Investigation of Multi-fidelity Co-Kriging Model for Hydraulic Conductivity in Sangamon Watershed

Maryam Ghadiri[1], Chien-Yung Tseng[2], and Hadi Meidani[2]

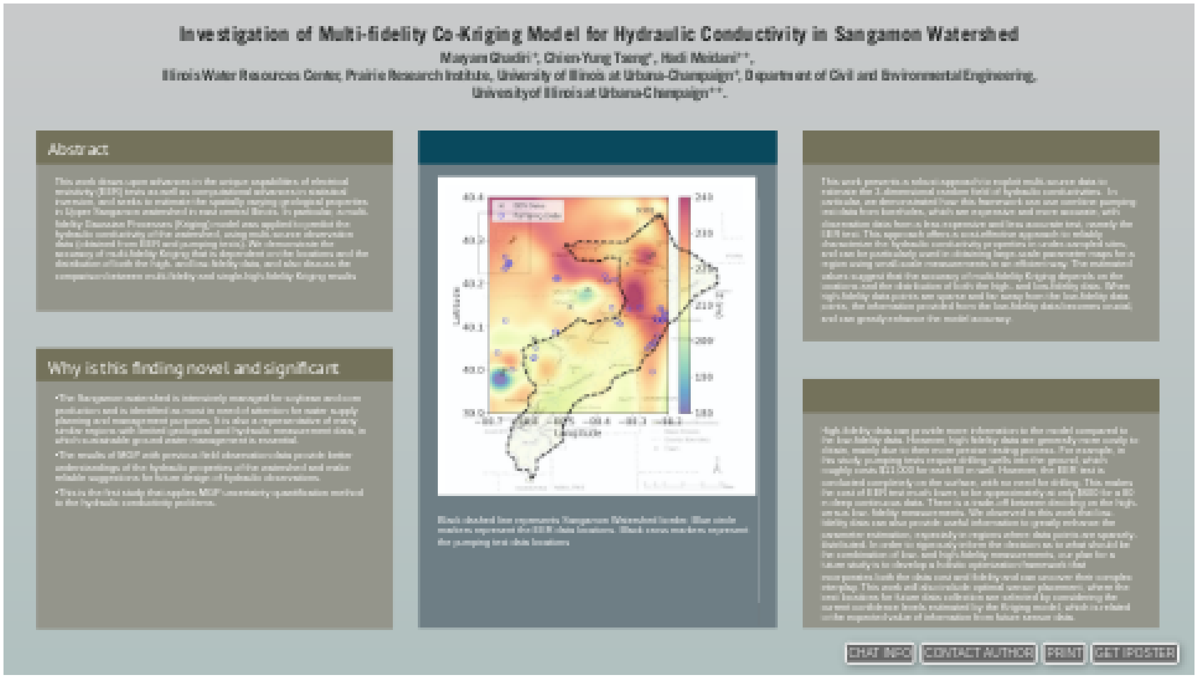
[1]Illinois Water Resources Center
[2]University of Illinois at Urbana Champaign

January 21, 2021

Abstract

Sangamon watershed is recognized as one of the most worth noting regions for water and environmental supply planning and management purposes according to its intensively management for soybean and corn production. It is also a representative area with limited geological and hydraulic measurement data, in which sustainable ground water and environmental management is essential. To better understand the hydraulic properties of the entire watershed, a multi-fidelity Gaussian Processes (Kriging) model was applied to predict the hydraulic conductivity of the upper Sangamon watershed, using previous multi-sources of field observation data (Electrical Earth Resistivity and pumping test data). The model also provided a quantification of uncertainty of the predicted values, which helps us to make reliable suggestions for the future design of hydraulic observations. The data fidelity effect to the model was discussed by comparing multi-fidelity and single-high-fidelity Kriging results. The model predicted values suggest that the accuracy of multi-fidelity Kriging depends on the locations and the distribution of both the high- and low-fidelity data. When high-fidelity data points are sparse and far away from the low-fidelity data points, the information provided from the low-fidelity data becomes extremely important, which can greatly enhance the model performance and accuracy. This study has paved the way to a more efficient parameter estimation in under-sampled sites by effectively estimating large-scale parameter maps using small-scale measurements and by applying uncertainty quantification method to a real watershed observation case. It will also draw upon and contribute to advances in Bayesian experimental design, and will optimally result in financial savings.

