# K L Consulting (UK) Limited
## Asset Integrity Management

**STATISTICAL AND PROBABILISTIC ANALYSES**

**GUIDELINES FOR USE DURING THE ASSESSMENT OF PLANT OR EQUIPMENT INTEGRITY**

**DOCUMENT KLC001**

**Document Review Cycle**

| REVIEW PERIOD | NEXT REVIEW | REVIEW BY | TECHNICAL AUTHORITY |
|---|---|---|---|
| 36 Months | December 2016 | KL | KL |

**KL Consulting (UK) Limited, Elrick. Westhill, Aberdeenshire; Tel 07545-455-355**

## DISTRIBUTION LISTING

| Copy Number | [Insert When Issued in Hardcopy] |
|---|---|
| Master (signed hardcopy) | Corporate Headquarters |
| Master (electronic) | Document Register |
| Hardcopy Distribution | |
| 001 | |
| 002 | |
| 003 | |
| 004 | |
| 005 | |
| 006 | |
| 007 | |
| 008 | |
| 009 | |
| 010 | |

## REVISION CONTROL

| Rev No. | Description of Change | Date |
|---|---|---|
| 00 | For Implementation | Dec 2010 |
| 01 | General Revision and Update | Dec 2013 |
| 02 | | |
| 03 | | |
| 04 | | |
| 05 | | |

## CONTENTS

## 1.0    INTRODUCTION

The Oil and Gas industry, as with other industries which involve the processing and transport of aggressive chemicals at high or relatively high pressure, remains under the constant threat of uncontrollable release.  The consequences of such releases are significant and can lead to a Major Accident Hazard event causing injury and death to personnel, as well as significant damage or destruction of equipment, assets and neighbouring properties.

The United Kingdom Continental Shelf is highly regulated and mandates that plant assets may only be operated following the submission and subsequent acceptance by the UK Health and Safety Executive of either a Safety Case (offshore facilities) or COMAH Safety Report (onshore facilities).  In addition, operators are required, in law to ensure that all equipment remains in a safe and fit-for-purpose condition, thus ensuring that hazards to personnel are minimised.

Central to the submission of either a Safety Case or COMAH Safety Report is a Quantitative Risk Assessment; the intent of which being to identify where the primary safety risks are and to develop effective methods to mitigate or remove as appropriate.  Removal of risks is a specific requirement if those risks are deemed to be 'unacceptable'; risks may be managed and/or mitigated if considered 'As Low As Reasonably Practicable' (ALARP), or no further management or mitigation methods are required if the risks are considered to be broadly acceptable ('Low').  QRA's are therefore ubiquitous throughout the Oil and Gas Industry, and are clearly appropriate in the context of the development of effective risk management and/or mitigation measures, nevertheless there are, in the authors opinion, considered to be significant weaknesses both in the QRA and in the Safety Management Measures which are so-based in either the Safety Case or COMAH Safety Report; these are:

1. QRA's tend to be performed by specialist third-party consultancy organisations who have the resources, expertise and experience necessary in the performance of such assessments, however as a consequence QRA's are often difficult, or more commonly than not, impossible to interrogate in many practical applications thus largely invalidating the compilation thereof[1];

2. The QRA makes use of standard and/or generic equipment failure probabilities that are unlikely to bear any resemblance to actual failure probabilities at a given installation (e.g. the generic failure probability of pipework in hydrocarbon service is typically stated as $10^{-5}$ per annum (reference)); any real resemblance would likely be purely coincidental.

3. The Safety Case or COMAH Safety Report typically state a risk of $10^{-3}$ per annum and above to be unacceptable, a risk of $10^{-6}$ and below to be broadly acceptable (requiring no other risk mitigation or management measures) and the region in-between to be the ALARP region.  In practice however, no such means or measures, nor indeed efforts are made to demonstrate that actual equipment risks are within the stated and/or required regions of acceptance.

---

[1] The precise details are not provided in the Safety Case; only summary results are presented.  It is not usually possible to determine the methodology used to calculate the (quantitative) risks, nor are the analytical tools used presented in any great detail either.

## 1.1 QRA Paradigm

Quantified Risk Assessment (QRA) is a means of identifying potential hazards to an asset, its plant, process equipment and personnel and assessing the likelihood of occurrence and the subsequent consequences.

In the UK offshore oil and gas industry, an operator is required to submit to the regulatory body (the Health and Safety Executive) a Safety Case for each of its installations in UK waters (COAMH report for onshore installations). Traditionally, a QRA is 'left on the shelf' until re-submission of the Safety Case is required. In general, the results are not widely understood or communicated; it is also the case that where the quantitative data are presented, the analyst is unable to further interrogate for use in practical applications as the narrative rarely, if ever, includes sufficient detail as to how the information was so derived. The data is collected and compiled simply to satisfy the Regulator, and is rarely if ever used for any other purpose; this is wasteful and largely invalidates the process.

The wider importance of QRA data to the effective through-life management of a large-scale asset cannot be underestimated. The optimal approach is to adopt an integrated methodology, where the QRA is prepared at the concept design stage, and subsequently refining repeatedly as more data becomes available. The 'Risk Profile' then more accurately represents the current state of the asset, so ensuring that, throughout the life of the asset, decisions are made on the basis of the best available information.

Quantified Risk Assessment (QRA) is therefore a central component of the Safety Case, as is a demonstration of the adequacy of a company's Safety Management System (SMS), established to ensure that the design and operation of the installation and its equipment are safe. The SMS sets out safety objectives, the system by which these objectives are to be achieved, performance standards to be met and the means by which adherence to these standards is to be monitored.

Zone A - Intolerable
Risk Reduction Required

$1 \times 10^{-3}$

INCREASING RISK

Zone B - ALARP Region
Risk Reduction Must Be Considered

$1 \times 10^{-6}$

Zone C - Tolerable
Risk Reduction Not Required

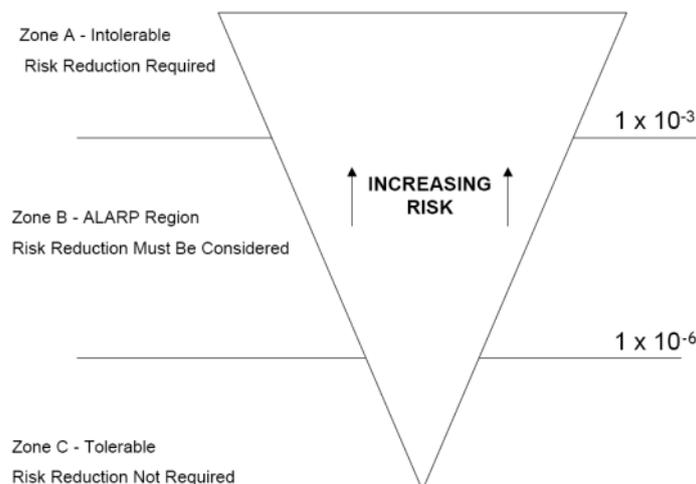**Figure 1: Quantitative Risks – Definitions of Tolerable Risks, Intolerable Risks and ALARP**

The above inverted triangle illustrating the regions of acceptability, ALARP and unacceptability are particularly informative; they provide clear boundaries within which the Safety Management System for a given installation must operate. It is clear however, that operations within certain of the above regions, and in

particular the region designated as 'unacceptable' is rarely, if ever, demonstrated. Anomalies for example are commonly assessed either in a qualitative or semi-quantitative fashion and whereas these approaches are acceptable, the primary issue is that whereas the Safety Case specifically states that a risk of $1 \times 10^{-3}$ per annum is unacceptable, there is rarely, if ever, any analyses performed to demonstrate that a given anomaly complies with this specific requirement. In extreme cases, this may result in equipment anomalies inadvertently over-step this pre-defined boundary.

These guidelines therefore, were written to assist analysts with the rather detailed and complex process of probabilistic assessment and in the use of Reliability methods for calculating quantitative failure probabilities. The analyses contained herein contain many specific examples to illustrate how the analytical methods work in practice rather than a detailed consideration of the theory which forms the foundations thereto; the literature often skips over the practicalities of employing certain analytical methods, where this document aims to present such in a detailed and methodical way. This document deliberately excludes the quantitative analyses of the consequences of a given event; this information is usually summarized in the accompanying Safety Case (or COMAH report) and due reference should be made thereto when calculating risk.

## 2.0   STATISTICAL ANALYSIS

Physical properties of materials, process operating conditions and the like will always vary; scatter in the observable data (i.e. from physical testing and/or parameter measurements) is always prevalent, however the extent will vary depending on the parameter being measured.  For example, corrosion test results often show a higher degree of scatter than many other measured parameters, largely because the observations made depend on a very wide variety of influencing factors.  Statistical analyses are extremely helpful in this respect as such facilitates the interpretation of meaningful results, especially where the data differs significantly.  Difficulties are often experienced when a variety of parameters are being investigated in order to determine the net effect of each; however statistical analysis methods provides a rational approach to such problems.

Modern technology (i.e. computer systems and software systems) readily facilitates sophisticated statistical analyses on large datasets with relative ease.   This capability alone permits the speedy investigation of innumerable statistical parameters and to determine whether associations exist between a large number of variables and if so, to develop quantitative analytical expressions describing the relationship(s).  Statistical analyses find wide application and are used extensively across any and all industries; for example, investment Fund Managers use statistical analyses to assess the various market sectors to determine whether the share prices of companies within differing sectors rise cooperatively, fall cooperatively or rise and fall cooperatively; indeed statistical assessments are used to determine the behaviour of entire market sectors relative to the market as a whole, or relative to other entire sectors; investment strategies are often developed using such analyses.

Statistical analysis methods are not difficult to comprehend or to assimilate, nor is their any difficulty with the application of such so that meaningful information may be obtained.  Innumerable tools are available to the analyst, from programmable calculators, computer spreadsheets and programmable software systems to name but a few; the preferred system however, is MathCAD, largely because of its great versatility and capability when compared to other systems that are available commercially.   The reader is free to make use of any system or tool that's available, however certain of the analyses included herein makes use of many of the in-built features which are only available in MathCAD (in particular the solve blocks 'find', 'minimize') and so the emphasis herein is reflected accordingly.  Although not critical to the statistical analyses per se, the methodology employed for calculating the relevant statistical parameters is nevertheless highly dependent on the use of these in-built MathCAD features and therefore in the absence of MathCAD the reader would be required to resort to the traditional (graphical) methods for the determination thereof.

## 3.0    STATISTICAL AND PROBABILISTIC ASSESSMENTS

Most observable phenomena contain a degree of uncertainty; that is, such phenomena can never be predicted with absolute certainty.  It is generally the case that repeated measurements of many physical phenomena generate multiple outcomes, however among such outcomes it is often the case that certain outcomes occur more frequently than others.  This behaviour is described in the context of random variables; randomness being defined as the possibility of observing more than one possible outcome and therefore the outcome itself is to some degree subject to chance.

The fundamental mathematical formulation of probability theory identifies all possible outcomes for a specific problem and defines the events in the context of these possibilities.

## 3.1    Axioms of Probability

There are three axioms of probability which govern the mathematical treatments which are discussed in the subsequent sections of this document; these are:

| Axiom | Description and Meaning |
|---|---|
| $P(E) \geq 0$ | This axiom states that the probability (P) of an event (E) will always be non-negative. |
| $P(\Omega) = 1;$ or $0 \leq P(E) \leq 1.0$ | This axiom states that the probability of the entire sample space ($\Omega$) is unity; or more simply that the probability (P) of an event (E) will lie between zero (0) and unity (1). |
| $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ | This axiom states that the probability (P) of the union of two mutually exclusive events (E$_1$ and E$_2$) is equivalent to the summation of their individual probabilities (for clarity, the mathematical term 'union' is equivalent to the logical 'OR' statement). |

## 3.2    Quantifying Randomness

The above axioms are central to all mathematical treatise concerning probabilities.  These axioms can be used in a variety of ways to determine the likelihood (or probability) that a given event will occur.  In reality, event probabilities are rarely known and therefore need to be estimated; however before the event probability can be estimated the uncertainty associated thereto must be understood and quantified.  The mathematical representation of a random variable is therefore a primary task in probabilistic formulation; this subject is discussed in more detail in this section.

Materials properties are known to be random variables, where values collected from testing and/or detailed analyses, and thus collectively constitute the sample space that describes the random variable, may then be used to extract quantitative information about the variable itself.  In order to demonstrate this, consider the following example that illustrates the variability of the Young's Modulus of structural steel beams; the analyses described herein are based on forty one

specific observations from samples taken from specimen structural steel beam sections.  The data are presented below in Table 1.

| Test No | Young's Modulus (GPa) | Test No | Young's Modulus (GPa) |
|---|---|---|---|
| 1 | 199.3 | 22 | 178.6 |
| 2 | 201.3 | 23 | 220.6 |
| 3 | 188.9 | 24 | 230.3 |
| 4 | 197.9 | 25 | 211.0 |
| 5 | 195.8 | 26 | 225.5 |
| 6 | 206.2 | 27 | 215.8 |
| 7 | 208.2 | 28 | 210.3 |
| 8 | 203.4 | 29 | 215.8 |
| 9 | 204.1 | 30 | 200.0 |
| 10 | 195.8 | 31 | 202.7 |
| 11 | 195.1 | 32 | 195.1 |
| 12 | 202.0 | 33 | 210.3 |
| 13 | 202.0 | 34 | 214.4 |
| 14 | 193.7 | 35 | 202.0 |
| 15 | 208.2 | 36 | 188.9 |
| 16 | 208.2 | 37 | 202.0 |
| 17 | 208.9 | 38 | 202.0 |
| 18 | 215.1 | 39 | 215.8 |
| 19 | 198.6 | 40 | 189.6 |
| 20 | 190.3 | 41 | 202.7 |
| 21 | 204.1 | | |

**Table 1: Young's Modulus for Structural Steel Beams**

In basis of design document the Young's Modulus ('E') for the above-mentioned structural steel beams was stated as 200 GPa, however the Table above clearly shows that this value rarely occurs; therefore it is clear in this case that the value E is a random variable and that this randomness needs to be modelled appropriately.

In the above Table, the minimum (measured) value for E is 178 GPa and the maximum 230 GPa; if either value is used in the basis of design then clearly the behaviour of a structure would not be accurately predictable since the actual value of E will in all likelihood be something different.  The stiffness of a structure would either be underestimated or overestimated; furthermore, these values may not be the absolute minimum or maximum values.  If more sample analysis had been undertaken the maximum and minimum values may likely change, therefore working on the basis of maximum or minimum values may not be desirable and usually is not accepted as good design practice.  In order to overcome this deficiency, one common-sense approach is to determine the mean or expected value of E; on the basis of the above data this is 203 GPa, however it is clear that this value alone does not provide complete information.  Information regarding the dispersion of the values with respect to the mean is needed; this is expressed in terms of the variance, standard deviation or coefficient of variation of the variable.  In addition, it is also particularly useful to know whether the dispersion is

| | Statistical & Probabilistic Analyses – Guidelines For Use During The Assessment Of Plant Or Equipment Integrity |
|---|---|

**K L Consulting (UK) Limited**
Asset Integrity Management

symmetrical or asymmetrical – this degree of symmetry is describable using the skew parameter.

