Imprecise probability analysis of steel structures subject to atmospheric corrosion

Hao Zhang[a,*], Loc Ha[a,1], Quanwang Li[b,2], Michael Beer[c,3]

[a] *School of Civil Engineering, The University of Sydney, NSW 2006, Australia*
[b] *Department of Civil Engineering, Tsinghua University, Beijing 100084, China.*
[c] *Institute for Computer Science in Civil Engineering, Leibniz University Hannover, Germany. Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK. International joint Research Center for Engineering Reliability and Stochastic Mechanics (ERSM), Tongji University, Shanghai, China.*

Abstract

Evaluating the behaviour of deteriorating steel structures is complicated by the inherent uncertainties in the corrosion process. Theoretically, these uncertainties can be modeled using a probabilistic approach. However, there are practical difficulties in identifying the probabilistic model for the deterioration process as the actual corrosion data are rather limited. Also, the dependencies between different random variables are often vaguely known and, thus, not included in the modeling. This paper proposes a probabilistic analysis framework for modeling the atmospheric corrosion of steel structures with incomplete information. The framework is based on the theory of imprecise probability and copula. Two examples are presented to illustrate the methodology. The role of epistemic uncertainties on structural reliability is investigated through the examples.

Keywords: atmospheric corrosion, corrosion, copula, deterioration,

[*] Corresponding author (H. Zhang). Tel.: +61293513923, E-mail address: hao.zhang@sydney.edu.au

[1] E-mail address: loc.ha@sydney.edu.au

[2] E-mail address: li_quanwang@tsinghua.edu.cn

[3] E-mail address: beer@bauinf.uni-hannover.de

imprecise probability, interval analysis, probability box, steel

1. Introduction

For the safety assessment of deteriorating steel structures, it is crucial to develop a reliable probabilistic model of deterioration to predict the temporal changes to structural resistance [1, 2]. The deterioration of steel structures is a stochastic process with high uncertainties and variabilities. Recent works have treated the uncertainties using a pure probabilistic approach [3, 4]. This approach requires that all statistical characteristics for each uncertainty can be determined reliably from sufficient observational data. In practice, however, available real-world data on structural corrosion are very limited, and the selection of probabilistic models (e.g., distribution type and/or distribution parameters) for uncertain variables is so generally based on limited information and/or subjective judgment.

It is thus advisable to consider the distribution itself as uncertain when the available data is limited. Statistical estimations provide us with distribution functions for the sampling uncertainty, which depends on the sample size. This uncertainty is reducible with an increasing amount of information/data. From this angle, it may be understood as epistemic uncertainty. Within a pure probabilistic framework, epistemic uncertainty can be handled with Bayesian approaches. Uncertain parameters of a probabilistic model can be described with prior distributions and updated by means of even limited data. They can then be modeled by Bayesian random variables and introduced formally, together with the remaining (aleatory) uncertainties, in the probabilistic analysis [5]. Judgmental information is needed to characterize the epistemic uncertainties. The characterization of the epistemic uncertainties can be substantiated by using the Bayesian updating rule when data become available. However, when the data is very limited, the result of the Bayesian approach remains as almost purely subjective.

Alternatively, an imprecisely known probability distribution can be modeled by a family of all candidate probability distributions which are compatible with available data. This is the idea of the theory of imprecise probabilities [6]. Dealing with a set of probability distributions is essentially different from a Bayesian approach. A practical way to represent the distribution family is to use a probability bounding approach by specifying the lower and upper bounds of the imprecise probability distribution. This corresponds to the use of an interval to represent an unknown but bounded number. Consequently, a unique failure probability cannot be determined. Instead, the failure probability is obtained as an interval whose width reflects the imprecision of the distribution model in the calculated reliability.

A popular uncertainty model using the probability bounding approach is the probability box (p-box for short) structure [7]. A p-box is closely related to other set-based uncertainty models such as random sets, fuzzy probabilities, Dempster-Shafer evidence theory and random intervals. In many cases, these uncertainty models can be converted into each other, and thus considered to be equivalent [7–10]. Therefore, the p-box approach presented in this paper is also applicable to other set-based uncertainty models. The approach of imprecise probability generally requires less subjective information than the Bayesian approach. It can be argued that, from a frequentist point of view, the epistemic uncertainties in the probability distribution can be more faithfully represented using a probability bounding approach [6, 7, 11].

Conventional probabilistic analysis often neglects the correlations and dependencies between random variables. This assumption is a common practice partly due to its mathematical convenience, but more likely due to the limited availability of data. It has been shown that the wrong assumption of dependence can lead to unreliable predictions for risk assessments [12]. Copula theory is a powerful tool for the dependence modeling of multivariate data. A copula is a joint cumulative distribution function (CDF) with uniform marginal. Copula theory has been used to model dependence in probability boxes. Ref. [12] proposed a dependence bounds convolution approach in which the uncertainties are modelled as Dempster-Shafer structures and the dependence is expressed as a given parametric copula. This method is useful for calculations of basic arithmetic operations with small numbers of variables. In [13], copula theory is combined with random sets for computing the lower and upper bounds of a failure probability.

This paper proposes a practical framework for uncertainty analysis using dependent p-boxes in which copulas describe the dependence. The Akaike Information Criterion is used to select the copula model that provides the best fit to the observational data. The confidence intervals of the copula parameter are estimated using the Bootstrap method. The dependent p-boxes are propagated through interval Monte Carlo (MC) simulation in order to assess structural reliability. The framework is applied to the time-dependent reliability analysis of steel structures subject to atmospheric correlations, and is demonstrated through two examples. The importance of epistemic uncertainty in the probabilistic modeling including dependencies is demonstrated on its influence on the reliability estimates.