Investigation of Multi-fidelity Co-Kriging Model for Hydraulic Conductivity in Upper Sangamon Watershed



Maryam Ghadiri*, Chien-Yung Tseng*, Hadi Meidani**,
Illinois Water Resources Center, Prairie Research Institute, University of Illinois at Urbana-Champaign*, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign**.

ENTER NAMES OF AFFILIATED INSTITUTIONS

PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1–17 December 2020

ABSTRACT

Accurate estimation of hydrogeological properties may be limited for large areas, and these areas are often under-sampled. It is critical to establish an estimation framework where information from affordable small-scale measurements can also be used to estimate hydrological properties over a large area. This should be done in an optimal way where information from **expensive and cheaper tests** can be combined in a rigorous approach. This study presents a numerical framework where information from different measurement sources is combined to characterize the 3-dimensional random field representing the hydraulic conductivities of a watershed. This work draws upon advances in the unique capabilities of electrical resistivity (EER) tests as well as computational advances in statistical inversion, and seeks to estimate the spatially varying geological properties in Sangamon watershed in east central Illinois. In particular, a multi-fidelity Gaussian Processes (Kriging) model was applied to predict the hydraulic conductivity of the watershed, using multi-source observation data (obtained from EER and pumping tests). We demonstrate the accuracy of multi-fidelity Kriging that is dependent on the locations and the distribution of both the high- and low-fidelity data, and also discuss the comparison between multi-fidelity and single-high-fidelity Kriging results. When high-fidelity data points are sparse and far away from the low-fidelity data points, it is shown that information provided from the low-fidelity data can enhance the parameter estimation. The proposed framework will also offer quantified uncertainty/error in the hydraulic conductivity estimations, which can be used to assess how the model precision can be improved by obtaining new observation data in a future study.

WHY IS THIS FINDING NOVEL AND SIGNIFICANT

•The Sangamon watershed is intensively managed for soybean and corn production and is identified as most in need of attention for water supply planning and management purposes. It is also a representative of many similar regions with limited geological and hydraulic measurement data, in which sustainable ground water management is essential.

•The results of MGP with previous field observation data provide better understandings of the hydraulic properties of the watershed and make reliable suggestions for future design of hydraulic observations.

•This is the first study that applies MGP uncertainty quantification method to the hydraulic conductivity problems.

