Rebuttal

Dear editor, referees and discussant,

thank you for your constructive comments and interest in our work. We revised the manuscript accordingly. In the following, we provide point-by-point replies (in blue) to the comments (in italic). While line numbers in the comments refer to the previous manuscript version, mentioned lines in the response relate to the revised manuscript. Attached to the response is a marked-up manuscript version with tracked changes.

With kind regards,

Alraune Zech
on behalf of the author-team.

Editor comment:

I would suggest the authors should try to improve a bit the overall scientific significance of their study by highlighting some novel aspects and advances of the present study with respect to the existing literature on the subject.

We followed the advice of the editor along the comments of the discussant and the referees. We updated the abstract as well as passages of the introduction (l 64 ff) and conclusion section (422ff) accordingly.

Response to Referee #1: (Received and published: 9 March 2020)

The manuscript presents a hierarchical approach for modeling flow and transport in heterogeneous aquifer. The approach is applied to the now classic MADE macrodispersion experiment, and it is focused on the modeling of longitudinal mass distribution, as observed during the course of the experiment. The paper is very well written and the method is clearly illustrated. The topic is relevant, and I do believe that approaches like the one envisioned here are very important to reduce the complexity of natural ground-water systems. I think that the work deserves publication. A few minor comments follow.

We want to thank the referee for his positive evaluation of our work. We appreciate the referee's time and effort he put into reviewing our manuscript. The paper will benefit from revising it according to his comments. Along the lines of the author comments published before, we here address all raised in combination with text modification in the manuscript.

- *Abstract: I find unusual to start new paragraphs within an abstract.*
  We decided for a structuring of the abstracts into paragraphs to improve readability.
- *Eq.1: why the ADE is presented in one spatial dimension? This may be misleading, also considering that hydraulic conductivity K(x) is variable in x only under such conditions (line 32)*
  Eq. 1 and related quantities are adapted to 3D.

- *Figure 1. The position of the boundary between orange and green is not clear: where does it derive?*
  The position of the boundary between blocks: Studying Figure 1 (left) shows that 20m down stream of the source (black dot) the head pattern changes abruptly. Along the orange line in the left figure there are 4 head isolines whereas along the green line of 40m length only one. This is a strong indication for a change of mean hydraulic conductivity. Thus, we chose this position as location for the interface of distinct material blocks. Figure 1 (right) indicates the vertical cross section where the choice of the coordinate system is along the one outline in the left figure.

  We added a comment in the caption of Figure 1. We expanded the explanation of Figure 1 in the text (section 2.1.1, L.130)

- *Line 155: is there any evidence of such large differences in hydraulic conductivity among zones (several orders of magnitude, line 157)?*
  This section focus on the general setup of Module A. It is designed to represent deterministic areas of high conductivity contrast in the conceptual model for sites with field evidence. We added an explanatory sentence to the paragraph (L. 163).
  For the MADE site, there is evidence of such large differences in hydraulic conductivity (Figure 1 and the explanation on that). Thus, the use of Module A in a conductivity conceptual setup is warranted.

- *Line 164: "reproduce" instead of "reprocude".* Corrected.

- *Line 175: why the choice of Ih and Iv?*
  We corrected the misleading formulation (L. 182ff): Figure 2 provides a visual example. The choice of I_h and I_v is thus arbitrary here. The specific parameters are transferred to the figure caption.

- *Line 181: please elaborate more on the "expert knowledge" for assessing the range of reasonable I_h values estimated.*
  We revised the paragraphs on the inclusion topology elaborating on how to determine reasonable estimates for horizontal inclusion length scales in the context of expert knowledge (L.184ff).

- *Line 189: please define ergodicity, and briefly explain (possibly with references) why it is assumed when the plume has travelled 10-100 characteristics lengths.*
  Intuitively speaking, the ergodic hypothesis for a system implies that all states of the ensemble are available in each realization [Dagan, 1989]. A figurative description in the context of transport is, that the plume sampled sufficient heterogeneity over its travel distance to be representative for the average behavior of the heterogeneous material structure. The value of 10-100 characteristics lengths follows from stochastic arguments of the sample size [Dagan, 1988,1989].

  We revised the paragraph adding a definition of ergodicity and we gave references to the relation to travel distance. (L. 198ff)

- *Line 248: I don't see a clear transition at x=20 from Fig.1*
  The point was emphasized in Figure 1 and section 2.1.1 (see also comment above).

- *Line 250: I don't recall the contrast described here in Boggs et al (1992), please elaborate more.*
  A statement on the conductivity contrast is repeatedly mentioned in the Boggs et al, 1992 paper, starting in the abstract: "This asymmetry was produced by accelerating groundwater flow along the plume travel path that, in turn, resulted from an approximate 2-order-of-magnitude increase in the mean hydraulic conductivity between the near-field and far-field regions of the site."
  In the manuscript's paragraph we state all relevant information and refer to the supporting information for further explanations. We do not know what additional information to provide here.
- *Line 257: "designed" instead of "design".* Corrected.
- *Line 272: please explain the heuristic approach with some more detail. Line 275: why 600 realizations?*
  The parameter I_h is the most difficult to extract from data, generally due to the very limited amount of information on horizontal structures and connectivity. Thus, a pragmatic, but also stochastic meaningful approach is necessary. We decided to combine estimates from the data (the range of $I_h \in [5m, 20m]$ deduced from vertical inclusion length and the anisotropy rate), with the approach of parametric uncertainty: instead of using only one value out of the range, we allow for 3 different: 5m, 10m and 20m. The different inclusion length produce distinct effects on connected pathways and thus on the mass distribution. In the combined ensemble the character of each inclusion length is thus integrated. The ensemble thus consists of 3* 200 realization of each inclusion length.
  We used 600 realizations to assure that the number is sufficiently large to ensure ensemble convergence. As stated in the manuscript, we found in preliminary convergence tests, that 200 realizations are sufficient to reproduce ensembles averages. Given the combination of different inclusion length (previous comment), we combined 3*200 realizations for the general ensemble representing model structure A+B.

  We reworked and expanded the paragraphs on the inclusion structure and the number of realizations accordingly. (L. 298ff) In this line, we modified the formulation heuristic approach which is misleading.
- *Line 294: so the model is 2d? why not working with the more realistic 3d setup? Do you expect differences in the results? I guess that the additional degree of freedom brought by 3d could make a difference.*
  Dimensionality of the model: We provided a detailed discussion in the (previous) author comment to the referee. A paragraph on that is added to the manuscript and we provide a detailed discussion along these lines in the Supporting information.
- *Line 296: how is the solute injected? Does the local injection rate depend on local hydraulic conductivity?*
  Solute injection follows the experimental description in Boggs et al., 1992. It is a flux related injection being the realistic representation of natural conditions. Thus the local distribution of tracer depends o the local heterogeneity.

  We added the explanation ("It is a flux related injection being the realistic representation of natural conditions.") to the manuscript.

- *Figure 6. Please introduce a legend.*
  We modified the text (A, A+b, A+B+C) to a legend and introduced a legend in Figure 7 as well.
- *Conclusions: the first item of the list (line 396) is a rather well known and general statement, I would not add it as one of the conclusive statements of the work.*
  We removed the sentence from the conclusive statements.

