UCI Samueli School of Engineering | Department of Mechanical and Aerospace Engineering

Multi-fidelity Modeling under Uncertainty

Ramin Bostanabad, Ph.D.
Assistant Professor
Mechanical and Aerospace Engineering
University of California, Irvine
Raminb@uci.edu

Advanced Modeling & Simulation (AMS) Seminar Series
NASA Ames Research Center, April 27th, 2023

https://pmacslab.eng.uci.edu/

OPTIMUM DATA COLLECTION

Design of M_2AX nanolaminate ternary alloys that exhibit many of the beneficial properties of both ceramic and metallic materials:

An early transition metal $M = \{Sc, Ti, V, Cr, Zr, Nb, Mo, Hf, Ta, W\}$ $\rightarrow$ 10
A main group element $A = \{Al, Si, P, S, Ga, Ge, As, Cd, In, Sn, Tl, Pb\}$ $\rightarrow$ 12
Either carbon or nitrogen $X = \{C, N\}$ $\rightarrow$ 2

$\rightarrow$ **240**

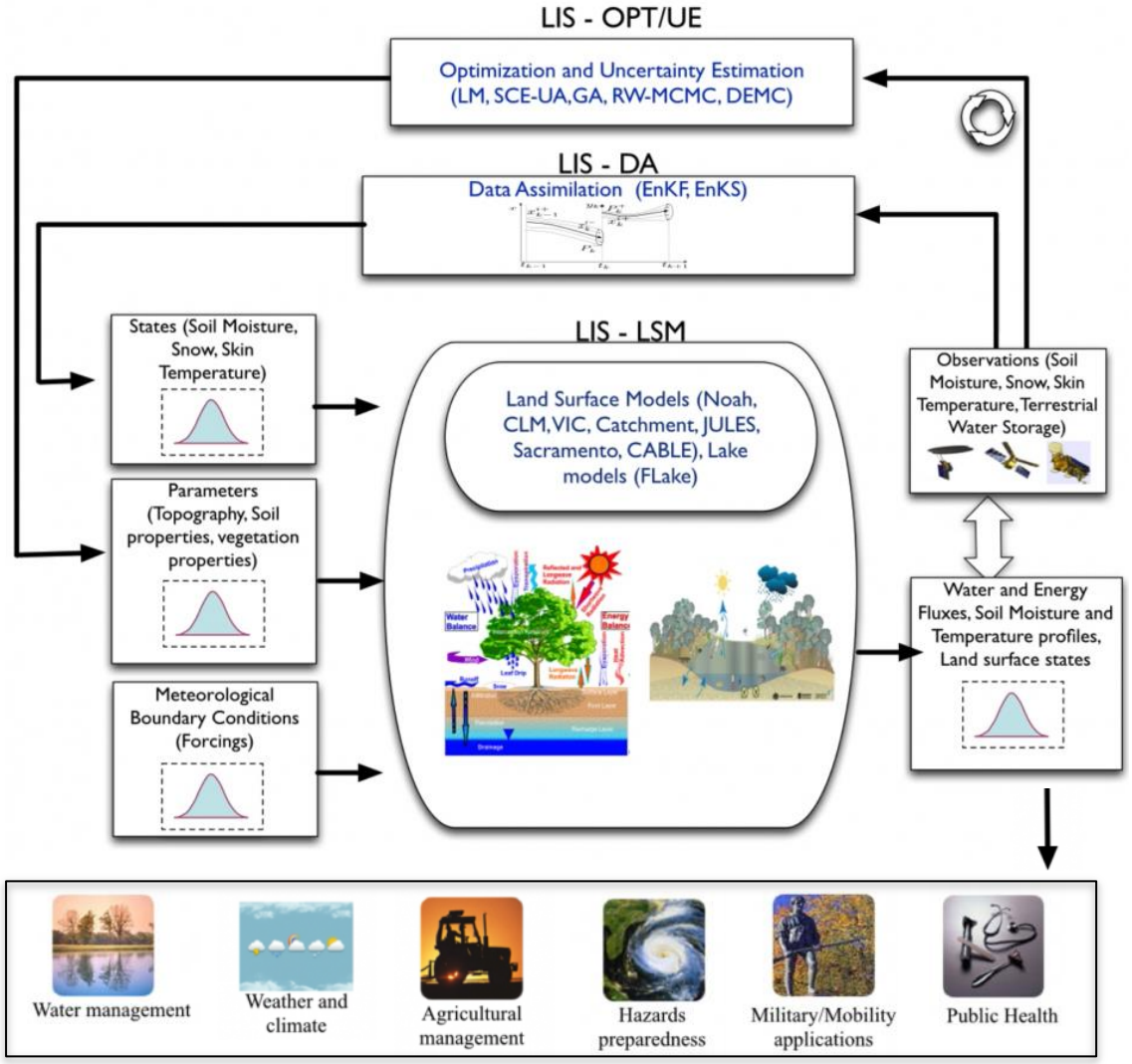
1. **How can we find the composition with the largest bulk modulus?**
2. **Can we use inaccurate but cheap simulations?**

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn

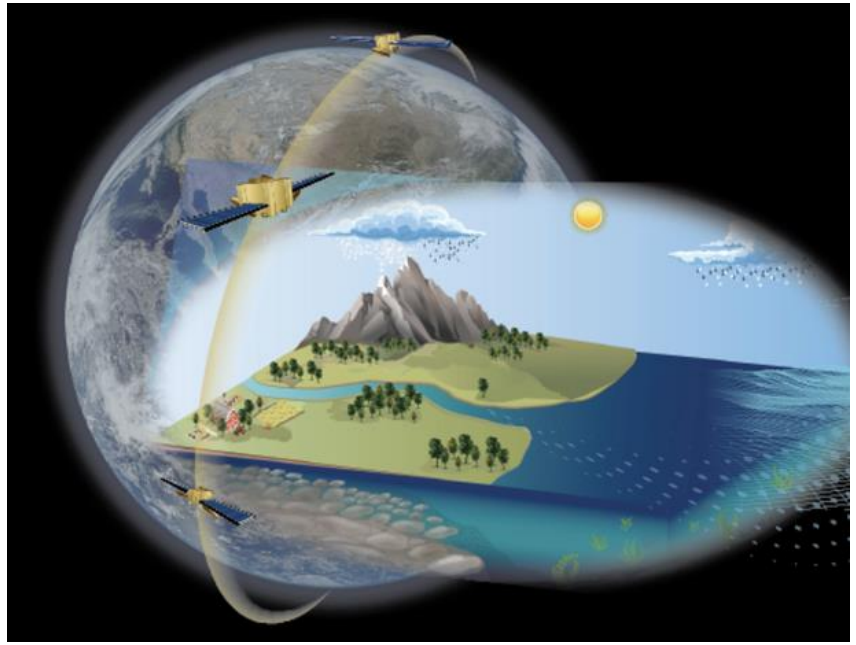
Zhou, Y.C., H.Y. Dong, X.H. Wang, and S.Q. Chen, Electronic structure of the layered ternary carbides ti2snc and ti2gec. Journal of Physics: Condensed Matter, 2000. 12(46): p. 9617.

PROBABILISTIC METAMODELING (EMULATION)

Land information system (LIS):



Freshwater component of LIS:



https://fwis.eis-hydro.mysmce.com/index.html

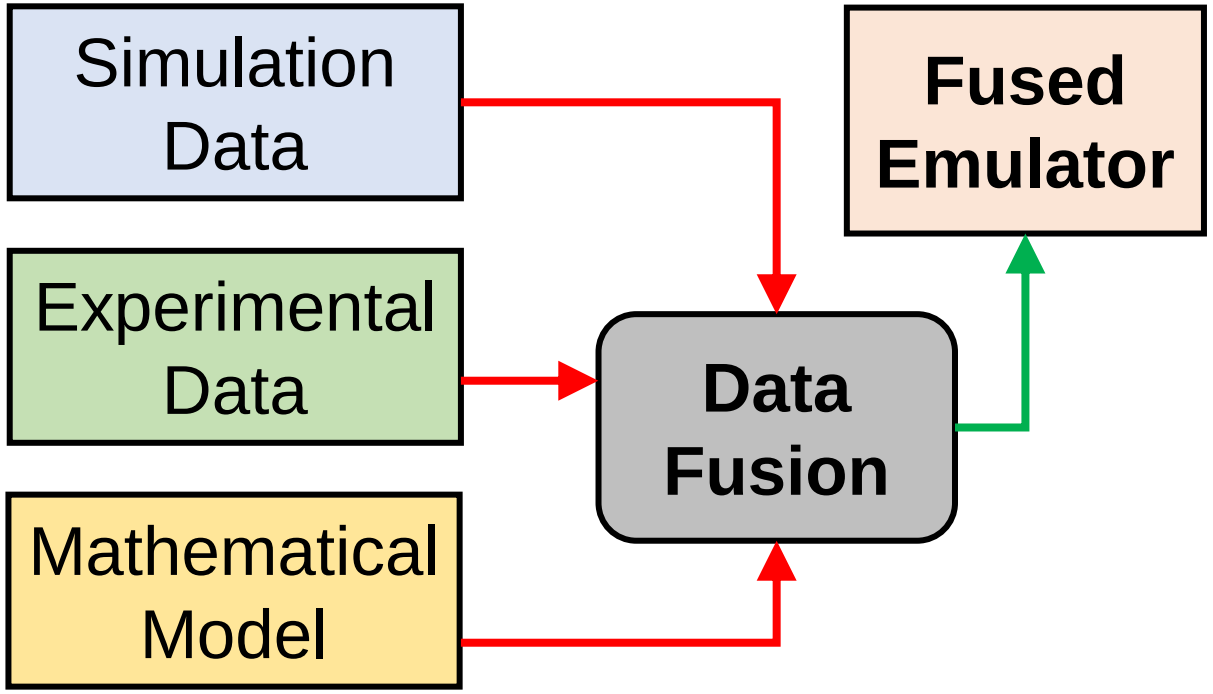
1. How can we reduce reliance on expensive simulations?
2. How can we detect anomalies?

https://pmacslab.eng.uci.edu/

Kumar, Sujay V., et al. "A Land Surface Data Assimilation Framework Using the Land Information System: Description and Applications." Advances in Water Resources 31.11 (2008): 1419-32. Print.

DATA FUSION OR MULTI-FIDELITY MODELING

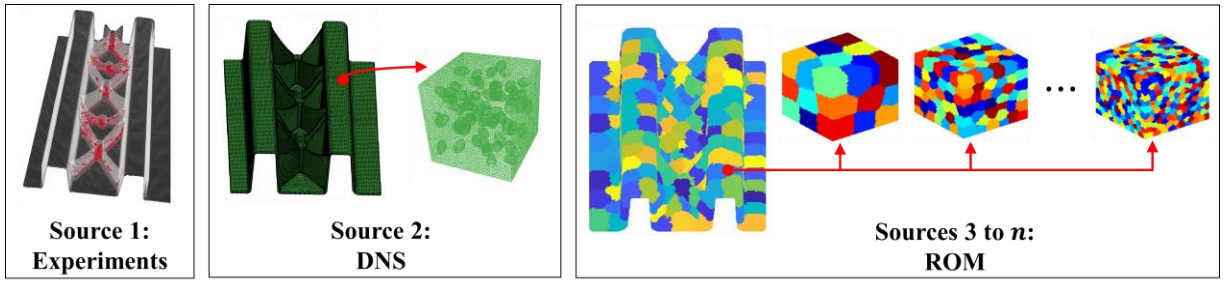
Learn from multiple data sources that model the same physical system but with **different fidelity levels** and **noise**!



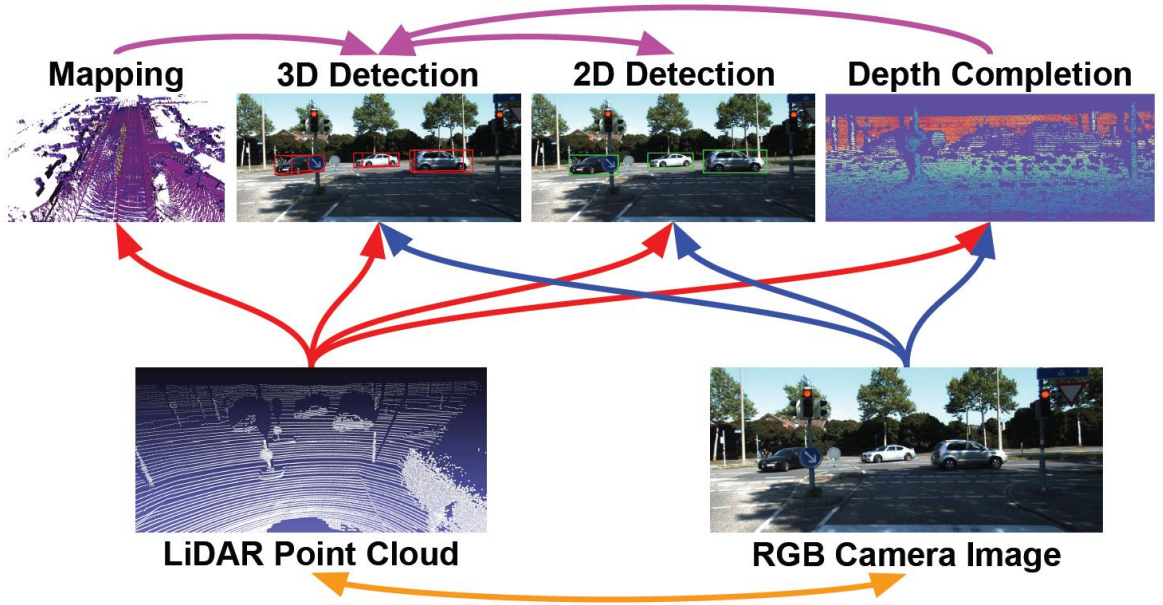
Eweis-Labolle, J.T., N. Oune, and R. Bostanabad, Data Fusion With Latent Map Gaussian Processes. Journal of Mechanical Design, 2022. 144(9).

https://pmacslab.eng.uci.edu/

Mechanics:



Self-driving cars:



Liang, M., et al. Multi-task multi-sensor fusion for 3d object detection. IEEE/CVF Conference, 2019.

ONE OF THE EXISTING TECHNOLOGIES

Consider the method of Kennedy and O'Hagan (KOH):

$$y_h(\boldsymbol{x}) = \rho y_l(\boldsymbol{x}, \boldsymbol{\theta}) + \delta(\boldsymbol{x}, \boldsymbol{\theta}) + \varepsilon$$

Inputs: $\boldsymbol{x}$ Calibration parameters: $\boldsymbol{\theta}$ scaling factor: ρ

High-fidelity source: y_h low-fidelity source: y_l noise: ε

Limitations:

1. Relies on explicit formulations
2. Accommodates only 2 sources

Kennedy, Marc C, and Anthony O'Hagan. "Bayesian Calibration of Computer Models." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63, no. 3 (2001): 425-64.

2. Procedure

The full data vector, $\boldsymbol{d} = [\boldsymbol{y}, \boldsymbol{z}]^T$, is normally distributed with conditional mean and variance:

$$E(\boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}) = \boldsymbol{m_d} = \boldsymbol{H\beta}, \text{ Eq. 2}$$

$$var(\boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}) = \boldsymbol{V_d}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{V_1}(\boldsymbol{D_1}) & \rho \boldsymbol{C_1}(\boldsymbol{D_1}, \boldsymbol{D_2}(\boldsymbol{\theta})) \\ \rho\left(\boldsymbol{C_1}(\boldsymbol{D_1}, \boldsymbol{D_2}(\boldsymbol{\theta}))\right)^T & \rho^2 \boldsymbol{V_1}(\boldsymbol{D_2}(\boldsymbol{\theta})) + \boldsymbol{V_2}(\boldsymbol{D_2}) + \lambda \boldsymbol{I_q} \end{pmatrix}, \text{ Eq. 3}$$

Or

$$var(\boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}) = \boldsymbol{V_d}(\boldsymbol{\theta}) = \begin{pmatrix} \sigma_1^2 \boldsymbol{S_1}(\boldsymbol{D_1}) & \rho \sigma_1^2 \boldsymbol{R_1}(\boldsymbol{D_1}, \boldsymbol{D_2}(\boldsymbol{\theta})) \\ \rho\left(\sigma_1^2 \boldsymbol{R_1}(\boldsymbol{D_1}, \boldsymbol{D_2}(\boldsymbol{\theta}))\right)^T & \rho^2 \sigma_1^2 \boldsymbol{S_1}(\boldsymbol{D_2}(\boldsymbol{\theta})) + \sigma_2^2 \boldsymbol{S_2}(\boldsymbol{D_2}) + \lambda \boldsymbol{I_q} \end{pmatrix},$$

where:

- $\boldsymbol{\beta} = [\beta_1, \beta_2]^T$ and $\boldsymbol{H} = \begin{bmatrix} \boldsymbol{1}_{p\times 1} & \boldsymbol{0}_{p\times 1} \\ \rho \boldsymbol{1}_{q\times 1} & \boldsymbol{1}_{q\times 1} \end{bmatrix}_{(p+q)\times 2} = \begin{bmatrix} \boldsymbol{hp} & \boldsymbol{0}_{p\times 1} \\ \rho \boldsymbol{hq} & \boldsymbol{hq} \end{bmatrix}_{(p+q)\times 2}$
- $\boldsymbol{\phi} = [\rho, \lambda, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2]$
- The (i, j) element of $\boldsymbol{V_1}(\boldsymbol{D_1})$ is $c_1\left((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{u}_j, \boldsymbol{t}_j)\right) = \sigma_1^2 r_1\left((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{u}_j, \boldsymbol{t}_j)\right)$.
- The (i, j) element of $\boldsymbol{S_1}(\boldsymbol{D_1})$ is $r_1\left((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{u}_j, \boldsymbol{t}_j)\right)$.
- The (i, k) element of $\boldsymbol{C_1}(\boldsymbol{D_1}, \boldsymbol{D_2}(\boldsymbol{\theta}))$ is $c_1((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{x}_k, \boldsymbol{\theta}))$.
- The (k, l) element of $\boldsymbol{V_1}(\boldsymbol{D_2}(\boldsymbol{\theta}))$ is $c_1((\boldsymbol{x}_k, \boldsymbol{\theta}), (\boldsymbol{x}_l, \boldsymbol{\theta}))$.
- The (k, l) element of $\boldsymbol{V_2}(\boldsymbol{D_2})$ is $c_2(\boldsymbol{x}_k, \boldsymbol{x}_l)$.

It must be noted that $\boldsymbol{V_d}$ also depends on the hyperparameters $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ but these are not explicitly stated for simplicity. The joint posterior distribution of the parameters conditioned on the data is then:

$$P(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}|\boldsymbol{d}) \propto P(\boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}) P(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}), \text{ Eq. 4}$$

Completing the square for $\boldsymbol{\beta}$ (i.e., making a normal distribution in terms of it by manipulating the above), would give its estimate:

$$\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{d} \sim N(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{W}(\boldsymbol{\theta}))$$

where

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{W}(\boldsymbol{\theta}) \boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{d}, \text{ Eq. 5}$$

$$\boldsymbol{W}(\boldsymbol{\theta}) = (\boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{H})^{-1}, \text{ Eq. 6}$$

Integrating $\boldsymbol{\beta}$ out and assuming the prior can be simplified to $P(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}) \propto P(\boldsymbol{\theta}) P(\boldsymbol{\phi})$:

$$P(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{d}) \propto |\boldsymbol{V_d}(\boldsymbol{\theta})|^{-0.5} |\boldsymbol{W}(\boldsymbol{\theta})|^{0.5} \exp\left(-\frac{1}{2}\left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right)^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right)\right) P(\boldsymbol{\theta}) P(\boldsymbol{\phi}), \text{ Eq. 7}$$

To arrive at $P(\boldsymbol{\theta}|\boldsymbol{d})$, one needs to marginalize Eq. 7 wrt $\boldsymbol{\phi}$. However, this is intractable and very computationally demanding. So, the MLE estimates of $\boldsymbol{\phi}$ are plugged into Eq. 7 to reduce it to:

$$P(\boldsymbol{\theta}|\boldsymbol{d}, \hat{\boldsymbol{\phi}}) \propto |\boldsymbol{V_d}(\boldsymbol{\theta})|^{-0.5} |\boldsymbol{W}(\boldsymbol{\theta})|^{0.5} \exp\left(-\frac{1}{2}\left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right)^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right)\right) P(\boldsymbol{\theta}), \text{ Eq. 8}$$

Where $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and $\boldsymbol{W}(\boldsymbol{\theta})$ are given in Eq. 5 and Eq. 6. Such estimates are derived in **two stages**.

2.1. Parameter estimation

Stage 1: Fit a GP model to $\boldsymbol{y}$ to estimate β_1 and $\boldsymbol{\psi}_1$ via, e.g., the MLE method. Then, we will have:

$$E\left(\eta(\boldsymbol{u}^*, \boldsymbol{t}^*)|\boldsymbol{y}, \hat{\beta}_1, \hat{\boldsymbol{\psi}}_1\right) = \hat{\beta}_1 + \boldsymbol{g}^T(\boldsymbol{u}^*, \boldsymbol{t}^*) \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \left(\boldsymbol{y} - \boldsymbol{h}\boldsymbol{p}\hat{\beta}_1\right), \text{ Eq. 9}$$

$$cov\left(\eta(\boldsymbol{u}^*, \boldsymbol{t}^*), \eta(\boldsymbol{u}', \boldsymbol{t}')\right) = c_1((\boldsymbol{u}^*, \boldsymbol{t}^*), (\boldsymbol{u}', \boldsymbol{t}')) - \boldsymbol{g}^T(\boldsymbol{u}^*, \boldsymbol{t}^*) \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \boldsymbol{g}(\boldsymbol{u}', \boldsymbol{t}') +$$
$$\left(1 - \boldsymbol{h}\boldsymbol{p}^T \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \boldsymbol{g}(\boldsymbol{u}^*, \boldsymbol{t}^*)\right)^T \left(\boldsymbol{h}\boldsymbol{p}^T \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \boldsymbol{h}\boldsymbol{p}\right)^{-1} \left(1 - \boldsymbol{h}\boldsymbol{p}^T \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \boldsymbol{g}(\boldsymbol{u}', \boldsymbol{t}')\right), \text{ Eq. 10}$$

$\boldsymbol{g}_{p\times 1}(\boldsymbol{u}^*, \boldsymbol{t}^*)$ is a column vector with i^{th} element $c_1((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{u}^*, \boldsymbol{t}^*))$.

NOTE: **In the codes**, σ_1^2 is NOT in $\boldsymbol{g}_{p\times 1}(\boldsymbol{u}^*, \boldsymbol{t}^*)$ and in the above equations, $\boldsymbol{S}_1$ is used instead of $\boldsymbol{V}_1$.

Stage 2: Estimate $\boldsymbol{\psi}_2, \lambda$ and ρ by integrating the conditional distribution wrt the prior distribution of $\boldsymbol{\theta}$:

1. $\boldsymbol{z}|\boldsymbol{y}, \hat{\boldsymbol{\psi}}_1, \boldsymbol{\psi}_2, \rho, \lambda, \boldsymbol{\theta}$ is normally distributed.
2. The likelihood can be constructed by integrating the mean vector and covariance matrix wrt the prior of $\boldsymbol{\theta}$.
   a. $E\left(z_k|\boldsymbol{y}, \hat{\boldsymbol{\psi}}_1, \boldsymbol{\psi}_2, \rho, \lambda\right) = \int E\left(z_k|\boldsymbol{y}, \hat{\boldsymbol{\psi}}_1, \boldsymbol{\psi}_2, \rho, \lambda, \boldsymbol{\theta}\right) P(\boldsymbol{\theta}) d\boldsymbol{\theta}$

      Using Eq. 1, Eq. 9, and Eq. 10:

      $E\left(z_k|\boldsymbol{y}, \hat{\boldsymbol{\psi}}_1, \boldsymbol{\psi}_2, \rho, \lambda\right) = \beta_2 + \rho\hat{\beta}_1 + \rho\left(\int \boldsymbol{g}^T(\boldsymbol{x}_k, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}\right) \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \left(\boldsymbol{y} - \boldsymbol{h}\boldsymbol{p}\hat{\beta}_1\right)$
      $\boldsymbol{g}_{p\times 1}(\boldsymbol{x}_k, \boldsymbol{\theta})$ is a column vector with i^{th} element $c_1\left((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{x}_k, \boldsymbol{\theta})\right)$.

      In vector form:

      $E\left(\boldsymbol{z}|\boldsymbol{y}, \hat{\boldsymbol{\psi}}_1, \boldsymbol{\psi}_2, \rho, \lambda\right) = \boldsymbol{h}\boldsymbol{q}\beta_2 + \rho\hat{\boldsymbol{\zeta}}$, Eq. 11
      $\hat{\boldsymbol{\zeta}} = \boldsymbol{h}\boldsymbol{q}\hat{\beta}_1 + \int \boldsymbol{G}^T P(\boldsymbol{\theta}) d\boldsymbol{\theta}\, \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \left(\boldsymbol{y} - \boldsymbol{h}\boldsymbol{p}\hat{\beta}_1\right)$, Eq. 12
      $\boldsymbol{G}$ ... is a matrix with (i, k) element $c_1\left((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{x}_k, \boldsymbol{\theta})\right)$

   b. Since η, δ, and ε are a priori independent, we have: $\boldsymbol{V} = \rho^2 \boldsymbol{C} + \boldsymbol{V_2}(\boldsymbol{D_2}) + \lambda \boldsymbol{I_q}$, Eq. 13

      where the (k, l) element of $\boldsymbol{C}$ is (simplified using $\boldsymbol{hp} = \boldsymbol{1}_{p\times 1}$ and $\boldsymbol{hq} = \boldsymbol{1}_{q\times 1}$):

      $C_{kl} =$
      $\int c_1\left((\boldsymbol{x}_k, \boldsymbol{\theta}), (\boldsymbol{x}_l, \boldsymbol{\theta})\right) P(\boldsymbol{\theta}) d\boldsymbol{\theta}$
      $-tr(\boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \int \boldsymbol{g}(\boldsymbol{x}_l, \boldsymbol{\theta}) \boldsymbol{g}^T(\boldsymbol{x}_k, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta})$
      $+tr(\boldsymbol{W_1}(\boldsymbol{D_1}))$
      $-tr(\boldsymbol{W_1}(\boldsymbol{D_1}) \boldsymbol{1}_{p\times 1}^T \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \int \boldsymbol{g}(\boldsymbol{x}_l, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta})$
      $-tr(\boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \boldsymbol{1}_{p\times 1} \boldsymbol{W_1}(\boldsymbol{D_1}) \int \boldsymbol{g}^T(\boldsymbol{x}_k, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta})$
      $+tr(\boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \boldsymbol{1}_{p\times 1} \boldsymbol{W_1}(\boldsymbol{D_1}) \boldsymbol{1}_{p\times 1}^T \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \int \boldsymbol{g}(\boldsymbol{x}_l, \boldsymbol{\theta}) \boldsymbol{g}^T(\boldsymbol{x}_k, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta})$, Eq. 14

      where
      $\boldsymbol{W_1}(\boldsymbol{D_1}) = \left(\boldsymbol{1}_{p\times 1}^T \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \boldsymbol{1}_{p\times 1}\right)^{-1} = \frac{1}{\boldsymbol{1}_{p\times 1}^T \boldsymbol{V_1}^{-1}(\boldsymbol{D_1}) \boldsymbol{1}_{p\times 1}}$ (a scalar)

      In matrix notation, $\boldsymbol{C}$ would be:

      $\boldsymbol{C} = \boldsymbol{C1} - \boldsymbol{C2} + \boldsymbol{C3} - \boldsymbol{C4} - \boldsymbol{C5} + \boldsymbol{C6}$

      NOTE: **In the codes**, σ_1^2 is NOT in $\boldsymbol{C}$.

   c. Using Eq. 11 and Eq. 13, one can derive the likelihood function and then maximize it to find the point estimates for $\boldsymbol{\psi}_2, \lambda$ and ρ:

      $P(\boldsymbol{\psi}_2, \rho, \lambda|\boldsymbol{d}, \hat{\boldsymbol{\psi}}_1) \propto |\boldsymbol{V}|^{-0.5} |\boldsymbol{W}_2|^{0.5} \exp\left(-\frac{1}{2}\left(\boldsymbol{z} - \boldsymbol{1}_{q\times 1}\hat{\beta}_2 - \rho\hat{\boldsymbol{\zeta}}\right)^T \boldsymbol{V}^{-1} \left(\boldsymbol{z} - \boldsymbol{1}_{q\times 1}\hat{\beta}_2 - \rho\hat{\boldsymbol{\zeta}}\right)\right) P(\boldsymbol{\psi}_2, \rho, \lambda)$, Eq. 15

      where β_2 has been integrated out (by making a complete square) to find:

      $\hat{\beta}_2 = \boldsymbol{W}_2 \boldsymbol{1}_{q\times 1}^T \boldsymbol{V}^{-1}(\boldsymbol{z} - \rho\hat{\boldsymbol{\zeta}})$, Eq. 16

      $\boldsymbol{W}_2 = \left(\boldsymbol{1}_{q\times 1}^T \boldsymbol{V}^{-1} \boldsymbol{1}_{q\times 1}\right)^{-1}$, Eq. 17

      The negative log-likelihood of Eq. 15 (multiplied by 2) is:

      $\ln|\boldsymbol{V}| + \ln|\boldsymbol{W}_2^{-1}| + \left(\boldsymbol{z} - \boldsymbol{1}_{q\times 1}\hat{\beta}_2 - \rho\hat{\boldsymbol{\zeta}}\right)^T \boldsymbol{V}^{-1} \left(\boldsymbol{z} - \boldsymbol{1}_{q\times 1}\hat{\beta}_2 - \rho\hat{\boldsymbol{\zeta}}\right)$

   d. Taking the integrals above given the prior of $\boldsymbol{\theta}$. Assuming:
      $\boldsymbol{\theta} \sim Normal(\boldsymbol{m}_\theta, \boldsymbol{V}_\theta)$
      $c\left((\boldsymbol{u}, \boldsymbol{t}), (\boldsymbol{u}', \boldsymbol{t}')\right) = \sigma_1^2 \exp\left(-(\boldsymbol{u} - \boldsymbol{u}')^T \boldsymbol{\Omega}_u (\boldsymbol{u} - \boldsymbol{u}') - (\boldsymbol{t} - \boldsymbol{t}')^T \boldsymbol{\Omega}_t (\boldsymbol{t} - \boldsymbol{t}')\right)$

      The integrals above become:

      $\int \boldsymbol{G}^T P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \left(\int \boldsymbol{G} P(\boldsymbol{\theta}) d\boldsymbol{\theta}\right)^T$ so:
      $\int G_{ik} P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int c_1\left((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{x}_k, \boldsymbol{\theta})\right) P(\boldsymbol{\theta}) d\boldsymbol{\theta}$
      $= \sigma_1^2 |\boldsymbol{I}_s + 2\boldsymbol{V}_\theta \boldsymbol{\Omega}_t|^{-0.5}$
      $\times \exp\left(-(\boldsymbol{u}_i - \boldsymbol{x}_k)^T \boldsymbol{\Omega}_u (\boldsymbol{u}_i - \boldsymbol{x}_k)\right)$
      $\times \exp\left(-(\boldsymbol{m}_\theta - \boldsymbol{t}_i)^T (2\boldsymbol{V}_\theta + \boldsymbol{\Omega}_t^{-1})^{-1} (\boldsymbol{m}_\theta - \boldsymbol{t}_i)\right)$

      $\int c_1\left((\boldsymbol{x}_k, \boldsymbol{\theta}), (\boldsymbol{x}_l, \boldsymbol{\theta})\right) P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \sigma_1^2 \exp\left(-(\boldsymbol{x}_k - \boldsymbol{x}_l)^T \boldsymbol{\Omega}_u (\boldsymbol{x}_k - \boldsymbol{x}_l)\right)$

      $\int \boldsymbol{g}(\boldsymbol{x}_l, \boldsymbol{\theta}) \boldsymbol{g}^T(\boldsymbol{x}_k, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is a $p \times p$ matrix with (i, j) element:
      $\sigma_1^4 |\boldsymbol{I}_s + 4\boldsymbol{V}_\theta \boldsymbol{\Omega}_t|^{-0.5} \exp\left(-(\boldsymbol{u}_i - \boldsymbol{x}_l)^T \boldsymbol{\Omega}_u (\boldsymbol{u}_i - \boldsymbol{x}_l) - (\boldsymbol{x}_k - \boldsymbol{u}_j)^T \boldsymbol{\Omega}_u (\boldsymbol{x}_k - \boldsymbol{u}_j)\right)$
      $\times \exp\left(-\frac{1}{2}(\boldsymbol{t}_i - \boldsymbol{t}_j)^T \boldsymbol{\Omega}_t (\boldsymbol{t}_i - \boldsymbol{t}_j) - \frac{1}{2}\left(\boldsymbol{m}_\theta - \frac{\boldsymbol{t}_i + \boldsymbol{t}_j}{2}\right)^T \left(\boldsymbol{V}_\theta + \frac{\boldsymbol{\Omega}_t^{-1}}{4}\right)^{-1} \left(\boldsymbol{m}_\theta - \frac{\boldsymbol{t}_i + \boldsymbol{t}_j}{2}\right)\right)$

      $\int \boldsymbol{g}(\boldsymbol{x}_l, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is a vector with i^{th} element of $\int G_{il} P(\boldsymbol{\theta}) d\boldsymbol{\theta}$.

