

Article

# Copula-Based Uncertainty Quantification (Copula-UQ) for Multi-Sensor Data in Structural Health Monitoring

He-Qing Mu <sup>1,2,3,\*</sup> , Han-Teng Liu <sup>1</sup> and Ji-Hui Shen <sup>1</sup>

<sup>1</sup> School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China; 201820106919@mail.scut.edu.cn (H.-T.L.); 201920106856@mail.scut.edu.cn (J.-H.S)

<sup>2</sup> State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou 510640, China

<sup>3</sup> Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150080, China

\* Correspondence: cthqmu@scut.edu.cn

Received: 29 August 2020; Accepted: 24 September 2020; Published: 6 October 2020



**Abstract:** The problem of uncertainty quantification (UQ) for multi-sensor data is one of the main concerns in structural health monitoring (SHM). One important task is multivariate joint probability density function (PDF) modelling. Copula-based statistical inference has attracted significant attention due to the fact that it decouples inferences on the univariate marginal PDF of each random variable and the statistical dependence structure (called copula) among the random variables. This paper proposes the Copula-UQ, composing multivariate joint PDF modelling, inference on model class selection and parameter identification, and probabilistic prediction using incomplete information, for multi-sensor data measured from a SHM system. Multivariate joint PDF is modeled based on the univariate marginal PDFs and the copula. Inference is made by combining the idea of the inference functions for margins and the maximum likelihood estimate. Prediction on the PDF of the target variable, using the complete (from normal sensors) or incomplete information (due to missing data caused by sensor fault issue) of the predictor variable, are made based on the multivariate joint PDF. One example using simulated data and one example using temperature data of a multi-sensor of a monitored bridge are presented to illustrate the capability of the Copula-UQ in joint PDF modelling and target variable prediction.

**Keywords:** copula; missing data; multivariate random variable; probability density distribution; structural health monitoring

## 1. Introduction

The problem of uncertainty quantification (UQ) for multi-sensor data has been one of the main concerns in nondestructive testing and structural health monitoring (SHM) over the years [1–9]. One important task is multivariate joint probability density function (PDF) modelling. Due to irregularities of multi-sensor data, the joint PDF can be too complicated to be modelled by traditional approaches. For example, traditional multivariate PDFs (such as a multivariate normal distribution) cannot model the PDF with multiple peaks. The multivariate mixture PDFs (such as multivariate normal mixture model), utilized in SHM and damage detection [10–13], rely on the proper choice of the number and the type of the mixture distributions and an initial value of parameter vector in optimization [14]. The Nataf distribution, utilized in SHM and structural reliability [15,16], relies on

the assumption that the transformed random variables, obtained from the marginal transformations of the original random variables, are multivariate normal distribution [17].

In recent years, copula-based statistical inference has attracted significant attention due to the fact that it decouples the inference on the univariate marginal PDF of each random variable and the statistics dependence structure (called copula) among the random variables. In the areas of the SHM and structural assessment, Zhang and Kim [18] investigated a way of detecting bridge damage for the long-term health monitoring by using the copula theory. Fan and Liu [19] predicted the dynamic reliability of a bridge system based on SHM data. Pan et al. [20] developed a copula-based approach to model the structural health of an operational metro tunnel in a dependent system. Liu et al. [21] considered the correlation between the fatigue equivalent stress and the stress cycle using the copula function in the fatigue reliability assessment. Srinivas et al. [22] proposed the multivariate simulation of dependent axle weights of different vehicle classes. Zhang et al. [23] investigated the specification of long-term design loads for offshore structures considering multiple environmental factors.

Although the copula-based statistical inference has been widely applied, there are two limitations in previous works related to the SHM and structural assessment. The first limitation is insufficient types of probabilistic model candidates for univariate marginal PDF modelling. From the parametrization point of view, there are parametric models and nonparametric models for PDF modelling. The former type, assuming that sample data come from a distribution that has a fixed set of parameters, is suitable for data with regular statistical pattern; the latter type, being not specified a priori but being instead adaptively determined from data, is suitable for data with an irregular statistical pattern. In the SHM, it is well known that the statistical regularities of data from multiple sensors can be significantly different from each other. Thus, due to the complexity of real SHM data, only considering one type of probabilistic model in PDF modelling bounds the solution space for UQ, leading to incapability of capturing a statistical pattern of data. However, this important issue was not realized in previous works, so parametric models and nonparametric models were not considered simultaneously. For example, References [18–21,23] solely adopted parametric models, while Reference [22] solely adopted nonparametric models for univariate marginal PDF modelling. Thus, this paper attempts to break through this limitation by including sufficient types of probabilistic models as candidates. The second limitation is negligence of probabilistic prediction using available information, especially in the case of using the incomplete information of the predictor variable due to missing data caused by a sensor fault issue. For the works of the research area of SHM using the copula [18–23], it had not been realized that the joint PDF can be utilized for probabilistic prediction on the target variable using the available information of the predictor variable. Even for the very recent work of another research area using the copula [24], probabilistic prediction on the target variable is limited to the case using the complete information of the predictor variable only. However, the case of incomplete information of the predictor variable, due to missing data caused by a sensor fault issue, is critical and common in the SHM. Thus, this paper attempts to break through this limitation by conducting computation of marginalization and conditioning based on the copula-based joint PDF, for prediction on the PDF of the target variable using the complete (from normal sensors) or incomplete information (due to missing data caused by sensor fault issue) of the predictor variable.

This paper proposes the copula-based UQ (Copula-UQ), composing multivariate joint PDF modelling, inference on model class selection and parameter identification, and probabilistic prediction using incomplete information, for multi-sensor data measured from a SHM system. The proposed Copula-UQ contains two stages. The first stage is the copula-based multivariate joint PDF modelling. It is based on the univariate marginal PDFs and the copula. The second stage is copula-based inference and prediction. Inference, including determination of optimal parameters and selection of optimal model classes, is made by combining the idea of the inference functions for margins (IFM) and the maximum likelihood estimate (MLE). Prediction on the PDF of the target variable, using the complete or incomplete information of the predictor variable, are made based on the copula-based multivariate joint PDF.

The structure of this paper is outlined as follows. Section 2 presents copula-based multivariate joint PDF modelling, including model class candidates for univariate marginal PDFs and copula. Section 3 presents copula-based inference and prediction, including inference on univariate marginal PDFs and copula, and prediction on the target variable. Section 4 presents illustrative examples. One example using simulated data and one example using temperature data of multi-sensor of a monitored bridge are presented to illustrate the capability of the proposed Copula-UQ in joint PDF modelling and target variable prediction.

## 2. Copula-Based Multivariate Joint PDF Modelling

Let  $p(x_1, x_2, \dots, x_D)$  denote the joint PDF of  $D$  random variables  $(X_1, X_2, \dots, X_D)$ , and  $\mathbf{x} \in \mathbb{R}^{D \times N}$  denote the measured data matrix with its component  $\mathbf{x}_{d,i}$  being the  $d$ -th dimension of the  $i$ -th data point, with  $d = 1, \dots, D$  and  $i = 1, \dots, N$ . The copula-based multivariate joint PDF is to model  $p(x_1, x_2, \dots, x_D)$  based on the univariate marginal PDFs  $p(x_d), d = 1, 2, \dots, D$  of each random variable and the statistics dependence structure (called copula) among the random variables, given the measured data matrix  $\mathbf{x}$ .

### 2.1. Univariate Marginal PDFs

For the  $d$ -th univariate random variable  $X_d$ , consider a set of  $N_M$  model class candidates, namely,  $\mathcal{M}_d^{m_d}, m_d = 1, 2, \dots, N_M$ . Each model class candidate represents a marginal PDF for  $X_d$ . It is obvious that a larger solution space for uncertainty quantification can be achieved with larger  $N_M$ . Let  $p(x_d | \theta_d^{m_d}, \mathcal{M}_d^{m_d})$  be the PDF of  $x_d$  given the parameter vector  $\theta_d^{m_d} \in \mathbb{R}^{\mathcal{N}(\theta_d^{m_d})}$  of model class candidate  $\mathcal{M}_d^{m_d}$ , where  $\mathcal{N}(\theta_d^{m_d})$  is the number of parameters depending on  $\mathcal{M}_d^{m_d}$ . As copula-based multivariate joint PDF modelling is flexible on the models for univariate marginal PDFs of different dimensions, a set of parametric and nonparametric models are introduced as model class candidates for selection of marginal PDFs. It is worth noting that the introduction of both parametric and nonparametric models simultaneously is necessary for breaking through the first limitation (i.e., insufficient types of probabilistic model candidates) described in Section 1 because it provides a large solution space for uncertainty quantification. This will be validated in Section 4—illustrative examples.

Parametric modeling methods assume that sample data come from a distribution that has a fixed set of parameters. In this paper, the parametric model class candidates include: Normal distribution, Lognormal distribution, Weibull distribution, Gamma distribution, Gumbel distribution and Uniform distribution.