Response to Referee #2:

(Received and published: 23 March 2020)

The manuscript presents a hierarchical approach for modeling flow and transport in heterogeneous aquifer. The key idea is combining large-scale deterministic structures and simple stochastic approaches. While the inclusion of a hierarchical structure to deal with heterogeneous structure is not new (some modelers have used similar ideas, yet not as structured as in this case), the authors introduce a formalism to make it understandable and efficient, I think is the main value of the manuscript. A significant point in the manuscript is the n-th try to model the data from the MADE side (here n isa very, very large number).

We'd like to thank the referee for taking his time to review our manuscript and appreciate his positive evaluation and constructive criticism. We revised the manuscript accordingly. Here we outline the changes following the discussion in the final author comments.

The aspect of novelty and that this model is yet another try to model the MADE side tracer test was also raised the a discussant and the editor. We addressed these issues by revising the abstract as well as introduction and conclusion section.

So, maybe the main comment I have is the issue of dimensionality. First, the very simple thing is that eq. (1) should be 3D, as this is a general idea, and no need to simplify the problem at this point (you can do that later).
Eq. 1 was adapted to 3D.

But, most importantly, your application is 1D.
We want to specify that the conceptual and numerical models are 2D. We post-process the calculated mass distributions to allow a comparison with the 1D reference data of the MADE 1 experiment: As outline in the manuscript, averaged over the directions perpendicular to the flow and aggregated over intervals of 10m. (L. 341).

I have seen many models trying to fit the 1D data of MADE; but after all these years, I have not yet seen the spatial distribution of values. - We neither.
Everybody reports the correspondence with transects (your figures 6 and 7). Transects are OK, but do not reflect the real picture at all. From l.235, "Concentrations were observed within a spatially dense monitoring network at several times after injection". Is this data available? Why nobody uses it in their models?
Unfortunately, no other mass data than the 1D transects is available to us. In correspondence with many colleagues, we figured that the raw data is not public. We regret this situation, but are not in the position to change it. So, I can just agree to the referee.
We added a comment on the data situation to section 3 (L. 245)

You start with Figure 1. Why such a simple concept, if we know that it is slightly more complicated.
That is actually one of the paper's targets: make use of the "simple concepts" and data which is often available such as piezometric surface maps to construct a reasonable heterogeneous hydraulic conductivity structure. In the application of the concept to MADE we wanted to show that by integrating and combining "basic" data, it is possible to reproduce apparently complex mass distribution patterns at least at the level of spatially integrated longitudinal mass distributions.
We added a comment on that to section 3 (L. 243).

But, in general I like the work, and I feel it is very well written. I loved in particular the section "Exemplary Model Aims". This is written in a very didactic way.

This is a tough one and I do not expect an answer. The model developed in Section 3.2 involves quite a number of decisions and parameters. Then you get a reasonable fit. Now, can you really calibrate this model with so many parameters at very different scales (variances, integral distances, p values, anisotropy ratios, directions of anisotropy,...)? I can see that being done manually for one-two parameters (e.g.,your line 276), but more? You would need a supercomputer and plenty of staff or students working on it, but this would be a waste. So, is there any automatic calibration approach that you envision in the future?
In section 3.2, we outlined how to derive the required parameters from hydraulic observations. We emphasize that the model is set up as a predictive model. There is no calibration involved.
We saw the need to emphasized this point (predictive and calibration-free model) in the revised manuscript: as e.g. in L. 83, 238, 404.

The choice of several values of inclusion length is not a calibration but an integration of parametric uncertainty. We did not calibrate the model to one of the values, but included random realization with all of these values to the ensemble.
The paragraph on the derivation of the inclusions structure and choice of horizontal inclusion length I_h, was reworked and expanded accordingly (L. 302ff).

Minor issues:

The problem inherent to hierarchy of scales is how do you assign variability to one scale or the other one. I mean, you can always claim that some general trends are nothing but randomness if we look at a larger scale. Some discussion about how to distinguish Modules (A), (B) and (C) in a general case could benefit the paper. I mean,should (B) always related to the transport features as suggested in l.166?
We fully agree with the referee that categorizing spatial variability observed at a specific site to scales is subject of discussion and uncertainty. Addressing the point "I mean, should (B) always related to the transport features as suggested in l.166?" (L 166: heterogeneous features at the same length scale as the plume transport itself") - Not per se. Generally, we relate the Modules to the typical length scale of material feature, also related to specific observation methods. However, Module (B) represents heterogeneity of a few meters length (up to some tenth of meters), which coincides with the typical length scale of a contaminant plumes. In this sense, Module (B) is prone to be representing the relevant heterogeneity.

We added a section (L. 227ff) discussing the hierarchy of scales and how to distinguish modules.

You could comment a bit on recharge, because probably recharge and transmissivity (in a 2D scenario) might be correlated. How does your hierarchy approach deal with this parameter?

Similarly, you could also comment a bit on the impact of porosity, if you think it is relevant (maybe it is not); it appears in the transport equation.
We add a notice on recharge and porosity in L. 259.
As a site note: In our model, we do not work with transmissivity since there are variations in conductivity and flow velocities in the vertical direction.

L79. In my opinion the models of Fiori (2013, 2017) are completely non-predictive (actually, they are based on wrong assumptions, as you show in your paper); outperforming those methods should not even be cited.
We refrain from removing the reference to Fiori et al. (2013, 2017). They offer an alternative model for the MADE site, which we consider as valuable contribution to the scientific discussion.

L90. Again, the use of word "macro-dispersion"; maybe you refer to "enhanced dispersion". The concept of "macro" refers to a specific quantity (since the original derivation of Gelhar and Axness, to all those by Dagan and so) that are never, ever, attained in real field conditions.
In the context here, we agree that macrodispersion is not the proper choice of words. We corrected accordingly (L. 96).

L116. Is this reference really needed here? I mean, the relevance of pumping testscomes from the 1930's if not earlier. And we teach them in class…- We removed the reference.

L 254. "Arrival" is misspelled. - Corrected.

L 312. This is equivalent only if a Gaussian distribution of concentrations is invoked. You could add this warning. - We added the warning (L. 353).

Response to Discussant:

(Received and published: 27 February 2020)

This article presents an approach re 'A hierarchical aquifer model which combines large-scale deterministic structures and simple stochastic approaches' in order to 'Predict Transport in a Heterogeneous Aquifers'.
As it is a research paper we may expect this to be a novel approach. If the paper is extending similar earlier work then references should be made. If it is routine application of existing methods the paper should be called a case history.
The paper presented here shows neither a novel nor an original approach. The approach presented was also earlier applied to the same site where the MADE project was conducted (Columbus Air Force Base, MS, USA).
The type of hierarchical deterministic/stochastic modelling of geological features and permeability distribution discussed in the paper, has been extensively used in the oil and gas industry since the mid 1980s. There is a vast body of literature on the methodology and applications. All this is completely ignored, ie. not referenced, in this paper. Plenty basic references (up to 1996) can be found in chapter 2 of ref 1 below.
This type of model is also not new for the Columbus Air Force Base area where the MADE experiment was conducted. I have personally published a PhD thesis and an article on a hierarchical deterministic/stochastic approach applied to tracer tests at Columbus Air Force Base (the site where the MADE experiment was conducted). The 4th listed author is well aware of all this, as he personally communicated with me, was reviewer of my PhD thesis (Ref 1 below), and

attended conferences where papers were presented (eg. ref 2).
Given this, the authors should thoroughly re-study existing literature and reference some key papers out of the oil and gas industry. They also should make very clear that this is not a novel/original approach but simply a standard application of what has done before and is routine in oil and gas reservoir modelling. The authors should also make clear reference to similar work already conducted∼25 years ago at the same site (Columbus Air Force Base test site where the MADE experiment was conducted), eg. ref 2.
The only reason why the material could be published, is that it finally may point out the scientific confusion and structural research mis-management around the MADE experiment and stochastic hydrology. The MADE experiment has led to numerous publications in journals, which all ignored to account for geological heterogeneity in an appropriate manner and ignored other work that would not fit the premises of stochastic hydrology (macro dispersion theory).

