geologic knowledge provides most of the necessary information or that samples denser than the integral scale do not significantly improve the characterization. Other conclusions are surprising such as the fact that (1) uncertainty and errors remain very large even in the presence of (unrealistically) many data, (2) conditioning to many T data may reduce uncertainty but increase the bias, and (3) conditioning to many T data seems to inhibit the inverse conditioning from improving the characterization.

Before going further, we observe that the reduction of uncertainty concomitant with an increased bias that has already been mentioned by Scheibe and Chien [34]. They compare multiGaussian characterization techniques based on real field data. The quality criterion was the reproduction of a measured tracer breakthrough curve. They showed that conditioning to a large number of small scale measurements did not significantly improve the model prediction and could even lead to biased and overly confident predictions. Our findings are fully consistent with these results and suggest that problems do not arise because of measurement errors, but are generated by erroneously applying multiGaussian characterization techniques to non-multiGaussian fields.

Why the multiGaussian assumption prevented providing accurate results when the number of data is large? When the mean distance d between the T data is small with respect to the integral scale of the reference, the inference of a variogram model is easy and accurate. The histograms are also well defined. Then, the high density of conditioning T data leads to a situation in which the variability between the simulations is very small. This can be seen on the histograms of the forecasted water levels in the pumping well or in the histogram of the forecasted outflow rates (Figs. 4m and n). The simulated fields are very strongly constrained by the T data, by the variogram model, as well as by the multiGaussian assumption used by the simulation technique. Therefore, the inverse technique has very little degrees of freedom. In case of a large amount of conditioning transmissivity data, the hydraulic head data can not modify the transmissivity field significantly, because the transmissivity data "freeze" the – wrongly postulated – multiGaussian model. Conditioning to head data has only very local impacts and cannot result in significant changes of the large scale structure of the field. This can be seen by comparing the fields before and after inversions in Fig. 3. Important changes are observed with 21 T data, while little changes are visible with 1000 T data.

These observations and the fact that the error norms did not reduce significantly when increasing the number of T conditioning data from 250 to 1000 samples, are in agreement with the findings of Grabow et al. [18] and van Leeuwen et al. [38]. Grabow et al. [18] showed that the reduction of the number of T conditioning data did not increase the error of flow and transport predictions. van Leeuwen et al. [38] showed that beyond a certain threshold of conditioning data density the performance did not improve further. Both Grabow et al. [18] and van Leeuwen et al. [38] explain their results by the screening effect due to redundancy of data. Note that van Leeuwen et al. [38] dealt with a reference multiGaussian field originally avoiding the connectivity problem. This redundancy effect is probably accentuated when narrow preferential flow paths exist.

The synthetic field used in this numerical study displays a bimodal distribution with a high connectivity of the large values. There are clear channels that completely cross the

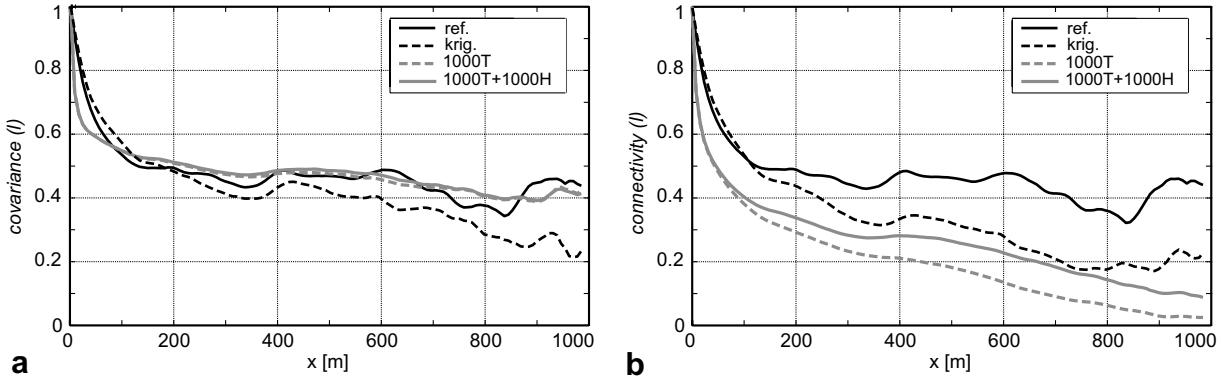

Fig. 7. Directional covariance (a) and connectivity (b) functions along the x-axis for 1000 T data. The lines corresponding to the simulations are ensemble averages over the 100 simulations (1000 T, and 1000 $T + 1000$ h).

domain (Fig. 1). The channels are globally oriented along the x axis and offer a high connectivity in that direction, but locally their orientation varies and they are generally not parallel to the x axis; and they have a wide range of width along the y direction. All this complexity cannot be captured by the variogram, and is smoothed out by the multiGaussian model. Even if the variograms are very well estimated with a large number of T data and even if the T map produced by kriging shows the locations of most of the main channels (Fig. 2i). The kriged map does not reproduce the connectivity because the thin sections of the channels are clogged by low transmissivity values (compare Figs. 2 and 3i).