Nonparametric models differ from parametric models in that the model structure is not specified a priori but is instead determined from data. The term nonparametric is not meant to imply that such models completely lack parameters but that the number and nature of the parameters are flexible and not fixed in advance. In this paper, the kernel density estimation method is used. Based on the  $d$ -th dimensional measured data  $\mathbf{x}_d = [\mathbf{x}_{d,1}, \mathbf{x}_{d,2}, \dots, \mathbf{x}_{d,N}]^T$ , the nonparametric kernel density function is adopted [25]:

$$p(x_d | \theta_d^{m_d}, \mathcal{M}_d^{m_d}) = \frac{1}{N\theta_d^{m_d}} \sum_{j=1}^N K_d^{m_d} \left( \frac{x_d - \mathbf{x}_{d,j}}{\theta_d^{m_d}} \right) \quad (1)$$

where  $\theta_d^{m_d}$  is the bandwidth ( $\theta_d^{m_d} > 0$ ) controlling the smoothness of the resulting probability density curve, and  $K_d^{m_d}(\cdot)$  is the kernel smoothing function depending on  $\mathcal{M}_d^{m_d}$  and  $\mathbf{x}_d$ . Let  $\varrho_{d,j}^{m_d} = (x_d - \mathbf{x}_{d,j})/\theta_d^{m_d}$ . Four kernel functions, including Normal kernel, Box kernel, Triangle kernel and Epanechnikov kernel, are introduced as follows:

$$\text{Normal kernel : } K_d^{m_d} \left[ \varrho_{d,j}^{m_d} \right] = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[ \varrho_{d,j}^{m_d} \right]^2 \right\} \quad (2)$$

$$\text{Box kernel : } K_d^{m_d} \left[ \varrho_{d,j}^{m_d} \right] = \frac{1}{2} \mathbf{1}_{|\varrho_{d,j}^{m_d}| \leq 1} \left[ \varrho_{d,j}^{m_d} \right] \quad (3)$$

$$\text{Triangle kernel : } K_d^{m_d} \left[ \varrho_{d,j}^{m_d} \right] = \left[ 1 - \varrho_{d,j}^{m_d} \right] \mathbf{1}_{|\varrho| \leq 1} \left[ \varrho_{d,j}^{m_d} \right] \quad (4)$$

$$\text{Epanechnikov kernel : } K_d^{m_d} \left[ \varrho_{d,j}^{m_d} \right] = \frac{3}{4} \left\{ 1 - \left[ \varrho_{d,j}^{m_d} \right]^2 \right\} \mathbf{1}_{|\varrho_{d,j}^{m_d}| \leq 1} \left[ \varrho_{d,j}^{m_d} \right] \quad (5)$$

where  $\mathbf{1}_{|\varrho_{d,j}^{m_d}| \leq 1} \left[ \varrho_{d,j}^{m_d} \right]$  is the indicator function.

## 2.2. Copula

Let  $P(x_d | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d})$  denote the marginal cumulative distribution function (CDF) of  $X_d$  based on  $\boldsymbol{\theta}_d^{m_d}$  and  $\mathcal{M}_d^{m_d}$ . The probability  $u_d^{m_d}$  of  $X_d \leq x_d$  can be obtained as:

$$u_d^{m_d} = P(x_d | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d}) \quad (6)$$

Applying the probability integral transform to  $X_d$ , the univariate random variable  $U_d^{m_d}$  can be obtained:

$$U_d^{m_d} = P(X_d | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d}) \quad (7)$$

where  $U_d^{m_d}$  follows standard uniform distribution on the interval [0,1].

The copula of  $(X_1, X_2, \dots, X_D)$  is defined as a joint CDF of  $(U_1^{m_1}, U_2^{m_2}, \dots, U_D^{m_D})$  [26]:

$$C(u_1^{m_1}, u_2^{m_2}, \dots, u_D^{m_D}) = P(U_1^{m_1} \leq u_1^{m_1}, U_2^{m_2} \leq u_2^{m_2}, \dots, U_D^{m_D} \leq u_D^{m_D}) \quad (8)$$

That is  $C : [0, 1]^D \rightarrow [0, 1]$  is a  $D$ -dimensional copula if  $C$  is a joint CDF of a  $D$ -dimensional random vector on the unit hypercube  $[0, 1]^D$ , with the marginal PDF of each component of the random vector following the standard uniform PDF on the interval [0,1]. The copula  $C$  describes exactly the dependence structure among the random variables.

Consider the joint CDF of  $(X_1, X_2, \dots, X_D)$ :

$$P(x_1, \dots, x_D) = P(X_1 \leq x_1, \dots, X_D \leq x_D) \quad (9)$$

Sklar's theorem states that there exists a  $D$ -dimensional copula, such that [27]:

$$P(x_1, \dots, x_D) = C(u_1^{m_1}, \dots, u_D^{m_D} | \boldsymbol{\psi}) = C(P(x_1 | \boldsymbol{\theta}_1^{m_1}, \mathcal{M}_1^{m_1}), \dots, P(x_D | \boldsymbol{\theta}_D^{m_D}, \mathcal{M}_D^{m_D}) | \boldsymbol{\psi}) \quad (10)$$

where  $\boldsymbol{\psi}$  is the parameter vector of the copula. If  $P(x_d | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d})$  are continuous, the copula is unique; otherwise, it is uniquely determined on the Cartesian product of the ranges of the marginal CDFs. Sklar's theorem clearly indicates that the joint CDF of random variables can be characterized by a copula in terms of the marginal CDFs.

Thus, the joint PDF,  $p(x_1, x_2, \dots, x_D)$ , can be derived from its joint CDF,  $P(x_1, x_2, \dots, x_D)$ , of Equation (10):

$$p(x_1, x_2, \dots, x_D) = c(u_1^{m_1}, u_2^{m_2}, \dots, u_D^{m_D} | \boldsymbol{\psi}) \cdot \prod_{d=1}^D p(x_d | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d}) \quad (11)$$

where  $c(u_1^{m_1}, u_2^{m_2}, \dots, u_D^{m_D} | \boldsymbol{\psi})$  is the copula density function:

$$c(u_1^{m_1}, u_2^{m_2}, \dots, u_D^{m_D} | \boldsymbol{\psi}) = \frac{\partial^D C(u_1^{m_1}, u_2^{m_2}, \dots, u_D^{m_D} | \boldsymbol{\psi})}{\partial u_1^{m_1} \partial u_2^{m_2} \dots \partial u_D^{m_D}} \quad (12)$$

and  $p(x_d | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d})$  is the marginal PDF of  $X_d$ .

In this paper, the multivariate Gaussian copula and the associated copula density function are introduced as follows [28]:

$$C(u_1^{m_1}, u_2^{m_2}, \dots, u_D^{m_D} | \boldsymbol{\psi}) = \Phi_{\boldsymbol{\rho}(\boldsymbol{\zeta})}(\boldsymbol{\zeta}) \quad (13)$$

$$c(u_1^{m_1}, u_2^{m_2}, \dots, u_D^{m_D} | \boldsymbol{\psi}) = |\boldsymbol{\rho}(\boldsymbol{\zeta})|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \boldsymbol{\zeta}^T [\boldsymbol{\rho}(\boldsymbol{\zeta})^{-1} - \mathbf{I}] \boldsymbol{\zeta}\right\} \quad (14)$$

$$\boldsymbol{\zeta} = [\Phi^{-1}(u_1^{m_1}), \Phi^{-1}(u_2^{m_2}), \dots, \Phi^{-1}(u_D^{m_D})]^T \quad (15)$$

where  $\Phi^{-1}(\cdot)$  is the inverse CDF of the univariate standard normal distribution function,  $\Phi_{\boldsymbol{\rho}(\boldsymbol{\zeta})}(\cdot)$  is the joint CDF of a  $D$ -dimensional normal distribution with mean vector zero and covariance matrix equal to the correlation coefficient matrix  $\boldsymbol{\rho}(\boldsymbol{\zeta})$  of  $\boldsymbol{\zeta}$  defined in Equation (15), the parameter vector  $\boldsymbol{\psi}$  is the collection of the off-diagonal elements of the upper triangular part of  $\boldsymbol{\rho}(\boldsymbol{\zeta})$  and  $\mathbf{I}$  is a  $D$ -dimensional identity matrix.

### 3. Copula-Based Inference and Prediction

#### 3.1. Inference on Univariate Marginal PDFs and Copula

This stage is to make inference on  $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1^{m_1}, \dots, \boldsymbol{\theta}_D^{m_D}\}$  (parameters of the marginal PDFs),  $\mathbf{M} = \{\mathcal{M}_1^{m_1}, \dots, \mathcal{M}_D^{m_D}\}$  (model class candidates of the marginal PDFs) and  $\boldsymbol{\psi}$  (parameters of the multivariate Gaussian-copula), based on the measured data matrix  $\mathbf{x} \in \mathbb{R}^{D \times N}$  and the probability matrix  $\mathfrak{U} \in \mathbb{R}^{D \times N}$ , with its component  $\mathfrak{U}_{d,i} = P(\mathbf{x}_{d,i} | \boldsymbol{\Theta}, \mathbf{M})$ . Under the idea of the inference functions for margins (IFM) [29],  $\boldsymbol{\Theta}$  (along with  $\mathbf{M}$ ) and  $\boldsymbol{\psi}$  can be determined separately.

For the univariate marginals, the optimal parameter values can be obtained by the MLE:

$$\hat{\boldsymbol{\theta}}_d^{m_d} = \operatorname{argmax}_{\boldsymbol{\theta}_d^{m_d}} \log\{Lp(\mathbf{x}_d | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d})\}, \quad d = 1, \dots, D \quad (16)$$

$$Lp(\mathbf{x}_d | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d}) = \prod_{i=1}^N [p(\mathbf{x}_{d,i} | \boldsymbol{\theta}_d^{m_d}, \mathcal{M}_d^{m_d})] \quad (17)$$

where  $\log\{\cdot\}$  is the logarithmic function. For most of the parametric models, analytical forms for the optimal parameters can be derived (for example, see Reference [30]). For nonparametric models, the optimal value,  $\hat{\boldsymbol{\theta}}_d^{m_d}$  (bandwidth), can be obtained by considering the asymptotic mean integrated squared error solution [31]:

$$\hat{\boldsymbol{\theta}}_d^{m_d} = \left(\frac{4\hat{\sigma}_d^5}{3N}\right)^{-\frac{1}{5}} \quad (18)$$

$$\hat{\sigma}_d = \frac{Q_d(0.75) - Q_d(0.25)}{1.349} \quad (19)$$

where  $Q_d(0.75)$  and  $Q_d(0.25)$  are the 75% and 25% quantiles of  $\mathbf{x}_d$ .

The optimal marginal PDFs ( $\hat{\mathbf{M}} = \{\hat{\mathcal{M}}_1, \dots, \hat{\mathcal{M}}_D\}$ ) are selected by comparing the optimal likelihood values of different  $m_d$ :

$$\hat{\mathcal{M}}_d = \operatorname{argmax}_{\mathcal{M}_d^{m_d}} Lp(\mathbf{x}_d | \hat{\boldsymbol{\theta}}_d^{m_d}, \mathcal{M}_d^{m_d}), \quad m_d = 1, \dots, N_M, d = 1, \dots, D \quad (20)$$

After selecting  $\hat{\mathbf{M}}$ , the optimal parameters associated with  $\hat{\mathbf{M}}$  are denoted as  $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_D\}$ . Based on  $\hat{\mathbf{M}}$  and  $\hat{\boldsymbol{\Theta}}$ , the component of the optimal probability matrix  $\hat{\mathfrak{U}}_{d,i} = P(\mathbf{x}_{d,i} | \hat{\boldsymbol{\Theta}}, \hat{\mathbf{M}})$  can be obtained.

