Comparing Well and Geophysical Data for Temperature Monitoring Within a Bayesian Experimental Design Framework

Robin Thibaut, Nicolas Compaire, Nolwenn Lesparre, Maximilian Ramgraber, Eric Laloy, Thomas Hermans

Comparing Well and Geophysical Data for Temperature Monitoring within a Bayesian Experimental Design Framework

Robin Thibaut[1], Nicolas Compaire[2], Nolwenn Lesparre[3], Maximilian Ramgraber[4], Eric Laloy[5], Thomas Hermans[1]

[1]Ghent University, Belgium
[2]Institut des Sciences de la Terre, Université Grenoble Alpes, France
[3]Institut Terre et Environnement de Strasbourg, Université de Strasbourg, France
[4]Massachusetts Institute of Technology, USA
[5]SCK CEN, Belgium

Key Points:

- We propose a method for optimizing the design of a 4D temperature field monitoring experiment.
- We use Bayesian Evidential Learning (BEL) to infer the posterior distribution of the temperature field for the experiment design.
- We demonstrate how BEL can combine temperature logs and ERT data to assist in the design of experiments for a 4D temperature monitoring.

Corresponding author: Robin Thibaut, `robin.thibaut@ugent.be`

Abstract

Temperature logs are an important tool in the geothermal industry. Temperature measurements from boreholes are used for exploration, system design, and monitoring. The number of observations, however, is not always sufficient to fully determine the temperature field or explore the entire parameter space of interest. Drilling in the best locations is still difficult and expensive. It is therefore critical to optimize the number and location of boreholes. Due to its higher spatial resolution and lower cost, four-dimensional (4D) temperature field monitoring via time-lapse Electrical Resistivity Tomography (ERT) has been investigated as a potential alternative. We use Bayesian Evidential Learning (BEL), a Monte Carlo-based training approach, to optimize the design of a 4D temperature field monitoring experiment. We demonstrate how BEL can take into account various data source combinations (temperature logs combined with geophysical data) in the Bayesian optimal experimental design (BOED). To determine the optimal data source combination, we use the Root Mean Squared Error (RMSE) of the predicted target in the low dimensional latent space where BEL is solving the prediction problem. The parameter estimates are accurate enough to use in BOED. Furthermore, the method is not limited to monitoring temperature fields and can be applied to other similar experimental design problems. The method is computationally efficient and requires little training data. For the considered optimal design problem, a training set of only 200 samples and a test set of 50 samples is sufficient.

Plain Language Summary

The design of experiments is a critical step in scientific research to ensure that the data collected is of sufficient quality to answer the research question. The design of an experiment is often optimized to minimize the cost while still fulfilling the desired accuracy in the prediction. In this contribution, we use an approach called Bayesian Evidential Learning to optimize the design of an experiment. The Bayesian philosophy is used to incorporate all available information into the design of the experiment and to quantify the uncertainty in the prediction. In particular, our approach makes it straightforward to use different data sources (sensors), and different types of data (e.g., temperature, pressure, and concentration measurements). We demonstrate the method in the design of a temperature field monitoring experiment to characterize the geothermal energy storage potential of an aquifer, a key step in the development of geothermal energy projects. This framework makes experiment design easier and faster, a prerequisite for robust risk analysis and decision-making. The method is not limited to temperature fields and can be applied to other similar experimental problems, such as the monitoring of a contaminant plume or saltwater intrusion.

1 Introduction

Geothermal systems, including borehole thermal energy storage (BTES) and shallow aquifer thermal energy storage systems (ATES) are becoming more popular as the world looks for ways to reduce greenhouse gas emissions. Such systems use thermal energy extracted from the ground or groundwater to heat or cool buildings, which necessitates some electrical energy input for the heat pump, while storing the excess heat or cold underground. The goal is to re-use this thermal energy during the next season in a cyclic utilization (Saner et al., 2010; Vanhoudt et al., 2011; Bayer et al., 2013; Duijff et al., 2021). The performance of BTES and ATES strongly depends on the subsurface properties. Many variables are involved in geothermal processes, including porosity, hydraulic conductivity, thermal conductivity, and heat capacity. Subsurface temperature fluctuations are strongly influenced by the spatial distribution of these parameters, the boundary conditions, and the aquifer's hydraulic gradient when modeling the underground response under thermal stress (Ferguson, 2007; Bridger & Allen, 2010; W. Sommer et

al., 2013; W. T. Sommer et al., 2014). Previous research demonstrated that time-lapse Electrical Resistivity Tomography (ERT) could monitor spatial temperature changes in the subsurface with a relatively large spatial coverage by utilizing variations in resistivity caused by temperature changes (Hermans et al., 2012, 2014; Arato et al., 2015; Hermans et al., 2015; Lesparre et al., 2019; Robert et al., 2019). In turn, ERT monitoring experiments can be used to predict the response of the subsurface to thermal exploitation (Hermans et al., 2018, 2019).

This paper presents a methodology for improving the design of field experiments using well and geophysical data by utilizing Bayesian Evidential Learning (BEL). BEL is a technique that combines Monte Carlo simulations and machine learning that can be used to improve estimation of predictions uncertainty (Hermans et al., 2016, 2018; Michel et al., 2020; Thibaut et al., 2021). It performs the Bayesian inference (usually within a low-dimensional latent space) using a direct relationship between predictor (data) and target (prediction) learned from a training set sampled from the prior distribution. Its efficiency has been demonstrated with extensive synthetic validation, but also against rejection sampling (Scheidt et al., 2015), McMC algorithms (Michel et al., 2020, 2022), field data (Hermans et al., 2019), and experimental design (Thibaut et al., 2021). Previous research has shown that BEL can estimate the posterior distribution of parameters in a variety of contexts, including geothermal systems (Hermans et al., 2018, 2019; Athens & Caers, 2019), contaminant transport (Satija & Caers, 2015; Scheidt et al., 2015), geophysical inversion (Hermans et al., 2016; Michel et al., 2020). In addition, the BEL framework has been successfully applied to a range of subsurface field cases, such as groundwater, shallow and deep geothermal and oil/gas predictions (J. Park & Caers, 2020; Pradhan & Mukerji, 2020; Tadjer & Bratvold, 2021).

To forecast the behavioral patterns of such processes, mathematical, physics-based models are frequently used. These models are typically described by partial differential equation systems, but are often too complex to solve analytically and must be solved numerically. These simulations produce a prediction defined on a grid of locations in space and/or time, which is conditioned on the uncertainty of the model parameters. Typically, model outputs are compared to field data via forward modeling in order to calibrate the model unknown parameters, which is known as inverse modeling (Tarantola, 2005, 2006). In prediction-focused approaches, such as BEL, the prediction is directly estimated from the data (Hermans, 2017). In both strategies, data must be collected through an experimental process. Because this process is typically costly and/or time-consuming, it is critical to optimize resource utilization and information gain by controlling the experimental conditions for data acquisition (Attia et al., 2018). This is referred to as optimal experimental design (OED), which is commonly viewed as an optimization problem (Ryan et al., 2016; Kleinegesse & Gutmann, 2020; Haan et al., 2021; Lykkegaard & Dodwell, 2022).

By combining prior knowledge of the model's unknown parameters with likelihood– the data's contribution to those parameters–Bayesian statistics produce the posterior distribution, from which it is possible to draw conclusions about the model's unknown parameters. Applying the Bayesian paradigm to OED at the definition of the objective function leads to a so-called Bayesian OED (BOED) approach. There are several approaches to solving (B)OED problems, all of which involve maximization or minimization of a utility function (Lindley, 1956; Chaloner & Verdinelli, 1995; Ryan et al., 2016). However, the computational burden is significant because BOED requires solving the inverse problem for each possible dataset, as the posterior distribution must first be estimated to compute the Bayesian utility function. Surrogate modeling and Bayesian Model Averaging (BMA) in combination with preposterior estimation are the two main simplifications that have been suggested in the field of Earth Sciences to make posterior computation tractable. Thibaut et al. (2021) provide an overview of these two simplifications. Iterative approaches requiring multiple forward problem runs, such as Markov chain Monte Carlo (McMC)

methods, are commonly used to solve the problem (Laloy & Vrugt, 2012; Vrugt, 2016; Ryan et al., 2016). Zhang et al. (2015), for example, use a McMC algorithm to solve the BOED of sampling well locations and source parameters identification of groundwater contaminants, but defined a surrogate for the contaminant transport equation to reduce the computational burden of McMC. Tarakanov and Elsheikh (2020) developped a BOED methodology aimed at subsurface flow problems relying on a polynomial chaos expansion surrogate model for the utility function embedded in a McMC algorithm.

The main benefit of BEL is that the inferred predictor-target relationship can be applied to any data set that is consistent the prior distribution. This is advantageous for BOED because the prediction problem can be solved quickly and without expensive computation for any newly proposed data set. Thibaut et al. (2021) demonstrated, for instance, how BEL can be used to predict the wellhead protection area (WHPA) surrounding pumping wells. The WHPA (target) was predicted using a small number of tracing experiments (predictors). To determine the optimal data source design, the predictive power of the posterior distribution was evaluated in the original high-dimensional space.

In this contribution, we propose a methodology for solving a BOED problem involving two different data types, using the BEL framework in a low-dimensional latent space, which alleviates the "curse of dimensionality" (Bellman, 1961) and reduces computational and memory demands. This study is based on the monitoring of a heat injection-storage-pumping experiment monitored by time-lapse Electrical Resistivity Tomography (ERT) and well data mimicking the field experiment of Lesparre et al. (2019). In addition to having the potential to improve subsurface information, the fusion of various data types can also be used to plan surveys and choose the optimal set of instruments (JafarGandomi & Binley, 2013). There is a tradeoff between non-invasive data acquisition methods such as ERT and invasive data acquisition methods such as drilling, with the latter being generally more expensive. It is critical to determine whether drilling is required or if a simple geophysical survey will suffice, depending on the specific problem at hand. It can also be useful to know whether it is worthwhile to conduct a long geophysical survey and mobilize a large amount of equipment if we can use an existing borehole in the area.

This contribution is unique in five ways when compared to previous applications of BEL to experimental design:

1. We use two different type of predictors, geophysical and borehole data, which vary in time, to predict a four-dimensional target, the temperature field magnitude in the aquifer over time. It will first be demonstrated how to predict the target using each data set separately.
2. Then, it will be demonstrated how to use them together. Because the predictors are of different types, the dimensionality reduction step in the pre-processing section is applied to each instance separately, and then concatenated before being fed to the learning algorithm.
3. The Transport Map method is used to sample in the low dimensional space.
4. The data utility function is calculated in the low dimensional latent space, without the need to back-transform the data to its original space.
5. We identify the optimal choice between two ERT protocols with our proposed methodology.

Overall, BEL is well-suited to solve the considered BOED problem at a low cost (considering both time and computational demand).

2 Methodology

2.1 BEL

The BEL framework has been extensively explained in contributions such as Hermans et al. (2018) or Scheidt et al. (2018), and we will only summarize the fundamental principles in this contribution. The objective of BEL is to determine the target H's posterior probability distribution $p(H|D_{obs})$, conditioned on an observed predictor D_{obs}. Both the target and the predictor are real, multidimensional random variables. First, training sets of H and D are generated by sampling the prior distribution through forward modeling. Then, the target and the predictor are separately subjected to dimensionality reduction through Principal Component Analysis (PCA) and projected to the new principal component (PC) space. Next, Canonical Correlation Analysis (CCA) analyses the PCs to find their underlying correlations while transforming them to new, maximally correlated canonical variates (CVs). Given that the CVs' relationships describe the behavior of each target dimension for each predictor dimension, an observed data point can be used to infer the posterior distribution of each unknown target dimension. To do this, we first project the observation onto the data CV axes using the same transformations we derived for the data samples. Then, for each target dimension, we use the corresponding bivariate distribution and the observed CV to derive the posterior distribution of that dimension. Thibaut et al. (2021), for example, demonstrated how this can be easily accomplished using multivariate Gaussian inference, provided that the CVs' bivariate distributions are both Gaussian and linear. Kernel Density Estimation (KDE) is another method for approximating the bivariate distribution for each CCA dimension without requiring such assumptions to be verified (e.g., Hermans et al. (2019); Michel et al. (2020)). KDE, on the other hand, has two parameters that must be adjusted: the kernel type, which defines the shape of the distribution at each coordinate, and the kernel bandwidth, which describes the size of the kernel at each position. Transport maps (Spantini et al., 2018) form another method that allows to calculate the posterior distribution and is used in this work (§2.3.3).

