Designs of experiments for computer code calibration

Adama BARRY

Thesis supervisors: François BACHOC[1] & Clémentine PRIEUR[2]

Promoters: Sarah BOUQUET[3] & Miguel MUNOZ ZUNIGA[3]

[1]Institut de Mathématiques de Toulouse

[2]Université Grenoble Alpes & INRIA Grenoble

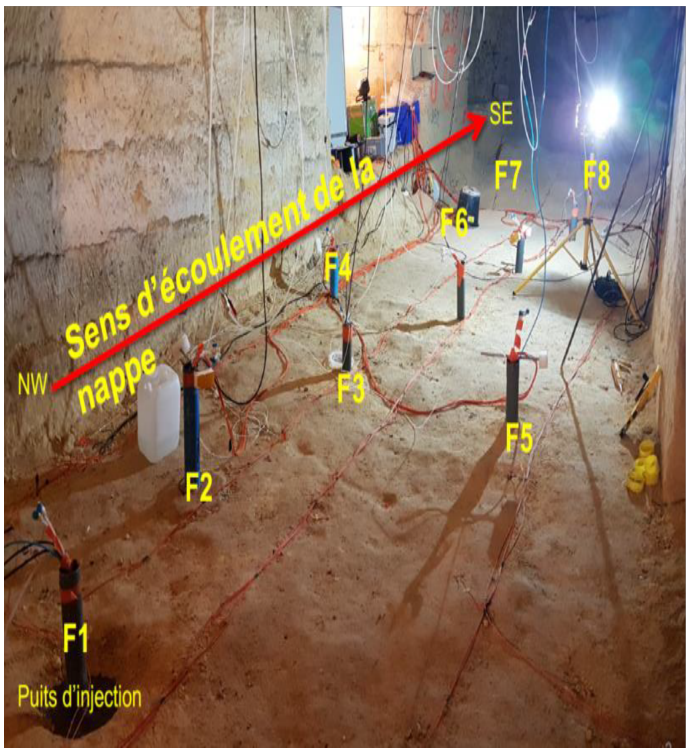
[3]IFP Énergies Nouvelles

ETICS - Octobre 2023

Plan

Introduction

Application context: monitoring a CO_2 leak



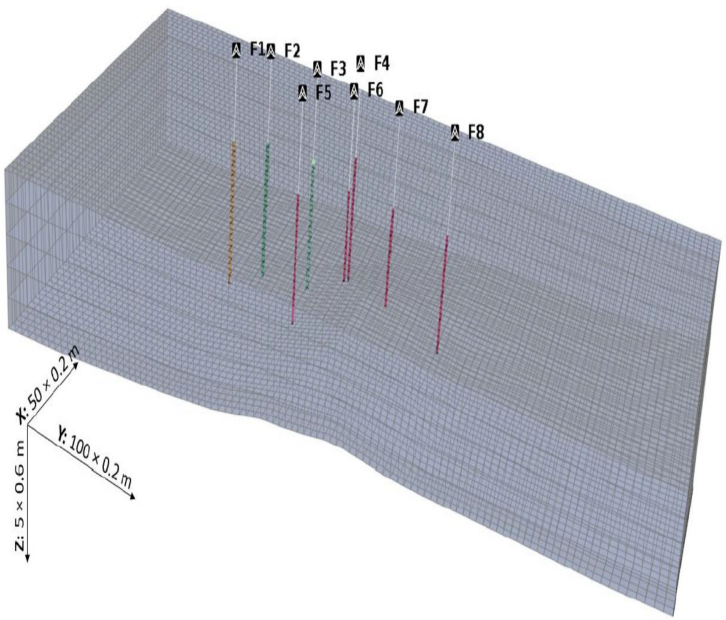
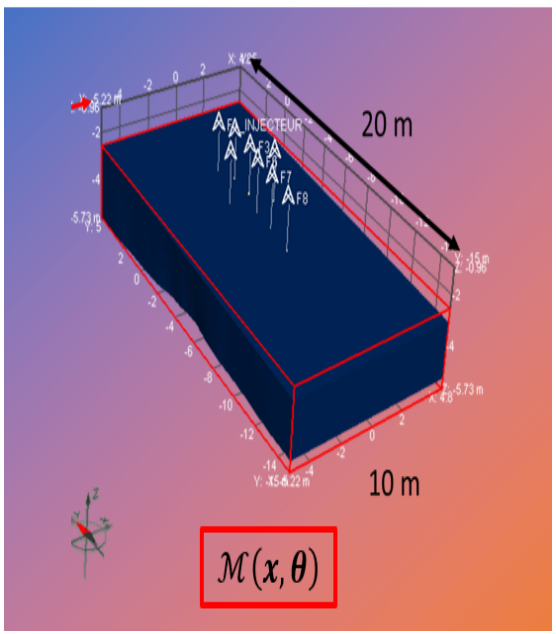


Figure 1: Monitoring of CO_2 concentration.

- Physical phenomenon: CO_2 concentration in observation wells over time.
- Control variables: position of 6 observation wells $(F_2, F_3 \ldots, F_7)$.

Application context: computer code



- Computer code: 3D flow model.
- Input variables:
  - Control variable x: position of observation wells.
  - Calibration parameters θ:

Parameter	Min	Max
disp. Longitudinal (m)	0.13	4.5
disp. Transversal (m)	0.04	1.5
Porosity	0.2	0.45
Permeability	0.1	48.24

Computer code as black box function

- $\mathcal{X}$: experimental domain.
- Θ: parameters domain.

