Tools and analysis of spatio-temporal dynamics in heterogeneous aquifers: Applications to artificial recharge and forced-gradient solute transport

by

Daniele Pedretti

A Dissertation Submitted to Doctorate School of the

Universitat Politecnica de Catalunya – UPC- Barcelona Tech

for the title of Doctor of Science

Advisor:

Prof. Xavier Sánchez-Vila

Co-advisors:

Prof. Daniel Fernàndez-Garcia

Prof. Diogo Bolster

President of the jury:

Prof. Jesús Carrera Ramirez

Examinators:

Prof. Maarten W. Saaltink

Prof. Alberto Bellin

Reviewers:

Prof. David. A. Benson

Prof. Tanguy Le Borgne

Grup d'Hidrologia Subterranea / Hydrogeology Group

Thesis abstract

This thesis deals with the development of tools and analysis to characterize and predict artificial recharge and radial convergent solute transport processes in heterogeneous media. The goal is to provide new insights to understand how heterogeneity, which is the main natural source of uncertainty in decision-making processes related with groundwater applications, can be controlled and its effects predicted for practical purposes in these topics. For hydrogeological applications, accurate modeling of phenomena is needed, but it is uncertain. Uncertainty is derived from the spatio-temporal random distribution of hydrodynamic (physical, chemical and biological) variables affecting groundwater processes, which is translated into random distribution of modeling parameters and equations. Such randomness is of two types: epistemic, when it can be reduced increasing the sample frequency of an experiment; aleatory, when it cannot be reduced when more information is analyzed. Sometimes hydrodynamic processes occur at so small scales that they become impossible to characterize with traditional methods, and from a practical perspective, this is analogous to deal with aleatoric model parameters. However, if some constitutive relationship (either empirically, theoretically or physically based) can be built between processes across different scales, then small-scale processes can be reproduced by equivalent large-scale model parameters. Uncertainty becomes amenable to be treated as epistemic randomness, and large-scale characterization techniques can be used to improve the description, interpretation or prediction of these processes. This thesis deals with these topics. The manuscript is composed by two main parts (the first on artificial recharge and the second on solute transport), each of them divided into three chapters. In chapter 1 of each part, a tool is developed to obtain quantitative information to model a selected variable at coarse grid resolutions. In the case of artificial recharge, satellite images are used to model the spatial variability of the infiltration capacity on top soils with a metric-scale detail. In the case of solute transport, a new method to estimate density from particle distribution is shown. In chapters 2, it is explored what processes occurring at the fine scales can affect the interpretation of artificial recharge and solute transport processes at larger scales. In the first part, a combined method that joins satellite images and field data along with a simple clogging model is used to display the equally-possible spatio-temporal mapping of the infiltration capacity of topsoil during artificial pond flooding activities. In the second part, numerical three-dimensional models are used to simulate transport in heterogeneous media under convergent radial flow to a well at fine scale. It is shown that an appropriate model framework can reproduce similar observations on contaminant temporal distribution at controlling section similar to those obtained in the field tracer tests. It is also provided a physical explanation to describe the so-called anomalous late-time behavior on breakthrough curves which is sometimes observed in the reality at larger scales. In the chapters 3, models are used to define the uncertainty around operating parameters in the optic of prediction and management on artificial recharge and solute transport. In the first case, a probability framework is built to define the engineering risk of management of artificial recharge ponds due to random variability of the initial distribution of infiltration, which controls several important clogging factors based on theoretical approaches. In the case of solute transport, it is discussed how equivalent parameters based on mass-transfer models can be related with the geometrical distribution of hydraulic parameters in anisotropic formation, when convergent flow tracer tests are used.

Resumen de tesis

Esta tesis está enfocada en el desarrollo de herramientas y análisis para caracterizar y predecir procesos de recarga artificial y de transporte de solutos convergente radial en medios heterogéneos. El objetivo es proporcionar nuevas perspectivas para comprender cómo la heterogeneidad, que es la principal fuente natural de incertidumbre en la toma de decisiones relacionadas con las aplicaciones en hidrogeología, puede ser controlada y sus efectos predichos para propósitos práctico. Para ello se precisa correctamente a los fenómenos, pero esta tarea es incierta. La incertidumbre se deriva de la distribución aleatoria espacio-temporal de las variables (físicas, químicas y biológicas) que afectan a los procesos hidrodinámicos en las aguas subterráneas, lo que se traduce en una distribución aleatoria de los parámetros de los modelos y de sus ecuaciones. Dicha aleatoriedad es de dos tipos: epistémica, cuando puede reducirse aumentando la frecuencia de muestreo de un experimento; puramente aleatoria, cuando no se puede reducir añadiendo más información. A veces, se producen procesos hidrodinámicos a escalas tan pequeñas e imposible de caracterizar con los métodos tradicionales que se convierten desde un punto de vista práctico en parámetros puramente aleatorios. Sin embargo, si alguna relación constitutiva (o bien empíricamente, teóricamente o de base física) puede ser construido entre los procesos a escalas diferentes, entonces los procesos a pequeña escala puede ser reproducido por parámetros de modelo equivalente. La incertidumbre se debe entonces a variabilidad epistémica, y técnicas de caracterización a gran escala se puede utilizar para mejorar la descripción, interpretación o predicción de procesos pequeña escala. Esta tesis se ocupa de estos temas. El manuscrito se compone de dos partes principales (el primero sobre la recarga artificial y la segunda sobre transporte de solutos), cada uno de ellos dividido en tres capítulos. En el capítulo 1 de cada parte, se documenta una herramienta para obtener información cuantitativa una variable seleccionada con modelos a resoluciones gruesas. En el caso de la recarga artificial, las imágenes satelitales se utilizan para modelar la variabilidad espacial de la capacidad de infiltración en los suelos superiores con un detalle de escala métrico. En el caso del transporte de solutos, se muestra un nuevo método para estimar la densidad de distribución de partículas. En los capítulos 2, se explora qué procesos ocurren a escalas finas y que pueden afectar a la interpretación de la recarga artificial y de los procesos de transporte de solutos con modelos a escalas más grandes. En la primera parte, un método combinado que une a las imágenes de satélite y datos de campo con un modelo de colmatación de poros simple se utiliza para visualizar la posible evolución espacio-temporal de la capacidad de infiltración del suelo durante la inundación de balsa de recarga. En la segunda parte, modelos numéricos tridimensionales se utilizan para simular el transporte en medios heterogéneos bajo flujo radial convergente a un pozo a gran escala de detalle. Se demuestra que modelos 3D son más apropiados que modelos 2D para reproducir observaciones similares sobre la distribución temporal contaminante en un punto de control. También se proporciona una explicación física para describir el llamado efecto de anomalía a largos tiempos sobre las curvas de llegada que a veces se observa en la realidad. En el capítulo 3, los modelos se utilizan para definir la incertidumbre entorno a los parámetros de los modelos, de cara a la predicción y a la gestión de la recarga artificial y transporte de solutos. En el primer caso, se propone un análisis de probabilidad para definir el riesgo ingenieril de gestión de las balsas de recarga artificial, debido a la aleatoriedad de la distribución inicial de infiltración, que controla varios factores importantes de colmatación de poros según enfoques teóricos. En el caso del transporte de solutos, se discute cómo los parámetros equivalentes basados en modelos de transferencia de masa pueden estar relacionados con la distribución geométrica de los parámetros hidráulicos en formaciones anisótropas, cuando se usan ensayos de trazadores con flujo convergente radial.

This thesis has been financed by the Spanish Ministry of Education through the FPU-MED scholarship. The author is also grateful to the Ministry for providing funding for the visiting period at the University of Notre Dame (USA) through the 'FPU-MED Estancia Breve' program and scholarship. The Author also acknowledges the initial funding provided by the Hydrogeological Group of the UPC through internal projects.

Alla mia famiglia.

Contents

Chapter 1

Introduction

In the last century, the Earth has suffered from the major depletion and pollution of aquifers. Groundwater is a poorly renewable and extremely vulnerable resource, yet it represents about the 40% of the total fresh water in the Earth. In some cases, it provided the 100% of totally available fresh water in some semi-arid areas of the world ([Foster 2006]).

In the recent years we started to acknowledge the importance of groundwater. Until now, however, uncontrolled and undocumented drilling activities have affected most of the urban and semiarid aquifers all around the world (e.g. [Carrera 2005, Custodio 2002, Llamas 2002, Foster 2006]). Because of the slow groundwater movement, most of the negative effects of such a long uncontrolled situation in the past decades are still to be seen in the incoming future.

Aquifers are very slowly naturally regenerated by recharge and highly vulnerable to direct contamination from the surface. To maximize the groundwater use and to minimize the risk of contamination in the most sustainable and cost-effective manner, strategic decisions about hydrogeological management at short, mid and long terms need to be accurately made. However, this represents an uncertain task in most cases because of the complex hydrodynamic physical and chemical distribution of properties in the subsurface, mainly due to geological and geochemical heterogeneities.

Typically, the impossibility of complete characterization of geological heterogeneities render hydraulic variables controlling flow and solute dynamics as random parameters for modelling purposes ([Tartakovsky 2008]). In this sense, a probabilistic approach is advisable ([Krzysztofowicz 2001]), since it could then be integrated into a larger scale probabilistic risk assessment studies ([Tartakovsky 2007, Bolster 2009, de Barros 2008]). Randomness can be essentially of two types ([Christakos 2000]): epistemic, when it can be reduced by increasing the number of experiments; or, aleatory, when it cannot be reduced by increasing the number of experiments. A statistical framework is most of the time need to define the level of uncertainty of a decision.

From a practical perspective, it is better to deal with epistemic randomness. In this case, poorly characterized site can be additionally sampled to obtain more exhaustive descriptions and the uncertainty reduced about a variable. Some hydrogeological processes occur, however, at spatio-temporal scales that are unfeasible to observe and behave as aleatoric variable at macroscale. Yet, if some mathematical physically-based formulation can be built to pass from non characterizable to practically characterizable scales, such as using upscaling or downscaling procedure,

then these variable these models can be used . Evidently, these working models need to be properly developed and checked, since the modeling scaling issue is very uncertain in most hydrogeological cases.

Example of these multiscale processes are touched in this thesis. Those defining clogging or anomalous solute transport mechanisms, for instance, occur at the pore scales, which cannot be measured because of lack of sufficient detail; those describing the full spatial variability of an hydraulic property, such as infiltration capacity or geological anisotropic structure, can occur and show continuity over tens or hundreds of meters, and thus require too much information that cannot be obtained for cost-limiting reasons.

An additional issue that is explored in this thesis is how efficient numerical tools, such as image processing of remote sensing images or particle tracking to reproduce solute transport, can be adopted to efficiently reproduce multiscale dynamics. Efficiency means finding the optimum trade-off between detail of he outcome and computational cost, which can be a high price to pay in most cases that render the mathematical treatment of a problem unfeasible to solve for in the practice.

1.1 Thesis goal and outline

This thesis is focused to illustrate tools to deal with hydraulic heterogeneity in an effective manner, and to understand processes in groundwater to reduce the effects of uncertainty in decision-making processes and risk assessment. Two topics are analyzed in details: aquifer artificial recharge by means of superficial ponds is analyzed in the first part of the thesis; radial convergent transport of solutes is studied in the second part. Each part is composed of three chapters, following this sequence.

(1) The first chapters describes the 'tools' used to obtain spatial and temporal distribution of variables. In the first part, satellite images are used to obtain spatial description of the infiltration capacity in preclogging states. In the second part, kernel density estimation are coupled to particle tracking to obtain accurate probability density functions of solute concentrations.

(2) The second chapters deal with the 'analysis' of the process occurring during artificial recharge activities and during the transport characterization of an aquifer by means of tracer tests, when heterogeneity is accounted for but it cannot be directly measured or reproducible. In the first part, the effects of an accurate selection of lumped temporal models for clogging are evaluated to obtain reliable spatio-temporal variability of the infiltration capacity from data. In the second part, analysis of processes generating macroscopic anomalies on breakthrough curves during convergent flow tracer tests are analyzed.

(3) The third chapters account for the 'application' of these finding for predictive purpose and accurate management of natural resources. In the first part, an engineering-risk-based analysis is developed to deal with the effects of spatial variability of topsoils hydraulic properties of artificial recharge ponds; in the second

part, the effects of anisotropic statistical distribution of geological properties are analyzed to evaluate the resulting distribution of apparent mass-transfer parameters used to upscale and predict solute transport in heterogeneous aquifers.

Each part of the thesis ends with the general conclusions obtained from each of the three chapters composing the part. A final global conclusion is also reported at the end of the thesis. The single chapters are based on manuscripts oriented for scientific journals. The references to each manuscript is given at the beginning of each chapter.

Note that in the thesis reference keys are always defined by the surname (or enterprise name) of the first authors and year, despite many authors can have contribute to the publication. The complete bibliographic references are reported in the Bibliography at the end of the thesis.

Part I

Aquifer artificial recharge

Prologue to part 1

Managing water resources under scarcity is a necessity in many arid and semi-arid regions worldwide [Gee 1988, Bouwer 2002, Gale 2005, Scanlon 2006]. Amongst the many practices that exist to increase groundwater availability as well as to improve water quality in a given area, artificial or induced recharge practices (AR) are viable options [Dillon 2002, Greskowiak 2005, Jha 2009]. A common system to induce recharge is via excavated surface ponds (SP). These facilities are popular in developed and developing countries (e.g. [O'Shea 1981, Asano 1985, Hofkes 1986, Ting 2002, Tuinhof 2003, Aish 2004, Stonestrom 2007, Cheng 2009]).

In an artificial SP facility, recharge is performed by diverting available water (e.g. reclaimed, storm water, river water) into the pond and letting it infiltrate naturally from the top surface to the underlying aquifer. Excavations are flooded with water coming from any available source (e.g. reclaimed water, stormwater, river water). In a properly designed facility, evaporation and other losses are typically negligible compared with the infiltration rates, and shallow water depths are typically preferable ([Bouwer 1988]).

The maximum rate at which water can infiltrate in the subsurface is known as the infiltration capacity (I_c). It regulates both the total amount of infiltration towards the aquifer and some characteristic times (such as the water residence time within an SP) which are needed to make appropriate management of the facilities (e.g. [Perez-Paricio 2000]).

Methods for assessment I_c are quite different than those estimating total aquifer recharge, such as the ones evaluated using e.g. energy or chemical mass balances using analytical or numerical methods([Hantush 1967, Caro 1981, Rao 1983, Latinopoulos 1986, Das Gupta 1988, Finnemore 1995]). In effect I_c is mostly controlled by the first few centimeters of soil. Most of the times, estimations of I_c are done directly, such as using infiltrometers ([Smith 1972]), or indirectly, such as from pore- or grain-size distributions (PSD, GSD) ([Brooks 1964, van Genuchten 1980]).

I_c varies in space and time due to soil heterogeneities, clogging processes and temperature fluctuations, as well as other processes that affect the aquifer topsoil (i.e $I_c \rightarrow I_c(\mathbf{x}, t)$). Most of these phenomena are uncertain to predict as they occur at not easily characterization scales.

Spatial descriptions of $I_c(\mathbf{x}, t)$ are usually uncertain since data are usually very limited to a few points in space, and are not exhaustive for adequate spatial description of the infiltration capacity of the topsoil is required to guarantee successful long-term performance of AR practices.

However, although $I_c(\mathbf{x}, t)$ generally varies in space and time, typically the variable of true interest in managing operations is its spatially-averaged counterpart $\overline{I_c}(t)$. At a given time, $\overline{I_c}(t)$ can be estimated either statistically from small-scale local measurements if a sufficient amount of spatially-distributed data are available, or experimentally with large-scale infiltration tests (e.g. [Abu-Taleb 1999, Barahona-Palomo 2010]).

Because of temporal processes that influence $I_c(\mathbf{x},t)$, however, such a map should be update periodically. The most common effects for temporal variation of the infiltration are natural reactions related to a variety of clogging mechanisms that together reduce $I_c(\mathbf{x},t)$ and $\overline{I}(t)$ during pond flooding stages. Clogging is a combination of physical, biological and chemical processes, which modify the properties of both the water and the soil. At a typical SP site, one observes a significant reduction in PSD and GSD within the first few centimeters of the soil due to clogging. Seasonal variations in the density and viscosity of water may contribute to the temporal variability of $\overline{I_c}(t)$, but these effects are expected to be secondary and act on much larger time scales (months) than those relevant to clogging (days). These topics have been widely discussed in the literature (e.g [Baveye 1998, Vandevivere 1995, Perez-Paricio 2000, Bouwer 2002, Greskowiak 2005, Civan 2007, Civan 2010])

The rate of natural clogging mechanisms at small scales is highly uncertain. Most approaches have been suggested in the past to assess the effect of independent individual mechanisms (a detailed analysis on clogging processes and models is offered in Chapter 4). However, a fundamental question is how to evaluate the effect on pore clogging that encompasses the mutual dependence of all clogging mechanisms. Although several published field and laboratory experiments exist that aim to understand and quantify their mutual interaction depending on soil type [Ernisee 1975, Avnimelech 1983, Rosowski 1986, Kim 2010, Hoffmann 2010], predicting the clogging rates via cumulation of individual mechanisms remains, in the practice, challenging. The reasons are that : (i) reactions leading to a (macroscale) clogging effect take place at microscales (e.g. [Baveye 1998]) that cannot be easily measured, or that cannot be easily upscaled to the field pond scale and (ii) the rate at which clogging mechanisms jointly develop depends on a variety of site-specific conditions and factors, such as the initial textural, chemical and density heterogeneities, or water temperature fluctuations ([Guin 1972, Tien 1979, Clement 1996, Greskowiak 2005, Civan 2007, Civan 2010, Zamani 2009]).

Making adequate decisions for managerial application about artificial recharge ponds relies therefore in the correct assessment of the spatio-temporal local variability of $I_c(\mathbf{x},t)$ and $\overline{I_c}(t)$, which is intrinsically uncertain. In the following chapters, three aspects are analyzed to deal with this issue.

In Chapter 2, a remote-sensing approach is used to estimate the spatial variability of the infiltration capacity in pre-clogging conditions (i.e. $I_{c0}(\mathbf{x})$. This method allows to infer the exhaustive description of this variable at any point of the domain, and its reliability is based on the quality of the correlation between data and pixel values, and on the resolution of the satellite image.

In Chapter 3, a method to couple direct measurements of $I_c(\mathbf{x},t)$ at varoius stage during the flooding stages of the pond, large-scale satellite information and a clogging model is presented. This method allows to obtain spatio-temporal maps of $I_c(\mathbf{x},t)$, which is a requirement for adequate management of these facilities. Uncertainty is due to both the quality of the correlation between data and satellite image, as well as with the temporal model chosen to reproduce clogging.

In Chapter 4, the relationship between heterogeneity and uncertain management

of these facilities is evaluated within a probabilistic framework. $I_c(\mathbf{x}, t)$ is treated as a spatio-temporal random field. It is assumed that the clogging model depends directly on the first estimated spatial distribution of the infiltration capacity (again, obtained for instance using the estimation from the satellite images). Uncertainty derives from the random distribution of I_c at the initial conditions, which of course can be very different from site to site. For this reason, both single realization and a Monte Carlo framework are used to carry on the uncertainty analysis.

Chapter 2

Using satellite images to map the infiltration capacity of topsoils

2.1 Introduction

The goal of this chapter is to explore the potential use of satellite images as an efficient, low-budget and fast method to assess the spatially-variable infiltration capacity of topsoils in preclogging conditions ($I_{c0}(\mathbf{x})$). The aim is to build a relationship between pixel intensities and hydraulic properties in order to estimate $I_{c0}(\mathbf{x})$ over large domains using satellite images. As an illustrative case, a case study has been studied based on a pilot SP facility in the municipality of Sant Vicenç dels Horts (Barcelona, Spain) located in the silico-clastic and highly heterogeneous Llobregat alluvial aquifer.

Spatial variations of the infiltration capacity occur at scales much below the size of an infiltration pond. While in most time a single value of infiltration for the full pond is needed, this means involving some homogenization process, and some representative equivalent infiltration value has to be used instead of the local-scale infiltration capacity.

In most cases averaged values are found integrating the information about local $I_{c0}(\mathbf{x})$. Direct measurement are typically scarce and sparse, if not non-existent. Thus, it may be desirable to rely on secondary information using geophysics, remote sensing, image analysis, or any combination of the above. These methods provide dense information over large scales about (secondary) variables that are related to the (primary) variable of interest. The secondary variables' spatial structure can be more easily evaluated and used to infer that of the primary variable. In fact, variograms or covariance functions about the secondary variables can be implemented to study the spatial structure of primary data (e.g. [Gooverts 1997]).

Remote sensing is a relatively well-understood, successful and cost-effective solution to obtain qualitative estimations of recharge or related hydrogeological variables over large scales (e.g. [Saraf 1998, Granger 2000, Milewskia 2009]). However, very few cases in the literature have documented the use of such approaches for quantitative assessments of infiltration capacity. The use of photographic images is restricted to a few studies (e.g. [Chica-Olmo 2000, Reddy 1989]). Historically there has been

[0]This chapter is based on D. Pedretti et. al (2012), *A fast and inexpensive method for the quantitative assessment of the spatially variable infiltration capacity using satellite images*, Journal of Hydrology, vol. 430-431, pp. 118-126, doi 10.1016/j.jhydrol.2012.02.008

a major economic barrier as the acquisition of a typical sequence of satellite images analysis [Bons 1996], was a prohibitely expensive step.

This situation has changed in the recent years. Today, satellite and aerial images of relatively high quality can be obtained at a high resolution and affordable prices (or often no cost) from a variety of sources such as popular Internet-based map providers, cartographic or geological surveys, military research institutions, national and international space agencies, etc.

Satellite or aerial images can cover entire geological basins, and so can be extremely useful for hydrological studies. Often they can be obtained at different resolutions (supports). An image is composed of a fixed number of pixels with varying intensities. Digital images can be made up of several bands (or 'channels'), both in the visible range of colors (red to blue colors) and of non-visible bands (such as the infrared one). The combination and superposition of the visible colors give rise to the typical image one observes on a computer screen. In a single band, the relative intensity of each pixel depends on the electromagnetic energy reflected by the land surface. This is then processed by an acquisition device with a given sensitivity. Understanding the interaction between soil reflectance and the acquisition device is key to deriving information from remote-sensing data [Goetz 1985].

While technical descriptions about the characteristic features of the capturing devices (cameras, video recorders, radars) and monitor quality can be easily obtained from the technical spec sheets, quantifying the soil reflectance in heterogeneous media remains challenging. Some conditioning factors are for instance, moisture content, iron-oxide mineral content, organic-matter content, surface roughness, thickness and colonial organization of the vegetation canopy and grain structure and organization (see [Irons 1989, Okin 2004] for details).

Hydraulic properties of topsoils are also dependent on the same factors as soil reflectance. For instance, the characteristic grain size of a soil is related to the soil permeability [Hazen 1882, Vukovic 1992], as well as the soil porosity [Kozeny 1927, Carman 1938, van Genuchten 1980]). Moisture content is also influenced by both clay and organic-matter content. All this reasoning suggests that soil reflectance can be used to obtain information about hydraulic parameters of the topsoil.

The chapter is organized as follows: section 2.2 describes the site and the available experimental data on $I_{c0}(\mathbf{x})$; section 2.3 discusses the image analysis; section 2.4 shows an application of the proposed methodology.

2.2 The artificial recharge facility in Sant Vicenç dels Horts

A SP was constructed in the municipality of Sant Vicenç dels Horts near Barcelona to study managed artificial recharge practices in the Llobregat River Lower Valley aquifer. The purpose of this pilot area is as a research facility to study the fate of micropollutants during infiltration practices. The facility is located in the prodelta region of the Llobregat River. The site is centered at UTM coordinates 418446.63

North and 4581658.18 East (zone 31T). Different high-resolution satellite photos are available for the site at different years. Fig. 2.1 shows an image available to the public through a popular Internet-based map server, and captured on November 15th, 2007.

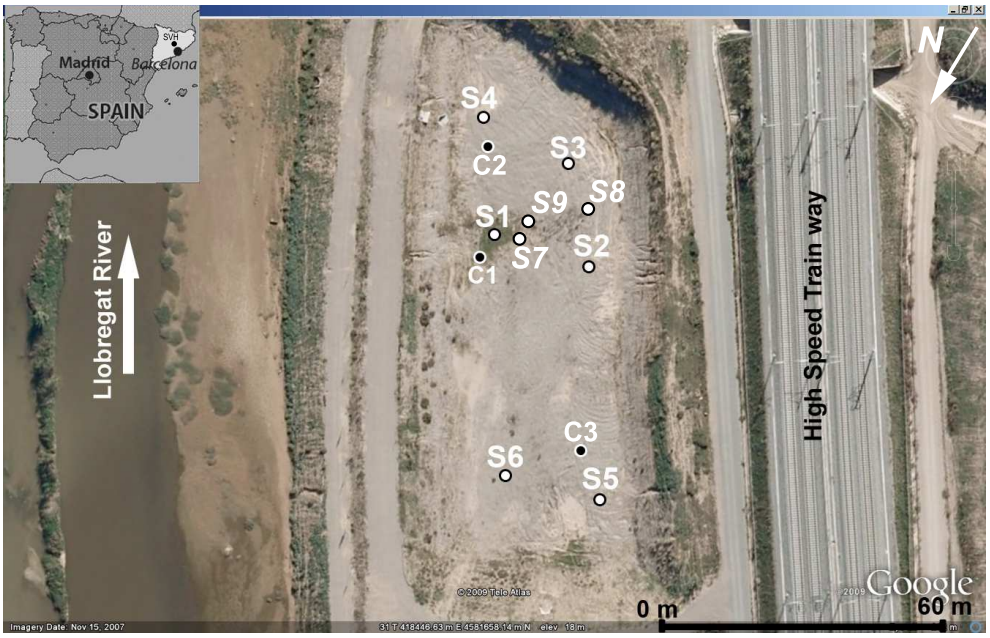

Figure 2.1: Aerial photo of Sant Vicenç dels Horts and the Barcelona municipality. The site is located in the Lower Valley, a few km up from the beginning of the delta. The Llobregat River can be traced in the picture. The UTM coordinates are 31T [418446.63 N , 4581658.18 E]

The geology of the area is a sequence of fine- and coarse-grained facies of siliciclastic materials, deposited according to the evolution of the paleoriver. The deposition of fine-grained materials occurs in low energy streams (minimum on the alluvial planes), while coarser material needs higher transport efficiency (maximum along the channel). Therefore the hydrogeological setting is composed of sandy-gravel or gravelly-sand (depending on the proportion of the average grain sizes), separated by non-continuous fine-grained layers. At the scale of the pond, the unsaturated zone has a thickness of between 8 to 10 meters. The excavation of the pond ranges from 4 to 6 meters below the actual ground surface, on the western edge of the Llobregat River. A series of field experiments were performed in the SP to assess the local heterogeneities of the topsoils. In the following sections, we summarize two of the most significant activities, the vertical geological description of the ground surface using open pits and a campaign of double rings infiltrometer tests. Both were performed before any MAR activity was carried out at the site (only natural rainfall actually infiltrated during this time).

2.2.1 Vertical geological profiles of topsoils

Three open pits were dug to study the vertical distribution of the geological materials at the upper meter measured from the bottom of the pond. Additionally, samples were taken to obtain the granulometry curve of the different materials described in the open pits. This enabled the qualitative inference of the hydraulic properties of the formation. Fig. 2.2 shows three vertical profiles obtained at locations C1-C2-C3 (see fig. 2.2). Visual abrupt changes of soil color (seen in the field) are marked in the figure by solid lines. These changes indicate that soil moisture content and grain distributions are layered in the top sections. Specifically, for C1, the 30-cm-thick top layer displays high moisture, clay and organic content, overlapping the other deeper layers which show coarse-grained materials at lower moisture content. Such differences are visible in the left most photo of fig. 2.2: the upper horizon is clearly darker than the rest of the outcrop. Outcrops from pits C2 and C3 show similar horizontal layering. In this case, fine-grained materials and organic content were not observed in the outcrop; the vertical variability is due to changes in sand or gravel relative content.

Just by looking at fig. 2.2 it follows that there should be a correlation between soil color and permeability, dark pixels being representative of the less permeable materials. This proposed correlation is explored in Section 2.3.

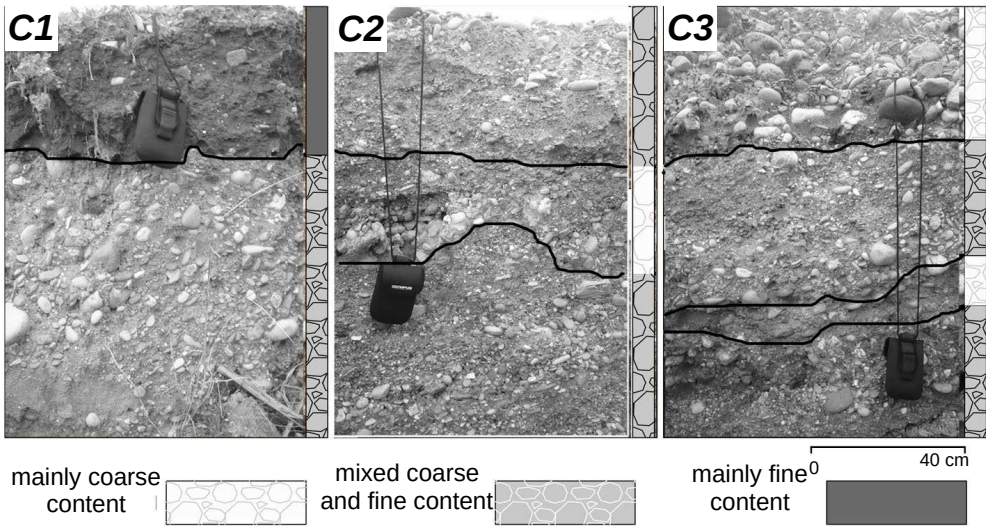

Figure 2.2: Vertical profiles at C1 (left), C2 (center) and C3 (right) excavated pits, located close to the double ring test zones S1, S4 and S5, respectively (see location of all these points in fig. 2.1). The right side of each picture shows the geological stratigraphic columns (legend at the bottom).

2.2.2 Field measurements of topsoil local infiltration capacity

Six double-ring infiltrometer tests [Bouwer 1986] were performed in February 2009 on the topsoil of the pond. The location of these experiments (S1 to S6 points in fig. 1) was randomly-selected. A double-ring infiltrometer technique was used. This technique has been well documented [Bouwer 1986] for the direct measurement of infiltration rates and its applicability has been assessed and validated for several ground conditions (e.g. [Bodhinayake 2004]).

We briefly describe the method here. The device consists of two concentric thin-walled metal cylinders, with an approximate height of 40 cm, in which a falling head test is carried out. The test consists of three parts: (a) the rings are pushed into the first (five to ten) centimeters of topsoil with minimum soil disturbance; (b) both rings are filled with water to the same initial level; (c) the change in water level (decrease) in the inner ring is measured over time. The purpose of the external ring is to minimize lateral flow occurring under the internal ring and ensures primarily vertical flow. After a standard time of about 2 hours, the soil is saturated and the infiltration rates (i.e., changes of water heights versus variable time intervals) tend asymptotically to a quasi-steady constant value. The actual infiltrated volume versus time curve is interpreted by means of a modified Kostiakov method [Smith 1972]. Vertical infiltration rate is then determined by the amount of water poured into the inner ring per unit of surface area and time. The inner ring diameter is 0.4 m and as such our tests provide the infiltration capacity of the S-location over a support area of 0.13 m^2.

For illustrative purposes we plot the infiltration curves for three locations (S1, S2 and S6) in fig. 2.3. These three are representative of the low (S1), middle (S2) and high (S6) rate infiltration areas. Complete results are summarized in the first two columns of Table 2.1. Note that local infiltration rates span over at least two orders of magnitude in this domain.

2.3 Image analysis

As reported in Section 2.2, heterogeneities occur at the site in both vertical and horizontal directions, with different characteristic scales. MAR facilities need to adequately map the spatial distribution of the local infiltration capacity $I_{c0}(\mathbf{x})$ over the entire pond since we are interested in total recharge as a function of time (i.e. the spatially-averaged infiltration capacity $\overline{I_{c0}(\mathbf{x})}(t)$).

The spatial structure of $I_{c0}(\mathbf{x})$ cannot be inferred with great confidence only from the information obtained at a few sparsely distributed data points (such as S1-S6 from the February 2009 campaign). Additional information, either on $I_{c0}(\mathbf{x})$ or else on a related secondary variable must be sought for this purpose. While the former can be expensive or challenging for a variety of reasons, secondary information can be used to condition the primary information.

Secondary variables are typically related to primary variables via mathematical and physically-based models; the density of secondary data is normally sufficient for

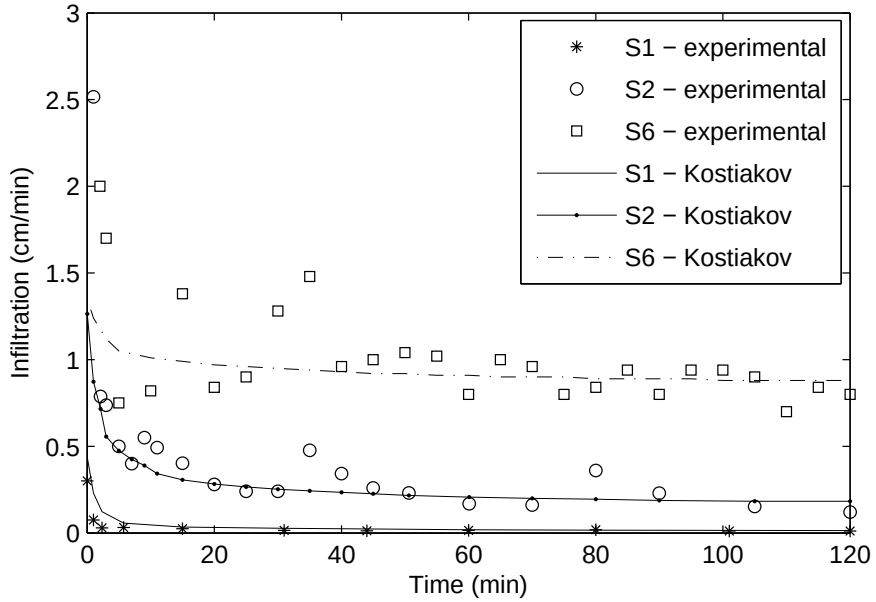

Figure 2.3: Experimental measurement of local infiltration rates from double-ring infiltration tests and interpretation using the modified Kostiakov method, at three location S1,S2 and S6. $I_{c0}(\mathbf{x})$ is calculated as the infiltration rate measured after 120 minutes. The analytical curves proposed by Kostiakov [Smith 1972] are shown for reference. The actual value used in Table 2.1 is that inferred from the analytical curve at t = 2h, rather than the measured values itself, in order to avoid noise.

detailed spatial descriptions, and with some approximations, can be directly used to model the primary information [Journel 1999]. In most cases the correlation between the two variables is not perfect and some error (either correlated or not) must be included in the model.

Our conjecture is to use satellite images of our pond to extract secondary information, and infer primary information based on the relationship between the measured $I_{c0}(\mathbf{x})$ and the pixel intensity of the image at the test locations. Recalling fig. 1 and Table 2.1, we can see that point S1 corresponds to the lowest $I_{c0}(\mathbf{x})$ value and is located in a green area in the two images (dark in gray colors). Such visual differences are no longer appreciable to the human eye (or at least our eyes) for locations with high $I_{c0}(\mathbf{x})$ values, since it is more difficult to distinguish bright colors. Nonetheless a relationship appears to exist.

2.3.1 Description of the digital data for the Sant Vicenç dels Horts site

We analyze the relationship that exists between $I_{c0}(\mathbf{x})$ and the pixel intensities of two types of digital images, coming from two different sources. The first is obtained

by an Internet-based map software, while the second is an image provided by the local cartographic institute. While the latter may suppose an initial, sometimes prohibitive, investment, we aim to achieve good-quality information with former, which is completely free and is suitable for for fast and cheap assessments.

In our first case, the Internet image (the same in fig. 1) is obtained by capturing an image as a standard RGB (Red, Green and Blue channels) raster (saved in TIFF file format). For convenience, we choose to work only on a rectangular portion of the image (marked in fig. 1), which corresponds roughly to an area of 45×100 m^2. This area is represented by 326×730 pixels, so that each pixel corresponds to 0.02 m^2. The raster is then composed of a total of $2.38 \cdot 10^5$ pixels per band. This method is deliberately quick and simple to illustrate that it can be cost effective and quick. However, it must be noted that the method can suffer from a lack of precise information due to image compression, resolution problems and filtering that may have been performed by the Internet map software programmer or with other processing software to save the TIFF file. In this case, this provider does not appear to provide sufficient information about the image to know exactly what filtering processes and image editing were performed.

On the other hand, the second digital image of the exact same space at the exact same time is provided directly by the local Cartographic Institute of Catalonia (ICC). The image is a non filtered RGB bands raster, in non compressed TIFF format, with a pixel resolution of 25 cm^2. We refer to this image as the 'original' image, since all specifications are well known from the source. In fact, it is known that the Internet map provider used images from the ICC to build their software and images.

For comparison purposes, both digital images are taken as the same moment in time and considered the same working area. The color depth of both digital images is of 8 bits, offering $2^8 = 256$ values on intensity per channel. The range is from 0 (black = minimum intensity) to 255 (white = maximum intensity) .

Fig. 2.4 displays histograms of the pixel intensities for each X color channel (X=R,G or B), and histograms of averaged values over the three bands, for both images. For the Internet image, the color intensities in all channels do not cover the entire range of possible values, demonstrating a potential filtering that took place. Moreover, the statistics for each of the three channels are different. Let $m(X)$ and $\sigma(X)$ be respectively the mean and the standard deviation of the pixel values for each X channel, calculated from the histograms. The red (R) color intensities range from 35 to 225 with $m(R) = 184.64$ and $\sigma_R = 16.8$. The Green (G) and Blue (B) channels exhibit similar standard deviations ($\sigma(G) = 16.0$ and $\sigma(B) = 16.9$) but with smaller means $m(G) = 174.1715$ and $m(B) = 162.4$). Notice that the actual distribution of pixel data is quite symmetric, showing slightly positively skeweness and a leptokurtic effect. To test it visually, we plot an equivalent Gaussian bell with the same $m(X)$ and $\sigma(X)$ in each band of the Internet based image.

The original ICC-provided image exhibits similar histograms of color intensities but with higher variances than the Internet-provided image. Here, $m(R) = 183.6$, $\sigma(R) = 36.7$; $m(G) = 179.5791$, $\sigma(G) = 34.8$; $m(B) = 157.7$, $\sigma(B) = 33.3$. These

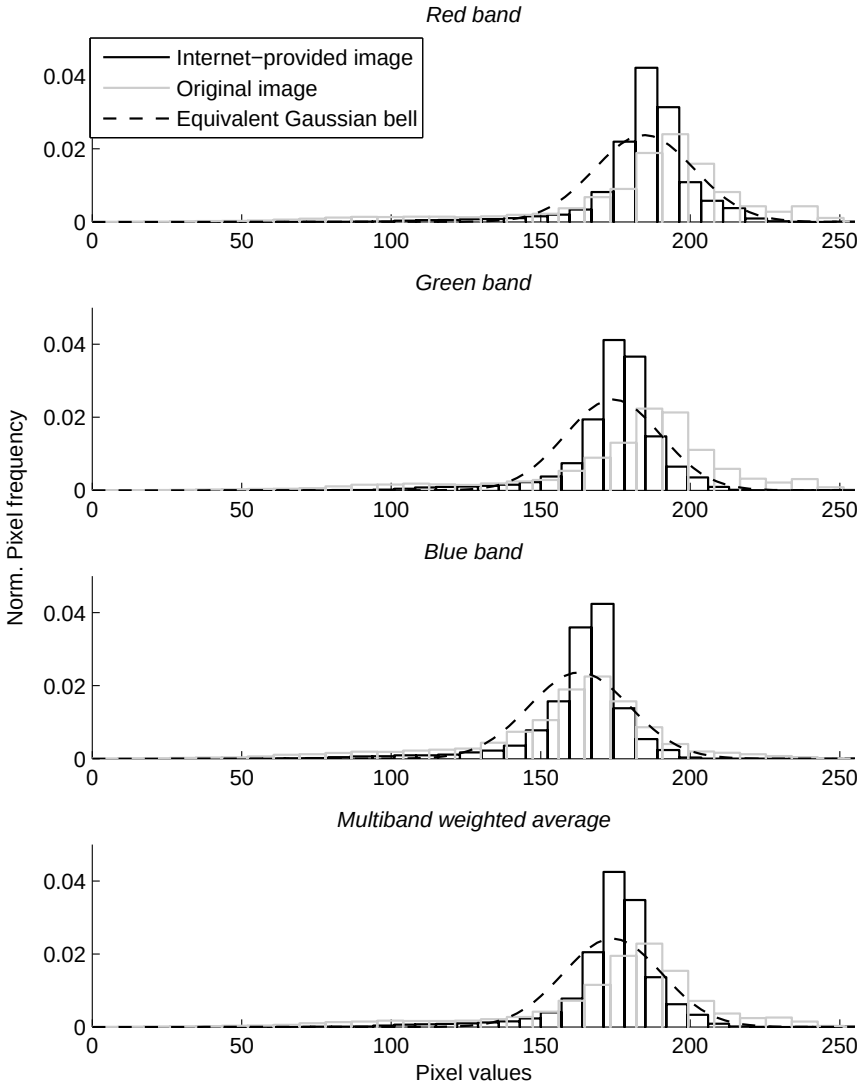

Figure 2.4: Density of pixels in the three bands of the sampled image. A Gaussian curve (dotted line) is fitted to the histograms of the Internet-provided images to test visually the symmetry of the distributions

histograms again display a positive skewness.

2.3.2 Accounting for support scales dissimilarities

A quantitative analysis of the correlation between $I_{c0}(\mathbf{x})(x,t)$ and color intensities from digital images can be visually inspected with a scatter plot. However, the spatial support of the infiltration test (0.13 m^2) is rather large compared to the pixel resolution (around 0.02 m^2). In order to make comparable predictions we averaged the pixel intensities in the digital images over a window of 3×3 pixels. The procedure is graphically explained in fig. 2.5).

For each Sj site (S1 to S9) and for each X band we estimated the local mean $\mu_{j,X}$ and relative local standard deviation $\sigma_{\mu_{j,X}}$ over a 3×3 pixel window centered at j. The variance is a measure of the quality of the estimation, which depends on the variations observed within the given window. These values are reported in Table 2.1.

2.3.3 Multiband regression model

We formulate a general regression model which considers the quality of data varying with each observation location in each color channel. This is done to include different degrees of confidence which are implicitly associated with the regression model of each color channel. In our case (fig. 2.4), we see that, for each observation window and image, the mean pixel values ($\mu_{j,X}$) are in most cases similar among color channels. However, the quality of the estimation of $\mu_{j,X}$ given by its standard deviation $\sigma_{\mu_{j,X}}$ varies with the observation, so that the regression coefficients depend on the image source and color channel. In an attempt to incorporate these effects, we generalized the previous methodology as it follows. Noting that the Pearson's correlation, r_X^2, measures the goodness of fit to a linear regression model in each X band, for each j-point we estimated the color intensity as a weighted average over the color channels, $\overline{\mu}_j$, such that

$$\mu_{j,\overline{X}} = \frac{r_R^2\ \mu_{j,R} + r_G^2\ \mu_{j,G} + r_B^2\ \mu_{j,B}}{r_R^2 + r_G^2 + r_B^2} \tag{2.1}$$

The results are shown in Table 2.2. We found that a linear regression model can satisfactorily describe the dependence of the natural logarithm of $I_{c0}(\mathbf{x})$ and averaged pixel intensity $\mu_{j,\overline{X}}$. The general form of the equation is

$$\ln(I_{c0}) = a(\overline{\mu}) + b + \varepsilon \tag{2.2}$$

Note that color intensities were found to follow a quite symmetric distribution (fig. 2.4). Despite they do not show an exactly Gaussian behavior we could in principle assume that the related infiltration capacity roughly follows a log-normal distribution, as given by (2.2). Log-normal distribution model is a typical adopted for hydraulic conductivities in soils (e.g. [Freeze 1975]), to which $I_{c0}(\mathbf{x})$ strictly depends.

