Journal of Hydrology (2008) 355, 76–90

available at www.sciencedirect.com

ScienceDirect

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

Regularized pilot points method for reproducing the effect of small scale variability: Application to simulations of contaminant transport

Andrés Alcolea [a,*], Jesús Carrera [b,1], Agustín Medina [a,2]

[a] School of Civil Engineering. Universitat Politecnica de Catalunya, Jordi Girona, 1-3, 08034 Barcelona, Spain
[b] Institut Jaume Almera de Ciencies de la Terra, Consejo Superior de Investigaciones Cientificas, Lluis Sole Sabaris st n/n, 08028 Barcelona, Spain

Received 13 August 2007; received in revised form 2 February 2008; accepted 5 March 2008

KEYWORDS
Small scale variability;
Heterogeneity;
Hydraulic conductivity;
Tailing;
Inverse problem;
Regularized pilot points method

Summary Small scale variability of hydraulic conductivity controls ground water contaminant transport and some of the subtle aspects of transport through heterogeneous media (e.g., mixing and tailing). Unfortunately, it cannot be identified. This paper addresses the question of whether the presence of high frequency fluctuations, which define small scale variability, impedes the characterization of large scale variability patterns. In parallel, we investigate whether including small scale variability allows us to reproduce tailing in breakthrough curves. To this end, we solve the inverse problem using the regularized pilot points method for simulating fields of hydraulic conductivity conditioned to available information. Heterogeneity of hydraulic conductivity is represented by two nested variograms simulating small scale (short range variogram) and large scale (long range) variability patterns. Calibrated fields are applied to the prediction of a transport problem. Application to four synthetic examples (with varying importance of the small scale variability) shows that, first, the calibrations reproduce the statistics of the ''true'' fields and, second, that a good characterization of the small scale variability is not critical for groundwater flow modeling. More important, small scale variability leads to increased tailing in solute breakthrough curves and needs to be acknowledged for proper transport prediction, which is rarely the case in aquifer modeling practice.

* Corresponding author. Present address: CHYN, Centre d'Hydrogéologie, Université de Neuchâtel, Rue Emile-Argand 11, CH-2009 Neuchâtel, Switzerland. Tel.: +41 32 718 25 97.
E-mail addresses: andres.alcolea@unine.ch (A. Alcolea), jcarrera@ija.csic.es (J. Carrera), agustin.medina@upc.edu (A. Medina).
[1] Tel.: +34 93 409 54 10.
[2] Tel.: +34 93 401 69 15.

0022-1694/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2008.03.004

Introduction

Characterization of heterogeneity is essential for contaminant transport. Solutions obtained using the advection-dispersion equation (ADE) while ignoring spatial variability display numerous departures from field observations (Carrera, 1993). These include the well known scale effects of dispersivity (Lallemand-Barres and Peaudecerf, 1978; Gelhar et al., 1985, 1992; Neuman, 1990), but also time and directional dependence of apparent porosity (Sanchez-Vila and Carrera, 1997; Neuman, 2005) and tailing in breakthrough curves (Adams and Gelhar, 1992; Kennedy and Lennox, 2001; Fernandez-Garcia et al., 2005). The literature shows consistent discrepancies between the breakthrough curves (BTCs) predicted by the ADE and the measured ones (Vallocchi, 1985; Carrera, 1993; Kosakowski, 2004). Tailing becomes critical for remediation systems and contaminant migration whenever small concentrations are of concern, because these may last much longer than predicted by the ADE. In fact, tailing, which proves the non-Fickian nature of solute transport, is ubiquitous and can be found even in ''homogeneous'' laboratory experiments (Silliman and Simpson, 1987; Levy and Berkowitz, 2003; Cortis and Berkowitz, 2004). These experiments show that tailing can be attributed to small scale heterogeneity, which causes solute mass to enter into low permeability regions to trail behind the bulk of the plume. This causes, first, spreading and, eventually, mixing. Therefore, understanding and simulating tailing does allow to separate these two concepts, which is important because mixing may control chemical reactions (De Simoni et al., 2007) while spreading controls plume extent (Dentz and Carrera, 2007).

Tailing in BTCs is usually simulated by adding terms representing the exchange of solute between mobile and immobile regions (multirate mass transfer) to the ADE (Rasmuson, 1985; Haggerty and Gorelick, 1995). These terms can be represented by means of a memory function (Carrera et al., 1998; Haggerty et al., 2000). The concept has been generalized by the continuous time random walk approach (see a review in Berkowitz et al., 2006), which allows for a systematic study of the transition from normal (Fickian) to anomalous (non-Fickian) transport behavior. However, none of these approaches links explicitly the additional terms to actual variability patterns, so that it is not possible to define them a priori.

While tailing may be attributed to heterogeneity, stochastic research has concentrated on explaining the scale growth of observed dispersivities (Dagan, 1986). In this context, the universal scaling theory of Neuman (1990) is particularly relevant to our work. According to this theory, hydraulic conductivity displays many scales of heterogeneity at any given sample size. In fact, it is this superposition what explains the observed scale dependence of dispersivity. Large scale variability can be characterized stochastically and, one hopes, identified with some accuracy using geological maps, geophysics, model calibration, etc. Small scale variability is defined by (high frequency) fluctuations at length scales smaller than typical distances between boreholes. Thus, its characterization is difficult, even hardly possible, with usually available data. As a result, small scale variability is often disregarded in hydrogeological modeling. In fact, Rubin et al. (2003) developed an approach to define dispersivity as a function of the scale of variability truncated by modeling. Unfortunately, it is not known if the superposition of variability scales reproduces tailing. Certainly, ignoring small scale variability does not improve the reproduction of late time behavior of solute transport.

Identification of heterogeneity can be achieved using conditional estimation or conditional simulation methods. The first group seeks a deterministic, though uncertain, optimum characterization in the sense of minimum estimation error, honoring all available data (typically, hydraulic conductivity and head measurements). This group includes linearized cokriging (Kitanidis and Vomvoris, 1983), conditional expectation (Dagan, 1985) and maximum likelihood estimation (Carrera and Neuman, 1986), among others. While formulations of this group are different, they do not vary from each other in the essence (Carrera et al., 2005). All of them neglect the effect of small scale variability, and the estimated fields are inherently smooth. This limitation can be overcome by conditional moment equations, where one seeks estimates of mean parameters while acknowledging the effect of small scale variability (Guadagnini and Neuman, 1999, 2001; Guadagnini et al., 2003; Hernandez et al., 2003, 2006). Conditional simulation methods are explicitly stochastic and can accommodate small scale variability. They yield a number of equally likely realizations of the unknown field conditioned to all available information (Sahuquillo et al., 1992; Gómez-Hernández et al., 1997; Capilla et al., 1997; Hendricks Franssen, 2001; Hendricks Franssen et al., 2003).

The pilot points method (de Marsily et al., 1984; RamaRao et al., 1995; Lavenue and Pickens, 1992) is a flexible parameterization technique, that can be used both for conditional estimation and for conditional simulation. The method facilitates accommodating small scale variability. It has been successfully applied to a number of problems (RamaRao et al., 1995; Vesselinov et al., 2001; Hernandez et al., 2003). This approach can help in realizing the hope of simulating hydraulic conductivity fields that are consistent with available large scale data and yet contain high frequency fluctuations.

Despite the outstanding levels of sophistication reached by the aforementioned techniques (e.g. considering increasingly complex systems and different types of conditioning data), little attention has been devoted to test their capabilities for reproducing tailing of breakthrough curves and the role of small scale variability of hydraulic conductivity. The objective of this paper is to present a step in this direction. Specifically, we aim at evaluating whether the presence of high frequency fluctuations impedes the characterization of large scale variability trends and whether including small scale fluctuations allows us to reproduce tailing in BTCs.

Procedure for representing small scale variability

The procedure is tested using four synthetic cases on a single domain with increasing level of small scale variability. In essence, the procedure follows the steps of Meier et al. (2001) and Hendricks Franssen (2001):

(1) Definition of the problem domain.
(2) Generation of four ''true'' $\log_{10}K$ (Y hereinafter) fields with increasing small scale variability.
(3) Generation of the available drawdown data.
(4) Estimation and simulations of Y field conditioned to Y and drawdown data using the regularized pilot points method (Appendix A).
(5) Application of the calibrated Y-fields to transport prediction.

These steps are outlined below.

Definition of the problem domain

The problem domain is a 400 m × 400 m square which is discretized using elements of 10 m × 10 m. This is enlarged to a global domain of 3800 m × 3800 m (Fig. 1) to avoid spurious boundary effects. Hydraulic tests and transport prediction take place in the central portion. Outside, the element size increases as the mesh progresses towards the boundaries. The whole domain is used for flow calibration, while only the central portion is used for transport predictions. All runs were performed using TRANSIN code (Medina and Carrera, 2003), as modified by Alcolea et al. (2006a). This code uses a second order Galerkin method for simulating transport, which requires a Peclet number smaller than 2 for stability.