2.2 Experimental setup

We present a method for estimating the posterior distribution of an unknown four-dimensional temperature field during a heat injection-storage-pumping experiment. The experiment has been described in Lesparre et al. (2019). In short, hot water was injected at 3 m^3/h into an aquifer at a temperature of 42 °C for six hours, followed by another injection at 14.5 °C for 20 minutes. Then, it was stored in the aquifer for 92 hours, and then pumped back out for 16 hours and 15 minutes at a flow rate of 3 m^3/h. The goal was to track the evolution of temperature distribution in the aquifer over time. The experiment was monitored using time-lapse ERT and temperature data in wells. The time-varying predictor (measured voltage for ERT, direct temperature for wells) and the target (temperature distribution in the aquifer) are high-dimensional, and their relationship is non-linear. By performing geothermal field experiments such as injection and hot water pumping tests, we can gain a better understanding of the aquifer's behavior (Palmer et al., 1992; Macfarlane et al., 2002; Vandenbohede et al., 2009, 2011; V. Wagner et al., 2014; Wildemeersch et al., 2014; B.-H. Park et al., 2015; Klepikova et al., 2016). Combined with these tests, geophysical and thermal monitoring can track heat transfer in the aquifer. The location of injection wells is the focal point of the hydrogeological models used in this study. These models' grids are 60 meters in length in the known direction of natural aquifer flow and 40 meters in length in the perpendicular direction to that direction (Figure 1A). The grid layers begin at the surface of the aquifer's saturated zone at a depth of 3 meters and conclude at the surface of the impermeable basement at 10 meters. The space step along the Z axis (depth) is 0.5 meters. The space step along the X and Y axes ranges from 2.5 centimeters at the injection point to 2.5 meters at the model's

edges, with a 0.25 centimeter refinement in a 3 meter radius around the injection points in the hydraulic flow direction (Y axis).

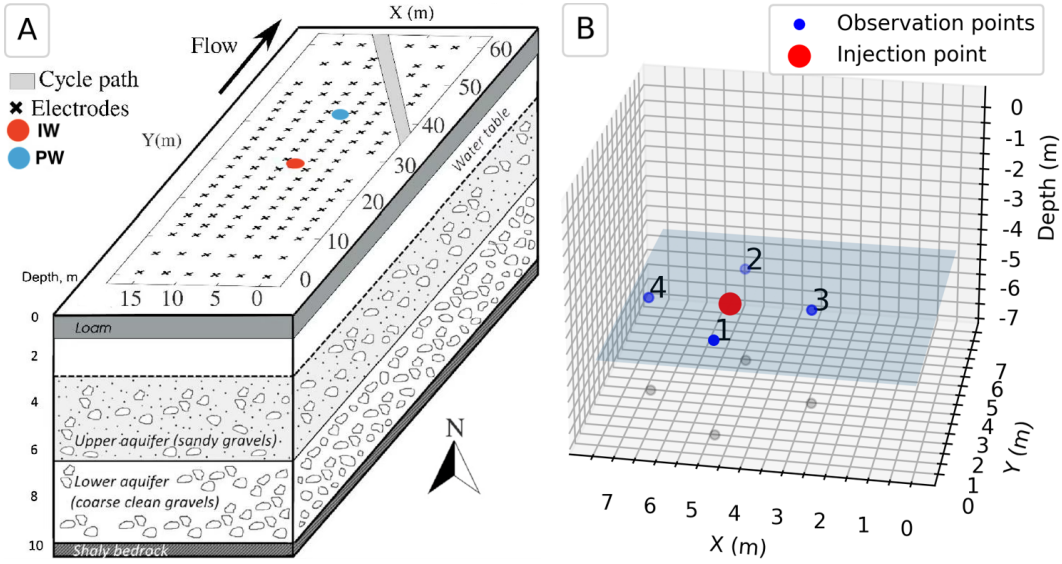

Figure 1. A. Model design (modified from Lesparre et al. (2019)). IW: Injecting well. PW: Pumping well. B. Positions of observation and injection wells.

The mean and variance of hydraulic conductivity (K), the anisotropy factor, and its direction in the horizontal plane and the range of the spherical variogram are used to build the models of the hydraulic conductivity distribution in the aquifer. In addition, the homogeneous porosity and the natural gradient are set as uncertain parameters. All other parameters including water and matrix thermal properties are set as constant. Bulk thermal properties are calculated based on porosity using the arithmetic average. The prior is the sum of all the definitions of the parameters under consideration. We detail the ranges of the variables in Table 1.

Uncertain parameter	**Range**
Log K mean	U[-4, -1], K in m/s
Log K variance	U[0.05,2], K in m/s
Effective porosity	U[0.05, 0.3]
Variogram main range	U[1, 10] m
Anisotropy ratio	U[0,0.5]
Orientation	U[0, π] rad
Natural gradient	U[0.05, 0.167] %

Table 1. Parameters of the prior model. $U[a, b]$ refers to the continuous uniform distribution bounded by the values a and b.

Each parameter is generated randomly and independently according to a uniform distribution. The sequential Gaussian simulation algorithm (Goovaerts, 1997) is then used to generate the hydraulic conductivity fields. In the direction of Y, the hydraulic flow at the grid boundaries is zero, and it respects the natural gradient in the X-direction. The direction of the flow due to the natural gradient is indicated on Figure 1A. For this investigation, 250 hydrogeological models were built and used to generate the temperature fields resulting from the injection-storage-pumping experiment. The HydroGeoSphere code was used to simulate the temperature field in the hydraulic conductivity grids during the injection-storage-pumping experiment (Brunner & Simmons, 2012). There are 106-time observations in each simulation. The temperature grids obtained are reduced to sub-grids around the injection well of size 16×16×14 of elementary volumes 0.5 m×0.5 m×0.5 m to reduce the amount of data (Figure 1B). The temperature is averaged when the subgrid elemental volumes contain several elemental volumes from the initial grid.

The following equations were used to transform the temperature T into conductivity (Hermans et al., 2014):

$$\frac{\sigma_{f,T}}{\sigma_{f,25}} = m_{f,25}(T-25)+1, \text{ and} \tag{1}$$

$$\Delta T = \frac{1}{m_{f,25}}\left[\frac{\sigma_{b2,T}}{\sigma_{b1}}\frac{\sigma_{f1}}{\sigma_{f,25}} - 1\right] + 25 - T_{init}, \tag{2}$$

where $m_{f,25}$ is the fractional change of the fluid conductivity per degree Celsius around the reference temperature of 25°C, σ_{b1} is the background conductivity, $\sigma_{b2,T}$ is the background conductivity at temperature T which we aim to determine at each time-step, σ_{f1} is the fluid conductivity at the initial state, $\sigma_{f,25}$ is the fluid conductivity at 25°C, and T_{init} is the initial temperature.

The synthetic study replicates a real ERT monitoring campaign performed on a well documented site (Dassargues, 1997; Derouane & Dassargues, 1998; Brouyère, 2001; Hermans et al., 2015; Klepikova et al., 2016). Therefore, the initial temperature was measured directly in the field, and the background resistivity σ_{b1} was determined using their inversion using real field data. The temperature and fluid conductivity trend from on-site water samples was calculated to be $m_{f,25}$=0.0194 and $\sigma_{f,25}$ was estimated to be 0.0791S/m (Hermans et al., 2015; Lesparre et al., 2019). σ_{f1} = 0.0614S/m was estimated from Equation 2 with an initial temperature of 13.44°C.

The forward modeling process went as follows:

1. Temperature field simulations on the discretized grid (HydroGeoSphere, Brunner and Simmons (2012)).
2. Computation of the conductivity with the petrophysical law for each simulation and time step (Equation 2).
3. Computation of the ERT profiles for each simulation and time step (EIDORS, Polydorides and Lionheart (2002)).

The ERT simulations are computed with the following experimental setup: 6 parallel profiles of 21 electrodes in the X direction, with a 2.5 meter spacing between the 17 central electrodes and a 5-meter spacing between the four electrodes at the profile's edge.Two protocols were used: multiple gradient (MG) and dipole-dipole (DD). The DD array was chosen due to its widespread use in ERT investigations, which is due in part to the low electromagnetic coupling between the circuits. The MG array was another appealing option due to its sensitivity distribution and high S/N ratio, as well as its robustness and multi-channel compatibility (Dahlin & Zhou, 2006). For each of the 106 temperature simulation time observations, an ERT simulation is generated for each model. At each time step, 1100 quadrupoles are used for the DD protocol and 848 quadrupoles are used for

the MG protocol. The target for each model is thus made of 16×16×14 temperature grids for the 106 observation times, and the predictor is made of the resistances measured by 1100 or 848 quadrupoles for the 106 observation times.

2.3 Heat prediction

2.3.1 Pre-processing

2.3.1.1 Target The target H is a three-dimensional temperature field subdivided into $\mathrm{n}_{rows}\times\mathrm{n}_{cols}\times\mathrm{n}_{lay}$ over n_{step} observation time steps. It is critical to reduce dimensionality because some small-scale variations of the target do not need to be perfectly reconstructed as they are already beyond the predictor's resolution. Before performing dimension reduction, the raw target is scaled to unit variance, because the dimension reduction step is sensitive to the scale of the data. The dimension reduction itself is done by linear PCA. The principal components are new variables produced by combining the initial variables in a linear way.

2.3.1.2 Predictor The geophysical predictor D_g is made up of resistance values measured by n_{quad} quadrupoles over n_{step} observation time steps, and the borehole predictor D_b is made of temperature curves measured at observation well's locations. Before performing dimension reduction, both raw predictors are scaled to unit variance. The dimension reduction is performed by linear PCA, and the number of components to keep is automatically set by setting the amount of variance that needs to be explained to 99%. When working with geophysical data, filtering higher dimensions allows to reduce the effect of noise on the prediction (Hermans et al. (2016); Michel et al. (2020)).

After dimensionality reduction, the PCs of both predictor and target are scaled to unit variance, because covariance matrices are sensitive to the scale of the data.

2.3.2 Training

After dimensionality reduction, the next step is to find the relationship between the predictor and the target in the reduced space. The CCA algorithm projects the data onto a new set of axes which maximize the correlation between the two data sets. The transformed variables are called Canonical Variates (CVs). Similarly to PCA, each pair of transformed variables is orthogonal (uncorrelated) with the other pairs. The CVs are the transformed variables representing the mutual information between the two data sets. Let δ be the number of PCs necessary to explain the required amount of variance in the predictor, and $\mathrm{n}_{training}$ be the number of pairs of predictors and targets used for training. In our case, δ is the sum of the number of components of the geophysical data and the borehole temperature curves, i.e., $\delta = \delta_g + \delta_b$. This is simply done by concatenating the geophysical and borehole temperature PCs into a single matrix. Following preprocessing, the model is trained with the CCA algorithm to establish a multivariate relationship between D and H in PC space. The number of CCA components is set to δ, the maximum number that can be used (Meloun & Militký, 2012). The number of components to keep depends on the dimensionality reduction of the predictor. To allow back-transformation, more components must be used for the predictor than for the target before learning the relationship between the two. However, to avoid overfitting and noise propagation, it is recommended that both have a similar number of components. It does not have to be strictly the same. Let the superscript c denote the canonical space. The canonical variates (CVs) pairs are stored in the $(n_{training} \times \delta)$ matrices

$$D_\delta^c = d_{i,1}^c, d_{i,2}^c, \ldots, d_{i,\delta}^c | i = 1, \ldots, n_{training} \tag{3}$$

$$H_\delta^c = h_{i,1}^c, h_{i,2}^c, \ldots, h_{i,\delta}^c | i = 1, \ldots, n_{training} \tag{4}$$

With the pairs of canonical variates $\left(d_{:,1}^c, h_{:,1}^c\right) \sim \pi_1$ to $\left(d_{:,\delta}^c, h_{:,\delta}^c\right) \sim \pi_\delta$ established, we may infer the posterior in the canonical subspaces by independently conditioning each

of the resulting bivariate joint distributions $\pi_j, j = 1, \ldots, \delta$ on a new observation projected into the canonical predictor space $d^c_{new,j}$. In this study, we consider a new approach based on triangular transport, which offers a good tradeoff between computational efficiency and accuracy.

2.3.3 Conditioning with transport methods

In essence, transport methods seek a monotone, invertible transport map $\mathbf{S}$ that transforms samples from a target distribution π_j into samples from a simpler, user-specified reference distribution η, typically a standard multivariate normal distribution $N(0, I)$, where I is the identity matrix. This map allows us to sample conditionals of the target distribution, and thus implements the conditioning operation we are principally interested in. Transport methods are a nuanced topic, and the interested reader is referred to (Villani, 2009; El Moselhy & Marzouk, 2012; Spantini et al., 2018) for a more detailed discussion of their theoretical properties. In this study, we focus only on the parts necessary for the conditioning operation. Following Spantini et al. (2022) and Baptista et al. (2022), the map $\mathbf{S}$ is triangular, meaning it has as many parameterized map components as there are dimensions in π_j (two in this case, so $\mathbf{S} : \mathbb{R}^2 \to \mathbb{R}^2$). Each map component depends on one more dimension of the target's probability density function (pdf) than its predecessor. In our setting, the map is structured as:

$$\mathbf{S}\left(d^c_{:,j}, h^c_{:,j}\right) = \begin{bmatrix} S_1\left(d^c_{:,j}\right) \\ S_2\left(d^c_{:,j}, h^c_{:,j}\right) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{z} \tag{5}$$

where $\mathbf{z}$ are samples distributed according to the reference pdf η. These map component functions are made up of user-specified combinations of polynomials or radial basis functions, and must be monotone in their last argument. This means that $\partial_{d^c_{:,j}} S_1\left(d^c_{:,j}\right) > 0$ and $\partial_{h^c_{:,j}} S_2\left(d^c_{:,j}, h^c_{:,j}\right) > 0$. This will ensure that the entire map $\mathbf{S}$ remains monotone and thus invertible. This monotonicity requirement can be ensured with diligent parameterization of the map components (e.g., Equation 2 in Baptista et al. (2022)).

