Sensitivity to statistical estimation uncertainties and probabilistic model identification

ETICS 2022

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

7th of October 2022

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

Table of contents

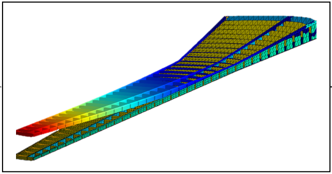
1. Context
2. Proposed approach
3. Sensitivity analysis
4. Illustration
5. Conclusion

Context - Uncertainty Quantification

Simulation code of a mechanical structure:

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

→



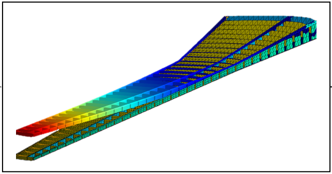
→

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

Context - Uncertainty Quantification

Simulation code of a mechanical structure:

d 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 :

$$\boldsymbol{X} \longrightarrow \boxed{\text{Black Box} \;\; \phi} \longrightarrow Y$$

with $\boldsymbol{X} = (X_1, ..., X_d)^t$ with values on the domain $\mathcal{X} \subseteq \mathbb{R}^d$ 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):

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

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):

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

Assuming $\tau = \mathrm{Id}$, the Monte Carlo (MC) estimator of this integral is given by:

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

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

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):

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

Assuming $\tau = \mathrm{Id}$, the Monte Carlo (MC) estimator of this integral is given by:

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

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

$$\boxed{\widetilde{\boldsymbol{X}} = \{\boldsymbol{X}^{(j)}, j = 1, \dots, N_{\boldsymbol{X}}\}} \xrightarrow{\Sigma} \boxed{\hat{\mu}^{MC}}$$

Context - 2^{nd} uncertainty source

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

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

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

with $N_{\boldsymbol{D}}$ the size of the database $\widetilde{\boldsymbol{D}}$. 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, 2021.

[2] K James, Lindsey and others. Parametric statistical inference. Oxford University Press, 1996.

[3] A J Izenman. Recent developments in nonparametric density estimation. Journal of the american statistical association, 1991.

Context - Small-Data

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

$$\downarrow$$

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

$$\downarrow$$

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

$$\Big\downarrow \Sigma$$

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

Context - Small-Data

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

$$\downarrow$$

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

$$\downarrow$$

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

$$\Sigma \downarrow$$

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

- The database is of limited size $N_{\boldsymbol{D}}$: the small-data context is imposed by costly experimental tests.

Context - Small-Data

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

$$\downarrow$$

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

$$\downarrow$$

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

$$\Big\downarrow \Sigma$$

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

- The database is of limited size $N_{\boldsymbol{D}}$: the small-data context is imposed by costly experimental tests.

- The MC sample is of limited size $N_{\boldsymbol{X}}$: the small-data context is imposed by the simulation time induced by the model complexity.

Context - Small-Data

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

$\downarrow$

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

$\downarrow$

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

$\Sigma \downarrow$

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

- The database is of limited size $N_{\boldsymbol{D}}$: the small-data context is imposed by costly experimental tests.
- The MC sample is of limited size $N_{\boldsymbol{X}}$: the small-data context is imposed by the simulation time induced by the model complexity.

$\downarrow$

Test-Simulation trade-off

Context - Conditioning on a given database

The expectation (1.1) and its estimator are thus written as following for a given database $\widetilde{\boldsymbol{D}} = \widetilde{\boldsymbol{d}}$:

$$\mathbb{E}_{\hat{f}_{\boldsymbol{X} \mid \widetilde{\boldsymbol{D}}=\widetilde{\boldsymbol{d}}}}\left[\phi\left(\boldsymbol{X}\right)\right] = \int_{\mathcal{X}} \phi\left(\boldsymbol{x}\right) \hat{f}_{\boldsymbol{X} \mid \widetilde{\boldsymbol{D}}=\widetilde{\boldsymbol{d}}}\left(\boldsymbol{x}\right) \mathrm{d}\boldsymbol{x} \qquad (1.3)$$

$$\approx \frac{1}{N_{\boldsymbol{X}}} \sum_{j=1}^{N_{\boldsymbol{X}}} \phi\left(\boldsymbol{X}^{(j)}\right), \qquad (1.4)$$

with $\boldsymbol{X}^{(j)} \overset{i.i.d.}{\sim} \hat{f}_{\boldsymbol{X} \mid \widetilde{\boldsymbol{D}}=\widetilde{\boldsymbol{d}}}$. The estimator (1.4) is now subject to the first uncertainty source conditioned on the database $\widetilde{\boldsymbol{D}} = \widetilde{\boldsymbol{d}}$.

