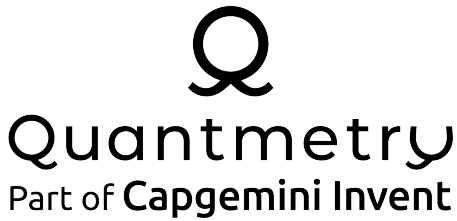
Conformal prediction for surrogate modelling in the UQ framework

Edgar Jaber, Vincent Blot

ETICS 2023 - PhD Seminar

Supervisors : N. Brunel (Quantmetry),
V. Chabridon, E. Remy, B. Iooss (EDF R&D),
D. Lucor (LISN), M. Mougeot (Centre Borelli)



Sommaire

Problem formulation

For a computer code $g : X \; ! \; Y$ used in industrial applications, the Uncertainty Quantification (UQ) methodology aims at evaluating how uncertainty on the inputs $X \; 2 \; X$ affects our knowledge of the output $g(X) \; 2 \; Y$ [De Rocquigny et al., 2008].

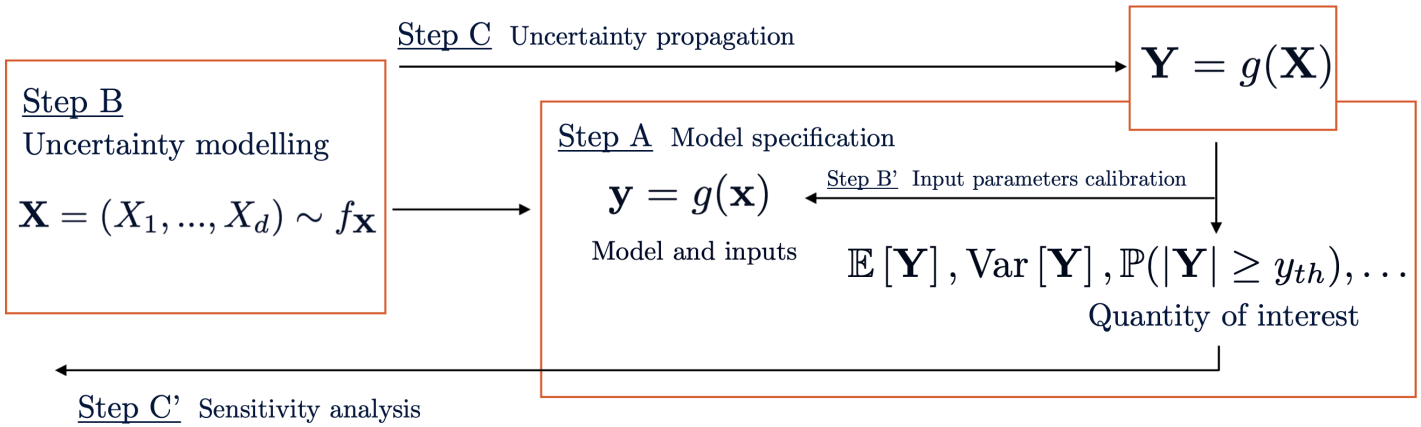


Figure – General UQ methodology.

Problem formulation

Since these codes are time-costly, surrogates g are constructed for performing heavy simulations like Monte-Carlo batch runs for Step C, C'.

For objective reasons, one needs to assess the *quality* of these surrogates.

We propose to use the Conformal prediction paradigm [Vovk et al., 2005] which is a generic, model-agnostic theory allowing to build *prediction* sets to these surrogates with frequentist coverage guarantees.

Sommaire

CP in regression setting

[Vovk et al., 2005]. We fix a probability space (; F ; P).

Definition

Let $X ; Y$ be metric spaces and $Z = X \quad Y$. Let $n \; 2$ N and $D = fZ_1; : : : ; Z_n g \; 2 \; 2^Z$ a training sample. For $\quad 2 \; (0; 1)$ a conformal predictor of coverage $\quad$ is any measurable function of the form :

$$\begin{aligned} C \; : 2^Z \quad X \; ! \; 2^Y \\ (D; X) \; 7! \; C_{n;} \; (X); \end{aligned} \tag{1}$$

such that for a new point $Z_{n+1} = (X_{n+1}; Y_{n+1}) \; 2 \; Z$:

$$\mathrm{P}\,(Y_{n+1} \; 2 \; C_{n;} \; (X_{n+1})) \quad 1 \quad : \tag{2}$$

Three main methods exist to estimate these set-functions : full-conformal, split-conformal and cross-conformal estimators. We focus on the latter.

