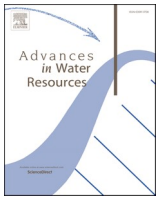
Advances in Water Resources 178 (2023) 104489

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres



Estimation of hydraulic conductivity in a watershed using sparse multi-source data via Gaussian process regression and Bayesian experimental design

Chien-Yung Tseng [a,b,c,*], Maryam Ghadiri [b,c], Praveen Kumar [c], Hadi Meidani [c]

[a] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States
[b] Illinois Water Resources Center, Prairie Research Institute, University of Illinois at Urbana-Champaign, Champaign, IL, United States
[c] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States

ARTICLE INFO

Keywords:
Bayesian experiment
Gaussian process
Multi-fidelity
Hydraulic conductivity
Hydro-geoinformatics
Optimization

ABSTRACT

Enhanced water management systems depend on accurate estimation of subsurface hydraulic properties. However, geologic formations can vary significantly, so information from a single source (e.g., widely spaced boreholes) is insufficient in characterizing subsurface aquifer properties. Therefore, multiple sources of information are needed to complement the hydrogeology understanding of a region. This study presents a numerical framework in which information from different measurement sources is combined to characterize the 3D random field in a multi-fidelity prediction model. Coupled with the model, a Bayesian experimental design was used to determine the best future sampling locations. The Upper Sangamon watershed in east-central Illinois was selected as the case study site, where the multi-fidelity Gaussian process model was used to estimate the hydraulic conductivity in the region of interest. Multi-source observation data were obtained from electrical resistivity and borehole pumping tests. The accuracy of the model prediction is dependent on the locations and the distribution of both high- and low-fidelity data. Furthermore, the multi-fidelity model was compared with the single-fidelity model. The uncertainties and confidence in the measurements and parameter estimates were quantified and used to design future cycles of data collection to further improve the confidence intervals.

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 given area of interest. Borehole cross-pumping test is a traditional, reliable method to directly measure subsurface hydro-geophysics properties such as hydraulic conductivity (Reinhart 2006; Hamm et al., 2007). There are several types of cross-pumping tests. Constant-rate pumping test is one of them, which directly measures the steady water flow underground by maintaining a constant hydraulic head gradient across the subsurface. However, although borehole testing can provide relatively high-confidence measurement results, drilling a borehole to obtain the information at just one specific location is expensive and time-consuming. On the other hand, electrical signal measurement such as earth electrical resistivity (EER) and electromagnetic induction have been broadly used in hydro-geophysics investigations (Lesmes and Friedman, 2005). These tests introduce electrical current into the subsurface and measure the resistivity via several receivers along the transect, which provides continuous hydro-geophysics information in the measured region without intruding into the ground. However, this type of measurement requires direct or inverse empirical relations between electrical and hydro-geophysical properties. Several studies have examined factors influencing relationships between electrical resistivity and hydraulic properties of aquifers (Kelly 1977; Mazáč et al., 1985; Niwas and Singhal 1985; Yadav 1995; Sikandar and Christen 2012). For example, Mazáč et al. (1990) studied the relationships between hydraulic conductivities and rock resistivities, and they examined the role of the distribution of hydraulic conductivity on dynamics of pollution spreading in rock medium. These relationships are sometimes case-specific and not universal for all locations. Also, soil saturation and temperature can affect these relationships as well (Khalil and Santos, 2009). Therefore, the estimation can cause tremendous errors if not used properly.

* Corresponding author.
E-mail address: tsengc@ornl.gov (C.-Y. Tseng).

https://doi.org/10.1016/j.advwatres.2023.104489
Received 22 September 2022; Received in revised form 8 June 2023; Accepted 13 June 2023
Available online 16 June 2023

Hydraulic conductivity of geologic formations can vary by orders of magnitude over relatively small spatial scales, so characterizing subsurface aquifer properties using only the information acquired from a widely spaced single measurement is challenging and potentially inaccurate. One method that has been widely employed is to use an integrated exploration approach in which borehole and other geophysical datasets are jointly interpreted (Lesmes and Friedman, 2005). Numerous studies have used information at different levels of fidelity to estimate hydraulic conductivity for groundwater system models (Asher et al., 2015; Zhang et al., 2018). These multi-fidelity (MF) models combine low-fidelity (LF) data with high-fidelity (HF) data to approximate the prediction with an accuracy that is better than that offered by a single-fidelity (SF) model (Peherstorfer et al., 2018; Fernández-Godino et al., 2016). The terms "HF data" and "LF data" refer to different levels of detail and accuracy of the data. HF data refers to data that is precise, accurate, and of high quality, with a high level of detail and granularity, while LF data, on the other hand, refers to data that is less accurate, less precise, and of lower quality, with a lower level of detail and granularity.

Forrester et al. (2007) first proposed a global optimization strategy, using MF surrogate models to include multiple levels of information into the predictions, which is called "MF kriging". In MF kriging models, data points obtained from sensors with different fidelity levels are fit with different surrogate models that provide estimation and prediction without the need to obtain a large number of expensive tests or run expensive numerical simulations. Among the surrogate models, the Gaussian Process (i.e., kriging) has been widely used in MF groundwater modeling (Zaytsev and Burnaev, 2017). Compared to the traditional co-kriging method that uses information from correlated variables to improve the accuracy of predictions, which involves modeling the correlation between two or more variables, MF kriging focuses on using data from multiple levels of fidelity to improve the accuracy of spatial predictions by incorporating information from multiple sources. The MF models facilitate the usage of data with different levels of fidelity by combining a HF function (a more accurate but expensive representation of a physical phenomenon) with a LF function (a less accurate but inexpensive representation of a physical phenomenon). Asher et al. (2015) and Fernández-Godino et al. (2016) extensively surveyed several data-driven methods of combining fidelities with a primary focus on kriging models for MF applications. This focus was particularly because such a Gaussian process entails an uncertainty structure that readily lends itself to an MF modeling approach (Fernández-Godino et al., 2016). A more recent study by Zheng et al. (2018) employed MF Gaussian surrogates 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.

However, numerous questions remain unanswered in predicting hydraulic properties of subsurface formations in watersheds. For example, how many field tests need to be conducted to achieve the desired accuracy of the estimation, and what if the existing data points are sparse? Many studies have confirmed the usage of MF kriging models for predicting the hydro-geophysical information in a specific region with abundant HF and LF data, but few studies have focused on discussing the effects of HF and LF data location distribution, especially in a data-sparse situations. Another subject that has not been fully addressed is to understand how data fidelity associated with different tests would affect future test locations. Recently, Menberg et al. (2020) used an MF approach along with Bayesian parameter estimation in subsurface heat and fluid transport models to include information from a more physically accurate but expensive HF model, as well as a large number of evaluations from a less accurate, less expensive LF model. The study demonstrated that the combined information from sources with different fidelities substantially improved the posterior distribution results, which may be important for determining future test locations to optimize the information gained from newly implemented data sources.

In this study, we present a quantitative MF framework to combine information from EER measurements and pumping (borehole) measurements with different fidelities and accuracies to enhance the understanding of hydro-geophysical characteristics. For the first time, we specifically address the effect of data locations from data with different levels of fidelity under sparse data distributions. Also, we investigated how future test locations with different fidelities should be conducted to optimally enhance our understanding of the geo-hydraulic properties of a region according to the new information gained. As a sample case study, we selected an intensively managed area located in the Upper Sangamon watershed in east-central Illinois as the study site. Traditional high-accuracy pumping test data were used with small-scale EER measurements to generate 2D maps of hydraulic conductivity over a large-scale region with quantified uncertainties in different depth layers. According to the MF kriging framework, the discussions focused on how the distribution of data with different fidelities would affect the model accuracy, especially with sparse data points. We further discussed how the MF model can learn from new sensors using probabilistic statistical tools to select the best locations for future data collection. The approach was based on the Bayesian experimental design, which selects the best locations from a set of candidate locations according to the value of information that each location is expected to offer (Norberg and Rosn, 2006). By relating the expected value of information from each location to the present levels of uncertainties in the MF kriging model, we thus can pick the best location with the most information gain. 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 theoretical background, which includes detailed information about site selection, observation data, lognormal ordinary kriging (LOK), MF LOK, 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 kriging. In Section 4, we discuss the effect of fidelity on the estimated field and the estimation accuracy followed by the distribution of LF and HF data points. The application of optimal Bayesian experimental design for obtaining optimal future sampling locations is also presented. Finally, Section 5 provides discussions and conclusions.

2. Method

2.1. Study site

The Sangamon River is a major tributary to the Illinois River with the confluence near Chandlerville in Cass County, Illinois. The watershed spreads 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. It 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 comprises nearly 90% of the land area (Keefer et al., 2005). As an intensively managed landscape, this region is at risk for deterioration of land and water systems. Therefore, more observations are needed to understand and predict the behavior of natural services (ecological, hydrological, and climatic services) that support basic human needs such as water, food, and energy.

2.2. Data description

Based on the geological properties of subsurface formation, hydraulic conductivity is spatially correlated but sometimes can vary significantly. Hence information from a single source is insufficient, indicating the need of multiple sources of information to complement the hydrogeology understanding of a region. In this study, two types of

field observation data, EER measurement and pumping tests, were used as data sources with different fidelities to estimate hydraulic conductivity of the Upper Sangamon watershed using a MF Gaussian Process model (more details in Section 2.4). Fig. 1 shows the data locations.

