Nurisp SP4 Meeting
Bayesian calibration methods

François Bachoc
Jean-Marc Martinez
Paris VII University supervisor: Josselin Garnier

CEA - DEN/DM2S/SFME/LGLS

November 2010

Introduction

Context

- Phd started in october in partnership between CEA and Paris VII university.
- CEA supervisor : Jean Marc Martinez.
- Paris VII supervisor : Josselin Garnier.

Subject

- Very general context of the probabilistic modelling of uncertainties in the industry (ex : French research group Mascot Num).
- For us : probabilistic modelling of the error between a computation code (or numerical model) and the real system.
- Goals : To calibrate the computation code and to improve its predictions.
- Scientific keywords : numerical simulation, gaussian processes, estimation, prediction, bayesian framework, model error.

Computation code and reality

A computation code, or parametric numerical model, is represented by a function f :

$$\begin{aligned} f \quad : \mathbb{R}^r \times \mathbb{R}^m &\rightarrow \mathbb{R} \\ (\boldsymbol{x}, \boldsymbol{\beta}) &\rightarrow f(\boldsymbol{x}, \boldsymbol{\beta}) \end{aligned}$$

The numerical phenomenon is represented by a function Y_{real}.

$$\begin{aligned} Y_{real} \quad : \mathbb{R}^r &\rightarrow \mathbb{R} \\ \boldsymbol{x} &\rightarrow Y_{real}(\boldsymbol{x}) \end{aligned}$$

- The inputs $\boldsymbol{x}$ are the experimental conditions (ex : geometric factors, limit conditions).
- The inputs $\boldsymbol{\beta}$ are the calibration parameters of the computation code (eg : physical laws parameters).
- The output $f(\boldsymbol{x}, \boldsymbol{\beta})/Y_{real}(\boldsymbol{x})$ is a quantity of interest (eg : a produced energy).

A computation code modelizes (gives an approximation of) a physical phenomenon.

Model error

Statistical modelling : The physical phenomenon is random and centered around the correctly parameterized computation code.

Equation of the statistical model

$$Y_{real}(\omega, \boldsymbol{x}) = f(\boldsymbol{x}, \boldsymbol{\beta}(\omega)) + Z(\omega, \boldsymbol{x})$$

- Equation that holds for a specific parameters vector $\boldsymbol{\beta}$. Called "the" parameter of the computation code.
  - No prior information case : $\boldsymbol{\beta}$ constant and unknown.
  - Prior information case (bayesian case) : $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_{prior}, \mathbf{Q}_{prior})$
- Z is a centered, stationnary, gaussian process. We denote by C_{mod} the covariance function of Z.
  - C_{mod} belongs to a parametric set : ▸ Illustration

Why a stationnary gaussian process ?

- Gaussian variables : most commonly used to represent errors, conserve themselves through conditional expectations and linear operations.
- Stationnarity : restrict the number of possible gaussian processes (statistical bias-variance trade-off). In statistical inference : replace sample repetition (iid case) by spatial repetition.

Measure error

- We also take possible measure errors into account.
- The different measure errors are represented by iid centered gaussian variables.
- The gaussian process for the measured, or observed, physical phenomenon is therefore :

$$Y_{obs}(\omega, \boldsymbol{x}) = Y_{real}(\omega, \boldsymbol{x}) + \epsilon(\omega, \boldsymbol{x})$$

- We denote by σ_{mes} the standard deviation of the measure error.
- For simplicity here, we suppose that we do not do more than one measure for the same vector of experimental conditions.

Goals associated to the modelling

Kinds of work to do :

1. The covariance function of the model error is known : Calibration and Prediction.
2. A covariance function is proposed : Model test.
3. The covariance function is unknown : Model selection.

Outline of studies using the modelling

- Step 1 : Estimation of the hyper-parameters of the covariance function.
- Step 2 : Plug-in of the estimated hyper-parameters to perform calibration and prediction.