Ref 1 – Herweijer, J.C.,1997.Sedimentary heterogeneity and flow towards a well. Ph.D. dissertation, Free University, Amsterdam (https://www.hydrology.nl/images/docs/dutch/1997.01.07_Herweijer.pdf)

Ref 2 – Herweijer, J.C, 1996.Use of sedimentology and geostatistical modeling to estimate uncertainty of groundwater models. Proc. International Conferenceon Calibration and Reliability in Groundwater modeling (ModelCARE96), Golden (CO,USA), September, 1996 (https://pdfs.semanticscholar.org/a5a5/25d8da8091bb59a59795262d932f0b4a6333.pdf)

Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2020-30/hess-2020-30-SC1-supplement.pdf

We want to thank Joost Herweijer for his interest in our work. We acknowledge the initiation of the scientific discussion on the subject matter of integrating aquifer heterogeneity to hydrogeological transport models as this was one of the goals of the work. As stated in the manuscripst’s last paragraph, we aim to contribute to bridging the gap between the advanced research in stochastic hydrogeology and its limited use by practitioners. In this line, we agree that hierarchical deterministic/stochastic modeling permeability is used in the oil and gas industry, but hardly found it’s way into applied hydrogeology. We acknowledge his effort in providing us publications on MADE which have not been available to us or of which we were not aware, respectively.

The points he raised are now addressed in the manuscript:

- it was specified that the presented work is an application of the hierarchical approach. Reformulated e.g. as “novel conceptualization strategy of aquifer heterogeneity in a hierarchical deterministic/stochastic framework”
- missing references are integrated: Herweijer, 1996, 1997, Bryant & Flint, 2009, Bianchi & Zheng, 2016
- We are aware that this study is not the first approach to model transport at the MADE site. However, it is one of the few predictive models (no calibration) for the MADE 1 experiment and conceptually very different from the other predictive approaches [e.g. Fiori et al. 2013, 2017, Bianchi & Zheng, 2016]. Where the Monte Carlo procedure is related to computational effort, the amount of required field data is limited making the approach attractive to less investigated sites.

- The mentioned work on “A hierarchical deterministic/stochastic approach applied to tracer tests at Columbus Air Force Base” was integrated and provides a valuable reference. The application of hierarchical approaches in the context of pumping test interpretation is perfectly in line with the suggestions in our study to make use of distinct interpretation methods for aquifer heterogeneity characterization. It is a great example to make use of pumping tests to determine connected areas of high conductivity. The application to the MADE experiment however relates to another tracer test setup, with forced flow between wells, again focusing on fast flow channels.

(Received and published: 9 March 2020)

2D vs 3D has been looked extensively, see e.g.

- *Static characterizations of reservoirs: refining the concepts of connectivity and continuity Joseph M. Hovadik and David K. Larue Petroleum Geoscience, Vol. 13 2007, pp.195–211 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.818.7201&rep=rep1&type=pdf*
- *King, P. R., 1990, The connectivity and conductivity of overlapping sand bodies. In, Buller Anthony T. et al.,eds, North Sea oil and gas reservoirs; II, Proceedings of theNorth Sea oil and gas reservoirs conference. [Book, Conference Document], Pages 353-362. https://link.springer.com/chapter/10.1007/978-94-009-0791-1_30*
- *Models in study @ MADE site mentioned in my previous comments were all run in 3D, as is was concluded that 2D models tend to ’suppress’ connectivity.*

The discussion on the model dimensionality (2D vs. 3D) is given along the response to referee #2.

A Field Evidence Model: How to Predict Transport in a 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]Utrecht University, Department of Earth Science, Utrecht, The Netherlands
[3]Eberhard Karls University Tübingen, Germany
[4]University of 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 presumably missing simple concepts have limited the incorporation of heterogeneity to many field transport models up to now.

We present a ~~hierarchical aquifer model which~~ novel conceptualization strategy for aquifer heterogeneity. The hierarchical approach combines large-scale deterministic structures and simple stochastic ~~approaches~~methods. Such a heterogeneous conductivity can easily be integrated into ~~a~~ numerical models. Depending on the modelling aim, the required structural complexity can be adapted. The same holds for the amount of available field data. The conductivity model is constructed step-wise following field evidence from observations; though relying on as minimal data as possible. Starting point are deterministic blocks, derived from head profiles and pumping tests. Then, sub-scale heterogeneity in form of random binary inclusions are introduced to each block. Structural parameters can be determined e.g. from flowmeter measurements.

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 ADE models are able to describe highly skewed tracer plumes by incorporating deterministic contrasts and effects of connectivity in a stochastic way even without using uni-modal heterogeneity 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 the 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), an increased urbanization with the resulting solid and liquid wastes 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 (Bear, 1972; Fetter, 2000).

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

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

given ~~here in one spatial dimension x~~ in space $\boldsymbol{x}=(x,y,z)$ and time t. ~~D is the macro-dispersion coefficient and $u(x,t)$~~ $\mathbf{D}$ 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 ~~$K(x)$~~ $K(\boldsymbol{x})$ through Darcy's Law. A proper description of the velocity field ~~$u(x,t)$~~ $\boldsymbol{u}(\boldsymbol{x},t)$, thus aquifer heterogeneity, is crucial for predicting the concentration distribution ~~$C(x,t)$~~$C(\boldsymbol{x},t)$.

The adequate parametrization of the heterogeneous conductivity ~~$K(x)$~~ $K(\boldsymbol{x})$ poses a significant challenge in practical model development due to the lack of data. 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 (Dagan, 1989; Gelhar, 1993; Koltermann and Gorelick, 1996).

On one hand, fully deterministic approaches use either uniform (effective) conductivities in large domains or maps of heterogeneity, created by interpolation, e.g. Kriging (Kitanidis, 2008). The former approach requires only few data to the price of neglecting local effects of heterogeneity. The latter requires a huge amount of observation data which is hardly ever available in practical cases. Furthermore, conductivity fields from interpolation result in smooth structures lacking geological realism. On the other hand, stochastic methods allow to resolve heterogeneity based on a limited amount of data. Thus, they are able to capture the uncertainty in flow and transport predictions caused by heterogeneity. Common methods as (i) Gaussian random fields (Freeze, 1975; Dagan, 1989; Gelhar, 1993; Zinn and Harvey, 2003); (ii) indicator/hydrofacies models (Carle and Fogg, 1996; Fogg et al., 2000); or (iii) multi-point statistics/training images (Renard et al., 2011; Linde et al., 2015) allow to create spatially distributed conductivity fields of higher geological realism. Modelling flow and transport in ensembles of heterogeneous fields (Monte Carlo approach) do not only provide mean behavior but also uncertainty ranges.

