Speaker: Amandine Marrel

Part of MS194 Gaussian Process Modeling for Inverse Problems

Hidden But Essential Recipes for Successful Gaussian Process Metamodeling to Support Uncertainty Quantification in Numerical Simulation

Amandine Marrel
CEA, DES, IRESNE, DER, Cadarache F-13108 Saint-Paul-Lez-Durance,
amandine.marrel@cea.fr

Bertrand Iooss
EDF R&D, 6 quai Watier, 78400, Chatou, France, bertrand.iooss@edf.fr

The value of using Gaussian Process (GP) regression to emulate costly computational codes for uncertainty management is now widely established. The probabilistic metamodel provided by GP-regression, in the sense that it provides a predictive distribution for each new evaluation point, offers great added value, particularly for safety, reliability or risk assessment studies. However, guaranteeing confidence in the use of this metamodel requires two crucial steps: its training on the available learning data and its validation (often by cross-validation process in our application context). We are particularly interested here in the context of given data, small data (number of model simulations limited to a few hundred) and large numbers of uncertain inputs (from a few dozen to a hundred). In this context, building a successful GP metamodel often calls for a preliminary variable selection. Kernel-based methods (HSIC) and associated independence tests are especially appropriate, for screening but also ranking the inputs. Then, particular care is required when estimating GP hyperparameters: going beyond simple maximum likelihood approaches may be wise. Finally, GP validation must include various criteria to assess the preditivity and reliability of the metamodel's entire predictive law.

The presentation will focus on recent advances in these three topics, with the aim of providing guidelines and recipes for successful GP metamodeling in the considered application context.

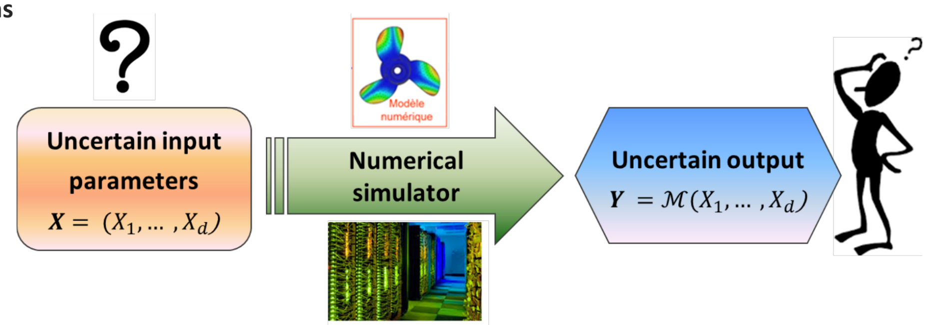
Hidden But Essential Recipes for Successful Gaussian Process Metamodeling to Support UQ in Numerical Simulation

Amandine MARREL*, Bertrand IOOSS‡
*CEA Energy Division, IRESNE, DER, Cadarache, France
‡EDF R&D, Chatou, France

2024 SIAM Conference on Uncertainty Quantification - March 1st 2024, Trieste.

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**

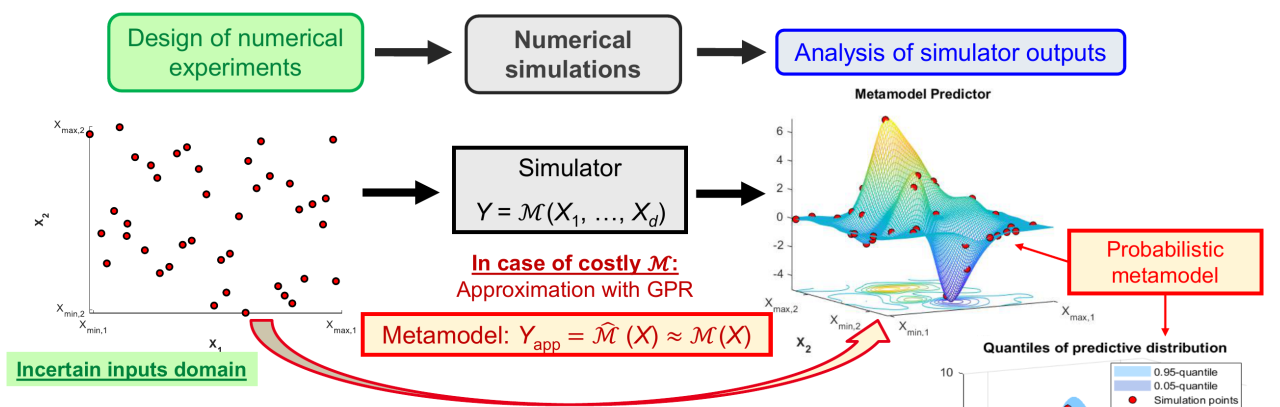


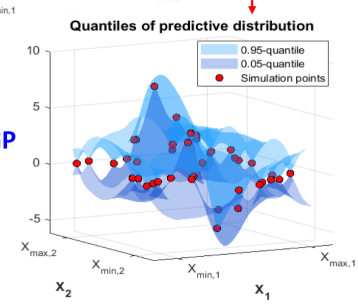
Risk assessment in nuclear accident analysis