      Note that there is no need to, when coding, calculate the above element-by-element. It can be done via matrix algebra.

   e. Taking the integrals above given the prior of $\boldsymbol{\theta}$. Assuming:
      $\boldsymbol{\theta} \sim Uniform$ with $\theta_v \epsilon [a_v, b_v]$ for $v = 1, \ldots, s$
      $c\left((\boldsymbol{u}, \boldsymbol{t}), (\boldsymbol{u}', \boldsymbol{t}')\right) = \sigma_1^2 \exp\left(-(\boldsymbol{u} - \boldsymbol{u}')^T \boldsymbol{\Omega}_u (\boldsymbol{u} - \boldsymbol{u}') - (\boldsymbol{t} - \boldsymbol{t}')^T \boldsymbol{\Omega}_t (\boldsymbol{t} - \boldsymbol{t}')\right)$

      The integrals above become:

      $\int \boldsymbol{G}^T P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \left(\int \boldsymbol{G} P(\boldsymbol{\theta}) d\boldsymbol{\theta}\right)^T$ so:

      $\int G_{ik} P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int c_1\left((\boldsymbol{u}_i, \boldsymbol{t}_i), (\boldsymbol{x}_k, \boldsymbol{\theta})\right) P(\boldsymbol{\theta}) d\boldsymbol{\theta}$
      $= \sigma_1^2 (2\pi)^{\frac{s}{2}} |0.5\boldsymbol{\Omega}_t^{-1}|^{0.5}$
      $\times \exp\left(-(\boldsymbol{u}_i - \boldsymbol{x}_k)^T \boldsymbol{\Omega}_u (\boldsymbol{u}_i - \boldsymbol{x}_k)\right)$
      $\times \prod_{v=1}^{s} \frac{1}{b_v - a_v} \int_a^b MVN(\boldsymbol{t}_i, 0.5\boldsymbol{\Omega}_t^{-1})$

      $\int c_1\left((\boldsymbol{x}_k, \boldsymbol{\theta}), (\boldsymbol{x}_l, \boldsymbol{\theta})\right) P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \sigma_1^2 \exp\left(-(\boldsymbol{x}_k - \boldsymbol{x}_l)^T \boldsymbol{\Omega}_u (\boldsymbol{x}_k - \boldsymbol{x}_l)\right)$

      $\int \boldsymbol{g}(\boldsymbol{x}_l, \boldsymbol{\theta}) \boldsymbol{g}^T(\boldsymbol{x}_k, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is a $p \times p$ matrix with (i, j) element:
      $\exp\left(-(\boldsymbol{u}_i - \boldsymbol{x}_l)^T \boldsymbol{\Omega}_u (\boldsymbol{u}_i - \boldsymbol{x}_l) - (\boldsymbol{x}_k - \boldsymbol{u}_j)^T \boldsymbol{\Omega}_u (\boldsymbol{x}_k - \boldsymbol{u}_j)\right)$
      $\times \sigma_1^4 (2\pi)^{\frac{s}{2}} \left|\frac{1}{4}\boldsymbol{\Omega}_t^{-1}\right|^{\frac{1}{2}} \prod_{v=1}^{s} \frac{1}{b_v - a_v} \int_a^b MVN\left(\frac{1}{2}(\boldsymbol{t}_i + \boldsymbol{t}_j), \frac{1}{4}\boldsymbol{\Omega}_t^{-1}\right) \times$
      $\exp\left(-\frac{1}{2}(\boldsymbol{t}_i - \boldsymbol{t}_j)^T \boldsymbol{\Omega}_t (\boldsymbol{t}_i - \boldsymbol{t}_j)\right)$

      $\int \boldsymbol{g}(\boldsymbol{x}_l, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is a vector with i element of $\int G_{il} P(\boldsymbol{\theta}) d\boldsymbol{\theta}$.

2.2. Posteriors

Once $\boldsymbol{\phi}$ is estimated, the posterior distribution of $\boldsymbol{\theta}$ can be calculated via Eq. 8 above:

$$P(\boldsymbol{\theta}|\boldsymbol{d}, \hat{\boldsymbol{\phi}}) \propto |\boldsymbol{V_d}(\boldsymbol{\theta})|^{-0.5} |\boldsymbol{W}(\boldsymbol{\theta})|^{0.5} \exp\left(-\frac{1}{2}\left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right)^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right)\right) P(\boldsymbol{\theta})$$

Where the estimate of the full mean vector is used (compare it to $\hat{\beta}_1$ and $\hat{\beta}_2$):

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{W}(\boldsymbol{\theta}) \boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{d}, \text{ Eq. 5}$$

$$\boldsymbol{W}(\boldsymbol{\theta}) = (\boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{H})^{-1}, \text{ Eq. 6}$$

To predict the experimental response at a future point $\boldsymbol{x}^*$ based on the estimated hyper-parameters:

$$E\left(z^*|\boldsymbol{\theta}, \boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right) = [\hat{\rho}\ 1] \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{T}^T(\boldsymbol{x}^*, \boldsymbol{\theta}) \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right), \text{ Eq. 18}$$

$$cov\left(z^*, z'|\boldsymbol{\theta}, d, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right) = \hat{\rho}^2 c_1\left((\boldsymbol{x}^*, \boldsymbol{\theta}), (\boldsymbol{x}', \boldsymbol{\theta})\right) + c_2(\boldsymbol{x}^*, \boldsymbol{x}') + \hat{\lambda} -$$
$$\boldsymbol{T}^T(\boldsymbol{x}^*, \boldsymbol{\theta}) \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{T}(\boldsymbol{x}', \boldsymbol{\theta}) + \left([\hat{\rho}\ 1]^T - \boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{T}(\boldsymbol{x}^*, \boldsymbol{\theta})\right)^T (\boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{H})^{-1} \left([\hat{\rho}\ 1]^T - \boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{T}(\boldsymbol{x}', \boldsymbol{\theta})\right), \text{ Eq. 19}$$

Where $\boldsymbol{T}(\boldsymbol{x}^*, \boldsymbol{\theta}) = \begin{bmatrix} \hat{\rho} \boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{\theta}) \\ \hat{\rho}^2 \boldsymbol{g}_{12}(\boldsymbol{x}^*, \boldsymbol{\theta}) + \boldsymbol{g}_{22}(\boldsymbol{x}^*) \end{bmatrix}$, $\boldsymbol{g}_{12}(\boldsymbol{x}^*, \boldsymbol{\theta})$ is a column vector of size $q \times 1$ with j^{th} element $c_1\left((\boldsymbol{x}_j, \boldsymbol{\theta}), (\boldsymbol{x}^*, \boldsymbol{\theta})\right)$, and $\boldsymbol{g}_{22}(\boldsymbol{x}^*)$ is a column vector of size $q \times 1$ with j^{th} element $c_2(\boldsymbol{x}_j, \boldsymbol{x}^*)$.

Now, the posterior distribution of $\boldsymbol{\theta}$ from Eq. 8 can be used to integrate (via a numerical method such as Legendre-Gauss quadrature) it out from the above:

$$E\left(z^*|\boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right) = E\left(E\left(z^*|\boldsymbol{\theta}, \boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right)\right), \text{ Eq. 20}$$

$$cov\left(z^*, z'|\boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right) = E\left(cov\left(z^*, z'|\boldsymbol{\theta}, \boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right)\right) +$$
$$cov\left(E\left(z^*|\boldsymbol{\theta}, \boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right), E\left(z'|\boldsymbol{\theta}, \boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right)\right), \text{ Eq. 21}$$

The posterior of the discrepancy function can be found similarly:

$$E\left(\delta(\boldsymbol{x}^*)|\boldsymbol{\theta}, \boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right) = [0\ 1] \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{T}_2^T(\boldsymbol{x}^*, \boldsymbol{\theta}) \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right), \text{ Eq. 22}$$

$$cov\left(\delta(\boldsymbol{x}^*), \delta(\boldsymbol{x}')|\boldsymbol{\theta}, \boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right) = c_2(\boldsymbol{x}^*, \boldsymbol{x}') - \boldsymbol{T}_2^T(\boldsymbol{x}^*, \boldsymbol{\theta}) \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{T}_2(\boldsymbol{x}', \boldsymbol{\theta}) +$$
$$\left([0\ 1]^T - \boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{T}_2(\boldsymbol{x}^*, \boldsymbol{\theta})\right)^T (\boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{H})^{-1} \left([0\ 1]^T - \boldsymbol{H}^T \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \boldsymbol{T}_2(\boldsymbol{x}', \boldsymbol{\theta})\right), \text{ Eq. 23}$$

Where $\boldsymbol{T}_2(\boldsymbol{x}^*, \boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{0}_{p\times 1} \\ \boldsymbol{g}_{22}(\boldsymbol{x}^*) \end{bmatrix}$, and $\boldsymbol{g}_{22}(\boldsymbol{x}^*)$ is a column vector of size $q \times 1$ with j^{th} element $c_2(\boldsymbol{x}_j, \boldsymbol{x}^*)$.

And that of the computer model:

$$E\left(\eta(\boldsymbol{x}^*)|\boldsymbol{\theta}, \boldsymbol{d}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2, \hat{\rho}, \hat{\lambda}\right) = [1\ 0] \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{T}_1^T(\boldsymbol{x}^*, \boldsymbol{\theta}) \boldsymbol{V_d}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right), \text{ Eq. 22}$$

ONE OF THE EXISTING TECHNOLOGIES

Consider the method of Kennedy and O'Hagan:

$$y_h(\boldsymbol{x}) = \rho y_l(\boldsymbol{x}, \boldsymbol{\theta}) + \delta(\boldsymbol{x}, \boldsymbol{\theta}) + \varepsilon$$

Inputs: $\boldsymbol{x}$ Calibration parameters: $\boldsymbol{\theta}$ scaling factor: ρ

High-fidelity source: y_h low-fidelity source: y_l noise: ε

Limitations:

1. Relies on explicit formulations
2. Accommodates only 2 sources
3. Does not provide simple and visualizable discrepancy metric

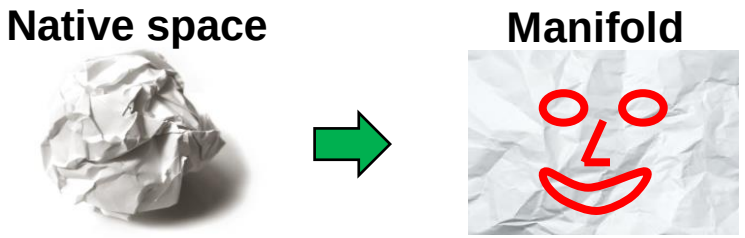
Neural networks inspired by such approaches have similar (& more) limitations.

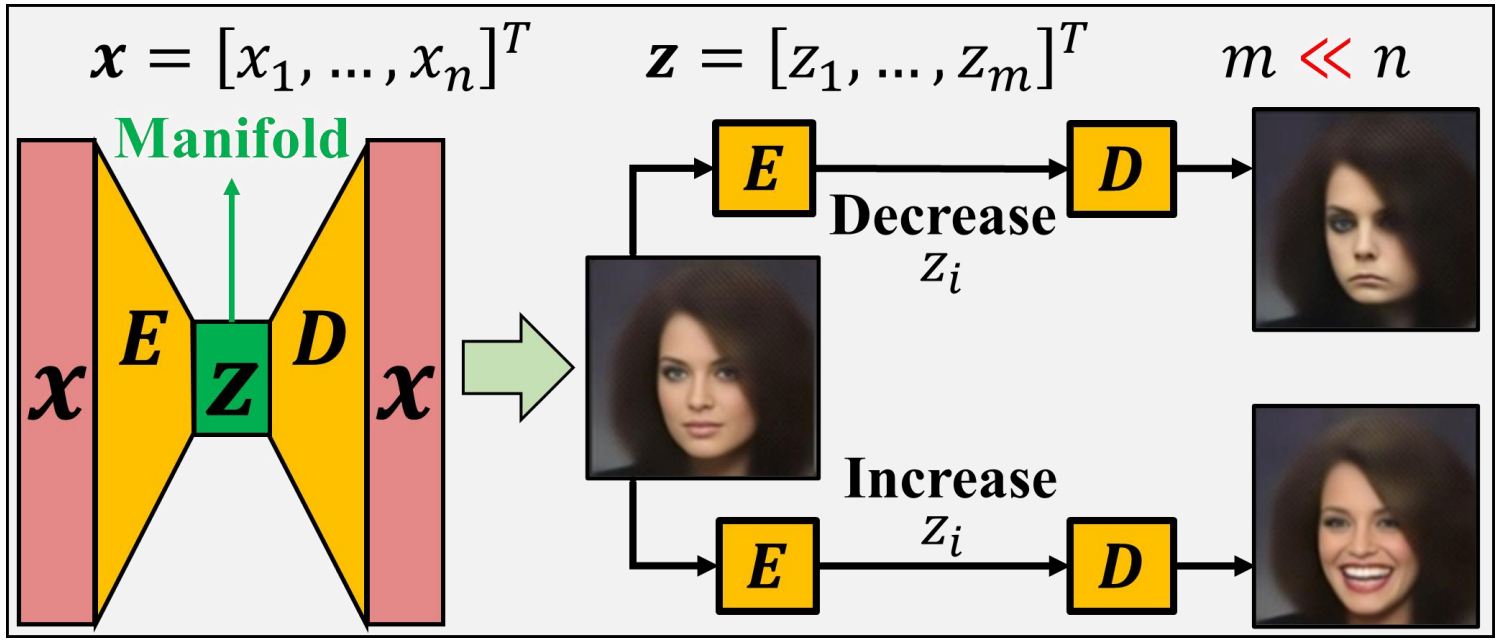
Kennedy, Marc C, and Anthony O'Hagan. "Bayesian Calibration of Computer Models." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63, no. 3 (2001): 425-64.

PROPOSED APPROACH

Our hypothesis: Complementary complex information are easier to fuse in **learnt nonlinear manifolds** (aka latent spaces).

Eweis-Labolle, J.T., N. Oune, and R. Bostanabad, Data Fusion With Latent Map Gaussian Processes. Journal of Mechanical Design, 2022. 144(9).





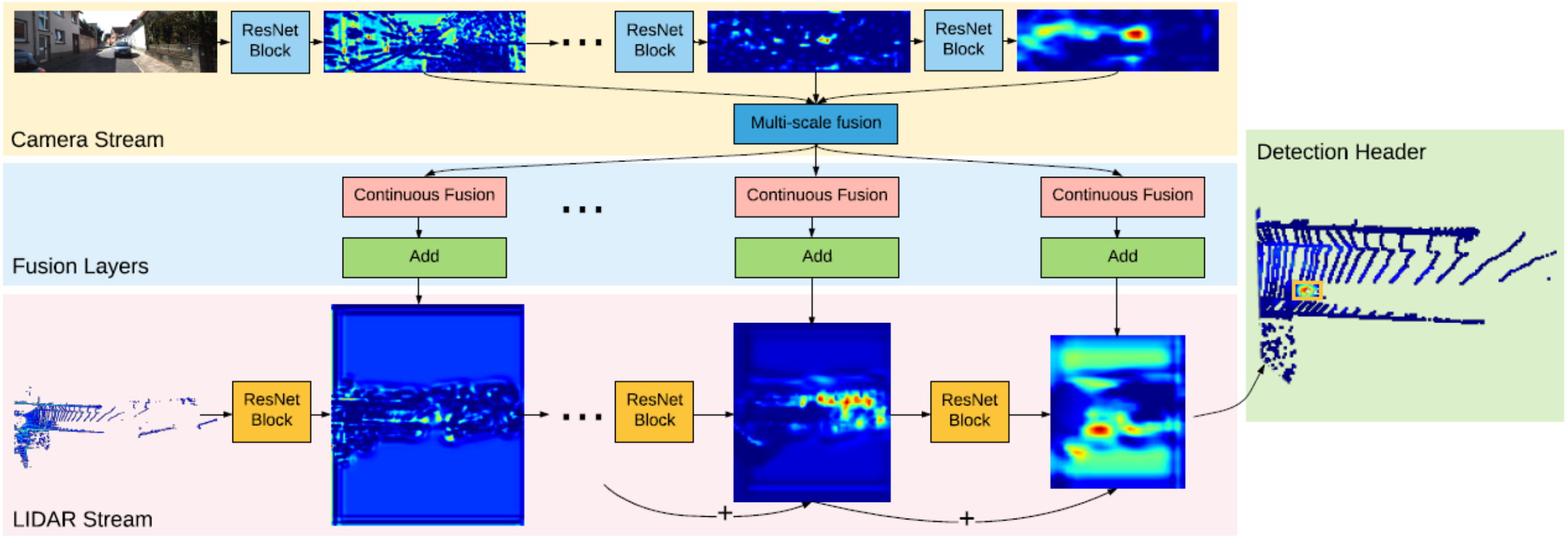
1. Coifman, R.R. and S. Lafon, Diffusion maps. Applied and Computational Harmonic Analysis, 2006. 21(1): p. 5-30.
2. Schmidt, M. and H. Lipson, Distilling free-form natural laws from experimental data. Science, 2009. 324(5923): p. 81-5.
3. Tenenbaum, J.B., V. de Silva, and J.C. Langford, A global geometric framework for nonlinear dimensionality reduction. Science, 2000. 290(5500): p. 2319-23.
4. Donoho, D.L. and C. Grimes, Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. Proc Natl Acad Sci U S A, 2003. 100(10): p. 5591-6.
5. Lawrence, N., Probabilistic non-linear principal component analysis with Gaussian process latent variable models. Journal of Machine Learning Research, 2005. 6(Nov): p. 1783-1816.

MULTISCALE DATA FUSION IN AUTONOMOUS CARS

Fusion is achieved via scale-wise feature <u>concatenation</u> (+ some operations)

Advantages: Intuitive, simple, amenable to pre-training, scalable, …

Disadvantages: Iterative, big (reliable) data, expensive to train, …

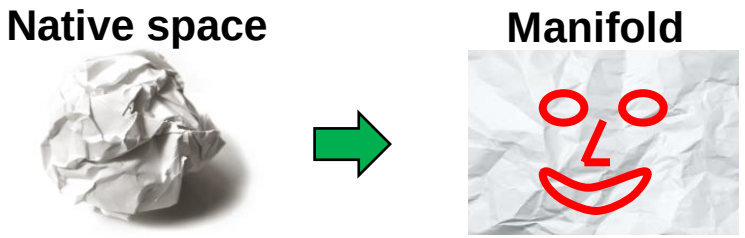


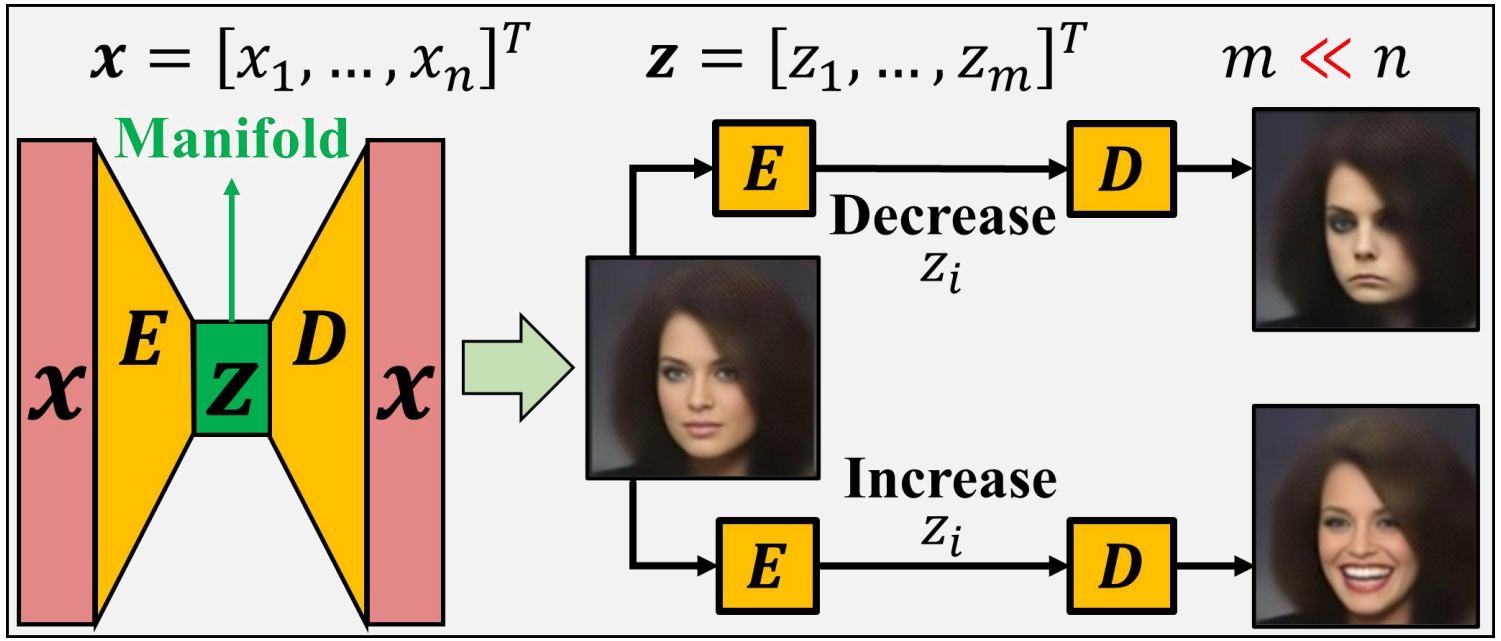
Liang, M., et al. Deep continuous fusion for multi-sensor 3d object detection. in Proceedings of the European conference on computer vision (ECCV). 2018.

PROPOSED APPROACH

Our hypothesis: Complementary complex information are easier to fuse in **learnt nonlinear manifolds** (aka latent spaces).

Eweis-Labolle, J.T., N. Oune, and R. Bostanabad, Data Fusion With Latent Map Gaussian Processes. Journal of Mechanical Design, 2022. 144(9).





How can we assimilate data on learnt manifolds while

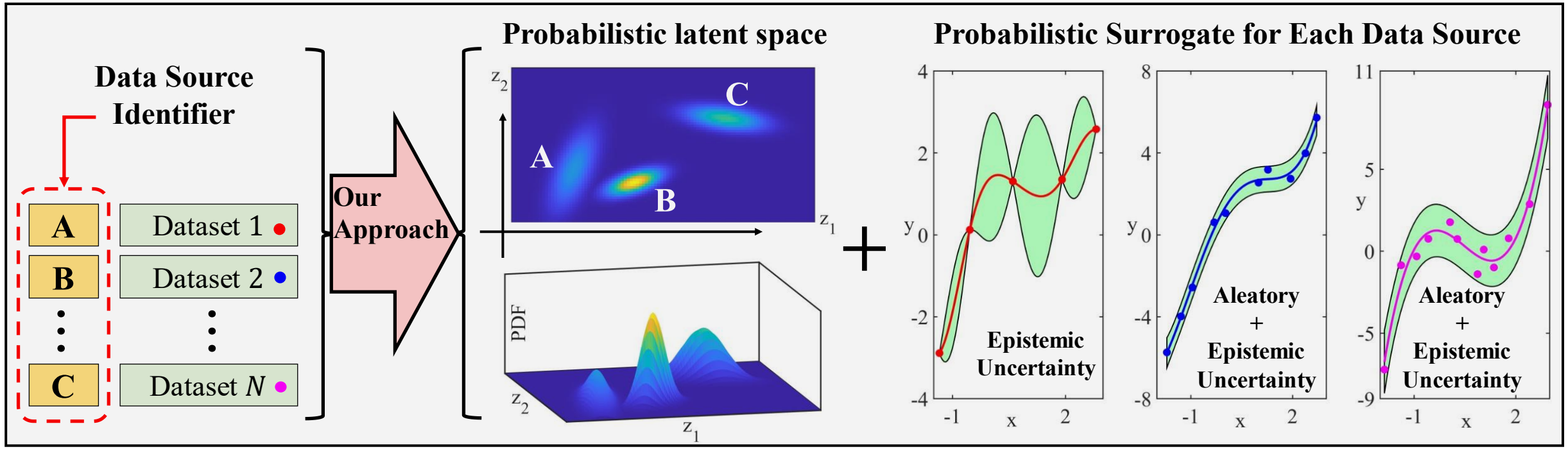
- Having small data
- Building emulators
- Minimizing architecture iterations
- Visualizing source discrepancies

EXPECTED RESULTS IN A GLANCE

Joint fusion of multiple models, multiple dataset.

Significance

- Visualize the (dis)similarity among data sources.
- Increased accuracy, flexibility, and modularity



Eweis-Labolle, J.T., N. Oune, and R. Bostanabad, Data Fusion With Latent Map Gaussian Processes. Journal of Mechanical Design, 2022. 144(9).

