

# Water Resources Research

## RESEARCH ARTICLE

10.1029/2020WR028672

### Key Points:

- Aquifer transport is predicted by six models of different heterogeneity conceptualization and field data input
- Key model inputs are mean velocity and heterogeneity representation including spatial correlation
- The macrodispersion experiment (MADE) site plume longitudinal mass distribution is a robust transport measure

### Correspondence to:

A. Zech,  
a.zech@uu.nl

### Citation:

Zech, A., Attinger, S., Bellin, A., Cvetkovic, V., Dagan, G., Dentz, M., et al. (2021). A comparison of six transport models of the MADE-1 experiment implemented with different types of hydraulic data. *Water Resources Research*, 57, e2020WR028672. <https://doi.org/10.1029/2020WR028672>

Received 3 SEP 2020  
Accepted 19 APR 2021

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## A Comparison of Six Transport Models of the MADE-1 Experiment Implemented With Different Types of Hydraulic Data

Alraune Zech<sup>1,2</sup> , Sabine Attinger<sup>2,3</sup> , Alberto Bellin<sup>4</sup> , Vladimir Cvetkovic<sup>5</sup> , Gedeon Dagan<sup>6</sup> , Marco Dentz<sup>7</sup> , Peter Dietrich<sup>2,8</sup> , Aldo Fiori<sup>9</sup> , and Georg Teutsch<sup>2</sup> 

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

**Abstract** Six conceptually different transport models were applied to the macrodispersion experiment (MADE)-1 field tracer experiment as a first major attempt for model comparison. The objective was to show that complex mass distributions in heterogeneous aquifers can be predicted without calibration of transport parameters, solely making use of structural and flow data. The models differ in their conceptualization of the heterogeneous aquifer structure, computational complexity, and use of conductivity data obtained from various observation methods (direct push injection logging, DPIL, grain size analysis, pumping tests and flowmeter). They share the same underlying physical transport process of advection by the velocity field solely. Predictive capability is assessed by comparing results to observed longitudinal mass distributions of the MADE-1 experiment. The decreasing mass recovery of the observed plume is attributed to sampling and no physical process like mass transfer is invoked by the models. Measures like peak location and strength are used in comparing the modeled and measured plume mass distribution. Comparison of models reveals that the predictions of the solute plume agree reasonably well with observations, if the models are underlain 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. The robustness of the results implies that conservative transport models with appropriate conductivity upscaling strategies of various observation data provide reasonable predictions of plumes longitudinal mass distribution, as long as key features are taken into account.

## 1. Introduction

Modeling 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; Fetter et al., 2018; Gelhar, 1993; Rubin, 2003). The task of predicting transport faces several 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 uncertain.

Under these circumstances, models play an important role: they help understanding the involved processes, analyzing field data, and making long term predictions. 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 their goals. We focus on transport of plumes of conservative solutes in steady flow, driven by the natural head gradient. At the considered aquifer scale, the main mechanism responsible for plume spreading is macrodispersion (Zech et al., 2015, 2019) due to the spatial variability of the hydraulic conductivity  $K(\mathbf{x})$ . The effect increases with higher  $K$  heterogeneity as quantified for instance by the log-conductivity variance  $\sigma_Y^2$ . We focus on the quantification of transport by the spatial relative mass distribution in mean flow direction  $x$ .

A few elaborate field experiments were conducted in the past, in order to gain understanding and validate models. A challenging one is at the macrodispersion experiment (MADE) site (Boggs et al., 1992) located near Columbus, Mississippi. It is situated in a highly heterogeneous aquifer, making it of relevance to many actual aquifers (Gómez-Hernández et al., 2017). An important feature of MADE was the application of different observation methods, in order to characterize the spatial distribution of hydraulic conductivity (e.g., Boggs et al., 1990; Rehfeldt et al., 1992; Bohling et al., 2016). Long-term tracer tests provide observed relative mass distribution in space at a few fixed times  $t$  (Boggs et al., 1990, 1995). It has motivated a flurry of works on structure characterization by field data and different modeling strategies (see e.g., Zheng et al., 2011).

We examine the ability of six transport models of different aquifer conceptualization to predict the observed relative mass distribution without calibration of transport parameters: (a) the first order approximation (FOA) (Fiori et al., 2017); (b) the multi-indicator model and self-consistent approximation (MIMSCA) (Fiori et al., 2013); (c) the time-domain random walk (TDRW) of Dentz et al. (2020); (d) the Binary Inclusions model (Zech et al., 2021); (e) a Binary Facies model, and (f) the flow reactors model (see Section 3 for details). We further examine the predictive power of the models by extending the time ( $t = 1,000$  days) beyond MADE observations ( $t_{\max} = 503$  days).

The selected models adopt upscaling: they do not aim to reproduce the actual spatial conductivity distribution, but to represent the aquifer structure at a simplified level of complexity, which still accounts for the effect of heterogeneity on transport. Within that frame, models cover a wide spectrum of aquifer configurations. They differ in conceptualizations of formation structure and of transport, in the use of field data, in computational methodology and in effort. Key features they share are the advection dominated subsurface transport, flow, and transport initial and boundary conditions, and a model setup independent of transport observations.

The limitation to transport models independent of measured concentrations is motivated by the aim of application to field situations, where transport tests at the relevant space and time scales are hardly available due to their duration and costs. Models use the independently inferred hydraulic data to predict advection-dominated transport and do not incorporate mass transfer as an additional mechanism. Although mass transfer models are useful to predict non-Gaussian plumes, particularly in fractured media, their application to field scale transport in porous media, such as MADE (Harvey & Gorelick, 2000), requires determining mass transfer rates from calibration by using concentration observations. We concentrate here only on predictive transport models relying solely on field data of aquifer properties and flow.

We believe that the comparison of the models which differ in type and underlying field data, conceptualization of conductivity structure, and complexity of computations by using 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.

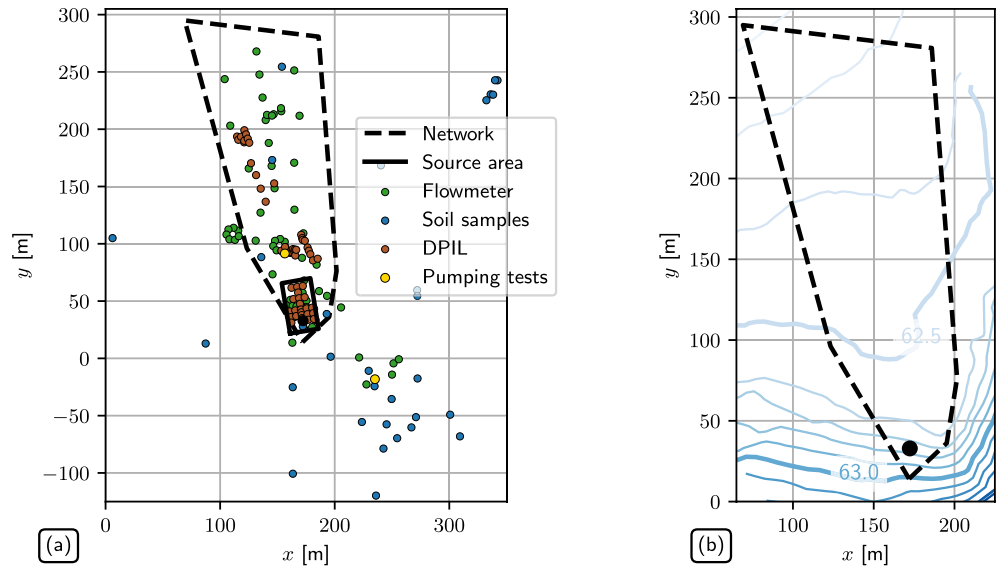
The plan of the study is as follows: Section 2 recapitulates the MADE aquifer characterization and transport experiment; Section 3 describes the methodology of the models, including their application to MADE; Section 4 is devoted to the model prediction for solute mass distribution in comparison to MADE observations, as well as at later time. Section 5 contains the general discussion on data comparison, while Section 6 concludes the study.

## 2. MADE Transport Field Experiment

The MADE experiment was the object of a large body of publications, dealing with aquifer characterization, structure, and conductivity analysis as well as transport interpretation; see for instance reviews of Zheng et al. (2011) and Gómez-Hernández et al. (2017). We recall in the following only those aspects of direct relevance to the examined models.

