Sequential Bayesian Experimental Design for Prediction in Physical Experiments Informed by Computer Models

Hao Zhu

Markus Hainy

Institute of Applied Statistics, Johannes Kepler University Linz, Austria

19 June 2026

JOHANNES KEPLER UNIVERSITÄT LINZ

mODa 14
Model-Oriented Data Analysis and Optimum Design

Sequential Bayesian Experimental Design for Prediction in Physical Experiments Informed by Computer Models

Hao Zhu

Markus Hainy

Institute of Applied Statistics, Johannes Kepler University Linz, Austria

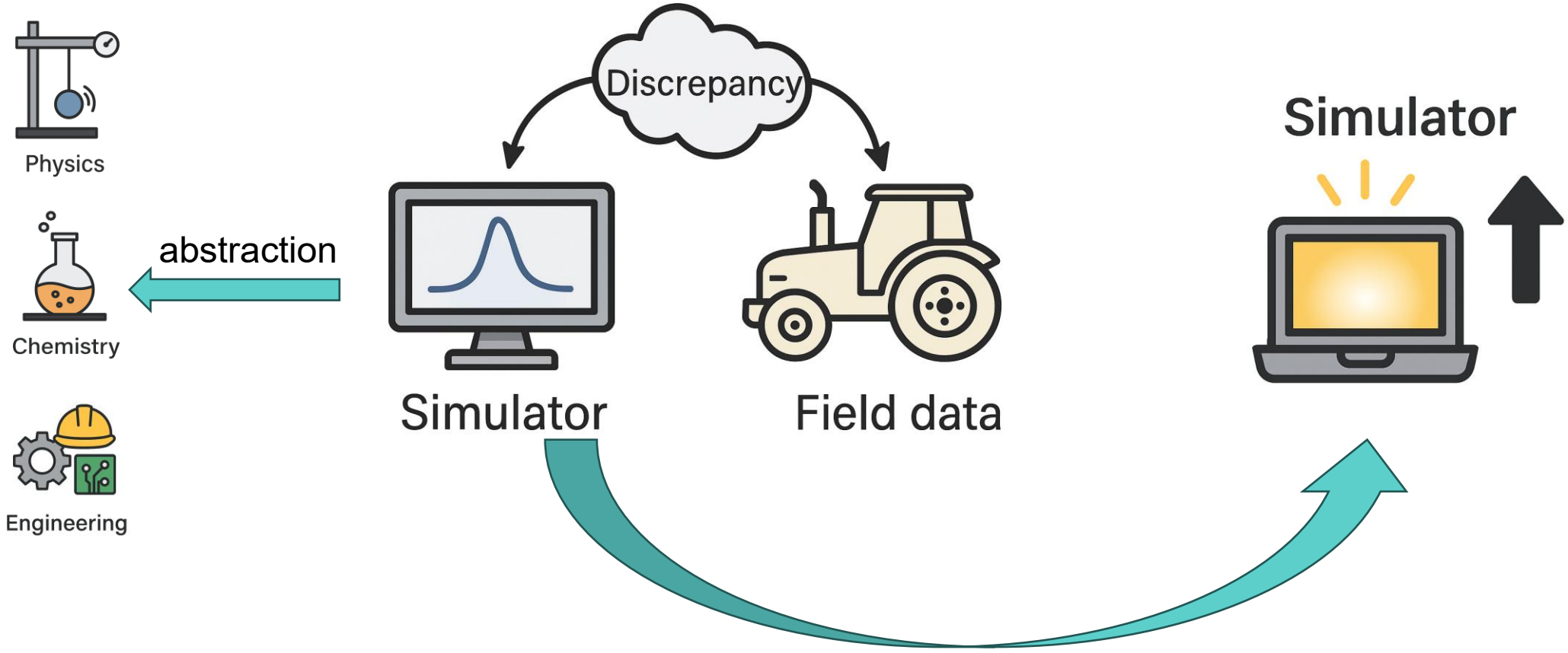
19 June 2026

JOHANNES KEPLER UNIVERSITÄT LINZ

mODa 14
Model-Oriented Data Analysis and Optimum Design

CONTENT

Background of Kennedy & O'Hagan (KOH) Model



Jointly modeling the simulator data and physical experimental data

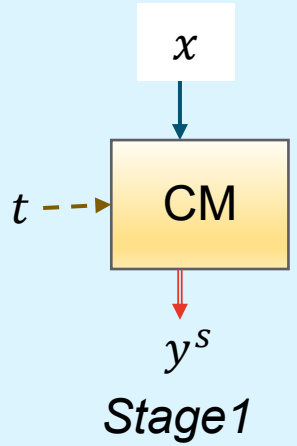
Kennedy & O'Hagan (KOH) Model

Computer Model ------ CM

Two types of input:

- *Experimenter-controlled input (known) --------* x
- *Calibration parameter input: capture unknown features of the system -------* t

Output: y^s



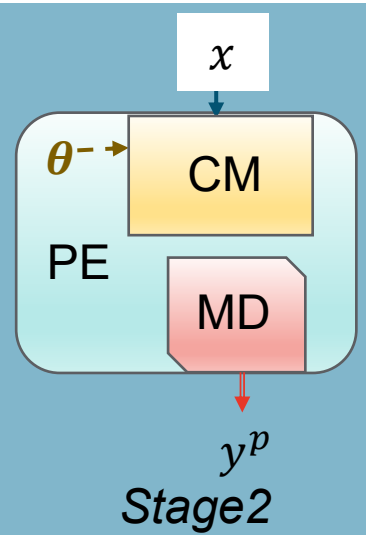
Physical Experiments ------ PE

Model Discrepancy ------ MD

Input: *Experimenter-controlled input (known) --------* x

Parameter: *Calibration parameter* θ

Output: y^p



Kennedy & O'Hagan (KOH) Model

Two-stage model (Kennedy and O'Hagan (2001))

$$y^p(x_i) = y^c(x_i, \theta) + \delta(x_i) + \epsilon_i, \qquad i = 1, \dots, n$$

$$y^s(x_j, t_j) = y^c(x_j, t_j), \qquad j = 1, \dots, m$$

$$y^c(\cdot,\cdot) \sim GP\left(m_c(\cdot,\cdot), K_{\phi_1}(\cdot,\cdot)\right)$$

$$\delta(\cdot) \sim GP\left(m_\delta(\cdot), K_{\phi_2}(\cdot,\cdot)\right)$$

θ: *design calibrated parameter*

design input $x = (x_1, \dots, x_r)$

calibration parameter input $t = (t_1, \dots, t_h)$

$y^s(x_j, t_j)$ *denotes the output from the simulator model*

$y^c(x_i, \theta)$ *denotes the simulator output at input* x *given calibration parameter* θ

