A new estimation algorithm for more reliable prediction in Gaussian Process Regression

Amandine MARREL*, Bertrand IOOSS‡

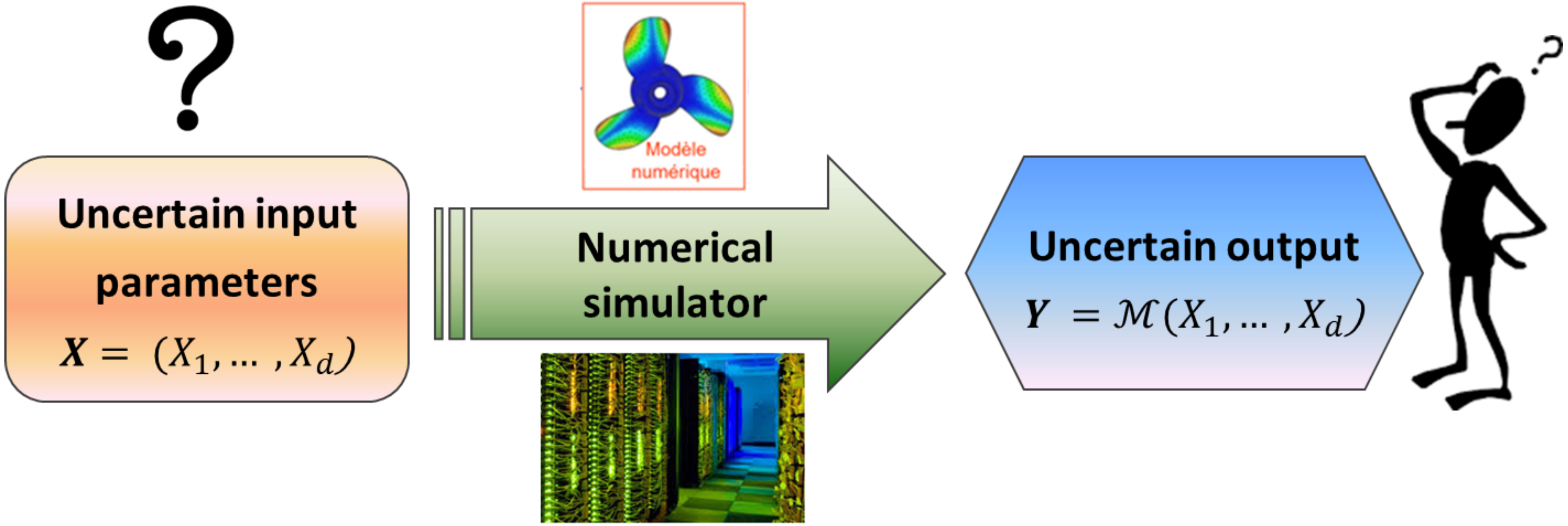
*CEA Energy Division, IRESNE, DER, Cadarache, France

‡EDF R&D, Chatou, France

Workshop of the SAMOURAI ANR project - December 2024, Paris.

Risk assessment in nuclear accident analysis

- **Safety studies:** compute a failure risk (margins, rare events) with validated computer/numerical models
- **Numerical simulators:** fundamental tools to understand, model & predict physical phenomena
- **Large number of input parameters,** related to physical and numerical modelling
- **Uncertainty on some inputs → uncertainty on output & safety margins**
- **BEPU (Best-Estimate-Plus-Uncertainties)**: realistic models + uncertain inputs → **Better assessment of the real margins**



Risk assessment in nuclear accident analysis

- **How to deal with uncertainties in numerical simulation?**

→ Probabilistic framework and Monte Carlo-based methods

→ **CPU-expensive simulator** ⇒ Use of **machine learning to mimic the simulator and propagate input uncertainties**

→ **Applicative constraints/framework:**

✓ **Given data for training**: a single inputs/output **sample** $D_S = \left(\boldsymbol{x}^{(i)}, y^{(i)}\right)_{1 \leq j \leq n}$ where $y^{(i)} = \mathcal{M}\left(\boldsymbol{x}^{(i)}\right)$

→ random or quasi-random sample

✓ **Small sample size**: $n \approx$ 100 to 1000 simulations

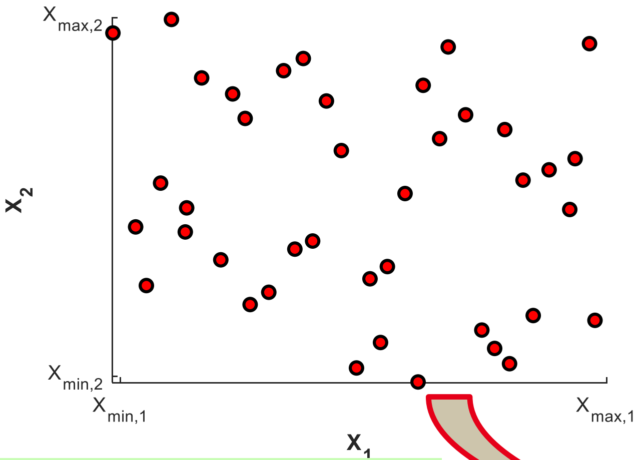
✓ **Large number of uncertain inputs**: d ≈ 10 to 100 inputs

✓ **Required UQ associated to each prediction**

→ Gaussian Process Regression (GPR): particularly well-suited tool ⇒ Very popular

Crucial use of GPR metamodel

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



Incertain inputs domain

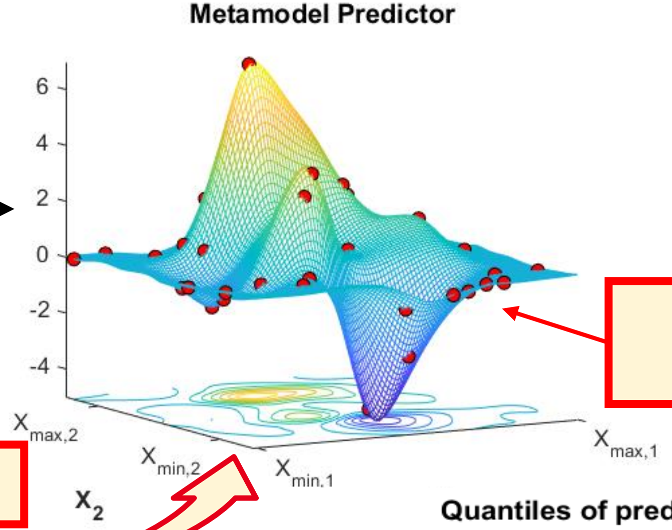
→ Simulator

$Y = \mathcal{M}(X_1, \ldots, X_d)$ →

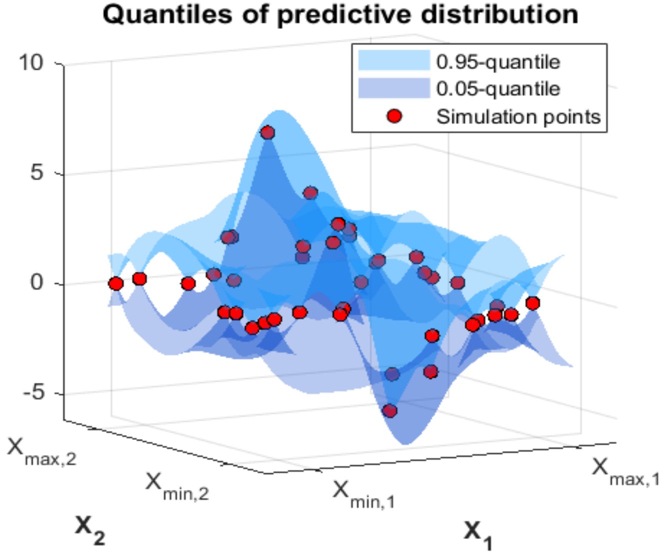
In case of costly $\mathcal{M}$:

Approximation with GPR

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



Probabilistic metamodel



✓ Build from the dataset, GPR mimics the true model $\mathcal{M}$, providing a **GP predictive distribution** for each new evaluation point

$\Rightarrow$ **Intrinsic quantification of prediction error!**

$\Rightarrow$ Very appealing, but in practice calls for **a few good practices**!