Full-Conformalized Ridge Regression (CRR)

[Vovk et al., 2005] A non-conformity score is any measurable function of the form :

$$\begin{aligned} A : 2^Z \quad Z \quad ! \quad \overline{\mathrm{R}} \\ (D;Z) \; 7! \; A(D;Z): \end{aligned} \tag{3}$$

Assume $X = Y = \mathrm{R}$ and R_D a regression rule learned on $D = fZ_1;:::;Z_ng = f(X_1;Y_1);:::;(X_n;Y_n:)g$. A straightforward non-conformity score is given by the di□erence between prediction and actual value, for all $i = 1;:::;n$:

$$A(D;Z_i) = Y_i \quad R_D(X_i) = Y_i \quad \text{Þ}_i \tag{4}$$

Full-Conformalized Ridge Regression (CRR)

For a new input $X_{n+1} \in \mathcal{X}$ and output $Y \in \mathcal{Y}$, we denote by $\mathcal{D}^{\theta} := Dat(X_{n+1}; Y)$. The full conformal predictor is given for all confidence level $0 < \alpha < 1$ by :

$$\hat{C}^{CRR}_{n;\alpha}(X_{n+1}) = \left\{ Y \in \mathcal{Y} : \frac{\#\{i : A(\mathcal{D}^{\theta}; Z_i) \geq A(\mathcal{D}^{\theta}; (X_{n+1}; Y))\}}{n} > \alpha \right\}$$

This estimator is based on the "dissimilarity" of the new pair inside the updated dataset. It achieves the required coverage property.

It is computationally intractable due to the full grid search on the label space $\mathcal{Y}$.

To tackle this, more advanced estimators are proposed in the following.

In the rest of this talk we suppose that $Y = g(X)$ where g is some deterministic function (e.g a numerical simulation code).

The Jackknife+ estimator

[Barber et al., 2021]. We train a metamodel $\hat{g}$ on D and n leave-one-out metamodels $\hat{g}_{-i}$ on $D_n(X_i; g(X_i))$. The Jackknife+ estimator is given by :

$$\hat{C}^{J+}_{n;}(X_{n+1}) = \quad q_{n;} \quad \hat{g}_{-i}(X_{n+1}) \quad R_i^{LOO} \quad ; q^{+}_{n;} \quad \hat{g}_{-i}(X_{n+1}) + R_i^{LOO} \tag{5}$$

Where the leave-one-out error defined by :

$$R_i^{LOO} := jg(X_i) \quad \hat{g}_{-i}(X_i)j \tag{6}$$

With this estimator we have a coverage guarantee of 1 2 .

However, these intervals have almost constant width for any new point.

CV+ estimator

Let $K \in \{1,\dots,n\}$. We divide $\mathcal{D}^{\text{train}}$ in K-disjoint subsets $S_1 \cup \dots \cup S_K$.

We fit K metamodels with the k-th fold removed : $\hat{g}_{-S_k}$.

We compute the conformity scores :

