Imprecise Subset simulation for reliability analysis

Dimitris Giovanis
Assistant Research Professor

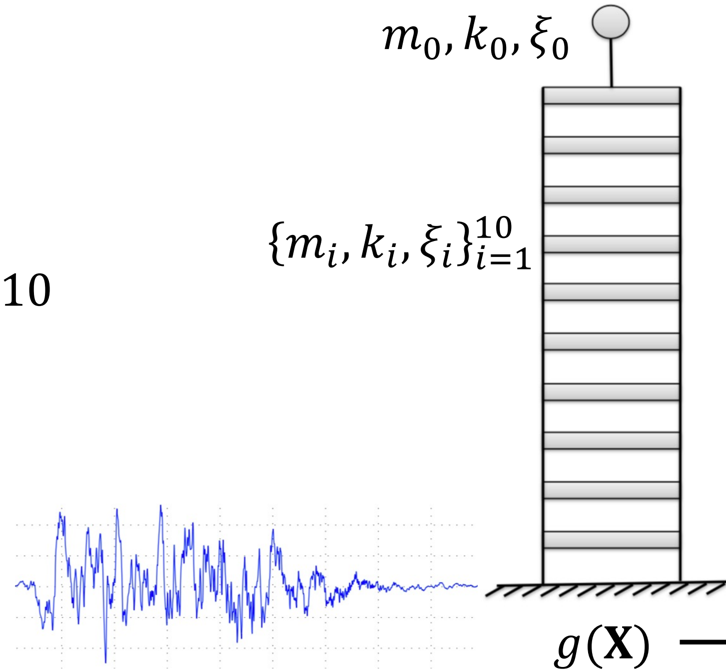
Michael D. Shields
Associate Professor

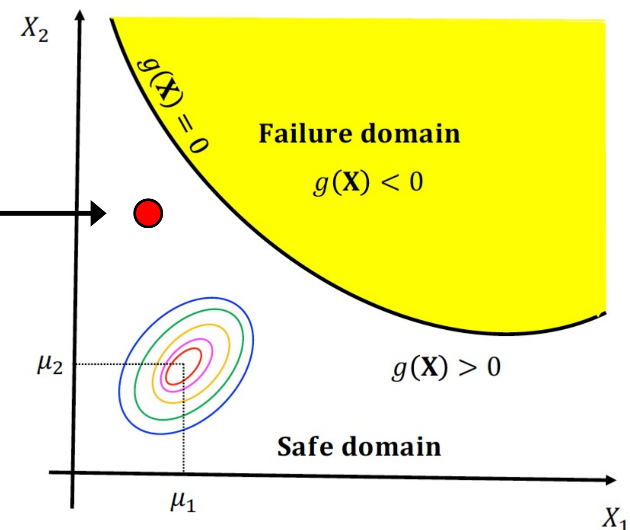
Dept. Of Civil and Systems Engineering
Johns Hopkins University

May 17-20, 2021

Effect of small dataset

$$\mathbf{X} = \begin{Bmatrix} m_i \\ k_i \\ \xi_i \end{Bmatrix}, i = 0, \dots, 10$$