If we are only interested in conditioning, we only need to define, optimize, and evaluate the second map component $S_2 : \mathbb{R}^2 \to \mathbb{R}$ (Spantini et al., 2022). We optimize this map component S_2 by minimizing the Kullback-Leibler divergence between the target pdf π_j, presumed to be known only through samples, and the map's approximation to the target pdf, which is obtained by sending the standard Gaussian reference η through the inverse map $\mathbf{S}^{-1}$. From this, we can derive the following objective function $\mathcal{J}$ (Spantini et al., 2018):

$$\mathcal{J}(S_2) = \sum_{i=1}^{n_{training}} \left(\frac{1}{2}\left(S_2(d^c_{i,j}, h^c_{i,j})\right)^2 - \log \frac{\partial S_2(d^c_{i,1}, h^c_{i,j})}{\partial h^c_{i,j}} \right). \tag{6}$$

The forward evaluation of S_2 is a simple evaluation of its constituent basis functions. We can also invert S_2 efficiently through an appropriate one-dimensional root finding algorithm. Note that while the forward map components S_1 and S_2 can be evaluated independently, their inverses must be evaluated in sequence:

$$\mathbf{S}^{-1}\left(\mathbf{z}\right) = \begin{bmatrix} S_1^{-1}\left(z_1\right) \\ S_2^{-1}\left(d^c_{:,j}, z_2\right) \end{bmatrix} = \begin{bmatrix} d^c_{:,j} \\ h^c_{:,j} \end{bmatrix} \tag{7}$$

where the inverse of the second map component S_2^{-1} depends on the outcome $d^c_{:,j}$ of the first inversion S_1^{-1}. This dependence makes triangular transport highly useful for Bayesian inference. In fact, the inverse map $\mathbf{S}^{-1}$ factorizes the target pdf π_j according to the ordering of the variables in Equation 5 as $\pi_j\left(d^c_{:,j}, h^c_{:,j}\right) = \pi_j\left(d^c_{:,j}\right)\pi_j\left(h^c_{:,j}|d^c_{:,j}\right)$, where the second term on the right-hand side $\left(\pi_j\left(h^c_{:,j}|d^c_{:,j}\right)\right)$ corresponds to the posterior and is sampled by the map component inverse S_2^{-1} (e.g., Section 7.1 of Spantini et al. (2018)).

In other words, this means that we can sample conditionals of π_j by simply replacing the argument $d^c_{:,j}$ during the map component inversion S_2^{-1} with any values on which we want to condition. For instance, supplying duplicates of the observation $d^c_{obs,j}\mathbf{1}^\top$, where $\mathbf{1}$ is a column vector of 1s, we can sample the desired conditional $\pi_j\left(h^c_{i,j}|d^c_{obs,j}\right)$:

$$\mathbf{t}^{c,\text{cond.}}_{:,j} = S_2^{-1}\left(z_2; d^c_{obs,j}\mathbf{1}^\top\right) \sim \pi_j\left(h^c_{i,j}|d^c_{obs,j}\right). \tag{8}$$

Note that the required reference samples z_2 can be either drawn from a standard Gaussian distribution, or (better) obtained from the forward map $z_2 = S_2\left(d^c_{:,j}, h^c_{:,j}\right)$ (see Spantini et al. (2022)). With the conditioned samples $\mathbf{t}^{c,\text{cond.}}_{:,j}$ for each pair of covariates $\left(d^c_{:,j}, h^c_{:,j}\right)$, we can then back-transform the posterior samples into the original target space by ascending from CCA, PCA, and undoing any transformations.

2.4 Experimental design

BEL can be used to assess the amount of information delivered by various data sources. The actual data can have any value within the prior data space, and data sources can be placed anywhere across the grid. To identify highly informative data sets based on their location, one data-utility function must be maximized or minimized (Thibaut et al., 2021). To restrict computation time and keep some realism, we will only consider four well positions around the injection well, which is sufficient to demonstrate the approach. With the geophysical data, a total of 31 different combinations of data sets are possible. These unique combinations are combined with a total of 50 unknown ground truths (test set). Each sample of the test set is sampled with 500 posterior samples. This number is arbitrarily chosen to be high enough to ensure that the posterior distribution is reasonably well-sampled while remaining low enough to keep the computation time manageable. We end up with an array of shape (31, 50, $n_{samples}$, δ) = (number of combinations, test set size, sample set size, number of canonical components).

One of the challenges of our case study is the high dimensionality of the problem. This is why we use the principal components to perform BOED, which is a novel approach in this context. The true target is transformed to the PC space and its RMSE with the predicted targets' PCs is computed. The advantage of working in a lower dimension is that we need to predict a smaller number of dimensions, which is computationally faster. Because the back-transformation from the PCs to the observation targets is a linear operation, it is simple to demonstrate that the prediction error is minimized if the distance between the PCs of the predicted and observed targets is also minimized. The minimization of the target prediction error is therefore equivalent to the minimization of the distance between the PCs of the predicted and the observed targets. The PCs are weighted by their explained variance ratio during distance computation because they account for the different importance of the different components.

We validate our approach with k-fold cross validation. We repeat our computations across 5-folds and average the metric results. Averaging across folds is not strictly necessary but will be used here to increase the robustness of the BOED.

The methodology is summarized below:

Input: training set (D_{train}, H_{train}), test set (D_{test}, H_{test}), data utility function UF
Output: Averaged cross-validation rankings
for all Fold f in 5-fold cross-validation **do**
  $H_{train,f} \leftarrow$ Target training data for fold f
  $H_{train,f} \leftarrow PCA_h.fit\_transform(H_{train,f})$
  $H_{test,f} \leftarrow$ Target test data for fold f
  $H_{test,f} \leftarrow PCA_h.transform(H_{test,f})$
  for all Possible Combinations **do**

$O \leftarrow Combination$
$D_{train,f,O} \leftarrow$ Predictor training data for fold f and combination O
$D_{train,f,O} \leftarrow PCA_d.fit\_transform(D_{train,f,O})$
$D_{test,f,O} \leftarrow$ Predictor test data for fold f and combination O
$D_{test,f,O} \leftarrow PCA_d.transform(D_{test,f,O})$
$TrainedModel \leftarrow CCA.fit(D_{train,f,O}, H_{train,f})$ ▷ Training step
for all Ground Truths in $(D_{test,f,O}, H_{test,f})$ **do**
$D_{true} \leftarrow$ True predictor
$D_{true} \leftarrow PCA_d.transform(D_{true})$
$H_{true} \leftarrow$ True target
$H_{true} \leftarrow PCA_h.transform(H_{true})$
$H_{posterior} \leftarrow TrainedModel.predict(D_{true})$ ▷ Predicting in PC space
$Utility \leftarrow UF(H_{posterior}, H_{true})$
$Results.add(Utility)$
end for
$Ranking \leftarrow add(Results)$
$Results.clear()$
end for
end for
$Final \leftarrow mean(Ranking)$

The method was developed in Python on a standard laptop with 16 GB of RAM and a 2.30 GHz 8-core Intel Core i9 processor. Profiling has been performed on the code for temperature prediction with the most complex combination (ERT combined with the four wells). It took 38 seconds to initialize the model, load the dataset, pre-process the data, and save the files. The prediction part of the experiment took 27 seconds, bringing the total time to 65 seconds, including the time required to save the results to a file. Throughout, the process used no more than 2.08 GB of RAM. Given that the entire implementation was created for research purposes, the implementation can be enhanced for speed and memory management (e.g., parallelization, low-level programming language implementation, etc.). Finally, the computational costs differ depending on the case study and the size and number of data sets used.

3 Application

3.1 Optimal protocol determination

Before determining the optimal sensor combination, we can determine the optimal protocol between MG and DD following our methodology. Common approaches to optimizing electrode arrays for ERT either seek the best measurement configuration on a given set of electrodes (e.g., Wilkinson et al. (2015)) or select electrode locations (e.g., F. M. Wagner et al. (2015)). Both strategies are based on the resolution matrix. Uhlemann et al. (2018) propose a new approach that combines these two strategies by introducing an additional weight that penalizes the addition of electrode locations to the optimized set. Qiang et al. (2022) take a different approach. Instead of focusing on image resolution, they propose that the ERT survey be optimized for a specific target of interest by maximizing the information gained from a target area using Bayesian experimental design.

In contrast to these methods, we seek to determine the optimal protocol for a given set of electrodes, and not the electrode locations or quadrupole configurations themselves. It could, however, be extended for that purpose. Furthermore, previous approaches use the resolution matrix, whereas our approach uses the predictive power of the posterior distribution in a low-dimensional latent space. Qiang et al. (2022) also makes use of the posterior distribution's predictive power, but in a physical, high-dimensional space. Lastly, these approaches are limited to a single target of interest, whereas our machine learn-

ing approach can generalize to unseen targets if the prior distribution is representative of the true distribution.

We consider four cases:

1. The MG array (848 quadrupole measures), to which PCA is applied to explain 99% of the variance.
2. The DD array (1100 quadrupole measures), to which PCA is applied to explain 99% of the variance.
3. Combination 1: concatenation of the MG and DD arrays to which PCA is applied to explain 99% of the variance.
4. Combination 2: union of the previously calculated principal components of the MG and DD configurations. Thus, it is equivalent to joining the principal components from items 1 and 2.

Combinations 1 & 2 are not among the available options for continuing the study, but they are included here for completeness. The computation is performed on the 50 test cases over 5 different folds for each of the four cases, and the averaged results are shown in Figure 2. The MG array, despite having fewer quadrupoles than the DD array, is the best single protocol. In fact, of the four cases, the DD array has the lowest predictive power. Therefore, the MG array is selected as the optimal protocol for the study. Interestingly, Combination 1 has a lower predictive power than the MG array, although it contains more quadrupoles. The MG and DD configurations are not equivalent and produce different outcomes. As a result, when using PCA, joining the two configurations in this manner resulted in information loss. Combination 2, however, significantly improves the predictive power over the individual cases. It is explained by the fact that the two configurations are complementary and that the union of the previously calculated principal components of the two configurations captures the most meaningful information for the problem. The posterior distribution is better constrained as a result of the additional information, improving predictive ability.

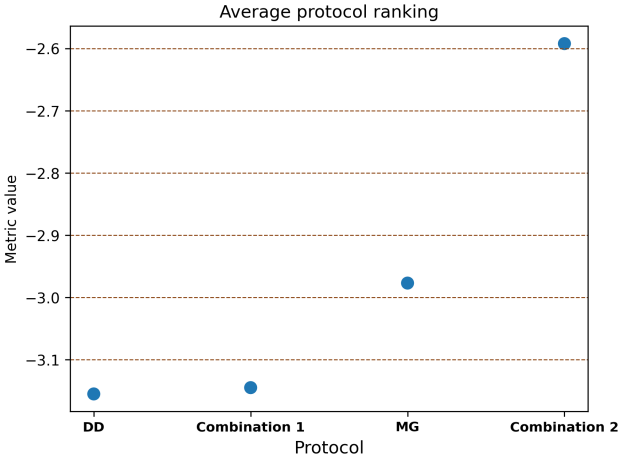

Figure 2. Average score of the different protocols over 5 folds. DD: PCA(Dipole-Dipole), MG: PCA(Multiple-Gradient). Combination 1: PCA(DD+MG), Combination 2: PCA(DD)+PCA(MG).

3.2 Target prediction

This section shows how to predict a single four-dimensional temperature field, using three distinct predictors: (i) the geophysical data, made of resistance measurements from 848 quadrupoles over 106 observation time steps, which are combined to form the predictor, (ii) a single temperature profile over 106 observation times, and (iii) the combination of the geophysical data with all 4 borehole temperature curves. We show that BEL can accurately estimate the target posterior distribution with varying uncertainty levels according to the type of predictor used. The dataset in this section has a total size of $n = 250$. The training set is then reduced to $n_{80\%} = 200$ models, with the remaining models being used to validate BEL's ability to predict the target, and later on for OED. Previous BEL applications have demonstrated that making accurate predictions with a dataset of this size is possible (Hermans et al., 2016, 2018, 2019; Athens & Caers, 2019; Michel et al., 2020; J. Park & Caers, 2020; Yin et al., 2020; Thibaut et al., 2021). While a small training set size is inevitable due to the time-consuming nature of the simulations, it is sufficient because the prediction is a temperature distribution that varies smoothly in both time and space and results from advection, diffusion, and dispersion processes. Such target is much simpler than the underlying K model. BEL is a Bayesian method that incorporates uncertainty. Therefore, a large training set size is not required and the method is more robust against overfitting. Furthermore, the use of cross-validation ensures the robustness of the results.