2. Dependent Probability boxes

2.1. Probability boxes with dependencies

Let $F_X(x)$ denote the cumulative distribution function (CDF) for a real-valued random variable X. A probability box is defined by a pair of CDFs,

$\underline{F}_X(x)$ and $\overline{F}_X(x)$, which form the envelopes of the probability family

$$\mathscr{P} = \{P|\forall x \in \mathbb{R}, \underline{F}_X(x) \leq F_X(x) \leq \overline{F}_X(x)\}. \tag{1}$$

A p-box thus represents an $F_X(\,)$ which is imprecisely known except that it is within the two bounding CDFs. It can be seen that $\underline{F}_X(\,)$ and $\overline{F}_X(\,)$ are the lower and upper probabilities of the event $X \leq x$. Detailed background can be found elsewhere [7]. There are various ways to define p-boxes such as utilizing Kologorox-Smirnow (K-S) confidence limits, Chebyshev's inequality, or by distributions with interval parameters, depending on the amount of available information [14].

The modeling of dependencies between probability boxes follows the concept of dependence between random variables. Both Pearson correlation and rank correlation have been adopted for p-boxes, but retaining their limitations known from probability theory. Thus, copula models have been suggested to describe dependence between p-boxes [15]. There are two main advantages of using copulas for this purpose. First, copulas can account for various types of dependencies. Second, the copula is flexible in selecting the appropriate dependence model independently from choosing the marginal distributions for each variable [16].

2.2. A brief introduction of copulas

A copula is a multivariate CDF for which the marginal distribution of each variable is uniform. According to Sklar's Theorem, a joint distribution can be expressed in terms of the marginal distribution functions and a copula which describes the dependence structure between the variables. Consider a d-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ with margins $F_i(x)$, $i = 1, \ldots, d$. There exists a copula C such that the joint CDF, denoted by

$F_{\mathbf{X}}(x_1, \ldots, x_d)$, can be written as

$$F_{\mathbf{X}}(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{2}$$

There are two common classes of copulas; Gaussian and Archimedean. The Gaussian copula is used for the normal dependence structure. This structure can be estimated from its only parameter of a correlation matrix [17]. In a non-normal case, Archimedean copulas are often used to model the dependence structure in the data. The class of copula has a closed-form of representation,

$$C(u_1, u_2, ..., u_d, \theta) = \varphi^{-1}\left(\varphi(u_1), \varphi(u_2), ..., \varphi(u_d, \theta)\right), \tag{3}$$

in which φ is a generator with φ^{-1} completely monotonic on $[0, \infty) \times [0, \infty) ... \times [0, \infty)$ (d-dimensional copula). The copula parameter, θ, can be related to various dependence structures of Archimedean copulas. The most common Archimedean copulas include Clayton, Gumbel and Frank copulas which are summarised in Table 1. Details about copulas can be found elsewhere, e.g., [18].

Table 1: Some common Archimedean copulas.

Copula	Form	Range of θ
Clayton	$\mathrm{C}(u_1, u_2, \theta) = \left(u_1{}^{-\theta} + u_2{}^{-\theta} - 1\right)^{-1/\theta}$	$(0, \infty)$
Frank	$\mathrm{C}(u_1, u_2, \theta) = -\theta^{-1} \log\left\{1 + \frac{\left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)}{e^{-\theta} - 1}\right\}$	$\mathbb{R}$
Gumbel	$\mathrm{C}(u_1, u_2, \theta) = \exp\left(-\left((-\log(u_1))^{\theta} + (-\log(u_1))^{\theta}\right)^{1/\theta}\right)$	$[1, \infty)$

2.3. Estimation of copula parameter

Different copulas represent different dependence structures on the data. Thus, we establish the copula model in two steps. Step 1 is devoted to estimate the parameters for a number of candidate copulas. The copulas considered in this paper (e.g., Clayton, Gumbel and Frank copulas) involve only one parameter, denoted by θ. The copula parameter θ can be estimated by the classical maximum likelihood estimation (MLE). The MLE yields a point estimate of θ.

In step 2 the best-fit copula model for the given observed data and (point-) estimated parameter is selected. This is realized based on the Akaike Information Criterion (AIC), which has particular suitability for best fit estimations when the samples are small [19]. The AIC is given by

$$\text{AIC} = -2 \log L + 2q, \tag{4}$$

in which $\log L$ is the log-likelihood function and q is the number of parameters of the copula model [20]. A copula model with a smaller AIC-value fits the data better.

When the observational data is quite limited, it is desirable to calculate an interval estimate of θ to indicate the range over which the copula may lie with a certain confidence. The present work uses the Bootstrap method [21] to construct confidence intervals of copula parameters.

Suppose we have n pairs of data points $(\{x_1, y_1\}, \ldots, \{x_n, y_n\})$ representing two dependent random variables X and Y. We aim to estimate the copula $C(X, Y)$ to characterize their dependence. The procedure of computing the $100(1-2\alpha)\%$ confidence interval for the copula parameter θ can be summarized as follows

1. Compute a point estimate, $\hat{\theta}$, for θ from the original dataset.

2. Construct a Bootstrap sample $(\{x_1^*, y_1^*\}, \dots, \{x_n^*, y_n^*\})$. Compute the copula parameter θ^* and the Bootstrap difference $\delta^* = \theta^* - \hat{\theta}$.
3. Repeat Step 2 for B times. Thus, we obtain $(\delta_1^*, \dots, \delta_B^*)$, in which δ_i^* represents the Bootstrap difference for the ith Bootstrap sample.
4. Determine the $100(\alpha)$th and $100(1-\alpha)$th percentile of $(\delta_1^*, \dots, \delta_B^*)$, denoted by δ_α^* and $\delta_{1-\alpha}^*$. Then $100(1-2\alpha)\%$ confidence interval for θ is calculated as $[\hat{\theta} - \delta_\alpha^*, \hat{\theta} - \delta_{1-\alpha}^*]$.

3. Interval Monte Carlo simulation with dependent p-boxes

We follow the concept of propagating p-boxes using simulation-based methods (e.g., interval Monte Carlo simulation or similar approaches) [22–25]. Consider a mapping $g : \mathbf{X} \to Y$, $\mathbf{X} = (X_1, X_2)$ are basic variables represented by p-boxes. Further, X_1 and X_2 are dependent through a copula C. The response quantity Y, as a function of X_1 and X_2, is another p-box. Let $F_Y(y)$ denote the CDF of Y. We are interested to determine the p-box structure for the response quantity Y, i.e., bounds on $F_Y(y)$.

