HAL open science

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

Emiliano Torre, Stefano Marelli, Paul Embrechts, Bruno Sudret

▶ To cite this version:

Emiliano Torre, Stefano Marelli, Paul Embrechts, Bruno Sudret. A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas. Probabilistic Engineering Mechanics, 2018, 10.1016/j.probengmech.2018.08.001 . hal-01901982

HAL Id: hal-01901982
https://hal.science/hal-01901982

Submitted on 23 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ETH zürich

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

E. Torre, S. Marelli, P. Embrechts, B. Sudret

Chair of Risk, Safety and Uncertainty Quantification
Stefano-Franscini-Platz 5
CH-8093 Zürich

Risk, Safety & Uncertainty Quantification

Data Sheet

Journal:

Report Ref.: RSUQ-2017-012

Arxiv Ref.: http://arxiv.org/abs/1709.08626 - [stat.CO]

DOI:

Date submitted: September 25, 2017

Date accepted: -

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

E. Torre, S. Marelli, P. Embrechts, B. Sudret

August 10, 2018

Abstract

Systems subject to uncertain inputs produce uncertain responses. Uncertainty quantification (UQ) deals with the estimation of statistics of the system response, given a computational model of the system and a probabilistic model of its inputs. In engineering applications it is common to assume that the inputs are mutually independent or coupled by a Gaussian or elliptical dependence structure (copula).

In this paper we overcome such limitations by modelling the dependence structure of multivariate inputs as vine copulas. Vine copulas are models of multivariate dependence built from simpler pair-copulas. The vine representation is flexible enough to capture complex dependencies. This paper formalises the framework needed to build vine copula models of multivariate inputs and to combine them with virtually any UQ method. The framework allows for a fully automated, data-driven inference of the probabilistic input model on available input data.

The procedure is exemplified on two finite element models of truss structures, both subject to inputs with non-Gaussian dependence structures. For each case, we analyse the moments of the model response (using polynomial chaos expansions), and perform a structural reliability analysis to calculate the probability of failure of the system (using the first order reliability method and importance sampling). Reference solutions are obtained by Monte Carlo simulation. The results show that, while the Gaussian assumption yields biased statistics, the vine copula representation achieves significantly more precise estimates, even when its structure needs to be fully inferred from a limited amount of observations.

Keywords: uncertainty quantification, input dependencies, vine copulas, reliability analysis, polynomial chaos expansions

1 Introduction

Uncertainty Quantification (UQ) estimates statistics of the response of a system subject to stochastic inputs. The system is usually described by a deterministic computational model $\mathcal{M}$ (*e.g.*, a finite element code). The input consists of M possibly coupled parameters, modelled by a random vector $\boldsymbol{X}$ with joint cumulative distribution function (CDF) $F_{\boldsymbol{X}}$ and probability density (PDF) $f_{\boldsymbol{X}}$. The computational model transforms $\boldsymbol{X}$ into an uncertain output $Y = \mathcal{M}(\boldsymbol{X})$, which here we take to be a univariate random variable. The extension to multivariate outputs is straightforward.

Of interest in UQ problems are various statistics of Y, such as its CDF F_Y, its moments, the probability of extreme events (*i.e.*, of small or large quantiles), the sensitivity of Y to the different components X_i of $\boldsymbol{X}$, and others. Because $\mathcal{M}$ is typically a complex model which is not known explicitly, analytical solutions are in general not available. The model behavior can only be known point-wise in correspondence with inputs $\boldsymbol{x}^{(j)}$ sampled from $F_{\boldsymbol{X}}$, where it produces responses $y^{(j)} = \mathcal{M}(\boldsymbol{x}^{(j)})$ (non-intrusive, or black-box approach). The classical and most general strategy to solve this class of problems is by Monte Carlo simulation (MCS). MCS draws the $\boldsymbol{x}^{(j)}$ as i.i.d samples from $F_{\boldsymbol{X}}$, which requires the sample size n to be large enough to cover the input probability space sufficiently well. When $\mathcal{M}$ is computationally expensive and the available computational budget is limited to a few dozens to hundreds of runs, alternative approximation techniques are used instead of MCS. Examples include the first and second order reliability methods (FORM (Hasofer and Lind, 1974), SORM (Fiessler et al., 1979)), importance sampling (IS, Melchers (1999)) and subset simulation (Au and Beck, 2001) in reliability analysis for the estimation of small failure probabilities (see also Ditlevsen and Madsen (1996); Lemaire (2009)), and polynomial chaos expansions (PCE, Li and Ghanem (1998)), Kriging (Matheron, 1967), and other metamodelling techniques for the estimation of the moments.

Since $\mathcal{M}$ is a deterministic code, all uncertainty in Y is due to the uncertainty in $\boldsymbol{X}$. Therefore, regardless of the approach (MCS or others) chosen to estimate the statistics of Y of interest, a suitable model of $F_{\boldsymbol{X}}$ is critical to obtain accurate estimates. Historically, the components X_i of $\boldsymbol{X}$ are assumed to be mutually independent, or to have the dependence structure of a multivariate elliptical distribution (Lebrun and Dutfoy, 2009a). Among the latter, Gaussian distributions are often employed because they are simple to model and to fit to data, since they only require the computation of pairwise correlation coefficients. In addition, some advanced UQ techniques take advantage of (or require) mutually independent inputs. These include FORM, SORM, IS, some types of subset simulation (*e.g.*, Papaioannou et al. (2015)), PCE. The most general transformation to map the input vector $\boldsymbol{X}$ onto a vector $\boldsymbol{Z}$ with independent components, the Rosenblatt transform (Rosenblatt, 1952), requires the computation of conditional PDFs, which are hardly known in practical applications. However, when $F_{\boldsymbol{X}}$ has a Gaussian dependence structure, this map is known and is equivalent to the well known Nataf transform (Nataf, 1962; Lebrun and Dutfoy, 2009a). The Gaussian assumption introduces thus a convenient representation of input dependencies. When the real dependence structure deviates from this assumption, it may however introduce a bias in the resulting estimates. The validity or the impact of the Gaussian assumption, though, are typically not quantified. Novel methodologies in UQ largely focus on providing better estimation techniques rather than on allowing for different probabilistic input models.

Recently, dependence modelling has seen significant advances in the mathematical community with the widespread adoption of copula models, and of vine copulas in particular. Copula theory allows to separately model the dependence (by multivariate copula functions) and the marginal behaviour (by univariate CDFs) of joint distributions. This provides a flexible way to build multivariate probability models by selecting each ingredient individually (Nelsen, 2006; Joe, 2015). Copulas have recently been used in various studies in engineering, such as in earthquake (Goda, 2010; Goda and Tesfamariam, 2015; Zentner, 2017) and sea waves (Michele et al., 2007; Masina et al., 2015; Montes-Iturrizaga and Heredia-Zavoni, 2016) engineering. Applications, however, are often limited to low-dimensional (typically bivariate) problems, or to relatively simple copula families, prominently the Gaussian or Archimedean families (Nelsen, 2006). In higher dimensions, building and selecting copulas that properly represent the coupling of the phenomena of interest may be a complex problem. Vine copulas, first established by Joe (1996) and Bedford and Cooke (2002), ease this construction by expressing multivariate copulas as a product of simpler bivariate copulas among pairs of random variables. As a result, vine models offer an easy interpretation and are extremely flexible. Vine copulas have been extensively employed, for instance, in financial applications (Aas, 2016). In engineering, these models have been, so far, largely overseen. Recently, Wang and Li (2017c,a) proposed their application in the context of reliability analysis, for the special case when only partial information (correlation coefficients) is available. In a later study, they used vine copulas in combination with MCS for reliability analysis (Wang and Li, 2017b).

This manuscript proposes a general framework to use vine copulas to model model input dependencies in UQ problems. The flexibility of these models guarantees an accurate description of the input dependence properties that shape the output statistics. Besides, since algorithms to compute the Rosenblatt transform of vine copulas are available, these dependence models are applicable also in combination with UQ techniques that work in probability spaces with independent variables. Algorithms to infer the structure and fit the parameters of vine models to data, for instance based on maximum likelihood or Bayesian estimation, also exist, making these models suitable for data driven applications (Aas et al., 2009; Schepsmeier, 2015).

After recalling fundamental results of copula and vine copula theory (Sections 2-3), we combine three established UQ methodologies, FORM, IS and PCE, with vine copula models of the input dependencies (Section 4). In Sections 5-6 we apply the methodology to two truss models. We show that modelling non-Gaussian input dependencies with the Gaussian copula yields wrong estimates of the failure probability and of the response moments. The problem cannot be amended by using different UQ methods, since it is inherent to the wrong representation of the input uncertainty. Reliable estimates are obtained instead by using a suitable vine representation of the input, also when the vine is purely inferred from available data. The method's advantages and current limitations are discussed in Section 7.

2 Copulas and vine copulas

Multivariate inputs in UQ problems are generally modelled as random vectors. The statistical properties of an M-dimensional random vector $\boldsymbol{X}$ are fully described by its joint CDF

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(X_1 \leq x_1, \ldots, X_M \leq x_M).$$

The joint CDF defines both the marginal CDF of each component X_i of $\boldsymbol{X}$, *i.e.*, $F_i(x_i) = F_{X_i}(x_i) = \mathbb{P}(X_i \leq x_i)$, $i = 1, \ldots, M$, and the dependence properties of the variables. As such, prescribed parametric families of joint CDFs dictate specific parametric forms for the marginal and joint properties of the random variables. More flexible models should be compatible with inference techniques, to be applicable when only a finite number of realisations of the input $\boldsymbol{X}$ is available. They should also optimally provide the isoprobabilistic map that decouples their random variables, such to be usable in combination with UQ techniques that assume mutually independent inputs. This section introduces vine copula models and illustrates how they meet the requirements listed above.

2.1 Copulas and Sklar's theorem

An M-copula is defined as an M-variate joint CDF $C : [0,1]^M \to [0,1]$ with standard uniform marginals, that is, such that

$$C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i \quad \forall u_i \in [0,1], \quad \forall i = 1, \ldots, M.$$

Sklar's theorem (Sklar, 1959) allows one to express joint CDFs in terms of their marginal distributions and a copula.

Theorem (Sklar). *For any M-variate* CDF *$F_{\boldsymbol{X}}$ with marginals $F_1, \ldots, F_M$, an M-copula $C_{\boldsymbol{X}}$ exists, such that for all $\boldsymbol{x} \in \mathbb{R}^M$*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = C_{\boldsymbol{X}}(F_1(x_1), \ldots, F_M(x_M)). \tag{1}$$

Besides, $C_{\boldsymbol{X}}$ is unique on $Ran(F_1) \times \ldots \times Ran(F_M)$, where Ran is the range operator. In particular, $C_{\boldsymbol{X}}$ is unique on $[0,1]^M$ if all F_i are continuous, and it is given by

$$C_{\boldsymbol{X}}(\boldsymbol{u}) = F_{\boldsymbol{X}}(F_1^{-1}(u_1), \ldots, F_M^{-1}(u_M)), \quad \boldsymbol{u} \in [0,1]^M. \tag{2}$$

Conversely, for any M-copula C and any set of M univariate CDF*s F_i with domain $\mathcal{D}_i$, $i = 1, \ldots, M$, the function $F : \mathcal{D}_1 \times \ldots \times \mathcal{D}_M \to [0, 1]$ defined by*

$$F(x_1, \ldots, x_M) := C(F_1(x_1), \ldots, F_M(x_M)) \tag{3}$$

is an M-variate CDF *with marginals $F_1, \ldots, F_M$.*

The representation (1) guarantees that any joint CDF can be expressed in terms of its marginals and a copula. In the following we work with joint CDFs $F_{\boldsymbol{X}}$ having continuous marginals F_i.

Copulas of known families of joint CDFs can be derived from (2). Finally, one can use (3) to build a multivariate CDF F by separately specifying and combining M univariate CDFs F_i and a copula C. The univariate CDFs describe the marginal behaviour, while the copula describes the dependence properties. Sklar's theorem thus allows one to split the problem of modelling the joint behaviour of the components of $\boldsymbol{X}$ into two separate problems. One first models the marginals F_i, then transforms the original components X_i into uniform random variables $U_i = F_i(X_i)$, leading to the transformation

$$\mathcal{T}^{(\mathcal{U})} : \boldsymbol{X} \mapsto \boldsymbol{U} = (F_1(X_1), \ldots, F_M(X_M))^{\mathrm{T}}. \tag{4}$$

The joint CDF of $\boldsymbol{U} = (U_1, \ldots, U_M)^{\mathrm{T}}$ is the associated copula.

Sklar's theorem can be re-stated in terms of probability densities. If $\boldsymbol{X}$ admits PDF $f_{\boldsymbol{X}}(\boldsymbol{x}) := \dfrac{\partial^M F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial x_1 \ldots \partial x_M}$ and copula density $c_{\boldsymbol{X}}(\boldsymbol{u}) := \dfrac{\partial^M C_{\boldsymbol{X}}(\boldsymbol{u})}{\partial u_1 \ldots \partial u_M}$, then the following relation holds:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = c_{\boldsymbol{X}}(F_1(x_1), \ldots, F_M(x_M)) \cdot \prod_{i=1}^{M} f_i(x_i). \tag{5}$$

2.2 Copula-based measures of dependence

Since copulas fully describe multivariate dependencies, it is natural to introduce dependence measures based on the copula only, and not on the marginals. Several such measures, also known as measures of concordance, exist. An example is Spearman's correlation coefficient, defined for a random pair (X_1, X_2) as

$$\rho_S(X_1, X_2) := \rho_P(F_1(X_1), F_2(X_2)),$$

where ρ_P is the classical Pearson correlation coefficient. Another example is Kendall's tau

$$\tau_K(X_1, X_2) := \mathbb{P}((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0) - \mathbb{P}((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0),$$

where $(\tilde{X}_1, \tilde{X}_2)$ is an independent copy of (X_1, X_2). If the copula of (X_1, X_2) is C, then

$$\rho_S(X_1, X_2) = 12 \iint_{[0,1]^2} C(u, v) du dv - 3 = 3 - 12 \iint_{[0,1]^2} u \frac{\partial C(u, v)}{\partial u}, \tag{6}$$

and

$$\tau_K(X_1, X_2) = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1 = 1 - 4 \iint_{[0,1]^2} \frac{\partial C(u, v)}{\partial u} \frac{\partial C(u, v)}{\partial v} du dv, \tag{7}$$

where the RHS in both equations is well defined if the copula partial derivatives exist and are not degenerate at the borders (Joe, 2015).

One can show that $\tau_K = 0$ and $\rho_S = 0$ if (X_1, X_2) are independent, that $\tau_K = 1 \Leftrightarrow \rho_S = 1 \Leftrightarrow X_1 = \alpha(X_2)$ for some strictly increasing $\alpha(\cdot)$, and that $\tau_K = -1 \Leftrightarrow \rho_S = 1 \Leftrightarrow X_2 = \beta(X_1)$ for some strictly decreasing $\beta(\cdot)$ (Embrechts et al., 1999). Other copula based measures of pairwise concordance exist (Scarsini, 1984), as well as multivariate extensions (Taylor, 2007). A discussion of such measures is beyond the scope of this paper.

Asymptotic tail dependence (hereinafter, simply tail dependence) of a random pair (X_1, X_2) is another example of dependence property that is completely described by the copula and not by the marginals. The joint distribution of (X_1, X_2) is said to be upper tail dependent if the probability that one of the two variables takes values in its upper tail (*i.e.*, high quantiles), given that the other has taken values in its upper tail, does not decay to zero. Lower tail dependence is defined analogously for low quantiles. Tail dependence thus allows for simultaneous extremes, and is for instance used to model systemic risks. Formally, (X_1, X_2) with marginals F_1 and F_2 are upper tail dependent if

$$\lim_{u \uparrow 1^-} \mathbb{P}(X_1 > F_1^{-1}(u) | X_2 > F_2^{-1}(u)) = \lambda_u > 0, \tag{8}$$

and are lower tail dependent if

$$\lim_{u \downarrow 0^+} \mathbb{P}(X_1 < F_1^{-1}(u) | X_2 < F_2^{-1}(u)) = \lambda_l > 0, \tag{9}$$

given that these limits exist; λ_u and λ_l are called the upper and lower tail dependence coefficients, and can be expressed in terms of the copula C of (X_1, X_2) by

$$\lambda_u = \lim_{u \uparrow 1^-} \frac{1 - 2u + C(u,u)}{1-u}, \quad \lambda_l = \lim_{u \downarrow 0^+} \frac{C(u,u)}{u}. \tag{10}$$

2.3 Copula examples

Here we provide three families of copulas that will be used in Section 5 and Section 6 to model different dependence structures among input loads on a truss model. A list of classical families of copulas and their properties can be found in Nelsen (2006); Joe (2015). A summary of 19 families of bivariate copulas used for inference in this study and of their dependence properties is provided in Tables 11-12.