$y^p(x_i)$ *denotes the observation from the physical experiment*

$\delta(x_i)$ *captures the model discrepancy*

$\epsilon_i \sim N(0, \lambda^2)$

Uncertainty Quantification

Problem:

- Field data are scarce → Small training data
- Point prediction is not reliable
- Safety margin / risk assessment
- …

Fully Bayesian GP (MCMC)

↓

Bayesian Design

Design Target

Improving **future predictions** by selecting the optimal points to carry out experiments

Candidate points: $\mathbf{D}^{cand} = \{d_1, \ldots, d_J\}$

Future prediction points: $\mathbf{D}^{pred} = \{d_1^{pred}, \ldots, d_K^{pred}\}$

Parameters: $\psi = (\theta, \phi)$ → Hyperparameter

Calibration parameter

Carry out the experiment

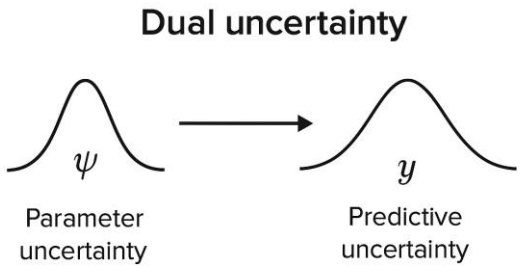
Initial field data : $\mathbf{D}$ → $y^p(\mathbf{D}^{cand})$ → KOH Model → $y^p(\mathbf{D}^{pred})$

Optimally predict the future observations

$$y^p(\mathbf{D}^{pred}) \mid y^p(\mathbf{D}), y^p(\mathbf{D}^{cand}), \psi \sim \mathcal{N}(\mu, \Sigma)$$

Bayesian Design Criteria

Driven by predictive uncertainty reduction:



Mutual information gain maximization:

$$U(d_t) = I\big(y^p(\mathbf{D}^{pred});\, y^p(d_t) \mid y^p(d_{1:t-1}), d_t\big)$$

$$= \int p\big(y^p(\mathbf{D}^{pred}), y^p(d_t) \mid y^p(d_{1:t-1}), d_t\big)\, \log\left(\frac{p\big(y^p(\mathbf{D}^{pred}), y^p(d_t) \mid y^p(d_{1:t-1}), d_t\big)}{p\big(y^p(d_t) \mid d_t, y^p(d_{1:t-1})\big)\, p\big(y^p(\mathbf{D}^{pred}) \mid y^p(d_{1:t-1})\big)} \right) \mathrm{d}\big(y^p(\mathbf{D}^{pred}), y^p(d_t)\big)$$

where

$$p\big(y^p(d_t) \mid d_t,\, y^p(d_{1:t-1})\big) = \int p\big(y^p(d_t) \mid d_t,\, y^p(d_{1:t-1}),\, \boldsymbol{\psi}\big)\, p\big(\boldsymbol{\psi} \mid y^p(d_{1:t-1})\big)\, \mathrm{d}\boldsymbol{\psi},$$

$$p\big(y^p(\mathbf{D}^{pred}) \mid y^p(d_{1:t-1})\big) = \int p\big(y^p(\mathbf{D}^{pred}) \mid y^p(d_{1:t-1}),\, \boldsymbol{\psi}\big)\, p\big(\boldsymbol{\psi} \mid y^p(d_{1:t-1})\big)\, \mathrm{d}\boldsymbol{\psi},$$

$$p\big(y^p(\mathbf{D}^{pred}), y^p(d_t) \mid y^p(d_{1:t-1}), d_t\big) = \int p\big(y^p(\mathbf{D}^{pred}), y^p(d_t) \mid y^p(d_{1:t-1}), d_t,\, \boldsymbol{\psi}\big)\, p\big(\boldsymbol{\psi} \mid y^p(d_{1:t-1})\big)\, \mathrm{d}\boldsymbol{\psi}$$

Complexity–MI Hybrid Criterion

MI-only limitation:

1. May over-sample in high variance regions.

2. Miss critical structure (extrema, inflection)

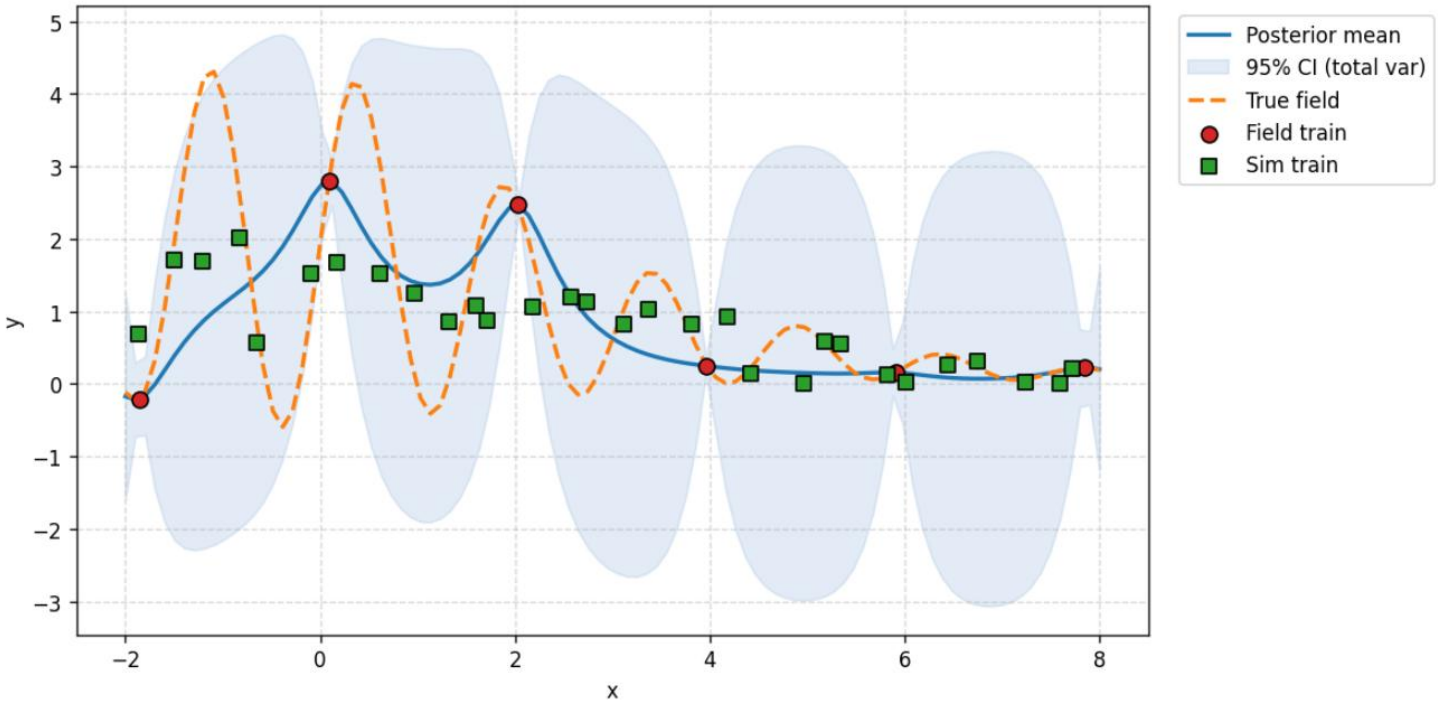
Compute complexity measure at each candidate point
- Based on $\|\nabla \mu(x)\|$
- Rate of slope change

↓

Standardize over candidates
- MI gain
- Complexity $\rightarrow \bar{C}(d)$

↓

Score
Choose design point with max Score



$$score_{mi+com}(d_t) = w_{\mathrm{MI}} \cdot \widehat{U}(d_t \mid d_{1:t-1}) + w_C \cdot C(d_t \mid d_{1:t-1})$$

Accelerating Mutual Information Computation via Gaussian Mixture Reduction

Direct double loop Monte Carlo is expensive

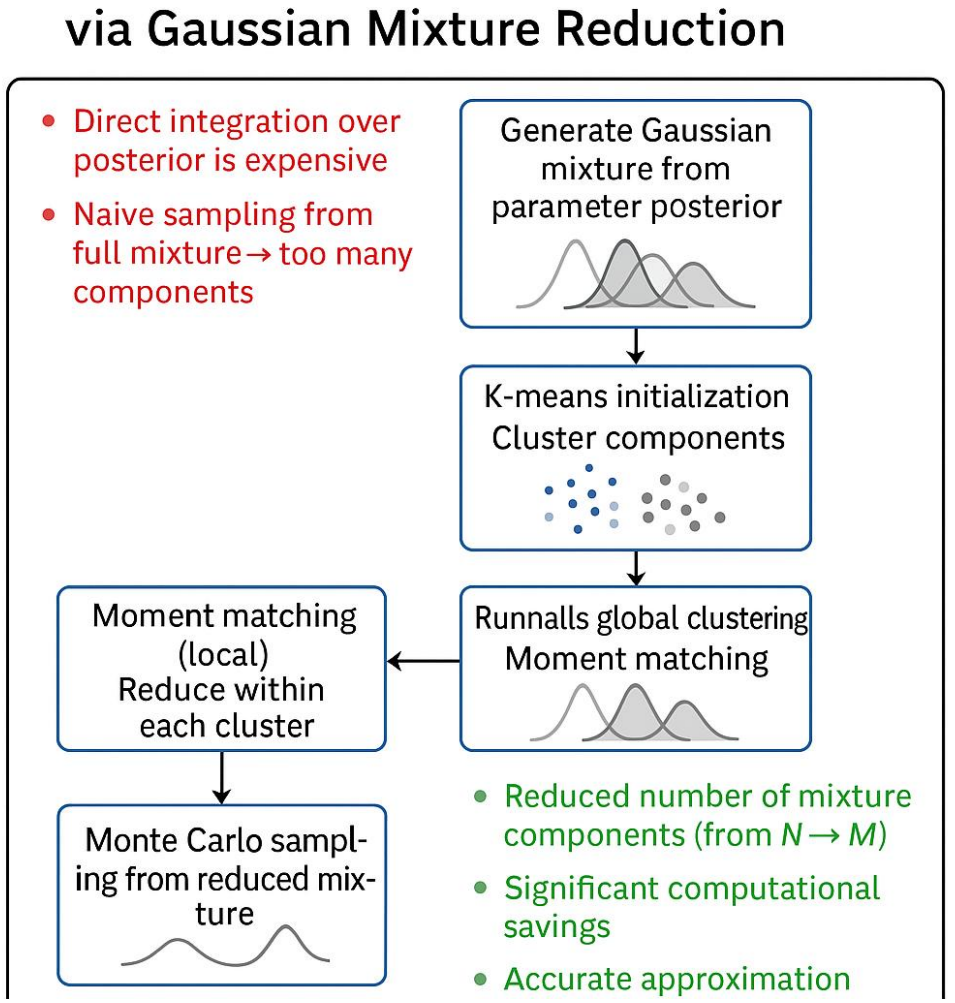
Gaussian Mixture Reduction:

If we have MCMC sample of 1000 sets of $\psi = (\theta, \phi)$

Each set of them will generate a predictive Gaussian distribution:

$$y^p(\mathbf{D}^{pred}) \mid y^p(\mathbf{D}), y^p(\mathbf{D}^{cand}), \psi \sim \mathcal{N}(\mu, \Sigma)$$

If 1000-component Gaussian mixture model can be reduced to 200,100,50….