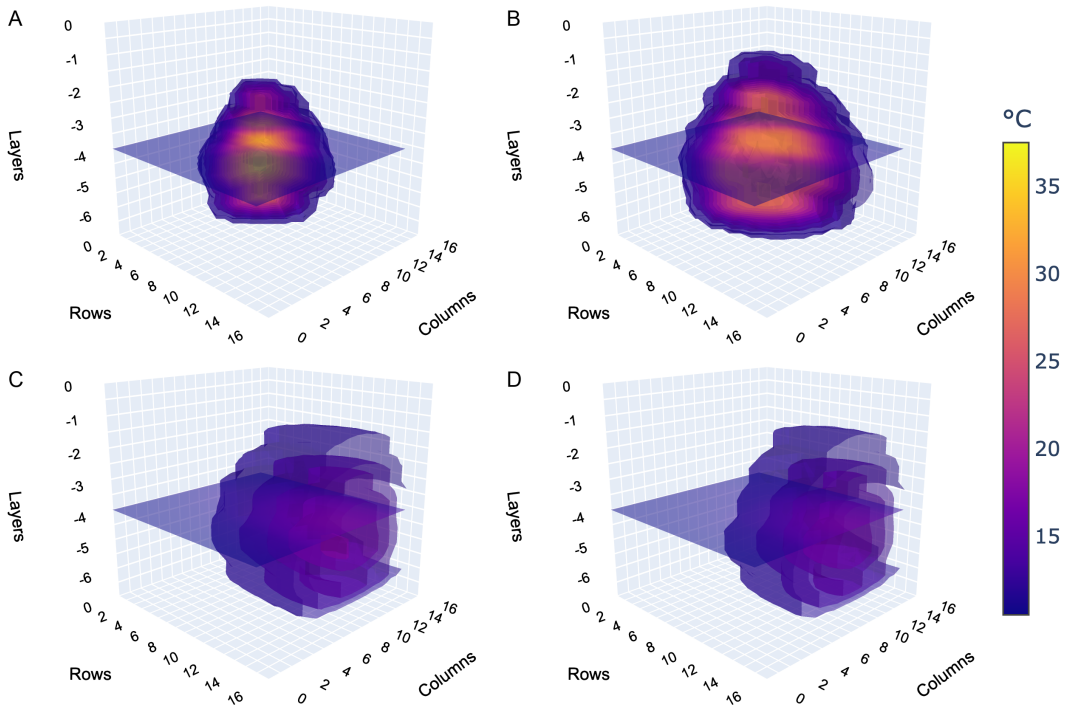

Figure 3. Snapshots of the temperature field at time steps 4 (A), 14 (B), 61 (C) and 74 (D) for one example. The injection well discharge is at (column, row, layer) = (9, 6, -5). The reference plane at the wells level is highlighted.

3.2.1 Pre-processing

The three predictors undergo the pre-processing described in §2.3.1. When geophysical data is combined with borehole temperature curves, the following steps are taken:

1. PCA is performed to explain 99% of the variance in the geophysical data to obtain the PCs $D_{G,99}$
2. PCA is performed to explain 99% of the variance in each of the n_b boreholes temperature curves to obtain the PCs $D^{(i)}_{B,99}$ for each borehole i. These PCs are then concatenated into a single array $D_{B,99}$.
3. The geophysical PCs are concatenated with the borehole temperature curves PCs to obtain the array $D_{G,B,99}$

We have to perform PCA separately for each predictor, because if we simply concatenated the predictor arrays and then applied PCA, the resulting predictor would be more representative of the geophysical data than the borehole temperature curves. Furthermore, the PCs are scaled to unit variance before concatenation because the PCs magnitude can vary greatly from predictor to predictor. The resulting principal components for all cases can directly be compared in Figure 4. The left column shows the predictor PCs, and the right column the target PCs for all cases. For each PC dimension (X-axis), the PC value for each of the 200 training instances is plotted (Y-axis). The example test instance is plotted on top of the training instances. The required number of PCs is chosen to account for 99 percent of the variance in the predictor. In parallel, PCA is applied to the target. In line with our methodology, the number of components to keep is set to the maximum number of PCs required by the predictor, and the corresponding amount of variance explained in the target is summarized in Table 2.

Parameter \Combination	**G**	**1**	1, 2	1, 2, 3	1, 2, 3, 4	**G, 1, 2, 3, 4**
δ	10	3	7	10	13	23
Explained variance (target)	77%	57%	71%	77%	80%	88%

Table 2. Effect of the number of PCs (δ) on the target PCA explained variance. **G** stands for geophysical data. **1, 2, 3, 4** stand for the borehole temperature curves. Case (i) is **G**, case (ii) is **1** and case (iii) is **G, 1, 2, 3, 4**.

When a single temperature profile is used as a predictor (Case (ii)), the target variance is not explained as well as when the entire set of predictor data is used (Case (iii)), hence the smaller number of target PCs required. When the predictor is a combination of geophysical data and all temperature profiles, some additional variance is captured. The remaining target PCs for each case are shown in Figure 4B, D and C. These additional target components are not used in training. They are, however, saved for subsequent use in OED.

3.2.2 Training and prediction

The CCA mapping allows the fusion of the low-dimensional representation of the predictor with the low-dimensional representation of the target in order to make predictions of the target distribution. It is run on all cases to find the canonical variates that define the relation between the predictor and the target, using a maximum of 23 PCs on the predictor. The first three canonical variate pairs for each case are shown in Figure 5. Each row corresponds to a single case, and each column to a single canonical variate pair. The variable ρ is the correlation between the canonical variates, and can be interpreted as the amount of mutual information shared between the target and predictor. For all cases, the predictor explains a large part of the target: ρ_1 (first pair) = 0.999, 0.959 and 1. for case (i), (ii) and (iii) respectively. Hence, CCA is well-suited for our purposes. The canonical variates for case (ii) provide a poorer explanation for the target (ρ_2

(second pair) = 0.462). This is because the temperature curves used as predictors are one-dimensional and only convey information about a small portion of our model over time. The strong correlations in the canonical variates pairs of cases (i) and (iii) is further indication of the suitability of CCA for this study. Transport map inference is run for each pair for each case, using the test predictor (on the X-axis (d^c) of Figure 5). The 500 samples of $p(h^c|d^c_*)$ in Figure 5 represent our predictions of the unknown target distribution in the canonical space. They can then be sequentially back-transformed to the principal component space (see posterior samples in Figure 4) and to the original space (see Figures 6 and 7 for 1D and 2D representations, respectively).

Figure 6 shows the temperature curves of the 500 samples at the location of the observation well number 2 for each case. This point was arbitrarily chosen to illustrate the one-dimensional temperature curve at one observation point. Across cases, the level of uncertainty, expressed by the spread of the temperature curves, is higher during the injection and storage phases than it is during the pumping phase, which is a positive development because this portion of the curve is the part that is typically inferred for ATES systems (energy recovery). The magnitude of the spread of the predicted samples in the canonical space (Figure 5) is sequentially transmitted through the principal component space (Figure 4) and the original space (Figure 6), since the data flows through linear transformations.

On Figure 7A, the true (test) temperature cross-section at layer 9 (at a depth of 4.5m) and time t=105.75h (pumping phase) is shown. Figure 7B shows one randomly drawn example from the 500 sampled temperature profiles when using the ERT predictor alone. The results are not only visually close to the truth, but the absolute temperature difference ranges from -0.8 to 0.4 degrees Celsius, which is the magnitude of accuracy we expect when using resistance data (Hermans et al., 2015). Figures 7C and 7D show the same results when a single borehole temperature profile and the full combination of ERT and temperature profiles are used, respectively. As expected, the single borehole image is the least accurate, while the combination of all temperature profiles and ERT is the most accurate.

Because the underlying internal behavior of the aquifer is nonlinear, given the number of parameters involved, the underlying physics of the model and the nature of the predictor (indirect geophysical data), such a level of uncertainty is expected (Hermans et al., 2015, 2016, 2018). Considering the level of complexity of the underlying physics and the relatively small training set, BEL is capable of inferring different possible outcomes with reasonable uncertainty.

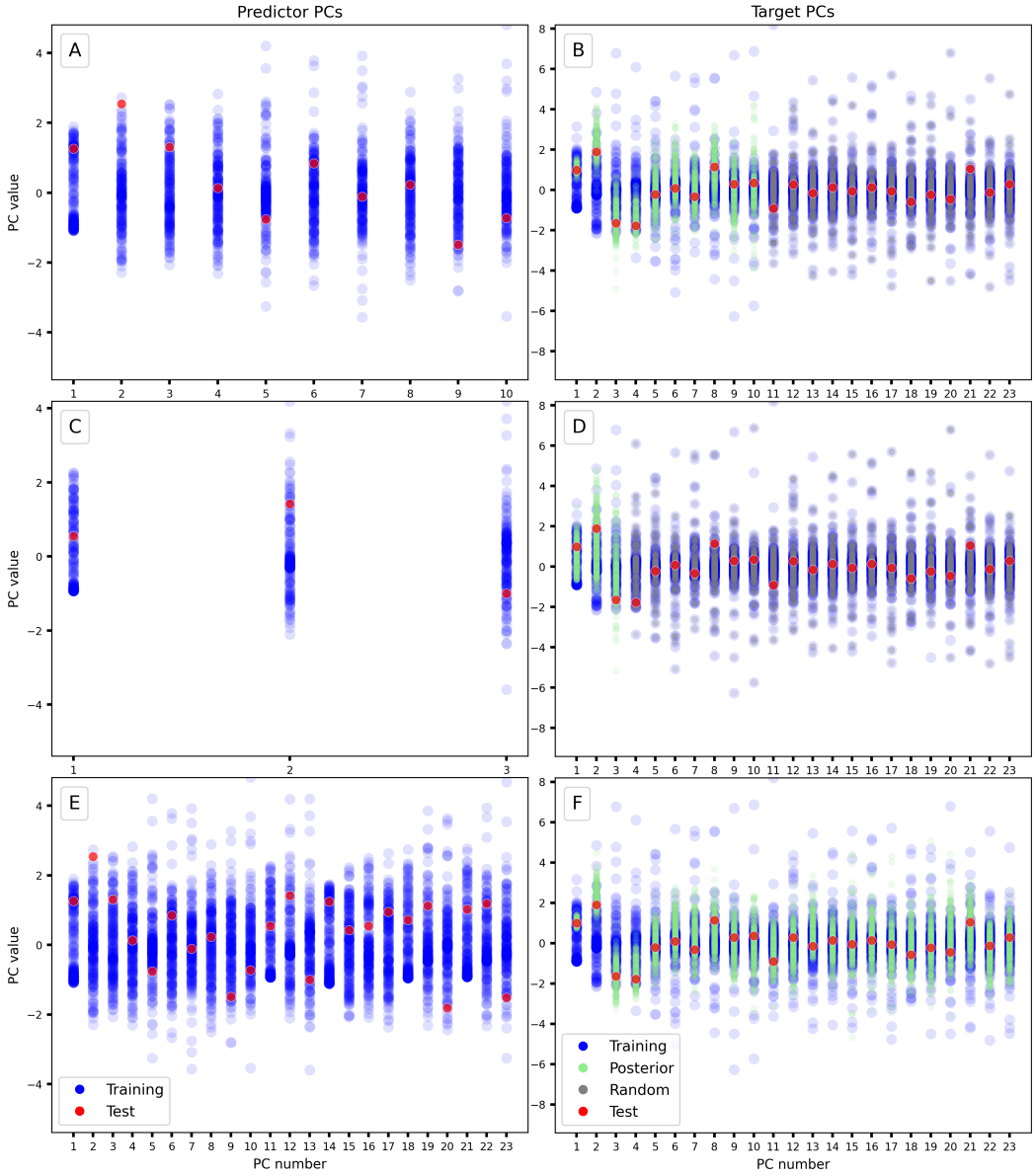

Figure 4. Principal Component Scores. The 'Random' PC samples are drawn at random from the target PC training set. They will be used in the OED section 4. A. Case (i). Predictor: ERT data, B. Case (i). Target. C. Case (ii). Predictor: Temperature profile from borehole 1. D. Case (ii). Target. E. Case (iii). Predictor: Full combination (ERT data + four boreholes temperature profiles). F. Case (iii). Target.