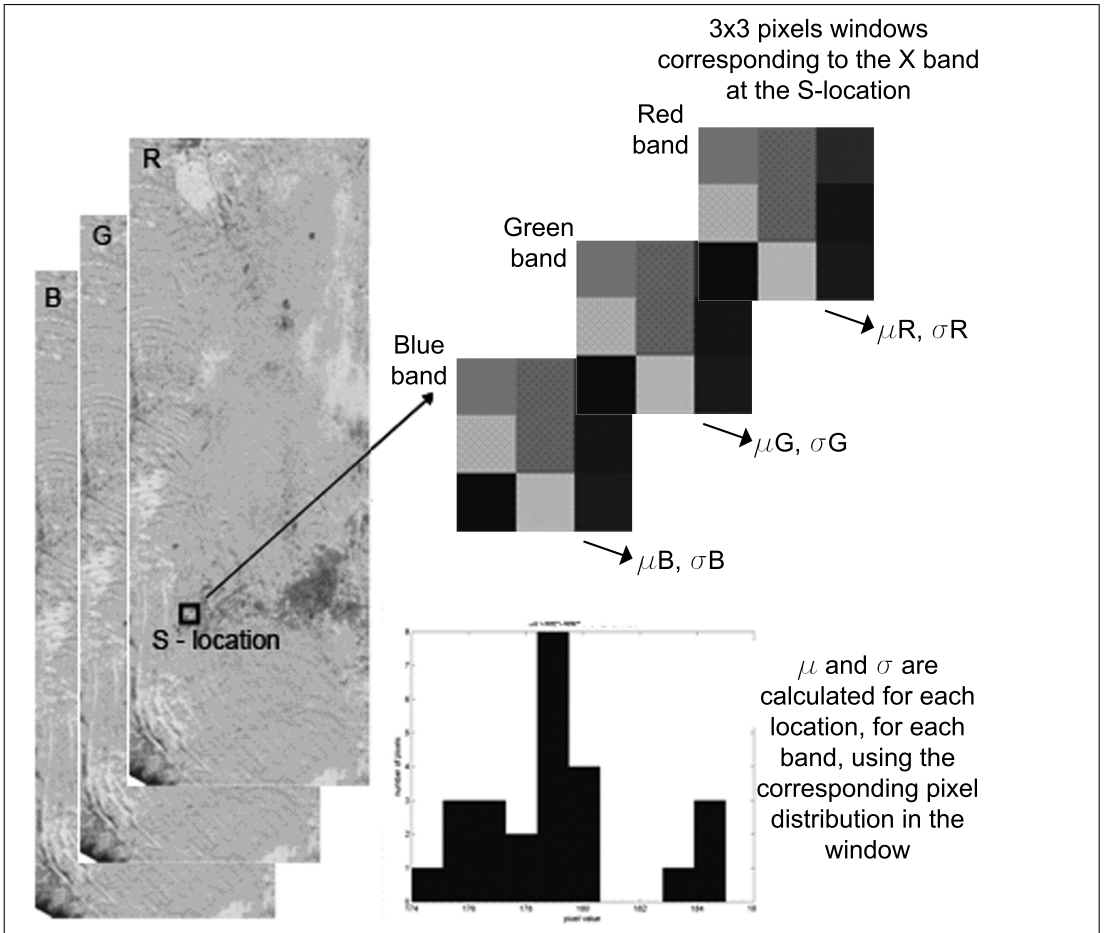

Figure 2.5: Assigning pixel intensities to a given test j-location (S_j) for each X band (R, G or B). The point location of the test is assigned to a given pixel in the image. Intensity values are obtained as the average over the 3×3 pixel window centered at this reference pixel ($\mu_{j,X}$); the standard deviation ($\sigma_{\mu_{j,X}}$) of the 9 pixel intensities referred to $\mu_{j,X}$ are reported in Table 2.1.

In (2.2), ε expresses the model error. Since the quality of the estimates of $\mu_{j,X}$ is not constant across observations, other regression methods such as weighted least squares should be used. In this method, the measurement error is weighted based on its corresponding degree of confidence. Here, we estimated these weights as inversely proportional to the quality of the local multiband estimation $\overline{\mu}_j$, such that

$$\overline{w}_j = \frac{1}{\sqrt{\sigma^2_{\overline{\mu}_j}}} \tag{2.3}$$

where $\sigma^2_{\overline{\mu}_j}$ is the multiband estimation variance calculated as

$$\sigma^2_{\overline{\mu}_j} = \frac{r_R^4\ \sigma_{\mu_{j,R}} + r_G^4\ \sigma_{\mu j,G} + r_B^4\ \sigma_{\mu j,B}}{\left(r_R^2 + r_G^2 + r_B^2\right)^2}, \tag{2.4}$$

Estimates of the variances for all measurements are reported in Table 2.1. Fig. 2.6 shows the resulting fitted regression models for each digital image source. For Internet-provided image

$$\ln(I_{c0}) = 0.0380\overline{\mu} - 5.244 \qquad r^2_{\overline{X}} = 0.85, \tag{2.5}$$

while for the original image we obtain

$$\ln(I_{c0}) = 0.0343\overline{\mu} - 4.466 \qquad r^2_{\overline{X}} = 0.93. \tag{2.6}$$

2.3.4 Model Validation

The regression model was validated against three independent double-ring infiltrometer measurements obtained during a second campaign. Their locations are also denoted in fig. 2.1 and marked by S7, S8 and S9 tags. These new experiments were specifically selected to fill gaps in the linear regression model. The infiltration tests for these 3 locations took place in June 2009.

Unfortunately, the new measures of $I_{c0}(\mathbf{x})$ could not be used directly. Between the two campaigns, a large flooding test took place in the pond, resulting in a net decrease of infiltration capacity due to clogging processes. The impact of flooding was analyzed by repeating the previous infiltration tests at the S1-S6 locations. A good linear relationship was found between the pre-clogged and post-clogged infiltration values, which allowed us to correct the June 2009 data to the values corresponding to February 2009 (see the following Chapter 3 for details). These corrected value are reported in Table 2.1. In fig. 2.6 they are indicated with a star beside the points. We see that they agree quite well with our regression model, lying within the region of confidence (expressed by the 95% confidence boundaries of the multiband regression model).

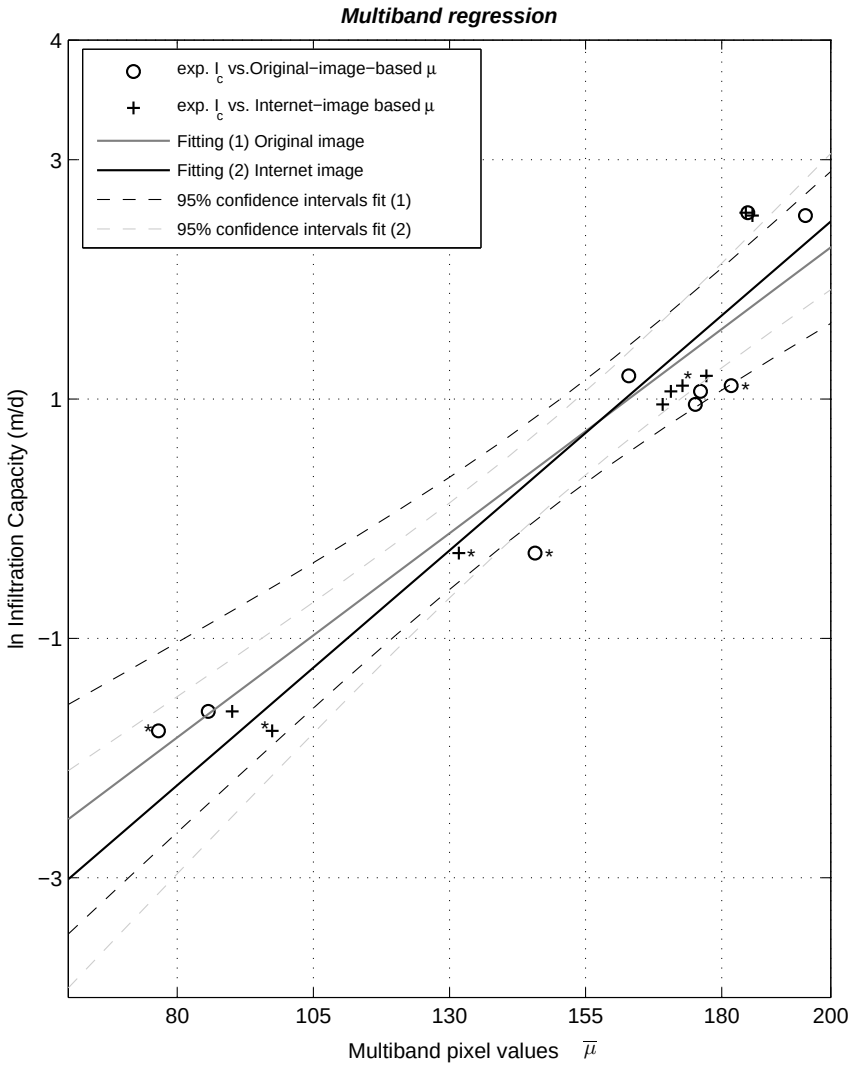

Figure 2.6: Linear fitting of the experimental tests of the infiltration capacity ($I_{c0}(\mathbf{x})$, in natural log scale) versus the multiband weighted average pixel values ($\mu_{\overline{X}}$), calculated at the 3×3 window around the S_j experimental site. Weights are given by the relative quality of the information, based on the multiband window variances $\sigma^2_{j,\overline{X}}$ - see text. The region of confidence is expressed by the 95% confidence boundaries.

2.3.5 Single band regression

A special case of the multiband model is when only single bands are accounted for. Such models (single-band regression models) are a general form of (2.2) and are especially useful when one specific channel or band provides a better fit than the others (provided physical conditions are fulfilled).

For instance, infrared bands (when available) have been used to evaluate specific patterns of soil moisture (e.g. [Price 1980]). This information could be related to some characteristic hydrodynamic property of the soil, thus making the correlation more robust. In our case, we explore the quality of a linear regression for the three visible-color bands, separately in each image. This is simply done by taking the single channels parameters in (2.2). The single band model is thus

$$\ln(I_{c0}) = a_X \overline{\mu}_X + b_X + \varepsilon_X \tag{2.7}$$

where the regression coefficients are now referred to the selected band X. We found that a good correlation exist for the three color channels in each image. In the Internet-provided image,

$$\left.\begin{array}{lll} a_R = 0.0383, & b_R = -5.327, & r_R^2 = 0.94, \\ a_G = 0.0452, & b_G = -6.342, & r_G^2 = 0.91, \\ a_B = 0.0393, & b_B = -4.985, & r_B^2 = 0.92, \end{array}\right\}$$

while for the original image,

$$\left.\begin{array}{lll} a_R = 0.0320, & b_R = -4.586, & r_R^2 = 0.90, \\ a_G = 0.0335, & b_G = -4.758, & r_G^2 = 0.878, \\ a_B = 0.0350, & b_B = -4.094, & r_B^2 = 0.840, \end{array}\right\}$$

We observe that the parameters differ slightly from one other due to small differences in the histrogram distributions ($m(X)$ and $\sigma(X)$). However, r_X^2 is greater than 84% for each of the three channels, which suggests that a good correlation exists for each case. r_B^2 for the Internet image is slightly larger than the for the other cases and perhaps this might suggest this is the preferred image and band. However, given that the differences in r^2 are so small, in the following, we apply the general multiband model for illustrative purposes.

2.4 Estimates of Infiltration Capacity at the pond scale

The methodology described above is used to estimate the spatial distribution of $I_{c0}(\mathbf{x})$ at the SP pilot site. Results are only shown for the multiband regression model, but in this particular example plots would be qualitatively identical and quantitatively similar using the single-band model. The resulting infiltration map stemming from the multiband regression analysis of the Internet digital image is

shown in fig. 2.7. The local infiltration capacity is calculated using (2.2) at each pixel in the whole image. As a way to evaluate the adequacy of the model, we compared estimates of the global infiltration rate as calculated from the Internet image with observations of the total maximum infiltration recorded during a flooding test performed at the SP site by local water authorities from March to May 2009. The global infiltration rate ($\overline{I_{c0}}$) is calculated as the spatial average of the local $I_{c0}(\mathbf{x})$ over the Ω area, as

$$\overline{I_{c0}} = \frac{1}{\Omega} \int_{\Omega} I_{c0}(\mathbf{x}) d\mathbf{x} \tag{2.8}$$

During this test, experimental value of the total infiltration rate were calculated using a water balance in the pond (conservation of mass in the pond including evaporation). The total maximum infiltration rate was reached after 40 days (when the pond was completely flooded) and was measured to be 3.6 m/day. Predictions using the multiband model are of $\overline{I_{c0}}$ = 4.47 m/d for the original image and $\overline{I_{c0}}$ = 3.92 m/d for the Internet image. We observe that the latter prediction provides a relatively good agreement with the experimental data, as the error is less than 10%. We deem this to be an acceptable error given the trade-off between the cost of the analysis and the estimation error. It is worth noting that a more proper and detailed assessment of errors should follow a rigorous evaluation of several factors, including errors in the calculation of the global mass balance. In the latter case, for instance, the combined effects gas clogging, incomplete pond flooding, errors in the measurements of the discharge rates of entry water, etc. lead to measurement uncertainty. This has not been considered in detail here.

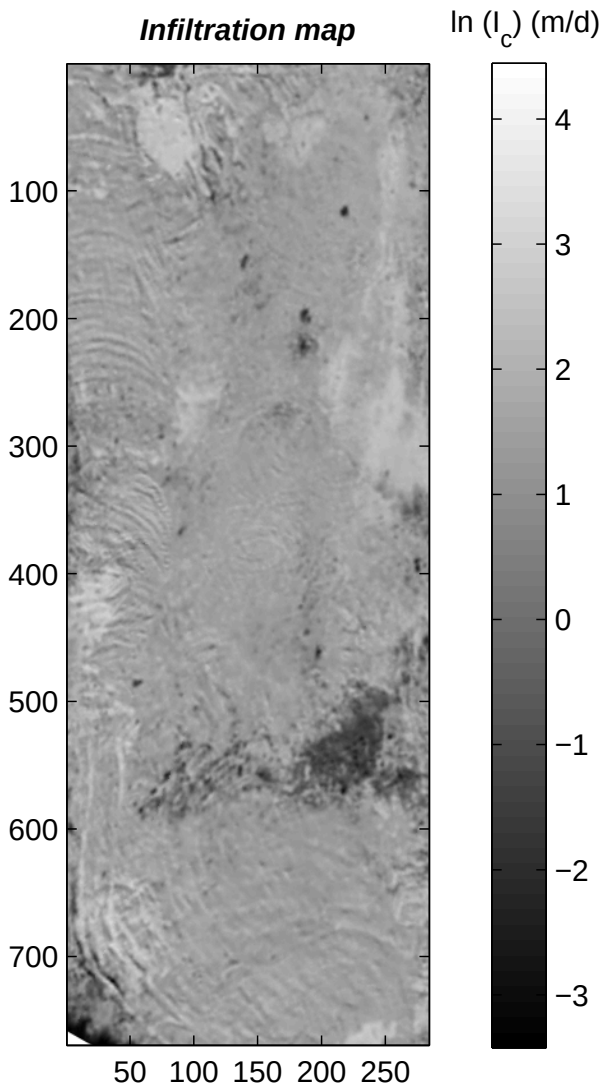

Figure 2.7: Map of the local infiltration capacities (in m/d) calculated using the multiband regression model based on the Internet image pixel values shown in Fig. 1. The natural log scale is used to highlight the heterogeneous distribution of soil hydraulic properties over the studied area.

Table 2.1: Field measurements of infiltration capacities ($I_{c0}(\mathbf{x})$, in m/d) from the double ring tests at the specified locations (S_j) and the correspondent average ($\mu_{X,j}$) and standard deviation ($\sigma_{X,j}$) of the 3x3 window pixel values for each X band (R,G or B) at the j-site (j=1,...,9). Above, the data refer to the Internet-provided image; below, the data corresponding to the original image are reported. For the full color map, M_j is the arithmetic mean of the pixel values when all bands are considered together at the j location, and w_j = weight from the regression analysis at the j location. The infiltration tests for S7, S8 and S9 were performed in June 2009 and have been corrected to the values of February 2009 by means of linear regression (see text).

Internet image							
Site	I_{c0}	Red band		Green band		Blue band	
		$\mu_{R,j}$	$\sigma_{R,j}$	$\mu_{G,j}$	$\sigma_{G,j}$	$\mu_{B,j}$	$\sigma_{B,j}$
S1	0.2	94.1	2.29	98.9	1.83	77.6	5.69
S2	2.6	179.1	1.83	168.1	4.39	160.3	4.82
S3	2.9	181.1	4.48	171.0	0.60	160.0	9.16
S4	3.3	187.3	0.50	177.9	5.21	166.4	8.81
S5	12.9	196.9	4.98	178.2	1.58	178.2	5.18
S6	12.6	196.0	1.58	185.0	5.05	176.0	6.17
*S7	0.17	103.2	5.21	104.7	1.80	84.7	4.53
*S8	3.04	183.3	1.80	172.3	5.42	162.7	5.00
*S9	0.75	142.4	5.41	132.4	7.12	120.4	9.28

Original image							
Site	I_{c0}	Red band		Green band		Blue band	
		$\mu_{R,j}$	$\sigma_{R,j}$	$\mu_{G,j}$	$\sigma_{G,j}$	$\mu_{B,j}$	$\sigma_{B,j}$
S1	0.2	91.2	6.96	93.3	4.53	71.9	2.03
S2	2.6	193.5	7.10	179.1	4.83	151.2	2.45
S3	2.9	179.6	10.32	191.6	6.94	156.2	4.39
S4	3.3	181.4	11.89	160.9	7.24	145.2	0.52
S5	12.9	198.9	4.68	193.1	5.18	160.9	5.21
S6	12.6	202.4	6.64	203.3	4.16	179.4	1.58
*S7	0.17	83.8	4.47	86.3	3.77	58.6	5.22
*S8	3.04	198.0	5.09	195.0	5.91	150.3	1.99
*S9	0.75	141.9	10.04	149.0	7.96	146.4	5.41

Table 2.2: Multiband average ($\overline{mu}_j$) and calculated multiband weight ($\overline{w}_j$) among the X band at the 3×3 windows around the j-th experimental S-sites (j=1,...,9). Data refer to the Internet and the original images

Site	Internet image		Original image	
	$\overline{mu}_j$	$\overline{w}_j$	$\overline{mu}_j$	$\overline{w}_j$
S1	90.1	0.50	85.7	0.18
S2	169.2	0.34	175.1	0.19
S3	170.7	0.28	176.1	0.12
S4	177.2	0.34	162.9	0.11
S5	184.4	0.24	184.8	0.21
S6	185.6	0.32	195.4	0.18
*S7	97.4	0.24	76.5	0.25
*S8	172.8	0.30	181.7	0.20
*S9	131.7	0.17	145.7	0.12

Chapter 3

Spatio-temporal mapping of infiltration capacity using limited experimental data and geostatistics

3.1 Introduction

Complete spatio-temporal mapping of $I_c(\mathbf{x}, t))$ is crucial for the optimal operation, management, and maintenance of the artificial pond (as shown in Chapter 4). However, it is a cumbersome task, since it is not realistically feasible using direct measurements and models are uncertain.

During recharge processes, $I_c(\mathbf{x}, t)$ reduces in time because of a variety of clogging processes that tend to modify the pore size distribution (PSD) and the grain-size distribution (GSD) of the top soil. The topsoil hydraulic conductivity $K_s(\mathbf{x}, t)$ depends on the PSD and GSD (e.g. [Kozeny 1927, Carman 1938, Hazen 1882]), which in turn controls $I_c(\mathbf{x}, t)$. To complicate matters even further, the reduction of infiltration also can depend on some initial properties of the soil. A detailed assessment of theories to compute clogging is given with references in the following section and in Chapter 4.

This chapter is devoted to explain how a geostatistical approach can be used to obtain exhaustive map of the spatio-temporal distribution of the infiltration capacity $(I_c(\mathbf{x}, t))$ on an artificial recharge pond during a flooding activity. The approach is developed to integrate a limited dataset of direct information about clogging development, secondary information about the initial infiltration capacity based on satellite images, and a physically-based lumped exponential model for clogging.

The chapter is structured as follows. In Section 3.2 we briefly recall some aspects concerning the use of clogging model and spatial analysis to evaluate the combined spatio.-temporal variability of I_c. In Section 3.3 we propose a local infiltration model with an exponential decay to an asymptotic value. Initial, asymptotic infiltration, and decay coefficients are considered to be correlated. In Section 3.4 we further develop the model by applying it directly to an artificial infiltration pond where

[0]This chapter is based on D. Pedretti et. al (2011), *Spatio-temporal assessment of soil infiltration capacity using physical-based models and geostatistical inference*, Stochastic Environmental Resources and Risk Assessment, vol. 25,n. 8, pp. 1065-1075, doi 10.1007/s00477-011-0486-4

local infiltration data is available at a few points and can be further obtained from a full mapping of a secondary variable. An external validation of the model from the evolution of the spatially averaged infiltration with time is also included.

3.2 Background

The quality of most of the information depends on the quality of geological characterizations, which is almost always poor in practice since direct investigations are costly, affected by serious errors and sometimes simply not feasible. Some models have been proposed in the past to correlate initial hydraulic conductivity profiles with the intensity of clogging.

Unfortunately, most of them are based on assumptions of homogeneities that over- or under-estimate the rate of clogging. For instance, assuming the soil pore sizes or the grain sizes to be uniformly (e.g. [Kozeny 1927, Carman 1938, Hazen 1882]) and non-uniformly (e.g. [Guin 1972]) distributed is essential to correctly estimate the development of clogging. Indeed, the use of uniform formulations implicitly include average macro-characteristic features of PSD and GSD, such as the total soil porosity ϕ or some characteristic grain size d_g, from which the seepage velocity is calculated. However, bioclogging does not grow uniformly but rather starts by developing local microcolonies ithin smaller pores, and grows to occupy the larger pores. On the other hand, physical mechanisms act differently.

According to the filter theory (see [Zamani 2009] for details) the clogging rates are inversely proportional with d_g but this relationship depends strongly on other factors such as the suspended solid size (d_s) and $I_c(t = 0)$. [Guin 1972] adopted a macroscopic Kozeny-like model where the porous medium is decomposed into a bundle of channels with different hydraulic radii R (ratio between porosity and the specific surface area of the soil, S_a). In the case of rapid particle deposition, the clogging rate of individual pores is proportional to the square of its specific surface area ($-\frac{\mathrm{d}S_a}{\mathrm{d}t} \propto S_a^2$). Thus, this implies that the area of larger pores (i.e. larger initial infiltration capacities) is decreased preferentially. In the case of slow particle deposition, $-\frac{\mathrm{d}S_a}{\mathrm{d}t} \propto S_a^{1/2}$. Experiments on bioclogging [Cunningham 1991, Vandevivere 1992] showed that the relative change of permeability (K_s/K_{s0}) depends, among other parameters, on the GSD.

It has been observed that all clogging mechanisms do not go on indefinitely, but rather end up providing an asymptotic infiltration value, which varies in space (e.g. [Baveye 1998]). The reasons that lead to an asymptotic clogging effect are controversial and depend on local conditions. For instance, bioclogging mechanisms roughly follow a Monod behavior [Monod 1949, Okubo 1979, Baveye 1998], which reproduce a microbial growth with a maximum asymptotic value of development. Physical clogging, on the other hand, varies in time since the governing forces change from volumetric to surface and vice-versa depending on flow velocity, available reactive surface areas, etc. (see [Zamani 2009] for a complete description of whole physical processes). While physical clogging could theoretically result in zero infil-

tration, bioclogging may help preventing additional physical clogging, so that the combination of both processes may still present a finite non-zero asymptotic value.

While assessment of initial and final permeabilities can certainly be done with the use of empirical or theoretical formulations, the question still remains about how to evaluate the actual evolution of infiltration with time. This is required for a proper risk assessment (this issue is addressed in Chapter 4).

Lumped solutions such as exponential decay formulations are often adopted at the field scale [Iwasaki 1937, Perez-Paricio 2000, Kim 2010, Hoffmann 2010]. The use of such models is appealing since only a few parameters are required for estimation purposes, but so far there is no widely accepted work on the relationship between the upscaled and the local infiltration models. While lumped (upscaled) models are routinely used as a way to assess when maintenance operations should take place at the full pond (usually drying the pond and cleaning), we contend that a detailed clogging model could be used to derive an ad-hoc remediation operation that targets only specific portions of the pond (similar to the concept of precise agriculture).

A major limitation of mapping infiltration variations locally in space and time is that primary information is limited and plagued by errors. A viable option to enhance the mapping of characteristic soil properties is to combine direct (primary) with secondary data (related to the primary ones). An example is to combine measurements from surface infiltrometers with satellite-images based secondary information, such as using the method shown in Chapter 2. Normally, secondary information is suitable for spatial assessment as it typically contains a denser dataset (e.g. [Gooverts 1997]), but the robustness of the method relies in the strength of the correlation existing between the two types of information.

3.3 Methodology

3.3.1 Problem Statement

Let us consider a given artificial SP recharge site in which direct measurements of the local infiltration rate, $I_c(\mathbf{x}, t)$, are known at sparse locations and at a few discrete times, i.e., $I_c(\mathbf{x}_i, t_j)$ $\{i = 1, ..., m\}$, $\{j = 1, ..., n\}$. This is often the case in most practical applications as the operation of an artificial recharge pond is typically done under flooding conditions and thereby any exhaustive characterization of the infiltration capacity is too costly and time-consuming. The support scale of the measurement is local and given, for example, by the size of an infiltrometer test.

We assume that an extensively sampled secondary variable is available at some support scale. An example would be data coming from the colour intensity of a satellite image, which can provide valuable information related to the soil hydraulic properties such as the moisture content of the soil, the vegetative canopy density on the ground [Chica-Olmo 2000, Granger 2000, Milewskia 2009] or the infiltration capacity at a specific time (see Chapter 2). Evidently, a limitation of satellite images is that image-based methods are restricted to non-flooded times, with at most two

measuring times, i.e., before and after flooding.

Under these conditions, we attempt to map the temporal evolution of the infiltration capacity at the site so that better management of the artificial recharge pond can be undertaken.

3.3.2 Clogging model

We assume that the temporal reduction of the infiltration capacity at every point $\mathbf{x}$ in the domain due to clogging processes follows a decaying exponential law. The model assumes that clogging takes place on the order of tens of days. Temperature can therefore be neglected as it fluctuates on two temporal scales, seasonally and daily, that do not affect clogging occurrence. Seasonality occurs at temporal scales much longer than clogging time, and night-day fluctuations are too short to affect clogging development permanently. The clogging model can be formulated in different ways. One possible formulation is

$$I_c(\mathbf{x},t) = R\ I_{c,0}(\mathbf{x}) \exp(-\lambda(\mathbf{x},t)t) = I_{c0}(\mathbf{x}) \exp(-\lambda(\mathbf{x},t)t), \tag{3.1}$$

where $I_c, 0(\mathbf{x})$ is the infiltration capacity at some initial operation stage, $\lambda(x,t)$ is the lumped clogging coefficient and R is a generic, instantaneous reduction function of infiltration caused by gas production and other mechanisms [Olsthoorn 1982, Bouwer 2002], which in general is close to 1. We can remove the impact of R by using $I_{c0}(\mathbf{x}) = R\ I_{c,0}(\mathbf{x})$ as the initial infiltration value. It is worth noting that in (3.1), λ changes over space and time, tending to zero as time increases.

Consequently, $I_c(\mathbf{x},t)$ values tend to an (spatially dependent) asymptotic value, $I_{c,f}$. Therefore, an alternative model to (3.1) that implicitly includes this asymptotic value $I_{c,f}$ can be formulated as

$$I_c(\mathbf{x},t) = I_{c,f}(\mathbf{x}) + (I_{c0}(\mathbf{x}) - I_{c,f}(\mathbf{x}))\ \exp(-\lambda(\mathbf{x})t), \tag{3.2}$$

where $\lambda(\mathbf{x})$ is now constant in time. This clogging model constitutes the basis of our approach to map the temporal evolution of the infiltration capacity. The fundamental advantage of this model is that parameterizes the temporal evolution of the infiltration capacity I_c by means of three constant-in-time variables/parameters: $I_{c,f}(\mathbf{x})$, $I_{c0}(\mathbf{x})$, and $\lambda(\mathbf{x})$, all three variable in space, and are therefore susceptible to simple geostatistical analysis.

Total infiltration capacity can be obtained by spatial averaging of either (3.1) or (3.2); i.e.

$$\overline{I}_c(t) = \frac{1}{V} \int_V I_c(\mathbf{x},t) \mathrm{d}\mathbf{x}. \tag{3.3}$$

In most cases $\overline{I}_c(t \to \infty) \equiv \overline{I}_{c,f}$ will be too small to be acceptable; that is, it will be below a pre-specified threshold value, I_t. Such small could be unacceptable in practical situations. Thus, the need for a model that can provide information about the spatial evolution of infiltration with time. It turns out that whenever $I_{c,f}(\mathbf{x})$

and $I_{c0}(\mathbf{x})$ are fully known, the temporal evolution of the local infiltration depend exclusively on $\lambda(\mathbf{x})$. We postulate that this non-time dependent clogging factor is directly correlated with some initial property of the soil, so that $\lambda(\mathbf{x}) = f(I_{c0}(\mathbf{x}))$. This is feasible since clogging develops at different rates according with the local distribution of PSD or GSD, both highly uncertain.

3.3.3 Mapping the soil infiltration capacity

If experimental measurements of I_c are limited to a few locations, secondary information can be incorporated to estimate the spatial distribution of the primary variable. For this purpose, several geostatistical techniques can be used, such as the collocated cokriging model. Yet, in this case, its direct application is cumbersome as one needs to estimate the evolution of the variogram matrix with time as clogging progresses. To overcome this problem, we propose a new approach.

As shown in Chapter 2 of this thesis, the exist a relationship such that $I_{c0} = \mathrm{f}_1(P_v)$ and $I_{c,f} = \mathrm{f}_2(P_v)$, where P_v are the color pixel values of a satellite image and f are two generic linear functions. By extension, and invoking a phenomenological approach, it is to be expected that there could exist a relationship between soil parameters and P_v, and thus also between λ and P_v.

This relationship should be constructed via theoretical or empirical methods. The theory suggests that some clogging mechanisms, such as physical clogging, can be modeled using a filter approach (e.g. [Zamani 2009]); thus, the clogging rate should be negatively correlated with some grain size representative diameter, d_g. Other mechanisms however do not necessarily rely on this assumption: biological clogging can be modeled using Monod-based growing models (e.g. [Clement 1996]), for which the rate of bioclogging is linearly proportional to the soil density ρ which is positively correlated with d_g (see Chapter 4 for derails). Therefore, in a real site, it is important to assess the relative importance of the two mechanisms, since this will control the relationship existing between λ and the soil properties. In the field though, the use of (3.2) gives lumped clogging factors, in which the single effects of each mechanism is somehow hidden. As such, any relationship existing between P_v and experimentally-based λ does not give any indication of the relationship existing between P_v and all the parameters characterizing the soil, including d_g, but also density, fraction of organic matter, etc. Thus, it is not *a priori* clear whether λ and P_v should be positively or negatively correlated. In the application example later we will use both possibilities.

Our approach starts from having a perfect knowledge of some secondary variable (obtained for example from satellite images), while there is little to no information on the three primary variables controlling infiltration: I_{c0}, $I_{c,f}$, and λ. Thus, it is possible to use some geostatistical approach involving either cokriging (to get a smooth estimate in a mean sense) or cosimulation (to be included in a Monte Carlo approach). Since data from the primary variables are always expected to be less than exhaustive, we chose a collocated cokriging approach under the Markov model I. Such a model limits the secondary variable to the data available at the estimation

location, and further estimates the cross-variograms by employing an underlying regression model. The advantage is that matrix instabilities caused by densely sampled secondary data (such as high-resolution satellite rasters of pixel values) are avoided and reduce the burden of modelling all variograms and cross-variograms in a cokriging system [Almeida 1996, Journel 1999]. In this case, the cokriging model only requires the knowledge of the variogram of the primary variable, the correlation coefficient, and the variance of the secondary variable. Thus, the modelling effort is almost the same as for kriging one variable. For simplicity, the cokriging of each variable is conducted independently.

Unfortunately, in most practical situations, the variogram of these primary variables cannot be directly estimated as too little information is available at a given SP site. To circumvent this problem, we suggest to completely rely on the variogram of the secondary variable (satellite image) that is always well characterized. For each primary variable, $\{I_{c0}, I_{c,f}, \lambda\}$, the auto-variogram can be estimated using regression models. In Chapter 2, it was shown that a linear regression satisfactorily correlates the natural logarithm of infiltration capacities (both Y_0=ln(I_{c0}) and Y_f=ln($I_{c,f}$)) with the color pixel values of an image (P_v). The model has the form

$$Y_0 = a_0(P_v) + b_0 + \varepsilon_{Y,0}, \tag{3.4}$$

$$Y_f = a_f(P_v) + b_f + \varepsilon_{Y,f}, \tag{3.5}$$

where a and b are the regression coefficients, and ε_Y represents regression model errors. To complete the picture we must specify a model for λ. Parsimony leads us to postulate a similar linear model in terms of $Y_\lambda = \ln(\lambda)$,

$$Y_\lambda = a_\lambda(P_v) + b_\lambda + \varepsilon_{Y,\lambda}. \tag{3.6}$$

From these simple models we can write the variogram functions as

$$\gamma_{Y_i}(\mathbf{h}) = a_i^2 \gamma_{P_v}(\mathbf{h}) + \gamma_{\varepsilon_{Y,i}}(\mathbf{h}), \qquad i = 0, f, \lambda \tag{3.7}$$

where $\mathbf{h}$ is the lag distance between data values. We can further assume that the regression model errors are uncorrelated both with errors at different locations or with the secondary variable (pure nugget). Once the spatial distributions of the three variables have been obtained, it is possible to use (3.2) to obtain the spatio-temporal distribution of the local infiltration rate, and consequently the temporally variable global infiltration. We now illustrate the method and highlight the limitations in a real site.

3.4 Application Example

3.4.1 Observations and modelling at the pond scale (global scale)

We apply this approach on the pilot SP located in Sant Vicenç dels Horts. The reader is referred to Chapter 2 for a detailed description of the site. We will consider here that fig. 2.7 represents the initial state of our system ($I_{c0}(\mathbf{x})$). A flooding

test was performed on this site between March and June 2009 in order to test the performance of the site for MAR operations. The following variables were recorded on a continuous basis: headwater at the pond, $h(t)$; discharge rate towards the aquifer, $Q(t)$ (recorded from water mass balance within the pond, disregarding evaporation); infiltration area, A (usually a direct function of h(t) to include the basin slopes); and distance from the surface to the water table, $L(t)$. Infiltration at the full pond scale is then obtained as

$$\overline{I}_c(t) \approx \left(\frac{Q(t)\ L(t)}{A\ h(t)} \right) \tag{3.8}$$

The recorded infiltration values can be matched by a simple exponential model

$$\overline{I}_c(t) = \overline{I}_{c,f} + (\overline{I}_{c,0} - \overline{I}_{c,f}) \exp\left(-\lambda^e t\right) \tag{3.9}$$

Fig. 3.1 shows that a good match between data and the best-fit approximation of the large-scale infiltration model (3.9) can be obtained. λ^e is an equivalent clogging parameter for the entire pond. Note that the asymptotic value is used to evaluate the minimum infiltration value of the pond, which is reached when clogging no longer develops. The solid line represents observations using (3.8), while the dotted line is the best fit model solution using (3.9) giving the following parameters: $\overline{I}_{c,f} = 0.44$, and $\lambda^e = 0.17\ \mathrm{d}^{-1}$. Note that infiltration data are dimensionless being normalized by the maximum initial value of the infiltration rates.

Note also that according to the calibration process using (3.9), we have a global model of the reduction of infiltration with time, that could now be used for management or risk evaluations. For example, it is found that reduction of infiltration to 50% of the initial value is reached after just roughly 7 days, but it will for example take 14 days to reduce to 40% of the original value. It is possible to obtain a compromise between allowing the system to work for larger periods without maintenance by just allowing less water to infiltrate.

In Fig. 3.1 the infiltration curve displays some daily oscillations. This is due to unaccounted causes, such as the impact of water temperature or atmospheric pressure with time. A detailed analysis of these causes could lead to a smoothing of the oscillations, leading to a better fit of the simple exponential model. However, this is not deemed essential here and the impact of these additional processes is not pursued in our work.

3.4.2 Observations at the local scale

Double-ring infiltrometers [Smith 1972] were used to measure the local infiltration capacities at sparse locations in the SP, before and after the flooding test was performed. A double-ring infiltrometer reproduces the local soil transition from unsaturated to saturated conditions that would take place under flooding conditions, and evaluates the infiltration capacity. In short, two metal rings are buried a few centimeters deep into the soil and filled with water (at constant or variable height).

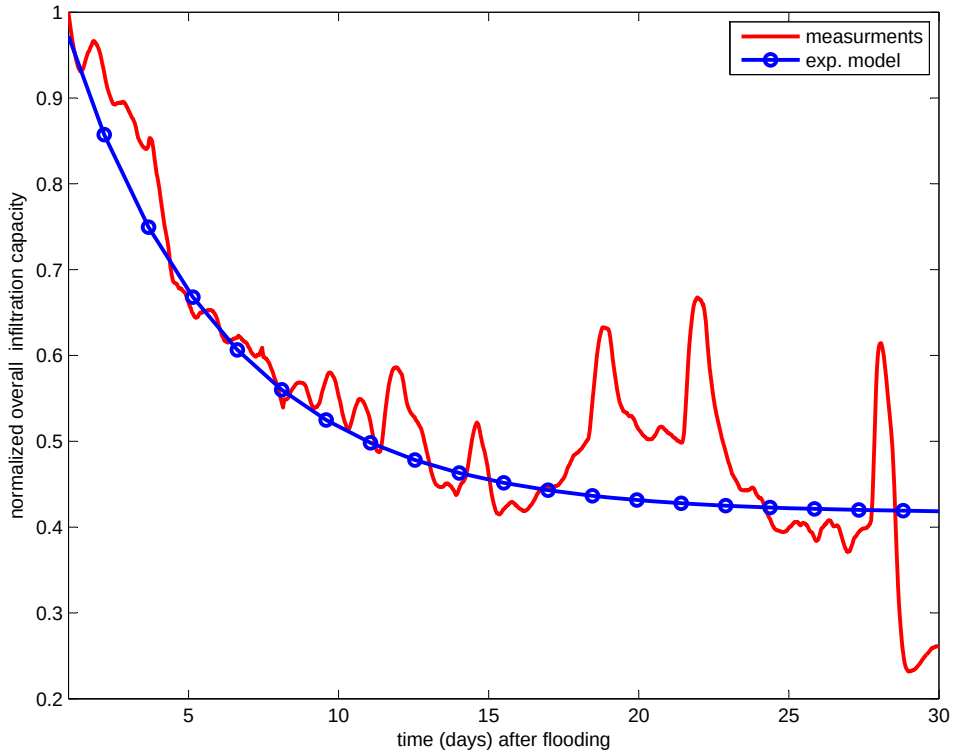

Figure 3.1: Experimental observations (solid line) of normalized $\overline{I_c}$ with time from the flooding test. The dotted curve is the best-fit approximation with $\overline{I}(t = \infty) = 0.44$ and λ^e=0.17 d^{-1}.

After a standard time of two hours, the infiltration rate approaches a steady state value, identified as $I_c(\mathbf{x}, t)$ at that location $\mathbf{x}$ and time t.

During the flooding experiments it is possible to obtain infiltration values at the local scale by using other devices. For consistency we did not use other methods and rely on two double-ring campaigns, This implies that the only available measurements are those of I_{c0} and $I_{c,f}$. Only a few points are available for each one of these two variables. The former was obtained at 6 spatial points and the latter at those same 6 points plus three additional ones (a total of 9).

It is worthwhile noting that double-ring tests provide a direct estimate of infiltration capacity at the scale of the device. In this case, they were representative of a small support scale (≈ 0.12 m^2). This may have a significant influence in the geostatistical analysis, since it is important to use a method that accounts if necessary for the difference in support between the different variables involved. In Figure 1 the "S" letters denote the locations where the double-ring infiltration tests were performed. The "C" letters refer to excavated pits where geology was directly observed and described.

Results from the analysis of the infiltration tests (using the modified Kostyakov method) were reported in Table 2.1 in Chapter 2, and are graphically plotted in Fig. 3.2. It is clear that all points analyzed show a reduction in infiltration between the initial and the final stage. It can be observed that this reduction is larger (both in absolute and in relative terms) for the points displaying higher I_{c0} values.

3.4.3 Local scale modelling

The results of the correlation between the red band of the image presented in Fig. 3.2 and the logarithm of the infiltration values at $t = t_0$ and at $t = t_f$ are found to be linear according to equations (3.4) and (3.5). A good linear correlation is obtained for both states of the artificial pond, with a Pearson's coefficient of $r^2 = 0.87$ for the February campaign ($t = t_0$) and $r^2 = 0.89$ for the June campaign ($t = t_f$).

3.4.4 Experimental variograms and Cross-variograms

Field experiments are usually difficult to conduct and costly. Thus, one can typically always expect a very low number of data points, which in most cases (certainly in ours), prevents the direct estimation of the variogram of the infiltration capacity. Instead, a well behaving sampled variogram may be available for some secondary variable, P_v in our case. Fig. 3.3 shows the resulting two directional variograms obtained for P_v in the principal directions. The direction of maximum correlation is oriented along the y axis. The model variograms show three different structures, whose combined formulation in terms of principal directions is

$$\gamma_{P_v}(h_x) = 297 \left[0.32 \cdot \text{Gauss}(\frac{h_x}{7}) + 0.44 \cdot \text{Sph}(\frac{h_x}{92}) + 0.24 \cdot \text{Exp}(\frac{h_x}{400}) \right] \qquad (3.10)$$

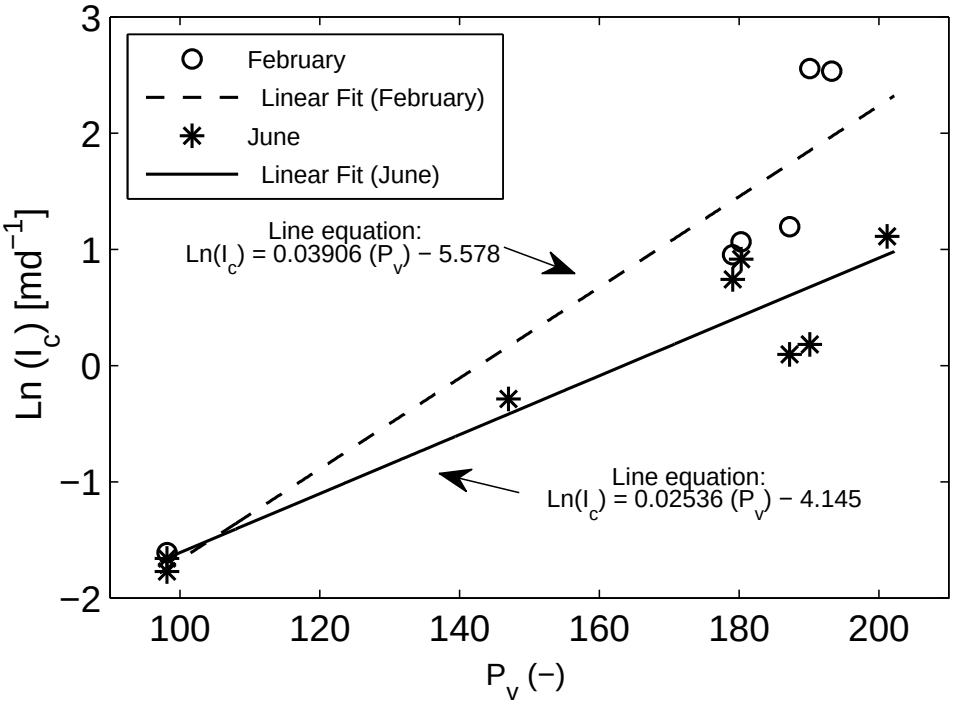

Figure 3.2: Correlation between the pixel values (P_v) from the red band of the satellite image (Figure 2.1) and measured infiltration capacities at selected locations in the SP at two different moments: pre-flooding (open circles), and post-flooding (stars). The lines show the linear fitting to the experimental data. The Pearson's coefficient for the dotted line (February dataset) is r^2=0.87; for the straight line (June dataset), r^2=0.89

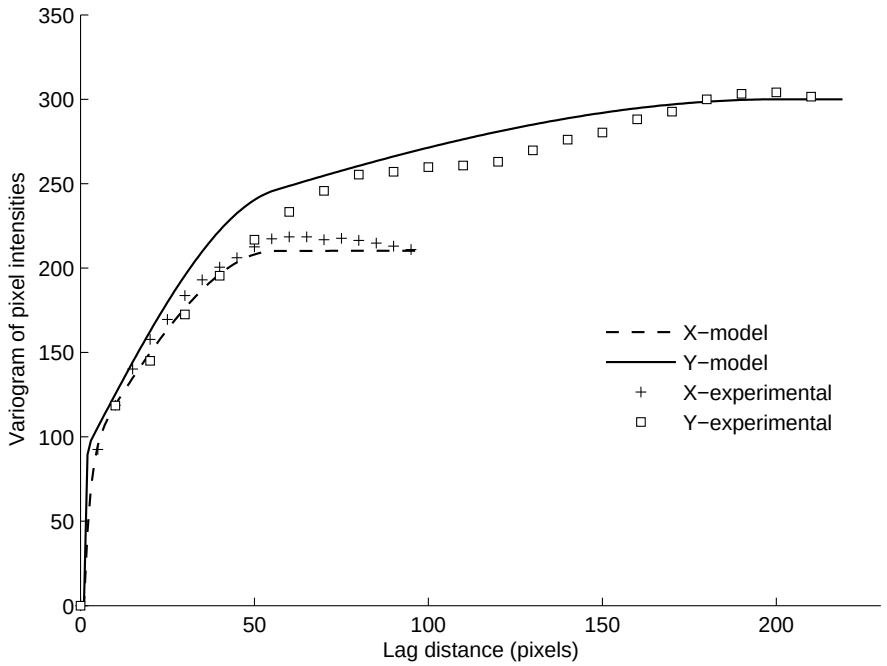

Figure 3.3: Directional variograms of the pixel values in the reference image (exhaustive dataset)

$$\gamma_{P_v}(h_y) = 297\left[0.32 \cdot \text{Gauss}(\frac{h_y}{7}) + 0.44 \cdot \text{Sph}(\frac{h_y}{68})\right] \tag{3.11}$$

where Gauss(.), Exp(.) and Sph(.) are the standard unitary variogram models [Deutsch 1998].