Generation of the ''true'' conductivity fields

First, we select 10 measurement locations at the central part of the domain and set a value of Y at those points (Y measurements, common for all four tests). Then, we generate four ''true'' Y fields (Fig. 2) by sequential Gaussian simulation (Deutsch and Journel, 1992), conditioned to the 10 error-free measurements and the geostatistical models presented in Table 1. These models are stationary with a variance of 2 [$\log_{10}$ m^2/s]. The spatial variability of Y fields is simulated by spherical variograms. These display linear behavior at the origin, which ensures some small scale variability. However, their slope (and, thus, the importance of small scale variability) is reduced when the range is increased. To control both the total range and relative importance of small scale variability, we defined variograms by superimposing two nested spherical structures of 40 and 200 m range, representing the small and large scales of variability, respectively. The four variograms differ on the contribution of the short range structure, ranging from none (NH) to 75% (HH). This contribution is measured as the ratio of the sill of the short range structure to the global one (Table 1). The procedure is such that measurements present a positive bias with respect to the ''true'' fields. This may reflect practical situations when either wells are drilled in known highly producing areas or when wells drilled in low permeability zones are abandoned.

Generation of the available drawdown data

Drawdown data come from three independent pumping tests in the most productive wells of the central part (pumping rates of 10^{-3} m^3/s at wells B1, B2 and B3 in Fig. 1). ''Real'' steady-state drawdowns were simulated at grid nodes (Fig. 3) using the four ''true'' Y fields and prescribing a zero drawdown as initial condition and at the boundaries. Drawdowns were calculated at the 10 measurement locations displayed in Fig. 1. A Gaussian white noise was added to those data to simulate measurement errors. Different noise levels were analyzed to test the effect of measurement errors in the flow calibrations and transport predictions. Results were not very sensitive to noise level. Finally, standard deviations of 0.25 m for the pumping test at well B1 and 0.15 m for the pumping tests at wells B2 and B3 were assigned as noise level (1% of the maximum drawdown at each test, the largest noise level tested).

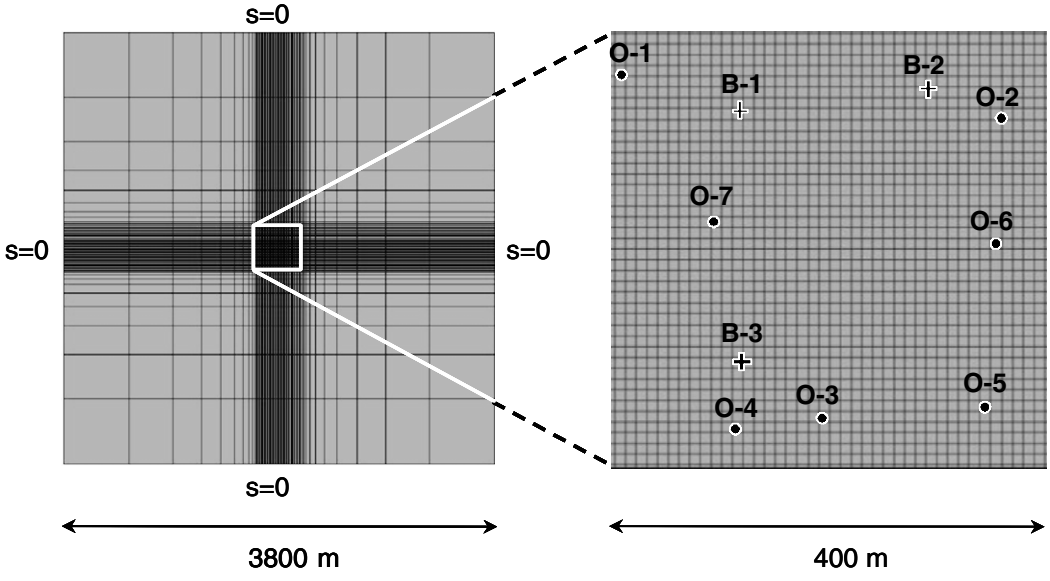

Figure 1 Flow domain and location of conditioning measurements. The inset bounds the zone of interest (model domain for transport prediction).

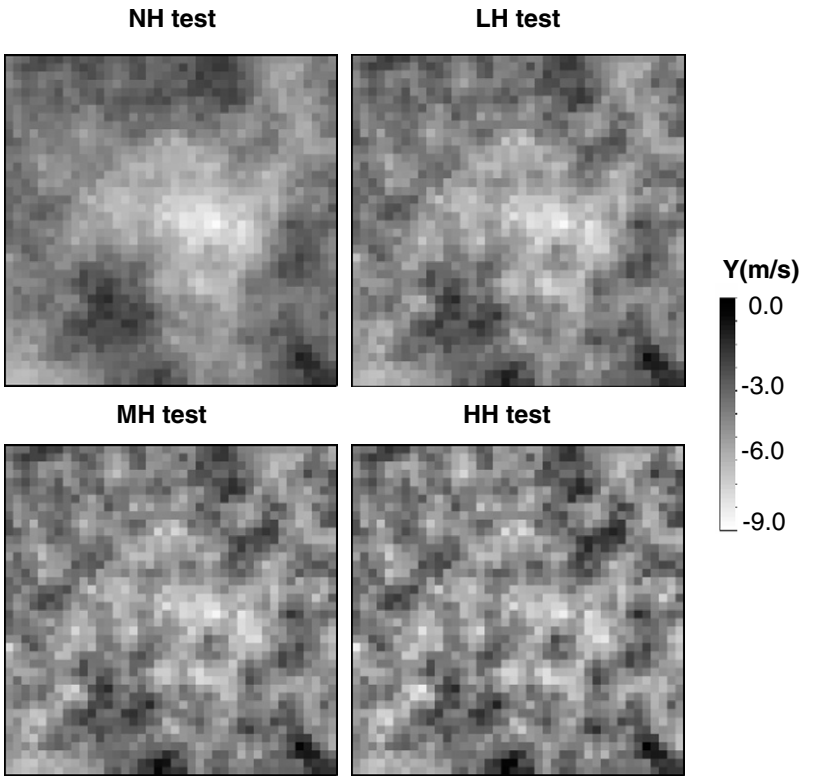

Figure 2 Zone of interest of the ''true'' Y fields. Notice that the large scale trends of NH (low Y in the middle, high Y in the lower right corner, etc.) are reproduced in all realizations. Still, small scale variability, as reflected by the granularity of the fields, increases from NH to HH.

Table 1 Statistical parameters of nested structures defining variograms used for the generation of the four ''true'' log-hydraulic conductivity fields

Synthetic test	Small scale variability	Nested structure defining			
		Small scale heterogeneity		Large scale heterogeneity	
		Range (m)	Sill ($\log_{10}$[m/s]^2)	Range (m)	Sill ($\log_{10}$[m/s]^2)
NH	None	–	–	200	2
LH	Low	40	0.5	200	1.5
MH	Medium	40	1.0	200	1.0
HH	High	40	1.5	200	0.5

Calibration of the Y fields using the regularized pilot points method

Conditional estimation and 20 conditional simulations of the Y field are obtained for each synthetic test (a total of 84 calibrations) by conditioning the model to the available Y and drawdown measurements, as well as their initial drifts (conditioned to the geostatistical model and Y measurements only). The geostatistical models are considered known and error-free (Table 1). The pilot points are arranged on a regular network of 81 points within the zone of interest (spacing of 50 m, equivalent to 4.5 pilot points per correlation range of the long range structure in each direction). Sixteen additional pilot points are located beyond the zone of interest (coarse discretization area in Fig. 1).

Following the methodology of Medina and Carrera (2003), the a posteriori statistical analysis (step 3 in the inversion methodology, see Appendix A) was performed for each realization of the Y field, finding the optimum weights for both conditional estimation and simulations. These weights were always 1.0, both for the term of state variables and for the plausibility term. This finding confirms the results of Alcolea et al. (2006b).

Application of the calibrated Y-fields to transport prediction

The transport problem used for predictions consists of the instantaneous invasion (i.e., a 9900 s injection pulse) of a solute in the zone of interest. To this end, we first simulate a steady state flow field in the finely gridded central domain (Fig. 1) by imposing no flow at the left and right segments and a head gradient of 1% between the lower and upper segments. The solute enters the domain with a concentration

