A Comparison of Six Transport Models of the MADE-1 Experiment implemented with Different Types of Hydraulic Data

Alraune Zech¹, Sabine Attinger², Alberto Bellin³, Vladimir Cvetkovic⁴, Gedeon Dagan⁵, Marco Dentz⁶, Peter Dietrich⁷, Aldo Fiori⁸, and Georg Teutsch⁹

¹Utrecht University
²UFZ Centre for Environmental Research
³Università di Trento
⁴KTH (Royal Institute of Technology)
⁵Tel Aviv University
⁶IDAEA-CSIC
⁷UFZ, Helmholtz Centre for Environmental Research
⁸Università degli Studi Roma Tre
⁹UFZ

November 21, 2022

Abstract

Six conceptually different models of steady groundwater flow and conservative transport are applied to the heterogeneous MADE aquifer. Their predictive capability is assessed by comparing the modelled and observed longitudinal mass distributions at different times of the plume in the MADE-1 experiment, as well as at a later time. The models differ in their conceptualization of the heterogeneous aquifer structure, computational complexity, and use of permeability data obtained from various observation methods (DPIL, Grain Size Analysis, Pumping Tests and Flowmeter). Models depend solely on aquifer structural and flow data, without calibration by transport observations. Comparison of model results by various measures, i.e. peak location, bulk mass and leading tail, reveals that the predictions of the solute plume agree reasonably well with observations if the models are underlined by a few parameters of close values: mean velocity, a parameter reflecting log-conductivity variability and a horizontal length scale related to conductivity spatial correlation. From practitioners perspective the robustness of the models is an important and useful property. The model comparison provides insight into relevant features of transport in heterogeneous aquifers. After further validation by additional field experiments or by numerical simulations, the results can be used to provide guidelines for users in selecting conceptual aquifer models, characterization strategies, quantitative models and implementation for particular goals.
A Comparison of Six Transport Models of the MADE-1 Experiment implemented with Different Types of Hydraulic Data

Alraune Zech\textsuperscript{1,2}, Sabine Attinger\textsuperscript{2,3}, Alberto Bellin\textsuperscript{4}, Vladimir Cvetkovic\textsuperscript{5}, Gedeon Dagan\textsuperscript{6}, Marco Dentz\textsuperscript{7}, Peter Dietrich\textsuperscript{2,8}, Aldo Fiori\textsuperscript{9}, Georg Teutsch\textsuperscript{2}

\textsuperscript{1}Department of Earth Science, Utrecht University, The Netherlands
\textsuperscript{2}Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany
\textsuperscript{3}Institute of Earth and Environmental Science-Geoecology, University Potsdam, Germany
\textsuperscript{4}Department of Civil, Environmental and Mechanical Engineering, University of Trento, Italy
\textsuperscript{5}Department of Water Resources Engineering, Royal Institute of Technology, Stockholm, Sweden
\textsuperscript{6}School of Mechanical Engineering, Tel Aviv University, Ramat Aviv, Israel
\textsuperscript{7}Institute of Environmental Assessment and Water Research (IDAÉA), Spanish National Research Council, Barcelona, Spain
\textsuperscript{8}Center of Applied Geoscience, University of Tübingen, Tübingen, Germany
\textsuperscript{9}Department of Engineering, Roma Tre University, Rome, Italy

Key Points:

- Predictions of transport in highly heterogeneous aquifer are achieved with six models of different conceptualization and field data input
- Models' predictions are reliable if they share similar values of mean velocity and permeability degree of variability and correlation scale
- The MADE site plume longitudinal mass distribution is a robust transport measure

Corresponding author: Alraune Zech, a.zech@uu.nl
Abstract

Six conceptually different models of steady groundwater flow and conservative transport are applied to the heterogeneous MADE aquifer. Their predictive capability is assessed by comparing the modelled and observed longitudinal mass distributions at different times of the plume in the MADE-1 experiment, as well as at a later time. The models differ in their conceptualization of the heterogeneous aquifer structure, computational complexity, and use of permeability data obtained from various observation methods (DPIL, Grain Size Analysis, Pumping Tests and Flowmeter). Models depend solely on aquifer structural and flow data, without calibration by transport observations. Comparison of model results by various measures, i.e. peak location, bulk mass and leading tail, reveals that the predictions of the solute plume agree reasonably well with observations if the models are underlined by a few parameters of close values: mean velocity, a parameter reflecting log-conductivity variability and a horizontal length scale related to conductivity spatial correlation. From practitioners perspective the robustness of the models is an important and useful property. The model comparison provides insight into relevant features of transport in heterogeneous aquifers. After further validation by additional field experiments or by numerical simulations, the results can be used to provide guidelines for users in selecting conceptual aquifer models, characterization strategies, quantitative models and implementation for particular goals.

1 Introduction

Modelling contaminant transport by groundwater is a topic of great interest that stimulated intensive research in the last four decades due to its relevance to aquifer pollution (Dagan, 1989; Gelhar, 1993; Fetter et al., 2018). The task of predicting transport faces a few difficulties: the processes are of long duration, measurements are scarce, the subsurface medium is of complex heterogeneous structure subjected to uncertainty and many times the geometry and the mass content of the contaminant source is also not known with certainty.

Under these circumstances models play an important role: they help understanding the involved processes, analysing field data and making long range prediction. Models developed in the past differ in conceptualization of the aquifer structure, in the required data, in quantification of transport, in the formulation of the governing equations and mechanisms they represent, in computational complexity and in models goals.

We focus on transport of plumes of conservative solutes in steady natural gradient flow, driven by a constant mean head gradient. Quantification of the spatial distribution is by $m(x,t)$, mass per unit length, where $x$ is the mean flow direction and $t$ the time. It encapsulates the process of longitudinal spreading in space and time. In practice it allows, for instance, to estimate the mass of solute pumped by wells or flowing into rivers or reservoirs. It also serves as a first step toward achieving other goals like determining the local concentration $C(x,t)$.

There is general agreement that spatial variability of the hydraulic conductivity $K(x)$ is the main growth mechanism responsible for plume spreading in aquifers, termed macrodispersion (Zech et al., 2015). The effect increases with higher $K$ heterogeneity as quantified for instance by the log-conductivity variance $\sigma_Y^2$.

A few elaborate field experiments were conducted in the past in order to gain understanding and validate models. The most challenging one is at the MADE site (Boggs et al., 1992), situated in a highly heterogeneous aquifer, making it of relevance to many actual aquifers (Gomez-Hernandez et al., 2017). An important feature of MADE was the application of different observation methods in order to characterize the $K$ spatial distribution (e.g. Boggs et al. (1990); Rehfeldt et al. (1992); Bohling et al. (2016)). Long-term tracer tests provide graphs of observed mass distribution $m(x,t)$ in space at a few
fixed \( t \) values (Boggs et al., 1990, 1995). It has motivated a flurry of works on structure characterization by field data and different modelling strategies (see e.g. Zheng et al. (2011)). We will elaborate on \textsc{made} in section 2.

We examine the ability of six conceptually different models to predict the observed mass \( m \). Subsequently, we examine the predictive power of the models by extending the time \( (t = 1000 \text{ days}) \) beyond \textsc{made} observations \( (t_{\text{max}} = 503 \text{ days}) \). The selected models, outlined in Section 3, differ in conceptualizations of formation structure and transport, in the use of field data, in initial conditions, in computational methodology and in effort. A few of the models were developed in the past while the others were formulated for use in the present paper. The models cover a wide spectrum of configurations. We concentrate here only on models that can predict transport based on field data of aquifer properties and flow; we do not consider models calibrated on prolonged transport tests which are generally of large duration and cost.