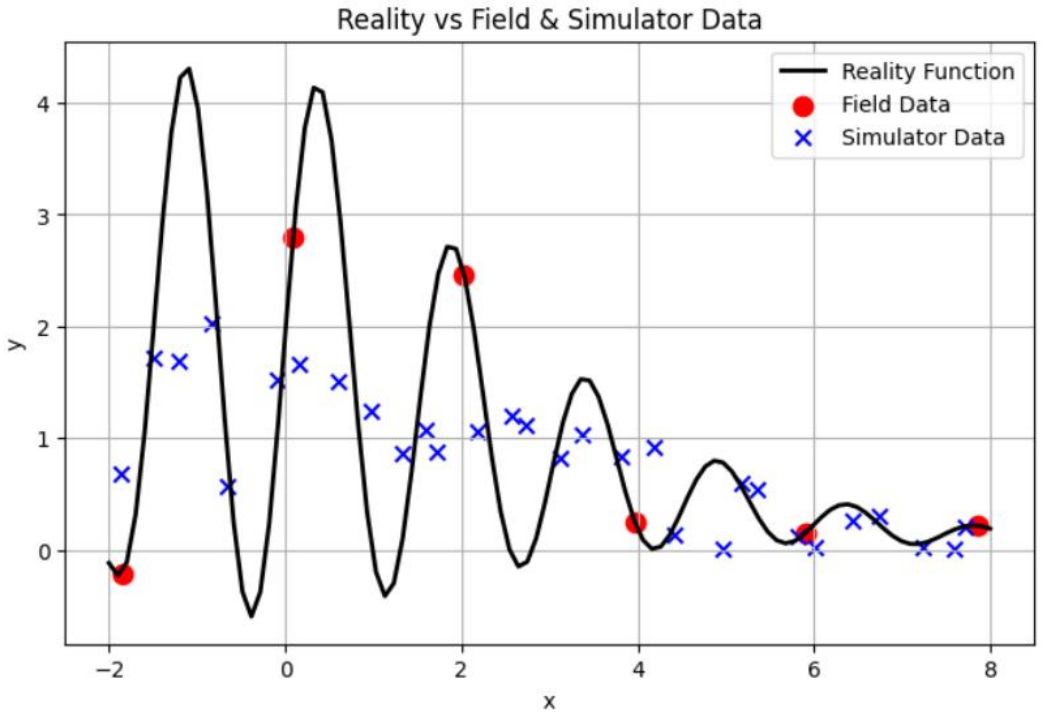
Schieferdecker, D., & Huber, M. F. (2009)

Simulation Example

One numerical example:

Computer model: $f(x;\alpha,\beta) = 10 \cdot Gamma(x + 2.2;\alpha,\beta)$ $\alpha: shape > 0, \beta: scale > 0$

Physical model: $f(x) = 10 \cdot Gamma(x + 2.2; 1.8, 2.0) + Gamma(x + 2.2; 1.8, 2.0) \cdot \sin(2\pi x) + \epsilon$



Data:
- Field data: 6 points
- Simulator data: 30 points

Input domain: $x \in (-2,8)$

Parameters:
- Shape α: True value 1.8
- Scale β: True value 2.0
- Kernel hyperparameters

Weakly informative priors

Gaussian process models:
1. Simulator data: stage 1 GP uses RBF kernel.
2. Discrepancy: stage 2 GP uses RBF kernel × periodic kernel.

Sequential & Adaptive Design

Sequential/Greedy (offline) and adaptive (online) design are both multi-round designs.

Initialize: set of selected design points $\mathbf{D}^* \leftarrow \emptyset$

For i (round) = 1,…,N:

1. **Evaluate criterion**: Compute $\boldsymbol{Criterion\,(d)}$ for all $d \in \mathbf{D}^{cand} \setminus \mathbf{D}^*$ using calibrated KOH Model.

2. **Select** $d^* = \arg\max \boldsymbol{Criterion\,(d)}$ (or $\arg\min \boldsymbol{Criterion\,(d)}$)

3. **Update:**
   If sequential (offline): $\mathbf{D}^* \cup \{d^*\}$
   if adaptive (online):
   1. Run experiment at d^*, obtain y^*
   2. Initial field data $\mathbf{D} \cup \{(d^*, y^*)\}$
   3. Recalibrate KOH model: update $\boldsymbol{\psi} = (\boldsymbol{\theta}, \boldsymbol{\phi})$
   4. $\mathbf{D}^* \cup \{d^*\}$

Setup

- Number of experimental rounds: 10 rounds

- Candidate design points $\mathbf{D}^{cand} = \{d_1, \ldots, d_J\}$ 50 points uniformly distributed over the interval (−1.98, 7.98).
- Prediction points DT: 100 points $\mathbf{D}^{pred} = \{d_1^{pred}, \ldots, d_K^{pred}\}$ uniformly distributed over the interval (−2, 8).

- Design criteria to compare:
  1. MI (mutual information maximization)
  2. MI-complexity hybrid
  3. IMSPE (Integrated mean squared prediction error minimization)
  4. IMSPE-complexity hybrid
  5. D-optimality: better calibration of model parameters θ
  6. Maximin distance: space-filling benchmark, greedily add the candidate whose minimal Euclidean distance to the existing design set is largest.

