Hydrology and
Earth System
Sciences

A field evidence model: how to predict transport in heterogeneous aquifers at low investigation level

Alraune Zech[1,2], **Peter Dietrich**[1,3], **Sabine Attinger**[1,4], and **Georg Teutsch**[1]

[1]Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany
[2]Department of Earth Science, Utrecht University, Utrecht, the Netherlands
[3]Center for Applied Geoscience, Eberhard Karls University Tübingen, Tübingen, Germany
[4]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

Correspondence: Alraune Zech (a.zech@uu.nl)

Abstract. Aquifer heterogeneity in combination with data scarcity is a major challenge for reliable solute transport prediction. Velocity fluctuations cause non-regular plume shapes with potentially long-tailing and/or fast-travelling mass fractions. High monitoring cost and a shortage of simple concepts have limited the incorporation of heterogeneity into many field transport models up to now.

We present an easily applicable hierarchical conceptualization strategy for hydraulic conductivity to integrate aquifer heterogeneity into quantitative flow and transport modelling. The modular approach combines large-scale deterministic structures with random substructures. Depending on the modelling aim, the required structural complexity can be adapted. The same holds for the amount of monitoring data. The conductivity model is constructed step-wise following field evidence from observations, seeking a balance between model complexity and available field data. The starting point is a structure of deterministic blocks, derived from head profiles and pumping tests. Then, subscale heterogeneity in the form of random binary inclusions is introduced to each block. Structural parameters can be determined, for example, from flowmeter measurements or hydraulic profiling.

As proof of concept, we implemented a predictive transport model for the heterogeneous MADE site. The proposed hierarchical aquifer structure reproduces the plume development of the MADE-1 transport experiment without calibration. Thus, classical advection–dispersion equation (ADE) models are able to describe highly skewed tracer plumes by incorporating deterministic contrasts and effects of connectivity in a stochastic way without using uni-modal hetero-geneity models with high variances. The reliance of the conceptual model on few observations makes it appealing for a goal-oriented site-specific transport analysis of less well investigated heterogeneous sites.

1 Introduction

Groundwater is extensively used worldwide as a major drinking water resource and consequently needs to be protected with respect to quantity and quality. Increasing pressure on the quality originates from the intensification of agriculture using agrochemicals (non-point sources) and an increased urbanization with the resulting solid and liquid waste and contaminant spills from industrial applications (point sources).

Essential for groundwater protection is the quantitative analysis of the fate and transport of various contaminants in the groundwater body. This can be either for a provisional risk assessment or for the clean-up of an already existing groundwater contamination. Numerical models are common tools to quantify the flow and transport, where partial differential equations are solved using initial and boundary conditions.

For simplicity, we restrict ourselves to saturated flow and transport of a dissolved, non-reactive contaminant. The governing equation for its concentration $C(\boldsymbol{x}, t)$ is the advection–dispersion equation (ADE) (Bear, 1972):

$$\frac{\partial C(\boldsymbol{x}, t)}{\partial t} = -\boldsymbol{u}(\boldsymbol{x}, t) \cdot \nabla C(\boldsymbol{x}, t) + \nabla (\mathbf{D} \cdot \nabla C(\boldsymbol{x}, t)), \qquad (1)$$

given in space $\boldsymbol{x} = (x, y, z)$ and time t. $\mathbf{D}$ is the dispersion coefficient tensor, and $\boldsymbol{u}(\boldsymbol{x}, t)$ is the Darcy velocity vector. The latter is a function of the hydraulic gradient J and the heterogeneous hydraulic conductivity $K(\boldsymbol{x})$ through Darcy's law. A proper description of the velocity field $\boldsymbol{u}(\boldsymbol{x}, t)$, thus aquifer heterogeneity, is crucial for predicting the concentration distribution $C(\boldsymbol{x}, t)$.

The adequate parameterization of the heterogeneous conductivity $K(\boldsymbol{x})$ poses a significant challenge in the practical model setup due to data scarcity. Numerous deterministic and stochastic approaches have been developed to incorporate the effects of spatial heterogeneity of conductivity on flow and transport, particularly in the context of stochastic subsurface hydrology (Koltermann and Gorelick, 1996). Representing conductivity by an effective uniform value is convenient for aquifers of low heterogeneity since it can be inferred from pumping tests with decent monitoring effort. But predicting transport in aquifers of significant variability fails when neglecting local effects of heterogeneity and preferential flow.

Stochastic methods allow the heterogeneity to be resolved and thus capture the induced uncertainty in flow and transport predictions. However, the amount of observation data required is usually high, depending on the method's complexity. Common methods are (i) Kriging (Kitanidis, 2008), (ii) Gaussian random fields (Dagan, 1989; Gómez-Hernández and Gorelick, 1989; Zinn and Harvey, 2003), potentially combined with Kriging for conditioning to observations; (iii) indicator/hydrofacies models (Journel and Gómez-Hernández, 1993; Carle and Fogg, 1996; Fogg et al., 2000); or (iv) multi-point statistics/training images (Strebelle, 2002; Renard et al., 2011; Linde et al., 2015).

For many unconsolidated sediments, field observations showed that conductivity is approximately log-normal (Delhomme, 1979; Gelhar, 1993; Rubin, 2003), characterized by the geometric mean K_G and the log-conductivity variance σ_Y^2. Variogram analysis provides structural parameters such as correlation length ℓ and anisotropy ratio e based on spatially distributed observations, e.g. from flowmeter, permeameter, or injection logging. Despite increased efficiency in exploration methods, data are not even sufficient for variogram analysis in most practical cases, thus hampering the practical application of Kriging and Gaussian random fields. Alternatively, hydrofacies models use indicator geostatistics with transition probability to generate geological heterogeneity structures. Although conceptually different, the required amount of input data is similarly high. Multi-point statistical methods provide heterogeneity structures of high geological realism, when training images are available. Although satellite data might provide areal training images, vertical structures rely on extensive literature on geology or outcrop studies. Both are hardly available at the scale representative of plume transport, impeding the method's use at hydrogeological sites.

A recent debates series (Rajaram, 2016; Fiori et al., 2016; Fogg and Zhang, 2016; Cirpka and Valocchi, 2016; Sanchez-

Vila and Fernàndez-Garcia, 2016) outlined the gap between the advanced research in stochastic subsurface hydrology and its application in the practice of groundwater flow and transport modelling. We see a significant reason in the lack of data for complex stochastic models. Thus, we advocate the use of hierarchical approaches, combining deterministic and stochastic hydraulic conductivity conceptualization. Hierarchical approaches are regularly used in reservoir modelling (Damsleth et al., 1992; Smith et al., 2001; Bryant and Flint, 2009), particularly for consolidated sediments. Aside from qualitative approaches for multi-scale heterogeneity representation, e.g. Neton et al. (1994), Herweijer (1997), or Koltermann and Gorelick (1996) (and references therein), only few quantitative approaches were proposed, such as generating sequences of facies assemblages using indicator geostatistics and transition probability at various scales (Weissmann and Fogg, 1999; Proce et al., 2004) or combining training images for large-scale facies realizations with variogram-based geostatistical methods for random intrafacies permeability (Huysmans and Dassargues, 2009). Both approaches show a high level of model complexity and required (hydro)geological input data.

Here, we present a parsimonious hierarchical heterogeneity conceptualization which is easy to apply in quantitative models for predicting flow and solute transport. In a deterministic–stochastic framework we combine descriptive zonation with statistical methods, following the lines of Gómez-Hernández and Gorelick (1989). The goal is to optimize the aquifer structure setup given the simulation target constrained by the available field data. Therefore, we aim to provide tools making aquifer heterogeneity more accessible for practical applications, including hands-on software. The approach is based on the fact that subsurface heterogeneity can be generally classified into (a) larger scale dominant features which primarily determine the general flow direction together with the average groundwater flow velocity and (b) smaller scale features which are responsible for the dispersion, more specifically, the spatial spreading of a solute.

We create a deliberate connection between the model parameterization requirements and the field characterization methods beyond a single monitoring method. Pumping tests, for example, are best suited to determine the spatially averaged transmissivity, i.e. hydraulic conductivity, even in a heterogeneous aquifer environment (Herweijer, 1996; Zech et al., 2016). Together with head gradients from piezometric level maps, this yields good estimates of the mean groundwater flow velocities. High-resolution, small-scale borehole logs of hydraulic conductivity (e.g. from flowmeter or direct push methods) provide information on conductivity variability and consequently the dispersion parameters needed. Here, we consider two stochastic methods representing spatial variability: Gaussian random fields, which require distributed observation data for a variogram analysis, and a simplistic binary structure, which relies only on a few (e.g. two–four) well logs but takes parametric uncertainty into account. The

latter is developed as an option for less investigated sites, only requiring a decent amount of field data from standard monitoring methods for heterogeneous aquifer modelling.