Monte Carlo simulation involves repeated random sampling from each input distribution and to observe the result. Since only the bounds of CDF's for $\mathbf{X}$ are known, it is not possible to generate point samples but only interval samples. Let $[\overline{F}_{X1}, \underline{F}_{X1}]$ and $[\overline{F}_{X2}, \underline{F}_{X2}]$ be upper and lower bounds of CDF for X_1 and X_2. Random interval samples from X_1 and X_2 can be generated as follows.

1. Generate a sample of N dependent uniform variate $(\{u_1^i, u_2^i\}, i = 1, .., N)$ from the specified copula C.

2. Generate pairs of dependent random interval samples by

$$[\underline{x}_1^i, \overline{x}_1^i] = [\overline{F}_{X1}^{-1}(u_1^i), \underline{F}_{X1}^{-1}(u_1^i)],$$
$$[\underline{x}_2^i, \overline{x}_2^i] = [\overline{F}_{X2}^{-1}(u_2^i), \underline{F}_{X2}^{-1}(u_2^i)],\ i = 1, 2, \ldots, N$$

in which $\overline{F}^{-1}$ and $\underline{F}^{-1}$ denotes the inverse functions of upper and lower bounds of a p-box.

For the sampling in the first step we utilize a common method, see e.g., [18]:

1. Generate two independent standard uniform variates u_1 and t.
2. Set $u_2 = c_u^{-1}(t)$, where c_u^{-1}denotes a quasi-inverse of c_u.
3. $\{u_1, u_2\}$ is a pair of uniform variates with the specified copula C.

Once the correlated random interval samples are generated, the empirical lower and upper bounds for the CDF of Y can be calculated as

$$\begin{aligned}\underline{F}_Y(y) &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{I}[\overline{g}(\mathbf{x}_i) \leq y],\\ \overline{F}_Y(y) &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{I}[\underline{g}(\mathbf{x}_i) \leq y],\end{aligned} \tag{5}$$

in which N = total number of simulations, $\mathbf{x}_i = ([\underline{x}_1^i, \overline{x}_1^i], [\underline{x}_2^i, \overline{x}_2^i])$, $\mathbf{I}[\,]$ is the indicator function, having the value 1 if $[\,]$ is "true" and the value 0 if $[\,]$ is "false". $\overline{g}$ and $\underline{g}$ represent a lower bound and an upper bound for $g(\mathbf{x}_i)$, i.e.,

$$\begin{aligned}\underline{g}(\mathbf{x}_i) &= \min\{g(X_1, X_2) : \underline{x}_1^i \leq X_1 \leq \overline{x}_1^i,\ \underline{x}_2^i \leq X_2 \leq \overline{x}_2^i\},\\ \overline{g}(\mathbf{x}_i) &= \max\{g(X_1, X_2) : \underline{x}_1^i \leq X_1 \leq \overline{x}_1^i,\ \underline{x}_2^i \leq X_2 \leq \overline{x}_2^i\}.\end{aligned} \tag{6}$$

Computing Eq. (6) involves the calculation of the range of function g when the inputs vary in certain closed intervals. The problem of finding the range of a function is solved on the basis of interval analysis [26]. A variety of solution techniques have been proposed, including the interval arithmetic approach, combinatorial method, perturbation method, and optimization method, etc. Reliable methods are available to compute the bounds of responses of structures with reasonable accuracy when the structural stiffness and geometrical properties and loads vary in relatively narrow intervals [27]. It should be noted that the burden of interval analysis can be reduced if the response quantity Y is monotonic with respect to the input variables.

4. Atmospheric corrosion model for steel structures

For steel structures, corrosion is considered as the most dominant form of deterioration. Corrosion is a product of the chemical reaction by electrochemical oxidation of metals and oxidant when a steel surface is left unprotected from the environment. This chemical reaction causes a reduction in net area of a member; thus, it leads to a reduction in the structural capacity of a steel member.

Depending on the environment where the steel is exposed, corrosion processes can be broadly classified as atmospheric corrosion, immersion corrosion and underground corrosion. The present paper considers the atmospheric corrosion in rural-urban environment to illustrate the proposed uncertainty analysis framework. It should be noted that corrosions due to salted air (marine atmospheric corrosion), de-icing chemical, etc. have higher impact to the failure of a structure. However, they are beyond the scope of the present study.

The available models for time-variant atmospheric corrosion of steel are commonly based on the mass loss or penetration depth loss from experiments. They include time variable and several regression coefficients in the form of power formula to capture the corrosion process. A widely-accepted model for long-term atmospheric corrosion of steel conforms to an equation of the form [28–30]:

$$c(t) = A \cdot t^B, \tag{7}$$

in which $c(t)$ is the corrosion loss after t years, A is the corrosion loss after one year, and B is a constant representing the slope of the logarithmic transformation of Eq. (7). The power function was derived based on the diffusional process of oxygen through rust layers. Due to its simplicity, the power function has a long history in modelling of atmospheric corrosion for steel structures [31, 32]. This model was adopted in the present study.

4.1. Uncertainties in the corrosion model

The coefficients A and B in Eq. (7) were studied in [29, 33]. The study showed that A and B are dependent on environmental parameters including ambient temperature, moisture of environment and presence of pollutants, etc. If the site-specific environment information is not available, the values of A and B can be estimated according to the general classification of environment, i.e., marine, urban and rural environment.

The present study focuses on the modelling of atmospheric corrosion of carbon steels exposed in rural-urban environments. A total of 62 atmospheric corrosion data in rural-urban environment are complied from the literature [34–38]. These corrosion data are fitted with Eq. (7) and the coefficients A and B are determined. The obtained statistics (point estimates) of A and B are summarised in Table 2.

It can be seen that the results from these studies are quite consistent. Most studies show that A has a mean of about 30 μm with a high COV (coefficient of variation) around 0.3, and the coefficient B has a mean around 0.55 with a COV varying between 0.1 to 0.3. (Note that there are only two samples in [37]. The data can be used to estimate the mean values, but is insufficient to estimate the variance.) If the data from the five sources are lumped together, A has a mean of 29.1 μm and a COV of 0.31, and B has a mean of 0.54 with a COV of 0.21. Many researchers assumed that A and B follow normal distributions, e.g., [39]. This assumption is adopted in the present study.