We believe that the comparison of the models which differ in type and underlying field data, conceptualization of \( K \) spatial structure and complexity of computations by using \textsc{made} as a platform is important in helping the research community to grasp transport issues and the users in selecting characterization strategies, goals of models and method implementation.

The plan of the paper is as follows: Sect. 2 recapitulates the \textsc{made} aquifer characterization and transport experiment; Sect. 3 describes the methodology of the models, including their application to \textsc{made}; Sect. 4 is devoted to the model prediction for solute mass \( m \) in comparison to \textsc{made} observations as well as time beyond. Sect. 5 contains the general discussion on data comparison while Sect. 6 concludes the paper.

2 MADE Transport Field Experiment

The \textsc{made} experiment was the object of a large body of publications dealing with the aquifer properties data collection and analysis as well as transport observations and interpretation (see for instance reviews Zheng et al. (2011); Gomez-Hernandez et al. (2017)). We recall in the following only those aspects of direct relevance to the examined models.

2.1 Hydraulic Conductivity Spatial Distribution

The \textsc{made} site aquifer is composed of highly heterogeneous alluvial terrace deposits. Measurements of hydraulic conductivity at multiple locations (see Fig. 1) were performed by granulometry of soil samples, flow meter, slug test, and Direct Push Injection Logger (DPIL) (Boggs et al., 1990; Rehfeldt et al., 1992; Bohling et al., 2016). Besides, two pumping tests provide equivalent conductivities \( K_{\text{eq}} \) of the volume surrounding the wells (Boggs et al., 1992). The use of different techniques at same site and subsequent application by different models offers an unique opportunity to examine their impact on transport prediction.

The spatially distributed observations (Fig. 1), carried out at different depths resulted in a large volume of data which served for geostatistical analysis. Fig. 2 summarizes their outcomes, which is of interest for the application of the different models (Sect. 3). The most reliable and extensively used data are those based on flowmeter (Boggs et al., 1990; Rehfeldt et al., 1989) with a total number of \( N = 2611 \) observations and more recently DPIL with \( N = 31123 \) (Bohling et al., 2012, 2016).

The differences in the geostatistical parameters, especially in the geometric mean \( K_G \), are a result of the different properties of the observation methods as well as the density and locations of observation points. Particularly, the difference between flowmeter and DPIL can be explained by the inability of flowmeter to detect low \( K \) values. While
Figure 1. Left: Map of MADE site according to Boggs et al. (1990); Bohling et al. (2016): Locations of hydraulic conductivity measurements devices (coloured dots); tracer test source area (black outline); and sampling network boundary (dashed outline) where bromide samples were collected. Right: Potentiometric surface map of head measurements according to Boggs et al. (1990). Black dot marks tracer test injection location.

the maximal values $K_{\text{max}}$ are close (around 0.005 m/s) the minimal values $K_{\text{min}}$ differ by two orders of magnitude. The effect manifested in Fig. 2 is that flowmeter $K_{G}$ is larger and $\sigma_{Y}^{2}$ is smaller. The difference in the longitudinal integral scales $I_h$ suggests that the zones of low $K$ values are less connected than those of higher magnitude. The analysis based on soil samples ($N = 214$) is less reliable. Still, it provided input data for conductivity conceptualization based on the lithofacies approach (e.g. Carle & Fogg (1996)) as presented by Bianchi & Zheng (2016) and herein for a simplified binary structure approach. The impact of these difference upon flow and transport are discussed in Sect. 4.

2.2 Flow

In Fig. 1 we reproduce the head contour lines map (Boggs et al., 1992, Fig. 3). The head gradient is not constant, but slowly varying in space. The mean head gradient is between $J \in [0.003, 0.0036]$ depending on the choice of boundary locations. The non-uniformity of the head contour density indicates the presence of large scale mean hydraulic conductivity trends.

2.3 Transport Experiment

We focus on the first tracer transport experiment, which was conducted in years 1986–1988 (Boggs et al., 1990, 1992; Rehfeldt et al., 1992; Adams & Gelhar, 1992). The tracer plume displayed a non-Gaussian longitudinal solute mass distribution with the bulk of the mass staying near the source, but with lower amounts spreading downgradient extensively.
Figure 2. Geostatistical measures for MADE from DPIL (direct push injection logging) (Bohling et al., 2016), flowmeter, grain size analysis, slug tests (Rehfeldt et al., 1992): log-conductivity variance $\sigma_{ln K}^2$, horizontal and vertical integral scale (correlation length) $I_{h, v}$, respectively. Visualization of geometric mean conductivity $K_G$, range of observed values from minimal to maximal ($[K_{min}, K_{max}]$), and range of one variance around mean ($[K_G \cdot e^{-\sigma^2}, K_G \cdot e^{\sigma^2}]$). $K_{eq}$ denotes the equivalent conductivity for two large scale pumping tests (Boggs et al., 1990).

Initial Conditions: A quantity of $M_0 \approx 25$ kg of Bromide dissolved in $10 \text{ m}^3$ of water was injected over a period of 48 hours through 5 wells of half a meter screen length. While the subsequent transport took place under practically mean uniform flow conditions, the tracer solution was forced radially into the aquifer during injection. As a result, the initial tracer body was much larger than the body adjacent to the wells screens, which can be seen in the early tracer plume snapshot at 9 days after injection (Adams & Gelhar, 1992, Fig. 5). This is important as far as ergodicity and setup of initial conditions are concerned. The apparent upstream tracer spread cannot be interpreted as a result of upstream dispersion. The injection mode also implies that the initial condition was flux proportional, with a preference of mass flowing in high conductivity channels.

Plume Detection and Data Aggregation (Upscaling): We reproduce the longitudinal mass distribution $\bar{m}(x, t)$ of Adams & Gelhar (1992, Fig. 7) at six times $T$: 49, 126, 202, 279, 370, and 503 days after beginning of injection. The computation of $\bar{m}$ is based on concentration $C(x, y, z, t)$ sampled in a dense MLS network, which thins out with distance to the source (Fig. 1). Subsequently, $C$ was numerically integrated over transverse planes $(y, z)$, accumulated and averaged over slices of 10 m length in the $x$ direction to obtain the upscaled longitudinal mass $\bar{m}(x, t)$ (Adams & Gelhar, 1992). The reported mass is displayed at the centers of the slices at $x = -5 \text{ m}$, 5 m, 15 m, . . . (see discussion of difference between the fine scale $m$ and the upscaled one $\bar{m}$ in the sequel).

The fact that the reported mass is an upscaled/aggregated quantity, was overlooked by many previous studies which compared $\bar{m}$ with modeled mass at fine scale, as mentioned by Fiori et al. (2019). The significance of data aggregation is discussed in section 4 herein.

Mass Recovery The reported mass $\bar{m}$ does not obey the mass conservation requirement $\int \bar{m}(x, t) \, dx = 1$ except at 126 days after injection. Mass apparently decreases over time after 126 days with recovery rates of 2.06, 0.99, 0.68, 0.62, 0.54, and 0.43, for the $T = 49, 126, 202, 279, 370$, and 503 days, respectively.

As discussed in Adams & Gelhar (1992), the excessive mass recovery at $t = 49$ days could be due to spurious hydraulic connections among the multilevel samplers. This is a result of the method of installation and is enhanced by the pressure of injection. Pref-
3 Subsurface Transport Models

3.1 Common Features and Prerequisites

We compare various models which were developed independently by the authors in the past or devised recently in the frame of the present study. Before describing the models specific properties, we recapitulate a few common features:

1. We examine only predictive models. They rely solely on structural and flow data which can be measured independently of transport. Models with parameters calibrated by transport experiments are not considered.