We demonstrate the methodology for MADE, a heterogeneous, well investigated research site (e.g. Boggs et al., 1990; Zheng et al., 2011; Gómez-Hernández et al., 2017). Following our adaptive approach, we use various amounts and types of hydraulic observation data for heterogeneity conceptualization to construct numerical transport models. Predictions are independently evaluated against field tracer data from the MADE-1 experiment (Boggs et al., 1992). We do not reconstruct the actual conductivity structure at MADE but predict tracer plume behaviour following a Monte Carlo approach devoid of calibration. Model results show good agreement with observed plume data, also compared to other predictive transport models for MADE (e.g. Salamon et al., 2007; Fiori et al., 2013, 2017; Bianchi and Zheng, 2016). In this line, we provide an alternative approach for predictive transport modelling at a significantly heterogeneous site with a simple conceptualization and decent observation effort.

The course of the paper is the following: Sect. 2 features the approach in light of different modelling aims. Section 3 is dedicated to the application of the methodology for the MADE aquifer. We close with a summary and conclusions in Sect. 4.

2 Approach

Large-scale hydraulic structures of hundreds of metres or more determine the groundwater flow direction and magnitude in combination with groundwater catchment boundaries. Subsequently, they set the mean transport velocity. This is the key parameter to predict the location of the bulk mass of substances dissolved in the groundwater when input conditions are known.

Variations of hydraulic properties on intermediate scale, in the range of tens of metres, generate spatially variable flow fields. They also render transport velocities variable at these scales, resulting in larger spreading of plumes. This is particularly important for modelling tailing or leading mass fronts. Fluctuations on scales smaller than these intermediate scales have a blending effect, generally increasing local mixing and enhancing dispersion (Werth et al., 2006).

Following this conceptual view, we generate hydraulic conductivity fields composed of three components: Module (A), (B), and (C), which capture the effects at large-, intermediate-, and small-scale heterogeneity, respectively. Each component is selected according to the model aim and the data at hand to parameterize the hydraulic conductivity for this component.

The procedure is exemplified for the MADE site. This significantly heterogeneous site was intensively investigated with various measurement devices providing many different data sets, such as pumping tests and flowmeter and DPIL (direct push injection logging) measurements (Boggs et al., 1990; Bohling et al., 2016). Detailed information on MADE can be found in Sect. 3 and the Supplement.

In the approach, we consider several steps:

1. specifying the aim of the model (what do we want to predict?);

2. selecting processes and process components which need to be accounted for in the model (what does this imply for the conceptualization of hydraulic conductivity?);

3. selecting suitable measurement methods (which method can deliver the data needed for parameterizing hydraulic conductivity with affordable effort?);

4. conceptualizing hydraulic conductivity;

5. calculating flow and transport.

Before specifying the hydraulic conductivity components, Modules (A), (B), and (C), we illustrate our concept, discussing two exemplary model aims.

2.1 Exemplary model aims

2.1.1 Model aim "mean arrival"

1. *Aim.* The mean arrival of a contaminant from a point source is predicted.

2. *Processes.* Regional groundwater movement and direction and magnitude of flow are estimated, making use of the groundwater flow equation and Darcy's law. Transport is modelled by advection. For the sake of simplicity, we do not consider reactivity.

3. *Field characterization.* Regionalized groundwater-level measurements provide the direction and magnitude of the hydraulic gradient. It is critical to outline areas of different gradients (zones), indicating regional hydraulic conductivity trends and large-scale heterogeneity. Pumping tests can provide independent values of effective transmissivity within each zone.

4. *Conceptualization of hydraulic conductivity.* Conductivity is considered homogeneous within each large-scale zone. Effects of heterogeneity are captured in effective parameters representing average flow behaviour, e.g. determined from pumping tests.

5. *Solving flow and transport.* Flow is solved either analytically, e.g. for one or two zones of different effective hydraulic conductivity, or numerically in the case of a more complex spatial distribution of zones. Transport can be determined, making use of analytical or numerical solutions of the ADE according to initial and boundary conditions.
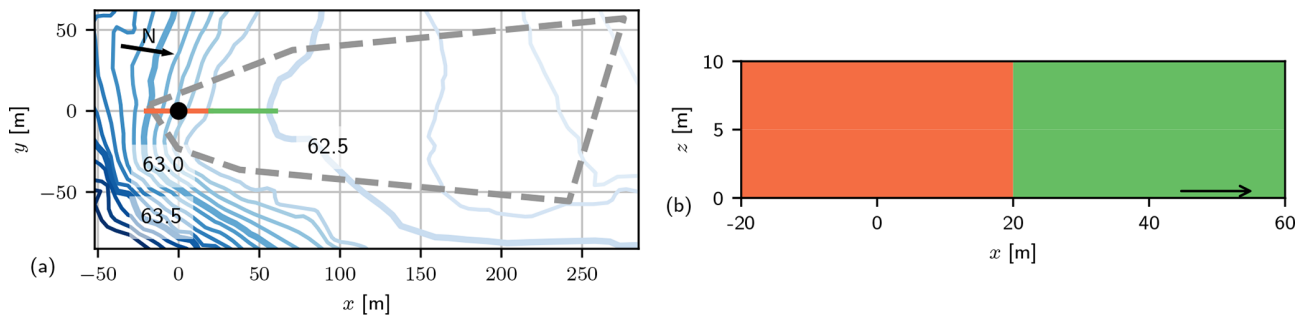
Figure 1. (a) Potentiometric surface map of head measurements according to Boggs et al. (1990). The orange–green line indicates the location of the cross section displayed in **(b)**: concept (Module A) for large conductivity structure with deterministic zones of low (orange) and high (green) conductivity. The arrow indicates the flow direction. The location of the interface between structures corresponds to a change in the hydraulic head pattern at $x = 20$ m.

Example MADE

The piezometric surface map of MADE (Boggs et al., 1992; Fig. 3) shows a significant non-uniform hydraulic head pattern. At 20 m downstream of the injection location, head isolines reduce abruptly. The reproduced head contours in Fig. 1a allow two major zones to be delineated: an area of low conductivity upstream (left) and high conductivity downstream (right). Two large-scale pumping tests confirm the contrast in mean conductivity of about 2 orders of magnitude (Boggs et al., 1992). Consequently, flow should be modelled with distinct mean conductivity in two vertical zones (Fig. 1b) when aiming to model mean arrival times for the MADE site.

2.1.2 Model aim "risk assessment"

1. *Aim.* The early or late arrival of contaminants commonly used in risk assessments is predicted.

2. *Processes.* Flow is described by Darcy's law, and transport is quantified through the ADE (Eq. 1). It is particularly relevant to capture variability in transport velocity to estimate the spreading behaviour of plumes.

3. *Field characterization.* High- and low-conductivity subsurface structures with a characteristic horizontal length scale of several metres need to be detected. Typical examples are channels formed in braided river systems. Typical investigation methods giving field evidence of such heterogeneity structures are small-scale slug tests, borehole flowmeter logs, or permeameter tests detecting strongly vertically varying conductivity.

4. *Conceptualization of hydraulic conductivity.* The nonuniform spatial structure of conductivity needs to be considered.

5. *Solving flow and transport.* Small variations in conductivity allow analytical solutions to be applied with effective measures, e.g. from first-order theory (Dagan,

1989). Spatially resolved heterogeneity requires numerical solutions of flow and transport using numerical tools (Monte Carlo approach).

Example MADE

Borehole flowmeter logs at MADE (Rehfeldt et al., 1989; Boggs et al., 1990) reveal horizontal layers with conductivity differences over 2–3 orders of magnitude. For instance, the flowmeter log F-40 shown in Fig. 2a has a bulk of high conductivity values, with about 15 % of values being 2 orders of magnitude smaller. Logs at other locations (F-09 and F-18) show the inverse behaviour: a bulk of low conductivity values with high conductivity inclusions embedded.

Such strong vertical variation indicates the presence of high conductivity channels acting as preferential flow paths and low conductivity zones with stagnant flow, which both impact plume spreading behaviour strongly. Consequently, when aiming to model early and late plume arrival, these features need to be accounted for in a flow and transport model for the MADE site.

2.2 Scale-dependent conductivity modules

Given the scale dependency of hydraulic conductivity features and their distinct relevance for flow and transport predictions, we propose three components: Module (A), (B), and (C), which capture large-, intermediate-, and small-scale heterogeneity effects, respectively. Given a certain model aim, components are selected (or not) with regard to the available field data. We shortly discuss the modules and motivations of their use based on the data of the MADE site example for different aims.

