Vine copulas for uncertainty quantification: why and how

Other Conference Item

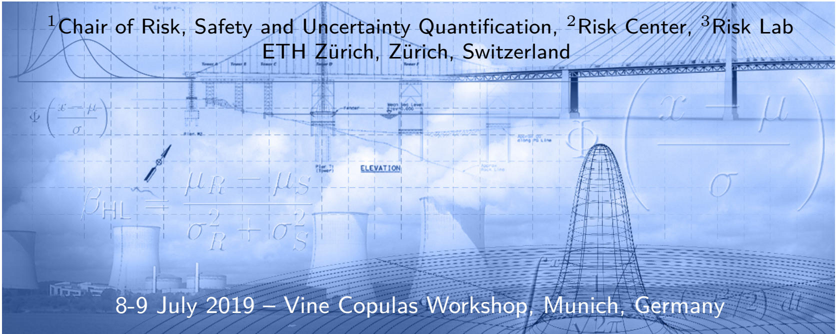
Author(s):
Torre, Emiliano; Marelli, Stefano; Embrechts, Paul; Sudret, Bruno

Publication date:
2019-07-08

Permanent link:
https://doi.org/10.3929/ethz-b-000353198

Vine copulas for Uncertainty Quantification: why and how

Emiliano Torre[1,2,3], Stefano Marelli[1],
Paul Embrechts[2,3], Bruno Sudret[1,2]



[1]Chair of Risk, Safety and Uncertainty Quantification, [2]Risk Center, [3]Risk Lab
ETH Zürich, Zürich, Switzerland

8-9 July 2019 – Vine Copulas Workshop, Munich, Germany

Outline

Outline

Problem statement

A model $\mathcal{M}$ subject to input $\boldsymbol{X}$ produces the response Y:

$$Y = \mathcal{M}(\boldsymbol{X}),$$

where:

- $\mathcal{M}$: known computational model (black box)
- $\boldsymbol{X} = (X_1, \dots, X_d)$: real-valued **random vector** with joint cdf $F_{\boldsymbol{X}}$
- $\Rightarrow$ Y uncertain: real-valued **random variable**

Uncertainty Quantification (UQ): Statistics of Y?

$\mathcal{M}$ computationally expensive $\Rightarrow$ Monte Carlo unaffordable $\Rightarrow$ faster methods.

Problem statement

A model $\mathcal{M}$ subject to input $\boldsymbol{X}$ produces the response Y:

$$Y = \mathcal{M}(\boldsymbol{X}),$$

where:

- $\mathcal{M}$: known computational model (black box)
- $\boldsymbol{X} = (X_1, \ldots, X_d)$: real-valued **random vector** with joint cdf $F_{\boldsymbol{X}}$
- $\Rightarrow$ Y uncertain: real-valued **random variable**

Uncertainty Quantification (UQ): Statistics of Y?

$\mathcal{M}$ computationally expensive $\Rightarrow$ Monte Carlo unaffordable $\Rightarrow$ faster methods.

UQ analyses and methods

Parameters of $Y = \mathcal{M}(\boldsymbol{X})$ of interest in UQ:

- Global parameters: $\mathbb{E}(Y^k)$, PDF f_Y, ...
  $\hookrightarrow$ Perturbation theory, surrogate modelling
- Tail probabilities, extreme quantiles: $\mathbb{P}(Y \geq y_{\text{critical}})$
  $\hookrightarrow$ Reliability analysis: FORM, SORM, subset simulation, ...
- Conditional tail probabilities: $\mathbb{P}(Y \geq y_{\text{critical}} | X_2 = x_2)$
  $\hookrightarrow$ Fragility analysis
- Sensitivity indices, importance factors
  $\hookrightarrow$ Sensitivity analysis
- ...

