Sequential Bayesian experimental design for calibration of expensive simulation models

Özge Sürer, Matthew Plumlee, & Stefan M. Wild

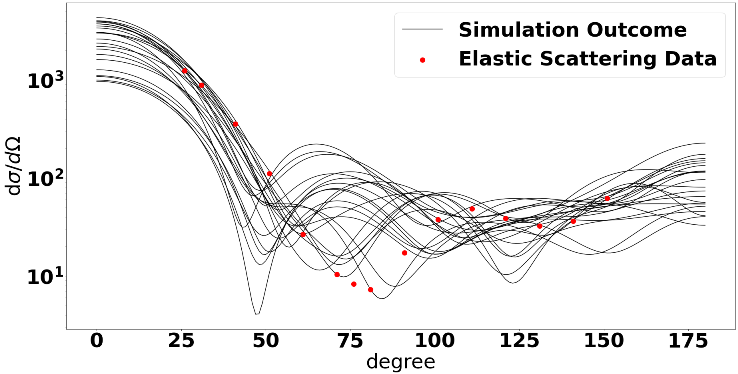
Miami University

ISNET, May 2023

Authors gratefully acknowledge support from the NSF CSSI program under grant OAC-2004601.

Simulation Models to Understand Complex Processes

- Requires significant computation time
  - more complex phenomena require more time-consuming simulations!



Model Calibration

- Let $(\mathbf{x}_1, \ldots, \mathbf{x}_d)$ be the design points where the data $\mathbf{y} = (y(\mathbf{x}_1), \ldots, y(\mathbf{x}_d))$ is collected
- Seek parameter vector(s) $\boldsymbol{\theta} \in \mathbb{R}^p$ to align simulator outcomes $\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta(\mathbf{x}_1, \boldsymbol{\theta}), \ldots, \eta(\mathbf{x}_d, \boldsymbol{\theta})) \in \mathbb{R}^d$ with observation $\mathbf{y} \in \mathbb{R}^d$
- Observation $\mathbf{y}$ can be modeled using the expensive simulation $\boldsymbol{\eta}(\boldsymbol{\theta})$

$$\mathbf{y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

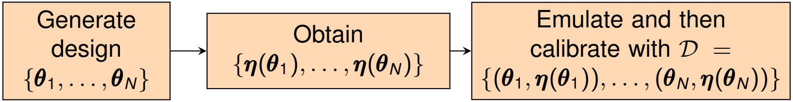
- Posterior distribution is computed by using Bayes' rule

$$\underbrace{p(\boldsymbol{\theta}|\mathbf{y})}_{\text{posterior}} \propto \tilde{p}(\boldsymbol{\theta}|\mathbf{y}) = \underbrace{p(\mathbf{y}|\boldsymbol{\theta})}_{\text{likelihood}} \underbrace{p(\boldsymbol{\theta})}_{\text{prior}}$$

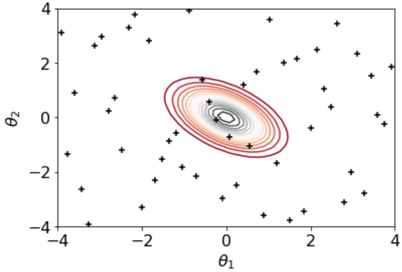
  - Prior $p(\boldsymbol{\theta})$ can be computed for any $\boldsymbol{\theta}$ (i.e., independent of simulation)
  - Likelihood requires expensive simulator evaluation for a given $\boldsymbol{\theta}$

Model Calibration

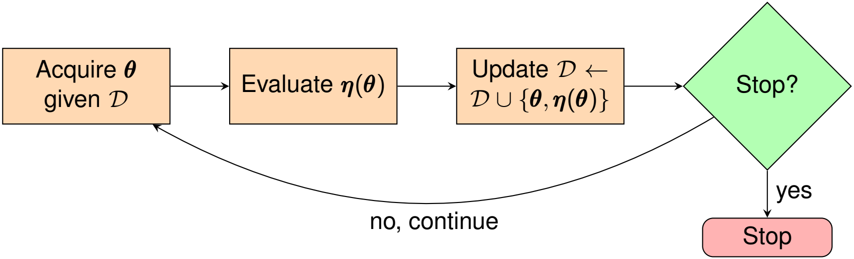
- MCMC methods require too many evaluations of expensive simulations
- Alternatively, one builds a cheaper emulator as a proxy to a simulator, and then leverages MCMC sampling to obtain draws from the posterior



