Rare event estimation using nonparametric Bernstein adaptive sampling

MASCOT-NUM 2023 - April 4, 2023

Elias Fekhari[1,2]

PhD supervisors: Bertrand Iooss[1,2], Vincent Chabridon[1], Joseph Muré[1]
PhD duration: 01/2021 - 01/2024

[1] EDF R&D - 6 quai Watier, Chatou, France
[2] Université Côte d'Azur - 28 Avenue de Valrose, Nice, France

Context and industrial motivations

Rare event estimation

Bernstein Adaptive Nonparametric Conditional Sampling (BANCS)

Offshore wind turbine application

Conclusions and limits

Industrial context

- EDF is a major actor of the offshore wind turbine development
- Take strategic decisions in uncertain conditions
  (e.g., chose a floater design, extend a wind farm operating time)
- EDF R&D participates to HIPERWIND[1] (EU research project)

HIPERWIND · IFP Energies nouvelles · DNV · ETH zürich · DTU

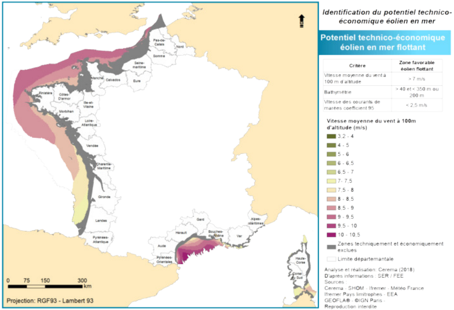


Figure 1: French floating wind energy potential (source: CEREMA).

[1] https://www.hiperwind.eu/

UQ on offshore wind turbine simulator

Environmental conditions data:

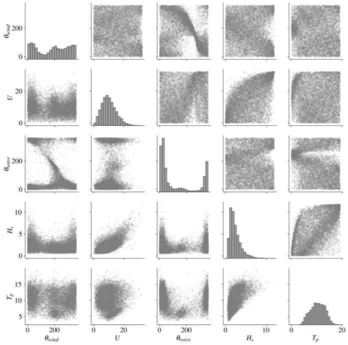


Figure 2: Copulogram[2] of the South Brittany environmental data ($N = 10^4$).

Numerical simulation model:

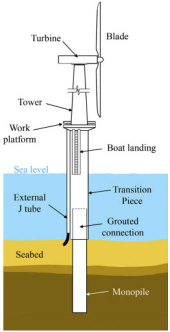


Figure 3: Monopile OWT diagram (source: Chen et al. 2018).

Variable of interest: Fatigue damage on a wind turbine (seabed level)

[2] https://github.com/efekhari27/copulogram

UQ on offshore wind turbine simulator

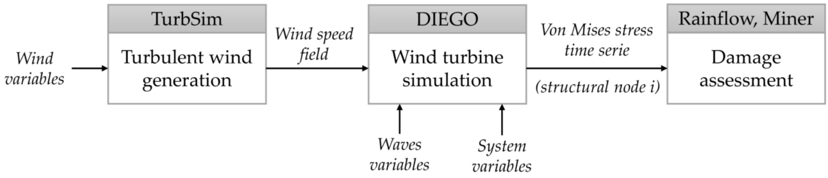


Figure 4: Diagram of the chained wind turbine simulation model[3].

Scientific challenges:

- Costly numerical models deployed on high performance computers facility (simulation CPU time: $\sim$15 min)
- Stochastic wind generation treated with repetitions
- Given-data uncertainty propagation with a complex dependency
- Rare event estimation with dedicated sensitivity analysis

[3] Kim et al. 2022.

Previous contributions

C1 Treatment of environment data

- ▷ Nonparametric uncertainty quantification[4]
  (joint work with DNV and DTU)
- ▷ Quantifying wake-induced perturbations within a wind farm[5]
  (joint work with IFPEN)

[4] Vanem et al. 2023.
[5] Lovera et al. 2023.
[6] https://efekhari27.github.io/otkerneldesign/master/index.html
[7] Fekhari, Iooss, et al. 2023.
[8] Fekhari, Chabridon, et al. 2023b.
[9] Fekhari, Chabridon, et al. 2023a.

Previous contributions

C1 Treatment of environment data

- ▷ Nonparametric uncertainty quantification[4]
  (joint work with DNV and DTU)
- ▷ Quantifying wake-induced perturbations within a wind farm[5]
  (joint work with IFPEN)

C2 Uncertainty propagation using Bayesian quadrature[6]

- ▷ Sequential metamodel validation[7]
  (joint work with L.Pronzato and M.J. Rendas)
- ▷ Given-data central tendency estimation[8]

---

[4] Vanem et al. 2023.
[5] Lovera et al. 2023.
[6] `https://efekhari27.github.io/otkerneldesign/master/index.html`
[7] Fekhari, Iooss, et al. 2023.
[8] Fekhari, Chabridon, et al. 2023b.
[9] Fekhari, Chabridon, et al. 2023a.

Previous contributions

C1 Treatment of environment data

- Nonparametric uncertainty quantification[4]
  (joint work with DNV and DTU)
- Quantifying wake-induced perturbations within a wind farm[5]
  (joint work with IFPEN)

C2 Uncertainty propagation using Bayesian quadrature[6]

- Sequential metamodel validation[7]
  (joint work with L.Pronzato and M.J. Rendas)
- Given-data central tendency estimation[8]

C3 Rare event estimation using nonparametric Bernstein adaptive sampling[9]

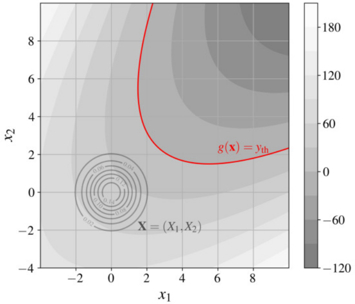
---

[4] Vanem et al. 2023.
[5] Lovera et al. 2023.
[6] `https://efekhari27.github.io/otkerneldesign/master/index.html`
[7] Fekhari, Iooss, et al. 2023.
[8] Fekhari, Chabridon, et al. 2023b.
[9] Fekhari, Chabridon, et al. 2023a.

Rare event problem:

Propagate a random input vector **X** through a deterministic limit-state function $g : \mathbb{R}^d \mapsto \mathbb{R}$ and estimate the following failure probability:

$$p_{\mathrm{f}} := \mathbb{P}(g(\mathbf{X}) \leq y_{\mathrm{th}}) = \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \,\mathrm{d}\mathbf{x}, \quad (y_{\mathrm{th}} \in \mathbb{R}, \mathcal{F}_{\mathbf{x}} \subset \mathbb{R}^d) \quad (1)$$



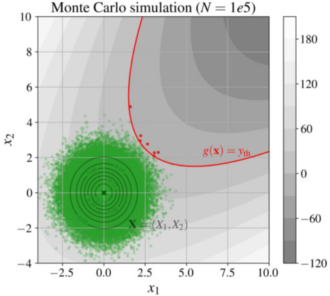
[10] Morio and Balesdent 2015.

Rare event problem:

Propagate a random input vector **X** through a deterministic limit-state function $g : \mathbb{R}^d \mapsto \mathbb{R}$ and estimate the following failure probability:

$$p_f := \mathbb{P}(g(\mathbf{X}) \leq y_{th}) = \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{F}_\mathbf{x}}(\mathbf{x}) f_\mathbf{X}(\mathbf{x})\, d\mathbf{x}, \quad (y_{th} \in \mathbb{R}, \mathcal{F}_\mathbf{x} \subset \mathbb{R}^d) \quad (1)$$

Rare event estimation methods[10]**:**



$\{\mathbf{X}^{(j)}\}_{j=1}^N \overset{\text{i.i.d}}{\sim} f_\mathbf{X}$

$$\widehat{p}_f^{\,MC} = \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\mathcal{F}_\mathbf{x}}(\mathbf{X}^{(j)})$$

[10] Morio and Balesdent 2015.

