Speaker: Bertrand Iooss

Part of M120 - Surrogate Modelling and Data-Driven Approaches for UQ

Gaussian process regression: new hyperparameter estimation algorithm for more reliable prediction

Bertrand Iooss
EDF R&D, 6 quai Watier, 78400, Chatou, France, bertrand.iooss@edf.fr
Amandine Marrel
CEA, DES, IRESNE, DER, Cadarache F-13108 Saint-Paul-Lez-Durance, amandine.marrel@cea.fr

In the framework of the emulation of CPU-time expensive numerical simulators with Gaussian process (GP) regression, we propose in this work a new algorithm for the estimation of GP covariance parameters, referred to as GP hyperparameters. The objective is twofold: to ensure a GP as predictive as possible w.r.t. to the output of interest, but also with reliable prediction intervals, i.e. representative of GP prediction error. To achieve this, we propose a new constrained multi-objective algorithm for the hyperparameter estimation. It jointly maximizes the likelihood of the observations as well as the empirical coverage function of GP prediction intervals, under the constraint of not degrading the GP predictivity. Cross validation techniques and advantageous update GP formulas are notably used. The benefit brought by the algorithm compared to standard algorithms is illustrated on a large benchmark of analytical functions (with dimensions from 1 to 20 input variables). Different designs of experiments and different covariance models are considered. An application on a real data test case modeling an aquatic ecosystem is also proposed: GP metamodeling within a log-kriging approach is used to predict the biomass of a species at a given time. The multi-objective algorithm performs better than standard algorithms and this particular metamodeling framework shows the crucial interest of well-estimated and reliable prediction variances in GP regression.

Gaussian process regression: new hyperparameter estimation algorithm for more reliable prediction

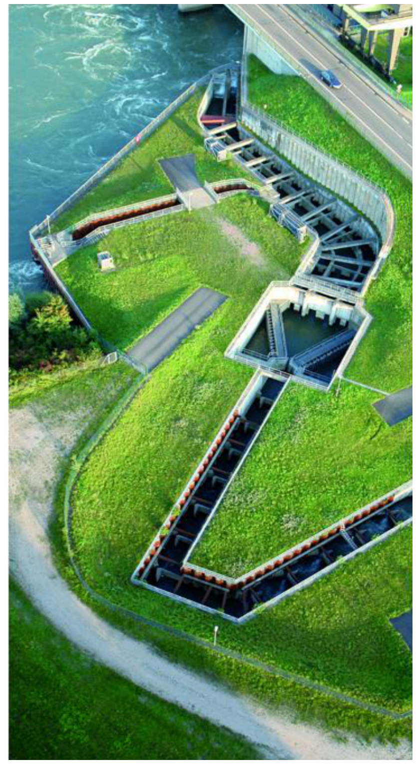
Application to an aquatic ecosystem model

Bertrand IOOSS - *EDF R&D, Chatou, France*

Amandine Marrel - *CEA/DES/IRESNE, Cadarache, France*

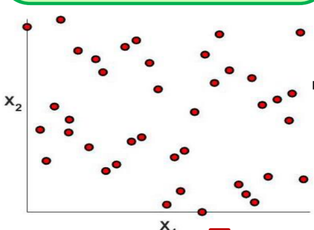
2024 SIAM Conference on Uncertainty Quantification

February, 28th, 2024

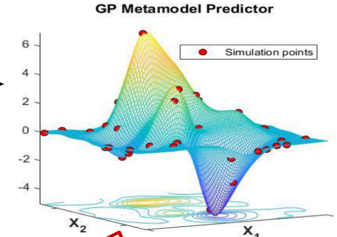


Use of models and metamodels in UQ

Design of numerical experiments (e.g. Monte Carlo) → **Numerical simulations** → Analysis of simulator outputs