2. Flow is steady and uniform in the mean (natural gradient), driven by the steady head gradient $J$ in the $x$ direction.

3. Transport is advective and spreading is caused by the spatial variability of $K(x)$. The effective porosity $\theta$ is assumed to be constant.

4. A plume of mass $M_0$ of a conservative solute is injected initially in the aquifer at $t = 0$. Spreading is quantified by the mass arrival at a control plane at $x$: $M_{\text{Total}}(x,t) = \theta \int_{-\infty}^{x} \int \int C(x',y,z,t) dx'dydz$, where $C$ is the concentration. For a fixed $x$, $M = M_{\text{Total}}/M_0$ is the BTC whereas for a fixed $t$ it is the relative mass accumulated beyond $x$.

5. We determine the fine scale relative mass spatial distribution $m(x,t) = -\partial M/\partial x$ at a few times $t$. However, in line with MADE observations, we calculate the (up-scaled) relative mass averaged over a medium slice of length $\Delta = 10 \text{ m}$ centered at $x$, which is given by $\bar{m}(x,t; \Delta) = (1/\Delta)\int (x-\Delta/2, t) - M(x+\Delta/2, t)$ such that $m = \lim_{\Delta \searrow 0} \bar{m}$.

6. The up-scaled relative mass $\bar{m}$ is derived for the MADE conditions at $t = 49, 126, 202, 279, 370,$ and $503$ days after injection toward comparison with measured $\bar{m}$. Additionally, models are applied to prediction of $\bar{m}$ at $t = 1000$ days for inter-comparison. As useful additional quantification we also consider the mass flux through the control plane, $\mu(x,t) = \partial M/\partial t$.

Remark on additional MADE transport Models: Several other transport models which are not discussed here have been presented in the literature for MADE. We do not consider models calibrated on transport observations since they are not predictive, such as the work of Barlebo et al. (2004). This further includes dual-domain models (e.g. Harvey & Gorelick (2000) and Feehley et al. (2000)) and the continuous time random walk (CTRW) model of Berkowitz & Scher (2001). The lithofacies approach of Bianchi & Zheng (2016) would have been a candidate but results are only available for the MADE-2 tracer experiment setting rather than MADE-1 considered here. The same holds for the work of Salamon et al. (2007). The conceptual framework is however underlying the facies model herein. Similarly, the fractional ADE model of Benson et al. (2001) is applied only to...
It is characterized completely by a random velocity field. The latter is obtained by expanding the mass conservation equation and Darcy’s law in power series in random velocity 

\[ \sigma_Y \]

Thus, their results are unsuitable for model comparison, particularly with prediction beyond simulation results are limited to a fraction of the total plume transport distance.

We briefly recall past results and recent advances relevant to this work. Graphs Dagan (1989); Gelhar (1993); Rubin (2003) leading to various analytical solutions. They generated detailed representations of the MADE site aquifer and subsequently solved the flow and transport equations. However, due to the computational effort, transport simulations results are limited to a fraction of the total plume transport distance.

3.2 Brief Description

3.2.1 First Order Approximation (FOA)

Background The solution of flow and transport in heterogeneous formations of a random \( Y = \ln K \) structure by a first order approximation in the log-conductivity variance \( \sigma_Y^2 \) was the topic of intensive research in the last four decades (see e.g. the monographs Dagan (1989); Gelhar (1993); Rubin (2003)) leading to various analytical solutions. We briefly recall past results and recent advances relevant to this work.

K-Structure The random \( Y \) field is regarded as stationary and multi-Gaussian. It is characterized completely by \( K_G \), \( \sigma_Y^2 \), and the two point covariance \( C_Y \) of horizontal integral scale \( I \) and vertical one \( I_v \). The anisotropy ratio \( \epsilon = I_v/I \) is generally smaller than unity.

Flow The mean velocity is given by \( U = K_{eff}/\theta \), with the effective conductivity \( K_{eff}/K_G = \text{func}(\sigma_Y^2, \epsilon) \) determined from the solution of the flow equations for the random velocity field. The latter is obtained by expanding the mass conservation equation and Darcy’s law in power series in \( Y' = Y - \langle Y \rangle \).

Transport Traditionally, the mean relative mass distribution \( \langle m(x, t) \rangle \) was derived for conditions of given initial deterministic resident concentrations \( C_0(x, 0) \) either in a volume \( V_0 \), with \( C_0 = M_0/(\theta V_0) \), or with mass concentrated on the plane \( x = 0 \) over an area \( A_0 \) and quantified by \( m_0 = M_0/A_0 \) (Kreft & Zuber, 1978). Similarly, detection was in the resident mode with the mean \( \langle M \rangle \) and \( \langle m \rangle \) satisfying the ADE

\[
\frac{\partial(M)}{\partial t} + U \frac{\partial(M)}{\partial x} = D_L(t) \frac{\partial^2(M)}{\partial x^2} \tag{1}
\]

The macrodispersion coefficient \( D_L = U \alpha_L \), with longitudinal macrodispersivity \( \alpha_L \), was determined in the Lagrangean framework with the aid of the solute particles trajectories. At first order \( \alpha_L(t) \) grows from zero at \( t = 0 \) to an asymptotic constant value \( \alpha_L = \sigma_Y^2 I \) (Dagan, 1989). The transient \( \alpha_L(t) \) was determined by a quadrature for an exponential covariance and is approximated accurately by an analytical expression as a function of mean flow velocity, time and aquifer statistics: \( \alpha_L(t)/\sigma_Y^2 I = \text{func}(U t/I, \epsilon) \) (Dagan & Cvetkovic, 1993, Eq. 20). The asymptotic value is attained after a travel distance of a few integral scales \( I \). The Gaussian solution approximates \( \langle m \rangle \) satisfactorily for field experiments in weakly heterogeneous aquifers like Cape Cod (Hess et al., 1992) and Borden Site (Sudicky, 1986) and/or far from the injection zone. However, it failed to model the highly skewed mass distribution observed at MADE close to the injection zone. This finding has motivated development of new nonlinear models.

Fiori et al. (2017) presented the solution of the ADE (1) for the more realistic conditions of flux proportional injection. The solution for the BTC is given by the inverse Gaussian distribution:

\[
\langle M \rangle = \frac{1}{2} \left\{ \text{erfc} \left( \frac{x - U t}{2 \sqrt{D_L t}} \right) + \exp \left( \frac{U x}{D_L} \right) \text{erfc} \left( \frac{x + U t}{2 \sqrt{D_L t}} \right) \right\} \tag{2}
\]
Unlike the Gaussian solution of Eq. (1), \( \langle m(x,t) \rangle = -\partial(M)/\partial x \) from Eq. (2) displays a skewed shape and lack of upstream dispersion. Jankovic et al. (2017) showed that the solution (2) with dispersion by first-order approximation \( D_L(x) \) represents accurately the results of numerical simulations even for the large value of \( \sigma_Y^2 \) = 8 except an under-estimation of a few percents of the mass in the tail of late arrival. We apply Eq. (2) to predict the MADE plume in the sequel as part of models comparison.

