Combining machine learning and surrogate modeling for data-driven uncertainty propagation in high-dimension

Author(s):
Lataniotis, Christos; Marelli, Stefano; Sudret, Bruno

Publication date:
2019-06-24

Permanent link:
https://doi.org/10.3929/ethz-b-000352572

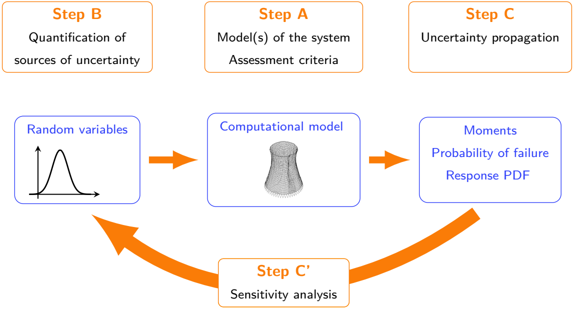
Combining machine learning and surrogate modeling for data-driven uncertainty propagation in high-dimension

C. Lataniotis, S. Marelli*, B. Sudret

Chair of Risk, Safety & Uncertainty Quantification – ETH Zürich

UNCECOMP 2019, Crete (Greece), 24.06.2019

High Dimensional, Data-Driven UQ



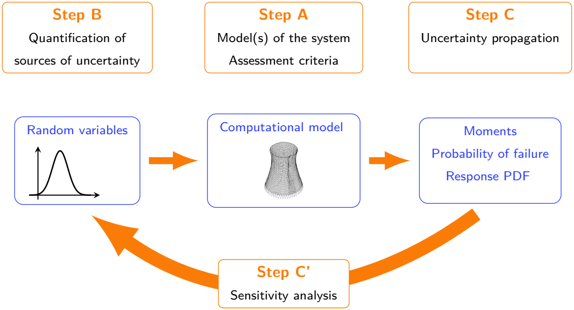
Standard Requirements

A. A computational model (e.g. FEM model or surrogate model)

B. Some representation of the input variability (e.g. $f_{\boldsymbol{X}}(\boldsymbol{x})$), or resampling capabilities

Sudret, B. (2007). Uncertainty propagation and sensitivity analysis in mechanical models - Contributions to structural reliability and stochastic spectral methods. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand.

High Dimensional, Data-Driven UQ



High dimensionality, purely data driven

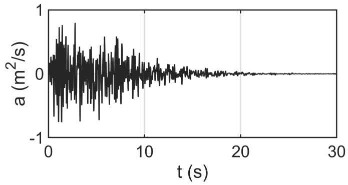
A. No computational model, no surrogate in high dimension ($M \sim 10^2 - 10^6$)

B. Only a limited input sample available, no inference possible

Goal: Do something better than just sample mean and sample variance

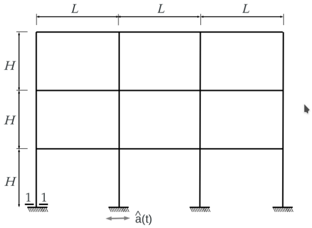
Why would this be useful?

Earthquake engineering



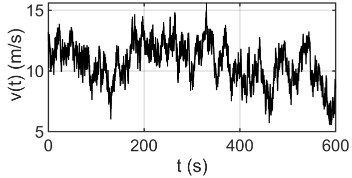
Earthquake databases

→



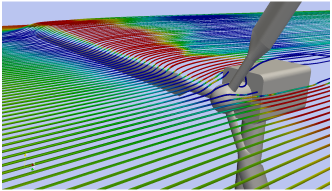
Building/Local/Regional damage

Structural health monitoring



Sensor readings

→



Residual lifespan/Power throughput

Outline

Outline

Principle of Reduced Dimension Resampling (RDR)

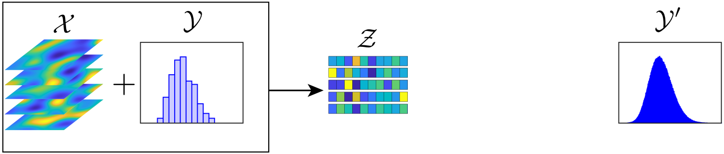
Goal: Estimate response PDF $f_Y(y)$ from a dataset $\mathcal{X}$,$\mathcal{Y}$

Large input dimension M

- Cannot infer $f_{\boldsymbol{X}}(\boldsymbol{x})$:
  no $\mathcal{X}$ enrichment
- Cannot surrogate $y = \mathcal{M}(\boldsymbol{x})$:
  no $\mathcal{Y}$ enrichment

Principle of Reduced Dimension Resampling (RDR)