→ Simulator $Y = \mathcal{M}(X_1, \ldots, X_d)$ →

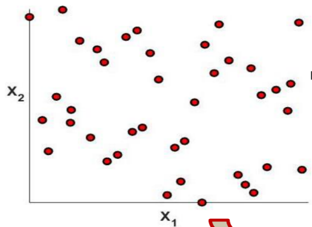


In case of costly $\mathcal{M}$:
Approximation with Machine Learning

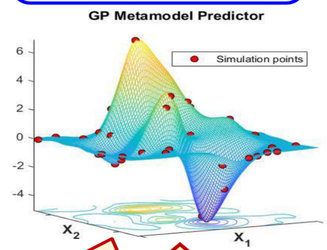
Metamodel: $Y_{app}(X) \approx \mathcal{M}(X)$

Use of models and metamodels in UQ

Design of numerical experiments → Numerical simulations → Analysis of simulator outputs



→ Simulator $Y = \mathcal{M}(X_1, \ldots, X_d)$ →



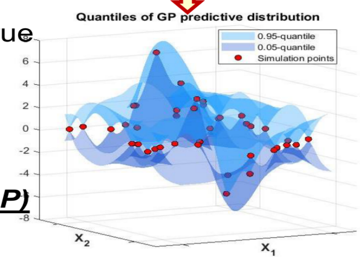
In case of costly $\mathcal{M}$:
Approximation with Machine Learning

Metamodel: $Y_{app}(X) \approx \mathcal{M}(X)$

Probabilistic metamodel

- Build from the dataset a predictor that mimics the true model $\mathcal{M}$, with good prediction capabilities (*validation*)
- Bonus: with reliable quantification of prediction error

Ex : Polynomials, random forests, …, ***Gaussian process (GP)***



Reminders on GP metamodel

► **Only a *n*-sample of simulations is available** (Monte-Carlo, LHS, space-filling…)

$$\boldsymbol{D_S} = \left(\boldsymbol{X_S^{(j)}}, \boldsymbol{Y_S^{(j)}}\right)_{1 \le j \le n} \text{ where } Y_S^{(j)} = \mathcal{M}\left(X_S^{(j)}\right)$$

► **Probabilistic surrogate model** : response is considered as a realization of a random GP *[RW06]*:

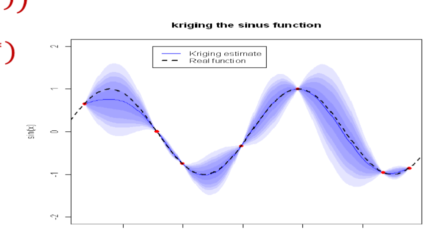
$Y(\boldsymbol{x}) \sim GP\left(\mu(\boldsymbol{x}), k(\boldsymbol{x}', \boldsymbol{x})\right)$ with $\mu(\boldsymbol{x})$ the mean and $k(\boldsymbol{x}', \boldsymbol{x})$ the covariance function.

⇒The predictive GP is the GP conditioned by the observations (X_S, Y_S) :

$$Y(\boldsymbol{x}^*)_{|Y(X_S)=Y_S} \sim GP\left(\hat{\mu}(\boldsymbol{x}^*), \hat{s}(\boldsymbol{x}', \boldsymbol{x}^*)\right)$$

With analytical formulations for $\hat{\mu}(\boldsymbol{x}^*)$ *and* $\hat{s}(\boldsymbol{x}', \boldsymbol{x}^*)$

⇒ Conditional mean $\hat{\mu}(\boldsymbol{x}^*)$ serves as the predictor at location $\boldsymbol{x}^*$



⇒ Prediction variance (*i.e.* mean squared error) is given by conditional covariance $\hat{s}(\boldsymbol{x}^*, \boldsymbol{x}^*)$

⇒ Prediction interval of any level α can be built at any location $\boldsymbol{x}^*$

Reminders on GP metamodel

In practice: parametric choices for trend function μ and covariance function k

$$Y(\boldsymbol{x}) \sim GP(\mu(\boldsymbol{x}), k(\boldsymbol{x}', \boldsymbol{x}))$$

⇒ For μ: either constant or linear basis

⇒ For k: tensorized 1-D covariance functions of ν-Matérn (with $h = |x - \tilde{x}|$)

$$k_{\sigma,\nu,\theta}(x, \tilde{x}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}h}{\theta}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}h}{\theta}\right)$$

⇒ Need to estimate from the dataset the correlation hyperparameters $\boldsymbol{\theta} \in \mathbb{R}^{+,d}$

➡ How to ensure that the estimated hyperparameters $\boldsymbol{\theta}$ yield good predictivity but also **reliable GP prediction intervals**?

