Feature scale and identifiability: How much information do point hydraulic measurements provide about heterogeneous head and conductivity fields?

Scott K. Hansen[1], Daniel O'Malley[2], and James P. Hambleton[3]

[1]Zuckerberg Institute for Water Research, Ben-Gurion University of the Negev, Israel
[2]EES-16, Los Alamos National Laboratory, USA
[3]Department of Engineering, Cambridge University, UK

Correspondence: Scott K. Hansen (skh@bgu.ac.il)

Abstract. We systematically investigate how the spacing and type of point measurements impacts the scale of subsurface features that can be identified by groundwater flow model calibration. To this end, we consider the optimal inference of spatially heterogeneous hydraulic conductivity and head fields based on three kinds of point measurements that may be available at monitoring wells, namely of head, permeability, and groundwater speed. We develop a general, zonation-free technique for Monte Carlo (MC) study of field recovery problems, based on Karhunen-Loève (K-L) expansions of the unknown fields whose coefficients are recovered by an analytical, continuous adjoint-state technique. This technique allows unbiased sampling from the space of all possible fields with a given correlation structure and efficient, automated gradient-descent calibration. The K-L basis functions have a straightforward notion of wavelength, revealing the relationship between feature scale and reconstruction fidelity, and they have an *a priori* known spectrum, allowing for a non-subjective regularization term to be defined. We perform automated MC calibration on over 1100 conductivity-head field pairs, employing a variety of point measurement geometries and evaluating the mean-squared field reconstruction accuracy, both globally and as a function of feature scale. We present heuristics for feature scale identification, examine global reconstruction error, and explore the value added by both the groundwater speed measurements and by two different types of regularization. We find that significant feature identification becomes possible as feature scale exceeds four times measurement spacing and identification reliability subsequently improves in a power law fashion with increasing feature scale.

1 Introduction

In the following, we quantify the reliability of reconstruction of heterogeneous hydraulic conductivity field features—meaning regions of locally similar hydraulic conductivity—of different scales, given different densities and kinds of point hydraulic measurements. We study a basic question that is inherent in all remote sensing, which is: how much information can be extracted, and at what resolution? A rigorous feature scale perspective, which we have not seen before in the context of the inverse problem of hydrogeology, is illuminating because model calibration may succeed in capturing large-scale features of hydrogeological relevance while the L2 (integral square) error of the recovered field relative to the true field remains mislead-

ingly high. In the middle column of Figure 1, we show two log-conductivity fields that have been reconstructed using gradient descent model calibration. While both reconstructed fields qualitatively match the true field, in both cases calibration was able to remove only 2/3 of the L2 error relative to the underlying field. However, the remaining error pertains to higher wavenumber (i.e., spatial frequency) features which may be of less importance.

1.1 Detailed motivation

For many forms of groundwater modeling, it is common to infer the spatially heterogeneous hydraulic conductivity and head fields which together determine groundwater flow from sparse, passively collected measurements at groundwater wells. While significant high-resolution information about heterogeneity may be obtained from a wide variety of interventions including geophysical measurements (Slater, 2007; Szabó, 2015), hydraulic tomography (Illman et al., 2007) and both push-pull (Hansen et al., 2017) and point-to-point (Irving and Singha, 2010) tracer tests, such data is often unavailable at scales of interest. Due to their prevalence in practical, field-scale model calibration, we focus here on three types of point information that are readily and commonly determined at monitoring wells. These are hydraulic head, $\tilde{h}$, obtained from direct piezometry, Darcy flux magnitude, $\tilde{q}$, obtained from point dilution tests, and log hydraulic conductivity, $\ln \tilde{K}$, from slug tests or Kozeny–Carman core sample analyses. Calibrating a spatially-nonuniform hydraulic conductivity field at high resolution from sparse measurements of these three types is generally a highly under-determined inverse problem that defies exact solution.

Model calibration is premised on the assumption that fitting the observable data will sufficiently constrain unobserved underlying fields to enable useful conclusions, even given major uncertainty. However, given only point measurements of a given resolution, it is not immediately clear how accurate the estimated head and log hydraulic conductivity fields will be. Recalling the well-known 1D result that the drop in head along a streamline segment is determined by the harmonic mean of the conductivities encountered, and not their order, we see it is not possible to recover information about small-scale variations from endpoint head, flux and conductivity. However, we may gain hope from knowledge that the head field is generally much smoother than the underlying conductivity field, and that a full suite of measurements at each well will provide us with both the head, $\tilde{h}$, and $\nabla \tilde{h}$: the first two terms of the Taylor expansion of this smooth function around each well. To the extent that it is possible to correctly reconstruct the head distribution at locations away from the wells, we *will* be able to partially constrain the conductivity and to identify its large-scale features—meaning contiguous regions of similar hydraulic conductivity.

Additionally, it may be possible to employ regularization techniques that do not depend on detailed site knowledge to further improve the reconstruction quality. The question is: how much can we learn, at which scales?

To our knowledge, the degree to which density and type of measurement relate to the scale of feature that can reliably be recovered, and to the global error of the recovered field, has not previously been examined systematically. We seek to address these matters in two ways. Firstly, we aim to derive a quantitative cutoff and scaling relationship relating feature scale identifiability to measurement density. Secondly, we wish to make concrete, empirically grounded statements about the relative value of the different types of point data and regularization schemes to accurate inference of hydraulic conductivity and head fields using standard gradient-descent model calibration.

1.2 Previous contributions

Being essential to reliable modeling, the inverse problem of reconstructing hydraulic conductivity fields from incomplete data has long interested subsurface hydrologists. The literature on this problem has become so large over the decades that it even contains a substantial *meta*-literature of review papers (e.g., Bagtzoglou and Atmadja, 2005; Carrera et al., 2005; Hendricks Franssen et al., 2009; Zhou et al., 2014). Interested readers are referred to these papers; we will only briefly summarize some key developments that contextualize our work.

From a purely mathematical standpoint, the hydraulic problem is ill-posed, being non-unique, and additional constraints are required to regularize to a unique, physically plausible solution. Neuman (1973) proposed an explicit regularization scheme employing explicit smoothness and boundedness constraints on the conductivity field. Yoon and Yeh (1976) proposed a similar scheme, in which the solution was expressed as a superposition of finite element shape functions, with the boundedness constraints applied to the shape functions themselves. An alternative dimension-reduction approach based on zonation with fixed parameters in each zone was proposed by Yeh and Yoon (1981). An alternative two-stage approach, limited to small conductivity variances (Zhou et al., 2014), is based on first identifying the geostatistical spatial correlations of head and permeability, followed by co-kriging to estimate conductivity at locations of interest (Kitanidis and Vomvoris, 1983; Hoeksema and Kitanidis, 1984). In the context of the general nonlinear conductivity calibration problem, where variance is not generally small and an initial guess may be far from the true solution, iterative variants on the geostatistical approach have been proposed Kitanidis (1995); Yeh et al. (1995); Cardiff et al. (2009). Explicitly Bayesian approaches that view model error as the key source of uncertainty (Carrera and Neuman, 1986a), and in which regularization is viewed as coming from a prior probability distribution (Kitanidis, 1986; Woodbury and Ulrych, 2000) have also been proposed. Analytical covariance relationships between head and permeability are valid for small perturbations. For work in large-dimensional spaces (such as calibration of high-resolution conductivity fields), reduction of non-uniqueness by use of principal component dimension reduction has also been proposed (Tonkin and Doherty, 2005; Kitanidis and Lee, 2014), representing a sort of pre-regularization.

The relative performance of the various techniques has not much been much studied. Exceptions we are aware of include the comprehensive numerical inter-comparison studies of Zimmerman et al. (1998) and Hendricks Franssen et al. (2009) and the bench-scale study of Illman et al. (2010), which only compared hydraulic tomography with simple averaging and kriging. Hendricks Franssen et al. (2009) reported little performance difference among the various techniques they compared.

Additionally, researchers have tried approaches that improve uniqueness by addition of other forms of physical data. These include hydraulic tomography, in which the head field is manipulated by pumping at wells surrounding the area of interest (e.g., Gottlieb and Dietrich, 1995; Yeh and Liu, 2000; Illman et al., 2007), use of transient head data (e.g., Carrera and Neuman, 1986a; Zhu and Yeh, 2005), flux measurements (Tso et al., 2016), chemical (Wagner, 1992; Michalak and Kitanidis, 2004; Xu and Gómez-Hernández, 2018; Delay et al., 2019) and thermal (Woodbury et al., 1987) tracer data. These are largely out of scope for the present work, which focuses on steady-state hydraulic inversion only, although we will re-examine use of local groundwater speed (i.e., flux) information.

As inversion is typically performed via gradient descent, it is necessary to estimate the gradient of the loss function with respect to the parameters. Where the loss function is expensive to compute and the parameter space is high dimensional, it is generally not feasible to estimate the gradient naively via successive parameter perturbation. Various solutions exist, including automatic differentiation Elizondo et al. (2002); Sambridge et al. (2007); Wu et al. (2023), and the adjoint state method, initially applied to the groundwater inverse problem by Sykes et al. (1985). Adjoint state formulations for groundwater model inversion commonly incorporate head data only (Sykes et al., 1985; Carrera and Neuman, 1986b; Lu and Vesselinov, 2015; Delay et al., 2019), although they have been formulated to incorporate other forms of information (Cirpka and Kitanidis, 2001). Both discrete and continuous adjoint-state formulations have been studied by Delay et al. (2017) and Hayek et al. (2019), with the former identifying advantages for the continuous approach. We have not seen formulations for a loss function that includes conductivity and speed measurements, or regularization.

1.3 Concrete goals and overview of contents

At high level, our goals for the research are two-fold. Primarily, we aim to quantify the spatial scale at which it becomes possible to identify coherent conductivity field features as a function of hydraulic point measurement spacing, and how our ability to identify these features increases with increasing measurement density. We do so in the context of a paradigmatic steady-state 2D system, but we expect that the results uncovered will be of more general applicability. Secondarily, we aim to study the relative value of three types of point measurement (conductivity, head, and groundwater speed) and two types of empirically-grounded regularization schemes (Tikhonov and series truncation) on field reconstruction performance.

To study the inter-relation of measurement density, feature scale and identifiability, we develop an automated gradient descent calibration Monte Carlo (MC) framework with the following properties:

1. Samples initial guesses in an unbiased fashion from the space of possible conductivity fields.

2. Avoids arbitrary *a priori* zonation and structural constraints.

3. Output easily interpreted in terms of feature scale reconstruction reliability.

4. Sufficiently low computational cost.

We satisfy the first three properties by generating fields with Karhunen-Loève (K-L) expansions and working directly with their coefficients rather than zonating the domain and calibrating constant conductivities on fixed regions. We select a covariance kernel such that the K-L basis functions can be determined analytically, and such that they each have an obvious spectral interpretation. As the basis functions are orthonormal, they admit a Fourier-type analysis of the reconstruction. The last property is satisfied deriving a continuous adjoint-state sensitivity of loss function to measurements, reducing the computational cost to the size of the measurement vector rather than the vector of K-L coefficients.

We stress that, although we develop a framework for quantifying the amount of learning that occurs during model calibration, we are not adding to the literature of bespoke techniques for the inverse problem of hydrogeology. Ultimately, our calibration

120 technique amounts to gradient descent minimization of a squared-error objective function, which is a standard approach in contemporary machine learning.

In Section 2 we derive the continuous adjoint-state form of the optimization equations for the steady-state groundwater flow equation subject to a loss function containing point data about head, flux, and conductivity, as well as an arbitrary regularization term. In Section 3 we derive the K-L basis functions and eigenvalues, present fitness metrics, discuss regularization, and

125 formalize two MC studies. In Section 4 we present results, derive a quantitative relation between measurement density and feature reconstruction reliability, discuss global error, and numerically compare the utility of various types of measurements and regularization approaches. In Section 5, we summarize our key conclusions and point towards future work.