### 2.1. Hydraulic Conductivity Distribution

The MADE site aquifer is composed of highly heterogeneous alluvial terrace deposits. Measurements of hydraulic conductivity at multiple locations (Figure 1a) were performed by flowmeter, slug test, and direct push injection logging (DPIL) (Boggs et al., 1990; Bohling et al., 2016; Rehfeldt et al., 1992). Granulometry

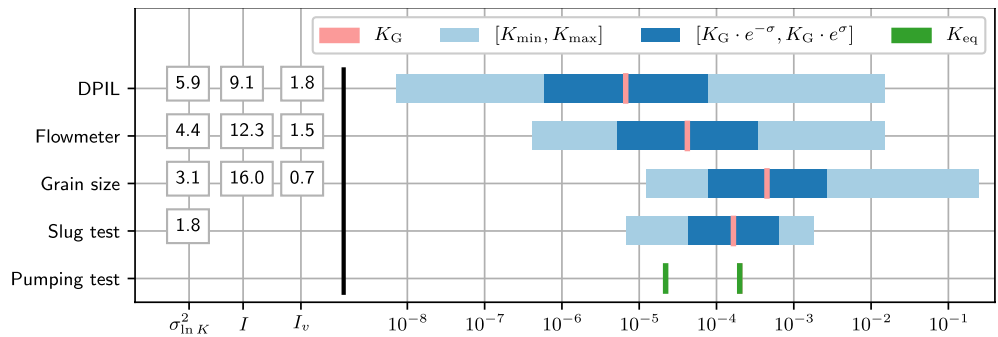


**Figure 1.** (a) Map of macrodispersion experiment (MADE) site according to Boggs et al. (1990) and Bohling et al. (2016): Locations of hydraulic conductivity measurements devices (colored dots); tracer test source area (black outline); and sampling network boundary (dashed outline) where bromide samples were collected. (b) Potentiometric surface map of head measurements according to Boggs et al. (1990). Black dot marks tracer test injection location.

analysis of soil samples provided additional estimates. Besides, two pumping tests provide equivalent conductivities  $K_{eq}$  of the volume surrounding the wells (Boggs et al., 1992). The use of different techniques at the same site and subsequent application by different models offers an opportunity to examine their impact on transport prediction.

The spatially distributed observations (Figure 1a), carried out at different depths resulted in a large volume of data which served for geostatistical analysis. Figure 2 summarizes their outcomes, which is of interest for the application of the different models (Section 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 2,611 observations and more recently DPIL with 31,123 samples (Bohling et al., 2012, 2016).

The differences in the geostatistical parameters, revealed by Figure 2, especially in the geometric mean conductivity  $K_G$ , are a result of the different properties of the observation methods as well as the density and locations of observation points. In particular, the difference between flowmeter and DPIL can be



**Figure 2.** Geostatistical measures of hydraulic conductivity  $K$  for macrodispersion experiment (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)  $I$  and  $I_v$ , respectively. Visualization of geometric mean  $K_G$ , range of observed values from minimal to maximal ( $[K_{min}, K_{max}]$ ), and range of one standard deviation around mean ( $[K_G e^{-\sigma}, K_G e^{\sigma}]$ ).  $K_{eq}$  denotes the equivalent conductivity for two large scale pumping tests (Boggs et al., 1990).

explained by the inability of flowmeter to detect low  $K$  values. While the maximal values  $K_{\max}$  are close (around 0.005 m/s) the minimal values  $K_{\min}$  differ by two orders of magnitude. The effect manifested in Figure 2 is that flowmeter  $K_G$  is larger and  $\sigma_Y^2$  is smaller. The difference in the longitudinal integral scales  $I$  suggests that the zones of low  $K$  values are less connected than those of higher magnitude. The analysis based on 214 soil samples 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 and Zheng (2016) and herein for a simplified binary structure approach. The impact of these difference upon flow and transport are discussed in Section 4.

## 2.2. Flow

In Figure 1b, we reproduce the head contour lines map of Boggs et al. (1992, their Figure 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 (Boggs et al., 1990).

## 2.3. Transport Experiment

We focus on the first tracer transport experiment, which was conducted in the years 1986–1988 (Adams & Gelhar, 1992; Boggs et al, 1990, 1992; Rehfeldt et al., 1992). The tracer plume displayed a non-Gaussian longitudinal solute mass distribution with the bulk of the mass staying near the source and small amounts spreading over large distances.

### 2.3.1. Initial Conditions

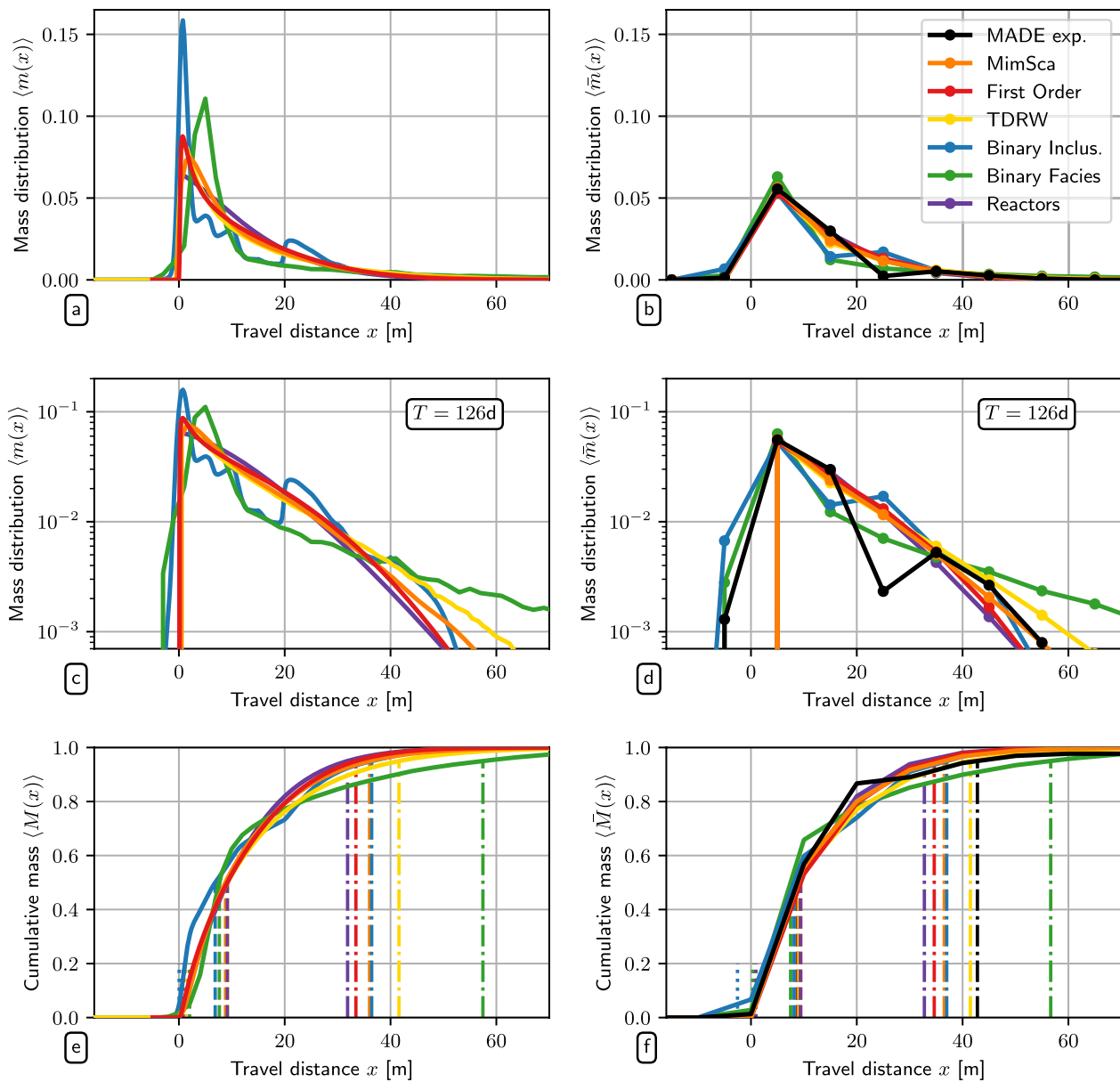
A solution obtained by dissolving  $M_0 = 25$  kg of bromide in  $10 \text{ m}^3$  of water was injected over a period of 48 h through five wells of 60 cm screen length. While the subsequent transport took place under natural head gradient conditions, the tracer solution was forced radially into the aquifer during injection. As a result, the initial tracer body was much larger than the expected extent of 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, Figure 5). The apparent upstream tracer spread is not a result of dispersive transport opposite to the main flow direction, but of tracer input history. This is important for models which assume ergodicity and for the setup of initial conditions. The injection mode also implies that the initial condition was flux proportional, with a preference of mass flowing in high-conductivity channels.

### 2.3.2. Plume Detection and Data Aggregation (Upscaling)

We reproduce the observed longitudinal relative mass distribution  $\bar{m}(x, t)$  of Adams and Gelhar (1992, Figure 7) at six times: 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 (Figure 1a). Subsequently,  $C$  was numerically integrated over transverse planes ( $y, z$ ), accumulated and averaged over slices 10-m long in the  $x$  direction. The obtained upscaled longitudinal relative mass  $\bar{m}(x, t) = \frac{1}{\delta} \int_{x-\delta/2}^{x+\delta/2} m(x', t) dx'$ , where  $m(x, t)$  is the relative longitudinal mass (Section 3), is assigned to the centers of the slices at  $x = -5 \text{ m}, 5 \text{ m}, 15 \text{ m}, \dots$  (Adams & Gelhar, 1992). The fact that the reported mass is an upscaled/aggregated quantity was overlooked by many previous studies which compared the observed  $\bar{m}$  with  $m$ , the modeled mass at fine scale, as mentioned by Fiori et al. (2019). The significance of data aggregation is discussed in Section 4.

### 2.3.3. Mass Recovery

The observed relative mass  $\bar{m}$  displays decreasing recovery rates  $M_{\text{total}}/M_0$  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. Thus, the observed relative mass does not obey the mass conservation requirement  $\int \bar{m}(x, t) dx = 1$ , except at 126 days after injection (Figure 3). We do not normalize observed mass  $\bar{m}$  by the total reported mass  $\tilde{m} = \bar{m}(x, t) / \int \bar{m}(x, t) dx$ , but use  $\bar{m}$  for comparison with theoretical models, as it is appropriate for a conservative solute.



