



HAL
open science

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

Bertrand Iooss, Amandine Marrel

► **To cite this version:**

Bertrand Iooss, Amandine Marrel. Gaussian process regression: new hyperparameter estimation algorithm for more reliable prediction. SIAM Conference on Uncertainty Quantification (UQ24), Feb 2024, Trieste, Italy. cea-04506405

HAL Id: cea-04506405

<https://cea.hal.science/cea-04506405>

Submitted on 15 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

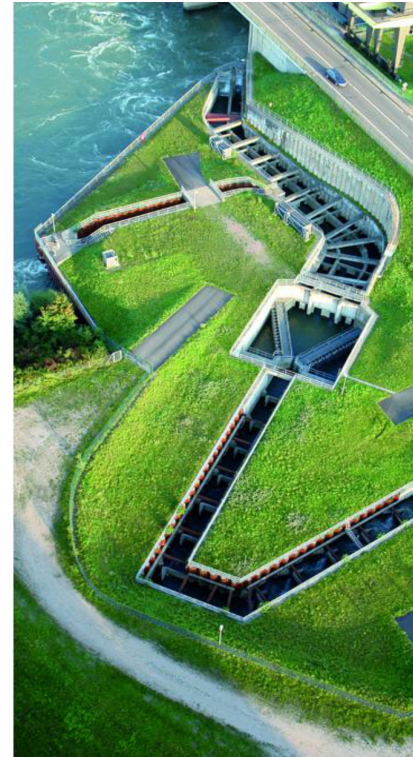
Application to an aquatic ecosystem model

Bertrand IOOSS - EDF R&D, Chatou, France

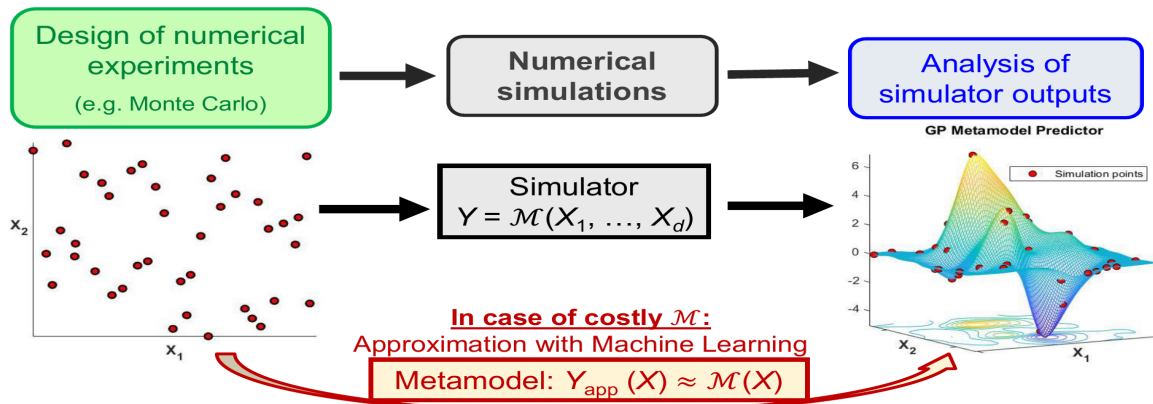
Amandine Marrel - CEA/DES/IRENE, Cadarache, France

