Dealing with Vague and Limited Information in Uncertainty Quantification

Michael Beer, Matteo Broggi
Institute for Risk and Reliability

Leibniz Universität Hannover

Ioannis Kougioumtzoglou
Department of Civil Engineering and Engineering Mechanics

Columbia University in the City of New York

Edoardo Patelli
Institute for Risk and Uncertainty

University of Liverpool

Thanks to:
Liam Comerford, Marco de Angelis, Zhang Mingqiang, …

GENERAL SITUATION

Endeavor

- numerical modeling – physical phenomena, structure, and environment

  ➡ prognosis – system behavior, hazards, safety, risk, robustness, economic and social impact, ...

» close to reality
» numerically efficient

Deterministic methods

- deterministic structural parameters
- deterministic computational models

⬌ Reality

Uncertainty ?

Imprecision ?

Vagueness ?

Fluctuations ?

Ambiguity ?

Variability ?

Indeterminacy ?

➡ Consequences ?

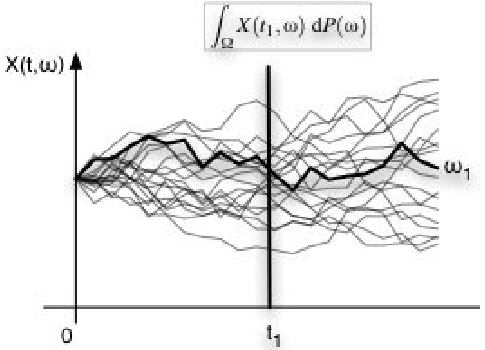
MULTI-DISCIPLINARY CHALLENGE

Major directions of development … and combinations thereof

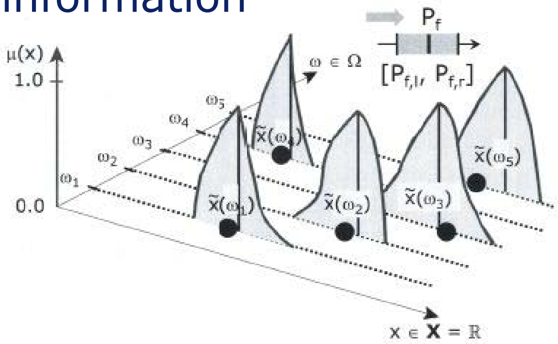
Advanced
Stochastic Modeling

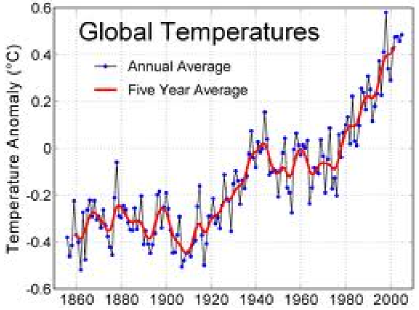
$$Z = \int Dx\, e^{iS[x]/\hbar}$$

Monte Carlo
Simulation methods



Generalized models
for vague and imprecise information

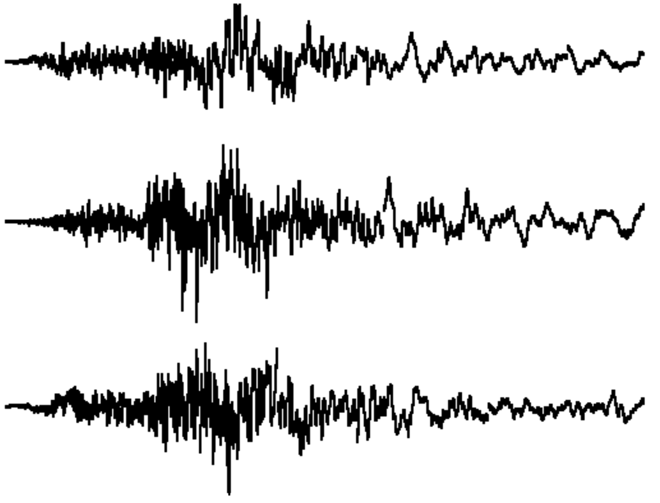




JOINT TIME-FREQUENCY ANALYSIS

Civil infrastructure

» seismic motions
» ocean waves
» winds
» blast events
» storms
» hurricanes



Extreme Events

JOINT TIME-FREQUENCY ANALYSIS

Excitation → System → Response

Excitation: non-stationary stochastic process

(earthquake, wind, ocean waves, blast events etc)

System: nonlinear and time-varying behavior due to severe dynamic excitation

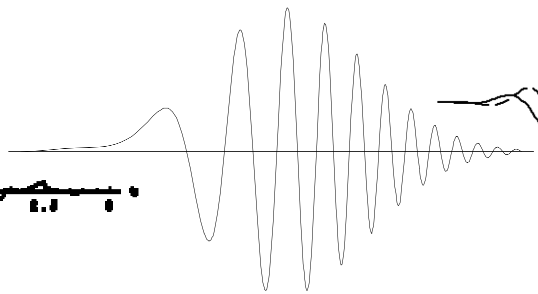
Response: non-stationary stochastic process

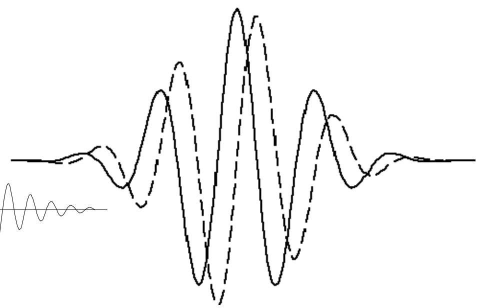
- excitations with time-varying intensity & frequency content
  → joint time-frequency analysis

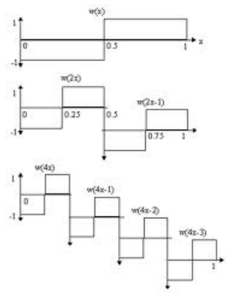
Wavelets

class of functions with compact support in both time and frequency domains used for the localization of a function in these domains

Alfred Haar (1909)





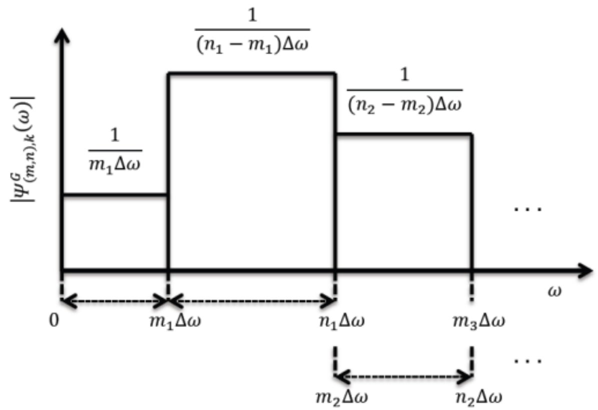


JOINT TIME-FREQUENCY ANALYSIS

Evolutionary Power Spectrum (EPS) estimation

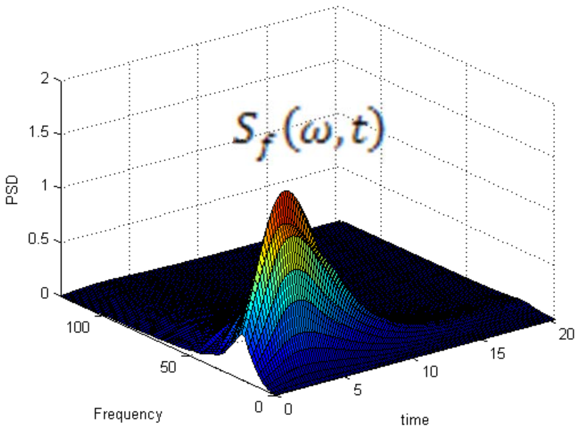
- mathematical model for random process representation
  - » representation of random properties in correspondence with the EPS
  - » locally stationary wavelet process

$$f(t) = \sum_j \sum_k w_{j,k} \psi_{j,k}(t) \xi_{j,k}$$



- Generalized Harmonic Wavelets
  - » orthogonality & additional parameter

➡ convenient frequency windowing
  - » flexible window size
  - » non-overlapping supports



