Sensitivity analysis to probabilistic model identification and statistical estimation

GST Mécanique et Incertain - Toulouse, France.

Charles Surget[1,2]
charles.surget@onera.fr
Sylvain Dubreuil[1], Jérôme Morio[1], Cécile Mattrand[2], Jean-Marc Bourinet[2], Nicolas Gayton[2].

19th of October 2023

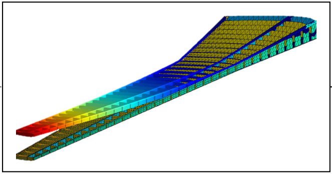
[1] ONERA/DTIS, Université de Toulouse, Toulouse, F-31055, France
[2] Université Clermont Auvergne, CNRS, SIGMA Clermont, Clermont-Ferrand, F-63000, France

Table of contents

Context - Uncertainty Quantification

Simulation code of a mechanical structure:

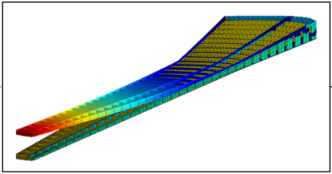
q parameters
$(\mathrm{E}, \nu, \mathrm{L}, ...)$



An output quantity
$(\sigma, \mathrm{u}, ...)$

Context - Uncertainty Quantification

Simulation code of a mechanical structure:

q parameters $(\mathrm{E}, \nu, \mathrm{L}, ...)$ →  → An output quantity $(\sigma, \mathrm{u}, ...)$

In an uncertainty quantification context, those parameters are considered as an input continuous random vector:

$\boldsymbol{X}$ → Black Box ϕ → Y

with $\boldsymbol{X} = (X_1, ..., X_q)^t$ with values on the domain $\mathcal{X} \subseteq \mathbb{R}^q$ and defined by a given Probability Density Function (PDF) $f_{\boldsymbol{X}}$.

Context - 1^{st} uncertainty source

One could be interested in assessing the following expectation of a particular function τ of $Y = \phi(\boldsymbol{X})$ (e.g. a mean or a probability of failure):

$$\mathbb{E}_{f_{\boldsymbol{X}}}\left[\tau\left(\phi\left(\boldsymbol{X}\right)\right)\right] = \int_{\mathcal{X}} \tau\left(\phi\left(\boldsymbol{x}\right)\right) f_{\boldsymbol{X}}\left(\boldsymbol{x}\right) \mathrm{d}\boldsymbol{x}. \quad (1.1)$$

Context - 1^{st} uncertainty source

One could be interested in assessing the following expectation of a particular function τ of $Y = \phi(\boldsymbol{X})$ (e.g. a mean or a probability of failure):

$$\mathbb{E}_{f_{\boldsymbol{X}}}\left[\tau\left(\phi\left(\boldsymbol{X}\right)\right)\right] = \int_{\mathcal{X}} \tau\left(\phi\left(\boldsymbol{x}\right)\right) f_{\boldsymbol{X}}\left(\boldsymbol{x}\right) \mathrm{d}\boldsymbol{x}. \tag{1.1}$$

The Monte Carlo (MC) estimator of this integral is given by (assuming $\tau = \mathrm{Id}$ for illustration):

$$\hat{\mu}^{MC} = \frac{1}{N_{\boldsymbol{X}}} \sum_{j=1}^{N_{\boldsymbol{X}}} \phi\left(\boldsymbol{X}^{(j)}\right), \tag{1.2}$$

with $\boldsymbol{X}^{(j)} \overset{i.i.d.}{\sim} f_{\boldsymbol{X}}$ and $N_{\boldsymbol{X}}$ the size of the sample of simulations.

Context - 1^{st} uncertainty source

One could be interested in assessing the following expectation of a particular function τ of $Y = \phi(\boldsymbol{X})$ (e.g. a mean or a probability of failure):

$$\mathbb{E}_{f_{\boldsymbol{X}}}\left[\tau\left(\phi\left(\boldsymbol{X}\right)\right)\right] = \int_{\mathcal{X}} \tau\left(\phi\left(\boldsymbol{x}\right)\right) f_{\boldsymbol{X}}\left(\boldsymbol{x}\right) \mathrm{d}\boldsymbol{x}. \tag{1.1}$$

The Monte Carlo (MC) estimator of this integral is given by (assuming $\tau = \mathrm{Id}$ for illustration):

$$\hat{\mu}^{MC} = \frac{1}{N_{\boldsymbol{X}}} \sum_{j=1}^{N_{\boldsymbol{X}}} \phi\left(\boldsymbol{X}^{(j)}\right), \tag{1.2}$$

with $\boldsymbol{X}^{(j)} \overset{i.i.d.}{\sim} f_{\boldsymbol{X}}$ and $N_{\boldsymbol{X}}$ the size of the sample of simulations. A first uncertainty source is related to this sample, defined as $\widetilde{\boldsymbol{X}}$ in the following process:

$$\widetilde{\boldsymbol{X}} = \{\boldsymbol{X}^{(j)}, j = 1, \ldots, N_{\boldsymbol{X}}\} \longrightarrow \hat{\mu}^{MC}$$

Context - 2^{nd} uncertainty source

In a realistic context, the PDF $f_{\boldsymbol{X}}$ may be unknown [1]. Thus, the probabilistic model has to be inferred from experimental tests:

$$\widetilde{\boldsymbol{D}} = \{\boldsymbol{D}^{(i)}, i = 1, \dots, N_{\boldsymbol{D}}\}$$
with $\boldsymbol{D}^{(i)} \overset{i.i.d.}{\sim} f_{\boldsymbol{X}}$

$\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}}$

with $N_{\boldsymbol{D}}$ the size of the sample of experiments $\widetilde{\boldsymbol{D}}$. The estimation $\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}}$ [2, 3] of the PDF $f_{\boldsymbol{X}}$ induces a second uncertainty source related to $\widetilde{\boldsymbol{D}}$.