⇒ Crucial for safety applications: a « reliable » confidence interval is required on an estimated quantity of interest related to safety (as a high-order quantile)

See application in nuclear safety in [IL19,ILG19]

➡ Especially in large dimension (d > 10) and small dataset ($n \approx 100$)

GP estimation and validation

We note $y_i = y(\boldsymbol{x}_i)$
$\hat{y}_i = \hat{\mu}(\boldsymbol{x}_i)$

► **Usual estimation methods** *[KO23,Mur21,Pet22]*

→ Maximum likelihood-based estimation (MLE) ⇔ minimization of NLL

→ Cross-validation-based estimation: minimization of $\text{RMSE}=\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right\}^{0.5}$

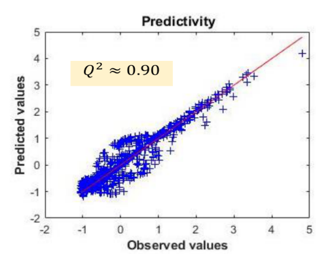
→ Bayesian approaches (CPU ++)

► **Validation criteria computed by cross-validation (LOO)**

→ Accuracy of the GP predictor: $Q^2 = 1 - \text{RMSE}^2/\text{Var}(Y)$

→ Accuracy of the predictive variance:

$$\text{PVA} = \left|\log \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{\hat{s}_{-i}^2}\right|$$



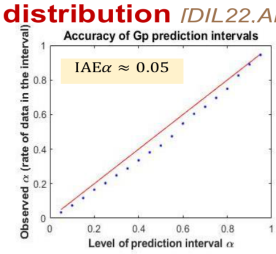
→ **Accuracy of the whole GP conditional distribution** *[DIL22,ABG23]*

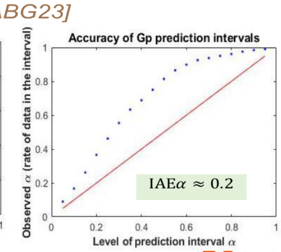
Empirical coverage function for $\alpha \in [0,1]$:

$$\hat{\Delta}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{y_i \in \mathcal{PI}_{\alpha,-i}(\mathbf{x}_i)\}$$

⇒ **Integrated Absolute Error on $\hat{\Delta}(\alpha)$**

$$\text{IAE}\alpha = \int_0^1 |\hat{\Delta}(\alpha) - \alpha| \, d\alpha$$





Study of criteria NLL, Q^2, PVA and $IAE\alpha$ on many test cases

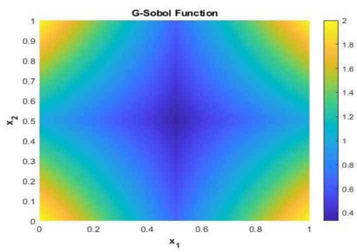
→ **Close behavior of NLL and Q^2 ⇒ keep NLL as the main estimation objective** to ensure the predictivity of the metamodel ⇒ Consistent with *[PBF+23,Pet22]*

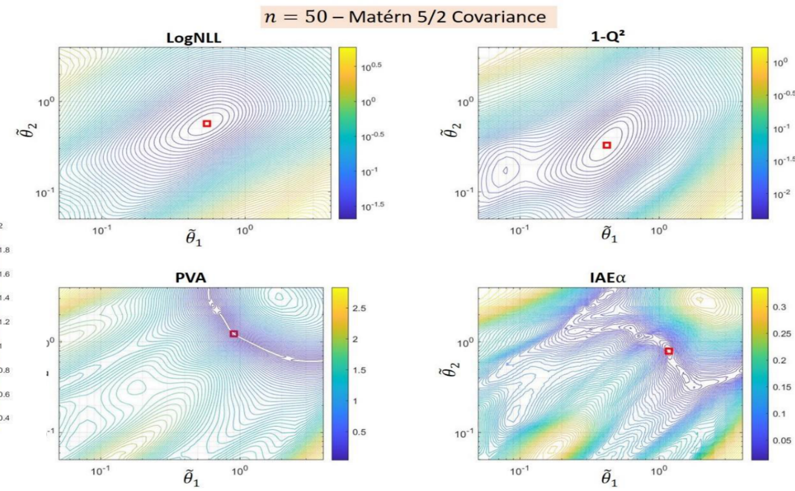
→ **Similar behavior of PVA and IAEα but more irregular w.r.t. $\boldsymbol{\theta}$**

