Global Sensitivity Analysis Reporting Tool for Easily Detecting Variable Impact and Interaction

ENBIS Spring Meeting – Dortmund, May 16, 2024

Elena Raponi (e.raponi@liacs.leidenuniv.nl), Leiden Institute of Advanced Computer Science, The Netherlands

Universiteit Leiden
The Netherlands

liacs Leiden Institute of Advanced Computer Science

Naco Natural Computing Computer Science Leiden University

enbis European Network for Business and Industrial Statistics

Index

1. Today's topic
2. Inspiration
   - Original Problem
   - Global Sensitivity Analysis
   - Morris' Method
   - Some Results
3. GSAreport
4. Application Case from Computational Mechanics
5. Final Remarks

Today's Topic:
Global Sensitivity Analysis Reporting Tool

GSA report

https://github.com/nikivanstein/GSAreport

- Application containing **many GSA methods**
- **Dimension and sample-size dependent**, it selects the appropriate methods to perform global sensitivity analysis.
- Works with **pre-generated datasets** or **generates the sample** itself based on problem specifications
- Developed within the **Natural Computing** Cluster - LIACS

Niki van Stein

Elena Raponi

Thomas Bäck

Naco Natural Computing Computer Science Leiden University

Inspiration

Original Problem

Single Nucleotide Polymorphism (SNP) Genotyping

- Plant breeding → Process to develop new plant varieties trough mutation and recombination to improve the genetic potential of plants.
- SNP contributions can be strong/moderate, independent of each other, or influence and modulate the effect of each other.
- We want to evaluate how SNP variations impact a target

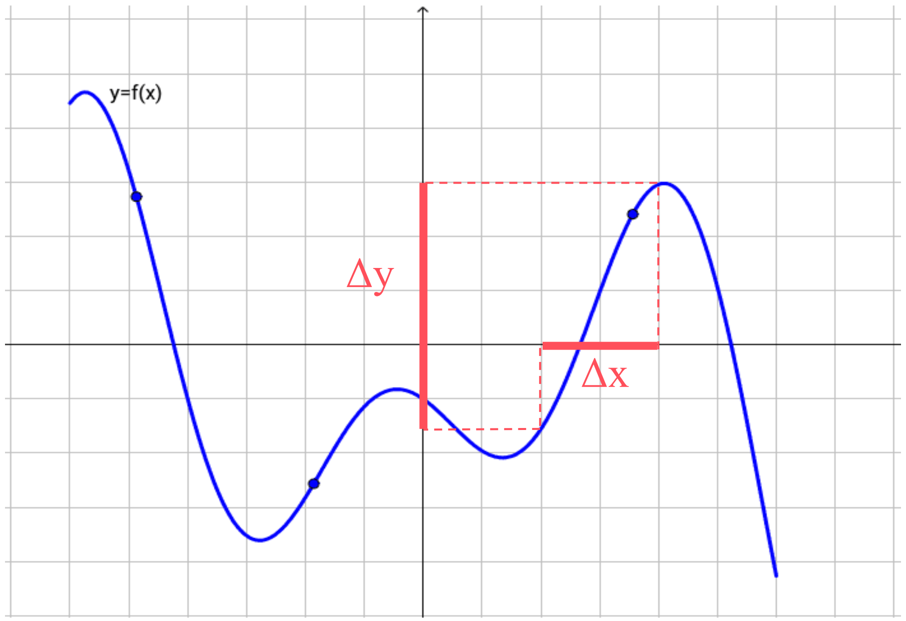
Sample 1	Sample 2	Sample 3	Sample 4
CCCTAAACCCTAA	CCCTAAACCCTAA	CCCTAAACCCTAA	CCCTAAACCCTAA
AACCTAAACCCTA	ACCCTAAACCCTA	AACCTAAACCCTA	AACCTAAACCCTG
AACCTCTGAATCC	AACCTGTGAATCC	AACCTGTGAATCC	AACCTCTGAATCC
TTAATCCCTAAAT	TTAATCCCTAAAT	TTAATCCCTAAAT	TTAATCCCTAAAT
CCCTAAATCTTTA	CCCTAAATCTTTA	CCCTAAATCTTTA	CCCTAAATCTTTA
ACTCTTACATCCA	ACTCCTACATCCA	AATCCTACATCCA	AATCCTACATCCA
TGAATCCCTAAAT	TGAATCCCTAAAT	TGAATCCCTAAAT	TGAATCCCTAAAT
ACGC	ACGC	ACGC	ACGC

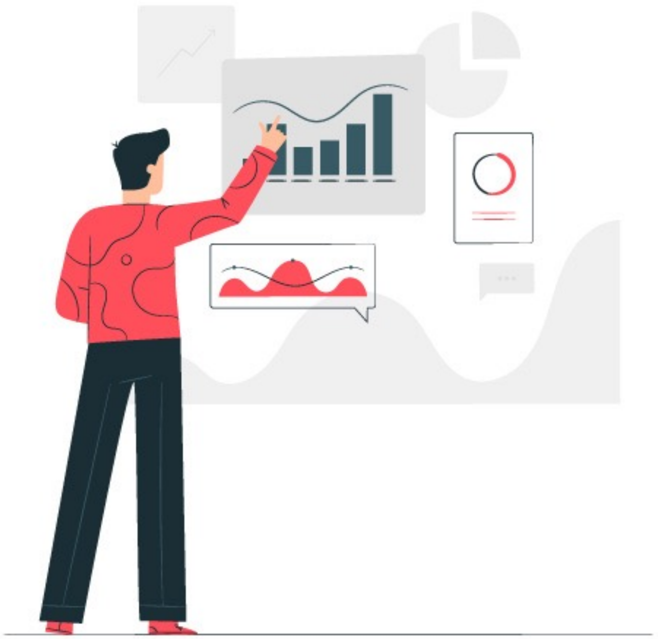
→ Trait value

Inspiration

Global Sensitivity Analysis

- Global sensitivity analysis (GSA) quantifies the importance of model inputs and their interactions with respect to model output.
- It measures the **uncertainty in output** based on the **changes in the input**.
- Global ≠ Local