The pumping test involved pumping from a test well at a controlled rate and monitoring the flow rate through the drawdown at different locations along the radial axis from the test well. Hydraulic conductivity values of aquifer material as determined from pump tests and aquifer tests varies spatially but not as much temporally. Repeated pump tests may show changes in hydraulic conductivity at the well skin of production wells (or within gravel packs immediately surrounding well screens) but repeated long-duration aquifer tests would provide similar values over time. The measured hydraulic conductivity values of aquifer material can provide convincing values. Therefore, pumping test measurements have been well recognized as one of the most reliable ways to measure soil hydraulic conductivity with high-fidelity, meaning that the data is accurate and precise with high quality and less errors or discrepancies, which provides a true and faithful representation of the true value in the real field. However, because of 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 2D resistivity model of the relationships between aquifer hydraulic and electrical properties (Kelly and Frohlich 1985; Slater 2007; Khalil and Santos 2009; Tizro et al., 2010). The measurement relies on testing with a dipole–dipole electrode configuration in a vertical 2D plane of the field, which is commonly used in 2D electrical resistivity surveys because it provides good resolution of subsurface structures with high sensitivity to lateral changes in resistivity. In the array, current is injected into the ground using two current electrodes, and voltage measurements are made using two pairs of potential electrodes placed at a distance from the current electrodes. The distance between the current and potential electrodes is increased systematically to obtain measurements at various depths. The dipole-dipole array is particularly useful for detecting planar variations vertically and horizontally in subsurface resistivity to infer the hydraulic conductivity. However, the accuracy depends on the equipment precision. Hence, the measurement has lower cost but also lower fidelity due to its lower accuracy compared with traditional pumping (borehole) testing. In general, EER values are known to vary. There are many factors that can affect the measured resistivity, like soil texture, clayey matters, pore structure, and temperature. The greatest variability generally comes from the changes in saturation conditions and temperature (Khalil and Santos, 2009). Within the Upper Sangamon watershed, the water table is within a few meters of the ground surface except in areas of very steep slopes and very coarse materials. These areas are a small percentage of the entire area and thus can be neglected at the scale of this study. There is some variation from temperature, but we neglected this effect because it is small within the context of this generalized study.

EER measurement provides a continuous estimation of hydraulic conductivity in a small vertical plane (~800 m long and ~80 m deep). To apply the EER data together with the pumping test data in the MF Gaussian Process model, we need to ensure that two data sources provide the same physical meanings. Therefore, EER has to be converted to hydraulic conductivity since pumping tests directly provide the hydraulic conductivity information. A study by Lu et al. (2019) demonstrated that the relationship between the 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 testing. According to their experiment results with sands, when K is a dependent variable, the given best fitted empirical parameters can be obtained by fitting to a comprehensive data set:

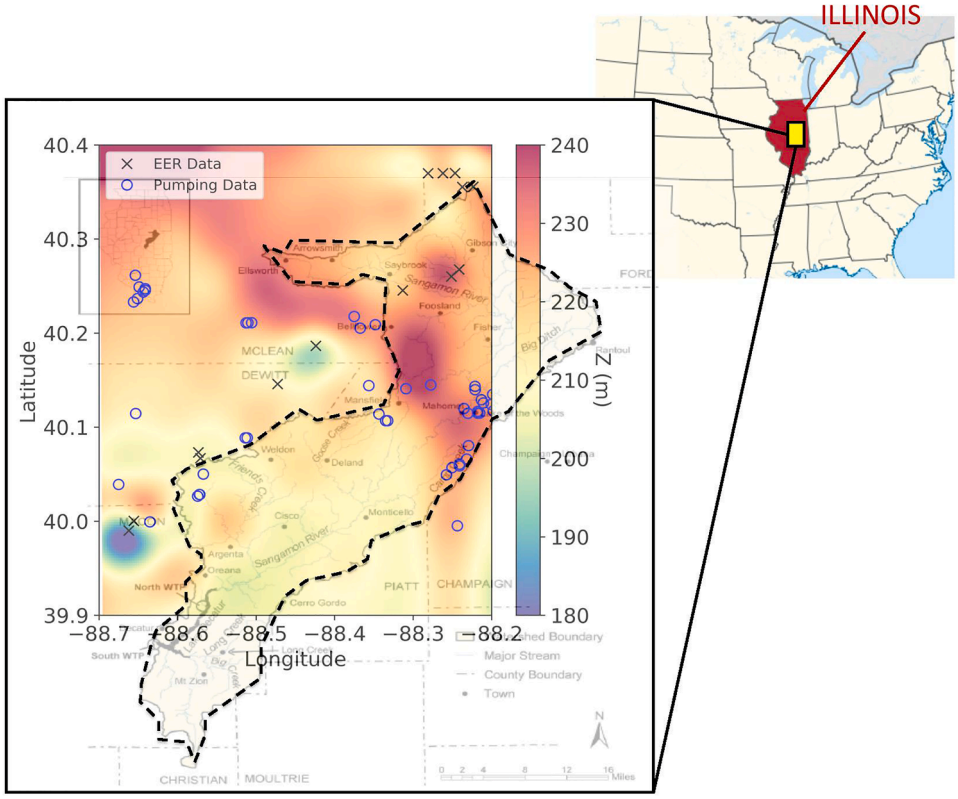


Fig. 1. Locations for data in the Upper Sangamon watershed. The Black dashed line represents the Upper Sangamon watershed. Blue circle markers represent the pumping test data locations. Black cross markers represent the EER data locations.

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

As mentioned previously, many factors can affect the empirical relationship between the measured resistivity and K, especially the soil pore size. Within the Upper Sangamon watershed, the dominant soil types found in this area are Mollisols and Alfisols whose pore sizes can range from 0.1 to 1.0 mm, which is similar to the pore size of the typical sands (0.2 to 2.0 mm). Therefore, we assumed that Eq. (1) by Lu et al. (2019) is valid within the context of this generalized study. In this study, as the soil deposition was accumulated layer by layer, resulting in a consistent geological composition horizontally, after converting the measured resistivity, σ_{EC}, into hydraulic conductivity, K, the horizontal mean value of K was set as the representative value in each depth for the LF data input in the MF kriging model.

Borehole pumping tests were conducted at specific locations with different depths. Unlike EER measurement, which provides continuous vertical information, pumping tests provide point information on subsurface properties from the measurement of flow velocities within soil pores. They offer higher accuracy on hydraulic conductivity, which was set as the HF data source in MF kriging. We obtained EER data from 15 locations with continuous depth and pumping test data from 68 locations with specific depths for each test. Compared to the entire horizontal study domain ($60km \times 50km$) with the relatively sparse distributions of the data points, both EER and pumping test can be viewed as point measurements in the model under nearly the same measurement scale in each layer. The EER and pumping tests were conducted by the 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.

2.3. Lognormal ordinary kriging

Lognormal Ordinary Kriging (LOK) is a commonly used 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 and Dengiz, 2018). Compared to the traditional ordinary kriging model, the LOK model can improve the calculation of statistics and weighted averages to avoid negative and extreme estimated values, which helps to reduce the impact of outliers on the estimated values (Roth, 1998). Also, the LOK model can be more appropriate when the variable being studied exhibits a positive skewed distribution, providing more accurate and realistic estimates, since it considers the asymmetry of the distribution. As shown in Fig. 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 symmetrical without any extreme data.

The LOK model algorithm follows the structure of Gaussian processes:

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

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 $\boldsymbol{x}$, $\overline{\mu}$ is the mean value of $ln(\boldsymbol{y})$ within the simulation domain, and $\boldsymbol{K} = \{K_{ij}\}$ is a symmetric matrix, which is constructed by the kriging function $k(x_i, x_j; \theta)$ with exponential variogram through the following equation:

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

where $\theta = (n, s, r)$ are the kriging parameters, namely nugget (n), sill (s), and range (r). The nugget (n) is related to the amount of a short range of the initial randomness or noise in the data. The range (r) represents the distance at which data are no longer correlated and the semivariance first flattens out and reaches the sill (s), the total variance where the empirical variogram appears to level off. The kriging parameters can be obtained by fitting the sample variogram to the semivariance, γ, with the given the observation data $\{\boldsymbol{x}, \boldsymbol{y}\}$, which can be expressed as

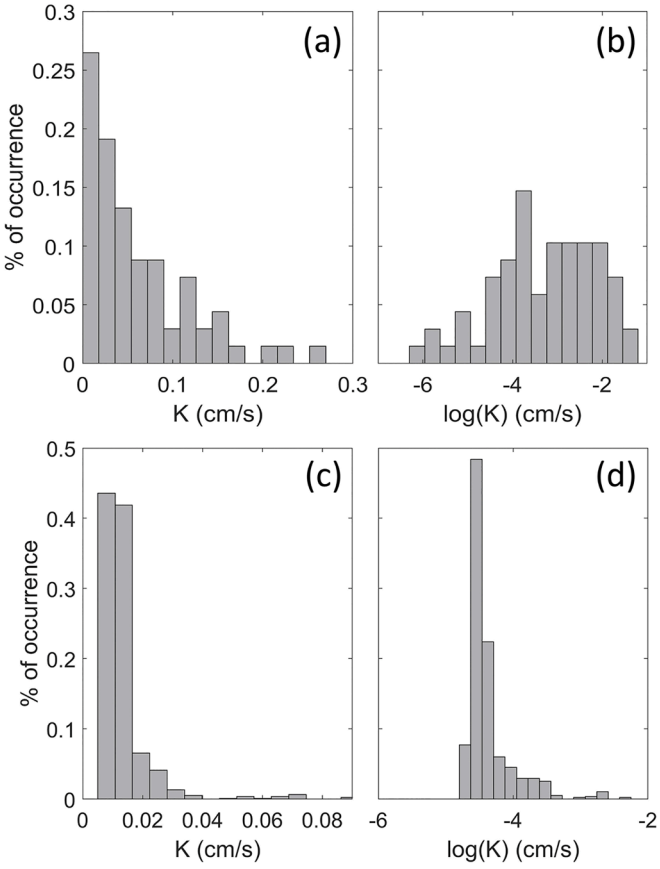


Fig. 2. Histogram of the measured hydraulic conductivities from the pumping tests under (a) normal scale and (b) log scale; and from the EER tests under (c) normal scale and (d) log scale.