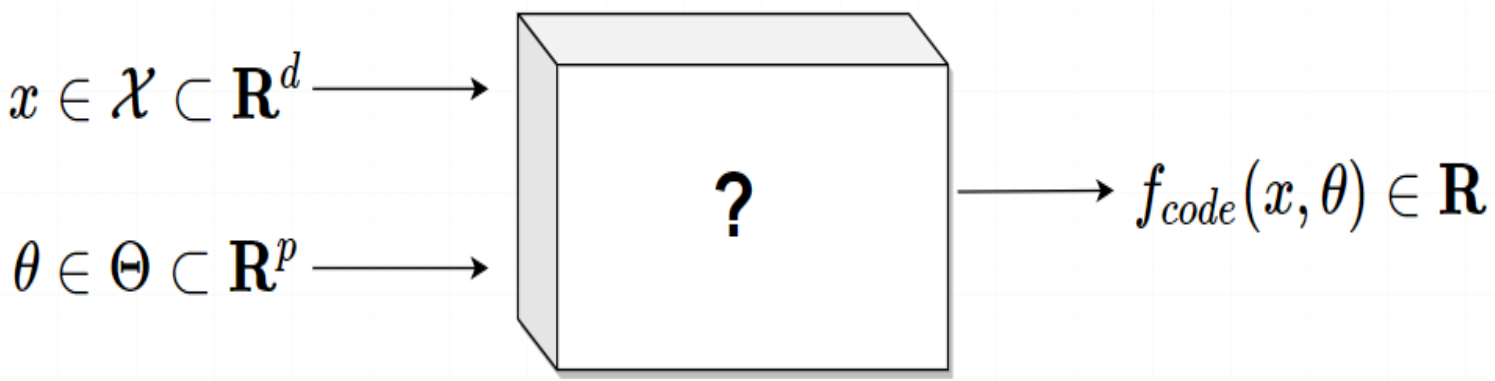


Figure 2: Scalar output computer code.

Statistical framework

- Relationship between physical observation and simulation:

$$\mathbf{Y}_{\text{obs}}(x) = f_{\text{code}}(x, \theta_0) + \varepsilon_x,$$

where $\theta_0 \in \Theta$ the true value assumed to be all influential, $\varepsilon_x \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and σ_ε^2 is assumed known.

- Design of physical experiments $X_{\text{obs}} = \left[x^{(1)}, \ldots, x^{(n)}\right]^T$.
- Physical observations at X_{obs}: $Y_{\text{obs}} = \left[y_{\text{obs}}^{(1)}, \ldots, y_{\text{obs}}^{(n)}\right]^T$.
- Statistical modelling of observation:

$$Y_{\text{obs}} = \{\mathbf{Y}_{\text{obs}}(x^{(1)}) = y_{\text{obs}}^{(1)}, \ldots, \mathbf{Y}_{\text{obs}}(x^{(n)}) = y_{\text{obs}}^{(n)}\}.$$

- Prior distribution: $\pi_0(\theta)$.
- Exact posterior distribution:

$$\pi(\theta \mid Y_{\text{obs}}) \propto \frac{1}{\sigma_\varepsilon^n} \exp\Big(-\frac{1}{2\sigma_\varepsilon^2} \text{SS}(\theta)\Big) \pi_0(\theta),$$

where $\text{SS}(\theta) = \sum_{i=1}^n \left(y_{\text{obs}}^{(i)} - f_{\text{code}}(x^{(i)}, \theta)\right)^2$.

Statistical framework

Goals

- The choice of a design of physical experiments $X_{\text{obs}} = \left[x^{(1)}, \ldots, x^{(n)}\right]^T$ to minimize uncertainty about θ_0.
- Estimate the vector of parameters θ_0.

Approaches:

- Construct a criteria for selecting the physical experiments to carried out.
- Reduce simulation cost by using a metamodel: need of a dedicated criterion for the selection of numerical experiments.

GP metamodel

Gaussian process metamodel

Goal

Build a metamodel for the computer code.

A priori: f_{code} is the realization of a Gaussian process

$$\mathbf{Y}_{\text{code}} \sim \mathbf{GP}(m_\beta, k),$$

where

- $m_\beta : (x, \theta) \longmapsto m_\beta(x, \theta) = h(x, \theta)^T \beta$: mean function,
- $k : \big((x, \theta), (x', \theta')\big) \longmapsto k\big((x, \theta), (x', \theta')\big)$: covariance function.

Consider the following:

- $D_M = \big[(x_1, \theta_1), \ldots, (x_M, \theta_M)\big]^T$ the design of numerical experiment,
- $f_{\text{code}}(D_M) = \big[f_{\text{code}}(x_1, \theta_1), \ldots, f_{\text{code}}(x_M, \theta_M)\big]^T$ the simulations.

A posteriori: $\mathbf{Y}^M_{\text{code}} := \Big[\mathbf{Y}_{\text{code}} \mid \mathbf{Y}_{\text{code}}(D_M) = f_{\text{code}}(D_M)\Big] \sim \mathbf{GP}\Big(\mu^M, k^M\Big)$,
where μ^M et k^M are respectively the a posteriori mean function and the a posteriori covariance function.

Approximation of the exact a posteriori distribution

Preliminary:

- The metamodel parameters are estimated using the modularization technique (Liu, Bayarri, and Berger 2009).
- Assumption: A priori on θ and f_{code} independent.