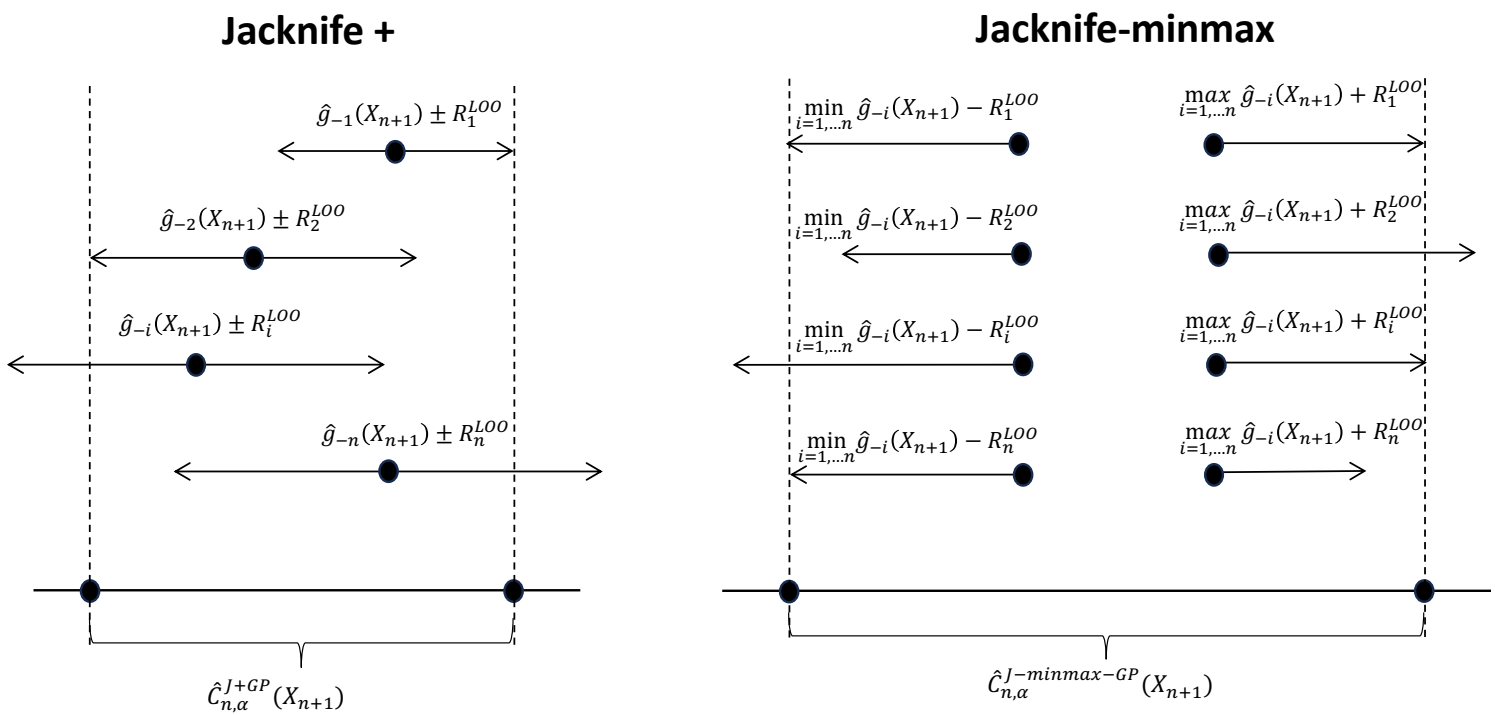
$$R_i^{CV} := |g(X_i) - \hat{g}_{-S_{k(i)}}(X_i)|, \quad \forall i \in \{1,\dots,n\}, \tag{7}$$

where $S_{k(i)}$ is the fold containing X_i.

We estimate our prediction intervals :

$$\hat{C}_{n,\alpha}^{\text{CV+}}(X_{n+1}) = \left[\hat{q}_{n,\alpha}^{-} \left\{ \hat{g}_{-S_{k(i)}}(X_{n+1}) - R_i^{CV} \right\}, \hat{q}_{n,\alpha}^{+} \left\{ \hat{g}_{-S_{k(i)}}(X_{n+1}) + R_i^{CV} \right\} \right] \tag{8}$$

J+/minmax



$$with\ R_i^{LOO} = |Y_i - \hat{g}_{-i}(X_i)|$$

Jacknife/CV minmax

Additionally, we can add minmax on both the J+ and the CV+ methods by replacing the metamodel prediction with :

$$\hat{y}_{-i}(X_{n+1}) \quad ! \quad \min_i \hat{y}_{-i}(X_{n+1}); \max_i \hat{y}_{-i}(X_{n+1}): \tag{9}$$

$$\hat{y}_{-S_k(i)}(X_{n+1}) \quad ! \quad \min_i \hat{y}_{-S_k(i)}(X_{n+1}); \max_i \hat{y}_{-S_k(i)}(X_{n+1}): \tag{10}$$