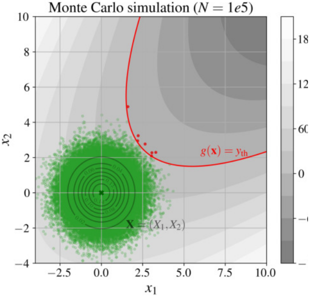
Brief review of the methods (1/2)

Rare event problem:

Propagate a random input vector $\mathbf{X}$ through a deterministic limit-state function $g : \mathbb{R}^d \mapsto \mathbb{R}$ and estimate the following failure probability:

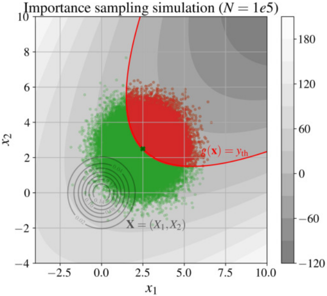
$$p_{\mathrm{f}} := \mathbb{P}(g(\mathbf{X}) \leq y_{\mathrm{th}}) = \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x}, \quad (y_{\mathrm{th}} \in \mathbb{R}, \mathcal{F}_{\mathbf{x}} \subset \mathbb{R}^d) \quad (1)$$

Rare event estimation methods[10]:



$\{\mathbf{X}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$

$$\widehat{p}_{\mathrm{f}}^{\,\mathrm{MC}} = \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}^{(j)})$$



$\{\mathbf{X}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} h_{\mathbf{X}}$

$$\widehat{p}_{\mathrm{f}}^{\,\mathrm{IS}} = \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}^{(j)}) \frac{f_{\mathbf{X}}(\mathbf{X}^{(j)})}{h_{\mathbf{X}}(\mathbf{X}^{(j)})}$$

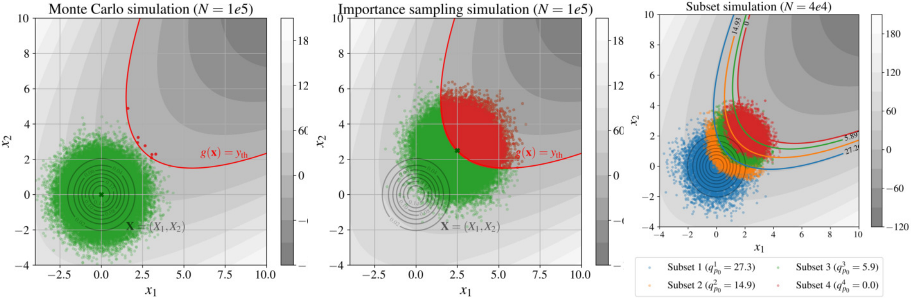
[10] Morio and Balesdent 2015.

Rare event problem:

Propagate a random input vector $\mathbf{X}$ through a deterministic limit-state function $g : \mathbb{R}^d \mapsto \mathbb{R}$ and estimate the following failure probability:

$$p_{\mathrm{f}} := \mathbb{P}(g(\mathbf{X}) \leq y_{\mathrm{th}}) = \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x}, \quad (y_{\mathrm{th}} \in \mathbb{R}, \mathcal{F}_{\mathbf{x}} \subset \mathbb{R}^d) \quad (1)$$

Rare event estimation methods[10]:



$\{\mathbf{X}^{(j)}\}_{j=1}^N \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ | $\{\mathbf{X}^{(j)}\}_{j=1}^N \overset{\text{i.i.d}}{\sim} h_{\mathbf{X}}$ | $\{\mathbf{X}^{(j)}_{[k]}\}_{j=1}^N \overset{\text{i.d}}{\sim} f_{\mathbf{X}}(\mathbf{x}|\mathcal{F}_{[k-1]})$

$$\widehat{p_{\mathrm{f}}}^{\mathrm{MC}} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}^{(j)})$$

$$\widehat{p_{\mathrm{f}}}^{\mathrm{IS}} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}^{(j)}) \frac{f_{\mathbf{X}}(\mathbf{X}^{(j)})}{h_{\mathbf{X}}(\mathbf{X}^{(j)})}$$

$$\widehat{p_{\mathrm{f}}}^{\mathrm{SS}} = p_0^{k-1} \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}^{(j)}_{[k]})$$

[10] Morio and Balesdent 2015.

Brief review of the methods (2/2)

Adaptive rare event estimation methods:

- Subset simulation[11,12,13] (SS):
  - ▷ using MCMC: dependent samples (limit for dedicated sensitivity analysis), MCMC convergence diagnostics
- Adaptive importance sampling
  - ▷ using cross-entropy optimization[14]: parametric, considers only one failure domain
  - ▷ using kernel density estimation[15] (NAIS): nonparametric, degenerates in high dimension

Alternative idea:

Plug a nonparametric copula estimator with an adaptive rare event algorithm to properly capture the dependence structure

---

[11] Au and Beck 2001.

[12] Cérou et al. 2012.

[13] Papaioannou et al. 2015.

[14] Kurtz and Song 2013.

[15] Morio 2011.

Empirical Bernstein Copula (1/4)

Multivariate modeling using copulas[16] (Sklar theorem):
Considering a random vector $\mathbf{X} \in \mathbb{R}^d$, with its distribution F and its marginals $\{F_i\}_{i=1}^d$, there exists a copula $C : [0,1]^d \to [0,1]$, such that:

$$F(x_1, \dots, x_d) = C\left(F_1(x_1), \dots, F_p(x_d)\right). \quad (2)$$

- When the joint distribution is continuous, this copula is unique
- One can divide the multivariate fitting problem into two independent problems: fitting marginals and fitting the copula

Copula estimation:

- Parametric: vines copula[17] (delicate choice of parametric family)
- Nonparametric: empirical Bernstein copula[18], B-splines copula[19]

---

[16] Joe 1997.
[17] Joe and Kurowicka 2011.
[18] Sancetta and Satchell 2004.
[19] Nagler, Schellhase, and Czado 2017.

Empirical Bernstein Copula (2/4)

Bernstein polynomial basis (degree *m*)

$$b_t^m(u) := \binom{m}{t} u^t (1-u)^{m-t} \qquad (3)$$

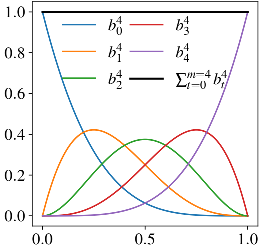


Figure 5: Bernstein polynomial basis of 4th degree.

Empirical Bernstein Copula (2/4)

Bernstein polynomial basis (degree m)

$$b_t^m(u) := \binom{m}{t} u^t (1-u)^{m-t} \qquad (3)$$

Bernstein 1D approx. ($\forall C : \mathbb{R} \mapsto \mathbb{R}$)

$$B_m(C)(u) := \sum_{t=0}^{m} C\left(\frac{t}{m}\right) b_t^m(u) \qquad (4)$$

- $\lim_{m\to\infty} B_m(C) = C$ uniformly on $[0,1]$
- Bezier curves are a weighted version

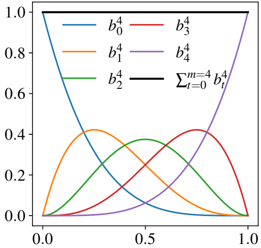


Figure 5: Bernstein polynomial basis of 4th degree.

Bernstein polynomial basis (degree m)

$$b_t^m(u) := \binom{m}{t} u^t (1-u)^{m-t} \qquad (3)$$

Bernstein 1D approx. ($\forall C : \mathbb{R} \mapsto \mathbb{R}$)

$$B_m(C)(u) := \sum_{t=0}^{m} C\left(\frac{t}{m}\right) b_t^m(u) \qquad (4)$$

- $\lim_{m\to\infty} B_m(C) = C$ uniformly on $[0,1]$
- Bezier curves are a weighted version

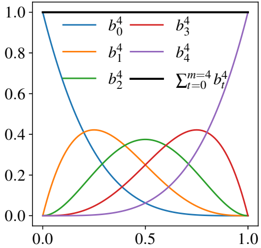


Figure 5: Bernstein polynomial basis of 4th degree.