Inspiration

Morris Method

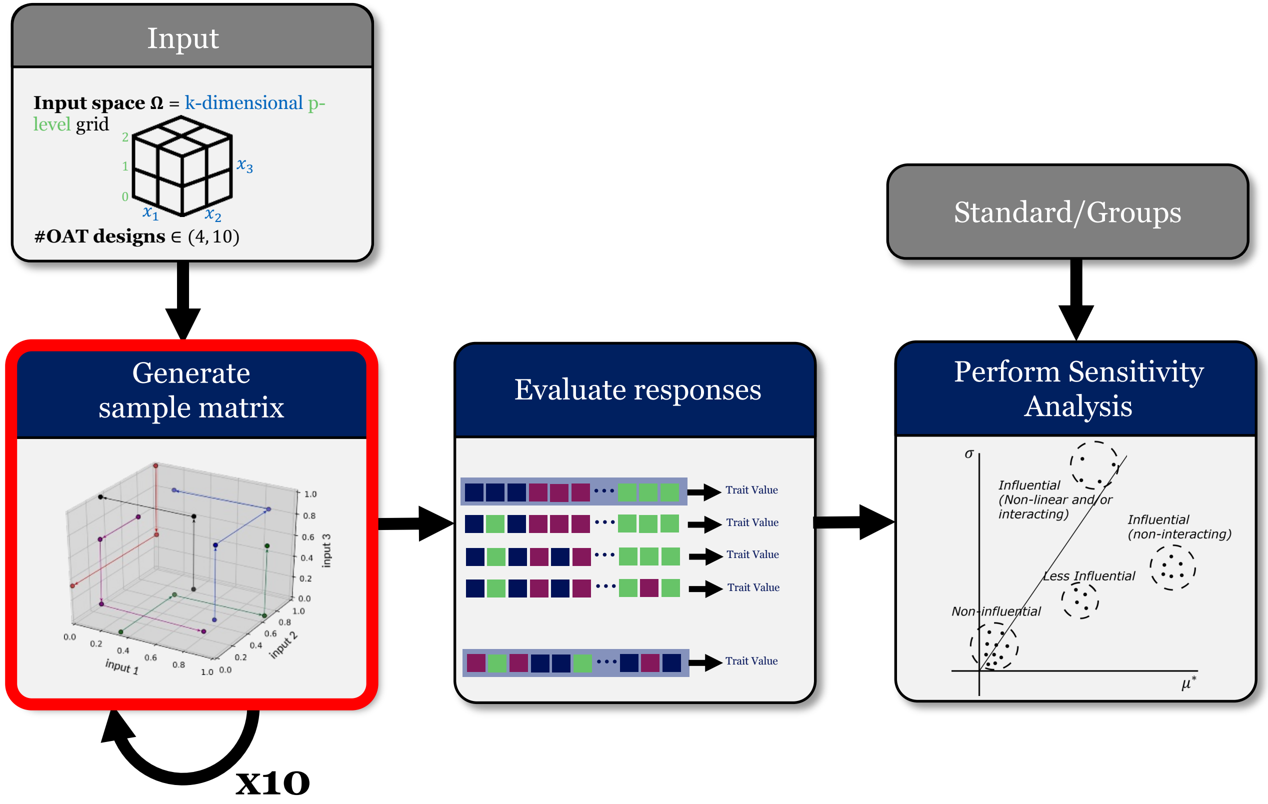
- Given a random sampling of a vector $X = (x_1, x_2, \ldots, x_k)$ from the grid Ω, a k-dimensional, p-level grid, we generate the the distribution of elementary effects F_i for every i^{th} input.

- The elementary effect is defined as

$$EE_i(X) = \frac{y(x_1, x_2, \ldots, x_i \pm \Delta, \ldots, x_k) - y(x_1, x_2, \ldots, x_k)}{\Delta}$$

where Δ is the *grid jump* on the trajectory.

- We consider $\Delta = 1$ and repeat this for every feature of the input and for $r = 10$ randomly sampled inputs.

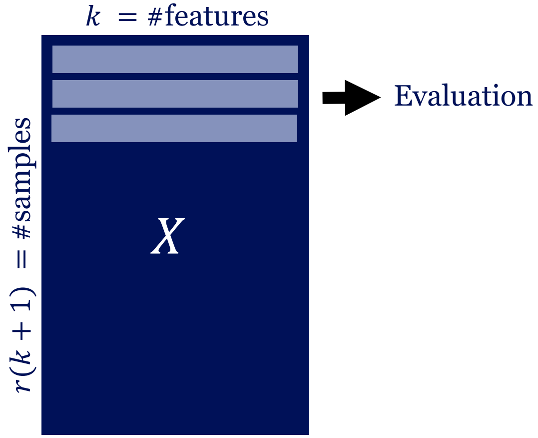


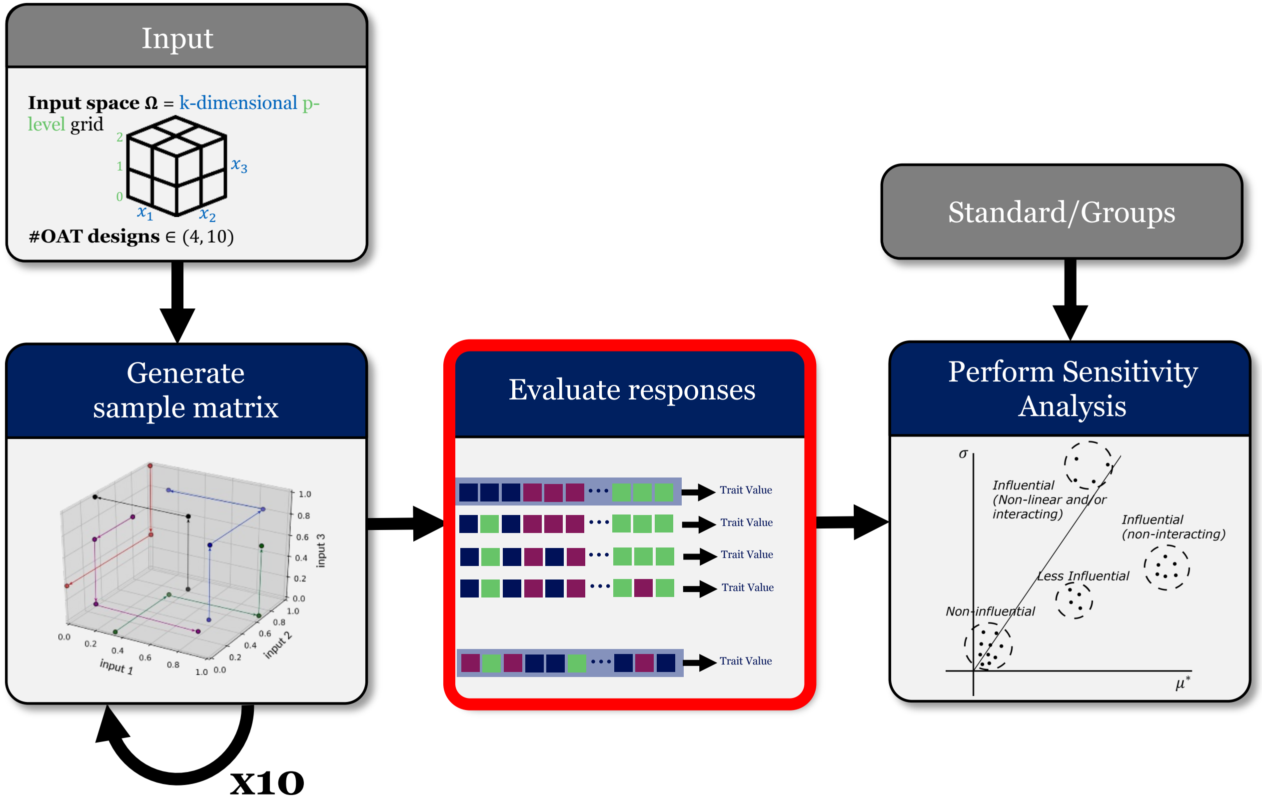
Inspiration

Morris Method

- We obtain a sample matrix X of dimensions $r(k+1)\times k$, where:

  - r = number of trajectories
  - $(k+1)$ = number of points for each trajectory





Inspiration

Morris Method

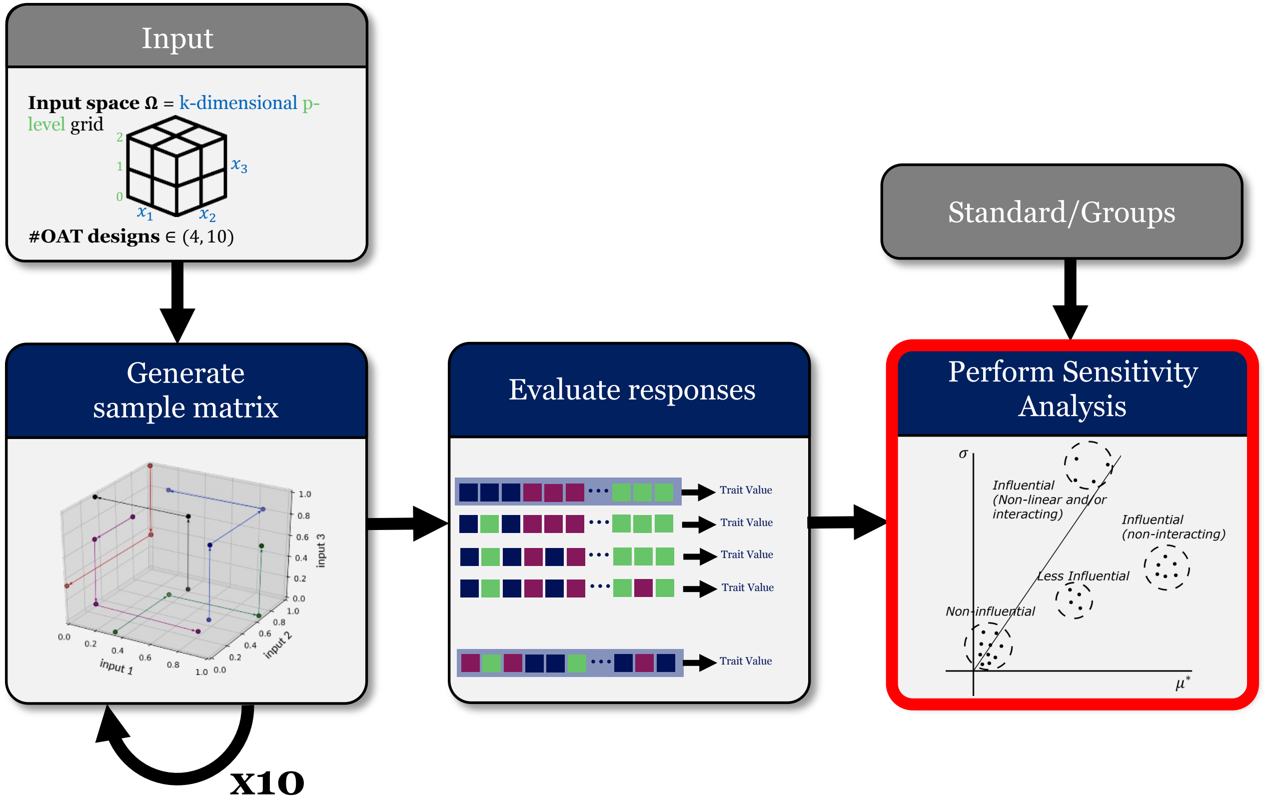
- We compute the sensitivity measures for each factor over the 10 trajectories:
  - **absolute mean $\boldsymbol{\mu_i^*}$** of the distribution F_i
  - **standard deviation $\boldsymbol{\sigma_i}$** of the distribution F_i

$$\mu_i^* = \frac{1}{r}\sum_{j=1}^{r}\left|EE_i^j\right|$$

→ robust to Type II errors

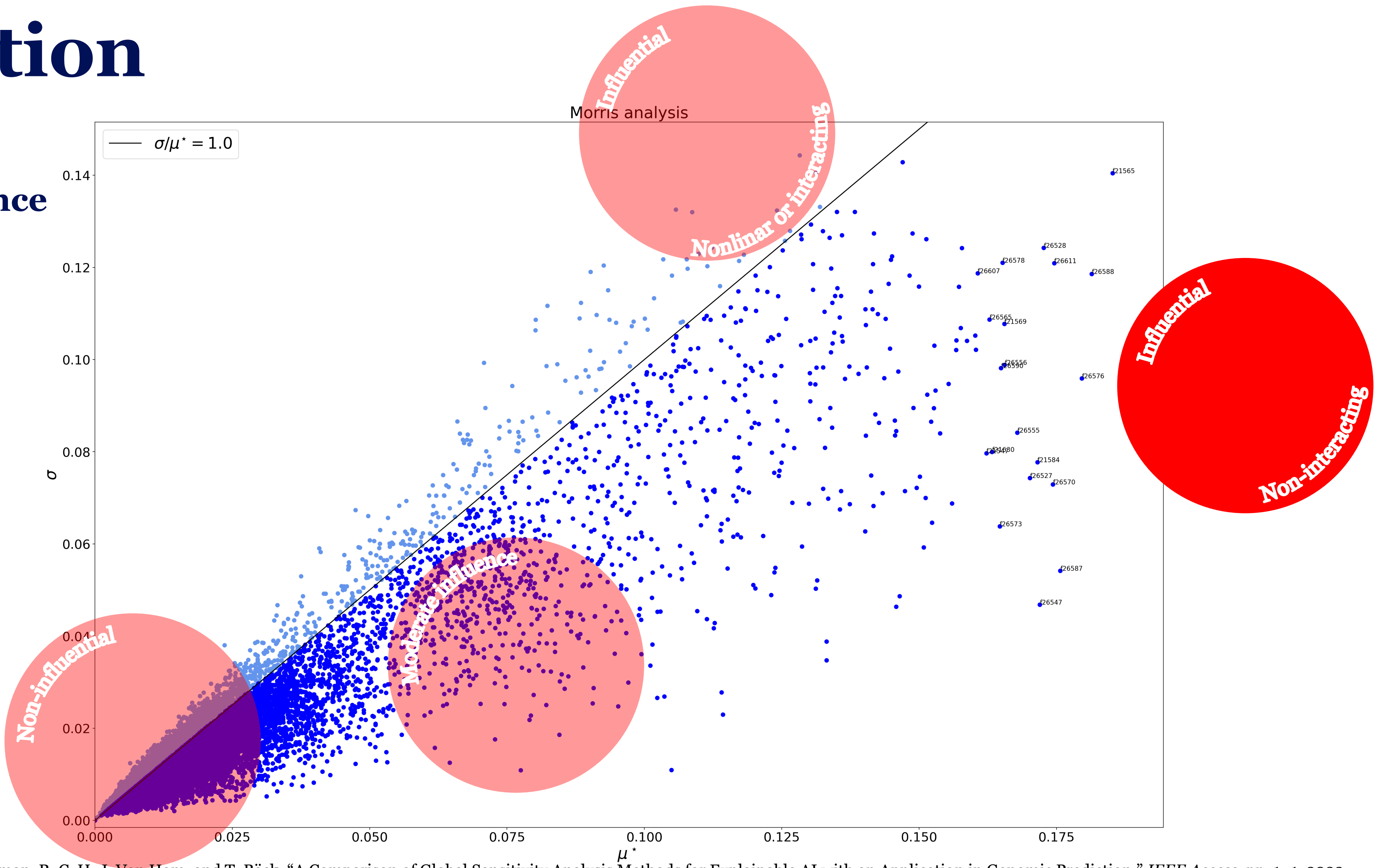
$$\sigma_i = \sqrt{\frac{1}{r}\sum_{j=1}^{r}\left(EE_i^j - \mu_i\right)^2}$$

where $i = 1, \dots, k$.