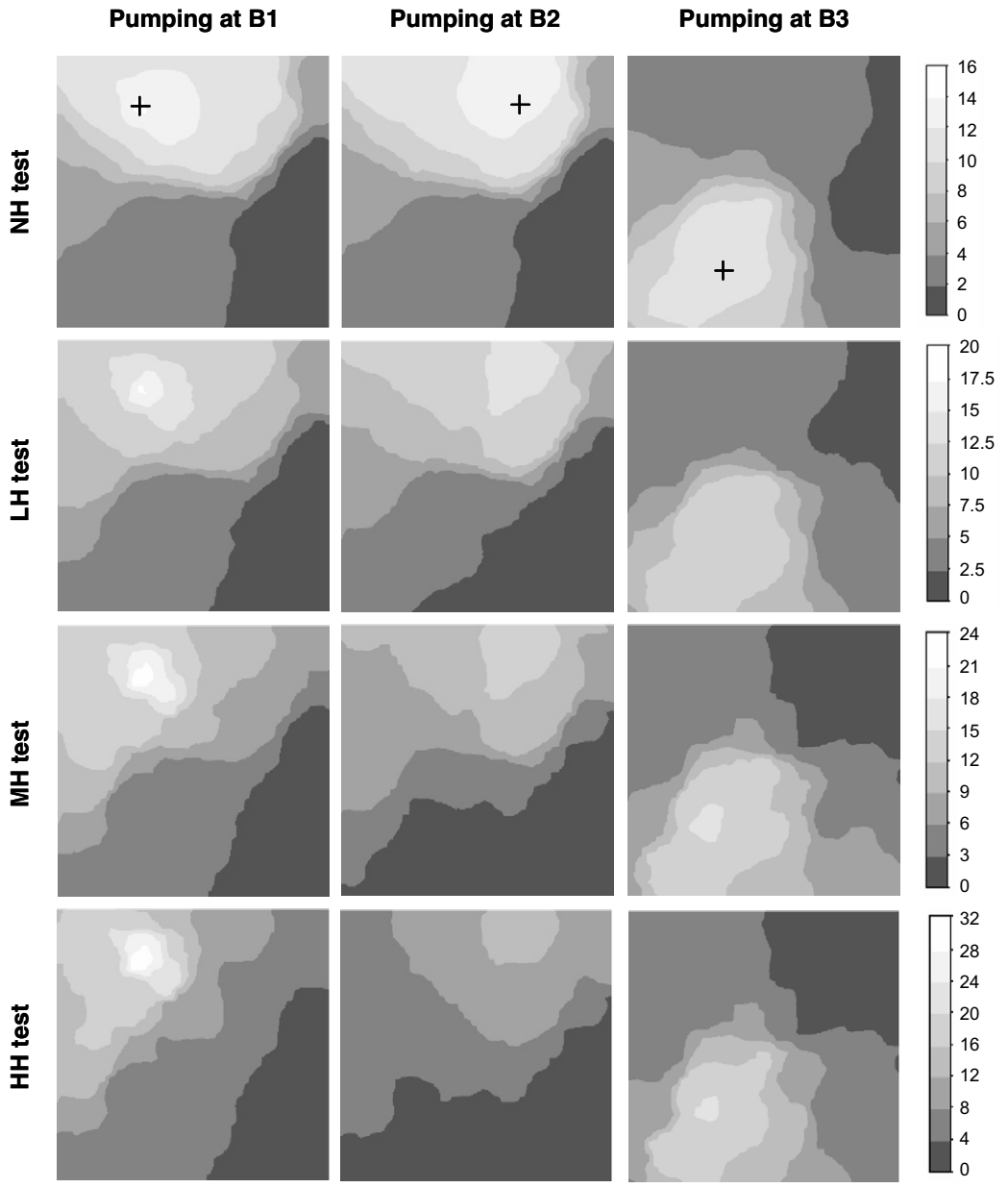

Figure 3 ''True'' steady state drawdowns (in m) at the zone of interest. Crosses at pictures in row 1 depict the location of pumping wells.

of 100 ppm through the lower segment. We choose an injection line instead of point injection to minimize the impact of the hydraulic conductivity value in the vicinity of the point of injection. Values of 8 m and 1 m were chosen for longitudinal and transverse dispersivities, respectively, in all runs. The same values were adopted for conditional simulations and for the true case because both attempt to reproduce the same scales of variability. Therefore, dispersivities reflect the same unmodeled sources of velocity fluctuations (i.e., high variance and extremely small range, smaller than grid size, variability of hydraulic conductivity). We could have adopted larger dispersivities for the conditional estimation cases because these do not attempt to reproduce small scale variability. However, we preferred to keep the same dispersivity values, both for simplicity and consistency in analyzing the role of small scale variability. In all cases, these values of dispersivities minimize the numerical instabilities. Porosity and saturated thickness are also constant, with values of 0.1 and 0.01 m, respectively. An image of the fingering nature of transport in the ''true'' domains is given in Fig. 4, which displays a continuous injection of tracer with a concentration of 1 ppm. Integrated flux averaged BTC at the upper segment will be used for evaluating the transport predictions.

Results

Results are evaluated both qualitatively (Y maps and histograms and concentration fits) and quantitatively. The

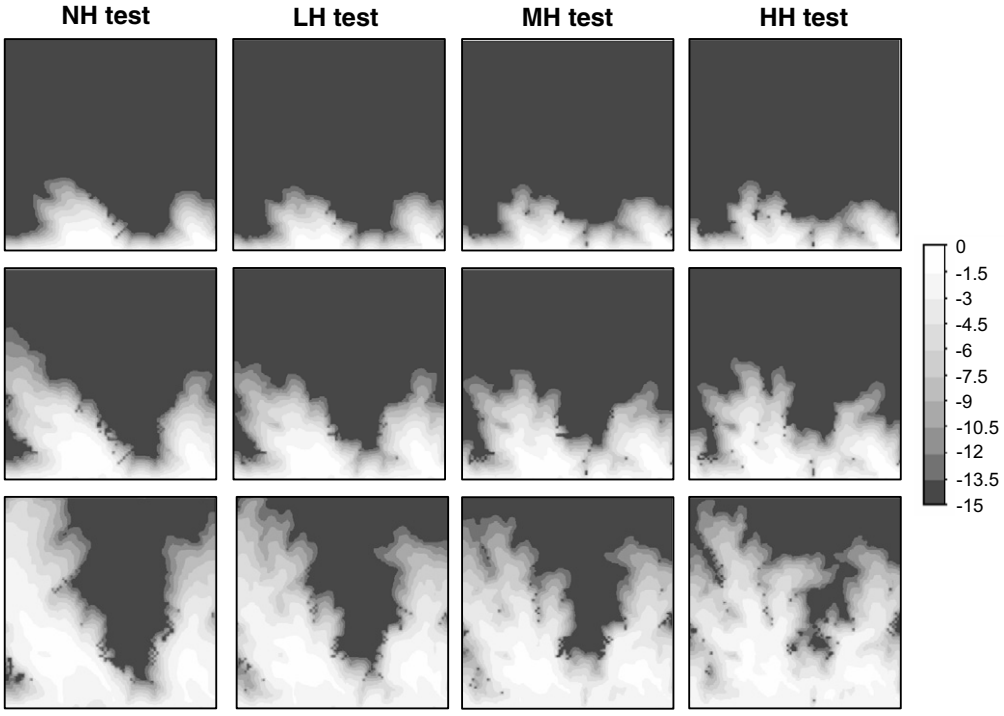

Figure 4 ''True'' $\log_{10}$ concentrations obtained after 9900 s (row 1), 39,000 s (row 2) and 1.4×10^5 s (row 3) of a continuous injection at the lower boundary (input concentration is 1 ppm).

quantitative analysis is performed using an error vector **e**, defined as the difference between calculated and ''true'' values of the Y field at each block of the zone of interest (1600 blocks of 10 m × 10 m). We use the following statistics for analyzing results from the estimation process:

(1) Total objective function (F in Eq. (A1)).
(2) Mean absolute error of the Y field:

$$e_Y = \frac{1}{1600}\sum_{i=1}^{1600}|e_i| = \frac{1}{1600}\sum_{i=1}^{1600}|Y^{true} - Y^{calc}| \quad (1)$$

(3) Root mean square error of the Y field ($RMSE_Y$) and the analogous magnitudes for drawdowns and concentrations ($RMSE_s$ and $RMSE_c$, respectively)

$$RMSE_Y = \left(\frac{1}{1600}\sum_{i=1}^{1600} e_i^2\right)^{1/2} \quad (2)$$

We test the performance of the calibrated Y fields for predicting transport BTCs using the peak concentration and arrival time as well as the slope of the tail. The latter is obtained through regression of late time concentrations. The late time portion of the BTC often displays a ''bumpy'' aspect due to the development of preferential flow paths (Fig. 4). As a result, the definition of this slope is somewhat arbitrary. We chose as ''late time'' the portion with concentrations three orders of magnitude below the peak.

Calibration results are presented first. The Y fields and estimation errors e_Y and $RMSE_Y$ are presented in Fig. 5. The histograms of the Y values and the corresponding statistical moments are depicted in Fig. 6 and Table 2, respectively. The match to drawdown data as measured by $RMSE_s$ is displayed in Fig. 7. With regard to transport prediction, the BTCs obtained with the calibrated Y fields are presented in Fig. 8. The corresponding flow and transport mass balances are depicted in Fig. 9. Peak concentrations, arrival times and the slope characterizing the late time behavior of the BTC are summarized in Table 3.

The first observation that becomes apparent from Fig. 5 is the strong effect of conditioning to drawdowns on estimated/simulated hydraulic conductivities. For any given test and realization (CE or CS), the field obtained by conditioning only to Y measurements is qualitatively worse than the one obtained by conditioning to drawdown and Y data. Therefore, calibration to drawdown data using the optimum weight of the plausibility term reduces estimation errors. The reduction is not dramatic in terms of estimation errors (both e_Y and $RMSE_Y$ are reduced by about 30%). However, conditioning to drawdowns improves the overall look of the computed field (it resembles vaguely the ''true'' field). In fact, estimation or simulations conditioned to hydraulic conductivity data only are of limited interest in practice, as they ignore the valuable information contained in the drawdowns (or any other dependent variable) data set.