All of the aforementioned parameters (mean, standard deviation etc.) are expressible mathematically. Consider a random variable X and that n observations of X are available. The mean, or expected value, of the random variable X, which is a measure of the central tendency of the data (often referred to as the first central moment), can be calculated as follows:

$$\mu_X = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \qquad (1)$$

It is noted however, that in equation (1) no distinction is made between the population and the sample mean; ordinarily it is implicitly assumed that the sample size is relatively large (the implications of small sample sizes are obvious).

The variance of X, which is a measure of the spread in the data about the mean (also known as the second central moment), can be calculated as follows:

$$Var(X) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \mu_X\right)^2 \qquad (2)$$

If the random variable X is expressed in GPa (Giga-Pascals), then the unit of variance will be in GPa². This dimensional problem can be avoided by taking the square root of the variance. This is the standard deviation and is calculated as:

$$\sigma_X = \sqrt{Var(X)} \qquad (3)$$

Although the standard deviation is expressed in the same units as the mean value, its absolute value does not clearly indicate the degree of dispersion in the random variable, without referring to the mean value. Since the mean and standard deviation are expressed in the same units, a non-dimensional term can be introduced by taking the ratio of the standard deviation and the mean; this parameter is the coefficient of variation (COV) and is calculated as:

$$\delta_X = \frac{\sigma_X}{\mu_X} \qquad (4)$$

For a deterministic variable the coefficient of variance is zero; the smaller the value of COV indicates a smaller degree of uncertainty of the primary variable. In many engineering applications a COV of 0.1 to 0.3 is common for a random variable.

The skewness (also known as the third central moment) is calculated as follows:

$$Skewness = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i - \mu_X\right)^3 \qquad (5)$$

Since skewness is a third moment it can be positive or negative; again to avoid dimensional problems a non-dimensional measure of skewness, known as the skewness coefficient can be introduced as follows:

$$\theta_x = \frac{Skewness}{\sigma_x^3} \qquad (6)$$

If $\theta_x$ is zero, the randomness of the primary variable is symmetric; if $\theta_x$ is positive then the dispersion is more above the mean value than below it and the converse is true if $\theta_x$ is negative.

In the context of the Young's Modulus data as shown in Table 1 above, and with reference to the above equations, the following quantitative information was obtained:

| Random Parameter | Young's Modulus |
|---|---|
| Mean | 203.9 GPa |
| Variance | 108 GPa$^2$ |
| Standard Deviation | 10.3 GPa |
| Coefficient of Variance | 0.05 |
| Skewness | 232.5 GPa$^3$ |
| Skewness Coefficient | +0.218 |

**Table 2: Statistical Parameters Associated with a Random Variable**

The above relatively simple analysis has indicated the following:

1. The uncertainty in the Young's Modulus for the structural steel beams is relatively small;

2. The randomness is asymmetrical (i.e. the dispersion is skewed about the mean);

3. The dispersion is skewed more above the mean value than below the mean.

Simple analysis of the data therefore has provided valuable information regarding the random variable, however the above is only considered a preliminary description of the randomness of the variable. A more complete description can be obtained by determining the analytical model which best describes the variable; there are a number of analytical models available which describe the randomness in the variable, these models are termed 'probability distributions'. For each probability distribution available (and there are numerous distributions available for use), there are two specific types of probability function:

1. Probability Density Function (PDF) – which is essentially a frequency curve; it does not directly provide information on probability but only indicates the nature of the randomness;

2. Cumulative Distribution Function (CDF) – this is the area beneath the PDF curve and as such directly provides the probability of a random variable having a value less than or equal to a specific, predefined value.

The relationship between the PDF and CDF is shown conceptually in the two Figures illustrated below. The Figures below have made use of the data as shown in Table 1 for the Young's Modulus of structural steel beams and these data have been

modelled using the Normal (or Gaussian) probability distribution, which was chosen in this instance because of its simplicity and wide-ranging application (it should be noted that further and more detailed analysis which is presented later in this section clearly indicates that the Gaussian distribution is not the best model to describe the randomness in E; moreover, a Gaussian distribution is symmetrical about the mean, but we already know from the preliminary analysis of the data that the randomness of E is actually skewed).
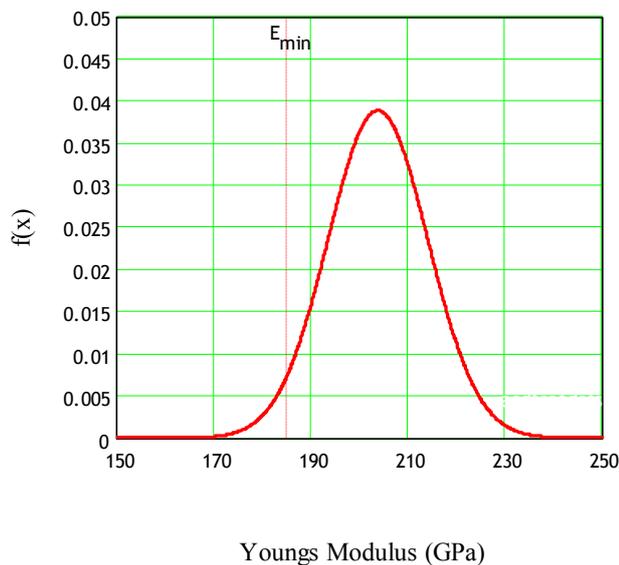


Youngs Modulus (GPa)

**Figure 2: Probability Density Function (PDF)**
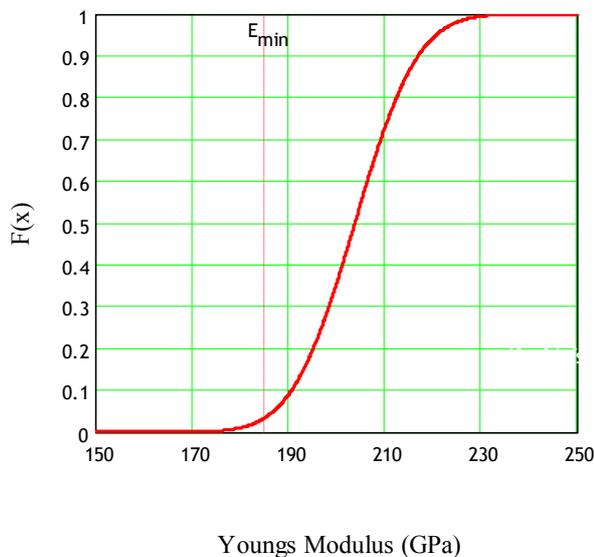


Youngs Modulus (GPa)

**Figure 3: Associated Cumulative Density Function (CDF)**

The probability distributions as shown above suggest, by way of example, the following:

1. Suppose the predefined minimum acceptable value of 'E' (which often may be stated in a basis of design document for example) is 185 GPa; the probability of values of 'E' being less than or equal to 185 GPa (shown graphically as $E_{min}$ in both Figures 1 and 2) is quantified by the point at which the line describing $E_{min}$ intersects the CDF curve, which in this instance is approximately 0.03, or 3%. Conversely, this means that the probability that the values of 'E' being greater than or equal to the stated 185 GPa minimum is 0.97, or 97%.

2. For the stated value $E_{min}$, the frequency of observation of this value is approximately 0.007; a relatively infrequent observation.

The above example helps to illustrate the relationship between the PDF and CDF for a given probability distribution. It should be noted however, that since the CDF in particular provides information on probability for a given random variable, it must satisfy all of the three previously stated axioms of probability; in particular, there are several important observations which can be made on the basis of the graphical illustrations (i.e. Figures 1 and 2):

1. The PDF must be a non-negative function; it can be zero, and theoretically its range is from -∞ to +∞.

2. The CDF must be zero at -∞ and 1.0 at +∞; using standard terminology this means $F_X(-\infty)=0$ and $F_X(+\infty)=1.0$.

3. The CDF is always greater than or equal to zero; that is $F_X(x) \geq 0$, and is a non-decreasing function of a random variable.

4. The CDF is continuous with the random variable. In addition, for continuous random variables, the CDF has a mathematical derivative.

Any mathematical function that satisfies all of the above criteria can be considered an acceptable CDF and can be used in risk and reliability analyses.

## 3.3 Modelling of Random Variables

There are a wide variety of probability distributions available for the modelling of random variables; the intention of this section however, is to illustrate how the modelling process is accomplished, rather than providing detailed descriptions of the models themselves – there are many texts available in the literature that may be referenced in this instance and it is recommended that the user review these in conjunction with this document. The intent of this document is merely to provide simple guidance rather than detailed narrative.

Procedures to calculate the probabilities of events can represent quite a significant challenge in its own right. To define a distribution uniquely, the associated parameters of a given distribution (or model) need to be estimated; this is accomplished using the available data. In practice, the choice of probability

distribution may be dictated by mathematical convenience, by familiarity with given distributions, or by standard engineering practices[2].

When sufficient data are available, numerous methods can be used to determine the underlying distribution which best 'fits' the random variable. However, it is often the case that more than one probability distribution may fit the available data and therefore the task becomes more difficult in that, first of all an accurate fit must be found, and secondly where two or more 'fits' are found, the best fit must be determined. Where 'fits' are concerned, often the naked eye suffices, however there are numerable examples where the 'goodness of fit' is such that the best fit is not discernible from the naked eye only, and so other analytical techniques are sought in such instances. This section concerns the modelling of the available data for a random variable, and determining the 'best fit'.

As an example of how random data are modelled, the data for the Young's Modulus (as presented in Table 1) will be used. The under-lying distribution can be established in several ways: (a) drawing a frequency diagram, (b) plotting the data on probability paper and (c) conducting some statistical tests known as goodness-of-fit tests for a given distribution. The methods (a), (b) and (c) are considered to be the traditional approach; extremely laborious and very time-consuming. They do not make use of the technology which is now available to the analyst, which allows an appropriate model or models to be determined very quickly, and determination of the 'best fit' model very quickly if necessary also.

The approach advocated by the author is analytical and requires MathCAD; the treatments presented herein are written in MathCAD, and in particular make use of the analytical expressions which define the CDF and/or PDF of a given probability distribution. The probability distributions used in this section are detailed in Table 3 below; definitions of the terms used in the expressions defining each probability distribution are presented below:

| Parameter | Definition |
|---|---|
| x | Random Variable |
| f(x) | Probability Density Function (PDF) describing the random variable 'x' |
| F(x) | Cumulative Distribution Function (CDF) describing the random variable 'x' |
| $\mu$ | Mean value |
| $\sigma$ | Standard Deviation |
| $\gamma$ | Euler-Mascheroni Constant (=0.57721) |
| exp or 'e' | Exponential function (where exp or e = 2.718) |
| ln or log(e) | Natural Logarithm (logarithm to the base 'e') |
| $\alpha$ | Location Factor |
| $\beta$ | Scale Factor |
| erf | Error Function (a standard integral function) |
| $\Gamma$ | Gamma function (a standard integral function) |

[2] For example, for pipeline risk and reliability assessments, it is acceptable engineering practice to assume, in the absence of actual test data, that the yield strength of the pipeline material is described by the log-normal probability distribution, with the mean value equal to 1.1 times the specified minimum value and COV equal to 0.04.

| Probability Distribution | Type | Description of Algorithm |
|---|---|---|
| Normal or Gaussian | PDF | $f(x) = \dfrac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\left[\frac{(x-\mu)^2}{2 \cdot \sigma^2}\right]}$ |
| | CDF | $F(x) = \dfrac{1}{2} \cdot \left(1 + \text{erf}\left(\dfrac{x - \mu}{\sigma \cdot \sqrt{2}}\right)\right)$ |
| Log-Normal | PDF | $f(x) = \dfrac{1}{\sqrt{2 \cdot \pi} \cdot \xi \cdot x} \cdot e^{\left[-\frac{1}{2} \cdot \left(\frac{\ln(x)-\lambda}{\xi}\right)^2\right]}$ |
| | CDF | $F(x) = \dfrac{1}{2} \cdot \left(1 + \text{erf}\left(\dfrac{\ln(x) - \lambda}{\xi \cdot \sqrt{2}}\right)\right)$ |
| Logistic | PDF | $f(x) = \dfrac{1}{4 \cdot \beta} \cdot \text{sech}^2\left(\dfrac{x - \alpha}{2 \cdot \beta}\right)$ |
| | CDF | $F(x) = \dfrac{1}{2} \cdot \left(1 + \tanh\left(\dfrac{x - \alpha}{2 \cdot \beta}\right)\right)$ |
| Gumbel | PDF | $f(x) = \dfrac{1}{\beta} \cdot \exp\left(\dfrac{x - \alpha}{\beta}\right) \cdot \exp\left(-\exp\left(\dfrac{x - \alpha}{\beta}\right)\right)$ |
| | CDF | $F(x) = 1 - \exp\left(-\exp\left(\dfrac{x - \alpha}{\beta}\right)\right)$ |
| Weibull | PDF | $f(x) = \left(\dfrac{x}{\beta}\right)^\alpha \cdot \dfrac{\alpha}{x} \cdot e^{-\left(\frac{x}{\beta}\right)^\alpha}$ |
| | CDF | $F(x) = 1 - e^{\left[-\left(\frac{x}{\beta}\right)^\alpha\right]}$ |

**Table 3: Commonly Used Probability Distributions**

The procedure for fitting random variables to the above five probability distributions is rather simple; whereas the analytical expressions may appear rather daunting, once the reader manages to assimilate the basic process involved the overall simplicity becomes more obvious.

