# Ensemble Kalman inversion of induced polarization data

Chak-Hau Michael Tso [1,3,4] Marco Iglesias [2] and Andrew Binley[3,4]

[1]*Environmental Data Science Group, UK Centre for Ecology and Hydrology, Lancaster LA*1 4*AP, UK*
[2]*School of Mathematical Sciences, University of Nottingham, Nottingham NG*7 2*RD, UK. E-mail:* marco.iglesias@nottingham.ac.uk
[3]*Lancaster Environment Centre, Lancaster University, Lancaster LA*1 4*YQ, UK*
[4]*Centre for Excellence in Environmental Data Science, Lancaster University and UKCEH, Lancaster LA*1 4*YQ, UK*

## SUMMARY

This paper explores the applicability of ensemble Kalman inversion (EKI) with level-set parametrization for solving geophysical inverse problems. In particular, we focus on its extension to induced polarization (IP) data with uncertainty quantification. IP data may provide rich information on characteristics of geological materials due to its sensitivity to characteristics of the pore–grain interface. In many IP studies, different geological units are juxtaposed and the goal is to delineate these units and obtain estimates of unit properties with uncertainty bounds. Conventional inversion of IP data does not resolve well sharp interfaces and tends to reduce and smooth resistivity variations, while not readily providing uncertainty estimates. Recently, it has been shown for DC resistivity that EKI is an efficient solver for inverse problems which provides uncertainty quantification, and its combination with level set parametrization can delineate arbitrary interfaces well. In this contribution, we demonstrate the extension of EKI to IP data using a sequential approach, where the mean field obtained from DC resistivity inversion is used as input for a separate phase angle inversion. We illustrate our workflow using a series of synthetic and field examples. Variations with uncertainty bounds in both DC resistivity and phase angles are recovered by EKI, which provides useful information for hydrogeological site characterization. Although phase angles are less well-resolved than DC resistivity, partly due to their smaller range and higher percentage data errors, it complements DC resistivity for site characterization. Overall, EKI with level set parametrization provides a practical approach forward for efficient hydrogeophysical imaging under uncertainty.

**Key words:** Induced polarization; Ensemble Kalman methods; Inversion; Data assimilation; Uncertainty quantification; Level sets.

## 1 INTRODUCTION

### 1.1 Motivation

The advances of hydrogeophysics have been motivated by a better quantitative understanding of the subsurface hydrological parameters inferred from geophysical data (Binley *et al.* 2015). A key motivation for hydrogeophysics is to leverage the high spatial coverage and imaging capabilities to provide insights to describe subsurface structure and process. However, conventional geophysical inversions naturally yield smooth images due to spatial regularization. The way the inverse problem is posed also causes the reporting of uncertainty estimates to be frequently ignored in geophysics. We seek to develop and demonstrate an efficient and flexible inversion method that is suitable for different geophysical data and provides some estimates of uncertainty. Such understanding of uncertainty is related to the analysis of information content (or data worth) (JafarGandomi

& Binley 2013) and is an emerging field of research in hydrogeophysical studies.

### 1.2 Geophysical inversion

Geophysical inversion is typically posed as an optimization problem: the goal is to update model parameters until the value of an objective function decrease to a certain pre-defined threshold. Typically, the model values are the geophysical properties at each point (or grid) of the model domain, while the objective function is the sum of data (inversely weighted by data uncertainty) and regularized model misfits. The relative strength of the two terms are also controlled by a scalar weight, which is usually determined by line search. The regularization is needed because this mesh-based inversion is underdetermined and ill-posed. A standard choice of regularization is to apply a roughness filter for neighbouring grid cells and this approach is commonly known as smoothness-constrained inversion (SCI). SCI is an extremely robust and efficient method

and thus is often seen as the default method for inversions of electrical geophysical data. However, this typical inversion approach has two main disadvantages. The first is that the standard choice for regularization often yields unrealistic overly smooth estimates of geophysical properties that cannot capture sharp discontinuities (e.g. from the presence of different geologic properties) that are commonly found in real geophysical settings. The second disadvantage of SCI is that its deterministic formulation does not provide measures of the uncertainty of the resultant model. Uncertainty in model estimates arises not only from the presence of measurement errors but also from the inherent non-uniqueness of the solution to the inverse problem.

Many approaches have been developed to circumvent the issues of SCI. This includes disconnecting the smoothness constraints at known locations (Slater & Binley 2006) or rerunning SCI with bootstrapping subsets of the data set to provide a measure of uncertainty (Yang *et al.* 2014; Fernández-Muñiz *et al.* 2019). However, to date, most inversion approaches generally only account for uncertainty that arises from measurement errors but not the model parameter uncertainties (Tso *et al.* 2017, 2019). Alternative regularization approaches that allow enhancements of sharp interfaces within an image (e.g. Farquharson & Oldenburg 1998), as well as geostatistical approaches (Bouchedda *et al.* 2017; Yeh *et al.* 2002), have also been explored.

Other global techniques, such as those discussed in Sen & Stoffa (2013), have also been used for geophysical inversions. For example, Bijani *et al.* (2017) used a genetic algorithm-based Pareto Multi-Objective Global Optimization (PMOGO) method to perform joint inversions and showed great flexibility and promising results avoiding minimal entrapment. There have also been attempts to improve SCI from some of its limitations, such as using convolutional wavelet transform to perform SCI inversion on the feature space (Pang *et al.* 2020), and the use of area-to-point kriging to obtain fine-scale geophysical properties fields from resolution-limited SCI images (Nussbaumer *et al.* 2019). Finally, complex priors, such as those from training images, can be encoded in a low-dimensional space to aid inversion (Lopez-Alvis *et al.* 2022).

## 1.3 Bayesian inversion

The Bayesian approach to inverse problems (see e.g. Stuart 2010) provide us with a framework to quantity uncertainty in the solution of an inverse problem which, in turn, can be posed in terms of computing the *posterior* distribution of the unknown model parameters given observed data. Since the posterior is, in general, not available in closed form, sampling methods such as Markov Chain Monte Carlo (MCMC) or Sequential Monte Carlo (SMC) are required to approximate statistics of the posterior via Monte Carlo estimates computed from samples. However, in order to compute accurate statistics via fully Bayesian sampling methods, very long chains are often required to obtain a sufficiently large number of decorrelated samples (Iglesias *et al.* 2013a, 2018). This is particularly the case for a wide class of problems in which the parameter that we wish to infer (and hence the posterior) is defined on a very high-dimensional space that arises from discretizing partial differential equations (Cotter *et al.* 2013). Unfortunately, many geophysical problems fall in the above class, because the property value of each grid cell (typically hundred thousands or even millions of them) needs to be inferred from the data. Since MCMC involves evaluating the forward model at every step of the chain, this method is computationally unfeasible for geophysical settings

unless these are either 1-D or low-resolution 2-D, or when the geoelectrical properties are parametrized in terms of a few (i.e. $\approx 10$) parameters. In the context of geophysical inversion, examples of the settings in which MCMC has been used to provide a fully Bayesian, probabilistic estimate of the resultant model include the works of Ramirez *et al.* (2005) and Irving & Singha (2010), with the former using voxel-based proposals, and the latter uses a binary facies-based parametrization.

In order to reduce the computational burden of sampling algorithms for geophysical inverse problems, recent work (Kang *et al.* 2021) has used deep learning (e.g. variational autoencoders) to build computationally efficient surrogates of the forward model that can be used within the sampling algorithm. Furthermore, variational inference approaches (Zhang & Curtis 2020) can be used to characterize the functional form of the posterior and to estimate its key statistics (e.g. mean and variance), rather than obtaining samples from it. Bayesian evidential learning (Scheidt *et al.* 2018) has been proposed to reduce the number of dimensions and map the constitutive statistical relationships between the reduced model parameters and the reduced data using canonical correlation analysis (CCA) in order to allow fast sampling of the posterior space (e.g. Hermans *et al.* 2016; Michel *et al.* 2020; Thibaut *et al.* 2022; Michel *et al.* 2022).

The performance of Bayesian methods is highly dependent on the prior distribution which plays a role analogous to that of regularization in deterministic approaches. While Gaussian priors can be a computationally convenient choice to characterize geophysical properties, the resulting samples from the prior (and hence the posterior) are overly smooth. More recent approaches have shown the use of geostatistics (Aleardi *et al.* 2021), training images (Oware *et al.* 2019), adaptive zone boundary delineation (de Pasquale *et al.* 2019) or Voronoi cells (Galetti & Curtis 2018) can provide a more flexible approach to generate prior samples of geophysical properties that have substantially different values on regions with unknown geometries.

## 1.4 Ensemble Kalman inversion

Ensemble Kalman inversion (EKI) is a computational framework for inverse problems. It comprises a class of algorithms that can be seen as derivate/Jacobian-free optimization methods (Iglesias *et al.* 2013b; Iglesias 2016) as well as sampling schemes that produce a Gaussian approximation of the posterior (Iglesias *et al.* 2018; Iglesias & Yang 2021). We refer the reader to the papers by Calvello *et al.* (2022) and Chada *et al.* (2021) which include extensive reviews with different variants formulations and recent theoretical developments. Methods that are similar to EKI include Ensemble Smoother with multiple data assimilation (Emerick & Reynolds 2013) and Kalman Ensemble Generator (Bobe *et al.* 2020). Here we adopt the Bayesian perspective of EKI and consider the classical EKI algorithm of Iglesias & Yang (2021). This algorithm starts with an initial ensemble of samples (often called particles) from the prior of model parameters. Then, each of these samples is iteratively updated according to a Kalman-based formula which maps the ensemble from the prior into an ensemble from the approximate posterior. The number of iterations (usually between 10 and 20 of them) is determined adaptively to ensure a smooth transition between prior and posterior. At each iteration, the main computational cost is that of running the forward model multiplied by the number of particles (between $10^2$ and $10^3$). Upon convergence, sample statistics can be computed from the posterior ensemble.

In contrast to fully Bayesian algorithms (e.g. MCMC) which are designed to sample from the target (in this context) the posterior distribution, EKI provides only a Gaussian approximation of the posterior. However, EKI does not suffer from the curse of dimensionality inherent to sampling methods while providing good approximations of the posterior. Indeed, numerical work on tractable dimensions has shown that, for a wide range of inverse problems, EKI provides accurate approximations of the posterior while incurring in a fraction of the cost of fully Bayesian methods (Iglesias *et al.* 2013b, 2018; Iglesias 2015).

Because of the need to run the forward model for each ensemble member, the computational cost of EKI is higher than conventional variational optimization methods such as SCI. However, as mentioned earlier, EKI does not require the Jacobian of the forward map (i.e. the input–output map that arises from running the forward model). Without the limiting requirement for Jacobians of the forward map, fresh ways to tackle the ill-posedness of geophysical inverse problems via realistic (often non-differentiable) parametrizations of the unknown quantities become possible. This advantage of EKI was exploited recently (Tso *et al.* 2021) where EKI approach with level set parametrization was used to invert ERT data. Instead of directly estimating the geophysical properties everywhere in the model domain, Tso *et al.* (2021) introduced a level-set function (also defined everywhere) that parametrized the geometry of different zones. By thresholding the level-set function, we obtained a (non-differentiable) map that acted as a classifier for a set of unknown zones/regions with different, and also unknown, values of geophysical properties. The inverse problem was then posed in terms of estimating the level-set function as well as the resistivity values on each zones. Using EKI to solve this inverse problem enabled us to delineate the geometry of structures in a geophysical property field.

One of the main advantages of the level-set parametrization used in Tso *et al.* (2021) is that it can describe a wide range of arbitrary geometries. However, this level of generality means that the inverse problem is high-dimensional because the level-set at every grid cell of the computational domain needs to be estimated. Fortunately, as stated earlier, the EKI framework is very robust and can handle very large problems as shown in Tso *et al.* (2021).

Using level-set parametrizations within EKI for generic inverse problems was first proposed in Iglesias *et al.* (2016) and further studied by Chada *et al.* (2018). This framework has also been applied by Muir & Tsai (2020) and Muir *et al.* (2022) in deep-earth geophysics applications. The level-set parametrization within EKI was also used recently to infer permeability and porosity during resin transfer moulding for composite manufacturing (Matveev *et al.* 2021; Iglesias *et al.* 2018) as well as to infer elastic properties of biological tissue via magnetic resonance elastography (Iglesias *et al.* 2022).

## 1.5 Induced polarization

Geoelectrical methods are one of the most commonly used techniques for near-surface geophysical investigations. Electrical resistivity measurements are sensitive to both pore volume and pore surface area properties, but their utility for permeability ($k$) estimation, for example, is inherently limited because the two contributions cannot be separated; meanwhile, induced polarization (IP) has unique sensitivity to interconnected pore surface area (Slater 2007; Kemna *et al.* 2012). The past two decades has seen a steady growth in IP applications for subsurface investigations, including hydrological (e.g. Kemna *et al.* 2004; McLachlan *et al.* 2020; Rucker *et al.* 2021), engineering (e.g. Slater & Binley 2006; Revil *et al.* 2020), and biogeochemical (Williams *et al.* 2005; Ntarlagiannis *et al.* 2005; Kessouri *et al.* 2019; Saneiyan *et al.* 2019) applications. In particular, IP has been used in a number of studies to quantify the distributions of $k$, such as delineating $k$ profiles from cross-borehole IP surveys (Binley *et al.* 2016), mapping $k$ of a riverbed (Benoit *et al.* 2018), identifying the contact of two lithological units in a river corridor (Slater *et al.* 2010; Mwakanyamale *et al.* 2012), or assess the variation in soil moisture and textural properties in studies of unstable hillslopes (Revil *et al.* 2020). Similarly, IP has been used to delineate the subsurface hydrocarbon contamination at a former industrial site (Flores Orozco *et al.* 2013).

## 1.6 Uncertainty propagation

The full value of the unique sensitivity of IP can only be shown when its inversion can provide some measures of uncertainty. Unfortunately, such analysis is rarely conducted for field-scale studies. A number of previous works have used Bayesian method such as MCMC to model IP (Chen *et al.* 2012; Madsen *et al.* 2017; Bérubé *et al.* 2017). Few studies have incorporated IP data for Bayesian hydrogeophysical analysis; however, they tend to use the smooth inverted resistivity and phase angle images deterministically as input for subsequent Bayesian analysis. For example, Wainwright *et al.* (2016) use inverted IP data to inform probabilistic mapping of biogeochemical hotspots using an indicator field approach. The extent to which uncertainty in IP imaging propagate to uncertainty in images of hydrological properties remain largely unknown. Concurrently, there is an interest to understand the way in which uncertainties in geophysical images are translated to maps of hydrological properties—this would require an understanding of the effects of uncertainties in the petrophysical relationships that link geophysical and hydrological variables. This can be done both numerically (Day-Lewis *et al.* 2005; Moysey *et al.* 2005; Singha & Moysey 2006; Tso *et al.* 2019) and empirically, for example using colocated data for comparison (Isunza Manrique *et al.* 2023). Brunetti & Linde (2018) showed that accounting for petrophysical uncertainty improves Bayesian hydrogeophysical model selection. Current understanding is lacking on the impact of uncertainty from inverted IP data on hydrological estimates. It is critical to assess the magnitude and sources of uncertainties within images of inferred physical properties, such as $k$, that are generated from IP data.

## 1.7 Paper overview

The EKI method outlined in Tso *et al.* (2021) has clear advantages over existing methods for ERT inversions of DC resistivity data, especially for resistivity fields with discontinuous regions. However, its benefits for IP data and for generating maps of hydrological properties have yet to be illustrated. The principal aim of this work is to apply and demonstrate the use of EKI with level set parametrization for IP imaging, and to advance the use of EKI in geophysics by investigating the effects of different prior formulations and data noise levels. The outcomes of this inversion can then be used in a workflow that we propose to estimate the spatial distribution of hydrological properties with uncertainty bounds from IP imaging. In section 2 we introduce the methods we use for which, in addition to the single level-set parametrization used in Tso *et al.* (2021), we consider a parametrization in terms of multiple level-sets which can allow us to parametrize (and hence infer) a larger range of geophysical scenarios. We report results from our example applications in

Section 3. Finally, we discuss our findings in Section 4 and provide our conclusions in Section 5.

## 2 METHODS

In this section, we introduce the EKI framework for IP which relies on a level-set parametrization. First, we introduce the forward model for IP in Section 2.1 as well as the SCI method. The Bayesian inversion approach is discussed in Section 2.2. The parametrization of geophysical properties in terms of level-sets is presented in Section 2.3 and the EKI approach to address the re-parametrized problem is developed in Section 2.4. Finally, in Section 2.5 we present implementation details.

### 2.1 Forward problem and SCI

In the frequency domain, IP can be represented as a frequency-dependent complex resistivity $\rho^*$. For a generalized 2-D or 3-D IP problem with a heterogeneous $\rho^*$ field, the measured (complex) potential $V^*$ due to current injection at electrode locations $x_A$ and $x_B$, with strength $I$, can be described by the following partial differential equation:

$$-\nabla \cdot \left( \frac{1}{\rho^*(\mathbf{x})} \nabla V^*(\mathbf{x}) \right) = I(\delta(x - x_A) - \delta(x - x_B)), \quad (1)$$

with appropriate boundary conditions (e.g. Binley & Slater 2020).

To infer the field $\rho^*$, a conventional SCI can be used. This procedure uses the Gauss–Newton method to minimizes the combined observed data, $d^*$ [which is the log-transformed complex impedance $\log(V^*/I)$ in this study], and the model misfit by minimizing the following objective function (e.g. Binley & Slater 2020):

$$\frac{1}{2} \| W_{d^*}(d^* - \mathcal{F}(\rho^*)) \|^2 + \frac{\beta}{2} \| W_{\rho^*} \rho^* \|^2, \quad (2)$$

where $\mathcal{F}$ is the forward (or parameter-to-output) map that predicts potentials (e.g. via eq. 1) for a given $\rho^*$, $W_{d^*}$ is a matrix that assigns weights (accuracy) to the data, $\beta$ is a tuning regularization parameter and $W_{\rho^*}$, often referred to as the model roughness matrix, is usually a differential operator that enforces smoothness in a minimizer of eq. (2) (e.g. gradient/Laplacian filters). The second term in eq. (2) can be seen as a form of Tikhonov regularization that stabilizes the inversion.

### 2.2 The Bayesian approach

Let us use the representation of the complex resistivity $\rho^*(\mathbf{x})$ in terms of its magnitude and phase angle denoted by $\rho(\mathbf{x})$ and $\varphi(\mathbf{x})$, respectively. We note that $\rho(\mathbf{x})$ is equivalent to the DC resistivity since, for most hydrogeophysics problems, we expect the phase angle to be small (typically a few tens of milliradians). For computational convenience, the complex vector of measured potential $d^*$ is rewritten as a real vector of the form $d^* = [d, \xi]$, where $d$ and $\xi$ denote the magnitude and phase angle of $d^*$, respectively.

We assume that the data and the unknown are related via

$$d = \mathcal{F}_d(\rho, \varphi) + \eta_d \quad (3)$$

$$\xi = \mathcal{F}_\xi(\rho, \varphi) + \eta_\xi, \quad (4)$$

where $\eta_d$ and $\eta_\xi$ are independent random measurement errors that follow Gaussian distributions with zero mean and covariance $\Sigma_d$ and $\Sigma_\xi$, respectively. Here $\mathcal{F}_d$ and $\mathcal{F}_\xi$ denote the two components of the forward map (magnitude and phase angle of potential's prediction).

