Parametric estimation of covariance function in Gaussian-process based Kriging models. Application to uncertainty quantification for computer models

François Bachoc

PhD advisor: Josselin Garnier
CEA advisor: Jean-Marc Martinez

CEA-Saclay, DEN, DM2S, STMF, LGLS, F-91191 Gif-Sur-Yvette, France
LPMA, Université Paris 7

October 3 2013

1. Kriging for calibration, improved prediction and metamodeling of computer models
   - Statistical model and method for calibration and improved prediction of computer models
   - Application to the FLICA 4 thermal-hydraulic code
2. Maximum Likelihood and Cross Validation for parametric covariance function estimation
3. Finite sample analysis of ML and CV under model misspecification
4. Asymptotic analysis of ML and CV in the well-specified case
   - Asymptotic framework
   - Consistency and asymptotic normality
   - Analysis of the asymptotic variance matrices
5. Conclusion and perspectives

Numerical code and physical system

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

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

Observations can be made of a physical system Y_{real}

$$\boldsymbol{x}_i \quad \to \boxed{Y_{real}} \to y_{obs,i}$$

- The inputs $\boldsymbol{x}$ are the experimental conditions
- The inputs $\boldsymbol{\beta}$ are the calibration parameters of the numerical code
- The outputs $f(\boldsymbol{x}_i, \boldsymbol{\beta})$ and $y_{obs,i}$ are the variable of interest

A numerical code modelizes (gives an approximation of) a physical system

Statistical model

Models for the observations and the physical system

$$y_{obs,i} = \underbrace{Y_{real}(\boldsymbol{x}_i)}_{\text{physical system}} + \underbrace{\epsilon_i}_{\text{measure error}}$$

$$Y_{real}(\boldsymbol{x}) = \underbrace{f(\boldsymbol{x}, \beta)}_{\text{parameterized code}} + \underbrace{Z(\boldsymbol{x})}_{\text{model error}}$$

- Unknown parameter β : frequentist or Bayesian framework
- Model error function Z modeled as the realization of a centered Gaussian process

Objectives and treatment

Objectives

- Calibration : estimation of β
- Prediction : prediction of $Y_{real}(\boldsymbol{x}_{new})$ for a new experimental condition $\boldsymbol{x}_{new}$

Treatment

Linear approximation of the code w.r.t β
$\Longrightarrow$ Universal Kriging model

$$y_{obs,i} = \sum_{j=1}^{m} h_j(\boldsymbol{x}_i)\beta_j + Z(\boldsymbol{x}_i) + \epsilon_i$$

Classical linear Gaussian framework $\Longrightarrow$ classical conditioning formula $\Longrightarrow$ conditional distribution of $Y_{real}(\boldsymbol{x}_{new})$ conditionally to $y_{obs,1}, ..., y_{obs,n}$ is Gaussian with explicit mean and variance

1. Kriging for calibration, improved prediction and metamodeling of computer models
   - Statistical model and method for calibration and improved prediction of computer models
   - Application to the FLICA 4 thermal-hydraulic code
2. Maximum Likelihood and Cross Validation for parametric covariance function estimation
3. Finite sample analysis of ML and CV under model misspecification
4. Asymptotic analysis of ML and CV in the well-specified case
   - Asymptotic framework
   - Consistency and asymptotic normality
   - Analysis of the asymptotic variance matrices
5. Conclusion and perspectives

Settings for the thermal-hydraulic code Flica 4

The experiment

- Pressurized and possibly heated water flowing through a cylinder
- We measure the pressure drop between the two ends of the cylinder
- Variable of interest : The part of the pressure drop due to friction : ΔP_{fri}

Two kinds of experimental conditions

- System parameters : Hydraulic diameter D_h, Friction height H_f, Channel width e
- Environment variables : Output pressure P_o, Flowrate G_e, Wall heat flux Φ_w, Liquid enthalpy h^l_e, Thermodynamic title X^e_{th}, Input temperature T_i

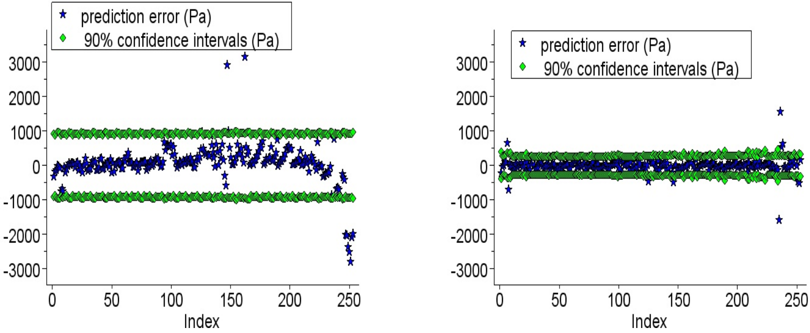
Experimental results

We dispose of 253 experimental results

Results with the thermal-hydraulic code Flica 4

Prediction results with 10-fold cross validation of the 253 experimental results :

	RMSE	90% Confidence Intervals
Calibrated code	$567 Pa$	$241/253 \approx 0.95$
Gaussian Processes	$196 Pa$	$241/253 \approx 0.95$



Influence of the linear approximation

- The Gaussian process model of the model error is also tractable in the non-linear case

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

- On the FLICA 4 data, we compare the linear approximation we use with the Bayes formula in the non-linear case for calibration and prediction.
  - Integrals are evaluated on a 5×5 grid in the calibration parameter space
  - The same grid is used for the linear-case
- We obtain
  - a 10% difference for calibration
  - a 1% difference for prediction

$\Longrightarrow$ In this case, we have shown that the model error compensates for the linearization error

