Virtual cross validation for numerical models calibration

François Bachoc†‡, Josselin Garnier‡ and Jean-Marc Martinez†

†CEA-Saclay, DEN, DM2S, SFME, LGLS, F-91191 Gif-sur-Yvette, France.

‡Laboratoire de Probabilités et Modèles Aléatoires & Laboratoire Jacques Louis Lions Université Paris VII

Introduction

- Context: Kriging in the context of numerical modelization of physical systems.
- Goal: Calibration of a numerical model from physical experiments and prediction of undone experiments.
- Here, we focus on the problem of hyper-parameters estimation for the Kriging process.

Theoretical Framework

We study the following Kriging model:

$$Y_{obs}(\omega, x) = \sum_{i=1}^{q} f_i(x)\beta_i(\omega) + Z(\omega, x) + \epsilon$$

where:

- $Y_{obs}(\omega, x)$ is the result of an experiment with experimental condition x.
- $\sum_{i=1}^{q} f_i(x)\beta_i$ is the result of the simulation of the physical system with experimental condition $x \in \mathcal{X}$, the numerical model being calibrated with parameters $(\beta_1, ..., \beta_q)$. We do the linear approximation of the numerical model.
- $\beta(\omega)$ is the correct parameter of the numerical model. In the frequentist case, β is an unknown constant and in the Bayesian case, β follows the $\mathcal{N}(\beta_{prior}, Q_{prior})$ law (known by expert judgement).
- $Z(\omega, x)$ is a centered Gaussian process on the set of experimental conditions $\mathcal{X} \in \mathbb{R}^d$ (model error).
- ϵ is an *iid* Gaussian centered measure error.

When the covariance function of the measure and model error process $Z + \epsilon$ is known the tasks of calibration (estimation of β) and prediction (prediction of $Y_{obs}(x_{new})$ according to a set of observations) can be solved by the analytical Kriging equations from the DACE methodology (e.g [SWN03]).
We are interested in finding a suitable covariance function for $Z + \epsilon$ in a parametric set $\{C_\theta, \theta \in \Theta\}$ of covariance functions on $\mathcal{X}$. We study the methods deriving from the maximum likelihood and the cross validation principle. We focus on the latter in the next part.

Virtual Leave One Out

For n observations, the functional model becomes,

$$y_{obs} = H\beta + z + \epsilon$$

with y_{obs}, z and ϵ n-dimensional Gaussian vectors and H an $n \times q$ matrix. When the covariance matrix R of $z + \epsilon$ is known we have, for any i, the expression of the prediction $\hat{y}_{obs,i}$ of $y_{obs,i}$ and predictive variance $\hat{\sigma}_i^2$ according to the vector $y_{obs,-i}$ of remaining observations. These scalar expressions can be gathered in a closed form matricial expression (see [Dub83]).
In the frequentist framework we have:
with $\epsilon_{loo,i} = y_{obs,i} - \hat{y}_{obs,i}$, $\sigma_{loo,i}^2 = \hat{\sigma}_i^2$ and $Q^{-} = R^{-1} - R^{-1}H(H^tR^{-1}H)^{-1}H^tR^{-1}$,

$$\boxed{\epsilon_{loo,i} = \frac{1}{Q^{-}_{i,i}}\left[Q^{-}y_{obs}\right]_i \quad \text{and} \quad \sigma_{loo,i}^2 = \frac{1}{Q^{-}_{i,i}}.}$$

In the Bayesian framework, the formula becomes, with $Q^{-1} = R^{-1} - R^{-1}H(H^tR^{-1}H + Q_{prior}^{-1})^{-1}H^tR^{-1}$:

$$\boxed{\epsilon_{loo,i} = \frac{1}{(Q^{-1})_{i,i}}\left[Q^{-1}(y_{obs} - H\beta_{prior})\right]_i \quad \text{and} \quad \sigma_{loo,i}^2 = \frac{1}{(Q^{-1})_{i,i}}}$$

Notice also that the frequentist formula is the limit of the Bayesian formula when Q_{prior}^{-1} tends to zero. The principle

of LOO based methods (see e.g [ZW10]) is therefore the following:

- For $\theta \in \Theta$ a vector of hyper-parameters for the covariance function, compute the covariance matrix R_θ of $z + \epsilon$.
- Compute the vectors of LOO errors $\epsilon_{loo,\theta}$ and predictive variance $\sigma_{loo,\theta}^2$ using formulas above (one $n \times n$ matrix inversion needed).
- "Optimize" criteria based on these vectors w.r.t θ.