Building an efficient GPR in practice

1. Dealing with the large input dimension

Dealing with the large input dimension

How to train the GP in large dimension? (d~10 to 100, e.g.)

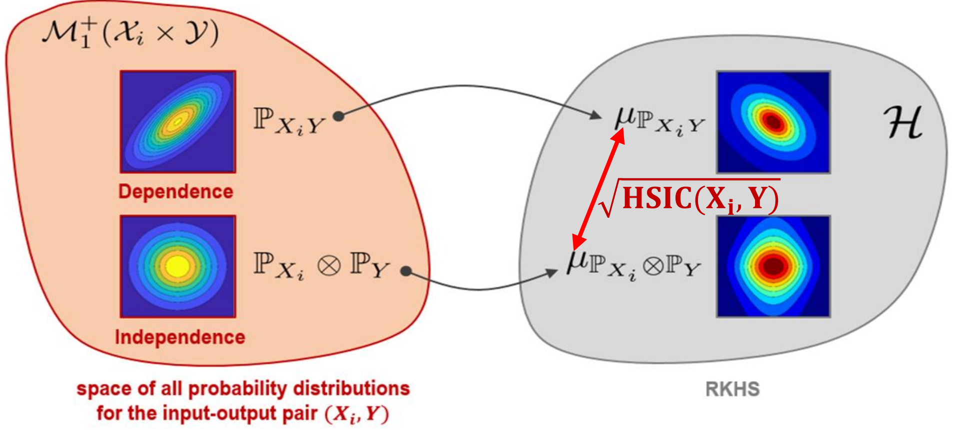
► **Curse of dimensionality $\Rightarrow$ too many GP hyperparameters have to be optimized!**

Preliminary SCREENING for input selection (and thus dimension reduction)

HSIC-based sensitivity measure [GFT+07] $\rightarrow$ dependence measure comparing the RKHS embeddings of joint distribution $\mathbb{P}_{X_i Y}$ and product of marginals $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$

$$HSIC(X_i, Y) = MMD^2(P_{X_i Y}, P_{X_i} \otimes P_Y) = \left\| \mu_{\mathbb{P}_{X_i Y}} - \mu_{\mathbb{P}_{X_i} \otimes \mathbb{P}_Y} \right\|^2$$

- ✓ HSIC can capture **a large spectrum of** input-output **relationships** (power of RKHS ☺)
- ✓ $\widehat{HSIC}$: Estimation from **a unique random sample**, robust in practice from $n \sim 100$

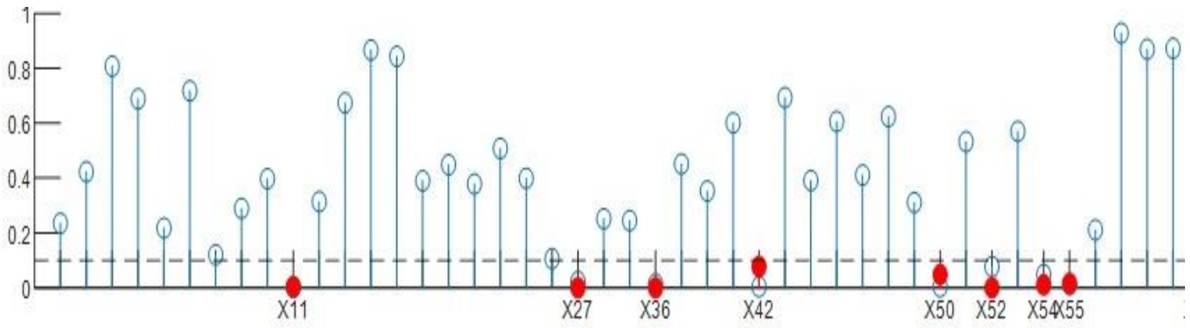


Extract from a presentation by G. Sarazin (CEA)

Dealing with the large input dimension

- **Screening with HSIC-based independence tests** [GFT+07]: $HSIC(X_i, Y) = 0 \Leftrightarrow X_i \perp Y$ (*with characteristic kernels!*)

P-value of global-HSIC tests



● Selected inputs

Selection of **significant inputs** (usually <20)

- ✓ **Explicative inputs** of GPR
- ✓ Non-significant influential inputs captured by an additional variance in GPR (nugget effect)

- **HSIC-based ranking with R²HSIC** [Dav15] : Inputs ordered by degree of influence

Can be used for **more robust sequential GPR estimation**

$\Rightarrow$ "Forward" estimation of GPR hyperparameters: successive inclusion of ordered inputs

See the **"ICSCREAM" methodology** [MIC22]

Building an efficient GPR in practice

1. **Dealing with the large input dimension**

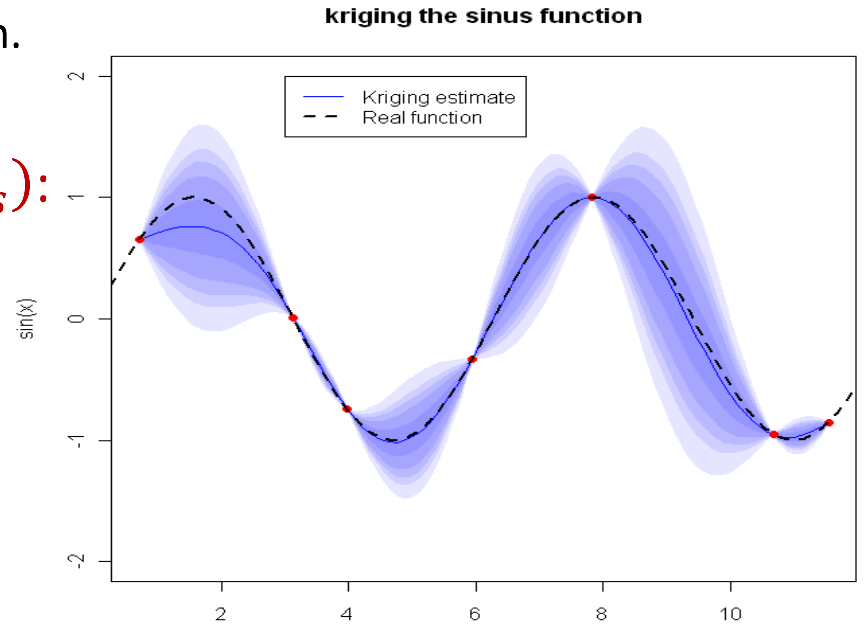
2. **Estimation of hyperparameters and validation**

Reminders on GPR

► **Probabilistic surrogate model**: response is considered as a realization of a random GP field [RW05,Gra21]

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

With $\mu(\boldsymbol{x})$ the mean and $k(\boldsymbol{x}', \boldsymbol{x})$ the covariance function.



