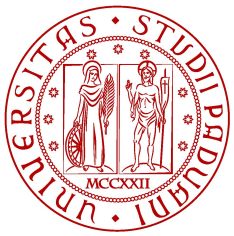


UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Universitá degli Studi di Padova

Dipartimento di Ingegneria Civile, Edile e Ambientale (ICEA)

SCUOLA DI DOTTORATO DI RICERCA IN
SCIENZE DELL'INGEGNERIA CIVILE ED AMBIENTALE
XXV CICLO

Tracer Test Data Assimilation for the Assessment of Local Hydraulic Properties in Heterogeneous Aquifers

Direttore della Scuola: Prof. Stefano Lanzoni

Supervisori: Prof. Paolo Salandin

Ing. Matteo Camporese

Dottoranda: Elena Crestani

Abstract

The hydraulic conductivity distribution in a natural porous media is characterized by a great heterogeneity that makes its spatial assessment problematic and expensive. At the same time, a detailed knowledge of the hydraulic properties, as porosity, storativity, transmissivity and hydraulic conductivity K, is fundamental for the prediction of groundwater flow and solute transport in natural formations. Among the hydraulic properties, being the subsurface transport phenomena in natural formations mainly controlled by the Darcy's law, the proper definition of the K spatial distribution at different scales plays a fundamental role to evaluate the evolution of a contaminant plume, to define the well-catchment areas or to monitor a landfill site. To estimate aquifer hydraulic properties, inverse models have long been studied and, beyond the traditional hydraulic conductivity and head measurements, tracer test analyses have been widely adopted in the past and their use have increased in the recent years thanks to a great improvement of geophysical techniques. Among others, the Electrical Resistivity Tomography (ERT) allows to monitor a tracer test injection, providing time-lapse informations about the plume evolution with limited cost.

Assuming that time-lapse spatially distributed data deduced from a tracer test are available, the present work investigates different approaches aimed to the estimation of the local K distribution. At this purpose, Kalman filter based data assimilation techniques are coupled with the Lagrangian transport model and applied in different synthetic contexts.

Sommario

La distribuzione di conducibilitá idraulica in un mezzo poroso naturale é caratterizzata da grande eterogeneitá, che rende la sua determinazione problematica e costosa. Allo stesso tempo, una approfondita conoscenza delle proprietá idrauliche, quali la porositá, l'immagazzinamento specifico e la conducibilitá idraulica K, é di fondamentale importanza per poter predire e analizzare il flusso sotterraneo e il trasporto di soluti in formazioni naturali. Poiché i fenomeni di trasporto sotterraneo che si realizzano negli acquiferi sono principalmente controllati dalla legge di Darcy, tra le diverse proprietá idrauliche sopraccitate, un'opportuna definizione della distribuzione spaziale di K gioca un ruolo fondamentale nella predizione del plume di inquinanti, e quindi assume particolare rilevanza in molte attivitá di pratico interesse, quali la definizione delle aree di salvaguardia dei pozzi o il monitoraggio di discariche. Le proprietá idrauliche degli acquiferi sono di norma stimate con l'ausilio di modelli inversi utilizzando, oltre le tradizionali misure di conducibilitá idraulica e piezometria, quelle derivanti da analisi di iniezioni controllate (test con traccianti o tracer test nella comune dizione anglosassone). I test con traccianti sono stati in diverse occasioni adottati nel passato ma il loro uso é aumentato negli anni recenti grazie agli sviluppi delle tecniche geofisiche che semplificano il monitoraggio delle prove in situ. Fra queste, la Tomografia Elettrica Resistiva (ERT) sembra essere la piú appropriata per misurare le quantitá di interesse nel caso di iniezioni di traccianti, essendo possibile acquisire un grande numero di informazioni sull'evoluzione spazio-temporale dell'evoluzione del plume, a costi relativamente limitati.

Partendo dal presupposto che siano disponibili misure derivanti da una iniezione controllata in pozzo, il presente lavoro suggerisce alcuni approcci che, sulla base dei dati de-

ducibili dalle misure ERT, permettono di stimare la distribuzione spaziale di K e verifica la loro effettiva capacitá predittiva. Tali modelli risultano dall'accoppiamento di tecniche basate sul filtro di Kalman con modelli di trasporto Lagrangiano: l'applicazione ad una estesa serie di casi sintetici ha permesso inoltre di ottenere utili indicazioni in relazione a vantaggi e svantaggi di ciascuna delle metodologie proposte.

Contents

List of Figures

Chapter 1

Introduction

This thesis investigates new techniques to assess the hydraulic conductivity distribution K at local scale from tracer test data.

At the local scale, that is the scale characterized in the vertical and horizontal direction by the order of magnitude of the aquifer thickness, the flow and transport processes are of a three-dimensional nature (Dagan, 1986), so that the detailed knowledge of the K distribution is fundamental for the prediction of groundwater flow and contaminant transport. Indeed, the non-Fickian evolution of the solute plumes is mainly controlled by K distribution and this means that its proper definition plays a fundamental role, for example, in defining the well-catchment areas or in monitoring a landfill site. However, in natural porous media the K is usually characterized by a great spatial variability, making its assessment hard to achieve.

Aquifer hydraulic properties are usually estimated through inverse models and, in recent years, the progresses in computational resources have opened new possibilities in their use. At a local scale, these inverse model use data measured, for example, during laboratory analysis of samples, slug tests, hydraulic tomographies and tracer tests. The use of the latter in the last years has increased thanks to the development of geophysical method, through which extensive data can be obtained with limited cost and non-intrusive monitoring. From geophysical interpretation of tracer tests, it is then possible to acquire low cost informations about the spazio-temporal evolution of the plume, that is useful not only to assess the aquifer parameters, but also to obtain a

better understanding of the overall transport phenomena.
Among the techniques available to solve an inverse problem, the Ensemble Kalman Filter (EnKF) and the Ensemble Smoother (ES) are here applied and coupled with a Lagrangian transport model to assess the local distribution of K. The solute injection is analyzed in terms of spatio-temporal evolution of the concentrations or as travel times monitored in a prescribed control plane. In both cases, data are considered as deduced from an Electrical Resistivity Tomography (ERT) survey, which provides extensive information, even if it requires the resolution of a specific geophysical inversion and an in-situ calibration of a petrophysical law to estimate the concentrations.
The thesis proves that it is possible to retrieve the hydraulic conductivity distribution at local scale with a good accuracy through the proposed method. Some improvement of the experimented procedure, mainly related to the reduction of numerical issues are also presented. A comparison between EnKF and ES is realized and the inadequacy of the ES for the analyzed problem is shown. It is also verify the possibility of avoiding the calibration of a petrophysical relation by using travel time data.

1.1 Literature review

1.1.1 Parameter estimation by Kalman filter based methods

Inverse problems have been studied for the last forty years to estimate aquifer hydraulic properties and with the recent computational resources progresses their use has developed. By using the definition of McLaughlin and Townley (1996), an inverse model concerns the estimation of spatially variable model parameters which have physical significance but are difficult to measure, so that, under certain circumstances, these parameters may be inferred from measurements related to the analyzed system.
The inverse model topic is really wide and it is difficult to give a complete overview about it. Some general aspects are given hereinafter with reference to the reviews by McLaughlin and Townley (1996), Zimmerman et al. (1998), Carrera et al. (2005), Vrugt et al. (2008) and Hendricks Franssen et al. (2009). After that, the attention will be focused on the use of Kalman filter based method for parameter estimation.

In groundwater hydrology usually we are interested in estimating the hydraulic conductivity and, at this purpose, the data may come from permeability measurements, hydraulic head measurements, tracer concentrations from tracer test measurements and geologic information on the characteristics of the formation (Zimmerman et al., 1998). In general, an inverse technique follows these steps: 1) define the way in which the parametrization, i.e., the spatial variability, of the hydraulic conductivity is described; 2) determine the forward equation it uses to relate the parameters to the measurements and the cross covariance between them; 3) choose an optimization method to find the parameter estimation and the criterion to define "good" the parameters estimate.
A schematic mathematical development of the inverse problem considers a model A, in which the discrete time evolution of the state vector $\mathbf{x}$ is described through:

$$\mathbf{x_{t+1}} = \boldsymbol{A}(\mathbf{x_t}, \mathbf{u}, \boldsymbol{\beta}) \tag{1.1}$$

where $\mathbf{u}$ represent the forcings (e.g., boundary conditions), β represents the vector of parameter values and t denotes time. The available measurements are collected in the vector $\mathbf{Z}^m = [\mathbf{z}_1^m, \ldots, \mathbf{z}_n^m]$ being n the number of observations. Through the model A and given an initial guess for β, the vector of model predictions $\mathbf{Z}(\boldsymbol{\beta}) = [\mathbf{z_1}(\boldsymbol{\beta}), \ldots, \mathbf{z_n}(\boldsymbol{\beta})]$ is computed. The aim of parameter estimation becomes finding those values of $\boldsymbol{\beta}$ such that the difference between the model-simulated output and measured data is forced to be as close to zero as possible. This corresponds to minimize an objective function, for which different formulations are available.
To solve an inverse problem, different approaches have been elaborated even if, as argued by Carrera et al. (2005), they do not really differ from each other in their essence, though they may differ in the computational details (for example, the search technique they use to find a minimum). Usually the parameters depend non-linearly on measurements, and most of the times non linear methods, as the maximum a posteriori methods, the nonlinear least squares methods, the maximum likelihood methods, and the pilot points methods, are used to solve the inverse problems. As stated by McLaughlin and Townley (1996), they all minimize some versions of the objective function so that they primarily differ in the search technique they use to find a minimum and in the adopted

parametrization. Less frequent is the use of linear methods that are obtained, for example, by approximating the operator that maps the state space $\mathbf{x}$ in the measurements $\mathbf{Z}$.

While dealing with inverse models, the main difficulties related to their application have to be mentioned: 1) non-identifiability, that occurs when more than one set of parameters leads to a given solution of the forward problem; 2) non-uniqueness, that is present when more than one set of parameters leads to a minima of the objective function; 3) instability, that is present when small changes in the observations lead to large changes in the estimated parameters. When the inverse model does not suffer of the previous problems it is said "well-posed". Anyway most of the inverse model are ill-posed.

In the last years, it has become common the solution of the inverse problem through the application of the Kalman filter based technique, as the Ensemble Kalman filter (EnKF) or the Ensemble Smoother (ES). Unlike the optimization processes that require the calculation of the derivatives of the objective function with respect to model parameters (and this process is hard to program and highly consuming), the EnKF and ES overcome this problems, by conditioning multiple equally likely realizations of the state variables to the measurements.

In the EnKF the time-series of measurements are not incorporated simultaneously, but the stochastic realizations are sequentially conditioned in an iterative process. Moreover, the efficiency of the EnKF is due to the fact that the updating process is not based on traditional sensitivity-based optimisation, that means the objective function derivatives, but on optimum weighting of model prediction and measurements. The first formulation of the EnKF is given in Evensen (1994) but the first application in groundwater problem is due to Chen and Zhang (2006). In their work, they apply the EnKF to update the hydraulic conductivity in synthetic bi- and three-dimensional fields by assimilating hydraulic head measurements, proving the capabilities of the method. Afterwards, Liu et al. (2008) use the EnKF is used to estimate both the hydraulic conductivity spatial distribution and various transport model parameter at the MADE site (Boggs et al., 1992), through hydraulic head and tracer concentration data col-

lected in the field. They then use the in situ flowmeter measurements to valuate the quality of the K estimation. In Hendricks Franssen and Kinzelbach (2008), the EnKF is used for parameter estimation (the K spatial distribution) in a synthetic study. The purpose is to solve regional groundwater flow problems by assimilating hydraulic head measurements. Moreover in their work, the filter inbreeding problems is investigated in details and some solutions are proposed for its reduction. In the work by Sun et al. (2009), a deterministic EnKF (DEnKF) is applied to assess the K field by hydraulic head measurements, to demonstrate the capabilities of the DEnKF compared to the traditional perturbation-based ensemble filters, as the EnKF. The difference between these two approaches is related to the fact the in the latter a random noise is added to each observation so that the desired analysis ensemble covariance is replicated, but for the authors the added observation noise becomes an extra source of inaccuracy, while in the deterministic ensemble filters the need for sample perturbation is bypassed. Nowak, in the work (Nowak, 2009), reformulated the EnKF such that only parameters are updated, introducing the need for an iterative procedure each time new measurement data are assimilated. If only parameters are updated, it is assumed that some other sources of uncertainty (e.g., model forcings) are negligible as compared to material parameter uncertainty, even if the approach does not exclude the option to parametrize uncertain model forcings and jointly update material and forcing parameters. Huang et al. (2009) in their work use the EnKF to estimate heterogeneous hydraulic conductivity field by assimilating hydraulic head measurements and to improve prediction of solute plume with unknown solute source condition by assimilating measurements of concentration at monitoring wells. In the application by Xie and Zhang (2010), the EnKF is applied in a physical-based distributed hydrological model at a local catchment scale for the hydrological process. The purpose is to estimate the curve number parameter, as well as the prognostic variables such as the runoff, soil water content and evapotranspiration by using mainly runoff measurements. Hendricks Franssen et al. (2011) apply the EnKF in a complex and realist flow situation that includes an unconfined aquifer, the unsatureted zone and river-aquifer interactions to estimate both hydraulic conductivity and leakage coefficients.

The work by Jafarpour and Tarrahi (2011) demonstrates the incapabilities of the EnKF in reducing the uncertainty of the variogram model parameters through the assimilation of measurements.

Some recent papers (e.g., Zhou et al., 2011; Li et al., 2012; Schoeniger et al., 2012) deal with the Gaussian hypothesis of the variables probability distribution function, that ensures the optimal working conditions for the Kalman filter based methods. In particular in these works it is applied a technique called Normal Score Transform through which the probability distribution function of the variables is made Gaussian, with the purpose of optimizing the EnKF performances.

Related to the increased use of geophysical methods, there is the paper by Camporese et al. (2011) in which the update of the hydraulic conductivity is realized by using electrical conductivity measurements derived from ERT imaging of a synthetic tracer test experiment.

Another Kalman filter based technique used for parameter estimation is the Ensemble Smoother. Compared to the EnKF, the ES is not a recursive methods and all the collected data are assimilated at the same time, that is in only one step. The main advantage of the ES is that it does not require an embedded procedure and it can be applied off-line, independently to the model adopted to describe the system evolution. So far, a limited number of ES applications is reported in literature and are related to the identification of the hydraulic conductivity distribution (Bailey and Baú, 2010, 2012) and to the irrigation procedure assessment (Bailey et al., 2012).

1.1.2 Tracer test analysis by geophysical methods

It is known and accepted that tracer tests are useful to characterize subsurface properties and to study the spreading of solutes in a range of scales that vary from laboratory scale to regional field-scale. The information derived from a tracer test can be used to estimate the transport parameter, to set up and calibrate flow and transport models, to identify contamination pathways or also to test forward transport predictions obtained from deterministic or stochastic modeling framework.

A general review about the application of tracer test for the investigation of heteroge-

neous porous media can be found in Ptak et al. (2004). In this work, the implications that the aquifer heterogeneity and the scale of the problem have on tracer testing are highlighted and proper methods to perform tracer tests are described. A large variety of the tracer compounds used in tracer experiments are listed and the advantages and disadvantages of natural or forced gradient tracer test are separately analyzed, by recalling literature applications. Anyway, in most of the literature cases described by Ptak et al. (2004), information on tracer solute transport is available in terms of measured concentration time series, i.e. breakthrough curves, at monitoring locations and a large number of monitoring wells is required to properly describe the three-dimensional spatio-temporal evolution of a plume. Usually, the evaluation approaches are based on the breakthrough curves deduced from a multilevel sampling developed in a single monitoring well, and the spatial moment analysis of concentrations cannot be applied due to of the limited number of measurement locations available.

As already said, there is a certain number of examples in literature papers where tracer test data are used in an inverse model for parameter estimation (see, just to give few examples, Barlebo et al. (2004), Vrugt et al. (2005), Fu and Gomez-Hernandez (2009), Huang et al. (2009), Liu et al. (2010), Dafflon et al. (2011), Camporese et al. (2011), Kowalsky et al. (2012), Li et al. (2012), Chen et al. (2012)). Moreover, in the last years the use of tracer test data have increased, thanks to the enhancement of geophysical methods that provide more detailed measurements compared to the traditional techniques. Indeed one of the main advantage of geophysical methods is that they provide higher spatial resolution data than the traditional in situ techniques, even if the signals acquired (e.g., geoelectrical, geomagnetic, seismic) do not give a direct information about the relevant hydraulic properties, requiring the identification of a link between geophysics and hydraulics. Among other geophysical methods, the application of the ground penetrating radar (GPR) (e.g., Annan, 2005) and the electrical resistivity tomography (ERT) (e.g., Binley et al., 2002; Kemna et al., 2002a) in monitoring subsurface phenomena has expanded. This is mainly due to their capability to monitor the changes in soil electrical properties that are linked to the variations in moisture content or solute concentration, or in other words, to achieve significant spatio-temporal

distributed data with a (relatively) limited cost.

In this work, the attention is focused on data that can be achieved by ERT.

An ERT data acquisition system acquires a series of voltage and current measurements from electrode arrays placed in surface or in wells. The electrode arrays consist of electrode dipoles that communicate with other dipoles and they are fastened at regular intervals to a supporting shaft or string. They can be spaced very close to each other, depending upon the resolution needed. The extensive data resulting from measurements taken between the electrode arrays are processed to produce electrical resistivity time-lapse images that show spatial variations in electrical resistivity using inversion algorithms. The inversion procedure is not straightforward, and requires proper implementation and solution techniques (Kemna et al., 2002b, e.g.,), because the current paths through the medium are dependent on the electrical conductivity structure, and has to be solved iteratively.

The link between electrical conductivity changes in soil properties is stated by a petrophysical relationship (Archie, 1942). The electrical conductivity σ_b derives from electrolytic conduction in the pore fluid (water), i.e., on moisture content and pore water salinity and can be also expressed as a function of porosity ϕ, water saturation S_w and electrical conductivity of the pore water σ_w, according to the Archie's law:

$$\sigma_b = \frac{1}{a}\phi^m S_w^n \sigma_w \tag{1.2}$$

with m and n empirical parameters specific for the site and a a proportionality constant of the order of one. Since σ_w is proportional to the salinity, there is a direct physical link between changes in electrical conductivity and changes in ionic solute concentration C ($\Delta\sigma_w \sim \Delta C$), so that it is possible to monitor a tracer test through an ERT (note that the water content and porosity variations affect the electrical conductivity changes). ERT data are commonly collected along two-dimensional (2D) profiles or image plane, but through a proper electrode arrays distributions, it is possible to obtain also three-dimensional (3D) images.

This work uses the hypothesis that 3D or 2D concentration time-lapse images, deduced from an ERT survey, are available and the specific issues related with geophysical

inversion (e.g., Camporese et al., 2011) are not considered.

1.2 Thesis outline

Chapter 2 gives the theoretical fundamentals of the methods and the techniques used. The first part describes the transport theory that handles the solute dispersion problem in subsurface with a stochastic approach. Then, the Kalman filter theory is developed, focusing on the techniques applied in this work, the Ensemble Kalman filter (EnKF) and the Ensemble Smoother (ES).

In Chapter 3, it is investigated the possibility of retrieving the local distribution of K from a tracer test cloud evolution by using an EnKF-based inversion approach. After a preliminary analysis that check the effectiveness of the EnKF in dealing with an inverse problem and explores the proper number of Monte Carlo runs, the developed model is applied in a three-dimensional size domain that reproduces an heterogeneous aquifer. To find a good compromise between satisfactory results and computational requirement, different scenarios are studied by considering the use of different variables in the updating and assimilation procedure, carrying out different numerical tests for each case. In a first scenario, all the measured concentrations are assimilated to update the hydraulic log-conductivity field. Other scenarios consider the assimilation of the plume spatial moments (second scenario), and the assimilation of concentration in a selected number of spatial positions (third scenario). The tests carried out in the first scenario have mainly the purpose to determine if the EnKF is able to correct an initial wrong guess of the variance and the integral scale respect to the ones of the true field. The results show that an error on the variance initial guess can be satisfactory arranged, while a wrong integral scale causes problems of convergence. With the second and third scenarios the possibility of reducing the computational cost are analyzed, but only the the third scenario seems to give acceptable results.

Chapter 4 focuses on the numerical issues related to the application of EnKF and investigates procedures specifically developed to limit the memory amount required and the degradation of the solution. Several numerical experiments are carried out simulating tracer tests in synthetically generated heterogeneous aquifers and assimilating

concentrations derived from a hypothetical ERT monitoring. Scenarios where all the available concentration measurements are assimilated and the entire hydraulic conductivity field is updated are compared to other scenarios, developed in order to improve the approach efficiency. In the latter the K values are updated only in a limited number of nodes, being the hydraulic conductivity in the rest of the domain the result of a subsequent conditional generation. The technique is tested in an aquifer characterized by medium and high heterogeneity. The performances in each different situation are analyzed in details and reciprocally compared in terms of root mean square error of the reconstructed hydraulic log-conductivity field and reproduced tracer plume evolution.
In Chapter 5, a comparison between EnKF and another Kalman Filter based technique, the Ensemble Smoother (ES), is developed. The ES ensures a computational time lower than the EnKF one, and it can be applied "off-line" to any transport model. The investigation refers to a two-dimensional aquifer in which a tracer test is simulated. The issues related to non-Gaussianity of the flow and transport variables are analyzed and different transformation of the pdf's are considered in order to evaluate their influence on the performances of the two method. The developed numerical analysis suggests that the EnKF gives in all the cases better result for the hydraulic conductivity estimation problem when concentration measurements are assimilated.
In Chapter 6, the EnKF technique is coupled with travel time approach, to overcome the need of a in situ calibration of the petrophysical relation to infer concentrations from resistivity distribution. In a synthetic anisotropic aquifer, a solute injection is realized and the resistivity variations are measured in the control plane CP (transversal to the solute moving direction). Only on the basis of the resistivity variation, i.e. without using the Archie's law, the cumulative distribution function of the solute particles crossing the CP is assimilated during the update procedure. From the analysis it seems that this approach represents an effective tool to estimate the main characteristics of the hydraulic conductivity distribution.
Finally the conclusions summarize the main results of the thesis.

Chapter 2

Theory and Model

2.1 Transport phenomena driven by heterogeneity

The evolution within an aquifer of a passive tracer, i.e., a solute that does not react neither with the porous matrix nor with the conveying fluid, can be described by the following form of the advection-diffusion equation:

$$\frac{\partial C}{\partial t} + \mathbf{V}\nabla C = \mathbf{D}\nabla^2 C, \tag{2.1}$$

where $C = C(\mathbf{x}, t)$ is the solute concentration in the position $\mathbf{x}$ at time t, $\mathbf{V}$ is the mean value of the effective velocity and $\mathbf{D}$ is the spatially constant dispersion tensor. Theoretical results reported in literature (Dagan, 1989) have found experimental confirmation of the validity of equation (2.1) in heterogeneous formations when:

i) the mean value of the effective velocity, that drives the solute clouds, is $\mathbf{V} =< \mathbf{q} > /\phi$, where ϕ is the porosity, $\mathbf{q}= \mathbf{q}(\mathbf{x})$ is the Darcy velocity, and the symbol $< >$ stands for ensemble mean. The Darcy velocity $\mathbf{q}$ is obtained, in a steady state context, from the solution of the fluid mass balance equation:

$$\begin{aligned} \nabla \mathbf{q} &= 0 \\ \mathbf{q} &= -K\nabla h(\mathbf{x}), \end{aligned} \tag{2.2}$$

where h is the piezometric head and $K = K(\mathbf{x})$ is the hydraulic conductivity. In natural sedimentary aquifers the spatial variability of ϕ is usually negligible with respect to that of K (e.g., Gelhar, 1993), so that the porosity can be considered constant in the definition of the effective velocity. Under these hypothesis the erratic spatial variation of q is due only to the erratic spatial variation of K

ii) the generic component of the dispersion tensor is:

$$D_{ij} = \frac{1}{2}\frac{dX_{ij}}{dt} \tag{2.3}$$

where $X_{ij} =< X_i' X_j' >$ is the moment of inertia of the solute trajectories $X_i =< X_i > + X_i'$, whose fluctuations are related only to the spatial variations of K.

The spatial invariance of the mean velocity and of the moment of inertia defined by eq.2.3 requires the fulfillment of the statistical homogeneity and stationarity of the Eulerian and Lagrangian flow fields respectively. The Lagrangian ($\mathbf{V}[\mathbf{X}_t]$) and the Eulerian ($\mathbf{q}$) flow fields are related through the well-know equation:

$$\mathbf{X}_t(t; \mathbf{a}, t_0) = \mathbf{a} + \mathbf{X} + \mathbf{X}_d = \mathbf{a} + \int_{t_0}^{t} \mathbf{V}[\mathbf{X}_t(t'; \mathbf{a}, t_0)]dt' + \mathbf{X}_d, \tag{2.4}$$

where the total trajectory $\mathbf{X}_t$ considers also the displacement $\mathbf{X}_d$ associated with the "Brownian motion" related to the pore scale effects of correlation length λ_d, whereas $\mathbf{X}$, the displacement originating from convection by the fluid, depends on the spatially erratic fluctuations of K with correlation length λ. The coordinate $\mathbf{a}$ is the initial position at the injection time t_0. Equation 2.4 is an approximation that results from the worthwhile separation of $\mathbf{V}_t$, that is the velocity of a fluid particle, in $\mathbf{V}_t = \mathbf{V} + \mathbf{v}_d$, where $\mathbf{v}_d$ is the Brownian motion component (Dagan, 1989). In natural aquifers the integral scale of the hydraulic conductivity usually prevails on that one that characterizes the pore-scale phenomena, so that the latter can be neglected when dealing with transport for high Peclet values. The validity of the relation 2.3 is rigorously limited to this case.

By describing the hydraulic conductivity K in a natural porous medium as a spatially

random erratic variable, this randomness reflects upon the Eulerian flow field and, through equation (2.4), upon the Lagrangian flow field. The spatial distribution of K and the solute evolution in the aquifer are therefore linked through relation 2.3. In other words, starting from a known statistical definition of the hydraulic conductivity spatial distribution, it is possible to describe the velocity driven dispersion phenomena.

2.1.1 The "macroscopic" concentration approach

Assuming that at time $t = t_0$ a solute body of concentration C_0 is injected in a volume V_0, being $C = 0$ everywhere for any $t < t_0$, the mass of solute dM contained in the infinitesimal volume $d\mathbf{a}$ inside the volume V_0 is

$$dM = \phi C_0(\mathbf{a}) d\mathbf{a} \tag{2.5}$$

and it moves along a trajectory of equation $\mathbf{x} = \mathbf{X}_t(t; \mathbf{a}, t_0)$. The relative concentration may be written as

$$\Delta C(\mathbf{x}, t; \mathbf{a}, t_0) = C_0(\mathbf{a}) \delta(\mathbf{x} - \mathbf{X}_t) d\mathbf{a}, \tag{2.6}$$

where δ is the Dirac's function and the porosity is assumed constant for simplicity in the whole domain. In eq.2.6 and in the followings, we refer to the total displacement $\mathbf{X}_t$. According with theoretical considerations of section 2.1 (negligibility of the pore scale dispersion contribute), it is $\mathbf{X}_t = \mathbf{X}$, i.e., the fluctuating velocity field is consequence of the erratic spatial variability of K. From a general point of view, the validity of the approach is kept also when the pore scale dispersion is non negligible. In this case, the relation between the K distribution and the velocity randomness would be partially shadowed by the pore scale contribute. Dealing with real applications, our interest is focused to the average concentration C over a voxel ΔV of finite size whose centroid is at $\mathbf{x}$, the concentration field resulting as:

$$\begin{aligned} C(\mathbf{x}, t; t_0) &= \frac{1}{\Delta V} \int_{\Delta V} \int_{V_0} \Delta C(\mathbf{x}', t; \mathbf{a}, t_0) d\mathbf{a}\, d\mathbf{x}' \\ &= \frac{1}{\Delta V} \int_{\Delta V} \int_{V_0} C_0(\mathbf{a}) \delta[\mathbf{x}' - \mathbf{X}_t(t; \mathbf{a}, t_0)] d\mathbf{a}\, d\mathbf{x}'. \end{aligned} \tag{2.7}$$

The size of the voxel depends on the physical device used to detect the value of C in real cases: it may be a well diameter in which mixing occurs or may be related to the resolution of a 3D ERT survey. In any case, equation (2.7) gives a direct connection between the value of the "resident concentration" (Dagan, 1989) and the Lagrangian flow field.