The requirement involves the input of a basic data set (i.e. in this instance, the data in Table 1 are used) and then calculating the mean and standard deviation for that particular data set. The above five probability distributions may be modelled

using the calculated mean and standard deviation for the actual data and solving the simultaneous equations which describe the equivalent mean and standard deviation for a given probability distribution; to simplify this process the MathCAD solve block is used. If MathCAD is not available to the user, then methods (a), (b) or (c) above would then be required. The analytical expressions that define the equivalent mean and standard deviation for a given probability distribution are presented below for reference:

| Probability Distribution | Parameter | Analytical Expression |
|---|---|---|
| Normal or Gaussian | $\mu$ | Not applicable |
| | $\sigma$ | Not applicable |
| Log-Normal | $\mu$ | $\mu = \exp\left(\lambda + \dfrac{\xi^2}{2}\right)$ |
| | $\sigma$ | $\sigma = \sqrt{\left(\exp\left(\xi^2\right) - 1\right)\cdot\exp\left(2\cdot\lambda + \xi^2\right)}$ |
| Logistic | $\mu$ | $\mu = \alpha$ |
| | $\sigma$ | $\sigma = \sqrt{\dfrac{\pi^2}{3}\cdot\beta^2}$ |
| Gumbel | $\mu$ | $\mu = \alpha + \beta\cdot\gamma$ |
| | $\sigma$ | $\sigma = \sqrt{\dfrac{\pi^2}{6}\cdot\beta^2}$ |
| Weibull | $\mu$ | $\mu = \beta\cdot\Gamma\left(1 + \dfrac{1}{\alpha}\right)$ |
| | $\sigma$ | $\sigma = \sqrt{\beta^2\cdot\left(\Gamma\left(1 + \dfrac{2}{\alpha}\right) - \Gamma\left(1 + \dfrac{1}{\alpha}\right)^2\right)}$ |

**Table 4: Analytical Expressions for Mean and Standard Deviations**

The procedure for mapping the actual statistical parameters (mean and standard deviation) to the above expressions in Table 4 and solving to determine the parameters of each probability distribution is very simple when making use of MathCAD. The procedure is detailed below:

### 3.3.1 Log-Normal Probability Distribution

Mapping the statistical parameters and solving for the Log-Normal probability distribution model parameters is accomplished using the MathCAD solve block. This is initiated in the MathCAD environment using the term 'Given'; it should be noted also that within the solve block, the simultaneous equations are solved using the

Boolean assign rather than the formal MathCAD definition function. It should also be noted that where values are unknown – as in the case of the solution to determine the probability distribution parameters – MathCAD requires the formal assignment of 'guesses' for those numerical values that it is solving for; the solve block will not work otherwise. The guesses used can be any numerical value; they don't require to be close to actual values at all, just that any numerical value (including zero) is assigned. Solving for the Log-Normal probability distribution is detailed as follows:

| Description | Syntax |
|---|---|
| Initial 'guesses' for the MathCAD solve block | $\lambda := 0.2$ $\xi := 0.2$ |
| Start of solve block | Given |
| Simultaneous equations | $\mu = \exp\left(\lambda + \dfrac{\xi^2}{2}\right)$ $\sigma = \sqrt{\left(\exp\left(\xi^2\right) - 1\right) \cdot \exp\left(2 \cdot \lambda + \xi^2\right)}$ |
| Solution for Log-Normal parameters | $\begin{pmatrix} \lambda \\ \xi \end{pmatrix} := \text{Find}(\lambda, \xi)$ |
| Log-Normal Parameter Values | $\lambda = 5.316$ $\xi = 0.05$ |

**Table 5: Log-Normal Probability Distribution - Syntax for Calculating Model Parameters**

The above Table illustrates the precise syntax (MathCAD) to be used for calculation of the Log-Normal probability distribution parameters; the methodology as shown calculates the parameters for the distribution based on the calculated mean and standard deviation for the input data (this is achieved by solving the simultaneous equations which define the mean and standard deviation which are functions of the model parameters – accomplished using the 'Given' and 'Find' functions within MathCAD).

The above Table illustrates the procedure for calculating the model parameters (in this case for the Log-Normal distribution); it must be borne in mind however, that the above procedure is not a demonstration that the data are describable in terms of the particular distribution. This is accomplished in a separate exercise and will be discussed below.

### 3.3.2 Logistic, Gumbel and Weibull Probability Distributions

Mapping the statistical parameters for the Logistic, Gumbel and Weibull probability distributions is very similar to that shown above for the Log-Normal probability distribution; the primary difference being the nature of the simultaneous equations. As before, this is accomplished using the MathCAD solve block.

| Description | Syntax |
|---|---|
| Initial 'guesses' for the MathCAD solve block | $\alpha := 0.2$ <br> $\beta := 0.2$ |
| Start of solve block | Given |
| Simultaneous equations | $\mu = \alpha$ <br><br> $\sigma = \sqrt{\dfrac{\pi^2}{3} \cdot \beta^2}$ |
| Solution for Logistic parameters | $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} := \text{Find}(\alpha, \beta)$ |
| Logistic Parameter Values | $\alpha = 203.924$ <br> $\beta = 5.658$ |

**Table 6: Logistic Probability Distribution - Syntax for Calculating Model Parameters**

| Description | Syntax |
|---|---|
| Initial 'guesses' for the MathCAD solve block | $\alpha := 0.2$ <br> $\beta := 0.2$ |
| Euler-Mascheroni Constant | $\gamma := 0.57721$ |
| Start of solve block | Given |
| Simultaneous equations | $\mu = \alpha + \beta \cdot \gamma$ <br><br> $\sigma = \sqrt{\dfrac{\pi^2}{6} \cdot \beta^2}$ |
| Solution for Gumbel parameters | $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} := \text{Find}(\alpha, \beta)$ |
| Gumbel Parameter Values | $\alpha = 199.305$ <br> $\beta = 8.002$ |

**Table 7: Gumbel Probability Distribution - Syntax for Calculating Model Parameters**

| Description | Syntax |
|---|---|
| Initial 'guesses' for the MathCAD solve block | $\alpha := 0.2$ <br> $\beta := 0.2$ |
| Start of solve block | Given |
| Simultaneous equations | $\mu = \beta \cdot \Gamma\left(1 + \dfrac{1}{\alpha}\right)$ <br><br> $\sigma = \sqrt{\beta^2 \cdot \left(\Gamma\left(1 + \dfrac{2}{\alpha}\right) - \Gamma\left(1 + \dfrac{1}{\alpha}\right)^2\right)}$ |
| Solution for Weibull parameters | $\begin{pmatrix}\alpha \\ \beta\end{pmatrix} := \text{Find}(\alpha, \beta)$ |
| Weibull Parameter Values | $\alpha = 24.783$ <br> $\beta = 208.455$ |

**Table 8: Weibull Probability Distribution - Syntax for Calculating Model Parameters**

## 3.4 Modelling Probability Distributions

In the section above, the procedure for calculating the various probability distribution function model parameters (using MathCAD) was clearly illustrated; this however, is not the same as modelling the actual observations and proving that those data are explicable by a given probability distribution. The procedure for accomplishing this is described in this section. It should be noted however, that the procedure as described requires MathCAD to accomplish; if unavailable, or if comparable software systems such as MATLAB are not available then the traditional procedure of using probability paper and plotting the data to determine best fit would be required.

The procedure requires the creation of an empirical probability distribution from the actual data and then modelling against a given probability distribution; the first step in the procedure involves ranking the data and computing the probability of occurrence from the ranked data, where the probability of occurrence is calculated as:

$$F(x) = \frac{\text{rank}(x)}{\max(\text{rank}(x)) + 1} \tag{7}$$

The ranking of the input data is accomplished using a relatively simple MathCAD programme:

$$r(x) := \begin{array}{|l} \text{sorted} \leftarrow \text{sort}(x) \\ \text{for } i \in 0.. \text{ length}(x) - 1 \\ \quad \begin{array}{|l} m \leftarrow \text{match}(x_i, \text{sorted}) + 1 \\ \text{rank}_i \leftarrow \text{mean}(m) \end{array} \\ \text{rank} \end{array} \tag{8}$$

The above ranking routine ranks the input data (in this case the vector x) from the smallest or largest or vice versa using the pre-defined function r(x) as shown above. This simple algorithm accounts for multiple flows of the same value by assigning each the average rank. The data are then sorted and the cumulative probability and recurrence values calculated as follows (note: the variable in the calculations below is $E_w$, which in this instance refers to the Young's Modulus data in Table 1 above):

| Description | Syntax |
|---|---|
| Definition of the range for the sorting routine | $i := 0.. \text{ length}\left(E_w\right) - 1$ |
| Rank the input data ($E_w$) | $r_{wt} := r\left(E_w\right)$ |
| Sorting of the ranks of the input data | $r_{swt} := \text{sort}\left(r_{wt}\right)$ |
| Sorting of the basic input data ($E_w$) | $s_{wt} := \text{sort}\left(E_w\right)$ |
| Calculate the cumulative probability | $\text{prob}_i := \dfrac{r_{swt_i}}{1 + \text{rows}\left(r_{swt}\right)}$ |
| Calculate recurrence | $\text{rec}_i := \dfrac{1}{\text{prob}_i}$ |

**Table 9: Syntax for Calculating Empirical Probability Distribution for Input Data**

The above routines were used specifically to develop an empirical probability distribution for the actual input data in Table 1; these data are then plotted with a given probability distribution using the model parameters calculated in section 3.3. The 'goodness of fit' should become immediately apparent. This is accomplished by plotting the parameters rec versus $s_{wt}$ (which is the empirical data set) and x versus F(x) (where the analytical expressions for each probability distribution defining F(x) are shown in Table 3 and for which describe the idealised analytical model).
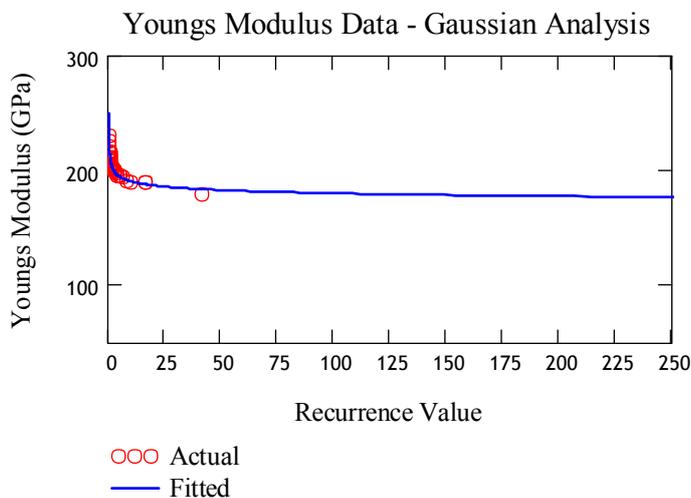


**Figure 4: Empirical and Analytical Modelling of E Data – Normal or Gaussian Distribution**
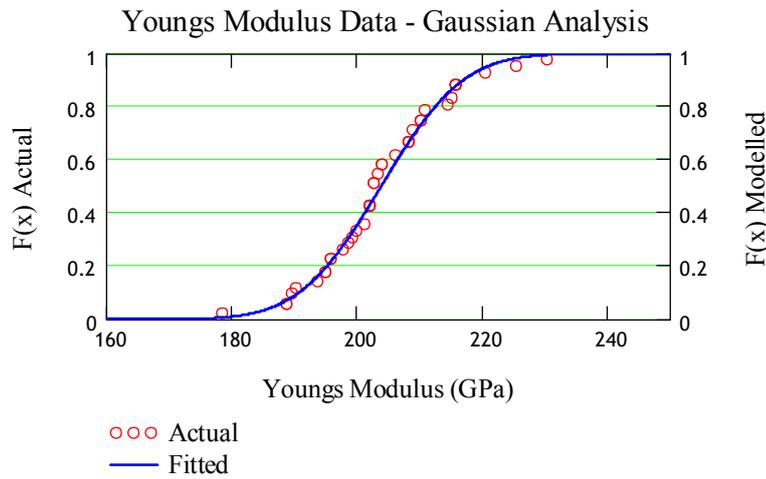
**Youngs Modulus Data - Gaussian Analysis**

○○○ Actual
—— Fitted

**Figure 5: Empirical and Analytical Modelling of E Data – Normal or Gaussian Distribution**

**Youngs Modulus Data - Log-Norm Analysis**

○○○ Actual
—— Fitted

**Figure 6: Empirical and Analytical Modelling of E Data – Log-Normal Distribution**

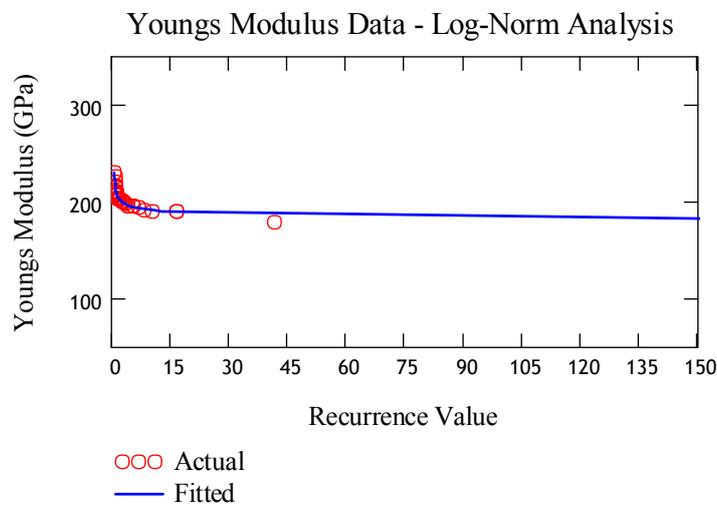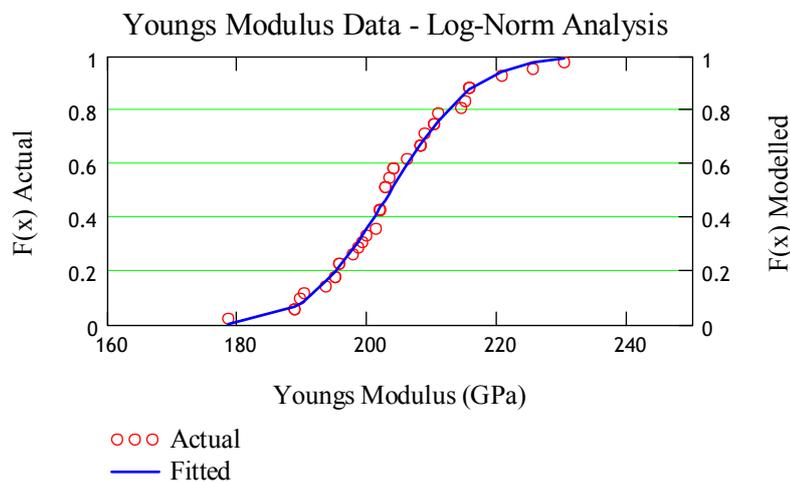**Youngs Modulus Data - Log-Norm Analysis**