A complete formulation of the inverse problem for IP data will consist of jointly inferring $\rho(\mathbf{x})$ and $\varphi(\mathbf{x})$ given both $d$ and $\xi$. Here we adopt a more practical approach in which we first invoke the assumption that $\varphi$ is small and compute, as in standard ERT, the DC resistivity from measurements $d$. Then, we use our estimate of DC resistivity to infer phase angle $\varphi$ given $\xi$. Since the imaginary component of complex resistivity is so small, the phase angle estimation can be seen as a final correction after DC resistivity is estimated.

We formulate both inverse problems from our two-step method via the Bayesian approach (Stuart 2010) in which we assume that $\rho(\mathbf{x})$ and $\phi(\mathbf{x})$ are random functions. We put a prior distribution for $(\rho, \varphi)$, denoted by $\mathbb{P}(\rho, \varphi)$ and, for simplicity, we assume independence (under the prior) so that $\mathbb{P}(\rho, \varphi) = \mathbb{P}(\rho)\mathbb{P}(\varphi)$. The prior comprises our knowledge of the unknown functions $\rho$ and $\varphi$ prior to the data.

The solution of the first Bayesian inverse problem (inferring $\rho$ given $d$) is the posterior on the DC resistivity $\mathbb{P}(\rho|d)$ which, from Bayes's rule the is given by

$$\mathbb{P}(\rho|d) \propto \mathbb{P}(d|\rho, \varphi = 0)\mathbb{P}(\rho), \quad (5)$$

where $\mathbb{P}(d|\rho, \varphi = 0)$ is the likelihood of $d$ evaluated at phase angle $\varphi = 0$.

For the second step we compute the marginal posterior of $\varphi$ given $\xi$. This is defined by

$$\mathbb{P}(\varphi|\xi) = \int \mathbb{P}(\varphi, \rho|\xi)d\rho, \quad (6)$$

where $\mathbb{P}(\varphi, \rho|\xi)$ is the joint posterior which, again from Bayes rule, is given by

$$\mathbb{P}(\varphi, \rho|\xi) \propto \mathbb{P}(\xi|\rho, \varphi)\mathbb{P}(\varphi)\mathbb{P}(\rho), \quad (7)$$

where $\mathbb{P}(\xi|\rho, \varphi)$ is the likelihood of $\xi$. In order to further simplify the problem, we propose to approximate $\mathbb{P}(\xi|\rho, \varphi) \approx \mathbb{P}(\xi|\rho = \overline{\rho}, \varphi)$, where $\overline{\rho}$ denotes the mean of DC resistivity posterior $\mathbb{P}(\rho|d)$ defined via eq. (5). In other words, we build (eq. 7) with the likelihood of $\xi$ evaluated at the mean of the DC resistivity posterior. Using this approximation in eq. (6) we find

$$\mathbb{P}(\varphi|\xi) \propto \int \mathbb{P}(\xi|\rho = \overline{\rho}, \varphi)\mathbb{P}(\varphi)\mathbb{P}(\rho)d\rho = \mathbb{P}(\xi|\rho = \overline{\rho}, \varphi)\mathbb{P}(\varphi). \quad (8)$$

From our Gaussian assumptions on $\eta_d$ and $\eta_\xi$ and expressions (3) and (4), it follows that

$$\mathbb{P}(\rho|d) \propto \exp\left[ -\frac{1}{2} \left\| \Sigma_d^{-1/2}(d - \mathcal{F}_d(\rho, 0)) \right\|^2 \right] \mathbb{P}(\rho) \quad (9)$$

and

$$\mathbb{P}(\varphi|\xi) \propto \exp\left[ -\frac{1}{2} \left\| \Sigma_\xi^{-1/2}(\xi - \mathcal{F}_\xi(\overline{\rho}, \varphi)) \right\|^2 \right] \mathbb{P}(\varphi). \quad (10)$$

Algorithmically, the proposed approximations amount to computing the marginal posterior of the DC resistivity (given $d$) while keeping the phase angle fixed $\varphi = 0$ in the likelihood. Then, the posterior marginal for the phase angle (given $\xi$) is computed using the mean of DC resistivity posterior marginal as a fixed estimate within the likelihood function. Our objective now is to use the EKI framework to produce samples from the approximate posterior from eqs (9) and (10). To this end, in the following subsection we introduce suitable parametrizations of the resistivity and the phase angle that will enable us to infer discontinuous properties.

## 2.3 Parametrization of spatial fields

As discussed in the introduction, the selection of the prior in Bayesian algorithms is crucial for their estimation performance. This is particularly challenging for subsurface properties with discontinuities, which makes common techniques to prescribe priors such as stationary Gaussian random fields not applicable. To tackle this challenge, this work utilizes two level-set parametrizations that enable us to delineate arbitrarily shaped zones of different permeability and to infer the values of the DC resistivity and phase angle within each of those zones. Note that through the paper, for simplicity, we refer to DC resistivity, or complex resistivity magnitude, as 'resistivity'.

The first level-set parametrization that we use is the one used in Tso *et al.* (2021) to invert ERT data via EKI. This parametrization requires only one level-set function and several thresholds to parametrize multiple zones. For completeness we include the description in Appendix B. By construction, this single level-set formulation does not allow more than two zones to intersect which may be a disadvantage in some geologic settings. To overcome this disadvantage, here we also consider a level-set approach that uses multiple level-set functions to allow the parametrization (and hence inference) of regions that can all intersect. This level-set approach was initially proposed in Litman (2005) for inverse scattering, and we introduce it below in the context of both DC and IP inversion. While this approach with multiple level-sets can handle more geophysical settings compared to the one used in Tso *et al.* (2021), it increases the number of unknowns and hence the dimension of the input space. Therefore, more samples may be needed within the inversion algorithm with the corresponding increase in computational cost. Hence, we recommend using the single level-set approach in Tso *et al.* (2021) as long as there is strong prior evidence (e.g. from preliminary stratigraphic analysis) suggesting that the underlying geophysical can be described with the single level-set parametrization.

### 2.3.1 Parametrization with multiple level-sets

Let us adapt the approach of Litman (2005) for our inverse problem with IP data. For brevity we describe only the parametrization of resistivity $\rho(\mathbf{x})$ while we use the analogous parametrization for the phase angle $\varphi(\mathbf{x})$. Also, for simplicity we consider a simple four-zone parametrization that relies on the assumption that the unknown resistivity takes only four (unknown) resistivity values $\rho_1$, $\rho_2$, $\rho_3$ and $\rho_4$ on (unknown) regions denoted by $\Omega_1$, $\Omega_2$, $\Omega_3$ and $\Omega_4$, respectively. These regions are, in turn, parametrized via thresholding two level-set functions, denoted by $\xi_1(\mathbf{x})$ and $\xi_2(\mathbf{x})$. In more detail, we assume those regions are defined by

$$
\begin{aligned}
\Omega_1 &= \{\mathbf{x} \,:\, \xi_1(\mathbf{x}) \leq \alpha_1, \, \xi_2(\mathbf{x}) \leq \alpha_2\}, \\
\Omega_2 &= \{\mathbf{x} \,:\, \xi_1(\mathbf{x}) > \alpha_1, \, \xi_2(\mathbf{x}) \leq \alpha_2\}, \\
\Omega_3 &= \{\mathbf{x} \,:\, \xi_1(\mathbf{x}) \leq \alpha_1, \, \xi_2(\mathbf{x}) > \alpha_2\} \\
\Omega_4 &= \{\mathbf{x} \,:\, \xi_1(\mathbf{x}) > \alpha_1, \, \xi_2(\mathbf{x}) > \alpha_2\},
\end{aligned}
\tag{11}
$$

where $\alpha_1$ and $\alpha_2$ are user defined parameters. In summary, the 4-zone characterization of the unknown resistivity is given by

$$
\rho(\mathbf{x}) = \begin{cases}
\rho_1, & \xi_1(\mathbf{x}) \leq \alpha_1, \, \xi_2(\mathbf{x}) \leq \alpha_2 \\
\rho_2, & \xi_1(\mathbf{x}) > \alpha_1, \, \xi_2(\mathbf{x}) \leq \alpha_2 \\
\rho_3, & \xi_1(\mathbf{x}) \leq \alpha_1, \, \xi_2(\mathbf{x}) > \alpha_2 \\
\rho_4, & \xi_1(\mathbf{x}) > \alpha_1, \, \xi_2(\mathbf{x}) > \alpha_2.
\end{cases}
\tag{12}
$$

The level-set parametrization above can be used to parametrize $2^N$ zones ($N > 1$) in terms $N$ level-set functions. If the number of zones is not a power of two we can simply split one of the zones. For example, in the case where we know that only three zones exist, we can use the 4-zone parametrization above with $\rho_3 = \rho_4$ and the third zone given by $\Omega_3 \cup \Omega_4$.

In principle, this multiple level-set parametrization requires us to know, *a priori*, the number of zones on which the geoelectric properties take different values. However, it is worth noticing that the value, $\rho_n$, that the property takes on the $n$th zone is also an unknown that we infer alongside the level-set functions. Therefore, as long as we choose $N$ to be the largest number of zones that we could expect, and provided the measurements are sufficiently informative, the EKI framework should infer the correct number of zones even if this is smaller than the original $N$. Nonetheless, the inference of the number of zones will be achieved indirectly by identifying the same (or very close) value of $\rho_n$'s on different zones.

With the aid of the parametrization in (12), the posterior for the DC resistivity can be rewritten in terms of the joint posterior of $\rho_1$, $\rho_2$, $\rho_3$, $\rho_4$ and the level-set functions $\xi_1(\mathbf{x})$ and $\xi_2(\mathbf{x})$ that determine the regions defined in eq. (11). Our aim is to choose a prior of Gaussian random fields (GRF) for $\xi_1(\mathbf{x})$ and $\xi_2(\mathbf{x})$ and then approximate the reparametrized marginal posteriors via EKI.

### 2.3.2 Parametrization of random fields

Iglesias (2016) and Chada *et al.* (2018) have shown that an accurate EKI implementation of a level-set parametrization requires to further parametrize the GRF (for the level-set function) in terms of hyperparameters which should be inferred within the EKI algorithm. To this end we take a further step and parametrize both $\xi_1(\mathbf{x})$ and $\xi_2(\mathbf{x})$ (or more when more than four regions are considered) by using the stochastic partial differential equations (SPDE) framework in which 2-D realizations of GRFs can be obtained by solving the following fractional SPDE:

$$
\begin{aligned}
&\left[ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \nabla \cdot \begin{pmatrix} L_{1,\alpha}^2 & 0 \\ 0 & L_{2,\alpha}^2 \end{pmatrix} \nabla \right]^{(\nu_\alpha + 1)/2} \xi_\alpha(\mathbf{x}) \\
&= \left[ \tau_\alpha^2 2\pi^{1/2} \frac{\Gamma(\nu_\alpha + 1/2)}{\Gamma(\nu_\alpha)} L_{1,\alpha} L_{2,\alpha} \right]^{1/2} \omega_\alpha(\mathbf{x})
\end{aligned}
\tag{13}
$$

with $\alpha = 1, 2$ and where $\nu_\alpha$ controls smoothness of the realization, $L_{1,\alpha}$ and $L_{2,\alpha}$ are intrinsic length scales along the horizontal and vertical direction, respectively, $\tau_\alpha$ is an amplitude scale, $\Gamma$ denotes the gamma function and $\omega_\alpha(\mathbf{x})$ is Gaussian white noise. Imposing appropriate boundary conditions for eq. (13) (Roininen *et al.* 2014) leads to solutions which are GRFs with Matérn isotropic covariance function (Lindgren *et al.* 2011). This parametrization of GRFs can be modified to include a degree of anisotropy along some preferential direction (Roininen *et al.* 2014). Other parametrizations of GRFs that can alternatively be used are those defined on simple domains, and for which the eigenfunctions and eigenvalues of prior covariance can be obtained closed-form (Dunlop *et al.* 2017).

Since the level-set functions $\xi_\alpha(\mathbf{x})$ are merely artificial functions (i.e. with no direct physical interpretation) that we use to parametrize the unknown geometry of the zones, we consider a fixed amplitude scale $\tau_\alpha = 1$ which means that, at every point of the domain $\mathbf{x}$, $\xi_\alpha(\mathbf{x})$ is a standard normal. Based on this selection, we can select threshold values $\alpha_1$ and $\alpha_2$ in (11) and (12) to ensure that, *a priori*, at every $\mathbf{x}$ there is a specific value for the probability of this point to

belong to each zone. For example, taking values $\alpha_1 = \alpha_2 = 0$ will ensure that there is a (prior) probability equal to 0.25 that any given point **x** belongs to any of the four zones.

We recall that the parameter $\nu_\alpha$ defines the smoothness of the level-set function $\xi_\alpha(\mathbf{x})$ and, thus, the smoothness (or roughness) of the interface between zones. Although we could include this parameter as part of the unknown input parameters that we wish to infer, for simplicity here we keep the parameter $\nu_\alpha = 2.0$ fixed. For the 2-D experiments that we present in the following section, our selection of $\nu_\alpha$ provides a sufficient degree of smoothness of the level-set function in order to capture well-defined zones.

Given those considerations let us now write the parametrizations above in a more compact form by noticing that eq. (13) defines an operator $\mathcal{P}^{GRF}$ that takes the hyperparameters, $L_{1,\alpha}, L_{2,\alpha}$ of the GRF (recall $\tau_\alpha = 1$ and $\nu_\alpha = 2$ are now fixed) together with the white noise $\omega_\alpha(\mathbf{x})$ in the right-hand side of eq. (13), into a realization of the GRF $\xi_\alpha(\mathbf{x})$, that is

$$\xi_\alpha = \mathcal{P}^{GRF}(L_{1,\alpha}, L_{2,\alpha}, \omega_\alpha). \tag{14}$$

On the other hand, note that eq. (12) defines a mapping of the form

$$\rho = \mathcal{P}^{LS}(\rho_1, \rho_2, \rho_3, \rho_4, \xi_1, \xi_2). \tag{15}$$

We can thus compose these two functions to finally arrive at $\rho = \mathcal{P}(u_\rho)$ where

$$u_\rho(\mathbf{x}) = \left( \rho_1, \rho_2, \rho_3, \rho_4, \{L_{1,\alpha}, L_{2,\alpha}, \omega_\alpha(\mathbf{x})\}_{\alpha=1}^2 \right) \tag{16}$$

and

$$\mathcal{P}(u_\rho) = \mathcal{P}^{LS}\left( \rho_1, \rho_2, \rho_3, \rho_4, \mathcal{P}^{GRF}(L_{1,1}, L_{2,1}, \omega_1), \mathcal{P}^{GRF}(L_{1,2}, L_{2,2}, \omega_2) \right). \tag{17}$$

We may use the same four-zone parametrization for the phase angle $\varphi = \mathcal{P}(u_\varphi)$, with parameters comprised in $u_\varphi$, analogous to those in eq. (16). Note that the parameters $L_{1\alpha}, L_{2\alpha}, \omega_\alpha(\mathbf{x})$, and hence the corresponding level-set functions $\xi_\alpha(\mathbf{x})$ that we inferred for the DC resistivity will be, in general, different from those for the phase angle. Consequently, the inferred geometry of the four zones obtained for the DC resistivity may not necessarily coincide with the geometry of the zones determined by the inferring phase angle.

Finally, it is worth mentioning that the proposed parametrization can be easily extended for a 3-D geometry and also to account for the case in which the unknown property (e.g. resistivity or phase angle) is spatially variable within each of the different regions (see ERT examples in Tso *et al.* 2021).

## 2.4 Solving the reparametrized Bayesian inverse problem via EKI

In this section, we use the parametrizations of the DC resistivity and phase angle to approximate the posteriors from eq. (9) and eq. (10). Let us define

$$\mathcal{G}_d(u_\rho) := \mathcal{F}_d(\rho, 0) = \mathcal{F}_d(\mathcal{P}(u_\rho), 0),$$

where we have used the parametrization $\rho = \mathcal{P}(u_\rho)$ from the previous subsection. We now wish to approximate the posterior on the parameter $u_\rho$ which from eq. (3) follows:

$$d = \mathcal{G}_d(u_\rho) + \eta_d, \qquad \text{with} \quad \eta_d \sim N(0, \Sigma_d), \tag{18}$$

From Bayes' rule we have that the sought posterior is

$$\mathbb{P}(u_\rho|d) \propto \exp\left[ -\frac{1}{2} \left\| \Sigma_d^{-1/2}(d - \mathcal{G}_d(u_\rho)) \right\|^2 \right] \mathbb{P}(u_\rho), \tag{19}$$

where $\mathbb{P}(u_\rho)$ denotes the prior on $u_\rho$. Once this prior is specified (see Section 2.4.1), the prior $\mathbb{P}(\rho)$ on the original physical property $\rho$ can be defined as the push-forward measure of $\mathbb{P}(u_\rho)$ under the map $\mathcal{P}$. This is denoted by $\mathcal{P}^\#\mathbb{P}(u_\rho)$. In simple words this means that samples, $\rho^{(j)}$, of $\mathbb{P}(\rho)$ can be obtained by sampling from $\mathbb{P}(u_\rho)$ and mapping those samples, say $u_\rho^{(j)}$, into the physical property via $\rho^{(j)} = \mathcal{P}(u_\rho^{(j)})$.

It can be shown that the above selection of the prior on $\rho$ as $\mathbb{P}(\rho) = \mathcal{P}^\#\mathbb{P}(u_\rho)$ implies $\mathbb{P}(\rho|d) = \mathcal{P}^\#\mathbb{P}(u_\rho|d)$, that is the sought marginal posterior on the DC resistivity can be obtained by pushing forward the posterior on the parameters $u_\rho$. Again, in terms of samples which are provided by the EKI discussed below, this simply means that once we obtain posterior samples for the parameter $u_\rho$, evaluating $\mathcal{P}(u_\rho)$ gives us samples from the posterior DC resistivity.

Similarly, by defining $\mathcal{G}_\xi(u_\varphi) := \mathcal{F}_\xi(\overline{\rho}, \varphi)$ we can formulate the posterior on the phase angle parameters

$$\mathbb{P}(u_\varphi|\xi) \propto \exp\left[ -\frac{1}{2} \left\| \Sigma_\xi^{-1/2}(\xi - \mathcal{G}_\xi(u_\varphi)) \right\|^2 \right] \mathbb{P}(u_\varphi), \tag{20}$$

where $\mathbb{P}(u_\varphi)$ denotes the prior on $u_\varphi$ which we assume have analogous form to that in eq. (23).