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

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:

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

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

$$f(\boldsymbol{x}^*|\boldsymbol{x}) \sim N(\boldsymbol{\mu}_l, \ \boldsymbol{\sigma}_l), \quad (6)$$

where

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

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

Since $f(\boldsymbol{x}^*)$ is in logarithmic scale, to estimate the parameter of interest (in this case, the hydraulic conductivity), we converted the logarithmic values, $\boldsymbol{\mu}_l$ and $\boldsymbol{\sigma}_l$, back to the actual mean and standard deviation values according to

$$\boldsymbol{\mu}^* = exp\left(\boldsymbol{\mu}_l + \frac{\boldsymbol{\sigma}_l^2}{2}\right), \tag{9}$$

$$\boldsymbol{\sigma}^* = \sqrt{[exp(\boldsymbol{\sigma}_l^2) - 1]exp(2\boldsymbol{\mu}_l + \boldsymbol{\sigma}_l^2)}. \tag{10}$$

2.4. Multi-fidelity lognormal ordinary kriging

To combine the observation data from EER measurements and pumping testing, the MF LOK model was used to perform 2D hydraulic conductivity mapping in different depth layers with smooth and continuous fusion of information from two sources with different levels of fidelity/precision. The MF kriging algorithm follows the structure proposed by Kennedy and O'Hagan (2000) and Forrester et al. (2007), assuming that

$$u_L(\boldsymbol{x}) \sim GP(\overline{\mu}, k_L(\boldsymbol{x}, \boldsymbol{x}; \theta_L)), \tag{11}$$

$$u_H(\boldsymbol{x}) \sim GP(\overline{\mu}, k_H(\boldsymbol{x}, \boldsymbol{x}; \theta_H)) \tag{12}$$

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:

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

where

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

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

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

k_L and k_H are the kriging functions (Eq. (3)) for the LF and HF data, respectively, and ρ is the MF constant, which was first proposed by Forrester et al. (2007) as a scaling factor to approximate the data with a LF contribution to the prediction. Following the auto-regressive model (Kennedy and O'Hagan, 2000), the idea is to approximate the high-fidelity function, $f_H(\boldsymbol{x})$, as the low-fidelity Gaussian Process surrogate, $u_L(\boldsymbol{x})$ multiplied by a scaling factor ρ plus a high-fidelity Gaussian Process surrogate $u_H(\boldsymbol{x})$ that represents the difference between $\rho u_L(\boldsymbol{x})$ and $u_H(\boldsymbol{x})$ to consider both contributions from the low-fidelity and high-fidelity data.

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:

$$\mathrm{f}_{mgp}(z) \sim N(\overline{\mu}, \boldsymbol{K}), \tag{17}$$

where

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

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

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

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

where N is the total number of the data points. In this study, we used 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 constructed the joint distribution:

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

where

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

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

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

where

$$\boldsymbol{\mu_m} = \overline{\mu} + \boldsymbol{q}^T\boldsymbol{K}^{-1}(\boldsymbol{y} - \overline{\mu}), \tag{24}$$

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

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

$$\boldsymbol{\mu}^* = exp\left(\boldsymbol{\mu}_m + \frac{\boldsymbol{\sigma}_m^2}{2}\right), \tag{26}$$

$$\boldsymbol{\sigma}^* = \sqrt{\left[exp\left(\boldsymbol{\sigma}_m^2\right) - 1\right]exp\left(2\boldsymbol{\mu}_m + \boldsymbol{\sigma}_m^2\right)}. \tag{27}$$

2.5. Optimal bayesian experimental design

Our experimental design addresses the challenge 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 and Rosn, 2006). For instance, in the context of hydraulic property estimation for aquifers, measurements collected from locations that are closely spaced will provide much less information compared with those obtained from locations that are farther 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 as the Kullback–Leibler divergence from the posterior distributions of the model parameter to the prior (Chaloner and Verdinelli, 1995). The best experiment among the ensemble of candidates is the one that maximizes the information gain, taken to be the Kullback–Leibler divergence from posterior to prior. Solving this optimization problem is numerically complicated because the evaluation of Kullback–Leibler divergence requires samples from the prior and posterior of the parameters. Here, we provide the technical background for this experimental design approach combined with the MF kriging model.

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

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

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

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

In this study, $\boldsymbol{\theta}$ is the sampled kriging parameters, including n, s, and r. n and s are constant values according to the MF model, and r represents the Gaussian distributed samples based on the fitted LF and HF range, r_L and r_H, with $\sigma_L = 0.01r_L$ and $\sigma_H = 0.01r_L$, which were selected to create data samples that are closely clustered around the mean range values r_L and r_H. More specifically, we aim to make the data samples to be very precise and consistent with little variations, so that the modeled variations of the prior distribution mainly come from $\boldsymbol{d}$, which 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

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

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

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

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

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

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

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

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

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

where the likelihood function, $p(d_i|\theta_j,s)$, uses the Gaussian radial basis likelihood function that consists of a exponential decaying function with the MF kriging model G:

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

Combining Eqs. (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 $U(s)$

$$s^* = \underset{s \in D}{argmax}[U(s)] = \underset{s \in D}{argmin}[-U(s)]. \tag{36}$$

To avoid being trapped by the local minimum points during optimization process, we applied a traditional brutal approach by setting the candidate locations for s every 1 km in x- and y- directions and then selected the one who had the minimum utility value $U(s)$ as s^*. The results of the sequential Bayesian experimental design application for future sampling locations are demonstrated in Section 4.3.

3. Results

3.1. Topography investigation

We used Lidar data from the US Geological Survey National Elevation Dataset for the Upper Sangamon watershed, along with the EER and pumping test data. The Lidar data are uniformly distributed in the rectangular region of the Upper Sangamon watershed, as shown in Fig. 1. Multi-quadratic radial basis function with Euclidean distance was used to interpolate the elevation between the Lidar data points. Fig. 3(a) shows that the topography of the watershed is generally flat, which is on average within a range of approximately 210 to 230 m. There is only a relatively low region in the southeastern region (~180 m). The flat topography suggests that a reasonable approach would be to represent the domain in a Cartesian coordinate system (x-y-z), denoting x-coordinate along the latitude, y-coordinate along the longitude, and the z-coordinate by depth (distance from the surface) to ignore the surface variation and set all the locations' surface as zero in depth for the z-value.

3.2. Single-fidelity results with multiple data sources

The Upper Sangamon watershed is in 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. Therefore, soil deposition patterns are expected to have a layer-by-layer distribution. The watershed contains mostly sand and gravel deposits concentrated in different layers, which are typically 10 to 15 m thick (Selkregg and Kempton, 1958). With an additional sensitivity analysis between the MF kriging results in five layers (Δz = 15 m) and the results in eight layers (Δz = 10 m) at similar depth, the comparison reconfirms that five layers with 15 m thickness is good enough to show the general geological property, such as hydraulic conductivity, in different layers in the watershed. Therefore, 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 and 85 m deep from the surface, as shown in Fig. 3(b). Within the same layer, soil and hydraulic properties (e.g., hydraulic conductivity) are similar and correlate across different locations. We constructed a 2D (horizontal) kriging model in different layers to construct a multilayer mapping of hydraulic conductivity.

SF kriging with multiple data sources was conducted as the reference to compare with the MF kriging model. In the SF kriging model, the data

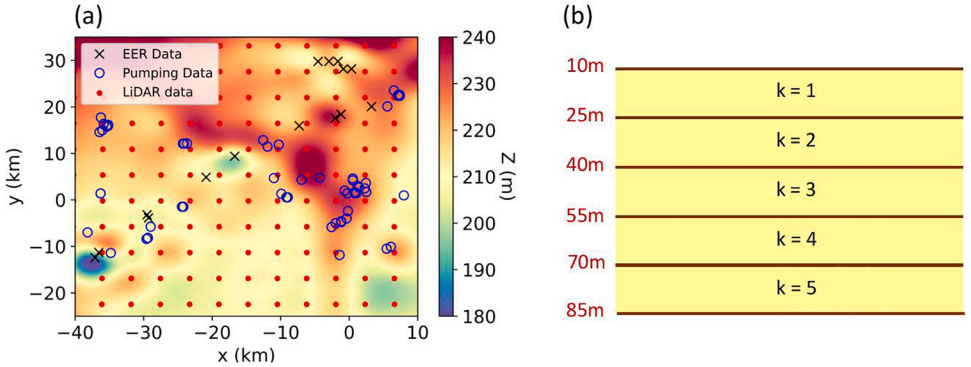


Fig. 3. (a) The surface elevation map of the rectangular region of the Upper Sangamon watershed shown in Fig. 1. Red dots represent the Lidar data. Blue circle markers represent the locations of EER data. Black cross markers represent the locations of pumping test data. Champaign City, Illinois (40°06′54″N, 88°16′22″W) was set as the origin point (x = 0 km, y = 0 km). (b) Sketch of the vertical layers setup, where k is the layer number.

sources were treated equally, ignoring their different fidelities. The exponential model is one of the most commonly used models to characterize cross spatial dependence in geological property data, which suggests that data spatial autocorrelation decreases exponentially with increasing distance based on prior knowledge of the phenomenon (Oliver and Webster, 2015). We thus used the exponential function-based variogram to fit the semivariance of the data (Eq. (4)), including EER and pumping test data on the sample variogram (Oliver and Webster, 1990). A Python-based fitting tool, using a nonlinear least squares algorithm, was applied for curve fitting. The fitted kriging parameters of $n = 0.02$, $s = 0.79$, and $r = 11.39$ were used.