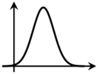
Uncertainty analysis: main steps

Step B
Quantification of
sources of uncertainty

Step A
Model(s) of the system
Assessment criteria

Step C
Uncertainty propagation

Random variables



→

Computational model

→

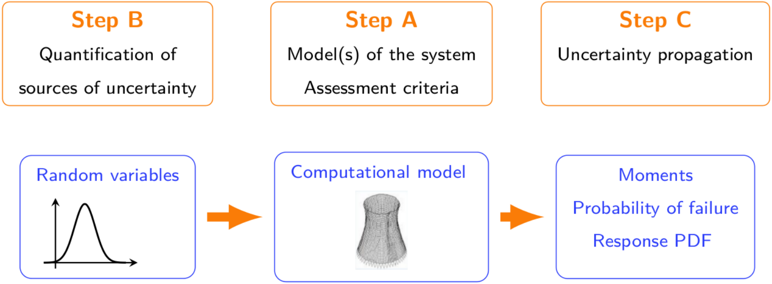
Moments
Probability of failure
Response PDF

- Most research in UQ: given $\mathcal{M}$ (step A) and F_X (step B), solve step C. Historically: $\{X_i\}$ assumed independent or with Gaussian copula.
- Focus here: choose a suitable F_X (step B)

  ↪ Especially a suitable copula. For large d: vines

  ↪ Map to specific probability spaces needed for some UQ methods

Uncertainty analysis: main steps



- Most research in UQ: given $\mathcal{M}$ (step A) and $F_{\boldsymbol{X}}$ (step B), solve step C. Historically: $\{X_i\}$ assumed independent or with Gaussian copula.
- Focus here: choose a suitable $F_{\boldsymbol{X}}$ (step B)

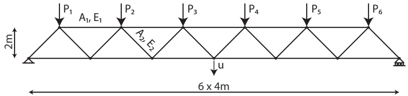
  $\hookrightarrow$ Especially a suitable copula. For large d: vines

  $\hookrightarrow$ Map to specific probability spaces needed for some UQ methods

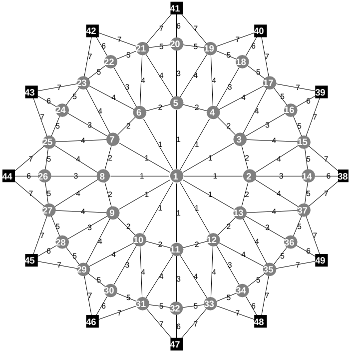
Outline

Examples: two finite element models

23 bar horizontal truss (bridge):



120 bar truss (dome):



$\mathcal{M}$: finite element model

$\boldsymbol{X}$: structural parameters and loads

Y : max deflection Δ

<u>Parameter of Δ of interest:</u>

- $\mathbb{P}(|\Delta| > \delta_{\mathrm{crit}})$

What is coming

We will:

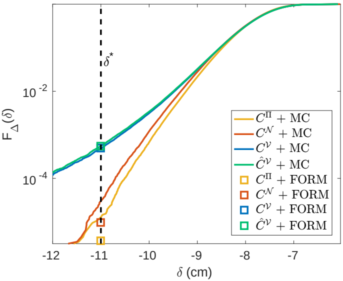
- Define the **true** joint PDF $f_{\boldsymbol{X}}$ of X (marginals + vine copula)
- Compute the **reference** solution by large MC simulation, or by FORM

Then, we will forget everything and

- Infer $f_{\boldsymbol{X}}$ from data (300 multivariate observations)
  In particular: Independence, Gaussian, or C-/D-vine copula
  $\hookrightarrow$Q1: does the copula matter?
- Estimate $\mathbb{P}(|\Delta| > \delta_{\text{crit}})$ by large MC simulation and by FORM for each copula
  $\hookrightarrow$Q2: does the "copula + UQ" framework work?