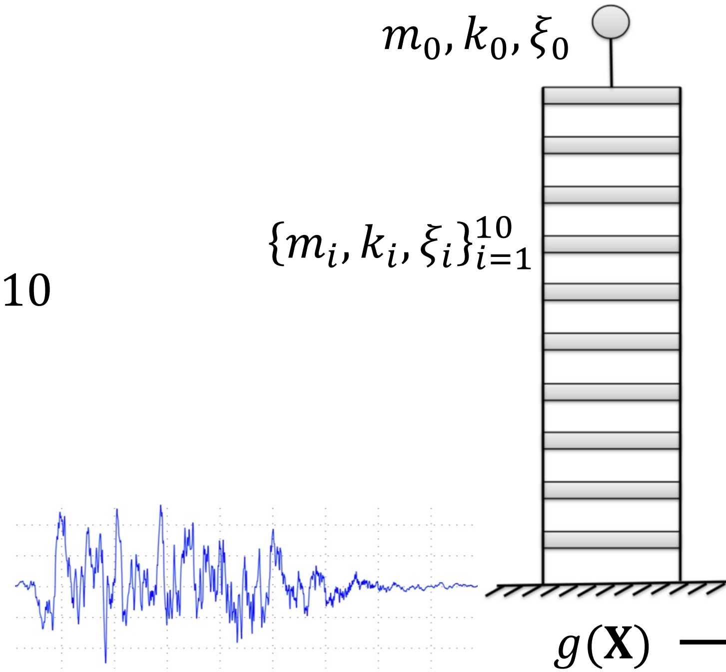




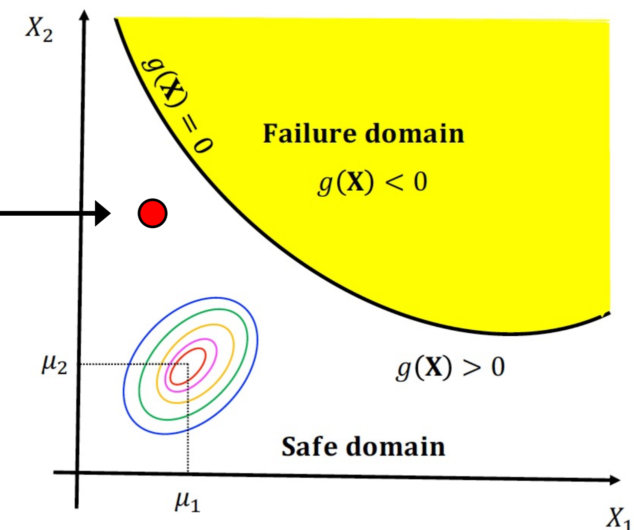
$$P_f = \mathbb{P}(g(\mathbf{X}) \leq 0) = \int_{\{\mathbf{X}: g(\mathbf{X}) \leq 0\}} p(\boldsymbol{x}) d\boldsymbol{x}$$

Effect of small dataset

$$\mathbf{X} = \begin{Bmatrix} m_i \\ k_i \\ \xi_i \end{Bmatrix}, i = 0, \dots, 10$$



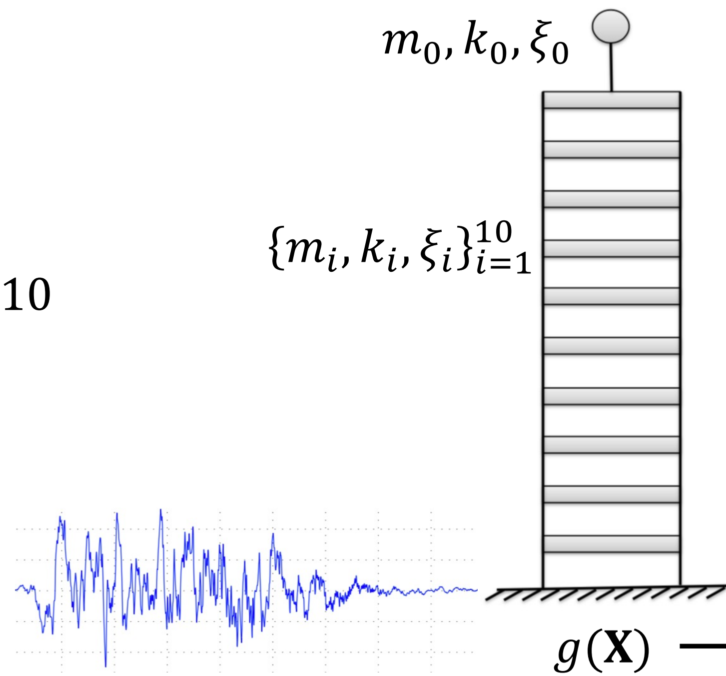
$g(\mathbf{X})$ →

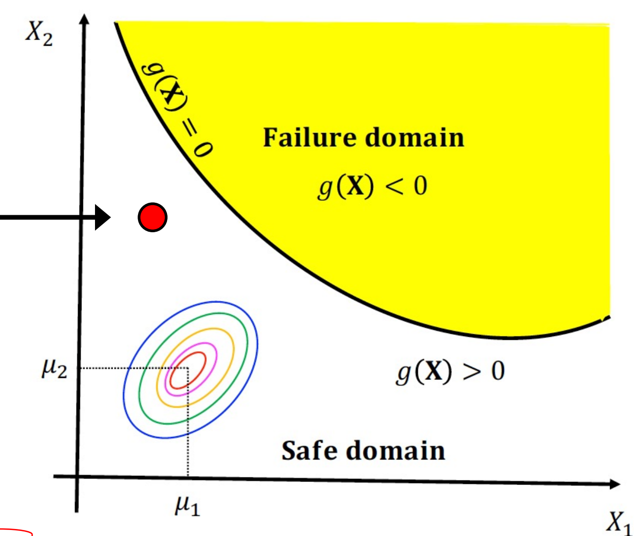


$$P_f = \mathbb{P}(g(\mathbf{X}) \leq 0) = \int_{\{\mathbf{X}: g(\mathbf{X}) \leq 0\}} \underbrace{p(\boldsymbol{x})}_{\text{Joint distribution function}} d\boldsymbol{x}$$

Effect of small dataset

$$\mathbf{X} = \begin{Bmatrix} m_i \\ k_i \\ \xi_i \end{Bmatrix}, i = 0, \ldots, 10$$





$$P_f = \mathbb{P}(g(\mathbf{X}) \leq 0) = \int_{\{\mathbf{X}: g(\mathbf{X}) \leq 0\}} p(\boldsymbol{x}) d\boldsymbol{x}$$

What is this distribution?
Where does it come from?

DATA $\mathcal{D}$

Observations

Obvious

We rarely have enough data

Observations

Obvious
We rarely have enough data

Consequence
Without sufficient data, we cannot assert probability distributions for reliability analysis.

Observations

<u>Obvious</u>
Where does it come from?

<u>Consequence</u>
Without sufficient data, we cannot assert probability distributions for reliability analysis.

<u>Questions</u>
Can we trust computational estimates of reliability?

If not, how can we build estimators that we do trust?... Or that we can at least quantify our confidence in?