2024 SIAM Conference on Uncertainty
Quantification

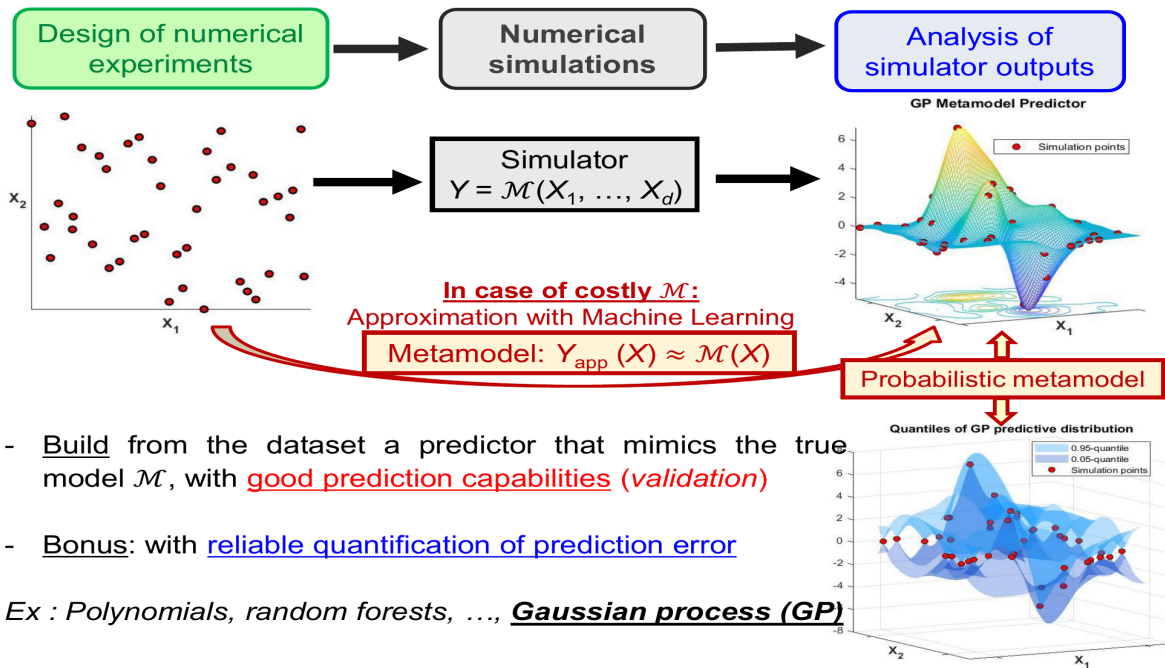
February, 28th, 2024



Use of models and metamodels in UQ



Use of models and metamodels in UQ



Reminders on GP metamodel

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

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

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

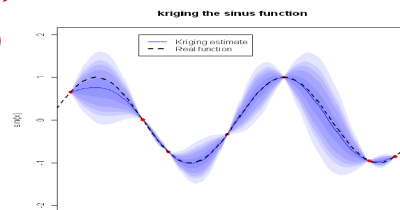
$$Y(\mathbf{x}) \sim GP(\mu(\mathbf{x}), k(\mathbf{x}', \mathbf{x})) \quad \text{with } \mu(\mathbf{x}) \text{ the mean and } k(\mathbf{x}', \mathbf{x}) \text{ the covariance function.}$$

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

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

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

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



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

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



Reminders on GP metamodel

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

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

⇒ 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 $\theta \in \mathbb{R}^{+,d}$

→ How to ensure that the estimated hyperparameters θ 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 $y_i = y(x_i)$
 $\hat{y}_i = \hat{\mu}(x_i)$

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

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

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

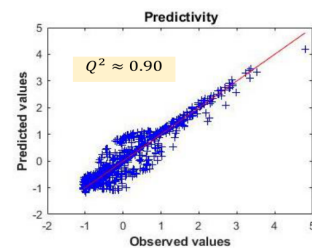
→ Bayesian approaches (CPU ++)

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

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

→ Accuracy of the predictive variance:

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



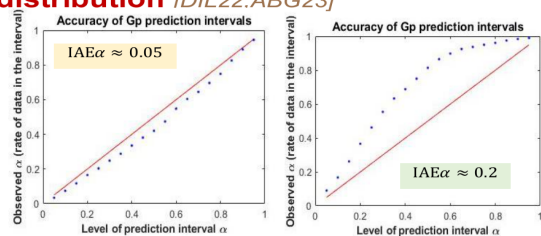
→ **Accuracy of the whole GP conditional distribution** [DIL22, ABG23]