Conclusion on calibration and improved prediction of computer models

- We propose to improve the prediction capability of the computer model by completing it with a statistical model
- Number of experimental results needs to be sufficient. In extrapolation (far from the experimental data) the prediction is simply given by the computer model

For more details

Bachoc F, Bois G, Garnier J and Martinez J.M, Calibration and improved prediction of computer models by universal Kriging, *Accepted in Nuclear Science and Engineering, http ://arxiv.org/abs/1301.4114v2*

1. Kriging for calibration, improved prediction and metamodeling of computer models
   - Statistical model and method for calibration and improved prediction of computer models
   - Application to the FLICA 4 thermal-hydraulic code
2. Maximum Likelihood and Cross Validation for parametric covariance function estimation
3. Finite sample analysis of ML and CV under model misspecification
4. Asymptotic analysis of ML and CV in the well-specified case
   - Asymptotic framework
   - Consistency and asymptotic normality
   - Analysis of the asymptotic variance matrices
5. Conclusion and perspectives

Parametric covariance function estimation

Parameterization

Covariance function model $\{\sigma^2 K_{\boldsymbol{\theta}}, \sigma^2 \geq 0, \boldsymbol{\theta} \in \Theta\}$ for the Gaussian Process Y.

- σ^2 is the variance hyper-parameter
- $\boldsymbol{\theta}$ is the multidimensional correlation hyper-parameter. $K_{\boldsymbol{\theta}}$ is a stationary correlation function.

Exemple : the Matérn $\frac{3}{2}$ covariance function on $\mathbb{R}$, parameterized by σ^2 and ℓ

$$\sigma^2 K_\ell(x_1, x_2) = \sigma^2 \left(1 + \sqrt{6}\frac{|x_1 - x_2|}{\ell}\right) e^{-\sqrt{6}\frac{|x_1 - x_2|}{\ell}}$$

Estimation

Y is observed at $\boldsymbol{x}_1, ..., \boldsymbol{x}_n \in \mathcal{X}$, yielding the Gaussian vector $\boldsymbol{y} = (Y(\boldsymbol{x}_1), ..., Y(\boldsymbol{x}_n))$.
Estimators $\hat{\sigma}^2(\boldsymbol{y})$ and $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ for the covariance hyper-parameters

"Plug-in" Kriging prediction

1. Estimate the covariance function
2. Assume that the covariance function is fixed and carry out the explicit Kriging equations

Maximum Likelihood

The most classical estimator for the covariance function : Maximum Likelihood (ML)

- Numerical optimization of an explicit matricial criterion

Cross Validation (Leave-One-Out)

Based on the Leave-One-Out prediction and predictive variances :

- $\hat{y}_{\boldsymbol{\theta},i,-i} = \mathbb{E}_{\sigma^2,\boldsymbol{\theta}}(Y(\boldsymbol{x}_i)|y_1,...,y_{i-1},y_{i+1},...,y_n)$
- $\sigma^2 c^2_{\boldsymbol{\theta},i,-i} = var_{\sigma^2,\boldsymbol{\theta}}(Y(\boldsymbol{x}_i)|y_1,...,y_{i-1},y_{i+1},...,y_n)$

Leave-One-Out estimation procedure we study :

$$\hat{\boldsymbol{\theta}}_{CV} \in \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \hat{y}_{\boldsymbol{\theta},i,-i})^2$$

and

$$\hat{\sigma}^2_{CV} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{y}_{\hat{\boldsymbol{\theta}}_{CV},i,-i})^2}{c^2_{\hat{\boldsymbol{\theta}}_{CV},i,-i}}$$

Virtual Leave One Out formula

Let $\mathbf{R}_{\boldsymbol{\theta}}$ be the covariance matrix of $\boldsymbol{y} = (y_1, ..., y_n)$ with correlation function $K_{\boldsymbol{\theta}}$ and $\sigma^2 = 1$

Virtual Leave-One-Out

$$y_i - \hat{y}_{\boldsymbol{\theta},i,-i} = \frac{\left(\mathbf{R}_{\boldsymbol{\theta}}^{-1}\boldsymbol{y}\right)_i}{\left(\mathbf{R}_{\boldsymbol{\theta}}^{-1}\right)_{i,i}} \quad \text{and} \quad c^2_{\boldsymbol{\theta},i,-i} = \frac{1}{(\mathbf{R}_{\boldsymbol{\theta}}^{-1})_{i,i}}$$

O. Dubrule, Cross Validation of Kriging in a Unique Neighborhood, *Mathematical Geology*, 1983.

Explicit matricial criteria for CV estimation

Using the virtual Cross Validation formula :

$$\hat{\boldsymbol{\theta}}_{CV} \in \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \frac{1}{n}\boldsymbol{y}^t\mathbf{R}_{\boldsymbol{\theta}}^{-1} diag(\mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-2}\mathbf{R}_{\boldsymbol{\theta}}^{-1}\boldsymbol{y}$$

and

$$\hat{\sigma}^2_{CV} = \frac{1}{n}\boldsymbol{y}^t\mathbf{R}_{\hat{\boldsymbol{\theta}}_{CV}}^{-1} diag(\mathbf{R}_{\hat{\boldsymbol{\theta}}_{CV}}^{-1})^{-1}\mathbf{R}_{\hat{\boldsymbol{\theta}}_{CV}}^{-1}\boldsymbol{y}$$

1. Kriging for calibration, improved prediction and metamodeling of computer models
   - Statistical model and method for calibration and improved prediction of computer models
   - Application to the FLICA 4 thermal-hydraulic code
2. Maximum Likelihood and Cross Validation for parametric covariance function estimation
3. **Finite sample analysis of ML and CV under model misspecification**
4. Asymptotic analysis of ML and CV in the well-specified case
   - Asymptotic framework
   - Consistency and asymptotic normality
   - Analysis of the asymptotic variance matrices
5. Conclusion and perspectives