2.2.1 Module A

The aquifer domain of interest is divided into deterministic zones of significantly different mean conductivity (i.e. more than 1 order of magnitude). The structure can comprise horizontal or vertical layering, simply in blocks or complex zone
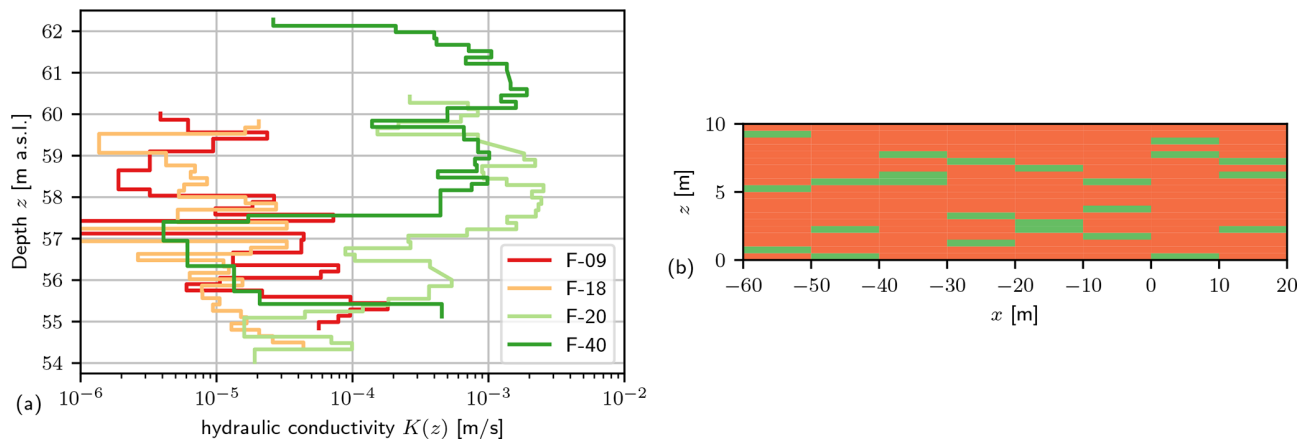
Figure 2. (a) Four flowmeter logs of hydraulic conductivity $K(z)$ versus depth z. The logs F-09 and F-18 are close to the tracer test injection location, and F-20 and F-40 are several tens of metres downstream (see Fig. 3). **(b)** Concept of binary inclusion structure (Module B) with 15 % high conductivity inclusions (green) embedded in the bulk of low conductivity (orange). Inclusion lengths are arbitrarily chosen to be $I_h = 5$ m and $I_v = 0.5 - 1$ m.

geometries, depending on the information available. The use of Module A is warranted when observation data indicate significant areal conductivity contrasts.

The zones represent large-scale geological structures exhibiting conductivity differences potentially over several orders of magnitude as a result of changes in deposition history or changes in the material's composition (Bear, 1972; Gelhar, 1993). Zones can be delineated using geologic maps, piezometric surface maps, and geophysical methods, providing information on aquifer structure, sedimentology, and genesis. Pumping tests are suitable for identifying mean conductivities for each zone due to their large detection scale. Flow simulations on the deterministic zone structure should reproduce the observed head pattern.

The MADE site is an example for which the concept of two zones of different mean hydraulic conductivity (Fig. 1b) can reproduce the hydraulic head pattern conceptually. Details will be discussed in Sect. 3.

2.2.2 Module B

When hydraulic conductivity shows heterogeneous features at the same length scale as the plume transport itself, they require proper resolution. A contaminant plume typically passes several of these intermediate-scale features but not enough to ensure ergodic transport behaviour. Thus, using effective parameters is not warranted. Since limited data availability precludes a deterministic representation of these features, stochastic approaches are best suited.

Binary stochastic models are a simple way to capture the effects of intermediate-scale features (Haldorsen and Lake, 1984; Dagan, 1986; Rubin, 1995). Figure 2b shows an example of how to conceptualize a medium with two K values: inclusions (K_2) are embedded in the bulk conductivity (K_1), with p characterizing the volume fraction of K_2. Inclusions

of high conductivity may represent preferential flow paths, whereas inclusions of low conductivity can be obstacles like clay lenses.

The inclusion topology is a matter of choice and data availability. A simple design is a distribution of non-overlapping blocks with horizontal length I_h and thickness I_v. Figure 2b provides an impression with an arbitrary choice of parameters. More complex layering structures can be adapted if additional topological information is available. However, the specific topology often plays a subordinate role. When not having any information on spatial correlation of heterogeneity, it is beneficial to assume some instead of sticking to a homogeneous model.

Characteristic length scales in a vertical direction I_v are detectable with low effort from a few borehole logs (Fig. 2a). Characteristic horizontal length measurements, I_h, are critical since they require spatially distributed observations. A parametric uncertainty approach can keep the effort low. A range of reasonable I_h values is estimated and applied in the random inclusion model. A sensitivity analysis reveals the impact of the parametric uncertainty of I_h on transport results. The estimates of I_h could result from auxiliary data such as vertical length scale in combination with anisotropy ratios. Another option is expert knowledge based on geological structures and similarities to outcrop studies. Methods such as diffusivity tests (Somogyvári et al., 2016) or novel approaches for pumping test interpretation (Zech et al., 2016) also offer options to gain estimates for I_h.

The binary structure as in Fig. 2b is beneficial in its plain stochastic concept, relying on few input data, simple implementation, and low computational requirements. It can be combined with Module (A) by implementing it within every deterministic zone preserving the mean conductivities. As for MADE, the inclusions represent the contrasting vertical layers as observed in flowmeter logs (Fig. 2a), from which the
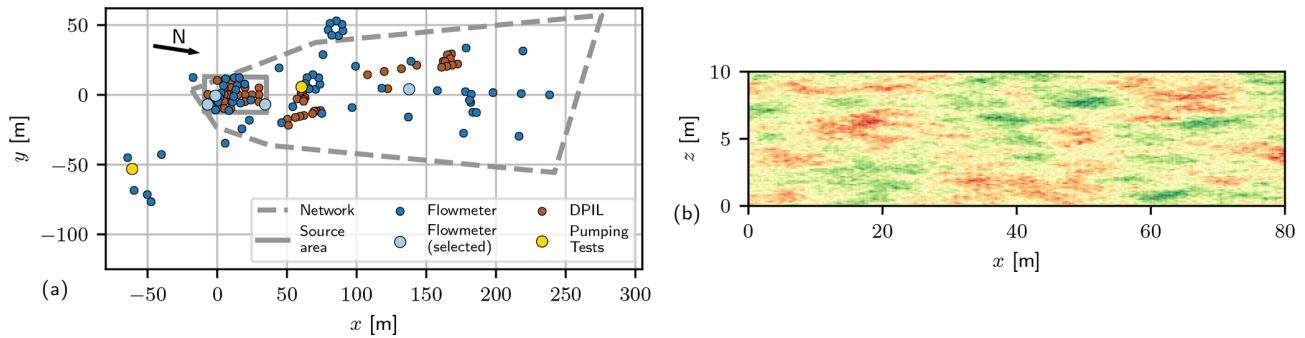
Figure 3. (a) Locations of measurements and tracer test observation network according to Boggs et al. (1990) and Bohling et al. (2016). **(b)** Gaussian random field with exponential co-variance structure as a conceptual module for small-scale conductivity (Module C).

inclusion parameters can be deduced for every deterministic zone (Sect. 3).

2.2.3 Module C

Variations in grain size and soil texture form small-scale heterogeneities of characteristic length scales up to 1 m. Their relevance for transport predictions depends on the degree of heterogeneity and ergodicity. A plume is considered ergodic when the behaviour within one realization is statistically representative, i.e. exchangeable with ensemble behaviour. Figuratively speaking, an ergodic plume has travelled long enough to sufficiently sample heterogeneity. This is usually assumed for transport distances of 10–100 characteristic lengths (Dagan, 1989), with higher values for an increasing degree of heterogeneity. When ergodic, effective parameters can capture effects of heterogeneity. Otherwise, the use of a spatial random representation is warranted.

If required, small-scale features can be conceptualized with a log-normal conductivity distribution $K(\boldsymbol{x}) \propto \mathcal{LN}(K_G, \sigma_Y^2)$ with a geometric mean K_G and log-variance σ_Y^2. Including a spatial correlation structure depends on the acquired complexity and the availability of two-point statistical data such as correlation length and anisotropy. Figure 3b gives an example.

Geostatistical parameters can be inferred from spatially distributed observations (Fig. 3a), e.g. permeameter, borehole flowmeter, or injection logging (Fig. 4). This is related to high effort and costs. Novel techniques like DPIL (Dietrich et al., 2008; Bohling et al., 2016) can provide a large amount of data at acceptable costs and time, but they are only accessible for shallow sites. Alternative approaches derive geostatistical parameters directly from pumping tests (Zech and Attinger, 2016; Zech et al., 2016) or dipole tracer tests (Zech et al., 2018). Note the discrepancy in geostatistical estimates among observation methods (Fig. 4), which is not uncommon for heterogeneous sites. Differences are attributed to scale effects as a result of different method characteristics, such as support volume and resolution. This underlines the caution

that has to be given to the appropriate use of observation data in conductivity conceptualization.

When combined with larger heterogeneity structures, small-scale fluctuations are subordinate. In the case of field evidence, Module (C) can be combined with Modules (A) and (B) by adding zero-mean fluctuations. According to Lu and Zhang (2002), the variances of heterogeneous substructures are additive. Thus, the log-normal variance relates to a "variance gap" between the total variance, e.g. from a geostatistical analysis of the entire domain, and the binary model's variance (Module B). It can be interpreted as the system's variance which is not captured by intermediate- and large-scale heterogeneity. The length scales for a correlation structure should be significantly smaller than the inclusion lengths of Module (B). Including small-scale heterogeneity enhances the realism of conductivity structure – however, at the expense of increasing investigation costs.

The MADE site is a rare example with geostatistics from multiple observation methods (Figs. 3a and 4). Methods well suited for small-scale heterogeneity show large variances from 4.5 up to 5.9. Given the high variance and the low mean conductivity, ergodic conditions cannot be assumed for transport within the range of a few hundred metres.