3.4.5 Mapping the clogging factor

Finally, the clogging factor λ should also be estimated. In our case, no direct field estimates about this parameter were measured. To overcome this problem, we assumed a perfect correlation between this parameter and the initial infiltration capacity Y_0,

$$Y_\lambda = a(Y_0) + b. \tag{3.12}$$

The coefficients a and b were estimated from the temporal evolution of the integrated infiltration capacity observed during the flooding test. The theory is unclear as to whether these two parameters should be positively or negatively correlated (e.g. [Guin 1972, Zamani 2009]). Thus, we explore in our site the potential range of parameters in the power law formula (3.12) that can lead to a fit of the observed global infiltration behavior. In particular, we analyze the sign of the coefficient a.

We will see how this may have a strong effect in devising potential remediation strategies.

Starting from given values a and b, the map of the λ values is obtained with (3.6), and then the local infiltration in space-time is obtained from (3.2). Finally, the integrated infiltration is obtained with (3.3). The method is repeated by changing the values of a and b until the curve matches that of Figure 3.1.

Several potential combinations of the two parameters might lead to reasonable fits. In Fig.3.4 we fix two different b values and explore the sensitivity of the global infiltration curves to a. It is found that we are able to obtain two sets of parameters (a, b), $(-0.4, -0.3)$ and $(0.3, -3.0)$ that lead to a similar fit. It can be observed that the curves are quite sensitive to the two parameters, providing a way to calibrate them. Calibration cannot be fully completed in our site since no data for intermediate times could be recorded. As a consequence we believe that there is a need to perform short flooding tests in real sites in order to obtain real values of λ that could be used to design an optimal management operation. This can only be obtained by performing test during the flooding period. The problem is that infiltration measurements are quite sensitive to the method used, and so it is difficult to be able to combine data coming from double-ring or seepage meters, for example.

While one can find different sets of (a, b) values, the key is the sign of the a parameter. Depending on this sign we will have enhanced clogging in regions of initially high infiltration capacities, or alternatively, clogging enhanced in points displaying initially small infiltration capacities. This can be seen in Fig. 3.5, which includes the maps corresponding to the two sets of (a, b) already presented before. By construction, the maps are visually highly correlated with those of initial or final infiltration, included also in the same figure.

3.4.6 Temporal evolution of the statistical distribution of local infiltration values

One of the main factors controlled by a is the actual shape of the probability density functions (PDFs) of λ. Thus, while the two sets of parameters lead to a very similar fit of the global infiltration curve, the actual values of λ are very different (see Fig.3.6).

A direct implication is that it is not possible to derive an equivalent λ value from the local ones capable of reproducing the full behavior. This can be observed in Fig. 3.7, where the evolution of infiltration with time using the harmonic, geometric and arithmetic means of the point values presented in Fig. 3.6 are computed. It is clear that upscaling from local values would be an error in general, and that some conditioning on real values is needed to produce a proper reproduction of the global behavior.

The evolution of the distributions of the local values of infiltration capacity with time is dependent on the map of λ values. This can be observed in Fig.3.8, where we plot the cumulative distribution functions (CDFs) of $I_c(\mathbf{x}, t)$. Positive a values imply a slower reduction in the infiltration values with time than a negative a value.

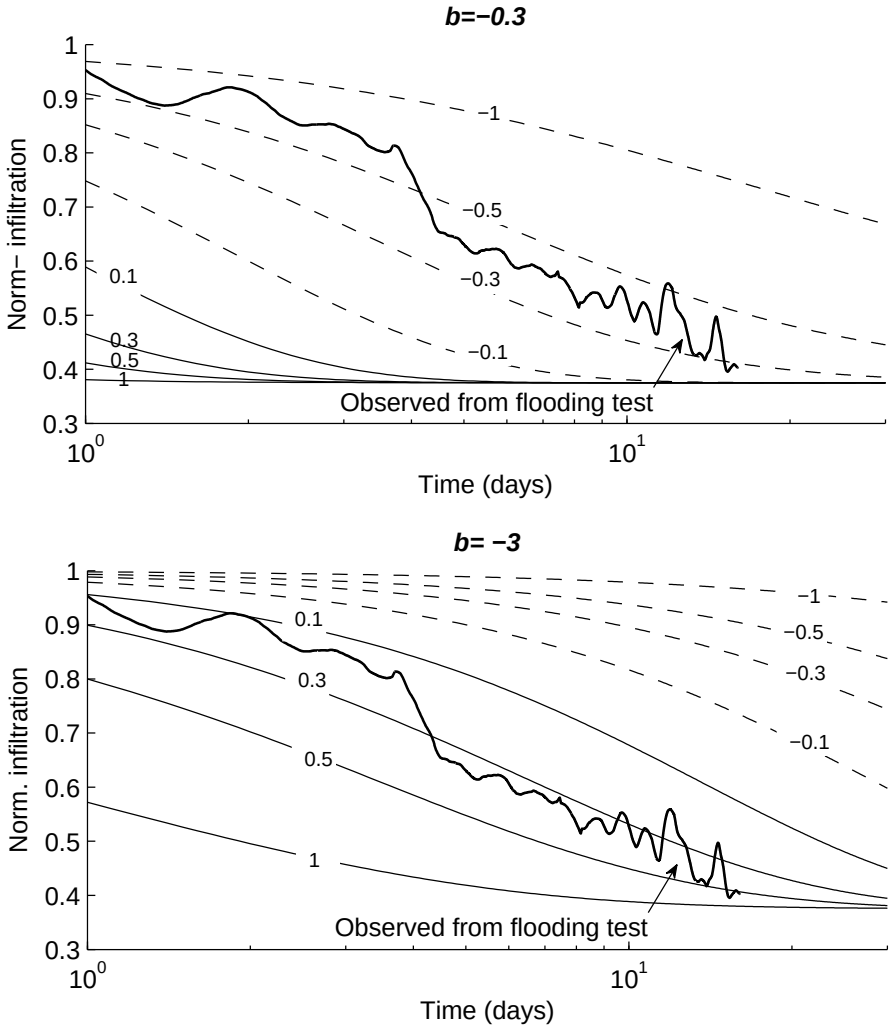

Figure 3.4: Sensitivity of the correlation model between $\ln(I_c)$ and $\ln(I_0)$ on the slope coefficient a and the y-intercept. The results are plotted for $b = -0.3$ (up) and $b = -3$ (down). The normalized observed values, calculated using the water balance equation, are also plotted for comparison with the curves.

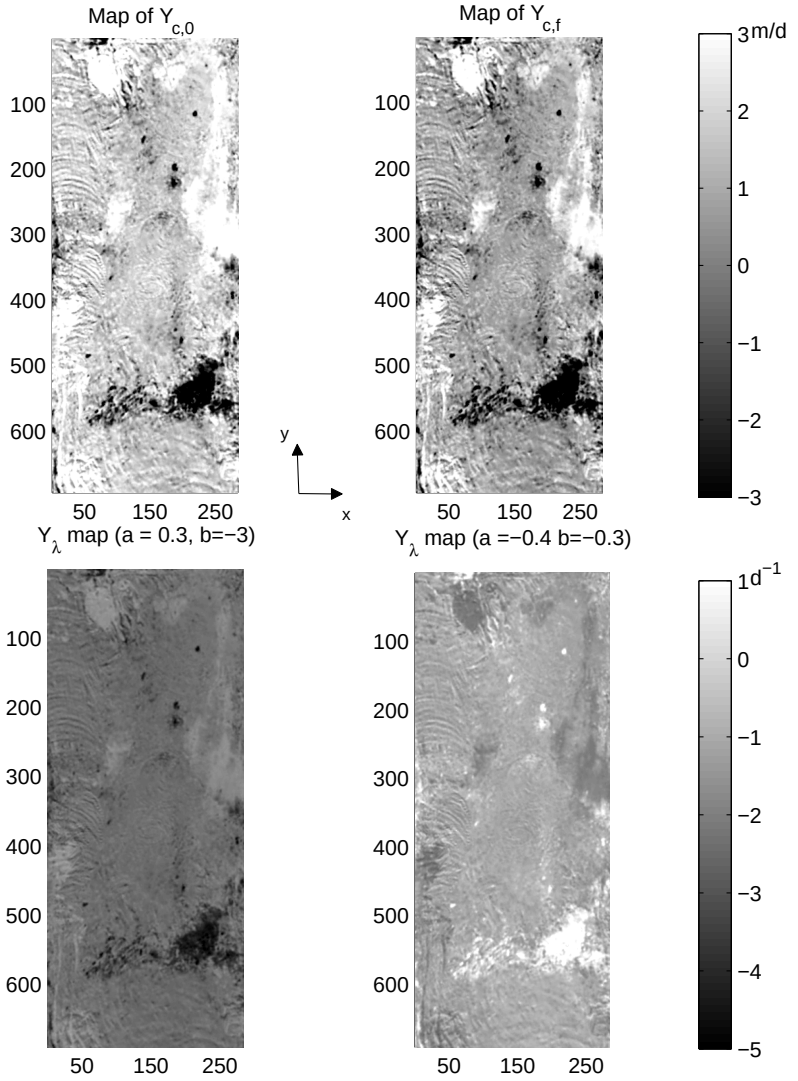

Figure 3.5: Cokriged maps of the logarithm of the infiltration capacities of the pilot SP at t_0 and t_f (up), and corresponding λ maps using positive a on the left and negative a on the right.

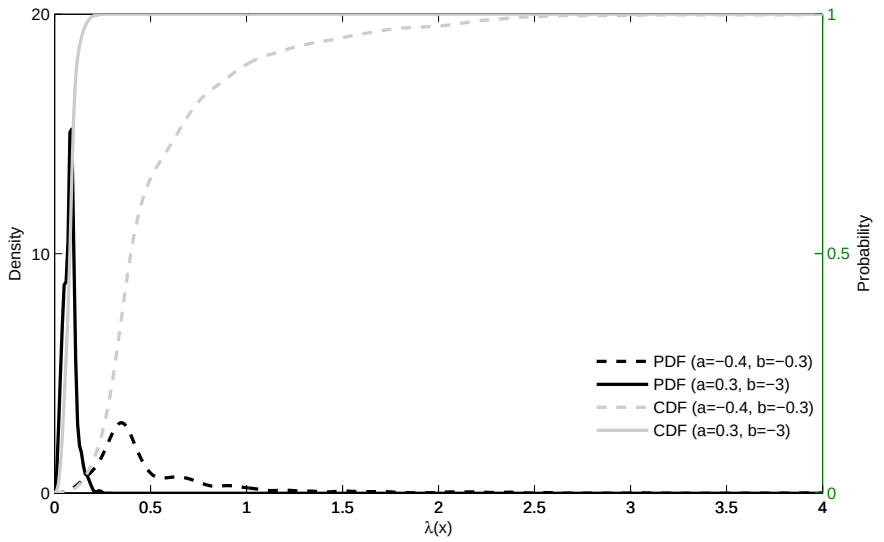

Figure 3.6: Probability density function (PDF, black lines) and cumulative frequency (CDF, gray lines) of $\lambda(\mathbf{x})$ calculated using different combination of a and b values, and mapped in fig.3.5(down).

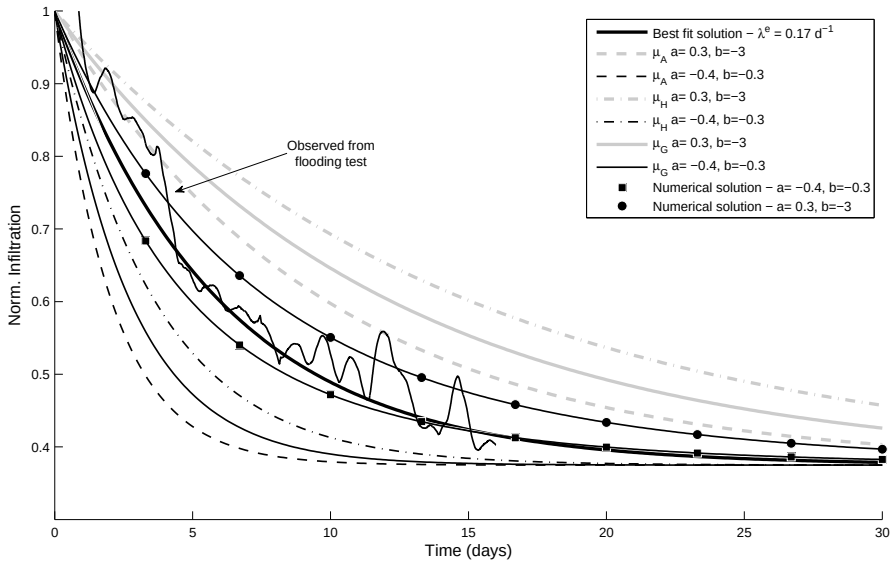

Figure 3.7: Evolution of the normalized mean infiltration capacity of the pond evaluated using different methods, versus the observed values and the equivalent solution using the effective $\lambda^e = 0.17$ d^{-1}. The numerical solutions (marked with line and symbols style) are obtained by spatial average of local values calculated using the reference local infiltration model. Analytical solutions using the geometrical (straight line), harmonic (dot-point line) and arithmetic (dot-dot) averaging of $\lambda(x)$ as effective decay rate are also reported for comparison purposes.

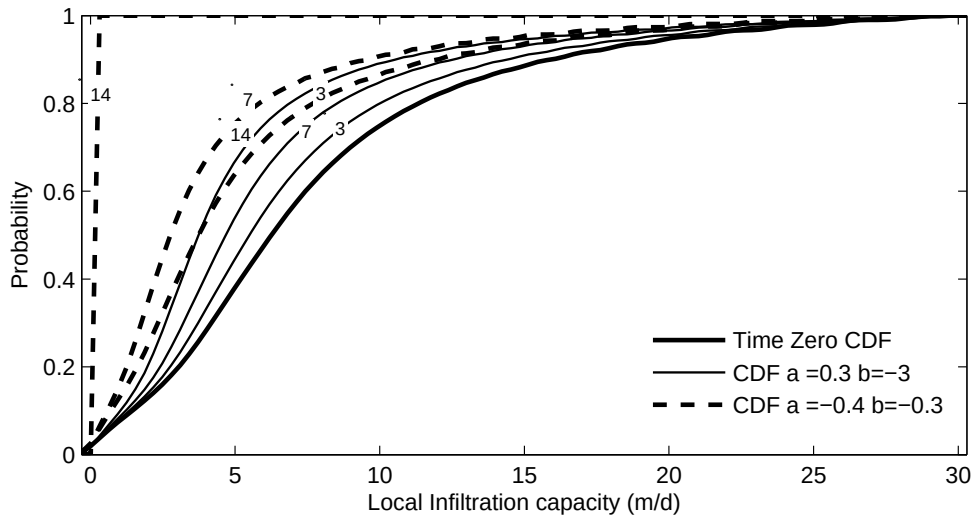

Figure 3.8: Evolution of the probability of local infiltration capacity values in the pond at different times (the tags express the days elapsed after the flooding test began) using two possible combination of relationships between λ and the initial infiltration capacity.

This is caused by the former leading to very small λ values in Fig. 3.6. These small values cause the infiltration to be reducing slowly with time. The picture completely changes for the alternative set of parameters $(-0.4, -0.3)$ where the local λ values are high, and thus infiltration reduces very fast with time.

These results will be significant when devising a remediation method. Notice how in Fig.3.8 it is observed that for the set $(0.3, -3.0)$ all values of infiltration decrease more or less equally in time. Thus, the best alternative for remediation would be to empty the full pond and restore the initial capacity if possible. On the other hand, for the $(-0.4, -0.3)$ we observe how the reduction of infiltration is mostly produced by the fast reduction in the more permeable areas. Thus, we could devise a method that targets the initially high infiltration areas. If clogging in these points is prevented, the global infiltration will remain high for a long period of time.

Chapter 4

A stochastic framework for the optimal maintenance of artificial recharge ponds under uncertainty

4.1 Introduction

Due to both soil heterogeneity, pore clogging, and lack of adequate soil and processes characterization, $I_c(\mathbf{x}, t)$ is highly uncertain. Inherent uncertainty in estimates of infiltration capacity in an artificial ponds (SP) and predictions of its temporal evolution introduces significant uncertainty into decisions about SP management, with important economical and environmental implications.

The goal of this chapter is to provide a framework to evaluate the engineering risk of making decisions in regards to optimal maintenance of SP, when spatio-temporal distributions of local hydraulic properties of the topsoils are uncertain. In this Chapter, $I_c(\mathbf{x}, t)$ is treated as a random field that renders the equations governing the spatio-temporal evolution of the artificial infiltration process stochastic. Clogging models are here derived from theoretical studies, most of them applied to other disciplines than hydrogeology, such as filter theories, biomass generation studies or biogas production.

In this chapter, the problem of management SP under uncertainty is formulated within a probabilistic framework context ([Drazen 1998, Orr 2005, Wagner 1987, Zenios 1998]). This in turn can form the core of a probabilistic risk analysis, which while a relatively new discipline in hydrogeology ([Batchelor 1998, Tartakovsky 2007, Winter 2008, de Barros 2008, de Barros 2009, Bolster 2009]), forms standard practice in other engineering disciplines(e.g. [NRC 1997]).

This chapter is structured as follows. Section 4.2 provides a justification for adopting this probabilistic approach. The main factors and their modeling formulations are described in section 4.3. Section 4.4 contains a general sensitivity analysis of the selected models. In section 4.5, this methodology is applied to one real and a few synthetic examples.

[0]This chapter is based on D. Pedretti et. al (2012), *Probabilistic Analysis of Maintenance and Operation of Artificial Recharge Ponds*, Advances in Water Resources, vol. 36, pp. 23-35, doi 10.1016/j.advwatres.2011.07.008

4.2 Operation of SP under Uncertainty

Depending on the overall intensity of the clogging mechanisms, SP can suffer from "aging" [NRC 1994], which is an appreciable reduction in the infiltration capacity in the first few days after flooding (e.g. see Chapter 3). To meet designed infiltration rates, SP must be periodically maintained with either preventive or corrective measures [ASCE 2001, CGWB 2007]. The speed with which the infiltration capacity approaches or drops below some critical threshold value (I_c^*)is the primary variable indicating when and what type of corrective measures should be taken.

The complexity of modeling soil clogging and the corresponding reduction in infiltration capacity, coupled with ubiquitously insufficient site characterization and soil heterogeneity, renders predictions of an infiltration pond's performance fundamentally uncertain. This challenge is partially alleviated by the fact that this performance is determined by the integrated infiltration capacity $\overline{I}_c(t)$ rather than the actual distribution of local infiltration values $I_c(\mathbf{x}, t)$.

Empirical evidence from several operating SP suggests an exponential decay behavior in the overall infiltration capacity (e.g. [Perez-Paricio 2000, Kim 2010, Hoffmann 2010])

$$\overline{I}_c(t) = \overline{I}_{c0} \mathrm{e}^{-\lambda_{\mathrm{eff}} t} \tag{4.1}$$

where the initial capacity $\overline{I}_{c0}$ and the effective decay (clogging) rate λ_{eff} are highly uncertain fitting parameters that are difficult to predict prior to the SP's operation. Actually, in general, the infiltration rate tends asymptotically to a non-zero value, but is most cases this asymptotic value is very small compared to the initial one and can thus be approximate by a zero value (e.g. the pilot SP in Spain discussed in Chapter 3).

A variety of maintenance activities with different scheduling plans can be applied to SP to control the reduction of $\overline{I}_c(t)$ with time. These can be subdivided into preventive or corrective measures ([Bedford 2001]). Preventive or maintenance activities are performed during the operation period in order to extend the system's life. Examples include pre-filtering input water to eliminate particles, scraping the soil surface before infiltration, using disinfectants to control algal growth, designing a large settling pond to remove organic matter, and controlling entry water temperature to avoid gas bubbling. Corrective measures must be taken if and when $\overline{I}_c(t)$ reaches its minimally acceptable level I_c. For these the operation of the SP must be temporarily stopped. Examples include scraping the bottom surface when the basins are dried out after specific recharge cycles, supplying additional disinfectants or chemicals to the water, or using underwater robots to scrape the soil surface during infiltration.

Regardless of the maintenance strategy, a SP's operating life is highly uncertain. Maintenance decisions have to be made under uncertainty, calling for a probabilistic approach. However, current practice is to schedule corrective measures based on experience and monitoring [CGWB 2007]. An optimal scheduling and selection of

maintenance measures affects the operational costs of SP. Forecasting these costs is subject to uncertainty and depends on many factors, such as the optimal performance of the chemical products, the input water quality and the performance of scraping machines.

4.3 Processes Contributing to SP Failure

We define a "system failure" at time t as the event "water infiltration capacity $\overline{I}_c(t)$ falls below a design value I_c". Among a large number of events that can lead to system failure are interruptions in water supply to the pond, deposition of extraneous impermeable materials at the pond's bottom, bad design and/or improper use of the pond, and its complete breakdown due to embankment slides, earthquakes, and acts of vandalism [Dillon 2002]. These and other similar events should be included into a complete probabilistic risk assessment of SP, but lie outside the scope of the present analysis.

Instead, we focus on system failures due to reduction in the soil's infiltration capacity caused by clogging. A mathematical model capturing the relationship between infiltration capacity and the soil hydraulic parameters is presented in section 4.3.1. Various clogging mechanisms are discussed, modeled, and combined together for an effective model of clogging in section 4.3.2.

4.3.1 Infiltration & hydraulic parameters

In a well designed SP, infiltration is controlled by the top soil layer. (A counterexample would be the occurrence of a very low-permeable layer with horizontal continuity, at some depth, limiting infiltration. The presence of such a layer is not considered here). When the top soil layer controlling the infiltration capacity is (nearly) fully saturated and the infiltration can be described by the one-dimensional Darcy law, we can write

$$I_c(\mathbf{x},t) = -K(\mathbf{x},t)\frac{\partial h(\mathbf{x},\mathbf{z},\mathbf{t})}{\partial z} \tag{4.2}$$

where $\mathbf{x} = (x,y)^T$ is the vector of horizontal coordinates, z is the vertical coordinate, h is the hydraulic head, and K is the saturated vertical hydraulic conductivity. The latter is defined as $K = \kappa\rho_w g/\mu_w$, where κ is the soil intrinsic permeability in the vertical direction, ρ_w and μ_w are the density and dynamic viscosity of water, respectively; and g is the gravity acceleration constant.

To account for changes in hydraulic conductivity due to pore clogging, we adopt the Kozeny-Carman law [Kozeny 1927, Carman 1938] according to which permeability κ varies with the soil porosity ϕ as

$$\kappa = \frac{d_m}{180}\frac{\phi^3}{(1-\phi)^2} \tag{4.3}$$

where d_m is the percentile of the cumulative distribution of the soil grain sizes.

The Kozeny-Carman law is applicable within a range of ϕ that includes soil grains coarser than fine sands. Such soils are typical for artificial recharge facilities, which typically are built in (highly permeable) coarse sandy sedimentary environments. However it is important to note that the Kozeny-Carman law is known to suffer from severe limitations that the user should be aware of prior to application (e.g. [Chapuis 2003]). When such limitations occur, alternative models can be integrated into this framework.

Assuming that the water properties and hydraulic gradient do not change over time, it follows from (4.2) that any reduction in infiltration capacity is linearly related to the reduction in K, i.e.,

$$\frac{I_c(\mathbf{x},t)}{I_{c0}(\mathbf{x},t)} = \frac{K(\mathbf{x},t)}{K_0(\mathbf{x})} \tag{4.4}$$

where $I_{c0}(\mathbf{x},t) = I(\mathbf{x},t=0)$ and $K_0(\mathbf{x}) = K(\mathbf{x},t=0)$ are the initial values of the infiltration capacity and the hydraulic conductivity, respectively. Combining (4.3) and (4.4), we obtain an equation relating the reduction in infiltration capacity to the reduction in porosity,

$$\frac{I_c(\mathbf{x},t)}{I_{c0}(\mathbf{x},t)} = \frac{\phi^3(\mathbf{x},t)}{\phi_0^3(\mathbf{x})}\frac{[1-\phi_0(\mathbf{x})]^2}{[1-\phi(\mathbf{x})]^2} \approx \frac{\phi^3(\mathbf{x},t)}{\phi_0^3(\mathbf{x})} \tag{4.5}$$

where $\phi_0(\mathbf{x}) = \phi(\mathbf{x},t=0)$ is the initial porosity before clogging started to occur.

The approximation in (4.5) is valid as long as temporal changes in ϕ remain small.

4.3.2 Mathematical models of clogging

A number of physical [Zamani 2009], biological [Baveye 1998, Clement 1996] and chemical [Greskowiak 2005] processes can contribute to clogging. Their complex interplay complicates the development of fully comprehensive mathematical models of clogging even though individual mechanisms are relatively well understood [Chang 2009, Civan 2005a, Civan 2005b, Tufenkji 2004]. Representative models of the three clogging mechanisms are discussed below.

4.3.2.1 Physical clogging

Physical clogging typically refers to filtration processes that reduce porosity ϕ through sedimentation and dragging of suspended particles [Iwasaki 1937, Zamani 2009]. Following [Iwasaki 1937], we employ a first-order mass transfer model,

$$\frac{\partial C(\mathbf{x},z)}{\partial z} = -\lambda_z(\mathbf{x})C(\mathbf{x},z), \tag{4.6}$$

to describe the vertical profile of the volumetric concentration of particles, C, that are removed from the suspension in the SP by trapping within the soil. The filtration coefficient λ_z combines volumetric and surface forces [Perez-Paricio 2000]. An exhaustive review of various forms of the filtration coefficient can be found in [Zamani 2009].

In a typical model, e.g., [Boyd 1974], λ_z is a combination of the following mechanisms:

- filtration induced by inertial forces, $\lambda_{ine} \propto d_s^{1.5}$;
- filtration due to interception mechanisms, $\lambda_{int} = (d_s/d_g)^n/d_g$;
- filtration caused by molecular (van der Waals) forces, $\lambda_{mol} \propto d_s^2$;
- filtration due to diffusion, $\lambda_{dif} = [\phi_0 K_b T/(\mu_w d_s d_g |\mathbf{q}|)]^{0.66}/d_g$;
- filtration due to sedimentation, $\lambda_{sed} = g\phi_0(\rho_k - \rho_w)d_s^2/(18\mu_w d_g|\mathbf{q}|)$.

d_s is the mean diameter of suspended particles in the water during the flooding stage, d_g is the characteristic grain size of the soil, n is the geometrical parameter for clogging interception mechanisms, K_b is the Boltzmann constant, T is the water temperature, $\mathbf{q}$ is the flow velocity of water, and ρ_k is the bulk density of the soil.

Empirical relationships can be used to relate the characteristic grain size d_g to the soil permeability κ. In the present analysis, we choose the Hazen formula [Hazen 1882],

$$\kappa = A d_{10}^2, \qquad A \approx 100, \tag{4.7}$$

in which $d_g = d_{10}$, the grain size corresponding to the tenth percentile in the cumulative distribution of grain sizes. The Hazen formula (4.7) is most appropriate for clean sands, which are typical in SP. It is worth emphasizing that the Hazen formula is used here for illustrative purposes, and can be replaced by other textural relations. Finally, we assume that d_{10} remains constant during an SP's operation, i.e., that the clogging material does not change the grain size distribution over time. This is deemed reasonable as clogging materials are typically an order of magnitude smaller (in the case of suspended particle) or have lower density (in the case of biomass) than the original material.

Let v_a denote an average particle attachment velocity to the soil matrix. Typical values of v_a can be obtained from the literature or from laboratory experiments. The experiments reported in [Perez-Paricio 2000] yield $v_a \approx 10^{-5}$ m/day, which is in agreement with reference values suggested in [Tien 1979]. Setting $t = z/v_a$ in (4.6) and integrating in time yields

$$C(\mathbf{x}, t) = C_0(\mathbf{x})\mathrm{e}^{-\lambda_p^*(\mathbf{x})t}, \qquad \lambda_p^* = \lambda_z v_a \tag{4.8}$$

where $C_0(\mathbf{x})$ is the initial concentration of particles in the soil column. Changes in the particle concentration, δC, cause changes in the porosity, $\delta\phi$. We postulate

a linear relation between the two, $\delta\phi/\phi = m_p \delta C/C$, where m_p is the coefficient of proportionality. It follows from (4.8) that

$$\phi(\mathbf{x},t) = \phi_0(\mathbf{x})\mathrm{e}^{-\lambda_p(\mathbf{x})t}, \qquad \lambda_p = m_p \lambda_p^* \tag{4.9}$$

where λ_p can be seen as the characteristic physical clogging factor. Substituting (4.9) into (4.5) yields a model describing the exponential reduction in the infiltration capacity due to physical clogging,

$$I_c(\mathbf{x},t) = I_{c0}(\mathbf{x})\mathrm{e}^{-3\lambda_p(\mathbf{x})t}. \tag{4.10}$$

4.3.2.2 Biological clogging

Biological activity, such as biomass growth and biogas generation, obstructs the pores and reduces both porosity and pore connectivity [Cunningham 1991, Baveye 1998]. Biological clogging is typically described with one of the three approaches: macroscopic models, micro-colony-based models, and biofilm-based models. Macroscopic transport equations resulting from all three are identical if biofilms and microcolonies are fully penetrating [Baveye 1989]. Furthermore, the three approaches yield acceptable predictions for coarse-textured materials (which are typical in SP operations), while poor predictions arise in fine-textured materials [Vandevivere 1995].

Bio-clogging manifests itself through a combination of factors: formation of a thin impermeable layer at the soil surface, biofilm formation on the soil grains, and precipitation of biomass that occludes the pores. We focus on the last two phenomena that act to reduce porosity. Specifically, we adopt a macroscopic approach and assume that all biomass growth leads to a direct reduction of porosity, $\phi(\mathbf{x},t) = \phi_0(\mathbf{x}) - \phi_b(\mathbf{x},t)$, so that

$$\frac{\phi(\mathbf{x},t)}{\phi_0(\mathbf{x})} = 1 - \frac{\phi_b(\mathbf{x},t)}{\phi_0(\mathbf{x})} \tag{4.11}$$

The fraction of the pore volume occupied by the biological mass, $\phi_b(\mathbf{x},t)$, can be expressed in terms of M_b, the relative biomass attached to the soil, and the biomass density ρ_b, [Clement 1996], such that

$$\phi_b(\mathbf{x},t) = \frac{\rho_k(\mathbf{x})}{\rho_b} M_b(\mathbf{x},t) \tag{4.12}$$

In general, biomass growth occurs in four stages: time-lag (adaptation), exponential growth (microbes have acclimated), stationary (limiting substrate), and decay (substrate exhausted) [Zwietering 1990]. While assessing the performance of SP, one is concerned with the initial stages of bio-clogging in which biomass grows exponentially,

$$M_b(\mathbf{x}) = M_b^*(\mathbf{x})[\mathrm{e}^{\lambda_s(\mathbf{x})t} - 1] \tag{4.13}$$

where $M_b^*(\mathbf{x})$ is the initial distribution of biomass and $\lambda_s(\mathbf{x})$ is the microbial growth parameter. Combining (4.11)–(4.13) yields

$$\frac{\phi(\mathbf{x},t)}{\phi_0(\mathbf{x})} = 1 - \frac{\rho_k(\mathbf{x})M_b^*(\mathbf{x})}{\rho_b\phi_0(\mathbf{x})}[\mathrm{e}^{\lambda_s(\mathbf{x})t} - 1] \tag{4.14}$$

For $\lambda_s t \ll 1$, (4.14) can be approximated by

$$\frac{\phi}{\phi_0} \approx 1 - \frac{\rho_k M_b^* \lambda_s}{\rho_b \phi_0} t \approx \mathrm{e}^{-\lambda_b(\mathbf{x})t}, \qquad \lambda_b = \frac{\rho_k M_b^* \lambda_s}{\rho_b \phi_0} \tag{4.15}$$

This approximation implicitly assumes the variation in porosity due to biofilm growth to be small relative to the initial biomass. It is adequate for risk assessment purposes, since a large reduction in porosity and permeability would make the SP operation not viable. In other words, standard SP operations would not allow clogging to develop beyond the exponential growth phase.

Substituting (4.15) into (4.5) yields a model describing the exponential reduction in the infiltration capacity due to biological clogging,

$$I_c(\mathbf{x},t) = I_{c0}(\mathbf{x},t)\mathrm{e}^{-3\lambda_b(\mathbf{x})t} \tag{4.16}$$

4.3.2.3 Chemical clogging

During the initial filling of a pond, gas can be generated within pores by a number of chemical processes, including microorganism activity, temperature effects, and the release of trapped bubbles [Christiansen 1944, Fry 1995]. As the air continues to occupy some pores (mainly large ones), water saturation and consequently hydraulic conductivity and infiltration rates decrease. Degassing can be rapidly reversed if the proper water temperature conditions are met [Perez-Paricio 2000]. The relationship between moisture content and conductivity is also characteristic of the grain size (i.e. type of soil); when this is similar to the pore distribution, the heterogeneous distribution of gas clogging factors would also show a spatial variability similar to the one of the grain size. However, direct measurements of the characteristic curves are challenging and costly. We take advantage of the fact that clogging due to gas formation takes place at small time scales (much smaller than those associated with physical and biological clogging). This allows us to employ an instantaneous reduction model of gas clogging,

$$I_c(\mathbf{x},t) = R(\mathbf{x})I_{c0}(\mathbf{x}) \tag{4.17}$$

where the reduction factor $R(\mathbf{x}) \in (0,1]$ is treated as an uncertain (random) fitting parameter.

4.3.2.4 Effective model of clogging

Although a few studies examining the interaction between processes exist (e.g. [Ernisee 1975, Rosowski 1986, Avnimelech 1983] for physical and biological clogging), most scarcely

provide quantitative information about the cumulative effect on reducing the infiltration capacity. For modeling purposes, we obtain the overall reduction in the infiltration capacity by linearly combining (4.10), (4.16), and (4.17), such that

$$I_c(\mathbf{x},t) = R(\mathbf{x}) I_{c0}(\mathbf{x}) \mathrm{e}^{-3[\lambda_p(\mathbf{x})+\lambda_b(\mathbf{x})]t} \tag{4.18}$$

4.4 Dependence of SP Performance on Soil Parameters

Predictions of clogging and the corresponding reduction in infiltration capacity depend on a number of parameters, whose values are highly uncertain due to soil heterogeneity. We adopt a probabilistic framework both to predict the effects of clogging on a SP's infiltration capacity and to quantify the predictive uncertainty.

Specifically, we focus on the impact of soil texture, as encapsulated by the grain size parameter d_g, on the reduction in the SP infiltration capacity due to various clogging mechanisms. Recall that d_g can be directly related to the initial infiltration values I_0 using (4.2) and (4.7), so that we could alternatively have selected $I_{c0}(\mathbf{x},t)$ or $K_0(\mathbf{x})$. In section 4.3, we demonstrated how the reduction in the infiltration capacity $I_c(\mathbf{x},t)$ in (4.18) can be related to the soil texture and to the soil particle size d_g. The following sensitivity analysis is used to explore further the question of how this soil parameter affects various clogging mechanisms and, via (4.18), the SP infiltration capacity.

The following grain sizes and soil densities were chosen to represent three different types of soils:

- $d_g = 1.5{\cdot}10^{0}$ cm and $\rho_k = 1.8$ g/cm^3 for soils of type (a),
- $d_g = 1.5{\cdot}10^{-2}$ cm and $\rho_k = 1.5$ g/cm^3 for soils of type (b),
- $d_g = 1.5{\cdot}10^{-4}$ cm and $\rho_k = 1.2$ g/cm^3 for soils of type (c).

The other parameters used for these simulations are reported in Table 4.1, taken as typical values encountered in the literature cited within this work. Note that the use of $\lambda_s = 2.5{\cdot}10^{-3}$ 1/day renders the approximation (4.15) applicable for t on the order of hundreds of days. Finally, we fix $\phi_0 = 0.3$ and $|\mathbf{q}| = 0.1$ m/day even though these parameters are expected to vary with d_g. This is done to isolate the relative importance of the physical and biological clogging, both of which depend exclusively on the grain size d_g.

Figure 4.1 illustrates the temporal evolution of the infiltration capacities, normalized by the corresponding initial values, considering individual and joint clogging process. We observe that biological processes play the dominant role in the overall clogging ($\lambda_p + \lambda_b$) for soils of type (a) (the coarser soils in our selection), while physical clogging is dominant in type (c) (the finer soils). For mid grain-sized soils, both mechanisms have similar impact on the overall clogging.

Table 4.1: Selected values for model entry parameters used for sensitivity analysis

Parameter	value	unit
d_s	$2{\cdot}10^{-3}$	cm
ρ_s	1.5	g/cm^3
n	2	-
T	298	K
ρ_w	1	g/cm^3
μ_w	$1.002{\cdot}10^{-7}$	N·s/cm^2
ρ_b	$2.5\ {\cdot}10^{-3}$	g/cm^3
λ_s	$2.5{\cdot}10^{-3}$	1/day
ϕ_0	0.3	-
$\|\mathbf{q}\|$	0.1	m/day

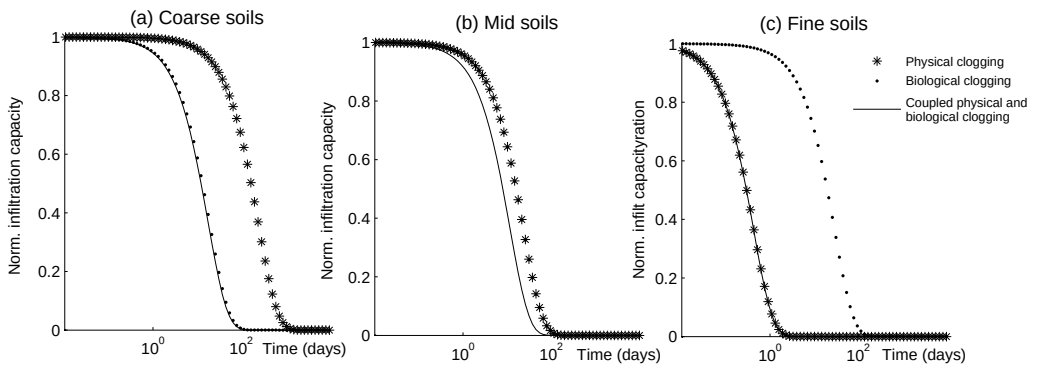

Figure 4.1: Reduction of the infiltration capacity of the SP topsoil (normalized with the initial value), calculated using Eq. (4.18), due to biological and physical clogging at the three different location of a SP characterized by locally uniform grain size distribution, corresponding respectively to (a) coarse, (b) mid-grained, and (c) fine soils. For mid-grained soils, physical clogging (stars) and biological clogging (points) practically overlap.

4.5 Applications

We demonstrate the applicability of the proposed approach with two examples. The first deals with field data collected at a pilot SP site (section 4.5.1). The second considers several synthetic examples that enable one to analyze the approach accuracy and robustness, and to quantify predictive uncertainty (section 4.5.2).

4.5.1 Pilot SP in Sant Vicenç dels Horts, Spain

We use the mathematical framework developed in Sections 4.3 and 4.4 to predict soil clogging and the corresponding reduction in the pond's infiltration capacity at a SP site located in Sant Vicenç dels Horts (herein, SVH) near the city of Barcelona, Spain (see Chapter 2)

The data analysis presented in Chapter 2 provides a rare example of SP, in which the initial local infiltration capacities are known with sufficient certainty. In general, this is not the case, and uncertainty (randomness) in $I_{c0}(\mathbf{x})$ should be accounted for in decision-making. In the analysis below, we explore both scenarios.

In this work we use Fig. 4.2A as our reference raster of initial infiltration capacities ($I_{c0}(\mathbf{x})$). Notice that the infiltration map here is a mirror version of fig. 2.7.

Figure 4.2B exhibits the characteristic grain size distribution $d_g(\mathbf{x})$ inferred from the distribution of $I_{c0}(\mathbf{x})$ in Fig. 4.2A by means of the Hazen formula (4.7). The logarithmic color scales in Fig. 4.2 highlight the high degree of spatial variability of the parameters considered in this study at the local (i.e., pixel) scale ($\approx 10^2$ cm^2). The parameters are inferred, and assumed to be constant, on a pixel basis.

The estimates of the grain-size distribution $d_g(\mathbf{x})$ in Fig. 4.2B rely on the relationship between the pixel intensity and the infiltration rate and on the Hazen formula (4.7), both of which are likely to introduce estimation errors. The impact of uncertainty on predictions of the SP infiltration capacity is quantified in Section 4.5.2 by conducting a series of Monte Carlo simulations. In the remainder of this subsection, we treat the grain-size distribution $d_g(\mathbf{x})$ in Fig. 4.2B deterministically, i.e., as a realization of the corresponding random field.

Figure 4.3 shows the estimates of physical clogging rates $\lambda_p(\mathbf{x})$ (Fig. 4.3A) and biological clogging rates $\lambda_b(\mathbf{x})$ (Fig. 4.3B) obtained from the estimates of the characteristic grain sizes $d_g(\mathbf{x})$ by following the procedure described in section 4.3.2.1 and 4.3.2.2. Note that the values of λ_p span four orders of magnitude, while λ_b is relatively uniform. In these calculations, we adopt the same default values for entry parameters as in Table 4.1. ρ_k was related to d_g according to the relationship established in Section 4.4.

To simplify the presentation, we take the reduction factor R due to gas clogging to be spatially homogeneous and set it to $R = 0.9$. Spatially-varying $R(\mathbf{x})$ can be obtained by invoking physical models combined with field measurements. Substituting $\lambda_p(\mathbf{x})$ and $\lambda_b(\mathbf{x})$ from Fig. 4.3 into (4.18), we compute the infiltration capacity $I_c(\mathbf{x}, t)$ in each pixel. Temporal snapshots of the resulting $I_c(\mathbf{x}, t)$ after $t = 7$, 14

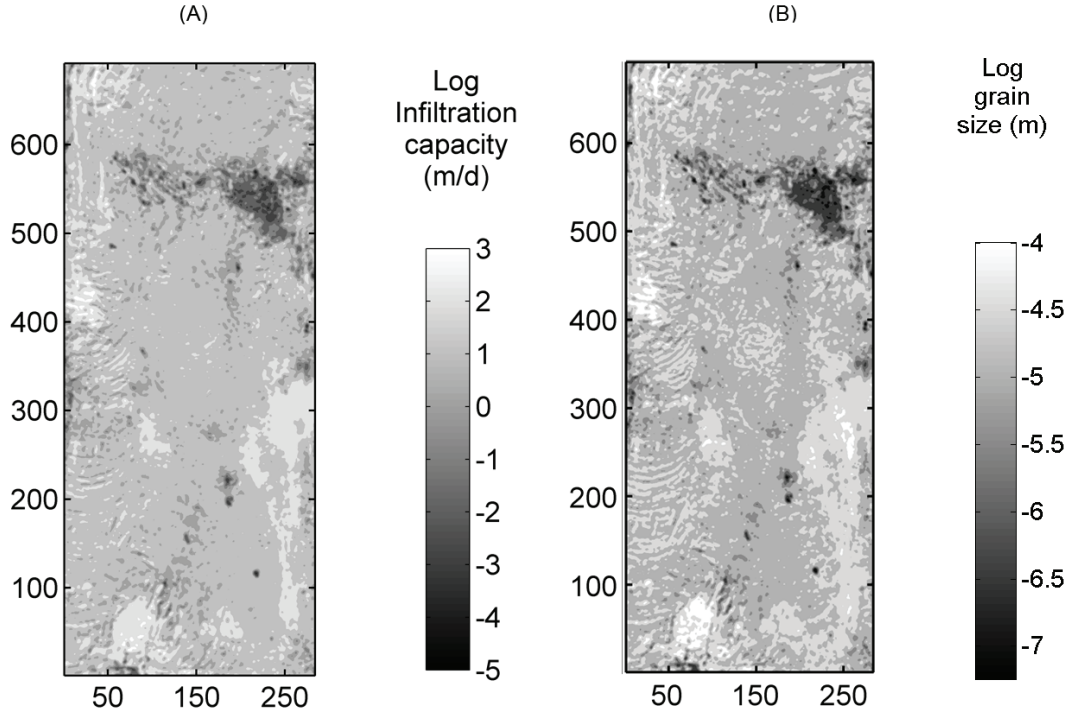

Figure 4.2: (A) Reference map of initial infiltration capacity distribution (modified from Chapter 2), and (B) relative grain-size distribution predicted with the Hazen formula (4.7). The size of the image is 286 × 694 pixels. Coordinates are also in pixels.