This visual observation is confirmed by the comparison of the connectivity functions of the reference field and of the kriging or the simulated fields (Fig. 7). We remind that the connectivity function represents the probability that a transmissive cell taken randomly in the domain is connected by a continuous path of adjacent cells of similar transmissivity with another transmissive cell located at a certain lag distance [35]. Fig. 7 shows that all the characterization methods reasonably reproduce the covariance function of the reference (Fig. 7a), but they systematically underestimate the connectivity of the reference (Fig. 7b). The lack of connectivity of the high values explains the tendency of underestimating the flux through the domain.

6. Conclusion

The goal of this work was to investigate and compare the reliability of direct and inverse multiGaussian techniques when applied to characterize fields that are not multiGaussian and exhibit preferential flow paths. For that purpose, the performances of kriging, stochastic simulations and sequential self-calibration method have been compared, as a function of an increasing number of samples of transmissivity and head data on a synthetic braided alluvial aquifer for predicting flow and transport. Multiple error indicators were used.

We found that up to a certain quantity of data, adding head or transmissivity measurements reduces the errors and the uncertainty. However, a limit is reached: when the density of samples becomes high, we observe unexpected outcomes such as increased bias with an increasing number of T data, or increased uncertainty estimation with an increased number of head data. The simulations results suggest that in case a large number of transmissivity data is used for conditioning, the hydraulic head data are less able to correct the consequences of the erroneous multiGaussian model (in terms of the evaluated performance measures). Particularly problematic is the bias with many T-data because the simultaneous decrease of the uncertainty may lead to overestimate the reliability of the results. In any case, errors remain important even with a large quantity of data.

These outcomes can be explained by the fact that two-point multiGaussian characterization techniques are unable to capture the correct connectivity, which plays a primary role in dictating flow and transport. Using a multiGaussian model when the reality is not multiGaussian may lead to inaccurate predictions. These results confirm and extend the opinion expressed by Gómez-Hernández and Wen [17] that modellers must be extremely careful when taking the decision of using a multiGaussian model.

In order to overcome these limitations, it is extremely important to use a stochastic model that allows reproducing the connectivity of the original field. Along that direction, we argue that techniques such as the multiple-point statistics [15,23,36] may allow that. Moreover, secondary information, such as geophysics, should be used whenever possible to infer the connectivity structures and to constrain the stochastic model.

Acknowledgements

The authors thank R Ababou, SM Gorelick, P Perrochet, and F Stauffer for their early comments on the manuscript, F. Cornaton for providing the GroundWater code, as well as M. Bierkens, A. Journel and an anonymous reviewer for their positive and constructive suggestions. Partial funding of this research was provided by the Swiss National Science Foundation (Grants: 207020-110017/1 and PP002-106557).

References

[1] Alcolea A, Carrera J, Medina A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Adv Water Res 2006;29(11):1678–89.
[2] Armstrong M, Galli A, Le Loc'h G, Geffroy F, Eschard R. Plurigaussian simulations in geosciences. Berlin: Springer; 2003.
[3] Bear J, Jacobs M. On the movement of water bodies injected into aquifers. J Hydrol 1965;3:37–57.
[4] Caers J. Petroleum geostatistics. Society of Petroleum Engineers; 2005.
[5] Caers J. Comparing the probability perturbation method with the gradual deformation method. Math Geol 2007;9(2).
[6] Caers J, Hoffman T. The probability perturbation method: a new look at Bayesian inverse modeling. Math Geol 2006;38(1):81–100.
[7] Capilla JE, Gomez-Hernandez JJ, Sahuquillo A. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric head data – 3. Application to the culebra formation at the waste isolation pilot plan (WIPP), New Mexico, USA. J Hydrol 1998;207(3–4):254–69.
[8] Capilla JE, Rodrigo J, Gómez-Hernández JJ. Simulation of non-Gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information. Math Geol 1999;31(7):907–27.
[9] Carrera J, Alcolea A, Agustín M, Hidalgo J, Slooten LJ. Inverse problem in hydrogeology. Hydrogeol J 2005;13:206–22.
[10] Cornaton F. Deterministic models of groundwater age, life expectancy and transit time distribution in advective-dispersive systems. PhD dissertation, Université de Neuchâtel; 2003.
[11] de Marsily G, Delay F, Gonçalvès J, Renard P, Teles V, Violette S. Dealing with spatial heterogeneity. Hydrogeol. J 2005;13(1):161–83.
[12] de Marsily G, Delhomme J-P, Delay F, Buoro A. 40 years of inverse problems in hydrogeology. CR Acad Sci Ser IIA Earth Planet Sci 1999;329(2):73–87.
[13] Delhomme JP. Spatial variability and uncertainty in groundwater flow parameters. Water Resour Res 1979;15(2):281–90.