Approximation of the exact posterior distribution

$$\pi(\theta \mid Y_{obs}, f_{\text{code}}(D_M)) \propto \mathcal{L}^c(Y_{obs} \mid f_{\text{code}}(D_M), \theta)\pi_0(\theta). \tag{1}$$

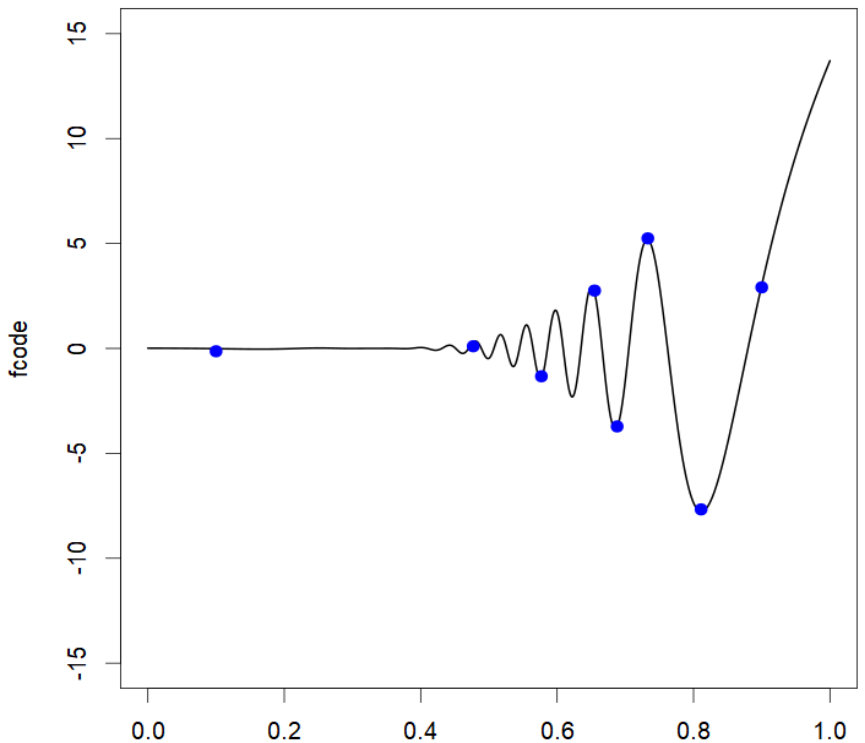
with

$$\begin{aligned}\mathcal{L}^c\left(Y_{obs} \mid \theta, f_{\text{code}}(D_M)\right) = & \frac{1}{(2\pi)^{n/2}\det\left(k^M(D_\theta)+\sigma_\varepsilon^2 I_n\right)^{1/2}} \\ & \times \exp\left[-\frac{1}{2}\,||\; Y_{obs}-\mu^M(D_\theta)\;||_{k^M(D_\theta)+\sigma_\varepsilon^2 I_n}\right],\end{aligned} \tag{2}$$

where $D_\theta = \{(x^{(1)},\theta),\ldots,(x^{(n)},\theta)\}$.

Design Of Physical Experiments

Motivation



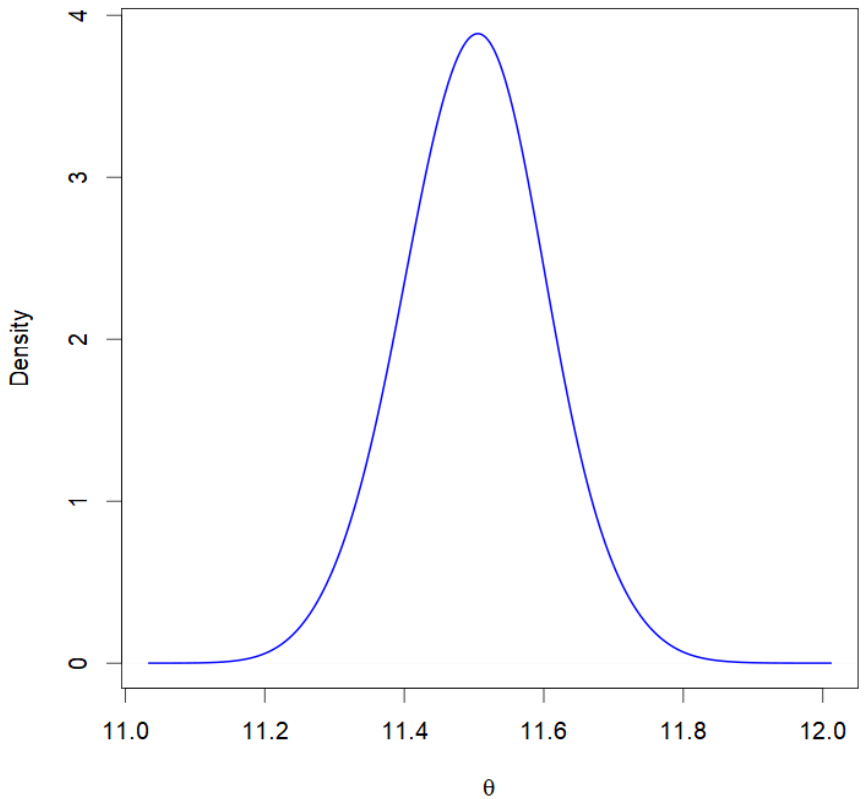


Figure 3: DOPE and a posteriori density ($\theta_0 = 12$).