We reiterate that $\mathcal{G}_\xi$ depends on the mean of the approximate marginal posterior $\mathbb{P}(\rho|d)$. Nonetheless, both eqs (19) and (20) can be written as:

$$\mathbb{P}(u|w) \propto \exp\left[ -\frac{1}{2} \left\| \Sigma^{-1/2}(w - \mathcal{G}(u)) \right\|^2 \right] \mathbb{P}(u), \tag{21}$$

where

$$(u, w, \mathcal{G}, \Sigma, \mathbb{P}(u)) = \begin{cases} (u_\rho, d, \mathcal{G}_d, \Sigma_d, \mathbb{P}(u_\rho)) \text{ to approx. (19)} \\ (u_\varphi, \xi, \mathcal{G}_\xi, \Sigma_\xi, \mathbb{P}(u_\varphi)) \text{ to approx. (20)}. \end{cases}$$

Therefore we can use a generic solver for Bayesian inverse problems to first approximate (19) and then eq. (20) as discussed earlier. To this end, here we apply the EKI framework from Iglesias *et al.* (2018) and Iglesias & Yang (2021) that we summarize in Algorithm 1 in Appendix A. The EKI algorithm is presented to approximate the posterior $\mathbb{P}(u|w)$ from eq. (21) in which observed/measured data is $w$, the unknown parameter is $u$, the forward map is $\mathcal{G}(u)$ and the noise covariance is $\Sigma$. As discussed in Section 1.4, the EKI algorithm defines a transition between the prior and the posterior via constructing a sequence of $q$ intermediate Gaussian measures:

$$\mathbb{P}(u) = \mathbb{P}_0(u) \rightarrow ... \rightarrow \mathbb{P}_{q+1}(u) = \mathbb{P}(u|w). \tag{22}$$

Each intermediate distribution $\mathbb{P}_m(u)$ ($m = 1, \ldots, q$) is approximated with an ensemble of particles, $\{u_m^{(j)}\}_{j=1}^J$. The EKI algorithm (Iglesias *et al.* 2018) consists of iteratively updating each particle according to an update formula that, at the iteration level $n$, can be derived from (i) linearizing the forward map around the mean of $\mathbb{P}_m(u)$, (ii) applying Bayes rule to the linearized problem involving the Gaussian approximation of $\mathbb{P}_m(u)$ and finally (iii) using covariance approximations for the derivatives from the linearized problem.

We follow the approach from Iglesias & Yang (2021) to compute the number of intermediate distributions, $q$, adaptively. Convergence of Algorithm 1 is controlled by the parameter $s_m$. The algorithm stops once $s_{m+1} = 1$, that is we set $q = m + 1$ and the corresponding particles $\{u_{m+1}^{(j)}\}_{j=1}^J$ provide a Gaussian approximation to the posterior $\mathbb{P}(u|w)$. The computational cost of EKI is given by $J \times q$ simulations which scales well with respect to the number of particles $J$. Hence, the computational burden of EKI can be amortized via the use of parallel computing since the prediction step Algorithm 1 is perfectly parallelizable.

Despite the large amount of recent progress in developing theory for EKI (see Calvello *et al.* 2022), the convergence of any existing variant of EKI algorithms can only be rigorously proved in the case when the forward map is linear and when the prior and the noise distributions are Gaussian. In particular, the seminal work of Schillings & Stuart (2017) introduced the continuous-time formulation of EKI which, in turn, lead to convergence proofs in the mean-limit settings albeit restricted to the case of linear forward models (which is not the case for ERT or IP). Notwithstanding, the EKI algorithm can be applied for non-linear forward models as well as any choice of distribution for the prior and the noise, and although further theory that ensures convergence of EKI for the non-linear and non-Gaussian cases is lacking, a large body of numerical evidence suggests that, for a broad range of problems, EKI can provide robust/stable estimates with the same accuracy of optimizer and/or fully-Bayesian samplers (such as MCMC or SMC) when the computational budget does not allow for large number of forward model solves (Iglesias *et al.* 2013a, 2018 ; Iglesias 2015). It is thus clear that EKI is a suitable choice for high-dimensional Bayesian inverse problems which cannot be solved via fully Bayesian samplers.

### 2.4.1 The prior

We assume that, under the prior, the components of $u_\rho$ [i.e. eq. (16) for the 4-zone case] are independent. Thus, can write the prior, for the general case of $2^N$-zones and $N$ level-set functions, as

$$\mathbb{P}(u_\rho) = \prod_{n=1}^{2^N} \mathbb{P}(\rho_n) \prod_{\alpha=1}^{N} \mathbb{P}(L_{1,\alpha}) \mathbb{P}(L_{2,\alpha}) \mathbb{P}(\omega_\alpha), \qquad (23)$$

where the terms in the right-hand side of eq. (23) are the priors of each component of $u_\rho$. As discussed earlier, here we choose the prior on the function $\omega_\alpha(\mathbf{x})$ (i.e. $\mathbb{P}(\omega_\alpha)$ in eq. 23) as Gaussian white noise, so that the corresponding samples of each level-set function, $\xi_\alpha(\mathbf{x})$, are realizations of GRFs.

Before we introduce the prior of the length scales of the level-set functions $L_{1,\alpha}$ and $L_{2,\alpha}$, let us first notice that these hyperparameters of the level-set function $\xi_\alpha(\mathbf{x})$ determine how rapid this function changes along each of the two spatial directions. For example, functions with large $L_{2,\alpha}$ and small $L_{1,\alpha}$ will enable to describe zones that are long along the vertical direction while shorter in the horizontal direction. When the geometry of the zones is completely unknown *a priori*, we suggest placing a uniform prior over a domain that covers both small and relatively long length scales. If we assume that the 2-D domain consist of the rectangular region [0, $D_1$] × [0, $D_2$], our numerical experiments show that the following uniform priors

$$\mathbb{P}(L_{1,\alpha}) = \mathbb{P}(L_{2,\alpha}) = \mathcal{U}\left[\frac{D_\alpha}{15}, \frac{D_\alpha}{5}\right], \qquad \alpha = 1, \dots, N \qquad (24)$$

cover a wide range of length scales (relative to the size of the domain) that can be used to characterize zonal geometries of different sizes. As shown in previous work on level-set parametrization for Bayesian inversion (Dunlop *et al.* 2017; Chada *et al.* 2018), if we do not infer these length scales within the EKI (i.e. if we keep them fixed) there is a high risk that the geometric features of the unknown may not be accurately captured with the inferred level-set parametrization.

For the priors of the values of the zones resistivities, that is $\rho_n$s we also select uniform priors

$$\mathbb{P}(\rho_n) = \mathcal{U}\left[\rho_{n*}, \rho_n^*\right], \qquad n = 1, 2, \dots, 2^N, \qquad (25)$$

where the lower and upper values $\rho_{n*}$ and upper values $\rho_n^*$ are specified according to each specific example that we discuss in the following section. These limit values may be informed with our prior knowledge of each zone's resistivity, but in the absence of this knowledge, we can simply use the same prior on each zone.

Furthermore, even though the priors are defined on specific intervals (e.g. $\left[\rho_{n*}, \rho_n^*\right]$ for resistivity) the posterior ensemble can take values outside these intervals since we do not impose any constraints on the updated particles (strictly speaking, this violates Bayes' formulation and we recommend re-running the inversion with larger prior intervals). Therefore, one can expect (as we confirm in our synthetic experiments) that when the prior interval is not well specified, measurements will be sufficiently informative to produce posteriors that are centred around the truth.

Finally, we make the choice of uniform priors for simplicity. Nonetheless, prior information on the site could suggest that different distributions are more suitable priors. There is, again, no restriction in terms of the type of distribution which can be used within the EKI algorithm.

### 2.4.2 Measures of performance and posterior estimates

The performance of the EKI algorithm can be monitored in terms of various quantities related to the data misfit as proposed in Iglesias & Yang (2021). More specifically, we consider:

$$\mathcal{D}_m = \frac{1}{M} \left\| \frac{1}{J} \sum_{j=1}^{J} \left[ \Sigma^{-1/2}\left(w - \mathcal{G}(u_m^{(j)})\right) \right] \right\|^2, \qquad (26)$$

where $M$ is the size of the vector, $w$, that is the observed data. To the first order, it is not difficult to see (e.g. from Iglesias & Yang 2021) that

$$\mathcal{D}_m = \frac{1}{M} \left\| \Sigma^{-1/2}\left[w - \frac{1}{J}\sum_{j=1}^{J}\mathcal{G}(u_m^{(j)})\right] \right\|^2$$
$$\approx \frac{1}{M} \left\| \Sigma^{-1/2}\left[w - \mathcal{G}(\overline{u}_m)\right] \right\|^2, \qquad (27)$$

where $\overline{u}_m$ denotes the ensemble mean at the iteration $m$. On the other hand, since our underlying observational model is of the form $w = \mathcal{G}(u) + \eta$, where $\eta \in \mathbb{R}^M$ is Gaussian with zero mean and covariance $\Sigma$. It then follows that

$$\left\| \Sigma^{-1/2}(w - \mathcal{G}(u)) \right\|^2 = \left\| \Sigma^{-1/2}\eta \right\|^2 \sim \chi_M,$$

where $\chi_M$ is the chi-squared distribution with $M$ degrees of freedom. Thus,

$$\mathbb{E}\left[ \left\| \Sigma^{-1/2}\eta \right\|^2 \right] = M.$$

Therefore, if the converged posterior ensemble yields output values close to the measurement error (or noise level), we should expect that the data misfit from eq. (26) achieve values close to one.

Upon convergence, we compute the means of the parameters' marginals posterior transformed into physical space. More specifically, we compute

$$\overline{u}_\rho = \frac{1}{J}\sum_{j=1}^{J} u_\rho^{(j)}, \quad \text{and} \quad \overline{u}_\varphi = \frac{1}{J}\sum_{j=1}^{J} u_\varphi^{(j)}, \qquad (28)$$

where $\{u_\rho^{(j)}\}_{j=1}^{J}$ and $\{u_\varphi^{(j)}\}_{j=1}^{J}$ denote the converged samples from the posteriors (i.e. eqs 19 and 20) obtained via EKI. Then, we compute

and display the transformed posterior means defined by

$$\rho^{\star} = \mathcal{P}(\overline{u}_{\rho}), \quad \text{and} \quad \varphi^{\star} = \mathcal{P}(\overline{u}_{\varphi}). \tag{29}$$

As discussed earlier, given the above converged samples of the parameters, samples from the marginal posteriors of DC resistivity and phase angle are obtained via

$$\rho^{(j)} = \mathcal{P}\left(u_{\rho}^{(j)}\right), \quad \text{and} \quad \varphi^{(j)} = \mathcal{P}\left(u_{\varphi}^{(j)}\right). \tag{30}$$

Note that the sample means computed from the individual posterior samples in eq. (30) do not coincide with the transformed means computed via eq. (29). Here we consider both quantities but for visualization and geological interpretation the latter provide us with a clear delineation of the interfaces between the zones. Indeed, by definition of the map $\mathcal{P}$ in eq. (17) (see also eqs 15 and 12) the estimates $\rho^{\star}$ and $\varphi^{\star}$ will display a sharp discontinuity at the zone interfaces due to the thresholding of the corresponding mean level-set function. In contrast, the posterior mean from samples (eq. 30), will not show such a clear delineation because of the uncertainty associated with the location of these interfaces. To quantify this uncertainty, however, we compute the sample standard deviation (STD) from the posterior samples in eq. (30).

Additional measures of the uncertainty in the geometry of the zones are given by zonal probabilities, namely the probability that, under the posterior, any given point **x** belongs to one of the zones. The computations of zone probabilities are described below for the case of the DC resistivity and analogously for the phase angle. Given an EKI-converged posterior parameter sample

$$u_{\rho}^{(j)}(\mathbf{x}) = \left(\rho_1^{(j)}, \rho_2^{(j)}, \rho_3^{(j)}, \rho_4^{(j)}, \left\{L_{1,\alpha}^{(j)}, L_{2,\alpha}^{(j)}, \omega_{\alpha}^{(j)}(\mathbf{x})\right\}_{\alpha=1}^{2}\right)$$

we use the corresponding level-set functions $\xi_{\alpha}^{(j)} = \mathcal{P}^{GRF}(L_{1,\alpha}^{(j)}, L_{2,\alpha}^{(j)}, \omega_{\alpha}^{(j)})$ (computed via solving eq. 13). Then, from eq. (11), zone 1 probability can be defined via:

$$P_1(\mathbf{x}) = \mathbb{P}(\xi_1(\mathbf{x}) \leq \alpha_1, \xi_2(\mathbf{x}) \leq \alpha_2) = \int_{-\infty}^{\alpha_1} \int_{-\infty}^{\alpha_2} \pi_{\xi_1, \xi_2}(y_1, y_2) dy, \tag{31}$$

where $\pi_{\xi_1, \xi_2}(y_1, y_2)$ denotes the posterior density of the random variable $(\xi_1(\mathbf{x}), \xi_2(\mathbf{x}))$ (for the fixed **x**). It is not difficult to see that $P_1(\mathbf{x})$ can be approximated from the ensemble of posterior level-set functions as follows:

$$P_1(\mathbf{x}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{I}_R(y_1, y_2) \pi_{\xi}(y_1, y_2) dy_1 \, dy_2 \approx \frac{1}{J} \sum_{j=1}^{J} \mathbb{I}_R(\xi_1^{(j)}(\mathbf{x}), \xi_2^{(j)}(\mathbf{x})), \tag{32}$$

where $R = (-\infty, \alpha_1] \times (-\infty, \alpha_2]$ and $\mathbb{I}_R(y_1, y_2) = 1$ if $(y_1, y_2) \in R$ while $\mathbb{I}_R(y_1, y_2) = 0$ otherwise). We use analogous definitions for probabilities of zones 2, 3 and 4, as well as the zone probabilities for phase angle.

It is important to note that the posterior zonal probabilities defined in eq. (32) depend on our choice of thresholds $\alpha_1$ and $\alpha_2$. As discussed in Section 2.3.2, one can choose these thresholds so that prior zonal probabilities take specific values that we select according to our prior knowledge, if available, of the likelihood of each zone. In the example given in Section 2.3.2, we noted that $\alpha_1 = \alpha_2 = 0$ (the choice for our 4-zone formulation) yields constant and equal prior zonal probability for each zone, but a different choice of $\alpha_1$ and $\alpha_2$ may assign higher probability to one of the zones under the prior. Hence, the interpretation of posterior zonal probabilities should be made within the context of our selection of prior zonal probabilities. This is particularly important when we compute zonal
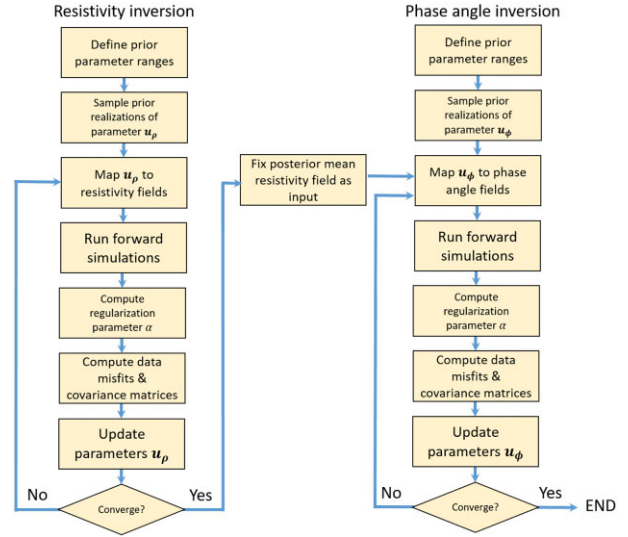


**Figure 1.** The flowchart of our sequential EKI algorithm for IP data inversion. Note that the parameters in the resistivity and phase angle inversions are independent from each other.

probabilities for the single level-set parametrization (for 3 zones) described in Appendix B and first used in Tso *et al.* (2021). For that case, since the prior of the level-set function has zero mean function, a choice of, for example $\alpha_1 = -\alpha_2$ with $\alpha_2 > 0$ will result in prior zonal probabilities equal to one (for all **x**) for zone 2 while zero for zone 1 and zone 3, which represents a prior field that everywhere (including those outside the imaging domain) belongs to the background zone (typically with the intermediate resistivity or phase angle ranges) with a probability of one.

## 2.5 Implementation details

As stated earlier, in our approach to EKI inversion of IP data we first use EKI to estimate the DC resistivity while keeping the phase angle fixed. After convergence, we fix the DC resistivity as the mean of the posterior resistivity obtained in the previous step and run EKI to infer phase angle only. See Fig. 1 for a flowchart that summarizes our IP inversion algorithm.

An alternative implementation of IP EKI inversion would be to jointly invert resistivity and phase angle. In other words, use EKI to iteratively update both properties simultaneously. In the joint approach, however, the number of unknown parameters doubles which results in substantially slower convergence of the EKI scheme. This sequential approach is more flexible than a joint inversion approach as it allows potential differences in the patterns of the resistivity and phase angle fields.

As in the EKI implementation of Tso *et al.* (2021) for ERT data, we use a forward modelling grid that extends laterally several times the dimension of the IP imaging area to simulate an infinite earth in field studies. For convenience, here we discretize the parameter grid used for inversion as a grid consisting of squares covering the entirety of the imaging area. At each iteration, the parameter grid is interpolated to the forward modelling grid to obtain simulated IP data. For this paper, we have implemented the EKI method in MATLAB$^{\textcircled{R}}$).

One of the zones is assigned to be the background zones such that its mean value is assigned for cells that are outside the parameter estimation grid. This is necessary to satisfy the infinite earth assumption in the modelling of most field geophysical problems.

In all the examples, we compare the inversion results from EKI to SCI (see e.g. Binley & Slater 2020). cR2 (https://www.es.lancs.ac.uk/people/amb/Freeware/cR2/cR2.htm) is used for forward modelling runs in EKI and SCI. Note that while the EKI runs invert resistivity and phase angles sequentially, the SCI runs inverts jointly the real and imaginary conductivities by solving the IP problem in the complex number domain. Where suitable, we also compare the results of using SCI with regularization disconnect.

## 3 EXAMPLE APPLICATIONS

We report the application of EKI for IP data in four example applications. In all cases, unless otherwise specified, we use the single level-set formulation outlined in Appendix B. We report the convergence statistics in Table 1. We also report the prior and posterior zonal resistivity/phase angle histograms and show the plots of a few example samples for each example case in the Supporting Information. Note that zone numbering for EKI, in general, is arbitrary. Since we invert DC resistivity and IP data sequentially, they have independent zone memberships. When results are reported in the following, zones are arranged in descending resistivity/ (negative) phase angle order. For brevity, we do not report results on the posterior estimates on the hyperparameters (i.e. length scales) but we emphasize that these were used to generate the plots of the posterior estimates for the spatial fields (resistivity and phase-angle). In addition, the plots of some of the samples shown in the Supporting Information show that our estimates of hyperparameters lead to physically consistent estimates of the geophysical properties that we infer.

### 3.1 Synthetic example (inclusions)

#### 3.1.1 Model setup

The synthetic survey contains 298 dipole–dipole measurements collected by 24 surface electrodes at 0.5 m separations. The background resistivity and phase angle for the true model (Figs 2a and b) are 100 Ωm and −10 mrad, respectively and a rectangular anomaly (1 m × 1.7 m) for both resistivity (10 Ωm) and phase angle (−15 mrad) near the surface. An additional rectangular phase angle anomaly (2 m × 1 m) of −20 mrad is at the lower right of the domain. Random noise of 2 per cent for resistivity and 2.0 mrad for phase angle (typical field data noise levels) are added to the synthetic data. The same level of measurement errors is assumed in the inversions.