Context - Conditioning on a given database

The expectation (1.1) and its estimator are thus written as following for a given database $\widetilde{\boldsymbol{D}} = \widetilde{\boldsymbol{d}}$:

$$\mathbb{E}_{\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}=\widetilde{\boldsymbol{d}}}}\left[\phi\left(\boldsymbol{X}\right)\right] = \int_{\mathcal{X}} \phi\left(\boldsymbol{x}\right) \hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}=\widetilde{\boldsymbol{d}}}\left(\boldsymbol{x}\right) \mathrm{d}\boldsymbol{x} \tag{1.3}$$

$$\approx \frac{1}{N_{\boldsymbol{X}}} \sum_{j=1}^{N_{\boldsymbol{X}}} \phi\left(\boldsymbol{X}^{(j)}\right), \tag{1.4}$$

with $\boldsymbol{X}^{(j)} \overset{i.i.d.}{\sim} \hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}=\widetilde{\boldsymbol{d}}}$. The estimator (1.4) is now subject to the first uncertainty source conditioned on the database $\widetilde{\boldsymbol{D}} = \widetilde{\boldsymbol{d}}$.

However, the uncertainty related to the database is not considered in the following variance :

$$\mathbb{V}_{\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}=\widetilde{\boldsymbol{d}}}}\left[\hat{\mu}^{MC}\right] = \frac{1}{N_{\boldsymbol{X}}} \mathbb{V}_{\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}=\widetilde{\boldsymbol{d}}}}\left[\phi\left(\boldsymbol{X}\right)\right]. \tag{1.5}$$

Context - Problems

$$\widetilde{\boldsymbol{D}} = \{\boldsymbol{D}^{(i)}, i = 1, \dots, N_{\boldsymbol{D}}\}$$

with $\boldsymbol{D}^{(i)} \overset{i.i.d.}{\sim} f_{\boldsymbol{X}}$

$$\downarrow$$

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

$$\downarrow$$

$$\widetilde{\boldsymbol{X}} = \{\boldsymbol{X}^{(j)}, j = 1, \dots, N_{\boldsymbol{X}}\}$$

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

$$\Sigma \downarrow$$

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

Context - Problems

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

$$\downarrow$$

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

$$\downarrow$$

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

$$\xrightarrow{\Sigma}$$

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

Problem A

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

Context - Problems

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

$$\downarrow$$

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

$$\downarrow$$

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

$$\Big\downarrow \Sigma$$

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

Problem A

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

Problem B

In order to improve efficiently the accuracy of the estimator, should the investment of data be made in the database or the MC sample?

Table of contents

1. Context
2. Proposed approach
3. Sensitivity analysis
4. Illustration
5. Conclusion

Proposed approach - Double integral expectation

Problem A

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

Proposed approach - Double integral expectation

Problem A

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

The following expectation takes into account the variation of the database:

$$\mathbb{E}_{f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}}\left[\phi\left(\boldsymbol{X}\right)\right] = \int_{\mathcal{X}^{N_D}} \int_{\mathcal{X}} \phi\left(\boldsymbol{x}\right) f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}(\boldsymbol{x},\widetilde{\boldsymbol{d}}) \mathrm{d}\boldsymbol{x} \; \mathrm{d}\widetilde{\boldsymbol{d}}. \tag{2.1}$$

Proposed approach - Double integral expectation

Problem A

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

The following expectation takes into account the variation of the database:

$$\mathbb{E}_{f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}}\left[\phi\left(\boldsymbol{X}\right)\right] = \int_{\mathcal{X}^{N_{\boldsymbol{D}}}} \int_{\mathcal{X}} \phi\left(\boldsymbol{x}\right) f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}(\boldsymbol{x},\widetilde{\boldsymbol{d}})\mathrm{d}\boldsymbol{x}\ \mathrm{d}\widetilde{\boldsymbol{d}}. \tag{2.1}$$

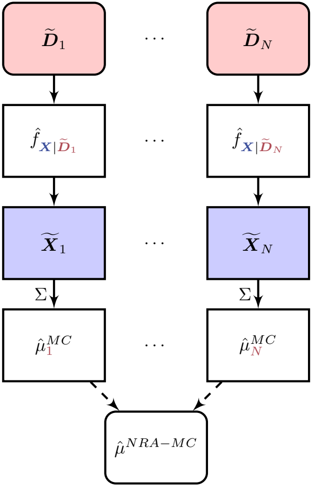
An estimator of this integral [4] is the following:

$$\hat{\mu}^{NRA-MC} = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{N_{\boldsymbol{X}}}\sum_{j=1}^{N_{\boldsymbol{X}}}\phi\left(\boldsymbol{X}_k^{(j)}\right) = \frac{1}{N}\sum_{k=1}^{N}\hat{\mu}_k^{MC}, \tag{2.2}$$

with $\boldsymbol{X}_k^{(j)} \overset{i.i.d.}{\sim} \hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}_k}$ and N the number of databases of size $N_{\boldsymbol{D}}$.