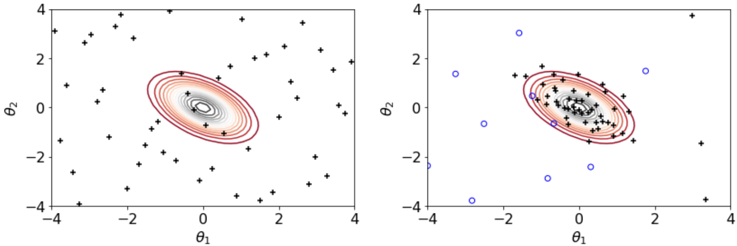
Example of 50 samples using LHS



Sequential Bayesian Experimental Design



LHS vs. proposed approach



Sequential Calibration Approach

Algorithm: Sequential Bayesian experimental design

Initialize $\mathcal{D}_1 = \{(\boldsymbol{\theta}_i, \boldsymbol{\eta}(\boldsymbol{\theta}_i)) : i = 1, \ldots, n_0\}$

for $t = 1, \ldots, n$ **do**

  Fit an emulator with $\mathcal{D}_t$

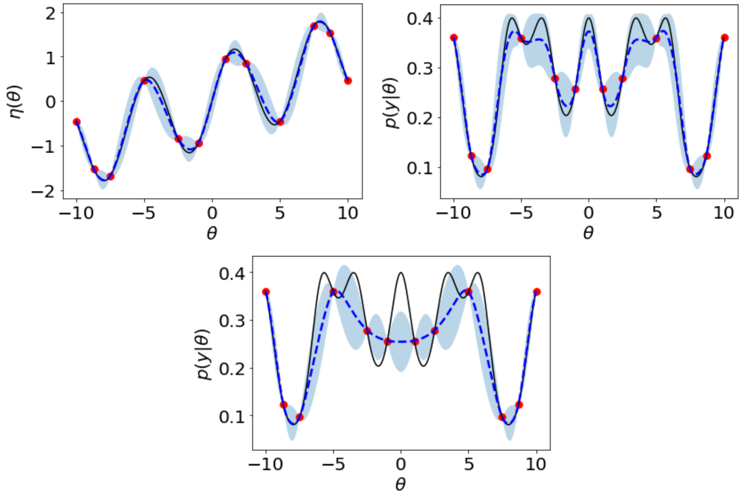
  Generate candidate solutions $\mathcal{L}_t$

  Select $\boldsymbol{\theta}^{\text{new}} \in \underset{\boldsymbol{\theta}^* \in \mathcal{L}_t}{\arg\min}\, \mathcal{A}_t(\boldsymbol{\theta}^*)$

  Evaluate $\boldsymbol{\eta}(\boldsymbol{\theta}^{\text{new}})$

  Update $\mathcal{D}_{t+1} \leftarrow \mathcal{D}_t \cup (\boldsymbol{\theta}^{\text{new}}, \boldsymbol{\eta}(\boldsymbol{\theta}^{\text{new}}))$

Emulation of Simulation Output



Gaussian Process Model for High-Dimensional Simulation Output

$$\Xi_t = \begin{bmatrix} \eta(\mathbf{x}_1, \boldsymbol{\theta}_1) & \cdots & \eta(\mathbf{x}_1, \boldsymbol{\theta}_{n_t}) \\ \eta(\mathbf{x}_2, \boldsymbol{\theta}_1) & \cdots & \eta(\mathbf{x}_2, \boldsymbol{\theta}_{n_t}) \\ \vdots & \vdots & \vdots \\ \eta(\mathbf{x}_d, \boldsymbol{\theta}_1) & \cdots & \eta(\mathbf{x}_d, \boldsymbol{\theta}_{n_t}) \end{bmatrix}$$

GP-based emulator that employs the basis vector approach that is now standard practice (e.g., Higdon et al., 2008)