Motivation

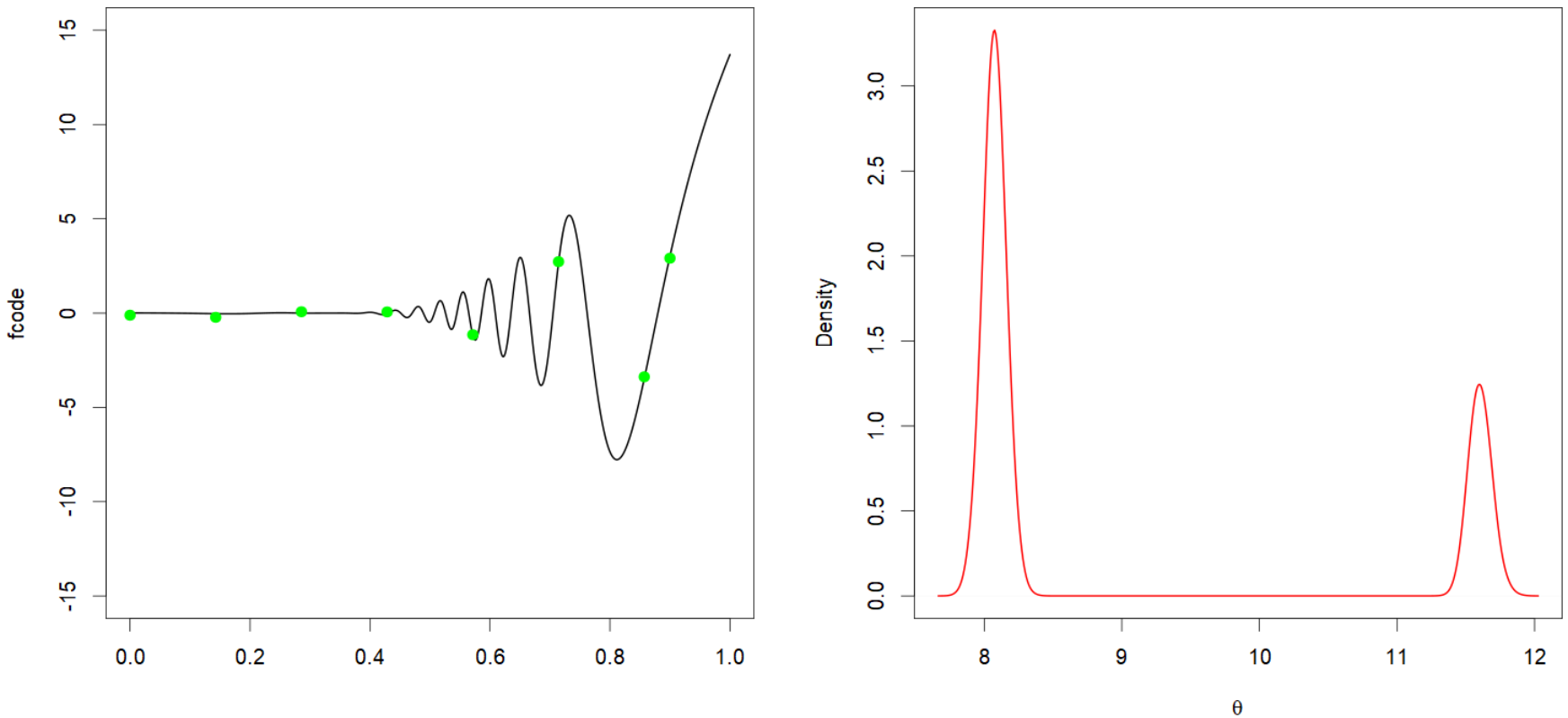


Figure 4: DOPE and a posteriori density ($\theta_0 = 12$).

Criterion of Design Of Physical Experiments (DOPE)

- $X = \{x^{(1)}, \ldots, x^{(n)}\} \in \mathcal{X}^n$ is a design of physical experiments.

Goal

Choose the optimal design of physical experiments X_{obs} such that:

$$X_{\text{obs}} = \arg\min_{X \in \mathcal{X}^n} \mathbf{C}(X) \text{ or } \arg\max_{X \in \mathcal{X}^n} \mathbf{C}(X),$$

where $\mathbf{C}(X)$ is the design criterion to be defined.

Two main types of criterion in litterature:

1. Criterion based on Fisher's information matrix.
2. Criterion based on posterior distribution (DAP).

Fisher information matrix's criteria for DOPE

- Computer code approximation: $\mathbf{Y}^M_{code} \sim \mathbf{GP}(\mu^M, k^M)$ a metamodel using $(D_M, f_{\text{code}}(D_M))$, where D_M is an initial design on $\mathcal{X} \times \Theta$,
- The information matrix in θ_0 is given by:

$$\mathbf{M}(X, \theta_0) = J(X, \theta_0)^T \Big[k^M\big((X, \theta_0), (X, \theta_0)\big) + \sigma_\varepsilon^2 I_n \Big]^{-1} J(X, \theta_0).$$

where $J(X, \theta_0) = \left(\dfrac{\partial \mu^M(x^{(i)}, \theta_0)}{\partial \theta_j} \right)_{1 \le i \le n, 1 \le j \le p}$ is the Jacobian.