[4] V Chabridon. Analyse de sensibilité fiabiliste avec prise en compte d'incertitudes sur le modèle probabiliste, 2018.

Proposed approach - Empirical variance



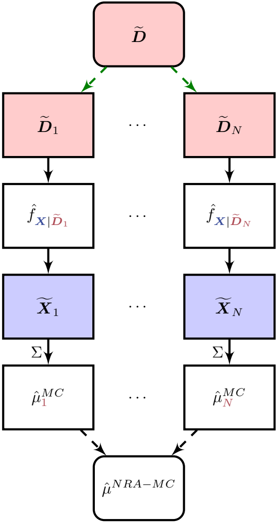
The process is repeated in order to estimate the variance with respect to the database.

Empirical variance:

$$\hat{\mathbb{V}}_{f_{(X,\widetilde{D})}}\left[\hat{\mu}^{MC}\right] = \frac{1}{N-1}\sum_{k=1}^{N}\left(\hat{\mu}_k^{MC} - \hat{\mu}^{NRA-MC}\right)^2 \quad (2.3)$$

(2.2)

Proposed approach - Small-data context



(2.3)

(2.2)

In a small-data context, only a database $\widetilde{\boldsymbol{D}}$ of limited size $N_{\boldsymbol{D}}$ is available.

Resampling method

Allows to generate N databases from the initial one. [5, 6] (e.g. Bootstrap)

Solution A

The nested estimator (2.2) is now conditioned on the initial database but the method allows to take into account the uncertainty related to it.

[5] C H Yu. Resampling methods: concepts, applications, and justification. Practical Assessment, Research, and Evaluation, 2002.

[6] B Efron. The jackknife, the bootstrap and other resampling plans. SIAM, 1982.

Table of contents

1 Context

2 Proposed approach

3 Sensitivity analysis

4 Illustration

5 Conclusion

Sensitivity analysis - ANOVA

Problem B

In order to improve efficiently the accuracy of the estimator, should the investment of data be made in the database or the MC sample?

Sensitivity analysis - ANOVA

Problem B

In order to improve efficiently the accuracy of the estimator, should the investment of data be made in the database or the MC sample?

An ANalysis Of VAriance [7, 8] is performed:

$$\begin{cases} S_{\widetilde{\boldsymbol{D}}} = \dfrac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{MC} | \widetilde{\boldsymbol{D}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{MC}\right]} \\ S_{\widetilde{\boldsymbol{X}}} = \dfrac{\mathbb{V}\left[\mathbb{E}\left[\hat{\mu}^{MC} | \widetilde{\boldsymbol{X}}\right]\right]}{\mathbb{V}\left[\hat{\mu}^{MC}\right]} \end{cases} . \tag{3.1}$$

Interpretation of Sobol' indices:

- $S_{\widetilde{\boldsymbol{D}}} \to$ proportion of variance due to the database,
- $S_{\widetilde{\boldsymbol{X}}} \to$ proportion of variance due to the MC sample.

[7] Ilya M Sobol'. Sensitivity analysis for non-linear mathematical models. Mathematical modelling and computational experiment, 1993.
[8] F Gamboa and others. Statistical inference for sobol pick-freeze monte carlo method. Statistics, 50(4):881–902, 2016.

Sensitivity analysis - Interpretation

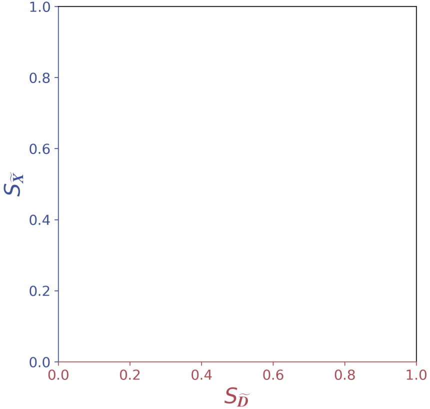


Figure 3.1: Interpretation of Sobol' indices associated to the database and the Monte Carlo sample.