2   Derivation of continuous adjoint-state optimization equations

We look to solve for a spatially-distributed hydraulic conductivity field, $K(\mathbf{x}; \mathbf{p})$, which is defined by some parameter vector,

130 $\mathbf{p}$, and exists throughout some domain, Ω, and where $\mathbf{x}$ represents the spatial coordinate within Ω. Optionally, we try to solve simultaneously for the specified-head boundary condition $h|_{\partial\Omega}$. We take for granted that our system obeys the steady-state groundwater flow equation $\nabla \cdot (K \nabla h) = 0$, where $h(\mathbf{x})$ represents hydraulic head and K, a scalar, represents an isotropic, local hydraulic conductivity. We further assume that we have accurate measurements of head, $\tilde{h}(\mathbf{x}_j)$, and/or Darcy flux magnitude, $\tilde{q}(\mathbf{x}_k)$, and log hydraulic conductivity, $\ln \tilde{K}_l$ at a number of discrete points, $\mathbf{x}_k$, within Ω.

135   We will be obtaining K by gradient descent on some loss function, J, representing mismatch between model predictions and point measurements, as a function of a vector of model parameters, $\mathbf{p}$, meaning we require the sensitivity vector $\frac{\partial J}{\partial \mathbf{p}}$. However, because h is measured and will appear explicitly in the formula for J we expect that we will need to determine $\frac{\partial h}{\partial \mathbf{p}}$, which is computationally intractable for large-dimensional $\mathbf{p}$. Consequently, we aim derive an adjoint equation that, when satisfied, eliminates the dependence of $\frac{\partial J}{\partial \mathbf{p}}$ on $\frac{\partial h}{\partial \mathbf{p}}$. We do this below.

140   We begin with the Dirichlet BVP for the steady-state groundwater flow equation:

$$\nabla \cdot (K \nabla h) \quad = \quad 0 \text{ in } \Omega, \tag{1}$$
$$h \quad = \quad g(\mathbf{x}) \text{ on } \delta\Omega, \tag{2}$$

where $g(\mathbf{x})$ is a known function.

2.1   Loss function

145 We attempt to minimize J, which consists of as many as four metrics:

1. the weighted square error of our modeled Darcy speeds, $K(\mathbf{x}_j; \mathbf{p}) \|\nabla h(\mathbf{x}_j; \mathbf{p})\|$, relative to measurements, $\tilde{q}(\mathbf{x}_k)$ at the measurement locations, $\mathbf{x}_j$

2. the weighted squared error of our modeled heads, $h(\mathbf{x}_k)$, relative to our measurements, $\tilde{h}(\mathbf{x}_k)$, at the measurement locations, $\mathbf{x}_k$,

5

3. the weighted squared error of our modeled log hydraulic conductivity, $\ln K(\mathbf{x}_l)$, relative to our measurements, $\ln \tilde{K}(\mathbf{x}_k)$, at the measurement locations, $\mathbf{x}_l$, and optionally

4. a regularization term, R, based on prior notions of how plausible a given underlying log-conductivity is.

This can be expressed mathematically as

$$
\begin{aligned}
J \equiv & \sum_j w_j \left(K(\mathbf{x}_j; \mathbf{p}) \|\nabla h(\mathbf{x}_j; \mathbf{p})\| - \tilde{q}(\mathbf{x}_j) \right)^2 \\
&+ \sum_k w_k \left(h(\mathbf{x}_k; \mathbf{p}) - \tilde{h}(\mathbf{x}_k) \right)^2 \\
&+ \sum_l w_l \left(\ln K(\mathbf{x}_l; \mathbf{p}) - \ln \tilde{K}(\mathbf{x}_l) \right)^2 \\
&+ R\left(K(\mathbf{x}; \mathbf{p}) \right)
\end{aligned}
\tag{3}
$$

Here we explicitly acknowledge the dependence of h and K on a vector of model parameters, $\mathbf{p}$, which generally encodes the model parameters and the boundary conditions. The weights w_j, w_k, and w_l, which may be zero, may reflect the relative areas of the Voronoi regions associated with each well, as well as a relative importance rating of the types of measurement. The magnitude of the regularization function may also vary by an arbitrary scalar to change its weight relative to measurement misfit. For simplicity we will generally not write this dependence explicitly.

2.2   Adjoint equation form

In Appendix A, we show the manipulations required to eliminate the sensitivity of h on each of the parameters defining the flow field in (3). Inspired by (A11), we propose to define ϕ^* to solve the BVP

$$
\nabla \cdot (K \nabla \phi^*) + 2 \sum_k w_k \left(h(\mathbf{x}_k) - \tilde{h}(\mathbf{x}_k) \right) = 0 \text{ in } \Omega,
\tag{4}
$$

$$
\phi^* = 0 \text{ on } \delta\Omega,
\tag{5}
$$

which we note is consistent with the adjoint BVP derived in Sykes et al. (1985). When ϕ^* is so defined, it eliminates two of the boundary integral terms and one of the volume integral terms in (A9), leaving the simplified sensitivity equation

$$
\begin{aligned}
\frac{\partial J}{\partial \mathbf{p}} = & \int_\Omega 2 \left(K \|\nabla h\| - \tilde{q} \right) \frac{\partial K}{\partial \mathbf{p}} \|\nabla h\| \sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j) \, dV \\
&+ \int_\Omega \frac{2}{K} \frac{\partial K}{\partial \mathbf{p}} \left(\ln K - \ln \tilde{K} \right) \sum_l w_l \delta(\mathbf{x} - \mathbf{x}_l) \, dV + \frac{\partial R(K)}{\partial \mathbf{p}} \\
&- \int_\Omega \frac{\partial K}{\partial \mathbf{p}} \nabla h \cdot \nabla \phi^* \, dV - \int_{\delta\Omega} K \phi \nabla \phi^* \cdot dS.
\end{aligned}
\tag{6}
$$

We show here the specific dependence of the regularization term on the K field alone.

2.3 Sensitivity to specific parameters

The vector $\mathbf{p}$ may contain two types of parameters: those defining the Type I (specified head) boundary condition, and those defining field $K(\mathbf{x})$, itself. We note that for any given parameter, p_i, the sensitivity expression (6) simplifies.

For parameters, p_i, defining the conductivity field, we observe that, because our domain has a specified-head boundary condition *which is governed by a different set of parameters* $\phi_i = 0$ in the final term of (6). Thus,

$$
\begin{aligned}
\frac{\partial J}{\partial p_i} = & \int_\Omega 2\left(K\|\nabla h\| - \tilde{q}\right) \frac{\partial K}{\partial p_i} \|\nabla h\| \sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j)\, dV \\
& + \int_\Omega \frac{2}{K} \frac{\partial K}{\partial p_i}\left(\ln K - \ln \tilde{K}\right) \sum_l w_l \delta(\mathbf{x} - \mathbf{x}_l)\, dV + \frac{\partial R(K)}{\partial p_i} \\
& - \int_\Omega \frac{\partial K}{\partial p_i} \nabla h \cdot \nabla \phi^*\, dV,
\end{aligned}
\tag{7}
$$

There are at least two convenient ways that the conductivity field can be parameterized. The first way is that the domain can be discretized, with p_i representing the constant value of K in some subdomain, Ω_i. In this case, $\frac{\partial K}{\partial p_i}$ is 1 in Ω_i, and 0 outside. Thus,

$$
\begin{aligned}
\frac{\partial J}{\partial p_i} = & \int_{\Omega_i} 2\left(K\|\nabla h\| - \tilde{q}\right) \|\nabla h\| \sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j)\, dV \\
& + \int_{\Omega_i} \frac{2}{K}\left(\ln K - \ln \tilde{K}\right) \sum_l w_l \delta(\mathbf{x} - \mathbf{x}_l)\, dV \\
& - \int_{\Omega_i} \nabla h \cdot \nabla \phi^*\, dV + \frac{\partial R(K)}{\partial p_i},
\end{aligned}
\tag{8}
$$

where only measurements inside the sub-domain participate in the sensitivity expression.

The second convenient approach is to write $\ln K(\mathbf{x}) = \sum_i p_i f_i(\mathbf{x})$, where the functions $f_i(\mathbf{x})$ form an orthogonal basis on Ω. This series type of representation has the advantage of being independent of any domain discretization. Here, $\frac{\partial}{\partial p_i}\{\ln K\} = f_i(\mathbf{x})$, so we may adapt (7)

$$
\begin{aligned}
\frac{\partial J}{\partial p_i} = & 2 \sum_j w_j \left[f_i\left(K\|\nabla h\| - \tilde{q}\right) K\|\nabla h\|\right]\Big|_{\mathbf{x}=\mathbf{x}_j} \\
& + 2 \sum_l w_l \left[f_i\left(\ln K - \ln \tilde{K}\right)\right]\Big|_{\mathbf{x}=\mathbf{x}_l} \\
& - \int_\Omega K f_i \nabla h \cdot \nabla \phi^*\, dV + \frac{\partial R(K)}{\partial p_i}.
\end{aligned}
\tag{9}
$$

For parameters defining the specified-head boundary condition, it follows that $\frac{\partial K}{\partial p_i} = 0$, so all but the final term of (6) disappear. Invoking (A5), it follows immediately that:

$$
\frac{\partial J}{\partial p_i} = -\int_{\delta\Omega} K \frac{\partial f}{\partial p_i} \nabla \phi^* \cdot dS.
\tag{10}
$$

As in the case of the conductivity field, the controlling parameters may either specify constant heads along individual boundary segments, or a series-type representation may be used. If Ω is convex, it is convenient to use a classic Fourier series, with an angular variable $\theta \in [0, 2\pi)$ sweeping the boundary like the hand of a clock.

3 Monte Carlo study of calibration

3.1 Overview

Using the sensitivity equation (9) and adjoint BVP (4-5) we study how well hydraulic conductivity and head fields can be reconstructed given point measurements of conductivity, head, and (optionally) groundwater velocity using adjoint-based gradient descent approaches, in the context of a paradigmatic 2D steady-state system. We aim to understand the relationship between the residual error of the fitted model and the L^2 (root integral square) error of the reconstructed conductivity field, and the degree to which this can be improved by the addition of (a) groundwater velocity information and (b) regularization based on statistical information about the conductivity field. In Section 4.4 we discuss the potential for these results to generalize to other systems.

We approach this question via a Monte Carlo study: generating many synthetic target "true" conductivity realizations, and for each target realization, performing gradient descent model calibration multiple times from different random initial guesses. In this way, we are able to assess the relative performance of different calibration methods, as well as of measurement resolution.

We employ two specialized techniques: (i) series expansion in K-L basis functions, which addresses the problem of data generation as well as model-independent conductivity field representation, and (ii) adjoint-state inversion by steepest descent from initial guesses generated using the approach in (i). This addresses the issue of unbiased sampling of the model space as well as the computational tractability, as the computational cost of the adjoint-state approach is dictated by the dimension of the observation space, rather than the (much larger) dimension of the model space for under-determined inverse problems.

