technische universität dortmund

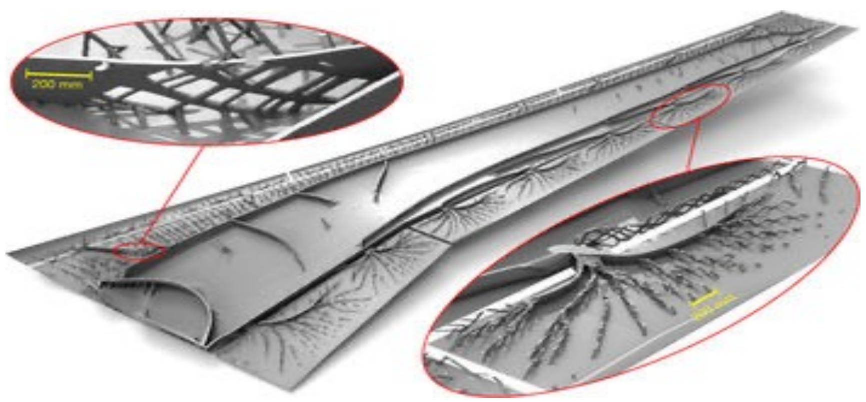
Chair for Reliability Engineering

Efficient numerical methods to deal with imprecise probabilities

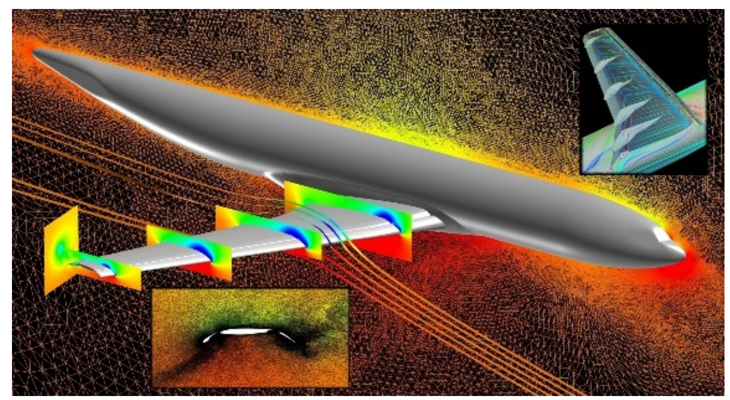
Prof. Dr. Matthias Faes
TU Dortmund, Chair for Reliability Engineering

Introduction

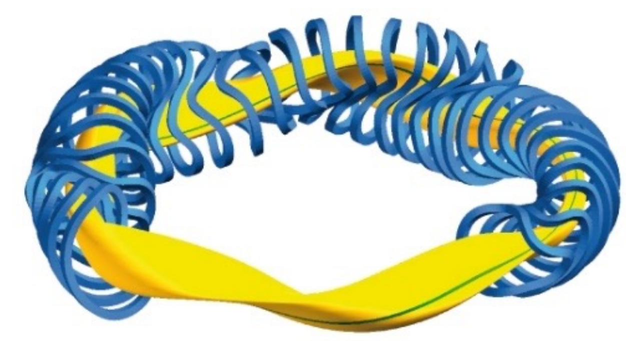
- modern design methods allow for assessment of structural quality and thorough design optimization long before first part has been produced



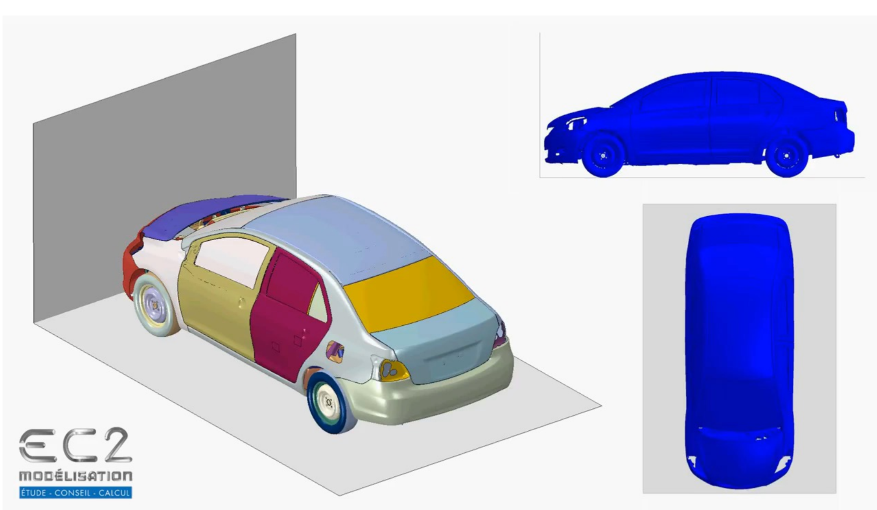
Topology optimised Airplane wing
(Nature)



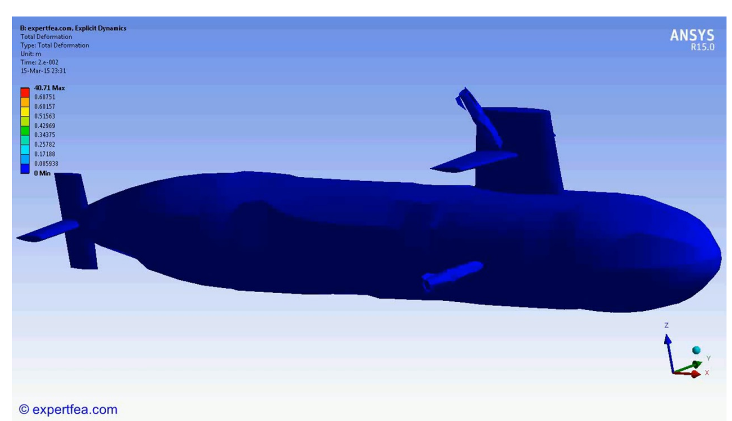
Airflow around Boeing 737 body
(NASA / Boeing)



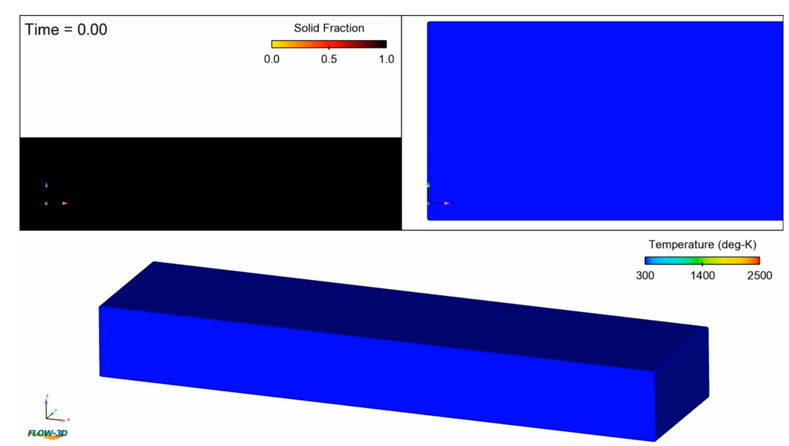
Wendelstein 7-X fusion reactor
(Max Planck Institute)



Car crash simulation
(Toyota Yaris)



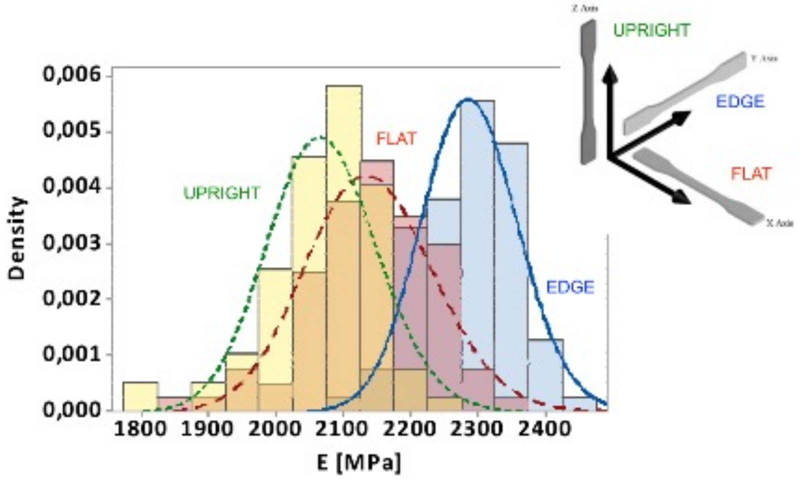
Dual torpedo impact
(Ansys)

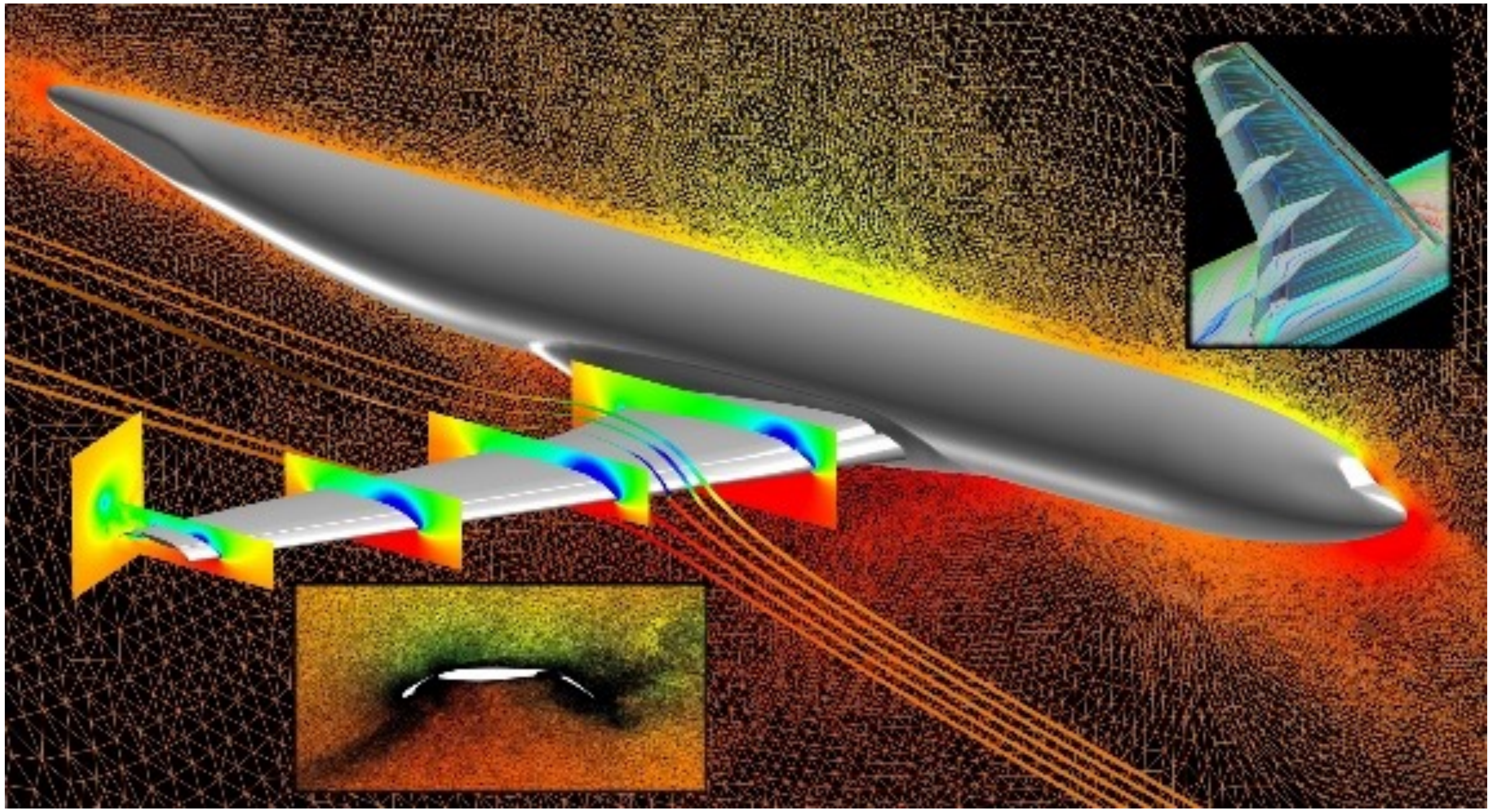


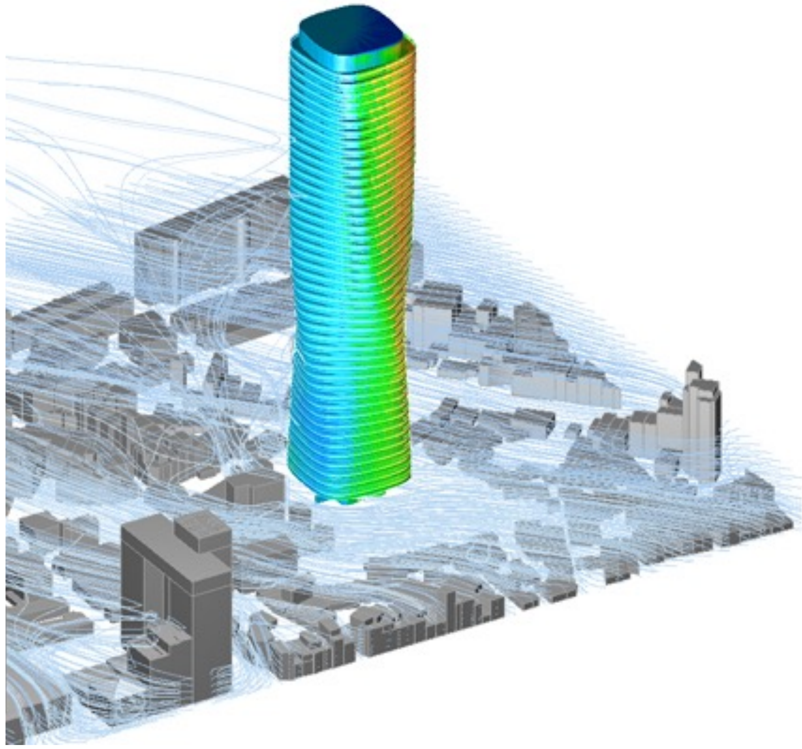
Laser Metal Deposition
(FlowScience)