#### 3.1.2 EKI IP inversion

We first perform IP inversion with EKI using a 2 and 3 zone formulation for resistivity and phase angle, respectively. The choice of 2 or 3 zones are based on the conceptualization of the site, as done in other geometry-based inversions (e.g. Bijani *et al.* 2017). Priors for the zonal resistivity ranges are as follows:

$$\mathbb{P}(\rho_1) = \mathcal{U}\big[80\,\Omega\text{m}, 300\,\Omega\text{m}\big],$$
$$\mathbb{P}(\rho_2) = \mathcal{U}\big[0.5\,\Omega\text{m}, 50\,\Omega\text{m}\big] \tag{33}$$

while those for phase angles are

$$\mathbb{P}(\varphi_1) = \mathcal{U}\big[-12\,\text{mrad}, -5\,\text{mrad}\big],$$
$$\mathbb{P}(\varphi_2) = \mathcal{U}\big[-17\,\text{mrad}, -13\,\text{mrad}\big],$$
$$\mathbb{P}(\varphi_3) = \mathcal{U}\big[-25\,\text{mrad}, -19\,\text{mrad}\big] \tag{34}$$

In practice, the ranges for these priors are selected based on *a priori* knowledge of the site. We reiterate that in the absence of prior knowledge we can assign the same prior to all zonal resistivities/phase angles but defined on a wider interval to account for large prior uncertainty. Conversely, we could choose different priors that assign higher probability to specific values if these were informed by prior knowledge.

Note that since the phase angle inversion is performed after the DC resistivity inversion, they can have different numbers of zones. The EKI results are compared with conventional SCI (Figs 2c and d). SCI recovered smooth targets, especially for phase angle; the lower target (−20 mrad) is not recovered at all. The posterior zonal resistivity values computed by EKI (from zone 1 to 2) are 99.39 and 5.06 Ωm; while those for phase angles (from zone 1 to 3) are −20.44, −14.15 and −6.15 mrad. The posterior mean for the resistivity field and phase angle computed from the samples given by eq. (30) are shown in Figs 2(e) and (f), while the corresponding estimates obtained via mapping the posterior mean of input parameters into resistivity and phase angle (i.e. via eq. 29) are displayed in Figs 2(g) and (h). We note that the target resistivity anomaly from the truth is at the correct location and the width and top boundary are almost exactly correct, although it appears to be not as deep as the true one. The recovered phase angle targets have geometries that roughly agree with the true ones, however, the upper left zone is much smaller while the lower right inclusion (−20 mrad) is much larger. In particular, the lower right target is less well resolved because of the sensitivity pattern of the surface measurements.

Uncertainty analysis of the posterior distribution is provided by zone 2 probability (Figs 2i–j) and standard deviation maps. For resistivity, the probability for (correctly) belonging to zone 1 and 2 is high. The most uncertain zone membership is observed near the lower boundary of the inclusion. In addition, slightly higher uncertainty is also observed near the lower left and lower right corners. For phase angle, note that a 3-zone formulation is used, with zone 2 being the background. High zonal probabilities (i.e. zone 2 probability close to 0 or 1) are observed in the top 5 m, correctly associating with the background and inclusion zone. Elsewhere there are regions where zones are poorly resolved (i.e. zone 2 probability around 0.5). These observations are reflected in the standard deviation maps (Figs 2k and l). For resistivity, standard deviation (*STD*) values are low everywhere except around the lower parts of the estimated inclusion. For phase angle, however, the *STD* pattern is more complex. With the exception of the top 2–3 m and the part of the lower right region, *STD* is high everywhere. The lower left corner region has *STD* values of up to 5 mrad. Such results highlight the estimation of phase angles is more challenging than that of resistivities due to their low signal-to-noise ratio, which is illustrated well by the greater variability in the phase angle posterior samples (see Supporting Information).

### 3.2 Synthetic example (3 layers)

#### 3.2.1 Model setup

The true model in this example is conceptualized as a 3-layer system with inclinations and pinch-out. All layers are assumed to extend laterally. From top to bottom, the three zones have resistivity (and phase angle) values of 250 Ωm (−10 mrad), 10 Ωm (−20 mrad) and 100 Ωm (−15 mrad) as shown in Figs 3(a) and (b). Random noise of 2 per cent for resistivity and 2.0 mrad for phase angle are

**Table 1.** Convergence performance of all EKI test cases presented in this paper in terms of the squared norm of the average residual, $\mathcal{D}_m$ from eq. (26). We report the values for the prior $m = 0$ and converged $m = q + 1$ ensemble. 300 forward model runs were used in all test cases in each iteration.

| | Resistivity | | | Phase angle | | |
|---|---|---|---|---|---|---|
| | $\mathcal{D}_0$ | $\mathcal{D}_{q+1}$ | Iterations $q$ | $\mathcal{D}_0$ | $\mathcal{D}_{q+1}$ | Iterations $q$ |
| Fig. 2 | 242.37 | 0.97 | 15 | 47.33 | 18.15 | 5 |
| Fig. 3 | 576.78 | 1.02 | 13 | 3.24 | 1.06 | 4 |
| Fig. 4 | 711.93 | 1.08 | 15 | 3.14 | 1.04 | 5 |
| Fig. 5 | 857.28 | 1.05 | 17 | 8.26 | 0.96 | 5 |
| Fig. 6 | 1719.21 | 1.22 | 16 | 8.61 | 1.39 | 4 |
| Fig. 7 | 1724.21 | 1.50 | 22 | 8.26 | 1.62 | 4 |
| Fig. 9 | $2.5 \times 10^4$ | 21.82 | 54 | 2.33 | 1.31 | 3 |
| Fig. 10 | $7.7 \times 10^4$ | 2.80 | 27 | 8.72 | 1.10 | 4 |
| Fig. 11 | $2.8 \times 10^4$ | 3.73 | 31 | 6.34 | 2.43 | 5 |
| Fig. 12 | $3.1 \times 10^5$ | 9.99 | 53 | 106.24 | 1.25 | 7 |

added to the synthetic data. The same level of measurement errors is assumed in the inversions.

### 3.2.2 EKI IP inversion

SCI provides a satisfactory estimate of the top layer; however, it shows no differentiation of the bottom two zones (Fig. 3c). They are depicted as a large smooth low-resistivity anomaly. For EKI, the priors for zonal resistivity values are

$$\mathbb{P}(\rho_1) = \mathcal{U}\big[0.5\ \Omega\text{m}, 50\ \Omega\text{m}\big],$$
$$\mathbb{P}(\rho_2) = \mathcal{U}\big[80\ \Omega\text{m}, 180\ \Omega\text{m}\big],$$
$$\mathbb{P}(\rho_3) = \mathcal{U}\big[200\ \Omega\text{m}, 300\ \Omega\text{m}\big]$$

while those for phase angles are the same as those in eq. (34). The estimated zonal resistivity values computed by EKI (from zone 1 to 3) are 249.5, 246.9 and 5.5 Ωm. Posterior mean resistivity and phase angle fields are shown in (Figs 3e and f). Resistivity and phase angle computed from mapping posterior mean for input parameters can be found in (Figs 3g and h). We notice that the topmost and bottommost zones are estimated as almost identical. Taking that into consideration, EKI recovers the geometry of the middle pinch-out zone very well.

Because of the close proximity in values for zone 2 and 3, the zonal probability map shows very low zone 2 probability in the middle pinch-out zone, but only a very small fraction of the area with a probability close to 1. Again, this is because the zone 2 and 3 resistivity values are too close to each other. For the *STD* map (Fig. 3i), zone 1 has very low uncertainty, while that for the bottommost zone and the boundary of the middle pinch-out zone is quite low too. The *STD* at the lower right corner is relatively high, highlighting it is the most difficult region to resolve.

SCI returns a smooth phase angle image with the upper half of the domain having less negative phase angles than the lower half (Fig. 3d). For EKI, three zones are clearly recovered, with zonal phase angles (from zone 1 to 3) equalling −19.6 mrad (pinch-out), −13.0 and −6.7 mrad (top), respectively (Fig. 3h). The top zone is recovered (although not as deep as the true one and the phase angle is less negative than the true one), while the pinch-out zone is wider and extends deeper than the estimated value. It is also noteworthy that there is a small strip of intermediate phase angle between the top and pinch-out zone. This is partly attributed to the implicit assumption that by using a single level set for three zones, the transition from the minimum to the maximum zone must include the intermediate zone.

Notably, Fig. 3(j) shows high zone 2 probability only on this strip of transitional values, while everywhere else has a low zone 2 probability. Together with the mean map, it shows the lower left corner (or most of the bottom zone in the domain) is not well resolved. For the *STD* map (Fig. 3l), the uncertainties in the top two layers are low, while that for the lower layer is high.

### 3.2.3 Effect of different level set formulations

### 3.2.4 Two level sets for three zones

An issue we observe in Fig. 3 is that since we are using a single level set to represent three zones, the zone with the lowest value cannot 'jump' to the zone with the highest value (e.g. Fig. 4h, and samples in the Supporting Information). In this particular example case, this is problematic since the pinch-out represents an intersection of the 3 zones. We circumvent this issue by repeating the inversion using the parametrization in terms of the two-level set functions that we introduced in Section 2.3.1 with three zones. The results shown in (Fig. 4) are much improved and the three distinctive zones for both DC resistivity and phase angle are better recovered. A useful feature that can be observed in Fig. 4(a) is that the top layer is estimated as an almost discrete value, while the second and third layers are relatively smooth. This shows that there is lower uncertainty on the value of the top layer. The posterior zonal resistivities obtained from the mean level sets returned (from zone 1 to 3) are 249.8, 187.0 and 7.3 Ωm; while those for phase angles are −20.1, −13.8 and −6.4 mrad.

### 3.2.5 Four zones formulation

Another issue that is worth investigating is the effect of incorrectly specifying the number of zones. In Fig. 5, we repeat EKI for this example with four zones, that is we allow the four values for resistivity (and the four values for phase angle) to be inferred. The priors for the fourth zone are specified as non-informative priors, spanning the range of the other three zones. The posterior zonal resistivities obtained from the mean level sets (c, d) returned (from zone 1 to 4) are 249.4, 123.4, 8.1 and 1.1 Ωm; while those for phase angles are −20.4, −14.0, −6.7 and −6.1 mrad. For resistivity, the middle layer does not extend as deep as the true field. For phase angles, the top zone is largely being identified correctly and the posterior is shown as four zones. However, it did not identify the bottommost, less negative phase angle zone at all. Instead, the middle zone extends
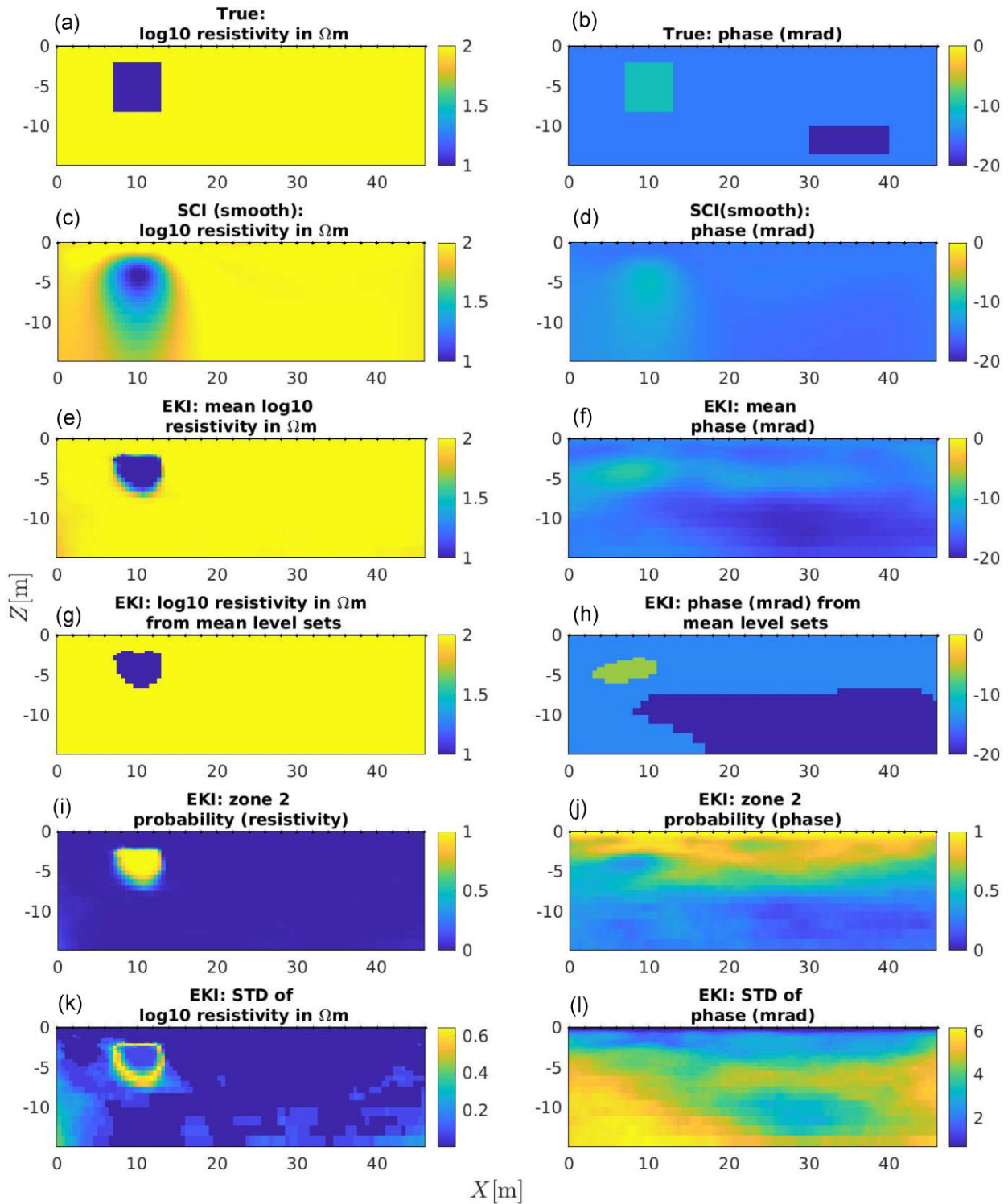
**Figure 2.** (a, b) True resistivity and phase angle models for the 2-D surface example, which comprises of an inclusion of identical geometry in both models and an additional inclusion in the phase angle model. (c, d) The resistivity model estimated by smoothness-constrained inversion. (e, f) The mean resistivity and phase angle model estimated by EKI across samples. (g, h) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (i, j) The prior estimated probability of zone 2 (i.e. not the background) (k, l) The posterior standard deviation for resistivity and phase angle.

to the bottom of the domain. The mean results (a, b) do, however, correctly show some indication that the resistivity and phase angle values in the lower left part of the domain lie somewhere between those in the top two zones.

*3.2.6 Effect of non-informative priors*

While in many cases informative first guesses of priors for each zone can be provided, it is worth considering the effects of using non-informative priors that are identical for each zone. Therefore, we
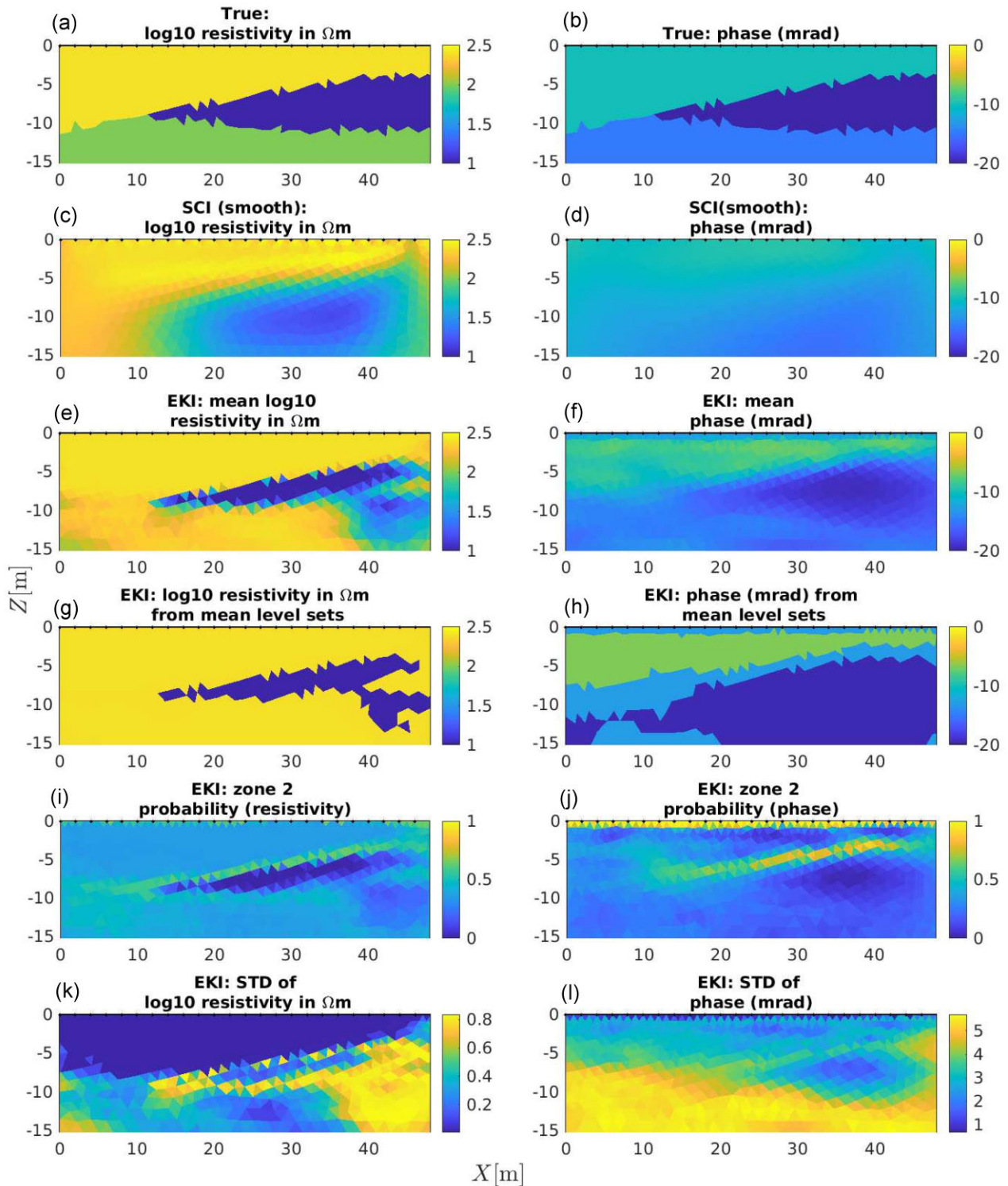
**Figure 3.** (a, b) True resistivity and phase angle models for the 2-D 3-layer example, comprises of three collocated non-horizontal layers. (c, d) The resistivity model estimated by smoothness-constrained inversion. (e, f) The mean resistivity and phase angle model estimated by EKI across samples. (g, h) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (i, j) The prior estimated probability of zone 2 (i.e. not the background) (k, l) The posterior standard deviation for resistivity and phase angle.