Goal: Estimate response PDF $f_Y(y)$ from a dataset $\mathcal{X}$,$\mathcal{Y}$



Large input dimension M

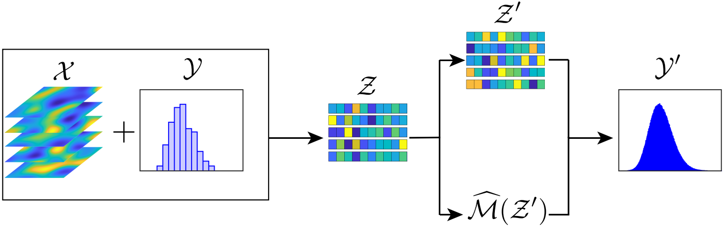
- Cannot infer $f_{\boldsymbol{X}}(\boldsymbol{x})$:
  no $\mathcal{X}$ enrichment
- Cannot surrogate $y = \mathcal{M}(\boldsymbol{x})$:
  no $\mathcal{Y}$ enrichment

The idea: RDR

1. Compression $\mathcal{X} \rightarrow \mathcal{Z}$

Principle of Reduced Dimension Resampling (RDR)

Goal: Estimate response PDF $f_Y(y)$ from a dataset $\mathcal{X}$,$\mathcal{Y}$



Large input dimension M

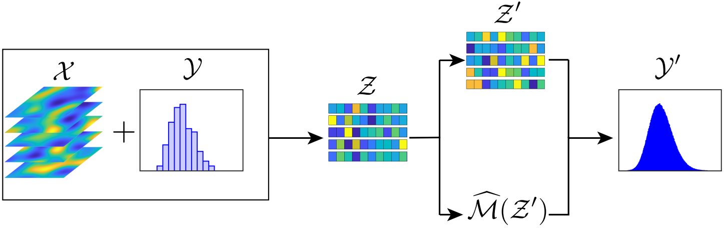
- Cannot infer $f_{\boldsymbol{X}}(\boldsymbol{x})$:
  no $\mathcal{X}$ enrichment
- Cannot surrogate $y = \mathcal{M}(\boldsymbol{x})$:
  no $\mathcal{Y}$ enrichment

The idea: RDR

1. Compression $\mathcal{X} \rightarrow \mathcal{Z}$
2. Reduced surrogate modeling: $\mathcal{Y} \approx \widehat{\mathcal{M}}(\mathcal{Z})$

Principle of Reduced Dimension Resampling (RDR)

Goal: Estimate response PDF $f_Y(y)$ from a dataset $\mathcal{X}$,$\mathcal{Y}$



Large input dimension M

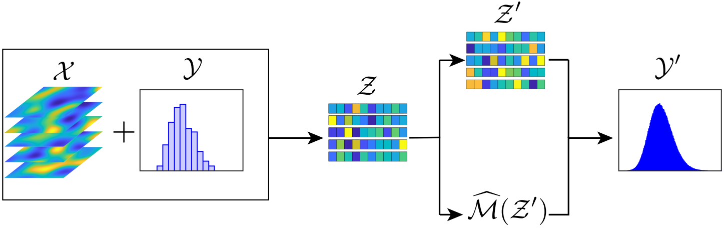
- Cannot infer $f_{\boldsymbol{X}}(\boldsymbol{x})$:
  no $\mathcal{X}$ enrichment
- Cannot surrogate $y = \mathcal{M}(\boldsymbol{x})$:
  no $\mathcal{Y}$ enrichment

The idea: RDR

1. Compression $\mathcal{X} \to \mathcal{Z}$
2. Reduced surrogate modeling: $\mathcal{Y} \approx \widehat{\mathcal{M}}(\mathcal{Z})$
3. Infer $f_{\boldsymbol{Z}}(\boldsymbol{z})$ from $\mathcal{Z}$

Principle of Reduced Dimension Resampling (RDR)

Goal: Estimate response PDF $f_Y(y)$ from a dataset $\mathcal{X},\mathcal{Y}$



Large input dimension M

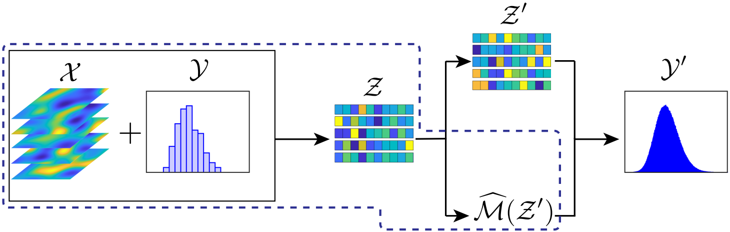
- Cannot infer $f_{\boldsymbol{X}}(\boldsymbol{x})$: no $\mathcal{X}$ enrichment
- Cannot surrogate $y = \mathcal{M}(\boldsymbol{x})$: no $\mathcal{Y}$ enrichment

The idea: RDR

1. Compression $\mathcal{X} \to \mathcal{Z}$
2. Reduced surrogate modeling: $\mathcal{Y} \approx \widehat{\mathcal{M}}(\mathcal{Z})$
3. Infer $f_{\boldsymbol{Z}}(\boldsymbol{z})$ from $\mathcal{Z}$
4. Resample $\mathcal{Z}'$ and evaluate $\mathcal{Y}' = \widehat{\mathcal{M}}(\mathcal{Z}')$

