Response to comments on *Feature scale and identifiability: How much information do point hydraulic measurements provide about heterogeneous head and conductivity fields?*

In this document we reproduce the editor, reviewer, and community comments in gray boxes, with our responses following, explaining our reasoning. Explanation of how specifically we have altered the MS in response to the comments appears in **bold**.

Editor's remarks

Two referees provided their assessment: one was very positive, the second one stressed some weaknesses. Basically, the questions posed by the more critical referee concern the fact that the results and discussions cannot be generalized in a straightforward way and the conclusions could somehow depend on the specific hydrogeological setup considered for the numerical test case.

I am well aware that any synthetic aquifer and hydrogeological setup used to test a new method can be subject to this kind of criticism. But, nevertheless, the authors should try to review sections 3 and 4, and conclusions in order to focus better which results can be considered of general validity and which results are more strictly dependent on the hydrogeological set-up considered for the test case. in fact, some sentences (e.g., lines 199 to 203) and the titles of the subsections in section 4 seem to introduce results of general validity.

We appreciate being given the opportunity to address these concerns.

We agree that the generality of the results is an important question which required further discussion. We do believe our results to be applicable beyond the specific system we have analyzed, however **in the revised manuscript, we have made some of our claims more precise, both in the lines identified, and in the conclusion. We now highlight more than once that our study is focused on a paradigmatic 2D steady-state flow scenario**. In this context, we believe it is clear that the headings in section 4 refer to the specific scenario we study. **We have also added a new subsection, 4.4 which addresses our understanding of the generality of our results and their applicability to other flow regimes**. We believe that some of the more critical reviewer's concerns may be related to their understanding of our paper as a model benchmark study rather than the basic science investigation that we understand it to be. Accordingly, **we have added material to the introduction that we believe better clarifies the nature and scope of the study** which may address some reviewer concerns.

We hope that our discussion in this document, as well as the new material added the manuscript will satisfy all outstanding concerns.

Reviewer 1

> The authors present an approach for using Karhunen-Loeve explanations to estimate spatial variability of properties and system states on observations of head, permeability, and velocity magnitude, both with and without regularization. They apply the adjoint approach for efficiently determining gradients of the objective function, allowing for more efficiency and, presumably, high-order approximations. The methods are explained thoroughly, including the assumptions and simplifications. I just have a few minor comments.

We appreciate the reviewer's overall positive assessment.

> 1. A substantial amount (about 20%) of the paper was devoted to developing the adjoint equation and the adjoint-based form of the derivatives of the objective function. Since that was just a tool to be used in the analysis, the detailed development seemed to detract from the main focus of the paper. Could Section 2 be moved to an appendix?

We agree that the flow of the paper is improved by moving many of the derivation details into an appendix. **We have moved the details of the derivation into the appendix**, retaining only the equations required for the main discussion in the body of the manuscript.

> 2. Figure 1 is very helpful as an example of the recovery of spatial distributions of head and permeability through the approach presented in the paper. All other figures show "error" between measures and fits, so Figure 1 is very useful as means of showing the reader the intermediate step. However, Figure 1 is out of place - it appears on p. 11, but it isn't mentioned in the text until page 20. Also more explanation can be provided to make the link between what appears in Figure 1 and how that is related to the data point plotted in the other figures. I wonder also why the two subplots of Figure 1 use different true ln K fields. If subplot a has regularization and subplot b does not, it would be more informative to see that results from the same true ln K field so that the reader can see the benefit of regularization. It would also be helpful to see the numerical value of "error" (the quantities that are plotted in the other figures) to get a sense for where these fits fall in those plots, compared to all other realizations that appear in the plots in Figures 2 and above.

We have added reference to Figure 1 in the introduction in our revised explanation of the aims and scope of the analysis. We have also revised Figure 1 so that both the fitting examples employ the same ground truth. We now also report and discuss the L2 error fitting error associated with the calibrations.

> 3. I would like to see some explanation of the practicality of this method. What measurement density is needed? Is it different for measurements of different quantities?

In our view, one of the most interesting outcomes of this work was the result that the features become identifiable as measurements become denser than 1/4 the feature scale. We also consider some different data (and regularization) combinations in Table 1, Figure 4, and the surrounding discussion. We think that these results may have been under-emphasized, so **we have augmented the discussion and added material to the introduction emphasizing these results**.

Reviewer 2

> This is a well written paper (until the discussion section) on the study of the worth of different types of data for inverse modeling in hydrogeology. Its aim is to try to identify the necessary sampling density in order to identify the underlying hydraulic conductivity structure. While I enjoyed reading the paper and the detailed derivations of the different equations, I had a hard time trying to find the novelty of the work. This question has been studied as early as in the late 1990s,

We are very surprised by these remarks. To our knowledge, our paper has two major original features:

1. The principal novelty of our study is the investigation of the connection between spatial scale of features in the hydraulic conductivity field and our ability to characterize them from a given density of point measurements. As far as we know, this has never been systematically investigated and no one has previously quantified the relationship between amount learned and relative measurement density, as we have. Our key result—that feature identification becomes possible above 4x measurement scale, with quality increasing in power law fashion with increasing feature scale above this threshold—has no real precedents.
2. The Karhunen-Loeve basis function framework we develop for assessing calibration accuracy without requiring the imposition of experimenter-specified a priori spatial zonation or subjective weights on the objective function is itself novel and can be applied (with suitable modifications) to analysis of a variety of different field recovery problems.

The reviewer did not respond to a request for elaboration as to what work in the 1990s they believe anticipated our analysis. Again, no previous study we know of has systematically considered the relationship between feature scale and feature identifiability by means of model calibration.

In reaction to these comments, as well as to some remarks by the first reviewer, **we have augmented our introduction section to clarify our science goals and key results, and to emphasize what is novel**.