However…

- high-fidelity models ⟺ complex structural behaviour
- many model variables subjected to uncertainty:
  - macro and micro scale inter- and intra-variability,
  - insufficiently known or variable loading,
  - approximation of complicated physics
- subjective human interpretation







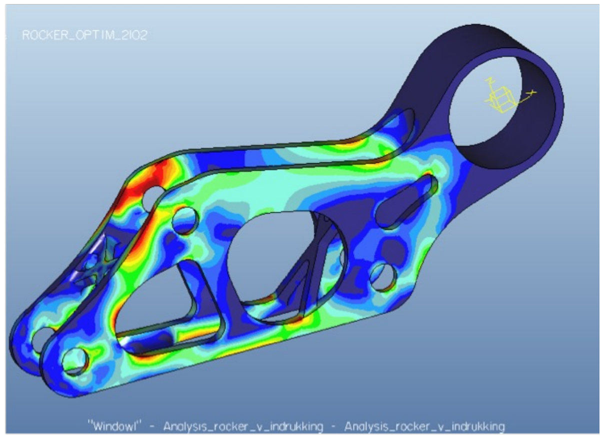
Model validity

- Advanced numerical engineering tools ⇝ only useful when models are valid
  - model validation and verification
  - model updating
- Inclusion of uncertainties:

does the deterministic model give results that are close enough to my experimental observations

⇩

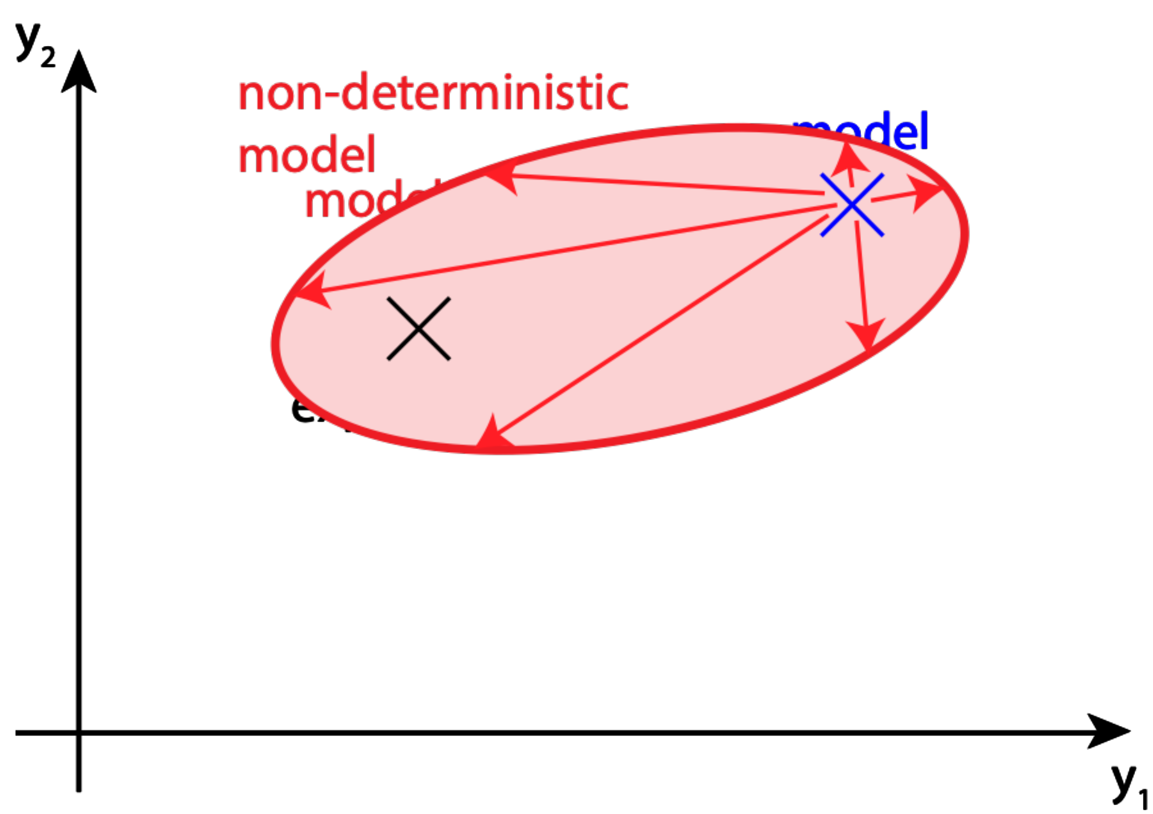
can the experimental observations be interpreted as a likely realisation of my non-deterministic model



⇔

Introduction

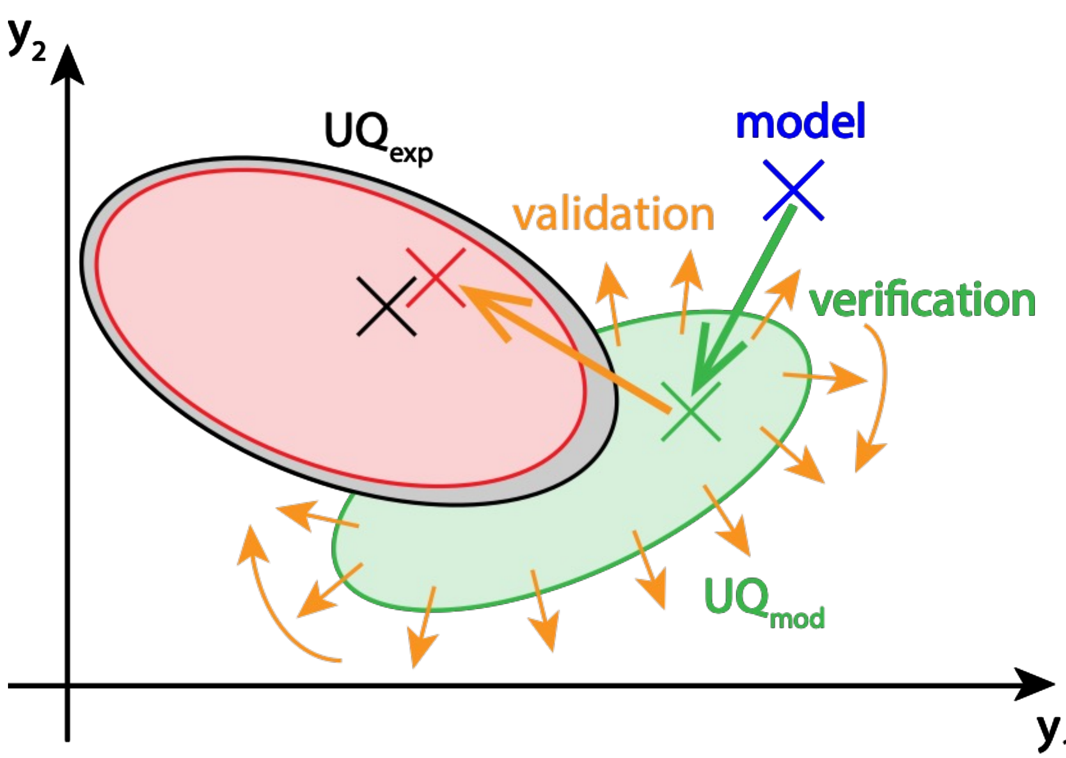
- **model validity** interpretation
  - modelling for exactness: bring model as close as possible to “reality”
    ⟶ what parameters do we tune? when is the result realistic? Is the problem well posed? is the solution unique?
  - modelling for robustness & reliability: include uncertainty that covers observation
    ⟶ what parameters? or non-parametric? what variability is realistic?



"A low-fidelity answer with known uncertainty bounds is more valuable than a high-fidelity answer with unknown uncertainty bounds" [NASA White Paper, 2002]

Introduction

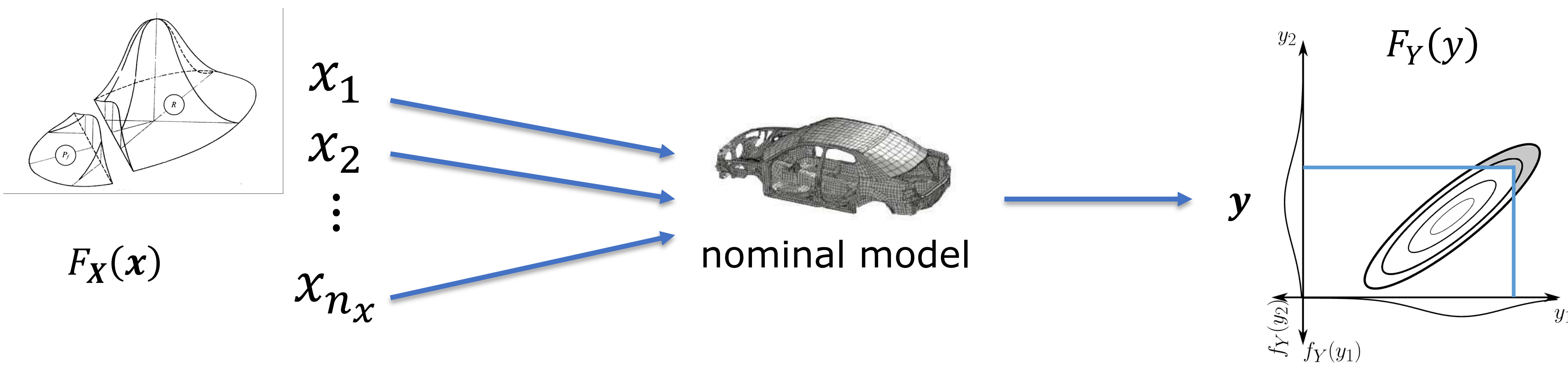
- **non-deterministic V&V**
  - verification deals with (reduces) error
  - model uncertainty quantification (UQ_{mod}) ⟶ uncertainty on numerical side
  - measurement uncertainty quantification (UQ_{exp}) ⟶ uncertainty on observations
  - validation now about matching UQ_{exp} and Uq_{mod}



How to model these uncertainties?

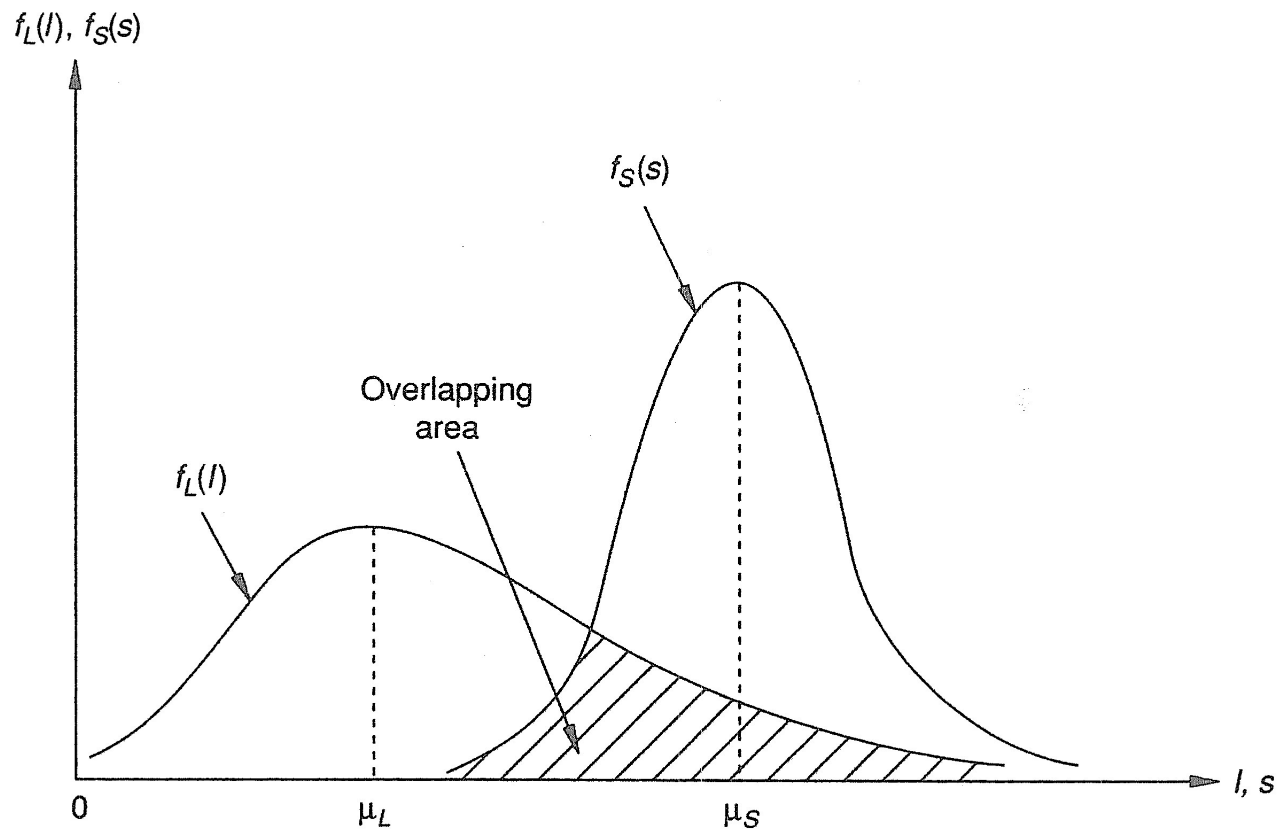
- Probability theory offers a complete framework to model variability
- Random variables $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_{n_x}\right)$ with support D_X
- Probability that $\boldsymbol{X}$ is less or equal than $\boldsymbol{x}$ is modelled as joint probability distribution function $F_{\boldsymbol{X}}(\boldsymbol{x}) = P\left(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_{n_x} < x_{n_x}\right)$ for $\boldsymbol{x} \in D_x$
- Joint probability density function f_X is the derivative of $F_{\boldsymbol{X}}$, i.e., $f_X = \frac{d}{dx} F_X(x)$

- Let $\mathcal{M}: \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_y}, \boldsymbol{x} \rightarrow \boldsymbol{y}$ denote a function representing the numerical model under consideration
- $F_{\boldsymbol{Y}}(\boldsymbol{y})$ represents the joint cumulative distribution function of the responses