Multimodel inference

Where do uncertainties in probability model arise?

- **Model-form uncertainty:** We don't know the form of M (e.g., *Gaussian*, *lognormal*, *Beta*). Bayes' rule can be used to compare multiple models M_j:

$$\hat{\pi}_j \equiv p\left(M_j \middle| \mathcal{D}\right) = \frac{p\left(\mathcal{D}|M_j\right) \cdot \pi_j}{\sum_{k=1}^{m} p(\mathcal{D}|M_k) \cdot \pi_k}$$

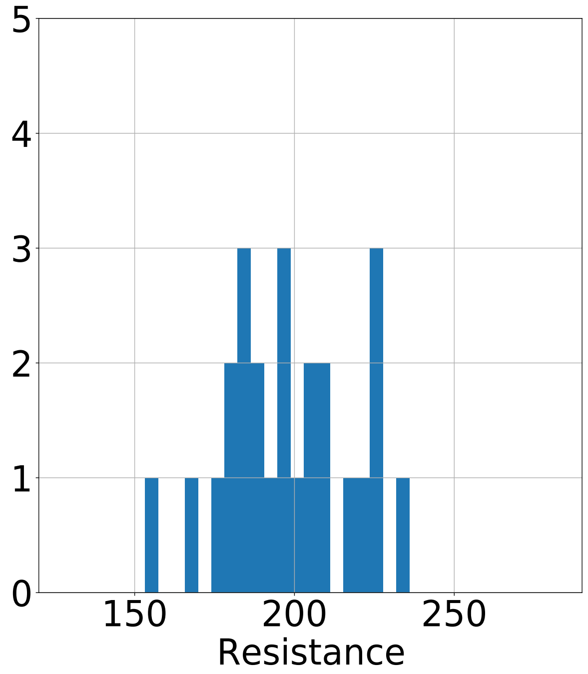
where $p\left(\mathcal{D}|M_j\right) = \int p\left(\mathcal{D}|M_j, \boldsymbol{\theta}_j\right) p\left(\boldsymbol{\theta}_j|M_j\right) \mathrm{d}\boldsymbol{\theta}_j$ is the model evidence.

- **Parameter uncertainty:** Even if we know the form of the M, we still need to estimates its parameters $\boldsymbol{\theta}$, given the data $\mathcal{D}$, i.e., $M\left(\boldsymbol{\theta}|\mathcal{D}\right)$

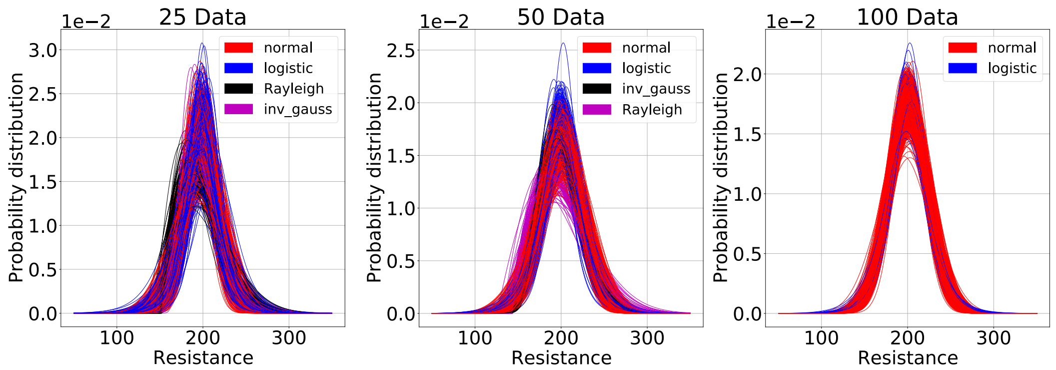
$$p\left(\boldsymbol{\theta}_j \middle| \mathcal{D}, M_j\right) = \frac{p\left(\mathcal{D}|\boldsymbol{\theta}_j, M_j\right) p\left(\boldsymbol{\theta}_j|M_j\right)}{p\left(\mathcal{D}|M_j\right)} \propto \mathcal{L}\left(\boldsymbol{\theta}_j|\mathcal{D}\right) p\left(\boldsymbol{\theta}_j|M_j\right)$$

This is performed using Markov Chain Monte Carlo.