- Transform d-dimensional Ξ_t into q-dimensional space
  - $\mathbf{W}_t = [\mathbf{w}_{t,1}, \ldots, \mathbf{w}_{t,q}] = \mathbf{B}_t^\top \Xi_t$
  - $\mathbf{B}_t = [\mathbf{b}_1, \ldots, \mathbf{b}_q]$ stores the orthonormal basis vectors
- Build q independent GPs on the latent space

Gaussian Process Model for High-Dimensional Simulation Output

An independent GP is used for each latent output $w_j(\cdot)$ with mean $m_{t,j}(\cdot)$ and variance $\varsigma^2_{t,j}(\cdot)$ such that

$$w_j(\boldsymbol{\theta})|\mathbf{w}_{t,j} \sim \mathrm{N}\left(m_{t,j}(\boldsymbol{\theta}), \varsigma^2_{t,j}(\boldsymbol{\theta})\right), \qquad j = 1, \dots, q,$$

where

$$m_{t,j}(\boldsymbol{\theta}) = \mathbf{k}_j(\boldsymbol{\theta}, \boldsymbol{\theta}_{1:n_t})\mathbf{K}_j(\boldsymbol{\theta}_{1:n_t})^{-1}\mathbf{w}_{t,j}$$
$$\varsigma^2_{t,j}(\boldsymbol{\theta}) = \mathbf{k}_j(\boldsymbol{\theta}, \boldsymbol{\theta}) - \mathbf{k}_j(\boldsymbol{\theta}, \boldsymbol{\theta}_{1:n_t})\mathbf{K}_j(\boldsymbol{\theta}_{1:n_t})^{-1}\mathbf{k}_j(\boldsymbol{\theta}_{1:n_t}, \boldsymbol{\theta})$$

Transfer predictions back to the original space such that predictive distribution of the emulator output is

$$\boldsymbol{\eta}(\boldsymbol{\theta})|\mathcal{D}_t \sim \mathrm{MVN}\left(\boldsymbol{\mu}_t(\boldsymbol{\theta}), \mathbf{S}_t(\boldsymbol{\theta})\right)$$

Posterior Inference

Goal is to better learn

$$\tilde{p}(\boldsymbol{\theta}|\mathbf{y}) = p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

where

$$p(\mathbf{y}|\boldsymbol{\theta}) = (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\eta}(\boldsymbol{\theta}))^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\eta}(\boldsymbol{\theta}))\right).$$

Lemma

Assuming that the covariance matrices $\boldsymbol{\Sigma}$ and $\mathbf{S}_t(\boldsymbol{\theta})$ are positive definite,

$$\mathbb{E}[\tilde{p}(\boldsymbol{\theta}|\mathbf{y})|\mathcal{D}_t] = f_{\mathcal{N}}\left(\mathbf{y};\, \boldsymbol{\mu}_t(\boldsymbol{\theta}),\, \boldsymbol{\Sigma} + \mathbf{S}_t(\boldsymbol{\theta})\right) p(\boldsymbol{\theta}),$$

$$\mathbb{V}[\tilde{p}(\boldsymbol{\theta}|\mathbf{y})|\mathcal{D}_t] = \left(\frac{1}{2^d\pi^{d/2}|\boldsymbol{\Sigma}|^{1/2}} f_{\mathcal{N}}\left(\mathbf{y};\, \boldsymbol{\mu}_t(\boldsymbol{\theta}),\, \frac{1}{2}\boldsymbol{\Sigma} + \mathbf{S}_t(\boldsymbol{\theta})\right) - \left(f_{\mathcal{N}}\left(\mathbf{y};\, \boldsymbol{\mu}_t(\boldsymbol{\theta}),\, \boldsymbol{\Sigma} + \mathbf{S}_t(\boldsymbol{\theta})\right)\right)^2\right) p(\boldsymbol{\theta})^2.$$

Expected Integrated Variance for Calibration

EIVAR criterion is calculated by

$$\mathcal{A}_t(\boldsymbol{\theta}^*) = \int_{\Theta} \mathbb{E}_{\boldsymbol{\eta}^*|\mathcal{D}_t} \left(\mathbb{V}[p(\mathbf{y}|\boldsymbol{\theta}) \,|(\boldsymbol{\theta}^*, \boldsymbol{\eta}^*) \cup \mathcal{D}_t] \right) p(\boldsymbol{\theta})^2 d\boldsymbol{\theta},$$

where $\boldsymbol{\eta}^* := \boldsymbol{\eta}(\boldsymbol{\theta}^*)$ represents the new simulation output at $\boldsymbol{\theta}^*$.