The independence copula

$$C^{(\Pi)}(\boldsymbol{u}) = \prod_{i=1}^{M} u_i \tag{11}$$

describes the case of mutual independence among the random variables. For $M = 2$, $C^{(\Pi)}$ has Spearman's rho $\rho_S^{(\Pi)} = 0$, Kendall's tau $\tau_K^{(\Pi)} = 0$, and tail dependence coefficients $\lambda_u^{(\Pi)} = \lambda_l^{(\Pi)} = 0$.

A Gaussian random vector $\boldsymbol{X}$ with correlation matrix $\mathbf{R} = (\rho_{ij})_{i,j=1}^{M}$ and marginals

$F_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \ldots, M$, has copula

$$C^{(\mathcal{N})}(\boldsymbol{u}) = \frac{1}{\sqrt{\det \mathbf{R}}} \exp\left(-\frac{1}{2} \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_M) \end{pmatrix}^T \cdot \left(\mathbf{R}^{-1} - \mathbf{I}\right) \cdot \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_M) \end{pmatrix} \right), \tag{12}$$

where Φ is the univariate standard normal CDF and $\mathbf{I}$ is the identity matrix of rank M. $C^{(\mathcal{N})}$ is called Gaussian copula or normal copula. One can prove that, if $M \geq 3$ variables are coupled by a Gaussian copula with correlation matrix $\mathbf{R}$, any pairs $(X_i,\, X_j)$ are coupled by a Gaussian pair copula with correlation matrix $\begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix}$. If so, their Spearman's rho is $\rho_S^{(\mathcal{N})} = \frac{6}{\pi} \arcsin(\frac{\rho_{ij}}{2})$, their Kendall's tau is $\tau_K^{(\mathcal{N})} = \frac{2}{\pi} \arcsin(\rho_{ij})$, and their tail dependence coefficients are $\lambda_u^{(\mathcal{N})} = \lambda_l^{(\mathcal{N})} = 0$. Therefore, multivariate Gaussian copulas assign negligible probabilities to joint extremes.

A pair copula that contemplates upper tail dependence is the bivariate Gumbel-Hougaard (or Gumbel, for brevity) copula

$$C^{(\mathcal{GH})}(u, v) = \exp\left(-\left[(-\log u)^\theta + (-\log v)^\theta \right]^{1/\theta} \right),\ \theta \in [1, +\infty). \tag{13}$$

In particular, if $\theta = 1$ then $C^{(\mathcal{GH})}(u, v) = uv$ (the independence copula). $C^{(\mathcal{GH})}$ has Kendall's tau $\tau_K^{(\mathcal{GH})} = (\theta - 1)/\theta$ and upper tail dependence coefficient $\lambda_u^{(\mathcal{GH})} = 2 - 2^{1/\theta}$, which increases from 0 to 1 as θ increases from 1 to $+\infty$. Finally, $\lambda_l^{(\mathcal{GH})} = 0$.

2.4 Vine copulas

When the input dimension M grows, defining a suitable M-copula which properly describes the pairwise and higher-order dependencies among the input variables becomes increasingly difficult. Multivariate extensions of several families of pair-copulas exist, but they rarely fit real data well. Bedford and Cooke (2002) proved that, instead, one may construct any M-copula by a product of simpler 2-copulas. Some are unconditional copulas among pairs of random variables, others are conditioned on the values taken by other variables. Here we briefly introduce this construction, known as pair copula or vine copula construction, and recall some important features. For details, we refer to the cited literature (see also Klüpperberg and Czado (date)). A recent review with a focus on financial applications can be found in Aas (2016).

Let $\boldsymbol{u}_{\bar{i}}$ be the vector obtained from the vector $\boldsymbol{u}$ by removing its i-th component, *i.e.*, $\boldsymbol{u}_{\bar{i}} = (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_M)^{\mathrm{T}}$. Similarly, let $\boldsymbol{u}_{\overline{\{i,j\}}}$ be the vector obtained by removing the i-th and j-th component, and so on. For a general subset $\mathcal{A} \subset \{1, \ldots,\, M\}$, $\boldsymbol{u}_{\overline{\mathcal{A}}}$ is defined analogously. Also, $F_{\overline{\mathcal{A}}|\mathcal{A}}$ and $f_{\overline{\mathcal{A}}|\mathcal{A}}$ indicate the joint CDF and PDF of the random vector $\boldsymbol{X}_{\overline{\mathcal{A}}}$ conditioned on $\boldsymbol{X}_{\mathcal{A}}$; $\mathcal{A} = \{i_1, \ldots, i_k\}$ and $\overline{\mathcal{A}} = \{j_1, \ldots,\, j_l\}$ form a partition of $\{1, \ldots,\, M\}$, that is, $\mathcal{A} \cup \overline{\mathcal{A}} = \{1, \ldots, M\}$ and $\mathcal{A} \cap \overline{\mathcal{A}} = \emptyset$. Using (5), $f_{\overline{\mathcal{A}}|\mathcal{A}}$ can be expressed as

$$
\begin{aligned}
f_{\overline{\mathcal{A}}|\mathcal{A}}(\boldsymbol{x}_{\overline{\mathcal{A}}}|\boldsymbol{x}_{\mathcal{A}}) =& c_{\overline{\mathcal{A}}|\mathcal{A}}(F_{j_1|\mathcal{A}}(x_{j_1}|\boldsymbol{x}_{\mathcal{A}}),\, F_{j_2|\mathcal{A}}(x_{j_2}|\boldsymbol{x}_{\mathcal{A}}), \ldots,\, F_{j_l|\mathcal{A}}(x_{j_l}|\boldsymbol{x}_{\mathcal{A}}))\times \\
&\times \prod_{j\in\overline{\mathcal{A}}} f_{j|\mathcal{A}}(x_j|\boldsymbol{x}_{\mathcal{A}}),
\end{aligned}
\tag{14}
$$

where $c_{\overline{\mathcal{A}}|\mathcal{A}}$ is an l-copula density – that of the conditional random variables $(X_{j_1|\mathcal{A}},\, X_{j_2|\mathcal{A}}, \ldots,\, X_{j_l|\mathcal{A}})^{\mathrm{T}}$ – and $f_{j|\mathcal{A}}$ is the conditional PDF of X_j given $\boldsymbol{X}_{\mathcal{A}}$, $j \in \overline{\mathcal{A}}$. Following Joe (1996), the univariate conditional distributions $F_{j|\mathcal{A}}$ can be further expressed in terms of any conditional pair copula $C_{ji|\mathcal{A}\setminus\{i\}}$ between $X_{j|\mathcal{A}\setminus\{i\}}$ and $X_{i|\mathcal{A}\setminus\{i\}}$, $i \in \mathcal{A}$:

$$
F_{j|\mathcal{A}}(x_j|\boldsymbol{x}_{\mathcal{A}}) = \frac{\partial C_{ji|A\setminus\{i\}}(u_j,\, u_i)}{\partial u_i}\Big|_{(F_{j|\mathcal{A}\setminus\{i\}}(x_j|\boldsymbol{x}_{\mathcal{A}\setminus\{i\}}),\, F_{i|\mathcal{A}\setminus\{i\}}(x_i|\boldsymbol{x}_{\mathcal{A}\setminus\{i\}}))}.
\tag{15}
$$

An analogous relation readily follows for conditional densities:

$$
\begin{aligned}
f_{j|\mathcal{A}}(x_j|\boldsymbol{x}_{\mathcal{A}}) =& \frac{\partial F_{j|\mathcal{A}}(x_j|\boldsymbol{x}_{\mathcal{A}})}{\partial x_j} \\
=& c_{ji|A\setminus\{i\}}(F_{j|\mathcal{A}\setminus\{i\}}(x_j|\boldsymbol{x}_{\mathcal{A}\setminus\{i\}}),\, F_{i|\mathcal{A}\setminus\{i\}}(x_i|\boldsymbol{x}_{\mathcal{A}\setminus\{i\}}))\times \\
&\times\, f_{j|\mathcal{A}\setminus\{i\}}(x_j|\boldsymbol{x}_{\mathcal{A}\setminus\{i\}}).
\end{aligned}
\tag{16}
$$

Substituting iteratively (15)-(16) into (14), Bedford and Cooke (2002) expressed $f_{\boldsymbol{X}}$ as a product of pair copula densities multiplied by $\prod_i f_i$. Recalling (5), it readily follows that the associated joint copula density c can be factorised into pair copula densities. Copulas expressed in this format are called vine copulas.

The factorisation is not unique: the pair copulas involved in the construction depend on the variables chosen in the conditioning equations (15)-(16) at each iteration. To organise them, Bedford and Cooke (2002) introduced a graphical model called the regular vine (R-vine). An R-vine among M random variables is represented by a graph consisting of $M-1$ trees $T_1,\, T_2, \ldots,\, T_{M-1}$, where each tree T_i consists of a set N_i of nodes and a set E_i of edges $e = (j, k)$ between nodes j and k. The trees T_i satisfy the following three conditions:

1. Tree T_1 has nodes $N_1 = \{1, \ldots, \mathrm{M}\}$ and $M-1$ edges E_1
2. for $i = 2, \ldots,\, M-1$, the nodes of T_i are the edges of T_{i-1} : $N_i = E_{i-1}$
3. Two edges in tree T_i can be joined as nodes of tree T_{i+1} by an edge only if they share a common node in T_i (proximity condition)

To build an R-vine with nodes $\mathscr{N} = \{N_1, \ldots, N_{M-1}\}$ and edges $\mathscr{E} = \{E_1, \ldots, E_{M-1}\}$, one defines for each edge e linking nodes $j = j(e)$ and $k = k(e)$ in tree T_i, the sets $I(e)$ and $D(e)$ as follows:

- If $e \in E_1$ (edge of tree T_1), then $I(e) = \{j,\, k\}$ and $D(e) = \emptyset$,
- If $e \in E_i$, $i \geq 2$, then $D(e) = D(j) \cup D(k) \cup (I(j) \cap I(k))$ and $I(e) = (I(j) \cup I(k)) \setminus D(e)$.

$I(e)$ contains always two indices j_e and k_e, while $D(e)$ contains $i-1$ indices for $e \in E_i$. One then associates each edge e with the conditional pair copula $C_{j_e,k_e|D(e)}$ between X_{j_e} and X_{k_e}

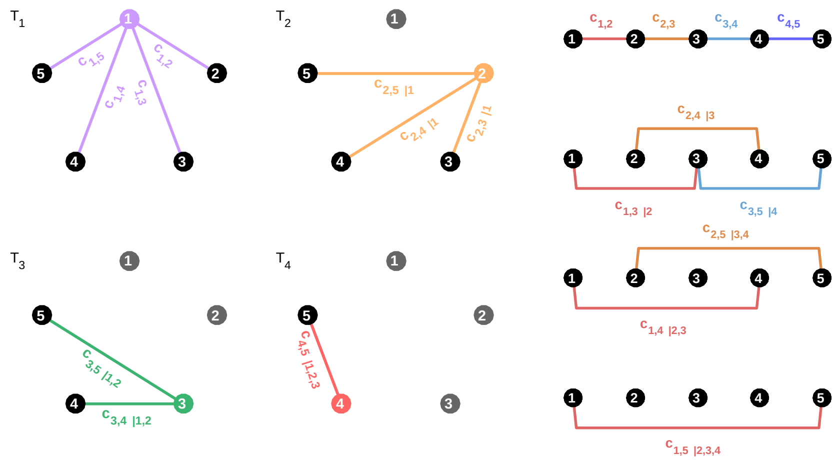

Figure 1: **Graphical representation of C- and D-vines.** The pair copulas in each tree of a 5-dimensional C-vine (left; conditioning variables are shown in grey) and of a 5-dimensional D-vine (right; conditioning variables are those between the connected nodes).

conditioned on the variables with indices in $D(e)$. An R-vine copula density with M nodes can thus be expressed as Aas (2016)

$$c(\boldsymbol{u}) = \prod_{i=1}^{M-1} \prod_{e \in E_i} c_{j_e,k_e|D(e)}(u_{j_e|D(e)}, u_{k_e|D(e)}). \tag{17}$$

Two special classes of R-vines are the drawable vine (D-vine, (Kurowicka and Cooke, 2005)) and the canonical vine (C-vine, (Aas et al., 2009)). Denoting $F(x_i) = u_i$ and $F_{i|\mathcal{A}}(x_i|\boldsymbol{x}_{\mathcal{A}}) = u_{i|\mathcal{A}}$, $i \notin \mathcal{A}$, a C-vine density is given by the expression

$$c(\boldsymbol{u}) = \prod_{j=1}^{M-1} \prod_{i=1}^{M-j} c_{j,j+i|\{1,\ldots,j-1\}}(u_{j|\{1,\ldots,j-1\}}, u_{j+i|\{1,\ldots,j-1\}}), \tag{18}$$

while a D-vine density is expressed as

$$c(\boldsymbol{u}) = \prod_{j=1}^{M-1} \prod_{i=1}^{M-j} c_{i,i+j|\{i+1,\ldots,i+j-1\}}(u_{i|\{i+1,\ldots,i+j-1\}}, u_{i+j|\{i+1,\ldots,i+j-1\}}). \tag{19}$$

The graphs associated to a 5-dimensional C-vine and to a 5-dimensional D-vine are shown in Figure 1. Note that this simplified illustration differs from the standard one introduced in Aas et al. (2009) and commonly used in the literature.

2.5 Vine inference in practice

Building a vine copula model that properly describes the dependencies among the inputs involves the following steps:

1. selecting the structure of the vine (for C- and D-vines: selecting the order of the nodes);

2. modelling each pair copula in the vine by a suitable parametric family (based on expert knowledge, when available, or learning from data);
3. assigning the copula parameters (from prior knowledge, or by fitting to data).

Steps 1-2 solve the representation problem, by providing a parametric model of the input dependencies. Step 3 uniquely determines the copula to be assigned to the inputs. We restrict our attention to the case where expert knowledge is not available and the vine has to be learned entirely from available data. Aas et al. (2009) provided algorithms to compute the likelihood of a C- or D-vine model given a sample $\{\hat{\boldsymbol{x}}^{(1)}, \ldots, \hat{\boldsymbol{x}}^{(n)}\}$ of observations. Joe (2015) presented a likelihood estimation algorithm for general R-vines. These algorithms enable, for a given parametric model (that is, once the vine structure and comprising pair copula families have been selected), parameter estimation (step 3) based on maximum likelihood.

The estimation could then in principle be iterated across all possible structures (step 1) and pair copula families (step 2) to find the most likely model describing the observed dependence properties. The number of possibilities to loop across, however, is extremely large: an M-copula density admits $2^{\binom{M-2}{2}-1}M!$ different R-vine factorisations (Morales-Nápoles, 2011), $M!$ of which are C- or D- vines. This approach is thus computationally demanding in the presence of even a moderate number of inputs. In the case studies examined in this work we take a different approach, originally proposed by Aas et al. (2009) and commonly preferred in applications, and first solve step 1 separately. The optimal vine structure is found heuristically by ordering the variables X_i such that pairs (X_i, X_j) with the strongest dependence are captured first, *i.e.*, fall in the first trees of the vine. The Kendall's tau $\tau_{K;ij}$ defined in (7) is taken as the measure of dependence. For a C-vine, this means selecting the central node in tree T_1 as the variable X_{i_1} that maximises $\sum_{j \neq i_1} \tau_{K;i_1 j}$, then the node of tree T_2 as the variable X_{i_2} which maximises $\sum_{j \notin \{i_1, i_2\}} \tau_{K;i_2 j}$, and so on. For a D-vine, this means ordering the variables $X_{i_1}, X_{i_2}, \ldots, X_{i_M}$ in the first tree so as to maximise $\sum_{k=1}^{M-1} \tau_{K;i_k i_{k+1}}$, which we solve as an open travelling salesman problem (OTSP) (Applegate et al., 2006). An open source Matlab implementation of a genetic algorithm to solve the OTSP is provided in Kirk (2014). An algorithm to find the optimal structure for R-vines has been proposed in Dißmann et al. (2013).

