Estimation of Hydraulic Conductivity in a Watershed Using Multi-source Data via Co-Kriging and Bayesian Experimental Design

Chien-Yung Tseng[1,2], Maryam Ghadiri[1,2,*], Timothy H. Larson[3], Praveen Kumar[2], Hadi Meidani[2]

[1] Illinois Water Resources Center, Prairie Research Institute, University of Illinois at Urbana-Champaign
[2] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign.
[3] Illinois State Geological Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign
* Corresponding author: Maryam Ghadiri (mghadiri@illinois.edu)

Abstract

Enhanced water management systems depend on accurate estimation of hydraulic properties of subsurface formations. This is while hydraulic conductivity of geologic formations could vary significantly. Therefore, using information only from widely spaced boreholes will be insufficient in characterizing subsurface aquifer properties. Hence, there is a need for other sources of information to complement our hydro-geophysics understanding of a region of interest. This study presents a numerical framework where information from different measurement sources is combined to characterize the 3-dimensional random field representing the hydraulic conductivity of a watershed in a Multi-Fidelity estimation model. Coupled with this model, a Bayesian experimental design will also be presented that is used to select the best future sampling locations. This work draws upon unique capabilities of electrical resistivity tests as well as statistical inversion. It presents a Multi-Fidelity Gaussian Processes (Kriging) model to estimate the geological properties in Upper Sangamon Watershed in east central Illinois, using multi-source observation data, obtained from electrical resistivity and pumping tests. We demonstrate the accuracy of Co-Kriging that is dependent on the locations and the distribution of both the high- and low-fidelity data, and also discuss its comparison with Single-High-Fidelity Kriging results. The uncertainties and confidence in the measurements and parameter estimates are then quantified and are in turn used to design future cycles of data collection to further improve the confidence intervals.

Highlights:

- A Multi-Fidelity Kriging model was designed to estimate the geological properties by different sources of data.
- Bayesian experimental design is used to select the best future sampling locations.
- We investigated how a more accurate model can "learn" from new sensors using probabilistic statistical tools.

1 Introduction

Reliable prediction of hydraulic properties of subsurface formations is a crucial step in improving water management systems. There are various testing approaches to obtain information from the area of interest. Among others, electrical methods such as electrical resistivity (ER) and electromagnetic induction (EMI) are broadly used in hydro-geophysics investigations (Lesmes, et al., 2005). These tests can be used individually or in combination with frequencies ranging from direct current (DC) to >1GHz to provide information about the subsurface (Lesmes, et al., 2005). Several studies have examined factors influencing relations between electrical resistivity and hydraulic properties of aquifers and aquifer materials (Kelly, 1977; Mazáč, et al., 1985). There are also studies in direct and inverse relationships between hydraulic conductivities, rock resistivities and the role of the distribution of hydraulic conductivity on dynamics of pollution spreading in rock medium (Mazáč, et al., 1990).

Hydraulic conductivity of geologic formations could vary by orders of magnitude over relatively small spatial scales, hence characterizing subsurface aquifer properties using just the information acquired from widely spaced boreholes is challenging (Lesmes, et al., 2005). One method employed by (Lesmes, et al., 2005) is to use an integrated exploration approach in which borehole and geophysical data sets are jointly interpreted. A closer look in the literature on predicting hydraulic properties of subsurface formations in watersheds still reveals a number of outstanding questions, e.g., How many field tests need

to be conducted to achieve the desired accuracy of our estimation? Where should be the location of the future tests, and how does the cost associated with different tests affect future test designs?

There are a number of studies on how to build groundwater models using information at different levels of fidelity (Asher, et al., 2015; Zhang, et al., 2018). Specifically, these Multi-Fidelity (MF) models combine both Low-Fidelity (LF) data with lower associated cost and accuracy with High-Fidelity (HF) data and approximate the output with an accuracy that is better than that offered by a Single-Fidelity (SF) model (Peherstorfer, et al., 2018; Fernández-Godino, et al., 2016). In general, in building MF models, one fits different Surrogate Models (SMs), to available data points obtained from sensors with different noise levels. SM approximations are models with minimal computational cost that can effectively offer estimation and prediction without the need to obtain a large number of expensive tests or to run expensive numerical simulations (Forrester, et al., 2007) (Fernández-Godino, et al., 2016).

Among the surrogate models, Gaussian Process Regression (Kriging) approach have been more widely used in MF groundwater modeling (Zaytsev, et al., 2017). Kriging properties facilitates the usage of MF models which combine HF function, a more accurate but expensive representation of a physical phenomenon, and a LF function, a less accurate but inexpensive representation of a physical phenomenon, while constructing a surrogate model (Zaytsev, et al., 2017). Review papers by (Asher, et al., 2015) and (Fernández-Godino, et al., 2016) extensively surveyed several data-driven methods of combining fidelities with the main focus on MF surrogate models among which Kriging has become a very popular surrogate particularly for MF applications. This is particularly due to the fact that Kriging entails an uncertainty structure that readily lends itself to MF modeling (Fernández-Godino, et al., 2016). Moreover, a recent research by (Zheng, et al., 2018) employed MF Gaussian surrogate to propose an adaptive MF ensemble smoother for data assimilation to reduce the high computational cost for characterization of model parameters in ensemble-based methods.

In a more recent study (Menberg, et al., 2020), authors have used MF approach to Bayesian parameter estimation in subsurface heat and fluid transport models for the first time. The study includes

information from a few physically more accurate (HF) but expensive parametric model outputs and a larger number of evaluations from a less accurate, less expensive LF model. They concluded that the results from the analytical and numerical model which combine low resolution model with the data from only a few runs of a higher resolution model substantially improved the posterior distribution results (Menberg, et al., 2020).

In this study, we have made the following contributions. We present a quantitative MF framework and for the first time combine information from Electrical Earth Resistivity (EER) tests and pumping tests (as two tests with different accuracies) to enhance the understanding about the geological and hydrological characteristics. Also, for the first time, we investigated how future tests with different fidelities should be conducted to optimally enhance our understanding about the hydraulic properties of a region. Specifically, we studied an intensively managed area located in the Upper Sangamon Watershed in Central Illinois, U.S.A., and generated 2D maps of hydraulic conductivity over a large-scale region with quantified uncertainties in different depth layers. In doing so, we made use of low cost, small-scale measurements obtained from the EER together with more accurate, more expensive pumping tests in a calibration framework based on Kriging. We also investigated how a more accurate model can "learn" from new sensors using probabilistic statistical tools, and how the best locations for future data collection can be selected.

Our approach is based on the Bayesian experimental design, which selects the best locations, out of a set of candidate locations, based on the value of information that each location is expected to offer (Norberg, et al., 2006). By relating the expected value of information from each location to the present levels of uncertainties in the MF Kriging model, we form and solve the experimental design problem. The proposed method can serve as a quantitative decision support framework to optimally conduct tests with different cost and accuracy levels.

The remainder of this paper is organized as follows. In Section 2, we provide the technical background, which includes detailed information about site selection, observation data, Lognormal Ordinary Kriging (LOK), MF Lognormal Ordinary Co-Kriging, and optimal Bayesian experimental design. In Section 3, we show the topography of the Upper Sangamon Watershed and discuss how the EER and pumping test data were obtained, and how the multi-source data were used in SF Kriging with multiple data sources and MF Co-Kriging. Discussions are provided in Section 4 for the effect of fidelity on the estimated field and the estimation accuracy followed by comments on the cost associated with the LF and HF data. Also, the application of optimal Bayesian experimental design for obtaining optimal future sampling locations is presented. Conclusions are provided in section 5.

2 Method

2.1 Site Selection

The Sangamon River is a major tributary to the Illinois River with the confluence near Chandlerville in Cass County. The drainage area of the Sangamon River at the Decatur dam is 925 square miles. The headwaters of the Sangamon River are in McLean County near the town of Ellsworth. The watershed lies across seven counties in east-central Illinois: Champaign, Christian, Dewitt, Ford, Macon, McLean, and Piatt. The major urban areas within the watershed are Decatur, Monticello, Mahomet, Rantoul, and Gibson City.

This watershed is intensively managed for soybean and corn production and is among the five watersheds in Illinois that are identified as most in need of attention for water supply planning and management (Mattia, et al., 2018). The predominant land use in the watershed is row crop agriculture, which composes nearly 90 percent of the land area (Keefer, et al., 2005)**.** As one of the Intensively Managed Landscapes (IMLs), this region is at risk for deterioration of land as well as water systems in our

121 environment. Hence, more observations are needed to understand and predict the behavior of natural
122 services (ecological, hydrological, and climatic services) which support basic human needs such as water,
123 food, and energy.

124 The geological formation of our study area consists of Glacial and other related deposits which
125 are grouped into four major lithostratigraphic units (from oldest to youngest); 1-Banner Formation which
126 is divided into three units; a lower, middle, and upper unit. The lower and middle Banner are typically
127 considered as one unit since the sediments of the lower Banner are not clearly distinguished from the
128 sediments of the middle Banner Formation. Sand and gravel deposits in the Mahomet Bedrock Valley and
129 those in the Mackinaw Bedrock Valley are the predominant sediments in the two bedrock valleys , 2- the
130 Glasford Formation which primarily consists of till deposits, 3- the Wedron Group which predominantly
131 consists of till deposits and 4- the Mason Group which primarily includes sand and gravel deposits
132 (Roadcap, et al., 2011).

133 Moreover, Hydro-geologic framework of Quaternary deposits in east-central Illinois consists of Mahomet
134 aquifer; aquifers in the upper Banner Formation; aquifers in the lower Glasford Formation; aquifers in the
135 upper Glasford Formation; Shallow and surficial aquifers (Soller, et al., 1999).

136 **2.2 Data Description**

137 Two field observation data, Electrical Earth Resistivity (EER) measurement and pumping test, are
138 used as different fidelity data sources to estimate hydraulic conductivity of USRW (see Figure 1 for the
139 data locations). The pumping test involves pumping from a test well at a controlled rate and monitoring the
140 flow rate through the drawdown at different locations along the radial axis from the test well. Hydraulic
141 conductivity values of aquifer material as determined from pump tests and aquifer tests will vary spatially
142 but wouldn't change much temporally. Repeated pump tests may show changes in hydraulic conductivity
143 at the well skin of production wells (or within gravel packs immediately surrounding well screens) but
144 repeated long-duration aquifer tests would give us similar values through time. The method is well

recognized as one of the most reliable ways of measuring soil hydraulic conductivity with HF. However, due to the high cost of drilling a well, limited data can be collected. EER measurement has also been widely applied to estimate hydraulic conductivity of the subsurface based on a two-dimensional resistivity model of the relations between aquifer hydraulic and electrical properties (Kelly, et al., 1985); (Slater, 2007); (Khalil, et al., 2009); (Tizro, et al., 2010). The measurement relies on testing with dipole-dipole electrode configuration in a vertical two-dimensional plane of the field to infer the hydraulic conductivity, whose accuracy depends on the equipment precision. It has lower cost, but also lower fidelity compared to the pumping test. In general, EER values are known to vary. The greatest variability is due to changes in saturation conditions. Within the Upper Sangamon River Watershed, water table is within a few meters of the ground surface except in areas of very steep slopes and very coarse materials. This is usually a small percentage of the entire area and is neglected at the scale of this study. There is some variation due to temperature, but we neglect this as being small within the context of this generalized study.