Inspiration

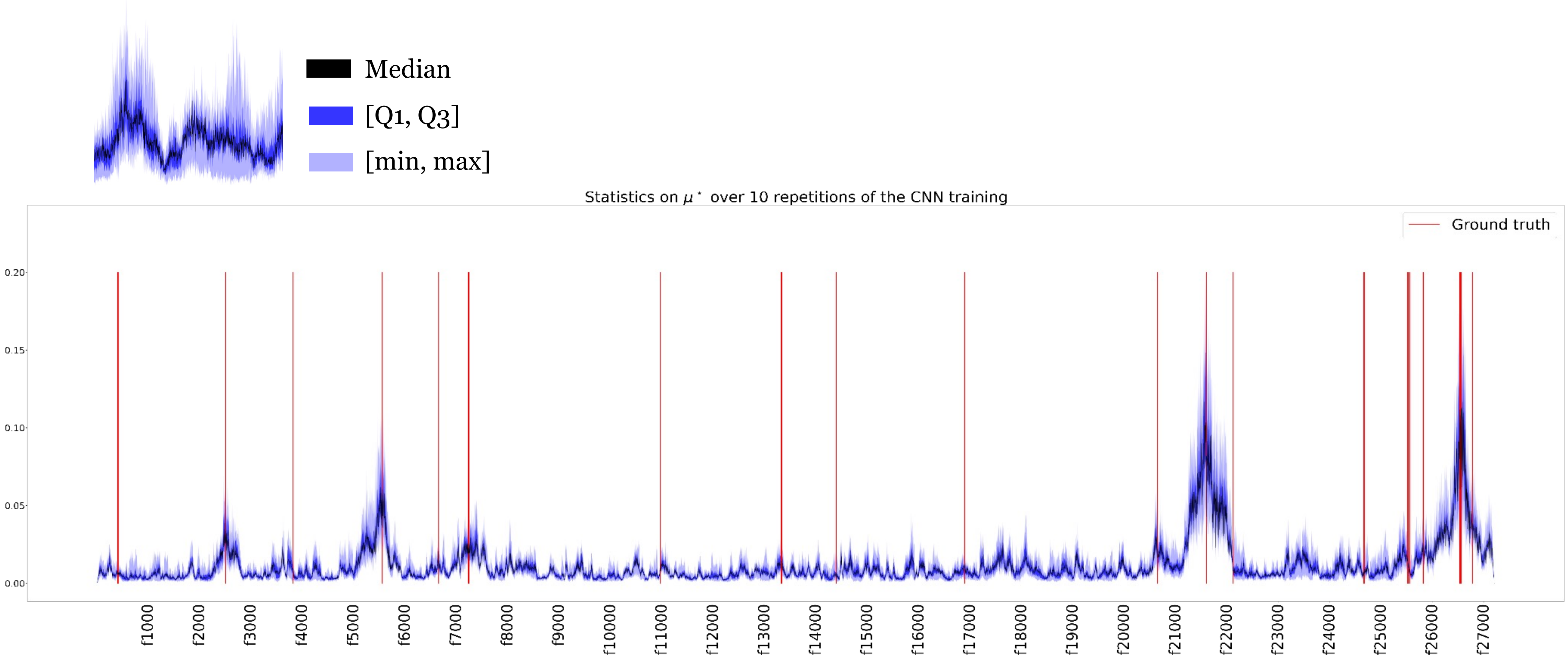
Some Results: Morris Covariance Plot



B. Van Stein, E. Raponi, Z. Sadeghi, N. Bouman, R. C. H. J. Van Ham, and T. Bäck, “A Comparison of Global Sensitivity Analysis Methods for Explainable AI with an Application in Genomic Prediction,” *IEEE Access*, pp. 1–1, 2022

Inspiration

Some Results: Lineplot for $\mu*$



B. Van Stein, E. Raponi, Z. Sadeghi, N. Bouman, R. C. H. J. Van Ham, and T. Bäck, "A Comparison of Global Sensitivity Analysis Methods for Explainable AI with an Application in Genomic Prediction," *IEEE Access*, pp. 1–1, 2022

GSA Report

GitHub

What do you need

GSA Report is easy to use. The report can be generated in a few steps:

- Generate a set of inputs

```
python GSAreport.py -p problem.json -d `pwd`/data --sample --samplesize 1000 -o pwd/output
```