Once the vine structure has been selected, steps 2 and 3 are solved together by an iterative procedure. For each pair copula composing the vine, and for each parametric families allowed for that copula, the parameters of the family are fitted to the available data based on maximum likelihood (other approaches, such as Bayesian estimation, may be followed (Gruber and Czado, 2015)). The parametric family which best fits the data is then chosen as the family that minimizes the Akaike information criterion (AIC)

$$\mathrm{AIC} = -2 \log L + 2k,$$

where k is the number of parameters of the pair copula and $\log L$ is its log-likelihood. The AIC penalises models with a larger number of parameters (which typically yield higher likelihood and would otherwise be preferred), thus preventing overfitting. Alternatives to the AIC have been proposed, for instance the Bayesian information criterion (BIC) and the copula information criterion (CIC; (Grønneberg and Hjort, 2014)). Also, one may alternatively opt for various goodness of fit tests (Schepsmeier, 2015; Fermanian, 2012). We did not consider these different approaches here. For a comparison of some of them, see Manner (2007). Optionally, once each pair copula has been separately selected by this iterative approach (sequential fitting), the selected pair-copula families are retained and the parameters of the vine are globally fitted to the data. This step, however, may be computationally very demanding if M is large.

To facilitate inference we rely on the commonly used *simplifying assumption* that the pair copulas $C_{j(e),k(e)|D(e)}$ in (17) only depend on the variables with indices in $D(e)$ through the arguments $F_{i(e)|D(e)}$ and $F_{j(e)|D(e)}$ (Czado, 2010). While being exact only in particular cases, this assumption is usually not severe (Haff et al., 2010). In Stöber et al. (2013) construction techniques for non-simplified vine copulas were proposed.

Table 11 shows the list of the 19 simplified pair copula families used for vine copula construction in this study, and implemented in the VineCopulaMatlab package by Kurz (2015). A summary of their properties is reported in Table 12. In addition to these copulas, their rotated versions were also considered. A rotation by 180° transforms a copula into its survival version. A rotation by 90° or 270° implements negative dependencies. Including the rotated copulas, 62 families were considered in total for inference.

3 Vine representations for UQ methods assuming independent inputs

Some advanced UQ techniques require or benefit from inputs $\boldsymbol{X}$ with independent components. For instance, PCE (Section 4.2) exploits independence to build a basis of polynomials orthonormal with respect to $F_{\boldsymbol{X}}$ by tensor product. This in turn simplifies the construction of a metamodel that expresses Y as a polynomial of the inputs. FORM and SORM (Section 4.3), as well as other reliability methods, take advantage of the probability measure of the standard normal space to approximate low probability mass regions. When the components of $\boldsymbol{X}$ are mutually dependent but independence is needed, it is therefore custom to transform $\boldsymbol{X}$ into a vector $\boldsymbol{Z}$ with independent components. The transformation $\mathcal{T}$ that performs this mapping thus changes the copula $C_{\boldsymbol{X}}$ of $\boldsymbol{X}$ into the independence copula $C^{(\Pi)}$ defined in (11). When $\mathcal{T}$ also makes $F_{\boldsymbol{Z}}$ rotationally invariant, it is called an isoprobabilistic transform. This section discusses existing isoprobabilistic transformations, relates them to copula theory, and highlights the existence of algorithms for their computation when $C_{\boldsymbol{X}}$ is expressed as an R-vine. By doing so, we demonstrate that vine copulas provide effective models of complex input dependencies also in combination with UQ approaches designed for independent inputs.

3.1 Compositional models for dependent inputs

Consider a generic UQ method that works in a probability space where input parameters are independent and have marginal distributions G_i. Assume that the input $\boldsymbol{X}$ to the model $\mathcal{M}$ has a joint CDF $F_{\boldsymbol{X}}$ for which an invertible isoprobabilistic transform $\mathcal{T} : \boldsymbol{X} \mapsto \boldsymbol{Z}$ is known, and that $\mathcal{T}^{-1} : \boldsymbol{Z} \mapsto \boldsymbol{X}$ is also known. Then, the system response $Y = \mathcal{M}(\boldsymbol{X})$ can be expressed as a function of $\boldsymbol{Z}$ by

$$Y = (\mathcal{M} \circ \mathcal{T}^{-1})(\boldsymbol{Z}). \tag{20}$$

The *compositional model* $\mathcal{M} \circ \mathcal{T}^{-1}$ can be seen as a black box model which combines the known map $\mathcal{T}^{-1}$ with the original computational model $\mathcal{M}$. The UQ method of choice can then be applied on the input $\boldsymbol{Z}$ and the model $\mathcal{M} \circ \mathcal{T}^{-1}$: the statistics of the output of $\mathcal{M} \circ \mathcal{T}^{-1}$ in response to $\boldsymbol{Z}$ are identical to the statistics of the output of $\mathcal{M}$ in response to $\boldsymbol{X}$.

Given $F_{\boldsymbol{X}}$ and $\mathcal{M}$, determining the compositional model requires then to determine $\mathcal{T}^{-1}$, which depends on $F_{\boldsymbol{X}}$. However, a general closed form expression for $\mathcal{T}^{-1}$ is in most cases unknown, even when $F_{\boldsymbol{X}}$ is known. This problem is associated exclusively to the copula $C_{\boldsymbol{X}}$ of $\boldsymbol{X}$, and not to its marginals F_i. Indeed, $\boldsymbol{X}$ can be mapped by the transformation $\mathcal{T}^{(\mathcal{U})}$ defined in (4) onto $\boldsymbol{U} \sim U([0,\, 1]^M)$, whose joint CDF is $C_{\boldsymbol{X}}$. Thus, one can always write

$$\mathcal{T} = \mathcal{T}^{(\mathcal{U})} \circ \mathcal{T}^{(\Pi)}, \tag{21}$$

where $\mathcal{T}^{(\mathcal{U})}$ – which depends on the marginals only – is known for a given $F_{\boldsymbol{X}}$, while $\mathcal{T}^{(\Pi)} : \boldsymbol{U} \mapsto \boldsymbol{Z}$ is to be determined.

3.2 Isoprobabilistic transforms and copulas

The most general isoprobabilistic transform $\mathcal{T}$, valid for any continuous $F_{\boldsymbol{X}}$, is the Rosenblatt transform (Rosenblatt, 1952), which reads

$$\mathcal{T}_1^{(\mathcal{R})} : \boldsymbol{X} \mapsto \boldsymbol{W}, \text{ where } \begin{cases} W_1 = F_1(X_1) \\ W_2 = F_{2|1}(X_2|X_1) \\ \vdots \\ W_M = F_{M|1,\ldots,M-1}(X_M|X_1,\ldots,\, X_{M-1}) \end{cases} . \tag{22}$$

Following (21), and as first noted in Lebrun and Dutfoy (2009a), one can rewrite $\mathcal{T}_1^{(\mathcal{R})} = \mathcal{T}_1^{(\Pi,\,\mathcal{R})} \circ \mathcal{T}^{(\mathcal{U})}$, where

$$\mathcal{T}_1^{(\Pi,\,\mathcal{R})} : \boldsymbol{U} \mapsto \boldsymbol{W}, \text{ with } W_i = C_{i|1,\ldots,i-1}(U_i|U_1,\ldots,U_{i-1}). \tag{23}$$

Here, $C_{i|1,\ldots,i-1}$ are conditional copulas of $\boldsymbol{X}$ (and therefore of $\boldsymbol{U}$), obtained from $C_{\boldsymbol{X}}$ by differentiation. The problem of obtaining an isoprobabilistic transform of $\boldsymbol{X}$ is thus reduced to the problem of computing derivatives of $C_{\boldsymbol{X}}$.

The variables W_i are mutually independent and marginally uniformly distributed in $[0, 1]$. To assign W_i any other marginal distribution Ψ_i, one can define the generalised Rosenblatt transform as a map $\mathcal{T}^{(\mathcal{R})} = \mathcal{T}_2^{(\mathcal{R})} \circ \mathcal{T}_1^{(\mathcal{R})} = \mathcal{T}_2^{(\mathcal{R})} \circ \mathcal{T}_1^{(\Pi,\, \mathcal{R})} \circ \mathcal{T}^{(\mathcal{U})}$, where

$$\mathcal{T}_2^{(\mathcal{R})} : \boldsymbol{W} \mapsto \boldsymbol{Z}, \text{ with } Z_i = \Psi_i^{-1}(W_i). \tag{24}$$

When $\Psi_i = F_i$ for all i, *i.e.*, $\mathcal{T}_2^{(\mathcal{R})} \equiv \left(\mathcal{T}^{(\mathcal{U})}\right)^{-1}$, $\mathcal{T}^{(\mathcal{R})}$ maps $\boldsymbol{X}$ onto a random vector $\boldsymbol{Z}$ with same marginals but independent components.

Each continuous joint CDF $F_{\boldsymbol{X}}$ defines multiple transforms of the type (23), one per permutation of the indices $\{1, \ldots, M\}$. However, these transforms involve conditional probabilities which are not generally available in closed form. A notable exception is the multivariate Gaussian distribution, where independence can be obtained by diagonalisation of the correlation matrix (*e.g.*, by Choleski decomposition). The Rosenblatt transform in this case (and in this case only, see Lebrun and Dutfoy (2009a)) is equivalent to the Nataf transform (Nataf, 1962), which is commonly used in engineering applications.

A generalized Nataf transform for elliptical copulas was proposed in Lebrun and Dutfoy (2009b). The generalization enables the mapping of random vectors with elliptical copulas into their standard spherical representative, having uncorrelated (but not mutually independent, except for the Gaussian case) components with elliptical, unit variance marginal distributions. Adopting the generalized Nataf transform instead of the Rosenblatt transform for inputs with non-elliptical copulas, of for inputs with elliptical copulas when a transformation to independent components is needed, may cause non-negligible errors on the estimates computed by UQ methods.

3.3 Rosenblatt transform and resampling for R-vines

Aas et al. (2009) provided algorithms to compute the Rosenblatt transform (23) and its inverse when $C_{\boldsymbol{X}}$ is a given C- or D-vine. Given the pair-copulas C_{ij} in the first tree of the vine, the algorithms first compute their derivatives $C_{i|j}$. Higher-order derivatives $C_{i|ijk}$, $C_{i|ijkh}, \ldots$ are obtained from the lower-order ones and their inverses by iteration. The derivatives of continuous pair copulas are available analytically in few cases (see, *e.g.*, Schepsmeier and Stöber (2014)) and numerically otherwise. Since these functions are monotone increasing distributions, their inverses are numerically cheap to compute by rootfinding, when not available analytically. An algorithm for the computation of the Rosenblatt and inverse Rosenblatt transforms for general R-vines was proposed in Schepsmeier (2015). These algorithms can be trivially implemented so as to process n samples in parallel.

In addition, $(\mathcal{T}^{(\mathcal{R})})^{-1}$ allows to sample from the vine model by transforming independent points uniformly distributed in $[0,\, 1]^M$. Space filling samples in the probability space can be obtained analogously (*e.g.*, by Sobol sequences or Latin Hypercube sampling, (McKay et al., 1979)). Given in particular a vine model of the input dependencies (obtained, *e.g.*, from expert knowledge or by inference from available data), the inverse Rosenblatt transform enables resampling from this model.

4 UQ for mutually dependent inputs

After recalling convergence properties of MC estimates, we summarise here three established UQ methods used in the numerical experiments carried out in Sections 5 and 6: PCE, FORM, and IS. Several other methods exist to solve the same problems, and we do not advocate for the ones considered here over others. Importantly, the framework demonstrated for these three methods extend to basically any UQ technique designed for problems with a finite number of coupled inputs.

PCE is a spectral method that expresses the system response as a polynomial of the input variables. It is used to estimate moments of the response, to compute sensitivity indices, or to perform resampling efficiently. FORM is a reliability analysis method designed to approximate small failure probabilities P_f numerically. IS is a stochastic sampling method that combines FORM with MC to obtain more robust estimates of P_f. When the computational budget is limited and only few runs of the computational model can be afforded, these methods provide significantly better estimates of their target statistics than MCS with the same number of observations. However, these methods strongly rely on an accurate representation of the input dependencies. Besides, some of them strongly benefit from the possibility of mapping the input random vector onto a vector with independent components.

Here we describe how to combine these methods with the vine representation of the input CDF illustrated in Section 2. The flexibility of R-vine models expands the applicability of these methods drastically.

4.1 Convergence of MC estimates

MC (or sample) estimates of a statistic $\eta = \eta(Y)$ are obtained as functions $\hat{\eta}_n = \hat{\eta}_n(Y)$ of n i.i.d realisations $\{\hat{y}^{(j)}\}_{j=1}^{n}$ of Y. Three statistics considered in the applications in Sections 5-6 are the mean $\mu(Y)$ of Y, its standard deviation $\sigma(Y)$, and failure probabilities of the type $P_{f;y^*}(Y) = \mathbb{P}(Y \geq y^*)$, where y^* is a critical threshold (see Table 1, first row). Their sample estimators are the sample mean $\hat{\mu}_n(Y)$, the corrected sample standard deviation $\hat{\sigma}_n(Y)$, and the sample survival function evaluated at y^*, $\hat{P}_{f;y^*,n}(Y)$. Their analytical expression is given in Table 1, second row.

If $\hat{\eta}_n(Y)$ is an unbiased estimator of $\eta(Y)$ and $\eta(Y) \neq 0$, the reliability of $\hat{\eta}_n(Y)$ can be quantified by its coefficient of variation (CoV), given by

$$\mathrm{CoV}(\hat{\eta}_n(Y)) = \frac{\sigma(\hat{\eta}_n(Y))}{\mu(\hat{\eta}_n(Y))} = \frac{\sigma(\hat{\eta}_n(Y))}{\eta(Y)}.$$

$\eta(\cdot)$:	$\mu(Y) = \int_{\mathbb{R}} y f_Y(y) dy$	$\sigma(Y) = \sqrt{\int_{\mathbb{R}} (y - \mu(Y))^2 f_Y(y) dy}$	$P_{f;y^*}(Y) = \int_{\{Y \geq y^*\}} f_Y(y) dy$
$\hat{\eta}_n(\cdot)$:	$\hat{\mu}_n(Y) = \frac{1}{n} \sum_j \hat{y}^{(j)}$	$\hat{\sigma}_n(Y) = \sqrt{\frac{1}{n-1} \sum_j \left(\hat{y}^{(j)} - \hat{\mu}_n(Y)\right)^2}$	$\hat{P}_{f;y^*}(Y) = \frac{1}{n} \sum_j \mathbf{1}_{\{\hat{y}^{(j)} > y^*\}}$
$\mathrm{CoV}(\hat{\eta}_n)$:	$\frac{\sigma(Y)}{\mu(Y)} \frac{1}{\sqrt{n}}$	$\approx \frac{1}{\sqrt{2n}}$	$\sqrt{\frac{1 - P_f}{n P_f}}$

Table 1: **Some MC sample estimates and their CoV**. The first row of the table defines the mean, standard deviation, and failure probability of the random variable Y. The second row shows their sample estimators, and the bottom row the CoV of such estimators (exact for $\hat{\sigma}_n(Y)$ only if Y is normally distributed).

It is common in engineering applications to accept estimates whose CoV is not larger than 0.1 (10%). The CoV of all statistics in Table 1, third row (approximate for $\hat{\sigma}_n(Y)$) is proportional to $1/\sqrt{n}$ and thus decays to 0 as n increases, however at a slow pace. The expression for $\mathrm{CoV}(\hat{\sigma}_n(Y))$ is obtained from the fact that $\sigma(\hat{\sigma}_Y) = \sigma(Y)/\sqrt{2N + O(N^2)}$ if Y is normally distributed (see Romanovsky (1925)).

4.2 Polynomial Chaos Expansion

PCE is a spectral method that represents a model $\mathcal{M}$ of finite variance as a linear sum of orthogonal polynomials (Ghanem and Spanos, 2003; Xiu and Karniadakis, 2002). As such, the parameters of the resulting representation have a statistical interpretation. For instance, the first two moments of the PCE model are encoded in the coefficients of the obtained polynomial. The model is also computationally cheap to evaluate, enabling an efficient evaluation of other global statistics of Y (higher order moments, the PDF, *etc.*) that would otherwise require an excessive number of runs of $\mathcal{M}$.

Building a PCE representation of the output is relatively simple, as recalled below, for independent inputs. For this reason, if $\boldsymbol{X}$ has non mutually independent components, it is convenient to first map it onto such a vector $\boldsymbol{Z}$ by an isoprobabilistic transform. Modelling the copula $C_{\boldsymbol{X}}$ of $\boldsymbol{X}$ as an R-vine provides the Rosenblatt transform (23)-(24) needed to this end.