**Application to MADE**  Application of Eq. (1) to MADE was based on the parameters derived from Bohling et al. (2016) by DPIL as follows: \( K_G = 0.58 \text{m/d}, \theta = 0.31 \), \( I = 9.1 \text{m}, I_v = 1.8 \text{m}, \sigma_Y^2 = 5.9 \). Rather than the first order approximation we used the more general formula by Zarlenza et al. (2018, Eq. 5) to arrive at \( K_{\text{eff}} = 2.28 \text{m/d} \). After identifying the representative mean head gradient in the plume zone (Fig. 1) as \( J = 0.0036 \), the mean velocity is given by \( U = K_{\text{eff}}J/\theta = 0.026 \text{m/d} \). Subsequently, the transient regime is taken into account by using a preasymptotic Dispersion \( D_L(x) = U \alpha_1(x) \), calculated according to Fiori et al. (2019, Eqs. C1,C2) based on \( \alpha_1 \) of (Dagan & Cvetkovic, 1993, Eq. 20), with the asymptotic value \( \alpha_1 = \sigma_Y^2 I = 53.7 \text{m} \). This was the information needed in order to derive \( \langle \tilde{m} \rangle \) based on Eq. (1) at the \( t \) and \( x \) values pertinent to MADE.

### 3.2.2 Multi-indicator Model and Self Consistent Approximation (MIM-SCA)

**Background** MIMSCA was developed in the last 15 years as an approximate model of flow and transport for aquifers of arbitrary degree of heterogeneity (Dagan & Fiori, 2003; Fiori et al., 2006; Cvetkovic et al., 2014). Its outcome has been compared with accurate numerical solutions for \( \sigma_Y^2 \) \( \leq \) 8 and applied to MADE. Fiori et al. (2019) recently applied the model to assess the uncertainty of prediction due to non-ergodic conditions or parametric uncertainty.

**K-Structure** The aquifer is modeled as an ensemble of rectangular inclusions tessellating the space similarly to layers of bricks. The elements are of dimension \( 2I \times 2I \), which is modelled by \( K_{\text{eff}} \). For \( M \) \( \leq \) 28 m/d. Subsequently, the transient regime is taken into account by using a preasymptotic Dispersion \( D_L(x) = U \alpha_1(x) \), calculated according to Fiori et al. (2019, Eqs. C1,C2) based on \( \alpha_1 \) of (Dagan & Cvetkovic, 1993, Eq. 20), with the asymptotic value \( \alpha_1 = \sigma_Y^2 I = 53.7 \text{m} \). This was the information needed in order to derive \( \langle \tilde{m} \rangle \) based on Eq. (1) at the \( t \) and \( x \) values pertinent to MADE.

**Flow** The mean velocity is given by \( U = K_{\text{eff}}J/\theta \). The mean effective conductivity \( K_{\text{eff}} \) is derived by the self consistent approximation (SCA), a well established method in the literature on heterogeneous aquifers (Dagan, 1989). Here, it consists in solving the flow equations for a generic inclusion of conductivity \( K \), submerged in a homogeneous matrix of the unknown \( K_{\text{eff}} \) and determining the latter by the SCA argument. \( K_{\text{eff}} \) follows as solution of a simple integral equation. Suribhatla et al. (2011) compared it with accurate numerical simulations with satisfactory agreement.

**Transport** Transport for the MIM was solved also with the SCA. Fiori et al. (2003) determined analytically the travel time required for a solute particle to move over an inclusion of conductivity \( K \), surrounded by a matrix of \( K_{\text{eff}} \). The BTC \( (M) \) follows as sum over the travel time random residuals pertaining to the inclusions of different \( K \) lying between the injection plane \( x = 0 \) and the control plane at \( x \) and the initial condition was of flux proportional injection at \( x = 0 \) while the mean BTC was derived by fast Fourier transform. Jankovic et al. (2003) showed a satisfactory agreement between the semi-analytical results and accurate 3D numerical simulations. Fiori et al. (2017) showed that the bulk of the BTC is well approximated by the FOA, while MIMSCA captures also the long tail of a few percents of the solute mass observed in numerical simulations.
Fiori et al. (2013) applied the method to the \textit{MADE} site transport setting for $\langle m \rangle$ and for $\langle \bar{m} \rangle$, with an update in Fiori et al. (2019) motivated by an update in geostatistical input parameters by Bohling et al. (2016). The needed parameters values are the same as those given above for FOA.

### 3.2.3 TDRW

\textbf{Background} Time-domain random walk and continuous time random walk approaches have been used extensively over the past two decades for the modeling of transport in heterogeneous porous media (Noetinger et al., 2016). Within this framework and based on the stochastic TDRW method of Comolli et al. (2019), Dentz et al. (2020) derived a predictive upscaled model that avoids calibration by transport observations. The basic idea is to quantify particle motion in spatially variable flow fields through a Markov processes for equidistant particle velocities, whose steady state distribution is given by the flux-weighted distribution of flow velocities. While details can be found in Dentz et al. (2020), we describe here the main features of the model. This modeling approach has been used and verified for the prediction of the evolution of particle velocity statistics, particle distributions, dispersion and breakthrough curves in pore and Darcy scale heterogeneous porous media (Hakoun et al., 2019; Comolli et al., 2019).

\textbf{K-Structure} Hydraulic conductivity is represented by a three-dimensional log-normally distributed multi-Gaussian spatial random field. Thus, the random $K$-field is characterized in terms of 4 parameters, similarly to the FOA (section 3.2.1): geometric mean conductivity $K_G$, ln $K$ variance, and correlation lengths $\ell_h$ and $\ell_v$. Random realizations were filtered such that the spatial mean and variance of the log hydraulic conductivity are within a 5% tolerance interval around the target values.

\textbf{Flow Model} Groundwater flow is the result of the steady state groundwater flow equation, which is solved numerically on multi-Gaussian hydraulic conductivity fields characterized by a log-normal marginal distribution. A unit head drop between inlet and outlet is considered and no-flux boundaries are specified at the horizontal domain boundaries. The target variable is the magnitude of the Eulerian velocity $v_e(x)$ (absolute value of the Eulerian velocity), which is characterized by a uni-variate distribution. It is obtained from the numerically obtained magnitude of the Darcy velocity $q(x)$ by multiplication with the magnitude $J$ of the head gradient, and geometric mean conductivity $K_G$, and division by (constant) porosity $\theta$, thus $v_e(x) = q(x)K_GJ/\theta$.

\textbf{Transport Model} Transport is modeled by a continuous time random walk. Thus, particles move at constant space increment at transition times that are obtained from the particle velocity. The plume mass distribution at a given time is equivalent to the particle distribution. The particle velocity is modeled as a stationary Markov process, whose steady state distribution $p(v)$ is given by the flux-weighted Eulerian flow velocity. The flux-weighting is due to the fact that in this framework particle velocities sample the flow velocity equidistantly along path lines. This is in contrast to isochrone sampling in classical Lagrangian frameworks, for which the steady state distribution of particle velocities is equal to the Eulerian velocity distribution (Dentz et al., 2016; Comolli et al., 2019).

The Markov process of particle velocities is modeled through an Ornstein-Uhlenbeck process for the normal scores. The normal scores are obtained by mapping the velocities first to a uniform and then to a unit Gaussian random variable. The model requires the Eulerian velocity distribution and advective tortuosity as inputs. The latter is given by the ratio of the mean Eulerian velocity and the mean Eulerian velocity component along the mean hydraulic gradient.
Application to MADE For application to MADE, the model is parameterized based on the description of experimental conditions and aquifer properties as head gradient of $J = 0.0036$ and porosity of 0.31 according to Boggs et al. (1992); Adams & Gelhar (1992). The retardation coefficient is set equal to one. The distribution of Eulerian velocity magnitude as the propagator of the upscaled transport model is derived using the geostatistical parameters of log-normal hydraulic conductivity reported by Bohling et al. (2016) (Fig. 2). The average velocity component in mean flow direction is given by $\overline{v}_1 = \frac{q}{G}J/\theta = 1.942 \times 10^{-7} \text{ m/s} = 0.0167 \text{ m/d}$. The average Eulerian velocity magnitude is $\overline{v}_e = \frac{q}{G}J/\theta = 2.234 \text{ m/s} = 0.0193 \text{ m/d}$.