Objectives

We want to study the cases of model misspecification, that is to say the cases when the true covariance function K_1 of Y is not in $\mathcal{K} = \left\{\sigma^2 K_{\boldsymbol{\theta}}, \sigma^2 \geq 0, \boldsymbol{\theta} \in \Theta\right\}$

In this context we want to compare Leave-One-Out and Maximum Likelihood estimators from the point of view of prediction mean square error and point-wise estimation of the prediction mean square error

We proceed in two steps

- When $\mathcal{K} = \left\{\sigma^2 K_2, \sigma^2 \geq 0\right\}$, with K_2 a correlation function, and K_1 the true unit-variance covariance function : theoretical formula and numerical tests
- In the general case : numerical studies

Bachoc F, Cross Validation and Maximum Likelihood estimations of hyper-parameters of Gaussian processes with model misspecification, *Computational Statistics and Data Analysis 66 (2013) 55-69, http ://dx.doi.org/10.1016/j.csda.2013.03.016*.

Case of variance hyper-parameter estimation

- $\hat{Y}_{new}$: prediction of $Y_{new} := Y(\boldsymbol{x}_{new})$ with fixed misspecified correlation function K_2
- $\mathbb{E}\left[(\hat{Y}_{new} - Y_{new})^2 \middle| \boldsymbol{y} \right]$: conditional mean square error of the prediction $\hat{Y}_{new}$
- One estimates σ^2 by $\hat{\sigma}^2$. $\hat{\sigma}^2$ may be $\hat{\sigma}^2_{ML}$ or $\hat{\sigma}^2_{CV}$
- Conditional mean square error of $\hat{Y}_{new}$ predicted by $\hat{\sigma}^2 c^2_{\boldsymbol{x}_{new}}$ with $c^2_{\boldsymbol{x}_{new}}$ fixed by K_2

Definition : the Risk

We study the Risk criterion for an estimator $\hat{\sigma}^2$ of σ^2

$$\mathcal{R}_{\hat{\sigma}^2, \boldsymbol{x}_{new}} = \mathbb{E}\left[\left(\mathbb{E}\left[(\hat{Y}_{new} - Y_{new})^2 \middle| \boldsymbol{y} \right] - \hat{\sigma}^2 c^2_{\boldsymbol{x}_{new}} \right)^2 \right]$$

Explicit expression of the Risk

Let, for $i = 1, 2$:

- $\boldsymbol{r}_i$ be the covariance vector of Y between $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ and $\boldsymbol{x}_{new}$ with covariance function K_i
- $\mathbf{R}_i$ be the covariance matrix of Y at $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ with covariance function K_i

Proposition : formula for quadratic estimators

When $\hat{\sigma}^2 = \boldsymbol{y}^t \mathbf{M} \boldsymbol{y}$, we have

$$\begin{aligned}\mathcal{R}_{\hat{\sigma}^2, \boldsymbol{x}_{new}} &= f(\mathbf{M}_0, \mathbf{M}_0) + 2c_1 tr(\mathbf{M}_0) - 2c_2 f(\mathbf{M}_0, \mathbf{M}_1) \\ &\quad + c_1^2 - 2c_1 c_2 tr(\mathbf{M}_1) + c_2^2 f(\mathbf{M}_1, \mathbf{M}_1)\end{aligned}$$

with

$$\begin{aligned} f(\mathbf{A}, \mathbf{B}) &= tr(\mathbf{A}) tr(\mathbf{B}) + 2 tr(\mathbf{AB}) \\ \mathbf{M}_0 &= (\mathbf{R}_2^{-1} \boldsymbol{r}_2 - \mathbf{R}_1^{-1} \boldsymbol{r}_1)(\boldsymbol{r}_2^t \mathbf{R}_2^{-1} - \boldsymbol{r}_1^t \mathbf{R}_1^{-1}) \mathbf{R}_1 \\ \mathbf{M}_1 &= \mathbf{M} \mathbf{R}_1 \\ c_i &= 1 - \boldsymbol{r}_i^t \mathbf{R}_i^{-1} \boldsymbol{r}_i, \qquad i = 1, 2 \end{aligned}$$

Corollary : ML and CV are quadratic estimators $\Longrightarrow$ we can carry out an exhaustive numerical study of the Risk criterion

Two criteria for the numerical study

Definition : Risk on Target Ratio (RTR)

$$RTR(\boldsymbol{x}_{new}) = \frac{\sqrt{\mathcal{R}_{\hat{\sigma}^2, \boldsymbol{x}_{new}}}}{\mathbb{E}\left[(\hat{Y}_{new} - Y_{new})^2 \right]} = \frac{\sqrt{\mathbb{E}\left[\left(\mathbb{E}\left[\left(\hat{Y}_{new} - Y_{new} \right)^2 \middle| \boldsymbol{y} \right] - \hat{\sigma}^2 c^2_{\boldsymbol{x}_{new}} \right)^2 \right]}}{\mathbb{E}\left[(\hat{Y}_{new} - Y_{new})^2 \right]}$$

Definition : Bias on Target Ratio (BTR)

$$BTR(\boldsymbol{x}_{new}) = \frac{\left| \mathbb{E}\left[(\hat{Y}_{new} - Y_{new})^2 \right] - \mathbb{E}\left(\hat{\sigma}^2 c^2_{\boldsymbol{x}_{new}} \right) \right|}{\mathbb{E}\left[(\hat{Y}_{new} - Y_{new})^2 \right]}$$