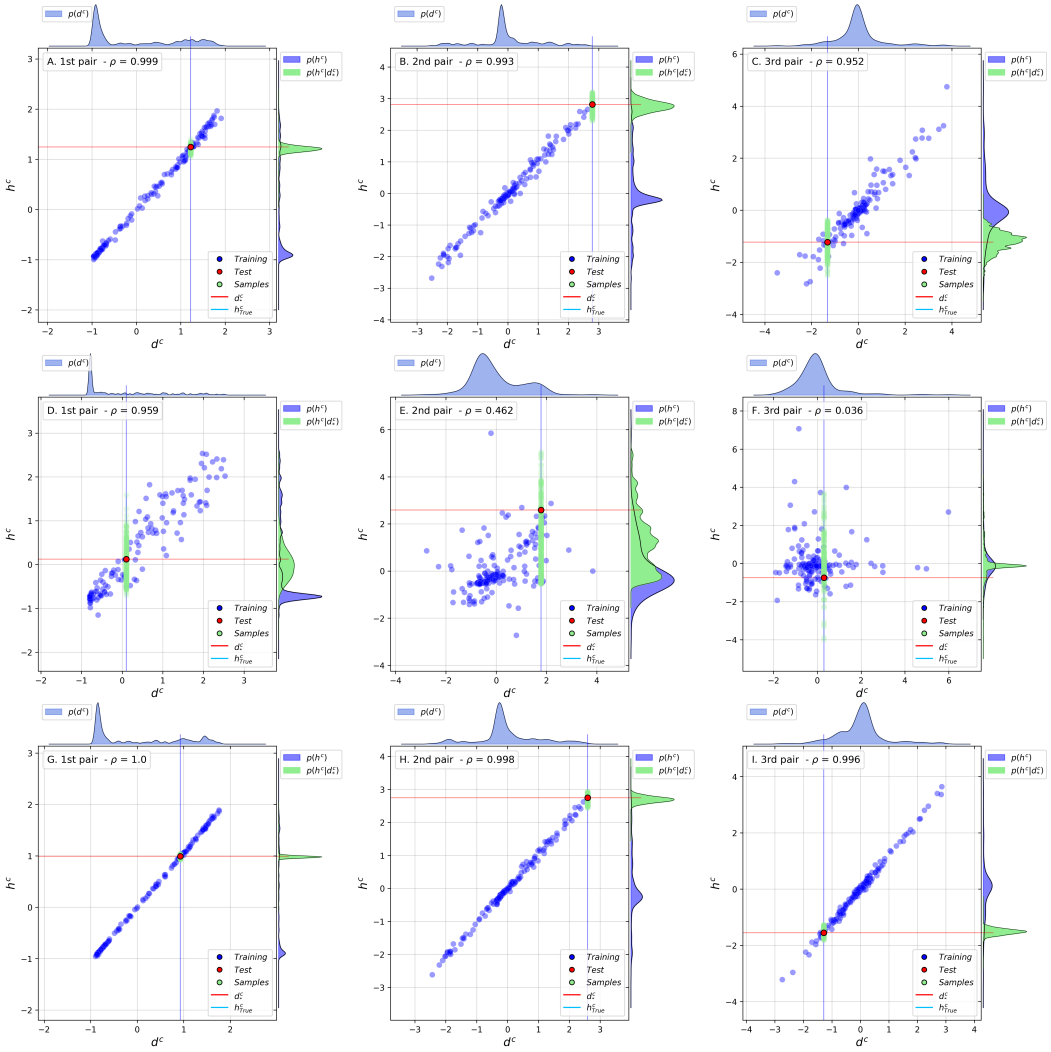

Figure 5. Canonical Variate Pairs (1 to 3). The first row (A, B, C), case (i): uses the geophysical predictor, the second row (D, E, F), case (ii): uses the borehole predictor, and the third row (G, H, I), case (iii): uses both predictors. The true point coordinates (Test) are highlighted by the two lines in each dimension.

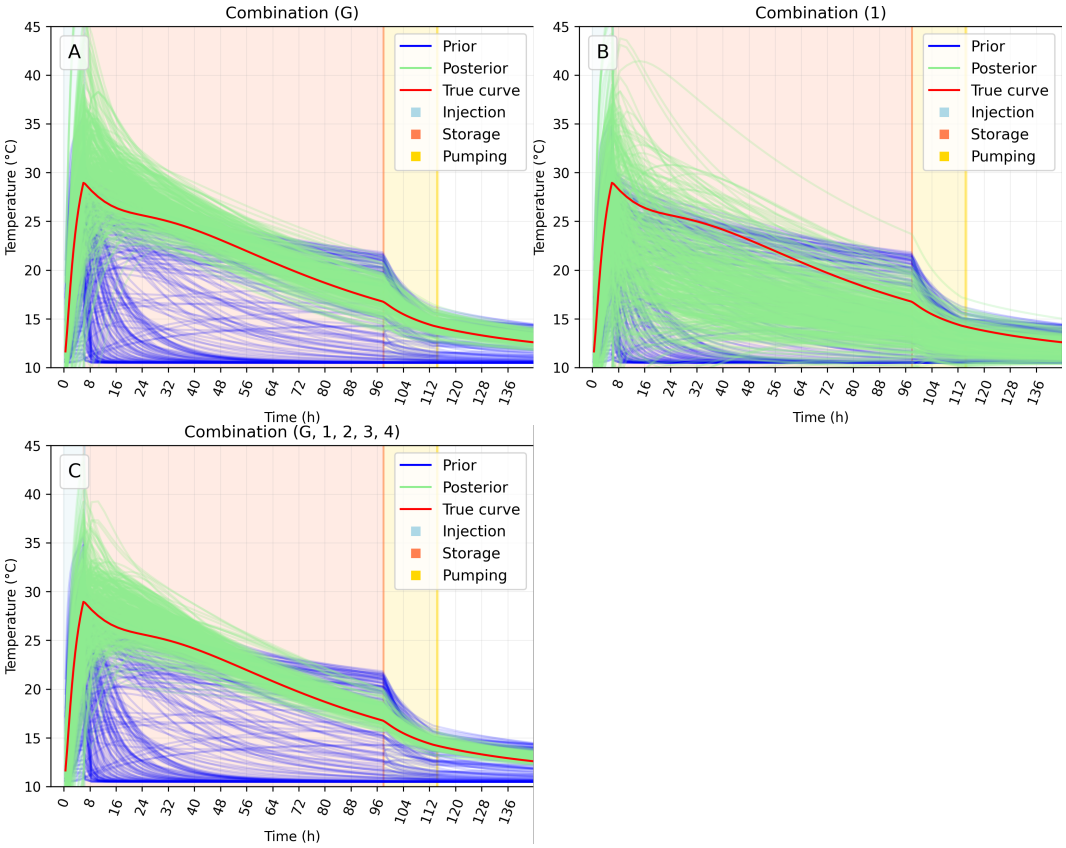

Figure 6. Temperature curves across all time steps, at the observation well 2. A. Using the ERT as predictor. B. Using the temperature curves at well 1 as predictor. C. Using the ERT and all the temperature curves at wells 1, 2, 3 and 4 as predictors.

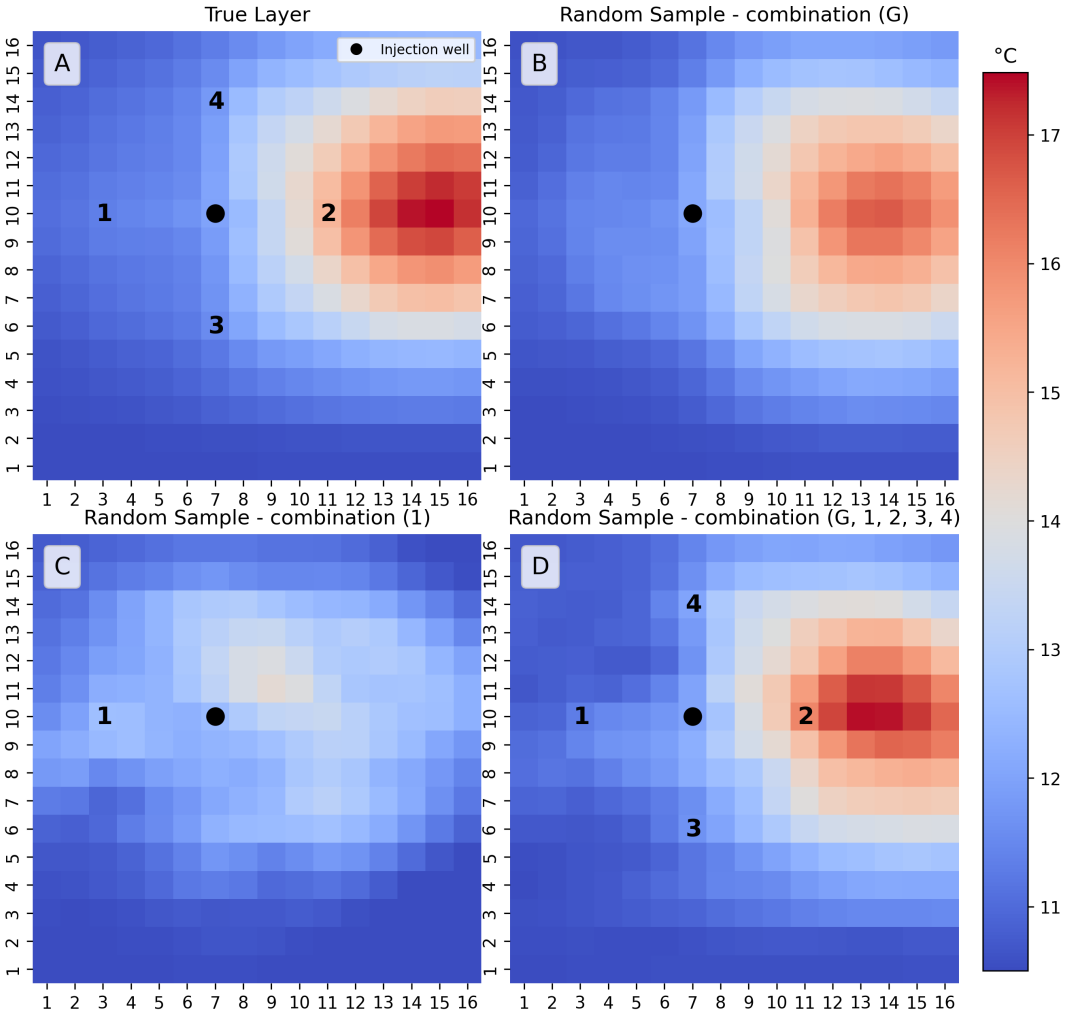

Figure 7. Cross-section of the temperature field at time step 74 (105.75 hours - pumping phase) and layer 9 heat injection level. A. Ground truth. B. Using the ERT as predictor. C. Using the temperature curves at well 1 as predictor. D. Using the ERT and all the temperature curves at wells 1, 2, 3 and 4 as predictors.

4 Results

To compare the different methods fairly, the same number of components must be used in each case. Random components from the training set are chosen for BOED, as shown in Figure 4. The extra random components, which were not used for training, could thus not be predicted, but they must be included to avoid bias in the results. Selecting them at random for each PC dimension ensures that the score of the combination will be representative of the actual prediction performance. The expected outcome is a ranking of combinations ranging from a small number of combinations to all considered data sources, with the most unfavourable case being the use of a single observation well and the most favourable case being the use of geophysical data and all available wells, as can be visually evaluated across cases on Figure 6.

To derive a robust BOED, we average the results over the test set of ground truth models. However, the results depend on the chosen ground truth models set, so a combination that is ranked as the best in one case may not be ranked as such for another

set, which is why we use a 5-fold cross-validation to validate our findings. We use k-fold cross-validation to produce predictions over 5 different training and test sets. We average the rankings across folds to obtain the final ranking. Figure 8A depicts the well locations for the default case. Figure 8B depicts the average results of all 31 combinations over the 5 folds for the default case. The geophysical data is labeled as 'G' and the well data are labeled by their well ID (1, 2, 3, 4). For example, 'G12' is the combination of geophysical data with wells 1 and 2. The results are consistent with what was expected, and using more data sources yields the best results, which is logical given the symmetry of our observation wells network. Our findings show that the ERT data alone provides the most information to the model, with a clear difference in metric value between the wells alone and the ERT data. To highlight that there is no overlap between the two, a darker background colour indicates when the ERT data is used, and a lighter background colour indicates when the wells alone are used. Our results corroborate previous findings that ERT is a valuable tool for monitoring the development of thermally affected zones in aquifers (Lesparre et al., 2019; Hermans et al., 2016).

In terms of observation wells, wells 1 and 4 are consistently ranked as the least favorable cases, and well 2 is ranked as the most favorable case. This is due to the direction of the heat plume over time. As shown on Figures 3, 6 and 7, the pumping causes the plume to move downstream, allowing well 2 to record more variation from the plume over time. The aquifer's natural gradient has an X component (see Figure 1A), which causes the plume to slowly drift towards well 3, which is why well 3 is ranked in the middle. However, because the plume is moving away from them, wells 1 and 4 are more likely to record redundant data from it (e.g., a flat temperature curve), which makes them less relevant for model training.

Similar observations were made in the case of solute transport in Thibaut et al. (2021): downstream wells provided the most useful data. In their settings, their tracer curves provided more information on their underlying hydraulic conductivity field, lowering their prediction's uncertainty. They did not investigate the information gained from any geophysical data, which in our case is sufficient to understand the evolution of the heat plume, but it is important to note that the overall information gain from the combination of all wells and ERT is the highest. Since ERT is an indirect data source, adding at least one borehole allows to provide direct information on the temperature and thus to reduce the uncertainty.

To corroborate our findings, we reproduce the experiment using four different well locations, but with the same data (the previous k-fold shuffle seed is reused). The alternative well positions are shown in Figure 8C and the results are shown in Figure 8D. The outcomes are in line with expectations. The worst-case scenario is well number 3. It is located upstream and is less affected by the heat plume. Because of its proximity to the heat injection, well number 1 is the best case. The combination of the various wells produces more insightful results than before. Wells 1 and 2 cover both upstream and downstream areas, and are far apart. As a group, they collect more data from the heat plume than the other group of two observation wells. Following the same logic, the combination of wells 1, 2, and 4 ('124') is the best case. Because alternative well locations are more dispersed and provide less redundant information, the increase in information provided by geophysical data is not as strong as in the previous case. In the case of alternative well locations, the full well combination '1234' is not the best case (Figure 8D). As described in section 3, feeding to the training algorithm an array of concatenated data from various sources, some of which provide redundant information, can reduce the model's prediction power.