Reliability analysis by MCS

Bridge:



Copula Model	Method (# runs)	P_f $(\times 10^{-4})$
$C^\mathcal{V}$	MC (10^7)	4.84
$\hat{C}^\mathcal{V}$	MC (10^7)	5.52
$\hat{C}^\mathcal{N}$	MC (10^7)	0.31
C^Π	MC (10^7)	0.13

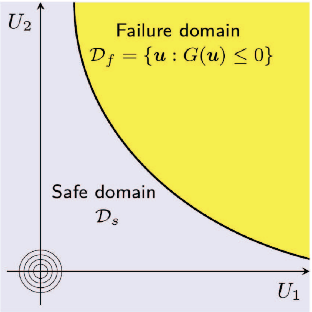
- The copula matters!

First Order Reliability Method (FORM)

Problem: system fails in unknown failure domain $\mathcal{D}_f$. Failure probability P_f?

- Assume $\mathcal{D}_f$ convex $\Rightarrow \exists!$ point $U^* \in \mathcal{D}_f$ closest to $\mathbf{0}$
- Assume $F_{\boldsymbol{X}} = \boldsymbol{\Phi}_d$, d-variate standard normal distribution

Then $P_f \approx 1 - F_{\boldsymbol{X}}(\boldsymbol{U}^*) = \boldsymbol{\Phi}_d(-||\boldsymbol{U}^*||)$.

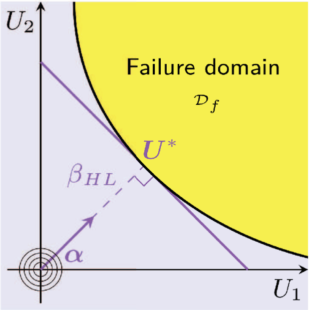


First Order Reliability Method (FORM)

Problem: system fails in unknown failure domain $\mathcal{D}_f$. Failure probability P_f?

- Assume $\mathcal{D}_f$ convex $\Rightarrow \exists!$ point $U^* \in \mathcal{D}_f$ closest to $\mathbf{0}$
- Assume $F_{\boldsymbol{X}} = \boldsymbol{\Phi}_d$, d-variate standard normal distribution

Then $P_f \approx 1 - F_{\boldsymbol{X}}(\boldsymbol{U}^*) = \boldsymbol{\Phi}_d(-||\boldsymbol{U}^*||)$. Task: find $\boldsymbol{U}^*$.

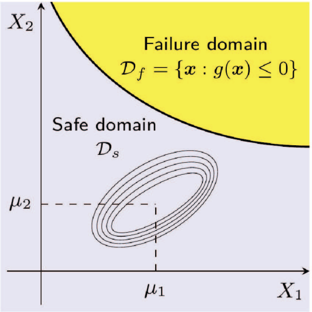


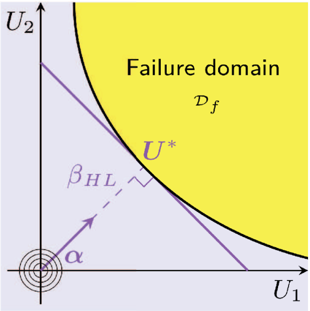
First Order Reliability Method (FORM)

Problem: system fails in unknown failure domain $\mathcal{D}_f$. Failure probability P_f?

- Assume $\mathcal{D}_f$ convex $\Rightarrow \exists!$ point $U^* \in \mathcal{D}_f$ closest to $\mathbf{0}$
- Assume $F_{\boldsymbol{X}} = \boldsymbol{\Phi}_d$, d-variate standard normal distribution

Then $P_f \approx 1 - F_{\boldsymbol{X}}(\boldsymbol{U}^*) = \boldsymbol{\Phi}_d(-||\boldsymbol{U}^*||)$. Task: find $\boldsymbol{U}^*$.