The EER measurement provides a continuous estimation of hydraulic conductivity in a small vertical plane (~800 m long and ~80 m deep). A recent study done by (Lu, et al., 2019) shows that the relationship between soil's hydraulic conductivity, K, and Electrical Conductivity, σ_{EC}, follows an exponential function form as: $K = ae^{-b\sigma_{EC}} + c$, where the parameters a, b and c can be estimated using the calculated σ_{EC} in soil layers which is the inverse of the resistivity data captured by EER test. According to (Lu, et al., 2019) when K is a dependent variable, the given best fitted empirical parameters can be obtained by fitting to a comprehensive data set:

$$\begin{cases} a = 299.6e^{-0.001147\sigma_{EC}} + 157 \\ b = 0.2061e^{-0.0001535\sigma_{EC}} + 0.004299 \\ c = 7.996e^{-0.0001264\sigma_{EC}} + 0.6567 \end{cases} \quad (1)$$

In this study, the horizontal mean value was set as the representative value in each depth for the LF data input in MF Co-Kriging model. The equipment uncertainty of EER measurement is of the order of 1 (ohm-m), giving the initial variance of K as 10^{-3} (cm/s) (Kelly, et al., 1985), which was set as the nugget value for the EER data in the Co-Kriging model. The pumping tests were conducted at specific locations with

different depths. Unlike EER measurement that provides continuous vertical information, pumping test gives point information of subsurface properties from the measurement of flow velocities within soil pores. It offers higher accuracy of hydraulic conductivity, which is later set as the HF data source in Co-Kriging. In this study we had EER data at 15 locations with continuous depth and pumping test data at 68 locations with specific depth for each one. The EER and pumping tests were conducted by Illinois State Geological Survey and Illinois State Water Survey, and the locations of the tests were originally selected to aid in Quaternary mapping projects and to develop communities' water supply planning and management.

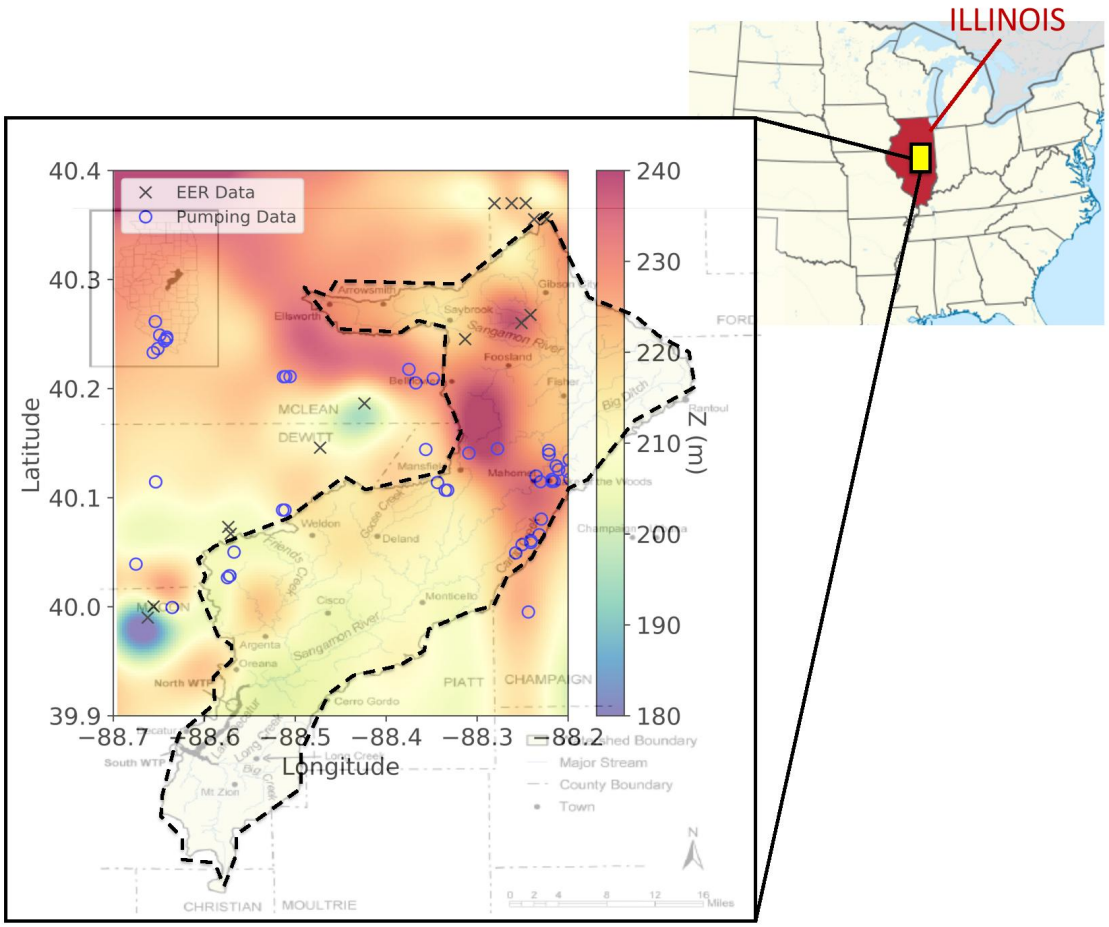


Figure 1. Locations for data in the USRW in Illinois, USA. The Black dashed line represents Sangamon River Watershed. Blue circle markers represent the pumping test data locations. Black cross markers represent the EER data locations.

2.3 Lognormal Ordinary Kriging

Lognormal Ordinary Kriging (LOK) is a popular geostatistical procedure that generates an estimated mapping of geo-properties from a scattered set of points with scalar values based on a logarithmic transformation of the estimators (Balaban, et al., 2018). Compared to the traditional Ordinary Kriging (OK) model, LOK model can reduce the data variance and improve the calculation of statistics and weighted averages that avoid negative or extreme estimated values (Roth , 1998). As shown in Figure 2, the positively skewed distribution of K can be observed in both pumping test data and EER data under a normal scale. However, after we transformed the data on a log scale, the data looks more symmetric and the variance is greatly reduced.

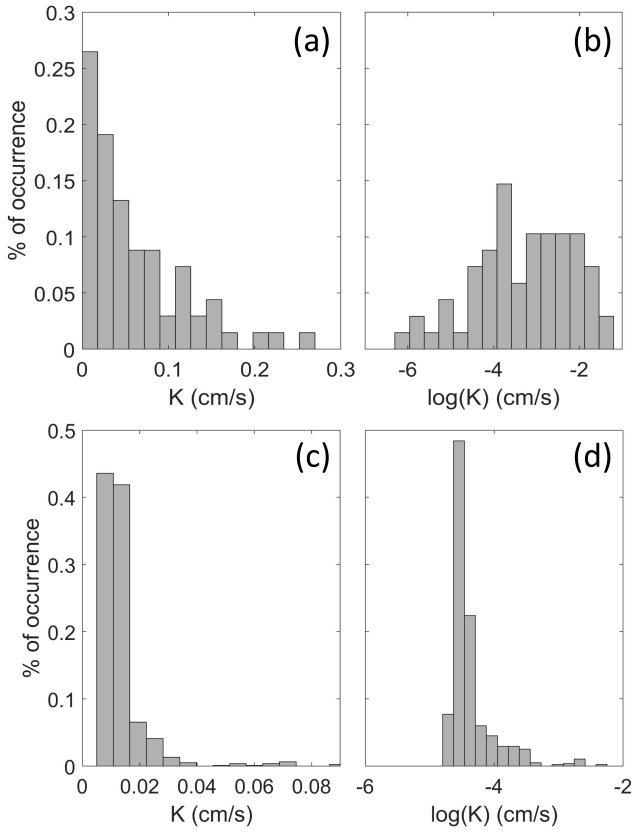


Figure 2. Histogram of the measured hydraulic conductivities from the pumping test under (a) normal scale and (b) log scale; from the EER test under (c) normal scale and (d) log scale.

The LOK model algorithm follows the structure of Gaussian Processes:

$$ln(\boldsymbol{y}) = f(\boldsymbol{x}) \sim GP(0, \boldsymbol{K}) \quad (2)$$

where $\boldsymbol{x} = \{x_i\}$ represents the locations of the data points, $\boldsymbol{y} = \{y_i\}$ represents the measured hydraulic conductivity corresponding to the locations $\boldsymbol{x}$, $\boldsymbol{K} = \{K_{ij}\}$ is a symmetric matrix, which is constructed by the Kriging function $k(x_i, x_j; \theta)$ with exponential variogram through the following equation:

$$K_{ij} = k(x_i, x_j; \theta) = n + s\left(1 - e^{-\frac{|x_i - x_j|}{r/3}}\right) \quad (3)$$

where $\theta = (n, s, r)$ are the Kriging parameters, namely Nugget (n), Sill (s), and Range (r). Nugget is usually specified according to the observation errors, while Sill and Range can be obtained by fitting the sample variogram according to the Kriging function.