Any Monte Carlo study must necessarily be simplified relative to reality. While we have simplified the problem to make it more amenable to mathematical analysis, we expect the patterns that emerge to hold more broadly. Although our mathematical analysis is fully general for steady-state groundwater flow, we restrict the current study to reconstruction to two spatial dimensions and employ a separable exponential covariance structure for which the K-L basis functions may be analytically determined in closed form. Explicit expansion of the fields in these basis functions is advantageous, because it allows straightforward spectral interpretation. We also assume that head boundary conditions are known, encoding a physically reasonable hydraulic gradient in the x-direction. We assume that a complete set of error-free, uniformly spaced measurements are available, allowing analysis to focus directly on the information content of the measurements and error induced by aliasing of unresolved high-wavenumber features. We further assume that true log-conductivity fields are Gaussian with a known correlation structure (equivalently, semivariogram), although except when regularization was employed, the statistics are only used for specification of initial guesses.
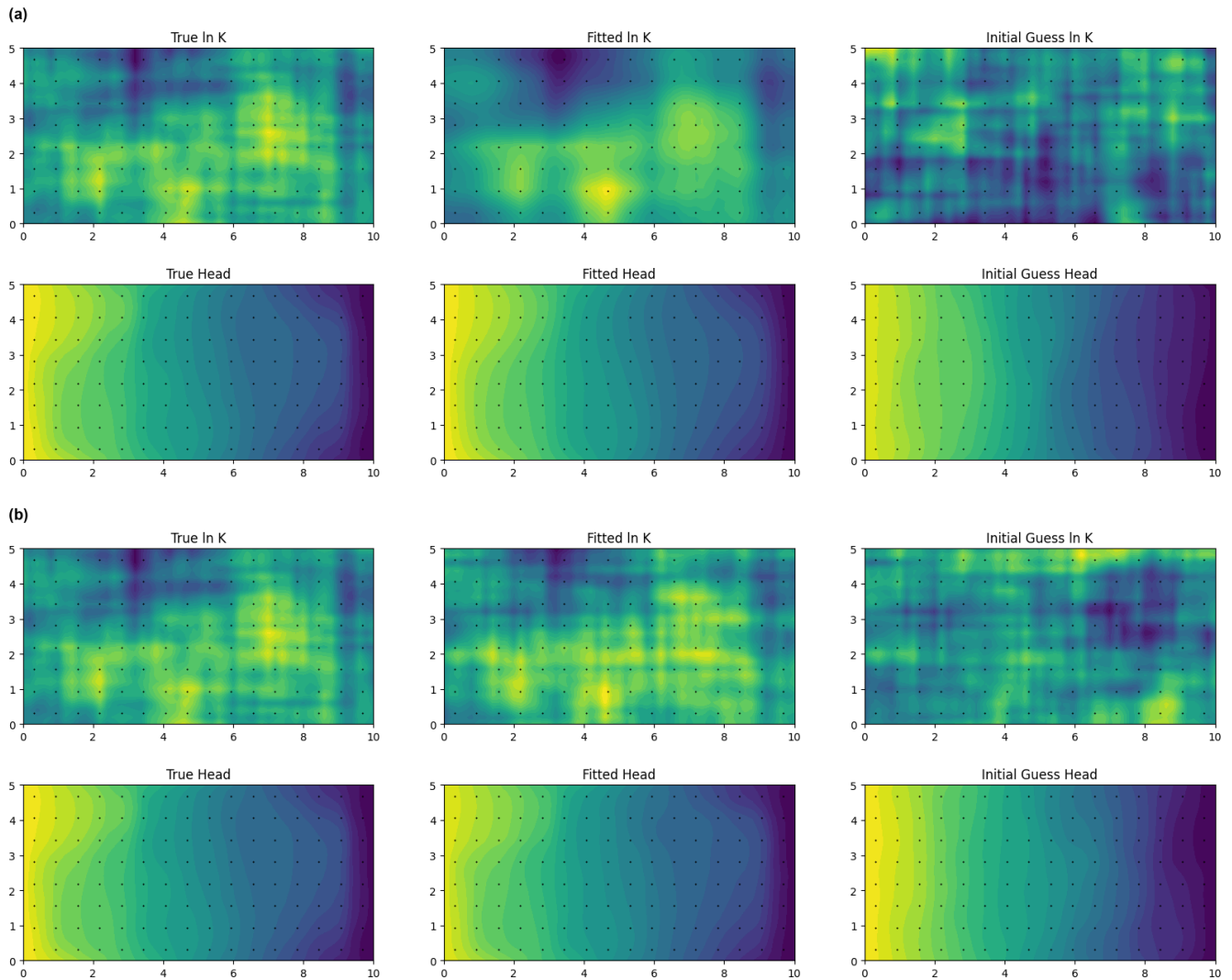
Figure 1. Examples of automated calibration (center column) of true fields $\ln \tilde{K}$ and $\tilde{h}$ (left column) after 500 optimization steps starting from an initially uncorrelated $\ln K$ field (right column), using the adjoint-state approach detailed in this paper. Color maps represent spatial variation of field values: higher value corresponds to lighter color within a given map, but numeric interpretation of a given color varies from map to map. Black dots in each image indicate the locations of simulated measurements. Two calibrations are shown, both featuring the same underlying field and the same data weights, as defined in (37-39), but differing in regularization approach: (a) regularization employed with $w_R = 1/2$, and (b) no regularization ($w_R = 0$). $\Psi = 0.10$ in both cases.

3.2 Representation of conductivity field and boundary conditions

We employ a generalized Fourier series type of representation of target log-conductivity fields as well as calibrated approximations to the target, meaning that fields are represented as vectors of coefficients and model calibration proceeds by iterative solution of (4-5) and (9). The analytic representation of the adjoint-state gradient descent equations combined with the series field representation means that any numerical discretization is completely abstracted out of the analysis, and forward modeling can be performed by a general finite element solver, in our case the FEniCS package for Python.

The actual orthonormal basis functions used are the K-L basis functions corresponding to covariance kernel (11)

$$\text{Cov}(\ln K(\mathbf{x}_1), \ln K(\mathbf{x}_2)) = \sigma_{\ln K}^2 \exp\left(-\frac{|x_1 - x_2|}{\eta_x} - \frac{|y_1 - y_2|}{\eta_y}\right), \tag{11}$$

where the position vectors, $\mathbf{x}_i = (x_i, y_i)$. Use of K-L basis function expansion has been recommended and previously employed successfully in the context of hydraulic inversion by Wang et al. (2021). On our 2D domain $\Omega = [0, L_x] \times [0, L_y]$, the n-th such function may be defined analytically (Zhang and Lu, 2004):

$$f_n(x, y) = \frac{1}{\sqrt{(\eta_x^2 w_{x,n}^2 + 1)L_x/2 + \eta_x}} [\eta_x w_{x,n} \cos(w_{x,n} x) + \sin(w_{x,n} x)] \times \tag{12}$$

$$\frac{1}{\sqrt{(\eta_y^2 w_{y,n}^2 + 1)L_y/2 + \eta_y}} [\eta_y w_{y,n} \cos(w_{y,n} y) + \sin(w_{y,n} y)], \tag{13}$$

where $w_{x,n}$ and $w_{y,n}$ are respectively the n-th positive solutions to

$$(\eta_x^2 w_{x,n}^2 - 1) \sin(w_{x,n} L_x) = 2\eta_x w_{x,n} \cos(w_{x,n} L_x), \tag{14}$$

$$(\eta_y^2 w_{y,n}^2 - 1) \sin(w_{y,n} L_y) = 2\eta_y w_{y,n} \cos(w_{y,n} L_y). \tag{15}$$

Viewing $w_{x,n}L$ as a single argument to both sine and cosine, it is clear that there will be infinitely many solutions, with one occurring in each interval $(k\pi, (k+1/2)\pi)$, where sine and cosine must cross, regardless of vertical scaling (with one extra solution occurring near 0). As $w_{x,n}$ grows, the solution point where $\eta_x w_{x,n} \sin(\cdot) \approx 2\cos(\cdot)$ must, on scale grounds, become very close to the zeroes of the sine function, namely the positive integer multiples of π. A similar argument applies for $w_{y,n}$. Thus, it makes sense to understand the K-L basis functions as analogous to the basis functions of Fourier series, with angular frequencies defined by $w_{x,n}$ and $w_{y,n}$. This makes it easy to interpret the several K-L basis function in terms of feature scale.

K-L basis functions, by definition, form an orthonormal basis so that

$$\int_0^{L_x} \int_0^{L_y} f_m(x, y) f_n(x, y) \, dy \, dx = \delta_{mn}, \tag{16}$$

and behave in such a way that any stochastic series

$$\ln \tilde{K} \equiv \sum_{n=1}^{\infty} \zeta_n \sqrt{\lambda_n} f_n(x, y) \tag{17}$$

where $\zeta_n \sim N(0,1)$, and eigenvalues λ_n are defined

$$\lambda_n = \frac{4\eta_x\eta_y\sigma_{\ln K}^2}{\left(\eta_x^2 w_{x,n}^2 + 1\right)\left(\eta_y^2 w_{y,n}^2 + 1\right)} \tag{18}$$

represents an equally likely realization of a multivariate Gaussian field with covariance structure (11). The eigenvalues, λ_n, decrease with increasing n, representing the decreasing importance of each subspace to explaining the variance of the Gaussian field.

For a truncated K-L series representation with largest coefficient N, we can derive an expression for the portion of the variance, θ, accounted for as follows. We define the total variance as the expected value of the integral of $\ln \tilde{K}$ over the entire domain. It follows from the definition of the pointwise variance, $\sigma_{\ln K}^2$, that

$$\mathrm{E}\left[\int_0^{L_x}\int_0^{L_y}\left(\ln \tilde{K}\right)^2 dy\, dx\right] = L_x L_y \sigma_{\ln K}^2 \tag{19}$$

Substituting in series representation (17), changing summation and integration order, and applying (16) yields

$$\int_0^{L_x}\int_0^{L_y}\left(\ln \tilde{K}\right)^2 dy\, dx = \int_0^{L_x}\int_0^{L_y}\left(\sum_{n=1}^{\infty}\zeta_n\sqrt{\lambda_n}f_n(x,y)\right)^2 dy\, dx \tag{20}$$

$$= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\zeta_m\zeta_n\sqrt{\lambda_m}\sqrt{\lambda_n}\int_0^{L_x}\int_0^{L_y}f_m(x,y)f_n(x,y)\, dy\, dx \tag{21}$$

$$= \sum_{n=1}^{\infty}\zeta_n^2\lambda_n. \tag{22}$$

Taking the expectation and observing that $\zeta_n \sim \mathcal{N}(0,1)$ yields a series expression for the total variance,

$$\mathrm{E}\left[\int_0^{L_x}\int_0^{L_y}\left(\ln \tilde{K}\right)^2 dy\, dx\right] = \sum_{n=1}^{\infty}\lambda_n. \tag{23}$$

Thus, by truncating the RHS of (23) to N terms and normalizing it via the RHS of (19), we arrive at the expression

$$\theta = \frac{1}{L_x L_y \sigma_{\ln K}^2}\sum_{n=1}^{N}\lambda_n. \tag{24}$$

Both target and initial guess fields are represented by length-N coefficient vectors, containing the coefficients ζ_n in (17). Because in our studies the only parameters being fit are the coefficients defining the log conductivity field, we can use the same K-L basis function expansion form to explicitly represent our calibrated approximations of the true $\ln \tilde{K}$:

$$\ln K(x) = \sum_n p_n \sqrt{\lambda_i} f_n(x). \tag{25}$$

While all solutions are expressed in terms of the K-L functions corresponding to the true covariance structure, this in itself does not prejudice the solution, as these functions form a complete basis. One may wonder if the initial guesses being drawn

from the same distribution as the true field in some way biases them towards the "correct" field. However, we are interested in how cost function J varies with L^2 reconstruction error during calibration, not the absolute magnitude of each, and the adjoint-state calibration equations (4-5) and (9) do not make use of the correlation information. This metric should not depend on the statistical structure of the initial guesses, as long as they are uncorrelated with their corresponding target field.