Log-normal random fields require a number of parameters like geometric mean, log-variance and spatial correlation lengths in horizontal and vertical directions. They result from geostatistical analysis of spatially distributed observations, e.g. from flowmeter, permeameter or injection logging, as DPIL (Dietrich et al., 2008). Despite increased efficiency in exploration methods, the cost and effort related to gather sufficient data hampers the application in practice. Alternatively, hydrofacies models use indicator geostatistics with transition probability to generate geological heterogeneity structures. Although conceptually different, the general amount of input data is similarly high. Training images are known for their geological realism, but depend strongly on the high resolution input data, e.g. reconstructed images from outcrop studies. Not only the availability of a training image limits their application, but particularly the question if it is representative for the larger aquifer domain where transport is modeled (Koltermann and Gorelick, 1996).

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 modeling. We see a significant reason in the lack of data for complex stochastic models. Thus, we ~~propose a novel approach which focuses on optimizing the available field data adaptive to~~ advocate the use of hierarchical approaches, combining deterministic and stochastic hydraulic conductivity conceptualization. In contrast to many application in the oil and gas industry (Bryant and Flint, 2009), they hardly found their way into applied hydrogeology (Herweijer, 1997).

Here, we present a novel conceptualization strategy of aquifer heterogeneity in a hierarchical deterministic/stochastic framework. Goal is to optimize the aquifer structure setup given the simulation target constrained by the available field data. Thereby, we aim to provide a tool making aquifer heterogeneity more accessible for practical applications.

Our 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, respectively the spatial spreading of a contaminant or solute.

We create a deliberate connection between the model parameterization requirements and the field characterization methods employed for measurement beyond a single method. Pumping tests, for example, are a recommended characterization method to determine the spatially averaged transmissivity respectively hydraulic conductivity, even in a heterogeneous aquifer environment ~~(Zech et al., 2016)~~(Herweijer, 1996; Zech et al., 2016). Together with the averaged gradient estimated from piezometric levels this yields good estimates of the mean groundwater flow velocities. On the other hand, high resolution, small-scale borehole logs of hydraulic conductivity (e.g. from flowmeter or DPIL) can provide the data needed to estimate the variability of the hydraulic conductivity field and consequently the dispersion parameters needed.

We demonstrate the methodology using field characterization data from MADE, a heterogeneous, well investigated research field site (e.g. Boggs et al. (1990); Zheng et al. (2011); Gomez-Hernandez et al. (2017)). Following our adaptive approach, we use a minimum of field data on aquifer properties to construct a numerical transport model~~and to predict tracer plume behavior following a Monte Carlo approach~~. Predictions are independently evaluated using field tracer data from the MADE-1 experiment (Boggs et al., 1992). ~~They~~ In contrast to most other MADE transport models, we predict tracer plume behavior following a Monte Carlo approach devoid of calibration. Model results shows good agreement with data, also compared to other complex predictive transport models for MADE (i.e. ~~g. Fiori et al. (2013, 2017)~~Fiori et al. (2013, 2017); Bianchi and Zheng (2016)).

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

2 Approach

Large scale hydraulic structures of hundreds or more meters 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 meters, generate spatially variable flow fields. They also render transport velocities variable at these scales resulting in a larger spreading of plumes. This is particularly important for modeling tailing or leading mass fronts. Fluctuations on scales smaller than these intermediate scales have a blending effect, generally increasing local mixing and ~~macro-dispersion~~ 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 parametrize 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, as pumping tests, flowmeter and DPIL measurements (Boggs et al., 1990; Bohling et al., 2016). Detailed information on MADE can be found in section 3 and the *Supporting Information*.

In the approach, we considers 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 minimal effort?
4. Conceptualizing hydraulic conductivity.
5. Calculating flow and transport.

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

2.1 Exemplary Model Aims

Model Aim "Mean Arrival"

1. **Aim**: Prediction of mean arrival of a contaminant from a point source.

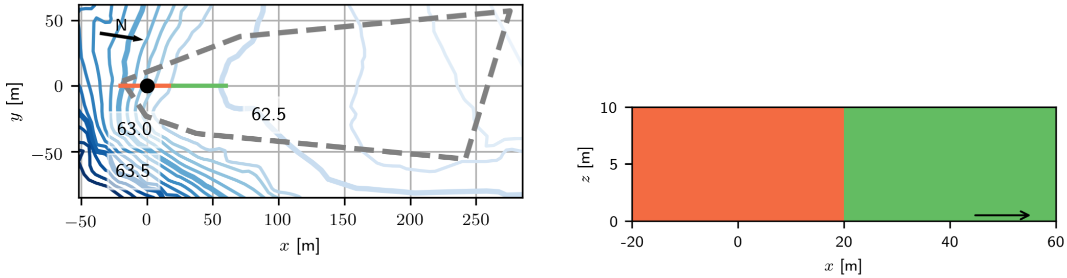

Figure 1. Left: Potentiometric surface map of head measurements according to Boggs et al. (1990). Orange-Green line indicates location of cross section displayed right: Concept (Module A) for large conductivity structure with deterministic zones of low (orange) and high (green) conductivity. Arrow indicates flow direction. Location of the interface between structures corresponds to change in hydraulic head pattern at $20m$.

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

3. **Field characterization:** Regionalized groundwater level measurements provide direction and magnitude of 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~~(Kruseman and de Ridder, 2000)~~.

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 behavior, 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 case of a more complex spatial distribution of zones. Transport can be determined making use of analytical or numerically solutions of the ADE according to initial and boundary conditions.

2.1.1 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 Figure 1a allow to delineate these two major zones: 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 two orders of magnitude (Boggs et al., 1992).

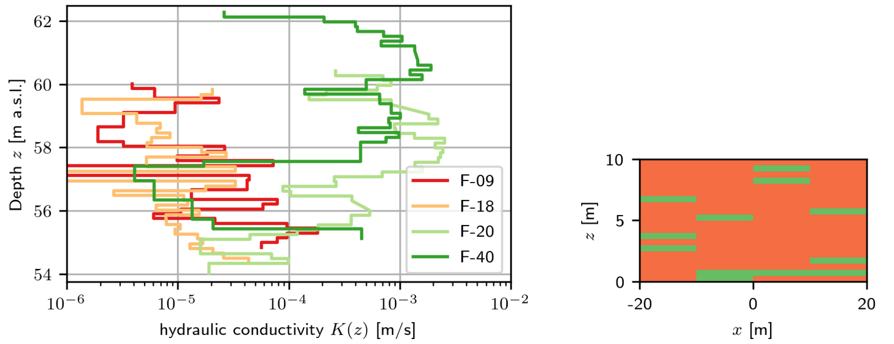

Figure 2. Left: Four flowmeter logs of hydraulic conductivity $K(z)$ versus depth z; the logs *F-09* and *F-18* are close to the tracer test injection location; *F-20* and *F-40* are several tenth of meters downstream (see Figure 3). Right: Concept of binary inclusion structure (Module B) with 15% high conductivity inclusions (green) embedded in the bulk of low conductivity (orange). Inclusion length are arbitrarily chosen as $I_h = 5\,\mathrm{m}$ and $I_v = 0.5 - 1\,\mathrm{m}$.