Reliability analysis

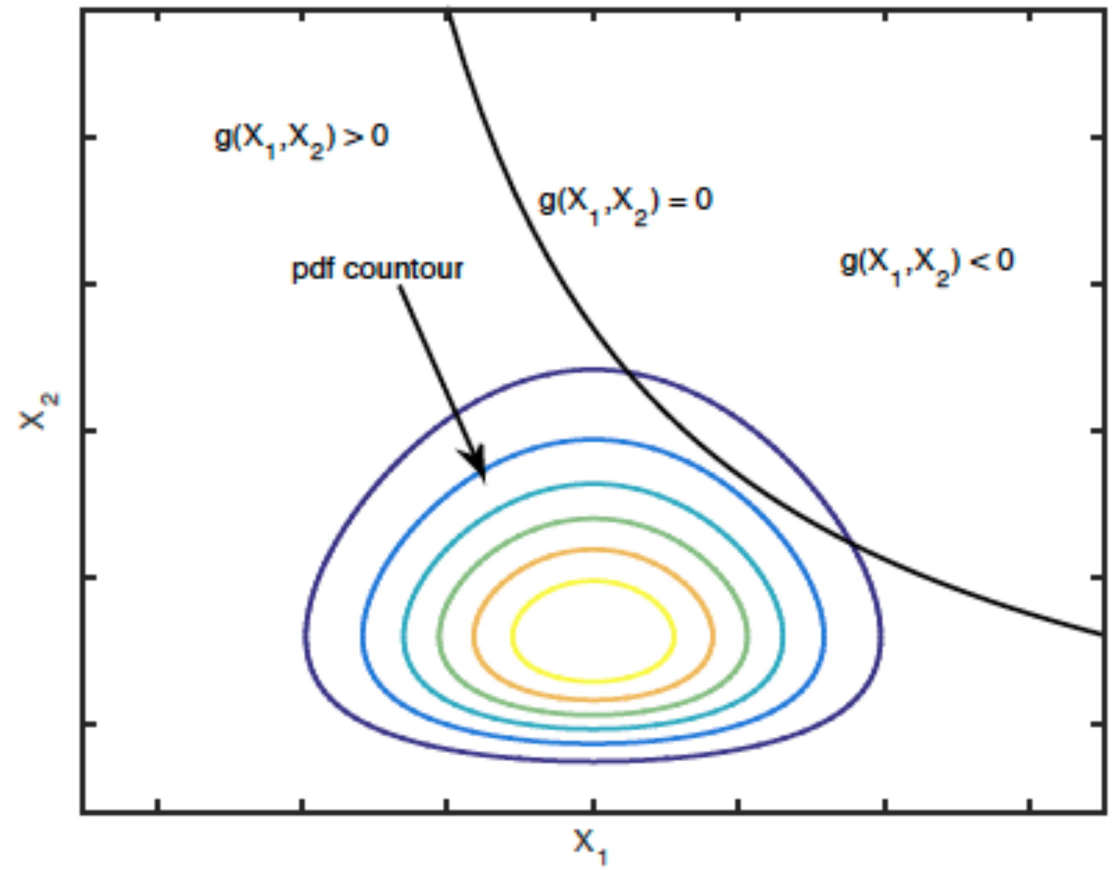
- Most simple case: strength and load are independent
- Load
  - the loading condition of the material $L = y(\boldsymbol{x})$ (e.g., tensile force)
  - Distributed as $L \sim f_L(l)$
  - Moments μ_L, σ_L
- Strength
  - Critical performance of the material $S = y_c$ (e.g., R_m)
  - Distributed as $S \sim f_S(s)$
  - Moments μ_S, σ_s
- Overlapping area: probability of failure

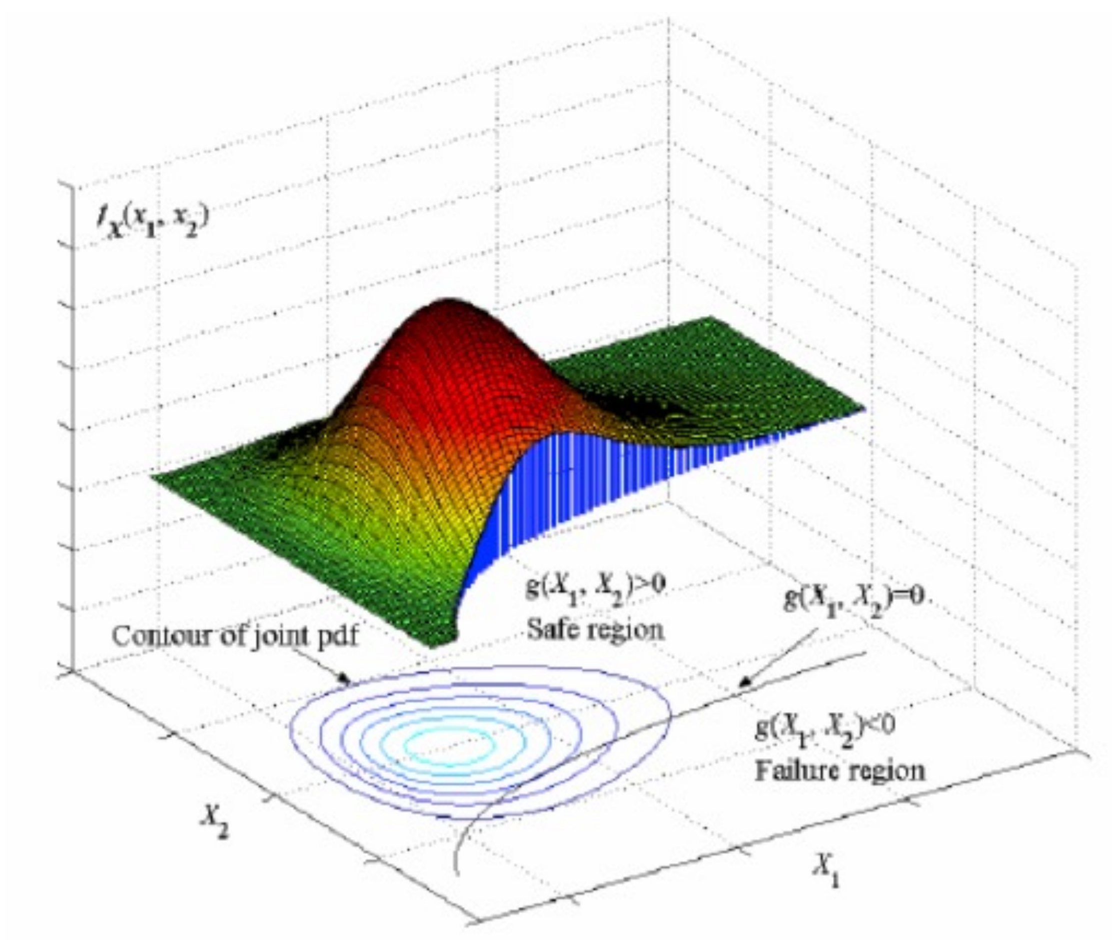


Reliability analysis

- In general multiple variable input quantities and failure modes
- Definition of performance function:
$$Z = g(\boldsymbol{X})$$
- Failure domain: region of the random variable space where $g \leq 0$
- Safe domain: region of the random variable space where $g > 0$
- Limit state function: $N-1$ dimensional curve for which $g(X_1, X_2, \ldots, X_N) = 0$
- Probability of Failure:

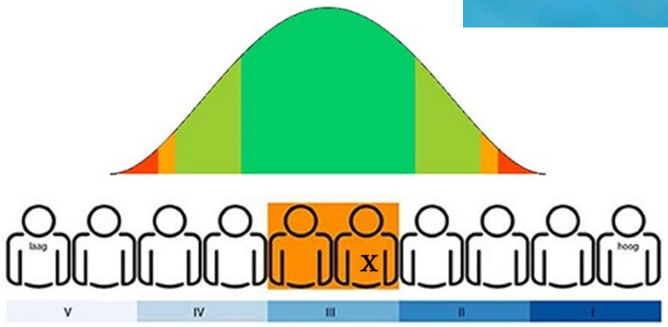
$$P_F = P(g(X_1, X_2, \ldots, X_N) \leq 0)$$
$$P_f = \int \int \ldots \int_{g(\boldsymbol{x})<0} f_X(\boldsymbol{x}) dx_1 dx_2 \ldots dx_n$$

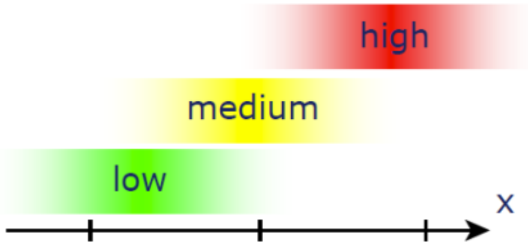


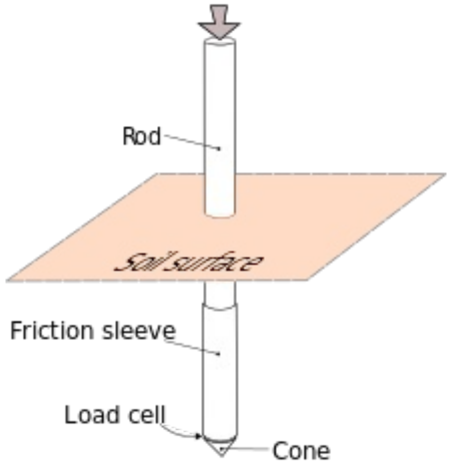


Reliability analysis: challenge

- $f_X(\boldsymbol{x})$ needs to be estimated to perform reliability analysis
- However, estimation is complicated by:
  - Imprecise measurements
  - Small sample set sizes
  - Incomplete expert elicitations
  - Changing environmental conditions
  - Vague or dubious information
  - Expert assessment / experience
  - Linguistic assessments
  - Conditional probabilities observed under unclear conditions
  - Only marginals are available
  - …







→ Mixtures of information from several sources with different levels of fidelity

Overview

Imprecise probabilities

How to deal with data issues in reliability analysis?

Propagation of p-boxes

Conclusions

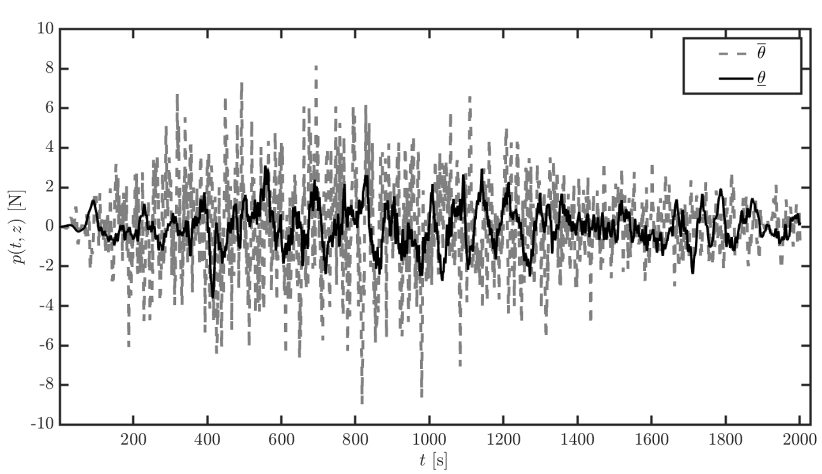
Imprecise probabilities

How to deal with data issues in reliability analysis?

Classification and modelling of imprecise information
By origins of uncertainty

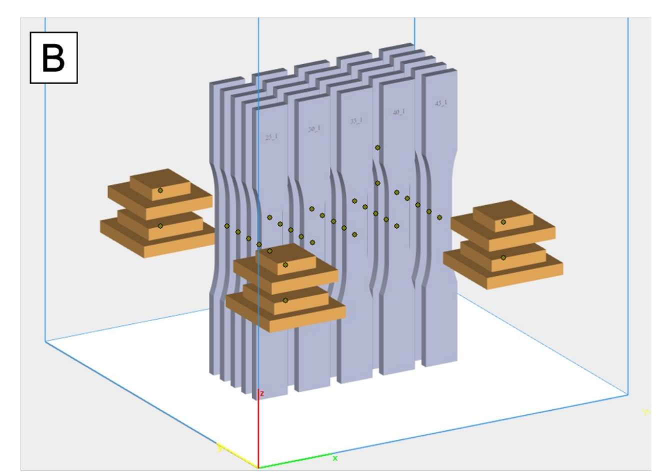
Aleatory uncertainty
- Irreducible uncertainty
- Caused by variability/fluctuations
- Property of the system
- Stochastic characteristics

→ Traditional probabilistic models
→ classic *variability*



Epistemic uncertainty
- Reducible uncertainty
- Caused by lack of knowledge
- Property of the analyst or analysis
- Inconsistency of information

→ No specific model predefined



Both sources tend to occur at the same time. How to integrate them both in our calculations?