3.3 Bayesian selection of misfit and regularization weights

The units of speed, conductivity, and head are incompatible, so the sets of coefficients $\{w_j\}$, $\{w_j\}$, and $\{w_k\}$ implicitly have different units. It is necessary to select defensible values for these, and to select the regularization term $R(\mathbf{p})$. To select these optimally, we adopt a maximum *a posteriori* Bayesian standpoint, and imagine that the goal of our calibration is to determine the most probable parameter vector $\arg\max_{\mathbf{p}} \pi_{\mathbf{P}|\mathbf{M}}(\mathbf{p}|\mathbf{m})$, where $\mathbf{m}$ is a vector of all point measurements. Here, $\pi_{\mathbf{P}|\mathbf{M}}$ represents a conditional probability density function, where $\mathbf{P}$ and $\mathbf{M}$ are understood as vector-valued random variables with $\mathbf{p}$ and $\mathbf{m}$ as respective possible values. We will use similar notation for all probability density functions. It follows from Bayes' formula that

$$\pi_{\mathbf{P}|\mathbf{M}}(\mathbf{p}|\mathbf{m}) \propto \pi_{\mathbf{M}|\mathbf{P}}(\mathbf{m}|\mathbf{p})\pi_{\mathbf{P}}(\mathbf{p}) \tag{26}$$

The maximum probability vector, $\mathbf{p}^{\max}$, equivalently minimizes the negative logarithm of the probability

$$\mathbf{p}^{\max} = \arg\min_{\mathbf{p}} -\ln \pi_{\mathbf{P}|\mathbf{M}}(\mathbf{m}|\mathbf{p}) - \ln \pi_{\mathbf{P}}(\mathbf{p}). \tag{27}$$

The first term on the RHS, the *likelihood*, will give rise to the weights w_j, w_j, and w_k, and the second term, the *prior*, will give rise to the regularization term. Let us consider it first. By nature of the K-L expansion, all of the coefficients are independent and identically distributed according to $\mathcal{N}(0,1)$,

$$\pi_{\mathbf{P}}(\mathbf{p}) \propto \prod_n \exp\left(-\frac{p_n^2}{2}\right), \tag{28}$$

$$-\ln \pi_{\mathbf{P}}(\mathbf{p}) = \sum_n \frac{p_n^2}{2} + \text{constant}, \tag{29}$$

where the constant is irrelevant from the perspective of minimization. We can adopt a similar approach for the likelihood term, defining the misfit-to measurement quantities

$$\Delta_j \equiv K(\mathbf{x}_j; \mathbf{p}) \|\nabla h(\mathbf{x}_j; \mathbf{p})\| - \tilde{q}(\mathbf{x}_j) \tag{30}$$

$$\Delta_k \equiv h(\mathbf{x}_k; \mathbf{p}) - \tilde{h}(\mathbf{x}_k) \tag{31}$$

$$\Delta_l \equiv \ln K(\mathbf{x}_l; \mathbf{p}) - \ln \tilde{K}(\mathbf{x}_l) \tag{32}$$

By a symmetry argument, all of these quantities are zero mean before optimization and, in the absence of regularization, remain so throughout optimization. This is seen via the following argument: As of the initial guess, $\mathbf{p}^0$, they are computed by

applying some fixed deterministic function to two independent, identically distributed random vectors ($\mathbf{p}^0$ and ζ) and taking the difference between them. The expected value of this is zero. For any given pairing of true coefficient vector ζ and initial guess vector $\mathbf{p}^0$, the pairing $-\zeta$ and $-\mathbf{p}^0$ is equally probable. Because the loss function only "sees" the squared deviation, it will respond in a symmetrical way to the original pairing as to the negated pairing: if `ParamIterStep`$(\zeta, \mathbf{p}^0) \to \mathbf{p}^1$, then `ParamIterStep`$(-\zeta, -\mathbf{p}^0) \to -\mathbf{p}^1$. Here, `ParamIterStep` represents the algorithm employed to iterate (improve) the parameter vector once by gradient descent. Thus, the expected value of the update vector is the zero vector. Inductively, the expected value of the deviations remains zero after any number of steps. Only the directional bias of a regularization term breaks the symmetry.

The variances of these quantities will vary in an unpredictable manner over the course of model calibration. Before calibration, by symmetry, we expect, e.g., $\mathrm{Var}(\Delta_j) = 2\,\mathrm{Var}(\tilde{h})$. After successful calibration, $\mathrm{Var}(\Delta_j) \approx 0$. In general, we approximate $\mathrm{Var}(\Delta_j) \approx \sigma_{\tilde{q}}^2$, $\mathrm{Var}(\Delta_k) \approx \sigma_{\tilde{h}}^2$, and $\mathrm{Var}(\Delta_k) \approx \sigma_{\ln \tilde{K}}^2$.

Assuming Δ_j, Δ_k, and Δ_l may be treated as Gaussian distributed, we write

$$\pi_{\mathbf{M}|\mathbf{P}}(\mathbf{m}|\mathbf{p}) \propto \prod_j \exp\left(-\frac{\Delta_j^2}{2\sigma_{\tilde{q}}^2}\right) \prod_k \exp\left(-\frac{\Delta_k^2}{2\sigma_{\tilde{h}}^2}\right) \prod_l \exp\left(-\frac{\Delta_l^2}{2\sigma_{\ln \tilde{K}}^2}\right). \tag{33}$$

$$-\ln \pi_{\mathbf{M}|\mathbf{P}}(\mathbf{m}|\mathbf{p}) = \sum_j \frac{\Delta_j^2}{2\sigma_{\tilde{q}}^2} + \sum_k \frac{\Delta_k^2}{2\sigma_{\tilde{h}}^2} + \sum_j \frac{\Delta_j^2}{2\sigma_{\ln \tilde{K}}^2} + \text{constant}. \tag{34}$$

Combining (26), (29), and (34) and inserting the definitions for Δ_j, Δ_k, and Δ_l gives us the expression to minimize

$$
\begin{aligned}
-\ln \pi_{\mathbf{M}|\mathbf{P}}(\mathbf{m}|\mathbf{p}) = {}& \sum_j \frac{1}{2\sigma_{\tilde{q}}^2} \left(K(\mathbf{x}_j;\mathbf{p}) \|\nabla h(\mathbf{x}_j;\mathbf{p})\| - \tilde{q}(\mathbf{x}_j)\right)^2 \\
& + \sum_k \frac{1}{2\sigma_{\tilde{h}}^2} \left(h(\mathbf{x}_k;\mathbf{p}) - \tilde{h}(\mathbf{x}_k)\right)^2 \\
& + \sum_j \frac{1}{2\sigma_{\ln \tilde{K}}^2} \left(\ln K(\mathbf{x}_l;\mathbf{p}) - \ln \tilde{K}(\mathbf{x}_l)\right)^2 \\
& + \sum_n \frac{1}{2} p_n^2 + \text{constant}.
\end{aligned}
\tag{35}
$$

The constant is irrelevant from the perspective of minimization. By comparison with loss function J defined in (3), we are thus motivated to define

$$R(\mathbf{p}) \equiv w_R \sum_n p_n^2, \tag{36}$$

315 and to define the various weighting coefficients.

$$w_j = \frac{1}{2\sigma_{\tilde{q}}^2} \qquad\qquad \forall j, \tag{37}$$

$$w_k = \frac{1}{2\sigma_{\tilde{h}}^2} \qquad\qquad \forall k, \tag{38}$$

$$w_l = \frac{1}{2\sigma_{\ln \tilde{K}}^2} \qquad\qquad \forall l, \tag{39}$$

$$w_R = \frac{1}{2}. \tag{40}$$

320 The relevant variances were determined by generating an ensemble of 20 realizations via K-L expansion, solving the groundwater flow equation (1-2) on them, and then generating simulated measurements. At each location, for each of the three types of measurement, the average value was computed, and the ensemble variances, $\sigma_{\ln K}^2$, $\sigma_{K||\nabla h||}^2$, and σ_h^2 were computed for each value about its appropriate mean value.

3.4 Mechanics of Monte Carlo study

325 A standard rectangular field was employed for all simulations, with dimensions $L_x = 10$, $L_y = 5$, and all simulated K fields featured the same separable 2D correlation structure as described in Section 3.2, with $\eta_x = \eta_y = 2$ and $\sigma_{\ln K}^2 = 2$, generated by K-L expansion. Enough basis functions were employed to account for 98.7% of the variance: the 2000 K-L basis functions with the largest eigenvalues were employed to generate the random fields. Any given realization is completely represented by a 2000-component vector, ζ, containing its K-L coefficients. The length scales L_x and L_y are artifacts of representing a 330 conceptually infinite random field on a computer, and we aimed to select them large enough that their values do not affect our conclusions about the information content of a given density of point measurements. Apart from L_x and L_y, the only physically significant length scales are the correlation lengths, η_x and η_y, and these are used in our analysis.

For each realization, simulated measurements were computed on uniformly-spaced square lattices with point separation Δ_m by solution of (1-2) using the FEniCS finite element system (Logg et al., 2012), employing a 100 by 50 finite element grid 335 with first-degree accurate square finite elements. Measurement density relative to field correlation length is represented by the dimensionless quantity

$$\Psi \equiv \frac{\Delta_m^2}{\eta_x \eta_y}. \tag{41}$$

We posit that effective measurement density controls our ability to reconstruct features of various scales, and that moderate, density-preserving perturbations in well locations will not substantially alter our conclusions. To the extent this is true, the 340 idealization of uniformly-spaced wells is harmless, and it provides a number of advantages for the study: (i) measurement spacing is represented by a single quantity, rather than a distribution, (ii) the number of truly distinct measurements equals the number of measurement locations, and (iii) the need to quantify and sample over an additional source of heterogeneity—in well configuration—is avoided.

The main MC study explored field reconstruction for ensembles of fields corresponding to five different values of Ψ, covering measurement spacings, Δ_m, ranging from dense to sparse relative to field correlation length. This study used a full complement of measurements, with values of speed, conductivity, and head at each measurement location, and also featured regularization. A second MC study was performed, using only the intermediate value $\Psi = 0.1$, which considered the impact of removing some information, with three ensembles neglecting one of regularization ($w_R = 0$), speed ($w_j = 0$), and conductivity ($w_l = 0$), respectively. A straightforward regularization-by-truncation scheme in which only leading (large eigenvalue) terms were calibrated and $w_R = 0$ was also considered.

Each ensemble employed its own distinct set of 50 unique "true" random K fields from which simulated measurements were collected, and calibration was run for each from at least two random initial parameter sets for each "true" random field, resulting in at least 100 automated calibrations per ensemble. The number of fields generated for each ensemble was verified to be adequate by comparing the results of the several ensembles and noting quantitatively and qualitatively similar results, despite each ensemble employing a unique set of fields (see, e.g., Section 4.3). Each calibration was performed by iterative evaluation of (4-5) and (9), over 500 steps (counted by runs of the forward model), using a straightforward adaptive step size algorithm which was found to work well: grow the step size by 10% after each step down gradient that successfully reduces the loss function, (3), and otherwise halve it until descent along the gradient reduces the loss function. 500 steps was found sufficient for gradient descent to achieve essentially all possible improvements in the calibrated K field (see Section 4.1), regardless of whether the objective function was locally minimized. For sparse measurements, this reduced computational cost relative to more common termination conditions based on convergence of the objective function. Each time the algorithm achieved a new minimum of the loss function, its value was stored, along with the parameter vector, $\mathbf{p}^i$, where i is the parameter iteration index. Iterations were counted only when there was a reduction in the loss function (once for each run of the adjoint model), so typically slightly fewer than 500 parameter iterations were stored due to occasional step size cuts that required additional forward model runs. The initial random parameter guess vector $\mathbf{p}^0$ was also stored for each calibration. The parameter vector corresponding to the smallest loss function was given the symbol $\mathbf{p}^{\text{final}}$, and used alongside $\mathbf{p}^0$ to quantify the improvement in the coefficient of each K-L basis function as a result of the automatic calibration. For performance reasons, fewer K-L functions were used in calibration than were used to generate the measurements, with the coefficients for the least significant (smallest eigenvalue) K-L functions fixed at zero. We denote the number of coefficients actually calibrated by ν. Typically, $\nu = 1350$ coefficients, representing 98.2% of variance, were employed, except in one trial, in which only $\nu = 12$ coefficients, representing 62.7% of variance were calibrated. The motivation for this selection depends on results from the main MC study, and is discussed in Section 4.