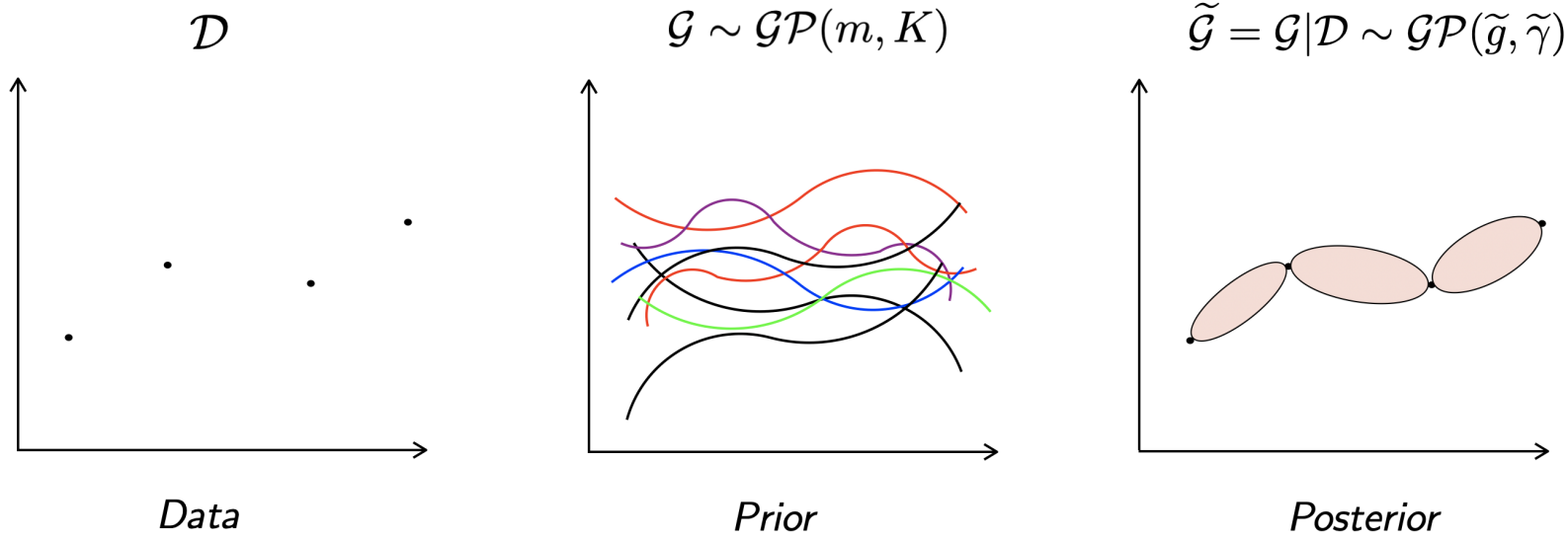
The intervals will not be centered anymore and we have the coverage guarantee [Barber et al., 2021] :

$$\mathrm{P}\left(g(X_{n+1}) \; 2 \; \hat{C}^{+\text{-mm}}_{n;}(X_{n+1}) \right) \quad 1 \quad : \tag{11}$$

Moreover, the resulting intervals will be more conservative.

Sommaire

Gaussian Process Regression (GPR)



Bayesian credibility intervals

Since we are in a Gaussian setting, we can have access to the confidence intervals for any new point $X_{n+1} \in X \setminus X$:

$$CR_\alpha(X_{n+1}) = \left[\hat{g}(X_{n+1}) \pm F^{-1}\left(1 - \frac{\alpha}{2}\right) \hat{e}(X_{n+1}) \right] \tag{12}$$

which in our case are credibility intervals. Here F is the CDF of the standard normal distribution.

If g was *truly* drawn from our posterior $GP(X; g(X))$, then we would have the exact coverage :

$$P(g(X_{n+1}) \in CR_\alpha(X_{n+1})) = 1 - \alpha \quad ; \tag{13}$$

and $CR_\alpha(X_{n+1})$ would be the true prediction interval for the true g.

Frequentist prediction intervals

In practice we don't have access to the true distribution $P_{(X;g(X))}$. If we dispose of a test dataset :

$$D^{test} = f(X_1; g(X_1)); : : : ; (X_m; g(X_m))g; \tag{14}$$

di erent from D, then in all generality we cannot expect to have :

$$\frac{1}{m} \sum_{i=1}^{m} 1\, fg(X_i)\ 2\ CR\ (X_i)g \quad 1 \quad ; \text{ a.s:} \tag{15}$$

Sommaire

The Burnaev-Wasserman program

[Burnaev and Vovk, 2014] Assume that X R^d, for all i, X_i 2 $L^2(\ \ )$ and the model g is truly gaussian. The credibility sets have exact coverage and output an interval of the form :

$$CR\ \ (X_{n+1}) = [B\ ; B\ ] \tag{16}$$

The CRR method with the GPR rule outputs a prediction interval of the form :

$$\mathrm{b}C_{n;}^{CRR}(X_{n+1}) = [C\ ; C\ ]: \tag{17}$$

A natural question is to compare the differences of the bounds of these two intervals and their asymptotic behaviour

An asymptotic result

See [Burnaev and Vovk, 2014] for a proof of the following.

