Tutorial of the Bayesian Experimental Design Coupling with Lognormal Multi-fidelity Gaussian Processes (Co-Kriging) – An Estimation of Hydraulic Conductivity in a Watershed

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

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

[2] Illinois State Geological Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign

[3] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign.

Abstract

This is a brief tutorial of Bayesian Experimental Design coupling with Multi-fidelity Gaussian Processes (Co-Kriging). An application example on estimation of hydraulics conductivity in a watershed is included in the package. The full code for this tutorial can be found on Github (https://github.com/Chienyungtseng/Multi-fidelity-Gaussian-Processes). More details can be found in the article "Estimation of Hydraulic Conductivity in a Watershed Using Multi-source Data via Co-Kriging Coupling with Bayesian Experimental Design" (submitted to Journal of Hydrology).

Lognormal Ordinary Kriging

Lognormal Ordinary Kriging (LOK) model algorithm follows the structure of Gaussian Processes:

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

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:

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

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 semivariogram according to the Kriging function. Given the observation data $\{\boldsymbol{x}, \boldsymbol{y}\}$, the semivariance, γ, of LOK model can be expressed as:

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

where $d_{ij} = |x_i - x_j|$ and $\mathrm{E}(.)$ 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:

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

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

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

where

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

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

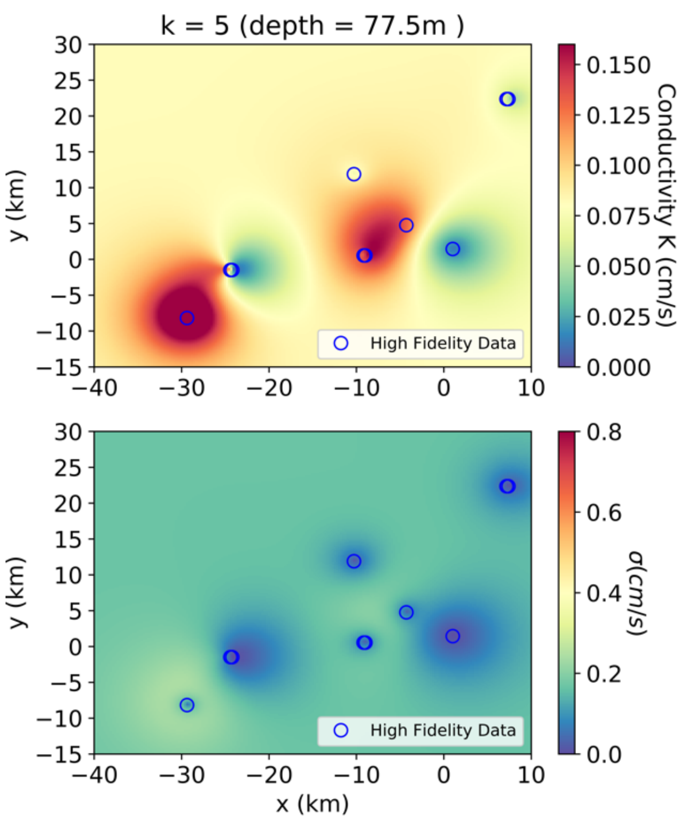
Since $f(\boldsymbol{x}^*)$ is in logarithmic scale, in order to estimate the parameter of interest, we need to convert the logarithmic values, $\boldsymbol{\mu_l}$ and $\boldsymbol{\sigma_l}$, back to the actual mean and standard deviation values according to:

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

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

Illustrative Example:

The figure shown below illustrates a Lognormal Ordinary Kriging application to the pumping test dataset in Upper Sangamon River Watershed. The upper panel shows the mean hydraulic conductivity and the lower panel shows the uncertainty deviations calculated from Kriging.



Illustrative Example: Lognormal Ordinary Kriging of the hydraulic conductivity and the corresponding standard deviation in the deepest layer (layer k = 5, depth = 77.5 m) of Upper Sangamon River Watershed. Blue circle markers represent the pumping test data locations.

Multi-fidelity Lognormal Ordinary Co-Kriging

The Co-Kriging algorithm assumes that

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

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

are two independent kriging functions. Then, the low-fidelity and high-fidelity 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:

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

where

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

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

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

where k_L and k_H are the Kriging functions for the LF and HF data, respectively, and ρ is the MF constant. Given the observation low-fidelity and high-fidelity 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 semivariogram according to the Kriging functions of the low-fidelity and high-fidelity data, respectively. To obtain the optimized ρ, normal distribution is applied:

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

where

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

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

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

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

where N is the total number of data points. For the estimations at a new set of points $\boldsymbol{x}^*$, we first construct the joint distribution:

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

where

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

Like the single-fidelity LOK model, according to the resulting conditional distribution, predictions can be estimated by

