



HAL
open science

Characterization of DNAPL source zones in clay-sand media via joint inversion of DC resistivity, induced polarization and borehole data

Xueyuan Kang, Christopher Power, Amalia Kokkinaki, André Revil, Jichun Wu, Xiaoqing Shi, Yaping Deng

► To cite this version:

Xueyuan Kang, Christopher Power, Amalia Kokkinaki, André Revil, Jichun Wu, et al.. Characterization of DNAPL source zones in clay-sand media via joint inversion of DC resistivity, induced polarization and borehole data. *Journal of Contaminant Hydrology*, 2023, 258, pp.104240. 10.1016/j.jconhyd.2023.104240 . hal-04261961

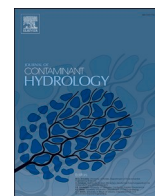
HAL Id: hal-04261961

<https://hal.science/hal-04261961v1>

Submitted on 27 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Characterization of DNAPL source zones in clay-sand media via joint inversion of DC resistivity, induced polarization and borehole data

Xueyuan Kang^a, Christopher Power^b, Amalia Kokkinaki^c, André Revil^d, Jichun Wu^{a,*}, Xiaoqing Shi^{a,*}, Yaping Deng^e

^a Key Laboratory of Surficial Geochemistry of Ministry of Education, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China

^b Department of Civil and Environmental Engineering, Western University, London, Ontario N6A 5B9, Canada

^c Department of Environmental Science, University of San Francisco, San Francisco, CA 94117, USA

^d Univ. Grenoble Alpes, Univ. Savoie Mont-Blanc, CNRS, UMR CNRS 5204, EDYTEM, 73370 Le Bourget du Lac, France

^e School of Natural Resources and Environmental Engineering, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Keywords:

DNAPL source zone characterization
Clay-sand environments
Induced polarization
Joint inversion
Hydrogeophysics

ABSTRACT

Toxic organic contaminants in groundwater are pervasive at many industrial sites worldwide. These contaminants, such as chlorinated solvents, often appear as dense non-aqueous phase liquids (DNAPLs). To design efficient remediation strategies, detailed characterization of DNAPL Source Zone Architecture (SZA) is required. Since invasive borehole-based investigations suffer from limited spatial coverage, a non-intrusive geophysical method, direct current (DC) resistivity, has been applied to image the DNAPL distribution; however, in clay-sand environments, the ability of DC resistivity for DNAPLs imaging is limited since it cannot separate between DNAPLs and surrounding clay-sand soils. Moreover, the simplified parameterization of conventional inversion approaches cannot preserve physically realistic patterns of SZAs, and tends to smooth out any sharp spatial variations. In this paper, the induced polarization (IP) technique is combined with DC resistivity (DCIP) to provide plausible DNAPL characterization in clay-sand environments. Using petrophysical models, the DCIP data is utilized to provide tomograms of the DNAPL saturation (S_N) and hydraulic conductivity (K). The DCIP-estimated K/S_N tomograms are then integrated with borehole measurements in a deep learning-based joint inversion framework to accurately parameterize the highly irregular SZA and provide a refined DNAPL image. To evaluate the performance of the proposed approach, we conducted numerical experiments in a heterogeneous clay-sand aquifer with a complex SZA. Results demonstrate the standalone DC resistivity method fails to infer the DNAPL in complex clay-sand environments. In contrast, the combined DCIP technique provides the necessary information to reconstruct the large-scale features of K/S_N fields, while integrating DCIP data with sparse but accurate borehole data results in a high resolution characterization of the SZA.

1. Introduction

With climate change and population growth, groundwater resources are becoming increasingly critical for economic development and potable water supply. However, contaminants from industrial activities threaten groundwater quality and pose a serious global environmental and social problem. Groundwater contamination caused by dense non-aqueous phase liquids (DNAPLs) is widespread at many industrialized sites (Kueper et al., 2014). As DNAPLs migrate through the unsaturated and saturated zones, the DNAPL may get trapped on local low-permeability media (e.g., clay lenses) as residual ganglia or pools

(Lenhard et al., 1989). The resulting DNAPL source zones (SZ) may exist as long-term groundwater contamination for decades.

An important prerequisite for designing effective remediation plans is to delineate the DNAPL source zone architecture (SZA) in sufficient detail (Abriola et al., 2012; Akyol et al., 2023; Akyol and Yolcubal, 2013; Karaoglu et al., 2019; Tick et al., 2022). Nevertheless, this is highly challenging as the DNAPL SZA is highly irregular and complicated due to the dependencies of DNAPL spatial distribution on subsurface properties and their spatial heterogeneity (Koch and Nowak, 2015; Mumford et al., 2022). Therefore, to accurately predict SZ behavior before and during clean-up efforts and estimate SZ longevity, characterization of

* Corresponding authors.

E-mail addresses: jcwu@nju.edu.cn (J. Wu), shixq@nju.edu.cn (X. Shi).

<https://doi.org/10.1016/j.jconhyd.2023.104240>

Received 21 April 2023; Received in revised form 26 July 2023; Accepted 29 August 2023

Available online 1 September 2023

0169-7722/© 2023 Elsevier B.V. All rights reserved.

hydraulic conductivity (K) and DNAPL saturation (S_N) are essential (Morrissy and Grismer, 1999).

Various characterization methods have been applied in DNAPL-contaminated sites. Conventional methods (e.g., cores from drilling, downstream monitoring wells, and direct push techniques) can only provide limited-spatial-resolution information of DNAPL content and lithology heterogeneity due to the sparsity of sampling points (Siegrist et al., 2006; Parker et al., 2008).

Geophysical methods are highly attractive as complementary techniques for DNAPL characterization since they are non-intrusive and have high sampling density. As a widely applied geophysical method, Direct Current (DC) resistivity has been successfully applied for DNAPL SZ detection (e.g., Chambers et al., 2010), characterizing dissolved NAPL plumes (e.g., Ntarlagiannis et al., 2016), and estimating corresponding hydraulic characteristics (e.g., Binley et al., 2002; Hinnell et al., 2010; Crestani et al., 2015).

However, for complex heterogeneous environments (e.g., sand-clay media), the ability of DC resistivity for DNAPLs imaging is limited. This is because DC resistivity measurements are influenced by both pore water properties, such as electrolytic conductivity, and interfacial electrochemical properties, such as cation exchange capacity (CEC) which can be related to lithology. Therefore, in clayey environments with surface conductivity contribution at the solid-pore water interface, DC resistivity is less effective in discriminating between DNAPL and surrounding soils, especially clayey materials. For instance, the DC resistivity alone cannot distinguish between highly saturated DNAPL in clayey soils (i.e., sand-clay mixture containing a certain proportion of clay grains) and moderately saturated DNAPLs in sandy soils (see Power et al., 2013; Almpanis et al., 2021). DNAPL presence in clay-sand soils is common since DNAPLs tend to accumulate on top of clayey formations. In situations with mixture of sand and clay, DNAPL SZs form networks of ganglia and pools that strongly depend on the soil heterogeneity. That in turn affects how long a DNAPL SZ will persist, contaminating surrounding and downstream groundwater, until it is completely dissolved.

Compared with DC resistivity, time-domain induced polarization (IP) may provide additional information about lithology and grain-surface properties by measuring the capacitive properties (i.e., chargeability) of the subsurface. It can be used to discriminate between pore water and interfacial properties (Vinegar and Waxman, 1984; Revil, 2013a; Maurya et al., 2018). This, in addition to the wealth of new petrophysical models that explain the polarization mechanism of a range of soils and rocks (Revil et al., 2017a; Abdusamad et al., 2019a, 2019b) partially saturated by oils (Schmutz et al., 2010, 2012; Deng et al., 2018) or dissolved metals (Vaudelet et al., 2011; Revil et al., 2017a), have resulted in the growing application of IP to DNAPL investigations (e.g., Sparrenbom et al., 2017; Nivorlis et al., 2019). Time-domain IP properties (chargeability) are usually measured simultaneously with the resistivity using the same electrode configuration. Therefore, this combined DC and IP is often referred to as DCIP. The measured chargeability permits additional information of the lithological distribution that controls DNAPL SZA, particularly in clayey environments (Almpanis et al., 2021).

Despite the benefit over standalone DC resistivity, DCIP may still only provide a low-resolution image of the SZA due to three main factors: (1) measurement errors, interferences and the uncertainties in petrophysical parameters (Brunetti and Linde, 2018); (2) low data sensitivity in deeper regions (Power et al., 2015); (3) missing spatial features of highly irregular SZA due to the simplified isotropic smoothing applied in traditional geophysical inversion (Kang et al., 2021).

To address the first two limitations, low-resolution but spatially extensive DCIP data can be integrated with high-accuracy but spatially sparse borehole data within a joint inversion framework; this allows for effective delineation of both large-scale features and identification of finer features that affect the collected measurements, maximizing the utility of complementary information in different datasets (e.g., Rossi

et al., 2015; Jardani et al., 2013).

However, traditional joint geophysical inversion methods rely on regularization of the unknowns by isotropic smoothing or geostatistical information, which may not be appropriate for the highly irregular S_N field (Kang et al., 2021; Arshadi et al., 2020). Therefore, traditional inversion approaches may lead to overly smooth estimations of SZA. To better represent the distribution of highly irregular S_N fields, a convolutional variational autoencoder (CVAE) can be used. CVAE is a neural network (NN) which can satisfactorily capture highly complex spatial patterns by learning these patterns from training data. Kang et al. (2021) demonstrated the utility of CVAE to parameterize the irregular DNAPL distribution characteristic of heterogeneous soil conditions. The CVAE-based parameterization approach has also been used to improve the inversion of ground-penetrating radar traveltime data (Lopez Alvis et al., 2022).

While the CVAE-based joint inversion framework has been shown to provide plausible characterization of DNAPL SZA with integrated hydrogeological and geophysical data, it has relied on standalone DC resistivity for SZA imaging, under a petrophysical assumption that neglects the surface conductivity contributions from solid phase (e.g., clayey medium) (Kang et al., 2021). This assumption is only valid for clay-free sand. For complex clay-sand environments, standalone DC resistivity cannot discriminate the DNAPL contaminant from surrounding soils, which limits its ability to identify DNAPL SZs. As chargeability measurements can better describe the lithology and discriminate between DNAPL and clayey soils, combined DCIP may improve the imaging of S_N and K fields. This would also improve the performance of joint inversion that uses these geophysical datasets, like the CVAE-based SZA inversion strategy in complex clay-sand environments.

This work aims to demonstrate that DCIP data can be critical in improving the characterization of DNAPL SZA and K distribution in complex clay-sand environments, where the conductive clayey soils hinder the effectiveness of standalone DC resistivity for DNAPL mapping. To achieve this, we present a two-step joint inversion framework: (1) DC and IP are combined to image the large-scale features of the S_N and K distributions, and (2) these DCIP-estimated S_N and K are used as conditioning data and subsequently combined with borehole measurements (i.e., K , S_N and downstream aqueous-phase concentration data) in a deep learning-based CVAE joint inversion method to provide a detailed estimation of the DNAPL SZA and its associated uncertainty. We demonstrate this framework in a synthetic clay-sand aquifer with an intricate DNAPL SZA.

2. Theory and methods

In this section, a coupled DNAPL-DCIP model is first presented to simulate the DCIP response of complex DNAPL SZs in clay-sand environments (Section 2.1). Intricate, realistic DNAPL SZAs are generated based on the underlying intrinsic K of the clay-sand mixture, and using the coupled DNAPL-DCIP model, the corresponding DCIP measurements are recorded. The obtained K/S_N fields and DCIP measurements are used as reference fields and observation data, respectively, for the inversion. The two-step joint inversion framework is then demonstrated, which comprises (1) a geostatistical approach for initial K/S_N imaging by inverting DCIP data (Section 2.2), and (2) a deep learning-based inversion method to improve DNAPL imaging by integrating the DCIP-estimated K/S_N fields and the borehole measurements (Section 2.3).

2.1. Coupled DNAPL-DCIP model for generating geophysical measurements

To simulate the DCIP response from DNAPL presence and the background soil matrix comprising clay and sand, a coupled DNAPL-DCIP model is developed by integrating a DNAPL percolation model, petrophysical models that consider both the electrical conductivity and chargeability, and a geo-electrical forward model. The proposed

DNAPL-DCIP model is summarized in following steps:

Step 1. Define intrinsic hydraulic conductivity (K).

We use a spectral-based method (Dietrich and Newsam, 1993) to generate a realistic, spatially-correlated subsurface distribution of hydraulic conductivity (K). Then the intrinsic permeability k can be calculated from K using $k = K\eta/\rho g$, where ρ , g , and η represent the pore-water density, gravitational acceleration, and dynamic viscosity, respectively. The permeability in each model cell determines the relative fraction of sand and clay in that cell, as described next.

Step 2. Determine the clay content (cl).

In this work, the lower threshold permeability of clay-free sand (k_{sd} [m^2]) is used to calculate the volumetric clay content (cl [-]). k_{sd} can be obtained by an empirical equation developed by Revil and Cathies (1999):

$$k_{sd} = \frac{d_{sd}^2 (\phi_{sd})^{3m}}{24} \quad (1)$$

where d_{sd} is the grain diameter of sand [m], ϕ_{sd} represents the porosity of sand [-], and m denotes the cementation exponent [-]. Eq. 1 was developed for the permeability of a clean sand. Permeability values that are larger than k_{sd} are considered to be clay-free sands, whereas others are clay-sand mixture containing a fraction of clay.