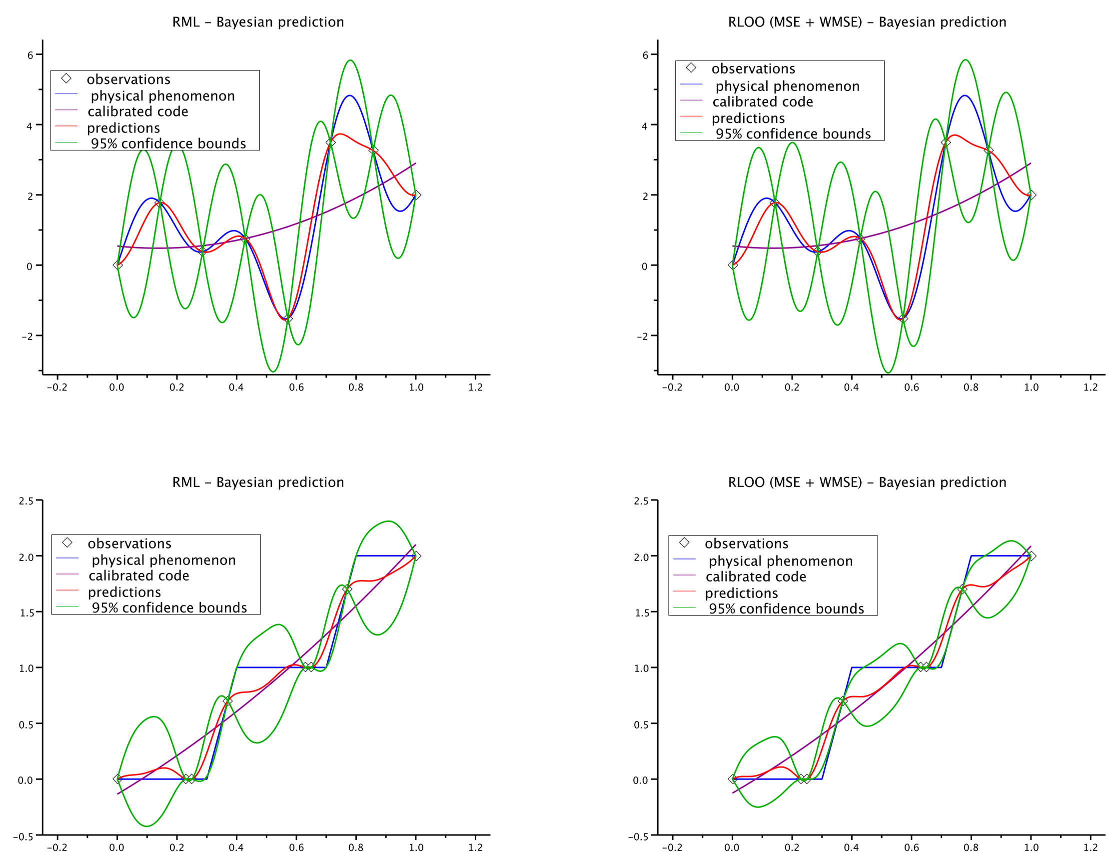
Mean square error	MSE	$\|\|\epsilon_{loo,\theta}\|\|^2$	to minimize
Weighted mean square error	WMSE	$\frac{1}{n}\sum_{i=1}^{n}\frac{(\epsilon_{loo,\theta,i})^2}{\sigma_{loo,\theta,i}^2}$	to set close to 1
Confidence bounds reliability	CBR	$\frac{1}{n}\sum_{i=1}^{n}\sharp\left\{i \mid \|\epsilon_{loo,\theta,i}\| \le t_\alpha\sigma_{loo,\theta,i}\right\}$	to set close to prob. α
Log predictive probability	LPP	$-\sum_{i=1}^{n}\left(\ln\sigma_{loo,\theta,i}^2 + \frac{(\epsilon_{loo,\theta,i})^2}{\sigma_{loo,\theta,i}^2}\right)$	to maximize

Name, acronyms, expression and goal behavior for some LOO criteria. Criterion LPP is presented in [RW06].