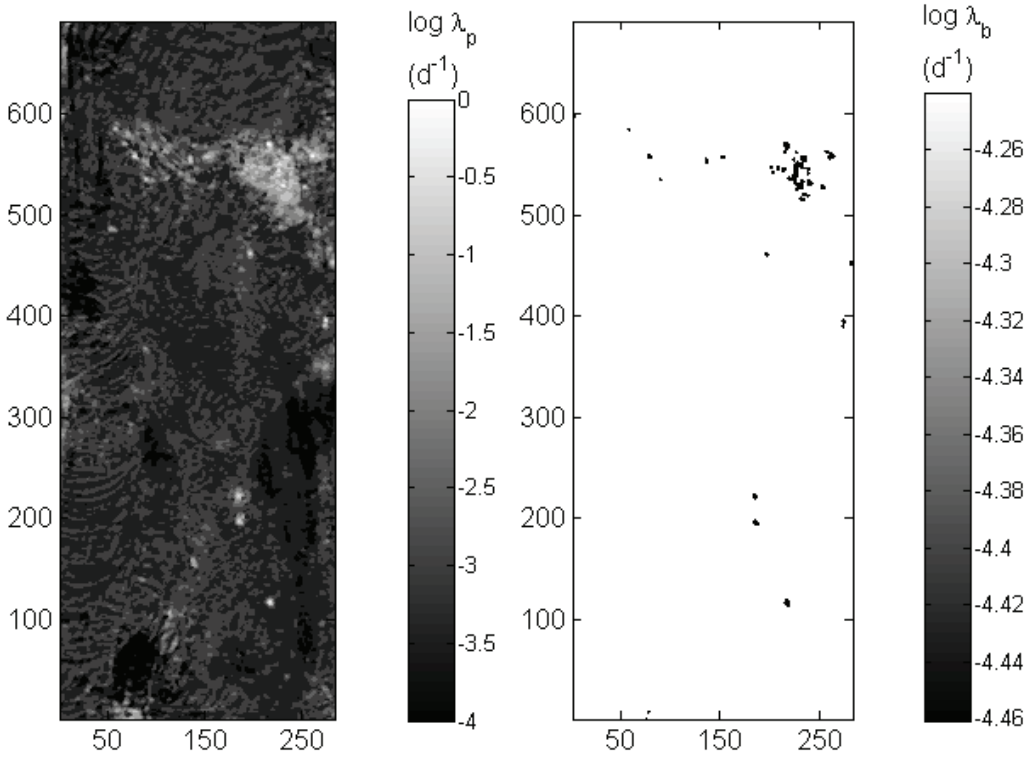

Figure 4.3: Spatial distributions of (A) the physical clogging rates λ_p and (B) the biological clogging rates λ_b, inferred from the spatial distribution of d_g in the reference domain. As such, each value has a pixel-size support volume. Notice scales are different in both plots.

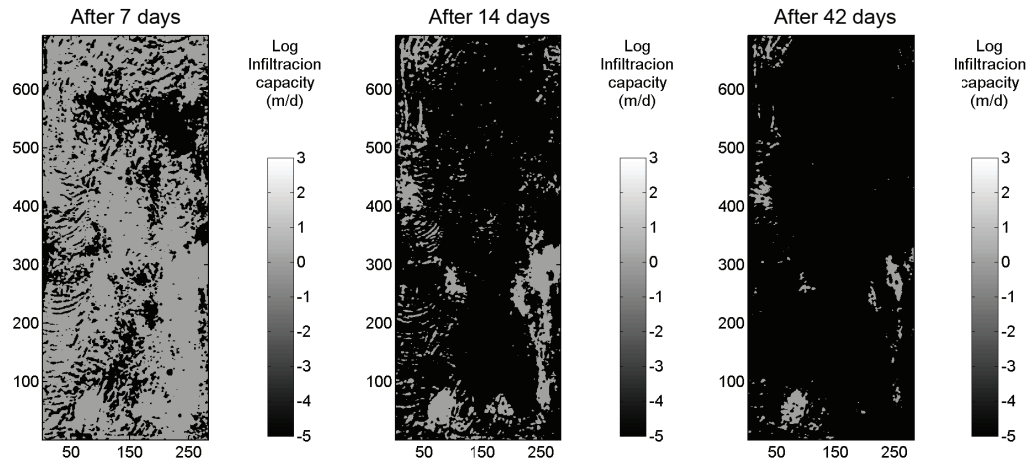

Figure 4.4: Spatial distribution of the infiltration capacities of the soil at different flooding stages (t = 7, 14 and 42 days) in the reference domain.

and 42 days of infiltration are shown in Fig. 4.4. A comparison of these snapshots with the initial infiltration capacity in Fig. 4.2A reveals a significant deterioration in performance of the SP. While a large fraction of the SP maintains s high infiltration capacity after 7 days, 42 days of SP operation results in a large reduction of the area corresponding to high infiltration.

Figure 4.5 summarizes the spatial variability of the infiltration capacity $I_c(\mathbf{x}, t)$ in Figs. 4.2A and 4.4 in the form of its probability density function (PDF) and cumulative distribution function (CDF) after 1 (t_1), 7 (t_7), 14 (t_{14}) and 42 (t_{42}) days of SP operation. These were computed from the corresponding histograms of pixel-level values of $I_c(\mathbf{x}, t)$. The infiltration-capacity PDFs at early times are broad, reflecting high uncertainty (spatial variability) of I. As time increases, the PDF shifts to the left, reflecting the decrease in infiltration rates. It also sharpens, indicating the decreased uncertainty. At all times, the PDF curves display asymmetric tailing (positively skewed curves)

4.5.1.1 On the use of effective clogging rates

Since previous observation show that large scale flooding experiments typically display an exponential decay in infiltration rate, we explore the feasibility of using an effective clogging rate λ_{eff} as a parameter to represent the various biological and physical clogging mechanisms over the whole SP footprint.

We compute the mean infiltration capacity $\overline{I}_c(t)$ by averaging the values of $I_c(\mathbf{x}, t)$ in (4.18) over all pixels. The resulting $\overline{I}_c(t)$, normalized with the average initial infiltration capacity $\overline{I}_{c0}$, is shown in Fig. 4.6a. Using a least-squares regression, the best fit exponential curve approximating the rate of change of the infiltration capacities gives a constant-in-time $\lambda_{\text{eff}} \approx 0.139$ d^{-1}, which means a characteristic

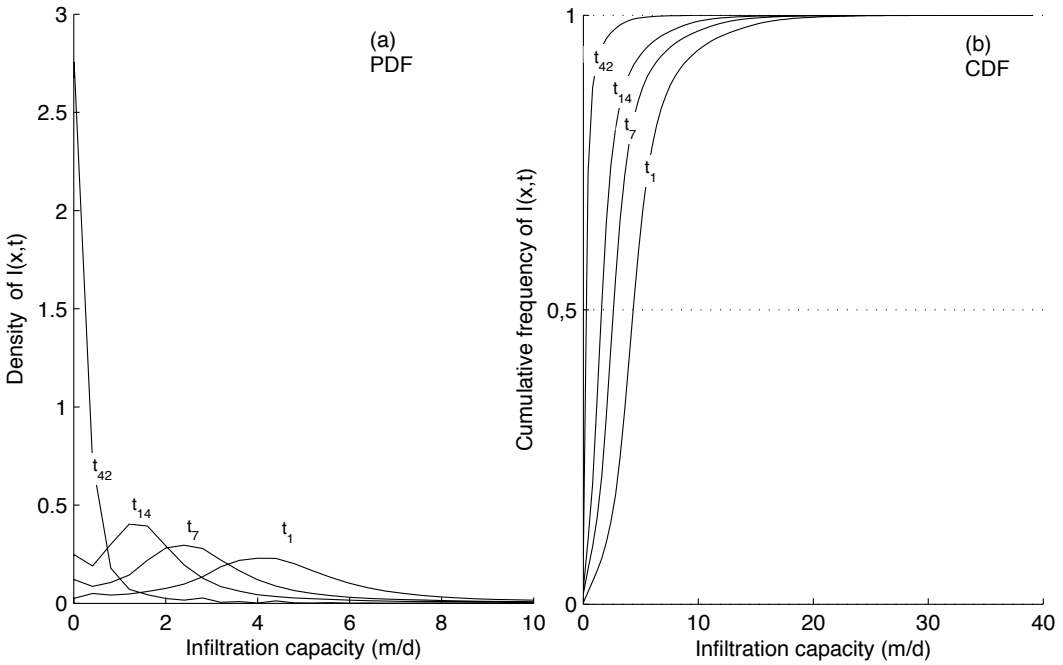

Figure 4.5: (a) Probability density function and (b) cumulative distribution function of the pixelized infiltration capacities $I_c(\mathbf{x}, t)$ after 1 (t_1), 7 (t_7), 14 (t_{14}), and 42 (t_{42}) days of SP operation at the SVH site.

clogging time of the water in the pond of 7.2 days .

Alternatively, we can estimate a variable-in-time $\lambda_{\text{eff}}(t)$ by invoking (4.1), such as

$$\lambda_{\text{eff}}(t) = -\frac{1}{t} \ln \left[\frac{\int_\Omega I_c(\mathbf{x}, t) \mathrm{d}\mathbf{x}}{\int_\Omega I_{c0}(\mathbf{x}) \mathrm{d}\mathbf{x}} \right] \tag{4.19}$$

The results of the two methods to calculate $\lambda_{\text{eff}}(t)$ are plotted together in Fig. 4.6b. The numerical evaluation of (4.19) suggests that in this case, λ_{eff} is monotonically decreasing in time, ranging from 0.155 d^{-1} to 0.125 d^{-1} over the chosen time interval (up to 100 days). This corresponds to characteristic clogging times between 6.45 days to 8 days (Fig. 4.6b). From a practical perspective, this difference may be negligible and a constant effective parameter λ_{eff}, exhibiting a "homogeneous" (or, more appropriately, homogenized) exponential decay, can be considered a good approximation (see the good fit in fig. 4.6).

4.5.1.2 SP management strategies

Next, we explore how spatial variability of the SP infiltration capacity $I(\mathbf{x}, t)$ affects the efficiency of the following four maintenance strategies:

- No maintenance activity is performed,

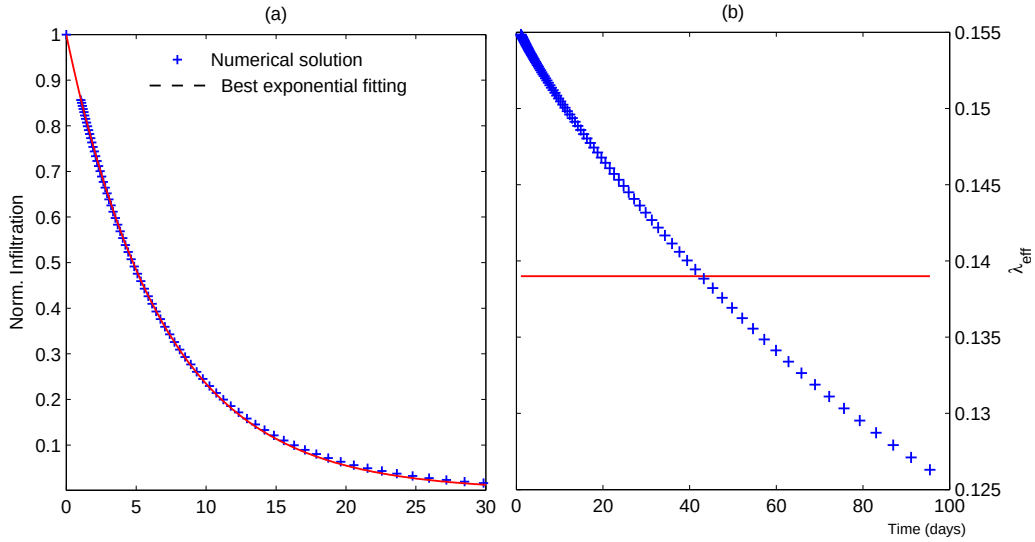

Figure 4.6: (a) Reduction in the overall infiltration capacity $\overline{I}_c(t)$ at the reference site,, computed by averaging its spatially-distributed counterpart $I_c(\mathbf{x}, t)$ (crosses) and the exponential fitting to the solution (solid line) with λ_{eff}=0.139 d^{-1}; (b) comparison between best-fit λ_{eff} (steady variable) and the time-dependent counterpart calculated from the numerical solution (4.19).

- All biological clogging mechanisms are remediated (Type A),
- All physical clogging mechanisms are remediated (Type B),
- Physical clogging mechanisms are remediated in selected areas of the SP footprint (Type C).

Type C maintenance can employ different criteria to identify the parts of the SP footprint where the remediation is to take place. One could clean an area selected purely on geometrical criteria (e.g., target a half of the area of the pond at a time). We on the other hand pursue a Type C maintenance strategy that relies on soil heterogeneity and hydraulic criteria to select clean-up areas. Specifically, we target the areas, wherein λ_p in Fig. 4.3A falls below a certain threshold value, e.g., 50% or 80% percentiles of the λ_p probability distribution (Fig. 4.7). Since λ_p is high for points displaying low I_0 and vice-versa, such maintenance strategies focus on the areas with the highest initial infiltration capacity I_0. This allows one to sustain higher infiltration rates longer, thus resulting in a decrease in the reduction of the mean infiltration capacity over long time. Complex geometries over which such a maintenance is to be performed present practical implementation difficulties. Thus, this approach has little practical significance at the current stage of SP practices but could be potentially interesting for future large scale facilities.

Figure 4.8 displays the average infiltration capacity $\overline{I}_c(t)$ achieved with the four maintenance strategies identified above. Unsurprisingly, all the strategies increase

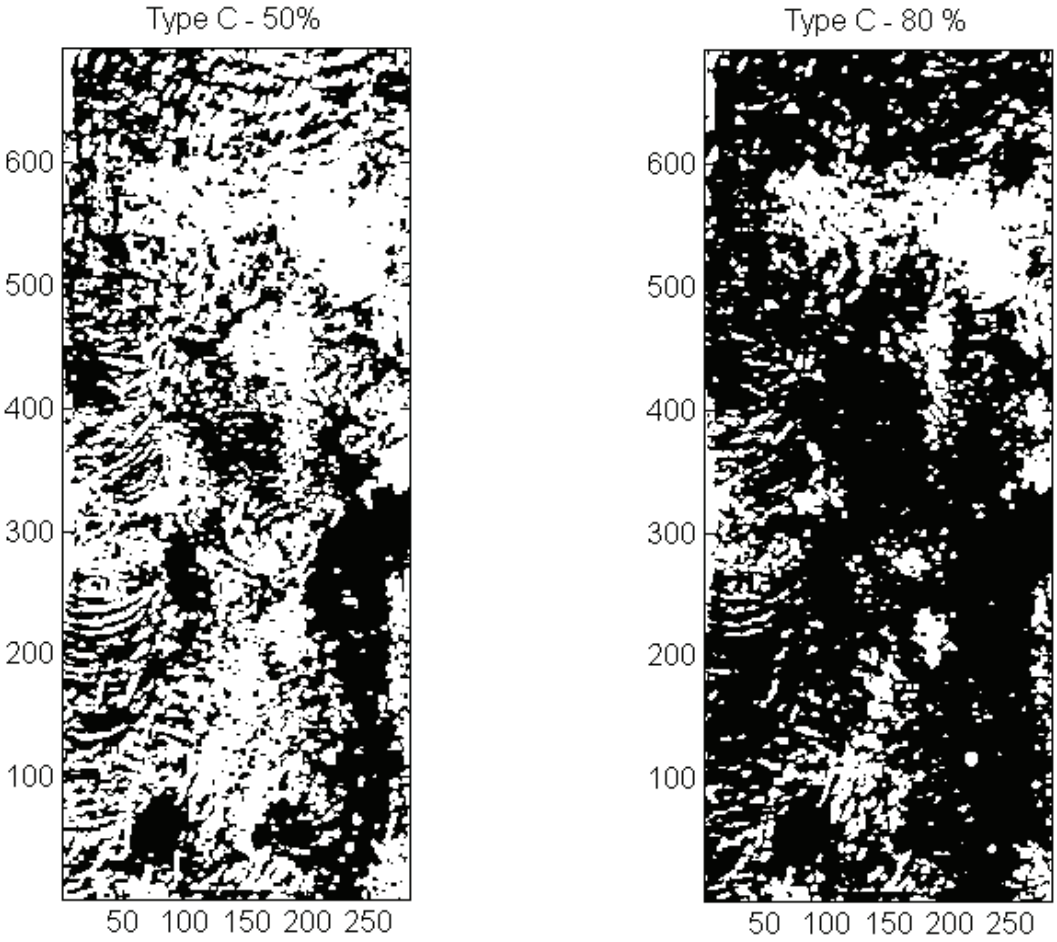

Figure 4.7: Areas corresponding to the 50% (left) and 80% (right) percentiles of the initial I_0 reference distribution. These black areas should be targeted in order to minimize the impact of physical clogging rate (Type C maintenance)

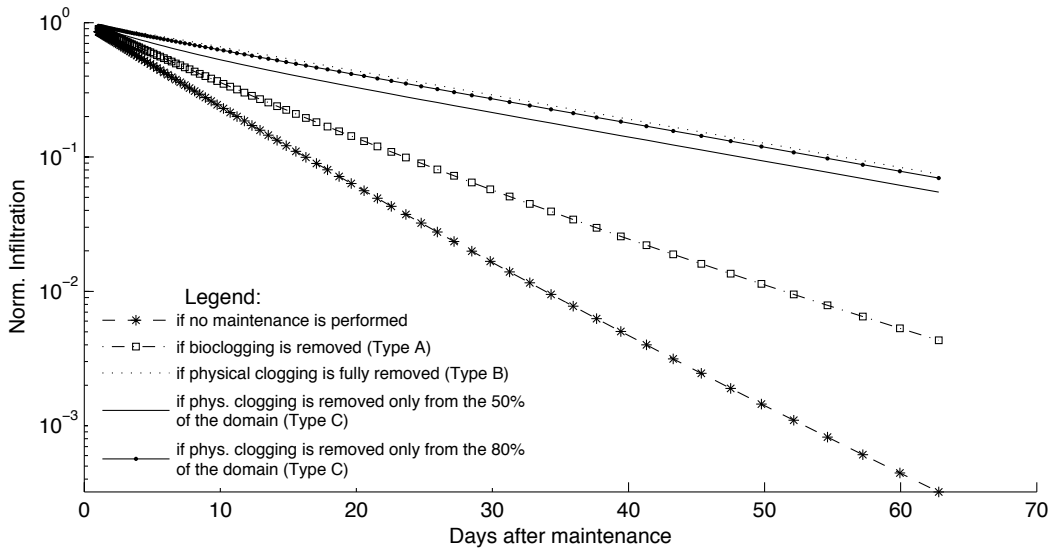

Figure 4.8: Reduction in the normalized mean infiltration capacity, $\overline{I}_c(t)/\overline{I}_{c0}$, of the reference site for the alternative maintenance strategies.

$\overline{I}_c(t)$ relative to its counterpart without any maintenance. If one defines 37% of the initial infiltration rate (corresponding to a characteristic time t=$1/\lambda_{\text{eff}}$) as the minimum acceptable infiltration capacity, then the SP's operational time without maintenance is about 7 days. Treating bio-clogging extends the operational time to about 9.5 days, and the full treatment of physical clogging increases the operational time to about 28 days. This analysis reveals that at this site physical clogging is the primary inhibiter for maintaining effective mean infiltration rates.

If only a partial maintenance of physical clogging is performed, the treatment of 50% of the domain extends the operational time to 18-20 days. A more extensive treatment, 80% of the domain, delivers a substantial gain over the untreated case of about 25-27 days. Thus, if feasible, a partial targeted maintenance offers significant gains in operational time (and thus could be a valid cost-efficient management alternative).

4.5.2 SP operations under uncertainty

In the preceding analysis on the pilot SP, we assumed that all the relevant parameters to compute (4.18) are known with certainty. Such a degree of certainty is uncommon in practical situations, where d_g, K_0, I_0, R, λ_b and λ_p are sparsely sampled and highly uncertain. This parametric uncertainty translates into uncertainty in predictions of whether and when the predicted mean infiltration rate $\overline{I}_c(t)$ falls below a critical value I_c.

To explore the effects of parametric uncertainty, we supplement our analysis

of the SVH site with a number of synthetic fields whose ensemble statistics are loosely based on the data obtained from this site. Rather than treating d_g as our random variable, we take the hydraulic conductivity, which we treat as deterministically linked to d_g as random. Specifically, as is commonly done, we assume that the initial natural log hydraulic conductivity $Y_0(\mathbf{x}) = \ln K_0(\mathbf{x})$ is a multi-Gaussian random field with zero mean $\overline{Y_0} = 0$, variance σ_Y^2, an exponential variogram, and correlation length I_S. We set h=0.5 m (constant in space and time) to compute I_0 from K_0 using (4.2) and A=100 to compute d_g from K_0 using (4.7). Since A and h are constant in space and time, the spatial structure is identical for d_g, K and I_0. Random realizations of $Y_0(\mathbf{x})$ were generated with the sequential simulation algorithm GSLIB [Deutsch 1998] on a computational domain comprised of 284 $\times$ 692 pixels. These realizations were not conditioned to data.

Section 4.5.2.1 contains an analysis of the infiltration capacities $\overline{I}_c(t)$ for single realizations of soil parameters. Its goal is to verify the robustness and generality of our findings at the SVH site. In Section 4.5.2.2, we perform Monte Carlo simulations (MCS) to quantify predictive uncertainty.

4.5.2.1 Analysis of individual (single) realizations

Consider four SP operating in soils, whose distributions of initial infiltration capacities $I_{c0}(\mathbf{x})$ are shown in Fig. 4.9. These fields are individual realizations of random fields obtained using different parameters, e.g. (a) $\sigma_Y^2 = 4$ and $I_S = 66$, (b) $\sigma_Y^2 = 4$ and $I_S = 6.6$, (c) $\sigma_Y^2 = 1$ and $I_S = 66$, and (d) $\sigma_Y^2 = 1$ and $I_S = 6.6$. In all cases the units for I_S are pixels.

Fig. 4.10 shows how the various maintenance strategies described in Section 4.5.1.2 affect the decay in the normalized infiltration capacity $\overline{I}_c(t)/\overline{I}_{c0}$ in the four synthetic fields (Fig. 4.9). Note here that when we say Type C maintenance it refers to the de-clogging of 50% of the pond surface.

For $\sigma_Y^2 = 4$ and $I_s = 66$, the results suggest that removal of biological clogging (Type A) yields better performance than applications on physical processes (Types B,C). For instance, 40 days after recharge has started, the reduction of infiltration from its initial value is about 80% using Type B and C maintenance (which practically overlap in this case), while there is a reduction of around 70% with Type A. For the same variance of $\sigma_Y^2 = 4$ but a comparatively small correlation length ($I_S = 6.6$), Type B and C offer slightly better performance than Type A, although after around 40-45 days of operation this behavior is inverted. However, the late time difference is quite small.

On the other hand, for the smaller variance (more homogeneous) system with $\sigma_Y^2 = 1$, biological clogging always appears to play a secondary role when compared to physical clogging effects in direct analogy to what we observed at the SVH site (Fig. 4.8). This is true for both the I_S = 66 and 6.6 realizations. Maintenance applied in the effort to remove biological clogging is practically useless relative to doing nothing. On the other hand both physical clogging maintenance approaches are much more effective.

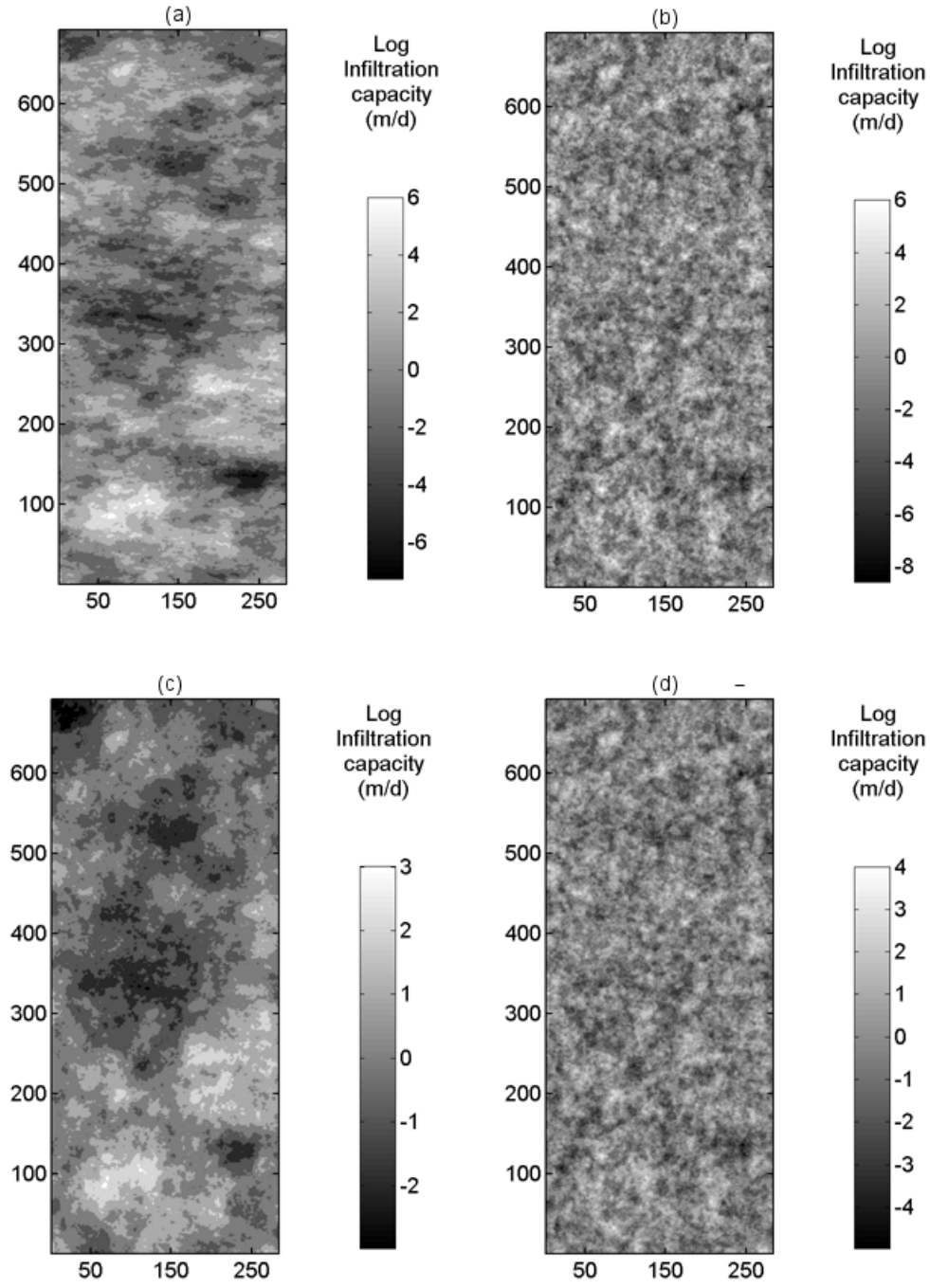

Figure 4.9: Individual realizations of the initial infiltration capacities $I_{c0}(\mathbf{x})$, using different variances of soil hydraulic conductivity σ^2_{Y} and correlation length I_S: (a) $\sigma^2_{\mathrm{Y}} = 4$ and $I_S = 66$, (b) $\sigma^2_{\mathrm{Y}} = 4$ and $I_S = 6.6$, (c) $\sigma^2_{\mathrm{Y}} = 1$ and $I_S = 66$, and (d) $\sigma^2_{\mathrm{Y}} = 1$ and $I_S = 6.6$. Units of I_S in pixels

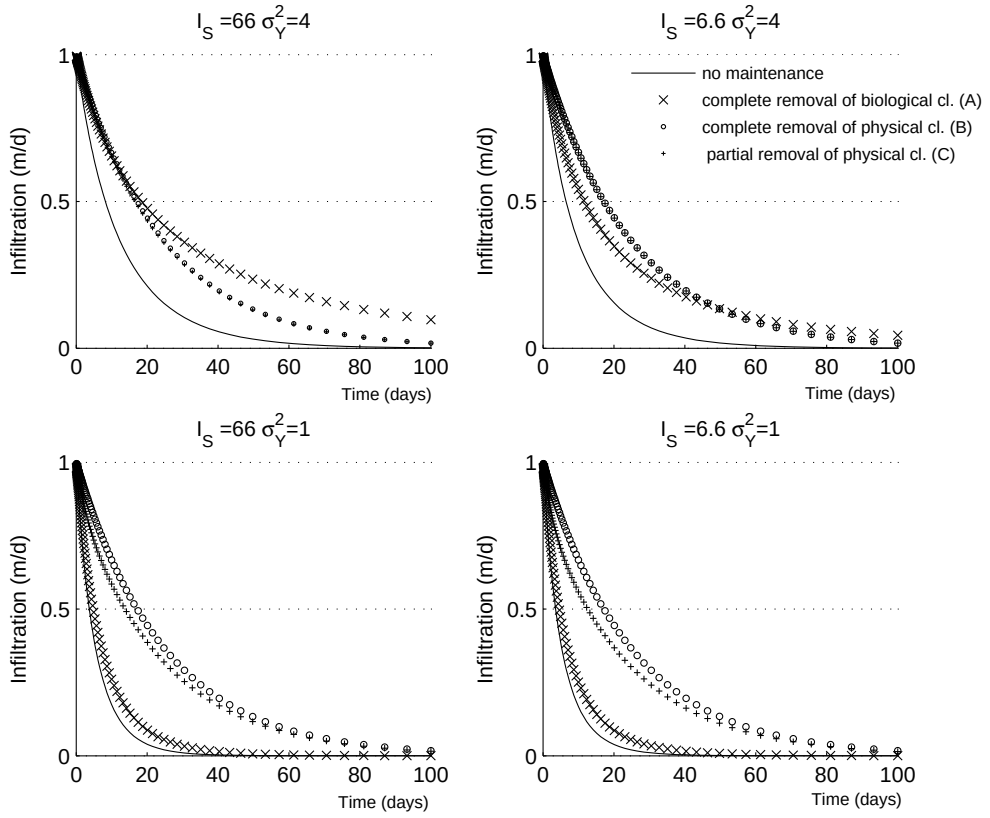

Figure 4.10: Impact of the different types of maintenance on the normalized infiltration capacity $\overline{I}_c(t)/\overline{I}_{c0}$ at the four SP sites shown in Fig. 4.9.

Note that in all cases the selective removal of physical clogging, i.e. Type C maintenance, is almost as effective as the complete removal maintenance, particularly for the high variance cases. This is because this maintenance keeps the dominant flow channels open and the low permeability regions affect the mean value very little.

4.5.2.2 Monte Carlo analysis

To quantify the uncertainty associated with predictions in the decline in the infiltration capacity of the four SP in Fig. 4.9, we conduct four sets of Monte Carlo simulations (MCS). Each set consists of 1000 realizations of the initial log-conductivity fields $Y_0(\mathbf{x})$ whose statistical properties are described in Section 4.5.2.1. As in the case of the single realizations, we use (4.2) and (4.7) to relate d_g and the initial infiltration to $Y_0(\mathbf{x})$.

Figures 4.11 and 4.12 display some of the most significant results of the analysis. The former displays the normalized spatially-averaged infiltration capacities against time ($\overline{I}_c(t)$) for each of the four statistical fields. The figure includes the ensemble mean of the predictions over all realizations, along with the degree of uncertainty

for each prediction expressed by the coefficient of variation (CV). Note that apart from the case with $I_S = 66$ and σ_Y^2 =4, the ensemble mean curves of Type B and C (marked by squares and circles) practically overlap.

One feature that stands out is that CV, initially zero (we assume perfect knowledge of I_{c0}) increases towards larger values at different rates, that depend on the type of maintenance and geometrical distribution of soil parameters. The most uncertain cases are those with biological maintenance and with no maintenance at all, while the uncertainty is generally quite small for the system with maintenance of physical clogging. This is in some sense a reflection of the fact that maintenance of physical clogging homogenizes the system. Interestingly, while biological processes are also heterogeneous, their range of heterogeneity is much smaller and so only treating biological processes maintains a high degree of uncertainty in the system with late time coefficients of variation in many cases reaching values of O(1). Comparing physical clogging maintenance schemes B and C demonstrates that the ensemble mean behavior is practically unchanged except with a small visible difference in the most heterogeneous system where $I_S = 66$ and σ_Y^2 =4. As already highlighted, the uncertainties are small for both cases with a largest value around $O(10^{-2})$ in the most heterogeneous systems. The uncertainty is always smallest for Type B maintenance, which is rather intuitive since it removes all physical clogging, while C only targets it partially, thus still maintaining a certain degree of physical clogging heterogeneity in the system. It is shown that biological maintenance results in the highest degree of uncertainty, which raises an important point for managers and decision makers: if dealing with a highly heterogeneous pond, one would have to decide whether to gamble with the uncertainty associated with maintenance A or take the more certain approach of maintenance against physical clogging.

To highlight the main observations from the Monte-Carlo simulations in a different way we present the cumulative distribution functions (CDF) of normalized infiltration at various times in figure 4.12. The CDFs are shown for each of the 16 cases (4 ponds with 4 maintenance strategies) at five different times t_1, t_7, t_{14}, t_{21} and t_{42}, where the subscript corresponds to the number of days elapsed. In figure 4.12 the cases, which suffer from the greatest degree of uncertainty, clearly stand out (e.g. the pond with $\sigma_Y^2 = 4$ and $I = 66$ with maintenance type A). On the contrary, many of the systems with low variance, particularly with maintenance types B and C have almost instantaneous jumps in their CDFs. These sudden jumps reflect a very high degree of certainty.

Finally to convey this information even more clearly we present selected probability density functions for some specific cases to highlight particular features in fig. 4.13. The PDFs are calculated as the best-fit non-parametric approximation to the discrete experimental histograms of infiltration capacities. We show the PDFs at various times for different maintenance activities, calculated for the case with greatest uncertainty ($I_S = 66$ pixels and $\sigma_Y^2 = 4$ m^2/day^2). The PDFs are shown at three different times, t_7, t_{14} and t_{42}.

The width of the PDFs reflects the degree of uncertainty and nicely visualizes some of the measures of uncertainty in figures 4.11 and 4.12. We would like to

highlight that in fig. 4.13, the shape of the PDFs for Type C is clearly different from the others; this is due to the fact that the type C maintenance selects to only maintain the higher infiltration rate channels, thus maintaining higher statistical mean infiltration capacity and inducing a negative skewness to the PDF. This is what increases the degree of uncertainty relative to full removal of physical clogging.

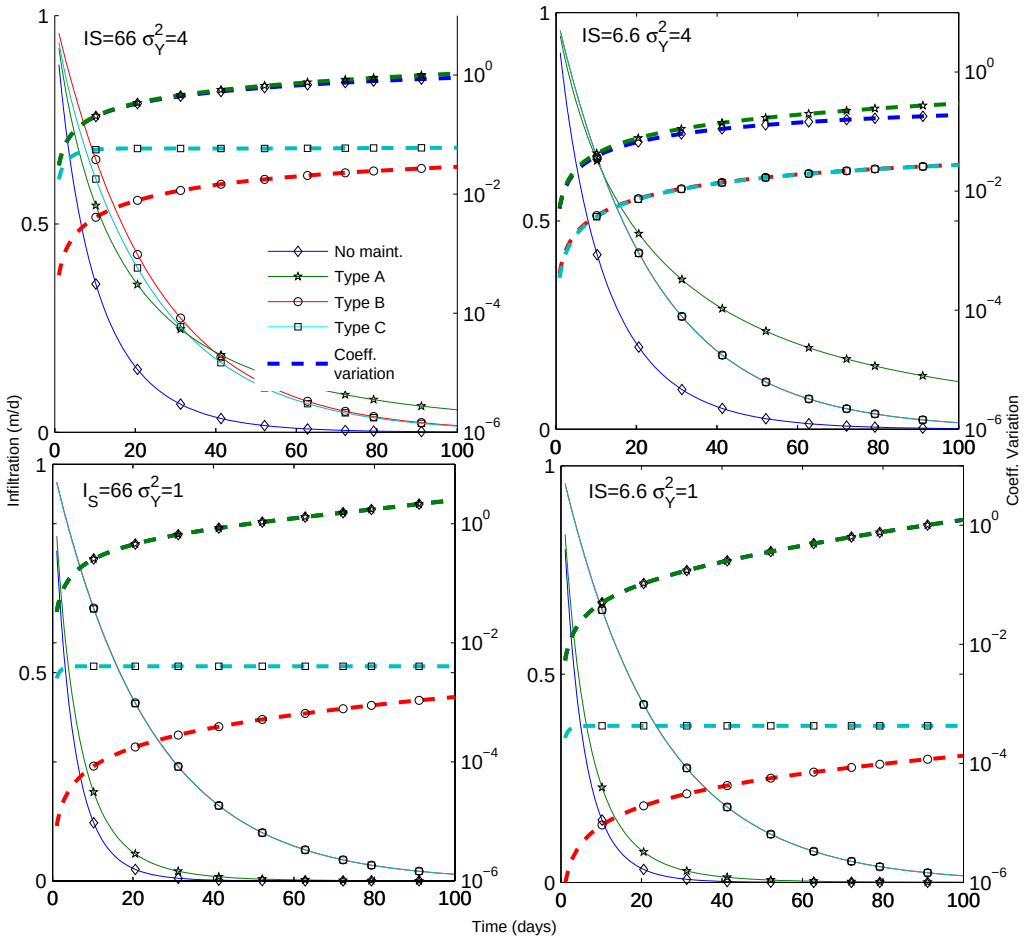

Figure 4.11: Ensemble average and coefficient of variation resulting from Monte Carlo analysis (1000 realizations) of the normalized overall infiltration capacity $\overline{I}_c(t)$ in time for each field geometry and variance, for different types of maintenance. Notice that Type B and C curves overlap in many cases, expressing a similar behavior.

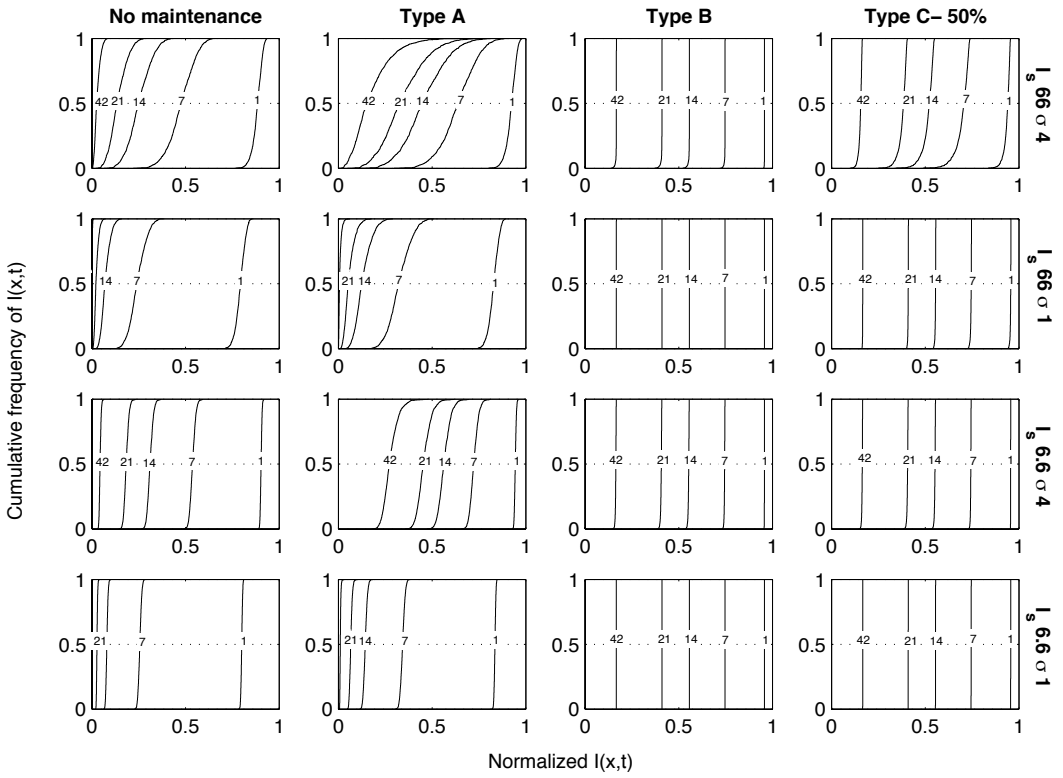

Figure 4.12: Cumulative distribution frequency (CDF) of the normalized overall infiltration capacity $\overline{I}_c(t)$ resulting from the Monte Carlo analysis. Results are shown at times t_1=1, t_7=7, t_{14}=14, t_{21}=21 and t_{42}=42 days. Results are shown for the four synthetic fields and four maintenance strategies.

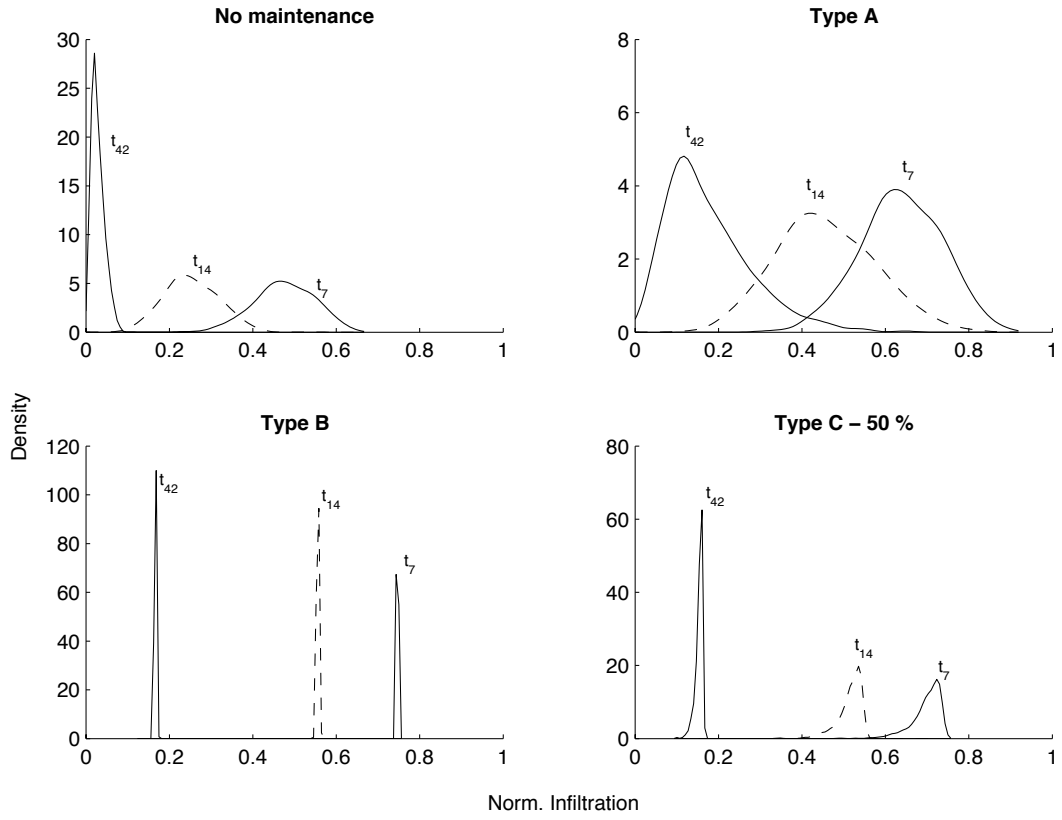

Figure 4.13: Probability density functions (PDFs) of the normalized overall infiltration rates, resulting from the Monte Carlo analysis. PDFs reflect the ensemble of fields synthetic realizations with $\sigma_Y^2 = 4$ and $I_S = 66$ pixels, for different types of maintenance applied on soils, at times t_7=7, t_{14}=14 and t_{42}=42 days.

Conclusions for part 1

The major conclusions that can be drawn from the first part are listed as follows and divided by chapters.

Using satellite images to map the infiltration capacity of topsoils

Dealing with spatial variability of soil hydraulic variables (like the infiltration capacity) in the field is always challenging and uncertain. This is due to the high cost of in situ site characterization. An alternative and low-cost method has been presented to map the infiltration capacity in heterogeneous alluvial formations from satellite images. Its was found that a relationship between infiltration capacity (measured in some random locations) and color intensity of two digital images can be well established to provide the spatial distribution of soil properties such as the infiltration capacity. The method was satisfactorily applied to an artificial recharge pond area close to the city of Barcelona.

The relationship we built appears to be robust, physically justifiable, and satisfies some premises like monotonicity of the correlation functions. In the test images, the ground surface appears not to be affected by objects or other disturbances (for instance, a presence of trees or shadows) that could have modified the natural color or reflectance in portions of the domain. This avoids the need to apply filters to the image to remove such imperfections, which would modify the original pixel organization, structure and intensities.

The physical validity is justified by noticing that the image depicts darker portions which have been associated with fine-grained portion of the soils, with substantial organic content and vegetation canopy, whereas brighter colors correspond to areas with cleaner sands and gravels of low fine content and vegetation. It is worth noting that the correspondence of dark soil to dark pixels and vice versa may not be universal (e.g. depending on the considered light spectrum range). Also different soil reflectance can depend on several factors such the type of light, solar ray inclination, time of day, season etc. As such, it is suggested that this method should be valid for a rapid assessment of areas that are geologically similar to the one we consider and for applications where the ground surface is free of covering materials. Both these conditions are characteristic of typical artificial recharge sites.

Most importantly, estimates of the total maximum infiltration rate obtained for the artificial pond site based on this relationship were consistent with a water mass balance performed during a first inundation experiment at the site. Predicted values match quite well with actual observations, according to official measurements made by the local public administration. Both digital images, the Internet digital image from a popular web based provider and the one officially provided by the Catalan Cartographic Institute, gave similar results, highlighting the robustness of the method. The discrepancies that do arise suggest that any estimates should be coupled with an uncertainty analysis to quantify the reliability of this method as

compared to other more sophisticated but more expensive ones.

Spatio-temporal mapping of infiltration capacity

Assessing the spatio-temporal variability of soil hydraulic variables is often needed to make optimal, effective and efficient decisions in many branches of the groundwater hydrology. A methodology to map the spatio-temproal distribution of the characteristic lumped temporal factors of clogging (λ) in a SP, with the aid of satellite images and a few experimental datasets based on local sparse measurements is provided.

