The Effects of Correlated Input Data on Estimates of Failure Probability

R A Ainsworth and Y Lei

British Energy Generation Ltd, Gloucester, GL4 3RS, UK

Abstract: *This paper uses R6 to examine the effects of correlated input data on estimates of failure probability. For 100% positive and 100% negative correlation between fracture toughness and yield stress, the failure probability can be evaluated 'exactly' by integrating the tails of distributions up to points which can be determined. Approximate estimates of failure probability are introduced, and shown to be upper and lower bound estimates for 100% negative correlation. The 'exact' and approximate estimates of failure probability are compared with some Monte Carlo simulations for a flawed plate under membrane loading. For 100% negative correlation, the upper and lower bound estimates are close to each other when the failure probability is low (~10^{-5}).*

INTRODUCTION

R6 [1] presents methods for estimating the probability of failure, P_f, allowing for variability in inputs such as defect size, fracture toughness and yield stress, using integration and Monte Carlo methods, or approximately from margins on the failure assessment diagram (FAD) [2]. With Monte Carlo methods, correlations between input quantities may be included [3].

The approximate methods in [1, 2] are restricted to cases where input quantities are not correlated. This paper addresses extension of simplified methods to include correlation. First, some background is given on correlations in Normal distributions. This is used to address effects of correlations on R6 assessment points and associated failure probabilities. Some special cases and approximate estimates of P_f are examined.

DISTRIBUTION FUNCTIONS WITH CORRELATION

Two random variables X, Y are said to be distributed as a bivariate Normal distribution with means and variances (m_x, m_y) and (σ_x^2, σ_y^2), respectively, and correlation ρ if the joint probability that X is less than or equal to x and Y is less than or equal to y is given, following [4], by:

$$P_r\{X \le x, Y \le y\} = \frac{1}{\sigma_x \sigma_y} \int_{-\infty}^{h} \int_{-\infty}^{k} g(s,t,\rho) ds\, dt \quad (1)$$

where, using standardised normal variates, with $-1<\rho<1$,

$$h = (x-m_x)/\sigma_x \qquad k = (y-m_y)/\sigma_y \quad (2)$$

$$g(s,t,\rho) = [2\pi(1-\rho^2)^{1/2}]^{-1} \exp\left[-\frac{1}{2}\left(\frac{s^2 - 2\rho st + t^2}{1-\rho^2}\right)\right] \quad (3)$$

In [4], some properties and numerical values for g are presented and three special cases are considered. These are discussed further here.

Uncorrelated Variables

The case $\rho = 0$ corresponds to X and Y being independent and

$$\Pr\{X \le x, Y \le y\} = P(h)P(k) \quad (4)$$

where P(x) is the cumulative normal distribution function. Thus the joint probability is simply the product of the probabilities that $X \le x$ and $Y \le y$.

100% Positive Correlation

For 100% positive correlation ($\rho = 1$) values $X>m_x$ are associated with values $Y>m_y$. If X takes a value x_0, Y takes the value

$$y_0 = m_y + (\sigma_y / \sigma_x)(x_0 - m_x) \quad (5)$$

and [4]

$$\begin{aligned} \Pr\{X \le x, Y \le y\} &= P(h), h \le k \\ \Pr\{X \le x, Y \le y\} &= P(k), k \le h \end{aligned} \quad (6)$$

Thus, the joint probability reduces to evaluation of a single distribution function since Eq. 5 ensures that if $X \le x_0$ then $Y \le y_0$.

100% Negative Correlation

For 100% negative correlation ($\rho = -1$) values $X>m_x$ are associated with values $Y<m_y$. If X takes a value, x_0, Y takes the value

$$y_0 = m_y - (\sigma_y / \sigma_x)(x_0 - m_x) \quad (7)$$

and [4]

$$\Pr\{X \le x, Y \le y\} = P(h) + P(k) - 1 \qquad h + k > 0 \tag{8}$$

$$\Pr\{X \le x, Y \le y\} = 0 \qquad h + k \le 0 \tag{9}$$

Eq. 9 shows that certain combinations are not possible: e.g. both X and Y cannot be below their mean values.

EFFECT OF CORRELATIONS ON R6 ASSESSMENT POINTS

The two variables X and Y are taken as the 0.2% proof stress, σ_y, and the fracture toughness, K_{mat}, described by normal distributions with means $\overline{\sigma}_y, \overline{K}_{mat}$ and variances $\sigma^2_{\sigma_y}, \sigma^2_{K_{mat}}$. Resulting variations in the R6 parameters L_r and K_r, for primary loading only are:

$$L_r = \overline{L}_r / (\sigma_y / \overline{\sigma}_y) \tag{10}$$

$$K_r = \overline{K}_r / (K_{mat} / \overline{K}_{mat}) \tag{11}$$

where $\overline{L}_r$ and $\overline{K}_r$ are the values based on mean properties.

