Parameter control in the presence of uncertainties

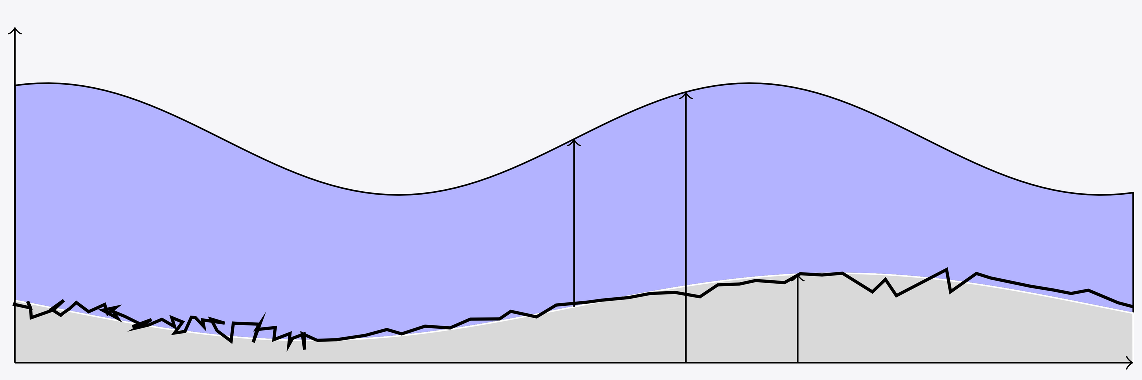
Victor Trappler, Élise Arnaud, Laurent Debreu, Arthur Vidard
victor.trappler@univ-grenoble-alpes.fr
AIRSEA Research team (Inria)– Laboratoire Jean Kuntzmann
Adjoint Workshop, Aveiro 2018

How can one calibrate a computer model so that it performs reasonably well for different random operating conditions ?

Objectives

- Define suitable definitions of robustness in the field of computer code calibration
- Develop efficient techniques and algorithms in order to estimate those parameters
- Deal with the high-dimension of the parameter spaces: Dimension reduction

Setting of the problem



X

The calibration problem is to be able to find a value of $\boldsymbol{k}$ denoted $\hat{\boldsymbol{k}}$ that matches the best the observations $\boldsymbol{y}^{\mathrm{obs}}$.

Inverse Problem

Environmental variables
$\boldsymbol{U} \in \mathcal{U}$ *random*

Control variable
$\boldsymbol{k} \in \mathcal{K}$

Direct Simulation

$M(\boldsymbol{k}, \boldsymbol{u})$

Inverse Problem

$\boldsymbol{y}^{\mathrm{obs}}$

$$J(\boldsymbol{k}) = \frac{1}{2}\|M(\boldsymbol{k}) - \boldsymbol{y}^{\mathrm{obs}}\|^2_{\Sigma^{-1}} \qquad \text{(Cost function)}$$

and we have to perform the following minimisation problem, usually with the help of the adjoint method

$$\hat{\boldsymbol{k}} = \operatorname*{arg\,min}_{\boldsymbol{k}\in\mathcal{K}} J(\boldsymbol{k})$$

Now, $\boldsymbol{u}$ sampled from $\boldsymbol{U}$ of density p_U and $\boldsymbol{y}^{\mathrm{obs}} = M(\boldsymbol{k}_{\mathrm{ref}}, \boldsymbol{u}_{\mathrm{ref}})$
The loss function is now

$$\underbrace{J(\boldsymbol{k}, \boldsymbol{U}) = \frac{1}{2}\|M(\boldsymbol{k}, \boldsymbol{U}) - \boldsymbol{y}^{\mathrm{obs}}\|^2_{\Sigma^{-1}}}_{\text{Random variable}}$$

- What criteria to use to "optimize" in a sense J ?
- How to deal with long computation ?

Which criterion to choose ? [1, 2]

