Calibration of computer models

Sequential Designs of Experiments

Pierre BARBILLON

Fall 2023, école ETICS

Outline

Considered framework

Model $\mathcal{M}_2$:

$$\mathcal{M}_2 : \ \forall i \in [\![1, \ldots, n_e]\!], \quad y_i^e = F(\boldsymbol{x}_i, \boldsymbol{\theta}) + \epsilon_i,$$

Goal: find DoNE in order to make $\pi(\boldsymbol{\theta}|\mathbf{y}^e, \mathbf{y}^c, \mathbf{X}^e, D^c) = \pi^C(\boldsymbol{\theta}|\mathbf{y}^e, f(D_M^c))$ as close as possible to $\pi(\boldsymbol{\theta}|\mathbf{y}^e)$ under a limited N.

Extension to M4

Possible if a priori on the discrepancy function.

Posterior consistency

Proposition

Under the following assumptions:

- $\pi(\boldsymbol{\theta})$ *has a bounded support* Θ*,*
- *the code output* $f(\boldsymbol{x}, \boldsymbol{\theta})$ *is uniformly bounded on* $\mathcal{X} \times \Theta$*,*
- *the correlation function (kernel) of the GP surrogate is a classical radial basis function*
- f *lies in the associated Reproducing Kernel Hilbert Space,*
- *the covering distances* $h_{D_M^c}$ *associated with the sequence of designs* $(D_M^c)_M$ *tends to* 0 *with* $M \to \infty$*,*

then, we have:

$$\lim_{M\to\infty} KL\Big(\pi(\boldsymbol{\theta}|\mathbf{y}^e)||\pi^C(\boldsymbol{\theta}|\mathbf{y}^e, f(D_M^c))\Big) = 0\,.$$

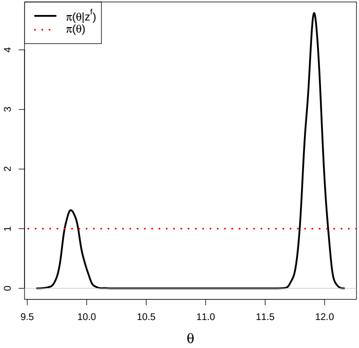
where

$$h_{D_M^c} = \max_{(\mathbf{x}',\boldsymbol{\theta}')\in\mathcal{X}\times\Theta}\ \min_{(\mathbf{x}_i,\boldsymbol{\theta}_i)\in D_M^c} \|(\mathbf{x}',\boldsymbol{\theta}') - (\mathbf{x}_i,\boldsymbol{\theta}_i)\| \underset{M\to\infty}{\longrightarrow} 0\,.$$

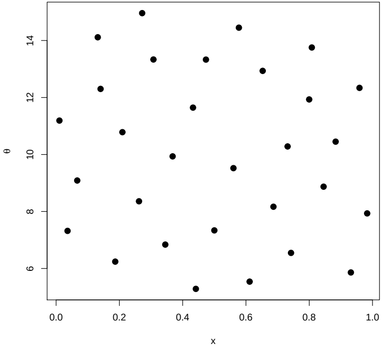
Motivation for adaptive designs in calibration

Quality of calibration (Bayesian or ML) is affected by choice in the numerical design.

- Calibration with unlimited runs of f

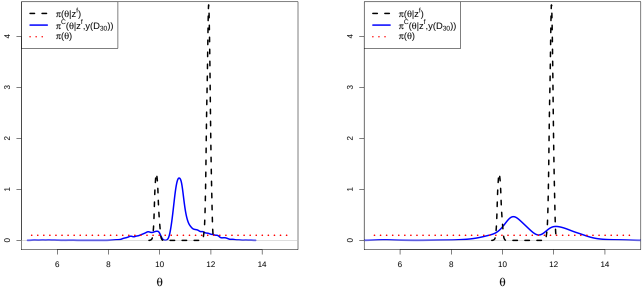


LHS maximin design



Motivation for adaptive designs in calibration

- Calibration with emulator built from a design with $M = 30$ calls to f



Outline

1. Apprixmate calibration
2. EGO enhanced design of numerical experiments for calibration

EI for calibration

Expected improvement criterion originally proposed by [Jones et al., 1998] for optimizing a black-box function

Optimization goal : maximize the likelihood $\Rightarrow$ Expected Improvement for calibration.