The large value in variance, as determined for MADE, could likely be the result of preferential flow and/or trends in mean conductivity. Thus, explicitly representing deterministic zones (Module A) and preferential flow paths (Module B) might render the representation of small-scale features (Module C) redundant. Modelling hydraulic conductivity as log-normal fields solely based on Module (C) seems warranted when there is no indication of deterministic zones or preferential pathways.

2.3 Hierarchy of scales

The hierarchy of scales poses an inherent problem for each groundwater model based on heterogeneous field data. Data interpretation often does not allow general trends to be clearly distinguished from randomness. The three modules provide a simple classification of transport-relevant hetero-
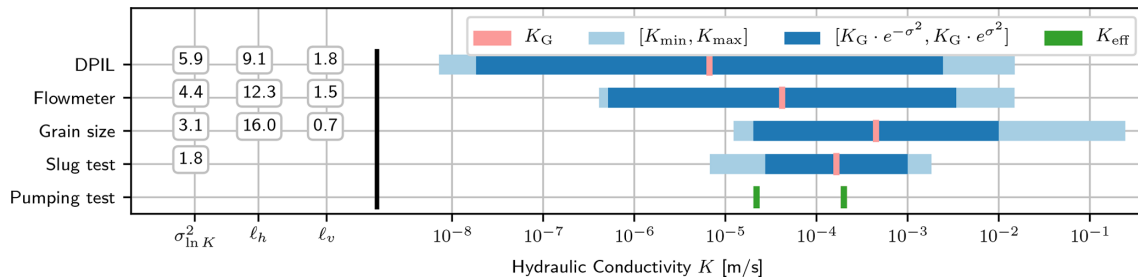
Figure 4. Geostatistical values for MADE from DPIL (direct push injection logging) (Bohling et al., 2016), flowmeters, grain size analysis, slug tests (Rehfeldt et al., 1992), and effective mean values (K_{eff}) of two large-scale pumping tests (Boggs et al., 1990): log-conductivity variance $\sigma_{\ln K}^2$ and horizontal and vertical correlation length ℓ_h and ℓ_v, respectively. Visualization of the range of observed values from minimal (K_{min}) to maximal (K_{max}), variance range, and geometric mean K_G.

geneity scales: (A) beyond the plume scale, i.e. above 100 m; (B) in the range of the plume scale (about 10–100 m); and (C) subscale (< 1 m). This classification might not hold for every field and transport situation but provides an orientation for developing site-specific heterogeneous conductivity structures.

Which module to integrate at a specific site depends on multiple aspects: (i) is there field data evidence for a heterogeneity structure of a certain length scale? (ii) Is there sufficient data to parameterize a conceptual heterogeneity representation? (iii) Is it necessary to present the heterogeneity given the travel distance of the plume (ergodicity)? Having a positive answer to each of these questions for a certain module warrants its consideration in the conductivity conceptual model.

3 Predictive transport model for MADE

We validate our approach by performing flow and transport calculations for the MADE setting without parameter calibration. Although many approaches to model the transport at the MADE site exist, including detailed aquifer conceptualizations (e.g. Herweijer, 1997; Julian et al., 2001, for a detailed review see Zheng et al., 2011), only few of them have a predictive character, i.e. devoid of calibration to transport results (Fiori et al., 2013, 2017; Dogan et al., 2014; Bianchi and Zheng, 2016).

Based on the scale-dependent conductivity modules (Sect. 2.2), we develop different conductivity structures according to the field evidence given the structural data at MADE. We thereby aim to identify the "most simple" of our concepts, which still provides a reasonable prediction of the complex observed mass distribution. The computed tracer plumes are compared to the MADE-1 transport experimental results (Boggs et al., 1992; Adams and Gelhar, 1992). Since the observed spatial concentration distribution is not available, we make use of 1D longitudinal mass transects at specified times.

Following the approach steps outlined in Sect. 2, we define our model aim broader then specified in Sect. 2.1: the goal is predicting the general plume behaviour. This might serve different purposes as, for example, remediation and includes the mean flow behaviour. The fact that there are no breakthrough curve data available for MADE inhibits the subject of arrival times being studied. Particularly critical is first arrival, as discussed in Adams and Gelhar (1992). Processes involved here are flow and transport governed by Darcy's law and the advection–dispersion equation (Eq. 1).

3.1 MADE field data

The MADE site is located on the Columbus Air Force Base in Mississippi, United States. The aquifer was characterized as shallow, unconfined, of about 10–11 m thickness (Boggs et al., 1992). It consists of alluvial terrace deposits composed of poorly sorted to well-sorted sandy gravel and gravely sand with significant amounts of silt and clay. The first extensive field campaign by Boggs et al. (1990) yielded a multitude of hydrogeological information, as, for example, piezometric surface maps and hydraulic conductivity observations from soil samples, flowmeters, and pumping tests (Fig. 4). Field campaigns in subsequent years supplemented observations and data interpretations. For an overview see, for example, Zheng et al. (2011), Bohling et al. (2016), or Table 1 in the Supplement. We apply a porosity of 0.31. Recharge is assumed uniform and very small (Boggs et al., 1990). Both quantities are kept constant due to the dominating effect of hydraulic conductivity given the significant variations and the uncertainty associated with observations (Fig. 4).

The MADE-1 transport experiment was conducted in the years 1986–1988 (Boggs et al., 1990, 1992; Rehfeldt et al., 1992; Adams and Gelhar, 1992). A pulse of bromide was injected over a period of 48.5 h, applying a flow rate of 3.5 l/min. The forced input conditions enlarged the tracer body at the source. Transport then took place under ambient flow conditions.

Concentrations were observed within a spatially dense monitoring network at several times after injection. We fo-
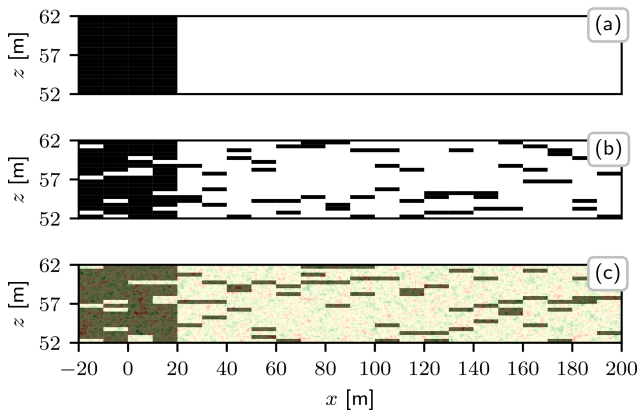
Figure 5. Realizations of hydraulic conductivity structures. **(a)** Deterministic zones (Module A), low K_1 in black, high K_2 in white. **(b)** Inclusions in deterministic zones (Modules A + B); volume fraction of inclusions $p = 15\%$, inclusion lengths $I_h = 10$ m, $I_v = 0.5$ m. **(c)** Inclusions in deterministic zones and subscale heterogeneity (Modules A + B + C); correlation lengths $\lambda_h = 2.5$ m, $\lambda_v = 0.125$ m.

cus on the reported longitudinal mass distribution of Adams and Gelhar (1992; Fig. 7) at six times: 49, 126, 202, 279, 370, and 503 d after injection. Values are integrated measures over transverse planes and accumulated over segments of 10 m length, given at the centres of segments at -5, 5, 15, …. The reported mass does not display mass recovery except at 126 d with recovery rates of 2.06, 0.99, 0.68, 0.62, 0.54, and 0.43, for the six times, respectively. We do not normalize the reported mass to recovered mass but stick to the actually observed values associating the mass loss with insufficient sampling in the downstream zone, as discussed in detail by Fiori (2014).

3.2 Hydraulic conductivity structures

Three hydraulic conductivity conceptualizations are designed in line with the specifications for MADE in Sect. 2, which serve different model aims. Modules (A), (B), and (C) are combined successively to capture the scale hierarchy of heterogeneity at the MADE site. Figure 5 illustrates examples for each conceptualization.

3.2.1 Deterministic zones (A)

Following the lines of Rehfeldt et al. (1992), we create conductivity zones based on the changes in the piezometric surface map (Fig. 1). We chose two vertically arranged deterministic zones (Fig. 5): a zone of low average conductivity, Z_1, from upstream of the tracer input location to $x = 20$ m downstream, and zone Z_2, an area of high average conductivity from 20 m downstream of the source (Sect. 2.1).

We fix mean conductivity values in the zones as $\bar{K}_{Z1} = 2e - 6$ m/s and $\bar{K}_{Z2} = 2e - 4$ m/s with a contrast of 2 orders of

magnitude as stated by Boggs et al. (1992). The specific values are chosen according to the two large-scale pumping tests (Boggs et al., 1992) and the head level rise during injection, which is particularly important for early plume development. Details are given in the Supplement.

When fixing regional conductivities from pumping tests, the model scale coincides with the measurement scale. This way, our structures are independent of the upscaling of the method-specific (and location-specific) geometric means reported for MADE (Fig. 4). The deterministic conductivity conceptualization is suitable for properly modelling the regional groundwater in line with the model aim "mean arrival" as specified in Sect. 2.1.

