Adaptive numerical designs for the calibration of computer codes

Guillaume Damblin[1,2]

[1]EDF R&D, *6 quai Watier 78401 Chatou*
[2]AgroParisTech/INRA UMR MIA 518, *16 rue Claude Bernard 75005 Paris*

MASCOT NUM 2015, April 8 2015

Outline

Calibration of costly computer codes

Adaptive designs based on the EI criterion

Outline

Notations

Let $r(\mathbf{x}) \in \mathbb{R}$ be a physical quantity of interest:

- $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$ is a vector of control variables,
- $z(\mathbf{x}) = r(\mathbf{x}) + \epsilon(\mathbf{x})$ is the physical measurement.

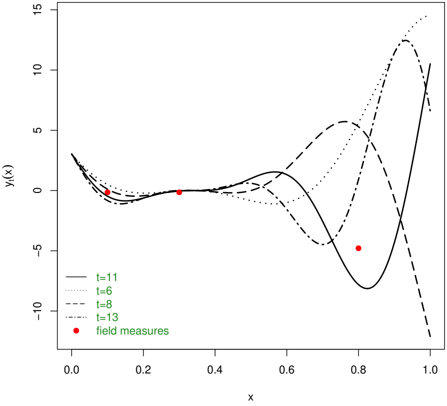
Let $y_{\mathbf{t}}(\mathbf{x}) \in \mathbb{R}$ be a computer code:

- $\mathbf{x}$ is aligned on the r input,
- $\mathbf{t} \in \mathcal{T} \in \mathbb{R}^p$ is a vector of code parameters (may have no counterpart in r).

What is the value of $\mathbf{t}$ making the best agreement between $r(\mathbf{x})$ and $y_{\mathbf{t}}(\mathbf{x})$?

Illustration

The function $y_t(x) = (6x-2)^2 \times \sin(tx-4)$ *on* $[0,1]$ *for several values of* $t \in [5,15]$. *Red dots are the physical measurements* $z(\mathbf{x_i})$.



The statistical modelling

- n physical experiments:
  - $\mathbf{x} = \{\mathbf{x_1}, \cdots, \mathbf{x_n}\}$,
  - $\mathbf{z} = \{z(\mathbf{x_1}), \cdots, z(\mathbf{x_n})\}$.
- $\exists \boldsymbol{\theta} \in \mathcal{T} \quad r(\mathbf{x_i}) = y_{\boldsymbol{\theta}}(\mathbf{x_i})$ (negligible model error),
- Recall $z(\mathbf{x}_i) = r(\mathbf{x_i}) + \epsilon(\mathbf{x_i})$,
- Hence, $z(\mathbf{x_i}) = y_{\boldsymbol{\theta}}(\mathbf{x_i}) + \epsilon$ where $\epsilon \overset{i.i.d}{\sim} \mathcal{N}(0, \lambda^2)$.

Statistical calibration consists in estimating θ in this regression model!

Bayesian inference of $\boldsymbol{\theta}$

Bayesian inference : $\Pi(\boldsymbol{\theta}|\mathbf{z}) \propto \mathcal{L}(\mathbf{z}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta})$

- $\Pi(\boldsymbol{\theta})$ is the *prior* distribution,
- $\mathcal{L}(\mathbf{z}|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\lambda} \exp\Big(- \frac{1}{2\lambda^2} SS(\boldsymbol{\theta}) \Big)$,

where $SS(\boldsymbol{\theta}) = ||\mathbf{z} - y_{\boldsymbol{\theta}}(\mathbf{x})||^2$.

Bayesian inference of $\boldsymbol{\theta}$

The code $y_{\boldsymbol{\theta}}(\mathbf{x})$ is non-linear:

$\Longrightarrow$ no closed form for $\Pi(\boldsymbol{\theta}|\mathbf{z})$,

$\Longrightarrow$ need for MCMC methods,

$\Longrightarrow$ need for hundreds of simulations $y_{\boldsymbol{\theta}_\mathbf{i}}(\mathbf{x}_\mathbf{i})$.