3.5 Scale and fitness metrics

To quantify the results of the parametric study, it is helpful to define some quantities relating to feature scale and to the accuracy of the reconstruction.

The eigenvalues, λ_n, are proxies for the significance of the various basis functions in the K-L expansion, but are difficult to interpret physically. For analysis, we define *normalized basis function wavelength* as the harmonic mean of the x- and

y-direction wavelengths, normalized to the measurement spacing, Δ_m:

$$T_n = \frac{1}{\Delta_m} \frac{2}{\frac{w_{x,n}}{2\pi} + \frac{w_{y,n}}{2\pi}}. \tag{42}$$

380   This quantity is a measure of the feature scale associated with K-L basis function $\phi_n(x,y)$, and was found to vary nearly monotonically, though non-linearly with eigenvalue. Precisely, if $T_2 > T_1$, then $\lambda_2 + \epsilon > \lambda_1$, for some small ϵ.

     To quantify the reconstruction error, we will generally employ the L^2 (root-mean square) norm, $||\cdot||_2$, to quantify the distance between the true log-conductivity field, $\ln \tilde{K}$, and the calibrated log-conductivity field, $\ln K$:

$$\left\| \ln \tilde{K} - \ln K \right\|_2 \equiv \sqrt{\int_0^{L_x} \int_0^{L_y} \left(\ln \tilde{K}(x,y) - \ln K(x,y) \right)^2 dy\, dx} \tag{43}$$

385

$$= \sqrt{\sum_n \lambda_i (\zeta_n - p_n)^2}, \tag{44}$$

where the last equality follows from Parseval's theorem. Implications of the theorem are that we can work directly with the coefficient vectors ζ and $\mathbf{p}$, and that we can thereby quantify the proportion of the error in the subspaces defined by a given K-L basis function or collection of basis functions. As we are especially interested in the reconstruction of features with large scales, which are captured by basis functions with large T (equivalently large eigenvalues, which have low indexes, n), we

390 define

$$L_N^2 \equiv \sqrt{\sum_{n=1}^N \lambda_i (\zeta_n - p_n^{\mathrm{final}})^2}, \tag{45}$$

where p_n^{final} is the n-th component of $\mathbf{p}^{\mathrm{final}}$. Where all components are employed ($N = 2000$), we sometimes drop the subscript, N, for clarity. We also define a normalized form of this quantity

$$\hat{L}_N^2 \equiv \frac{L_N^2}{\frac{\sum_{n=1}^N \lambda_n}{\sqrt{L_x L_y \sigma_{\ln K}^2}}}, \tag{46}$$

395 which has the interpretation of root-mean-square error divided by the proportion of the L^2 norm of the entire log-conductivity field, i.e., $\|\ln K\|_2$ included in the considered subspace. We are also interested in the amount of learning that occurs in the subspace defined by each K-L basis function over the course of calibration. We define

$$r_n \equiv \left\langle \frac{(\zeta_n - p_n^{\mathrm{final}})^2}{(\zeta_n, 0 - p_n^0)^2} \right\rangle_G, \tag{47}$$

where p_n^0 is the n-th component of initial coefficient vector $\mathbf{p}^0$, and $\langle \cdot \rangle_G$ represents the geometric mean over all the calibration

400 trials in an ensemble. If $r_n = 1$, measurements on average provided no information about the coefficient for K-L basis function $f_n(x,y)$, and if $r_n = 0$, then perfect identification occurs.

4 Results and discussion

4.1 Reduction in global L^2 error by calibration

Because the loss function is based on point and / or proxy measurements for the true fields of interest, the relationship between reductions in the loss and improved accuracy of the underlying $\ln \tilde{K}$ field is only indirect. Head fields are generally smooth relative to their underlying conductivity fields, observations are sparse, and conductivities along a streamline may be permuted without altering the effective conductivity along that streamline: permutation is detectable only indirectly through the deformation of neighboring streamlines. Thus, it is not *a priori* obvious how well calibration of point measurements (i.e., reducing loss function J) will succeed in constraining conductivity values remote from the measurement locations. To examine this, we consider the global L^2 error of the entire reconstructed conductivity field, $\ln K$, relative to the known field, $\ln \tilde{K}$ from which the measurements are generated. This metric weights accuracy at all locations in the domain equally, so if reduction in the loss function truly results in learning about the underlying field, reduction in J should correspond to reduction in L^2. In Figure 2, we present results comparing the global error, L^2 with J for each iteration in each automated calibration for each of five ensembles in the main MC study featuring different measurement densities from $\Psi = 0.02$ to 1.56. In general, a steep decline was observed in L^2 versus J for early iterations, which plateaued in later iterations, as further reduction in the objective function ceased to cause significant improvements in model fidelity. The ultimate level of L^2 error at the plateau was found to increase with Ψ, as the average distance to the nearest measurement point increased and the measurement set contained less relevant information to constrain the conductivity and head fields there.

4.2 Feature scale and information gain due to calibration

A second illuminating analysis that can be performed on the main MC study is to consider the average improvement each individual K-L coefficient, indexed by T, the basis function wavelength normalized to the measurement spacing. This allows us to determine the ability to identify features of a particular scale by calibration, and to determine a cutoff beneath which small-scale variability is essentially invisible to the calibration process. In Figure 3, the average ratio of squared coefficient error after vs. before calibration is shown at a function of T for each calibrated K-L basis function, for all five ensembles, each for a different value of Ψ. For small T, we observe a relatively flat (relative to T) regime in which only minor improvement is observed. We attribute the improvement to the w_R regularization term, for reasons discussed below. Around $T = 4$, we observe a sharp regime change, in which r begins to rapidly decrease in power law fashion with increasing T, we identify the following empirical relation for this regime,

$$r_n \approx 10\, T_n^{-2.5}, \qquad T_n > 4, \tag{48}$$

shown as the dotted line in Figure 3. A consequence of this analysis is that positive identification becomes possible only once feature scale exceeds four times the measurement spacing. We note good coherence between the various ensembles, suggesting that T is an adequate metric for capturing relative basis function scale.
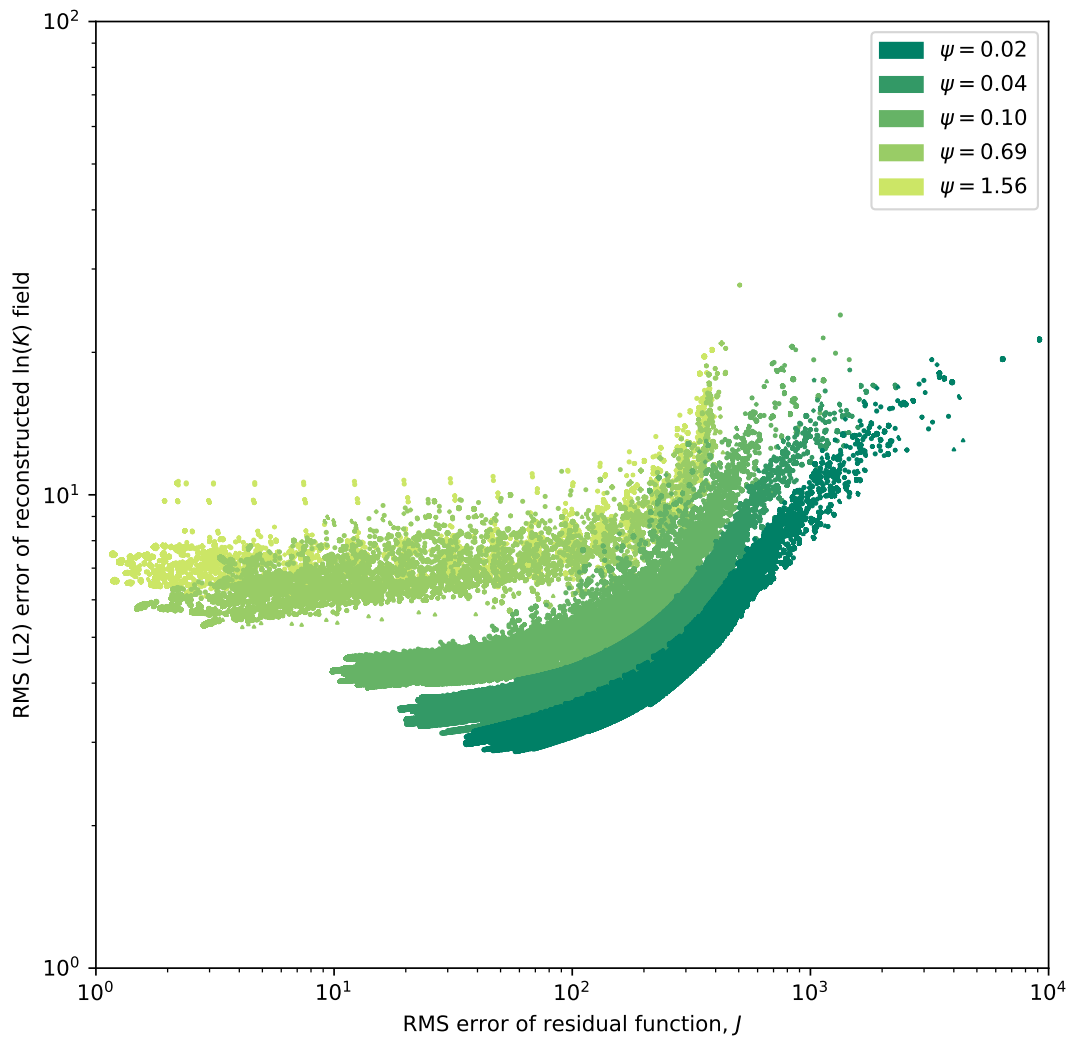
Figure 2. Comparison of the L^2 error of the reconstructed hydraulic conductivity field versus the residual function, J, defined in (3) for several different measurement densities, Ψ, in the main MC study. Weights are as defined in (37-40). Each ensemble is plotted with a unique color, and within each ensemble, each parameter iteration for each calibration trial (i.e., unique pairing of ζ and $\mathbf{p}^0$) is shown.
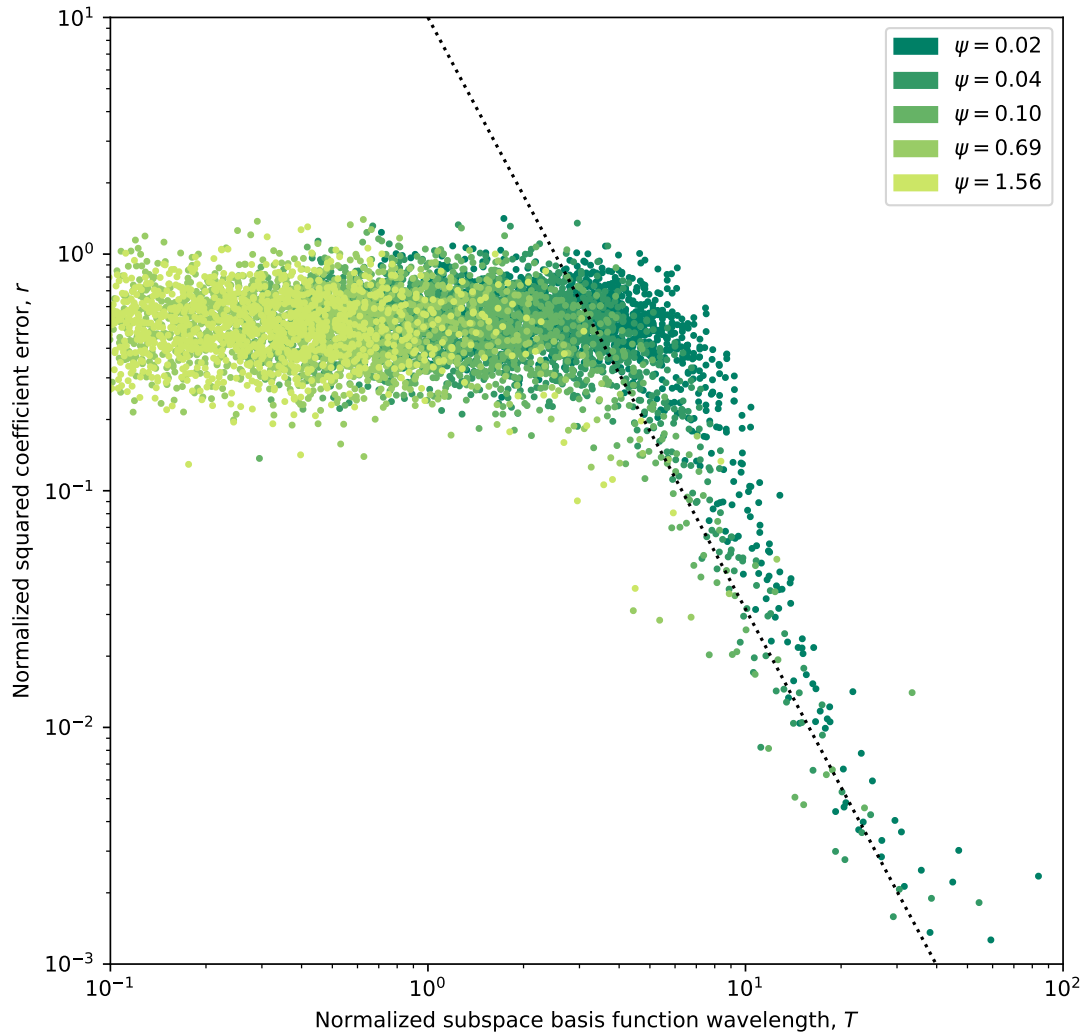
Figure 3. Geometric mean squared relative coefficient error improvement, r for each K-L basis function, $f_n(x)$, across all calibration trials (i.e., unique pairing of ζ and $\mathbf{p}^0$) in each ensemble (corresponding to a different value of ψ) in the main MC study. Each basis function is represented by its normalized wavelength, T, rather than its index. A dashed trend line scaling as $T^{-2.5}$ is shown in the large-wavelength region.