Given an isoprobabilistic transform $\mathcal{T}$ such that $\boldsymbol{Z} = \mathcal{T}(\boldsymbol{X})$, it follows that $Y = (\mathcal{M} \circ \mathcal{T}^{-1})(\boldsymbol{Z}) = \mathcal{M}'(\boldsymbol{Z})$. In the following, Y is assumed to have finite variance. The PCE of $Y = \mathcal{M}'(\boldsymbol{Z})$ is defined as

$$Y = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} y_\alpha \Psi_{\boldsymbol{\alpha}}(\boldsymbol{Z}), \tag{25}$$

where the $\Psi_{\boldsymbol{\alpha}}$ are multivariate polynomials orthonormal with respect to $f_{\boldsymbol{Z}}$, *i.e.*,

$$\int_{\mathcal{D}_{\boldsymbol{Z}}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{z}) \Psi_{\boldsymbol{\beta}}(\boldsymbol{z}) f_{\boldsymbol{Z}}(\boldsymbol{z}) d\boldsymbol{z} = \delta_{\boldsymbol{\alpha\beta}}.$$

Here, $\delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ is the Kroenecker delta symbol.

Since $\boldsymbol{Z}$ has independent components, each $\Psi_{\boldsymbol{\alpha}}$ can be obtained as a tensor product of M univariate polynomials $\phi^{(i)}_{\alpha_i}(x_i)$ orthonormal with respect to the marginals g_i of Z_i:

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{z}) = \prod_{i=1}^{M} \phi^{(i)}_{\alpha_i}(z_i).$$

The polynomial basis is guaranteed to exist if the marginals distributions all have finite moments of any order. A unique representation exists if additionally the marginals are uniquely represented by the sequence of their moments. For details, as well as for sufficient conditions that guarantee uniqueness, see Ernst et al. (2012). For instance, the $\phi^{(i)}_{\alpha_i}$ are Hermite polynomials if Z_i is standard normal, *i.e.*, if $g_i(z) = \varphi(z) = \exp(-z^2/2)/\sqrt{2\pi}$. In the applications illustrated in Section 5 we work with this choice, although other choices may be favoured in different applications. An investigation of optimal choices for the marginal distributions of $\boldsymbol{Z}$ is an open question that will be investigated in a future study. Classical families of polynomials are described in Xiu and Karniadakis (2002).

The sum in (25) comprises an infinite number of terms. For practical purposes, it is truncated to a finite sum (Marelli and Sudret, 2017). Given a truncation scheme and the corresponding set $\mathcal{A}$ of multi-indices, the coefficients $y_{\boldsymbol{\alpha}}$ in

$$Y_{PC}(\boldsymbol{Z}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} y_{\alpha} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{Z}) \tag{26}$$

are evaluated on a set $\{(\hat{\boldsymbol{z}}^{(j)} = \mathcal{T}^{-1}(\hat{\boldsymbol{x}}^{(j)}),\, \hat{y}^{(j)}\}_{j=1}^{n}$ of observations (the *experimental design*). Many strategies exist to accomplish this task, such as projection methods based on Gaussian (Le Maître et al., 2001) or sparse quadrature (Keese and Matthies, 2003; Xiu, 2010), least-squares minimisation (Berveiller et al., 2006), and different adaptive sparse methods (Doostan and Owhadi, 2011; Jakeman et al., 2015), hybridised into a single methodology by (Blatman and Sudret, 2011). The latter is particularly suitable when M is large, because it achieves a sparse basis out of a very large initial set of possible polynomials, and is therefore the method of choice in this study.

Once a PCE metamodel (26) of the compositional model $\mathcal{M}'$ is built, the first two moments of the response are encoded in the coefficients of the expansions. Indeed, due to orthonormality of the polynomial basis,

$$\mu(Y_{PC}) = y_{\boldsymbol{0}}, \quad \sigma^2(Y_{PC}) = \sum_{\alpha \in \mathcal{A} \setminus \{\boldsymbol{0}\}} y^2_{\boldsymbol{\alpha}}. \tag{27}$$

Higher-order moments, as well as other statistics, can be efficiently estimated by simulation.

4.3 First order reliability method

Let $Y = \mathcal{M}(\boldsymbol{X})$ be the uncertain, scalar output of the computational model $\mathcal{M}$ in response to an uncertain M-variate input $\boldsymbol{X}$ with joint CDF $F_{\boldsymbol{X}}$, joint PDF $f_{\boldsymbol{X}}$ and domain $\mathcal{D}_{\boldsymbol{X}}$.

Suppose that the system fails if $Y \geq y^*$, where y^* is a critical threshold. The failure condition is usually rewritten as $g(\boldsymbol{x}) \leq 0$, with $g(\boldsymbol{x}) = y^* - \mathcal{M}(\boldsymbol{x})$.

Reliability analysis concerns the evaluation of the failure probability $P_f = \mathbb{P}(g(\boldsymbol{X}) \leq 0)$, *i.e.*, of the probability mass over the failure domain $\mathcal{D}_f = \{\omega : g(\boldsymbol{X}(\omega)) \leq 0\}$. $\mathcal{D}_f$ is typically known only implicitly, preventing a direct estimation of P_f.

Using the indicator function

$$\mathbf{1}_{\mathcal{D}_f}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } g(\boldsymbol{x}) \leq 0 \\ 0 & \text{if } g(\boldsymbol{x}) > 0 \end{cases},$$

P_f can be expressed as

$$P_f = \int_{\mathcal{D}_{\boldsymbol{X}}} \mathbf{1}_{\mathcal{D}_f}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} = \mu(\mathbf{1}_{\mathcal{D}_f}(\boldsymbol{X})), \tag{28}$$

where $\mu(\cdot)$ is the mean operator with respect to $f_{\boldsymbol{X}}$.

If $\boldsymbol{X}$ is multivariate normal with independent components, FORM (Hasofer and Lind, 1974; Hohenbichler and Rackwitz, 1983) approximates $\mathcal{D}_f$ with an hyperplane tangent to the limit-state surface $\{\omega : g(\boldsymbol{X}(\omega)) = 0\}$ in its point closest to the origin (the *design point* $\boldsymbol{x}^*$). The rationale is that the standard normal density is a fast decaying function of the norm of its argument, so that – assuming the uniqueness of the design point – the probability mass of $\mathcal{D}_f$ concentrates around $\boldsymbol{x}*$ (see also Der Kiureghian and Liu (1986)).

If $\boldsymbol{X}$ is not multivariate normal, but has a normal copula, the Nataf transform (which is equivalent to the Rosenblatt transform in the Gaussian case, see Lebrun and Dutfoy (2009a) and Section 3.3) is used map $\boldsymbol{X}$ into a standard normal random vector $\boldsymbol{Z} = \mathcal{T}(\boldsymbol{X})$, and FORM can then be used to search for the design point $\boldsymbol{z}^*$ in the standard normal space. If $\boldsymbol{X}$ has a more general elliptical copula, the generalized Nataf transform can be employed to map $\boldsymbol{X}$ into a vector $\boldsymbol{Z}$ whose components are uncorrelated (but not independent) and have elliptical marginals with unit variance (Lebrun and Dutfoy, 2009b). In this case, the probability density of $\boldsymbol{Z}$ is again a rapidly decreasing function of the norm of its argument, and FORM can be used analogously to the standard normal case.

If $\boldsymbol{X}$ has a non-elliptical copula, employing the generalized Nataf transform would yield biased estimates of the failure probability. One has to resort to different isoprobabilistic transformations, the most general being the Rosenblatt transform, to map $\boldsymbol{X}$ into a standard normal (or into a spherical elliptical) random vector $\boldsymbol{Z} = \mathcal{T}(\boldsymbol{X})$, thus reconducting the problem to one of the two cases above. To treat this more general case, we express $F_{\boldsymbol{X}}$ in terms of its marginals F_i and of its copula $C_{\boldsymbol{X}}$ as in (1), and we model $C_{\boldsymbol{X}}$ as an R-vine (see Section 2.4). The Rosenblatt transform of the latter is available (see Section 3.3), allowing $\boldsymbol{X}$ to be mapped onto the standard normal space, where the classical version of FORM can be used.

4.4 Importance sampling

In the context of reliability analysis, IS is used to combine the convergence speed of FORM with the robustness of MC sampling. For a general, non-standard-normal input vector $\boldsymbol{X}$ admitting Rosenblatt transform $\boldsymbol{Z} = \mathcal{T}^{(\mathcal{N})}(\boldsymbol{X})$, (28) can be recast as

$$P_f = \int_{\mathbb{R}^M} \mathbf{1}_{\mathcal{D}_f} \left(\mathcal{T}^{-1}(\boldsymbol{z})\right) \frac{\varphi_M(\boldsymbol{z})}{\psi(\boldsymbol{z})} \psi(\boldsymbol{z}) d\boldsymbol{z} = \mu_\psi \left(\mathbf{1}_{\mathcal{D}_f} \left(\mathcal{T}^{-1}(\boldsymbol{Z})\right) \frac{\varphi_M(\boldsymbol{Z})}{\psi(\boldsymbol{Z})} \right), \tag{29}$$

where $\varphi_M(\cdot)$ is the M-variate standard normal density, $\psi(\cdot)$ is a suitable M-variate density (the *importance density*) and μ_ψ is the mean operator with respect to ψ. Melchers (1999) recommends to assign ψ as the standard normal density centered at the design point found by FORM: $\psi(\boldsymbol{z}) = \varphi_M(\boldsymbol{z} - \boldsymbol{z}^*)$.

Given a sample $\{\hat{\boldsymbol{z}}^{(1)}, \ldots, \hat{\boldsymbol{z}}^{(n)}\}$ of $\psi(\boldsymbol{Z})$, the IS estimator of P_f is the sample estimator

$$\hat{P}_{f;\mathrm{IS}} = \frac{1}{N} \exp(||\boldsymbol{z}^*||^2/2) \sum_{j=1}^{n} \mathbf{1}_{\mathcal{D}_f} \left(\mathcal{T}^{-1}(\hat{\boldsymbol{z}}^{(j)})\right) \exp(-\hat{\boldsymbol{z}}^{(j)} \cdot \boldsymbol{z}^*),$$

which has variance

$$\sigma^2(\hat{P}_{f;\mathrm{IS}}) \approx \frac{1}{n(n-1)} \sum_{j=1}^{n} \left(\mathbf{1}_{\mathcal{D}_f} \left(\mathcal{T}^{-1}(\hat{\boldsymbol{z}}^{(j)})\right) \frac{\varphi_M(\hat{\boldsymbol{z}}^{(j)})}{\varphi_M(\hat{\boldsymbol{z}}^{(j)} - \boldsymbol{z}^*)} - \hat{P}_{f;\mathrm{IS}} \right)^2$$

and $\mathrm{CoV}(\hat{P}_{f;\mathrm{IS}}) \approx \sigma(\hat{P}_{f;\mathrm{IS}})/\hat{P}_{f;\mathrm{IS}}$. Since the latter is given in terms of $\hat{P}_{f;\mathrm{IS}}$, it is unknown until $\hat{P}_{f;\mathrm{IS}}$ is computed. One can progressively increase the sample size n until $\mathrm{CoV}(\hat{P}_{f;\mathrm{IS}})$ drops below a desired level.

5 Results on a horizontal truss model

We first demonstrated the analysis workflow developed above on a horizontal truss model. Estimates based on advanced and computationally efficient UQ techniques were compared to reference MCS estimates. Earlier work on vine representations combined with MCS can be found in Wang and Li (2017b), limited to reliability analysis.

The analysis was run in Matlab (The Mathworks Inc., 2016). Specifically, the vine copula inference was performed using the open source package VineCopulaMatlab (Kurz, 2015). We enriched this toolbox with functionalities for the computation of the Rosenblatt and inverse Rosenblatt transforms of C- and D- vines, and for the calculation of the optimal D-vine structure on data, implemented as an open travelling salesman problem (Applegate et al., 2006). The UQ analyses were performed with the free Matlab-based software UQLab (Marelli and Sudret, 2014).

5.1 Computational model

The horizontal truss model, already used in Blatman and Sudret (2011), comprises 23 bars connected at 6 upper nodes, as shown in Figure 2. The structure is 24 meters long and 2

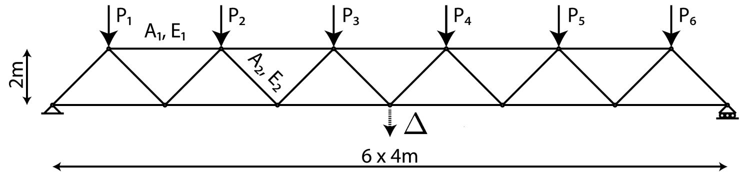

Figure 2: **Scheme of the horizontal truss model.** Modified from Blatman and Sudret (2011).

meters high. Six random loads P_1, P_2, …, P_6 were applied onto the structure, one on each upper node. As a result, the structure exhibited a downward vertical displacement at each node. The largest displacement Δ was always at the center. Excess displacement leads to failure.

5.2 Probabilistic input model

We considered the case of uncertain loads P_i, causing uncertainty in the output response Δ. The bar properties of the truss, differently from Blatman and Sudret (2011), were kept constant. The loads $\boldsymbol{X} = (P_1, P_2, \ldots, P_6)$ were modelled by assigning separately their marginals F_i and their copula $C_{\boldsymbol{X}}$. We fixed the marginals to univariate Gumbel CDFs with mean $\mu = 5 \times 10^4$ N and standard deviation $\sigma = 0.15\mu = 7.5 \times 10^3$ N:

$$F_i(x; \alpha, \beta) = e^{-e^{-(x-\alpha)/\beta}}, \quad x \in \mathbb{R}, \, i = 1, 2, \ldots, 6, \tag{30}$$

where $\beta = \sqrt{6}\sigma/\pi$, $\alpha = \mu - \gamma\beta$, and $\gamma \approx 0.5772$ is the Euler-Mascharoni constant.

We then investigated how different copulas affect the statistics of the truss response. First, we employed the independence copula $C^{(\Pi)}$ defined by (11), which implies independence among the loads.

Loads on a truss structure may be expected to be positively correlated: higher loads on one node increase the chance to have higher loads on other nodes (e.g. due to snow or traffic jam on a bridge). To account for this, we selected next a 6-dimensional Gaussian copula $C^{(\mathcal{N})}$ (12). We assigned the copula parameters ρ_{1j}, $j = 2, \ldots, 6$, such that the Spearman's correlation coefficients (6) would be $\rho_{S;1j} = 0.135$, resulting in $\rho_{1j} = 0.141$ (and Kendall's tau $\tau_{K;1j} \simeq 0.0904$). Besides, we set $\rho_{S;ij|1} = 0$ for each $i \neq j$, $i, j \neq 1$, so that all loads other than P_1 would be conditionally independent given P_1.

Beside being positively correlated, in a realistic scenario loads are likely to be upper tail dependent: an extremely large load on one node increases the chance to have large loads elsewhere (*e.g.*, when due to a heavy snowfall or to a traffic jam). Therefore, we last investigated a scenario with tail dependent loads (see Section 2.2). We modelled upper tail dependence by means of a C-vine $C^{(\mathcal{V})}$. We selected P_1 as the first node of $C^{(\mathcal{V})}$, and we set the pair copulas in the first tree to bivariate Gumbel-Hougaard copulas $C^{(\mathcal{GH})}_{1j}$ (13) between

P_1 and P_j, $j = 2, \ldots, 5$. We took the parameter $\theta_{1j} = 1.1$, $j = 2, \ldots, 6$, yielding Spearman's correlation coefficients $\rho_{S;1j} = 0.135$ as for the Gaussian copula. This choice resulted in $\tau_{K;1j} = 0.0909$ (close to the value determined by the Gaussian copula) but also determined an upper tail dependence coefficient $\lambda_{u;1,j} = 0.122$ between P_1 and P_j. We further set the other pair copulas of the vine, *i.e.*, all conditional pair copulas, to the independence copula, ensuring conditional independence of (P_i, P_j) given P_1 for each i, $j \neq 1$. The resulting vine $C^{(\mathcal{V})}$ had density

$$c^{(\mathcal{V})}(u_1, \ldots, u_6) = \prod_{j=2}^{6} c_{1j}^{(\mathcal{GH})}(u_1, u_j). \tag{31}$$

Note that other vine structures could have been used to model tail dependent loads: for instance, a D-vine whose first tree couples the loads on neighbouring nodes of the truss. Expert knowledge and available input data may provide guidance in this selection process.