Linear code and observations : notations (1/3)

Linear code w.r.t the parameters

$$\forall \boldsymbol{x} : \; f(\boldsymbol{x}, \boldsymbol{\beta}) = \sum_{i=1}^{m} h_i(\boldsymbol{x})\beta_i$$

Observations

- We observe the physical phenomenon $Y_{real}(\boldsymbol{x})$ for n (different) inputs $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)}$. Associated random vectors :

$$\boldsymbol{y}_{obs} = \begin{pmatrix} Y_{obs}(\boldsymbol{x}^{(1)}) \\ \vdots \\ Y_{obs}(\boldsymbol{x}^{(n)}) \end{pmatrix}, \boldsymbol{y}_{real} = \begin{pmatrix} Y_{real}(\boldsymbol{x}^{(1)}) \\ \vdots \\ Y_{real}(\boldsymbol{x}^{(n)}) \end{pmatrix}$$

- We want to predict the value of the phenomenon $Y_{real}(\boldsymbol{x})$ for a new input $\boldsymbol{x}^{(0)}$. Associated random variable :

$$y_0 = Y_{real}(\boldsymbol{x}^{(0)})$$

Linear code and observations : notations (2/3)

Matrix associated to the code and the observations

We define the $n \times m$ matrix $\mathbf{H}$ by :

$$H_{i,j} = h_j(\boldsymbol{x}^{(i)}) \quad i = 1, ..., n \;\; j = 1, ..., m$$

The vector of outputs of the code, parameterized by $\boldsymbol{\beta}$ for $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)}$:

$$\begin{pmatrix} f(\boldsymbol{x}^{(1)}, \boldsymbol{\beta}) \\ \vdots \\ f(\boldsymbol{x}^{(n)}, \boldsymbol{\beta}) \end{pmatrix} = \mathbf{H}\boldsymbol{\beta}$$

Vector associated to the code and $\boldsymbol{x}^{(0)}$

We define the m size vector, $\boldsymbol{h}^{(0)}$ by :

$$\boldsymbol{h}_i^{(0)} = h_i(\boldsymbol{x}^{(0)}) \quad i = 1, ..., m$$

The output of the code, parameterized by $\boldsymbol{\beta}$, for $\boldsymbol{x}^{(0)}$:

$$f(\boldsymbol{x}^{(0)}, \boldsymbol{\beta}) = (\boldsymbol{h}^{(0)})^T \boldsymbol{\beta}$$

Linear code and observations : notations (3/3)

Model error vector

Vectors of the model and measure error for the observations :

$$\boldsymbol{z} = \begin{pmatrix} Z(\boldsymbol{x}^{(1)}) \\ \vdots \\ Z(\boldsymbol{x}^{(n)}) \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon(\boldsymbol{x}^{(1)}) \\ \vdots \\ \epsilon(\boldsymbol{x}^{(n)}) \end{pmatrix}$$

Model error at $\boldsymbol{x}^{(0)}$:

$$z_0 = Z(\boldsymbol{x}^{(0)})$$

Covariance matrix

$\mathbf{R}$: Covariance matrix of $\boldsymbol{z} + \boldsymbol{\epsilon}$:

$$\mathbf{R}_{i,j} = C_{mod}(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}) + \sigma^2_{mes}\mathbf{1}_{i=j} \quad i = 1, ..., n \;\; j = 1, ..., n$$

$\boldsymbol{r}^{(0)}$: Covariance vector between $\boldsymbol{z} + \boldsymbol{\epsilon}$ and z_0 :

$$r_i^{(0)} = C_{mod}(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(0)}) \quad i = 1, ..., n$$

Matrix equation of the statistical model

The statistical model becomes, for the inputs $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)}$:

$$\boldsymbol{y}_{obs} = \mathbf{H}\boldsymbol{\beta} + \boldsymbol{z} + \boldsymbol{\epsilon}$$

With $\boldsymbol{z} + \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{R})$.

- No prior information case
  - When $\mathbf{R} = \sigma^2 \mathbf{I}_n$ (nugget covariance function) : Linear regression model.
- Prior information case

$$\boldsymbol{y}_{obs} \sim \mathcal{N}(\mathbf{H}\boldsymbol{\beta}_{prior}, \mathbf{R} + \mathbf{H}\mathbf{Q}_{prior}\mathbf{H}^T)$$

- Main interest of the correlation : Efficient prediction of the phenomenon when it does not have the same shape as the computation code.

Calibration (1/2)

Calibration problem $=$ Statistical estimation problem

Estimation of $\boldsymbol{\beta}$

- An estimator of $\boldsymbol{\beta}$ is a function $\hat{\boldsymbol{\beta}} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$.
- $\hat{\boldsymbol{\beta}}(\boldsymbol{y}_{obs})$ is the estimation of $\boldsymbol{\beta}$ according to the vector of observations $\boldsymbol{y}_{obs}$.
- Quality measure of an estimator : Mean square error : $\mathbb{E}_{\boldsymbol{y}_{obs},\boldsymbol{\beta}}\left[||\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\boldsymbol{y}_{obs})||^2\right]$.