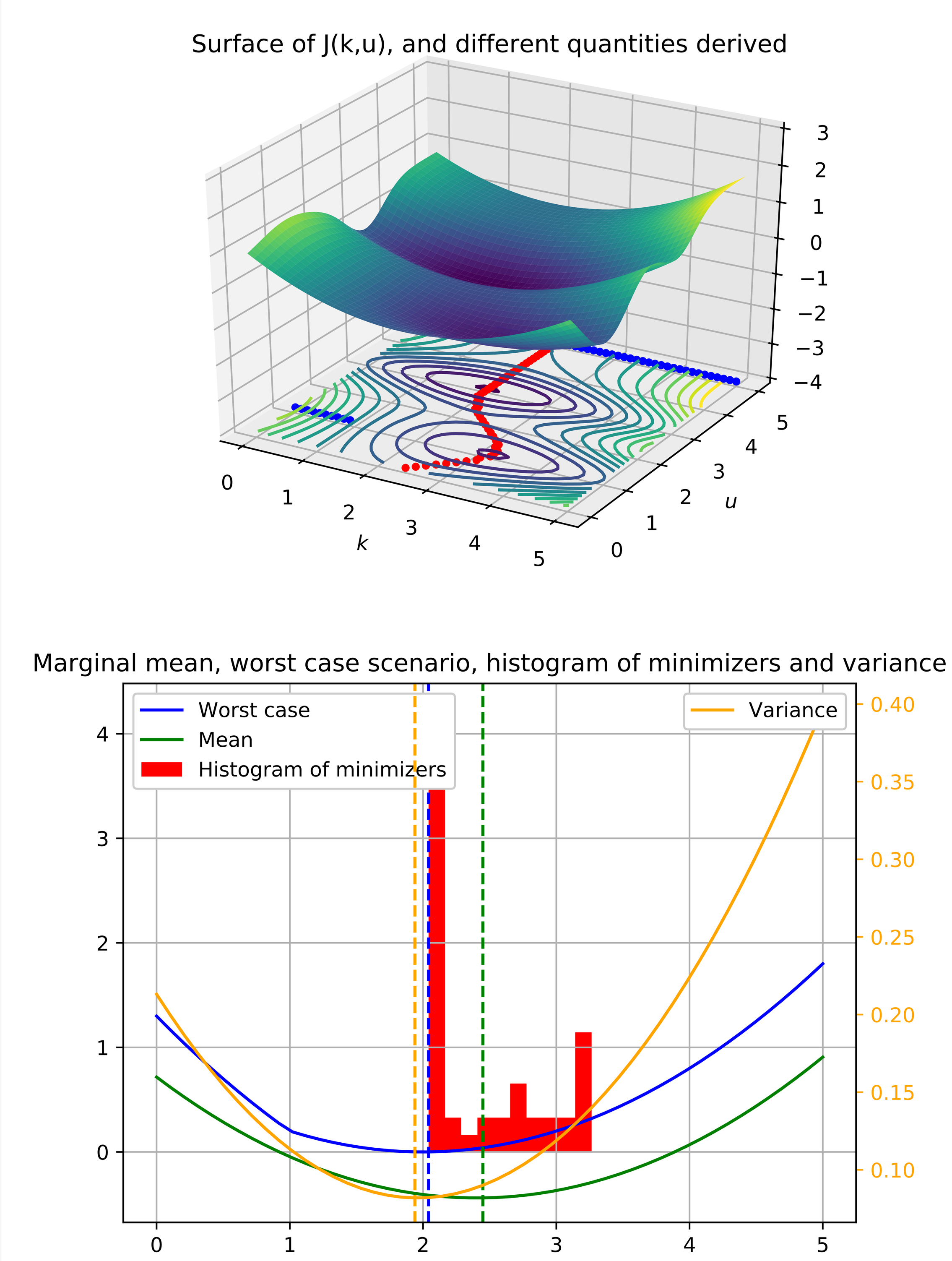
Different approaches

- Consider the worst-case scenario [3]
$$J_{\mathrm{w}}(\boldsymbol{k}) = \max_{\boldsymbol{u}\in\mathcal{U}} J(\boldsymbol{k}, \boldsymbol{u}) \quad \text{and} \quad \hat{\boldsymbol{k}}_{\mathrm{wc}} = \operatorname*{arg\,min}_{\boldsymbol{k}} J_{\mathrm{w}}(\boldsymbol{k})$$
- The solution gives good results on average:
$$\mu(\boldsymbol{k}) = \mathbb{E}_U[J(\boldsymbol{k}, \boldsymbol{U})] \quad \text{and} \quad \hat{\boldsymbol{k}}_{\mu} = \operatorname*{arg\,min}_{\boldsymbol{k}} \mu(\boldsymbol{k})$$
- The estimate gives steady results:
$$\sigma^2(\boldsymbol{k}) = \mathbb{V}\mathrm{ar}_U[J(\boldsymbol{k}, \boldsymbol{U})] \quad \text{and} \quad \hat{\boldsymbol{k}}_{\sigma^2} = \operatorname*{arg\,min}_{\boldsymbol{k}} \sigma^2(\boldsymbol{k})$$
- Compromise between Mean and Variance: multiobjective optimization problem:
$$\text{Pareto front of } (\mu(\boldsymbol{k}), \sigma^2(\boldsymbol{k}))$$
- Reliability analysis: Probability of being below threshold T
$$R_T(\boldsymbol{k}) = \mathbb{P}\left[J(\boldsymbol{k}, \boldsymbol{U}) \leq T\right], \quad \hat{\boldsymbol{k}}_{R_T} = \operatorname*{arg\,max} R_T(\boldsymbol{k})$$
- Special case: $T_{\min} = T(\boldsymbol{U}) = \min_{\boldsymbol{k}} J(\boldsymbol{k}, \boldsymbol{U})$
$$\mathbb{P}\left[J(\boldsymbol{k}, \boldsymbol{U}) \leq T_{\min}\right] = \mathbb{P}\left[\boldsymbol{k} = \operatorname*{arg\,min}_{\tilde{\boldsymbol{k}}} J(\tilde{\boldsymbol{k}}, \boldsymbol{U})\right]$$