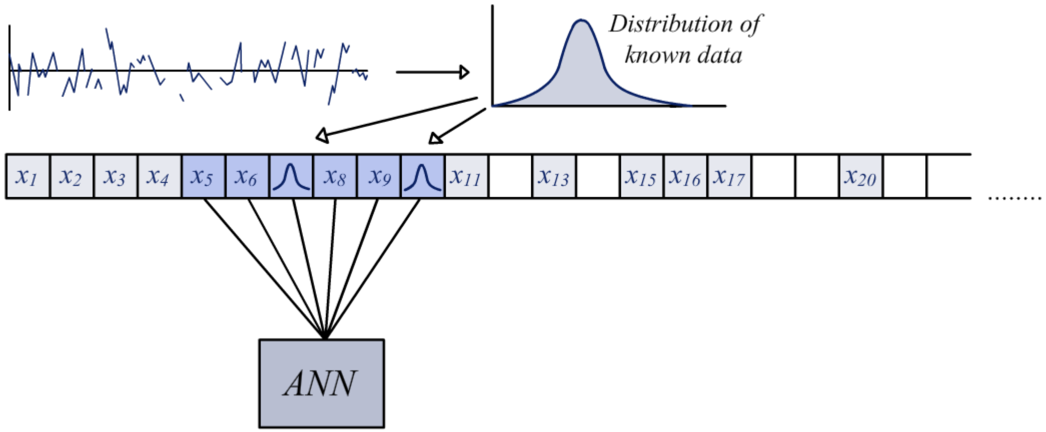
➡ distribution of “power” or “energy” of the underlying non-stationary stochastic process over different frequencies

basis for most engineering load models

JOINT TIME-FREQUENCY ANALYSIS

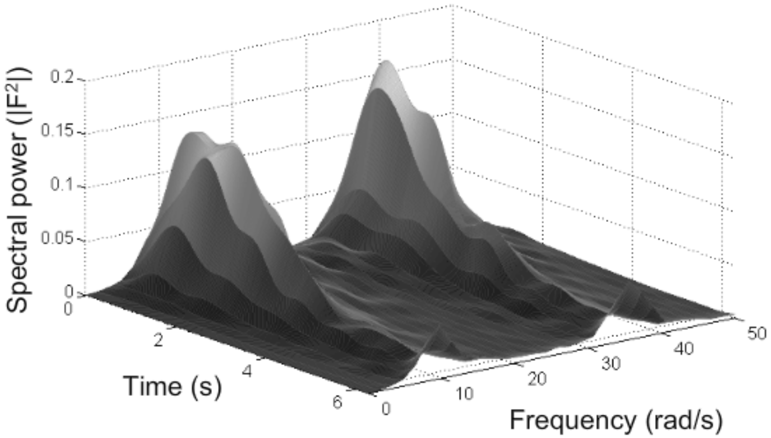
Incomplete data

- process records with gaps
  - » sensor failures
  - » data corruption
  - » power outages
  - » ...
- artificial intelligence technology



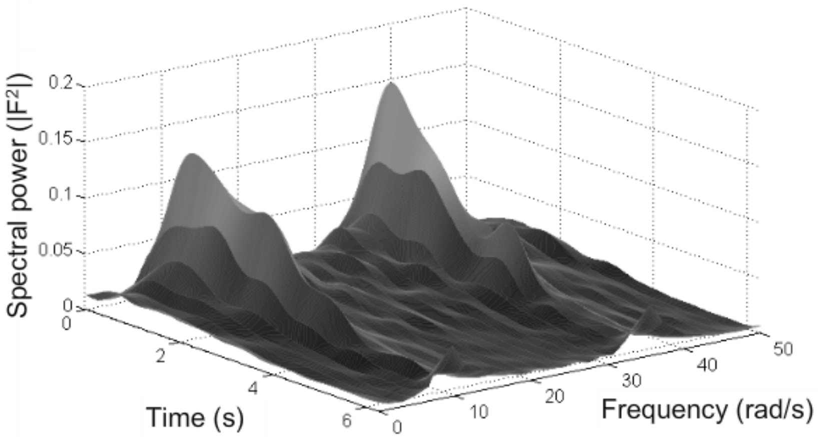
Harmonic wavelet estimated spectrum

25 records of length 512 - **no missing data**



Estimated spectrum of ANN reconstructions

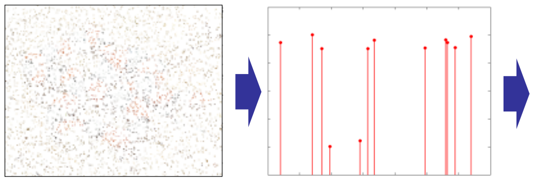
25 records of length 512 - **50% missing data in random locations**