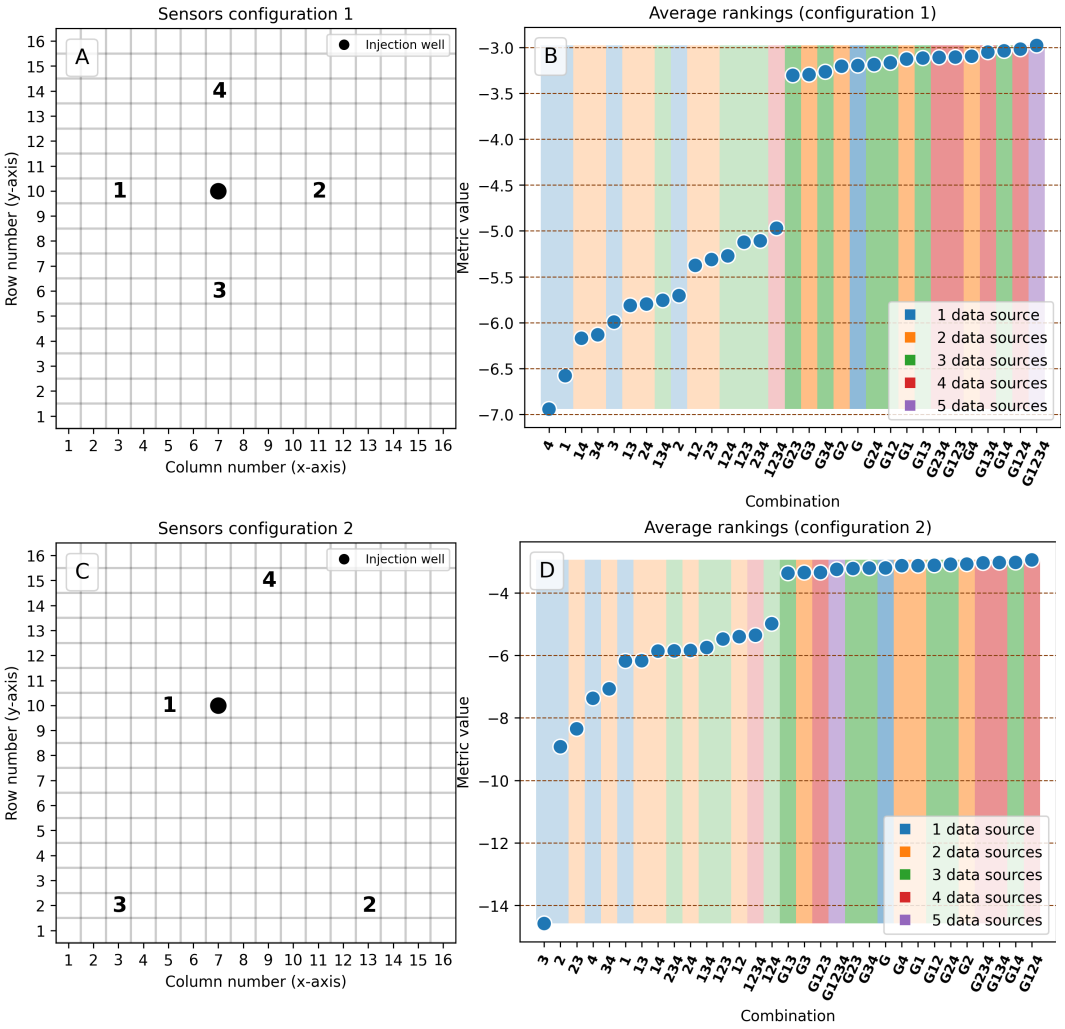

Figure 8. Sensors position for the two cases and ranking of the different combinations of data sources. To visualize a higher score for the best combination, the metric (RMSE) values opposite are displayed. The use of ERT data (**G**) is indicated by a darker background shade, whereas the use of wells alone is indicated by a lighter shade. A. Default well locations. B. Average ranking of 5 folds for the default well locations. C. Alternative well locations. D. Average ranking of 5 folds for the alternative well locations.

5 Discussion

A fundamental question is how we will put our framework into action in order to make the best decisions. Testing for all well positions and combinations is possible in our synthetic study, but time-consuming, because each combination requires a training and testing phase. As a reminder, the number of possible combinations is $\frac{n!}{w!(n-w)!}$, where n is the number of possible positions and w is the number of wells. For a number of 3 wells and 256 possible positions (number of cells for one plane in the grid), the number of possible combinations is $2,763,520$. This is assuming uniform grid cells and no hydrogeological property scaling issues. Indeed, petrophysical laws may not be viable at different scales (see Singha et al. (2015) for a discussion on this topic). However, regardless of which combination is selected, the computations will be based on the same 250 heat flow and transport simulations. This ensures that the most time-consuming com-

ponent will not be performed more than once. As a result, running the experimental design for a representative subset of these combinations is feasible. Constraints such as at least 5 m between wells, at least one well upstream and one downstream, and so on can be used to reduce the number of possible combinations.

One ideal situation in which to apply our framework is when potential well locations are predetermined and limited by factors such as land occupation. The situation where the wells are already positioned, and we are looking for the single next-best well position is an alternative ideal situation. This topic, however, is beyond the scope of this contribution. This would entail sequentially updating our prior knowledge from the existing well(s), as each new measurement would necessitate rerunning the forward model with updated parameters. This would be a challenging task given our computationally expensive forward model. Future research will investigate the ability of the framework to sequential BOED, and solve the problem of allocating new sensors placement adaptively such as, e.g., Haan et al. (2021), Lykkegaard and Dodwell (2022), Wang et al. (2022).

In addition to temperature monitoring and borehole sampling optimization, the method as it is presented here can be used to solve a wide range of data sources fusion and location problems with any utility function. Haan et al. (2021) mention that solving such allocation problems helps answer questions such as: Which additional sensor type would yield the highest gain? What is the most effective data source combination? How many and where should sensors be activated given logistical or energy constraints?

In the field of Earth Sciences, Eidsvik et al. (2015) integrates these concepts into a decision-theoretic framework in which the utility function is defined as Value of Information (VOI), which is the expected utility of the information gained from the experiment. It encompasses the notions that the information gained must be (i) informative, (ii) result in material (practical) decisions, and ultimately (iii), lead to economic decisions. These notions are sequentially conditional, such that, in order to make an economic decision, the information must first lead to a material decision, which depends on the information gained. The idea of VOI was applied and developed in works such as Liu et al. (2012) to the remediation of groundwater sites, Trainor-Guitton et al. (2013) to monitor the detection of CO_2 leaks, Nenna and Knight (2014) to evaluate the benefits of acquiring geophysical data as part of a groundwater management strategy, and a geophysical perspective on VOI is presented by Trainor-Guitton (2014). In addition, a new concept called efficacy of information (EOI), which is similar to VOI but without financial rewards or costs, was recently introduced by Caers et al. (2022). Our approach does not take into account the financial or logistical costs of the experiment, but instead focuses on the information gain from the experiment; however, decision-theoretic metrics can easily be incorporated into our framework into a VOI or EOI framework, because information gain is the first stage of the VOI framework.

Several other dynamic processes can be monitored using time-lapse electrical techniques (Singha et al., 2015; Hermans et al., 2018), and the methodology could be applied to cases such as the monitoring of a contaminant plume (Power et al., 2014; Robinson et al., 2020; Tso et al., 2020; Nazifi et al., 2022) or saltwater intrusion (SWI) (Johnson et al., 2015; Palacios et al., 2020). For example, to assess the efficiency of a remediation strategy for a contaminated aquifer, the position of a well can be optimized to minimize the cost of drilling, while still fulfilling the desired accuracy in the prediction of the contaminant concentration. In SWI-prone areas, it could also be used to minimize the uncertainty in the most vulnerable areas in the prediction of the saltwater front by optimizing the number and positions of measurements. Other geophysical monitoring applications could also benefit from the methodology. Wilkinson et al. (2015) review some examples of geophysical monitoring applications. Newer developments include applications such as CO_2 sequestration (Auken et al., 2014; Lu et al., 2015; Pezard et al., 2016; Schmidt-Hattenberger et al., 2016), soil properties for agriculture applications (Blanchy

et al., 2020), or even erosion (Masi et al., 2020) and ground-loosening processes (Kim et al., 2022).

We also point out that the approach can be used to combine datasets of any type and any dimension in order to stochastically solve geophysical inverse problems. Although different geophysical methods are complementary, JafarGandomi and Binley (2013) argue that inconsistency in their acquisition geometries (i.e., 1D, 2D, or 3D) makes the data fusion process difficult. In their inversion strategy, they re-casted the subsurface model to a set of locally layered 1D earth-models to overcome the incompatibility of the acquisition geometries and also to reduce the computational cost of the 2D and 3D inversions. Our fusion method easily combines datasets of any dimension because it first reduces their dimensionality before concatenating them and feeding them to the learning algorithm.

6 Conclusion

In this paper, we present a method for optimizing 4D temperature field monitoring experiments using a Bayesian approach. The proposed methodology uses a combination of ODE and Bayesian inference to identify informative observation well locations. We apply our method to a synthetic case study involving the prediction of a four-dimensional temperature field from data collected by electrical resistivity tomography (ERT) and four observation wells recording the temperature over time. These predictors of different nature are combined in the principal component space to form a new predictor. Using our method, the optimal ERT electrode configuration is determined prior to optimization, and we found that, in our case, the multiple-gradient array outperforms the dipole-dipole array in terms of information gain. Following the training step, targets are sampled from the inferred posterior distribution in a low-dimensional latent space using transport map methods, which are a powerful tool in our Bayesian inference framework for sampling an unknown target given a known predictor. Applying a simple metric (RMSE) to the principal components of the predicted and true targets allows to determine the locations of observation wells that minimize uncertainty. This method can be used to optimize the design of 4D temperature field monitoring experiments and to reduce the cost of data collection by choosing a threshold between precision and number and nature of the data.

Our findings indicate that the placement of observation wells must take into account the direction of the heat plume, and that an observation network that includes wells both downstream and upstream of the original injection well is optimal. A combination of various observation wells and geophysical data always yields the best results. We also noticed that increasing the number of observation wells does not always improve prediction accuracy if the wells are not placed in the proper locations and thus provide redundant or no information. Careful consideration of the information provided by the different data types and how they should be combined is necessary prior to optimization. Our method provides a straightforward procedure for examining these trade-offs, and is thus a useful technique for determining the optimal sensor combination.

One of our framework's limitations is that it requires prior knowledge of the unknown model parameters as well as knowledge of the physics of the problem to allow forward modeling, which can be computationally expensive. However, with a small number of examples (200 for training and 50 for testing), the method is able to predict the posterior distribution of the temperature field with a reasonable accuracy and to find the best combination of data sources among 4 prescribed well positions and their combinations, with the caveat that the training set should be as representative as possible of the real data. As a more general drawback, OED can not guarantee assumptions such as experiment effectiveness, model form validity, and criterion reflecting experiment objectives. To satisfy the aforementioned assumptions, OED must carefully consider the experiment (Smucker et al., 2018).

The proposed method is not limited to the monitoring of temperature fields and can be applied to the prediction of any type of high-dimensional target from a fusion of multiple predictors in a wide range of contexts, such as geophysical, environmental, and engineering applications.

7 Open Research

The software developments related to this paper are distributed in the software SKBEL (v2.0.0) published on GitHub under the BSD 3-Clause License (Thibaut & Ramgraber, 2021). Additionally, the dataset is available on Kaggle (Lesparre et al., 2022) under the terms of the CC BY-SA 4.0 License. It includes the data used in the paper (Temperature field monitoring with ERT data) and Jupyter notebooks to reproduce the results.

References

Arato, A., Boaga, J., Comina, C., Seta, M. D., Sipio, E. D., Galgaro, A., ... Mandrone, G. (2015, 8). Geophysical monitoring for shallow geothermal applications – two italian case histories. *First Break*, *33*. Retrieved from `https://www.earthdoc.org/content/journals/0.3997/1365-2397.33.8.82010` doi: 10.3997/1365-2397.33.8.82010

Athens, N. D., & Caers, J. K. (2019, 12). A monte carlo-based framework for assessing the value of information and development risk in geothermal exploration. *Applied Energy*, *256*, 113932. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S0306261919316198` doi: 10.1016/j.apenergy.2019.113932

Attia, A., Alexanderian, A., & Saibaba, A. K. (2018, 9). Goal-oriented optimal design of experiments for large-scale bayesian linear inverse problems. *Inverse Problems*, *34*, 095009. doi: 10.1088/1361-6420/aad210

Auken, E., Doetsch, J., Fiandaca, G., Christiansen, A. V., Gazoty, A., Cahill, A. G., & Jakobsen, R. (2014, 2). Imaging subsurface migration of dissolved co2 in a shallow aquifer using 3-d time-lapse electrical resistivity tomography. *Journal of Applied Geophysics*, *101*, 31-41. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S0926985113002681` doi: 10.1016/j.jappgeo.2013.11.011

Baptista, R., Marzouk, Y., & Zahm, O. (2022). On the representation and learning of monotone triangular transport maps. *arXiv*. Retrieved from `https://arxiv.org/abs/2009.10303v2` doi: 10.48550/ARXIV.2009.10303

Bayer, P., Rybach, L., Blum, P., & Brauchler, R. (2013, 10). Review on life cycle environmental effects of geothermal power generation. *Renewable and Sustainable Energy Reviews*, *26*, 446-463. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S1364032113003420` doi: 10.1016/j.rser.2013.05.039

Bellman, R. (1961). *Adaptive control processes: A guided tour. (a rand corporation research study).* Princeton, N. J.: Princeton University Press, XVI, 255 p. (1961).