We finally investigated the viability of the vine representation when the vine itself is not known and has to be fully inferred from data. To this end, we sampled $m = 300$ realisations from $C^{(\mathcal{V})}$ and learned from them the vine structure, its pair copula families and their parameters, as detailed in Section 2.5. The pair copulas were chosen among the parametric families listed in Table 11 and their rotated versions defined by (33). The inferred pair copulas comprising $\hat{C}^{(\mathcal{V})}$ are summarised in Table 2, along with their Kendall's tau and upper tail dependence coefficients. In real applications, the input observations needed for the inference procedure may be obtained by monitoring of the loads themselves, or may be estimated from available data (*e.g.*, weather or traffic conditions), and do not require any model evaluation. The resulting C-vine $\hat{C}^{(\mathcal{V})}$ had a different structure, only two of the five pair copulas in the first tree were of the Gumbel-Hougaard family, and one of the conditional copulas in the second tree was not the independence copula. All other pair copulas were correctly found to be independence copulas. Despite the differences from the true vine $C^{(\mathcal{V})}$, using $\hat{C}^{(\mathcal{V})}$ provided very good quality estimates of the statistics of the truss deflection, as shown below.

5.3 Analysis of the moments for different load couplings

For each probabilistic model of the loads, we analysed the mean $\mu(\Delta)$ and standard deviation $\sigma(\Delta)$ of the resulting system response by MCS and by PCE.

The MC estimates were computed as sample estimates on $\{\hat{\delta}_i = \mathcal{M}(\hat{\boldsymbol{x}}_i)\}_{i=1}^{10^7}$, where $\{\hat{\boldsymbol{x}}_i\}_{i=1}^{10^7}$ was a set of i.i.d input realisations. The vertical deflections δ_i were computationally affordable to compute due to the simplicity of the model. The results are summarised in Table 3, together with the CoV of the estimates (see Table 1).

While $\mu(\Delta)$ was virtually identical across the input model, $\sigma(\Delta)$ exhibited non-negligible changes. For instance, it increased by almost 10% from the independence to the vine copula. As a consequence, if $C^{(\mathcal{V})}$ were the true copula among the loads, an MC estimate of $\sigma(\Delta)$

Copula	Family	Parameter values	τ_K	λ_u
C_{13}	Clayton, rotated 180	$\theta = 0.1806$	0.0828	0.0215
C_{12}	Tawn-2	$\theta_1 = 1.439$, $\theta_2 = 0.3406$	0.1522	0.1975
C_{16}	Gumbel	$\theta = 1.103$	0.0934	0.1254
C_{14}	Tawn, rotated 180	$\theta_1 = 7.506$, $\theta_2 = 0.05197$, $\theta_3 = 0.2982$	0.0454	0
C_{15}	Clayton, rotated 180	$\theta = 0.3794$	0.1595	0.1609
$C_{35\|1}$	Tawn	$\theta_1 = 11.05$, $\theta_2 = 0.1338$, $\theta_3 = 0.1178$	0.0658	0.1151

Table 2: **Pair copulas of inferred C-vine for loads on horizontal truss.** Pair copulas of the C-vine model $\hat{C}^{(\mathcal{V})}$ for the loads on the horizontal truss model, obtained from 300 samples of $C^{(\mathcal{V})}$. The pair copulas found to be independence copulas are not shown. The last two columns indicate the Kendall's tau and the upper tail dependence coefficient of each pair copula.

	$C^{(\Pi)}$	$C^{(\mathcal{N})}$	$C^{(\mathcal{V})}$	$\hat{C}^{(\mathcal{V})}$
$\hat{\mu}_{\mathrm{MC}}(\Delta)$ (cm):	$\hat{\mu}_{\mathrm{MC}}^{(\Pi)} = 7.78$	$\hat{\mu}_{\mathrm{MC}}^{(\mathcal{N})} = 7.78$	$\hat{\mu}_{\mathrm{MC}}^{(\mathcal{V})} = \mathbf{7.78}$	$\hat{\mu}_{\mathrm{MC}}^{(\hat{\mathcal{V}})} = 7.78$
$\mathrm{CoV}(\hat{\mu}_{\mathrm{MC}}(\Delta))$ ($\times 10^{-5}$):	2.1	2.3	2.4	2.4
$\hat{\sigma}_{\mathrm{MC}}(\Delta)$ (cm):	$\hat{\sigma}_{\mathrm{MC}}^{(\Pi)} = 0.528$	$\hat{\sigma}_{\mathrm{MC}}^{(\mathcal{N})} = 0.566$	$\hat{\sigma}_{\mathrm{MC}}^{(\mathcal{V})} = \mathbf{0.581}$	$\hat{\sigma}_{\mathrm{MC}}^{(\hat{\mathcal{V}})} = 0.593$
$\mathrm{CoV}(\hat{\sigma}_{\mathrm{MC}}(\Delta))$ ($\times 10^{-4}$):	2.2	2.2	2.2	2.2

Table 3: **Moments of horizontal truss deflection for different load couplings.** MC estimates $\hat{\mu}_{\mathrm{MC}}(\Delta)$ and $\hat{\sigma}_{\mathrm{MC}}(\Delta)$ for different copulas of the loads, based on 10^7 samples, and their CoV. Reference solutions in bold.

based on the independence assumption would be biased. Conversely, fitting a Gaussian copula or a C-vine to data yielded more accurate estimates.

MC estimates converge slowly (see Section 4.1) and therefore need to be computed on large samples. Figure 3 shows by solid lines the errors on MC estimates drawn for the four different copulas on 10, 100, $\ldots, 10^5$ samples. The reference solutions were the MC estimates $\hat{\mu}_{\mathrm{MC}}^{(\mathcal{V})}$, $\hat{\sigma}_{\mathrm{MC}}^{(\mathcal{V})}$ obtained through 10^7 samples under $C^{(\mathcal{V})}$. The errors of estimates $\tilde{\mu}$, $\tilde{\sigma}$ were defined as

$$E_{\mathrm{rel}}^{(\mu)} = \left| \frac{\tilde{\mu}}{\hat{\mu}_{\mathrm{MC}}^{(\mathcal{V})}} - 1 \right|, \quad E_{\mathrm{rel}}^{(\sigma)} = \left| \frac{\tilde{\sigma}}{\hat{\sigma}_{\mathrm{MC}}^{(\mathcal{V})}} - 1 \right|. \tag{32}$$

Note that, due to the CoV of the reference solutions, reported in Table 3, the errors shown in Figure 3 are reliable only down to approximately 10^{-4} for the means and 10^{-3} for the standard deviations.

We further estimated for each copula the error on $\mu(\Delta)$ and $\sigma(\Delta)$ yielded by PCE,

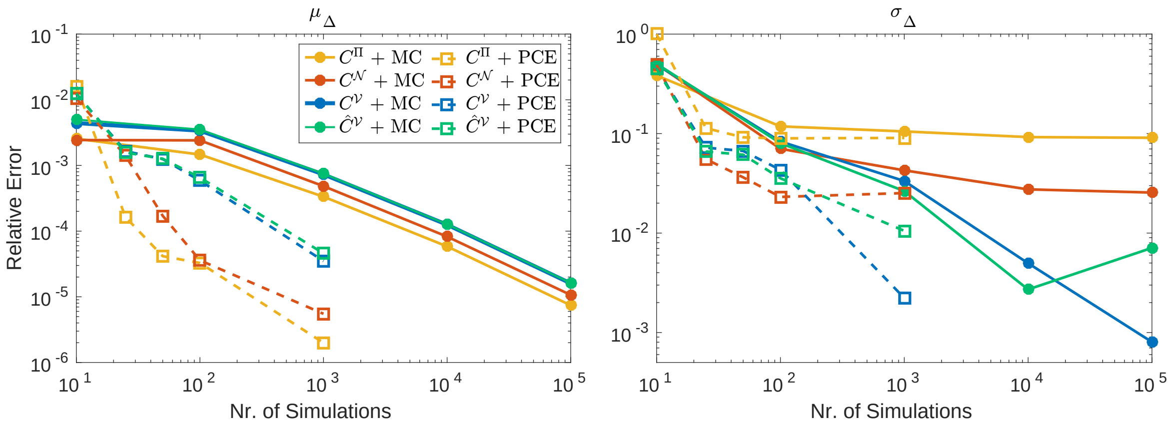

Figure 3: **Estimates of the deflection moments.** Estimates of the errors on $\mu(\Delta)$ (left panel) and on $\sigma(\Delta)$ (right panel) obtained using an increasing number of samples by MCS (solid lines) and by PCE (dashed lines), for loads coupled by $C^{(\Pi)}$ (yellow), $C^{(\mathcal{N})}$ (red), $C^{(\mathcal{V})}$ (blue) and $\hat{C}^{(\mathcal{V})}$ (green). Reference solution: MC estimate obtained on 10^7 samples with copula $C^{(\mathcal{V})}$.

which is known to converge faster than MCS (see Section 4.2). We increased the size of the experimental design from 10 to 1000 sample points. The errors on the PCE estimates are shown in Figure 3 by dashed lines (again, reliable only down to the above mentioned precision). Notably, for the same number n of samples the PCE error is significantly smaller than the MCS error, demonstrating that the vine representation is fully compatible with PCE metamodelling.

5.4 Reliability analysis for different load couplings

The truss model was set to fail if the deflection Δ reached or exceeded the critical threshold $\delta^* = 11\,\text{cm}$. Reliability analysis was performed to evaluate the failure probability $P_f = \mathbb{P}(\Delta \geq \delta^*) = 1 - F_\Delta(\delta^*)$.

For each probabilistic input model (i.e. for each copula $C_{\boldsymbol{X}}$, combined with the marginals in (30)), we first obtained reference solutions by MCS. Using the $n = 10^7$ i.i.d. realisations $\{\hat{\delta}_i = \mathcal{M}(\hat{\boldsymbol{x}}_i)\}_{i=1}^{10^7}$ obtained for the analysis of the moments, we estimated P_f as the fraction of observed deflections $\hat{\delta}_i$ larger than 11 cm. Then, we drew estimates by FORM, applied on the compositional model resulting from decoupling the loads via Rosenblatt transformation (see Sections 3 and 4.3). The results are summarised in Table 4.

The failure probability estimated by MCS with the independence copula $C^{(\Pi)}$ was $\hat{P}_f^{(\Pi)} = (1.5 \pm 0.1) \times 10^{-5}$. The FORM estimate was $\hat{P}_{f;\text{FORM}}^{(\Pi)} = 0.37 \times 10^-5$, obtained by 219 runs of the computational model. Figure 4B shows, in yellow, the empirical survival function of Δ obtained under $C^{(\Pi)}$, for values of δ ranging from 6 cm to 12 cm. The vertical dashed line marks the critical threshold δ^*, whereas the square indicates the FORM estimate of P_f.

The MC estimate of P_f under the Gaussian copula $C^{(\mathcal{N})}$ was $\hat{P}_{f;\text{MC}}^{(\mathcal{N})} = (3.4 \pm 0.2) \times 10^{-5}$

Method:	MCS				FORM			
Copula:	$C^{(\Pi)}$	$C^{(\mathcal{N})}$	$C^{(\mathcal{V})}$	$\hat{C}^{(\mathcal{V})}$	$C^{(\Pi)}$	$C^{(\mathcal{N})}$	$C^{(\mathcal{V})}$	$\hat{C}^{(\mathcal{V})}$
$\hat{P}_f$ ($\times 10^{-4}$):	0.15 ± 0.01	0.34 ± 0.02	$\mathbf{5.04 \pm 0.07}$	3.30 ± 0.06	0.037	0.10	4.88	2.94
CoV($\hat{P}_f$) (%):	8.2	5.4	1.4	2.3	—	—	—	—
# runs:	10^7	10^7	10^7	10^7	219	219	108	128

Table 4: **Estimates of the truss failure probability.** Estimates of P_f obtained with different copulas and methods (for MCS with standard deviation of the estimator; reference solution in bold), CoV of the MC estimate (see Table 1), and number of runs of the computational model needed to obtain the solution.

(Figure 4, orange line: empirical survival function of Δ under $C^{(\mathcal{N})}$). Compared to the independent case, the failure probability increased by a factor of over 2, as a result of the positive correlations among the loads. The FORM estimate was $\hat{P}^{(\mathcal{N})}_{f;\mathrm{FORM}} = 1.0 \times 10^{-5}$, obtained again by 219 runs.

The MC estimate of P_f under the C-vine $C^{(\mathcal{V})}$ was $\hat{P}^{(\mathcal{V})}_{f;\mathrm{MC}} = (5.04 \pm 0.07) \times 10^{-4}$ (Figure 4, blue line: empirical survival function of Δ under $C^{(\mathcal{V})}$). This value was over 33 times larger than the case of independent loads and 14 times larger than the Gaussian case, despite the marginal distributions of the loads being identical across all cases, and the Spearman's correlation coefficients being identical between $C^{(\mathcal{V})}$ and $C^{(\mathcal{N})}$. The FORM estimate using $C^{(\mathcal{V})}$ was $\hat{P}^{(\mathcal{V})}_{f;\mathrm{FORM}} = 4.88 \times 10^{-4}$, obtained by only 108 runs.

Finally, the MC estimate of P_f assuming $\hat{C}^{(\mathcal{V})}$ was $\hat{P}^{(\hat{\mathcal{V}})}_{f;\mathrm{MC}} = 3.30 \times 10^{-4}$, about 35% smaller than the reference solution $\hat{P}^{(\mathcal{V})}_{f;\mathrm{MC}}$ (Figure 4, green line: empirical survival function of Δ under $\hat{C}^{(\mathcal{V})}$). The FORM estimate, obtained with 128 runs of the computational model, was $\hat{P}^{(\hat{\mathcal{V}})}_{f;\mathrm{FORM}} = 2.94 \times 10^{-4}$ (42% smaller than $\hat{P}^{(\mathcal{V})}_{f;\mathrm{MC}}$).

In light of these results, in a scenario where the true dependence among the loads is described by (31), assuming independence or a Gaussian copula would cause a severe underestimation of the failure probability of the system, even when relying on a large MCS strategy. Properly capturing the dependencies (in particular, the tail dependencies) among the inputs is thus more critical towards getting accurate estimates of P_f than using more precise estimation algorithms (FORM combined with $C^{(\mathcal{V})}$ outperforms MCS on 10^7 samples combined with $C^{(\mathcal{N})}$). Furthermore, the error on P_f remains small when the vine is entirely inferred from available input data, because the tail dependencies are properly captured (see Table 2). This demonstrates the viability of the vine copula modelling framework in reliability analysis for purely data driven inference.

The results above show that the failure probability is heavily misestimated when inputs coupled by a C-vine with tail dependencies are modelled by a Gaussian copula. A natural

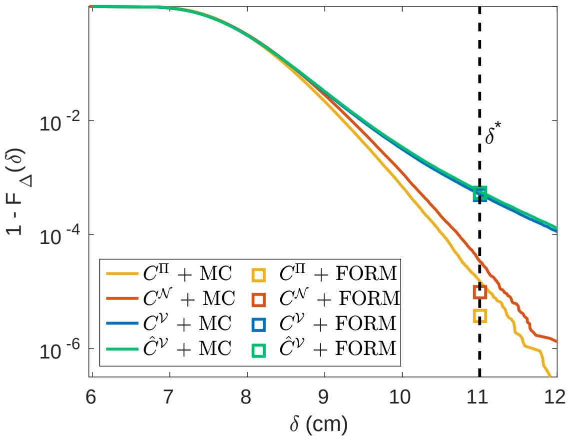

Figure 4: **Reliability analysis of the truss structure.** Solid lines: MC estimate of the survival function $\mathbb{P}(\Delta > \delta)$ for loads coupled by the independence copula $C^{(\Pi)}$ (yellow), the Gaussian copula $C^{(\mathcal{N})}$ (red), the C-vine with known parameters $C^{(\mathcal{V})}$ (blue, mostly overlapping with green), and the C-vine with fitted parameters $\hat{C}^{(\mathcal{V})}$ (green). The vertical dashed line marks the critical threshold δ^*. Squares: estimates of P_f obtained by FORM.

question that arises is the following: is the opposite also true? In other words, how well can the C-vine family capture input dependencies described by a Gaussian copula? To answer this question, we performed additional simulations with $C^{(\mathcal{N})}$ as the true input copula (having associated failure probability $\hat{P}^{(\mathcal{N})}_{f;\mathrm{MC}} = (3.4 \pm 0.2) \times 10^{-5}$). We sampled 300 observations from $C^{(\mathcal{N})}$, and inferred a C-vine from those. The pair copula families were selected as before by AIC among the parametric families listed in Table 11 and their rotated version, and their parameters were fitted by maximum likelihood. The resulting estimate of P_f was $(2.2 \pm 0.2) \times 10^{-5}$, only 35% smaller than the reference value. This demonstrates that the C-vine family is an effective dependence model also in the presence of simple Gaussian dependencies. In conclusion, this class of models covers a larger range of dependence scenarios than Gaussian (or elliptical) copulas, enabling UQ also in cases where the classical use of the Nataf transform would lead to wrong estimates.