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

In Fig. 4, the SF kriging result shows a relatively uniform distribution of mapped hydraulic conductivity, K, in the upper three layers (depth

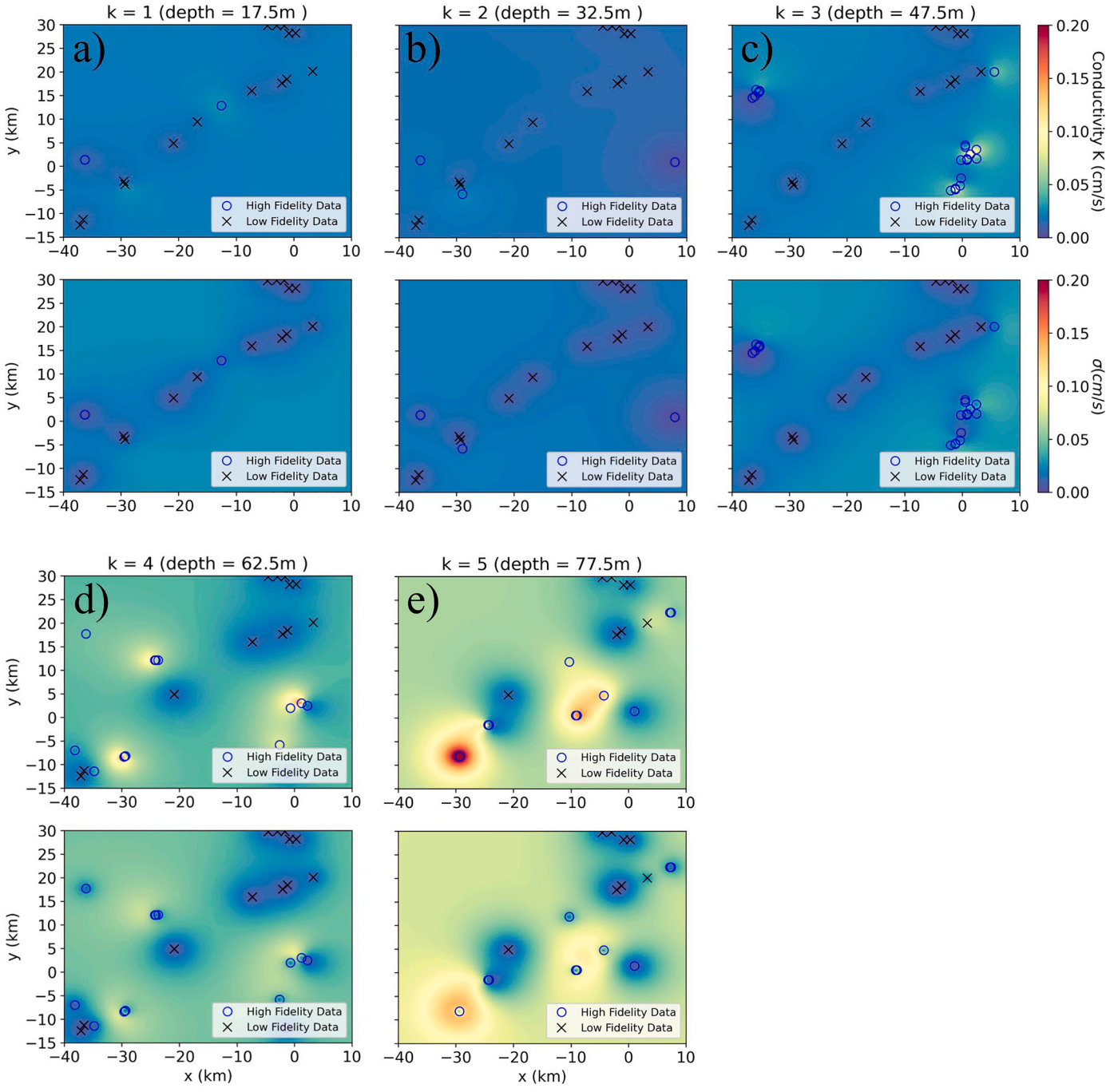


Fig. 4. SF kriging of the hydraulic conductivity and corresponding standard deviation in the Upper Sangamon watershed in different depth layers. (a) Layer $k = 1$, depth = 17.5 m. (b) Layer $k = 2$, depth = 32.5 m. (c) Layer $k = 3$, depth = 47.5 m. (d) Layer $k = 4$, depth = 62.5 m. (e) Layer $k = 5$,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

<50 m), and some peak values can be observed in the lower two layers (depth >50 m). This result suggests that more varying soil properties exist in the deeper layers of the watershed. The uncertainty in the estimated properties is presented by the standard deviations, σ. However, we can see that some regions with high standard deviation around the HF data point shown in Fig. 4d and 4e. It is because when converting the standard deviation, σ^*, from the lognormal scale back to normal scale, the magnitude of σ^* is also related to the magnitude of the mean value, μ_l (Eq. (10)). At the region where the estimated mean magnitude of K is high will result in a relatively high deviation. However, if we look closer to the lower-left point (inside the blue circle) in Fig. 4e, the deviation around the data point remains low (blue to yellow) due to the unbiased estimate from the model near the data point, but quickly becomes higher (orange) due to the high mean K region (red region in the upper panels).

3.3. Multi-fidelity results

The SF kriging model did not account for the fact that data are from different sources, so they were considered with the same uncertainty. However, since different data sources typically have different uncertainty/error ranges depending on equipment, methods, and human factors, the fidelity of these data sources should also be incorporated

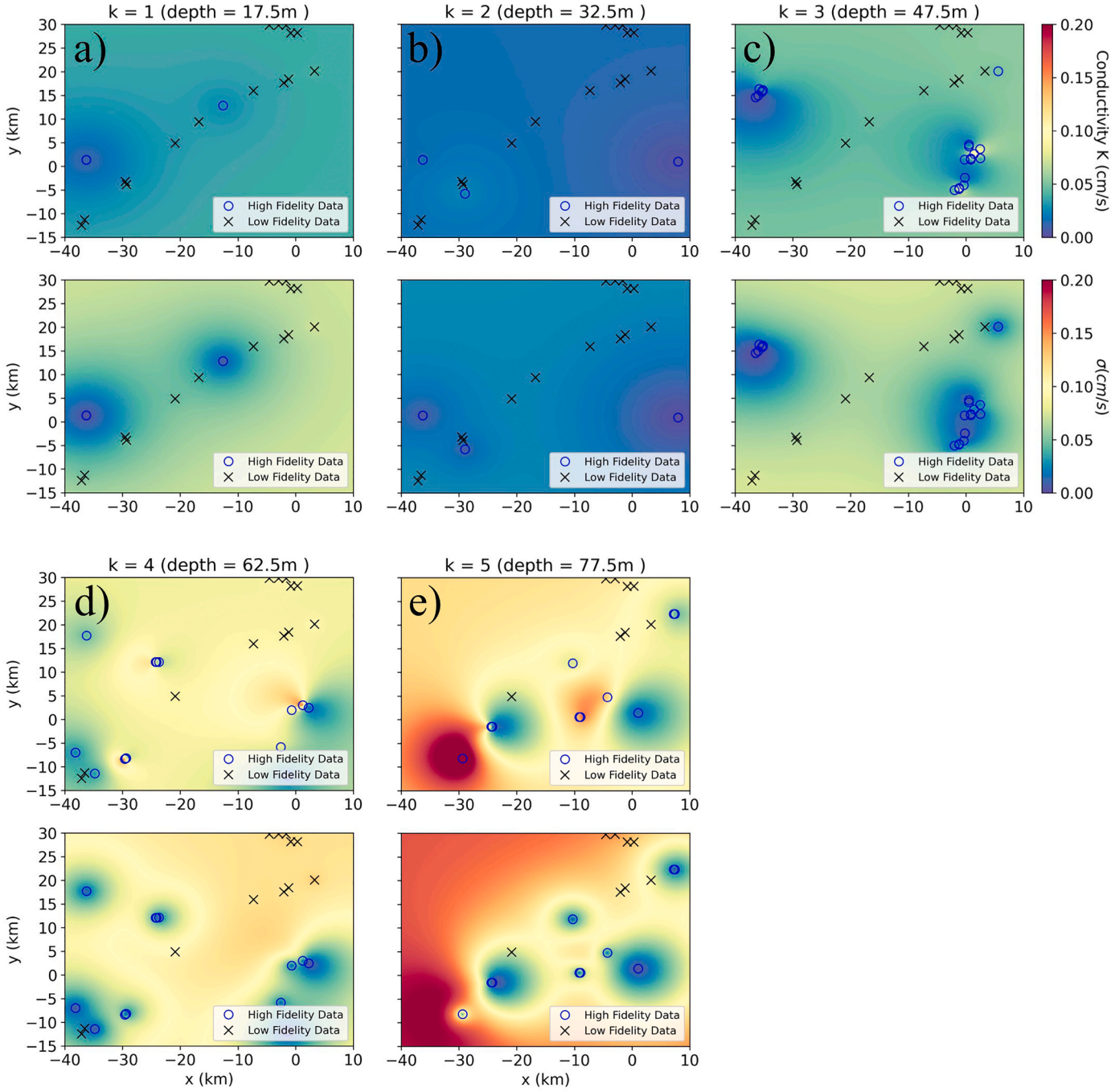


Fig. 5. MF kriging of the hydraulic conductivity and corresponding standard deviation in the Upper Sangamon watershed in different depth layers. (a) Layer $k = 1$, depth = 17.5 m. (b) Layer $k = 2$, depth = 32.5 m. (c) Layer $k = 3$, depth = 47.5 m. (d) Layer $k = 4$, depth = 62.5 m. (e) Layer $k = 5$, depth = 77.5 m. The value of depth shown on top of each panel is the center z-location in each layer (Fig. 3(b)). Blue circles represent the HF data locations. Black cross markers represent the LF data locations.

into the model. Incorporation of these factors makes MF kriging models a more accurate approach than SF kriging models when multiple sources of data are available. Thus, not much confidence can be placed in the SF kriging result shown in Fig. 4. Data with different levels of fidelity should also be treated separately in the sample variogram for two sets of fitted kriging parameters.