LOCATION AND METHOD

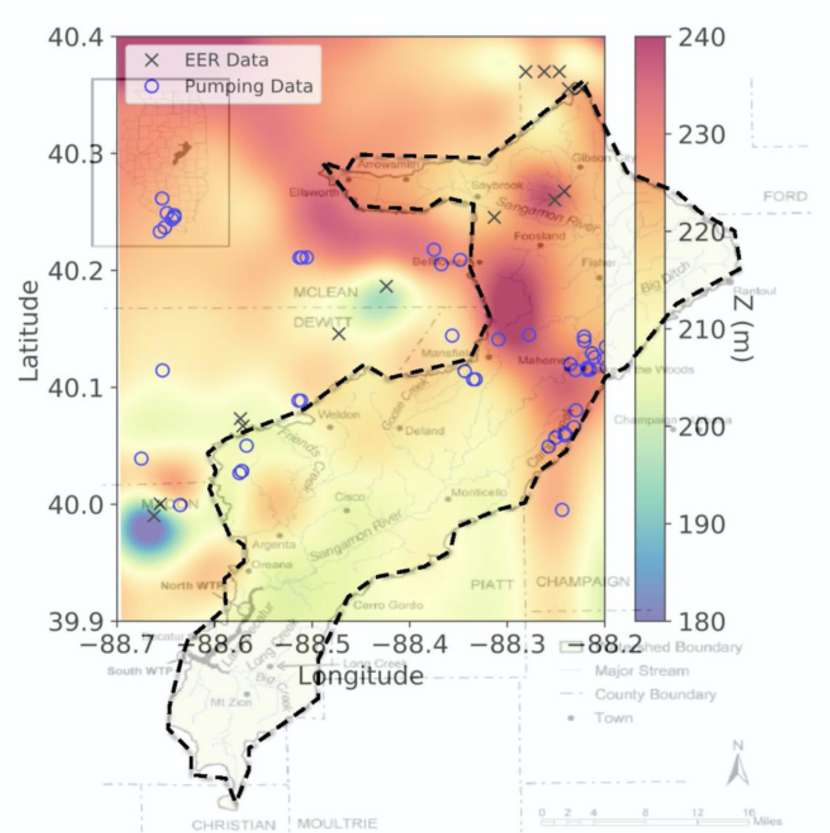


Figure 1. Black dashed line represents Upper Sangamon Watershed border. Blue circle markers represent the EER data locations. Black cross markers represent the pumping test data locations

-Upper Sangamon Watershed is chosen as the study site for the current study. Sangamon watershed is recognized as one of the most worth noting regions for water supply planning and management purposes according to its intensively management for soybean and corn production.

- EER measurement has long 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 and Frohlich, 1985; Slater, 2007; Khalil and Santos, 2009; Gomez et al., 2019)

- Each EER measurement was tested with dipole- dipole electrode configuration in a vertical two- dimensional plane in the Upper Sangamon Watershed as shown in Figure 2.1. The horizontal mean estimated hydraulic conductivity values along the z-direction were set as the representative values for the large-scale Kriging calculation. The equipment uncertainty of EER measurement is in an order of 1 (ohm-m), giving the initial variance of K as 10-3 (cm/s) (Kelly and Frohlich, 1985), which later will be used as the nugget value for the low-fidelity data in Co-Kriging.

Multi-fidelity Lognormal Ordinary Co-Kriging

To combine the observation data from EER measurement and pumping test, the Multi-fidelity Gaussian Processes Model (Co-Kriging) is used to perform a three-dimensional hydraulic conductivity mapping with smooth and continuous fusion of information from two sources with different fidelity/precision.

- In Multi-fidelity Kriging model, we treat EER and pumping test data separately when Kriging parameters were obtained from each semivariogram.