Blanchy, G., Watts, C. W., Richards, J., Bussell, J., Huntenburg, K., Sparkes, D. L., ... Binley, A. (2020, 1). Time-lapse geophysical assessment of agricultural practices on soil moisture dynamics. *Vadose Zone Journal*, *19*. Retrieved from `https://onlinelibrary.wiley.com/doi/10.1002/vzj2.20080` doi: 10.1002/vzj2.20080

Bridger, D. W., & Allen, D. M. (2010, 1). Heat transport simulations in a heterogeneous aquifer used for aquifer thermal energy storage (ates). *Canadian Geotechnical Journal*, *47*, 96-115. Retrieved from `http://www.nrcresearchpress.com/doi/10.1139/T09-078` doi: 10.1139/T09-078

Brouyère, S. (2001). *Etude et modélisation du transport et du piégeage des solutés en milieu souterrain variablement saturé (study and modelling of transport and*

retardation of solutes in variably saturated media). (PhD thesis ed.). Retrieved from https://hdl.handle.net/2268/40804

Brunner, P., & Simmons, C. T. (2012, 3). Hydrogeosphere: A fully integrated, physically based hydrological model. *Ground Water*, *50*, 170-176. doi: 10.1111/j.1745-6584.2011.00882.x

Caers, J., Scheidt, C., Yin, Z., Wang, L., Mukerji, T., & House, K. (2022, 6). Efficacy of information in mineral exploration drilling. *Natural Resources Research*, *31*, 1157-1173. doi: 10.1007/s11053-022-10030-1

Chaloner, K., & Verdinelli, I. (1995, 8). Bayesian experimental design: A review. *Statistical Science*, *10*. doi: 10.1214/ss/1177009939

Dahlin, T., & Zhou, B. (2006). Multiple-gradient array measurements for multichannel 2d resistivity imaging. *Near Surface Geophysics*, *4*, 113-123. doi: 10.3997/1873-0604.2005037

Dassargues, A. (1997, 3). Modeling baseflow from an alluvial aquifer using hydraulic-conductivity data obtained from a derived relation with apparent electrical resistivity. *Hydrogeology Journal*, *5*, 97-108. Retrieved from http://link.springer.com/10.1007/s100400050125 doi: 10.1007/s100400050125

Derouane, J., & Dassargues, A. (1998, 11). Delineation of groundwater protection zones based on tracer tests and transport modeling in alluvial sediments. *Environmental Geology*, *36*, 27-36. Retrieved from http://link.springer.com/10.1007/s002540050317 doi: 10.1007/s002540050317

Duijff, R., Bloemendal, M., & Bakker, M. (2021, 12). Interaction effects between aquifer thermal energy storage systems. *Groundwater*, gwat.13163. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/gwat.13163 doi: 10.1111/gwat.13163

Eidsvik, J., Mukerji, T., & Bhattacharjya, D. (2015). *Value of information in the earth sciences.* Cambridge University Press. Retrieved from https://www.cambridge.org/core/product/identifier/9781139628785/type/book doi: 10.1017/CBO9781139628785

El Moselhy, T. A., & Marzouk, Y. M. (2012, October). Bayesian inference with optimal maps. *Journal of Computational Physics*, *231*(23), 7815–7850. Retrieved 2022-03-01, from https://linkinghub.elsevier.com/retrieve/pii/S0021999112003956 doi: 10.1016/j.jcp.2012.07.022

Ferguson, G. (2007, 7). Heterogeneity and thermal modeling of ground water. *Ground Water*, *45*, 485-490. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/j.1745-6584.2007.00323.x doi: 10.1111/j.1745-6584.2007.00323.x

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation.* New York: Oxford University Press.

Haan, S., Ramos, F., & Müller, R. D. (2021, 1). Multiobjective bayesian optimization and joint inversion for active sensor fusion. *Geophysics*, *86*, ID1-ID17. Retrieved from https://library.seg.org/doi/10.1190/geo2019-0460.1 doi: 10.1190/geo2019-0460.1

Hermans, T. (2017, 1). Prediction-focused approaches: An opportunity for hydrology. *Groundwater*, *55*, 683-687. doi: 10.1111/gwat.12548

Hermans, T., Lesparre, N., De Schepper, G., & Robert, T. (2019, August). Bayesian evidential learning: a field validation using push-pull tests. *Hydrogeology Journal*, *27*(5), 1661–1672. Retrieved 2020-11-21, from http://link.springer.com/10.1007/s10040-019-01962-9 doi: 10.1007/s10040-019-01962-9

Hermans, T., Nguyen, F., Klepikova, M., Dassargues, A., & Caers, J. (2018, 1). Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using bayesian evidential learning. *Water Resources Research*, *54*, 2931-2948. doi: 10.1002/2017WR022135

Hermans, T., Nguyen, F., Robert, T., & Revil, A. (2014, 8). Geophysical methods for monitoring temperature changes in shallow low enthalpy geothermal

systems. *Energies*, *7*, 5083-5118. Retrieved from http://www.mdpi.com/1996-1073/7/8/5083 doi: 10.3390/en7085083

Hermans, T., Oware, E., & Caers, J. (2016, 9). Direct prediction of spatially and temporally varying physical properties from time-lapse electrical resistance data. *Water Resources Research*, *52*, 7262-7283. Retrieved from http://doi.wiley.com/10.1002/2016WR019126 doi: 10.1002/2016WR019126

Hermans, T., Vandenbohede, A., Lebbe, L., & Nguyen, F. (2012, 1). A shallow geothermal experiment in a sandy aquifer monitored using electric resistivity tomography. *GEOPHYSICS*, *77*, B11-B21. Retrieved from https://library.seg.org/doi/10.1190/geo2011-0199.1 doi: 10.1190/geo2011-0199.1

Hermans, T., Wildemeersch, S., Jamin, P., Orban, P., Brouyère, S., Dassargues, A., & Nguyen, F. (2015). Quantitative temperature monitoring of a heat tracing experiment using cross-borehole ert. *Geothermics*, *53*, 14-26. doi: 10.1016/j.geothermics.2014.03.013

JafarGandomi, A., & Binley, A. (2013, 9). A bayesian trans-dimensional approach for the fusion of multiple geophysical datasets. *Journal of Applied Geophysics*, *96*, 38-54. doi: 10.1016/j.jappgeo.2013.06.004

Johnson, T., Versteeg, R., Thomle, J., Hammond, G., Chen, X., & Zachara, J. (2015). Four-dimensional electrical conductivity monitoring of stage-driven river water intrusion: Accounting for water table effects using a transient mesh boundary and conditional inversion constraints. *Water Resources Research*, *51*, 6177-6196. doi: 10.1002/2014WR016129

Kim, B., Joung, I., Cho, A., Shin, D., Han, Y., & Nam, M. (2022). Monitoring the perturbation zone near a foundation excavation with electrical resistivity tomography: Comparison between time-lapse 3d and 2d inversions in single-profile study. *Journal of Applied Geophysics*, *205*. doi: 10.1016/j.jappgeo.2022.104772

Kleinegesse, S., & Gutmann, M. U. (2020, 13–18 Jul). Bayesian experimental design for implicit models by mutual information neural estimation. In H. D. III & A. Singh (Eds.), *Proceedings of the 37th international conference on machine learning* (Vol. 119, pp. 5316–5326). PMLR. Retrieved from https://proceedings.mlr.press/v119/kleinegesse20a.html

Klepikova, M. V., Borgne, T. L., Bour, O., Dentz, M., Hochreutener, R., & Lavenant, N. (2016, 7). Heat as a tracer for understanding transport processes in fractured media: Theory and field assessment from multiscale thermal push-pull tracer tests. *Water Resources Research*, *52*, 5442-5457. Retrieved from http://doi.wiley.com/10.1002/2016WR018789 doi: 10.1002/2016WR018789

Laloy, E., & Vrugt, J. A. (2012, jan). High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. *Water Resources Research*, *48*(1). Retrieved from http://doi.wiley.com/10.1029/2011WR010608 doi: 10.1029/2011WR010608

Lesparre, N., Compaire, N., Hermans, T., & Thibaut, R. (2022). *4D Temperature Monitoring* [Dataset]. Kaggle. Retrieved from https://www.kaggle.com/dsv/3819983 doi: 10.34740/KAGGLE/DSV/3819983

Lesparre, N., Robert, T., Nguyen, F., Boyle, A., & Hermans, T. (2019, 1). 4d electrical resistivity tomography (ert) for aquifer thermal energy storage monitoring. *Geothermics*, *77*, 368-382. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0375650517303772 doi: 10.1016/j.geothermics.2018.10.011

Lindley, D. V. (1956, 12). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, *27*, 986-1005. doi: 10.1214/aoms/1177728069

Liu, X., Lee, J., Kitanidis, P. K., Parker, J., & Kim, U. (2012, 4). Value of

information as a context-specific measure of uncertainty in groundwater remediation. *Water Resources Management*, *26*, 1513-1535. doi: 10.1007/s11269-011-9970-3

Lu, C., Zhang, C., Hunag, H., & Johnson, T. C. (2015, 2). Monitoring co ¡sub¿2¡/sub¿ sequestration into deep saline aquifer and associated salt intrusion using coupled multiphase flow modeling and time-lapse electrical resistivity tomography. *Greenhouse Gases: Science and Technology*, *5*, 34-49. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/ghg.1437 doi: 10.1002/ghg.1437

Lykkegaard, M. B., & Dodwell, T. J. (2022, 6). Where to drill next? a dual-weighted approach to adaptive optimal design of groundwater surveys. *Advances in Water Resources*, *164*, 104219. doi: 10.1016/j.advwatres.2022.104219

Macfarlane, A., Förster, A., Merriam, D., Schrötter, J., & Healey, J. (2002, 12). Monitoring artificially stimulated fluid movement in the cretaceous dakota aquifer, western kansas. *Hydrogeology Journal*, *10*, 662-673. Retrieved from http://link.springer.com/10.1007/s10040-002-0223-7 doi: 10.1007/s10040-002-0223-7

Masi, M., Ferdos, F., Losito, G., & Solari, L. (2020). Monitoring of internal erosion processes by time-lapse electrical resistivity tomography. *Journal of Hydrology*, *589*. doi: 10.1016/j.jhydrol.2020.125340

Meloun, M., & Militký, J. (Eds.). (2012). *Statistical data analysis: a practical guide* (Reprinted ed.). New Delhi: WPI, Woodhead Publ. India Pvt. Ltd. (OCLC: 696087077)

Michel, H., Hermans, T., & Nguyen, F. (2022, 9). Iterative prior resampling and rejection sampling to improve 1d geophysical imaging based on bayesian evidential learning (bel1d). *Geophysical Journal International*. doi: 10.1093/gji/ggac372

Michel, H., Nguyen, F., Kremer, T., Elen, A., & Hermans, T. (2020, 5). 1d geological imaging of the subsurface from geophysical data with bayesian evidential learning. *Computers & Geosciences*, *138*, 104456. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0098300419306028 doi: 10.1016/j.cageo.2020.104456

Nazifi, H. M., Gülen, L., Gürbüz, E., & Pekşen, E. (2022). Time-lapse electrical resistivity tomography (ert) monitoring of used engine oil contamination in laboratory setting. *Journal of Applied Geophysics*, *197*. doi: 10.1016/j.jappgeo.2022.104531

Nenna, V., & Knight, R. (2014, 1). Demonstration of a value of information metric to assess the use of geophysical data for a groundwater application. *GEOPHYSICS*, *79*, E51-E60. doi: 10.1190/geo2012-0474.1

Palacios, A., Ledo, J. J., Linde, N., Luquot, L., Bellmunt, F., Folch, A., ... Carrera, J. (2020). Time-lapse cross-hole electrical resistivity tomography (chert) for monitoring seawater intrusion dynamics in a mediterranean aquifer. *Hydrology and Earth System Sciences*, *24*, 2121-2139. doi: 10.5194/hess-24-2121-2020

Palmer, C. D., Blowes, D. W., Frind, E. O., & Molson, J. W. (1992, 10). Thermal energy storage in an unconfined aquifer: 1. field injection experiment. *Water Resources Research*, *28*, 2845-2856. Retrieved from http://doi.wiley.com/10.1029/92WR01471 doi: 10.1029/92WR01471