**Figure 3.** Longitudinal mass distribution for models at  $T = 126$  days in fine model resolution (left column) and upscaled (aggregated) form ( $\Delta x = 10$  m, right column) against macrodispersion experiment (MADE)-1 experiment data at linear scale (first row), log-scale (second row) and in cumulative mass (third row). Vertical lines in third 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 m.

The discrepancy in mass recovery rate is indicative of uncertainty of the data. As discussed in Adams and 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 buildup during injection. Preferential sampling from high conductivity regions is also possible. In such case, the assumption of uniform tracer distribution employed in the spatial integration easily leads to an overestimated mass recovery. In contrast, in line with Fiori (2014), we attribute the decreasing mass recovery at later times to insufficient sampling in the downstream zone. The high heterogeneity implies possible channeling effects at MADE and mass trapping in low-conductivity parts which are hardly sampled. Similarly, the inadequate coverage due to the much lower sampling density downstream leads to observation errors, particu-

larly at the edges of a plume. Thus, the much lower sampling density makes it difficult to sample the leading edge of the plume, which may contain a significant fraction of the plume mass at MADE (Fiori et al., 2019).

### 3. Subsurface Transport Models

We compare six subsurface transport models, which were developed independently by the authors in the past or devised recently in the frame of the present study. Before providing a short description of each model, we outline the key features shared by all models.

#### 3.1. Common Features and Prerequisites

The following features determined the selection of the models considered herein:

1. We consider 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 driven by the natural head gradient  $J$  in the  $x$  direction.
3. Transport of solutes is the result of advection by the Eulerian velocity field with neglect of local dispersion, which has a negligible effect on  $\bar{m}$ . Spreading is caused by the spatial variability of conductivity  $K(x)$ , while the effective porosity  $\theta$  is constant. Along Equation 1, we do not consider dual-domain or transport models which use mass transfer parameters. Mass transfer effects are modeled indirectly here by transport in low conductivity zones.
4. A plume of mass  $M_0$  of a conservative solute is injected in the aquifer at  $t = 0$ , through the five injection wells. Spreading is quantified by the mass arrival at a control plane at  $x$ :  $M_{total}(x, t) = \theta \int_x^\infty \int C(x', y, z, t) dx' dy dz$ , where  $C$  is the concentration. For a fixed  $x$ ,  $M = M_{total}/M_0$  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 (upscaled) relative mass averaged over a medium slice of length  $\Delta = 10$  m centered at  $x$ , which is given by  $\bar{m}(x, t; \Delta) = (1/\Delta)[M(x - \Delta/2, t) - M(x + \Delta/2, t)]$  such that  $m = \lim_{\Delta \ll l} \bar{m}$ .
6. The upscaled relative mass  $\bar{m}$  is derived for the MADE conditions at  $T = 49, 126, 202, 279, 370$ , and 503 days after injection toward comparison of models outcome with measured  $\bar{m}$ . Additionally, models are applied to prediction of  $\bar{m}$  at  $t = 1,000$  days for models intercomparison.
7. The analysis is focused on the prediction of the mass ensemble mean for comparison with measured mass. Although a complete analysis would also comprise measures of uncertainty, this task is not undertaken here because of the limited extent of the paper and its aims.

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). Similarly, we exclude dual-domain models (e.g., Feehley et al., 2000 and Harvey & Gorelick, 2000) and the continuous time random walk (CTRW) model of Berkowitz and Scher (2001). The lithofacies approach of Bianchi and Zheng (2016) or the analog data-based approach of Pirot et al. (2015) would have been candidates, 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). Their conceptual framework is, however, underlying the facies model herein. Similarly, the fractional advection-dispersion model of Benson et al. (2001) was applied only to MADE-2 data. Furthermore, not all parameters, such as the skewness parameter  $\beta$ , are fully predictive based on structural data only. Dogan et al. (2014) presented a predictive fully numerical transport model for the MADE-1 experiment based on flowmeter and DPIL measurements. They generated detailed representations of the  $K$  field conditioned on observations in a small sector of the MADE site aquifer and subsequently solved the flow and transport equations. However, due to the computational effort, transport simulation results are limited to a fraction of the total plume transport distance. Thus, their results are unsuitable for model comparison, particularly with prediction beyond observation times.

### 3.2. Brief Description of Models

Here, we summarize the main features of the six models, their application to MADE and the use of observational  $K$  data.

#### 3.2.1. First Order Approximation (FOA)

**Background:** The solution of flow and transport in heterogeneous formations of a random log-normal conductivity distribution, that is of normal  $Y = \ln K$ , by a first order approximation in the log-conductivity variance  $\sigma_Y^2$  was the topic of intensive research in the last four decades, for example, Dagan (1989); Gelhar (1993); Rubin (2003), which led to various analytical solutions.

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

**Flow:** The mean velocity is given by  $U = K_{\text{eff}}J/\theta$ , with the effective conductivity  $K_{\text{eff}}/K_G = \text{func}(\sigma_Y^2, e)$  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 of  $Y' = Y - \langle Y \rangle$ .

**Transport:** The mean relative mass distribution  $\langle m(x, t) \rangle$  was derived for an initial condition of given deterministic resident concentration  $C_0(x, 0)$  either in a volume  $V_0$  with  $C_0 = M_0/(\theta V_0)$ , or with mass concentrated over an area  $A_0$  at  $x = 0$  and quantified by  $m_0 = M_0/A_0$  (Kreft & Zuber, 1978). Similarly, detection was in the resident mode with the mean cumulative mass  $\langle M \rangle$  and the distributed one  $\langle m \rangle$ , satisfying the advection-dispersion equation (ADE);

$$\frac{\partial \langle M \rangle}{\partial t} + U \frac{\partial \langle M \rangle}{\partial x} = D_L(t) \frac{\partial^2 \langle M \rangle}{\partial x^2} \quad (1)$$

The solution for the resident mode is given by  $\langle M_R \rangle = \frac{1}{2} \text{erfc} \left( \frac{x - Ut}{2\sqrt{D_L t}} \right)$  and

$\langle m_R \rangle = -\frac{\partial \langle M_R \rangle}{\partial x} = \frac{1}{2\sqrt{\pi D_L t}} \exp \left( -\frac{(x - Ut)^2}{4D_L t} \right)$  (Fiori et al., 2017; Kreft & Zuber, 1978). The Gaussian solution  $\langle m_R \rangle$  applies to weakly heterogeneous aquifers and/or far from the injection zone.

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

$$\langle M \rangle = \frac{1}{2} \text{erfc} \left( \frac{x - Ut}{2\sqrt{D_L t}} \right) + \frac{1}{2} \exp \left( \frac{Ux}{D_L} \right) \text{erfc} \left( \frac{x + Ut}{2\sqrt{D_L t}} \right) \quad (2)$$

The spatial mass distribution for a tracer pulse follows from Equation 2 with  $\langle m(x, t) \rangle = -\partial \langle M \rangle / \partial x$ . In contrast to the Gaussian solution  $\langle m_R \rangle$ , it displays a skewed shape and lack of upstream dispersion, thus providing a more realistic representation of transport in heterogeneous aquifers.

The macrodispersion coefficient  $D_L(t) = \alpha_L \cdot U$  in the ADE Equation 1, 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). Preasymptotic  $\alpha_L(t)$  was determined by a quadrature for an exponential covariance. It is approximated accurately by an analytical expression as function of mean flow velocity  $U$ , time  $t$  and aquifer statistics  $I$  and  $e$  (Dagan & Cvetkovic, 1993, Equation 20). The asymptotic value is attained after a travel distance of a few integral scales  $I$ .

Jankovic et al. (2017) showed that the solution Equation 2 with a dispersion coefficient  $D_L(x)$  given by the first order approximation represents accurately the results of numerical simulations even for the large value of  $\sigma_Y^2 = 8$ , except an underestimation of a few percent of the mass in the tail of late arrival.

**Application to MADE:** Application of solution Equation 2 to MADE was based on the parameters derived from Bohling et al. (2016) by DPIL as follows:  $K_G = 0.58$  m/d,  $\theta = 0.31$ ,  $I = 9.1$  m,  $I_v = 1.8$  m,  $\sigma_Y^2 = 5.9$ . Rather

than the first order approximation, we used the more general formula by Zarlenga et al. (2018, Equation 5) to arrive at  $K_{\text{eff}} = 2.28$  m/d. After identifying the representative mean head gradient in the plume zone (Figure 1b) as  $J = 0.0036$ , the mean velocity is given by  $U = K_{\text{eff}}J/\theta = 0.026$  m/d. Transient preasymptotic dispersion  $D_L(x) = U\alpha_L(x)$  is calculated according to Fiori et al. (2019, Equations C1 and C2) based on  $\alpha_L$  of (Dagan & Cvetkovic, 1993, Equation 20), with the asymptotic value  $\alpha_L = \sigma_Y^2 I = 53.7$  m. Solving Equation 2 with  $D_L(x)$  provides  $\langle M \rangle$  and the related  $\langle \bar{m} \rangle$  at the  $t$  and  $x$  values pertinent to MADE. Calculation of mass just takes seconds with an implementation of the analytical solution in a computer algebra system.