6 Results on a dome truss model under asymmetric loads

We further considered a more complex model of a three-dimensional 120-bar dome truss. The model was used to demonstrate the applicability of our novel copula-based UQ framework on a more realistic case study than the one previously analysed. Due to the computational complexity of this model ($\sim$ 15 seconds/run), large MCS was not affordable.

Copula	Family	Parameter values	τ_K	λ_u
C_{12}	Partial Frank	$\theta = 1.363$	0.1189	0
C_{13}	Partial Frank	$\theta = 1.320$	0.1166	0
C_{14}	IterFGM	$\theta_1 = 0.8832$, $\theta_2 = -0.8688$	0.1536	0
C_{15}	AMH	$\theta = 0.4297$	0.1080	0
C_{16}	Gaussian	$\theta = 0.1726$	0.1104	0

Table 5: **Pair copulas of inferred C-vine for horizontal truss, under Gaussian assumption.** Pair copulas of the C-vine model for the loads on the horizontal truss model, obtained from 300 samples of $C^{(\mathcal{N})}$. The pair copulas found to be independence copulas are not shown. The last two columns indicate the Kendall's tau and the upper tail dependence coefficient of each pair copula.

6.1 Computational model

The dome structure, illustrated in Figure 5A from the top and in Figure 5B from the front, consists of 120 bars connected to a total of 49 nodes. Nodes 1 to 37 (grey dots) are unsupported and therefore, when subject to vertical loading, exhibit a displacement from the original position in possibly all directions. The spatial dimensions of the structure are reported in panel B.

The computational model was implemented in the finite element software Abaqus (Smith, 2009). This structure was previously analysed in Kaveh and Talatahari (2009) to obtain optimal sizing variables so as to minimise the total structural weight. The authors distinguished 7 groups of bars, and optimised the cross-sections of each group to minimise the total structural weight of the structure under 4 different types of stress and displacement constraints. We considered in particular their case 2, where stress and displacement constraints (± 5 mm) in the $x-$ and y- directions were enforced. In Kaveh and Talatahari (2009) it was further assumed that each unsupported node is subject to vertical loading, taken as 60 kN at node 1, 30 kN at nodes 2-14, and 10 kN at nodes $15-37$. Under these conditions, the optimal cross sections reported in Table 6 were obtained, yielding a total structural weight of $89,35$ kN.

6.2 Probabilistic input model

Here we were interested in analysing the displacement of the nodes, considered as a risk factor potentially leading to failure.

First, we assigned uncertainty to the bar cross-sections. We modelled the 7 previously identified groups by independent log-normal random variables with mean given by the values in Table 6 and CoV $\sigma/\mu = 0.03$. We assigned the bars in each of the 7 groups identical cross-

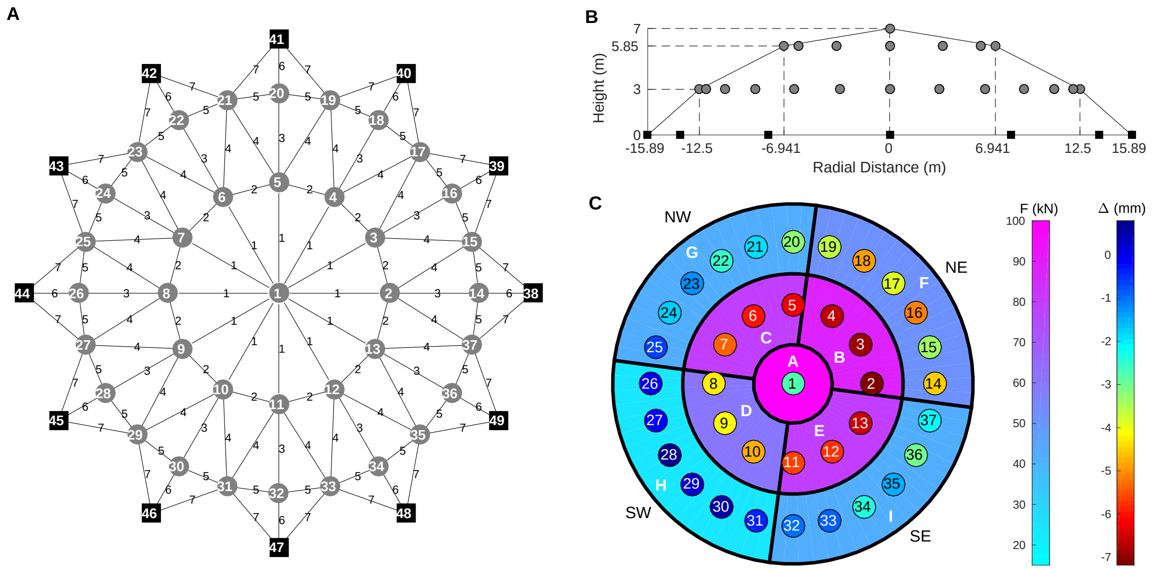

Figure 5: **Dome structure and average response to loads. A)** Truss model, consisting of 37 central nodes (grey dots, 1-37), 12 support nodes (black squares, 37-49) and 120 bars connecting them, divided into 7 groups. **B)** Profile of the dome with spatial dimensions. **C)** Sectors of the dome surface and nodes in each sector. The color of each sector represents the total average load weighing on each node in the sector (left color bar). The color of each node marks the average vertical displacement of that node in response to the loading, calculated over 1000 Monte Carlo simulations.

	Optimal cross-sectional areas (cm^2)							Weight (kN)	Loads (kN)			
Group:	1	2	3	4	5	6	7	89.35	Node:	1	2-13	14-37
Area:	24.38	21.79	26.61	17.64	10.38	22.79	16.38		Load:	60	30	10

Table 6: **Dome's structural parameters and loads**. Cross-section of each group of bars that minimises the total structural weight under stress and x-, y- displacement constraints, resulting structural weight, and additional loads on each node. Values provided by Kaveh and Talatahari (2009).

sections.

We further assigned uncertainty to the loads applied on each node. We modelled a scenario where loads are distributed asymmetrically over the structure. A preliminary analysis showed that asymmetric loads yielded higher maximum vertical displacement compared to symmetric loads. We divided the 37 central nodes into 9 groups, named A to I, corresponding to different sectors of the surface of the dome. Sector A contains solely node 1, sector B contains nodes 2, 3 and 4, and so on, as shown in Figure 5 and in the upper two rows of Table 7. We considered the nodes in each group to be subject to the same load, and modelled the loads on the 9 groups by a 9-dimensional random vector with prescribed marginals and copula. The marginals were taken to be Gumbel distributions (30), whose moments, shown in the bottom two rows of Table 7, were determined as follows. We assigned to each sector a Gumbel-distributed load, having mean $1\,\mathrm{kN/m^2}$ for the top and north-east sectors A,B,F, $0.5\,\mathrm{kN/m^2}$ for the north-west and south-east sectors C, E, G and I, and $0.25\,\mathrm{kN/m^2}$ for the south-west sectors D and H. The different mean values could model, for instance, snow falling on the dome from the north-east direction. The average external weight on each node (third-last row of the table) was obtained by multiplication with the total area of the node's sector (fourth row) and by division with the number of nodes in that sector. The CoV of each distribution was set to 0.2. Finally, deterministic service loads similar to those suggested by Kaveh and Talatahari (2009) were added: $60\,\mathrm{kN}$ on node 1, $30\,\mathrm{kN}$ on nodes 2-13, $10\,\mathrm{kN}$ on nodes $14-37$.

We coupled the 9 loads by three different copulas: the independence copula $C^{(\Pi)}$ (11), the Gaussian copula $C^{(\mathcal{N})}$ (12), and a 9-dimensional C-vine $C^{(\mathcal{V})}$ (31). The C-vine consisted of Gumbel-Hougaard pair-copulas (13), each with parameter $\theta = 5$, between sector A and sectors $B, \ldots, I$ for the first tree, and independence conditional pair-copulas for the other trees. This choice assigns the loads between nodes in sector A and any other loads a Kendall's correlation coefficient $\tau_K = 0.8$ and an upper tail dependence coefficient $\lambda_u = 0.85$. Thus, $C^{\mathcal{V}}$ assigns a strong positive correlation to the loads and a high probability of having joint extremes if one of the loads takes values in its upper tail. The Gaussian copula was taken

Sector		A	B	C	D	E	F	G	H	I
Nodes		1	2–4	5–7	8–10	11–13	14–19	20–25	26–31	32–37
$\mathrm{L_{fix}}$/Node	(kN)	60	30 each				10 each			
Sector's Area	($\mathrm{m^2}$)	37.84	257.00 each				496.38 each			
Area/Node	($\mathrm{m^2}$)	37.84	86.33 each				82.73 each			
$\mathrm{L_{ext}/m^2}$	(kN)	1	1	0.5	0.25	0.5	1	0.5	0.25	0.5
$\mu(\mathrm{L_{ext}}/\mathrm{Node})$	(kN)	37.84	21.58	10.79	5.40	10.79	20.68	10.34	5.17	10.34
$\mu(\mathrm{L_{tot}}/\mathrm{Node})$	(kN)	97.84	51.58	40.79	35.40	40.79	30.68	20.34	15.17	20.34
$\sigma(\mathrm{L_{tot}}/\mathrm{Node})$	(kN)	7.57	4.32	2.16	1.08	2.16	4.14	2.07	1.03	2.07

Table 7: **Load statistics on each dome sector**. For each node sector from A to I: nodes in the sector, structural load $\mathrm{L_{fix}}$ per node , average external load $\mathrm{L_{ext}}$ per node, moments of the total load $\mathrm{L_{tot}}$.

such that its correlation matrix would match the correlation coefficients determined by the C-Vine.

We further inferred a C-vine $\hat{C}^{(\mathcal{V})}$ from 300 samples obtained from $C^{(\mathcal{V})}$. The resulting vine $\hat{C}^{(\mathcal{V})}$, whose comprising pair copulas are listed in Table 8, had the same structure as $C^{(\mathcal{V})}$, Gumbel-Hougaard copulas C_{AB}, C_{AC}, $\ldots$, C_{AH}, and Tawn-2 copula C_{AI}. The Tawn-2 copula is a generalization of the Gumbel copula with right-skewed asymmetry in relation to the main diagonal. It is obtained from the three-parameters Tawn copula (Tawn, 1988) by setting one of its two asymmetry parameters to 1 (see Tables 11-12, row 17). All conditional copulas of $\hat{C}^{(\mathcal{V})}$, finally, were correctly found to be independence copulas.

6.3 Model response to the uncertain input

The output of the model in response to a single instance of the input is a list of displacements in the x-, y- and z- directions, one per node, as well as the tension (or compression) of each bar. We restricted our attention to displacements only, which, if excessive (11 mm in any direction), lead to failure of the structure.

We first performed a preliminary Monte-Carlo analysis based on 1000 simulations of the input (with loads coupled by $C^{\mathcal{V}}$) and corresponding output displacements. Figure 5C shows in two different color codes the average weights on the nodes in each sector (left color bar) and the resulting average vertical displacement of each node (right color bar). Negative displacement indicates that the node moved downwards. Some nodes exhibited positive displacements, *i.e.*, uplifting.

Copula	Family	Parameter values	τ_K	λ_u
C_{AB}	Gumbel	$\theta = 5.093$	0.8037	0.8542
C_{AC}	Gumbel	$\theta = 4.836$	0.7932	0.8459
C_{AD}	Gumbel	$\theta = 5.151$	0.8059	0.8560
C_{AE}	Gumbel	$\theta = 4.775$	0.7906	0.8438
C_{AF}	Gumbel	$\theta = 4.631$	0.7841	0.8385
C_{AG}	Gumbel	$\theta = 5.018$	0.8007	0.8519
C_{AH}	Gumbel	$\theta = 4.712$	0.7878	0.8415
C_{AI}	Tawn-2	$\theta_1 = 5.257$, $\theta_2 = 0.967$	0.7875	0.8445

Table 8: **Pair copulas of inferred C-vine for loads on dome structure.** Pair copulas of the C-vine model $\hat{C}^{(\mathcal{V})}$ for the loads on the dome structure, obtained from 300 samples of $C^{(\mathcal{V})}$. The pair copulas found to be independence copulas are not shown. The last two columns indicate the Kendall's tau and the upper tail dependence coefficient of each pair copula.

For all simulations and all nodes, the vertical displacement always exceeded in absolute value the displacement in the x- and y- directions. This was expected, considering that the average bar's cross-sections were optimised to minimise the latter two. Besides, the absolute vertical displacement was always maximal at node 2, except for 17 out of 1000 simulations where the maximal absolute displacement was observed at node 3, but was never critical (that is, was always $< 11\,\text{mm}$). Thus, we reduced the model's response to the vertical displacement Δ of node 2:

$$\Delta = \mathcal{M}(\boldsymbol{X}), \quad \boldsymbol{X} = (A_1, \ldots, A_7, L_A, \ldots, L_I).$$

6.4 Analysis of the moments

For each copula mentioned above, we evaluated the mean $\mu(\Delta)$ and the standard deviation $\sigma(\Delta)$ of the deflection Δ at node 2 both by MCS and by PCE. The estimates were based on samples of size n increasing from 10 to 1000. Due to the generally faster convergence of PCE with respect to MCS for small sample sizes, the PCE estimates built on 1000 samples were taken as reference values for each of the four copula models (see Table 9). The values obtained indicate that the independence, Gaussian and vine copulas yielded for Δ similar means but different standard deviations.

Taken in particular $C^{(\mathcal{V})}$ to be the true copula among the loads, and the corresponding PCE estimates $\hat{\mu}^{(\mathcal{V})}_{\text{PCE}}$ and $\hat{\sigma}^{(\mathcal{V})}_{\text{PCE}}$ based on 1000 points to be the reference solutions, we computed the relative error of all other estimates. The errors, defined analogously to (32), are shown in Figure 6. From these results, three main conclusions can be drawn. First, if $C^{(\mathcal{V})}$

	$C^{(\Pi)}$	$C^{(\mathcal{N})}$	$C^{(\mathcal{V})}$	$\hat{C}^{(\mathcal{V})}$
$\hat{\mu}_{\text{PCE}}(\Delta)$ (mm):	$\hat{\mu}^{(\Pi)}_{\text{PCE}} = -7.193$	$\hat{\mu}^{(\mathcal{N})}_{\text{PCE}} = -7.183$	$\hat{\mu}^{(\mathcal{V})}_{\text{PCE}} = \mathbf{-7.182}$	$\hat{\mu}^{(\hat{\mathcal{V}})}_{\text{PCE}} = -7.182$
$\hat{\sigma}_{\text{PCE}}(\Delta)$ (mm):	$\hat{\sigma}^{(\Pi)}_{\text{PCE}} = 1.164$	$\hat{\sigma}^{(\mathcal{N})}_{\text{PCE}} = 0.588$	$\hat{\sigma}^{(\mathcal{V})}_{\text{PCE}} = \mathbf{0.552}$	$\hat{\sigma}^{(\hat{\mathcal{V}})}_{\text{PCE}} = 0.560$

Table 9: **Moments of dome's deflection for different load couplings.** PCE estimates of $\mu(\Delta)$ and $\sigma(\Delta)$ for different copulas among the loads, based on 1000 observations. Reference solutions in bold.