[1] G Sarazin. Analyse de sensibilité fiabiliste en présence d'incertitudes épistémiques introduites par les données d'apprentissage. PhD thesis, Toulouse, ISAE, 2021.
[2] James K Lindsey *et al.* Parametric statistical inference. Oxford University Press, 1996.
[3] A J Izenman. Review papers: Recent developments in nonparametric density estimation. Journal of the american statistical association, 86(413):205-224, 1991.

Context - 2^{nd} uncertainty source

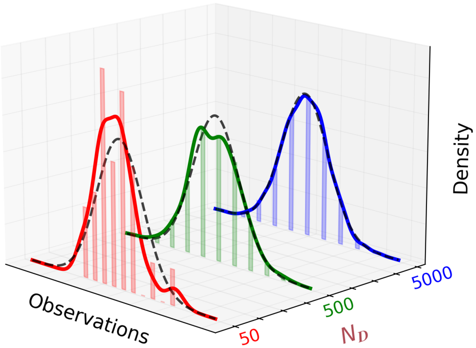


Figure 1.1: Impact of the size of the sample of experiments (a) on the identification of the probabilistic model and (b) on the mean estimate for an univariate Gaussian distribution.

Context - Problematics

$$\widetilde{\boldsymbol{D}} = \{\boldsymbol{D}^{(i)}, i = 1, \dots, N_{\boldsymbol{D}}\}$$
with $\boldsymbol{D}^{(i)} \overset{i.i.d.}{\sim} f_{\boldsymbol{X}}$

↓

$$\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}}$$

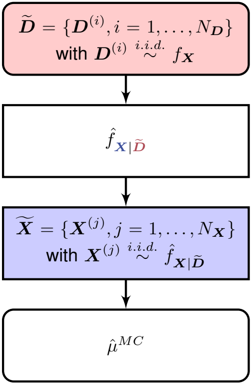
↓

$$\widetilde{\boldsymbol{X}} = \{\boldsymbol{X}^{(j)}, j = 1, \dots, N_{\boldsymbol{X}}\}$$
with $\boldsymbol{X}^{(j)} \overset{i.i.d.}{\sim} \hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}}$

↓

$$\hat{\mu}^{MC}$$

Context - Problematics



Problem A

How to take into account the uncertainty of the sample of experiments in the variance of the estimator?

Context - Problematics

$\widetilde{\boldsymbol{D}} = \{\boldsymbol{D}^{(i)}, i = 1, \ldots, N_{\boldsymbol{D}}\}$ with $\boldsymbol{D}^{(i)} \overset{i.i.d.}{\sim} f_{\boldsymbol{X}}$

↓

$\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}}$

↓

$\widetilde{\boldsymbol{X}} = \{\boldsymbol{X}^{(j)}, j = 1, \ldots, N_{\boldsymbol{X}}\}$ with $\boldsymbol{X}^{(j)} \overset{i.i.d.}{\sim} \hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}}$

↓

$\hat{\mu}^{MC}$

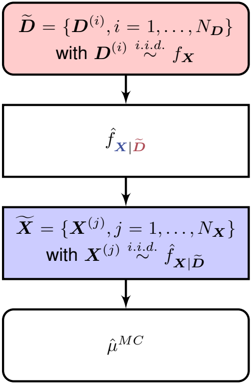
Problem A

How to take into account the uncertainty of the sample of experiments in the variance of the estimator?

Problem B

In order to improve efficiently the accuracy of the estimator, should the data enrichment be made in the sample of experiments or the sample of simulations?

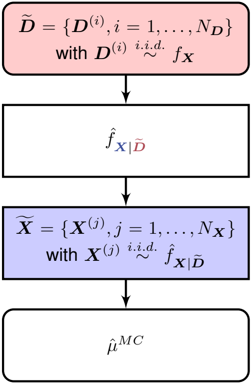
Context - Small-Data



- Limited size $N_{\boldsymbol{D}}$:
  the small-data context is imposed by costly physical experiments.

- Limited size $N_{\boldsymbol{X}}$:
  the small-data context is imposed by the simulation time induced by the model complexity.

Context - Small-Data

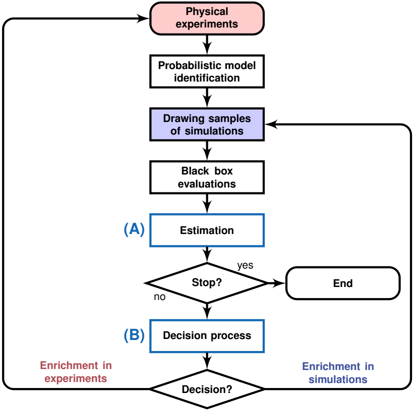


- Limited size $N_{\boldsymbol{D}}$:
  the small-data context is imposed by costly physical experiments.

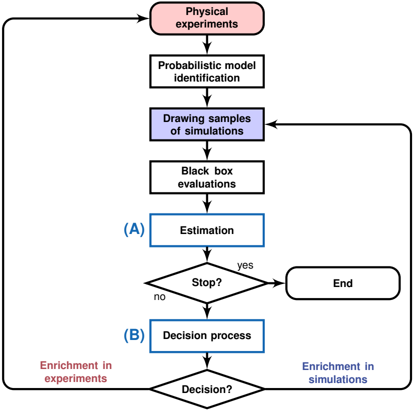
- Limited size $N_{\boldsymbol{X}}$:
  the small-data context is imposed by the simulation time induced by the model complexity.

Physical experiments-Simulations
trade-off

Context - Goal



Context - Goal

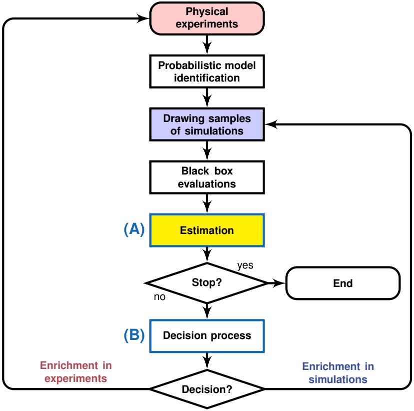


The desired method gathers:

- A trade-off with negligible numerical cost.
- A guarantee of the robustness of the estimate.
- A reuse of data at subsequent enrichment stages.