Integrated versions over the prediction domain $\mathcal{X}$

$$IRTR = \sqrt{\int_{\mathcal{X}} RTR^2(\boldsymbol{x}_{new}) d\mu(\boldsymbol{x}_{new})}$$

and

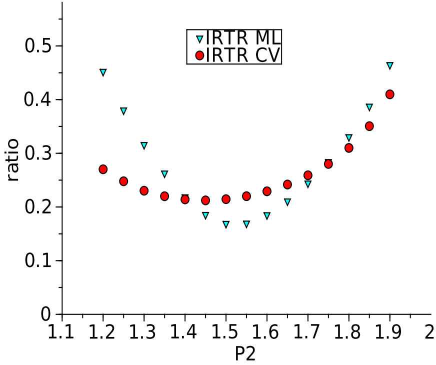
$$IBTR = \sqrt{\int_{\mathcal{X}} BTR^2(\boldsymbol{x}_{new}) d\mu(\boldsymbol{x}_{new})}$$

CV more robust than ML to covariance model misspecification (1/6)

70 observation points on $[0, 1]^5$. Mean over LHS-Maximin samplings.
K_1 and K_2 are power-exponential covariance functions,

$$K_i(x, y) = \exp\left(-\sum_{j=1}^{5}\left(\frac{|x_j - y_j|}{\ell_i}\right)^{p_i}\right),$$

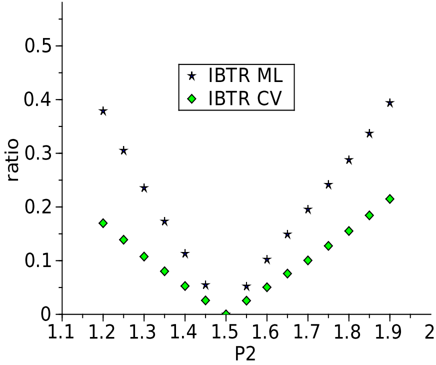
with $\ell_1 = \ell_2 = 1.2$, $p_1 = 1.5$, and p_2 varying.



70 observations on $[0,1]^5$. Mean over LHS-Maximin samplings.
K_1 and K_2 are power-exponential covariance functions,

$$K_i(x,y) = \exp\left(- \sum_{j=1}^{5} \left(\frac{|x_j - y_j|}{\ell_i} \right)^{p_i} \right),$$

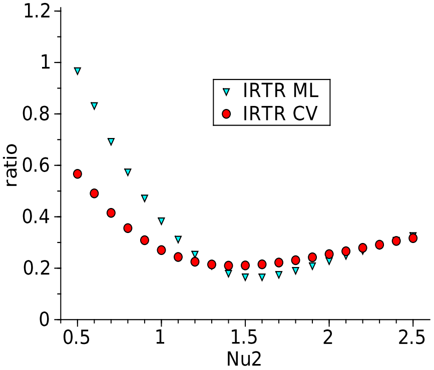
with $\ell_1 = \ell_2 = 1.2$, $p_1 = 1.5$, and p_2 varying.



70 observations on $[0,1]^5$. Mean over LHS-Maximin samplings.
K_1 and K_2 are Matérn covariance functions,

$$K_i(x,y) = \frac{1}{\Gamma(\nu_i)2^{\nu_i-1}} \left(2\sqrt{\nu_i} \frac{||x-y||_2}{\ell_i} \right)^{\nu_i} K_{\nu_i} \left(2\sqrt{\nu_i} \frac{||x-y||_2}{\ell_i} \right),$$

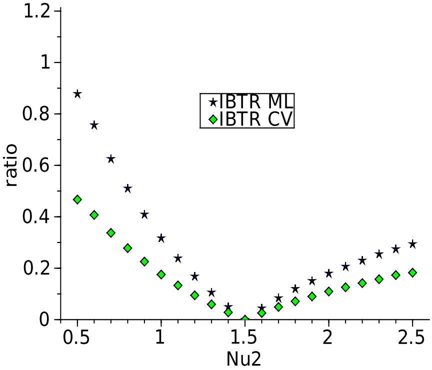
with Γ the Gamma function and K_{ν_i} the modified Bessel function of second order.
We use $\ell_1 = \ell_2 = 1.2$, $\nu_1 = 1.5$, and ν_2 varying.



70 observations on $[0,1]^5$. Mean over LHS-Maximin samplings.
K_1 and K_2 are Matérn covariance functions,

$$K_i(x,y) = \frac{1}{\Gamma(\nu_i)2^{\nu_i - 1}} \left(2\sqrt{\nu_i} \frac{||x-y||_2}{\ell_i} \right)^{\nu_i} K_{\nu_i} \left(2\sqrt{\nu_i} \frac{||x-y||_2}{\ell_i} \right),$$

with Γ the Gamma function and K_{ν_i} the modified Bessel function of second order.
We use $\ell_1 = \ell_2 = 1.2$, $\nu_1 = 1.5$, and ν_2 varying.

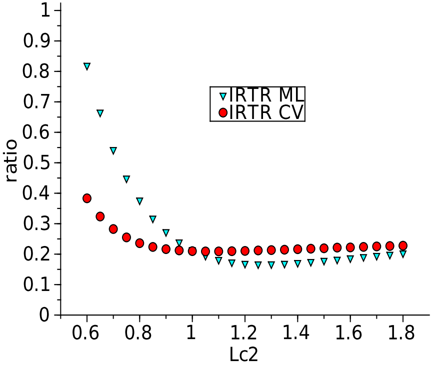


CV more robust than ML to covariance model misspecification (5/6)

70 observations on $[0,1]^5$. Mean over LHS-Maximin samplings.
K_1 and K_2 are Matérn covariance functions,

$$K_i(x,y) = \frac{1}{\Gamma(\nu_i)2^{\nu_i-1}} \left(2\sqrt{\nu_i}\frac{||x-y||_2}{\ell_i}\right)^{\nu_i} K_{\nu_i}\left(2\sqrt{\nu_i}\frac{||x-y||_2}{\ell_i}\right),$$

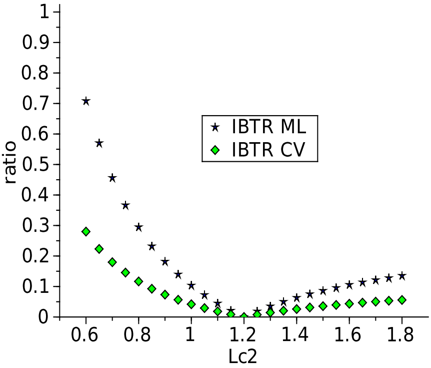
with Γ the Gamma function and K_{ν_i} the modified Bessel function of second order.
We use $\nu_1 = \nu_2 = \frac{3}{2}$, $\ell_1 = 1.2$ and ℓ_2 varying.



