Imprecise Probabilities of Engineering System Failure from Random and Fuzzy Set Reliability Analysis

Jim Hall
Department of Civil Engineering,
University of Bristol, UK
jim.hall@bristol.ac.uk

Jonathan Lawry
Department of Engineering Mathematics,
University of Bristol, UK
j.lawry@bristol.ac.uk

Abstract

Reliability analysis of engineering systems conventionally represents the system state variables as precise probability distributions and generates precise estimates of the probability of system failure. It is demonstrated how this conventional approach can be extended to handle imprecise knowledge about the system state variables, represented in general as random sets, in order to generate bounds on the probability of failure. The conventional assumption of a precise limit state function is then relaxed. A new method based on linguistic covering of the state variable space with fuzzy set labels is introduced and is used to generate an imprecise limit state function from very scarce experimental data.

Keywords. Reliability analysis, imprecise failure probabilities, random sets, linguistic labels.

1 Introduction

Engineering reliability analysis is conventionally based on the use of probabilistic information about the loads and responses of an engineering system to estimate the system probability of failure. Whilst use of reliability methods is now widespread, they have been criticised on several grounds [2,3,12]. Amongst the most significant are the constraints that the information input into the analysis has to be in a precise probabilistic format and the model (known as the limit state function) through which this information is extended is a precise model. The former constraint has been addressed quite widely by reformulating reliability calculations to accept information in a range of formats, including probability intervals [7], fuzzy sets [2,4,6,20] convex modelling [12] and random sets [22,23].

The latter problem of the precise form of the limit state function is more profound than the format of the parameters themselves. Conventionally, uncertainty in the limit state function has been addressed by adding another random variable to the state variable set to represent uncertainty. However, the empirical meaning of this variable is far from clear and its precise probabilistic format can be hard to justify.

It is interesting to note that whilst reliability theory first emerged for analysis of structural and mechanical systems it has been most challenged in rather more complex engineering domains. This helps to explain why it is in fields such as geotechnical/rock engineering [8,23] and hydraulic engineering [13,15] that reworkings of conventional reliability theory have been proposed to address some of the difficulties outlined above. This paper uses examples of reliability analysis of flood defence dike to demonstrate how imprecise parameters and limit state functions can be used in reliability calculations.

2 Formulation of the reliability problem

Reliability analysis calculates p_f, the probability of failure of a system characterised by a vector $\mathbf{x} = (x_1,\dots x_n)$ of basic variables on $X = R^n$. The resistance r of the system can be expressed as $r = g_r(\mathbf{x})$ and the loading effect s as $s = g_s(\mathbf{x})$. The probability of failure p_f is the probability p of $(r \le s)$:

$$p_f = p(r \le s) \qquad (1)$$

or in general

$$p_f = p(g(\mathbf{x}) \le 0) \qquad (2)$$

where g is termed the 'limit state function' and the probability of failure is identical to the probability of limit state violation. The resistance r and load effect s are generally implicit in $\mathbf{x}$.

If $f_X(\mathbf{x})$ is the joint probability density function over the basic variables, then

$$p_f = p(g(\mathbf{x}) \le 0) = \int_{g(\mathbf{x})\le 0} f_X(\mathbf{x})d\mathbf{x}\,. \tag{3}$$

If the n basic variables are independent then

$$f_X(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)\,. \tag{4}$$

3 Reliability calculations with random set variables

Whilst it is conventional to represent the basic variables $\mathbf{x}$ as precise random variables, in practice the information about these variables may be quite imprecise, suggesting that the probability distribution on X should be extended to the power set $\mathcal{P}(X)$. The probability mass distribution across the power set can then in general be handled through random set theory, providing a coherent structure for handling both probabilistic and possibilistic parameters [10,11]. The reliability problem then becomes that of finding the bounds on

$$p_f = p\left(g(\mathbf{x}) \le 0\right) \tag{5}$$

subject to the available knowledge restricting the allowed values of $\mathbf{x}$. The dependency between $(x_1,\ldots x_n)$ can be expressed as a random relation R, which is a random set $(\mathfrak{R}, \rho)$ on the Cartesian product $X_1\times\ldots\times X_n$, in which case the range of g is the random set $(\mathcal{F}, m)$ such that [11]:

$$\mathcal{F} = \{g(R_i) \mid R_i \in \mathfrak{R}\}, \quad g(R_i) = \{g(\mathbf{x}) \mid \mathbf{x} \in R_i\} \tag{6}$$

$$m(A) = \sum_{R_i : A = g(R_i)} \rho(R_i)$$

If the set of failed states is labelled $F \subseteq X$, the upper and lower bounds on the probability of failure are then the Plausibility $Pl(F)$ and Belief $Bel(F)$ respectively:

$$Bel(F) \le p_f \le Pl(F) \tag{7}$$

where

$$Bel(F) = \sum_{A_i : A_i \subseteq F} m(A_i) \tag{8}$$

$$Pl(F) = \sum_{A_i : A_i \cap F \neq \varnothing} m(A_i) \tag{9}$$

Equations (6) to (9) form the basis for evaluation of the bounds on system reliability with random set variables.

3.1 Special cases

A number of interesting special cases of Equation (6) are addressed below, by considerations of the situation when $\mathbf{x} = (x_1, x_2)$, so g is a mapping $X_1\times X_2 \rightarrow Z$. The extension to several variables in $\mathbf{x}$ is straightforward.

(i) *Set-valued variables*
When $(\mathfrak{R}, \rho)$ is such that $\mathfrak{R} = \{A\times B\}$, then (6) gives

$$g(A,B) = \{g(x_1,x_2) \mid x_1 \in A, x_2 \in B\} \tag{10}$$

This is a fundamental definition of interval analysis [19]. A particular case is when the variables $\mathbf{x}$, are a combination of l intervals,

$$(x_1,\ldots, x_l) = \mathbf{u} \in [\mathbf{a}, \mathbf{b}] \tag{11}$$

and n - l probability distributions,

$$(x_{l+1},\ldots, x_n) = \mathbf{v}. \tag{12}$$

The probability of failure can be evaluated at each vertex of the space of interval variables $x_1,\ldots,x_l$:

$$p_f \in \{ \int_{g(\mathbf{u},\mathbf{v})\le 0} f_V(\mathbf{v})\, d\mathbf{v} \mid \mathbf{u} \in [\mathbf{a}, \mathbf{b}]\} \tag{13}$$

and the bounds on the probability of failure derived from the vertex method [9]. The vertex method is numerically straightforward alongside the integration of probability distributions, and can in practice be used to reduce the dimensionality of the convolution integral (Equation (3)) to manageable proportions.