Illustation on 1D cases

Covariance model	hyper-parameter estimation method	real function	estimated hyper-parameter
$\sigma^2\exp\left(-\frac{(x-y)^2}{l_c^2}\right)$	RML	smooth	$\sigma = 1.77$ $l_c = 0.09$
$\sigma^2\exp\left(-\frac{(x-y)^2}{l_c^2}\right)$	RLOO (MSE + WMSE)	smooth	$\sigma = 1.76$ $l_c = 0.09$
$\sigma^2\exp\left(-\left(\frac{x-y}{l_c}\right)^p\right)$	RML	piecewise affine	$\sigma = 0.23$ $l_c = 0.078$ $p = 2$
$\sigma^2\exp\left(-\left(\frac{x-y}{l_c}\right)^p\right)$	RLOO (MSE + WMSE)	piecewise affine	$\sigma = 0.13$ $l_c = 0.065$ $p = 2$

Hyper-parameter estimation for different real functions and covariance models using Restricted LOO and ML. For Resticted LOO we optimized both criteria MSE and WMSE.



Result of hyper-parameters estimation for calibration and prediction. Top: smooth real function. Bot: piecewise affine real function.

Calibration of the thermohydraulic code FLICA-4

- Physical context: Thermohydraulic studies of nuclear reactors carried out at Nuclear Energy Division of CEA
- Dimensions and number of experiments: $q = 2$, $d = 6$, $n = 108$.
- From a statistical point of view: design of experiments is not space filling. (4 of the 6 experimental conditions take less than 5 values and 2 of these 4 are quasi-linearly correlated).
- We place ourselves in the Bayesian case w.r.t the 2-dimensional parameter β of the code.

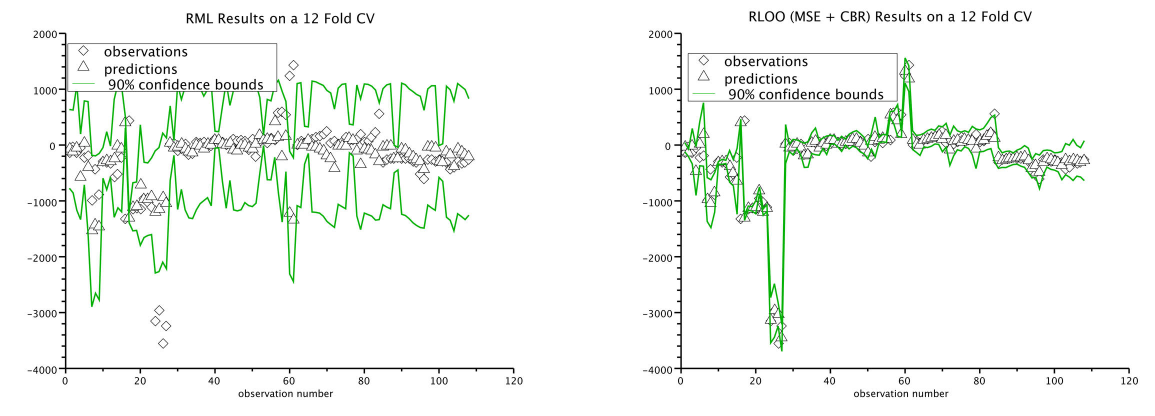
We present here the results obained with the isotropic exponential covariance model (most interesting results because of the irregularity of the data).

Step 1: We see the behavior of 5 different methods on the 108 observations.

Method	RML	ML	LOO(PLP)	RLOO (MSE+WMSE)	LOO (MSE+CBR)
$\hat{\sigma}$	686	664	760	3476	390
$\hat{l}_c$	0.011	0.01	0.013	0.59	0.50

Hyper-parameter estimation for different LOO and ML methods. We see 2 groups of models proposed (one with $l_c \approx 0.01$ and one with $l_c \approx 0.5$). Between the 2 last columns, the criterion CBR seems preferable to us because of 2 outliers in the experiments (close in distance but very different for experimental results).

Step 2: We keep the RML and LOO(MSE+CBR) methods and see the results they give on a 12-fold cross validation procedure.



Plot of the observations, predictions and confidence bounds for the 2 methods. Left: Result of the RML algorithm: $MSE = 646^2$, 93% of 90%-confidence bounds are valid.Right: Result of the LOO(MSE+CBR) algorithm: $MSE = 281^2$, 83% of 90%-confidence bounds are valid.

Conclusion

- Cross validation is not much more costly than Maximum Likelihood. Good results on this thermohydraulic code FLICA-4 case study.
- Questions: Which criterion to use? Which covariance model to choose?

References

[Dub83] O. Dubrule. Cross validation of kriging in a unique neighborhood. *Mathematical Geology*, 15, 1983.

[RW06] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[SWN03] T.J. Santner, B.J. Williams, and W.I. Notz. *The design and analysis of computer experiments*. Springer, 2003.

[ZW10] H. Zhang and Y. Wang. Kriging and cross validation for massive spatial data. *Environmetrics*, 21, 2010.