Probability boxes

Idea: provide set of possible distribution functions $F_{\boldsymbol{X}}(\boldsymbol{x})$ bounded by lower CDF $\underline{F}_{\boldsymbol{X}}(x) \in \mathbb{F}$ and upper CDF $\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) \in \mathbb{F}$, with $\mathbb{F}$ the set of all CDFs on $D_X \subseteq \mathbb{R}$

- Formally, a p-box is defined as the set $\{F_X(x) \in \mathbb{F} | \underline{F}_{\boldsymbol{X}}(x) \leq F_X(x) \leq \overline{F}_X(x), x \in D_X\}$

- Epistemic uncertainty on $F_X(x)$ is accounted for explicitly by assigning an interval $[\underline{F}_X(x), \overline{F}_X(x)]$ for each value of $x \in D_X$

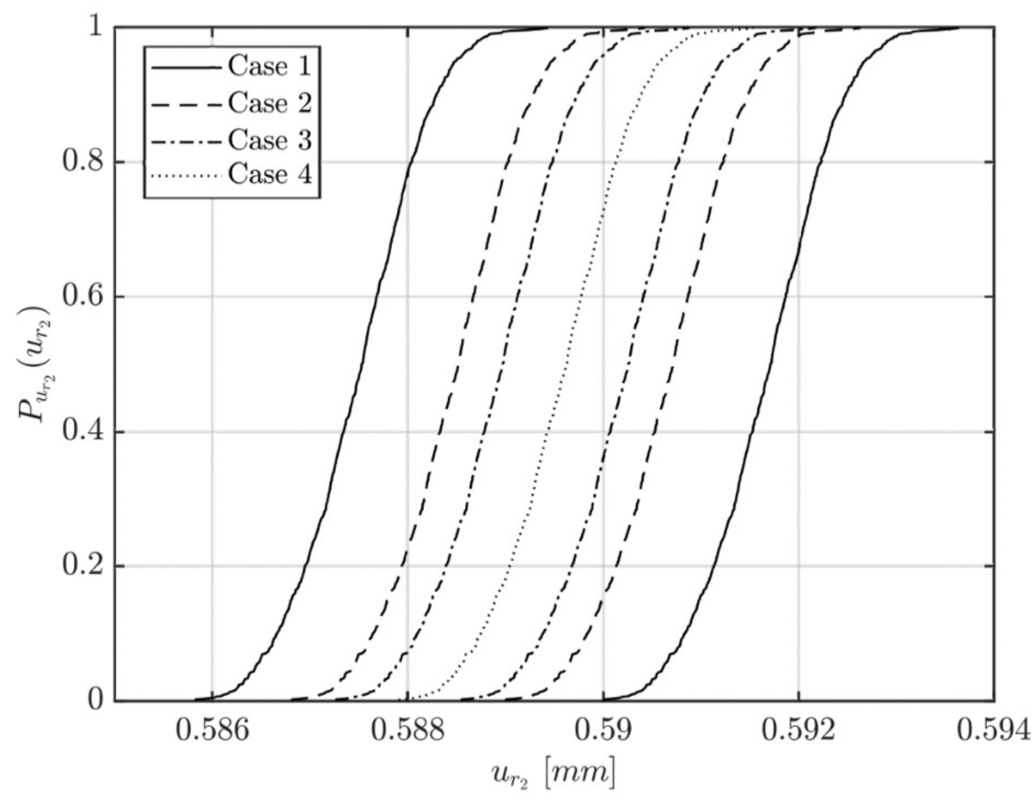
- Small epistemic uncertainty: $[\underline{F}_X(x), \overline{F}_X(x)]$ is a tight interval

  → Large confidence in CDF and results

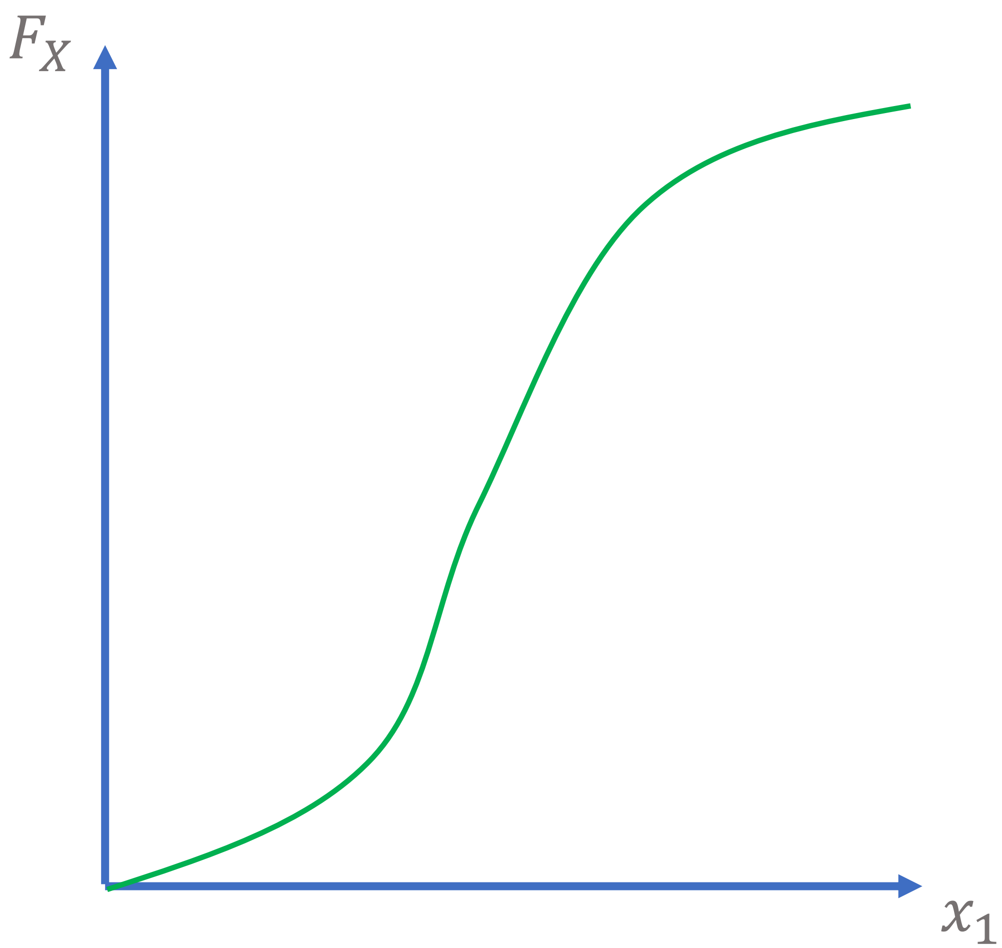
- Large epistemic uncertainty: $[\underline{F}_X(x), \overline{F}_X(x)]$ is a wide interval

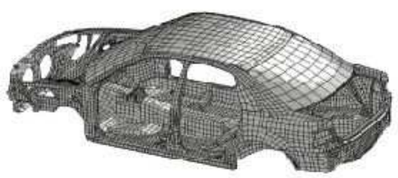
  → Low confidence in CDF and results

  → collect more data



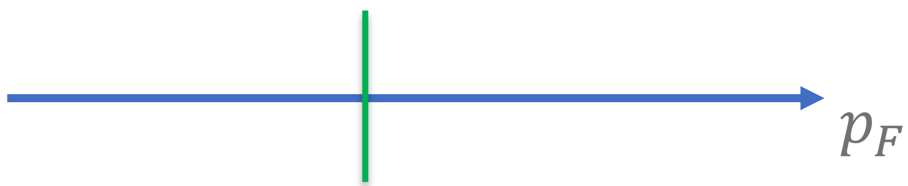
Reliability analysis





nominal model

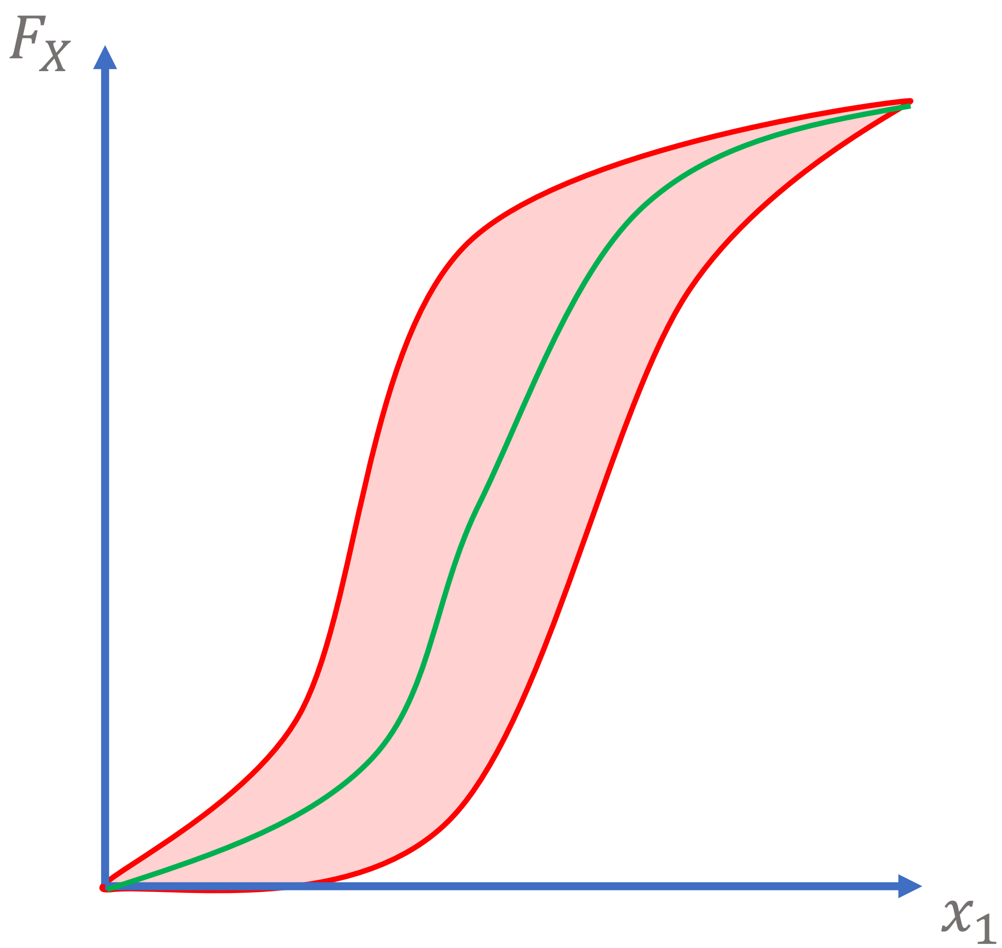
$$p_F = \int \int \dots \int_{g(\boldsymbol{x})<0} f_X(\boldsymbol{x}) dx_1 dx_2 \dots dx_n$$

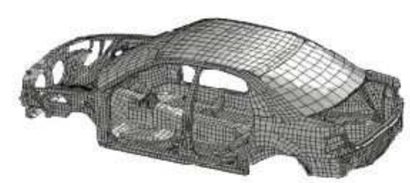


Imprecise information

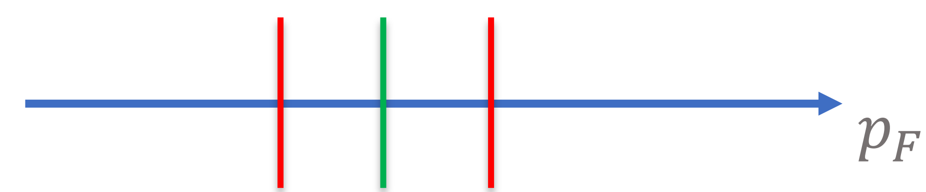
$$\underline{p}_F = \min_{f_X} \int \int \dots \int_{g(\boldsymbol{x})<0} f_X(\boldsymbol{x}) dx_1 dx_2 \dots dx_n$$

$$\overline{p}_F = \max_{f_X} \int \int \dots \int_{g(\boldsymbol{x})<0} f_X(\boldsymbol{x}) dx_1 dx_2 \dots dx_n$$





nominal model



Measure for “trustworthiness” of results

Propagation of p-boxes

General idea

- Pure probabilistic context (i.e., only aleatory uncertainty):

$$\mathcal{P} = E[\mathcal{H}(X)] = \int_{D_x} \mathcal{H}(\boldsymbol{x}) f_{\boldsymbol{X}}(x) d\boldsymbol{x}$$

- In case $\boldsymbol{X}$ represents a p-box, $\underline{\mathcal{P}} \leq \mathcal{P} \leq \overline{\mathcal{P}}$ are obtained as:

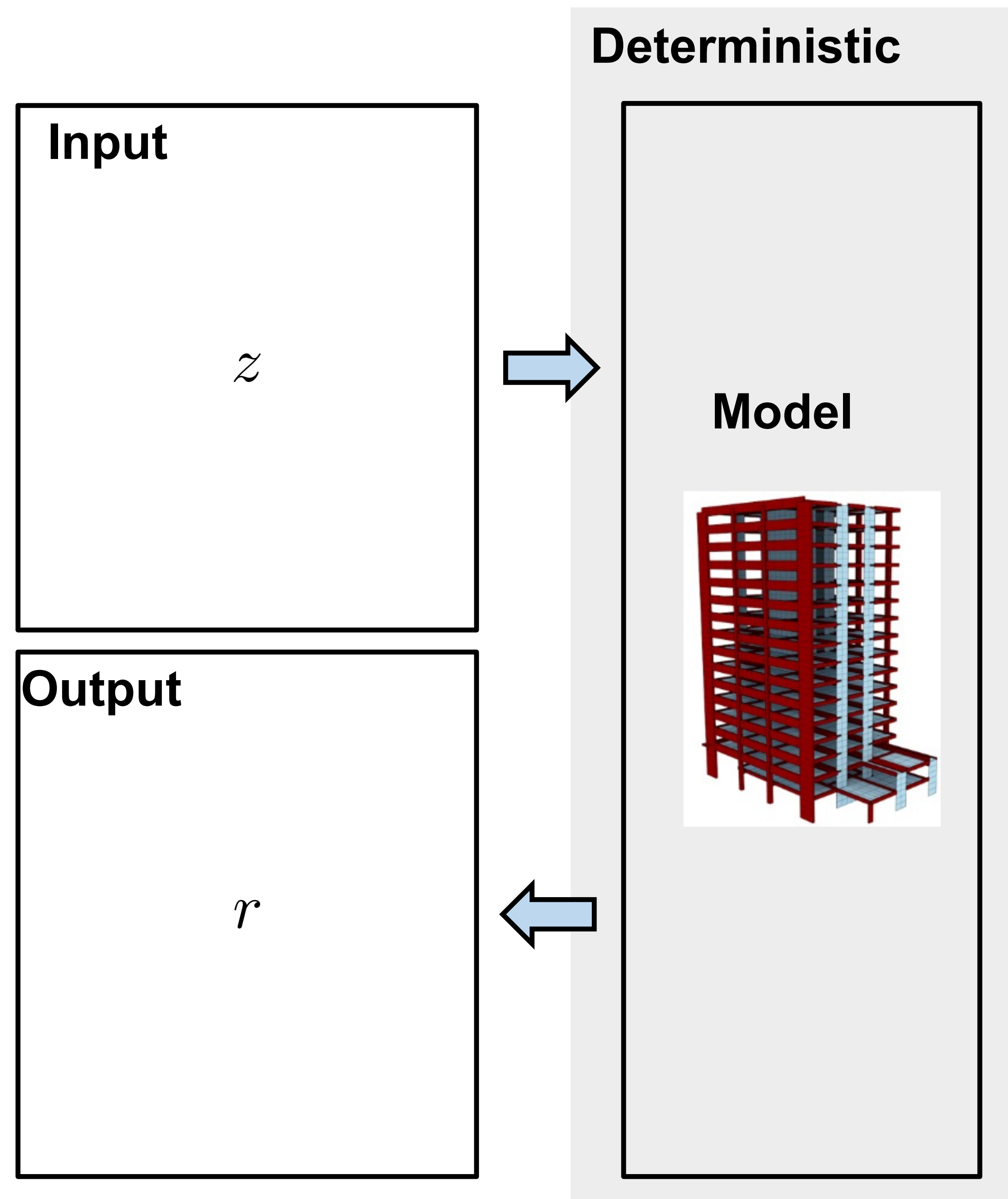
$$\underline{\mathcal{P}} = \min_{f_X} \int_{D_x} \mathcal{H}(\boldsymbol{x}) f_{\boldsymbol{X}}(x) d\boldsymbol{x}$$

$$\overline{\mathcal{P}} = \max_{f_X} \int_{D_x} \mathcal{H}(\boldsymbol{x}) f_{\boldsymbol{X}}(x) d\boldsymbol{x}$$