In the MF kriging model, we treated EER and pumping test data separately where kriging parameters were obtained from each sample variogram. The fitted kriging parameters based on the exponential function-based variogram are $n = 0.06$, $s = 0.08$, and $r = 2.91$ for EER measurement; and $n = 0.49$, $s = 0.88$, and $r = 21.12$ for pumping tests. According to the fitted kriging parameters, the range (r) of the High-fidelity pumping test dataset, $r = 21.12$, which is much higher than the range (r) of the Low-fidelity EER dataset, $r = 2.91$, indicating that EER data has weaker spatial correlation due to its low fidelity of the data accuracy and quality.

Fig. 5 shows the MF kriging result of the hydraulic conductivity and the corresponding standard deviation in the Upper Sangamon watershed. Compared with the SF kriging results shown in Fig. 4, MF kriging puts more weight on the HF data (shown by circle markers). Thus, the estimated K and σ distribution patterns generally follow the distribution of the pumping test data. Furthermore, regions near the HF data points (blue circles in Fig. 5) have lower standard deviation. This means that the model assesses higher confidence in the estimates in those regions.

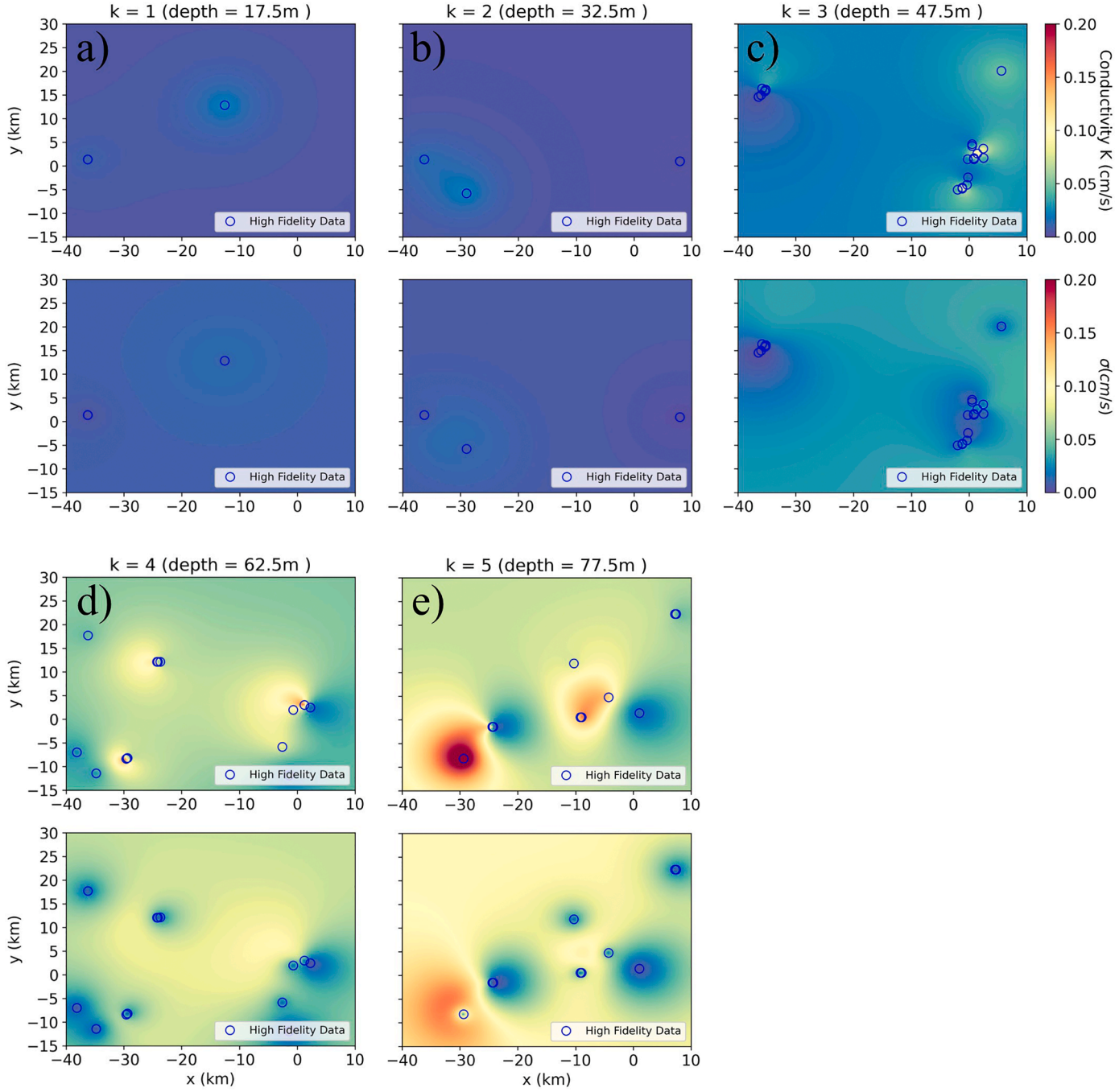


Fig. 6. SHF kriging of the hydraulic conductivity and the corresponding standard deviation in the Upper Sangamon watershed with only HF data (pumping test data) in different depth layers. (a) Layer $k = 1$, depth = 17.5 m. (b) Layer $k = 2$, depth = 32.5 m. (c) Layer $k = 3$, depth = 47.5 m. (d) Layer $k = 4$, 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 (Fig. 3(b)). Blue circles represent the locations of HF data.

The LF data (shown by cross markers), however, do not help reduce uncertainty levels in a large area but nevertheless provide local hydraulic information in regions far from the HF data locations.

4. Discussion

4.1. Fidelity effect on the predicted field

In MF kriging, pumping test data were selected as the HF data source because they offer a more reliable measurement method than EER data. Hence, the estimated K and σ values based on MF kriging (Fig. 5) are mostly dominated by the HF data (pumping test data). To further study the fidelity effect, we excluded the LF data and only considered the HF data in the kriging model (Fig. 6) to compare with the MF results (Fig. 5). Fig. 6 shows that in regions near the HF data, the estimated K and σ values are similar to those in Fig. 5. However, in regions far from the HF data points, the models provide very different K and σ estimates, especially in the upper three layers (depth <50 m) where HF data points are scarce. The higher σ estimations are because of the additional information provided by the LF data. However, the higher estimated σ does not suggest that the LF data provide incorrect information; instead, the different estimations of K suggest that the LF data do provide valuable information about the hydraulic conductivity properties for regions where expensive HF tests are not available or economically not feasible.

4.2. Fidelity effect on the estimation accuracy

To evaluate the estimated K values in the MF kriging model, we focused on the lower two layers (depth >50 m) and removed HF data points in each layer from the estimation model. Since HF data was measured directly by the pumping test with high accuracy, we assumed its measured K is the true value and used the removed HF data points as the reference to compare with the model estimated values. In Fig. 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 an LF data location to assess the accuracy of LF data contributions. The calculated standard deviation values did not differ significantly between the MF model and the SHF model. More HF data can increase the confidence levels in both cases. However, a difference in estimated K values was not observed in this comparison.