Compression and surrogate: a tight bond

Step 1: Compression and surrogate modeling



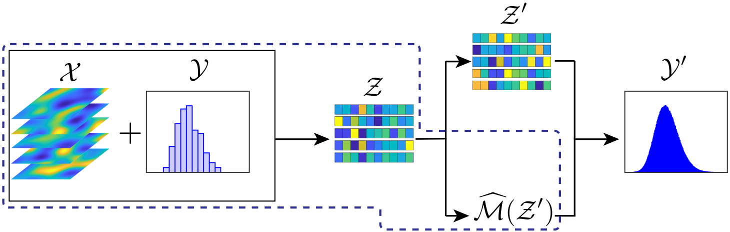
- **The goal:** identify $g, \boldsymbol{w}, \widehat{\mathcal{M}}$ and $\boldsymbol{\theta}$ such that:

$$\boldsymbol{z} = g(\boldsymbol{X}, \boldsymbol{w})\,,\ \boldsymbol{Z} \in \mathbb{R}^m\,,\ m \ll M$$

$$\widehat{\mathcal{M}}\left(\boldsymbol{Z}; \boldsymbol{\theta}\right) \approx \mathcal{M}\left(\boldsymbol{X}\right)$$

Compression and surrogate: a tight bond

Step 1: Compression and surrogate modeling



- **The goal:** identify $g, \boldsymbol{w}, \widehat{\mathcal{M}}$ and $\boldsymbol{\theta}$ such that:

$$\boldsymbol{z} = g(\boldsymbol{X}, \boldsymbol{w}) \, , \; \boldsymbol{Z} \in \mathbb{R}^m \, , \; m \ll M$$

$$\widehat{\mathcal{M}}\left(\boldsymbol{Z}; \boldsymbol{\theta}\right) \approx \mathcal{M}\left(\boldsymbol{X}\right)$$

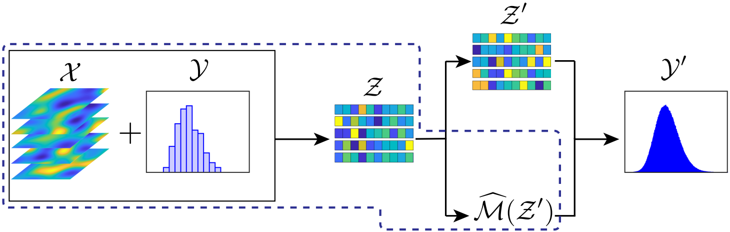
DRSM algorithm: Lataniotis et al., 2018

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w} \in \mathcal{D}_{\boldsymbol{w}}} \widehat{\varepsilon}_{\text{gen}}(\boldsymbol{w}; \widehat{\boldsymbol{\theta}}(\boldsymbol{w}), \mathcal{X}, \mathcal{Y}),$$

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\theta}}} \widehat{\varepsilon}_{\text{gen}}(\boldsymbol{\theta}; \boldsymbol{w}, \mathcal{X}, \mathcal{Y})$$

Compression and surrogate: a tight bond

Step 1: Compression and surrogate modeling



- **The goal:** identify $g, \boldsymbol{w}, \widehat{\mathcal{M}}$ and $\boldsymbol{\theta}$ such that:

$$\boldsymbol{z} = g(\boldsymbol{X}, \boldsymbol{w}), \; \boldsymbol{Z} \in \mathbb{R}^m, \; m \ll M$$

$$\widehat{\mathcal{M}}(\boldsymbol{Z}; \boldsymbol{\theta}) \approx \mathcal{M}(\boldsymbol{X})$$

DRSM algorithm: Lataniotis et al., 2018

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w} \in \mathcal{D}_{\boldsymbol{w}}} \widehat{\varepsilon}_{\text{gen}}(\boldsymbol{w}; \widehat{\boldsymbol{\theta}}(\boldsymbol{w}), \mathcal{X}, \mathcal{Y}),$$

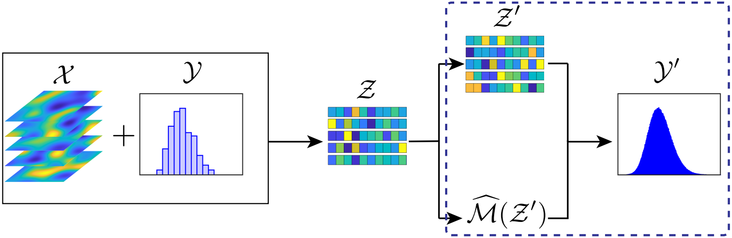
$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \mathcal{D}_{\boldsymbol{\theta}}} \widehat{\varepsilon}_{\text{gen}}(\boldsymbol{\theta}; \boldsymbol{w}, \mathcal{X}, \mathcal{Y})$$