Bernstein multivariate approx. ($\forall C : \mathbb{R}^d \mapsto \mathbb{R}$)

$$B_{\mathbf{m}}(C)(\mathbf{u}) := \sum_{t_1=0}^{m_1} \cdots \sum_{t_d=0}^{m_d} C\left(\frac{t_1}{m_1}, \ldots, \frac{t_d}{m_d}\right) \prod_{j=1}^{d} b_{t_j}^{m_j}(u_j) \qquad (5)$$

Empirical Bernstein copula:

The empirical Bernstein copula[20] (EBC) is a the Bernstein approximation of the empirical copula C_n (ranked data with size n)

Properties:

- $B_{\mathbf{m}}(C_n)(\mathbf{u}) \to C(\mathbf{u}), \quad \forall u_j \in]0,1[$ if $\frac{m^{d/2}}{n} \to 0$, when $m, n \to \infty$
- When $m \nearrow$, the bias $\searrow$ and the variance $\nearrow$
- Asymptotic optimal tuning minimizing Eq. (6): $m_{\mathrm{AMISE}} = \lceil n^{2/(d+4)} \rceil$

$$\mathbb{E}\left[\| B_{\mathbf{m}}(C_n) - C \|_2^2 \right] \tag{6}$$

- Beta[21] tuning: $m_{beta} = n$

[20] Sancetta and Satchell 2004.

[21] Segers, Sibuya, and Tsukahara 2017.

South Brittany data

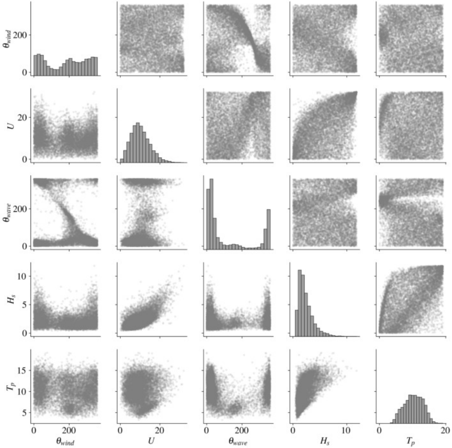


Figure 6: Copulogram[22] of the South Brittany environmental data ($N = 10^4$).

Simulated data on EBC ($m_{AMISE} = 8$)

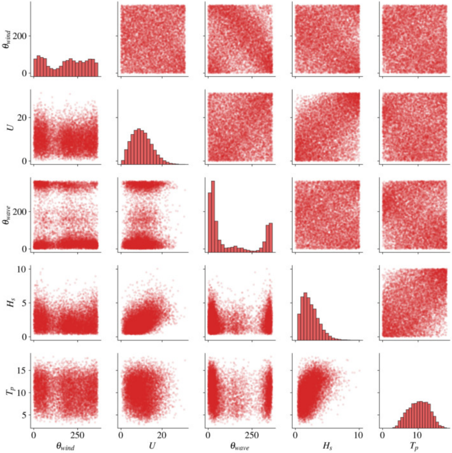


Figure 7: Copulogram of simulated data on noparametric model (copula fitted by EBC and marginals by KDE) ($n = 10^4$).

[22] `https://github.com/efekhari27/copulogram`

South Brittany data

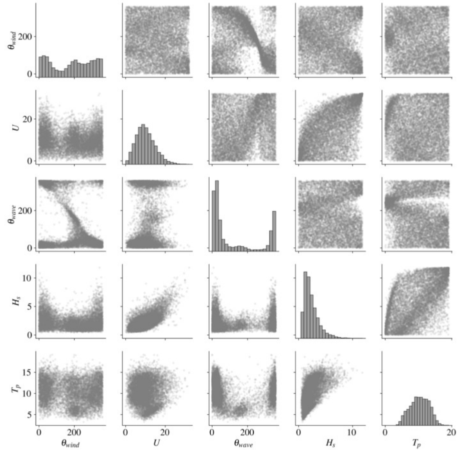


Figure 6: Copulogram[22] of the South Brittany environmental data ($N = 10^4$).

Simulated data on EBC ($m = 40$)

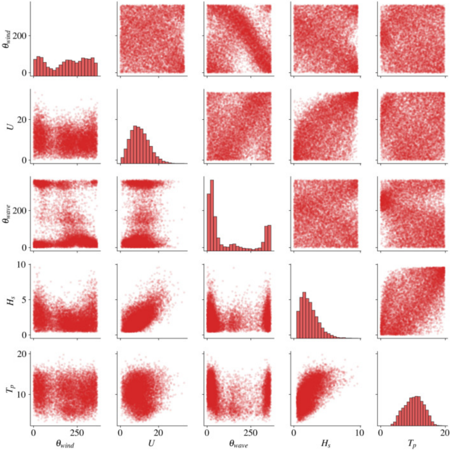


Figure 7: Copulogram of simulated data on noparametric model (copula fitted by EBC and marginals by KDE) ($n = 10^4$).

[22] https://github.com/efekhari27/copulogram

Empirical Bernstein Copula (4/4)

South Brittany data

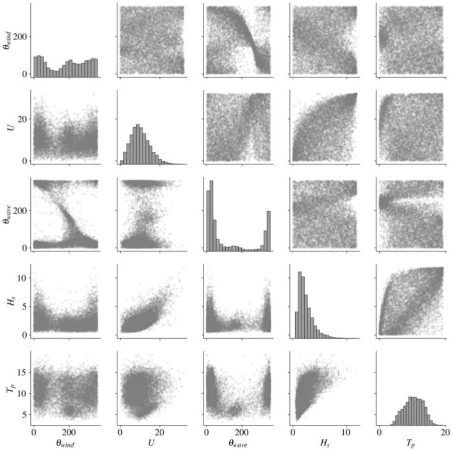


Figure 6: Copulogram[22] of the South Brittany environmental data ($N = 10^4$).

Simulated data on EBC ($m = 100$)

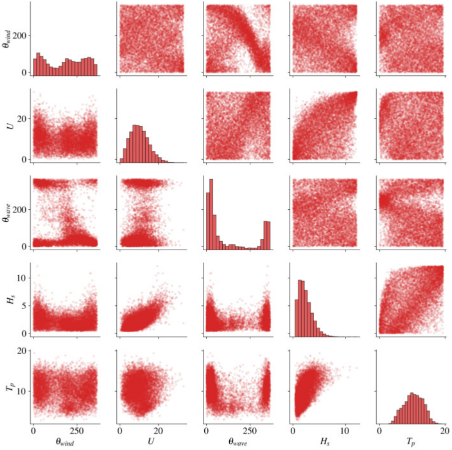


Figure 7: Copulogram of simulated data on noparametric model (copula fitted by EBC and marginals by KDE) ($n = 10^4$).

[22] https://github.com/efekhari27/copulogram

Empirical Bernstein Copula (4/4)

South Brittany data

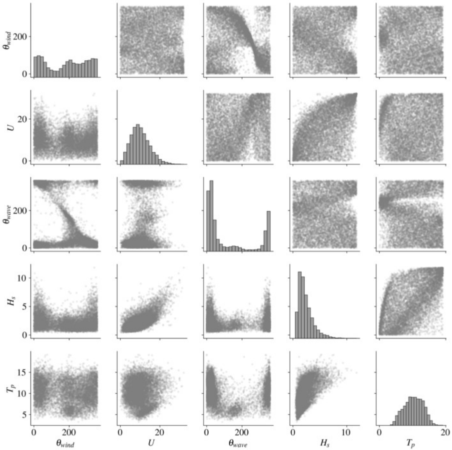


Figure 6: Copulogram[22] of the South Brittany environmental data ($N = 10^4$).

