Bayesian calibration and validation of a numerical model: an overview

Merlin Keller[1],Pierre Barbillon[2], Mathieu Carmassi[1], Matthieu Chiodetti[1], Guillaume Damblin[3], Cédric Goeury[1], Kaniav Kamary[4], Éric Parent[2]

[1]EDF R&D, France
[2]Agroparistech, France
[3]CEA, France
[4]Centrale Supélec, France

UQSAY Seminar, Centrale-Supélec

Industrial Context

Computer experiments are widely used in industrial studies :

- to complement / replace costly physical experiments
- in many applications : conception, reliability, risk assessment, long-term strategy planning . . .

Main concern : How reliable are the results of a numerical simulation ?

- Many uncertainty sources to account for : numerical, parametric, modelisation, extrapolation, . . .
- Need to quantify these sources and their influence on the results of the study

Adressing these issueds is the primary goal of the verification, validation and uncertainty quantification (VVUQ) framework [NRC, 2012]

Verification, Validation, Calibration

Definition of Validation [AIAA, 1998]

"[. . .] the process of determining the degree to which a model or simulation and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model."

- **Calibration** seeks to reduce parametric uncertainties, preliminary to validation ; both rely on the comparison of simulation results to physical measures
- **Verification** assesses that the numerical resolution of the equations defining the simulation model is accurate enough, prior to calibration and validation ([Roy and Oberkampf, 2010], outside the scope of this talk)

General Notations

- $r(x)$ is the **physical process of interest**, seen as an unkown function, where
- $x \in \mathbb{R}^d$ are the controlled (observed) inputs
- if not stated otherwise, $r(x)$ is scalar, and deterministic

- $y_c(x, \theta)$ is the **numerical model** (or computer code), seen as a mathematical 'black-box' function emulating r, where
- $\theta \in \mathbb{R}^p$ is the set of code parameters (uncertain physical constants defining the model)

- **Physical measures** of the physical process are available :

$$y_{exp,i} \quad = \quad r(x_{exp,i}) + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

Usual assumptions

Most often, code is assumed without bias [Cox et al., 2001], meaning that :

$$r(x) = y_c(x, \theta),$$

leading to the 'usual' calibration model :

Model 1

$$\mathcal{M}_1 : y_{exp,i} = y_c(x_i, \theta) + \varepsilon_i,$$

whose log-likelihood is obtained as :

$$\ell_1(\mathbf{y}_{exp}) \quad = \quad cte - \frac{n}{2} \log \sigma^2 - \frac{||\mathbf{y}_{exp} - \mathbf{y}_c(\mathbf{x}, \theta)||^2}{2\sigma^2}.$$

Consequence : Maximum likelihood estimation (MLE) boils down to ordinary least-squares (OLS). Assuming $y_c(x_{exp,i}, \theta) = h(x_{exp,i})^\top \theta$, the OLS estimator becomes explicit :

$$\widehat{\theta}_{OLS} \quad = \quad (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{y}_{exp}$$

with $\mathbf{H} = (h(x_{exp,1}), \ldots, h(x_{exp,n}))^\top$.

Bayesian calibration in the linear case

Still assuming $y_c(x_{exp,i},\theta) = h(x_{exp,i})^\top \theta$, we may define the Normal Inverse-Gamma conjugate prior for (θ, σ^2) with hyperparameters $(\theta_0, Q_0, a_0, b_0)$:

$$\begin{aligned} \theta|\sigma^2 &\sim \mathcal{N}(\theta_0, \sigma^2 Q_0^{-1}) \\ \sigma^2 &\sim \mathcal{IG}(a_0, b_0). \end{aligned}$$

Then, the posterior density of (θ, σ^2) is still Normal Inverse-Gamma, with updated hyperparameters $(\widehat{\theta}_n, Q_n, a_n, b_n)$:

$$\begin{aligned} \widehat{\theta}_n &= (Q_0 + \mathbf{H}^\top \mathbf{H})^{-1}(Q_0\theta_0 + \mathbf{H}^\top \mathbf{y}_{exp}) \\ Q_n &= (Q_0 + \mathbf{H}^\top \mathbf{H})^{-1} \\ a_n &= a_0 + \frac{n}{2} \\ b_n &= b_0 + \frac{1}{2}\left(y_{exp}^\top y_{exp} + \theta_0^\top Q_0 \theta_0 - \theta_n^\top Q_n \theta_n\right) \end{aligned}$$

- OLS coincides with the ‘non-informative’ limiting case $Q_0 = a_0 = b_0 = 0$
- Apart from this simple linear case, code calibration is never explicit !