JOINT TIME-FREQUENCY ANALYSIS

Incomplete data

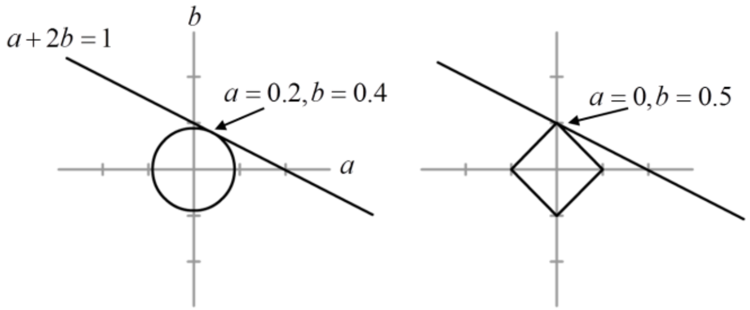
- Compressive sensing approach



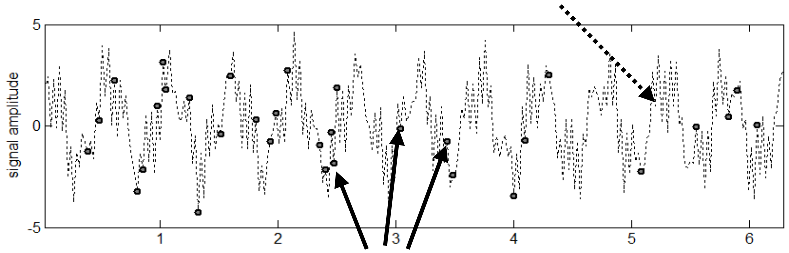
Assume sparsity in a known basis

- Minimization of $\Sigma_i|x_i|$ in $Ax = y$ promotes sparse solutions

Least squares **L1 minimization**

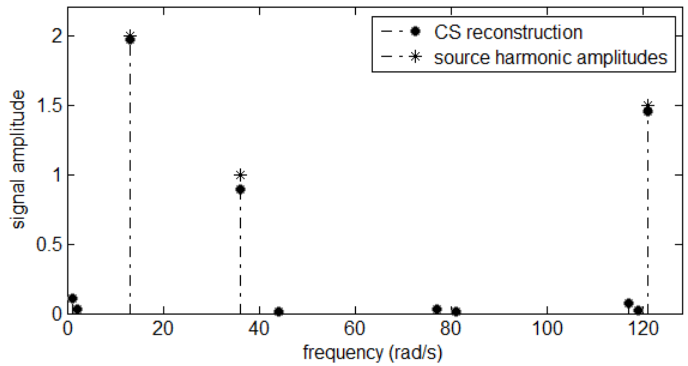


Stochastic process with 3 harmonics and white noise



Limited number of sample points available

Assume sparsity in harmonic basis to identify key frequencies

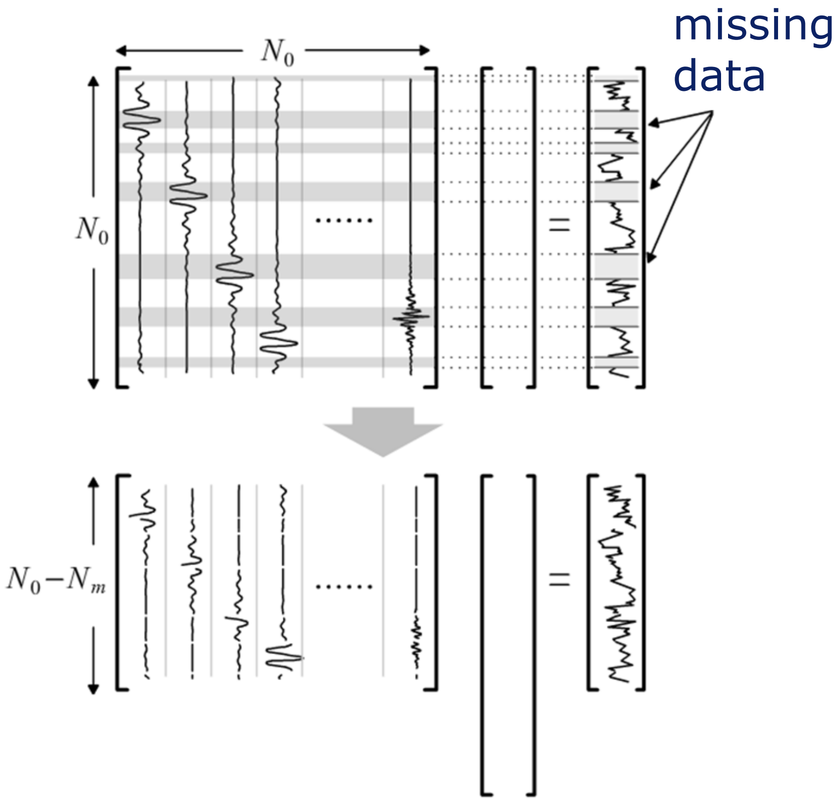


JOINT TIME-FREQUENCY ANALYSIS

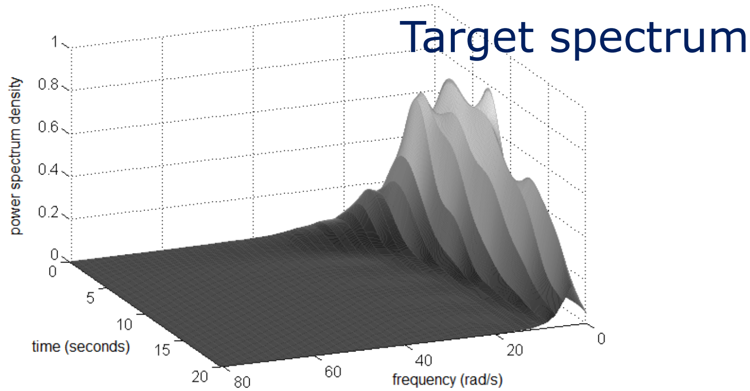
Incomplete data

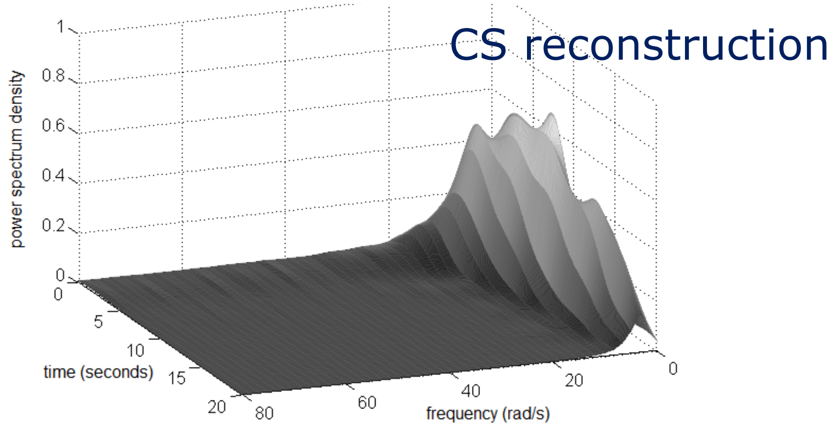
- Compressive sensing approach
  - » Harmonic basis construction

Wavelet basis (example)



Example for non-separable, non-stationary process: (50% randomly distributed missing data over 25 records using CS with a harmonic wavelet basis)

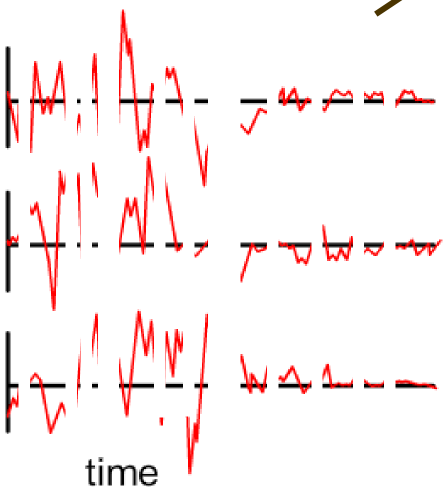




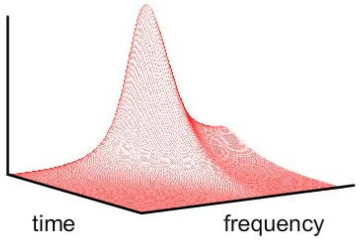
EXAMPLE: RELIABILITY ANALYSIS

Incomplete Earthquake Records

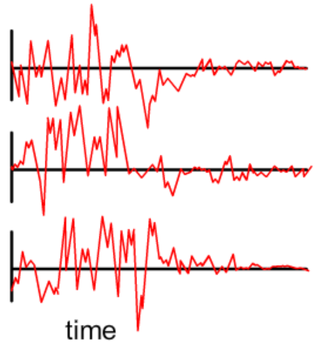
Incomplete
non-stationary
earthquake
records