4.3 Information gain due to velocity measurements and regularization

In the second MC study, we fixed $\Psi = 0.10$ but modified the objective function in various ways to ascertain how the modifications affect the L^2 accuracy of the ultimate reconstruction (L^2_{2000}), and also the ability to calibrate large-scale features in the 12-dimensional subspace spanned by the K-L basis functions satisfying the previously identified criterion $T > 4$ (L^2_{12}). The specifics of the studies and the resulting metrics are tabulated in Table 1. A visual representation of the amount of learning that occurs about the coefficients of each K-L basis function, sorted by T, analogous to the one for the main study is shown in Figure 4.

In this study, the base case (featuring all available information) was shared with the main MC study. In this data set, calibration was performed beginning from a pair of different random $\mathbf{p}^0$ vectors for each true field vector, ζ. Visual examination of the calibrated head fields for this data set showed that the fitted $\ln K$ and h fields were in all cases visually identical for both of the paired initial parameter guesses. This suggests that regularized objective function was convex, and that the unique solution is essentially found in all cases, with no influence of initial conditions. However, even when the unique regularized solution is achieved, the remaining relative error, $\hat{L}^2_{2000}$, is still 42%, and 12% in the subspace spanned by the most significant (largest wavelength) basis functions. Figure 1 (a) visually illustrates the results: note how large-scale features are captured, but there is still significant infidelity between the smooth, regularized $\ln K$ and the true $\ln \tilde{K}$. By calibration, L2 error is reduced to 32% of its value in the initial guess.

Where regularization is disabled, pairwise visual comparison reveals some residual effect of initial guess vector, although large-scale features of the true fields are plainly always captured. Unregularized calibrated L^2 error was only slightly worse than for regularized calibration in the most significant subspace (again see Table 1), attributable to the fact that the data is highly informative about these K-L coefficients, and there is limited work for regularization to accomplish. For the unregularized ensemble, we attribute residual L^2 error to errors in high-wavenumber subspaces from the initial guess $\mathbf{p}^0$ that manifest at locations remote from the measurements and which are left over after large-scale features have been identified. This is an entirely different cause of error than the smoothing which generates unique solutions in the w_R-regularized case. Figure 1 (b) illustrates unregularized calibration using the same target field and measurement data as used in (a). By calibration, L2 error is reduced to 34% of its value in the initial guess, and the final absolute L2 error for (b) is 24% higher than for (a).

Without regularization from w_R, $r \approx 1$ in the $T < 4$ subspace corresponding to small-scale features, indicating that the measurements provide no information about these coefficients. We attribute the improved performance with the regularization term to the fact that the coefficients of ζ and $\mathbf{p}^0$ are all independent, identically distributed random variables $\sim \mathcal{N}(0,1)$ it is easy to show that $\mathrm{Var}\,(\zeta_n - p_n^0) > \mathrm{Var}\,(\zeta_n) = 1$, so by pulling each p_n towards zero, the w_R term in J decreases average square error of p_n, even absent any information from the measurements regarding its value.

A previous study (Tso et al., 2016), performed using a different co-kriging-based calibration approach suggested a significant gain in reconstruction reliability over use of head-conductivity co-kriging. We tested whether groundwater speed would provide similar extra information in the context of gradient descent model calibration by performing a calibration ensemble in which its weight was zero. We found performance essentially identical to that for the base case. Given that head and conductivity

fields jointly determine the flux everywhere, and regularization is sufficient to determine a unique solution, the fact that speed information was found to be redundant implies that the regularized solution correctly identifies $\tilde{h}$ and $\tilde{K}$ at the measurement locations.

470   Given the idea that a correct speed and head field could uniquely determine $\tilde{K}$ everywhere, we examined the extent to which speed measurements could substitute for hard-to-determine point measurements of $\tilde{K}$ by eliminating the weight on these measurements in J. This was found to be possible to some extent, but performance was significantly reduced relative to all other approaches, with identification in the most significant, large wavelength subspaces, as measured by L_{12}^2, significantly degraded. A possible explanation for this result is that calibration of $\tilde{K}$ becomes sensitive to errors in ∇h as well as in h.

475   Finally, motivated by the idea that the large-wavelength subspace ($T > 4$) is readily identifiable without regularization, and that the measurements contain little or no information about the small-T (small-scale feature) subspace, we attempted a naive truncation-based regularization approach in which only the most significant, $T > 4$ K-L coefficients are calibrated, and the rest are fixed to 0, and no regularization term is applied (i.e., $\nu = 12$ and $w_R = 0$ for this ensemble). Performance of this approach was not as strong as the other trials, with $\hat{L}_{12}^2 = 0.21$, about 50% greater than in the base case, but identification
480 of the large-scale subspace remained respectable and still exceeded the regularized trial with no $\ln \tilde{K}$ measurements. When the K-L field representation is employed, $\partial K / \partial p_n$ does not collapse to unity, so there remains an expensive basis-function-specific integration in (9). Thus, the computational cost of computing the gradient using naive truncation was reduced more than 100-fold in this ensemble relative to the base case and this approach merits further consideration.

Table 1. Descriptive statistics of five ensemble calibration experiments in the second MC study, corresponding to $\Psi = 0.1$

	w_j	w_k	w_l	w_r	ν	L_{12}^2	$\hat{L}_{12}^2$	L_{2000}^2	$\hat{L}_{2000}^2$
Base case	2	225	0.25	0.25	1350	1.09	0.14	4.25	0.42
No regularization	2	225	0.25	0.0	1350	1.19	0.15	4.93	0.49
No $\tilde{q}$ information	0	225	0.25	0.25	1350	1.14	0.14	4.29	0.43
No $\ln \tilde{K}$ information	2	225	0	0.25	1350	3.58	0.45	6.04	0.60
Truncated	2	225	0.25	0.0	12	1.68	0.21	6.19	0.62

4.4   Can we expect these results to generalize?

485 Both the identification cutoff of four times measurement scale and the observation of power law improvement of identifiability with increased feature scale are empirical results of an MC study. As with any empirical investigation, it is natural to wonder how general the results are, or whether they are dependent in some way on the specific setting chosen for the study. In our investigations, these relations were found to be robust to changed point measurement resolution, and also to zero weighting of certain types of measurements and different regularization schemes, suggesting robustness to alteration of the objective
490 function.

Regarding the specifics of correlation structure and average gradient, it is helpful to recall that the point measurements functionally constrain a low-order 2D Taylor series expansion of the head near each measurement location, and that an accurate
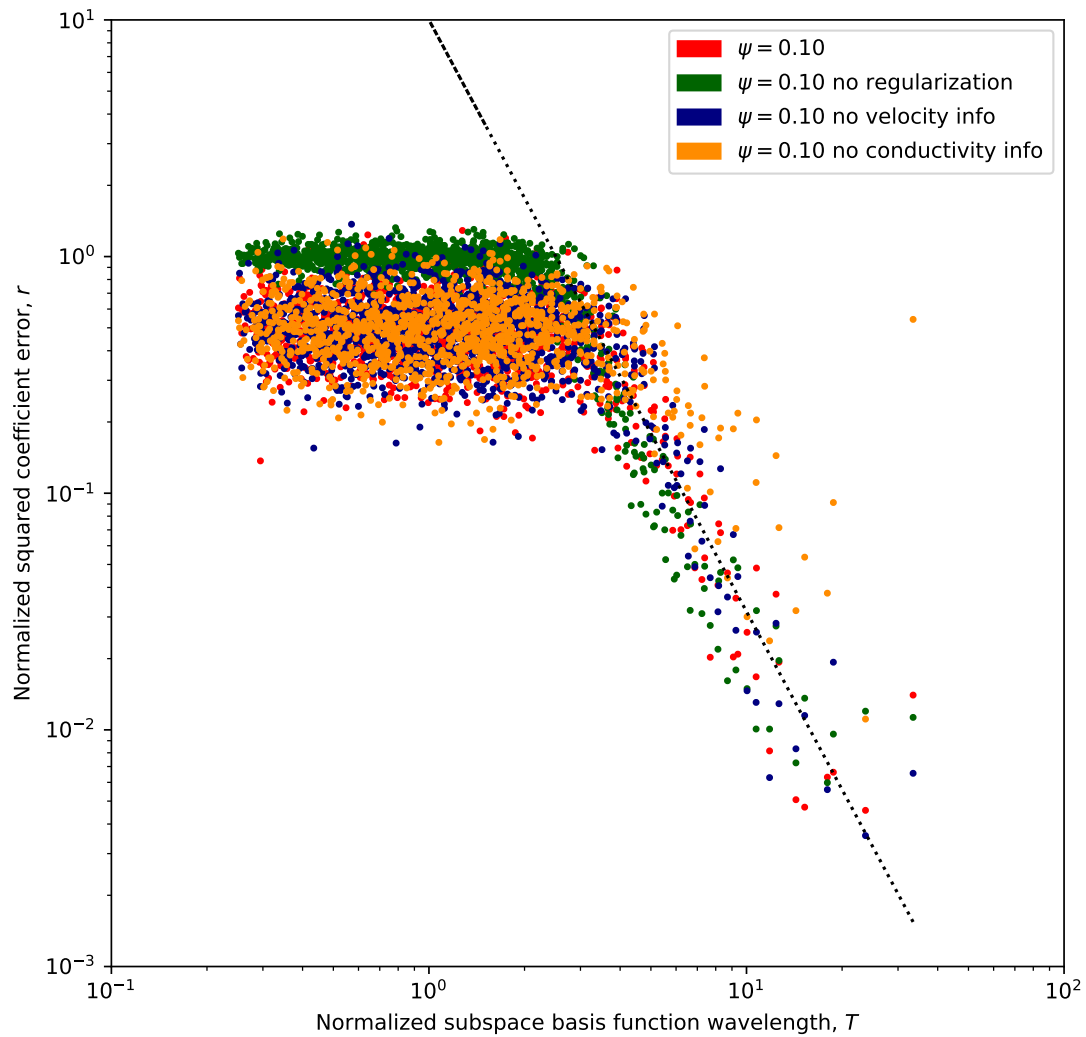
Figure 4. Geometric mean squared relative coefficient error improvement, r, across all calibration trials in each ensemble in the second MC study. Results are presented as a function of normalized K-L basis function wavelength, T, with each ensemble representing different terms present in the loss function. A dashed trend line scaling as $T^{-2.5}$ is shown in the large wavelength region.

head field can be uniquely inverted to recover the underlying conductivity field. It is thus reasonable to attribute the weaker identification of smaller scale features to their proportionally larger impact on higher-order terms of the head field, which do not impact the observed quantities. The decreased visibility of smaller-scale features due to non-unique impact on observables on this analysis is a direct result of feature scale normalized to measurement scale (i.e., T). Consequently, we do not expect the details of correlation structure or large-scale flow regime to impact our results.