- Evaluate outputs

```
python generate_outputs.py
```

- Extract Sensitivity Information

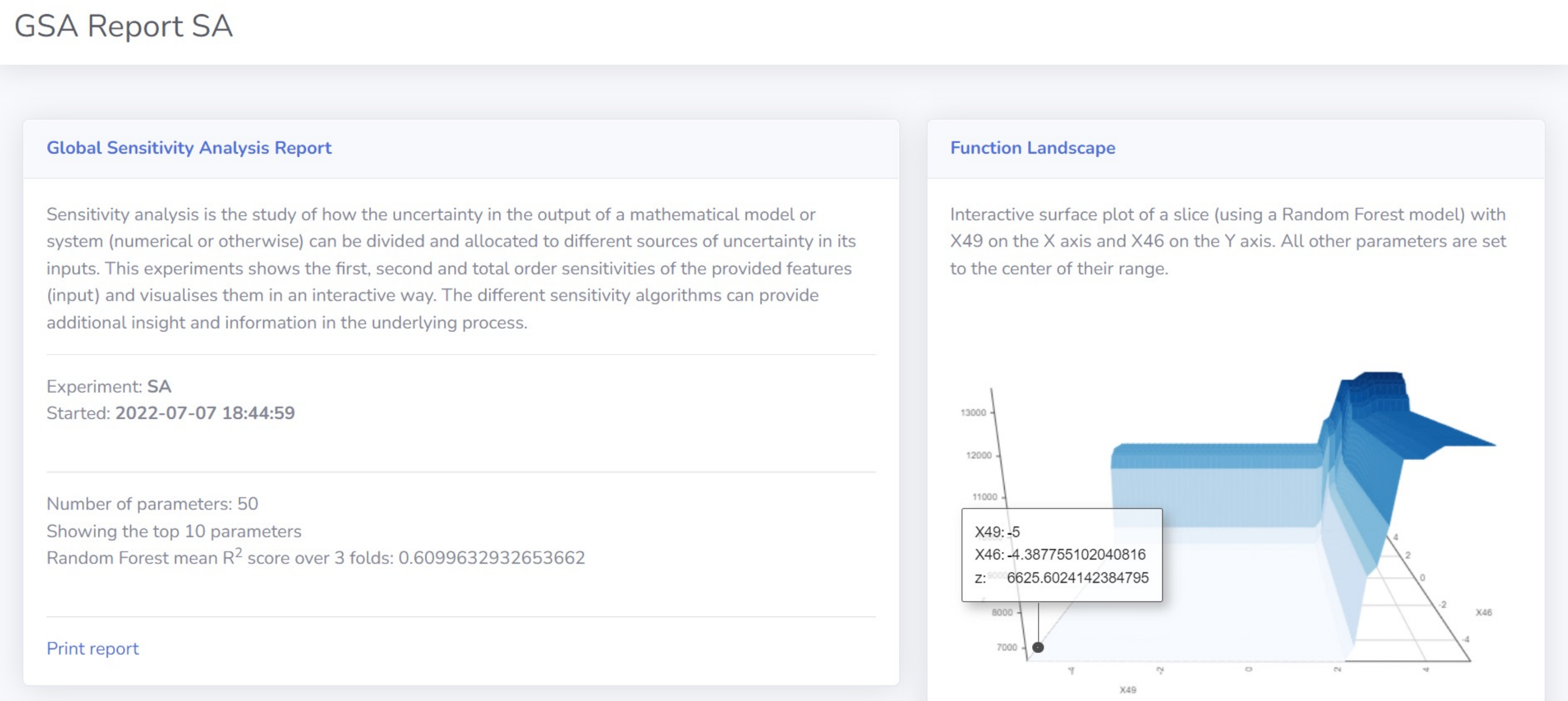
```
python GSAreport.py -p problem.json -d data_dir -o output_dir
```

B. Van Stein, E. Raponi, Z. Sadeghi, N. Bouman, R. C. H. J. Van Ham, and T. Bäck, “A Comparison of Global Sensitivity Analysis Methods for Explainable AI with an Application in Genomic Prediction,” *IEEE Access*, pp. 1–1, 2022

GSAreport

What do you get

In a few seconds, you get an interactive report containing charts of the most used GSA methods that are suitable for your specific application.



Variance-based methods
- Sobol
- FAST
- RBD-FAST

Derivative-based methods
- Morris
- DGSM

Density-based methods
- DELTA
- PAWN

Model-based methods
- Linear models
- Random Forest
- Shapley and SHAP

B. Van Stein, E. Raponi, Z. Sadeghi, N. Bouman, R. C. H. J. Van Ham, and T. Bäck, “A Comparison of Global Sensitivity Analysis Methods for Explainable AI with an Application in Genomic Prediction,” *IEEE Access*, pp. 1–1, 2022

GSAreport

What do you get

TABLE 2. Comparison of different GSA and XAI methods based on different aspects and support. A + indicates that the feature is present for a method. For the performance in accuracy and time a distinction is made between the best methods (++), good methods (+), worse methods (–) and the worst method for a given situation (- -). *The TreeSHAP method is model dependent, however classical Shapley values are model independent. ** The linear model has an advantage due to the experimental setup and is therefore not included in this overview regarding accuracy performance. * The TreeSHAP method was not included.**

	Variance-based			Derivative-based		Density-based		Model-based		
	Sobol	Fast	RBD-FAST	Morris	DGSM	DELTA	PAWN	Linear	RF	TreeSHAP
First order sens.	+	+	+	+	+	+		+		
Second order sens.	+			+						
Total order sens.	+	+		+	+	+	+		+	+
Direction of effect								+		+
Confidence indication	+	+	+	+	+	+	+			
Grouping support	+			+	+					
Model independence	+	+	+	+	+	+	+			*
Sampling scheme independence			+			+	+	+	+	+
No min. sample size required							+	+	+	+
Multidimensional averaging	+	+	+	+	+	+	+	+	+	+
Performance in low dim.	+	-	+	++	- -	+	+	**	+	***
Performance in high dim.	-	- -	+	++	-	+	-	**	+	***
Performance with small sample sizes.	-	- -	+	++	-	-	-	**	-	***
Computational efficiency.	+	-	+	+	+	- -	-	++	-	***

B. Van Stein, E. Raponi, Z. Sadeghi, N. Bouman, R. C. H. J. Van Ham, and T. Bäck, "A Comparison of Global Sensitivity Analysis Methods for Explainable AI with an Application in Genomic Prediction," *IEEE Access*, pp. 1–1, 2022

Application Case from Computational Mechanics

Drawability Assessment of Deep-Drawn Components

- In the early-stage development of sheet metal parts, key design properties of new structures must be specified.
- This affects drawability.
- GSA can provide insight into the impact of various changes in drawing configurations on drawability.
- We test GSAreport on this industrial test case.

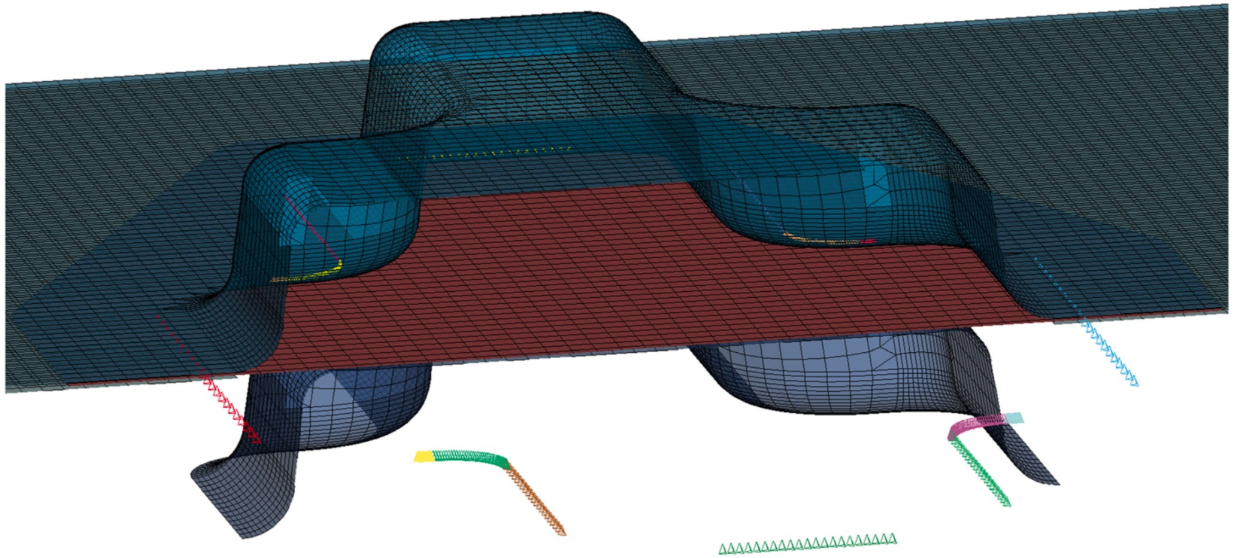


Figure: Arrangement of geometries in a sheet metal forming procedure.