The main difference between conditional estimation and simulation stems from the variability, which is best analyzed through the histograms (Fig. 6). Histograms of the conditional estimation to Y measurements, CE(Y), are almost symmetric around the mean value of Y measurements (Fig. 6), as measured by the small skewness of the distribution (Table 2). This effect becomes increasingly apparent with increasing relevance of small scale variability, as measured by the variances (decrease), the skewness (tend to zero) and the kurtosis coefficients (increase) in Table 2. Thus, histograms of CE(Y) (for any given relevance of the

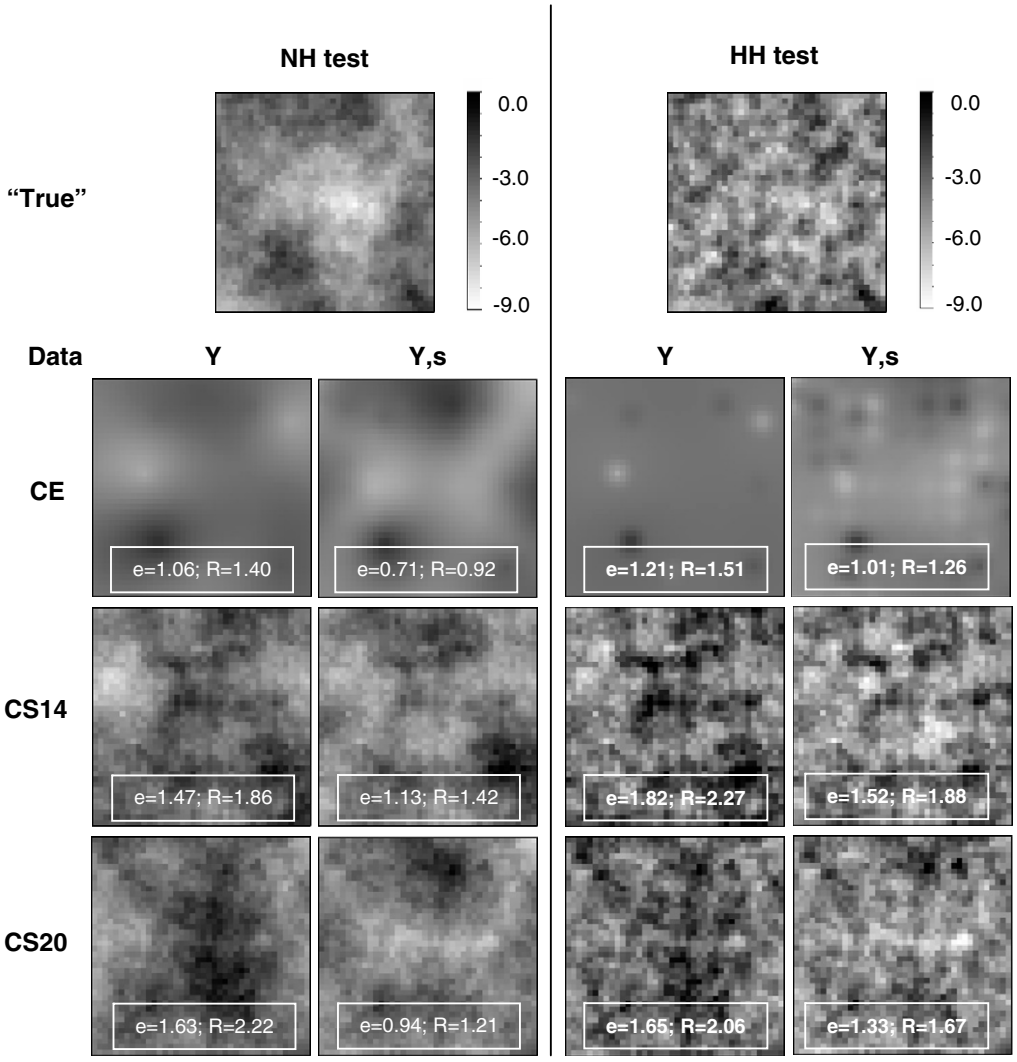

Figure 5 Y fields obtained for the largest relevance of the small scale variability (HH test; subset of pictures at the right of the Figure) and when this variability is neglected (NH test). Two columns are displayed for each subset. The left column displays the Y field conditioned to Y measurements only. The right column displays the Y field conditioned to Y and drawdown measurements. Row 1 contains the ''true'' Y fields. Row 2 contains the estimated fields and rows 3 and 4 contain the results of two out of 20 simulations. In the insets, e and R denote mean absolute error and root mean square error (e_Y and $RMSE_Y$ in Eqs. (1) and (2)).

small scale variability) are centered on the measurements mean and are too sharp (and more so as the degree of small scale variability increases). The same effect, though less notable, can be observed in the fields characterized by conditional estimation to Y and drawdown data, CE(Y, s). However, these outcomes are better as these fields are biased towards the mean value of the corresponding ''true'' field. In summary, conditional estimation to Y measurements only leads to fields that are too homogeneous (narrow histogram) and centered around direct measurements. Adding drawdowns to conditioning data broadens the histogram and displaces it towards the ''true'' mean, but not sufficiently.

Conversely, simulations (regardless of the type of conditioning data) yield more realistic results (Fig. 5), even though their estimation errors are larger than the ones obtained by conditional estimation (as expected). Simulations conditioned only to Y data, CS(Y), overestimate hydraulic conductivities (Fig. 6) because measurements present a positive bias. Fortunately, conditioning to drawdown data helps to alleviate this problem and the ensembles of simulations conditioned to Y and drawdown data, CS(Y, s), resemble (for any given relevance of the small scale variability) the histogram of the ''true'' fields (Fig. 6) and the first two moments of the corresponding distributions (Table 2).

Analyzing the role of small scale variability on calibration is complex. We use drawdown fits ($RMSE_s$ in Fig. 7) to analyze the improvement caused by conditioning (similar conclusions can be reached by analyzing the total objective function F, not shown here). Best results are most often obtained when the relevance of small scale variability is negligible (NH test). This case yields the smallest values of $RMSE_s$ (and F) in most cases, because the ''true'' field is smoothest (Fig. 2) and, thus, easiest to be characterized. However, results do not degrade monotonically with increasing relevance of small scale variability. In any case, $RMSE_s$ are very small for all realizations (even smaller than

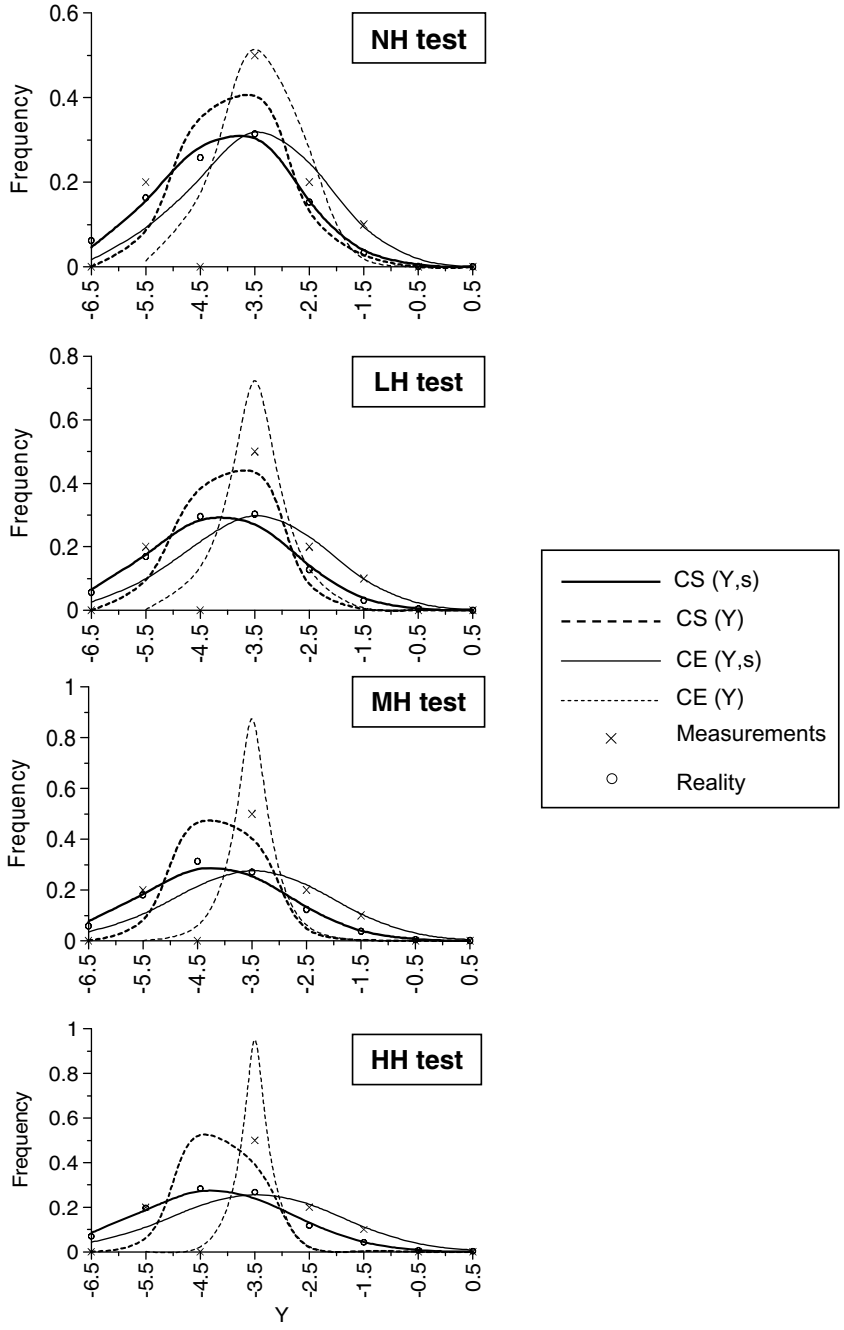

Figure 6 Histogram of Y fields. Results of conditional simulation are the ensemble of 20 realizations. In the legend, CE and CS denote conditional estimation and conditional simulation, respectively; Y and s denote the type of conditioning data (hydraulic conductivity and drawdown, respectively).

