Adaptive numerical designs for the calibration of computer codes

Pierre BARBILLON[1]
Joint work with: Guillaume DAMBLIN[1,2], Merlin KELLER[2], Éric PARENT[1], Alberto PASANISI[2]

[1] AgroParisTech / INRA UMR MIA Paris, Université Paris-Saclay
[2] EDF R.& D.

Journées MEXICO-MASCOT NUM 22-23 novembre 2016

Outline

Outline

Outline

Field data

- Field data provided by physical experiments:

$$\mathbf{y}^F = y^F(\mathbf{x}_1), \ldots, y^F(\mathbf{x}_n)\,,$$

- n is small, $\mathbf{x}_1, \ldots \mathbf{x}_n \in \mathcal{X}$ hard to set, sometimes uncontrollable, included in a small domain...

- Model:

$$y^F(\mathbf{x}_i) = \zeta(\mathbf{x}_i) + \epsilon(\mathbf{x}_i)\,,$$

where

  - $\zeta(\cdot)$ real physical process (unknown),
  - $\epsilon(\mathbf{x}_i)$ often assumed i.i.d. $\mathcal{N}(0, \sigma^2)$,
  - σ^2 sometimes treated as known...

Computer model / simulator

Computer experiments:

Computer model (simulator) $(\mathbf{x}^*, \boldsymbol{\theta}) \mapsto f(\mathbf{x}^*, \boldsymbol{\theta}) \in \mathbb{R}^s$ where

- **physical parameters**: $\mathbf{x}^* \in \mathcal{X} \subset \mathbb{R}^m$ observable and often controllable inputs
  - $\mathbf{x}^*$ same meaning as in field data,
  - extrapolation if $\mathbf{x}^* > \max(\mathbf{x}_i)$ or $\mathbf{x}^* < \min(\mathbf{x}_i)$.
- **simulator parameters**: $\boldsymbol{\theta} \in \mathcal{T} \subset \mathbb{R}^d$ non-observable parameters, required to run the simulator.

  2 types:
  - "calibration parameters": physical meaning but unknown, necessary to make the code mimic the reality,
  - "tuning parameters": no physical interpretation.

f designed to mimic the unknown physical process $\zeta(\cdot)$ for a value of $\boldsymbol{\theta}$.
The simulator is often an **expensive black-box function**.
$\Rightarrow$ limited number M of runs of the simulator.

Relationship between the simulator and the data

for $i = 1, \ldots, n$,

- if the simulator sufficiently represents the physical system:

$$y_i^F = f(\mathbf{x}_i, \boldsymbol{\theta}^*) + \epsilon(\mathbf{x}_i)\,,$$

  i.e. for the unknown value $\boldsymbol{\theta} = \boldsymbol{\theta}^*$: $f(\mathbf{x}, \theta^*) = \zeta(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{X}$,

- if the field observations are inconsistent with the simulations (irreducible model discrepancy):

$$y_i^F = f(\mathbf{x}_i, \boldsymbol{\theta}^*) + \delta(\mathbf{x}_i) + \epsilon(\mathbf{x}_i)\,.$$

  $\delta(\cdot)$ models the difference between the simulator and the physical system:

$$\delta(\mathbf{x}) = \zeta(\mathbf{x}) - f(\mathbf{x}, \theta^*)\,,$$

  but
  - What does θ^* mean ?
  - A best fitting ?
  - identifiability issues ?
  - usually assumed to be smoother than the real physical process $\zeta(\cdot)$

Ref.: Kennedy and O'Hagan (2001), Hidgon et al. (2005)...

Outline

A calibration example

Hypotheses:

- The simulator represents sufficiently well the physical system:

$$y^F(\mathbf{x}_i) = f(\mathbf{x}_i, \boldsymbol{\theta}^*) + \epsilon_i, \quad i = 1, \ldots, n.$$

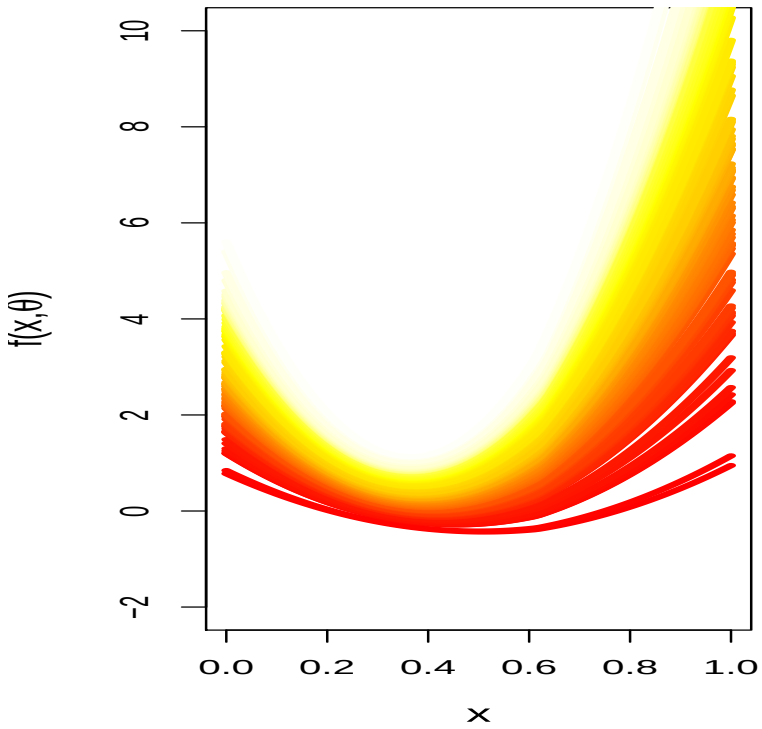
- But unknown $\boldsymbol{\theta}^*$.
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d. with known σ^2.

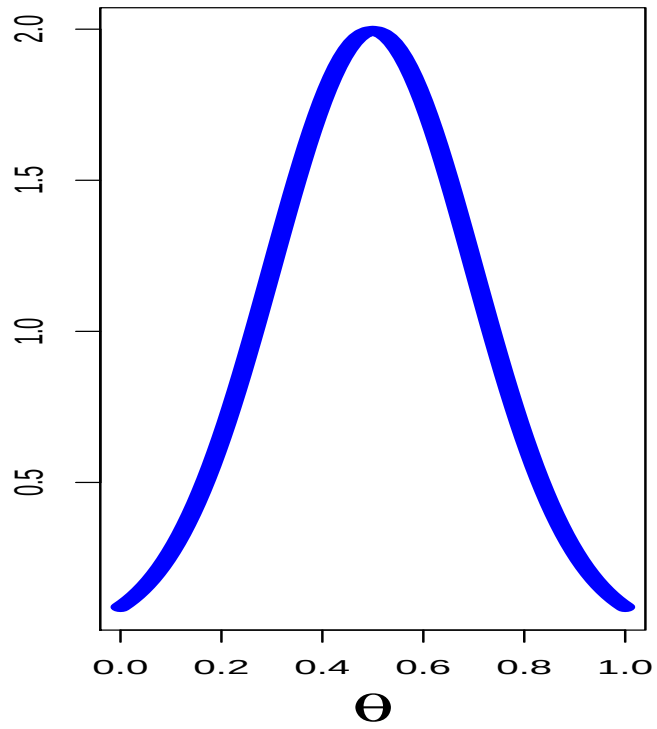
In the next example:

- $\sigma^2 = 0.3$
- $n = 6$,
- $\theta^* = 0.6$

A calibration example

Prior:
prior distribution on unknown $\boldsymbol{\theta}$: $\pi(\cdot)$
from expert judgment, past experiments...
Possible choice $\pi(\boldsymbol{\theta}) = \mathcal{N}(\theta_0, \sigma_0^2) = \mathcal{N}(0.5, 0.04)$.





A calibration example

Data:
Couples $(\mathbf{x}_1, y_1^F), \ldots, (\mathbf{x}_n, y_n^F)$ from physical experiments.

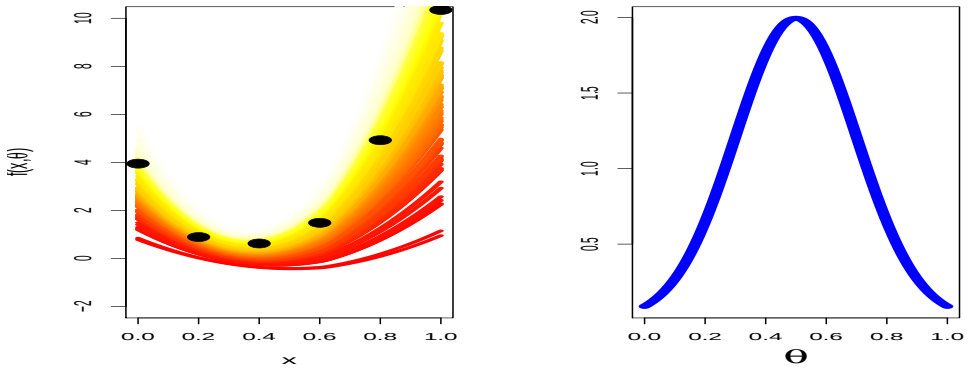
Posterior distribution:

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{y}^F) &\propto l(\boldsymbol{\theta}; \mathbf{y}^F) \cdot \pi(\boldsymbol{\theta}) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y^F(\mathbf{x}_i) - f(\mathbf{x}_i, \boldsymbol{\theta}))^2 - \frac{1}{2\sigma_0^2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^2\right)\end{aligned}$$

- Analytical posterior if $\boldsymbol{\theta} \mapsto f(\mathbf{x}, \boldsymbol{\theta})$ is a linear map,
- Otherwise MH sampling to simulate according to the posterior distribution.