⇒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 GPR

▶ **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: stationary covariance built-upon tensorized 1-D covariance functions of ν-Matérn

1-Dim ⟶ $k_{\sigma,\nu,\theta}(x,\tilde{x}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}h}{\theta}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}h}{\theta}\right)$ ⟶ **3/2 or 5/2 Matérn covariances** offer good properties and « intermediate » regularity

d-Dim ⟶ $k_{\sigma,\nu,\boldsymbol{\theta}}(\mathbf{x},\tilde{\mathbf{x}}) = \sigma^2 \prod_{i=1}^{d} k_{1,\nu,\theta_i}(x_i - \tilde{x}_i)$ with $h = |x - \tilde{x}|$

Hyperparameters
$\boldsymbol{\theta} \in \mathbb{R}^{+,d}$

⇒ Additional variance (nugget effect → nugget hyperparameter $\lambda \in \mathbb{R}^+$)

	$v = \frac{1}{2}$	$v = \frac{3}{2}$	$v = \frac{5}{2}$	$v = +\infty$
Usual name	exponential	3/2-Matérn	5/2-Matérn	Gaussian
$k_{\sigma,\nu,\theta}(x,\tilde{x})$	$\sigma^2 e^{-\frac{h}{\theta}}$	$\sigma^2(1+\sqrt{3}\frac{h}{\theta})e^{-\sqrt{3}\frac{h}{\theta}}$	$\sigma^2\left(1+\sqrt{5}\frac{h}{\theta}+\frac{5}{3}\left(\frac{h}{\theta}\right)^2\right)e^{-\sqrt{5}\frac{h}{\theta}}$	$\sigma^2 e^{-\frac{1}{2}\left(\frac{h}{\theta}\right)^2}$
Differentiability of GP trajectories	$\mathcal{C}^0$	$\mathcal{C}^1$	$\mathcal{C}^2$	$\mathcal{C}^\infty$

Estimation of GPR hyperparameters

$\Rightarrow$ **How to robustly estimate the hyperparameters $\boldsymbol{\theta} \in \mathbb{R}^{+,\boldsymbol{d}}$ from the learning sample?**

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

$\Rightarrow$ Crucial for safety applications

Especially in **« medium » dimension** ($d \in [10, 20]$) and **small dataset** ($n \in [100, 1000]$)

Estimation of GPR hyperparameters

▶ **Usual estimation methods** [KO22,Mur21,Pet22,PBF+23]

→ **Maximum likelihood (MLE)** ⇔ minimization of NLL

→ **Cross-validation and Mean Squared Error** :

minimization of $\mathrm{RMSE} = \left\{\frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}_{-i}(\mathbf{x}^{(i)})\right)^2\right\}^{0.5}$

where $\hat{y}_{-i}(\mathbf{x}^{(i)})$ is the metamodel predictor in $\mathbf{x}^{(i)}$ when $(\mathbf{x}^{(i)}, y^{(i)})$ is removed from the learning sample.

Ill-posedness of MLE, problem of **flatness** of functions to be minimized

→ **Bayesian approaches**

CPU ++, delicate choice of priors
Except RobustGAsp method of [GWB18]

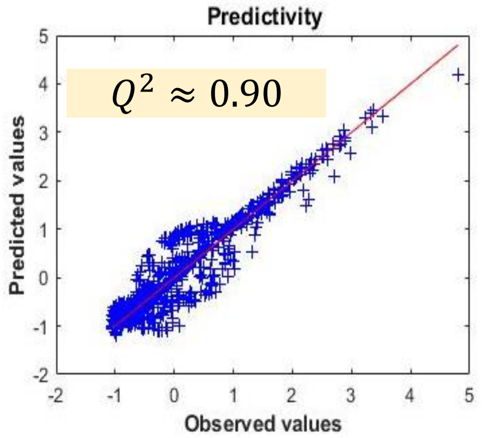
⇒ Could we do better?

⇒ How to check that estimated hyperparameters lead to a "good" GPR metamodel?

Validation of GPR

▶ **Validation criteria computed by cross-validation (LOO or K-fold CV)** [DIG+21, ABG23, MI24a]

→ **Accuracy of the GP predictor (only):** $Q^2 = 1 - \frac{RMSE^2}{\frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \frac{1}{n}\sum_{i=1}^{n} y^{(i)}\right)^2}$



→ **Accuracy of the predictive variance:** $\text{PVA} = \left|\log \frac{1}{n}\sum_{i=1}^{n} \frac{\left(y^{(i)} - \hat{y}_{-i}(\mathbf{x}^{(i)})\right)^2}{\hat{s}_{-i}^2}\right|$

→ **Accuracy of the whole GP conditional distribution**

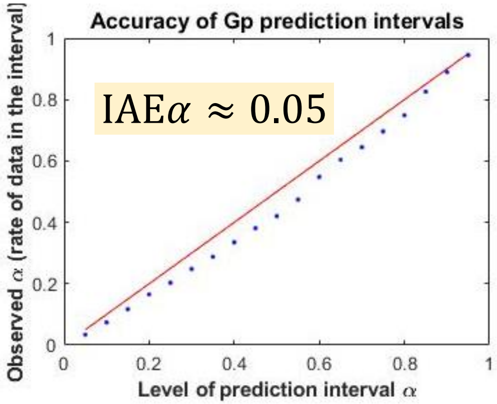
From empirical coverage function for α∈[0,1]: $\hat{\Delta}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y^{(i)} \in PI_{\alpha,-i}(\mathbf{x}^{(i)})\}$

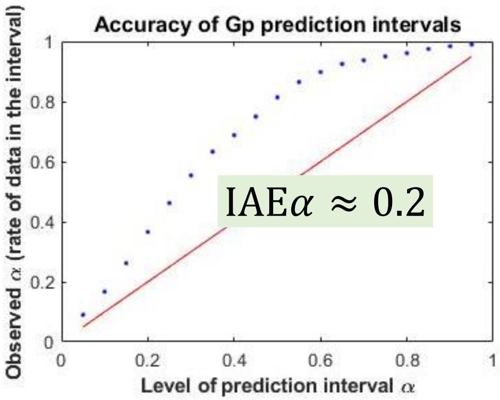
with $PI_{\alpha,-i}(\mathbf{x}^{(i)})$ the α-level GP prediction interval for $\mathbf{x}^{(i)}$ with $(\mathbf{x}^{(i)}, y^{(i)})$ removed from learning sample

⇒ **α-PI Plot**

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

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





Building an efficient GPR in practice

1. Dealing with the large input dimension

2. Estimation of hyperparameters and validation

3. New hyperparameter estimation algorithm

From the analysis of estimation & validation criteria...

- **Study of criteria NLL, Q^2, PVA and $IAE\alpha$ on a large benchmark of analytical tests**

  → **Close behavior of NLL and Q^2 $\Rightarrow$ keep NLL as the main estimation objective** to ensure predictivity