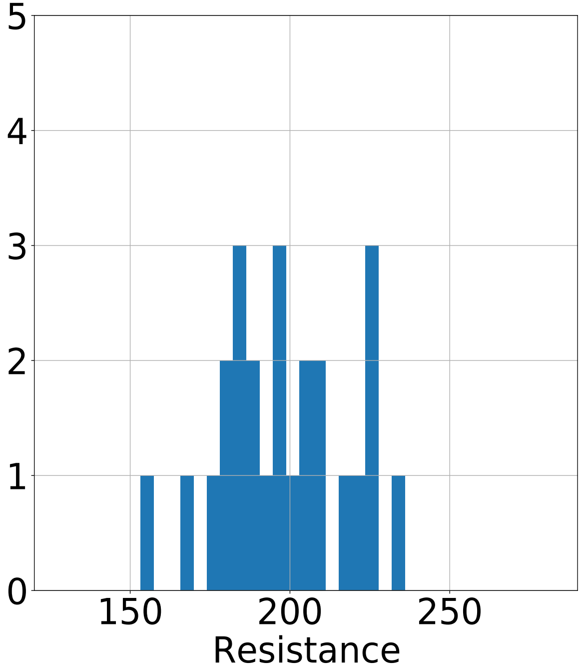
Multimodel UQ



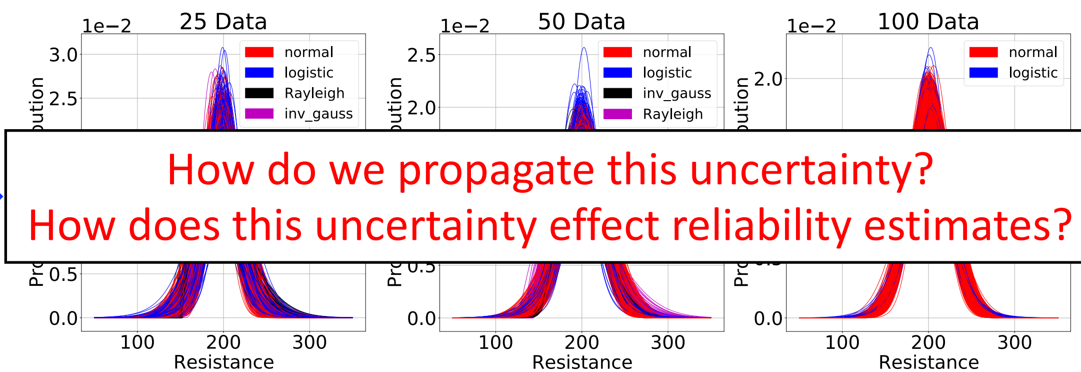
Candidate probability models for varying dataset size



Multi-model UQ



Candidate probability models for varying dataset size



How do we propagate this uncertainty?

How does this uncertainty effect reliability estimates?

Subset simulation – Key Ingredients

<u>Subset Simulation</u>

Subdivision of the failure event $g(\mathbf{X}) \leq 0$ to m sub-events (subsets)

$$g_i = \{\mathbf{X}: g(\mathbf{X}) \leq b_i\}, i = 1, \dots, m$$

Where $b_1 > b_2 > \cdots > b_m$

The probability of failure is given by:

$$P_f = \mathbb{P}(g_m) = P_1 \cdot \prod_{i=2}^{m} P_i,$$

where

$$P_1 = \mathbb{P}(g_1), \quad P_i = \mathbb{P}(g_i | g_{i-1})$$

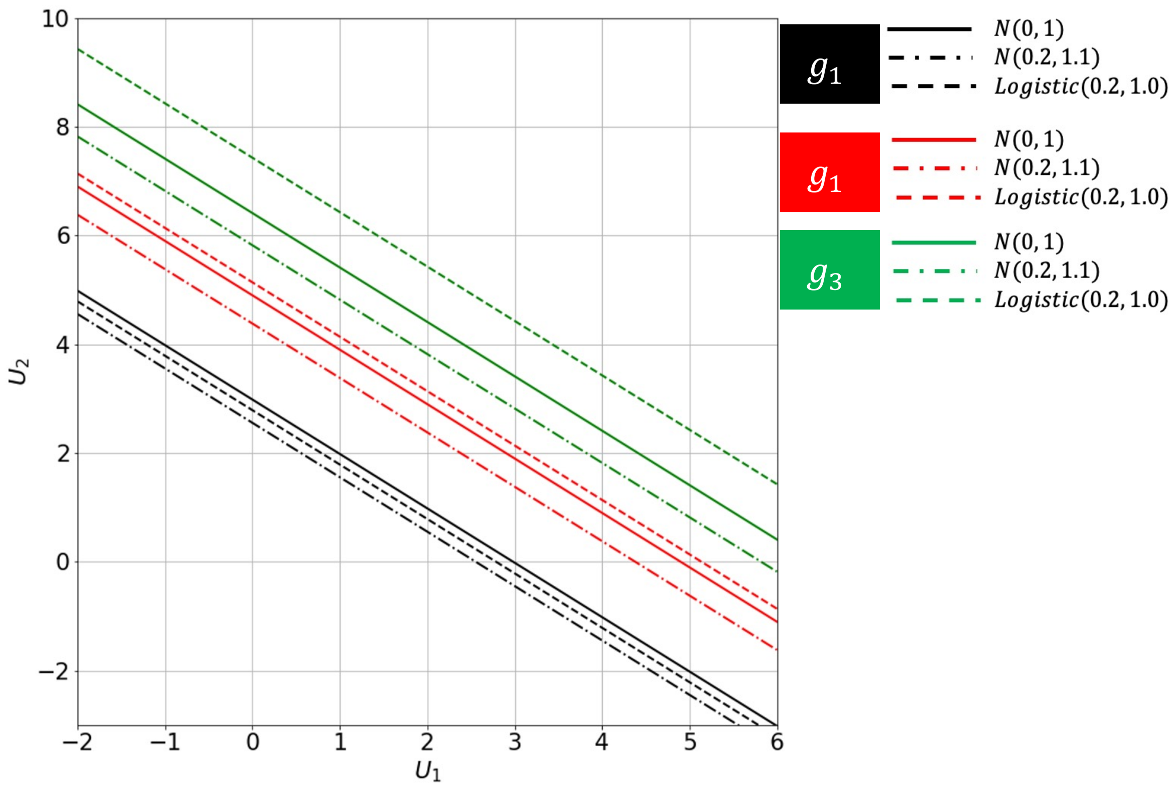
Motivation

Simple structural model: $$g(\mathbf{U}) = \beta\sqrt{2} - \sum_{j=1}^{2} U_j$$

$M_1 \sim N(0,1)$

$M_2 \sim N(0.2, 1.1)$

$M_3 \sim Logistic(0.2, 1)$



Imprecise Subset simulation

Bayesian Model Selection
$\{(M_r, p_r)\}, r = 1, \ldots, n$

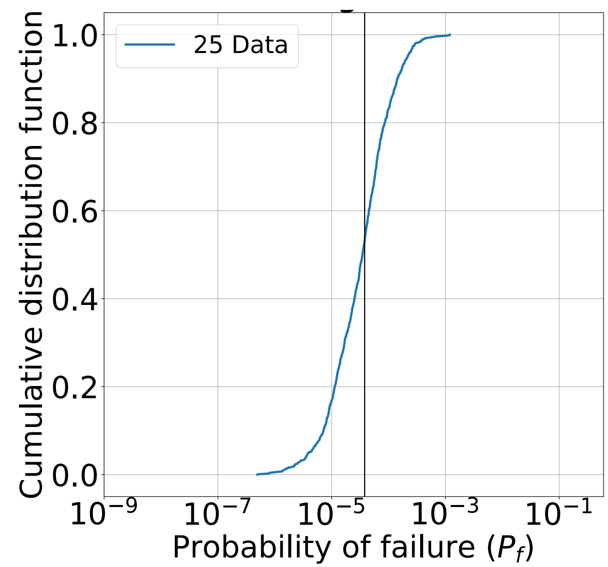
Bayesian parameter estimation $\{\boldsymbol{\theta}_r^{(j)}\}, j = 1, \ldots, n_c$

Select N models
$p_t(\cdot), t = 1, \ldots, N$

Subset run #1

Subset run #2

Subset run #N



Weighted Imprecise Subset simulation

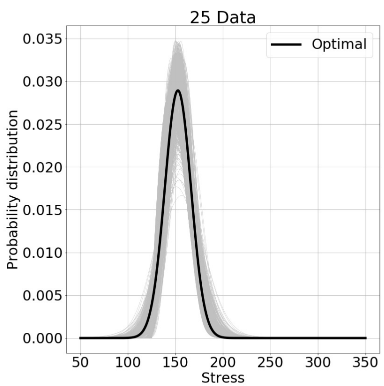
Bayesian Model Selection
$\{(M_r, p_r)\}, r = 1, \ldots, n$

Bayesian parameter estimation $\{\boldsymbol{\theta}_r^{(j)}\}, j = 1, \ldots, n_c$

Select N models
$p_t(\cdot), t = 1, \ldots, N$

Identify optimal marginal density
$q_j^*(x)$

$$q_j^*(x) = \sum_{i=1}^{n} \pi_t \mathbb{E}_\theta[p_t(x|\theta)]$$



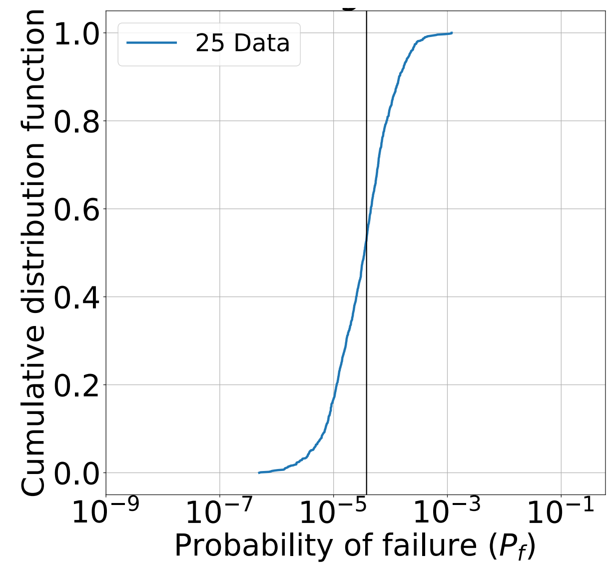
Run one subset simulation with target pdf
$q_d^*(x) = \prod_{j=1}^{d} q_j^*(x)$