Given the observation data $\{\boldsymbol{x}, \boldsymbol{y}\}$, the semivariance, γ, of LOK model can be expressed as:

$$\gamma(d_{ij}) = \frac{1}{2}E\left[ln(y_i) - ln(y_j)\right]^2 \quad (4)$$

where $d_{ij} = |x_i - x_j|$ and E(.) is the expectation operator that returns the mean value. Then for the estimations at a set of new locations of points $\boldsymbol{x}^*$, normal distribution is applied:

$$\begin{bmatrix} f(\boldsymbol{x}^*) \\ f(\boldsymbol{x}) \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} k(\boldsymbol{x}^*, \boldsymbol{x}^*; \theta) & k(\boldsymbol{x}^*, \boldsymbol{x}; \theta) \\ k(\boldsymbol{x}, \boldsymbol{x}^*; \theta) & \boldsymbol{K} \end{bmatrix}\right) \quad (5)$$

According to the resulting conditional distribution, estimations at a given point is given by

$$f(\boldsymbol{x}^*|\boldsymbol{x}) \sim N(\boldsymbol{\mu_l}, \boldsymbol{\sigma_l}) \quad (6)$$

where

$$\boldsymbol{\mu_l} = k(\boldsymbol{x}^*, \boldsymbol{x}; \theta)K^{-1}\boldsymbol{y} \quad (7)$$

$$\boldsymbol{\sigma_l} = k(\boldsymbol{x}^*, \boldsymbol{x}^*; \theta) - k(\boldsymbol{x}^*, \boldsymbol{x}; \theta)K^{-1}k(\boldsymbol{x}, \boldsymbol{x}^*; \theta) \quad (8)$$

Since $f(\boldsymbol{x}^*)$ is in logarithmic scale, in order to estimate the parameter of interest (in our case the hydraulic conductivity), we need to convert the logarithmic values, $\boldsymbol{\mu_l}$ and $\boldsymbol{\sigma_l}$, back to the actual mean and standard deviation values according to:

$$\boldsymbol{\mu}^* = exp\left(\boldsymbol{\mu_l} + \frac{\boldsymbol{\sigma_l}^2}{2}\right) \quad (9)$$

$$\boldsymbol{\sigma}^* = \sqrt{[exp(\boldsymbol{\sigma_l}^2) - 1]exp(2\boldsymbol{\mu_l} + \boldsymbol{\sigma_l}^2)} \quad (10)$$

2.4 Multi-Fidelity Lognormal Ordinary Co-Kriging

To combine the observation data from EER measurement and pumping test, the MF Lognormal Co-Kriging model is used to perform a two-dimensional hydraulic conductivity mapping in different depth layers with smooth and continuous fusion of information from two sources with different fidelity/precision. The Co-Kriging algorithm follows the structure proposed by (Kennedy, et al., 2000) (Forrester, et al., 2007), assuming that

$$u_L(\boldsymbol{x}) \sim GP\big(0, k_L(\boldsymbol{x}, \boldsymbol{x}; \theta_L)\big) \quad (11)$$

$$u_H(\boldsymbol{x}) \sim GP\big(0, k_H(\boldsymbol{x}, \boldsymbol{x}; \theta_H)\big) \quad (12)$$

are two independent kriging functions. Then, the LF and HF LOK functions can be modeled as $f_L(\boldsymbol{x}) = u_L(x)$ and $f_H(\boldsymbol{x}) = \rho u_L(\boldsymbol{x}) + u_H(\boldsymbol{x})$, respectively, which can be expressed as a multi-output LOK:

$$\begin{bmatrix} f_L(\boldsymbol{x}) \\ f_H(\boldsymbol{x}) \end{bmatrix} \sim GP\left(0, \begin{bmatrix} k_{LL}(\boldsymbol{x}, \boldsymbol{x}; \theta_L) & k_{LH}(\boldsymbol{x}, \boldsymbol{x}; \theta_L, \rho) \\ k_{HL}(\boldsymbol{x}, \boldsymbol{x}; \theta_L, \rho) & k_{HH}(\boldsymbol{x}, \boldsymbol{x}; \theta_L, \theta_H, \rho) \end{bmatrix}\right) \quad (13)$$

where

$$k_{LL}(\boldsymbol{x}, \boldsymbol{x}; \theta_L) = k_L(\boldsymbol{x}, \boldsymbol{x}; \theta_L) \quad (14)$$

$$k_{LH}(\boldsymbol{x}, \boldsymbol{x}; \theta_L, \rho) = k_{HL}(\boldsymbol{x}, \boldsymbol{x}; \theta_L, \rho) = \rho k_L(\boldsymbol{x}, \boldsymbol{x}; \theta_L) \quad (15)$$

$$k_{HH}(\boldsymbol{x}, \boldsymbol{x}; \theta_L, \theta_H, \rho) = \rho^2 k_L(\boldsymbol{x}, \boldsymbol{x}; \theta_L) + k_H(\boldsymbol{x}, \boldsymbol{x}; \theta_H) \quad (16)$$

where k_L and k_H are the Kriging functions (Equation (3)) for the LF and HF data, respectively, and ρ is the MF constant.

Given the observation LF and HF data, $\{\boldsymbol{x_L}, \boldsymbol{y_L}\}$ and $\{\boldsymbol{x_H}, \boldsymbol{y_H}\}$, the Kriging parameters θ_L and θ_H can be fitted by the sample variogram according to the Kriging functions of the LF and HF data, respectively. To obtain the optimized ρ, normal distribution is applied:

$$\mathrm{f}_{mgp}(\boldsymbol{z}) \sim N(0, \boldsymbol{K}) \quad (17)$$

where

$$\boldsymbol{z} = \begin{bmatrix} \ln(\boldsymbol{y_L}) \\ \ln(\boldsymbol{y_H}) \end{bmatrix} \quad (18)$$

$$\boldsymbol{K} = \begin{bmatrix} k_{LL}(\boldsymbol{x_L}, \boldsymbol{x_L}; \theta_L) & k_{LH}(\boldsymbol{x_L}, \boldsymbol{x_H}; \theta_L, \rho) \\ k_{HL}(\boldsymbol{x_H}, \boldsymbol{x_L}; \theta_L, \rho) & k_{HH}(\boldsymbol{x_H}, \boldsymbol{x_H}; \theta_L, \theta_H, \rho) \end{bmatrix} \quad (19)$$

and the optimized constant ρ can be trained by minimizing the negative log marginal likelihood (NLML):

$$NLML(\theta_L, \theta_H, \rho) = \frac{1}{2}\boldsymbol{y^T} K^{-1}\boldsymbol{y} + \frac{1}{2} ln|\boldsymbol{K}| + \frac{N}{2} ln(2\pi) \quad (20)$$

where N is the total number of the data points. In this study, we use TNC, a truncated Newton algorithm minimization method (Nash, 1984) to obtain the optimized constant ρ. For the estimations at a new set of points $\boldsymbol{x}^*$, we first construct the joint distribution:

$$\begin{bmatrix} f_H(\boldsymbol{x}^*) \\ \boldsymbol{z} \end{bmatrix} \sim N\left(0, \begin{bmatrix} k_{HH}(\boldsymbol{x}^*, \boldsymbol{x}^*; \theta_L, \theta_H, \rho) & \boldsymbol{q}^T \\ \boldsymbol{q} & \boldsymbol{K} \end{bmatrix}\right) \quad (21)$$

where

$$\boldsymbol{q}^T = [k_{HL}(\boldsymbol{x}^*, \boldsymbol{x_L}; \theta_L, \rho), k_{HH}(\boldsymbol{x}^*, \boldsymbol{x_H}; \theta_L, \theta_H, \rho)\,] \quad (22)$$

Like the SF LOK model, according to the resulting conditional distribution, predictions can be estimated by

$$f_H(\boldsymbol{x}^*|\boldsymbol{z}) \sim N(\boldsymbol{\mu_m}, \boldsymbol{\sigma_m}) \quad (23)$$

where

$$\boldsymbol{\mu_m} = \boldsymbol{q}^T \boldsymbol{K}^{-1} \boldsymbol{y} \qquad (24)$$

$$\boldsymbol{\sigma_m} = k_{HH}(\boldsymbol{x}^*, \boldsymbol{x}^*) - \boldsymbol{q}^T \boldsymbol{K}^{-1} \boldsymbol{q} \qquad (25)$$

Finally, we back transform the mean $\boldsymbol{\mu_m}$ and the standard deviation $\boldsymbol{\sigma_m}$ of the MF model back into normal domain:

$$\boldsymbol{\mu}^* = exp\left(\boldsymbol{\mu_m} + \frac{\boldsymbol{\sigma_m}^2}{2}\right) \qquad (26)$$

$$\boldsymbol{\sigma}^* = \sqrt{[exp(\boldsymbol{\sigma_m}^2) - 1] exp(2\boldsymbol{\mu_m} + \boldsymbol{\sigma_m}^2)} \qquad (27)$$

2.5 Optimal Bayesian Experimental Design

Our experimental design concerns the problem of identifying the best locations for future tests or data collections. These locations are identified based on the value of information that each location is expected to offer (Norberg, et al., 2006). For instance, in the context of hydraulic property estimation for aquifers, measurements that are collected from locations that are closely spaced, will provide much less information compared to those obtained from locations that are sufficiently apart. In establishing a quantitative framework that captures these facts, a Bayesian experimental design procedure can be used. This begins by quantifying the value of information. Specifically, the value of information is defined as the information gain conditioned on the design variables. The information gain is formally defined to be the Kullback-Leibler (KL) divergence from the posterior distributions of the model parameter to prior (Chaloner, et al., 1995). The best experiment, among the ensemble of candidates is the one that maximizes the information gain, taken to be the Kullback-Leibler (KL) divergence from posterior to its prior. Solving this optimization problem is numerically complicated, as the evaluation of KL divergence requires samples from the prior and posterior of the parameters. In what follows, we provide the technical background for this experimental design approach combined with the MF Co-Kriging model.