the standard deviation of measurement errors, which suggests a certain overparameterization) and display little dependence on the level of small scale variability. In summary, small scale variability does not control the behavior of steady-state heads, which is ruled by the large scale heterogeneity patterns. This implies that (steady-state) hydraulic data will not suffice for characterizing high frequency fluctuations.

Small scale variability becomes important for modeling contaminant transport. The late time behavior of the BTCs depends to a large extent on small scale variability (Carrera, 1993; Berkowitz et al., 2006). Regardless of the type and number of conditioning data, conditional estimation, which yields smooth Y fields, does not match the slope of the late-time portion of BTCs (Fig. 8). The error in calculated slope increases as the relevance of the small scale variability becomes more important (Table 3). Conversely, most conditional simulations reproduce this slope, even when only Y measurements are used for conditioning. In this case, breakthroughs are too fast and peak concentrations too high, reflecting the bias of Y measurements. Moreover, simulated BTCs span a very broad range. Yet, all simulations display a tail similar to the ''true'' BTC.

Table 2 Statistical moments of the distribution of measurements, ''true'' and estimated/simulated fields (in bold, values of mean and variance which are closest to the ''true'' ones)

Measurements		Mean	Variance	Skewness	Kurtosis
		−3.94	1.50	0.04	0.69
NH test	''True'' field	−4.60	1.57	−0.32	−0.14
	CE (Y)	−3.90	0.47	−0.36	0.10
	CE (Y, s)	−4.33	0.73	0.29	−0.31
	CS (Y)	−3.95	1.50	−0.05	−0.04
	CS (Y, s)	**−4.53**	**1.52**	−0.16	0.16
LH test	''True'' field	−4.65	1.47	−0.11	−0.01
	CE (Y)	−3.93	0.25	−0.28	1.06
	CE (Y, s)	−4.48	0.53	0.18	−0.32
	CS (Y)	−3.99	**1.73**	−0.04	0.01
	CS (Y, s)	**−4.68**	1.75	−0.11	0.14
MH test	''True'' field	−4.68	1.55	0.05	−0.04
	CE (Y)	−3.94	0.14	−0.14	3.78
	CE (Y, s)	−4.58	0.42	0.15	−0.10
	CS (Y)	−4.01	2.02	−0.04	0.03
	CS (Y, s)	**−4.74**	**1.88**	−0.10	0.07
HH test	''True'' field	−4.71	1.71	0.12	−0.08
	CE (Y)	−3.95	0.07	−0.02	12.87
	CE (Y, s)	−4.63	0.33	0.23	0.32
	CS (Y)	−4.01	2.33	−0.05	0.04
	CS (Y, s)	**−4.79**	**2.07**	−0.07	0.04

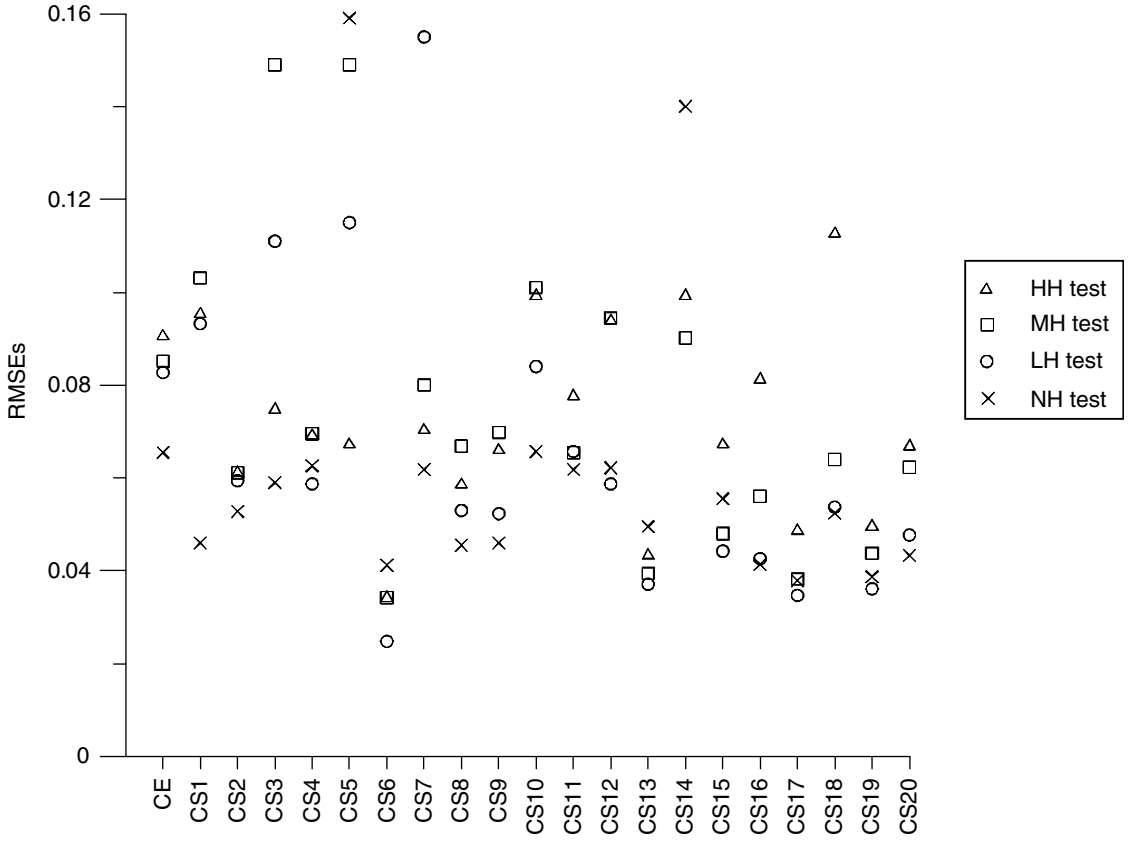

Figure 7 Root mean square error of drawdowns ($RMSE_s$) for all calibrations, conditioned to hydraulic conductivity and drawdown data.