70 observations on $[0,1]^5$. Mean over LHS-Maximin samplings.
K_1 and K_2 are Matérn covariance functions,

$$K_i(x,y) = \frac{1}{\Gamma(\nu_i) 2^{\nu_i - 1}} \left(2\sqrt{\nu_i} \frac{||x-y||_2}{\ell_i} \right)^{\nu_i} K_{\nu_i} \left(2\sqrt{\nu_i} \frac{||x-y||_2}{\ell_i} \right),$$

with Γ the Gamma function and K_{ν_i} the modified Bessel function of second order.
We use $\nu_1 = \nu_2 = \frac{3}{2}$, $\ell_1 = 1.2$ and ℓ_2 varying.



Summary of numerical results

For variance hyper-parameter estimation

- For not too regular design of experiments : CV is more robust than ML to misspecification
  - Larger variance but smaller bias for CV
  - The bias term becomes dominant in the model misspecification case
- For regular design of experiments, CV is less robust to model misspecification

For variance and correlation hyper-parameter estimation

- Numerical study on analytical functions
- Confirmation of the results of the variance estimation case

For more details :

Bachoc F, Cross Validation and Maximum Likelihood estimations of hyper-parameters of Gaussian processes with model misspecification, *Computational Statistics and Data Analysis 66 (2013) 55-69, http ://dx.doi.org/10.1016/j.csda.2013.03.016*.

1. Kriging for calibration, improved prediction and metamodeling of computer models
   - Statistical model and method for calibration and improved prediction of computer models
   - Application to the FLICA 4 thermal-hydraulic code
2. Maximum Likelihood and Cross Validation for parametric covariance function estimation
3. Finite sample analysis of ML and CV under model misspecification
4. Asymptotic analysis of ML and CV in the well-specified case
   - Asymptotic framework
   - Consistency and asymptotic normality
   - Analysis of the asymptotic variance matrices
5. Conclusion and perspectives

Framework and objectives

Estimation

We do not make use of the distinction $\sigma^2, \boldsymbol{\theta}$. Hence we use the set $\{K_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ of stationary covariance functions for the estimation.

Well-specified model

The true covariance function K of the Gaussian Process belongs to the set $\{K_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$. Hence

$$K = K_{\boldsymbol{\theta}_0}, \boldsymbol{\theta}_0 \in \Theta$$

Objectives

- Study the consistency and asymptotic distribution of the Cross Validation estimator
- Confirm that, asymptotically, Maximum Likelihood is more efficient
- Study the influence of the spatial sampling on the estimation

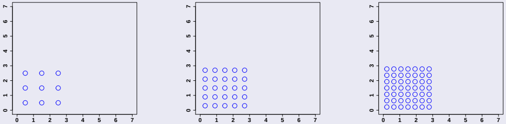
- Spatial sampling : Initial design of experiment for Kriging
- It has been shown that irregular spatial sampling is often an advantage for hyper-parameter estimation
  - Stein M, Interpolation of Spatial Data : Some Theory for Kriging, *Springer, New York, 1999. Ch.6.9*.
  - Zhu Z, Zhang H, Spatial sampling design under the infill asymptotics framework, *Environmetrics 17 (2006) 323-337*.
- Our question : Is irregular sampling always better than regular sampling for hyper-parameter estimation ?

Two asymptotic frameworks for hyper-parameter estimation

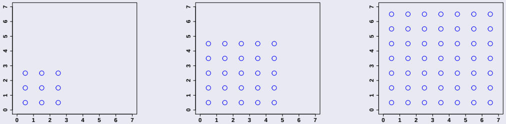
Asymptotics (number of observations $n \to +\infty$) is an area of active research (Maximum-Likelihood estimator)

Two main asymptotic frameworks

- fixed-domain asymptotics : The observations are dense in a bounded domain



- increasing-domain asymptotics : A minimum spacing exists between the observation points $\longrightarrow$ infinite observation domain.



Choice of the asymptotic framework

Comments on the two asymptotic frameworks

- fixed-domain asymptotics
  From 80'-90' and onwards. Fruitful theory

  Stein, M., Interpolation of Spatial Data Some Theory for Kriging, *Springer, New York, 1999.*

  However, when convergence in distribution is proved, the asymptotic distribution does not depend on the spatial sampling $\longrightarrow$ Impossible to compare sampling techniques for estimation in this context
- increasing-domain asymptotics :
  Asymptotic normality proved for Maximum-Likelihood under restricted conditions

  Sweeting, T., Uniform asymptotic normality of the maximum likelihood estimator, *Annals of Statistics 8 (1980) 1375-1381.*

  Mardia K, Marshall R, Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika 71 (1984) 135-146.*

  (no results for CV)

We study increasing-domain asymptotics for ML and CV under irregular sampling

Bachoc F, Asymptotic analysis of the role of spatial sampling for hyper-parameter estimation of Gaussian processes, *Submitted to Journal of Multivariate Analysis, available at http ://arxiv.org/abs/1301.4321.*