Calibration (2/2)

No prior information case

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, according to the vector of observations $\boldsymbol{y}_{obs}$ at $(\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)})$ is :

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\boldsymbol{y}_{obs}$$

- Unbiased estimator ($\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$)
- If $\boldsymbol{y}_{obs} = \mathbf{H}\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}(\boldsymbol{y}_{obs}) = \boldsymbol{\beta}$

Prior information case

In the prior information case, the conditional law of $\boldsymbol{\beta}$, according to the observations $\boldsymbol{y}_{obs}$ is gaussian with mean $\boldsymbol{\beta}_{post}$, where

$$\boldsymbol{\beta}_{post} = \boldsymbol{\beta}_{prior} + (\mathbf{Q}_{prior}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta}_{prior}).$$

- Best predictor according to the mean square error.
- When $\mathbf{Q}_{prior}^{-1} \rightarrow 0$ (Uninformative prior) we find the prediction of the no prior information case.

Prediction (1/2)

Prediction of y_0

- A predictor of y_0 is a function $\langle y_0 \rangle : \mathbb{R}^n \longrightarrow \mathbb{R}$.
- $\langle y_0 \rangle(\boldsymbol{y}_{obs})$ is the prediction of y_0 according to the vector of observations $\boldsymbol{y}_{obs}$.
- Quality measure of a predicor : Mean square error : $\mathbb{E}_{\boldsymbol{y}_{obs}, y_0}\left[|y_0 - \langle y_0 \rangle(\boldsymbol{y}_{obs})|^2\right]$.

Prediction (2/2)

No prior information case

The unbiased predictor of y_0 at $\boldsymbol{x}^{(0)}$, linear with respect to the vector of observations $\boldsymbol{y}_{obs}$ at $(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)})$, which minimizes the mean square error is :

$$\langle y_0 \rangle = (\boldsymbol{h}^{(0)})^T \hat{\boldsymbol{\beta}} + (\boldsymbol{r}^{(0)})^T \mathbf{R}^{-1} (\boldsymbol{y}_{obs} - \mathbf{H}\hat{\boldsymbol{\beta}})$$

with $\hat{\boldsymbol{\beta}}$ the maximum likelihood estimator of $\boldsymbol{\beta}$.

- We do not have access to the best predictor, because its expression makes use of the unknown parameter $\boldsymbol{\beta}$.
- The prediction expression is decomposed into a calibration term and a gaussian inference term of the model error.

Prior information case

The conditional law of y_0 according to the observations $\boldsymbol{y}_{obs}$ is gaussian with mean $\langle y_0 \rangle$, with :

$$\langle y_0 \rangle = (\boldsymbol{h}^{(0)})^T \boldsymbol{\beta}_{post} + (\boldsymbol{r}^{(0)})^T \mathbf{R}^{-1} (\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta}_{post})$$

- Best predictor.
- When $\mathbf{Q}_{prior}^{-1} \to 0$ (uninformative prior) we find the predictor of the no prior information case.

Illustration of calibration (1/3)

- Physical phenomenon : $Y(x) = x^2$.
- Computation code : $f(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x$.
- Covariance function of the model error : $C_{mod}(x - y) = \sigma^2 \exp\left(-\frac{|x-y|^2}{l_c^2}\right)$. $\sigma = 0.3$, $l_c = 0.5$ (known).
- Measure error : $\sigma_{mes} = 0.1$.
- Bayesian case with :

$$\boldsymbol{\beta}_{prior} = \begin{pmatrix} 0.2 \\ 1 \end{pmatrix}, \mathbf{Q}_{prior} = \begin{pmatrix} 0.09 & 0 \\ 0 & 0.09 \end{pmatrix}$$

- Observations : $x_1 = 0.2$, $x_2 = 0.4$, $x_3 = 0.6$ and $x_4 = 0.8$.

Illustration of calibration (2/3) (unnoised case)

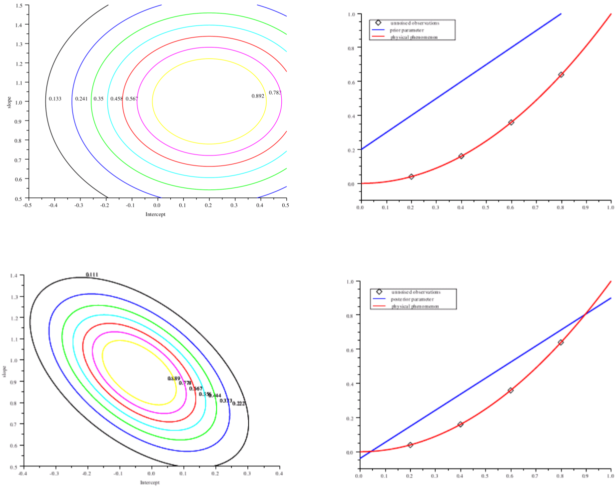


FIG.: Up-left : Prior distribution of the parameter β. Down-left : Posterior distribution of the parameter β. Right : plot of the code response corresponding to prior and posterior mean of the code parameter.