The equation by Revil and Cathies (1999) is then used to calculate the volumetric clay content, cl (ratio of clay) based on the permeability of clay-sand mixtures (k [m^2]):

$$cl = \frac{k_{sd}^{\frac{1}{3m}} - k^{\frac{1}{3m}}}{k_{sd}^{\frac{1}{3m}} \left(\frac{1 - \phi_{cl}}{\phi_{sd}} \right)} \quad (2)$$

where ϕ_{cl} is the porosity of pure clay [-]. For the grids with $k > k_{sd}$, the cl is assumed to be 0. Otherwise, cl can be calculated from Eq. 2. Notably, Eq. 2 was adapted from a physics-based permeability-prediction model developed by Revil and Cathies (1999). This model has been validated to provide close agreement between predicted permeabilities and those measured over 11 orders of magnitude from available laboratory and core measurements (Revil and Cathies, 1999). The validation data were collected from a variety of materials (e.g., unconsolidated sand and beads, natural sandstones) (Chauveteau and Zaitoun, 1981; Chilindar, 1964; Bear, 1988; Johnson et al., 1986).

Step 3. Determine the total porosity (ϕ).

The total porosity (ϕ [-]) is obtained from the clay content (cl) by (Revil and Cathies, 1999):

$$\phi = \phi_{sd}(1 - cl) + \phi_{cl}cl \quad (3)$$

Note that all the petrophysical parameters we used for the numerical experiments, such as ϕ_{cl} , ϕ_{sd} , and d_{sd} , were summarized in Table 1 in section 3.1.

Step 4. Simulation of DNAPL percolation with k and ϕ .

DNAPL percolation in a 3D heterogeneous field is simulated via a stochastic invasion-percolation algorithm developed by Koch and Nowak (2015), which was adapted from Glass et al. (2001). DNAPL infiltration is simulated by modelling invasion into each numerical grid based on its entry pressure. In this way, the stochastic invasion-percolation model can consider the dependencies of the final SZA on the effective permeabilities and porosities. To account for the unresolved pore-scale heterogeneity, a stochastic component is also introduced during the invasion modelling. The stochastic invasion-percolation model provides a computationally efficient way to generate intricate SZAs that are physically consistent with the spatial distribution of permeability. Using DNAPL release information (e.g., the location of release point and release mass), k , and ϕ , the DNAPL SZA (i.e., distribution of S_N) can be generated. For a more detailed description of the adapted stochastic invasion-percolation algorithm, please see Text S1 in supporting information or Koch and Nowak (2015).

Step 5. Determine electrical conductivity (σ) and chargeability (M).

Table 1

Reference model parameters for the numerical experiment.

Parameter	Value
Domain size (m)	20 × 20 × 5
Grid spacing (m) (Δx , Δy , Δz)	0.5 × 0.5 × 0.5
<i>Geostatistical parameters for generating reference ln K field</i>	
Covariance kernels for ln K_i	$q_K(x, x') = \sigma_K^2 \exp(- x - x' /D)$
Correlation lengths for ln K_i (m)	$l_x, l_y, l_z = 5.0, 5.0, 1.2$
Variance for ln K_i	$\sigma_K^2 = 2.0$
Mean for ln K_i (ln(m/s))	ln $K_{mean} = -14.5$
<i>DNAPL source zone</i>	
Spill point (m)	(10, 10, 0)
DNAPL density (kg/m ³) ^a	1462
DNAPL viscosity (Pa·s) ^a	5.9×10^{-4}
DNAPL residual saturation ^a	0.2
Total DNAPL release mass (tons)	7.15
Release rate (m ³ /s) ^b	5×10^{-6}
Brooks-Corey shape	2
<i>Groundwater flow and transport</i>	
Mean hydraulic gradient	0.01
DNAPL solubility (g/L)	1.1
Molecular diffusion coefficient	0
Longitudinal dispersivity (m)	0.05
Horizontal transverse dispersivity (m)	0.01
<i>Petrophysical model</i>	
Electrical conductivity of the aqueous phase (S/m)	0.2
Cementation exponent ^c	2
Saturation exponent ^c	2
d_{sd} (m) ^d	1×10^{-4}
k_{sd} (m ²) ^d	1.07×10^{-12}
B (m ² V ⁻¹ s ⁻¹) ^c	3.0×10^{-9}
λ (m ² V ⁻¹ s ⁻¹) ^c	3.0×10^{-10}
Cation exchange capacity (C/kg) ^c	1.45(sand) 48,000(clay)
Porosity ^c	0.37(sand) 0.56(clay)

^a doi: Cohen et al. (1993).

^b doi: Power et al. (2013).

^c doi: Revil et al. (2021).

^d doi: Revil and Cathies (1999).

With the physics-based petrophysical models developed by Revil et al. (2017a), the hydrogeologic properties (e.g., S_N and cl) can be converted to the geophysical properties (e.g., σ , M). For a clean sand medium, one can safely ignore the surface conductivity contribution when applying Archie's law (Archie, 1942) to relate the hydrogeologic parameters to geophysical properties. However, for clayey materials, one should account for the surface conductivity contribution from the solid-pore water interface (e.g., Waxman and Smits, 1968). In this work, the low-frequency complex conductivity models by Revil et al. (2017a, 2020) is used to consider the impact of DNAPL presence and background soils (clay-sand mixtures) on σ and M :

$$\sigma = \theta^m \sigma_w + \theta^{m-1} \rho_g (B - \lambda) \text{CEC} \quad (4)$$

$$M_n = \theta^{m-1} \rho_g \lambda \text{CEC} \quad (5)$$

where σ represents the DC electrical conductivity of the clay-sand mixtures [S/m]. Note that even though we are using DC resistivity method, we present the resistivity parameter in its reciprocal (i.e., electrical conductivity, σ). M_n denotes the normalized chargeability [S/m]. M is the chargeability ($M = M_n / \sigma$). Based on the previous study on petrophysical model, this work assumes that the cementation exponent (m) is equal to the saturation exponent (n), and that the quadrature conductivity exponent (p) equals the saturation exponent (n) minus 1 (see justification in the Figs. 18 and 20 of Revil, 2013b). θ denotes the volumetric water content [-], which can be obtained by $\theta = \phi S_w$. S_w is water saturation (in this work, $S_w = 1 - S_N - S_G$). σ_w is the pore water conductivity [S/m], ρ_g denotes the mass density of the solid phase (2650 kg/m³ for both sand and clay minerals). B denotes the apparent mobility of the counterions for surface conduction [m²V⁻¹ s⁻¹], and λ represents

the apparent mobility of the counterions for polarization [$m^2V^{-1} s^{-1}$]. The CEC [C/kg] denotes the cation exchange capacity of the subsurface media (1 meq/100 g = 96,320C/kg). The effective CEC of clay-sand mixture can be calculated from (Almpanis et al., 2021):

$$CEC = CEC_{sd} \cdot (1 - cl) + CEC_{cl} \cdot cl \tag{6}$$

Notably, the electrochemical properties of the interface between water and DNAPL fluid are assumed to be negligible in this work. Although Revil et al. (2011) suggested that the presence of electrical double layers on water-NAPL interfaces would influence both surface conduction (resistivity) and polarization (chargeability) and possibly change the DCIP response of the subsurface, there is still a lack of knowledge to explicitly model the impact of EDL properties of DNAPL.

Step. 6 Geophysical forward model for generating DCIP measurements.

Using the geophysical properties (σ and M_n) from the petrophysical models, the DCIP measurements (i.e., electrical potentials V , and apparent chargeability M_a) can be computed through the following

equations (see Oldenburg and Li, 1994):

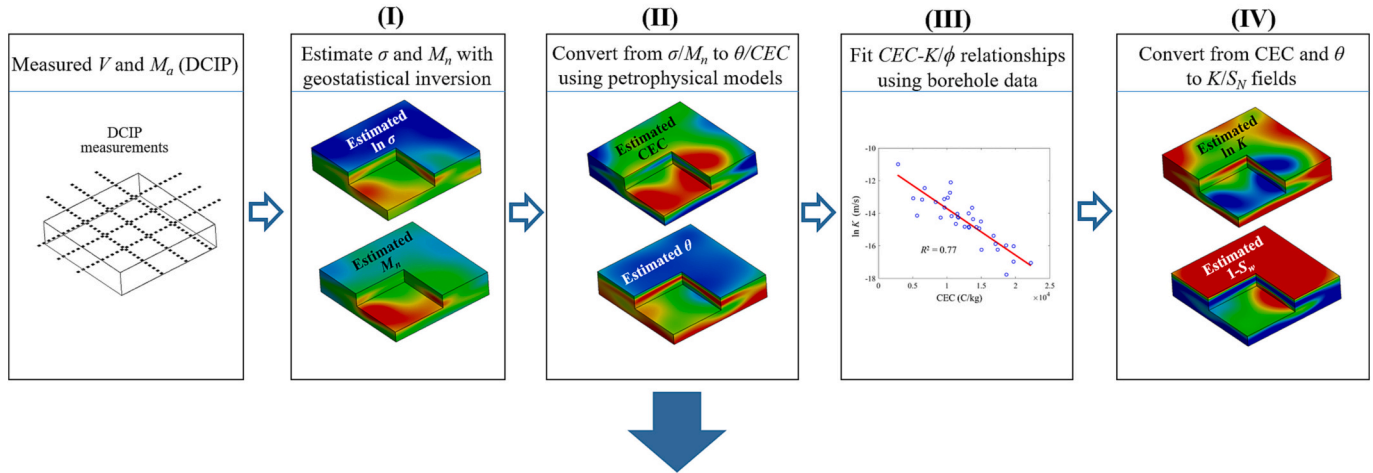
$$\nabla \bullet (\sigma \nabla V) = -I \delta(\mathbf{r}) \rightarrow V = \Psi(\sigma) \tag{7}$$

$$M_a = \frac{\Psi(\sigma_{IP}) - \Psi(\sigma)}{\Psi(\sigma_{IP})} \tag{8}$$

where $\delta(\mathbf{r})$ represents the Dirac delta function, M_a denotes the apparent chargeability [mV/V]. Ψ is the forward model operator for DC resistivity, see Eq. 7. I denotes the current intensity. σ is the electrical conductivity, $\sigma_{IP} = \sigma(1 - M)$. In this work, the IP-measured apparent chargeability is approximated as a small perturbation of the resistivity model. To obtain this perturbation (i.e., apparent chargeability, M_a), the forward model for DC resistivity (i.e., Ψ) is performed twice by using the original conductivity σ and the induced polarization-affected conductivity $\sigma(1-M)$, respectively (Oldenburg and Li, 1994).

In the coupled DNAPL-DCIP model, there are many parameters that are difficult to obtain in the field (e.g., the parameters about DNAPL release, and the petrophysical parameters k_{sd} , d_{sd} , ϕ_{sd} , ϕ_{cl} , CEC_{sd} , CEC_{cl}).

Step A. Geophysical estimation of K and S_N



Step B. Deep learning-based joint inversion: incorporating multi-source information

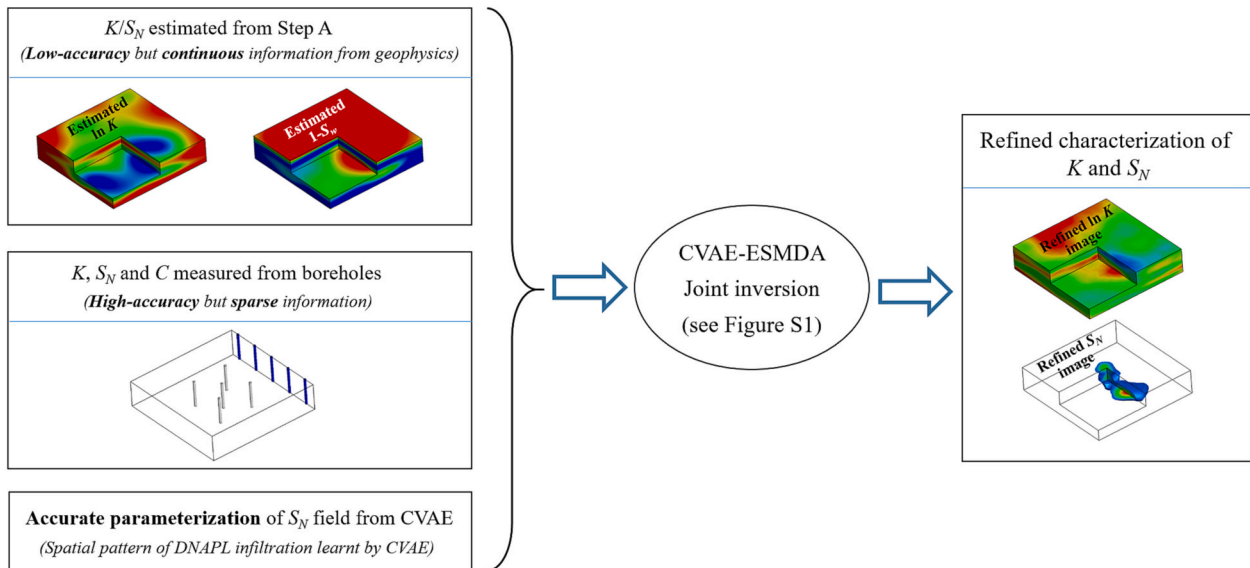


Fig. 1. Workflow of the two-step joint inversion framework: (a) Using DCIP to estimate the K and S_N . (b) Deep learning-based joint inversion. K and S_N represent the hydraulic conductivity and DNAPL saturation, respectively. V and M_a represent the electrical potential and apparent chargeability measured by DCIP, respectively. σ , M_n , CEC, θ , and ϕ denote the electrical conductivity, normalized chargeability, cation exchange capacity, volumetric water content, and porosity, respectively. C represent the downstream dissolved DNAPL concentration. CVAE-ESMDA is the Convolutional Variational AutoEncoder – Ensemble Smoother with Multiple Data Assimilation method.