The removed HF data points provide a reference value of $K = 0.078$ cm/s (Fig. 7a) and 0.026 cm/s (Fig. 7d) in the fourth layer and $K = 0.081$ cm/s (Fig. 7g) in the fifth layer at the data locations. After the data points were removed, the MF model provided a prediction of $K = 0.081$ cm/s (Fig. 7b) and 0.033 cm/s (Fig. 7e) in the fourth layer and $K = 0.107$ cm/s (Fig. 7h) in the fifth layer at the data locations. The SHF model provided a prediction of $K = 0.052$ cm/s (Fig. 7c) and 0.032 cm/s (Fig. 7f) in the fourth layer and $K = 0.086$ cm/s (Fig. 7i) in the fifth layer at the data locations. These results provide an estimated accuracy. To

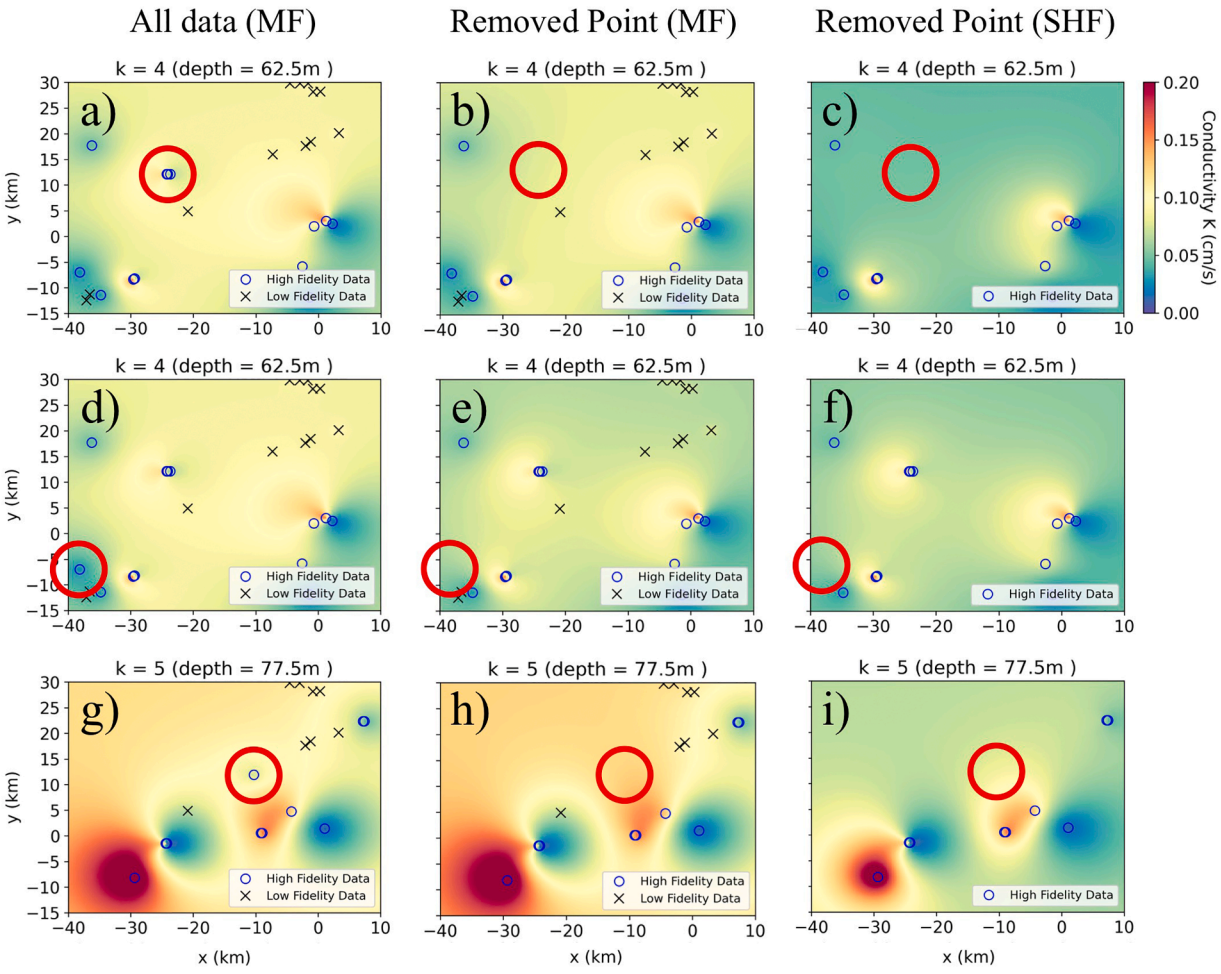


Fig. 7. Comparisons between MF kriging and SHF kriging with specific points removed. a), d), g): MF kriging of the hydraulic with all data points in the last two layers. b), e), h): MF kriging of the hydraulic with specific points removed in the lower two layers. c), f), i) HF kriging of the hydraulic conductivity with specific points removed in the lower two layers. Blue circle markers represent the HF data locations. Black cross markers represent the LF data locations. Red circles highlight the removed HF data points.

obtain a more precise measure of accuracy, we defined a_k as

$$a_k = 1 - \left(K_{pred} - K_{ref}\right)/K_{ref} \quad (37)$$

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% (Fig. 7b) and 73% (Fig. 7e) in the fourth layer and 68% (Fig. 7h) in the fifth layer, and for the SHF model, the accuracy was 67% (Fig. 7c) and 77% (Fig. 7f) in the fourth layer and 94% (Fig. 7i) in the fifth layer. The accuracy shows that when the removed points were far from the other HF data points, LF data provided important information to enhance the estimation of the kriging model (Fig. 7a, b, c). When the removed points were relatively close to the other HF data points, LF data were not necessary, and the predictions are dominated by the information provided from the surrounding HF data points with nearly the same accuracy in MF and SHF models (Figure d, e, f). When the removed points were in similar distances from the surrounding HF and LF data points, LF data might even provide information with higher variance due to its relatively low accuracy compared with the HF information (Fig. 7g, h, i). Thus, the accuracy of the MF model depends on the locations and the distribution of both the LF and HF data. When HF data points are scarce and far from the LF data points, the information provided from LF data becomes important and can enhance model performance and accuracy.

To further investigate the fidelity effect, we choose the deepest (fifth) layer as the test case and consecutively remove HF data points one by one within or close to the LF data points. Fig. 8 shows the estimated hydraulic conductivity field by MF kriging and SHF kriging under four-point removal scenarios: keep all the HF data; remove one point; remove two points; and remove three points. Comparing the results of MF kriging versus SHF kriging, as more data points were removed, SHF kriging showed a relatively lower and more uniform estimated K field. However, since LF data still provided the surrounding information in MF kriging, they provided estimated K values closer to the original estimates where all the data points were present.

According to the definition in Eq. (37), Fig. 9 shows the accuracy of K estimates at locations pt-1, pt-2, and pt-3 under MF kriging and SHF kriging when no data points, one data point, two data points, and three data points were removed. When all data points were present, the accuracy was 100% at all three locations. When one point was removed (at location pt-1), the accuracy remained 100% at locations pt-2 and pt-3, but at location pt-1, SHF kriging shows a greater accuracy compared with MF kriging because of the far distance between the removed point and the LF data points, as discussed in Fig. 7. When data at two points (locations pt-1 and pt-2) were removed, MF kriging began to show higher accuracy at locations pt-1 and pt-2, and location pt-3 remained at 100% accuracy. When all three data points were removed, MF kriging showed obviously higher accuracy at all the locations compared with SHF kriging. The results again confirm that when HF data become scarcer, the information provided by LF data becomes more critical in MF kriging and can lead to better estimation of hydraulic conductivity.

4.3. Future data collection using bayesian experimental design

We applied the Bayesian experimental design along with the MF kriging model to determine the future sampling locations for the HF data (pumping test) measurement. We chose the deepest (fifth) 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 (Fig. 1). Once the current optimal point was obtained, the hydraulic conductivity value was then predicted by the MF kriging model at that location. The current estimated optimal point with its predicted hydraulic conductivity was then put back in the MF kriging model as one of the synthetic measurement data points to update the model and train the new optimized constant ρ for the next optimal sampling location.

The final optimal result is shown in Fig. 10. The optimal locations are denoted by the red triangles with the numbers indicating the sequential order. The sampling points were located in the region where σ was high, indicating the need of future measurements to enhance the confidence of the prediction and understanding of the region of interest. The future sampling points provided more information to the region near the suggested locations, where variances were greatly reduced. Variances were slightly increased in regions far from the suggested locations because of the unbalanced information entered into the model. However, according to Bayesian experimental design, those regions were relatively less efficient for future measurements compared with 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

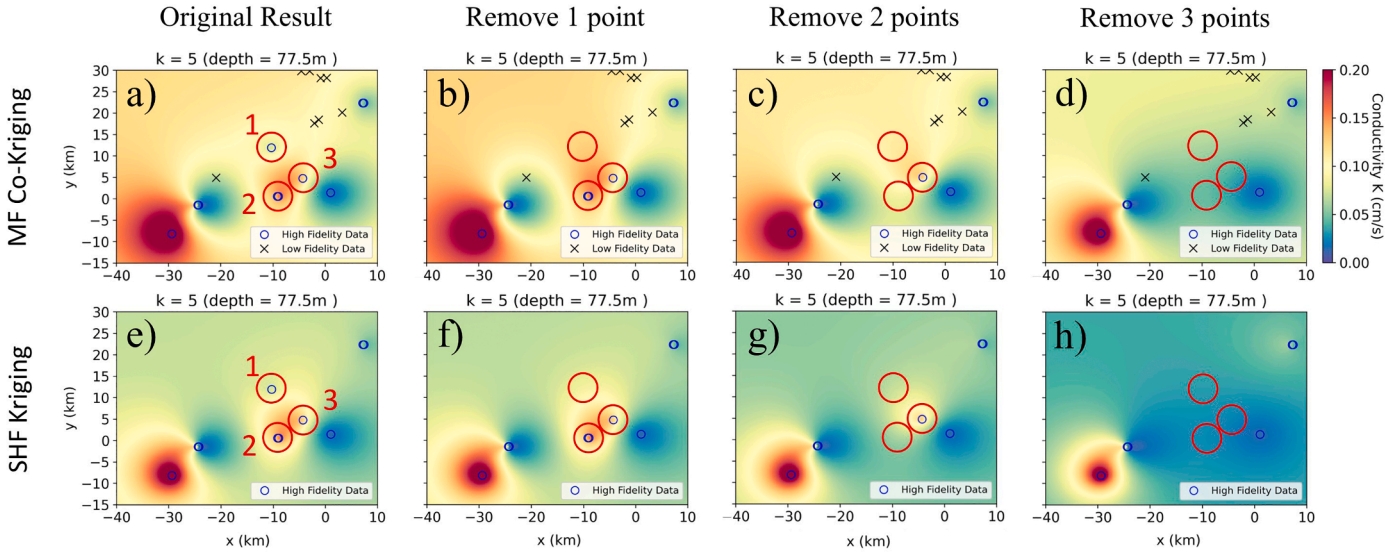


Fig. 8. Comparisons between MF kriging and SHF kriging in the deepest layer (depth >70 m) with three consecutive points removal. MF kriging of the hydraulic conductivity with (a) all data points, (b) one point removed, (c) two points removed, and (d) three points removed. SHF kriging of the hydraulic conductivity with (e) all data points, (f) one point removed, (g) two points removed, and (h) three points removed. Blue circles represent the HF data locations. Black cross markers represent the LF data locations. Red circles highlight the removed HF data points, and the nearby red numbers in (a) and (e) show the removal order of the points.