Illustration of calibration (3/3) (noised case)

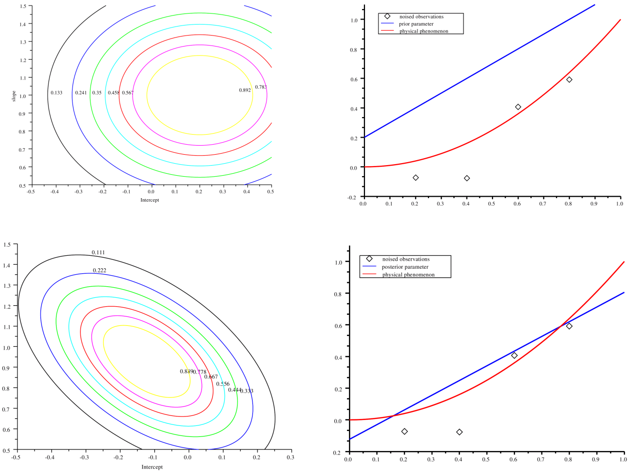
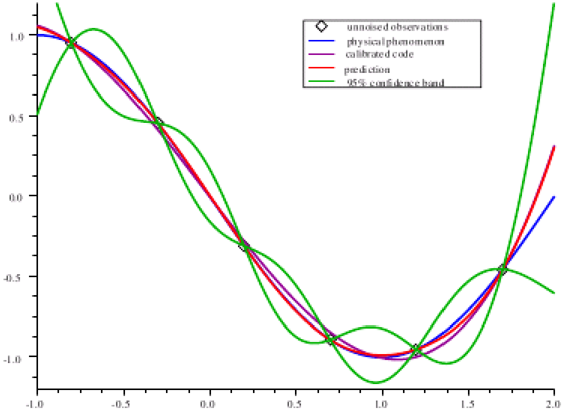


FIG.: Up-left : Prior distribution of the parameter β. Down-left : Posterior distribution of the parameter β. Right : plot of the code response corresponding to prior and posterior mean of the code parameter.

Illustration of prediction (1/3)

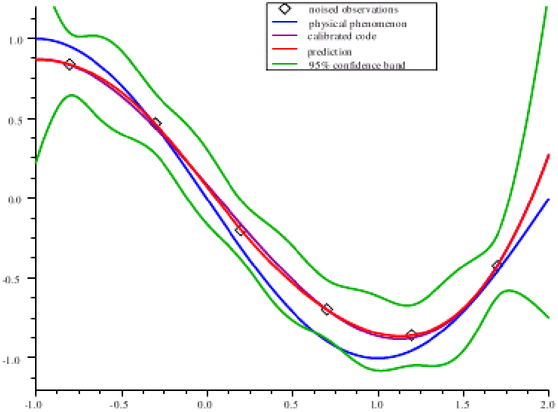
- Physical phenomenon : $Y(x) = -sin(\frac{\pi x}{2})$.
- Computation code : $f(x, \beta) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.
- Covariance function of the model error :
  $C_{mod}(x - y) = \sigma^2 \exp\left(-\frac{|x-y|^2}{l_c^2}\right)$. $\sigma = 0.3$, $l_c = 0.5$ (known).
- Measure error : $\sigma_{mes} = 0.1$.
- No prior information case.
- 6 Observations regularly sampled between -0.8 and 1.7.

Illustration of prediction (2/3) (unnoised case)



- The use of the model error improves the prediction given by the computation code.

Illustration of prediction (3/3) (noised case)



- The measure error deteriorates the quality of the predictions.
- The confidence bands are however still reliable.

Framework

- The calibration and prediction methods presented above give good results because we used a reasonnable covariance function.
- The model selection is a statistical parameter estimation problem.