(ii) *Consonant random Cartesian products*
Consonant random Cartesian products correspond to decomposable fuzzy relations, in which case $\mathfrak{R}$ is the set of the level-cuts of the equivalent fuzzy relation R. In this case R is consonant and μ_R is given by

$$\mu_R(x_1, x_2) = \sum_{(x_1,x_2)\in R_i} m(R_i) \tag{14}$$

These level cuts are Cartesian products if and only if $\exists$ F_1, F_2 fuzzy sets on X_1 and X_2 such that

$$\mu_R(x_1, x_2) = \min\left(\mu_{F_1}(x_1), \mu_{F_2}(x_2)\right) \tag{15}$$

R is then a fuzzy Cartesian product denoted $F_1\times F_2$.

When $(\mathfrak{R}, \rho)$ is a consonant random relation, then $(\mathcal{F}, m)$ obtained through Equation (6) is consonant and equivalent to the fuzzy set $g(R)$ defined by:

$$\mu_{g(R)}(z) = \begin{cases} \sup\left(\mu_R(x_1, x_2) : z = g(x_1, x_2)\right) \\ 0 \text{ if } g^{-1}(z) = \varnothing \end{cases} \tag{16}$$

For consonant random Cartesian products, putting Equation (15) into Equation (16) gives:

$$\mu_{g(R)}(z) = \mu_{g(F_1,F_2)}(z) \tag{17}$$

$$= \begin{cases} \sup\left(\min\left(\mu_{F_1}(x_1), \mu_{F_2}(x_2)\right) : z = g(x_1, x_2)\right) \\ 0 \text{ if } g^{-1}(z) = \varnothing \end{cases}$$

This is Zadeh's extension principle [25], *i.e.* the fundamental equation of fuzzy arithmetic with non-interactive variables.

(iii) *Stochastically decomposable Cartesian products*
$(\mathfrak{R}, \rho)$ is a stochastically decomposable random Cartesian product on $X_1 \times X_2$ when

$$\forall R \in \mathfrak{R}, \exists\, A_1 \subseteq X_1, A_2 \subseteq X_2, R \subseteq X_1 \times X_2 \tag{18}$$

and

$$\forall A_1, A_2, \rho_{12}(A_1 \times A_2) = m_1(A_1).\, m_2(A_2). \tag{19}$$

A stochastically decomposable random Cartesian product can be specified by means of two stochastically independent random sets S_1 and S_2, with the joint basic assignment ρ_{12} being given by (19). In this case the general extension principle in (6) becomes

$$m(A) = \sum_{A_1, A_2 : g(A_1, A_2) = A} \rho_{12}(A_1, A_2)\,. \tag{20}$$

(iv) *Joint probability distributions*
Another particular case of (6) is when $\forall R \in \mathfrak{R}, \exists\, x_1, x_2, R = \{(x_1, x_2)\}$, in which case ρ defines the joint probability assignment on $X_1 \times X_2$ and $(\mathcal{F}, m)$ as defined by (6) gives a probability assignment on X such that

$$p(z) = m(\{z\}) = \sum_{x_1, x_2 : g(x_1, x_2) = z} \rho(x_1, x_2) \tag{21}$$

and in this case

$$p_f = p(g(x_1, x_2) \le 0) = \sum_{x_1, x_2 : g(x_1, x_2) \le 0} \rho(x_1, x_2) \tag{22}$$

which is equivalent to Equation (3).

When $(\mathfrak{R}, \rho)$ is a joint probability distribution which is stochastically decomposable, *i.e.*

$$p(x_1, x_2) = p_1(x_1).p_2(x_2), \tag{23}$$

p_1 and p_2 are probability assignments of stochastically independent variables and

$$p(z) = \sum_{x_1, x_2 : g(x_1, x_2) = z} p_1(x_1).p_2(x_2) \tag{24}$$

so

$$p_f = p(g(x_1, x_2) \le 0) = \sum_{x_1, x_2 : g(x_1, x_2) \le 0} m(x_1).m(x_2) \tag{25}$$

which is equivalent to Equation (4).

3.2 Consonant approximations to random sets

In general, whilst the parameters of **x** will often each be one of the special cases introduced above, z $(=g(\mathbf{x}))$ will be a non-consonant random set on Z. However, it is possible to benefit from the computational efficiency of Equation (17) by identifying consonant approximations to non-consonant marginal variables in **x**. The approach is particularly attractive during reliability-based systems optimisation [22], when multiple calls are made to the reliability function.

Proposition 1 (see [11]). *If $(\mathcal{F}_1, m_1)$ and $(\mathcal{F}_2, m_2)$ are the images through function g: $X_1 \times X_2 \to Z$ of random relations $(\mathfrak{R}_1, \rho_1)$ and $(\mathfrak{R}_2, \rho_2)$ on $X_1 \times X_2$, respectively, then $(\mathcal{F}_1, m_1) \subseteq (\mathcal{F}_2, m_2)$ as soon as $(\mathfrak{R}_1, \rho_1) \subseteq (\mathfrak{R}_2, \rho_2)$.*

Consequently

$$[Bel_1(A), Pl_1(A)] \subseteq [Bel_2(A), Pl_2(A)], \tag{26}$$

which can be used as an efficient mechanism for bracketing the probability of failure.