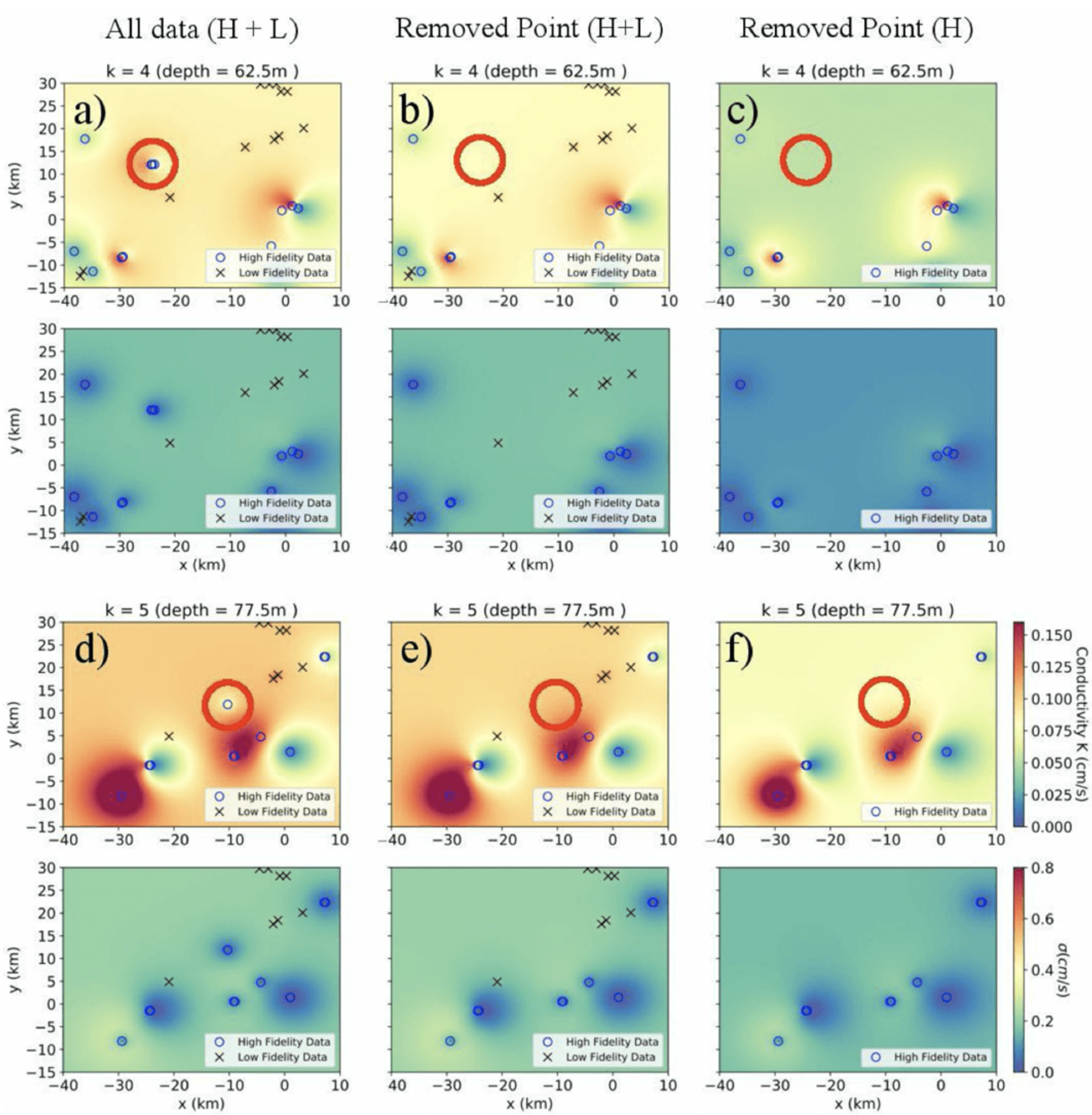


Figure 2. Comparisons between Multi-fidelity Kriging and Single-high-fidelity Kriging with specific points removal. a) and d) Multi-fidelity Kriging of the hydraulic conductivity and the corresponding standard deviation with all data points in the last two layers. b) and e) Multi-fidelity Kriging of the hydraulic conductivity and the corresponding standard deviation with specific point removals in the last two layers. c) and f) High-fidelity Kriging of the hydraulic conductivity and the corresponding standard deviation 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.

Optimal Future Sampling Locations

-Bayesian experimental design along with the multi-fidelity Kriging model was applied to infer the optimal future sampling locations. We chose the deepest (5th) layer, which has more uniform distribution of both high- and low-fidelity data points.

-5 optimal sampling locations were estimated one by one with the initial guess of the sampling locations that were uniformly assigned within the simulation domain, focusing on the watershed region

-Once the current optimal point was obtained, the hydraulic conductivity value can be predicted by the multi-fidelity Kriging model at that location

-The optimal locations are denoted by the red triangles with the numbers indicating the sequential order.