Variance and correlation component

We suppose that the model error covariance is $C_{mod}(\boldsymbol{x}, \boldsymbol{y}) = \sigma^2_{mod} Corr_{mod,\alpha}(\boldsymbol{x} - \boldsymbol{y})$. Then the covariance function C of $Z + \epsilon$ takes the form :

$$C(\boldsymbol{x}, \boldsymbol{y}) = (\sigma^2_{mod} + \sigma^2_{mes}) \left(\frac{\sigma^2_{mod}}{\sigma^2_{mod} + \sigma^2_{mes}} Corr_{mod,\alpha}(\boldsymbol{x} - \boldsymbol{y}) + (1 - \frac{\sigma^2_{mod}}{\sigma^2_{mod} + \sigma^2_{mes}}) \mathbf{1}_{\boldsymbol{x}=\boldsymbol{y}} \right)$$

Therefore we can reparameterize with a variance parameter and a correlation parameter.

So, in the sequel, we suppose that the covariance of $Z + \epsilon$ has the form :

$$C(\boldsymbol{x} - \boldsymbol{y}) = \sigma^2 Corr_\theta(\boldsymbol{x} - \boldsymbol{y}).$$

We present 2 methods for model selection : Maximum likelihood and cross validation.

Maximum likelihood (No prior information case)

We denote by $\mathbf{R}_{corr,\theta}$ the correlation matrix of the observations.
The log-likelihood of the observations $\boldsymbol{y}_{obs}$ is :

$$\ell(\boldsymbol{\beta},\sigma,\theta) \propto -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(|\mathbf{R}_{corr,\theta}|) - \frac{1}{2\sigma^2}(\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta})^T(\mathbf{R}_{corr,\theta})^{-1}(\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta})$$

We denote :

$$\hat{\boldsymbol{\beta}}(\theta) = (\mathbf{H}^T(\mathbf{R}_{corr,\theta})^{-1}\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{R}_{corr,\theta})^{-1}\boldsymbol{y}_{obs}$$

and

$$\hat{\sigma}^2(\theta) = \frac{1}{n}(\boldsymbol{y}_{obs} - \mathbf{H}\hat{\boldsymbol{\beta}}(\theta))^T(\mathbf{R}_{corr,\theta})^{-1}(\boldsymbol{y}_{obs} - \mathbf{H}\hat{\boldsymbol{\beta}}(\theta)).$$

The maximum likelihood estimator of $\hat{\theta}$, $\hat{\sigma}$ and $\hat{\boldsymbol{\beta}}$ of θ, σ and $\boldsymbol{\beta}$ are :

$$\boxed{\hat{\theta} \in \arg\min_{\theta} |\mathbf{R}_{corr,\theta}|^{\frac{1}{n}}\hat{\sigma}(\theta)^2}$$

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\theta}),$$

and

$$\hat{\sigma} = \hat{\sigma}(\hat{\theta}).$$

Maximum likelihood (Bayesian case)

Recall that in the bayesian case :

$$\boldsymbol{y}_{obs} \sim \mathcal{N}(\mathbf{H}\boldsymbol{\beta}_{prior}, \sigma^2 \mathbf{R}_{corr,\theta} + \mathbf{H}\mathbf{Q}_{prior}\mathbf{H}^T)$$

- The log-likelihood is not anymore a function of $\boldsymbol{\beta}$.
- The gaussian process Y_{obs} is not anymore stationnary $\rightarrow$ No separation of σ^2 and θ in the maximisation of the likelihood.

By denoting :

$$\mathbf{Q}_{\sigma,\theta} = \sigma^2 \mathbf{R}_{corr,\theta} + \mathbf{H}\mathbf{Q}_{prior}\mathbf{H}^T$$

we have the log-likelihood :

$$\ell(\sigma,\theta) \propto -\frac{1}{2}\log(det(\mathbf{Q}_{\sigma,\theta})) - \frac{1}{2}(\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta}_{prior})^T(\mathbf{Q}_{\sigma,\theta})^{-1}(\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta}_{prior})$$

and :

$$\boxed{(\hat{\sigma},\hat{\theta}) \in \arg\max_{\sigma,\theta} \ell(\sigma,\theta)}$$

Cross validation (1/3)

The prediction procedure (Bayesian or non-bayesian framework) leads to a parametric metamodel : $\boldsymbol{x}^{(0)} \longrightarrow \langle y_0 \rangle_{\sigma,\theta}$.

- It is a function that approximates the physical phenomenon.
- It is built according to the observations ($\approx$ learning set).