GAUSSIAN PROCESSES (GPs)

Completely characterized by mean and covariance functions:

$$Y(\boldsymbol{x}) \sim GP(m(\boldsymbol{x}), c(\boldsymbol{x}, \boldsymbol{x}'))$$

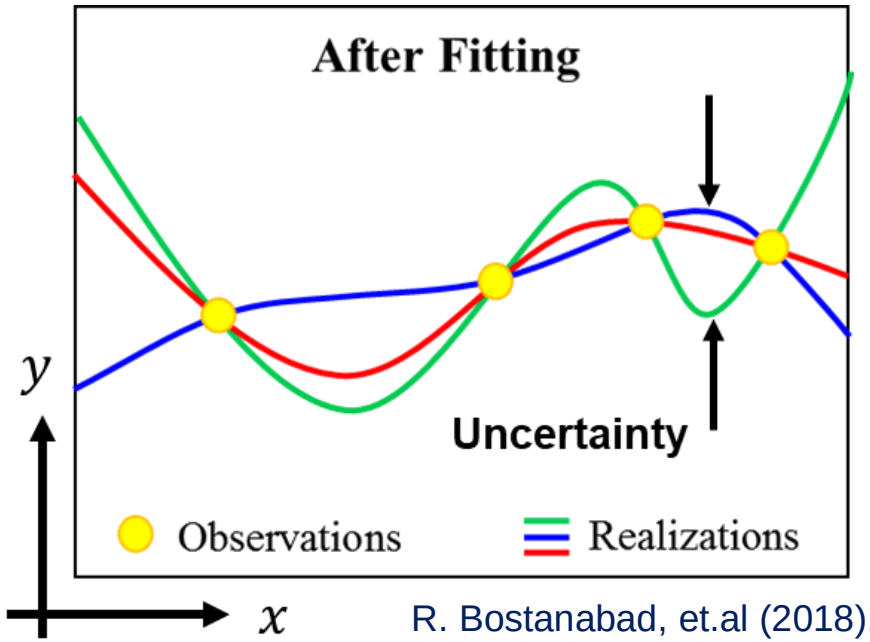
Mean Function

Covariance Function

Parametric forms for **mean** and **covariance** functions:

$$m(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})\boldsymbol{\beta}$$

$$c(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left(-(\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{\Omega}(\boldsymbol{x} - \boldsymbol{x}')\right) = \sigma^2 r(\boldsymbol{x}, \boldsymbol{x}')$$



R. Bostanabad, et.al (2018). *Int J Numer Meth Eng* 114 501-516.

Prediction upon fitting:

$$\mathbb{E}[y^*] = \boldsymbol{f}(\boldsymbol{x}^*)\widehat{\boldsymbol{\beta}} + \boldsymbol{g}^T(\boldsymbol{x}^*)\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}\right)$$

$$cov(y^*, y') = c(\boldsymbol{x}^*, \boldsymbol{x}') - \boldsymbol{g}^T(\boldsymbol{x}^*)\boldsymbol{V}^{-1}\boldsymbol{g}(\boldsymbol{x}') + \boldsymbol{h}(\boldsymbol{x}^*)^T(\boldsymbol{F}^T\boldsymbol{V}^{-1}\boldsymbol{F})^{-1}\boldsymbol{h}(\boldsymbol{x}')$$

Planas, R., N. Oune, and R. Bostanabad, Evolutionary Gaussian Processes. Journal of the Mechanical Design, 2021.

GPS: THE GOOD, THE BAD, THE UGLY

Advantages: Uncertainty quantification, interpretable, fast, robust to overfitting and noise, …

Disadvantages: Scalability, non-Gaussian uncertainties, reversion to the mean, handling categorical/qualitative inputs, …

Why focus on this issue?

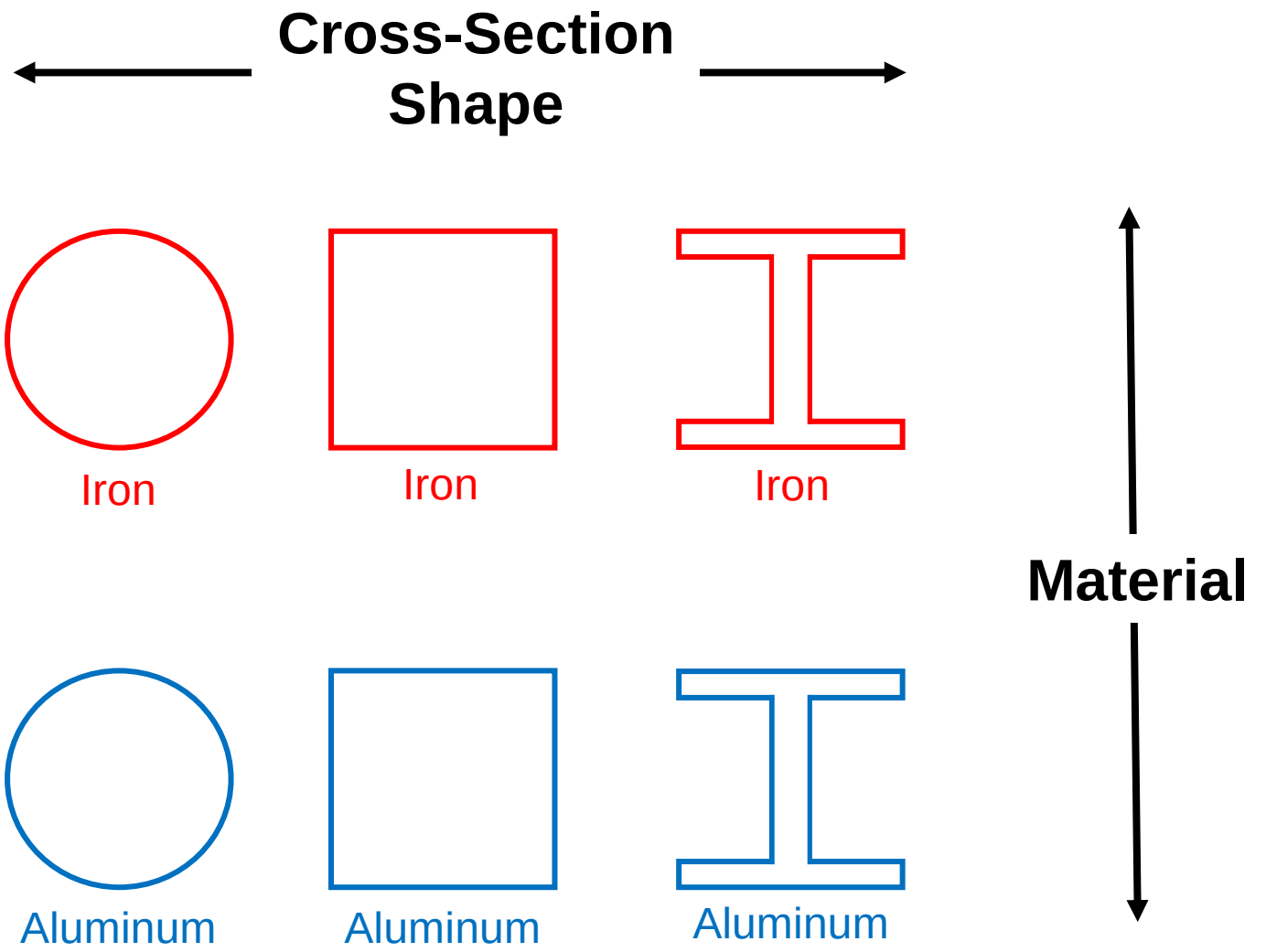
1. Planas, R., N. Oune, and R. Bostanabad, Evolutionary Gaussian Processes. Journal of the Mechanical Design, 2021.
2. Bostanabad, R., et al. "Leveraging the Nugget Parameter for Efficient Gaussian Process Modeling." International Journal for Numerical Methods in Engineering 114.5 (2018): 501-16. Print.
3. Bostanabad, R., Y.C. Chan, L.W. Wang, P. Zhu, and W. Chen, Globally approximate gaussian processes for big data with application to data-driven metamaterials design. Journal of Mechanical Design, 2019. 141(11).
4. Zhang, Yichi, Siyu Tao, Wei Chen, and Daniel W. Apley. "A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors.". Technometrics 62, no. 3 (Aug 17 2019): 291-302.

MIXED DATASET: BREAK DOWN OF TRADITIONAL GP

Inputs: $\boldsymbol{w} = [\boldsymbol{x}, \boldsymbol{t}]^T$ Output

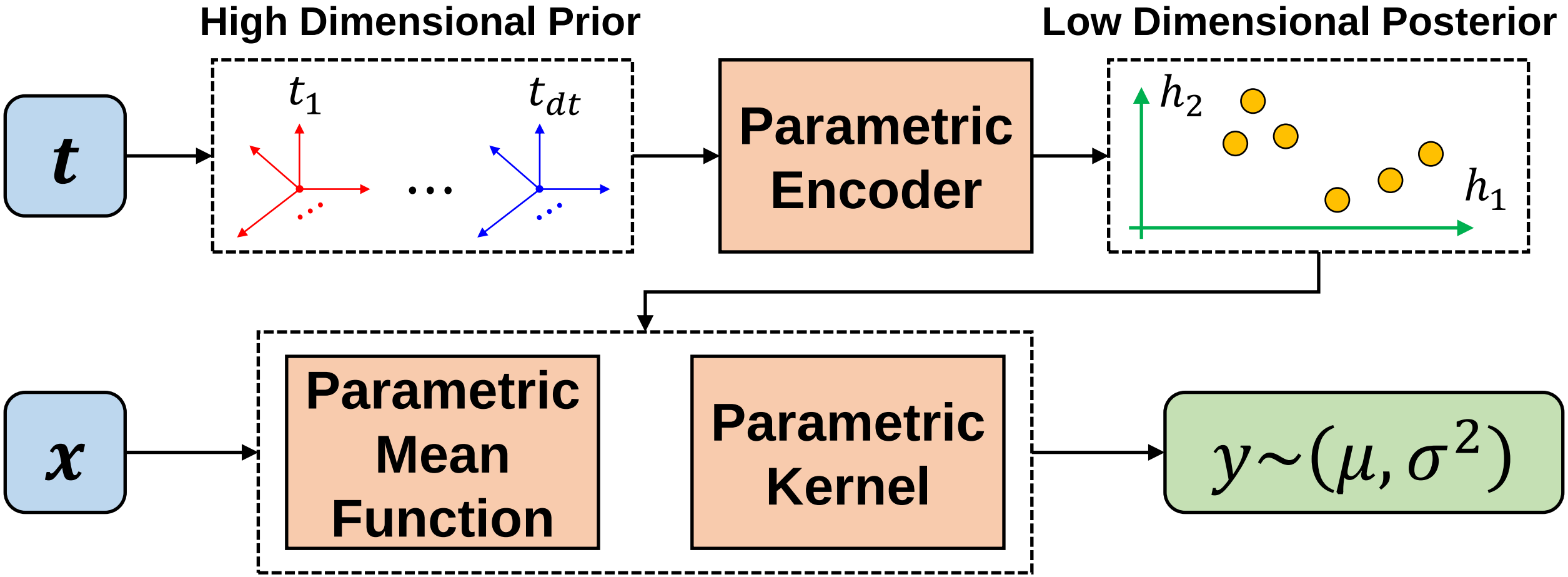
x_1	x_2	t_1	t_2	y
200	15	AL	Disk	1.2
219	21	Fe	Disk	2.5
193	52	Fe	Sq.	10
236	9	Fe	I	11.9
301	22	AL	I	3.1
202	42	Fe	Sq.	0.9

⋮



Oune, N., & Bostanabad, R. (2021). Latent map Gaussian processes for mixed variable metamodeling. Computer Methods in Applied Mechanics and Engineering, 387, 114128.

LATENT MAP GAUSSIAN PROCESS (LMGP)



- **All** parameters are **jointly** learned via MLE or MAP. **Or** fully Bayesian!
- Can have multiple posterior space (i.e., manifolds).

https://pmacslab.eng.uci.edu/

Oune, N., & Bostanabad, R. (2021). Latent map Gaussian processes for mixed variable metamodeling. Computer Methods in Applied Mechanics and Engineering, 387, 114128.

LATENT MAP GP (LMGP) TRAINING: STEP 1

The **basic idea** is to:

1. Put a **quantitative prior** on each combination of categorical variables.

Initial Data Matrix

$$\overbrace{\begin{bmatrix} x_1^1 & x_2^1 & S & F \\ x_1^2 & x_2^2 & C & A \\ x_1^3 & x_2^3 & I & A \\ \vdots & \vdots & \vdots & \vdots \\ x_1^n & x_2^n & S & F \end{bmatrix}}^{\boldsymbol{x}\quad\quad\boldsymbol{t}}_{n\times 4} \Rightarrow \overbrace{\begin{bmatrix} \boldsymbol{x^1} & \xi_{SF} \\ \boldsymbol{x^1} & \xi_{CA} \\ \boldsymbol{x^1} & \xi_{IA} \\ \vdots & \vdots \\ \boldsymbol{x^1} & \xi_{SF} \end{bmatrix}}^{\boldsymbol{x}\quad\boldsymbol{\xi}}$$

Example:

Level	Shape	Material
1	Circle (C)	Iron (F)
2	Square (S)	Aluminum (A)
3	I-Beam (I)	

$$\Rightarrow \overbrace{\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}}^{\text{Shape}\quad\text{Material}}$$

Oune, N., & Bostanabad, R. (2021). Latent map Gaussian processes for mixed variable metamodeling. Computer Methods in Applied Mechanics and Engineering, 387, 114128.

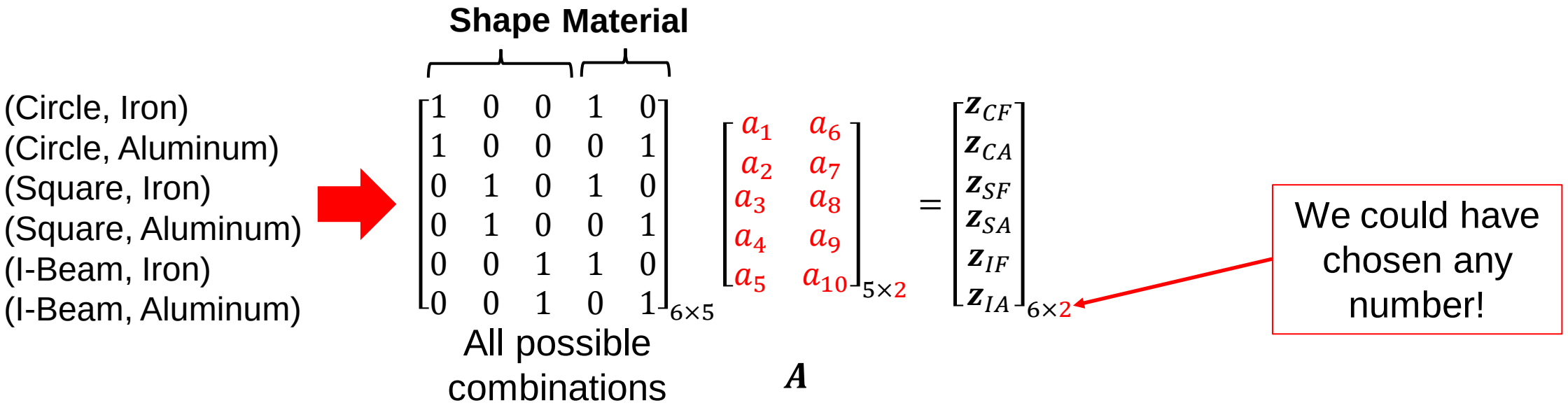
LATENT MAP GP (LMGP) TRAINING: STEP 2

The **basic idea** is to:

1. Put a **quantitative prior representation** on each combination of categorical variables.
2. Push the prior representations forward by a **parametric function**.

   For example:

$$\left(\boldsymbol{A}_{p\times q}\right)^{T}\boldsymbol{\xi}_{p\times 1}=\boldsymbol{z}_{q\times 1}$$



Oune, N., & Bostanabad, R. (2021). Latent map Gaussian processes for mixed variable metamodeling. Computer Methods in Applied Mechanics and Engineering, 387, 114128.

LATENT MAP GP (LMGP) TRAINING: STEP 3

The **basic idea** is to:

1. Put a **quantitative prior representation** on each combination of categorical variables.
2. Push the prior representations forward by a **parametric function**.
3. Estimate **all** parameters via MAP.

Shape **Material**

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}_{6\times5}}_{\text{All possible combinations}} \underbrace{\begin{bmatrix} a_1 & a_6 \\ a_2 & a_7 \\ a_3 & a_8 \\ a_4 & a_9 \\ a_5 & a_{10} \end{bmatrix}_{5\times2}}_{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{z}_{CF} \\ \boldsymbol{z}_{CA} \\ \boldsymbol{z}_{SF} \\ \boldsymbol{z}_{SA} \\ \boldsymbol{z}_{IF} \\ \boldsymbol{z}_{IA} \end{bmatrix}_{6\times2}$$

New Correlation Function

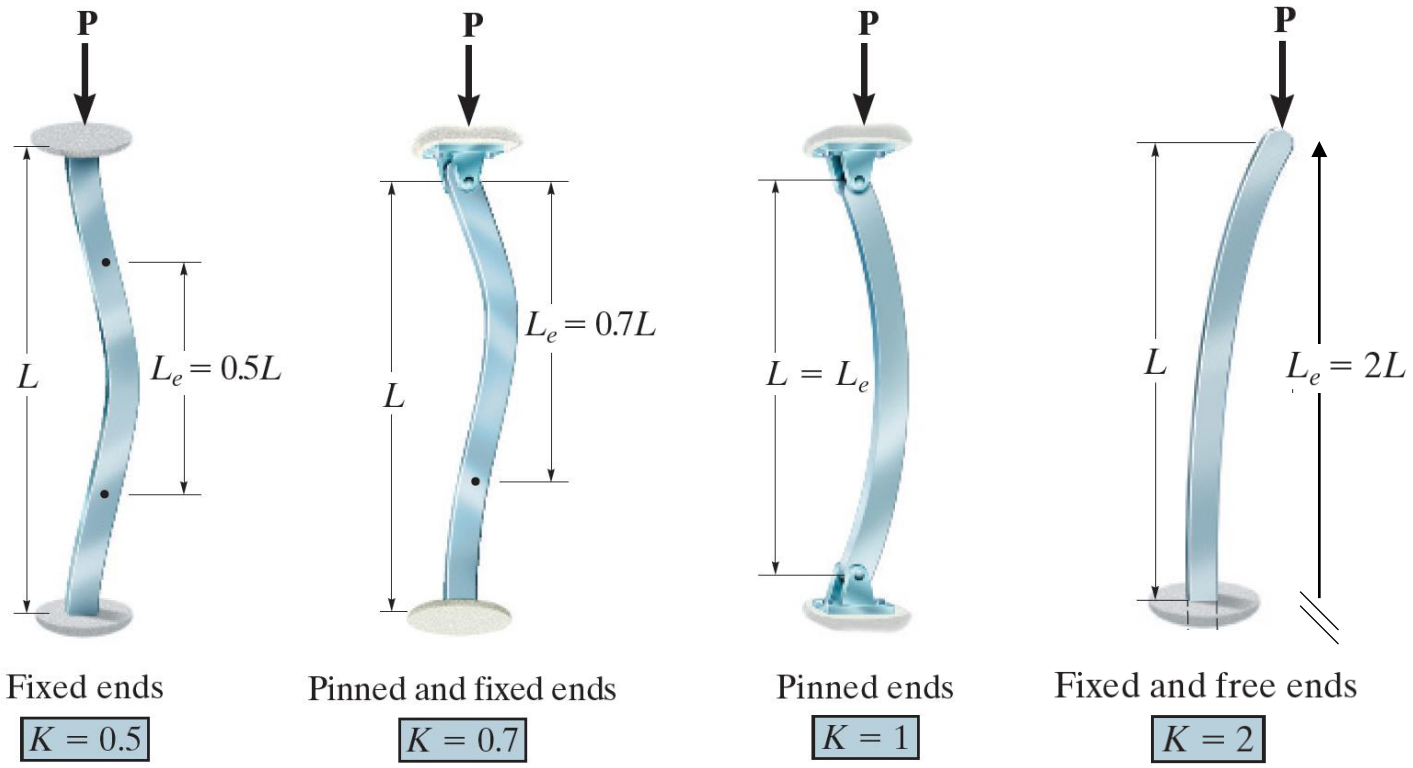
$$r(\boldsymbol{w}, \boldsymbol{w}') = exp\{-(\boldsymbol{x}-\boldsymbol{x}')^T \boldsymbol{\Omega} (\boldsymbol{x}-\boldsymbol{x}') - \|\boldsymbol{z}(\boldsymbol{t}) - \boldsymbol{z}(\boldsymbol{t}')\|_2^2\}$$

$$\boldsymbol{w} \equiv [\boldsymbol{x}, \boldsymbol{t}]^T$$

Oune, N., & Bostanabad, R. (2021). Latent map Gaussian processes for mixed variable metamodeling. Computer Methods in Applied Mechanics and Engineering, 387, 114128.

CONTROLLED EXPERIMENT: BUCKLING

- Euler's formula: $$P_{cr} = \frac{\pi^2 EI}{(KL)^2} = \frac{\pi^2 EI}{L_e^2}$$



Converted to Categorical:

$$I = \begin{cases} 12.1 \times 10^6 & mm^4 \\ 29.5 \times 10^6 & mm^4 \\ 9.41 \times 10^6 & mm^4 \end{cases}$$

$$K = \begin{cases} 0.5 \\ 0.7 \\ 1 \\ 2 \end{cases}$$

$$E = \begin{cases} 200 & GPa \\ 73.1 & GPa \end{cases}$$

Quantitative:

$$0.5 \leq L \leq 1.5 \quad m$$

- Generate 50 samples → LMGP does **NOT** know the underlying values.

Oune, N., & Bostanabad, R. (2021). Latent map Gaussian processes for mixed variable metamodeling. Computer Methods in Applied Mechanics and Engineering, 387, 114128.

SAMPLE DATASET

Numerical Inputs **Output**

L	E	K	I	P_{cr}
1.00	73.1	0.50	12.1	2.860
1.05	73.1	0.70	29.5	1.616
1.50	200.0	2.00	9.91	0.647
1.25	73.1	0.50	12.1	1.830
1.15	200.0	0.70	29.5	3.686
1.20	73.1	1.00	12.1	0.497
⋮	⋮	⋮	⋮	⋮
1.40	200.0	2.00	9.91	0.743

Mixed Inputs **Output**

L	E	K	I	P_{cr}
1.00	A	a	1	2.860
1.05	A	b	2	1.616
1.50	B	d	3	0.647
1.25	A	a	1	1.830
1.15	B	b	2	3.686
1.20	A	c	1	0.497
⋮	⋮	⋮	⋮	⋮
1.40	B	d	3	0.743

Input to LMGP

$$E = \begin{cases} 73.1 & Aluminum \rightarrow A \\ 200 & Steel \rightarrow B \end{cases}$$

$$K = \begin{cases} 0.5 & Fixed\ ends \rightarrow a \\ 0.7 & Fixed - pinned\ ends \rightarrow b \\ 1 & Pinned\ ends \rightarrow c \\ 2 & Fixed - free\ ends \rightarrow d \end{cases}$$

$$I = \begin{cases} 12.1 & I - Beam \rightarrow 1 \\ 29.5 & C - Channel \rightarrow 2 \\ 9.49 & Equal - leg\ Angle \rightarrow 3 \end{cases}$$

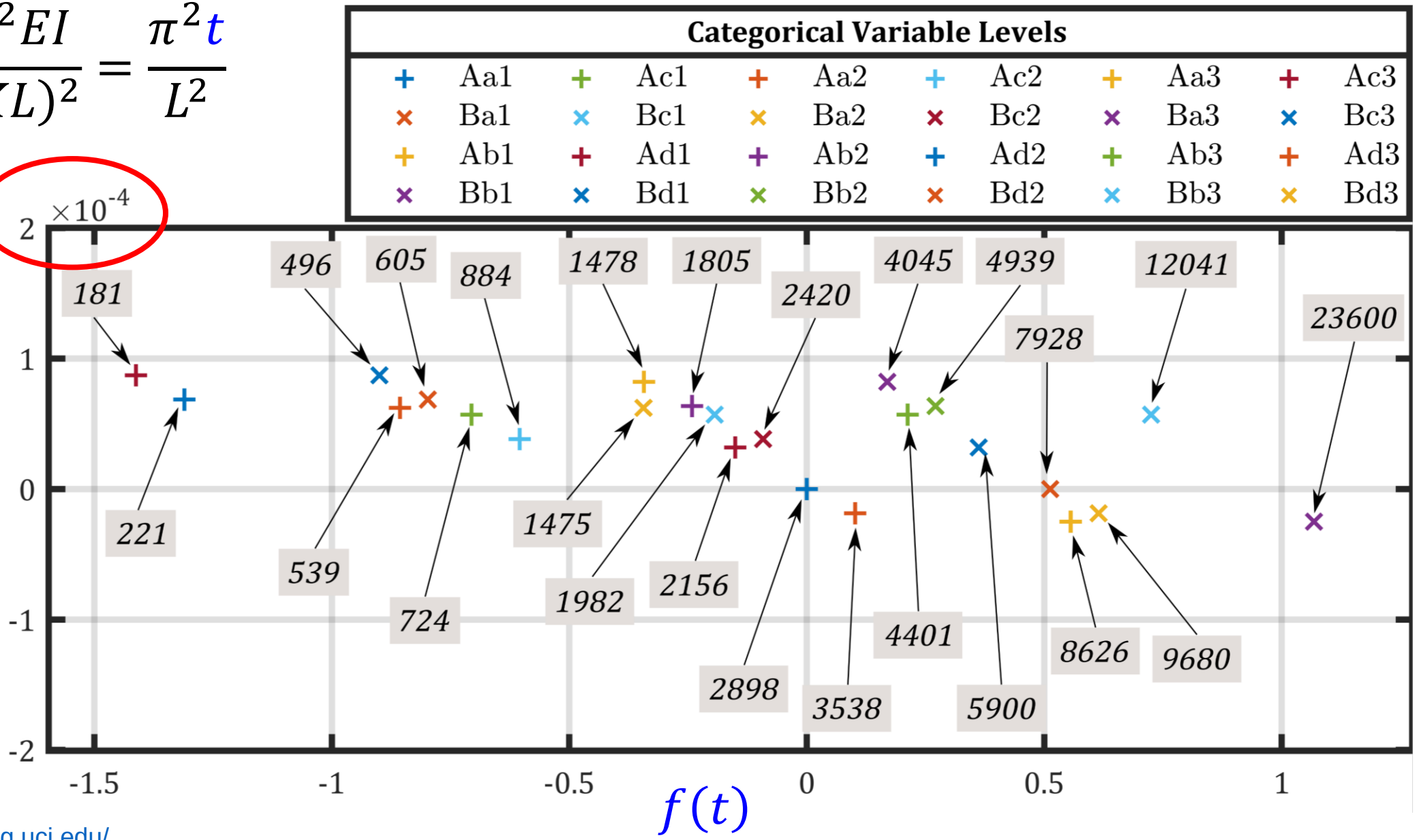
$$0.5 \leq L \leq 1.5$$

Oune, N., & Bostanabad, R. (2021). Latent map Gaussian processes for mixed variable metamodeling. Computer Methods in Applied Mechanics and Engineering, 387, 114128.

BUCKLING: LATENT SPACE

$$P_{cr} = \frac{\pi^2 EI}{(KL)^2} = \frac{\pi^2 t}{L^2}$$

SMALL!



MULTI-FIDELITY MODELING WITH LMGP

1 Augment Inputs with Data Source Index (assuming 3 sources)

$$[x_h] \Rightarrow [x_h \quad a]$$

$$[x_{l_1}] \Rightarrow [x_{l_1} \quad b]$$

$$[x_{l_2}] \Rightarrow [x_{l_2} \quad c]$$

2 Concatenate Datasets

$$x = \begin{bmatrix} x_h & a \\ x_{l_1} & b \\ x_{l_2} & c \end{bmatrix}$$

$$y = \begin{bmatrix} y_h \\ y_{l_1} \\ y_{l_2} \end{bmatrix}$$

3 Fit LMGP to the ALL data



Eweis-Labolle, J. T., Oune, N. & Bostanabad, R. Data Fusion With Latent Map Gaussian Processes. Journal of Mechanical Design 144, 1-41.

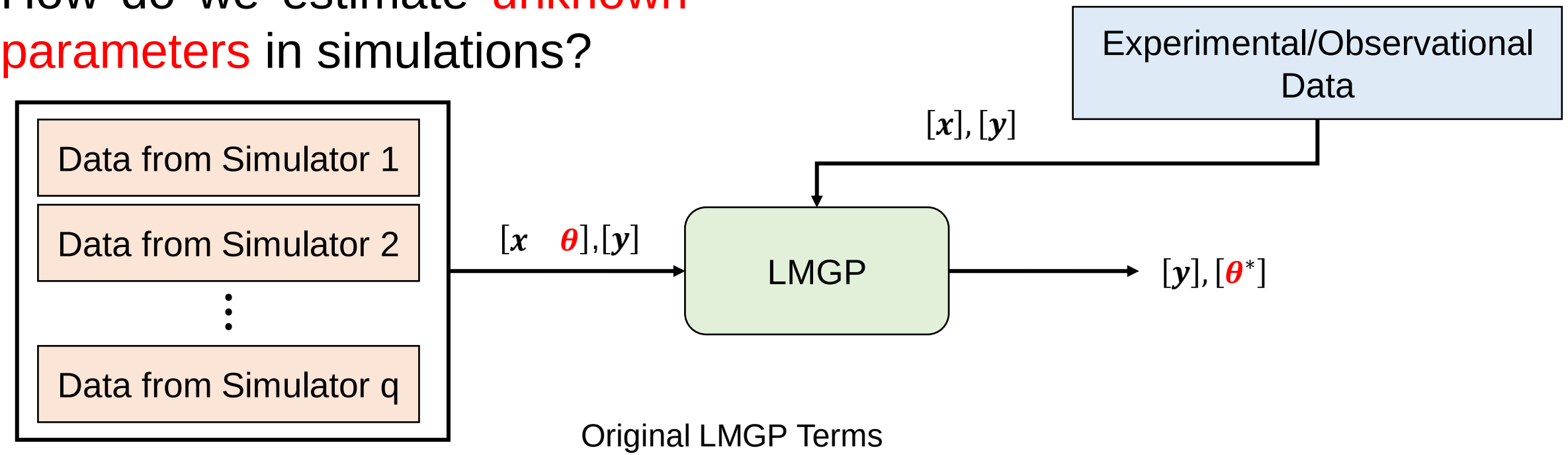
Why should this work?

Oune, N. and R. Bostanabad, Latent map Gaussian processes for mixed variable metamodeling. Computer Methods in Applied Mechanics and Engineering, 2021. 387: p. 114128.