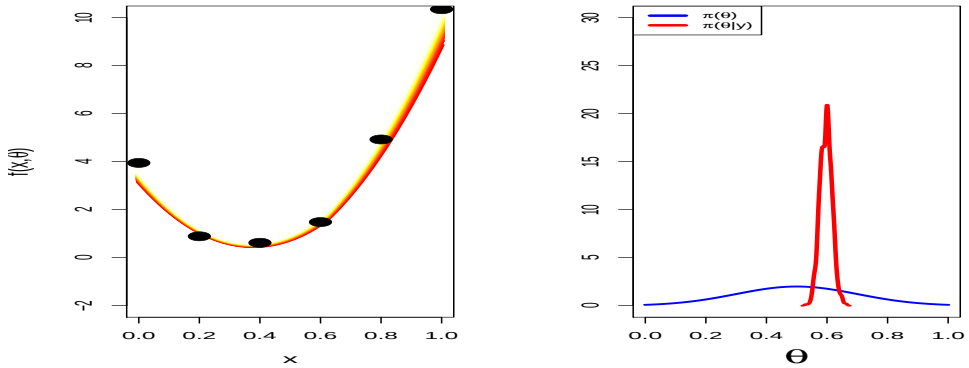
A calibration example

Prior with data:



⇓ Metropolis-Hastings algorithm ⇓

Posterior on θ:



More details on the MH algorithm

Initialisation:
θ^0 chosen.

Update:
iterations $t = 1, \ldots,$

1. Proposal: $\tilde{\theta}^{t+1} = \theta^t + \mathcal{N}(0, \tau^2)$.
2. Compute
$$\alpha(\theta^t, \tilde{\theta}^{t+1}) = \frac{\pi(\tilde{\theta}^{t+1}|\mathbf{y}^F)}{\pi(\theta^t|\mathbf{y}^F)}$$
3. Acceptation:
$$\theta^{t+1} = \begin{cases} \tilde{\theta}^{t+1} & \text{with probability } \alpha(\theta^t, \tilde{\theta}^{t+1}) \\ \theta^t & \text{otherwise.} \end{cases}$$

Note that the ratio $\alpha(\theta^t, \tilde{\theta}^{t+1})$ needs several computations of $f(\mathbf{x}, \boldsymbol{\theta})$ at each step since

$$\pi(\theta|\mathbf{y}^F) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y^F(\mathbf{x}_i) - f(\mathbf{x}_i, \boldsymbol{\theta}))^2 - \frac{1}{2\sigma_0^2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^2\right).$$

Outline

Expensive black-box computer code

- Run the simulator for a given $(\mathbf{x}^*, \boldsymbol{\theta})$ is time-consuming / expensive.
- The simulator is a black-box, no intrusive methods are possible.

$\Rightarrow$ Only few runs of the simulator are possible then we cannot apply algorithms (as in Bayesian calibration) which make a massive use of simulator runs.

Using an emulator / metamodel / coarse model / approximation of the simulator which is fast to compute, but:

- loss on precision of prediction,
- new uncertainty source: accuracy of the model approximation,
- taken into account.

Choosing a design of numerical experiments

Desing of experiments $\mathbf{D_M}$: Choosing M couples

$$(\mathbf{x}_j^*, \boldsymbol{\theta}_j)$$

- space filling for x,
- with respect to the prior distribution on $\boldsymbol{\theta}$,
- $\mathbf{x}_j^* = \mathbf{x}_i$?

where the simulator is called.

Emulator using Gaussian Process:

- Very popular in computer experiments.
- integrated in a Bayesian framework: appears in the likelihood function and a prior on the parameters of the Gaussian process are chosen.
- model uncertainty coming from approximation of f.
- After the calibration step, used in prediction for a new point $\mathbf{x}$.

Meta-modeling: prior distribution on f

Sacks et al. (1989).
Gaussian process prior F on function f:
$\forall (\mathbf{x}^*, \boldsymbol{\theta}) \in E$,

$$F(\mathbf{x}^*, \boldsymbol{\theta}) = \sum_{k=1}^{Q} \beta_k h_k(\mathbf{x}^*, \boldsymbol{\theta}) + Z(\mathbf{x}^*, \boldsymbol{\theta}) = H(\mathbf{x}^*, \boldsymbol{\theta})^T \boldsymbol{\beta} + Z(\mathbf{x}^*, \boldsymbol{\theta}),$$

where

- $h_1, \ldots, h_Q$ regression functions and $\boldsymbol{\beta}$ parameters vector,
- Z centered Gaussians process with covariance function:

$$\mathrm{Cov}(Z(\mathbf{x}_1^*, \boldsymbol{\theta}_1), Z(\mathbf{x}_2^*, \boldsymbol{\theta}_2)) = \sigma^2 K((\mathbf{x}_1^*, \boldsymbol{\theta}_1), (\mathbf{x}_2^*, \boldsymbol{\theta}_2)),$$

  where K is correlation kernel.

Hypotheses

- $K((\mathbf{x}_1^*, \boldsymbol{\theta}_1), (\mathbf{x}_2^*, \boldsymbol{\theta}_2)) = \sigma_K^2 \exp(-\xi_{\mathbf{x}^*} \sum |\mathbf{x}_1^* - \mathbf{x}_2^*|^\alpha - \xi_{\boldsymbol{\theta}} \sum |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|^\alpha)$
- parameters $\phi = (\boldsymbol{\beta}, \sigma^2, K$ parameters) assumed fixed (in practice, maximum likelihood estimators);

Meta-modeling: posterior

- $v_1 = f((\mathbf{x}^*, \boldsymbol{\theta})_1), \ldots, v_M = f((\mathbf{x}^*, \boldsymbol{\theta})_M)$ evaluations of f on a design $\mathbf{D}_M$
- Process F^{D_M}: Conditioning F to $F((\mathbf{x}_1^*, \boldsymbol{\theta}_1)) = v_1, \ldots, F(\mathbf{x}_M^*, \boldsymbol{\theta}_M)) = v_M$. Gaussian Process with mean $m(\mathbf{x}^*, \boldsymbol{\theta})$ and covariance $C((\mathbf{x}^*, \boldsymbol{\theta}), (\mathbf{x}^*, \boldsymbol{\theta})') \; \forall (\mathbf{x}^*, \boldsymbol{\theta}), (\mathbf{x}^*, \boldsymbol{\theta})'$.

  For all $(\mathbf{x}^*, \boldsymbol{\theta}) \in E$,
  - $m(\mathbf{x}^*, \boldsymbol{\theta})$ approximates $f(\mathbf{x}^*, \boldsymbol{\theta})$,
  - $C((\mathbf{x}^*, \boldsymbol{\theta}), (\mathbf{x}^*, \boldsymbol{\theta}))$ uncertainty on this approximation.

For all $(\mathbf{x}_i^*, \boldsymbol{\theta}_i) \in D_M$,
- $m(\mathbf{x}_i^*, \boldsymbol{\theta}_i) = f(\mathbf{x}_i^*, \boldsymbol{\theta}_i)$,
- $C((\mathbf{x}_i^*, \boldsymbol{\theta}_i), (\mathbf{x}_i^*, \boldsymbol{\theta}_i)) = 0$.

Gaussian process emulator: illustration

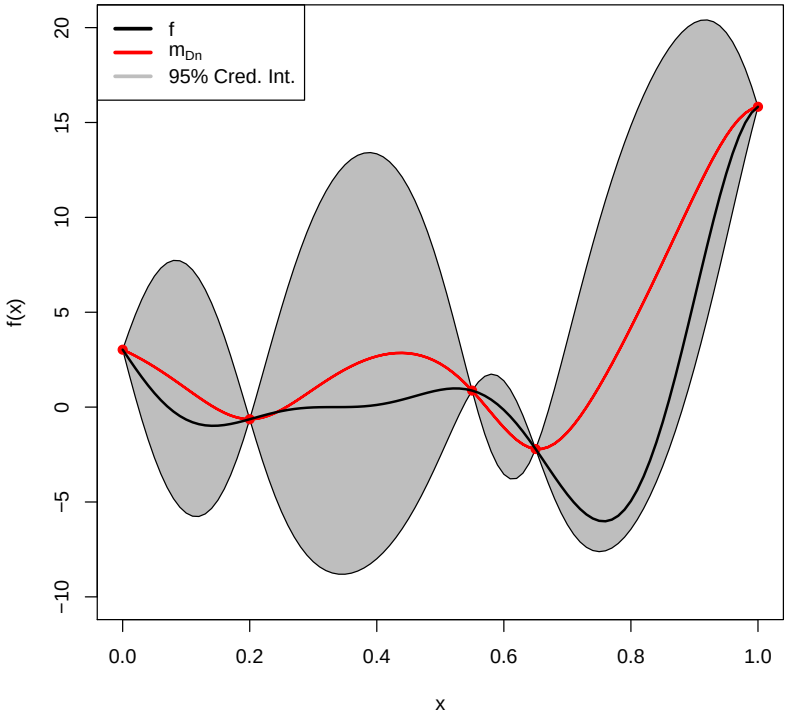


Figure: Posterior mean and pointwise credible interval

Outline

Likelihood with a Gaussian process hypothesis on f

- $\mathbf{z} = (\mathbf{y}^F_1, \ldots, \mathbf{y}^F_n, f(\mathbf{x}^*_1, \boldsymbol{\theta}_1), \ldots, f(\mathbf{x}^*_M, \boldsymbol{\theta}_M)) = (\mathbf{y}^F, f(\mathbf{D}_M))$
- likelihood on $\mathbf{z}$