The optimal values of the parameters of the multivariate Gaussian-copula  $\hat{\psi}$  can be determined by considering the optimization on  $Lc(\hat{\mathbf{u}}|\psi)$ :

$$\hat{\psi} = \underset{\psi}{\operatorname{argmax}} \log\{Lc(\hat{\mathbf{u}}|\psi)\} \quad (21)$$

$$Lc(\hat{\mathbf{u}}|\psi) = \prod_{i=1}^N [c(\hat{\mathbf{u}}_{1,i}, \hat{\mathbf{u}}_{2,i}, \dots, \hat{\mathbf{u}}_{D,i}|\psi)] \quad (22)$$

where  $c(\hat{\mathbf{u}}_{1,i}, \hat{\mathbf{u}}_{2,i}, \dots, \hat{\mathbf{u}}_{D,i}|\psi)$  is obtained by substituting  $\hat{\mathbf{u}}_{1,i}, \hat{\mathbf{u}}_{2,i}, \dots, \hat{\mathbf{u}}_{D,i}$  into Equation (14). For the multivariate Gaussian copula, the optimal parameter  $\hat{\psi}$  is the collection of the off-diagonal elements of the upper triangular part of  $\rho(\zeta)$ , with each component being Pearson's correlation coefficient.

### 3.2. Prediction on the Target Variable Given Complete or Incomplete Information (Due to Missing Data Caused by a Sensor Fault Issue)

Let  $p(x_1, x_2, \dots, x_D | \hat{\psi}, \hat{\Theta}, \hat{\mathbf{M}})$  denote the multivariate joint PDF obtained by substituting  $\hat{\psi}, \hat{\Theta}, \hat{\mathbf{M}}$  into Equation (11). Let the target variable be the set containing the selected components of  $X_1, X_2, \dots, X_D$  for prediction, and the predictor variable be the complement of the target variable. As the joint PDF contains all the statistical information about the random variables  $(X_1, X_2, \dots, X_D)$ , prediction on the PDF of the target variable can be obtained using the complete or incomplete information of the predictor variable. Let  $x_{ta}, x_o$  and  $x_{uo}$  denote the target variable, observed predictor variable and unobserved predictor variable, respectively. It is worth noting that the existence of unobserved predictor variable  $x_{uo}$  is very common in the SHM as it represents missing data of the corresponding channels of fault sensors. However, the very recent work of copula-based prediction [24] was still incapable of tackling the existence of  $x_{uo}$  in its prediction phase. Here, by conducting computation of marginalization and conditioning on the copula-based multivariate joint PDF, the prediction on the PDF of  $x_{ta}$  based on the observation  $x_o = \tilde{x}_o$  only (that is, available information only) can be obtained by:

$$p(x_{ta}|x_o = \tilde{x}_o) = \frac{p(x_{ta}, x_o = \tilde{x}_o)}{p(x_o = \tilde{x}_o)} = \frac{\int p(x_1, x_2, \dots, x_D | \hat{\psi}, \hat{\Theta}, \hat{\mathbf{M}}) \Big|_{x_o = \tilde{x}_o} dx_{uo}}{\iint p(x_1, x_2, \dots, x_D | \hat{\psi}, \hat{\Theta}, \hat{\mathbf{M}}) \Big|_{x_o = \tilde{x}_o} dx_{uo} dx_{ta}} \quad (23)$$

where  $p(x_1, x_2, \dots, x_D | \hat{\psi}, \hat{\Theta}, \hat{\mathbf{M}}) \Big|_{x_o = \tilde{x}_o}$  is the copula-based multivariate joint PDF with substituting  $x_o = \tilde{x}_o$ . The predicted PDF  $p(x_{ta}|x_o = \tilde{x}_o)$  contains all the statistical information about  $x_{ta}$  based on the information  $x_o = \tilde{x}_o$ . Accordingly, the predicted value (mean) and the associated uncertainty (standard deviation) of  $x_{ta}$  can be obtained.

## 4. Illustrative Examples

One example of simulation data and one example of real SHM data are demonstrated. For the simulation data example, the design of it is to validate the following three critical issues: (1) the necessary introduction of both parametric and nonparametric models for breaking through the first limitation (i.e., insufficient types of probabilistic model candidates), (2) the capability of the multivariate joint PDF modelling of the proposed Copula-UQ (as the true joint PDF is known) and (3) the performance of the proposed Equation (23) for prediction on the target variable given complete or incomplete information (due to missing data caused by a sensor fault issue). For the real SHM data example, the performance of the proposed Copula-UQ for prediction on the target variable under complete (normal sensors) or incomplete (fault sensors) information is further validated by considering the following two cases:

(1) the test dataset is identical to the training dataset, and (2) the test dataset is different from the training dataset.

#### 4.1. Simulation Data

This example applies the proposed Copula-UQ for multivariate joint PDF modelling and prediction of five-dimensional random variables,  $\mathbf{X} = [X_1, \dots, X_5]^T$ . First, five uncorrelated random variables,  $\mathbf{Z} = [Z_1, \dots, Z_5]^T$ , with different marginal PDFs are constructed (see Table 1). Then, the random variables  $\mathbf{X} = [X_1, \dots, X_5]^T$  are obtained by applying an affine transformation  $\mathbf{X} = \mathbf{AZ}$  with  $\mathbf{A}$  given as:

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.2 & 0 & 0.2 & 0.1 \\ 0 & 0.6 & 0 & 0.2 & 0.2 \\ 0 & 0 & 1 & 0 & 0 \\ 0.3 & 0 & 0 & 0.7 & 0 \\ 0 & 0.3 & 0 & 0 & 0.7 \end{bmatrix} \quad (24)$$

Thus, the analytical form of the joint PDF of  $\mathbf{X}$  is:

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{p_{\mathbf{Z}}(\mathbf{z} = \mathbf{A}^{-1}\mathbf{x})}{\det|\mathbf{A}|} \quad (25)$$

where  $p_{\mathbf{Z}}(\mathbf{z} = \mathbf{A}^{-1}\mathbf{x})$  is the joint PDF of  $\mathbf{Z}$  with  $\mathbf{z} = \mathbf{A}^{-1}\mathbf{x}$ . Figure 1 shows the scatter plot of the simulated data for  $X_1$  to  $X_5$  ( $N = 500$ ). The correlation coefficient matrix for  $\mathbf{X}$  is:

$$\text{corr}(\mathbf{X}, \mathbf{X}) = \begin{bmatrix} 1 & 0.6935 & 0.0322 & 0.8028 & 0.1451 \\ 0.6935 & 1 & -0.0185 & 0.7096 & 0.4532 \\ 0.0322 & -0.0185 & 1 & 0.0118 & -0.0184 \\ 0.8028 & 0.7096 & 0.0118 & 1 & -0.0155 \\ 0.1451 & 0.4532 & -0.0184 & -0.0155 & 1 \end{bmatrix} \quad (26)$$

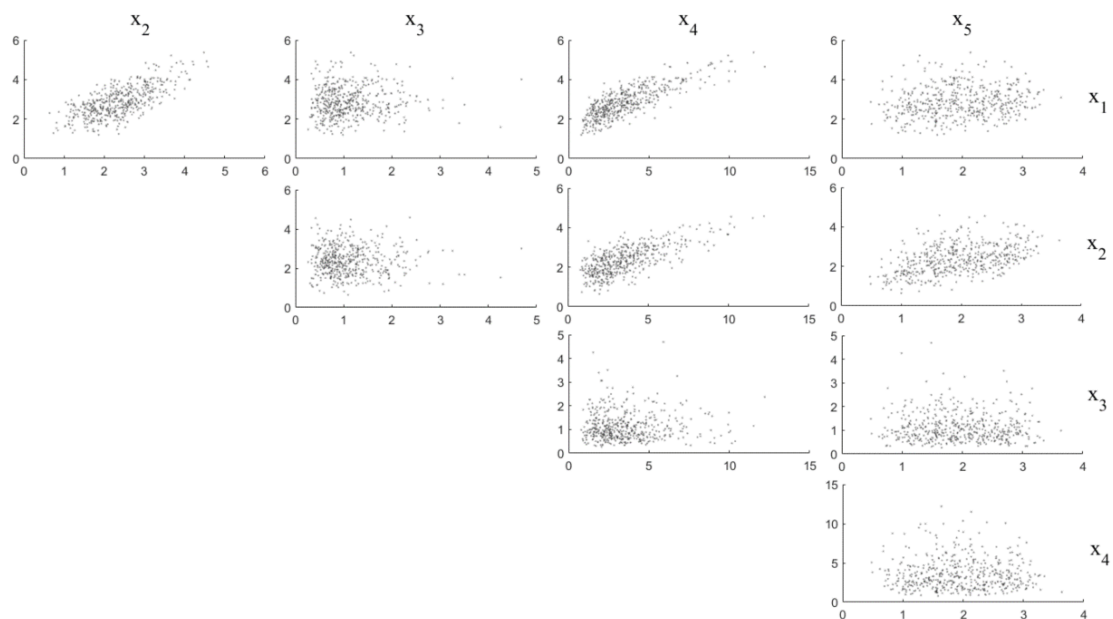


Figure 1. Scatter plot of the simulated data of  $X_1$  to  $X_5$  ( $N = 500$ ).

High correlation (for example, between  $x_1$  and  $x_4$ ), medium correlation (for example, between  $x_2$  and  $x_5$ ) and low correlation (for example, between  $x_1$  and  $x_3$ ) can be found in this case.