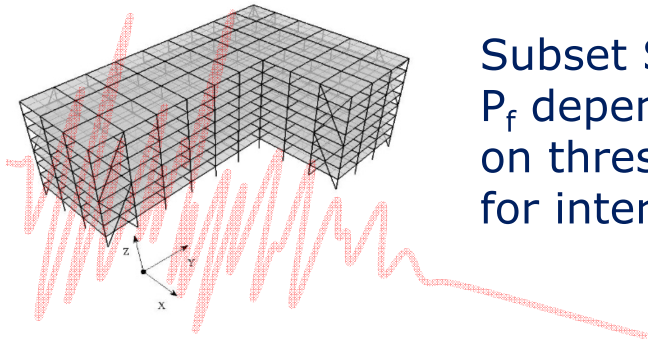
Spectrum estimation



Simulated
realizations



Reliability Analysis



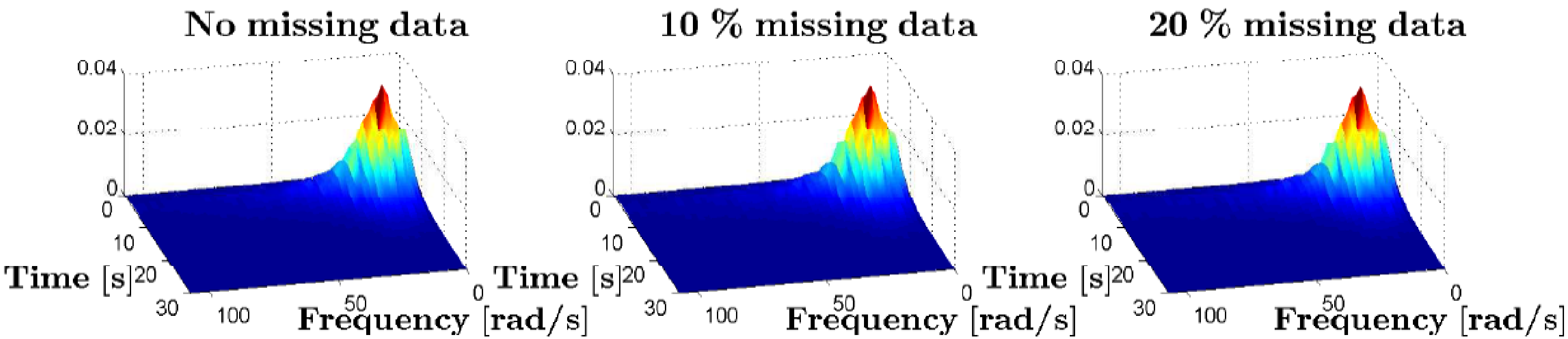
Subset Sampling;
P_f depending
on threshold
for interstory drift

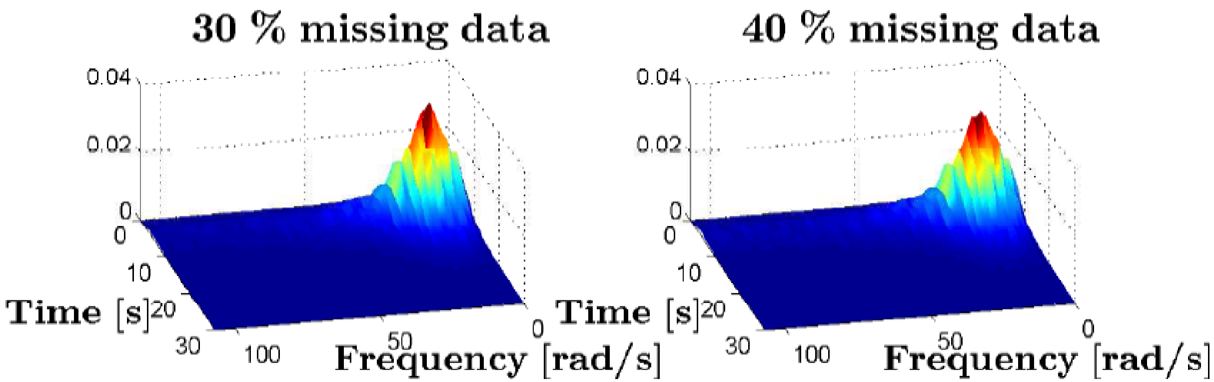
Collaboration with
Professor Hector Jensen

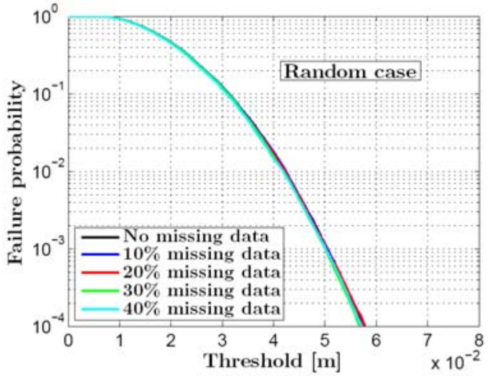
EXAMPLE: RELIABILITY ANALYSIS

Incomplete Earthquake Records

- Scenario 1: data removed at random positions (10%, 20%, 30%, 40%)





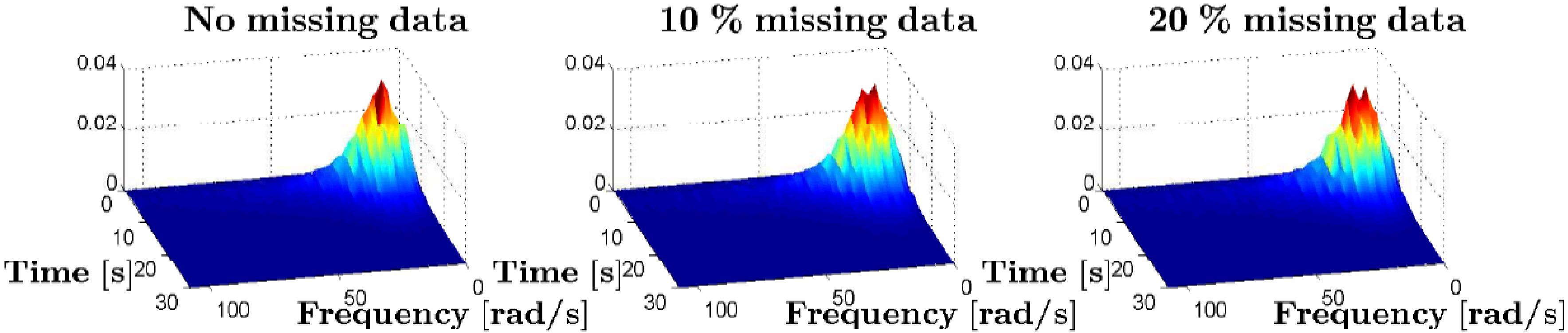


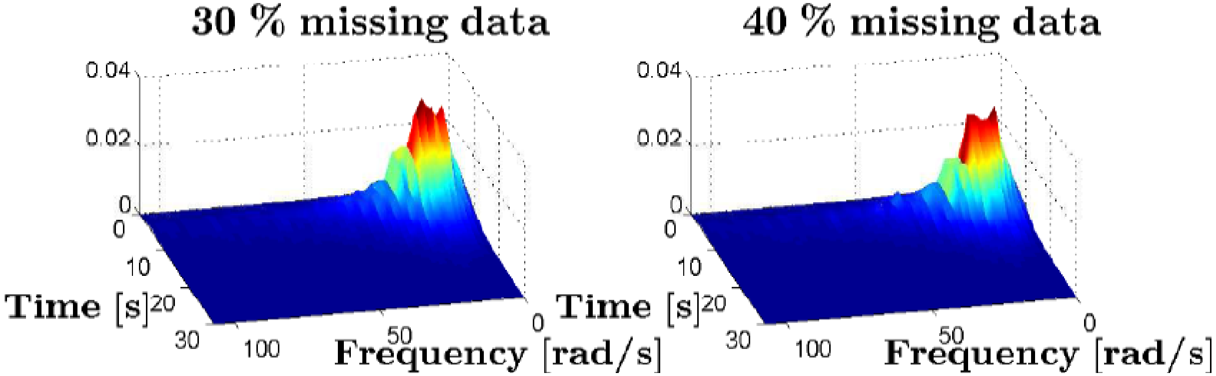
Collaboration with
Professor Hector Jensen