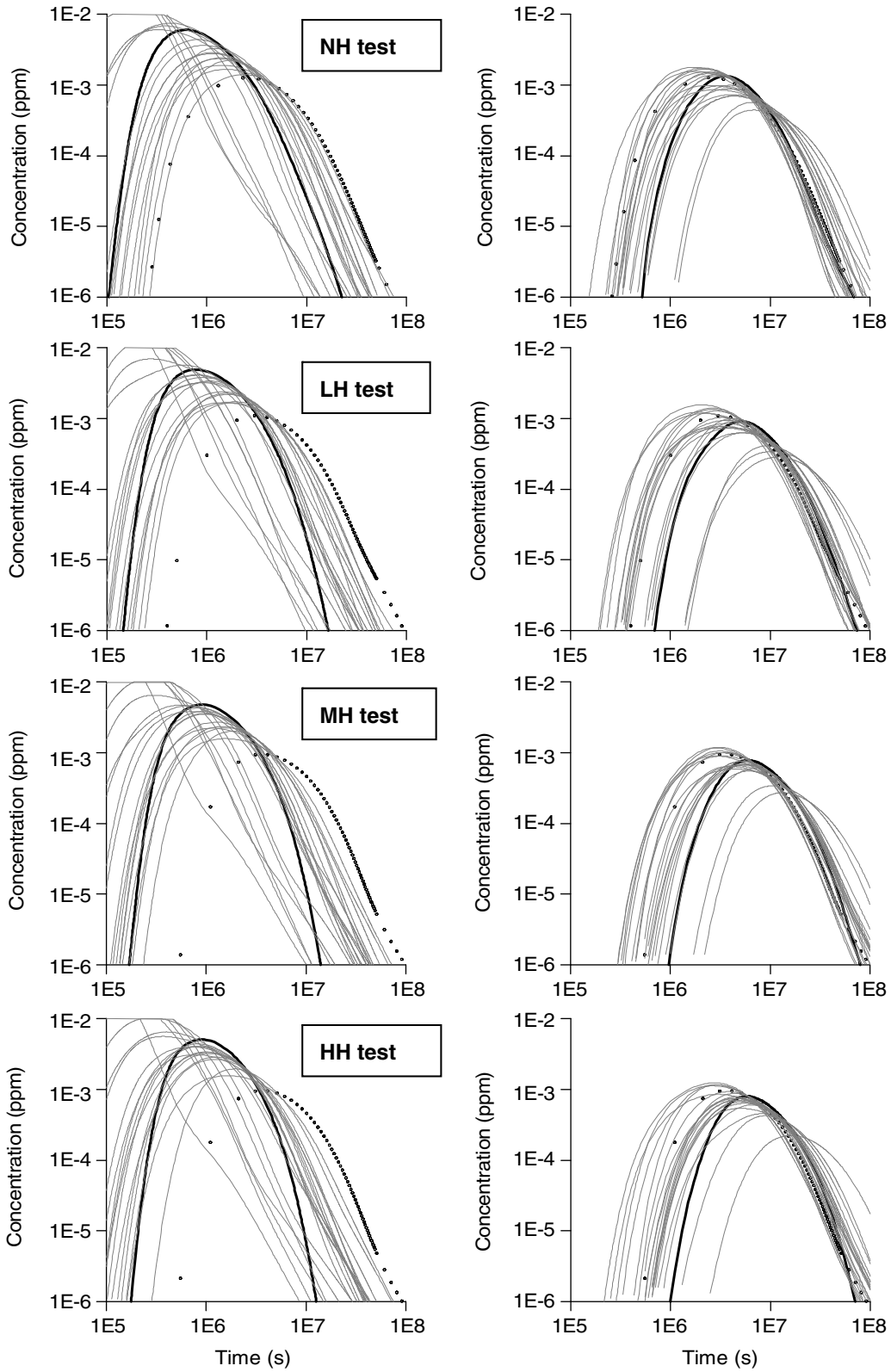

Figure 8 Predicted BTCs for a pulse injection using the Y fields conditioned to Y measurements only (left column) and to Y and drawdown measurements (right column). ''True'' BTCs are depicted with dots, the BTC corresponding to the conditional estimation with thick line and the 20 BTCs corresponding to conditional simulation with thin grey line.

The impact of using also drawdowns for conditioning is manifested again in Fig. 8 and Table 3. First, errors in peak time, peak concentration, final slope and $RMSE_c$ decrease as expected (both for conditional estimation and simulation). Second, the uncertainty of the predicted BTCs using conditional simulations is reduced substantially, as measured by

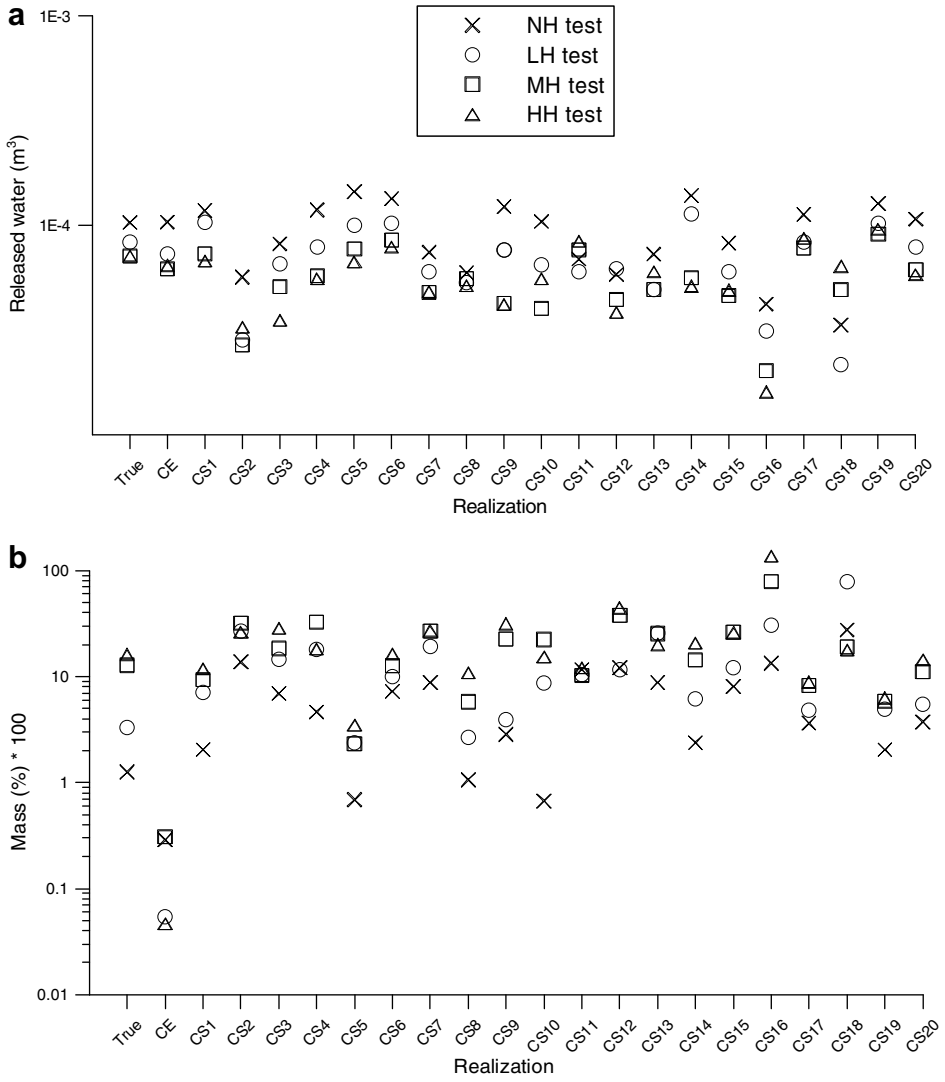

Figure 9 Mass balance of transport predictions. (a) Released volume of water through the upper boundary of the zone of interest and (b) tracer mass remaining in the aquifer at the end of the simulation.

Table 3 Peak time, peak concentration and late-time slope of the BTCs for the different realizations (mean value and variance for conditional simulations)

		Conditioned to Y				Conditioned to Y, drawdown			
		$\log_{10}$ Peak time	$\log_{10}$ Peak conc.	Slope	$RMSE_c$	$\log_{10}$ Peak time	$\log_{10}$ Peak conc.	Slope	$RMSE_c$
NH test	''True''	6.40	−2.90	−2.96	–	6.40	−2.90	−2.96	–
	CE	5.78	−2.22	−7.85	$9.34 \cdot 10^{-4}$	6.54	−2.88	−4.16	$8.16 \cdot 10^{-5}$
	CS (mean)	5.75	−2.26	−3.27	$3.55 \cdot 10^{-3}$	6.49	−2.96	−2.91	$8.88 \cdot 10^{-5}$
LH test	''True''	6.48	−2.96	−3.24	–	6.48	−2.96	−3.24	–
	CE	5.90	−2.31	−16.39	$6.73 \cdot 10^{-4}$	6.71	−3.04	−5.53	$9.21 \cdot 10^{-5}$
	CS (mean)	5.77	−2.28	−3.54	$3.35 \cdot 10^{-3}$	6.59	−3.08	−2.97	$8.64 \cdot 10^{-5}$
MH test	''True''	6.56	−3.02	−2.44	–	6.56	−3.02	−2.44	–
	CE	5.95	−2.32	−27.78	$6.29 \cdot 10^{-4}$	6.77	−3.11	−4.62	$9.26 \cdot 10^{-5}$
	CS (mean)	5.75	−2.27	−3.80	$3.40 \cdot 10^{-3}$	6.68	−3.14	−2.76	$7.49 \cdot 10^{-5}$
HH test	''True''	6.56	−3.02	−2.29	–	6.56	−3.02	−2.29	–
	CE	5.95	−2.29	−39.59	$6.51 \cdot 10^{-4}$	6.76	−3.10	−5.27	$9.24 \cdot 10^{-5}$
	CS (mean)	5.73	−2.25	−4.23	$3.55 \cdot 10^{-3}$	6.68	−3.14	−2.74	$7.90 \cdot 10^{-5}$

the statistics listed in Table 3. Qualitatively, one can observe how the ''true'' BTCs are bounded by the set of predicted BTCs obtained with the simulated fields, for any given relevance of the small scale variability. This is corroborated by the reduction of errors in late time slope ($RMSE_c$ in Table 3). In short, bias effects introduced by conditioning only to Y measurements are removed by conditioning also to drawdowns. It is also important to notice that such conditioning reduces uncertainty (as measured by the width of the envelope of simulated BTCs) in all cases. Reduction is most significant for the NH case (i.e., no short range variability), which reflects again that drawdowns contain little information about small scale variability patterns.