Pseudo Bayesian

Setup

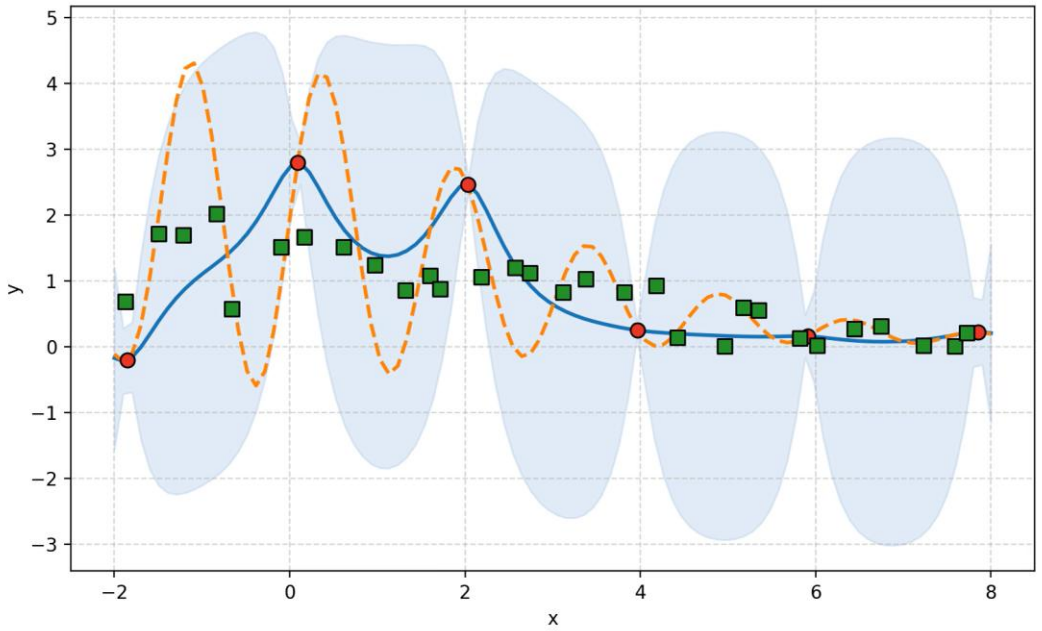
- Software:
  Pytorch, Pyro

- MCMC:
  NUTS HMC algorithm
  1000 posterior parameter samples of hyperparameters and calibration parameters $\psi = (\theta, \phi)$

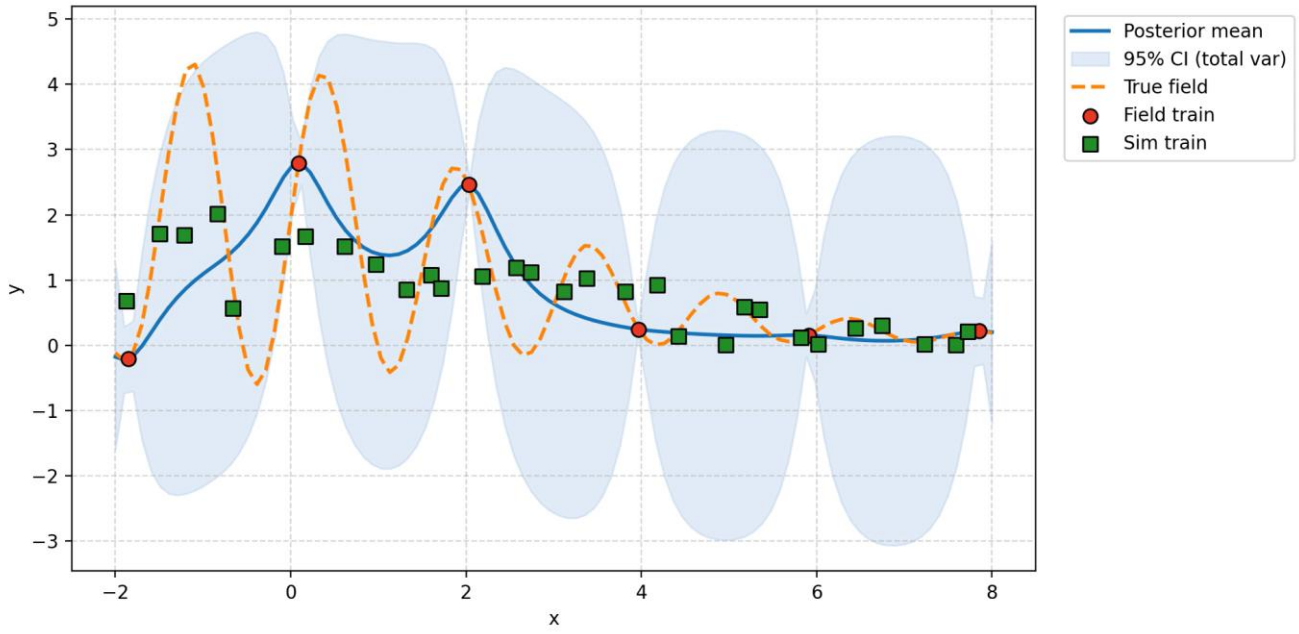
- Outer sample for y: 10000

- Computation:
  1. Use **vectorized operations** to avoid explicit for loops
  2. Cache big matrices and pre-computations in sequential (offline) design (e.g. Schur complement)
  3. **Gaussian Mixture Reduction via clustering**
  ….