Uncorrelated Variables

When σ_y and K_{mat} are uncorrelated, L_r and K_r are independent. An approximate failure probability, P_f^A, in this case has been given [2] as

$$P_f^A = P(h_1) + P(k_1) \tag{12}$$

where

$$h_1 = \overline{\sigma}_y [1 - F^\sigma] / \sigma_{\sigma_y} F^\sigma \tag{13}$$

$$k_1 = \overline{K}_{mat} [1 - F^K] / \sigma_{K_{mat}} F^K \tag{14}$$

F^σ and F^K are reserve factors on toughness and yield stress, respectively, for the point $(\overline{L}_r, \overline{K}_r)$ based on mean values. These are depicted in Fig. 1.

For distributions other than normal, the terms $P(h_1)$ and $P(k_1)$ in Eq. 12 are replaced by the integrals of the tails of the yield stress and fracture toughness distributions up to $(\overline{\sigma}_y / F^\sigma)$ and $(\overline{K}_{mat} / F^K)$, respectively.

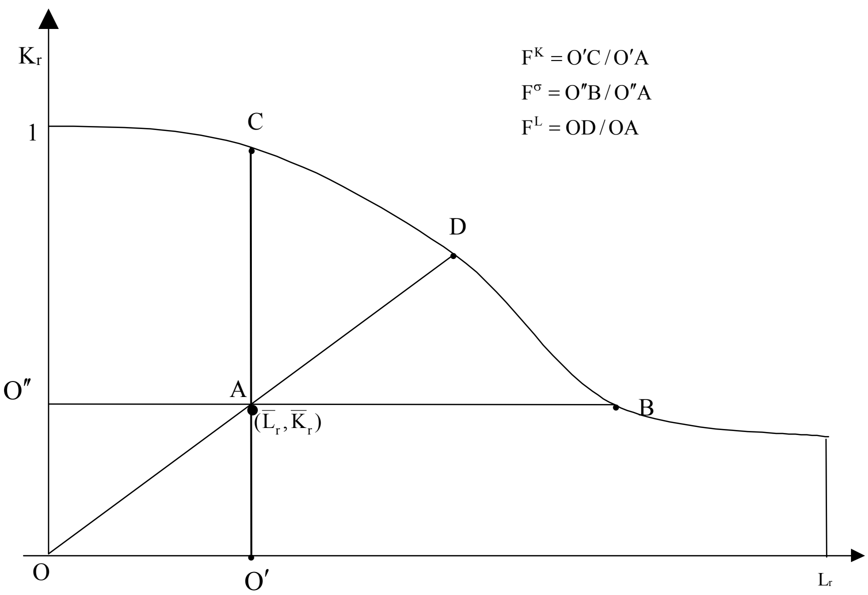

Figure 1 R6 reserve factors for simple primary loading

100% Positive Correlation

Combining Eqs 5, 10 and 11 defines the locus on the FAD:

$$L_r = \frac{\bar{L}_r}{\left\{1 + \left(\frac{\sigma_{\sigma_y}/\bar{\sigma}_y}{\sigma_{K_{mat}}/\bar{K}_{mat}}\right)\left(\frac{\bar{K}_r}{K_r} - 1\right)\right\}} \tag{15}$$

When the yield stress is fixed, the locus is the vertical line O'AC in Fig. 1 and P_f is $P(k_1)$ with k_1 given by Eq. 14. Similarly, when the toughness is fixed, P_f is $P(h_1)$ with h_1 given by Eq. 13. When the toughness and yield stress have equal coefficients of variation, Eq. 15 is the straight line OAD in Fig. 1. Then P_f is obtained from the load factor F^L in Fig. 1 as

$$P_f(\rho = 1) = P(h_2) \tag{16}$$

where

$$\begin{aligned} h_2 &= \bar{\sigma}_y[1 - F^L]/(\sigma_{\sigma_y} F^L) \\ &= \bar{K}_{mat}[1 - F^L]/(\sigma_{K_{mat}} F^L) \end{aligned} \tag{17}$$

Some results for equal coefficients of variation with 100% positive correlation have been reported in [3], using Monte Carlo methods. A range of load factors were generated by varying the crack size in a simple tension geometry. Table 1 compares the results with Eq. 16. Agreement is generally well within 1%. At low failure probabilities (~10^{-5}) the larger differences are due to the finite sampling in the Monte Carlo approach.