Four types of copula were examined to represent the dependence between A and B: Frank, Clayton, Gumbel and Gaussian copulas. The copula parameters were estimated using the maximum likelihood method. Among the four candidate copulas, the Frank copula with a parameter $\theta = -1.85$ yields the smallest value of AIC, thus it provides the best-fit to the dependence structure of A and B.

Table 2: Statistics for corrosion coefficients A and B (rural-urban environment).

References	A		B		No. samples
	Mean (μm)	COV	Mean	COV	
[34]	24.5	0.22	0.501	0.23	12
[35]	28.36	0.29	0.571	0.19	14
[36]	30.74	0.32	0.583	0.07	19
[37]	23.5	–	0.516	–	2
[38]	32	0.31	0.502	0.30	15
Combined	29.1	0.31	0.54	0.21	62

In addition to the randomness in A and B, we next consider the additional uncertainty (epistemic uncertainty) arising from the inaccuracies in the estimation of distribution parameters due to limited data, namely, the mean values of A and B (denoted by μ_A and μ_B), and the copula parameter θ. The imperfect knowledge about μ_A, μ_B and θ can be modelled by interval bounds constructed from confidence intervals. The 95% confidence intervals for μ_A and μ_B are $\mu_A = [26.83, 31.37]$ and $\mu_B = [0.51, 0.57]$. Using the Bootstrap method, the 95% confidence interval for θ is found to be $\theta = [-3.35, -0.35]$.

Depending on the modeling of the distribution parameters, the present study considers 6 cases, as summarised in Table 3. Case 1 uses the point estimates for μ_A, μ_B and θ. This case represents the customary practice in which the epistemic uncertainties due to small sample size are not considered. Case 2 considers the interval estimate of θ, while μ_A and μ_B are point estimates. In Case 3, both μ_A and μ_B are modeled as intervals, while θ is a point estimate. To examine the sensitivities of each of the parameters (μ_A, μ_B and θ) on the failure probability P_f, Case 3 is further divided into Case 3a and Case 3b. In Case 3a, μ_A is an interval estimate, while μ_B and θ are point estimates. In Case 3b, μ_B is an interval estimate, while μ_A and θ are point estimates. Thus, by comparing Case 2, 3a and 3b, the impacts of μ_A, μ_B and θ on P_f can be quantified, respectively. In Case 4, all parameters, μ_A, μ_B and θ, are modeled as intervals.

Table 3: Six cases for modeling A (unit: μm) and B.

	A		B		
	μ_A	σ_A	μ_B	σ_B	copula parameter
Case 1	29.1	9.13	0.54	0.11	$\theta = -1.85$
Case 2	29.1	9.13	0.54	0.11	$\theta = [-3.35, -0.35]$
Case 3	[26.83, 31.37]	9.13	[0.51, 0.57]	0.11	$\theta = -1.85$
Case 3a	[26.83, 31.37]	9.13	0.54	0.11	$\theta = -1.85$
Case 3b	29.1	9.13	[0.51, 0.57]	0.11	$\theta = -1.85$
Case 4	[26.83, 31.37]	9.13	[0.51, 0.57]	0.11	$\theta = [-3.35, -0.35]$

5. Examples

5.1. Example 1: a steel plate

A steel plate in tension is studied. The problem is adopted from [40]. The limit state function is given by:

$$g = R(t) - S, \tag{8}$$

in which S is the applied tensile load, and $R(t)$ is the time-variant resistance of the plate. Let b and d denote the nominal width and thickness of the plate, respectively. The plate is assumed to be corroded in the rural-urban environment on two sides, thus the temporal change to the plate thickness is $d - 2c(t)$, in which $c(t)$ represents the corrosion loss after t years. The time-dependent structural resistance is given by:

$$R(t) = f_y b(d - 2c(t)), \tag{9}$$

in which f_y is the yield stress. The applied load S is assumed to be a normal distribution with a mean of 200 kN and a standard deviation of 23 kN. The yield stress f_y is a normal random variable with a mean of 300 MPa and a standard deviation of 10 MPa. The width b and the original thickness d are deterministic, and $b = 250$ mm, $d = 4$ mm. The corrosion loss c is computed according to Eq. (7).

To verify the reliability results from the interval Monte Carlo method, a double-loop Monte Carlo procedure is utilized to compute the bounds of P_f for Case 2 and Case 3 at $t = 20$ year and $t = 50$ year. In the double-loop Monte Carlo procedure, the interval parameter(s) is assumed to uniformly distribute between its lower and upper bounds. Two hundred samples of the interval parameter was generated in the outer loop using the Latin Hypercube sampling technique. With each sampled parametric value, the failure probability was then evaluated in the inner loop. Thus a sample of P_f can be obtained; its lower and upper bounds are then compared with the P_f bounds computed using the interval Monte Carlo method. Tables 4 and 5 compare the bounds of P_f in $t = 20$ year and $t = 50$ year obtained from the double-loop Monte Carlo and the interval Monte Carlo methods. It can be seen that the results from the two methods agree reasonably well. The bounds of P_f from the interval Monte Carlo method are slightly wider than those of the double-loop Monte Carlo method.

Table 4: Comparison of double-loop Monte Carlo and interval Monte Carlo methods: P_f in $t = 20$ yr, (Example 1).

	Case 2		Case 3	
	$\underline{P}_f(\%)$	$\overline{P}_f(\%)$	$\underline{P}_f(\%)$	$\overline{P}_f(\%)$
Interval MC	0.049	0.727	0.116	0.410
Double-loop MC	0.051	0.682	0.121	0.387

Table 5: Comparison of double-loop Monte Carlo and interval Monte Carlo methods: P_f in $t = 50$ yr, (Example 1).