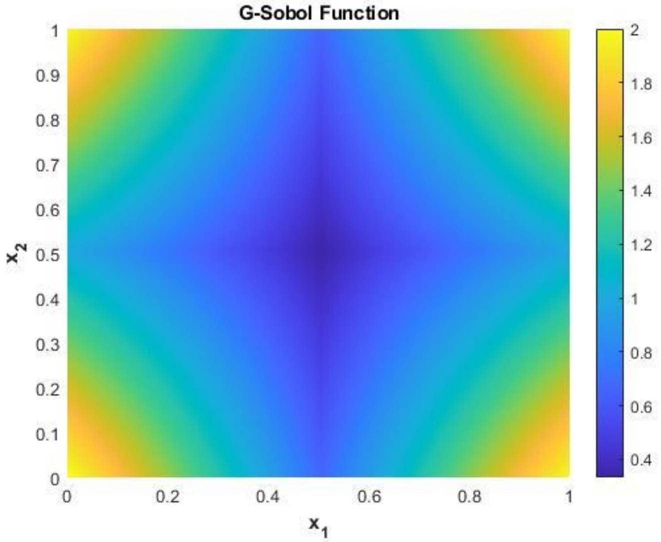
  → *Consistent with [PBF+23,Pet22]*

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

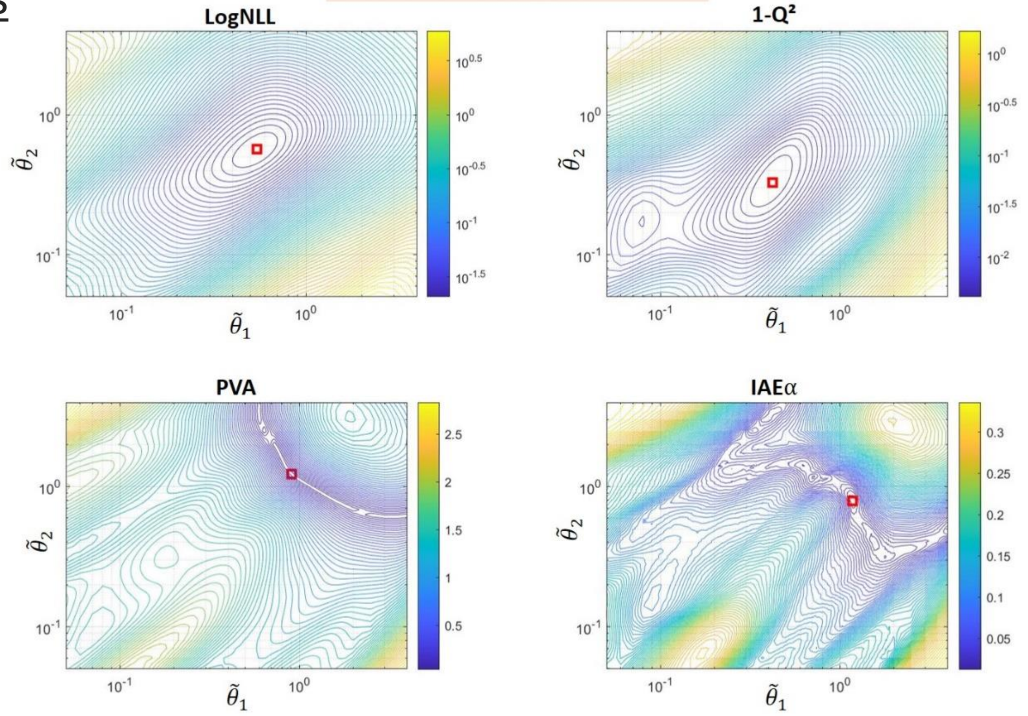
  $\Rightarrow$ Some local minima compatible with optimal values of the other criteria

  $\Rightarrow$ But No to be optimized independently of the others

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



$n = 50$ – Matérn 5/2 Covariance



To a new estimation algorithm!

- **Study of criteria NLL, Q^2, PVA and $IAE\alpha$ on a large benchmark of analytical tests**

→ **Close behavior of NLL and Q^2 $\Rightarrow$ keep NLL as the main estimation objective** to ensure predictivity

→ **$\mathbf{IAE}\alpha$ 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 $\mathbf{IAE}\alpha$**, but need to control the possible degradation of Q^2 value, which guarantees the predictivity

$\Rightarrow$ **Optimization based on NLL and $IAE\alpha$ + Control of Q^2**

($IAE\alpha$ and Q^2 estimated by cross validation + use of LOO Dubrule formulas)

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

Algorithm flowchart

All details in Marrel and B. Iooss, *Probabilistic surrogate modeling by Gaussian process: A new estimation algorithm for more robust prediction*, Reliability Engineering and System Safety, Volume 247, July 2024, 110120.

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

↓

Pareto front of n_{Pareto} non-dominated solutions

$\left(\tilde{\boldsymbol{\theta}}_{NGSA,1}^{new}, \dots, \tilde{\boldsymbol{\theta}}_{NGSA,n_{pareto}}^{new}\right)$

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

⇒ $\hat{Q}_{LOO}^{2,new}$

→ ⇒ 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

↓

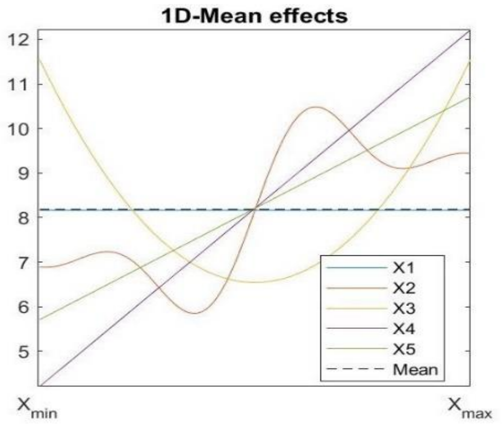
$\tilde{\boldsymbol{\theta}}_{NSGA}^{BestC1}$

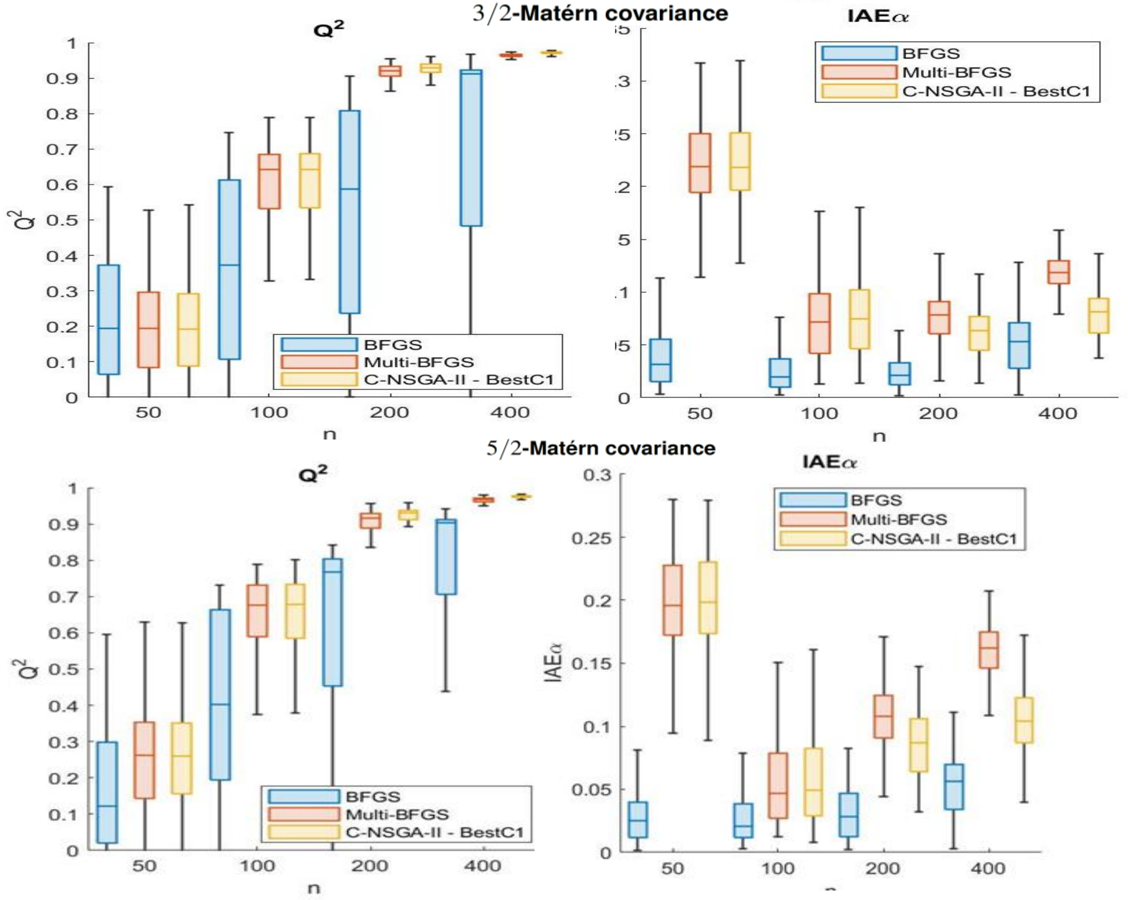
Intensive benchmark on analytical test functions

► Comparison with usual algorithms based on NLL optimization only (BFGS/multistart)

d = 2 to 20, ≠ covariance, ≠ sample sizes, ≠ DoE, with/without nugget effect

Example on Marrel-d20 function : $Y(\boldsymbol{X}) = a_1 \sin[6\pi X_1^{\frac{5}{2}} \left(X_2 - \frac{1}{2}\right) + a_2 \left(X_3 - \frac{1}{2}\right)^2 + a_3 X_4 + a_4 X_5 + r_{X_6,\ldots,X_{15}}$