Table 1 P_f for 100% Positive Correlation

Defect Size a/w	$\overline{L}_r$	$\overline{K}_r$	F^L	**P_f from Monte Carlo [3]**	**P_f from Eq. 16**
0.23	0.434	0.255	2.29	$1.16\ 10^{-5}$	$1.23\ 10^{-5}$
0.34	0.506	0.366	1.83	$3.26\ 10^{-4}$	$3.32\ 10^{-4}$
0.39	0.547	0.432	1.64	$1.72\ 10^{-3}$	$1.71\ 10^{-3}$
0.43	0.585	0.493	1.49	$6.64\ 10^{-3}$	$6.69\ 10^{-3}$
0.48	0.641	0.585	1.31	$3.61\ 10^{-2}$	$3.62\ 10^{-2}$
0.53	0.709	0.694	1.15	0.165	0.165
0.58	0.794	0.825	0.99	0.517	0.517
0.62	0.877	0.943	0.88	0.841	0.840

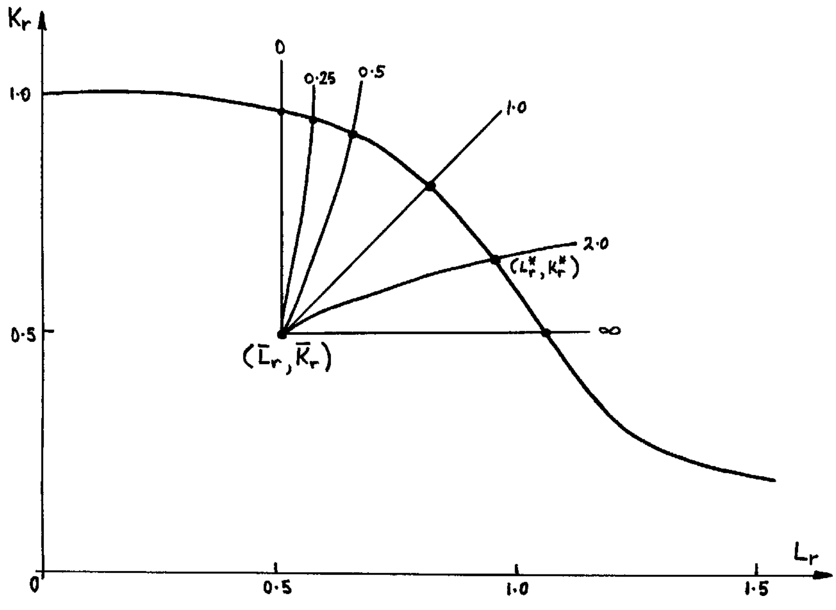

Figure 2 100% positive correlation: the curves are for the different ratios of the coefficients of variation for yield stress and toughness

When the coefficients of variation for toughness and yield stress differ, the locus of assessment points defined by Eq. 15 is a curve on the FAD, Fig. 2. If this intersects the FAD at the point (L_r^*, K_r^*), then

$$P_f(\rho=1) = P(h_3) \tag{18}$$

where

$$\begin{aligned} h_3 &= \overline{\sigma}_y(\overline{L}_r / L_r^* - 1)/\sigma_{\sigma_y} \\ &= \overline{K}_{mat}(\overline{K}_r / K_r^* - 1)/\sigma_{K_{mat}} \end{aligned} \tag{19}$$

100% Negative Correlation

Combining Eqs 7, 10 and 11 gives the locus of assessment points

$$L_r = \frac{\overline{L}_r}{\left\{1 - \left(\dfrac{\sigma_{\sigma_y}/\overline{\sigma}_y}{\sigma_{K_{mat}}/\overline{K}_{mat}}\right)\left(\dfrac{\overline{K}_r}{K_r} - 1\right)\right\}} \tag{20}$$

shown in Fig. 3 for equal coefficients of variation for yield stress and toughness.

One possibility in Fig. 3 is that the locus always lies outside the FAD. This corresponds to the result of Eq. 9 where there is zero probability that the toughness and yield stress are both sufficiently high to avoid failure.