Maximize the likelihood $\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}^e)$ over $\boldsymbol{\theta} \Leftrightarrow$ Minimize $SS(\boldsymbol{\theta}) = \|\mathbf{y}^e - f(\mathbf{X}^e, \boldsymbol{\theta})\|^2$ over $\boldsymbol{\theta}$.

For given:

- field experiments $\mathbf{y}^e = y^e(\mathbf{x}_1^e), \dots, y^e(\mathbf{x}_n^e)$,
- D_k^c numerical design on $\mathcal{X} \times \Theta$ with M points,
- m_k current minimal value of $SS(\boldsymbol{\theta})$.

EI criterion:

$$EI_{D_k^c}(\boldsymbol{\theta}) = \mathbb{E}_{D_k^c}\left((m_k - SS(\boldsymbol{\theta}))^+\right) \ ,$$

to be maximized.

EI criterion is applied to a function of f.

EI computation

$$
\begin{aligned}
EI_{D_k^c}(\boldsymbol{\theta}) &= \int_{B(0,\sqrt{m_k})} (m_k - SS(\boldsymbol{\theta}))\, dF_{D_M} \\
&= m_k \cdot \mathbb{P}_{D_M}(SS(\boldsymbol{\theta}) \leq m_k) - \mathbb{E}_{D_M}\left(SS(\boldsymbol{\theta})\mathbb{I}_{SS(\boldsymbol{\theta})\leq m_k}\right)
\end{aligned}
$$

- no close form computation,
- $\mathbb{P}_{D_M}(SS(\boldsymbol{\theta}) \leq m_k)$ is an upper bound and easier to compute,
- importance sampling may be used for the second term.

Algorithm

Initialization

- Build an initial numerical design $D_0^c \subset \mathcal{X} \times \Theta$ of size M_0.
- Run the code over D_0^c, then construct an initial GPE based on $f(D_0^c)$.
- Compute $\hat{\boldsymbol{\theta}}_1$ as the posterior mean $\mathbb{E}[\boldsymbol{\theta}|\mathbf{y}^e, f(D_0^c)]$.
- $D_1^c = D_0^c \cup \{(\mathbf{x}_i^e, \hat{\boldsymbol{\theta}}_1)\}_{1\leq i\leq n_e}$.
- Update the GPE distribution after running the code over $\{(\mathbf{x}_i^e, \hat{\boldsymbol{\theta}}_1)\}_{1\leq i\leq n_e}$.
- Compute $m_1 := SS(\hat{\boldsymbol{\theta}}_1)$.

From $k = 1$, repeat the following steps as long as $M_0 + n \times (k+1) \leq M$.

Step 1 Find an estimate $\hat{\boldsymbol{\theta}}_{k+1}$ of $\boldsymbol{\theta}^\star_{k+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, EI_{D_k^c}(\boldsymbol{\theta})$.

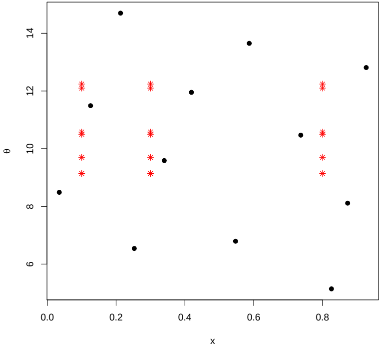
Step 2 $D_{k+1}^c = D_k^c \cup \{(\mathbf{x}_i^e, \hat{\boldsymbol{\theta}}_{k+1})\}_{1\leq i\leq n_e}$.

Step 3 Run the code over all new locations $\{(\mathbf{x}_i^e, \hat{\boldsymbol{\theta}}_{k+1})\}_{1\leq i\leq n_e}$.

Step 4 Update the GPE distribution based on $f(D_{k+1}^c)$.

Step 5 Compute $m_{k+1} := \min\{m_1, \cdots, m_k, SS(\hat{\boldsymbol{\theta}}_{k+1})\}$.