Table of contents

Proposed approach - Estimation



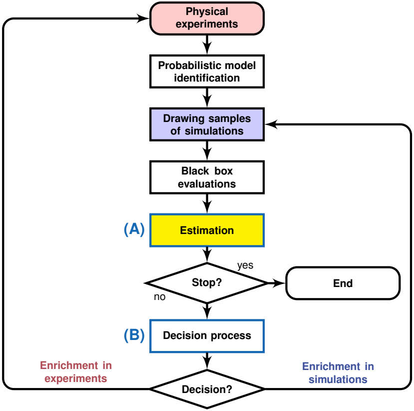
Problem A

How to take into account the uncertainty of the sample of experiments in the variance of the estimator?

[4] V Chabridon. Analyse de sensibilité fiabiliste avec prise en compte d'incertitudes sur le modèle probabiliste-Application aux systèmes aérospatiaux. PhD thesis, UCA(2017-2020), 2018.

[5] A Owen and Y Zhou. Safe and effective importance sampling. Journal of the American Statistical Association,95(449):135-143, 2000.

Proposed approach - Estimation



[4] V Chabridon. Analyse de sensibilité fiabiliste avec prise en compte d'incertitudes sur le modèle probabiliste-Application aux systèmes aérospatiaux. PhD thesis, UCA(2017-2020), 2018.

[5] A Owen and Y Zhou. Safe and effective importance sampling. Journal of the American Statistical Association,95(449):135-143, 2000.

Problem A

How to take into account the uncertainty of the sample of experiments in the variance of the estimator?

Expectation with double integral:

$$\mathbb{E}_{f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}}\left[\phi\left(\boldsymbol{X}\right)\right] = \int_{\mathcal{X}^{N_D}} \int_{\mathcal{X}} \phi\left(\boldsymbol{x}\right) f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}(\boldsymbol{x},\widetilde{\boldsymbol{d}})\mathrm{d}\boldsymbol{x}\ \mathrm{d}\widetilde{\boldsymbol{d}}.$$

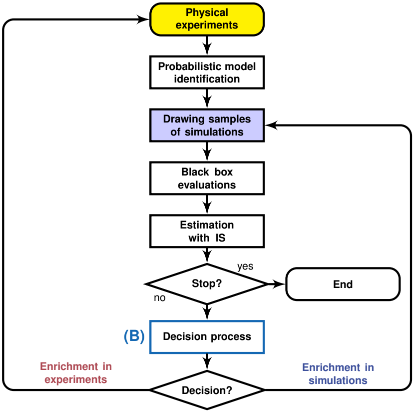
Nested estimator with Importance Sampling:

$$\hat{\mu}^{N-IS} = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{N_{\boldsymbol{X}}}\sum_{j=1}^{N_{\boldsymbol{X}}}\phi\left(\boldsymbol{X}_k^{(j)}\right)\frac{\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}_k}(\boldsymbol{X}_k^{(j)})}{g(\boldsymbol{X}_k^{(j)})}$$

$$= \frac{1}{N}\sum_{k=1}^{N}\hat{\mu}_k^{IS},$$

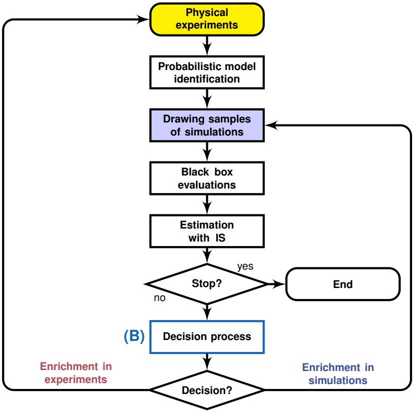
with $\boldsymbol{X}_k^{(j)} \overset{i.i.d.}{\sim} g$ and N the number of samples of experiments [4, 5].

Proposed approach - Small-data context



In a small-data context, only one N_D-sample $\widetilde{D}$ of limited size is available.

Proposed approach - Small-data context



In a small-data context, only one N_D-sample $\widetilde{D}$ of limited size is available.

Resampling method

Allows to generate N samples of experiments from an initial one [6, 7]. (e.g. Bootstrap [BS])

Solution A

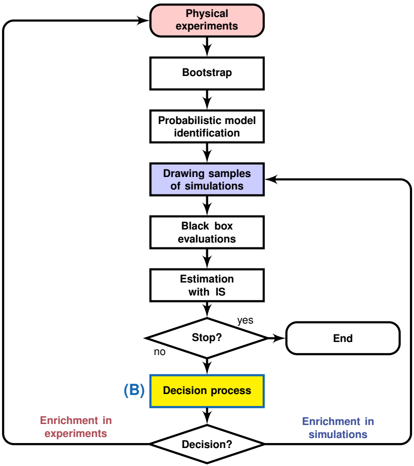
The nested estimator is conditioned on the initial sample of experiment but the uncertainty related to it is considered.

Variance of estimation:

$$\mathbb{V}_{f_{(\boldsymbol{X},\widetilde{D})}}\left[\hat{\mu}^{N-IS}\right] = \frac{1}{N}\mathbb{V}_{f_{(\boldsymbol{X},\widetilde{D})}}\left[\hat{\mu}^{IS}\right].$$

[6] C H Yu. Resampling methods: concepts, applications, and justification. Practical Assessment, Research, and Evaluation, 8(1):19, 2002.
[7] B Efron. The jackknife, the bootstrap and other resampling plans. SIAM, 1982.

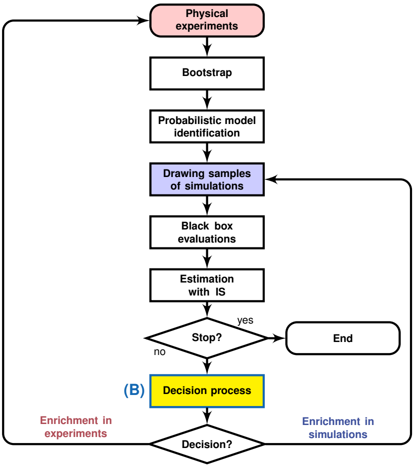
Proposed approach - Sensitivity analysis



Problem B

In order to improve efficiently the accuracy of the estimator, should the data enrichment be made in the sample of experiments or the sample of simulations?