have repeated the runs reported in Figs 4 and 5 using the following priors for all zones:

$$\mathbb{P}(\rho) = \mathcal{U}[0.5 \, \Omega\text{m}, 300 \, \Omega\text{m}],$$

$$\mathbb{P}(\varphi_1) = \mathcal{U}[-25 \, \text{mrad}, -5 \, \text{mrad}],$$

The switch to non-informative priors certainly have caused has led to some deterioration of EKI results (Figs 6 and 7). For resistivity, the results are no worse than that in Fig. 3, generally identifying a low resistivity pinch-out in a resistive background. In both cases, the
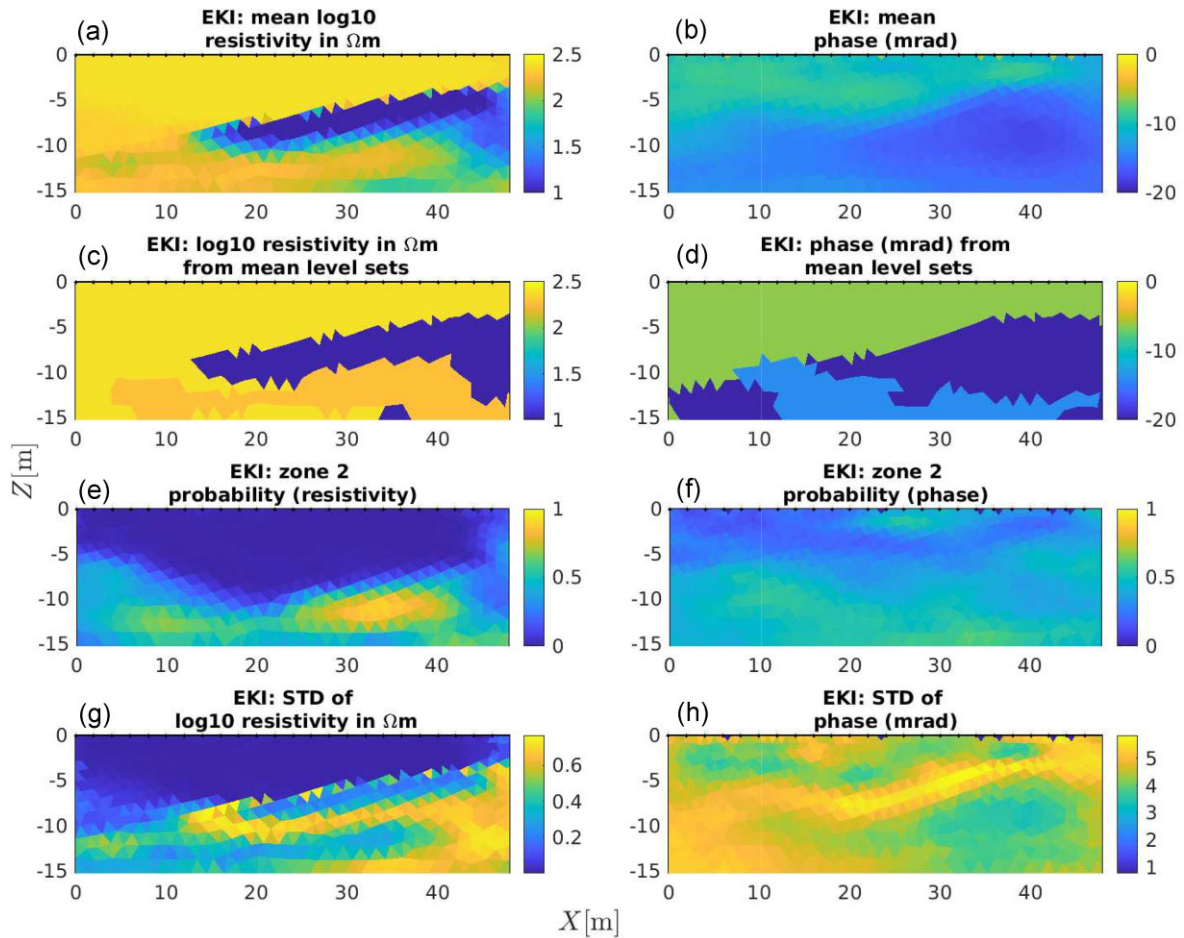
**Figure 4.** A repeat of Fig. 3 but with two level sets functions, thus allowing jumps from the lowest to the highest-value zones in a three-zone formulation. (a, b) The mean resistivity and phase angle model estimated by EKI across samples. (c, d) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (e, f) The prior estimated probability of zone 2 (i.e. not the background) (g, h) The posterior standard deviation for resistivity and phase angle.

top layer is clearly and accurately identified as a discrete, resistive zone. In Fig. 6, the shape of the bottom of the pinch-out was not clearly identifiable but Fig. 6(a) indicates the bottom zone is likely to be less resistive than the top zone. The posterior zonal resistivities obtained from the mean level sets returned (from zone 1 to 3) are 250.4, 241.8 and 17.4 Ωm. Fig. 7 shows clearly the shape of the pinch-out, although the middle zone extends deeper than the true geometry. The posterior zonal resistivities obtained from the mean level sets returned (from zone 1 to 4) are 249.4, 12.8, 12.8 and 8.0 Ωm.

For phase angles, in both cases all zones estimated by the mean level sets are close to −10.0 mrad (i.e. −12.18, −9.56, −7.68 mrad for Fig. 6, and −11.84, −10.66, −9.51 and −8.67 mrad for Fig. 7 ), thus not returning very useful results for this problem. This is likely to be caused by more uncertain DC resistivity as input, as well as the greater dependence on informative priors of phase angles due to their lower signal-to-noise ratio.

## 3.3 Pow Beck subcatchment

### 3.3.1 Site and data description

The Pow Beck subcatchment is located within the Eden Valley in northwest England (south of Carlisle). Mejus (2014) used multiple

geophysical methods for hydrogeological characterization. Among them, four surface IP surveys were conducted at the site. In this paper, we focus on the data for the survey IP09 of Mejus (2014), which consists of 48 electrodes at 2 m separation and 338 pairs of transfer resistances and phase angles after data filtering. After characterization of data errors, data errors of 0.5 per cent are assumed for resistivity and while that for phase angle is based on a curve-fitting, with the resultant formula being $0.0343\varphi^2 - 0.1397\varphi + 1.0$ mrad.

The geology at the site is dominated by Quaternary superficial deposits and glacial till dominates the Quaternary cover. Borehole records (1995–2006) from the British Geological Survey (BGS) with borehole log records showing that the thickness of superficial deposits varies ranging from 0 to 25 m. The Quaternary cover overlies a bedrock belonging to the Sherwood Sandstone group, which is comprised of cemented consolidated sediments. A percussive direct penetration test was also conducted at one location along the IP survey transect to reveal the vertical variation in material properties (see Fig. 8a). We also include the classification of lithological units based on field-based electrical properties by Mejus (2014) (Fig. 8b). Mejus (2014) also measured IP properties on laboratory samples of lithologies at the site and noted a rank in polarization (given by the imaginary component of electrical conductivity) of overburden (lowest); clayey till; sandstone (highest).
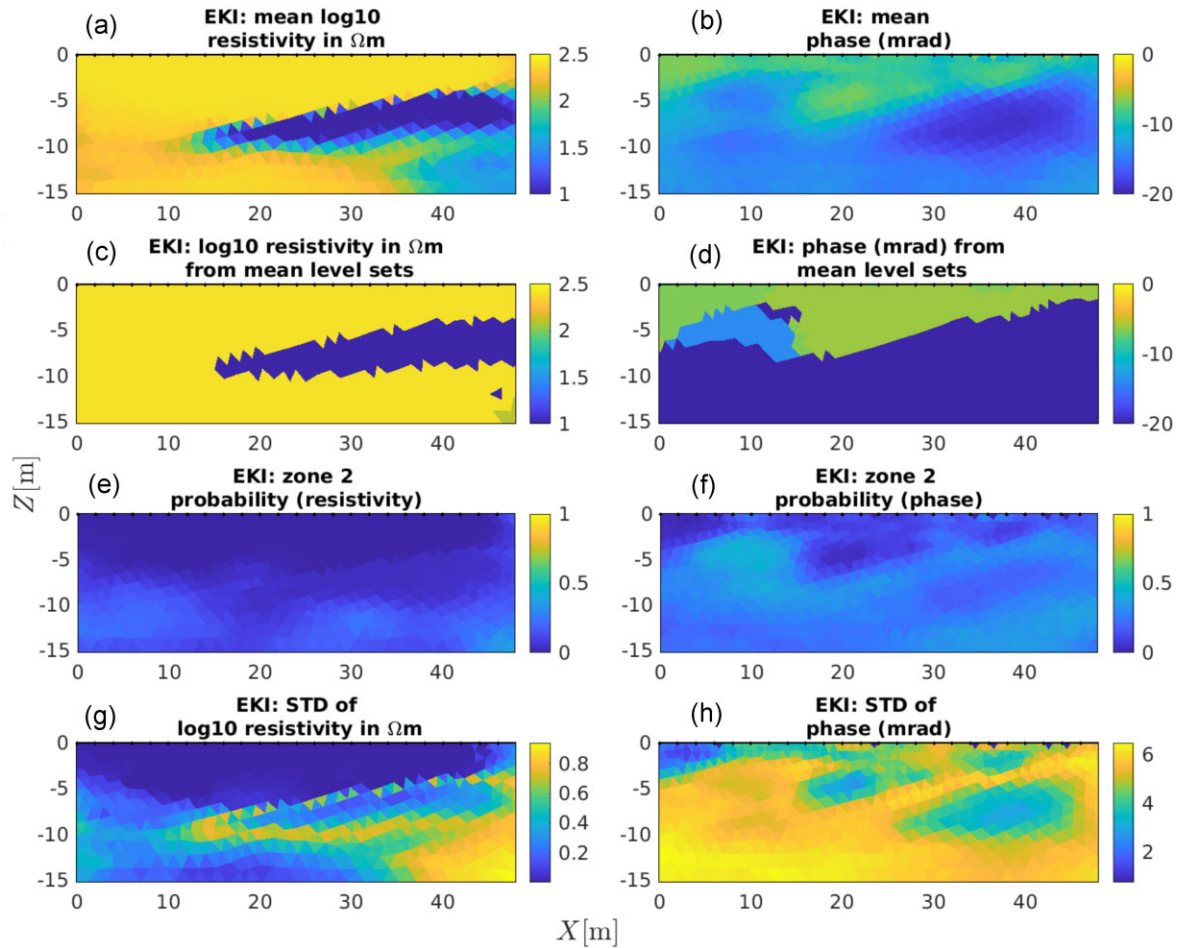
**Figure 5.** A repeat of Fig. 4 but using a four-zone formulation. (a, b) The mean resistivity and phase angle model estimated by EKI across samples. (c, d) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (e, f) The prior estimated probability of zone 2 (g, h) The posterior standard deviation for resistivity and phase angle.

### 3.3.2 EKI IP inversion

We perform IP inversion with EKI using a 3 zone formulation for both resistivity and IP and compare its results with conventional SCI (Fig. 9). Prior zonal values are

$$\mathbb{P}(\rho_1) = \mathcal{U}[10\,\Omega\text{m}, 100\,\Omega\text{m}],$$
$$\mathbb{P}(\rho_2) = \mathcal{U}[150\,\Omega\text{m}, 250\,\Omega\text{m}],$$
$$\mathbb{P}(\rho_3) = \mathcal{U}[250\,\Omega\text{m}, 1000\,\Omega\text{m}] \tag{35}$$

while those for phase angles are

$$\mathbb{P}(\varphi_1) = \mathcal{U}[-2\,\text{mrad}, 0\,\text{mrad}],$$
$$\mathbb{P}(\varphi_2) = \mathcal{U}[-4.5\,\text{mrad}, -2.5\,\text{mrad}],$$
$$\mathbb{P}(\varphi_3) = \mathcal{U}[-15\,\text{mrad}, -5\,\text{mrad}] \tag{36}$$

Fig. 9 shows the IP inversion results. The SCI result shows a relatively resistive surface soil layer overlaying a fairly conductive zone extended to a depth of about 10 m, which overlies a slightly more resistive background. The mean estimate of EKI, while also showing a resistive upper layer, suggests that the underlying conductive layer extends to a greater depth and is variable horizontally. This may be a result of greater depth resolution in EKI than SCI, as previously reported in Tso *et al.* (2021). It is, however, important to emphasize here that our EKI approach benefits from the 3-zone parametrization (via the level-set function) and the corresponding

priors of resistivity on each zone. Since this prior information is not encoded in SCI, it comes as expected that the resulting overly smooth field has limited depth resolution.

For phase angle, the SCI results show an area of less negative phase angle than the background above a depth of 10 m, which is in agreement with the EKI results. However, this zone is more laterally extensive in the latter. The posterior zonal resistivities obtained from the mean level sets returned by EKI (from zone 1 to 3) are 344.3, 275.1 and 55.9 $\Omega$m; while those for phase angles are $-8.7$, $-3.7$ and $-0.3$ mrad. The lack of DC resistivity contrast in the EKI results is almost certainly due to a lack of contrast in this property between the superficial sediments and the upper zone of the sandstone. However, the phase angle does show a contrast, suggesting a variation in electrical capacitive properties, which illustrates the potential value of IP in this example.

Uncertainty analysis of the posterior distribution is provided by zonal probabilities and standard deviation maps. Zonal probabilities in most of the domain are near 0 or 1, which provide us an indication of how likely (under the posterior) it is for a point $\mathbf{x}$ to belong to each zone. It is worth reiterating that zonal probabilities depend on the threshold $\alpha_1$ and $\alpha_2$ in the level-set function parametrization. For the 3-zone formulation that we use here (see Appendix B) the values for these thresholds are $\alpha_1 = -0.1$ and $\alpha_2 = 0.1$ and, thus, one should expect that, under the prior, all points will likely to belong to zone 2 (since the prior mean of the level-set is zero). Given the
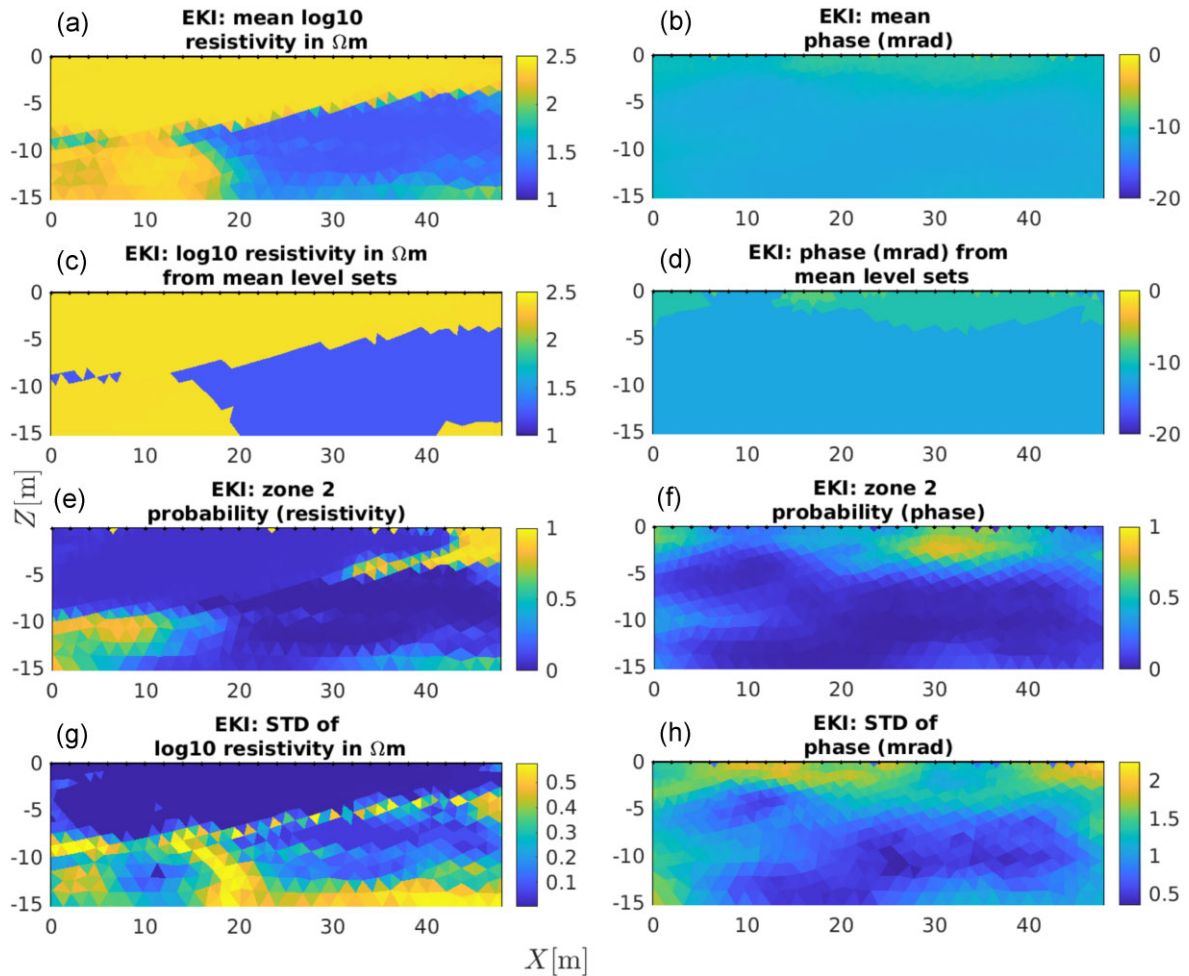
**Figure 6.** A repeat of Fig. 4 but with non-informative priors. (a, b) The mean resistivity and phase angle model estimated by EKI across samples. (c, d) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (e, f) The prior estimated probability of zone 2 (g, h) The posterior standard deviation for resistivity and phase angle.

low depth-resolution intrinsic to ERT, it comes as no surprise that the deeper part of the domain either remains to be zone 2 and zone 3 (given the proximity in resistivity values between zone 2 and with zone 3). However, it is remarkable that zone 1 (the low resistivity region) which is unlikely under the prior, is present in the central region with high probability under the posterior.

Zonal probabilities are also useful to quantity uncertain near the transition between zones, which in this case is quite noticeable at around 10 m depth for phase angle. The standard deviation for resistivity is low, with higher values observed near zonal boundaries. The standard deviation for phase angle are quite high in most areas since the range of estimated values are small. The EKI results agree well with percussive penetration test results (Fig. 8), that is a clear distinction in electrical properties (in particular the phase angle) at the lithological boundaries.

### 3.4 Permeable reactive barrier

A permeable reactive barrier (PRB) is an *in situ* technology for the remediation of a range of groundwater contaminants (chlorinated hydrocarbons, heavy metals, nitrate, etc.). The barrier is installed downgradient of the contaminant plume; *in situ* treatment is achieved by geochemical or biogeochemical reactions. Zero-valent

iron PRBs are the most common type of such technology, and are used for remediating chlorinated hydrocarbon contaminated groundwater. PRBs are typically installed using trenching, although more recently injection type installation has also been adopted. Ensuring adequate emplacement of the barrier at installation is important. Furthermore, given that any PRB must retain its enhanced permeability relative to the host aquifer, it is necessary to monitor the efficiency of the PRB over time to ensure satisfactory performance. Slater & Binley (2006, 2003) reported the use of electrical imaging to characterize the integrity and monitor the geochemical alteration of a zero-valent iron PRB over time at the US Department of Energy Kansas City Plant in Missouri, USA.