Here, these parameters are only used for generating the DCIP responses, which are then used as synthetic observation data in inversion. In other words, in real-world applications, one can directly measure the DCIP signals and apply a field-scale DNAPL inversion, rather than generating DCIP measurements by running the coupled DNAPL-DCIP model with the aforementioned parameters.

2.2. Geostatistical inversion of DCIP data for initial K/S_N imaging

In this section, the geostatistical inversion of DCIP data for initial K/S_N imaging is described. The main procedure is demonstrated in Fig. 1 (Step A) and summarized as follows:

2.2.1. Geostatistical inversion of σ and M_n

σ is first estimated by using electrical potential (V) data collected at the ground surface, and then the conductivity tomogram and M_a data are used for the inversion of the M_n . The inversion of σ and M_n is solved by the principal component geostatistical approach (PCGA, Kitanidis and Lee, 2014) and linear inversion (Oldenburg and Li, 1994), respectively. The PCGA is computationally efficient for large-scale inversion (*i.e.*, the number of unknowns is large), since it approximates the dense prior covariance with a low-rank matrix, while employing a matrix-free method to circumvent the explicit computation of the Jacobian matrix. The PCGA has been employed for various hydrogeophysical investigations (*e.g.*, Lee and Kitanidis, 2014; Soueid Ahmed et al., 2018; Kang et al., 2020). For the details of the PCGA, please refer to supporting information (Text S2).

2.2.2. Convert from σ/M_n to θ/CEC using petrophysical models

We can compute the distribution of the θ and CEC from M_n and σ according to following equations (Revil, 2013b; Revil et al., 2017a, 2020, 2021):

$$\theta = \left[\frac{1}{\sigma_w} \left(\sigma + M_n - \frac{M_n}{R} \right) \right]^{1/m} \quad (9)$$

$$CEC = \frac{M_n}{\theta^{m-1} \rho_g \lambda} \quad (10)$$

where R is a dimensionless number introduced by Revil et al. (2017a, 2020, 2021), with $R = \lambda/B \approx 0.10$ (independent of temperature and saturation). In this work, the pore water conductivity is assumed to be known and is uniform in the domain. Eqs. 9 and 10 are physics-based and not site-specific. They have been successfully validated in a variety of materials, such as weathered limestones (Revil et al., 2020), clayey and sandy soils (Revil et al., 2017a, 2021; Deng et al., 2018, 2020; Abdulsamad et al., 2019a), mudbricks (Abdulsamad et al., 2020), and rocks (Soueid Ahmed et al., 2020; Panwar et al., 2021; Abdulsamad et al., 2019b). These equations have been widely applied to quantify the hydrogeological properties (*e.g.*, θ , porosity, and CEC), see the above-mentioned references.

2.2.3. Establish CEC- ϕ and CEC- K relationships by fitting borehole data

Even with the θ estimated from the last step, S_w and S_N are still unknown, as knowledge of the heterogeneous porosity distribution is absent ($S_w = \theta/\phi$). Eqs. 3 and 6 indicate that both CEC and ϕ are related to the clay content cl . Therefore, one may establish a site-specific empirical relationship between CEC and ϕ by fitting the DCIP-estimated CEC to the borehole measured ϕ . Besides being used to estimate ϕ , the CEC is also useful for inferring K . Semi-empirical equations have been developed to relate the permeability k with CEC (Sen et al., 1990; Revil et al., 2020; Soueid Ahmed et al., 2020). Instead of assuming an exactly known CEC- K relationship, a similar strategy with CEC- ϕ fitting to get an approximate CEC- K relationship is employed.

2.2.4. Convert from CEC and θ to K/S_N fields

Utilizing the established empirical CEC- ϕ and CEC- K relationships, one can convert the DCIP-estimated CEC field to ϕ and K distributions, and image the S_w and S_N using $S_w = \theta/\phi$, and $S_N = 1 - S_w$ (only valid for saturated zone below the water table).

Note that what we present here is a “sequential” inversion strategy (also called uncoupled strategy, see Camporese et al., 2015), *i.e.*, the flow of information is one way: from geophysics to hydrology (see Step A in Fig. 1). Although the coupled DNAPL-DCIP model in section 2.1 is used to generate synthetic observation data, this model is not embedded in the inversion framework. Therefore, the information of DNAPL transport is not explicitly considered through the geostatistical inversion, and the DNAPL release information is not required to be known for the inversion.

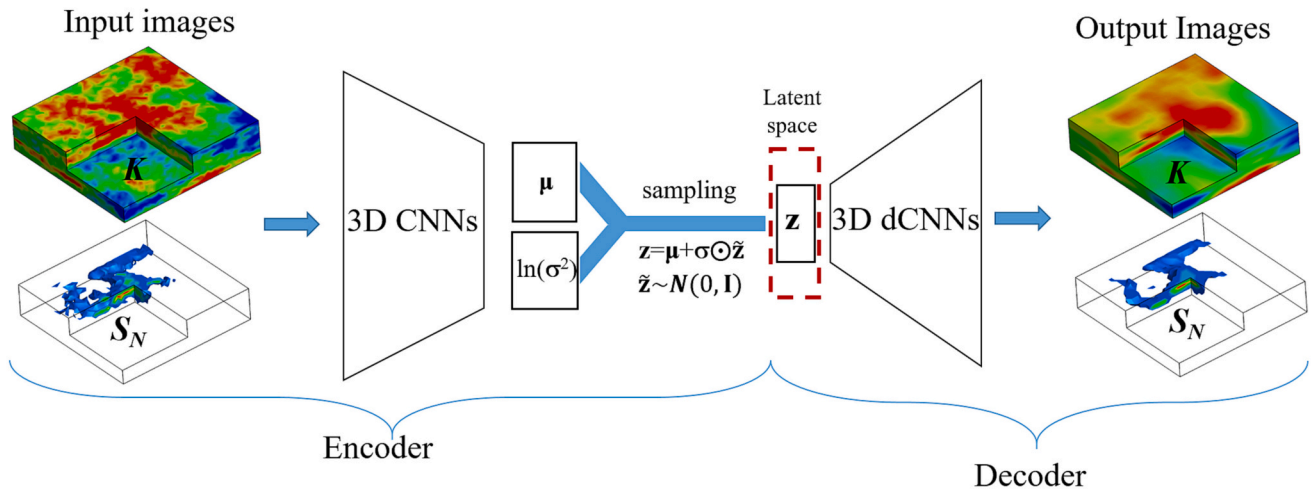
2.3. Deep learning-based joint inversion

From Step A, an initial K/S_N imaging may be obtained by DCIP inversion. However, this initial imaging may be low-resolution due to (1) uncertainty of IP data interpretation (*e.g.*, large measurement errors, petrophysical uncertainty, and diminished data sensitivity in deeper regions), and (2) an over-simplified representation of the highly irregular SZA imposed by a geostatistical prior. To further improve the initial K/S_N characterization, the convolutional variational autoencoder - ensemble smoother with multiple data assimilation (CVAE-ESMDA, Kang et al., 2021) is used as a joint inversion framework to incorporate the following multi-source information: (1) low-resolution but continuous information of K/S_N from DCIP inversion (Step A in Fig. 1), (2) high-accuracy but sparse borehole measurements (*e.g.*, K , S_N and downstream aqueous-phase concentration data), and (3) the accurate representation of the complex DNAPL infiltration pattern provided by CVAE (see Step B in Fig. 1). For the CVAE-ESMDA method, the CVAE network and the ESMDA are used for parameterization and inversion, respectively. In this section, the accurate representation of K/S_N provided by the CVAE is presented. For details on the ESMDA method, refer to the supporting information (Text S3) and Emerick and Reynolds (2013).

The CVAE is a neural network based method suitable for generating realizations with complex probability distributions in spaces that are high dimensional, as in the case of models with numerical grids larger than 1000 grid points. With DNAPL SZA being highly irregular, the CVAE is used to represent its probability distribution; it first utilizes an unsupervised technique to learn the structure of the unknowns (*i.e.*, K/S_N fields) so it can generate new realizations with similar spatial features as training data used to develop the CVAE network. In this way, the CVAE can implicitly capture the underlying physics of DNAPL infiltration without multiphase modelling. The CVAE network receives many pairs of K and S_N images for training and comprises an encoder and decoder (Fig. 2). The encoder maps these high-dimensional irregular K/S_N images to low-dimensional normal-distributed latent vector \mathbf{z} , and the decoder reconstructs the irregular K/S_N images from \mathbf{z} (Fig. 2). This encoder-decoder architecture, or as formally called, the CVAE parameterization, serves as a “bottleneck” structure that ensures only the most important features (*i.e.*, the low-dimensional latent vectors, \mathbf{z}) can be preserved and be reconstructed. Fig. 2 shows the main contents of the CVAE-ESMDA method, which is used for Step B in Fig. 1. For more information on the CVAE, see Kang et al. (2021).

By integrating the CVAE parameterization and ESMDA inversion, a joint inversion framework can be formulated to estimate the highly irregular K/S_N fields by incorporating the multisource information (see the Step B in Fig. 1). First, the CVAE network is trained on a set of K and S_N realizations generated by geostatistical and multiphase modelling (see the details in section 3.3). This way, the CVAE network captures the highly irregular spatial patterns of SZs. After CVAE training (Fig. 2a), the CVAE-ESMDA inversion process begins using the latent variables \mathbf{z} as the unknowns to be estimated, which are reconstructed to the original K/S_N

(a) Training CVAE



(b) CVAE-ESMDA Inversion

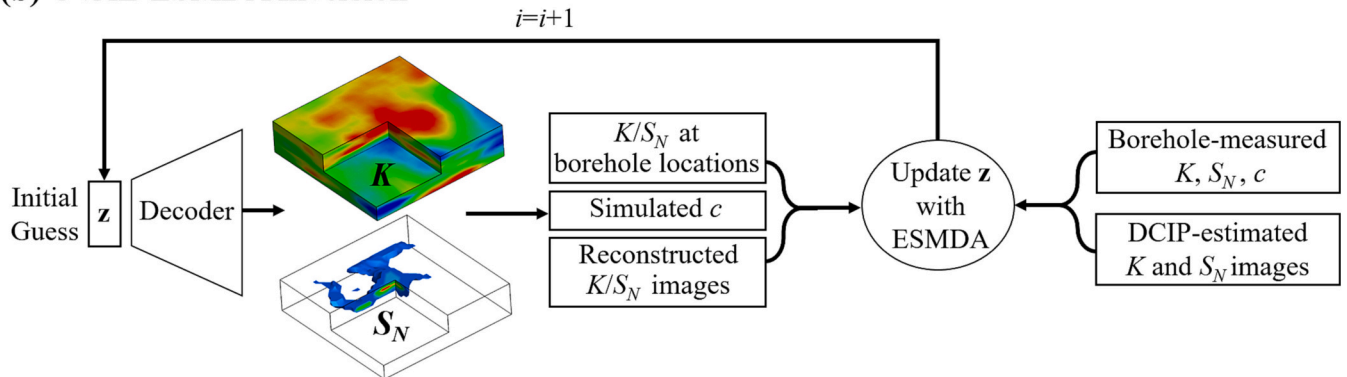


Fig. 2. Workflow of the CVAE-ESMDA inversion framework (i.e., Step B in Fig. 1). (a) using CVAE to parameterize K and S_N fields. (b) using CVAE-ESMDA for joint inversion. Note that 3D CNN and dCNN denotes the three-dimensional convolutional neural network and deconvolutional network, respectively. K , S_N and c represent the hydraulic conductivity, DNAPL saturation and downstream aqueous phase DNAPL concentration, respectively.

fields by the decoder to run the forward model. More specifically, the CVAE-ESMDA inversion procedure is:

- (1) Sample an initial ensemble of latent variables \mathbf{z} from $\mathcal{N}(\mathbf{0}, \mathbf{I})$;
- (2) Using the CVAE decoder, reconstruct the corresponding K and S_N fields from \mathbf{z} . In this way, the decoded S_N fields are forced to honor the DNAPL infiltration patterns (i.e., the connected and finger-like architectures).
- (3) Obtain the simulated $K/S_N/c$ borehole measurements and K/S_N images for each realization. c represents the downstream dissolved DNAPL concentration. The forward model for c can be found in the supporting information (Text S4);
- (4) Employ the ESMDA to update the latent variables \mathbf{z} for the i -th iteration by conditioning on the borehole $K/S_N/c$ measurements and DCIP-estimated K/S_N images;
- (5) Repeat (2) to (4) until the last iteration. Note that the number of iteration (N) is a predefined number in ESMDA (see Emerick and Reynolds, 2013).
- (6) Convert the best estimates for \mathbf{z} to $\mathbf{x} = [S_N, K]^T$ through the decoder.

In summary, to harness the complementarity of multisource data (i.e., DC resistivity, chargeability, and borehole measurements) and utilize the information from an accurate representation of the highly complex SZA provided by CVAE, a two-step joint inversion framework is proposed (Fig. 1). In Step A, the DCIP data is inverted/interpreted with a

geostatistical inversion method, and sophisticated EDL-based petrophysical models, to reconstruct the large-scale features of the K and S_N distributions. In Step B, the DCIP-estimated S_N and K are incorporated with borehole measurements in the CVAE-ESMDA, which accurately parameterizes the highly irregular SZA and provides a refined DNAPL image.