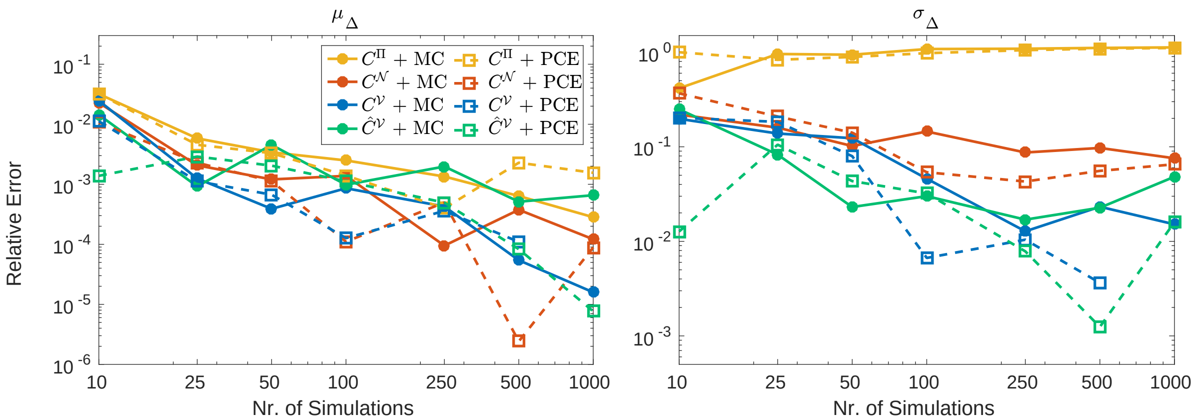

Figure 6: **Errors on moments of dome's deflection Δ.** Estimates of the errors on $\mu(\Delta)$ (left panel) and on $\sigma(\Delta)$ (right panel) obtained using an increasing number of samples by MCS (solid lines) and by PCE (dashed lines), for loads coupled by $C^{(\Pi)}$ (yellow), $C^{(\mathcal{N})}$ (red), $C^{(\mathcal{V})}$ (blue) and $\hat{C}^{(\mathcal{V})}$ (green). Reference solutions: PCE estimates $\tilde{\mu}^{(\mathcal{V})}$, $\tilde{\sigma}^{(\mathcal{V})}$ obtained on 1000 samples with copula $C^{(\mathcal{V})}$.

was the true copula among the loads, neither $C^{(\Pi)}$ nor $C^{(\mathcal{N})}$ would offer adequate alternative representations, since the associated standard deviations differ significantly (by 111% and 6.5%, respectively) from their reference value. Second, by employing the inferred vine $\hat{C}^{(\mathcal{V})}$ in combination with MCS (green solid line) it is possible to approximate the moments with higher precision than by the Gaussian (red) or independence (yellow) copulas. Finally, PCE combines well with the vine representation (dashed lines), yielding the smallest errors. It is worth noting, however, that using a proper copula model (a vine instead of a Gaussian or independence copula) is more important to obtain accurate estimates (particularly for $\sigma(\Delta)$) than using a more advanced UQ method (PCE instead of MC).

6.5 Reliability analysis

The dome structure was further set to fail if the displacement Δ was equal to or lower than the critical threshold $\delta^* = -11$ mm. We performed reliability analysis to estimate the failure

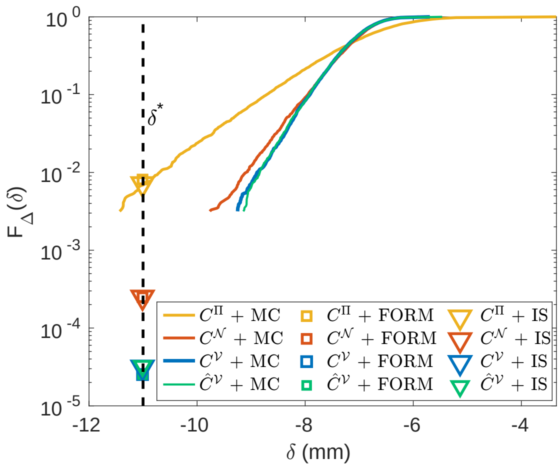

Figure 7: **Dome failure probability.** The solid lines show the MC estimates of the response CDF $F_\Delta(\delta) = \mathbb{P}(\Delta \leq \delta)$ (for values down to 10^{-3}) under loads coupled by $C^{(\Pi)}$ (yellow), $C^{(\mathcal{N})}$ (red), $C^{(\mathcal{V})}$ (blue) and $\hat{C}^{(\mathcal{V})}$ (green). The vertical dashed line represents the critical threshold δ^*. The squares and triangles on it indicate the FORM and IS estimates of the failure probability, respectively, obtained for each copula model.

probability $P_f = \mathbb{P}(\Delta \leq \delta^*) = F_\Delta(\delta^*)$ of excessive downward vertical displacement.

We performed 5000 simulations by MCS for each copula of the input model, keeping the marginals identical across the models. Figure 7 shows the CDFs resulting from copulas $C^{(\mathcal{V})}$ (blue), $\hat{C}^{(\mathcal{V})}$ (green), $C^{(\mathcal{N})}$ (red), and $C^{(\Pi)}$ (yellow), evaluated for probabilities down to 10^{-3}.

The MC estimate of P_f under $C^{(\Pi)}$ was $\hat{P}^{(\Pi)}_{f;\mathrm{MC}} = (6.8{\pm}1.2){\times}10^{-3}$. For copulas $C^{(\mathcal{N})}$, $C^{(\mathcal{V})}$ and $\hat{C}^{(\mathcal{V})}$ no simulations led to values of Δ below δ^*. We then resorted to FORM to evaluate P_f for each copula, obtaining the estimates $\hat{P}^{(\Pi)}_{f;\mathrm{FORM}} = 8.0 \times 10^{-3}$, $\hat{P}^{(\mathcal{N})}_{f;\mathrm{FORM}} = 2.50 \times 10^{-4}$, $\hat{P}^{(\mathcal{V})}_{f;\mathrm{FORM}} = 2.53 \times 10^{-5}$, and $\hat{P}^{(\hat{\mathcal{V}})}_{f;\mathrm{FORM}} = 2.54 \times 10^{-5}$. Since MC estimates were not available or not reliable here in light of the small sample being available, we further performed IS to improve the FORM estimates and to get confidence intervals (see Section 4.4). We increased the IS sample size in steps of 100, until the CoV of the estimate was lower than 10%. We obtained the estimates $\hat{P}^{(\Pi)}_{f;\mathrm{IS}} = (7.13 \pm 0.70) \times 10^{-3}$, $\hat{P}^{(\mathcal{N})}_{f;\mathrm{IS}} = (2.47 \pm 0.24) \times 10^{-4}$, $\hat{P}^{(\mathcal{V})}_{f;\mathrm{IS}} = (3.13 \pm 0.29) \times 10^{-5}$, and $\hat{P}^{(\hat{\mathcal{V}})}_{f;\mathrm{IS}} = (3.27 \pm 0.30) \times 10^{-5}$.

The results, summarized in Table 10, show that the failure probability of the structure decreases by an order of magnitude from $C^{(\Pi)}$ to $C^{(\mathcal{N})}$, and by another order of magnitude from $C^{(\mathcal{N})}$ to $C^{(\mathcal{V})}$ and $\hat{C}^{(\mathcal{V})}$. Highly asymmetric loads (as due to $C^{(\Pi)}$ and, to a minor extent, to $C^{(\mathcal{N})}$) may create a deformation mechanism in the structure that favours large displacements of the most heavily loaded nodes (here, node 2). In contrast, the more symmetric loading determined by the C-vine results in a more evenly distributed load path that ultimately leads to a safer structure. For loads actually coupled by $C^{(\mathcal{V})}$, assuming

Method:	MCS				FORM				IS			
Copula:	$C^{(\Pi)}$	$C^{(\mathcal{N})}$	$C^{(\mathcal{V})}$	$\hat{C}^{(\mathcal{V})}$	$C^{(\Pi)}$	$C^{(\mathcal{N})}$	$C^{(\mathcal{V})}$	$\hat{C}^{(\mathcal{V})}$	$C^{(\Pi)}$	$C^{(\mathcal{N})}$	$C^{(\mathcal{V})}$	$\hat{C}^{(\mathcal{V})}$
$\hat{P}_f$ ($\times 10^{-5}$):	680 ± 120	–	–	–	800	25	2.53	2.56	713 ± 70	24.7 ± 2.4	$\mathbf{3.13 \pm 0.29}$	3.27 ± 0.30
CoV($\hat{P}_f$) (%):	17.6	–	–	–	–	–	–	–	9.9	9.9	9.3	9.3
tot. # runs:	5000	5000	5000	5000	182	349	276	181	482	749	876	781

Table 10: **Estimates of the dome failure probability.** Estimates of P_f obtained with different copulas and methods (reference solution in bold), CoV of the MC and IS estimates, and number of runs needed for the estimation.

the independence or Gaussian copulas thus leads to highly overestimating P_f. Conversely, building the vine by purely data-driven inference recovers the reference solution $\hat{P}_{f;\mathrm{IS}}^{(\mathcal{V})}$ with high precision. Again, the input model used for the analysis is more important to get an accurate estimate than the particular UQ method (FORM or IS) employed.

7 Discussion

We proposed a general framework that enables uncertainty quantification (UQ) for problems where the input parameters of the system exhibit complex, non-Gaussian, non-elliptical dependencies (copulas). The joint CDF of input parameters is expressed in terms of marginals and a copula, which are modelled separately. The copula is further modelled as a vine copula, *i.e.*, a product of simpler 2-copulas. This specification eases its construction, especially in high dimension, and offers a simple interpretation of the dependence model. A wide range of different dependence structures can be modelled using this approach.

Our framework focuses in particular on regular (R-) vines, for which algorithms exist to compute the likelihood on available data, thus enabling parameter fitting and data driven inference. In addition, R-vines offer algorithms to compute the associated Rosenblatt transform and its inverse on data, used to map the original input random vector into a vector with independent components and back. Thus, UQ techniques that benefit from input independence can be applied to any inputs coupled by R-vines. In this work we restricted our attention to inputs with continuous marginals, which cover a large class of engineering problems. Extensions of R-vines to discrete (*e.g.*, categorical and count) data have been recently proposed (Panagiotelis et al., 2012, 2017).

The methodology was first demonstrated on a simple horizontal truss model, for which Monte Carlo solutions were computationally affordable, and then replicated on a more complex truss model of a dome. Both structures deflected in response to loads on different nodes. Changing the copula among the loads from the independence to a Gaussian to a tail-dependent C-vine copula changed the statistics of the deflection, in particular its variance and upper quantiles. Taken the vine copula as the true dependence structure among the loads, the independence and Gaussian assumptions thus led to biased estimates of these statistics. The failure probability of the two systems, in particular, was mis-estimated by one to two orders of magnitude. This was true regardless of the particular UQ method used for the estimation. Using instead a vine copula model of the input dependencies and fitting the model to relatively few input observations yielded far better estimates.

These results demonstrate that using a proper dependence model for the inputs can be more critical to get high-accuracy estimates of the output statistics than employing a superior UQ algorithm. Our framework encompasses both aspects, allowing highly flexible probabilistic models of the input to be combined with virtually any UQ technique designed to solve problems characterized by (finitely many) coupled inputs. Also, we demonstrated that a suitable vine representation can be properly inferred on data also in the presence of simple Gaussian dependencies. Thus, this class of dependence models effectively covers a broader range of problems than the Nataf transform (also in its generalized form by Lebrun and Dutfoy (2009b)) does.

Selecting a vine that properly represents the dependencies of multivariate inputs may be challenging. We discussed and employed existing methods to perform fully automated inference on available data. When the dimension of the input is large or the parametric families of pair copulas considered for the vine construction are many, this approach may become computationally prohibitive. A-priori information on the input statistics may be used to ease the selection, for instance by Bayesian methods (Gruber and Czado, 2015). The problem of selecting suitable vines, however, remains open in very high dimension (say, > 50) or on very large samples. Also, computing the Rosenblatt and inverse Rosenblatt transforms in these cases may be computationally demanding or lead to numerical instability. Separating the inputs into mutually independent subgroups, by expert knowledge or by statistical testing, and inferring a (vine) copula for each separately, may reduce this problem significantly. Additionally, vine inference on samples of large size can become computationally demanding. Estimation techniques based on parallel computing have been recently proposed to solve this issue (Wei et al., 2016). Additional work is foreseen to address these challenges.

Acknowledgements

The work has been funded by the ETH Foundation through the ETH Risk Center Seed Project SP-RC 07-15 "Copulas for Big Data Analysis in Engineering Sciences", and by the RiskLab from the Department of Mathematics of the ETH Zurich. The authors thank Moustapha Maliki for providing access to the Abaqus implementation of the dome model and to the UQLink module of UQLab needed to run it, and for helpful discussions on the interpretation of the dome response.

References

Aas, K. (2016). Pair-copula constructions for financial applications: A review. *Econometrics 4*(4), 43.

Aas, K., C. Czado, A. Frigessi, and H. Bakkend (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics 44*(2), 182–198.

Applegate, D. L., R. E. Bixby, V. Chvátal, and W. J. Cook (2006). *The Traveling Salesman Problem: A Computational Study*. NewJersey: Princeton University Press.

Au, S. and J. Beck (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Prob. Eng. Mech. 16*(4), 263–277.

Bedford, T. and R. M. Cooke (2002). Vines – a new graphical model for dependent random variables. *The Annals of Statistics 30*(4), 1031–1068.

Berveiller, M., B. Sudret, and M. Lemaire (2006). Stochastic finite elements: a non intrusive approach by regression. *Eur. J. Comput. Mech. 15*(1-3), 81–92.

Blatman, G. and B. Sudret (2011). Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *J. Comput. Phys 230*, 2345–2367.

Czado, C. (2010). *Pair-Copula Constructions of Multivariate Copulas*, pp. 93–109. Berlin, Heidelberg: Springer Berlin Heidelberg.

Der Kiureghian, A. and P. Liu (1986). Structural reliability under incomplete probability information. *J. Eng. Mech. 112*(1), 85–104.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis 59*, 52–69.

Ditlevsen, O. and H. Madsen (1996). *Structural reliability methods*. J. Wiley and Sons, Chichester.

Doostan, A. and H. Owhadi (2011). A non-adapted sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys. 230*(8), 3015–3034.

Embrechts, P., A. McNeil, and D. Straumann (1999). Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and beyond*, pp. 176–223. Cambridge University Press.

Ernst, O. G., A. Mugler, H.-J. Starkloff, and E. Ullmann (2012). On the convergence of generalized polynomial chaos expansions. *ESAIM: M2AN 46*(2), 317–339.

Fermanian, J.-D. (2012). An overview of the goodness-of-fit test problem for copulas. In P. Jaworski, F. Durante, and W. K. Härdle (Eds.), *Copulae in Mathematical and Quantitative Finance*, Volume 213, Chapter 4, pp. 61–89. Springer.

Fiessler, B., H.-J. Neumann, and R. Rackwitz (1979). Quadratic limit states in structural reliability. *Journal of Engineering Mechanics 105*(4), 661–676.

Ghanem, R. and P. Spanos (2003). *Stochastic Finite Elements : A Spectral Approach*. Courier Dover Publications.

Goda, K. (2010). Statistical modeling of joint probability distribution using copula: Application to peak and permanent displacement seismic demands. *Structural Safety 32*, 112–123.

Goda, K. and S. Tesfamariam (2015). Multi-variate seismic demand modelling using copulas: Application to non-ductile reinforced concrete frame in Victoria, Canada. *Structural Safety 56*, 39–51.

Grønneberg, S. and N. Hjort (2014). The copula information criteria. *Scand. J. Stat. 41*, 436–459.

Gruber, L. and C. Czado (2015). Sequential Bayesian model selection of regular vine copulas. *Bayesian Analysis 10*(4), 937–963.

Haff, I. H., K. Aas, and A. Frigessi (2010). On the simplified pair-copula construction – simply useful or too simplistic? *Journal of Multivariate Analysis 101*, 1296–1310.

Hasofer, A.-M. and N.-C. Lind (1974). Exact and invariant second moment code format. *J. Eng. Mech. 100*(1), 111–121.

Hohenbichler, M. and R. Rackwitz (1983). First-order concepts in system reliability. *Structural Safety 1*(3), 177–188.

Jakeman, J., M. Eldred, and K. Sargsyan (2015). Enhancing ℓ_1-minimization estimates of polynomial chaos expansions using basis selection. *J. Comput. Phys. 289*, 18–34.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In L. Rüschendorf, B. Schweizer, and M. D. Taylor (Eds.), *Distributions with fixed marginals and related topics*, Volume 28 of *Lecture Notes–Monograph Series*, pp. 120–141. Institute of Mathematical Statistics.

Joe, H. (Ed.) (2015). *Dependence modeling with copulas*. CRC Press.

Kaveh, A. and S. Talatahari (2009). Particle swarm optimizer, and colony strategy and harmony search scheme hybridized for optimization of truss structures. *Computers and Structures 87*, 267–283.

Keese, A. and H. Matthies (2003). Numerical methods and smolyak quadrature for nonlinear stochastic partial differential equations. Technical report, Technical Report, Technische Universität Braunschweig.

Kirk, J. (2014). Traveling salesman problem – genetic algorithm. `https://ch.mathworks.com/matlabcentral/fileexchange/13680-traveling-salesman-problem-genetic-algorithm`.

Klüpperberg, C. and C. Czado (no date). Vine copula models. `https://www.statistics.ma.tum.de/forschung/vine-copula-models/`. [Online; accessed 21-September-2017].

Kurowicka, D. and R. M. Cooke (2005). Distribution-free continuous Bayesian belief nets. In A. Wilson, N. Limnios, S. Keller-McNulty, and Y. Armijo (Eds.), *Modern Statistical and Mathematical Methods in Reliability*, Chapter 10, pp. 309–322. World Scientific Publishing.