Park, B.-H., Bae, G.-O., & Lee, K.-K. (2015, 11). Importance of thermal dispersivity in designing groundwater heat pump ({GWHP}) system: {Field} and numerical study. *Renewable Energy*, *83*, 270-279. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S096014811500316X doi: 10.1016/j.renene.2015.04.036

Park, J., & Caers, J. (2020, October). Direct forecasting of global and spatial model parameters from dynamic data. *Computers & Geosciences*, *143*, 104567. Re-

trieved 2021-01-30, from https://linkinghub.elsevier.com/retrieve/pii/S0098300420305562 doi: 10.1016/j.cageo.2020.104567

Pezard, P. A., Denchik, N., Lofi, J., Perroud, H., Henry, G., Neyens, D., ... Levannier, A. (2016, 5). Time-lapse downhole electrical resistivity monitoring of subsurface co 2 storage at the maguelone shallow experimental site (languedoc, france). *International Journal of Greenhouse Gas Control*, *48*, 142-154. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1750583615301523 doi: 10.1016/j.ijggc.2015.12.005

Polydorides, N., & Lionheart, W. R. B. (2002, 12). A matlab toolkit for three-dimensional electrical impedance tomography: a contribution to the electrical impedance and diffuse optical reconstruction software project. *Measurement Science and Technology*, *13*, 1871-1883. doi: 10.1088/0957-0233/13/12/310

Power, C., Gerhard, J., Karaoulis, M., Tsourlos, P., & Giannopoulos, A. (2014). Evaluating four-dimensional time-lapse electrical resistivity tomography for monitoring dnapl source zone remediation. *Journal of Contaminant Hydrology*, *162-163*, 27-46. doi: 10.1016/j.jconhyd.2014.04.004

Pradhan, A., & Mukerji, T. (2020, 6). Seismic bayesian evidential learning: estimation and uncertainty quantification of sub-resolution reservoir properties. *Computational Geosciences*, *24*, 1121-1140. Retrieved from http://link.springer.com/10.1007/s10596-019-09929-1 doi: 10.1007/s10596-019-09929-1

Qiang, S., Shi, X., Kang, X., & Revil, A. (2022, 7). Optimized arrays for electrical resistivity tomography survey using bayesian experimental design. *GEOPHYSICS*, *87*, E189-E203. doi: 10.1190/geo2021-0408.1

Robert, Paulus, Bolly, Lin, K. S., & Hermans. (2019, 9). Heat as a proxy to image dynamic processes with 4d electrical resistivity tomography. *Geosciences*, *9*, 414. Retrieved from https://www.mdpi.com/2076-3263/9/10/414 doi: 10.3390/geosciences9100414

Robinson, J., Johnson, T., & Rockhold, M. (2020). Feasibility assessment of long-term electrical resistivity monitoring of a nitrate plume. *Groundwater*, *58*, 224-237. doi: 10.1111/gwat.12899

Ryan, E. G., Drovandi, C. C., McGree, J. M., & Pettitt, A. N. (2016, 4). A review of modern computational algorithms for bayesian optimal design. *International Statistical Review*, *84*, 128-154. doi: 10.1111/insr.12107

Saner, D., Juraske, R., Kübert, M., Blum, P., Hellweg, S., & Bayer, P. (2010, 9). Is it only co2 that matters? a life cycle perspective on shallow geothermal systems. *Renewable and Sustainable Energy Reviews*, *14*, 1798-1813. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1364032110001061 doi: 10.1016/j.rser.2010.04.002

Satija, A., & Caers, J. (2015, March). Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. *Advances in Water Resources*, *77*, 69–81. Retrieved 2021-01-30, from https://linkinghub.elsevier.com/retrieve/pii/S0309170815000044 doi: 10.1016/j.advwatres.2015.01.002

Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying Uncertainty in Subsurface Systems*. Hoboken, NJ, USA: John Wiley & Sons, Inc. Retrieved 2020-09-29, from http://doi.wiley.com/10.1002/9781119325888 doi: 10.1002/9781119325888

Scheidt, C., Renard, P., & Caers, J. (2015, February). Prediction-Focused Subsurface Modeling: Investigating the Need for Accuracy in Flow-Based Inverse Modeling. *Mathematical Geosciences*, *47*(2), 173–191. Retrieved 2021-01-30, from http://link.springer.com/10.1007/s11004-014-9521-6 doi: 10.1007/s11004-014-9521-6

Schmidt-Hattenberger, C., Bergmann, P., Labitzke, T., Wagner, F., & Rippe, D. (2016). Permanent crosshole electrical resistivity tomography (ert) as an established method for the long-term co¡inf¿2¡/inf¿ monitoring at the ketzin pilot

site. *International Journal of Greenhouse Gas Control*, *52*, 432-448. doi: 10.1016/j.ijggc.2016.07.024

Singha, K., Day-Lewis, F. D., Johnson, T., & Slater, L. D. (2015, 3). Advances in interpretation of subsurface processes with time-lapse electrical imaging. *Hydrological Processes*, *29*, 1549-1576. doi: 10.1002/hyp.10280

Smucker, B., Krzywinski, M., & Altman, N. (2018, 8). Optimal experimental design. *Nature Methods*, *15*, 559-560. doi: 10.1038/s41592-018-0083-2

Sommer, W., Valstar, J., van Gaans, P., Grotenhuis, T., & Rijnaarts, H. (2013, 12). The impact of aquifer heterogeneity on the performance of aquifer thermal energy storage. *Water Resources Research*, *49*, 8128-8138. Retrieved from `http://doi.wiley.com/10.1002/2013WR013677` doi: 10.1002/2013WR013677

Sommer, W. T., Doornenbal, P. J., Drijver, B. C., van Gaans, P. F. M., Leusbrock, I., Grotenhuis, J. T. C., & Rijnaarts, H. H. M. (2014, 2). Thermal performance and heat transport in aquifer thermal energy storage. *Hydrogeology Journal*, *22*, 263-279. Retrieved from `http://link.springer.com/10.1007/s10040-013-1066-0` doi: 10.1007/s10040-013-1066-0

Spantini, A., Baptista, R., & Marzouk, Y. (2022). Coupling techniques for nonlinear ensemble filtering. *SIAM Review, in press*. Retrieved from `https://arxiv.org/abs/1907.00389v2`

Spantini, A., Bigoni, D., & Marzouk, Y. (2018). Inference via low-dimensional couplings. *arXiv*. Retrieved from `https://arxiv.org/abs/1703.06131` doi: 10.48550/arxiv.1703.06131

Tadjer, A., & Bratvold, R. B. (2021). Managing uncertainty in geological co2 storage using bayesian evidential learning. *Energies*, *14*. doi: 10.3390/en14061557

Tarakanov, A., & Elsheikh, A. H. (2020, October). Optimal bayesian experimental design for subsurface flow problems. *Computer Methods in Applied Mechanics and Engineering*, *370*, 113208. Retrieved 2021-07-14, from `https://linkinghub.elsevier.com/retrieve/pii/S0045782520303935` doi: 10.1016/j.cma.2020.113208

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics. Retrieved from `http://epubs.siam.org/doi/book/10.1137/1.9780898717921` doi: 10.1137/1.9780898717921

Tarantola, A. (2006, 8). Popper, bayes and the inverse problem. *Nature Physics*, *2*, 492-494. doi: 10.1038/nphys375

Thibaut, R., Laloy, E., & Hermans, T. (2021, 12). A new framework for experimental design using bayesian evidential learning: The case of wellhead protection area. *Journal of Hydrology*, *603*, 126903. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S0022169421009537` doi: 10.1016/j.jhydrol.2021.126903

Thibaut, R., & Ramgraber, M. (2021, September). *SKBEL - Bayesian Evidential Learning framework built on top of scikit-learn* [Software]. Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.6205242` doi: 10.5281/zenodo.6205242

Trainor-Guitton, W. J. (2014, 3). A geophysical perspective of value of information: examples of spatial decisions for groundwater sustainability. *Environment Systems and Decisions*, *34*, 124-133. doi: 10.1007/s10669-013-9487-9

Trainor-Guitton, W. J., Ramirez, A., Yang, X., Mansoor, K., Sun, Y., & Carroll, S. (2013, 10). Value of information methodology for assessing the ability of electrical resistivity to detect co2/brine leakage into a shallow aquifer. *International Journal of Greenhouse Gas Control*, *18*, 101-113. doi: 10.1016/j.ijggc.2013.06.018

Tso, C.-H., Johnson, T., Song, X., Chen, X., Kuras, O., Wilkinson, P., ... Binley, A. (2020). Integrated hydrogeophysical modelling and data assimilation for

geoelectrical leak detection. *Journal of Contaminant Hydrology*, *234*. doi: 10.1016/j.jconhyd.2020.103679

Uhlemann, S., Wilkinson, P. B., Maurer, H., Wagner, F. M., Johnson, T. C., & Chambers, J. E. (2018, 7). Optimized survey design for electrical resistivity tomography: combined optimization of measurement configuration and electrode placement. *Geophysical Journal International*, *214*, 108-121. doi: 10.1093/gji/ggy128

Vandenbohede, A., Hermans, T., Nguyen, F., & Lebbe, L. (2011, 10). Shallow heat injection and storage experiment: Heat transport simulation and sensitivity analysis. *Journal of Hydrology*, *409*, 262-272. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0022169411005701 doi: 10.1016/j.jhydrol.2011.08.024

Vandenbohede, A., Louwyck, A., & Lebbe, L. (2009, 1). Conservative solute versus heat transport in porous media during push-pull tests. *Transport in Porous Media*, *76*, 265-287. Retrieved from http://link.springer.com/10.1007/s11242-008-9246-4 doi: 10.1007/s11242-008-9246-4

Vanhoudt, D., Desmedt, J., Bael, J. V., Robeyn, N., & Hoes, H. (2011, 12). An aquifer thermal storage system in a belgian hospital: Long-term experimental evaluation of energy and cost savings. *Energy and Buildings*, *43*, 3657-3665. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0378778811004427 doi: 10.1016/j.enbuild.2011.09.040

Villani, C. (2009). *Optimal Transport* (Vol. 338). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved 2022-03-01, from http://link.springer.com/10.1007/978-3-540-71050-9 doi: 10.1007/978-3-540-71050-9

Vrugt, J. A. (2016, January). Markov chain monte carlo simulation using the DREAM software package: Theory, concepts, and matlab implementation. *Environmental Modelling & Software*, *75*, 273–316. Retrieved 2021-02-05, from https://linkinghub.elsevier.com/retrieve/pii/S1364815215300396 doi: 10.1016/j.envsoft.2015.08.013

Wagner, F. M., Günther, T., Schmidt-Hattenberger, C., & Maurer, H. (2015, 3). Constructive optimization of electrode locations for target-focused resistivity monitoring. *GEOPHYSICS*, *80*, E29-E40. Retrieved from https://library.seg.org/doi/10.1190/geo2014-0214.1 doi: 10.1190/geo2014-0214.1

Wagner, V., Li, T., Bayer, P., Leven, C., Dietrich, P., & Blum, P. (2014, 2). Thermal tracer testing in a sedimentary aquifer: field experiment (lauswiesen, germany) and numerical simulation. *Hydrogeology Journal*, *22*, 175-187. Retrieved from http://link.springer.com/10.1007/s10040-013-1059-z doi: 10.1007/s10040-013-1059-z

Wang, Y., Zechner, M., Mern, J. M., Kochenderfer, M. J., & Caers, J. K. (2022, 8). A sequential decision-making framework with uncertainty quantification for groundwater management. *Advances in Water Resources*, *166*, 104266. doi: 10.1016/j.advwatres.2022.104266

Wildemeersch, S., Jamin, P., Orban, P., Hermans, T., Klepikova, M., Nguyen, F., ... Dassargues, A. (2014, 11). Coupling heat and chemical tracer experiments for estimating heat transfer parameters in shallow alluvial aquifers. *Journal of Contaminant Hydrology*, *169*, 90-99. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0169772214001247 doi: 10.1016/j.jconhyd.2014.08.001

Wilkinson, P. B., Uhlemann, S., Meldrum, P. I., Chambers, J. E., Carrière, S., Oxby, L. S., & Loke, M. (2015, 10). Adaptive time-lapse optimized survey design for electrical resistivity tomography monitoring. *Geophysical Journal International*, *203*, 755-766. Retrieved from https://academic.oup.com/gji/article-lookup/doi/10.1093/gji/ggv329 doi: 10.1093/gji/ggv329

Yin, Z., Strebelle, S., & Caers, J. (2020, 2). Automated monte carlo-based quan-

tification and updating of geological uncertainty with borehole data (autobel v1.0). *Geoscientific Model Development*, *13*, 651-672. Retrieved from https://gmd.copernicus.org/articles/13/651/2020/ doi: 10.5194/gmd-13-651-2020

Zhang, J., Zeng, L., Chen, C., Chen, D., & Wu, L. (2015, 1). Efficient bayesian experimental design for contaminant source identification. *Water Resources Research*, *51*, 576-598. Retrieved from http://doi.wiley.com/10.1002/2014WR015740 doi: 10.1002/2014WR015740