Flow mass balance is somewhat independent of the relevance of the small scale variability (Fig. 9a) and both conditional estimation and conditional simulation methods yield flow mass balances close to the ''true'' ones (in fact, CE results are better than many CS's). This result confirms that flow behavior is controlled by the large scale patterns of heterogeneity. Conversely, transport mass balance depends to a large extent on, first, the type of conditioning (estimation versus simulation), and second, on the relevance of small scale variability. All conditional simulations reproduce the ''true'' mass balance better than conditional estimation. The smoothness of conditional estimation Y fields favors that most solute mass has been washed away by the end of the simulation. As variability increases, increasing portions of mass remain retained in low permeability areas. This left-over mass is what caused tailing in the first place. It is also what explains the large differences in solute mass balances in Fig. 9b. In fact, reproducing trailing solute mass allows proper separation between mixing and spreading, which was one of our original motivations.

Conclusions

This work was motivated by the concept that one may be able to identify rather accurately the large scale (low frequency) trends of spatial variability, but not the high frequency components. Yet, these are relevant for properly simulating and understanding solute transport through heterogeneous media (or any other process that depends non-linearly on K, for that matter). Accepting that small scale variability cannot be identified, we follow on the steps of Gómez-Hernández et al. (1997), Hendricks Franssen (2001) and RamaRao et al. (1995). That is, we first simulate fields conditioned to all available direct measurements and conceptual constraints. Here, direct measurements were exact point measurements of $\log_{10}K$ and the only conceptual constraint was the assumption that the ''true'' field was a stationary random field with two nested variograms. The resulting random drift is then perturbed so as to ensure that observations (here, drawdowns) are well fitted by the model, using the regularized pilot points method. The question is whether this approach does indeed allow accurate transport simulations. The application leads to the following conclusions:

(1) Adding a component of small scale variability (i.e., simulating $\log_{10}K$ with two nested variograms) leads to increased tailing in transport simulations. The tail slope is much larger than that observed in the ''true'' breakthrough curves (receeding limb too steep). Yet, our results suggest that the slope may be decreased by adding more nested variograms. This is in agreement with the universal scaling theory of Neuman.
(2) Simulated fields reproduce the statistics of the ''true'' field. This confirms the results of Gómez-Hernández et al. (1997), Hendricks Franssen (2001), and RamaRao et al. (1995).
(3) Simulated fields reproduce the main features of ''true'' BTCs (arrival time, peak concentration and tail slope). This confirms the main conjecture that motivated this work, namely that one does not need to identify small scale variability, but to simulate its presence. This implies that reliable transport predictions require a good stochastic description, but not precise knowledge, of small scale variability.
(4) When small scale variability is ignored (as in the case of conditional estimation), simulated BTCs reproduce arrival time and peak concentration, but not the tail. This is not critical when the small scale component (NH test) is negligible. In such cases, optimal (smooth) estimation of hydraulic conductivity yields good results. However, in view of the ubiquity of tailing, we feel that this will be rarely, if ever, the case in practice. Thus, the use of conditional estimation, which neglects the small scale variability (with the exception of moment equations), is not recommended if one seeks meaningful transport predictions.
(5) When conditioning to drawdowns (in fact, to dependent variables) is not performed, simulated breakthrough curves reproduce the shape of ''true'' BTCs (in $\log t - \log c$ scale), but they may be biased if direct measurements are biased, which we fear is often the case.

Much remains to be done. Here we assumed that the structure of variability is known (known variograms, Kerrou et al., 2008). Moreover, the adopted structure is relatively simple (stationary random field defined by only two nested variograms). Dealing with more complex structures will be needed and will require overcoming several difficulties. In this context, the results presented here should be viewed as a hopeful step in the direction of simulating transport through heterogeneous media in a realistic manner.

Acknowledgements

This work was funded by ENRESA (Spanish Agency for Nuclear Waste Disposal), MEC (Spanish Ministry for Education and Science), the European Commission through project FUNMIG and the SNSF (Swiss National Science Foundation) through contract PP002-106557. The authors gratefully acknowledge X. Sanchez-Vila and M. Dentz (UPC) for their help and two anonymous reviewers for their valuable comments.

Appendix A. Inversion methodology: the regularized pilot points method

The inversion technique used in this work is a modification of the pilot points method (de Marsily et al., 1984; Lavenue

and de Marsily, 2001), to include a plausibility term. Algorithmic details of this methodology can be found in previous works (Alcolea et al., 2006a,b). The procedure can be summarized as follows:

(1) *Parameterization*. The unknown hydraulic property (typically $\log_{10}K$) is expressed as the superposition of two fields: a drift and an uncertain residual. The latter is a linear combination of the model parameters (value of the property at pilot points locations). The drift can be calculated by conditional estimation or conditional simulation, in which case the drift is a random function. In both cases, all available information (direct measurements, geophysics, geological data, etc.) can be used for conditioning.

(2) *Optimization of model parameters*. The optimum set of model parameters minimizes an objective function F that quantifies the misfit between calculated and measured data:

$$F(\mathbf{p}) = \sum_{i=1}^{nstat} \beta_i (\mathbf{s}_i - \mathbf{s}_i^*)^t \mathbf{V}_{s_i}^{-1} (\mathbf{s}_i - \mathbf{s}_i^*) + \sum_{j=1}^{ntypar} \mu_j (\mathbf{p}_j - \mathbf{p}_j^*)^t \mathbf{V}_{p_j}^{-1} (\mathbf{p}_j - \mathbf{p}_j^*)$$

where the first term measures the misfits between calculated ($\mathbf{s}_i$) and measured ($\mathbf{s}_i^*$) 'nstat' types of state variables and the second is a plausibility term which measures the departure of the model parameters $\mathbf{p}_j$ from their prior information $\mathbf{p}_j^*$ ('ntypar' denotes the number of types of parameters). $\mathbf{V}_s$ and $\mathbf{V}_p$ (block matrices containing $\mathbf{V}_{s_i}$ and $\mathbf{V}_{p_j}$, respectively) are the best guess of the corresponding covariance matrices and β_i and μ_j are weighting scalars correcting the specification of $\mathbf{V}_{s_i}$ and $\mathbf{V}_{p_j}$, respectively. Prior estimates of model parameters can be calculated by conditional estimation or simulation to available measurements. For the case of conditional estimation, $\mathbf{V}_p$ is the kriging error covariance matrix ($\mathbf{V}_k$). $\mathbf{V}_p$ is corrected if conditional simulation is performed ($\mathbf{V}_p = 2\mathbf{V}_k$; see Appendix B).

(3) *Finding the optimum weighting scalars β_i and μ_j (a posteriori statistical analysis)*. The optimization process is repeated using different values of the weighting scalars. Assigning low weights (μ_j) to model parameters disregards their prior information, but leads to the best fit of the measured state variables. Conversely, assigning large weights to model parameters disregards the measured state variables, biasing the solution towards prior information. The optimal values are the ones leading to the maximum expected likelihood of the parameters given the data (Medina and Carrera, 2003).

Appendix B. A priori covariance matrix of parameters

The formulation of the regularized pilot points method requires the specification of the a priori covariance matrix of parameters. If kriging is used for defining the drift (step 1 of the inversion methodology), this matrix is the kriging covariance matrix. If simulation is performed, the covariance matrix is calculated as follows.

Let $\mathbf{Y}$ be the vector of true values of the field at the n_e points/blocks to be estimated. Let $\mathbf{Y}^K$ and $\mathbf{Y}^{CS}$ be the vectors of kriged and simulated values, respectively. Let $\mathbf{V}_K$ and $\mathbf{V}_{CS}$ be the kriging and conditional simulation error covariance matrices, respectively. $\mathbf{M}$ is a matrix such that $\mathbf{V}_K = \mathbf{M} \cdot \mathbf{M}^t$ and $\mathbf{u}$ is a vector of independent Gaussian variables u_i with zero mean and unit variance.

Conditional simulation can be expressed as

$$\mathbf{Y}_{CS} = \mathbf{Y}^K + \mathbf{M} \cdot \mathbf{u} \tag{A2}$$

The $m-n$th component of the covariance matrix of CS errors is

$$(\mathbf{V}_{CS})_{m,n} = \mathrm{cov}[Y_m^{CS} - Y_m, Y_n^{CS} - Y_n] \tag{A3}$$

Substituting Eq. (A2) in (A3) and using the definition of covariance:

$$\begin{aligned}(\mathbf{V}_{CS})_{m,n} &= E\left[\left(Y_m^K - Y_m + \sum_{j=1}^{ne} M_{mj}u_j\right)\left(Y_n^K - Y_n + \sum_{k=1}^{ne} M_{nk}u_k\right)\right] \\ &= E[(Y_m^K - Y_m)(Y_n^K - Y_n)] + \sum_{j=1}^{ne}\sum_{K=1}^{ne} M_{mj}M_{nk}E[u_j u_K] \\ &\quad + \sum_{j=1}^{ne} M_{mj}E[u_j(Y_n^K - Y_n)] + \sum_{K=1}^{ne} M_{nk}E[u_k(Y_m^K - Y_m)]\end{aligned} \tag{A4}$$

The last two terms are zero because $\mathbf{u}$ is independent of $\mathbf{Y}$. The first term in the right hand side is the definition of kriging error covariance matrix. Finally, the second term equals $\mathbf{M} \cdot \mathbf{M}^t = \mathbf{V}_K$ because the components of $\mathbf{u}$ are independent with unit variance. Therefore,

$$\mathbf{V}_{CS} = 2\mathbf{V}_K \tag{A5}$$

References

Adams, E.E., Gelhar, L.W., 1992. Field study of dispersion in a heterogeneous aquifer. 2. Spatial moment analysis. Water Resour. Res. 28 (12), 3293–3308.