Using Bayesian inference, the posterior distribution of model parameters $p(\boldsymbol{\theta}|\boldsymbol{d},s)$ can be expressed as

$$p(\boldsymbol{\theta}|\boldsymbol{d},s)=\frac{p(\boldsymbol{\theta}|s)p(\boldsymbol{d}|\boldsymbol{\theta},s)}{p(\boldsymbol{d}|\boldsymbol{s})} \quad (28)$$

where $p(\boldsymbol{\theta}|s)$ is the prior distribution, $p(\boldsymbol{d}|\boldsymbol{\theta},s)$ is the likelihood, $p(\boldsymbol{d}|\boldsymbol{s})$ is the evidence, which can be considered as a normalizing constant

$$p(\boldsymbol{d}|s)=\int p(\boldsymbol{d}|\boldsymbol{\theta},s)\,p(\boldsymbol{\theta}|s)d\boldsymbol{\theta} \quad (29)$$

In this study, $\boldsymbol{\theta}$ is the sampled Kriging parameters, including *n*, *s*, and *r*. *n* and *s* are considered as constant values according to the MF model, while *r* is considered as Gaussian distributed samples based on the fitted LF and HF range, r_L and r_H, with $\sigma_L=0.01r_L$ and $\sigma_H=0.01r_L$. $\boldsymbol{d}$ is the sampled observation data whose probability distribution can be assumed Gaussian-like with the model-estimated μ and σ. s represents the designed future sampling location. Since the prior knowledge of $\boldsymbol{\theta}$ is not affected by s, the prior distribution

$$p(\boldsymbol{\theta}|s)=p(\boldsymbol{\theta}) \quad (30)$$

The expected utility in Bayesian experimental design can be expressed as (Lindley, 1956)

$$U(s)=\int u(s,\boldsymbol{d},\boldsymbol{\theta})p(\boldsymbol{\theta},\boldsymbol{d}|s)\,d\boldsymbol{\theta}d\boldsymbol{d} \quad (31)$$

where $u(s,\boldsymbol{d},\boldsymbol{\theta})$ is the utility function. Following the algorithm proposed by (Zhang, et al., 2015), the relative entropy from the prior to the posterior is chosen as the utility function (Lindley, 1956), which considers the expected gain in Shannon information (Shannon, 1948) given by the experiment

$$u(s,\boldsymbol{d},\boldsymbol{\theta})=\int p(\boldsymbol{\theta}|\boldsymbol{d},s)\ln[\frac{p(\boldsymbol{\theta}|\boldsymbol{d},s)}{p(\boldsymbol{\theta}|s)}]d\boldsymbol{\theta} \quad (32)$$

According to Bayes' theorem and Monte Carlo approach, the integral in Equation (31) can be approximated by the sum of the discrete values

$$U(s)\approx\frac{1}{n}\sum_{i=1}^{N}\{ln[p(d_i|\theta_i,s)]-ln[p(d_i|s)]\} \quad (33)$$

where d_i is each of the sampling data point, and N is the total number of the sampling data points. From Equation (29) and (30), the evidence $p(d_i|s)$ can also be approximated by the Monte Carlo approach

$$p(d_i|\boldsymbol{s}) = \int p(d_i|\boldsymbol{\theta}, s)\, p(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \frac{1}{n}\sum_{j=1}^{n} p\left(d_i|\theta_j, s\right) \quad (34)$$

where the likelihood $p\left(d_i|\theta_j, s\right)$ cas be expressed by a radial basis, exponential decaying function with the MF Co-Kriging model G:

$$p\left(d_i|\theta_j, s\right) = \exp\left(-\frac{1}{2}(d_i - G(\theta_j, s))^2\right) \quad (35)$$

Combining Equation (33), (34), and (35), the optimal sampling location s^* can be obtained by maximizing the expected utility $U(s)$ over the design domain D, which can be achieved by minimizing the negative $U(s)$

$$s^* = arg\max_{s \in D}[U(s)] = arg\min_{s \in D}[-U(s)] \quad (36)$$

The results of sequential Bayesian experimental design application for future sampling locations will be demonstrated in Section 4.3.

3 Results

3.1 Topography Investigation

We used the LiDAR data from the U.S. Geological Survey National Elevation Dataset (USGS NED) for USRW along with the EER and pumping test data. The LiDAR data is uniformly distributed in the rectangular region of USRW as shown in Figure 1. Multi-quadratic radial basis function with Euclidean distance was used to interpolate the elevation between the LiDAR data points. Figure 3 (a) shows that the topography of the watershed is generally flat, which is on average within a range of 210 m ~ 230 m. There is only a relatively low region in the southeastern region (~180m). The flat topography suggests that a reasonable approach would be to represent the domain in a Cartesian Coordinate system (x-y-z) denoting

z-coordinate by depth (distance from the surface), thus ignore the surface variation and set all the locations' surface as zero in depth for the z-value.

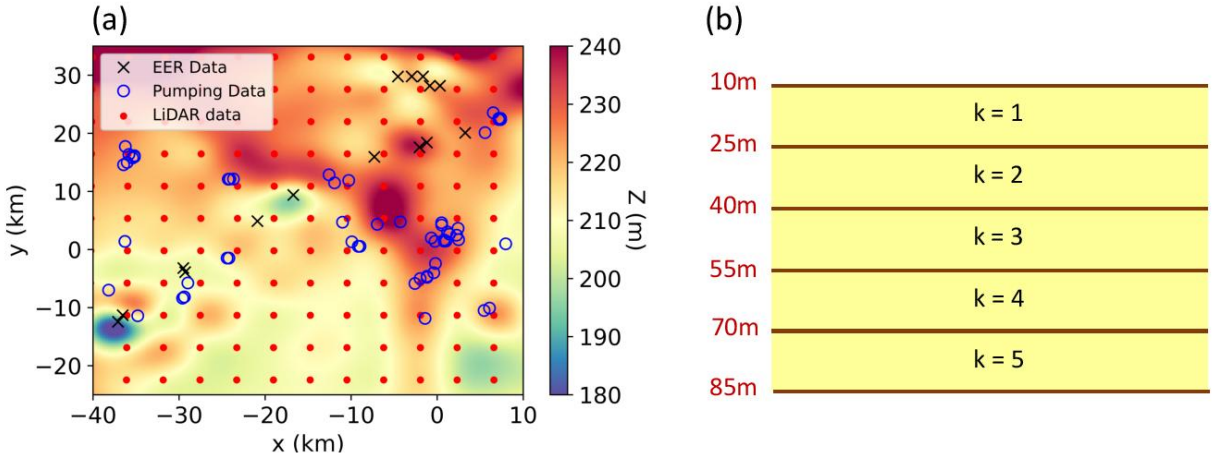


Figure 3. (a) The surface elevation map of the rectangular region of USRW shown in Figure 1. Red dots represent the LiDAR data. Blue circle markers represent the locations of EER data. Black cross markers represent the locations of pumping test data. Champaign city, IL (40°06'54"N, 88°16'22"W) is set as the origin point (x = 0 km, y = 0 km). (b) Sketch of the vertical layers setup, where k is the layer number.

3.2 Single-Fidelity Kriging Results with Multiple Data Sources

The USRW is a typical glaciated Midwest River Basin, which shows characteristic low-relief landscapes and reflects glacial deposition patterns, except for regions modified by stream processes in valleys. Hence, soil deposition patterns are expected to be a layer-by-layer distribution. The watershed contains mostly sand and gravel deposits concentrated in different layers which are typically 15 m thick (Selkregg, et al., 1958). Hence, we divided the 75 m thick domain region into five 15 m thick layers, where EER and pumping test data are located in a range between 10 m to 85 m deep from the surface as shown in Figure 3 (b). Within the same layer, soil and hydraulic properties (e.g., hydraulic conductivity) are similar and correlated across different locations. We constructed a two-dimensional (horizontal) Kriging model in different layers to construct a multi-layer mapping of hydraulic conductivity.

337 SF Kriging with multiple data sources was conducted as the reference to compare with the MF Co-
338 Kriging model. In the SF Kriging model, the data sources were treated equally, ignoring their different
339 fidelity. The exponential function-based variogram is used to fit the semivariance data (Equation 4)
340 including both EER and pumping test data on the sample variogram (Oliver, et al., 1990). The python-based
341 fitting tool, using a non-linear least squares algorithm is applied for curve fitting. The fitted Kriging
342 parameters of s = 0.61 and r = 14.36 by setting n = 0, assuming there is no initial uncertainty range from
343 the measurement. The exponential model one of the most commonly used models, which suggests that data
344 spatial autocorrelation decreases exponentially with increasing distance based on prior knowledge of the
345 phenomenon (Oliver, et al., 2015).

346 When measurements are done at irregular grid-points, setting a bandwidth, lag-tolerance, and angle
347 tolerance to account for the directional influence (anisotropic effects) can be helpful to statistically quantify
348 and analyze sample contributions in different ranges depending on the direction. However, since there is
349 limited number of the representative observation data from EER and pumping tests, we assumed isotropic
350 contribution from all the measurements without setting a bandwidth and tolerance to ensure enough data
351 points in the sample variogram.

352 In Figure 4, the SF Kriging result shows a relatively uniform distribution of mapped hydraulic
353 conductivity, K, in the first three layers (depth < 50 m), while some peak values can be observed in the last
354 two layers (depth > 50 m). This suggests more varying soil properties exist in the deeper layers of the
355 watershed. The uncertainty in the estimated properties is presented by the standard deviations, σ.