#### 3.4.1 Site and data description

Here, we re-invert the field data collected by Slater & Binley (2006). The cross-borehole ERT and IP survey was conducted using 63 electrodes distributed in three boreholes (see Fig. 6 for location of the electrodes). In total, 2223 quadrupoles are used after filtering. A 5 per cent data error is assumed for DC resistivity, while a 2.0 mrad data error is assumed for phase angle.

The site includes a conductive, L-shaped, zero-valent iron PRB. The extent of the PRB is assumed known since it is an engineered
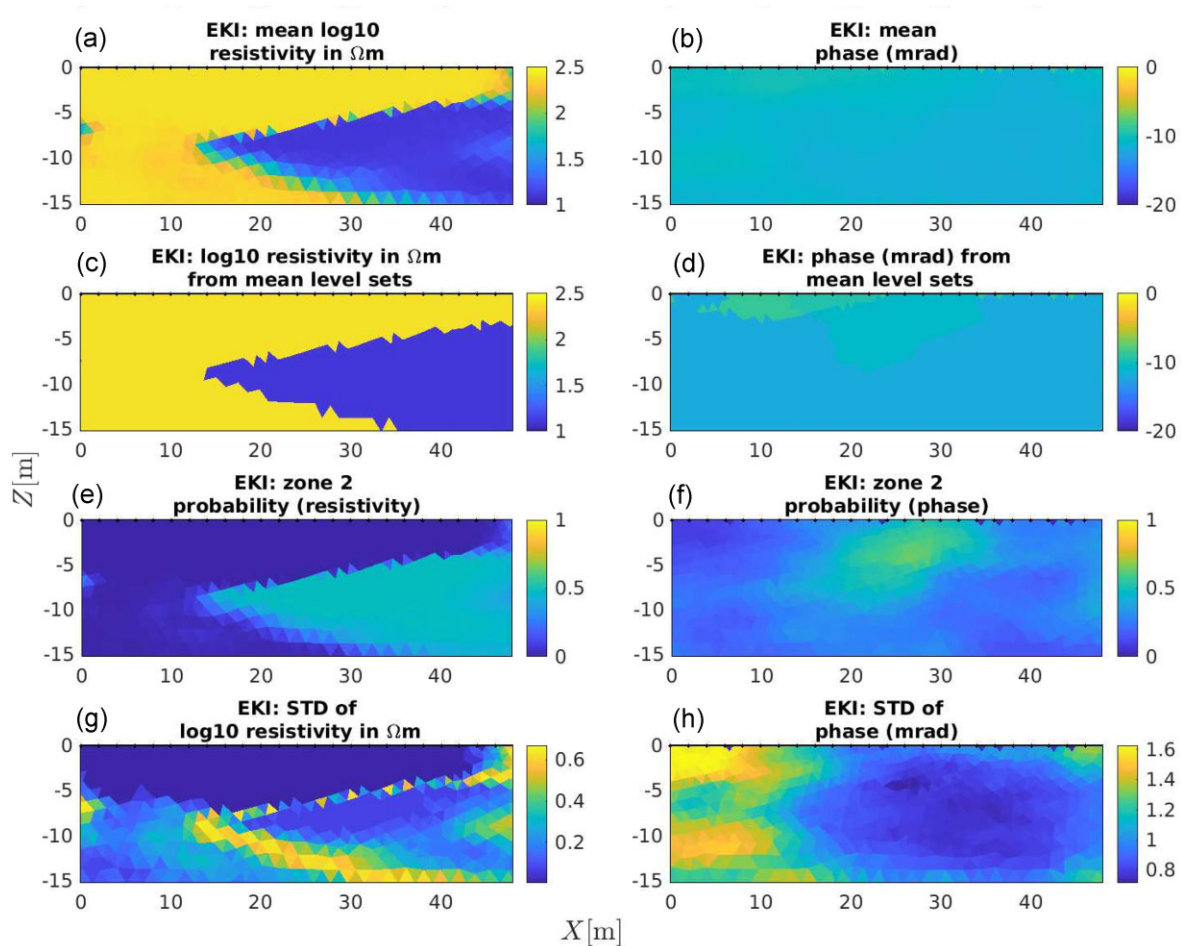
**Figure 7.** A repeat of Fig. 5 but using non-informative priors. (a, b) The mean resistivity and phase angle model estimated by EKI across samples. (c, d) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (e, f) The prior estimated probability of zone 2 (g, h) The posterior standard deviation for resistivity and phase angle.
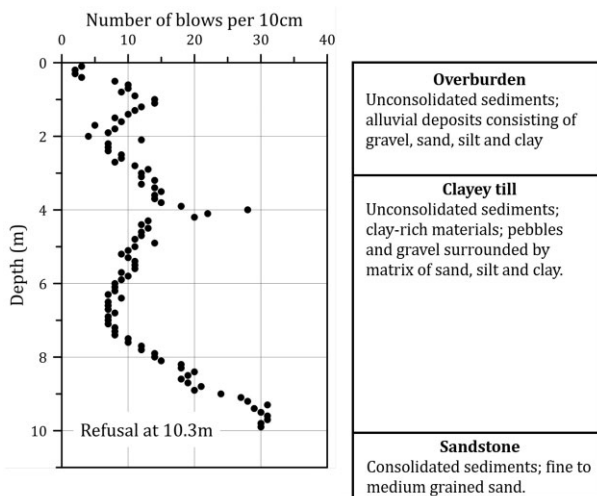


**Figure 8.** Direct penetration test (DPT) data from the Pow subcatchment taken at approximately $X = 42$ m along the IP survey line and a simplified profile of the lithological units. (source: Mejus 2014).

structure. In Slater & Binley (2006), the regularization at the boundary between the PRB and the background is disconnected, meaning

there is no smoothing applied across the boundary of the two materials. In contrast, for the EKI inversions reported below, we have not supplied any prior information on the boundary.

### 3.4.2 EKI IP inversion

For EKI runs in this section, priors for zonal resistivity are

$$\mathbb{P}(\rho_1) = \mathcal{U}[80\,\Omega\text{m}, 300\,\Omega\text{m}],$$
$$\mathbb{P}(\rho_2) = \mathcal{U}[0.5\,\Omega\text{m}, 50\,\Omega\text{m}] \tag{37}$$

while those for phase angles are

$$\mathbb{P}(\varphi_1) = \mathcal{U}[-7\,\text{mrad}, -1\,\text{mrad}],$$
$$\mathbb{P}(\varphi_2) = \mathcal{U}[-20\,\text{mrad}, -12\,\text{mrad}] \tag{38}$$

Our initial inversion shows that, unlike previous examples, setting homogeneous length scales for priors yield better results. Therefore, for all PRB examples we set the length scales in x-direction to be identical to that in the z-direction after running eq. (24).

To assess the utility of EKI, we first conduct inversions using a synthetic model, with a background resistivity of 100 $\Omega$m and phase angle of $-2.5$ mrad; as well as a PRB resistivity and phase angle, respectively, of 0.1 $\Omega$m and $-15$ mrad (Fig. 10 a-b). Random noise of the same amount as the assumed data error levels are added to the synthetic data. In Slater & Binley (2006), the inversion is
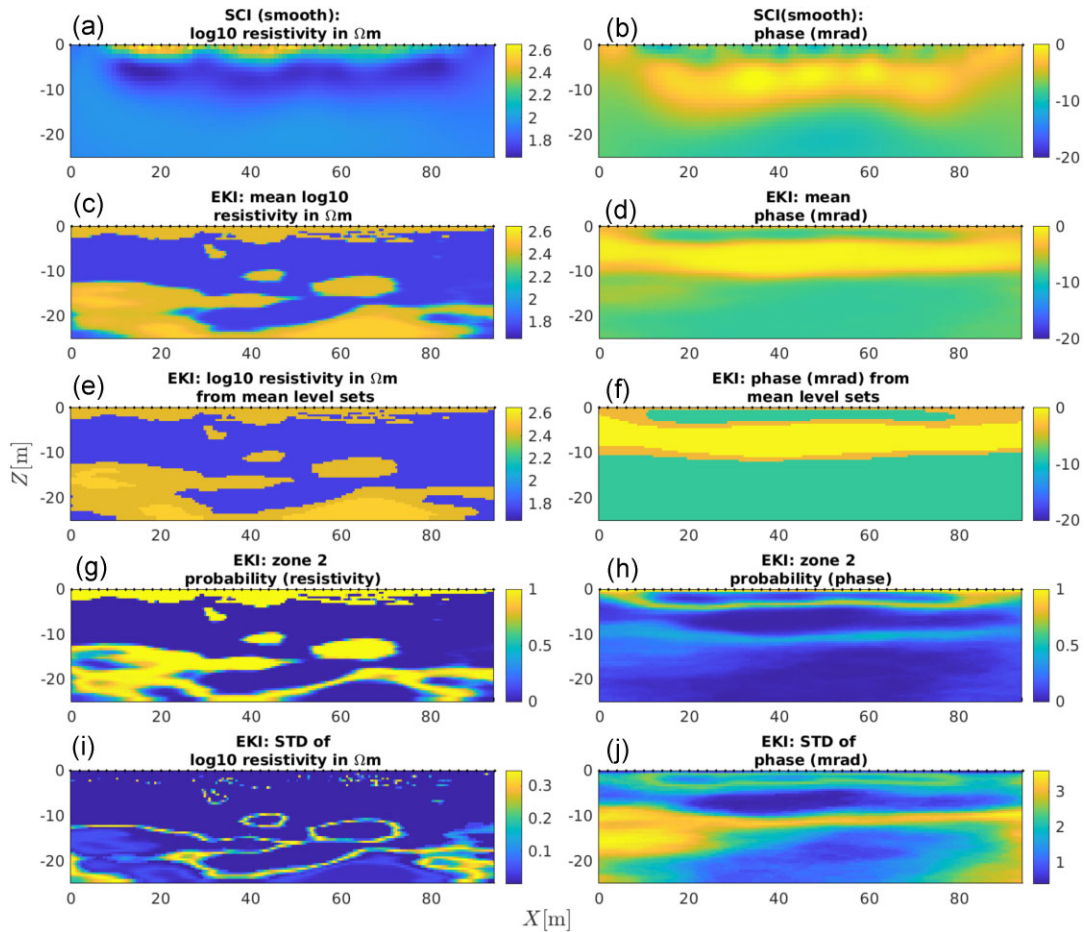
**Figure 9.** Pow Beck subcatchment: (a, b) The resistivity model estimated by smoothness-constrained inversion. (c, d) The mean resistivity and phase angle model estimated by EKI across samples. (e, f) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (g, h) The posterior Zone 2 probabilities for resistivity and phase angle. (i, j) The posterior standard deviation.

performed using SCI with regularization disconnect at the known boundary of the PRB. Therefore, we consider SCI results with and without the regularization disconnect. The SCI IP inversion (Figs 10c and d) recovers an 'L-shape' conductive anomaly that is consistent with the true PRB position, however, the phase angle recovery of the barrier region is weak (Fig. 10d). In contrast, SCI with regularization disconnect return the 'L-shape' zone perfectly (Figs 10e and f).

The posterior zonal resistivity values recovered by EKI from the mean level sets are 22.61 Ωm (zone 2) and 0.25 Ωm (zone 1) (Fig. 10i). The estimated PRB zone overlap with the true region very well, except its top part is slightly thinner. The estimated bottom extent of the PRB is lower the expected, and it includes a fine, discontinuous, vertical feature. However, these slight artefacts are unlikely to affect assessment of PRB integrity. The phase angle values recovered by EKI from the mean level sets are −2.49 (zone 2) and −13.5 (zone 1) mrad (Fig. 10j). As observed in the EKI field inversion, only the low part of the 'L-shaped', more negative phase angle feature associated with the PRB is recovered, but not the upper part. The zone 2 probability and *STD* maps also show the zonal uncertainty is low everywhere, except slightly higher along the estimated boundaries.

For the inversion of field data, the SCI results with and without regularization disconnect are shown in Figs 11(a)–(d). Without the disconnect approach, the inverted resistivity and phase angle

images does not resemble a PRB. The smooth images have high resistivities and highly negative phase angles in the middle of the model domain. With the disconnect applied, very high resistivities and highly negative phase angles are observed within the L-shaped boundary. However, smooth, higher resistivities and highly negative phase angles zones are still observed in the middle of the model domain. This may suggest the PRB extent may be slightly different than assumed, or the impact of non-Gaussian noise (which is most likely due to modelling errors not being fully taken into account).

The EKI results are reported in Figs 11(e)–(l). Without using any prior knowledge of the PRB boundaries, the returned resistivity patterns shows roughly the 'L' geometry, although the 'L' extends lower and its top part is wider than assumed. The recovered phase angle map shows a highly negative phase angle zone with a roughly round structure dipping to the right. Note that its left boundary is almost identical as assumed and the groundwater flow gradient is left to right. The zone 2 probability and *STD* maps also show the zonal uncertainty is low everywhere, except slightly higher along the estimated boundaries. The posterior zonal resistivity from the mean level sets (from zone 1 to 2) are 0.4 and 73.7 Ωm, while for phase angles they are (from zone 1 to 2) −21.0 and −2.5 mrad.

Based on our finding in this example, geoelectrical imaging with EKI may be a useful tool for probabilistic PRB integrity assessments. While SCI returns deterministic, overly smooth images that may limit the interpretation of PRB-related features, regularization
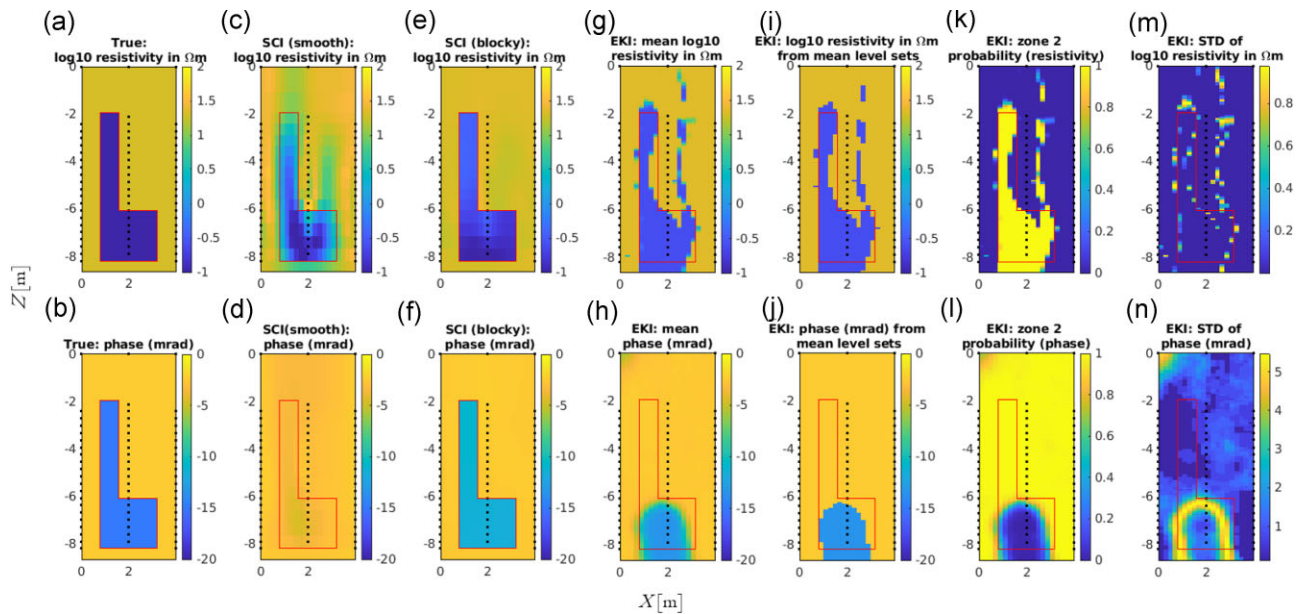
**Figure 10.** (a, b) True resistivity and phase angle models for the 2-D PRB example, which comprises of an L-shaped target that represents the PRB. (c, d) The resistivity and phase angle model estimated by smoothness-constrained inversion (without regularization disconnect). (e, f) The resistivity and phase angle model estimated by smoothness-constrained inversion (with regularization disconnect). (g, h) The mean resistivity and phase angle model estimated by EKI across samples. (i, j) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (k, l) The prior estimated probability of zone 2 (i.e. not the background) (m, n) The posterior standard deviation for resistivity and phase angle. The red lines denotes the true PRB boundary. The three vertical arrays of black dots indicate the borehole electrode positions.
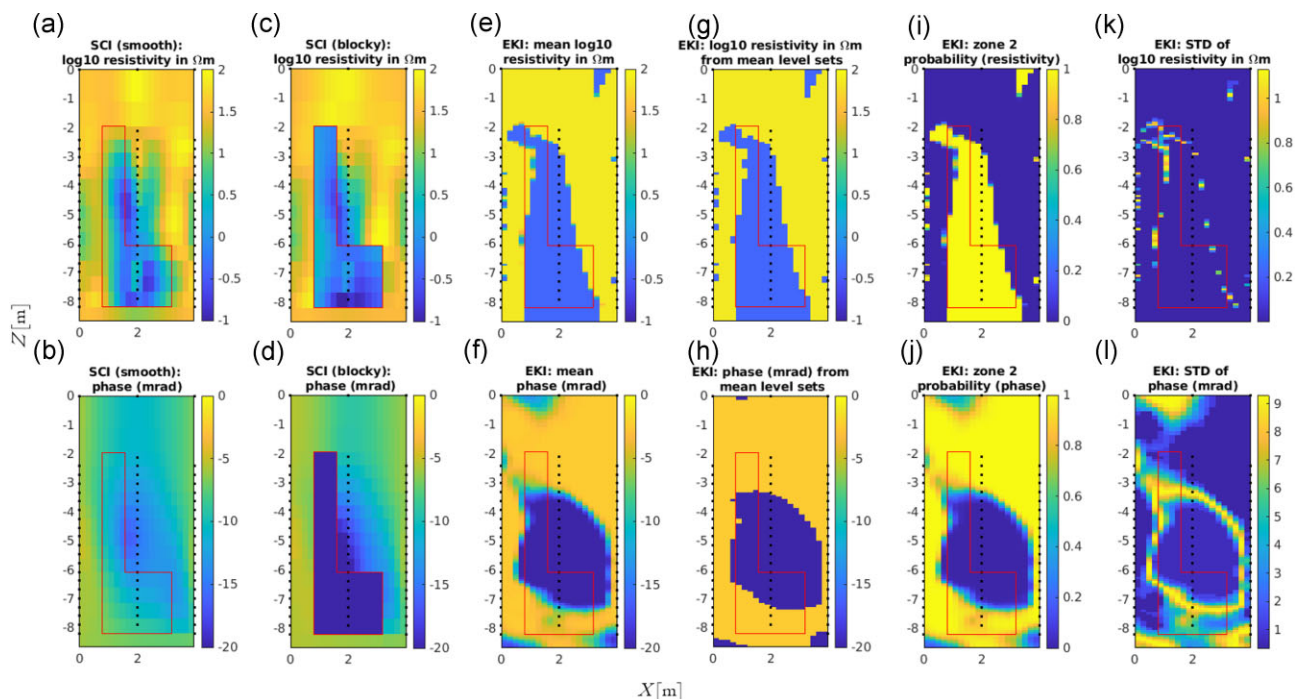


**Figure 11.** (a, b) The resistivity model estimated by smoothness-constrained inversion for the PRB field example (without regularization disconnect). (c, d) The resistivity model estimated by smoothness-constrained inversion for the PRB field example(with regularization disconnect). (e, f) The mean resistivity and phase angle model estimated by EKI across samples. (g, h) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (i, j) The prior estimated probability of zone 2 (i.e. not the background) (k, l) The posterior standard deviation for resistivity and phase angle. The red lines denotes the assumed true PRB boundary. The three vertical arrays of black dots indicate the borehole electrode positions.

disconnect relies on very strong assumptions on the location of the PRB. Such an assumption may not be desirable because the PRB may not be constructed exactly as planned, or may have experienced transformation over time. EKI provides an additional way to invert ERT and IP data and it relaxes this assumption and returns a map of mean zonal resistivity together with estimates of the posterior zonal probability that can be useful to monitor the shape and integrity of the PRB non-invasively over time. Although we recognize that

measures of uncertainty in terms of zonal probabilities are highly reliant on the choice of the level-set parametrization and the thresholds that we selected *a priori*, our work can be extended to infer more optimal thresholds from the ERT and IP data to better inform the uncertainty in the estimates that we produce.