Results without nugget effect

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

⇒ **Improvement of $IAE\alpha$ 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** (e.g. large dimension *d* + tensorized anisotropic stationary covariance)

Building an efficient GPR in practice

1. Dealing with the large input dimension

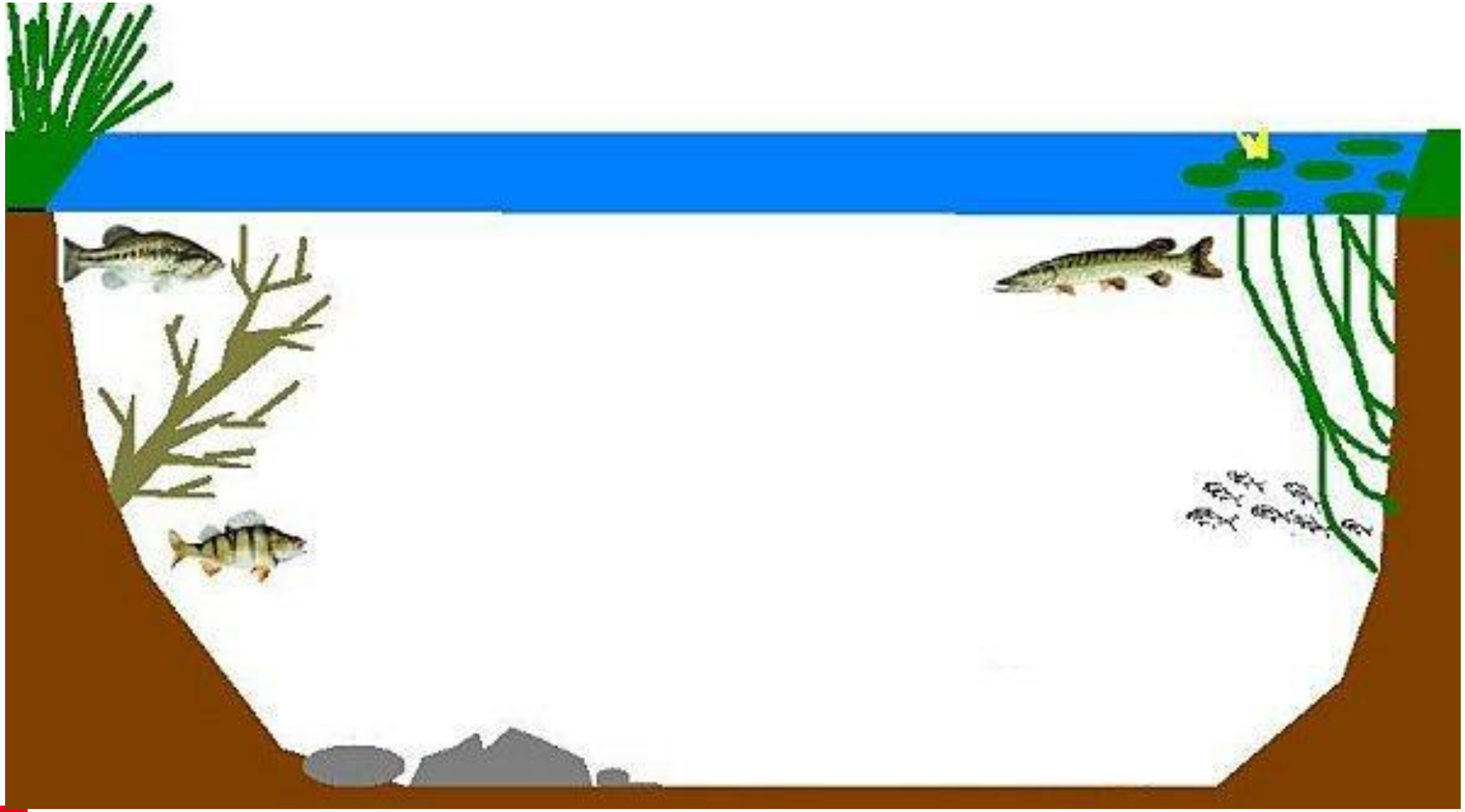
2. Estimation of hyperparameters and validation

3. New hyperparameter estimation algorithm

4. Illustration on aquatic prey-predator chain model

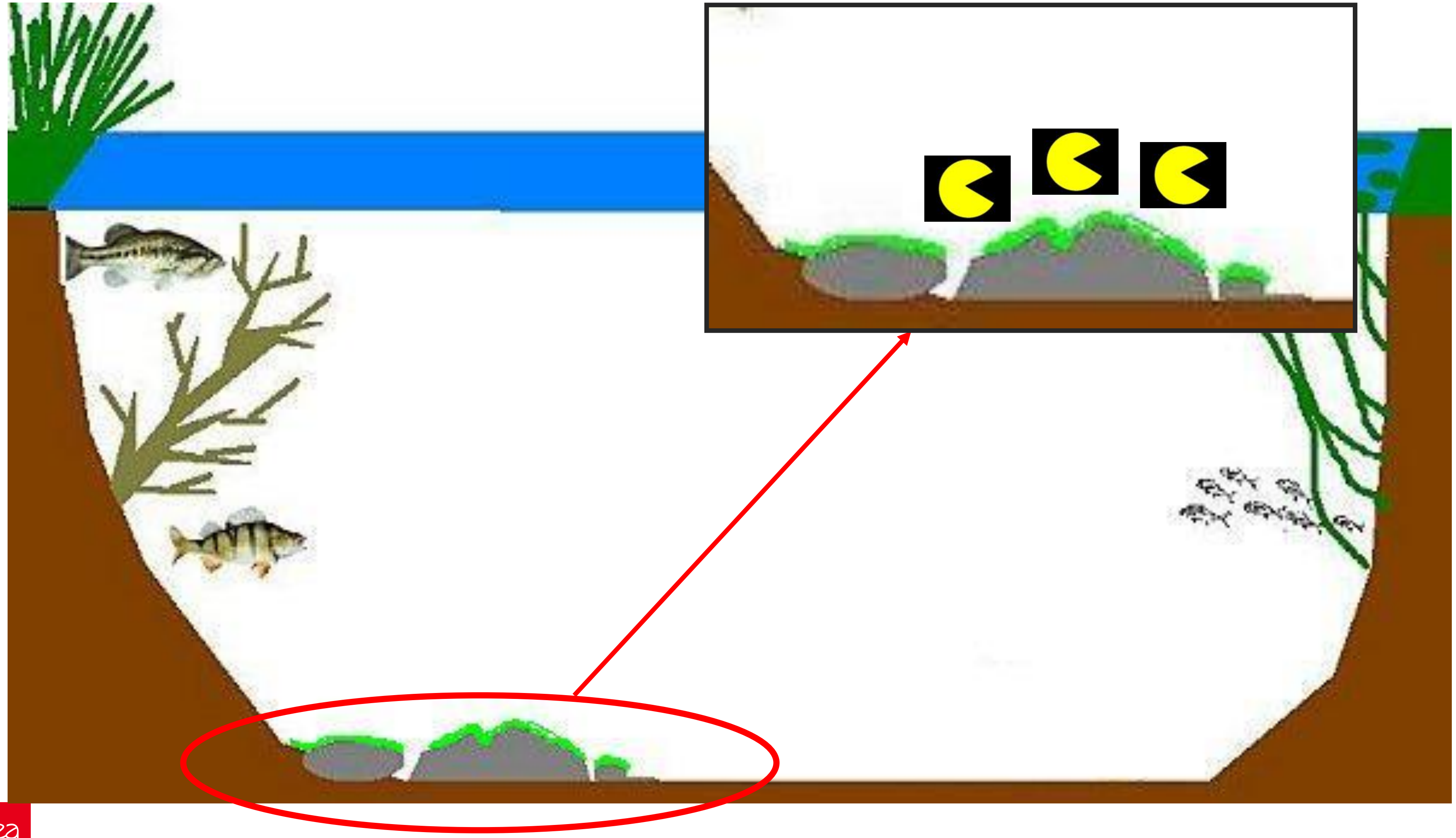
Application: aquatic prey-predator chain model