Bayesian calibration in the non-linear case

Two main solutions :

Laplace approximation

Asymptotically, $\pi(\theta,\sigma^2|y_{exp})$ is well approached by $\mathcal{N}(\widehat{\theta}_{MAP},\widehat{\Sigma})$, with :

- $\widehat{\theta}_{MAP} = \arg\min_\theta \left(\ell_1(\mathbf{y}_{exp}) + \log\pi(\theta,\sigma^2)\right)$ (maximum *a posteriori*)
- $\widehat{\Sigma} = -\left(\partial^2_\theta \left(\ell_1(\mathbf{y}_{exp}) + \log\pi(\theta,\sigma^2)\right)|_{\theta=\widehat{\theta}_{MAP}}\right)^{-1}$ (Hessian matrix inverse)

MCMC

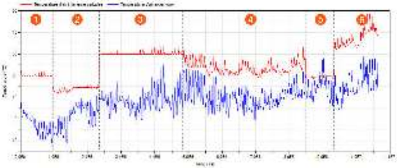
Letting $\eta_t := (\theta_t,\sigma_t^2)$, Generate $(\eta_t)_{t\geq 1}$ with a Markov kernel $Q(\eta_t|\eta_{t-1})$ targeting $\pi(\eta|y_{exp})$, for instance by Metropolis-Hastings :

- Generate proposal $\eta^\star \sim q(\eta^\star|\eta_{t-1})$
- Compute acceptance probability $\alpha = \min\left\{1; \frac{q(\eta_{t-1}|\eta^\star)}{q(\eta^\star|\eta_{t-1})}\frac{\pi(\eta^\star|y_{exp})}{\pi(\eta_{t-1}|y_{exp})}\right\}$

Comments

- Laplace approximation works often well, and should be tested first
- MCMC more generic and numerically stable

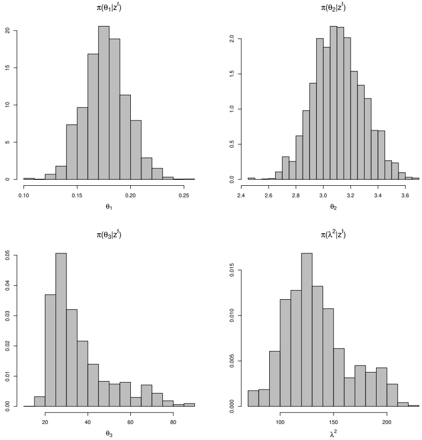
A simple illustration : BESTLAB [Damblin, 2015]



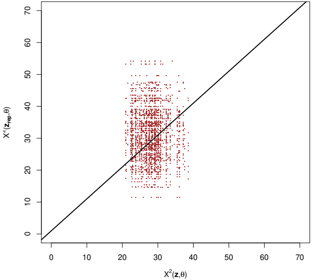
Calibration of a Dymola energy consumption model $y_c(x, \theta)$ for a single cell equiped with an AC system at EDF's BESTLAB test facility, based on experimental measures :

- y_c computes the power needed to maintain a target temperature in the celle
- x contains meteorological variables (outside temperature, irradiance, humidity, ...) measurements (assumed independent per time step)
- θ reduced to three parameters : albedo, thermal bridge, convection factor

BESTLAB calibration results



Posterior density of model parameters, for uniform priors



Predictive vs. realized χ^2 discrepancies :

$$\chi^2(\mathbf{y}_{exp}, \theta, \sigma^2) = \frac{||\mathbf{y}_{\exp} - y_c(\mathbf{x}_{exp}, \theta)||^2}{\sigma^2}.$$

Posterior p-value for goodness of fit is the proportion of points above the black line (here 60%).

Costly computer code

Assume $y_c(x,\theta)$ is costly to evaluate, and replaced by a Gaussian process emulator $Y_c(x,\theta)$ [Sacks et al., 1989], based on a numerical design $\mathcal{D} = (x_j, \theta_j)_{1 \le j \le N}$:

$$Y_c(x,\theta) \sim \mathcal{GP}(m_{\mathcal{D}}(\cdot,\cdot); c_{\mathcal{D}}(\cdot,\cdot;\cdot,\cdot)).$$

Then, we obtain a second calibration model :

Model 2

$$\mathcal{M}_2 : y_{exp,i} = Y_c(x_i,\theta) + \varepsilon_i.$$

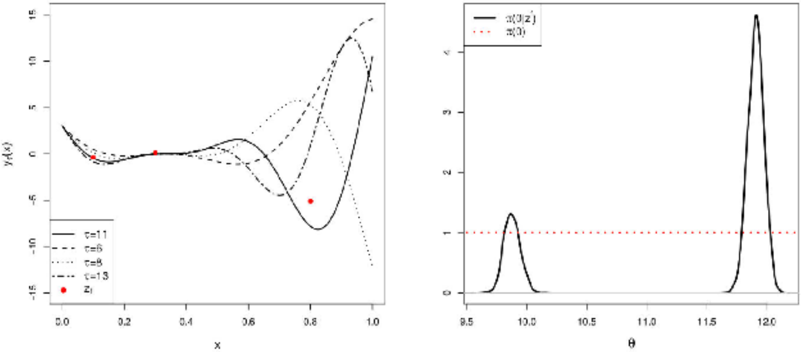
Integrating $Y_c(x,\theta)$ out yields the surrogate likelihood :