Table 2 shows the maximum log-likelihood value of different univariate marginal PDFs of  $X_1$  to  $X_5$ . Using Equation (20), the optimal univariate marginal PDF of each dimension can be determined, and they are indicated by “\_” (underline) in Table 2. The optimal PDFs of  $X_1$  to  $X_5$  are Normal kernel, Triangle kernel, Lognormal distribution, Lognormal distribution and Triangle kernel, respectively. In order to compare the fitting capacities of different PDFs shown in Table 2, Figure 2 shows the univariate marginal PDFs of  $X_1$  to  $X_5$ . Each subplot shows the data histogram, the top ranking PDF (that is, the optimal marginal PDF in Table 2; line style as “dash-dot line”), an intermediate ranking PDF (that is, an intermediate ranking PDF in Table 2; line style as “dashed line”) and a low ranking PDF (that is, a low ranking PDF in Table 2; line style as “dotted line”). From each subplot, it can be reconfirmed that the optimal marginal PDF of each dimension in Table 2 is the best model for uncertainty quantification of the corresponding component of  $X$ . It is worth noting that, from Table 2, even though  $X$  is a linear mapping of  $Z$  only composing very regular types of distributions described in Table 1, the optimal univariate marginal PDFs of  $X$  are not only from parametric but also from nonparametric models. This result shows that the introduction of both parametric and nonparametric models is necessary for breaking through the first limitation (i.e., insufficient types of probabilistic model candidates) described in Section 1 because it provides a large solution space for uncertainty quantification.

**Table 1.** Probability density functions (PDFs) of  $Z_1$  to  $Z_5$  (Simulation data).

Random Variable	Distribution Type	PDF
$Z_1$	Normal	$f(z_1) = \frac{1}{\sqrt{2\pi\sigma_1}} e^{-\frac{(z_1-\mu_1)^2}{2\sigma_1^2}}, \mu_1 = 3, \sigma_1 = 1$
$Z_2$	Normal	$f(z_2) = \frac{1}{\sqrt{2\pi\sigma_2}} e^{-\frac{(z_2-\mu_2)^2}{2\sigma_2^2}}, \mu_2 = 2, \sigma_2 = 0.8$
$Z_3$	Lognormal	$f(z_3) = \begin{cases} 0 & z_3 < 0 \\ \frac{1}{\sqrt{2\pi\sigma_3 z_3}} e^{-\frac{(\ln z_3 - \mu_3)^2}{2\sigma_3^2}} & z_3 \geq 0 \end{cases}, \mu_3 = 0, \sigma_3 = 0.5$
$Z_4$	Gamma	$f(z_4) = \begin{cases} 0 & z_4 \leq 0 \\ \frac{\beta_4^{\alpha_4} x^{\alpha_4-1} e^{-\beta_4 z_4}}{\Gamma(\alpha_4)} & z_4 > 0 \end{cases}, \alpha_4 = 2, \beta_4 = 2$
$Z_5$	Uniform	$f(z_5) = \begin{cases} 0 & z_5 \leq a_5, z_5 \geq b_5 \\ \frac{1}{b_5-a_5} & a_5 < z_5 < b_5 \end{cases}, a_5 = 0.5, b_5 = 3.5$

**Table 2.** The maximum log-likelihood value of different univariate marginal PDFs of  $X_1$  to  $X_5$ .

PDF	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
Normal kernel	-553.3514	-532.5992	-349.9818	-965.5745	-480.0222
Uniform kernel	-556.9401	-535.5691	-354.8022	-969.9313	-481.3604
Triangle kernel	-553.4045	<u>-532.4293</u>	-349.6769	-965.5385	<u>-479.7110</u>
Epanechnikov kernel	-555.2073	-534.0821	-351.9858	-967.8933	-480.6938
Normal distribution	-566.1014	-538.0549	-445.3173	-1045.1071	-494.6102
Lognormal distribution	-560.4108	-546.4166	<u>-349.4094</u>	<u>-960.0402</u>	-517.6582
Weibull distribution	-574.5840	-538.9670	-393.3779	-990.7425	-488.3424
Gamma distribution	-556.5564	-535.3848	-361.7671	-968.1071	-501.1078
Gumbel distribution	-638.5664	-600.1647	-623.2454	-1177.3065	-521.1552
Uniform distribution	-713.8654	-685.9265	-743.3989	-1217.4954	-574.5622

<sup>1</sup> “\_” indicates the optimal univariate marginal PDF.



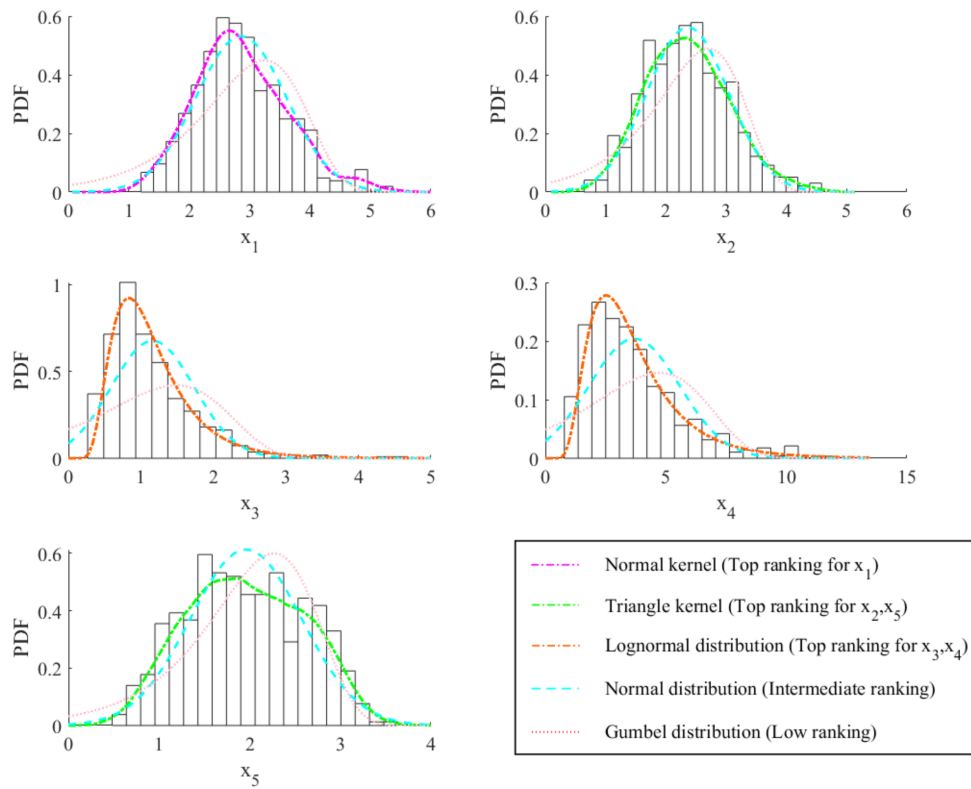


Figure 2. Univariate marginal PDFs of  $X_1$  to  $X_5$ .

The multivariate joint PDF of  $X$  is determined by Equation (11) with substituting the optimal marginal PDFs  $\hat{M}$  along with the associated optimal parameters  $\hat{\Theta}$  and the optimal parameter  $\hat{\psi}$  of the multivariate Gaussian copula. Figure 3 shows the projections of the multivariate joint PDF of  $X_1$  to  $X_5$ . Each subplot represents the projection of the multivariate joint PDF between two specific components of  $X$ . The black contour is the true PDF of Equation (25) while the green contour is the joint PDF by the proposed Copula-UQ. It can be shown that, even though the shape of the true PDF is irregular, the proposed Copula-UQ is capable of describing the statistical dependency structure.

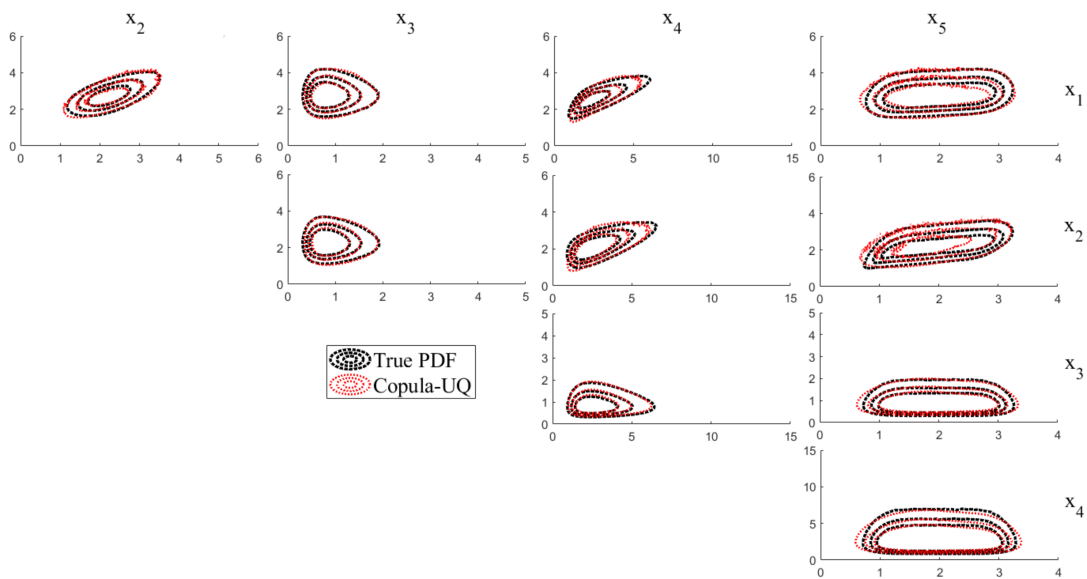


Figure 3. Projections of the multivariate joint PDF of  $X_1$  to  $X_5$ .