	Case 2		Case 3	
	$\underline{P}_f(\%)$	$\overline{P}_f(\%)$	$\underline{P}_f(\%)$	$\overline{P}_f(\%)$
Interval MC	0.297	10.082	1.394	5.719
Double-loop MC	0.307	9.245	1.455	5.356

The all six cases in Table 3 are then considered. The failure probability P_f for the six cases are plotted in Figs. 1 and 2 as a function of time. The results for $t = 20$ yr and 50 yr are also summarised in Table 6.

In Table 6, P_f is a point estimate for Case 1; it is 0.208% for $t = 20$ year and 2.910% for $t = 50$ year. This point estimate of P_f does not provide

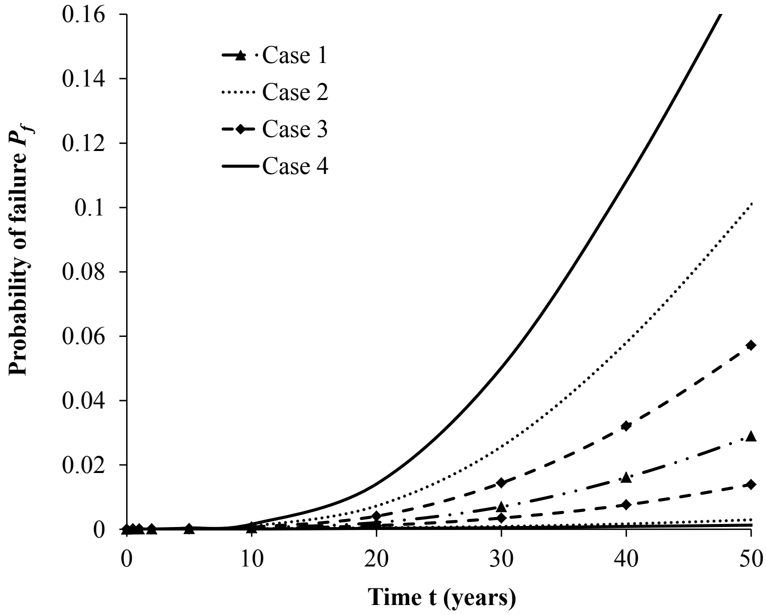

Figure 1: Bounds of the failure probabilities of Example 1 (Case 1-4).

Table 6: Probability of failure (Example 1).

	$t = 20$ years		$t = 50$ years	
	$\underline{P}_f$	$\overline{P}_f$	$\underline{P}_f$	$\overline{P}_f$
Case 1	0.208%		2.910%	
Case 2	0.049%	0.727%	0.297%	10.082%
Case 3	0.116%	0.410%	1.394%	5.719%
Case 3a	0.166%	0.283%	2.212%	3.733%
Case 3b	0.149%	0.308%	1.847%	4.449%
Case 4	0.029%	1.382%	0.131%	17.650%

information about the confidence in the result of the reliability estimate. The role of epistemic uncertainty on P_f is clearly demonstrated in the bounds of P_f shown in Figure 1 and Table 6. The width of the interval P_f shows the effect of epistemic uncertainty on the results of the reliability estimate. For example, the upper bound of P_f for $t = 50$ year is 10.082% for Case 2. This probability is an order of magnitude greater than the point estimate from Case 1. A point estimate without considering the epistemic uncertainty may significantly underestimate the true risk.

It can be seen from Figure 1 that the interval bounds of P_f become wider as time increases. Take Case 3 for example, P_f varies from $[0.116\%, 0.410\%]$ at $t = 20$ yr, and increases significantly to $[1.394\%, 5.719\%]$ at $t = 50$ yr. It is also observed that the width of P_f for Case 2 is much wider than that of Case 3. For instance, at $t = 50$ year, P_f is $[0.297\%, 10.082\%]$ for Case 2, and $[1.394\%, 5.719\%]$ for Case 3. The width of the former is more than twice of the latter. This suggests that the epistemic uncertainty in the copula modeling the dependence between A and B has a more significant effect on P_f than the epistemic uncertainty in the mean values of A and B. As expected, the width of P_f becomes wider when the analysis incorporates more epistemic uncertainties, i.e., the interval failure probabilities for Case 2 and Case 3 are enclosed in the P_f for Case 4.

To study the sensitivity of the parameters, μ_A, μ_B and θ, on the failure probability, we compare the results of Case 2, Case 3a and Case 3b. The time-dependent probabilities of failure for the three cases are plotted in Figure 2. The results for $t = 20$ yr and $t = 50$ yr are also presented in Table 6. It is observed from Figure 2 that the width of P_f for Case 2 is much wider than those of Case 3a and Case 3b. This suggests that the uncertainty in the dependence between A and B has a more significant effect on P_f than the epistemic uncertainties in μ_A and μ_B. Table 6 and Fig. 2 also shows that the widths of P_f for Case 3a and Case 3b are comparable, implying that the uncertainties in μ_A and μ_B have similar effects on P_f. This sensitivity study shows that to improve the confidence in the reliability estimates, additional

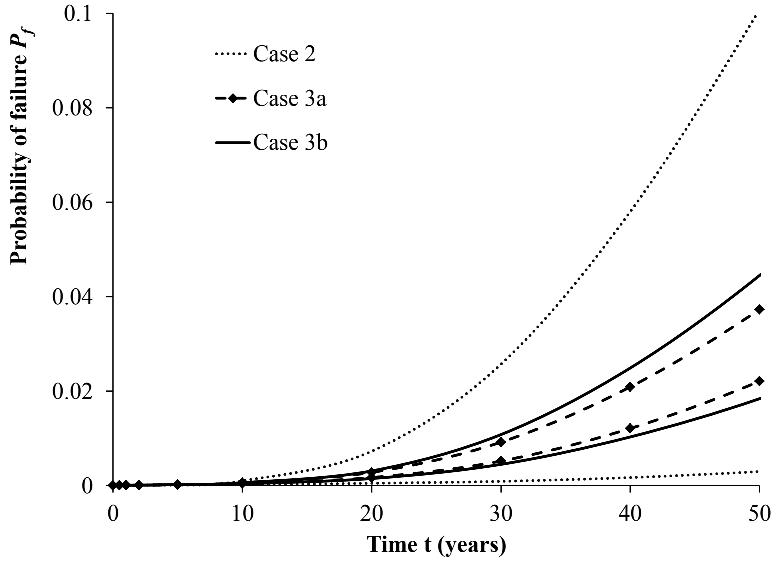

Figure 2: Bounds of the failure probabilities of Example 1 (Case 2, Case 3a and Case 3b).

data should be collected, particularly for modeling the dependence between A and B.