Issue : the code is costly $\Longrightarrow M << \infty$ simulations are allocated!

A possible solution : replacing the code by a Gaussian process emulator!

The Gaussian process emulator (GPE)

Prior hypothesis:

$$y_{\mathbf{t^j}}(\mathbf{x^j}) = y(\mathbf{x^j}, \mathbf{t^j}) \sim Y = \mathcal{PG}(m_{\boldsymbol{\beta}}(.), \Sigma_{\boldsymbol{\Psi}}(.)).$$

Design of numerical experiments:

$$\mathbf{D_M} := \{(\mathbf{x^1}, \mathbf{t^1}), \cdots, (\mathbf{x^M}, \mathbf{t^M})\} \subset \mathcal{X} \times \mathcal{T}$$

$$\Longrightarrow$$

$$\mathbf{y}(\mathbf{D_M}) := \{y(\mathbf{x^1}, \mathbf{t^1}), \cdots, y(\mathbf{x^M}, \mathbf{t^M})\}$$

GPE emulator:

$$Y^M := Y|\mathbf{y}(\mathbf{D_M}) \sim \mathcal{PG}(\mu^{\mathbf{M}}_{\boldsymbol{\beta}}(.), V^{\mathbf{M}}_{\boldsymbol{\Psi}}),$$

which gives a stochastic prediction of $y_{\mathbf{t}}(\mathbf{x})$ over $\mathcal{X} \times \mathcal{T}$.

The approximated likelihood based on a GPE

It is given by the conditional likelihood $\mathcal{L}^{\mathcal{C}}(\mathbf{z}|\mathbf{y}(\mathbf{D_M}), \boldsymbol{\theta}, \hat{\beta}, \hat{\boldsymbol{\Psi}})$

$$\mathcal{L}^{C}(\mathbf{z}|\boldsymbol{\theta}, y(\mathbf{D_M}), \hat{\beta}, \hat{\boldsymbol{\Psi}}) \propto |V_{\hat{\boldsymbol{\Psi}}} + \lambda^2 \mathbf{I}_n|^{-1/2} \exp -\frac{1}{2}\Big[\mathbf{z} - \boldsymbol{\mu}^{\mathbf{M}}_{\hat{\boldsymbol{\beta}}}(\mathbf{x}, \boldsymbol{\theta})^{\mathrm{T}})$$

$$(V_{\hat{\boldsymbol{\Psi}}} + \lambda^2 \mathbf{I}_n)^{-1}(\mathbf{z} - \boldsymbol{\mu}^{\mathbf{M}}_{\hat{\boldsymbol{\beta}}}(\mathbf{x}, \boldsymbol{\theta}))\Big].$$

where $(\hat{\beta}, \hat{\boldsymbol{\Psi}}) = \underset{(\boldsymbol{\beta}, \boldsymbol{\Psi})}{\operatorname{argmax}} \, \mathcal{L}^{M}(\mathbf{y}(\mathbf{D_N})|\boldsymbol{\beta}, \boldsymbol{\Psi})$

Approximated Bayesian calibration of $\boldsymbol{\theta}$

- $\Pi^C(\boldsymbol{\theta}|\mathbf{z}, \mathbf{D_M}) \propto \mathcal{L}^C(\mathbf{z}|\boldsymbol{\theta}, \mathbf{D_M})\Pi(\boldsymbol{\theta})$,

Approximated Bayesian calibration of $\boldsymbol{\theta}$

- $\Pi^C(\boldsymbol{\theta}|\mathbf{z}, \mathbf{D_M}) \propto \mathcal{L}^C(\mathbf{z}|\boldsymbol{\theta}, \mathbf{D_M})\Pi(\boldsymbol{\theta})$,
- $\Pi^C(\boldsymbol{\theta}|\mathbf{z}, \mathbf{D_M})$ is cheap to evaluate $\Longrightarrow$ MCMC methods OK!