⇒ Some local minima compatible with optimal values of the other criteria

⇒ <u>No to be optimized independently of the others</u>

Illustration with a test on a 2D G-Sobol fct





New estimation algorithm for GP hyperparameters

<u>From these understandings:</u>

→ **Close behavior of NLL and Q^2 ⇒ keep NLL as the main estimation objective** to ensure the predictivity of the metamodel

→ **IAEα more directly related to reliable predictive intervals, than PVA**

→ **In the neighborhood of the optimal NLL point, existence of better points $\boldsymbol{\theta}$ w.r.t IAEα**, but need to control the possible degradation of Q^2 value, which guarantees the predictivity

<u>We propose the following algorithm:</u>

⇒ **Optimization based on NLL and $IAE\alpha$ + Control of Q^2**
($IAE\alpha$ and Q^2 estimated by cross validation)

⇒ **Proposition of a multi-objective NSGA-II algorithm with constraint on Q^2**

EDF

Algorithm flowchart

Step 1: Initial MLE-based estimation of $\tilde{\theta}$ with standard algorithm

Estimation of hyperparameters by **minimizing the NLL** with a **multistart BFGS algorithm**

→ $\tilde{\boldsymbol{\theta}}_{MLE}^{init}$

→ Computation of LOO estimator of Q^2 of the GP built-upon $\tilde{\boldsymbol{\theta}}_{MLE}^{init}$: → $\hat{Q}_{LOO}^{2,init}$

Step 2: Estimation of $\tilde{\theta}$ with constrained multi-objective algorithm

- **2 objectives** to be minimized: **NLL and IAEα** (IAEα computed by LOO)
- **1 constraint** on LOO estimator of Q^2: $\hat{Q}_{LOO}^{2,new} \geq \hat{Q}_{LOO}^{2,init} - \gamma$ with $\gamma = 0.05$, e.g.
- Optimization performed by **NSGA-II algorithm, with initial population based on $\tilde{\boldsymbol{\theta}}_{MLE}^{init}$**

→ Computation of LOO estimator of Q^2 for each candidate $\tilde{\boldsymbol{\theta}}_{NGSA}^{new}$: → $\hat{Q}_{LOO}^{2,new}$

Pareto front of n_{Pareto} non-dominated solutions $\left(\tilde{\boldsymbol{\theta}}_{NGSA,1}^{new}, \dots, \tilde{\boldsymbol{\theta}}_{NGSA,n_{pareto}}^{new}\right)$

→ Each Pareto solution verifies $\hat{Q}_{LOO,i}^{2,new} \geq \hat{Q}_{LOO}^{2,init} - \gamma$ with Q^2 computed for $\tilde{\boldsymbol{\theta}}_{NGSA,i}^{new}$, for $i = 1, \dots, n_{pareto}$

Step 3: Selection of the "best" solution in Pareto front

3.1 Clustering of Pareto front set of solutions according to the values of both objectives (NLL and IAEα)
→ Number of clusters k optimized with **Elbow method**

3.2 Selection of the cluster C1 with optimal NLL (cluster with lower values of NLL)

3.3 Selection of the "best" solution in C1: solution of C1 with optimal (i.e. lowest) value of IAEα

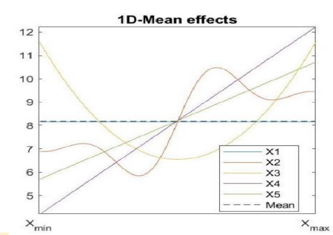
→ $\tilde{\boldsymbol{\theta}}_{NSGA}^{BestC1}$