Consequently, flow should be modelled with distinct mean conductivity in two vertical zones (Figure 1b) when aiming to model mean arrival times for the MADE site.

Model Aim "Risk Assessment"

1. **Aim:** Prediction of early or late arrival of contaminants commonly used in risk assessments.

2. **Processes:** Flow and transport equations; it is particularly relevant to capture variability in transport velocity to estimate spreading behavior of plumes.

3. **Field characterization:** Detecting and delineating high and low conductivity subsurface structures with a characteristic horizontal length scale of several meters. 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:** Spatially structured non-uniform conductivity.

5. **Solving flow and transport:** Small variations in conductivity allow to apply analytical solutions with effective measures, e.g. from first order theory (Dagan, 1989). Spatially resolved heterogeneity requires numerical solution of flow and transport with numerical tools (Monte Carlo approach).

2.1.2 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 Figure 2a has a bulk of high conductivity values with about 15% of values being two orders of magnitude smaller. Logs at other locations (*F-09* and *F-18*) show the inverse behavior: a bulk of low conductivity values with embeddings of high conductivity.

Such strong vertical variation indicate the presence of high conductivity channels acting as preferential flow path and low conductivity zones with stagnant flow which both impact strongly on plume spreading behavior. Consequently, when aiming to model early and late plume arrival these feature 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 motivate their use based on the data of the MADE site example for different aims.

Module A

The aquifer domain of interest is divided into deterministic zones of significantly different mean conductivity (i.e. more than one order of magnitude). The structure can comprise horizontal or vertical layering simply in blocks or complex zone geometries depending on information available. The use of Module A is warranted when observation data indicates significant areal conductivity contrasts.

The zones represent large scale geological structures exhibiting conductivity differences potentially over several orders of magnitude as a results 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 where the concept of two zones of different mean hydraulic ~~conducitivity~~ conductivity (Figure 1b) can ~~reprocude~~ reproduce conceptually the hydraulic head pattern. Details will be discussed in section 3.

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 behavior. Thus, using effective parameters is not warranted. Since limited data availability precludes from a deterministic representation of these features, stochastic approaches suit best.

Binary stochastic models are the simplest way to capture the effects of intermediate scale features. Figure 2b shows an example how to conceptualize a medium with two K values: inclusions (K_2) are embedded in the bulk conductivity (K_1), with p characterizing the percentage 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~~as in~~. Figure 2b ~~with $p = 15\%$, $I_h = 5\,\mathrm{m}$ and $I_v = 0.5 - 1\,\mathrm{m}$~~provides an impression with 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 vertical direction I_v are detectable with low effort from a few borehole logs (Figure 2a). Characteristic horizontal length as 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 ~~(e.g. using expert knowledge)~~ 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 results 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 Somogyvari 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 Figure 2b is beneficial in its plain stochastic concept relying on few input data, simple implementation and low computational requirement. 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 (Figure 2a), from which the inclusion parameters can be deduced for every deterministic zone (section 3).

Module C

Variations in grain size and soil texture form small scale heterogeneities of characteristic length scales up to one meter. Their relevance for transport predictions depends on ~~ergodicity and thus, on~~ the degree of heterogeneity ~~. Ergodicity is usually assumed when the plume has traveled~~ 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 ~~. Then~~(Dagan, 1989), which increasing value for 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 $K(\boldsymbol{x}) \propto \mathcal{LN}(K_\mathrm{G}, \sigma_Y^2)$ with 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 as correlation length and anisotropy. Figure 3b gives an example.

Geostatistical parameters can be inferred from spatially distributed observations (Figure 3a), e.g. permeameters, borehole flowmeter, or injection logging (Figure 4). This is related to high effort and costs. Novel techniques like DPIL (Dietrich et al.,

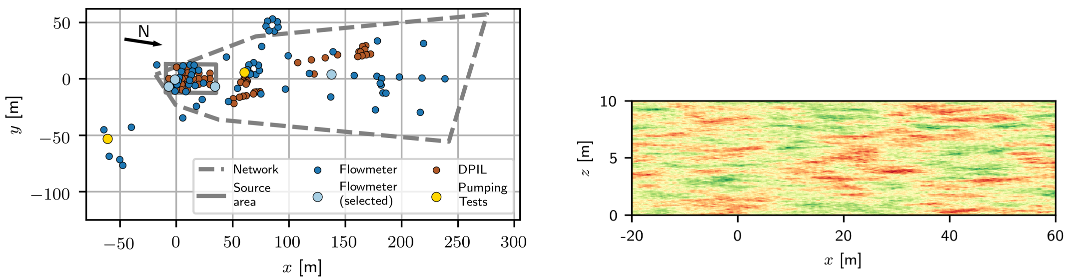

Figure 3. Left: Locations of measurements and tracer test observation network according to Boggs et al. (1990); Bohling et al. (2016). Right: Gaussian random field with exponential co-variance structure as conceptual module for small scale conductivity (Module C).

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. Alternatives can be approaches which derive geostatistical parameters directly from pumping tests (Zech and Attinger, 2016; Zech et al., 2016) or dipole tracer test (Zech et al., 2018).

When combining with larger heterogeneity structures, small scale fluctuations are subordinate. In 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 sub-structures is 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, on the expanse of increasing investigation costs.

The MADE site is a rare example with geostatistics from multiple observation methods (Figures 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 meters.

The large value in variance, as determined for MADE, can 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. Modeling hydraulic conductivity as log-normal fields solely based on Module (C) seems warranted when there is no indication for deterministic zones or preferential pathways.

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 to clearly distinguish general trends from randomness.

The three modules provide a simple classification of transport relevant heterogeneity scales: (A) – beyond plume scale, i.e. above 100m; (B) – range of plume scale (about 10-100m); and (C) – sub-scale (<1m). It will not be appropriate for every field and transport situation, but provides an orientation for developing site-specific heterogeneous conductivity structures.

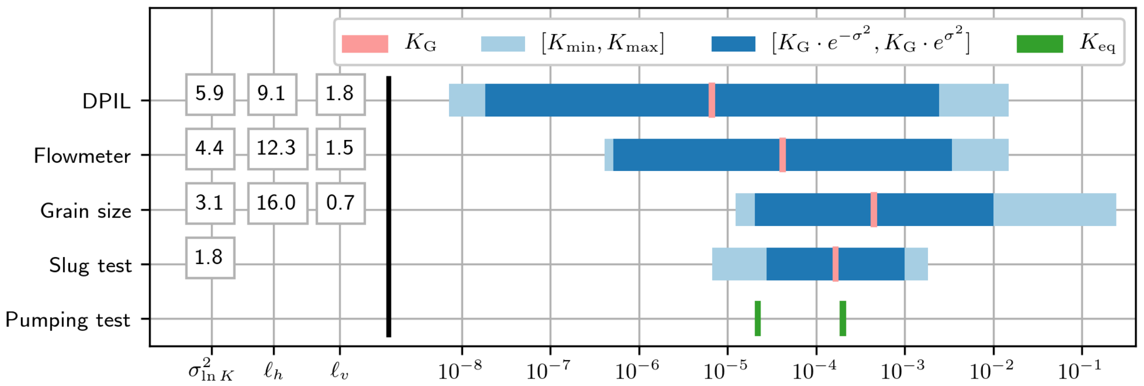

Figure 4. Geostatistical measures for MADE from DPIL (direct push injection logging) (Bohling et al., 2016), flowmeter, grain size analysis, slug tests (Rehfeldt et al., 1992) and effective mean values (K_{eff}) of two large scale pumping tests (Boggs et al., 1990): log-conductivity variance $\sigma^2_{\ln K}$, horizontal and vertical correlation length ℓ_h and ℓ_v, respectively. Visualization of range of observed values from minimal (K_{min}) to maximal (K_{max}), variance range and geometric mean K_{G}.