Theorem

Under the above assumptions we get

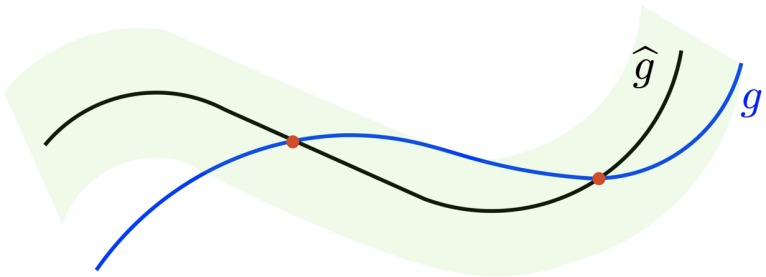
$$p_{\overline{n}}(B \quad C) \overset{Law}{\underset{n \to 1}{!}} N(0; h(\ )); \tag{18}$$

and similarly for the lower-bound.

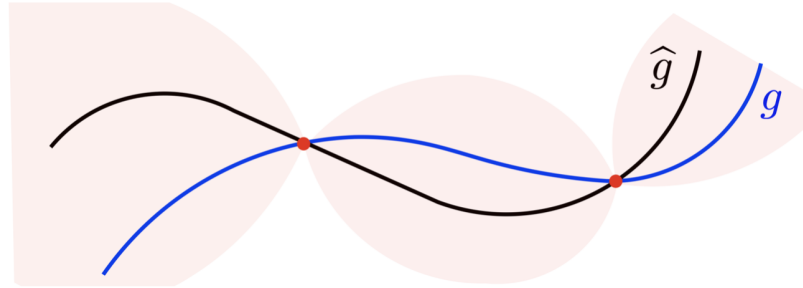
Here h is a function of the $1 \quad =2$-quantile of the standard normal distribution and of the mean and variance of the input distribution.

Problem formulation

Conformal Prediction



Gaussian Processes



Adaptive Conformal Prediction for better Uncertainty Quantification

$$\mathrm{P}\ \ g(X)\ 2\ C_{\widehat{g};}\ \ (X) \qquad 1 \qquad : \qquad (19)$$

J+GP

The idea is to adapt the Jackknife+ method presented earlier to GPR metamodels to have adaptive prediction intervals. We denote these predictors as $\hat{C}_{n;}^{J+GP}$ and can prove the following theorem :

Theorem

Assume $D = (X; g(X))$ *is exchangeable. For a new point* X_{n+1} 2 X nX *and a coverage level* 2 $(0; 1)$ *:*