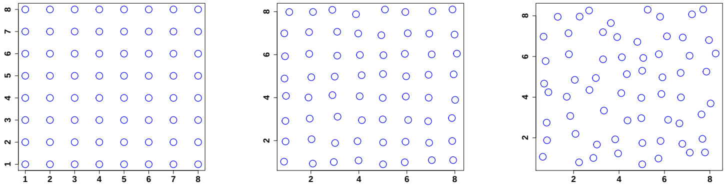
The randomly perturbed regular grid that we study

- Observation point i :

$$\boldsymbol{v}_i + \epsilon X_i$$

  - $(\boldsymbol{v}_i)_{i \in \mathbb{N}^*}$: regular square grid of step one in dimension d
  - $(X_i)_{i \in \mathbb{N}^*}$: *iid* with symmetric distribution on $[-1, 1]^d$
- $\epsilon \in (-\frac{1}{2}, \frac{1}{2})$ is the regularity parameter of the grid.
  - $\epsilon = 0 \longrightarrow$ regular grid.
  - $|\epsilon|$ close to $\frac{1}{2} \longrightarrow$ irregularity is maximal

Illustration with $\epsilon = 0, \frac{1}{8}, \frac{3}{8}$



1. Kriging for calibration, improved prediction and metamodeling of computer models
   - Statistical model and method for calibration and improved prediction of computer models
   - Application to the FLICA 4 thermal-hydraulic code
2. Maximum Likelihood and Cross Validation for parametric covariance function estimation
3. Finite sample analysis of ML and CV under model misspecification
4. Asymptotic analysis of ML and CV in the well-specified case
   - Asymptotic framework
   - Consistency and asymptotic normality
   - Analysis of the asymptotic variance matrices
5. Conclusion and perspectives

Consistency and asymptotic normality

Under general summability, regularity and identifiability conditions, we show

Proposition : for ML

- a.s convergence of the random Fisher information : The random trace $\frac{1}{n} Tr \left(\mathbf{R}_{\boldsymbol{\theta}_0}^{-1} \frac{\partial \mathbf{R}_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}_i} \mathbf{R}_{\boldsymbol{\theta}_0}^{-1} \frac{\partial \mathbf{R}_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}_j} \right)$ converges a.s to the element $(\mathbf{I}_{ML})_{i,j}$ of a $p \times p$ deterministic matrix $\mathbf{I}_{ML}$ as $n \to +\infty$
- asymptotic normality : With $\Sigma_{ML} = 2\mathbf{I}_{ML}^{-1}$

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0 \right) \to \mathcal{N} \left(0, \Sigma_{ML} \right)$$

Proposition : for CV

Same result with more complex formulas for asymptotic covariance matrix Σ_{CV}

$\Sigma_{ML,CV}$ depends only on the regularity parameter ϵ.

$\boxed{\longrightarrow}$ in the sequel, we study the functions $\epsilon \to \Sigma_{ML,CV}$

Main ideas for the proof

- A central tool : because of the minimum distance between observation points : the eigenvalues of the random matrices involved are uniformly lower and upper bounded
- For consistency : bounding from below the difference of M-estimator criteria between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ by the integrated square difference between $K_{\boldsymbol{\theta}}$ and $K_{\boldsymbol{\theta}_0}$
- For almost-sure convergence of random traces : block-diagonal approximation of the random matrices involved and Cauchy criterion
- For asymptotic normality of criterion gradient : almost-sure (with respect to the random perturbations) Lindeberg-Feller Central Limit Theorem
- Conclude with classical M-estimator method

1. Kriging for calibration, improved prediction and metamodeling of computer models
   - Statistical model and method for calibration and improved prediction of computer models
   - Application to the FLICA 4 thermal-hydraulic code
2. Maximum Likelihood and Cross Validation for parametric covariance function estimation
3. Finite sample analysis of ML and CV under model misspecification
4. Asymptotic analysis of ML and CV in the well-specified case
   - Asymptotic framework
   - Consistency and asymptotic normality
   - Analysis of the asymptotic variance matrices
5. Conclusion and perspectives

Objectives

The asymptotic covariance matrix $\Sigma_{ML,CV}$ depend only on the regularity parameter ϵ.
$\boxed{\longrightarrow}$ in the sequel, we study the functions $\epsilon \to \Sigma_{ML,CV}$

Small random perturbations of the regular grid

We study $\left(\frac{\partial^2}{\partial \epsilon^2} \Sigma_{ML,CV} \right)_{\epsilon=0}$
Closed form expression for ML for $d = 1$ using Toeplitz matrix sequence theory

Large random perturbations of the regular grid

We study $\epsilon \to \Sigma_{ML,CV}$
Closed form expression for ML and CV for $d = 1$ and $\epsilon = 0$ using Toeplitz matrix sequence theory

Matèrn model in dimension one

$$K_{\ell,\nu}(x,y) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(2\sqrt{\nu}\frac{|x-y|}{\ell} \right)^{\nu} K_{\nu} \left(2\sqrt{\nu}\frac{|x-y|}{\ell} \right),$$

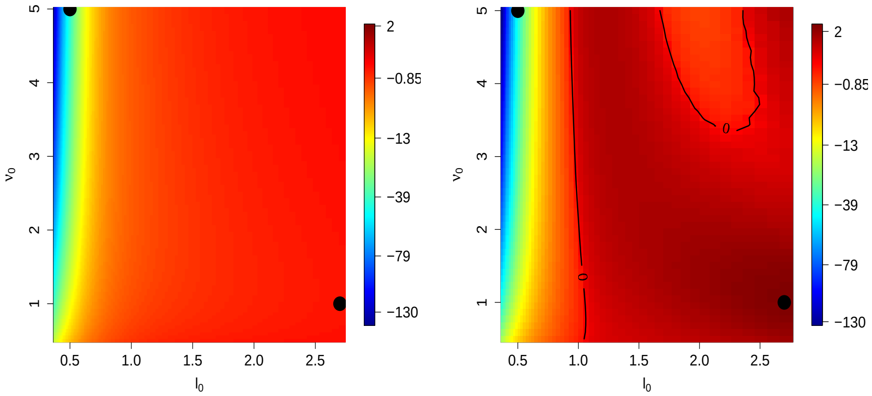
with Γ the Gamma function and K_ν the modified Bessel function of second order