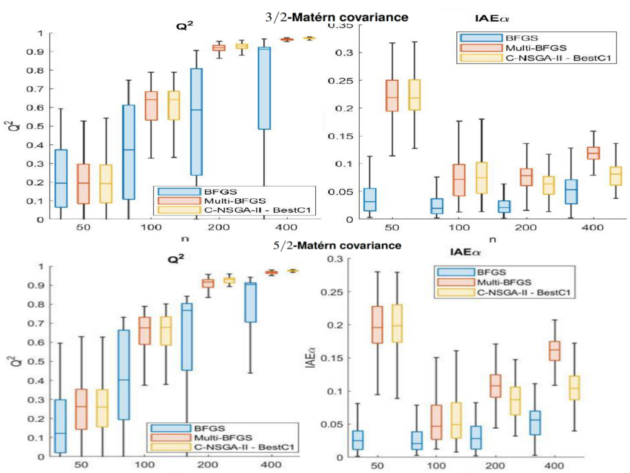
Intensive benchmark on test functions

d = 2 to 20, ≠ covariance, ≠ sample sizes, ≠ DoE, with/without nugget effect
Comparison with usual algorithms based on NLL optimiz. only (BFGS/multistart)

Example: Marrel-d20 function

$$Y = a_1 \sin[6\pi X_1^{\frac{5}{2}} \left(X_2 - \frac{1}{2}\right) + a_2 \left(X_3 - \frac{1}{2}\right)^2 + a_3 X_4 + a_4 X_5 + r_{X_{\dots}}$$





(no nugget)

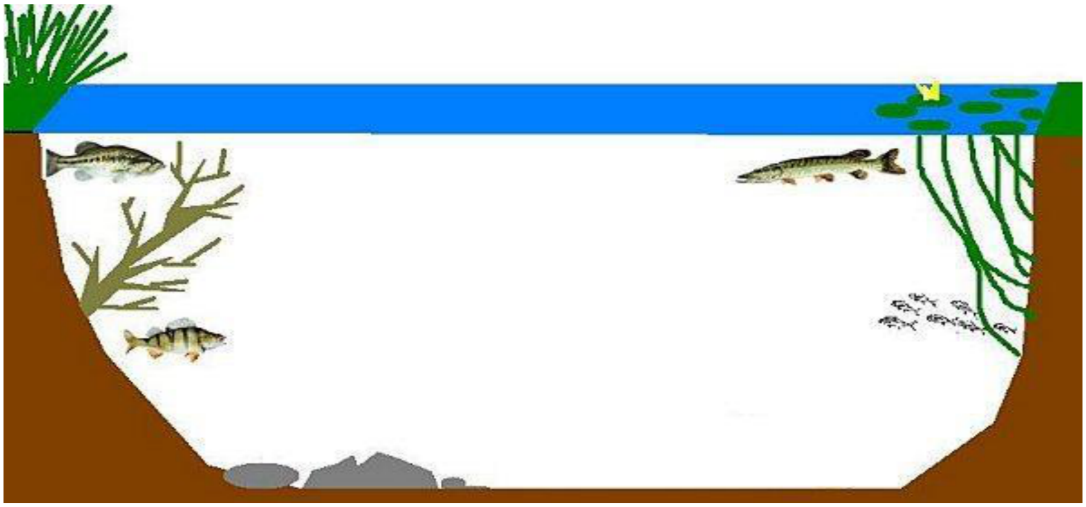
⟹ **Predictivity with Constr-NSGA-II algorithm at least as good as the simple NLL optimization**

⟹ **Improvement of *IAEα* especially if** :

- **The model is misspecified, i.e. if the covariance does not match the regularity of the function**
- **When the number of hyperparameters is large** (case of large dimension *d* + tensorized anisotropic stationary covariance)

