



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}, Hadi Meidani²

¹Illinois Water Resources Center, Prairie Research Institute, University of Illinois at Urbana-Champaign

²Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

Contact information: cytseng2@illinois.edu

Acknowledgements

This research is funded under the provisions of section of the Water Resources Research Act annual base grants program made possible and distributed through the Illinois Water Resources Center and United States Geological Survey. CT and HM acknowledges support from Department of Civil and Environmental Engineering University of Illinois at Urbana Champaign.

Highlights

- We present a numerical framework where information from different field measurement sources is combined to characterize different layers of the 2-dimensional hydraulic conductivity field of the Upper Sangamon River Watershed (USRW), Illinois, USA, in a Multi-Fidelity estimation model.
- Enhanced water management systems depend on estimation of hydraulic properties (e.g., hydraulic conductivity) of geologic formations, which could vary over small spatial scales.
- A Multi-Fidelity (MF) Co-Kriging model was designed to estimate the geological properties by different sources of data.
- We investigated how a more accurate model can "learn" from new sensors using probabilistic statistical tools.
- Bayesian experimental design is used to select the best future sampling locations.

Method

Site selection:

- The Sangamon River is a major tributary to the Illinois River in U.S.A.
- 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.

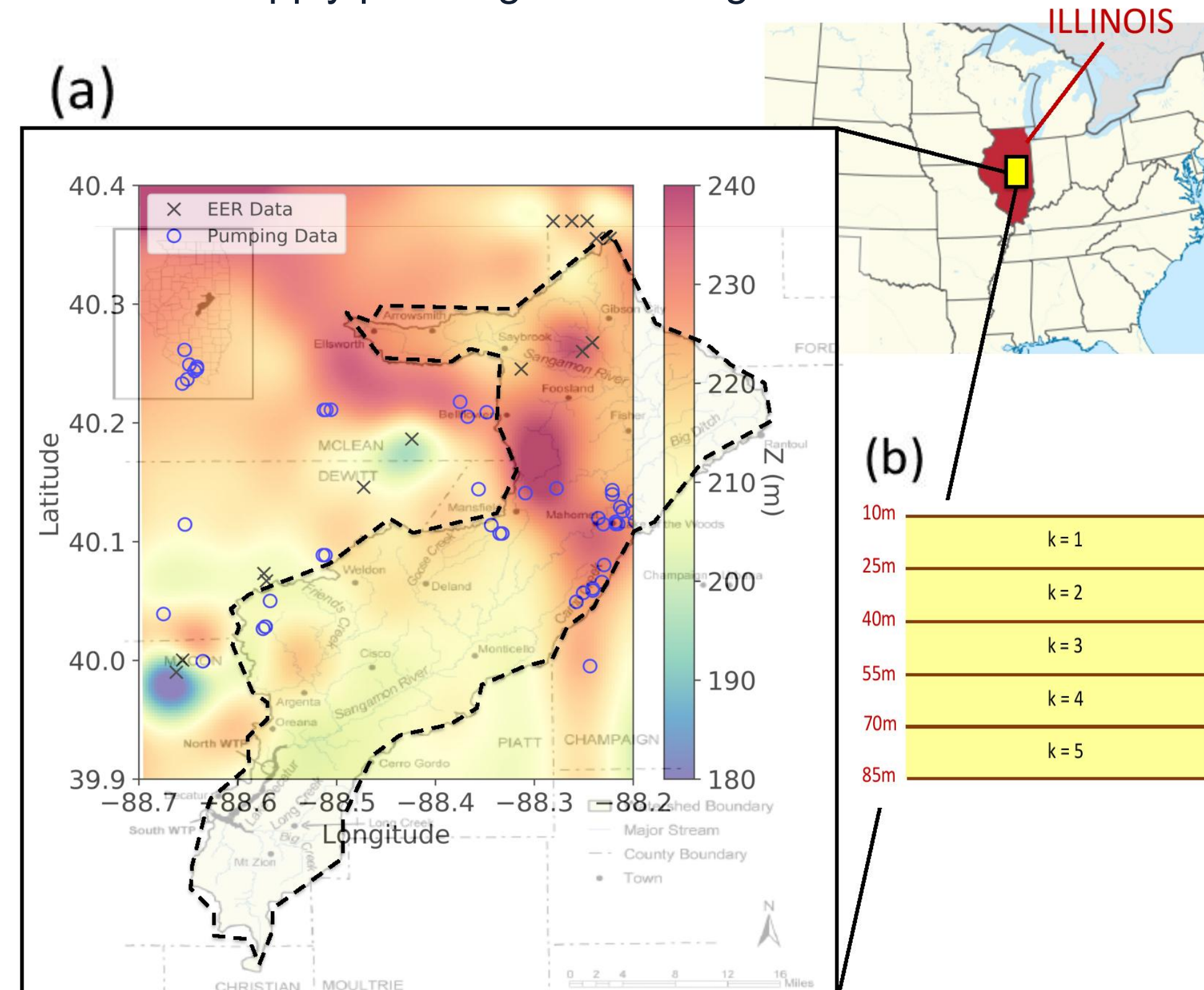