$(\mathcal{F}_1, m_1) \subseteq (\mathcal{F}_2, m_2)$ in this case signifies *strong inclusion* [10], for which the following three conditions must hold

(i) $\forall A \in \mathcal{F}_1, \exists\, B \in \mathcal{F}_2, A \subseteq B$,
(ii) $\forall B \in \mathcal{F}_2, \exists\, A \in \mathcal{F}_1, A \subseteq B$,
(iii) there is a non-negative assignment matrix W with entries $W(A,B)$, $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$,

$$\forall A \in \mathcal{F}_1,\ m_1(A) = \sum_{B : A \subseteq B} W(A, B)$$

$$\forall B \in \mathcal{F}_2,\ m_2(B) = \sum_{A : A \subseteq B} W(A, B)$$

where $W(A,B) = 0$ as soon as $A \not\subset B$.

$(\mathcal{F}_1, m_1)$ is *weakly included* [10] in $(\mathcal{F}_2, m_2)$ if $\forall A \in X$ $Bel_1(A) \ge Bel_2(A)$.

Note, however, that it is not sufficient to demonstrate inclusion of the marginal variables for Proposition 1 to apply, since the proposition is expressed in terms of the *relations* $(\mathfrak{R}_1, \rho_1) \subseteq (\mathfrak{R}_2, \rho_2)$. Thus, for example [10], if F_1 and F_2 are fuzzy sets on X_1 and X_2 respectively, such that they are outer approximations to random sets U_1 and U_2, respectively, then generally the fuzzy Cartesian product $F_1 \times F_2$ defined by Equation (15) is *not* an outer approximation of the stochastically decomposable random relation $U_1 \times U_2$ defined by Equation (19). If U_{F1} and U_{F2} are consonant random sets equivalent to F_1 and F_2, then the random set $U_{F1} \times U_{F2}$ is a strong outer approximation to $U_1 \times U_2$, but is no longer consonant. The minimal outer approximation R of $U_{F1} \times U_{F2}$ whose focal sets are in the set R of level cuts of $F_1 \times F_2$ has the membership function:

$$\mu = \min[\mu_{F_1}.(2 - \mu_{F_1}), \mu_{F_2}.(2 - \mu_{F_2})] \tag{27}$$

The best lower bound approximation (in the sense of weak inclusion) of $U_{F1}\times U_{F2}$ is the fuzzy set $F_1.F_2$ defined by $\mu_{F_1.F_2} = \mu_{F_1}.\mu_{F_2}$ [10].

Besides generalising reliability calculations to handle a range of set-valued variables, the above approach can also be used to bracket Monte Carlo simulation results [11]. The approach has been demonstrated in reliability analysis of rock mass response [23] and can drastically reduce the computational burden of reliability calculations, as well as providing an explicit evaluation of the error involved in the calculation.

3.3 Numerical example of reliability analysis with random set variables

Systems of dikes provide protection against flooding in many countries. Whilst in the past design and safety assessment of flood defence systems has been based on deterministic methods and factors of safety, reliability methods are being increasingly widely adopted as the basis for risk-based decision-making. Reliability methods are particularly attractive in the context of flood defence because the main hydraulic loads (rainfall, flood water levels, wave heights) are well described as random processes. However, there is much less statistical data about other determinants of system reliability, for example dike soil strengths.

The example presented here is based on a previous conventional reliability analysis of a dike on the Frisian coast in the Netherlands, along the Wadden Sea [14]. The behaviour of the concrete block revetment on the seaward slope of the dike is described by basic variables $\mathbf{x} = (\Delta, D, H_s, \alpha, M, s_{op})$ where

- Δ is the density of the revetment blocks,
- D is the diameter of the revetment blocks,
- H_s is the significant wave height,
- α is the slope of the revetment,
- M is a model parameter and
- s_{op} is the offshore peak wave steepness.

The limit state function $g(\mathbf{x})$ is given by

$$g(\mathbf{x}) = \Delta D - H_s \frac{\xi_{op}}{M \cos\alpha} \tag{28}$$

where $\xi_{op} = s_{op}^{-0.5}\tan\alpha$. The wave height H_s in shallow water is related to the water depth. Given a particular water level h, the wave height is normally distributed with

$$\mu_{H_s|h} = 0.224h + 0.117, \quad \sigma_{H_s|h} = 0.04h - 0.05 \tag{29}$$

where h is Gumbel distributed with parameters $\alpha = 0.36$, $\xi = 2.91$ and there are 3 storm events each year. The wave height therefore also conforms reasonably closely to a Gumbel distribution with $\alpha = 0.12$ and $\xi = 0.94$. The other parameters in the original analysis were assumed to be normally distributed and independent (Table 1). The situation therefore corresponds to the special case of stochastically independent state variables (Section 3.1(iv)). The probability of failure of 9×10^{-4} per year is calculated according to Equations (3) and (4), which are conveniently solved using first order second moment (FOSM) or Monte Carlo methods [18].

Variable	μ	σ	Lower bound	Upper bound
s_{op}	0.036	0.004	-	-
$\tan\alpha$	0.33	0.01	0.32	0.34
M	4.06	0.698	3.0	5.2
Δ	1.62	0.02	1.60	1.65
D	0.70	0.02	0.68	0.72

Table 1: Means and standard deviations of normally distributed parameters [16]. Interval bounds for imprecise parameters

3.3.1 Interval measurements of imprecisely known parameters and probability distribution of random loading parameters

Whilst the hydraulic loading parameters H_s and s_{op} are derived from statistical analysis of the variable loads at the site, the other parameters, which relate to the strength of the system, are based on imprecise measurements of parameters that notionally have an exact value at any given cross-section of dike. To distinguish between these two types of uncertainty H_s and s_{op} can be described by their probability distributions, whilst $\tan\alpha$, M, Δ and D are allocated the interval values given in Table 1. The bounds on M have been obtained from the bounds on experimental measurements, whilst the bounds on $\tan\alpha$, Δ and D can be obtained from knowledge of construction tolerances or (imprecise) measurements. The situation therefore corresponds to the case described in Section 3.1(i) and the bounds on the probability of failure can be calculated according to Equation (13) to generate an imprecise value of the probability of failure $p_f \in [1\times10^{-7}, 2\times10^{-3}]$. The width of this bound is dominated by the bounds on M.