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? And (iii) is it necessary to present the heterogeneity given the travel distance of the plume (ergodicity)? Having a positive answer to each of the question 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 calculation for the MADE setting ~~.~~ without parameter calibration. Although, many approaches to model the transport at the MADE site exist (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 (section 2.2), we ~~derive~~ 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 ~~plume is~~ plumes are compared to the MADE-1 transport ~~experiment (Boggs et al., 1992; Adams and Gelhar, 1992).~~ 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 section 2, we define our model aim broader then specified in section 2.1: The target is predicting the general plume behavior. This might serve different purposes as e.g remediation and includes the mean flow behavior. The fact that there is no break-through curve data available for MADE, inhibits to study the subject of arrival times.

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, U.S.A. The aquifer was characterized as shallow, unconfined, of about $10-11$ m ~~thick with a porosity of 0.32~~ 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 hydro-geological information, as e.g. piezometric surface maps and hydraulic conductivity observations from soil samples, flowmeter and pumping tests (Figure 4). Field campaigns in subsequent years supplemented observations and data interpretations. For an overview see e.g. Zheng et al. (2011); Bohling et al. (2016) or Table 1 in the *Supporting Information*. 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 (Figure 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 $48.5h$ 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 focus on the reported longitudinal mass distribution of Adams and Gelhar (1992, Fig.7) at six times: 49, 126, 202, 279, 370, and 503 days after injection. Values are integrated measures over transverse planes and accumulated over slices of 10 m length, given at the centers of slices at -5 m, 5 m, 15 m, ... The reported mass does not display mass recovery except at 126 days 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 to insufficient sampling in the downstream zone as discussed in details by Fiori (2014).

3.2 Hydraulic Conductivity Structures

Three hydraulic conductivity conceptualizations are designed in line with the specifications for MADE in section 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 illustrated examples for each conceptualization.

3.2.1 Deterministic Zones (A)

As indicted by the piezometric surface map (Figure 1, section 2.1.1), we chose two vertically arranged deterministic zones (Figure 5): a low in average conductivity zone Z_1 from upstream of the tracer input location to $x = 20$ m downstream and zone Z_2 as high-in-the-average conductivity area from 20 m downstream of the source. At $x = 20$ m is an abrupt change in the head isoline pattern.

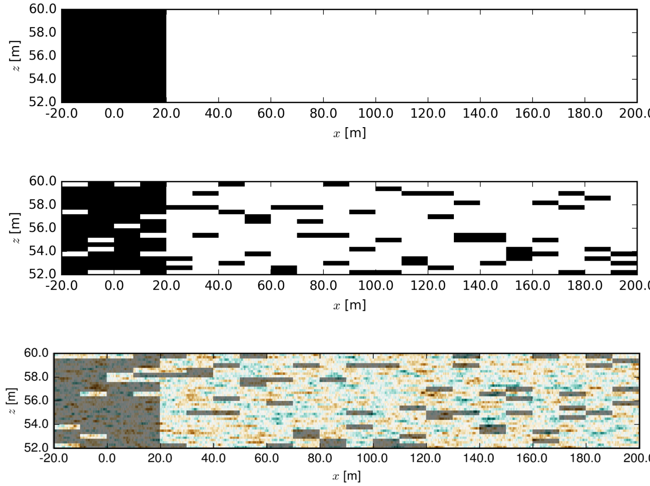

Figure 5. Realizations of hydraulic conductivity structures: (top) Deterministic zones (Module A), low K_1 in black, high K_2 in white. (center) Inclusions in deterministic zones (Modules A+B); amount of inclusions $p = 15\%$, inclusion lengths $I_h = 10\,\text{m}$, $I_v = 0.5\,\text{m}$. (bottom) Inclusions in deterministic zones and sub-scale heterogeneity (Modules A+B+C); correlation lengths $\lambda_h = 2.5\,\text{m}$, $\lambda_v = 0.125\,\text{m}$.

We fix average conductivity values of $\bar{K}_{Z1} = 2e-6$ m/s and $\bar{K}_{Z2} = 2e-4$ m/s with a contrast of two orders of magnitude as stated by Boggs et al. (1992). The specific values are chosen according to the two large scale pumping test (Boggs et al., 1992) and the head level rise during injection which is particularly important for early plume development. Details are given in the *Supporting Information*. This deterministic conductivity conceptualization is suitable for properly modelling the regional groundwater in line with the model aim "Mean ~~Arrivel~~Arrival" as specified in section 2.1.

3.2.2 Inclusion Structure in Zones (A+B)

Flowmeter logs from MADE show a significant discontinuous heterogeneity in the layering (Figure 2). We represent these structures making use of the simple binary inclusion structured described in section 2.2.

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 (Figure 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 (Figure 2a). Being from both zones Z_1 and Z_2, the profiles differ significantly in their average value. However, all show a tendencies of binary behavior with a significant spread over depth. The data is grouped into high and low values being at least two orders of magnitude apart. Then, p is the fraction of values in the minor group, which is $10-20\%$ for the MADE flowmeter profiles (Figure 2a) leading to $p = 15\%$ as default value.

The inclusions structure in both zones is ~~design~~ designed according to the simplified block structure outlined in paragraph 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 (Figure 2a). Thus, we chose $I_v = 0.5$ m. In combination with a inclusion percentage 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 ~~an heuristic approaches in combination~~ a pragmatic, but stochastic meaningful approach by combining estimates with parametric uncertainty to rely on as little data as possible: ~~The anisotropy ratio~~ 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))~~is $e = 0.1 - 0.025$~~. Combining it with the inclusion thickness of $I_v = 0.5$ m gives a range of $I_h \in [5\,m, 20\,m]$. To cover parametric uncertainty we use three different values of I_h, namely 5 m, 10 m and 20 m instead of only one. The different inclusion length produce distinct effects on connected pathways and thus on the mass distribution. In the combined ensemble the character of each inclusion length is thus integrated. Figure 5b shows an example structure for $I_h = 10$ m. Note that inclusion can touch, so some inclusions are thicker (e.g. $2I_v = 1$ m) and longer (e.g. $2I_h = 20$ m).