3.2.2 Inclusion structure in zones (A + B)

Flowmeter logs from MADE show a significant discontinuous heterogeneity in the layering (Fig. 2). We represent these structures, making use of the binary inclusion structure described in Sect. 2.2.2. We assume little to no information on horizontal structures and connectivity to mimic typical field situations – thereby deliberately ignoring the large amount of data at MADE. We make use of solely four flowmeter logs (Fig. 2a).

The binary conductivity distribution is constructed for the entire domain comprising both deterministic zones. The upstream zone Z_1 consists of a bulk of low conductivity K_1 with a percentage p of high conductivity K_2 inclusions, the downstream zone Z_2 vice versa (Fig. 5).

We identify the specific values of K_1 and K_2 from the statistical relationship for binary structures (Rubin, 1995): $\ln \bar{K}_{Z1} = (1 - p) \cdot \ln K_1 + p \cdot \ln K_2$ and $\ln \bar{K}_{Z2} = p \cdot \ln K_1 + (1 - p) \cdot \ln K_2$ using the mean conductivities of the zones $\bar{K}_{Z1} = 2e - 6$ m/s and $\bar{K}_{Z2} = 2e - 4$. p is deduced from the flowmeter profiles (Fig. 2a). Being from both zones Z_1 and Z_2, the profiles differ significantly in their average value. However, all show a tendency to binary behaviour, with a significant spread over depth. The data are grouped into high and low values, being at least 2 orders of magnitude apart. Then, p is the fraction of values in the minor group, which is 10 %–20 % for the MADE flowmeter profiles (Fig. 2a), leading to $p = 15\%$ as a default value.

The inclusions' structure in both zones is designed according to the simplified block structure outlined in Sect. 2.2.2. The domain is divided into horizontal blocks of length I_h. Each block contains randomly located inclusions of thickness I_v. The flowmeter logs at MADE indicate a change in vertical layering every 0.25–1 m (Fig. 2a). Thus, we chose $I_v = 0.5$ m. In combination with a volume fraction of inclusions of $p = 15\%$ and an aquifer thickness of 10 m, this gives three inclusions per block.

The parameter I_h is the most difficult to extract from data, due to the limited amount of information on horizontal structures and connectivity. We specify I_h through a pragmatic but stochastic meaningful approach by combining estimates with
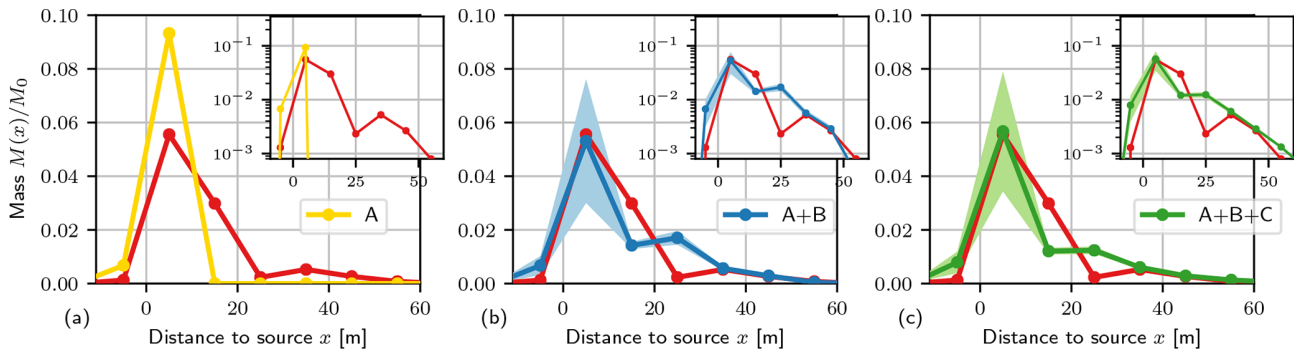
Figure 6. Longitudinal mass distribution at $T = 126$ days for conductivity concepts: (A) deterministic zones **(a)**, (A + B) inclusions in zones **(b)**, and (A + B + C) inclusion in zones with subscale heterogeneity **(c)** (Fig. 5). Shaded areas (light blue and green) indicate parametric uncertainty bands. Mass distribution observed at MADE experiment in red; the semi-log scale is in the subplot.

parametric uncertainty to rely on as few data as possible. A first guess results from auxiliary data analysis: an anisotropy ratio of $e = 0.1 - 0.025$ is given from the large-scale pumping tests (Boggs et al., 1990). Combining it with the inclusion thickness of $I_v = 0.5$ m gives a range of $I_h \in [5\,\mathrm{m}, 20\,\mathrm{m}]$. To cover parametric uncertainty we use three different values of I_h, namely 5, 10, and 20 m, instead of only one. The different inclusion lengths produce distinct effects on connected pathways and thus on the mass distribution. A combined ensemble integrates the character of each inclusion length. Figure 5b shows an example structure for $I_h = 10$ m. Note that inclusions can touch, so some inclusions are thicker (e.g. $2I_v = 1$ m) and longer (e.g. $2I_h = 20$ m).

For the Monte Carlo approach, we create ensembles of 600 individual random realizations, with 200 realizations of each inclusion length I_h, while all other parameters are fixed. Preliminary investigations showed that 200 realizations are sufficient to ensure ensembles convergence. Reported flow and transport results for the inclusion structure in zones (A + B) are ensemble means.

3.2.3 Subscale heterogeneity in zones (A + B + C)

We combine Modules (A), (B), and (C) to an inclusion structure in deterministic zones with small-scale fluctuations (A+B+C), depicted in Fig. 5c. Structural aspects of Modules (A) and (B) are the same as described before, including parametric uncertainty for the inclusion length $I_h \in \{5, 10, 20\}$ m. Module C is integrated as log-normal-distributed conductivity fluctuations (Sect. 2.2.3). The characterizing parameters for Module (C) depend on the statistics of the super-ordinate modules (A) and (B).

The log-normal fluctuations $\ln Y(x)$ are generated using *gstools* (Müller and Schüler, 2019) with zero mean, since the overall mean conductivity refers to $\bar{K}_{Z1}$ and $\bar{K}_{Z2}$ of the deterministic zones. The log-conductivity variance σ_Y^2 follows from the variance gap, as a difference between the variance of the inclusion structure and the overall variance. The binary inclusions for the chosen setting have a variance of $\sigma_Z^2 =$

5.52 resulting from $\sigma_Z^2 = p \cdot (1 - p) \cdot (\ln K_1 - \ln K_2)^2$ (Rubin, 1995). With an overall variance of $\sigma_F^2 = 5.9$ as indicated by Bohling et al. (2016) (Fig. 4), we arrive at a fluctuation variance of $\sigma_Y^2 \approx 0.5$. We apply an exponential co-variance function with length-scale parameters being a fraction of the inclusion length scales: $\lambda_h = 1/4 I_h$ and $\lambda_v = 1/4 I_v$. Testing several ratios, we saw that its impact on transport behaviour is negligible. Ensembles consist of 600 realizations.

3.3 Numerical model settings

Flow and transport are calculated making use of the finite element solver OpenGeoSys (Kolditz et al., 2012) in the *ogs5py* Python framework (Müller et al., 2020). The simulation domain is a 2D cross section within $x \in [-20, 200]$ m and $z \in [52, 62]$ m, generously comprising the area of the MADE-1 tracer experiment (Boggs et al., 1992). We applied constant head boundary conditions at the left and right margin, with a mean head gradient of $J = 0.003$. Tracer is injected at a well located at $x = 0$, with a central screen of 0.6 m depth. Injection takes place over a period of 48.5 h with an injection rate of $Q_{in} = 1.166e - 5\,\mathrm{m}^3/\mathrm{s}$ according to the initial conditions reported by Boggs et al. (1992). We use a flux-related injection, representing natural conditions. For technical details, the reader is referred to the Supplement.