### 3.2.2. Multi-Indicator Model and Self Consistent Approximation (MIMSCA)

**Background:** MIMSCA was developed in the last 15 years as an approximate model of flow and transport for aquifers of arbitrary degree of heterogeneity (Cvetkovic et al., 2014; Dagan & Fiori, 2003; Fiori et al., 2006). Its outcome has been compared with accurate numerical solutions for  $\sigma_Y^2 \leq 8$  and applied to MADE. Fiori et al. (2019) recently used the model to assess the uncertainty of prediction due to nonergodic 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  $2I \times 2I \times 2I_v$ . The block  $K$  values are assigned independently with random values from a univariate pdf  $f(K)$  and the structure is designated as the MIM. The univariate  $f(K)$  was chosen to be lognormal of parameters  $K_G, \sigma_Y^2$ . This way, the random  $K$  field is completely defined in terms of four parameters, similar to the FOA (Section 3.2.1).

**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:** Solute transport in the multi-indicator conductivity structure was also solved with the SCA. The starting point is the travel time of a solute particle for the passage through an inclusions of conductivity  $K$  embedded in a matrix of  $K_{\text{eff}}$ , which Fiori et al. (2003) determined analytically. The solute BTC  $\langle M \rangle$  at  $x$  follows as sum over the travel-time random residuals pertaining to the inclusions of different  $K$  lying between the injection and the control plane. Initial condition was of flux proportional injection at  $x = 0$  while the mean BTC was derived by fast Fourier transform. Jankovic et al. (2003) arrived at 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.

**Application to MADE:** Fiori et al. (2013) applied the method to the MADE site transport setting to derive  $\langle m \rangle$  and  $\langle \bar{m} \rangle$  with an update in Fiori et al. (2019), motivated by a reevaluation of geostatistical input parameters by Bohling et al. (2016). The needed parameter values are the same as those given above for FOA. Calculating the mass distribution with the semianalytical solution of Fiori et al. (2013) required a few minutes in a computer algebra system.

### 3.2.3. Time-Domain Random Walk (TDRW)

**Background:** TDRW and CTRW 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 against transport observations. The basic idea is to quantify particle motion in spatially variable flow fields through a Markov process 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 (Comolli et al., 2019; Hakoun et al., 2019).



**K-Structure:** Hydraulic conductivity is represented by a 3D spatial random field of log-normal distribution. Thus, the random  $K$ -field is characterized in terms of four parameters, similar to the FOA (Section 3.2.1): conductivity geometric mean  $K_G$ ,  $\ln K$  variance, and correlation lengths  $\ell_h$  and  $\ell_v$ . Random  $K$ -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.

**Flow:** The steady state groundwater flow equation is solved numerically on an ensemble of random  $K$ -realizations of specified conductivity statistics. 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 univariate distribution. Thus,  $v_e(x) = q(x)K_G J / \theta$ , where  $q$  is the numerically derived flux.

**Transport:** Transport is modeled by a continuous time random walk. We provide the key aspects of the particular conceptualization. For a detailed model description, the reader is referred to Dentz et al. (2020). 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  $p_v(v)$  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 (Comolli et al., 2019; Dentz et al., 2016). The Markov process describing 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:** 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) and Adams and 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) (Figure 2). The average velocity component in mean flow direction is given by  $\bar{v}_1 = \bar{q}_1 K_G J / \theta = 1.942 \times 10^{-7}$  m/s = 0.0167 m/d. The average Eulerian velocity magnitude is  $\bar{v}_e = \bar{q} K_G J / \theta = 2.234$  m/s = 0.0193 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). Computation of particles transport in a C-implementation requires minutes on a usual computer.

### 3.2.4. Binary Facies

**Background:** An alternative approach to the continuous univariate  $Y$  distribution, used in the previous models is a hydrofacies model, which generates hydrofacies distributions compatible with the observational data. The method uses hydrofacies classification, instead of hydraulic conductivity as observational data. 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 and Fogg (1996). It enjoys flexibility in handling juxtapositional tendencies among hydrofacies and the availability of the software T-PROGS (Carle, 1999). In the context of the studies of MADE, Bianchi and Zheng (2016) presented an application to the MADE-2 experiment by using five 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 into a single low conductivity hydrofacies.

**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, for example, 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 hydrofacies class. 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 Kozeny-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  $L_{z,1}$ ,  $L_{z,2}$ ,  $L_{x,1}$ , and  $L_{x,2}$  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 framework by making use of numerical solvers, namely Modflow 2005 in combination with particle tracking (Pollock, 2012). Mean velocity  $U$  and mean relative mass distribution  $\langle \bar{m} \rangle$  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 MIMSCA models. The parameters values pertinent to MADE were identified as:  $p_1 = 0.145$ ,  $p_2 = 0.855$ ,  $K_1 = 190$  m/d,  $K_2 = 1.49$  m/d,  $L_{z,1} = 0.702$  m,  $L_{z,2} = 4.15$  m. The low borehole density prevented reliable estimate of the transition probability in the horizontal directions. Thus, isotropy is assumed in horizontal direction and the characteristic lengths were based on the integral scales identified by Rehfeldt et al. (1992) and assuming the relationship  $L_{1,z}/L_{1,h} = L_{2,z}/L_{1,h} = e$ , with  $e$  being the anisotropy ratio. Thus, horizontal length scales are obtained by dividing the vertical ones, by the estimate of  $e = 0.0437$  (Rehfeldt et al., 1992), arriving at  $L_{h,1} = 16.0$  m, and  $L_{h,2} = 94.8$  m, respectively. In the vertical direction, the Markov model of transition probability fitted to the experimental one was used. As already mentioned above, the horizontal length scale values shall be regarded as estimates, affected by uncertainty. A check of the results obtained with different, smaller, values (not shown) led to similar plume mass distributions except for the long distance tail of minute mass. An ensemble of over 200 independent Monte Carlo realizations of the hydrofacies distribution were thus generated.

Flow was solved numerically with Modflow 2005 in a computational domain of  $300 \times 100 \times 10$  m<sup>3</sup> 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  $\theta = 0.31$ . Advective transport was simulated by tracking 1,000 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  $U = 0.079$  m/day with a standard deviation of 0.0046 m/d. Computations in a Modflow-flopy environment allow to perform ensemble simulations within two hours on a common PC.

### 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. (2021), in which the stochastic conceptualization of the binary hydraulic conductivity took large scale deterministic information into account.

**K-Structure:** The aquifer is modeled as a 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 modeled as nonoverlapping conductivity blocks of length  $l$  and width  $l_v$ , placed randomly within the computational domain. 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  $\langle \bar{m} \rangle$  are obtained by ensemble averaging.

**Application to MADE:** Zech et al. (2021) 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, Figure 3), on two large scale pumping tests 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  $\ell_v = 0.5$  m and length  $\ell_h \in [5, 10, 20]$  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. Three inclusions of 0.5 m thickness are placed at random in the vertical direction, while the 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 three 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 simulated for each random  $K$ -structure by solving the groundwater flow equation, and subsequently the ADE with the Finite Elements Method software OpenGeoSys in a 2D cross section of  $220 \times 10$  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 the solution at a constant flow rate of  $Q_{in} = 1.15 \cdot 10^{-5}$  m<sup>3</sup>/s during a period of  $T_{in} = 48.5$  h (Boggs et al., 1992). Porosity  $\theta = 0.32$  and local dispersivities  $\alpha_L = 0.01$  m and  $\alpha_V = 0.01$  m were assumed. The resulting ensemble mean velocity was  $U = 0.0254$  m/d with a standard deviation of 0.02 m/d. Computation time on a common PC was 200 h of CPU time for the entire ensemble.

### 3.2.6. Flow Reactors

**Background:** The Flow Reactor model, or short Reactors, 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 for example, by Molin and Cvetkovic (2010).

**$K$ -Structure:** Hydraulic conductivity is conceptualized as stationary log-normal spatial random function  $Y$  with continuous two-point correlation structure. The 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:  $U = K_{eff} \cdot J/\theta$ . The effective conductivity  $K_{eff}$  is a function of the geostatistical parameters of log conductivity.