$$\begin{aligned}\ell_2(\mathbf{y}_{exp}|\theta,\sigma^2,y_c(\mathcal{D})) &= cte - \frac{1}{2}\log|\sigma^2\mathbf{I}_n + \mathbf{c}_{\mathcal{D}}(\theta)| \\ &\quad -\frac{1}{2}(\mathbf{y}_{exp} - \mathbf{m}_{\mathcal{D}}(\theta))^\top(\sigma^2\mathbf{I}_n + \mathbf{c}_{\mathcal{D}}(\theta))^{-1}(\mathbf{y}_{exp} - \mathbf{m}_{\mathcal{D}}(\theta)),\end{aligned}$$

and the corresponding surrogate posterior :

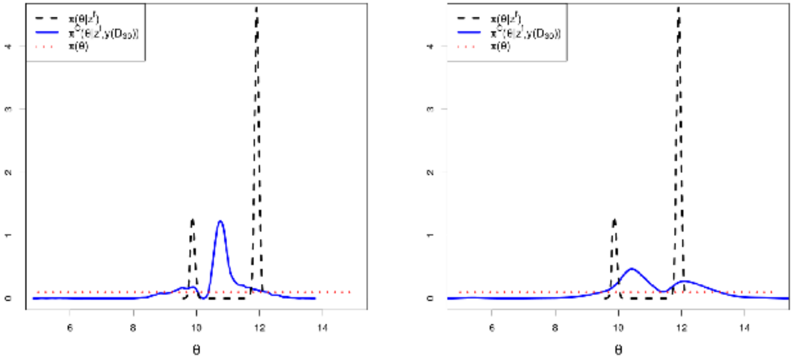
$$\pi(\theta,\sigma^2|y_{exp},y_c(\mathcal{D})) \propto \ell_2(\mathbf{y}_{exp}|\theta,\sigma^2,y_c(\mathcal{D}))\ \pi(\theta,\sigma^2)$$

Toy example



Left : $y_c(x, \theta) = (6x - 2)^2 \times \sin(\theta x - 4)$ on $[0, 1]$ for several values of $\theta \in [5, 15]$. Red dots are the field measurements. Right : exact posterior distribution

Toy example : LHS



Sampling of surrogate posterior $\pi(\theta, \sigma^2 | y_{exp}, y_c(\mathcal{D}))$ from two different maximin LHD of size $N = 30$ (using the R library MCMCpack).

Using results from approximation theory, we can show under certain conditions that [Damblin et al., 2017] : $\lim_{N \to \infty} KL\left(\pi(\theta|y_{exp}) || \pi(\theta|y_{exp}, y_c(\mathcal{D}))\right) \to 0$
Question : How to choose $\mathcal{D}$ in order to minimize KL for $N < \infty$?

Sequential strategies [Damblin et al., 2017]

Assuming k simulations $y_c(\mathbf{x}_{exp},\theta_j))_{1\leq j\leq k}$ have already been run, let $Y_k(x,\theta)$ denote the current GP emulator, and find

$$\begin{aligned}\theta_{k+1} &= \arg\max_{\theta} EI_k(\theta) \\ &= \mathbb{E}\left[(m_k - SS_k(\theta))_+\right],\end{aligned}$$

where :

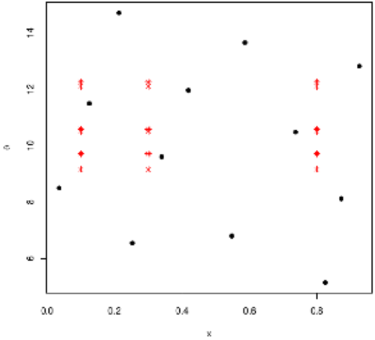
- $m_k = \min(SS(\theta_1),\ldots,SS(\theta_k))$ with SS : sum of squares for the actual code
- $SS_k(\theta)$ sum of squares replacing $y_c(\mathbf{x}_{exp},\theta)$ by $Y_k(\mathbf{x}_{exp},\theta)$

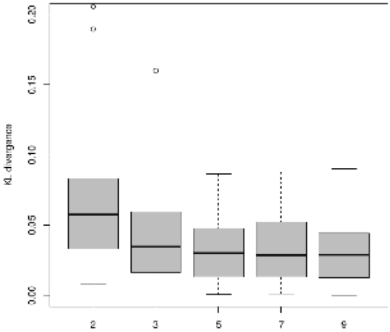
Comments :