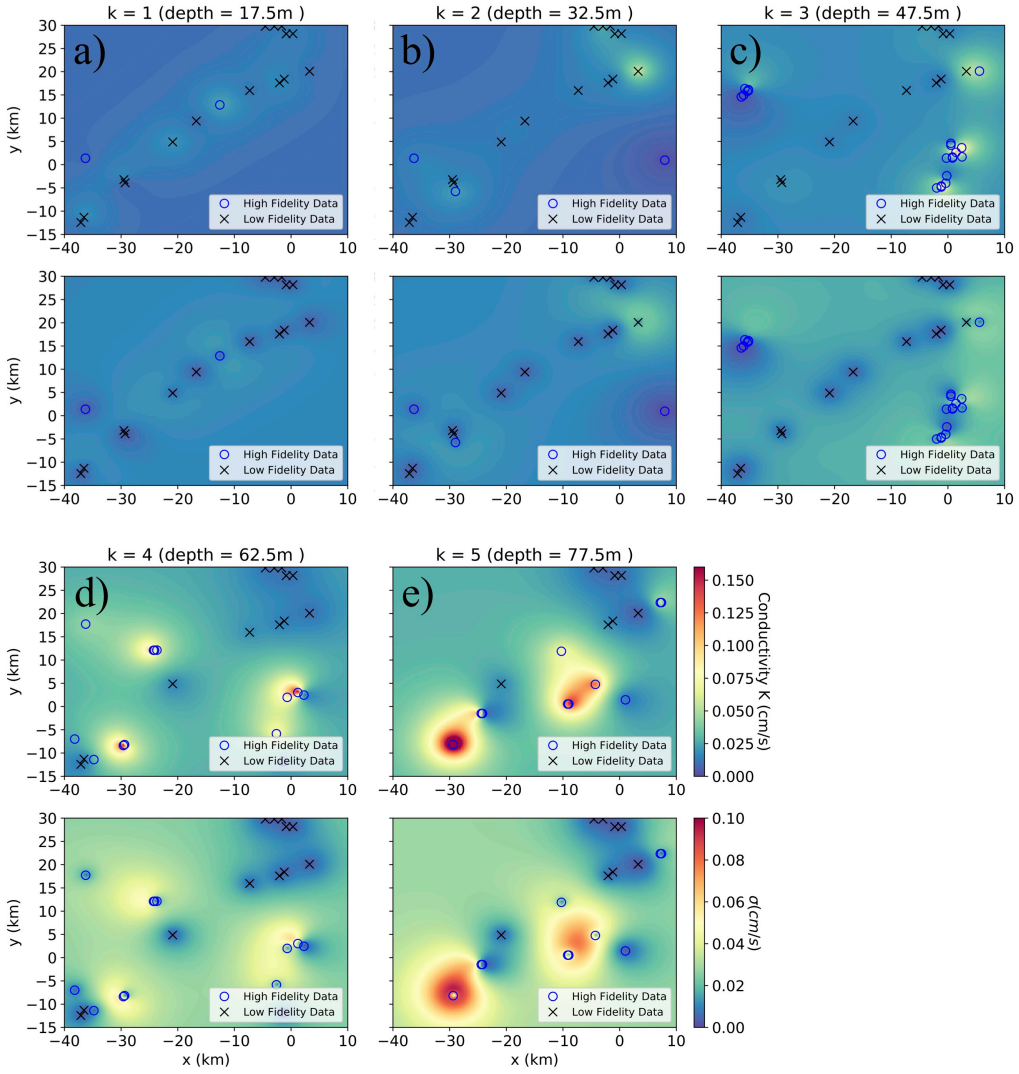


Figure 4. Single-Fidelity Kriging of the hydraulic conductivity and the corresponding standard deviation in the USWR in different depth layers. a) layer k = 1, depth = 17.5 m. b) layer k = 2, depth =32.5 m. c) layer k = 3, depth = 47.5 m. d) layer k = 4, depth = 62.5 m. e) layer k = 5, depth = 77.5 m. The value of depth shown on top of each panel is the center z-location in each layer (Figure 3 (b)). Blue circle markers represent the EER data locations. Black cross markers represent the pumping test data locations.

3.3 Multi-Fidelity Co-Kriging Results

364 The SF Kriging model didn't account for the fact that data are from different sources, and therefore,
365 they are considered with the same uncertainty. However, since different data sources typically have
366 different uncertainty/error ranges due to equipment, methods, and human factors, the fidelity of these data
367 sources should also be incorporated into the model. This makes MF Co-Kriging models a more accurate
368 approach compared to the Single Kriging model, when multiple sources of data are available. Hence, not
369 much confidence can be placed on the SF Kriging result shown in Figure 4. Data with different fidelity
370 should also be treated separately in the sample variogram for two sets of fitted Kriging parameters.

371 In MF Co-Kriging model, we treat EER and pumping test data separately when Kriging parameters
372 are obtained from each sample variogram. The fitted Kriging parameters based on the exponential function-
373 based variogram are: s = 0.18 and r = 1.93 by setting n = 9.9×10^{-5} for EER; s = 1.35 and r = 12.45 by
374 setting n = 0 for pumping test. For the EER test, n = 9.9×10^{-5} was set according to the natural log variance
375 of the initial variance (10^{-3} (cm/s)) of K from the equipment error.

376 Figure 5 shows the Multi-Fidelity Co-Kriging result of the hydraulic conductivity and the
377 corresponding standard deviation in the USRW. Compared to the Single Kriging result shown in Figure 5,
378 Multi-Fidelity Co-Kriging puts more weight on the HF data (shown by circle markers). Hence, it can be
379 seen that the estimated K and σ distribution patterns generally follow the distribution of the pumping test
380 data. It can also be seen that regions near the HF data points (blue circles in Figure 5) have lower standard
381 deviation. This means that the model assesses higher confidence in the estimates in those regions. The LF
382 data (shown by cross markers), however, do not help reduce uncertainty levels in a large area, but
383 nonetheless provide local hydraulic information in regions far away from the HF data locations.

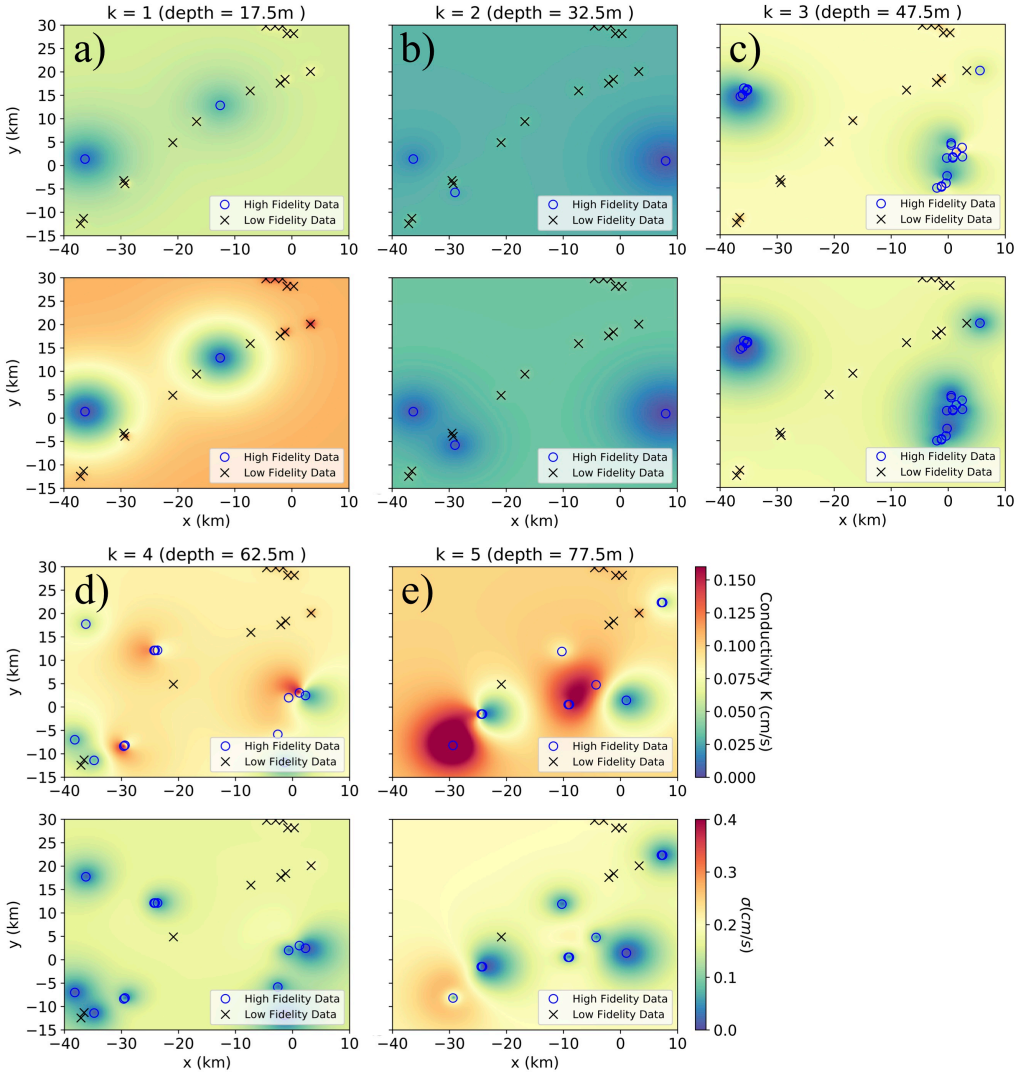
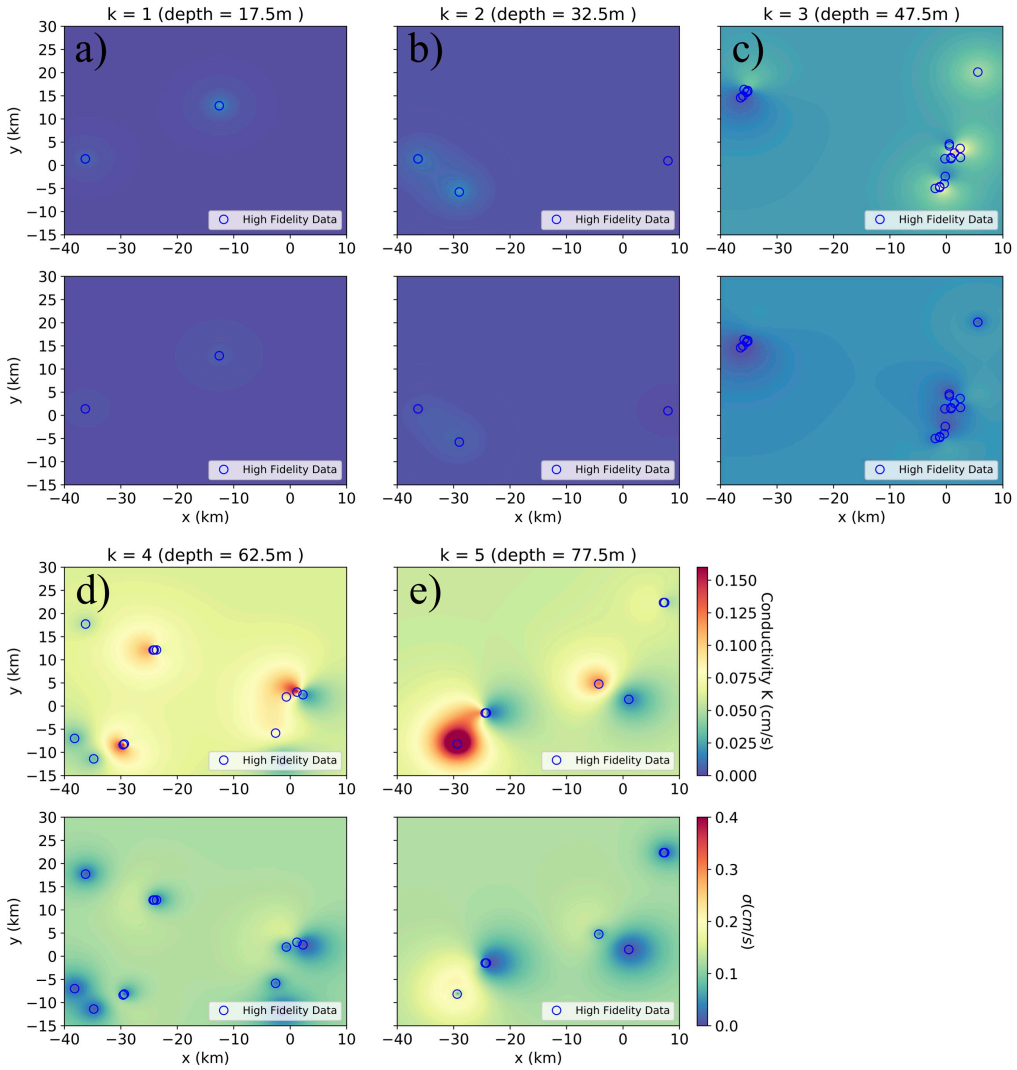


Figure 5. Multi-Fidelity Co-Kriging of the hydraulic conductivity and the corresponding standard deviation in the USRW in different depth layers. a) layer k = 1, depth = 17.5 m. b) layer k = 2, depth =32.5 m. c) layer k = 3, depth = 47.5 m. d) layer k = 4, depth = 62.5 m. e) layer k = 5, depth = 77.5 m. The value of depth shown on top of each panel is the center z-location in each layer (Figure 3 (b)). Blue circles represent the High-Fidelity data locations. Black cross markers represent the Low-Fidelity data locations.