A point source particle distribution is assumed. Following Boggs et al. (1992) and Fiori et al. (2013), the initial mass distribution is approximately flux-weighted. Thus, in this modeling framework, the initial distribution of particle velocities is set equal to the flux-weighted Eulerian velocity distribution (Dentz et al., 2020).

3.2.4 Binary Facies

Background An alternative approach to adopting continuous univariate $Y$ distribution consists in representing the media as an assemblage of hydrofacies. Among the geostatistical methodologies adopted for this purpose, the one based on the combined use of transition probability and Markov chain has been mostly employed since its introduction by Carle & Fogg (1996). It enjoys flexibility in handling justapositional tendencies among hydrofacies and the availability of the software T-PROGS (Carle, 1999).

In the context of the studies of MADE, Bianchi & Zheng (2016) presented an application to the MADE-2 experiment by adopting representation by 5 hydrofacies. Here we apply the methodology to the MADE-1 experiment in a more parsimonious way. We maintain the highly conductive hydrofacies and combine the remaining four, of low conductivity, into a single hydrofacies.

$\textit{K-Structure}$ The porous media is modeled by using two or more facies, each one of constant hydraulic conductivity. The spatial distribution of the facies is generated randomly based on transition probabilities, e.g. by using T-PROGS. Hydrofacies identification is usually based on granulometry analysis. Here we focus on two hydrofacies of probabilities of occurrence $p_1$ and $p_2$, respectively. Hydraulic conductivities $K_{1,2}$ are the weighted arithmetic means of the hydraulic conductivity of all samples belonging to the same granulometry class, i.e. hydrofacies. The hydraulic conductivity of the samples are calculated from the characteristic diameters $d_{10}$ and $d_{25}$ of the sediments by using a modified version of the Közeni-Carman expression proposed by Riva et al. (2010). Transition probabilities between hydrofacies are obtained by fitting a Markov model to the experimental transition probabilities (Carle & Fogg, 1996). They are expressed with the aid of the characteristic thickness and length for each facies, denoted as $L_{x,1}$, $L_{x,2}$, $L_{z,1}$, and $L_{z,2}$ respectively. Commonly, samples density is relatively low in the horizontal directions, leading to uncertain experimental transition probabilities in these directions, even after assuming isotropy in the horizontal plane.

Flow and Transport Flow and transport are solved repetitively in a Monte Carlo frame by making use of numerical solvers, Modflow 2005 in combination with particle tracking (Pollock, 2012). Mean velocity $\langle \overline{\text{U}} \rangle$ and mean relative mass distribution $\langle \overline{\text{m}} \rangle$ are obtained by ensemble averaging.

Application to MADE The binary hydrofacies model was based on the granulometry of 214 samples taken from the 38 boreholes at MADE (Boggs et al., 1990), which is a data set completely independent from those of DPIL used by the FOA and MIM-SCA4 models. The parameters values pertinent to MADE were identified as: $p_1 = 0.145$, $p_2 = 0.855$, $K_1 = 190 \text{ m/d}$, $K_2 = 1.49 \text{ m/day}$, $L_{z,1} = 0.702 \text{ m}$, $L_{z,2} = 4.15 \text{ m}$. The low borehole density prevented reliable estimate of the transition probability in the horizon-
Flow was solved numerically with Modflow 2005 in a computational domain of 300×100 ×10 m³ with grid spacing of 1 m in horizontal and 0.5 m in vertical directions for each hydrofacies realization. Constant head boundary conditions were applied with a mean head gradient of J = 0.003. The adopted porosity value was θ = 0.31. Advection transport was simulated by tracking 1000 particles, distributed according to the mass distribution measured at day 9 since the beginning of the tracer test and advected by the random velocity (in the absence of local dispersion) by means of the ModPath 6 package. The resulting ensemble mean velocity is \( U = 0.079 \text{ m/day} \) with a standard deviation of 0.0046 m/d.

### 3.2.5 Binary Inclusions

**Background** Further simplification of the previous binary facies model is achieved by representing the high conductivity zones as rectangular inclusions submerged in the low conductivity matrix. Furthermore, for the high length to thickness ratio of the inclusions it was found that flow and transport can be modeled approximately as two-dimensional. We describe here briefly the application of the model to MADE presented in Zech et al. (2020), in which the stochastic conceptualization of the binary hydraulic conductivity took large scale deterministic information into account.

**K-Structure** The aquifer is modeled as binary random structure. The two possible conductivity values \( K_1 \) and \( K_2 \) represent areas of high and low conductivity, respectively. The spatial structure is given as non-overlapping conductivity blocks of length \( I \) and width \( I_v \). Site specific topological features, are integrated as deterministic structures, such as layers or blocks of different average conductivity. The MADE-specific K-structure model is outlined below.

**Flow and Transport** Flow and transport is solved for every random \( K \) realization numerically making use of Darcy’s Law and ADE solvers. Mean velocity \( U \) and mean relative mass distribution \( \langle m \rangle \) are obtained by ensemble averaging.

**Application to MADE** Zech et al. (2020) adopted the K-structure for MADE and its characterizing parameter values based on the inspection of the piezometric surface map of Boggs et al. (1992, Fig. 3), on two large scale pumping tests (Table 1) and on a few flow meter logs (Boggs et al., 1992). According to observations, the area near the source is dominated by low conductivity \( K_2 \) with inclusions of high \( K_1 \). The relative area of the inclusions is \( p = 15\% \) determined from flow meter log analysis. Beyond a distance of 20 m downstream of the injection location the distribution is inverted: the bulk is dominated by the high conductivity \( K_1 \) with \( p = 15\% \) inclusions of low \( K_2 \). The inclusions are of thickness \( \ell_v = 0.5 \text{ m} \) and length \( \ell_h \in [5, 10, 20] \text{ m} \). The former was determined from flowmeter observations while the latter are subjected to parametric uncertainty. The range was determined from the ratio of thickness and observed anisotropy value. The random component of the model comprises of the locations of the three vertical inclusions of 0.5 m thickness, while length and horizontal position of inclusions are...
fixed for every realization. Given the inclusion occurrence of \( p = 15\% \), an aquifer thickness of 10 m and a thickness of 0.5 m, a total number of 3 inclusions per block are randomly placed in vertical position with equal probability. Altogether, an ensemble of 600 conductivity structures was created with random inclusion structures, with groups of 200 realizations of same length inclusion length \( \ell_h \) of 5 m, 10 m or 20 m.

Flow and transport was calculated for every random \( K \)-structure solving the groundwater flow equation and subsequently the ADE with the FEM software OpenGeoSys in a 2D cross section of 220 m \( \times \) 10 m. The boundary conditions were of constant head on the vertical boundaries defined by the selected head gradient \( J = 0.003 \). The initial condition for transport was imposed by injecting a solute discharge \( Q_{in} = 1.15 \times 10^{-5} \text{ m}^3/\text{s} \) during a period of \( T_{in} = 48.5 \text{ h} \) in the injection well with a screen length of 0.6 m (Boggs et al., 1992). Porosity is \( \theta = 0.32 \), local dispersivity was \( \alpha_L = 0.01 \text{ m} \), not impacting the overall mass distribution. The resulting ensemble mean velocity was \( U = 0.0254 \text{ m/d} \) with a standard deviation of 0.02 m/d.