3. Numerical experiments methodology

3.1. Synthetic aquifer and source zone architecture

We evaluated the performance of the proposed methodology in a 3D heterogeneous aquifer with intricate source zone architecture. The research domain ($20 \text{ m} \times 20 \text{ m} \times 5 \text{ m}$, Fig. 3) was discretized into $40 \times 40 \times 10 = 16,000$ numerical grids. The length of each grid was 0.5 m. The depth of groundwater table was 0.5 m. The water saturation (S_W) in the vadose zone (above the groundwater table) was set as 0.4. The side boundaries were governed by constant hydraulic head condition, thus conducting an average hydraulic gradient of 0.01 in the positive x -direction. A no-flux boundary was imposed at the bottom of the aquifer.

To generate the synthetic intrinsic K fields, we used a spectral-based method (Dietrich and Newsam, 1993) with the parameters listed in Table 1. The generated reference K field (Fig. 3a) had a mean of $5.17 \times 10^{-7} \text{ m/s}$ and a variance of 2.0 (for $\ln K$), representing a strongly heterogeneous aquifer with sand-clay mixture. The initial entry pressure (before NAPL infiltration) ranged from 6.62×10^4 to $5.12 \times 10^6 \text{ Pa}$. This

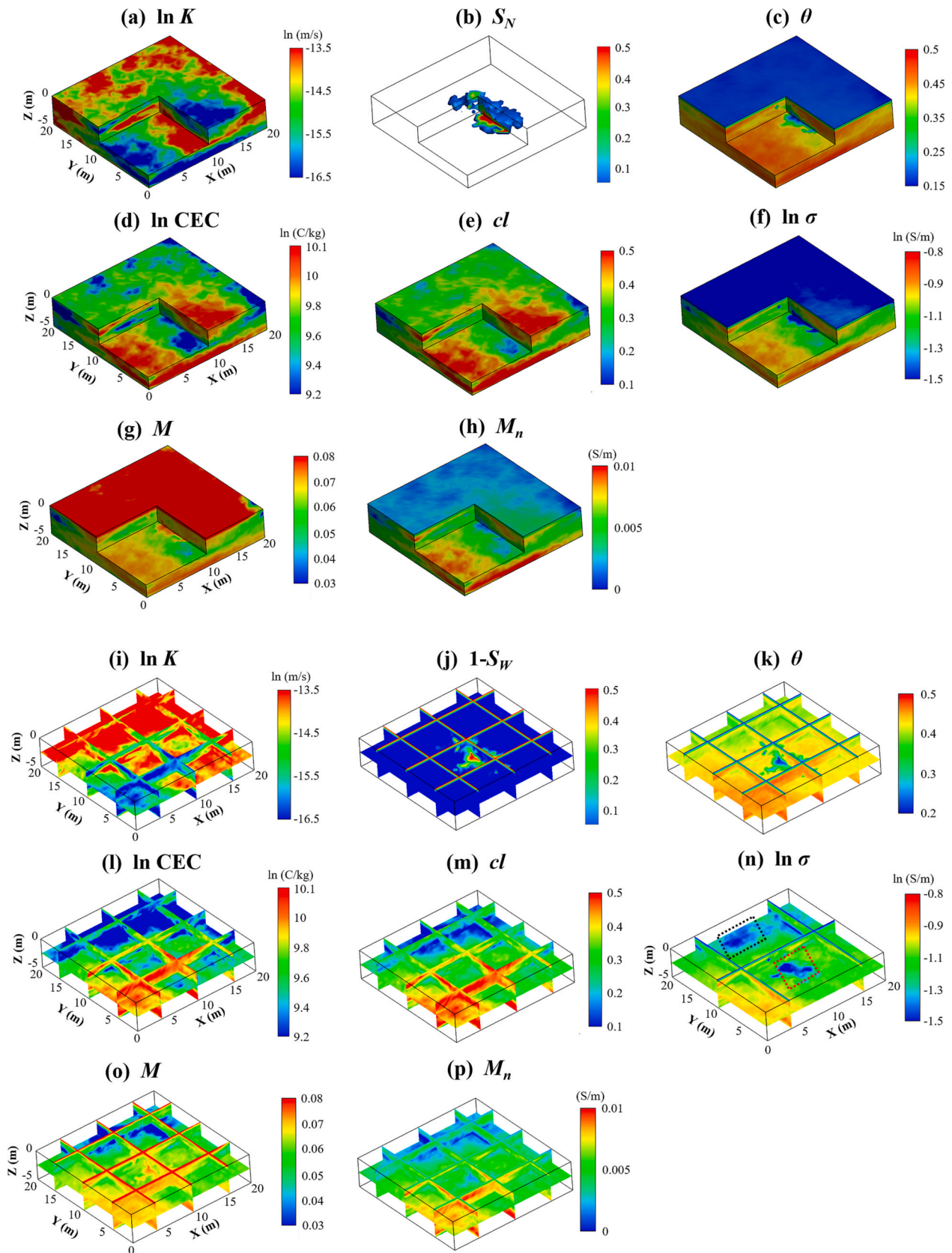


Fig. 3. The reference parameter fields of the synthetic clay-sand aquifer. (a)-(h): 3D surface images. (i)-(p): corresponding 3D slices. Note that K , S_N , θ , CEC, cl , σ , M , M_n represents the hydraulic conductivity, DNAPL saturation, volumetric water content, cation exchange capacity, clay content, electrical conductivity, chargeability, and normalized chargeability, respectively. The red and black dotted boxes in Fig. 3n denote the insulating DNAPL region and the background low-conductivity sandy soil region, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

range is similar with previous experimental data on clay-sand porous media (Brooks and Corey, 1964; Thomas et al., 1968; Schowalter, 1979). Based on the reference K field, the stochastic invasion-percolation modelling algorithm described in Section 2.1 (Step 4) is used to calculate the S_N distribution. The resulting reference DNAPL saturation field is shown in Fig. 3b. The reference SZs are pool-dominated (ganglia to pool ratio is 0.3). Notably, this work assumes the SZs contains only a single-component contaminant, namely oil-phase trichloroethylene (TCE).

Then, based on the coupled DNAPL-DCIP model presented in Section 2.1, the corresponding distributions of θ , CEC, cl , σ , M , and M_n are obtained (see Fig. 3 c-h). To better demonstrate the reference parameter fields, we also plot the corresponding y - z slices ($x = 10$ m) (see Fig. S1 in supporting information). The cl images (Fig. 3e and Fig. 3m) indicate that a large proportion of the domain is rich in clay grains ($cl > 0.4$). The slices of σ (Fig. 3n and Fig. S1f) show that using DC resistivity technique alone cannot discriminate between insulating DNAPL (the red dotted box in the center) and background low-conductivity sandy soil (the black dotted box). The slices of M (Fig. 3o) show that the chargeability is also controlled by both the S_W (S_N) and the clay content. In contrast, the normalized chargeability distribution highly resembled the spatial pattern of clay content (see Fig. 3m, Fig. 3p, Fig. S1e and Fig. S1h). This indicates that the M_n was mainly controlled by lithology (clay content) rather than DNAPL SZ, which was also observed in Almpanis et al. (2021). Therefore, the M_n may provide additional knowledge about the spatial variation in lithology, indicating that DCIP has the potential to discriminate the contributions of DNAPL presence and clay presence on electrical signals.

3.2. Hydrogeophysical experiment design

To obtain the characterization data, five intrusive boreholes were assigned in the vicinity of the SZ (grey cylinders in Fig. 4a). Fifty direct measurements of K/S_N were collected from different depths of these five boreholes ($z = -0.5, -1.0, \dots, -5.0$ m). To establish a site-specific empirical CEC- ϕ relationship, thirty-five ϕ measurements were also collected from these boreholes ($z = -0.5, -1.0, \dots, -3.5$ m). Fifty samples of dissolved DNAPL concentrations were collected from a downstream transect that is set orthogonal to the x -direction (see the five blue cylinders in Fig. 4a).

To record the DCIP response of DNAPL SZs in clay-sand mixtures, 180 electrodes were set on the ground surface (black rectangular points in Fig. 4a). Along each line, the inline electrode spacing was 1 m, while the offset distance (or interline spacing) between lines was 5 m. The DCIP measurements were recorded using a pole-dipole configuration (e.g., Chambers and Ogilvy, 2002), resulting in 6444 V and 6444 M_a

measurements. The sensitivity image of the DCIP survey is shown in Fig. 4b, which was computed following an accumulated-sensitivity-based method of Kemna (2000). As expected, the sensitivity decreased dramatically to the bottom, indicating significant loss of resolution in this region.

The multisource conditioning dataset (i.e., K , S_N , ϕ , V , M_a , and c) can be collected, with S_N , ϕ , V and M_a calculated with the coupled DNAPL-DCIP model, and c obtained through DNAPL-dissolution and solute-transport models described in supporting information (Text S4). Adding Gaussian noises with standard deviations of 0.2 ln (m/s), 0.005, 0.001 g/L, 0.15 V and 0.003 to the reference ln K , ϕ , c , V and M_a data, respectively, ensured more realistic noisy measurements. The noise for V and M_a corresponded to 1.5% of the maximum electrical-potential change and 5% of the average values of the apparent chargeabilities, respectively, which were typical levels of noise used in synthetic studies (e.g., Oldenburg and Li, 1994; Pollock and Cirpka, 2010; Camporese et al., 2015). As measurements of DNAPL saturation are difficult and prone to error, instead of using exact DNAPL saturation values, we use three categories to indicate no DNAPL regions ($S_N = 0$), low-saturation regions ($0 < S_N \leq 0.2$), and high-saturation regions ($S_N > 0.2$).

3.3. CVAE network training

The CVAE's structure for parameterization is shown in Fig. 2a and detailed in Table S1. The input data of training were 3D K and S_N images, with a dimension of $40 \times 40 \times 10 \times 2 = 32,000$. For CVAE training, the K/S_N reconstruction problem is posed as a classification task to represent the highly irregular K/S_N fields (Kang et al., 2021). Specifically, the continuous S_N values were separated into 10 discrete classes (i.e., $S_N = 0$, $0 < S_N < 0.06$, $0.06 \leq S_N < 0.12$, $0.12 \leq S_N < 0.18$, $0.18 \leq S_N < 0.30$, $0.30 \leq S_N < 0.40$, $0.40 \leq S_N < 0.50$, $0.50 \leq S_N < 0.60$, $0.60 \leq S_N < 0.70$, and $S_N \geq 0.70$). The resulting discrete S_N and K images were parameterized into the latent variables z through an encoder composed of six fully convolutional layers (see Table S1). Taking z as input, the K/S_N images can be reconstructed by the decoder.

The training set contained 40,000 K and S_N images (realizations) generated by the stochastic invasion-percolation algorithm (Step 4 in Section 2.1). Each K/S_N image has a dimension of $40 \times 40 \times 10 = 16,000$. Furthermore, another 4000 test realizations were generated to test the reconstruction performance of CVAE. The reference S_N and K images of the inversion (Fig. 3a-b) were not included in the training set. Instead, they were from the testing set. It took 10.6 h to generate these 44,000 K/S_N images, using 20 CPUs in parallel. Notably, at real-world sites, the longitudinal correlation length of the K field (λ_x, λ_z), the location of DNAPL release points (x_0, y_0, z_0), release rates (v_0), and the total release mass (M_0), would be unknown. Thus, all these seven

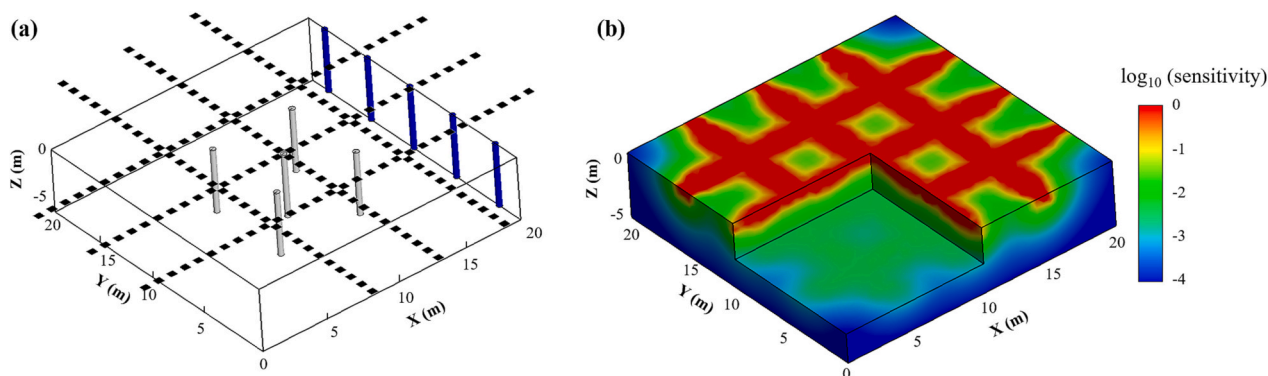


Fig. 4. (a) Sketch of boreholes and the electrodes distribution for DCIP. (b) Sensitivity map of the DCIP survey. Note that the grey and blue cylinders represent the boreholes in the vicinity of the SZ and the downstream boreholes for collecting aqueous-phase concentration data, respectively. The black rectangular points denote the electrodes on the ground surface for DCIP survey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

variables were treated as random variables when generating training realizations, *i.e.*, $\lambda_x \in (2.5, 7.5)$ m, $\lambda_z \in (0.5, 1.9)$ m, $x_0 \in (5, 15)$ cm, $y_0 \in (5, 15)$ m, $z_0 \in (-1, 0)$ m, $v_0 \in (5.4 \times 10^{-6}, 9.0 \times 10^{-6})$ m³/s, $M_0 \in (5.4, 9.0)$ tons. The ranges these variables were referred to previous studies on DNAPL contaminant modelling (*e.g.*, Maji et al., 2006; Koch and Nowak, 2016).