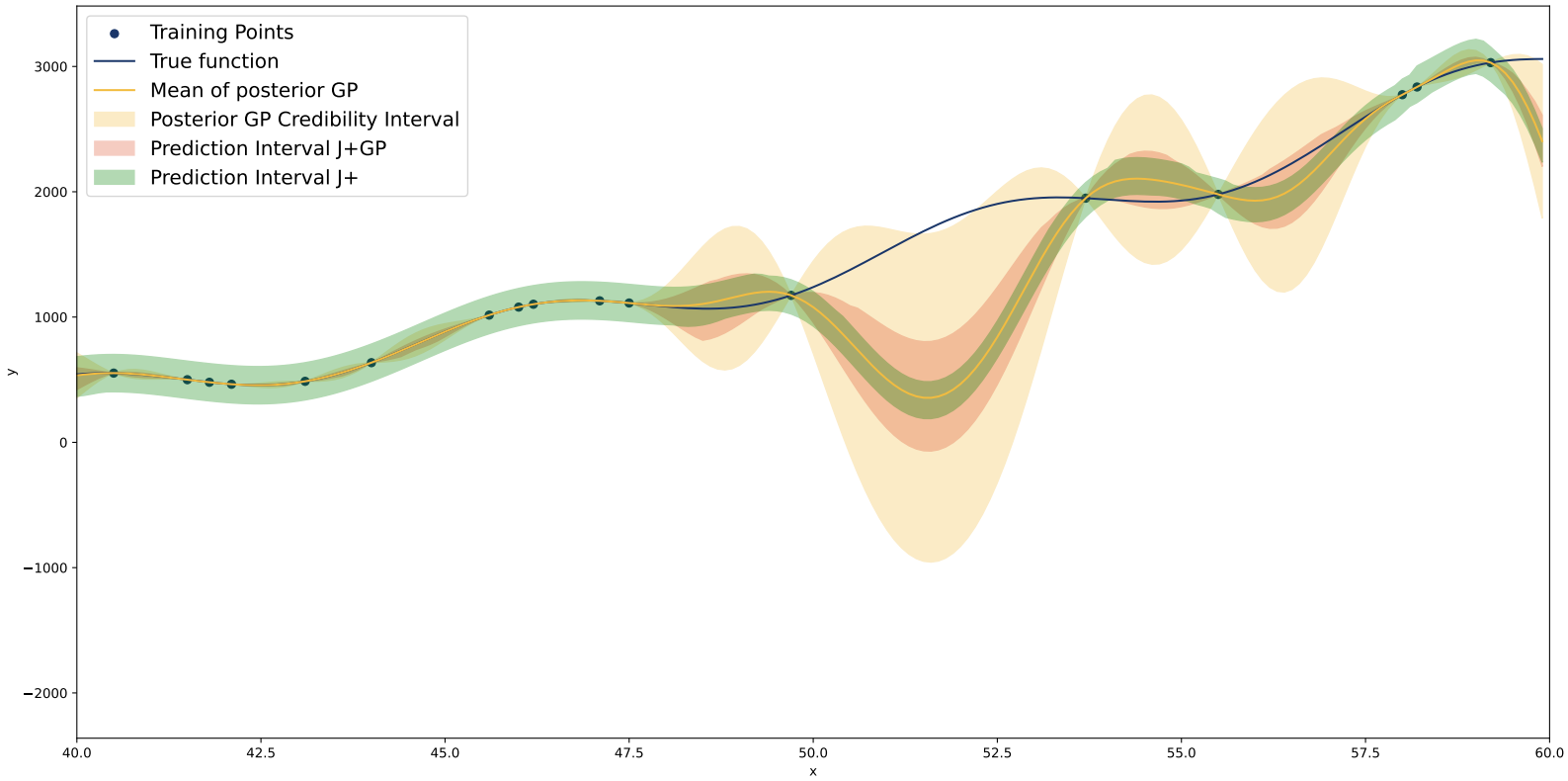
$$P \quad g(X_{n+1}) \; 2 \; \hat{C}_{n;}^{J+GP}(X_{n+1}) \qquad 1 \quad 2 \; : \tag{20}$$

Example on a toy function

The Vincent Blot© highly original g_{VB} :

$$g_{VB}(x) = 3x\sin(x) \quad 2x\cos(x) + \frac{x^3}{40} \quad \frac{1}{2}x^2 \quad 10x; \quad (21)$$

which we sample at ease.



Sommaire

Conclusion

Conformal prediction is a method for performing distribution-free uncertainty quantification of machine learning algorithms.

It can be used in a Bayesian regression setting for testing the soft-assumption on the original model.

Prediction intervals for Gaussian Processes using an adaptation of the CP algorithms can be built (upcoming paper [Jaber and Blot, 2023] - in progress).

A more robust uncertainty quantification of this type of surrogates can be achieved and can thus serve in assessing their quality for the purpose of performing industrial UQ.

References I

Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani., R. J. (2021). Predictive inference with the jackknife+. *Annals of Statistics, 486-507*, 49.

Burnaev, E. and Vovk, V. (2014). Efficiency of conformalized ridge regression. In *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 605–622.

De Rocquigny, E., Devictor, N., and Tarantola, S. (2008). *Uncertainty in industrial practice - A guide to quantitative uncertainty managment*. Wiley and Sons.

Jaber, E. and Blot, V. (2023). Conformalizing Gaussian Processes For More Robust Uncertainty Quantification. *In progress*.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer, springer edition.