**Transport:** Transport is modeled analytically as series of flow reactors. 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  $\Delta x$ . We view the transport domain as  $x/\Delta x$  flow reactors, each with a mean turnover time  $\Delta x/U$ . The residence time within each reactor is an exponential function  $\exp(-t/\Delta\tau) = \exp(-tU/\Delta x)$ . Thus, the residence time pdf of the reactor series is the Gamma function:

$$f(t, x) = \frac{\exp\left(-\frac{tU}{\Delta x}\right) t^{\frac{x}{\Delta x}-1} \left(\frac{U}{\Delta x}\right)^{x/\Delta x}}{\Gamma\left(\frac{x}{\Delta x}\right)} \quad (3)$$

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

$$F(t, x) = \int_0^t f(\theta, x) d\theta = 1 - \gamma\left(x / \Delta x, tU / \Delta x\right) \quad (4)$$

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 from  $p(x, t) = -\frac{\partial F(t, x)}{\partial x} = \frac{\partial \gamma(x / \Delta x, tU / \Delta x)}{\partial x}$  as:

$$p(x,t) = \frac{1}{\Delta x \Gamma\left(\frac{x}{\Delta x}\right)} \left[ G_{2,3}^{3,0} \left( \frac{tU}{\Delta x} \mid \begin{matrix} 1,1 \\ 0,0, \frac{x}{\Delta x} \end{matrix} \right) + \Gamma\left(\frac{x}{\Delta x}, \frac{tU}{\Delta x}\right) \left( \log\left(\frac{tU}{\Delta x}\right) - \psi^{(0)}\left(\frac{x}{\Delta x}\right) \right) \right] \quad (5)$$

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

Application to MADE: Mean flow velocity  $U = 0.026$  m/d is derived analogously to the FOA model from DPIL-based geostatistical  $K$ -parameters (Bohling et al., 2016), resulting in  $K_{\text{eff}} = 2.28$  m/d (Zarlanga et al., 2018), head gradient of  $J = 0.0036$  and porosity of  $\theta = 0.31$ . The mass distribution follow from applying the pdf Equation 5 with  $\Delta x = 2.5l$  making use of the correlation length  $l = 9.1$  m from DPIL observations (Bohling et al., 2016). Calculation of mass with an implementation of the pdf Equation 5 in a computer algebra system is a matter of seconds.

## 4. Results and Comparison

### 4.1. Data Preparation and Display

Visualization of modeled mass distributions allows qualitative assessment and comparison of models. Perception of model performance is subjective depending on the display mode with critical aspects being the display scale, data upscaling (aggregation), and normalization. We present longitudinal mass distributions  $\langle \bar{m} \rangle$  at linear and logarithmic scales as well as in a cumulative form  $\langle M \rangle$ , to achieve a comprehensive display. The various display modes of the spatial mass distributions allow to interpret a few plume's specific features: (a) mass peak location and bulk behavior; (b) the tails (forefront and trailing zones), and (c) mass recovery. They are relevant to specific goals such as risk assessment and remediation.

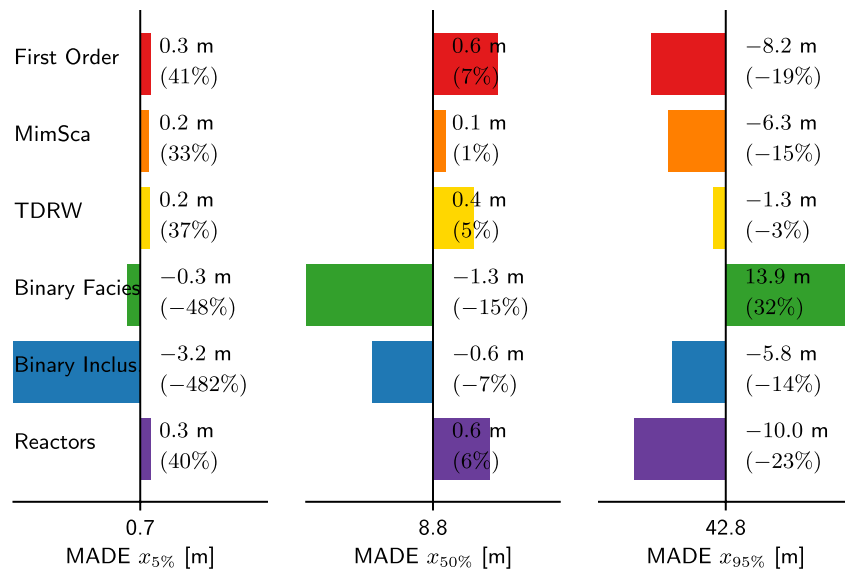
In line with the MADE experimental results, we remind that  $\langle \bar{m}(x, t) \rangle$  is aggregated over intervals of 10 m. We illustrate the effect by displaying the model outcomes at  $T = 126$  days in both forms, the fine scale  $\langle m(x, t) \rangle$ , and the upscaled  $\langle \bar{m}(x, t) \rangle$  (Figure 3) as well as  $\langle M \rangle$ . All modelled longitudinal mass distributions are normalized with respect to the injected mass such that the area beneath  $m$  or  $\bar{m}$  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 here for two reasons: (a) the skewed mass distribution leads to biased results when using a moment analysis and (b) the decreasing mass recovery in combination with the observation data uncertainty of the leading tail. The latter precludes the use of common fitness measures. Instead, we compare in Figure 4 recovery locations at  $T = 126$  days for 5%, 50%, 95% as predicted by models relative to MADE values of  $x = 0.7, 8.6, 42.8$  m. The recovery location  $x_{\alpha\%}$  is defined as the distance for which  $\alpha\%$  of the mass is upstream of  $x$ .

### 4.2. Comparison Between Models Prediction and MADE Experiment

Figure 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 the 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  $x_{95\%} \cong 60$  m, while for all the others models  $30 \lesssim x_{95\%} \lesssim 40$  m (see Figure 3e).



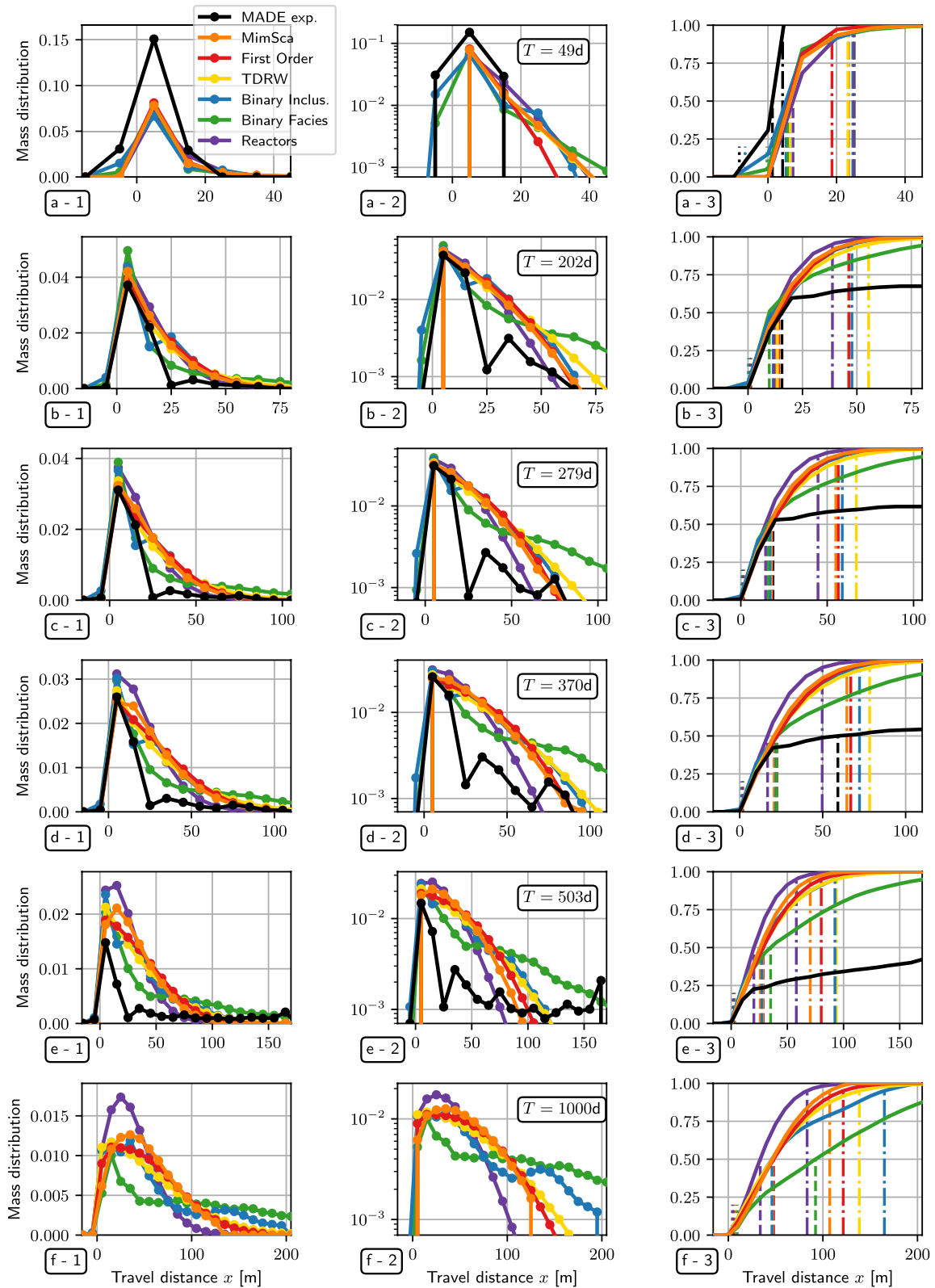
**Figure 4.** Model-experiment-comparison of recovery locations  $x$  for 5%, 50%, and 95% mass recovery (columns) at 126 days after injection where mass recovery was 99% in the experiment. Macrodispersion experiment (MADE) values at the  $x$ -axis (in m) and model values as absolute (in m) and relative (in %) difference in colored bars and in numbers.

The comparison between  $\langle m \rangle$  and the upscaled  $\langle \bar{m} \rangle$  in Figure 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 model predictions are greatly reduced. In particular, the predicted  $\langle \bar{m} \rangle$  agrees quite well with MADE, much better than  $\langle m \rangle$ , as far as visual inspection reveals. This is expectable though in the past models prediction at fine scale were compared with MADE (e.g., Dogan et al., 2014; Harvey & Gorelick, 2000). 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 m. It could be erroneously interpreted as upstream dispersion.