A schematic representation of the problem is given in figure 2.1

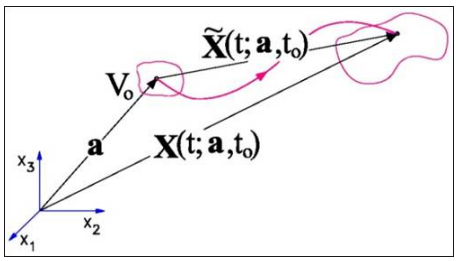


Figure 2.1: Schematic representation of the concentration estimation problem

In this thesis, the 3D Lagrangian transport is simulated by means of a finite volume solver for steady state groundwater flow, coupled with the Pollock's post-processing algorithm for the computation of the particle trajectories (Pollock, 1988). The Eulerian velocity field is used for the computation of the trajectories of a set of N particles suitable for the simulation of a contaminant release. Defined as M the tracer total mass injected at time $t = t_0$

$$M = \phi \int_{V_0} C_0(\mathbf{a}) d\mathbf{a} = \phi C_0 V_0 \tag{2.8}$$

the total mass uniformly injected in the system, a dimensionless concentration field is computed as (Crestani et al., 2010; Camporese et al., 2011):

$$\begin{aligned} \tilde{C}(\mathbf{x}, t; t_0) &= \frac{\phi}{M} \int_{\Delta V} \int_{V_0} C_0(\mathbf{a}) \delta[\mathbf{x}' - \mathbf{X}_t(t; \mathbf{a}, t_0)] d\mathbf{a}\, d\mathbf{x}' \\ &= \frac{1}{N \Delta V} \int_{\Delta V} \sum_{i=1}^{N} \delta[\mathbf{x}' - \mathbf{X}_t(t; \mathbf{a}, t_0)] \Delta\mathbf{a}\, d\mathbf{x}'. \end{aligned} \tag{2.9}$$

In the tracer test simulations that will be realized in this work, only the preasimptotic (non-Fickian) regime is taken into account. As a consequence, the concentration distri-

bution is always non-Gaussian and a noticeable tail is manifest for large values of the hydraulic log-conductivity variance.

To simplify the notation, in the following we will denote as C the dimensionless concentration $\tilde{C}$, defined in (2.9).

In the macrodispersion process (Rubin, 2003), instead of a smoothed evolution of the solute cloud typical of Brownian motion, we observe a fragmentation of the plume driven by the velocity field controlled by the spatial variability of K. In this context, a useful representation of the plume dispersion may be given by the knowledge of its spatial moments.

If M is defined as in 2.8, the solute mass conservation leads to the centroid position $\mathbf{R}$ expressed as the spatial mean on the entire domain Ω

$$\begin{aligned}
\mathbf{R}(t;t_0) &= \frac{\phi}{M}\int_{\Omega}\mathbf{x}\,C(\mathbf{x},t;t_0)\,d\mathbf{x} \\
&= \frac{\phi}{M}\int_{\Omega}\int_{V_0}\mathbf{x}\,C_0(\mathbf{a})\,\delta[\mathbf{x}-\mathbf{X}_t(t;\mathbf{a},t_0)]d\mathbf{a}\,d\mathbf{x} \\
&= \frac{\phi}{M}\int_{V_0}\mathbf{X}_t\,(t;\mathbf{a},t_0)\,C_0(\mathbf{a})\,d(\mathbf{a}) \qquad (2.10)
\end{aligned}$$

and to the second spatial moment S_{ij}

$$\begin{aligned}
S_{ij}(t;t_0) &= \frac{\phi}{M}\int_{\Omega}[x_i - R_i(t;t_0)][x_j - R_j(t;t_0)]C(\mathbf{x},t;t_0)\,d\mathbf{x} \\
&= \frac{\phi}{M}\int_{\Omega}\int_{V_0}[x_i - R_i(t;t_0)][x_j - R_j(t;t_0)]\,C_0(\mathbf{a})\,\delta[\mathbf{x}-\mathbf{X}_t(t;\mathbf{a},t_0)]d\mathbf{a}\,d\mathbf{x} \\
&= \frac{\phi}{M}\int_{V_0}[X_{t,i}(t;\mathbf{a},t_0) - R_i(t;t_0)][X_{t,j}(t;\mathbf{a},t_0) - R_j(t;t_0)]\,C_0(\mathbf{a})d(\mathbf{a}) \qquad (2.11)
\end{aligned}$$

In a statistically homogeneous velocity field, the injection of a solute volume V_0 large enough with respect to the correlation length λ ensures the fulfillment of the ergodic requirements (e.g., Dagan, 1991; Papoulis and Pillai, 2002). In this case, the spatial mean $\mathbf{R}$ and the second spatial moment S_{ij} approach the expected value $\langle\mathbf{R}\rangle$ and the displacement tensor $X_{ij} = \langle(X_i - \langle X_i\rangle)(X_j - \langle X_j\rangle)\rangle$, i.e., the first and the second central moment of the Lagrangian velocity field, respectively. In the literature it is well documented that the X_{ij} time behavior exhibits in 3D a strong asymmetry, being its

longitudinal component greater than the transversal components (Dagan, 1989).

2.1.2 The travel time approach

The "macroscopic concentration" approach considers spatial distributions of the concentrations at different times. The solute transport can be described also by a different approach. Each solute particle moving from the source, crosses a plane (CP) located downstream in a prescribed position $\bar{x}$ perpendicular to the mean flow direction at time τ (figure 2.2) . The knowledge of the statistics of τ through which the transport can be described is named "travel time" approach (Dagan and Nguyen, 1989; Dagan et al., 1992; Cvetkovic et al., 1992). According with the scheme of figure 2.2, the particle crosses the CP at the coordinates $\boldsymbol{\eta}$=(η , ζ). Then, the solute transport is described through the knowledge of $\boldsymbol{\eta}$ and τ in a CP located in a prescribed position downstream the source.

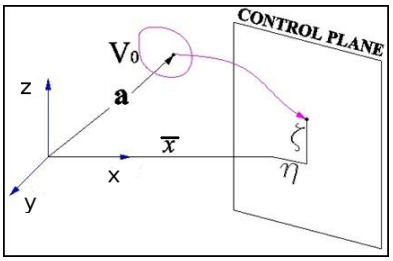


Figure 2.2: Schematic representation of the travel time estimation problem

A single solute particle of mass $\Delta m = C_0 d\mathbf{a}$, that comes from an instantaneous injection of a solute body of volume V_0, is considered. The particle starts from $\mathbf{x} = \mathbf{a}$ at time t_0 and crosses the CP in $\mathbf{y} = \boldsymbol{\eta}$ ($\mathbf{y} = (y\,,\, z)$) at time $t = \tau$, being both $\mathbf{y}$ and t random variables. Due to the randomness of the velocity field, the solute flux transfered through the CP by the particle is a random function of $\mathbf{y}$ and t and can be expressed as:

$$\Delta q(y, z, t; \bar{x}, \mathbf{a}, t_0) = \Delta m\, \delta(y - \eta)\, \delta(z - \zeta)\, \delta(t - \tau) \tag{2.12}$$

The total discharge is obtained by integration on the entire CP

$$\Delta Q(t; \bar{x}, \mathbf{a}, t_0) = \int_{CP} \Delta q(\boldsymbol{\eta}, t; \bar{x}, \mathbf{a}, t_0)\, d\boldsymbol{\eta} = \Delta m\, \delta(t - \tau) \tag{2.13}$$

and the total mass of solute that has crossed CP is

$$\Delta M(t;\bar{x},\mathbf{a},t_0) = \int_{t_0}^{t}\int_{CP} \Delta q(\boldsymbol{\eta},t';\bar{x},\mathbf{a},t_0)\, d\boldsymbol{\eta}\, dt' = \Delta m\, H(t-\tau) \tag{2.14}$$

where $H(t-\tau)$ is the Heaviside step function ($H=1$ for $t \geq 0$ and $H=0$ for $t<0$). The statistical moments definition of Δq, ΔQ and ΔM (all are random variables because of $\boldsymbol{\eta}$ and τ), let the development of the travel time approach. In particular, the first moment is considered. Defined as $g_1(\boldsymbol{\eta},\tau;x,\mathbf{a})$ the probability distribution function (pdf) of the particle originating at $\mathbf{x}=\mathbf{a}$ and $t=t_0$ to cross the CP in $\mathbf{y}=\boldsymbol{\eta}$ at $t=\tau$ and $G_1(t;\bar{x},\mathbf{a})$ the corresponding cumulative distribution function (CDF), by definition the expected value becomes then:

$$\langle \Delta q \rangle = \int_{CP}\int_{t_0}^{\infty} \Delta q\, g_1 d\boldsymbol{\eta}\, d\tau = \Delta m\, g_1(y,t;\bar{x},\mathbf{a},t_0) \tag{2.15}$$

that is, the expected value of the solute flux is proportional to the transverse displacement and travel time joint pdf. In the same way, we get:

$$\langle \Delta Q \rangle = \Delta m\, g_1(t;\bar{x},\mathbf{a},t_0) \tag{2.16}$$

and

$$\langle \Delta M \rangle = \Delta m\, G_1(t;\bar{x},\mathbf{a},t_0). \tag{2.17}$$

The previous relations determined for one particle can be generalized for any input finite injection on a volume V_0. The expected value of the total mass is

$$\langle M(t;\bar{x},V_0,t_0)\rangle = \int_{V_0} d\mathbf{a}\int_{t_0}^{t} dt'\int_{t_0}^{\infty} d\tau \int_{CP} C_0(\mathbf{a})\delta(\mathbf{y}-\boldsymbol{\eta})\delta(t-\tau) g_1(\boldsymbol{\eta},\tau;\bar{x},\mathbf{a}) d\boldsymbol{\eta} \tag{2.18}$$

where, by following Dagan et al. (1992)the joint pdf g_1 can be assumed as the product of the two independent $g_2 2$ and g_3

$$g_1(\boldsymbol{\eta},\tau;\bar{x},\mathbf{a}) = g_2(\tau;\bar{x},\mathbf{a})\, g_3(\boldsymbol{\eta};\bar{x},\mathbf{a}) \tag{2.19}$$

Without loss of generality, the CP can be viewed as the sum of N_{sCP} sub-control planes, each of area A_i, that do not intersect each other, being $A = \sum_i A_i$ the total area of the CP, so that eq. 2.18 can be written as:

$$\begin{aligned}\langle M(t;\bar{x},V_0,t_0)\rangle &= \int_{V_0} d\mathbf{a}\int_{t_0}^{t} dt'\int_{t_0}^{\infty} d\tau \sum_{i=1}^{N_{sCP}}\int_{A_i} C_0(\mathbf{a})\,\delta(\mathbf{y}-\boldsymbol{\eta})\,\delta(t-\tau)\,g_1(\boldsymbol{\eta},\tau;\bar{x},\mathbf{a})\,d\boldsymbol{\eta}\\ &= \int_{V_0} d\mathbf{a}\int_{t_0}^{t} dt' \sum_{i=1}^{N_{sCP}} C_0(\mathbf{a}) g_1(t;\mathbf{y}|A_i,\bar{x},\mathbf{a})\\ &= \int_{V_0} d\mathbf{a} \sum_{i=1}^{N_{sCP}} C_0(\mathbf{a}) G_1(t;\mathbf{y}|A_i,\bar{x},\mathbf{a}) \qquad (2.20)\end{aligned}$$

In a stationary velocity field, by using the eq. 2.19 we can write

$$\langle M(t;\bar{x}-a_x)\rangle = \int_{V_0} d\mathbf{a} \sum_{i=1}^{N_{sCP}} C_0(\mathbf{a})\, g_3[(y-a_y\,,\,z-a_z)|A_i;\bar{x}-a_x]\, G_2(t;\bar{x}-a_x) \quad (2.21)$$

If the injection volume reduces to an area in a plane perpendicular to the coordinate x (the mean flow direction), we have:

$$\langle M(t;\bar{x}-a_x)\rangle = G_2(t;\bar{x}-a_x)\int_{V_0} C_0(a_y a_z)\, g_3[(y-a_y\,,\,z-a_z)|A_i,\bar{x}-a_x]\, da_y da_z \quad (2.22)$$

and when the C_0 can be assumed constant on the injection area, $\int_{V_0} g_3[(y-a_y\,,\,z-a_z)\,da_y da_z = P(V_0|A_i)$ is the probability of a particle moving from V_0 to cross the CP in the area A_i at any time. Thus, the probability is related to the transverse dispersion of the plume at time

$$t_x = \frac{t\langle U\rangle}{x-a} \qquad (2.23)$$

In a stationary context, the three dimensional transverse dispersion is limited and the transverse displacement of the path lines can be considered negligible. Consequently $P(V_0|A_i)$ can be assumed proportional to $A_{V_0,i}$, being the latter the projection of the injection volume V_0 on the CP, according to the mean flow direction.

Note finally that from the solute breakthrough curves, a variety of arrival time measures, such as peak arrival times, temporal moments of concentration, and arrival time quantiles, can be calculated. In the following, arrival time quantiles are used. At each

monitoring location, the quantiles τ_q are given implicitly by the normalized cumulative concentration,

$$Q(\tau) = \frac{\int_0^\tau C(t)dt}{\int_0^\infty C(t)dt} \tag{2.24}$$

where C is concentration and t is the time. The q_t quantile τ_q is the value of time τ at which $Q(\tau_q) = q$. For example, the 0.5 quantile or median arrival time $\tau_{0.5}$ is the time when half of the solute has gone past a monitoring location.

2.2 The Kalman filter based techniques

According to Evensen (2009b), the combined parameter and state estimation problem for a dynamical model can be formulated as finding the joint pdf of the parameters and model state, given a set of measurements and a dynamical model with known uncertainties. Using Bayes theorem, the problem can be written in the simplified form

$$f(\mathbf{y}, \boldsymbol{\alpha}|\mathbf{z}) = \gamma f(\mathbf{y}, \boldsymbol{\alpha}) f(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha}), \tag{2.25}$$

where $f(\mathbf{y}, \boldsymbol{\alpha})$ is the joint pdf for the model state $\mathbf{y}$ (as function of space and time) and the parameters $\boldsymbol{\alpha}$, $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha})$ is the likelihood function of the measurements $\mathbf{z}$, and γ is a normalization constant whose computation requires the evaluation of the integral of (2.25) over the multi-dimensional solution and parameter space. Note that in writing equation (2.25) we implicitly assume that the only uncertainty in the model formulation lies in the parameters, while boundary and initial conditions are perfectly known. If, as usual, we work with a model state that is discretized in time, we can represent $\mathbf{y}$ at fixed time intervals as $\mathbf{y}_i = \mathbf{y}(t_i)$ with $i = 0, 1, ..., k$. If we further assume that the model is a first-order Markov process, we can define the pdf for the model integration from time t_{i-1} to t_i as $f\left(\mathbf{y}_i|\mathbf{y}_{i-1}, \boldsymbol{\alpha}\right)$. Let us now assume that also the measurements $\mathbf{z}$ can be divided into subsets of measurement vectors $\mathbf{z}_i$, collected at the same time steps of the model t_i and that the measurement errors are uncorrelated in time. Under

these hypotheses and from Bayes theorem, equation (2.25) becomes (Evensen, 2009b)

$$f\left(\mathbf{y}_1, ..., \mathbf{y}_k, \boldsymbol{\alpha}|\mathbf{z}\right) = \gamma f\left(\boldsymbol{\alpha}\right) \prod_{i=1}^{k} f\left(\mathbf{y}_i|\mathbf{y}_{i-1}, \boldsymbol{\alpha}\right) f\left(\mathbf{z}_i|\mathbf{y}_i, \boldsymbol{\alpha}\right), \tag{2.26}$$

in which the pdf of the parameters $f\left(\boldsymbol{\alpha}\right)$ is expressed explicitly. Rewriting (2.26) as a sequence of iterations and integrating out the state variables at all previous times, we obtain

$$f\left(\mathbf{y}_i, \boldsymbol{\alpha}|\mathbf{z}_1, ..., \mathbf{z}_i\right) = \gamma f\left(\boldsymbol{\alpha}\right) f\left(\mathbf{y}_i|\mathbf{y}_{i-1}, \boldsymbol{\alpha}\right) f\left(\mathbf{z}_i|\mathbf{y}_i, \boldsymbol{\alpha}\right). \tag{2.27}$$

The combined parameter and state estimation can thus be formulated sequentially using Bayesian statistics, under the condition that measurement errors are independent in time and the dynamical model is a Markov process. Equations (2.25) and (2.27) can be solved numerically by means of a Monte Carlo approach, which approximates the probabilistic information conveyed by the conditional pdfs of the state, parameters, and measurements with an ensemble of realizations of size NMC. Compared to the traditional Kalman filter, the main advantage of this approach is that it can deal with non-linear models and measurement operators, thanks to its capability to approximate the probability distributions of the variables with an ensemble of realizations.
Each of the NMC state vectors is propagated in time according to the forecast model, which can be expressed as a vector-valued discrete-time state equation:

$$\mathbf{y}^j(t) = A[\mathbf{y}^j(\tau), \boldsymbol{\alpha}^j, t, \tau]; \quad t_0 \leq \tau < t; \quad \mathbf{y}^j(t_0) = \mathbf{y}_0^j, \tag{2.28}$$

where $\mathbf{y}^j(t)$ is the j^{th} $(j = 1, \ldots, NMC)$ state vector predicted by the model at the time t, $\boldsymbol{\alpha}^j$ is the j^{th} set of model parameters, A is the operator relating the system state at the current time t to the system state at the previous time τ, and $\mathbf{y}_0^j$ is the initial condition at time t_0.
At time t_i, m_i measurements are available and the model describing how these measurements and the system state are related is also expressed as a vector-valued discrete-time measurement equation:

$$\mathbf{z}^j(t_i) = \mathbf{H}\mathbf{y}_{true}(t_i) + \mathbf{w}^j(t_i), \tag{2.29}$$

where $\mathbf{H}$ is the operator that maps the model state to the measurements, $\mathbf{y}_{true}(t_i)$ is the true state, and $\mathbf{z}^j(t_i)$ is the j^{th} vector of observations at the time t_i, which is obtained perturbing the m_i measurements with a random noise $\mathbf{w}^j(t_i)$, representing the measurement errors. Here we assume that $\mathbf{w}^j(t_i)$ is normally distributed with expected value equal to zero and assigned variance σ^2_{meas}. Under the hypothesis that the pdfs for the model prediction as well as the likelihood are Gaussian, it is possible to update each realization of the system state according to the following equation, which is obtained by minimizing the model error covariance matrix (Evensen, 2009b):

$$\mathbf{y}^j_{upd}(t_i) = \mathbf{y}^j + \mathbf{P}_e\mathbf{H}^T(\mathbf{H}\mathbf{P}_e\mathbf{H}^T + \mathbf{R}_e)^{-1}(\mathbf{z}^j(t_i) - \mathbf{H}\mathbf{y}^j(t_i)), \tag{2.30}$$

where $\mathbf{y}^j_{upd}(t_i)$ is the j^{th} realization of the updated system state, $\mathbf{P}_e$ is the prior estimate of the system state error covariance matrix, which is computed by sampling the ensemble statistics, and $\mathbf{R}_e$ is the measurement error covariance matrix.

Two different techniques are applied in this work to solve equation (2.30): the Ensemble Kalman Filter (EnKF) and the Ensemble Smoother (ES).

Within the formulation of the ES, equation (2.25) requires that the pdfs are approximated from an integration of the ensemble through the whole assimilation time period. In other words, all of the data are processed by equation (2.30) in one step, and the solution is updated as a function of space and time, using the space-time covariances estimated from the ensemble of model realizations.

On the other hand, within the EnKF formulation, equation (2.27) is a sequential expression that can be solved through incremental updates given by (2.30) with the data at the current time step only.

Note that in both cases, when the model pdfs and the likelihood function are Gaussian, the Bayesian formulation corresponds to the minimization of a quadratic cost function (Evensen, 2009b). The only difference is that in ES the time dimension is included in the optimization, while in EnKF the minimization is performed at each assimilation time. However, when the pdfs are not Gaussian, equation (2.30) is an approximation and no longer yields optimal updates in terms of a cost function minimization, although the general Bayesian formulation remains valid, i.e., sampling of a posterior pdf is still

performed.
The computation of the update step implemented in this study follows initially (Chapter 3) the algorithm elaborated by Keppenne (2000) and, after that, the square root algorithm introduced by Evensen (2004).

2.3 The coupled inverse model

The NMC realizations of the random function K are generated out by an improved sequential Gauss simulation algorithm (Baú and Mayer, 2008; Crestani et al., 2010). The hydraulic conductivity field is assumed log-normally distributed ($Y = lnK$) with expected value $\langle Y \rangle$, variance σ_Y^2, and exponential isotropic correlation structure $\rho_Y = exp(-|\boldsymbol{\xi}|/\lambda)$, being $\boldsymbol{\xi} = \mathbf{x}_2 - \mathbf{x}_1$ the lag distance. The flow field is then calculated by the finite volume solver at steady state with appropriate boundary conditions that ensure a mean gradient $\mathbf{J} = const$. The Eulerian velocity field is used, by the Pollock's post-processing algorithm, for the computation of the trajectories of a number N of particles suitable for the simulation of a contaminant release.
The estimation of the hydraulic log-conductivity field is realized by measurements assimilation and using an augmented system in which only the model parameters, i.e., the Y values, are considered:

$$\mathbf{y}^j(t) = [Y_1, \ldots, Y_{cp}]^j, \tag{2.31}$$

In (5.1), cp is a chosen number of voxels (nodes) that can be at most equal to the total number of nodes discretizing the domain and $Y_1, \ldots, Y_{cp}$ are the Y values in the cp nodes.
In this thesis, three types of measurements are used in the assimilation procedure:

- the concentration C. A snapshot of the plume is realized at specified times and all the non-zero C values are taken into account (see equation 2.9);

- the spatial moments. The moments R_x, R_y, R_z, S_x, S_y, S_z are measured and assimilated at specified times (see equations 2.10 and 2.11);

- the cumulative distribution function of the travel times for each subplanes in which the control plane is divided. At specified times, all the CDF values (time-discretized) till that time is assimilated.

The effectiveness of the Y values update is related to the cross-correlation between Y and C, expressed by the product $\mathbf{P_e H}^T$ in equation (2.30).
At time t_0, NMC realizations of the hydraulic log-conductivity field are generated with assigned statistical structure and the state vectors are built as in equation (5.1). Starting with the the same initial concentration C_0, for each Y field the solute plume is propagated forward in time to the first measurement time t_1, according to the Lagrangian transport model. At t_1 the hydraulic log-conductivities $Y_1, \ldots, Y_{cp}$ are updated to reflect the effect of the measurements $\mathbf{z}_1$.
The transport problem is then solved in the updated fields for the time period $[t_0, t_2]$. The whole process continues sequentially: first a propagation step over each interval between t_0 and measurement time t_i and then an update step at each measurement time t_i. It has to be stressed that, in order to ensure mass conservation throughout the simulation, after every assimilation step the plume evolves in the updated Y field from t_0 with the same initial concentration field. This recursive application of the Kalman filter differs from the classic sequential one commonly adopted. By restarting the plume evolution with the same initial concentration in updated Y fields distribution, we apply the predictive analysis expressed by the Kalman gain on a system state improved at each assimilation step. This procedure is similar to the advanced first-order second-moment (AFOSM) method adopted in risk analysis (Yen et al., 1986). As in the AFOSM method, to overcome nonlinearity problems and the lack of Gaussianity, the solution is recursively approached with improved values of the estimator (the Kalman gain) corresponding to improved estimates of the solution itself.
In the ES application, there is no need for sequential updates and the parameters are estimated in a single off-line step. At time t_0, each realization of the ensemble of NMC Y fields is initialized with the same concentration distribution C_0 and the solute plume is propagated forward in time until the last assimilation time. The measurement vectors

are thus assembled as

$$\mathbf{z}^j = [\mathbf{z}(t_1), \mathbf{z}(t_2), \ldots, \mathbf{z}(t_{tm})]^{jT}, \tag{2.32}$$

i.e., the perturbed measurements for all the measurement times (tm) are stored in a vector of dimension $[\sum_{i=1}^{tm} m_i]$, for each Monte Carlo simulation. Since the ES does not require sequential updates, the computational time is expected to be significantly lower than in the EnKF, even though the dimension of the vectors and matrices in equation (2.30) is significantly larger than in EnKF.

Chapter 3

Hydraulic conductivity assessment via tracer test data assimilation

3.1 Introduction

Natural aquifers are characterized by a great variability of the hydraulic properties, such as porosity, hydraulic conductivity, and storage coefficient: as a consequence, an exact definition of these properties is practically never possible.

In this Chapter, the applicability of the coupled model EnKF-Lagrangian transport, developed to characterize the hydraulic conductivity distribution from a tracer test injection at a local scale, is analyzed.

Initially, a preliminary test is performed in a quasi-one dimensional domain to analyze the EnKF feasibility in inverse problems. After that, the same approach is applied in a three-dimensional finite size domain reproducing an heterogeneous aquifer.

Defined as L an appropriate unit length, the domain dimensions are set to $8\ L \times 8\ L \times 8\ L$ and eight concentration images of the traveling plume are recorded and used as error-free measurements. Different applications of the coupled Lagrangian transport - EnKF model (as described in section 2.3) are carried out, by making different choices in the system state variables and by assimilating different quantities.

The main objective of this work is to explore the performance of the coupled approach in different assimilation scenarios and to find an optimal choice for both the variables to be assimilated and to be updated. Not only to limit the computer memory resources and CPU time required by the method, but also to assess the implications of the assimilation scheme for future applications with real ERT data.

3.2 Description of the scenarios

Three different scenarios of system states and assimilation variables, with their relevant results, are taken into account.

In this Chapter, the EnKF is used as in inverse model tool and also for data assimilation. This means that the system state vector is given by the parameters (Y values) and the variables, so that, at each update, not only the Y fields but also the variables estimated by the model are corrected, by means of the measurements.

In the first scenario, the system state $\mathbf{y}^j(t_i)$ consists of the concentration values C and the log-conductivity values Y for all the n discretization nodes of the domain. For each realization, the system state vector is:

$$\mathbf{y}^j(t_i) = [C_1, \ldots, C_n, Y_1, \ldots, Y_n]^T, \tag{3.1}$$

where the subscript refers to the node numbers.

Each vector $\mathbf{y}^j(t_i)$ is updated every time concentration measurements are available, using equation (2.30), in which the measurement vector is $\mathbf{z}(t_i) = [C_1, \ldots, C_m]$, where $m \leq n$ is the total number of nodes in which the concentration is not zero. The updated Y field is then used to restart the plume evolution till the next time for which measurements are available.

In the second scenario, we drastically reduce the number of variables by limiting the the system state dimension according to the following expression:

$$\mathbf{y}^j(t_i) = [R_x, R_y, R_z, S_x, S_y, S_z, Y_1, \ldots, Y_{cp}]^T, \tag{3.2}$$

being cp a selected number of log-conductivity values ($cp \leq n$). The update is realized by assimilating R_x, R_y, R_z and S_x, S_y, S_z that are the first and second spatial moments, respectively, along the principal directions x, y, and z (see equations 2.10 and 2.11). After each update, the whole Y field is re-computed through a conditioned generation on the basis of the cp Y values updated by the EnKF assimilating the (six) plume moments.