By contrast, it is well-documented that behavior of physical processes such as Darcy flow may be affected by system dimension (e.g., Chiogna et al., 2014). So while we follow the example of much previous research in focusing on reconstruction of 2D conductivity (equivalently, transmissivity) fields, a valuable follow-up would consider identifiability of full 3D conductivity fields.

5   Summary and conclusions

We computed the sensitivity of a loss function based on three types of point hydraulic measurement and a general regularization term to an underlying, heterogeneous hydraulic conductivity field via a continuous adjoint-state approach. Using K-L expansion, we created a general, zonation-free technique for analyzing the quality of inversion from point data to an underlying heterogeneous scalar field. Using K-L basis functions which feature clearly-defined wavelengths, we analyzed how reliably conductivity and head fields were recovered, both as a function of feature scale and globally, given different densities and types of point measurements.

The following are our key conclusions:

➢ Substantially better than random identification accuracy was observed once K field feature scales exceed four times the measurement spacing, with relative squared error after calibration scaling as $T^{-2.5}$, where T is the feature scale normalized to measurement spacing. This relation provides a useful guideline for needed measurement density to achieve desired field reconstruction accuracy. For higher resolution characterization, non-point geophysical or tracer data may be required.

➢ Even at very coarse measurement resolutions relative to K field correlation length, the point-wise loss function was found informative globally (in an L^2 sense) about the underlying conductivity field, and calibration offers better than random feature identification at all scales when regularization is employed.

➢ Groundwater speed point measurements were found to partially substitute for point conductivity measurements, but contra the sequential linear estimation study of Tso et al. (2016), provide little additional information in the context of regularized gradient descent optimization of point prediction errors, provided reliable head and conductivity point measurements already exist.

➢ Gradient descent calibration ultimate L^2 error was found to be dependent on measurement density and displays quickly diminishing returns on optimization. We posit that large-scale features are identified quickly and the loss function is less sensitive to point mismatch remote from the measurement locations. Thus, later iterations tend to optimize field values

near the well: K-field infidelity remote from the measurement locations that exists by the time large-scale features have been identified tends to persist.

The approach adopted here has a number of useful features that may be helpful for further investigations. The analytical sensitivity expression for a loss function containing many different types of data and regularization can be used for model fitting, and its zonated form (8) is particularly efficient for use with finite volume and finite element codes. Also, the sort of 530 zonation-free K-L series Monte Carlo calibration study could be used to explore the identifiability of other geophysical fields, beyond the hydraulic conductivity/transmissivity fields considered in this study. Naturally, as the relations uncovered herein are empirically based, it is possible they may depend in some way on the specific model chosen for the study. Exploration in 3D may in particular be profitable.

Appendix A: Adjoint sensitivity equation derivation

535 The sensitivity of J (3) to any given model parameter, p_i, can be computed directly:

$$
\begin{aligned}
\frac{\partial J}{\partial p_i} = & \sum_j 2w_j \left(K(\mathbf{x}_j) \|\nabla h(\mathbf{x}_j)\| - \tilde{q}(\mathbf{x}_j) \right) \left[\frac{\partial K}{\partial p_i} \|\nabla h(\mathbf{x}_j)\| + K \frac{\nabla h \cdot \nabla \phi_i}{\|\nabla h\|} \right] \\
& + \sum_k 2w_k \phi_i(\mathbf{x}_k) \left(h(\mathbf{x}_k) - \tilde{h}(\mathbf{x}_k) \right) \\
& + \sum_l 2w_l \frac{1}{K(\mathbf{x}_l)} \frac{\partial K}{\partial p_i} \left(\ln K(\mathbf{x}_l) - \ln \tilde{K}(\mathbf{x}_l) \right) \\
& + \frac{\partial R}{\partial p_i},
\end{aligned}
\tag{A1}
$$

where we define the sensitivity $\phi_i \equiv \partial h / \partial p_i$. Employing vector form and rewriting in terms of integrals over the volume Ω:

$$
\begin{aligned}
\frac{\partial J}{\partial \mathbf{p}} = & \int_\Omega 2 \left(K \|\nabla h\| - \tilde{q} \right) \frac{\partial K}{\partial \mathbf{p}} \|\nabla h\| \sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j) \, dV \\
& + \int_\Omega 2 \left(K \|\nabla h\| - \tilde{q} \right) K \frac{\nabla h \cdot \nabla \phi}{\|\nabla h\|} \sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j) \, dV \\
& + \int_\Omega 2\phi \left(h - \tilde{h} \right) \sum_k w_k \delta(\mathbf{x} - \mathbf{x}_k) \, dV \\
& + \int_\Omega 2 \frac{1}{K} \frac{\partial K}{\partial \mathbf{p}} \left(\ln K - \ln \tilde{K} \right) \sum_l w_l \delta(\mathbf{x} - \mathbf{x}_l) \, dV \\
& + \frac{\partial R}{\partial \mathbf{p}}.
\end{aligned}
\tag{A2}
$$

Simplifying the second integral by employing integration by parts to remove the dependence on $\nabla\phi$ and then employing (1) yields

$$
\frac{\partial J}{\partial \mathbf{p}} = \int_\Omega 2\left(K\|\nabla h\| - \tilde{q}\right)\frac{\partial K}{\partial \mathbf{p}}\|\nabla h\|\sum_j w_j \delta(\mathbf{x}-\mathbf{x}_j)\, dV
$$

$$
- \int_\Omega 2\phi K \nabla h \cdot \nabla\left[\left(K - \frac{\tilde{q}}{\|\nabla h\|}\right)\sum_j w_j \delta(\mathbf{x}-\mathbf{x}_j)\right]\, dV
$$

$$
+ \int_\Omega 2\phi\left(h - \tilde{h}\right)\sum_k w_k \delta(\mathbf{x}-\mathbf{x}_k)\, dV
$$

$$
+ \int_\Omega 2\frac{1}{K}\frac{\partial K}{\partial \mathbf{p}}\left(\ln K - \ln \tilde{K}\right)\sum_l w_l \delta(\mathbf{x}-\mathbf{x}_l)\, dV
$$

$$
+ \frac{\partial R}{\partial \mathbf{p}}. \tag{A3}
$$

.

A1 Weak-form equation for sensitivity

We may differentiate equations (1-2) with respect to $\mathbf{p}$ so as to generate a BVP for the sensitivity vector, ϕ:

$$
\nabla \cdot \left(\frac{\partial K}{\partial \mathbf{p}}\nabla h\right) + \nabla \cdot (K\nabla\phi) = 0 \text{ in } \Omega, \tag{A4}
$$

$$
\phi = \frac{\partial f}{\partial \mathbf{p}} \text{ on } \delta\Omega. \tag{A5}
$$

Sensitivity equation (A4) can be placed in weak form by expressing it as an integral with respect to ϕ^*, an arbitrary, smooth test function known as the *adjoint*:

$$
\int_\Omega \left[\nabla \cdot \left(\frac{\partial K}{\partial \mathbf{p}}\nabla h\right)\right]\phi^*\, dV + \int_\Omega \left[\nabla \cdot (K\nabla\phi)\right]\phi^*\, dV = 0. \tag{A6}
$$

We then integrate the first integral by parts once and the second by parts twice, to arrive at the five-term expression

$$
0 = \int_{\delta\Omega}\frac{\partial K}{\partial \mathbf{p}}\nabla h\phi^* \cdot dS - \int_\Omega \frac{\partial K}{\partial \mathbf{p}}\nabla h \cdot \nabla\phi^*\, dV
$$

$$
+ \int_{\delta\Omega} K\nabla\phi\phi^* \cdot dS - \int_{\delta\Omega} K\phi\nabla\phi^* \cdot dS + \int_\Omega \phi\nabla\cdot(K\nabla\phi^*)\, dV. \tag{A7}
$$

25

A2 Removing derivatives of ϕ and simplifying

Our goal is to derive an equation for $\frac{\partial J}{\partial \mathbf{p}}$ that does not explicitly depend on ϕ. To this end, we add (A3) and (A7) to generate

$$
\begin{aligned}
\frac{\partial J}{\partial \mathbf{p}} =& \int_\Omega 2\left(K\,\|\nabla h\| - \tilde{q}\right)\frac{\partial K}{\partial \mathbf{p}}\,\|\nabla h\|\sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j)\,dV \\
& - \int_\Omega 2\phi K\nabla h \cdot \nabla\left[\left(K - \frac{\tilde{q}}{\|\nabla h\|}\right)\sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j)\right]\,dV \\
& + \int_\Omega 2\phi\left(h - \tilde{h}\right)\sum_k w_k \delta(\mathbf{x} - \mathbf{x}_k)\,dV \\
& + \int_\Omega \frac{2}{K}\frac{\partial K}{\partial \mathbf{p}}\left(\ln K - \ln \tilde{K}\right)\sum_l w_l \delta(\mathbf{x} - \mathbf{x}_l)\,dV + \frac{\partial R}{\partial \mathbf{p}} \\
& + \int_{\delta\Omega} \frac{\partial K}{\partial \mathbf{p}}\nabla h \phi^* \cdot dS - \int_\Omega \frac{\partial K}{\partial \mathbf{p}}\nabla h \cdot \nabla \phi^*\,dV \\
& + \int_{\delta\Omega} K\nabla\phi\phi^* \cdot dS - \int_{\delta\Omega} K\phi\nabla\phi^* \cdot dS + \int_\Omega \phi\nabla\cdot(K\nabla\phi^*)\,dV.
\end{aligned} \tag{A8}
$$

We ultimately want to remove the dependence on ϕ, so we combine the volume integrals whose integrands are proportional to it:

$$
\begin{aligned}
\frac{\partial J}{\partial \mathbf{p}} =& \int_\Omega \phi\left[\nabla\cdot(K\nabla\phi^*) - 2K\nabla h \cdot \nabla\left(\left(K - \frac{\tilde{q}}{\|\nabla h\|}\right)\sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j)\right) + 2\left(h - \tilde{h}\right)\sum_k w_k \delta(\mathbf{x} - \mathbf{x}_k)\right]\,dV \\
& + \int_\Omega 2\left(K\,\|\nabla h\| - \tilde{q}\right)\frac{\partial K}{\partial \mathbf{p}}\,\|\nabla h\|\sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j)\,dV \\
& + \int_\Omega \frac{2}{K}\frac{\partial K}{\partial \mathbf{p}}\left(\ln K - \ln \tilde{K}\right)\sum_l w_l \delta(\mathbf{x} - \mathbf{x}_l)\,dV + \frac{\partial R}{\partial \mathbf{p}} \\
& - \int_\Omega \frac{\partial K}{\partial \mathbf{p}}\nabla h \cdot \nabla\phi^*\,dV + \int_{\delta\Omega} \frac{\partial K}{\partial \mathbf{p}}\nabla h \phi^* \cdot dS + \int_{\delta\Omega} K\nabla\phi\phi^* \cdot dS - \int_{\delta\Omega} K\phi\nabla\phi^* \cdot dS.
\end{aligned} \tag{A9}
$$

The first integral (and its dependence on the ϕ) is removed if the expression in the square brackets is zero, implying that

$$
\int_\Omega \nabla\cdot(K\nabla\phi^*) - 2K\nabla h \cdot \nabla\left(\left(K - \frac{\tilde{q}}{\|\nabla h\|}\right)\sum_j w_j \delta(\mathbf{x} - \mathbf{x}_j)\right) + 2\left(h - \tilde{h}\right)\sum_k w_k \delta(\mathbf{x} - \mathbf{x}_k)\,dV = 0 \tag{A10}
$$

Applying integration by parts to the second term of the integrand and then applying (1) allows the simplification

$$
\int_\Omega \nabla\cdot(K\nabla\phi^*) + 2\left(h - \tilde{h}\right)\sum_k w_k \delta(\mathbf{x} - \mathbf{x}_k)\,dV = 0 \tag{A11}
$$

This motivates solution of an adjoint problem in ϕ^* to eliminate the dependence of (A9) on ϕ. See Section 2.2 for the result.