$$l(\boldsymbol{\theta}, \sigma^2, \phi; \mathbf{z}) \propto |\Sigma_{\mathbf{z}}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T \Sigma_{\mathbf{z}}^{-1} (\mathbf{z}-\boldsymbol{\mu})\right)$$

where

- $\boldsymbol{\mu} = (\mu(\mathbf{x}_1, \boldsymbol{\theta}), \ldots, \mu(\mathbf{x}_n, \boldsymbol{\theta}), \mu(\mathbf{x}^*_1, \boldsymbol{\theta}_1), \ldots, \mu(\mathbf{x}^*_M, \boldsymbol{\theta}_M))^T$ is the mean of the Gaussian process,
- $$\Sigma_{\mathbf{z}} = \Sigma_f + \begin{pmatrix} \Sigma_y & 0 \\ 0 & 0 \end{pmatrix}$$

  with $\Sigma_y = \sigma^2 I_n$ and Σ_f is obtained as the covariance matrix corresponding to the points: $(\mathbf{x}_1, \boldsymbol{\theta}), \ldots, (\mathbf{x}_n, \boldsymbol{\theta}), (\mathbf{x}^*_1, \boldsymbol{\theta}_1), \ldots, (\mathbf{x}^*_M, \boldsymbol{\theta}_M)$.

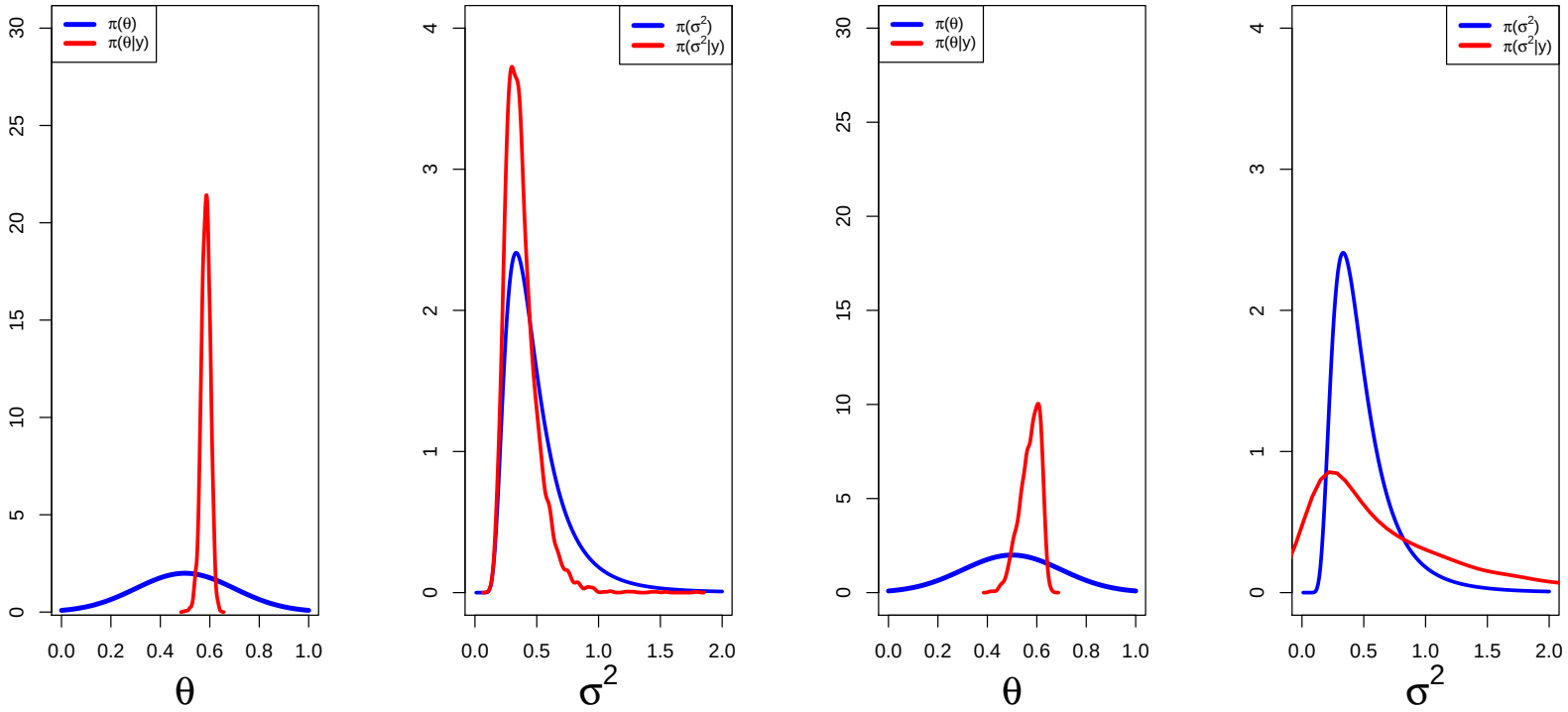
Dealing with GP parameters ϕ

- prior distribution on μ and covariance parameters Hidgon et al. (2005)
  $\Rightarrow$ MCMC inference

- MLE estimators for ϕ Kennedy and O'Hagan (2001)
  - plug-in approach,
  - only computer data $f(\mathbf{D}_M)$ are used ($n < M$) for MLE
    $\Rightarrow$ Modularity approach Liu et al. (2009).
  - likelihood $l(\boldsymbol{\theta}, \sigma^2; \mathbf{y}^F | f(\mathbf{D}_M))$:

$$l(\boldsymbol{\theta}, \sigma^2; \mathbf{y}^F | f(\mathbf{D}_M)) \propto |\tilde{\Sigma}_{\mathbf{y}^F}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y}^F - m(\mathbf{x}, \boldsymbol{\theta}))^T \tilde{\Sigma}_{\mathbf{y}^F}^{-1} (\mathbf{y}^F - m(\mathbf{x}, \boldsymbol{\theta}))\right)$$

    where
    - $m(\cdot)$ is the mean of the GP conditioned to simulator data,
    - $\tilde{\Sigma}_{\mathbf{y}^F} = \Sigma_{\mathbf{y}^F} + \tilde{\Sigma}_f = \sigma^2 I_n + \tilde{\Sigma}_f$ where $\tilde{\Sigma}_f$ is constructed with the covariance function C of the conditioned GP.



unlimited runs versus $M = 12$

Posterior consistency

Proposition

Under the following assumptions:

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

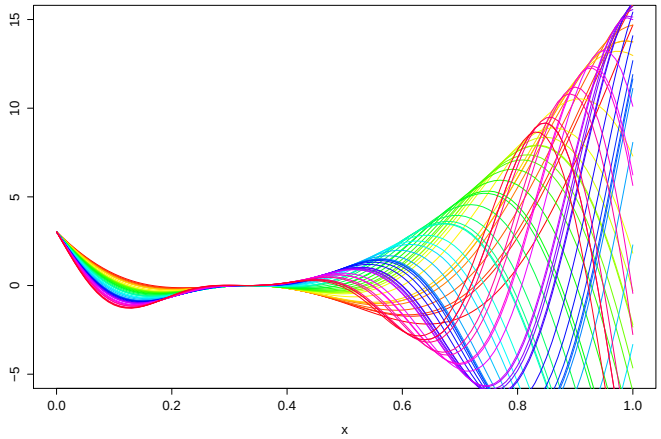
then, we have:

$$\lim_{M \to \infty} KL\left(\pi(\boldsymbol{\theta}|\boldsymbol{y}^F) || \pi^C(\boldsymbol{\theta}|\boldsymbol{y}^F, f(\boldsymbol{D}_M))\right) = 0\,. \tag{1}$$

Motivation for adaptive designs in calibration

Example:

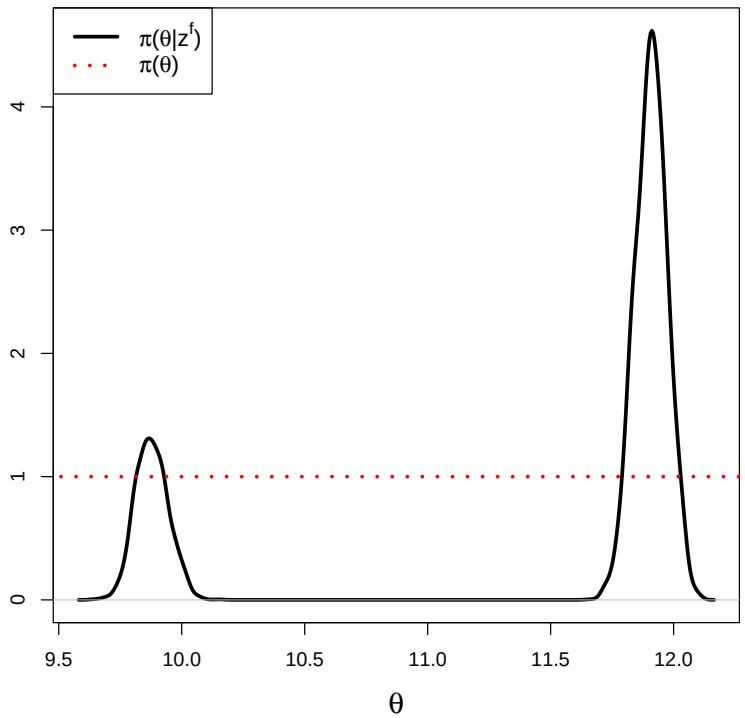
- $\theta = 12$,
- $(x_1, x_2, x_3) = (0.1, 0.3, 0.8)$,
- $f(x, \theta) = (6 \cdot x - 2)^2 \cdot \sin(\theta \cdot x - 4) + \epsilon$,
- $\epsilon_i \sim \mathcal{N}(0, 0.1^2)$ i.i.d.,
- prior $\theta \sim \mathcal{U}[5, 15]$,
- $y_i = f(x_i, \theta) + \epsilon_i$.