- **How to deal with uncertainties in numerical simulation?**
  - → Probabilistic framework and Monte Carlo-based methods
  - → **CPU-expensive simulator** ⇒ Use of **machine learning to propagate input uncertainties**
  - → **Applicative constraints/framework:**
    - ✓ **Given data**: a single inputs/output **sample** $(\boldsymbol{x}^{(i)}, \boldsymbol{y}_i)_{1 \le j \le n}$ **where** $\boldsymbol{y}_j = \mathcal{M}(\boldsymbol{x}^{(i)})$ to be used for multi-purpose (sensitivity analysis, uncertainty propagation... And **training a metamodel**)
    - ✓ **Small sample size**: $n \approx 100$ to 1000 simulations
    - ✓ **Large number of uncertain inputs**: $d \approx 10$ to 100 inputs
    - ✓ **Required UQ associated to each prediction**

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

Crucial use of GPR metamodel





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

⇒ **Intrinsic quantification of prediction error!**

⇒ A very appealing theory, but in practice calls for **a few good practices**!

Challenges for 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!**

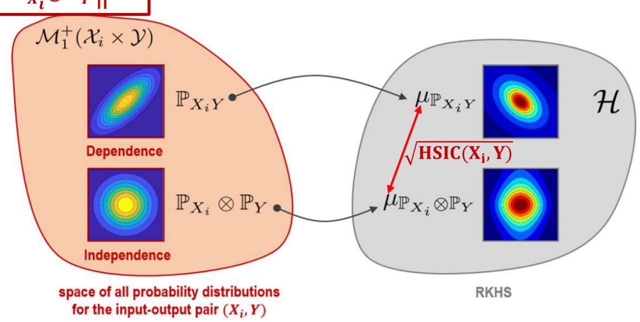
Use of **preliminary SCREENING** for input selection (and thus dimension reduction).

- **Sensitivity measure from HSIC** (Hilbert-Schmidt Independence Criterion [GFT+07]), built-upon RKHS embeddings

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

✓ $\widehat{HSIC}$: Estimation of from a **unique random sample**, efficient in practice from $n \sim 100$

✓ HSIC "summarizes" the cross-cov between feature maps $\Rightarrow$ Large panel of input-output dependency captured



Extract from a presentation by G. Sarazin (CEA)

MMD: Maximum Mean Discrepancy

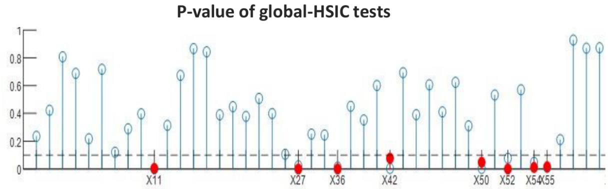
Dealing with the large input dimension

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

- ✓ **Null hypothesis**: $\mathcal{H}_0 : X_i \perp Y$ against $\mathcal{H}_1 : X_i \not\perp Y$
- ✓ **Test statistics**: $n\widehat{\text{HSIC}}(X_i, Y)$
- ✓ **Decision rule**: $\mathcal{H}_0$ rejected iff $n\widehat{\text{HSIC}}(X_i, Y) > q_{1-\alpha}$ where $q_{1-\alpha}$ is the $(1-\alpha)$quantile of $\boldsymbol{n\widehat{\text{HSIC}}(X_i, Y)}$ under $\mathcal{H}_0$ (several available procedures to estimate $q_{1-\alpha}$ [GFT+07, EM24])

⇒ Use for screening of inputs

P-value of global-HSIC tests



● **Selected inputs**

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

↓

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

cea

Dealing with the large input dimension

▶ **HSIC-based ranking** [Dav15]

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

See the **"ICSCREAM" methodology** [MIC21]

cea

Dealing with the large input dimension

► HSIC can capture **a large spectrum of relationships** (power of RKHS ☺)

► Able to deal with **many types of variables and purposes**:

- **Goal-oriented for safety studies** [MC21], to measure the influence in a restricted domain: $Y \in \mathcal{C}$
- **Functional output** $\Rightarrow$ **definition of specific kernels** [EM24]

► More powerful tests based on SupHSIC [EM24] and HSIC-ANOVA indices [SMD+23]