Studies of biological contamination of rivers



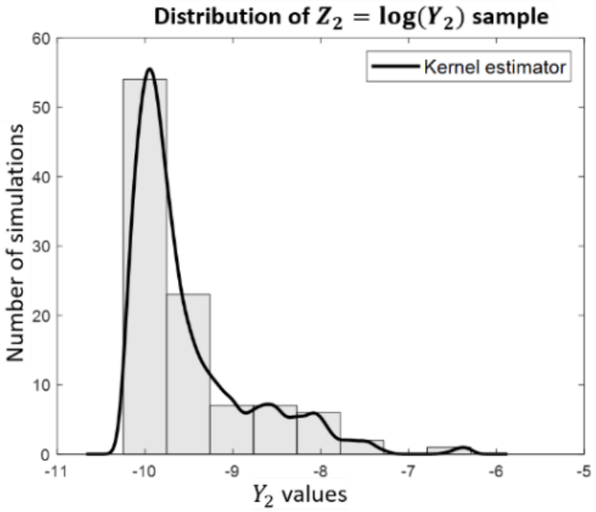
Application: aquatic prey-predator chain model

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



Application: aquatic prey-predator chain model

- **Simulator: MELODY with *d = 20*** uncertain inputs:
  - Periphyton: photosynthesis/mortality/excretion rates, survival temperature, saturation constants, …
  - Grazers: consumption/assimilation/ mortality/excretion rate, survival temperature, …
- **2 outputs of interest**: Periphyton (Y_1) and Grazers (Y_2) biomasses at day 60
- Sample of ***n = 100*** simulations of the model MELODY (from 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

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

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

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

Application: aquatic prey-predator chain model

⇒ **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 more robust to modeling choices** (prior choice of GPR covariance)

Conclusions and remaining challenges

- ✓ GPR benefits greatly from **preliminary HSIC-based screening**
- ✓ GPR calls for **robust estimation of hyperparameters:** considering validation criteria of the whole GP distribution when estimating hyperparameters $\Rightarrow$ enables more robust estimation !
- ✓ Particular attention must be paid to **GP validation**

  $\Rightarrow$ Part of a more general effort to **ensure confidence in machine learning for UQ**

► **Some interesting challenges for UQ applications**

- ✓ Use **more powerful tests** based on SupHSIC [EM24] and HSIC-ANOVA indices [SMD+23]
- ✓ **Screening-free approaches for high dimensional problems** (e.g. beyond 30 to 50 inputs)
- ✓ Learning **outputs with highly irregular**, or even **chaotic behavior** (due to physical threshold phenomena and phenomenological bifurcations, for instance)

References 1/2

Reference of this work

[MI24a] A. Marrel and B. Iooss, Probabilistic surrogate modeling by Gaussian process: A review on recent insights in estimation and validation, Reliability Engineering and System Safety, Volume 247, July 2024, 110120.

[MI24b] A. Marrel and B. Iooss, Probabilistic surrogate modeling by Gaussian process: A new estimation algorithm for more robust prediction, - Reliability Engineering and System Safety, Volume 247, July 2024, 110094.

Work funded by ANR SAMOURAI research project

General references

[ABG23] Acharki, N., Bertoncello, A., and Garnier, J. (2023). Robust prediction interval estimation for GP by cross-validation method. Computational Statistics Data Analysis, 178:107597.

[Dav15] Da Veiga (2015). Global sensitivity analysis with dependence measures, Journal of Statistical Computation and Simulation, 85:1283-1305, 2015.

[DIG+21] Demay, C., Iooss, B., Gratiet, L., and Marrel, A. (2022). Model selection for GP regression: an application with highlights on the model variance validation. QREI Journal, 38:1482-1500.

[EM24] El Amri and. Marrel (2024). More powerful HSIC based independence tests, extension to space filling designs and functional data. International Journal for Uncertainty Quantification14(2): 69-98.

References 2/2

[Gra21] B. Gramacy (2021) Gaussian Process Modeling, Design, and Optimization for the Applied Sciences. Chapman and Hall/CRC.

[GFT+07] Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B. & Smola, A. (2007). A kernel statistical test of independence. Advances in Neural Information Processing Systems, 2007.

[GWB18] Gu, M., Wang, X., and Berger, J. O. (2018). Robust gaussian stochastic process emulation. The Annals of Statistics, 46(6A):3038 – 3066.

[KO22] Karvonen & Oates (2022). Maximum Likelihood Estimation in GP is ill-posed. Preprint.

[MIC22] Marrel, Iooss and Chabridon, (2022). The ICSCREAM Methodology: Identification of Penalizing Configurations in Computer Experiments Using Screening and Metamodel – Applications in Thermal Hydraulics, Nuclear Science and Engineering, 196:3,301-321.

[Mur21] Muré (2021). Propriety of the reference posterior GP distribution. The Annals of Statistics. 49(4):2356-2377.

[Pet22] Petit S. (2022). Improved Gaussian process modeling. Application to Bayesian optimization. PhD University Paris-Saclay.

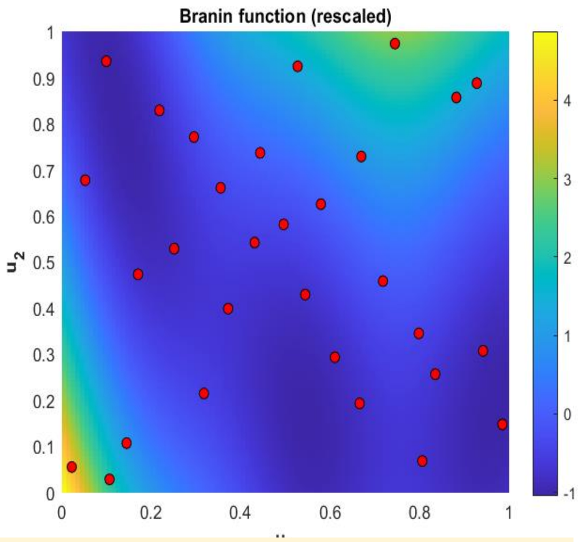
[PBF+23] Petit, S., Bect, J., Feliot, P., and Vazquez, E. (2023). Model parameters in GP interpolation: an empirical study of selection criteria. SIAM/ASA Journal on Uncertainty Quantification, 11(4), 1308-1328.

[RW05] C.E. Rasmussen and C.K.I. Williams (2006). Gaussian processes for machine learning. MIT Press.

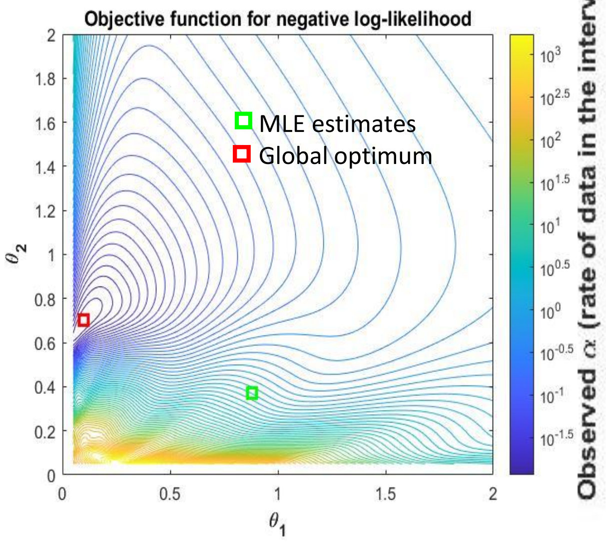
[SMD+23] Sarazin, G., Marrel, A., Da Veiga, S. & Chabridon (2023). New insights into the feature maps of Sobolev kernels: application in global sensitivity analysis. https://cea.hal.science/cea-04320711.