### 3.4.3 Effect of data noise

We repeat the inversion of the synthetic PRB case by lowering the data noise levels to half of the original or more (Fig. 12). Specifically, the noise levels are 2 per cent for DC resistivity and 0.5 mrad for phase angle.

The number of iterations required to converge has about doubled. This is expected as data is assumed to be of higher quality and are given more weight. The DC resistivity inversion improves from Fig. 12. In particular, there is an increased number of correctly identified low-resistivity PRB pixels, and fewer misidentified ones outside the PRB. The EKI posterior resistivity from the mean level sets is estimated to be 29.33 and 0.28 $\Omega$m. Phase angle results also show improvements. While Fig. 10 phase angle estimates do not show the PRB to resemble an 'L' shape, here it shows some indication that the feature is wider at lower depths. The posterior zonal phase angle is $-14.45-2.51$ mrad. While there have been improvements in resolving features with lower data noise, the bottom of the 'L' shape PRB remains difficult to resolve clearly. The low resolution in this region may be mitigated by using a deeper borehole electrode array.

### 3.5 Summary of convergence performance

Table 1 shows the convergence performance of the example cases in terms of, $D_m$, the square norm of the average residuals from eq. (26). As discussed in Section 2.4.2 we expect this measure $D_m$ to achieve values close to one for the converged (posterior) ensemble (i.e. when $m = q + 1$). All cases show a large reduction from the initial value $D_0$ computed from the prior ensemble, which confirms that our prior is rather uninformative and displays large uncertainty. In almost all of the reported cases, $D_m$ converges close to the value 1.0 which, again, is indicative that the data can be explained, up to noise level (measurement errors) by the predictions made by the posterior ensemble. Cases where $D_{q+1}$ achieves value substantially larger than one could represent an underestimation of the variance of the measurement error (e.g. Figs 7 and 11 for resistivity) which we use in eq. (26) to weight the residual. Although we considered realistic measurement errors for both the magnitude and phase angle components of the potential's measurements, phase angle measurement errors were larger (relative to the size of the measurements) than the magnitude's errors used for resistivity inversions. It comes as no surprise that way fewer iterations were required for phase angle inversion since larger errors (and so large variance) means the data can be easier explained, albeit with larger uncertainty, with these larger measurement errors.

We also report the fraction of cells with their zone membership being correctly identified in Table 2 in cases where the true zone geometry is known (1.0 implies perfect identification). Zone membership is obtained from the mean level sets by EKI. Note that this measure is highly dependent on the geometry of the problem and problems with simple geometry and clear background region tend to have high scores. This measure is particularly useful to evaluate reruns of the same problem. For instance, it shows an improvement with (i) using two level sets functions (higher score from Fig. 4

than Fig. 3) and (ii) with lower data noise levels (higher score from Fig. 12 than Fig. 10). Generally, in all cases the zone membership is correctly identified for a large fraction of cells (i.e. >0.7, with the exception of phase in Fig. 7, which is due to the effect of using non-informative priors on phase angles) and the score for resistivity is better than phase angle.

## 4 DISCUSSION

We have demonstrated the extension of EKI with level set parametrization from ERT to IP data. This confirms the potential to extend the method to different geophysical modalities (Tso *et al.* 2021), alongside with concurrent developments in its application in seismic studies (Muir & Tsai 2020; Muir *et al.* 2022). In general, this method is a highly flexible framework that is suitable for a wide range of subsurface applications that has sharp changes in properties and its advantage to solve the inverse problem using only evaluations of the forward solver (and not its Jacobian) means it can be readily applied to new problems and even coupled or joint models.

EKI is a robust approach for solving large-scale high-dimensional Bayesian inverse problems for which fully Bayesian sampling methods such as MCMC are computationally unfeasible. However, it is important to reiterate that EKI relies on Gaussian approximations as well as the use of a small number of realizations. Hence, the estimated uncertainty may be underestimated and, thus, care should be taken to interpret the images (and their uncertainties) returned by EKI and the underlying assumptions should be taken into consideration (e.g. survey geometry, prior ranges, conceptual model). For instance, when considering the uncertainty maps, the resolution pattern of the measurement array (e.g. lower resolution at greater depths for a surface array) and prior formulation (e.g. choice of level-set thresholds $\alpha$ in Section 2.4.2 and Appendix B) should be taken into account. Meanwhile, we have shown that since the range of values for phase angles are small as well as their lower signal-to-noise ratio, their percentage uncertainty can be very high. Users should consider these caveats when interpreting the resultant uncertainty estimates.

We have adopted a sequential approach to invert IP data in which we first infer resistivity, then we infer phase angle. While a joint inversion via EKI can also be performed, the dimension of the input space increases substantially. Thus a larger ensemble size may be needed which, in turn, increases the computational cost of the inversion algorithm. Also, for joint inversions both the magnitude and phase angle of the potential measurements need to be jointly inverted. Given that the phase angle measurements are smaller than the magnitude measurements, there is a high risk that the former will be overshadowed by the latter, and hence not significantly contribute to the estimation of the geophysical properties.

We also note the misfit improvements for phase angles tend to be significantly smaller than resistivity, and the posterior zonal probability patterns are less clear. This does not imply a lack of value for incorporating IP data. Rather, it highlights an independent phase angle inversion requires good estimates of resistivity values as inputs.

Based on the issues arisen in Fig. 3, we have performed an in-depth investigation on using alternative level-set formulations for EKI, such as using two level-set functions for three zones and using a four-zone formulation. We find that using two level-sets and thus allowing jumps from the lowest-value zone to the highest-value zone greatly improve results in that example. We also find
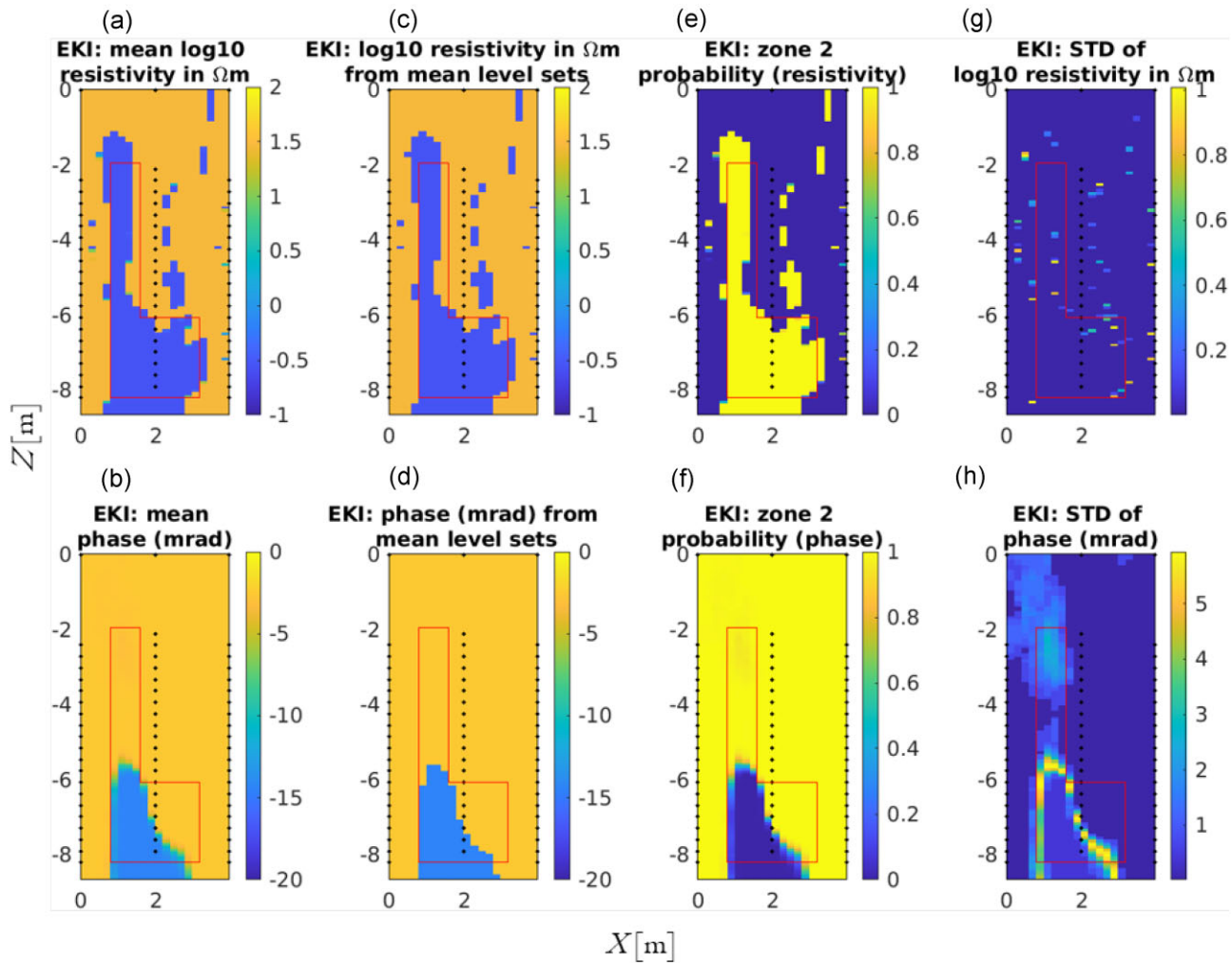
**Figure 12.** Repeat of Fig. 10 with lower data noise. (a, b) The mean resistivity and phase angle model estimated by EKI across samples. (c, d) The resistivity and phase angle model obtained from the mean level sets estimated by EKI. (e, f) The prior estimated probability of zone 2 (i.e. not the background) (g, h) The posterior standard deviation for resistivity and phase angle. The red lines denotes the true PRB boundary. The three vertical arrays of black dots indicate the borehole electrode positions.

**Table 2.** Fraction of cells being correctly identified for their zone membership based on the mean level set. The corrected values are obtained by manually adjusting for zone membership in the estimated field to the true field. This is to account for cases where two or more estimated zonal values are close to each other.

|  | Resistivity (corrected) | Phase (corrected) |
|---|---|---|
| Fig. 2 (inclusions) | 0.98 | 0.82 |
| Fig. 3 (3 layers) | 0.53 (0.87) | 0.51 |
| Fig. 4 (3 layers, 2 level sets) | 0.83 (0.91) | 0.83 |
| Fig. 5 (3 layers, 4 zones) | 0.15 (0.72) | 0.38 (0.63) |
| Fig. 6 (3 layers, 2 level sets, non-uniform priors) | 0.67 | 0.29 |
| Fig. 7 (3 layers, 4 zones, non-uniform priors) | 0.02 (0.74) | 0.18 |
| Fig. 10 (PRB synthetic) | 0.86 | 0.80 |
| Fig. 11 (PRB field) | 0.81 | 0.68 |
| Fig. 12 (PRB synthetic low noise) | 0.89 | 0.81 |

that incorrectly specifying a three-zone problem with four zones only slightly affected EKI results. Using non-informative priors that are identical for all zones somewhat impacted DC resistivity results, but had a very large impact on phase angle results. Finally, we note that each of these alterations to the level set formulation increases the prior uncertainty of the EKI problem, as indicated by the increases in the prior residuals $\mathcal{D}_0$ (relative to that in Fig. 3) in

Table 1. Therefore, we recommend starting with a more basic, more informative prior formulation when encountering new problems and considering whether EKI results can be improved by these alternative formulations incrementally.

As discussed earlier, since IP measures the electrical polarizability of subsurface materials, it provides additional information on, for example, hydraulic properties that cannot be obtained from

ERT. However, since phase angle variations are generally small, it is perhaps even more important to interpret phase angle estimates alongside its uncertainty estimates. EKI provides an effective method for geophysical inversion for both resistivity and phase angles, especially if they are in arbitrarily shaped zones and there is an abrupt jump in property values. Joint interpretation of resistivity and IP images, alongside their uncertainty estimates, can provide useful and reliable information to delineate the spatial distribution of hydraulic properties.

In addition to visual inspection, EKI for IP results can also be used to obtain probabilistic maps of properties of interest, for example permeability, given a suitable petrophysical model, such as the Weller *et al.* (2015) model. In the Supporting Information, we have provided an example to use the results from this paper and uncertainty propagation methods (Tso *et al.* 2019) to obtain maps of mean and standard deviation of permeability, *k*. An advantage of this method is that a breakdown for each of the petrophysical and geophysical parameters can be obtained. For IP inversion, despite the existence of petrophysical relationships between *k* and IP for decades, this is one of the first work to invert field IP data and convert it to *k* fields with uncertainty bounds. It represents a step towards making more meaningful hydrological predictions from IP data. Römhild *et al.* (2022) has recently used cross-hole IP data to complement hydraulic test to image near-surface aquifer. In theory, the EKI level set approach demonstrated herein can be further extended for similar applications.

## 5 CONCLUSIONS

We have demonstrated the use of EKI with level set parametrization for IP data. Unlike commonly used SCI, it can effectively estimate resistivity and phase angle structure with arbitrarily shaped zones. Importantly, it also provides estimation of uncertainty in terms of zone membership and standard deviation. These uncertainty estimates can be propagated to obtain uncertainty bounds of hydrological properties of interest (e.g. via petrophysical relationships). Our findings highlight the added value of using the EKI approach to invert ERT and IP data, not only for recovering geophysical structures but also for making probabilistic assessment of hydrological properties. We have also provided in-depth investigation on the effects of more advanced level set formulations, the use of non-informative priors, and data noise on EKI performance. Percentage uncertainties for phase angles also tend to be higher than that for resistivity. Overall, EKI provides highly valuable information to delineate the juxtaposition of contrasting material properties.

## SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

**suppl_data**

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

## ACKNOWLEDGMENTS

## DATA AVAILABILITY

The data underlying this article are available in Zenodo (Tso & Iglesias, 2023), at https://zenodo.org/doi/10.5281/zenodo.8256492 or GitHub (https://github.com/cmtso/EKI_IP) . The repository contains input data files for each test case and example MATLAB® codes to implement the method.

## REFERENCES

Aleardi, M., Vinciguerra, A. & Hojat, A., 2021. A geostatistical Markov chain Monte Carlo inversion algorithm for electrical resistivity tomography, *Near Surf. Geophys.,* **19**(1), 7–26.

Benoit, S., Ghysels, G., Gommers, K., Hermans, T., Nguyen, F. & Huysmans, M., 2018. Characterization of spatially variable riverbed hydraulic conductivity using electrical resistivity tomography and induced polarization, *Hydrogeol. J.,* **27**(1), 395–407.

Bérubé, C.L., Chouteau, M., Shamsipour, P., Enkin, R.J. & Olivo, G.R., 2017. Bayesian inference of spectral induced polarization parameters for laboratory complex resistivity measurements of rocks and soils, *Comput. Geosci.,* **105**, 51–64.

Bijani, R., Lelièvre, P.G., Ponte-Neto, C.F. & Farquharson, C.G., 2017. Physical-property-, lithology- and surface-geometry-based joint inversion using Pareto Multi-Objective Global Optimization, *Geophys. J. Int.,* **209**(2), 730–748.

Binley, A. & Slater, L., 2020. *Resistivity and Induced Polarization: Theory and Applications to the Near-Surface Earth,* Cambridge Univ. Press.

Binley, A., Hubbard, S.S., Huisman, J.A., Revil, A., Robinson, D.A., Singha, K. & Slater, L.D., 2015. The emergence of hydrogeophsyics for improved understanding of subsurface processes over multiple scales, *Water Resour. Res.,* **51**, 3837–3866.

Binley, A., Keery, J., Slater, L., Barrash, W. & Cardiff, M., 2016. The hydrogeologic information in cross-borehole complex conductivity data from an unconsolidated conglomeratic sedimentary aquifer, *Geophysics,* **81**(6), E409–E421.

Bobe, C., Hanssens, D., Hermans, T. & Van De Vijver, E., 2020. Efficient probabilistic joint inversion of direct current resistivity and small-loop electromagnetic data, *Algorithms,* **13**(6), 144 , doi:10.3390/a13060144.

Bouchedda, A., Bernard, G. & Gloaguen, E., 2017. Constrained electrical resistivity tomography Bayesian inversion using inverse Matérn covariance matrix, *Geophysics,* **82**(3), E129–E141.

Brunetti, C. & Linde, N., 2018. Impact of petrophysical uncertainty on Bayesian hydrogeophysical inversion and model selection, *Adv. Water Resour.,* **111**(July 2017), 346–359.

Calvello, E., Reich, S. & Stuart, A.M., 2022. Ensemble Kalman methods: a mean field perspective, doi:10.48550/arXiv.2209.11371.

Chada, N.K., Iglesias, M.A., Roininen, L. & Stuart, A.M., 2018. Parameterizations for ensemble Kalman inversion, *Inverse Problems,* **34**(5), doi:10.1088/1361-6420/aab6d9.

Chada, N.K., Chen, Y. & Sanz-Alonso, D., 2021. Iterative ensemble kalman methods: A unified perspective with some new variants, *Found. Data Sci.,* **3**(3), 331–369.

Chen, J., Hoversten, G.M., Key, K., Nordquist, G. & Cumming, W., 2012. Stochastic inversion of magnetotelluric data using a sharp boundary parameterization and application to a geothermal site, *Geophysics,* **77**(4), E265 , doi:10.1190/geo2011-0430.1.

Cotter, S.L., Roberts, G.O., Stuart, A.M. & White, D., 2013. MCMC methods for functions: modifying old algorithms to make them faster, *Stat. Sci.,* **28**(3), 424–446.

Day-Lewis, F.D., Singha, K. & Binley, A.M., 2005. Applying petrophysical models to radar travel time and electrical resistivity tomograms: resolution-dependent limitations, *J. geophys. Res.,* **110**(8), 1–17.

de Pasquale, G., Linde, N., Doetsch, J & Holbrook, W.S., 2019. Probabilistic inference of subsurface heterogeneity and interface geometry using geophysical data, *Geophys. J. Int.,* **217**(2), 816–831. doi:10.1093/gji/ggz055.

Dunlop, M.M., Iglesias, M.A. & Stuart, A.M., 2017. Hierarchical Bayesian level set inversion, *Stat. Comput.,* **27**(6), 1555–1584.