Figure 1 (a) Locations for data in the Upper Sangamon River Watershed (USRW) in Illinois, USA. (b) Sketch of the vertical layer setup.

Multi-Fidelity Lognormal Co-Kriging:

$$\begin{aligned} & f_H(x) = \rho u_L(x) + u_H(x), \begin{cases} u_L(x) \sim GP(0, k_L(x, x; \theta_L)) \\ u_H(x) \sim GP(0, k_H(x, x; \theta_H)) \end{cases} \\ & [f_L(x)] \sim GP\left(0, \begin{bmatrix} k_{LL}(x, x; \theta_L) & k_{LH}(x, x; \theta_L, \rho) \\ k_{HL}(x, x; \theta_L, \rho) & k_{HH}(x, x; \theta_H, \rho) \end{bmatrix}\right) \\ & \begin{cases} k_{LL}(x, x; \theta_L) = k_L(x, x; \theta_L) \\ k_{LH}(x, x; \theta_L, \rho) = k_{HL}(x, x; \theta_L, \rho) = \rho k_L(x, x; \theta_L) \\ k_{HH}(x, x; \theta_H, \rho) = \rho^2 k_L(x, x; \theta_L) + k_H(x, x; \theta_H) \end{cases} \\ & K_{ij} = k(x_i, x_j; \theta) = n + s \left(1 - \exp(3|x_i - x_j|/r)\right) \\ & \theta = (n, s, r) \rightarrow \text{Nugget } (n), \text{ Sill } (s), \text{ and Range } (r) \\ & \text{Optimization negative log marginal likelihood (NLML):} \\ & NLML(\theta_L, \theta_H, \rho) = \frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} + \frac{1}{2} \ln |\mathbf{K}| + \frac{N}{2} \ln(2\pi) \end{aligned}$$

Optimal Bayesian Experimental Design:

- Expected gain in Shannon information by the utility function $u(s, \mathbf{d}, \theta)$ with Bayes' theorem and Monte Carlo approach:
$$U(s) \approx \frac{1}{n} \sum_{i=1}^n \{\ln[p(d_i|\theta_i, s)] - \ln[p(d_i|s)]\}$$
- The optimal sampling location s^* can be obtained by maximizing the expected utility $U(s)$ over the design domain D :
$$s^* = \arg \max_{s \in D} [U(s)] = \arg \min_{s \in D} [-U(s)]$$

Results and discussion

Multi-Fidelity Co-Kriging:

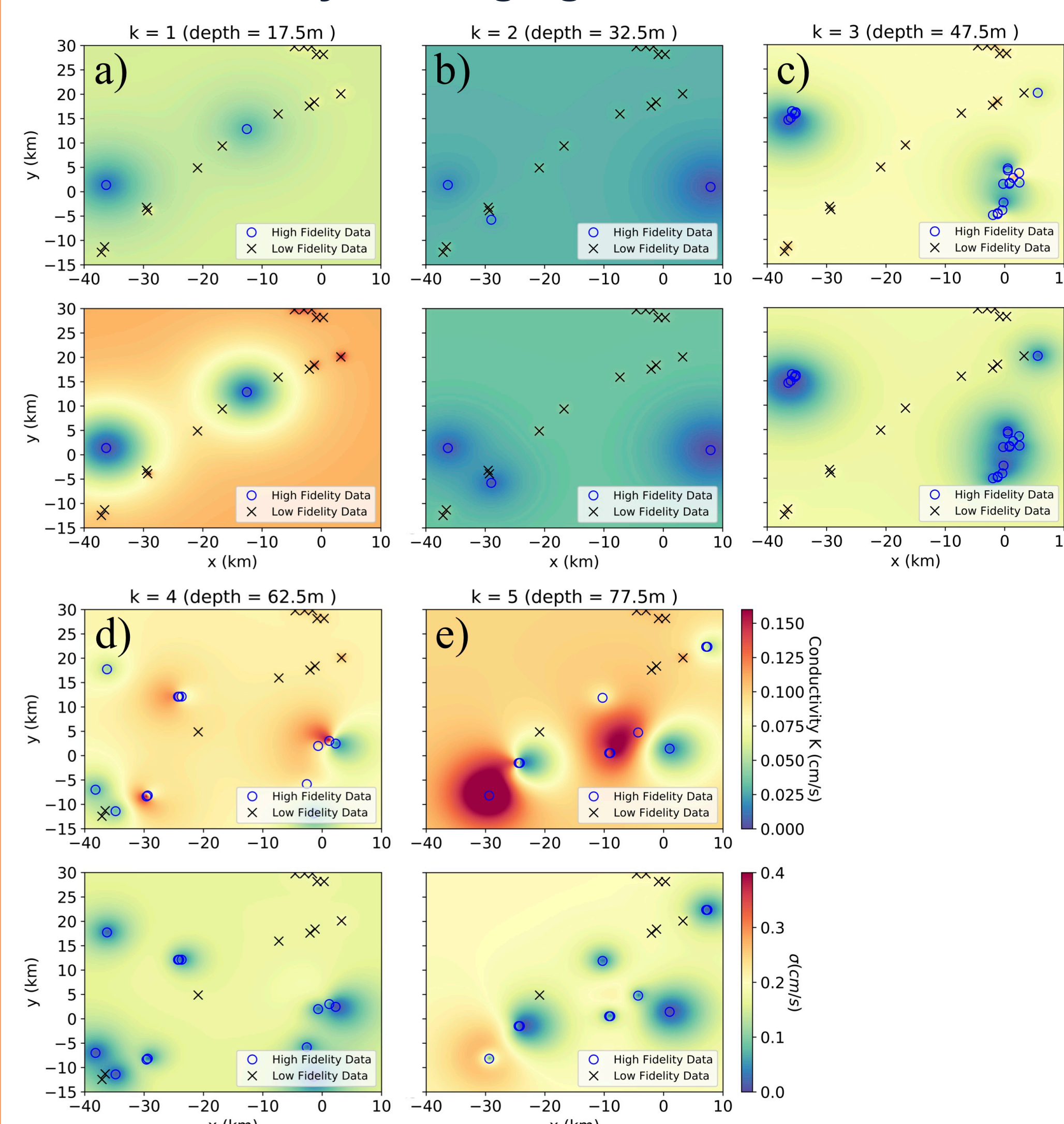


Figure 2 Multi-Fidelity Co-Kriging of the hydraulic conductivity and the corresponding standard deviation in the USRW in different depth layers. a) layer $k=1$, depth=17.5 m. b) layer $k=2$, depth=32.5 m. c) layer $k=3$, depth=47.5 m. d) layer $k=4$, depth=62.5 m. e) layer $k=5$, depth=77.5 m. The value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

Single-High-Fidelity Kriging:

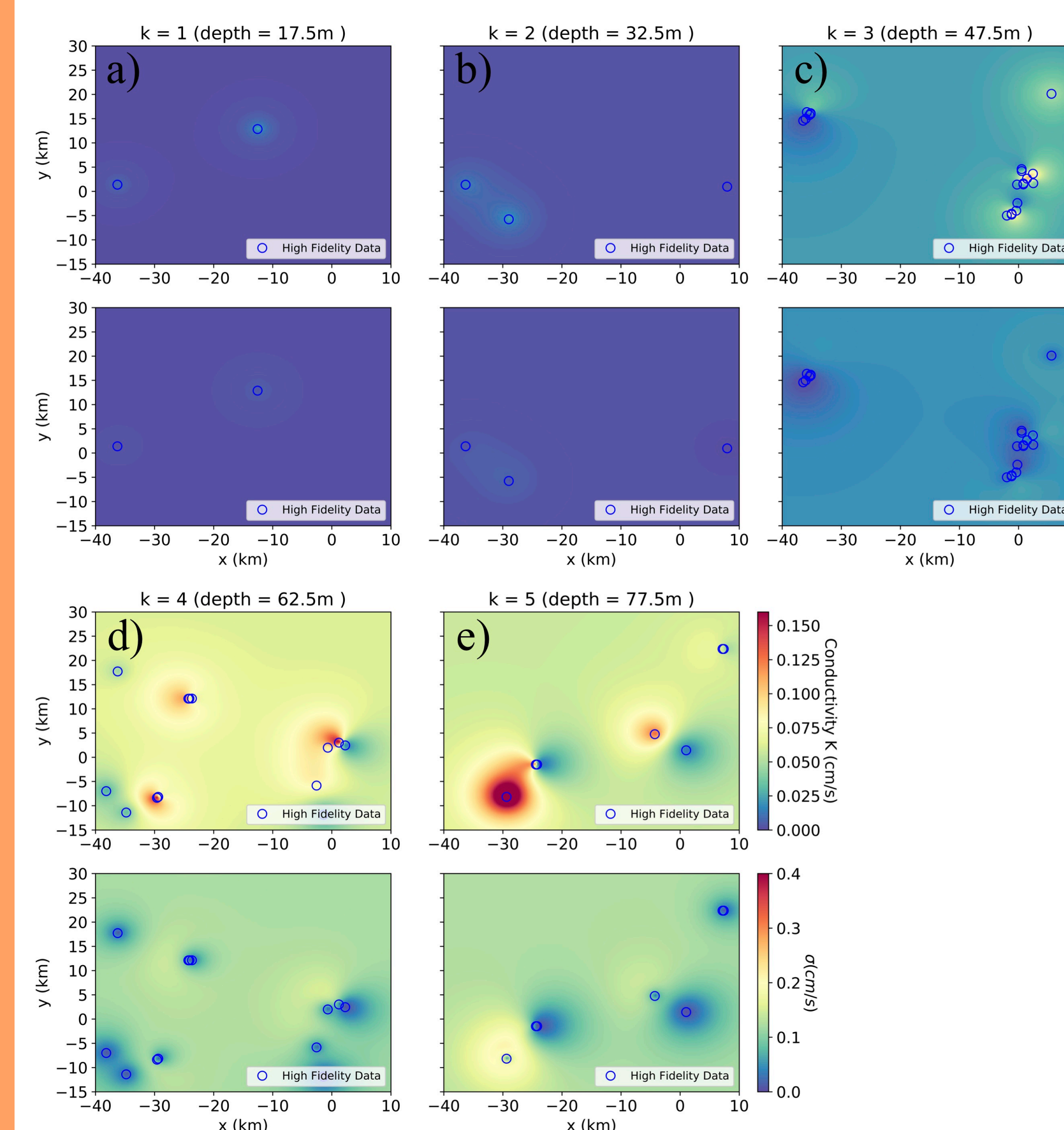


Figure 3 Single-High-Fidelity Kriging of the hydraulic conductivity and the corresponding standard deviation in the USRW in different depth layers. a) layer $k=1$, depth=17.5 m. b) layer $k=2$, depth=32.5 m. c) layer $k=3$, depth=47.5 m. d) layer $k=4$, depth=62.5 m. e) layer $k=5$, depth=77.5 m. The value of depth shown on top of each panel is the center z-location in each layer.