5.2. *Example 2: a ten-bar truss*

Figure 3 shows a ten-bar planar steel truss subjected to two concentrated loads P. The example is adopted from [41]. The truss members are circular hollow section (CHS) with three different sections, A_1, A_2 and A_3, for the horizontal, vertical and diagonal members, respectively. The nominal section sizes (outer diameter) and thickness are summarised in Table 7. The basic random variables include the load P, the thickness of CHS r_i, $i = 1, 2, 3$, and the Young's modulus E. These random variables are assumed to be mutually statistically independent normal distributions, with the statistics summarised in Table 8. The outer diameters of the sections are assumed to be deterministic and equal to their nominal values.

The limit state of interest is the stress in the diagonal member 1. The

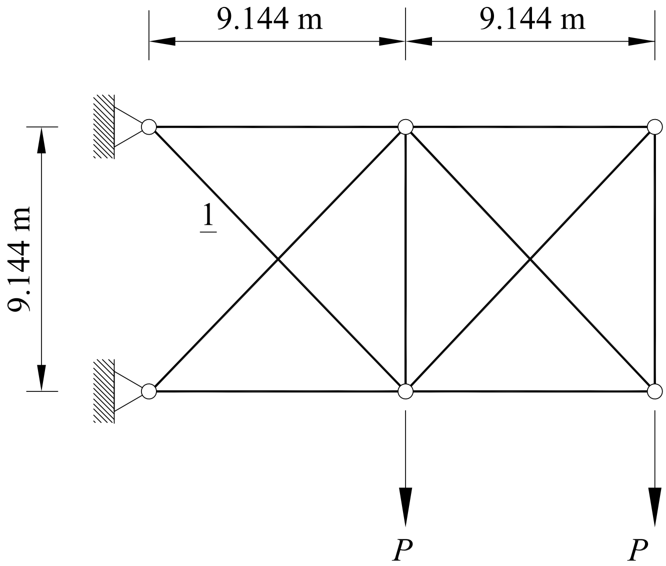

Figure 3: A ten-bar steel truss (adopted from [41]).

Table 7: Nominal section sizes for the ten-bar truss.

Section	d (outer diameter)	r (thickness)
A_1	243.8 mm	8.8 mm
A_2	193 mm	2.9 mm
A_3	243.8 mm	7.4 mm

Table 8: Random variables for the ten-bar truss.

Variable	Unit	Mean	COV (%)	Distribution
P	kN	444.8	20	Normal
r_1	mm	8.8	8	Normal
r_2	mm	2.9	8	Normal
r_3	mm	7.4	8	Normal
E	GPa	205	6	Normal

stress, σ_1, is given by a closed-form solution [41]:

$$\sigma_1 = \frac{P}{A_3}\left(2 + \frac{\sqrt{2}A_1A_2A_3\left(2\sqrt{2}A_1 + A_3\right)}{D}\right), \tag{10}$$

where

$$D = 4{A_2}^2\left(8{A_1}^2 + {A_3}^2\right) + 4\sqrt{2}A_1A_2A_3\left(3A_1 + 4A_2\right) + A_1{A_3}^2\left(A_1 + 6A_2\right). \tag{11}$$

The limit state function, $g(\cdot)$, is defined as

$$g = \sigma_a - \sigma_1 \tag{12}$$

in which the allowable stress σ_a is 250 MPa.

Considering the atmospheric deterioration of the steel and assuming two-sided corrosion loss, the cross-section areas $A_i\left(t\right)$ change with time t:

$$A_i\left(t\right) = \left(\frac{{d_i}^2\pi}{4} - \frac{\left(d_i - 2r_i\right)^2\pi}{4}\right) - \left(\frac{{d_i}^2\pi}{4} - \frac{\left(d_i - 2c\left(t\right)\right)^2\pi}{4}\right), \quad i = 1, 2, 3. \tag{13}$$

in which d_i and r_i denote the outer diameter and thickness for the CHS members, and $c(t)$ is the corrosion loss after t years.

Cases 2, 3a and 3b are first studied to examine the sensitivity of each parameter, μ_A, μ_B and θ, on the failure probability. The lower and upper bounds of the failure probabilities for the three cases are plotted in Fig. 4. Next, all six cases listed in Table 3 are studied. Figure 5 plots the lower and upper bounds of P_f as a function of time for Case 1-4. Table 9 presents the probability of failure for $t = 30$ year and $t = 50$ year. From Figs. 4 and 5, and Table 9, similar observations can be made as in Example 1. Case 4 has the widest bounds, followed by Case 2, 3 and 1. The results confirm that 1) the epistemic uncertainty has a significant impact on the reliability estimates, 2) the epistemic uncertainty in the copula parameter θ has a far

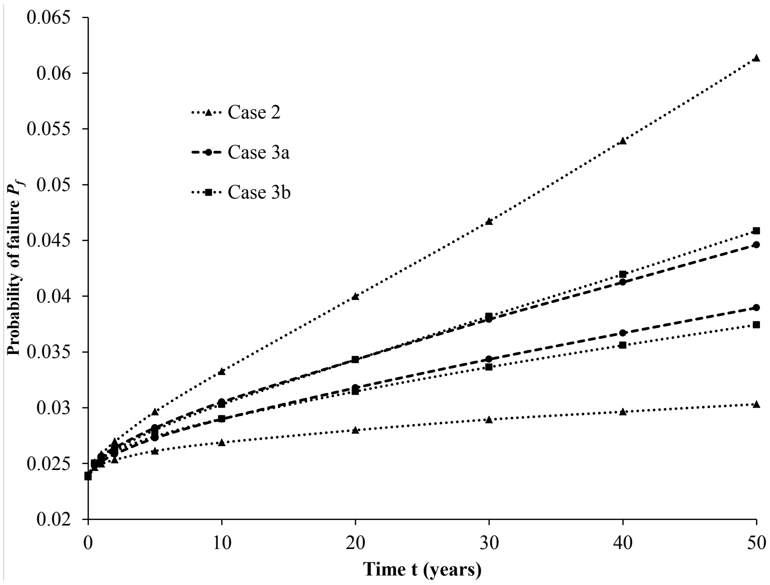

Figure 4: Bounds of the failure probabilities of Example 2 (Case 2, Case 3a and Case 3b).