- Optimization over all possible $f_{\boldsymbol{X}}$ consistent with the definition of the p-box
- Nested optimization
  - Inner loop: reliability problem
  - Outer loop: (non-convex) optimization problem
- Three classes of methods
  - Double-loop approaches
  - Decoupling approaches
  - Surrogate modeling schemes

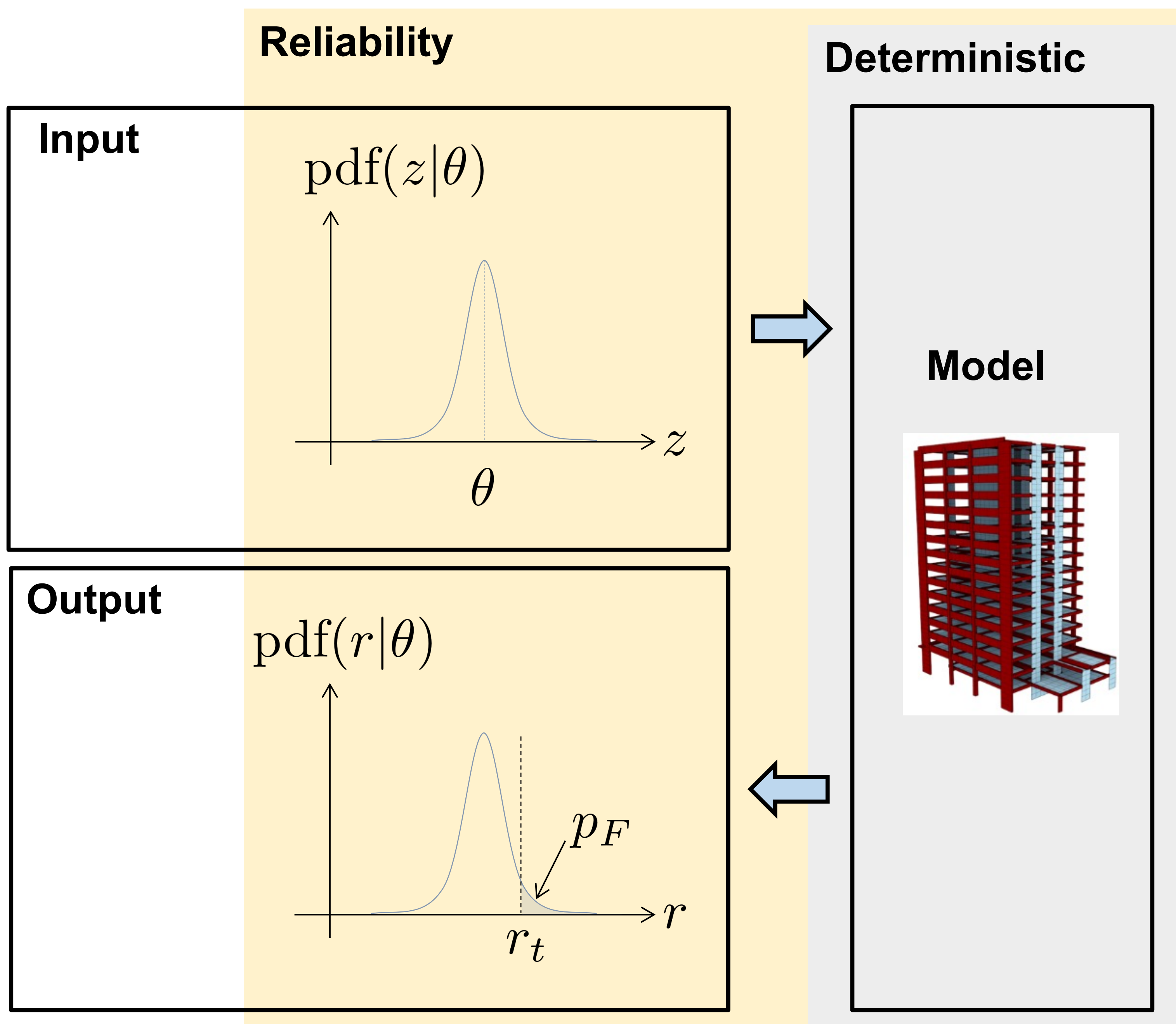
Motivation – Deterministic Analysis

- Engineering system represented by numerical **model** (e.g. finite elements)
- Model depends on inputs z (forces, material properties, etc.)
- Response of interest r



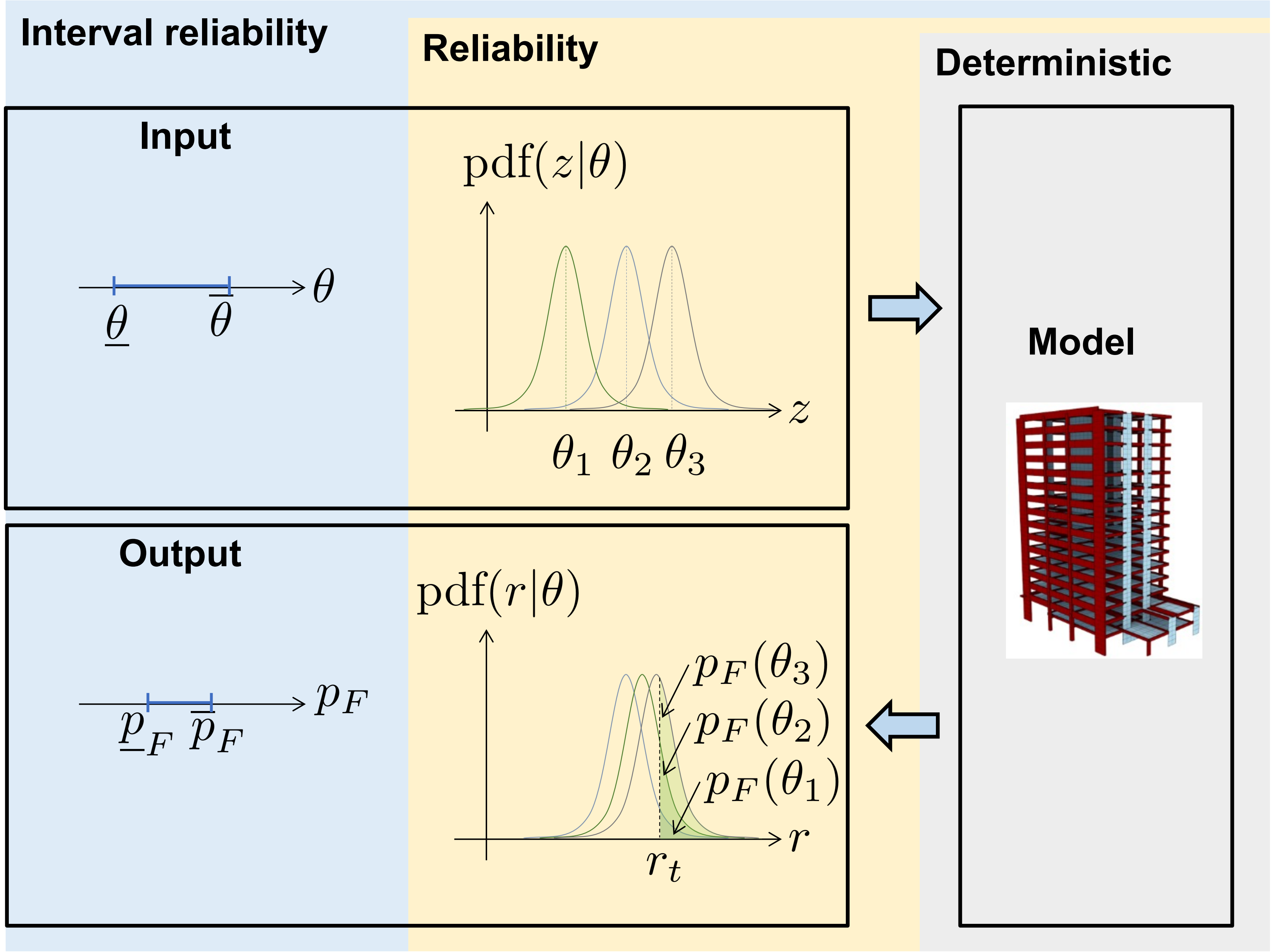
Motivation – Reliability Analysis

- **Aleatory uncertainty** of input modeled with probability distribution, depends on parameter θ (e.g. mean)
- Response must not exceed threshold r_t
- Probability of undesirable behavior: p_F



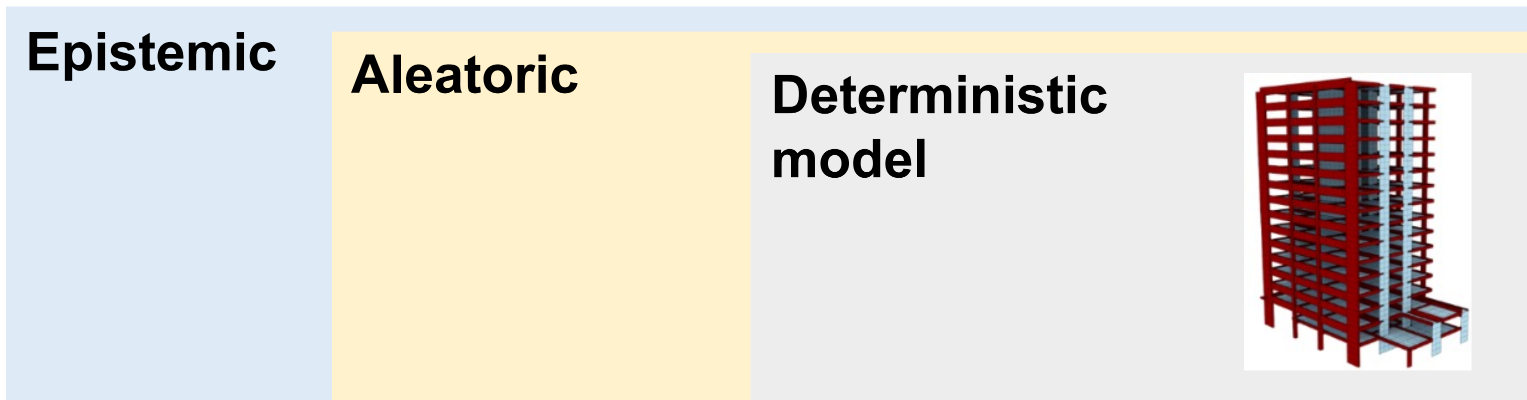
Motivation – Interval Reliability Analysis

- **Epistemic uncertainty** on θ modeled as interval (***parametric p-box***)
- p_F belongs to an interval



Decoupling approaches

- **Coping** with aleatoric and epistemic uncertainty: **huge challenge!**



- what we would like: decoupling of the uncertainty

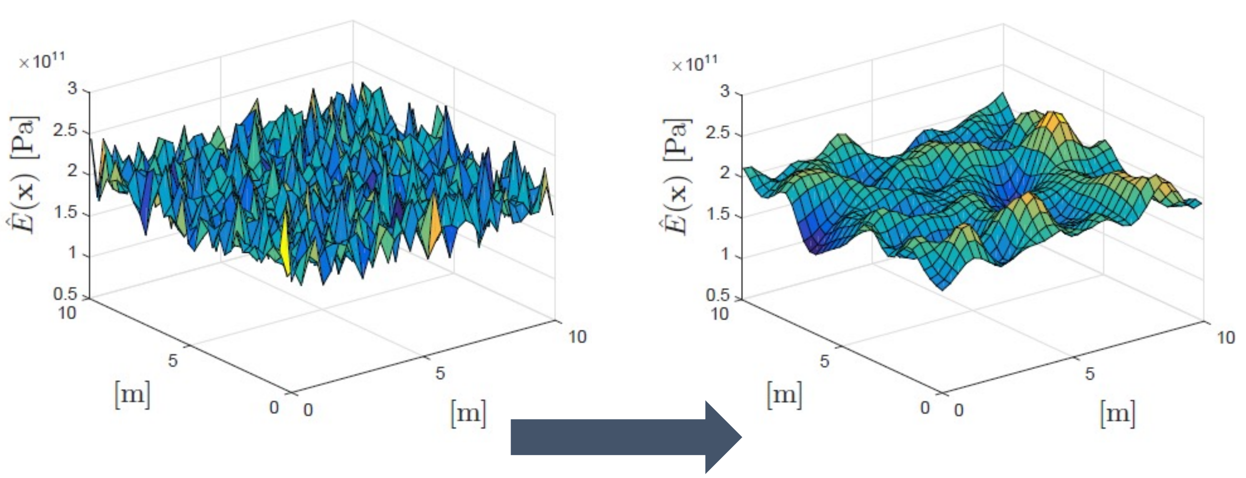
Scope

- Linear systems subject to epistemic ($\boldsymbol{\theta}_s$) and aleatory ($\boldsymbol{y}$) uncertainties
- Gaussian forces affected subject to ($\boldsymbol{\theta}_f$) and aleatory ($\boldsymbol{z}$) uncertainty

response | system | force

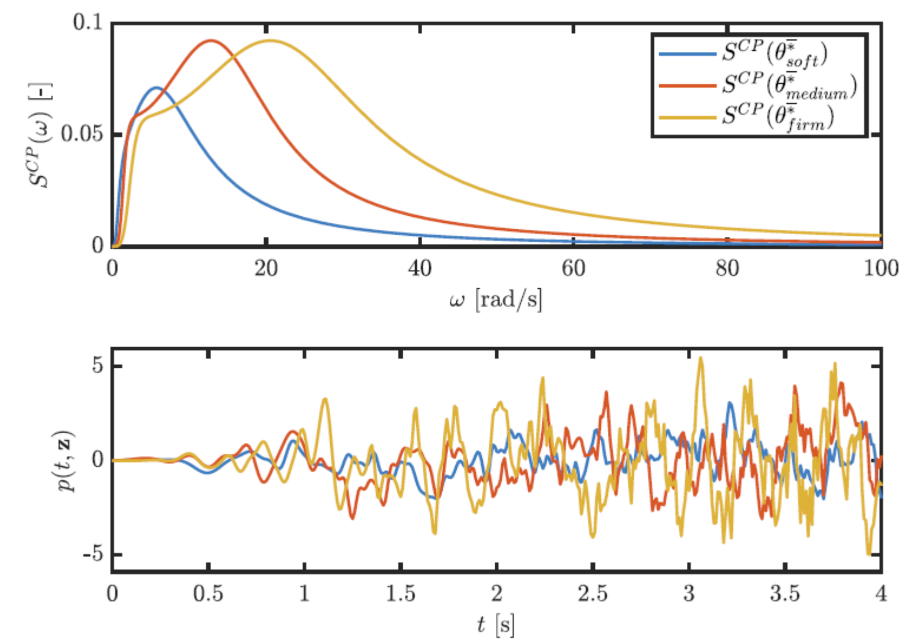
$$\boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{A}(\boldsymbol{\theta}_s, \boldsymbol{y}) \boldsymbol{f}(\boldsymbol{\theta}_f, \boldsymbol{z})$$