Eweis-Labolle, J.T., N. Oune, and R. Bostanabad, Data Fusion With Latent Map Gaussian Processes. Journal of Mechanical Design, 2022. 144(9): p. 1-41.

MODEL CALIBRATION

How do we estimate unknown parameters in simulations?



Original LMGP Terms

$$r\left(\begin{bmatrix} x \\ t \\ \theta \end{bmatrix}^i, \begin{bmatrix} x \\ t \\ \theta \end{bmatrix}^j\right) = exp\left\{-\left(x^i - x^j\right)^T \Omega_x \left(x^i - x^j\right)\right\} \times exp\left\{-\left(z^i - z^j\right)^T \left(z^i - z^j\right)\right\} \times exp\left\{-\left(\theta^i - \theta^j\right)^T \Omega_\theta \left(\theta^i - \theta^j\right)\right\}$$

Quantitative Inputs

Source Identifier

Additional Term

Eweis-Labolle, J., N. Oune and R. Bostanabad (2022). "Data Fusion with Latent Map Gaussian Processes." Journal of Mechanical Design: 1-41.

OPTIMUM DATA COLLECTION

Design of M_2AX nanolaminate ternary alloys that exhibit many of the beneficial properties of both ceramic and metallic materials:

An early transition metal $M = \{Sc, Ti, V, Cr, Zr, Nb, Mo, Hf, Ta, W\} \rightarrow 10$
A main group element $A = \{Al, Si, P, S, Ga, Ge, As, Cd, In, Sn, Tl, Pb\} \rightarrow 12$
Either carbon or nitrogen $X = \{C, N\} \rightarrow 2$

$\rightarrow \mathbf{240}$

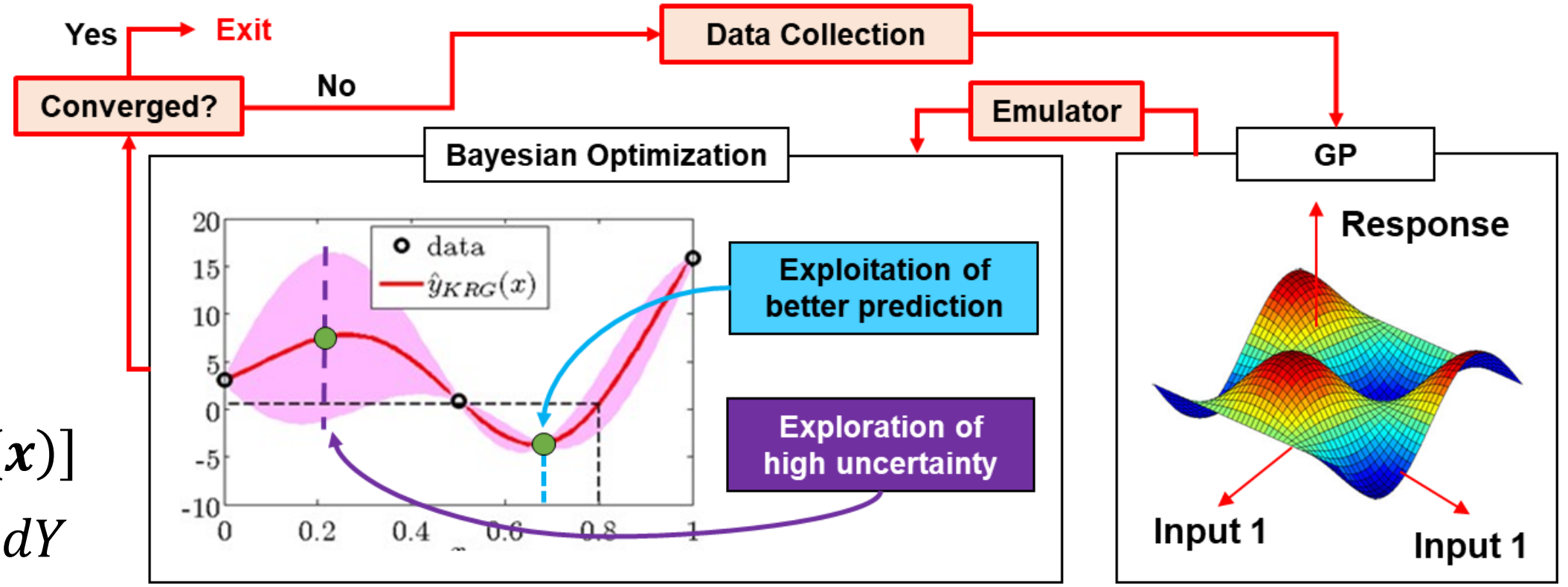
1. **How can we find the composition with the largest bulk modulus?**
2. **Can we use inaccurate but cheap simulations?**

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn

Zhou, Y.C., H.Y. Dong, X.H. Wang, and S.Q. Chen, Electronic structure of the layered ternary carbides ti2snc and ti2gec. Journal of Physics: Condensed Matter, 2000. 12(46): p. 9617.

BAYESIAN OPTIMIZATION (BO)

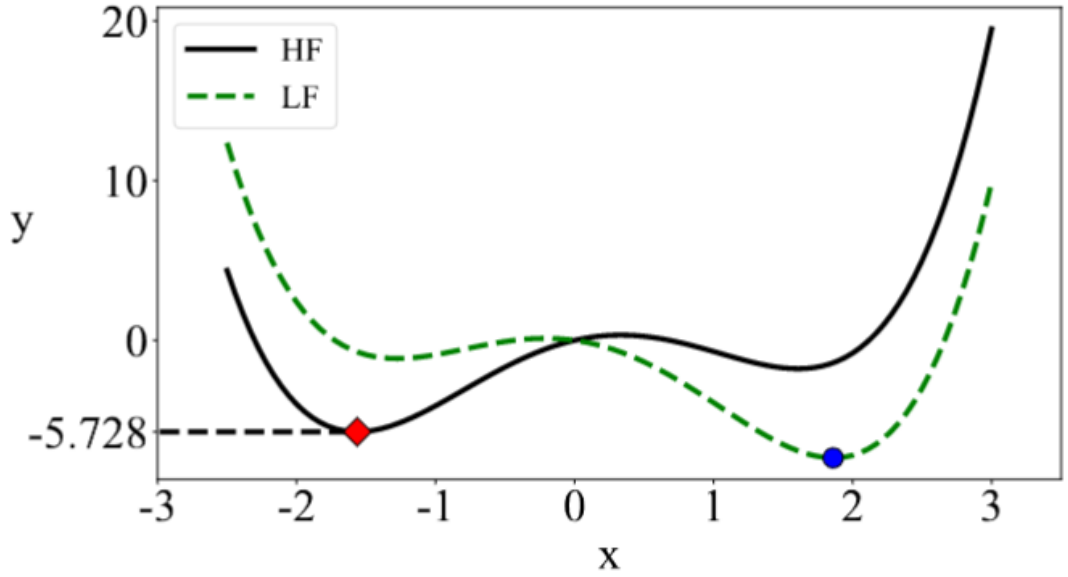
1. Fit an **emulator** to an initial training data.
2. Use the emulator in the **acquisition function** $\alpha(\boldsymbol{x})$ that balances exploration and exploitation to identify the next candidate.
3. Add the candidate to the data and repeat until convergence.

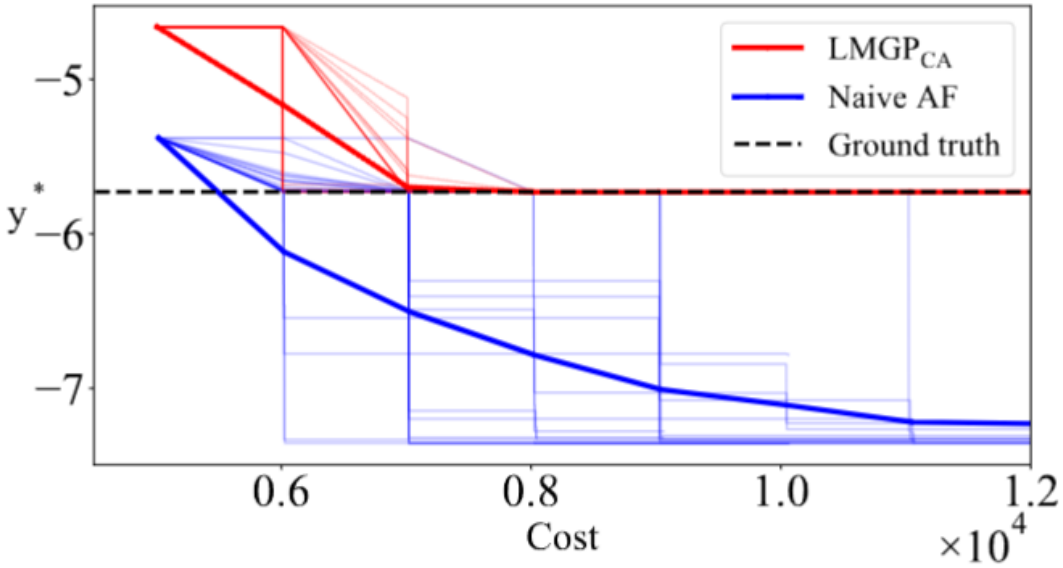


$$\alpha(\boldsymbol{x}) = \mathbb{E}[I(\boldsymbol{x})] = \int_{\mathbb{X}} I(\boldsymbol{x}) dY$$

CHALLENGES

1. Categorical inputs
2. Small data
3. Multi-fidelity
    - Low-fidelity sources have different optima!





Zanjani Foumani, Z., et al., Multi-fidelity cost-aware Bayesian optimization. Computer Methods in Applied Mechanics and Engineering, 2023. 407: p. 115937.

NEW ACQUISITION FUNCTION

Our idea: Expected utility of low- and high-fidelity data should be measured differently!

Low-fidelity: Exploration

High-fidelity: Exploitation

Low-fidelity: For a maximization problem, we define normalized improvement as:

$$\frac{y_j(\boldsymbol{x}) - \mu_j(\boldsymbol{x})}{\sigma_j(\boldsymbol{x})} = z \sim \mathcal{N}(0,1) \Rightarrow y_j(\boldsymbol{x}) - y_j^* = (\mu_j(\boldsymbol{x}) - y_j^*) + \sigma_j(\boldsymbol{x})z$$

$$I_{LF}(\boldsymbol{x}; j) = \begin{cases} \sigma_j(\boldsymbol{x})z & y_j(\boldsymbol{x}) > y_j^* \\ 0 & y_j(\boldsymbol{x}) \leq y_j^* \end{cases}$$

$$\alpha_{LF}(\boldsymbol{x}; j) = \int_{-\infty}^{\infty} I_{LF}(\boldsymbol{x}; j)\phi(z)\mathrm{d}z = \int_{-\infty}^{\infty} \sigma_j(\boldsymbol{x})z\phi(z)\mathrm{d}z = \sigma_j(\boldsymbol{x})\phi\left(\frac{y_j^* - \mu_j(\boldsymbol{x})}{\sigma_j(\boldsymbol{x})}\right)$$

Zanjani Foumani, Z., et al., Multi-fidelity cost-aware Bayesian optimization. Computer Methods in Applied Mechanics and Engineering, 2023. 407: p. 115937.

NEW ACQUISITION FUNCTION

Our idea: Expected utility of low- and high-fidelity data should be measured differently!

Low-fidelity: Exploration

High-fidelity: Exploitation

Low-fidelity: Use the exploration part of expected improvement (EI)

$$\alpha_{LF}(\boldsymbol{x}; j) = \sigma_j(\boldsymbol{x})\phi(\frac{y_j^* - \mu_j(\boldsymbol{x})}{\sigma_j(\boldsymbol{x})})$$

High-fidelity: Use probability of improvement (PI)

$$\alpha_{HF}(\boldsymbol{x}; l) = \Phi(\frac{\mu_l(\boldsymbol{x}) - y_l^*}{\sigma_l(\boldsymbol{x})})$$

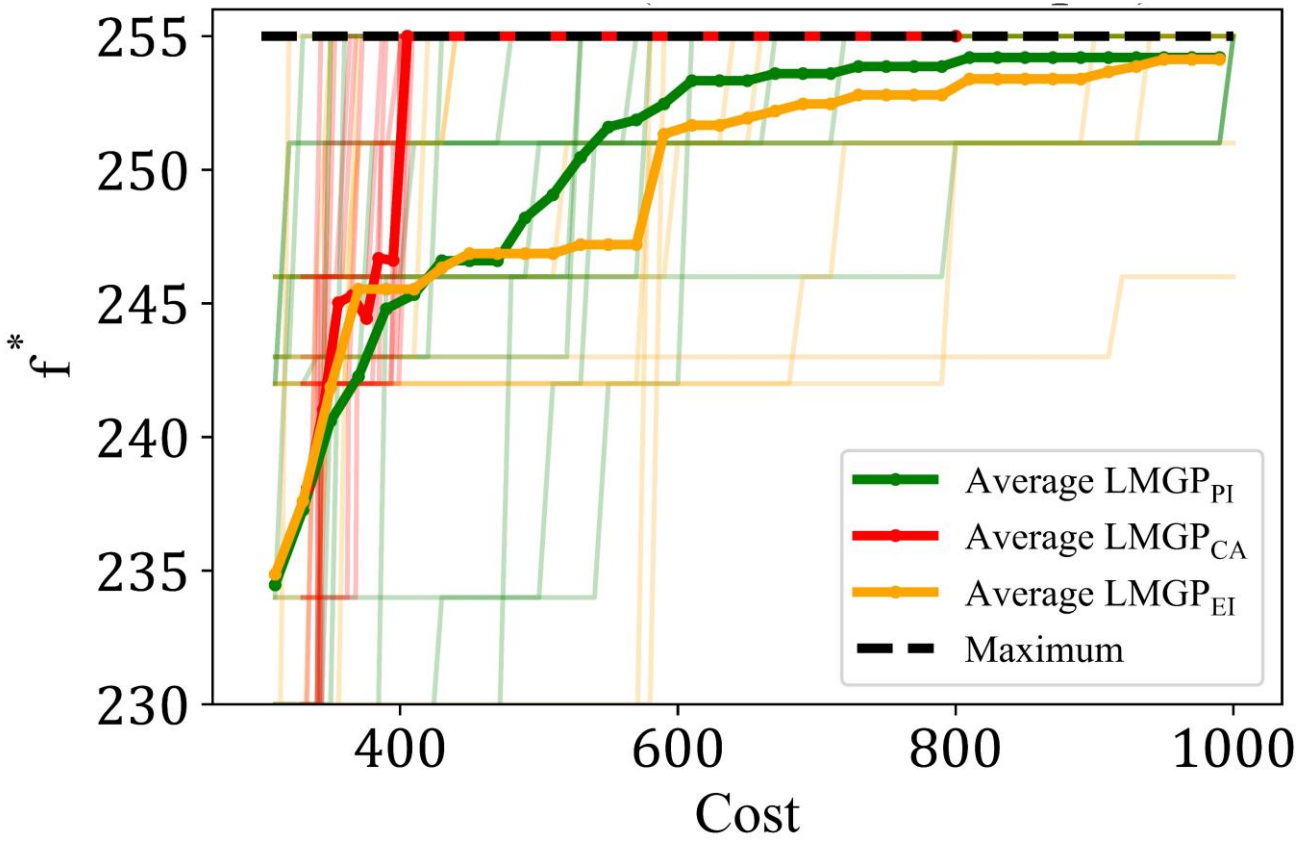
Final acquisition function (AF):

$$\alpha_{MFCA}(\boldsymbol{x}; j) = \begin{cases} \alpha_{LF}(\boldsymbol{x}; j)/O(j) & j = [1, \cdots, ds] \ \ \& \ \ j \neq l \\ \alpha_{HF}(\boldsymbol{x}; l)/O(l) & j = l \end{cases}$$

Zanjani Foumani, Z., et al., Multi-fidelity cost-aware Bayesian optimization. Computer Methods in Applied Mechanics and Engineering, 2023. 407: p. 115937.

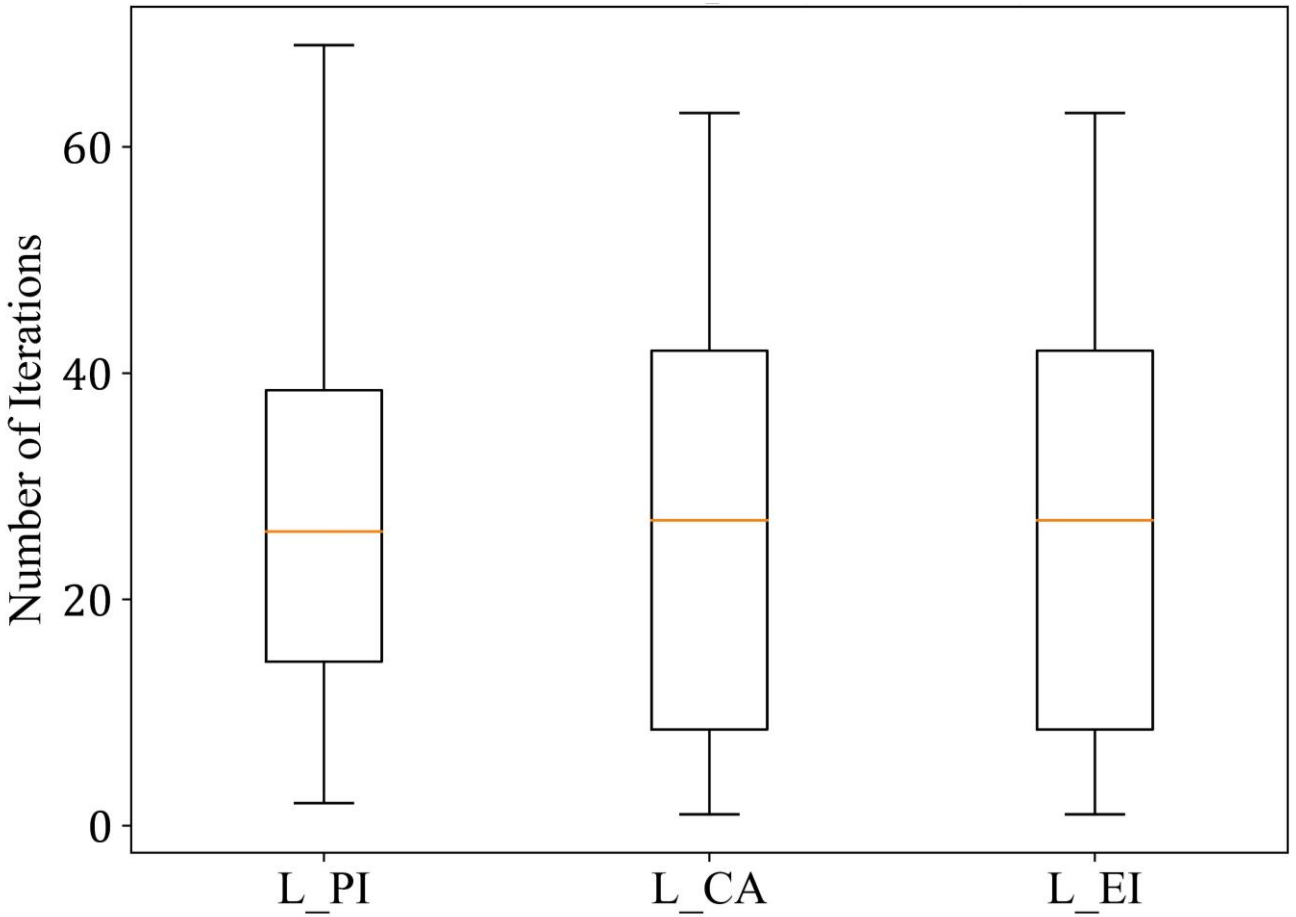
RESULTS: CONVERGENCE HISTORY

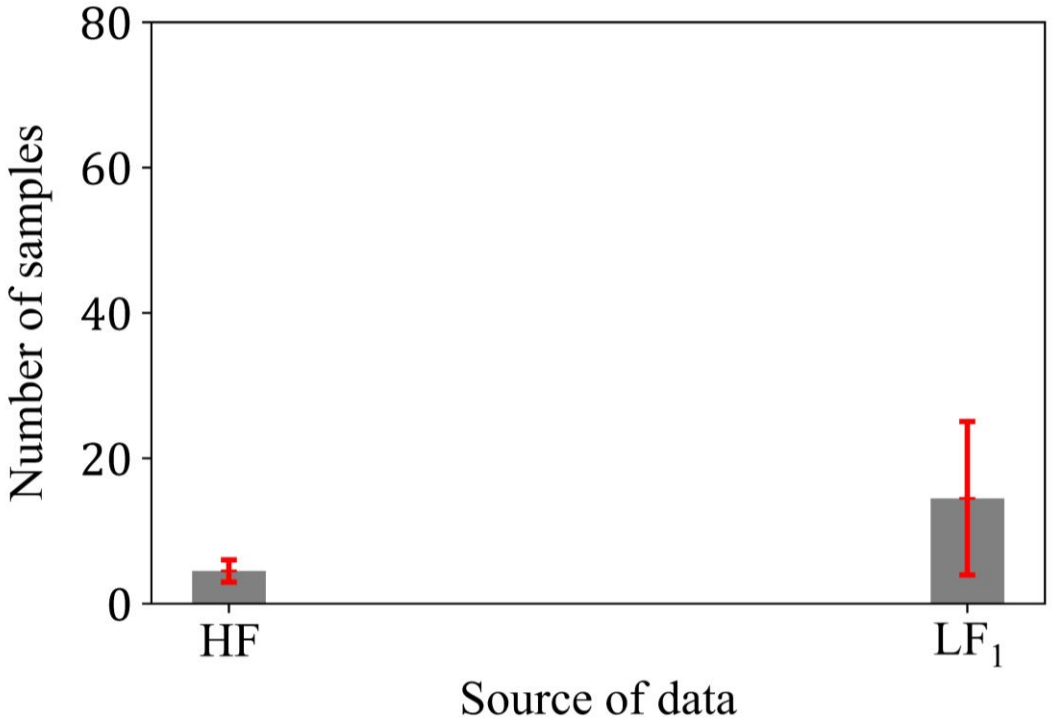
- High fidelity source is 10 times more expensive (10/1).
- BO is initialized with 5HF/30LF samples (15 repetitions).
  - Multi-fidelity BO has slightly higher initial cost.
- Multi-fidelity BO converges much faster.



RESULTS: USE OF LOW-FIDELITY SOURCE

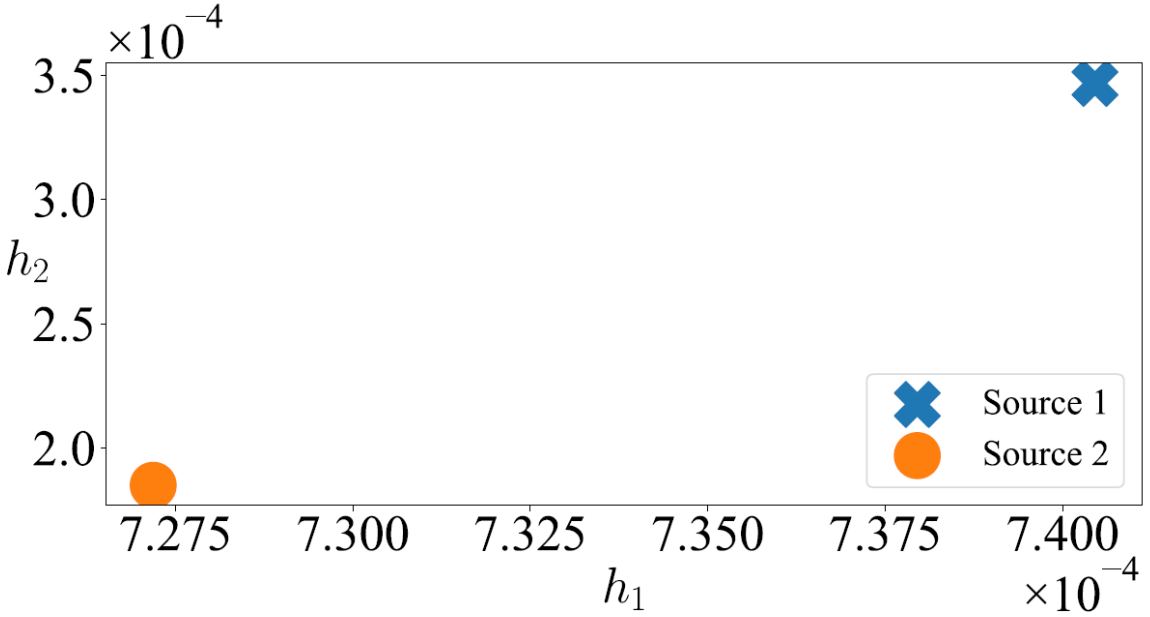
- Total number of iterations are comparable across different methods.
  - Our multi-fidelity BO samples more from low-fidelity source.



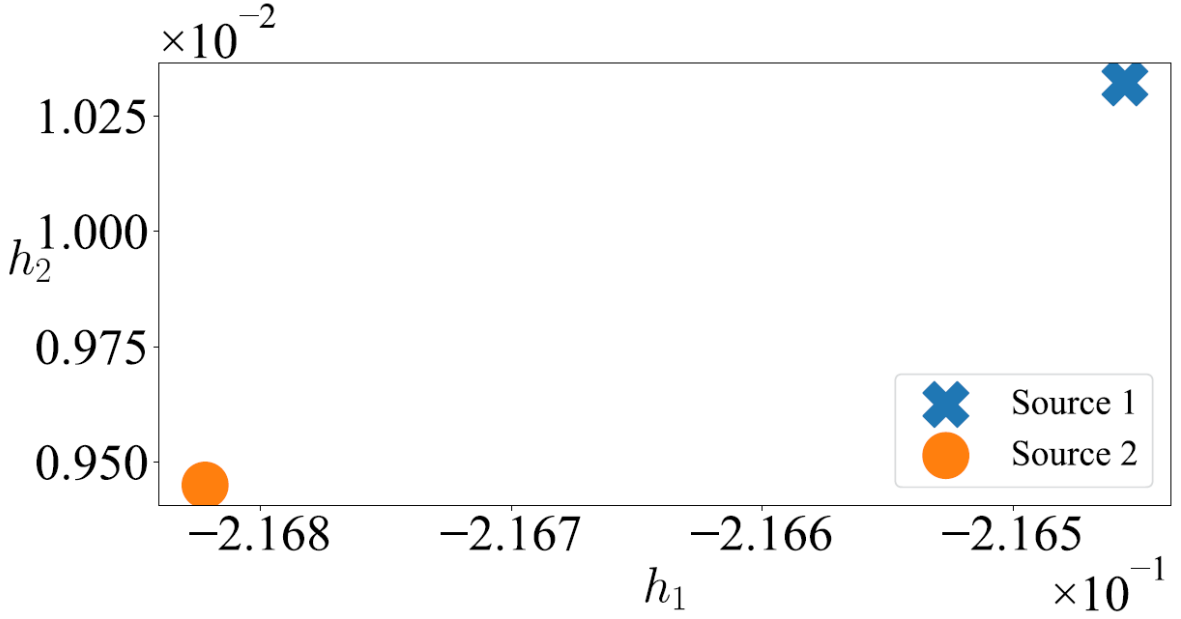


LEARNT MANIFOLDS: DATA SOURCE

At the end of BO



Using all the data



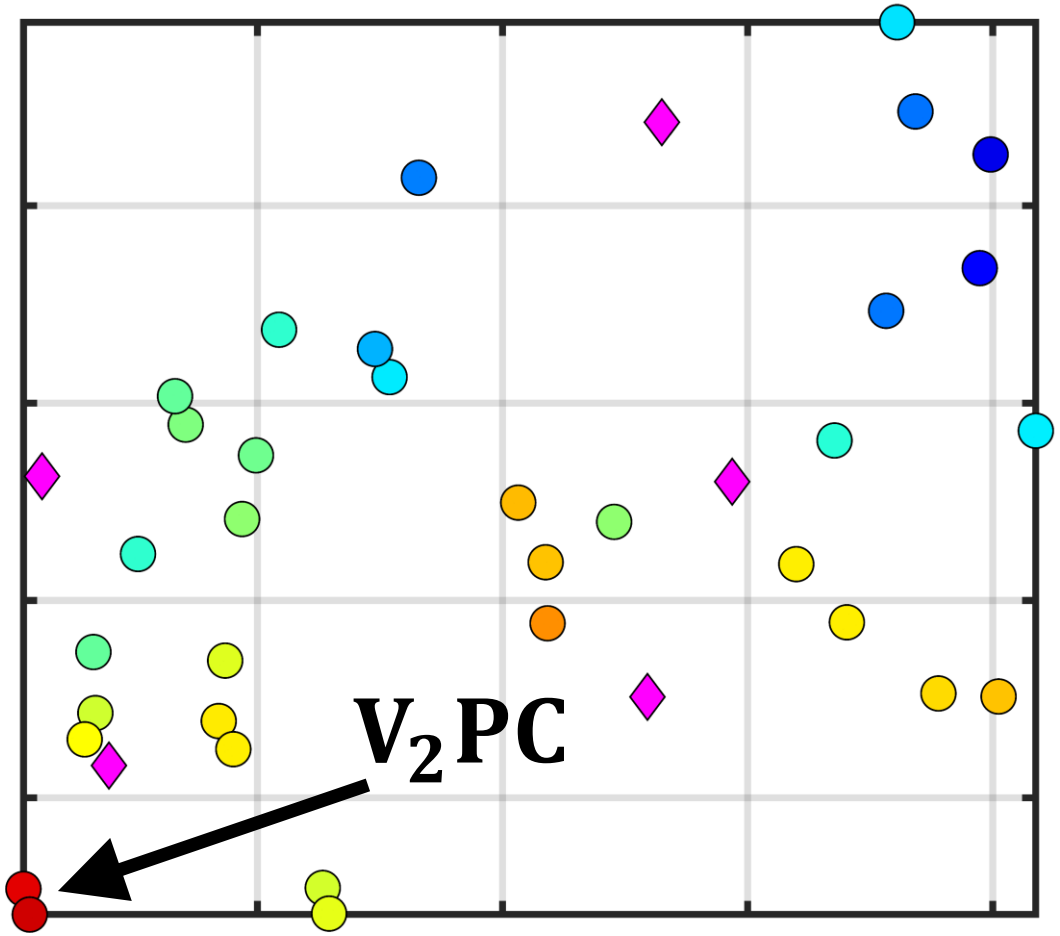
Cross-source correlation based on <u>nonlinearly</u> learnt latent variable:

$$\exp(-10^{-4}) \sim 1$$

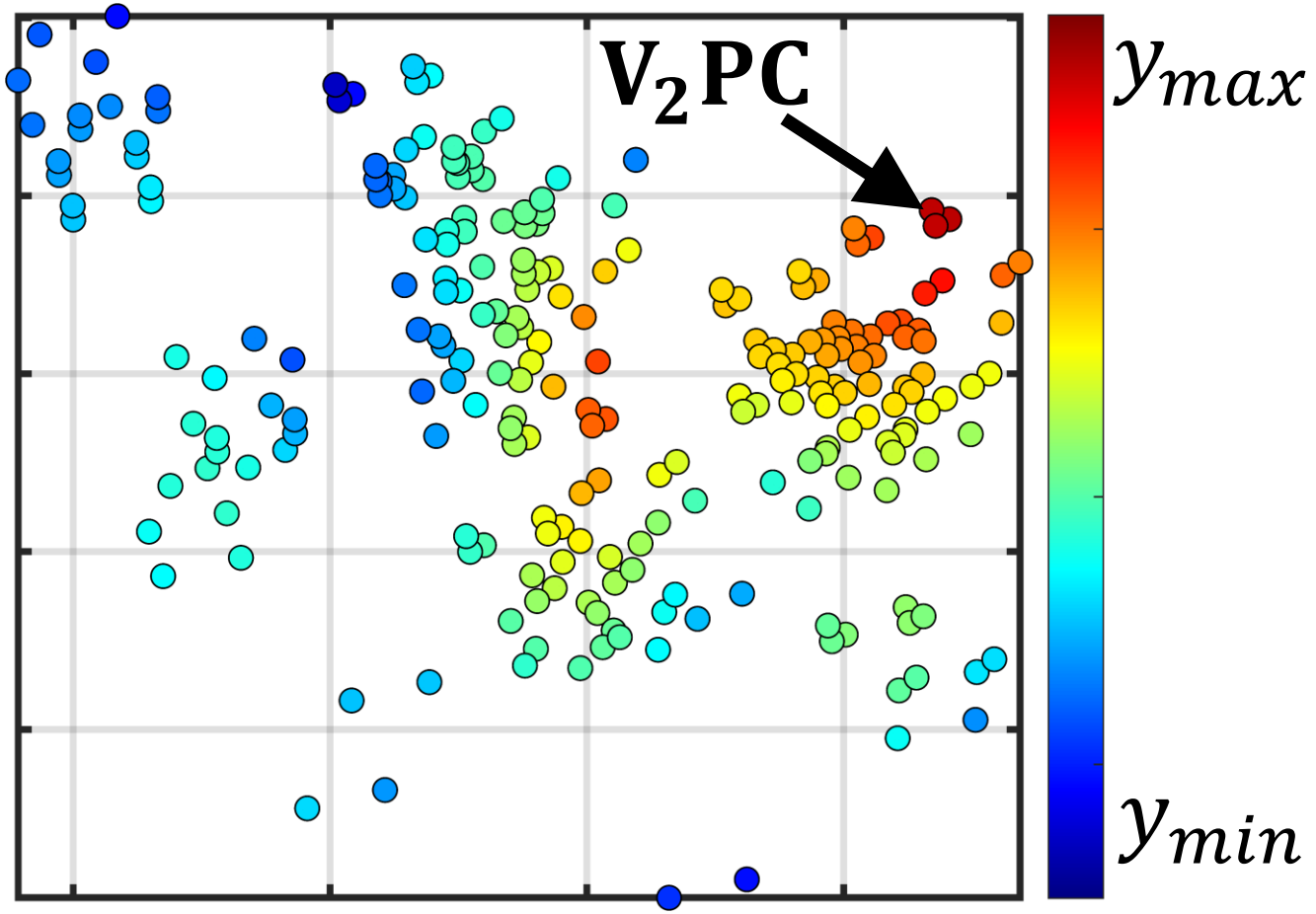
$$\exp(-10^{-2}) \sim 1$$

LEARNT MANIFOLDS: COMPOSITION

At the end of BO



Using all the data



https://pmacslab.eng.uci.edu/

Will Multi-fidelity BO outperform Single-fidelity BO?

Zanjani Foumani, Z., et al., Multi-fidelity cost-aware Bayesian optimization. Computer Methods in Applied Mechanics and Engineering, 2023. 407: p. 115937.

HYBRID ORGANIC-INORGANIC PEROVSKITE (HOIP)

HOIPs are crystals with the structural formula ABX_3:

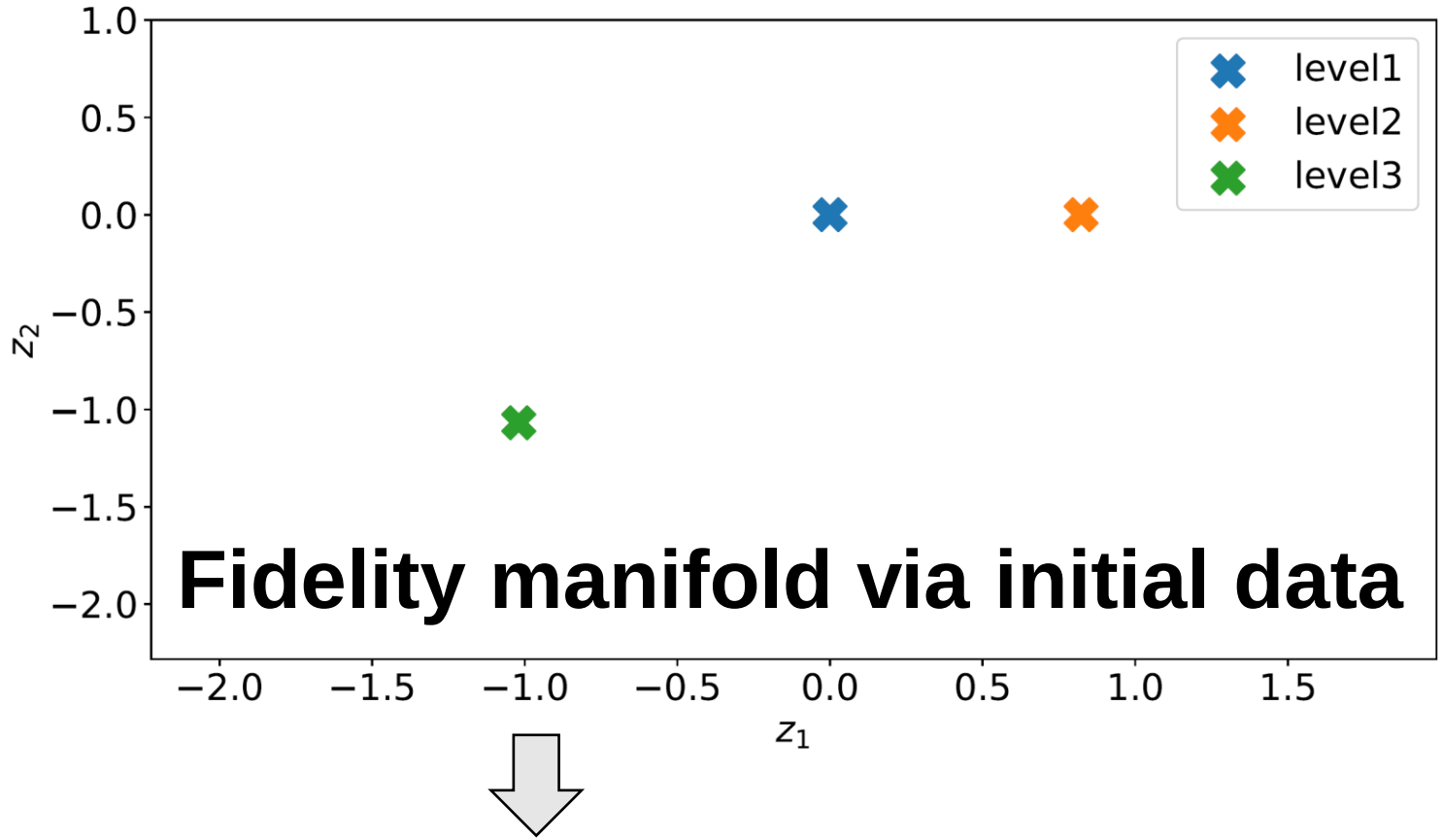
- $A \rightarrow 10$ levels
- $B \rightarrow 3$ levels
- $X \rightarrow 16$ levels

- 3 data sources (DFT)

Initial data: $15 - 20 - 15$

Cost: $15 - 10 - 5$

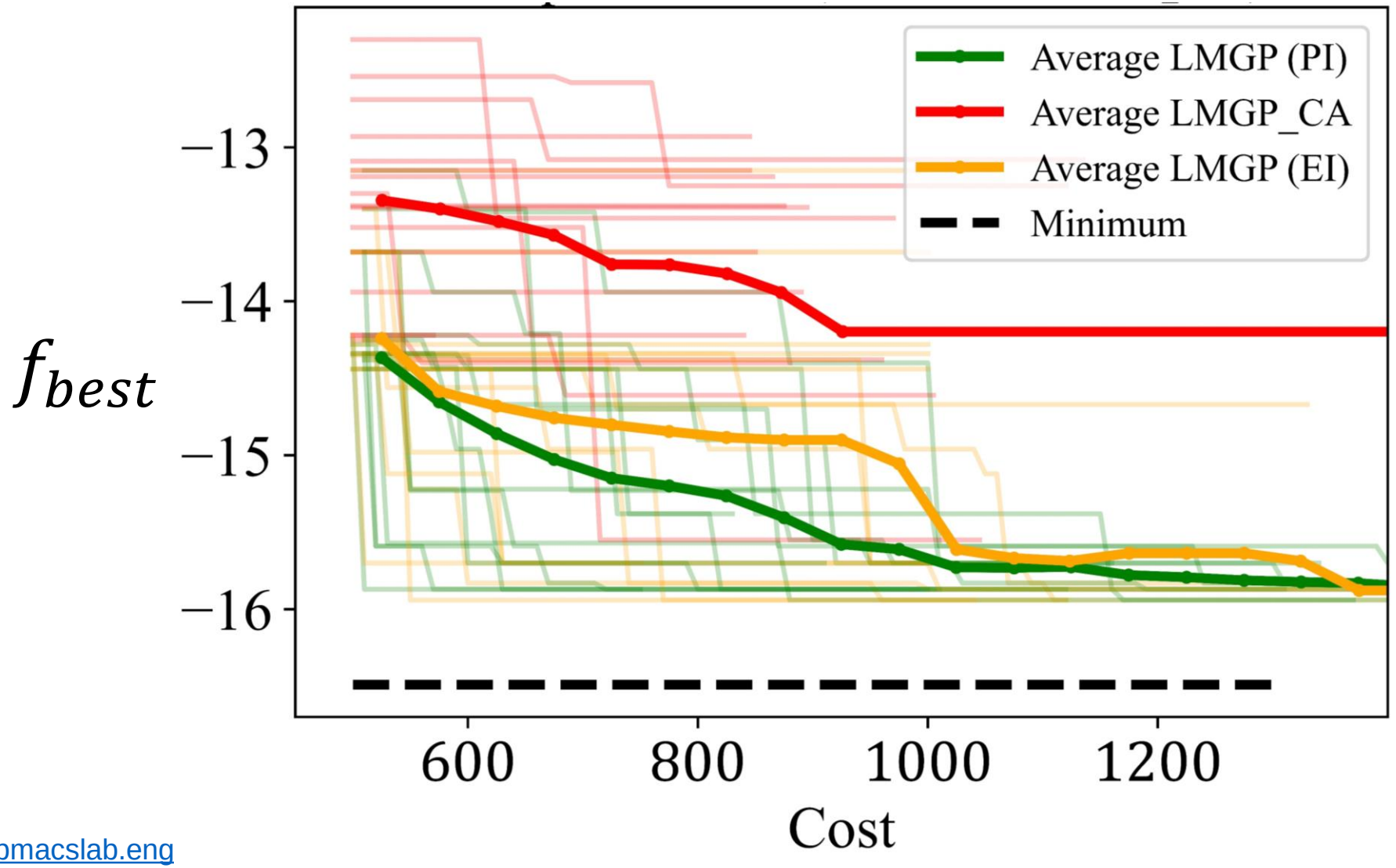
Reps: 20



Low-fidelity sources will NOT help!

BO CONVERGENCE

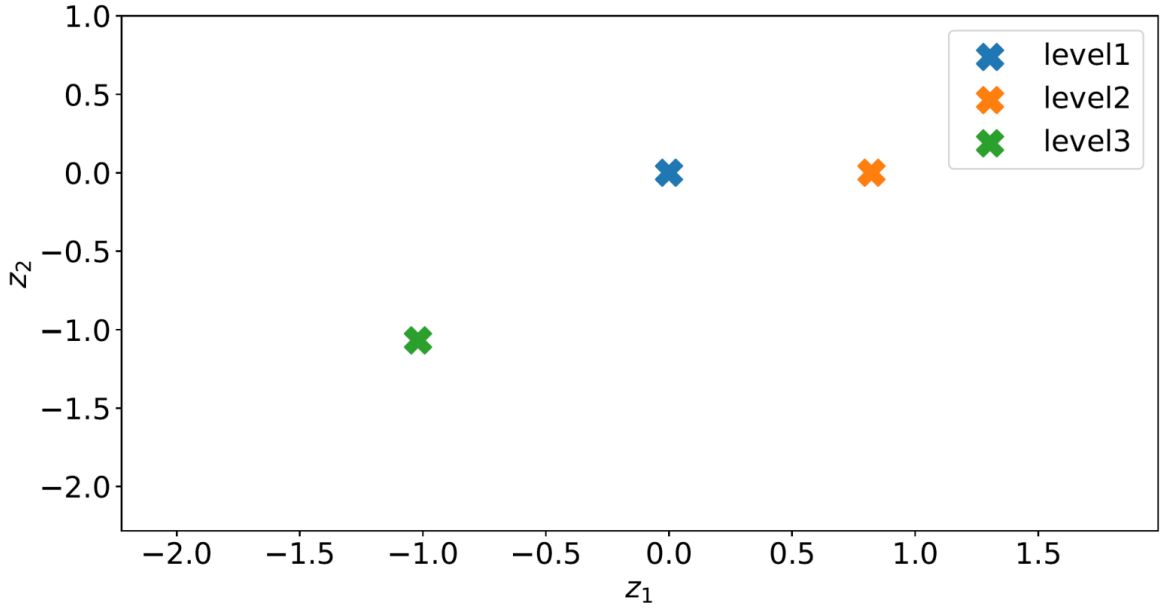
As expected, single-fidelity outperforms multi-fidelity BO!



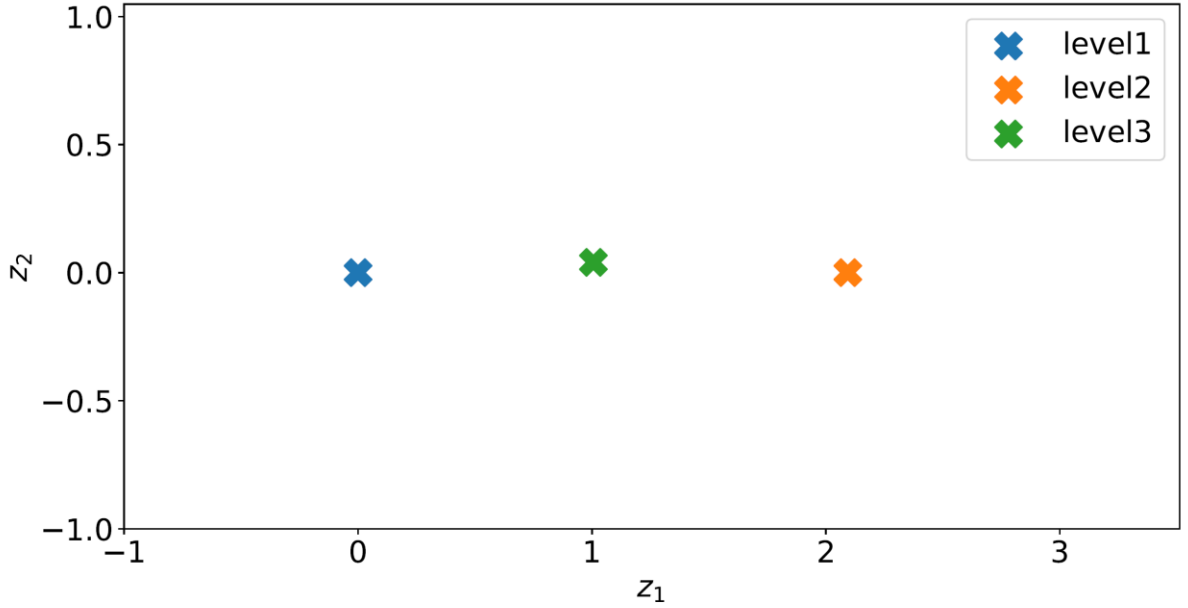
FIDELITY MANIFOLD COMPARISON (HOIP)

Can we trust the fidelity manifold learnt on small initial data?

Fidelity manifold via initial data



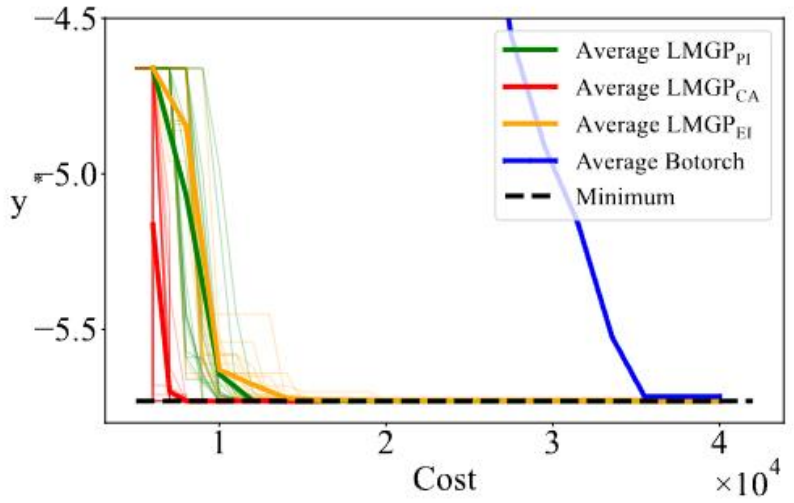
Fidelity manifold via all data



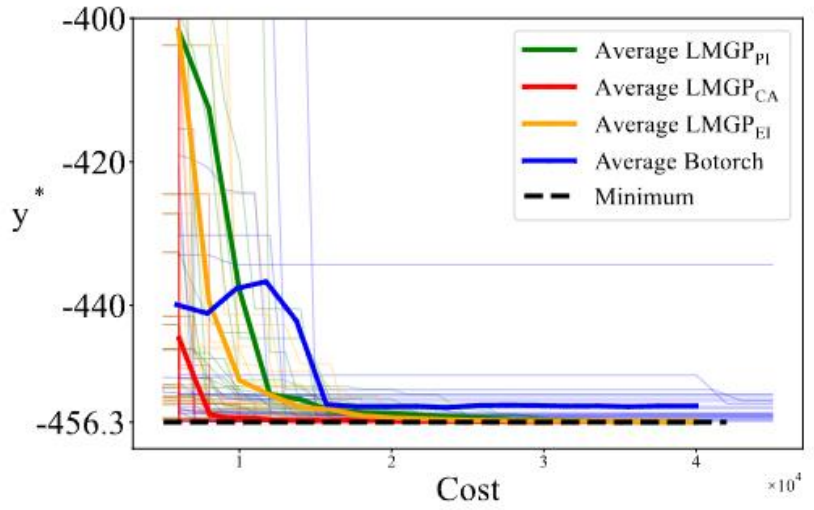
Zanjani Foumani, Z., et al., Multi-fidelity cost-aware Bayesian optimization. Computer Methods in Applied Mechanics and Engineering, 2023. 407: p. 115937.

COMPARISON STUDIES: ANALYTIC FUNCTIONS

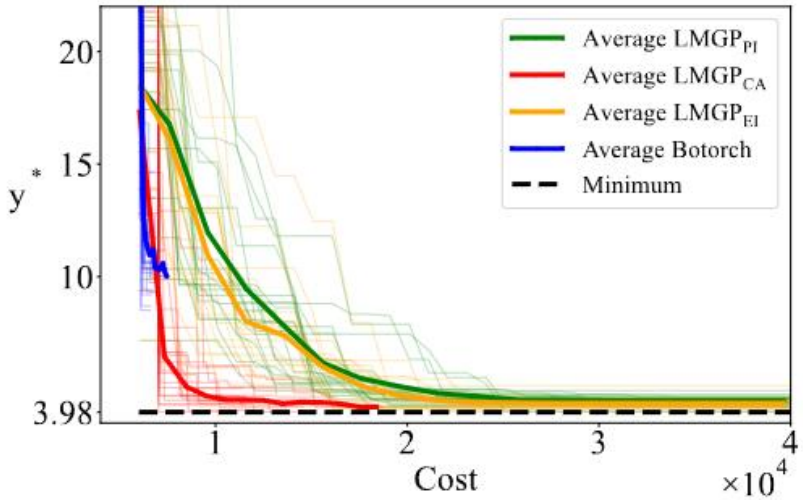
Our approach outperforms BoTorch (which is lookahead).



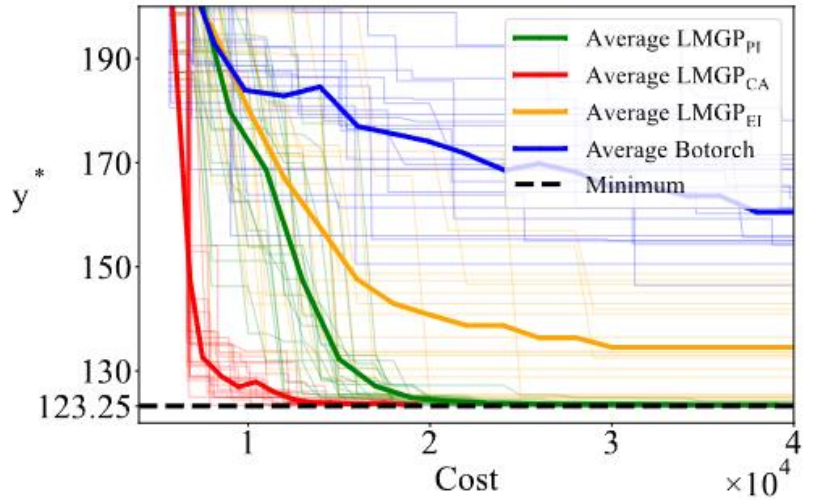
(a) Double-well Potential



(b) Rosenbrock



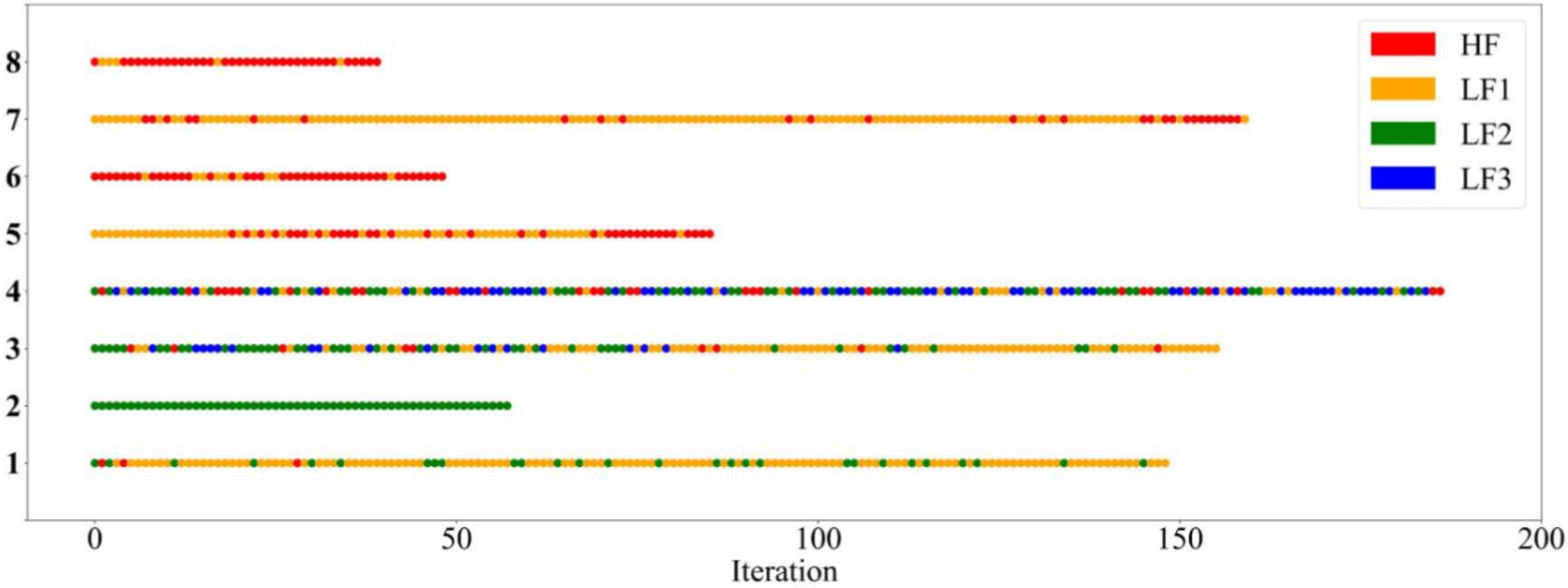
(c) Borehole



(d) Wing

Zanjani Foumani, Z., et al., Multi-fidelity cost-aware Bayesian optimization. Computer Methods in Applied Mechanics and Engineering, 2023. 407: p. 115937.

EXPLORATION VS EXPLOITATION FOR DATA SOURCE



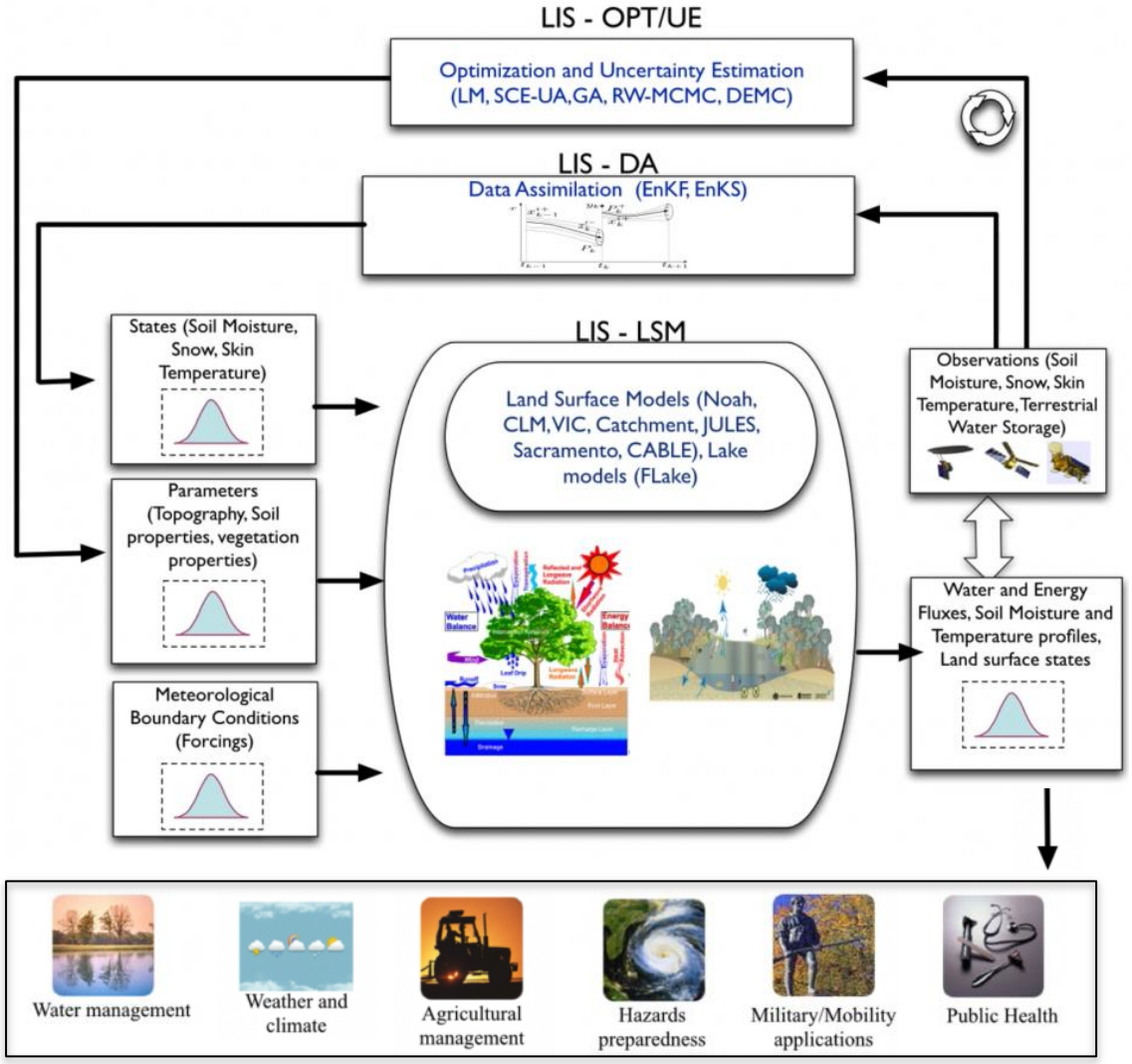
1: Borehole ($LMGP_{CA}$) **3**: Wing ($LMGP_{CA}$) **5**: Rosenbrock ($LMGP_{CA}$) **7**: Double well ($LMGP_{CA}$)

2: Borehole (BoTorch) **4**: Wing (BoTorch) **6**: Rosenbrock (BoTorch) **8**: Double well (BoTorch)

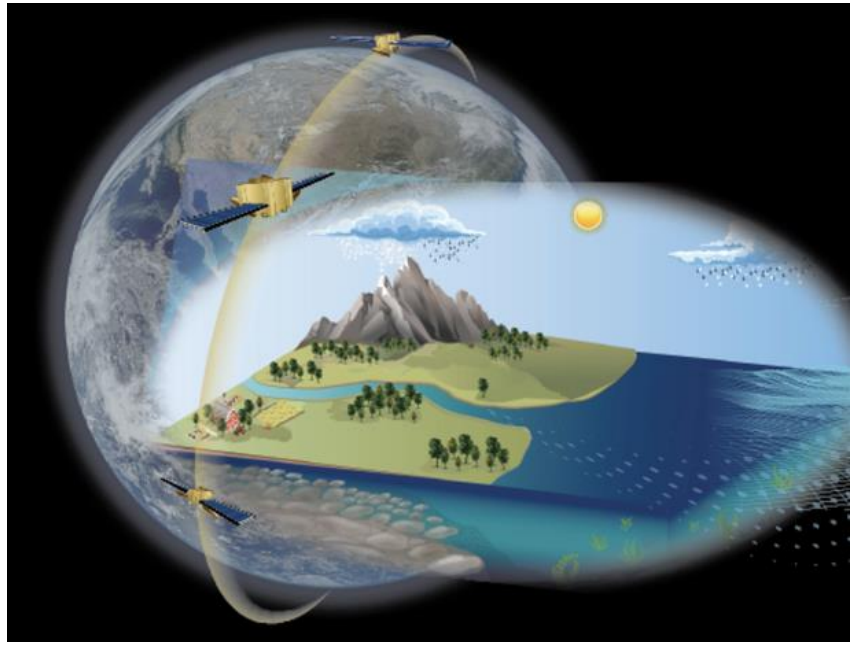
Zanjani Foumani, Z., et al., Multi-fidelity cost-aware Bayesian optimization. Computer Methods in Applied Mechanics and Engineering, 2023. 407: p. 115937.