Proposed approach - Sensitivity analysis



Problem B

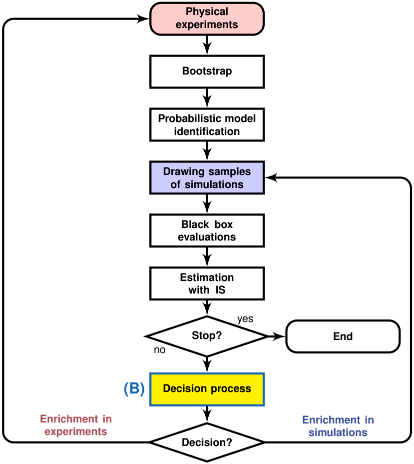
In order to improve efficiently the accuracy of the estimator, should the data enrichment be made in the sample of experiments or the sample of simulations?

An ANalysis Of VAriance [8, 9] is performed:

[8] I M Sobol'. Sensitivity analysis for non-linear mathematical models. Mathematical modelling and computational experiment, 1:407-414, 1993.

[9] F Gamboa *et al.* Statistical inference for Sobol pick-freeze Monte Carlo method. Statistics, 50(4):881–902, 2016.

Proposed approach - Sensitivity analysis



Problem B

In order to improve efficiently the accuracy of the estimator, should the data enrichment be made in the sample of experiments or the sample of simulations?

An ANalysis Of VAriance [8, 9] is performed:

$$S_{\boldsymbol{D}} = \frac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{IS} | \widetilde{\boldsymbol{D}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{IS}\right]} \qquad S_{\boldsymbol{X}} = \frac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{IS} | \widetilde{\boldsymbol{X}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{IS}\right]}$$

Interpretation of Sobol' indices:

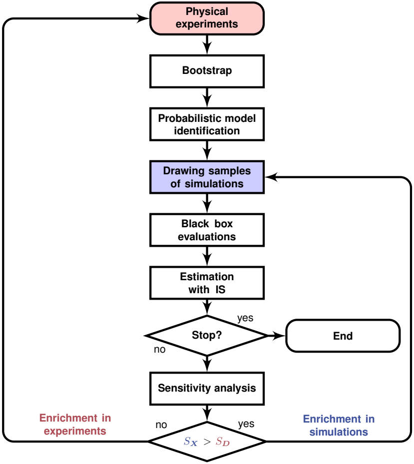
$S_{\boldsymbol{D}} \rightarrow$ proportion due to the sample of experiments

$S_{\boldsymbol{X}} \rightarrow$ proportion due to the sample of simulations

[8] I M Sobol'. Sensitivity analysis for non-linear mathematical models. Mathematical modelling and computational experiment, 1:407-414, 1993.

[9] F Gamboa *et al.* Statistical inference for Sobol pick-freeze Monte Carlo method. Statistics, 50(4):881–902, 2016.

Proposed approach - Synthesis



The proposed approach:

A) Takes into account the uncertainty of the sample of experiments.

B) Guides the data enrichment in the driving source of uncertainty.

C) Faces a minimized cost, equivalent to a classic MC method.

D) Updates itself to guarantee the robustness of the estimate.

Table of contents

Illustrations - Cantilever Beam

Mean deflection of the free end of a cantilever beam:

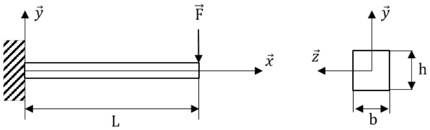


Figure 3.1: Representation of a cantilever beam where F is the transverse load applied on the free end of the beam of length L, Young's modulus E and cross-section bh.

$$\phi\left(F, L, E, b, h\right) = \frac{4FL^3}{Ebh^3}$$

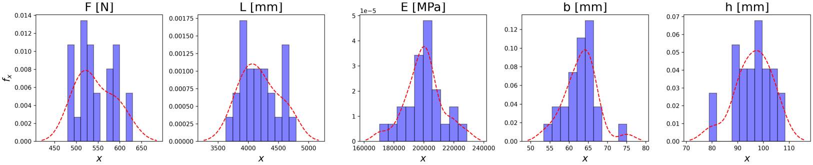
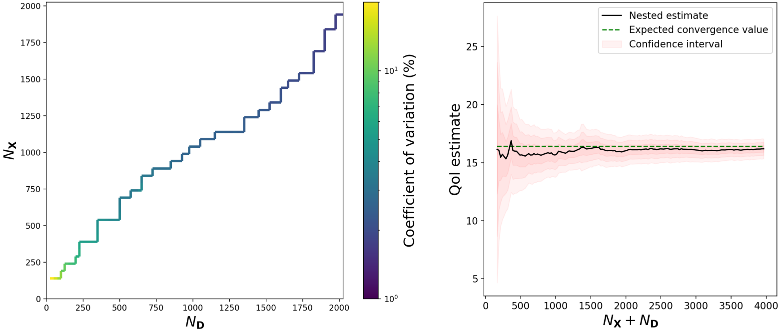


Figure 3.2: Histogram of each marginal distribution built on a sample of experiments of size $N_D = 25$ [10].
A probabilistic model is identified on this sample (in red).

[10] L Baoyu *et al*. Reliability analysis based on a novel density estimation method for structures with correlations. Chinese Journal of Aeronautics, 30(3):1021-1030, 2017.

Illustrations - Trajectory of enrichment

Cantilever beam toy case ($q = 5$) with non-parametric identification (Mean):



(a) **(b)**

Figure 3.3: Approach applied to a test case of a cantilever beam of dimension $q = 5$ with a sample of experiment of initial size $N_D = 25$ and a sample of simulations of initial size $N_X = 140$. The enrichment step is set to 20. (a) Trajectory of enrichment taken by the algorithm for the cantilever beam toy case. (b) Estimation of the Quantity of Interest (QoI) during data enrichment. The expected convergence value is $\mathbb{E}[\phi(\boldsymbol{X})] \approx 16.403$[mm].