Alcolea, A., Carrera, J., Medina, A., 2006a. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Adv. Water Res. 29 (11), 1678–1689.

Alcolea, A., Carrera, J., Medina, A., 2006b. Inversion of heterogeneous parabolic-type equations using the pilot points method. Int. J. Numer. Meth. Fluids 51 (9–10), 963–980.

Berkowitz, B., Cortis, A., Dentz, M., Scher, H., 2006. Modeling non-fickian transport in geological formations as a continuous time random walk. Rev. Geophys. 44, 1–49.

Capilla, J.E., Gómez-Hernández, J.J., Sahuquillo, A., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 2. Demonstration on a synthetic aquifer. J. Hydrol. (203), 175–188.

Carrera, J., 1993. An overview of uncertainties in modelling groundwater solute transport. J. Contam. Hydrol. 13, 23–48.

Carrera, J., Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady-state conditions. 1. Maximum likelihood method incorporating prior information. Water Resour. Res. 22 (2), 199–210.

Carrera, J., Sanchez-Vila, X., Benet, I., Medina, A., Galarza, G., Guimera, J., 1998. On matrix diffusion: formulations, solution methods and qualitative effects. Hydrogeol. J. 6, 178–190.

Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., Slooten, L.J., 2005. Inverse problem in hydrogeology. Hydrogeol. J. 13, 206–222.

Cortis, A., Berkowitz, B., 2004. Anomalous transport in ''classical'' soil and sand columns. Soil Sci. Soc. Am. J. 68, 1539–1548.

Dagan, G., 1985. Stochastic modelling of groundwater flow by unconditional and conditional probabilities: the inverse problem. Water Resour. Res. 21 (1), 65–72.

Dagan, G., 1986. Statistical theory of groundwater flow ant transport: pore to laboratory, laboratory to formation and formation to regional scale. Water Resour. Res. 22 (9), 120S–134S.

de Marsily, G.H., Lavedan, G., Boucher, M., Fasanino, G., 1984. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In: Verly et al. (Eds.), Geostatistics for Natural Resources Characterization. Part 2. D. Reidel Pub. Co., pp. 831–849.

De Simoni, M., Sanchez-Vila, X., Carrera, J., Saaltink, M.W., 2007. A mixing ratios-based formulation for multicomponent reactive transport. Water Resour. Res. 43, W07419. doi:10.1029/2006WR00525.

Dentz, M., Carrera, J., 2007. Mixing and spreading in stratified flow. Phys. Fluids 19 (1). Art. 01710.

Deutsch, C.V., Journel, A.G., 1992. GSLIB, Geostatistical Software Library and User's Guide. Oxford University Press, New York.

Fernandez-Garcia, D., Rajaram, H., Illangasekare, T.H., 2005. Assessment of the predictive capabilities of stochastic theories in a three-dimensional laboratory test aquifer: effective hydraulic conductivity and temporal moments of breakthrough curves. Water Resour. Res. 41 (4). doi:10.1029/2004WR00352.

Gelhar, L.W., Mantoglou, A., Welty, C., Rehfeldt, K.R., 1985. A review of field-scale physical solute transport processes in saturated and unsaturated porous media. Rep. EPRI-EA-4190, Proj. 2485-5, Electr. Power. Res. Inst., Palo Alto, California.

Gelhar, L.W., Welty, C., Rehfeldt, K.R., 1992. A critical review of data on field-scale dispersion in aquifers. Water Resour. Res. 28 (7), 1955–1974.

Gómez-Hernández, J.J., Sahuquillo, A., Capilla, J.E., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. Part 1. Theory. J. Hydrol. 204 (1–4), 162–174.

Guadagnini, A., Neuman, S.P., 1999. Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains. 1. Theory and computational approach. Water Resour. Res. 35 (10), 2999–3018.

Guadagnini, A., Neuman, S.P., 2001. Recursive conditional moment equations for advective transport in randomly heterogeneous velocity fields. Trans. Porous Media 42 (1–2), 37–67.

Guadagnini, A., Riva, M., Neuman, S.P., 2003. Three-dimensional steady state flow to a well in a randomly heterogeneous bounded aquifer. Water Resour. Res. 39 (3), 1048. doi:10.1029/2002WR00144.

Haggerty, R., Gorelick, S.M., 1995. Multi-rate mass transfer for modelling diffusion and surface reactions in media with pore-scale heterogeneity. Water Resour. Res. 31 (10), 2383–2400.

Haggerty, R., McKenna, S.A., Meigs, L.C., 2000. On the late-time behavior of tracer test breakthrough curves. Water Resour. Res. 36 (12), 3467–3479.

Hendricks Franssen, H.J., 2001. Inverse stochastic modeling of groundwater flow and mass transport. Ph.D. Thesis, Technical University of Valencia, Spain.

Hendricks Franssen, H.J., Gómez-Hernández, J.J., Sahuquillo, A., 2003. Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data. J. Hydrol. 281 (4), 281–295.

Hernandez, A.F., Neuman, S.P., Guadagnini, A., Carrera, J., 2003. Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion. Stochast. Environ. Res. Risk. Assess. 17 (5), 329–338.

Hernandez, A.F., Neuman, S.P., Guadagnini, A., Carrera, J., 2006. Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media. Water Resour. Res. 42 (5). doi:10.1029/2005WR00444.

Kennedy, C.A., Lennox, W.C., 2001. A stochastic interpretation of the tailing effect in solute transport. Stochast. Environ. Res. Risk. Assess. 15, 325–340.

Kerrou, J., Renard, Ph., Hendricks Franssen, H.J., Lunati, I., 2008. Issues in characterizing heterogeneity and connectivity in non-multiGaussian media. Adv. Water Resour. 31, 147–159.

Kitanidis, P.K., Vomvoris, E.G., 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one dimensional simulations. Water Resour. Res. 19 (3), 677–690.

Kosakowski, G., 2004. Anomalous transport of colloids and solutes in a shear zone. J. Contam. Hydrol. 72 (1–4), 23–46.

Lallemand-Barres, P., Peaudecerf, P., 1978. Recherche des rélations entre la valeur de la dispersivité macroscopique d'un milieu aquifère, ses autres caracteristiques et les conditions de mesure. Étude bibliographique. Hydrogéol. Géol. Ingénieur, 277–284.

Lavenue, A.M., Pickens, J.F., 1992. Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. Water Resour. Res. 28 (6), 1543–1569.

Lavenue, M., de Marsily, G., 2001. Three-dimensional interference test interpretation in a fractured aquifer using the pilot-point inverse method. Water Resour. Res. 37 (11), 2659–2675.

Levy, M., Berkowitz, B., 2003. Measurement and analysis of non-Fickian dispersion of solutes in porous media. J. Contam. Hydrol. 64 (3–4), 203–226.

Medina, A., Carrera, J., 2003. Geostatistical inversion of coupled problems: dealing with computational burden and different types of data. J. Hydrol. (281), 251–264.

Meier, P., Medina, A., Carrera, J., 2001. Geostatistical inversion of cross-hole pumping tests for identifying preferential flow channels within a shear zone. Ground Water 39 (1), 10–17.

Neuman, S.P., 1990. Universal scaling of hydraulic conductivities and dispersivities in geological media. Water Resour. Res. 22 (8), 1479–1758.

Neuman, S.P., 2005. On the tensorial nature of advective porosity. Adv. Water Resour. 28 (2), 149–159.

RamaRao, B.S., Lavenue, M., de Marsily, G.H., Marietta, M.G., 1995. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 1. Theory and computational experiments. Water Resour. Res. 31 (3), 475–493.

Rasmuson, A., 1985. The influence of particle shape on the dynamics of fixed beds. Chem. Eng. Sci. 40, 1115–1122.

Rubin, Y., Bellin, A., Lawrence, A.E., 2003. On the use of block-effective macrodispersion for numerical simulations of transport in heterogeneous formations. Water Resour. Res. 39 (9), 1242. doi:10.1029/2002WR00172.

Sahuquillo, A., Capilla, J., Gómez-Hernández, J.J., Andreu, J., 1992. Conditional simulation of transmissivity fields honouring piezometric data. In: Blain, W.R., Cabrera, E. (Eds.), Fluid Flow Modelling. Comput. Mech., Billerica, MA, pp. 201–212.

Sanchez-Vila, X., Carrera, J., 1997. Directional effects on convergent flow tracer tests. Math. Geol. 29 (4), 551–569.

Silliman, S.E., Simpson, E.S., 1987. Laboratory evidence of the scale effect in dispersion of solutes in porous media. Water Resour. Res. 23 (8), 1667–1673.

Vallocchi, A.J., 1985. Validity of the local equilibrium assumption for modeling sorbing solute transport through homogeneous soil. Water Resour. Res. 21 (6), 808–820.

Vesselinov, V.V., Neuman, S.P., Illman, W.A., 2001. Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff. 2. Equivalent parameters, high-resolution stochastic imaging and scale effects. Water Resour. Res. 37 (12), 3019–3041.