The method allows to estimate the spatio-temporal variability of the infiltration capacities of the topsoil on the pilot artificial recharge area in Spain. Although the developed model introduces some simple hypotheses and simplifications in this illustrative example, it is shown that we were able to map the evolution of local infiltration with time. To validate the method, the measured total infiltration capacities at the SP scale is compared with the ones calculated by integration of calculated local values using a geostatistical-physically-mixed approach. Numerical results satisfactorily agree with the observations

It is found that arguably the most significant parameter is the sign of the a parameter in the power law model relating λ and the initial infiltration. Depending on this sign it is possible to see the practical relevance of designing a remediation method that targets only parts of the domain (acting on the high infiltrating areas during operation), or else it is better to use the classical approach of treatment after emptying the facility. This alternative might have significant impact in managing an SP site.

Maintenance of artificial recharge ponds under uncertainty

A stochastic framework to evaluate the infiltration rate in an artificial recharge pond over time is formulated and used to manage operations of maintenance of the pond under uncertainty. The framework focuses on physical, biological and retardation-like clogging mechanisms, although additional processes can easily be included. The general methodology is applied to two sets of single realization examples (a real SIP and four synthetic ones), and within a Monte Carlo framework.

Biological and physical processes was combined to produce a joint reduction of infiltration in time. The relative impact of one versus the other is governed by some typical parameter of the soil like the grain size. A simple sensitivity analysis on the models demonstrates that physical clogging is more sensitive to soil heterogeneity than biological clogging.

Detailed knowledge of the geological heterogeneities of the soil are a determining factor for optimal decision making in maintenance operations of the artificial ponds. Actions can be taken aimed at reducing the impact of biological or physical clogging independently. Some of them can be selectively applied to target portions of the domain, instead of being applied over the whole domain. A proper initial

characterization can significantly help to accurately predict the infiltration rate in a pond. Such an initial characterization can for example be conducted using primary or secondary information, such as satellite images.

The single realization approach considers the spatial distribution of initial infiltration capacity to be fully known at any location of the pond. It demonstrates how in a heterogeneous system the mean infiltration rate can be described with an effective homogeneous decay rate despite the large degree of heterogeneity.

The Monte Carlo framework shows that the risk of making optimal decisions depends on soil heterogeneity. Uncertainty mainly depends on the global variability of the hydraulic properties of the soil (risk of design failure increases as the variance of the hydraulic conductivity increases). The geometric distribution (evaluated by the correlation length) of such parameters also plays a role, but in most cases is secondary.

Maintenance actions that target physical clogging mechanisms significantly reduce the uncertainty in predicting the temporal evolution of the infiltration rate relative to only maintenance in relation to biological clogging or no maintenance at all. On the other hand, some soil characterization parameter combinations lead to a larger value of the ensemble infiltration capacity with time. Thus a decision maker must weight the relative gain in infiltration capacity to the gamble of uncertainty.

Part II

Solute transport

Prologue to part 2

Solute transport in heterogeneous media is highly uncertain to model and predict. The main reason is related to the existence of heterogeneity of hydrodynamic soil properties that determines random groundwater flow dynamics. Accurate modeling of such dynamics is needed to make decisions in the optic of risk assessment and prevention of hazardous situations management of remediation activities (e.g. [Tartakovsky 2007, de Barros 2008, Bolster 2009]).

For predictions to be reliable, models need to show accuracy and capacity to reproduce experimental characterization tests. However, defining an appropriate model for this purpose is challenging. Typical characterization scales are done at the order of 10^0 to 10^3 meters ([Gelhar 1992, Gelhar 1993]), while most of the processes controlling flow and transport fluctuations occur at the pore scale (10^{-6} to 10^{-2} meters). Possibly, in the next future, practical characterization methods will be able to explore microscopic levels of the soils, and at the same time practical models will be able to extensively and efficiently simulate directly flow and solute dynamics at such microscopic scales. Today, only very specific characterization methods and models have remarkably showed that this approach may work (e.g. [Morris 1997, Tartakovsky 2006]), indicating that this path must be followed and investigated. However, they are still in an embryonal stage as they require huge costs and high demanding computational efforts that impede their current use for practical purposes.

Nowadays, macroscopic models are required, invoking some scaling-up (e.g. averaging) procedure. For upscaling method to be effective, they require in turn an extensive knowledge about the evolution of the processes controlling flow and transport over several spatio-temporal scales. A classical macroscopic model that has been largely used so far is the advection-dispersion equation (ADE), which can be developed considering parallel or radial coordinates, depending on the geometry of the flow field ([Bear 1972, Moench 1989]). ADE relies on the Fickian laws and works for homogeneous cases, which are however not very common in nature. Most field-scale experiments showed that this model presents serious limitation to reproduce solute transport, and some author from the 1960's has been invoking the existence of additional apparent process that can control transport at large observation scales ([Coats 1964]). Transport that cannot be simulated by ADE is generally called 'non-Fickian', or 'anomalous'.

A macroscopic evidence for non-Fickianity are the non symmetric breakthrough curves (BTCs) resulting in most tracer test experiments. Many authors have postulated that BTCs non-Fickianity is a consequence of the presence of physical and chemical soil heterogeneities (e.g. [Levy 2003, Salamon 2007, Riva 2008]). Phenomenological evidences suggests that heterogeneity determines heavy-tailed distributions on BTCs that sometimes scale very similarly to a power-law (PL) distribution of the form $c \approx t^{-m}$ at late times (long after the peak is observed). The parameter m is also known as the BTCs slope, since PL distributions scale

as straight lines in double logarithmic plots; m has been reported to range between 1 and 3 in several cases (e.g. [Becker 2000, Becker 2003, McKenna 2001, Hoehn 1998, Fernàndez-Garcia 2004, Gouze 2008]). Other distributions than PL also display an apparent straight slope in double-log plots ([Clauset 2007]) and can be used to fit late-time behavior of experimental BTCs (e.g. the lognormal or the gamma probability density functions). PL distribution are, however, very attractive since some authors found a strict relationship between PL-shaped BTCs and some specific spatio-temporal distribution of some physical soil properties (e.g. [Dentz 2003, Bijeljic 2006, Willmann 2008, Gouze 2008] and reference therein). Nonetheless, the common thing in these works is that consider nonstationary fields or mean uniform flow conditions; for many other circumstance where heavily-tailed BTCs similar to PL distribution are observed, such as under radial convergent flow and in stationary sandy aquifers (e.g. [Fernàndez-Garcia 2004]), a clear relationship between physical parameters and BTCs scaling has been given yet.

Heavy-tailed BTCs are in most cases found to be very similar to those produced by 'mass-transfer' processes. Mass-transfer-based methods have been commonly used as an upscaling method to simulate non Fickian transport. While these methods were designed to simulate processes where part of solute was actually uptake to or stagnant in less mobile zones ([Coats 1964, Van Genuchten 1976, Brusseau 1990]), recent theories (e.g. [Haggerty 1995, Carrera 1998, Sanchez-Vila 2004]) and applications ([McKenna 2001, Gouze 2008]) showed that diffusive mechanisms can also lead to similar mass-transfer processes.

Pore-scale velocity contrasts from fluctuation of the hydraulic conductivity can also lead to late-time tailing on BTCs. This aspect has been observed for instance using detailed numerical techniques under uniform flow conditions ([Fiori 2006]). Mass-transfer-based models can be still used in this case for fitting purposes, but a direct relationship between mass-transfer parameters cannot be easily built. First of all, because of the fundamental difference between advective-dominated and diffusion-dominated processes. In the first case apparent mass-transfer is actually due to local flow fluctuations and while in the second scale it is driven by concentration gradients. From a mathematically perspective, diffusion is similar to first-order rate mass transfer (e.g. [Nkedi-Kizza 1984]).

Several authors sought for a possible apparent phenomenological link between mass-transfer parameters and advective-dominated anomalous transport (e.g. [Zinn 2003, Willmann 2008, Flach 2012]). To find such a link, these authors relied on the ability of two-dimensional numerical models under uniform flow conditions to reproduce anomalous tailing, in order to understand what dynamics control the shape of BTCs and and consequently control mass-transfer parameters. However, this model configuration presents limitations, both because of the model dimensionality and also because of the geometry of the flow field.

Two dimensional models are in most cases not able to reproduce flow and transport dynamics when injection-extraction distances are of the order of magnitude are vertical model dimensions. In addition, when horizontal integral scales of aquifer heterogeneity are similar to the vertical ones, three-dimensional models

are suggested (e.g. [Dagan 1989], Part 4). Comparing data reviews in [Gelhar 1992, Gelhar 1993], it seems that these conditions are usually met in most experimental sites. [Fernàndez-Garcia 2004] for instance indicate that anomalous behavior develops at a distance of a few horizontal integer scales under convergent radial flow; here three-dimensional tanks were employed and filled with clean sands of variable conductivity and without artificially-imposed connected channels. When injection distances are similar to the horizontal integer scales, flow and transport may tend to behave as in a stratified medium ([Dagan 1989], p.293). Three-dimensional models are also suggested to better reproduce connectivity patterns ([Fiori 2012]), which can have a large impact on anomalous solute transport [Zinn 2003, Willmann 2008, Trinchero 2008]. Two-dimensional models of stationary Multigaussian fields are very limited to simulate connected channels ([Sanchez-Vila 1996]); in effect, [Zinn 2003, Willmann 2008] had to artificially modify their Multigaussian fields to obtain long-connected structures and thus reproduce anomalous transport.

The aim of this second part of the thesis is to understand the nature of BTCs tailing when anomalous transport generating during conservative tracer tests in three-dimensional, stationary, Multigaussian and unconditional simulations of log-transformed hydraulic conductivity fields under radial convergent flow to a well. Knowing the nature of tailing is fundamental to understand whether mass-transfer models can be effectively applied to predict transport in these situations, where a sound mathematical relationship between solute transport parameters and mass-transfer parameters does not apparently exist.

The second part of this thesis is thus organized as follows.

In Chapter 5 a new tool that improve the estimation of BTCs from particle tracking distribution is illustrated. Particle tracking is an attractive method to accurately simulate transport in heterogeneous aquifers, but a reliable estimation of densities is usually cumbersome to obtain using traditional methods.

In Chapter 6, processes causing tailing on BTCs are analyzed. It is shown that tailing can be caused as a combination of three-dimensional radial convergent flow, stratification of the tracer after being injected in the aquifer and connectivity between the injection and the extraction well. The shape of these tails is very similar to those found during experiments and also very similar to tailing caused by mass-transfer processes.

Chapter 5

An automatic locally-adaptive method to estimate heavily-tailed breakthrough curves from particle distributions

5.1 Introduction

Among the several numerical methods existing to reproduce solute transport, particle tracking methods (PTM) provide an efficient and versatile method (e.g. [Prickett 1981, Kinzelbach 1987, Pollock 1988]). PTM are especially indicated to simulate Lagrangian advective and dispersive/diffusive transport in heterogeneous porous media under complex boundary conditions, which naturally occur during typical tracer tests for aquifer characterizations (e.g. [Tompson 1990]). In heterogeneous domains, flow velocity contrasts can be very high and traditional approaches such as Eulerian models fail as they would require a very fine grid discretization to catch the actual pore-scale dynamics controlling solute variability (e.g [Liu 2004]), which is most cases not feasible from a computational perspective in most common applications. On the other hand, PTM can reproduce diffusive/dispersive processes in absence of typical numerical problems encountered in other methods, such as oscillation or dispersion, and thus improving the modelling exercise in a cost-effective manner ([Salamon 2006b]).

PTM are based on the estimation of Lagrangian velocities taken by solute particles while travelling between known locations in the aquifer (e.g. [Kaasschieter 1995, Bensabat 2000, Pokrajac 2002]). Their application in many groundwater problem has been widely reported in the literature, especially for application in risk assessment and solute transport remediation and protection ([Shapiro 1988, Dagan 1992, Berglund 1995, Zimmerman 1998, Andricevic 1996, Bellin 2004, Andricevic 2008]). A general conclusion from these works is that, for a proper estimation of the risk, the entire shape of travel time probability density functions (PDFs) at a controlling section (i.e. the breakthrough curves, BTCs) is very important. For application involving design and maintenance of aquifer restoration and protection activities, PDFs also need to be smooth to adequately reproduce gradients

[0]This chapter is based on D. Pedretti and D. Fernandez-Garcia et al. (submitted to Advances in Water Resources), *An automatic locally-adaptive method to estimate heavily-tailed breakthrough curves from particle distributions*

(e.g. [Fernàndez-Garcia 2011]). Temporal gradients are used for instance to infer physical and chemical properties of the soils from PDFs, such as connectivity, fractional parameters, matrix-diffusion or mass-transfer parameters ([Carrera 1998, Benson 2000, Haggerty 2000, Berkowitz 2006, Trinchero 2008]).

A major difficulty using PTM relies in the estimation of PDFs. For computational reasons, a generic mass can be discretized into a limited number of particles (n), usually ranging from $n = 10^4$ to $n = 10^7$ (e.g. [Riva 2008]). An estimation based of a rough estimation of the histogram or using some naive estimator determines discontinuous density functions, based an arbitrarily selected bandwidths ([Silverman 1986]). Classical approaches lead to unrealistic PDFs peaks and discontinuities that impede the calculation of PDFs derivative.

Kernel density estimators (KDE) offer a valid option to obtain smooth density functions from a finite distribution of samples ([Silverman 1986, Hardle 1990]). While several KDE methods exist, they can be divided into two main groups, depending on the criteria chosen to find the bandwidth size (the key parameter controlling the degree of smoothness of the estimated densities). The first class of methods is called global estimators, since the bandwidth size is constant and independent from the distribution shape. The second class corresponds to adaptive estimators, where the bandwidth size is still scaled and depend on the distribution shape. According to [Silverman 1986], the selection of one of the two method depends on the specific goal of the study, since each of them is more oriented to estimate a specific portion of the curves, but penalizing other parts. Adaptive methods, for instance, are suggested to fit tails on heavy-tailed distribution, while global estimators provide a more robust option for symmetric curves.

For application involving the reconstruction of solute travel time PDFs from PTM model in heterogeneous aquifers, however, it is not clear what method should be selected, since the shape of BTCs is not known 'a priori' in most cases. In this Chapter, we develop a general method that provide accurate estimation of the full BTCs independently from their final shape. This approach combines the ability of both global and adaptive methods and takes advantage of the experimental cumulative distribution function (CDFs) of the particles travel times, which is a natural output from each model run. To check the improvement of our method, we first analyze traditional methods against some benchmark solutions. Then, we show how our method predicts PDFs in a much more accurate and cost-effective manner than the other ones.

This Chapter is structured as follows. In section 5.2 we introduce the concept of KDE methods and how particle travel time PDFs and BTCs are linked together. In Section 5.3, we define KDE statistics and the automatic and optimum selection of the global bandwidth. In the following Section 5.4, we show how adaptive methods are derived from the global methods, and introduce a standard method to compute adaptive estimates. After observing the limitation of both methods, we introduce in the last Section 5.5 the new method, and compare qualitative and quantitative advantages with respect with the other tested methods.

5.2 Particle Tracking Methods and Kernel Density Estimation

Particle tracking methods simulate transport by injecting a large number of particles into the system. Each particle carries a portion of the solute mass which moves with groundwater and reacts according to some fundamental mechanisms. A large variety of mechanisms can be chosen to efficiently simulate different transport phenomena. These mechanisms range from pure advective and random walk motions to complex reaction and mass transfer processes (e.g. [Tompson 1990, Salamon 2006b, Moroni 2007, Benson 2009]).

In all cases, the outcome of PTM is typically expressed as a distribution of particle travel times at control locations and/or positions observed at different times. Since this information is discrete in nature a reconstruction process is mandatory to finally obtain concentrations. Reconstructing concentrations from particle distributions is normally seen as the main disadvantage of PTM. This process is based on the relationship between particle distributions and concentrations, which determines that concentrations are proportional to particle densities; for instance, flux concentrations observed at a given well location, c_f, are related to the particle travel times PDF, $p(t)$, by

$$p(t) = \frac{Q}{m_0} c_f \qquad m_0 = \int_0^\infty Q c_f dt \tag{5.1}$$

where Q is the total flow at the extraction well, and m_0 is the area underneath the BTC. Unfortunately, the application of such simple relationship is plugged with difficulties. An infinite number of particles is never used to properly fulfill the validity of this equation. Kernel Density Estimation (KDE) methods have been introduced in the late '50 ([Rosenblatt 1956]) as a manner to improve the density estimation of the probability density function. For a complete and exhaustive review of these methods we refer to [Silverman 1986] or [Hardle 1990].

The need of KDE appears crucial at the moment of estimating smoothed PDFs with frequency histograms, which basically consists in counting the number of particles falling into a given time interval $B_j = [t_0 + (j-1)h, t_0 + jh]$, where h is the time interval (size of the bin) and t_0 is the origin of the histogram. The frequency at the jth bin is determined by

$$p_j \approx \widehat{p}_j \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} I\{t_i \in B_j\}, \tag{5.2}$$

where t_i is the particle travel time data $(t_1, ..., t_n)$, and $I\{\cdot\}$ is an indicator function defined as one if t_i is inside the bin B_j and zero otherwise. By its definition, the use of indicator functions provides a discontinuous box model that prevents the estimation of smooth probability density functions. Moreover, the histogram depends on its origin t_0 and the selection of the bandwidth h. Kernel density estimators is a natural extension of the histogram estimator in which the fixed box model is replaced by moving weighting function K,

Table 5.1: Possible type of kernel functions (from [Silverman 1986])

Type of kernel	$K(\tau)$
Biweight	$\begin{cases} \frac{3(1-(\frac{1}{t}\tau^2))}{4\sqrt{5}} & \text{for} \lvert t\rvert < 1 \\ 0 & \text{otherwise} \end{cases}$
Epanechnikov	$\begin{cases} \frac{15(1-t^2)^2}{16} & \text{for} \lvert t\rvert < \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp(\frac{-\tau^2}{2})$
Rectangular	$\begin{cases} \frac{1}{2} & \text{for} \lvert t\rvert < 1 \\ 0 & \text{otherwise} \end{cases}$
Triangular	$\begin{cases} 1-\lvert t\rvert & \text{for} \lvert t\rvert < 1 \\ 0 & \text{otherwise} \end{cases}$

$$\widehat{p}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{t-t_i}{h}\right), \tag{5.3}$$

where h is the deterministic histogram lag and K is a weighting function that should be chosen from symmetric functions of unitary area, i.e.

$$\int K(\tau)\mathrm{d}\tau = 1. \tag{5.4}$$

Typical kernel functions are reported in Table 5.2. While Gaussian kernels are usually preferred for computational advantages (and will be used in all the applications shown in this Chapter), the choice of a specific type of kernel does not influence substantially the final results. The 'kernel window width' (or simply, 'kernel bandwidth') h is the smoothing parameter that depends on the data distribution and varies from datum to datum. We discuss the selection of h in the following section, being the focus of this Chapter.

Based on this, the density estimation can be seen as an attempt to reconstruct an histogram where every point is the center of the sampling lag interval h. The area of influence of the ith datum is restricted to a fixed area, determined by the arbitrary choice of h. If h is too large, final densities can display discontinuities instead of more continuous shapes. If h is too small, there cannot be sufficient data to find the proper density and the outcome will fluctuate. Therefore, the outcome density is not built accounting for the relative importance of the single datum, but it is based upon an 'a priori' selection of the bandwidth size.

5.3 Automatic global selection of bandwidth

The parameter h controls the degree of smoothing and relates the area of influence around a datum. When h is small, less data can effectively influence the estimate at a given time location, leading to fluctuations. When h is large, many particles can contribute to that estimate. To select an optimal h value, we first need to evaluate the goodness of the estimator given by ((5.3)) with an objective criterion. A common criterion is the Mean Integrated Squared Error (MISE), formally written as

$$MISE(h) = E \int (p(t) - \widehat{p}(t))^2 \, dt, \tag{5.5}$$

and expressed in terms of the bias and variance of the estimator as ([Parzen 1962])

$$MISE(h) = \int \text{bias}^2(\widehat{p}(t))dt + \int \text{var}(\widehat{p})(t)dt, \tag{5.6}$$

Assuming p to be twice differentiable, an approximated expression of the bias is

$$\text{bias}(\widehat{p}(t)) = \frac{h^2}{2} p''(t)\mu_2(K) + \mathcal{O}(h^2), \tag{5.7}$$

where $\mu_2(K)$ is the second moment of K and $p''(t)$ is the second derivative of the estimated function. Notice that the bias error is quadratic in h. A similar approximation can be written for the variance

$$\text{var}(\widehat{p}(t)) = \frac{1}{nh} \|K\|_2^2 p(t) + \mathcal{O}\left((nh)^{-1}\right), \tag{5.8}$$

here the symbol $\|\cdot\|_2^2$ is the L_2-norm operator, i.e., $\|p(t)\|^2 = \int p(t)^2 \mathrm{d}t$. Notice that the variance error is inversely proportional to the product of the number of particle and the kernel bandwidth.

An optimal global value of h can be obtained by minimizing (5.6). Imposing that the derivative of MISE with respect to h is zero in the limit when $h \to 0$ and $(nh) \to \infty$ and recalling that $\int p(t)\mathrm{d}t = 1$, we obtain ([Parzen 1962]) the global optimum kernel bandwidth $\widehat{h}_G$ reading

$$\widehat{h}_G = \left(\frac{\|K\|_2^2}{n\|p''(t)\|_2^2 \mu_2(K)} \right)^{1/5} \tag{5.9}$$

Note that from (5.9), $\widehat{h}$ is inversely proportional to the number of particles. That means that the larger the number of particle, the smaller the bandwidth size is. It can be observed that $\widehat{h}$ depends on the second derivative of the density function $(p''(t))$, i.e. on the shape of the density function in itself, which is unknown. To deal with this issue, several methods have been proposed in the literature to estimate $p''(t)$. A common method is the so-called 'plug-in' methods, which consists of finding a direct estimation of $p''(t)$ by first using a reference distribution and iteratively applying successive approximation to a convergence criteria. Details of

the method can be found in e.g. [Park 1990, Hardle 1990]. Plug-in methods have been applied for instance by [Fernàndez-Garcia 2011] to estimate reactions rates from concentration gradients from PTM simulations. This method is used in the applications shown in this chapter.

Replacing (5.9) into (5.3), the global bandwidth method (GB) is obtained such that

$$\widehat{p}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\widehat{h}_G} K\left(\frac{t-t_i}{\widehat{h}_G}\right), \tag{5.10}$$

In order to evaluate the performance of GB to reproduce heavily-tailed BTCs from particle distribution we compare analytical solutions of the multi-rate mass transfer model embedded in STAMMT-L ([Haggerty 2002]) with those reconstructed from randomly distributed particles, of different population sizes, generated using a known particle tracking code, RW3D ([Fernandez-Garcia 2005, Salamon 2006a]). STAMMT-L simulates one dimensional uniform-flow transport affected by advective, dispersive and rate-limited mass-transfer mechanisms. We chose a power-law distribution of mass-transfer coefficients so as to generate PDFs with different peakness and tailing. The shape of these BTCs mainly depends on two dimensionless numbers: the Peclet number (Pe), which relates advective and dispersive characteristic times (no molecular diffusion is accounted for), and the Damköhler number (Da), which relates advective and mass-transfer characteristic times. Both numbers can be written as

$$Pe = \frac{L}{a} \qquad Da = \frac{(1+\beta)L\alpha_h}{v} \tag{5.11}$$

where L is the domain length, a is the local dispersivity, β is the capacity coefficient, α_h in the harmonic mean of mass-transfer rates and v in the advective velocity. For a power-law function, α_h depends on the cut-off rates α_{min} and α_{max} and the power-law slope k. For most combination of parameters, α_h is computed numerically (e.g. [Haggerty 2000]).

We chose to fix $Pe = 100$ and to vary Da. The complete set of resulting parameters can be found in Table 5.3 and at the title of each plots from fig. 5.1 to 5.6. When Da is low, such as for $k = 1, \beta = 1$, advection dominates over mass-transfer mechanisms, and BTCs have very peaked shapes with long tails. For increasing Da, the shapes tend to be more symmetric, as mass-transfer increases its relative importance over advection. Analytical solutions are reported from fig. 5.1 to fig. 5.6 with dotted line.

The RW3D particle tracking code move particles by a combination of a drift term that accounts for velocity and a Brownian motion that is responsible for dispersion. In addition to this, mass transfer processes are simulated by tracking the state of a particle that can change with time. The state of a particle is an attribute that defines the domain at which the particle is present at a given time within the overlapping multiple porosity media. The change from one state to another is determined using transition probabilities.

The results of the application of GB are shown in fig. 5.1 for two different distribution sizes. Solid thick (blue) line indicate the results for $n = 10^4$, while solid thinner (red) line for the case with $n = 10^5$. The analytical solution is plotted in black dotted line. We can appreciate that the estimation by GB satisfactorily matches the early arrival and peak behavior of the reference PDFs. However, it performs poorly at late times (i.e. after the peak), where the numerical solution oscillates and eventually tends to a fictitious asymptotic plateau. At the case $k = 3$, $\beta = 5$ (high Da), the solution does not fluctuate as in the other cases, reflecting that GB works better to reproduce more symmetric distributions. The starting time for oscillation is linked due to amount of particles used to obtain the PDFs: when $n = 10^5$, the oscillation looks like more stable over a wider time length than in the case of $n = 10^4$.

To obtain a quantitative measure of the fitting quality for this method, we measured the relative local error between the estimated and actual density functions at the i-th datum, ε, reading i.e.

$$\varepsilon_i = \left| \frac{\widehat{p}(t_i) - p(t_i)}{p(t_i)} \right| \tag{5.12}$$

In fig. 5.3, the black lines with circles refer to ε for GB (the other curves indicate the errors for the other methods, and they will be describe in the following sections). Note that the plot scales are logarithmic to emphasise that ε can span over several orders of magnitude. For GB, ε is minimum around the concentration peaks and it is generally lower at shorter times, where the number of particle used for the estimation is large. As the time elapses and the mass used to estimate the particle density diminishes, ε becomes dramatically higher (above $\varepsilon = 10^1$); this is especially true for heavy-tailed PDFs (low Da).

The use of GB is therefore robust and reliable to estimate early time and peak behavior of the PDFs outcomes, but it generates spurious effects on tail that can be alleviate only by increasing the number of particles (at expense of additional computational time). For limited number of particles, tailing is only partially reproduced, especially for heavy-tailed distributions. In the following sections, alternative adaptive methods are tested to cope with this issue.

5.4 Globally adaptive bandwidth method

Adaptive methods are considered an efficient manner to deal with long-tailed distributions ([Silverman 1986]). Long-tailed distributions implicate that the masses are smudged out over a wider range than one in the main part of the distribution. Intuitively, as the number of particles used to estimate their density diminishes, the smoothing parameter (h) should increase, to allow a smoother estimation to in the regions where data are lacking. To partially overcome the problem derived from GB on heavy-tailed distribution, one common procedure consists in varying h so that it increases where data are lacking. This is formally written as

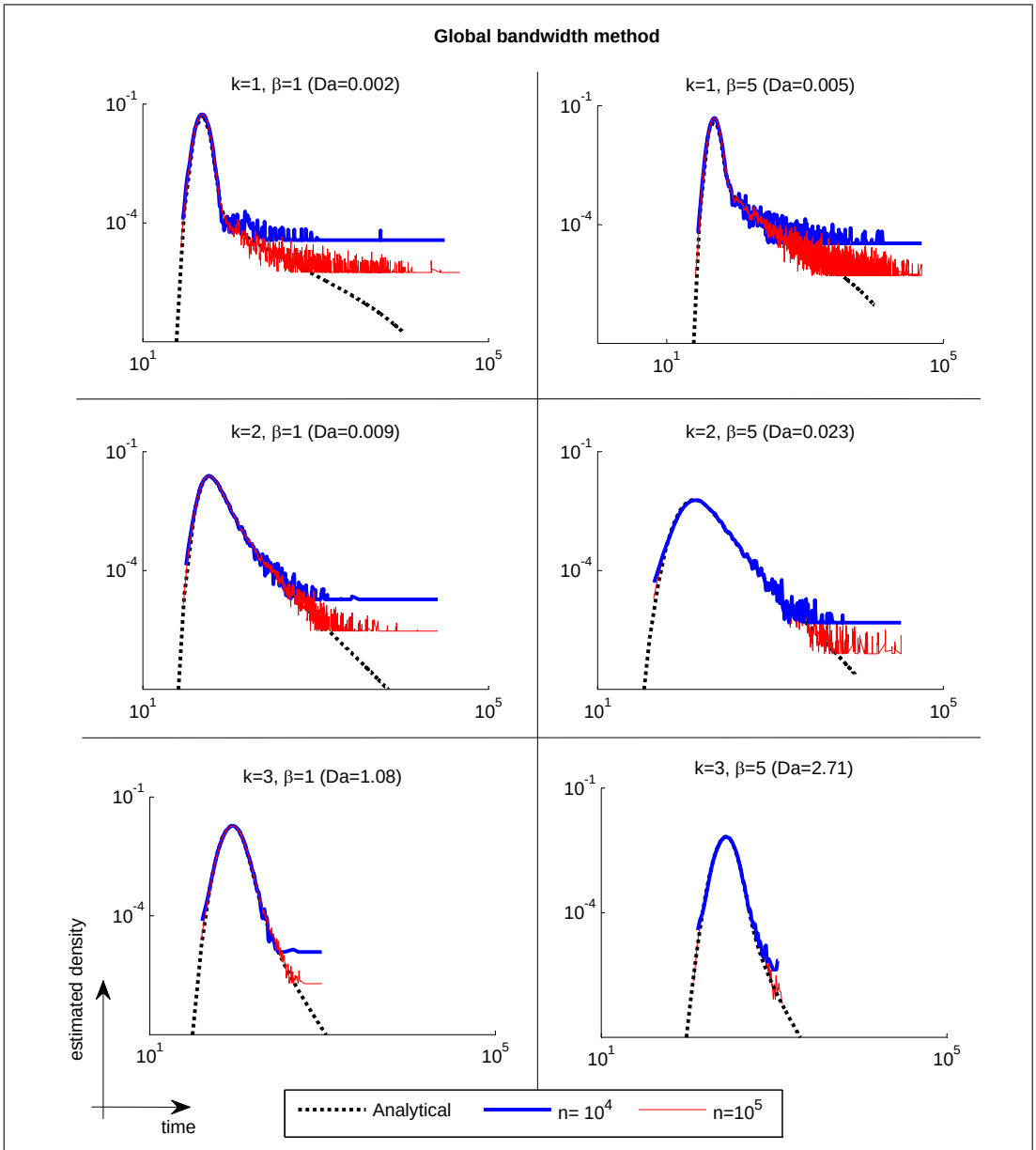

Figure 5.1: Comparison between reference PDFs from analytical solutions and estimated particle density functions using the global bandwidth method. Simulations are run using different number of particle datasets. Parameters k indicate the slope of the PDFs and β indicating the position of center of mass; Da is the Damkohler number.

Table 5.2: Parameters for analytical solutions. v= pore-scale velocity; α=longitudinal dispersivity; ϕ=total porosity; L=domain length; k=distribution power; β =capacity coefficient; *=only for random walk simulation

Variable	value [dimensions]
v	1 [LT^{-1}]
a	0.05 [L]
L	5 [L]
Pe	100
k	1; 2; 3 [-]
β	1; 5 [-]
α_{min}	10^{-4} [T^{-}1]
α_{max}	1 [T^{-}1]
Da (k=1,$\beta = 1$)	0.002
Da (k=1,$\beta = 5$)	0.005
Da (k=2,$\beta = 1$)	0.009
Da (k=2,$\beta = 5$)	0.023
Da (k=3,$\beta = 1$)	1.08
Da (k=3,$\beta = 5$)	2.71

$$\widehat{p}(t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}K\left(\frac{t-t_i}{h_i}\right), \tag{5.13}$$

A common approach is suggested by [Silverman 1986] and it is tested here against the reference distributions shown previously. This approach is based on a few iterative steps:

1. finding the initial estimation of the PDFs using GB (i.e. using (5.13));
2. calculating a new weighting function h_i, such that

$$h_i = \widehat{h}_G\lambda_i \qquad \lambda_i = \left(\frac{\widehat{p}_i(t)}{F_G}\right)^{-\alpha} \tag{5.14}$$

where $\widehat{p}_i(t)$ is the first estimate of the density function at the i-th datum, F_G is its geometric mean of $\widehat{p}_i(t)$, i.e.

$$F_G = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\ln(\widehat{p}_i(t))\right) \tag{5.15}$$

where α is a sensitivity parameter fulfilling $0 \leq \alpha \leq 1$;

Note that h_i is no longer constant to $\widehat{h}_G$ such as in GB, but stirred by the first estimated to change from datum to datum. The new estimation can be obtained

substituting (5.14) into (5.13). This approach is therefore is called 'globally adaptive bandwidth' method (GA).

We highlight that the new density estimation does not follows a second optimization process, but relies on the first estimated $\widehat{h}_G$. The selection of α determines how important the shape of the first guess is in the selection of the second estimation. There is not a well-defined automatic criterion to obtain α; only [Abramson 1982] found a theoretical response to choose α=0.5, but for the other values such a sound mathematical background is lacking.

The comparison between the reference and the simulated curves using GA is given in fig.5.2 for $\alpha = 0.5$. It can be observed at a first qualitative glance that this method seems to fit quite well the heavy-tailed reference PDFs cases (the one showing low Da). As Da increases, however, GA determine underestimated peaks and the late-time behaviors no longer matches from the reference distributions. Especially, while tails seem quite smoothed and continuous at late-time, their gradients are different from the ones showed by the reference distributions. This is issue is particularly visible for the cases with $k = 3$, which are much more symmetric than the rest. Here, the kernel estimation is erroneously overweighting the late-time tailing and sacrificing too much the early portions of the curves, while in reality the weights should be more equally distributed.

Surprisingly, the method seems to work better for lower n where Da is lower (more asymmetric curves). This fact can be explained considering that in (5.9) the errors diminish as n increases, but from the use of (5.14) (noticing the negative sign on α) the relationship is inverted.

While this analysis is qualitatively suggesting that this method may not perform satisfactorily to reproduce BTCs, this conclusion is additionally supported after observing fig. 5.3. Here we report the temporal variation of the local errors (ε) in comparison with GB. For completeness, we also add the error analysis for $\alpha = 0.25$, $\alpha = 0.5$ and $\alpha = 1$. It can be appreciate that at larger times and for low Da, ε for GA tend to lower values than for GB, meaning that GA increase the accuracy of the estimation. For the other cases (higher Da and/or earlier y times) the local errors are in general dramatically high ($\varepsilon > 10^1$), reflecting the oversmoothing effects around the peak and that these methods perform poorly for more symmetric distributions.

The use of GA poses serious concerns about the reliability of the PDFs outcomes, as it can give misleading solute transport parameters that can have a catastrophic impact on decision-making processes (e.g. in designing risk assessment or aquifer remediation activities). Moreover, the fact that the results worsen while increasing the number of particles renders unclear what best dataset size should be applied to improve the fitting exercise. To address this issue, an alternative solution is presented in the next section.

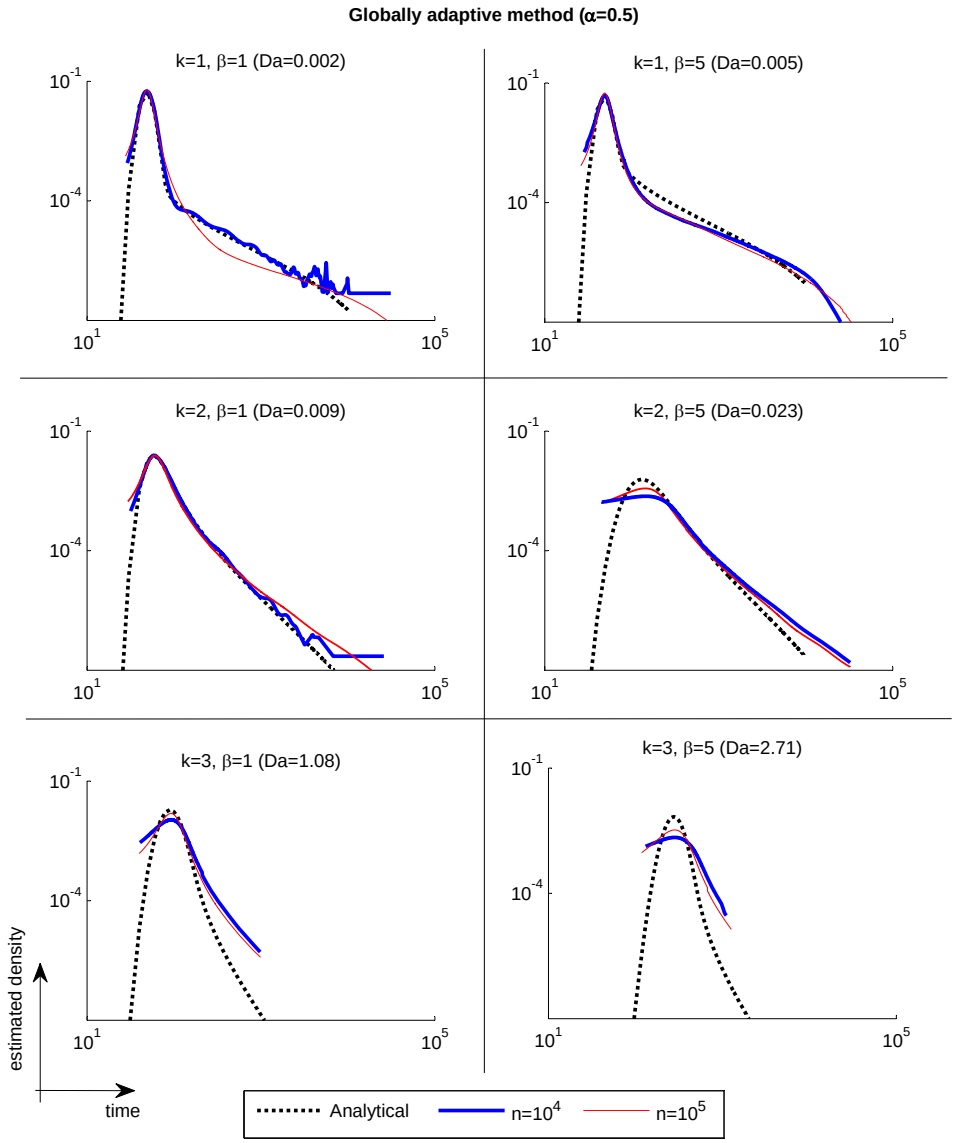

Figure 5.2: Comparison between reference PDFs from analytical solutions and estimated particle density functions using the adaptive method.

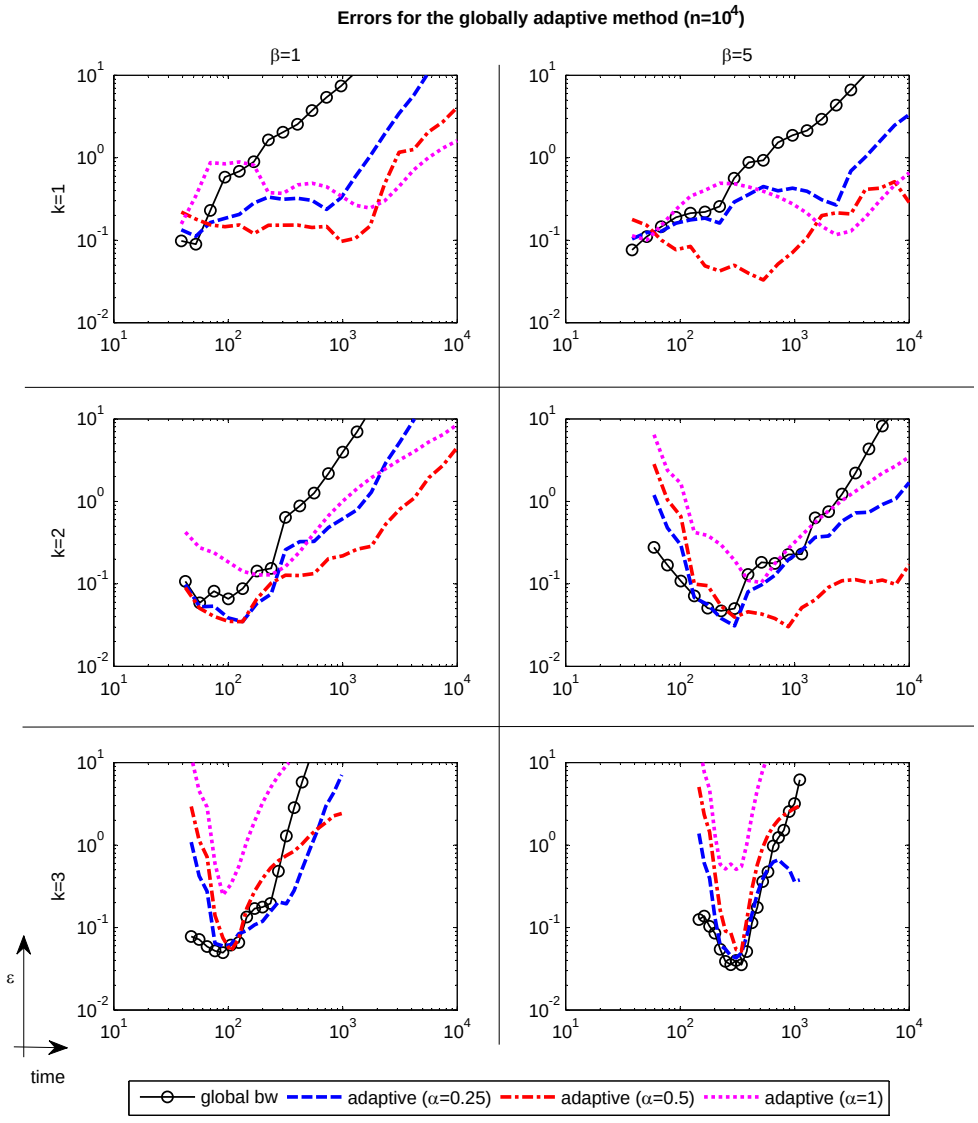

Figure 5.3: Absolute relative errors (ε) between the benchmark solutions and the estimated PDFs using global bandwidth method and adaptive methods.

5.5 An Improved Locally Adaptive Method

The analysis of the two standard kernel methods described so far suggests that the choice of the most proper estimator should be indicatively done depending on the type and the goal of the problem. When the estimation of early-time or peak behavior is relatively more important than the late-time behavior, GB should be preferred; if the late-time distribution is the key feature to analyze and very skewed PDFs are most likely to be observed, then GA should be implemented.

Unfortunately, in most hydrogeological applications the final PDFs shapes are not known 'a priori' and the selection of the kernel estimator is uncertain. During tracer tests performed in heterogeneous aquifers, for instance, the shape of BTCs measured after injecting solutes from different positions is highly variable and can be very dissimilar to ideal parametric distributions such as those simulated using MRMT models with PL distribution (e.g. [Fernàndez-Garcia 2004]). Dimensionless transport parameters controlling the BTCs shapes such Peclet, Damköhler or β numbers are partially obtained by 'a posteriori' model fitting ([Harvey 1995, Willmann 2008]) but they would be necessary for an 'a priori' selection of the type of kernel.

Therefore, we look for a common and general procedure to be universally applied independently from the final expected shape of the BTCs. We propose here a new KDE weighting function, which is derived after coupling the ability of GB and GA methods. The idea is the following. We notice from the previous analysis that (1) the weighting factor h_i should be similar to the global estimated bandwidth at early portion ($h_i \to \widehat{h}_G$ as $t \to 0$) and around the peak of PDFs, while (2) it should increase only at late-time i.e. on the tails, where the estimation is based on a more limited number of particles, similarly to the weight given by GA ($h_i \to \widehat{h}_G \lambda_i$ as $t \to \infty$). This means finding a new function that drives the behavior of h_i, but to avoid additional complexity of the bandwidth selection, we look for a solution to be applied without any need of calibration.

A suitable and appealing function displaying these characteristics and that is directly obtained from the numerical outputs is the experimental cumulative density function, $P(t)$. A quick manner to obtain $P(t)$ is by sorting the vector of arrival time distribution from the shorter (t_1) to the largest arrival time (t_n); $P(t)$ results

$$P(t_i) = \frac{i - C}{n} \tag{5.16}$$

where C gives the functional form of P for a datum (e.g. for a triangle shape, $C = 0.5$; for a box shape, $C = 0$). $P(t)$ always takes values comprised between $[0, 1]$, being lower around the early time and increasing in the regions where less mass is used to estimate the PDFs for larger times. The new weighting function to be substituted into (5.13) becomes

$$h_i = (1 - P(t_i))\widehat{h}_G + P(t_i)\widehat{h}_G \lambda_i \tag{5.17}$$

We call this method 'locally adaptive bandwidth method' (LA), because h_i is locally modified by a function that is independent from the first estimated PDFs. The results from the application of this method are shown in fig. 5.4, fig. 5.5 and fig. 5.6 respectively for $\alpha = 0.25$, $\alpha = 0.5$ and $\alpha = 1$. We note from these figure that LA sensibly improves the estimation for each combination of parameters, and works satisfactorily well for for heavy-tailed and less skewed curves. More specifically, using $\alpha = 0.5$ and $\alpha = 1$ not only the oversmoothing effects around the BTC peak disappear but tailing is also captured, matching the one of the reference analytical solutions.