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

where

$$\boldsymbol{\mu_m} = \boldsymbol{q}^T\boldsymbol{K}^{-1}\boldsymbol{y}$$

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

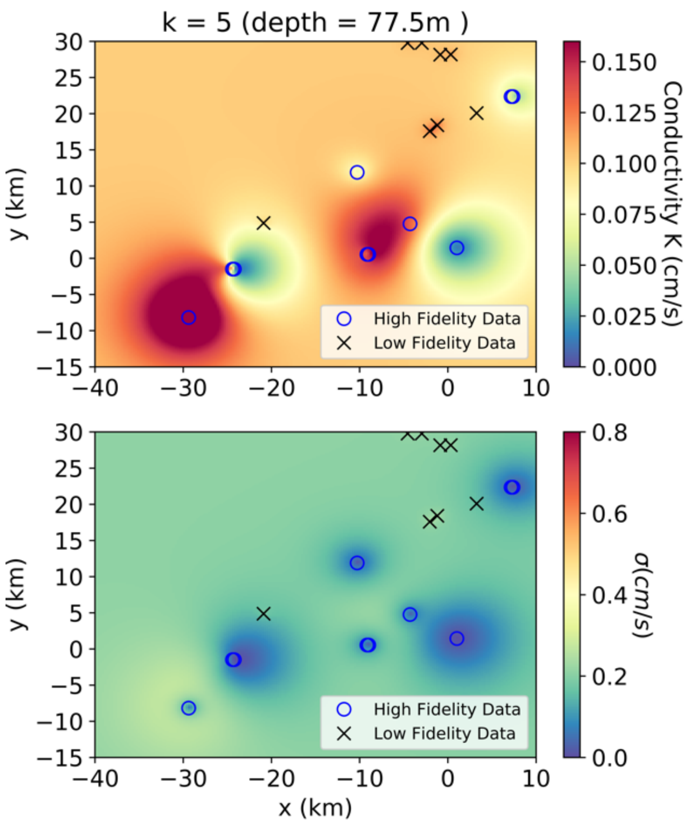
Finally, we backtransform the mean $\boldsymbol{\mu_m}$ and the standard deviation $\boldsymbol{\sigma_m}$ of the multi-fidelity model back into normal domain:

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

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

Illustrative Example:

The figure shown below illustrates a Multi-fidelity Lognormal Ordinary Co-Kriging application to the multi-source dataset: pumping test (high-fidelity) and Electricity Resistivity (low-fidelity) dataset in Upper Sangamon River Watershed. The upper panel shows the mean hydraulic conductivity and the lower panel shows the uncertainty deviation calculated from Kriging.



Illustrative Example: Multi-fidelity Lognormal Ordinary Co-Kriging of the hydraulic conductivity and the corresponding standard deviation in the deepest layer (layer k = 5, depth = 77.5 m) of Upper Sangamon River Watershed. Blue circle markers represent the Electricity Resistivity data locations. Black cross markers represent the pumping test data locations.

Optimal Bayesian Experimental Design

The Bayesian approach was applied to infer the optimal future sampling locations based on previous measurement data with the multi-fidelity Co-Kriging model. According to Bayesian inference, the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{d}, s)$ can be expressed as

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

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

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

Here, $\boldsymbol{\theta}$ is the sampled Kriging parameters, including n, s, and r. n and s are considered as constant values according to the multi-fidelity model, while r is considered as Gaussian distributed samples based on the fitted low-fidelity and high-fidelity 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 $\boldsymbol{\theta}$ does not affect by s, the prior distribution

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

The expected utility in Bayesian experimental design can be expressed as

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

where $u(s, \boldsymbol{d}, \boldsymbol{\theta})$ is the utility function. The relative entropy from the prior to the posterior is chosen as the utility function, which considers the expected gain in Shannon information given by the experiment.

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

According to Baye's theorem and Monte Carlo approach, the integral in the above equation can be approximated by the sum of the discrete values

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

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

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

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

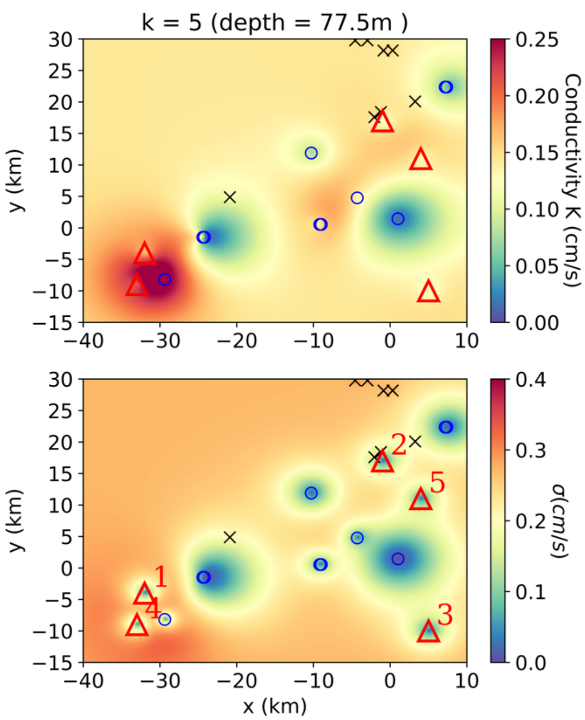
$$p(d_i|\theta_j, s) = \exp\left(-\frac{1}{2}(d_i - G(\theta_j, s))^2\right)$$

Combining all the equations, 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 $U(s)$

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

Illustrative Example:

The figure shown below illustrates a Bayesian Experimental Design combining with Multi-fidelity Lognormal Ordinary Co-Kriging application to the multi-source dataset: pumping test (high-fidelity) and Electricity Resistivity (low-fidelity) data in Upper Sangamon River Watershed. The upper panel shows the mean hydraulic conductivity and the lower panel shows the uncertainty deviation calculated from Kriging. 4 optimal sampling locations were estimated one by one by the Bayesian Experiment. Once the current optimal point was obtained, the hydraulic conductivity value was predicted by the multi-fidelity Co-Kriging model at the location. The current estimated optimal point was then put into the Kriging model as one of the measurement data point to update the model for the next optimal sampling location.



Illustrative Example: Updated mean and variance with all 4 sequential optimal future sampling points by the Bayesian experimental design with the multi-fidelity Co-Kriging model. 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.