Calculate the IS weights

$$w_t(x_i^{opt}) = \frac{p_t(x_i^{opt})}{q_d^*(x_i^{opt})}$$

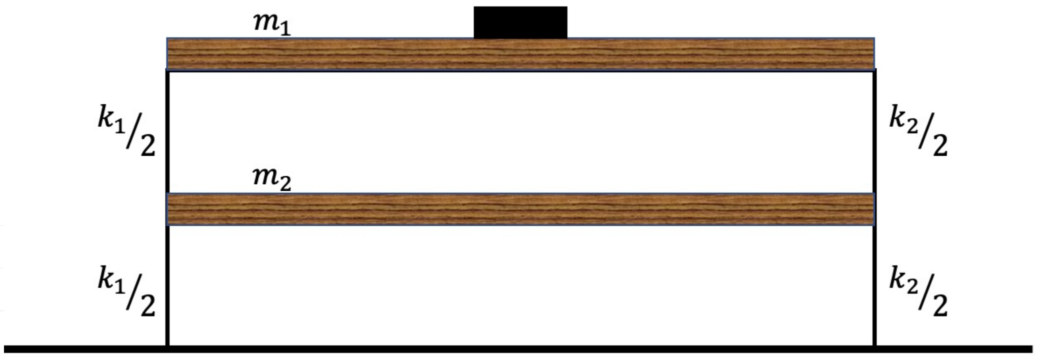
Weight the probabilities

$$P_i^t = \sum w_t(x_i^{opt}) I_{g_i^{opt}}(x_l^{opt}) \Rightarrow P_f = \prod P_i^t$$



Numerical example

Two-story linear structure subjected to ground acceleration

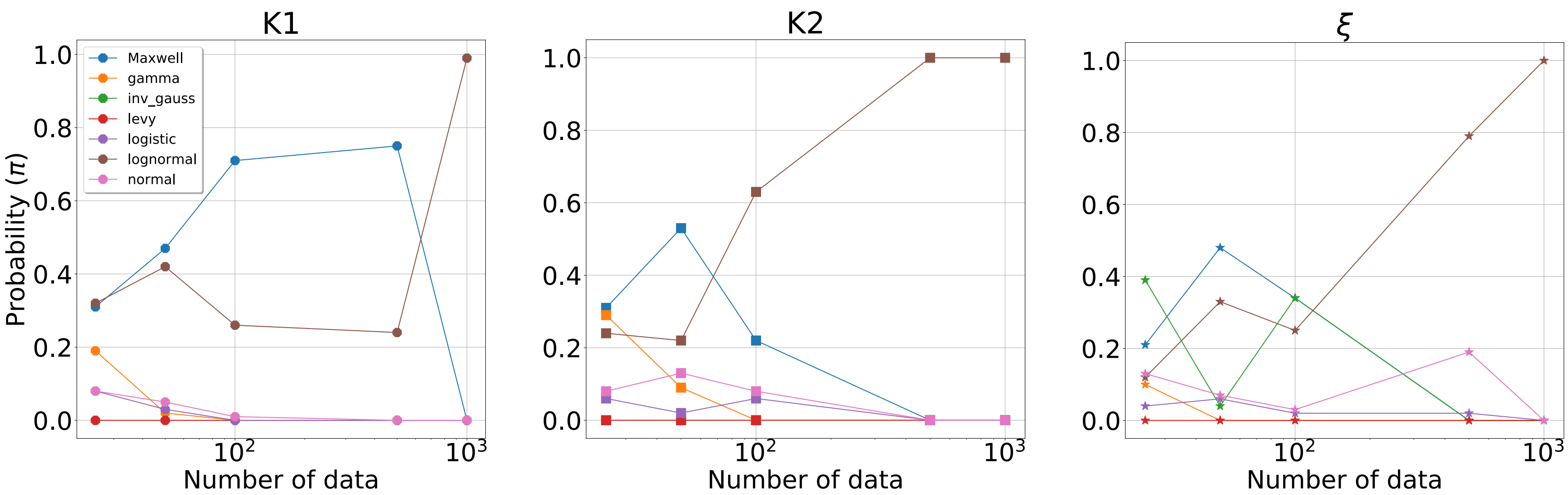


Sources of uncertainty

Variable	Mean value	c.o.v	Description
k_1	3000	0.2	Stiffness parameter
k_2	3000	0.2	Stiffness parameter
ξ	0.03	0.15	Damping ratio