### 3.2.6 Reactors

**Background** The reactor model conceptualizes subsurface solute transport as series of reactors which are linked to aquifer structure. The concept allows to model transport in a setting where tracer is injected into a low permeability zone, capturing strongly constrained downstream movement and skewed plume shapes. The concept was applied e.g. by Molin & Cvetkovic (2010).

**K-Structure** Hydraulic conductivity is conceptualized as log-normal spatial random function \( Y \) with continuous two-point correlation structure. The multi-Gaussian geostatistical parameters are those reported by Bohling et al. (2016) based on DPIL for the model application to MADE of same values as those adopted in the First Order and MIMSCA models (sections 3.2.1 and 3.2.2).

**Flow** Mean uniform flow velocity is calculated analytically based on Darcy’s Law: \( U = K_{\text{eff}} \cdot J/\theta \). \( K_{\text{eff}} \) for MADE is derived analogously to the MIMSCA model from geostatistical parameters (Zarlenga et al., 2018), resulting in \( U = 0.026 \text{ m/d} \).

**Transport** Transport is modelled analytically as series of flow reactors, each described by an exponential function \( \exp(-t/\Delta \tau) \) of the mean turnover time \( \Delta \tau \). While this model cannot be related directly to the permeability distribution, it is of interest to examine the outcome of a conceptually different model. The number of reactors is a function of the ratio \( x/\Delta x \) of the transport domain size \( x \) and the velocity fluctuations length scale \( \Delta x \). Each reactor has a mean turnover time of \( \Delta \tau = \Delta x/U \) and thus a residence time pdf of \( \exp(-tU/\Delta x) \).

The residence time pdf in the \( x/\Delta x \) series of reactors is then a Gamma function:

\[
f(t, x) = \frac{e^{-\frac{tU}{\Delta x}} \cdot \left( \frac{tU}{\Delta x} \right)^{x/\Delta x}}{\Gamma \left( \frac{x}{\Delta x} \right)}
\]

The cumulative density function of residence time is obtained by integration as

\[
F(t, x) = \int_0^t f(\theta, x) d\theta = 1 - \frac{\gamma \left( \frac{x}{\Delta x}, \frac{tU}{\Delta x} \right)}{\Gamma \left( \frac{x}{\Delta x} \right)}
\]

where \( \gamma \) is the lower incomplete Gamma-function.
The spatial tracer distribution, as the tracer position pdf $p(x; t)$ [1/L] at time $t$ follows as:

$$p(x, t) = -\frac{\partial F(t, x)}{\partial x} = \frac{1}{\Delta x} F\left(\frac{x}{\Delta x}\right) \left[ G^{3.0}_{2.3} \left( \frac{tU}{\Delta x}, 0, 0, \frac{x}{\Delta x} \right) + \Gamma \left( \frac{x}{\Delta x}, \frac{tU}{\Delta x} \right) \left( \log \left( \frac{tU}{\Delta x} \right) - \psi^{(0)} \left( \frac{x}{\Delta x} \right) \right) \right]$$

where $G$ is the Meijer g-function, and $\psi^{(0)}$ is the Polygamma function. With unit tracer mass, we have $m(x, t) = p(x, t)$.

Note that in the limit $\Delta x \to 0$, the number of reactors tends to infinity $x/\Delta x \to \infty$, and we recover plug flow $f(t, x) \to \delta(t - x/U)$.

Like in the MIMSCA model, the length parameter $\Delta x$, the integral scale of the velocity fluctuations, is assumed to be in the range of two to three log-conductivity horizontal correlation length $I$.

Application to MADE The value of $U = 0.026 \text{ m/d}$ is identical with the one used in the FOA or MIMSCA models and similarly $I = 9.1 \text{m}$ is based on Bohling et al. (2016).

4 Prediction and Inter-comparison using MADE Data

4.1 Results Presentation

The visual presentation of results concerning the mass spatial distribution significantly influences the perception of model performance. Critical aspects are the display scale, data upsampling (aggregation) and normalization. We present longitudinal mass distributions $\langle \tilde{m} \rangle$ at linear and logarithmic scales as well as in a cumulative form $\langle \tilde{m} \rangle$, to achieve a comprehensive display. The various display modes of the spatial mass distributions allow to interpret a few plume’s specific features: (i) mass peak location and bulk behavior; (ii) the tails (forefront and trailing zones) and (iii) mass recovery. They are relevant to specific goals such as risk assessment and remediation.

In line with the MADE experimental results, we remind that $\langle \tilde{m}(x, t) \rangle$ is aggregated over intervals of 10 m. We illustrate the effect by displaying the model outcomes at $T = 126$ days in both forms, the fine scale $\langle \tilde{m}(x, t) \rangle$ and the upscaled $\langle \tilde{m}(x, t) \rangle$ (Fig. 3) as well as $\langle M \rangle$. All longitudinal mass distributions are normalized with respect to the injected mass such that the area beneath $m$ or $\tilde{m}$ is unity. For the MADE data this is only true for $t = T = 126$ days where mass recovery is around 99%.

Spatial moments are commonly used to quantify the comparison between different models and measurements. However, this is not appropriate because of the skewed shape of $\langle \tilde{m} \rangle$ and the large impact of the uncertainty of the tail. Instead, we compare in Fig. 4 recovery locations at $T = 126$ days for 5%, 50%, and 95% as predicted by models relative to $\tilde{m}$ values of $x = 0.7$, 8.6, 42.8 meters. The definition of the recovery location is for instance for $x_{95\%}$, the position for which 95% of the total mass is upstream of $x_{95\%}$, i.e. $\langle M \rangle = 0.95$.

Spatial moments are commonly used to quantify the comparison between different models and measurements. However, this is not appropriate because of the skewed shape of $\langle \tilde{m} \rangle$ and in particular the large impact of the minute and uncertain mass fraction in the forefront tail, upon the second spatial moment (Fiori et al., 2017). Instead, we compare in Fig. 4 recovery locations at $T = 126$ days for 5%, 50%, and 95% as predicted by models relative to $\tilde{m}$ values of $x = 0.7$, 8.6, 42.8 meters. The definition of the recovery location is for instance for $x_{95\%}$, the position for which 95% of the total mass is upstream of $x_{95\%}$, i.e. $\langle M \rangle = 0.95$.

-13-
Figure 3. Longitudinal mass distribution for models at $T = 126$ d in fine model resolution (left column) and upscaled (aggregated) form ($\Delta x = 10$ m, right column) against MADE-1 experiment data at linear scale (1st row), log-scale (2nd row) and in cumulative mass (3rd row). Vertical lines in 3rd column indicate locations of 5% (dotted), 50% (dashed), and 95% (dashed dotted) recovered mass. Note that for the Binary Facies model the fine scale refers to a grid resolution of 2 meters.
4.2 Comparison between Models Prediction and MADE Experiment

Fig. 3 displays the longitudinal mass distribution for all models and the MADE experimental data at \( T = 126 \) days after injection: \( \langle m \rangle \) at model’s fine resolution and \( \langle \bar{m} \rangle \) aggregated over 10 m intervals, including MADE. Direct comparison is most revealing at that time since the experimental recovery rate is 99%.