The network was trained by using the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 1×10^{-3} and a batch size of 32. The number of epochs was set as 200. The latent space dimension were chosen as 500. These hyperparameters (*e.g.*, number of epochs, learning rates, and batch size) were set by trial-and-error to test the performance of different architecture designs. The training took 4.6 h with an NVIDIA Tesla V100 GPU.

3.4. Setup of the synthetic cases

Table 2 summarizes the five synthetic cases demonstrated in this study. Case Geosta-DC and Geosta-DCIP, were conducted first to validate the benefits of extending from DC resistivity to DCIP. In Case Geosta-DC, the geostatistical method (PCGA) assimilated 6444 DC measurements (V) for σ imaging. Since the σ is controlled by both pore water (related to S_N) and lithology (related to K and clay content), one cannot infer K/S_N by using DC-estimated σ alone. In Case Geosta-DCIP, DC measurements (6444 V) were combined with IP chargeability data (6444 M_a) to estimate both K and S_N fields, following the procedure in Section 2.2. Thirty-five borehole measurements of K/ϕ were used to develop site-specific CEC- K/ϕ relationships. In Case Geosta-DCIP, the unknowns were 16,000 K and 16,000 S_W . Note that as the DCIP data suffer from diminished sensitivity with depth (see Fig. 4b), we only utilized the DCIP estimation in high-sensitivity regions for the subsequent inversion (*i.e.*, \log_{10} sensitivity > -3 , as suggested by Kemna, 2000). In addition, since DCIP cannot discriminate between S_G (saturation of gas phase) and S_N , it cannot provide information for the DNAPL in the unsaturated zone. Given its properties, the majority of DNAPL mass infiltrates below the water table, and therefore only DCIP-estimated S_N below the water table were used.

To demonstrate the benefit of integrating multisource datasets (*i.e.*, DCIP-estimated K/S_N and borehole data), the CVAE-ESMDA was used as joint inversion scheme instead of traditional geostatistical methods, since the CVAE can better represent the highly irregular S_N fields and improve the accuracy for DNAPL imaging. The cases CVAE-Bore, CVAE-DCIP, and CVAE-Joint were conducted, where different datasets were assimilated in the CVAE-ESMDA. The S_N and K fields were estimated using: (i) only borehole measurements (*i.e.*, direct measurements of K , S_N , and downstream dissolved-phase concentrations) in Case CVAE-Bore, (ii) only DCIP data in Case CVAE-DCIP, and (iii) both datasets in Case CVAE-Joint.

For Cases Geosta-DC and Geosta-DCIP, when applying PCGA for σ inversion, the number of principle components was chosen as 100. For the CVAE-ESMDA cases, the number of the iterations was set to be $N_I = 4$, and the ensemble size were set to be $N_e = 500$.

The normalized root-mean-square error (NRMSE) was used as the accuracy criterion for inversion.

Table 2

Parameter sets for the five cases of the numerical experiment.

Case	Observation Data	Unknowns
Geosta-DC	6444 V	16,000 σ
Geosta-DCIP	6444 $V/M_a + 35 \phi$	16,000 K/S_N
CVAE-DCIP	K/S_N from Geosta-DCIP	500 z
CVAE-Bore	50 $K/S_N/c$	500 z
CVAE-Joint	① K/S_N from Geosta-DCIP ②50 $K/S_N/c$	500 z

K , S_N , c , ϕ , V , M_a represent the measurement of hydraulic conductivity, DNAPL saturation, downstream dissolved-phase concentration, porosity, electrical potential and apparent chargeability, respectively. z is the latent variables of the CVAE network.

$$NRMSE = \frac{1}{x_{max} - x_{min}} \sqrt{\frac{1}{N_u} \sum_{j=1}^{N_u} (x_j - \tilde{x}_j)^2} \quad (12)$$

where N_u is the number of unknowns in the research domain. x and \tilde{x} are the actual and the best estimate of the unknown parameters (*i.e.*, $\ln K$ and S_N), respectively. x_{max} and x_{min} are the maximum and minimum values of the parameters, respectively.

4. Inversion results

4.1. Geostatistical inversion: DC-only versus combined DCIP

Figure 5 demonstrates the $\ln \sigma$ and M_n fields estimated from cases Geosta-DC and Geosta-DCIP shown in Table 2. In Case Geosta-DC, standalone DC resistivity can only provide the σ estimation (blue dotted frame in Fig. 5). The DC-estimated $\ln \sigma$ can reconstruct the main spatial pattern of the reference electrical conductivity field (*e.g.*, the low conductivity near the ground surface and the resistive sandy region at $x = 7-13$ m, $y = 16-20$ m), but cannot reveal a significant low-conductivity anomaly in the DNAPL region, partly due to the deviation of the true spatial distribution of $\ln \sigma$ from the Gaussian prior covariance model used in PCGA (Kang et al., 2020). From the x - y slice ($z = -2.5$ m) of the estimated $\ln \sigma$ (see Fig. S2), a region of slightly lower conductivity was found that may be caused by DNAPL presence (*i.e.*, dotted region in Fig. S2). However, using σ alone cannot discriminate between insulating DNAPL and background resistive sandy region, since σ is governed by both pore water and lithology.

In Case Geosta-DCIP, both DC and IP measurements (*i.e.*, V and M_a) were used with PCGA and IP linear inversion method to estimate the $\ln \sigma$ and M_n fields. Fig. 5 demonstrates the DCIP-estimated M_n can recover the large-scale features of the true M_n field, while it overestimates the M_n values at $x = 0-10$ m, $y = 0-8$ m. Based on the DCIP-estimated σ and M_n fields, the distributions of θ and CEC can be computed using the petrophysical models described in Eqs. 9 and 10. The DCIP-derived θ can identify the location of the DNAPL SZ (black dotted rectangles in the third column of Fig. 5). Moreover, the estimated CEC successfully delineated the clayey regions (corresponding to high CEC values) and sandy regions (low CEC values). Note that the NRMSEs of M_n , θ , and CEC are higher than the NRSME of $\ln \sigma$, due to a propagation of error from σ to derived variables (*i.e.*, θ and CEC).

For Case Geosta-DCIP, we then developed site-specific empirical relationships of CEC- ϕ and CEC- K by fitting the DCIP-derived CEC to the borehole measured ϕ/K . Note that only 35 pairs of CEC- K/ϕ data were used in the shallow part of the borehole ($z \geq -3.5$ m, corresponding to the high sensitivity region in Fig. 4b) to ensure reliability of the DCIP-derived CEC. Similar calibration strategies have been applied to determine θ - σ relationship, *i.e.*, calibrating field-based inverted electrical conductivity with *in situ* measurements of soil moisture (*e.g.*, Garré et al., 2013). Fig. 6 presents the fitting results for CEC- ϕ/K relationships. Linear relationships between CEC and ϕ/K were obtained with coefficients of determination (R^2) higher than 0.7. Although the calibrated CEC- ϕ/K relationships were based on synthetic data, studies based on laboratory experiments also show that CEC is highly related to the permeability, porosity, and clay content of the porous medium (Sen et al., 1990; Revil et al., 2017a, 2017b, 2020; Soueid Ahmed et al., 2020).

Using the established CEC- K/ϕ relationships, the DCIP-derived CEC/ θ can be converted to K and S_W ($S_W = \theta/\phi$). The corresponding K/S_N estimations are shown in Fig. 7. The DCIP-estimated K field reflected the large-scale spatial features of the reference K distribution (*e.g.*, the high-permeability regions at $x = 5-20$ m, $y = 12-20$ m, $z = -2 \sim -4$ m), but overestimated the extent of the low-permeability regions ($x = 0-13$ m, $y = 0-10$ m, $z = -2 \sim -4$ m). The estimated $(1-S_W)$ field successfully revealed the location of high S_N region and the unsaturated layer near

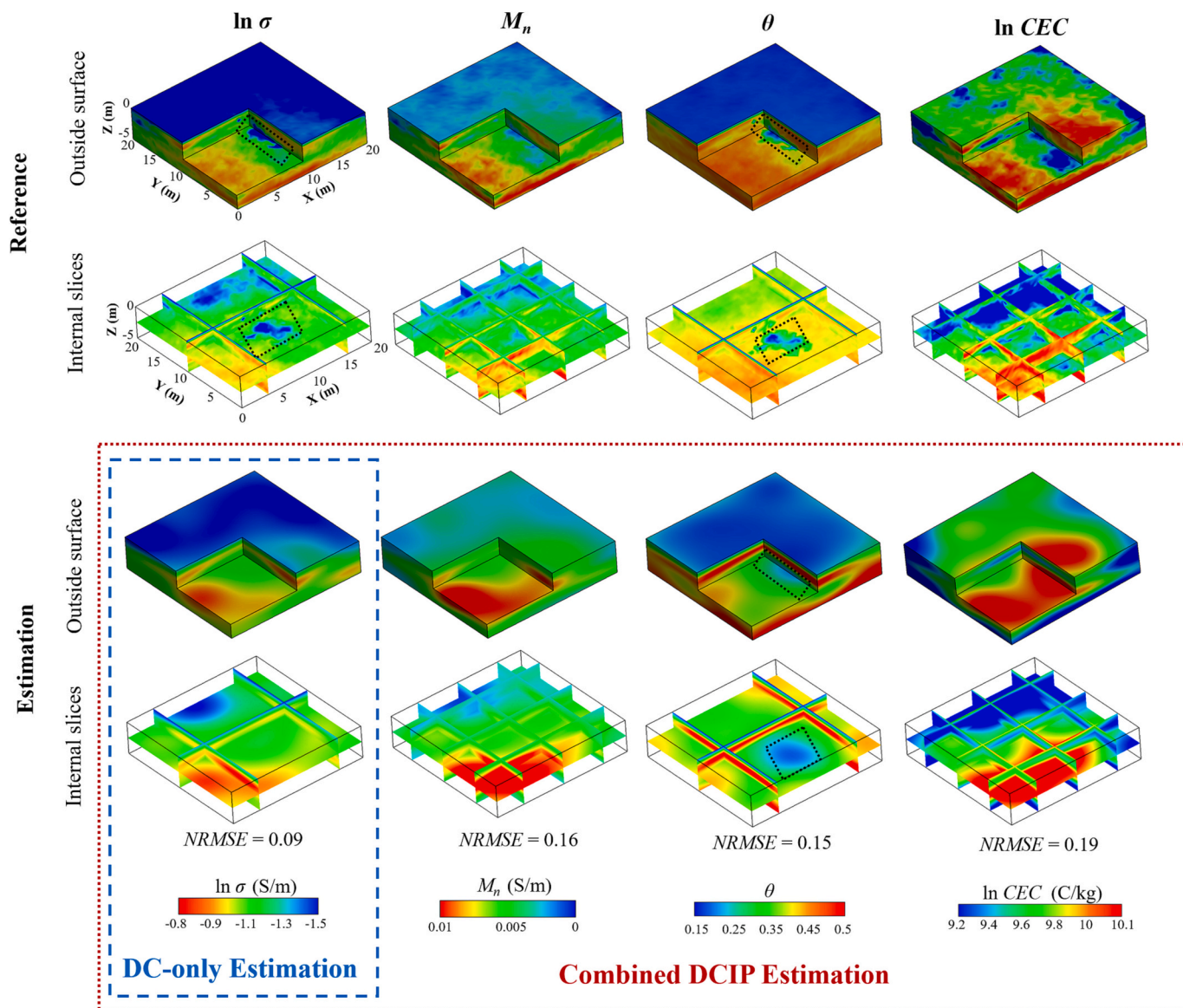


Fig. 5. The reference and estimated electrical conductivity (σ), normalized chargeability (M_n), water content (θ), cation exchange capacity (CEC) from Cases Geosta-DCIP (using both resistivity and chargeability data) and Case Geosta-DC (using DC resistivity alone). Note that the DC resistivity may only provide the estimation of σ (i.e., the first column in Fig. 5), while the combined DCIP can provide the estimations of σ , M_n , θ and CEC (i.e., the first to fourth columns). The black dotted boxes denote the DNAPL occupied region.

the ground surface. Nevertheless, the estimated $(1-S_w)$ field is relatively smooth, generally overestimating the spread of DNAPL SZ. This is mainly due to three factors: (1) incomplete information contained in the DCIP data (insufficient for high-resolution DNAPL imaging); (2) the simplified prior covariance model considered in geostatistical inversion; (3) approximation error from the CEC- K/ϕ relationships.

We compared the results from Cases Geosta-DC and Geosta-DC in Fig. 5. It indicates that standalone DC resistivity failed to infer the DNAPL SZs in complex clay-sand environments, while combined DCIP can favorably separate the contributions of insulating DNAPL from background resistive sandy soils on the electrical signals, and reconstruct the large-scale features of both K and S_N fields (Fig. 7). Since the DCIP method is easy-to-implement and only requires limited *in situ* core samples for petrophysical calibration, one can apply this approach before the intrusive borehole-based investigation, to get a preliminary knowledge of the DNAPL distribution and subsurface heterogeneity.

4.2. Joint inversion of DCIP and borehole data

Figure 8 shows the estimated K and S_N fields for Cases CVAE-Bore, CVAE-DCIP, and CVAE-Joint. The grey cylinders in the 3D S_N images represent the locations of the boreholes. Fig. S4 in supporting information depicts the corresponding posterior standard deviations for these cases. Table 3 summarized the NRMSEs of each case. In Case CVAE-Bore, only borehole data (i.e., K , S_N and c) were used. The estimated K revealed the main pattern of the heterogeneous conductivity field (Fig. 8). However, borehole data alone failed to reconstruct the shape and size of the reference SZA due to limited spatial coverage of the sampling. Specifically, pure-phase DNAPL was not detected in any boreholes except the borehole in the center of the domain ($x = 10$ m, $y = 10$ m). Therefore, the estimated SZA was limited to the region within these five boreholes. Although the CVAE can provide an additional physics-based information (i.e., spatial pattern of DNAPL SZA) to constrain the estimation, there was still a lack of spatially continuous information to resolve such a complex SZA.