PROBABILISTIC METAMODELING (EMULATION)

Land information system (LIS):



Freshwater component of LIS:



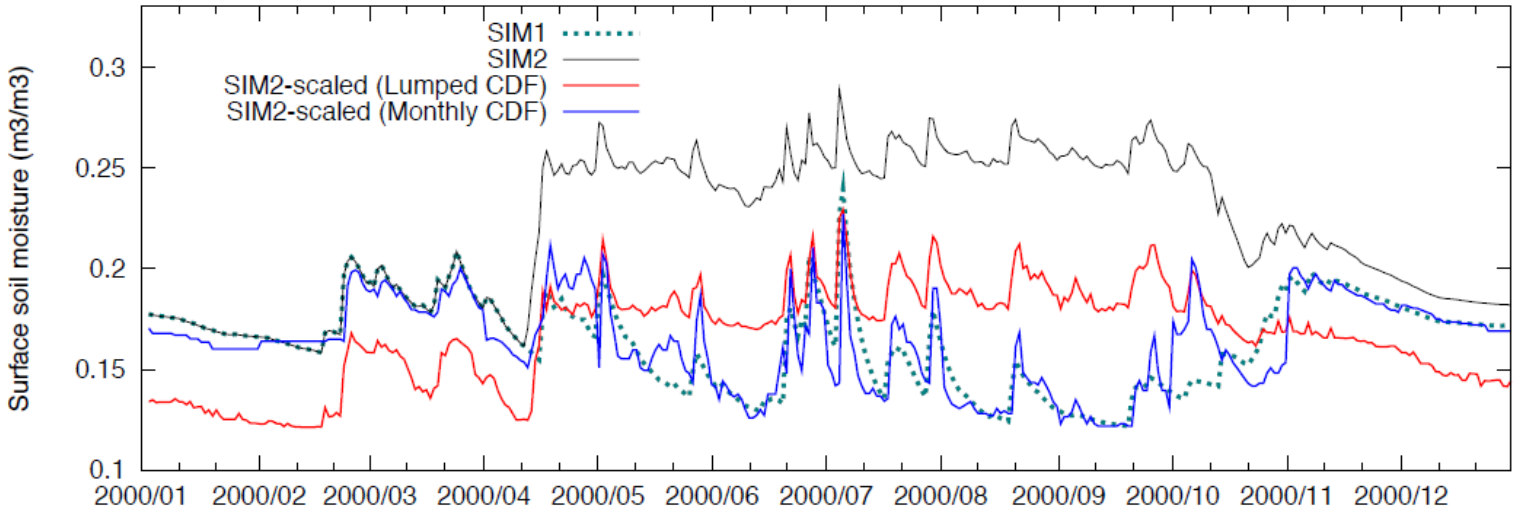
https://fwis.eis-hydro.mysmce.com/index.html

1. How can we reduce reliance on expensive simulations?
2. How can we detect anomalies?

https://pmacslab.eng.uci.edu/

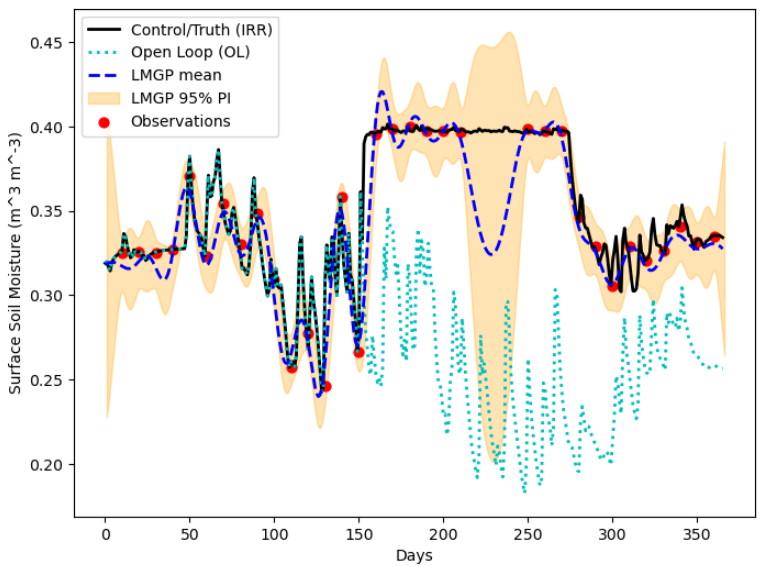
Kumar, Sujay V., et al. "A Land Surface Data Assimilation Framework Using the Land Information System: Description and Applications." Advances in Water Resources 31.11 (2008): 1419-32. Print.

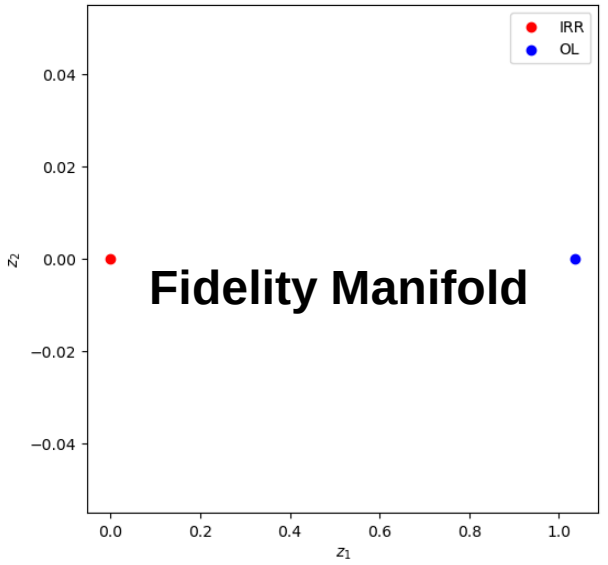
DATA ASSIMILATION FOR LIS



Kumar, S. V., et al. "Evaluating the Utility of Satellite Soil Moisture Retrievals over Irrigated Areas and the Ability of Land Data Assimilation Methods to Correct for Unmodeled Processes." Hydrol. Earth Syst. Sci. 19.11 (2015): 4463-78

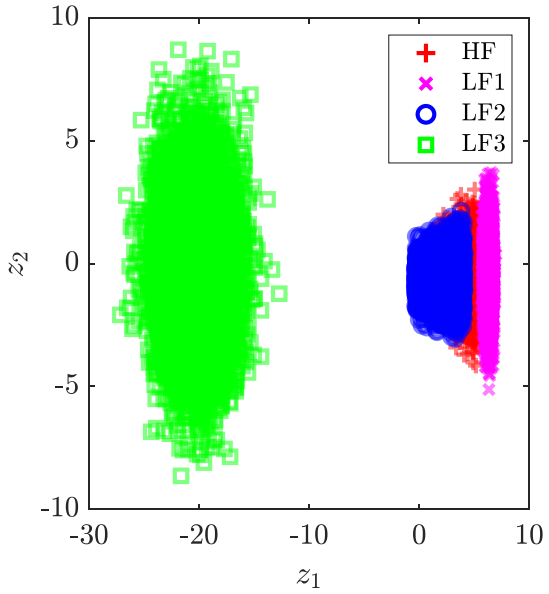
Figure 7. Same as Fig. 2 and including the surface soil moisture time series from rescaling SIM2 to the SIM1 climatology. The red and blue lines represent the SIM2 integration rescaled to the SIM1 climatology using lumped and monthly CDF matching, respectively.

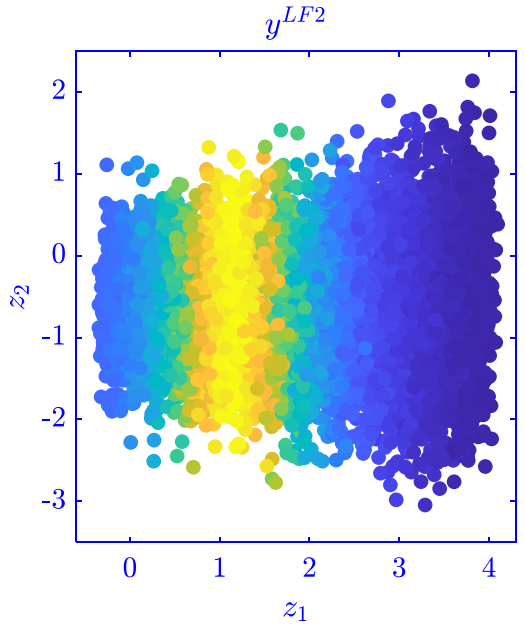


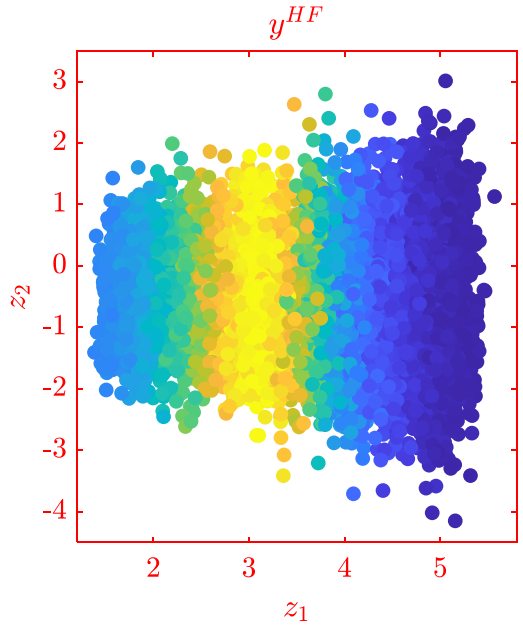


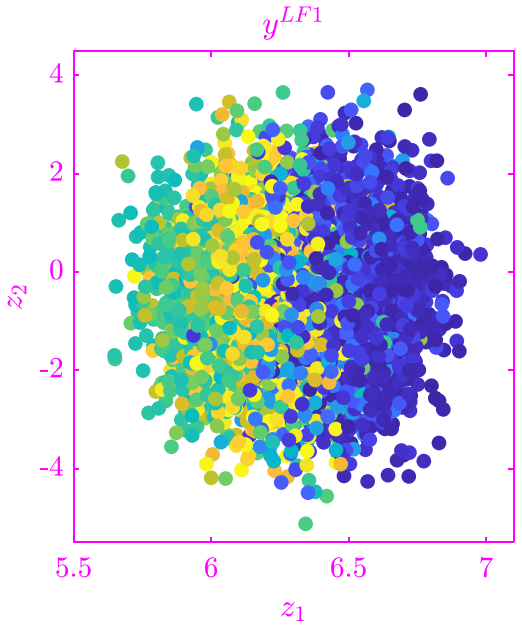
ONGOING (UNPUBLISHED) WORKS

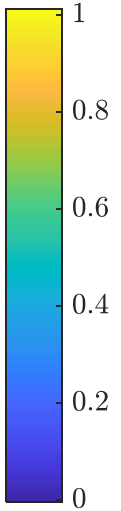
1. **Novelty or anomaly detection**
2. **Non-Gaussian uncertainty quantification**
3. **Learning spatial model discrepancies**











https://pmacslab.eng.uci.edu/

ANOMALY DETECTION

Identification of rare items, events, or observations which deviate significantly from the majority of the samples (outlier detection or novelty detection).

Applications:

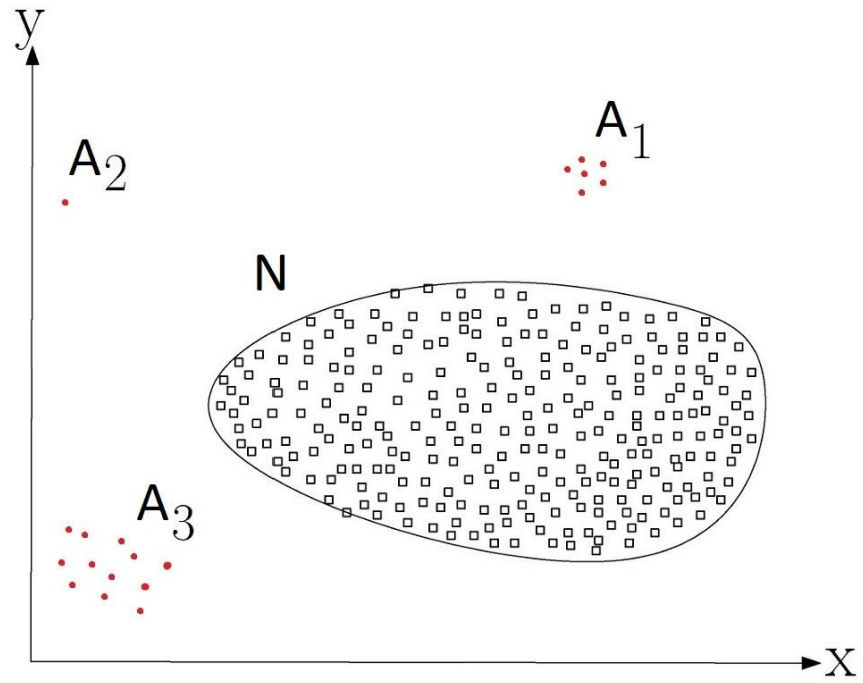
- Fault detection
- System health monitoring
- Fraud detection (Ex: manufacturing, credit card fraud, etc.)
- Intrusion detection
- Detect ecosystem disturbances
- Anomaly detection for product quality

Given:

Data with mixed anomalous and normal samples.

Goal:

Finding anomalous samples.



A simple example of anomalies in a two-dimensional dataset.

MANUFACTURING DATA

Aluminum high-pressure die-casting (HPDC)

- **1495 observations**, each sample representing a production casting, with **79 input variable** data collected from 12 HPDC work cells and 20 die casting tools in the production of engine block castings at FCA Kokomo Casting Plant.
- **Output:** ultimate tensile strength (UTS).
- **Outliers:** 21 data points, manually selected.

Can we find these anomalous data automatically?

Part of the data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1		SerialNur	AvgFastH	AvgFastR	AvgInterm	AvgInterm	AvgSlowH	AvgSlowR	BiscuitLe	CavityFill	DieClose	DieClose	EndofSho	FastShot\	FinalInte	Intensific	IntensPre	IntensVe	Intermed	MetalTen	SlowShot	TieBarTor	TieBarTor	TieBarTor	TieBarTor	TieBarTor	TipLubeTi	CycleTime	DwellTim	DwellTim	DieOpenT	ExtractRol	SprayRob	LinerLoac	C
2	0	0018C603	1336	1348	2049	3849	1881	3749	2.6	104	22.1	101	38.9	187.8	6338	0.83	201	6	10.2	1229	6	863	831	809	845	3347	27.5	144.35	14.024	14	14.475	18.075	60.35	1.93	
3	1	0028C603	1322	1354	2065	3841	1890	3741	2.4	101	21.7	105	39.1	193.6	6427	0.8	201	6	10	1225	6.2	857	835	818	863	3375	27.5	145.12	14.024	14	14.475	17.875	60.475	1.922	
4	2	0028C603	1333	1382	2067	3864	1883	3766	2.4	102	21.7	105	39.1	191	6416	0.86	210	6	10.1	1224	6.2	857	835	820	862	3373	27.5	144.32	14.024	14	14.6	18	60.35	1.935	
5	3	0028H503	1246	1098	2005	3410	2021	3448	2.7	105	26.1	102	38.6	191.3	6816	0.95	129	8	9.4	1221	3.8	841	850	862	844	3397	27.5	143.23	14.014	14.014	11.311	16.068	64.35	1.378	
6	4	0038C603	1366	1367	2058	3880	1885	3744	2.3	106	21.2	105	39.2	185.3	6355	0.84	203	6	10.2	1229	6.4	862	836	821	867	3386	27.5	144.52	14	14	14.45	17.949	60.15	1.937	
7	5	0038C603	1348	1363	2061	3909	1883	3760	2.5	106	21.2	105	39.1	185.1	6314	0.81	201	6	10.3	1227	6.2	859	834	824	865	3382	27.5	144.5	14.001	14.024	14.5	18.075	60.275	1.942	
8	6	0038H503	1279	1163	2014	3387	2024	3403	2.8	108	26.1	102	38.5	190	6888	1	136	8	9.7	1220	3.7	836	839	858	845	3378	27.5	143.31	14.014	14.015	10.946	16.224	64.455	1.372	
9	7	0038H503	1279	1163	2014	3387	2024	3403	2.8	108	26.1	102	38.5	190	6888	1	136	8	9.7	1220	3.7	836	839	858	845	3378	27.5	143.31	14.014	14.015	10.946	16.224	64.455	1.372	
10	8	0048C502	1303	1202	2136	4153	2177	4255	2.6	102	21.5007	95	39	193.8	6482	0.86	157	2	10.1	1213	4	794	828	843	824	3289	20	143.33	14.013	14.025	15.001	14.153	57.878	1.616	
11	9	0048C502	1301	1183	2124	4158	2163	4259	2.6	103	21.5377	93	38.9	191.6	6490	0.87	164	2	10.1	1211	4	794	824	843	814	3275	20	142.86	14.017	14.016	14.976	14.281	57.753	1.616	
12	10	0048H503	1237	1113	2014	3391	2031	3463	2.8	113	26.1	99	38.4	185.8	6887	1.04	141	7	10.7	1222	5.6	828	836	846	837	3347	27.5	141.9	14.013	14.014	10.869	16.12	64.375	1.378	
13	11	0058C502	1279	1177	2105	4148	2145	4258	2.6	103	21.50995	93	38.8	191.4	6408	0.9	160	2	10.2	1220	4	805	830	843	814	3292	20	143.05	14.005	14.021	14.878	14.14	57.955	1.614	
14	12	0058C502	1283	1186	2108	4147	2143	4269	2.6	102	21.5007	94	38.9	192.3	6428	0.84	179	2	10.3	1226	4	800	822	844	818	3284	20	142.64	14.009	14.009	14.801	14.205	57.755	1.611	
15	13	0058H503	1212	1096	1985	3388	2006	3424	3	122	26.1	100	38.3	185.2	6946	0.96	129	7	10.3	1216	5.7	832	842	855	844	3375	27.5	142.06	14.014	14.014	11.05	15.963	64.272	1.383	
16	14	0068C605	1316	1367	2001	3931	1429	2930	2.6	105	19.88541	93	39.1	185.2	6351	0.85	141	2	9.5	1227	4.4	852	847	848	872	3419	20	142.8	14.011	14.017	12.987	16.148	57.428	2.069	
17	15	0078C502	1291	1242	2138	4132	2178	4272	2.8	103	21.2528	95	38.7	194	6516	0.85	176	2	10.2	1219	4	796	820	841	811	3268	20	143.3	14.025	14.026	14.607	14.252	57.653	1.618	
18	16	0078C502	1301	1243	2138	4114	2177	4267	2.8	102	21.23245	95	38.7	194.3	6483	0.86	178	2	10.2	1221	4.1	799	815	842	811	3267	20	143.28	14.013	14.006	14.621	14.172	57.856	1.617	
19	17	0078C502	1315	1217	2139	4140	2181	4255	2.8	104	21.2454	95	38.7	190.9	6512	0.87	175	2	10.2	1221	4.1	798	811	841	819	3269	20	143.35	14.009	14.009	14.688	14.22	57.855	1.608	
20	18	0078C502	1296	1241	2140	4127	2177	4272	2.8	103	21.2269	95	38.7	194.1	6494	0.84	196	2	10.2	1218	4.1	796	825	842	808	3271	20	143.56	14.022	14.016	14.65	14.237	57.991	1.61	
21	19	0078F606	1890	1847	1917	1868	1966	1919	2.3	101	21	88	39.3	194.8	6458	0.83	231	7	10.1	1210	4	858	846	852	875	3431	30	135.07	14.004	14	10.543	0	53.094	1.67	

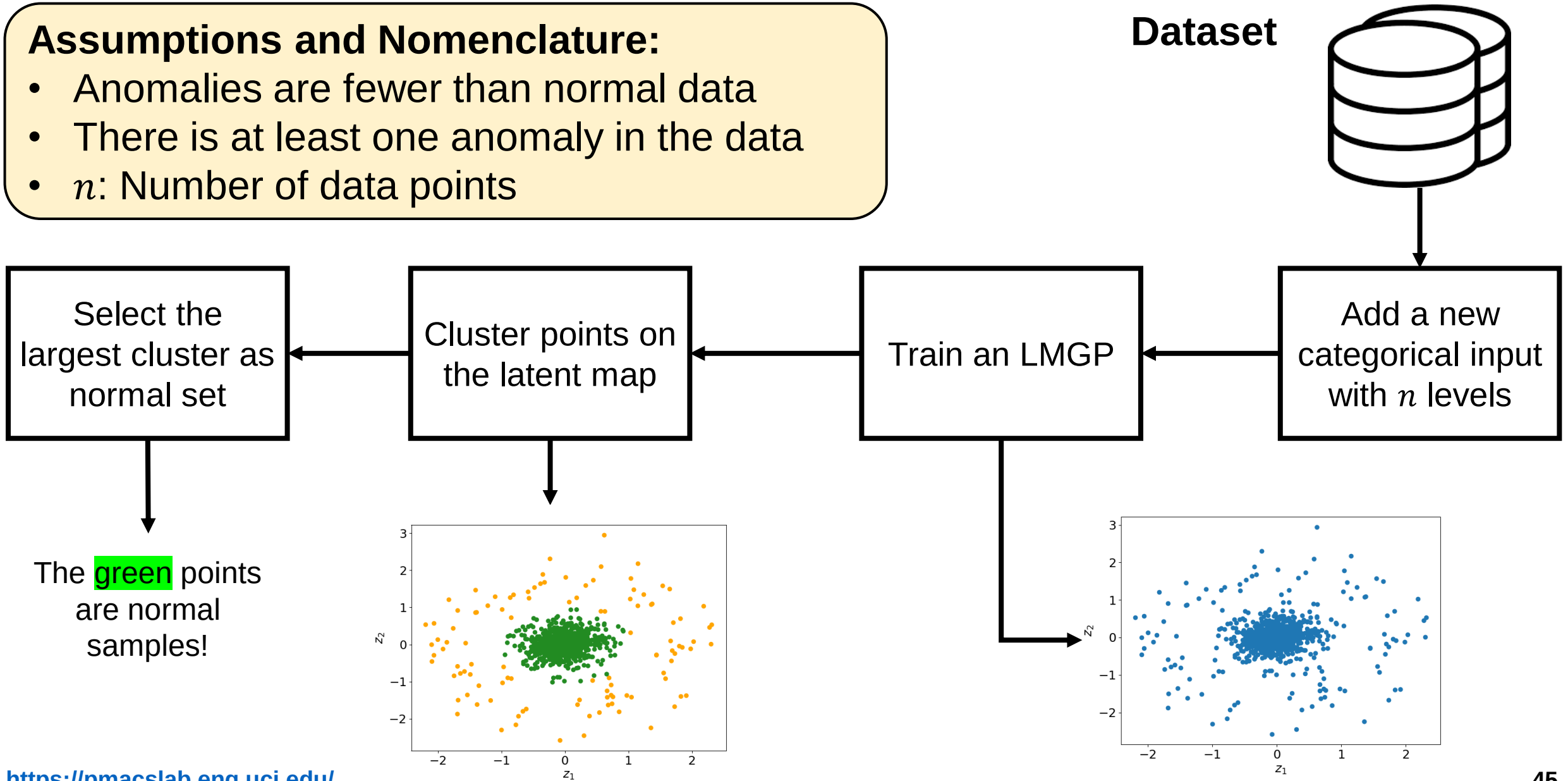
Kopper, Adam, et al. "Model selection and evaluation for machine learning: deep learning in materials processing." Integrating Materials and Manufacturing Innovation 9.3 (2020): 287-300.

https://pmacslab.eng.uci.edu/

ANOMALY DETECTION VIA LMGP

Assumptions and Nomenclature:

- Anomalies are fewer than normal data
- There is at least one anomaly in the data
- n: Number of data points



https://pmacslab.eng.uci.edu/

MANUFACTURING DATA: RESULTS

Aluminum high-pressure die-casting (HPDC)

- 1495 observations, 79 input variable, and 1 output
- Outliers: 21 data points

$$\text{True positive rate (TPR)} = \frac{\text{True Positive}}{\text{True Positive}+\text{False Negative}}$$

$$\text{True negative rate (TNR)} = \frac{\text{True Negative}}{\text{True Negative}+\text{False Positive}}$$

$$G_{mean} = \sqrt{\text{TPR} + \text{TNR}}$$

Number of anomalous	TPR		TNR		G-mean	
	LMGP	Isolation forest	LMGP	Isolation forest	LMGP	Isolation forest
21	0.994	0.973	1.000	0.100	0.997	0.312
100	0.993	0.978	0.990	0.030	0.991	0.170
200	0.966	0.973	0.935	0.030	0.950	0.171

CONCLUDING REMARKS

Scarce data scenarios are ubiquitous in many real world applications:

- Judicious data collection → Adaptive sampling
- **Probabilistic learning** → GP/LMGP
- **Data fusion** → LMGP
- Multi-fidelity BO → LMGP + New acquisition function

ACKNOWLEDGMENT



Tammer Eweis-Labolle
PhD Student



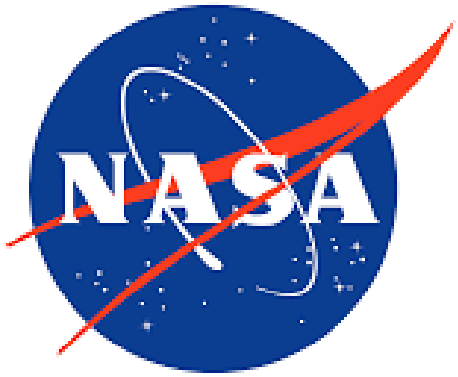
Sanaz Zanjani
PhD Student



Amin Yousefpour
PhD Student



Mehdi Shishehbor
Postdoc







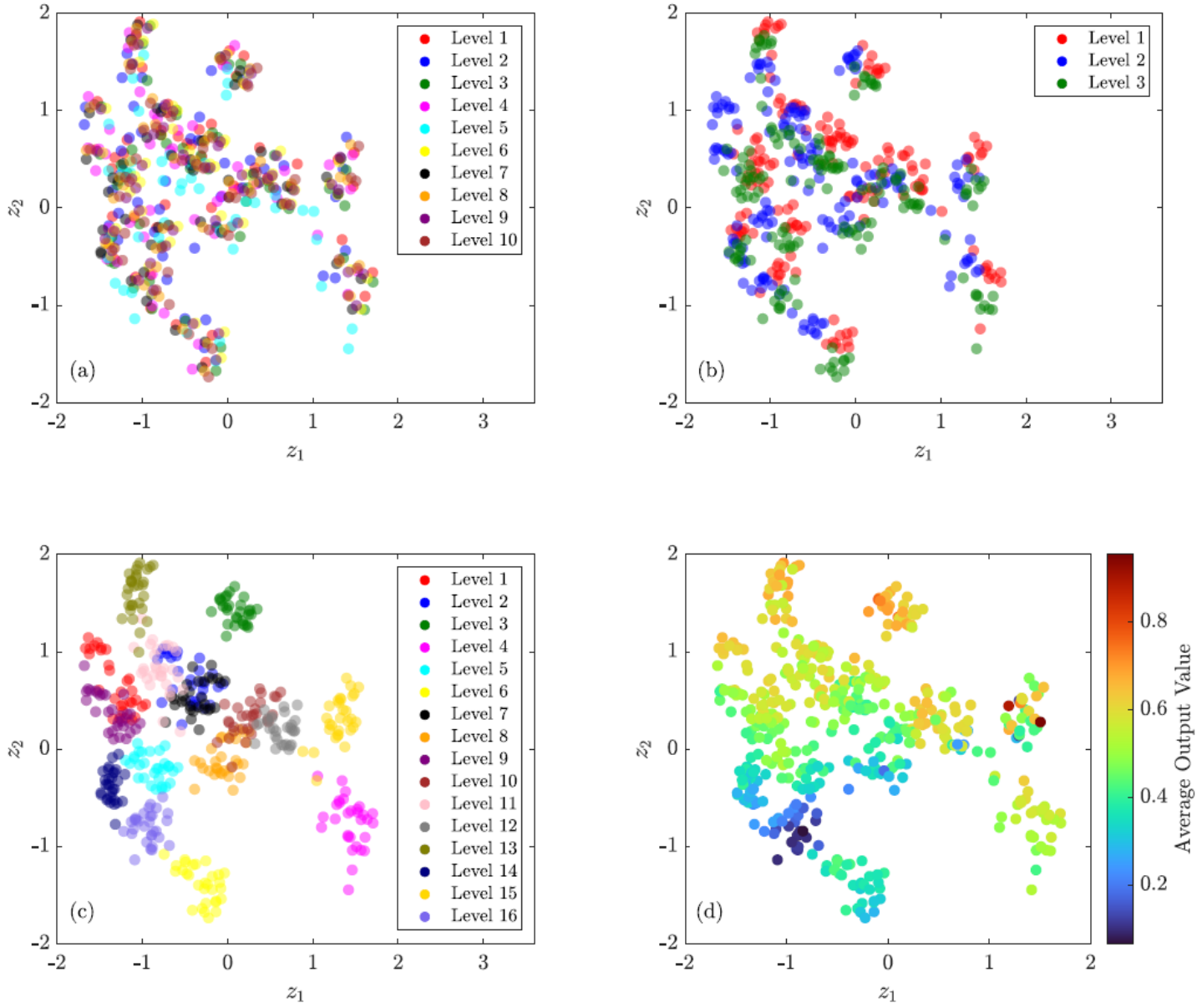


Thank you!