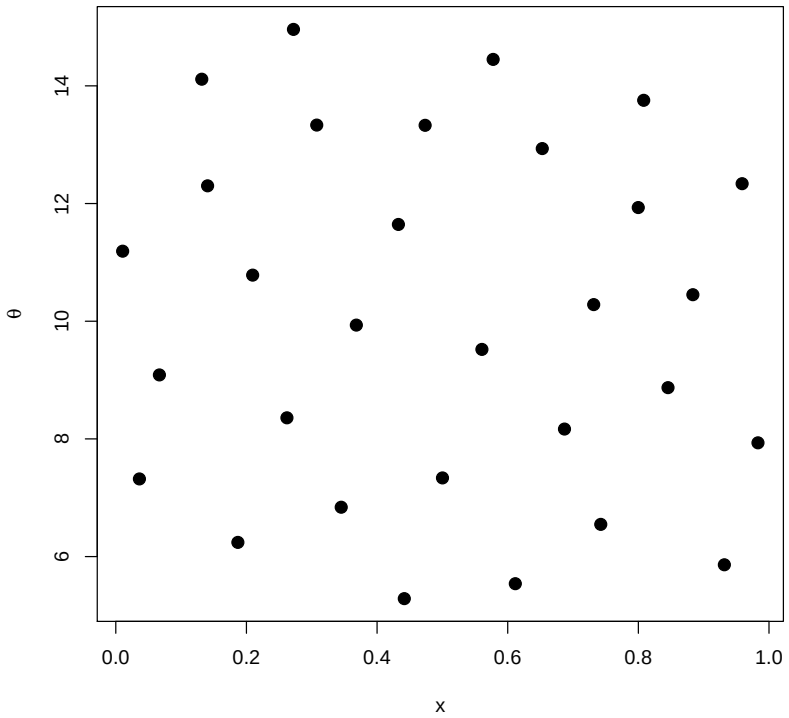
Motivation for adaptive designs in calibration

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

- Calibration with unlimited runs of f

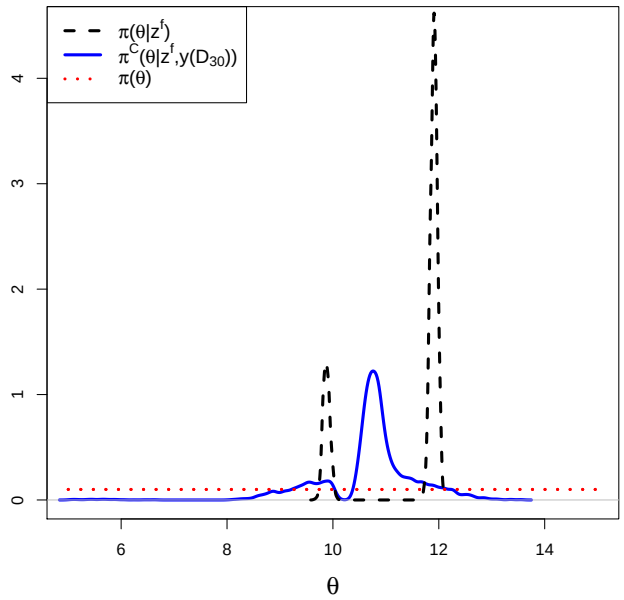


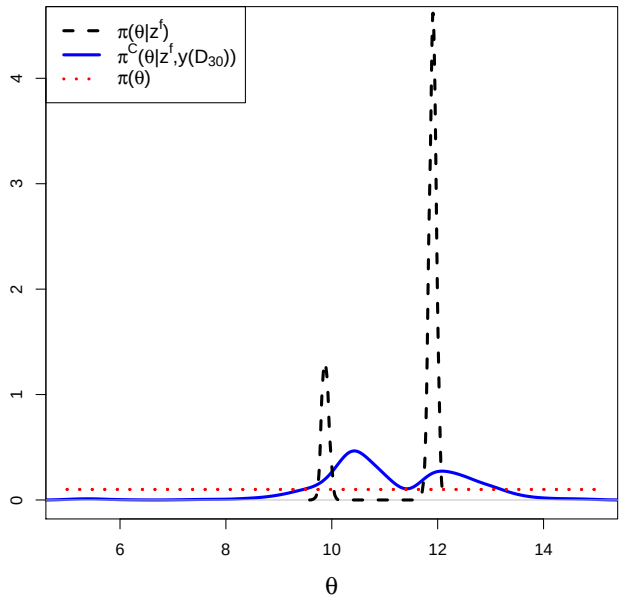
LHS maximin design



Motivation for adaptive designs in calibration

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





Outline

Outline

Efficient Global Optimization

- Goal: Find the global extremum (here minimum e.g.) of f,
- Expected improvement criterion proposed by Jones et al. (1998):

$$EI_n(\mathbf{x}) = \mathbb{E}((min_n - F(\mathbf{x}))^+ | F(D_n)),$$

where min_n is the current minimum value:

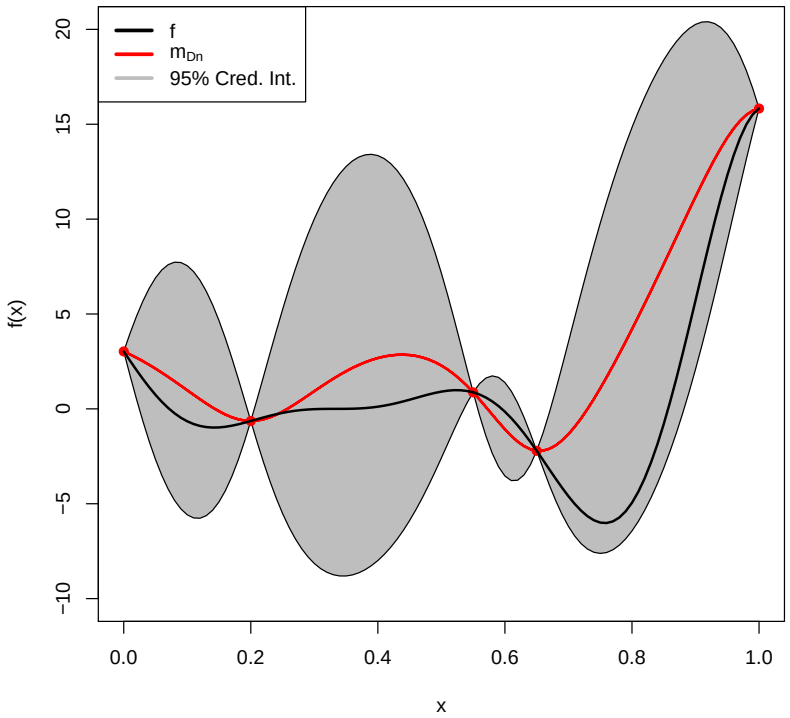
$$min_n = \min_{1,\dots,n} f(\mathbf{x}_i)$$

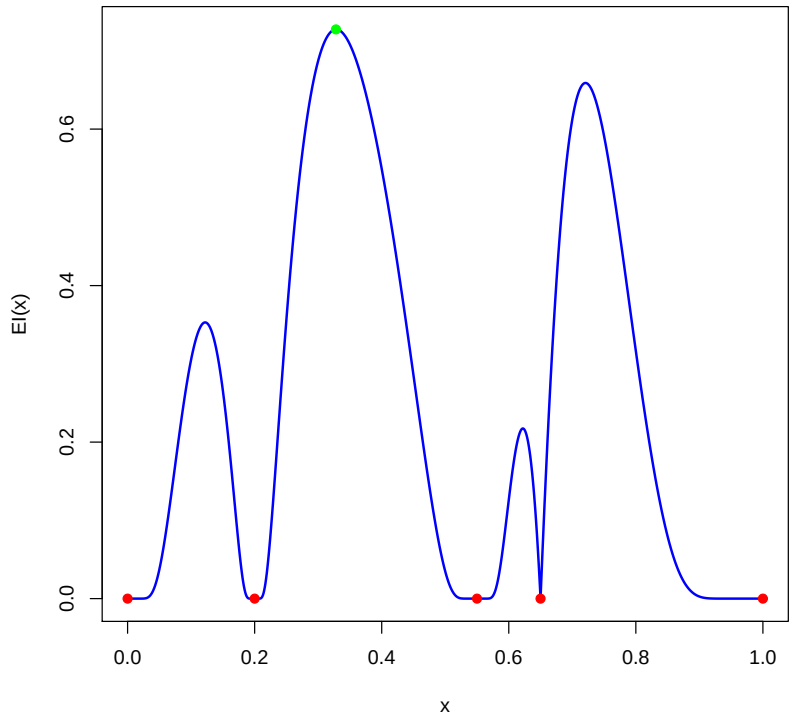
- Closed-form computation:

$$EI_n(\mathbf{x}) = (min_n - m_{D_n}(\mathbf{x}))\Phi\left(\frac{min_n - m_{D_n}(\mathbf{x})}{\sqrt{C_{D_n}(\mathbf{x},\mathbf{x})}}\right) + \sqrt{C_{D_n}(\mathbf{x},\mathbf{x})}\phi\left(\frac{min_n - m_{D_n}(\mathbf{x})}{\sqrt{C_{D_n}(\mathbf{x},\mathbf{x})}}\right)$$