We checked the impact of dimensionality. A detailed discussion is provided in the Supplement. We found almost no differences between 2D and 3D simulation setups, where the binary structure (Module B) dominates. Extending the binary structure in the horizontal direction perpendicular to main flow does not provide additional degrees of freedom for the flow. Thus, extending the model hardly impacts the flow and thus transport pattern while significantly increasing computational effort. However, dimensionality effects hold for conductivity conceptualization, with prevailing log-normal distribution, i.e. dominated by Module C. The option of complexity reduction by using 2D instead of 3D models is warranted for this application by the fact that conductivity con-
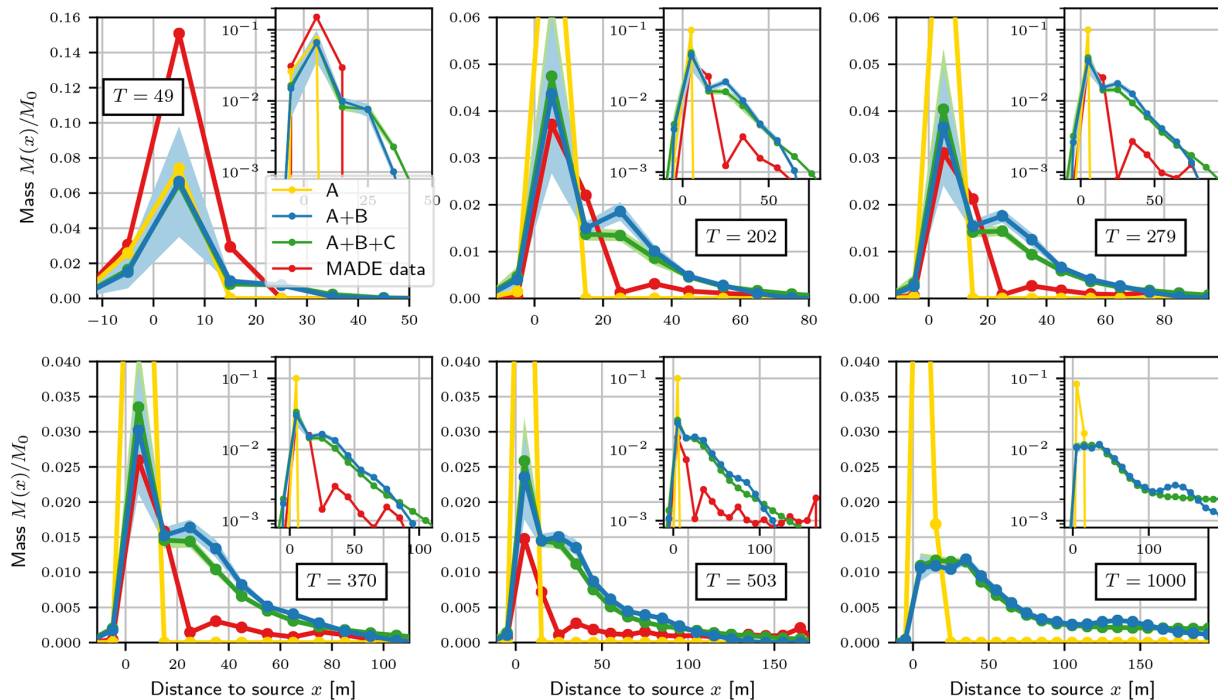
Figure 7. Mass distributions at times $T = 49, 202, 279, 370, 503$, and $1000\,\text{d}$. The colour code is as follows: red denotes the MADE-1 experiment; yellow denotes concept (A); blue denotes concept (A + B); green denotes concept (A + B + C). Shaded areas (light blue and green) indicate parametric uncertainty bands; the semi-log scale is in the subplot.

ceptualizations are dominated by the binary structure (module B).

Simulation results are processed like the MADE-1 experimental data. Longitudinal mass distributions are vertical averages and accumulated horizontally over 10 m segments. Note that the simulated distributions show a full mass recovery. Besides spatial mass distributions for the six times for which experimental data are available, we present the breakthrough curves (BTCs) as temporal mass evolution at critical distances, although no BTC data are reported for the MADE-1 experiment.

3.4 Simulation results

Figure 6 shows the simulated longitudinal mass distributions $M(x)/M_0$ of the specified conductivity conceptualizations (Sect. 3.2) at $T = 126$ days after injection. They are compared to the MADE-1 experiment data, which had a mass recovery of 99 % at that time.

The mass distribution for the deterministic structure (concept A, yellow) shows a sharp peak close to the injection location and no mass downstream. The conductivity structures with inclusions in deterministic zones (A + B, blue) and with subscale heterogeneity (A + B + C, green) result in skewed mass distributions, with a peak close to the injection area and a small amount of mass ahead of the bulk. Shaded areas indicate parametric uncertainty due to the variable inclusion length I_h. The shade area margins refer to ± 3 ensemble

standard deviations, which is similar to the 99 % confidence intervals, considering a Gaussian distribution of variations.

A direct comparison of the mass distributions $M(x)/M_0$ for the structures is depicted in Fig. 7 for six temporal snapshots, including $T = 1000\,\text{d}$, for which no experimental data are available. The general form of the mass distributions is persistent in time for all conductivity structures.

Figure 8 shows simulated breakthrough curves (BTCs) for the deterministic block and inclusion conductivity structure at three distances to the injection location. The results for concept (A + B + C) are very close to those of concept (A + B), thus are not displayed. Apparent differences to the longitudinal mass distributions as in Fig. 7 are due to the spatial data aggregation. The BTC for Module A has the expected Gaussian shape with a late breakthrough at $x = 5\,\text{m}$ given the very low conductivity in the injection area. The stochastic models have an earlier breakthrough and strong tailing at all distances.

BTCs are not available for the MADE-1 transport experiment. However, we added the aggregated mass values at the three locations for the six reported times in a subplot to indicate a trend of temporal mass development. Note that mass values of the BTCs and those at MADE are at different scales due to data aggregation and mass recovery.
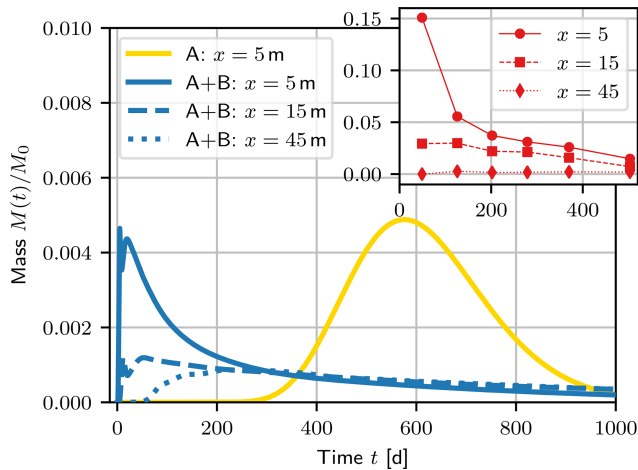
Figure 8. Breakthrough curves: total mass $M(t)/M_0$ versus time at selected control plane locations for inclusion structure (A + B), (blue) at $x = 5$ m (solid), $x = 15$ m (dashed), and $x = 45$ m (dotted) and for deterministic structure (A) at $x = 5$ m (solid yellow line). Reported mass values for MADE at the three locations (red markers) are given in the subplot. Note the difference in scale due to the spatial averaging of experimental data.

3.5 Discussion

All conductivity structures were able to reproduce the skewed hydraulic head distribution as observed at MADE (Fig. 1a). The corresponding mean flow velocity determines the travel time. As a result, all models properly reproduced the spatial position of the mass peak (Fig. 6).

The deterministic block structure (A) failed to reproduce the skewed mass distribution observed at MADE. The leading front mass travelling through fast flow channels could not be predicted (Fig. 7) solely using average K values in zones. In line with the model aim "mean arrival" (Sect. 2.1), the simple structure allows the regional groundwater movement to be estimated and the location of the bulk mass to be predicted. However, in the case of the "risk assessment" aim, the arrival times of mass would be significantly underestimated, as clearly observable when comparing BTCs (Fig. 8).

Tracer transport in a binary conductivity structure with inclusions (concept A + B) reproduces the observed mass, both for the peak near the injection site and the leading front. The simulated longitudinal mass distribution shows a second peak downstream (Fig. 7), which increases with time. The position is related to the interface between the low and high conductivity zones at 20 m distance to the source. Such a second peak is absent in the observed MADE plume; however it might be associated with the mass loss for the later times. The skewed mass distribution is related to significantly smaller first arrival times as can be seen for the BTCs in Fig. 8 compared to the deterministic structure. The BTCs are clearly non-Gaussian with heavy tailing. It shows the same temporal behaviour as the MADE experiment data.

The horizontal inclusion length I_h for structure (A + B) was not fixed but was varied over the range of $I_h \in \{5, 10, 20\}$ m. The uncertainty bands in Fig. 6b indicate that I_h mostly influences the height of the mass peak close to the source. I_h characterized the connectivity of the source area Z_1 to the high conductivity zone Z_2. Thus, it determines the distance of the bulk mass being trapped in the low conductivity area. The larger the I_h, the higher the amount of mass transported downstream. The shape of the leading front is less impacted by I_h given that its value does not influence the effect of the inclusions as preferential flow per se.

The predicted plume shape for the conductivity structure with inclusions and subscale heterogeneity (A + B + C) is almost similar to the one without subscale heterogeneity (A + B). Consequently, the inclusion structure is the one which determines the shape of the distribution, whereas the impact of subscale heterogeneity is minor. Given the model aim of plume prediction, the additional effort for determining characterizing geostatistical parameters for the subscale heterogeneity is not warranted.

The binary conductivity conceptualization (A + B) was derived for MADE with few observations from standard methods, as can be expected to be present at many field sites. The price for the limited amount of data is parametric uncertainty. A sensitivity study revealed that the mass distribution resulting from the binary conductivity structure is very robust against the choice of parameters. The inclusion length I_h and the choice of the K contrast between the zones show the highest impact. The latter was expected as the mean conductivity determines the average flow velocity and by that the peak location and the general distribution shape. The impact of I_h is represented in the uncertainty bands (Figs. 6b, 7). Other parameters, such as the volume fraction of inclusions p, and subscale heterogeneity parameters, such as the variance, have minor effects. For details, the reader is referred to the Supplement. In this regard, the binary structure is very stable towards parametric uncertainty.