3.3.2 Interval measurement and fuzzy sets of imprecisely known parameters and probability distribution of random loading parameters

The parameter M was subsequently modelled as the fuzzy set illustrated in Figure 1, while the hydraulic loading parameters H_s and s_{op} were represented as probability distributions and the imprecisely known parameters relevant to the dike, $\tan\alpha$, Δ and D, were represented as intervals, as previously. This combination induces a non-consonant random set on the parameter space. The bounds of the probability of failure can then calculated using Equation (13) by considering M as a

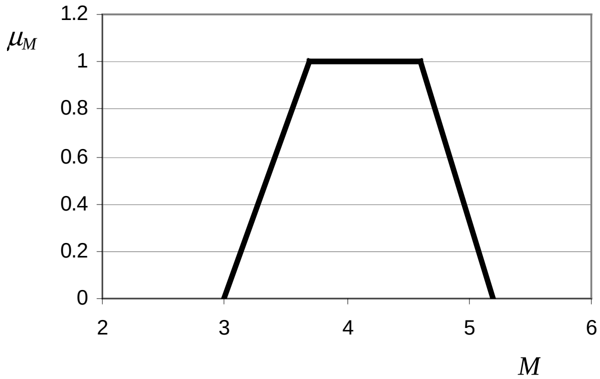

Figure 1: Fuzzy set for M

weighted set of α-cuts, generating a probability of be failure of $p_f \in [5\times10^{-7}, 6\times10^{-4}]$. The rather narrower bounds on the probability of failure reflect the increased information content in the fuzzy set relative to the interval bounds. The result no longer bounds the precise probability of failure generated by the conventional reliability method, because of differing assumptions about the probability/possibility of the lower tail of M. These assumptions are scrutinised and improved upon in Section 4.2.

4 Model uncertainty

In the preceding section it has been demonstrated how conventional reliability analysis with random variables can be generalised to random set-valued variables in order to generate bounds on the probability of system failure that reflect the uncertainty in the system variables. The analysis has assumed the existence of some limit state function $z = g(\mathbf{x})$, however, the nature of this function has not been examined.

In conventional reliability analysis the function is assumed to be a precise mapping, even though knowledge of the engineering system behaviour is often too scarce for the mapping to be substantiated by experimental evidence. Indeed in some circumstances, particularly the more challenging engineering fields mentioned in the Introduction, all that may be available is some vague relationship based on the judgement of a few experts and perhaps a handful of experimental measurements. In any case, there will always be a range of dependability of limit state functions, with some being highly dependable and others being much less dependable. Indeed most real engineering systems can fail by several different mechanisms, each of which will have a corresponding limit state function, so even within the analysis of a single system there will be a range of levels of uncertainty associated with the different limit state functions. To represent each of these functions as the same precise mapping misrepresents the engineer's variable state of knowledge about the system.

The conventional probabilistic approach to handling uncertainty in the limit state function (usually referred to as 'model uncertainty') is to introduce another random variable in the limit state equation to represent model uncertainty. In the example introduced in Section 3.2, the parameter M (Equation (28)) was modelled as a normally distributed random variable in order to represent the model uncertainty. There are, however, a number of criticisms of this approach:

(i) Parameterisation of uncertainty in the limit state function involves an implicit assumption of the form of the relationship between the basic variables (linearity, in the example introduced above). This assumption may not be justified on the basis of the scarce available knowledge.

(ii) There may not be sufficient information to identify the form of the distribution of the uncertainty parameter. Often a normal distribution is assumed without empirical justification.

(iii) If model uncertainty is represented by a precise random variable in all cases then it is not possible to represent varying states of knowledge about the limit state function.

Moreover, Blockley [3] argues that reducing model uncertainty to a single parameter is inadequate because the level of sophistication of handling such a difficult and important part of the total uncertainty is very much less than for the relatively straightforward issue of uncertainty in the system variables.

In this section, two approaches to generalising the conventional formulation of the limit state function are presented. The first deals with the imprecision in the definition of the states 'failed' and 'not failed'. A new approach is then introduced that uses a fuzzy set classification of the system behaviour to develop a limit state function on the basis of scarce experimental data. The approach is demonstrated by application to the example introduced above.

4.1 Fuzzy failure surface

It has been recognised that the distinction between 'failed' and 'not failed' states is seldom as crisp as the formulation of the limit state function suggests. This has led to the development of multi-state structure function [5]. It is natural therefore to fuzzify the boundary by adding a failure level index α, such that $\alpha = 1$ represents 'complete failure' and $\alpha = 0$ represents 'complete survival'[16]. A given failure level α therefore corresponds to a system response $g(\mathbf{x}) = z_\alpha$, so the probability of this failure level can be calculated as

$$p[\alpha] = p[g(\mathbf{x}) \le z_\alpha] = \int_{g(\mathbf{x})\le z_\alpha} f_X(\mathbf{x})d\mathbf{x} \,. \qquad (30)$$

The approach, however, has not gained general acceptance because of the difficulty in defining fuzzy failure levels. In practice crisp failure levels are defined through a process of collective judgement and negotiation in code committees. The imprecise nature of these criteria is well recognised but from the point of view of the designer who has to make a decision, the benefit of explicitly fuzzifying the condition has not been widely accepted.

4.2 Imprecise knowledge of the limit state function

A more practical problem is the situation in which the engineer has only limited information on which to base the limit state function. As was explained above, the conventional probabilistic approach is to introduce another random variable to represent the model uncertainty. An alternative approach is introduced here, which involves constructing random sets over the parameter space, which, for any given point in that space, represent the available evidence that the system has failed or has not failed. At any given point a probability mass of unity is distributed between three focal sets: {*failed*}, {*not failed*} and {*failed, not failed*}. The approach proceeds as follows.

For each variable a finite set of labels *LA* is defined that form a linguistic covering of the state space [17].