Illustrations - Near optimal trajectory

Cantilever beam toy case ($q = 5$) with non-parametric identification (Mean):

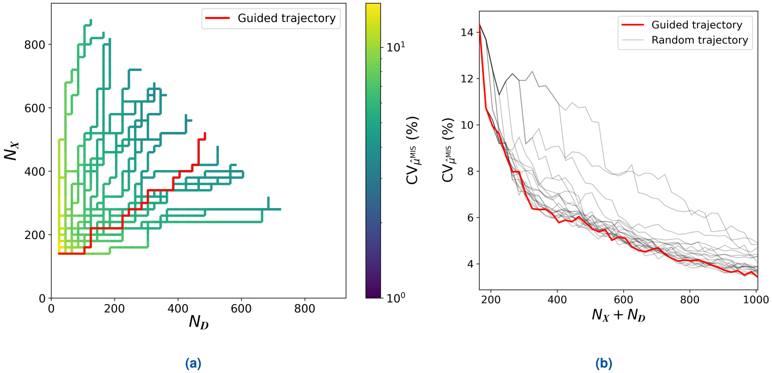


Figure 3.4: Comparison between the guided trajectory and random trajectories of enrichment. (a) Trajectory of enrichment guided by sensitivity analysis (in red) alongside 20 random trajectories. (b) Coefficient of variation assessed during data enrichment. *The coefficient of variation related to the guided trajectory (in red) is one of the lowest.*

Illustrations - Near optimal trajectory

Cantilever beam toy case ($q = 5$) with non-parametric identification (Mean):

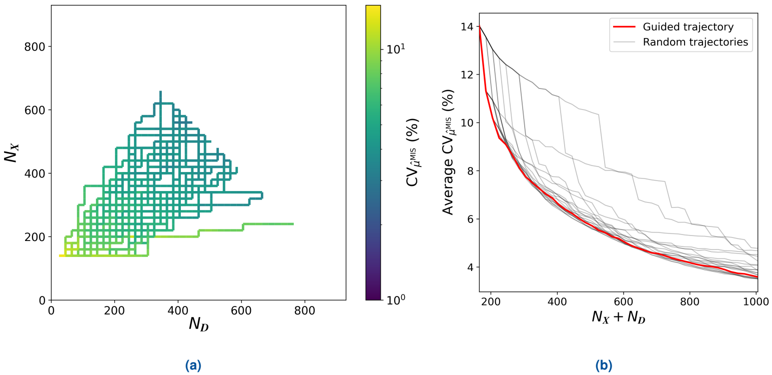
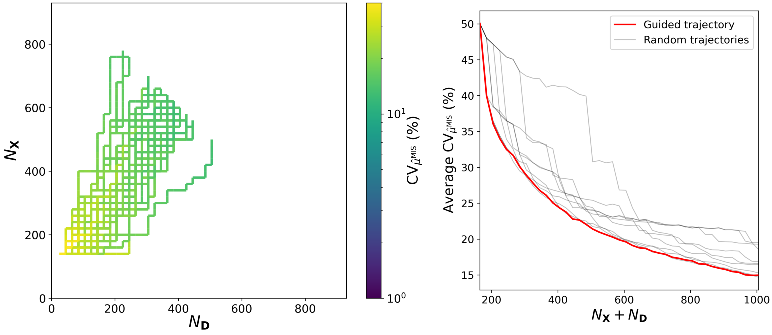


Figure 3.5: Repetition of the approach. (a) Trajectory of enrichment guided by sensitivity analysis obtained on 40 repetitions. (b) Average coefficient of variation assessed on 40 repetitions during data enrichment. *The coefficient of variation related to the guided trajectory (in red) is one of the lowest*.

Illustrations - Impact of the QoI: failure probability

Virkler experiment toy case ($q = 3$) with parametric identification (Probability):



(a) **(b)**

Figure 3.6: Repetition of the approach on a Virkler experiment toy case. (a) Trajectory of enrichment guided by sensitivity analysis obtained on 20 repetitions. (b) Average coefficient of variation assessed on 20 repetitions during data enrichment. *The expected value* $\mathbb{P}[\tau(\phi(\boldsymbol{X})) \leqslant N_s] \approx 0.126$ *is obtained with a* 10^7 *MC sample.*

Table of contents

Conclusion

Framework

- The probabilistic model is **unknown** and is inferred from physical experiments,
- A **small-data context** is imposed by costly physical experiments and a costly black box function.

Conclusion

Framework

- The probabilistic model is **unknown** and is inferred from physical experiments,
- A **small-data context** is imposed by costly physical experiments and a costly black box function.

Current method

A) Takes into account the uncertainty of the sample of experiments,

B) Answers the physical experiments-simulations trade-off by guiding the investment of data in the driving source of uncertainty.

C) Faces a minimized cost, equivalent to a classic Monte Carlo method.

D) Updates itself to guarantee the robustness of the estimate.

Conclusion

Framework

- The probabilistic model is **unknown** and is inferred from physical experiments,
- A **small-data context** is imposed by costly physical experiments and a costly black box function.

Current method

A) Takes into account the uncertainty of the sample of experiments,

B) Answers the physical experiments-simulations trade-off by guiding the investment of data in the driving source of uncertainty.

C) Faces a minimized cost, equivalent to a classic Monte Carlo method.

D) Updates itself to guarantee the robustness of the estimate.

Perspectives

- Improvement of the approach: data quantification, budget, ...
- Application to buckling tests of thin-shell structures.

References I

[1] Gabriel Sarazin.
Analyse de sensibilité fiabiliste en présence d'incertitudes épistémiques introduites par les données d'apprentissage.
PhD thesis, Toulouse, ISAE, 2021.

[2] James K Lindsey et al.
Parametric statistical inference.
Oxford University Press, 1996.

[3] Alan Julian Izenman.
Review papers: Recent developments in nonparametric density estimation.
Journal of the american statistical association, 86(413):205–224, 1991.

References II

[4] Vincent Chabridon.
Analyse de sensibilité fiabiliste avec prise en compte d’incertitudes sur le modèle probabiliste-Application aux systèmes aérospatiaux.
PhD thesis, Université Clermont Auvergne(2017-2020), 2018.