Interesting DRSM features

- Non-intrusive
- Optimal DR w.r.t. the surrogate performance
- Outcome
  - Surrogate $\widehat{\mathcal{M}}(\boldsymbol{z}; \widehat{\boldsymbol{\theta}})$
  - $\mathcal{Z} = g(\mathcal{X}; \widehat{\boldsymbol{w}})$

Methodology (2/2)

Step 2: Reduced Dimension Resampling

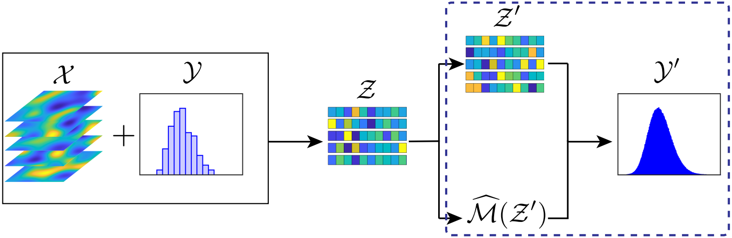


Probabilistic modeling

- **The goal:** infer the joint distribution of the reduced input: $Z \sim f_{Z}(Z)$
- **2-steps inference** Torre et al., 2019
  - Marginal fitting
  - Advanced copula inference

Methodology (2/2)

Step 2: Reduced Dimension Resampling



Probabilistic modeling

- **The goal:** infer the joint distribution of the reduced input: $Z \sim f_Z(Z)$
- **2-steps inference** Torre et al., 2019
  - Marginal fitting
  - Advanced copula inference

Resampling + MCS

- **The goal:** enrich the original response sample $\mathcal{Y} \rightarrow \mathcal{Y}'$
- **Resampling + surrogate**
  - resample $\mathcal{Z}'$ from $f_Z(z)$ (e.g. Rosenblatt, rejection sampling, MCMC)
  - evaluate $\mathcal{Y}' = \widehat{\mathcal{M}}(\mathcal{Z}')$

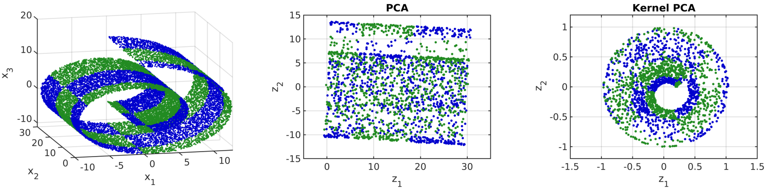
Outline

Dimensionality reduction

Kernel Principal Component Analysis (KPCA)

Schölkopf et al., 1999

A non-linear extension of PCA



- First map the input to a high dimensional space (feature space):

$$\boldsymbol{x} \in \mathcal{R}^M \mapsto \Phi(\boldsymbol{x}; \boldsymbol{w}') \in \mathcal{H}$$

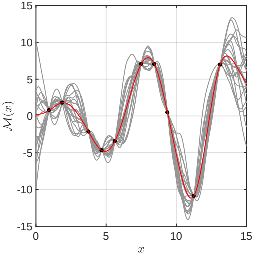
  then perform PCA in this space

- The mapping is implicit, using appropriate kernels:

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{w}) = \Phi(\boldsymbol{x}_i; \boldsymbol{w}') \cdot \Phi(\boldsymbol{x}_j; \boldsymbol{w}')$$

- Calculates $\mathcal{Z}$ as projections onto the first m eigenvectors of $\boldsymbol{C}_{\mathcal{H}} = \text{cov}\left[\Phi(\mathcal{X})\right]$

Surrogate modeling methods



Universal Kriging:

$$\widehat{\mathcal{M}}(\boldsymbol{X}) = \boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{X}) + \sigma^2 Z(\boldsymbol{X}, R(\boldsymbol{x}, \boldsymbol{x}_{\text{ED}}; \widehat{\boldsymbol{\theta}}))$$

$$\widehat{\boldsymbol{\theta}} = \arg\min J(\widehat{\boldsymbol{\theta}})$$

Objective function varies depending on the estimation method (*e.g.* maximum likelihood, cross-validation, *etc.*)

Sparse polynomial chaos expansions:

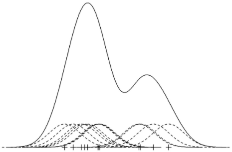
$$\widehat{\mathcal{M}}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \theta_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$$

$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$ are multivariate polynomials orthonormal with respect to $f_{\boldsymbol{X}}(\boldsymbol{x})$

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{A}|}}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{\boldsymbol{\alpha} \in \mathcal{A}} \theta_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(i)}) - y^{(i)} \right)^2 + \lambda \|\boldsymbol{\theta}\|_1$$

Probabilistic input model inference (data-driven)

Marginal distributions inference



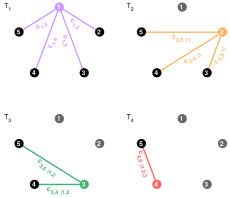
- Using kernel density estimation

$$\widehat{f}_Z(z) = \frac{1}{N\,h} \sum_{i=1}^{N} \kappa\left(\frac{z - z^{(i)}}{h}\right)$$