4 Discussion

4.1 Fidelity Effect on the Predicted Field

In Multi-fidelity Co-Kriging, pumping test data was selected as High-fidelity data source due to it being a more reliable measurement method compared to EER data. Hence, the estimated K and σ values based on MF Co-Kriging (Figure 5) are mostly dominated by the HF data (pumping test data). To further study the fidelity effect, we exclude the LF data and only consider the HF data in the Kriging model (Figure 6) to compare with the MF results (Figure 5). Figure 6 shows that in regions near the HF data, both the estimated K and σ values are similar as that in Figure 5. However, in regions far from the HF data points, the models provide much different K and σ estimates, especially in the first three layers (depth < 50 m) where HF data points are scarce. The higher σ estimations are due to the additional information provided by the LF data. However, the higher estimated σ does not mean that the LF data provides wrong information, instead, the different estimations of K suggest that LF data do provide valuable information about the hydraulic conductivity properties for regions where expensive HF tests are not available or economically not feasible. We will test the accuracy of the estimated K between MF Co-kriging and Single-High-Fidelity (SHF) Kriging in the following subsection (Section 4.2).



406

407 *Figure 6. Single High-Fidelity Kriging of the hydraulic conductivity and the corresponding standard*
408 *deviation in the USRW with only High-Fidelity data (pumping test data) in different depth layers. a) layer*
409 *k = 1, depth = 17.5 m. b) layer k = 2, depth =32.5 m. c) layer k = 3, depth = 47.5 m. d) layer k = 4, depth*
410 *= 62.5 m. e) layer k = 5, depth = 77.5 m. The value of depth shown on top of each panel is the center z-*
411 *location in each layer (Figure 3 (b)). Blue circles represent the locations of HF data.*

412

413 **4.2 Fidelity Effect on the Estimation Accuracy**

To evaluate the estimated K values in the MF Co-Kriging model, we focus on the last two layers (depth > 50 m) and remove HF data points in each layer from the estimation model. We then use these removed data points as the validation data to be compared with the estimated values. In Figure 7, the red circles show the locations of the removed HF data points in each layer. The removed data points were selected based on their locations. Specifically, we preferred HF locations that were close to a LF data location in order to assess the accuracy of LF data contributions. It can be seen that the calculated standard deviation values do not differ significantly between MF model and SHF model. More HF data can increase the confidence levels in both cases. However, we do see the difference on the estimated K value in these figures.

The removed HF data points provide a reference value of $K = 0.078\ cm/s$ in the 4th layer and $K = 0.081\ cm/s$ in the 5th layer at the data locations. After the data points were removed, the MF model provided a prediction of $K = 0.081\ cm/s$ in the 4th layer and $K = 0.107\ cm/s$ in the 5th layer at the data location, while the SHF model provided an estimate of $K = 0.052\ cm/s$ in the 4th layer and $K = 0.086\ cm/s$ in the 5th layer at the data location. The above result provides an estimated accuracy. Let us now use a more precise measure of accuracy, defined as

$$a_k = 1 - (K_{pred} - K_{ref})/K_{ref} \qquad (37)$$

where K_{pred} is the predicted K from the MF or SHF model, K_{ref} is the reference K, i.e., the removed HF data. For the MF model, the accuracy is 96% in the 4th layer and 68% in the 5th layer, while for the SHF model, the accuracy is 67% in the 4th layer and 94% in the 5th layer. In the 4th layer, the removed points are far away from the other HF data points. Under this condition, LF data provides important information to enhance the estimation of the Kriging model. However, in the 5th layer, the removed points are relatively close to the other HF data points. Hence, LF data is not necessary and might provide information with higher variance due to its LF information compared to the surrounding HF information.

As a result, the accuracy of MF model depends on the locations and the distribution of both the LF and HF data. When HF data points are scarce and far away from the LF data points, the information provided from LF data becomes important and can enhance the model performance and accuracy.

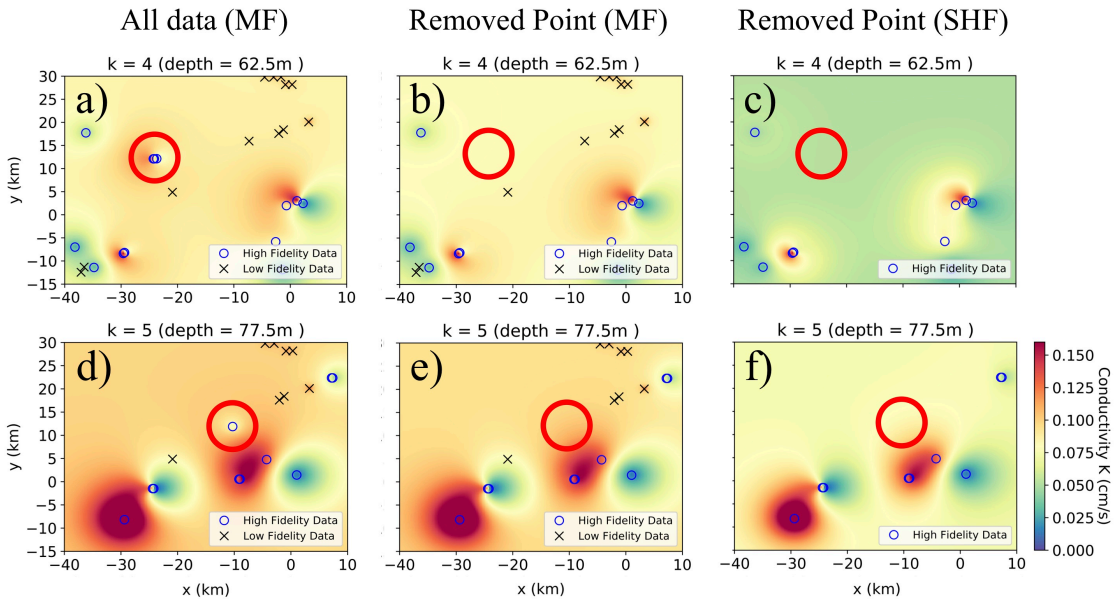


Figure 7. Comparisons between Multi-Fidelity Co-Kriging and Single-High-Fidelity Kriging with specific points removal. a) and d) Multi-Fidelity Co-Kriging of the hydraulic with all data points in the last two layers. b) and e) Multi-Fidelity Co-Kriging of the hydraulic with specific point removals in the last two layers. c) and f) High-Fidelity Kriging of the hydraulic conductivity with specific point removals in the last two layers. Blue circle markers represent the High-Fidelity data locations. Black cross markers represent the Low-Fidelity data locations. Red circles highlight the removed High-Fidelity data points.

Next, in order to further investigate the fidelity effect, we choose the deepest (5^{th}) layer as the test case and consecutively remove HF data points one by one within or close to the LF data points. Figure 8 shows the estimated hydraulic conductivity field by MF Co-Kriging and SHF Kriging under four different point-removal scenarios: keep all the HF data, remove point 1, remove points 1 and 2, and remove points 1, 2, and 3. Comparing the results of MF Co-Kriging versus SHF Kriging, it can be seen that as more data points are removed, SHF Kriging shows a relatively lower and more uniform estimated *K* field. However,

since LF data still provides the surrounding information in MF Co-Kriging, it provides estimated K values closer to the original estimates where all the data points were present.

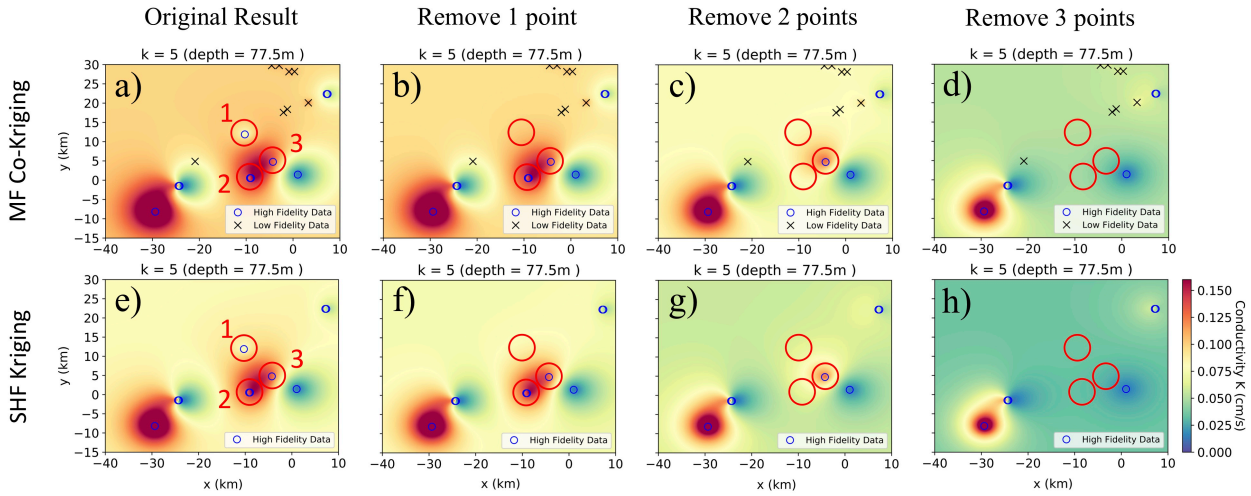


Figure 8. Comparisons between Multi-Fidelity Co-Kriging and Single-High-Fidelity Kriging in the last layer (depth >70 m) with three consecutive points removal. Multi-Fidelity Co-Kriging of the hydraulic conductivity with a) all data points. b) 1 point removal. c) 2 points removal. d) 3 points removal. Single-High-Fidelity Kriging of the hydraulic conductivity with e) all data points. f) 1 point removal. g) 2 points removal. h) 3 points removal. Blue circles represent the High-Fidelity data locations. Black cross markers represent the Low-Fidelity data locations. Red circles highlight the removed High-Fidelity data points, and the nearby red numbers in a) and e) show the removal order of the points.