Fidelity Effect on the Estimation Accuracy:

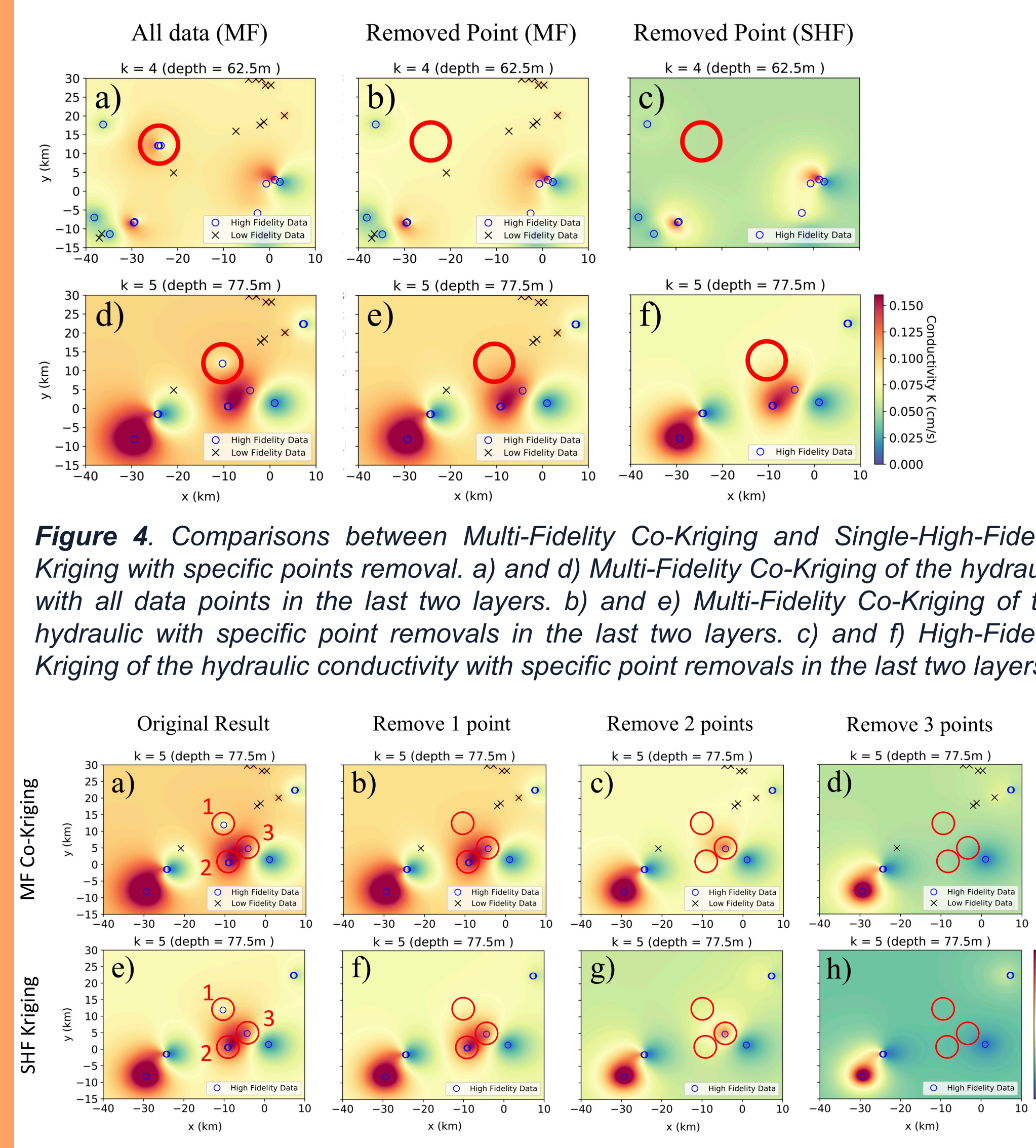


Figure 5 Comparisons between Multi-Fidelity Co-Kriging and Single-High-Fidelity Kriging in the last layer (depth > 70 m) with three consecutive points removal. Multi-Fidelity Co-Kriging of the hydraulic conductivity with a) all data points. b) 1 point removal. c) 2 points removal. d) 3 points removal. 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.