- Non-parametric technique, suitable for data-driven applications

Copula inference

Torre et al., 2019



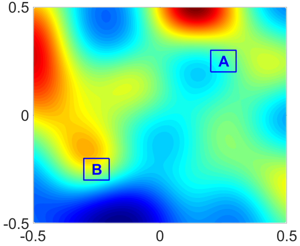
- Using canonical vine copulas
- Constructed as a product of pair copulas
- Copula structure and parameters are inferred based on Aas et al. 2009
- Allow for efficient implementation of the Rosenblatt transform

Outline

2D heat diffusion: problem statement

Input

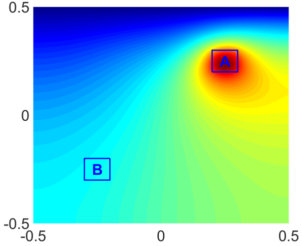
$d(\boldsymbol{v})$



- Heat diffusion coefficient $d(\boldsymbol{v})$
- Lognormal random field

Computational model

$T(\boldsymbol{v})$



$$-\nabla\left(d(\boldsymbol{v})\nabla T(\boldsymbol{v})\right) = 500\, I_A(\boldsymbol{v})$$

- **Output**: $Y = \frac{1}{|B|}\int_{\boldsymbol{v}\in B} T(\boldsymbol{v})\ [^oC]$

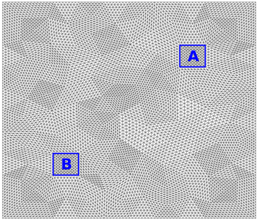
Training: 800 samples **Validation**: 10^5 samples

Li, C.C and Der Kiureghian, A. (1993). Optimal discretization of random fields. J. Eng. Mech. 119 (6), pp. 1136-1154.

2D heat diffusion: problem statement

Input

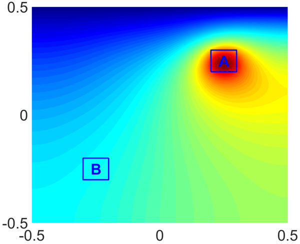
$d(\boldsymbol{v})$



- Heat diffusion coefficient $d(\boldsymbol{v})$
- Lognormal random field
- **Dimensionality** $\mathbf{M = 16,000}$

Computational model

$T(\boldsymbol{v})$



$$-\nabla\left(d(\boldsymbol{v})\nabla T(\boldsymbol{v})\right) = 500\, I_A(\boldsymbol{v})$$

- Output: $Y = \frac{1}{|B|}\int_{\boldsymbol{v}\in B} T(\boldsymbol{v})\ [^oC]$

Training: 800 samples **Validation**: 10^5 samples

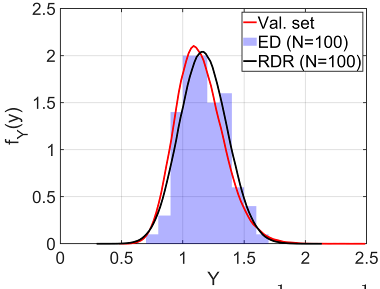
Li, C.C and Der Kiureghian, A. (1993). Optimal discretization of random fields. J. Eng. Mech. 119 (6), pp. 1136-1154.

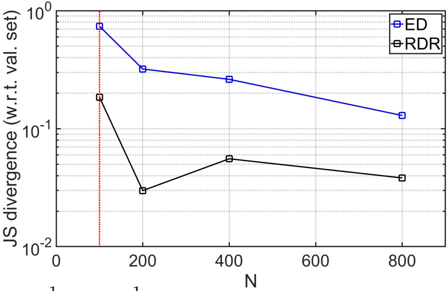
2D heat diffusion: response PDF estimation

DRSM results:	Compression scheme:	KPCA (Polynomial kernel)
	Optimal reduced dimension:	$m = 20$
	Surrogate model:	Kriging

2D heat diffusion: response PDF estimation

DRSM results:	Compression scheme:	KPCA (Polynomial kernel)
	Optimal reduced dimension:	$m = 20$
	Surrogate model:	Kriging