4 Summary and conclusions

We introduce a modular concept of heterogeneous hydraulic conductivity for predictive modelling of field-scale subsurface flow and transport. The central idea is to combine deterministic structures with simple stochastic approaches to rely on few measurements and to forgo calibration. The scale hierarchy of hydraulic conductivity induces three structure modules which represent (A) deterministic large-scale features like facies, (B) intermediate-scale heterogeneity-like preferential pathways or low conductivity inclusions, and (C) small-scale random fluctuations. Field evidence of heterogeneity features and module's input parameters are provided by observation methods with the appropriate detection scale. The specific form of the scale-dependent features depends on the site characteristics and field data. We propose a determin-

istic model for large-scale features, a simple binary statistical model for intermediate-scale features, and a log-normal random model for small-scale features. However, the integration of alternative conductivity structures is possible. Thus, the concept is easily adaptable to any field site, making aquifer heterogeneity accessible for practical applications.

An illustrative example is given for the heterogeneous MADE site. Three modular conductivity structures are constructed, based on two observations: (i) the existence of distinct zones of mean flow velocity and (ii) high conductivity contrasts in depth profiles, suggesting local inclusions acting as fast flow channels. The structures are used in a predictive flow and transport model which is free of calibration. The comparison of results to the MADE-1 field tracer experiment showed that all conceptualizations can be of value, depending on the modelling aim. However, predicting the mass plume behaviour required heterogeneity to be taken into account.

The combination of deterministic and simple binary stochastic showed the best results given the trade-off between transport prediction and need for measurements. Realizations of hydraulic conductivity were composed of binary inclusions in two blocks with different average conductivity. Details on the topology are thereby secondary, since binary structures show robustness towards the choice of specific parameters.

The simple binary structure was able to capture the overall characteristics of the MADE tracer plume with reasonable accuracy, requiring only a small number of observations. Among the few predictive transport models for the MADE site, the approach presented shows a higher level of simulation effort due to the Monte Carlo simulations. However, the lower level of data requirements makes it attractive for application at less investigated sites. Note that when applying the proposed heterogeneity conceptualization in other modelling applications, a 3D model setup should be considered first, in particular when heterogeneity is conceptualized by a log-normal distribution (Module C). A complexity reduction to 2D models is warranted when the heterogeneous conductivity conceptualizations do not impact the flow pattern in the transverse horizontal direction, such as the binary structure. The generality of the binary concept makes it easily transferable to other sites, particularly when focusing on a few, but scale-related measurements.

A hierarchical conductivity structure allows for balance between complexity and available data. Large-scale structures determine the mean flow behaviour, which is most critical for flow predictions. They can be integrated into a model with reasonably low effort. Structural complexity increases with a decreasing heterogeneity scale, where small-scale features have the highest demand on observation data. However, even with limited information on the conductivity structure, simple stochastic modules can be used to incorporate the effect of heterogeneity. Considering small-scale features, the

conductivity structure can be extended by including modules when additional measurements are available.

Distinguishing the effects of the scale-specific features on flow and transport also allows the need for further field investigations and potential strategies to be identified. The adaptive construction based on scale-specific modules allows a conductivity structure model to be created as complex as necessary but as simple as possible.

The use of simple binary models is very powerful when dealing with strongly heterogeneous aquifers. They require fewer observation data compared to uni-modal heterogeneity models, as log-normal conductivity with high variances. Binary models also allow effects of dual-domain transport models to be incorporated without the drawback of having non-measurable input parameters which require model calibration. Our work shows that highly skewed solute plumes can be reproduced with classical ADE models by incorporating deterministic contrasts and effects of connectivity statistically. In summary, we conclude the following:

– Modular concepts of conductivity structure allow the multiple scales of heterogeneity to be separated. Scale-related investigation methods provide field evidence and conductivity model parameters. A hierarchical approach for conductivity can thus reduce observation effort by focusing on the model aim.

– Site-specific heterogeneous hydraulic conductivity can be easily constructed with simple methods, taking the (limited) amount of data into account. For aquifers with a high conductivity contrast, we recommend combining large-scale deterministic structures and simple binary stochastic models.

– The application example at MADE showed that complex field structures can be represented appropriately for transport predictions with an economic use of investigation data.

This work aims to contribute to bridging the gap between the advanced research in stochastic hydrogeology and its limited use by practitioners, being a subject of recent debate (e.g. Rajaram, 2016). We advocate the use of heterogeneity in transport models for successfully predicting solute behaviour, particularly in complex aquifers. This can be done with few data and simple tools: adaptive structures allowing deterministic, random binary, and geostatistical models to be combined, depending on the available data and the site-specific modelling aim.

in the random inclusion structure adapted to the MADE-1 settings. The Python API *ogs5py* (Müller, 2019) and the geostatistics packages *gstools* (Müller and Schüler, 2019) used in this study are both available at https://github.com/GeoStat-Framework. Data on the MADE aquifer can be accessed via the stated literature sources. Data generated for this study are available upon request to the corresponding author.

References

Adams, E. E. and Gelhar, L. W.: Field study of dipersion in a heterogeneous aquifer: 2. Spatial moments analysis, Water Resour. Res., 28, 3293–3307, https://doi.org/10.1029/92WR01757, 1992.

Bear, J.: Dynamics of Fluids in Porous Media, Elsevier, New York, 1972.

Bianchi, M. and Zheng, C.: A lithofacies approach for modeling non-Fickian solute transport in a heterogeneous alluvial aquifer, Water Resour. Res., 52, 552–565, https://doi.org/10.1002/2015WR018186, 2016.

Boggs, J. M., Young, S., Benton, D., and Chung, Y.: Hydrogeologic Characterization of the MADE Site, Tech. Rep. EN-6915, EPRI, Palo Alto, CA, 1990.

Boggs, J. M., Young, S. C., Beard, L. M., Gelhar, L. W., Rehfeldt, K. R., and Adams, E. E.: Field study of dispersion in a heterogeneous aquifer: 1. Overview and site description, Water Resour. Res., 28, 3281–3291, https://doi.org/10.1029/92WR01756, 1992.

Bohling, G. C., Liu, G., Dietrich, P., and Butler, J. J.: Reassessing the MADE direct-push hydraulic conductivity data using a re-

vised calibration procedure, Water Resour. Res., 52, 8970–8985, https://doi.org/10.1002/2016WR019008, 2016.

Bryant, I. D. and Flint, S. S.: Quantitative Clastic Reservoir Geological Modelling: Problems and Perspectives, John Wiley & Sons, Ltd, 1–20, https://doi.org/10.1002/9781444303957.ch1, 2009.

Carle, S. F. and Fogg, G. E.: Transition probability-based indicator geostatistics, Math. Geol., 28, 453–476, https://doi.org/10.1007/BF02083656, 1996.

Cirpka, O. A. and Valocchi, A. J.: Debates – Stochastic subsurface hydrology from theory to practice: Does stochastic subsurface hydrology help solving practical problems of contaminant hydrogeology?, Water Resour. Res., 52, 9218–9227, https://doi.org/10.1002/2016WR019087, 2016.

Dagan, G.: Statistical theory of groundwater flow and transport: Pore to laboratory, laboratory to formation, and formation to regional scale, Water Resour. Res., 22, 120S–134S, https://doi.org/10.1029/WR022i09Sp0120S, 1986.

Dagan, G.: Flow and Transport on Porous Formations, Springer, New York, 1989.

Damsleth, E., Tjolsen, C. B., Omre, H., and Haldorsen, H. H.: A Two-Stage Stochastic Model Applied to a North Sea Reservoir, J. Petrol. Technol., 44, 402–486, https://doi.org/10.2118/20605-PA, 1992.

Delhomme, J. P.: Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach, Water Resour. Res., 15, 269–280, https://doi.org/10.1029/WR015i002p00269, 1979.

Dietrich, P., Butler, J. J., and Faiss, K.: 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, 2008.

Dogan, M., Van Dam, R. L., Liu, G., Meerschaert, M. M., Butler, J. J., Bohling, G. C., Benson, D. A., and Hyndman, D. W.: Predicting flow and transport in highly heterogeneous alluvial aquifers, Geophys. Res. Lett., 41, 7560–7565, https://doi.org/10.1002/2014GL061800, 2014.

Fiori, A.: Channeling, channel density and mass recovery in aquifer transport, with application to the MADE experiment, Water Resour. Res., 50, 9148–9161, https://doi.org/10.1002/2014WR015950, 2014.

Fiori, A., Dagan, G., Jankovic, I., and Zarlenga, A.: The plume spreading in the MADE transport experiment: Could it be predicted by stochastic models?, Water Resour. Res., 49, 2497–2507, https://doi.org/10.1002/wrcr.20128, 2013.

Fiori, A., Cvetkovic, V., Dagan, G., Attinger, S., Bellin, A., Dietrich, P., Zech, A., and Teutsch, G.: Debates – Stochastic subsurface hydrology from theory to practice: The relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. What is characterization and stochastic theory good for?, Water Resour. Res., 52, 9228–9234, https://doi.org/10.1002/2015WR017525, 2016.

Fiori, A., Zarlenga, A., Jankovic, I., and Dagan, G.: Solute transport in aquifers: The comeback of the advection dispersion equation and the First Order Approximation, Adv. Water Resour., 110, 349–359, https://doi.org/10.1016/j.advwatres.2017.10.025, 2017.