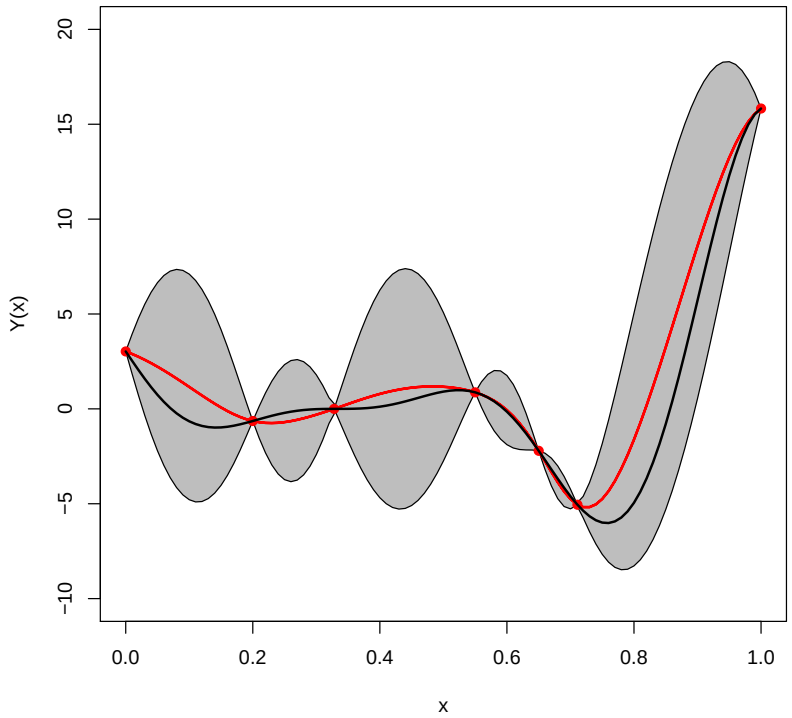
where Φ and ϕ are respectively the cdf and the pdf of $\mathcal{N}(0,1)$.