Adaptive design



Algorithm one at a time

Algorithm (step k $\longrightarrow$ step k + 1) :

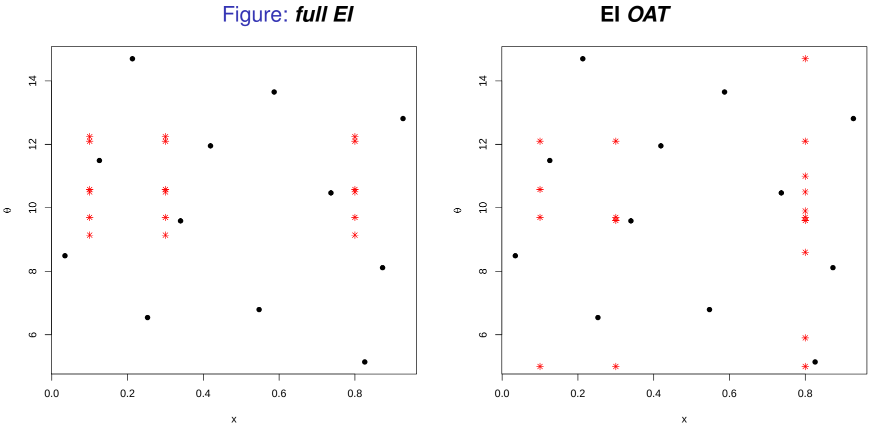
1. $\boldsymbol{\theta}_{k+1} = \underset{\boldsymbol{\theta}}{\text{argmax}}\, EI_k(\boldsymbol{\theta})$,
2. $D^c_{k+1} = D^c_k \cup (\mathbf{x}^\star, \boldsymbol{\theta}_{k+1})$ where $\mathbf{x}^\star \in \mathbf{X}^F = \left[\mathbf{x}^e_1, \cdots, \mathbf{x}^e_n\right]^T$,
3. $f(D^c_{k+1}) = f(D^c_k) \cup \{f(\mathbf{x}^\star, \boldsymbol{\theta}_{k+1})\}$,
4. $F^{D_{k+1}} = F|f(D^c_{k+1})$,
5. $m_{k+1} := \min\left\{\mathbb{E}[SS_{k+1}(\boldsymbol{\theta}_1)], \cdots, \mathbb{E}[SS_{k+1}(\boldsymbol{\theta}_k)], \mathbb{E}[SS_{k+1}(\boldsymbol{\theta}_{k+1})]\right\}$.

$$\boxed{\textbf{Only 1 simulation to compute } \mathbf{m_{k+1}}\textbf{!}}$$

where a criterion for step 2 is:

$$\mathbf{x}^\star = \underset{\mathbf{x} \in \{\mathbf{x}^e_1, \ldots, \mathbf{x}^e_{n_e}\}}{\text{argmax}} \left(\frac{\mathrm{Var}_F\left(F^{D^c_k}(\mathbf{x}^e_i, \boldsymbol{\theta}_{k+1})\right)}{\max\limits_{i=1,\cdots,n} \mathrm{Var}_F\left(F^{D^c_k}(\mathbf{x}^e_i, \boldsymbol{\theta}_{k+1})\right)} \times \frac{\mathrm{Var}_{\boldsymbol{\theta}}(m^k(\mathbf{x}^e_i, \boldsymbol{\theta}))}{\max\limits_{i=1,\cdots,n} \mathrm{Var}_{\boldsymbol{\theta}}(m^k(\mathbf{x}^e_i, \boldsymbol{\theta}))} \right)$$

Comparison full EI / EI one at a time



Recall that:

$$\pi(\boldsymbol{\theta}|\mathbf{y}^e) \propto \pi(\boldsymbol{\theta}) \cdot \exp(-SS(\boldsymbol{\theta})/2\sigma^2)$$

is high where $\boldsymbol{\theta} \mapsto SS(\boldsymbol{\theta})$ is small.

$$\mathrm{KL}\big(\pi(\boldsymbol{\theta}|\mathbf{y}^e)||\pi^C(\boldsymbol{\theta}|\mathbf{y}^e, f(D_M^c))\big) = \underbrace{K - K_M}_{(A)} + \int_\Theta \pi(\boldsymbol{\theta}|\mathbf{y}^e)\underbrace{(C - C_M(\boldsymbol{\theta}))}_{(B)}\mathrm{d}\boldsymbol{\theta}$$

$$+ \frac{1}{2}\int_\Theta \pi(\boldsymbol{\theta}|\mathbf{y}^e)\underbrace{\Big((\mathbf{y}^e - m(\mathbf{X}^e, \boldsymbol{\theta}))^T\tilde{\Sigma}_{\mathbf{y}^e}^{-1}(\mathbf{y}^e - m(\mathbf{X}^e, \boldsymbol{\theta}))) - SS(\boldsymbol{\theta})/\sigma^2\Big)}_{(C)}\mathrm{d}\boldsymbol{\theta}$$