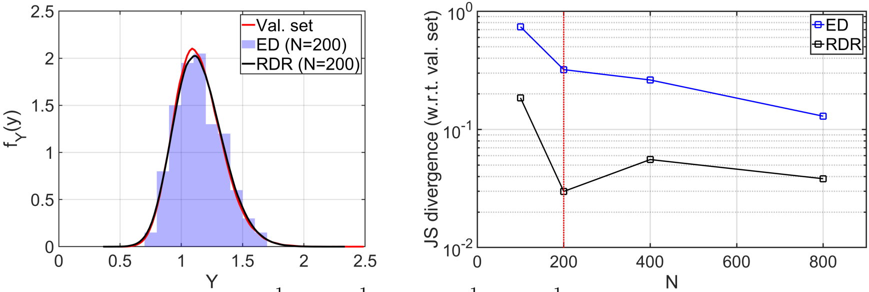
$$JS(p \parallel q) = \frac{1}{2} KL(p \parallel \frac{1}{2}(p+q)) + \frac{1}{2} KL(q \parallel \frac{1}{2}(p+q))$$

$$KL(p \parallel q) = \int_{-\infty}^{+\infty} p_X(x) \log\left(\frac{p_X(x)}{q_X(x)}\right) dx \approx \sum_{i=1}^{n_b} \tilde{p}_i \log\left(\frac{\tilde{p}_i}{\tilde{q}_i}\right)$$

Lin, J (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory, 37 (1), pp. 145-151.

2D heat diffusion: response PDF estimation

DRSM results:	Compression scheme:	KPCA (Polynomial kernel)
	Optimal reduced dimension:	$m = 20$
	Surrogate model:	Kriging



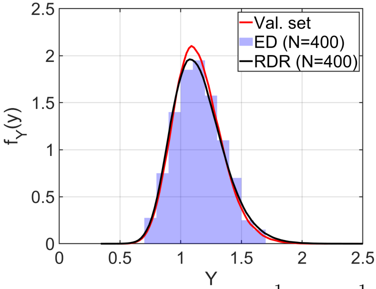
$$JS(p \parallel q) = \frac{1}{2} KL(p \parallel \frac{1}{2}(p+q)) + \frac{1}{2} KL(q \parallel \frac{1}{2}(p+q))$$

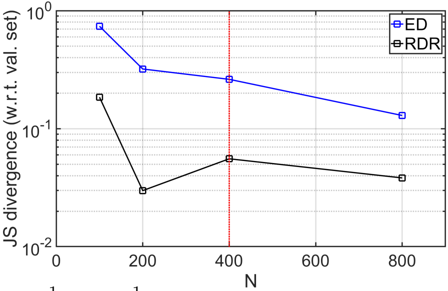
$$KL(p \parallel q) = \int_{-\infty}^{+\infty} p_X(x) \log\left(\frac{p_X(x)}{q_X(x)}\right) dx \approx \sum_{i=1}^{n_b} \tilde{p}_i \log\left(\frac{\tilde{p}_i}{\tilde{q}_i}\right)$$

Lin, J (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory, 37 (1), pp. 145-151.

2D heat diffusion: response PDF estimation

DRSM results:	Compression scheme:	KPCA (Polynomial kernel)
	Optimal reduced dimension:	$m = 20$
	Surrogate model:	Kriging





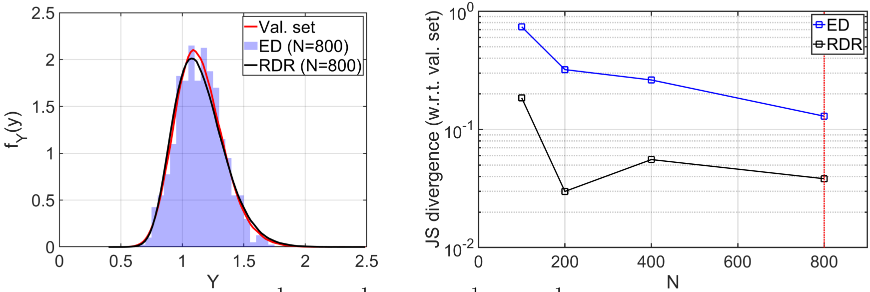
$$JS(p \parallel q) = \frac{1}{2} KL(p \parallel \frac{1}{2}(p+q)) + \frac{1}{2} KL(q \parallel \frac{1}{2}(p+q))$$

$$KL(p \parallel q) = \int_{-\infty}^{+\infty} p_X(x) \log\left(\frac{p_X(x)}{q_X(x)}\right) dx \approx \sum_{i=1}^{n_b} \tilde{p}_i \log\left(\frac{\tilde{p}_i}{\tilde{q}_i}\right)$$

Lin, J (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory, 37 (1), pp. 145-151.

2D heat diffusion: response PDF estimation

DRSM results:	Compression scheme:	KPCA (Polynomial kernel)
	Optimal reduced dimension:	$m = 20$
	Surrogate model:	Kriging