T. Lehrer, A. Kaps, I. Lepenies, E. Raponi, M. Wagner, and F. Duddeck, "Complementing Drawability Assessment of Deep-Drawn Components With Surrogate-Based Global Sensitivity Analysis," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, vol. 10, no. 3, p. 031204, Sep. 2024, doi: 10.1115/1.4065143.

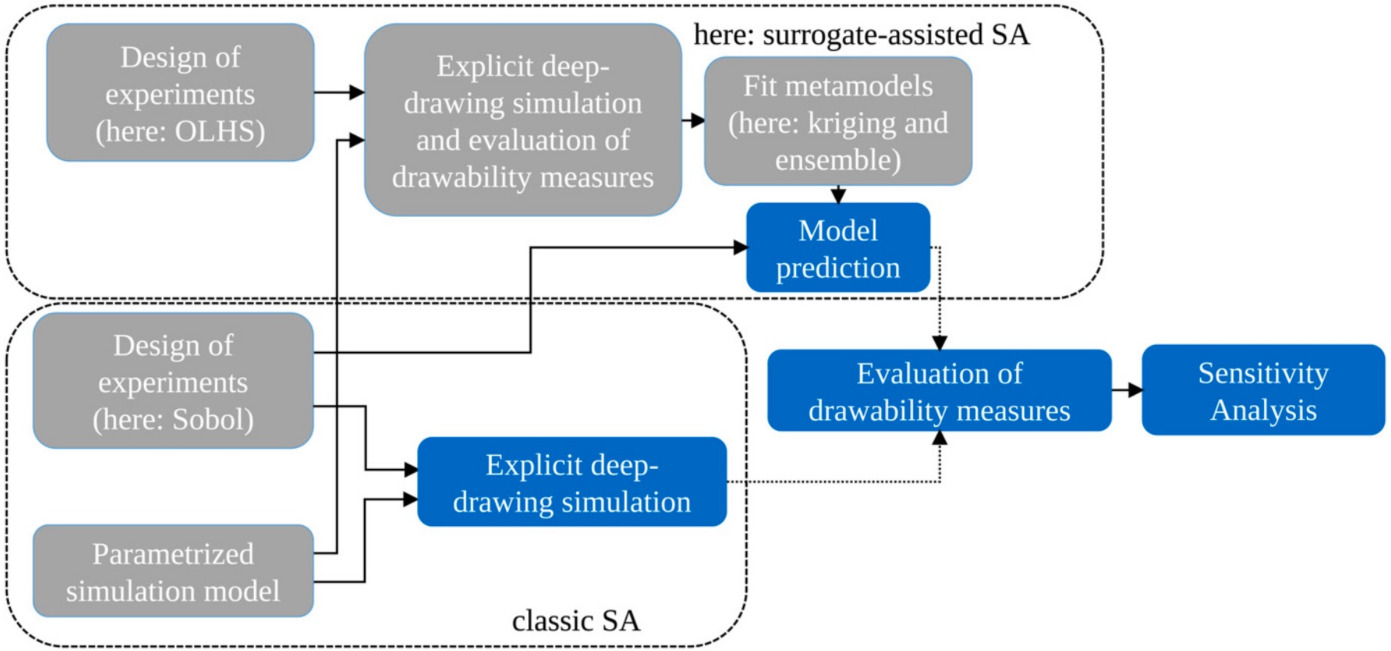
Application Case from Computational Mechanics

Workflow

- Schematic representation of the workflow for sensitivity analysis.

- Compared are a “classic” simulation-based approach and surrogate-assisted approaches based on a Kriging and an Ensemble metamodel

- Other baselines:
  - UQpyLab
  - GSAreport



T. Lehrer, A. Kaps, I. Lepenies, E. Raponi, M. Wagner, and F. Duddeck, “Complementing Drawability Assessment of Deep-Drawn Components With Surrogate-Based Global Sensitivity Analysis,” *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, vol. 10, no. 3, p. 031204, Sep. 2024, doi: 10.1115/1.4065143.

Application Case from Computational Mechanics

Problem Setup

- Parameters

Symbol	Parameter range	Parameter
t	(0.8–2.0) mm	Sheet thickness
r	(0.8–2.5)	Lankford coefficient
σ_y	(140.0–180.0) MPa	Yield strength of the blank
f_{blkh}	(130.0–190.0) kN	Blankholder force
μ_{d}	(0.05–0.12)	Dynamic coefficient of friction

- Targets: drawability measures
  - Average thickness reduction

$$f_{\mathrm{thinning}} = \left(\frac{1}{N} \sum_{e=1}^{N} (h_{\mathrm{t}}^{e} - h_0)^2 \right)^{\frac{1}{2}}$$

N = number of model elements
h_0 = initial thickness of the blank sheet
h_t^e = element thickness at the end

  - Average distance of bad elements from their closest limit curve

$$f_{\mathrm{distances}} = \begin{cases} f_{\mathrm{n-d}}, & \text{if } f_{\mathrm{n-d}} > f_{\mathrm{dt}} \\ f_{\mathrm{d}}, & \text{otherwise} \end{cases}$$

f_{n-d} = non-drawable configurations
f_d = drawable configurations
f_{dt} = drawability threshold

T. Lehrer, A. Kaps, I. Lepenies, E. Raponi, M. Wagner, and F. Duddeck, “Complementing Drawability Assessment of Deep-Drawn Components With Surrogate-Based Global Sensitivity Analysis,” *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, vol. 10, no. 3, p. 031204, Sep. 2024, doi: 10.1115/1.4065143.

Application Case from Computational Mechanics

Results

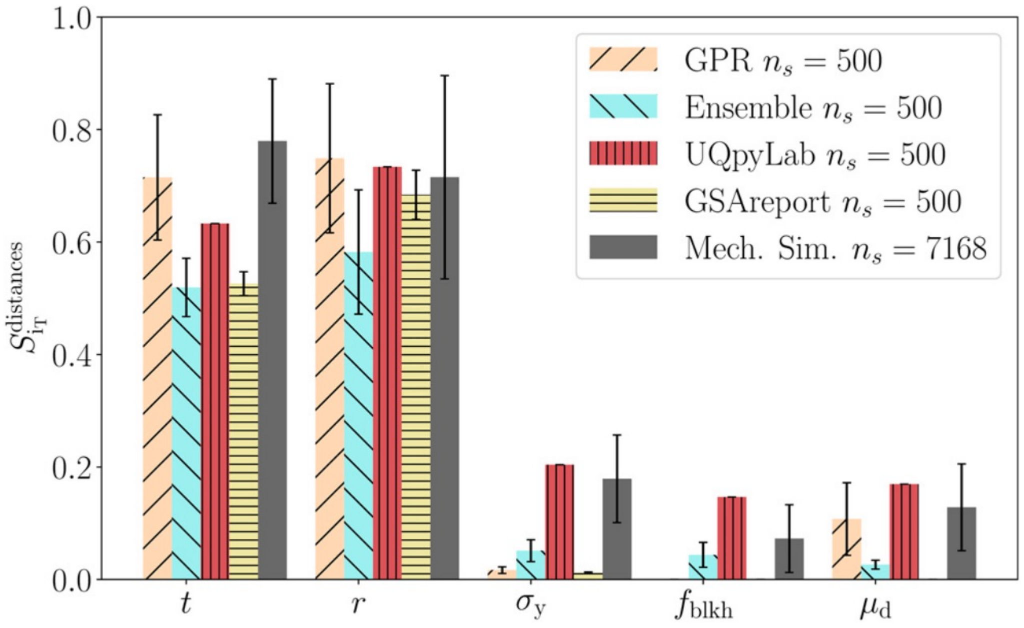


Figure: SI for the ensemble model, GPR surrogate, UQpyLab, and GSAreport. The SI based on the mechanical simulation is given as reference. Results are shown on the **distances objective function** for every parameter.