○○○ Actual
—— Fitted

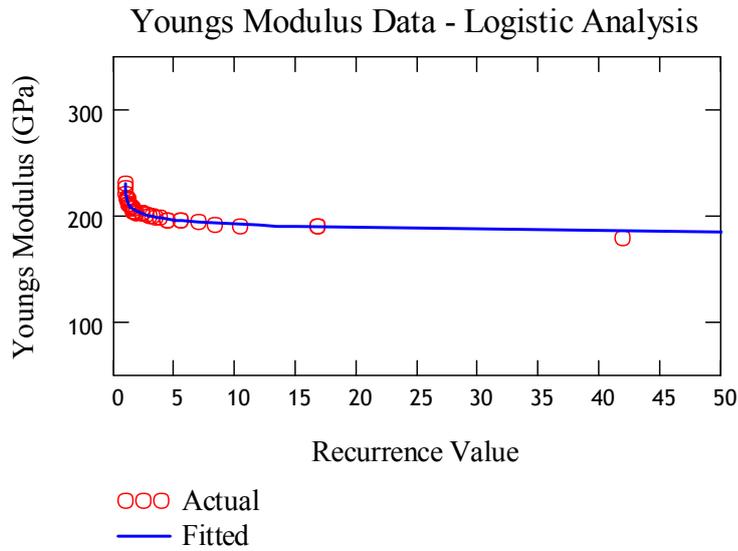**Figure 7: Empirical and Analytical Modelling of E Data – Log-Normal Distribution**

**Figure 8: Empirical and Analytical Modelling of E Data – Logistic Distribution**
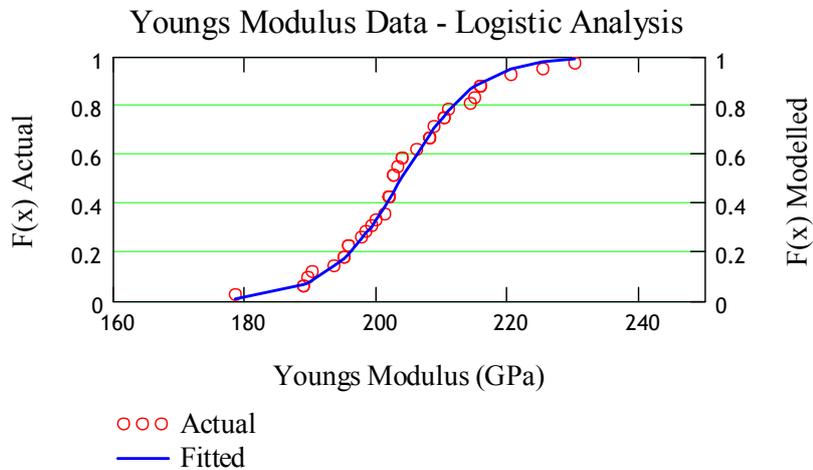


**Figure 9: Empirical and Analytical Modelling of E Data – Logistic Distribution**
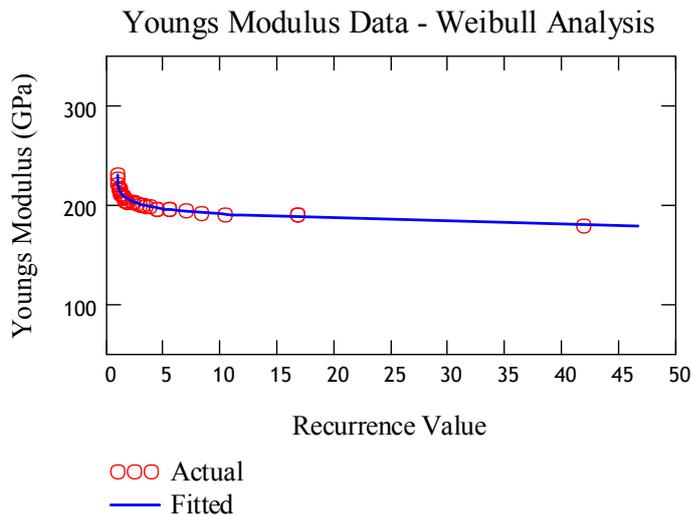
**Figure 10: Empirical and Analytical Modelling of E Data – Weibull Distribution**

Youngs Modulus Data - Weibull Analysis



Actual
Fitted
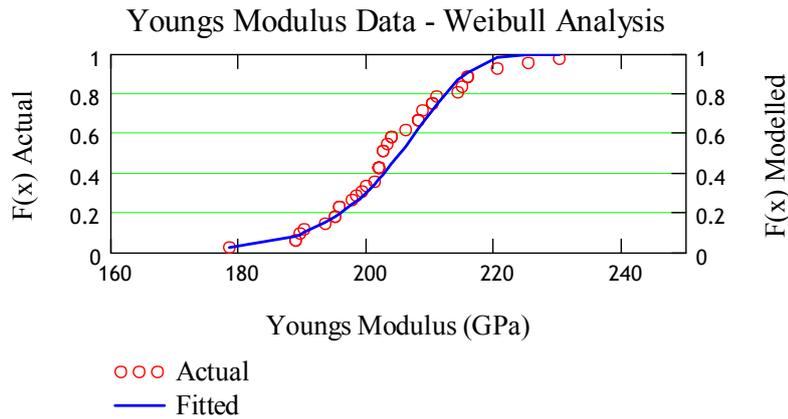
**Figure 11: Empirical and Analytical Modelling of E Data – Weibull Distribution**

Youngs Modulus Data - Gumbel Analysis



Actual
Fitted
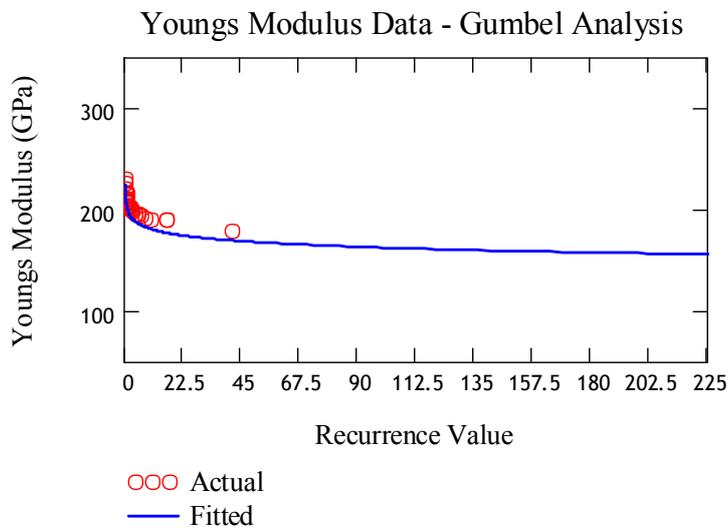
**Figure 12: Empirical and Analytical Modelling of E Data – Gumbel Distribution**

Youngs Modulus Data - Gumbel Analysis



Actual
Fitted

**Figure 13: Empirical and Analytical Modelling of E Data – Gumbel Distribution**

## 3.5 Determining the 'Best-Fit' Probability Distribution

The procedure thus far has involved the calculation of the basic statistical parameters from a given data set (in this case Young's Modulus data), determining model parameters for a series of probability distributions based on the basic statistical data and then plotting empirical probability distributions and analytical distributions to determine a 'visual' goodness of fit.
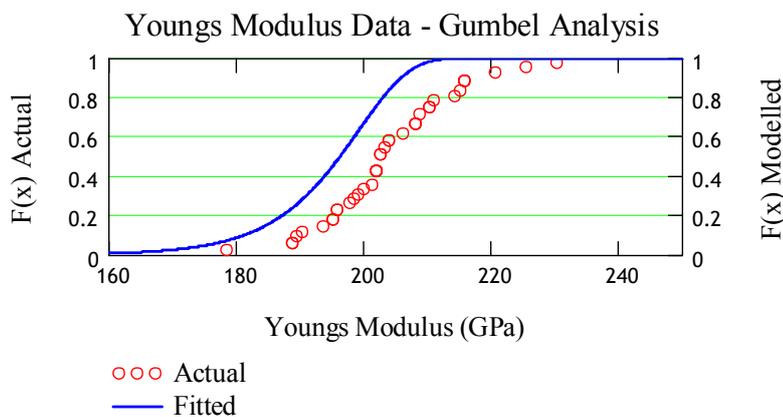
With reference to Figures 3-12 above, the following observations are worthy of note:

1. There appears to be a very good fit of the Young's Modulus data with the Normal (Gaussian), Log-Normal and Logistic probability distributions. Visual goodness of fit is good and it is not possible to discern the 'best fit' visually; notwithstanding the issue regarding the best overall fit, it is reasonable to conclude that the Young's Modulus for the structural steel beams is describable by all three.

2. The Weibull distribution describes the Young's Modulus reasonably well, however the overall fit is clearly better described by either the Normal, Log-Normal or Logistic models.

3. The Gumbel distribution is clearly unacceptable; the poor fit of the model with the actual data is clear and obvious.

There is an analytical method available to the analyst which allows the determination of the best fit to the data; a procedure which is particularly valuable when the fits appear good visually, as they do with the Normal, Log-Normal and Logistic models for the Young's Modulus data. The procedure for accomplishing this is termed the Kolmogorov-Smirnov test; this procedure is discussed in the following narrative.

### 3.5.1 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test is a simple test to determine the 'fit' of a given statistical distribution to a sample set of actual data. The test involves the determination of the maximum difference in values between two cumulative distribution functions, as follows:

$$D_n = \max \left| F_x(x_i) - S_n(x_i) \right| \qquad (9)$$

Where $F_x(x_i)$ is the theoretical CDF of the assumed distribution at the ith observation or ordered samples $x_i$, and $S_n(x_i)$ is the corresponding stepwise CDF of the observed ordered samples. The CDF of $D_n$ can be related to a significance level $\varepsilon$ as follows:

$$P\left(D \le D_n^{\varepsilon}\right) = 1 - \varepsilon \qquad (10)$$

The values of $D_n^{\varepsilon}$ at various significance levels are illustrated in the matrix below for reference. According to the K-S test, if the maximum difference $D_n$ is less than or equal to the tabulated value $D_n^{\varepsilon}$, the assumed distribution is acceptable at the significance level $\varepsilon$.

$$
\begin{pmatrix}
\text{Samples} & D_n^{\,0.1} & D_n^{\,0.05} & D_n^{\,0.01} \\
5 & 0.51 & 0.563 & 0.669 \\
6 & 0.47 & 0.521 & 0.618 \\
7 & 0.438 & 0.486 & 0.577 \\
8 & 0.411 & 0.457 & 0.543 \\
9 & 0.388 & 0.432 & 0.514 \\
10 & 0.368 & 0.409 & 0.486 \\
11 & 0.352 & 0.391 & 0.468 \\
12 & 0.338 & 0.375 & 0.45 \\
13 & 0.325 & 0.361 & 0.433 \\
14 & 0.314 & 0.349 & 0.418 \\
15 & 0.304 & 0.338 & 0.404 \\
20 & 0.264 & 0.294 & 0.352 \\
25 & 0.24 & 0.264 & 0.32 \\
30 & 0.22 & 0.242 & 0.29 \\
35 & 0.21 & 0.23 & 0.27 \\
40 & 0.19 & 0.21 & 0.25 \\
45 & 0.18 & 0.2 & 0.24 \\
50 & 0.17 & 0.19 & 0.23 \\
\text{">50"} & \dfrac{1.22}{\sqrt{n}} & \dfrac{1.36}{\sqrt{n}} & \dfrac{1.63}{\sqrt{n}}
\end{pmatrix}
$$

**Table 10: Stated Maximum $D_n^{\,\varepsilon}$ for a Given Significance Level $\varepsilon$ (n is no samples)**

The syntax for applying the above technique for the Normal (or Gaussian) probability distribution to the Young's Modulus sample set is detailed as follows:

| Description | Syntax |
|---|---|
| Determine analytical CDF based on stated E values (Normal probability distribution) | $F\left(E_w, \mu, \sigma\right) := \dfrac{1}{2} \cdot \left( 1 + \mathrm{erf}\left( \dfrac{E_w - \mu}{\sigma \cdot \sqrt{2}} \right) \right)$ |
| Sort values | $F_X := \mathrm{sort}\left( F\left(E_w, \mu, \sigma\right) \right)$ |
| Determine difference per equation 9 above | $D := F_X - \mathrm{prob}$ |
| Calculate the maximum difference value (which may be max or min) | $D_n := \max\left( D_{max}, D_{min} \right)$ <br> $D_n = 0.077$ |
| Determine the number of samples 'n' in sample set | $n := \mathrm{rows}\left( \mathrm{prob} \right)$ <br> $n = 41$ |
| Decide significance level | $\varepsilon_{ks} := 1\%$ |
| K-S Values for $D_n^{\,\alpha}$ | $D_N = \text{"See Table 10"}$ <br> $D_N := 0.25$ |
| K-S Test (per equation 10) | $KS := \begin{cases} \text{"Acceptable"} & \text{if } D_n \le D_N \\ \text{"Unacceptable"} & \text{otherwise} \end{cases}$ <br> $KS = \text{"Acceptable"}$ |

**Table 11: Syntax for the Kolmogorov-Smirnov (K-S) Test**

The above analyses (Table 11) shows that the maximum acceptable difference, for 1% significance and sample size of 41 readings, is stated as 0.25; for the Normal (or Gaussian) probability distribution, the above analysis shows that the maximum difference is 0.077 and therefore the data are acceptable in the context of the Kolmogorov-Smirnov test and therefore the data may be acceptably described by the Normal probability distribution.
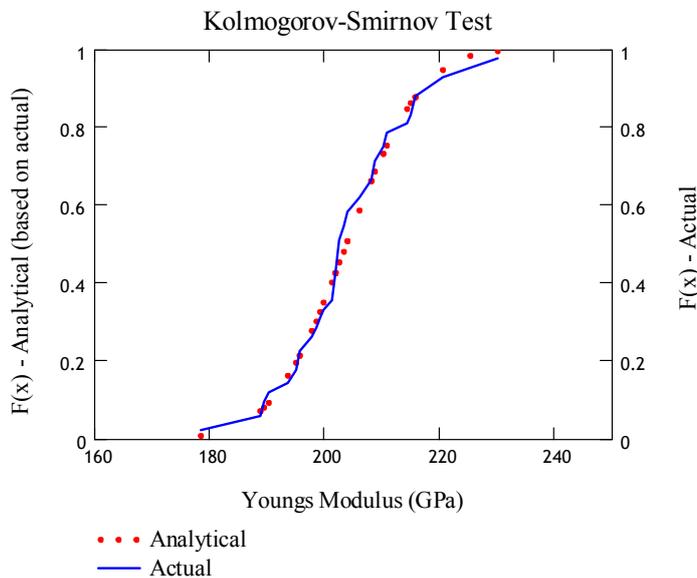


**Figure 14: Kolmogorov-Smirnov Test (Normal Probability Distribution)**

There are two primary benefits of using the Kolmogorov-Smirnov Test:

1. Demonstrates that a given probability distribution is acceptable for use, based on a pre-defined level of significance.