- For each new value θ_k, need to compute $y_c(x_{exp,i},\theta_k)_{1\leq i\leq n}$ (Algorithm 1)
- Alternatively, add a single point $x_{exp,i}$ that maximizes $Crit_k(x_{exp,i})$ (Algorithm 2), for instance :

$$Crit_k(x_{exp,i}) = \mathbb{V}\left[Y_k(x_{exp,i},\theta_{k+1})\right]$$

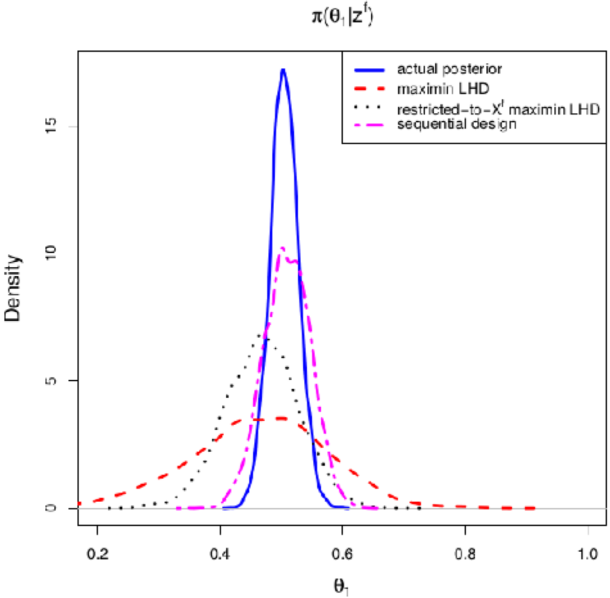
Results on toy example





Left : Algorithm 1 starting form a maximin LHD.
Right : Comparative KL divergences for a restricted to $\mathbf{x}_{exp}$ maximin LHD (2) and different sequential strategies

Results in 6 dimensions (g-Sobol function) for $N = 150$



Biased code

In [Kennedy and O'Hagan, 2001], an additional **code bias** (or **discrepancy**) term is considered : $\delta(x) = r(x) - y_c(x, \theta)$

- Main assumption :

$$\delta(x) \sim \mathcal{GP}(m_\delta(x); c_\delta(x, x'))$$

- **two possible interpretations :** latent variable distribution / prior on functional parameter

This leads to two additional calibration models [Carmassi, 2018] :

Model 3

$$\mathcal{M}_3 : y_{exp,i} = y_c(x_i, \theta) + \delta(x) + \varepsilon_i.$$

Model 4

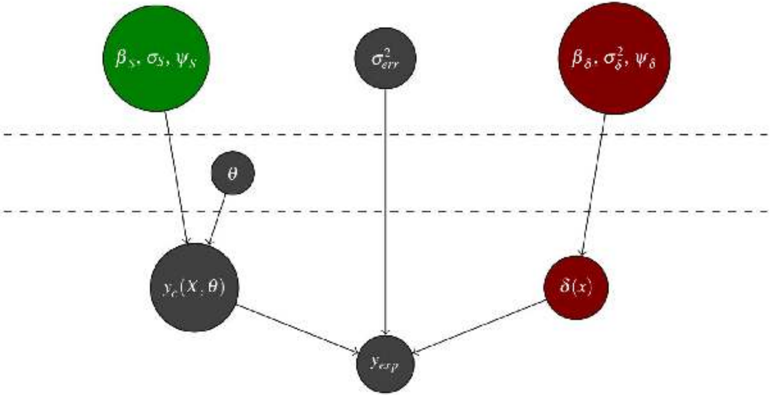
$$\mathcal{M}_4 : y_{exp,i} = Y_c(x_i, \theta) + \delta(x) + \varepsilon_i.$$

Comments

- If $y_c(x, \theta)$ is linear in θ, then Model 3 boils down to Universal Kriging (UK)

$\hookrightarrow$ calibration is explicit conditional on $(m_\delta(\cdot); c_\delta(\cdot, \cdot))$

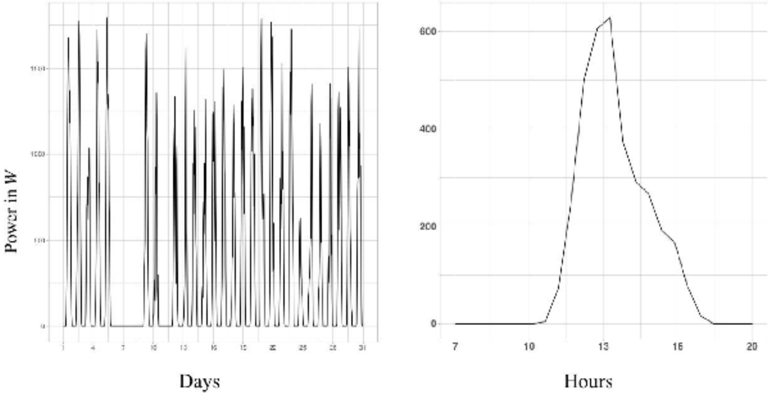
DAG of the four calibration models [Carmassi, 2018]