Adaptive Design for MI and MI-Complexity Hybrid Criteria

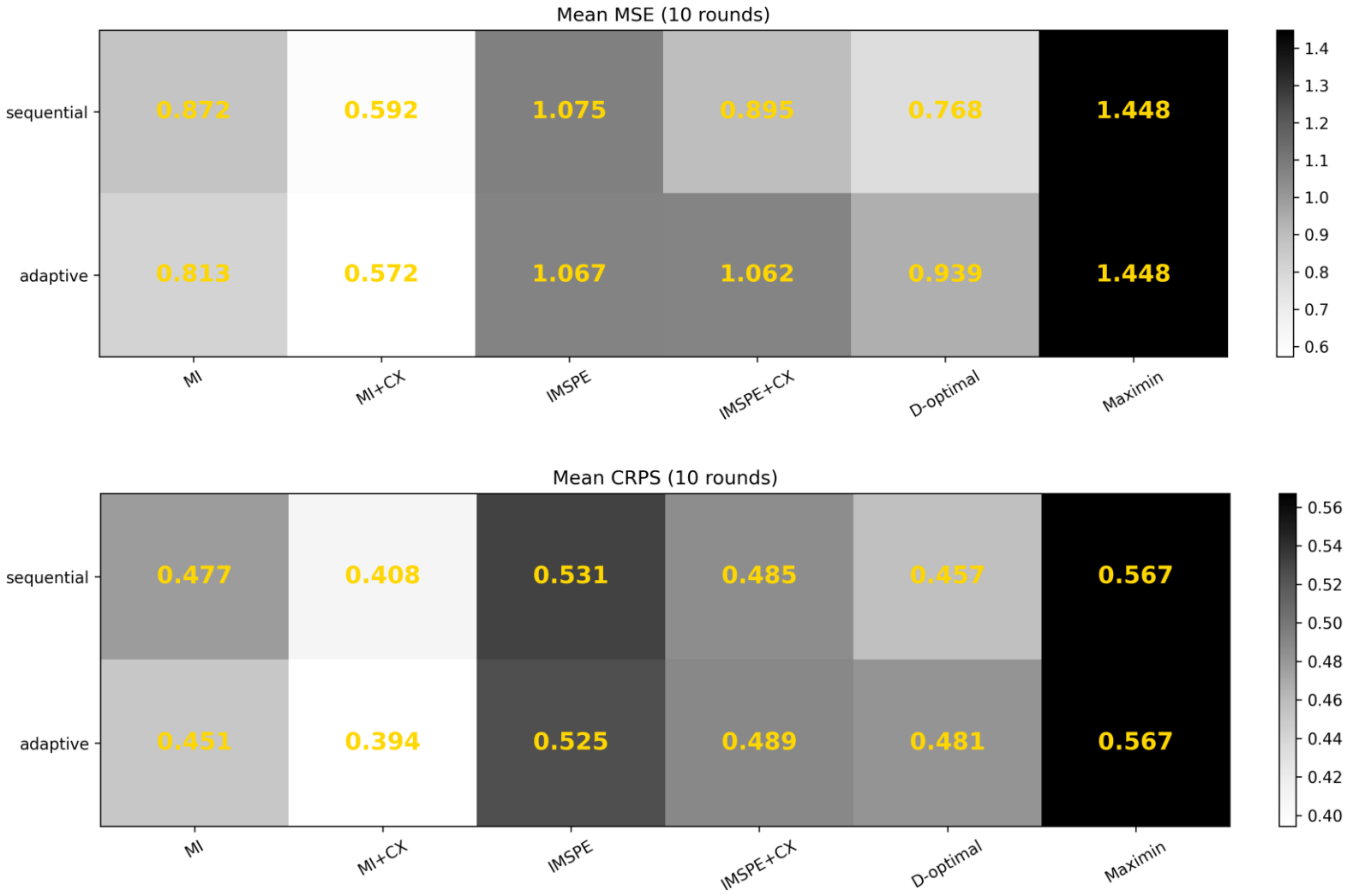


MI

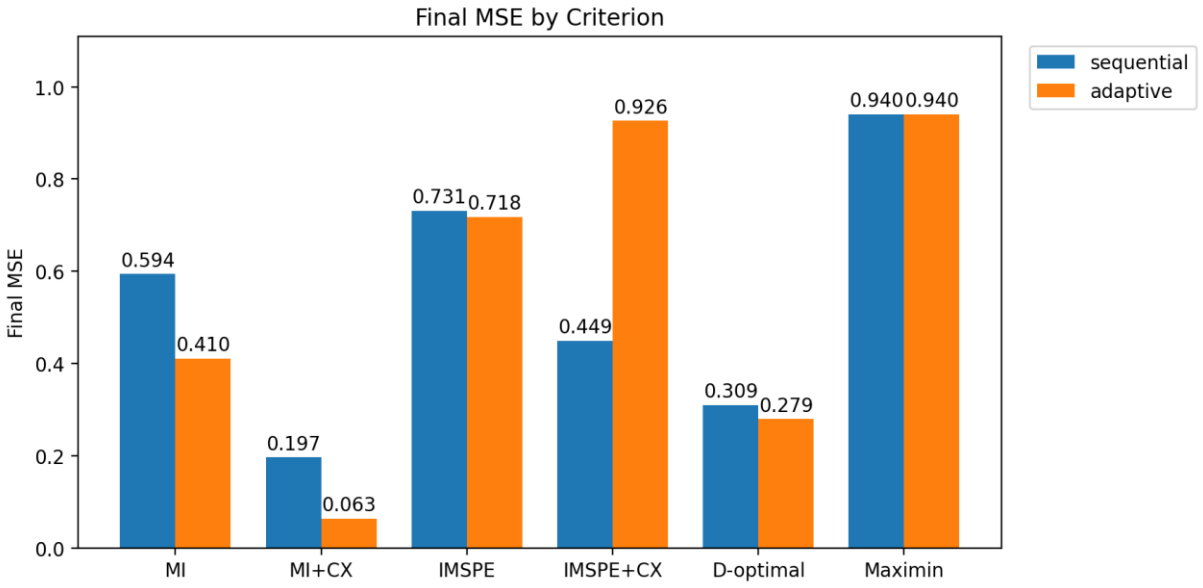


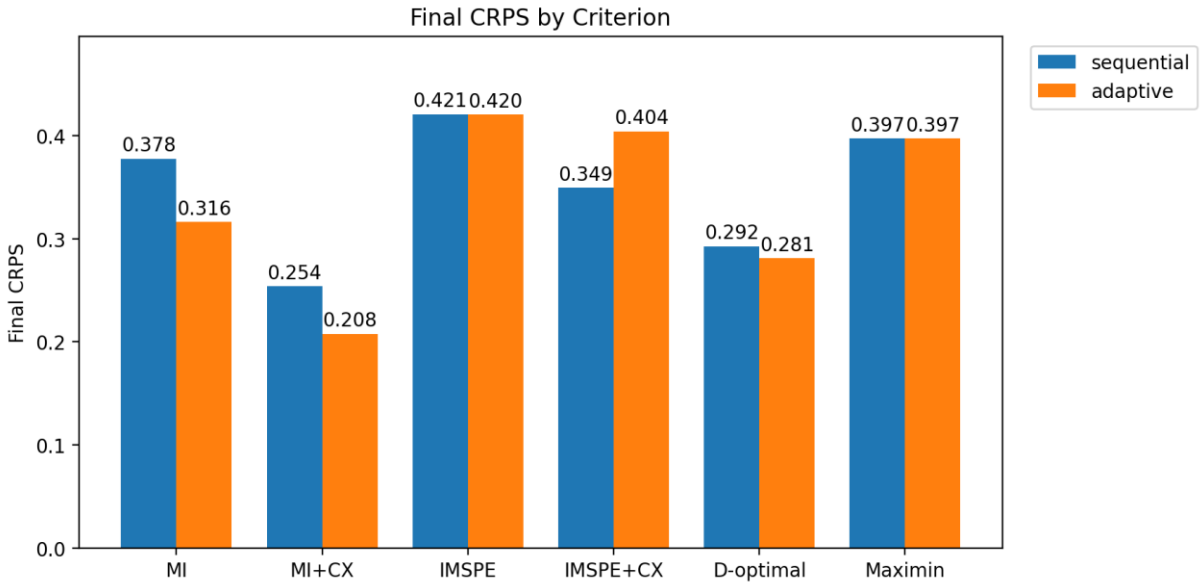
MI-Complexity hybrid

Average Performance Across Multiple Criteria



Final-Round Performance Across Multiple Criteria





Criteria	Running Time (10 rounds sequential design)
MI (naive NMC)	901.4012s
MI (Gaussian mixture reduction)	142.0512s
MI-complexity hybrid (naive NMC)	902.8288s
MI-complexity hybrid (Gaussian mixture reduction)	143.1869s
IMSPE	10.0309s
IMSPE-complexity hybrid	10.7311 s
D-optimality	295.0461s
Maximin distance	0.0168s

Comparison of running time of 10 rounds sequential/greedy (offline) design for each criterion