In fig. 5.8, we compare the weighting functions (h_i, normalized in the plots by $\widehat{h}_G$) for GA (Eq. (5.14)) and for the locally adaptive method (Eq. (5.17)) applied to the case $k = 2$, $\beta = 1$. We note that at early times the weights for the locally adaptive method tend to $\widehat{h}_G$ for all α, while they dramatically increase in the case of the globally adaptive method (generating oversmoothing effects). At late times, both solutions match, so that the locally adaptive method provide the same accuracy at the globally adaptive methods on tails. Table 5.5 resumed all the weighting functions used in this analysis.

The question is now what α parameter should be selected. The goal is to find a general method that is independent from any calibrating parameters. From a quantitative perspective, the ε criterion (shown in fig.5.7) helps to indicate which α provide the lowest local error and more accurate local estimation for each reference cases. At shorter times, all the errors collapse to the one of GB, while at larger time ε is always lower using LA than using GB. Observing the plots, $\alpha = 1$ apparently is the best option for $k = 1$ and $k = 2$, while $\alpha = 0.5$ offers the best tested option for $k = 3$.

However, it is observing the full shape of the BTcs that we suggest $\alpha = 0.5$ as most robust parameter to be used for universal estimations. In effect: (a) using $\alpha = 0.25$ (fig.5.4), the weighting function provide very few benefits with regard to GB at late time, where estimated PDF still oscillate; (b) using $\alpha = 1$ (fig.5.6) fluctuations almost disappears and local errors are in most cases very low, but concentration gradients are not very well captured as the solutions departs from the reference distribution at large times for the most symmetric distributions (higher Da); (c) using $\alpha = 0.5$ (fig.5.5), an excellent agreement is found for high Da, while for low Da tailing are correctly estimated and fluctuation are reduced for a sufficient time to correctly capture the late-time gradients.

Another important point concerns with the efficiency of the method. LA is much more efficient than using the standard methods (GB and GA), as a first qualitatively analysis observing for instance that the use of GB and GA estimators with $n = 10^5$ perform less accurately than LA with $\alpha = 0.5$ and $n = 10^4$. To quantitatively check this behavior with more detail, we consider a global error of the estimated PDFs in terms of the root mean squared errors (RMSE) for several distributions of variable size. The $RMSE$ is calculated as

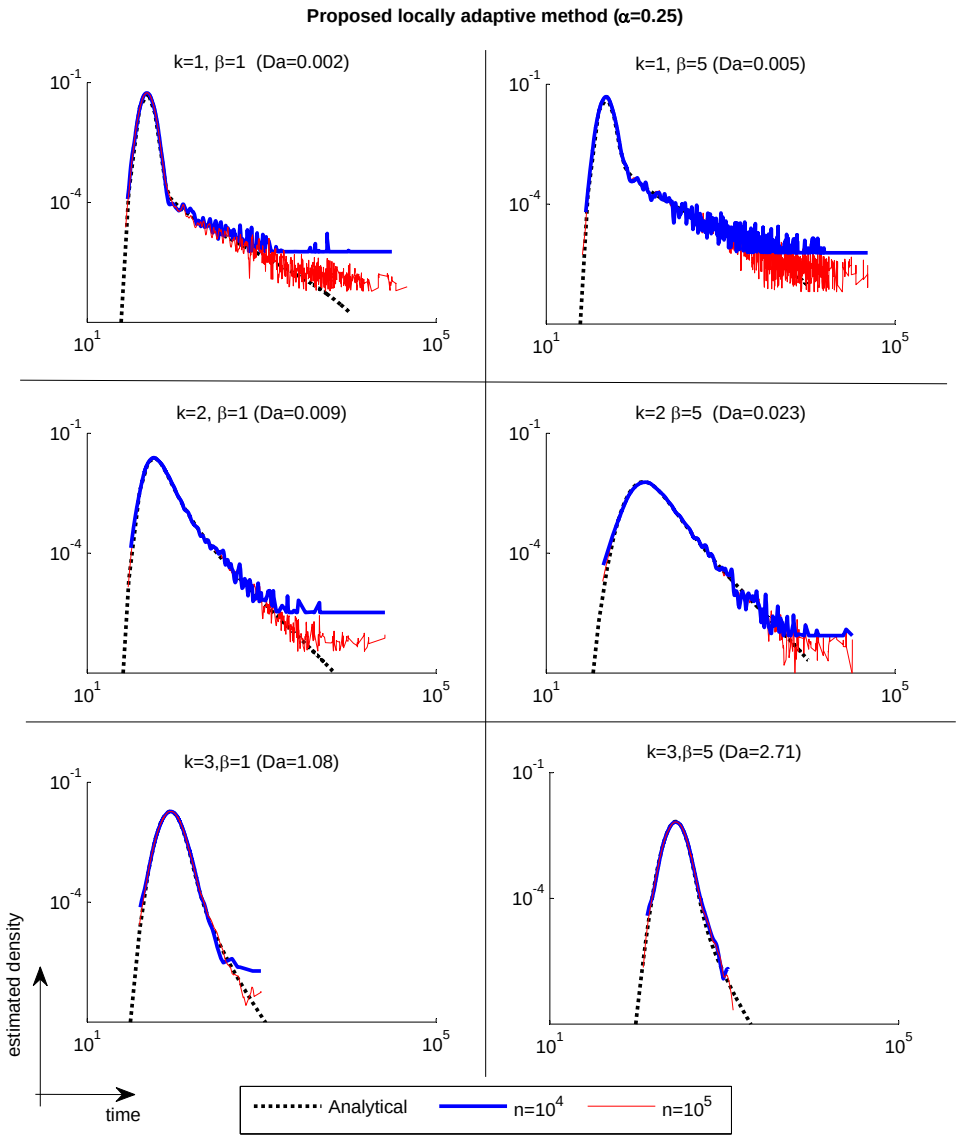

Figure 5.4: Comparison between reference PDFs from analytical solutions and estimated particle density functions using the proposed method with $\alpha = 0.25$.

Table 5.3: Weighting functions used in the three kernel density models. See the text for the definition of each parameter.

Method	h_i
Fixed bandwidth	$\widehat{h}_G$
Globally adaptive bandwidth	$\widehat{h}_G(\tilde{p}(t)/F_G)^{-\alpha}$
Locally adaptive	$\widehat{h}_G((1-P(t))+P(t)(\tilde{p}(t)/F_G)^{-\alpha})$

$$RMSE = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}\left(\varepsilon(t_i)\right)^2\right)} \tag{5.18}$$

In fig. 5.9 we show the results for GB, GA and LA for $\alpha = 0.5$ and $\alpha = 1$ and using different distribution sizes. It is found that LA with $\alpha = 0.5$ with $n = 10^5$ always gives the smallest overall error, independently from the final shape of the BTCs. It can be observed that for the case with low Da, GB with $n = 10^5$ (which takes ≈ 780 seconds on a computer equipped with an Intel® Xeon ® with 2.80 GHz and 6 Gb of RAM memory) perform generally worse than LA with $n = 10^4$ (which takes ≈ 20 seconds per simulation on the same machine) and in some cases even for $n = 10^3$ (which takes less than 3 seconds). For high Da, GB and LA methods perform similarly for larger datasets. It is noticeable again that the overall quality of the estimation using GA worsen as n, for the reasons we discussed above.

From these results, we can thus suggest the use of LA with $\alpha = 0.5$ as a general method to accurately reproduce the shape of any PDFs, and especially accurately capturing early time, peak as well as tailing gradients. While to improve the quality of the assessment $n = 10^5$ provide the best fit, in the optic of performing a large number of simulations (e.g. within a Monte Carlo framework) and considering the long computational times required to evaluate large datasets, a good trade-off between computational time and quality of the results can be found using $\alpha = 0.5$ and reducing the number of particles to $n = 10^4$.

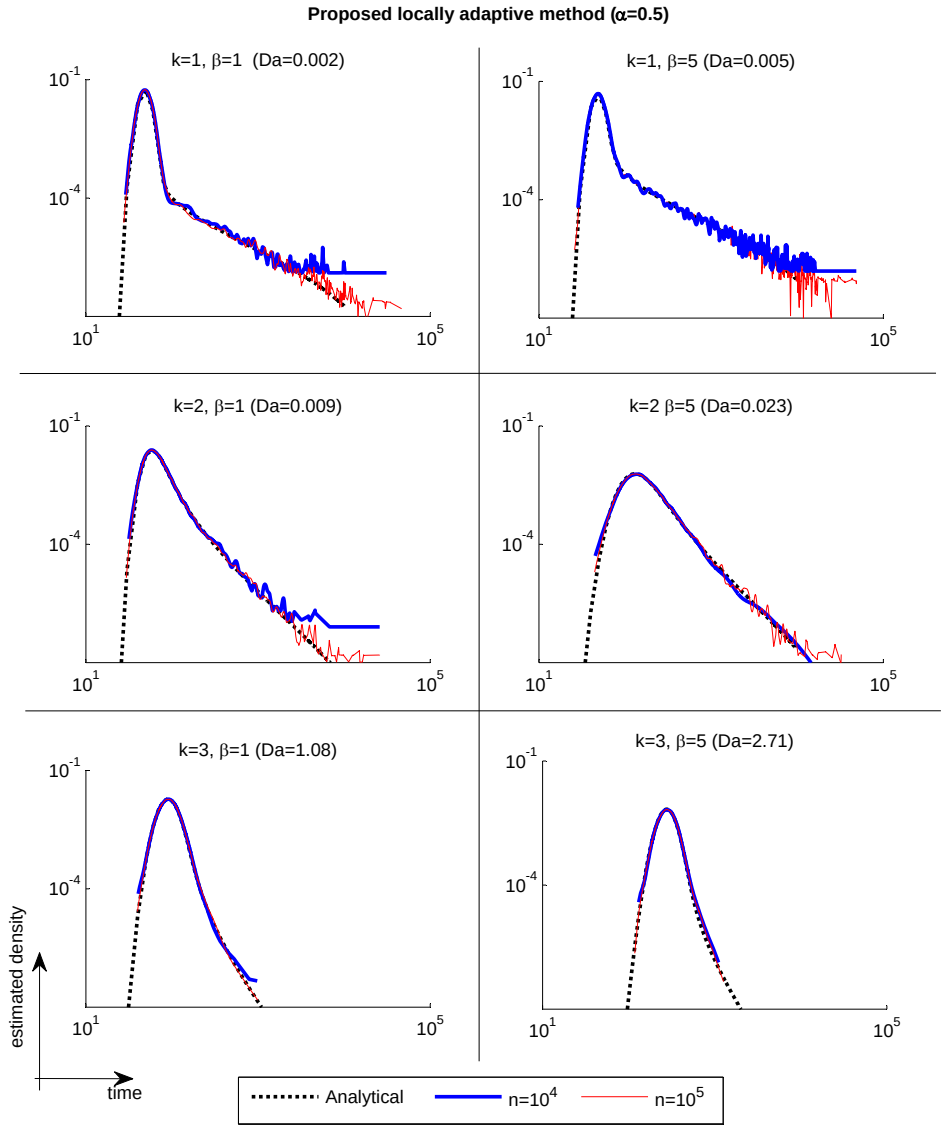

Figure 5.5: Comparison between reference PDFs from analytical solutions and estimated particle density functions using the proposed method with $\alpha = 0.5$.

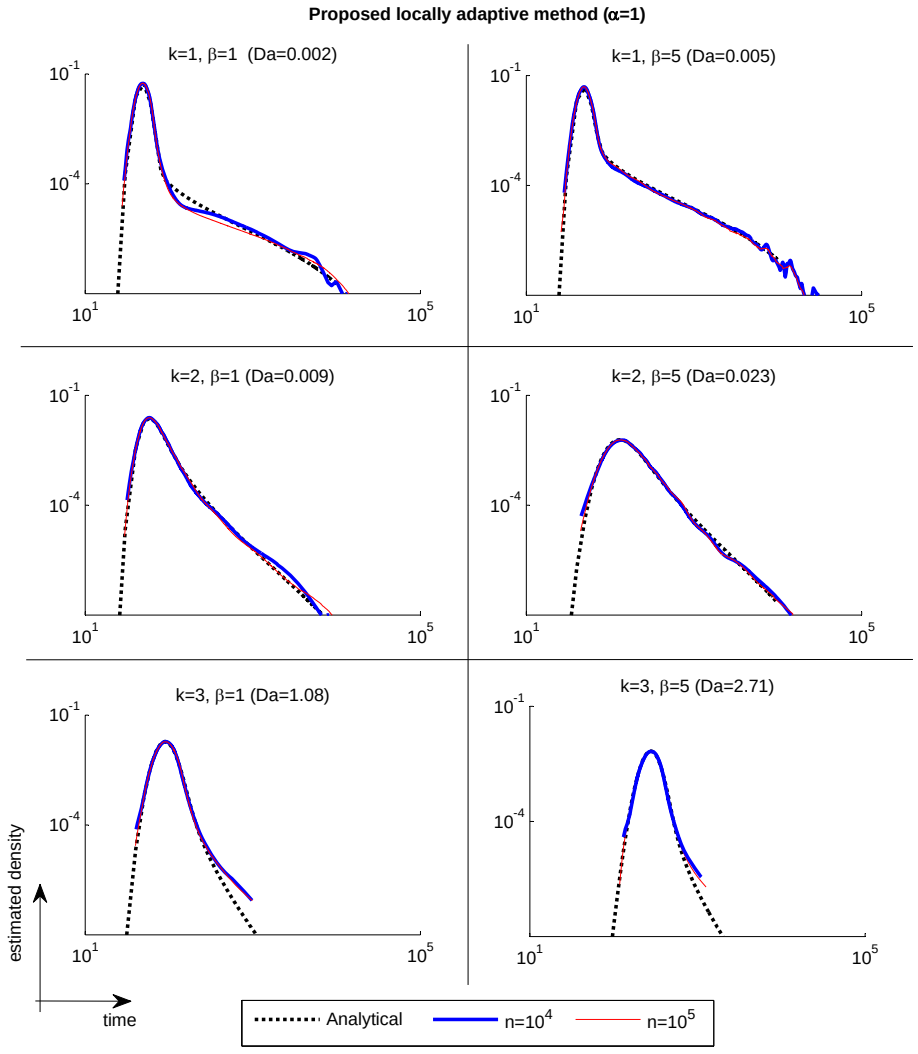

Figure 5.6: Comparison between reference PDFs from analytical solutions and estimated particle density functions using the proposed method with $\alpha = 1$.

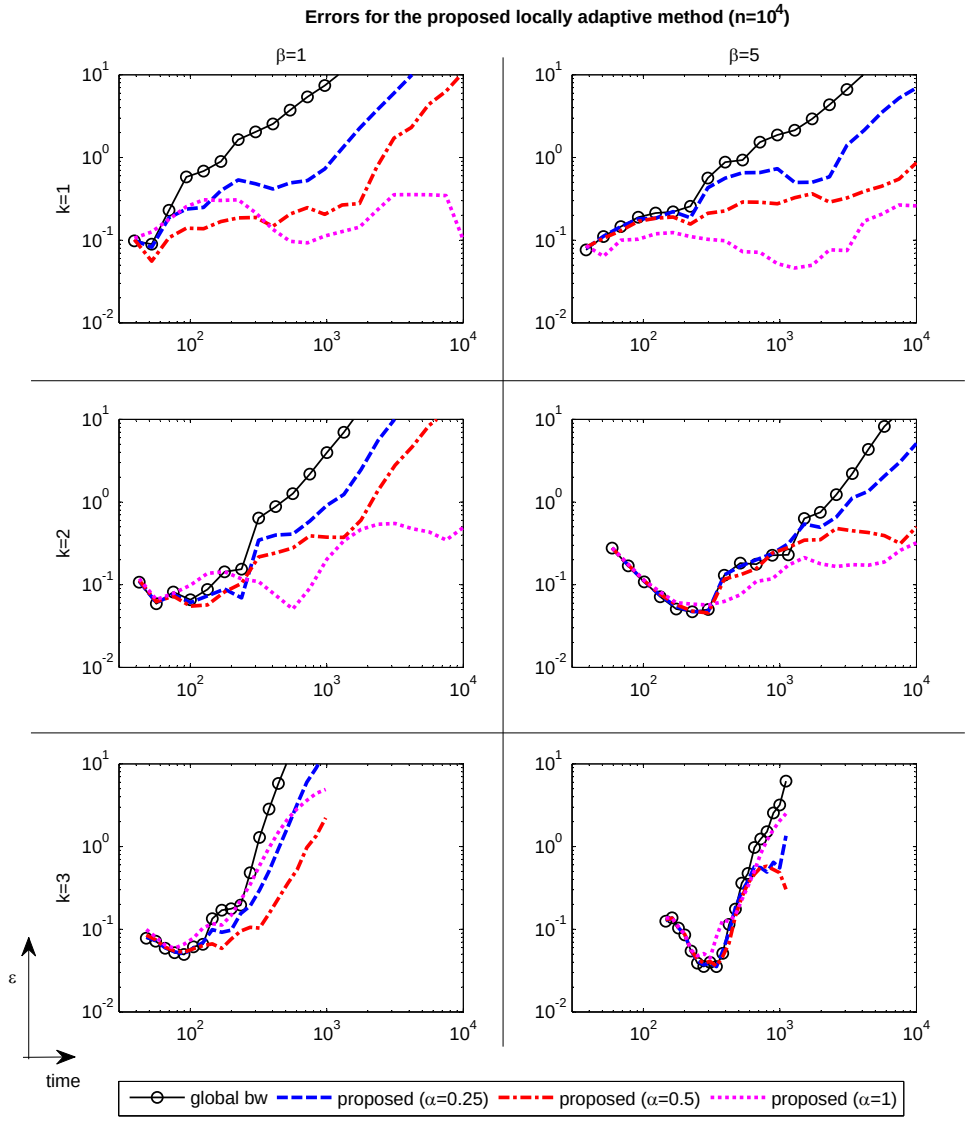

Figure 5.7: Absolute relative errors (ε) between the benchmark solutions and the estimated PDFs using global bandwidth method and proposed methods.

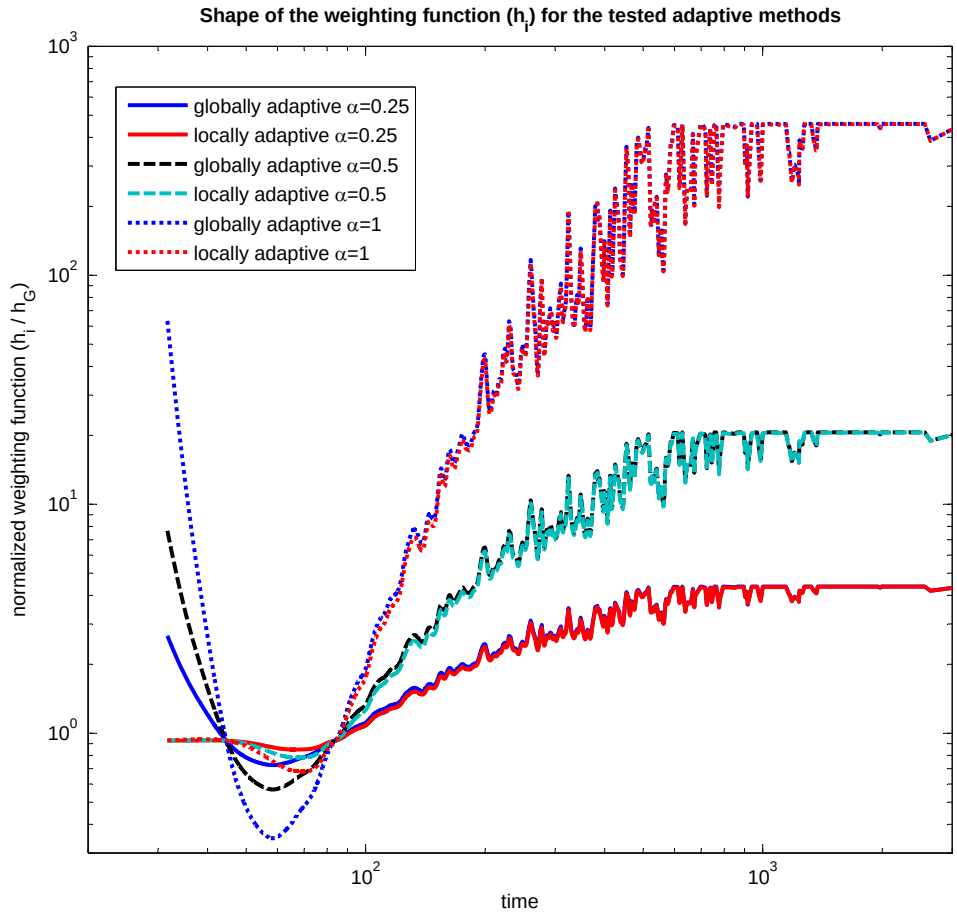

Figure 5.8: Shape of the weighting function between global bandwidth method (blue dotted line constant to the optimum global bandwidth size h, for the case $k = 2$,$\beta = 1$ using $n = 10^4$. Notice that the log-scale emphasise how the weight are close to h for the proposed method at early time and around the peak, while they tends to the values for the adaptive method as the time elapses.

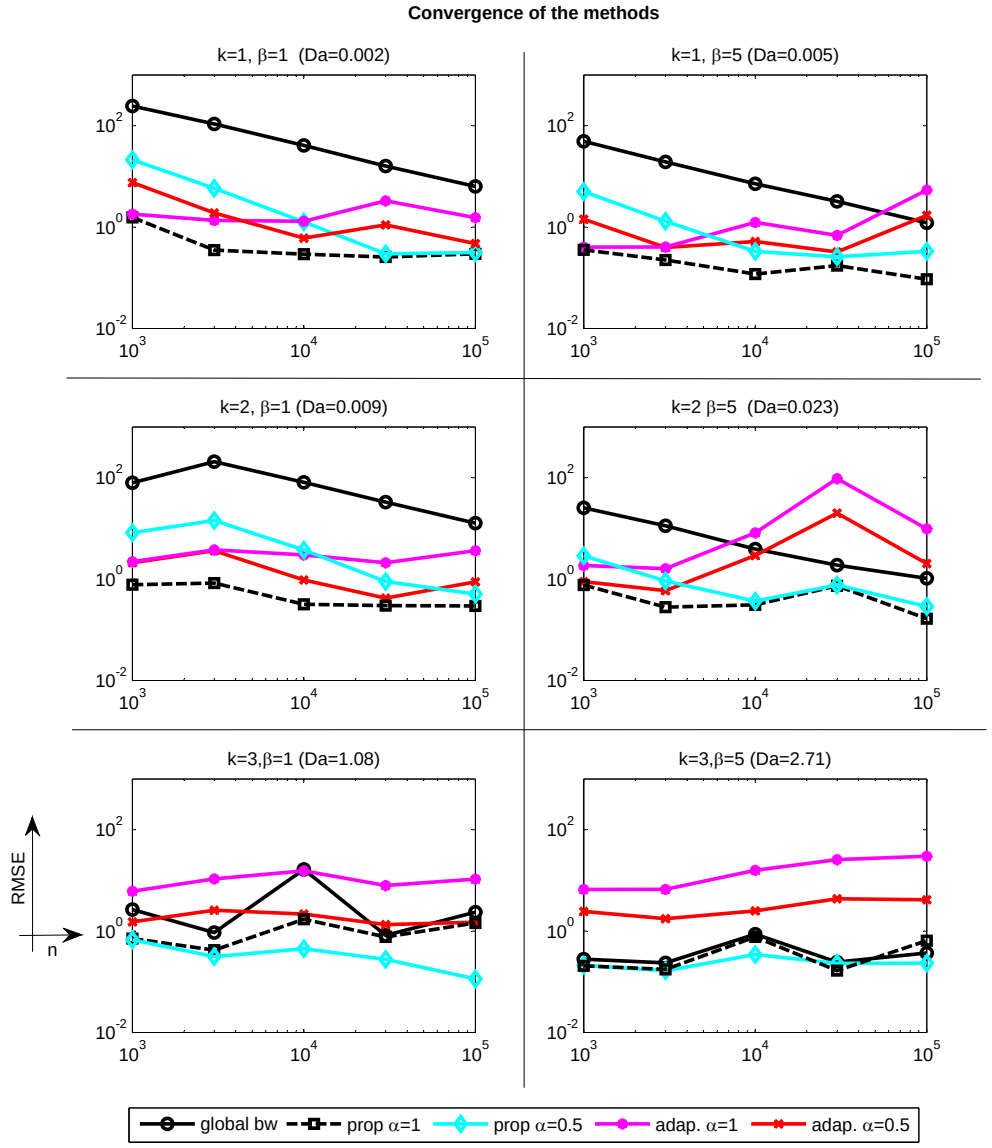

Figure 5.9: Root mean square error ($RMSE$) of the different methods depending on the number of particles used for the estimation of the PDFs.

Chapter 6

Development of anomalous tailing under radial convergent transport in three-dimensional heterogeneous sandy aquifers

6.1 Introduction

This Chapter is devoted to analyze the formation the heavy-tailed breakthrough curves (BTCs) with late-time distributions similar to those obtained during tracer tests in typical experimental conditions and stationary hydraulic conductivity fields. The goal is to understand what modeling parameters, similar to physical counterparts, are at the origin of the development of heavy-tailed BTCs.

Tailing on BTCs is a macroscopic evidence of heterogeneity and mass-transfer processes at the pore scale (e.g. [Levy 2003, Sanchez-Vila 2004, Salamon 2007, Riva 2008]). Being able to accurate reproducing and modeling it becomes important to understand the dynamics generating anomalous transport and give a physical meaning to model parameters (these methods are discussed in the next Chapter).

Several authors (e.g. [Zinn 2003, Willmann 2008, Flach 2012] used 2D numerical simulations to reproduce anomalous transport under the assumptions of uniform flow conditions and finite correlated log-normal hydraulic conductivity structures (i.e fields with log-normal distributed hydraulic conductivity). They conclude that these models cannot reproduce connectivity patterns, which are the most influent hydrodynamic parameter controlling BTC late-time behavior. In these works, connectivity is defined as the ratio between spatially-averaged and effective parameters defining flow and transport ([Knudby 2006]). We refer to the recent work by [Renard 2011] for an extensive review on connectivity concepts. These findings are in disagreement with some theoretical and experimental works. proved the existence of tailing on BTCs for short travel distances in multigaussian fields with finite integer scales of log-transformed hydraulic conductivity.

The experimental evidences reported by [Fernàndez-Garcia 2004] also indicate that transport under finite correlated structure can lead to anomalous transport.

[0]This chapter is based on D. Pedretti et. al (submitted to Water Resources Research), *On the formation of heavily-tailed breakthrough curves during convergent flow tracer tests in three-dimensional heterogeneous sandy aquifers*

The latter reproduced in laboratory and controlled conditions similar hydraulic conductivity fields and simulate convergent, divergent and uniform flow tracer tests in an intermediate-scale three-dimensional aquifers. Four of the experimental BTCs obtained during the convergent flow tests (CFTTs) using a deep-penetration source injection ([Fernàndez-Garcia 2003]) are shown in fig.6.1 and clearly display non-uniform transport behavior. We can observe from fig.6.1 that: (a) the BTC obtained after injecting at E1 (situated two integer scales away from the pumping well, displayed a heavy-tailed distribution after the peak, well approximated by a power-law function with $m = 1$; (b) at the same radial distance from the pumping well. the BTCs obtained injecting at E2 and E3 were quite symmetric, being E3 much more similar to a Gaussian bell than E2 (on which a PL fit could lead to low m values); (c) E4 is much more irregular and only approximately similar to a power-law with $m = 1$.

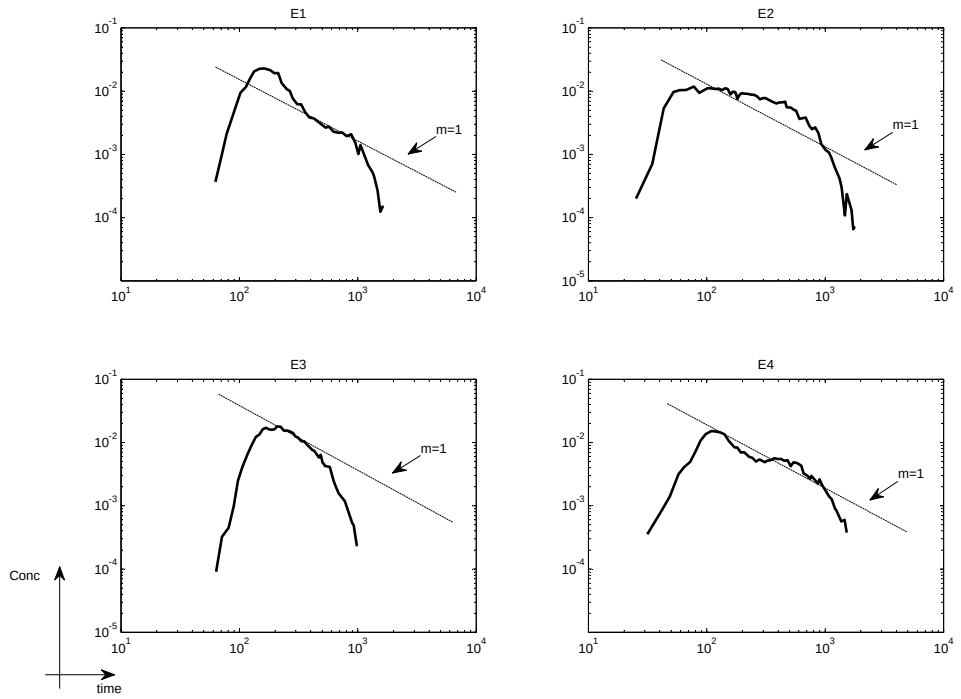

Figure 6.1: Four of the experimental BTCs obtained by [Fernàndez-Garcia 2004] during CFTT in heterogeneous 3D tank. Notice that the part of the BTCs showing heavy-tailed distribution scale following $c \approx t^{-1}$ (at E1 and at E4), while at E2 and E3 the shape is more symmetric. The injecting location varies within the 3D tank, keeping the radial distance from the well ($r_\lambda = \approx 2$)

The two-dimensional uniform flow simulations are hardly comparable with radial converging flow tracer tests setting. First, there is a fundamental difference in the two flow geometries. For instance, tailing naturally arise on BTCs under radial flow in homogeneous fields ([Gelhar 1971, Moench 1989, Welty 1994]), while it does not

under uniform flow Second, because two-dimensional (2D) models are not suitable to reproduce typical tracer tests settings, where the distance between the injection and the controlling section is on the range of the heterogeneity representative scales such as the integer scale (I) in a multigaussian K field. In this case, whenever possible, 3D models are preferable (e.g. [Dagan 1989]).

The dimensionality is also a key factor when studying connectivity. Flow connectivity is largely enhanced by 3D configurations than by 2D counterparts,according to the detailed analysis by [Fiori 2012]. This is especially true if unconditional Sequential Gaussian Simulations (SGS) are used to simulate stochastic hydraulic conductivity fields ([Fernàndez-Garcia 2010]). In fact, [Willmann 2008] had to heavily condition their simulations to generate highly conductive non-stationary structures leading to heavy-tailed BTCs similar to those observed in the field. It should be reminded that the numerical 3D flow and transport simulations by [Fogg 1986] showed the most influencing factor controlling flow and transport is the lenses connectivity rather than the relative K values of the lenses themselves. That means that not only high conductivity zone have a large impact on flow dynamics, but also poorly permeable zones if well-connected. The reliability of 2D SGS to adequately reproduce transport connectivity patterns have been extensively debated in the past ([Sanchez-Vila 1996, Gomez-Hernandez 1998]). However, 3D models have being much less studied than 2D counterparts, probably because of the still high-demanding computational cost of 3D simulations.

Several question arise: is BTCs scaling found by the experimental evidences reported in [Fernàndez-Garcia 2004] (power-law behavior with slope $m = 1$) due to some physical reasons, or is it just a random output? More in general, are there any physical reasons controlling BTCs scaling at large time after injection? What is the impact of the model dimensionality on the formation of tailing?

In this Chapter, we aim to give an answer to these questions. It will be show that 3D numerical realizations under radial convergent flow conditions help to reproducing solute transport similar to the experimental observations. We simulate synthetic heterogeneous fields drawn from a multilognromal K distribution, which is characterized by different combination of finite-scale correlation and variances, and use a special technique to measure the peak concentration time of each layer composing the three-dimensional simulations as an indicator of variability of point-to-point connectivity ([Trinchero 2008]) at various depth in the aquifer. The use of concentration peak times to estimate statistics of travel time was already adopted for instance by [Bellin 2004] and reference therein. The aim to provide new insights to explain how BTCs behave in typical CFTT field settings, and to find the key link between needed to interpret the results obtained from field tracer tests in real applications.

The Chapter is structured as follows. In Section 2, we describe the numerical approach we used to reproduce CFTTs in synthetic heterogeneous aquifers. Numerical output are used to estimate BTCs, both as 'depth-integrated' and as 'multilevel' distribution of concentrations at the extracting well. In Section 3, we show the full results in which we highlight how BTCs tailing develops in different heterogeneous

fields, and assuming injections to take place at variable positions within the same heterogeneous fields. The Chapter ends with a final discussion on the similarities between 3D simulations and their corresponding 2D counterparts, in Section 4, and the conclusions.

6.2 Numerical simulation of CFTT in different geological settings

We consider a typical CFTT scenario, in which a passive injection well (or piezometer) is located at a distance (r) from an extraction well. The latter is activated to withdraw groundwater at a constant discharge rate (Q). Once steady-state conditions are established at the injection well, a known mass (M) of a nonreactive compound is introduced into the aquifer through the injection well. The injected mass is potentially fully recoverable in ideal conditions; in real tests, the combination of mass losses plus the finite recording time leads to partial recovery of the injected mass.

In heterogeneous aquifers, under forced-gradient conditions, the horizontal flow velocity can vary along the vertical column by several order of magnitude. Therefore, the amount of tracer mass injected into the aquifer at different depths is proportional to the local horizontal velocity found at the different horizon along the vertical injecting column. This combination of injection and extraction proportional to local flow forces one to pose the problem in terms of flux-averaged concentrations rather than resident concentrations ([Kreft 1978, Parker 1984]).

BTCs can be measured both at the extraction well and, if existing, at some controlling section between the injection and the extraction location. We assume, without lack of generalization, that only the extraction well is used for measuring concentrations. At the extraction well, concentrations are usually observed as 'depth-integrated' BTCs over the entire screened section of the well, taking samples of the water once withdrawn and pushed to the surface. Alternatively, measurements can be made at different intervals along the screened section of the well using multilevel samplers (MLS), giving raise to 'multilevel' BTCs. These methods have been widely used in the past to provide useful indication about stratification of transport properties of the aquifers and to indicate the existence of preferential flow paths (e.g. [Ptak 1996, Ptak 2004, Bianchi 2011]).

With these concepts in mind, we adopt a numerical approach consisting of (a) generation of a number of realization of Y-fields from a predefined geostatistical model; (b) solution of the groundwater flow problem in each of the Y-fields by setting appropriate boundary conditions to impose forced-gradient behavior to a well; (c) solution of the transport problem (d) calculation of 'depth-integrated' and 'multilevel' BTCs at the well.

Details of the different steps are provided in the subsequent subsections.

6.2.1 Simulation of heterogeneous hydraulic conductivity fields and convergent flow solution

We start by generating three stochastic realizations of Y-fields using a sequential Gaussian simulator 'sgsim' included in the geostatistical modeling software SGEMS ([Remy 2009]). The support grid consists of a regular 3D lattice composed of N_L=100 planar layers, each of which is formed by $N_C = 251$ horizontal squared cells per side. We use i, j to identify a cell respectively in the x, y direction ($i, j = 1, ..., N_C$) and k to identify a specific layer ($k = 1, ..., N_L$). We assume Y to be isotropic in each cell, with the exception of the central vertical column, which represent the wells and inherits large vertical conductivities. Each cell has unit size, in all three directions.

By construction, all Y-fields have zero mean log-transformed hydraulic conductivity ($\overline{Y} = 0$). Three exponential variograms with unit variance ($\sigma_Y^2 = 1$) and variable integral scales, I, are used. Thus, indicating the three directional integer scale as I_x,I_y and I_z, we build:

- Field (A): I_x=I_y=I_z=4;
- Field (B): I_x=I_y=40, I_z=4;
- Field (C): I_x=I_y=I_z=40.

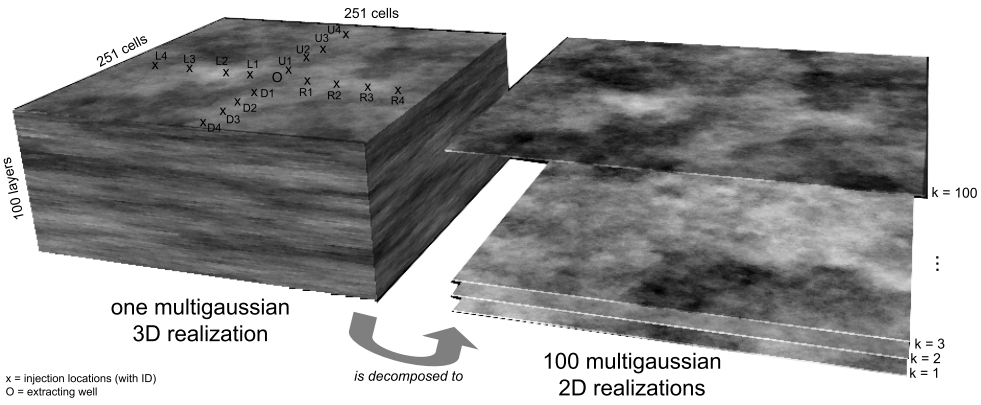

Figure 6.2: On the left, the 3D anisotropic field (B) used in the simulations, along with the position of the injection locations at various radial distances and angles around the well. On the right, schematic decomposition of a 3D field into 100 layers, each of them representing a 2D field with the same planar spatial correlation as the 3D original counterpart.

Field (B), reported on the left of fig. 6.2, displays statistical axisymmetric anisotropy, with ratios $I_x = I_y$ and $I_x/I_z = I_y/I_z = 10$). Field (A) and (C) are statistically isotropic, but with different ratio between integral scales and domain size to explore the impact of the relative distance of injection. Notice that since the

same fields will be used afterwords to explore the effects of having higher variances (increase degree of heterogeneity) by simple scaling, fig. 6.2 does not report the legend on purpose.

To obtain pumping conditions in each heterogeneous field, we proceed as follows. In each realization and at each cell, Y is back-transformed to arithmetic values ($K = \exp(Y)$), and set as property conditions within the finite-difference numerical code Modflow-2000 ([Harbaugh 2000]). Flow simulations are run under steady-state conditions and the aquifer is considered to be confined. To simulate radial flow conditions, we set Dirichlet boundary conditions at the lateral domain sides, and assign a sink term to the bottom cell of the pumping well, situated at the center of the domain. The flow effects driven by a pumping well are achieved by setting a large anisotropy ratio in the local hydraulic conductivity between horizontal and vertical direction ($K_z/K_x = 10^5$) at the pumping well central column.

Finally, to simulate other Y-fields, characterized by the same heterogeneous architecture but different variances, we multiply each cell of the Y-field by a factor of 2 and $2\sqrt{2}$ to obtain new synthetic fields displaying σ_Y^2=4 and 8 respectively for fields A,B and C. No numerical convergence problems were found regardless the variance numbers.

6.2.2 Design of 3D transport simulations and estimation of depth-integrated BTCs

We simulate conservative solute transport using the random-walk particle-tracking code RW3D ([Fernandez-Garcia 2005, Salamon 2006a]) which is efficiently coupled with Modflow-2000. In this algorithm, M is discretized into N_p particles (the mass of each particle resulting $m_p = M/N_p$). Based on the Langevin equation, each particle moves according to a drift displacement, based on the radial flow velocities calculated at each cell of the domain by Modflow-2000, and a Brownian motion that accounts for local dispersive process. In our simulations, we set $N_P = 3 \cdot 10^4$ and a local dispersivity value $\alpha = 0.25$, isotropic in all directions. We used a constant porosity value equals to $\phi = 0.1$ in all our simulations.

Injection wells are simulated at 16 different locations around the extracting well, indicated in fig. 6.2, such that

- injection boreholes at points U1,R1,L1,D1 are located at a distance of 5 cells from the extraction well, oriented along the coordinate axis
- injection boreholes at points U2,R2,L2,D2 are located 12.5 cells away from the extraction well;
- injection boreholes at points U3,R3,L3,D3 are located 25 cells away from the extraction well;
- injection boreholes at points U4,R4,L4,D4 are located 75 cells away from the extraction well;

Each tracer test is independent from one another and by injecting sequentially in a vertical line at each of the 16 injection wells.The number of particles at each layer is assumed to be proportional to the local Darcy's fluxes, to better represent flux-averaged conditions. Particles are introduced in the system from layer k=5 to layer k=95; a few cells are skipped at the top and at the bottom, to avoid rebounds and other uncontrolled boundary effects. Injection takes place as a pulse, at the initial time ($t = 0$). Notice that our analysis is dimensionless.

We obtain 'depth-integrated' BTCs after estimating the density of the distribution of travel times of the ensemble of released particles. As such, we do not keep track of where (at which depth) particles were injected. Since each particle carries the same amount of mass (m_p), the estimated density distribution of travel time is equivalent to the estimation of a normalized BTCs.

The estimation of the density functions is found using the kernel density estimators described in Chapter 5. The numerical solution was successfully tested under homogeneous conditions against the formulae by [Moench 1989] and [Gelhar 1971], to ensure the transport was not affected by boundary conditions.

6.2.3 An illustrative result of a depth-integrated BTC

Let us first analyze the key aspects involved in the formation of a BTC obtained in one specific simulation. This will help us to highlight the main phenomenological feature need to understand subsequent simulations results. We focus on the traveling paths of the particles within the heterogeneous domain and its impact upon BTCs tailing.

We consider injection taking place at location L4 in field (B) (fig. 6.2) with $\sigma_Y^2 = 4$. Fig. 6.3 is divided in four subplots, each of them representing the position of 1000 particles (above) and the 'depth-integrated' BTCs (below) collected at the well at different temporal snapshots ranging from $t = 0$ to $t = 7 \cdot 10^5$. The background colors represent the distribution of the (log) hydraulic conductivities on the vertical section parallel to the x-axis, passing through both the injection and the extraction well.

At $t = 0$ (initial time) the particles are more concentrated in the highly conductive areas along the vertical section where Darcy's velocities are higher (red pixels) while only a few particles are located in less conductive areas (green pixels).

At $t = 2.5 \cdot 10^4$, particles located initially within the higher Y-zones have traveled longer distances (eventually reaching the extracting well) compared to those located in initially lower permeability areas. Since a large number of particles were initially onjected in relatively high Y areas, the concentration peak is clearly displayed at a relative short time ($t \approx 2.5 \cdot 10^4$).

At $t = 1 \cdot 10^5$, most of the particles (i.e. mass) have been collected, and only a few still remain in the system. Notice that the BTC scales very differently before and after this time. Especially, a well-defined tail with behavior similar to $c \approx t^{-1}$ is clearly visible from this moment, scaling from $t = 7 \cdot 10^5$ on with a heavy-tailed distribution similar to a PL with slope $m = 1$.

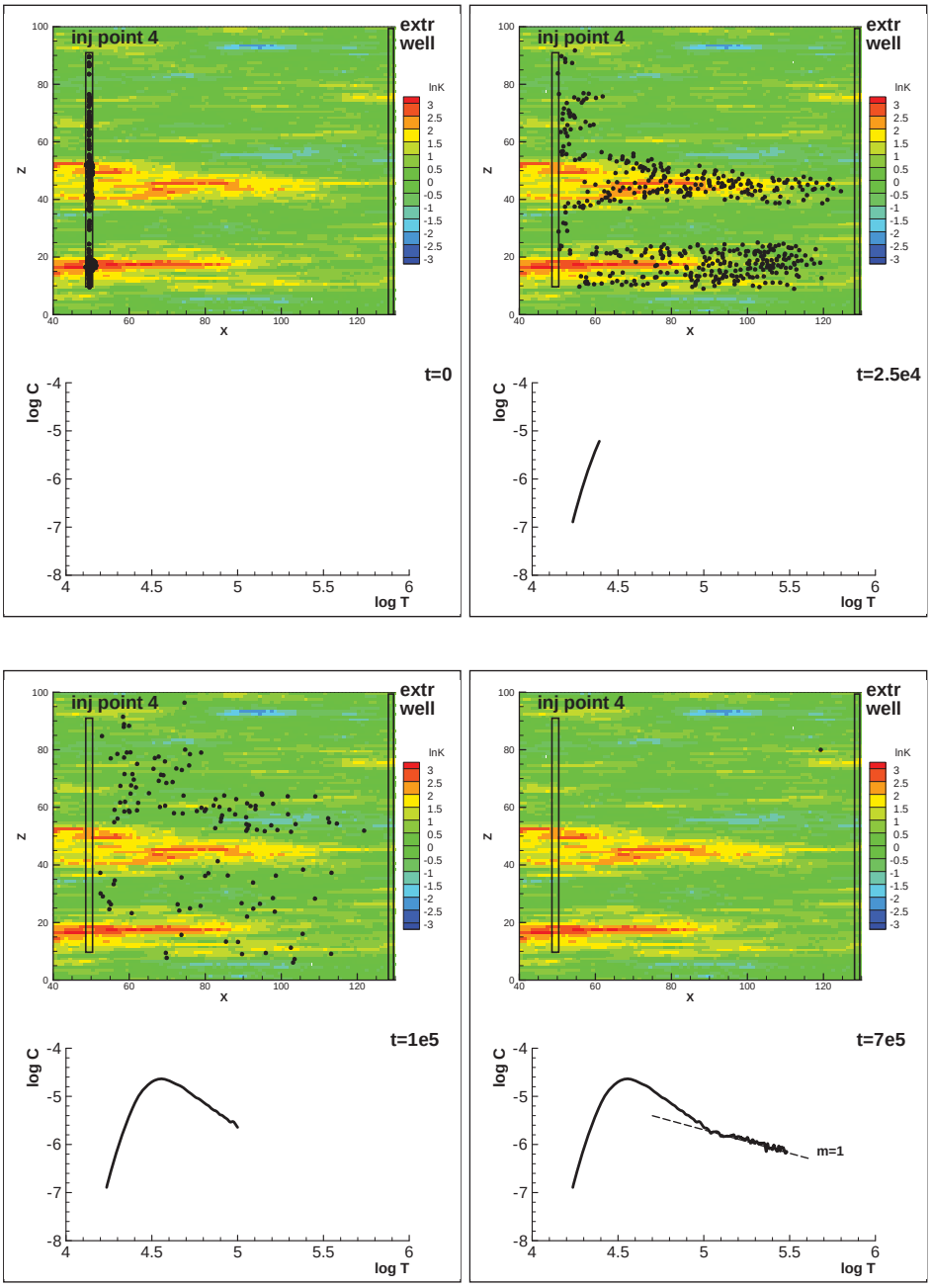

Figure 6.3: Particle displacement at different temporal steps during the simulation of a CFTT in the anisotropic field (B) with $\sigma_Y^2 = 4$, after injecting from position L4. In the upper plots, particles are plotted along with the distribution of log-transformed hydraulic conductivities, Y, at the central section of the domain. In the bottom, the evolution of the BTC estimated as 'depth-integrated' particle mass density at the extraction well.