We consider

- The estimation of ℓ when ν_0 is known
- The estimation of ν when ℓ_0 is known

$\Longrightarrow$ We study scalar asymptotic variances

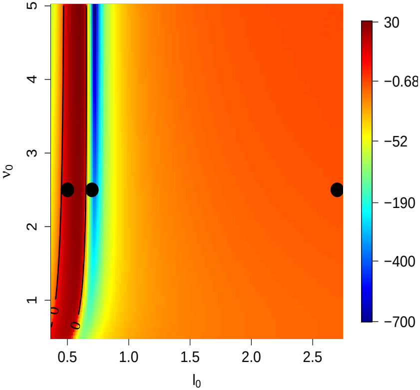
Estimation of ℓ when ν_0 is known.
Level plot of $(\partial_\epsilon^2 \Sigma_{ML,CV})/\Sigma_{ML,CV}$ in $\ell_0 \times \nu_0$ for ML (left) and CV (right)

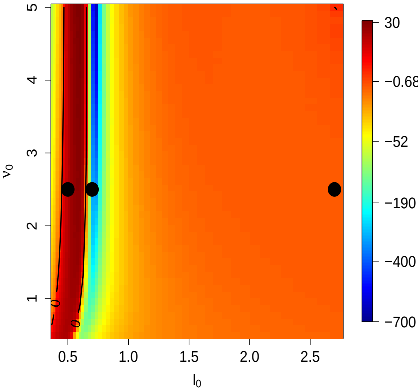


Small perturbations are always beneficial for ML. They can however deteriorate the CV estimation

Small random perturbations of the regular grid (2/2)

Estimation of ν when ℓ_0 is known.
Level plot of $(\partial_\epsilon^2 \Sigma_{ML,CV})/\Sigma_{ML,CV}$ in $\ell_0 \times \nu_0$ for ML (left) and CV (right)

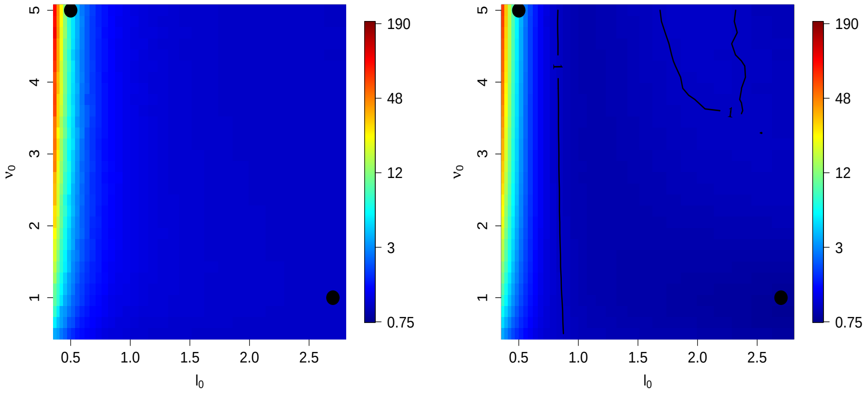




There exist cases of degradation of the estimation for small perturbations for ML around $\ell_0 \approx 0.5$.
Because the Matérn correlation function at $t = 0.7$ is almost independent of ν for $\ell_0 \approx 0.5$.

Large random perturbations of the regular grid (1/2)

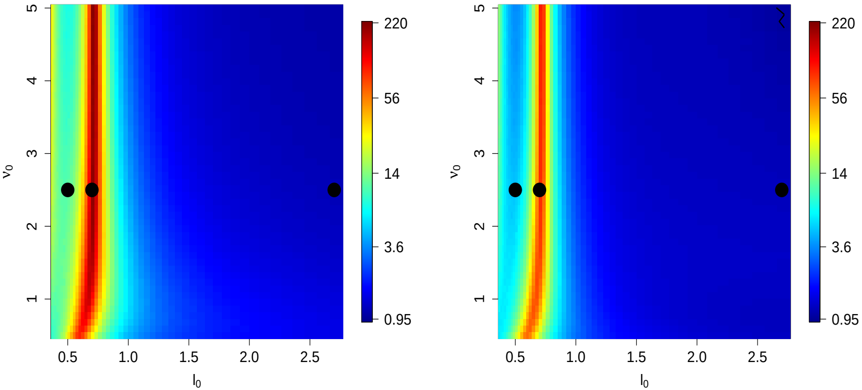
Estimation of ℓ when ν_0 is known.
Level plot of $\left[\Sigma_{ML,CV}(\epsilon=0)\right] / \left[\Sigma_{ML,CV}(\epsilon=0.45)\right]$ in $\ell_0 \times \nu_0$ for ML (left) and CV (right)



Strong perturbations are always beneficial for ML

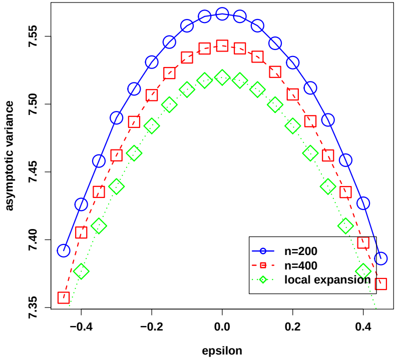
Large random perturbations of the regular grid (2/2)

