Adaptive numerical designs for the calibration of computer codes

Guillaume Damblin

CEA Saclay,
DEN/DANS/DM2S/STMF/LGLS

(from my PhD work at EDF R&D/AgroParisTech with Pierre Barbillon, Merlin Keller, Alberto Pasanisi and Eric Parent)

Summer School ETICS - June 2016

Calibration with fixed physical system

Basic assumptions :

- for any $\boldsymbol{x} \in \mathcal{X}$, there exists a unique "true" physical quantity of interest $\phi(\boldsymbol{x}) \in \mathbb{R}$,
- no discrepancy, i.e. $\phi(\boldsymbol{x}) = f(\boldsymbol{x}, \boldsymbol{\beta}_0)$,
- $f(\boldsymbol{x}, \boldsymbol{\beta}_0)$ might be strongly non-linear.

Statistical model :

- based on n physical measurements $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$,
- $\boldsymbol{x}_i \in \mathbb{R}^d \rightarrow \boxed{\text{true physical quantity } \phi(\boldsymbol{x}_i)} \rightarrow y_i \in \mathbb{R}$, then :

$$\begin{aligned} y_i &= \phi(\boldsymbol{x}_i) + \epsilon_i \\ &= f(\boldsymbol{x}_i, \boldsymbol{\beta}_0) + \epsilon_i \end{aligned} \tag{1}$$

where $\epsilon_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_m^2)$ with σ_m^2 is assumed known.

Calibration goal : estimation of $\boldsymbol{\beta}_0$ from Equation (1),

GP-based Bayesian calibration

$\boldsymbol{z} = (z_1, \cdots, z_n)^T$, $\mathbf{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n]^T$

Target posterior distribution :

$$\pi(\boldsymbol{\beta}_0|\boldsymbol{z}) \propto \frac{1}{(\sqrt{2\pi}\sigma_m)^n} \exp\left[-\frac{1}{2\sigma_m^2} SS(\boldsymbol{\beta}_0)\right] \pi(\boldsymbol{\beta}_0).$$

where

$$SS(\boldsymbol{\beta}_0) = ||\boldsymbol{z} - f(\mathbf{X}, \boldsymbol{\beta}_0)||^2 \tag{2}$$

- Computer experiments $f(\boldsymbol{x}, \boldsymbol{\beta})$ are time-consuming $\Longrightarrow$ MCMC sampling is infeasible.

GP-based posterior distribution :

$$\pi^C(\boldsymbol{\beta}_0|\mathbf{z}, f(\mathbf{D}_M)) \propto |V^M_{\hat{\boldsymbol{l}},\hat{\sigma}^2}(\boldsymbol{\beta}_0) + \sigma_m^2 \mathbf{I}_n|^{-1/2} \exp\Big\{ -\frac{1}{2}\Big[(\boldsymbol{z} - \mu^M_{\hat{\mathbf{a}},\hat{\boldsymbol{l}}}(\mathbf{D}_{\boldsymbol{\beta}_0}))^{\mathrm{T}}) \\ (V^M_{\hat{\boldsymbol{l}},\hat{\sigma}^2}(\boldsymbol{\beta}_0) + \sigma_m^2 \mathbf{I}_n)^{-1}(\boldsymbol{z} - \mu^M_{\hat{\mathbf{a}},\hat{\boldsymbol{l}}}(\mathbf{D}_{\boldsymbol{\beta}_0}))\Big]\Big\} \pi(\boldsymbol{\beta}_0) \tag{3}$$

where $\mathbf{D}_{\boldsymbol{\beta}_0} = [(\boldsymbol{x}_1, \boldsymbol{\beta}_0), \cdots, (\boldsymbol{x}_M, \boldsymbol{\beta}_0)]^T$.

The Kullback-Leibler divergence

The KL divergence measures how far a probability distribution is from a reference [2] :

$$\mathrm{KL}\left(\pi(\boldsymbol{\beta}_0|\boldsymbol{z})||\pi^C(\boldsymbol{\beta}_0|\boldsymbol{z}, f(\mathbf{D}_M))\right) = \int \pi(\boldsymbol{\beta}_0|\boldsymbol{z})\Big(\log\left(\pi(\boldsymbol{\beta}_0|\boldsymbol{z})\right) - \log\left(\pi^C(\boldsymbol{\beta}_0|\boldsymbol{z}, f(\mathbf{D}_M))\right)\Big)\mathrm{d}\boldsymbol{\beta}_0.$$

Main goal of the work : minimize this KL-divergence !