→ **More robust selection of inputs**

→ Efficiency demonstrated in **numerous industrial applications**, especially with small sample size n and large dimension d

cea

Challenges for an efficient GPR in practice

1. Dealing with the large input dimension

2. Reliable estimation of GPR hyperparameters

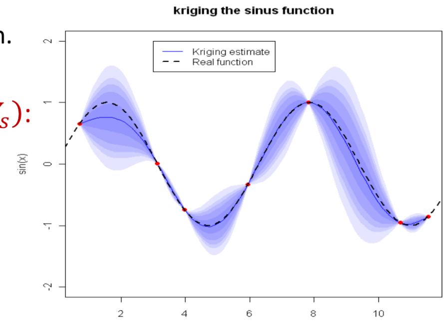
cea

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 $\boldsymbol{x}^*$

Recommendations for parametric choices

▶ **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 GP hyperparameters

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

For good **predicitivity** + **reliable GP prediction intervals**

$\Rightarrow$ Crucial for safety applications

Estimation of GP hyperparameters

▶ **Usual estimation methods** [KO22,Mur21,Pet22,PBF+22]

$\rightarrow$**Maximum likelihood**

$\rightarrow$**Cross-validation Mean Squared Error**

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

$\rightarrow$**Bayesian approaches**

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

Proposition of a new multi-objective estimation algorithm for more reliable GP prediction intervals in [MI24b]

Bertrand Iooss' talk on Wednesday, MS 120: "*Gaussian process regression: new hyperparameter estimation algorithm for more reliable prediction - Application to an aquatic ecosystem model*"

Work supported by French ANR SAMOURAI Project

Challenges for an efficient GPR in practice

1. Dealing with the large input dimension

2. Reliable estimation of GPR hyperparameters

3. Careful GPR validation for confident use

GPR validation

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

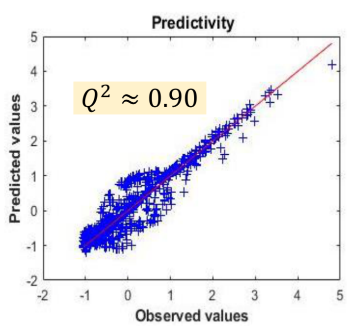
→ **Accuracy of the GP predictor (only)**:

$$Q^2 = 1 - \frac{MSE}{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right)^2} \quad \text{with } \mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_{-i}(\mathbf{x}^{(i)})\right)^2$$

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.

The closer to one the Q^2, the better the accuracy of the metamodel predictor.

Values	Interpretation
High value, close to 1	Good predictive capability of metamodel predictor for unobserved points.
Low value ($Q^2 \leq 0.5$, e.g.)	Poor predictive capability. Some possible reasons: - unsuitable or poorly estimated model; - very poorly predicted areas (Q^2 sensitive to highest or extreme errors) - Insufficient learning sample to properly explore the space of input



GPR validation

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

→ **Accuracy of the whole GP conditional distribution** [DIG+21,ABG23]

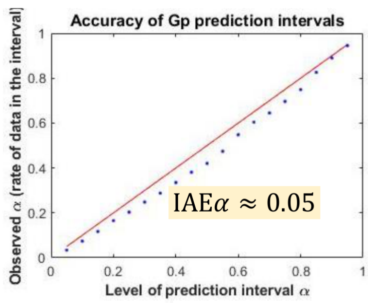
From empirical coverage function for α∈[0,1]: $\hat{\Delta}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{y_i \in PI_{\alpha,-i}(\mathbf{x}^{(i)})\}$

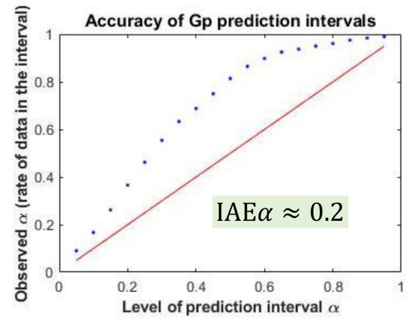
with $PI_{\alpha,-i}(\mathbf{x}^{(i)})$ the α-level GP prediction interval for $\mathbf{x}^{(i)}$ when $(\mathbf{x}^{(i)}, y_i)$ is removed from learning sample

⇒ **α-PI Plot**

⇒ **Integrated Absolute Error on $\hat{\Delta}(\alpha)$** [MI24a]