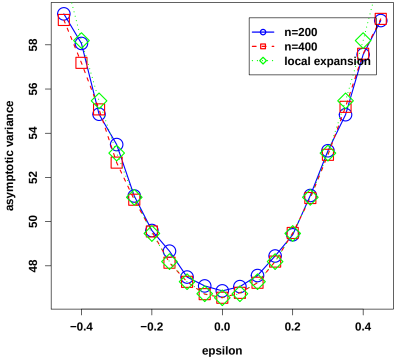
Estimation of ν when ℓ_0 is known.
Level plot of $\left[\Sigma_{ML,CV}(\epsilon = 0)\right] / \left[\Sigma_{ML,CV}(\epsilon = 0.45)\right]$ in $\ell_0 \times \nu_0$ for ML (left) and CV (right)



Strong perturbations are always beneficial for ML and CV

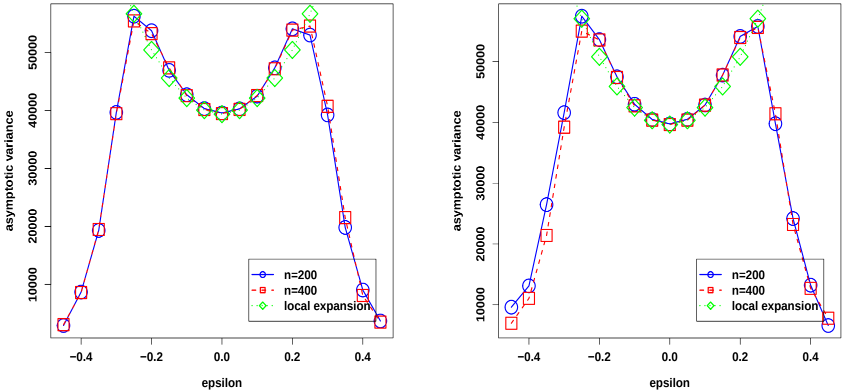
Estimation of ℓ when ν_0 is known, for $\ell_0 = 2.7$, $\nu_0 = 1$.
Plot of $\epsilon \to \Sigma_{ML,CV}$ for ML (left) and CV (right)





The asymptotic variance of CV is significantly larger than that of ML

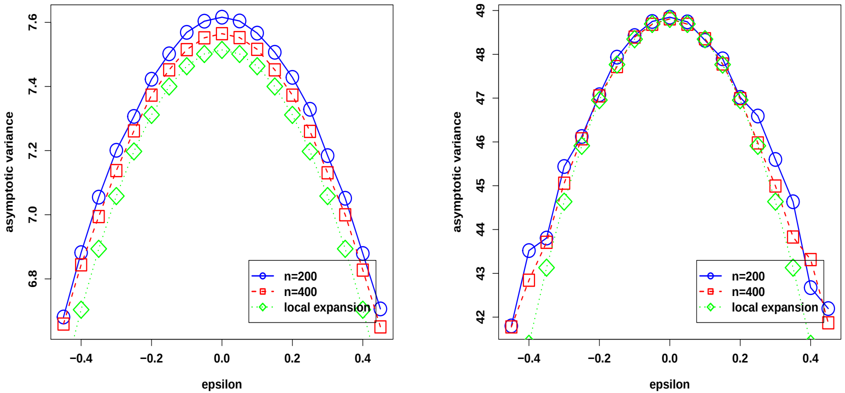
Estimation of ν when ℓ_0 is known, for $\ell_0 = 0.5$, $\nu_0 = 2.5$.
Plot of $\epsilon \rightarrow \Sigma_{ML,CV}$ for ML (left) and CV (right)



The asymptotic variances are increasing with $|\epsilon|$ for $|\epsilon| < 0.2$ (particularity of the Matérn model).
For $|\epsilon| > 0.2$ the asymptotic variances are strongly decreasing functions of $|\epsilon|$

Some particular functions $\epsilon \to \Sigma_{ML,CV}$ (3/3)

Estimation of ν when ℓ_0 is known, for $\ell_0 = 2.7$, $\nu_0 = 2.5$.
Plot of $\epsilon \to \Sigma_{ML,CV}$ for ML (left) and CV (right)



The asymptotic variance of CV is significantly larger than that of ML

Conclusion on the well-specified case

- CV is consistent and has the same rate of convergence than ML
- We confirm that ML is more efficient
- Strong irregularity in the sampling is an advantage for covariance function estimation
  - With ML, irregular sampling is more often an advantage than with CV
  - We show that, however, regular sampling is better for prediction with known covariance function $\Longrightarrow$ motivation for using space-filling samplings augmented with some clustered observation points

Z. Zhu and H. Zhang, Spatial Sampling Design Under the Infill Asymptotics Framework, *Environmetrics 17 (2006) 323-337*.

L. Pronzato and W. G. Müller, Design of computer experiments : space filling and beyond, *Statistics and Computing 22 (2012) 681-701*.

For further details :

Bachoc F, Asymptotic analysis of the role of spatial sampling for hyper-parameter estimation of Gaussian processes, *Submitted to Journal of Multivariate Analysis, available at http ://arxiv.org/abs/1301.4321*.

Conclusion and perspectives on covariance function estimation

General conclusion

- ML preferable to CV in the well-specified case
- In the misspecified-case, with not too regular design of experiments : CV is preferable because of its smaller bias
- In both misspecified and well-specified cases : the estimation benefits from an irregular sampling
- The variance of CV is larger than that of ML in all the cases studied.

Perspectives

- Designing other CV procedures (LOO error weighting, decorrelation and penalty term) to reduce the variance
- Expansion-domain asymptotic analysis of the misspecified case
- Start studying the fixed-domain asymptotics of CV, in the particular cases where it is done for ML

Thank you for your attention !