Appendix

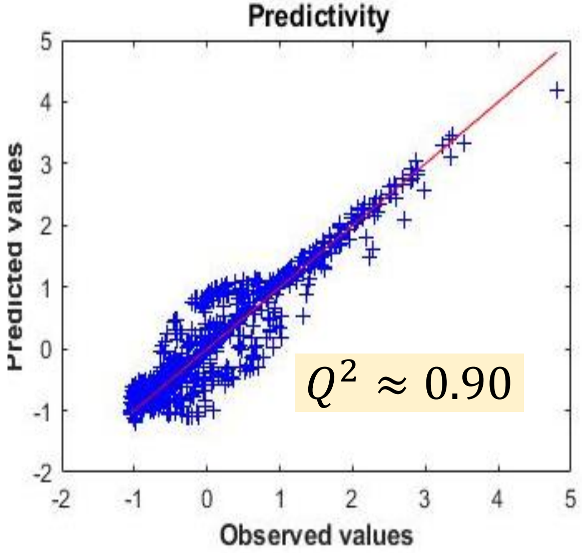
Illustration of criteria for GPR validation [MI24a]

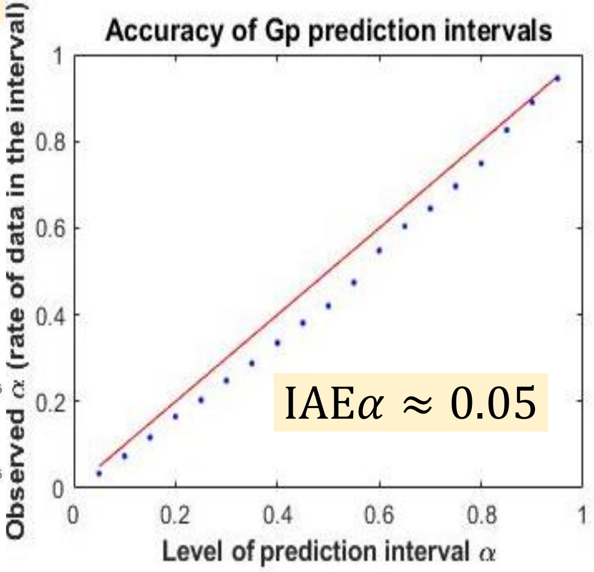


n=30, GPR with constant mean and Gaussian covariance

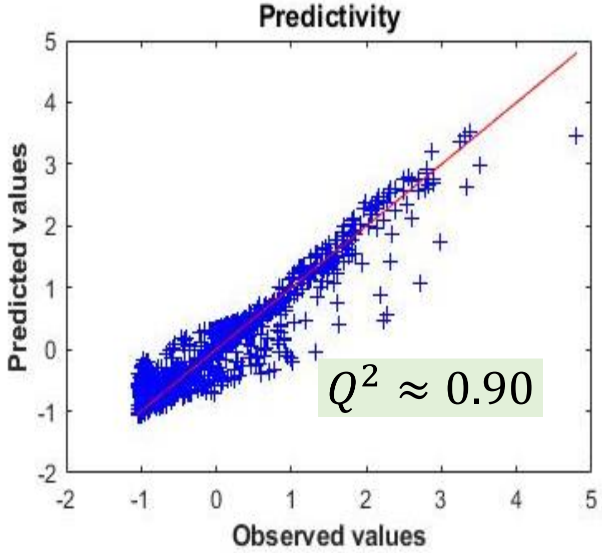


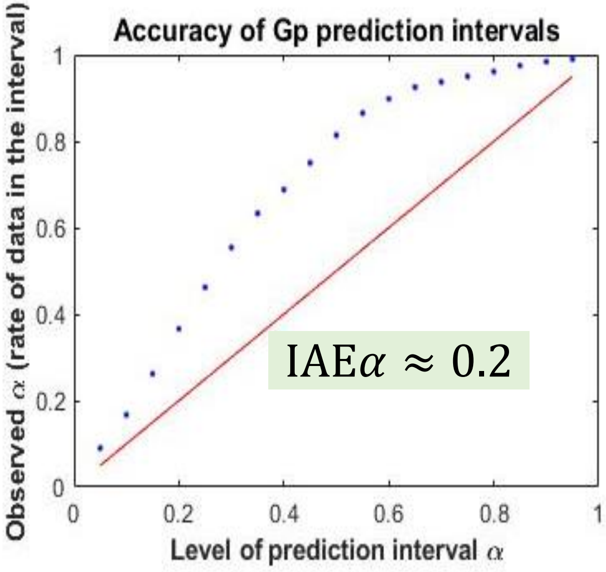
$[\theta_1, \theta_2] = [1.12 \quad 0.8]$



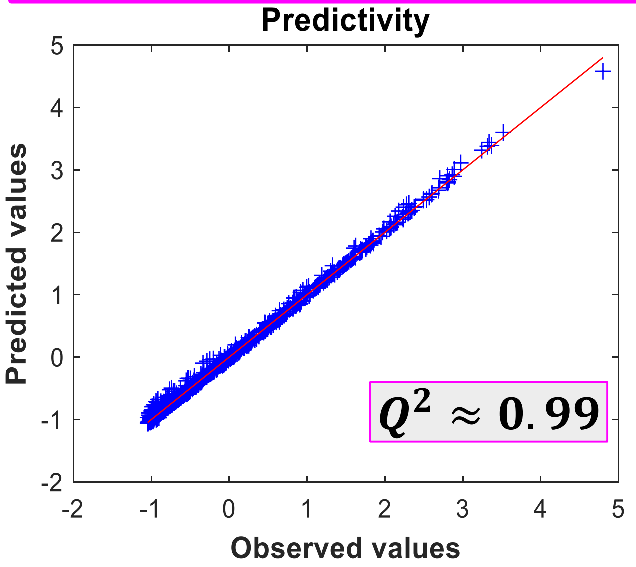


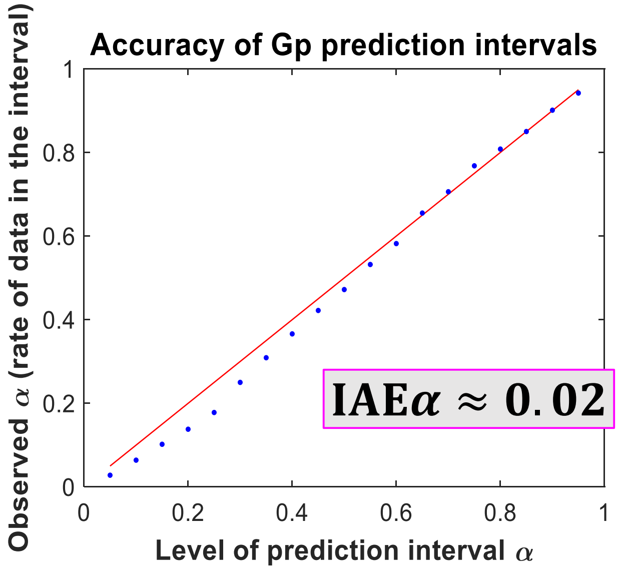
$[\theta_1, \theta_2] = [0.78 \quad 0.52]$





MLE estimates: $[\theta_1, \theta_2] = [0.88 \;\; 0.37]$





Dealing with the large input dimension

- **HSIC-based ranking** [Dav15] **:**

$$R^2_{HSIC} = \frac{HSIC\ (X,Y)}{\sqrt{HSIC(X,X)HSIC\ (Y,Y)}}$$

⇨ $R^2_{HSIC} \in [0,1]$ for easier interpretation

$$\textbf{Influence}(X_{[1]}) > \textbf{Influence}(X_{[2]}) > \cdots > \textbf{Influence}(X_{[d]})$$