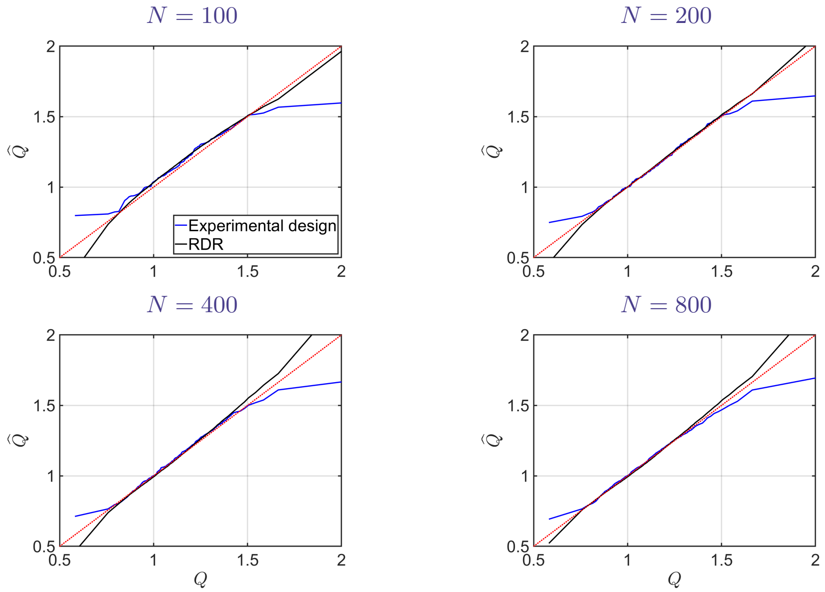


$$JS(p \parallel q) = \frac{1}{2} KL(p \parallel \frac{1}{2}(p+q)) + \frac{1}{2} KL(q \parallel \frac{1}{2}(p+q))$$

$$KL(p \parallel q) = \int_{-\infty}^{+\infty} p_X(x) \log\left(\frac{p_X(x)}{q_X(x)}\right) dx \approx \sum_{i=1}^{n_b} \tilde{p}_i \log\left(\frac{\tilde{p}_i}{\tilde{q}_i}\right)$$

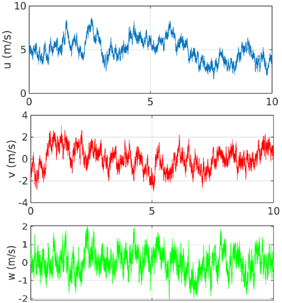
Lin, J (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory, 37 (1), pp. 145-151.

2D heat diffusion: Quantile estimation



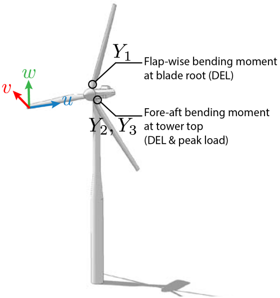
Structural health monitoring of a wind turbine (SHM)

Input



- Inflow wind speed realisations generated by TurbSim — Jonkman, 2009
- 10 min. each, sampled at 20 Hz
- $M = 12,000$ (u only), or $M = 36,000$ (u, v, w)

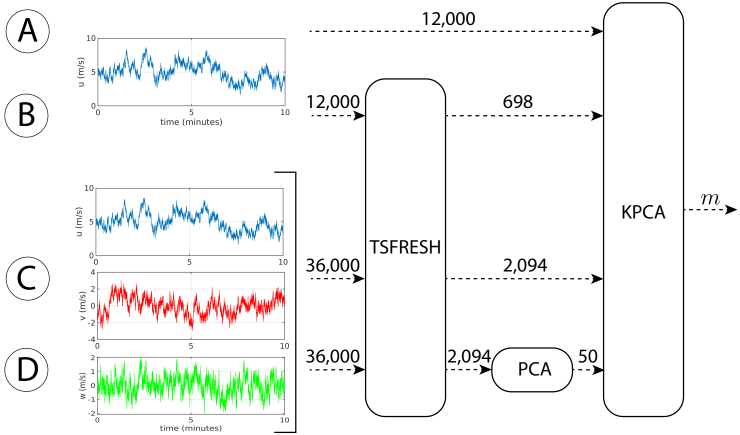
Computational model



- Aero-servo-elastic-simulation using OpenFAST — Jonkman, 2013
- Fatigue accumulation estimated by damage equivalent loads (DEL) — IEC 61400-1 standard

Training: $1,000$ samples

SHM: a more advanced set of compression schemes



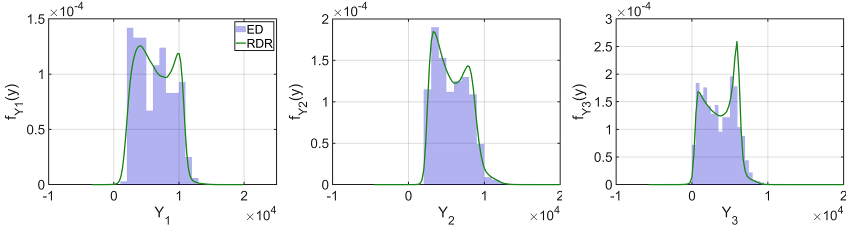
Christ, M., et al. (2018). Time Series FeatuRe Extraction on basis of scalable hypothesis tests (tsfresh - A Python package). Neurocomputing 307, pp. 72-77.