Comments

- Bayesian estimation is possible for each model, either using Laplace approximation or MCMC algorithms
- modularization used for GP on $Y_c(x, \theta)$ [Liu et al., 2009]

Case-study : PV power plant

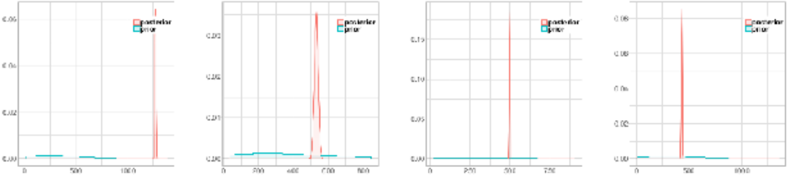


Main characteristics

- $y_c(x, \theta)$ represents the instantaneous power produced by EDF R& D 'PVZen' experimental power plant with :
- x weather covariates (irradiance, temperature), θ plant performance parameters

Comparison of the four calibration models

Calculations done with the CaliCo R package [Carmassi, 2018]



From left to right : Measurement errors σ^2 posterior densities in Models 1, 2, 3 and 4

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
coverage rate at 90% (in %)	91	32	87	23
RMSE of the instantaneous power (W)	12.69	21.61	5.91	18.7

Predictive performances of the four models computed by cross validation
$\hookrightarrow$ probable model discrepancy detected

Extensions

Recent methodological advances

- Sensitivity Analysis [Spagnol et al., 2018]
- Probabilistic Inversion [Fu, 2012], [de Crécy, 1997]
- Functional Code Calibration [Brown and Atamturktur, 2016]
- Orthogonal Gaussian processes [Plumlee, 2016]
- High-Dimensional Parameters [Carmassi, 2018]

Other Application Domains

- Metrology : Bayesian Calibration and Prediction of a Measure Process
- Probabilistic Seismic Hazard Analysis (PSHA) : Bayesian Model Selection (BMA) of Source Models
- Photovoltaïc Plant Long-Term Producible Forecast for Project Design and Contract Bidding
- Wind Turbine Plant : Long-Term Producible Forecast for Contract Bidding
- Historical Seismicity : Reconstructing Events from Past Testimonies
- Hydaulic modeling of flood risk with TELEMAC 2D
- ...

Short Review

Popular approaches

- Validation Metrics [Oberkampf and Barone, 2006] : Exploratory ways to compute the Calculation / Measure Gap
- A framework for validation of computer models [Bayarri et al., 2007] : Builds upon [Kennedy and O'Hagan, 2001] to obtain prediction bounds within Model 4 (biased, costly)

To be or not to be. . . biased ?

- **Better predictive performances** when accounting for code bias [Bayarri et al., 2007]. . .
- **Introduces non-identifiability** : code and bias terms cannot be discriminated without adding prior constraints (regularizing effect of Bayes inference)
- **Comes at a cost** : two correlated Gaussian processes need to be inferred instead of one

Bayesian Model Selection [Damblin et al., 2016]

Consider the task of selecting between two mutually exclusive assumptions :
$\mathcal{H}_0 : \delta(x) \equiv 0$ vs. $\mathcal{H}_1 : \exists x, \delta(x) \neq 0$ given y_{exp}
Principle : consider unknown assumption $\mathcal{H}$ as a parameter to be estimated, given a prior distribution on $\{\mathcal{H}_j\}_{0 \leq j \leq 1}$.
Applying Bayes' theorem :

$$\mathbb{P}\left[\mathcal{H} = \mathcal{H}_j | y_{exp}\right] = \frac{p(y_{exp}|\mathcal{H}_j)\mathbb{P}\left[\mathcal{H} = \mathcal{H}_j\right]}{\sum_{j'} p(y_{exp}|\mathcal{H}_{j'})\mathbb{P}\left[\mathcal{H} = \mathcal{H}_j\right]} \quad (1)$$

with $p(y|\mathcal{H}_j)$ the *marginal likelihood*, or *evidence*, for model $\mathcal{M}_j$:

$$p(y_{exp}|\mathcal{H}_j) = \int p(y_{exp}|\theta, \mathcal{H}_j)\pi(\theta)d\theta \quad (2)$$

Bayesian Model Selection, cont'd

Main steps :

1. Select most probable model, ie, maximize $p(y_{exp}|\mathcal{H}_j)\mathbb{P}\left[\mathcal{H}=\mathcal{H}_j\right]$,
2. Compute selected model's posterior distribution :

$$\pi(\theta|y_{exp},\mathcal{H}_j) \quad \propto \quad p(y_{exp}|\theta,\mathcal{H}_j)\pi(\theta) \qquad (3)$$

Typically sampled using Monte-Carlo Markov chain (MCMC) techniques [Robert and G., 2004].