This example indicates that our numerical settings are able to produce anomalous transport which gives rise to heavy-tailed BTCs, similar to experimental evidences (e.g. compare fig.6.3 with the E1 case in fig. 6.1). We can now take advantage of the numerical model to see whether we can provide a physical explanation for $m \approx 1$. To address this issue in detail, which will be more thoroughly discussed in Section 4, we study the impact of imperfect stratification in the 3D model.

6.2.4 Simulation of 'multilevel' transport

To evaluate the variability of solute mass arriving at different depths in a fully penetrating pumping well we propose the following approach, which is graphically resumed in fig. 6.2.

A proper evaluation of multilevel BTCs could be obtained by measuring the density of particle arrival time at each cell of the controlling section (the extraction well). Unfortunately, this would require a huge number of particle to inject in order to obtain an adequate estimate of the density at less connected or conductive level. In effect, observing fig. 6.3, particles are massively arriving at a few location along the well, while only a few particles arrive at other depths. This figure was generated with only 100 random particles as a graphical examples, but it is very well illustrative of this problem.

To partially overcome this issue, we opt for an alternative measure of multilevel BTCs. We account for the fact that in radial convergent flow, planar flow is dominant with respect to its vertical component. Moreover, at relative short injection distances compared with with the horizontal integral scales ($r/I_x \approx 1$), plumes migrate practically independently from one planar layer to the next. This effect is enhanced in fields displaying axisymmetric anisotropy.

Noticing that negligible vertical velocities and transverse dispersion leads to particle paths that exclusively follow the horizontal plane in which particles are initially injected, we proceed as follows. We separate each k layer composing the 3D block and use them as independent 2D Y-fields. The new 2D fields have 251 cells per side with unit thickness ($b = 1$) and same I_x, I_y, and σ_Y^2 than the 3D original source.

In each of the $N_k = 100$ decomposed layers, we reproduced 2D flow and transport and estimate BTCs, with a similar procedure described in previous sections. We inject the same number of particles ($N_P = 3 \cdot 10^5$) and the same mass ($M = 1$) per layer as a pulse release directly at the same 16 positions used for the 3D simulations and extract at a discharge rate Q (same as the one used for the 3D simulations). We thus obtained 100 BTCs (one for each layer) for each injection position. Herein, BTCs obtained from this methodology are called as '2D-BTCs'.

As an example of the output, we consider again the injection taking place at location L4 in field (B) with $\sigma_Y^2 = 4$. In fig. 6.4, in the big window, the same depth-integrated BTC obtained from 3D transport (fig. 6.3, bottom-right) is again reported as a thick, black line in the big picture. Along with that, the gray curves represent the 100 BTCs obtained after injecting in each layer composing the 3D

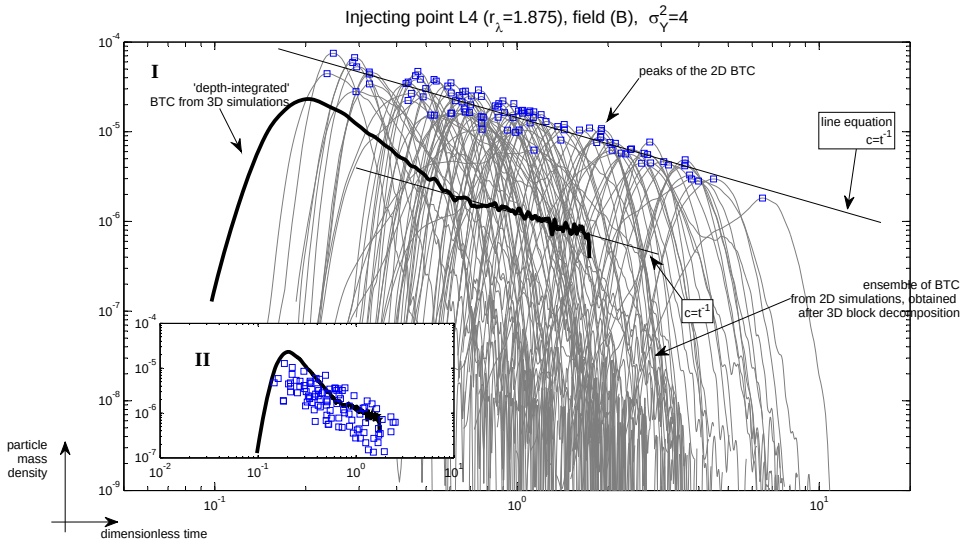

Figure 6.4: Superposition of particles pdfs (BTC) obtained after injecting at L4 in the the anisotropic field (B) with σ^2_Y=4 and $r_\lambda = 1.875$. 3D BTC is plotted with a dark strong line. 2D BTC from 'layer injection' are plotted with gray colors. A red 2D curve is plotted to emphasize one of the rare examples of heavy-tailed 2D BTC. The peaks of the 2D are marked in blue. Two lines with slope $c \approx t^{-1}$ are overlapping the 2D peak cloud and the late-time of the 3D BTCs to highlight this late-time scaling behavior.

block (i.e. the 2D BTCs). Distance and time are expressed with the following dimensionless variables

- Dimensionless time (t_c), which is obtained such that

$$t_c = \frac{t}{t_{adv}} \tag{6.1}$$

  where t_{adv} is the advective time, such that

$$t_{adv} = \frac{\pi r^2 b \phi}{Q} \tag{6.2}$$

  where b is the aquifer thickness ($b = 1$ for 2D simulations and $b = 100$ for 3D simulations).

- Dimensionless injection distance (r_λ), which is obtained to compare BTCs obtained from different realizations, showing different integral scales. We normalize r by the horizontal integral scale of each field, I_x, such that

$$r_\lambda = \frac{r}{I_x} \tag{6.3}$$

Note that in Eq.(6.2), we have used r rather than r_λ, with the purpose of emphasizing the effect of heterogeneity on the solute travel time.

We observe from 6.4) that, differently from the 3D integrated case, 2D-BTCs never show heavy-tailed BTCs, but are mostly symmetric (with rare exceptions). This means that the behavior of 2D fields is much more similar to the one found in homogeneous fields rather the one found in heterogeneous fields, implying that our 2D settings are not suitable to reproduce anomalous transport, despite the synthetic fields are obtained using the same horizontal variogram statistics than the 3D counterparts. Maybe, this is due to the use of unconditional 2D SGS simulations, similarly to what observed by [Willmann 2008].

Another remarkable effect is that 2D-BTCs look like convolved over several time lengths. This is indicated by the spread of the maximum peak of concentrations on each BTCs (marked by a blue square), which will be extensively used as a travel time indicator in the following lines The peak time is representing the position of most of these each symmetric curves, expect with the few exceptions showing asymmetric patterns. It can be noticed that the ensemble of maximum concentrations for the 2D BTCs (indicated by squares in fig. 6.4) scale similarly to a PL with unit slope, i.e. of the form $\widehat{p}(t) \approx t_c^{-1}$.

The reader should have noticed, however, that there is a fundamental difference between the estimated concentration for 2D and depth-integrated BTCs. This is due to the fact that 2D and 3D simulations account for the same N_P and M. However, in 3D simulation the mass is released along the vertical injection line is $M = 1$, but because of the flux-weighted injection scheme at each layer, the total mass at each layer is split differently among the layers . Thus, In 2D simulations, one expects that the injected mass should be equal to the one injected at each corresponding layer in 3D simulations. To do so, one needs to divide the estimated mass density at each 2D simulation corresponding to a layer k ($\widehat{p_k}(t)$) with the total mass injected in the 3D simulation at that layer (M_k). This is obtained as

$$\widehat{p_k}(t) = \frac{1}{D_F(k)} \frac{QC(t)}{\int QC(t)\mathrm{d}t} \tag{6.4}$$

where $D_F(k)$ is a dilution factor that is proportional to the mass injected in the system at each layer in 2D simulations and and 3D simulations, and $C(t)$ is the resident concentration at the well locations. $D_F(k)$ can be defined as

$$D_F(k) = \frac{q_x(k)}{\sum_{k=1}^{N_L} q_x(k)} \tag{6.5}$$

where $q_x(k)$ is the local seepage velocity at each injection locations. The effects of the mass normalization are shown in the small window in fig. 6.4. We only plot the 2D-BTC peaks for graphical purposes. It can be seen that the concentration peaks are found at similar concentrations as the 3D counterparts, confirming that the mass distribution among layer comes from a convolution of layer transport giving raise to the depth-integrated counterpart. The scattering between 2D peaks and 3D curves are due to the intrinsic different structure of 2D and 3D flow simulations (i.e.

we totally neglect the vertical advective and dispersive component), while they are expected to disappear using a proper multilevel sampling.

We present in the next section the most relevant findings extracted from the additional simulations, in order to find the impact of the degree of heterogeneity in the different slopes of the BTCs.

6.3 Comparison of different cases

We describe other significant results of our numerical analysis, to be compared with the illustrated case in fig. 6.4. The purpose is to show how exportable is the scaling as $\widehat{p}(t) \approx t_c^{-1}$ on the depth-integrated BTC tailing at late times, and to link this slope to some characteristic physical patterns, such as the distribution of peak times.

6.3.1 Evaluation of the anisotropic case (field B) with σ_Y^2=4

We compare here the results for other injection locations within the anisotropic field (B) with σ_Y^2=4. In addition to injection at L4, in fig. 6.5 we plot the results for other three injection locations, specifically at L1, L2 and L3 (mainly accounting for the different distances to the pumping well). In the plot, the depth-integrated BTCs (black line) are reported with a cloud of points representing the peaks of the 2D BTCs simulations (resulting after applying the 'layer decomposition' methodology described above, for each injection point).

We observe that depth-integrated BTC always show heavy-tailed distributions, but the BTCs slopes change as r_λ increases. Specifically, we observe that injecting at short distances (L1,L2) where the normalized injection distance is very small ($r_\lambda \approx 0.1, 0.3$), PL tailing is observed with constant slope, but different from (larger than) $\widehat{p}(t) \approx t_c^{-1}$. On the other hand, as soon as the injecting distance increases PL behaves as $\widehat{p}(t) \approx t_c^{-1}$ (L3,L4). Moreover, as the distance increases, the slope manifest for larger times. Notice that L3 and L4 curves display a shape very similar to the one of (respectively) E3 and E1 cases in fig. 6.1.

Note that the peak concentrations are not normalized using (6.4). This is done to emphasise that the cloud of 2D peaks, which is aligned to $\widehat{p}(t) \approx t_c^{-1}$ for all injection locations, has a relative spread depending on the injection position. It is being narrower at L1 and wider at L4, with L2 and L3 as intermediate cases. The spread is calculated as

$$\sigma_t^2 = \frac{1}{N_k - 1} \sum_{k=1}^{N_k} \left(\ln(t_{pk}(k)) - \overline{\ln(t_{pk})} \right) \tag{6.6}$$

where $\ln(t_{pk}(k))$ represent the log-transformed value of dimensionless temporal coordinate at which the peaks are observed on a specific 2D BTCs (i.e. at a specific layer k) and $\overline{\ln(t_{pk})}$ is the average of $\ln(t_{pk}(k))$. A low variance (e.g. $\sigma_t^2 \to 0$) would indicate that transport is very similar in all the horizons (e.g. homogeneous medium).

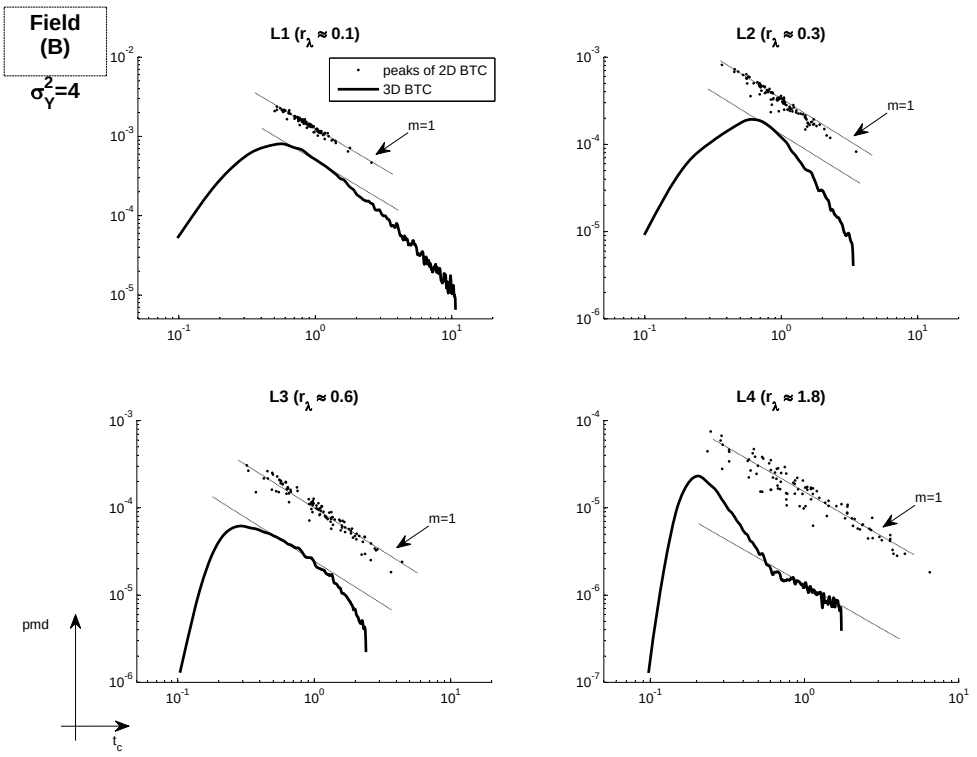

Figure 6.5: 3D BTCs (and relative peak distributions) obtained after injecting in the anisotropy field (B) with $\sigma^2 Y$=4, at L-points. Notice that the behavior $\widehat{p}(t) \approx t_c^{-1}$ can be fitted only at L4 ($r_\lambda \approx 1.8$) , while the other locations display a more symmetric behavior.

At L1, $\sigma_t^2 = 0.076$. This low value is due to the short normalized injection distances, which prevent the tracer to display a vertical stratification of the concentration measurable in terms of 'depth-integrated' BTCs. At L2 and L3, σ_t^2=0.165 and 0.324, respectively, emphasizing the development of vertical stratification of the plume, and enhancing difference between layers. At L4, where r_λ is close to the horizontal integral scale, $\sigma_t^2 = 0.561$, indicating that the plume is more stratified than in the other cases, and it is totally controlled by the different properties of each layer.

6.3.2 Comparison with isotropic fields

Considering that field (B) (fig.6.2) has anisotropic correlated structure of the hydraulic conductivity, a reasonable question is whether the stratification of the plume is controlled by the stratified distribution of Y. We therefore evaluate the behavior of BTCs 2D and 3D Y fields (A,C) constructed from isotropic models.

The curves for field (A) are plotted in fig. 6.6. We plot only the curves corresponding to the injection wells R1 to R4, since the other arrays (L,U,D) show similar

behavior. Being $I_x = 4$ in this field, the normalized injection distance is close to the unit for R1 ($r_\lambda \approx 1.8$) and increases to higher values (up to $r_\lambda \approx 18$ for R4).

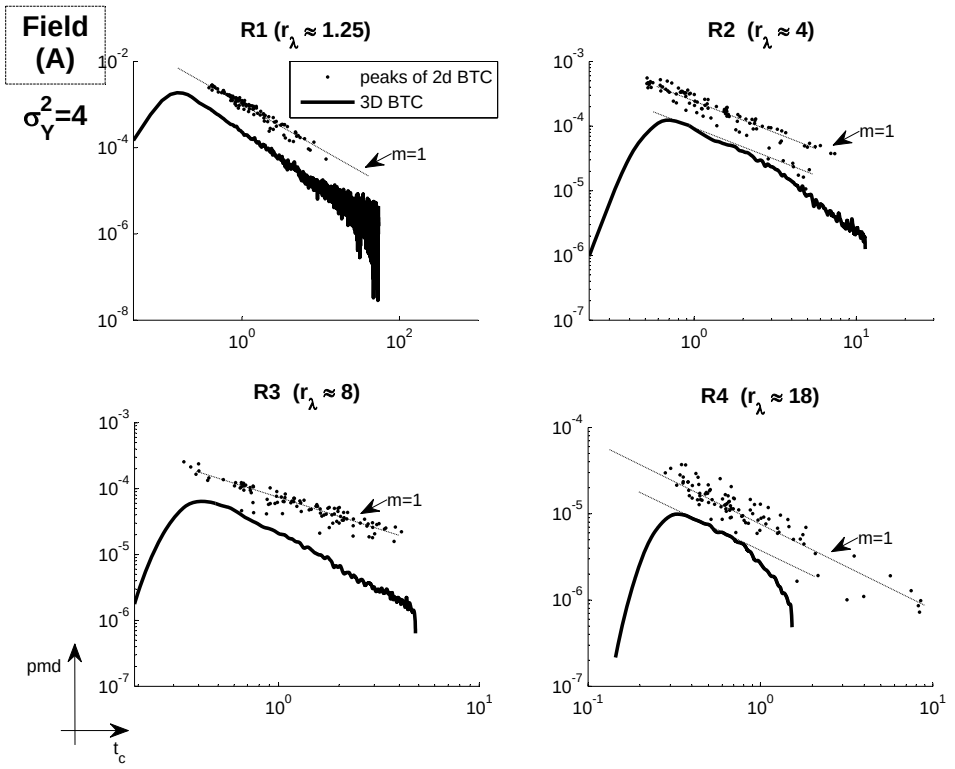

Figure 6.6: 3D BTCs (and relative peak distributions) obtained after injecting in the isotropic field (A) with $\sigma^2 Y$=4, at R-points. Notice that the behavior $\widehat{p}(t) \approx t_c^{-1}$ is very similar to the 3D BTC slope at L1 ($r_\lambda \approx 1.8$) , but tend to diverge as long as the injection distance increases. At R4, the BTCs is more symmetric.

.

A behavior very similar to $\widehat{p}(t) \approx t^{-1}$ is observed at R1 but as soon as the injection distance increases, this behavior tends to vanish, until it is no longer observable at R4. From R1 to R3, σ_t^2 decreases from 0.621 to 0.372. At R4, σ_t^2 is slightly higher than at in the other cases (0.717); it should be noticed however from fig. 6.6 that this value can be biased by the presence of a few outliers, since a few layers observed very high peak time. Disregarding the 5 larges values ot peak times (5% ot the total layers), for instance, σ_t^2 reduces to 0.341. This last point could indicate that $\widehat{p}(t) \approx t_c^{-1}$ is observed on the 3D BTC not only when σ_t^2 is high, but also that the condition '$r_\lambda \approx 1$ must be accomplished.

The curves for field (C) are plotted in fig. 6.7. This field shows the same horizontal integral scale (I_x) as field (B), but a relevant point is that the ratio between aquifer thickness and vertical integer scales is small ($b/I_z = 2.5$). Therefore, it should be expected a more homogenization in the BTCs. We found that, at very

short injection distances, such as R1, the distribution of 2D peaks is very narrow ($\sigma_t^2 = 0.07$), and no heavy-tailed behavior is clearly observed. However, it becomes very important as long as the distance increases (at R4, $\sigma_t^2 = 0.25$), similarly to the anisotropic field (B) (whose BTCs tends to scale with $\widehat{p}(t) \approx t_c^{-1}$ at R4).

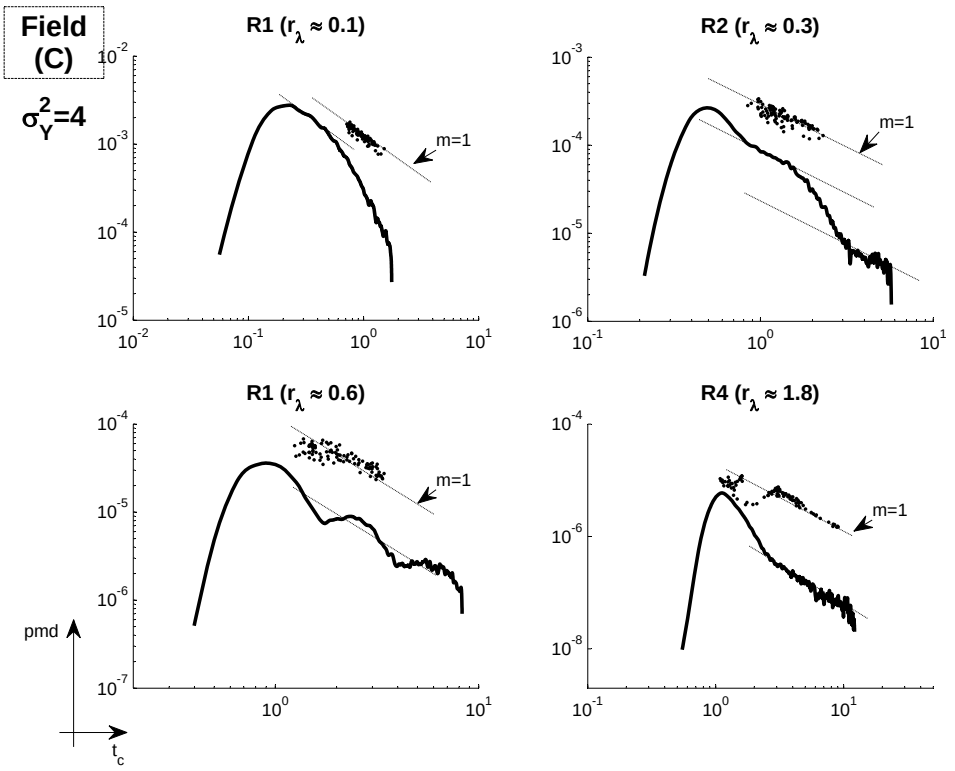

Figure 6.7: 3D BTCs (and relative peak distributions) obtained after injecting in the anisotropy field (>C) with $\sigma^2 Y$=4, at R-points. Notice that the behavior $\widehat{p}(t) \approx t_c^{-1}$ can be fitted at L4 ($r_\lambda \approx 1.8$). From R1 to R4, the 3D BTCs increasingly tend to heavy-tailed distributions; at R2 and R3, a PL behavior with $m = 1$ can be inferred at different intermediate portions of the BTCs.

This latter result suggests that, in our single realizations, the initial vertical stratification of the tracer is quite relevant, independently on the statistical structure of the media, when the injection takes place at a distance similar to the horizontal integral scale. As a consequence of (A) and (C), it looks like there is only a middle range of injection distances where $\widehat{p}(t) \approx t_c^{-1}$. If the distance is very small, the system acts as practically homogeneous and solutes tend to arrive at similar times, so that the BTC shape is controlled by dispersion, leading to quite symmetric distributions. Also, this happens for very large systems, where we can consider that the concentration are homogenized and no heavy-tailed behavior is observed.

Contrarily, for intermediate cases and when the system is highly stratified, BTCs develop a PL tail with unit slope value.

6.3.3 Effect of injection position on power-law scaling

In single stochastic realizations and especially under radial flow conditions, the solute behavior is non ergodic even when injection distance is located at tens of integrals scales apart from the well (e.g. [Matheron 1967, Rubin 2003]). Non ergodicity of the plume means that BTCs can vary very differently from realization to realization.

A direct consequence of ergodicity would be that all curves would not depend on the injection location, but only on its distance to the well. In our realizations, this was not the case, and so we compare here the solutions obtained for the BTCs recorded at points located at the same distance but in different directions, showing a significant non-ergodic effect.

In fig.6.2 we plot the depth-integrated BTC with relative 2D peaks obtained after mass injection at locations D2, L2, D4 and L4 in the three fields (A,B,C). For these simulations, we keep σ_Y^2=4. Results are shown in fig. 6.8. In the plot, features corresponding to injection points 'L' are displayed in red, while those corresponding to injection points 'D' are displayed in blue.

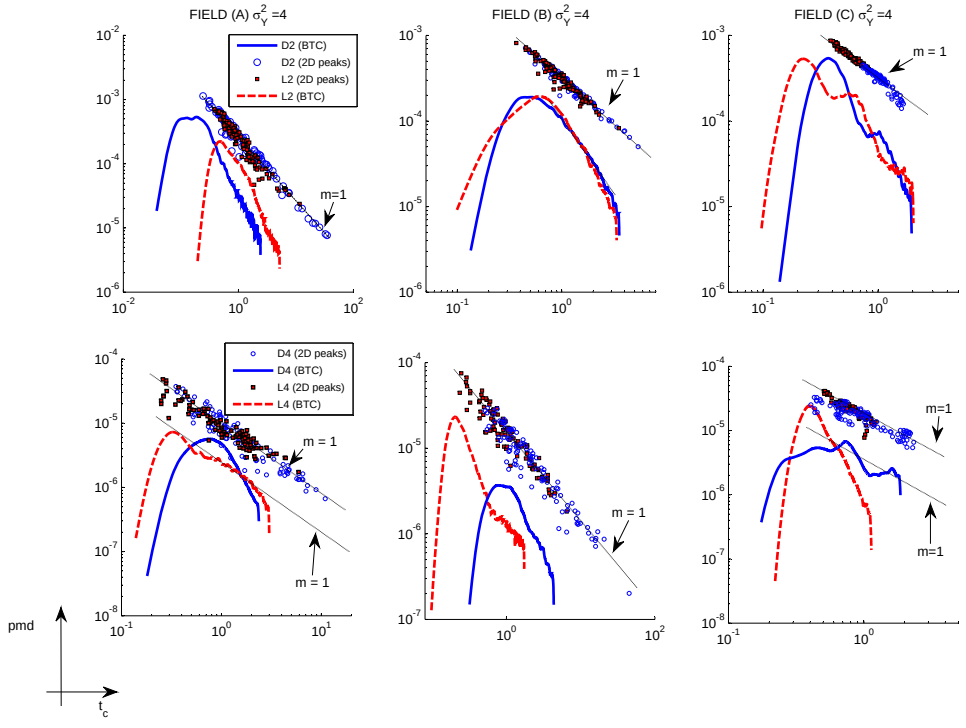

Figure 6.8: 3D BTCs (and relative peak distributions) after injecting in the the three fields A,B,C with σ_Y^2=4, at different locations. Red slopes indicate injections taking place from L2 (top) and L4 (bottom); in blue, injections taking place from D2 (top) adn D4 (bottom). The line indicate which BTcs display a behavior like $\widehat{p}(t) \approx t_c^{-1}$.

We observe that in field (A) (left) the shapes of BTC are slightly different for the injection locations L2 and D2 ($r_\lambda \approx 3$) and no PL behavior similar to $\widehat{p}(t) \approx t_c^{-1}$

is observed. At L4 ($r_\lambda \approx 18$), BTCs show a well-defined behavior $\widehat{p}(t) \approx t_c^{-1}$ at late time; here, the spread of the distribution of 2D peaks is close to the unit (σ_t^2 at L4 = 0.701). At D4, the variance is slightly larger σ_t^2=0.533, but analogous to what is seen for R4 (fig. 6.6), BTC does not scale with $\widehat{p}(t) \approx t_c^{-1}$.

Observing field (B) (center), the late-time behavior on BTCs at D2 and L2 ($r_\lambda \approx 0.6$) is practically the same, showing no constant PL tailing. Here, the spread of the distribution of 2D peaks is narrow ($\sigma_t^2 = 0.263$ for D2 and $\sigma_t^2 = 0.165$ for L2, indicating that once again the BTC does not develop for low vertical variability of the plume behavior. A more pronounced PL effect is found at D4, but not in the case of injection from L4. Again, despite both injection points are located at $r_\lambda \approx 1.8$, the spread of the 2D peaks distribution is much higher at D4 (σ_t^2=1.007) than at L4 (σ_t^2=0.562)

In field (C) (right), for injection locations D2 and L2, BTCs are found to follow a similar, symmetric behavior at late time ($\sigma_t^2 = 0.045$ at D2 and $\sigma_t^2 = 0.036$ at L2). At D4, the BTC shows a heavy-tailed distribution (roughly following a PL with $m \approx 1$, despite the disturbing presence of multiple peaks), and at L4 the BTC is more symmetric. Once again, σ_t^2 the spread of the peaks is higher when BTCs displays PL tailing ($\sigma_t^2 = 0.252$ late D4) than than where no tailing occur (σ_t^2= 0.056 at L4).

Once again, the combination of high σ_t^2 and r_λ close to the unit determine scaling similar $\widehat{p}(t) \approx t_c^{-1}$, independently from their position and type of geological setting.

6.3.4 Effect of Y variance on power-law scaling

The last parameter evaluated is the effect of the total variance of the log-transformed hydraulic conductivity, σ_Y^2. In our analysis, we consider the effect of σ_Y^2=1,4, and 8 on the anisotropic field (B) (fig.6.2 b), after injecting at two different locations (L2 and L4).

The results are summarized in fig. 6.9. The upper plots display the BTC from 3D simulations (along with the relative peaks of concentration of 2D BTCs) for σ_Y^2=1. The central plots refer to σ_Y^2=4 and the bottom plots for σ_Y^2=8. On the left, the plots indicate BTC (and peaks) obtained after injecting in L2, while on the right we plot the results on an injection from L4.

The general effect of σ_Y^2 is to increase σ_t^2. At L2, we can observe that the variance of the 2D peaks increases from σ_t^2=0.038 for σ_Y^2=1 to σ_t^2=0.329 for σ_Y^2=8. In no case we appreciate $c \approx t^{-1}$, which can be attributable to the fact that injection take place very close to the well ($r_\lambda < 1$).

At L4 (where $r_\lambda \approx 1.8$), BTCs are much more asymmetric. While for σ_Y^2=1 (having σ_t^2=0.143) BTCs do not show PL behavior as $\widehat{p}(t) \approx t^{-1}$, this effect is much more clearly show for the case σ_Y^2=4 and at σ_Y^2=8. In the latter cases, σ_t^2 is respectively 0.562 and 1.101.

From this last analysis, we have observed that the variance is a key controlling factor for the development of PL tailing on our BTCs; however, two other conditions (injection distance close the integral scale and high spread of the 2D peaks) need to

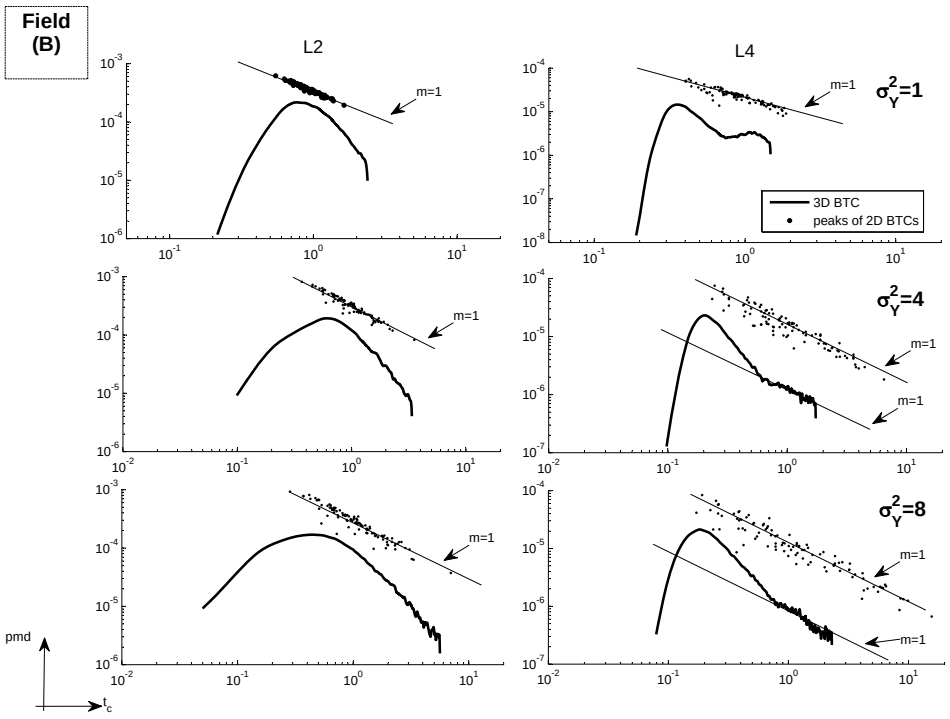

Figure 6.9: 3D BTCs (and relative peak distributions) for the anisotropic field (B) for variable variance (σ_Y^2=1,4,8, respectively from to the top to the bottom of the plot), at two injection locations, L2 and L4. Note that BTC tends to scale as a PL with $m \approx 1$ for L4 in case of variance $\sigma_Y^2 = 4$ and 8, which also show distribution of 2d peaks more spread than in the case where m is not found close to the unit.

be accomplished for PL tailing to occur.

6.4 Discussion: what does it mean $c \approx t^{-1}$ scaling?

From our results, we have observed that in most cases a breakthrough curve with a well-defined PL tailing of the form $\widehat{p}(t) \approx t_c^{-1}$ can develop during CFTTs in sandy aquifers. This behavior is highly variable, but common patterns seem to be identified to phenomenologically explain this behavior. Especially, we learned that the normalized injection distance for PL to take place on BTCs is $r_\lambda \approx 1$. Once this PL shape develops, its effect extend for larger times with increasing σ_Y^2.

To explain $m = 1$ on depth-integrated BTCs, we first consider why the slope of the peaks of 2D BTCs must also follow $m = 1$. We account for the fact that the architecture of the tracer, once it has been injected into the well, is stratified. This is well visible from fig. 6.3. Plume stratification is due to the existence of flow heterogeneities, that condition (I) the mass injected in each layer, which depend exclusively on the flow velocities at the cells located along the injection column and

(II) the mean travel time along each horizon.

The larger the integral scale, the further away the impact of layering at the injection well extends within the aquifer. On the other hand, when the injection point is far away from the pumping well, transport tends somehow to homogenize, despite it never reaches ergodic conditions.

The limit case is the one of a perfectly stratified medium , where each layer acts independently (assuming transversal dispersion is negligible). In each k-th layer, convergent radial transport can be approximated by the analytical solution of [Welty 1994] (their Eq. 26). After some manipulation, it can be seen from this equation that the peak of concentration c_{pk}, is

$$c_{pk} = \frac{M}{2Qt_{pk}} \left(\frac{4}{3}\pi \frac{1}{Pe} \right) \tag{6.7}$$

where Pe is the Peclet number, defined as

$$Pe = \frac{r}{\alpha_L} \tag{6.8}$$

in which α_L is the longitudinal dispersivity. This solution is valid for $Pe >\approx 10$, which is usually found in the field, according to [Gelhar 1992]. It highlights that, for each layer, the maximum concentration in each layer scales linearly with the layer-specific injected mass and inversely proportional to the discharge rate applied to this .

In the 2D-BTCs, injected mass is equal in all realizations so that the peak concentration is governed by Q and t_{pk}. Pe is constant for a given injection distance. Hence, from (6.7) we can say that

$$c_{pk} t_{pk} \propto B \frac{M}{Q} \tag{6.9}$$

where B is a constant. This relationship is sometimes used to design the total mass of solute tracer to be injected during experiments.

The question is how to translate this concept to depth-integrated BTCs. Here, a different problem occurs. Assuming again a perfectly stratified medium, we notice that for a given layer and due to boundary conditions (flux-averaged injection), the ratio M/Q should be constant. Thus, in each layer it is also satisfied that

$$c_{pk} \propto t_{pk}^{-1} \tag{6.10}$$

However, the further away from the ideal condition, the more the solution deteriorates. For instance, a less stratified medium (I_z/I_x not tending to zero) would manifest differently in the curves. A similar explanation can be sought for large injection distances. Due to the quasi-radial flow geometry, particles injected very far away from the well move initially with low velocity (compared with the mean aquifer velocity), so that dispersion dominates (small Pe). Then, the individual 2D BTCs would be less symmetric and show significant tailing (e.g. [Moench 1989]).

In section 3 we have shown that while 2D-BTCs exhibit a shape similar to a Gaussian bell, their corresponding peak time can drastically vary from one layer to another (measured by σ_t^2). From (6.9) this variability causes that for radial distances closes to the integral scale a power-law tailing behavior with $m = 1$ develops. This variability reflects the fact that the specific structure of the hydraulic conductivity contained within the particle path connecting the well with the injection is different for each layer. This is very similar to the concept of point-to-point connectivity ([Trinchero 2008]), who explained the importance of the structure and organization of K values along the particle path to predict travel times. Based on this, we can attribute large σ_t^2 to large variability of point-to-point connectivity patterns along the vertical line injection. In this line, it is worthwhile noticeable that one should consider an effective discharge rate Q to be used in 2D simulations. Such a value should be proportional to the effective discharge rate obtained at each layer of the well column in 3D simulations, and used in each of the 2D simulations. It should be noticed that the effect of rescaling Q at each layer would not affect the peak distributions but only σ_t. In effect Q only affects t_{pk} (because of (6.2)) and not the constant factor $c_{pk}t_{pk}$.

Conclusions of part 2

The main conclusions from this part are shown in the following paragraphs.

A locally-adaptive algorithm for particle density estimation

Particle tracking algorithms provide an attractive method to model solute transport in the subsurface, but several problems arise at the moment of estimating particle density functions. Here it is shown that various existing method to estimate these densities that using different number of particle and different classes of kernel densities fail to correctly reproduce various degree of peakness and tailing, when tested against several reference distributions.

We observed that kernel density estimators based on a global bandwidth method, which have a sound mathematical background, can qualitatively reproduce the general shape of skewed densities only if a huge number of particles is adopted. If a reduced number of particles is used, the estimated BTCs is not completely reliable, especially on the tails, which appear disturbed by numerical effects (such as oscillations). In addition, the computational time can be very high if several realizations are needed.

Classical adaptive methods, such as the one proposed by [Silverman 1986], provide a better fitting of the tails on heavy-tailed distributions, which appear smoothed and thus gradients could be more easily determined; however, they underestimated peak concentrations and perform worse in symmetric distributions. Since the shape of density functions is in most cases unknown, adaptive methods are not very reliable for typical applications in hydrogeology.

A new method has been proposed that uses the experimental cumulative distribution of particles to rescale the bandwidth size and adapt to the shape of the first estimated density. In this sense, an universal method that improves the fitting for any type of symmetry degree has been succesfully obtained and sensibly improve the fitting test against benchmark solutions. Only one calibration parameter is needed, and its selection may depend on the purpose of the study. The recommendation to use this method are the following:

1. If only a quick estimation is needed to determine the shape of the distribution, it can be sufficient to perform a single simulation using a large number of particles combined with the proposed method with $\alpha = 0.5$.

2. If a more accurate estimation is needed, three simulations are needed: the first, with a large number of particles and the fixed-bandwidth method and the other two with less particles and the proposed method with $\alpha = 0.5$ and $\alpha = 1$; thus, the correct parameter is chose depending on which of the two proposed methods is more similar to the fixed-bandwidth estimation;

3. If more simulations are needed, such as in a Monte Carlo framework, the proposed method can be used once α has been chosen after testing a few

realizations with large number of particles using the three methods, and then using the proposed one with a few number of particles.

Development of tailing

The processes involved in the development of heavy-tailed BTCs in heterogeneous porous media have been studied using a numerical approach based on single realizations of stochastic multigaussian simulations of log-normal hydraulic conductivity fields. A three-dimensional approach under convergent radial flow has been adopted, which appears a logical conceptual setting to adequately reproduce the usual configurations of real tracer tests. Nonetheless, the topic has received poor attention in the literature.

The focus on this Chapter has been to focus on the physical causes leading to power-law tails with low exponent (m=1), sometimes observed in real tests.

A first conclusion of this work is that in realistic three-dimensional settings, the late-time distribution of the concentrations observed in convergent flow field tracer tests is mainly controlled by thedegree of (pseudo) stratification whether it is real (physical) or caused by strong variantion in the hdyraulic conductivity.

The main conclusions of the work are that, for a given random field with predefined geostatistical properties, different late-time behaviors are observed for breakthrough curves obtained at different injection locations. When the following four conditions are accomplished:

1. large vertical variability of the connectivity for the layers composing the 3D formation;

2. disordered systems (with mid to high variance);

3. injection distance comparable with the planar horizontal integral scale of the heterogeneity;

the tendency of our synthetic fields is to display BTC scaling at late time with $c \approx t^{-1}$. The reason is that the maximum concentrations of breakthrough curves under general flow conditions depend inversely on the arrival time of the peak value. When these conditions are not or only partially fulfilled, BTC scale with a PL slope comprised between 1 and 4, which is similar to the observation reported in literature.

This work suggests also that three-dimensional models are necessary to reproduce skewed BTC similar to the ones observed in the field using sequential Gaussian simulations. Tailing does not naturally occurred in 2D transport simulations, unless new phenomena are accounted for such as nonlocality of the transport at the pore-scale. We can also suggest that the nature of the memory functions associated to these models in sandy aquifers are mostly related to the vertical architecture of the connectivity patterns between the injection location and the pumping well.

Anisotropy

The existence of a quantitative link relating the capacity coefficient β of mass-transfer models and hydraulic anisotropy has been shown. The results provide

a quantitative and robust evaluation of the strict dependence of β and physical parameters of the aquifers. Especially, a quick method to estimate β based on ratio between the concentration peak time from each breakthrough curve and the first moment of the arrival time distribution showed that β is largely controlled by the connectivity between injection and extraction wells, which is highly dependent upon anisotropy.

The work shows how important a characterization of the anisotropic structure of the aquifer is to predict the macroscopic behavior of β and to define the possible range of variability of this variable. More specifically, important conclusions that can be inferred from this work are listed as follow:

1. Connectivity has been found to be a key parameter controlling the development of apparent mass-transfer behavior and thus controlling β. From connectivity it depends how likely is the system to behave as an (apparent) mobile/immobile domain.

2. When connectivity is high, solute is injected through a few selected paths and injection distances are comparable with the horizontal integer scales (three conditions met by fields showing high permeable channels embedded in a relative less permeable matrix and tracer injection takes place as flux-averaged scheme from well located along the main direction of anisotropy), a relationship can be established between β and some connectivity indicators, such as the concentration peak time.

3. When connectivity is high but injection takes place along planes, such as in the case of fractures oriented along the well column, solute are more quickly mixed after being injected into the aquifer and they reach the controlling section without developing an (apparent) mass-transfer behavior. In this case, a relationship between anisotropy and β is not well displayed

The main conclusion is that the behavior of β during CFTTs can be predicted in as statistical sense when anisotropic formations are being characterized. These finding helped to find a direct link between mass-transfer upscaling parameter simulating apparent anomalous transport and physical soil parameters, (such as connectivity) which is in most of the time missing. While in specific field departures from ensemble mean can be important, this work shed new light on the possibility of render these effective solutions useful for predictive purposes and upscaling of solute transport modeling.

Conclusions

This thesis has dealt with several problems occurring during artificial recharge of aquifers and solute transport and mainly relate to hydrodynamic heterogeneities occurring at all scales in aquifers. The focus has been to develop tools to analysis and predict the effects of heterogeneity on the decisions in a more efficient and effective manner, as well as to understand what dynamics control macroscopic evidences of heterogeneity and the effects on the selection of effective modeling solutions.

The major common conclusions drawn from the two parts are listed as follows.

The technical tools that have been presented improve the modeling of artificial recharge and solute transport dynamics and can be used to obtain better estimation of random variables. Satellite images has been proved to bea new and powerful tool to characterize the uncertain spatial distribution of the infiltration capacity of aquifer on broad scales, in a cheap and quick manner. A new type of locally-adaptive kernel density estimators has been developed to increase the accuracy of particle density estimation from travel time distributions, reducing computational time and increasing the estimation of breakthrough curves from heterogenous aquifers.

The analysis on the macroeffects of heterogeneity on clogging and solute transport have shown that analytical models to upscale variability of the infiltration capacity and solute transport spatio-temporal dynamics should account for the nature of heterogeneity itself. It has been observed that in the presence of only few data about the infiltration capacity at the beginning and at the end of a flooding test, the wrong selection of an upscaling model can deliver uncertain spatio-temporal description of the variable, if the relationship between physical and model properties is not clear. When data do not exist, but the spatial structure of the clogging variables and the initial condition of the ponds are known, the relative intensity of each process (assuming that it depends exclusively on the initial conditions) can be highly variable and the decisions to be made about the maintenance of the ponds can be weighted depending on initial conditions of the pond, which can be for instance inferred by means of satellite images.