[5] Art Owen and Yi Zhou.
Safe and effective importance sampling.
Journal of the American Statistical Association, 95(449):135–143, 2000.

[6] Chong Ho Yu.
Resampling methods: concepts, applications, and justification.
Practical Assessment, Research, and Evaluation, 8(1):19, 2002.

References III

[7] Bradley Efron.
The jackknife, the bootstrap and other resampling plans.
SIAM, 1982.

[8] Ilya M Sobol'.
Sensitivity analysis for non-linear mathematical models.
Mathematical modelling and computational experiment, 1:407–414, 1993.

[9] Fabrice Gamboa, Alexandre Janon, Thierry Klein, A Lagnoux, and Clémentine Prieur.
Statistical inference for Sobol pick-freeze Monte Carlo method.
Statistics, 50(4):881–902, 2016.

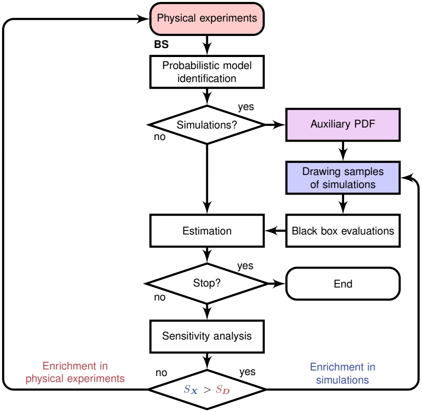
References IV

[10] LI Baoyu, Leigang Zhang, ZHU Xuejun, YU Xiongqing, and MA Xiaodong.
Reliability analysis based on a novel density estimation method for structures with correlations.
Chinese Journal of Aeronautics, 30(3):1021–1030, 2017.

[11] Anthony Brockwell.
Universal residuals: A multivariate transformation.
Statistics & probability letters, 77(14):1473–1478, 2007.

[12] Régis Lebrun and Anne Dutfoy.
Do rosenblatt and nataf isoprobabilistic transformations really differ?
Probabilistic Engineering Mechanics, 24(4):577–584, 2009.

Appendix A - IS and cost reduction



An update of the probabilistic model does not affect the sample of simulations $\widetilde{\mathbf{X}}$. An additional cost reduction is possible with Bootstrap (BS).

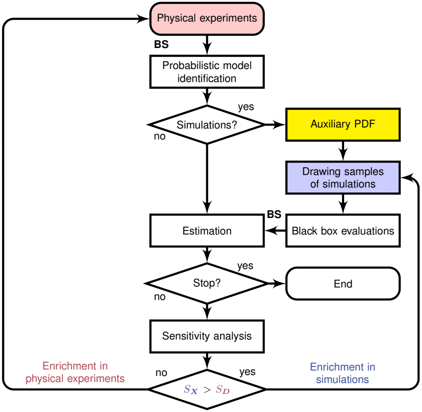
Resampling method

Bootstrap (BS) allows to generate N sample of simulations from the initial one.

Solution C

The cost of the whole approach is reduced to $N_{\mathbf{X}}$. It is **equivalent** to a MC simulation that does not consider the uncertainty of the sample of experiments.

Appendix B - MIS



Problem D

How to define an adequate auxiliary PDF as the probabilistic model evolves?

Multiple IS (MIS) [5]:

$$\hat{\mu}^{N-MIS} = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{N_{\boldsymbol{X}}}\sum_{j=1}^{N_{\boldsymbol{X}}}\phi\left(\boldsymbol{X}_k^{(j)}\right)\frac{\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}_k}(\boldsymbol{X}_k^{(j)})}{\frac{1}{N_{\boldsymbol{X}}}\sum_{l=1}^{N_{\boldsymbol{X}}}\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}_l}(\boldsymbol{X}_k^{(j)})}$$
$$= \frac{1}{N}\sum_{k=1}^{N}\hat{\mu}_k^{MIS},$$

with $\boldsymbol{X}_k^{(j)} \overset{i.i.d.}{\sim} \frac{1}{N_{\boldsymbol{X}}}\sum_{l=1}^{N_{\boldsymbol{X}}}\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}_l}$. The choosen auxiliary PDF updates itself and remains close to target PDFs.

Variance of estimation:

$$\mathbb{V}_{f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}}\left[\hat{\mu}^{N-MIS}\right] = \frac{1}{N}\mathbb{V}_{f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}}\left[\hat{\mu}^{MIS}\right].$$

[5] A Owen and Y Zhou (2000). Safe and effective importance sampling. Journal of the American Statistical Association.

Appendix C - Transformation for a MC procedure

Isoprobabilistic transformation

The transformation $\mathcal{T}_{\boldsymbol{D}}$ [11, 12] is performed here to work with an independent sample $\widetilde{\boldsymbol{U}} = \{\boldsymbol{U}^{(j)}, j = 1, ..., N_{\boldsymbol{X}}\}$:

$$\mathcal{T}_{\boldsymbol{D}} : \left| \begin{array}{ccc} [0,1]^d & \longrightarrow & \mathcal{X} \\ \boldsymbol{U} & \longmapsto & \boldsymbol{X} \end{array} \right. , \tag{4.1}$$

with $\boldsymbol{U}^{(j)} \overset{i.i.d.}{\sim} \mathcal{U}\,[0,1]^d$.

[11] AE Brockwell. Universal residuals: A multivariate transformation. Statistics probability letters, 2007.
[12] R Lebrun *et al*. Do rosenblatt and nataf isoprobabilistic transformations really differ? Probabilistic Engineering Mechanics, 2009.