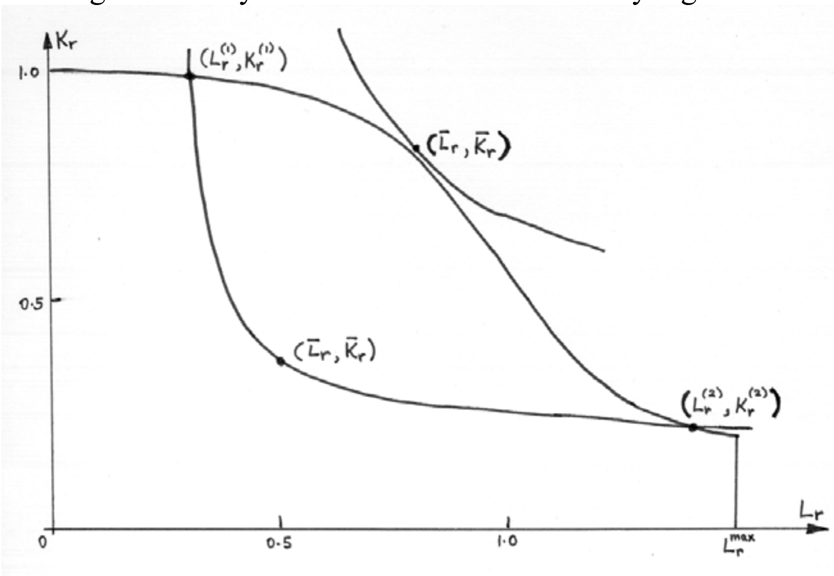

Figure 3 Loci of assessment points for 100% negative correlation

A more likely event is that the point $(\bar{L}_r, \bar{K}_r)$ lies inside the FAD. Then, the locus intersects the FAD at two points $(L_r^{(1)}, K_r^{(1)})$ and $(L_r^{(2)}, K_r^{(2)})$, shown in Fig. 3. This corresponds to the result of Eq. 8 and

$$P_f(\rho = -1) = P(h_4) + P(k_4) \tag{21}$$

where

$$h_4 = \bar{\sigma}_y(\bar{L}_r / L_r^{(2)} - 1)/\sigma_{\sigma_y} \tag{22}$$

$$k_4 = \bar{K}_{mat}(\bar{K}_r / K_r^{(1)} - 1)/\sigma_{K_{mat}} \tag{23}$$

Comparing Figs. 1 and 3, for 100% negative correlation

$$P_f(\rho = -1) < P_f^A \tag{24}$$

Therefore, while the estimates in [3] are approximations in general, for 100% negative correlation the estimates provide an upper bound to P_f. A lower bound is obtained, noting that $L_r^{(2)} \le L_r^{max}$ and $K_r^{(1)} \le 1$, as

$$P_f(\rho = -1) > P_f^B \equiv P(h_5) + P(k_5) \tag{25}$$

where

$$h_5 = \bar{\sigma}_y(\bar{L}_r / L_r^{max} - 1)/\sigma_{\sigma_y} \tag{26}$$

$$k_5 = \bar{K}_{mat}(\bar{K}_r - 1)/\sigma_{K_{mat}} \tag{27}$$

since $K_r < 1$ and $L_r < L_r^{max}$ are limits to the FAD. For the lower curve in Fig. 3 this approximation would be accurate since $K_r^{(1)} \simeq 1$ and $L_r^{(2)} \simeq L_r^{max}$.

Results from [3], using Monte Carlo methods for $\rho = -1$ are compared with the above approaches in Table 2. For loci fully outside the FAD, P_f is 1.0*. For other cases, P_f is given by Eq. 21 which gives results close to those from Monte Carlo methods. The results demonstrate the bounding properties of inequalities (24) and (25). The bounds can be close when P_f is dominated by fracture or collapse.

Table 2 Failure Probability for 100% Negative Correlation $(\rho = -1)$

Defect Size a/w	$\mathbf{K_r^{(1)}}$	$\mathbf{L_r^{(2)}}$	**P_f from Monte Carlo [3], $\rho = -1$**	**P_f from Eq. 21**	**P_f^A from Eq. 12**	**P_f^B from Eq. 25**
0.23	0.991	1	$1.03\ 10^{-5}$	$1.11\ 10^{-5}$	$1.11\ 10^{-5}$	$1.11\ 10^{-5}$
0.34	0.986	1	$1.06\ 10^{-4}$	$1.08\ 10^{-4}$	$1.08\ 10^{-4}$	$1.07\ 10^{-4}$
0.39	0.982	1	$3.58\ 10^{-4}$	$3.57\ 10^{-4}$	$3.66\ 10^{-4}$	$3.54\ 10^{-4}$
0.43	0.977	1	$1.04\ 10^{-3}$	$1.04\ 10^{-3}$	$1.14\ 10^{-3}$	$1.01\ 10^{-3}$
0.48	0.966	1	$5.04\ 10^{-3}$	$5.10\ 10^{-3}$	$7.51\ 10^{-3}$	$4.49\ 10^{-3}$
0.53	0.943	1	$3.84\ 10^{-2}$	$3.82\ 10^{-2}$	0.107	$2.53\ 10^{-2}$
0.58	-	-	1.0	1.0*	1.07	0.155
0.64	-	-	1.0	1.0*	2.0	0.809