The various models display some differences in their mass distribution \( \langle m \rangle \) at fine scale. Particularly, the peak value is higher for the Flow Reactors by a factor of 2 than the other models. The Binary Facies model displays plume tailing downstream of the other models, with \( x_{95\%} \approx 60\) m while for all the others models \( 30\) m \( \lesssim x_{95\%} \lesssim 40\) m (see Fig. 3).

The comparison between \( \langle m \rangle \) and the upscaled \( \langle \bar{m} \rangle \) in Fig. 3 reveals a few interesting features: upscaling reduces the peak values of \( \langle m \rangle \) by a factor of around 2, the spreading zone is expanded, and the differences between models prediction are greatly reduced. In particular, the predicted \( \langle m \rangle \) agrees quite well with MADE, much better than \( \langle \bar{m} \rangle \), as far as visual inspection reveals. This is expectable though in the past models prediction at fine scale were compared with MADE (e.g. Harvey & Gorelick (2000); Dogan et al. (2014)). The upstream spread of the aggregated \( \langle \bar{m} \rangle \) for MADE is partly an artifact of upscaling: it smears the upstream forced injected mass of the initial solute body over 10 meters. It could be erroneously interpreted as upstream dispersion.

The quantitative results in Fig. 4 on recovery locations relative to MADE strengthen the conclusions from the visual inspection. All models are in good agreement with MADE for the location \( x_{50\%} \). We attribute that to the closeness of the mean velocity \( U \) of various models (see previous Section). The Binary Facies differs by 15% due to the differ-
Figure 5. Longitudinal mass distribution for models and MADE-1 experiment at 49, 202, 279, 370, and 503 days after injection at linear scale (1st column), log-scale (2nd column) and in cumulative form (3rd column) with recovery locations (Fig. 3). Observe the different recovery rates of 2.06, 0.68, 0.62, 0.54, and 0.43, respectively.
Figure 6. Longitudinal mass prediction of models at 1000 days at linear scale (1st column), log-scale (2nd column) and in cumulative form (3rd column) with recovery locations.

ence in the value of $U$. The agreement is still good for $x_{5\%}$ with some deviations for the

$Binary Inclusion$ model. Last, the models prediction relative to $MADE$ of $x_{95\%}$, reflect-

ing the "fast" moving solute, is more variable, but still within acceptable differences in

practice.

Fig. 5 summarizes the model and $MADE$ results of $\langle \bar{m} \rangle$ for all times for which $MADE$

experimental data is available. The comparison at these times with $MADE$ is more dif-

ficult than for $T=126$ days because of the variable mass recovery, which is visible in

Fig. 5 in the cumulative mass panel $\langle M \rangle$.

Our interpretation of the apparent "mass loss" for later times is the less dense sam-

pling in the downstream zone as illustrated by Fig. 1 on one hand and the not-sampled

solute quickly moving in high conductivity channels on the other hand, as already men-

tioned before. This is clearly visible in Fig. 5 displaying the larger predicted mass than

the measured one downstream of the peak. Despite that, it is remarkable that for $T =

202, 270, 370$ days all models agree quite with both measured peak value and its distance

from the injection zone. At the largest time $T = 503$ days the low mass recovery (43

%) makes the comparison between data and models prediction quite problematic. Still,

the peak location and even the value are within acceptable differences in practice. As

for inter-comparison of models prediction, inspection of the robust cumulative mass $\langle M \rangle$

at different times shows remarkable closeness except for $Binary Facies$. The latter over-
estimates the location $x_{95\%}$ pertaining to fixed values of $\langle M \rangle > 0.4$. This enhanced tail-
ing is attributed in part to the larger integral horizontal scale identified from granulom-

etry analysis (Fig. 2) as compared to that resulting from DPIL, which was used by the

most of the models. In addition, it may be related to channels of high conductivity present

in some realizations of the binary $K$ field. It is remarkable that for both $\langle \bar{m} \rangle$ and $\langle M \rangle$, the predictions by $FOA$, $MIMSCA$ and $TDRW$ are very close for all $x$ (see discussion

in Sect 5).

4.3 Prediction Beyond $MADE$ Experiment

The important role of models is to provide prediction of future solute plumes de-

velopment. In order to compare the outcome of the 6 different models considered in the

present study, we have used them for the same $MADE$ conditions, but at larger time than

$T=503$ days. Thus, the long term (at $T = 1000$ days) predicted plume mass spatial dis-

tribution is displayed in Fig. 6. All models agree in the peak travel distance as a con-

sequence of similar flow velocities $U$. However, the differences between the peak value

are more pronounced but still within a factor of two and even less for all models, except

the $Reactors$, which predicts a higher peak and reduced tail. As for prediction of fore-

front tailing associated with fast moving solute, both $Binary Facies$ and $Binary Inclu-
sions$ display longer tails with $80 \lesssim x_{95\%} \lesssim 140$ meters. Inspection of the cumulative
distribution reveals again that prediction of models except *Binary Facies* are within a relatively narrow distribution.

5 Discussion

5.1 Modelling

The main aim of the study is to compare the prediction of six different models, with *MADE* serving as a platform. However, based on conceptual similarity as well as prediction, the six models can be divided in three groups as follows: (i) *FOA* (First Order Approximation), *MIMSCA* (Multi Indicator and Self Consistent Approximation), and *TDRW* (Time Domain Random Walk); (ii) *BF* (Binary Facies) and *BI* (Binary Inclusions); and (iii) *Reactors*. Herein a discussion of the main results for each group.

(i) FOA, MIMSCA, and TDRW

For the three models $K$ is modeled as a multi-Gaussian stationary random field, completely characterized by $K_G$, $\sigma^2_Y$, $I$, $I_v$, the flow mean velocity being given by $U = K_{eff} J/\theta$. They lead to approximate distributions of $\langle m \rangle$, $\langle \bar{m} \rangle$ and $\langle M \rangle$ as functions of $x, t$.

The three models differ in conceptualization and computational complexity. *FOA* is analytical, leading to an Inverse Gaussian $\langle m \rangle$ which satisfies an ADE with macrodispersivity derived analytically by *FOA*. *MIMSCA* is semi-analytical, based on summation of travel time through inclusions of random $K$. *TDRW* is semi-numerical, with the velocity field derived by *Monte Carlo* simulations while transport is based on an approximation of the Lagrangian velocity field.

One of our main result is that the solutions for $\langle \bar{m} \rangle$ and $\langle M \rangle$ by the three models are very close and in good agreement with the bulk of *MADE* experimental data. Thus they are very robust and prediction depends primarily on $U$ *Fiori* et al. (2017), as well as $\sigma^2_Y$ and $I$ and much less on models methodology.

(ii) BF and BI

The hydraulic conductivity heterogeneity is modeled by two values $K_1, K_2$ of volume fractions $p_1$ and $p_2 = 1 - p_1$. Two length scales $L_x, L_z$ characterize the $K$-facies, whose geometry has random elements. Flow and transport are solved numerically and repeatedly, by *Monte Carlo* simulations; besides the mean values $U$, $\langle \bar{m} \rangle$ and $\langle M \rangle$, the statistical moments of these parameters can be also obtained.

The two models differ in few respects: *BF* is three-dimensional; the random facies geometry is generated by transitional probability using the *TROGS* code, which requires the knowledge of two more length scales for the second $K$-facies being different from the first. All the structural parameters are obtained from granulometry measurements. In contrast, the simpler *BI* model is two-dimensional and consists of identical rectangular inclusions of conductivity $K_1$ submerged in the $K_2$ matrix for specified deterministic regions. The inclusions lengths assume 3 different values of same probability, their elevation being random. The structural parameters $K_1, K_2$ and $p_1$, as well as the length scales, are derived from pumping tests and few flowmeter measurements.