Imprecise Random field

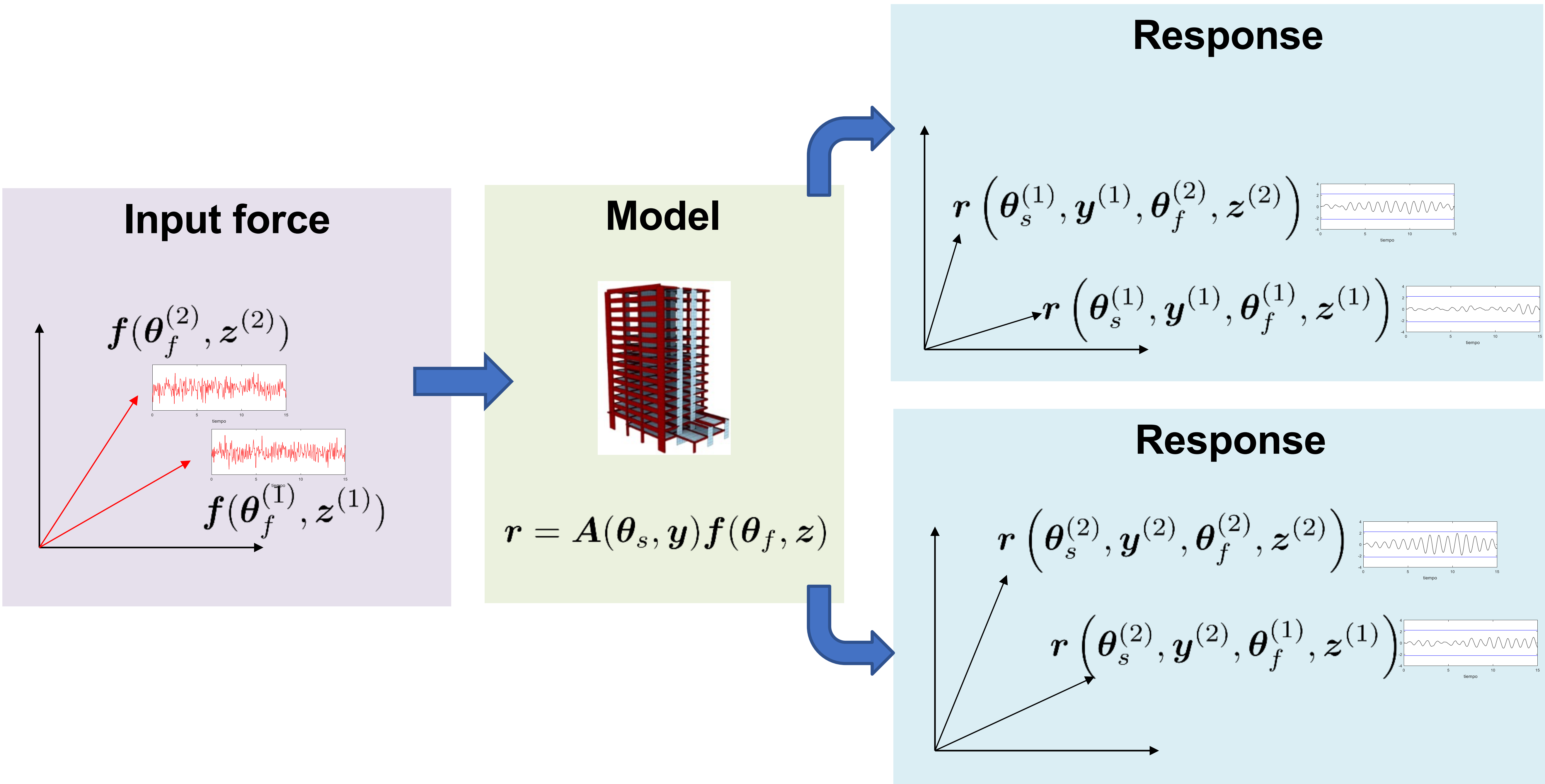


Correlation length increases

Imprecise PSD

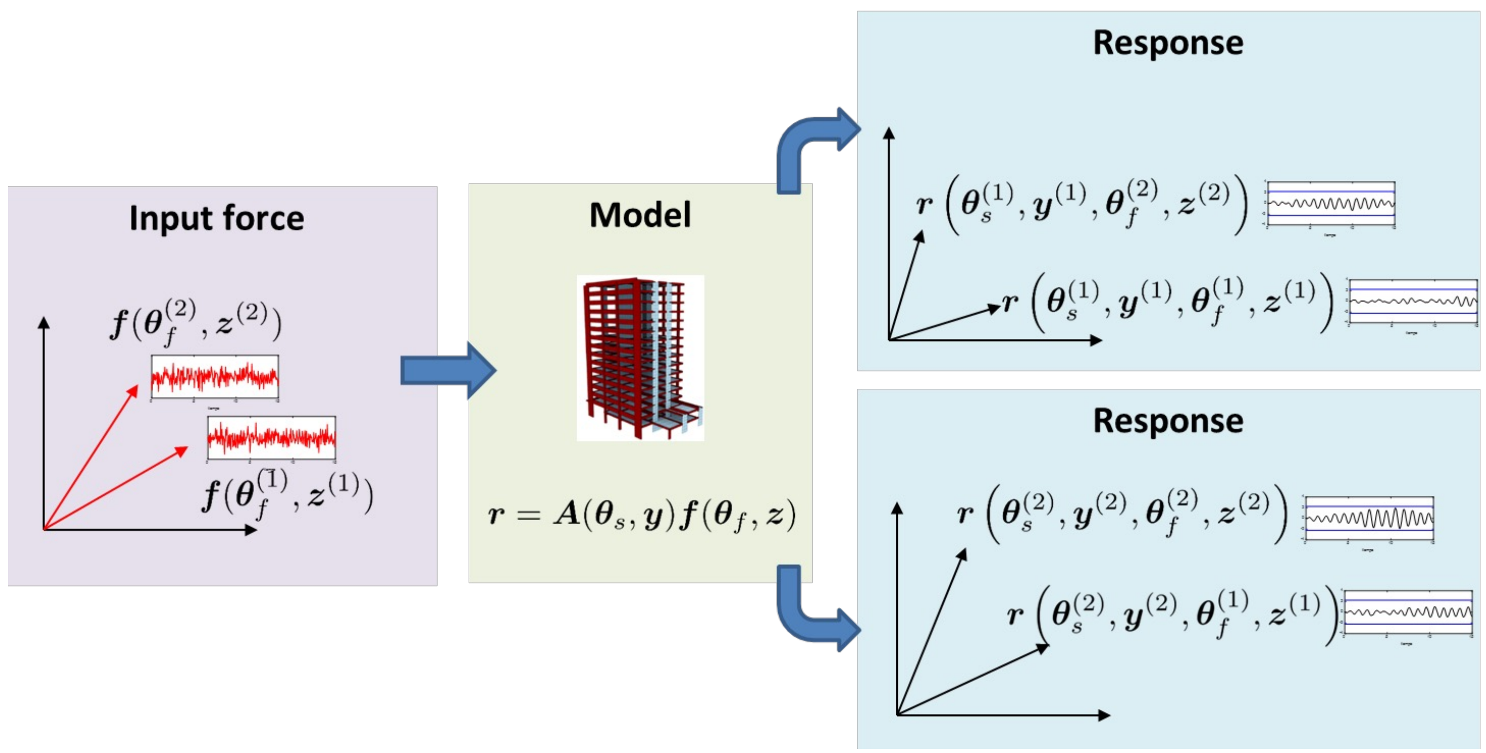


Operator Norm Theorem (1/2)



- Response $\boldsymbol{r}$ is the result of ***stretching*** loading $\boldsymbol{f}(\boldsymbol{\theta}_f, \boldsymbol{z})$ by $\boldsymbol{A}(\boldsymbol{\theta}_s, \boldsymbol{y})$
- <u>Less stretching</u> leads to **smaller** $\boldsymbol{p_F}$; <u>more stretching</u> leads to **higher** $\boldsymbol{p_F}$

Operator Norm Theorem (2/2)



- The amount of ***stretching*** induced by $\boldsymbol{A}$ can be bounded by **operator norm theorem**

$$||\boldsymbol{A}||_{\infty,2} = \inf\{c \geq 0 : ||\boldsymbol{A}\boldsymbol{v}||_{\infty} \leq c\,||\boldsymbol{v}||_2\}$$

- In this case, operator norm corresponds to **maximum standard deviation of response** $\sigma_{\max}(\boldsymbol{\theta})$

Proposed approach

Optimization (epistemic parameters)

Deterministic model

+

Reliability (aleatory Parameters)

Deterministic model

$$\underline{\boldsymbol{\theta}}^* = \operatorname*{argmin}_{\boldsymbol{\theta}\in[\underline{\boldsymbol{\theta}},\overline{\boldsymbol{\theta}}]} \left(\sigma_{\max}(\boldsymbol{\theta})\right)$$

$$\overline{\boldsymbol{\theta}}^* = \operatorname*{argmax}_{\boldsymbol{\theta}\in[\underline{\boldsymbol{\theta}},\overline{\boldsymbol{\theta}}]} \left(\sigma_{\max}(\boldsymbol{\theta})\right)$$

Determined <u>deterministically</u> via linearization
⇝ only required for stochastic system matrices

$\underline{p}_F$ $\overline{p}_F$ p_F

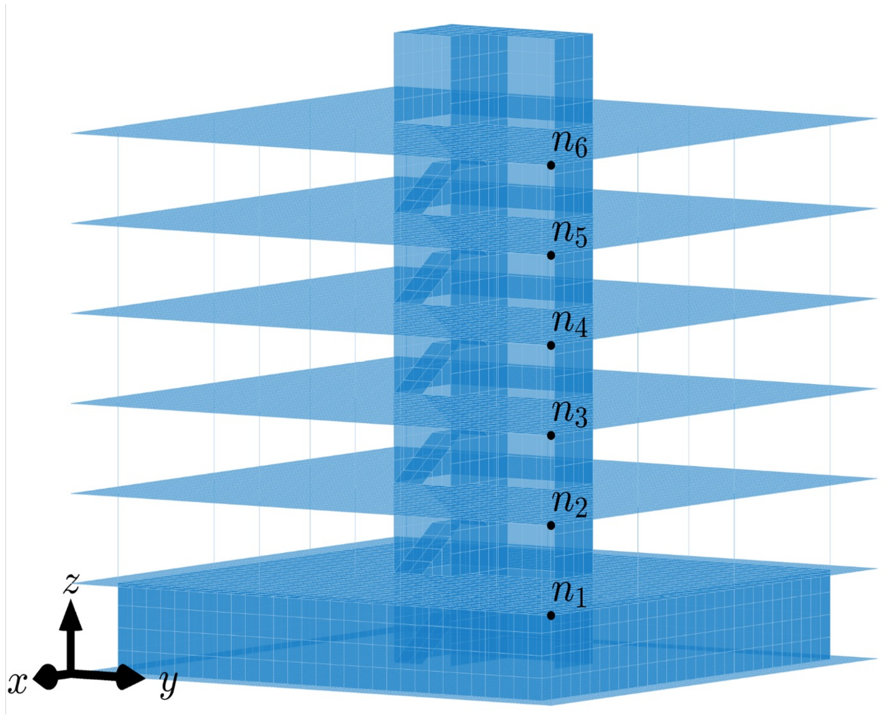
- <u>Proposed approach involves</u>:
  - **Two deterministic optimization** problems
  - **Two reliability** problems