> and the application proposed, while very elegant, is limited to a very narrow and little interesting case of steady groundwater flow in an aquifer with prescribed head at the boundary and with an underlying hydraulic conductivity field drawn from a multi Gaussian random function with an exponential covariance. For the paper to be worth to be considered for publication it should have addressed a tougher problem: transient, non-Gaussian field, larger variance, anisotropic, generic boundary conditions, addressing the issue of measurements taken at different supports. As is, the a paper is a very nice mathematical exercise with little practical interest.

We are surprosed to see the argument that the manuscript would only be worthwhile to publish if it considered a "tougher problem" with, e.g., an esoteric non-Gaussian statistical correlation structure, or high variance, or transient flow. This seems strange and perhaps represents a misunderstanding of what we are studying.

If our work were an algorithm benchmarking study where we were trying to select the most robust approach from amongst candidates, we could see the merit in choosing a "tough" benchmark. But our work is not a benchmarking study; rather, it concerns the information content of hydraulic measurements. This basic science question is best answered using a canonical flow model that exhibits the physics we wish to study, employing realistic parameters and standard assumptions but otherwise shedding needless complexity. Therefore, the more niche modeling scenarios suggested by the reviewer are counterproductive for our purposes. We stress: our study is not "about" any specific flow configuration; rather we are studying how reliably groundwater flow transports information about conductivity features to remote measurement locations. Given that the feature scale / identifiability relationship has never been quantified in any scenarios at all, we believe that choosing a common flow scenario as a basis for our analyses is a good choice.

We do take the point that it is possible that the relations we have uncovered may in some sense specific to the assumptions we have made, and realize that it is important to discuss the possible generality of the results. **We discuss the possible generality of the results in a new subsection, 4.4**.

> As a minor point, there was an earlier benchmarking of inverse models not mentioned in the opening, which addressed a tougher problem than the one discussed here published in Water Resources Research by Zimmerman et al. In 1988.

We appreciate that the reviewer brought this systematic inversion technique benchmarking paper to our attention. **We now mention the Zimmerman paper in the introduction as an example of the few papers performing rigorous comparison of inversion approaches**, preceding the works of Franssen et al. (2009) and Illman et al. (2010).

Community comment

> General comments
>
> Good theoretical research with implication on groundwater flow modelling and the engineering of the reservoirs where the geological flow heterogeneities are of paramount importance. Please, follow my guidance to improve the manuscript.

We appreciate the commenter's positive overall assessment.

> Specific comments
>
> Line 6. I suggest "This technique allows unbiased". Add the word "technique".

We have done this.

> Lines 20-22. Mini-permeameter, slug, packer and pumping tests can be also used to identify flow heterogeneities and determine the hydraulic conductivity. Specify this point.
>
> Line 22. "Point-to-point tracer tests" to detect flow heterogeneities. Please, add recent literature on the topic:
>
> - Deleu, R., Frazao, S. S., Poulain, A., Rochez, G., & Hallet, V. (2021). Tracer Dispersion through Karst Conduit: Assessment of Small-Scale Heterogeneity by Multi-Point Tracer Test and CFD Modeling. Hydrology, ((4)
>
> - Lorenzi, V., Banzato, F., Barberio, M. D., Goeppert, N., Goldscheider, N., Gori, F., Lacchini A., Manetta M., Medici G., Rusi S., Petitta, M. (2024). Tracking flowpaths in a complex karst system through tracer test and hydrogeo-

chemical monitoring: Implications for groundwater protection (Gran Sasso, Italy). Heliyon, 10(2)

– Poulain, A., Rochez, G., Van Roy, J.P., Dewaide, L., Hallet, V. and De Sadelaer, G., 2017. A compact field fluorometer and its application to dye tracing in karst environments. Hydrogeology Journal, 25

Lines 48-88. The literature on the topic is much broader.

[....]

Line 514. Please, add relevant literature on the topic.

The commmenter suggested substantially increasing the literature review. Realistically, the literature somewhat related to our topic is so broad that we could plausibly have surveyed hundreds or thousands of papers. We have kept the bibliography largely as it was, save for adding two references in response to remarks by Reviewers 1 and 2. The suggested literature on karst terrain in particular is quite far afield: we discuss heterogeneous, but continuous permeability fields exclusively in our manuscript.

Line 84. Disclose the 3 to 4 specific objectives of your research by using numbers (e.g., i, ii and iii).

We agree it is useful to outline the objectives around here. In response to this, and to remarks from the reviewers, **we have added a new paragraph at the start of subsection 1.3 enumerating our specific high-level objectives.**

Line 90. “Feature scale”. This expression is difficult to understand. Do you mean “observation scale”?

We see that this was not clear in the original draft. **We now define what we mean by the term in the first sentence of the introduction of the revised paper.**

Line 199. “theoretical observations”. Can you re-call the key equations instead?

We now reference the key equations explicitly here.

Line 501. “Other geophysical fields”. (i) remind to the reader that the principal implications are in the calibration of groundwater flow models, (ii) which other implications/applications in geophysics? You can look at my general comments.

Regarding (i) **we now explicitly recall the fields studied**. Regarding (ii), we feel that it is best to leave the possible applications outside our field of study general.

Figures and tables

Figure 1. Very nice figure, but it needs some changes. (i) Make the rectangles closer, (ii) make words and numbers larger, and (iii) thicker the black nodes.

Figures 2-4. Make words and figures larger on x and y axes.

We believe the figure labels are large enough that they should be legible even if a figure is shrunk to single-column size; we have used the same formatting previously in other articles and it has rendered well in the published version. In terms of other stylistic comments, we think the figures are effective and readable as they are. Thus, we have not altered the figures, except for revised Figure 1 in response to Reviewer 1.