Real-World Example

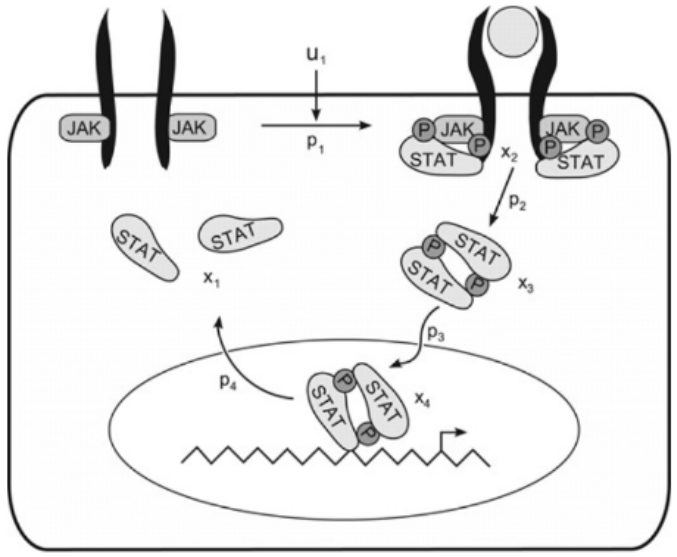


Fig. 2. Model of the JAK-STAT pathway. In this model u_1 serves as driving input, while the total concentration of STAT ($x_1+x_2+2x_3$) and the total concentration of phosphorylated STAT in the cytoplasm (x_2+2x_3) were measured. Note that the step from x_4 back to x_1 is associated with a delay.

Swameye et al. 2003

Biological background:

We study the JAK–STAT5 signalling pathway (Swameye et al. 2003) in which

$$v_1, v_2, v_3, v_4$$

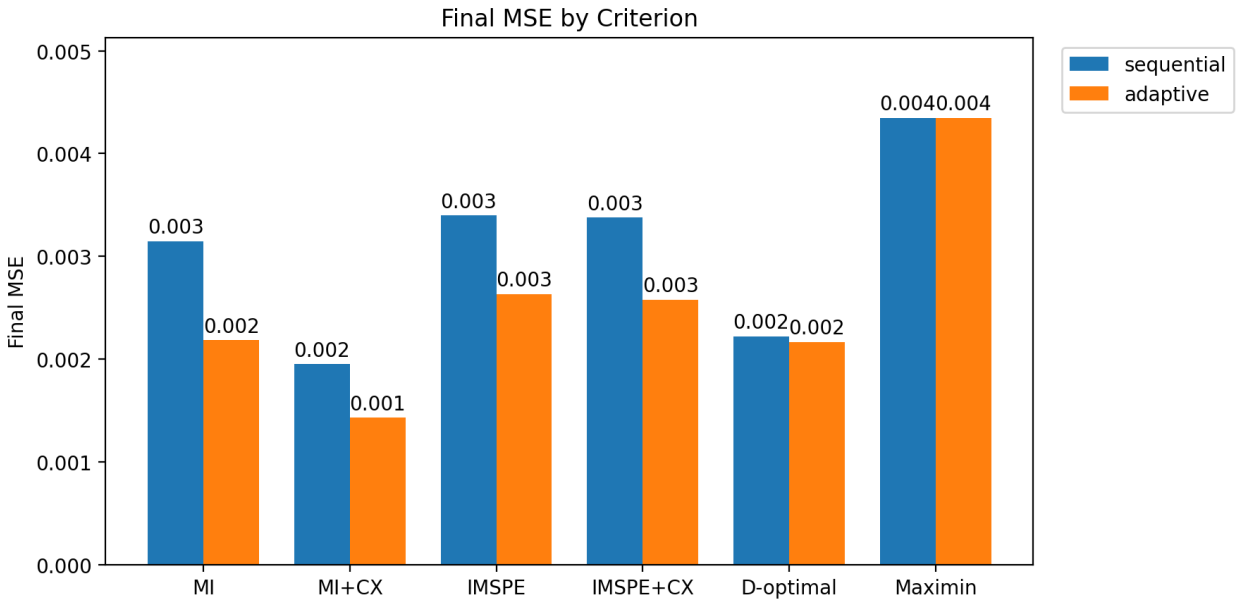
are cytoplasmic unphosphorylated STAT5, phosphorylated STAT5 monomer, dimer, and nuclear STAT5, respectively.