Questions?

https://pmacslab.eng.uci.edu/

HOIP: LEARNT COMPOSITION MANIFOLD



https://pmacslab.eng.uci.edu/

CALIBRATION: ANALYTICAL EXAMPLE

Source 1: $$y_h = 0.036 S_\omega^{0.758} W_{f\omega}^{0.0035} \left(\frac{A}{cos^2(\Lambda)}\right)^{0.6} \theta_1^{0.006} \theta_2^{0.04} \left(\frac{100\theta_3}{\cos(\Lambda)}\right)^{-0.3} (\theta_4 W_{dg})^{0.49} + S_\omega W_p$$

Source 2: $$y_{l_1} = 0.036 S_\omega^{0.758} W_{f\omega}^{0.0035} \left(\frac{A}{cos^2(\Lambda)}\right)^{0.6} \theta_1^{0.006} \theta_2^{0.04} \left(\frac{100\theta_3}{\cos(\Lambda)}\right)^{-0.3} (\theta_4 W_{dg})^{0.49} + W_p$$

Source 3: $$y_{l_2} = 0.036 S_\omega^{0.800} W_{f\omega}^{0.0035} \left(\frac{A}{cos^2(\Lambda)}\right)^{0.6} \theta_1^{0.006} \theta_2^{0.04} \left(\frac{100\theta_3}{\cos(\Lambda)}\right)^{-0.3} (\theta_4 W_{dg})^{0.40} + W_p$$

Source 4: $$y_{l_3} = 0.036 S_\omega^{0.900} W_{f\omega}^{0.0035} \left(\frac{A}{cos^2(\Lambda)}\right)^{0.6} \theta_1^{0.006} \theta_2^{0.04} \left(\frac{100\theta_3}{\cos(\Lambda)}\right)^{-0.3} (\theta_4 W_{dg})^{0.49} + 0$$

Error

$$\boldsymbol{x}^T = [S_\omega, W_{f\omega}, A, \Lambda, W_{dg}, W_p], \boldsymbol{\theta}^T = [\theta_1, \theta_2, \theta_3, \theta_4],$$

$$\min(\boldsymbol{x}) = [150, 220, 6, -10, 1700, 0.025],$$

$$\max(\boldsymbol{x}) = [200, 300, 10, 10, 2500, 0.08],$$

$$\min(\boldsymbol{\theta}^T) = [16, 0.5, 0.08, 2.5], \max(\boldsymbol{\theta}^T) = [45, 1, 0.18, 6]$$

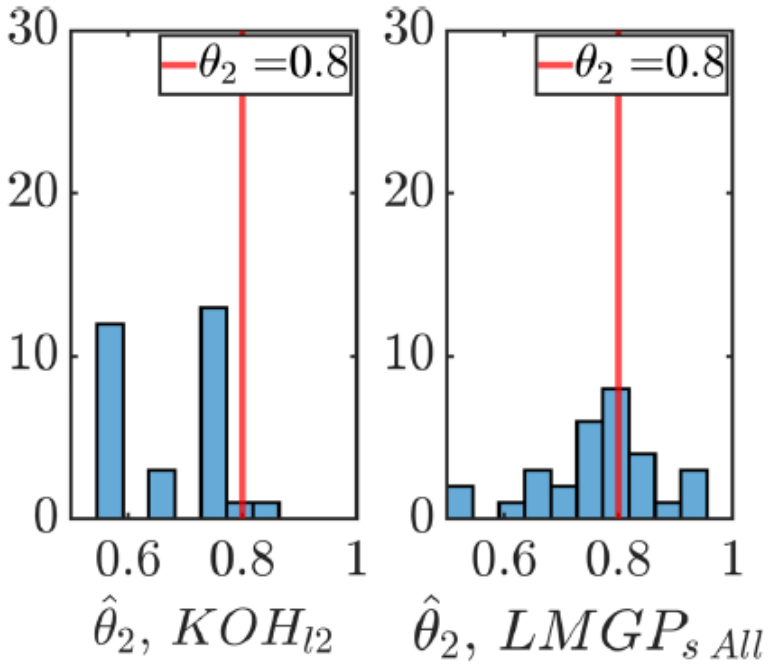
For Source 1: $$\boldsymbol{\theta}^{T^*} = [40, 0.8, 0.17, 3]$$

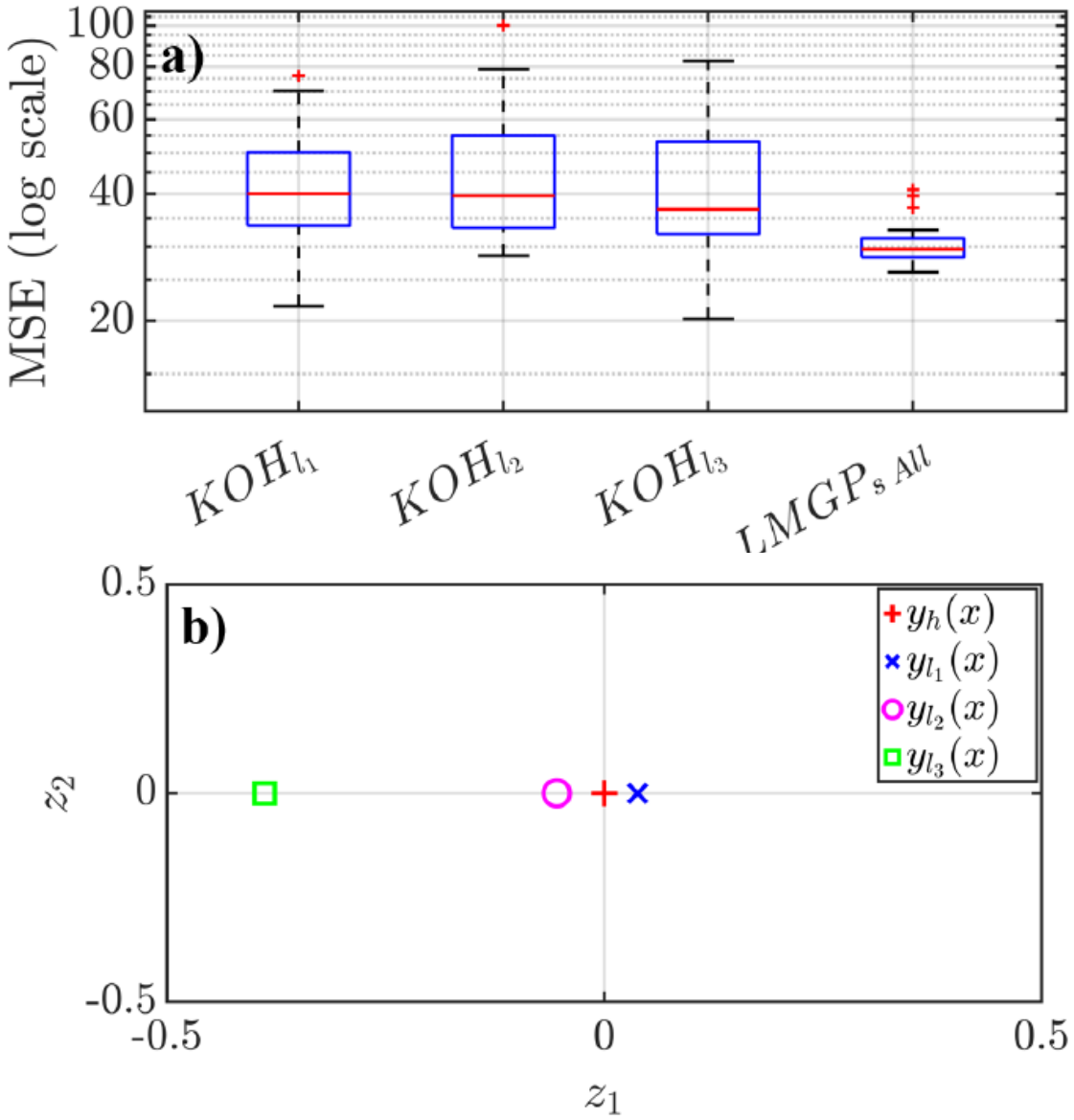
Training process

- **15** high fidelity samples from source 1
- **50** low fidelity samples from sources 1, 2, 3
- For all sources: $\sigma = 5$
- Repeat 30 times

CALIBRATION RESULTS

- MSE is lower than competing methods.
- Learnt latent space is again useful.
- Distribution of estimated parameters are more representative of uncertainties.





https://pmacslab.eng.uci.edu/

Eweis-Labolle, J., N. Oune and R. Bostanabad (2022). "Data Fusion with Latent Map Gaussian Processes." Journal of Mechanical Design: 1-41.

MODIFIED BOREHOLE

- Input dimensionality: 8
- # of information sources: 5
- Initial data:
  - High-fidelity : 5
  - Low-fidelity 1: 5
  - Low-fidelity 2: 5
  - Low-fidelity 3: 5
  - Low-fidelity 4: 50
- Cost: 1000/100/100/100/10
- Minimum of $y_1(\boldsymbol{x})$: **3.98**

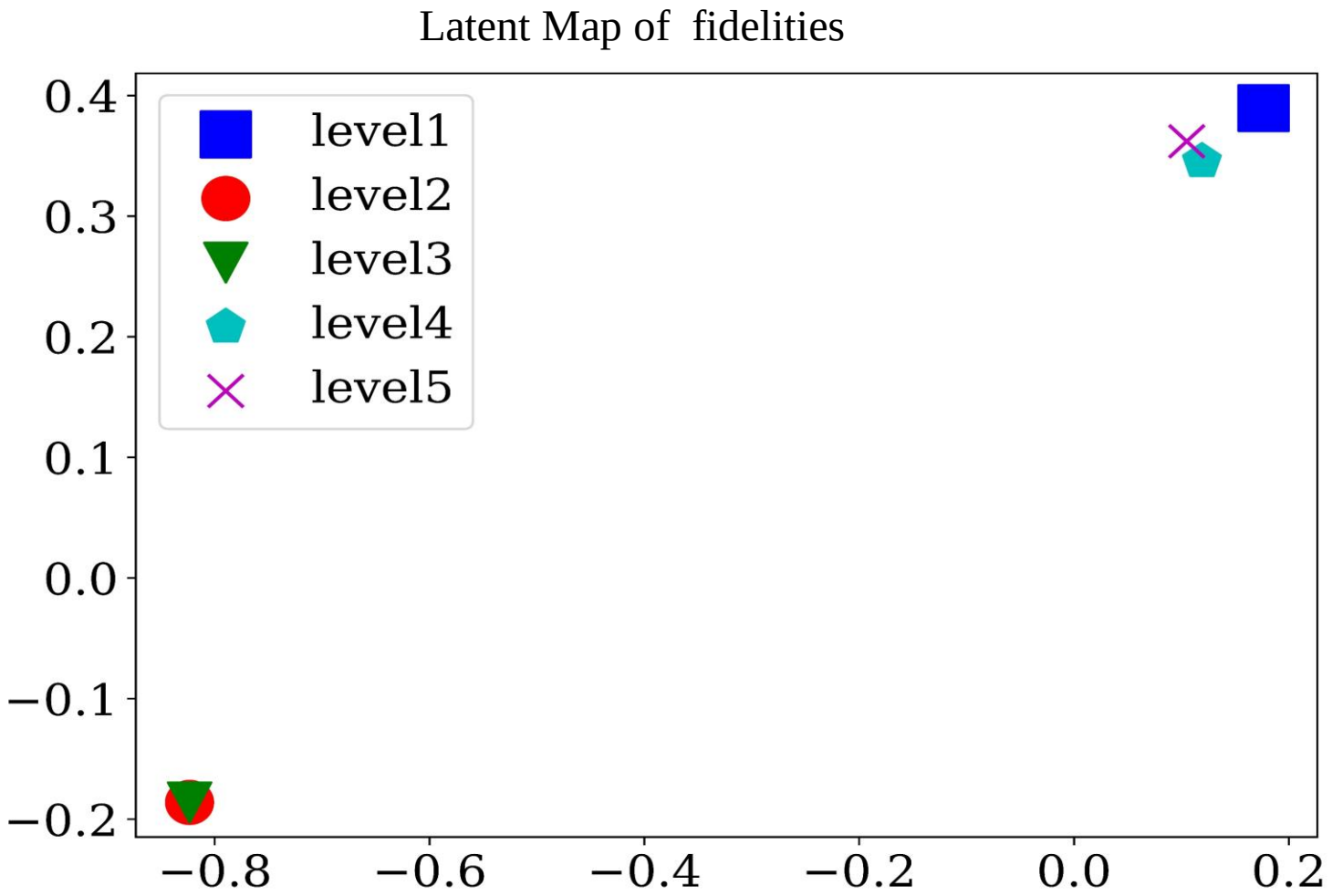
Source 1 (High-Fidelity):	$y_1 = \frac{2\pi T_u (H_u - H_l)}{\ln\left(\frac{r}{r_w}\right)\left(1 + \frac{2\,LT_u}{\ln\left(\frac{r}{r_w}\right) r_w^2 k_w} + \frac{T_u}{T_l}\right)}$
Source 2:	$y_2 = \frac{2\pi T_u (H_u - 0.8\,H_l)}{\ln\left(\frac{r}{r_w}\right)\left(1 + \frac{1\,LT_u}{\ln\left(\frac{r}{r_w}\right) r_w^2 k_w} + \frac{T_u}{T_l}\right)}$
Source 3:	$y_3 = \frac{2\pi T_u (H_u + 3H_l)}{\ln\left(\frac{r}{r_w}\right)\left(1 + \frac{8\,LT_u}{\ln\left(\frac{r}{r_w}\right) r_w^2 k_w} + 0.75\,\frac{T_u}{T_l}\right)}$
Source 4:	$y_4 = \frac{2\pi T_u (1.1\,H_u - H_l)}{\ln\left(\frac{4r}{r_w}\right)\left(1 + \frac{3\,LT_u}{\ln\left(\frac{r}{r_w}\right) r_w^2 k_w} + \frac{T_u}{T_l}\right)}$
Source 5:	$y_5 = \frac{2\pi T_u (1.05\,H_u - H_l)}{\ln\left(\frac{2r}{r_w}\right)\left(1 + \frac{2\,LT_u}{\ln\left(\frac{r}{r_w}\right) r_w^2 k_w} + \frac{T_u}{T_l}\right)}$

Should we use all information sources?

Do all of them help us?

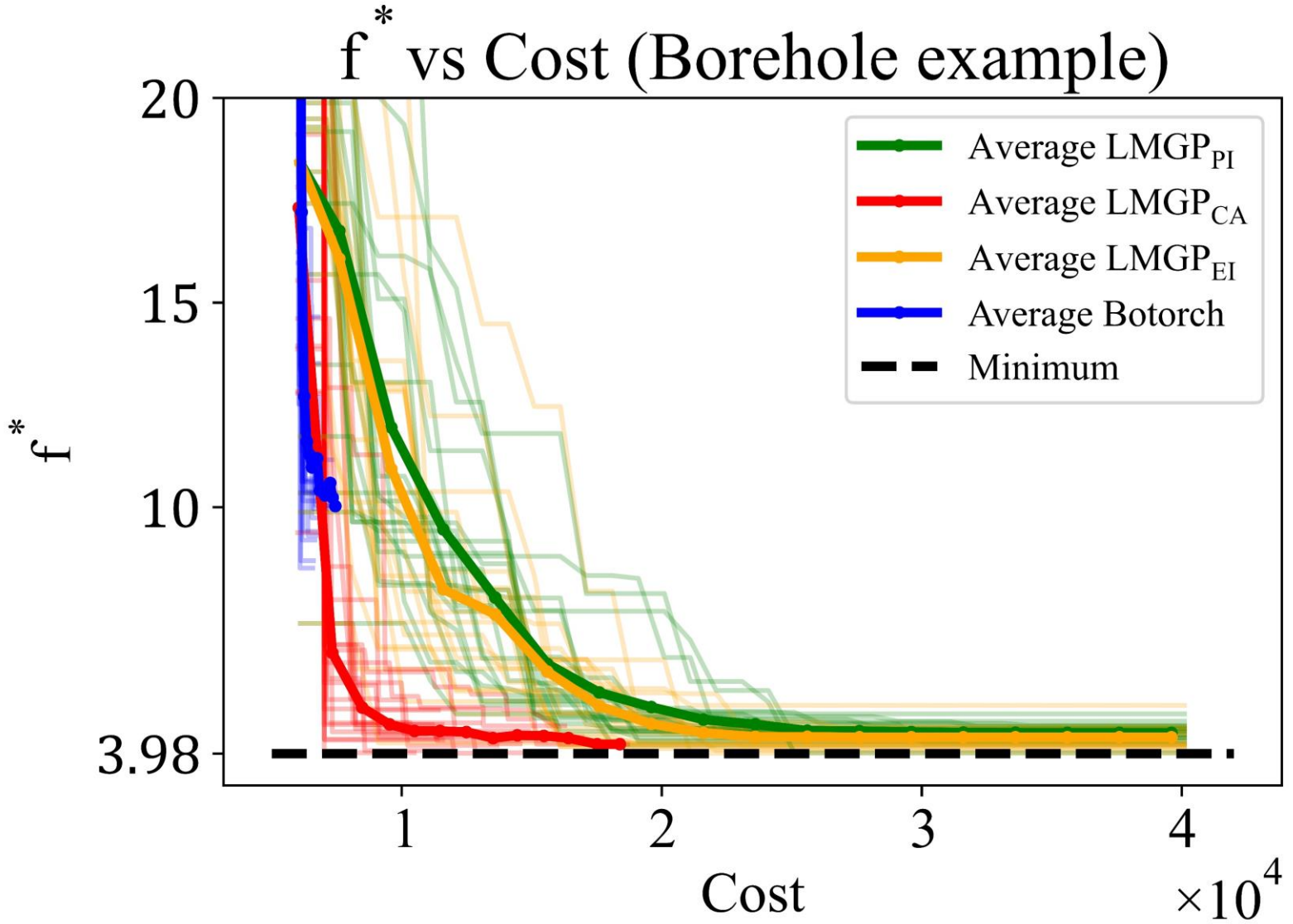
→ **Latent Map of LMGP**

WHICH SOURCE TO INCLUDE?



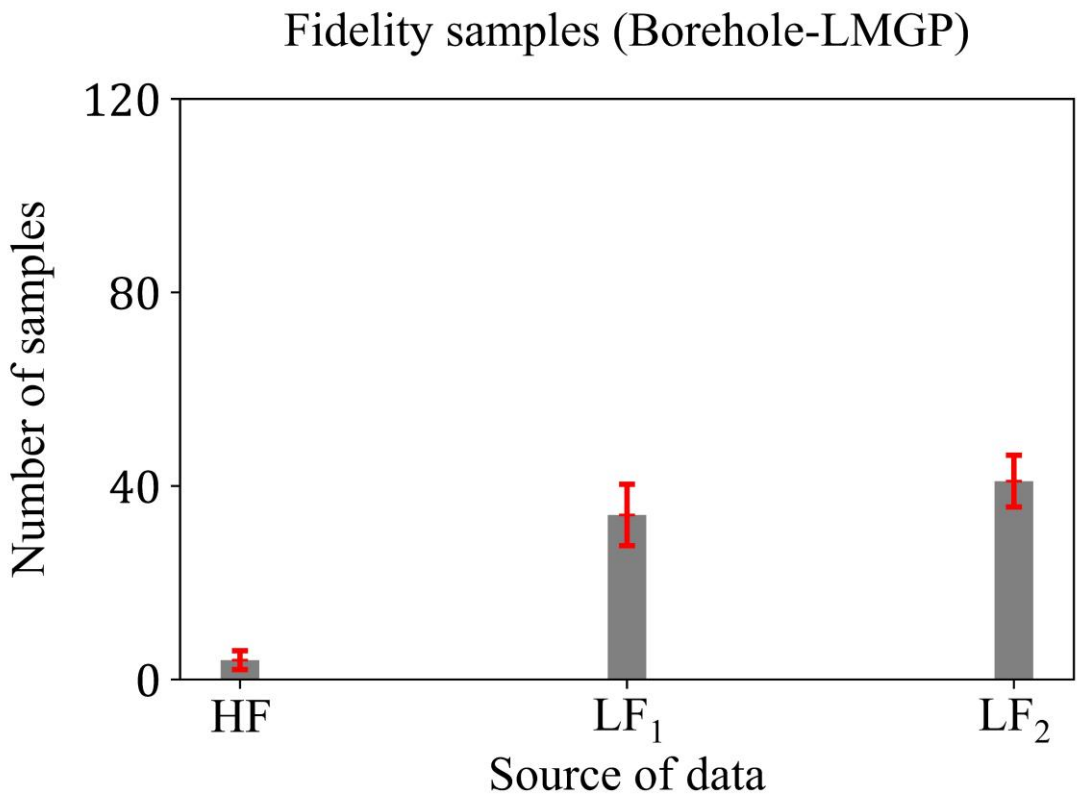
BO FOR THE BOREHOLE EXAMPLE

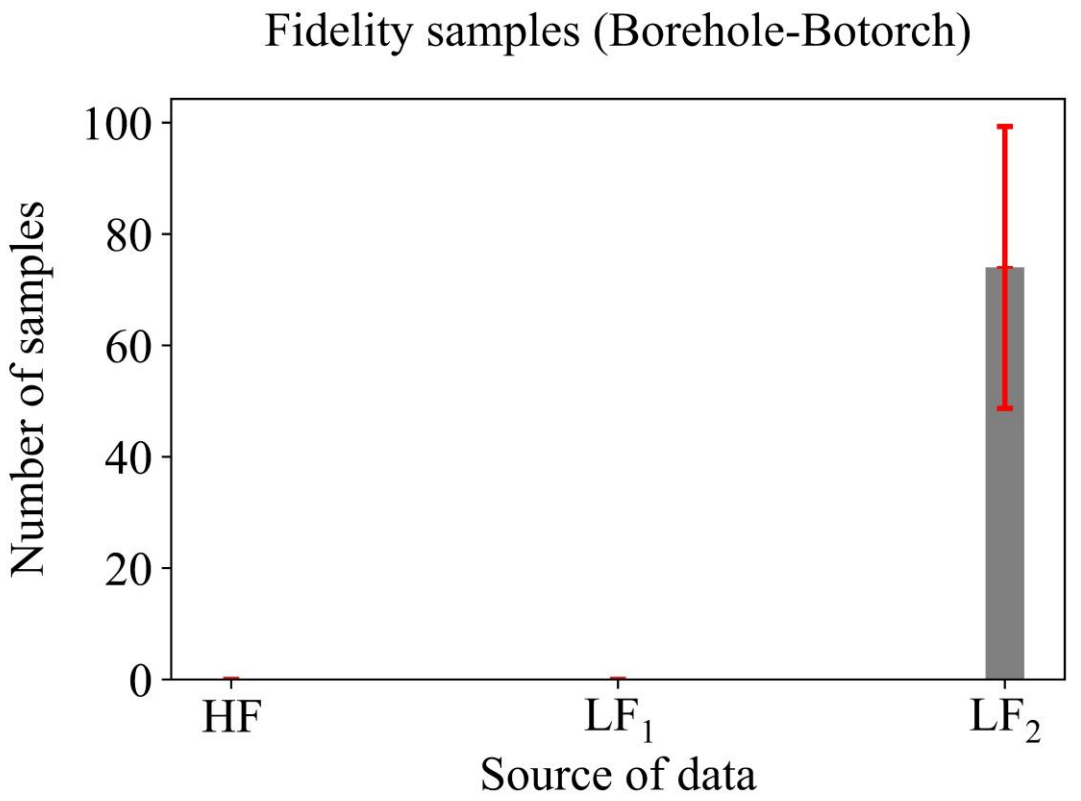
BOTroch is using knowledge gradient (lookahead)



SAMPLING

- Since LF2 is very inexpensive, BOTorch only samples from that.