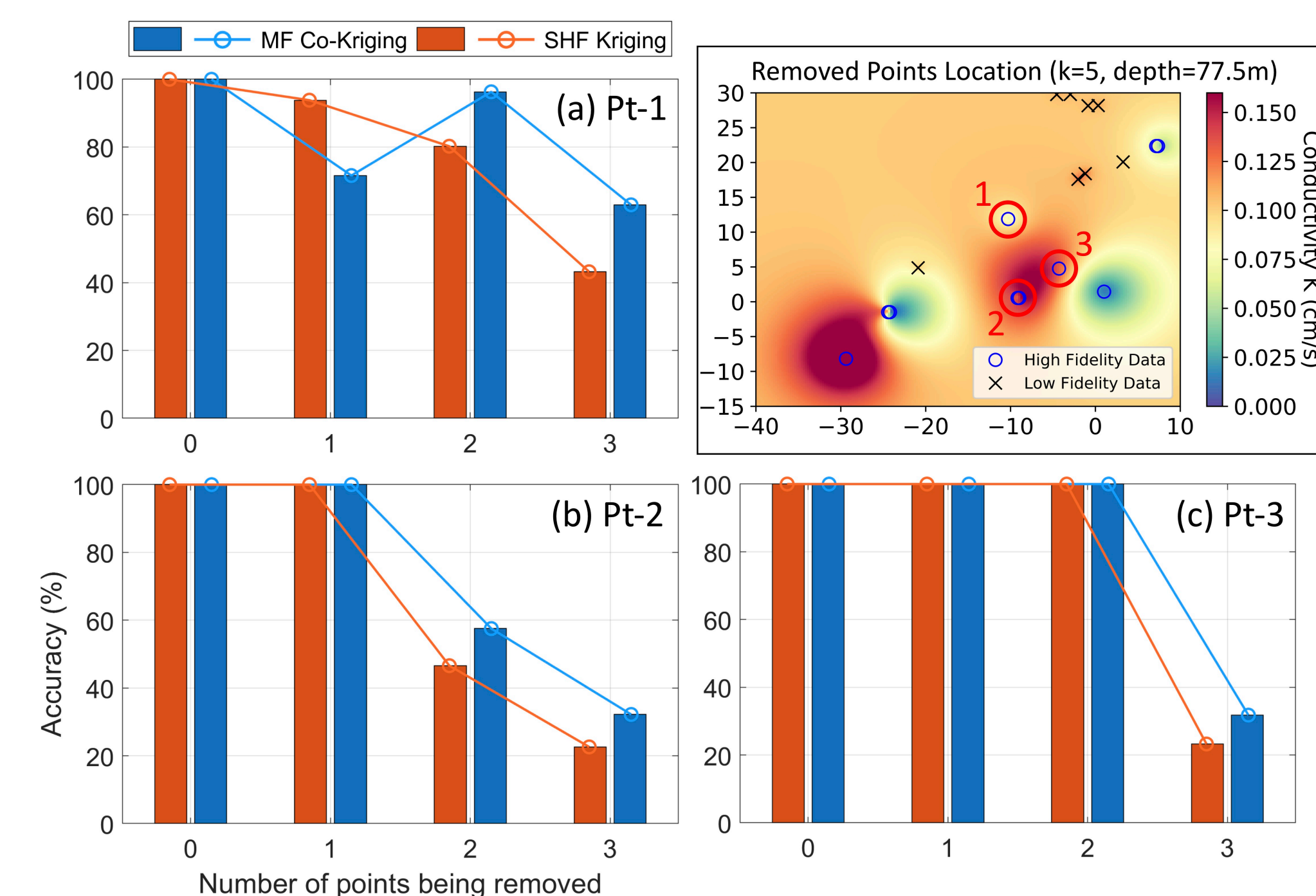


Figure 6 The accuracy of (a) point 1, (b) point 2, and (c) point 3 under Multi-Fidelity Co-Kriging and Single-High-Fidelity Kriging when removing 0 point, 1 point, 2 points, and 3 points. The removed points 476 location is shown in the top-right panel, and the points' removing order follows the denoted number of the points.