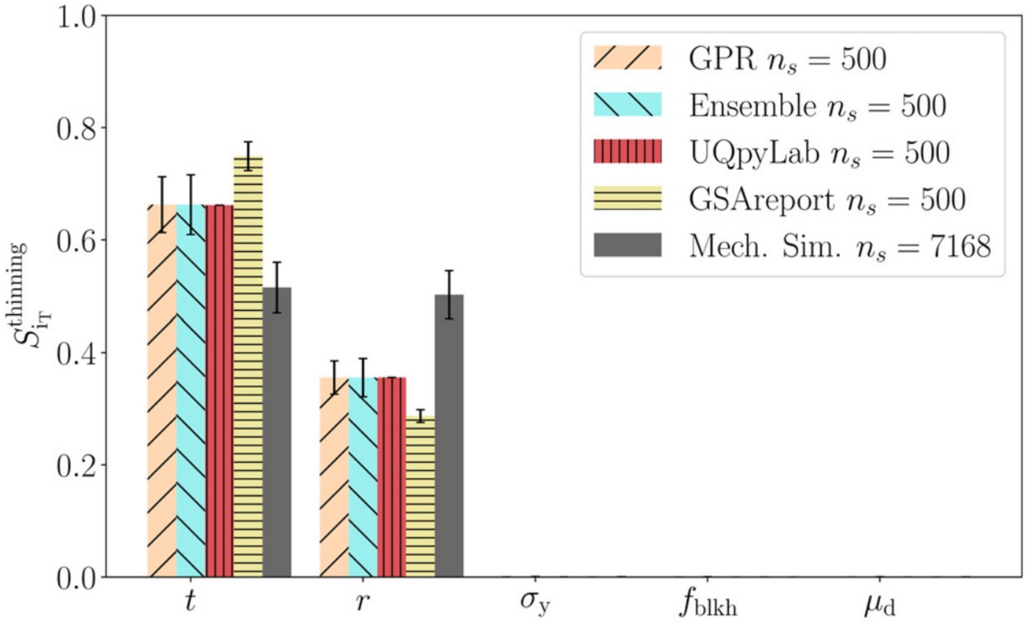


Figure: SI for the ensemble model, GPR surrogate, UQpyLab, and GSAreport. The SI based on the mechanical simulation is given as reference. Results are shown on the **thinning objective function** for every parameter.

T. Lehrer, A. Kaps, I. Lepenies, E. Raponi, M. Wagner, and F. Duddeck, "Complementing Drawability Assessment of Deep-Drawn Components With Surrogate-Based Global Sensitivity Analysis," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, vol. 10, no. 3, p. 031204, Sep. 2024, doi: [10.1115/1.4065143](https://doi.org/10.1115/1.4065143).

Application Case from Computational Mechanics

A broader analysis through GSAreport

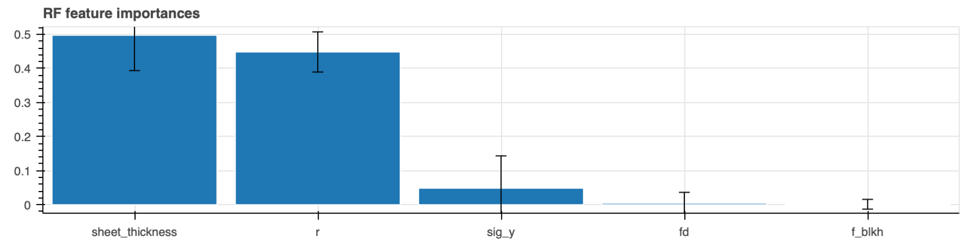
GSA Report SA

Global Sensitivity Analysis Report

Sensitivity analysis is the study of how the uncertainty in the output of a mathematical model or system (numerical or otherwise) can be divided and allocated to different sources of uncertainty in its inputs. This experiments shows the first, second and total order sensitivities of the provided features (input) and visualises them in an interactive way. The different sensitivity algorithms can provide additional insight and information in the underlying process.

Experiment: **SA**
Started: **2023-10-04 15:03:33**

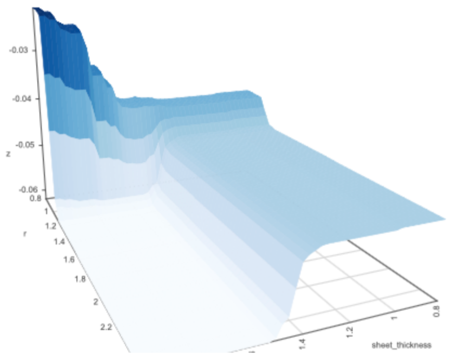
Number of parameters: 5
Showing the top 5 parameters
Random Forest mean R^2 score over 3 folds: 0.5227423370867968