Kurz, M. (2015). Vine copulas with matlab. `https://ch.mathworks.com/matlabcentral/fileexchange/46412-vine-copulas-with-matlab`.

Le Maître, O., O. Knio, H. Najm, and R. Ghanem (2001). A stochastic projection method for fluid flow – I. Basic formulation. *J. Comput. Phys. 173*, 481–511.

Lebrun, R. and A. Dutfoy (2009a). Do Rosenblatt and Nataf isoprobabilistic transformations really differ? *Prob. Eng. Mech. 24*, 577–584.

Lebrun, R. and A. Dutfoy (2009b). A generalization of the Nataf transformation to distributions with elliptical copula. *Prob. Eng. Mech. 24*(2), 172–178.

Lemaire, M. (2009). *Structural reliability.* Wiley.

Li, R. and R. Ghanem (1998). Adaptive polynomial chaos expansions applied to statistics of extremes in nonlinear random vibration. *Prob. Eng. Mech. 13*(2), 125–136.

Manner, H. (2007). Estimation and model selection of copulas with an application to exchange rates. Research Memorandum 056, Maastricht University, Maastricht Research School of Economics of Technology and Organization (METEOR).

Marelli, S. and B. Sudret (2014). UQLab: A framework for uncertainty quantification in Matlab. In *Vulnerability, Uncertainty, and Risk (Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom)*, pp. 2554–2563.

Marelli, S. and B. Sudret (2017). UQLab user manual – Polynomial chaos expansions. Technical report, Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich. Report # UQLab-V1.0-104.

Masina, M., A. Lamberti, and R. Archetti (2015). A copula based approach for estimating the joint probability of water levels and waves. *Coastal Engineering 97*, 37–52.

Matheron, G. (1967). Kriging or polynomial interpolation procedures. *Canadian Inst. Mining Bull. 60*, 1041.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 2*, 239–245.

Melchers, R. (1999). *Structural reliability analysis and prediction.* John Wiley & Sons.

Michele, C., G. Salvadori, G. Passoni, and R. Vezzoli (2007). A multivariate model of sea storms using copulas. *Coastal Engineering 54*, 734–751.

Montes-Iturrizaga, R. and E. Heredia-Zavoni (2016). Reliability analysis of mooring lines using copulas to model statistical dependence of environmental variables. *Applied Ocean Research 59*, 564–576.

Morales-Nápoles, O. (2011). Counting vines. In D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*, Chapter 9, pp. 189–218. World Scientific Publisher Co.

Nataf, A. (1962). Détermination des distributions dont les marges sont données. *C. R. Acad. Sci. Paris 225*, 42–43.

Nelsen, R. (2006). *An introduction to copulas* (Second ed.). Springer Series in Statistics. Springer-Verlag New York.

Panagiotelis, A., C. Czado, and H. Joe (2012). Pair copula constructions for multivariate discrete data. *J. Amer. Statist. Assoc. 107*, 1063–1072.

Panagiotelis, A., C. Czado, H. Joe, and J. Stöber (2017). Model selection for discrete regular vine copulas. *Computational Statistics & Data Analysis 106*, 138–152.

Papaioannou, I., W. Betz, K. Zwirglmaier, and D. Straub (2015). MCMC algorithms for subset simulation. *Probabilistic Engineering Mechanics 41*, 89–103.

Romanovsky, V. (1925). On the moments of the standard deviation and of the correlation coefficient in samples from normal. *Metron 5*(4), 3–46.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*, 470–472.

Scarsini, M. (1984). On measures of concordance. *Stochastica 8*(3), 201–218.

Schepsmeier, U. (2010). Maximum likelihood estimation of c-vine pair copula constructions based on bivariate copulas from different families. Master's thesis, Technische Univärsität München.

Schepsmeier, U. (2015). Efficient information based goodness-of-fit tests for vine copula models with fixed margins. *Journal of Multivariate Analysis 138*(C), 34–52.

Schepsmeier, U. and J. Stöber (2014). Derivatives and Fisher information of bivariate copulas. *Statistical Papers 55*(2), 525–542.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris 8*, 229–231.

Smith, M. (2009). *ABAQUS/Standard User's Manual, Version 6.9.* Simulia.

Spanhel, F. and M. S. Kurz (2016). The partial copula: Properties and associated dependence measures. *Statistics and Probability Letters 119*, 76–83.

Stöber, J., H. Joe, and C. Czado (2013). Simplified pair copula constructions — limitations and extensions. *Journal of Multivariate Analysis 119*, 101–118.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika 75*(3), 397–415.

Taylor, M. D. (2007). Multivariate measures of concordance. *Annals of the Institute of Statistical Mathematics 59*(4), 789–806.

The Mathworks Inc. (2016). *Matlab version 9.0.0.341360 (R2016a).* Natick, Massachusetts: The Mathworks Inc.

Wang, F. and H. Li (2017a). Stochastic response surface method for reliability problems involving correlated multivariates with non-gaussian dependence structure: Analysis under incomplete probability information. *Computers and Geotechnics 89*, 22–32.

Wang, F. and H. Li (2017b). System reliability under prescribed marginals and correlations: Are we correct about the effect of correlations? *Reliability Engineering & System Safety*, –. In press.

Wang, F. and H. Li (2017c). Towards reliability evaluation involving correlated multivariates under incomplete probability information: A reconstructed joint probability distribution for isoprobabilistic transformation. *Structural Safety 69*, 1–10.

Wei, Z., D. Kim, and E. M. Conlon (2016). Parallel computing for copula parameter estimation with big data: A simulation study. *arXiv 1609.05530.*

Xiu, D. (2010). *Numerical methods for stochastic computations – A spectral method approach.* Princeton University press.

Xiu, D. and G. Karniadakis (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput. 24*(2), 619–644.

Zentner, I. (2017). A general framework for the estimation of analytical fragility functions based on multivariate probability distributions. *Structural Safety 64*, 54–61.

A Some families of pair copulas and their properties

Table 11 lists the 19 parametric families of pair copulas implemented in the VineCopulaMatlab toolbox (Kurz, 2015) used here for vine inference. Each pair copula in the inferred vines was chosen among these families and their rotated versions defined by (33), by selecting the family yielding the lowest AIC. The rotations of a pair-copula distribution C are defined, here and in most references, by

$$\begin{aligned} C^{(90)}(u,v) &= v - C(1-u,v), \\ C^{(180)}(u,v) &= u + v - 1 + C(1-u,1-v), \\ C^{(270)}(u,v) &= u - C(u,1-v). \end{aligned} \tag{33}$$

(Note that $C^{(90)}$ and $C^{(270)}$ are obtained by flipping the copula density c around the horizontal and vertical axis, respectively; some references provide the formulas for actual rotations: $C^{(90)}(u,v) = v - C(v,1-u)$, $C^{(270)}(u,v) = u - C(1-v,u)$). Including the rotated copulas, 62 families were considered in total for inference in our study.

The analytical expressions for the Kendall's tau and for the coefficients λ_l, λ_u of lower and upper tail dependence of the non-rotated families, when available, are reported in Table 12. We derived ourselves a few of these expressions, as indicated in the table, since we could not find them in the existing literature (see notes (a) and (c) in the table's caption). Note also that λ_l and λ_u switch when a copula density is rotated by 180° and becomes its survival version. This allows copulas with lower tail dependence to be used to model upper tail dependence, and vice versa, by 180° rotation. Copulas rotated by 90° and 270° model negative dependence.

ID	Name	CDF	Parameter range
1	AMH	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\theta \in [-1, 1]$
2	AsymFGM	$uv\left(1+\theta(1-u)^2 v(1-v)\right)$	$\theta \in [0, 1]$
3	BB1	$\left(1+\left((u^{-\theta_2}-1)^{\theta_1}+(v^{-\theta_2}-1)^{\theta_1}\right)^{1/\theta_1}\right)^{-1/\theta_2}$	$\theta_1 \geq 1$, $\theta_2 > 0$
4	BB6	$1-\left(1-\exp\left\{-\left[(-\log(1-(1-u)^{\theta_2}))^{\theta_1}+(-\log(1-(1-v)^{\theta_2}))^{\theta_1}\right]^{1/\theta_1}\right\}\right)^{1/\theta_2}$	$\theta_1 \geq 1$, $\theta_2 \geq 1$
5	BB7	$\varphi(\varphi^{-1}(u)+\varphi^{-1}(v))$, where $\varphi(w)=\varphi(w;\theta_1,\theta_2)=1-\left(1-(1+w)^{-1/\theta_1}\right)^{1/\theta_2}$	$\theta_1 \geq 1$, $\theta_2 > 0$
6	BB8	$\frac{1}{\theta_1}\left(1-\left(1-\frac{(1-(1-\theta_1 u)^{\theta_2})(1-(1-\theta_1 v)^{\theta_2})}{1-(1-\theta_1)^{\theta_2}}\right)^{1/\theta_2}\right)$	$\theta_1 \geq 1$, $\theta_2 \in (0, 1]$
7	Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$\theta > 0$
8	FGM	$uv(1+\theta(1-u)(1-v))$	$\theta \in (-1, 1)$
9	Frank	$-\frac{1}{\theta}\log\left(\frac{1-e^{-\theta}-(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$	$\theta \in \mathbb{R}\backslash\{0\}$
10	Gaussian	$\Phi_{2;\theta}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)$ $^{(a)}$(see (12), with $d=2$)	$\theta \in (-1, 1)$
11	Gumbel	$\exp\left(-((-\log u)^{\theta}+(-\log v)^{\theta})^{1/\theta}\right)$	$\theta \in [1, +\inf)$
12	Iterated FGM	$uv(1+\theta_1(1-u)(1-v)+\theta_2 uv(1-u)(1-v))$	$\theta_1, \theta_2 \in (-1, 1)$
13	Joe/B5	$1-\left((1-u)^{\theta}+(1-v)^{\theta}+(1-u)^{\theta}(1-v)^{\theta}\right)^{1/\theta}$	$\theta \geq 1$
14	Partial Frank	$\frac{uv}{\theta(u+v-uv)}(\log(1+(e^{-\theta}-1)(1+uv-u-v))+\theta)$	$\theta > 0$
15	Plackett	$\frac{1+(\theta-1)(u+v)-\sqrt{(1+(\theta-1)(u+v))^2-4\theta(\theta-1)uv}}{2(\theta-1)}$	$\theta \geq 0$
16	Tawn-1	$(uv)^{A\left(\frac{\log v}{\log(uv)};\theta_1,\theta_3\right)}$, where $A(w;\theta_1,\theta_3)=(1-\theta_3)w+\left[w^{\theta_1}+(\theta_3(1-w))^{\theta_1}\right]^{1/\theta_1}$	$\theta_1 \geq 1$, $\theta_3 \in [0, 1]$
17	Tawn-2	$(uv)^{A\left(\frac{\log v}{\log(uv)};\theta_1,\theta_2\right)}$, where $A(w;\theta_1,\theta_2)=(1-\theta_2)(1-w)+\left[(\theta_2 w)^{\theta_1}+((1-w))^{\theta_1}\right]^{1/\theta_1}$	$\theta_1 \geq 1$, $\theta_2 \in [0, 1]$
18	Tawn	$(uv)^{A(w;\theta_1,\theta_2,\theta_3)}$, where $w=\frac{\log v}{\log(uv)}$ and	
		$A(w;\theta_1,\theta_2,\theta_3)=(1-\theta_2)(1-w)+(1-\theta_3)w+\left[(\theta_2 w)^{\theta_1}+(\theta_3(1-w))^{\theta_1}\right]^{1/\theta_1}$	$\theta_1 \geq 1$, $\theta_2, \theta_3 \in [0, 1]$
19	t-	$t_{2;\nu,\theta}\left(t_{\nu}^{-1}(u),t_{\nu}^{-1}(v)\right)$ $^{(b)}$	$\nu > 1$, $\theta \in (-1, 1)$

Table 11: **Distributions of bivariate copula families used for inference of vine copulas.** The copula IDs are reported as assigned in the VineCopulaMatlab toolbox used here (Kurz, 2015). (a) Φ is the univariate standard normal distribution, and $\Phi_{2;\theta}$ is the bivariate normal distribution with zero means, unit variance and correlation parameter θ. (b) t_{ν} is the univariate t distribution with ν degrees of freedom, and $t_{\nu,\theta}$ is the bivariate t distribution with ν degrees of freedom and correlation parameter θ.

ID	Name	τ_K	λ_l	λ_u	Special cases
1	AMH	$1 - \frac{2\theta + 2(1-\theta)^2 \ln(1-\theta)}{3\theta^2}$	$0.5 \cdot \mathbf{1}_{\{\theta=1\}}$	0	—
2	AsymFGM	$\frac{\theta}{18}$ $^{(a)}$	0	0	—
3	BB1	$1 - \frac{2}{\theta_1(\theta_2+2)}$	$2^{-1/(\theta_1\theta_2)}$	$2 - 2^{1/\theta_1}$	Clayton ($\theta_1 = 1$), Gumbel ($\theta_2 \downarrow 0^+$)
4	BB6	numerical	0	$2 - 2^{1/(\theta_1\theta_2)}$	Joe ($\theta_1 = 1$), Gumbel ($\theta_2 = 1$)
5	BB7	see (Schepsmeier, 2010)	$2^{-1/\theta_1}$	$2^{-1/\theta_2}$	Joe ($\theta_1 \downarrow 0^+$), Clayton ($\theta_2 = 1$)
6	BB8	numerical	0	0 for $\theta_1 \neq 1$	Joe ($\theta_1 \downarrow 0^+$), Frank ($\theta_2 = 1$)
7	Clayton	$\frac{\theta}{\theta+2}$	$2^{-1/\theta}$	0	—
8	FGM	$\frac{2\theta}{9}$	0	0	—
9	Frank	$1 + \frac{4}{\theta}(\frac{1}{\theta}\int_0^\theta t(e^t-1)^{-1}dt - 1)$	0	0	—
10	Gaussian	$\frac{2}{\pi}\arcsin(\theta)$	0	0	—
11	Gumbel	$\frac{\theta-1}{\theta}$	0	$2 - 2^{1/\theta}$	—
12	Iterated FGM	$\frac{2\theta_1}{9} + \frac{(25+\theta_1)\theta_2}{450}$ $^{(a)}$	0	0	FGM ($\theta_2 = 0$)
13	Joe/B5	$1 + \frac{2}{2-\theta}(\digamma(2) - \digamma(\frac{2}{\theta}+1))$ $^{(b)}$	0	$2 - 2^{1/\theta}$	—
14	Partial Frank (Spanhel and Kurz, 2016)	numerical	0	0	—
15	Plackett	numerical	0	0	—
16	Tawn-1	numerical	0 $^{(c)}$	$1 + \theta_3 - \left(1 + \theta_3^{\theta_1}\right)^{1/\theta_1}$ $^{(c)}$	Gumbel ($\theta_3 = 1$)
17	Tawn-2	numerical	0 $^{(c)}$	$1 + \theta_2 - \left(1 + \theta_2^{\theta_1}\right)^{1/\theta_1}$ $^{(c)}$	Gumbel ($\theta_2 = 1$)
18	Tawn	numerical	0 $^{(c)}$	$\theta_2 + \theta_3 - \left(\theta_2^{\theta_1} + \theta_3^{\theta_1}\right)^{1/\theta_1}$ $^{(c)}$	Tawn-1 ($\theta_2 = 1$), Tawn-2 ($\theta_3 = 1$), Gumbel ($\theta_2 = \theta_3 = 1$)
19	t-	$\frac{2}{\pi}\arcsin(\theta)$	$\lambda_l = \lambda_u =$ $^{(d)}$ $= 2t_{\nu+1}\left(-\sqrt{(\nu+1)(1-\theta)/(1+\theta)}\right)$		—

Table 12: **Some properties of the considered pair copulas.** Kendall's tau, tail dependence coefficients, subfamilies of pair copulas that obtain for specific parameter values. (a) We derived the analytical expression of τ_K for the asymmetric and iterated FGM copulas using the RHS of (7). (b) $\digamma$ is the digamma function. (c) We derived the analytical expression of the tail dependence coefficients by using (10), by noting that $A(w) = 1 + \frac{1}{2}\left((\theta_2^{\theta_1} + \theta_3^{\theta_1})^{1/\theta_1} - (\theta_2 + \theta_3)\right)$ when $u = v$ and, for λ_u, by calculating the limit through first order Taylor expansion. (d) t_ν is the univariate t distribution with ν degrees of freedom.