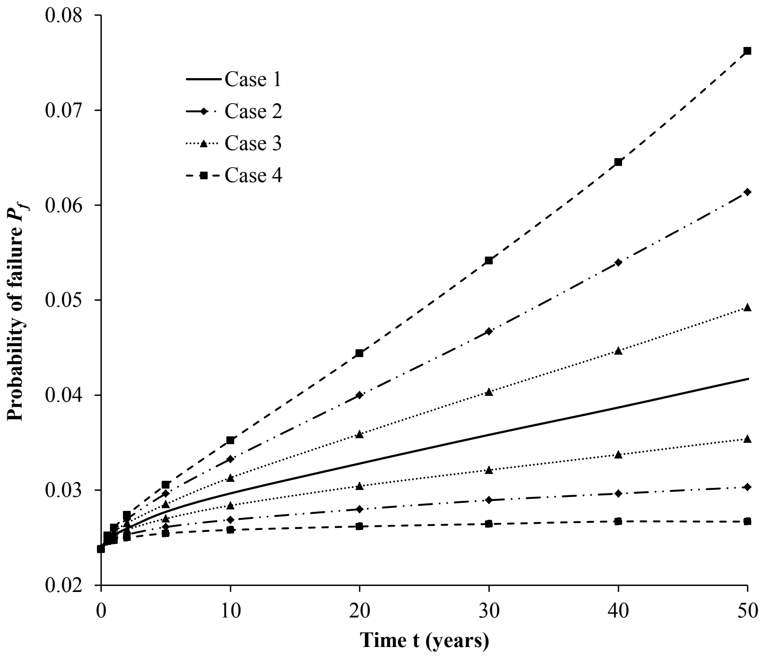

Figure 5: Bounds of the failure probabilities of Example 2 (Case 1-4).

Table 9: Probability of failure of member 1 of the 10-bar truss.

	$t = 30$ yr		$t = 50$ yr	
	$\underline{P}_f$	$\overline{P}_f$	$\underline{P}_f$	$\overline{P}_f$
Case 1	0.0359		0.0414	
Case 2	0.0288	0.0466	0.0302	0.0612
Case 3	0.0321	0.0402	0.0352	0.0491
Case 3a	0.0343	0.0378	0.0389	0.0445
Case 3b	0.0336	0.0382	0.0374	0.0456
Case 4	0.0265	0.0541	0.0268	0.0764

more significant effect on P_f than the epistemic uncertainties in the means of A and B, and 3) the epistemic uncertainties in the mean values of A of B have comparable effects on P_f.

6. Conclusions

Significant epistemic uncertainties exist in the current models for atmospheric corrosion of steel structures due to the limited availability of reliable corrosion data. Probability-box is a useful tool to model the uncertain corrosion process, accounting for both the aleatory and epistemic uncertainties. In the present study, the epistemic uncertainties are vested in the estimates of the first-order statistics (mean) of the corrosion coefficients A and B, and also the dependence structure between A and B. By examining available corrosion data, it is found that the dependence between A and B can be modeled by a Frank copula. The confidence intervals of the copula parameter are estimated using the Bootstrap method. Interval Monte Carlo method are used to compute the lower and upper bounds of probability of failure.

The probability-box analysis framework was applied to the time-dependent reliability analysis of a steel plate and a steel truss structures. In both examples, similar observations are made. The epistemic uncertainties play an important role in the reliability assessment. A point estimate of P_f without considering any epistemic uncertainty may lead to a false impression of the reliability. The interval bounds of P_f become wider as time increases. It was also found that the epistemic uncertainty in the dependence between A and B (vested in the copula parameter θ) has a far more significant effect on P_f than the epistemic uncertainty in the means of A and B. The epistemic uncertainties in the mean values of A and B have comparable effects on P_f. The importance of collecting more corrosion data, particularly for modeling the dependence of A and B, is demonstrated if the confidence in the reliability assessment is to be improved.

Acknowledgments

Support for this research was provided, in part, by the Lawrence and Betty Brown PhD Research Scholarship in Civil Engineering, the Civil Engineering Research Development Scheme from the University of Sydney, and the International Program Development Fund from the University of Sydney. These supports are gratefully acknowledged. The authors would like to acknowledge the thoughtful suggestions of two anonymous reviewers, which substantially improved the present paper.

References

[1] Q. Li, C. Wang, B. Ellingwood, Time-dependent reliability of aging structures in the presence of non-stationary loads and degradation, Struct. Saf. 52A (2015) 132–141.

[2] H. Dai, B. Zhang, W. Wang, A multiwavelet support vector regression method for efficient reliability assessment, Reliab. Engng. Syst. Saf. 136 (2015) 132–139.

[3] Q. Li, C. Wang, Updating the assessment of resistance and reliability of existing aging bridges with prior service loads, J. Struct. Engrg., ASCE 141 (12) (2015) 04015072.

[4] H. Dai, H. Zhang, W. Wang, A multiwavelet neural network-based response surface method for structural reliability analysis, Compu.-Aided Civ. Infrastruct. Engrg. 30 (2) (2015) 151–162.

[5] B. R. Ellingwood, K. Kinali, Quantifying and communicating uncertainty in seismic risk assessment, Struct. Saf. 31 (2009) 179–187.

[6] P. Walley, Statistical reasoning with imprecise probabilities, Chapman and Hall, London, 1991.

[7] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, K. Sentz, Constructing probability boxes and Dempster-Shafer structures, Tech. Rep. SAND2002-4015, Sandia National Laboratories (2003).