Approximated Bayesian calibration of $\boldsymbol{\theta}$

- $\Pi^C(\boldsymbol{\theta}|\mathbf{z}, \mathbf{D_M}) \propto \mathcal{L}^C(\mathbf{z}|\boldsymbol{\theta}, \mathbf{D_M})\Pi(\boldsymbol{\theta})$,
- $\Pi^C(\boldsymbol{\theta}|\mathbf{z}, \mathbf{D_M})$ is cheap to evaluate $\Longrightarrow$ MCMC methods OK!
- The larger $\mathbf{D_M}$, the closer $\mathcal{L}^C(\boldsymbol{\theta}|\mathbf{z}, \mathbf{D_M})$ to $\mathcal{L}(\boldsymbol{\theta}|\mathbf{z})$

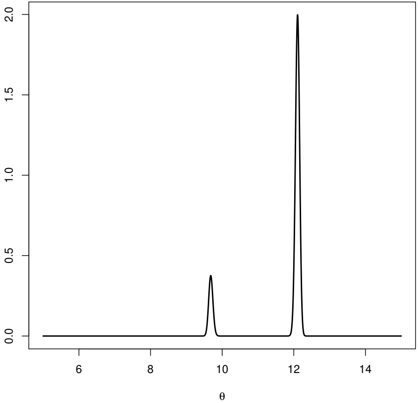
Approximated Bayesian calibration of $\boldsymbol{\theta}$

- $\Pi^C(\boldsymbol{\theta}|\mathbf{z},\mathbf{D_M}) \propto \mathcal{L}^C(\mathbf{z}|\boldsymbol{\theta},\mathbf{D_M})\Pi(\boldsymbol{\theta})$,
- $\Pi^C(\boldsymbol{\theta}|\mathbf{z},\mathbf{D_M})$ is cheap to evaluate $\Longrightarrow$ MCMC methods OK!
- The larger $\mathbf{D_M}$, the closer $\mathcal{L}^C(\boldsymbol{\theta}|\mathbf{z},\mathbf{D_M})$ to $\mathcal{L}(\boldsymbol{\theta}|\mathbf{z})$
- $\mathrm{KL}(\Pi^C(\boldsymbol{\theta}|\mathbf{z},\mathbf{D_M})||\Pi(\boldsymbol{\theta}|\mathbf{z})) \underset{M\to\infty}{\longrightarrow} 0$

When M is small, $\mathrm{KL}(\Pi^C(\boldsymbol{\theta}|\mathbf{z},\mathbf{D_M})||\Pi(\boldsymbol{\theta}|\mathbf{z}))$ may be high !

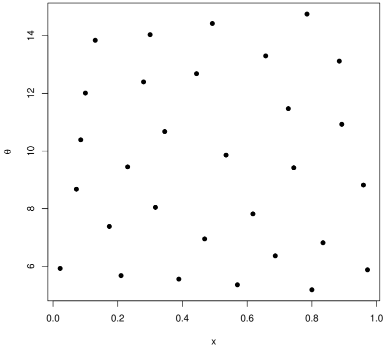
Artificial example

Left: The target posterior distribution $\Pi(\boldsymbol{\theta}|\mathbf{z})$



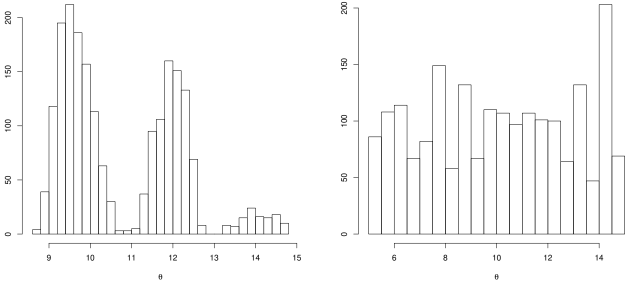
Toy example

Two maximin Latin Hypercube Design $\mathbf{D_M}$



Toy example

The corresponding $\Pi^C(\theta|\mathbf{z}, \mathbf{D_M})$ *according to* $\mathbf{D_M}$



Outline

Calibration of costly computer codes

Adaptive designs based on the EI criterion

Adaptive calibration

- **Question**: How choosing $\mathbf{D_M}$ to reduce $\mathrm{KL}(\Pi^C(\theta|\mathbf{z}, \mathbf{D_M})||\Pi(\theta|\mathbf{z}))$?