- The Bayesian information matrix:

$$\mathbf{M}_b(X, \theta_0) = \mathbf{M}(X, \theta_0) + \mathbf{M}_0(\theta_0),$$

with $\mathbf{M}_0(\theta_0) = \frac{\partial^2}{\partial \theta^2} \log \pi_0(\theta_0)$ the precision matrix.

- Optimality function:

$$\psi : \mathbf{M} \in \mathcal{S}_p^+(\mathbf{R}) \longmapsto \psi(\mathbf{M}) \in \mathbf{R}$$

Examples: determinant, trace, maximum eigenvalue, minimum eigenvalue, etc.

Fisher information matrix's criteria for DOPE

1. Eψ-optimality criterion:

$$\mathbf{C}(X) = \mathbf{E}_{\theta_0}\left[\psi(\mathbf{M}_b(X, \theta_0))\right].$$

2. Robuste min-max-optimality criterion:

$$\mathbf{C}(X) = \min_{\theta_0 \in \Theta} \psi(\mathbf{M}_b(X, \theta_0)) \text{ ou } \max_{\theta_0 \in \Theta} \psi(\mathbf{M}_b(X, \theta_0)).$$

Examples:

- $\psi(\mathbf{M}) = \det(\mathbf{M})$, ED-optimality, to be maximized (Fedorov 1980; Pronzato and Walter 1985),
- $\psi(\mathbf{M}) = \mathrm{Tr}(\mathbf{M}^{-1})$ MMIT-optimality, to be maximized (Pronzato and Walter 1985).

Inconveniences

- Linear hypothesis: we use linear approximation of the code in the neighbourhood of θ_0.
- Posterior distribution not taken into account.

Kullback Leibler criterion for DOPE

Definition (Divergence de Kullback-Leibler)

Let p and q be two probability distributions on the same Ω space. The Kullback-Leibler divergence is defined as:

$$\mathbf{KL}(p \mid\mid q) = \int_{\Omega} \log \frac{p(x)}{q(x)} p(dx).$$

Information measurement of design X (Abellan and Noetinger 2010):

$$\mathbf{KL}\left[\pi(\theta \mid \mathbf{Y}^s(X, \theta_0)) \mid\mid \pi_0(\theta)\right]$$

where

- $\mathbf{Y}^s(X, \theta_0) = \mu^M(X, \theta_0) + \left[k^M\big((X, \theta_0), (X, \theta_0)\big)^{1/2}\right]^T \varepsilon_{GP} + \varepsilon_X$, where $\varepsilon_{GP} \sim \mathcal{N}(0, I_n)$ and $\varepsilon_X \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_n)$.

We can then define the EKL-optimality criteria:

$$\mathbf{C}_{\mathsf{KL}}(X) = \mathbf{E}_{\theta_0}\left[\mathbf{E}_{\mathbf{Y}^s}\left(\mathbf{KL}\left[\pi(\theta \mid \mathbf{Y}^s(X, \theta_0)) \mid\mid \pi_0(\theta)\right]\right)\right].$$

Posterior distribution's criteria for DOPE

- Simulation of physical observations:

$$\mathbf{Y}^s(X,\theta_0) = \mu^M(X,\theta_0) + \left[k^M\big((X,\theta_0),(X,\theta_0)\big)^{1/2}\right]^T \varepsilon_{GP} + \varepsilon_X$$

where $\varepsilon_{GP} \sim \mathcal{N}(0, I_n)$ and $\varepsilon_X \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_n)$.

- Assessment of design of physical experiments quality:

$$\psi\Big(\mathrm{Cov}\big(\theta \mid \mathbf{Y}^s(X,\theta_0)\big)\Big),$$
$$\mathbf{E}_\theta\Big[l(\theta,\theta_0) \mid \mathbf{Y}^s(X,\theta_0)\Big],$$

where ψ is the optimality function and l is a loss function.

Examples:

- $\psi(\mathbf{M}) = \mathrm{Tr}(\mathbf{M})$: sum of a posteriori variances,
- $l(\theta,\theta_0) = \frac{1}{n} \mid\mid \theta - \theta_0 \mid\mid_1$: Mean Absolute Error a posteriori,
- $l(\theta,\theta_0) = \frac{1}{n} \mid\mid \theta - \theta_0 \mid\mid_2^2$: Mean Square Error a posteriori.

Posterior distribution's criteria for DOPE

- Considering uncertainties:

$$\mathbf{C}_{\text{cov}}(X) = \mathbf{E}_{\theta_0}\left\{\mathbf{E}_{\mathbf{Y}^s}\left[\psi\left(\text{Cov}\left(\theta \mid \mathbf{Y}^s(X,\theta_0)\right)\right)\right]\right\}, \tag{3}$$

$$\mathbf{C}_{\text{loss}}(X) = \mathbf{E}_{\theta_0}\left\{\mathbf{E}_{\mathbf{Y}^s}\left[\mathbf{E}_{\theta}\left(l(\theta,\theta_0) \mid \mathbf{Y}^s(X,\theta_0)\right)\right]\right\}. \tag{4}$$

- Monte Carlo calculation costs: $O(N_{\theta_0} N_{\mathbf{Y}^s} N_{\theta})$.
- Fast calculation of criterion: $O(N_{\theta_0})$
  - Joint random sampling of $\theta_0, \varepsilon_{GP}$ and ε,
  - Unique uniform sampling of θ then weighting by the posterior distribution $\pi(\theta \mid \mathbf{Y}^s(X,\theta_0))$.
- Sequential versions of all these criteria are possible.