The third scenario can be considered as a compromise between the first and the second ones. Concentration is assimilated only in a limited number of selected cells and the log-conductivity values are updated only in those same cells, their position properly chosen within the area affected by the solute evolution. The system state in this case is thus:

$$\mathbf{y}^j(t_i) = [C_1, \ldots, C_{cp}, Y_1, \ldots, Y_{cp}]^T. \tag{3.3}$$

As in the second scenario, also in this case the Y field after each update is generated by conditioning on the cp updated Y values.

All the scenarios are summarized in table 3.1.

	assimilated data	system state	vector size
scenario 1	concentration values $C_1, \ldots, C_m$	$\mathbf{y}^j(t_i) = [C_1, \ldots, C_m, Y_1, \ldots, Y_n]^T$	$2n$
scenario 2	first and second spatial concentration moments $R_x, R_y, R_z, S_x, S_y, S_z$	$\mathbf{y}^j(t_i) = [R_x, \ldots, S_z, Y_1, \ldots, Y_{cp}]^T$	$\leq (6+n)$
scenario 3	concentration values $C_1, \ldots, C_{cp}$	$\mathbf{y}^j(t_i) = [C_1, \ldots, C_{cp}, Y_1, \ldots, Y_{cp}]^T$	$\leq 2n$

Table 3.1: Summary of the scenarios carried out.

The main objective of this work is to explore the performance of the coupled approach in the different assimilation scenarios and to find an optimal choice for both the variables to be assimilated and to be updated. Not only to limit the computer memory resources and CPU time required by the method, but also to assess the implications of the assimilation scheme for future applications with real ERT data.

3.3 Results and discussion

3.3.1 Model setup

Defined as L the unit length, the simulated domain has dimensions 8 $L \times 8$ $L \times 8$ L, discretized along each direction with $L/4$ side cubic cells, for a total of $33 \times 33 \times 33 = 35937$ nodes. The boundary conditions are as follows. Pressure head is imposed (Dirichlet) in $x = 0$ and $x = 8l$, resulting in a mean gradient $J = 0.6$ along the main flow direction, while no-flow boundary conditions (Neumann) are imposed at the remaining sides of the domain. A reference Y field has been generated with mean $< Y >= 0.025$, variance $\sigma_Y^2 = 0.5$, and isotropic exponential correlation structure with integral scale $\lambda = 1$ L. Its representation is given in figure 3.1.

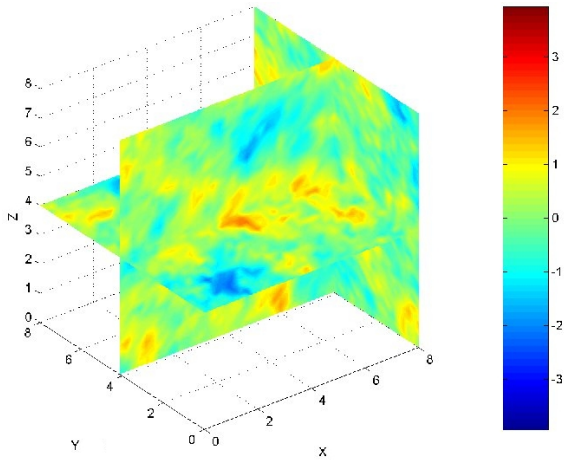


Figure 3.1: The hydraulic log-conductivity spatial distribution of the reference field

The solute is assumed as instantaneously injected in a well with diameter of 0.5 L, vertical size of 5.7 L, and centered in x=0.875 L and y=4.125 L. The injection is simulated by 1104 particles distributed along the well. The concentration is proportional to the ratio between the number of particles within a cell of the domain and the total number of injected particles and is recorded every 0.5 T (being T an appropriate time unit), for a total of eight concentration images.

A number of synthetic experiments are carried out to verify the capabilities of the proposed approach to reproduce the reference Y field in a variety of test cases. For each test case, we start from an ensemble of log-conductivity realizations characterized by

initial statistical properties obviously different from those used to create the reference field.

It is important to notice that the measurements are always perturbed with an assigned noise (σ^2_{meas}) to represent the errors affecting the assimilated data. The value of σ^2_{meas} has clearly a certain impact on the solutions. In particular, the algorithm used in this study (Keppenne, 2000) is characterized by poor performances when σ^2_{meas} tends to zero.

3.3.2 Preliminary analysis

Before applying the model to the domain above mentioned, a preliminary analysis is performed with the purpose of verifying if the EnKF can deal with the parameter estimation problem we want to solve. At this aim, a quasi one-dimensional case, obtained by simplifying the previous three-dimensional one, is realized. The true field is characterized by only two values of Y extracted from a normal distribution with mean zero and variance 0.5. The first value is $Y_{1,true}$=-0.3066, the second one is $Y_{2,true}$=0.2421 (see figure 3.2). An injection is realized and the concentrations are measured, only till the plume is in the first portion ($Y_{1,true}$) of domain, for eight instants.

Different scenarios are elaborated by varying the mean and the variance of the normal distribution from which the two Y values of the field are extracted. A summary is given in Table 3.2 that gives the characteristic of the probability distribution function from which the Y values are extracted.

	NMC	$Y_1 \sim N(< Y_1 >, \sigma^2_{Y_1})$	$Y_2 \sim N(< Y_2 >, \sigma^2_{Y_2})$	σ^2_{meas}
reference field	/	$Y_1 \sim N(0, 0.5)$	$Y_2 \sim N(0, 0.5)$	/
preliminary test 1	2000	$Y_1 \sim N(0.5, 0.75)$	$Y_2 \sim N(-0,5, 0.75)$	0.01
preliminary test 2	2000	$Y_1 \sim N(-0.3066, 0.75)$	$Y_2 \sim N(-0,5, 0.75)$	0.01
preliminary test 3	2000	$Y_1 \sim N(-0.3066, 0.75)$	$Y_2 \sim N(-0,5, 0.75)$	0.01
preliminary test 4	2000	$Y_1 \sim N(-0.0323;, 0.75)$	$Y_2 \sim N(-0.0323;, 0.75)$	0.01

Table 3.2: Summary of the preliminary tests carried out.

The results are analyzed by visualizing how the initially wrong values of Y evolve during the measurements assimilation procedure. In Figure 3.3(a) and 3.3(b) the results for preliminary test 2 and 3 respectively are shown (preliminary test 1 and preliminary test

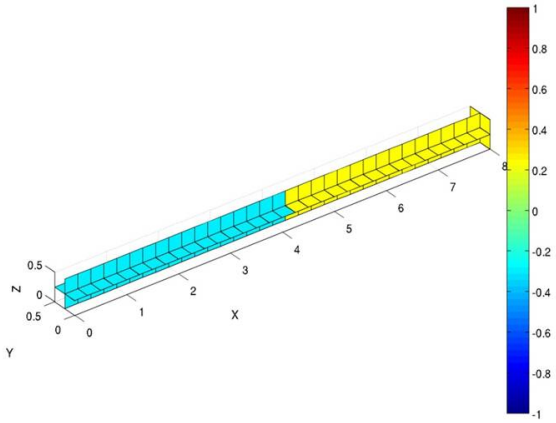
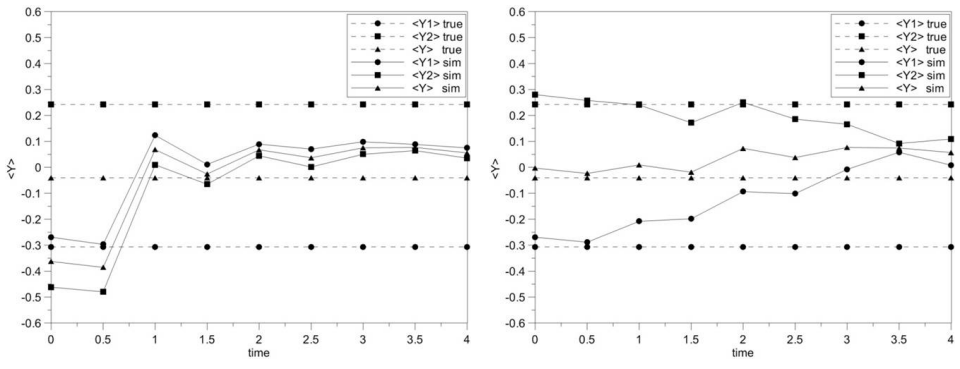


Figure 3.2: True field of the preliminary analysis

4 are not shown being analogous). In each graph it is displayed how the values of Y_1, Y_2 and Y, that is the average between Y_1 and Y_2, are corrected during the assimilation procedure. The true values are also reported. It can be seen that, despite the fact that the concentration measurements involve only the first part of the domain, also the Y_2 is corrected and, in particular, both the Y_1 and the Y_2 tend to converge toward the true mean value.

This simple case shows that the EnKF is effective in correcting the parameters, since the two values of the field are modified and made to converge towards the true mean one that is necessary to guarantee that, on average, the simulated evolution of the plume is consistent with the true one.



(a) (b)

Figure 3.3: Y_1, Y_2 and mean Y trend values after each update for (a) preliminary test 2 and (b) preliminary test 3. The true Y values are also reported.

This preliminary analysis demonstrates the EnKF capabilities in correcting an initial wrong Y distribution, in accordance with the plume movement. Obviously, in this case the plume does not spread and different values of Y can not be recognized in the space. Anyway, the mean value is assessed consistently with the only information available, that is the movement of the plume centroid.

3.3.3 First scenario

As already mentioned, in the first scenario the system state is given by the values of concentration and log-conductivity in all the discretization nodes and concentration measurements are assimilated.

Different experiments are carried out, varying initial variance and integral scale with respect to the reference field. Table 3.3 summarizes all the tests. Notice that the first three experiments are designed to study the effect of the Monte Carlo ensemble size. For each experiment the initial approximation of the Y field is homogeneous, as

	NMC	$< Y >$	σ^2_Y	integral scale λ_Y	σ^2_{meas}
reference field	/	0	0.5	1	/
test 1	500	0.025	0.75	1	0.01
test 2	1000	0.025	0.75	1	0.01
test 3	2000	0.025	0.75	1	0.01
test 4	2000	0.025	0.25	1	0.01
test 5	2000	0.025	0.25	1	0.0001
test 6	2000	0.025	0.75	0.5	0.01
test 7	2000	0.025	0.75	2	0.01

Table 3.3: Summary of the experiments carried out for scenario 1.

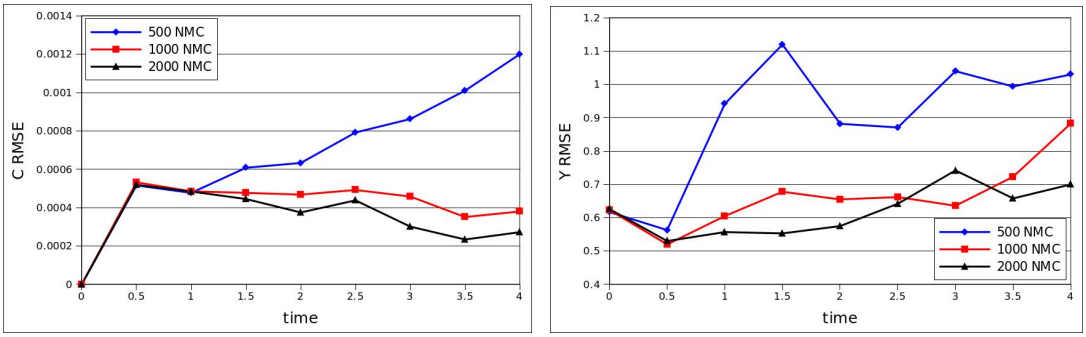
a result of the average over NMC Monte Carlo realizations, with mean and variance as reported in Table 3.3. At each update we expect that the average log-conductivity field converges towards the reference field.

In order to assess the ensemble size (NMC) that ensures a good performance at a reasonable computational cost (we recall here that the convergence of Monte Carlo processes is theoretically proportional to $1/\sqrt{NMC}$), the evolution of the root mean square error ($RMSE$) is evaluated for experiments 1, 2, and 3. For a generic variable

ψ, $RMSE$ is calculated as:

$$RMSE(t) = \sqrt{\sum_{i=1}^{nod}(\psi_{i,enkf}(t) - \psi_{i,tf}(t))^2/nod}, \tag{3.4}$$

where nod is total number of nodes considered for the computation, $\psi_{i,enkf}(t)$ is the ensemble mean, and $\psi_{i,tf}(t)$ is the "true" value (i.e., the reference field value), where subscript i indicates the node number. $RMSE(t)$ is computed only for the nodes included in a limited vertical slice of the domain, where the solute plume is expected to evolve, assuming that the log-conductivity field outside this portion does not influence the dispersive process. Figures 3.4(a) and 3.4(b) show the time evolution of $RMSE(t)$, i.e. the RMSE value after each update step, for the concentration and the log-conductivity, respectively. As expected, as the ensemble size increases, the error is



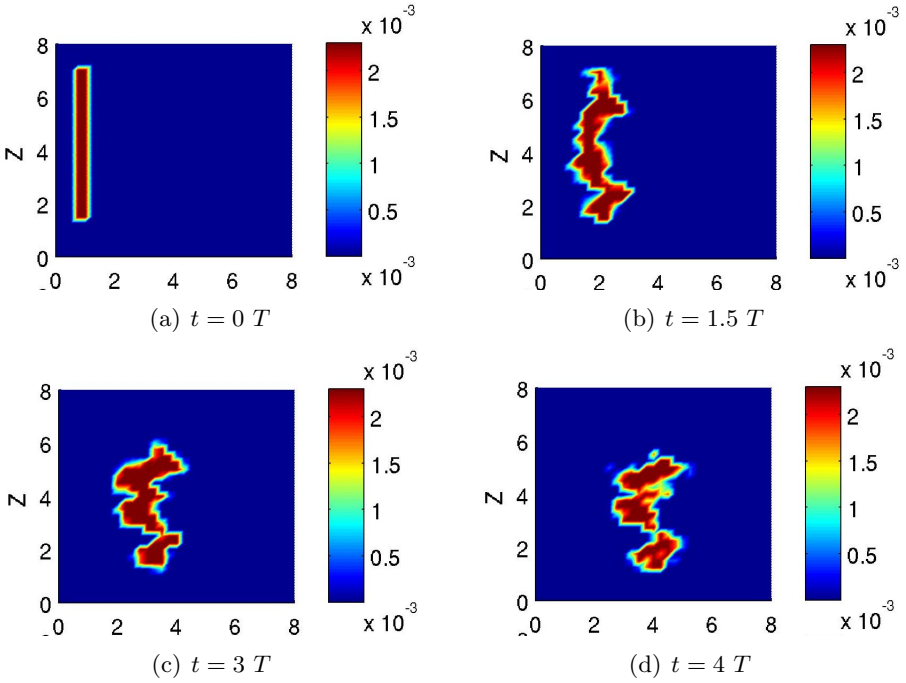
(a) (b)

Figure 3.4: Time evolution of (a) the concentration root mean square error and (b) the hydraulic log-conductivity root mean square for tests 1, 2, and 3, in a portion of the domain limited around the plume path.

reduced. The solution accuracy improves significantly by increasing NMC from 500 to 1000, to a lesser extent for a variation from 1000 to 2000. On the basis of these findings, an ensemble size of 2000 has been chosen for all the subsequent tests (from 4 to 7), computational costs being still affordable.

Figures 3.5 and 3.6 show the concentration profiles along the vertical cross section at $y = 4.125\ L$, for the reference case and test 3, respectively. The snapshots are taken at times $t = 0\ T$, $1.5\ T$, $3.0\ T$, and $4.0\ T$ (corresponding to the third, sixth, and eighth update, respectively). No visible difference can be evinced between the reference con-

centration field and the one obtained at the end of the assimilation experiment. Figure 3.7 shows the retrieved hydraulic log-conductivity field in test 3 at the end of the simulation, i.e., after eight updates. By comparing the latter to the reference field (figure 3.1), even though from a purely qualitative standpoint, it is manifest the capability of the proposed approach to reproduce correctly the main features of the reference field.



(a) $t = 0\ T$ (b) $t = 1.5\ T$

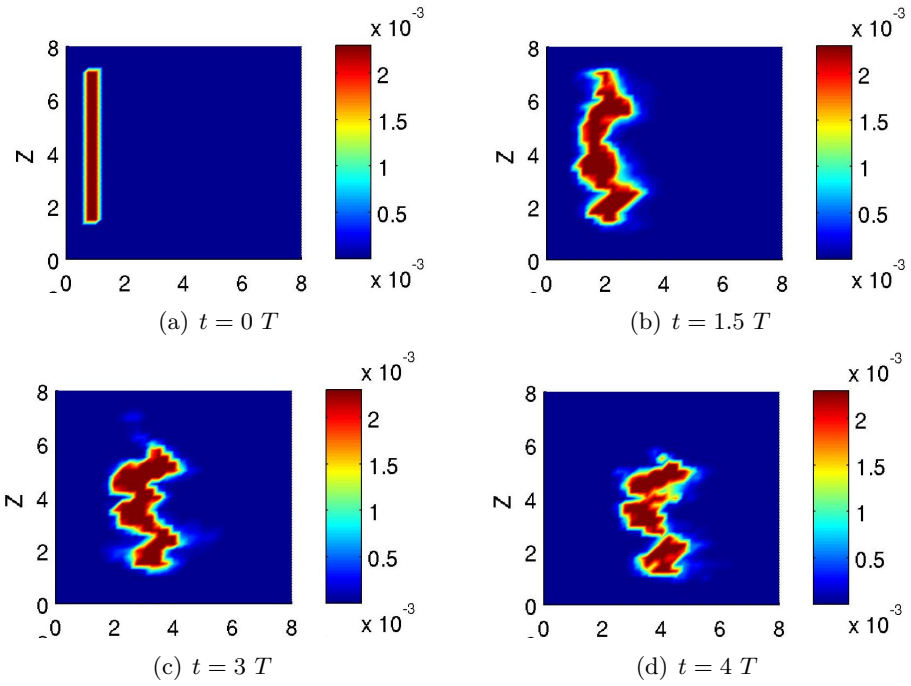
(c) $t = 3\ T$ (d) $t = 4\ T$

Figure 3.5: True concentration distribution at $t = 0\ T$, $1.5\ T$, $3\ T$, and $4\ T$ along the vertical cross-section at $y = 4.125\ L$.

A typical effect of the EnKF is the progressive decrease of the ensemble variance, which is computed here as:

$$\sigma^2_{Y,i} = \frac{1}{NMC} \sum_{mc=1}^{NMC} (Y_{i,mc})^2 - \left(\frac{1}{NMC} \sum_{mc=1}^{NMC} Y_{i,mc} \right)^2 , \tag{3.5}$$

where $Y_{i,mc}$ is the log-conductivity value at the i-th node in the mc-th realization. Figure 3.8 shows four snapshots of the ensemble variance spatial variability for experiment 3, at times $t = 0\ T$, $1.5\ T$, $3\ T$, and $4\ T$. While for $t = 0\ T$ the variance is equal, on average, to the theoretical value initially assigned ($\sigma^2_Y = 0.75$), in the subsequent times it is progressively reduced, broadly following in time and space the plume evolu-



(a) $t = 0\ T$ (b) $t = 1.5\ T$

(c) $t = 3\ T$ (d) $t = 4\ T$

Figure 3.6: Test 3: concentration distribution obtained during the assimilation procedure at $t = 0\ T$, $1.5\ T$, $3\ T$, and $4\ T$ along the vertical cross-section at $y = 4.125\ L$.

tion. In other words, the application of the EnKF progressively reduces the uncertainty on the log-conductivity values, trying to mimic the available measurements. The field obtained reproduces well the reference one in the area affected by the plume evolution, while some discrepancies are manifest in positions where the ensemble variance reduction is caused only by the propagation of the updates linked to the correlated structure of the log-conductivity field. It must be stressed however that the variance reduction does not guarantee anywhere the convergence of the simulation to the true solution, i.e., the reference Y field.

A good reconstruction of both the plume and the log-conductivity field is obtained also in experiment 4 (not shown here) by assigning an initial variance lower than that of the reference simulation.

In experiments 6 and 7 the initial value of the integral scale is varied, as well as the variance, with respect to the characteristics of the reference field. In experiment 7, in particular, the initial integral scale is twice as great as that of the true field, thus the effect of the assimilation is propagated to a greater number of nodes around the

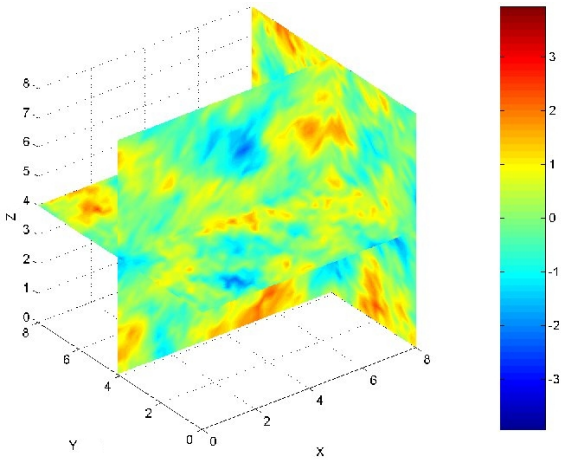


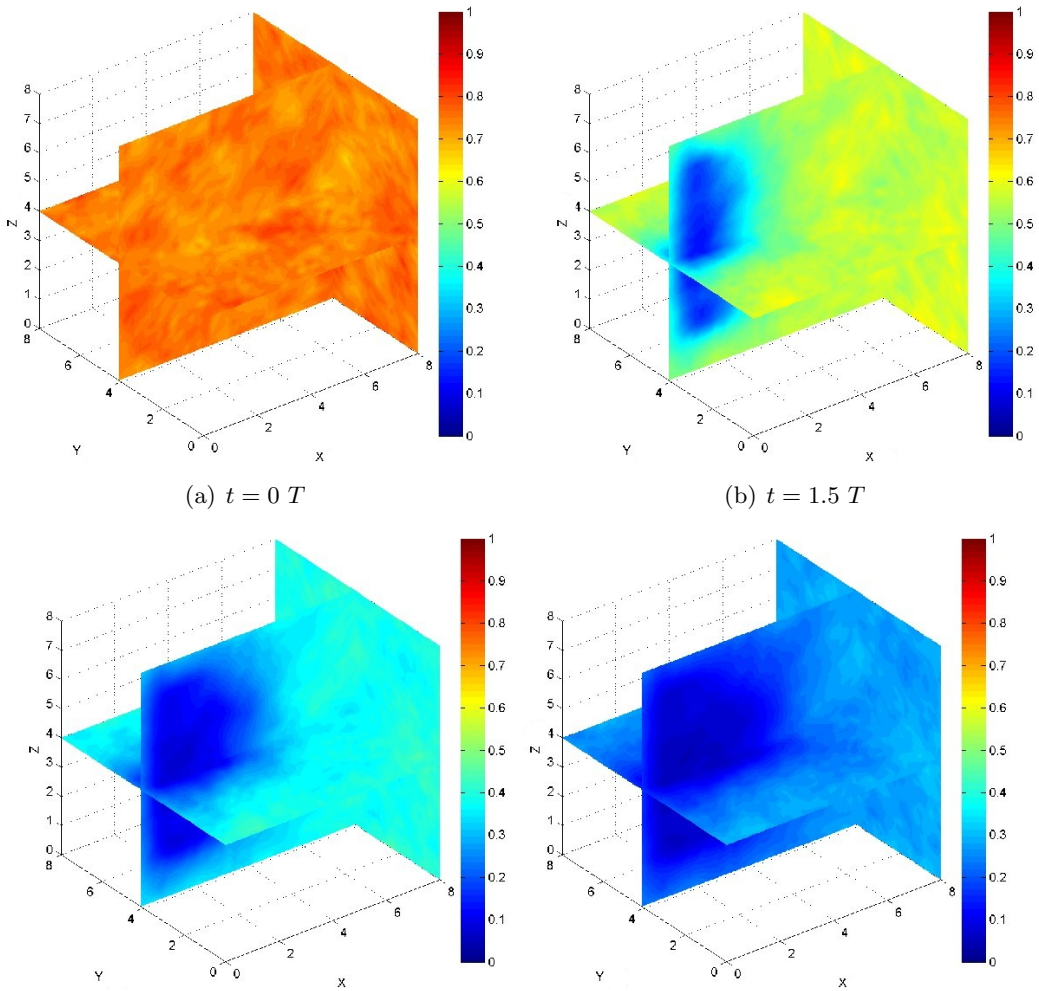
Figure 3.7: The hydraulic log-conductivity spatial distribution obtained at the end of test 3.

plume. As a consequence, the early times assimilation procedure influences areas that are far from the natural evolution of the plume in the reference field. The final solutions of C and Y could be thus defined as "spurious", with exaggerated spatial variations of log-conductivity, both in terms of spatial extent and contrast between blocks with different Y (Figure 3.9).

In experiment 6 (not shown), taking an initially halved integral scale with respect to the reference field leads to equally unsatisfactory results. In this case the assimilation procedure affects only a limited part of the domain and the final result shows discrepancies similar to those of the test case 7, even though for opposite reasons.

3.3.4 Second and third scenarios

To reduce the total number of variables and to make the coupled approach easier to handle, we developed a second scenario, where only the first and second spatial moments of the contaminant plume along the principal directions x, y, and z are assimilated. The system state consists of these moments and of the conditioning values of hydraulic conductivity, suitably selected among the total number of cells. Unfortunately the results obtained are not satisfactory. The application of equation (2.30) to the system state obtained by considering only the concentration first and second statistical moments seems to have a poor influence on the updating procedure. The space-time behavior of the reproduced plume differs significantly from the reference one



(a) $t = 0\ T$ (b) $t = 1.5\ T$

(c) $t = 3\ T$ (d) $t = 4\ T$

Figure 3.8: Snapshots of the ensemble variance spatial distribution for test 3 at times at $t = 0\ T$, $1.5\ T$, $3\ T$, and $4\ T$.

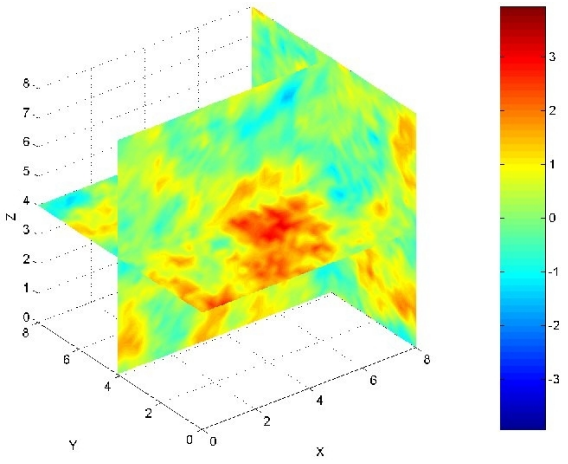


Figure 3.9: The hydraulic log-conductivity spatial distribution obtained at the end of test 7.

and, as a result, the reproduced log-conductivity field is even worse. Several reasons can explain this failure: among others, it is likely that the knowledge of only the first two concentration moments is not enough to reproduce the non-Fickian dispersion that dominates our test cases.

For this reason we shifted to the third scenario, where the concentration in a limited number of cells is assimilated. Again, only a selected number of log-conductivity values is updated and the position of the assimilated concentration values is properly chosen. The application of this scenario is related to the definition of the conditioning points statistical properties. The main issue concerns the evaluation of the variance and the integral scale of the conditioning points: indeed these parameters are used during the conditioned generation to define the statistical properties of the whole domain. They are calculated from the variogram of the conditioning points through an interpolation of the experimental variogram with the theoretical one (obviously with an exponential trend, as assumed). This operation is however not linear and thus is the source of some difficulties for the determination of these parameters, related for example to the optimal search algorithm used to perform this operation or to identify the ideal number of conditioning points for a satisfactory experimental variogram.