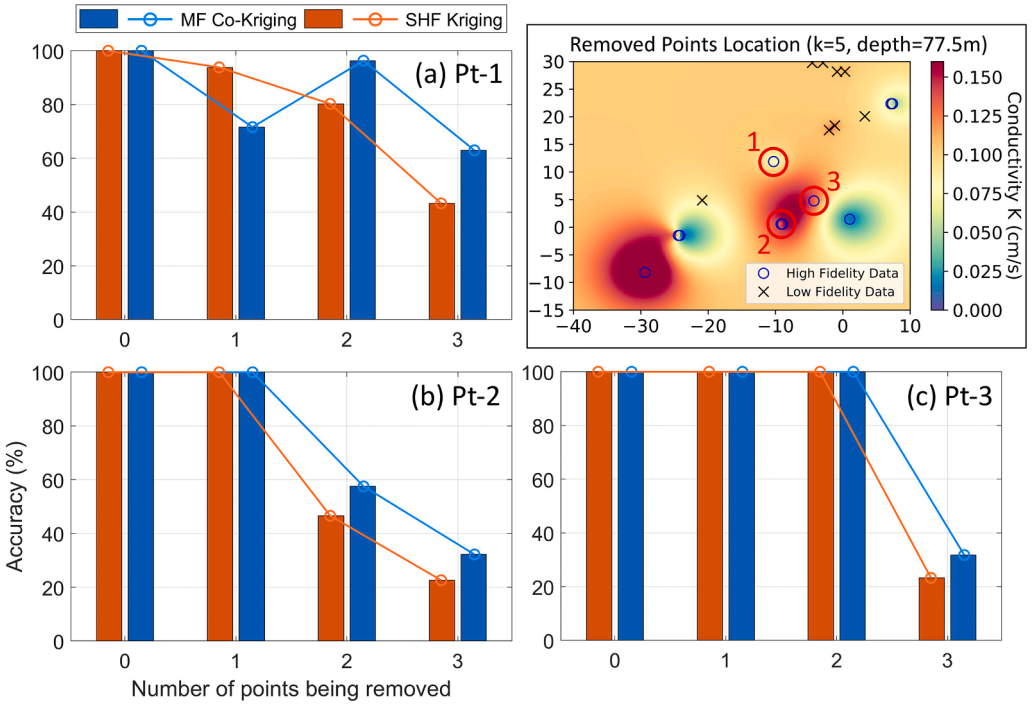


Fig. 9. The accuracy of (a) location 1, (b) location 2, and (c) location 3 under MF kriging and SHF kriging when removing no data points, one data point, two data points, and three data points. The removed points' locations are shown in the top-right panel, and the removal order follows the denoted number of the points.

pumping test data and EER measurement data. However, because the pumping test data (HF data) are dominant in the MF kriging model and the pumping test is more expensive and 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 used in future studies on developing a more robust optimization framework that incorporates data cost and fidelity and can reveal their complex interactions.

5. Discussions and conclusions

This work presents a robust approach to exploit multi-source data to estimate the 3D random field of hydraulic conductivities. We demonstrated how the described framework can use the combination of pumping test data from boreholes, which are expensive and highly accurate, with observation data from less expensive and less accurate EER measurements. This approach offers a cost-effective approach to reliably characterize the hydraulic conductivity properties, specifically in undersampled sites, and can be particularly used in obtaining large-scale parameter maps for a region using small-scale measurements in an efficient way. For the first time, we studied the distribution effect of different fidelity data, showing that the estimation accuracy of MF kriging depends on the locations of both the LF and HF data. When HF data points are sparse and the location is far from the other HF data points, the information provided from the LF data becomes crucial and can greatly enhance model accuracy.

This study suggests that HF data can provide more information to the model compared with LF data. However, HF data are generally more expensive to obtain, mainly because of their more precise testing process. For example, pumping tests require drilling wells into the ground, which costed approximately $11,000 for each 80 m well for each data point in this study. However, the EER measurements were conducted completely on the surface, with no need for drilling. This makes the cost of EER testing much lower, at approximately only $600 for 80 m deep continuous data. Since the results also demonstrate that LF data can provide useful information to enhance the model estimation, especially in regions where data points are sparsely distributed, there is a trade-off between deciding on HF versus LF measurements. By implementing Bayesian experimental design along with the current confidence levels from the kriging model, optimal sensor placement locations for future data collection are suggested, which were related to the expected value of information from future sensor data. To rigorously inform the decision as to what should be the combination of LF and HF measurements, future study is needed to develop a more holistic optimization framework that incorporates both data cost and fidelity and evaluates their complex interplay.

Some future works are required according to the limitations of the model and the assumptions made for data usage. For example, EER measurements can correlate very differently with hydraulic conductivity on different scales and different soil mediums since soil porosity affects the correlation. The relationships between EER and hydraulic conductivity used in the study were obtained from an experiment with sand (Lu et al., 2019), which might not be representative of other types of geologic materials. This consideration requires additional experiments to obtain a more universal empirical relationship between EER and hydraulic conductivity, which is outside the scope of the current study. Also, kriging is limited to the condition of when correlations are non-local in a continuous field. Sometimes, in a sparsely sampled area or in cases when the field is discontinuous because of rivers, fractures, or faults, kriging might not be necessarily accurate. Therefore, more measurements are needed to further confirm our approach for this study site. However, the current study presents a robust framework that reveals important facts about the usage of MF models along with Bayesian statistics tools to enhance understanding and future observation design of hydro-geology properties based on the existing data. Once the data provide certain levels of confidence, MF kriging models can extract useful information from the data according to different levels of fidelity to make reasonable predictions with corresponding assumptions.