Drawbacks :

- No generic way to evaluate the marginal likelihood $p(y_{exp}|\mathcal{H}_j)$ (see [Kass and Raftery, 1995] for instance)
- What if models are equally probable ?

Bayesian Model Averaging (BMA) [Hoeting et al., 1999]

Principle : integrate out $\mathcal{H}$ from joint posterior distribution of $(\theta, \mathcal{H})$ yielding the 'model-averaged' posterior for θ :

$$\begin{aligned} \pi(\theta|y_{exp}) &= \int \pi(\theta|y_{exp}, \mathcal{H})\pi(\mathcal{H}|y_{exp})d\mathcal{H}d\mathcal{H} \\ &\propto \sum_j P\left[\mathcal{H} = \mathcal{H}_j\right] p_j(\theta|y_{exp},)\pi_j(\theta|y), \end{aligned} \tag{4}$$

Main steps :

1. Compute posterior probabilities $p(\mathcal{H}_j|y_{exp}) \propto p(y_{exp}|\mathcal{H}_j)P\left[\mathcal{H} = \mathcal{H}_j\right]$ for all models
2. Remove all models with negligible posterior probabilities ;
3. Compute posterior density for all remaining models,
4. Deduce model-averaged posterior (4)

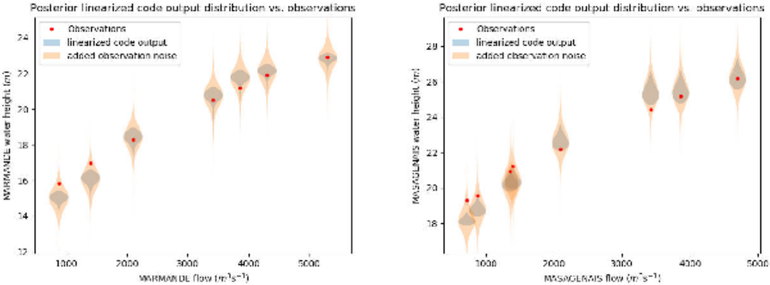
Drawbacks : Still depends on the hard to obtain marginal likelihood $p(y_{exp}|\mathcal{H}_j)$

Mixture Modeling [Kamary et al., 2014a]

Principle : Embed competing models $\mathcal{H}_j$ in an encompassing mixture model $\mathcal{H}_\alpha$ [Kamary et al., 2014b] :

$$y_{exp,i}|\alpha,\theta,\mathcal{H}_\alpha \overset{iid}{\sim} \sum_j \alpha_j p(y_{exp,i}|\theta,\mathcal{H}_j), \quad (5)$$

BMA coresponds to special case of a single observation : the dataset



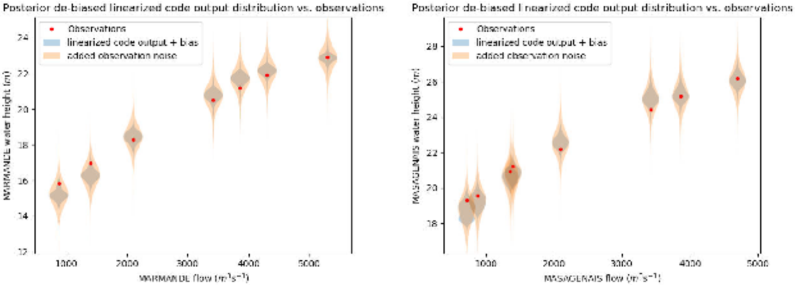
Application to TELEMAC 2D case-study : Bayesian bias detection

Mixture Modeling [Kamary et al., 2014a]

Principle : Embed competing models $\mathcal{H}_j$ in an encompassing mixture model $\mathcal{H}_\alpha$ [Kamary et al., 2014b] :

$$y_{exp,i}|\alpha, \theta, \mathcal{H}_\alpha \overset{iid}{\sim} \sum_j \alpha_j p(y_{exp,i}|\theta, \mathcal{H}_j), \tag{5}$$

BMA coresponds to special case of a single observation : the dataset



Application to TELEMAC 2D case-study : Bayesian bias detection

Mixture Modeling [Keller and Kamary, 2017]

Now we formalize the BMA / mixture modeling equivalence. Observe that the BMA posterior (4) can be re-written as :

$$\begin{aligned}
\pi(\theta|y_{exp}) &\propto \sum_j P\left[\mathcal{H} = \mathcal{H}_j\right] p(y_{exp}|\mathcal{H}_j)\pi(\theta|y_{exp}, \mathcal{H}_j) \\
&\propto \sum_j \mathbb{P}\left[\mathcal{H} = \mathcal{H}_j\right] p(y_{exp}|\mathcal{H}_j)\frac{p(y_{exp}|\theta)\pi(\theta)}{p(y_{exp}|\mathcal{H}_j)} \\
&\propto \left\{\sum_j P\left[\mathcal{H} = \mathcal{H}_j\right] p(y|\theta, \mathcal{H})\right\}\pi(\theta). \qquad (6)
\end{aligned}$$

Theorem (BMA as mixture modeling)

Bayesian model averaging is equivalent to the mixture of all prior-weighted candidate models, considering the complete dataset as a single observation.