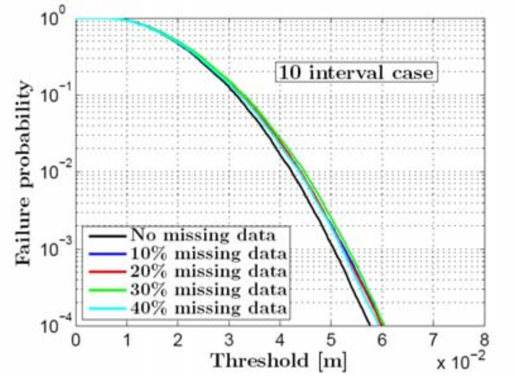
EXAMPLE: RELIABILITY ANALYSIS

Incomplete Earthquake Records

- Scenario 2: data removed at 10 intervals located at random positions (10%, 20%, 30%, 40%)





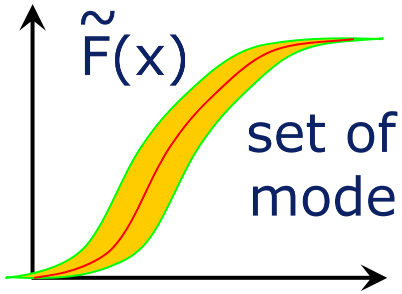


Collaboration with
Professor Hector Jensen

CHALLENGE

statistical analysis of imprecise and rare data

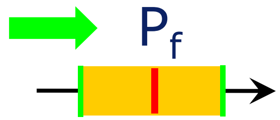
Is it safe ?



set of plausible models

reliability analysis

$\tilde{P}_f$



$[P_{f,l}, P_{f,r}]$
imprecision reflected in P_f

Is the reliability analysis still reliable ?

Effects on P_f ?

epistemic uncertainty

Sensitivity of P_f to imprecision ?

CLASSIFICATION AND MODELING

According to sources

- aleatory uncertainty
  - » irreducible uncertainty
  - » property of the system
  - » fluctuations / variability

  stochastic characteristics

  ➡ traditional probabilistic models

- epistemic uncertainty
  - » reducible uncertainty
  - » property of the analyst
  - » lack of knowledge or perception

  collection of all problematic cases, inconsistency of information

  ➡ no specific model

According to information content

- uncertainty
  - » probabilistic information

  ➡ traditional and subjective probabilistic models

- imprecision
  - » non-probabilistic characteristics

  ➡ set-theoretical models

In view of the purpose of the analysis

- averaged results, value ranges, worst case, etc. ?

CLASSIFICATION AND MODELING

Simultaneous appearance of uncertainty and imprecision

- separate treatment of uncertainty and imprecision in one model
- generalized models combining probabilistics and set theory

➡ **concepts of imprecise probabilities**

» interval probabilities
» sets of probabilities / p-box approach
» random sets
» fuzzy random variables / fuzzy probabilities
» evidence theory / Dempster-Shafer theory

common basic feature:

set of plausible probabilistic models over a range of imprecision

(set of models which agree with the observations)

➡ bounds on probabilities for events of interest

IMPLEMENTING EPISTEMIC UNCERTAINTY

Subjective probabilities

- Bayesian update

$$f_{\Theta|X_1,\ldots,X_n}\left(\theta|x_1,\ldots,x_n\right)=\frac{\left[\prod_{i=1}^{n} f\left(x_i|\theta\right)\right]\cdot g_{\Theta}\left(\theta\right)}{\int\left[\prod_{i=1}^{n} f\left(x_i|\theta\right)\right]\cdot g_{\Theta}\left(\theta\right)d\theta}$$

prior distribution

posterior distribution

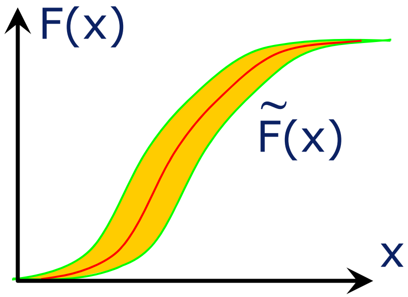
» approach result from "inside the epistemic uncertainty"

Imprecise probabilities

- set-theoretical models for indeterminacy / imprecision (imprecise data, vague conditions or copulas etc.)

$$F(x)=\left\{\left[F_l(x_i),F_u(x_i)\right]\ \forall x_i\right\}$$

e.g., set-valued parameters in probabilistic models



» approach result from "outside the epistemic uncertainty"

➡ choice depending on available information and purpose of the analysis

➡ approaches not competing but complementary and can be combined (e.g. set of priors, update with imprecise data)

FUZZY PROBABILITIES

Model

- fuzzy random variables $\tilde{X}: \Omega \rightarrow \mathbb{F}(\mathbb{R})$ (frequentist view)

  ➡ fuzzy set of real-valued random variables $\tilde{X} = \left\{ \left(X_j,\ \mu\left(X_j\right) \right) \ \middle|\ x_{ji} \in \tilde{x}_i \ \forall i \right\}$

  ➡ fuzzy set of cdfs $\tilde{F}(x) = \left\{ \left(F_j(x),\ \mu\left(F_j(x)\right) \right) \ \middle|\ X_{ji} \in \tilde{X},\ \mu\left(F_j(x)\right) = \mu\left(X_j\right) \ \forall j \right\}$

Relationship to other concepts of imprecise probabilities

- evidence theory
  - » focal subsets represented by fuzzy sets
  - » frequentist view: probability of focal subsets induced by elementary events
- p-box approach
  - » fuzzy set of p-boxes $\tilde{F}(x) = \left\{ \left(F_\alpha(x),\ \mu\left(F_\alpha(x)\right) \right) \ \middle|\ F_\alpha(x) = \left[F_{\alpha l}(x), F_{\alpha u}(x) \right],\ \mu\left(F_\alpha(x)\right) = \alpha \ \forall \alpha \in (0,1] \right\}$