Sensitivity analysis - Interpretation

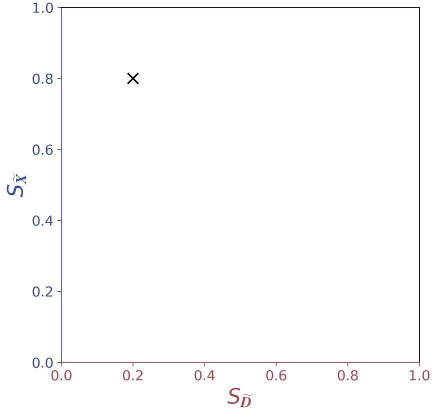


Figure 3.1: Interpretation of Sobol' indices associated to the database and the Monte Carlo sample.

Sensitivity analysis - Interpretation

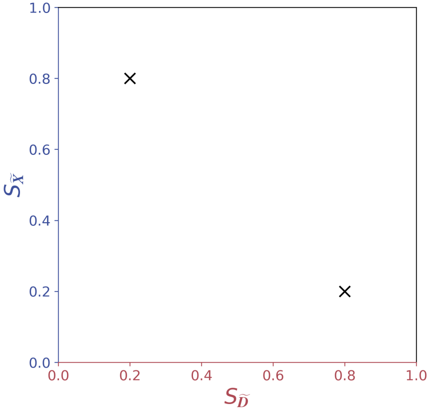


Figure 3.1: Interpretation of Sobol' indices associated to the database and the Monte Carlo sample.

Sensitivity analysis - Interpretation

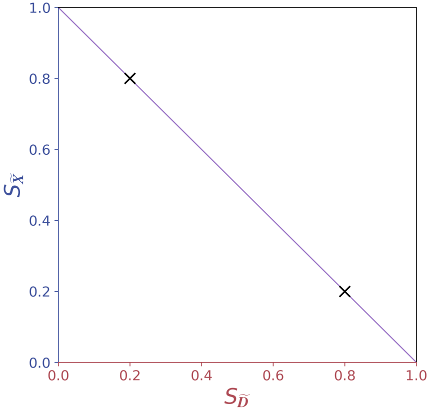


Figure 3.1: Interpretation of Sobol' indices associated to the database and the Monte Carlo sample.

Sensitivity analysis - Interpretation

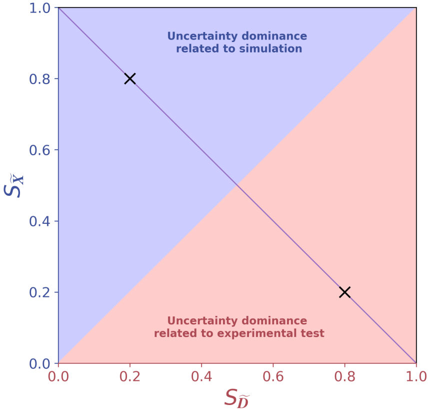


Figure 3.1: Interpretation of Sobol' indices associated to the database and the Monte Carlo sample.

Sensitivity analysis - Interpretation

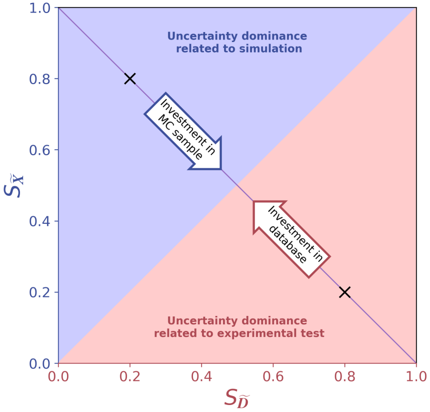


Figure 3.1: Interpretation of Sobol' indices associated to the database and the Monte Carlo sample.

Sensitivity analysis - Independance of inputs

However, the dependance of $\widetilde{X}$ to $\widetilde{D}$ is an issue for the sensitivity analysis.

Sensitivity analysis - Independance of inputs

However, the dependance of $\widetilde{\boldsymbol{X}}$ to $\widetilde{\boldsymbol{D}}$ is an issue for the sensitivity analysis.

Isoprobabilistic transformation

The transformation $\mathcal{T}_{\boldsymbol{D}}$ [9, 10, 11] is performed here to work with an independent sample $\widetilde{\boldsymbol{U}} = \{\boldsymbol{U}^{(j)}, j = 1, ..., N_{\boldsymbol{X}}\}$:

$$\mathcal{T}_{\boldsymbol{D}} : \left| \begin{array}{ccc} [0,1]^d & \longrightarrow & \mathcal{X} \\ \boldsymbol{U} & \longmapsto & \boldsymbol{X} \end{array} \right. , \tag{3.2}$$

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

[9] M Rosenblatt. Remarks on a multivariate transformation. The annals of mathematical statistics, 1952.