Simulated data on EBC ($m_{beta} = 10^4$)

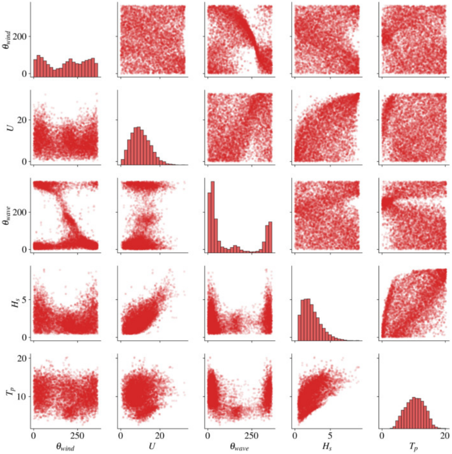


Figure 7: Copulogram of simulated data on noparametric model (copula fitted by EBC and marginals by KDE) ($n = 10^4$).

---

[22] `https://github.com/efekhari27/copulogram`

BANCS: algorithm for rare event estimation

Splitting a failure domain $\mathcal{F}_{\mathbf{x}}$ into nested subsets $\mathcal{F}_{[1]} \supset \ldots \supset \mathcal{F}_{[k_\#]} = \mathcal{F}_{\mathbf{x}}$

$$p_{\mathrm{f}} = \mathbb{P}(\mathcal{F}_{\mathbf{x}}) = \mathbb{P}(\cap_{k=1}^{k_\#} \mathcal{F}_{[k]}) = \prod_{k=1}^{k_\#} \mathbb{P}(\mathcal{F}_{[k]} | \mathcal{F}_{[k-1]}) \tag{7}$$

BANCS: algorithm for rare event estimation

Algorithm 1: Subset simulation

N, number of samples per iteration
$m \in \mathbb{N}$, parameter of the EBC fitting
$p_0 \in]0, 1[$, empirical quantile order (rarity parameter)
Set $k = 0$ and $f_{[0]} = f_{\mathbf{X}}$
Sample $\mathbf{X}_{[0],N} = \{\mathbf{X}_{[0]}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} f_{[0]}$
Evaluate $G_{[0],N} = \{g(\mathbf{X}_{[0]}^{(j)})\}_{j=1}^{N}$
Estimate the empirical p_0-quantile $\widehat{q}_{[0]}^{p_0}$ of the set $G_{[0],N}$
while $\widehat{q}_{[k]}^{p_0} > y_{\mathrm{th}}$ **do**

- Subset $\mathbf{A}_{[k+1],n} = \{\mathbf{X}_{[k]}^{(j)} \subset \mathbf{X}_{[k],N} | g(\mathbf{X}_{[k]}^{(j)}) > \widehat{q}_{[k]}^{p_0}\}_{j=1}^{n}$
- Sample by MCMC $\mathbf{X}_{[k+1],N} = \{\mathbf{X}_{[k+1]}^{(j)}\}_{j=1}^{N} \overset{\text{i.d}}{\sim} f_{\mathbf{X}|F_{[k+1]}}$
  (with $\mathbf{A}_{[k+1],n}$ as initialization points)
- Evaluate $G_{[k+1],N} = \{g(\mathbf{X}_{[k+1]}^{(j)})\}_{j=1}^{N}$
- Estimate the empirical p_0-quantile $\widehat{q}_{[k+1]}^{p_0}$ of $G_{[k+1],N}$
- Set $k = k + 1$

Set total iteration number $k_{\#} = k - 1$
Estimate $\widehat{p}_{\mathrm{f}} = (1-p_0)^{k_{\#}} \cdot \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\{g(\mathbf{X}_{[k_{\#}]}^{(j)}) \geq y_{\mathrm{th}}\}}(\mathbf{X}_{[k_{\#}]}(j))$
$\widehat{p}_{\mathrm{f}}$, estimate of p_{f}

BANCS: algorithm for rare event estimation

Algorithm 2: Subset simulation

N, number of samples per iteration
$m \in \mathbb{N}$, parameter of the EBC fitting
$p_0 \in]0, 1[$, empirical quantile order (rarity parameter)
Set $k = 0$ and $f_{[0]} = f_{\mathbf{X}}$

Sample $\mathbf{X}_{[0],N} = \{\mathbf{X}_{[0]}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} f_{[0]}$

Evaluate $G_{[0],N} = \{g(\mathbf{X}_{[0]}^{(j)})\}_{j=1}^{N}$

Estimate the empirical p_0-quantile $\widehat{q}_{[0]}^{p_0}$ of the set $G_{[0],N}$

while $\widehat{q}_{[k]}^{p_0} > y_{\text{th}}$ **do**

- Subset $\mathbf{A}_{[k+1],n} = \{\mathbf{X}_{[k]}^{(j)} \subset \mathbf{X}_{[k],N} | g(\mathbf{X}_{[k]}^{(j)}) > \widehat{q}_{[k]}^{p_0}\}_{j=1}^{n}$
- Sample by MCMC $\mathbf{X}_{[k+1],N} = \{\mathbf{X}_{[k+1]}^{(j)}\}_{j=1}^{N} \overset{\text{i.d}}{\sim} f_{\mathbf{X}|F_{[k+1]}}$ (with $\mathbf{A}_{[k+1],n}$ as initialization points)
- Evaluate $G_{[k+1],N} = \{g(\mathbf{X}_{[k+1]}^{(j)})\}_{j=1}^{N}$
- Estimate the empirical p_0-quantile $\widehat{q}_{[k+1]}^{p_0}$ of $G_{[k+1],N}$
- Set $k = k + 1$

Set total iteration number $k_{\#} = k - 1$

Estimate $\widehat{p}_{\text{f}} = (1 - p_0)^{k_{\#}} \cdot \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\{g(\mathbf{X}_{[k_{\#}]}^{(j)}) \geq y_{\text{th}}\}}(\mathbf{X}_{[k_{\#}]^{(j)}})$

$\widehat{p}_{\text{f}}$, estimate of p_{f}

Algorithm 3: BANCS

N, number of samples per iteration
$m \in \mathbb{N}$, parameter of the EBC fitting
$p_0 \in]0, 1[$, empirical quantile order (rarity parameter)
Set $k = 0$ and $f_{[0]} = f_{\mathbf{X}}$

Sample $\mathbf{X}_{[0],N} = \{\mathbf{X}_{[0]}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} f_{[0]}$

Evaluate $G_{[0],N} = \{g(\mathbf{X}_{[0]}^{(j)})\}_{j=1}^{N}$

Estimate the empirical p_0-quantile $\widehat{q}_{[0]}^{p_0}$ of the set $G_{[0],N}$

while $\widehat{q}_{[k]}^{p_0} > y_{\text{th}}$ **do**

- Subset $\mathbf{A}_{[k+1],n} = \{\mathbf{X}_{[k]}^{(j)} \subset \mathbf{X}_{[k],N} | g(\mathbf{X}_{[k]}^{(j)}) > \widehat{q}_{[k]}^{p_0}\}_{j=1}^{n}$
- Fit marginals of the subset $\mathbf{A}_{[k+1],n}$ by KDE $\{\widehat{F}_i\}_{i=1}^{d}$
- Fit the copula of the subset $\mathbf{A}_{[k+1],n}$ by EBC $B_{\mathbf{m}}(C_n)$
- Build a CDF $\widehat{F}_{[k+1]}(\mathbf{x}) = B_{\mathbf{m}}(C_n)(\widehat{F}_1(x_1), \dots, \widehat{F}_d(x_d))$
- Sample $\mathbf{X}_{[k+1],N} = \{\mathbf{X}_{[k+1]}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} \widehat{f}_{[k+1]}$
- Evaluate $G_{[k+1],N} = \{g(\mathbf{X}_{[k+1]}^{(j)})\}_{j=1}^{N}$
- Estimate the empirical p_0-quantile $\widehat{q}_{[k+1]}^{p_0}$ of $G_{[k+1],N}$
- Set $k = k + 1$