In addition, this last scenario is more elaborated than the first one: the number and the position of the cells in which the conditioning values are assigned must be conveniently chosen within the domain. Obviously when the *cp* number of conditioning positions

reach the total node number n of the entire domain, scenario 3 falls into scenario 1, whose performance is shown in the previous section.
This scenario will be deeply analyzed in the next chapter.

3.4 Final remarks

It has been suggested a novel inverse modeling approach, developed to recover the spatial distribution of hydraulic conductivity in heterogeneous aquifers at the local scale, and based on the interpretation of tracer tests with a Lagrangian transport model coupled to the ensemble Kalman filter data assimilation technique. A series of tests, carried out with reference to a synthetic test case and several different scenarios of system state assembling and variables assimilated, gives the following findings.
The number of Monte Carlo simulations, i.e., the ensemble size used for EnKF, must be properly chosen to ensure an accurate reproduction of the covariance matrices, while maintaining a reasonable computational cost. In the case examined, the analysis of the root mean square error shows that a NMC value of 500 is not sufficient to ensure an adequate solution, which can be obtained instead by $NMC = 2000$.
In all the tests concerning the first scenario, where the system state is defined by the non-zero nodal concentrations augmented by all the log-conductivity values, the plume retrieval is generally satisfactory and substantially unaffected by errors in the initial definition of the aquifer properties.
When only the initial variance is varied with respect to the reference field, the method also provides a good representation of the hydraulic conductivity field.
A less satisfactory log-conductivity field solution is obtained when an error on the initial estimate of the integral scale λ is considered. The convergence toward the "true" solution is slow when the initial λ value is underestimated, while if the same value is overestimated, spurious solutions can be obtained, with abnormal spatial variations of the log-conductivity compared to the spatial behavior of the reference field. However, it must be emphasized that we are dealing with an "inverse problem", i.e, there are several fields of Y that respect the plume concentration distribution and it is thus not possible to ensure the uniqueness of the solution.

Different scenarios have been analyzed to reduce the number of variables and the computational burden. An attempt to limit the assimilated variable only to the first and second statistical moments of the concentration field and to consider in the augmented state only a few log-conductivity values in a properly chosen number of cells seems to be inadequate. A more elaborated scenario, where concentration is assimilated in a limited number of cells and the conditioning Y values are updated in the same selected positions.

Nevertheless, the proposed approach seems to be an effective tool for estimating the distribution of hydraulic conductivity at the local scale, although further research is necessary, especially to identify the best combination of quantities to assimilate.

Chapter 4

Numerical issues and possible improvements

4.1 Introduction

The results of a detailed analysis developed to overcome numerical issues related to the use of EnKF is reported in this Chapter. The possibility to reduce the required memory amount by limit the number of positions where the concentration data are assimilated and the hydraulic conductivity data are updated is investigated.
Within the same Lagrangian approach already described, the concentration data are assimilated by means of the EnKF to deduce the K spatial distribution, focusing on both theoretical and numerical perspectives of the hydrological inversion. We analyze the ability of the method to reconstruct the local distribution of K in mild and strong non-linear and non-Gaussian contexts. Moreover, we consider the opportunity to reduce the size of the system state required by the EnKF algorithm at each update, limiting, at the same time, the filter divergence, that implicitly affects EnKF applications. To this aim, only a limited number of selected K values is updated and the remaining portion of the field, whose assessment is necessary to reproduce the plume evolution, is obtained by a conditional generation, which is then used only to fill the skeleton obtained by the assessed K values in the conditioning points. Although the latter technique could be included in the wide family of the "pilot-point" geostatistical inversion methods (PPM)

(e.g., Certes and DeMarsily, 1991; Ramarao et al., 1995; Gomez-Hernandez et al., 1997; Alcolea et al., 2006; Doherty et al., 2010), the fundamental difference is that the classic PPM is a model-fitting method (Rubin et al., 2010), while here the EnKF algorithm, which is a Bayesian approach, is used to update the parameters in the conditioning points, instead of an objective function that deterministically minimizes the mismatch between predictions and observations.

A tracer test is simulated in two synthetic aquifers, characterized by mild and strong heterogeneity distribution of K, assuming as known the concentration distribution thanks to a hypothetical ERT monitoring experiment. For each of the two test cases, three scenarios are carried out: in the first scenario all the available concentration measurements are assimilated and the entire hydraulic conductivity field is updated while in the other two scenarios K is updated only in a limited number of nodes by assimilating the concentrations in these same nodes, being the remaining portion of the field conditioned to the updated K values.

4.2 The inversion model

The general methodology used is the same that the one described in section 2.3 but hereinafter more details to describe the application of this Chapter are given.

We estimate the hydraulic log-conductivity by assimilating concentration measurements and using an augmented system in which only the model parameters, i.e., the Y values, are considered:

$$\mathbf{y}^j(t) = [Y_1, \ldots, Y_{cp}]^j, \tag{4.1}$$

In (5.1), cp is a chosen number of voxels (nodes) that can be at most equal to the total number of nodes discretizing the domain and $Y_1, \ldots, Y_{cp}$ are the Y values in the cp nodes.

At time t_0, NMC realizations of the hydraulic log-conductivity field are generated with assigned statistical structure and the state vectors are built as in equation (5.1). Starting with the same initial concentration C_0, for each Y field the solute plume is propagated forward in time to the first measurement time t_1, according to the La-

grangian transport model. At t_1 the hydraulic log-conductivities $Y_1, \ldots, Y_{cp}$ are updated to reflect the effect of the concentration measurements $\mathbf{z}(t_1)$. If $cp < tnod$ the hydraulic log-conductivity values in the remaining $(tnod - cp)$ nodes are obtained by a generation conditioned to the updated $Y_1, \ldots, Y_{cp}$ values. In other words, $Y_1, \ldots, Y_{cp}$ correspond to the conditioning points through which, after each assimilation time, the whole Y field is reproduced. Clearly, if $cp = tnod$, at each assimilation step all the Y values are updated and there is no need for a conditional generation. The transport problem is then solved in the updated field for the time period $[t_0, t_2]$. The whole process continues sequentially: first a propagation step over each interval between t_0 and measurement time t_i and then an update step at each measurement time t_i (Figure 4.1) plus a conditional generation, if required. It has to be stressed that, in order to ensure mass conservation throughout the simulation, after every assimilation step the plume evolves in the updated Y field from t_0 with the same initial concentration field.

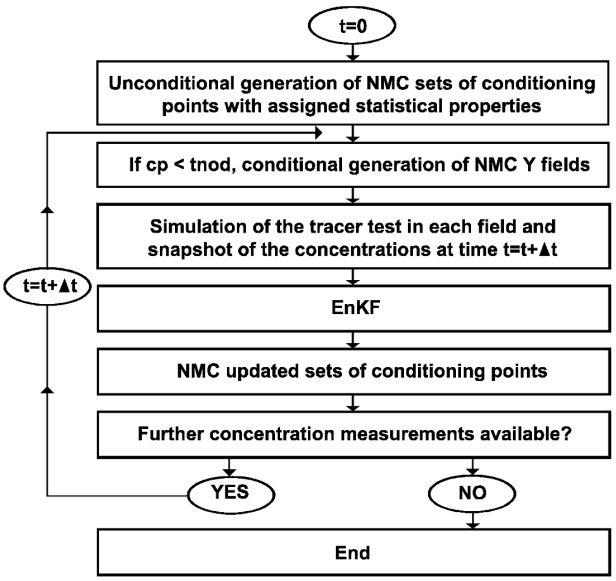


Figure 4.1: Flowchart of the proposed inversion approach.

Since one of the key points of the model is to use conditional generations, two main issues related to the conditioning points need to be addressed. The first issue concerns the number and position of the conditioning points: to guarantee a meaningful conditional generation there must be at least one conditioning point every integral scale.

In other words, cp has to be large enough to represent properly the EnKF covariance matrices, while, on the other hand, the larger is cp, the larger is the memory required by the scheme to build, store, and manipulate these covariance matrices. To assess the best combination of number and position that assures a good solution and the trade-off with the computational effort, various scenarios are tested, whose results are discussed in the following section. The second issue concerns the geostatistical characterization of the conditioning points. After each assimilation step, the $Y_1, \ldots, Y_{cp}$ values are the result of the EnKF update and to recover the remaining $(tnod - cp)$ values of Y, the variance σ_Y^2 and the integral scale λ of the updated Y values have to be estimated. The former is computed as:

$$\sigma_Y^2 = \sum_{mc=1}^{NMC} \sum_{i=1}^{cp} \frac{Y_{mc,i}^2}{NMC \cdot cp} - \left(\sum_{mc=1}^{NMC} \sum_{i=1}^{cp} \frac{Y_{mc,i}}{NMC \cdot cp} \right)^2, \tag{4.2}$$

where $Y_{mc,i}$ is the Y value at the i^{th} conditioning point for the mc^{th} realization, while the integral scale λ is evaluated by means of a linear regression on the ensemble averaged experimental correlation structure.

4.3 Numerical experiments

4.3.1 Model setup

As previously mentioned, two different reference fields with medium and relatively high σ_Y^2 are defined to test the proposed inversion model in a number of synthetic experiments.

The first reference field has dimensions of 8 $L \times 8$ $L \times 8$ L, where L is an arbitrary and consistent unit length, and is discretized along each directions into $L/4$ sided cubic volumes (voxels), for a total of $33 \times 33 \times 33 = 35937$ corresponding nodes. Dirichlet boundary conditions are applied at $x = 0$ L (piezometric head $h = 100.0$ L) and at $x = 8$ L ($h = 95.2$ L), while Neumann no-flow boundary conditions are imposed along the remaining sides of the domain. The Y distribution of the reference field is the result of a single unconditional generation with isotropic and exponential covariance

structure with integral scale $\lambda = 1\ L$, resulting in spatial mean $\langle Y \rangle = 0.03$ and variance $\sigma_Y^2 = 0.43$. A representation of the reference field is given in Figure 4.2. A tracer test is simulated by assuming an instantaneous solute injection in a well of 0.5 L diameter, vertical size of 6 L, and centered in $x = 0.875\ L$ and $y = 4.125\ L$ (Pos0 in Figure 4.3a and b). The solute is simulated by 4992 particles uniformly distributed in the horizontal and vertical directions (Figure 4.3b and c), which correspond to 52 particles within each voxel of the injection volume (uniform density of 3328 particles/L^3). On the basis of a preliminary analysis and considering the discretization of the domain and, consequently, of the velocity field, this number of particles is adequate to properly describe the plume evolution. The concentration, computed according to equation (2.9), i.e., proportional to the ratio between the number of particles within a cell of the domain and the total number of injected particles, is recorded every 0.5 T, where T is any consistent time unit, for a total of eight concentration images. These concentration measurements are used during the assimilation procedure.

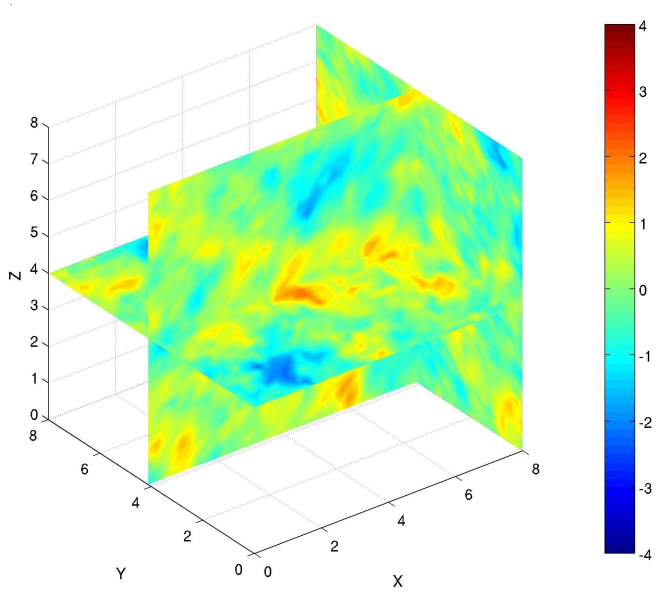


Figure 4.2: The hydraulic log-conductivity spatial distribution for the first reference field (8 $L \times 8\ L \times 8\ L$).

The generation of the second reference field, isotropic with integral scale $\lambda = 1\ L$, results in spatial mean and variance values of $\langle Y \rangle = -0.08$ and $\sigma_Y^2 = 1.59$. The injection well position and the time intervals for the concentration measurements are

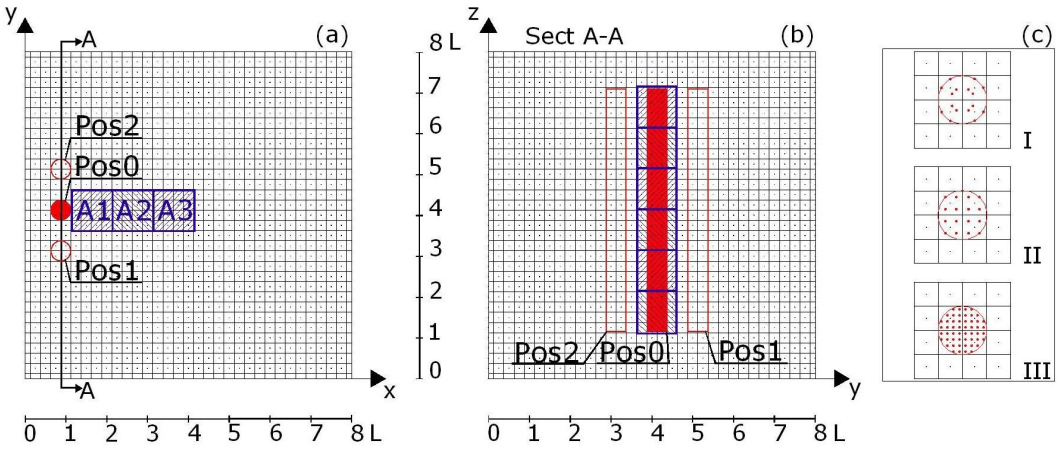


Figure 4.3: Numerical discretization of the first test case domain with the injection positions Pos0, Pos1 and Pos2. A1, A2, and A3 define the zone of the domain used for the result analysis. A detail of the planimetric particle distribution is shown on the right.

the same as in the previous case. Considering the dependency of the plume dispersion on σ_Y^2 (Dagan, 1989), with a larger variance we expect that the solute covers a larger portion of the domain. The main objective of the second test case is then to investigate the impact of a more spreaded plume on the inversion procedure. Since with large σ_Y^2 values the trajectory distribution is manifestly non-Gaussian and, in the Fickian regime, it spreads along x with a positive skewness (Salandin and Fiorotto, 1998), we extend the longitudinal domain size to 16 L, in order to prevent the loss of particles from the downstream boundary. However, the retrieval of the hydraulic conductivity field is assessed only in the first half portion of the domain (i.e., up to $x = 8\ L$), for comparison with the results of the first test case. A proper choice of the Dirichlet boundary conditions ($h = 100.00\ L$ at $x = 0\ L$ and $h = 94.24\ L$ at $x = 16\ L$) ensures the same mean velocity as in the first test case. Neumann no-flow boundary conditions are imposed along the remaining sides of the domain.

The hydraulic log-conductivity probability distribution that initializes the procedure is characterized by mean and variance values different from the ones of the reference field. Number and spacing of the conditioning points, prior geostatistical parameters, and measurement uncertainty (i.e., the variance of the measurement error σ_{meas}^2) of the various scenarios are reported in Tables 4.1 and 4.2. All the scenarios are simulated with an ensemble size (NMC) of 2000. This number of realizations is computationally

affordable and guarantees a proper description of the dispersion phenomena for the range of log-conductivity variance used here (Bellin et al., 1992; Salandin and Fiorotto, 1998). Smaller ensemble sizes worsen the solution (Crestani et al., 2010), while the improvements obtained with larger NMCs (e.g., NMC=5000) are not significant enough to justify the increased computational cost (see next section 4.3.2).

Table 4.1: Main characteristics of the numerical experiments carried out for the first test case.

	conditioning points *cp*						prior statistics			
	number			spacing (L)			$< Y >$	σ^2_Y	λ (L)	σ^2_{meas}
	x	y	z	x	y	z				
Reference field 1	/	/	/	/	/	/	0.03	0.43	1	/
Scenario 1	33	33	33	0.25	0.25	0.25	-0.50	0.75	1	0.01
Scenario 2	17	17	17	0.50	0.50	0.50	-0.50	0.75	1	0.01
Scenario 3	11	11	11	0.75	0.75	0.75	-0.50	0.75	1	0.01

Table 4.2: Main characteristics of the numerical experiments carried out for the second test case.

	conditioning points *cp*						prior statistics			
	number			spacing (L)			$< Y >$	σ^2_Y	λ (L)	σ^2_{meas}
	x	y	z	x	y	z				
Reference field 2	/	/	/	/	/	/	-0.08	1.59	1	/
Scenario 4	65	33	33	0.25	0.25	0.25	-0.50	2.25	1	0.34
Scenario 5	33	17	17	0.50	0.05	0.50	-0.50	2.25	1	0.34
Scenario 6	22	11	11	0.75	0.75	0.75	-0.50	2.25	1	0.34

The retrieval of the hydraulic conductivity fields in the various simulated scenarios is assessed by means of the root mean square error RMSE, computed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\Psi_{enkf,i} - \Psi_{true,i})^2}{n}}, \tag{4.3}$$

where n is the total number of nodes considered for the calculation and $\Psi_{enkf,i}$ and $\Psi_{true,i}$ are the estimated and the true variable values at the i^{th} node, respectively. However, due to the erratic spatial behavior of hydraulic log-conductivity and to the stochastic characteristics of the inversion approach adopted, it seems appropriate to quantify the error in retrieving the Y field not only pointwise, but also on volumes of finite size. To this aim, an aggregate RMSE for the hydraulic log-conductivity is calculated by considering Y values averaged over small portions of the domain, i.e., cubic subvolumes whose side is chosen equal to an integral scale. Defined as nsv the

number of discretization nodes included in a subvolume, the values averaged over the subvolumes are estimated as

$$\Psi_{true}^{subv} = \frac{\sum_{j=1}^{nsv} \Psi_{true,j}}{nsv} \qquad \Psi_{enkf}^{subv} = \frac{\sum_{j=1}^{nsv} \Psi_{enkf,j}}{nsv} \tag{4.4}$$

and are used in equation (5.2), in which n is now the number of subvolumes considered in the computation of the aggregate RMSE.
We assess also the plume evolution as simulated in the reconstructed Y fields by means of the concentration RMSE.

4.3.2 Sensitivity analyses

We present in this section some preliminary analyses carried out to determine the optimal ensemble size and the number and distribution of particles to be released for the simulation of the tracer test. To this aim, we test our approach in the first reference field (8 $L \times 8$ $L \times 8$ L) with a number of conditioning points *cp* equal to the total number of nodes of the domain discretization. In this case, the size of the EnKF system state vector is maximum and the Y values in each node of the domain are updated. This means that a conditional generation is not necessary after the updates.
Sposito and Dagan (1994) showed that the initial state of a solute plume, as defined by its concentration distribution, affects the subsequent solute evolution due to the velocity field. Thus, the first analysis is carried out to understand how different initial particle distributions in the injection volume can affect the model results. The three cases considered are illustrated in Figure 4.3c. In the first simulation, the solute is described by 1152 particles non-uniformly distributed in the horizontal cross-section (configuration I in Figure 4.3c) while the vertical spacing is $\Delta a = 1/12\ L = constant$). The second simulation is carried out with the same number of injected particles, but with a uniform horizontal distribution (configuration II in Figure 4.3c). In both cases there are the same number of particles (12) for each element of the injection well, ensuring the same initial concentration distribution. In the third simulation the number of uniformly distributed particles is increased to 4992: the vertical spacing is now $\Delta a = 1/16\ L = constant$ with 52 particles within each element of the injection volume.

For all the tests an ensemble of 2000 NMC is used.

Figure 4.4, which shows the comparison between the Y and C RMSE resulting from these assimilation experiments, demonstrates that using a uniform particle distribution gives an improved retrieval of the hydraulic conductivity, while there are no significant differences for the solute concentration. The best performance, both for Y and C, is achieved in the last case, with uniform distribution and the highest number of particles. For this reason, in all the following tests we use as the initial tracer distribution the configuration III of Figure 4.3c.

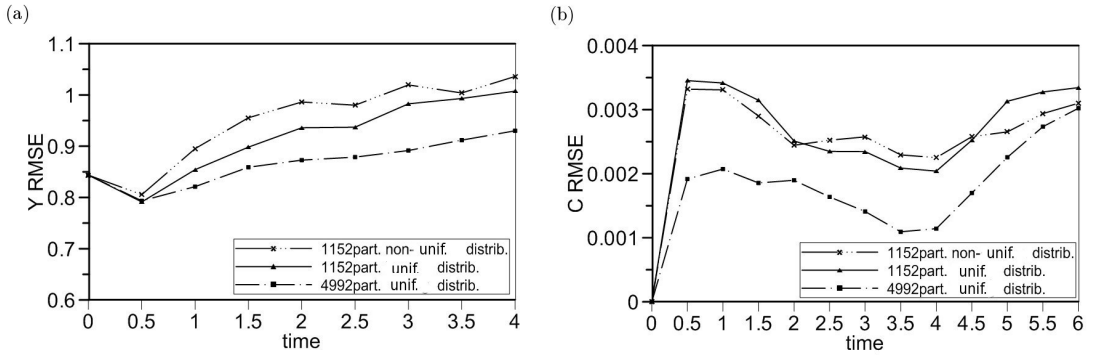


Figure 4.4: Root mean square error of (a) hydraulic log-conductivity and (b) tracer concentration, computed for the whole domain as a function of particle number and distribution.

Note that the simulation of the plume evolution is protracted, using the final Y field, also beyond the last assimilation step, resulting in a C RMSE that increases after $t = 4\ T$. This issue will be discussed in detail in the following section and it is a preliminary clue that the corrections made to the hydraulic conductivity field by the EnKF are somewhat limited to the portion of the domain actually covered by the mean solute passage.

The second analysis concerns the sensitivity of the inversion model to the ensemble size NMC. The Y and C RMSE are compared for two simulations in which the ensemble size are 2000 and 5000 (Figure 4.5). As expected, using $NMC = 5000$ gives better results than using $NMC = 2000$ for the retrieval of the hydraulic conductivity field but with an unaffordable computational effort and with limited differences in the plume reproduction.

Considering the results of these preliminary analyses, all the numerical experiments

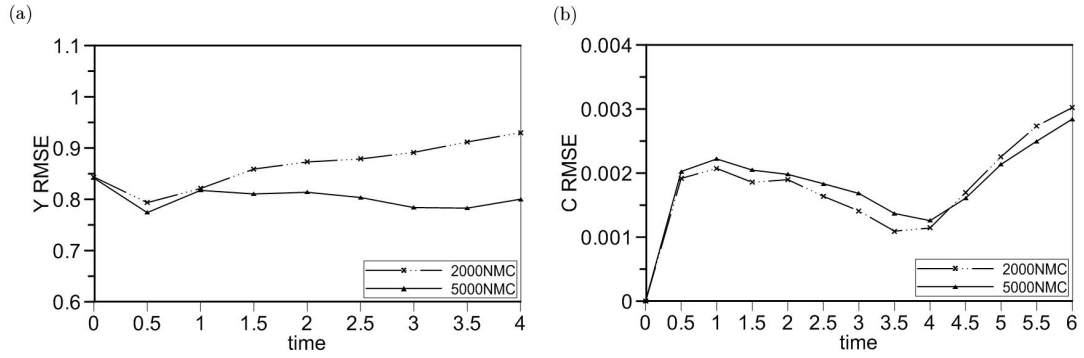


Figure 4.5: Root mean square error of (a) hydraulic log-conductivity and (b) tracer concentration, computed for the whole domain as a function of the ensemble size.

presented in the following sections are carried out with an initial tracer distribution of 4992 particles uniformly distributed and an ensemble size of 2000.

4.3.3 Scenarios with the first reference field

Three scenarios are run with the first reference field ($8\ L \times 8\ L \times 8\ L$), varying only number and spacing of the conditioning points, while the other simulation parameters (prior geostatistical information and measurement error variance σ^2_{meas}) are kept constant (Table 4.1).

Figure 4.6 shows the spatial distribution of the ensemble mean and variance of Y after the fourth and eighth (last) assimilation in scenario 1, for which all Y values are updated (no conditional generation). The expectation of the lnK prior fields, i.e., the Y fields before any update, is uniform, being the ensemble average of a statistically homogeneous field. After each measurement assimilation, the Y values are updated and the zones with different hydraulic log-conductivity are progressively delineated. By comparison with the reference field (Figure 4.2), it can be seen how the inversion model is able to identify the heterogeneities mainly in the portion of the domain "sampled" by the plume, i.e., in the vertical $x - z$ cross section corresponding to the injection well. The ensemble variance of Y, initially uniform and equal to the prior value, shows a gradual decrease at each assimilation, broadly following in time and space the plume evolution. The reduction is stronger in the aquifer portion interested by the plume, where a significant improvement of the Y field has been achieved. This indicates that

the EnKF finds its best estimate in that portion, while, outside the influence area of the plume, the uncertainty is still high. On the other hand, the low variances outside the influence area of the plume, after the last assimilation, can be related to the propagation of the filter inbreeding.

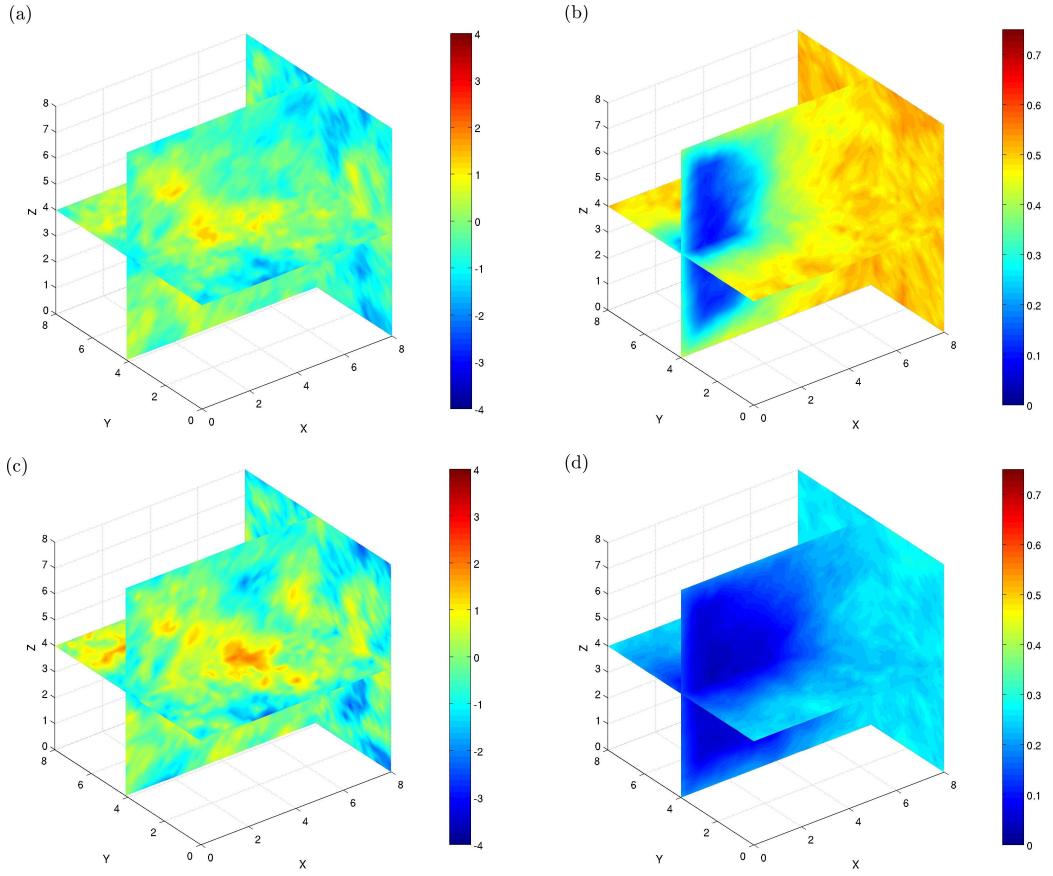


Figure 4.6: Scenario 1: the hydraulic log-conductivity field after (a) the fourth and (c) the last assimilation step and Y ensemble variance after (b) the fourth and (d) the last assimilation step.