2. Allows the analyst to determine the best overall fit when a number of probability distributions may be acceptable.

In the context of the Young's Modulus data, the Normal, Log-Normal, Logistic and Weibull distributions were all acceptable, based on a 1% significance level; the Gumbel distribution however, was unacceptable. The following K-S test data are summarised for reference:

| Model | $D_n = \max \left| F_x(x_i) - S_n(x_i) \right|$ |
|---|---|
| Normal (or Gaussian) | 0.077 |
| Log-Normal | 0.067 |
| Logistic | 0.076 |
| Weibull | 0.137 |

**Table 12: K-S Test Data (Young's Modulus)**

The K-S Test data in Table 12 above illustrate that whereas all probability distributions are acceptable (i.e. the maximum difference is less than the stated value for a 1% significance value) the 'best fit' model is clearly shown to be the Log-Normal distribution; this is because the Log-Normal model predicted values

yield the smallest overall difference when compared with the actual distribution values.

## 3.6   Summary

This section has illustrated the procedure for modelling data sets with standard probability distributions using MathCAD.  The narrative further considers methods to determine 'best fit' models of data sets either visually and/or analytically and confirming the validity of a given model thereto, for a pre-defined level of significance level.

The above analyses are very useful for understanding the behaviour of a given variable; in the context of Integrity Assessments, the above analyses are the precursors to the assessment of failure probabilities and the assessment of risk. These analyses are generally more complicated when considered probabilistically (quantitatively); such treatments and the general concepts are discussed in the next section of these guidelines.

## 4.0    FAILURE PROBABILITY

This section discusses and details the various approaches available for calculating the failure probability of a given component; the intent is to present an overview of the procedures available, rather than to provide detailed narrative of the theory covering each.  The analytical approaches in this section make use of Reliability models, and whereas understanding the axioms of Reliability would clearly be of considerable benefit, it was considered inappropriate to include within this section; this document is meant to serve as a guide only, and moreover a guide to the detail of how each of the available Reliability models work in practice.

The intent is to provide users with a modest appreciation of the method concept(s) although avoids lengthy discussion of the methodology which is presented elsewhere (and for which it is strongly recommended that readers make reference to in conjunction with this document).  The following sections detail the approach for the solution of the various reliability algorithms for uncorrelated random variables only; for correlated random variables, various corrections must be applied and these are not discussed here.

References in the literature present many complex analytical expressions which in themselves are not useable as they are presented (there are innumerable complex integrals for example) and must be solved or modified so that the user can make use of the formulae when applying such to practical applications.  This section attempts to clarify and assist users in this respect.

### 4.1    Reliability Methods

The basic concept of classical reliability theory involves the evaluation of the probability of failure by considering specific performance criteria and the associated load and resistance parameters; it is the functional relationships that define load and resistance that is of primary interest, where mathematically this is generally defined by the following 'limit state' equation:

$$g(x) = R - S \tag{11}$$

where R is 'Resistance' (or strength) and S is 'Load'.  The conditions for which failures will occur are those when the random parameter variables which define the loads exceed those which define variable resistances (i.e. when $g(x) < 0$); the points at which failures are likely to commence therefore, are where the load and resistance variables are equal.  This is the limit state, which if over-stepped, would cause a failure to occur and is definable for all $g(x) = 0$; the points all along $g(x) = 0$ defines the so-called 'failure surface'.  The 'failure surface' delineates the boundary between the safe and unsafe regions in variable parameter space and represents a state beyond which an equipment item (such as a pipeline) can no longer fulfil the function for which it was originally designed.

On the basis of equation (11), and the fact that failures will therefore occur when $g(x) < 0$, the probability of failure ($P_f$) is calculated as:

$$P_f = \int \ldots\ldots \int_{g(x)<0} \phi_X\left(x_1, x_2, \ldots x_n\right) . dx_1 . dx_2 \ldots dx_n \tag{12}$$

Where $\phi_X(x_1, x_2, \ldots x_n)$ is the joint probability density function for the basic random variables $X_1$, $X_2$, $\ldots X_n$ and that the integration is performed over the entire failure region ($g(x) < 0$).  Equation (12) is considered to be the fundamental equation of reliability analysis, but in practice is virtually impossible to obtain for two reasons:

firstly, the joint probability density distributions can not generally be obtained and secondly, the integration in equation (12) is extremely difficult.

The approach of choice is to use analytical approximations of the integral as defined in equation (12) that are simpler to compute. Reliability methods were chosen for this specific reason as they readily provide very good estimates of failure probability: there are several methods available, however these guidelines will focus on First Order Second Moment (FOSM), First-Order Reliability Method (FORM) and the Second Order Reliability Method (SORM).

The 'limit state' equation g(x) can be either linear or non-linear functions of the basic variables; however the First Order second Moment (FOSM) and the First Order Reliability Method (FORM) assumes the failure surface to be a linear function. In a lot of cases FORM in particular can dependably be used as an approximation of the integration defined in equation (12) for the calculation of failure probabilities. In certain instances however, the curvature of the failure surface may result in significant under or over-estimations of failure probability and therefore the Second Order Reliability Method (SORM) may then be required to be used to determine a more accurate estimate. These methods will be discussed in a more practical context as to how best they can be used, with specific examples of their application, rather than an overall discussion of the theory of Reliability which may adequately be found within the literature. During the course of the narrative, where specific examples are presented, they are done so using standard MathCAD syntax, and may therefore be readily incorporated within a given MathCAD sheet for use.

## 4.2    First Order Second Moment Reliability Method (Hasofer-Lind)

The Hasofer-Lind method is applicable for normal (or Gaussian) distributed random variables. It first defines the reduced variables as:

$$X_{i\_r} = \frac{X_i - \mu_{Xi}}{\sigma_{Xi}} \qquad (13)$$

$$(i = 1, 2 .. \ n)$$

Where $X_{i\_r}$ is a random variable with zero mean and unit standard deviation. The above expression is used to transform the original limit state $g(X) = 0$ to the reduced limit state $g(X_r) = 0$. The coordinate system of $X_i$ is referred to as the original coordinate system and $X_{i\_r}$ the 'transformed' or reduced coordinate system (in other words, $X_i$ is 'normal' and $X_r$ is 'standard normal').

The Hasofer-Lind method (as with other Reliability methods) require the computation of a 'safety index', which in this case shall be designated $\beta_{HL}$; this index is defined as the minimum distance from the origin of the axes in the 'reduced' coordinate (standard normal) system to the limit state surface (failure surface - definable as $g(X) = R - S$ for all $g(X) = 0$).

The Hasofer-Lind method can be explained by considering the following limit state equation:

$$g(X) = R - S = 0 \qquad (14)$$

Where 'R' is termed the Resistance (or strength) of the component and 'S' is the load the component experiences; as both are random variables definable by probability distributions, then no single value exists where R=S, or where R-S=0. Instead there exists a range of variables which satisfy equation (14) above; this range of variables is the limit, or point beyond which failures are possible. As this is a range (or series), the point along which R=S (or where R-S=0) is termed the so-called 'failure surface'. The Hasofer-Lind method assumes that the probability distributions for both 'R' and 'S' are normally distributed, which as the example in section 3.0 above, is not always the case, therefore limiting the usefulness of this Reliability method. The procedure involved with calculating failure probabilities using the Hasofer-Lind method is nevertheless worthwhile, given it's relative simplicity and the fact that it leads nicely into the more complicated methods that are available and indeed are actually more appropriate for use.

The Hasofer-Lind method introduces the 'reduced' variables as:

$$R_r = \frac{R - \mu_R}{\sigma_R}$$

$$ \tag{15}$$

$$S_r = \frac{S - \mu_S}{\sigma_S}$$

By substitution, the above limit state equation (equation 14) then becomes:

$$g(X) = R_r \cdot \sigma_R - S_r \cdot \sigma_S + \mu_R - \mu_S = 0 \tag{16}$$

The coordinates of the intercepts of the above limit state equation on the (transformed) $R_r$ and $S_r$ axes can be shown to be as follows:

$$I_R = \left[ \frac{-(\mu_R - \mu_S)}{\sigma_R}, 0 \right]$$

$$ \tag{17}$$

$$I_S = \left( 0, \frac{\mu_R - \mu_S}{\sigma_S} \right)$$

Therefore from simple trigonometry, the distance from the origin to the failure surface as defined by the limit state equation is as follows:

$$\beta_{HL} = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \tag{18}$$

This distance (as described by the parameter $\beta_{HL}$) is referred to as the 'reliability index'. In addition, the point x* is often used in references discussing FOSM (and indeed FORM/SORM) which represents the coordinates of the minimum distance from the origin of the joint probability distribution to the limit state (often referred to as the Most Likely Failure Point (MLFP) or 'design point').

The Hasofer-Lind reliability index can be used to calculate a first-order approximation of the failure probability as:

$$P_f = \Phi\left(-\beta_{HL}\right) \tag{19}$$

Where $\Phi$ refers to the integral of the standard normal density function along the chord joining the origin of the 'reduced' axes to the 'MLFP' x*. From basic analyses, it becomes obvious that the nearer x* is to the origin, the larger is the failure probability. Thus the point of minimum distance from the origin to the limit state surface (x*), represents the worst-case combination of the stochastic variables and is appropriately named the 'most probable point of failure' or 'design point'.

At the 'MLFP' (x*), the direction cosines, along the coordinate axes (in 'reduced' coordinates) are calculated as follows:

$$\alpha_i = \left. \frac{\left(\dfrac{\partial}{\partial X_{i\_r}} g(X)\right)^{\text{¤}}}{\sqrt{\displaystyle\sum_{i=1}^{n} \left(\dfrac{\partial}{\partial X_{i\_r}} g(X)\right)^{2\text{¤}}}} \right| \tag{20}$$

for l = 1,2..n

note: the symbol "¤" denotes that the derivatives are to be determined at the MLFP. In the space of the original coordinates, the MLFP x* can be calculated as:

$$x^{\text{¤}} = \mu_{Xi} - \alpha_i \cdot \sigma_{Xi} \cdot \beta_{HL} \tag{21}$$

If we assume R and S are (normal) random variables which represent 'resistance' and applied 'load' and that the limit state equation in the original coordinate system is definable by equation (14), then the limit state equation in the reduced coordinate system is definable by equation (16); Differentiating (16) with respect to $R_r$ and $S_r$ yields the following:

$$\frac{\partial}{\partial R_r} g(X) = \sigma_R$$

$$\tag{22}$$

$$\frac{\partial}{\partial S_r} g(X) = -\sigma_S$$

The direction cosines are then calculated in accordance with the expression above as follows:

$$\alpha_R = \frac{\dfrac{\partial}{\partial R_r}g(X)}{\sqrt{\left(\dfrac{\partial}{\partial R_r}g(X)\right)^2 + \left(\dfrac{\partial}{\partial S_r}g(X)\right)^2}} \tag{23}$$

$$\alpha_S = \frac{\dfrac{\partial}{\partial S_r}g(X)}{\sqrt{\left(\dfrac{\partial}{\partial R_r}g(X)\right)^2 + \left(\dfrac{\partial}{\partial S_r}g(X)\right)^2}} \tag{24}$$

The above expressions in (23) and (24) reduce, on substitution, to the following:

$$\alpha_R = \frac{\sigma_R}{\sqrt{\sigma_R^2 + \sigma_S^2}}$$

$$\tag{25}$$

$$\alpha_S = \frac{-\sigma_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}$$

Where:

$$\alpha_R = \cos\left(\theta_R\right)$$
$$\alpha_S = \cos\left(\theta_S\right) \tag{26}$$

From the above expression ($x^{\prime} = \mu_{Xi} - \alpha_i \cdot \sigma_{Xi} \cdot \beta_{HL}$) in 'reduced' coordinates where $\mu_{Xi} = 0$, the coordinates of the 'MLFP' can be calculated as follows:

$$r_r^{\prime} = -\alpha_R \cdot \beta_{HL} = \frac{-\sigma_R}{\sqrt{\sigma_R^2 + \sigma_S^2}} \cdot \beta_{HL}$$

$$\tag{27}$$

$$s_r^{\prime} = -\alpha_S \cdot \beta_{HL} = \frac{\sigma_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \cdot \beta_{HL}$$

Substituting back into the limit state equation (in 'reduced' form) gives the following:

| | Statistical & Probabilistic Analyses – Guidelines For Use During The Assessment Of Plant Or Equipment Integrity |
|---|---|

**K L Consulting (UK) Limited**
Asset Integrity Management

$$\beta_{HL} = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \tag{28}$$

Thus indicating that the computational method used works correctly. Further, re-using the above expression ($x^\square = \mu_{Xi} - \alpha_i \cdot \sigma_{Xi} \cdot \beta_{Hl}$) the 'MLFP' in the original coordinates is given as:

$$r^\square = \mu_R - \left(\frac{\sigma_R}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) \cdot \sigma_R \left(\frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) = \frac{\mu_R \cdot \sigma_S^2 + \mu_S \cdot \sigma_R^2}{\sigma_R^2 + \sigma_S^2} \tag{29}$$

$$s^\square = \mu_S - \left(\frac{-\sigma_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) \cdot \sigma_S \cdot \left(\frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) = \frac{\mu_R \cdot \sigma_S^2 + \mu_S \cdot \sigma_R^2}{\sigma_R^2 + \sigma_S^2}$$

In this case $r^\square = s^\square$ thus indicating that the 'MLFP' lies on the limit state line. Mean-Value FOSM (MVFOSM) methods are readily usable in that the reliability index and failure probabilities are calculable as follows:

$$\beta = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \tag{30}$$

$$P_f = \Phi(-\beta)$$

There are however, several important observations that may be made:

1. As long as the limit state equation of R and S is linear and all variables are normally distributed, the safety indices and failure probabilities calculated using MVFOSM and Hasofer-Lind methods will be exactly the same. However, MVFOSM does not make use of any information with regard to the distribution of R and S, whereas Hasofer-Lind is applicable when they are normally distributed.

2. The most important difference is that in the MVFOSM method, the MLFP is at the mean values of R and S, indicating that they are not on the limit state line, whereas with the Hasofer-Lind method the 'MLFP' is on the limit state line. This can be shown through the following example.