[8] P. Walley, Towards a unified theory of imprecise probability, Int. J. Approximate Reasoning 24 (2000) 125–148.

[9] B. Möller, M. Beer, Engineering computation under uncertainty-capabilities of non-traditional models, Comput. Struct. 86 (10) (2008) 1024–1041.

[10] C. Baudrit, D. Dubois, N. Perrot, Representing parametric probabilistic models tainted with imprecision, Fuzzy Sets Syst. 159 (2008) 1913–1928.

[11] P. Walley, T. Fine, Towards a frequentist theory of upper and lower probability, Ann. Stat. 10 (1982) 741–761.

[12] S. Ferson, R. B. Nelsen, J. Hajagos, D. J. Berleant, J. Zhang, W. T. Tucker, L. R. Ginzburg, W. L. Oberkampf, Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis, Report SAND2004-3072, Sandia National Laboratories, Albuquerque, NM (2004).

[13] D. A. Alvarez, J. E. Hurtado, An efficient method for the estimation of structural reliability intervals with random sets, dependence modeling and uncertain inputs, Comp. Struc. 142 (2014) 54–63.

[14] M. Zhang, M. Beer, S. Quek, Y. Choo, Comparison of uncertainty models in reliability analysis of offshore structures under marine corrosion, Struct. Saf. 32 (6) (2010) 425–432.

[15] G. Salvadori, C. De Michele, Frequency analysis via copulas: Theoretical aspects and applications to hydrological events, Water Resources Research 40 (12).

[16] C. Genest, A. C. Favre, Everything you always wanted to know about copula modeling but were afraid to ask, Journal of Hydrologic Engineering 12 (4) (2007) 347–368.

[17] B. Renard, M. Lang, Use of a gaussian copula for multivariate extreme value analysis: some case studies in hydrology, Advances in Water Resources 30 (4) (2007) 897–912.

[18] R. B. Nelsen, An introduction to copulas, Springer, 1999.

[19] R. H. Shumway, D. S. Stoffer, D. S. Stoffer, Time series analysis and its applications, Springer, 2011.

[20] U. Cherubini, E. Luciano, W. Vecchiato, Copula methods in finance, John Wiley & Sons, 2004.

[21] B. Efron, Computers and the theory of statistics: thinking the unthinkable, SIAM Review 21 (4) (1979) 460–480.

[22] D. A. Alvarez, On the calculation of the bounds of probability of events using infinite random sets, Int. J. Approximate Reasoning 43 (2006) 241–267.

[23] H. Zhang, R. L. Mullen, R. L. Muhanna, Interval Monte Carlo methods for structural reliability, Struct. Saf. 32 (2010) 183–190.

[24] H. Zhang, H. Dai, M. Beer, W. Wang, Structural reliability analysis on the basis of small samples: an interval Monte Carlo method, Mech. Systems Signal Process. 37 (2013) 137–151.

[25] H. Zhang, Interval importance sampling method for finite element-based structural reliability assessment under parameter uncertainties, Struct. Saf. 38 (2012) 1–10.

[26] R. E. Moore, Interval Analysis, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1966.

[27] H. Zhang, Nondeterministic linear static finite element analysis: an interval approach, Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA, USA (2005).

[28] D. De la Fuente, I. Daz, J. Simancas, B. Chico, M. Morcillo, Long-term atmospheric corrosion of mild steel, Corrosion Science 53 (2) (2011) 604–617.

[29] S. Feliu, M. Morcillo, S. Feliu Jr, The prediction of atmospheric corrosion from meteorological and pollution parameters–II. long-term forecasts, Corrosion Science 34 (3) (1993) 415–422.

[30] R. H. McCuen, P. Albrecht, Composite modeling of atmospheric corrosion penetration data, ASTM Special Technical Publication 1194 (1994) 65–65.

[31] M. Morcillo, B. Chico, I. Daz, H. Cano, D. De la Fuente, Atmospheric corrosion data of weathering steels. a review, Corrosion Science 77 (2013) 6–24.

[32] ISO-9223, Corrosion of metals and alloys-Corrosivity of atmospheres. Classification, determination and estimation, International Organization for Standardization (2012).

[33] S. Feliu, M. Morcillo, S. Feliu Jr, The prediction of atmospheric corrosion from meteorological and pollution parameters–I. annual corrosion, Corrosion Science 34 (3) (1993) 403–414.

[34] F. Haynie, J. B. Upham, Effects of atmospheric pollutants on corrosion behavior of steels, Materials Protection and Performance 10 (1971) 18–21.

[35] L. Atteraas, S. Haagenrud, Atmospheric corrosion testing in Norway, John Wiley & Son, New York, USA, 1982, pp. 873–891.

[36] V. Kucera, S. Haagenrud, L. Atteraas, J. Gullman, S. Dean, T. Lee, Corrosion of steel and zinc in scandinavia with respect to the classification of the corrosivity of atmospheres, Degradation of Metals in Atmospheres, ASTM STP 965 (1987) 264–281.

[37] H. D. Hui, P. Strekalov, Y. N. Mikhailovskii, D. T. BIN, A. Mikhailov, The corrosion-resistance of steels, zinc, copper, aluminum, and alloys in the humid tropics of vietnam-the results of 5-year tests, Protection of Metals 30 (5) (1994) 437–443.

[38] M. Morcillo, J. Simancas, S. Feliu, Long-term atmospheric corrosion in spain: results after 13 to 16 years of exposure and comparison with worldwide data, ASTM special technical publication 1239 (1995) 195–195.

[39] M. Chakravorty, D. M. Frangopol, R. L. Mosher, J. E. Pytte, Time-dependent reliability of rock-anchored structures, Reliab. Engng. Syst. Saf. 47 (3) (1995) 231–236.

[40] R. E. Melchers, Probabilistic model for marine corrosion of steel for structural reliability assessment, J. Struct. Engrg., ASCE 129 (11) (2003) 1484–1493.

[41] S.-K. Choi, R. V. Grandhi, R. A. Canfield, Reliability-based structural design, Springer Science & Business Media, 2006.