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



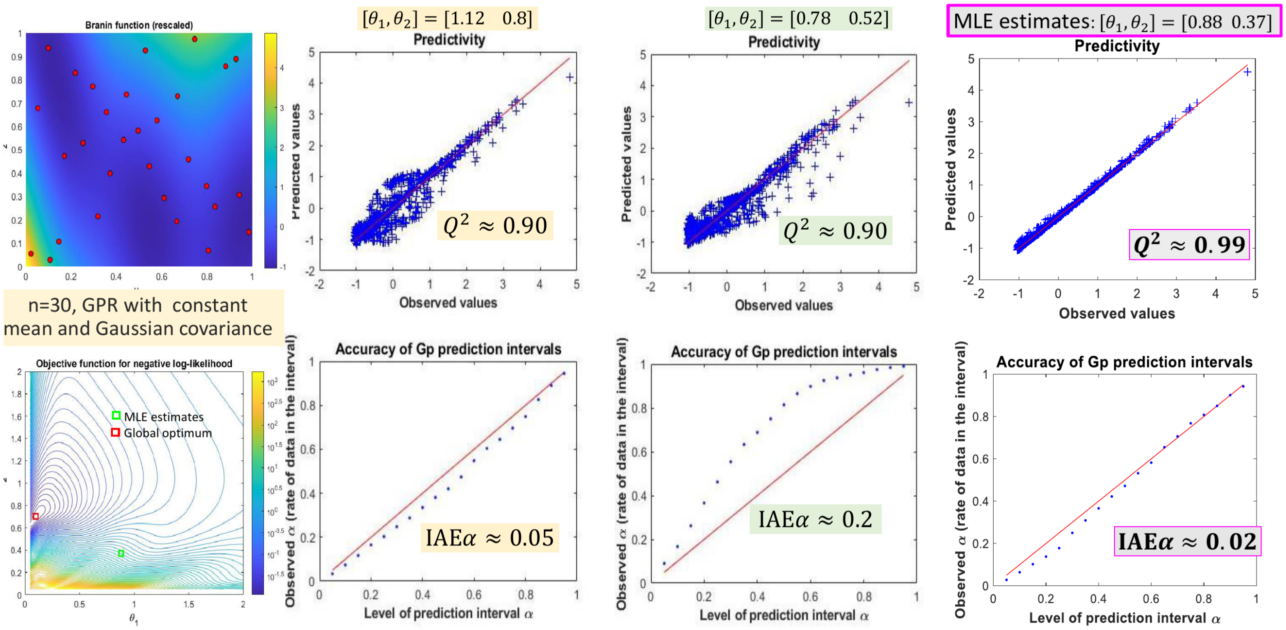


GPR validation

► **One message: Joint interpretation of Q^2 and IAEα !** [MI24a]

Criterion	Values	Interpretation
IAEα	**Value close to 0**	**Only if Q^2 is also high**, reliable predicted confidence intervals
	High values, close to 0.5 e.g.	Unreliable predicted prediction intervals ("underconfident" or "overconfident" model) ⇒ **Explanation from cross-interpretation with Q^2 and α-PI plot**

Illustration of criteria for GPR validation [MI24a]



Conclusions and remaining challenges

Conclusions

► **Recommendations for an efficient GP Regression:**

- ✓ GPR benefits greatly from **preliminary HSIC-based screening**
- ✓ GPR calls for **robust estimation of hyperparameters**
- ✓ Particular attention must be paid to **GP validation**

⇒ All these recipes are integrated in **ICSCREAM methodology** [MIC21]

⇒ Part of a more general effort to **ensure confidence in machine learning for UQ**

► **Interesting challenges for UQ applications**

- ✓ **High dimensional** problems (for example beyond 30 to 50 inputs and screening-free)
- ✓ Extension to **more complex inputs** (graph structure, e.g.)
- ✓ Learning **outputs with highly irregular**, or even **chaotic behavior** (due to physical threshold phenomena and phenomenological bifurcations, for instance)

cea

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, Preprint https://hal.science/cea-04322810**

[MI24b] 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

Work partly funded by ANR SAMOURAI research project

anr agence nationale de la recherche SAMOURAI

EDF SAFRAN IFP Energies nouvelles cea MINES Saint-Étienne CentraleSupélec GERAD

Simulation **A**nalytics and **M**eta-model-based solutions for **O**ptimization, **U**ncertainty and **R**eliability **A**nalys**I**s

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.

cea

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.

[MIC21] Marrel, Iooss and Chabridon, (2021). Statistical identification of penalizing configurations in high-dimensional thermalhydraulic numerical experiments: The ICSCREAM methodology, Nuclear Science & Engineering.

[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+22] Petit, S., Bect, J., Feliot, P., and Vazquez, E. (2022). Model parameters in GP interpolation: an empirical study of selection criteria. Preprint - https://hal-centralesupelec.archives-ouvertes.fr/hal-03285513.

[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