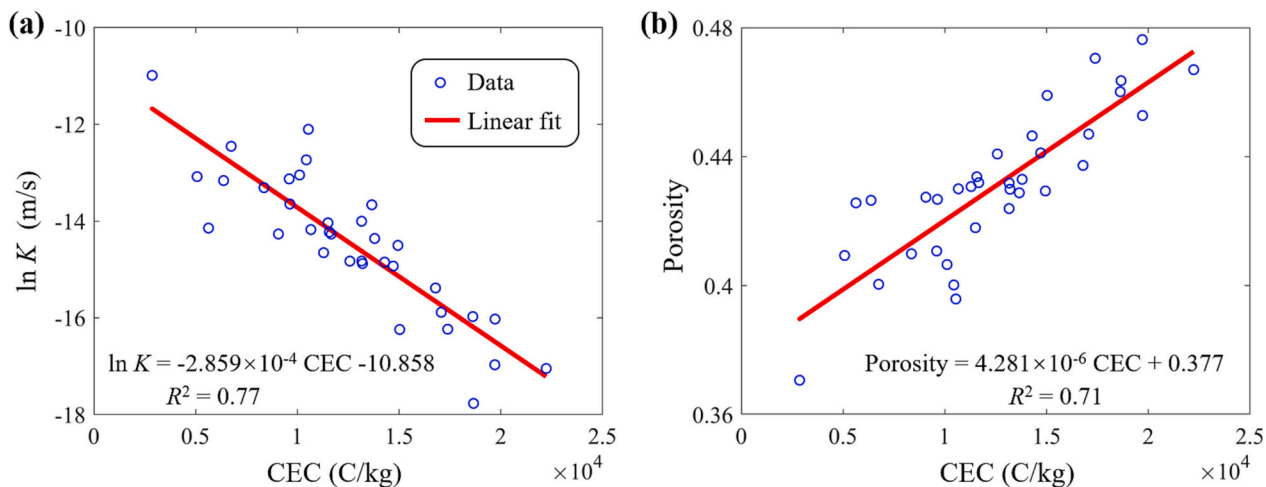


Fig. 6. Hydraulic conductivity and porosity versus CEC. (a) Measured $\ln K$ versus DCIP-derived CEC. (b) Measured porosity versus DCIP-derived CEC.

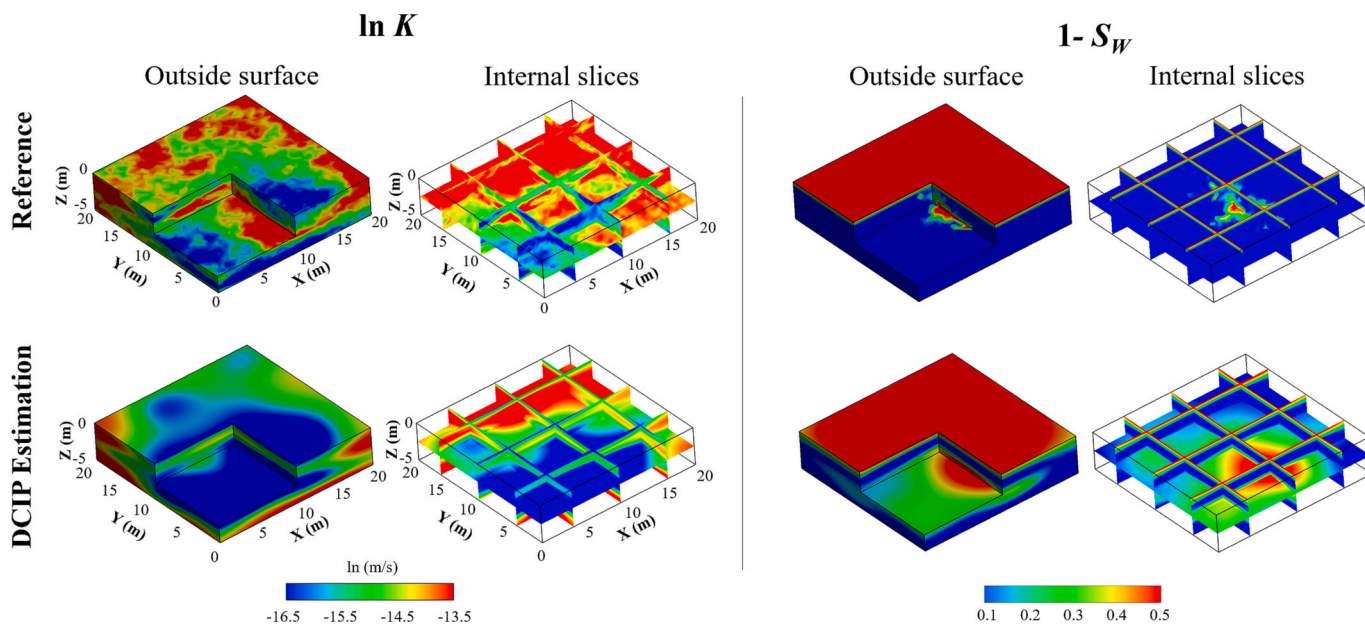


Fig. 7. The reference and estimated $\ln K$ and $(1 - S_w)$ from geostatistical inversion of DCIP data (Case Geosta-DCIP). Note that S_w represents the water saturation.

In CVAE-DCIP, the DCIP-derived K/S_N dataset was used as conditioning data to reconstruct the K and S_N fields. The best estimate of K can recover large-scale features of reference K distribution, and the estimated S_N can capture the extent of highly saturated DNAPL ($S_N > 0.4$) on the y -axis. Moreover, the CVAE-DCIP-estimated K/S_N can achieve a slightly higher accuracy than the results from Case Geosta-DCIP (see Fig. 7 and the *NRMSE* in Table 3). This superiority is mainly because: (1) CVAE can better represent the highly irregular K/S_N fields; (2) in CVAE-DCIP, only the more reliable DCIP-derived K/S_N data from high-sensitivity region (see Fig. 4b) were used to avoid artifacts caused by diminished sensitivity. Nevertheless, results from CVAE-DCIP still overestimated the extent of low-permeability region, and the spread of SZ in general. In other words, combined DCIP cannot accurately capture the shape and extent of the SZ on its own due to its lower resolution.

In Case CVAE-Joint, all three types of information for the inversion were integrated, i.e., DCIP-estimated K/S_N images, borehole measurements of $K/S_N/C$, and the accurate representation of K/S_N provided by CVAE. The results indicates that the CVAE-Joint reconstructed the K/S_N field with more details compared to using a single dataset (see Fig. 8). Among all three information sources, DCIP provided low-resolution but

spatially continuous information. As a supplement, a small amount of high-accuracy borehole data served as hard constraints for the DCIP-derived overspread SZ. Finally, the representation of K/S_N provided by CVAE can force the SZA estimation to honor the spatial pattern of DNAPL infiltration (Kang et al., 2021). Therefore, by utilizing the three complementary information, a SZ pattern similar to the reference can be obtained. The *NRMSE* of the K/S_N estimations (Table 3) also indicates the superior performance of CVAE-Joint.

Seven spatial metrics of S_N (i.e., DNAPL mass and spatial moments) were also evaluated for further comparative analysis of CVAE-Bore, CVAE-DCIP and CVAE-Joint (see Fig. 9). Fig. 9a demonstrates that using borehole data alone tended to underestimate the DNAPL mass due to the sparsity of sampling data, while using DCIP data alone led to a significant overestimation of DNAPL mass. In contrast, CVAE-Joint can produce a more accurate estimation of DNAPL mass, along with a lower uncertainty level (Fig. 9a). CVAE-Joint also better captured the location and spread of the SZ (see the first and second spatial moments of mass in Fig. 9 b-g). These metrics are important for remediation strategy designs (Christ et al., 2006, 2010).

Regarding the uncertainty of the estimates (see Fig. S4), the

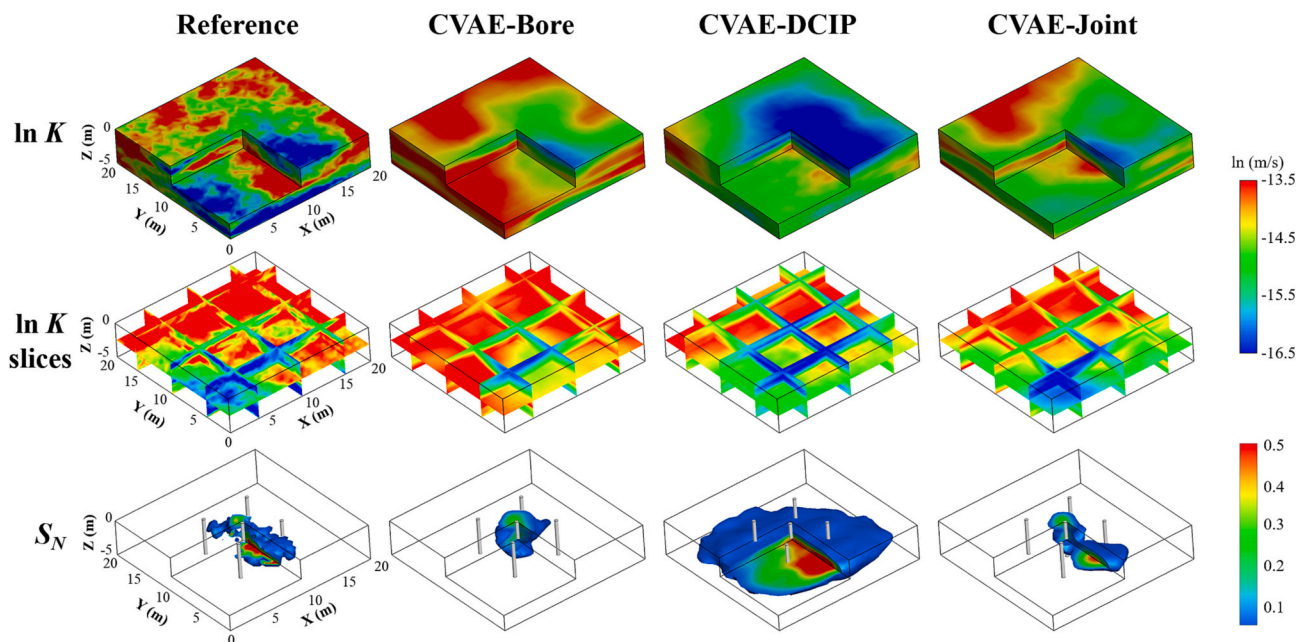


Fig. 8. The estimated ensemble means of K and S_N for Cases CVAE-Bore, CVAE-DCIP, and CVAE-Joint. Note that the grey cylinders in S_N images represent the locations of intrusive boreholes.

Table 3
 NRMSE of the estimated K/S_N and computational times for each case.

Case	$NRMSE_S$	$NRMSE_K$	Computational time (min)
Geosta-DC	/	/	241
Geosta-DCIP	0.25	0.53	243
CVAE-DCIP	0.19	0.41	1
CVAE-Bore	0.09	0.36	315
CVAE-Joint	0.08	0.31	326

uncertainty of K/S_N produced by joint inversion was lower than that from using single dataset (e.g., borehole data), indicating that the additional DCIP data can help reduce the uncertainty level and alleviate the non-uniqueness issue. In addition, although the uncertainty in empirical CEC- ϕ relationship (Fig. 6) may translate to uncertainty in the DCIP-estimated S_N through $S_N = 1-\theta/\phi$, the DCIP-estimated K/S_N are treated with a relatively lower weight in the joint inversion. Therefore, by considering high-accuracy borehole data (K , S_N and c) in the joint inversion, one can better constrain the estimated K/S_N and reduce the

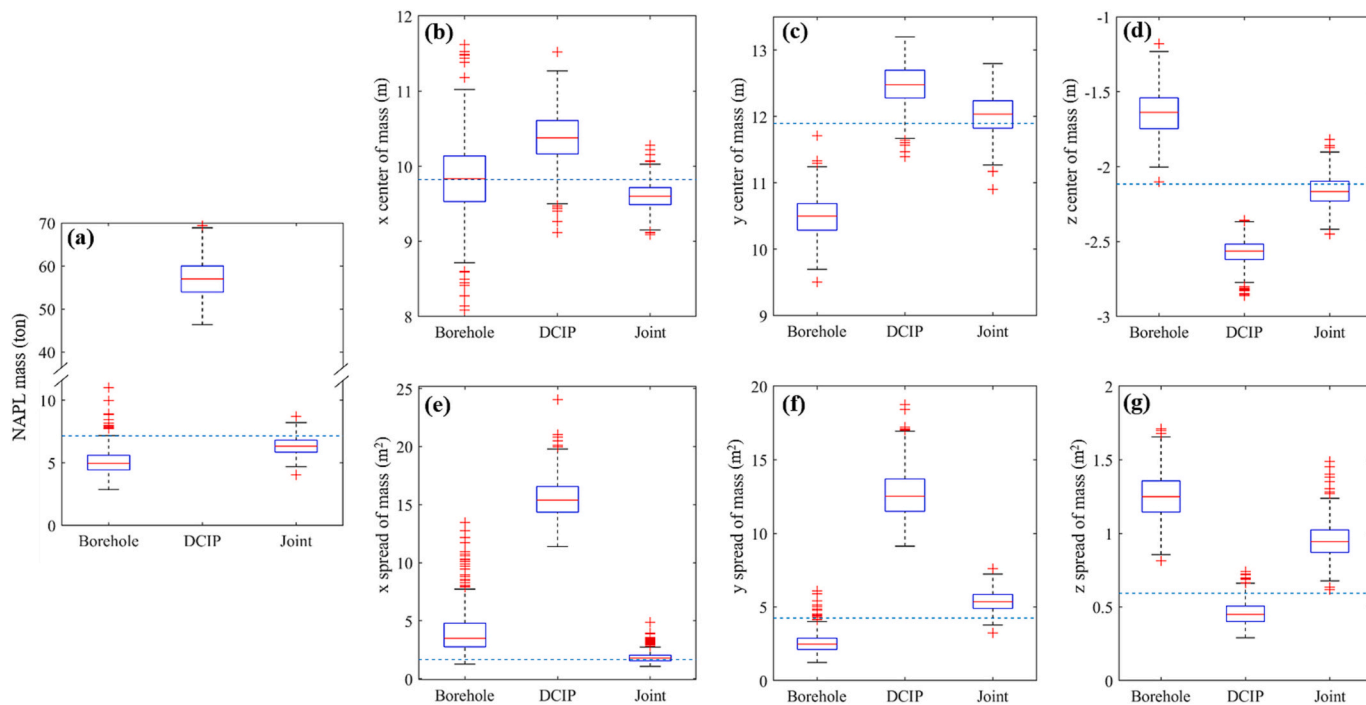


Fig. 9. Estimated versus reference SZ metrics by using different types of data. (a) NAPL mass (b) x center of mass (c) y center of mass (d) z center of mass (e) x spread of mass (f) y spread of mass (g) z spread of mass. The horizontal blue dotted lines denote the true (reference) metrics. The boxplots shows the distribution of the SZ metrics for all posterior realizations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

uncertainty raised by petrophysical fitting (*i.e.*, CEC- ϕ).