Set total iteration number $k_{\#} = k - 1$

Estimate $\widehat{p}_{\text{f}} = (1 - p_0)^{k_{\#}} \cdot \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\{g(\mathbf{X}_{[k_{\#}]}^{(j)}) \geq y_{\text{th}}\}}(\mathbf{X}_{[k_{\#}]^{(j)}})$

$\widehat{p}_{\text{f}}$, estimate of p_{f}

Numerical results on toy-cases

Parabolic toy-case:

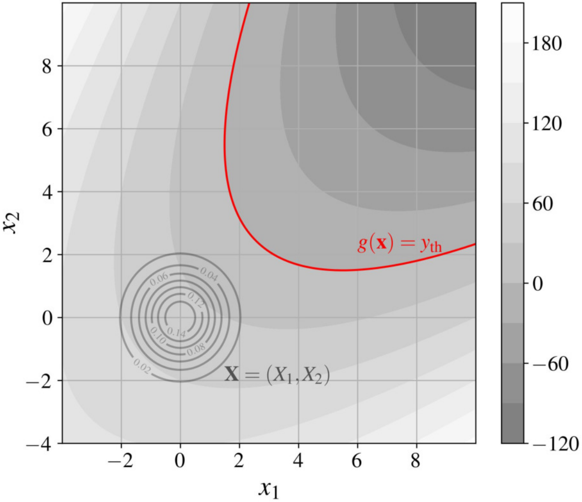


Figure 8: BANCS sampling steps - illustration of the iterations on a parabolic case.

Numerical results on toy-cases

Parabolic toy-case:

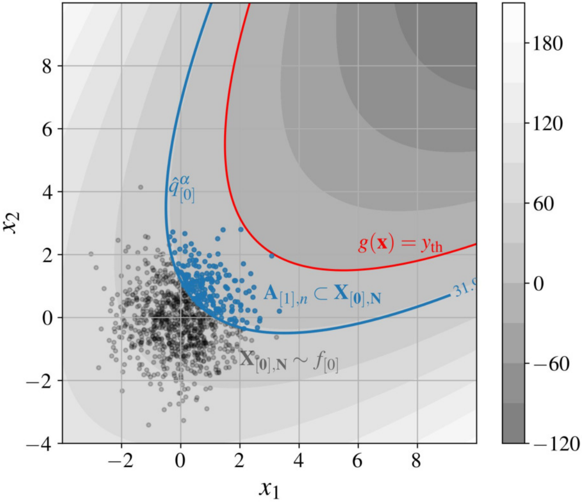


Figure 8: BANCS sampling steps - illustration of the iterations on a parabolic case.

Parabolic toy-case:

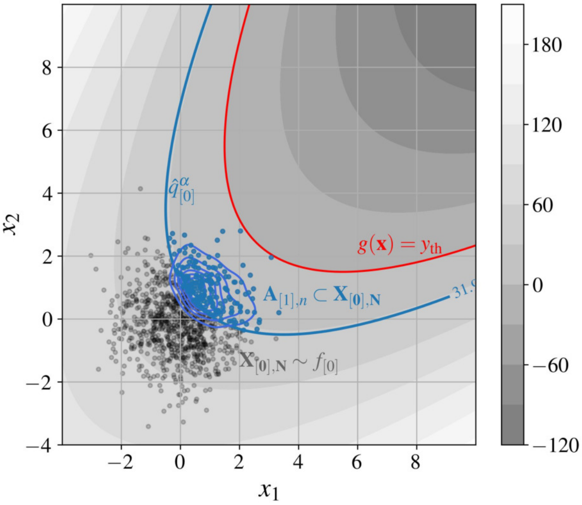


Figure 8: BANCS sampling steps - illustration of the iterations on a parabolic case.

Parabolic toy-case:

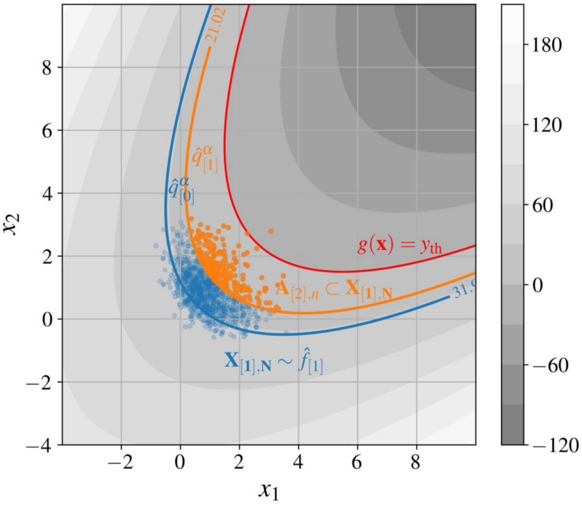


Figure 8: BANCS sampling steps - illustration of the iterations on a parabolic case.

Parabolic toy-case:

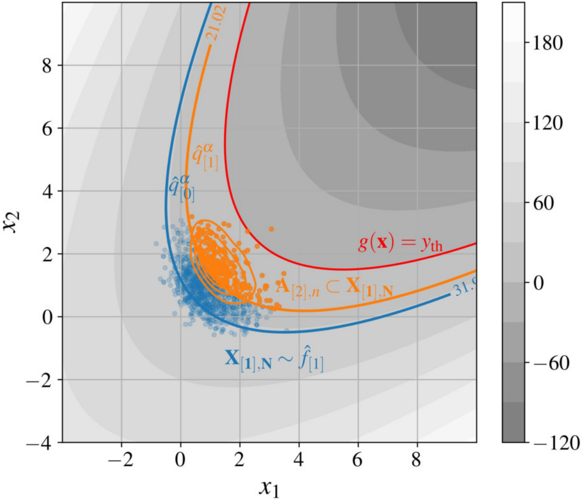


Figure 8: BANCS sampling steps - illustration of the iterations on a parabolic case.

Numerical results on toy-cases

Parabolic toy-case:

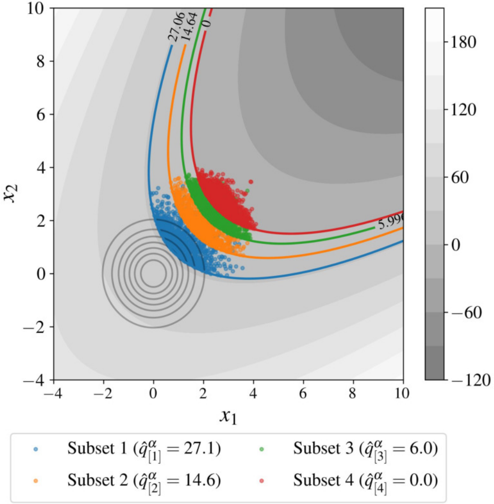


Figure 8: BANCS sampling steps - illustration of the iterations on a parabolic case.

Numerical results on toy-cases:

Toy-case #1: Four-branch ($p_f^{\text{ref}} = 2.21 \times 10^{-4}$)

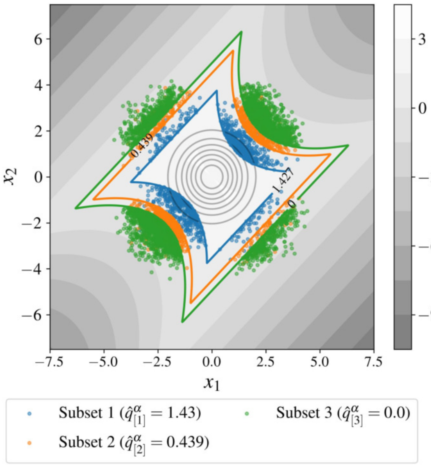


Figure 9: BANCS sampling steps - toy-case #1.