Print report

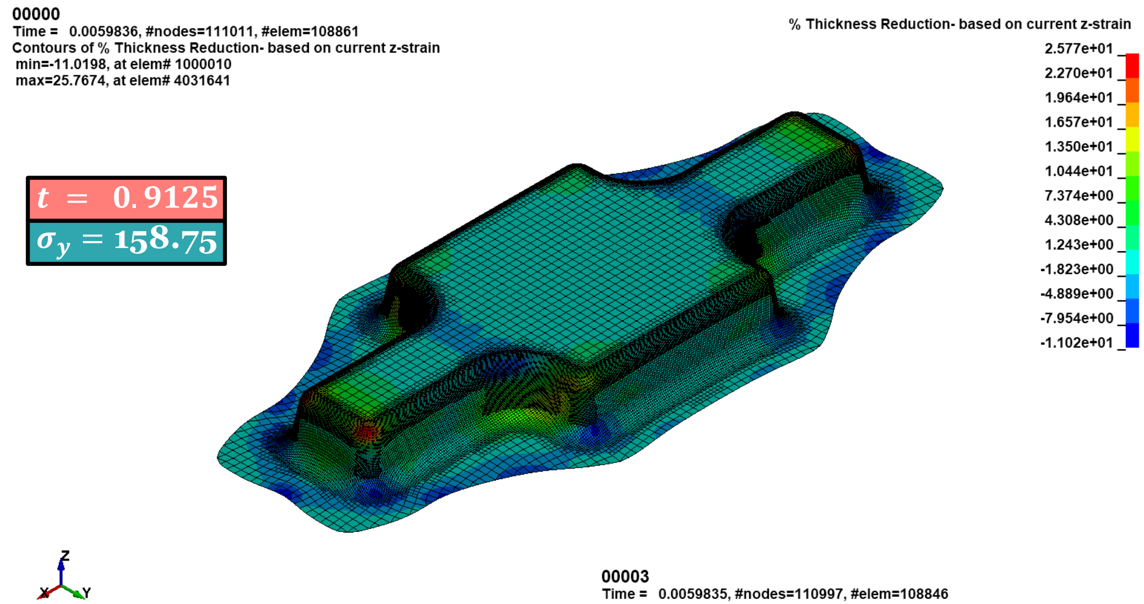
Function Landscape

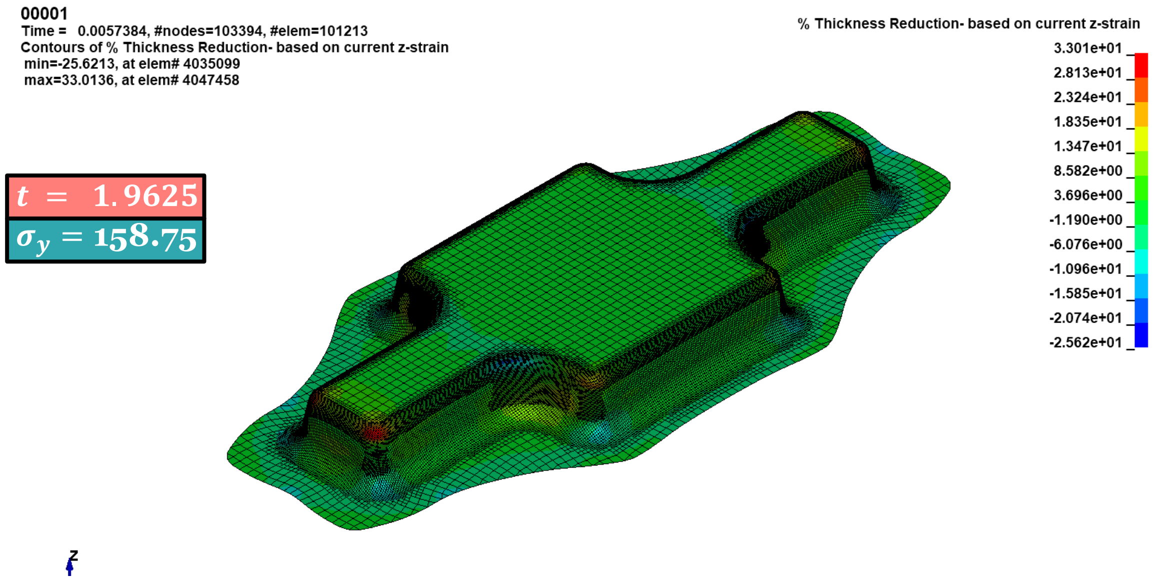
Interactive surface plot of a slice (using a Random Forest model) with sheet_thickness on the X axis and r on the Y axis. All other parameters are set to the center of their range.

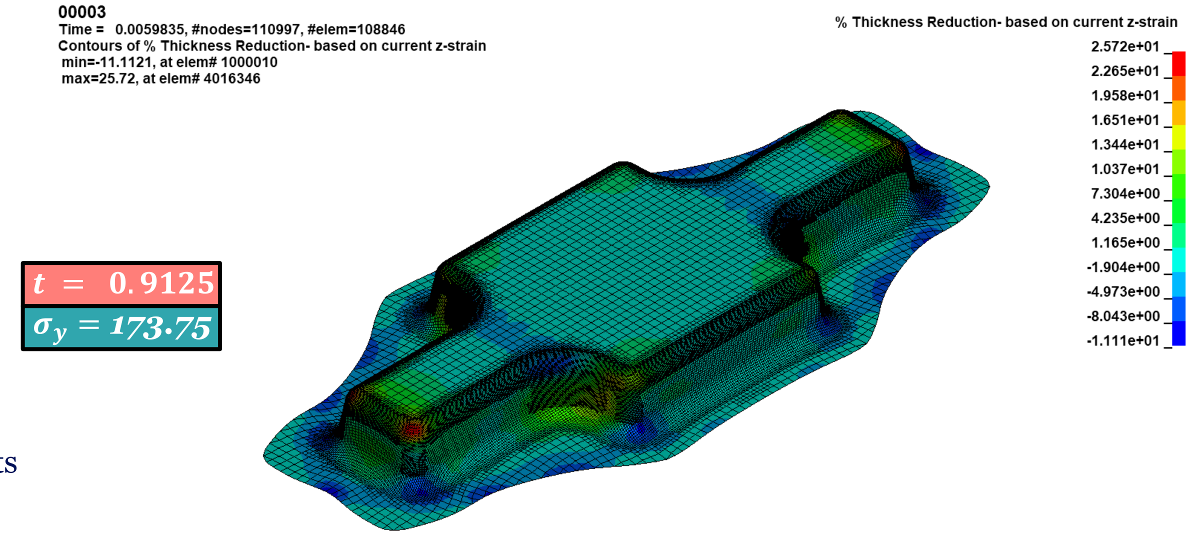


file:///Users/elenaraponi/Documents/LIACS_Mac/Talks/2024_ENBIS_spring_meeting/gsa_reports/distances_500/output/2023-10-04T15-03-SA-report.html

Application Case from Computational Mechanics







Influential
Non-influential

Figure: Mechanical simulation outputs of “extreme” drawing configurations.

Final Remarks

- **GSA** studies how the **uncertainty in the output** of a model can be apportioned to different sources of **uncertainty in the model input**.
- **Crucial during** the **early stages** of investigations to determine the influential parameters in laboratory or simulation-based experiments.
- **GSAreport**, a **user-friendly**, **knowledge-agnostic** tool that generates an interactive report using different GSA techniques.
- **Useful** tool **for** (1) **non-professionals** and (2) **experts in the field** who need a **fast** preliminary **analysis** of the problem.
- Performance demonstrated on application case from mechanics: a **drawability assessment of deep-drawn components**.
- Relatively new tool → **Feedback** is **welcome**.

Let's make Industrial Data Science Trustworthy together! 💪

GSA

GitHub

[1] B. Van Stein, **E. Raponi**, Z. Sadeghi, N. Bouman, R. C. H. J. Van Ham, and T. Bäck, "A Comparison of Global Sensitivity Analysis Methods for Explainable AI with an Application in Genomic Prediction," *IEEE Access*, pp. 1–1, 2022

[2] T. Lehrer, A. Kaps, I. Lepenies, **E. Raponi**, M. Wagner, and F. Duddeck, "Complementing Drawability Assessment of Deep-Drawn Components With Surrogate-Based Global Sensitivity Analysis," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, vol. 10, no. 3, p. 031204, Sep. 2024, doi: 10.1115/1.4065143.

Elena Raponi (LIACS)
e.raponi@liacs.leidenuniv.nl