Figure 4 shows the comparisons of observed values and predicted values of  $X_2$  to  $X_5$ . The 45-degree reference line represents that the observed values and predicted values are identical. Each subplot shows the predicted value of the target variable, determined based on Equation (23), using the incomplete information (yellow dots) and complete information (blue dots) of the predictor variable. For example, for the yellow dots of the subplot in the upper left (for  $x_2$ ), the target variable, observed predictor variable and unobserved predictor variable are  $x_{ta} = x_2$ ,  $x_o = x_1$ ,  $x_{uo} = \{x_3, x_4, x_5\}$ , respectively. For the blue dots of the subplot in the upper left (for  $x_2$ ),  $x_{ta} = x_2$ ,  $x_o = \{x_1, x_3, x_4, x_5\}$ ,  $x_{uo}$  is an empty set. By comparing the scatter plots of yellow and blue dots, one can observe the evolution of the predicted value changes with respect to the amount of information given by the predictor variable. It can be anticipated that the predicted values can be improved (that is, the dots distributing closer to the 45-degree reference line) when given more information from the predictor variable. This conclusion can be confirmed from the subplots in the upper left (for  $x_2$ ), lower left (for  $x_4$ ) and lower right (for  $x_5$ ) of Figure 4. Note that there is insignificant improvement of the predicted values of the subplots located in the upper right (for  $x_3$ ), this is because of low correlations between  $x_3$  and other components, shown in Equation (26). This result shows that the proposed formulation of Equation (23), breaking through the second limitation (i.e., negligence of probabilistic prediction using available information) by conducting computation of marginalization and conditioning, is capable of making predictions even though the information of the predictor variable is incomplete.

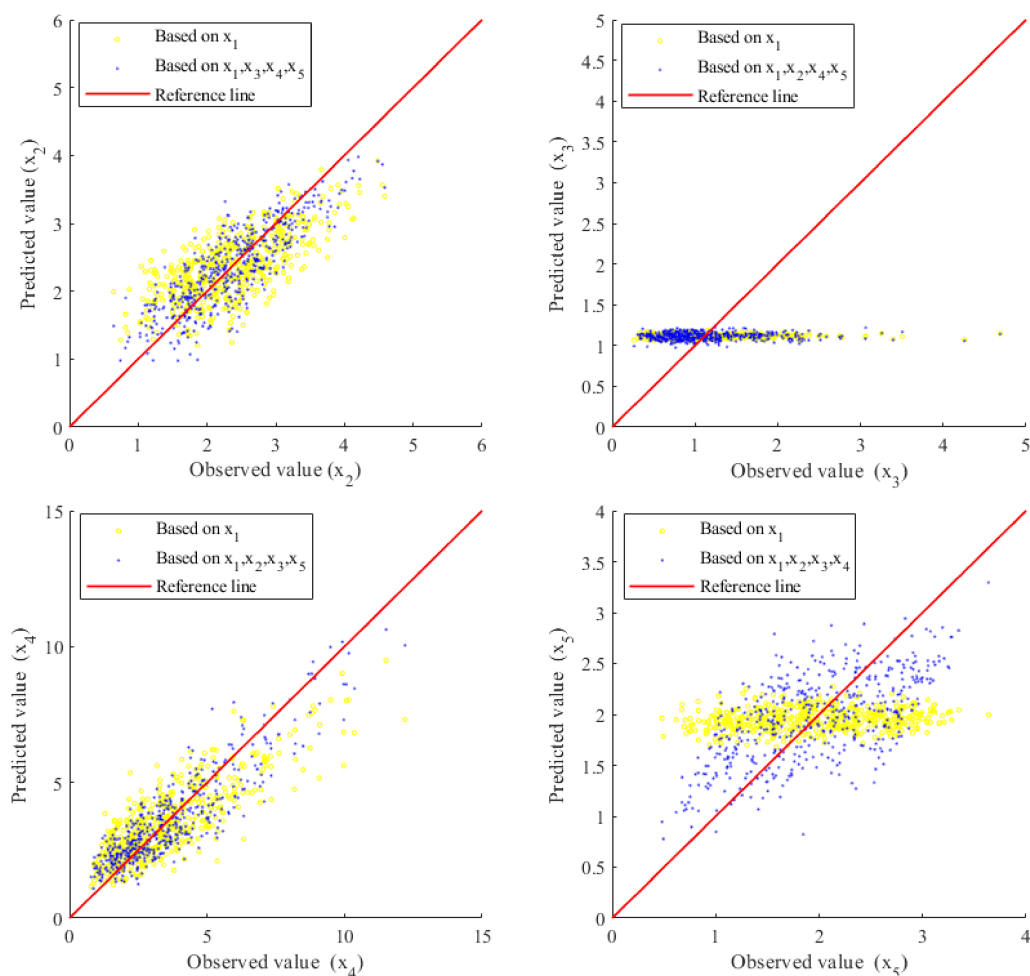


Figure 4. Comparisons of observed values and predicted values of  $X_2$  to  $X_5$ .

#### 4.2. Temperature Data of Multi-Sensor of a Monitored Bridge

Temperature is a critical loading factor for structures [32]. Variation of temperatures in structures significantly influences the material properties (for example, Young's modulus [32]), static characteristics (for example, deflection and deformation [32]), dynamic characteristics (for example, structural frequencies [33–35], damping ratios and mode shapes [36]) and boundary conditions [37]. Temperatures, including ambient air temperature and structural component temperature, of a multi-sensor of a structure are uncertain due to the fact that they are affected by not only the ambient factors, including air temperature variation, solar radiation intensity, humidity and wind speed, but also the complex processes of heat transfer [38]. Practically, UQ in temperatures are conducted based on temperature data measured from multiple sensors installed in different locations of a monitored structure [38–42]. As these works utilized traditional PDF modelling approaches, and modelling of temperature-related random variables was limited to two-dimensional. Here, due to the capacity of the multivariate joint PDF modelling of the proposed Copula-UQ, the dimension can be extended to  $D$ -dimensional, where  $D$  is the number of temperature sensors selected in the analysis.

This study utilized the proposed Copula-UQ to analyze temperature data of the multi-sensor of the Dowling Hall Footbridge [36]. The bridge, located on the Medford campus of Tufts University, has a two-span continuous steel frame (each span is 22 m) and a reinforced concrete deck. Temperatures of different locations are monitored using the type T thermocouples manufactured by Omega Engineering (measurement ranging from  $-250$  to  $+350$  °C). Multi-sensor layout for temperature monitoring can be referred to in Figure 7 of Reference [43]. There are in total ten temperature sensors and they can be divided into two sensor clusters according to their locations: the west span cluster and the east span cluster. The west span cluster includes sensors for pier temperature ( $C_1$ ), bridge deck temperature ( $C_2$ ), steel temperature at the south side ( $S_1$ ), steel temperature at the north side ( $S_3$ ) and air temperature ( $A_1$ ). The east span cluster includes sensors for pier temperature ( $C_4$ ), bridge deck temperature ( $C_3$ ), steel temperature at the south side ( $S_2$ ), steel temperature at the north side ( $S_4$ ) and air temperature ( $A_2$ ).

The temperature data can be accessed from Reference [44]. Figure 5 shows time histories of ten temperature sensors beginning on January 5 2010 and ending on May 2 2010. In each subplot, two sensors monitor the same type of temperature, but these two sensors belong to the west span cluster and the east span cluster, respectively. For example, in the first subplot, both  $C_1$  and  $C_4$  monitored pier temperature, but  $C_1$  and  $C_4$  belong to the west span cluster and the east span cluster, respectively. It can be observed that there is insignificant difference in measurement between two sensors monitoring the same type of temperature even though they belong to two different clusters. It is worth noting that there is difference between the steel temperature at the south and north sides of the bridge. The reason is due to the fact that the effects of sunlight to the south and north side are different. During the daytime hours, the sensor on the south side ( $S_1$  and  $S_2$ ) was significantly warmer than the sensor on the north side ( $S_3$  and  $S_4$ ) [36]. Therefore, temperature data of five sensors from the west span cluster ( $C_1, C_2, S_1, S_3, A_1$ ) are utilized for UQ. The corresponding correlation coefficient matrix is:

$$corr = \begin{bmatrix} 1 & 0.9049 & 0.8698 & 0.9601 & 0.9691 \\ 0.9049 & 1 & 0.8748 & 0.9345 & 0.9212 \\ 0.8698 & 0.8748 & 1 & 0.9522 & 0.9357 \\ 0.9601 & 0.9345 & 0.9522 & 1 & 0.9952 \\ 0.9691 & 0.9212 & 0.9357 & 0.9952 & 1 \end{bmatrix} \quad (27)$$

Table 3 shows the maximum log-likelihood value of the univariate marginal PDFs of  $C_1, C_2, S_1, S_3$  and  $A_1$ . It can be observed that the optimal PDFs of  $C_1, C_2, S_1, S_3$  and  $A_1$  are distributed as the nonparametric model with the Normal kernel. Figure 6 shows the univariate marginal PDFs of  $C_1, C_2, S_1, S_3$  and  $A_1$ . It is obvious that in each subplot, the top-ranking model fits the frequency histogram better than the intermediate- and low-ranking models, reconfirming the model class selection results in Table 3. Figure 7 shows the projections of the multivariate joint PDF of  $C_1, C_2, S_1, S_3$  and  $A_1$ .

It can be shown that the contours by the Copula-UQ are capable of quantifying the uncertainty of the multivariate temperature data.

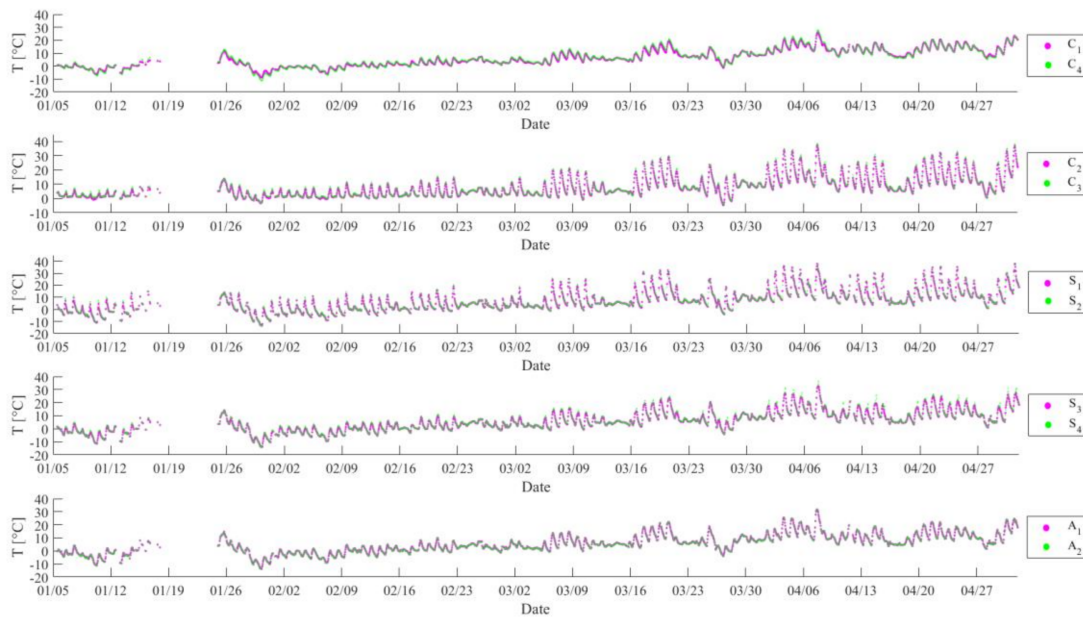


Figure 5. Time histories of ten temperature sensors beginning on January 5 2010 and ending on May 2 2010.