Cross validation (LOO)

- Given a vector of hyper-parameters (σ, θ).
- For i from 1 to n we learn $\boldsymbol{x}^{(0)} \longrightarrow \langle y_0 \rangle_{\sigma,\theta}$ with the reduced observations vector $\{(x_1, y_{obs,1}), \ldots, (x_{i-1}, y_{obs,i-1}), (x_{i+1}, y_{obs,i+1}), \ldots, (x_n, y_{obs,n})\}$
- we compute the LOO errors by :

$$\epsilon_{LOO,i}(\sigma, \theta) = y_{obs,i} - \langle y_i \rangle_{\sigma,\theta}(\boldsymbol{y}_{obs,-i}).$$

Then, the LOO estimator of σ and θ is :

$$\boxed{(\hat{\sigma}, \hat{\theta}) \in \arg\min_{\sigma,\theta} ||\boldsymbol{\epsilon}_{LOO}(\sigma, \theta)||^2}$$

Cross validation (2/3) Non-bayesian framework

We have a closed formula for the LOO errors vector :
With :

$$\mathbf{Q}^{-}(\sigma,\theta) = \frac{1}{\sigma^2}\left(\mathbf{R}_{corr,\theta}^{-1} - \mathbf{R}_{corr,\theta}^{-1}\mathbf{H}(\mathbf{H}^T\mathbf{R}_{corr,\theta}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}_{corr,\theta}^{-1}\right)$$

We have :

$$\boldsymbol{\epsilon}(\sigma,\theta) = (diag(\mathbf{Q}^{-}))^{-1}\mathbf{Q}^{-}$$

$\boldsymbol{\epsilon}$ actually does not depends on σ.
Therefore we have :

$$\hat{\theta} \in \arg\min_{\theta} ||\boldsymbol{\epsilon}_{LOO}(\theta)||^2$$

and we keep the estimator for σ :

$$\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y}_{obs} - \mathbf{H}\hat{\beta}(\hat{\theta}))^T(\mathbf{R}_{corr,\hat{\theta}})^{-1}(\boldsymbol{y}_{obs} - \mathbf{H}\hat{\beta}(\hat{\theta}))$$

Cross validation (3/3) Bayesian framework

We also have a closed formula for the LOO errors vector :
Recall :

$$\mathbf{Q}_{\sigma,\theta} = \sigma^2 \mathbf{R}_{corr,\theta} + \mathbf{H}\mathbf{Q}_{prior}\mathbf{H}^T$$

We have :

$$\boldsymbol{\epsilon}(\sigma,\theta) = (diag(\mathbf{Q}^{-1}))^{-1}\mathbf{Q}^{-1}$$

This time, $\boldsymbol{\epsilon}$ depends on both σ ans θ.
We do the optimization w.r.t σ and θ.

Conclusion and prospects

Conclusion

- For prediction, expert knowledge is taken into account via the bayesian framework.
- The predictions are similar than the ones given by non-probabilistic kernels methods.
- The statistical model gives confidence bands for predictions and calibration.
- The hyper-parameter estimation step is crucial.

Prospects of the Phd

- Model selection procedure : asymptotic study of the maximum likelihood and LOO.
- Non linear codes.

Thank you for your attention.

Model error illustration

- Exemples of covariance functions

  "Nugget" model $C_{mod}(\boldsymbol{x} - \boldsymbol{y}) = \sigma^2 \delta_{\boldsymbol{x}-\boldsymbol{y}}$

  Gaussian model $C_{mod}(\boldsymbol{x} - \boldsymbol{y}) = \sigma^2 \exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{y}||^2}{l_c^2}\right)$

  Generelized exponential model $C_{mod}(\boldsymbol{x} - \boldsymbol{y}) = \sigma^2 \exp\left(-(\frac{|\boldsymbol{x}-\boldsymbol{y}|_1}{l_c})^p\right)$
- All 3 are stationnary and parametric. We call hyper-parameters the parameters of the covariance functions. Here hyper-parameters are σ, (σ, l_c) and (σ, l_c, p).
- Exemples of realizations with gaussian covariance function.

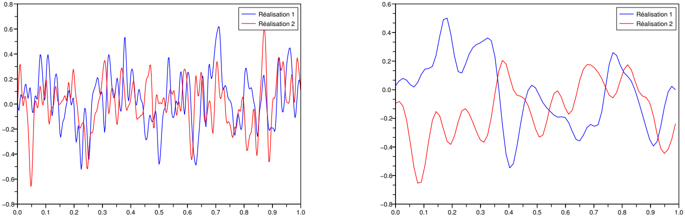


FIG.: Left : $\sigma = 0.2$, $l_c = 0.01$ Right : $\sigma = 0.2$, $l_c = 0.05$