SHM: estimates of the response PDFs

DRSM Results:	**F-W DEL** (Y_1)	**F-A DEL** (Y_2)	**F-A Peak Load** (Y_3)
Compression scheme	TSFRESH+PCA +KPCA	TSFRESH+KPCA	TSFRESH+KPCA
Input time series	$(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$	$\boldsymbol{u}$	$(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$
Reduced dimension	$m_1 = 20$	$m_2 = 10$	$m_3 = 25$
Surrogate	sparse PCE		

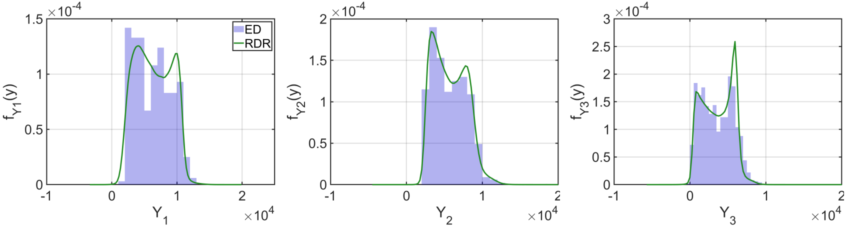
SHM: estimates of the response PDFs

DRSM Results:	**F-W DEL** (Y_1)	**F-A DEL** (Y_2)	**F-A Peak Load** (Y_3)
Compression scheme	TSFRESH+PCA +KPCA	TSFRESH+KPCA	TSFRESH+KPCA
Input time series	$(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$	$\boldsymbol{u}$	$(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$
Reduced dimension	$m_1 = 20$	$m_2 = 10$	$m_3 = 25$
Surrogate	sparse PCE		



SHM: estimates of the response PDFs

DRSM Results:	**F-W DEL** (Y_1)	**F-A DEL** (Y_2)	**F-A Peak Load** (Y_3)
Compression scheme	TSFRESH+PCA +KPCA	TSFRESH+KPCA	TSFRESH+KPCA
Input time series	$(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$	$\boldsymbol{u}$	$(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$
Reduced dimension	$m_1 = 20$	$m_2 = 10$	$m_3 = 25$
Surrogate	sparse PCE		



A multimodal behavior is identified

Outline

Summary & Outlook

Summary

- We introduced a novel approach for data-driven uncertainty propagation with high dimensional inputs
- Extracts additional information form the available data (PDF, quantiles)
- Can be used in the presence of multiple types/sources of data

Summary & Outlook

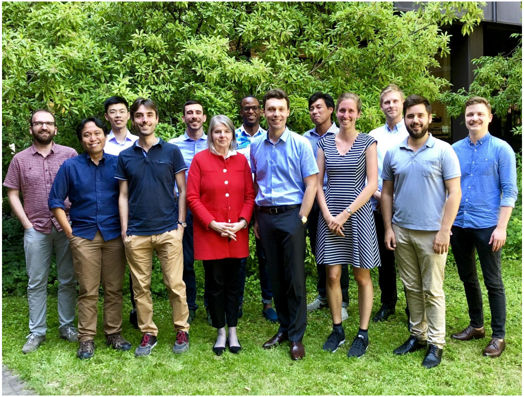
Summary

- We introduced a novel approach for data-driven uncertainty propagation with high dimensional inputs
- Extracts additional information form the available data (PDF, quantiles)
- Can be used in the presence of multiple types/sources of data

Outlook

- Exploration of other applications (e.g. reliability analysis)
- Further investigation of the structural health monitoring application
  - Estimate the probability of a component failure within a pre-defined timeline
  - Fuse additional inputs
  - Sensory readings instead of simulated data

Questions?



Chair of Risk, Safety & Uncertainty Quantification

www.rsuq.ethz.ch

The Uncertainty Quantification Software

www.uqlab.com

Thank you very much for your attention!

References

Christ, Maximilian, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. 2018. "Time Series FeatuRe Extraction on basis of scalable hypothesis tests (tsfresh–A Python package)." *Neurocomputing* 307:72–77.

Jonkman, Bonnie J. 2009. *TurbSim user's guide: Version 1.50.* Technical report. National Renewable Energy Lab.(NREL), Golden, CO (United States).

Jonkman, Jason. 2013. "The new modularization framework for the FAST wind turbine CAE tool." In *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition,* 202.

Lataniotis, Christos, Stefano Marelli, and Bruno Sudret. 2018. "Extending classical surrogate modelling to ultrahigh dimensional problems through supervised dimensionality reduction: a data-driven approach." *arXiv preprint arXiv:1812.06309.*

Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller. 1998. "Nonlinear Component Analysis As a Kernel Eigenvalue Problem." *Neural Comput.* 10 (5): 1299–1319.

Sudret, B. 2007. *Uncertainty propagation and sensitivity analysis in mechanical models - Contributions to structural reliability and stochastic spectral methods.* Habilitation Thesis. Université Blaise Pascal, Clermont-Ferrand, France.

Torre, E., S. Marelli, P. Embrechts, and B. Sudret. 2018. "A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas." *Prob. Eng. Mech.* 55:1–16.