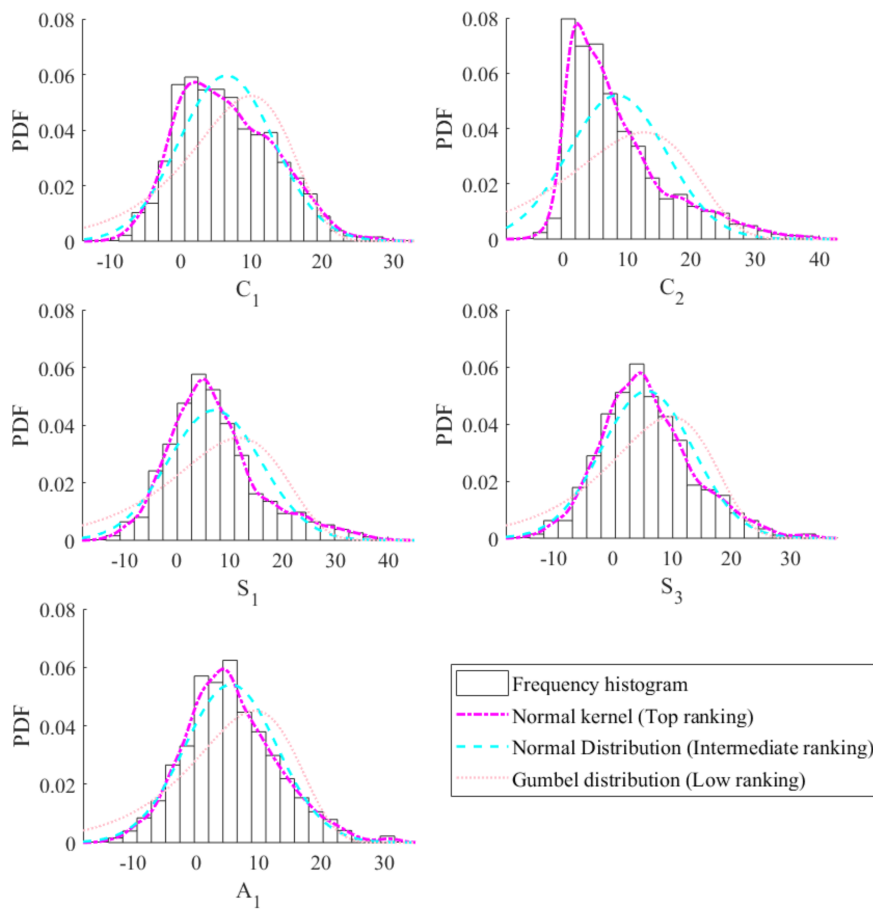


Figure 6. Univariate PDFs of  $C_1, C_2, S_1, S_3$  and  $A_1$ .

**Table 3.** The maximum log-likelihood value of the univariate marginal PDFs of  $C_1, C_2, S_1, S_3$  and  $A_1$ .

PDF	$C_1$	$C_2$	$S_1$	$S_3$	$A_1$
Normal kernel	<u>-8355.3340</u>	<u>-8216.0978</u>	<u>-8968.6926</u>	<u>-8737.8576</u>	<u>-8635.4290</u>
Uniform kernel	-8359.5995	-8246.7759	-8979.0882	-8750.4474	-8645.4155
Triangle kernel	-8355.4031	-8222.3736	-8968.8041	-8738.2284	-8635.7345
Epanechnikov kernel	-8358.0363	-8232.9844	-8973.6371	-8743.8474	-8640.5933
Normal distribution	-8442.4746	-8790.1815	-9149.7283	-8813.6539	-8691.3643
Gumbel distribution	-8752.8354	-9405.0940	-9648.7378	-9218.5528	-9072.9188
Uniform distribution	-9210.6580	-9614.1435	-10,153.6737	-9855.9299	-9716.0754

<sup>1</sup> “\_” indicates the optimal marginal PDF.

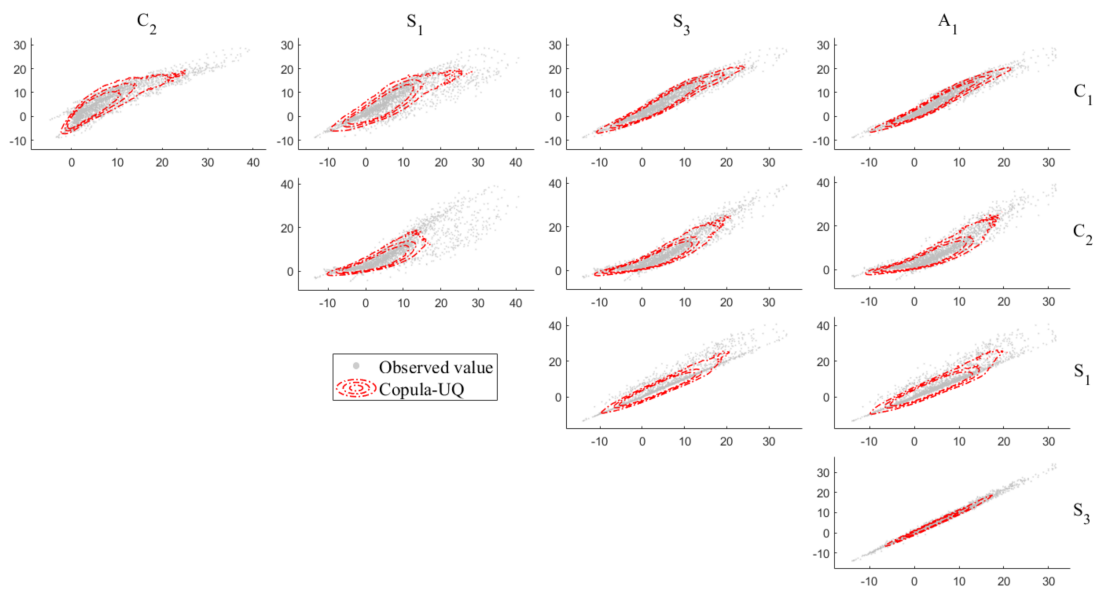
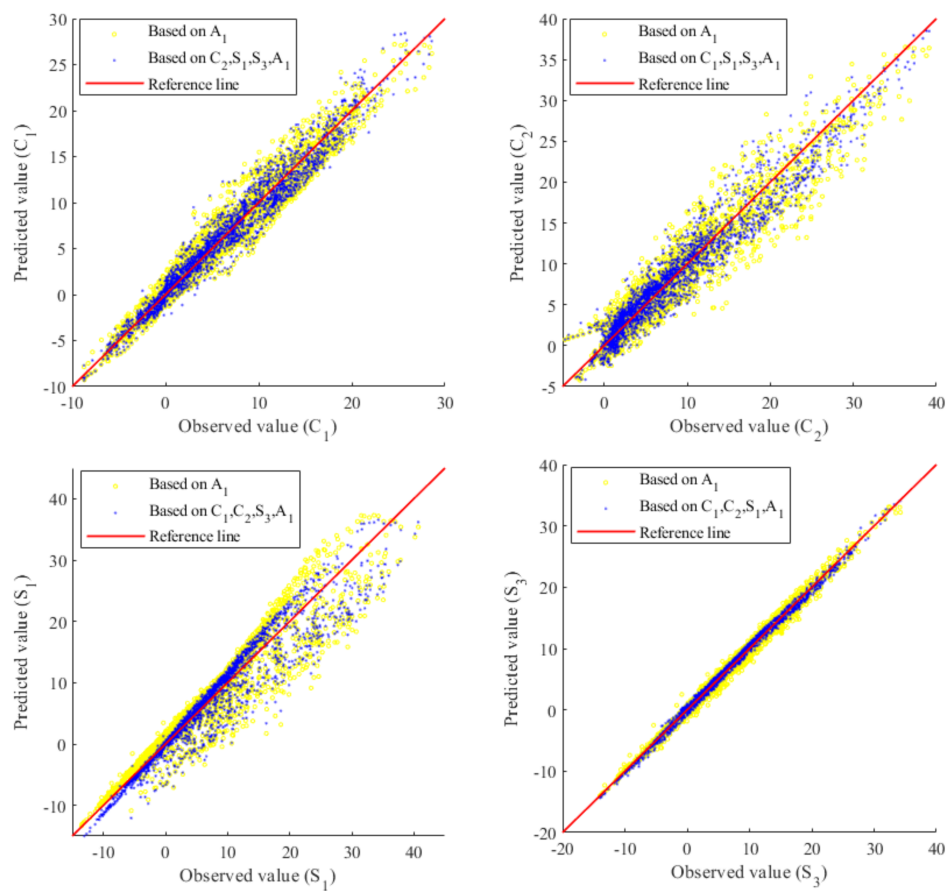
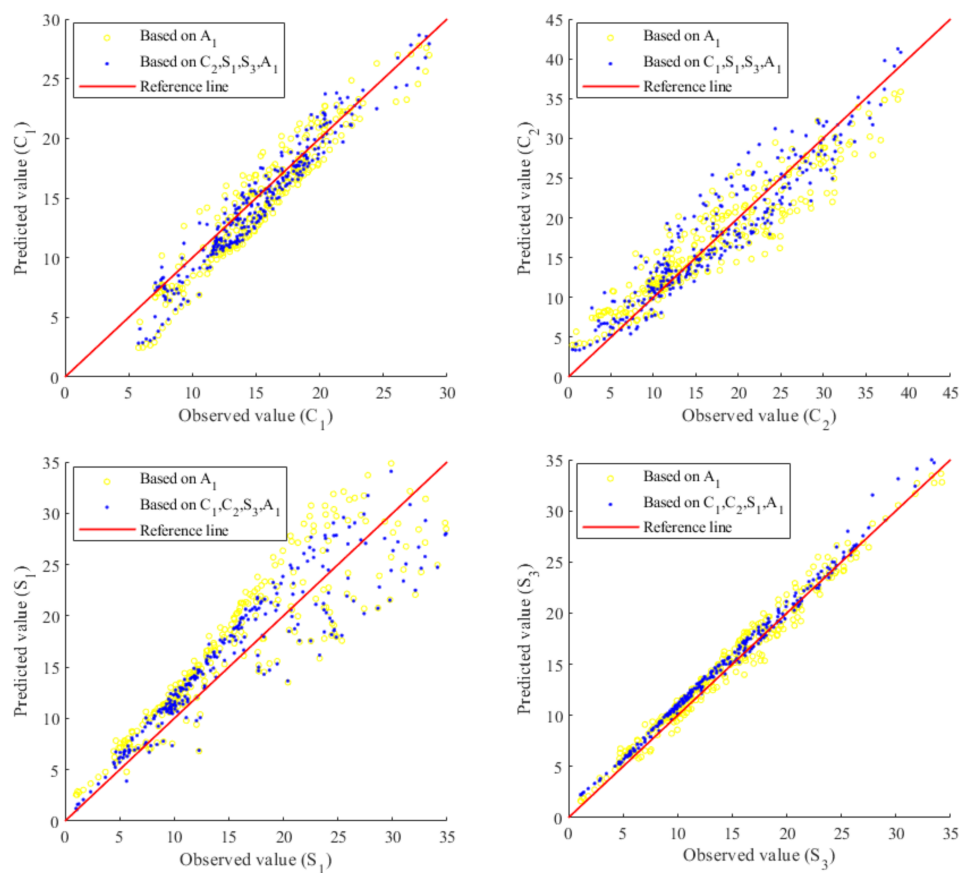
**Figure 7.** Projection of the multivariate joint PDF of  $C_1, C_2, S_1, S_3$  and  $A_1$ .