Numerical results on toy-cases:

Toy-case #1: Four-branch ($p_f^{ref} = 2.21 \times 10^{-4}$)

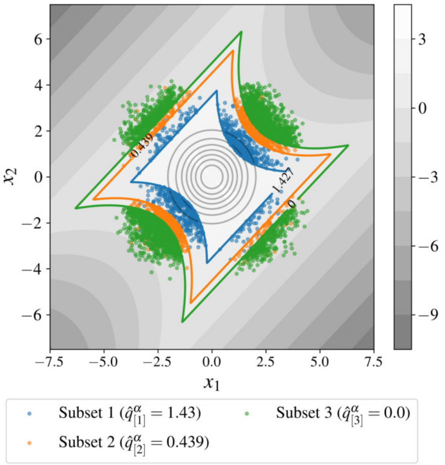


Figure 9: BANCS sampling steps - toy-case #1.

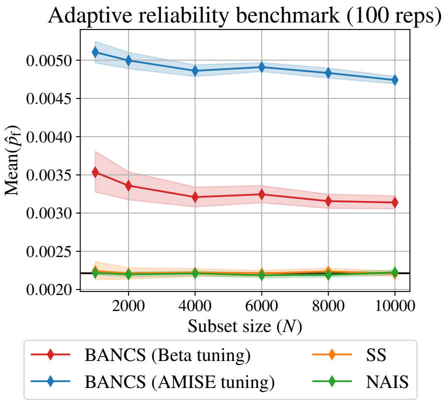


Figure 10: Benchmark results - toy-case #1 (IC built by bootstrap on 100 repetitions).

Toy-case #1: Four-branch ($p_f^{ref} = 2.21 \times 10^{-4}$)

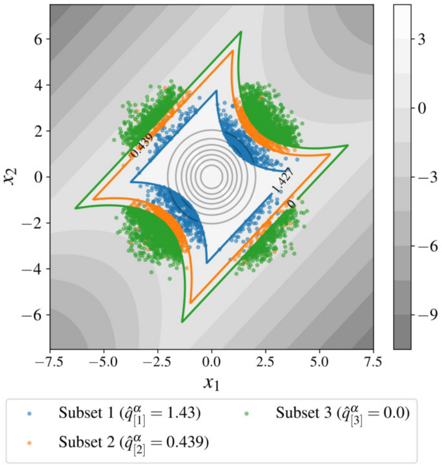


Figure 9: BANCS sampling steps - toy-case #1.

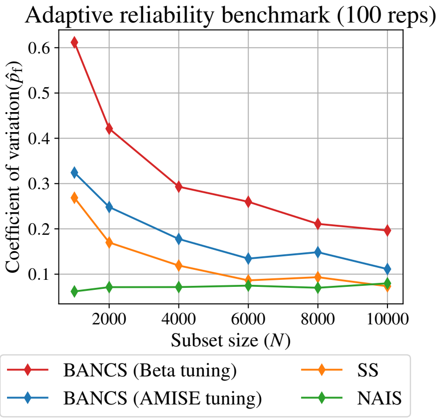


Figure 10: Benchmark results - toy-case #1 (IC built by bootstrap on 100 repetitions).

Numerical results on toy-cases

Toy-case #2: 7D reliability problem[23] ($p_\mathrm{f}^\mathrm{ref} = 8.10 \times 10^{-3}$)

$$g_2(\mathbf{x}) = 15.59 \times 10^4 - \frac{x_1 x_3^2}{2x_3^2} \frac{x_2^4 - 4x_5 x_6 x_7^2 + x_4(x_6 + 4x_5 + 2x_6 x_7)}{x_4 x_5 (x_4 + x_6 + 2x_6 x_7)} \tag{7}$$

Toy-case #2: 7D reliability problem[23] ($p_{\mathrm{f}}^{\mathrm{ref}} = 8.10 \times 10^{-3}$)

$$g_2(\mathbf{x}) = 15.59 \times 10^4 - \frac{x_1 x_3^2}{2x_3^2} \frac{x_2^4 - 4x_5 x_6 x_7^2 + x_4(x_6 + 4x_5 + 2x_6 x_7)}{x_4 x_5 (x_4 + x_6 + 2x_6 x_7)} \quad (7)$$

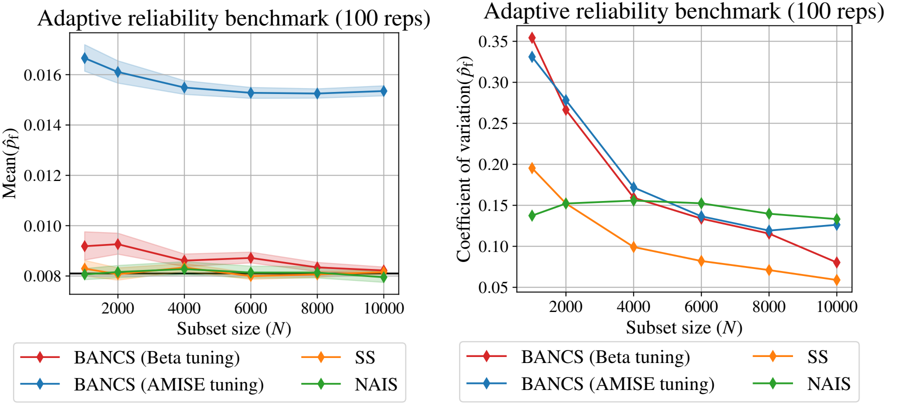


Figure 11: Benchmark results - toy-case #2 (IC built by bootstrap on 100 repetitions)

Discussion

Numerical results:[24]

	$p_\mathrm{f}^{\mathrm{ref}}$	$\widehat{p}_\mathrm{f}^{\mathrm{BANCS}}$	$\widehat{\delta}^{\mathrm{BANCS}}$	$\widehat{p}_\mathrm{f}^{\mathrm{SS}}$	$\widehat{\delta}^{\mathrm{SS}}$	$\widehat{p}_\mathrm{f}^{\mathrm{NAIS}}$	$\widehat{\delta}^{\mathrm{NAIS}}$
#1	2.21×10^{-4}	3.13×10^{-4}	19%	2.20×10^{-4}	7%	2.20×10^{-4}	7%
#2	8.10×10^{-3}	8.22×10^{-3}	7%	8.16×10^{-3}	6%	7.95×10^{-3}	14%

Table 1: Results of the numerical experiments (subset sample size $N = 10^4$, $p_0 = 0.1$).

Remarks

- SS and NAIS algorithms use the `OpenTURNS`[25] implementation
- Beta tuning seems favorable for BANCS (see Beta copula[26])
- BANCS performs well on the medium dimension toy-case #2
- The algorithm seems to introduce a bias

[24] `https://github.com/efekhari27/bancs`

[25] `https://openturns.github.io/www/`

[26] Segers, Sibuya, and Tsukahara 2017.

Application to the offshore wind turbine case

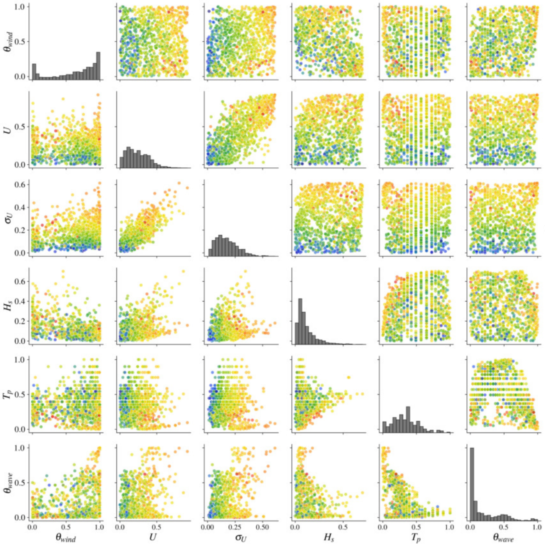


Figure 12: Copulogram with outputs in color on the Teesside case ($n = 2000$). The highest values are in red.