The RMSE, as defined in equation (5.2), is used to analyze the difference between the reference and the retrieved Y fields. Figure 4.7 shows the RMSE for scenario 1, as well as for scenarios 2 and 3, which are reported for convenience but will be discussed later on in this section. The results of the open loop simulation, i.e., a tracer test carried out with the prior expected Y field, are also reported. Even though the visual comparison between Figure 4.2 and Figure 4.6 shows a manifest improvement of the estimated Y field at each time step, the RMSE obtained considering the values of the whole domain

exhibits, after an initial decrease, an increasing trend and reaches values larger than the open loop (Figure 4.7a).

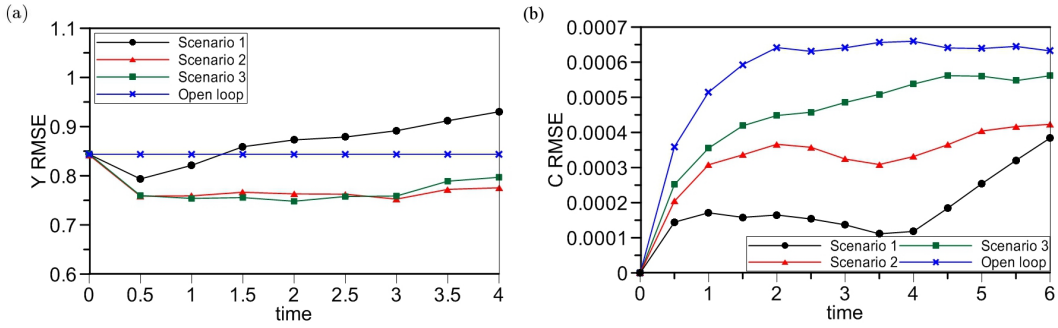


Figure 4.7: Root mean square error of (a) hydraulic log-conductivity and (b) tracer concentration computed pointwise for the whole domain in scenarios 1, 2, 3, and open loop.

However, as we are dealing with an ill-posed inverse problem, larger Y errors do not necessarily imply larger errors in the plume reproduction. To highlight this aspect, the concentration root mean square error (C RMSE), is computed according to equation (5.2), where $\Psi_{enkf,i}$ and $\Psi_{true,i}$ are the concentration values in the $i-th$ node of the solute cloud in the retrieved (obtained after the last assimilation) and in the true fields, respectively. The C RMSE analysis is extended until $t = 6\ T$, to underline its behavior after the time corresponding to the last assimilation ($t = 4\ T$). The results, reported in Figure 4.7b, show a decreasing RMSE trend until the time corresponding to the last update, followed by a significant increase, confirming that the correction of the Y field is likely to be limited to the portion of the domain sampled by the plume. As previously stated, we expect that laterally only a portion of the domain (where the plume is presumed to move) is affected by the concentration measurements and thus effectively updated. To investigate this aspect, a local analysis is carried out in a limited domain portion with an extension of three integral scales (we recall that $\lambda = 1\ L$) along x, from $x = 1.125\ L$ to $x = 4.125\ L$ (immediately downstream of the well), one integral scale along y, from $y = 3.625\ L$ to $y = 4.625\ L$, and six integral scales (the height of the well) along z, from $z = 1.125\ L$ to $z = 7.125\ L$ (Figure 4.3). This zone is subdivided into three subzones (defined as A1, A2, A3), each with an extension of one integral scale along the main flow direction, to evaluate in detail the effect of the evolving plume on

the updating of the Y values. We denote as "A" the union of A1, A2, A3.

Figure 4.8a reports the Y RMSE computed for each subzone. To underline the relation between the RMSE in each subzone and the passage of the plume, Figure 4.8a displays also the longitudinal component of the plume first moment R_1, i.e., the location of the plume center of mass along the x direction, and a measure of the tracer spreading S_{11} in the form of error bars (see equations 2.10 and 2.11). As a consequence of the first two updates, a general RMSE decrease is manifest in all the subzones. At $t = 1\ T$, the plume is still inside the A1 subzone: while the reduction in A1 is directly affected by the measured data, the RMSE decrease in A2 and A3 is probably related to the effects of the correction operated by the EnKF on the initially wrong Y ensemble mean. The subsequent updates, when the plume leaves A1, have no significant effects on the RMSE of subzone A1, which remains almost constant. For both subzones A2 and A3 it is interesting to note that after the initial decrease, the corresponding RMSE remains almost constant or tends to increase until the plume does not affect the specific subzone. In the result analysis, it has to be considered that the solute cloud extension is here approximated by the centroid position and the error bar is proportional to the standard deviation. Overall, in comparison with the results of Figure 4.7 we can note a marked improvement of the Y estimation in the portion of the domain directly sampled by the tracer and a definite temporal trend of the RMSE in the three subzones that broadly follows the plume evolution.

If we consider subzones with extension $2\ L$ along y, i.e., laterally doubled with respect to the previous cases, we obtain the results shown in Figure 4.8b. Even though the improvement of the Y field estimation is still manifest, the root mean square error appears to be larger than the one calculated previously and the link between the RMSE trend and the plume evolution is smoothed. Due to the lack of the Lagrangian velocity correlation, the plume spreading in the transverse direction is limited and the efficacy of the retrieved procedure in the y direction is limited to approximately one integral scale.

Above findings are confirmed by an analysis of the plume evolution when the solute is injected in the two alternative positions Pos1 and Pos2 (Figure 4.3), laterally shifted

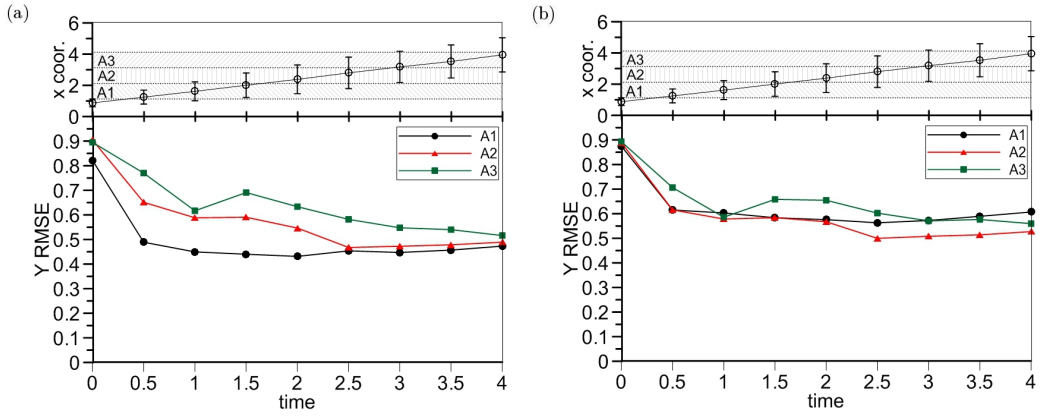


Figure 4.8: Scenario 1: root mean square error of hydraulic log-conductivity computed pointwise in subzones A1, A2, and A3. Top graphs show the location of the true plume center of mass along the x direction (R_1) and the spreading (error bars equal to $\pm 2\sqrt{S_{11}}$). The extension along y of the subzones is (a) 1 L and (b) 2 L.

with respect to the Pos0 position used for the assimilation procedure. The plumes resulting from injections carried out in Pos1 and Pos2, both in the reference and in the retrieved Y fields, are compared in terms of the concentration RMSE (Figure 4.9). With respect to the plume evolution starting from Pos0, the error is significantly larger for the tracer released in Pos1 and Pos2. After the last assimilation time $t = 4\ T$, the RMSE related to the injection in Pos0 gradually increases, approaching the RMSE values of the plumes injected in Pos1 and Pos2. This behavior reflects the improvement of the retrieved Y field, which is related to the dispersion characteristics: while the plume stretches out in the longitudinal direction, affecting also a portion of the domain beyond the centroid position, the concentration drops suddenly along the transverse direction, with a consequent lack of lateral data.

Overall, using the proposed approach we are able to achieve a satisfactory retrieval of the hydraulic log-conductivity field. Even if the RMSE computed using nodal Y values gives a satisfactory error measure, it seems to be more appropriate to evaluate the aggregate Y RMSE as defined by equations (5.2) and (4.4) with reference to unitary subvolumes (1 $L \times 1\ L \times 1\ L$), corresponding to one cubic integral scale. Also this RMSE evaluation refers to zone A and subzones A1, A2, and A3, and the results are reported in Figure 4.10. Compared to the pointwise RMSE (Figure 4.8a), the aggregate

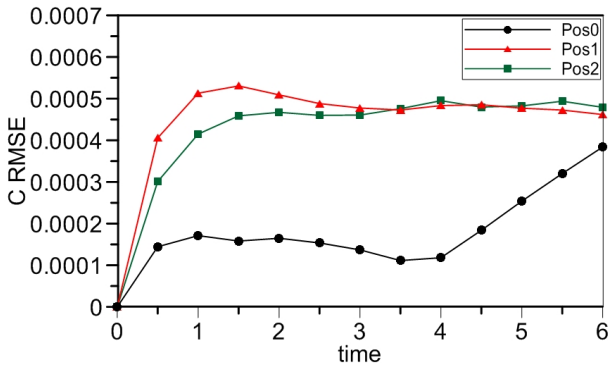


Figure 4.9: Scenario 1: concentration root mean square error computed for simulations using the retrieved hydraulic conductivity field when the solute is injected in Pos0, Pos1, and Pos2.

Y root mean square error reaches smaller values, as expected. The time evolution of the aggregate errors is still manifest as the plume travels through the subzones but the effect of the first two updates is highlighted for all the subzones, including A3, which is sampled by the tracer only at late times. The aggregate Y RMSE is also displayed for the entire zone A and reveals an average behavior between the three subzones.

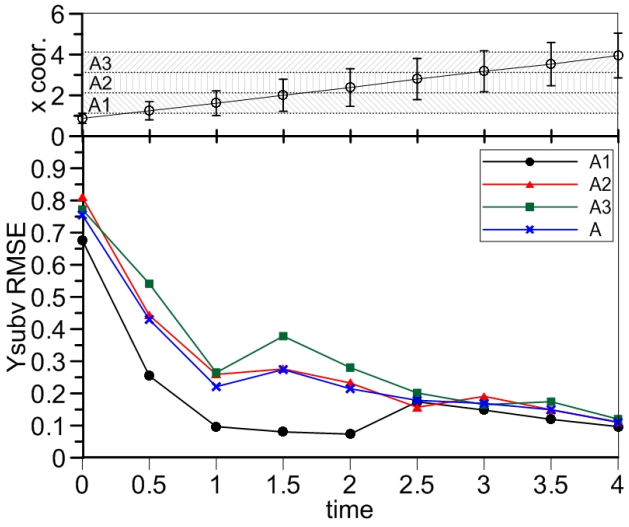


Figure 4.10: Scenario 1: aggregate root mean square error of hydraulic log-conductivity computed in zone A and subzones A1, A2, and A3. The top graphs shows the location of the true plume center of mass along the x direction (R_1) and the spreading (error bars equal to $\pm 2\sqrt{S_{11}}$).

These results indicate that inherent uncertainties in the inversion procedure affect the

reproduction of pointwise Y values, so that a more suitable parameter to evaluate the retrieval is the RMSE aggregated over subvolumes representative of the Y integral scale. In scenario 2 only one out of two nodes is included in the system state vector (Table 4.1) and the concentration measurements are sampled only in these nodes. Therefore, these nodes are located 0.5 L (half integral scale) apart, resulting in a number of conditioning points $cp = 17 \times 17 \times 17 = 4913$ and reducing the system state dimension to seven times less than scenario 1, where $cp = 35937$ (equation (5.1)). Figure 4.11 shows the spatial distribution of the ensemble mean and variance of Y after the fourth and last (eight) assimilation in scenario 2 (compare with Figure 4.6 for scenario 1). The conditional generation of the hydraulic log-conductivity values significantly limits the loss of Y ensemble variance inherent in the EnKF procedure. The trade-off is that the features of the estimated Y field are not as well delineated as in scenario 1 and, as a consequence, the concentration field may be reproduced less satisfactorily. The time evolution of the Y and C (pointwise) RMSE of scenario 2 are reported in Figure 4.7 in comparison with other scenarios. While the C RMSE in scenario 2 is larger than in scenario 1, the Y RMSE is smaller: this fact can be ascribed to a smoothed spatial distribution of the hydraulic log-conductivity but can be related also to the limited performance of the pointwise RMSE evaluated in the whole domain. The loss of ensemble variance can be effectively counteracted by updating only a fraction of the parameters and using these updated parameters to generate the remaining fraction. Figure 4.12 shows the aggregate Y RMSE within zone A and subzones A1, A2, and A3 for scenario 2. Compared with scenario 1 (Figure 4.10), the correspondence between the plume evolution and the RMSE trend is now less evident, even though the RMSE trend for each subzone suggests again that the updates somehow "follow" the tracer.

In scenario 3 the hydraulic log-conductivity of one out of three nodes is included in the system state vector and, as a consequence, the measurement vector dimension is further reduced. The relevant nodes are located 0.75 L apart, i.e., 3/4 of an integral scale, resulting in a number of conditioning points (system state dimension) $cp = 11 \times 11 \times 11 = 1331$, 27 times less than in scenario 1. The spatial distribution of the ensemble mean and variance of Y for scenario 3 are shown in Figure 4.13 after the

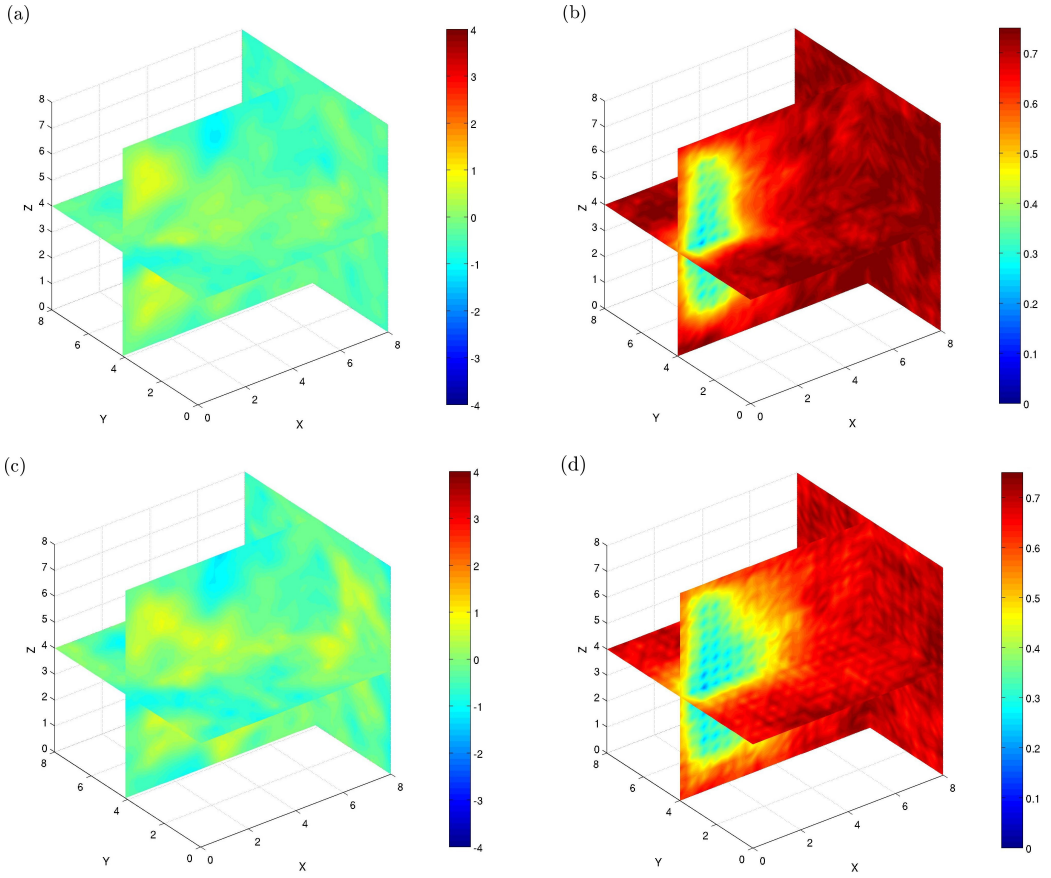


Figure 4.11: Scenario 2: the hydraulic log-conductivity field after (a) the fourth and (c) the last assimilation step and Y ensemble variance after (b) the fourth and (d) the last assimilation step.

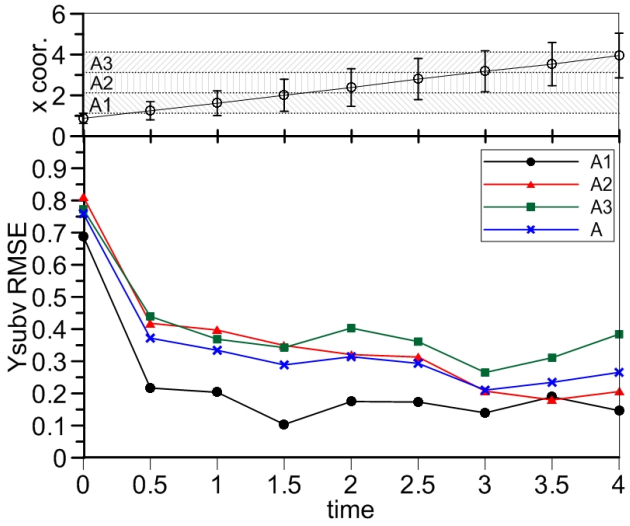


Figure 4.12: Scenario 2: aggregate root mean square error of hydraulic log-conductivity computed in zone A and subzones A1, A2, and A3. The top graph shows the location of the true plume center of mass along the x direction (R_1) and the spreading (error bars equal to $\pm 2\sqrt{S_{11}}$).

fourth and last (eighth) assimilation. With respect to scenario 2, the estimated Y field is further smoothed and the loss of ensemble variance is slowed down, as a result of the conditional generation of a larger number of nodes. As expected, the pointwise Y RMSE computed over the whole domain (reported in the above mentioned Figure 4.7a, does not significantly differ from scenario 2, but the reproduction of the concentration field is further worsened, even though the C RMSE is still smaller than the one of the open loop simulation (Figure 4.7b. Figure 4.14 shows the Y aggregate RMSE within zone A and subzones A1, A2, and A3. The RMSE trend is no longer manifestly correlated to the plume position, but, on average, the Y RMSE decrease in A1 precedes the one in A2, and the RMSE decrease in A2 precedes the one in A3.

Overall, we can conclude that, in our retrieval procedure, limiting the update to a fraction of the total number of parameters (Y values) and estimating the remaining ones through a conditional generation have the advantage to reduce the filter inbreeding. On the other hand, the convergence toward the optimal estimation is slowed down and the tracer plume is likely to be estimated less satisfactorily (Figure 4.7b.

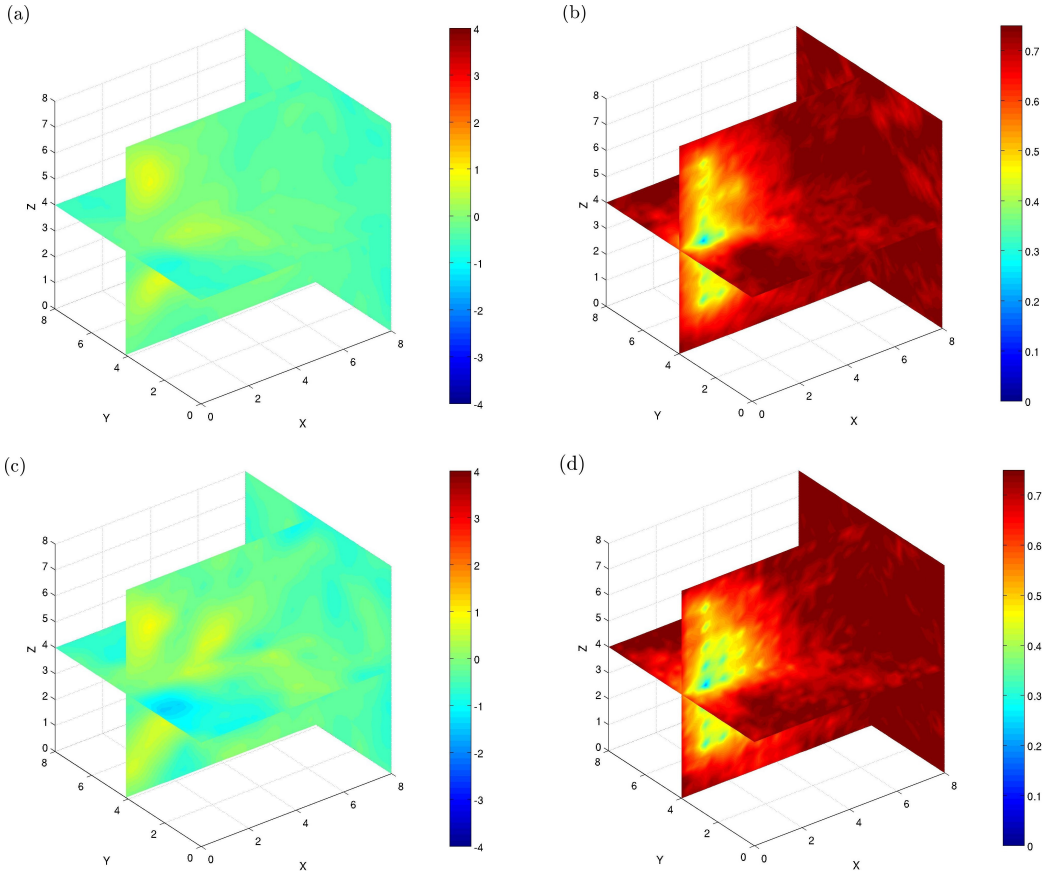


Figure 4.13: Scenario 3: the hydraulic log-conductivity field after (a) the fourth and (c) the last assimilation step and Y ensemble variance after (b) the fourth and (d) the last assimilation step.

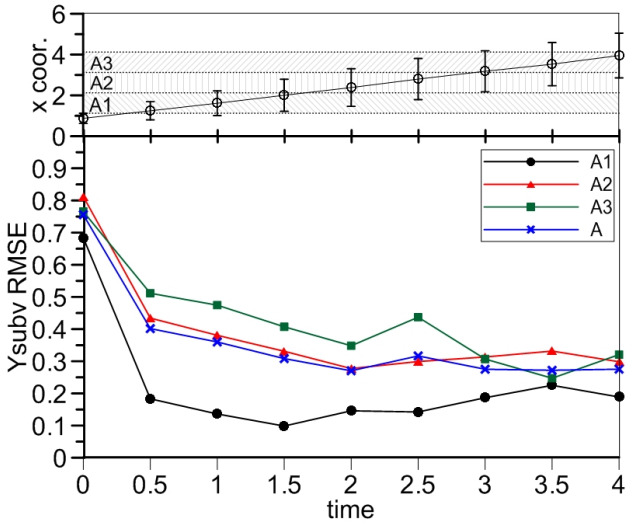


Figure 4.14: Scenario 3: aggregate root mean square error of hydraulic log-conductivity computed in zone A and subzones A1, A2, and A3. The top graph shows the location of the true plume center of mass along the x direction (R_1) and the spreading (error bars equal to $\pm 2\sqrt{S_{11}}$).

4.3.4 Scenarios with the second reference field

The proposed approach is tested also with respect to the second reference field (16 $L \times 8\ L \times 8\ L$) with three numerical experiments analogous to those carried out for the first reference field. Aside from the differences between the two reference fields that are described in section 5.2.1, another distinction is that the σ^2_{meas} value assigned to the measurements during the simulations is now increased (Table 4.2). In fact, it is reasonable to expect that a real ERT inversion applied to a more dispersed plume (as the one resulting in this second test case) would yield concentration data affected by more uncertainty, due to the inherent proneness of classical electrical inverse methods to overdispersion (e.g., Camporese et al., 2011). In choosing the value of σ^2_{meas} since the developed scenarios are synthetic, we decide to maintain the same level of confidence of the model prediction with respect to the measurements as the scenarios in the first reference field, and thus to prevent the first updates from being overestimated. This choice can be explained by considering equation (2.30), rewritten in the scalar case:

$$y_{upd}(t) = y(t) + \frac{P_e}{P_e + R_e}(z(t) - H\, y(t)), \tag{4.5}$$

where the symbols have the same meaning as in equation (2.30) but now represent scalar quantities. The Kalman gain is now $P_e/(P_e + R_e)$ where P_e and R_e are the variances of the system state and the measurement error, respectively. In our multi-dimensional case, increasing the prior value of σ^2_Y and keeping σ^2_{meas} unchanged corresponds to an increase of P_e and hence of the Kalman gain, resulting in augmented updates. The new σ^2_{meas} is thus defined assigning to the ratio $\sigma^2_K/\sigma^2_{meas}$ the same value as in the first test case, being σ^2_K the hydraulic conductivity variance which is directly related to the concentration evolution and equal to $(e^{\sigma^2_Y} - 1)e^{2\langle Y\rangle+\sigma^2_Y}$. We deem this choice better than keeping constant the ratio $\sigma^2_Y/\sigma^2_{meas}$, which may be a reasonable approximation only for small σ^2_Y values.

Three numerical experiments corresponding to scenarios 4, 5, and 6 are carried out, and they are designed to obtain approximately the same travel times of the tracer plume center of mass. The Y ensemble average and variance fields are not shown, being their general trends similar to the ones obtained in the first test case, and the results are only analyzed in term of RMSE.

For comparison with Figure 4.7, in Figure 4.15 the pointwise Y RMSE computed for the whole domain in scenarios 4, 5, 6, and the open loop is shown. Scenarios 5 and 6, reported for convenience in the same Figure, will be discussed later on in this section. Figure 4.15a shows that scenario 4, compared with the results of scenario 1 (Figure 4.7a, which is the corresponding scenario in the first test case (in both cases no conditional generation is used), exhibits values always smaller than the initial (open loop) Y RMSE.

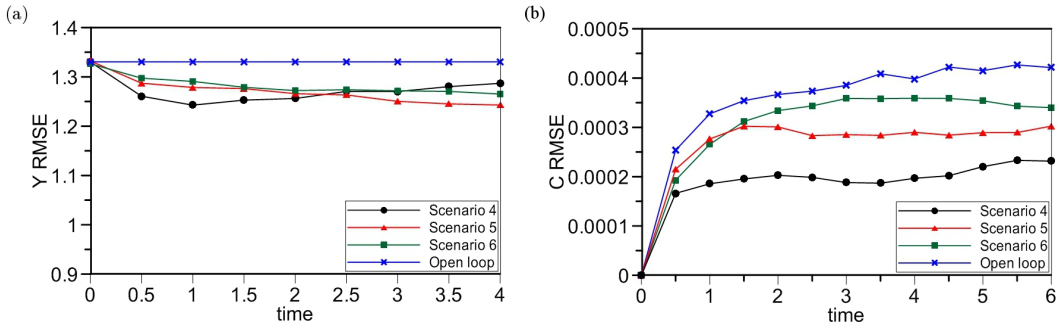


Figure 4.15: Root mean square error of (a) hydraulic log-conductivity and (b) tracer concentration computed pointwise for the whole domain in scenarios 4, 5, 6, and open loop.