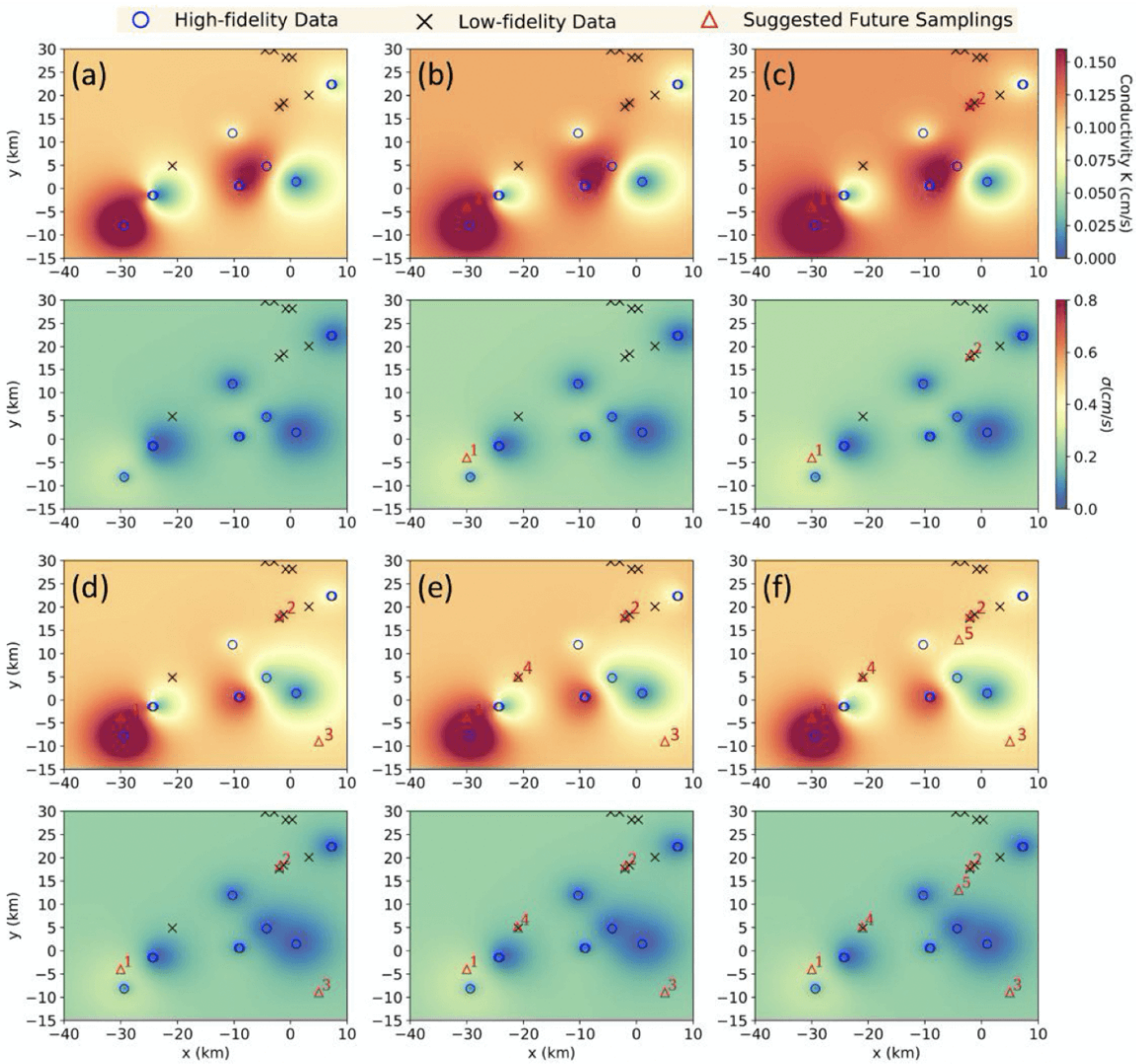


Figure 4. The suggested sequential optimal sampling locations using Bayesian experimental design with the Multi-fidelity Kriging model. (a) Initial Kriging result. (b) Updated mean and variance with the 1st observation point. (c) Updated mean and variance with the 1st and 2nd observation points. (d) 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 the 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.

DISCUSSION

- This work presents a robust approach to exploit multi-source data to estimate the 3-dimensional random field of hydraulic conductivities. In particular, we demonstrated how this framework can use combine 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 multi-fidelity Kriging depends on the locations and the distribution of both the high- and low-fidelity data.
- When high-fidelity data points are sparse and far away from the low-fidelity data points, the information provided from the low-fidelity data becomes crucial, and can greatly enhance the model accuracy.

CONCLUSION

- High-fidelity data can provide more information to the model compared to the low-fidelity data. However, high fidelity 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 high- versus low- fidelity measurements. We observed in this work that low-fidelity data can also provide useful information to greatly enhance the parameter estimation, especially in regions where data points are sparsely-distributed.
- In order to rigorously inform the decision as to what should be the combination of low- and high-fidelity measurements, our plan for a future study is to develop a holistic optimization framework that incorporates both the data cost and fidelity and can uncover their complex interplay.
- This work will also include 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.

What new questions does this lead to:

- For future usage of the model, we should also consider the estimated cost of the measurement and then select the specific region where more data investigations are needed. Finally applying the Bayesian model for the optimal sampling locations to make a more economical decision for the data.