Numerical processing

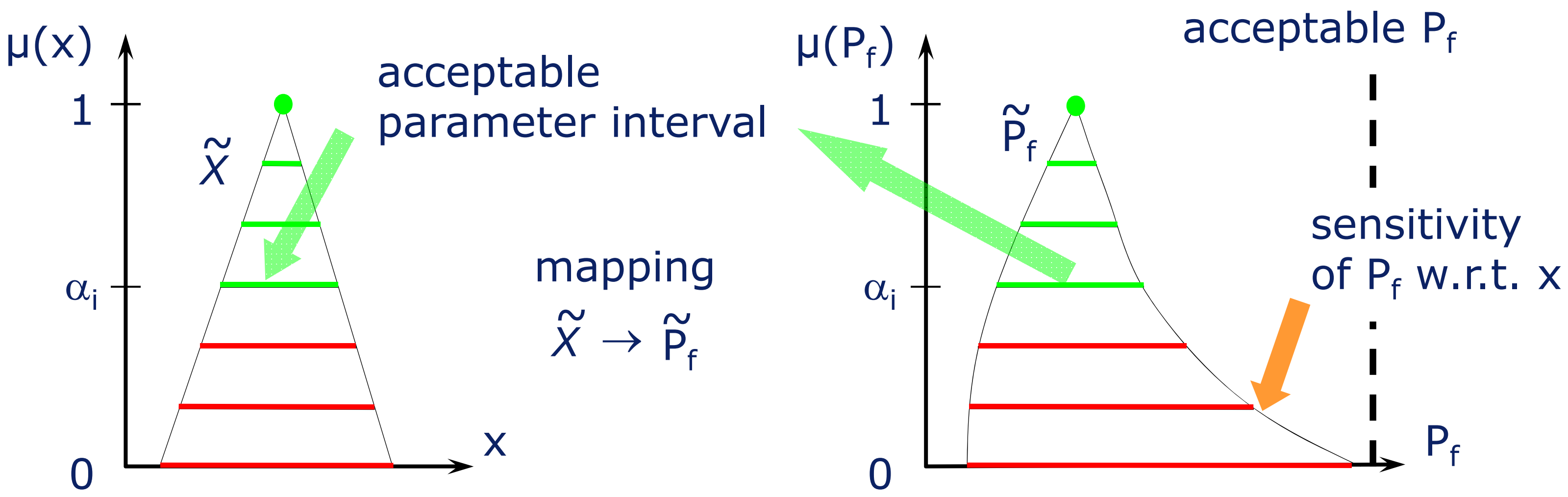
- stochastic techniques combined with interval / fuzzy analysis
  - » input: fuzzy set of plausible probabilistic models
  - » output: bounds on probabilities/probabilistic parameters for events/quantities of interest for several intensities of imprecision

FUZZY PROBABILITIES IN RELIABILITY ANALYSIS

Fuzzy parameters

- structural parameters
- probabilistic model parameters

Failure probability



- ➡ coarse specifications of design parameters & probabilistic models
- ➡ attention to / exclude model options leading to large imprecision of P_f
- ➡ acceptable imprecision of parameters & probabilistic models
- ➡ indications to collect additional information
- ➡ definition of quality requirements
- ➡ robust design

μ(.) not important, but analysis with various intensities of imprecision

EXAMPLE: RELIABILITY ANALYSIS

Fixed jacket platform

- imprecision in the models for
  - » wave, drag and ice loads
  - » wind load
  - » corrosion
  - » joints of tubular members
  - » foundation
  - » possible damage

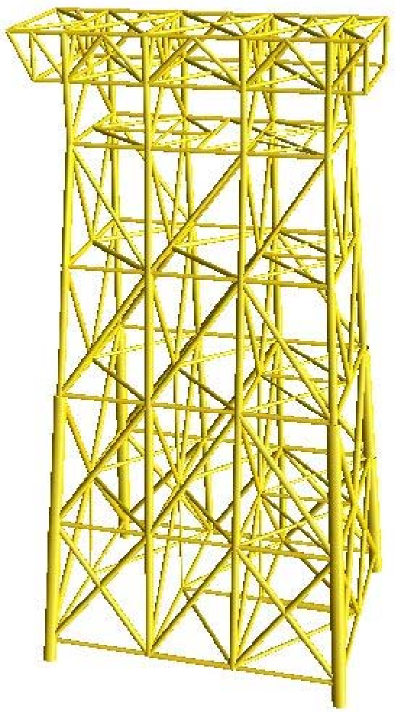
Collaboration with Profs.
Quek Ser Tong
Choo Yoo Sang

EXAMPLE: RELIABILITY ANALYSIS

Collaboration with Profs.
Quek Ser Tong
Choo Yoo Sang

Fixed jacket platform

- dimensions
  - » height: 142 m
  - » top: 27 X 54 m
  - » bottom: 56 X 70 m

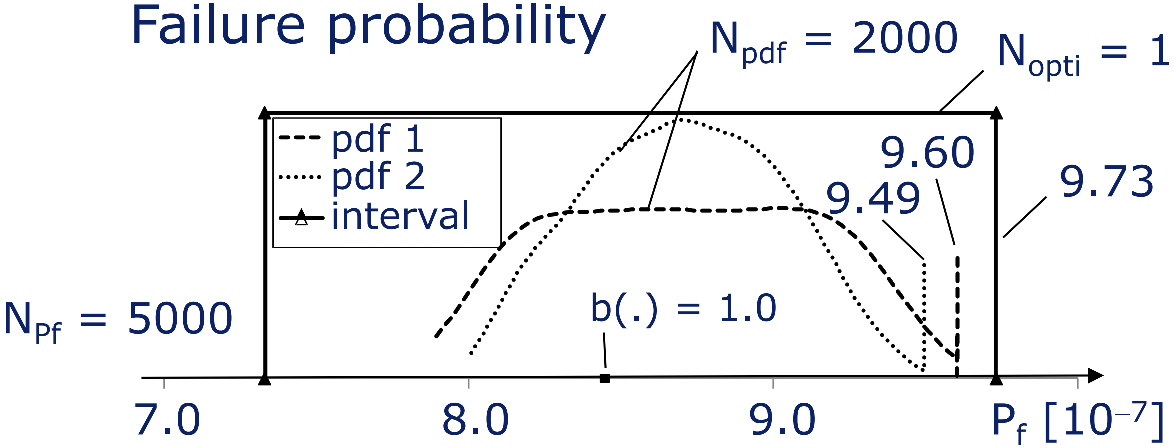


- loads, environment
  - » T = 15°C, t = 5 a
  - » random: wave height, current, yield stress, and corrosion depth c(t,E)
  - » pdf / interval for imprecise parameters

Reliability analysis

- Monte Carlo simulation with importance sampling and response surface approximation