Using the definition in Equation (37), Figure 9 shows the accuracy of K estimates at points 1, 2, and 3 under MF Co-Kriging and SHF Kriging when zero data point, one data point, two data points, and three data points are removed. When all the points are present, the accuracy is 100% at all three points. When point 1 is removed, the accuracy remains 100% at points 2 and 3, but at point 1, the SHF Kriging shows a greater accuracy compared to MF Co-Kriging, because of the far distance between point 1 and the LF data points as discussed in Figure 7. When both data at points 1 and 2 are removed, MF Co-Kriging shows higher accuracy both at points 1 and 2, while point 3 remains at 100% accuracy. When all three points are removed, MF Co-Kriging shows obviously higher accuracy at all the points' locations compared

to SHF Kriging. The results reconfirm that when HF data becomes scarcer, the information provided by LF data becomes more critical in MF Co-Kriging and can lead to better estimation of hydraulic conductivity.

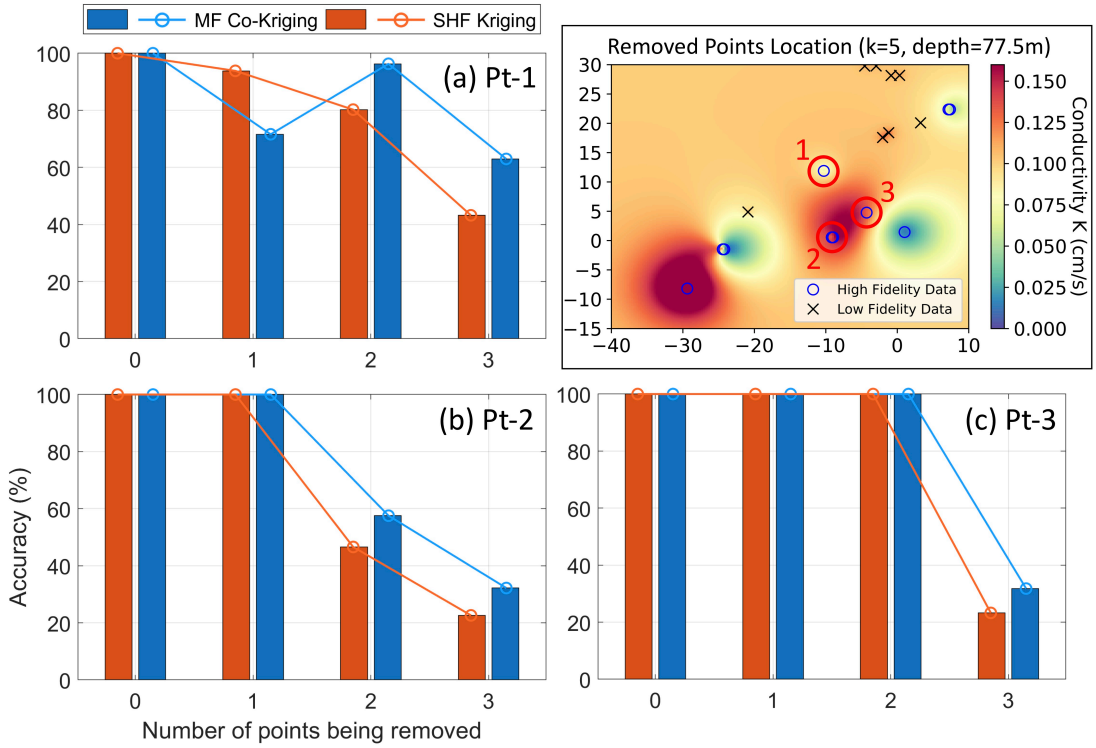


Figure 9. The accuracy of (a) point 1, (b) point 2, and (c) point 3 under Multi-Fidelity Co-Kriging and Single-High-Fidelity Kriging when removing 0 point, 1 point, 2 points, and 3 points. The removed points location is shown in the top-right panel, and the points' removing order follows the denoted number of the points.

4.3 Future Data Collection Using Bayesian Experimental Design

In this section, we apply the Bayesian experimental design along with the MF Co-Kriging model to determine the future sampling locations for the HF data (pumping test) measurement. We chose the deepest (5^{th}) layer, which has more uniform distribution of both LF and HF data points. Five optimal sampling locations for future pumping test data were estimated one by one with the initial guesses of the sampling location uniformly assigned in the simulation domain (Figure 1). Once the current optimal point was

obtained, the hydraulic conductivity value was then predicted by the MF Co-Kriging model at that location. The current estimated optimal point with its predicted hydraulic conductivity was then put back to the MF Co-Kriging model as one of the synthetic measurement data point to update the model and train the new optimized constant ρ for the next optimal sampling location.

The final optimal result is shown in Figure 10. The optimal locations are shown by the red triangles with the numbers indicating the sequential order. The sampling points are located in the region where σ is high, indicating the need of future measurements to enhance the confidence of the prediction and understanding of the region of interest. We can see that the future sampling points provide more information to the region nearby the suggested locations, where variances are greatly reduced. Variances are slightly increased in regions far away from the suggested locations due to the unbalanced information entered into the model. However, according to Bayesian Experimental Design, those regions are relatively less efficient for future measurements compared to the suggested locations when considering the expected gain in Shannon information (see Section 2.5).

The Bayesian experimental design model can be carried out for both pumping test data and EER data. However, because the pumping test data (HF data) is dominant in the MF Co-Kriging model and the pumping test is more expensive thus more limited, the appropriate future optimal locations are more critical than EER test locations from an economical perspective. Therefore, in this study, we performed the Bayesian experimental design to infer the optimal locations for future pumping tests, which can provide more valuable information, as suggested by the model. This work can be the first step of future studies on developing a more robust optimization framework that incorporates both the data cost and fidelity and can uncover their complex interplay.

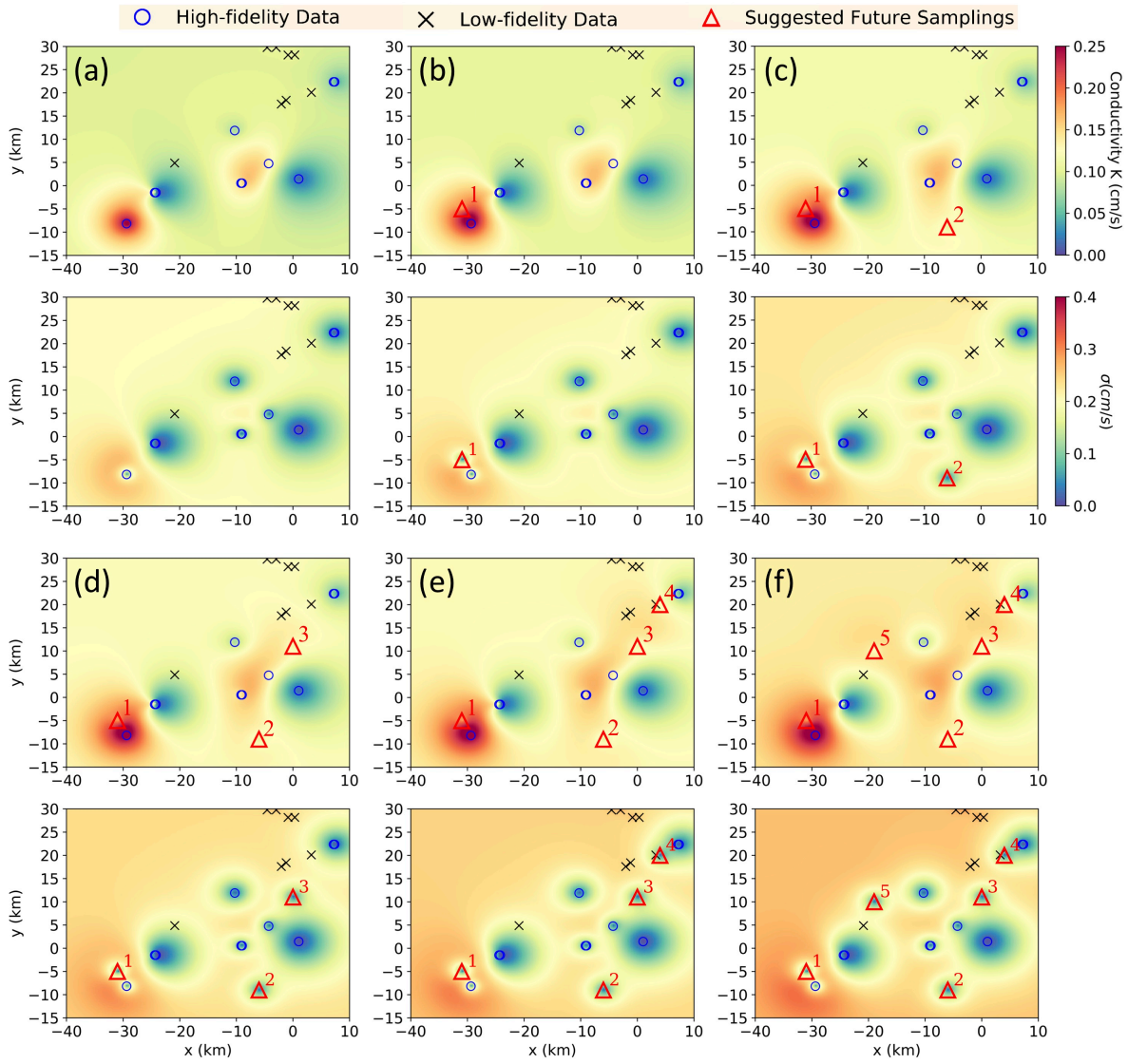


Figure 10. The suggested sequential optimal sampling locations using Bayesian experimental design with the Multi-Fidelity Co-Kriging model for the deepest (5th) layer. (a) Initial Kriging result. (b) Updated mean and variance with the 1st observation point. (c) Updated mean and variance with the 1st and 2nd observation points. (d) Updated mean and variance with the 1st, 2nd, and 3rd observation points. (e) Updated mean and variance with the 1st, 2nd, 3rd, and 4th observation points. (f) Updated mean and variance with all 5 optimal observation points. Blue circle markers represent the High-Fidelity data locations. Black cross markers represent the Low-Fidelity data locations. Red triangles represent the suggested optimal future sampling locations. The red numbers represent the order of the samplings.