Nonuniqueness is one of the major challenges of the high-dimensional inverse problem. The proposed CVAE-based joint inversion of DCIP and borehole data can effectively alleviate the non-uniqueness issue from the following aspects:

- (1) **Sufficient data provided by DCIP:** the non-intrusive geophysical survey can provide much more data (6444 DCIP data) than intrusive boreholes (only 50 K/S_N data).
- (2) **An accurate prior provided by CVAE:** the deep learning-based parameterization can provide a physics-based prior, which can capture the physics of DNAPL infiltration and accurately represent the complex spatial pattern of the SZs. This physics-based prior can serve as an additional constraint to help the estimate converge to a narrow parameter space.

5. Conclusions

In this work, an integrated hydrogeophysical approach is proposed to characterize DNAPL saturation (S_N) along with the distribution of the hydraulic conductivity (K) in complex clay-sand environments. In clay-sand environments, the ability of the popular DC resistivity technique is limited due to its difficulty in discriminating DNAPL from surrounding soils. To address this limitation, DC resistivity and TDIP were combined to image the large-scale features of the S_N and K fields. Based on the image of the S_N and K fields estimated from DCIP, borehole measurements (*i.e.*, K , S_N and downstream aqueous-phase concentration data) were incorporated within a deep learning-based inversion method to further improve the accuracy of S_N and K characterization.

To evaluate the effectiveness of the proposed methodology, we performed numerical experiments in a 3D heterogeneous clay-sand aquifer with an intricate DNAPL SZA. While standalone DC resistivity failed to infer the DNAPL SZs from clay-sand environments, combined DCIP was better able to discriminate between insulating DNAPL and surrounding resistive sandy soils. DCIP was shown to successfully reveal the large-scale spatial features of K/S_N fields; however, due to low sensitivity it was found to miss the finer features. Comparing the usage of different types of dataset, we found that combining DCIP and borehole measurements can more accurately characterize the SZA compared to using a single dataset. DCIP provides low-resolution but spatially continuous information, with high-accuracy but spatially sparse borehole data serving as hard constraints for the overestimated DCIP-derived SZ. Moreover, the deep learning-based joint inversion framework (CVAE-ESMDA) can satisfactorily incorporate the information content from multisource data and the physically constrained prior provided by the trained CVAE, which further enhances the DNAPL image in clay-sand environments.

Although this study is built on numerical experiments, the source zones and heterogeneous sand-clay soils resemble real-world conditions. In addition, the hydrogeophysical setups are practical for real-world fields, as the DCIP survey only required surface-based electrodes and only five boreholes were considered for invasive sampling. However, real-world situations may be more complex than numerical experiments. For example, the wettability of DNAPLs may change from non-wetting to wetting. Surfactants may be produced by the activity of microorganisms (Ron and Rosenberg, 2001) which can cause complex conduction and polarization mechanisms that may introduce uncertainty into the petrophysical models for converting σ/M_n to θ /CEC. This petrophysical uncertainty may affect the hydrogeophysical inversion results. Future work should further investigate the petrophysical relationship between chargeability and DNAPL saturation under different wetting conditions. Alternatively, this petrophysical uncertainty can be incorporated in a Bayesian framework. In other words, one can infer both the unknowns and the petrophysical uncertainty (Brunetti and Linde, 2018).

In addition, the assumption of uniform pore water conductivity (σ_w) may be violated when the SZs undergo extensive dechlorination (*e.g.*,

the *in-situ* chemical oxidation remediation process). In this case, one can use the borehole-measured σ_w to interpolate the spatial distribution of σ_w , and then apply the petrophysical relationships (Eqs. 9–10) to convert the DCIP response to K/S_N estimation. Although this interpolation may introduce some uncertainty for the joint hydrogeophysical inversion results, the DCIP still holds the potential to help locate the highly saturated DNAPL region in clay-sand environments.

CRedit authorship contribution statement

Xueyuan Kang: Conceptualization, Methodology, Software, Investigation, Data curation, Writing – original draft. **Christopher Power:** Visualization, Software, Writing – review & editing. **Amalia Kokkinaki:** Writing – review & editing. **André Revil:** Writing – review & editing. **Jichun Wu:** Supervision, Resources, Writing – review & editing, Funding acquisition. **Xiaoqing Shi:** Supervision, Resources, Writing – review & editing. **Yaping Deng:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2022YFC3703101) and the National Natural Science Foundation of China (41977157, 42272276, and 42202267). We are grateful to the Editor, Christian Griebler and anonymous reviewers for their insightful comments and suggestions, which significantly improve the quality of this work.

References

- Abdulsamad, F., Revil, A., Soueid Ahmed, A., Coperey, A., Karaoulis, M., Nicaise, S., Peyras, L., 2019a. Induced polarization tomography applied to the detection and the monitoring of leaks in embankments. *Eng. Geol.* 254, 89–101. <https://doi.org/10.1016/j.enggeo.2019.04.001>.
- Abdulsamad, F., Revil, A., Ghorbani, A., Toy, V., Kirilova, M., Coperey, A., Duvalidard, P. A., Ménard, G., Ravanel, L., 2019b. Complex conductivity of graphitic schists and sandstones. *J. Geophys. Res. Solid Earth* 124 (8), 8223–8249. <https://doi.org/10.1029/2019JB017628>.
- Abdulsamad, F., Revil, A., Prime, N., Gnonnoue, P.Y., Schmutz, M., Plé, O., 2020. Complex conductivity of rammed earth. *Eng. Geol.* 273, 105697 <https://doi.org/10.1016/j.enggeo.2020.105697>.
- Abriola, L.M., Christ, J.A., Pennell, K.D., Ramsburg, C.A., 2012. Source remediation challenges. In: Kitanidis, P.K., McCarly, P.L. (Eds.), *Delivery and Mixing in the Subsurface: Processes and Design Principles for In Situ Remediation*. Springer, New York, NY, pp. 239–276.
- Akyol, N.H., Yolcubal, I., 2013. Oxidation of nonaqueous phase trichloroethylene with permanganate in Epikarst. *Water Air Soil Pollut.* 224 (5) <https://doi.org/10.1007/s11270-013-1573-6>.
- Akyol, N.H., Demiray, Z., Copty, N.K., Acilioğlu, T., Akyol, G., 2023. Mass-flux reduction and mass removal relationships (MFR-MR) of pool-dominated multicomponent DNAPLs in a heterogeneous geological system. *Global NEST J.* 25 (4), 43–55. <https://doi.org/10.30955/gnj.004289>.
- Almpanis, A., Gerhard, J., Power, C., 2021. Mapping and monitoring of DNAPL source zones with combined direct current resistivity and induced polarization: a field-scale numerical investigation. *Water Resour. Res.* 57 <https://doi.org/10.1029/2021WR031366> e2021WR031366.
- Archie, G.E., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics. *Transact. Am. Inst. Min. Metall. Eng.* 146, 54–61.
- Arshadi, M., Kaluzu, C.D.P., Miller, E.L., Abriola, L.M., 2020. Subsurface source zone characterization and uncertainty quantification using discriminative random fields. *Water Resour. Res.* 56 <https://doi.org/10.1029/2019WR026481> e2019WR026481.
- Bear, J., 1988. *Dynamics of Fluids in Porous Media*, p. 764. Dover, Mineola, N.Y.
- Binley, A., Cassiani, G., Middleton, R., Winship, P., 2002. Vadose zone flow model parameterisation using cross-borehole radar and resistivity imaging. *J. Hydrol.* 267 (3–4), 147–159.