Consequence

- (6) explicit, depence on $p(y_{exp}|\mathcal{H}_j)$ removed

$\hookrightarrow$ can be used to sample the model-averaged posterior by MCMC

Generic algorithm for Bayesian model averaging and selection

Main steps :

1. Generate posterior sample $(\theta_{1:S})$ from BMA mixture model (6) using, eg, MCMC
2. Estimate posterior probability of $\mathcal{H}_j$ by :

$$\widehat{\pi}(\mathcal{H}_j|y_{exp}) = \frac{1}{S}\sum_{s=1}^{S} w_j(\theta_s). \quad (7)$$

with $w_j(\theta_s) := \frac{p(y_{exp}|\theta_s,\mathcal{H}_j)\mathbb{P}[\mathcal{H}=\mathcal{H}_j]}{\sum_{j'} p(y_{exp}|\theta_s,\mathcal{H}_{j'})\mathbb{P}[\mathcal{H}=\mathcal{H}_{j'}]}$

3. Posterior expectations in model j can be obtained by importance sampling, using the weighted sample $(\theta_s, w_j(\theta_s))_{1\leq s\leq S}$.

Equation (7) is justified by the fact that :

$$\begin{aligned}\pi(\mathcal{H}_j|y_{exp}) &= \int \pi(\mathcal{H}_j|\theta, y_{exp})\pi(\theta|y_{exp})d\theta \\ &= \int \frac{p(y_{exp}|\theta,\mathcal{H}_j)\mathbb{P}[\mathcal{H}=\mathcal{H}_j]}{\sum_{j'} p(y_{exp}|\theta,\mathcal{H}_{j'})\mathbb{P}[\mathcal{H}=\mathcal{H}_{j'}]}\pi(\theta|y)d\theta.\end{aligned}$$

Numerical performance

Quality of importance sampling estimate of posterior distribution $\pi(\theta|y_{exp}, \mathcal{H}_j)$ measured by the effective sample size (ESS) [Arnaud Doucet, 2001] :

$$ESS_j = \frac{\left(\sum_s w_j(\theta_s)\right)^2}{\sum_s w_j(\theta_s)^2}. \quad (8)$$

Proposition (Quality of BMA estimates)

1. *The variance of the posterior probability estimate (7) is bounded above by :*

$$\mathbb{V}\left[\widehat{\pi}(\mathcal{H}_j|y)\right] \leq \frac{1}{S}\pi(\mathcal{H}_j|y)(1-\pi(\mathcal{H}_j|y))$$

2. *ESS_j is bounded below by :*

$$ESS_j \geq S \times \widehat{\pi}(\mathcal{H}_j|y)$$

Proof : see [Keller and Kamary, 2017].

Illustration : joint calibration/validation of linear code

Goal : Explain experimental data y_{exp} from explanatory variables x_{exp}, using linear code $h(x_{exp})\theta$

$$\mathcal{H}_0 : y_{exp} = h(x_{exp})\theta + \varepsilon, \quad (9)$$

or, accounting for possible model bias [Kennedy and O'Hagan, 2001],

$$\mathcal{H}_1 : y_{exp} = h(x_{exp})\theta + \delta(x_{exp}) + \varepsilon, \quad (10)$$

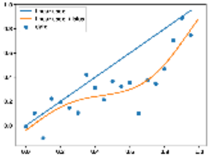
where :

$$\begin{aligned} \varepsilon|\kappa, \sigma^2 &\sim \mathcal{N}(0, \kappa\sigma^2) \\ \delta(\cdot)|\sigma^2 &\sim GP(m(\cdot), \sigma^2 K(\cdot,\cdot)) \\ \pi(\theta) &\propto 1 \\ \pi(\sigma^2) &\propto \sigma^{-2} \\ \pi(\kappa) &\sim \mathcal{U}([0,1]) \end{aligned}$$

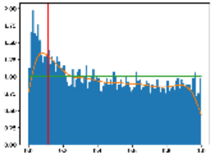
Setting already investigated in [Damblin et al., 2016], using intrinsic Bayes factors [Berger and Pericchi, 1996].