Definition 1. *Linguistic covering. A set of fuzzy sets $F_1,\ldots, F_n$ forms a linguistic covering of X if and only if $\forall x \in X$ max $(\mu_{F_1},\ldots,\mu_{F_n}) = 1$.*

For any $x \in X$, a unit mass can be distributed over the set of labels covering that point, according to the fuzzy memberships of the labels. Suppose that $\{l_1,\ldots, l_k\} = \{l_i \in LA \mid \mu_{l_i}(x) > 0\}$ and that $\{l_1,\ldots, l_k\}$ is ordered such that

$$\mu_{l_i}(x) \geq \mu_{l_{i+1}}(x) \text{ for } i = 1,\ldots, k\text{-}1. \tag{31}$$

If $L_i = \{l_1,\ldots, l_i\} \mid 1 \leq i \leq k$ then the mass distribution at x can be written as a random set ($\mathcal{F}_x$, m_x), where

$$\mathcal{F}_x = \{L_i \mid i = 1,\ldots, k\} \tag{32}$$

$$m_x(L_i) = \begin{cases} \mu_{l_i}(x) - \mu_{l_{i+1}}(x) \mid i = 1,\ldots,k-1 \\ \mu_{l_k}(x) \quad\quad\quad\quad \mid i = k \end{cases}$$

This is referred to as the label description of x. For example if the set of labels {small}, {medium} and {large} provides a linguistic covering of the space of $x \in [0,100]$ as shown Figure 2, then $\mathcal{F}_{30}$ = {*small, medium*}: 0.5, {*small*}: 0.5.

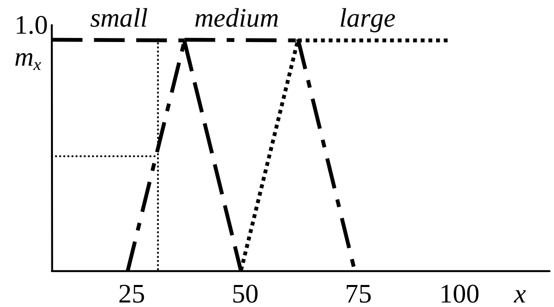

Figure 2: Linguistic covering of $x \in [0,100]$

Notice that in order for $\mathcal{F}_x$ to be a normalised random set, in the sense that zero mass is allocated to the empty set for every x, then the set of labels *LA* must form a linguistic covering as given in Definition 1. Normalised random sets are desirable in this context since otherwise mass is associated with the possibility that none of the 'words' in *LA* are appropriate as labels for some x and this makes prediction more problematic.

The idea of a label description of a point can be extended to obtain a label description of a database of measurements, so that each element in the database will be described by a random set signifying a mass value for each subset of *LA*.

Definition 2. *Joint density estimate on labels. Suppose that each of the n elements in the database has two attributes that can be classified against label sets LA_1 and LA_2 respectively, then $\forall (S_1, S_2) \subseteq LA_1 \times LA_2$ the label description of the database D*

$$D = \{(x(i), y(i)) \mid i = 1,\ldots, n\} \tag{33}$$

on $2^{LA_1} \times 2^{LA_2}$ *is defined by*

$$m_D(S_1, S_2) = \frac{1}{n}\sum_{i=1}^{n} m_{x(i)}(S_1).m_{y(i)}(S_2) \tag{34}$$

This approach can be applied to the classification of the variable space in a reliability problem based on incomplete information. Suppose knowledge about the system behaviour comprises a database of n tests of system response and in each test the response has been categorised as belonging to one of three classes, C_1 = {*failed*}, C_2 = {*not failed*}, C_3 = {*failed, not failed*} *i.e.* 'unknown'. Now consider the sub-database of instances with class C_j,

$$D_j = \{(x(i), y(i)) \mid C(i) = C_j\} \tag{35}$$

m_{Dj} can be evaluated according to Equation (34) to find the mass assignment on $LA_1 \times LA_2$ describing class C_j.

For reliability analysis, the distribution on $X_1 \times X_2$, rather than on $LA_1 \times LA_2$, is required. By Bayes theorem

$$p(x \mid L_i) = \frac{p(L_i \mid x)\, p(x)}{\int_{x \in X} p(L_i \mid x)\, p(x) dx} \tag{36}$$

Definition 3. *The Posterior density from labels is defined as*

$$p(x \mid L_i) = \frac{m_x(L_i)\, p(x)}{\int_{x \in X} m_x(L_i)\, p(x) dx} \tag{37}$$

Clearly $p(x)$ is an unknown prior distribution and we must model this uncertainty in some way. One approach would be to identify a family of distributions that must contain p(x) [25] and then obtain interval probabilities by taking the upper and lower bounds. Unfortunately in this context we have no information that would allow us to restrict this family of distributions. Taking upper and lower bounds across all distributions gives $p(I|L_i) \in [0,1]$ for any I, a measurable subset of X, for which

$$\int_I \mu_{L_i}(x) dx > 0\ . \tag{38}$$

Instead we adopt a more traditional approach and assume that $p(x)$ is the uniform distribution. This has the advantage of being the maximum entropy distribution and hence introduces minimum prior information into the model. Such a property in itself would seem to lend some justification to this choice as we are giving maximum possible weighting to the information contained in the data. Assuming a uniform prior distribution on X gives

$$\forall x \in X, \quad p(x \mid L_i) = \frac{m_x(L_i)}{\int_{x \in X} m_x(L_i) dx}\ . \tag{39}$$

This can be used to obtain a density on X conditional on the data described by m_D,

$$\forall x \in X, \quad p(x \mid m_D) = \sum_{i=1}^{k} m_D(L_i)\, p(x \mid L_i)\ , \tag{40}$$

which can be used to develop a distribution at x over the different system states. By Bayes theorem

$$p(C_j \mid x, y) = \frac{p(x, y \mid C_j)\, p(C_j)}{p(x, y)}\ , \tag{41}$$

Therefore, if $p(x, y \mid m_{Dj})$ is used as an estimate of $p(x, y \mid C_j)$ the distribution at point (x, y) over the states C_j is given by

$$m_{x,y}(C_j) = \frac{p(x, y \mid m_{D_j})\, p(C_j)}{\sum_{j=1}^{3} p(x, y \mid m_{D_j})\, p(C_j)}\ , \quad \sum_{j=1}^{3} m_{x,y}(C_j) = 1 \tag{42}$$

Each point (x, y) then has an imprecise conditional probability of failure

$$p_f(x, y) \in [m_{x,y}(C_1),\ m_{x,y}(C_1) + m_{x,y}(C_3)] \tag{43}$$

The conventional reliability problem in Equation (3) becomes

$$p_f = p(g(\mathbf{x}) \le 0) = \int_{g(\mathbf{x}) \le 0} f_X(\mathbf{x}).p_f(\mathbf{x}) d\mathbf{x}\ , \tag{44}$$

which will yield bounds on the probability of failure even if the basic variables are precise probability distributions. The expression can be further extended to include random-set valued variables as explained in Section 3 above.