Failure probability



STOCHASTIC SAMPLING

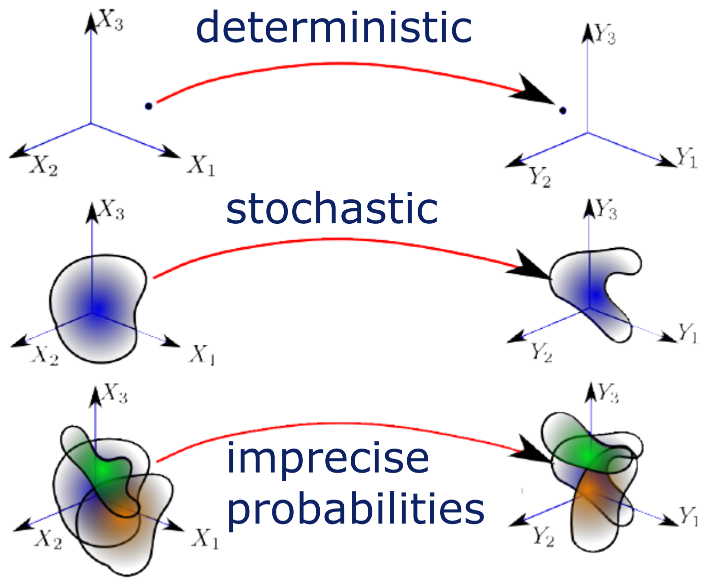
Features

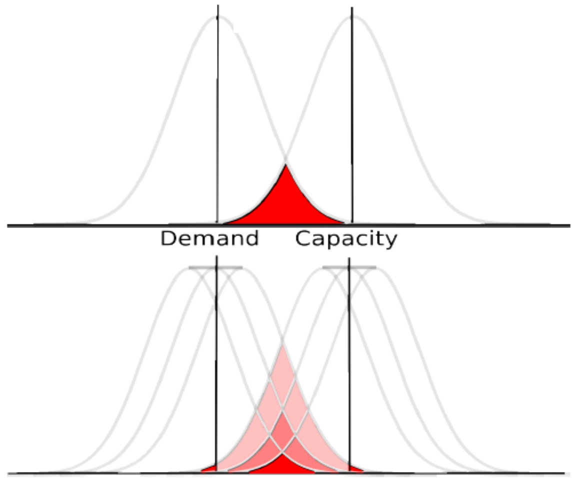
- generally applicable
  - » repeated analysis of deterministic model
  - » statistical evaluation of model response
- high numerical effort

➡ implementations to increase numerical efficiency
  - » high performance computing
  - » efficient (variance reducing) sampling schemes

➡ developments to further advance
  - » importance sampling
  - » subset sampling
  - » line sampling

in combination with subjective and imprecise probabilities

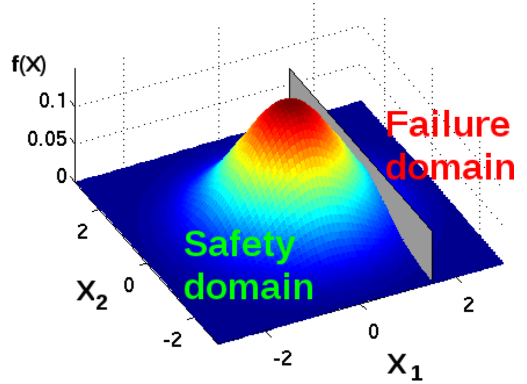




ADVANCED SAMPLING SCHEMES

Line sampling

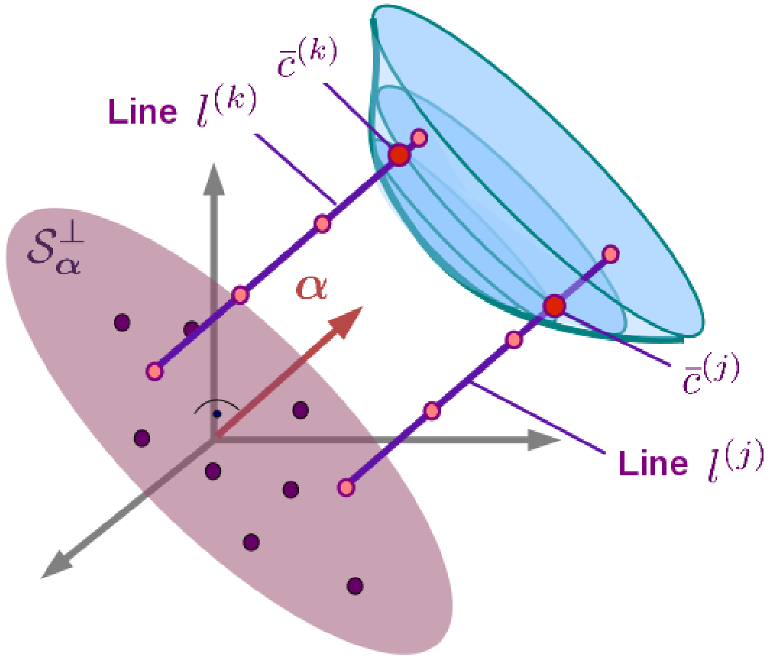
- concept
  - » in the standard normal space approximate P_f through distances to the limit state at random locations



- practical implementation
  - » build a subspace $S_\alpha^\perp$ perpendicular to the „important direction" α
  - » sample in $S_\alpha^\perp$ and find the distance $c^{(i)}$ to the limit state at the random points
  - » estimate P_f as average of the $p_f^{(i)}$ associated with the $c^{(i)}$

$$\hat{P}_f = \frac{1}{N_L} \sum_{i=1}^{N_L} p_f^{(i)} \qquad \hat{P}_f = \frac{1}{N_L} \sum_{i=1}^{N_L} \phi\left(-\bar{c}^{(i)}\right)$$

  - » independent of
    - dimensionality and
    - magnitude of P_f

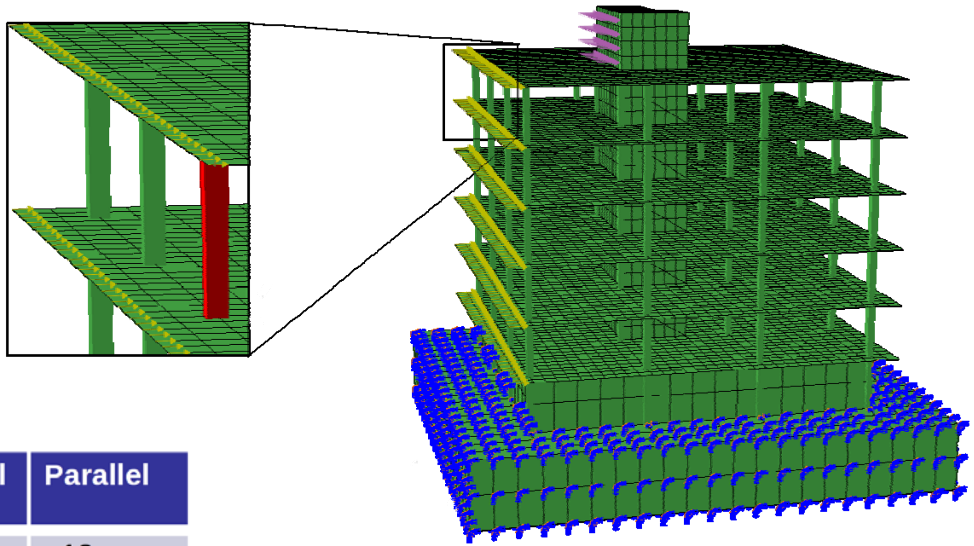


EXAMPLE: RELIABILITY ANALYSIS

Multi-storey building

- model
  - » 8,200 finite elements, 66,300 dof
  - » 244 random variables and 5 intervals (parameters of some rv)

- reliability analysis
  - » component failure
  - » line sampling
  - » interval analysis with global optimization
  - » distributed computing



Analysis type	Sequential	Parallel
Direct Monte Carlo Simulation 10000 samples	> 6 days	~13 hours
Advanced MCS – Line Sampling 100 samples	1h 32 min	7.8 min

Method	Speed-up
Distributed computing	11.7
Advanced Simulation	~ 100
Advanced Simulation + Distributed computing	> 1000

HIGH-DIMENSIONAL PROBLEMS

Advanced Monte Carlo simulation with interval probabilities

- global optimization problem

$$\underline{p_f} = \inf_{x,p} \int_{\Omega_f(x)} h_d(\xi, p)\, d\Omega$$

$$\overline{p_f} = \sup_{x,p} \int_{\Omega_f(x)} h_d(\xi, p)\, d\Omega$$

p – distribution parameters
ξ – random variables
x – intervals

Ω_f depends on intervals x !