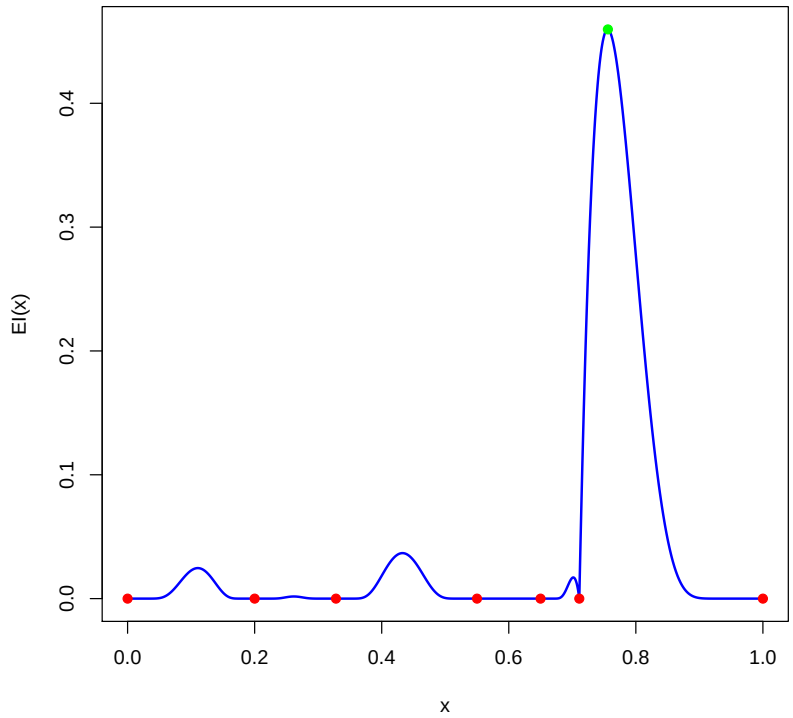
Example step 1



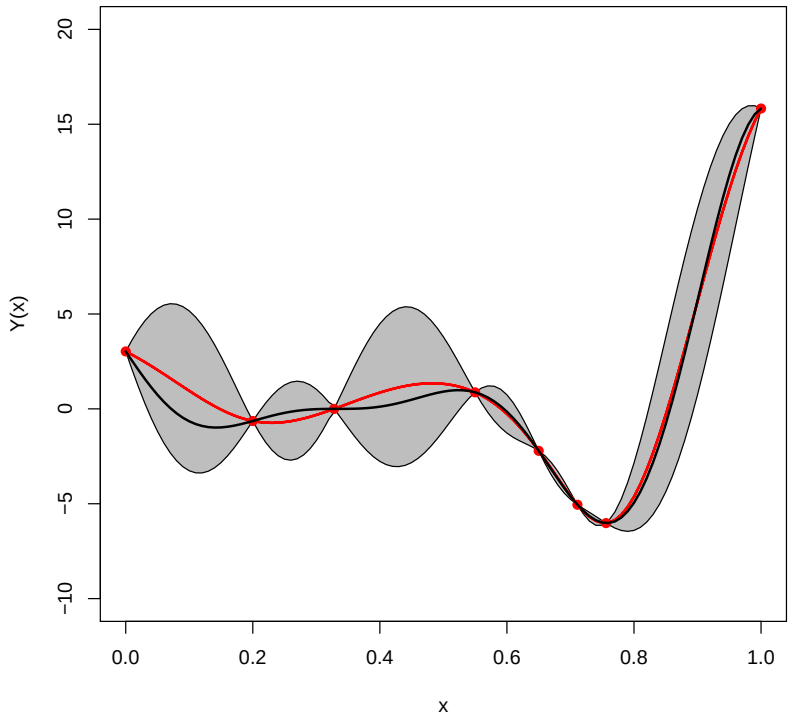


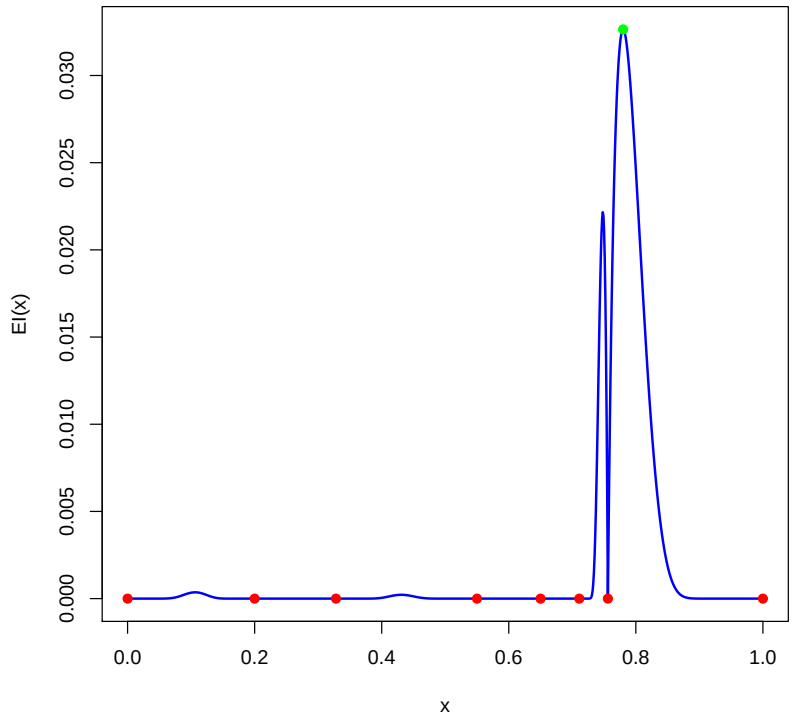
Example step 2



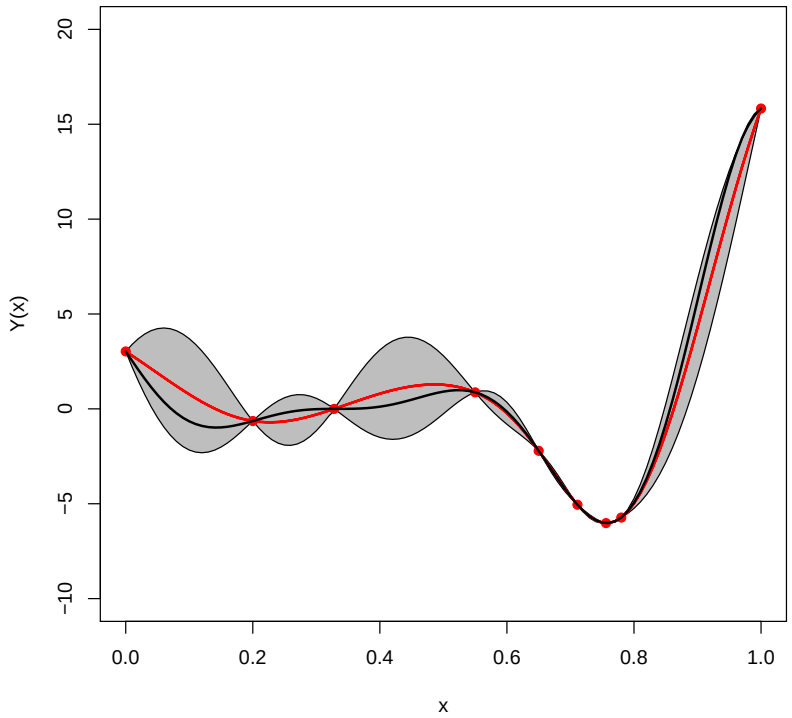


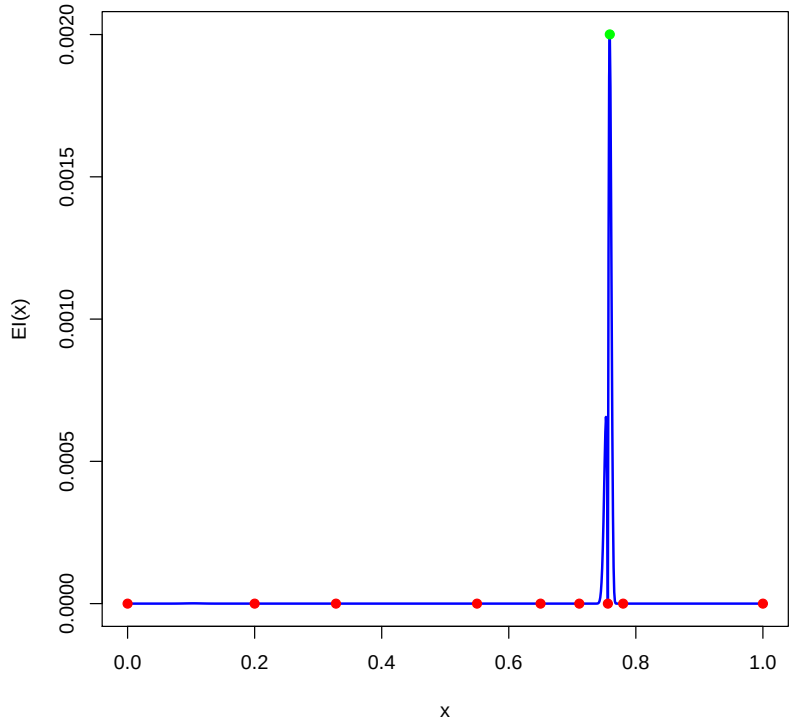
Example step 3



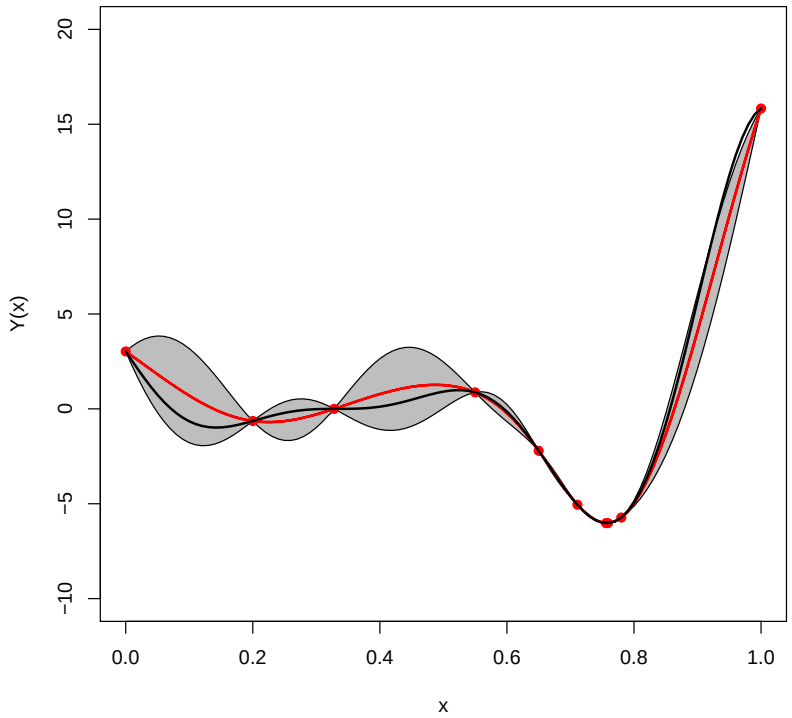


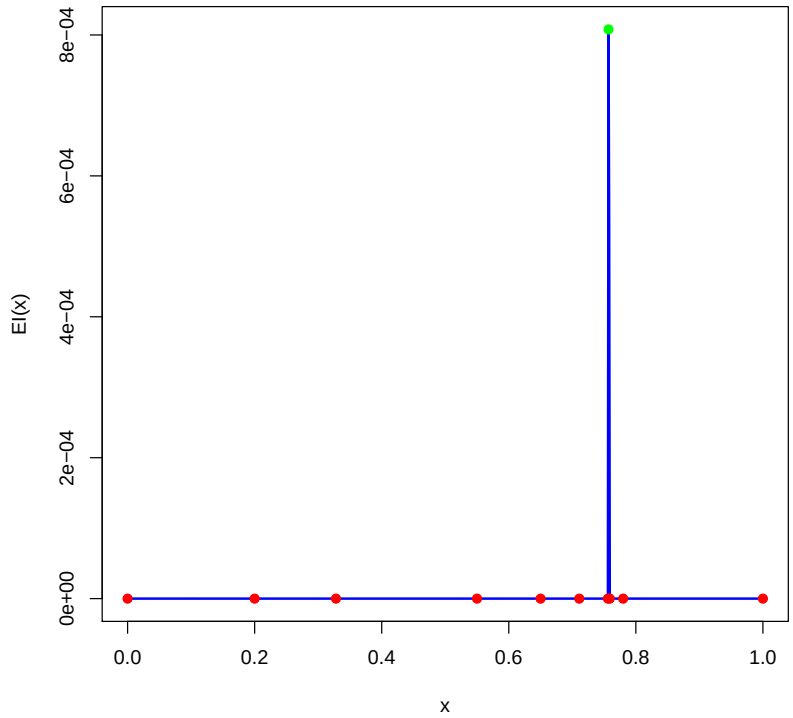
Example step 4





Example step 5





Outline

EI for calibration

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

Maximize the likelihood $l(\boldsymbol{\theta}; \mathbf{z})$ over $\boldsymbol{\theta} \Leftrightarrow$ Minimize $SS(\boldsymbol{\theta}) = \|\mathbf{y}^F - F(\mathbf{x}, \boldsymbol{\theta})\|^2$ over $\boldsymbol{\theta}$.

For given:

- field experiments $\mathbf{y}^F = y^F(\mathbf{x}_1), \ldots, y^F(\mathbf{x}_n)$,
- $\mathbf{D}_k$ numerical design on $\mathcal{X} \times \mathcal{T}$ with M points,
- m_k current minimal value of $SS(\boldsymbol{\theta})$.

EI criterion:

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

to be maximized.

EI criterion is applied to a function of f.

EI computation

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

- 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 $\mathbf{D}_0 \subset \mathcal{X} \times \mathcal{T}$ of size M_0.
- Run the code over $\mathbf{D}_0$, then construct an initial GPE based on $f(\mathbf{D}_0)$.
- Compute $\hat{\theta}_1$ as the posterior mean $\mathbb{E}[\theta|\mathbf{z}^f, f(\mathbf{D}_0)]$.
- $\mathbf{D}_1 = \mathbf{D}_0 \cup \{(\mathbf{x}_i^f, \hat{\theta}_1)\}_{1 \leq i \leq n}$.
- Update the GPE distribution after running the code over $\{(\mathbf{x}_i^f, \hat{\theta}_1)\}_{1 \leq i \leq n}$.
- Compute $m_1 := SS(\hat{\theta}_1)$.

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

Step 1 Find an estimate $\hat{\theta}_{k+1}$ of $\theta_{k+1}^\star = \underset{\theta}{\operatorname{argmax}}\, EI_{\mathbf{D}_k}(\theta)$.