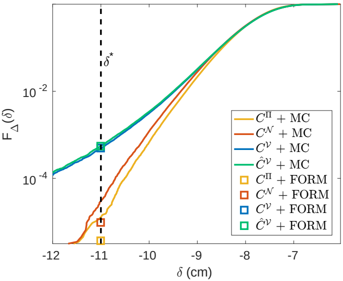




- If $f_{\boldsymbol{X}} \neq \boldsymbol{\Phi}_d$: First map $\boldsymbol{X} \mapsto \boldsymbol{X}' \sim \boldsymbol{\Phi}_d$

Reliability analysis: MC vs FORM results

Bridge:

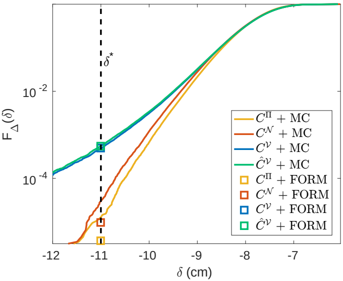


Copula Model	Method (# runs)	P_f $(\times 10^{-4})$
$C^{\mathcal{V}}$	MC (10^7)	4.84
$\hat{C}^{\mathcal{V}}$	MC (10^7)	5.52
$\hat{C}^{\mathcal{N}}$	MC (10^7)	0.31
C^{Π}	MC (10^7)	0.13
$C^{\mathcal{V}}$	FORM (108)	4.88
$\hat{C}^{\mathcal{V}}$	FORM (128)	5.44
$\hat{C}^{\mathcal{N}}$	FORM (219)	0.10
C^{Π}	FORM (219)	0.04

- The copula matters!
- FORM works well in combination with vine copulas

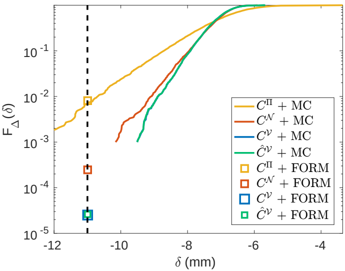
Reliability analysis: MC vs FORM results

Bridge:



Copula Model	Method (# runs)	P_f $(\times 10^{-4})$
$C^{\mathcal{V}}$	MC (10^7)	4.84
$\hat{C}^{\mathcal{V}}$	MC (10^7)	5.52
$\hat{C}^{\mathcal{N}}$	MC (10^7)	0.31
C^{Π}	MC (10^7)	0.13
$C^{\mathcal{V}}$	FORM (108)	4.88
$\hat{C}^{\mathcal{V}}$	FORM (128)	5.44
$\hat{C}^{\mathcal{N}}$	FORM (219)	0.10
C^{Π}	FORM (219)	0.04

Dome:



Copula Model	Method (# runs)	P_f $(\times 10^{-5})$
$C^{\mathcal{V}}$	MC (5000)	–
$\hat{C}^{\mathcal{V}}$	MC (5000)	–
$\hat{C}^{\mathcal{N}}$	MC (5000)	–
C^{Π}	MC (5000)	680
$C^{\mathcal{V}}$	FORM (1988)	2.53
$\hat{C}^{\mathcal{V}}$	FORM (3184)	2.54
$\hat{C}^{\mathcal{N}}$	FORM (1199)	25
C^{Π}	FORM (1199)	800

Conclusions

- Vine copulas are dependence models suitable to UQ problems:
  - $\hookrightarrow$ Enable use of advanced UQ methods when input $\boldsymbol{X}$ exhibits complex dependencies
  - $\hookrightarrow$ Work well in combination with inference
  - $\hookrightarrow$ Demonstrated applications in reliability analysis (FORM)
- Code soon available via UQlab:
  `www.uqlab.com`
- For more details:

ELSEVIER
Probabilistic Engineering Mechanics
Volume 55, January 2019, Pages 1-16

A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas

Emiliano Torre [a, b, c], Stefano Marelli [b], Paul Embrechts [c, a, d], Bruno Sudret [b, a]

Conclusions

- Vine copulas are dependence models suitable to UQ problems:
  - ↪ Enable use of advanced UQ methods when input $\boldsymbol{X}$ exhibits complex dependencies
  - ↪ Work well in combination with inference
  - ↪ Demonstrated applications in reliability analysis (FORM)
- Code soon available via UQlab: `www.uqlab.com`

THANK YOU!