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

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

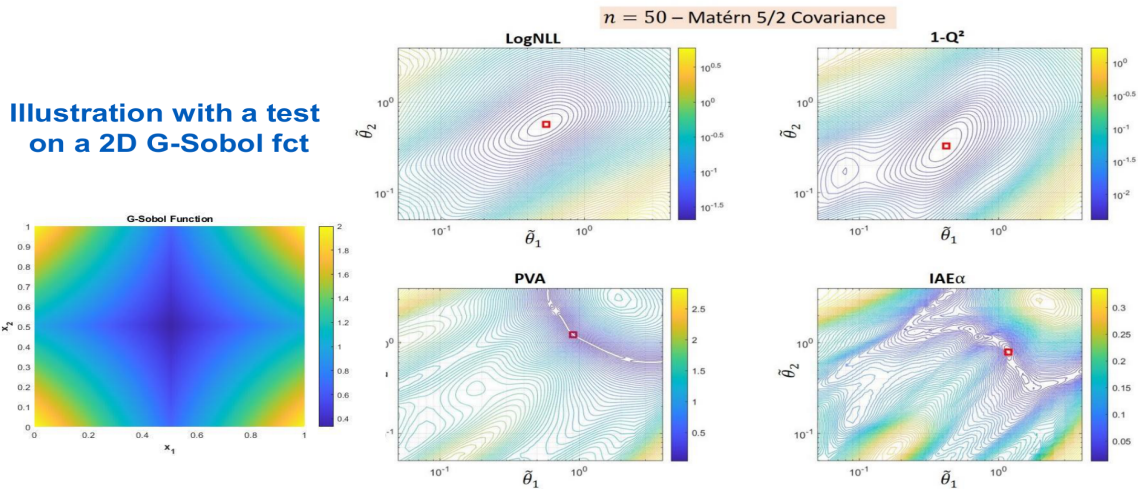
⇒ **Integrated Absolute Error on $\hat{\Delta}(\alpha)$**

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



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

- Close behavior of NLL and $Q^2 \Rightarrow$ keep NLL as the main estimation objective to ensure the predictivity of the metamodel \Rightarrow Consistent with [PBF*23,Pet22]
- Similar behavior of PVA and $IAE\alpha$ but more irregular w.r.t. θ
 - \Rightarrow Some local minima compatible with optimal values of the other criteria
 - \Rightarrow No to be optimized independently of the others



New estimation algorithm for GP hyperparameters

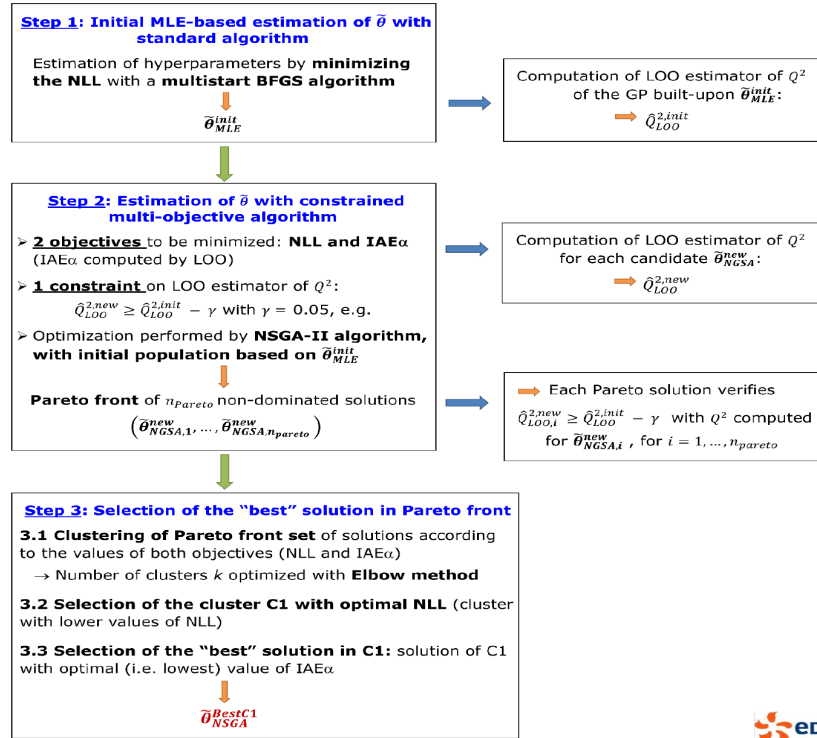
From these understandings:

- Close behavior of NLL and $Q^2 \Rightarrow$ keep NLL as the main estimation objective to ensure the predictivity of the metamodel
- $IAE\alpha$ more directly related to reliable predictive intervals, than PVA
- In the neighborhood of the optimal NLL point, existence of better points θ w.r.t $IAE\alpha$, but need to control the possible degradation of Q^2 value, which guarantees the predictivity

We propose the following algorithm:

- \Rightarrow Optimization based on NLL and $IAE\alpha$ + Control of Q^2
($IAE\alpha$ and Q^2 estimated by cross validation)
- \Rightarrow Proposition of a multi-objective NSGA-II algorithm with constraint on Q^2

Algorithm flowchart

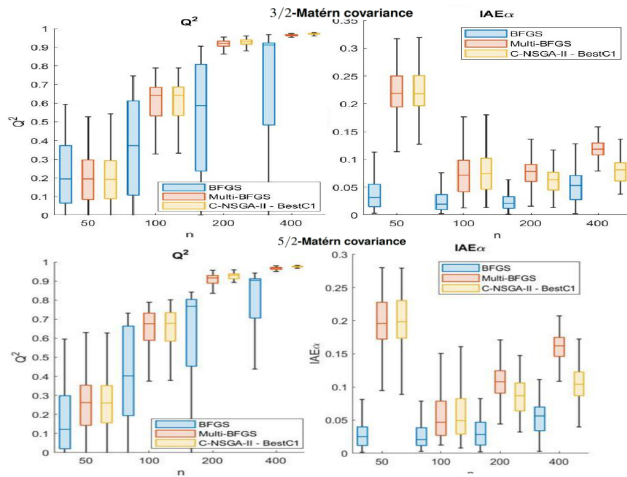
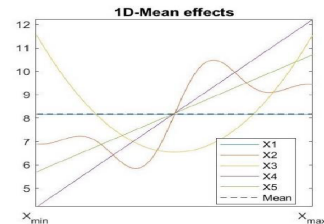


Intensive benchmark on test functions

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

Example: Marrel-d20 function

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



(no nugget)

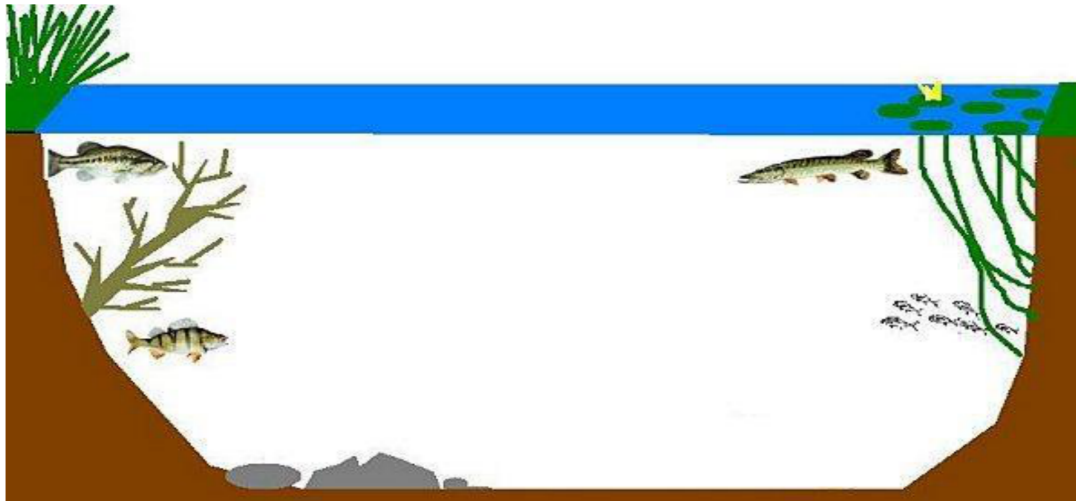
⇒ **Predictivity with Constr-NSGA-II algorithm** at least as good as the simple NLL optimization