Appendix D - ANOVA interpretation

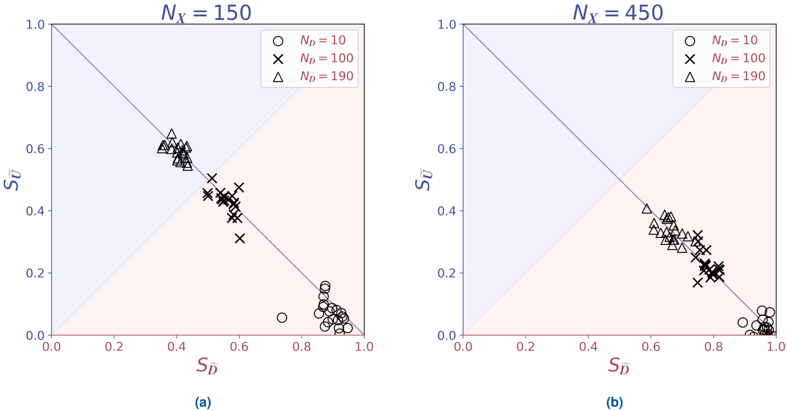


Figure: Evolution of Sobol' indices for the cantilever beam toy-case at $N_{\boldsymbol{D}} = [10, 100, 190]$ and (a) $N_{\boldsymbol{X}} = 150$ (b) $N_{\boldsymbol{X}} = 450$. Estimation of $n = 20$ indices for each combination.

Appendix E - Update of auxiliary PDF

An investment of $h_{\boldsymbol{X}}$ data performed in the input sample. When the reference database $\widetilde{\boldsymbol{D}}$ has been updated by $p_{\boldsymbol{D}}$ data prior to this investment, the new PDFs are hence estimated from more relevant databases and are added to the mixture so that

$$g = \frac{1}{N_{\boldsymbol{X}} + h_{\boldsymbol{X}}} \left(\sum_{i=1}^{N_{\boldsymbol{X}}} \hat{f}_i^{(N_{\boldsymbol{D}})} + \sum_{i=N_{\boldsymbol{X}}+1}^{N_{\boldsymbol{X}}+h_{\boldsymbol{X}}} \hat{f}_i^{(N_{\boldsymbol{D}}+h_{\boldsymbol{D}})} \right), \tag{4.2}$$

where $\hat{f}_i^{(n)}$ is the PDF estimated from the database $\widetilde{\boldsymbol{D}}_i$ of size n.

Appendix F - Global N-MIS estimator

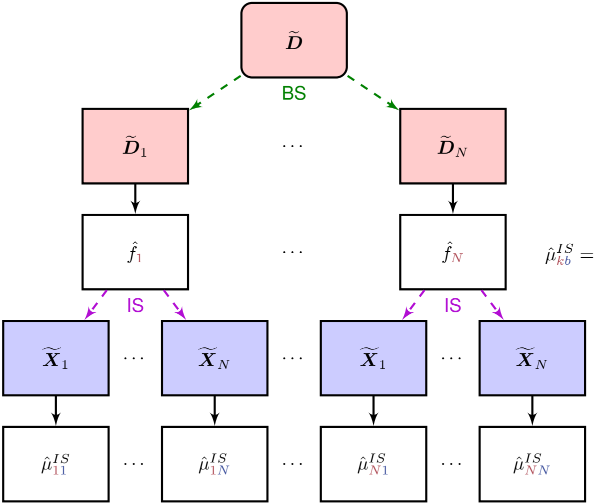
The global N-MIS estimator at each investment step is therefor given by

$$\hat{\mu}^{N-MIS}_{(M,P)} = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{N^{(P)}_{\boldsymbol{X}}} \sum_{j=1}^{N^{(P)}_{\boldsymbol{X}}} \phi\left(\boldsymbol{X}^{(j)}_{k}\right) \frac{\hat{f}^{((M))}_{k}(\boldsymbol{X}^{(j)}_{k})}{g^{((P))}(\boldsymbol{X}^{(j)}_{k})}, \quad (4.3)$$

$$\text{with} \begin{cases} g^{(0)} = \dfrac{1}{N_{\boldsymbol{X}}{}^{(0)}} \displaystyle\sum_{i=1}^{N_{\boldsymbol{X}}{}^{(0)}} \hat{f}^{(0)}_{i} \\ g^{(P)} = \dfrac{N^{(P-1)}_{\boldsymbol{X}}}{N^{(P)}_{\boldsymbol{X}}} g^{(P-1)} + \dfrac{1}{N^{(P)}_{\boldsymbol{X}}} \displaystyle\sum_{i=1+N^{(P-1)}_{\boldsymbol{X}}}^{N^{(P)}_{\boldsymbol{X}}} \hat{f}^{(M)}_{i} \end{cases},$$

where $\boldsymbol{X}^{(j)}_{k} \overset{i.i.d.}{\sim} g^{(P)}$ and $N_{\boldsymbol{D}}{}^{(M)}$ is the size at the M-th step of investment in the reference database whereas $N^{(P)}_{\boldsymbol{X}}$ is the size at the P-th step of investment in the reference input sample.

Appendix G - Direct sensitivity analysis



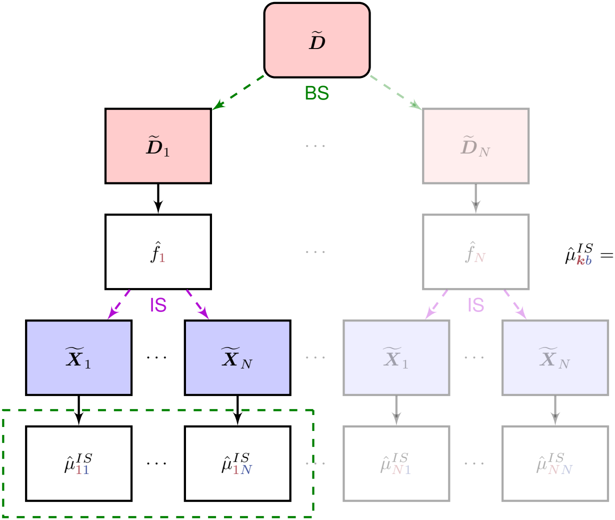
Sobol indices:

$$S_{\boldsymbol{D}} = \frac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{IS}|\widetilde{\boldsymbol{D}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{IS}\right]}$$

$$S_{\boldsymbol{X}} = \frac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{IS}|\widetilde{\boldsymbol{X}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{IS}\right]}$$

$$\hat{\mu}^{IS}_{kb} = \frac{1}{N_{\boldsymbol{X}}} \sum_{j=1}^{N_{\boldsymbol{X}}} \phi\left(\boldsymbol{X}_b^{(j)}\right) \frac{\hat{f}_k\left(\boldsymbol{X}_b^{(j)}\right)}{g\left(\boldsymbol{X}_b^{(j)}\right)}$$