### 4.2.1  Example: Mean-Value FOSM versus Hasofer-Lind

Suppose a cable of resistance R has to support a weight S. Assuming both R and S are normally distributed where (units are psi):

Mean Resistance (psi):               $\mu_R := 120000$

Standard Deviation (Resistance (psi)):   $\sigma_R := 18000$

Mean Load (psi): $\qquad \mu_S := 50000$

Standard Deviation (Load (psi)): $\qquad \sigma_S := 12000$

In this case, as both are normally distributed (and assuming that the limit state equation is linear) then:

$$\beta_{MVFOSM} = \beta_{HL}$$

Calculate the Reliability Index:

$$\beta := \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}$$

$$\beta = 3.236$$

Define the Standard Normal Probability Distribution PDF:

$$f(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(\frac{x^2}{-2}\right)$$

Calculate Reliability:

$$R_c := \int_{-\infty}^{\beta} f(x)\, dx$$

$$R_c = 0.99939$$

Calculate Failure Probability:

$$P_{fail} := 1 - \int_{-\infty}^{\beta} f(x)\, dx$$

$$P_{fail} = 6.066 \times 10^{-4}$$

The coordinates of the 'MLFP' according to MVFOSM however are (120 ksi, 50 ksi); according to Hasofer-Lind the coordinates of the 'MLFP' are:

Coordinates of the 'Resistance' (Hasofer-Lind Method):

$$r := \frac{\mu_R \cdot \sigma_S^2 + \mu_S \cdot \sigma_R^2}{\sigma_R^2 + \sigma_S^2}$$

$$r = 7.154 \times 10^4$$

Coordinates of the 'Load' (Hasofer-Lind Method):

$$s := \frac{\mu_R \cdot \sigma_S^2 + \mu_S \cdot \sigma_R^2}{\sigma_R^2 + \sigma_S^2}$$

$$s = 7.154 \times 10^4$$

The coordinates for the 'MLFP' are (71.54 ksi, 71.54 ksi), according to the Hasofer-Lind method, indicating that it is on the limit state equation.

## 4.3  First Order Reliability Method (FORM) - Introduction

The First Order Reliability Method (FORM) involves the transformation of a set of basic random variables '$x_i$' to a set of equivalent standard normally distributed variables '$u_i$'.  The transformation is achieved as:

$$u_i = \Phi^{-1}\left[F_i(x_i)\right]$$

and the inverse transformation:

$$x_i = \left(F_i\right)^{-1}\left(\Phi(u_i)\right)$$

Where    represents the cumulative standard normal distribution, $F_i(x_i)$ is the cumulative probability distribution of the variable $x_i$ and   $^{-1}$ and $F^{-1}$ are the corresponding inverse functions.  The method requires that the transformation to equivalent standard normal variables be achieved such that:

$$F(x_i) = F(u_i) \ \&$$

$$x_i = F^{-1}\left(F(u_i)\right)$$

This transformation produces a 'space' of variables which are distributed such that the joint probability density distribution is rotationally symmetric.   The joint distribution is then a function of the distance 'r' from the origin and has the explicit form:

$$\phi(u) = \frac{\exp\left(\frac{-r^2}{2}\right)}{2\pi}$$

The notation used in the above expressions is conventional where the lower case symbol $\phi$ denotes density distribution and the upper case symbols    and F denotes cumulative distributions.

The point in standard normal space (or u-space) where the proximity of the failure surface (g(x)=0) to the origin of the joint probability density distribution is a minimum, is considered to have the highest probability density and consequently is the most likely point of failure.  The First Order Reliability Method involves locating this point in u-space, where r(u) is determined as:

$$r(u) = \left[(u_1)^2 + (u_2)^2 + (u_3)^2 + \ldots\ldots(u_i)^2\right]^{\frac{1}{2}}$$

and the point at which r(u) is a minimum (in u-space) is defined as the First Order Reliability Index 'β' and this represents the most likely point of failure.

The First Order Reliability Method approximates the 'failure surface' as a tangent line at the most likely failure point (i.e. where $min|r(u)|=\beta$), as illustrated in Figure 15; since the joint probability density distribution is rotationally symmetric about the origin, the coordinates are transformed by rotating about the origin to produce a new axis '$z_i$' such that it passes through the most likely failure point. The failure probability ($P_f$) is then evaluated as:

$$P_f = \int_{\beta}^{\infty} \phi(z_i).dz_i$$

Or, in conventional notation:

$$P_f = \Phi(-\beta)$$

In essence therefore, FORM involves searching for the minimum value of r(u) (in u-space) and calculating a failure probability based on the integration detailed above.
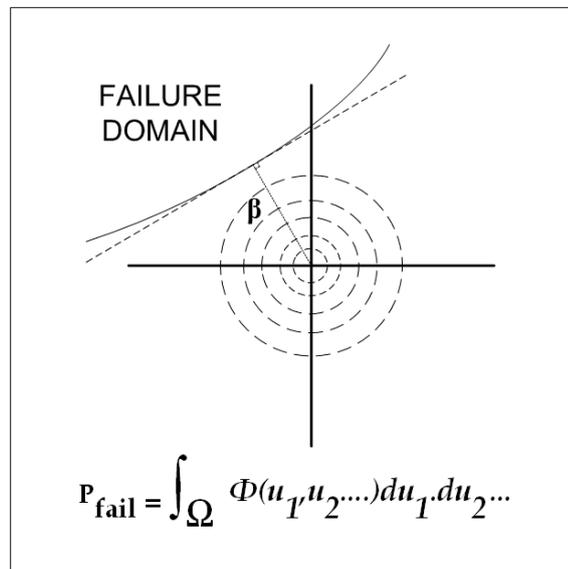


**Figure 15: FORM – Transformation of Variables and Calculation of the Most Likely Failure Point (or 'Design Point').**

### 4.3.1  Example Calculation: FORM by Constrained Minimisation

There are two methods available for use, and these are: (i) Constrained Minimisation and (ii) Newton-Raphson recursive type method.  This section deals with the procedure for solving the FORM equations using the method of Constrained Minimisation.

The procedure for determining the solution to the equations is relatively simple; as before, the syntax used below is written in MathCAD format and is therefore easily transferable into an analysts calculation sheet as it is shown herein.  The procedure also makes use of the MathCAD solve block, and the 'Minimize' function; in the absence of MathCAD therefore, alternative analytical methods would be required to solve the equations.

The FORM technique requires location of the most likely point of failure. This is achieved through a process of constrained minimisation, expressed as Minimise r(u), subject to the constraint M(u) = 0; in this case the function r(u) is defined as follows:

$$r(u) = \sqrt{\left(u_1\right)^2 + \left(u_2\right)^2 + \ldots\ldots\ldots \left(u_N\right)^2}$$

where r(u) is the distance from the origin (in u-space) and M(u) is the safety margin. Consider the following scenario for a component with known strength characteristics which is subject to known variable loadings, defined as follows:

Mean Resistance (ksi): $\mu_R := 22$

Standard Deviation (Resistance (ksi)): $\sigma_R := 2.2$

Mean Load (ksi): $\mu_S := 10$

Standard Deviation (Load (ksi)): $\sigma_S := 3$

The procedure for solving the FORM equations is detailed as follows:

| Step | Description | MathCAD Syntax |
|---|---|---|
| 1 | Define the cumulative distribution function (standard normal cumulative distribution function): | $\Phi(z) := cnorm(z)$ |
| 2 | Transform the basic input variables into equivalent standard normal variables for Resistance 'R' and Load 'S': | $R(z) := qnorm\left(\Phi(z), \mu_R, \sigma_R\right)$ $S(z) := qnorm\left(\Phi(z), \mu_S, \sigma_S\right)$ |
| 3 | Define the point of constrained minimisation by setting the initial values to zero (M(u) = 0): (note: the suffixes used in $u_1$ and $u_2$ are matrix references and not sub-scripts; the solution will not function correctly if matrix references are not used) | $u_1 := 0$ $u_2 := 0$ |
| 4 | Define the function which describes the most likely point of failure (in u-space): | $r(u) := \sqrt{\left(u_1\right)^2 + \left(u_2\right)^2}$ |
| 5 | Find the minimum value of r(u) using the technique of constrained minimisation: | Given $R\left(u_1\right) - S\left(u_2\right) = 0$ $r(u) := Minimize(r, u)$ |
| 6 | Coordinates of Most Likely Failure Point (MLFP) in u-space: (note: the 'T' in the expression for r(u) simply tells MathCAD to 'transpose' the matrix output from a single column to single row matrix) | $r(u)^T = (-1.908 \quad 2.601)$ |
| 7 | Calculate the First Order Reliability Index, β: (note: the suffixes used in $r(u)_1$ and $r(u)_2$ are matrix references and not sub-scripts; the solution will not function correctly if matrix references are not used) | $\beta := \sqrt{\left(r(u)_1\right)^2 + \left(r(u)_2\right)^2}$ $\beta = 3.226$ |

8      Define the standard normal function:

$$f(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}$$

9      Calculate the failure probability:

$$P_f := 1 - \int_{-\infty}^{\beta} f(x)\, dx$$

$$P_f = 6.285 \times 10^{-4}$$

It is possible to display the above calculations graphically, similar to the idealised illustration in Figure 14; the solution of the above equations and the plotting of the failure surface was achieved using a separate analytical routine (not shown) and plots the failure surface for the above example based on the limit state equation $g(x) = R – S$ with the results being presented in standard normal space (or u-space); this is illustrated graphically below:
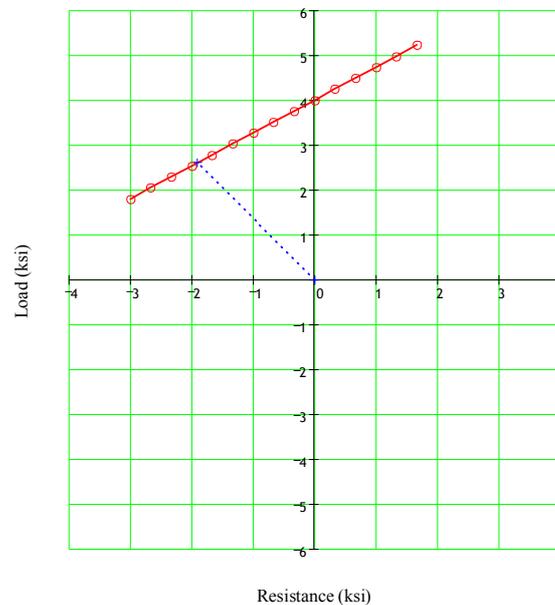


**Figure 16: Failure Surface Plotted in Standard Normal Space (Co-ordinates of β are illustrated with respect to the failure surface also)**

The above illustrates the coordinates of the Most Likely Failure Point (MLFP), which is the minimum distance in u-space from the origin of the joint probability distribution (the origin is at zero in the above plot); the dotted blue line illustrated above is the First Order Reliability Index (β).

**General Comments**

It must be recognised that FORM, and indeed MVFOSM and Hasofer-Lind methods, are approximations; they merely provide estimates of Reliability (or probability). The aforementioned methods make assumptions regarding the nature of the limit state equation (i.e. the failure surface is assumed to be linear with FORM); inaccuracies resulting from FORM are often manifested as a direct consequence of the non-linear variation of all points along $g(x) = 0$; in such cases, the more detailed

Second-Order Reliability Method (SORM) is often required. Indeed, statistical errors may also result in FORM or SORM when transforming the basic variables to equivalent standard normal variables. Transformation of variables and the assumptions in respect of the nature of the limit state equation are not actually required when using the Monte Carlo simulation method; consequently, this method is considered to be the primary yard-stick upon which to determine the accuracy and validity of FORM (and SORM, as appropriate).

The Monte Carlo numerical technique, the Monte Carlo simulation, involves the random sampling of the 'load' and 'resistance' variables to artificially simulate a large number of experiments and to observe the outcome thereof. In the case of the above example this involves sampling each of the random variables from their respective distributions and evaluating the failure surface (definable for all g(x)=0) as the limit state for those values to determine whether a failure situation (which is definable for all g(x)<0), is likely to occur. This artificial experiment is repeated many times, each time with a new random vector of variables. Therefore for N trials, the probability of failure is determined as:

$$P_f = \frac{n(g(x) < 0)}{N}$$

where n(g(x)<0) denotes the number of trials for which g(x)<0.

In the case of Monte Carlo simulation, there is an obvious relationship between the number of trials N and the degree of accuracy on $P_f$. By performing a large number of iterations, the ratio of the number of failure outcomes to the total number of iterations tends to the exact probability of failure. The primary limitation is the number of iterations required, and although Monte Carlo simulations are easily applied, they are more or less limited to determining failure probabilities in the vicinity of $10^{-5}$ per year and above (bear in mind though that for a $10^{-5}$ per year failure probability this would require approximately $10^7$ iterations to be performed).

The following MathCAD routine was developed to illustrate the Monte Carlo simulation method, and to illustrate the general accuracy of FORM in this particular instance. As before the syntax used is standard MathCAD syntax and is therefore easily transposed into an analysts calculation sheet as it is:

| Step | Description | MathCAD Syntax |
|---|---|---|
| 1 | Define the number of sample experiments to be performed in the simulation | $n := 50000$ |
| 2 | Specify the calculation range | $j := 1 .. n$ |
| 3 | Random number generators (the two routines generate random numbers on the interval 0,1) | $e_j := rnd(1)$ |
| | (note: the suffixes used in $e_j$ and $m_j$ are matrix references and not sub-scripts; the solution will not function correctly if matrix references are not used) | $m_j := rnd(1)$ |
| 4 | Generate random vectors for 'R' and 'S' | $R_j := qnorm(e_j, \mu_R, \sigma_R)$ |
| | | $S_j := qnorm(m_j, \mu_S, \sigma_S)$ |
| 5 | Define the limit state equation | $M_j := R_j - S_j$ |

6    Define a condition statement for the condition where the load exceeds the resistance (i.e. the failure condition) and sum the total number of samples where this condition is observed

$$I_j := \begin{vmatrix} 1 & \text{if } M_j < 0 \\ 0 & \text{otherwise} \end{vmatrix}$$

$$n_f := \sum_j I_j$$

$$n_f = 38$$

7    Calculate the probability of failure (which is the ratio of the total number of samples where the failure condition is satisfied and the total number of samples)

$$Pfail := \frac{n_f}{n}$$

$$Pfail = 7.6 \times 10^{-4}$$

The Monte Carlo simulation method is illustrated graphically below for the above example; this illustrates the outcome of each sample and where each resides relative to the failure condition (the failure condition being the sample results which extend beyond the boundary indicated by the blue dotted line):
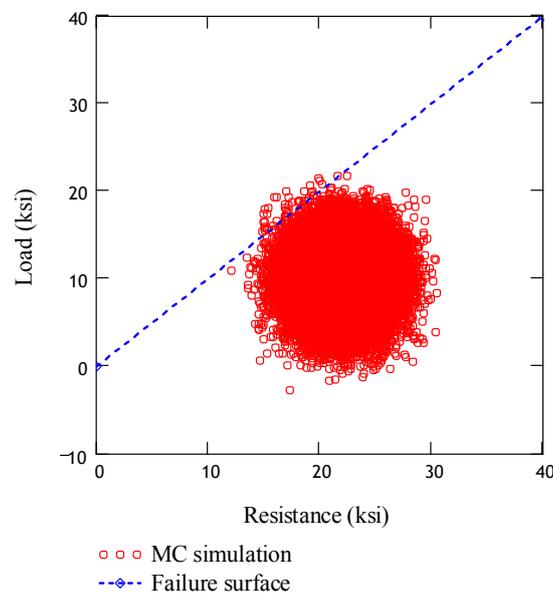


**Figure 17: Monte Carlo Simulation**

The algorithms used in the Monte Carlo simulation method are very simple as can be seen; care must nevertheless be taken to ensure that the random vector of variables being generated during the simulation are correctly defined.  One very simple check is to determine whether the random vector of variables generates the pre-defined means of the known input data for R and S, as follows:

| Step | Description | MathCAD Syntax |
|---|---|---|
| 1 | Calculate the mean Resistance 'R' from the vector generated in the simulation routine | $Rmean := \frac{1}{n} \cdot \left( \sum_j R_j \right)$ |
| | | $Rmean = 21.992$ |

2    Calculate the mean Load 'S' from the vector generated in the simulation routine

$$\text{Smean} := \frac{1}{n} \cdot \left( \sum_j S_j \right)$$

$$\text{Smean} = 9.991$$

The outputs illustrated above show that the vector of random variables provides samples with mean Resistance of 21.992 ksi and mean Load of 9.991 ksi; these values compare very favourably with the pre-defined values of 22 ksi and 10 ksi, respectively, thus confirming that the random samples generated are reliable.