Bayesian approach

Include beliefs upon $\boldsymbol{K}$ and $\boldsymbol{U}$ through priors: Bayes' theorem

$$p_{K,U|Y}(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}}) \propto p_U(\boldsymbol{u})p_K(\boldsymbol{k})\overbrace{p_{Y|K,U}(\boldsymbol{y}|\boldsymbol{k}, \boldsymbol{u})}^{\exp(-J(\boldsymbol{k},\boldsymbol{u}))}$$
$$\propto p_U(\boldsymbol{u})\underbrace{p_{K|Y,U}(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u})}_{=f(\boldsymbol{k},\boldsymbol{u})}$$



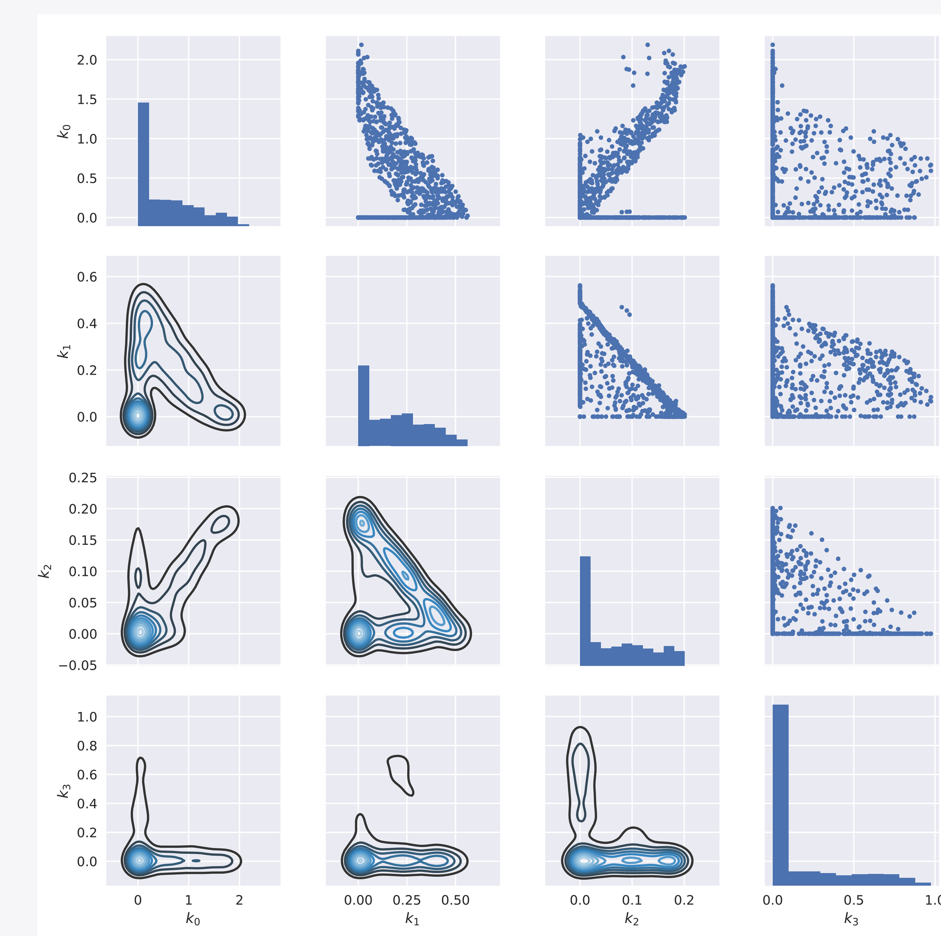
Methods

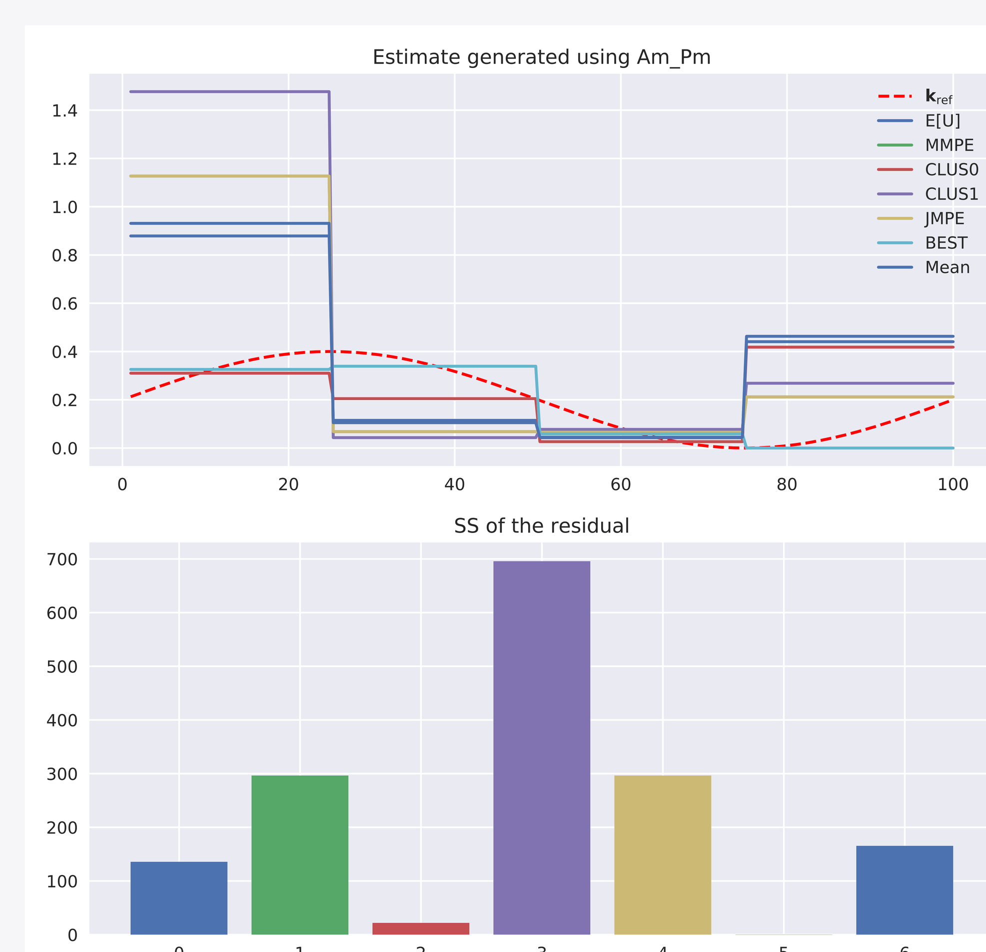
- MCMC based methods to sample from the posterior distribution $p_{K,U|Y}(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}})$, and/or marginalize
  - State Augmentation for Marginal Estimation (SAME) [4] to get $\hat{\boldsymbol{k}}_{\mathrm{MMAP}}$
  - Hamiltonian/Langevin Monte Carlo: Improve convergence of MCMC via the information brought by the gradient of the posterior.
- Estimation of $\boldsymbol{K}_{\arg\max}$
  - Need for efficient optimization (importance of the gradient)
  - Kernel Density Estimation
  - Clustering/ Mode-seeking algorithms
- Metamodelling
  - Build surrogate model cheap to evaluate
  - Lead to adaptative sampling strategies