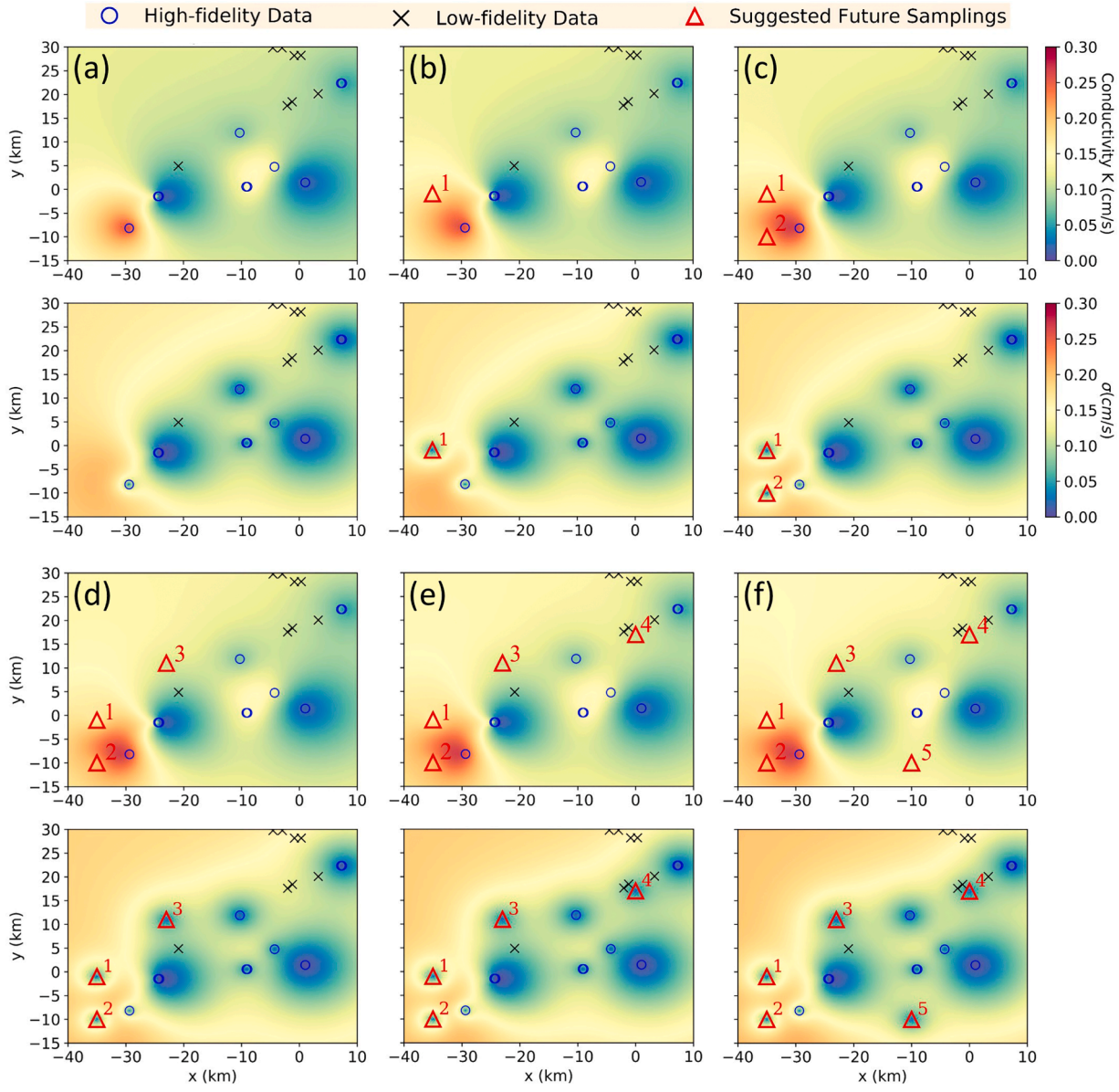


Fig. 10. The suggested sequential optimal sampling locations using Bayesian experimental design with the MF kriging model for the deepest (fifth) layer. (a) Initial kriging result. (b) Updated mean and variance with the first observation point. (c) Updated mean and variance with the first and second observation points. (d) Updated mean and variance with the first , second, and third observation points. (e) Updated mean and variance with the first, second, third, and fourth observation points. (f) Updated mean and variance with all five optimal observation points. Blue circle markers represent the HF data locations. Black cross markers represent the LF data locations. Red triangles represent the suggested optimal future sampling locations. The red numbers represent the order of the samplings.

CRediT authorship contribution statement

Chien-Yung Tseng: Methodology, Formal analysis, Investigation, Software, Visualization, Writing – original draft, Writing – review & editing. **Maryam Ghadiri:** Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Praveen Kumar:** Resources, Funding acquisition, Writing – review & editing. **Hadi Meidani:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

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

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 US Geological Survey. C.-Y.T. acknowledges support from the Illinois Water Resources Center, the Department of Civil and Environmental Engineering at University of Illinois at Urbana Champaign, and the Environmental Sciences Division at Oak Ridge National Laboratory. M. G. acknowledges support from the Illinois Water Resources Center and the U.S. Geological Survey. P.K. acknowledges support from the National Science Foundation Grant EAR1331906 for the Critical Zone Observatory for Intensively Managed Landscapes, a multi-institutional collaborative effort. H.M. acknowledges support from the Department of Civil and Environmental Engineering at University of Illinois at Urbana Champaign. Instrumentation and technical support were provided by

the Illinois State Geological Survey. Special thanks are given to Dr. Timothy Larson and Dr. Andrew Stumpf at the Illinois State Geological Survey, and Daniel R. Hadley at the Illinois State Water Survey, for providing data and supporting this research. Codes and data presented in this article are available on GitHub through Zenodo Data Repository (https://zenodo.org/record/7098032). C.-Y.T. is an employee of UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US DOE. Accordingly, the US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript or allow others to do so, for US Government purposes.

References

Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51 (8), 5957–5973.

Balaban, M., Dengiz, B., 2018. Lognormal ordinary kriging metamodel in simulation optimization. Int. J. (ORAJ) 5 (1), 1–12.

Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: a review. Statist. Sci. 273–304.

Fernández-Godino, M.G., C. Park, N.H. Kim, and R.T. Haftka. 2016. "Review of multi-fidelity models." *arXiv preprint* (arXiv:1609.07196).

Forrester, A.I.J., Sóbester, A., Keane, A.J., 2007. Multi-fidelity optimization via surrogate modelling. Mathemat., Phys. Engin. Sci. 463 (2088), 3251–3269.

Hamm, S.Y., Kim, M., Cheong, J.Y., Kim, J.Y., Son, M., Kim, T.W., 2007. Relationship between hydraulic conductivity and fracture properties estimated from packer tests and borehole data in a fractured granite. Eng. Geol. 92 (1–2), 73–87.

Keefer, Laura, and Erin Bauer. 2005. *Watershed monitoring for the Lake Decatur watershed: 2000-2003.* ISWS Contract Report CR-2005-09.

Kelly, W.E., 1977. Geoelectric sounding for estimating aquifer hydraulic conductivity. Groundwater 15 (6), 420–425.

Kelly, W.E., Frohlich, R.K., 1985. Relations Between Aquifer Electrical and Hydraulic Properties. Groundwater 23 (2), 182–189.

Kennedy, M.C., O'Hagan, A., 2000. Predicting the output from a complex computer code when fast approximations are available. Biometrika 87 (1), 1–13.

Khalil, M.A., Santos, F.A.M., 2009. Influence of degree of saturation in the electric resistivity– hydraulic conductivity relationship. Surv. Geophys 30 (6), 601.

Lesmes, D.P., Friedman, S.P., 2005. Relationships between the Electrical and Hydrogeological Properties of Rocks and Soils. In: Hydrogeophysics, 50. Springer, pp. 87–128.

Lindley, D.V., 1956. On a measure of the information provided by an experiment. Annals Mathemat. Stat. 27 (4), 986–1005.

Lu, C., Lu, J., Zhang, Y., Puckett, M.H., 2019. A convenient method to estimate soil hydraulic conductivity using electrical T conductivity and soil compaction degree. J. Hydrol. (Amst) 575, 211–220.

Mattia, C.M., Lovell, S.T., Davis, A., 2018. Identifying barriers and motivators for adoption of multifunctional perennial cropping systems by landowners in the Upper Sangamon River Watershed, Illinois. Agrofores. Syst. 92 (5), 1155–1169.

Mazáč, O., Císlerová, M., Kelly, W.E., Landa, I., Venhodová, D., 1990. Determination of hydraulic conductivities by Surface Geoelectrical Methods. Geotechnical and Environmental Geophysics: Volume II, Environmental and Groundwater. Society of Exploration Geophysicists, p. 352.

Mazáč, O., Kelly, W.E., Landa, I., 1985. A hydrogeophysical model for relations between electrical and hydraulic properties of aquifers. J. Hydrol. (Amst) 79 (1–2), 1–19.

Menberg, K., Bidarmaghz, A., Gregory, A., Choudhary, R., Girolami, M., 2020. Multi-fidelity approach to Bayesian parameter estimation in subsurface heat and fluid transport models. Sci. Total. Environ. 745, 140846.

Nash, S.G., 1984. Newton-Type Minimization Via the Lanczos Method. SIAM. J. Numer. Anal. 21, 770–778.

Niwas, S., Singhal, D.C., 1985. Aquifer transmissivity of porous media from resistivity data. J. Hydrol. (Amst) 82 (1–2), 143–153.

Norberg, T., Rosn, L., 2006. Calculating the optimal number of contaminant samples by means of data worth analysis. Environmetrics 17 (7), 705–719.

Oliver, M.A., Webster, R., 2015. Basic Steps in geostatistics: the Variogram and Kriging. SpringerBriefs in Agriculture, New York.

Oliver, M.A., Webster, R., 1990. Kriging: a method of interpolation for geographical information systems. Int. J. Geograph. Inform. Syst. 4 (3).

Peherstorfer, B., Willcox, K., Gunzburger, M., 2018. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. SIAM. Rev. 60 (3), 550–591.

Reinhart, D.R., 2006. Estimating the hydraulic conductivity of landfilled municipal solid waste using the borehole permeameter test. *J. Environmen. Engin.* (ASCE) 132 (6), 645.

Roth, C., 1998. Is lognormal kriging suitable for local estimation? Math. Geol 30 (8), 999–1009.

Selkregg, L.F., Kempton, J.P., 1958. Groundwater geology in east-central Illinois, a preliminary geologic report. Illinois. State. Geological. Survey. Circular no. 248.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Techn. J. 27 (3), 379–423.

Sikandar, P., Christen, E.W., 2012. Geoelectrical sounding for the estimation of hydraulic conductivity of alluvial aquifers. Water Resour. Manage. 26 (5), 1201–1215.

Slater, L., 2007. Near surface electrical characterization of hydraulic conductivity: from petrophysical properties to aquifer geometries—a review. Surv. Geophys 28 (2–3), 169–197.

Tizro, A.T., Voudouris, K.S., Salehzade, M., Mashayekhi, H., 2010. Hydrogeological framework and estimation of aquifer hydraulic parameters using geoelectrical data: a case study from West Iran. Hydrogeology (18), 917–929.

Yadav, G.S., 1995. Relating hydraulic and geoelectric parameters of the Jayant Aquifer, India. J. Hydrol. (Amst) 167 (1–4), 23–38.

Zaytsev, A., Burnaev, E., 2017. Large scale variable fidelity surrogate modeling. Annals Mathem. Artific. Intellig.-Spring. 81, 167–186.

Zhang, J., Man, J., Lin, G., Wu, L., Zeng, L., 2018. Inverse Modeling of Hydrologic Systems with Adaptive Multifidelity Markov Chain Monte Carlo Simulations. Water Resour. Res. 54 (7), 4867–4886.

Zhang, J., Zeng, L., Chen, C., Chen, D., Wu, L., 2015. Efficient Bayesian experimental design for contaminant source identification. Water Resour. Res 51 (1), 576–598.

Zheng, Q., Zhang, J., Xu, W., Wu, L., Zeng, L., 2018. Adaptive Multifidelity Data Assimilation for Nonlinear Subsurface Flow Problems. Water Resour. Res. 55 (1), 203–217.