4.3 Numerical example of reliability analysis including imprecise failure surface

In Section 3.2 the uncertainty in Equation (28) was addressed by using the conventional approach of representing it with a model parameter (M in this case), though the approach was then extended to represent M as a fuzzy set. In the following analysis the uncertainty parameter M is removed and the limit state function described in terms of a mass assignment to the space of system variables.

From Equation (28) we have that when $g(\mathbf{x}) = 0$

$$\frac{H_s}{\Delta D} = \frac{M \cos\alpha}{\xi_{op}} \tag{45}$$

Now $H_s \ge 0$ so, if the limit state function is described by Equation (28), in a plot of $H_s/\Delta D$ against $\cos\alpha/\xi_{op}$ the limit state function will be a line passing through the origin with a slope M. There are physical arguments constraining the limit state function to pass through the origin. There are no physical reasons for it to be linear, though the assumption of linearity is explicit in Equation (28). Data on failure of specific dike revetments is scarce and very expensive to obtain. At the site in question there were only ten relevant measurements (Figure 3), from which it is clear that assumptions of a linear relationship and normal distribution of residuals around that linear relationship are rather difficult to justify.

Each experimental point can be seen as dividing the parameter space into ‘failed’, ‘not failed’ and ‘unknown’ regions. A minimum assumption about the limit state function (which is much weaker than the linear assumption) is that it is monotonic for each of the variables groups $H_s/\Delta D$ and $\cos\alpha/\xi_{op}$. Thus an increase in $H_s/\Delta D$ for constant or reducing value of $\cos\alpha/\xi_{op}$ will make the structure more prone to failure. Similarly an increase in $\cos\alpha/\xi_{op}$ for constant or reducing $H_s/\Delta D$ will make the structure less prone to failure. Each experimental point therefore divides the space into a ‘failed’ quadrant, a ‘not failed’ quadrant and two ‘unknown’ quadrants. This is illustrated for one point in Figure 3. Even with this minimal assumption of monotonically, the data still shows some conflict.

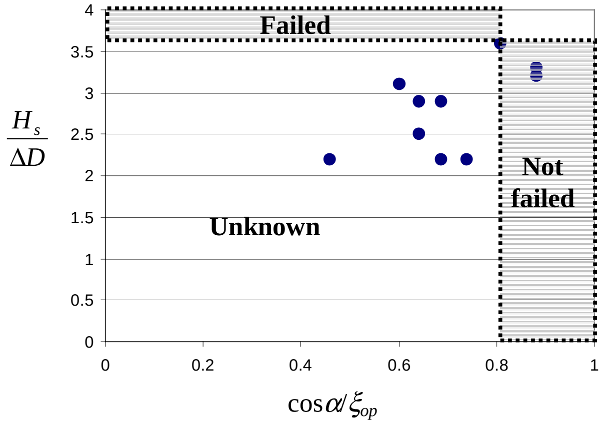

Figure 3: Experimental measurements of $H_s/\Delta D$ and $\cos\alpha/\xi_{op}$ at failure ($g(\mathbf{x}) = 0$)

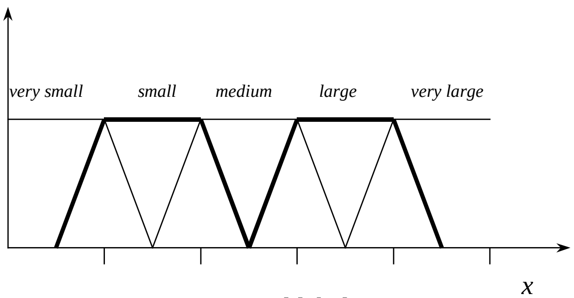

Figure 4: Marginal label set on x

To implement the method introduced in Section 4.2 the variable space was covered with regular grid of 100 points (x, y) on $[0,1]\times[0,4]$, where $x = \cos\alpha/\xi_{op}$ and $y = H_s/\Delta D$. Each point on the grid was then classified as belonging to C_1, C_2, or C_3 on the basis of each experimental point in turn. There were ten experimental points, so the database $D = \{(x(i), y(i)) \mid i = 1,\ldots, n\}$ contained 1000 points, *i.e.* $n = 1000$. The label sets LA_1 and LA_2 for each of the two groups of variables corresponded to five uniform trapezoidal fuzzy sets on the respective universes. These could be thought of as corresponding to the labels *very small, small, medium, large, very large*, which formed a linguistic covering, according to Definition 1, on each of the axes. Figure 4 illustrates the label set on x. The mass assigned to the label set was calculated according to Equation (32) for each point on the grid of 100 points. The database of 1000 points could then be used to calculate the joint mass assignment on the label set for the three classes, C_1, C_2, and C_3, according to Equations (34) and (35). Using Equations (39), (40) and (42) the posterior density $m_{x,y}(C_j)$ was calculated for $j = 1$ to 3. In Figure 5 each point in the grid has been given classification C_c where

$$m_{x,y}(C_c) = \max_j (m_{x,y}(C_j)). \tag{46}$$

Note how the extent of the 'unknown' region reflects the distribution of data points. The distribution of $m_{x,y}(C_1)$ is illustrated in Figure 6.