Figure 8, in the same fashion as Figure 4, shows the comparisons of observed values and predicted values of  $C_1, C_2, S_1$  and  $S_3$  (training dataset: full monitoring dataset, test dataset: full monitoring dataset). Again, for each subplot, it can be observed that the predicted values can be improved when given more information of the predictor variable. Note that the yellow dots correspond to incomplete information of the predictor variable due to a sensor fault. For example, for the yellow dots of the subplot in the upper left (for  $C_1$ ), the target variable, observed predictor variable and unobserved predictor variable are  $x_{ta} = C_1, x_o = A_1, x_{uo} = \{C_2, S_1, S_3\}$ , respectively. That is, the yellow dots show the predicted value of the target variable  $C_1$  using the information from the observed variable of normal sensor  $A_1$ , but without using the information from the unobserved variable of fault sensors  $C_2, S_1, S_3$  because of the fault status of these three sensors. For the blue dots of the subplot in the upper left (for  $C_1$ ),  $x_{ta} = C_1, x_o = \{A_1, C_2, S_1, S_3\}, x_{uo}$  is an empty set. From the four subplots of Figure 8, although the sensor fault issue enlarges the fluctuation of the yellow dots, the proposed Copula-UQ gives satisfactory results as the available information of sensor  $A_1$  is properly utilized for making predictions on the target variable.



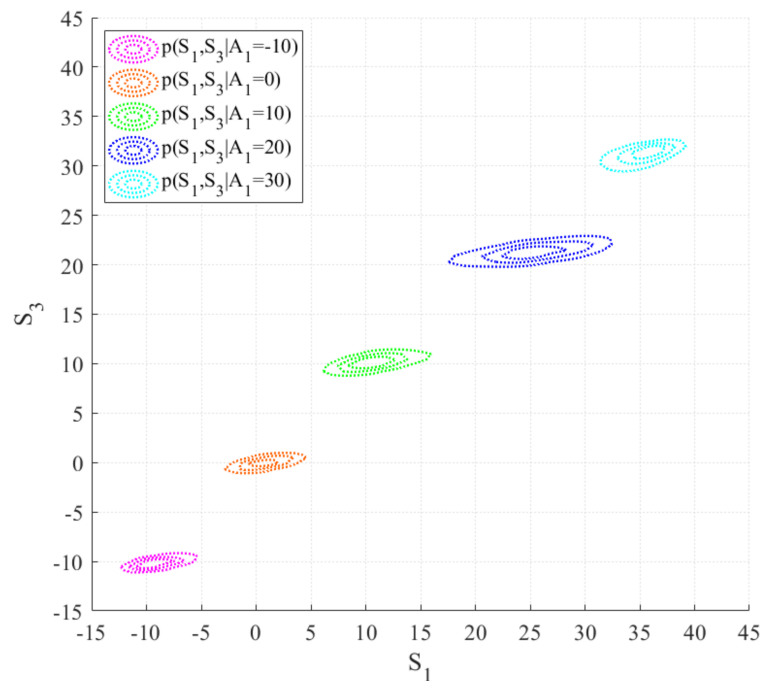
**Figure 8.** Comparisons of observed values and predicted values of  $C_1, C_2, S_1, S_3$  (training dataset: full monitoring dataset, test dataset: full monitoring dataset). <sup>2</sup> Note that the yellow dots correspond to incomplete information of the predictor variable due to missing data caused by a sensor fault issue.

For further validating the prediction capacity of the proposed Copula-UQ under data missing by sensor fault issue, a new computation is conducted as follows: (1) the monitored dataset was divided into the training dataset (data covering first 90% of days out of total monitoring period) and the test dataset (complement of the training dataset), (2) the marginal PDF along with the copula model was inferred based on the training dataset and (3) the prediction capacity of the trained copula model was validated based on the test dataset with or without data missing by sensor fault issue. Figure 9, in the same fashion as Figure 8, shows comparisons of observed values and predicted values of  $C_1, C_2, S_1$  and  $S_3$  (training dataset: dataset of first 90% of total number of the monitoring days, test dataset: complement of training dataset). It can be observed that the fluctuations of both the blue dots (corresponding to complete information from normal sensors) and the yellow dots (corresponding to incomplete information due to data missing by sensor fault issue) are acceptable. Therefore, it can be concluded that even though the training dataset is different from the test dataset, the proposed Copula-UQ still gives satisfactory results in multivariate PDF modelling and target variable prediction.



**Figure 9.** Comparisons of observed values and predicted values of  $C_1, C_2, S_1, S_3$  (training dataset: data covering first 90% days out of total monitoring period, test dataset: complement of training dataset).  
<sup>2</sup> Note that the yellow dots correspond to incomplete information of the predictor variable due to missing data caused by a sensor fault issue

Figure 10 shows the predicted joint PDFs between  $S_1$  and  $S_3$  under incomplete information with different given values of  $A_1$  only, and without observing information of fault sensors  $C_1$  and  $C_2$ . That is, the joint PDF  $p(S_1, S_3 | A_1 = \bar{A}_1)$  shows how the steel temperatures of the south and north sides evolve with changing the air temperature. As  $\bar{A}_1$  increases, the optimal values of  $p(S_1, S_3 | A_1 = \bar{A}_1)$  increased accordingly. It can be observed that the differences between the steel temperature ( $S_1$  or  $S_3$ ) and the air temperature ( $\bar{A}_1$ ) become more significant as  $\bar{A}_1$  increases. The reason is as follows: higher  $\bar{A}_1$  associates with higher solar radiation intensity. Given that the specific heat capacity of steel is higher than that of air, the increase of temperature of steel is more significant than that of air. It can also be observed that the temperature of  $S_3$  is lower than that of  $S_1$ . This result coincides with the on-site situation of sunlight of the Dowling Hall Footbridge, in that the sunlight intensity to the north side ( $S_3$ ) is lower than that to the south side ( $S_1$ ) [36]. The predicted joint PDFs among the temperature of different locations of the structure are important pieces of information for uncertain thermal loading and can be utilized for thermal-induced structural response assessment.



**Figure 10.** Predicted joint PDFs between  $S_1$  and  $S_3$  under incomplete information with given different values of  $A_1$  only, and without observing information of fault sensors  $C_1$  and  $C_2$ .

## 5. Conclusions

This paper proposed the Copula-UQ for multivariate joint PDF modelling, inference on model class selection and parameter identification, and probabilistic prediction using incomplete information, and presented one example using simulated data and one example using temperature data of a multi-sensor of a monitored bridge. For inference on univariate marginal PDFs, the results show that, in general cases, the optimal univariate marginal PDFs of different dimensions are different, so the introduction of both parametric and nonparametric models is necessary because it provides a large solution space for uncertainty quantification. For prediction on the target variable using the complete (from normal sensors) or incomplete information (due to missing data caused by a sensor fault issue) of the predictor variable, the proposed Copula-UQ is capable of obtaining the PDF of the target variable. The proposed methodology can be extended to tackle different multivariate joint PDF modelling problems in SHM with emphasizing the prediction purpose under incomplete information with a sensor fault issue. This important piece of information of the PDF of the target variable can be utilized for uncertainty propagation in further analysis.

**Author Contributions:** Conceptualization, methodology, formal analysis, investigation, resources, writing—original draft preparation, supervision, project administration, funding acquisition, H.-Q.M.; formal analysis, Software, validation, data curation, writing—review and editing, visualization, H.-T.L. and J.-H.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Pearl River S&T Nova Program of Guangzhou (201806010172), Scientific Research Fund of Institute of Engineering Mechanics, China Earthquake Administration (Grant No. 2019 EEEVL0401), Science and Technology Program of Guangzhou (201804020069).