At each stage t, the next parameter is chosen to minimize (approximately) the acquisition function $\mathcal{A}_t$ such that

$$\boldsymbol{\theta}^{\text{new}} \in \underset{\boldsymbol{\theta}^* \in \mathcal{L}_t}{\arg\min}\, \mathcal{A}_t(\boldsymbol{\theta}^*).$$

Expected Integrated Variance for Calibration

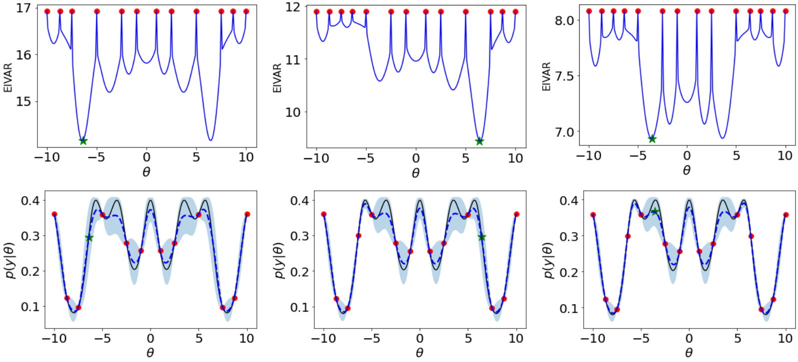
Lemma

Under the conditions of Lemma 1,

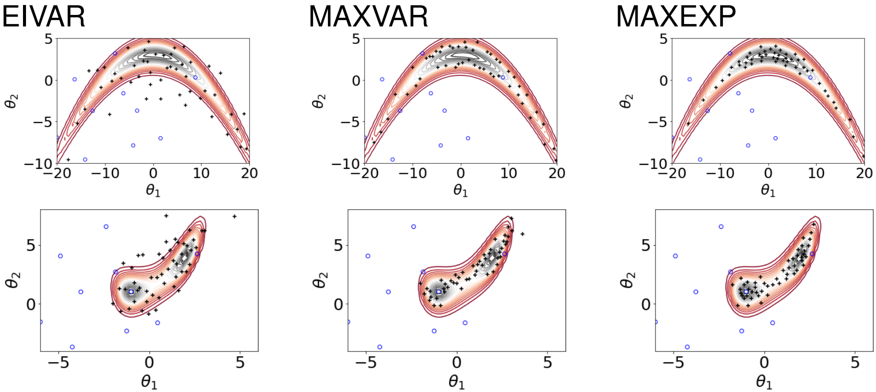
$$\int_{\Theta} \mathbb{E}_{\boldsymbol{\eta}^* | \mathcal{D}_t} \left(\mathbb{V}[p(\mathbf{y}|\boldsymbol{\theta}) \,|(\boldsymbol{\theta}^*, \boldsymbol{\eta}^*) \cup \mathcal{D}_t] \right) p(\boldsymbol{\theta})^2 d\boldsymbol{\theta}$$
$$= \int_{\Theta} \left(\frac{f_{\mathcal{N}}\left(\mathbf{y};\, \boldsymbol{\mu}_t(\boldsymbol{\theta}),\, \frac{1}{2}\boldsymbol{\Sigma} + \mathbf{S}_t(\boldsymbol{\theta})\right)}{2^d \pi^{d/2} |\boldsymbol{\Sigma}|^{1/2}} - \frac{f_{\mathcal{N}}\left(\mathbf{y};\, \boldsymbol{\mu}_t(\boldsymbol{\theta}),\, \frac{1}{2}\left(\boldsymbol{\Sigma} + \mathbf{S}_t(\boldsymbol{\theta}) + \boldsymbol{\phi}_t(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\right)\right)}{2^d \pi^{d/2} |\boldsymbol{\Sigma} + \mathbf{S}_t(\boldsymbol{\theta}) - \boldsymbol{\phi}_t(\boldsymbol{\theta}, \boldsymbol{\theta}^*)|^{1/2}} \right) p(\boldsymbol{\theta})^2 d\boldsymbol{\theta}.$$