Optimization algorithm for optimal DOPE

- **Optimization problem:**

$$X^* = \arg\max_{X \in \mathcal{X}^n} \mathbf{C}(X)$$

We recall $\mathcal{X} \subset \mathbf{R}^d$.

- **Some challenges:**
  - The Design $X \in \mathcal{X}^n \subset \mathbf{R}^{d \times n}$ is a matrix:

$$X = \begin{pmatrix} x_1^{(1)} & \dots & x_d^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(n)} & \dots & x_d^{(n)} \end{pmatrix},$$

  - An evaluation of the criteria $\mathbf{C}(X)$ takes time,
  - No analytical expression for the gradient or the Hessian of $\mathbf{C}(X)$.

$\Longrightarrow$ A variant of the simulated annealing optimization algorithm.

Simulated Annealing for optimal DOPE

- Key idea: sub-optimal initial matrix + random perturbation per row.

1. **Initialization**:
   - Choose k_{max} the maximum number of iterations, T_0 the initial temperature, σ^2 the variance parameter and X_0 the initial matrix by **forward optimization algorithm**.
2. **While** $k \leq k_{max}$ **do**:
   - Random perturbation:

$$X_{prop} = X_k + \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ \varepsilon_1 & \dots & \varepsilon_d \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix},$$

where $(\varepsilon_1, \dots, \varepsilon_d) \sim \mathcal{N}(0, \sigma^2 I_d)$ such that $X_{prop} \in \mathcal{X}^n$

Simulated Annealing for optimal DOPE

2. Continuation of step 2:
   - Evaluating deterioration $\Delta_k = C(X_k) - C(X_{prop})$
   - Calculate $p = \min(e^{-\Delta_k / T_k}, 1)$
   - Draw $u \sim \mathcal{U}_{[0,1]}$
   - Accept-reject step:

$$X_{k+1} \longleftarrow \begin{cases} X_{prop} & \text{if } p \geq u \\ X_k & \text{else.} \end{cases}$$

   - Update the temperature $T_{k+1} = cT_k$, with $0 \leq c \leq 1$
   - Update $k \longleftarrow k+1$

   End While.

3. Return $X_{k_{max}}$.

Forward Optimisation Algorithm: FOA

The FOA proposed by ABTINI 2018 provide a near-optimal solution.

The steps of the algorithm are as follows:

1. **Initialization**: Take $X_0 = \{\}$ as the initial solution.
2. **For** k ranging from 1 to n **do**:
   - Determine
   $$x_k^* = \arg\max_{x \in \mathcal{X} \setminus X_k} \mathbf{C}(X_k \cup \{x\}).$$
   - Update $X_{k+1} = X_k \cup \{x_k^*\}$.

   End For.
3. Return X_n.

N.B: We compute $T_0 = -\frac{\Delta f}{\log P_0}$ or $T_0 = -\frac{\Delta f}{\log \frac{1-P_0}{P_0}}$ for initial temperature.

Design Of Numerical Experiments

Design Of Numerical Experiments (DONE)

- At this stage, physical data $(X_{\text{obs}}, Y_{\text{obs}})$ are available.

Goal

Choose the design of numerical experiments $D_M = \left[(x_1, \theta_1), \ldots, (x_M, \theta_M)\right]^T$.

Two approaches:

- Space filling design such as LHS-maximin (see Pronzato and Müller 2012),
- Sequential design based on minimization of some criterion.

Damblin et al. 2018 criterion for DONE

$$D_M = \underset{D \in (\mathcal{X} \times \Theta)^M}{\arg\min} \ \mathbf{KL}(\pi(\theta \mid Y_{\text{obs}}) \mid\mid \pi(\theta \mid Y_{\text{obs}}, f_{\text{code}}(D_M))).$$

Heuristic: KL minimization

- Choose

  $D_M = D_0 \cup D_{M-M_0}$ where $D_0 \in (\mathcal{X} \times \Theta)^{M_0}$ and $D_{M-M_0} \in (X_{\text{obs}} \times \Theta)^{M-M_0}$,

  M_0: the size of the initial design for building an initial metamodel.

- Resolve the following optimization problem

$$\min_{\theta \in \Theta} \quad \text{SS}(\theta) = \sum_{i=1}^{n} \left(y_{\text{obs}}^{(i)} - f_{\text{code}}(x^{(i)}, \theta) \right)^2.$$

$\Longrightarrow$ **Resolution of the optimization problem by the EGO algorithm based on the EI criterion.**

Results

Test function

The test function playing the role of computer code:

$$f_{\text{code}} : [-2, 6]^2 \times [2, 4] \times [0, 8] \to \mathbf{R}$$

$$(x_1, x_2, \theta_1, \theta_2) \longmapsto \left(\exp(-x_1^2 - x_2^2) + \frac{\sin(2\pi x_2)}{x_2} \right) \left(\theta_1^2 x_1 + \frac{\theta_1 \theta_2}{100} \right)$$