UQLab

- For more details:

ELSEVIER

Probabilistic Engineering Mechanics
Volume 55, January 2019, Pages 1-16

A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas

Emiliano Torre [a, b, c], Stefano Marelli [b], Paul Embrechts [c, a, d], Bruno Sudret [b, a]

UQLab: facts and figures

- ≈ **2,000 users**
- **930^+ active users from 77 countries**

- Release of V0.9 on July 1st, 2015 (beta version)
- V1.0 on April 28th, 2017
  UQLabCore + modules (PCE, Kriging, Sensitivity, Rare events)
- V1.1 on July 1st, 2018
  Support vector machines, UQLink
- V1.2 on February 22nd, 2019
  Bayesian inversion, UQLib

Country	**# Users**
United States	354
China	224
France	223
Switzerland	180
Germany	134
United Kingdom	92
Italy	73
India	57
Canada	53
Brazil	51

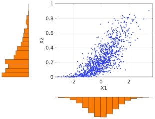
As of May 27, 2019

www.uqlab.com

UQLab: The Uncertainty Quantification Software

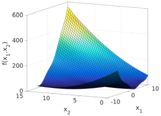
Probabilistic Input Modelling

- Common marginals
- Support for user-defined marginals
- Support for bounds on all distributions (including user-defined)
- Gaussian copula



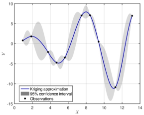
Modelling facilities

- Simple text strings
- Matlab m-files
- Matlab handles
- UQLink: easily connect UQLab to third party modelling software



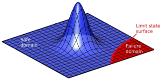
Advanced metamodelling

- Sparse degree-adaptive Polynomial Chaos Expansions
- Gaussian process modelling (Kriging)
- Polynomial-Chaos Kriging
- Low-rank tensor approximations
- Support vector machines



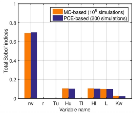
Reliability Analysis (Rare Event Estimation)

- FORM/SORM approximation
- Monte Carlo Simulation (MCS)
- Importance Sampling
- Subset Simulation
- Adaptive Kriging (AK-MCS)



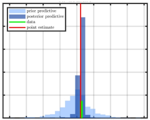
Sensitivity Analysis

- Correlation-based indices
- Standard Regression Coefficients
- Cotter measure
- Morris indices
- Sampling-based Sobol' indices
- PCE- and LRA-based Sobol' indices
- Borgonovo δ indices
- Support for dependent inputs

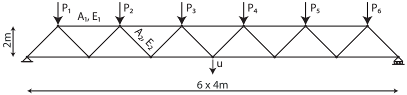


Bayesian inversion

- Intuitive problem statement
- Advanced MCMC algorithms
- Multi-model support (joint inversion)
- Support for custom likelihood



UQLab example: Inference of Truss input model



$$\boldsymbol{X} = (E_1, E_2, A_1, A_2, P_1, \ldots, P_6)$$

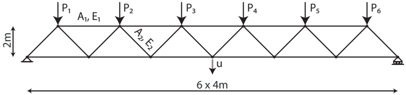
```
for ii = 1:2       % Assign uncertain Young moduli
    iOpts.Marginals(ii).Type = 'LogNormal';
    iOpts.Marginals(ii).Moments = [2.1e11, 2.1e10];
end
for ii = 3:4       % Assign uncertain cross-sections
 iOpts.Marginals(ii).Inference.Data = A(:,ii-2);
end
for ii = 5:10      % Assign uncertain loads
    iOpts.Marginals(ii).Inference.Data = L(:,ii-4);
end

% Define the input copula: product of Independent(dim:4) copula
% and D-Vine copula inferred from data matrix L
iOpts.Copula(1).Type = 'Independent';
iOpts.Copula(1).Parameters = eye(4);

iOpts.Copula(2).Type = 'DVine';
iOpts.Copula(2).Inference.Data = L;

% Create Input
myInput = uq_createInput(iOpts);
```