References

Bagtzoglou, A. and Atmadja, J.: Mathematical methods for hydrologic inversion: The case of pollution source identification, Water Pollution, 5, 65–96, https://doi.org/10.1007/b11442, 2005.

Cardiff, M., Barrash, W., Kitanidis, P. K., Malama, B., Revil, A., Straface, S., and Rizzo, E.: A potential-based inversion of unconfined steady-state hydraulic tomography, Ground Water, 47, 259–270, https://doi.org/10.1111/j.1745-6584.2008.00541.x, 2009.

Carrera, J. and Neuman, S. P.: Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 3. Application to Synthetic and Field Data, Water Resources Research, 22, 228–242, https://doi.org/10.1029/WR022i002p00199, 1986a.

Carrera, J. and Neuman, S. P.: Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 1. Maximum Likelihood Method Incorporating Prior Information, Water Resources Research, 22, 199–210, https://doi.org/10.1029/WR022i002p00199, 1986b.

Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., and Slooten, L. J.: Inverse problem in hydrogeology, Hydrogeology Journal, 13, 206–222, https://doi.org/10.1007/s10040-004-0404-7, 2005.

Chiogna, G., Rolle, M., Bellin, A., and Cirpka, O. A.: Helicity and flow topology in three-dimensional anisotropic porous media, Advances in Water Resources, 73, 134–143, https://doi.org/10.1016/j.advwatres.2014.06.017, 2014.

Cirpka, O. A. and Kitanidis, P. K.: Sensitivity of temporal moments calculated by the adjoint-state method and joint inversing of head and tracer data, Advances in Water Resources, 24, 89–103, www.elsevier.com/locate/advwatres, 2001.

Delay, F., Badri, H., Fahs, M., and Ackerer, P.: A comparison of discrete versus continuous adjoint states to invert groundwater flow in heterogeneous dual porosity systems, Advances in Water Resources, 110, 1–18, https://doi.org/10.1016/j.advwatres.2017.09.022, 2017.

Delay, F., Badri, H., and Ackerer, P.: Heterogeneous hydraulic conductivity and porosity fields reconstruction through steady-state flow and transient solute transport data using the continuous adjoint state, Advances in Water Resources, 127, 148–166, https://doi.org/10.1016/j.advwatres.2019.03.014, 2019.

Elizondo, D., Cappelaere, B., and Faure, C.: Automatic versus manual model differentiation to compute sensitivities and solve non-linear inverse problems, Computers and Geosciences, 28, 309–326, http://www.inria.fr/safir/SAM/, 2002.

Gottlieb, J. and Dietrich, P.: Identification of the permeability distribution in soil by hydraulic tomography, Inverse Problems, 11, 353–360, https://doi.org/10.1088/0266-5611/11/2/005, 1995.

Hansen, S., Vesselinov, V., Reimus, P., and Lu, Z.: Inferring subsurface heterogeneity from push-drift tracer tests, Water Resources Research, 53, https://doi.org/10.1002/2017WR020852, 2017.

Hayek, M., RamaRao, B. S., and Lavenue, M.: An Adjoint Sensitivity Model for Steady-State Sequentially Coupled Radionuclide Transport in Porous Media, Water Resources Research, 55, 8800–8820, https://doi.org/10.1029/2019WR025686, 2019.

Hendricks Franssen, H. J., Alcolea, A., Riva, M., Bakr, M., van der Wiel, N., Stauffer, F., and Guadagnini, A.: A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments, Advances in Water Resources, 32, 851–872, https://doi.org/10.1016/j.advwatres.2009.02.011, 2009.

Hoeksema, R. J. and Kitanidis, P. K.: An Application of the Geostatistical Approach to the Inverse Problem in Two-Dimensional Groundwater Modeling, Water Resources Research, 20, 1003–1020, 1984.

Illman, W. A., Liu, X., and Craig, A.: Steady-state hydraulic tomography in a laboratory aquifer with deterministic hetero-geneity: Multi-method and multiscale validation of hydraulic conductivity tomograms, Journal of Hydrology, 341, 222–234, https://doi.org/10.1016/j.jhydrol.2007.05.011, 2007.

610   Illman, W. A., Zhu, J., Craig, A. J., and Yin, D.: Comparison of aquifer characterization approaches through steady state groundwater model validation: A controlled laboratory sandbox study, Water Resources Research, 46, 1–18, https://doi.org/10.1029/2009WR007745, 2010.

Irving, J. and Singha, K.: Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, Water Resources Research, 46, https://doi.org/10.1029/2009WR008340, 2010.

Kitanidis, P. K.: Parameter Uncertainty in Estimation of Spatial Functions: Bayesian Analysis, Water Resources Research, 22, 499–507,
615   https://doi.org/10.1029/WR022i004p00499, 1986.

Kitanidis, P. K.: Quasi-linear geostatistical theory for inversing, Water Resources Research, 31, 2411–2419, 1995.

Kitanidis, P. K. and Lee, J.: Principal Component Geostatistical Approach for large- dimensional inverse problems, Water Resources Research, 50, 5428–5443, https://doi.org/10.1002/2013WR014979, 2014.

Kitanidis, P. K. and Vomvoris, E. G.: A Geostatistical Approach to the Inverse Problem in Groundwater Modeling (Steady State) and One-
620   Dimensional Simulations., Water Resources Research, 19, 677–690, 1983.

Logg, A., Mardal, K.-A., and Wells, G.: Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book, Springer, Berlin, https://doi.org/10.1007/978-3-642-23099-8, 2012.

Lu, Z. and Vesselinov, V.: Analytical sensitivity analysis of transient groundwater flow in a bounded model domain using the adjoint method, Water Resources Research, 51, 5060–5080, https://doi.org/10.1111/j.1752-1688.1969.tb04897.x, 2015.

625   Michalak, A. M. and Kitanidis, P. K.: Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, Water Resources Research, 40, 1–14, https://doi.org/10.1029/2004WR003214, 2004.

Neuman, S. P.: Calibration of Distributed Parameter Groundwater Flow Models Viewed as a Multiple-Objective Decision Process under Uncertainty, Water Resources Research, 9, 1006–1021, 1973.

Sambridge, M., Rickwood, P., Rawlinson, N., and Sommacal, S.: Automatic differentiation in geophysical inverse problems, Geophysical
630   Journal International, 170, 1–8, https://doi.org/10.1111/j.1365-246X.2007.03400.x, 2007.

Slater, L.: Near surface electrical characterization of hydraulic conductivity: From petrophysical properties to aquifer geometries - A review, Surveys in Geophysics, 28, 169–197, https://doi.org/10.1007/s10712-007-9022-y, 2007.

Sykes, J. F., Wilson, J. L., and Andrews, R. W.: Sensitivity Analysis for Steady State Groundwater Flow Using Adjoint Operators, Water Resources Research, 21, 359–371, 1985.

635   Szabó, N. P.: Hydraulic conductivity explored by factor analysis of borehole geophysical data, Hydrogeology Journal, 23, 869–882, https://doi.org/10.1007/s10040-015-1235-4, 2015.

Tonkin, M. J. and Doherty, J.: A hybrid regularized inversion methodology for highly parameterized environmental models, Water Resources Research, 41, 1–16, https://doi.org/10.1029/2005WR003995, 2005.

Tso, C., Zha, Y., Yeh, T., and Wen, J.: The relative importance of head, flux, and prior information in hydraulic tomography analysis, Water
640   Resources Research, 52, 3–20, https://doi.org/10.1111/j.1752-1688.1969.tb04897.x, 2016.

Wagner, B. J.: Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling, Journal of Hydrology, 135, 275–303, 1992.

Wang, N., Chang, H., and Zhang, D.: Deep-Learning-Based Inverse Modeling Approaches: A Subsurface Flow Example, Journal of Geophysical Research: Solid Earth, 126, 1–36, https://doi.org/10.1029/2020JB020549, 2021.

645   Woodbury, A. D. and Ulrych, T. J.: A full-Bayesian approach to the groundwater inverse problem for steady state flow, Water Resources Research, 36, 2081–2093, https://doi.org/10.1029/2000WR900086, 2000.

Woodbury, A. D., Smith, L., and Dunbar, W. S.: Simultaneous inversion of hydrogeologic and thermal data: 1. Theory and application using hydraulic head data, Water Resources Research, 23, 1586–1606, https://doi.org/10.1029/WR023i008p01586, 1987.

Wu, H., Greer, S. Y., and O'Malley, D.: Physics-embedded inverse analysis with algorithmic differentiation for the earth's subsurface, Scientific Reports, 13, https://doi.org/10.1038/s41598-022-26898-1, 2023.

Xu, T. and Gómez-Hernández, J. J.: Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble Kalman filter, Advances in Water Resources, 112, 106–123, https://doi.org/10.1016/j.advwatres.2017.12.011, 2018.

Yeh, T. C. J. and Liu, S.: Hydraulic tomography: Development of a new aquifer test method, Water Resources Research, 36, 2095–2105, https://doi.org/10.1029/2000WR900114, 2000.

Yeh, T.-C. J., Gutjahr, A. L., and Jin, M.: An Iterative Cokriging-Like Technique for Ground-Water Flow Modeling, Ground Water, 33, 33–41, 1995.

Yeh, W. W. and Yoon, Y. S.: Aquifer Parameter Identification with Optimum Dimension in Parameterization, Water Resources Research, 17, 664–672, 1981.

Yoon, Y. S. and Yeh, W. G.: Parameter Identification in an Inhomogeneous Medium With the Finite-Element Method., Soc Pet Eng AIME J, 16, 217–226, https://doi.org/10.2118/5524-pa, 1976.

Zhang, D. and Lu, Z.: An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loeve and polynomial expansions, Journal of Computational Physics, 194, 773–794, https://doi.org/10.1016/j.jcp.2003.09.015, 2004.

Zhou, H., Gómez-Hernández, J. J., and Li, L.: Inverse methods in hydrogeology: Evolution and recent trends, Advances in Water Resources, 63, 22–37, https://doi.org/10.1016/j.advwatres.2013.10.014, 2014.

Zhu, J. and Yeh, T. C. J.: Characterization of aquifer heterogeneity using transient hydraulic tomography, Water Resources Research, 41, 1–10, https://doi.org/10.1029/2004WR003790, 2005.

Zimmerman, D. A., De Marsily, G., Gotway, C. A., Marietta, M. G., Axness, C. L., Beauheim, R. L., Bras, R. L., Carrera, J., Dagan, G., Davies, P. B., Gallegos, D. P., Galli, A., Gómez-Hernández, J., Grindrod, P., Gutjahr, A. L., Kitanidis, P. K., Lavenue, A. M., McLaughlin, D., Neuman, S. P., RamaRao, B. S., Ravenne, C., and Rubin, Y.: A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, Water Resources Research, 34, 1373–1413, https://doi.org/10.1029/98WR00003, 1998.