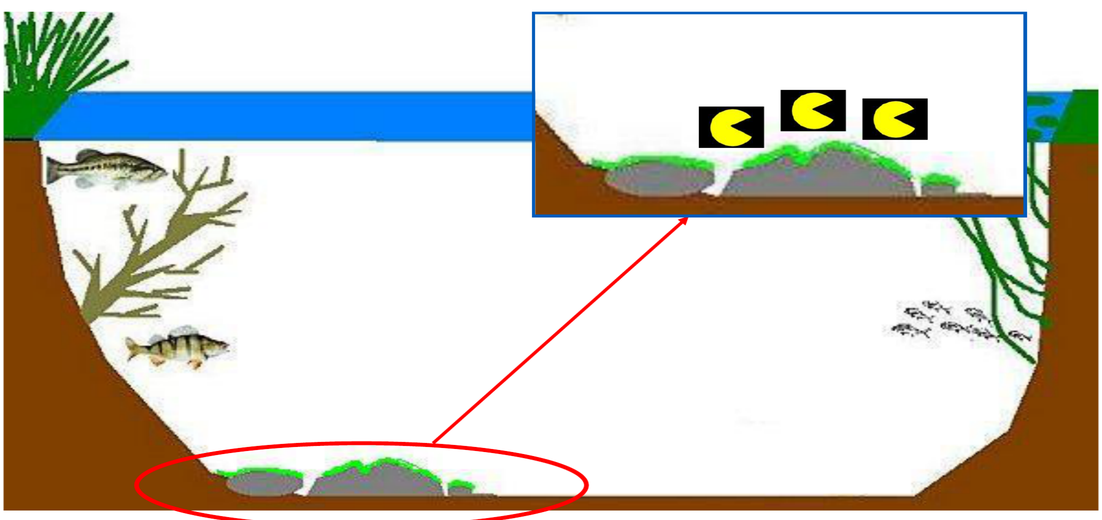
Application: aquatic prey-predator chain model

Studies of biological contamination of rivers



Application: aquatic prey-predator chain model

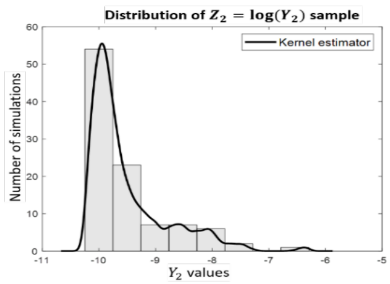
EDO-type equations describing the growth of microorganisms, periphyton, grazing and prey-predator interactions



Application to an aquatic ecosystem model

- MELODY model: prey-predator chain in an aquatic ecosystem *[IPB+12]*
- ***d* = 20** uncertain inputs:
  - Periphyton: photosynthesis/mortality/excretion rates, survival temperature, saturation constants, …
  - Grazers: consumption/assimilation/mortality/excretion rates, survival temperature, …
- **2 outputs of interest**: Periphyton (Y_1) and Grazers (Y_2) biomasses at day n°60
- Sample of ***n* = 100** simulations of the model MELODY (space-filling design)

- Need of **preliminar logarithmic transformation**



⇒ **Lognormal-kriging modeling:**

- Emulation of $Z_i = \log(Y_i)$ with GP regression
- Lognormal-kriging back-transformations to obtain metamodel for Y_i

$$\hat{y}_i(\mathbf{x}) = e^{\left(\hat{z}_i(\mathbf{x}) + 0.5\hat{s}^2_{z_i}(\mathbf{x})\right)}$$

$$\hat{s}^2_Y(\mathbf{x}) = \left(e^{\hat{s}^2_{z_i}(\mathbf{x})} - 1\right) e^{\left(2\hat{z}_i(\mathbf{x}) + \hat{s}^2_{z_i}(\mathbf{x})\right)}$$

Results

Additional comparison with **Bayesian RobustGaSP approach** *[GWB18]*

⇒ **With** nugget effect (included in the set of GP hyperparameters to be estimated)

Data	Covariance	Predictivity Coefficient Q^2			IAEα		
		Multi-BFGS	C-NSGA-II-BestC1	RobustGaSP	Multi BFGS	C-NSGA-II-BestC1	RobustGaSP
Y_2	Matern3/2	0,70	0,74	0,25	0,10	0,07	0,04
	Matern5/2	0,77	**0,82**	0,66	0,09	**0,02**	0,07
	Gaussian	0,75	**0,79**	0,66	0,08	**0,02**	0,06