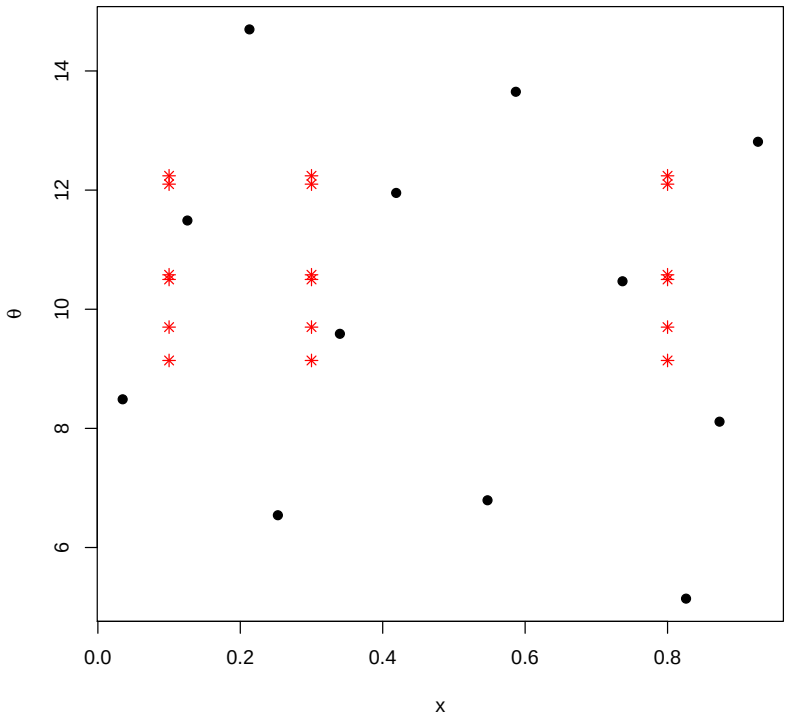
Step 2 $\mathbf{D}_{k+1} = \mathbf{D}_k \cup \{(\mathbf{x}_i^f, \hat{\theta}_{k+1})\}_{1 \leq i \leq n}$.

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

Step 4 Update the GPE distribution based on $f(\mathbf{D}_{k+1})$.

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

Adaptive design



Algorithm one at a time

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

- $\theta_{k+1} = \underset{\theta}{\operatorname{argmax}}\, EI_k(\theta)$,
- $\mathbf{D}_{k+1} = \mathbf{D}_k \cup (\mathbf{x}^\star, \theta_{k+1})$ where $\mathbf{x}^\star \in \mathbf{X}^f = \left[\mathbf{x}_1^f, \cdots, \mathbf{x}_n^f\right]^T$,
- $f(\mathbf{D}_{k+1}) = f(\mathbf{D}_k) \cup \{f(\mathbf{x}^\star, \theta_{k+1})\}$,
- $F^{D_{k+1}} = F|f(\mathbf{D}_{k+1})$,
- $m_{k+1} := \min\{\mathbb{E}[SS_{k+1}(\theta_1)], \cdots, \mathbb{E}[SS_{k+1}(\theta_k)], \mathbb{E}[SS_{k+1}(\theta_{k+1})]\}$.

Only 1 simulation to compute m_{k+1}!

Algorithm one at a time

Two criteria to choose $\mathbf{x}^{\star}_{k+1}$:

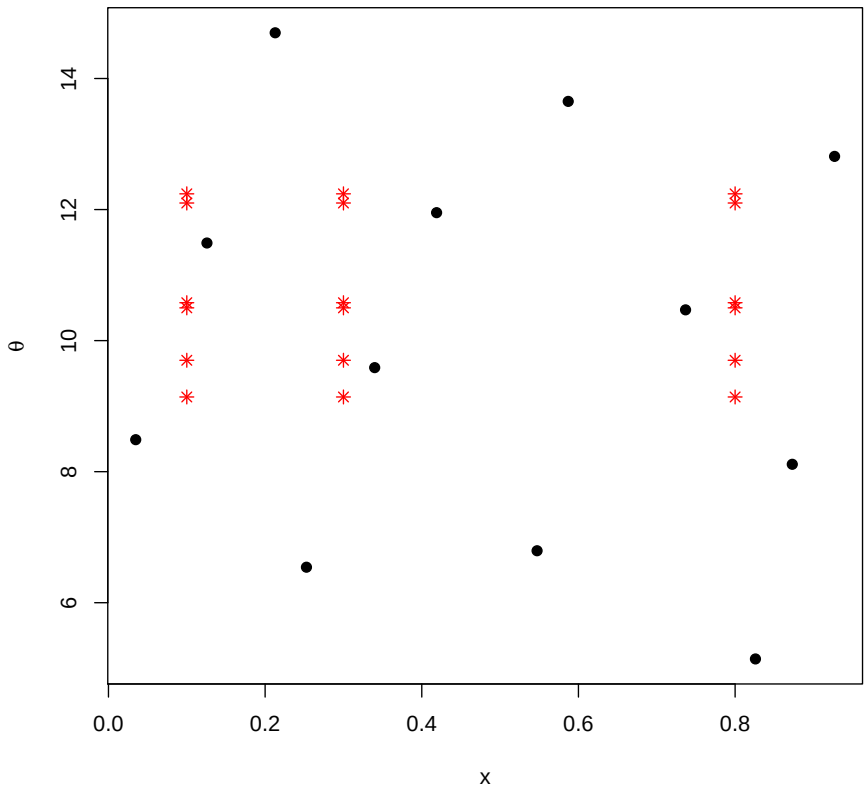
1. $\mathbf{x}^{\star} = \underset{\mathbf{x}_i^f}{\operatorname{argmax}} \operatorname{Var}(F^{\mathbf{D}_k}(\mathbf{x}_i^f, \boldsymbol{\theta}_{k+1}))$,
2. $\mathbf{x}^{\star} = \underset{\mathbf{x}_i^f}{\operatorname{argmax}} \left(\frac{\operatorname{Var}\left(F^{D_k}(\mathbf{x}_i^f, \boldsymbol{\theta}_{k+1})\right)}{\max\limits_{i=1,\cdots,n} \operatorname{Var}\left(F^{D_k}(\mathbf{x}_i^f, \boldsymbol{\theta}_{k+1})\right)} \times \frac{\operatorname{Var}(m^k(\mathbf{x}_i^f, \boldsymbol{\theta}))}{\max\limits_{i=1,\cdots,n} \operatorname{Var}(m^k(\mathbf{x}_i^f, \boldsymbol{\theta}))} \right)$

where $\operatorname{Var}(m^k(\mathbf{x}_i^f, \boldsymbol{\theta}))$ is computed with respect to $\pi(\boldsymbol{\theta})$.

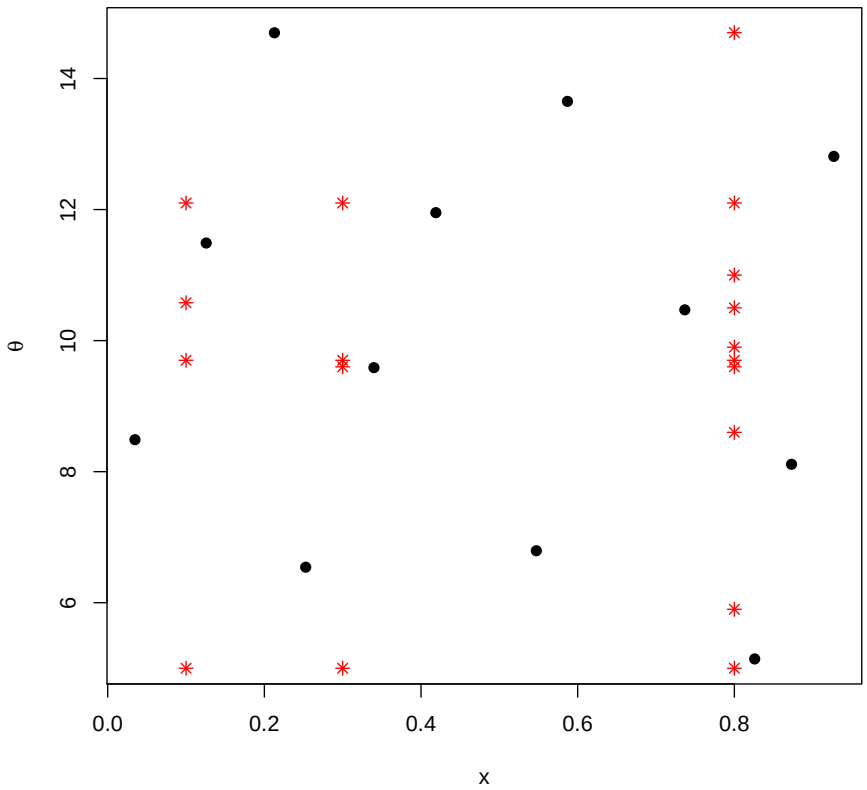
criterion 2 = trade-off between uncertainty on $\mathbf{F^{D_k}}$ for $(\mathbf{x_i^f}, \boldsymbol{\theta}_{\mathbf{k+1}})$ and sensibility of $\mathbf{F^{D_k}}(\mathbf{x_i^f}, \boldsymbol{\theta})$ to $\boldsymbol{\theta}$

Comparison full EI / EI one at a time

Figure: ***full EI***

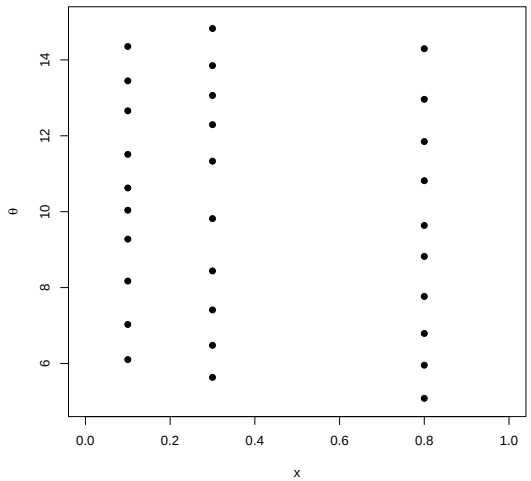


EI *OAT*



Additional comments

- Choice of $\mathbf{D}_0$: *restricted*-to-$\mathbf{X}^f$ maximin LHD, so called because they are maximin on $\mathbf{X}^f \times \mathcal{T}$ and LH on $\mathcal{T}$.