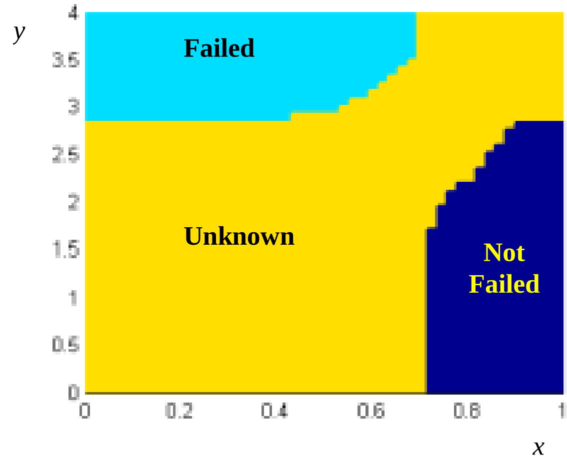

Figure 5: Imprecise limit state function

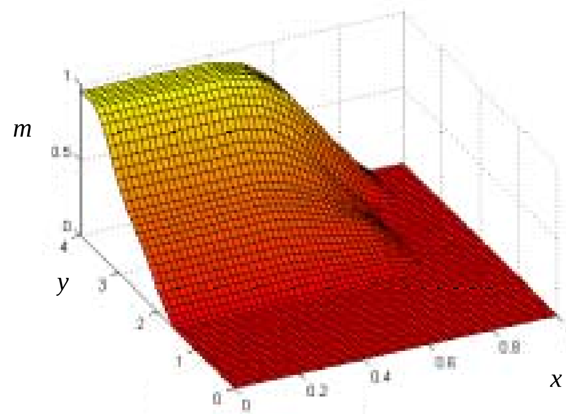

Figure 6: Distribution of $m_{x,y}(C_1)$

Further insight into the information carried on the label set can be obtained by considering alternative strategies for reallocating the points classified as 'unknown'. Two strategies were tested:

(i) The points classified as 'unknown' were removed from the database and the mass assignments $m_{x,y}(C_1)$ and $m_{x,y}(C_2)$ normalised to $m_{x,y}{}^*(C_1)$ and $m_{x,y}{}^*(C_2)$, respectively, by setting $m_{x,y}{}^*(C_1) = m_{x,y}(C_1)/(m_{x,y}(C_1)+m_{x,y}(C_2))$ and $m_{x,y}{}^*(C_2) = m_{x,y}(C_2)/(m_{x,y}(C_1)+m_{x,y}(C_2))$. At points where $m_{x,y}(C_1) = m_{x,y}(C_2) = 0$, the normalisation is undefined, so $m_{x,y}{}^*(C_1)$ and $m_{x,y}{}^*(C_2)$ were set to 0.5.

(ii) $m_{x,y}(C_3)$ was reallocated equally between $m_{x,y}(C_1)$ and $m_{x,y}(C_2)$, so $m_{x,y}{}^*(C_1) = m_{x,y}(C_1)+m_{x,y}(C_3)/2$ and $m_{x,y}{}^*(C_2) = m_{x,y}(C_2)+m_{x,y}(C_3)/2$. This corresponds to Smets' pignistic probability distribution [21] and Baldwin's least prejudiced distribution [1].

The classification between C_1 and C_2 is the same for both of the reallocation strategies and is illustrated in Figure 7. The zone marked as 'unknown' corresponds to the area where $m_{x,y}(C_1) = m_{x,y}(C_2) = 0$. Compare this with the

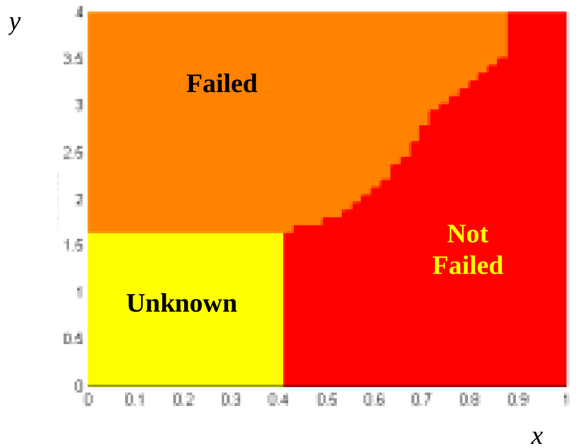

Figure 7: Precise limit state function after reallocation of uncertain probability

conventional probabilistic approach to parameterising the uncertainty with M, which corresponds to a straight classification boundary passing through the origin with gradient 4.06. The fuzzy classification mechanism has learnt a rather more subtle classification. Whilst the classification of the space with the two reallocation methods is the same, the probability distribution is significantly different. Figure 8 shows the normalised distribution and Figure 9 shows the least prejudiced distribution.

The limit state function that has been learnt from the data is combined with the distributions of the system state variables to generate probability of system failure. Each of the variables $\mathbf{x} = (H_s, \Delta, D, \tan\alpha, s_{op})$ has been modelled probabilistically, for comparison with the conventional reliability analysis, using the distributions and parameters detailed in Table 1 and Equation (29). Integrating this distribution joint distribution of the basic variables $\mathbf{x}$ over the imprecise limit state function (Equation (44)) yields a probability of system failure of $[2\times10^{-5}, 0.87]$. These rather wide bounds reflect that relative weakness of the monotonicity assumption about the limit state function when compared with the linear assumption implicit in the conventional probabilistic method (which yielded a point probability of failure of 2×10^{-4}). Nonetheless, as was argued previously, there are scant grounds to substantiate the assumptions of linearity and normality. The implications of plausibly weakening these assumptions are dramatic.