⇒ Best results with **Constr-NSGA-II algorithm**: **better Q^2 and IAEα**

⇒ **Without** nugget effect

Data	Covariance	Predictivity Coefficient Q^2			IAEα		
		Multi-BFGS	C-NSGA-II-BestC1	RobustGaSP	Multi BFGS	C-NSGA-II-BestC1	RobustGaSP
Y_2	Matern3/2	0,70	**0,75**	0,47	0,10	0,06	0,03
	Matern5/2	0,78	**0,84**	**0,83**	0,08	**0,02**	0,07
	Gaussian	0,70	**0,72**	**0,89**	0,06	**0,03**	0,06

⇒ **Better behavior of RobustGasp without nugget : best Q^2 but not IAEα**

⇒ **Constr-NSGA-II algorithm is also more robust to modeling choices**

Conclusions & Perspectives

- High benefits of considering validation criteria of the **whole GP distribution** when estimating hyperparameters ⇒ enables **more robust estimation !**

- Particular attention must be paid to **GP validation**

- Part of a more general effort to ensure confidence in machine learning models

- Perspectives: some are methodological but the main ones concern software & industrial deployment of GP!

eDF

Thank you!

Funding by SAMOURAI ANR project

anr agence nationale de la recherche SAMOURAI

Simulation Analytics and Meta-model-based solutions for Optimization, Uncertainty and Reliability Analysis

eDF SAFRAN ifp Energies nouvelles cea MINES Saint-Étienne CentraleSupélec GERAD

Reference of this work: A.Marrel and B. Iooss, *Probabilistic surrogate modeling by Gaussian process: A new estimation algorithm for more robust prediction*, Preprint https://hal.science/cea-04322818

See also: A. Marrel and B. Iooss, *Probabilistic surrogate modeling by Gaussian process: A review on recent insights in estimation and validation*, Preprint https://hal.science/cea-04322810

and  's talk "Hidden but essential recipes for successful GP metamodeling to support UQ in numerical simulation" on Friday, MS 194, 10:00AM

eDF

References

[RW06] Rasmussen & Williams, *Gaussian processes for machine learning*, MIT Press, 2006
[IM19] Iooss & Marrel, Advanced methodology for uncertainty propagation in computer experiments with large number of inputs, *Nuclear Technology*, 205:1588-1606, 2019
[ILG19] Iooss & Le Gratiet, Uncertainty and sensitivity analysis of functional risk curves based on Gaussian processes, *Reliability Engineering and System Safety*, 187:58-66, 2019
[KO23] Karvonen & Oates, Maximum Likelihood Estimation in GP is ill-posed, *Journal of Machine Learning research*, 24:1-47, 2023
[Mur21] Muré, Propriety of the reference posterior GP distribution, *The Annals of Statistics*, 49:2356-2377, 2021
[Pet22] Petit, *Improved Gaussian process modeling. Application to Bayesian optimization*, PhD Thesis, University Paris-Saclay, 2022
[DIL22] Demay, Iooss, Le Gratiet & Marrel, Model selection for Gaussian Process regression: an application with highlights on the predictive variance, *Quality and Reliability Engineering International Journal*, 38:1482-1500, 2022
[ABG23] Acharki, Bertoncello, & Garnier, Robust prediction interval estimation for GP by cross-validation method, *Computational Statistics Data Analysis*, 178:107597, 2023
[PBF+23] Petit, Bect, Feliot, & Vazquez, Parameter selection in GP interpolation: an empirical study of selection criteria, *SIAM/ASA Journal of Uncertainty Quantification*, 11, 2023
[IPB+12] Iooss, Popelin, Blatman, Ciric, Gamboa, Lacaze & Lamboni, Some new insights in derivative-based global sensitivity measures, *ESREL 2012 Conference*, ESRA, Helsinki, Finland, June 2012
[GWB18] Gu, Wang, and Berger, Robust Gaussian stochastic process emulation, *The Annals of Statistics*, 46(6A):3038 – 3066, 2018