[10] AE Brockwell. Universal residuals: A multivariate transformation. Statistics probability letters, 2007.

[11] R Lebrun and others. Do rosenblatt and nataf isoprobabilistic transformations really differ? Probabilistic Engineering Mechanics, 2009.

Sensitivity analysis - Solution

Solution B

The investment of data is guided by the highest index:

- $S_{\widetilde{\boldsymbol{D}}} > S_{\widetilde{\boldsymbol{U}}}$ $\longrightarrow$ Investment in the database (experimental tests),
- $S_{\widetilde{\boldsymbol{D}}} < S_{\widetilde{\boldsymbol{U}}}$ $\longrightarrow$ Investment in the MC sample (simulation).

Table of contents

1 Context

2 Proposed approach

3 Sensitivity analysis

4 Illustration

5 Conclusion

Illustration - Cantilever Beam

Mean deflection of the free end of a cantilever beam:

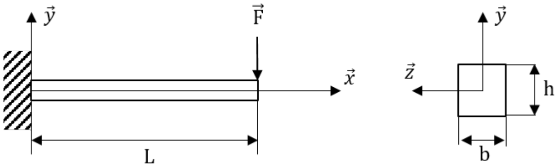


Figure 4.1: Representation of a cantilever beam where F is the transverse load applied on the free end of the beam of length L, Young's modulus E and cross-section bh.

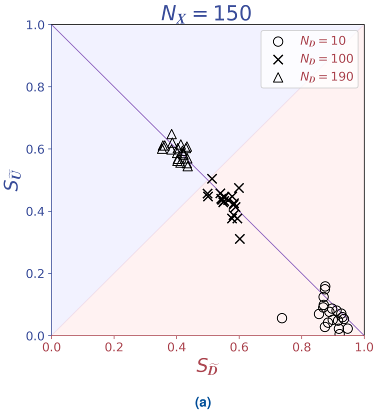
$$\phi\left(F, L, E, b, h\right) = \frac{4FL^3}{Ebh^3}$$

Table 4.1: Probabilistic models associated to independent input variables for a cantilever beam toy-case. [12]

Input variable	Distribution	Mean	Coefficient of variation
F	LogNormal	556.8 [N]	0.08
L	Normal	4290 [mm]	0.1
E	LogNormal	2.10^5 [MPa]	0.06
b	Normal	62 [mm]	0.1
h	Normal	98.7 [mm]	0.1

[12] L Baoyu and others. Reliability analysis based on a novel density estimation method for structures with correlations, 2017.

Illustration - Results



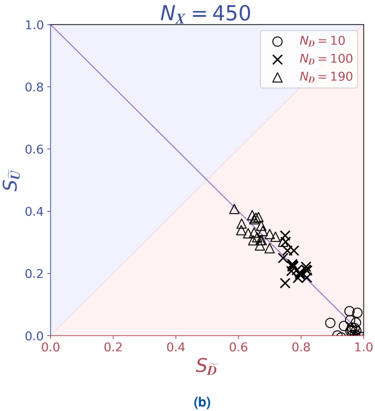


Figure 4.2: Evolution of Sobol' indices for the cantilever beam toy-case at $N_{\boldsymbol{D}} = [10, 100, 190]$ and (a) $N_{\boldsymbol{X}} = 150$ (b) $N_{\boldsymbol{X}} = 450$. Estimation of $n = 20$ indices for each combination.

Table of contents

1 Context

2 Proposed approach

3 Sensitivity analysis

4 Illustration

5 Conclusion

Conclusion

Framework

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

Conclusion

Framework

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

Current method

A) Takes into account the database uncertainty in the variance of the estimator,

B) Answers the test-simulation trade-off by guiding the investment of data in the driving source of uncertainty.

Conclusion

Framework

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

Current method

A) Takes into account the database uncertainty in the variance of the estimator,

B) Answers the test-simulation trade-off by guiding the investment of data in the driving source of uncertainty.

Perspectives

- Reduction of the computational burden with importance sampling methods, [13]
- Quantification of the amount of data to invest while considering cost differences.

[13] A Owen, Y Zhou. Safe and effective importance sampling. Journal of the American Statistical Association, 2000.

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

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

References III

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

[8] 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.

[9] Murray Rosenblatt.
Remarks on a multivariate transformation.
The annals of mathematical statistics, 23(3):470–472, 1952.

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

References IV

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

[12] 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.

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