Fogg, G. E. and Zhang, Y.: Debates – Stochastic subsurface hydrology from theory to practice: A geologic perspective, Water Resour. Res., 52, 9235–9245, https://doi.org/10.1002/2016WR019699, 2016.

https://doi.org/10.5194/hess-25-1-2021

Hydrol. Earth Syst. Sci., 25, 1–15, 2021

Fogg, G. E., Carle, S. F., and Green, C.: Connected-network paradigm for the alluvial aquifer system, Geol. S. Am. S., 348, 25–42, https://doi.org/10.1130/0-8137-2348-5.25, 2000.

Gelhar, L.: Stochastic Subsurface Hydrology, Prentice Hall, Englewood Cliffs, N. Y., 1993.

Gómez-Hernández, J., Butler, J. J., Fiori, A., Bolster, D., Cvetkovic, V., Dagan, G., and Hyndman, D.: Introduction to special section on Modeling highly heterogeneous aquifers: Lessons learned in the last 30 years from the MADE experiments and others, Water Resour. Res., 53, 2581–2584, https://doi.org/10.1002/2017WR020774, 2017.

Gómez-Hernández, J. J. and Gorelick, S. M.: Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance and recharge, Water Resour. Res., 25, 405–419, https://doi.org/10.1029/WR025i003p00405, 1989.

Haldorsen, H. and Lake, L.: A New Approach to Shale Management in Field-Scale Models, Soc. Petrol. Eng. J., 24, 447–457, https://doi.org/10.2118/10976-PA, 1984.

Herweijer, J. C.: Constraining uncertainty of groundwater flow and transport models using pumping tests, in: Calibration and Reliability in Groundwater Modelling, vol. 237, 473–482, IAHS Publ. no. 237, Golden, Colorado, 1996.

Herweijer, J. C.: Sedimentary Heterogeneity and Flow Towards a Well, Ph.D. thesis, Vrije Universiteit Amsterdam, Amsterdam, 1997.

Huysmans, M. and Dassargues, A.: Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium), Hydrogeol. J., 17, 1901–1911, https://doi.org/10.1007/s10040-009-0495-2, 2009.

Journel, A. G. and Gómez-Hernández, J. J.: Stochastic Imaging of the Wilmington Clastic Sequence, SPE Formation Evaluation, 8, 33–40, https://doi.org/10.2118/19857-PA, 1993.

Julian, H. E., Boggs, J. M., Zheng, C., and Feehley, C. E.: Numerical Simulation of a Natural Gradient Tracer Experiment for the Natural Attenuation Study: Flow and Physical Transport, Ground Water, 39, 534–545, https://doi.org/10.1111/j.1745-6584.2001.tb02342.x, 2001.

Kitanidis, P.: Introduction to Geostatistics: Applications in Hydrogeology, Cambridge University Press, Cambridge, New York, 2008.

Kolditz, O., Bauer, S., Bilke, L., Böttcher, N., Delfs, J.-O., Fischer, T., Görke, U. J., Kalbacher, T., Kosakowski, G., McDermott, C. I., Park, C. H., Radu, F., Rink, K., Shao, H., Shao, H. B., Sun, F., Sun, Y. Y., Singh, A. K., Taron, J., Walther, M., Wang, W., Watanabe, N., Wu, Y., Xie, M., Xu, W., and Zehner, B.: OpenGeoSys: An open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media, Environ. Earth Sci., 67, 589–599, https://doi.org/10.1007/s12665-012-1546-x, 2012.

Koltermann, C. E. and Gorelick, S. M.: Heterogeneity in Sedimentary Deposits: A Review of Structure-Imitating, Process-Imitating, and Descriptive Approaches, Water Resour. Res., 32, 2617–2658, https://doi.org/10.1029/96WR00025, 1996.

Linde, N., Renard, P., Mukerji, T., and Caers, J.: Geological realism in hydrogeological and geophysical inverse modeling: A review, Adv. Water Resour., 86, 86–101, https://doi.org/10.1016/j.advwatres.2015.09.019, 2015.

Lu, Z. and Zhang, D.: On stochastic modeling of flow in multimodal heterogeneous formations, Water Resour. Res., 38, 1190, https://doi.org/10.1029/2001WR001026, 2002.

Müller, S.: ogs5py v1.0.5, https://doi.org/10.5281/zenodo.3546035, 2019.

Müller, S. and Schüler, L.: GSTools: Reverberating Red (Version v1.1.1), Zenodo, https://doi.org/10.5281/zenodo.3532946, 2019.

Müller, S., Zech, A., and Heße, F.: ogs5py: A Python-API for the OpenGeoSys 5 scientific modeling package, Ground Water, https://doi.org/10.1111/gwat.13017, 2020.

Neton, M. J., Dorsch, J., Olson, C. D., and Young, S. C.: Architecture and directional scales of heterogeneity in alluvial-fan aquifers, J. Sediment. Res., 64, 245–257, https://doi.org/10.1306/D4267FA0-2B26-11D7-8648000102C1865D, 1994.

Proce, C. J., Ritzi, R. W., Dominic, D. F., and Dai, Z.: Modeling Multiscale Heterogeneity and Aquifer Interconnectivity, Groundwater, 42, 658–670, https://doi.org/10.1111/j.1745-6584.2004.tb02720.x, 2004.

Rajaram, H.: Debates – Stochastic subsurface hydrology from theory to practice: Introduction, Water Resour. Res., 52, 9215–9217, https://doi.org/10.1002/2016WR020066, 2016.

Rehfeldt, K. R., Hufschmied, P., Gelhar, L. W., and Schaefer, M.: Measuring hydraulic conductivity with the borehole flowmeter, Tech. Rep. EN-6511, EPRI, Palo Alto, CA, 1989.

Rehfeldt, K. R., Boggs, J. M., and Gelhar, L. W.: Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity, Water Resour. Res., 28, 3309–3324, https://doi.org/10.1029/92WR01758, 1992.

Renard, P., Straubhaar, J., Caers, J., and Mariethoz, G.: Conditioning Facies Simulations with Connectivity Data, Math Geosci., 43, 879–903, https://doi.org/10.1007/s11004-011-9363-4, 2011.

Rubin, Y.: Flow and Transport in Bimodal Heterogeneous Formations, Water Resour. Res., 31, 2461–2468, https://doi.org/10.1029/95WR01953, 1995.

Rubin, Y.: Applied Stochastic Hydrogeology, Oxford Univ. Press, New York, 2003.

Salamon, P., Fernàndez-Garcia, D., and Gòmez-Hernández, J. J.: Modeling tracer transport at the MADE site: The importance of heterogeneity, Water Resour. Res., 43, W08404, https://doi.org/10.1029/2006WR005522, 2007.

Sanchez-Vila, X. and Fernàndez-Garcia, D.: Debates – Stochastic subsurface hydrology from theory to practice: Why stochastic modeling has not yet permeated into practitioners?, Water Resour. Res., 52, 9246–9258, https://doi.org/10.1002/2016WR019302, 2016.

Smith, R., Bard, W., Corredor, J., Herweijer, J., McGuire, S., Antunez, A., Block, T., and Lazarde, N.: Geostatistical Modeling and Simulation of a Compartmentalized Deltaic Sequence, Ceuta Tomoporo Field, Lake Maracaibo, Venezuela, in: Proceedings on SPE Latin American and Caribbean Petroleum Engineering Conference, Society of Petroleum Engineers, https://doi.org/10.2118/69572-MS, 2001.

Somogyvári, M., Bayer, P., and Brauchler, R.: Travel-time-based thermal tracer tomography, Hydrol. Earth Syst. Sci., 20, 1885–1901, https://doi.org/10.5194/hess-20-1885-2016, 2016.

Strebelle, S.: Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics, Math. Geol., 34, 1–21, https://doi.org/10.1023/A:1014009426274, 2002.

Weissmann, G. S. and Fogg, G. E.: Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework, J. Hydrol., 226, 48–65, https://doi.org/10.1016/S0022-1694(99)00160-2, 1999.

Werth, C. J., Cirpka, O. A., and Grathwohl, P.: Enhanced mixing and reaction through flow focusing in heterogeneous porous media, Water Resour. Res., 42, W12414, https://doi.org/10.1029/2005WR004511, 2006.

Zech, A. and Attinger, S.: Technical note: Analytical drawdown solution for steady-state pumping tests in two-dimensional isotropic heterogeneous aquifers, Hydrol. Earth Syst. Sci., 20, 1655–1667, https://doi.org/10.5194/hess-20-1655-2016, 2016.

Zech, A. and Müller, S.: GeoStat-Examples/Binary_Inclusions, https://doi.org/10.5281/zenodo.4134627, 2020.

Zech, A., Müller, S., Mai, J., Heße, F., and Attinger, S.: Extending Theis' Solution: Using Transient Pumping Tests to Estimate Parameters of Aquifer Heterogeneity, Water Resour. Res., 52, 6156–6170, https://doi.org/10.1002/2015WR018509, 2016.

Zech, A., D'Angelo, C., Attinger, S., and Fiori, A.: Revisitation of the dipole tracer test for heterogeneous porous formations, Adv. Water Resour., 115, 198–206, https://doi.org/10.1016/j.advwatres.2018.03.006, 2018.

Zheng, C., Bianchi, M., and Gorelick, S. M.: 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, 2011.

Zinn, B. and Harvey, C. F.: When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields, Water Resour. Res., 39, 1051, https://doi.org/10.1029/2001WR001146, 2003.