For ~~our~~ the Monte Carlo Approach, we create ensembles of 600 individual random realizations~~. Allowing for parametric uncertainty, we consider three possible I_h values of 5 m, 10 m and 20 m when generating random realizations. The~~ , with 200 realizations of each inclusion length I_h, while all other parameters are fixed~~to the values outlined above~~. 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. ~~We checked for ensemble convergence and found that 200 realizations are already sufficient.~~

3.2.3 Sub-scale 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 Figure 5, bottom. 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 (section 2.2). The characterizing parameters for Module (C) depend on the statistics of the super-ordinate modules (A) and (B).

The log-normal fluctuations $\ln Y(\boldsymbol{x})$ are generated 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 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) (Figure 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 behavior is negligible. Ensembles consist of 600 realizations.

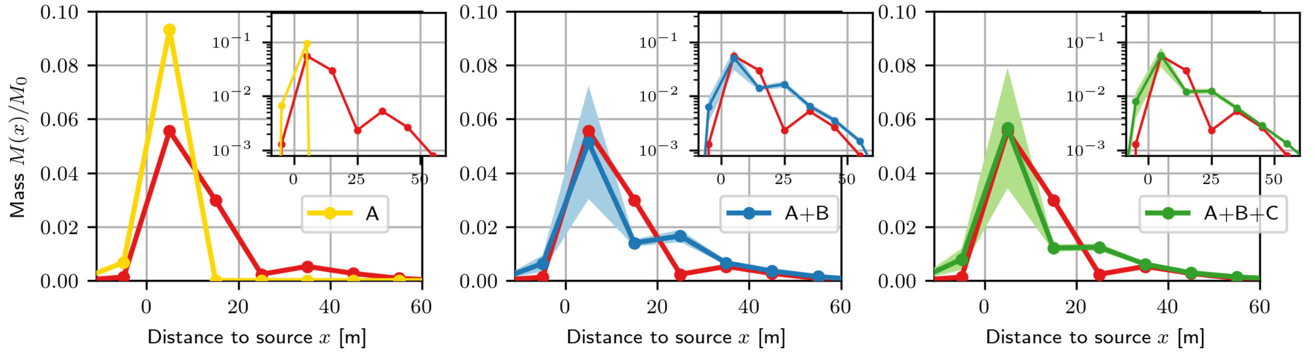

Figure 6. Longitudinal mass distribution at $T = 126$ days for conductivity concepts: (A) deterministic zones, (A+B) inclusions in zones, (A+B+C) inclusion in zones with sub-scale heterogeneity (Figure 5). Shaded areas (light blue and green) indicate parametric uncertainty bands. Mass distribution observed at MADE experiment in red. Linear scale and log-scale in subplot.

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 had 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 ~~$Q_{in} = 1.166e-5$~~ $Q_{\text{in}} = 1.166e-5$ m^3/s according to the initial conditions reported by Boggs et al. (1992). It is a flux related injection being the realistic representation of natural conditions. For technical details, the reader is referred to the *Supporting Information*.

We checked the impact of dimensionality (2D instead of 3D) and found almost no differences between 2D and 3D simulation setups. This is in contrast to known results for log-normal distributed fields, but can be explained by the conceptualization of the heterogeneous binary structure. 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. A detailed discussion is provided in the *Supporting Information*.

Simulation results are processed like the MADE-1 experimental data. Longitudinal mass distributions are vertical averages and accumulated horizontally over 10 m slices. Note that the simulated distributions show a full mass recovery. Besides spatial mass distributions for the six times where experimental data is available, we present the break through curves (BTCs) as temporal mass evolution at critical distances, although no BTCs data is 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 (section 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 sub-scale 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 are depicted in Figure 7 for six temporal snapshots, including $T = 1000$d, where no experimental data is 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 not displayed. Apparent differences to the longitudinal mass distributions as in Figure 7 are due to the spatial data aggregation. The BTC for Module A has the expected Gaussian shape with a late breakthrough at $x = 5$ 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.

3.5 Discussion

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

The deterministic block structure (A) failed to reproduce the skewed mass distribution observed at MADE. The leading front mass traveling through fast flow channels could not be predicted (Figure 7) solely using average K values in zones. In line with model aim "Mean Arrival" (section 2.1), the simple structure allows to estimate the regional groundwater movement and to predict the location of the bulk mass. However, in case of aiming at "Risk Assessment", the arrival times of mass would be significantly underestimated, as clearly be observable comparing BTCs (Figure 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 (Figure 7), which increases with time. The position is related to the interface between the low and high conductivity zones

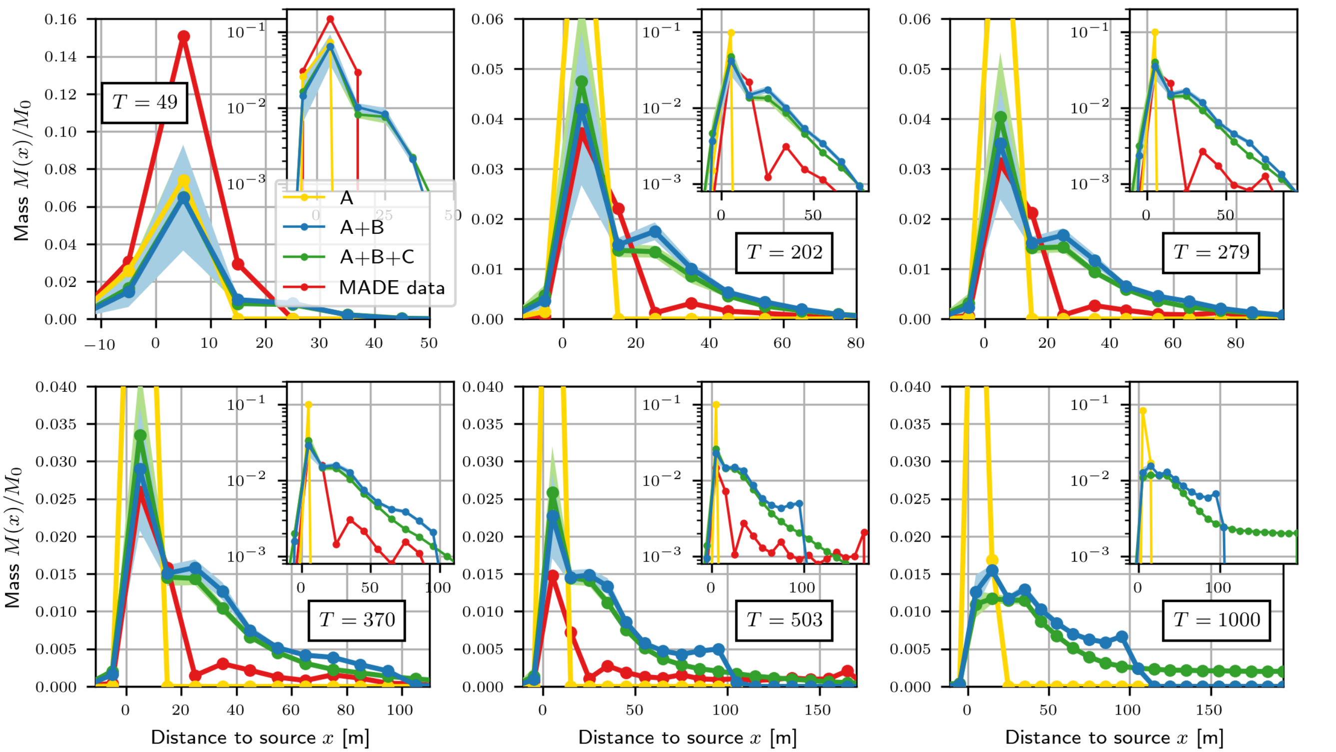

Figure 7. Mass distributions at times $T = 49$, 202, 279, 370, 503, and 1000 days (panels): red = MADE-1 experiment; yellow = concept (A); blue = concept (A+B); green = concept (A+B+C). Shaded areas (light blue and green) indicate parametric uncertainty bands; semi-log scale in subplot.