Where order $[\cdot]$ is such that $\hat{R}^2_{H,X_{[1]}} > \hat{R}^2_{H,X_{[2]}} > \cdots > \hat{R}^2_{H,X_{[d]}}$

⇒ Use for ranking of inputs

Inputs ordered by degree of influence

Can be used for **more robust sequential GPR estimation**

⇒ "forward" estimation of GPR hyperparameters: successive inclusion of ordered inputs

HSIC review: a kernel-based GSA method

- **MMD² applied between $P_{X_i Y}$ and $P_{X_i} \otimes P_Y \Rightarrow HSIC(X_i, Y)_{\mathcal{H}_{X_i}, \mathcal{H}_Y}$**

$\mathcal{H}_{X_i}$ and $\mathcal{H}_Y$ **RKHS** associated to X_i and Y, resp :

Kernel $k_{X_i}: \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ with feature space $\mathcal{H}_{X_i}$ and feature map φ_{X_i}

Kernel $k_Y: \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ with feature space $\mathcal{H}_Y$ and feature map φ_Y

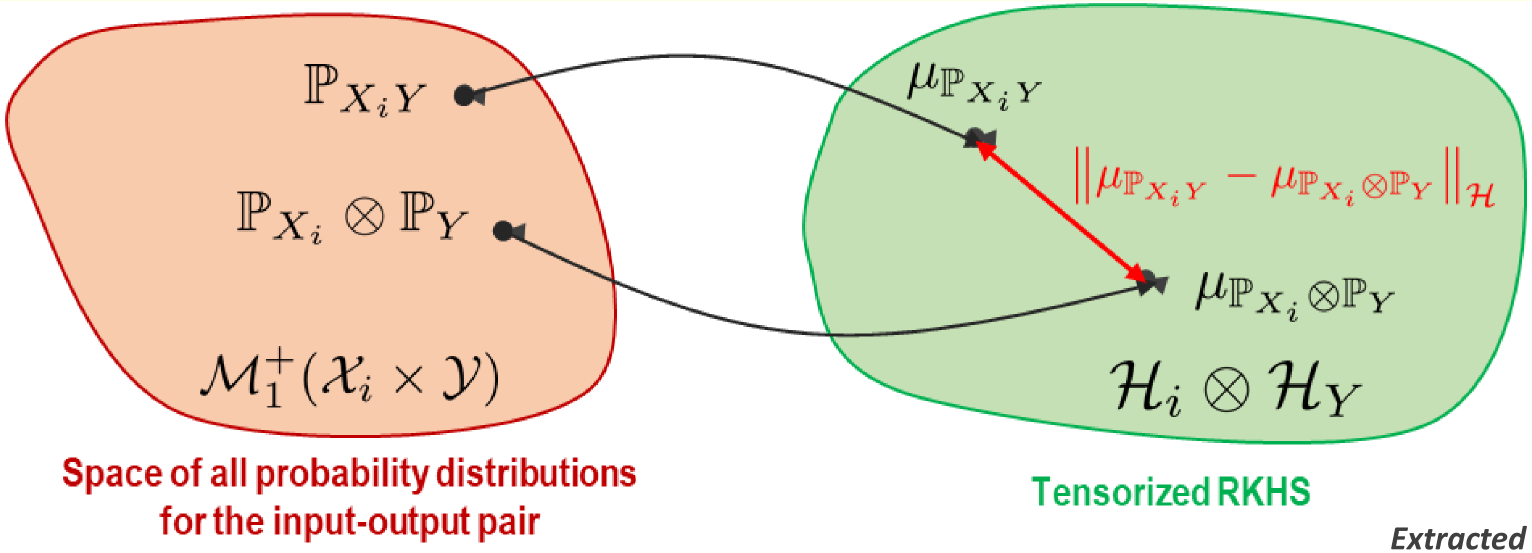
$K_{X_i}(x, x') = \langle \varphi_{X_i}(x), \varphi_{X_i}(x') \rangle_{\mathcal{H}_{X_i}}$ and $K_Y(y, y') = \langle \varphi_Y(y), \varphi_Y(y') \rangle_{\mathcal{H}_Y}$

kernel defines the inner product in the RKHS

HSIC = distance in the RKHS between the images of the two distributions of interest

Gretton et al. [2005]

$$\Rightarrow HSIC(X_i, Y)_{\mathcal{H}_{X_i}, \mathcal{H}_Y} = MMD^2_{\mathcal{H}_{X_i}, \mathcal{H}_Y}(P_{X_i Y}, P_{X_i} \otimes P_Y) = \left\| \mu_{\mathbb{P}_{X_i Y}} - \mu_{\mathbb{P}_{X_i} \otimes \mathbb{P}_Y} \right\|^2_{\mathcal{H}_{X_i}, \mathcal{H}_Y}$$



Extracted from G. Sarazin's (CEA) slides

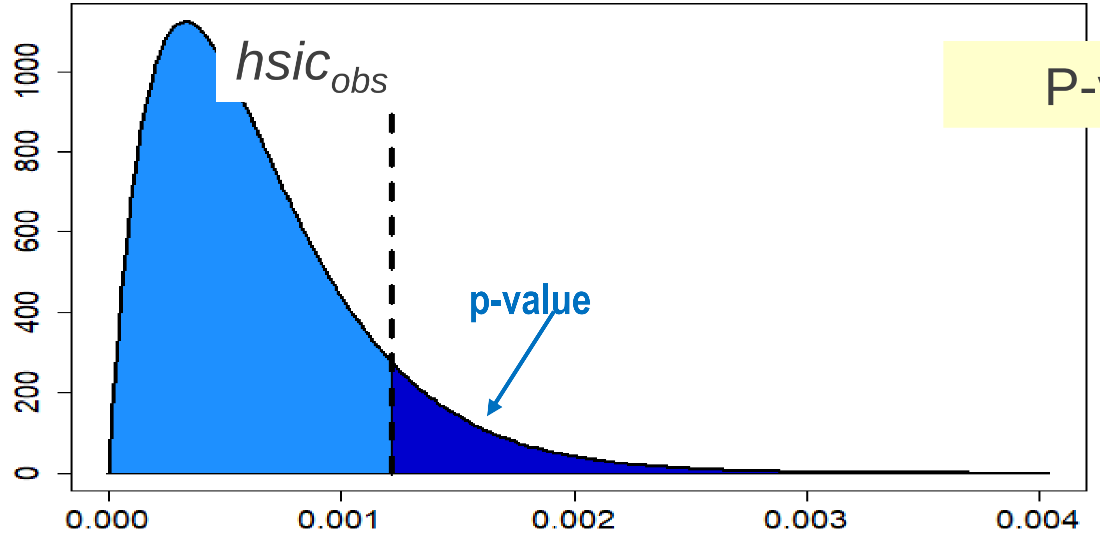
HSIC review: a kernel-based GSA method

HSIC-based independence tests for screening

How to have the distribution $n\widehat{\mathrm{HSIC}}(\mathrm{X_i}, \mathrm{Y})$ under $\mathcal{H}_0$ to compute *p-value*?

- ► **If *n* large: asymptotic test based** on approximation with Gamma law *(Gretton et al. (2008])*
- ► **If *n* small: Permutation-based** approximation *(De Lozzo & Marrel [2016a], Meynaoui [2019], El Amri & Marrel [2021a])*

Gamma distribution



P-value = $Pr\,[\,\widehat{\mathrm{HSIC}}(X_i, Y) > hsic_{obs}\,]$

<u>Interpretation of *p-value* for a level α ($\alpha = 5\%$ or 10%) for screening:</u>

➢ **<u>pval < α</u>** ⇒ H_0 (Independence) rejected ⇒ X_i **is significantly influential**