General Case

In the general case, simple loci of the type depicted in Figs. 2 and 3 are not produced. Instead, there is a distribution of values of K_r associated with any value of L_r. P_f may be evaluated using the general solution of Eq. 3 noting that normal distributions of σ_y and K_{mat} lead to normal distributions of $(1/L_r)$ and $(1/K_r)$. P_f follows by integrating the function g of Eq. 3 over the region inside the transformation of the FAD into $1/L_r$, $1/K_r$ space. The lower bound of Eq. 25 may be evaluated noting that $K_r<1$ and $L_r < L_r^{max}$ transform to $(1/K_r) > 1$ and $(1/L_r) > 1/L_r^{max}$.

Applications for Non-Normal Distributions

The results above have been presented for Normal distributions. However, other forms such as Weibull or log-normal are likely to be used. 100% positive or 100% negative correlation corresponds to a one-to-one relationship between toughness and yield stress. This leads to a one-to-one relationship between L_r and K_r. Therefore, loci of the type shown in Figs. 2 and 3 may readily be constructed and the intersection(s) with the FAD obtained. For 100% positive correlation, there are values σ_y^*, K_{mat}^* corresponding to the point (L_r^*, K_r^*) in Fig. 2. Then P_f is

$$P_f(\rho = 1) = \int_0^{\sigma_y^*} p_\sigma(\sigma_y) d\sigma_y = \Phi_\sigma(\sigma_y^*) \qquad (28)$$

where $p_\sigma(\sigma_y)$ is the probability density function of yield stress and Φ_σ is the associated cumulative function. Eq. 28 is the generalisation of Eq. 18 to non-normal distributions. For 100% negative correlation, Eq. 21 becomes

$$P_f(\rho=-1) = \Phi_\sigma(\sigma_y^{(2)}) + \Phi_K(K_{mat}^{(1)}) \quad (29)$$

where Φ_K is the cumulative function of toughness and $\sigma_y^{(2)}, K_{mat}^{(1)}$ are the values at points $L_r^{(2)}$ and $K_r^{(1)}$ in Fig. 3.

The approximate estimates of Eqs 12 and 25 become

$$P_f^A = \Phi_\sigma(\overline{\sigma}_y / F^\sigma) + \Phi_K(\overline{K}_{mat} / F^K) \quad (30)$$

$$P_f^B = \Phi_\sigma(\overline{L}_r \overline{\sigma}_y / L_r^{max}) + \Phi_K(\overline{K}_r \overline{K}_{mat}) \quad (31)$$

As 100% negative correlation leads to values of $\sigma_y > \overline{\sigma}_y$ associated with $K_{mat} < \overline{K}_{mat}$, and vice versa, the loci of assessment points have the same shape as depicted in Fig. 3 for non-normal distributions. Therefore, again

$$P_f^A \le P_f(\rho=-1) \le P_f^B \quad (32)$$

CLOSING REMARKS

This paper has indicated how correlations between yield stress and toughness can be treated using the R6 FAD. Exact methods have been given for 100% negative and 100% positive correlation. Upper and lower bound estimates have been given for 100% negative correlation. The upper bound is identical to an approximation currently in R6 for uncorrelated data.

REFERENCES

[1] Assessment of the Integrity of Structures Containing Defects, British Energy Generation Procedure R6 Revision 4 (2001).

[2] Wilson, R., Budden, P.J. and Mitchell, B.J., An investigation into the relationship between R6 reserve factor and conditional failure probability, Magnox Electric Report TE/GEN/REP/0050/96 (1996).

[3] Wilson,R., A preliminary study of the effect of correlation between fracture toughness and tensile properties on failure probability, Nuclear Electric Memorandum TEM/MEM/0004/95 (1995).

[4] Abramowitz, M. and Stegun, I.A. (Eds), *Handbook of Mathematical Functions*, Dover, New York (1972).