Adaptive calibration

- **Question**: How choosing $\mathbf{D_M}$ to reduce $\mathrm{KL}(\Pi^C(\theta|\mathbf{z},\mathbf{D_M})||\Pi(\theta|\mathbf{z}))$?
- **An idea**: reduce the difference $|\mathcal{L}^C(\theta|\mathbf{z},\mathbf{D_M}) - \mathcal{L}(\theta|\mathbf{z})|$ where $\mathcal{L}(\theta|\mathbf{z})$ is high,

Adaptive calibration

- **Question**: How choosing $\mathbf{D_M}$ to reduce $\mathrm{KL}(\Pi^C(\theta|\mathbf{z}, \mathbf{D_M})||\Pi(\theta|\mathbf{z}))$?
- **An idea**: reduce the difference $|\mathcal{L}^C(\theta|\mathbf{z}, \mathbf{D_M}) - \mathcal{L}(\theta|\mathbf{z})|$ where $\mathcal{L}(\theta|\mathbf{z})$ is high,
- **An equivalent idea** : reduce the uncertainty of the GPE at locations $\{(\mathbf{x_i}, \theta)\}$ where $SS(\theta)$ is low.

Adaptive calibration

- **Question**: How choosing $\mathbf{D_M}$ to reduce $\mathrm{KL}(\Pi^C(\theta|\mathbf{z}, \mathbf{D_M})||\Pi(\theta|\mathbf{z}))$?
- **An idea**: reduce the difference $|\mathcal{L}^C(\theta|\mathbf{z}, \mathbf{D_M}) - \mathcal{L}(\theta|\mathbf{z})|$ where $\mathcal{L}(\theta|\mathbf{z})$ is high,
- **An equivalent idea** : reduce the uncertainty of the GPE at locations $\{(\mathbf{x_i}, \theta)\}$ where $SS(\theta)$ is low.
- **A solution**: $\mathbf{D_M}$ is sequentially built thanks to the EI criterion applied to $SS(\theta)$.

$$\mathbf{D_1} \longrightarrow \cdots \longrightarrow \mathbf{D_k} \xrightarrow{\mathrm{EI}} \mathbf{D_{k+1}} \longrightarrow \cdots \longrightarrow \mathbf{D_M}$$

The step k

- $Y^k := Y|\mathbf{y}(\mathbf{D_k})$ constructed from $\mathbf{D_k}$,
- $m_k := \min\{SS(\boldsymbol{\theta_1}), \cdots, SS(\boldsymbol{\theta_{k-1}}), SS(\boldsymbol{\theta_k})\}$,
- $\mathbf{D_k} = \{(\mathbf{x_i}, \boldsymbol{\theta_j})\}_{1\leq i\leq n, 1\leq j\leq k}$ is a grid.

How to choose the next input locations $\{(\mathbf{x_i}, \boldsymbol{\theta_{k+1}})\}_{1\leq i\leq n}$ where the code is run ?