- A GPE can be used to emulate $\boldsymbol{\theta} \rightarrow SS(\boldsymbol{\theta})$ instead of f (Pratola et al., 2013) but OAT strategies not available.

Outline

Sobol function

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

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

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 $\mathcal{T}$:

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

Designs

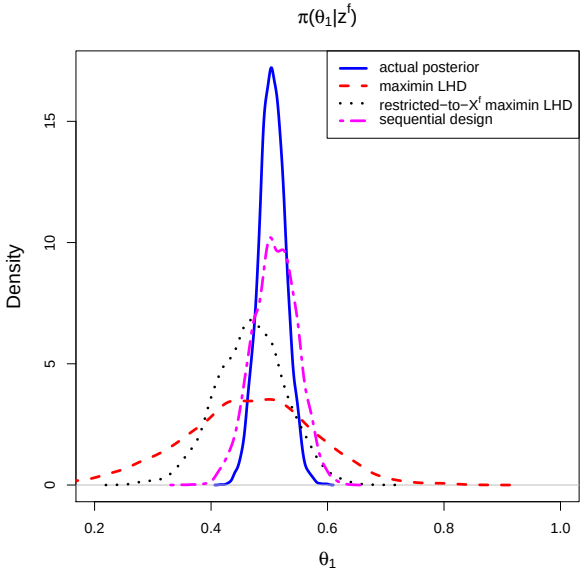
Number of simulations $M = 150$.

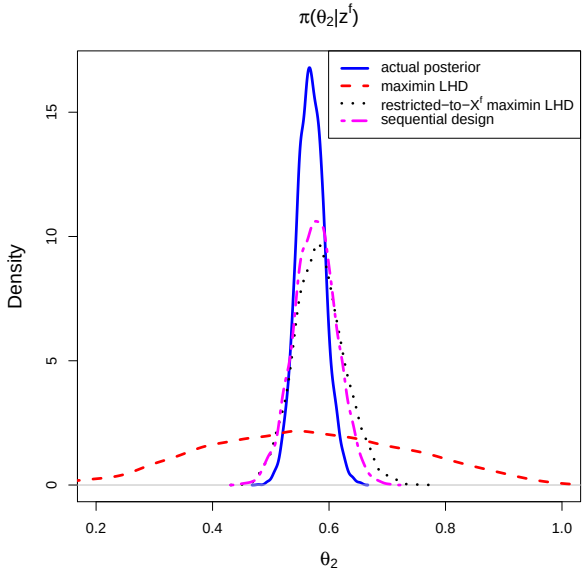
Comparison of 4 designs.

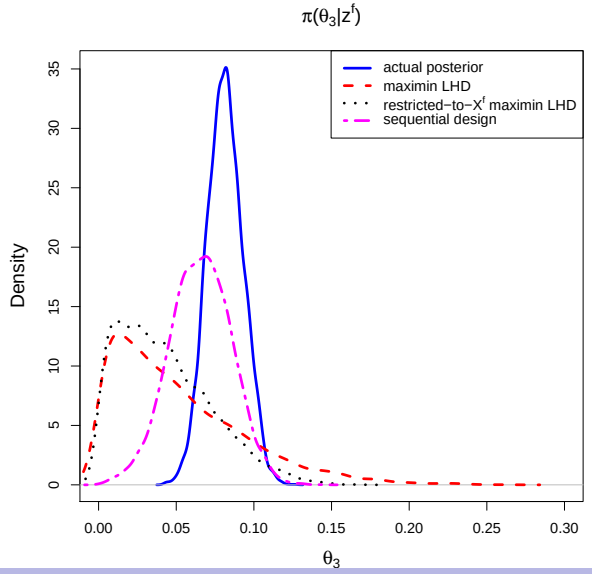
1. Maximin LHD in 6D: $\mathcal{X} \times \mathcal{T} = [0, 1]^6$.
2. *Restricted*-to-$\mathbf{X}^f$ maximin LHD.
3. Sequential designs OAT with GPE variance criterion for choosing $\mathbf{x}_{k+1}^\star$.
4. Sequential designs OAT with trade-off (GPE-variance, variability of f w.r.t. $\mathbf{x}$)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

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



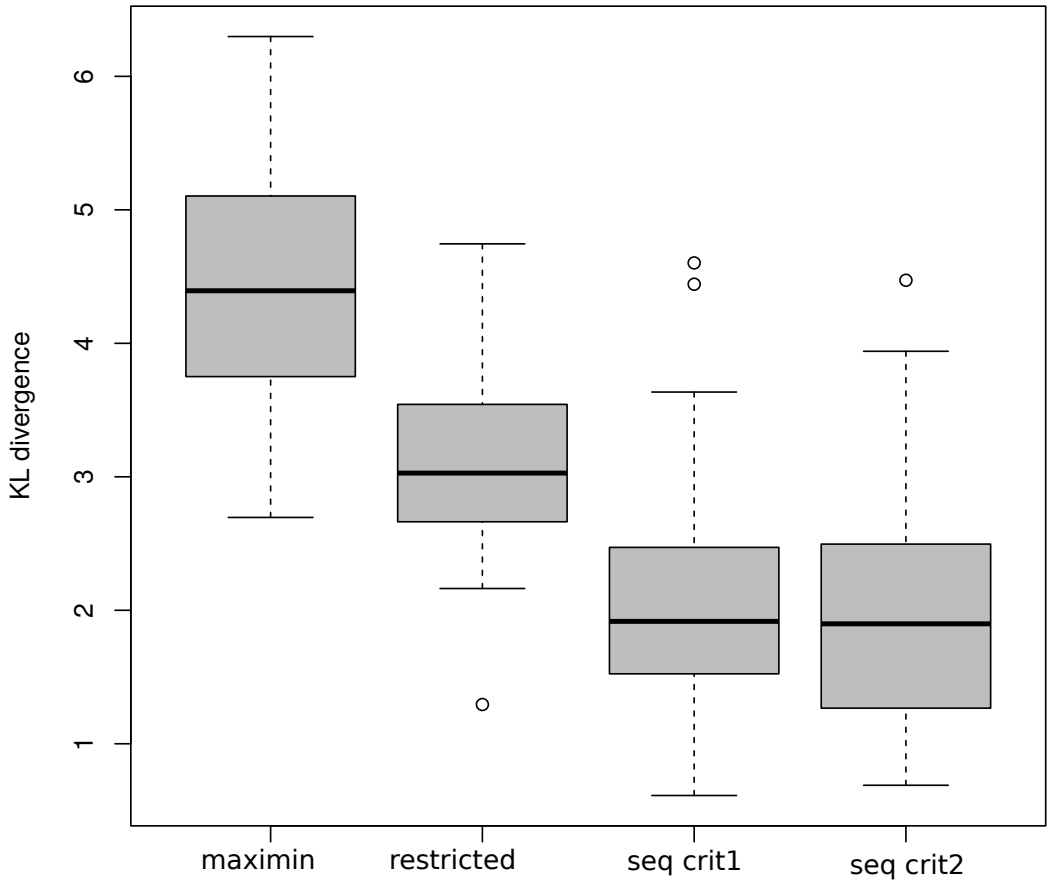




Assessing Robustness

- Simulation of 50 data set with θ sampled uniformly $[0,1]^3$.
- For each data set, surrogate posterior distribution sampled 10 times according to different GPEs constructed with $M = 150$ simulations.
- Performance of Designs evaluated in terms of the KL divergence between the surrogate posterior distribution and the actual posterior distribution.
- Boxplots are made with 50 values (one per data set), each of them being calculated as the mean of the 10 KL divergence values.

Robustness


KL divergence
1
2
3
4
5
6
maximin
restricted
seq crit1
seq crit2

Outline

Conclusion and perspective

What is done:

- Designs of numerical experiments adapted to calibration purpose (restriction of the space and sequential strategy),
- Robustness in calibration.

Open questions:

- Higher dimension questions, number of field experiments, efficient search of $\theta^\star$.
- New field experiments ?
- discrepancy issues ?

References

Damblin, G., Barbillon, P., Keller, M., Pasanisi, A. et Parent, É (2015). Adaptive numerical designs for the calibration of computer codes.

arXiv:1502.07232.

Calibration of computer models

- Dave Hidgon et al., 2005. Combining Field Data and Computer Simulations for Calibration and Prediction. SIAM 26(2).
- Marc Kennedy and Anthony O'Hagan, 2001. Bayesian Calibration of Computer Models. Journal of the Royal Statistical Society B 68.
- Jenny Brynjarsdóttir and Anthony O'Hagan, 2013. J. of Uncertainty Quantification.
- Liu, F., Bayarri, M., and Berger, J. (2009). Modularization in Bayesian analysis, with emphasis on analysis of computer models. Bayesian Analysis, 4(1):119–150.
- Pratola, M., Sain, S., Bingham, D., Wiltberger, M., and Rigler, E. (2013). Fast sequential computer model calibration of large nonstationary spatial-temporal processes. Technometrics, 55(2):232–242.

Optimization

- Donald Jones et al., 1998. Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization 13(4).

Gaussian Process emulator:

- Thomas Santner et al., 2003. The Design and Analysis of Computer Experiments. Springer-Verlag.
- Kai-Tai Fang et al., 2006. Design and Modeling for Computer Experiments. Computer Science and Data Analysis. Chapman & Hall/CRC.

R

- packages Dice....