[14] Diersch HG. Interactive, graphics-based finite element simulation system FEFLOW for modeling groundwater flow, contaminant mass and heat transport. WASY Institute for Water Resource Planning and System Research Ltd., Berlin, Germany; 1996.
[15] Feyen L, Caers J. Quantifying geological uncertainty for flow and transport modelling in multi-modal heterogeneous formations. Adv Water Res 2006;29(6):912–29.
[16] Gómez-Hernández JJ, Srivastava RM. ISIM3D: an ANSI-C three-dimensional multiple indicator conditional simulation model. Comp Geosci 1990;16(4):395–440.
[17] Gómez-Hernández JJ, Wen X-H. To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. Adv Water Resour 1998;21(1):47–61.
[18] Grabow GL, Mote CR, Sanders WL, Smoot JL, Yoder DC. Groundwater monitoring network design using minimum well density. Water Sci Technol 1993;28(3–5):327–35.
[19] Hendricks Franssen H-J, Gómez-Hernández JJ, Sahuquillo A. Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data. J Hydrol 2003;281(4):281–95.
[20] Hendricks Franssen H-J. Inverse stochastic modeling of groundwater flow and mass transport. PhD dissertation, Technical University of Valencia; 2001.
[21] Hu LY, Le Ravalec M, Blanc G. Gradual deformation and iterative calibration of truncated Gaussian simulations. Petrol Geosci 2001;7:S25–30.
[22] Janssen GMCM, Valstar JR, van der Zee SEATM. Inverse modeling of multimodal conductivity distributions. Water Resour Res 2006;42(W03410):doi:doi:10.1029/2005WR004356.
[23] Journel A, Zhang T. The necessity of a multiple-point prior model. Math Geol 2006;38(5):591–610.
[24] Journel AG, Deutsch CV. Entropy and spatial disorder. Math Geol 1993;25(3):329–55.
[25] Journel AG, Huijbregts CJ. Mining geostatistics. Academic Press; 1978.
[26] Koltermann CE, Gorelick SM. Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. Water Resour Res 1996;32(9):2617–58.
[27] Larocque M, Delay F, Banton O. A comparison of two stochastic inverse methods in a field-scale application. Ground Water 2003;41(1):15–23.
[28] Lavenue M, de Marsily G. Three-dimensional interference test interpretation in a fractured aquifer using the pilot point inverse method. Water Resour Res 2001;37(11):2659–75.
[29] Mariethoz G, Renard P, Cornaton F, Jaquet O. High resolution stochastic modelling of aquifers, example for a contamination migration problem. In: Proceedings of the Swiss geoscience meeting. Bern, Switzerland; November 24–25th, 2006.
[30] Mosley MP. Analysis of the effect of changing discharge on channel morphology and instream uses in a braided river, Ohau River, New Zealand. Water Resour Res 1982;18(4):800–12.
[31] Renard P, Gomez-Hernandez JJ, Ezzedine S. Characterisation of porous and fractured media. In: Anderson M, editor. Encyclopedia of hydrological sciences. John Wiley & Sons Ltd; 2005.
[32] Rosgen DL. A classification of natural rivers. CATENA 1994;22:169–99.
[33] Sánchez-Vila X, Carrera J, Girardi JP. Scale effects in transmissivity. J Hydrol 1996;183(1-2):1–22.
[34] Scheibe TD, Chien Y. An evaluation of conditioning data for solute transport prediction. Ground Water 2003;41(2):128–41.
[35] Stauffer D, Aharony A. Introduction to percolation theory. London: Taylor and Francis; 1994.
[36] Strebelle S. Conditional simulation of complex geological structures using multiple-point statistics. Math Geol 2002;34(1):1–21.
[37] Uffink GJM. Application of the Kolmogorov's backward equation in random walk simulation of groundwater contaminant transport. In: Kobus HE, Kinzelbach W, editors. Contaminant transport in groundwater. Rotterdam, The Netherlands: A.A. Balkema; 1989.
[38] van Leeuwen M, Butler AP, te Stroet CBM, Tompkins JA. Stochastic determination of well-capture zones conditioned on regular grids of transmissivity measurements. Water Resour Res 2000;36(4):949–57.
[39] Wen X-H, Gómez-Hernández JJ. Numerical modeling of macrodispersion in heterogeneous media: a comparison of multiGaussian and non-multiGaussian models. J Contam Hydrol 1998;30(1-2):129–56.
[40] Wen X-H, Gómez-Hernández JJ, Capilla JE, Sahuquillo A. Significance of conditioning to piezometric head data for predictions of mass transport in groundwater modeling. Math Geol 1996;28(7):951–68.
[41] Zimmerman DA, de Marsily G, Gotway CA, Marietta MG, Axness CL, Beauheim RL, et al. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour Res 1998;34(6):1373–413.
[42] Zinn B, Harvey CF. 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 2003;39(3):1051. doi:1010.1029/2001WR001146.