5 Conclusions

This work presents a robust approach to exploit multi-source data to estimate the 3-dimensional random field of hydraulic conductivities. In particular, for the first time we demonstrated how this framework can use the combination of pumping test data from boreholes, which are expensive and more accurate, with observation data from a less expensive and less accurate test, namely the EER test. This approach offers a cost-effective approach to reliably characterize the hydraulic conductivity properties in under-sampled sites and can be particularly used in obtaining large-scale parameter maps for a region using small-scale measurements in an efficient way. The estimated values suggest that the accuracy of MF Co-Kriging depends on the locations and the distribution of both the LF and HF data. When HF data points are sparse and far away from the LF data points, the information provided from the LF data becomes crucial, and can greatly enhance the model accuracy.

HF data can provide more information to the model compared to the LF data. However, HF data are generally more costly to obtain, mainly due to their more precise testing process. For example, in this study, pumping tests require drilling wells into the ground, which roughly costs $11,000 for each 80 m well. However, the EER test is conducted completely on the surface, with no need for drilling. This makes the cost of EER test much lower, to be approximately at only $600 for a 80 m deep continuous data. There is a trade-off between deciding on the HF versus LF measurements. We observed through this work that LF data can also provide useful information to greatly enhance the parameter estimation, especially in regions where data points are sparsely-distributed. This work also includes optimal sensor placement, where the best locations for future data collection are selected by considering the current confidence levels estimated by the Kriging model, which is related to the expected value of information from future sensor data. In order to rigorously inform the decision as to what should be the combination of LF and HF measurements, our plan for a future study is to develop a more holistic optimization framework that incorporates both the data cost and fidelity and can uncover their complex interplay.

Acknowledgments

This research is funded under the provisions of section 104 of the Water Resources Research Act annual base grants (104b) program made possible and distributed through the Illinois Water Resources Center and United States Geological Survey. C-Y.T. acknowledges support from Illinois Water Resources Center and Department of Civil and Environmental Engineering University of Illinois at Urbana Champaign. M.G. acknowledges support from Illinois Water Resources Center and United States Geological Survey. T.H.L. acknowledges support from United States Geological Survey. P.K. acknowledges support from the National Science Foundation Grant EAR1331906 for the Critical Zone Observatory for Intensively Managed Landscapes (IML-CZO), a multi-institutional collaborative effort. H.M. acknowledges support from Department of Civil and Environmental Engineering University of Illinois at Urbana Champaign. Instrumentation and technical support were provided by Illinois State Geological Survey. Special thanks are given to Dr. Andrew Stumpf in Illinois State Geological Survey and Daniel R. Hadley in Illinois State Water Survey for providing data and supporting this research.

Codes and data presented in this article are available on GitHub through Zenodo.

References

Asher M. J. [et al.] A review of surrogate models and their application to groundwater modeling [Journal] // Water Resourch Research. - [s.l.] : Water Resources Research, 2015. - 8 : Vol. 51. - pp. 5957-5973.

Asher M. J. [et al.] A review of surrogate models and their application to groundwater modeling [Journal] // Water Resourch Research. - 2015. - 8 : Vol. 51. - pp. 5957-5973.

Balaban Muzaffer and Dengiz Berna Lognormal ordinary kriging metamodel in simulation optimization [Journal] // Operations Research and Applications: An International Journal (ORAJ). - 2018. - 5(1).

Chaloner K. and Verdinelli I. Bayesian experimental design: A review [Journal] // Statistical Science. - 1995. - pp. 273-304.

Fernández-Godino M. Giselle [et al.] Review of multi-fidelity models [Journal] // arXiv preprint. - 2016. - arXiv:1609.07196.
Forrester Alexander I. J., Sóbester András and Keane Andy J. Multi-fidelity optimization via surrogate modelling [Journal] // The royal Society. - 2007. - Vol. 463. - pp. 3251–3269.
Keefer Laura and Bauer Erin Watershed monitoring for the Lake Decatur watershed: 2000-2003 [Report]. - [s.l.] : ISWS Contract Report CR-2005-09, 2005.
Kelly William E. and Frohlich Reinhard K. Relations Between Aquifer Electrical and Hydraulic Properties [Journal] // Groundwater. - 1985. - 23(2). - pp. 182-189.
Kelly William E. Geoelectric sounding for estimating aquifer hydraulic conductivity [Journal] // Ground Water. - 1977. - 15(6). - pp. 420-425.
Kennedy Marc C. and O'Hagan Anthony Predicting the output from a complex computer code when fast approximations are available [Journal] // Biometrika. - 2000. - 87(1). - pp. 1-13.
Khalil Mohamed Ahmed and Santos Fernando A. Monterio Influence of degree of saturation in the electric resistivity– hydraulic conductivity relationship [Journal] // Surveys in Geophysics. - 2009. - 6 : Vol. 30. - p. 601.
Larson D. R. and Herzog B. L. and Larson, T. H. Groundwater Geology of DeWitt, Piatt, and Northern Macon Counties, Illinois. [Report]. - Champaign : Illinois State Geological Survey Environmental Geology 155, 2003.
Lesmes David P. and Friedman Shmulik P. Relationships between the Electrical and Hydrogeological Properties of Rocks and Soils [Book Section] // Hydrogeophysics. - [s.l.] : Springer, 2005. - Vol. 50.
Lindley D. V. On a measure of the information provided by an experiment [Journal] // The Annals of Mathematical Statistics. - 1956. - pp. 986-1005.
Lu Chengpeng [et al.] A convenient method to estimate soil hydraulic conductivity using electrical T conductivity and soil compaction degree [Journal]. - [s.l.] : Journal Of Hydrology , 2019. - Vol. 575.
Mattia Chloe M., Lovell Sarah Taylor and Davis Adam Identifying barriers and motivators for adoption of multifunctional perennial cropping systems by landowners in the Upper Sangamon River Watershed, Illinois [Journal]. - [s.l.] : Agroforestry Systems, 2018. - 92(5). - pp. 1155-1169.
Mazáč O., Kelly W. E. and Landa I. A hydrogeophysical model for relations between electrical and hydraulic properties of aquifers [Journal] // Journal Of Hydrology. - 1985. - 1-2 : Vol. 79. - pp. 1-19.
Mazáč Oldřich [et al.] Determination of Hydraulic Conductivities by Surface Geoelectrical Methods [Book Section] // Geotechnical and Environmental Geophysics: Volume II, Environmental and Groundwater. - [s.l.] : Society of Exploration Geophysicists, 1990.
Menberg Kathrin [et al.] Multi-fidelity approach to Bayesian parameter estimation in subsurface heat and fluid transport models [Journal] // Sciene of the Total Environement. - [s.l.] : Science of the Total Environment, 2020. - 25 : Vol. 745. - p. 140846.
Nash S. G. Newton-Type Minimization Via the Lanczos Method [Journal] // SIAM Journal of Numerical Analysis. - 1984. - Vol. 21. - pp. 770-778.
Nickerson Cynthia [et al.] Trends in US farmland values and ownership [Report]. - [s.l.] : United States Department of Agriculture, 2012.
Norberg Tommy and Rosn Lars Calculating the optimal number of contaminant samples by means of data worth analysis [Journal] // Environmetrics. - 2006. - pp. 705–719.
Oliver M. A. and Webster R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging [Journal]. - [s.l.] : Catena, 2014. - 113.
Oliver M. A. and Webster R. Basic steps in geostatistics: the variogram and kriging [Book]. - New York : SpringerBriefs in Agriculture, 2015.
Oliver M. A. and Webster R. Kriging: a method of interpolation for geographical information systems [Journal]. - [s.l.] : International Journal of Geographical Information System, 1990. - 4(3).
Peherstorfer Benjamin, Willcox Karen and Gunzburger Max Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization [Journal] // Siam Review. - 2018. - 3 : Vol. 60. - pp. 550-591.

618 **Roadcap George S., Knapp H. Vernon and Wehrmann H. Allen and Larson, David R.** Meeting East-
619 Central Illinois Water Needs to 2050: Potential Impacts on the Mahomet Aquifer and Surface Reservoirs
620 [Report]. - [s.l.] : Illinois State Water Survey, Prairie Research Institute , 2011.
621 **Roadcap George S., Knapp H. Vernon and Wehrmann H. Allen and Larson, David R.** Meeting East-
622 Central Illinois Water Needs to 2050: Potential Impacts on the Mahomet Aquifer and Surface Reservoirs
623 [Report]. - Champaign-Urbana : lllinois State Water Survey, Prairie Research Institute, 2011.
624 **Roth Chris** Is lognormal kriging suitable for local estimation? [Journal] // Mathematical Geology. -
625 1998. - 30(8). - pp. 999-1009.
626 **Selkregg L. F. and Kempton J. P.** Groundwater geology in east-central Illinois, a preliminary geologic
627 report [Journal]. - Urbana : Illinois State Geological Survey, 1958. - Vol. Circular no. 248.
628 **Shannon C. E.** A mathematical theory of communication [Journal] // The Bell system technical journal. -
629 1948. - 27(3). - pp. 379-423.
630 **Slater Lee** Near surface electrical characterization of hydraulic conductivity: From petrophysical
631 properties to aquifer geometries—A review [Journal] // Surveys in Geophysics. - 2007. - 28(2-3). - pp.
632 169-197.
633 **Soller D. R., Price S. D. and Kempton J. P. and Berg, R. C.** Three-Dimensional Geologic Maps of
634 Quaternary Sediments in East-Central Illinois. [Report]. - Urbana : United States Geological Survey
635 Geologic Investigations Series Map I-2669, 1999.
636 **Tizro Abduulah Taheri [et al.]** Hydrogeological framework and estimation of aquifer hydraulic
637 parameters using geoelectrical data: a case study from West Iran [Journal] // Hydrogeology. - 2010. - 18. -
638 pp. 917-929.
639 **Zaytsev A. and Burnaev E.** Large scale variable fidelity surrogate modeling [Journal] // Annals of
640 Mathematics and Artificial Intelligence-Springer. - 2017. - Vol. 81. - pp. 167–186.
641 **Zhang Jiangjiang [et al.]** Efficient Bayesian experimental design for contaminant source identification
642 [Journal] // Water Resources Research. - 2015. - 51(1). - pp. 576-598.
643 **Zhang Jiangjiang [et al.]** Inverse Modeling of Hydrologic Systems with Adaptive Multifidelity Markov
644 Chain Monte Carlo Simulations [Journal] // Water Res. Research. - 2018. - 7 : Vol. 54. - pp. 4867-4886.
645 **Zheng Qiang [et al.]** Adaptive Multifidelity Data Assimilation for Nonlinear Subsurface Flow Problems
646 [Journal] // Water Res. Research. - 2018. - 1 : Vol. 55. - pp. 203-217.
647