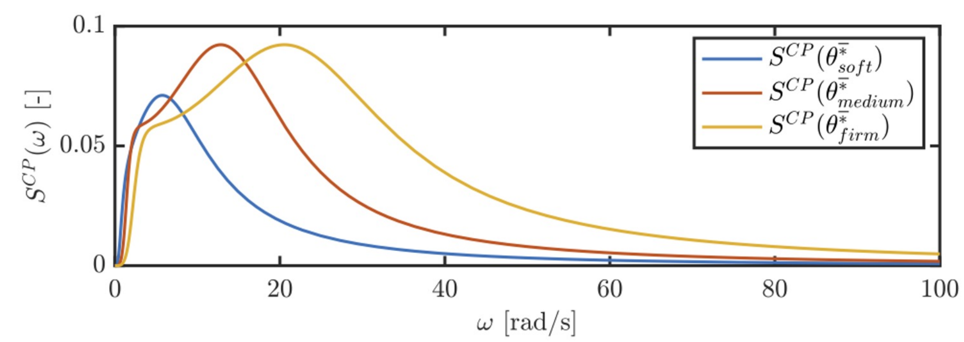
Example: a six-story building

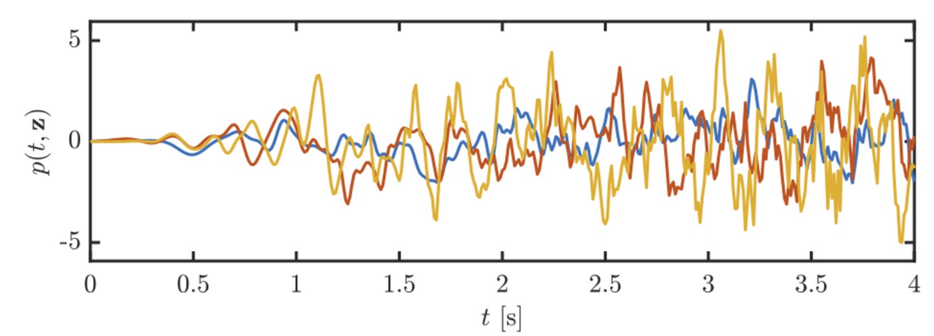
- 6 story building
  - Reinforced concrete
  - 9500 shell & beam elements
- QOI: inter-story drift
- Load: earthquake, which is modelled as stochastic process:
  - Gaussian stochastic process
  - Autocorrelation governed by modulated Clough-Penzien spectrum:

$$S_{CP}(\omega) = \frac{\omega_g^4 + (2\zeta_g\omega_g\omega)^2}{(\omega_g^2 - \omega^2)^2 + (2\zeta_g\omega_g\omega)^2} \cdot \frac{\omega^4}{(\omega_f^2 - \omega^2)^2 + (2\zeta_g\omega_g\omega)^2} \cdot S_0$$

Soil type	ω_g [rad/s]	ζ_g	ω_f [rad/s]	ζ_f
Firm	8π	0.60	0.8π	0.60
Medium	5π	0.60	0.5π	0.60
Soft	2.4π	0.85	0.24π	0.85





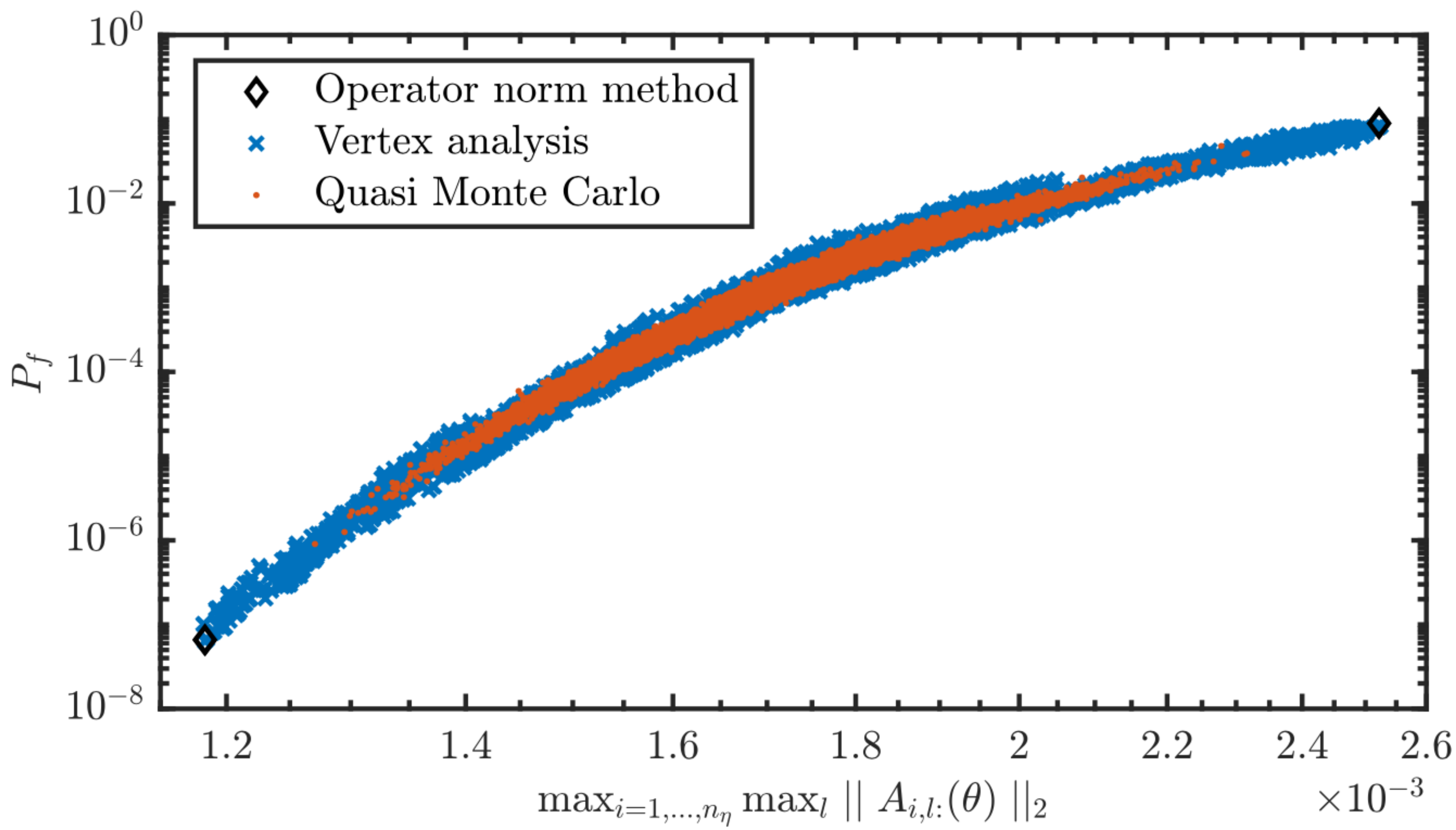


Example: results

- Optima in operator norm correspond to optima in failure probability
- Large reduction in computational cost:
  - Quasi Monte Carlo: 5.000.000 FE simulations
  - Vertex analysis: 4.096.000 FE simulations

  - Operator norm: 3500 FE simulations

Factor 1000 gain in efficiency

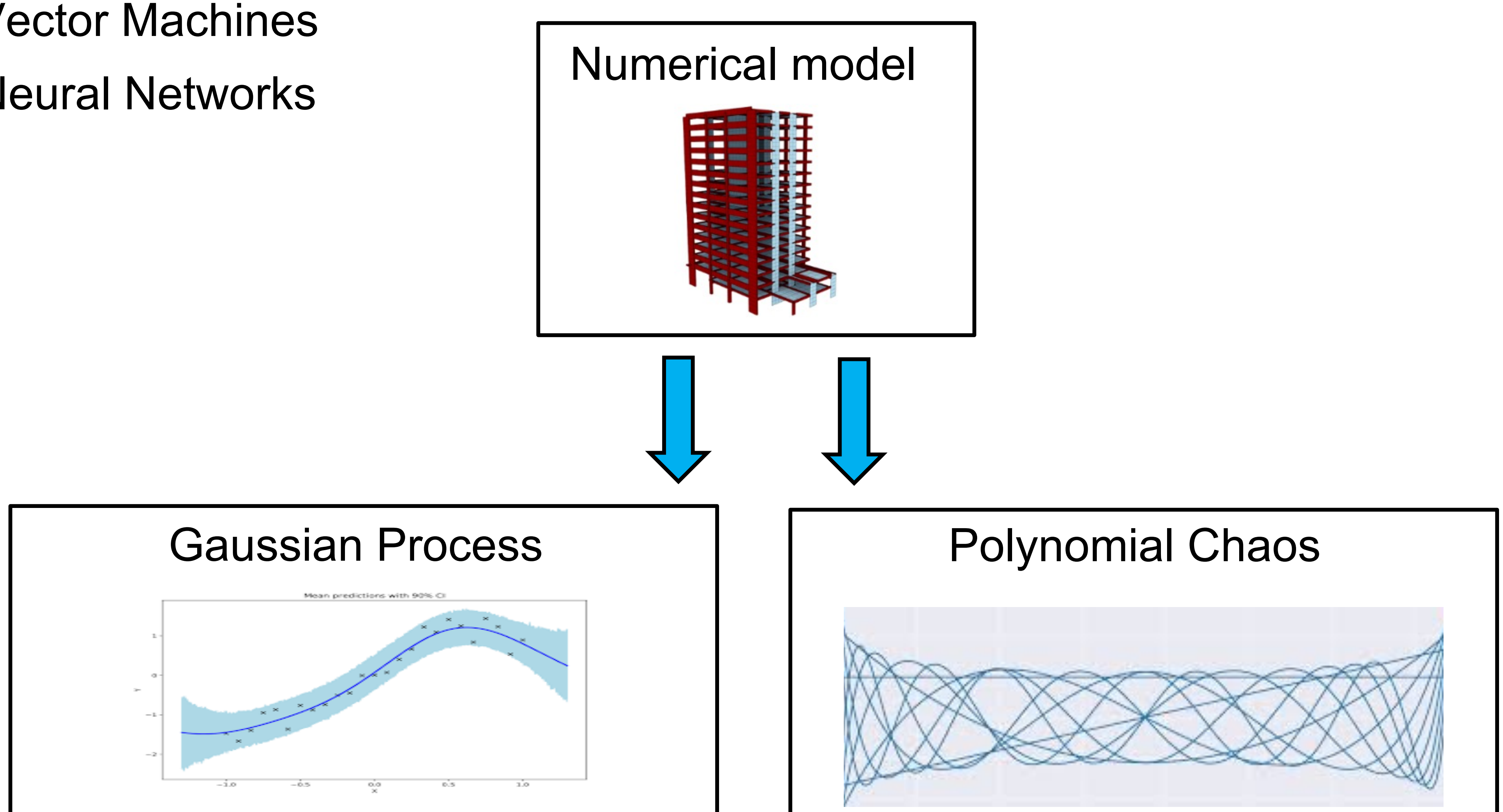


Further reading

- Faes, Matthias and Marcos A. Valdebenito. 2020. ‘Fully Decoupled Reliability-Based Design Optimization of Structural Systems Subject to Uncertain Loads’. *Computer Methods in Applied Mechanics and Engineering* 371: 113313.
- Faes, Matthias, Marcos A. Valdebenito, David Moens, and Michael Beer. 2020. ‘Bounding the First Excursion Probability of Linear Structures Subjected to Imprecise Stochastic Loading’. *Computers & Structures* 239: 106320.
- Faes, Matthias and Marcos A. Valdebenito. 2021. ‘Fully Decoupled Reliability-Based Optimization of Linear Structures Subject to Gaussian Dynamic Loading Considering Discrete Design Variables’. *Mechanical Systems and Signal Processing* 156: 107616.
- Faes, Matthias, Marcos A. Valdebenito, David Moens, and Michael Beer. 2021. ‘Operator Norm Theory as an Efficient Tool to Propagate Hybrid Uncertainties and Calculate Imprecise Probabilities’. *Mechanical Systems and Signal Processing* 152: 107482.
- Ni, Peihua et al. 2022. ‘Operator Norm-Based Statistical Linearization to Bound the First Excursion Probability of Nonlinear Structures Subjected to Imprecise Stochastic Loading’. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 8(1): 04021086.
- Fina, Marc et al. 2023. ‘Bounding Imprecise Failure Probabilities in Structural Mechanics Based on Maximum Standard Deviation’. *Structural Safety* 101: 102293.
- Ni, Peihua, et al. 2023. ‘Probability of failure of nonlinear oscillators with fractional derivative elements subject to imprecise Gaussian loads‘. *Publication in preparation*
(Contact me for a copy of the draft)

Surrogate Modeling

- Propagation of imprecise probabilities demands repeated simulations
- Numerical cost may grow rapidly
- Solution: replace model $\mathcal{M}$ with a surrogate model that is cheaper to evaluate
  - Polynomial chaos expansions
  - Gaussian processes
  - Support Vector Machines
  - Artificial Neural Networks
  - …



Surrogate modelling approaches

- **Coping** with <u>aleatoric</u> and <u>epistemic</u> uncertainty: **huge challenge!**

Epistemic

Aleatoric

Deterministic model

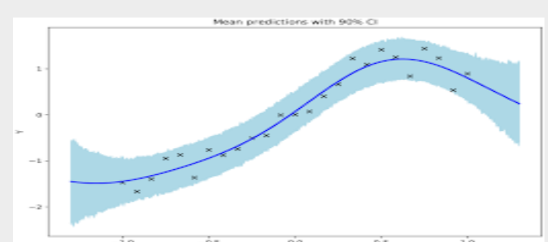
$\sim$ [min] – [hour]

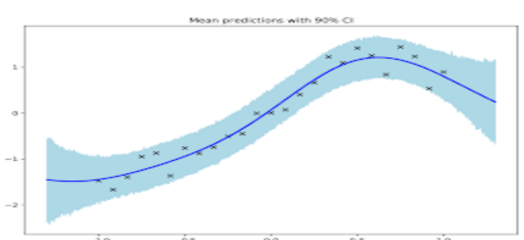
- what we would like: a cheaper numerical model to evaluate