The EI criterion: from $\mathbf{D}_k$ to $\mathbf{D}_{k+1}$

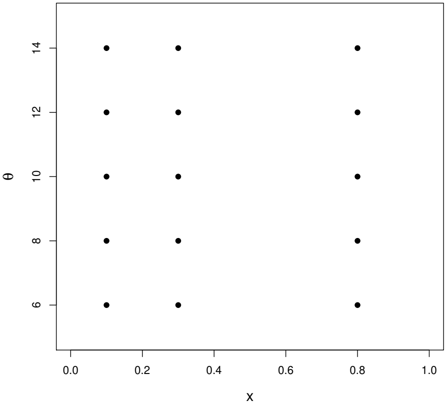
$$EI^k(\boldsymbol{\theta}) = \mathbb{E}\left[\left(m_k - SS_k(\boldsymbol{\theta})\right)\mathbf{1}_{SS_k(\boldsymbol{\theta}) \leq m_k} | Y^k\right] \quad \in \quad [0, m_k],$$

Then,

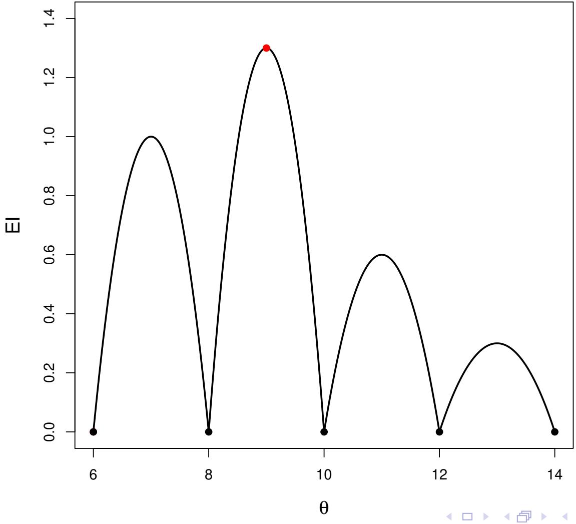
- $\boldsymbol{\theta}_{\mathbf{k+1}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, EI^k(\boldsymbol{\theta})$,
- $\mathbf{D}_{k+1} = \mathbf{D_k} \cup \{(\mathbf{x_i}, \boldsymbol{\theta}_{\mathbf{k+1}})\}_{1 \leq i \leq n}$.

To construct $\mathbf{D_M}$, repeat the EI criterion for $1 \leq k \leq M$!

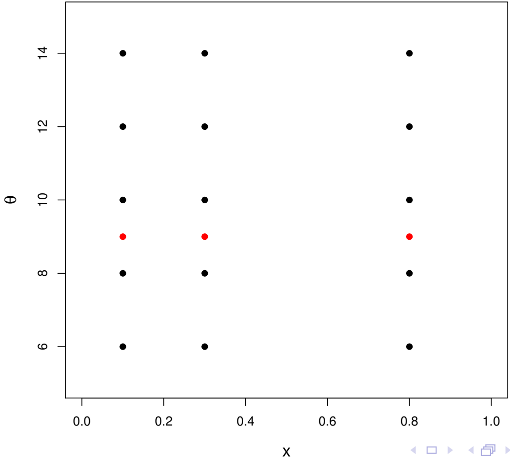
Design $\mathbf{D_k}$



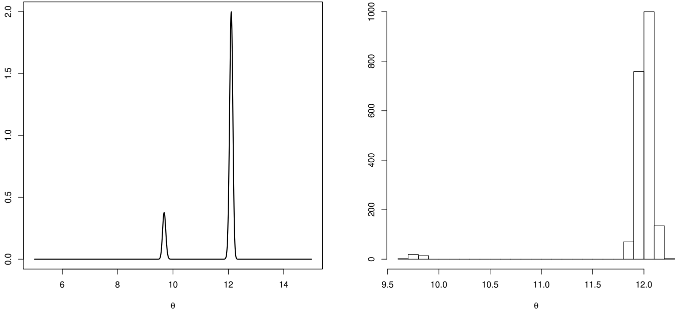
Optimization of the EI criterion



Design $\mathbf{D_{k+1}}$



Approximated calibration using $\mathbf{D_M}$



$\Longrightarrow$ low KL value !