The quantitative results in Figure 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). This conclusion is in agreement with the results of Eggleston and Rojstaczer (2000), who found significant differences in the prediction of  $m(x, t)$  at MADE by different interpolation schemes for  $K$ ; such schemes led to quite different estimates of  $U$ , as manifested by the different arrival times of the peak. The Binary Facies differ by 15% due to the higher velocity  $U$ , as a result of the different conductivity statistics of the grain size analysis (Figure 2). The agreement is still good for  $x_{5\%}$  with some deviations for the Binary Inclusion model. Last, the models prediction relative to MADE of  $x_{95\%}$ , reflecting the “fast” moving solute, is more variable, but still within acceptable differences in practice.

Figure 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 difficult than for  $T = 126$  days because of the variable mass recovery, which is visible in Figure 5 in the cumulative mass panel  $\langle M \rangle$ . For none of the displayed times, experimental mass recovery is complete, explaining the apparent mismatch between experiment and models.

In spite of the incomplete mass recovery for times  $T \geq 202$  days, models agree reasonably well with both measured peak value and its distance from the injection zone. While for the portion upstream the peak position recovery was likely achieved, disagreement is visible at distances beyond 20 m from the injection zone (Figures 5b3–5e3). We attribute the reduced mass recovery for later times to the less dense sampling in the downstream zone (Figure 1a) in combination with not-sampled solute quickly moving in high conductivity channels (discussed in detail in Section 2.3). For late times, such as  $T = 503$  days the low mass recovery (43%) makes the comparison more problematic. Still, the peak location and also its value are within



**Figure 5.** Longitudinal mass distribution of models and macrodispersion experiment (MADE)-1 experiment at observation times 49, 202, 279, 370, and 503 days after injection (a–e) and for models at  $t = 1,000$  days (f) at linear scale (first column) and log-scale (second column). Third column displays cumulative mass with recovery locations of 5%, 50%, and 95% mass. Model results and MADE-1 experimental data differ in the recovery rates: 1 for models at all times while 2.06, 0.68, 0.62, 0.54, and 0.43 for the MADE-1 experimental data, respectively.

acceptable differences in practice. Despite variable mass recovery through time, we find the agreement of model predictions and experimental results on these key plume parameters to be encouraging.

As for intercomparison of models prediction, inspection of the robust cumulative mass  $\langle M \rangle$  at different times shows remarkable closeness except for Binary Facies. The latter overestimates the location pertaining to values of  $\langle M \rangle > 0.4$ . This enhanced tailing is attributed in part to the larger integral horizontal scale identified from granulometry analysis (Figure 2), as compared to that resulting from DPIL, which was used by 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 Section 5).

### 4.3. Prediction Beyond MADE Experiment

The important role of models is to provide prediction of future solute plumes development. In order to compare the outcome of the six 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  $T = 1,000$  days) predicted plume mass spatial distribution is displayed in Figure 5f. All models agree in the peak travel distance as a consequence 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 forefront tailing associated with fast moving solute, both Binary Facies and Binary Inclusions display longer tails with  $80 \lesssim x_{95\%} \lesssim 140$  m. Inspection of the cumulative distribution reveals again that prediction of models, except Binary Facies, are within a relatively narrow distribution.

## 5. Discussion

Longitudinal mass distributions (Figures 3 and 5) showed good agreement of all models in the general plume shape and mass peak, but observable differences among models. We attribute these differences to different conceptualizations of aquifer heterogeneity in combination with conductivity parameters from different field observation methods. Both aspects are discussed in detail in the following subsections. All modes share the same controlling physical processes (Section 3.1), despite various mathematical formulations.

In light of the different conductivity conceptualization and parametrization, we attribute the similarities in plume shape to the fact that models share the key features: mean flow velocity, a parameter reflecting log-conductivity variability, and taking into account horizontal correlations of conductivity values.

### 5.1. Modeling

The six models applied to the MADE setting can be divided into three groups based on conceptual similarities and in line with simulation results: (a) FOA, MIMSCA, and TDRW; (b) Binary Facies and Binary Inclusions; and (3) reactors.

#### 5.1.1. FOA, MIMSCA, and TDRW

The three models regard conductivity as a multi-Gaussian stationary random field, completely characterized by geostatistical parameters  $K_G$ ,  $\sigma_Y^2$ ,  $I$ ,  $I_v$ . The mean flow velocity is given by  $U = K_{\text{eff}}/\theta$  (Zarlenga et al., 2018). Models differ in conceptualization and computational complexity. FOA is analytical, leading to an inverse Gaussian mass distribution  $\langle m \rangle$  which satisfies an ADE (Equation 1) with macrodispersivity derived analytically from first order approximation. MIMSCA is semianalytical, based on summation of travel time through inclusions of random  $K$ . TDRW is seminumerical, with the velocity field derived by Monte Carlo simulations, while transport is based on an approximation of the Lagrangian velocity field. Computational effort for solution of transport is low for all models, being in the order of minutes for a common PC.

The longitudinal mass distributions of all three models are very close and in good agreement with the bulk of MADE experimental data. Plume travel distances are similar (Figure 4), the same holds for mass peak locations and values even without data aggregation (Figure 3), which we attribute to the use of similar flow velocities.

Hence, models with a multi-Gaussian conductivity conceptualization provide a robust prediction of plume behavior. Reliable predictions depend less on the specific model implementation but rather on selecting appropriate velocity  $U$  and heterogeneity representative parameters  $\sigma_Y^2$  and  $I$ .

### 5.1.2. Binary Facies and Binary Inclusions

Both numerical models use a binary conceptualization of heterogeneity: hydraulic conductivity is modeled by adopting 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. Mean values of flow velocity  $U$  and mass distributions are achieved from ensemble averaging, which also allows obtaining the statistical moments of these quantities.

The two models differ in dimensionality, binary structure generation, and measurement input data: binary Facies is 3D and random facies geometries are generated by transitional probability using TPROGS (Carle, 1999). It can be interpreted as a simplified form of the facies approach applied by Bianchi and Zheng (2016) for MADE2, by considering two facies instead of five. All the structural parameters are obtained from granulometry measurements. The binary inclusions approach simplifies further the 3D random structure by considering regular inclusions in two dimensions. It consists of identical rectangular inclusions of conductivity  $K_2$  submerged in the  $K_1$  matrix for two dominant deterministic zones of different mean conductivity which could be delineated from head observations. Structural parameters are derived from pumping tests and few flowmeter measurements in a procedure adapted to the particular features of the MADE site.

Both models agree with the MADE1 experimental data. They are particularly able to reproduce the spreading of solute in the upstream zone of the plume, as well as migration of mass at low concentrations far downstream. While both models predict the peak mass and location, the Binary Facies displays enhanced downstream transport of tracer which is related to the higher mean flow velocity implied by the grain size conductivity statistics (Figure 2). The mean flow velocity of the Binary Inclusions model is slightly smaller than that used by FOA, MIMSCA, and TDRW from DPIL data due to the explicit representation of the low conductivity injection zone. The favorable comparison with MADE1 shows that numerical ADE models provide robust predictions of plume mass distribution derived through a Monte Carlo approach. This confirms results of Dogan et al. (2014) obtained with the aid of a detailed representation of conductivity distribution of a portion of the MADE site. Furthermore, the results for binary structures shows that the key element is accounting for conductivity contrasts rather than high resolution and a detailed specification of conductivity through conditioning.

The semianalytical models are relatively simple for computations but use a more extensive DPIL data set, whereas Binary Inclusions and Binary Facies modeling approaches are heavier on computation but use more readily accessible data sets.

### 5.1.3. Reactors

This model of different conceptualization was motivated by its general use in engineering and convenience for reactive transport modeling. Central parameter is the velocity  $U$ , which is derived from effective conductivity  $K_{\text{eff}}$  (Zarlenga et al., 2018). The second parameter, the velocity longitudinal correlation length is related to the conductivity integral scale  $I$ . The degree of heterogeneity quantified by  $\sigma_Y^2$  is not included. Instead, the series of reactors aggregate the inherent dispersion making the model not a general candidate for modeling advective transport in a heterogeneous aquifer. Its analytical nature makes the application of the model very simple, allowing calculations of mass within seconds.

The mass distribution and the aggregated one agrees well with the observed one at MADE, though it overvalues the peak, particularly at larger times, for example, by a factor of 1.5 compared to all other models for  $T = 1,000$  days. The transferability of the reactors model to other sites need to be tested. However, the findings confirm the predominant role of the mean velocity  $U$  and correlation scale  $I$  for robust transport prediction.



## 5.2. Data Selection

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

The semi-analytical models FOA, MIMSCA, TDRW, and Reactors used DPIL observations to estimate the flow velocity and the plume spreading. The Binary Facies model used granulometry data while the Binary Inclusions model used pumping test estimates, information from head maps and a few flowmeter data.