Avoid MCMC by studying $\boldsymbol{K}_{\arg\max}$

- Sample $\boldsymbol{u}^{(i)}$ from $\boldsymbol{U}$ of density p_U.
- Using adjoint method,
$$\boldsymbol{k}^{(i)}_{\arg\max} = \operatorname*{arg\,max} p_{K|Y,U}(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u}^{(i)})$$
- Once the set of samples $(\boldsymbol{k}^{(i)}_{\arg\max})$ is sufficient
  - Either KDE and perform a direct optimization on the estimate
  - Either perform Clustering analysis

Results





Future Work/Perspectives

- Use of Surrogate Models ?
- Dimension reduction of $\mathcal{K}$, $\mathcal{U}$

References

[1] Jeffrey S. Lehman, Thomas J. Santner, and William I. Notz. Designing computer experiments to determine robust control variables. *Statistica Sinica*, pages 571–590, 2004.

[2] Janis Janusevskis and Rodolphe Le Riche. Simultaneous kriging-based sampling for optimization and uncertainty propagation. Technical report, July 2010.

[3] Abraham Wald. Statistical Decision Functions Which Minimize the Maximum Risk. *Annals of Mathematics*, 46(2):265–280, 1945.

[4] Arnaud Doucet, Simon J. Godsill, and Christian P. Robert. Marginal maximum a posteriori estimation using Markov chain Monte Carlo. *Statistics and Computing*, 12(1):77–84, January 2002.