Epistemic

Aleatoric

Surrogate model

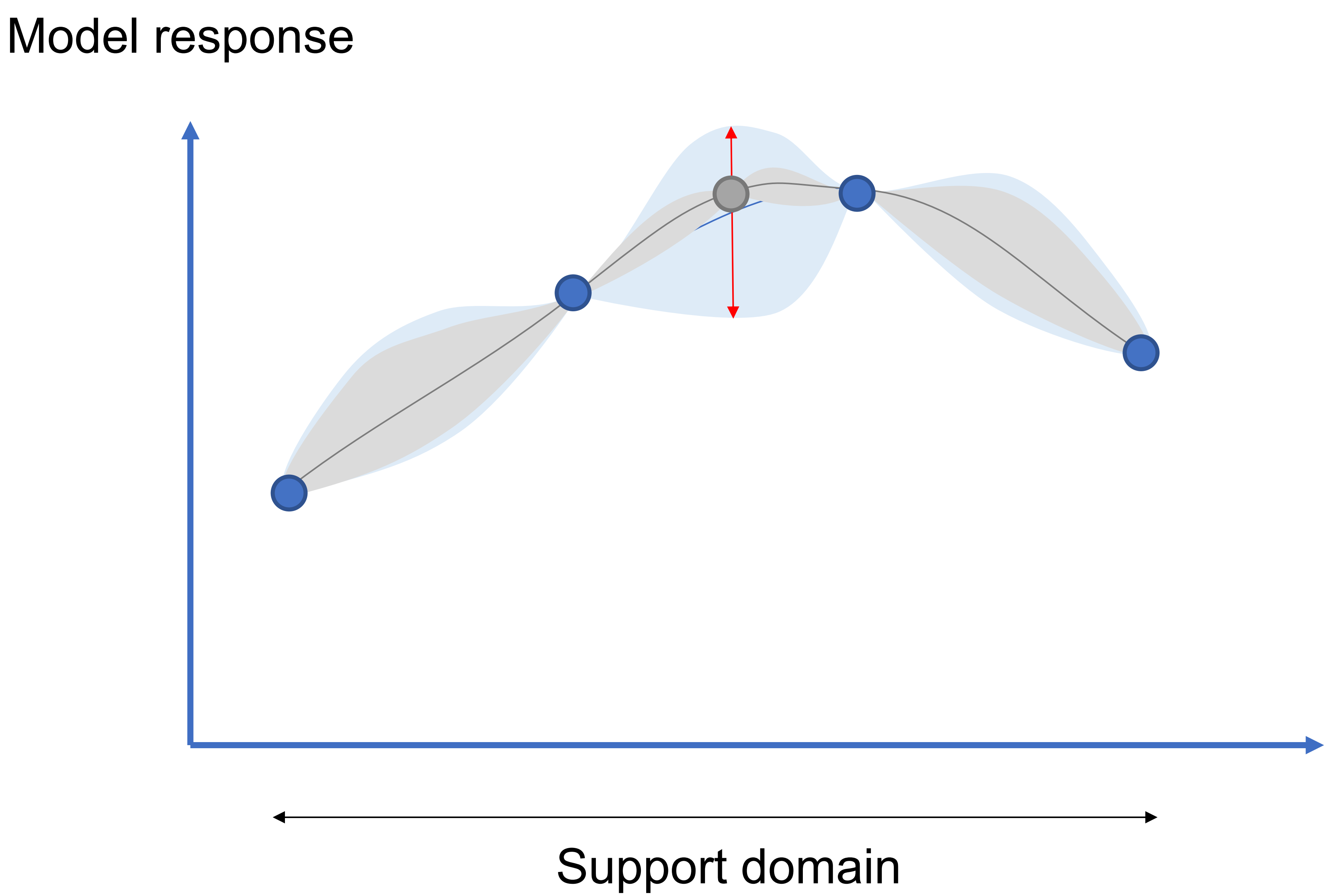




$\sim$ [sec]

Surrogate Modeling: Adaptive Gaussian Process Regression

- Key ingredient for training a surrogate is active learning



(1) Design of experiments / full system analyses

(2) Interpolation & confidence bounds

(3) Active learning / locate critical point

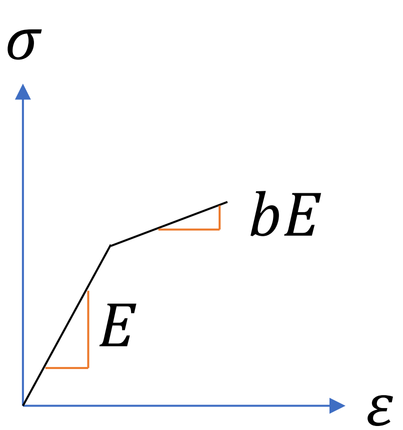
(4) Additional system analysis

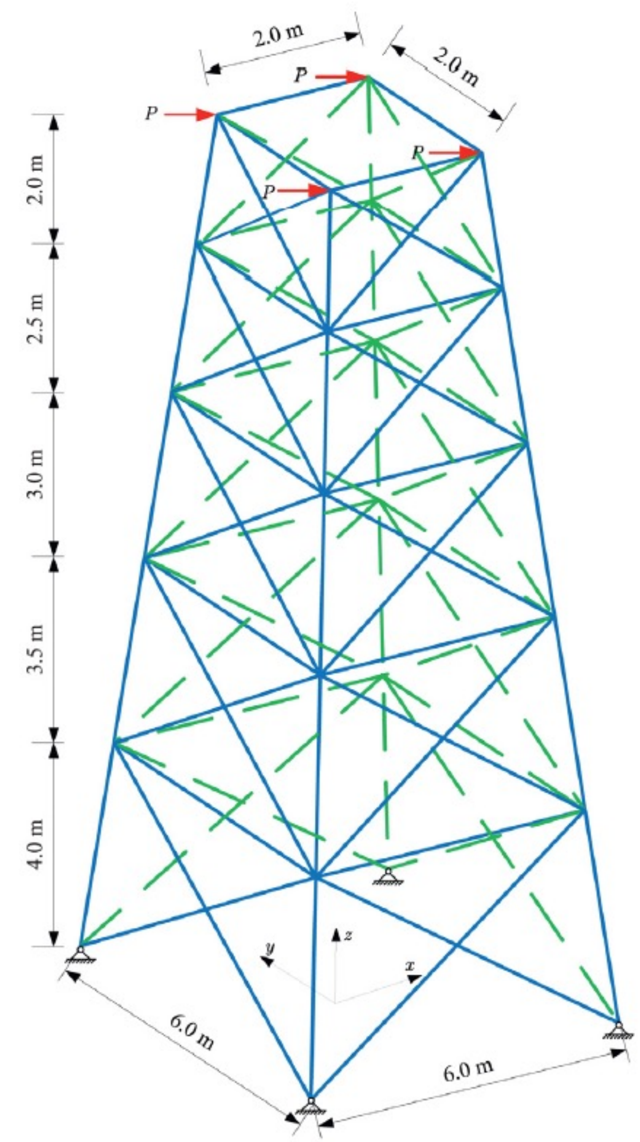
(5) Improve surrogate

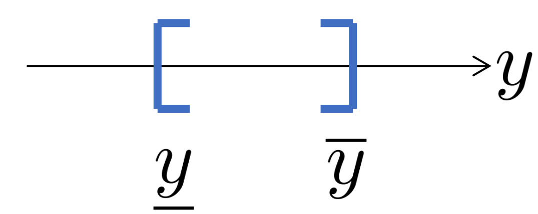
Surrogate Modeling: Adaptive Gaussian Process Regression

- Test example

Variable	Description	Interval	Unit
P	Wind load	[100,200]	kN
θ	Angle between the load direction and the x-axis	$[-45, 45]$	°
F_y	Yield strength of steel	[300,400]	MPa
E	Young's modulus of steel	$[1.8, 2.4] \times 10^5$	MPa
b	Strain hardening ratio	[0.015,0.025]	-
A_1	Cross-sectional area of the column members	[4000,5000]	mm^2
A_2	Cross-sectional area of the diagonal members	[3000,4000]	mm^2
A_3	Cross-sectional area of the horizontal members	[2000,3000]	mm^2



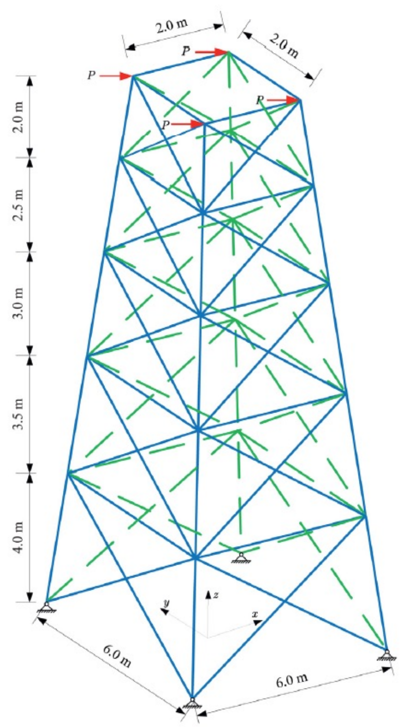


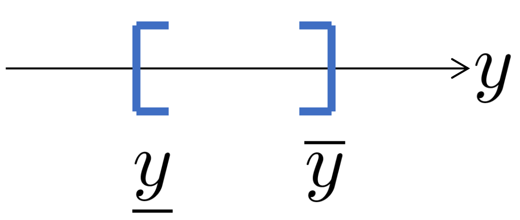


Input parameters (intervals) → Numerical model $\boldsymbol{y} = \mathcal{M}(\boldsymbol{\theta})$ → Response (interval)

Surrogate Modeling: Adaptive Gaussian Process Regression

- Test example





Response (interval)

$$\underline{y} = \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left(\mathcal{M}(\boldsymbol{\theta}) \right)$$

$$\overline{y} = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left(\mathcal{M}(\boldsymbol{\theta}) \right)$$

Method		Lower bound/mm	Upper bound/mm	N
Particle swarm optimization ($q = 10$)		11.9592	57.2421	5760
Proposed method (T-PBGO)	$q = 10$	11.9760	57.2388	70

Two orders of magnitude

Further reading

- C. Dang, P. Wei, M. G. R. Faes, M. A. Valdebenito, and M. Beer, ‘Interval uncertainty propagation by a parallel Bayesian global optimization method’, *Applied Mathematical Modelling*, vol. 108, pp. 220–235, Aug. 2022, doi: 10.1016/j.apm.2022.03.031.
- C. Dang, P. Wei, M. G. R. Faes, M. A. Valdebenito, and M. Beer, ‘Parallel adaptive Bayesian quadrature for rare event estimation’, *Reliability Engineering & System Safety*, vol. 225, p. 108621, Sep. 2022, doi: 10.1016/j.ress.2022.108621.
- C. Dang, P. Wei, M. G. R. Faes, and M. Beer, ‘Bayesian probabilistic propagation of hybrid uncertainties: Estimation of response expectation function, its variable importance and bounds’, *Computers & Structures*, vol. 270, p. 106860, Oct. 2022, doi: 10.1016/j.compstruc.2022.106860.
- C. Dang, M. A. Valdebenito, M. G. R. Faes, P. Wei, and M. Beer, ‘Structural reliability analysis: A Bayesian perspective’, *Structural Safety*, vol. 99, p. 102259, Nov. 2022, doi: 10.1016/j.strusafe.2022.102259.
- van Mierlo, C., Persoons, A., Faes, M., Moens, D. (2023). Robust design optimization of expensive stochastic simulators under lack-of-knowledge. ASCE/ASME Journal for Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering. vol. 9(2). p. 021205. 10.1115/1.4056950
- Van Mierlo, C., Persoons, A., Faes, M., Moens, D. (2023). Robust design optimisation under lack-of-knowledge uncertainty. *Computers & Structures. Volume 275,* 106910 10.1016/j.compstruc.2022.106910

Conclusions

Conclusions

- Traditional probabilistic approaches may not be sufficient to tackle uncertainty in engineering design in a consistent way
- Non-traditional approaches have shown their merit
  - Extreme scarce or non-probabilistic problems: interval analysis
  - Incomplete data on probabilistic descriptors: p-boxes
- Many efficient and flexible propagation schemes have been introduced to tackle these problems

- Further reading:
  - General overview on imprecise probabilities: Beer, M., Ferson, S., & Kreinovich, V. (2013). Imprecise probabilities in engineering analyses. *Mechanical Systems and Signal Processing*, *37*(1–2), 4–29. https://doi.org/http://dx.doi.org/10.1016/j.ymssp.2013.01.024
  - Review paper on interval and fuzzy analysis: Faes, M., & Moens, D. (2020). Recent Trends in the Modeling and Quantification of Non-probabilistic Uncertainty. *Archives of Computational Methods in Engineering*, *27*(3), 633–671. https://doi.org/10.1007/s11831-019-09327-x
  - Overview of computational methods for p-box analysis: Faes, M., Daub, M., Marelli, S., Patelli, E., Beer, M. (2021). Engineering analysis with probability boxes: a review on computational methods. *Structural Safety, 93,* 102092. https://doi.org/10.1016/j.strusafe.2021.102092

Acknowledgement & big thank you

- None of this work would have been possible without the collaborations with many mentors, colleagues, PhD students and friends over the last years

- David Moens
- Michael Beer
- Dirk Vandepitte
- Pol Spanos
- Kok-Kwak Phoon
- Enrico Zio
- Marcos Valdebenito
- Matteo Broggi
- Edoardo Patelli
- Mona Dannert
- Rodolfo Fleury
- Marc Fina
- Koen van Mierlo
- Lars Bogaerts
- Robin Callens
- Konstantinos Ypsilantis
- Jasper Cerneels
- Maurice Imholz
- Marco Daub
- Stefano Marelli
- Xiukai Yuan
- Pengfei Wei
- Jingwen Song
- Chao Dang
- Sifeng Bi
- Vasilis Fragkoulis
- Peihua Ni
- Danko Jerez
- Jonathan Sadeghi
- Mauricio Misraji
- Nataly Manque

and of course many others :)

Prof. Dr. Matthias Faes

- Born in Lier, Belgium 10.1991
- M.Sc. In Manufacturing Engineering Technology in 2013
- Ph.D. in Uncertainty Quantification from KU Leuven in 2017
- Winner of the ECCOMAS Award for the Best PhD Thesis of 2017 on Computational Methods in Applied Sciences and Engineering
- FWO postdoc at KU Leuven
- Alexander von Humboldt Fellow at Leibniz University Hannover
- Professor in Reliability Engineering at TU Dortmund since 02.2022

Contact details: Chair for Reliability Engineering

Prof. Dr. Matthias Faes
Chair for Reliability Engineering
TU Dortmund
Leonhard-Euler-Straße 5
44227 Dortmund

matthias.faes@tu-dortmund.de

lre.mb.tu-dortmund.de

We are always happy to host guests from other groups to set-up and maintain joint collaborations.