Comments

- no closed-form for $EI^k(\boldsymbol{\theta})$,

Comments

- no closed-form for $EI^k(\boldsymbol{\theta})$,
- $\mathbf{D_k}$ is a grid design,

Comments

- no closed-form for $EI^k(\boldsymbol{\theta})$,
- $\mathbf{D_k}$ is a grid design,
- unsuitable when n is large,

Comments

- no closed-form for $EI^k(\boldsymbol{\theta})$,
- $\mathbf{D_k}$ is a grid design,
- unsuitable when n is large,
- need of one at a time strategies:
  - maximize the EI criterion $\Longrightarrow \boldsymbol{\theta}_{k+1}$,
  - pick up a single pair $(\mathbf{x}^\star, \boldsymbol{\theta}_{k+1})$ where $\mathbf{x}^\star \in \{\mathbf{x_1}, \cdots, \mathbf{x_n}\}$.

Two criteria for one at a time strategies

- First criterion to reduce the uncertainty of the GPE:

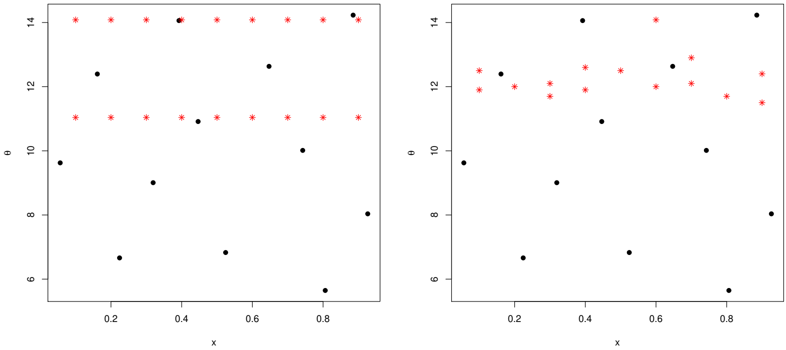
$$\mathbf{x}^\star = \max_{\mathbf{x_i}} \mathbb{V}(Y^k(\mathbf{x_i}, \boldsymbol{\theta}_{k+1}))$$

- Second criterion to compromise with the calibration goal:

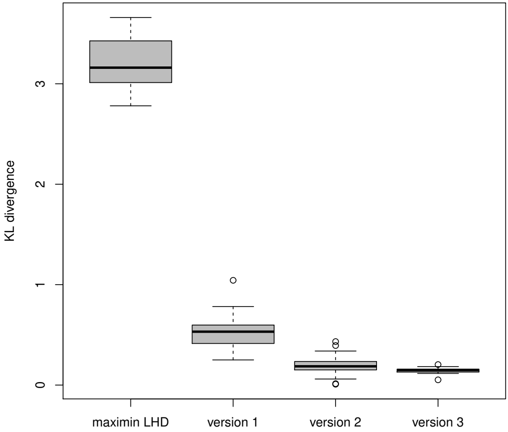
$$\mathbf{x}^\star = \max_{\mathbf{x_i}} \left(\frac{\mathbb{V}\big(Y^k(\mathbf{x_i}, \boldsymbol{\theta}_{k+1})\big)}{\max\limits_{i=1,\cdots,n} \mathbb{V}\big(Y^k(\mathbf{x_i}, \boldsymbol{\theta}_{k+1})\big)} \times \frac{\mathbb{V}(\mu_\beta^k(\mathbf{x_i}, \mathcal{T}))}{\max\limits_{i=1,\cdots,n} \mathbb{V}(\mu_\beta^k(\mathbf{x_i}, \mathcal{T}))} \right)$$

Design comparison

Black dots are the initial design. Red stars are the new experiments selected from the EI criterion.



Robustness in terms of the KL divergence



Main references

G. Damblin, P. Barbillon, M. Keller, A. Pasanisi, and E. Parent.
Adaptive numerical designs for the calibration of computer models.
Submitted and http://arxiv.org/abs/1502.07252.

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

M. Kennedy and A. O'Hagan.
Bayesian calibration of computer models.
Journal of the Royal Statistical Society, Series B, Methodological, 63:425–464, 2001(a).