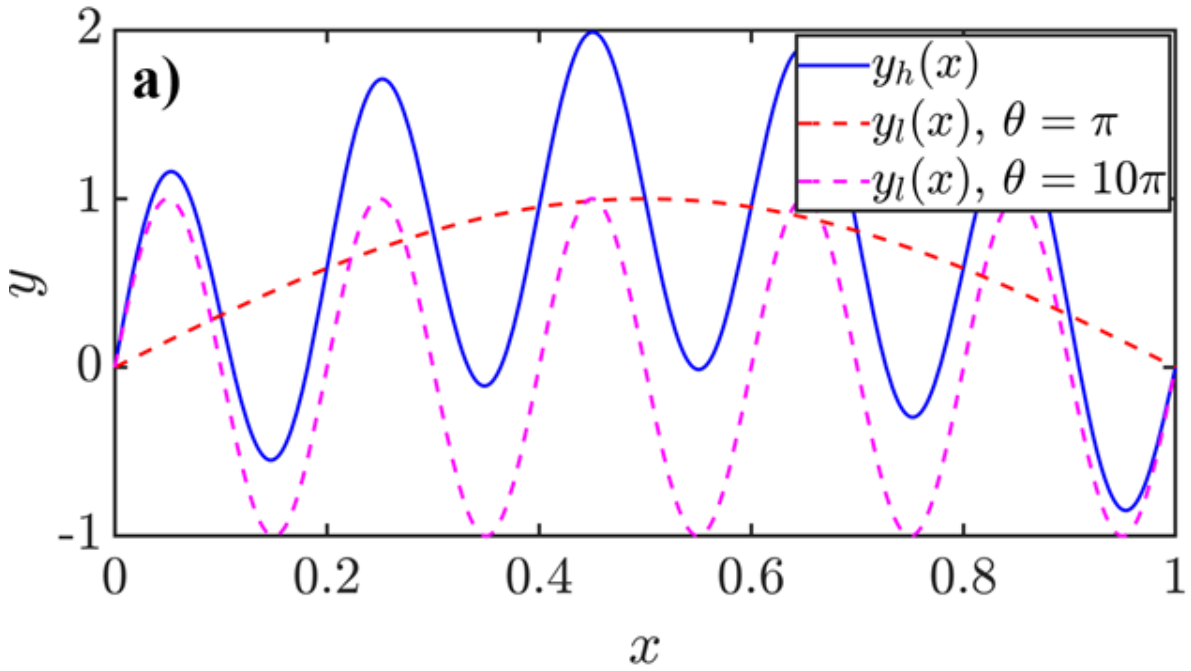


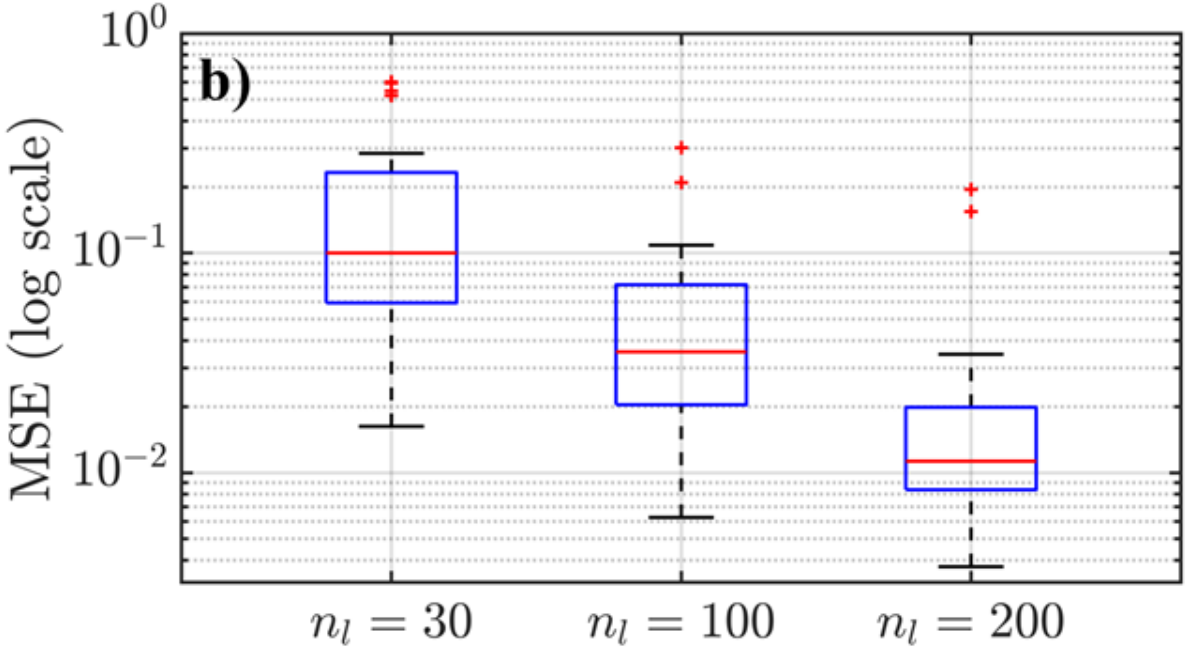
IDENTIFIABILITY IN CALIBRATION

$$y_h(x) = \sin(\pi x) + \sin(10\pi x), \qquad 0 \le x \le 1$$

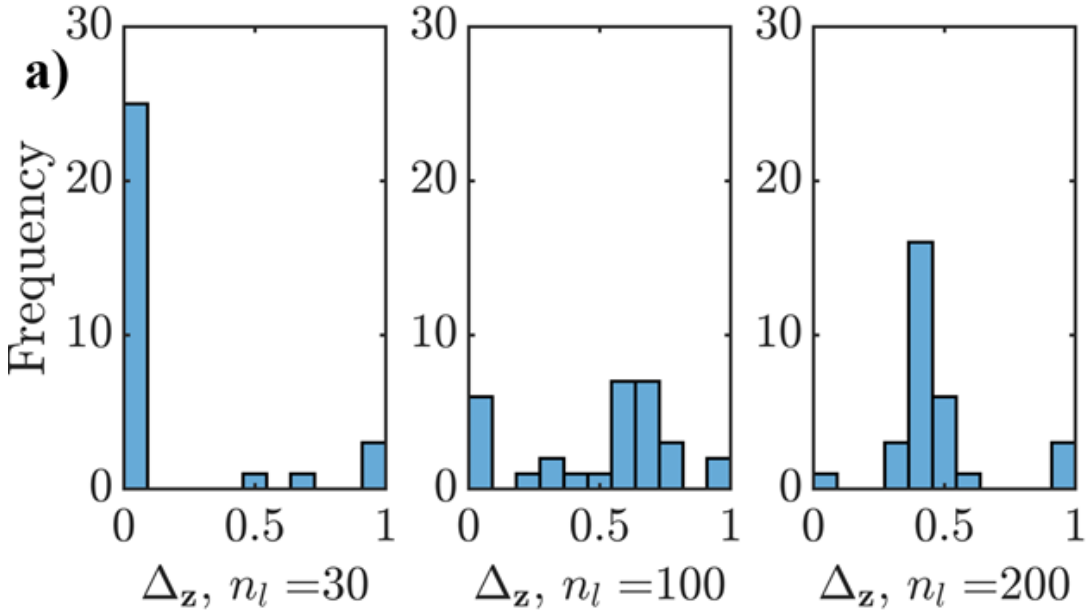
$$y_l(x) = \sin(\theta x), \qquad 0 \le x \le 1 \text{ and } \pi - 2 \le \theta \le 10\pi + 2$$

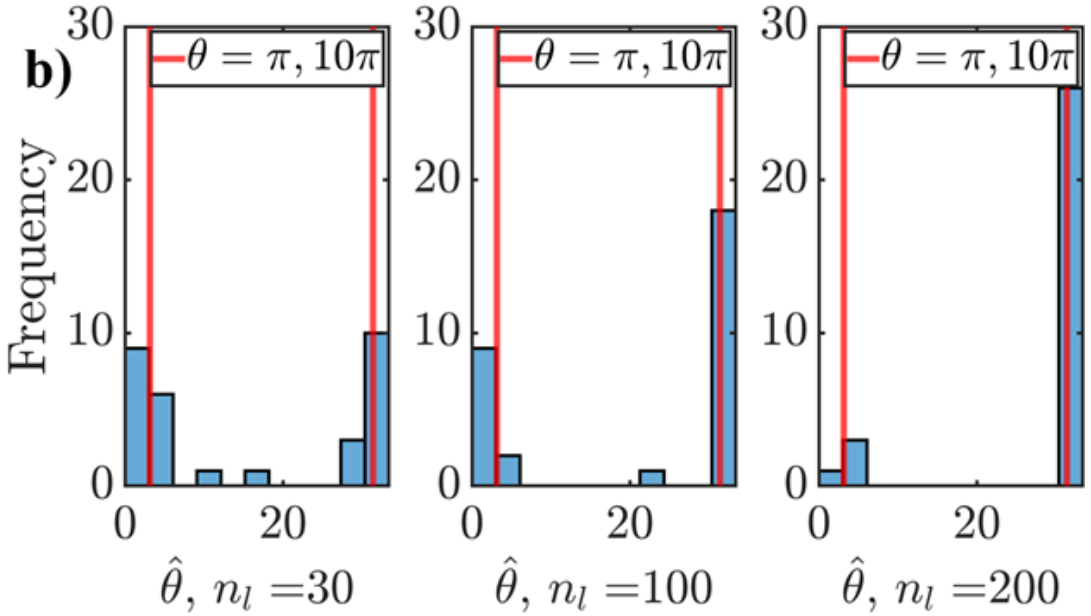
$$n_h = 15$$

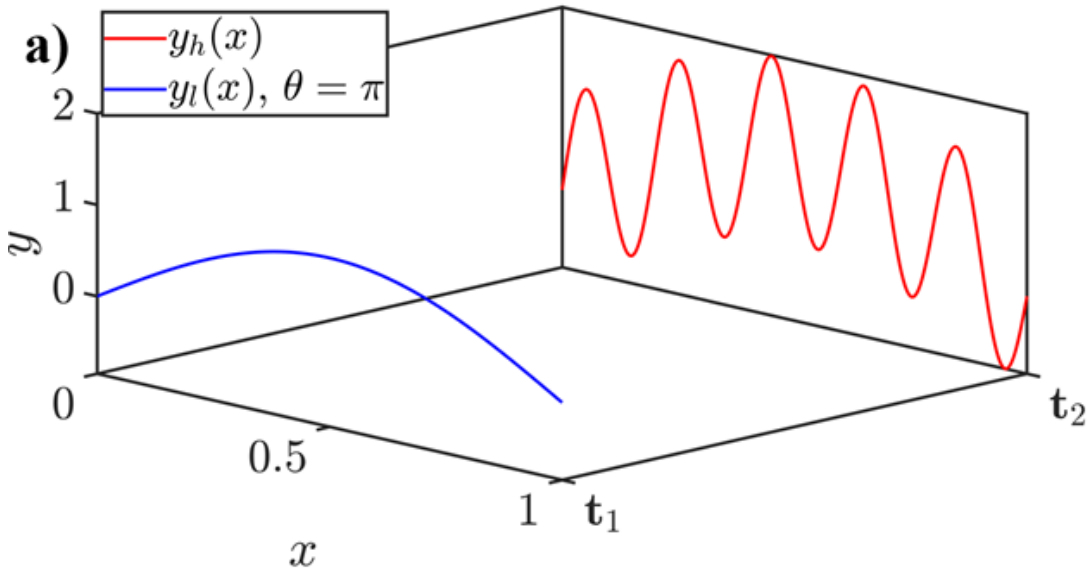


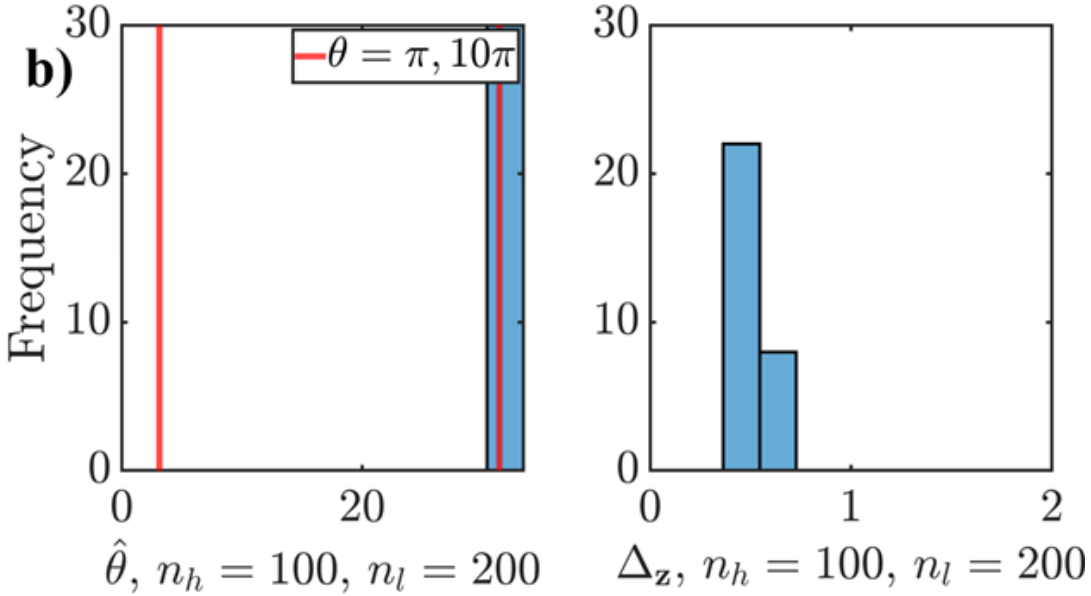


RESULTS: EFFECT OF DATA SIZE



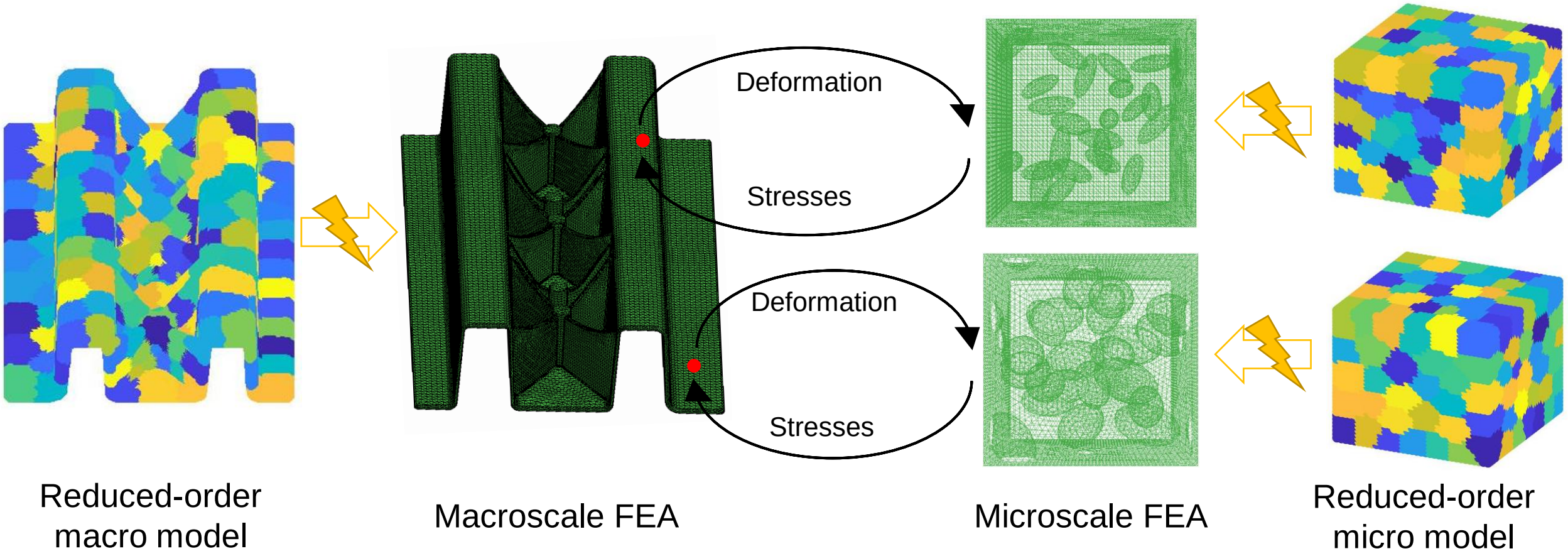






PHYSICS-BASED REDUCED ORDER MODELS (ROMs)

Strike a balance between accuracy and costs by systematically reducing the degrees of freedom.



https://pmacslab.eng.uci.edu/

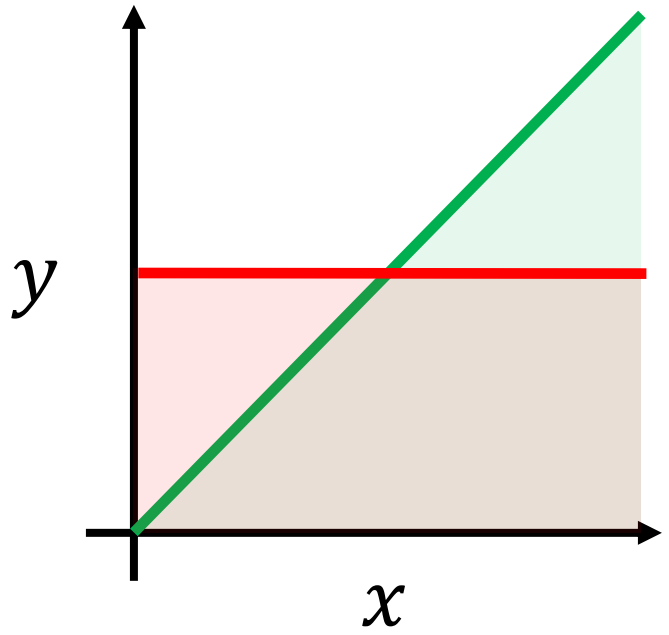
Deng S., et al, Reduced-Order Multiscale Modeling of Plastic Deformations in 3D Alloys with Spatially Varying Porosity by Deflated Clustering Analysis (2022). Computational Mechanics. Accepted.

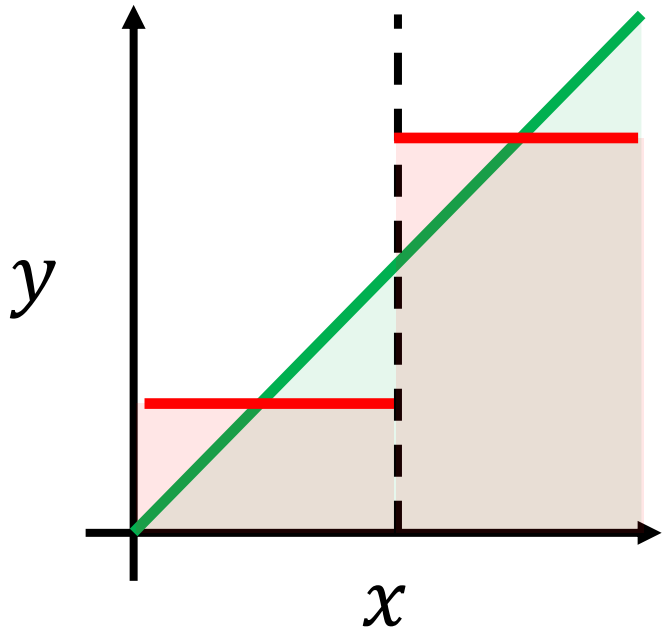
EFFECTS OF CLUSTERING

Clustering reproduces stress/strain fields in an average sense.

Now consider integration: The **average** behavior (i.e., the integral) is accurately captured but the **local** behavior (i.e., max/min value) is not.

Fracture is a localized phenomenon → clustering artificially strengthens the material!

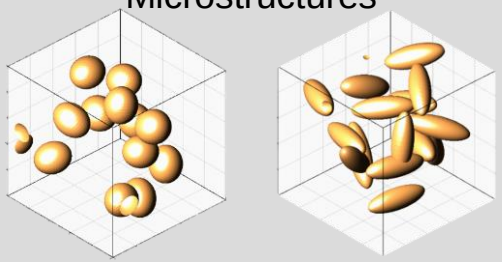




FLOWCHART OF OUR FRAMEWORK

Concurrent multiscale damage model (FE^2)

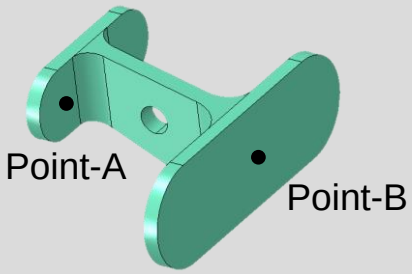
Microstructures



Point-A Point-B

$\mathbf{F}^{\mathrm{M}}$ $\boldsymbol{\sigma}^{\mathrm{M}}_{\mathrm{FEM}}$

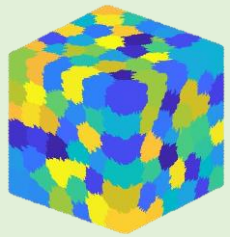
Macroscale component with spatially varying RVE



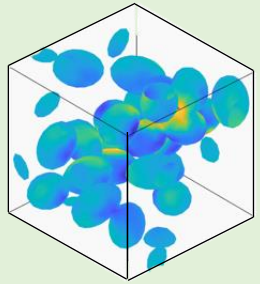
$\boldsymbol{\sigma}^{\mathrm{M}}_{\mathrm{ROM}}$

$\mathbf{F}^{\mathrm{M}}$

Computational analyses via reduced-order model (ROM)

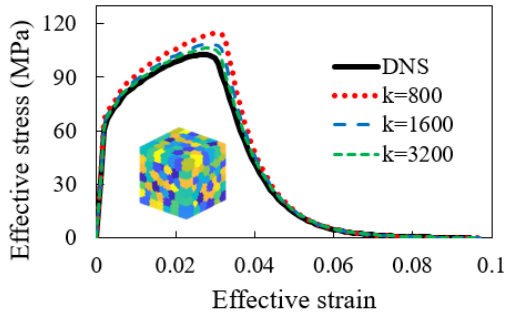


Material clusters with number of clusters (k)



Strain concentrations

Microstructural effective responses



- Calibrated microscale damage parameters $\widehat{\boldsymbol{\theta}}$
- GP emulation of effective tangents $\mathbb{C}^{\mathrm{M}}$

Data-driven calibration via latent map Gaussian process (LMGP)

Microstructure descriptors ($\boldsymbol{q}$)	Data source indicator $t = \{1, 2, 3, 4\}$	Damage (calibration) para. ($\boldsymbol{\theta}$)

LMGP-based surrogate (η)

Damage responses ($\boldsymbol{y}$)

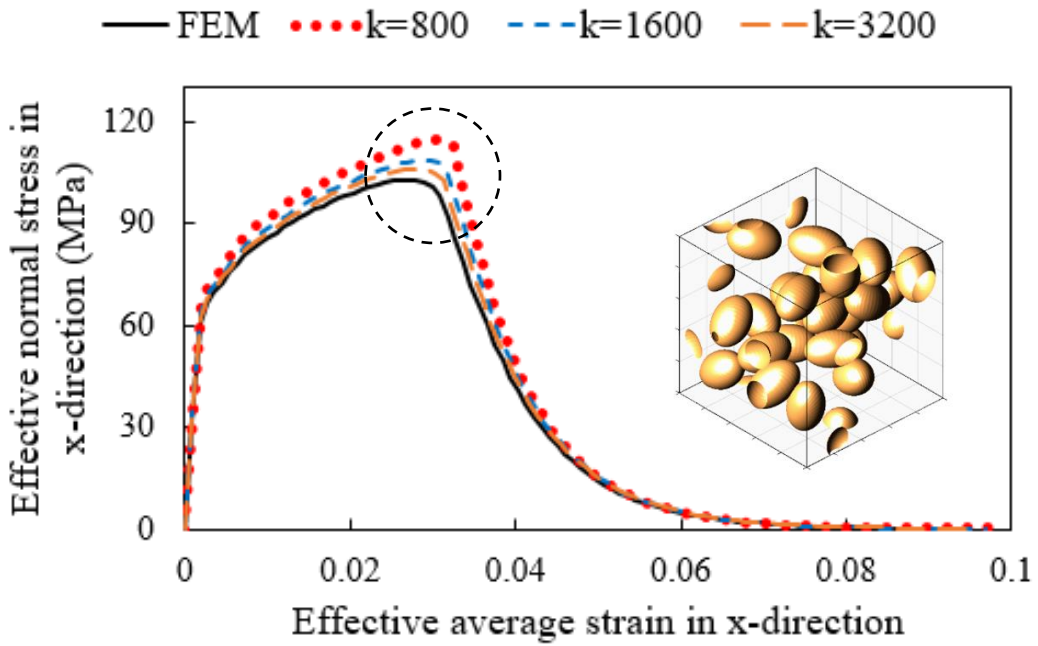
$$\widehat{\boldsymbol{\theta}}(\boldsymbol{q}, j) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\| \eta_{\boldsymbol{y}}(t = j, \boldsymbol{q}, \boldsymbol{\theta}) - \eta_{\boldsymbol{y}}(t = 4, \boldsymbol{q}, \boldsymbol{\theta}_{\boldsymbol{DNS}}) \right\|_2^2$$

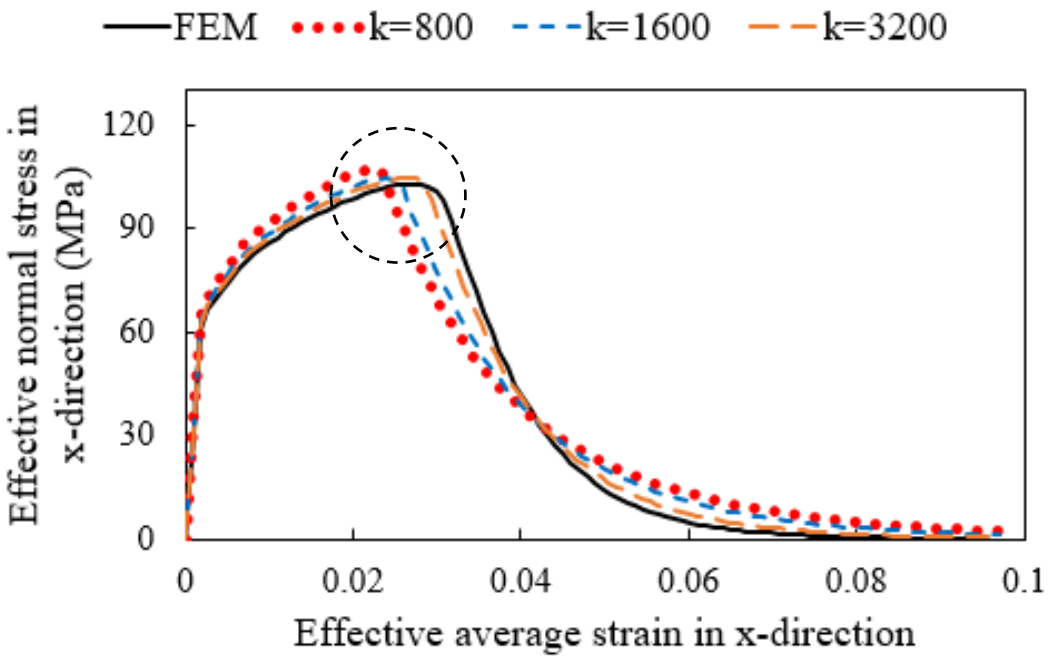
$$\mathrm{ROM}: j = 1, 2, 3;\ \mathrm{DNS}: t = 4$$

Deng, Shiguang, et al. "Data-Driven Calibration of Multi-Fidelity Multiscale Fracture Models." Journal of Mechanical Design. Accepted (2022).

https://pmacslab.eng.uci.edu/

ACCURACY IMPROVEMENT





ROM errors **before** calibration:

- ROM=800: 15.6%
- ROM=1600: 7.99%
- ROM=3200: 4.86%

ROM errors **after** calibration:

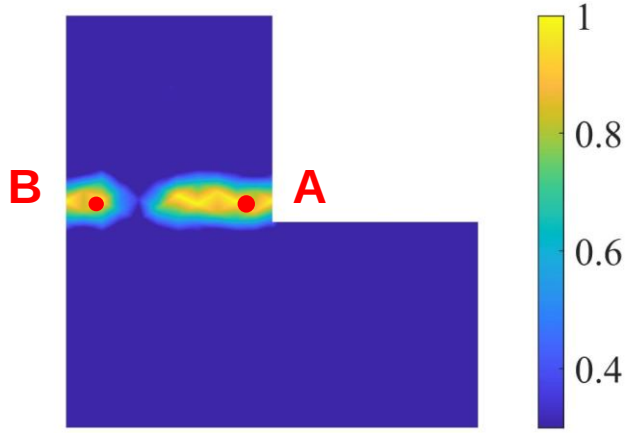
- ROM=800: 3.6%
- ROM=1600: 1.9%
- ROM=3200: 1.7%

Deng, Shiguang, et al. "Data-Driven Calibration of Multi-Fidelity Multiscale Fracture Models." Journal of Mechanical Design. Accepted (2022).

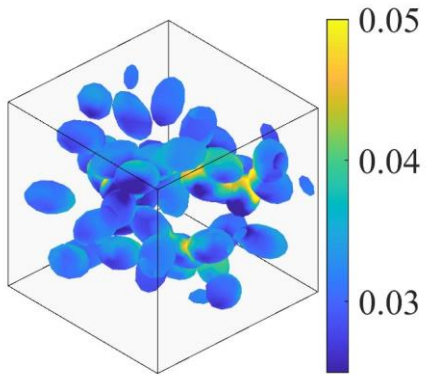
https://pmacslab.eng.uci.edu/

MULTISCALE DAMAGE ANALYSIS

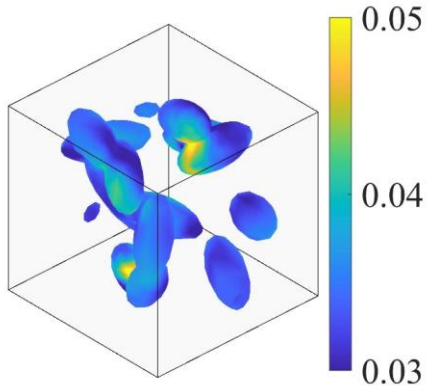
Multiscale solutions:



Macroscale damage pattern

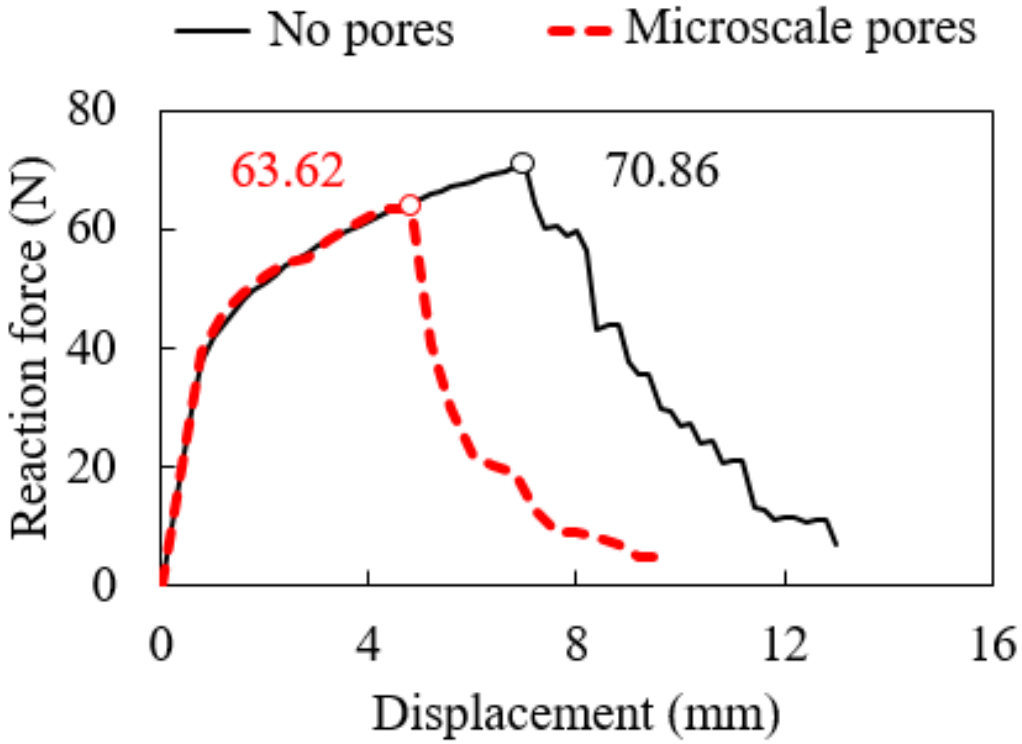


Point-A: micro strain



Point-B: micro strain

Load-displacement response:



Max. load carrying capacity:

- Simulation without pores: 70.9N
- Multiscale with pores: 63.6N (10.2% ↓)

https://pmacslab.eng.uci.edu/

Deng, S., Apelian, D. & Bostanabad, R. Concurrent Multiscale Damage Analysis with Adaptive Spatiotemporal Dimension Reduction. arXiv preprint arXiv:2205.12149 (2022).