Emerick, A.A. & Reynolds, A.C., 2013. Ensemble smoother with multiple data assimilation, *Comput. Geosci.,* **55,** 3–15.

Farquharson, C.G. & Oldenburg, D.W., 1998. Non-linear inversion using general measures of data misfit and model structure, *Geophys. J. Int.,* **134**(1), 213–227.

Fernández-Muñiz, Z., Khaniani, H. & Fernández-Martínez, J.L., 2019. Data kit inversion and uncertainty analysis, *J. appl. Geophys.,* **161,** 228–238.

Flores Orozco, A., Williams, K.H. & Kemna, A., 2013. Time-lapse spectral induced polarization imaging of stimulated uranium bioremediation, *Near Surf. Geophys.,* (11), 531–544.

Galetti, E. & Curtis, A., 2018. Transdimensional electrical resistivity tomography, *J. geophys. Res.,* **123,** 6347–6377. doi:10.1029/2017JB015418.

Hermans, T., Oware, E. & Caers, J., 2016. Direct prediction of spatially and temporally varying physical properties from time-lapse electrical resistance data, *Water Resour. Res.,* **52**(9), 7262–7283.

Iglesias, M. & Yang, Y., 2021. Adaptive regularisation for ensemble Kalman inversion, *Inverse Problems,* **37**(2), doi:10.1088/1361-6420/abd29b.

Iglesias, M., Lu, Y. & Stuart, A., 2016. A Bayesian level set method for geometric inverse problems, *Interf. Free Bound.,* **18**(2), 181–217.

Iglesias, M., Park, M. & Tretyakov, M.V., 2018. Bayesian inversion in resin transfer molding, *Inverse Problems,* **34**(10), doi:10.1088/1361-6420/aad1cc.

Iglesias, M., McGrath, D.M., Tretyakov, M.V. & Francis, S.T., 2022. Ensemble Kalman inversion for magnetic resonance elastography, *Phys. Med. Biol.,* **67,** doi:10.1088/1361-6560/ac9fa1.

Iglesias, M.A., 2015. Iterative regularization for ensemble data assimilation in reservoir models, *Comput. Geosci.,* **19**(1), 177–212.

Iglesias, M.A., 2016. A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems, *Inverse Problems,* **32**(2), doi:10.1088/0266-5611/32/2/025002.

Iglesias, M.A., Law, K. J.H. & Stuart, A.M., 2013a. Evaluation of Gaussian approximations for data assimilation in reservoir models, *Computational Geosciences,* **17**(5), 851–885.

Iglesias, M.A., Law, K. J.H. & Stuart, A.M., 2013b. Ensemble Kalman methods for inverse problems, *Inverse Problems,* **29**(4), doi:10.1088/0266-5611/29/4/045001.

Irving, J. & Singha, K., 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, *Water Resour. Res.,* **46**(11), doi:10.1029/2009WR008340.

Isunza Manrique, I., Caterina, D., Nguyen, F. & Hermans, T., 2023. Quantitative interpretation of geoelectric inverted data with a robust probabilistic approach, *Geophysics,* **88**(3), 1–93, doi:10.1190/geo2022-0133.1.

JafarGandomi, A. & Binley, A., 2013. A Bayesian trans-dimensional approach for the fusion of multiple geophysical datasets, *J. appl. Geophys.,* **96,** 38–54.

Kang, X., Kokkinaki, A., Power, C., Kitanidis, P.K., Shi, X., Duan, L., Liu, T. & Wu, J., 2021. Integrating deep learning-based data assimilation and hydrogeophysical data for improved monitoring of DNAPL source zones during remediation, *J. Hydrol.,* **601,** doi:10.1016/j.jhydrol.2021.126655.

Kemna, A., Binley, A.M. & Slater, L., 2004. Cross-borehole IP imaging for engineering and environmental applications., *Geophysics,* **69**(1), 97–107.

Kemna, A. *et al.*, 2012. An overview of the spectral induced polarization method for near-surface applications, *Near Surf. Geophys.,* **10**(6), 453–468.

Kessouri, P. *et al.*, 2019. Induced polarization applied to biogeophysics: recent advances and future prospects, *Near Surf. Geophys.,* **17**(6), 595–621.

Lindgren, F., Rue, H. & Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *J. R. Stat. Soc., B,* **73**(4), 423–498.

Litman, A., 2005. Reconstruction by level sets of n-ary scattering obstacles, *Inverse Problems,* **21**(6), S131 , doi:10.1088/0266-5611/21/6/S10.

Lopez-Alvis, J., Nguyen, F., Looms, M.C. & Hermans, T., 2022. Geophysical inversion using a variational autoencoder to model an assembled spatial prior uncertainty, *J. geophys. Res.,* **127**(3) , doi:10.1029/2021JB022581.

Madsen, L.M., Fiandaca, G., Auken, E. & Christiansen, A.V., 2017. Time-domain induced polarization – an analysis of Cole–Cole parameter resolution and correlation using Markov Chain Monte Carlo inversion, *Geophys. J. Int.,* **211**(3), 1341–1353.

Matveev, M., Endruweit, A., Long, A., Iglesias, M. & Tretyakov, M., 2021. Bayesian inversion algorithm for estimating local variations in permeability and porosity of reinforcements using experimental data, *Comp. Part A: Appl. Sci. Manufact.,* **143,** doi:10.1016/j.compositesa.2021.106323.

McLachlan, P., Chambers, J., Uhlemann, S. & Binley, A., 2020. Limitations and considerations for electrical resistivity and induced polarization imaging of riverbed sediments: observations from laboratory, field, and synthetic experiments, *J. appl. Geophys.,* **183,** doi:10.1016/j.jappgeo.2020.104173.

Mejus, L., 2014. Using multiple geophysical techniques for improved assessment of aquifer vulnerability *PhD thesis*, Lancaster University, UK.

Michel, H., Nguyen, F., Kremer, T., Elen, A. & Hermans, T., 2020. 1D geological imaging of the subsurface from geophysical data with Bayesian evidential learning, *Comput. Geosci.,* **138,** 104456 , doi:10.1016/j.cageo.2020.104456.

Michel, H., Hermans, T. & Nguyen, F., 2022. Iterative prior resampling and rejection sampling to improve 1-D geophysical imaging based on Bayesian evidential learning (BEL1D), *Geophys. J. Int.,* **232**(2), 958–974.

Moysey, S., Singha, K. & Knight, R., 2005. A framework for inferring field-scale rock physics relationships through numerical simulation, *Geophys. Res. Lett.,* **32**(8), 1–4.

Muir, J.B. & Tsai, V.C., 2020. Geometric and level set tomography using ensemble Kalman inversion, *Geophys. J. Int.,* **220,** 967–980.

Muir, J.B., Clayton, R.W., Tsai, V.C. & Brissaud, Q., 2022. Parsimonious velocity inversion applied to the Los Angeles Basin, CA, *J. geophys. Res.,* **127**(2) , doi:10.1029/2021JB023103.

Mwakanyamale, K., Slater, L., Binley, A. & Ntarlagiannis, D., 2012. Lithologic imaging using complex conductivity: lessons learned from the Hanford 300 Area, *Geophysics,* **77**(6), E397 , doi:10.1190/geo2011-0407.1.

Ntarlagiannis, D., Williams, K.H., Slater, L. & Hubbard, S., 2005. Low-frequency electrical response to microbial induced sulfide precipitation, *J. geophys. Res.,* **110**(G2), doi:10.1029/2005JG000024.

Nussbaumer, R., Linde, N., Mariethoz, G. & Holliger, K., 2019. Simulation of fine-scale electrical conductivity fields using resolution-limited tomograms and area-to-point kriging, *Geophys. J. Int.,* **218**(2), 1322–1335.

Oware, E.K., Irving, J. & Hermans, T., 2019. Basis-constrained Bayesian Markov-chain Monte Carlo difference inversion for geoelectrical monitoring of hydrogeologic processes, *Geophysics,* **84**(4), A37–A42.

Pang, Y., Nie, L., Liu, B., Liu, Z. & Wang, N., 2020. Multiscale resistivity inversion based on convolutional wavelet transform, *Geophys. J. Int.,* **223**(1), 132–143.

Ramirez, A.L. *et al.*, 2005. Stochastic inversion of electrical resistivity changes using a Markov Chain Monte Carlo approach, *J. geophys. Res.,* **110**(B2), doi:10.1029/2004JB003449.

Revil, A., Soueid Ahmed, A., Coperey, A., Ravanel, L., Sharma, R. & Panwar, N., 2020. Induced polarization as a tool to characterize shallow landslides, *J. Hydrol.,* **589,** doi:10.1016/j.jhydrol.2020.125369.

Roininen, L., Lasanen, S., Orispää, M. & Särkkä, S., 2014. Sparse approximations of fractional Matérn fields, *Scand. J. Stat.,* **45**(1), 194–216.

Römhild, L., Fiandaca, G., Hu, L., Meyer, L. & Bayer, P., 2022. Imaging hydraulic conductivity in near-surface aquifers by complementing cross-borehole induced polarization with hydraulic experiments, *Adv. Water Resour.,* **170,** doi:10.1016/j.advwatres.2022.104322.

Rucker, D.F., Tsai, C.-H., Carroll, K.C., Brooks, S., Pierce, E.M., Ulery, A. & Derolph, C., 2021. Bedrock architecture, soil texture, and hyporheic zone characterization combining electrical resistivity and induced polarization imaging, *J. appl. Geophys.,* **188,** doi:10.1016/j.jappgeo.2021.104306.

Saneiyan, S., Ntarlagiannis, D., Ohan, J., Lee, J., Colwell, F. & Burns, S., 2019. Induced polarization as a monitoring tool for in-situ microbial induced carbonate precipitation (MICP) processes, *Ecol. Eng.,* **127,** 36–47.

Scheidt, C., Li, L. & Caers, J., 2018. *Quantifying Uncertainty in Subsurface Systems,* Geophysical Monograph Series, John Wiley & Sons, Inc.,

Schillings, C. & Stuart, A.M., 2017. Analysis of the ensemble Kalman filter for inverse problems, *SIAM J. Numer. Anal.,* **55**(3), 1264–1290.

Sen, M.K. & Stoffa, P.L., 2013. *Global Optimization Methods in Geophysical Inversion,* Cambridge Univ. Press.

Singha, K. & Moysey, S., 2006. Accounting for spatially variable resolution in electrical resistivity tomography through field-scale rock-physics relations, *Geophysics,* **71**(4), A25 , doi:10.1190/1.2209753.

Slater, L., 2007. Near surface electrical characterization of hydraulic conductivity: from petrophysical properties to aquifer geometries - a review, *Surveys in Geophysics,* **28**(2–3), 169–197.

Slater, L. & Binley, A., 2003. Evaluation of permeable reactive barrier (PRB) integrity using electrical imaging methods, *Geophysics,* **68**(3), 911–921.

Slater, L. & Binley, A., 2006. Synthetic and field-based electrical imaging of a zero valent iron barrier: implications for monitoring long-term barrier performance, *Geophysics,* **71**(5), B129–B137.

Slater, L.D. *et al.,* 2010. Use of electrical imaging and distributed temperature sensing methods to characterize surface water-groundwater exchange regulating uranium transport at the Hanford 300 Area, Washington, *Water Resour. Res.,* **46**(10), 1–13.

Stuart, A.M., 2010. Inverse problems: a Bayesian perspective, *Acta Numerica,* **19**(May), 451–559.

Thibaut, R., Compaire, N., Lesparre, N., Ramgraber, M., Laloy, E. & Hermans, T., 2022. Comparing well and geophysical data for temperature monitoring within a Bayesian experimental design framework, *Water Resour. Res.,* **58**(11), doi:10.1029/2022WR033045.

Tso, C.-H.M. & Iglesias, M.A., 2023. *Example Scripts for Ensemble Kalman Inversion of Induced Polarization Data (v1.0.0),* Zenodo , doi:10.5281/zenodo.8256492.

Tso, C.-H.M. *et al.,* 2017. Improved characterisation and modelling of measurement errors in electrical resistivity tomography (ERT) surveys, *J. Appl. Geophys.,* **146,** 103–119.

Tso, C.-H.M., Kuras, O. & Binley, A., 2019. On the field estimation of moisture content using electrical geophysics: the impact of petrophysical model uncertainty, *Water Resour. Res.,* **55**(8), 7196–7211.

Tso, C.-H.M., Iglesias, M., Wilkinson, P., Kuras, O., Chambers, J. & Binley, A., 2021. Efficient multiscale imaging of subsurface resistivity with uncertainty quantification using ensemble Kalman inversion, *Geophys. J. Int.,* **225**(2), 887–905.

Wainwright, H.M., Flores Orozco, A., Bücker, M., Dafflon, B., Chen, J., Hubbard, S.S. & Williams, K.H., 2016. Hierarchical Bayesian method for mapping biogeochemical hot spots using induced polarization imaging, *Water Resour. Res.,* **52**(1), 533–551.

Weller, A., Slater, L., Binley, A., Nordsiek, S. & Xu, S., 2015. Permeability prediction based on induced polarization: Insights from measurements on sandstone and unconsolidated samples spanning a wide permeability range, *Geophysics,* **80**(2), D161–D173.

Williams, K.H., Ntarlagiannis, D., Slater, L.D., Dohnalkova, A., Hubbard, S.S. & Banfield, J.F., 2005. Geophysical imaging of stimulated microbial biomineralization, *Environ. Sci. Technol.,* **39**(19), 7592–7600.

Yang, X., Chen, X., Carrigan, C.R. & Ramirez, A.L., 2014. Uncertainty quantification of CO2 saturation estimated from electrical resistance tomography data at the Cranfield site, *Int. J. Greenhouse Gas Contl*., **27,** 59–68.

Yeh, T.-C.J., Liu, S., Glass, R.J., Baker, K., Brainard, J.R., Alumbaugh, D. & LaBrecque, D., 2002. A geostatistically based inverse model for electrical resistivity surveys and its applications to Vadose Zone Hydrology, *Water Resour. Res.,* **38**(12), doi:10.1029/2001WR001204.

Zhang, X. & Curtis, A., 2020. Seismic tomography using variational inference methods, *J. geophys. Res.,* **125**(4) , doi:10.1029/2019JB018589.

## APPENDIX A: EKI ALGORITHM

---

**Algorithm 1** Ensemble Kalman inversion algorithm for a generic inverse problem.

---

**Inputs:** (1) $w$: measurements; (2) $\Sigma$ measurements' error covariance; (3) $\{u_0^{(j)}\}_{j=1}^{J}$: initial ensemble from the prior $\mathbb{P}(u)$; (4) forward map $\mathcal{G}$.

**Output:** $\{u^{(j)}\}_{j=1}^{J}$: posterior ensemble (ensemble from the approximate posterior $\mathbb{P}(u|d)$.

Set $s_0 = 0$

**while** $s_n < 1$ **do**

(1) **Prediction step.** Evaluate

$\mathcal{G}_n^{(j)} = \mathcal{G}(u_n^{(j)}), \qquad j \in \{1, \ldots, J\},$

and define $\overline{\mathcal{G}}_n = \frac{1}{J} \sum_{j=1}^{J} \mathcal{G}_n^{(j)}.$

(2) **Compute regularisation parameter $\alpha_n$:**

Compute $\alpha_n^* = \frac{1}{M} \frac{1}{J} \sum_{j=1}^{J} ||\Sigma^{-1/2}(w - \mathcal{G}_n^{(j)})||^2$

**if** $s_n + \frac{1}{\alpha_n^*} \geq 1$ **then** Set $\alpha_n = \frac{1}{1-s_n}, s_{n+1} = 1$;

**else~**Set $\alpha_n = \alpha_n^*, s_{n+1} = s_n + \frac{1}{\alpha_n}$.

**end if**

(3)**Analysis step.** Define $C_n^{u\mathcal{G}}, C_n^{\mathcal{GG}}$ by

$$C_n^{\mathcal{GG}} = \frac{1}{J-1} \sum_{j=1}^{J} (\mathcal{G}(u_n^{(j)}) - \overline{\mathcal{G}}_n)(\mathcal{G}(u_n^{(j)}) - \overline{\mathcal{G}}_n)^T,$$

$$C_n^{u\mathcal{G}} = \frac{1}{J-1} \sum_{j=1}^{J} (u_n^{(j)} - \overline{u}_n)(\mathcal{G}(u_n^{(j)}) - \overline{\mathcal{G}}_n)^T.$$

Update each ensemble member:

$u_{n+1}^{(j)} = u_n^{(j)} + C_n^{u\mathcal{G}}(C_n^{\mathcal{GG}} + \alpha_n \Sigma)^{-1}(w - \mathcal{G}_n^{(j)} + \eta_n^{(j)}),$

where $j \in \{1, \ldots, J\}$ and $\eta_n^{(j)} \sim N(0, \Sigma)$.

$n + 1 \to n$

**end while**

---

## APPENDIX B: SINGLE LEVEL-SET PARAMETRIZATION

As for the two-level-set case, we describe only the formulation for the parametrization of resistivity $\rho(\mathbf{x})$. For simplicity we consider a simple three-zone parametrization that relies on the assumption that the unknown resistivity takes only three (unknown) resistivity values $\rho_1$, $\rho_2$ and $\rho_3$ on (unknown) regions denoted by $\Omega_1$, $\Omega_2$ and $\Omega_3$, respectively. These regions are, in turn, parametrized via thresholding the level-set function, denoted by $\xi(\mathbf{x})$. In other words, we assume those regions are defined by

$$\begin{aligned} \Omega_1 =& \{\mathbf{x} : \xi(\mathbf{x}) \leq \alpha_1\}, \\ \Omega_2 =& \{\mathbf{x} : \alpha_1 < \xi(\mathbf{x}) \leq \alpha_2\}, \quad \text{(B1)} \\ \Omega_3 =& \{\mathbf{x} : \xi(\mathbf{x}) > \alpha_2\} \end{aligned}$$

where $\alpha_1$ and $\alpha_2$ are user defined parameters. In summary, the 3-zone characterization of the unknown resistivity is given by

$$\rho(\mathbf{x}) = \begin{cases} \rho_1, & \xi(\mathbf{x}) \leq \alpha_1 \\ \rho_2, & \alpha_1 < \xi(\mathbf{x}) \leq \alpha_2 \\ \rho_3, & \xi(\mathbf{x}) > \alpha_2 \end{cases} \qquad (B2)$$

where for the results presented here we choose $\alpha_1 = -0.1$ and $\alpha_2 = 0.1$. It is worth noticing that the interface between $\Omega_1$ and $\Omega_2$ corresponds to the $\alpha_1$-level-set of $\xi(x)$ (i.e. $\{x : \xi(x) = \alpha_1\}$) while the interface between $\Omega_2$ and $\Omega_3$ is defined by the $\alpha_2$-level-set of $\xi(\mathbf{x})$. The implicit requirement that $\alpha_1 < \alpha_2$ in (11) means that,

under these modelling assumptions, $\Omega_1$ and $\Omega_3$ cannot intersect. The modelling framework here is suitable, for example, in the case where $\Omega_2$ corresponds to a background media with medium resistivity while $\Omega_1$ and $\Omega_3$ consist of non-overlapping regions of high and low resistivity, respectively. This limitations is addressed by the two-level set formulation discussed in Section 2.3. However, with the increase flexibility comes an increase of dimension of the input space and, thus, potential increase in computational cost for the EKI algorithm.