UQLab example: Inference of Truss input model



$$\boldsymbol{X} = (E_1, E_2, A_1, A_2, P_1, \ldots, P_6)$$

```
% Create computational model
mOpts.mHandle = uq_truss_model;  % Defined in external m-file
myModel = uq_createModel(mOpts);

% Create probabilistic input model
myInput = ... % see previous slide

% Perform FORM analysis
aOpts.Type = 'Reliability';
aOpts.Method = 'FORM';
myAnalysis = uq_createAnalysis(aOpts);
```

Polynomial Chaos Expansion (PCE)

PCE model of Y to $\boldsymbol{X}$: $$Y_{\rm PC} = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} y_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}),$$

where $\Psi_{\boldsymbol{\alpha}}$ are multivariate polynomials orthonormal wrt $f_{\boldsymbol{X}}$.

Theorem

If $\mathbb{V}(Y) < \infty$ *and* $\mathbb{E}(X^k) < \infty \ \forall k \geq 1$, *then* $\exists \{y_{\boldsymbol{\alpha}}, \Psi_{\boldsymbol{\alpha}}\} : Y_{PC} \xrightarrow{L^2} Y = \mathcal{M}(X)$.[1]

- Properties of coefficients $y_{\boldsymbol{\alpha}}$: (A) $\mathbb{E}(Y_{\rm PC}) = y_{\boldsymbol{0}}$, (B) $\mathbb{V}(Y_{\rm PC}) = \sum_{\boldsymbol{\alpha} \neq \boldsymbol{0}} y_{\boldsymbol{\alpha}}^2$
- (B) guarantees compressibility $\Rightarrow$ parsimonious parametric model

  $\hookrightarrow$ for large d: sparse regression (e.g. least angle regression)
- Provides good estimates of "global" statistics of Y: $\mathbb{E}(Y), \mathbb{V}(Y)$, f_Y ...

How to build the orthonormal basis $\{\Psi_{\alpha}, \boldsymbol{\alpha} \in \mathbb{N}^d\}$?

- If $f_{\boldsymbol{X}} = \prod_{i=1}^d f_{X_i}$: $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) := \prod_{i=1}^d \phi_{\alpha_i}^{(i)}(x_i)$, where $\phi_{\alpha_i}^{(i)}$ are univariate polynomials orthonormal wrt f_{X_i}.
- Otherwise: Rosenblatt transform $\boldsymbol{X} \mapsto \boldsymbol{X}'$, followed by PCE $Y_{\rm PC}(\boldsymbol{X}')$

[1] Ernst, Mugler, Starkloff, Ullmann (2012), ESAIM:M2AN, 46(2):317-339

Polynomial Chaos Expansion (PCE)

PCE model of Y to $\boldsymbol{X}$: $$Y_{\mathrm{PC}} = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d} y_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}),$$

where $\Psi_{\boldsymbol{\alpha}}$ are multivariate polynomials orthonormal wrt $f_{\boldsymbol{X}}$.

Theorem

If $\mathbb{V}(Y) < \infty$ *and* $\mathbb{E}(X^k) < \infty\ \forall k \geq 1$, *then* $\exists\{y_{\boldsymbol{\alpha}}, \Psi_{\boldsymbol{\alpha}}\} : Y_{PC} \xrightarrow{L^2} Y = \mathcal{M}(X)$.[1]

- Properties of coefficients $y_{\boldsymbol{\alpha}}$: (A) $\mathbb{E}(Y_{\mathrm{PC}}) = y_{\mathbf{0}}$, (B) $\mathbb{V}(Y_{\mathrm{PC}}) = \sum_{\boldsymbol{\alpha}\neq\mathbf{0}} y_{\boldsymbol{\alpha}}^2$
- (B) guarantees compressibility $\Rightarrow$ parsimonious parametric model

  $\hookrightarrow$ for large d: sparse regression (e.g. least angle regression)
- Provides good estimates of "global" statistics of Y: $\mathbb{E}(Y), \mathbb{V}(Y)$, f_Y ...