- if M is large enough, we proved that,

$$\lim_{M\to\infty} \mathrm{KL}\left(\pi(\boldsymbol{\beta}_0|\boldsymbol{z})||\pi^C(\boldsymbol{\beta}_0|\boldsymbol{z}, f(\mathbf{D}_M))\right) \tag{4}$$

- if M is small, we need to construct a numerical design $\mathbf{D}_M$ with care ! How to build $(\mathbf{D}_M, f(\mathbf{D}_M))$?

Main idea : apply the ***Expected Improvement (EI)*** criterion to $SS(\boldsymbol{\beta}_0)$ [1].

The EI criterion designed for $SS(\beta)$

EI criterion :

$$EI_k(\boldsymbol{\beta}) = \mathbb{E}\left[(m_k - SS_k(\boldsymbol{\beta}))\mathbf{1}_{SS_k(\boldsymbol{\beta})\leq m_k}\right] \tag{5}$$

where

- $m_k := \min\{SS(\boldsymbol{\beta}_1), \cdots, SS(\boldsymbol{\beta}_{k-1}), SS(\boldsymbol{\beta}_k)\}$,
- $SS_k(.)$ denotes the sum of squares of the residuals where f is replaced with the GPE conditional on $f(\mathbf{D}_k)$:

$$f_k(.) := f(.)|f(\mathbf{D}_k).$$

Maximization : $\boldsymbol{\beta}_{k+1} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, EI_k(\boldsymbol{\beta})$

Two versions of the EGO algorithm :

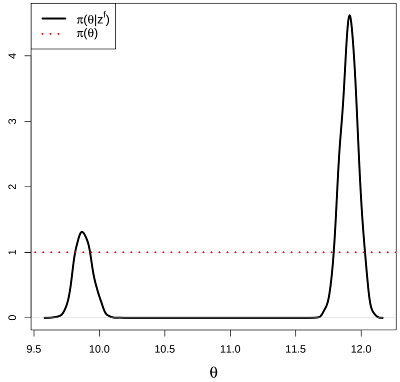
- Exact one (sequence of n runs) : $\mathbf{D}_{k+1} = \mathbf{D}_k \cup \{(\boldsymbol{x}_1, \boldsymbol{\beta}_{k+1}), \cdots, (\boldsymbol{x}_n, \boldsymbol{\beta}_{k+1})\}_{1\leq i\leq n}$,
- Approximated one (one at a time) : $\mathbf{D}_{k+1} = \mathbf{D}_k \cup (\boldsymbol{x}^\star, \boldsymbol{\beta}_{k+1})$ where $\boldsymbol{x}^\star \in \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$.

Interest : **new simulations $f(\boldsymbol{x}_i, \boldsymbol{\beta}_k)$ are run to reduce the uncertainty of the GPE where $SS(\boldsymbol{\beta}_0)$ is small ($\Leftrightarrow$ regions of high posterior probability when $\pi(\boldsymbol{\beta}_0)$ is non-informative).**