Appendix G - Direct sensitivity analysis



Sobol indices:

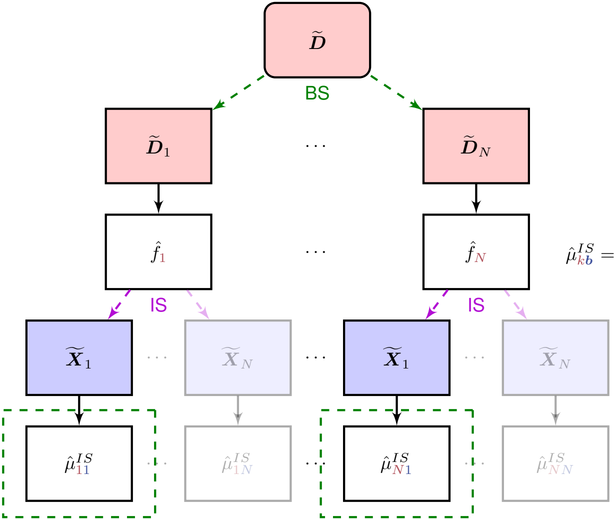
$$S_{\boldsymbol{D}} = \frac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{IS}|\widetilde{\boldsymbol{D}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{IS}\right]}$$

$$S_{\boldsymbol{X}} = \frac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{IS}|\widetilde{\boldsymbol{X}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{IS}\right]}$$

$$\hat{\mu}^{IS}_{\boldsymbol{k}b} = \frac{1}{N_{\boldsymbol{X}}}\sum_{j=1}^{N_{\boldsymbol{X}}}\phi\left(\boldsymbol{X}_b^{(j)}\right)\frac{\hat{f}_{\boldsymbol{k}}\left(\boldsymbol{X}_b^{(j)}\right)}{g\left(\boldsymbol{X}_b^{(j)}\right)}$$

$k = 1, ..., N$
estimations of
$\mathbb{E}\left[\hat{\mu}^{IS}|\widetilde{\boldsymbol{D}}\right]$

Appendix G - Direct sensitivity analysis



Sobol indices:

$$S_{\boldsymbol{D}} = \frac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{IS} \mid \widetilde{\boldsymbol{D}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{IS}\right]}$$

$$S_{\boldsymbol{X}} = \frac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{IS} \mid \widetilde{\boldsymbol{X}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{IS}\right]}$$

$$\hat{\mu}^{IS}_{k\boldsymbol{b}} = \frac{1}{N_{\boldsymbol{X}}} \sum_{j=1}^{N_{\boldsymbol{X}}} \phi\left(\boldsymbol{X}_{\boldsymbol{b}}^{(j)}\right) \frac{\hat{f}_k\left(\boldsymbol{X}_{\boldsymbol{b}}^{(j)}\right)}{g\left(\boldsymbol{X}_{\boldsymbol{b}}^{(j)}\right)}$$

$b = 1, ..., N$ estimations of $\mathbb{E}\left[\hat{\mu}^{IS} \mid \widetilde{\boldsymbol{X}}\right]$

Appendix H - Impact of the output behavior

Firespread toy case ($q = 10$) with non-parametric identification (Mean):

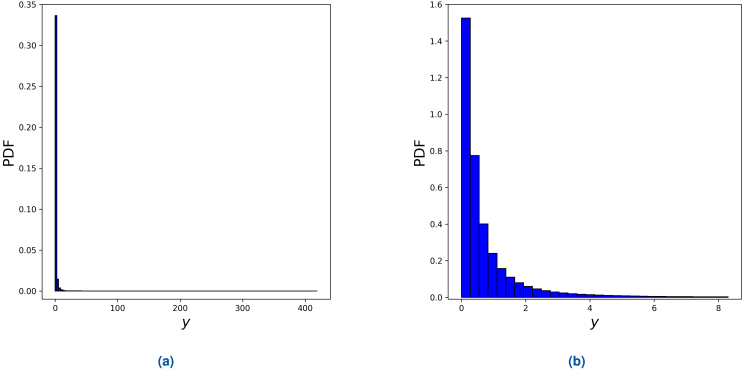
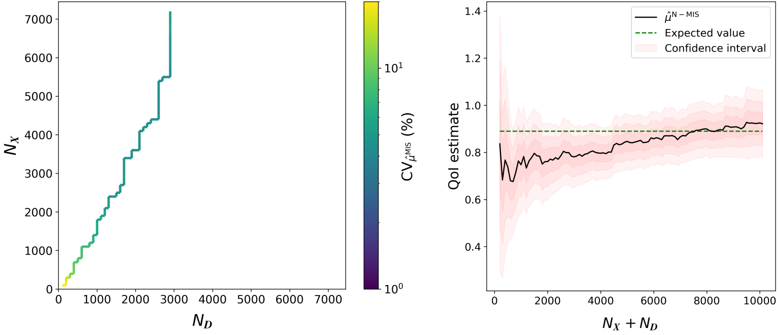


Figure: (a) Histogram of the output of Rothermel's modified model and (b) a focus on values lower than the 99%-quantile. The maximum value obtained as the output of the model is $R_{\text{max}} = 419\text{cm s}^{-1}$. The mean is estimated at $\mathbb{E}[\phi(X)] \approx 0.89\text{cm s}^{-1}$ with a standard deviation $\sigma_Y = 2.32\text{cm s}^{-1}$.

Appendix H - Impact of the output behavior

Firespread toy case ($q = 10$) with non-parametric identification (Mean):



(a) (b)

Figure: Approach applied to a test case of a firespread of dimension $q = 10$ with a sample of experiment of initial size $N_{\boldsymbol{D}} = 25$ and a sample of simulations of initial size $N_{\boldsymbol{X}} = 140$. The enrichment step is set to 100. (a) Trajectory of enrichment taken by the algorithm for the cantilever beam toy case. (b) Estimation of the QoI during data enrichment.