Results

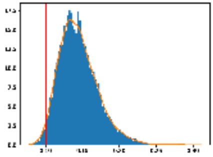
$\widehat{\pi}(\mathcal{H}_0 \mid y)$	95%LCL	95%UCL
38.6%	38.5%	38.8%



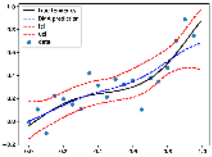
Ground truth : $\mathcal{H}_1$



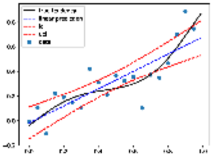
κ posterior distribution



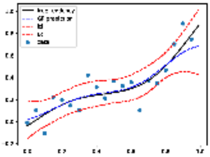
σ posterior distribution



BMA predictions



$\mathcal{H}_0$ predictions



$\mathcal{H}_1$ predictions

Discussion

- First results of joint calibration / validation algorithm are encouraging
- further tests needed, using real-life datasets and comparison to [Kamary et al., 2014a] on hydraulic case-study
- Combine with Jerome's framework for distributional robustness

$\hookrightarrow$ robust joint calibration-validation

- Any comments, suggestions ? **THANKS !**

References I

AIAA (1998).
Guide for the verification and validation of computational fluid dynamics simulations.
Reston VA : American Institute of Aeronautics and Astronautics, AIAA-G-077-1998.

Arnaud Doucet, N. d. F. . N. G., editor (2001).
Sequential Monte Carlo Methods in Practice.
Springer.

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

Berger, J. O. and Pericchi, L. R. (1996).
The intrinsic bayes factor for model selection and prediction.
Journal of the American Statistical Association, 91(433) :109–122.

References II

Brown, D. A. and Atamturktur, S. (2016).
Nonparametric Functional Calibration of Computer Models.
arXiv e-prints, page arXiv :1602.06202.

Carmassi, M. (2018).
Uncertainty quantification and calibration of a photovoltaic plant model : warranty of performance and robust estimation of the long-term production.
PhD thesis, University Paris-Saclay.

Cox, D., Park, J., and Clifford, E. (2001).
A statistical method for tuning a computer code to a data base.
Computational Statistics and Data Analysis, 37(1) :77–92.

Damblin, G. (2015).
Contributions statistiques au calage et à la validation des codes de calculs.
PhD thesis, University Paris-Saclay.

Damblin, G., Barbillon, P., Keller, M., Pasanisi, A., and Parent, E. (2017).
Adaptive numerical designs for the calibration of computer codes.
SIAM/ASA J. UNCERTAINTY QUANTIFICATION, 6(1) :151–179.

References III

Damblin, G., Keller, M., Barbillon, P., Pasanisi, A., and Parent, E. (2016).
Bayesian model selection for the validation of computer codes.
Quality and Reliability Engineering International, 32(6) :2043–2054. qre.2036.

de Crécy, A. (1997).
Circé : a tool for calculating the uncertainties of the constitutive relationships of cathare 2.
In *NURETH-8*.

Fu, S. (2012).
Inversion probabiliste bayésienne en analyse d'incertitude.
Thèse de l'Université Paris-Sud.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999).
Bayesian Model Averaging : A Tutorial.
STATISTICAL SCIENCE, 14(4) :382–417.

Kamary, K., Keller, K., Barbillon, P., Gœury, C., and Parent, E. (2014a).
Testing hypotheses via a mixture estimation model.
ArXiv e-prints.

References IV

Kamary, K., Mengersen, K., Robert, C. P., and Rousseau, J. (2014b).
Testing hypotheses via a mixture estimation model.
ArXiv e-prints.

Kass, R. E. and Raftery, A. E. (1995).
Bayes factors.
Journal of the American Statistical Association, 90(430) :773–795.

Keller, M. and Kamary, K. (2017).
Bayesian model averaging by mixture modeling.
ArXiv e-prints.

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

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.

References V

NRC (2012).
Assessing the Reliability of Complex Models : Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification.
Washington, D.C. : The National Academies Press.

Oberkampf, W. L. and Barone, M. F. (2006).
Measures of agreement between computation and experiment : Validation metrics.
Journal of Computational Physics, 217(1) :5–36.
Uncertainty Quantification in Simulation Science.

Plumlee, M. (2016).
Bayesian calibration of inexact computer models.
Journal of the American Statistical Association, 112.

Robert, C. and G., C. (2004).
Monte-Carlo Statistical Methods.
Springer-Verlag.

References VI

Roy, C. and Oberkampf, W. (2010).
A complete framework for verification, validation, and uncertainty quantification in scientific computing.
Technical report, American Institute of Aeronautics and Astronautics. AIAA 2010-124.

Sacks, J., W.J., W., Mitchell, T., and Wynn, H. (1989).
Design and analysis of computer experiments.
Statistical Science, 4(4) :409–423.

Spagnol, A., Le Riche, R., and Da Veiga, S. (2018).
Global sensitivity analysis for optimization with variable selection.
arXiv e-prints, page arXiv :1811.04646.