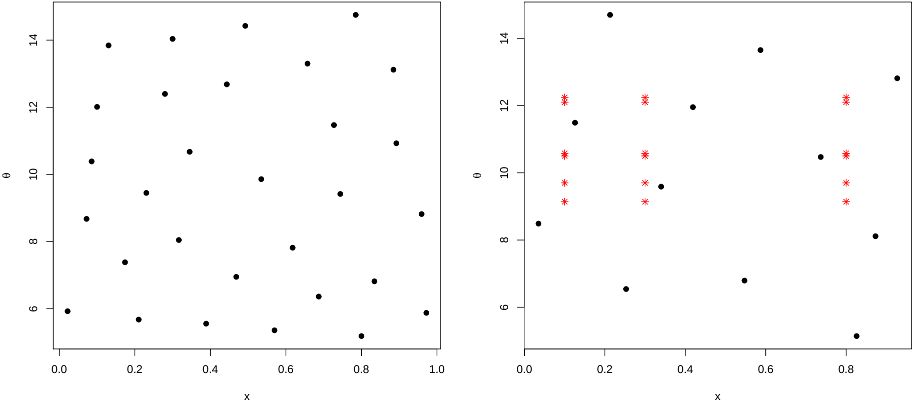
Illustration on a toy example

- $f(x, \beta) = (6x - 2)^2 \times \sin(\beta x - 4)$,
- $\mathbf{X} = [x_1 = 0.1, x_2 = 0.3, x_3 = 0.8]$,
- $\epsilon_i \sim \mathcal{N}(0, 0.3^2)$,
- $y_i = f(x_i, \beta_0) + \epsilon_i$ where $\boldsymbol{\beta_0} = \mathbf{12}$,
- flat prior on β_0 : $\pi(\beta_0) = \mathbf{1}_{[5,15]}(\beta_0)$.

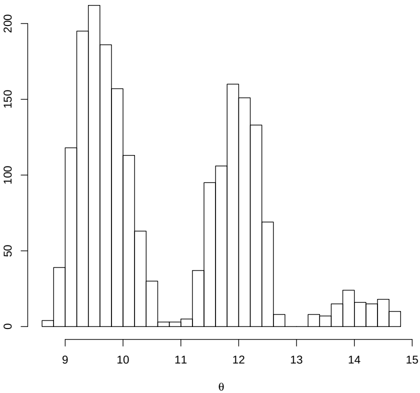
FIGURE : *Target posterior distribution* $\pi(\beta_0|\mathbf{y})$



One shot design vs EGO-based design : size($\mathbf{D}_M$) $= 30$



Corresponding GP-based posterior distributions



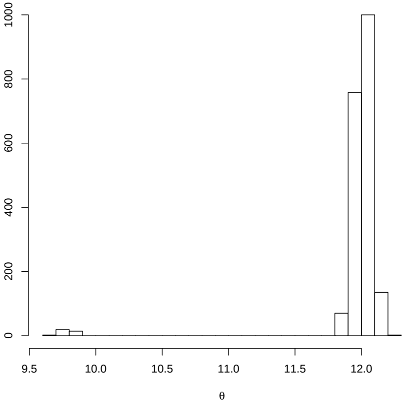
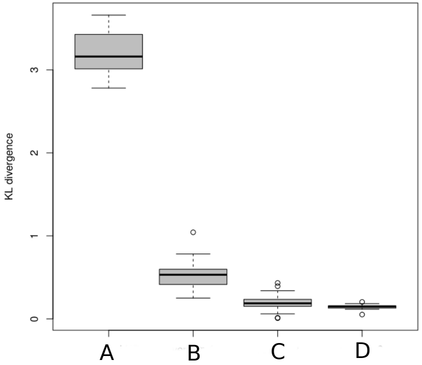


FIGURE : *Boxplots of the KL divergence computed between* $\pi^C(\beta_0|\boldsymbol{z}, y(\mathbf{D}_M))$ *and* $\pi(\beta_0|\boldsymbol{z})$

Impact of $\mathbf{D}_M$ in terms of the KL-divergence ?

- A : maximin LHD,
- B : exact EGO,
- C and D : one at a time EGO.



Conclusions

- Improve the way to maximize the EI criterion,
- Take into account the prior distribution in the writing of the EI criterion (when it is informative),
- Apply the algorithm to an industrial case,
- Extend the method when an unknown discrepancy function is inserted between the code output and the physical system,
- Build $\mathbf{D}_M$ from the KL-divergence instead of $SS(.)$?

DAMBLIN, G. ; BARBILLON, P. ; KELLER, M. ; PASANISI, A. ; PARENT, E., *Adaptive numerical designs for the calibration of computer codes*, http ://arxiv.org/abs/1502.07252 (2015).

JONES, D.R. ; SCHONLAU, M. ; WELCH, W.J., *Efficient Global Optimization of Expensive Black-Box Functions*, Journal of Global Optimization 13(4) (1998).

KULLBACK, S., *Information Theory and Statistics*, Wiley (1959).