As the corresponding Figure 4.7b, Figure 4.15b shows the concentration RMSE obtained by repeating the tracer injection in Pos0 (Figure 4.3) in the reconstructed Y field. In this scenario, the C RMSE increases after $t = 4\ T$ but the trend is less evident than in scenario 1, because of the greater plume spreading that affects a larger portion of the domain downstream of the plume centroid.

The local analysis on the hydraulic log-conductivity RMSE is carried out in the same portion of the domain (zone A with subzones A1, A2, and A3) as in the first reference field but, in this case, the longitudinal plume spreading is more manifest, as a result of the increased heterogeneity.

Figure 4.16 shows the aggregate Y root mean square error for scenario 4 in zone A and subzones A1, A2, and A3. The effect of the plume evolution is much less apparent if compared to the analogous scenario 1, even though it has to be remarked that subzone A3 starts from a RMSE lower than A1 and A2. The improvement in A3 is thus marginal for the first updates, while RMSE in A1 is smaller than in A2. This is in agreement with the results observed in scenario 1, although in this case the higher dispersion of the plume makes more difficult to interpret the RMSE trend. The plume spreading around the centre of mass grows with σ_Y^2, the assimilated data affecting a portion of the domain larger than in the first aquifer, and the skewness of the tracer cloud smooths away the effect due to the passage of the center of mass across different aquifer portions. For this reason, the reduction of the Y RMSE due to the passage of the plume center of mass across different zones is less apparent. Moreover, the positive skewness of the concentration spatial distribution raises also issues of non-Gaussianity for the EnKF so that the updates are affected by further approximations, in addition to those related to the finite size of the ensemble.

This further contributes to obscure the effect on the time evolution of Y RMSE that can be instead better observed in a less heterogeneous case.

In scenarios 5 and 6 only one out of two and one out of three nodes are considered for the update, analogously to scenarios 2 and 3, respectively.

The pointwise Y RMSE calculated in the whole domain (reported in the previously discussed Figure 4.15a does not show relevant differences between the two scenarios,

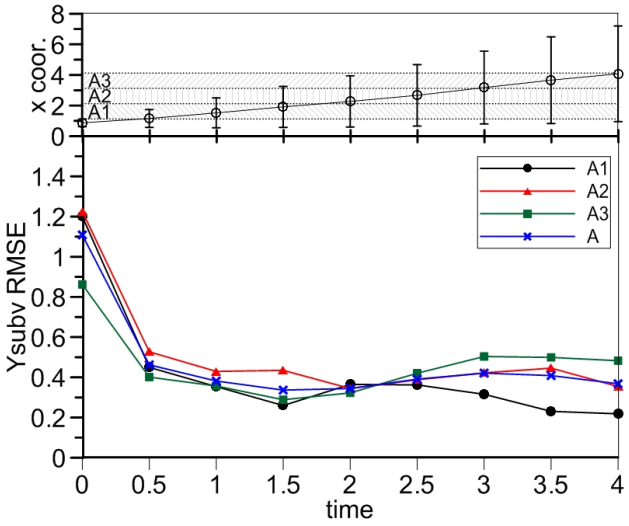


Figure 4.16: Scenario 4: aggregate root mean square error of hydraulic log-conductivity computed in zone A and subzones A1, A2, and A3. The top graph shows the location of the true plume center of mass along the x direction (R_1) and the spreading (error bars equal to $\pm 2\sqrt{S_{11}}$).

being the values smaller than the one of the open loop in both cases. With reference to the C RMSE behavior, its value increases as the number of conditioning points is reduced (Figure 4.15b, as in the first test case.

Figures 4.17 and 4.18 show the time evolution of the aggregate Y RMSE in zone A and subzones A1, A2, and A3 for scenarios 5 and 6, respectively. The results are broadly consistent with those of scenario 4 (compare Figure 4.16), except for a progressive smoothing due to the reducing of the updated Y values. This is again in agreement with the results obtained in the counterparts of the first reference field, i.e, scenarios 2 and 3 (Figures 4.12 and 4.14 versus 4.10). The C RMSE for scenarios 5 and 6 (Figure 4.15b does not exhibit the same increasing trend that is manifest in scenarios 2 and 3 after $t = 4\ T$ (Figure 4.7b. Again, this effect is probably due to the large spreading that affects the experiments in the second reference field.

Overall, it seems that the adverse effects of poor delineation of the estimated Y field related to the conditional generation reduce as the heterogeneity increases, as shown by less manifest RMSE differences among scenarios 4, 5, and 6 compared to the differences among scenarios 1, 2, and 3. At the same time, the main advantages, i.e., the filter

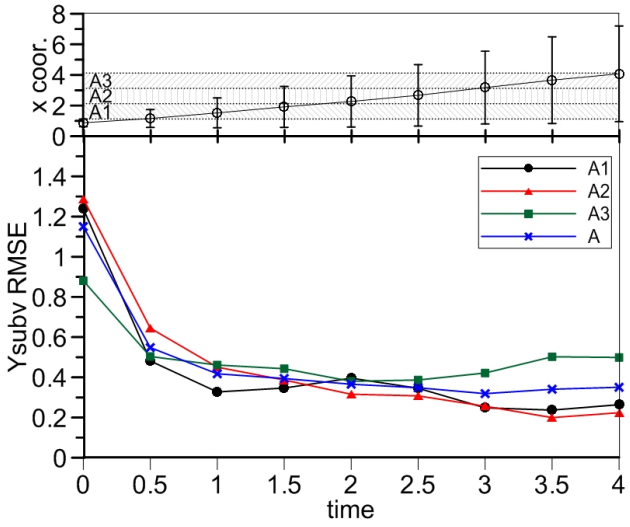


Figure 4.17: Scenario 5: aggregate root mean square error of hydraulic log-conductivity computed in zone A and subzones A1, A2, and A3. The top graph shows the location of the true plume center of mass along the x direction (R_1) and the spreading (error bars equal to $\pm 2\sqrt{S_{11}}$).

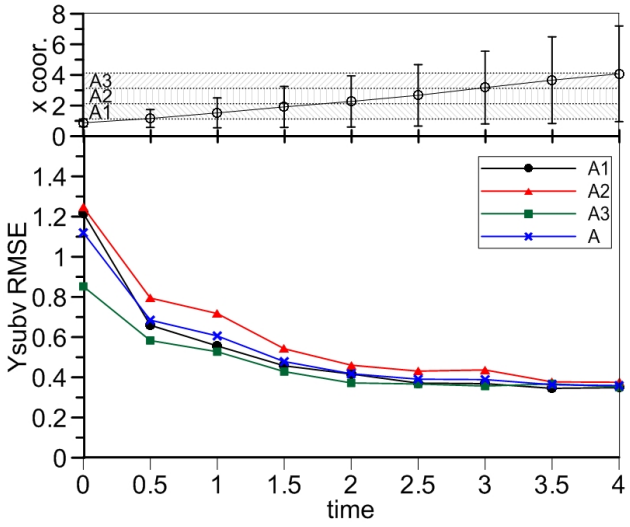


Figure 4.18: Scenario 6: aggregate root mean square error of hydraulic log-conductivity computed in zone A and subzones A1, A2, and A3. The top graph shows the location of the true plume center of mass along the x direction (R_1) and the spreading (error bars equal to $\pm 2\sqrt{S_{11}}$).

inbreeding and memory requirement reduction, are still maintained.

4.4 Final Remarks

It is investigated the theoretical and numerical behavior of a hydrological inversion procedure to assess heterogeneous hydraulic conductivity fields at the local scale from the knowledge of spatio-temporal concentration distributions obtained by tracer tests. In particular, in order to contrast filter inbreeding and to reduce the amount of memory required by the EnKF to store and manage the system state vectors and covariance matrices, we investigate also the use of conditional generation to complete the reconstruction of hydraulic log-conductivity fields starting from a limited and properly selected number of updated Y values.

After a preliminary sensitivity analysis we elected to use an ensemble size of 2000 Monte Carlo realizations and a uniform density of 3328 injected particles per cubic integral scale as a good compromise between computational effort and quality of results.

For each synthetic aquifer, three different scenarios are analyzed, in which the number of selected nodes where the C values are assimilated and Y values are updated, is progressively reduced.

In all the scenarios the reconstructed hydraulic conductivity field results in a tracer plume evolution with a concentration root mean square error (C RMSE) smaller than the one obtained in the open loop experiment, i.e., using the expectation of lnK prior fields. Consistently with previous results (e.g., Hendricks Franssen et al., 2003; Camporese et al., 2011), we notice that the retrieval of the hydraulic conductivity is limited to a portion of the domain that is directly sampled by the tracer plume. As demonstrated by a local analysis of the Y RMSE, taking into account the whole domain results in a RMSE that is always larger than the one calculated using a limited zone covered by the plume, due to the rapid decay of correlation between C and Y outside the evolving solute cloud. Because of its spatial distribution this phenomenon is more manifest in the transverse direction than in the longitudinal one. The analysis of the root mean square error carried out on Y values averaged over cubic subvolumes with side equal to an integral scale suggests that, for our inverse modeling framework, an

aggregate RMSE evaluation is more suitable than a pointwise approach, as it can better assess the retrieval of the blocks with different hydraulic log-conductivity values.
Reducing the number of updated points results in a smoothed behavior of the Y RMSE and in a less satisfactory reproduction of the tracer plume evolution. On the other hand, a limited number of nodes for the assimilation and the update has the advantage to slow down the understimation of ensemble variance that commonly affects the ensemble Kalman filter.
A larger Y variance, increasing the plume dispersion, limits the deterioration of the Y field retrieval that affects the conditional procedure. Moreover, the procedure maintains the main advantages related to the filter inbreeding and the memory requirement reduction.

Chapter 5

Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation

5.1 Introduction

Given the importance that ensemble Kalman filter (EnKF) and ensemble smoother (ES) are acquiring as parameter estimation modeling tools in groundwater hydrology, there is the need to investigate in more detail their capabilities and the theoretical implications related to their use for optimal estimation of only system states (Nowak, 2009, e.g.,). The objective of this work is thus to compare the EnKF and ES capabilities to retrieve the hydraulic conductivity spatial distribution in a groundwater flow and transport modeling framework. Moreover, since a fundamental hypothesis for the application of Kalman filter-based methods is that all the variables must be distributed as a Gaussian probability density function (pdf), the issues related to various transformations of the relevant pdfs are analyzed. EnKF and ES are here implemented in the same Lagrangian transport modeling framework already described, in order to estimate the

hydraulic conductivity field by assimilating concentration measurements derived from a tracer test in a two-dimensional synthetic aquifer (the full spatio-temporal evolution of the solute plume is known).
This problem is solved by considering a system state in which we retain only the model parameters:

$$\mathbf{y}^j = [Y_1, \ldots, Y_n]^j = \boldsymbol{\alpha}^j, \tag{5.1}$$

where $Y_1, \ldots, Y_n$ are the log-transformed hydraulic conductivity values ($Y = lnK$) at the n nodes discretizing the domain.

5.2 Numerical experiments

5.2.1 Model setup

A two-dimensional reference Y field is taken into account to compare the proposed inversion models in a number of numerical experiments. The domain has dimensions of 8 $L \times 8$ L, where L is an arbitrary and consistent unit length, and it is discretized along each directions into $L/4$ sided cells, for a total of $33 \times 33 = 3297$ corresponding nodes. By assuming $Y = ln(K)$, the multivariate normal Y distribution of the reference field is the result of a single unconditional generation fitting to an isotropic exponential covariance model with spatial mean $\langle Y \rangle = 0.35$, variance of $\sigma_Y^2 = 0.42$, and integral scale $\lambda = 1$ L. The random function Y is generated by an improved sequential Gaussian simulation algorithm (Baú and Mayer, 2008). The flow field is simulated using a standard finite volume solver at steady state with appropriate boundary conditions that ensure a constant mean gradient. Dirichlet boundary conditions are applied at $x = 0$ L ($h = 100.0$ L) and at $x = 8$ L ($h = 95.2$ L), while Neumann no-flow boundary conditions are imposed along the remaining sides of the domain. A graphic representation of the reference field is given in Figure 5.1. A tracer test is simulated by assuming an instantaneous solute injection with initial transversal size of 6 L (from $y = 1$ L to $y = 7$ L) and longitudinal size of 0.5 L (centered in $x = 0.875$ L). The solute is simulated by 16983 particles uniformly distributed, with the particle trajectories computed by means of the Pollock's particle tracking post-processing algorithm

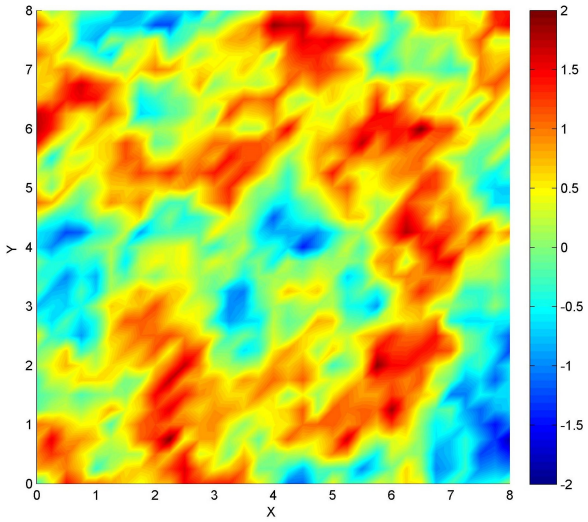


Figure 5.1: Spatial distribution of log-transformed hydraulic conductivity in the reference field.

and the concentration computed according to equation (2.9) every 0.5 T, where T is any consistent time unit. Figure 5.2 shows the plume evolution in the reference field at $t = 2\ T$ and $t = 4\ T$.

In this work, measurements of concentration are the data assimilated during the inversion procedures. In particular, we include in the measurement vectors all the concentration data greater than zero, in order to use all the available information on the plume evolution. The performances of EnKF and ES in retrieving the Y fields are compared to one another in a number of different scenarios described in Table 5.1. In these scenarios we also analyze the implications of the concentration non-Gaussianity through various marginal transformations of the pdf. The prior geostatistical parameters and the measurement uncertainty are kept constant in all scenarios, as the objective is to study the issues related to the effect of the pdf of the model variables on the EnKF and the ES performances in terms of parameter estimation. In scenario 1 the concentrations are assimilated without any transformation, i.e., with their original pdf, whereas in scenarios 2, 3, and 4 different pdf transformations are used to evaluate how they affect the parameter estimation. All scenarios are simulated using an ensemble size of 2000.

It should be noted that the ensemble of prior Y fields is synthetically generated by the

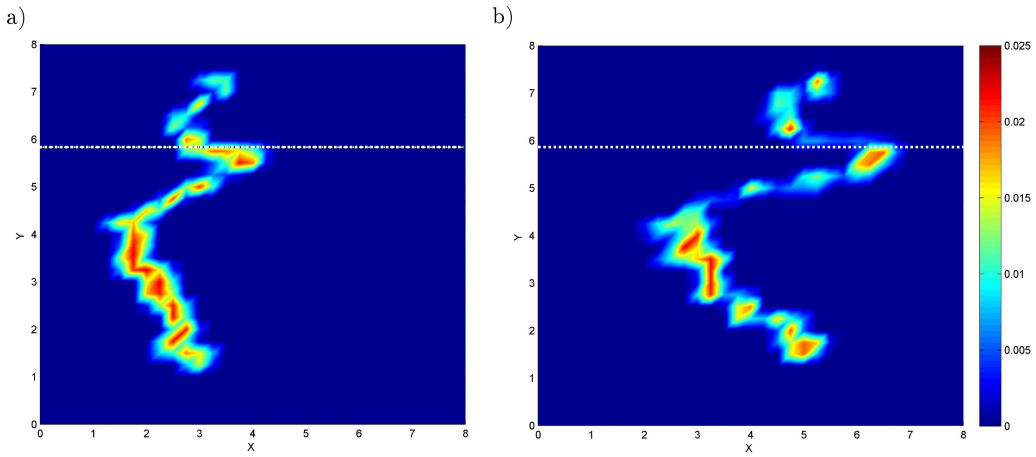


Figure 5.2: Plume evolution in the reference field at a) $t = 2\ T$ and b) $t = 4\ T$. The color bar denotes dimensionless concentration C. Dotted lines show the direction along which the cross-correlation structure between Y and C is evaluated.

Table 5.1: Description, prior geostatistical parameters, and measurement errors of the numerical experiments. [a] Normal score transform applied independently to the ensemble of concentration data in each assimilation node. [b] Normal score transform applied to the ensemble of concentration data over all assimilation nodes.

			prior statistics			
	assimilated data	technique	$< Y >$	σ_Y^2	$\lambda\ (L)$	CV
Reference field	/	/	0.03	0.43	1	/
Scenario 1a	C	EnKF	-0.50	0.75	1	0.01
Scenario 1b	C	ES	-0.50	0.75	1	0.01
Scenario 2a	lnC	EnKF	-0.50	0.75	1	0.01
Scenario 2b	lnC	ES	-0.50	0.75	1	0.01
Scenario 3a	NST(C)[a]	EnKF	-0.50	0.75	1	0.01
Scenario 3b	NST(C)[a]	ES	-0.50	0.75	1	0.01
Scenario 4a	modNST(C)[b]	EnKF	-0.50	0.75	1	0.01
Scenario 4b	modNST(C)[b]	ES	-0.50	0.75	1	0.01

same algorithm used to create the true field and that the application of the Lagrangian transport model in the true field yields exactly the true concentration distribution. The knowledge of the true state allows us to select the concentration measurements used during the assimilation and to evaluate the performance of EnKF and ES with respect to a known reference.

The estimate of the hydraulic conductivity fields in the various simulated scenarios is assessed by means of the root mean square error ($RMSE$), computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{sim,i} - Y_{true,i})^2}{n}}, \tag{5.2}$$

where n is the total number of nodes of the discretized domain, $Y_{sim,i}$ is the ensemble mean of the Y values estimated at the i^{th} node, and $Y_{true,i}$ is the true Y value at the i^{th} node.

Finally, we also assess the plume evolution as simulated in the reconstructed Y fields by means of the concentration $RMSE$, whose formulation is analogous to that of Y.

5.2.2 Results

5.2.2.1 Scenarios with untransformed concentration pdfs

In scenarios 1a and 1b, the concentration values are assimilated in the update procedure without modifying their original pdf, for both the EnKF and the ES. The spatial distributions of Y resulting from the inversions are reported in Figure 5.3. The comparison between the retrieved fields and the reference field (Figure 5.1) shows that the EnKF is quite effective in estimating the lnK field, whereas the ES performs rather poorly.

In these scenarios, both EnKF and ES are affected by the approximations related to the Gaussian assumption of equation (2.30), as the tracer concentration pdf at early travel times departs significantly from the normal pdf (e.g., Salandin and Fiorotto, 1998). Nevertheless, EnKF can handle these approximations better than ES, as, at each update, the realizations are steered toward the true solution and the Gaussian increments of the ensemble members lead to an approximately Gaussian ensemble dis-

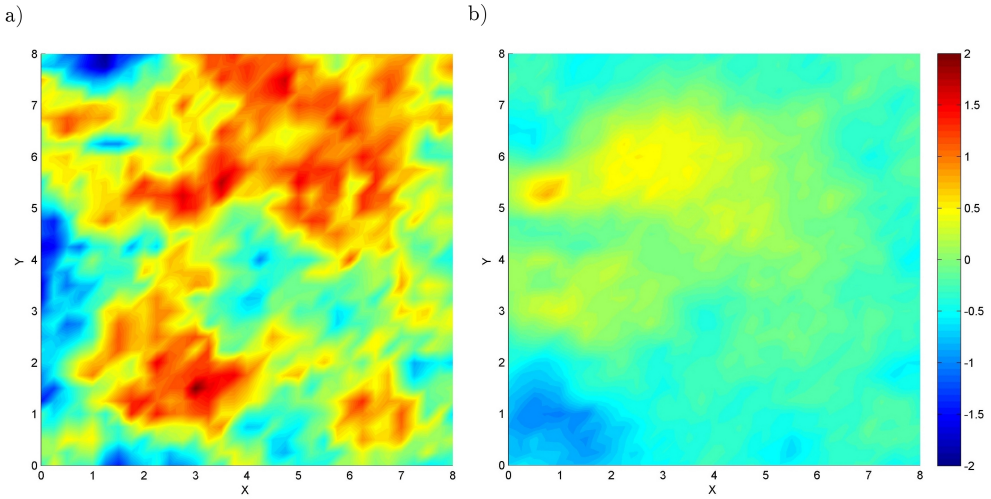


Figure 5.3: Scenario 1: log-transformed hydraulic conductivity field retrieved by a) EnKF and b) ES assimilating untransformed concentration data.

tributed around the true solution. This property of the sequential updating is not exploited in the ES, where realizations evolve freely until the end of the simulation, exacerbating the effects related to non-Gaussian ensemble distributions. A further explanation for the good performance of the EnKF can be found in our recursive application, through which, at each assimilation step, the plume is restarted with the same initial concentration in the updated Y fields. This procedure is analogous to the advanced first-order second-moment (AFOSM) method adopted in risk analysis (Yen et al., 1986). As in AFOSM, also in our EnKF application the lack of Gaussianity is overcome by approaching recursively the solution with improved values of the estimator (the Kalman gain) corresponding to improved estimates of the solution itself.

5.2.2.2 Scenarios with log-transformed concentration pdfs

In order to evaluate if a proper rearrangement of the concentration pdf can improve the results, in scenarios 2a and 2b we assimilate log-transformed concentration values. Although previous analyses highlight that Beta-type pdfs can be effectively used to reproduce the concentration distribution (Caroni and Fiorotto, 2005), the log-normal pdf is easier to handle and often represents a reasonable assumption (e.g. Bellin et al., 1994; Zhang et al., 2000).

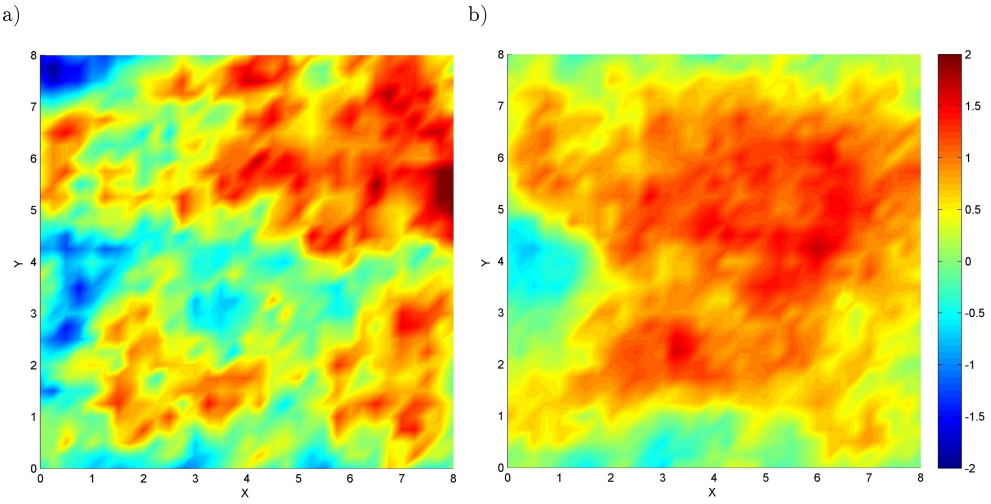


Figure 5.4: Scenario 2: log-transformed hydraulic conductivity field retrieved by a) EnKF and b) ES assimilating log-transformed concentration data.

Figure 5.4 shows the Y fields estimated by EnKF and ES when assimilating log-transformed concentration measurements. The EnKF produces again a good reproduction of the true hydraulic conductivity field, with only small differences with the results of scenario 1a, while the ES confirms its difficulty in retrieving the Y spatial distribution. Nonetheless, the comparison between scenarios 1b and 2b shows a slight improvement of the ES solution, indicating that the ES is more sensitive than the EnKF to the pdf of the assimilated variable.

5.2.2.3 Scenarios with normal score-transformed concentration pdfs

Since the log-transformation of concentration values does not ensure a normal pdf in all cases, in scenarios 3a and 3b another type of transformation is applied to the concentration data. Here a Gaussian pdf of the C values is obtained by applying a normal score transform (NST) (Zhou et al., 2011). The NST is a tool through which any cumulative probability distribution function (CDF) $F(x)$ is mirrored to the standard Gaussian CDF $G(y)$. In other words, the generic variable x of the $F(x)$ distribution can be transformed into the corresponding normally distributed variable y through the relation $F(x) = G(y)$, i.e., $y = G^{-1}[F(x)]$. In this case $x = C$ and a CDF is built for each node of the domain with the ensemble of C values simulated by the model, using

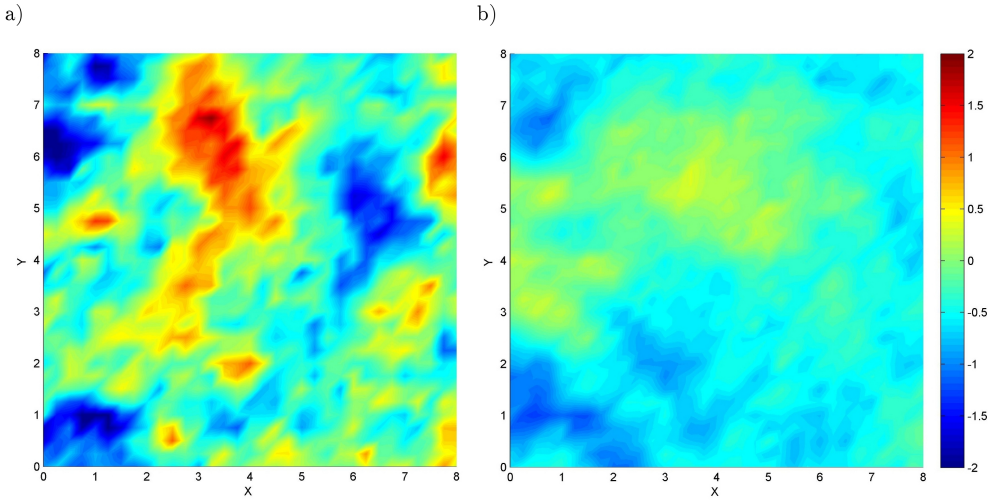


Figure 5.5: Scenario 3: log-transformed hydraulic conductivity field retrieved by a) EnKF and b) ES assimilating normal score-transformed concentration data.

the Hazen formula $F(C) = (i - 0.5)/NMC$, where $i = 1, \ldots, NMC$ is the rank of the concentration values after sorting the data in ascending order. For each node, different values of C are univocally associated to the CDF values, which are always the same and depend only on NMC.

The results obtained by assimilating normal score-transformed concentration data are reported in Figure 5.5 for both EnKF and ES. Despite the Gaussian distribution of the model variables, the retrieved fields are not satisfactory for both the techniques and, compared to the results of scenarios 1a and 2a, even the EnKF performs poorly.

Since in scenario 3 the approximations related to the Gaussian assumption are removed, the results seem to suggest that the NST might corrupt the cross-correlation structure between Y and C in the measurement locations. Indeed, when the C values are log-transformed, the relation that maps the original pdf of C to the transformed one is the same for all the different positions in space and thus the correlation structure is conserved. This is not the case for scenario 3, where a different NST is independently applied to the concentration ensemble at each node.

In order to illustrate this point, the $C-Y$ cross-correlation structures, evaluated in the measurement nodes by means of the product PH^T in equation (2.30), are compared for scenarios 1 and 3. The cross-correlation structures for scenario 1 at time $t = 2\ T$

and $t = 4\ T$ are reported in panels a) and b) of Figures 5.6 and 5.7, respectively, while panels c) and d) of the same Figures refer to scenario 3. The analysis considers only the longitudinal behavior of the cross-correlation, by reporting in the figures only the results calculated along the direction shown with a dotted line in Figure 5.2. Firstly, we note that there are relevant differences between the EnKF and ES cross-correlation structures. The EnKF cross-correlation behavior shows always higher values at the origin (zero lag) and a rapid decay with increasing lag, i.e., moving away from the measurement location. This confirms that the effectiveness of the EnKF is limited in a portion of the domain around the measurement location (Camporese et al., 2011). With ES, the peak of correlation at the measurement location is usually smaller than with EnKF, and the cross-correlation structure is more spread out. As the cross-correlation between Y and C is usually significant only for a limited lag distance, proportional to the product of the Y integral scale and the length of the area covered by the plume, the relatively high correlation values characterizing the ES results at large lags are probably spurious. Also for this reason the result of the ES inversion is a smoothed Y field. Secondly, when the NST is applied (compare panels c) and d) with a) and b) in Figures 5.6 and 5.7), there is an overall decrease of the peaks and the cross-correlation structures are even more smoothed, showing the significant alterations operated by the transformation.