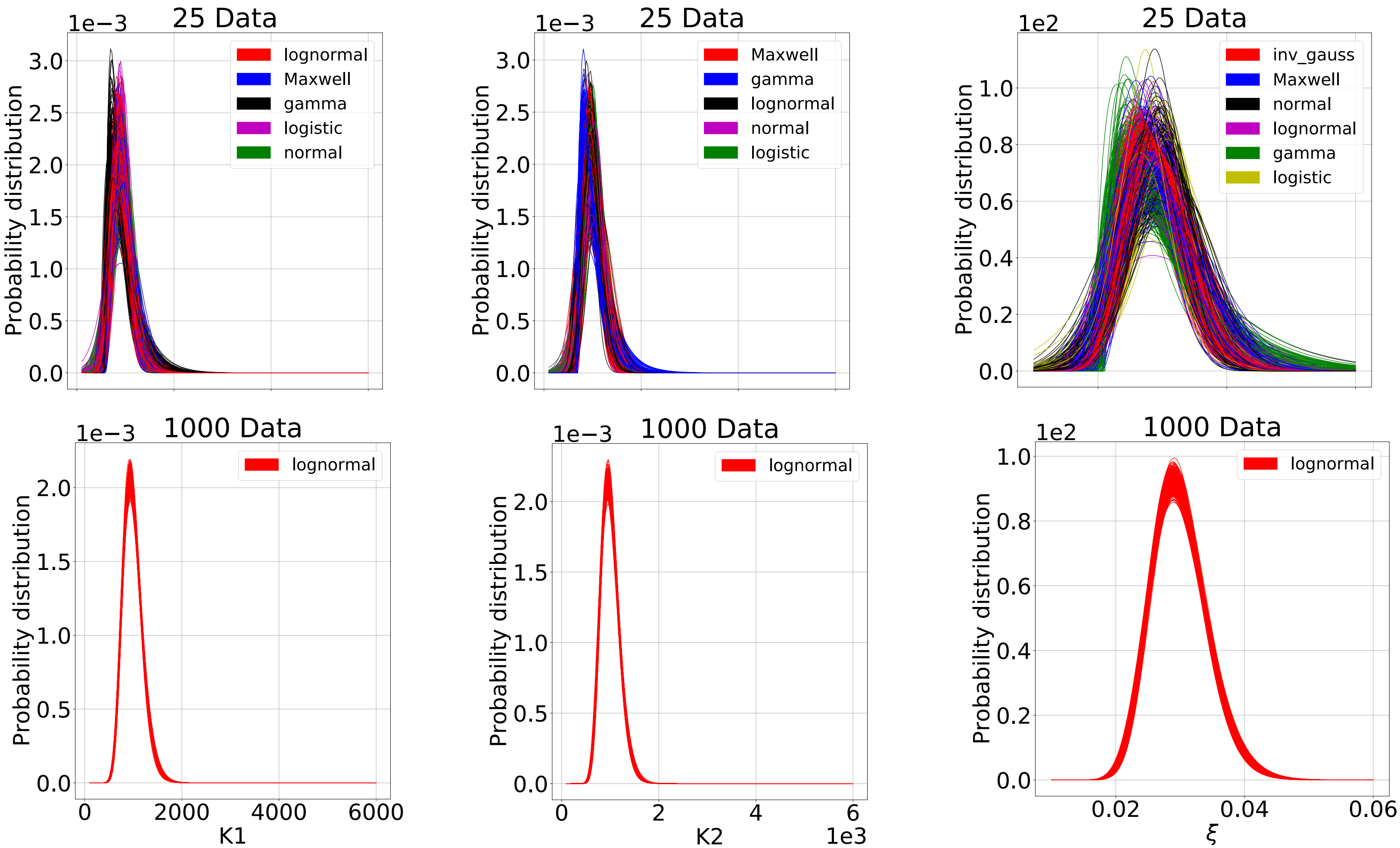
Numerical example

Probability of each model for increasing dataset size



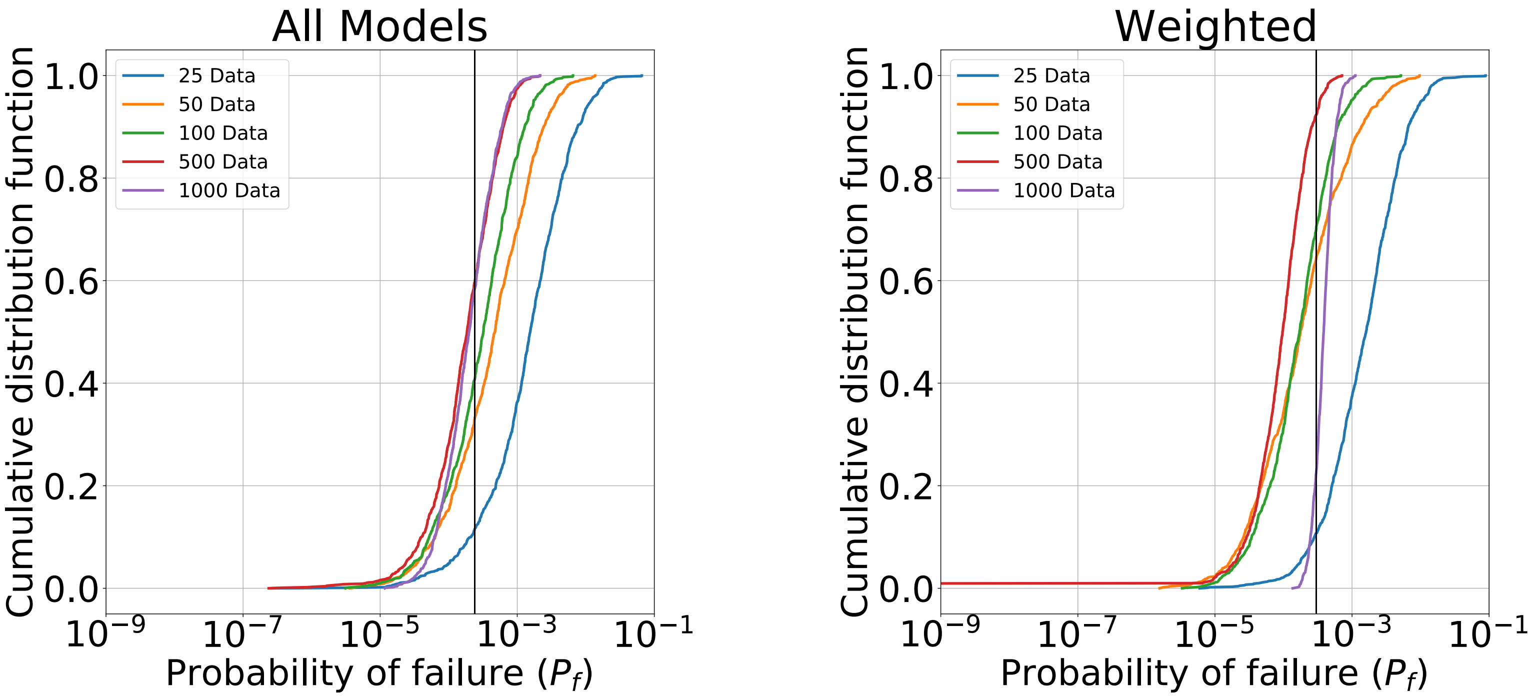
Numerical example

Sample sets of probability distributions for 25 and 1000 data, for k_1, k_2, ξ



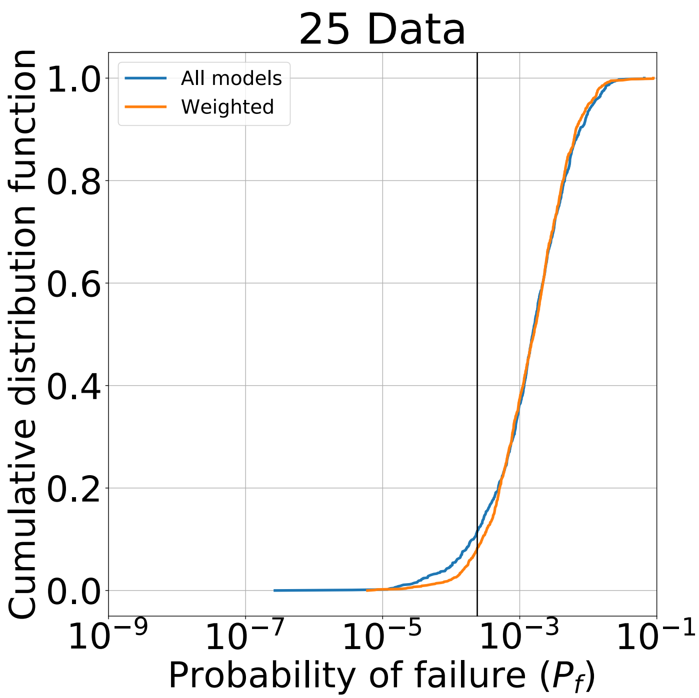
Numerical example

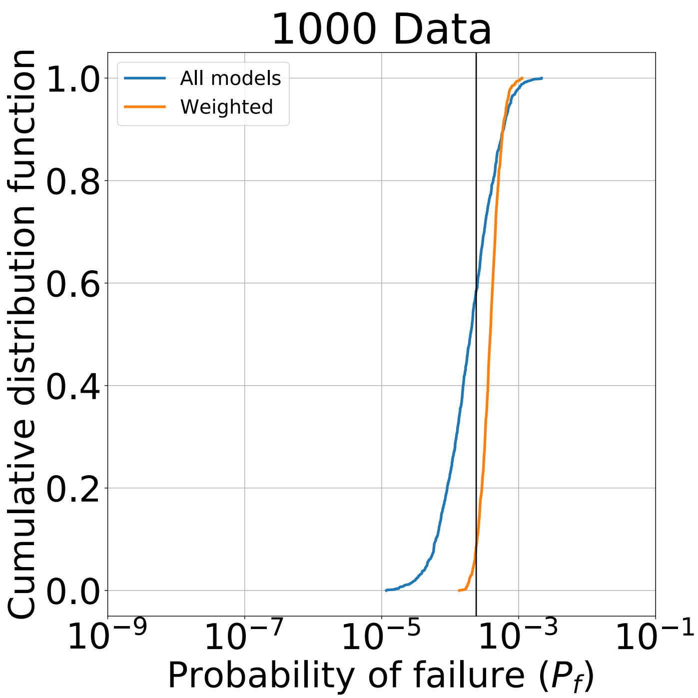
Empirical cdfs of the probability of failure P_f for different dataset sizes. The true $P_f = 2.4\times10^{-4}$ (obtained by means of MCS) is indicated by a vertical black line.



Numerical example

One-to-one comparison of the two approaches for 25 and 1000 data





Conclusions

- Investigate the quality of structural reliability estimates using Subset Simulation, when data for quantifying input uncertainty are scant.
- When the probability model is uncertain and represented by multiple models, the location of each conditional level is uncertain.
- An approach that leverages a Bayesian-Information Theoretic multimodel inference process for estimating distribution form and parameter uncertainty.
- An efficient estimator of reliability uncertainty from a single SuS simulation.

Conclusions

Thank you for your attention!