⇒ **Improvement of IAE α 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)

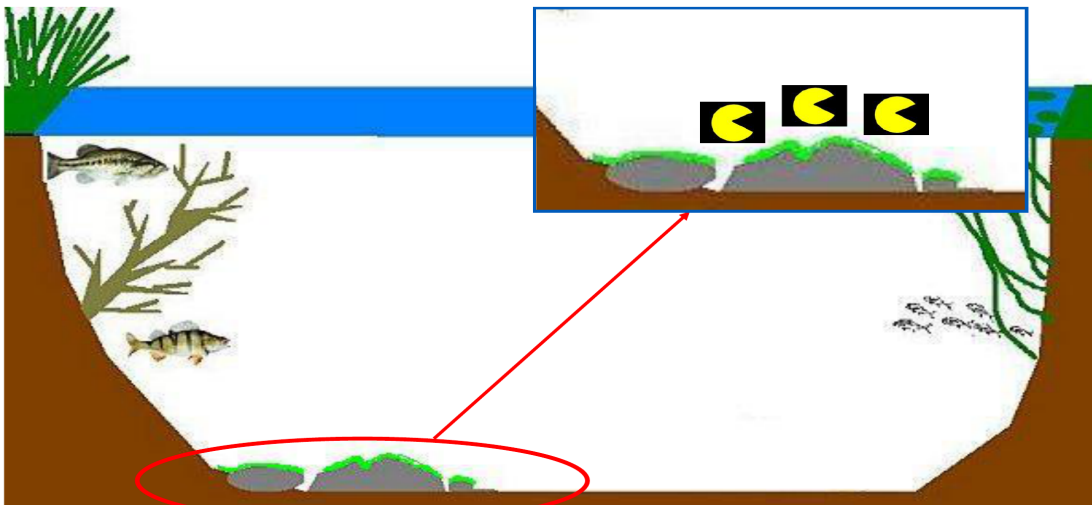
Application: aquatic prey-predator chain model

Studies of biological contamination of rivers



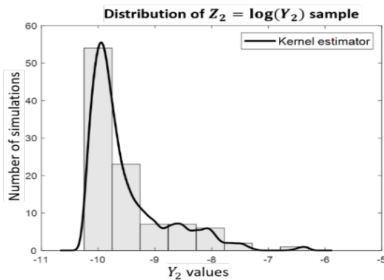
Application: aquatic prey-predator chain model

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

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

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



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			IAE α		
		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 IAE α**

⇒ **Without** nugget effect

Data	Covariance	Predictivity Coefficient Q^2			IAE α		
		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 IAE α**

⇒ **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 \Rightarrow 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!**



Thank you!

Funding by SAMOURAI ANR project



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



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



References

- [RW06] Rasmussen & Williams, *Gaussian processes for machine learning*, MIT Press, 2006
- [IM19] looss & Marrel, Advanced methodology for uncertainty propagation in computer experiments with large number of inputs, *Nuclear Technology*, 205:1588-1606, 2019
- [ILG19] looss & Le Gratiot, 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, looss, Le Gratiot & 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, Bertonecello, & 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] looss, 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