where K and K_M correspond to the normalizing constants:

$$K = -\log\left(\int_\Theta \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}^e)\pi(\boldsymbol{\theta})\right), \quad K_M = -\log\left(\int_\Theta \mathcal{L}^C(\boldsymbol{\theta}; \mathbf{y}^e|f(D_M^c))\pi(\boldsymbol{\theta})\right),$$

$$C = -\frac{n}{2}\log\sigma_{err}^2, \quad C_M(\boldsymbol{\theta}) = -\frac{1}{2}\log|\tilde{\Sigma}_{\mathbf{y}^e}^{-1}| = -\frac{1}{2}\log\left(|\Sigma_{exp,exp}(\mathbf{X}^e, \boldsymbol{\theta}) + \sigma_{err}^2\mathbf{I}_{n_e}\right)^{-1}|.$$

and

$$SS(\boldsymbol{\theta}) = \|\mathbf{y}^e - f(\mathbf{x}, \boldsymbol{\theta})\|^2.$$

Sobol function

$\mathbf{x} \in \mathcal{X} = [0,1]^3$, $\boldsymbol{\theta} \in \Theta = [0,1]^3$

$$f_{\boldsymbol{\theta}} : \mathbf{x} \in \mathcal{X} \longrightarrow f_{\boldsymbol{\theta}}(\mathbf{x}) = \prod_{i=1}^{3} \frac{|4x_i - 2| + \theta_i}{1 + \theta_i} .$$

Field measurements $\mathbf{y}^f$ chosen according to a maximin LHD on $\mathcal{X}$ of size $n = 60$. For $1 \leq i \leq 60$,

$$y_i^f = f_{\boldsymbol{\theta}}(x_i^f) + \epsilon_i ,$$

where $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.05^2)$ and $\boldsymbol{\theta} = (0.55, 0.55, 0.1)$.

GPE is fitted with a constant mean $m_{\boldsymbol{\beta}} = m$ and a Matérn $5/2$ correlation function.

Prior distribution $\pi(\boldsymbol{\theta})$ on Θ:

$$\pi(\boldsymbol{\theta}) \propto \mathbf{1}_{[0,1]^3}(\boldsymbol{\theta}).$$

Designs

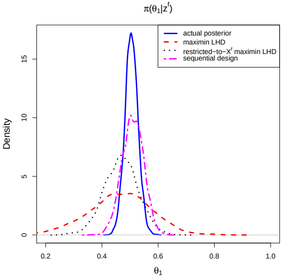
Number of simulations $M = 150$.

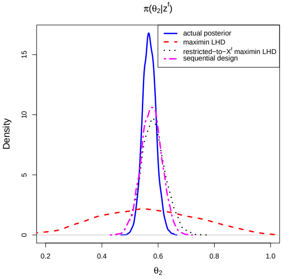
Comparison of 4 designs.

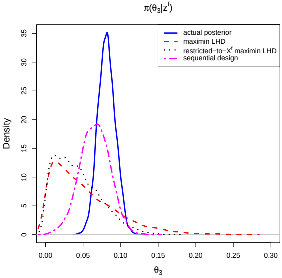
1. Maximin LHD in 6D: $\mathcal{X} \times \Theta = [0, 1]^6$.
2. *Restricted*-to-$\mathbf{X}^f$ maximin LHD.
3. Sequential designs OAT with GPE variance criterion for choosing $\mathbf{x}^\star_{k+1}$.
4. Sequential designs OAT with trade-off (GPE-variance, variability of f w.r.t. $\mathbf{x}$) (variance criterion for choosing $\mathbf{x}^\star_{k+1}$.

Sequential designs based on an initial design with $M_0 = 75$ points chosen as a *Restricted*-to-$\mathbf{X}^f$ maximin LHD.

Marginal posterior distributions







see also

[Sürer et al., 2023]

Jones, D. R., Schonlau, M., and Welch, W. J. (1998).
Efficient global optimization of expensive black-box functions.
Journal of Global optimization, 13(4):455–492.

Sürer, Ö., Plumlee, M., and Wild, S. M. (2023).
Sequential bayesian experimental design for calibration of expensive simulation models.
arXiv preprint arXiv:2305.16506.