- Input distribution:
  $\hookrightarrow$fitted by EBC
- Limit-state function:
  $\hookrightarrow$fitted by metamodel g
- Reliability problems:
  $\hookrightarrow$ wind turbine problem #1:
  $p_f^1 = \mathbb{P}(g(\mathbf{X}) \leq q_{99\%}) = 10^{-2}$
  $\hookrightarrow$wind turbine problem #2:
  $p_f^2 = \mathbb{P}(g(\mathbf{X}) \leq q_{99.9\%}) = 10^{-3}$

Application to the offshore wind turbine case

Numerical results:

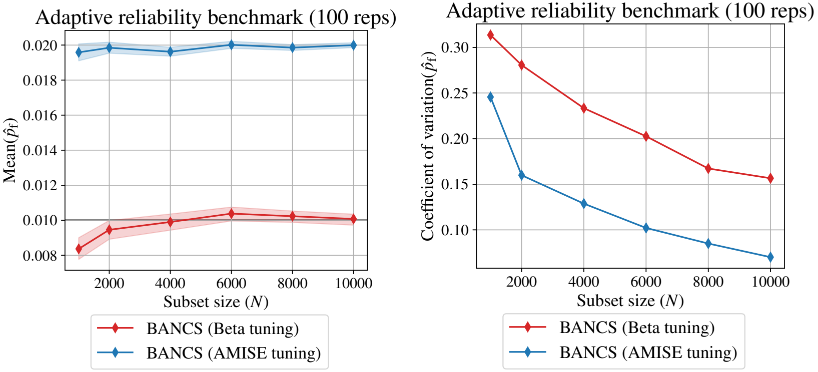


Figure 13: Numerical results - Industrial case (IC built by bootstrap on 100 repetitions)

	p_f^{ref}	$\widehat{p}_f^{\text{BANCS}}$	$\widehat{\delta}^{\text{BANCS}}$
wind turbine problem #1	10^{-2}	1.00×10^{-2}	16%
wind turbine problem #2	10^{-3}	1.60×10^{-3}	23%

Table 2: Numerical results - Industrial case (subset sample size $N = 10^4$, $p_0 = 0.1$).

Numerical results:

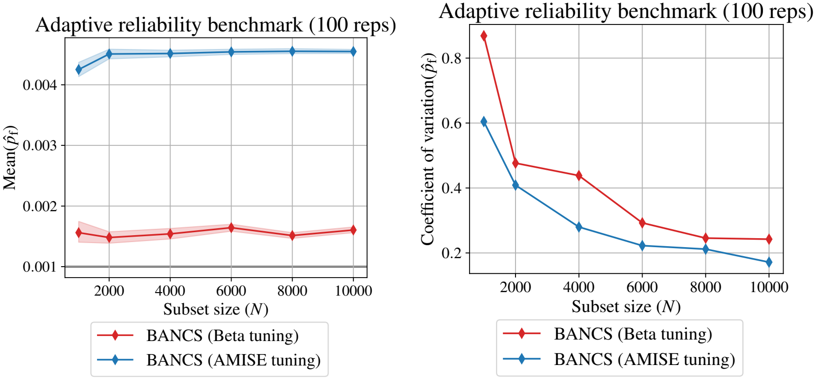


Figure 13: Numerical results - Industrial case (IC built by bootstrap on 100 repetitions)

	p_f^{ref}	$\widehat{p}_f^{\text{BANCS}}$	$\widehat{\delta}^{\text{BANCS}}$
wind turbine problem #1	10^{-2}	1.00×10^{-2}	16%
wind turbine problem #2	10^{-3}	1.60×10^{-3}	23%

Table 2: Numerical results - Industrial case (subset sample size $N = 10^4$, $p_0 = 0.1$).

Conclusions:

- Nonparametric copula estimation allows a lot of flexibility
- BANCS gives promising results (med. dimension) and i.i.d samples
  - $\hookrightarrow$ needed for dedicated sensitivity analysis
- BANCS allows rare event estimation directly in the physical space (i.e., without transformation)
  - $\hookrightarrow$ useful for complex inputs

[27] Marrel and Chabridon 2021.

Conclusions and limits

Conclusions:

- Nonparametric copula estimation allows a lot of flexibility
- BANCS gives promising results (med. dimension) and i.i.d samples
  - ↪ needed for dedicated sensitivity analysis
- BANCS allows rare event estimation directly in the physical space (i.e., without transformation)
  - ↪ useful for complex inputs

Limits and perspectives:

- BANCS estimator presents a bias, especially in small dimension
  - ↪ explore EBC tuning with other tools (e.g., Csiszár divergence)
  - ↪ explore penalized-EBC, B-splines
- BANCS estimator misses an asymptotic variance
- BANCS samples can be used for dedicated sensitivity analysis
  - ↪ which inputs influence the failure? (e.g., Target-HSIC indices[27])

[27] Marrel and Chabridon 2021.

Bibliography

- Au, S.-K. and J. L. Beck (2001). "Estimation of small failure probabilities in high dimensions by subset simulation". In: *Probabilistic Engineering Mechanics* 4, pp. 263–277.
- Cérou, F. et al. (2012). "Sequential Monte Carlo for rare event estimation". In: *Statistics and computing*, pp. 795–808.
- Chen, T. et al. (2018). "Fatigue bending test on grouted connections for monopile offshore wind turbines". In: *Marine Structures*, pp. 52–71.
- Fekhari, E., V. Chabridon, et al. (2023a). "Bernstein adaptive nonparametric conditional sampling: a new method for rare event probability estimation". preprint: hal-04052861v1.
- — (2023b). "Fast given-data uncertainty propagation in offshore wind turbine simulator using Bayesian quadrature". preprint: hal-04052859v1.
- Fekhari, E., B. Iooss, et al. (2023). "Model predictivity assessment: incremental test-set selection and accuracy evaluation". In: *Studies in Theoretical and Applied Statistics, SIS 2021, Pisa, Italy, June 21-25*. Ed. by N. Salvati et al. Springer Cham, pp. 315–347.
- Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall.
- Joe, H. and D. Kurowicka (2011). *Dependence modeling: vine copula handbook*. World Scientific.
- Kim, T. et al. (2022). "A comprehensive code-to-code comparison study with the modified IEA15MW-UMaine Floating Wind Turbine for H2020 HIPERWIND project". In: *Journal of Physics: Conference Series* 2265.4, p. 042006.
- Kurtz, N. and J. Song (2013). "Cross-entropy-based adaptive importance sampling using Gaussian mixture". In: *Structural Safety*, pp. 35–44.
- Lovera, A. et al. (2023). "Quantifying and clustering the wake-induced perturbations within a wind farm for load analysis". In: *Journal of Physics: Conference Series*. IOP Publishing.
- Marrel, A. and V. Chabridon (2021). "Statistical developments for target and conditional sensitivity analysis: Application on safety studies for nuclear reactor". In: *Reliability Engineering & System Safety*, p. 107711.
- Morio, J. (2011). "Non-parametric adaptive importance sampling for the probability estimation of a launcher impact position". In: *Reliability Engineering and System Safety*, pp. 178–183.
- Morio, J. and M. Balesdent (2015). *Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems: A Practical Approach*. Woodhead Publishing, Elsevier.
- Nagler, T., C. Schellhase, and C. Czado (2017). "Nonparametric estimation of simplified vine copula models: comparison of methods". In: *Dependence Modeling*, pp. 99–120.
- Papaioannou, I. et al. (2015). "MCMC algorithms for Subset Simulation". In: *Probabilistic Engineering Mechanics*, pp. 89–103.
- Sancetta, A. and S. Satchell (2004). "The Bernstein copula and its applications to modeling and approximations of multivariate distributions". In: *Econometric Theory*, pp. 535–562.
- Segers, J., M. Sibuya, and H. Tsukahara (2017). "The empirical beta copula". In: *Journal of Multivariate Analysis*, pp. 35–51.
- Vanem, E. et al. (2023). "A joint probability distribution model for multivariate wind and wave conditions". In: *Proceedings of the ASME 2023 42th International Conference on Ocean, Offshore and Arctic Engineering*.
- Yun, W. et al. (2018). "An efficient global reliability sensitivity analysis algorithm based on classification of model output and subset simulation". In: *Structural Safety*, pp. 49–57.