Solute transport in synthetic multigaussian fields characterized by fine-grid flow and particle-tracking transport methods displayed anomalous dynamics similar to those produce by mas-transfer models. The missing link between anomloaus transport in this type of fields seems to be related mostly to connectivity and flow stratification, which are in turn mostly controlled by the anisotropic structure of aquifers become stratified, the effects of connectivity are enhanced, confirming that both aspects are fundamental for the development of tailing in this cases. Connectivity has also been observed to control the apparent directional distribution of the capacity coefficient in statistically anisotropic aquifers. This behavior is emphasized observing in detail single realization as well as ensemble of realizations. The latter result provides new insights to obtain prediction about the distribution of the capacity coefficient in presence of anisotropic structure, which can be essential to make

accurate decisions concerning the characterization of solute transport properties in heterogeneous media.

Bibliography

[Abramson 1982] Ian S. Abramson. *On Bandwidth Variation in Kernel Estimates-A Square Root Law.* The Annals of Statistics, vol. 10, no. 4, pages 1217–1223, 1982.

[Abu-Taleb 1999] M. F. Abu-Taleb. *The use of infiltration field tests for groundwater artificial recharge.* Environ. Geol., vol. 37, no. 1-2, pages 64–71, 1999.

[Aish 2004] A. Aish and F. de Smedt. *Modeling of a groundwater mound resulting from artificial recharge in the Gaza Strip, Palestine.* In Water for life in the Middle East, 2nd Israeli-Palestinian international conference, Turkey, 10-14 October, 2004.

[Almeida 1996] A. Almeida and A. Journel. *Joint simulation of multiple variables with Markov-type coregionalization model.* Mathematical Geology, vol. 26, no. 5, pages 565–588, 1996.

[Andricevic 1996] R. Andricevic and V Cvetkovic. *Evaluation of Risk from Contaminants Migrating by Groundwater.* Water Resources Research, vol. 32, no. 3, pages 611–621, 1996.

[Andricevic 2008] R. Andricevic and V Cvetkovic. *Exposure concentration statistics in the subsurface transport.* Advances in Water Resources, vol. 31, pages 714–725, 2008.

[Aris 1956] R Aris. *On the Dispersion of a Solute by Diffusion, Convection and Exchange between Phases.* Proc R Soc Lond, vol. 252, no. 1271, pages 538–550, 1956.

[Asano 1985] T. Asano. *Overview: artificial recharge of groundwater.* Rapport technique, Butterworth Publisher, Stoneham, MA. Central Groundwater Board Guide on artificial recharge to groundwater. Ministry of water resources, New Delhi, India, 1985.

[ASCE 2001] ASCE. *Standard Guidelines for Artificial Recharge of Ground Water.* Rapport technique, EWRI/ASCE 34-01 (ASCE Standard No. 34-01), 2001.

[Avnimelech 1983] Y. Avnimelech and R. G. Menzel. *Biologically controlled flocculation of clay in lakes.* In EGU General Assembly 2010, editeur, Developments in Ecology and Environmental Quality. H.I. Shuval (Ed.), 2: 257-265. Balaban, Rehovot., volume 12 of *EGU2010-5326*, 1983.

[Barahona-Palomo 2010] M. Barahona-Palomo, D. Pedretti and X. Sanchez-Vila. *Infiltration tests at the Sant Vicenç dels Horts artificial recharge experimental site.* In EGU General Assembly 2010, editeur, Geophysical Research Abstracts, volume 12 of *EGU2010-5326*, 2010.

[Batchelor 1998] B. Batchelor, J. Valdes and V. Araganth. *Stochastic risk assessment of sites contaminated by hazardous wastes.* J. Environ. Eng., vol. 124, pages 380–388, 1998.

[Baveye 1989] P. Baveye and V. A. Valocchi. *An evaluation of mathematical models of the transport of biologically reacting solutes in saturated soils and aquifers.* Water Resour. Res., vol. 25, no. 6, pages 1413–1421, 1989.

[Baveye 1998] P. Baveye, P. Vandevivere, B. L. Hoyle, P. C. de Leo and D. de Lozada Sanchez. *Environmental Impact and Mechanisms of the Biological Clogging of Saturated Soils and Aquifer Materials.* Critical Rev. Environ. Sci. Tech., vol. 28, no. 2, pages 123–191, 1998.

[Bear 1972] J. Bear. Dynamics of fluids in porous media. Elsevier, New York, 1972.

[Becker 2000] M. Becker and A. Shapiro. *Tracer transport in fractured crystalline rock: Evidence of nondiffusive breakthrough tailing.* Water Resour. Res., vol. 36, pages 1677–1686, 2000.

[Becker 2003] M. Becker and A M Shapiro. *Interpreting tracer breakthrough tailing from different forced-gradient tracer experiment configurations in fractured bedrock.* Water Resources Research, vol. 39, no. 1, page W1014, 2003.

[Bedford 2001] T. Bedford and R. Cook. Probabilistic risk analysis: Fundations and methods. Cambridge University Press, New York, 2001.

[Bellin 2004] Alberto Bellin and Yoram Rubin. *On the use of peak concentration arrival times for the inference of hydrogeological parameters.* Water Resources Research, vol. 40, no. 7, pages 1–13, 2004.

[Bensabat 2000] J Bensabat, Q Zhou and J. Bear. *An adaprtive pathline-based particle tracking algorithm for the Euleriam-Lagrangian method.* Advances in Water Resources, vol. 23, pages 383–397, 2000.

[Benson 2000] DA Benson, SW Wheatcraft and MM Meerschaert. *Application of a fractional advection-dispersion equation.* Water Resour. Res., vol. 36, no. 6, pages 1403–1412, 2000.

[Benson 2009] D A Benson and M M Meerschaert. *A simple and efficient random walk solution of multi-rate mobile/immobile mass transport equations.* Advances in Water Resources, vol. 32, pages 532–539, 2009.

[Berglund 1995] S Berglund and V Cvetkovic. *Pump and treat remediation of heterogeneous aquifers: Effects of rate-limited mass transfer.* Ground Water, vol. 33, no. 4, pages 675–685, 1995.

[Berkowitz 2006] B Berkowitz, A Cortis, M Dentz and H Scher. *Modeling non-Fickian transport in geological formations as a continuous time random walk.* Reviews of Geophysics, vol. 44, page RG2003, 2006.

[Bianchi 2011] M. Bianchi, C. Zheng, G. Tick and S. Gorelick. *Investigation of Small-Scale Preferential Flow with a Forced-Gradient Tracer Test.* Ground Water, vol. 49, no. 4, pages 503–514, 2011.

[Bijeljic 2006] B Bijeljic and MJ Blunt. *Pore-scale modeling and continuous time random walk analysis of dispersion in porous media.* Water Resources Research, vol. 42, page W01202, 2006.

[Bodhinayake 2004] W. Bodhinayake, B.C. Si and K. Noborio. *Determination of Hydraulic Properties in Sloping Landscapes. Tension and Double-Ring Infiltrometers.* Vadose Zone Journal, vol. 3, pages 964–970, 2004.

[Bolster 2009] D. Bolster, M. Barahona-Palomo, M. Dentz, D. Fernàndez Garcia, X. Sanchez-Vila, P. Trinchero, C. Valhondo and D. M. Tartakovsky. *Probabilistic Risk Assessment applied to contamination scenarios in porous media.* Water Resour. Res., vol. 45, pages W06413, doi:10.1029/2008WR007551, 2009.

[Bons 1996] P. Bons and M. W. Jessell. *Image analysis of microstructures in natural and experimental samples.* In Declan G. De Paor, editeur, Structural Geology and Personal Computers, volume 15 of *Computer Methods in the Geosciences*, pages 135–166. Pergamon, 1996.

[Bouwer 1986] H. Bouwer. *Intake Rate: Cylinder Infiltrometer.* In A. Klute, editeur, Methods of Soil Analysis. Part 1. hysical and Mineralogical Methods", volume 32 of *P. Agronomy*, pages 825–844. Amer. Soc. Agronomy & Soil Sci. Soc. America, Wisconsin, 1986.

[Bouwer 1988] H. Bouwer. *Design and Management of Infiltration Basins for Artificial Recharge of Ground Water.* In 32nd Annual New Mexico Conference on Ground Water Management. Albuquerque, NM November 5-6, 1987, 1988.

[Bouwer 2002] H. Bouwer. *Artificial recharge of groundwater: hydrogeology and engineering.* Hydrogeol. J., 2002.

[Boyd 1974] R. H. Boyd and M. M. Ghosh. *An investigation of the influence of some physicochemical variables on porous-media filtration.* J. Am. Water Works Assn., vol. 66, no. 2, pages 94–98, 1974.

[Brooks 1964] R. H. Brooks and A. T. Corey. *Hydraulic properties of porous media.* Colorado St. Univ. Hydrol. Paper, vol. 3, page 27, 1964.

[Brusseau 1990] M L Brusseau and P S C Rao. *Modeling Solute transport in Structured Soils: A Review.* Geoderma, vol. 46, pages 169–192, 1990.

[Cameron 1997] D R Cameron and A Klute. *Convective-dispersive solute transport with a combined equilibrium and kinetic adsorption model.* Water Resources Research, vol. 13, no. 1, pages 183–188, 1997.

[Carman 1938] P.C. Carman. *The determination of the specific surface of powders.* J. Soc. Chem. Ind. Trans, vol. 57, page 225, 1938.

[Caro 1981] R. Caro and P.S. Eagleson. *Estimating aquifer recharge due to rainfall.* J. Hydrol., vol. 53, no. 1, pages 185–211, 1981.

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

[Carrera 2005] J. Carrera, E. Vazquez-Sune, E. Abarca, B. Capino, D. Gamez, A. Simo, J.M. Ninerola and E. Queralt. El baix llobregat. historia i actualitat ambiental d'un riu. ., chapitre Les aigues subterranies del Baix Llobregat., pages 72–92. 2005.

[CGWB 2007] CGWB. *Manual of artificial recharge of ground water.* Rapport technique, Government of India. Ministry of water resources. Central Grond Water Board. New Delhi., 2007.

[Chang 2009] Y. I. Chang, W.Y. Cheng and H.C. Chan. *A proposed correlation equation for predicting filter coefficient under unfavorable deposition conditions.* Sep. Purif. Technol., vol. 65, no. 3, pages 248–250, 2009.

[Chapuis 2003] R P Chapuis and M Aubertin. *On the use of the Kozeny-Carman equation to predict the hydraulic conductivity of soils.* Can. Geotech. J., vol. 40, pages 616–628, 2003.

[Cheng 2009] Y. Cheng, C.H. Lee, Y.C. Tan and H.F. Yeh. *An optimal water allocation for the Ailiao irrigation district in Pingtung Country, Taiwan.* Irrigation and Drainage Systems, vol. 58, no. 3, pages 287–306, 2009.

[Chica-Olmo 2000] M. Chica-Olmo and F. Abarca-Hernandez. *Computing geostatistical image texture for remotely sensed data classification.* Computers & Geosciences, vol. 26, pages 373–383, 2000.

[Christakos 2000] G Christakos. Modern spatiotemporal geostatistics. Oxford University Press, New York, 2000.

[Christiansen 1944] J. E. Christiansen. *Effect of Entrapped Air Upon the Permeability of Soils.* Soil Sci. Soc. Am. J., vol. 58, no. 5, pages 355–366, 1944.

[Civan 2005a] F. Civan and V. Nguyen. *Modeling particle migration and deposition in porous media by parallel pathways with exchange.* In K. Vafai, editeur, Handbook of Porous Media, pages 457–484. CRC Press, Boca Raton, FL, 2nd édition, 2005.

[Civan 2005b] F. Civan and M. L. Rasmussen. *Analytical models for porous media impairment by particles in rectilinear and radial flows.* In K. Vafai, editeur,

Handbook of Porous Media, pages 485–542. CRC Press, Boca Raton, FL, 2nd édition, 2005.

[Civan 2007] F. Civan. *Temperature effect on power for particle detachment from pore wall described by an Arrhenius-type equation.* Trans. Porous Med., vol. 67, no. 2, pages 329–334, 2007.

[Civan 2010] F. Civan. *Non-isothermal Permeability Impairment by Fines Migration and Deposition in Porous Media including Dispersive Transport.* Transp. Porous Med., pages doi:10.1007/s11242–010–9557–0, 2010.

[Clauset 2007] A Clauset, CR Shalizi and ME Newman. *Power-law distribution in empirical data.* SIAM Review, vol. 51, pages 661–703, 2007.

[Clement 1996] T. P. Clement, B. S. Hooker and R. S. Skeen. *Macroscopic Models for Predicting Changes in Saturated Porous Media Properties Caused by Microbial Growth.* Groundwater, vol. 34, no. 5, pages 934–942, 1996.

[Coats 1964] K H Coats and B D Smith. *Dead-End Pore Volume and Dispersion in Porous Media.* Society of Petroleum Engineers Journal, vol. 4, no. 1, pages 73–84, 1964.

[Cunningham 1991] A B Cunningham, F Characklis W G Abedeen and D Crawford. *Influence of biofilm accumulation on porous media hydrodynamics.* Environmental Science and Technology, vol. 25, no. 7, pages 1305–1311, 1991.

[Custodio 2002] E. Custodio. *Aquifer overexploitation: what does it mean?* Hydrogeology Journal, vol. 10, pages 254–277, 2002. 10.1007/s10040-002-0188-6.

[Dagan 1989] G Dagan. Flow and transport in porous formations. Springer-Verlag, Berlin, 1989.

[Dagan 1992] G Dagan, V Cvetkovic V and A M Shapiro. *A solute flux approach to transport in heterogeneous formations: 1. The general framework.* Water Resources Research, vol. 28, no. 5, pages 1369–1376, 1992.

[Das Gupta 1988] A. Das Gupta and G.N. Paudyal. *Estimating aquifer recharge and parameters from water level observations.* J. Hydrol., vol. 99, pages 103–116, 1988.

[de Barros 2008] F. P. J. de Barros and Y. Rubin. *A risk-driven approach for subsurface site characterization.* Water Resour. Res., vol. 58, page W01414, 2008.

[de Barros 2009] F. P. J. de Barros, Y. Rubin and R.M Maxwell. *The concept of comparative information yield curves and their application to risk-based site characterization.* Water Resour. Res., vol. 45, page W06401, 2009.

[Dentz 2003] M Dentz and B Berkowitz. *Transport behavior of a passive solute in continuous time random walks and multirate mass transfer.* Water Resour. Res., vol. 39, no. 5, pages 1111–1131, 2003.

[Deutsch 1998] C. Deutsch and A. Journel. *GSLIB: Geostatistical Software Library and User's Guide.* Rapport technique, Oxford University Press, New York, 340 p., 1998.

[Dillon 2002] P. Dillon, editeur. Management of Aquifer Recharge for Sustainability: Proceedings of the 4th International Symposium on Artificial Recharge of Groundwater, Adelaide, Sep 2001., New York, 2002. Taylor & Francis.

[Drazen 1998] A. Drazen and E. Helpman. *Economic effects on the government budgets.* In E. Helpman, A. Razin and E. Sadka, editeurs, Stabilization with exchange rates management under uncertainty, chapitre 16. MIT Press, Boston, 1998.

[Ernisee 1975] J. J. Ernisee and W. H. Abott. *Binding of mineral grains by a species oh Thalassiosira.* Nova Hedwigia Beih., vol. 53, pages 241–252, 1975.

[Fernàndez-Garcia 2002] D. Fernàndez-Garcia, X. Sanchez-Vila and TH Illangasekare. *Convergent-flow tracer tests in heterogeneous media: combined experimental-numerical analysis for determination of equivalent transport parameters.* Journal of Contaminant Hydrology, vol. 57, no. 1-2, pages 129 – 145, 2002.

[Fernàndez-Garcia 2003] D. Fernàndez-Garcia. *Scale-dependence of non-reactive and sorptive transport parameters estimated from radial and uniform flow tracer tests in heterogeneous formations: experimental and numerical inverstigations.* PhD thesis, Colorado School of Mines, 2003.

[Fernàndez-Garcia 2004] D. Fernàndez-Garcia, TH Illangasekare and H. Rajaram. *Conservative and sorptive forced-gradient and uniform flow tracer tests in a three-dimensional laboratory test aquifer.* Water Resources Research, vol. 40, page W10103, 2004.

[Fernandez-Garcia 2005] D Fernandez-Garcia, H. Rajaram and T. Illangasekare. *Assessment of the predictive capabilities of stochastic theories in a three-dimensional laboratory test aquifer: Effective hydraulic conductivity and temporal moments of breakthrough curves.* Water Resources Research, vol. 41, page W04002, 2005.

[Fernàndez-Garcia 2010] D. Fernàndez-Garcia, P. Trinchero and X. Sanchez-Vila. *Conditional stochastic mapping of transport connectivity.* Water Resour. Res., vol. 46, page W10515, 2010.

[Fernàndez-Garcia 2011] D. Fernàndez-Garcia and X. Sanchez-Vila. *Optimal reconstruction of concentrations, gradients and reaction rates from particle distributions.* Journal of Contaminant Hydrology, vol. 120-121, pages 99–114, 2011.

[Finnemore 1995] E. J. Finnemore. *A program to calculate groundwater mound heights.* J. Hydrol., vol. 33, no. 1, pages 139–143, 1995.

[Fiori 2006] A. Fiori, I. Jankivic and G Dagan. *Modeling flow and transport in highly heterogeneous three-dimensional aquifers: Ergodicity, Gaussianity, and anomalous behavior-2. Approximate semianalytical solution.* Water Resources Research, vol. 42, page W06D13, 2006.

[Fiori 2012] A. Fiori and I. Jankovic. *On prefential flow, channeling and connectivity in heterogeneous porous formations.* Math Geosci, vol. 44, pages 133–145, 2012.

[Flach 2012] G. Flach. *Relationship between dual-domain parameters and practical characterization data.* Ground Water, vol. 50, no. 2, pages 216–229, 2012.

[Fogg 1986] G E Fogg. *Groundwater-flow and sand body interconnectedness in a thick,multiple-aquifer system.* Water Resources Research, vol. 11, no. 5, pages 679–694, 1986.

[Foster 2006] S. Foster and D P. Loucks, editeurs. Non-renewable groundwater resources. a guidebook on socially-sustainable management for water-policy makers. United Nations Educational, Scientific and Cultural Organization, Paris (France), 2006.

[Freeze 1975] R. A. Freeze. *A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media.* Water Resour. Res., vol. 11, no. 5, pages 725–741, 1975.

[Fry 1995] V. A. Fry, J. D. Istok, L. Semprini, K. T. O'Reilly and T. E. Buscheck. *Retardation of Dissolved Oxygen Due to a Trapped Gas Phase in Porous Media.* Groundwater, vol. 33, no. 3, pages 391–398, 1995.

[Gale 2005] I. Gale. *Strategies for managed aquifer recharge (MAR) in semi-arid areas.* Rapport technique, IAH-MAR, UNESCO IHP, Paris, France, 2005.

[Gee 1988] Glendon W. Gee and Daniel Hillel. *Groundwater recharge in arid regions: Review and critique of estimation methods.* Hydrological Processes, vol. 2, no. 3, pages 255–266, 1988.

[Gelhar 1971] L. Gelhar and MA Collins. *General analysis of longitudinal dispersion in nonuniform flow.* Water Reso, vol. 7, no. 6, pages 1511–1521, 1971.

[Gelhar 1992] L.W. Gelhar, C. Welty and K.R. Rehfeldt. *A critical review of data on field-scale dispersion in aquifers.* Water Resour. Res., vol. 28, pages 1955–1974, 1992.

[Gelhar 1993] L W Gelhar. Stochastic subsurface hydrology. Prentice Hall, NJ, 1993.

[Goetz 1985] A.F.H. Goetz, G. Vane, J.E. Solomon and B.N. Rock. *Imaging spectrometry for earth remote sensing.* Science, vol. 228, no. 4704, pages 1174–1153, 1985.

[Goltz 1987] M N Goltz and Roberts P V. *Using the method of moments to analyze three-dimensional diffusion-limited solute transport from temporal and spatial perspectives.* Water Resources Research, vol. 23, no. 8, pages 1575–1585, 1987.

[Gomez-Hernandez 1998] JJ. Gomez-Hernandez and XH. Wen. *To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology.* Advances in Water Resources, vol. 21, no. 1, pages 47–61, 1998.

[Gooverts 1997] P. Gooverts. Geostatistics for environmental applications. Oxford University Press, USA, 1997.

[Gouze 2008] P Gouze, Y Melean, T Le Borgne, M Dentz and J Carrera. *Non-Fickian dispersion in porous media explained by heterogeneous microscale matrix diffusion.* Water Resources Research, vol. 44, page W11416, 2008.

[Granger 2000] R.J. Granger. *Satellite-derived estimates of evapotranspiration in the Gediz basin.* Journal of Hydrology, vol. 229, pages 70–76, 2000.

[Greskowiak 2005] J. Greskowiak, H. Prommer, G. Massmann, C. D. Johnston, G. Nützmann and A. Pekdeger. *The impact of variable saturated conditions on the hydrochemistry during artificial recharge of groundwater – A field study.* Appl. Geochem., vol. 20, pages 1409–1426, 2005.

[Guin 1972] James A. Guin. *Clogging of Nonuniform Filter Media.* Industrial & Engineering Chemistry Fundamentals, vol. 11, no. 3, pages 345–349, 1972.

[Haggerty 1995] R Haggerty and S Gorelick. *Multiple-rate mass transfer for modeling diffusion and surface reactions in media with pore-scale heterogeneity.* Water Resour. Res., vol. 31, no. 10, pages 2383–2400, 1995.

[Haggerty 2000] R Haggerty, SA McKenna and LC Meigs. *On the late-time behavior of tracer test breakthrough curves.* Water Resour. Res., vol. 36, no. 12, pages 3467–3479, 2000.

[Haggerty 2002] R Haggerty and PC Reeves. *STAMMT-L 1.0, Formulation and User's Guide, Technical Report ERMS 520308.* Rapport technique, Sandia National Laboratories, Albuquerque, NM, USA, 2002.

[Hantush 1967] M. S. Hantush. *Growth and decay of groundwater mounds in response to uniform percolation.* Water Resour. Res., vol. 3, pages 227–234, 1967.

[Harbaugh 2000] A.W. Harbaugh, E.R. Banta, Mary C. Hill and M. G. McDonald. *MODFLOW-2000, the U.S. Geological Survey modular ground-water model – User guide to modularization concepts and the Ground-Water Flow Process.* Rapport technique, U.S. Geological Survey Open-File Report 00-92, 121 p, 2000.

[Hardle 1990] W. Hardle. Smoothing techniques with implementation in s. Springer-Verlag, New York, 1990.

[Harvey 1995] C F Harvey and S Gorelick. *Temporal Moment-Generating Equations: Modeling Transport and Mass Transfer in Heterogeneous Aquifers.* Water Resources Research, vol. 31, no. 8, pages 1895–1911, 1995.

[Hazen 1882] A. Hazen. *Some physical properties of sands and gravels, with special reference to their use in filtration.* 24th Annual Rep., Massachusetts State Board of Health, Pub. Doc., vol. 34, pages 539–556, 1882.

[Hoehn 1998] E Hoehn, J. Eikenberg, T Fierz, W Drost and E Reichlmayr. *The Grimsel Migration Experiment : field injection-withdrawal experiments in fractured rock with sorbing tracers.* Journal of Contaminant Hydrology, vol. 34, pages 85–106, 1998.

[Hoffmann 2010] A. Hoffmann and G. Gunkel. *Bank filtration in the sandy littoral zone of Lake Tegel (Berlin): Structure and dynamics of the biological active filter zone and clogging processes.* Limnologica - Ecology and Management of Inland Waters, vol. In Press, Corrected Proof, pages –, 2010.

[Hofkes 1986] E.H. Hofkes and J.T. Visscher. *Artificial groundwater recharge for water supply of medium-size communities in developing countries.* In International Reference Centre for Community Water Supply and Sanitation, The Hague, 1986.

[Irons 1989] J.R. Irons, R.A. Weismiller and G.W. Petersen. Theory and applications of optical remote sensing. g. asrar, ed., chapitre Soil reflectance, pages 66–106. Wiley-Interscience, New York, 1989.

[Iwasaki 1937] T. Iwasaki. *Some Notes on Sand Filtration.* J. Am. Water Works Assn., vol. 29, pages 1597–1602, 1937.

[Jha 2009] M.K. Jha, Y. Kamii and K. Chikamori. *Cost-effective Approaches for Sustainable Groundwater Management in Alluvial Aquifer Systems.* Water Resources Management, vol. 23, no. 2, pages 219–233, 2009.

[Journel 1999] A. Journel. *Markov models for crosscovariances.* Mathematical Geology, vol. 31, no. 8, pages 955–964, 1999.

[Kaasschieter 1995] E F Kaasschieter. *Mixed finite elements for accurate particle tracking in saturated groundwater.* Advances in Water Resources, vol. 18, no. 5, pages 277–294, 1995.

[Kim 2010] Jung-Woo Kim, Heechul Choi and Yakov A. Pachepsky. *Biofilm morphology as related to the porous media clogging.* Water Research, vol. 44, no. 4, pages 1193 – 1201, 2010. Transport and Fate of Colloids and Microbes in Granular Aqueous Environments.

[Kinzelbach 1987] W Kinzelbach. *The random walk method in pollutant transport simulation. Advances in analytical and numerical groundwater flow and quality modelling.* Rapport technique, In: Custodio, E., et al. (Eds.), NATO ASI Series C, vol. 224, pp. 227-246., 1987.

[Knudby 2006] Christen Knudby and Jesús Carrera. *On the relationship between indicators of geostatistical, flow and transport connectivity.* Advances in Water Resources, vol. 28, no. 4, pages 405 – 421, 2006.

[Koltermann 1996] C E Koltermann and S Gorelick. *Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches.* Water Resources Research, vol. 32, no. 9, pages 2617–2658, 1996.

[Kozeny 1927] J. Kozeny. *Uber kapillare leitung des wassers im boden.* Sitzungsber. Akad. Wiss. Wien, vol. 136, pages 271–306, 1927.

[Kreft 1978] A Kreft and A Zuber. *On the physical meaning of the dispersion equation and its solutions for different initial and boundary conditions.* Chemical Engineering Science, vol. 33, pages 1471–1480, 1978.

[Krzysztofowicz 2001] R. Krzysztofowicz. *The case for probabilistic forecasting in hydrology.* Journal of Hydrology, vol. 249, pages 2–9, 2001.

[Latinopoulos 1986] P. Latinopoulos. *Analytical solutions for strip basin recharge to aquifers with Cauchy boundary conditions.* J. Hydrol., vol. 83, no. 3-4, pages 197–206, 1986.

[Levy 2003] M Levy and B Berkowitz. *Measurement and analysis of non-Fickian dispersion in heterogeneous porous media.* Journal of Contaminant Hydrology, vol. 64, no. 3-4, pages 203–226, 2003.

[Liu 2004] G Liu, C Zheng and S Gorelick. *Limits of applicability of the advection-dispersion model in aquifers containing connected high-conductivity channels.* Water Resources Research, vol. 40, page W08308, 2004.

[Llamas 2002] R Llamas and E Custodio. Intensively exploited aquifers main concepts, relevant facts and some suggestions. UNESCO, Paris, 2002.

[Matheron 1967] G Matheron. Elements pour une theorie des milieux poreux. Masson et Cie, Paris, 1967.

[McKenna 2001] S. McKenna, LC Meigs and R Haggerty. *Tracer tests in a fractured dolomite. 3. Double-porosity, multiple-rate mass transfer processes in convergent flow tracer tests.* Water Resour. Res., vol. 37, no. 5, pages 1143–1154, 2001.

[Milewskia 2009] A. Milewskia, M. Sultana, E. Yanb, R. Beckera, A. Abdeldayemc, F. Solimand and K. Gelil. *A Remote Sensing Solution for Estimating Runoff and Recharge in Arid Environments.* Journal of Hydrology, vol. 373, no. 1-2, pages 1–14, 2009.

[Moench 1989] A. Moench. *Convergent Radial Dispersion: A Laplace Transform Solution for Aquifer Tracer Testing.* Water Resour. Res., vol. 25, no. 3, pages 439–447, 1989.

[Monod 1949] J Monod. *The Growth of Bacterial Cultures.* Annual Review of Microbiology, vol. 3, pages 371–394, 1949.

[Moroni 2007] M Moroni, N Kleinfelter and J H Cushman. *Analysis of dispersion in porous media via matched-index particle tracking velocimetry experiments.* Advances in Water Resources, vol. 30, pages 1–15, 2007.

[Morris 1997] J P Morris, P J Fox and Y Zhu. *Modeling Low Reynolds Number Incompressible Flows Using SPH.* Journal of Computational Physics, vol. 136, no. 1, pages 214–226, 1997.

[Neuman 2005] S P Neuman. *On the tensorial nature of advective porosity.* Advances in Water Resources, vol. 28, pages 149–169, 2005.

[Nkedi-Kizza 1984] P Nkedi-Kizza, J W Biggar, H M Selim, M Th Van Genuchten, P J Wierenga, J M Davidson and D R Nielsen. *On the Equivalence of Two Conceptual Models for Describing Ion Exchange During Transport Through an Aggregated Oxisol.* Water Resources Research, vol. 20, no. 8, page 1123, 1984.

[NRC 1994] NRC. *Ground Water Recharge Using Waters of Impaired Quality.* Rapport technique, National Research Council, Committee on Ground Water Recharge, Water Science and Technology Board, Commission on Geosciences, Environment, and Resources., National Academy Press, Washington, D.C., 1994.

[NRC 1997] NRC. *Review of recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts.* Rapport technique, National Research Council, Natl. Acad. Press, Washington, D.C., 1997.

[Okin 2004] G.S. Okin and T.H Painter. *Effect of grain size on remotely sensed spectral reflectance of sandy desert surfaces.* Remote Sensing of Environment, vol. 89, pages 272–280, 2004.

[Okubo 1979] T Okubo and J Matsumoto. *Effect of infiltration rate on biological clogging and water quality changes during artificial recharge.* Water Resources Research, vol. 15, no. 6, page 1536, 1979.

[Olsthoorn 1982] T.N. Olsthoorn. *The clogging of recharge wells, main subjects.* Rapport technique, Working group on recharge wells. Riswijk, The Netherlands. 136 pp., 1982.

[Orr 2005] S. Orr and A. M. Meystel. *Approaches to optimal aquifer management and intelligent control in a multiresolutional decision support system.* Hydrogeol. J., vol. 13, no. 1, pages 223–246, 2005.

[O'Shea 1981] C.M.J. O'Shea, K. M. Baxter and A. N. Charalambous. *The hydrogeology of the Enfield-Haringey artificial recharge scheme, north London.* J. Hydrol., vol. 53, no. 1, pages 185–211, 1981.

[Park 1990] U Park and J S Marron. *Comparison of Data-Driven Bandwidth Selectors.* Journal of the American Statistical Association, vol. 85, no. 409, pages 66–72, 1990.

[Parker 1984] J C Parker and M Th Van Genuchten. *Flux-averaged and volume-averaged concentrations in continuum approaches to solute tranport.* Water Resources Research, vol. 20, no. 7, pages 866–872, 1984.

[Parzen 1962] E. Parzen. *On estimation of a probability density function and mode.* Ann. Math. Statist., vol. 33, pages 1065–1076, 1962.

[Perez-Paricio 2000] A. Perez-Paricio. *Integrated Model of Clogging Processes in Artificial Groundwater Recharge.* PhD thesis, Department of Geotechnical Engineering and Geosciences, Universitat Politecnica de Catalunya, BarcelonaTech, 2000.

[Pokrajac 2002] D Pokrajac and R Lazic. *An efficient algorithm for high accuracy particle tracking in finite elements.* Advances in Water Resources, vol. 25, pages 353–369, 2002.

[Pollock 1988] D. W. Pollock. *Semianalytical Computation of Path Lines for Finite-Difference Models.* Ground Water, vol. 26, pages 743–750, 1988.

[Price 1980] J C Price. *The potential of remotely sensed thermal infrared data to infer surface soil moisture and evaporation.* Water Resources Research, vol. 16, no. 4, pages 787–795, 1980.

[Prickett 1981] T A Prickett, T G Naymik and C G Longquist. *A random walk solute transport model for selected groundwater quality evaluations.* Rapport technique, Illinois State Water Survey. Bulletin, vol. 65. 103 pp., 1981.

[Ptak 1996] T Ptak and G Schmid. *Dual-tracer transport experiments in a physically and chemically heterogeneous aquifer material: effective transport parameters and spatial variability.* Journal of Hydrology, vol. 183, no. 1-2, pages 117–138, 1996.

[Ptak 2004] T Ptak, M Piepenbrink and E Martac. *Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport - a review of some recent developments.* Journal of Hydrology, vol. 294, pages 122–163, 2004.

[Rao 1983] N. H. Rao and P. B. S. Sarma. *Recharge to finite aquifer from strip basins.* J. Hydrol., vol. 66, pages 245–256, 1983.

[Reddy 1989] C.S.S. Reddy, D.J. Campagna and D.W. Levandowski. *Digital Image Processing Of Multitemporal Landsat Data And Its Applications In Ground Water Exploration.* In Geoscience and Remote Sensing Symposium,. IGARSS'89. 12th Canadian Symposium on Remote Sensing. Vol. 2. Pag. 544-547, 1989.

[Remy 2009] N. Remy, A. Boucher and J. Wu. *Applied Geostatistics with SGeMS. A User's Guide*, new york cambridge university press. édition, 2009.

[Riva 2008] Monica Riva, Alberto Guadagnini, D Fernandez-Garcia, Xavier Sanchez-vila and Thomas Ptak. *Relative importance of geostatistical and transport models in describing heavily tailed breakthrough curves at the Lauswiesen site.* Journal of Contaminant Hydrology, vol. 101, pages 1–13, 2008.

[Rosenblatt 1956] Murray Rosenblatt. *Remarks on Some Nonparametric Estimates of a Density Function.* The Annals of Mathematical Statistics, vol. 27, no. 3, pages pp. 832–837, 1956.

[Rosowski 1986] J. R. Rosowski, K. D. Hoagland and J. E. Aloi. *Structural morphology of diatom-dominated stream biofil communities under the impact of soil erosion.* In EGU General Assembly 2010, editeur, L. V. Evans and K. D Hoagland (Eds.). Algal Biofouling, Elsevier, Amsterdam (The Netherlands). Pp. 247-297, volume 12 of *EGU2010-5326*, 1986.

[Rubin 1997] Y. Rubin, M A Cushey and A Wilson. *The moments of the breakthrough curves of instantaneously and kinetically sorbing solutes in heterogeneous geologic media: Prediction and parameter inference from field measurements.* Water Resources Research, vol. 33, no. 11, pages 2465–2481, 1997.

[Rubin 2003] Y Rubin. Applied stochastic hydrogeology. Oxford University Press, USA, 2003.

[Salamon 2006a] P. Salamon, D. Fernàndez-Garcia and J. Gomez-Hernandez. *Modeling mass transfer processes using random walk particle tracking.* Water Resour. Res., vol. 42, page W11417, 2006.

[Salamon 2006b] P. Salamon, D. Fernàndez-Garcia and J. Gomez-Hernandez. *A review and numerical assessment of the randomwalk particle tracking method.* Journal of Contaminant Hydrology, vol. 87, no. 1-3, pages 277–305, 2006.

[Salamon 2007] P. Salamon, D. Fernàndez-Garcia and J. Gomez-Hernandez. *Modeling tracer transport at the MADE site: The importance of heterogeneity.* Water Resources Research, vol. 43, page W08404, 2007.

[Sanchez-Vila 1996] X. Sanchez-Vila, J. Carrera and JP. Girardi. *Scale effects in transmissivity.* Journal of Hydrology, vol. 183, no. 1-2, pages 1–22, 1996.

[Sanchez-Vila 1997] X. Sanchez-Vila and J. Carrera. *Directional effects on convergent flow tracer tests.* Mathematical Geology, vol. 29, no. 4, pages 551–569, 1997.

[Sanchez-Vila 2004] X. Sanchez-Vila and J Carrera. *On the striking similarity between the moments of breakthrough curves for a heterogeneous medium and a homogeneous medium with a matrix diffusion term.* Journal of hydrology, vol. 294, no. 1-3, page 164/175, 2004.

[Saraf 1998] A. K. Saraf and Choudhury P.R. *Integrated remote sensing and GIS for groundwater exploration and identification of artificial recharge sites.* Int. J. Remote Sensing, vol. 19, no. 10, pages 1825–1841, 1998.

[Scanlon 2006] B. R. Scanlon, K. E. Keese, A. L. Flint, L. E. Flint, G. B. Gaye, W. M. Edmunds and I. Simmers. *Global synthesis of groundwater recharge in semiarid and arid regions.* Hydrol. Processes, vol. 20, pages 3335–3370, 2006.

[Shapiro 1988] A M Shapiro and V Cvetkovic. *Stochastic analysis of solute arrival time in heterogeneous porous media.* Water Resources Research, vol. 24, no. 10, pages 1711–1718, 1988.

[Silverman 1986] B. W. Silverman. Density estimation for statistics and data analysis. Chapman & Hall CRC, 1986.

[Smith 1972] R.E. Smith. *The infiltration envelope: results from a theoretical infiltrometer.* Journal of Hydrology, vol. 17, pages 1–21, 1972.

[Stonestrom 2007] D.A. Stonestrom, J. Constantz, T.P.A. Ferre and S.A. Leake. *Ground-water recharge in the arid and semiarid southwestern United States.* Rapport technique, BUS Geological Survey Professional Paper 1703, 2007.

[Tartakovsky 2006] A M Tartakovsky and P Meakin. *Pore scale modeling of immiscible and miscible fluid flows using smoothed particle hydrodynamics.* Advances in Water Resources, vol. 29, no. 10, pages 1464–1478, 2006.

[Tartakovsky 2007] D. M. Tartakovsky. *Probabilistic risk analysis in subsurface hydrology.* Geophys. Res. Lett., vol. 34, pages L05404, doi:10.1029/2007GL029245, 2007.

[Tartakovsky 2008] D. M. Tartakovsky and C. L. Winter. *Uncertain future of hydrogeology.* ASCE J. 394 Hydrologic Engrg., vol. 13, no. 1, pages 37–39, 2008.

[Tien 1979] C. Tien and A. C. Payatakes. *Advances in Deep Bed Filtration.* AIChE J., vol. 25, no. 5, pages 737–759, 1979.

[Ting 2002] C.S. Ting, J.M. Tsai and M.K. Chien. *Pilot Study for Artificial Recharge of Groundwater Using High-infiltration Basins.* In 5th Conference on Groundwater Resource and Water Quality Protection:G45-50 (in English with Chinese abstract), 2002.

[Tompson 1990] A F B Tompson and L W Gelhar. *Numerical-Simulation of Solute Transport in 3-Dimensional, Randomly Heterogeneous Porous-Media.* Water Resources Research, vol. 26, pages 2541–2562, 1990.

[Trinchero 2008] P. Trinchero, X. Sanchez-Vila and D. Fernàndez-Garcia. *Point-to-point connectivity, an abstract concept or a key issue for risk assessment studies?* Advances in Water Resources, vol. 31, no. 12, pages 1742–1753, 2008.

[Tufenkji 2004] N. Tufenkji and M. Elimelech. *Correlation equation for predicting single-collector efficiency in physicochemical filtration in saturated porous media.* Environ. Sci. Technol., vol. 38, no. 2, pages 529–536, 2004.

[Tuinhof 2003] A. Tuinhof and J. P. Heederik, editeurs. Management of aquifer recharge and subsurface storage. Netherlands National Committee - International Association of Hydrogeology, No. 4. NNCIAH Publication, 2003.

[Valocchi 1985] A J Valocchi. *Validity of the local equilbrium assumption for modeling sorbing solute trasport through homogeneous soils.* Water Resources Research, vol. 21, no. 6, page 808, 1985.

[Van Genuchten 1976] M Th Van Genuchten and P J Wierenga. *Mass Transfer Studies in Sorbing Porous Media I. Analytical Solutions.* Soil Sci. Soc. Am. J., vol. 40, no. 4, pages 473–480, 1976.

[van Genuchten 1980] M. T. van Genuchten. *A closed-form equation for predicting hydraulic conductivity of unsaturated soils.* Soil Sci. Soc. Am. J., vol. 44, pages 892–898, 1980.

[Vandevivere 1992] P. Vandevivere and P. Baveye. *Saturated Hydraulic Conductivity Reduction Caused by Aerobic Bacteria in Sand Columns.* Soil Science Society of America Journal, vol. 56, no. 1, pages 1–13, 1992.

[Vandevivere 1995] P. Vandevivere, P. Baveye, D. S. de Lozada and P. DeLeo. *Microbial Clogging of Saturated Soils and Aquifer Materials: Evaluation of Mathematical Models.* Water Resour. Res., vol. 31, no. 9, pages 2173–2180, 1995.

[Vukovic 1992] M. Vukovic and A. Soro, editeurs. Determination of hydraulic conductivity of porous media from grain-size composition. Water Resources Publications, Littleton, CO, 1992.

[Wagner 1987] B. Wagner and S. Gorelick. *Optimal Groundwater Quality Management Under Parameter Uncertainty.* Water Resour. Res., vol. 23, no. 7, pages 1162–1174, 1987.

[Welty 1994] C. Welty and LW Gelhar. *Evaluation of longitudinal dispersivity from nonuniform flow tracer tests.* Journal of Hydrology, vol. 153, no. 1, pages 71–102, 1994.

[Willmann 2008] M. Willmann, J. Carrera and X. Sanchez-Vila. *Transport upscaling in heterogeneous aquifers: What physical parameters control memory functions?* Water Resour. Res., vol. 44, page W12437, 2008.

[Winter 2008] C. L. Winter and D. M. Tartakovsky. *A reduced complexity model for probabilistic risk assessment of groundwater contamination.* Water Resour. Res., vol. 44, no. 1, page W06501, 2008.

[Zamani 2009] A. Zamani and B. Maini. *Flow of dispersed particles through porous media – Deep bed filtration.* J. Petrol. Sci. Eng., vol. 69, pages 71–88, 2009.

[Zenios 1998] S. A. Zenios, R. M. Holmer, R. McKendall and C. Vassiadou-Zeniou. *Dynamic models for fixed-income portfolio management under uncertainty.* J. Econ. Dynam. Control, vol. 22, no. 10, pages 1517–1544, 1998.

[Zimmerman 1998] D A et al Zimmerman. *A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow.* Water Resources Research, vol. 34, no. 6, pages 1373–1413, 1998.

[Zinn 2003] B Zinn and C F Harvey. *When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields.* Water Resources Research, vol. 39, no. 3, page 1051, 2003.

[Zwietering 1990] M. H. Zwietering, I. Jongenburger, F. M. Rombouts and K. Van T Riet. *Modeling of the Bacterial Growth Curve.* Applied and Environmental Microbiology, vol. 56, no. 6, pages 1875–1881, 1990.

Appendix A

Derivation of Equation (7.6) in Chapter 7

We explain a possible derivation for (7.6) in Chapter 8, which can be obtained following [Haggerty 2000]. The sink-source term in (7.5) can be written as

$$\Gamma(x,t) = \int_0^t g(\tau)\frac{\partial C_m(t-\tau)}{\partial \tau}\mathrm{d}\tau \tag{A.1}$$

where $g(\tau)$ is the memory function ([Haggerty 1995, Carrera 1998]). (A.1) can also be written as

$$\Gamma(x,t) = c * \frac{\partial g(\tau)}{\partial \tau} + g_0 C_m(t-\tau) + C_{m0} g(\tau) \tag{A.2}$$

where the subscript 0 indicate the initial amounts fore each variable and '* is the convolution operator. Let's assume that $C_{m0} = C_{im0} = 0$, and that an injection of a unit mass takes place as an instantaneous pulse. [Haggerty 2000] show that at late times, the sink-source term becomes

$$\Gamma(x,t) \approx m_0 \frac{\partial g(\tau)}{\partial \tau} \tag{A.3}$$

where m_0 is the zeroth moment of the mass. We obtain

$$\frac{\partial C_m}{\partial t} + \beta \frac{\partial g(\tau)}{\partial \tau} = L(C_m) \tag{A.4}$$

where $L(C_m)$ is mechanical advective-dispersive operator. At late time, it can be considered that the dispersive flux is negligible with regards with the advective flux, i.e.

$$L(C_m) \approx v_m(r)\frac{\partial C_m}{\partial r} \tag{A.5}$$

and also that the derivative of the memory function is equal to the derivative of the mobile concentration. It results that

$$(1+\beta)\frac{\partial C_m}{\partial t} = v_m(r)\frac{\partial C_m}{\partial r} \tag{A.6}$$

Let's now move from a Eulerian to a Lagrangian picture of the solute transport. Using the following relationships between velocity and travel times

$$t_m = \int \frac{\mathrm{d}r}{v_m(r)} \qquad t_{adv} = \int \frac{\mathrm{d}r}{v_a(r)} \tag{A.7}$$

where t_{adv} is also equal to the first temporal moment μ_t^1, and $v_a(r) = \partial r/\partial t$ is the advective velocity. Then (A.6) can be rewritten as

$$\mu_t^1 = t_m(1+\beta) \tag{A.8}$$

or as Eq. (7.6) in Chapter 7.