The main results for these two models are as follows: the agreement with *MADE* data is reasonable; while the characterization effort is less demanding than for the previous models, the numerical solutions and the *Monte Carlo* simulations of both flow and transport are quite involved, especially for *BF*.

The main conclusions are: the favorable comparison with *MADE* is an additional proof of models robustness; the simplified structures and characterization are adapted
to the particular features of the MADE site for which two dominant zones could be delineated.

(iii) Reactors

This model has a different conceptualization from the previous ones and it was motivated by its general use in engineering and convenience for reactive transport modelling. The model parameters are $U$, which is derived from the solution of flow via $K_{\text{eff}}$, and $\Delta x$, the velocity longitudinal correlation length which is related to $I$. The degree of heterogeneity quantified by $\sigma^2_I$ is not included, instead the series of reactors aggregate the inherent dispersion of a flow reactor; thus the model is not a general candidate for modelling advective transport in a heterogeneous aquifer. Still, it was found of interest to compare the analytical solution for $\langle \bar{m} \rangle$ with MADE plume. The surprising result is that $\langle \bar{m} \rangle$ and even more so $\langle \bar{m} \rangle$ agree reasonably well with MADE, though the predicted peak of $\langle \bar{m} \rangle$ is larger by a factor of 1.5 than prediction by other models for $T = 1000$ days.

The finding strengthens the conclusion about the robustness of $\langle \bar{m} \rangle$ and $\langle M \rangle$ in predicting the measured MADE plume and its future development, with the predominant role of two parameters, the mean velocity $U$ and correlation scale $I$; whether the reactors model can be used for prediction requires its further development and comparison with more cases.

5.2 Data Selection

The $K$ distribution at the MADE site as inferred from different characterisation methods is summarized in Fig. 2: (i) DPIL measurements are a novel, affordable technique for shallow aquifers (Dietrich et al., 2008) by which the most comprehensive data set was obtained suitable for a geostatistical interpretation, including two-point statistics: (ii) Granulometry (or grain size analysis) is a standard method in hydrogeology that yields highly uncertain conductivity estimates (Vienken & Dietrich, 2011); (iii) Flowmeter measurements and pumping tests are standard methods for hydrogeological characterisation, the main limitation being accuracy of low $K$ values (Fig. 2).

The (semi-)analytical models FOA, MIMSCA, TDRW, and Reactors used DPIL observations to estimate the flow velocity and the plume spreading. The model Binary Facies (BF) used granulometry data and the Binary Inclusions (BI) model used pumping test estimates, information from head maps and a few flowmeter data. In essence, BF simplifies the facies approach applied by Bianchi and Zheng (2016) MADE2 by considering 2 facies instead of 5, where the binary $K$ values are inferred from granulometry. The BI approach simplifies further the 3D random structure by considering regular inclusions in two dimensions. Thus the semi-analytical models are relatively simple for computations but use a more extensive DPIL data set whereas BI and BF modelling approaches are heavier on computation but use more readily accessible data sets.

The derivation of the mean velocity $U$ by the different models imply the use of estimates of the measured mean head gradient $J$ and the effective porosity $\theta$, which are approximate. The models differ primarily in the use of the $K$ data. Still, the resulting estimates of the mean velocity $U$ are relatively close as revealed by the values appearing in Sect. 3.2 (0.026m/d for FOA, MIMSCA, 0.019m/d for TDRW, 0.079m/d for BF and 0.025m/d for BI). Even the deviation for BF is within an acceptable range, various approximations notwithstanding. Thus, the estimates of $U$ are quite robust, which explains the relative closeness of the predicted and measured locations of the peak of $\langle \bar{m} \rangle$ in Figures 3, 5, 6.

Similarly, the prediction of spreading as quantified by the $\langle \bar{m} \rangle$ and $\langle M \rangle$ distributions is quite robust, as already discussed above. An interesting finding which may somewhat explain the relative closeness of the distributions for different data characteriza-
tion methods is the magnitude of the FOA asymptotic longitudinal macrodispersivity 
\( \alpha_L = \frac{\sigma_Y^2}{I} \) based on the different values of \( \sigma_Y^2 \) and \( I \) in Fig. 2. The resulting values of \( \alpha_L \) are 53.7m, 54.1m and 49.6m for DPIL, Flowmeter and Grain Size, respectively.

Although above observations strictly apply to the MADE site only, they are nevertheless encouraging and motivate similar comparative analysis e.g., for less heterogeneous aquifers for which experimental data are available.

6 Summary and Conclusions

With a variety of hydraulic data available, MADE provides a unique opportunity for a comparative analysis of predictive modelling, from a wide range of (semi)-analytical models (FOA, MIMSCA and TDRW) that utilise extensive DPIL data for inferring geostatistical parameters, to numerical models (BF and BI) with relatively simple (binary) structures that utilise much less extensive data sets (granulometry and pumping tests+ flowmeter). The present paper takes advantage of these possibilities offered by MADE, and focuses on comparing predictions of the plume spreading by six different models. The models differ in theoretical formulations, in the conceptualization of aquifer structure, in the field data input, and in the computational effort. Common features of the models are: flow is steady, uniform in the mean and driven by a head gradient \( J \); solute spreading is caused by aquifer conductivity heterogeneity; models rely on structural data and flow data with no calibration on transport observations i.e. the models are predictive; plume mass behavior is assumed ergodic, i.e. the mean relative mass distribution \( \left\langle \bar{m} \right\rangle \) derived by the model is compared with the measured mass \( \bar{m} \) at a few times \( T \). Moreover, the apparent loss of mass of measured mass at MADE1 for \( T > 126 \) days (attributed to limited sampling) is not taken into account by the models. Model comparison at \( T = 1000 \) days, i.e., beyond the period of measurements, is also included.

The main and encouraging result for practitioners is that all model prediction agree reasonably well with MADE1 mass distributions and the same for the comparison at \( T=1000 \) days. Thus, the measures of the solute plumes are robust and models are reliable as long as they are underlined by a few basic parameters: mean velocity \( U \), a parameter reflecting log-conductivity variability and one taking horizontal correlations in conductivity into account. However, the reasonable agreement in model results is also related to the particular quantity under examination: the longitudinal mass distribution which is aggregated over spatial intervals is quite robust itself. If other measures are employed, such as local concentrations, results might differ.

To render the above conclusions of general validity, the study shall be extended by application to other cases than MADE. As a first step synthetic examples can be considered like formations of log-normal conductivity, with different connectivities as analyzed for instance by Fiori et al. (2017), or typical facies structures (e.g. Carle & Fogg (1996)). Similarly, one needs to test the hypothesis that a consistent comparison would have been obtained even for a less heterogeneous aquifer.

Acknowledgments

We thank Marco Bianchi and Boris Baeumer for their support during the development of the study. AF and AB acknowledge funding from the Italian Ministry of Education, University and Research (MIUR) in the frame of the Departments of Excellence Initiative 2018-2022 granted to Dept. of Engineering of Roma Tre University, and to the Dept. of Civil, Environmental and Mechanical Engineering of the University of Trento, respectively. MD acknowledges funding of the European Research Council (ERC) through the project MHetScale (contract number 617511), and the Spanish Research Agency (AEI) through the project HydroPore (contract number PID2019-106887GB-C31). Data on mod-
els and MADE observations can be found in the references. Scripts used in the paper are available upon request from the corresponding author.

References


Water Resources Research, 51(9), 7662–7679. doi: 10.1002/2015WR017220