Thank you for your attention

Teesside wind farm

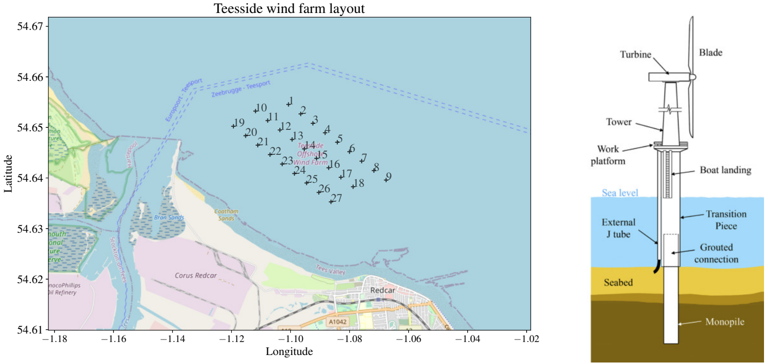


Figure 14: Wind farm layout (Teesside, UK); Monopile OWT diagram[28].

- Bottom-fixed monopile foundation with 2.3 MW Siemens wind turbines;
- Site very close to shore.

[28] Chen et al. 2018.

Chained numerical simulation models

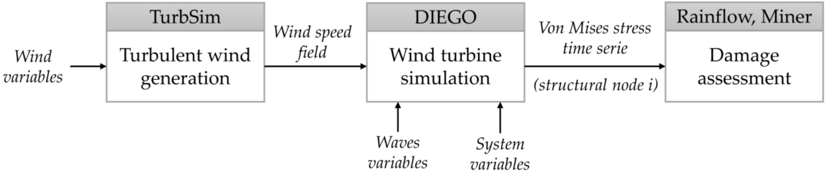


Figure 15: Diagram of the chained wind turbine simulation model

Chained numerical models

$$g\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$$
$$(\mathbf{x}, \mathbf{z}) \mapsto g(\mathbf{x}, \mathbf{z}).$$

- **X** environmental random vector with its joint distribution $f_{\mathbf{X}}(\cdot)$
- **Z** system random vector with its joint distribution $f_{\mathbf{Z}}(\cdot)$
- Simulation CPU time: $\sim$15 min; Simulated period: 10 min;
- Deployed on EDF R&D High Performance Computers facility.

TurbSim: turbulent wind field simulation

TurbSim is a stochastic, full-field, turbulence simulator (NREL)

- **inputs**: mean wind, wind direction, wind shear, turbulence model, turbulence intensity, hub height, simulation time, etc.
- **outputs**: wind speed field

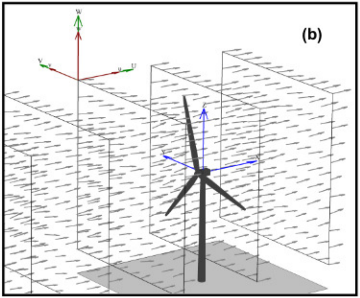


Figure 16: Illustrative wind speed field simulated

Simulations are replicated for 11 pseudo-random seed.

DIEGO: Hydro-Aero-Servo-Elasto simulation

Dynamique Intégrée des Eoliennes et Génératrices Offshore, DIEGO is WT multi-physics simulator (EDF R&D[29])

- **inputs**: TurbSim's output, waves properties, WT geometry, material properties, soil stiffness, controller properties.
- **outputs**: power production, mechanical stress, displacements, etc.

Figure 17: Illustrative structural and aero-dynamic mesh from DIEGO

[29] Kim et al. 2022.

Mechanical damage assessment

Damage computed at specific points of the structure (e.g., the mudline):

1. Equivalent Von Mises stress time series;
2. Fatigue stress cycles identification using Rainflow counting;
3. Damage computation using Miner's rule.

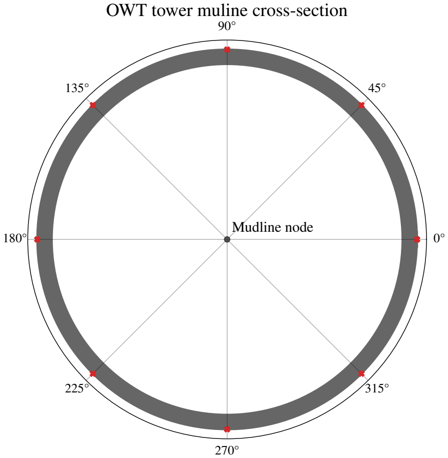


Figure 18: Horizontal cross-section of the OWT structure and the mudline

Random environmental conditions

Considered environmental random variables:

Mean wind speed	U	Weibull	10-min. average horizontal at 10m
Turbulence	σ_s	Log-normal	10-min. standard deviation
Wind direction	θ_{wind}	non-parametric	Wind directions
Significant wave height	H_s	Weibull	Significant wave height per hour
Peak wave period	T_p	Log-normal	Peak 1-hour spectral wave period
Wave direction	θ_{wave}	non-parametric	Wave directions
Mean shear	α	Normal	10-min mean shear exponent
Air density	δ	-	-

Table 3: Marginal distributions of the environmental random variables

- Large on-site SCADA data available
- Challenging dependency structure for a parametric model

Quantities of interest summary

Considering the random vectors **X** and **Z** as independent

$\hookrightarrow$ Different means

Conditioned by **Z**	$\mathbb{E}\left[g(\mathbf{X},\mathbf{Z})\vert\mathbf{Z}\right] = \int_{\mathcal{D}_{\mathbf{X}}} g(\mathbf{x},\mathbf{Z}) f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \phi(\mathbf{Z})$
Conditioned by **X**	$\mathbb{E}\left[g(\mathbf{X},\mathbf{Z})\vert\mathbf{X}\right] = \int_{\mathcal{D}_{\mathbf{Z}}} g(\mathbf{X},\mathbf{z}) f_{\mathbf{Z}}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \psi(\mathbf{X})$

$\hookrightarrow$ Different failure probabilities:

Joint	$p_{\mathrm{f}} = \mathbb{E}\left[\mathbb{1}_{\{g(\mathbf{X},\mathbf{Z})\leq D_{\mathrm{cr}}\}}\right] = \iint_{\mathcal{D}_{\mathbf{X}}\times\mathcal{D}_{\mathbf{Z}}} \mathbb{1}_{\{g(\mathbf{X},\mathbf{Z})\leq D_{\mathrm{cr}}\}} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}}(\mathbf{z})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{z}$
On **Z**	$p_{\mathrm{f}_{\mathbf{Z}}} = \mathbb{E}\left[\mathbb{1}_{\{\phi(\mathbf{Z})\leq D_{\mathrm{cr}}\}}\right] = \int_{\mathcal{D}_{\mathbf{Z}}} \mathbb{1}_{\{\phi(\mathbf{Z})\leq D_{\mathrm{cr}}\}} f_{\mathbf{Z}}(\mathbf{z})\,\mathrm{d}\mathbf{z}$
On **X**	$p_{\mathrm{f}_{\mathbf{X}}} = \mathbb{E}\left[\mathbb{1}_{\{\psi(\mathbf{X})\leq D_{\mathrm{cr}}\}}\right] = \int_{\mathcal{D}_{\mathbf{X}}} \mathbb{1}_{\{\psi(\mathbf{X})\leq D_{\mathrm{cr}}\}} f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x}$

- Bound these quantities? Concentration inequalities?
- Aleatory and epistemic random vectors with unexpected tags?