More importantly however, the above calculations show that the failure probabilities are in good agreement, with FORM indicating a failure probability of 6.3. x $10^{-4}$ and the Monte Carlo Simulation 7.6 x $10^{-4}$. This confirms the validity of FORM for estimating failure probabilities; moreover FORM requires considerably less computing power than does the Monte Carlo simulation, thus making the method more preferable for practical applications.

### 4.3.2  Example Calculation: FORM by Newton-Raphson Type Recursive Method

The following example is very simple, although the precise calculation methods used are somewhat involved.  All calculations are presented to the extent possible for the first iteration only (a recursive method involves a number of iterations until convergence of the pre-defined solution criteria become satisfied); thereafter, the user progresses through the calculations until the convergence criteria are satisfied.

The example below illustrates how the Newton-Raphson type recursive method works; the calculations below considers a structural steel bar which is subject to a deterministic bending moment; the properties of the bar are pre-defined with the probability distributions which define the yield stress and plastic modulus being considered as log-normal and normal (or Gaussian), respectively.

| | |
|---|---|
| Mean Yield Stress (ksi): | $F_y := 38$ |
| Standard Deviation (Yield (ksi)): | $\sigma_{Fy} := 3.8$ |
| Mean Plastic Modulus (in$^3$): | $Z := 54$ |
| Standard Deviation (Plastic Modulus (in$^3$)): | $\sigma_z := 2.7$ |
| Deterministic Applied Bending Moment (kip-in): | $M := 1140$ |

There are two ways in which this problem can be solved; first is to consider the limit state equation in the context of the 'strength' relationship (where the limit state is $g(x) = F_y Z - 1140$) or secondly, to consider the limit state equation in the context of the 'stress' relationship (where the limit state equation is $g(x) = F_y - \dfrac{1140}{Z}$); for the purposes of these guidelines, the solution presented considers the limit state equation in the context of the 'strength' relationship.

| | Statistical & Probabilistic Analyses – Guidelines For Use During The Assessment Of Plant Or Equipment Integrity |
|---|---|

K L Consulting (UK) Limited
Asset Integrity Management

The procedure for solving the FORM equations is detailed as follows:

| Step | Description | MathCAD Syntax |
|---|---|---|
| 1 | **Step 1**: Input the Limit State Equation | $g := F_y \cdot Z - 1140$ |
| 2 | **Step 2**: Evaluate the Limit State | $g = 912$ |
| 3 | **Step 3**: Transform non-normal variables to equivalent normal variables and determine the initial coordinates of the MLFP | |

$F_y$ is log-normal and therefore must be transformed to equivalent normal values. This is achieved as follows:

The log-normal distribution for parameter x

$$f_x(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \xi \cdot x} \cdot e^{\frac{-1}{2} \cdot \left(\frac{\ln(x) - \lambda_x}{\xi_x}\right)^2}$$

$$0 \le x < \infty$$

Where

$$\lambda_x = E(\ln(x)) = \ln(\mu_x) - \frac{1}{2} \cdot \xi_x^2$$

$$\xi_x^2 = Var(\ln(x)) = \ln\left[1 + \left(\frac{\sigma_x}{\mu_x}\right)^2\right] = \ln\left(1 + \delta_x^2\right)$$

Giving

$$\xi_x = \sqrt{\ln\left(1 + \delta_x^2\right)}$$

The equivalent normal mean and standard deviation at the 'MLFP'

(note: the suffix N denotes normal)

$$\sigma_{x\_N} = \xi_x \cdot x^{\square}$$

$$\mu_{x\_N} = x^{\square} \cdot \left(1 - \ln\left(x^{\square}\right) + \lambda_x\right)$$

Initially, the coordinates of the 'MLFP' are assumed to be the means values of the variables. Thus, transforming the variables gives the following:

Determine Parameter $\xi$

$$\xi_{Fy} := \sqrt{\ln\left(1 + \delta_{Fy}^2\right)}$$

$$\xi_{Fy} = 0.0997513$$

Determine Parameter $\lambda$

$$\lambda_{Fy} := \ln\left(F_y\right) - \frac{1}{2} \cdot \xi_{Fy}^2$$

$$\lambda_{Fy} = 3.632611$$

Determine Equivalent Normal Mean

$$\mu_{Fy\_N} := F_y \cdot \left(1 - \ln\left(F_y\right) + \lambda_{Fy}\right)$$

$$\mu_{Fy\_N} = 37.811$$

Determine Equivalent Normal
Standard Deviation

$$\sigma_{Fy\_N} := \xi_{Fy} \cdot F_y$$

$$\sigma_{Fy\_N} = 3.791$$

Compute the mean and standard
deviation at the 'MLFP' of the
equivalent normal distribution for
variables which are non-normal.  This
is calculated as:

$$x_{i\_N}^{¤} = \frac{x_{i\_N} - \mu_{i\_N}}{\sigma_{x\_N}}$$

Using the above expression, the
coordinates of the 'MLFP' in the
equivalent standard normal space are:

$$f_{¤\_N} := \frac{F_y - \mu_{Fy\_N}}{\sigma_{Fy\_N}}$$

$$f_{¤\_N} = 0.05$$

and:

$$z_{¤\_N} := \frac{Z - Z}{\sigma_z}$$

$$z_{¤\_N} = 0$$

4    **Step 4:** Evaluate the partial derivatives at the 'MLFP'

$$\frac{\partial g}{\partial F_y} \qquad\qquad \frac{\partial g}{\partial Z}$$

Therefore the derivatives at the 'MLFP' are:

$$\left(\frac{\partial g}{\partial F_y}\right)^{¤} = z_¤ \qquad \text{and} \qquad \left(\frac{\partial g}{\partial Z}\right)^{¤} = f_¤$$

Derivative of Z (plastic modulus) at the
MLFP

$$z_¤ := \frac{\partial}{\partial F_y}\left(F_y \cdot Z - 1140\right)$$

$$z_¤ = 54$$

Derivative of F (yield stress) at the MLFP

$$f_¤ := \frac{\partial}{\partial Z}\left(F_y \cdot Z - 1140\right)$$

$$f_¤ = 38$$

5    **Step 5**:  Compute the partial derivatives in the equivalent standard normal space at the MLFP:

$$\frac{\partial g}{\partial x_{i\_N}}$$

This is achieved using the chain rule of differentiation;

$$\frac{\partial g}{\partial x_{i\_N}} = \frac{\partial g}{\partial x_i} \cdot \frac{\partial x_i}{\partial x_{i\_N}} = \frac{\partial g}{\partial x_i} \cdot \sigma_{x\_N}$$

Thus:

$$\frac{\partial g}{\partial F_{y\_N}} = \left( \frac{\partial g}{\partial F_y} \right) \cdot \sigma_{Fy\_N} \qquad \text{and} \qquad \frac{\partial g}{\partial Z_N} = \left( \frac{\partial g}{\partial Z} \right) \cdot \sigma_z$$

Putting

$$p = \left( \frac{\partial g}{\partial F_y} \right) \cdot \sigma_{Fy\_N} \qquad \text{and} \qquad q = \left( \frac{\partial g}{\partial Z} \right) \cdot \sigma_z \qquad \text{(note: Z is already normal)}$$

$$p := \left[ \frac{\partial}{\partial F_y} \left( F_y \cdot Z - 1140 \right) \right] \cdot \sigma_{Fy\_N} \qquad q := \left[ \frac{\partial}{\partial Z} \left( F_y \cdot Z - 1140 \right) \right] \cdot \sigma_z$$

$$p = 204.69 \qquad\qquad q = 102.6$$

6    **Step 6**: Determine the coordinates of the new 'MLFP' in the equivalent standard normal space.

This is achieved using the Newton-Raphson type recursive algorithm as follows:

$$x_{k+1}^{N\square} = \frac{1}{\left| \nabla g(x_k^{N\square}) \right|^2} \left[ \nabla g(x_k^{N\square}) . x_k^{N\square} - g(x_k^{N\square}) \right] \nabla g(x_k^{N\square})$$

The above expression is a very generalised formula for non-linear limit state equations where:

$\nabla g(x_k^{N\square})$  is the gradient vector of the limit state equation

$g(x_k^{N\square})$  is the limit state function evaluated at the 'MLFP'.

For clarity, the above parameters are definable in this example as:

$$\left| \nabla g(x_k^{N\square}) \right|^2 = \left\{ \left( \frac{\partial g}{\partial F_{y\_N}} \right)^2 + \left( \frac{\partial g}{\partial Z} \right)^2 \right\}^{\square}$$

$$\nabla g(x_k^{N\square}) . x_k^{N\square} = \left( \frac{\partial g}{\partial F_{Y\_N}} \right)^{\square} . f_{\square\_N} + \left( \frac{\partial g}{\partial Z} \right)^{\square} . z_{\square\_N}$$

**where "¤" denotes 'MLFP'.** The new coordinates of the MLFP therefore become:

$$\left\{f_{\alpha\_N}\right\} = \frac{1}{\left[\left\{\left(\frac{\partial g}{\partial F_{y\_N}}\right)^2 + \left(\frac{\partial g}{\partial Z}\right)^2\right\}^{\square}\right]} \cdot \left[\left(\frac{\partial g}{\partial F_{y\_N}}\right)^{\square} \cdot f_{\alpha\_N} + \left(\frac{\partial g}{\partial Z}\right)^{\square} \cdot z_{\alpha\_N} - g\right] \cdot \left\{\frac{\partial g}{\partial F_{y\_N}}\right\}^{\square}$$

$$\left\{z_{\alpha}\right\} = \frac{1}{\left[\left\{\left(\frac{\partial g}{\partial F_{y\_N}}\right)^2 + \left(\frac{\partial g}{\partial Z}\right)^2\right\}^{\square}\right]} \cdot \left[\left(\frac{\partial g}{\partial F_{y\_N}}\right)^{\square} \cdot f_{\alpha\_N} + \left(\frac{\partial g}{\partial Z}\right)^{\square} \cdot z_{\alpha\_N} - g\right] \cdot \left\{\frac{\partial g}{\partial Z}\right\}^{\square}$$

Therefore the new coordinates are calculated as follows:

$$\text{New\_f}_{\alpha\_N} := \frac{1}{p^2 + q^2} \cdot \left(p \cdot f_{\alpha\_N} + q \cdot z_{\alpha\_N} - g\right) \cdot p \qquad \text{New\_z}_{\alpha\_N} := \frac{1}{p^2 + q^2} \cdot \left(p \cdot f_{\alpha\_N} + q \cdot z_{\alpha\_N} - g\right) \cdot q$$

$$\text{New\_f}_{\alpha\_N} = -3.521 \qquad\qquad\qquad \text{New\_z}_{\alpha\_N} = -1.765$$

7    **Step 7**:  Calculate the distance from the origin to the new 'MLFP'.

This is calculated as follows:

$$\beta = \sqrt{\sum_{i=1}^{n} \left( x_{i\_N}^{\,¤} \right)^2}$$

Therefore the First Order Reliability Index is:

$$\beta := \sqrt{New\_f_{¤\_N}^{\,2} + New\_z_{¤\_N}^{\,2}}$$

$$\beta = 3.939$$

Define the normal distribution of variable 'x'

$$f(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left( \frac{x^2}{-2} \right)$$

Calculate the failure probability:

$$P_f := 1 - \int_{-\infty}^{\beta} f(x) \, dx$$

$$P_f = 4.099 \times 10^{-5}$$

8    **Step 8**:  Determine the new coordinates of the iteration point in the original coordinates.  This is easily calculated using the following:

$$x_i^{\,¤} = \mu_{i\_N} + \sigma_{i\_N} \cdot x_{i\_N}^{\,¤}$$

Giving

$$New\_F_y := \mu_{Fy\_N} + \sigma_{Fy\_N} \cdot New\_f_{¤\_N} \qquad\qquad New\_Z := Z + \sigma_z \cdot New\_z_{¤\_N}$$

$$New\_F_y = 24.464 \qquad\qquad\qquad New\_Z = 49.235$$

These are the revised values of both $F_y$ and Z.  The limit state equation is then evaluated using these revised coordinates as follows:

$$New\_g := New\_F_y \cdot New\_Z - 1140$$

$$New\_g = 64.5$$

This is the end of iteration one; the solution to the revised limit state equation after iteration one still does not yield zero, or even a value close thereto (in this case New_g = 64.5); consequently, a number of iterations are required until the revised limit state equation converges to zero. This convergence will yield the coordinates of the Most Likely Failure Point (MLFP) (or design point) and hence failure probability.

The Newton-Raphson type recursive method is clearly an iterative process and as such the calculations as detailed above will be required to be repeated until pre-defined convergence criteria are satisfied. Usually, the criteria for convergence of g is a tolerance of 0.01, or less. Therefore, the calculation as shown above will require repetition until this criteria is satisfied. Using the above calculations, repeating all of the steps is very easy (if tedious) and simply involves proceeding to the next iteration by going back to step 3 and reassigning revised values for both f and z.