Future Data Collection Using Bayesian Experimental Design:

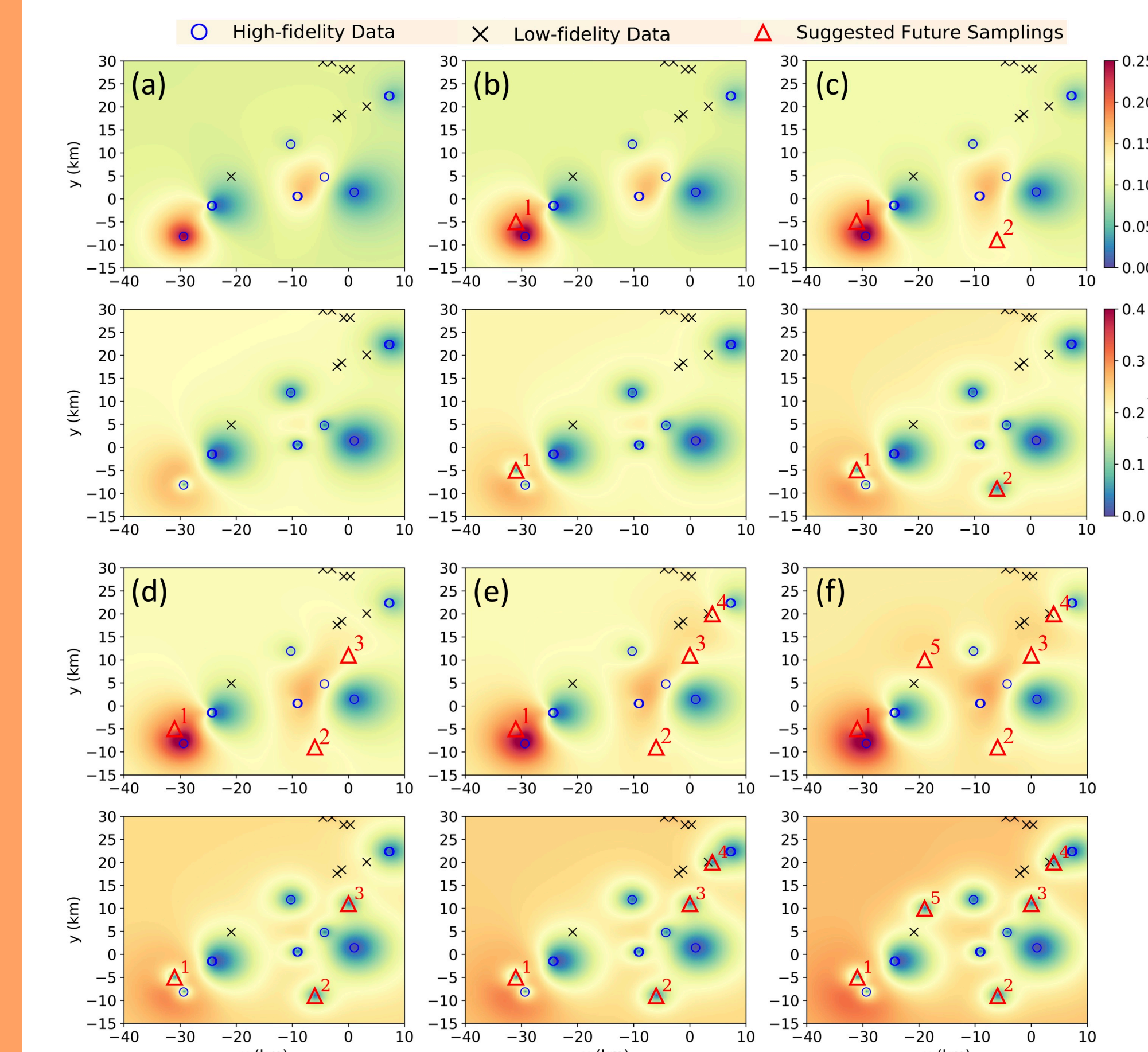


Figure 7 The suggested sequential optimal sampling locations using Bayesian experimental design with the Multi-Fidelity Co-Kriging model for the deepest (5th) layer. (a) Initial Kriging result. (b) Updated mean and variance with the 1st observation point. (c) Updated mean and variance with the 1st and 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.

Conclusions

- The estimated values suggest that the accuracy of MF Co-Kriging depends on the locations and the distribution of both the Low-Fidelity (LF) and High-Fidelity (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.
- Future work to rigorously inform the decision should combine LF and HF measurements, to develop a more holistic framework that incorporates both the data cost and fidelity and can uncover their complex interplay.