**Acknowledgments:** The authors thank all the reviewers and editors for their significant help and useful suggestions.

**Conflicts of Interest:** The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.



## References

1. Yuen, K.-V. *Bayesian Methods for Structural Dynamics and Civil Engineering*; John Wiley & Sons (Asia) Pte Ltd.: Singapore, 2010; ISBN 9780470824542.
2. Bernieri, A.; Ferrigno, L.; Laracca, M.; Molinara, M. Crack shape reconstruction in eddy current testing using machine learning systems for regression. *IEEE Trans. Instrum. Meas.* **2008**, *57*, 1958–1968. [[CrossRef](#)]
3. Memmolo, V.; Maio, L.; Boffa, N.D.; Monaco, E.; Ricci, F. Damage detection tomography based on guided waves in composite structures using a distributed sensor network. *Opt. Eng.* **2015**, *55*, 11007. [[CrossRef](#)]
4. Memmolo, V.; Ricci, F.; Boffa, N.D.; Maio, L.; Monaco, E. Structural health monitoring in composites based on probabilistic reconstruction techniques. *Procedia Eng.* **2016**, *167*, 48–55. [[CrossRef](#)]
5. Burgos, D.A.T.; Vargas, R.C.G.; Pedraza, C.; Agis, D.; Pozo, F. Damage identification in structural health monitoring: A brief review from its implementation to the use of data-driven applications. *Sensors* **2020**, *20*, 733. [[CrossRef](#)] [[PubMed](#)]
6. Kralovec, C.; Schagerl, M. Review of Structural Health Monitoring Methods Regarding a Multi-Sensor Approach for Damage Assessment of Metal and Composite Structures. *Sensors* **2020**, *20*, 826. [[CrossRef](#)] [[PubMed](#)]
7. Mu, H.Q.; Yuen, K.V. Novel outlier-resistant extended Kalman filter for robust online structural identification. *J. Eng. Mech.* **2015**, *141*, 04014100. [[CrossRef](#)]
8. Mu, H.Q.; Kuok, S.C.; Yuen, K.V. Stable robust extended Kalman filter. *J. Aerosp. Eng.* **2017**, *30*. [[CrossRef](#)]
9. Beck, J.L. Bayesian system identification based on probability logic. *Struct. Control. Heal. Monit.* **2010**, *17*, 825–847. [[CrossRef](#)]
10. Krishnan Nair, K.; Kiremidjian, A.S. Time Series Based Structural Damage Detection Algorithm Using Gaussian Mixtures Modeling. *J. Dyn. Syst. Meas. Control.* **2007**, *129*, 285–293. [[CrossRef](#)]
11. Qiu, L.; Fang, F.; Yuan, S.; Boller, C.; Ren, Y. An enhanced dynamic Gaussian mixture model-based damage monitoring method of aircraft structures under environmental and operational conditions. *Struct. Heal. Monit.* **2019**, *18*, 524–545. [[CrossRef](#)]
12. Santos, A.; Figueiredo, E.; Silva, M.; Santos, R.; Sales, C.; Costa, J.C.W.A. Genetic-based EM algorithm to improve the robustness of Gaussian mixture models for damage detection in bridges. *Struct. Control. Heal. Monit.* **2017**, *24*, e1886. [[CrossRef](#)]
13. Qiu, L.; Yuan, S.; Mei, H.; Fang, F. An improved Gaussian mixture model for damage propagation monitoring of an aircraft wing spar under changing structural boundary conditions. *Sensors* **2016**, *16*, 291. [[CrossRef](#)] [[PubMed](#)]
14. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT press: Cambridge, MA, USA, 2012; ISBN 0262304325.
15. Wist, H.T.; Myrhaug, D.; Rue, H. Statistical properties of successive wave heights and successive wave periods. *Appl. Ocean. Res.* **2004**, *26*, 114–136. [[CrossRef](#)]
16. Sagrilo, L.V.S.; Naess, A.; Doria, A.S. On the long-term response of marine structures. *Appl. Ocean. Res.* **2011**, *33*, 208–214. [[CrossRef](#)]
17. Lebrun, R.; Dutfoy, A. An innovating analysis of the Nataf transformation from the copula viewpoint. *Probabilistic Eng. Mech.* **2009**, *24*, 312–320. [[CrossRef](#)]
18. Zhang, Y.; Kim, C.W. Use of copula theory in the long-term health monitoring for deteriorated bridges. In Proceedings of the 12th International Conference on Structural Safety and Reliability, Vienna, Austria, 6–10 August 2017; pp. 2050–2059.
19. Fan, X.P.; Liu, Y.F. Time-variant reliability prediction of bridge system based on BDGCM and SHM data. *Struct. Control. Heal. Monit.* **2018**, *25*, e2185. [[CrossRef](#)]
20. Pan, Y.; Ou, S.; Zhang, L.; Zhang, W.; Wu, X.; Li, H. Modeling risks in dependent systems: A Copula-Bayesian approach. *Reliab. Eng. Syst. Saf.* **2019**, *188*, 416–431. [[CrossRef](#)]
21. Liu, Y.; Zhang, H.; Li, D.; Deng, Y.; Jiang, N. Fatigue reliability assessment for orthotropic steel deck details using copulas: Application to Nan-Xi Yangtze River Bridge. *J. Bridg. Eng.* **2018**, *23*, 4017123. [[CrossRef](#)]
22. Srinivas, S.; Menon, D.; Meher Prasad, A. Multivariate simulation and multimodal dependence modeling of vehicle axle weights with copulas. *J. Transp. Eng.* **2006**, *132*, 945–955. [[CrossRef](#)]
23. Zhang, Y.; Beer, M.; Quek, S.T. Long-term performance assessment and design of offshore structures. *Comput. Struct.* **2015**, *154*, 101–115. [[CrossRef](#)]

24. Panamtash, H.; Zhou, Q.; Hong, T.; Qu, Z.; Davis, K.O. A copula-based Bayesian method for probabilistic solar power forecasting. *Sol. Energy* **2020**, *196*, 336–345. [[CrossRef](#)]
25. Hollander, M.; Wolfe, D.A.; Chicken, E. *Nonparametric Statistical Methods*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 751, ISBN 1118553292.
26. Cherubini, U.; Luciano, E.; Vecchiato, W. *Copula Methods in Finance*; John Wiley & Sons: Hoboken, NJ, USA, 2004; ISBN 0470863455.
27. Sklar, M. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. Stat. univ. Paris* **1959**, *8*, 229–231.
28. Nelsen, R.B. *An Introduction to Copulas*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007; ISBN 0387286780.
29. Bouyé, E.; Durrleman, V.; Nikeghbali, A.; Riboulet, G.; Roncalli, T. Copulas for finance—a reading guide and some applications. *SSRN Electron. J.* **2000**. [[CrossRef](#)]
30. Mu, H.Q.; Hu, Q.; Guo, H.Z.; Zhang, T.Y.; Su, C. Uncertainty Quantification of Load Effects under Stochastic Traffic Flows. *Int. J. Struct. Stab. Dyn.* **2019**, *19*, 1–18. [[CrossRef](#)]
31. Bowman, A.W.; Azzalini, A. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*; Oxford University Press: New York, NY, USA, 1997; Volume 18, ISBN 0198523963.
32. Xu, Y.L.; Xia, Y. *Structural Health Monitoring of Long-Span Suspension Bridges*; CRC Press: Boca Raton, FL, USA, 2011; ISBN 0415597935.
33. Mu, H.-Q.; Yuen, K.-V.; Kuok, S.-C. Relevance feature selection of modal frequency–ambient condition pattern recognition in structural health assessment for reinforced concrete buildings. *Adv. Mech. Eng.* **2016**, *8*, 1–12. [[CrossRef](#)]
34. Mu, H.-Q.; Zheng, Z.-J.; Wu, X.-H.; Su, C. Bayesian network-based modal frequency–multiple environmental factors pattern recognition for the Xinguang Bridge using long-term monitoring data. *J. Low Freq. Noise Vib. Act. Control* **2020**, *39*, 545–549. [[CrossRef](#)]
35. Mu, H.-Q.; Yuen, K.-V. Modal frequency–environmental condition relation development using long-term structural health monitoring measurement: Uncertainty quantification, sparse feature selection and multivariate prediction. *Measurement* **2018**, *130*, 384–397. [[CrossRef](#)]
36. Moser, P.; Moaveni, B. Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge. *Mech. Syst. Signal. Process.* **2011**, *25*, 2336–2357. [[CrossRef](#)]
37. Peeters, B.; De Roeck, G. One-year monitoring of the Z24-Bridge: Environmental effects versus damage events. *Earthq. Eng. Struct. Dyn.* **2001**, *30*, 149–171. [[CrossRef](#)]
38. Ding, Y.; Zhou, G.; Li, A.; Wang, G. Thermal field characteristic analysis of steel box girder based on long-term measurement data. *Int. J. Steel Struct.* **2012**, *12*, 219–232. [[CrossRef](#)]
39. Xu, Y.L.; Chen, B.; Ng, C.L.; Wong, K.Y.; Chan, W.Y. Monitoring temperature effect on a long suspension bridge. *Struct. Control. Heal. Monit.* **2010**, *17*, 632–653. [[CrossRef](#)]
40. Wang, G.; Ding, Y.; Wang, X.; Yan, X.; Zhang, Y. Long-Term Temperature Monitoring and Statistical Analysis on the Flat Steel-Box Girder of Sutong Bridge. *J. Highw. Transp. Res. Dev. English Ed.* **2014**, *8*, 63–68. [[CrossRef](#)]
41. Gaoxin, W.; Youliang, D. Research on monitoring temperature difference from cross sections of steel truss arch girder of Dashengguan Yangtze Bridge. *Int. J. Steel Struct.* **2015**, *15*, 647–660. [[CrossRef](#)]
42. Yang, D.; Yi, T.; Li, H.; Zhang, Y. Monitoring and analysis of thermal effect on tower displacement in cable-stayed bridge. *Measurement* **2018**, *115*, 249–257. [[CrossRef](#)]
43. Moser, P.; Moaveni, B. Design and Deployment of a Continuous Monitoring System for the Dowling Hall Footbridge. *Exp. Tech.* **2013**, *37*, 15–26. [[CrossRef](#)]
44. Continuous Monitoring of the Dowling Hall Footbridge. Available online: [https://engineering.tufts.edu/cee/shm/research\\_BM\\_continuousMonitoring.asp](https://engineering.tufts.edu/cee/shm/research_BM_continuousMonitoring.asp) (accessed on 1 August 2020).