- map intervals x to augmented probability space

$$\Omega \times X \to \Theta : \quad x \to \eta \in \mathbb{C}_x = \left\{ h_n\left(\eta; \underline{\overline{\mu_x}}, \sigma_x\right) \middle| \underline{\overline{\mu_x}} = \underline{\overline{x}} \right\}$$

- exploit topological properties of Θ for line sampling

sampling direction $-\nabla g$

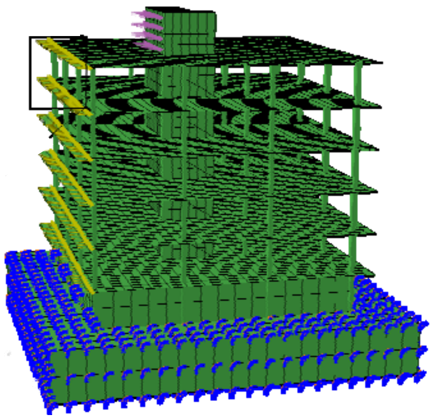
➡ optimal points $\left(p^u, x^u\right) = \psi^u\left(-\nabla g\right), \quad \left(p^l, x^l\right) = \psi^l\left(-\nabla g\right)$

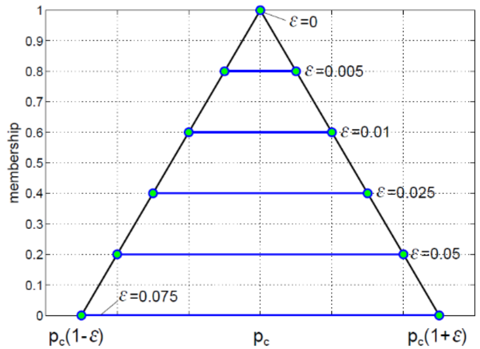
➡ $\overline{p_f} = \int_{\Omega_f(x^u)} h_d\left(\xi, p^u\right) d\Omega, \qquad \underline{p_f} = \int_{\Omega_f(x^l)} h_d\left(\xi, p^l\right) d\Omega$

EXAMPLE: ROBUST RELIABILITY ASSESSMENT

Multi-storey building – component failure analysis

- structure
  - » 8,200 finite elements, 66,300 dof
- imprecise probabilistic input
  - » 488 fuzzy parameters for 244 fuzzy random variables





± 7.5 %
tolerance range

SV #	Prob. dist.	$\overline{p} = p_c\ [1-\epsilon,\ 1+\epsilon]$		Description	Units
1	$\mathrm{N}(\overline{\mu},\ \overline{\sigma})$	$\mu_c = 0.1$	$\sigma_c = 0.01$	Columns' strength	GPa
2 – 193	$\mathrm{Unif}(\overline{a}, \overline{b})$	$a_c = 0.36$	$b_c = 0.44$	Sections' size	m
194 – 212	$\mathrm{LN}(\overline{m},\ \overline{v})$	$m_c = 35$	$v_c = 12.25$	Young's modulus	GPa
213 – 231	$\mathrm{LN}(\overline{m},\ \overline{v})$	$m_c = 2.5$	$v_c = 6.25\ 10^{-2}$	Material's density	$\mathrm{kg/dm}^3$
232 – 244	$\mathrm{LN}(\overline{m},\ \overline{v})$	$m_c = 0.25$	$v_c = 6.25\ 10^{-4}$	Poisson's ratio	-

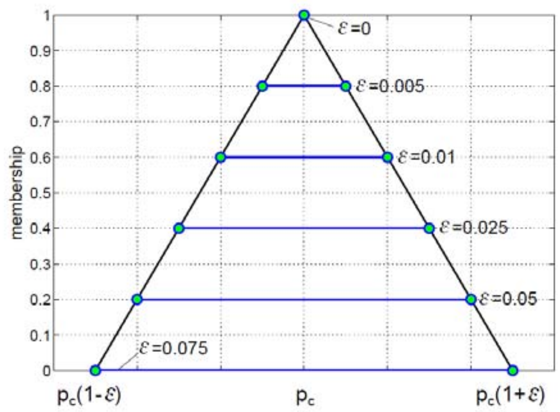
EXAMPLE: ROBUST RELIABILITY ASSESSMENT

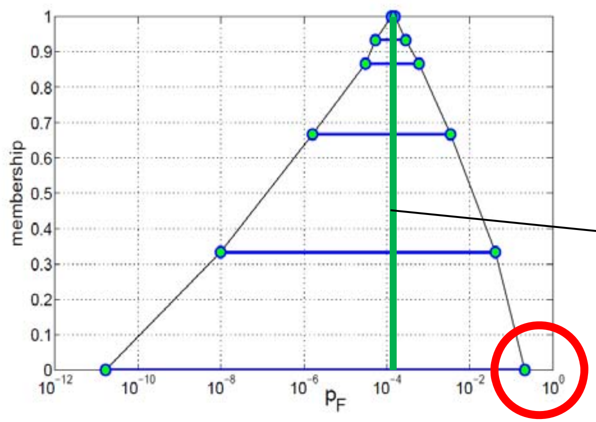
Multi-storey building – results

- advanced line sampling with pre-identified optimal points in Θ

	Lower Bound		Upper Bound		
ϵ	$\underline{p}_F$	CoV	$\overline{p}_F$	CoV	Ns
0.000	$1.42\ 10^{-4}$	$9.2\ 10^{-2}$	$1.42\ 10^{-4}$	$9.2\ 10^{-2}$	126
0.005	$5.75\ 10^{-5}$	$8.7\ 10^{-2}$	$2.63\ 10^{-4}$	$7.1\ 10^{-2}$	257
0.010	$4.57\ 10^{-5}$	$33.6\ 10^{-2}$	$5.30\ 10^{-4}$	$11.5\ 10^{-2}$	250
0.025	$1.75\ 10^{-6}$	$8.8\ 10^{-2}$	$3.22\ 10^{-3}$	$5.3\ 10^{-2}$	253
0.050	$2.27\ 10^{-8}$	$57.0\ 10^{-2}$	$3.88\ 10^{-2}$	$5.4\ 10^{-2}$	255
0.075	$1.88\ 10^{-11}$	$12.2\ 10^{-2}$	$2.02\ 10^{-1}$	$3.5\ 10^{-2}$	254

sample size
1,395





sensitivity of
failure probability

$P_f \approx 10^{-4}$

safety alerts

$\approx 10^{-1}$!

RESUMÉ

Major directions of development

$$\underbrace{\begin{array}{l}\bullet\ \text{advanced stochastic modeling}\\ \bullet\ \text{generalized models for vague and imprecise information}\\ \bullet\ \text{Monte Carlo simulation methods}\end{array}}_{\text{combinations}}$$

➡ realistic models

➡ efficient numerical analysis

➡ UQ for industry-sized structures and systems

➡ improved design, performance and reliability