The mean velocity  $U$  of all models rely on the measured mean head gradient  $J$  and the porosity  $\theta$ . The models differ primarily in the use of the  $K$  data. Still, the resulting estimates of the mean velocity  $U$  are relatively close: 0.026 m/d for FOA, MIMSCA, and Reactors; 0.019 m/d for TDRW, 0.025 m/d for Binary Inclusions, and 0.079 m/d for Binary Facies. Even the deviation for Binary Facies is within an acceptable range, various approximations notwithstanding. Thus, the estimates of velocity are quite robust, which explains the relative closeness of the predicted and measured locations of the peak of  $\bar{m}$  in Figures 3 and 5.

Similarly, the prediction of plume spreading is quite robust. A potential explanation for the relative closeness of model results despite different data use is the closeness of the magnitude of the asymptotic longitudinal macrodispersivity  $\alpha_L = \sigma_Y^2 I$ . The resulting values for the different values of log-conductivity variance  $\sigma_Y^2$  and the horizontal correlation scale  $I$  (Figure 2) are 53.7, 54.1 and 49.6 m for DPIL, flowmeter and grain size, respectively.

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

## 6. Summary and Conclusions

With a variety of hydraulic data available, the MADE site provides a unique opportunity for a comparative analysis of predictive modeling. The present study takes advantage of these possibilities and compares predictions of plume transport by six previously developed models that use different information to characterize the spatial variability of hydraulic conductivity. Thus, the FOA (Fiori et al., 2017), the MIM and self-consistent approximation (MIMSCA) (Fiori et al., 2013), and the TDRW (Dentz et al., 2020) models utilize extensive hydraulic profiling data (Bohling et al., 2016), for inferring geostatistical parameters. In contrast, the binary inclusions model (Zech et al., 2021) uses pumping tests and a few flowmeter data whereas the Binary Facies model relies on granulometric data.

Models differ in the conceptualization of aquifer structure and theoretical formulations. They further rely on different field data input and differ in their computational effort. Common features of the models are: flow is steady, uniform in the mean and driven by a constant head gradient; solute spreading is caused by aquifer conductivity heterogeneity; models rely on structural and flow data with no calibration against transport observations; plume mass distribution is assumed to be ergodic, meaning that the modeled mean relative mass distribution is compared with the upscaled measured mass at MADE at a few times. Model predictions of advective transport is compared to the MADE-1 transport experiment findings as quantified by the longitudinal mass distributions for time snapshots from 49 days upto 503 days. A model comparison is included at 1,000 days, which is beyond the period of measurements. Despite decreasing mass recovery of MADE-1 observations, mass data is not normalized. The seeming mass loss for times beyond 126 days is attributed to limited sampling and thus not taken into account in the models. The discrepancy between models and observed total mass strongly impacts the perception of match of model and experimental data at late times. However, mass distribution peak value and location as well as general plume shape at MADE are well predicted by all models.

The main and encouraging result for practitioners is that models conceptualized by relatively simple flow and transport equations, but explicitly taking aquifer heterogeneity into account, predict the complex MADE1 mass distribution reasonably well. The adopted measures of the solute plumes are robust and models are reliable as long as they are underlain by a few basic parameters: mean velocity  $U$ , a parameter reflecting log-conductivity variability and one reflecting the horizontal degree of log-conductivity correlation. 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.

The purpose of comparing models was not to recommend adopting a particular one, but to show that the robustness of model predictions allows the use of a variety of observation data and representations of hydraulic conductivity structures. The choice of models compared here is dictated by the requirement of their being predictive without reliance on transport data. Consequently, classes of models relying on parameter calibration by transport experiments, such as mass transfer models in porous media are not considered. However, our model choice does not preclude investigations of model classes which use calibration to transport observations when they are available.

Although the model comparison is limited to the data and setting at MADE, we surmise that it indicates applicability to other sites as well as for predictive modeling of transport. Rendering the above conclusions of general validity requires extending the study to other field sites. Synthetic examples, for instance, log-normal conductivity structures with different connectivities (Fiori et al., 2017) or facies structures (Carle & Fogg, 1996) would be an alternative given the limitation of controlled experiments at highly heterogeneous sites. Our results can provide guidance to the design of such new subsurface transport experiments.

## Data Availability Statement

Data on MADE observations can be found in the references. Longitudinal mass distributions of all presented transport models and scripts used in the paper are available in the GitHub repository AlrauneZ/MADE-model-comparison (Zech, 2021).

## Acknowledgments

The authors thank Marco Bianchi and Boris Baeumer for their support during the development of the study. A. Fiori and A. Bellin 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. M. Dentz 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).

## References

- Adams, E. E., & Gelhar, L. W. (1992). Field study of dispersion in a heterogeneous aquifer: 2. Spatial moments analysis. *Water Resources Research*, 28, 3293–3307. <https://doi.org/10.1029/92WR01757>
- Barlebo, H. C., Hill, M. C., & Rosbjerg, D. (2004). Investigating the macrodispersion experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model. *Water Resources Research*, 40, W04211. <https://doi.org/10.1029/2002WR001935>
- Benson, D. A., Schumer, R., Meerschaert, M. M., & Wheatcraft, S. W. (2001). Fractional Dispersion, Lévy Motion, and the MADE Tracer Tests. *Transport in Porous Media*, 42, 211–240. [https://doi.org/10.1023/A:100673300213110.1007/978-94-017-1278-1\\_11](https://doi.org/10.1023/A:100673300213110.1007/978-94-017-1278-1_11)
- Berkowitz, B., & Scher, H. (2001). The role of probabilistic approaches to transport theory in heterogeneous media. *Transport in Porous Media*, 42, 241–263. [https://doi.org/10.1023/A:100678501897010.1007/978-94-017-1278-1\\_12](https://doi.org/10.1023/A:100678501897010.1007/978-94-017-1278-1_12)
- Bianchi, M., & Zheng, C. (2016). A lithofacies approach for modeling non-Fickian solute transport in a heterogeneous alluvial aquifer. *Water Resources Research*, 52, 552–565. <https://doi.org/10.1002/2015WR018186>
- Boggs, J. M., Schroeder, J., & Young, S. (1995). *Data to support model development for natural attenuation study* (Tech. Rep. No. WR28-2-520-197). Norris, Tennessee: Tennessee Valley Authority Engineering Laboratory.
- Boggs, J. M., Young, S., Benton, D., & Chung, Y. (1990). *Hydrogeologic characterization of the MADE site* Tech. Rep. No. EN-6915. Palo Alto, CA: EPRI.
- Boggs, J. M., Young, S. C., Beard, L. M., Gelhar, L. W., Rehfeldt, K. R., & Adams, E. E. (1992). Field study of dispersion in a heterogeneous aquifer: 1. Overview and site description. *Water Resources Research*, 28, 3281–3291. <https://doi.org/10.1029/92WR01756>
- Bohling, G. C., Liu, G., Dietrich, P., & Butler, J. J. (2016). Reassessing the MADE direct-push hydraulic conductivity data using a revised calibration procedure. *Water Resources Research*, 52, 8970–8985. <https://doi.org/10.1002/2016WR019008>
- Bohling, G. C., Liu, G., Knobbe, S. J., Reboulet, E. C., Hyndman, D. W., Dietrich, P., & Butler, J. J. (2012). Geostatistical analysis of centimeter-scale hydraulic conductivity variations at the MADE site. *Water Resources Research*, 48, W02525. <https://doi.org/10.1029/2011WR010791>
- Carle, S. F. (1999). *T-PROGS: Transition probability geostatistical software, version 2.1* (p. 84). Livermore, CA: University of California, Davis.
- Carle, S. F., & Fogg, G. E. (1996). Transition probability-based indicator geostatistics. *Mathematical Geology*, 28, 453–476. <https://doi.org/10.1007/BF02083656>
- Comolli, A., Hakoun, V., & Dentz, M. (2019). Mechanisms, upscaling, and prediction of anomalous dispersion in heterogeneous porous media. *Water Resources Research*, 55, 8197–8222. <https://doi.org/10.1029/2019WR024919>
- Cvetkovic, V., Fiori, A., & Dagan, G. (2014). Solute transport in aquifers of arbitrary variability: A time-domain random walk formulation. *Water Resources Research*, 50, 5759–5773. <https://doi.org/10.1002/2014WR015449>
- Dagan, G. (1989). *Flow and transport on porous formations* (p. 465). New York: Springer. <https://doi.org/10.1007/978-3-642-75015-1>