How to build the orthonormal basis $\{\Psi_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathbb{N}^d\}$?

- If $f_{\boldsymbol{X}} = \prod_{i=1}^d f_{X_i}$: $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) := \prod_{i=1}^d \phi_{\alpha_i}^{(i)}(x_i)$, where $\phi_{\alpha_i}^{(i)}$ are univariate polynomials orthonormal wrt f_{X_i}.
- Otherwise: Rosenblatt transform $\boldsymbol{X} \mapsto \boldsymbol{X}'$, followed by PCE $Y_{\mathrm{PC}}(\boldsymbol{X}')$

[1] Ernst, Mugler, Starkloff, Ullmann (2012), ESAIM:M2AN, 46(2):317-339

Polynomial Chaos Expansion (PCE)

PCE model of Y to $\boldsymbol{X}$: $$Y_{\mathrm{PC}} = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} y_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}),$$

where $\Psi_{\boldsymbol{\alpha}}$ are multivariate polynomials orthonormal wrt $f_{\boldsymbol{X}}$.

Theorem

If $\mathbb{V}(Y) < \infty$ *and* $\mathbb{E}(X^k) < \infty \ \forall k \geq 1$, *then* $\exists \{y_{\boldsymbol{\alpha}}, \Psi_{\boldsymbol{\alpha}}\} : Y_{PC} \overset{L^2}{\to} Y = \mathcal{M}(X)$.[1]

- Properties of coefficients $y_{\boldsymbol{\alpha}}$: (A) $\mathbb{E}(Y_{\mathrm{PC}}) = y_{\mathbf{0}}$, (B) $\mathbb{V}(Y_{\mathrm{PC}}) = \sum_{\boldsymbol{\alpha} \neq \mathbf{0}} y_{\boldsymbol{\alpha}}^2$
- (B) guarantees compressibility $\Rightarrow$ parsimonious parametric model
  $\hookrightarrow$ for large d: sparse regression (e.g. least angle regression)
- Provides good estimates of "global" statistics of Y: $\mathbb{E}(Y), \mathbb{V}(Y)$, f_Y ...

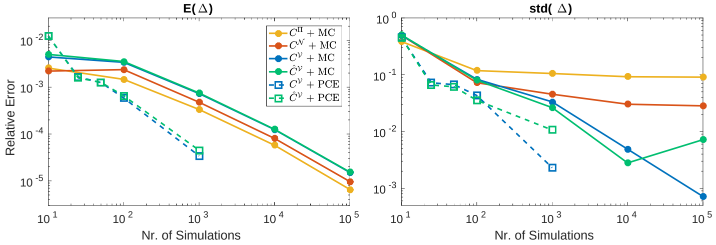
How to build the orthonormal basis $\{\Psi_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathbb{N}^d\}$?

- If $f_{\boldsymbol{X}} = \prod_{i=1}^d f_{X_i}$: $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) := \prod_{i=1}^d \phi_{\alpha_i}^{(i)}(x_i)$, where $\phi_{\alpha_i}^{(i)}$ are univariate polynomials orthonormal wrt f_{X_i}.
- Otherwise: Rosenblatt transform $\boldsymbol{X} \mapsto \boldsymbol{X}'$, followed by PCE $Y_{\mathrm{PC}}(\boldsymbol{X}')$

[1] Ernst, Mugler, Starkloff, Ullmann (2012), ESAIM:M2AN, 46(2):317-339

Vine representations + PCE: errors on $\mathbb{E}(\Delta)$ and $\mathrm{std}(\Delta)$

Bridge (uncertain inputs: 4 structural parameters + 6 loads)



Dome (uncertain inputs: 7 structural parameters + 9 loads)