5.2.2.4 Scenarios with modified normal score-transformed concentration pdfs

In order to maintain the original $Y - C$ cross-correlation structure and, at the same time, to work with Gaussian pdfs, a modified application of the NST is proposed. At every time step, only one cumulative distribution function is built by using the concentration values simulated in all the measurement nodes. We underline that this is different from the previous application of the NST, in which a relation between C and its CDF is defined in each node independently. Now the CDF $F(C)$, estimated at time t_i, is $F(C) = (i - 0.5)/(NMC \times m_i)$, where m_i is the number of measurement locations and $i = 1, \ldots, NMC \times m_i$ is the rank of the concentration values after sorting

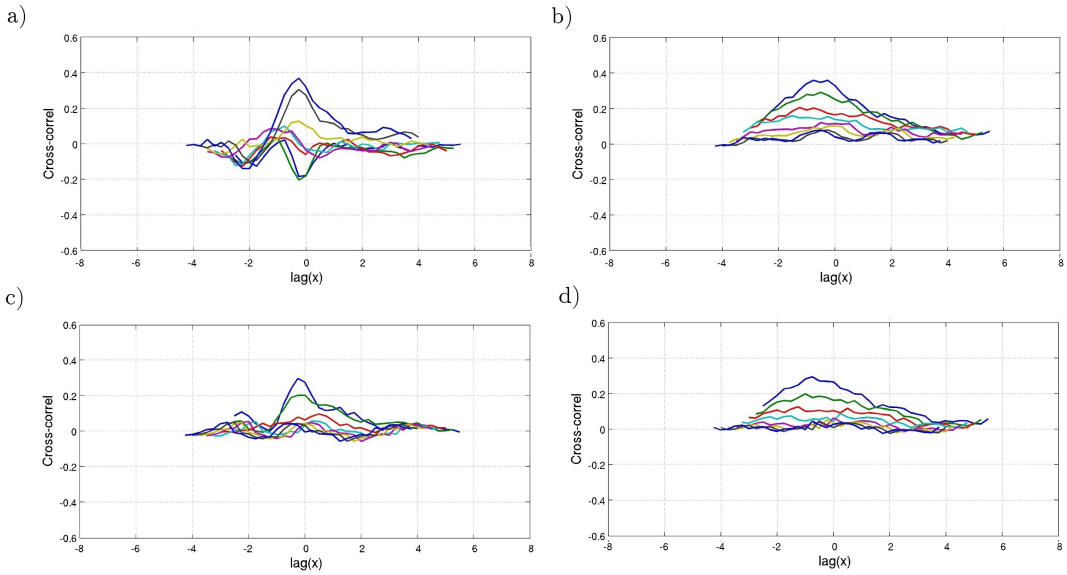


Figure 5.6: $Y - C$ cross-correlation at $t = 2\ T$ for the nodes located at $y = 5.75\ L$ in scenario 1a (subpanel a), scenario 1b (subpanel b), scenario 3a (subpanel c) and scenario 3b (subpanel d). Each color corresponds to a correlation structure centered at a different node sampled by the plume (see Figure 5.2, subpanel a)

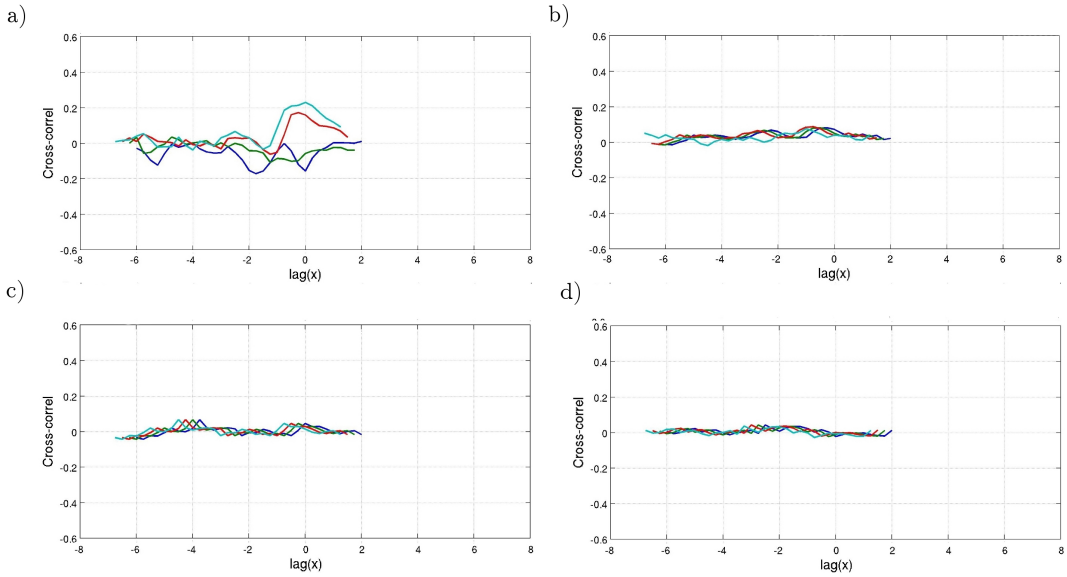


Figure 5.7: $Y - C$ cross-correlation at $t = 4\ T$ for the nodes located at $y = 5.75\ L$ in scenario 1a (subpanel a), scenario 1b (subpanel b), scenario 3a (subpanel c) and scenario 3b (subpanel d). Each color corresponds to a correlation structure centered at a different node sampled by the plume (see Figure 5.2, subpanel b)

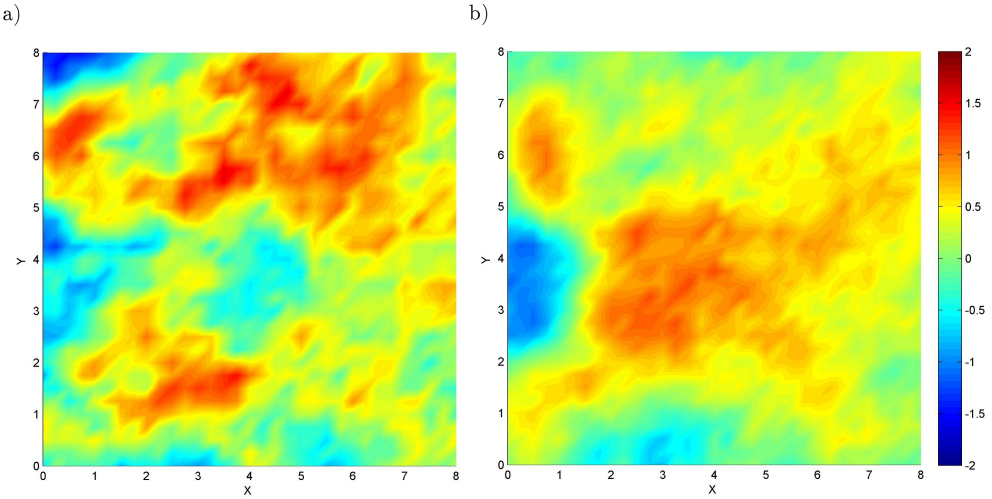


Figure 5.8: Scenario 4: log-transformed hydraulic conductivity field retrieved by a) EnKF and b) ES assimilating normal score-transformed concentration data.

the data in ascending order. With this modified application of the NST we obtain a satisfactory reproduction of the Gaussian distribution in each node, and, by using an invariant transformation, we do not alter the $Y - C$ cross-correlation structure.
The results of the inversions, reported in Figure 5.8 for both EnKF and ES, demonstrate the effectiveness of the modified NST, as the estimated Y fields are now showing an improvement (with respect to the prior fields) comparable to that in scenario 2 and look better than those of scenario 3. As in the previous scenarios, EnKF outperforms ES, which, however, shows significant improvements and seems to benefit more from the application of the modified NST.

5.2.3 Discussion

In Figure 5.9, the results of all the scenarios are summarized and compared in terms of root mean square errors of both Y and C vs time. All of the observations made in the previous sections, which were based merely on visual comparison, are confirmed by the $RMSE$ profiles. The EnKF consistently outperforms the ES, regardless of the adopted concentration pdfs, except for scenario 3a, in which the NST deteriorates the EnKF solution due to alterations of the $Y - C$ cross-correlation structure. This result, which reveals the inadequacy of the NST for this application, is in accordance with the

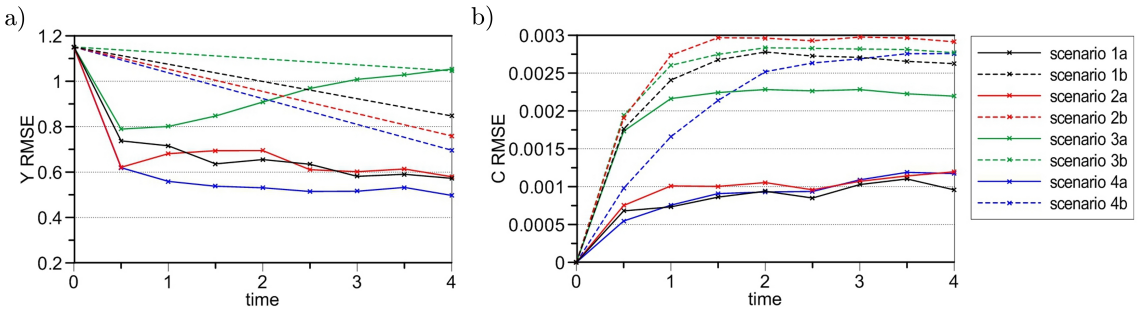


Figure 5.9: Root mean square error of a) the retrieved log-transformed hydraulic conductivity field and b) the concentration distribution in the retrieved field for the all scenarios.

conclusions drawn by Schoeniger et al. (2012). In their work, Schoeniger et al. applied the NST in conjunction with the EnKF to assimilate aquifer drawdown measurements and showed that the dependence between these data and the parameters is higher than the one between concentrations and parameters. The ES is more sensitive to the transformations operated to the assimilated data, as indicated by the $RMSE$ of Y in the various scenarios, even though the same sensitivity is not reflected by the C $RMSE$ (Figure 5.9).

Overall, the problem of retrieving the hydraulic conductivity field through the assimilation of concentration measurements is better handled by EnKF, probably due to the following reasons. First, due to the violation of Gaussianity investigated earlier with the different scenarios.

Second, due to the high nonlinearity of the problem we are dealing with. With EnKF, the Y fields are progressively updated and the simulated plumes gradually converge towards the true one. With ES, the plumes evolve freely in the prior fields until the end of the simulation and, consequently, their evolution is very different from the true one, especially at late times. To highlight this point, the true plume evolution at $t = 2\ T$ and $t = 4\ T$ (Figure 5.2) is compared with the evolution of the ensemble mean of the plumes simulated in the prior fields (Figure 5.10). In Figures 5.11 and 5.12 we also show the evolution of the ensemble average of the plumes as simulated in the Y fields estimated at $t = 2\ T$ and $t = 4\ T$, respectively, in scenario 1a (EnKF with original concentration pdfs). In other words, in Figure 5.11 EnKF is applied only until $t = 2\ T$ and thereafter the plumes are let to evolve in the estimated Y fields. The progressive

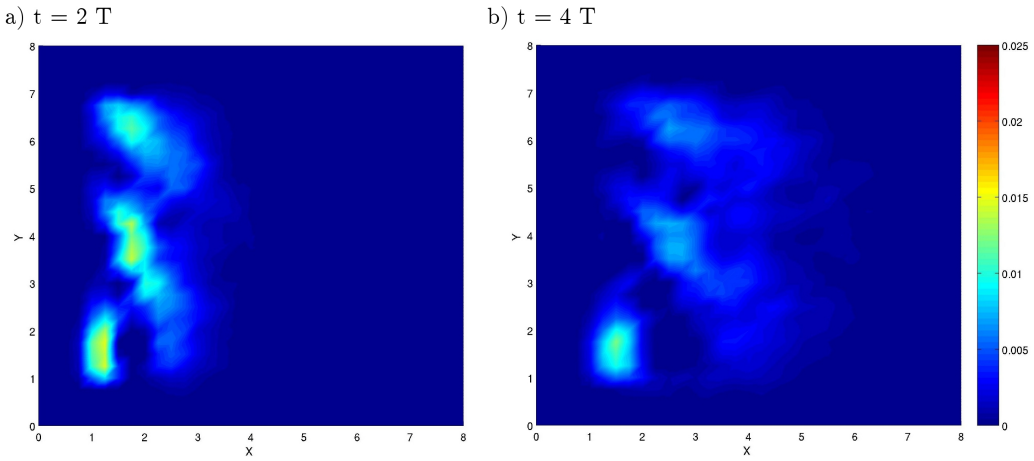


Figure 5.10: Ensemble mean of the plume evolution at $t = 2\ T$ and $t = 4\ T$ in the prior Y fields. The color bar denotes concentration.

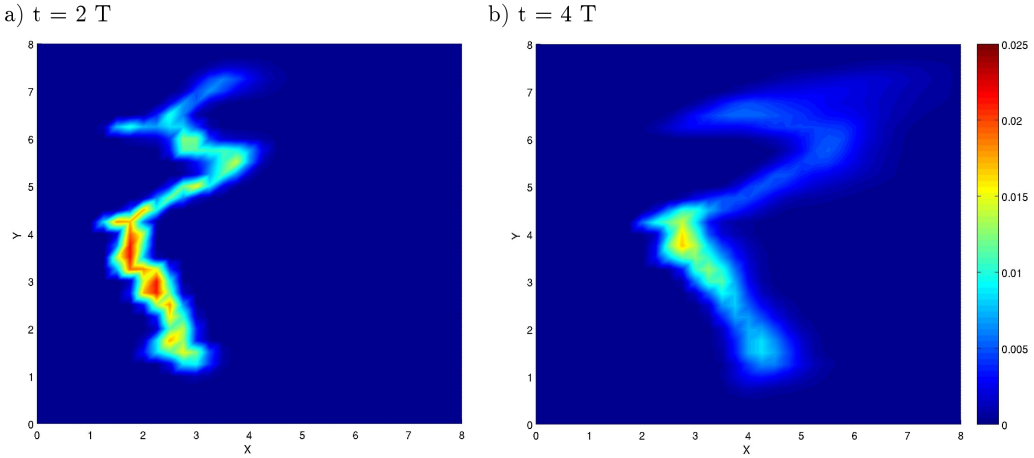


Figure 5.11: Ensemble mean of the plume evolution at $t = 2\ T$ and $t = 4\ T$ in the Y field estimated at $t = 2\ T$ by EnKF in scenario 1a. The color bar denotes concentration.

correction of the mean simulated plume is evident and already at $t = 2\ T$ it is very similar to the true one. With ES, instead, there are no recursive updates that modify the ensemble to resemble the true Y field and the nonlinearity of the problem cannot be captured.

5.3 Final remarks

The objective was to compare the performance of the two techniques and to analyze the effects of the lack of Gaussianity in the system variables. In the first scenarios

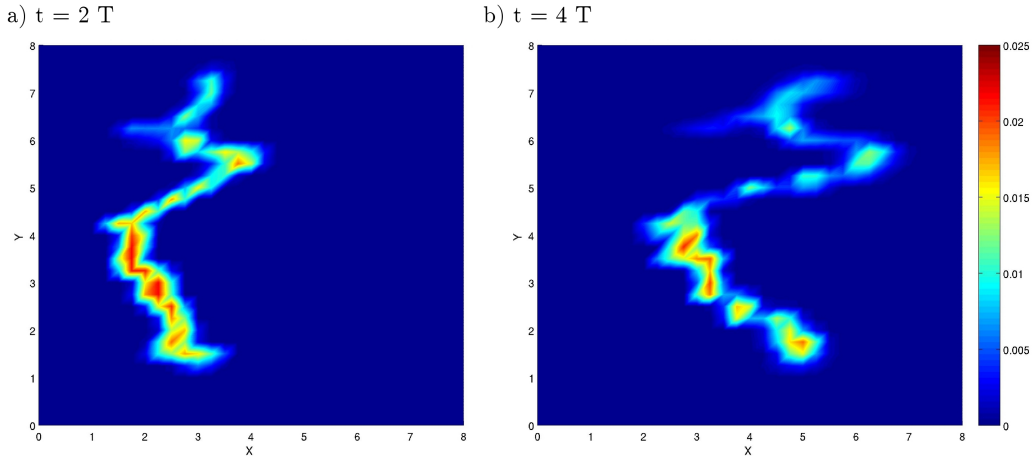


Figure 5.12: Ensemble mean of the plume evolution at $t = 2\ T$ and $t = 4\ T$ in the Y field estimated at $t = 4\ T$ by EnKF in scenario 1a. The color bar denotes concentration.

the concentrations were assimilated in the model without any manipulation, while in the other scenarios we took into account log-transformed data and two variants of the normal score transform (NST), analyzing also the cross-correlation structure between log-transformed hydraulic conductivity Y and concentration C.

The main conclusion of our study is that EnKF can reproduce with good accuracy the hydraulic conductivity field, consistently outperforming ES regardless of the probability distribution of the concentrations. This is due to two reasons: i) the lack of Gaussianity is overcome owing to the recursive Gaussian increments given by the EnKF updates that eventually lead to an ensemble of members normally distributed around the true solution; ii) the same recursive procedure continuously pulls the realizations toward the true solution, easing the inversion problem of the strong nonlinearity of the dispersion process. The only case in which EnKF does not work properly is when NST is applied to ensure that the concentration pdf is Gaussian in each node of the domain. This is due to the consequent alteration of the $Y - C$ cross-correlation structure, which instead must be correctly evaluated in order to assure the effectiveness of the EnKF inversion procedure. This suggests that NST must be applied with caution in any Kalman-filter based inversion scheme, checking at any time for possible corruptions of the cross-correlation between parameters and assimilation variables.

ES performs always worse than EnKF as it does not involve recursive updates of the

Y fields. This has two consequences: i) the solute plumes are free to evolve in the prior fields without corrections, eventually leading to significant differences from the true plume evolution; ii) non-Gaussian contributions in the concentration pdf are not kept under control.

Chapter 6

Overcoming the need of in situ calibrated petrophysical relation

6.1 Introduction

Borehole ERT monitoring of saline tracer tests allows to collect time-lapse geophysical data as changes occur in an aquifer as a consequence of dynamical variations in the hydrological state of the subsurface, and, as previously shown, it seems to be a promising tool for the hydrological characterization of natural aquifers. Nevertheless, ERT measurements are not directly related to the hydraulic parameters needed to predict flow and transport in porous media. The electrical conductivity field must be reconstructed by means of a geophysical inversion on the basis of current and voltage measurements, and the use of a petrophysical law (e.g., Archie's law) is required to deduce the solute concentration and the plume evolution closely linked to the distribution of the hydraulic properties. To retrieve the hydraulic conductivity distribution K from an ERT monitored saline tracer test, and to overcome the need for the knowledge of the concentration spatio-temporal evolution, it is proposed a novel approach that couples travel time modeling of transport with the ensemble Kalman filter (EnKF) used as an inversion tool. The definition of the solute transport in terms of travel (or residence) times allows to analyze the sequence of changes in electrical resistivity deduced from a ERT survey without converting the electrical data into concentrations (see section

2.1.2), since in ERT monitored tracer tests, electrical conductivity changes in time act as a proxy for the concentration travel times. Moreover, in the case of a multiple well saline tracer test, only two dimensional images of electrical conductivity defining the control planes (CP) need to be reconstructed by the geophysical inversion from current and voltage measurements, with a noticeable saving of computer resources with respect to the fully 3D case. Obviously, by using a single CP, only average informations can be achieved in the travel distance from the source and the CP, but this limitation can be theoretically overcome using more that one CP located at different travel distances.
To demonstrate the ability of the proposed approach, we consider a synthetic 3D case, where a control plane, defined by a pair of ERT monitored boreholes, is located downstream from an injection well perpendicular to the mean flow direction. The CP is properly subdivided in subareas, each of them characterized by a specific travel time distribution. The assimilation of the travel times in a Lagrangian model of transport whose K distribution represents the system state of the EnKF allows to update the average hydraulic conductivity for each subarea and hence to determine the K spatial variability with a resolution related to the one of the control plane discretization. We report on the capability of the inversion procedure as a function of the discretization in subareas.

6.2 Numerical experiment

6.2.1 Model setup

The reference field has dimensions of 8 $L \times 8$ $L \times 8$ L, where L is an arbitrary and consistent unit length, and is discretized along each directions into $L/4$ sided cubic volumes (voxels), for a total of $33 \times 33 \times 33 = 35937$ corresponding nodes. Dirichlet boundary conditions are applied at $x = 0$ L (piezometric head $h = 100.0$ L) and at $x = 8$ L ($h = 95.2$ L), while Neumann no-flow boundary conditions are imposed along the remaining sides of the domain. The Y distribution of the reference field is obtained through a direct Fourier-transform method (Robin et al., 1993). An anisotropic and exponential covariance structure is assigned with the horizontal integral scale of $\lambda_h = 8$ L

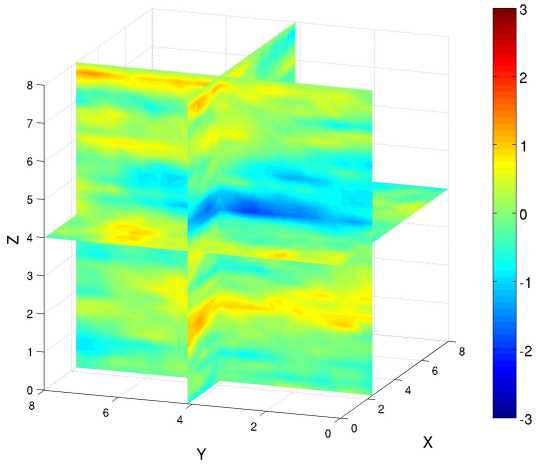


Figure 6.1: The true log-hydraulic conductivity spatial distribution

and the vertical one of $\lambda_v = 1\ L$. The resulting field has spatial mean $\langle Y \rangle = -0.18$ and variance $\sigma^2_Y = 0.24$ and its representation is given in Figure 6.1.

A tracer test is simulated by assuming an instantaneous solute injection in a well of $0.5\ L$ diameter, vertical size of $6\ L$, and centered in $x = 0.875\ L$ and $y = 4.125\ L$. The solute is simulated by 4992 particles uniformly distributed in the horizontal and vertical directions. The CP, perpendicular to the flow direction, is located at $x = 2.375\ L$ and is divided in a proper number n of sub-control plane (sCP), i.e., in a number of sCP that should be high enough to characterize the Y vertical variations (see figure 6.2). This choice is not known a priori, since the only available data are the changes of resistivity in the CP. Starting from a guess based on the recognized areas with different resistivity values, the proper number of sCP will be assessed on the basis of a trial procedure. The plume is then made to evolve, and every $0.5\ T$, where T is any consistent time unit, the travel time CDF value is estimated for each sCP, till all the plume has crossed the CP (at $t = 14\ T$ in this case). It has to be highlighted that the travel time of all the particles that compose the solute injection is computed by Pollock's technique in a discrete flow field (finite volume methods) and that the CDFs of the travel time are obtained as the breakthrough curve of the plume solute mass crossing each sCP. Indeed, as explained in section 2.1.2, for example we have the median arrival time τ_{50} when half of the solute has crossed the CP.

During the simulations, we start with NMC Y fields with prior statistics different

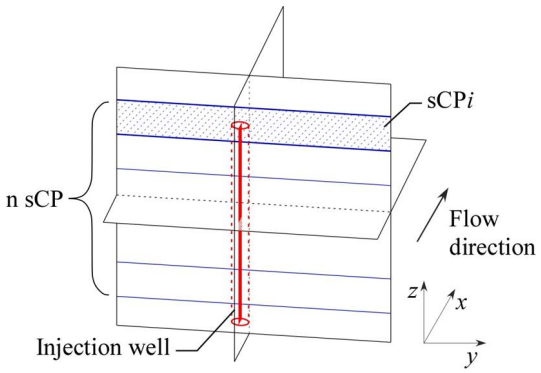


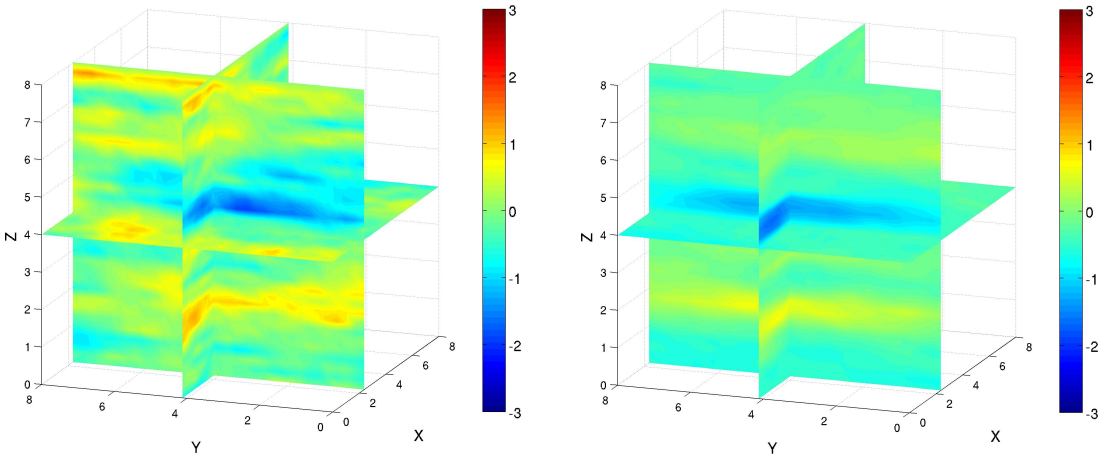
Figure 6.2: Schematic representation of the problem: the injection well, the control plane CP and its division in sCP are marked

from the true ones. In each field, the transport problem is solved for the time period $[t_0; t_1]$, being t_1 the first measurement time and the CDF value is estimated at the corresponding time for each sCP. The measurement vector is then $z(t_1) = [CDF_{sCP1}(t_1), \ldots, CDF_{sCPn}(t_1)]$, where $CDF_{sCPi}(t_j)$ is the CDF value for the $i-th$ sCP at time t_j. By using the measurements, the Y fields are updated and the transport problem is solved again for the time period $[t_0; t_2]$. The CDF values are evaluated in each sCP for times t_1 and t_2 and the measurement vector becomes $z(t_2) = [CDF_{sCP1}(t_1), \ldots, CDF_{sCPn}(t_1), CDF_{sCP1}(t_2), \ldots, CDF_{sCPn}(t_2)]$, i.e., all the CDF values till the considered measurements time are assimilated for each sCP. Note that this application of the EnKF is different from the one previously adopted and it may be considered as a restarted Ensemble Kalman Smoother (Evensen, 2009a). The whole process goes on by updating the Y fields, till measurements are available.

The tests carried out have the purposes to determine the proper number of sCP to use in order to identify the Y field. The proposed tests are reported in table 6.1.