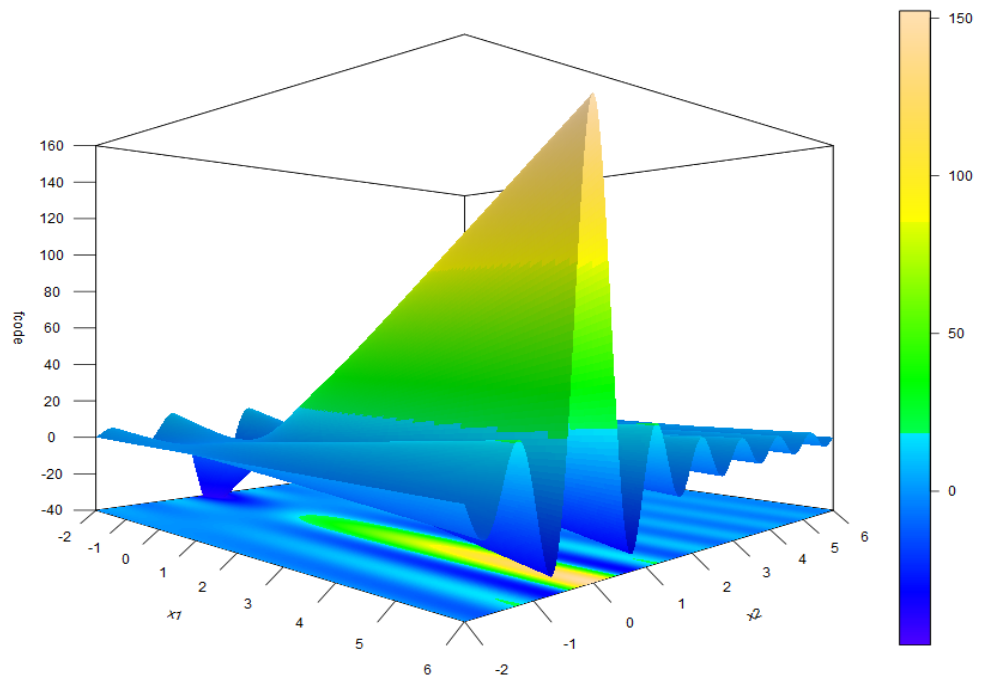


Figure 5: Test function at true value $\theta_0 = (2, 5)$.

Hyperparameters & experiments

We choose:

- The true value $\theta_0 = (2, 5)$
- The size of physical design of experiments $n = 10$
- The size of numerical design $M = 80$ with $M_0 = 60$
- The covariance function: Matern $5/2$
- The mean function is constant
- The measurement noise level $\sigma^2 = 5\% \, Var(f_{\text{code}})$

1. Build the metamodel using the computer code evaluations on design of numerical experiments of size M_0 (LHS-maximin + VAR criterion),
2. Choose the DOPE X^* by some $\mathcal{C}_+$: LHS-maximin,$\mathcal{C}_{det}$,$\mathcal{C}_{trace}$,$\mathcal{C}_{sov}$, $\mathcal{C}_{mse}$ or $\mathcal{C}_{kl}$,

Experiments

3. Repeat for $l = 1, \ldots, L$, with $L = 100$:
   - Simulate the physical observation by $\mathbf{Y}(X) = f_{\text{code}}(X, \theta_0) + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_n)$
   - Choose the DONE of size $M - M_0 = 20$ (OAT algorithm)
   - Compute the following comparison metrics:

$$\textbf{RMSE} = \frac{1}{L} \sum_{l=1}^{L} || \theta_0 - \hat{\theta}_{map}^{(l)} ||_2,$$

$$\textbf{ALCI} = \frac{1}{2L} \sum_{1=1}^{L} || \hat{\theta}_{sup}^{(l)} - \hat{\theta}_{inf}^{(l)} ||_1,$$

where $\hat{\theta}_{map}^{(l)} = \mathbf{E}\big[\theta \mid Y_{\text{obs}}^{(l)}(X^*), f_{\text{code}}(D_M)\big]$, $\hat{\theta}_{sup}^{(l)}$ and $\hat{\theta}_{inf}^{(l)}$ are respectively the upper and lower bounds of the CI of $\pi(\theta \mid Y_{\text{obs}}^{(l)}(X^*), f_{\text{code}}(D_M))$.