Approximate EIVAR is given as

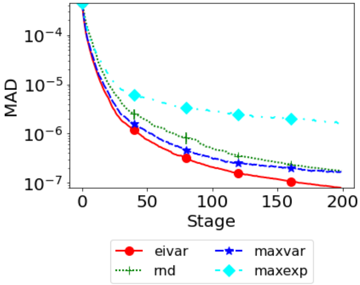
$$\frac{1}{|\Theta_{\mathrm{ref}}|} \sum_{\boldsymbol{\theta} \in \Theta_{\mathrm{ref}}} p(\boldsymbol{\theta})^2 \left(\frac{f_{\mathcal{N}}\left(\mathbf{y};\, \boldsymbol{\mu}_t(\boldsymbol{\theta}),\, \frac{\boldsymbol{\Sigma}}{2} + \mathbf{S}_t(\boldsymbol{\theta})\right)}{2^d \pi^{d/2} |\boldsymbol{\Sigma}|^{1/2}} - \frac{f_{\mathcal{N}}\left(\mathbf{y};\, \boldsymbol{\mu}_t(\boldsymbol{\theta}),\, \frac{1}{2}\left(\boldsymbol{\Sigma} + \mathbf{S}_t(\boldsymbol{\theta}) + \boldsymbol{\phi}_t(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\right)\right)}{2^d \pi^{d/2} |\boldsymbol{\Sigma} + \mathbf{S}_t(\boldsymbol{\theta}) - \boldsymbol{\phi}_t(\boldsymbol{\theta}, \boldsymbol{\theta}^*)|^{1/2}} \right).$$



EIVAR
17
16
15
−10
−5
0
5
10
θ
12
11
10
8.0
7.5
7.0
p(y|θ)
0.4
0.3
0.2
0.1


EIVAR
MAXVAR
MAXEXP
θ_2
θ_1
5
0
−5
−10
−20
−10
0
10
20

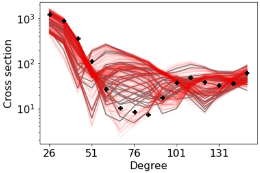
Comparison of Different Acquisition Functions



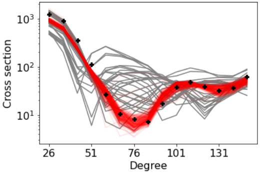
Application to a Reaction Model

- Reaction code `FRESCOX` takes the optical model parameters as input and generates the corresponding cross sections across angles from 0° to 180°
- Case study uses the elastic scattering data for the 48Ca(n,n)48Ca reaction

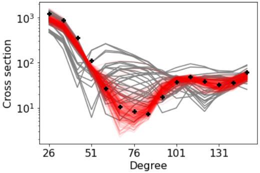
RND



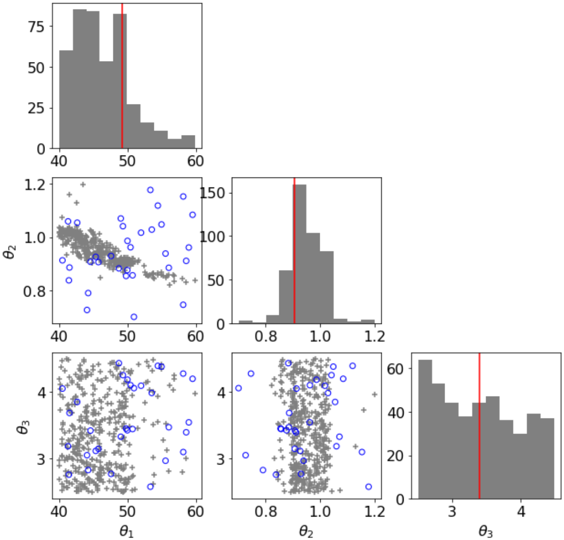
MAXVAR



EIVAR



Application to a Reaction Model



Questions?

The authors gratefully acknowledge support from the NSF grant OAC 2004601