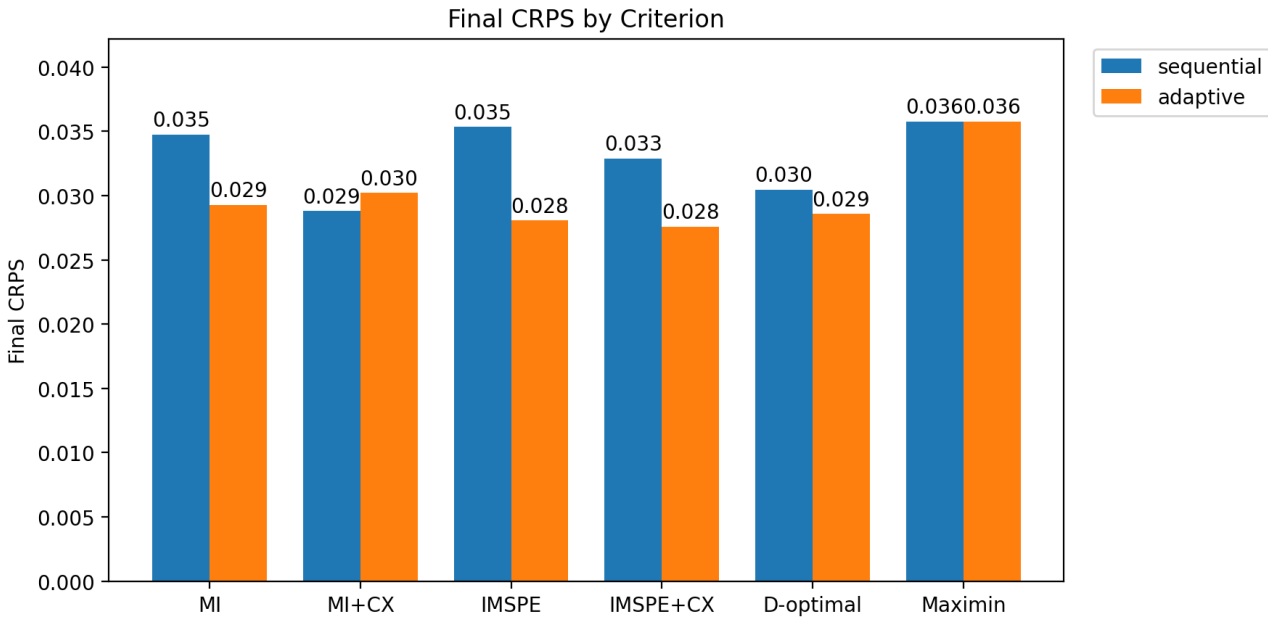
ODE system (simulator):

$$\frac{dv_1}{dt} = -p_1\, u_1\, v_1 + 2\, p_4\, v_4,$$
$$\frac{dv_2}{dt} = p_1\, u_1\, v_1 - p_2\, v_2^2,$$
$$\frac{dv_3}{dt} = -p_3\, v_3 + 0.5\, p_2\, v_2^2,$$
$$\frac{dv_4}{dt} = p_3\, v_3 - p_4\, v_4,$$

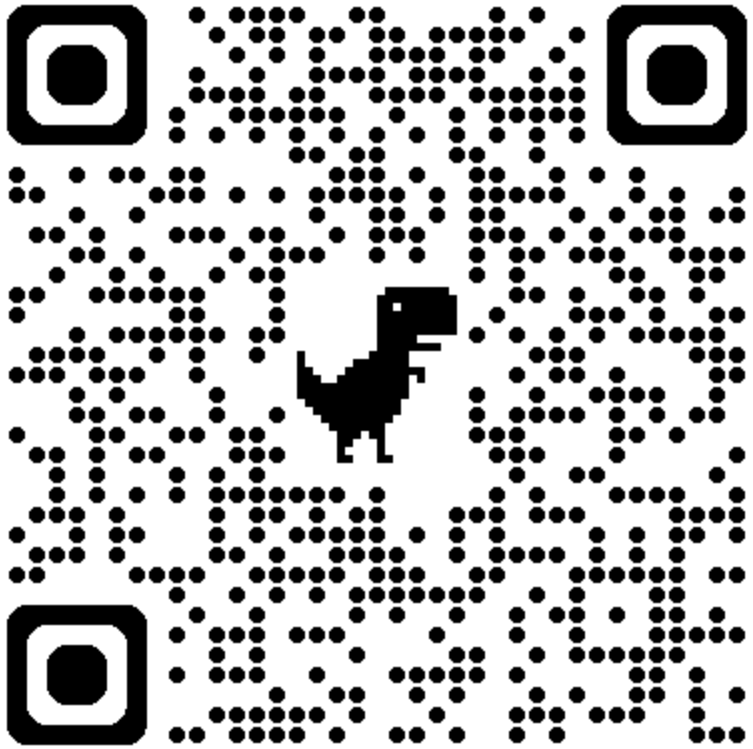
Here u_1 is the cytokine input, and $\theta = (p_1, p_2, p_3, p_4)$ are calibration parameters.

Final-Round Performance Across Multiple Criteria





Scan for Full Paper



References

Kennedy, M.C. & O'Hagan, A. (2001): Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464

Schieferdecker, D. & Huber, M.F. (2009): Gaussian mixture reduction via clustering. In: *Proceedings of the 12th International Conference on Information Fusion (FUSION 2009)*, pages 1536 – 1543, IEEE: Seattle (WA)

Swameye, I., Müller, T.G., Timmer, J., Oehler, S. & Klingmüller, U. (2003): Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences of the United States of America* 100(3):1028–1033

Zhu, H. & Hainy, M. (2026): Sequential Bayesian experimental design for prediction in physical experiments informed by computer models. *arXiv:2603.16756*