Results

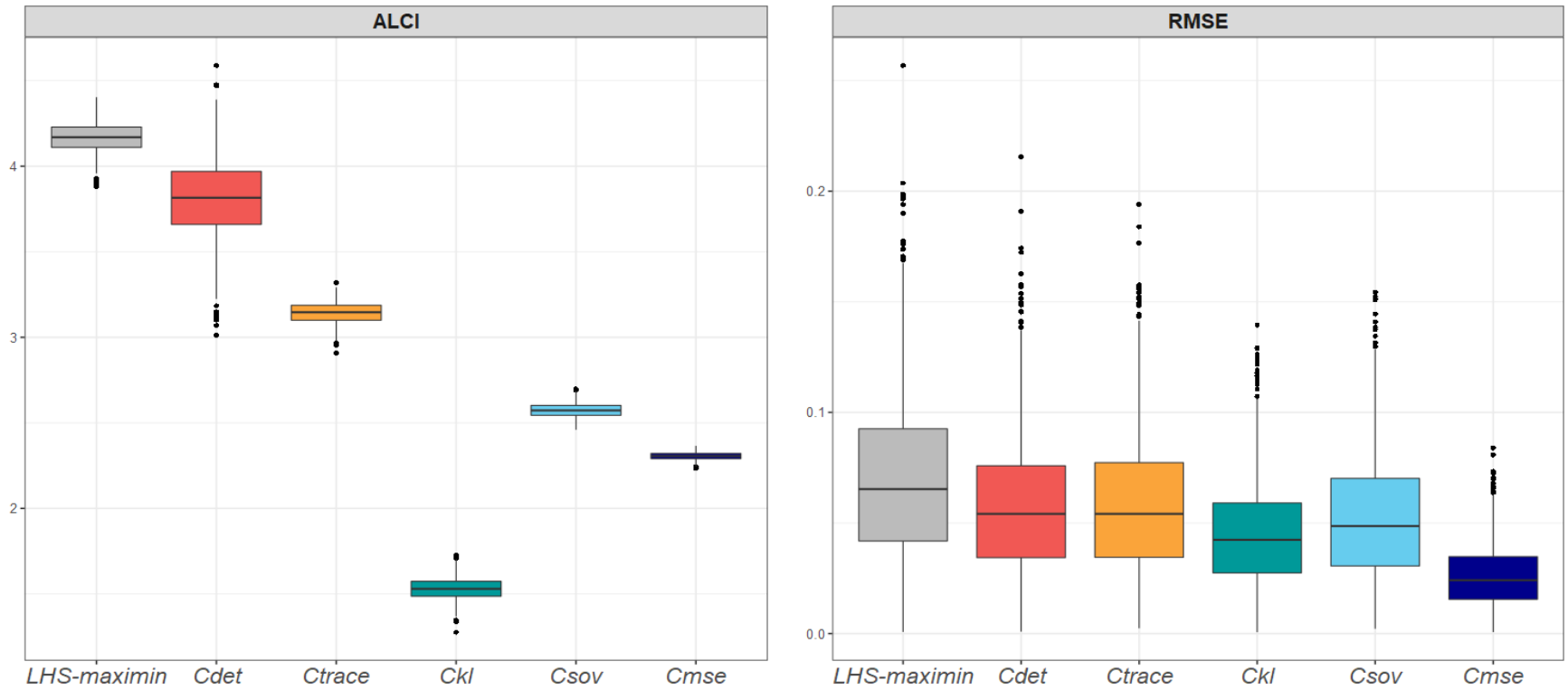


Figure 6: Boxplot of metrics.

$\Longrightarrow$ In this example, it makes sense to design the physical experiments.

Conclusions

Conclusions and perspectives

Conclusions

- Improvement of calibration with the proposed criteria for design of physical experiments.
- The advantages of global cirteria ($\mathcal{C}_{\mathbf{sov}}, \mathcal{C}_{\mathbf{mse}}$ and $\mathcal{C}_{\mathbf{kl}}$) over local criteria ($\mathcal{C}_{\mathbf{det}}$ and $\mathcal{C}_{\mathbf{trace}}$).

Perspectives

- Empirical comparison of criteria on different test functions.
- Apply the chosen strategy on simplified scalar output CO_2 application case.
- Extend the approach to functionnal output computer codes.
- Handle the case of biased computer codes and integrate sensitivity analysis.

Thanks for your attention !

Questions and suggestions are welcome !

References I

[1] A. Abellan and Benoit Noetinger. "Optimizing Subsurface Field Data Acquisition Using Information Theory". In: *Mathematical geosciences* 42 (Aug. 2010), pp. 603–630. DOI: 10.1007/s11004-010-9285-6.

[2] Mona ABTINI. "Plans prédictifs à taille fixe et séquentiels pour le krigeage". PhD thesis. 2018.

[3] Mathieu Carmassi et al. "Bayesian calibration of a numerical code for prediction". working paper or preprint. Jan. 2018. URL: https://hal.archives-ouvertes.fr/hal-01677167.

References II

[4] Guillaume Damblin et al. "Adaptive Numerical Designs for the Calibration of Computer Codes". In: *SIAM/ASA Journal on Uncertainty Quantification* 6.1 (2018), pp. 151–179. ISSN: 2166-2525. DOI: 10.1137/15m1033162. URL: http://dx.doi.org/10.1137/15M1033162.

[5] Valery V. Fedorov. "Convex design theory 1". In: *Statistics* 11 (1980), pp. 21–43.

[6] Marc Kennedy and Anthony O'Hagan. "Bayesian Calibration of Computer Models". In: *Journal of the Royal Statistical Society Series B* 63 (Feb. 2001), pp. 425–464. DOI: 10.1111/1467-9868.00294.

References III

[7] F. Liu, M. Bayarri, and J. Berger. "Modularization in Bayesian analysis, with emphasis on analysis of computer models". In: *Bayesian Analysis* 4 (Mar. 2009). DOI: 10.1214/09-BA404.

[8] Luc Pronzato. "Synthèse d'expériences robustes pour modèles à paramètres incertains". Thèse de doctorat dirigée par Walter, Éric Informatique Paris 11 1986. PhD thesis. 1986, 1 vol. (222 p.) URL: http://www.theses.fr/1986PA112260.

[9] Luc Pronzato and Werner G. Müller. "Design of computer experiments: space filling and beyond". In: *Statistics and Computing* 22 (2012), pp. 681–701.

References IV

[10] Luc Pronzato and Eric Walter. "Robust experiment design via stochastic approximation". In: *Mathematical Biosciences* 75.1 (1985), pp. 103–120. ISSN: 0025-5564. DOI: https://doi.org/10.1016/0025-5564(85)90068-9. URL: https://www.sciencedirect.com/science/article/pii/0025556485900689.

[11] Olivier Roustant, David Ginsbourger, and Yves Deville. "DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization". In: *Journal of Statistical Software* 51.1 (2012), pp. 1–55. URL: https://www.jstatsoft.org/v51/i01/.