- Brooks, A.N., Corey, A.T., 1964. Hydraulic properties of porous media. In: Corey, A.T., Dils, R.E., Yevjevich, V.M. (Eds.), *Hydrology Papers*. Colorado State University, Fort Collins, CO, pp. 1–25.
- Brunetti, C., Linde, N., 2018. Impact of petrophysical uncertainty on Bayesian hydrogeophysical inversion and model selection. *Adv. Water Resour.* 111, 346–359. <https://doi.org/10.1016/j.advwatres.2017.11.028>.
- Camporese, M., Cassiani, G., Deiana, R., Salandini, P., Binley, A., 2015. Coupled and uncoupled hydrogeophysical inversions using ensemble Kalman filter assimilation of ERT-monitored tracer test data. *Water Resour. Res.* 51 (5), 3277–3291. <https://doi.org/10.1002/2014WR016017>.
- Chambers, J., Ogilvy, O., 2002. 3D electrical imaging of known targets at a controlled environmental test site. *Environ. Geol.* 41 (6), 690–704. <https://doi.org/10.1007/s00254-001-0452-4>.
- Chambers, J.E., Wilkinson, P.B., Wealhall, G.P., Loke, M.H., Dearden, R., Wilson, R., Allen, D., Ogilvy, R.D., 2010. Hydrogeophysical imaging of deposit heterogeneity and groundwater chemistry changes during DNAPL source zone bioremediation. *J. Contam. Hydrol.* 118 (1–2), 43–61. <https://doi.org/10.1016/j.jconhyd.2010.07.001>.
- Chauveteau, G., Zaitoun, A., 1981. Basic rheological behavior of xanthan polysaccharide solutions in porous media: Effect of pore size and polymer concentration. In: Fayers, F.J. (Ed.), *Enhanced Oil Recovery*. Elsevier, New York, pp. 197–212.
- Chilindar, G.V., 1964. Relationship between porosity, permeability and grain size distribution of sands and sandstones. In: Van Straaten, L.M.J.U. (Ed.), *Deltaic and Shallow Marine Deposits*, vol. I. Elsevier, New York, pp. 71–75.
- Christ, J.A., Ramsburg, C.A., Pennell, K.D., Abriola, L.M., 2006. Estimating mass discharge from dense nonaqueous phase liquid source zones using upscaled mass transfer coefficients: an evaluation using multiphase numerical simulations. *Water Resour. Res.* 42 (W1142011) <https://doi.org/10.1029/2006WR004886>.
- Christ, J.A., Ramsburg, C.A., Pennell, K.D., Abriola, L.M., 2010. Predicting DNAPL mass discharge from pool-dominated source zones. *J. Contam. Hydrol.* 114 (1–4), 18–34. <https://doi.org/10.1016/j.jconhyd.2010.02.005>.
- Cohen, R.M., Mercer, J.W., Matthews, J., 1993. *DNAPL Site Evaluation*. In: Smoley, C.K. (Ed.), CRC Press, Fla, Boca Raton.
- Crestani, E., Camporese, M., Salandini, P., 2015. Assessment of hydraulic conductivity distributions through assimilation of travel time data from ERT-monitored tracer tests. *Adv. Water Resour.* 84, 23–36. <https://doi.org/10.1016/j.advwatres.2015.07.022>.
- Deng, Y., Shi, X., Revil, A., Wu, J., Ghorbani, A., 2018. Complex conductivity of oil-contaminated clayey soils. *J. Hydrol.* 561, 930–942. <https://doi.org/10.1016/j.jhydrol.2018.04.055>.
- Deng, Y., Shi, X., Zhang, Z., Sun, Y., Wu, J., Qian, J., 2020. Application of spectral induced polarization for characterizing surfactant-enhanced DNAPL remediation in laboratory column experiments. *J. Contam. Hydrol.* 230, 103603 <https://doi.org/10.1016/j.jconhyd.2020.103603>.
- Dietrich, C.R., Newsam, G.N., 1993. A fast and exact method for multidimensional gaussian stochastic simulations. *Water Resour. Res.* 29 (8), 2861–2869. <https://doi.org/10.1029/93WR01070>.
- Emerick, A.A., Reynolds, A.C., 2013. Ensemble smoother with multiple data assimilation. *Comput. Geosci.* 55 (3), 3–15.
- Garré, S., Coteur, I., Wongleechareon, C., Kongkaew, T., Diels, J., Vanderborght, J., 2013. Noninvasive monitoring of soil water dynamics in mixed cropping systems: a case study in Ratchaburi Province, Thailand. *Vadose Zone J.* 12 (2), j2012-j2129. <https://doi.org/10.2136/vzj2012.0129>.
- Glass, R.J., Conrad, S.H., Yarrington, L., 2001. Gravity-destabilized nonwetting phase invasion in macroheterogeneous porous media: near-pore-scale macro modified invasion percolation simulation of experiments. *Water Resour. Res.* 37 (5), 1197–1207. <https://doi.org/10.1029/2000WR900294>.
- Hinnell, A.C., Ferré, T., Vrugt, J.A., Huisman, J.A., Moysey, S., Rings, J., Kowalsky, M.B., 2010. Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. *Water Resour. Res.* 46 (4), W00D40.
- Jardani, A., Revil, A., Dupont, J.P., 2013. Stochastic joint inversion of hydrogeophysical data for salt tracer test monitoring and hydraulic conductivity imaging. *Adv. Water Resour.* 52, 62–77. <https://doi.org/10.1016/j.advwatres.2012.08.005>.
- Johnson, D.L., Plona, T.J., Kojima, H., 1986. Probing porous media with 1st sound, 2nd sound, 4th sound and 3rd sound. In: Jayanthi, R., Banavar, J., Winkler, K.W. (Eds.), *Physics and Chemistry of Porous Media*, vol. II. AIP, New York, pp. 243–277.
- Kang, X., Kokkinaki, A., Kitanidis, P.K., Shi, X., Revil, A., Lee, J., Soueid Ahmed, A., Wu, J., 2020. Improved characterization of DNAPL source zones via sequential hydrogeophysical inversion of hydraulic-head, self-potential and partitioning-tracer data. *Water Resour. Res.* 56 <https://doi.org/10.1029/2020WR027627>.
- Kang, X., Kokkinaki, A., Kitanidis, P.K., Shi, X., Lee, J., Mo, S., Wu, J., 2021. Hydrogeophysical characterization of nonstationary DNAPL source zones by integrating a convolutional Variational autoencoder and ensemble smoother. *Water Resour. Res.* 57 <https://doi.org/10.1029/2020WR028538>.
- Karaoglu, A.G., Copty, N.K., Akyol, N.H., Kilavuz, S.A., Babaei, M., 2019. Experiments and sensitivity coefficients analysis for multiphase flow model calibration of enhanced DNAPL dissolution. *J. Contam. Hydrol.* 225, 103515 <https://doi.org/10.1016/j.jconhyd.2019.103515>.
- Kemna, A., 2000. *Tomographic Inversion of Complex Resistivity: Theory and Application*. Der Andere Verlag, Osnabrück.
- Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. *arXiv e-prints*, arXiv:1412.6980.
- Kitanidis, P.K., Lee, J., 2014. Principal component geostatistical approach for large-dimensional inverse problems. *Water Resour. Res.* 50 (7), 5428–5443. <https://doi.org/10.1002/2013WR014630>.
- Koch, J., Nowak, W., 2015. Predicting DNAPL mass discharge and contaminated site longevity probabilities: conceptual model and high-resolution stochastic simulation. *Water Resour. Res.* 51 (2), 806–831. <https://doi.org/10.1002/2014WR015478>.
- Koch, J., Nowak, W., 2016. Identification of contaminant source architectures—a statistical inversion that emulates multiphase physics in a computationally practicable manner. *Water Resour. Res.* 52 (2), 1009–1025. <https://doi.org/10.1002/2015WR017894>.
- Kueper, B.H., Stroo, H.F., Vogel, C.M., Ward, C.H., 2014. *Chlorinated Solvent Source Zone Remediation*. Springer, New York.
- Lee, J., Kitanidis, P.K., 2014. Large-scale hydraulic tomography and joint inversion of head and tracer data using the principal component geostatistical approach (PCGA). *Water Resour. Res.* 50 (7), 5410–5427. <https://doi.org/10.1002/2014WR015483>.
- Lenhard, R.J., Parker, J.C., Kaluarachchi, J.J., 1989. A model for hysteretic constitutive relations governing multiphase flow: 3. Refinements and numerical simulations. *Water Resour. Res.* 25 (7), 1727–1736.
- 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.* Solid Earth 127 (3). <https://doi.org/10.1029/2021JB022581>.
- Maji, R., Sudicky, E.A., Panday, S., Teutsch, G., 2006. Transition probability/Markov chain analyses of DNAPL source zones and plumes. *Ground Water* 44 (6), 853–863. <https://doi.org/10.1111/j.1745-6584.2005.00194.x>.
- Maurya, P.K., Balbarini, N., Møller, I., Ronde, V., Christiansen, A.V., Bjerg, P.L., Auken, E., Fiandaca, G., 2018. Subsurface imaging of water electrical conductivity, hydraulic permeability and lithology at contaminated sites by induced polarization. *Geophys. J. Int.* 213 (2), 770–785. <https://doi.org/10.1093/gji/gyy018>.
- Morrissey, F.A., Grismer, M.E., 1999. Kinetics of volatile organic compound sorption/desorption on clay minerals. *J. Contam. Hydrol.* 36 (3–4), 291–312.
- Mumford, K.G., Bryck, S., Kueper, B.H., Mancini, S., Kavanaugh, M., Reynolds, D., 2022. Virtual site investigation to evaluate conceptual site model development at DNAPL-impacted sites. *Groundwater Monitor. Remed.* 42 (3), 44–58. <https://doi.org/10.1111/gwmr.12537>.
- Nivorlis, A., Dahlin, T., Rossi, M., Höglund, N., Sparrenbom, C., 2019. Multidisciplinary characterization of chlorinated solvents contamination and in-situ remediation with the use of the direct current resistivity and time-domain induced polarization tomography. *Geosciences* 9 (12), 487. <https://doi.org/10.3390/geosciences9120487>.
- Ntarlagiannis, D., Robinson, J., Soupios, P., Slater, L., 2016. Field-scale electrical geophysics over an olive oil mill waste deposition site: evaluating the information content of resistivity versus induced polarization (IP) images for delineating the spatial extent of organic contamination. *J. Appl. Geophys.* 135, 418–426. <https://doi.org/10.1016/j.jappgeo.2016.01.017>.
- Oldenburg, D.W., Li, Y.G., 1994. Inversion of induced polarization data. *Geophysics* 59 (9), 1327–1341. <https://doi.org/10.1190/1.1443692>.
- Panwar, N., Revil, A., Sharma, R., Schmutz, M., Duveillard, P.A., Garcia, B., Cerepi, A., Garcia Artigas, R., Vaudelet, P., Malet, E., Jailet, S., 2021. Induced polarization of carbonates. *J. Geophys. Res.* Solid Earth 126 (6). <https://doi.org/10.1029/2021JB022029>.
- Parker, B.L., Chapman, S.W., Guilbeault, M.A., 2006. Plume persistence caused by back diffusion from thin clay layers in a sand aquifer following TCE source-zone hydraulic isolation. *J. Contam. Hydrol.* 102 (1–2), 86–104. <https://doi.org/10.1016/j.jconhyd.2008.07.003>.
- Pollock, D., Cirkpa, O.A., 2010. Fully coupled hydrogeophysical inversion of synthetic salt tracer experiments. *Water Resour. Res.* 46 <https://doi.org/10.1029/2009WR008575>.
- Power, C., Gerhard, J.I., Tsourlos, P., Giannopoulos, A., 2013. A new coupled model for simulating the mapping of dense non-aqueous phase liquids using electrical resistivity tomography. *Geophysics* 78 (4), N1–N15. <https://doi.org/10.1190/GEO2012-0395.1>.
- Power, C., Gerhard, J.I., Tsourlos, P., Soupios, P., Simyrdanis, K., Karoulis, M., 2015. Improved time-lapse electrical resistivity tomography monitoring of dense non-aqueous phase liquids with surface-to-horizontal borehole arrays. *J. Appl. Geophys.* 112, 1–13. <https://doi.org/10.1016/j.jappgeo.2014.10.022>.
- Revil, A., 2013a. On charge accumulations in heterogeneous porous materials under the influence of an electrical field. *Geophysics* 78 (4), D271–D291. <https://doi.org/10.1190/GEO2012-0503.1>.
- Revil, A., 2013b. Effective conductivity and permittivity of unsaturated porous materials in the frequency range 1 mHz–1GHz. *Water Resour. Res.* 49 (1), 306–327. <https://doi.org/10.1029/2012WR012700>.
- Revil, A., Cathies, L.M., 1999. Permeability of shaly sands. *Water Resour. Res.* 35 (3), 651–662. <https://doi.org/10.1029/98WR02700>.
- Revil, A., Schmutz, M., Batzle, M.L., 2011. Influence of oil wettability upon spectral induced polarization of oil-bearing sands. *Geophysics* 76 (5), A31–A36. <https://doi.org/10.1190/geo2011-0006.1>.
- Revil, A., Coperey, A., Shao, Z., Florsch, N., Fabricius, I.L., Deng, Y., Delsman, J.R., Pauw, P.S., Karaoulis, M., de Louw, P.G.B., van Baaren, E.S., Dabekaussen, W., Menkovic, A., Gunnink, J.L., 2017a. Complex conductivity of soils. *Water Resour. Res.* 53 (8), 7121–7147. <https://doi.org/10.1002/2017WR020655>.
- Revil, A., Sleevi, M.F., Mao, D., 2017b. Induced polarization response of porous media with metallic particles - part 5: influence of the background polarization. *Geophysics* 82 (2), E77–E96. <https://doi.org/10.1190/geo2016-0388.1>.
- 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, 125369 <https://doi.org/10.1016/j.jhydrol.2020.125369>.
- Revil, A., Schmutz, M., Abdulsamad, F., Balde, A., Beck, C., Ghorbani, A., Hubbard, S.S., 2021. Field-scale estimation of soil properties from spectral induced polarization

- tomography. *Geoderma* 403, 115380. <https://doi.org/10.1016/j.geoderma.2021.115380>.
- Ron, E.Z., Rosenberg, E., 2001. Natural roles of biosurfactants. *Environ. Microbiol.* 3 (4), 229–236.
- Rossi, M., Manoli, G., Pasetto, D., Deiana, R., Ferraris, S., Strobbia, C., Putti, M., Cassiani, G., 2015. Coupled inverse modeling of a controlled irrigation experiment using multiple hydro-geophysical data. *Adv. Water Resour.* 82, 150–165. <https://doi.org/10.1016/j.advwatres.2015.03.008>.
- Schmutz, M., Revil, A., Vaudelet, P., Batzle, M., Femenía Viñao, P., Werkema, D.D., 2010. Influence of oil saturation upon spectral induced polarization of oil bearing sands. *Geophys. J. Int.* 183, 211–224. <https://doi.org/10.1111/j.1365-246X.2010.04751.x>.
- Schmutz, M., Blondel, A., Revil, A., 2012. Saturation dependence of the quadrature conductivity of oil-bearing sands. *Geophys. Res. Lett.* 39, L03402. <https://doi.org/10.1029/2011GL050474>.
- Schowalter, T.T., 1979. Mechanics of secondary hydrocarbon migration and entrapment. *AAPG Bull.* 63, 723–760. <https://doi.org/10.1306/2F9182CA-16CE-11D7-8645000102C1865D>.
- Sen, P.N., Straley, C., Kenyon, W.E., Whittingham, M.S., 1990. Surface-to-volume ratio, charge density, nuclear magnetic relaxation, and permeability in clay-bearing sandstones. *Geophysics* 55 (1), 61–69. <https://doi.org/10.1190/1.1442772>.
- Siegrist, R.L., Lowe, K.S., Crimi, M.L., Urynowicz, M.A., 2006. Quantifying PCE and TCE in DNAPL source zones: effects of sampling methods used for intact cores at varied contaminant levels and media temperatures. *Groundwater Monitor. Remed.* 26 (2), 114–124.
- Soueid Ahmed, A., Revil, A., Jardani, A., Chen, R., 2018. 3D geostatistical inversion of induced polarization data and its application to coal seam fires. *Geophysics* 83 (3), E133–E150. <https://doi.org/10.1190/geo2017-0232.1>.
- Soueid Ahmed, A., Revil, A., Abdulsamad, F., Steck, B., Vergnault, C., Guihard, V., 2020. Induced polarization as a tool to non-intrusively characterize embankment hydraulic properties. *Eng. Geol.* 271, 105604 <https://doi.org/10.1016/j.enggeo.2020.105604>.
- Sparrenbom, C.J., Åkesson, S., Johansson, S., Hagerberg, D., Dahlin, T., 2017. Investigation of chlorinated solvent pollution with resistivity and induced polarization. *Sci. Total Environ.* 575, 767–778. <https://doi.org/10.1016/j.scitotenv.2016.09.117>.
- Thomas, L.K., Katz, D.L., Tek, M.R., 1968. Threshold pressure phenomenon in porous media. *SPE J.* 243, 174–184. <https://doi.org/10.2118/1816-PA>.
- Tick, G., Slavic, D.R., Akyol, N.H., Zhang, Y., 2022. Enhanced-solubilization and dissolution of multicomponent DNAPL from homogeneous porous media. *J. Contam. Hydrol.* 247, 103967 <https://doi.org/10.1016/j.jconhyd.2022.103967>.
- Vaudelet, P., Revil, A., Schmutz, M., Franceschi, M., Bégassat, P., 2011. Changes in induced polarization associated with the sorption of sodium, lead, and zinc on silica sands. *J. Colloid Interface Sci.* 360, 739–752.
- Vinegar, H.J., Waxman, M.H., 1984. Induced polarization of shaly sands. *Geophysics* 49 (49), 1267.
- Waxman, M.H., Smits, L.J.M., 1968. Electrical conductivities in oil-bearing shaly sands. *Soc. Pet. Eng. J.* 8 (02), 107–122. <https://doi.org/10.2118/1863-a>.