Integrating over the precise limit state functions derived from normalisation and from the least prejudiced distribution yields system probabilities of failure of 2×10^{-4} and 0.43 respectively. The normalisation assumption is analogous to the conventional probabilistic approach of insisting upon a precise limit state function,

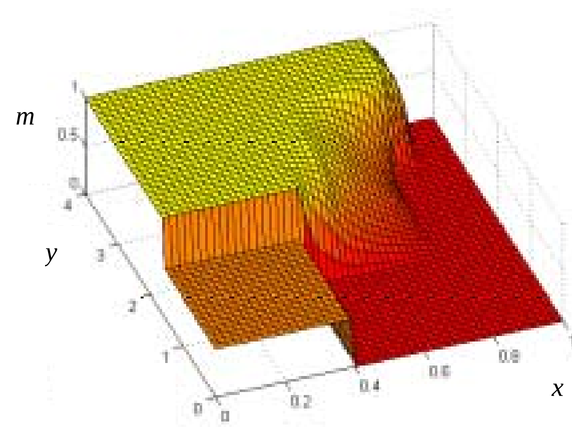

Figure 8 Normalised distribution of $m_{x,y}^*(C_1)$

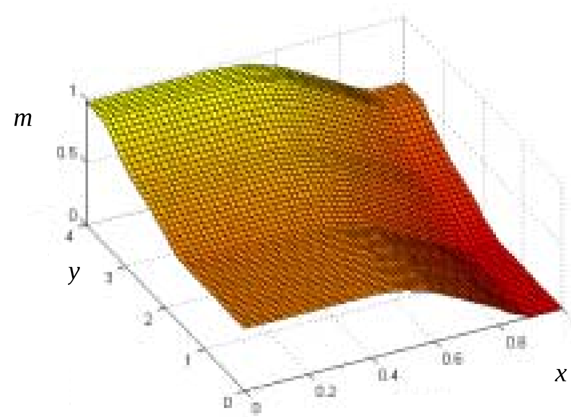

Figure 9 Least prejudiced distribution of $m_{x,y}^*(C_1)$

and yield a rather similar probability of failure. The least prejudiced assumption yields a probability of failure mid-way between the bounds obtained from the imprecise limit state function.

5 Conclusions

It has been demonstrated how plausible relaxations of some of the assumptions implicit in conventional reliability calculations result in rather wide bounds on the probability of system failure. Conventional reliability calculations insist on the system state variables being represented as a precise probability distributions. However, reliability calculations can be generalised to handle a range of random set-valued arguments, through the random set extension principle introduced by Dubois and Prade [11].

Conventionally, the system state variables are extended through a crisp limit state function. A new method for constructing an imprecise limit state function from scarce data based on minimal assumptions about the underlying system behaviour has been introduced. Application to a case study of reliability analysis of a flood defence dike has demonstrated how the form of the imprecise limit state function reflects the scarce knowledge about system behaviour. The approach has provided new insights into the sources of uncertainty and the assumptions implicit in the conventional probabilistic approach.

Acknowledgements

Data for the application were provided by Pieter van Gelder of Delft University of Technology.

References

[1] J.F. Baldwin, T.P. Martin, B.W. Pilsworth. *Fril - Fuzzy and Evidential Reasoning in AI*, Research Studies Press, Taunton, 1995.

[2] D.I. Blockley. *The Nature of Structural Design and Safety*, Ellis Horwood, Chichester, 1980.

[3] D.I. Blockley. Risk based structural reliability methods in context. *Structural Safety*, 21:335-348, 1999.

[4] C.B. Brown. A fuzzy safety measure. *ASCE J. Eng. Mech. Div.,* 105(EM5):855-872, 1980.

[5] B. Capelle. Multistate structure functions and possibility theory. In Kerre, E.E., editor, *Communications and Cognition*, 252-293, Gent, 1991.

[6] C. Cremona and Y. Gao. The possibilistic reliability theory: theoretical aspects and applications. *Structural Safety*, 19(2):173-201, 1997.

[7] W. Cui and D.I. Blockley. On the bounds for structural system reliability. *Structural Safety*, 9:247-259, 1991.

[8] G.R. Dodagoudar and G. Venkatachalam. Reliability analysis of slopes using fuzzy set theory. *Computers and Geotechnics*, 27(2):101-115, 2000.

[9] W. Dong and H.C. Shah. Vertex method for computing functions of fuzzy variables. *Fuzzy Sets & Syst.*, 24:65-78, 1987.

[10] D. Dubois and H. Prade. Consonant approximations of belief functions. *Int. J. Approx. Reasoning*, 4:419-449, 1990.

[11] Dubois, D. and H. Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets & Syst.*, 42:87-101, 1991.

[12] I. Elishakoff and Y. Ben-Haim. *Convex Models of Uncertainty in Applied Mechanics*. Elesevier, Amsterdam, 1990.

[13] J.W. Hall, J.P. Davis and D.I. Blockley. Uncertainty analysis of coastal projects. In B.L. Edge (editor) *Coastal Engineering 1998*, Vol.2:1461-1474, ASCE, New York, 1999.

[14] M. Hussaarts, J.K. Vrijling, P.H.A.M. van Gelder, H. de Loof and C. Blonk. The probabilistic optimisation of revetment on the dikes along the Frisian coast. In Losada, editor, *Coastal Structures '99*, 325-329, Balkema, Rotterdam, 2000.

[15] P.A. Johnson. Quantification of bridge scour under uncertainty. In Ayyub, B.M. and Gupta M.M., editors, *Uncertainty Modelling and Analysis: Theory and Applications*, 407-419, Elsevier, Amseterdam, 1994.

[16] K-L Lai and B.M. Ayyub. Generalized uncertainty in structural reliability assessment. *Civil Eng. Syst.*, 11:81-110, 1994.

[17] J. Lawry. Label prototypes for modelling with words. *North American Fuzzy Information Processing Society Conference,* submitted, 2001.

[18] R.E. Melchers. *Structural Reliability Analysis and Prediction*. Wiley, Chichester, 1999.

[19] R.E. Moore. *Interval Analysis*, Prentice Hall, Englewood Cliffs, 1966.

[20] D. Singer. A fuzzy set approach to fault tree and reliability analysis. *Fuzzy Sets & Syst.*, 34:145-155, 1990.

[21] P. Smets. Constructing the pignistic probability function in a context of uncertainty. In M. Henrion, R.D.. Schachter, L.N. Kanal, J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*, 29-39, North-Holland, New York, 1990.

[22] F. Tonon and A Bernardini. A random set approach to optimisation of uncertain structures. *Computers and Structures*, 68:583-600, 1998.

[23] F. Tonon, A Bernardini and A. Mammino. Reliability analysis of rock mass response by means of random set theory. *Reliability Eng. and Syst. Safety*, 70(3):263-282, 2000.

[24] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[25] L.A. Zadeh. The concept of a linguistic variable, and its application to approximate reasoning. *Inform. Sci.* 8:199-249 & 9:43-89, 1975.