- Dagan, G., & Cvetkovic, V. (1993). Spatial moments of a kinetically sorbing solute plume in a heterogeneous aquifer. *Water Resources Research*, 29, 4053–4061. <https://doi.org/10.1029/93WR02299>
- Dagan, G., & Fiori, A. (2003). Time-dependent transport in heterogeneous formations of bimodal structures: 1. The model. *Water Resources Research*, 39(5), 1112. <https://doi.org/10.1029/2002WR001396>
- Dentz, M., Comolli, A., Hakoun, V., & Hidalgo, J. J. (2020). Transport upscaling in highly heterogeneous aquifers and the prediction of tracer dispersion at the MADE site. *Geophysical Research Letters*, 47, e2020GL088292. <https://doi.org/10.1029/2020GL088292>
- Dentz, M., Kang, P. K., Comolli, A., Le Borgne, T., & Lester, D. R. (2016). Continuous time random walks for the evolution of Lagrangian velocities. *Physical Review Fluids*, 1, 074004. <https://doi.org/10.1103/PhysRevFluids.1.074004>
- Dietrich, P., Butler, J. J., & Faiss, K. (2008). A rapid method for hydraulic profiling in unconsolidated formations. *Ground Water*, 46, 323–328. <https://doi.org/10.1111/j.1745-6584.2007.00377.x>
- Dogan, M., Van Dam, R. L., Liu, G., Meerschaert, M. M., Butler, J. J., Bohling, G. C., et al. (2014). Predicting flow and transport in highly heterogeneous alluvial aquifers. *Geophysical Research Letters*, 41, 7560–7565. <https://doi.org/10.1002/2014GL061800>
- Eggleston, J. R., & Rojstaczer, S. A. (2000). Can we predict subsurface mass transport? *Environmental Science & Technology*, 34, 4010–4017. <https://doi.org/10.1021/es000903s>
- Feehley, C. E., Zheng, C., & Molz, F. J. (2000). A dual-domain mass transfer approach for modeling solute transport in heterogeneous aquifers: Application to the macrodispersion experiment (MADE) site. *Water Resources Research*, 36, 2501–2515. <https://doi.org/10.1029/2000WR900148>
- Fetter, C. W., Boving, T., & Kremer, D. (2018). *Contaminant hydrogeology* (3rd ed., p.647). Long Grove, IL: Waveland Pr Inc.
- Fiori, A. (2014). Channeling, channel density and mass recovery in aquifer transport, with application to the MADE experiment. *Water Resources Research*, 50, 9148–9161. <https://doi.org/10.1002/2014WR015950>
- Fiori, A., Dagan, G., Jankovic, I., & Zarlenga, A. (2013). The plume spreading in the MADE transport experiment: Could it be predicted by stochastic models? *Water Resources Research*, 49, 2497–2507. <https://doi.org/10.1002/wrcr.20128>
- Fiori, A., Jankovic, I., & Dagan, G. (2003). Flow and transport through two-dimensional isotropic media of binary conductivity distribution. Part 1: Numerical methodology and semi-analytical solutions. *Stochastic Environmental Research and Risk Assessment*, 17(6), 370–383. <https://doi.org/10.1007/s00477-003-0166-0>
- Fiori, A., Janković, I., & Dagan, G. (2006). Modeling flow and transport in highly heterogeneous three-dimensional aquifers: Ergodicity, Gaussianity, and anomalous behavior-2. Approximate semianalytical solution. *Water Resources Research*, 42, W06D13. <https://doi.org/10.1029/2005WR004752>
- Fiori, A., Zarlenga, A., Bellin, A., Cvetkovic, V., & Dagan, G. (2019). Groundwater contaminant transport: Prediction under uncertainty, with application to the MADE transport experiment. *Frontiers in Environmental Science*, 7. <https://doi.org/10.3389/fenvs.2019.00079>
- Fiori, A., Zarlenga, A., Jankovic, I., & Dagan, G. (2017). Solute transport in aquifers: The comeback of the advection dispersion equation and the First Order Approximation. *Advances in Water Resources*, 110, 349–359. <https://doi.org/10.1016/j.advwatres.2017.10.025>
- Gelhar, L. (1993). *Stochastic subsurface hydrology* (p. 390). : Prentice Hall, Englewood Cliffs.
- Gómez-Hernández, J. J., Butler, J. J., Fiori, A., Bolster, D., Cvetkovic, V., Dagan, G., & Hyndman, D. (2017). Introduction to special section on modeling highly heterogeneous aquifers: Lessons learned in the last 30 years from the MADE experiments and others. *Water Resources Research*, 53, 2581–2584. <https://doi.org/10.1002/2017WR020774>
- Hakoun, V., Comolli, A., & Dentz, M. (2019). Upscaling and prediction of Lagrangian velocity dynamics in heterogeneous porous media. *Water Resources Research*, 55, 3976–3996. <https://doi.org/10.1029/2018WR023810>
- Harvey, C., & Gorelick, S. M. (2000). Rate-limited mass transfer or macrodispersion: Which dominates plume evolution at the Macrodispersion Experiment (MADE) site? *Water Resources Research*, 36, 637–650. <https://doi.org/10.1029/1999WR900247>
- Jankovic, I., Fiori, A., & Dagan, G. (2003). Flow and transport in highly heterogeneous formations: 3. Numerical simulations and comparison with theoretical results. *Water Resources Research*, 39(9), 1270. <https://doi.org/10.1029/2002WR001721>
- Jankovic, I., Maghrebi, M., Fiori, A., & Dagan, G. (2017). When good statistical models of aquifer heterogeneity go right: The impact of aquifer permeability structures on 3D flow and transport. *Advances in Water Resources*, 100, 199–211. <https://doi.org/10.1016/j.advwatres.2016.10.024>
- Kreft, A., & Zuber, A. (1978). On the physical meaning of the dispersion equation and its solutions for different initial and boundary conditions. *Chemical Engineering Science*, 33, 1471–1480. [https://doi.org/10.1016/0009-2509\(78\)85196-3](https://doi.org/10.1016/0009-2509(78)85196-3)
- Molin, S., & Cvetkovic, V. (2010). Microbial risk assessment in heterogeneous aquifers: 1. Pathogen transport. *Water Resources Research*, 46. <https://doi.org/10.1029/2009WR008036>
- Noetinger, B., Roubinet, D., Russian, A., Le Borgne, T., Delay, F., Dentz, M., et al. (2016). Random walk methods for modeling hydrodynamic transport in porous and fractured media from pore to reservoir scale. *Transport in Porous Media*, 115, 345–385. <https://doi.org/10.1007/s11242-016-0693-z>
- Pirot, G., Renard, P., Huber, E., Straubhaar, J., & Huggenberger, P. (2015). Influence of conceptual model uncertainty on contaminant transport forecasting in braided river aquifers. *Journal of Hydrology*, 531, 124–141. <https://doi.org/10.1016/j.jhydrol.2015.07.036>
- Pollock, D. W. (2012). *User guide for MODPATH version 6: A particle-tracking model for modfl*. Reston, VA: U.S. Geological Survey. <https://doi.org/10.3133/tm6A41>
- Rehfeldt, K. R., Boggs, J. M., & Gelhar, L. W. (1992). Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity. *Water Resources Research*, 28, 3309–3324. <https://doi.org/10.1029/92WR01758>
- Rehfeldt, K. R., Hufschmied, P., Gelhar, L. W., & Schaefer, M. (1989). *Measuring hydraulic conductivity with the borehole flowmeter (Tech. Rep. No. EN-6511)*. Palo Alto, CA: EPRI.
- Riva, M., Guadagnini, L., & Guadagnini, A. (2010). Effects of uncertainty of lithofacies, conductivity and porosity distributions on stochastic interpretations of a field scale tracer test. *Stochastic Environmental Research and Risk Assessment*, 24, 955–970. <https://doi.org/10.1007/s00477-010-0399-7>
- Rubin, Y. (2003). *Applied stochastic hydrogeology* (p. 391). New York: Oxford University Press. <https://doi.org/10.1093/oso/9780195138047.003.0006>
- Salamon, P., Fernández-García, D., & Gómez-Hernández, J. J. (2007). Modeling tracer transport at the MADE site: The importance of heterogeneity. *Water Resources Research*, 43, W08404. <https://doi.org/10.1029/2006WR005522>
- Suribhatla, R., Jankovic, I., Fiori, A., Zarlenga, A., & Dagan, G. (2011). Effective conductivity of an anisotropic heterogeneous medium of random conductivity distribution. *Multiscale Modeling and Simulation*, 9, 933–954. <https://doi.org/10.1137/100805662>
- Vienken, T., & Dietrich, P. (2011). Field evaluation of methods for determining hydraulic conductivity from grain size data. *Journal of Hydrology*, 400, 58–71. <https://doi.org/10.1016/j.jhydrol.2011.01.022>

- Zarlenga, A., Janković, I., Fiori, A., & Dagan, G. (2018). Effective hydraulic conductivity of three-dimensional heterogeneous formations of lognormal permeability distribution: The impact of connectivity. *Water Resources Research*, *54*, 2480–2486. <https://doi.org/10.1002/2017WR022141>
- Zech, A. (2021). *Made-model-comparison*. Zenodo. <https://doi.org/10.5281/zenodo.4436610>
- Zech, A., Attinger, S., Bellin, A., Cvetkovic, V., Dietrich, P., Fiori, A., et al. (2019). A critical analysis of transverse dispersivity field data. *Ground Water*, *57*, 632–639. <https://doi.org/10.1111/gwat.12838>
- Zech, A., Attinger, S., Cvetkovic, V., Dagan, G., Dietrich, P., Fiori, A., et al. (2015). Is unique scaling of aquifer macrodispersivity supported by field data? *Water Resources Research*, *51*, 7662–7679. <https://doi.org/10.1002/2015WR017220>
- Zech, A., Dietrich, P., Attinger, S., & Teutsch, G. (2021). A field evidence model: How to predict transport in heterogeneous aquifers at low investigation level. *Hydrology and Earth System Sciences*, *25*, 1–15. <https://doi.org/10.5194/hess-25-1-2021>
- Zheng, C., Bianchi, M., & Gorelick, S. M. (2011). Lessons learned from 25 years of research at the MADE site. *Ground Water*, *49*, 649–662. <https://doi.org/10.1111/j.1745-6584.2010.00753.x>