at $20\,\mathrm{m}$ 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 Figure 8 compared to the deterministic structure. The BTCs are clearly non-Gaussian with heavy tailing. It shows the same temporal 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 Figure 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 I_h the higher is the amount of mass transported downstream. The shape of the leading front is less impacted by I_h giving 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 sub-scale heterogneity (A+B). consequently, the inclusion structure is the one which determines the

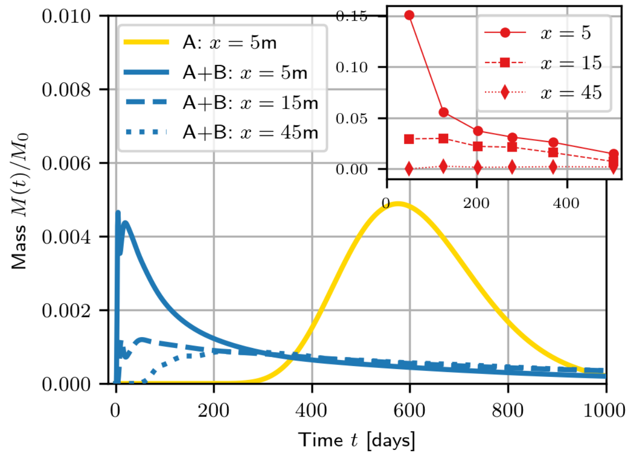

Figure 8. Breakthrough curves: Total mass $M(t)/M_0$ versus time at selected control plane locations for inclusion structure (A+B), (blue) at $x = 5$ m (solid), $x = 15$ m (dashed), $x = 45$ m (dotted); and for deterministic structure (A) at $x = 5$ m (solid yellow line). Reported mass values for MADE at the three locations (red markers) given in subplot. Regard the difference in scale due to the spatial averaging of experimental data.

shape of the distribution, whereas the impact of sub-scale heterogeneity is minor. Given the model aim of plume prediction, the additional effort for determining characterising geostatistical parameters for the sub-scale heterogeneity is not warranted.

The binary conductivity conceptualization (A+B) was derived for MADE with minimal data from field investigations, thus with a high 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 (Figures 6b, 7). Other parameters as amount of inclusion p and sub-scale heterogeneity parameters as the variance have minor effects. For details, the reader is referred to the *Supporting Information*. In this regard, the binary structure is very stable towards parametric uncertainty.

4 Summary and Conclusions

When aquifer heterogeneity is at a similar scale as solute transport, predictive transport models need to incorporate spatially distributed hydraulic conductivity. We introduce a modular concept of heterogeneous hydraulic conductivity for predictive modeling of field scale subsurface flow and transport. Central idea is to combine deterministic structures with simple stochastic approaches to rely on a minimal amount of 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; (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. Generally, we propose a deterministic model for large-scale features, a simple binary statistical model for intermediate and a geostatistical model for small-scale features. However, the integration of alternative conductivity structures is possible. Thereby, 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 ~~values depending on~~ value depending on the modelling aim. However, predicting the mass plume behaviour required to take heterogeneity into account.

The combination of deterministic and simple stochastic showed the best result given the trade-off between transport prediction and need for measurements. Realizations of hydraulic conductivity 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. This rather simple structure was able to capture the overall characteristics of the MADE tracer plume with reasonable accuracy requiring only a small amount of observations. Among the few predictive transport models for the MADE site, the presented approach 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. The generality of the concept makes it easily transferable to other sites; particularly when focusing on a few, but scale-related measurements.

A hierarchical conductivity structure allows to balance between complexity and available data. Large scale structures determine the mean flow behavior, which is most critical for flow predictions. They can be integrated to a model with reasonable low effort. Structural complexity increases with 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 feature, 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 to identify the need for further field investigations and potential strategies. The adaptive construction based on scale-specific modules allows to create a conductivity structure model 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 less observation data compared to uni-modal heterogeneity models, as log-normal conductivity with high variances. Binary models also allow to incorporate effects of dual-domain transport models 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 stochastically. specific transport analysis of less well investigated heterogeneous sites.

In summary, we conclude:

– ~~When aquifer heterogeneity is at a similar scale as solute transport, predictive transport models need to incorporate spatially distributed hydraulic conductivity.~~

– Modular concepts of conductivity structure allow to separate the multiple scales of heterogeneity. Scale related investigation methods provide field evidence and characterizing model parameters. A hierarchical approach for conductivity can thus minimize the 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 high conductivity contrast, we recommend combining large-scale deterministic structures and simple binary stochastics 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 behavior, particularly in complex aquifers. This can be done with few data and simple tools: adaptive structures allowing to combine deterministic, simple stochastic and geostatistical models depending on the available data and the site-specific modelling aim.

Code and data availability. Transport simulation and geo-statistical code used in this study is available on *https://github.com/GeoStat-Framework*. Data on the MADE aquifer can be accessed via the stated literature sources. Data generated for this study is available upon request to the corresponding author.

Author contributions. All authors contributed to developing the approach and writing the paper. Simulations and figure preparation was performed by AZ.

Competing interests. No competing interests.

Acknowledgements. The authors like to thank Jeff Bohling for background information on Flowmeter and DPIL data.

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 revised 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, pp. 1–20, John Wiley & Sons, Ltd, https://doi.org/10.1002/9781444303957.ch1, 2009.

Carle, S. F. and Fogg, G. E.: Transition probability-based indicator geostatistics, Mathematical Geology, 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.: Flow and Transport on Porous Formations, Springer, New York, 1989.

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.

Fetter, C. W.: Applied Hydrogeology, Prentice Hall, 2000.

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.

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

Freeze, R. A.: A stochastic-conceptual analysis of the one-dimensional groundwater flow in nonuniform homogeneous media, Water Resour. Res., 11, 725–742, https://doi.org/10.1029/WR011i005p00725, 1975.

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

Gomez-Hernandez, 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.

Herweijer, J. C.: Constraining uncertainty of groundwater flow and transport models using pumping tests, in: Calibration and Reliability in Groundwater Modelling, vol. 237, pp. 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.

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.

Kruseman, G. and de Ridder, N., eds.: Analysis and Evaluation of Pumping Test Data, ILRI publication, Wageningen, 47 edn., 2000.

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., Zech, A., and Heße, F.: `ogs5py`: A Python-API for the OpenGeoSys 5 scientific modeling package, Ground Water, under revision, 2020.

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.

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.

Somogyvari, 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.

Werth, C. J., Cirpka, O. A., and Grathwohl, P.: Enhanced mixing and reaction through flow focusing in heterogeneous porous media, Water Resour. Res., 42, W12 414, 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., 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, 2003.