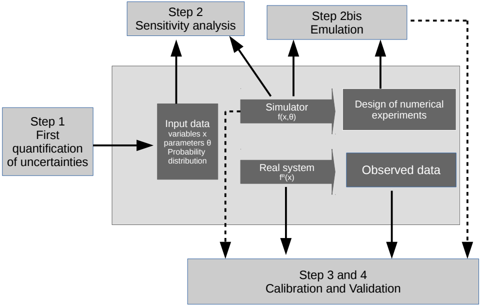
Calibration of computer models

Introduction

Pierre BARBILLON

Fall 2023, école ETICS

Uncertainty Quantification / Model Uncertainty



In this course, we will focus on calibration and validation.

Calibration of a computer code

Computer experiments:

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

- **physical parameters**: $\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^p$ observable and often controllable inputs
- **simulator parameters**: $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ 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.

Goal:
Calibrate the code: finding "best" or "true" $\boldsymbol{\theta}$ from real observations / field data (provided by physical experiments):

$$\mathbf{y}^e = \{y_1^e = \zeta(\mathbf{x}_1^e), \dots, y_n^e = \zeta(\mathbf{x}_n^e)\},$$

where ζ is the real physical phenomenon.

Validation

- Validation (rather than verification) is considered,
- Does the computer simulator correspond to field data?

$$\exists \boldsymbol{\theta}^*?, \text{ s.t., } \forall \mathbf{x}, \quad f(\mathbf{x}, \boldsymbol{\theta}^*) \approx y(\mathbf{x})$$

- This question is related with intended use of the simulator: range of $\mathbf{x}$, required precision...
- Biased computer model, no setting of calibrated parameters leads to outputs close to field data
  $\Rightarrow$ discrepancy.
- Do we want to validate the computer model itself or the computer model with the bias / discrepancy correction?

Framework

KOH framework chosen for this course:
[Kennedy and O'Hagan, 2001, Higdon et al., 2004, Bayarri et al., 2007].

History Matching not consider but may be relevant depending on the objective
[Craig et al., 1997, Vernon et al., 2010, Boukouvalas et al., 2014, Andrianakis et al., 2017]

Deterministic or stochastic simulator

In most contributions, f is considered as deterministic.

But some recent work considered f as stochastic ([Baker et al., 2022]):

- f uses stochastic approximations (MC,...) but the modeled phenomenon is deterministic ζ...
- f models a stochastic phenomenon $\zeta(\cdot)$ is stochastic.

Notations

- $\zeta(\mathbf{x})$ real physical or biological phenomenon,
- $f(\mathbf{x}, \boldsymbol{\theta})$ numerical code/model with $\mathbf{x}$ bservable or controllable input variable, $\boldsymbol{\theta}$ model parameter (no counterpart in the real phenomenon.
- DoNE: Design of Numerical Experiments: $D^c = \{(\mathbf{x}_1, \boldsymbol{\theta}_1), \ldots, (\mathbf{x}_N, \boldsymbol{\theta}_N)\}$ with corresponding evaluations of the computer model (time-consuming):

$$\mathbf{y}^c = f(D^c) = \{f(\mathbf{x}_1, \boldsymbol{\theta}_1), \ldots, f(\mathbf{x}_N, \boldsymbol{\theta}_N)\}.$$

- DoFE: Design of Field Experiments: $D^e = \{\mathbf{x}_1^e, \ldots, \mathbf{x}_{n_e}^e\}$ with corresponding noisy observation of ζ:

$$\mathbf{y}^e = \{y_1 = \zeta(\mathbf{x}_1^e) + \epsilon_1, \ldots, y_n = \zeta(\mathbf{x}_{n_e}^e) + \epsilon_{n_e}\}.$$

Outline

1. Calibration KOH and extensions
2. Sequential design of experiments
3. Focus on the discrepancy function and validation
4. Extensions to calibration of stochastic simulator

Andrianakis, I., McCreesh, N., Vernon, I., McKinley, T. J., Oakley, J. E., Nsubuga, R. N., Goldstein, M., and White, R. G. (2017).
Efficient history matching of a high dimensional individual-based HIV transmission model.
SIAM/ASA Journal on Uncertainty Quantification, 5(1):694–719.

Baker, E., Barbillon, P., Fadikar, A., Gramacy, R. B., Herbei, R., Higdon, D., Huang, J., Johnson, L. R., Ma, P., Mondal, A., et al. (2022).
Analyzing stochastic computer models: A review with opportunities.
Statistical Science, 37(1):64–89.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007).
A framework for validation of computer models.
Technometrics, 49(2):138–154.

Boukouvalas, A., Sykes, P., Cornford, D., and Maruri-Aguilar, H. (2014).
Bayesian precalibration of a large stochastic microsimulation model.
IEEE Transactions on Intelligent Transportation Systems, 15(3):1337–1347.

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1997).
Pressure matching for hydrocarbon reservoirs: a case study in the use of bayes linear strategies for large computer experiments.
In Case Studies in Bayesian Statistics, pages 37–93. Springer.

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004).
Combining field data and computer simulations for calibration and prediction.
SIAM Journal on Scientific Computing, 26(2):448–466.

Kennedy, M. C. and O'Hagan, A. (2001).
Bayesian calibration of computer models.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3):425–464.

Vernon, I., Goldstein, M., Bower, R. G., et al. (2010).
Galaxy formation: a Bayesian uncertainty analysis.
Bayesian Analysis, 5(4):619–669.