Table 6.1: Main characteristics of the numerical experiments realized

		prior statistics		
	number of sCP	$< Y >$	σ^2_Y	λ_h (L)
Reference field	as in the corresponding test	0.18	0.24	1
test 1	4	-0.50	0.75	8
test 2	8	-0.50	0.75	8
test 3	12	-0.50	0.75	8
test 4	8	-0.50	0.75	∞
test 5	8	-0.50	0.75	4



(a) (b)

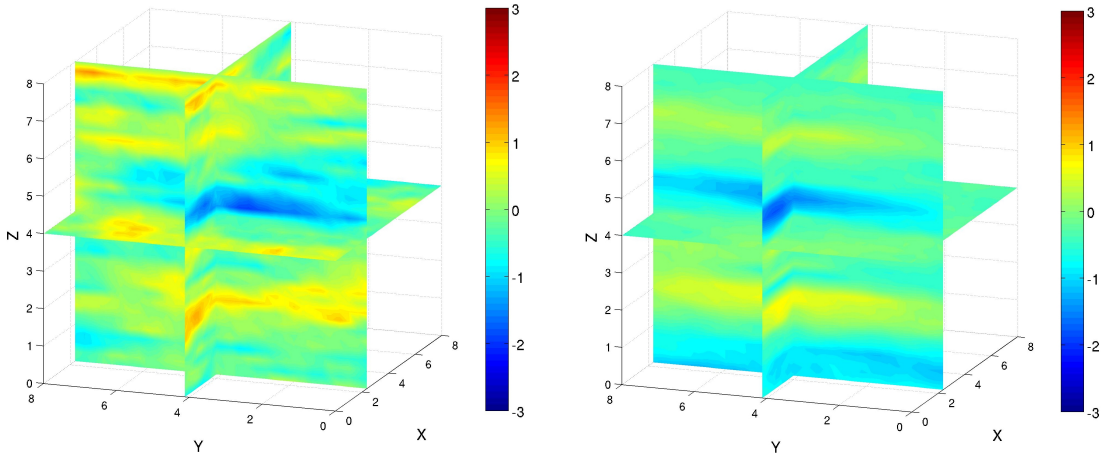
Figure 6.3: Test 1: comparison between a) the true and b) the retrieved log-hydraulic conductivity field

The effectiveness of the model is analyzed through the retrieved Y field and the root mean square error of the CDFs and of the plume in the retrieved Y field. In the all tests the vertical integral scale is assumed equal to the true one.

6.3 Results

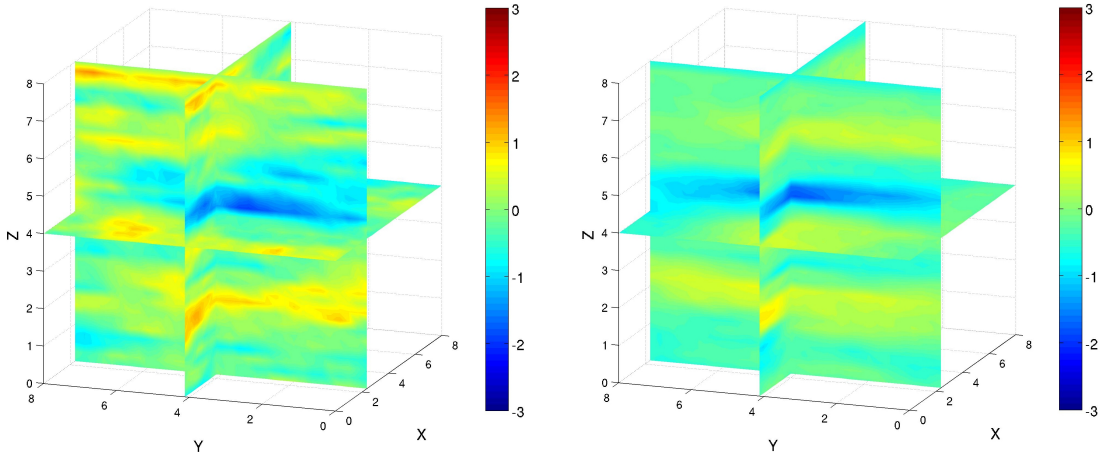
The influence of the number of sCP in capturing the main Y variations along the vertical direction is assessed through the tests 1, 2 and 3, where CP is divided in 4, 8 and 12 sCP respectively. The number of sCP is not further increased because we consider unhelpful to give a more refined discretization, and being also the vertical integral scale equal to one. The retrieved fields are reported from figure 6.3 to figure 6.5 and the first noticeable result is that the main Y vertical variations are recognized and reproduced by the method even with a rough discretization of the CP. Secondly, as expected, by increasing the number of sCP, the definition of Y seems to improve, since area with different Y values are more defined.

To prove the convergence of the model, i.e., if the assimilation of the CDF values improves the assessment of the Y field, a first analysis observes how the simulated CDFs converge toward the true one. We choose, as example, the fourth sCP in test 2 and the all NMC simulated CDFs (one for each Y field) are represented after every update



(a) (b)

Figure 6.4: Test 2: comparison between a) the true and b) the retrieved log-hydraulic conductivity field



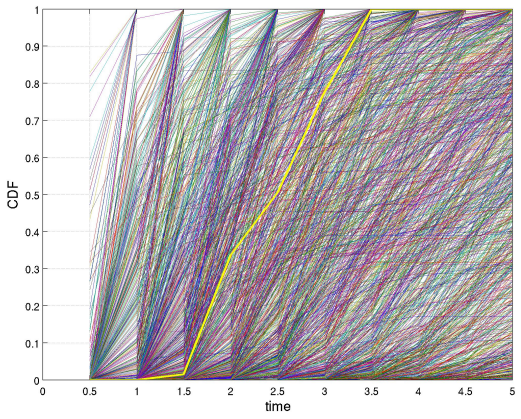
(a) (b)

Figure 6.5: Test 3: comparison between a) the true and b) the retrieved log-hydraulic conductivity field

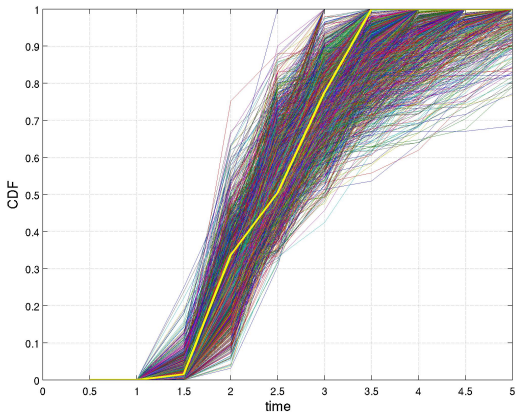
step. From figure 6.6(a) to figure 6.6(c), the simulated CDF are represented with the true CDF (i.e, the CDF obtained in true field and in the same sCP when 8sCP are used), before any update (figure 6.6(a)), after $t = 7/T$ (figure 6.6(b)) and after the last update (figure 6.6(c)). As it can be seen, during the computed CDFs approach the true one and after the last update, the dispersion around the true CDF is really limited. This analysis reveal the capabilities of the model and that it is possible the update of the Y values through the assimilation of the CDF measurements.

To evaluate the goodness of the results, the root mean square error (RMSE) of the CDF as it comes in the retrieved Y field is calculated. It is also estimated the concentration RMSE by realizing the injection in the retrieved field. This latter analysis can give useful information about the capabilities of the model and it can help in identifying the influences of the considered "variables" (number of sCP and integral scale) on the performances of the method but, obviously, it can be realized only during synthetic analysis, when the concentrations are known. During an in-situ test, only the travel time pdf is available since the C estimation and the use of a petrophysical relationship is skipped. Figure 6.7 gives the RMSE calculated for both the CDF and the C reported, for convenience, for the all cases. It can be seen that the differences among test 1, 2 and 3 in the CDF RMSE are limited, although the performances of test 2 are better than those of test 1 and 3, when the concentration RMSE is considered (6.7b). The differences in the RMSE analysis of CDF and C can be also ascribed to the pointwise evaluation of the latter (see, example, 5.2), while to compare test cases characterized by different horizontal integral scale and sCP number, the CDF RMSE must be evaluated by properly subdividing the CP. Here a first attempt was made evaluating the CDF RMSE on each element (1/4 of the vertical integral scale) of the CP. Further investigations have to be realized.

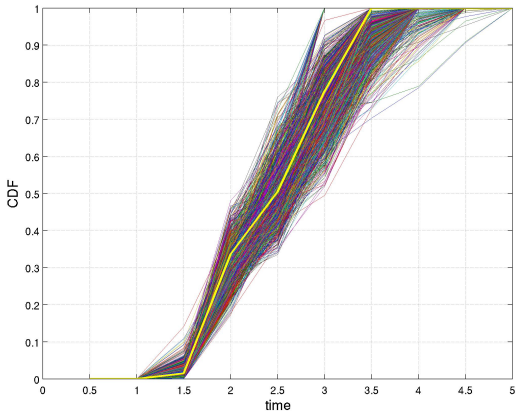
In all the tests examined, the solute is injected in a well that can be considered as a vertical line solute source. In a 3D domain the transversal dispersion is negligible respect to the longitudinal one, and the geometric characteristics of the injection volume dominate the transversal size of the downstream portion of the aquifer affected by the non-Fickian plume evolution. In our single well tracer test analysis, a vertical plane of



(a) $t = 0\ T$



(b) $t = 7\ T$



(c) $t = 14\ T$

Figure 6.6: CDFs of the travel time for the fourth sCP in test 2 at times a) $t = 0\ T$, b) $t = 7\ T$, c) $t = 14\ T$. The yellow thick line represents the CDF in the true field. All the other thin lines represents the CDF in each simulated field

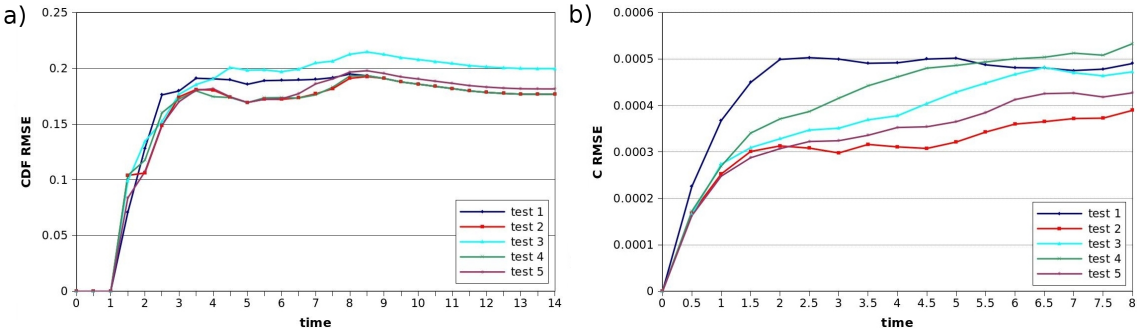
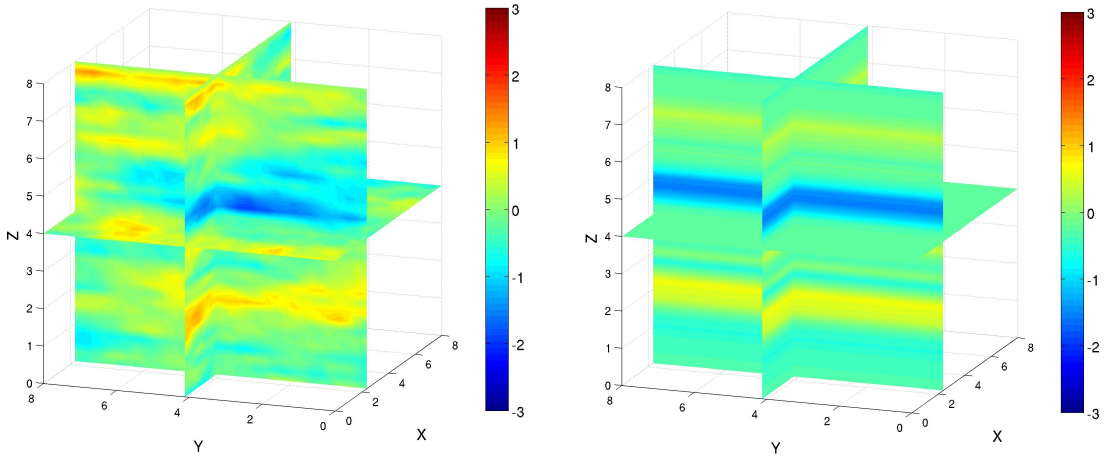


Figure 6.7: Root mean square error RMSE of a) the CDF in the retrieved field and b) the C in the retrieved field

the aquifer, parallel to the mean flow direction, is crossed by the solute that spreads according to the hydraulic conductivity vertical distribution. Consequently, none information can be retrieved in the direction transversal to the mean flow direction, unless an array of multiple wells is considered as injection volume. Obviously these considerations are strictly valid if the well injection geometry approaches a planimetric point source, that is the horizontal integral scale of Y overcomes the well diameter. Because of the finite planimetric size of our injection, tests 4 and 5 are realized with the purpose to determine how an error in the definition of the horizontal integral scale influences the goodness of the results. In both cases, 8sCP are used, since it seems to be a proper value from the previous tests. Test 4 considers the ideal case of a perfectly-stratified aquifer, i.e., with $\lambda_h = \infty$. Figure 6.8 shows that the main characteristics of the field are still caught and the RMSE analysis confirms this results. Even when the horizontal integral scale is underestimated (test 5), the resulting field gives a good approximation of the true one. This result confirms that a tracer test analysis of a single well injection can give information only about the vertical distribution of Y.

6.4 Final remarks

Although the results reported in this Chapter are preliminary and more analysis have to be developed, the travel time approach that allows to overcame the need of a petrophysical law, seems to be a promising technique in subsurface investigations. Respect to the approach that uses the concentrations spatio-temporal evolution, the



(a) (b)

Figure 6.8: Test 4: comparison between a) the true and b) the retrieved log-hydraulic conductivity field

travel time approach, besides to skip the in situ calibration of the Archie's law, requires a less computational expensive 2D geophysical inversion, with manifest advantages in real world applications.

Till now, it is possible to say that the comparison between the retrieved and the true Y fields shows that the developed method is always able to identify the main vertical variations of Y, independently to the assigned horizontal characteristics. The effectiveness of the model is also evident through the progressive correction of all the simulated CDFs that during the assimilation procedure tend to converge toward the corresponding true CDF.

The RMSE analysis shows non appreciable differences in the reproduction of the travel time CDFs, independently to the number of sCP.

The RMSE of the plume evolution in the retrieved Y field suggests a poorer performance of the concentration spatial distribution when a rough sCP subdivision is adopted.

According to the stochastic transport theory, the proposed approach is able to capture with a good approximation the vertical distribution of Y in the area surrounding the injection also with a simple model of perfectly stratified aquifer.

Summary and Conclusions

In this thesis several inversion procedure to assess heterogeneous hydraulic conductivity fields at the local scale from the knowledge of data obtained by tracer tests are proposed and critically analyzed. The main goal is to suggest a method to recognize the hydraulic conductivity spatial distribution at local scale in real world cases under the hypothesis that the tracer test evolution is mainly controlled by the latter. The relevance of the work is strictly related to the definition of well-catchment areas or to the monitoring of landfill sites, where the plume evolution has to be defined at early or mean travel time. In this condition the transport cannot be considered neither Gaussian nor ergodic, and the solution is not always straightforward.

In the all cases presented Kalman filter techniques are coupled with the Lagrangian transport model: the main differences in the proposed inversion model are related to the Kalman filter technique used (EnKF or ES) and to the assimilated data (macroscopic concentrations or travel time), being more subtle distinctions based on the number and location of the assimilated variables. In all the scenarios analyzed, it is assumed that the available measurements are deduced from a tracer test ERT survey. In other words, the data assimilation procedure considers three-dimensional or two-dimensional time lapse spatially distributed information, characterized by a prescribed degree of uncertainty. For this reason, the focus is only on the inverse model since the ERT results are deduced from synthetic experiments and considered as "yet available" data (this means that in this work troubles related to the geophysical inversion of ERT data are not considered). A preliminary real field ERT analysis was developed in the experimental site of Settolo - Valdobbiadene (Perri et al., 2012) and new experiments are currently in preparation, but a specific application of the proposed inversion tool to these real data

has not been developed until this time.
The main results of my work, already discussed at the end of each Chapter, can be summarized as:

i) proof of effectiveness of the proposed tool based on the EnKF and Lagrangian transport model in retrieving the hydraulic conductivity spatial distribution;

ii) pros and cons of the suggested procedures, mainly related to the reduction of numerical issues, like memory amount requirements and filter inbreeding;

iii) comparison between "on-line" (EnKF) and "off-line" (ES) data assimilation techniques that shows the limits of the ES whose performance is always worse than the EnKF one. Moreover it is demonstrated that the EnKF can overcome the lack of Gaussianity in the system variables thanks to its recursive procedure

iv) formulation of a tool, based on EnKF and travel time analysis, able to retrieve the main hydraulic conductivity field characteristics avoiding the need of any petrophysical relation (e.g. Archie's law).

Bibliography

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

Annan, A. P., 2005. Hydrogeophysics. Vol. 50. of Water Sci. Technol. Library. Springer, New York.

Archie, G. E., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics. Petrol. Tech. 1, 55–62.

Bailey, R., Baú, D., 2010. Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution. Water Resour. Res. 46, W12543.

Bailey, R., Baú, D., 2012. Estimating geostatistical parameters and spatially-variable hydraulic conductivity within a catchment system using an ensemble smoother. Hydrol. Earth Syst. Sci 16 (2), 287–304.

Bailey, R. T., Baù, D. A., Gates, T. K., 2012. Estimating spatially-variable rate constants of denitrification in irrigated agricultural groundwater systems using an ensemble smoother. J. Hydrol. 468–469 (0), 188–202.

Barlebo, H. C., Hill, M. C., Rosbjerg, D., 2004. Investigating the macrodispersion experiment (made) site in columbus, mississippi, using a three-dimensional inverse flow and transport model. Water Resour. Res. 40(4) (W04211).

Baú, D., Mayer, A., 2008. Optimal design of pump-and-treat systems under uncertain hydraulic conductivity and plume distribution. J. Contam. Hydrol. 100, 30–46.

Bellin, A., Rubin, Y., Rinaldo, A., 1994. Eulerian-lagrangian approach for modeling of flow and transport in heterogeneous geological formations. Water Resour. Res. 30 (11), 2913–2924.

Bellin, A., Salandin, P., Rinaldo, A., 1992. Simulation of dispersion in heterogeneous porous formations: Statistics, first order theories, convergence of computations. Water Resour. Res. 28, 2211–2227.

Binley, A., Cassiani, G., Middleton, R., Winship, P., 2002. Vadose zone flow model parameterisation using cross-borehole radar and resistivity imaging . J. Hydrol. 267, 147–159.

Boggs, J., Young, S., Beard, L., Gelhar, L., Rehfeldt, K., Adams, E., 1992. Field study of dispersion in a heterogeneous aquifer: 1. overview and site description. Water Resour. Res. 28 (12), 3281–3291.

Camporese, M., Cassiani, G., Deiana, R., Salandin, P., 2011. Assessment of local hydraulic properties from electrical resistivity tomography monitoring of a three-dimensional syntetic tracer test experiment. Water Resour. Res. 47 (W12508).

Caroni, E., Fiorotto, V., 2005. Analysis of concentration as sampled in natural aquifers. Trans. Porous Media 59 (1), 19–45.

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

Certes, C., DeMarsily, G., 1991. Application of the pilot point method to the identification of aquifer transmissivities. Adv. Water Resour. 14, 284–300.

Chen, X., Murakami, H., Hahn, M. S., Hammond, G. E., Rockhold, M. L., Zachara, J. M., Rubin, Y., 2012. Three-dimensional bayesian geostatistical aquifer characterization at the hanford 300 area using tracer test data. Water Resour. Res. 48 (W06501).

Chen, Y., Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble kalman filter. Adv. Water Resour. 29, 1107–1122.

Crestani, E., Camporese, M., Salandin, P., 2010. Hydraulic conductivity assessment via tracer test data assimilation: a comparison of updating techniques. In: Proceedings of the IAHR International Groundwater Symposium 2010.

Cvetkovic, V., Shapiro, A., Dagan, G., 1992. A solute flux approach transport heterogeneous formations 2. uncertainty analysis. Water Resour. Res. 28 (5), 1377–1388.

Dafflon, B., Barrash, W., Cardiff, M., Johnson, T., 2011. Hydrological parameter estimations from a conservative tracer test with variable-density effects at the boise hydrogeophysical research site. Water Resour. Res. 47 (W12513), 199–210.

Dagan, G., 1986. Statistical - theory of groundwater - flow and transport - pore to laboratory, laboratory to formation, and formation to regional scale. Water Resour. Res. 22, S120–S134.

Dagan, G., 1989. Flow and Transport in Porous Formations. Springer.

Dagan, G., 1991. Dispersion of a passive solute in non-ergodic transport by steady velocity fields in heterogeneous formations. J. Fluid Mech. 233, 197–210.

Dagan, G., Cvetkovic, V., Shapiro, A., 1992. A solute flux approach transport heterogeneous formations 1. the general framework. Water Resour. Res. 28 (5), 1369–1376.

Dagan, G., Nguyen, V., 1989. A comparison of travel time and concentration approaches to modeling transport by groundwater. Journal of Contaminant Hydrology 4 (1), 79–91.

Doherty, J. E., Fienen, M. N., Hunt, R. J., 2010. Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions. Scientific Investigations Report 2010 5168, U.S. Geological Survey.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. J. Geophys. Res. 99(C5), 10143–10162.

Evensen, G., 2004. Sampling strategies and square root analysis schemes for the EnKF. Ocean Dynam. 54, 539–560.

Evensen, G., 2009a. Data Assimilation. The ensemble Kalman filter. Springer, New York.

Evensen, G., 2009b. The ensemble Kalman filter for combined state and parameter estimation. IEEE Contr. Syst. Mag. 29 (3).

Fu, J., Gomez-Hernandez, J. J., 2009. Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking markov chain monte carlo method. J. Hydrol. 364 (3-4), 328–341.

Gelhar, L., 1993. Stochastic Subsurface Hydrology. Prentice Hall.

Gomez-Hernandez, J., Sahuquillo, A., Capilla, J., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data - i. theory. J. Hydrol. 203, 162–174.

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

Hendricks Franssen, H. J., Gomez-Hernandez, J., Sahuquillo, A., 2003. Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data. J. Hydrol. 281, 281–295.

Hendricks Franssen, H. J., Kaiser, H. P., Kuhlmann, U., Bauser, G., Stauffer, F., Muller, R., Kinzelbach, W., 2011. Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. Water Resour. Res. 47 (W02532).

Hendricks Franssen, H. J., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res. 44, 1–21.

Huang, C., Hu, B., Li, X., Ye, M., 2009. Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source. Stochastic Environmental Research and Risk Assessment 23, 1155–1167.

Jafarpour, B., Tarrahi, M., 2011. Assessing the performance of the ensemble kalman filter for subsurface flow data integration under variogram uncertainty. Water Resour. Res. 47 (W05537).

Kemna, A., Vanderborght, J., Kulessa, B., Vereecken, H., 2002a. Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ERT) and equivalent transport models. J. Hydrol. 267, 125–146.

Kemna, A., Vanderborght, J., Kulessa, B., Vereecken, H., 2002b. Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ERT) and equivalent transport models. J. Hydrol. 267, 125–146.

Keppenne, C. L., 2000. Data assimilation into a Primitive-Equation Model with a parallel Ensemble Kalman Filter. Mon. Weather Rev. 128(6), 1971–1981.

Kowalsky, M. B., Finsterle, S., Williams, K. H., Murray, C., Commer, M., Newcomer, D., Englert, A., Steefel, C. I., Hubbard, S. S., 2012. On parameterization of the inverse problem for estimating aquifer properties using tracer data. Water Resour. Res. 48 (W06535).

Li, L., Zhou, H., Gomez-Hernandez, J. J., Franssen, H. J. H., 2012. Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble kalman filter. J. Hydrol. 428, 152–169.

Liu, G., Chen, Y., Zhang, D., 2008. Investigation of flow and transport processes at the MADE site using ensemble Kalman filter. Adv. Water Resour. 31(7), 975–986.

Liu, X., Cardiff, M., Kitanidis, P., 2010. Parameter estimation in nonlinear environmental problems. Stochastic Environmental Research and Risk Assessment 24, 1003–1022.

McLaughlin, D., Townley, L. R., 1996. A reassessment of the groundwater inverse problem. Water Resour. Res. 32(5), 1131–1161.

Nowak, W., 2009. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. Water Resour. Res. 45, 1–17.

Papoulis, A., Pillai, S., 2002. Probability, random variables, and stochastic processes. McGraw-Hill.

Perri, M., Cassiani, G., Gervasio, I., Deiana, R., Binley, A., 2012. A saline tracer test monitored via both surface and cross-borehole electrical resistivity tomography: Comparison of time-lapse results. J. Appl. Geophys. 79, 6–16.

Pollock, D. W., 1988. Semianalytical computation of path lines for finite-difference models. Ground Water 26 (6), 743–750.

Ptak, T., Piepenbrink, M., Martac, E., 2004. Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport - a review of some recent developments. J. Hydrol. 294, 122–163.

Ramarao, B., Lavenue, A., DeMarsily, G., Marietta, M., 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.

Robin, M., Gutjahr, A., Suducky, E., Wilson, J., 1993. Cross-correlated random-field generation with the direct fourier-transform method. Water Resour. Res. 29 (7), 2385–2397.

Rubin, Y., 2003. Applied stochastic hydrogeology. Oxford University Press.

Rubin, Y., Chen, X., Murakami, H., Hahn, M., 2010. A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields. Water Resour. Res. 46 (W10523).

Salandin, P., Fiorotto, V., 1998. Solute transport in highly heterogeneous aquifers. Water Resour. Res. 34, 949–961.

Schoeniger, A., Nowak, W., Hendricks, H. F., 2012. Parameter estimation by ensemble kalman filters with transformed data: Approach and application to hydraulic tomography. Water Resour. Res. 48 (W04502).

Sposito, G., Dagan, G., 1994. Predicting plume evolution in heterogeneous porous formation. Water Resour. Res. 30(2), 585–589.

Sun, A. Y., Morris, A., Mohanty, S., 2009. Comparison of deterministic ensemble kalman filters for assimilating hydrogeological data. Adv. Water Resour. 32, 280–292.

Vrugt, J. A., Robinson, B. A., Vesselinov, V. V., 2005. Improved inverse modeling for flow and transport in subsurface media: Combined parameter and state estimation. Geophysical Research Letters 32 (18), n/a–n/a.

Vrugt, J. A., Stauff, P. H., Wöhling, T., Robinson, B. A., Vesselinov, V. V., 2008. Inverse Modeling of Subsurface Flow and Transport Properties: A Review with New Developments. Vadose Zone J. 7 (2).

Xie, X., Zhang, D., 2010. Data assimilation for distributed hydrological catchment modeling via ensemble kalman filter. Adv. Water Resour. 33 (6), 678 – 690.

Yen, B., Cheng, A., Melching, C., 1986. First order reliability analysis, in Stochastic and risk analysis in hydraulic engineering. Water Resources Publications, iSBN 0918334-57-8.

Zhang, D., Andricevic, R., Sun, A., Hu, X., He, G., 2000. Solute flux approach to transport through spatially nonstationary flow in porous media. Water Resour. Res. 36 (8), 2107–2120.

Zhou, H., Gómez-Hernández, J. J., Hendricks Franssen, H.-J., Li, L., 2011. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Adv. Water Resour. 34, 844–864.

Zimmerman, D. A., de Marsily, G., Gotway, C. A., Marietta, M. G., Axness, C. L., Beauheim, R. L., Bras, R. L., Carrera, J., Dagan, G., Davies, P. B., Gallegos, D. P.,

Galli, A., Gomez-Hernandez, J., Grindrod, P., Gutjahr, A. L., Kitanidis, P. K., Lavenue, A. M., McLaughlin, D., Neuman, S. P., RamaRao, B. S., Ravenne, C., Rubin, Y., 1998. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour. Res. 34(6), 1373–1413.