The convergence criteria are satisfied in this case following eight (8) iterations; the solution is detailed as follows:

(i) coordinates of the MLFP (in u-space): $f_N$ = -4.465     $z_N$ = -2.568


(ii) First Order Reliability Index   $\beta$ = 5.151


(iii) Failure probability   $P_f$ = 1.297 x $10^{-7}$



## 4.4    Second Order Reliability Method (SORM)

The Second Order Reliability Method, SORM, basically takes the FORM approach further by accounting for the curvature in the limit state (it is the curvature of the limit state around the Most Likely Failure Point (MLFP) which determines the accuracy of the FORM approximation - the reliability indexes will be the same but SORM modifies the failure probability to account for the curvature of the limit state).

The first task is to compute the FORM reliability Index. There are two methods available to determine the MLFP and corresponding reliability index using the FORM technique – shown above. The SORM calculations therefore require the FORM Reliability Index and associated direction cosines as inputs.

Calculations performed during the SORM treatment are designed primarily to account for limit state functions which are non-linear; for linear limit states the FORM estimate will provide an accurate estimate of failure probability. However, where the limit state is non-linear, the failure probability will be erroneous largely on account of the fact that the curvature has an effect on the actual failure domain (where this curvature is ignored in FORM). In other words, SORM improves the result by including information about the curvature of the limit state.

### 4.4.1  Calculation of the Direction Cosines

Direction cosines are not required in the solution to the FORM equations; they however important pre-cursors to the solution of the SORM equations and so the procedure for the calculation thereof is presented within this section of narrative.

Direction cosines reflect the fact that the primary parameters are actually Vectors – they have magnitude as well as direction; the direction in which a force is applied is reflected in the associated direction cosine for the said force. The direction

cosines in the above example (as detailed in Section 4.3.2) are calculated as follows:

General expression for calculating the direction cosine for a given Vector is:

$$\alpha_{Xi} = \frac{\left(\frac{\partial g}{\partial X_i}\right)^{\square} \sigma_{Xi\_N}}{\sqrt{\sum_{i=1}^{n}\left(\left(\frac{\partial g}{\partial X_i}.\sigma_{Xi\_}\right)^{\square}\right)^2}}$$

For the limit state equation defined in Section 4.3.2:

$$g(x) = F_Y.Z - 1140$$

Direction cosine for F$_y$ is:

$$\alpha_{Fy} = \frac{\left(\frac{\partial g}{\partial F_y}\right)^{\square} \sigma_{Fy\_N}}{\sqrt{\left(\left(\frac{\partial g}{\partial F_y}.\sigma_{Fy\_N}\right)^{\square}\right)^2 + \left(\left(\frac{\partial g}{\partial Z}.\sigma_{z\_N}\right)^{\square}\right)^2}}$$

Direction cosine for Z is:

$$\alpha_{Z} = \frac{\left(\frac{\partial g}{\partial Z}\right)^{\square} \sigma_{z\_N}}{\sqrt{\left(\left(\frac{\partial g}{\partial F_y}.\sigma_{Fy\_N}\right)^{\square}\right)^2 + \left(\left(\frac{\partial g}{\partial Z}.\sigma_{z\_N}\right)^{\square}\right)^2}}$$

Using the definitions for 'p' and 'q' as illustrated above in Section 4.3.2

That is: $\quad p = \left(\frac{\partial g}{\partial F_y}\right).\sigma_{Fy\_N} \quad$ and $\quad q = \left(\frac{\partial g}{\partial Z}\right).\sigma_z$

Direction cosine for F$_y$ is:

$$\alpha_{Fy} := \frac{p}{\sqrt{p^2 + q^2}}$$

$$\alpha_{Fy} = 0.894$$

Direction cosine for Z is:

$$\alpha_z := \frac{q}{\sqrt{p^2 + q^2}}$$

$$\alpha_z = 0.448$$

## 4.4.2  SORM Calculation

The failure probability using SORM was derived by Breitung (1984) using the theory of asymptotic approximations and is given as:

$$P_f = \Phi(-\beta).\prod_{i=1}^{n-1}(1 + \beta\kappa_i)^{-\frac{1}{2}} \qquad (31)$$

Where $\kappa$ denotes the principal curvature of the limit state at the MLFP and $\beta$ is the reliability index calculated using FORM.

Brietung's SORM method makes use of a parabolic approximation and as such does not use a general second order approximation in that it ignores mixed terms and their derivatives in the Taylor series approximation of the non-linear limit state equation. Instead the approach makes use of the theory of asymptotic approximation, which in reality is only accurate for larger values of $\beta$.

Once the value of $\beta$ becomes lower, the SORM estimate can become inaccurate, in which case the user should then make use of the Tvedt approaches (*Tvedt, L., "Distribution of Quadratic Forms in Normal Space - Application to Structural Reliability", Journal of Engineering Mechanics, ASCE, Vol. 116, No.6, pp1183-1197, (1990)*). The approach detailed below covers Brietung's approach.

When evaluating $\kappa$, the variables are rotated to another set of variables, where this involves the orthogonal transformation from Y to $Y_n$ space, where:

$$Y' = R.Y$$

The transformation involves the so-called Gram-Schmidt Orthogonalisation procedure. The first step is to construct a matrix $R_0$:

General procedure for construction of the $R_0$ matrix is as follows:

$$R_0 = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \cdots & \cdots & \alpha_n \end{pmatrix}$$

The number of elements of the $R_0$ matrix clearly depends on the number of variables in the limit state equation; in this example there are merely two variables, Fy and Z, therefore the $R_0$ matrix is constructed as:

$R_0$ matrix (example calculation)

$$R_0 = \begin{pmatrix} 1 & 0 \\ \alpha_{Fy} & \alpha_z \end{pmatrix}$$

The Gram-Schmidt Orthogonalisation procedure (G-S) considers a matrix with row vectors $r_{01}, r_{02}, \ldots r_{0n}$, transforms to a matrix R whose row vectors $r_1, r_2, \ldots r_n$ are orthogonal to each other, with the nth row the same as in matrix $R_0$; i.e. $r_n = r_{0r}$.

**Step 1:** Gram-Schmidt Orthogonalisation procedure

General Gram-Schmidt formula

$$r_k = r_{0k} - \sum_{j=k+1}^{n} \frac{r_j.r_{0k}^T}{r_j.r_j^T} r_j$$

Where the superscript 'T' means the 'transpose' of the row vector (note: the rows of R have to be computed in reverse order, from n to 1). In this case, the rows of R are required to be orthonormal, in which case each of the rows requires to be normalised upon completion (so that the row vector is of unit length).

In this example, the second row is $r_2 = (0.867, 0.498)$. To compute the first row, the above expression is used, where in this case k=1 and n=2. Thus:

Calculating the row vectors:

$$r_{0k} := (1 \quad 0) \qquad\qquad r_j := \begin{pmatrix} \alpha_{Fy} & \alpha_z \end{pmatrix}$$

$$r_{0k}{}^T = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad\qquad r_j{}^T = \begin{pmatrix} 0.894 \\ 0.499 \end{pmatrix}$$

Gram-Schmidt treatment:

$$r_1 := r_{0k} - \frac{r_j \cdot r_{0k}{}^T}{r_j \cdot r_j{}^T} \cdot r_j$$

$$r_1 = (0.237 \quad -0.425)$$

Thus, normalising the elements of r1 to produce a unit vector:

$$r_1 := \begin{pmatrix} \alpha_z & -\alpha_{Fy} \end{pmatrix}$$

$$r_1 = (0.499 \quad -0.894)$$

Therefore the matrix R becomes:

$$R := \begin{pmatrix} \alpha_z & -\alpha_{Fy} \\ \alpha_{Fy} & \alpha_z \end{pmatrix}$$

**Step 2:** Derive Matrix 'A'

The matrix 'A' is derived, whose elements which are denoted $a_{ij}$ is calculated as follows:

$$a_{ij} = \frac{\left( R \cdot D \cdot R^T \right)_{ij}}{\left| \Delta G(y^\alpha) \right|}$$

$$i,j = 1,2.. \ n-1$$

Where D is the n x n second-derivative matrix of the limit state equation in the standard normal space evaluated at the MLFP, R is the rotation matrix and $\Delta G(y^\alpha)$ is the length of the gradient vector in the standard normal space.

In the rotated space, the last variable $(Y_n)$ coincides with the β-vector calculated in FORM. In the following treatments the last row in the transformed $Y_n$ vector is dropped to take account of this.

The limit state can then be re-written in terms of the second order approximation as:

$$y_n' = \beta + \frac{1}{2} \cdot y_n'^T \cdot A \cdot y'$$

The curvatures $\kappa_i$ as used in Breitung's formula are calculated as the eigenvalues of the matrix A.

**Step 3:** SORM Solution

Input the coordinates of the MLFP in the standard normal space, as determined above when the convergence criteria are satisfied:

$$f_{\alpha\_N} = -4.465$$

$$z_{\alpha\_N} = -2.568$$

The rotation matrix is R

$$R = \begin{pmatrix} 0.499 & -0.894 \\ 0.894 & 0.499 \end{pmatrix}$$

Construction of the D matrix which contains the second derivatives of the limit state equation in the standard normal space is achieved as follows (using the chain rule of differentiation):

$$f_{*y} = \mu_{Fy\_N} + \sigma_{Fy\_N}.f_{*y\_N}$$

$$z_* = \mu_{z\_N} + \sigma_{z\_N}.z_{*N}$$

Note: the above expressions define the MLFP in the original coordinates and not standard normal coordinates. The derivatives of the above expressions are calculated as follows:

$$\frac{\partial^2 g}{\partial F_{y\_N}^2} = \frac{\partial}{\partial F_y}\left\{ \left( \frac{\partial g}{\partial F_y}.\frac{\partial F_y}{\partial F_{y\_N}} \right) \right\}.\frac{\partial F_y}{\partial F_{y\_N}} = \frac{\partial}{\partial F_y}\left( Z\sigma_{Fy\_N} \right)\sigma_{Fy\_N} = 0.$$

$$\frac{\partial^2 g}{\partial Z_N^2} = \frac{\partial}{\partial Z}\left\{ \left( \frac{\partial g}{\partial Z}.\frac{\partial Z}{\partial Z_N} \right) \right\}.\frac{\partial Z}{\partial Z_N} = \frac{\partial}{\partial Z}\left( F_y.\sigma_z \right)\sigma_z = 0$$

$$\frac{\partial^2 g}{\partial F_{y\_N}.\partial Z_N} = \frac{\partial}{\partial F_y}\left\{ \left( \frac{\partial g}{\partial Z}.\frac{\partial Z}{\partial Z_N} \right) \right\}.\frac{\partial F_y}{\partial F_{y\_N}} = \frac{\partial}{\partial F_y}(F_y.\sigma_z)\sigma_{Fy\_N} = \sigma_z.\sigma_{Fy\_N}$$

Thus, the equivalent normal mean and standard deviation of $F_y$ at the MLFP is given by $\mu_{f¤}$ and $\sigma_{f¤}$.

(where ¤ in the suffixes denotes MLFP). It is these values that must be used when calculating the matrix D.

Calculate matrix 'D'

$$D := \begin{pmatrix} 0 & \sigma_z.\sigma_{f¤} \\ \sigma_z.\sigma_{f¤} & 0 \end{pmatrix}$$

$$D = \begin{pmatrix} 0 & 6.523 \\ 6.523 & 0 \end{pmatrix}$$

The length of the gradient vector at the MLFP is then calculated. In order to calculate this two derivatives need to be evaluated:

$$\frac{\partial g}{\partial F_{y\_N}} = \frac{\partial g}{\partial F_y}.\frac{\partial F_y}{\partial F_{y\_N}} = Z.\sigma_{Fy\_N}$$

$$\frac{\partial g}{\partial Z_N} = \frac{\partial g}{\partial Z}.\frac{\partial Z}{\partial Z_N} = F_y.\sigma_z$$

| | Statistical & Probabilistic Analyses – Guidelines For Use During The Assessment Of Plant Or Equipment Integrity |
|---|---|

K L Consulting (UK) Limited
Asset Integrity Management

Therefore the gradient vector becomes:

$$\Delta g := \begin{pmatrix} z_{\alpha\_i} \cdot \sigma_{f\alpha} \\ f_{\alpha\_i} \cdot \sigma_z \end{pmatrix}$$

$$\Delta g = \begin{pmatrix} 113.717 \\ 65.396 \end{pmatrix}$$

The magnitude of the gradient vector is then:

$$\left| \Delta g \right| = 131.18$$

Therefore the matrix A is calculated as:

$$A := \frac{1}{\left| \Delta g \right|} \cdot R \cdot D \cdot R^T$$

$$A = \begin{pmatrix} -0.044 & -0.027 \\ -0.027 & 0.044 \end{pmatrix}$$

The rotation of coordinates makes the last variable coincide with the $\beta$-vector. Therefore the last row and last column of the matrix A are dropped from future consideration. Therefore the eigenvalue of the remaining element $a_{00}$ is simply $\kappa_i$

Principal curvature of the limit state equation is given as:

$$\kappa := A_{0,0}$$

$$\kappa = -0.044$$

Therefore the SORM failure probability becomes:

$$P_{f\_SORM} := cnorm\left(-\beta_i\right) \cdot \left(1 + \beta_i \cdot \kappa\right)^{-0.5}$$

$$P_{f\_SORM} = 1.476 \times 10^{-7}$$

For comparison with FORM, the failure probability can be converted back to an equivalent value of $\beta$ as follows:

$$\beta_{SORM} := -qnorm\left(P_{f\_SORM}, 0, 1\right)$$

$$\beta_{SORM} = 5.126$$

Comparison of results (FORM versus SORM):

$\beta_{FORM} = 5.151$

$\beta_{SORM} = 5.126$

$P_{f\_FORM} = 1.297 \times 10^{-7}$

$P_{f\_SORM} = 1.476 \times 10^{-7}$

The above analysis highlights the general approach to solving the SORM equations. The treatments are rather complex mathematically, and will vary with the limit state equation. The narrative deliberately excludes much of the detail and theory relating to SORM as there is a vast wealth of information readily available in the literature for the analyst to peruse; the intention of this document is merely as a guideline on how the equations are solved using simple example calculations. Often a lot of the narrative available in the literature does not include example calculations, and it is the authors considered opinion that example calculations provide significant insight to the novice of how the formulae may be used in

practice and it was this approach which has been adopted more or less throughout these guidelines.

## 5.0   SUMMARY

This document represents a series of practical guidelines.  It allows an analyst to:

1. Model random parameter variables in the context of a given probability distribution.

2. Allows the determination of the 'best fit' probability distribution to a given series of observations or test data by accepted analytical means.

3. Allows a user to determine quantitative failure probabilities on the basis of given observations and/or test data.

4. Provides a considerable lead in to quantitative risk analysis; it does not provide an output of risk given that it does not consider quantitative analyses of the consequences of a given event.  This information is usually already available in either the Safety Case or COMAH report; due reference for the quantitative consequence analysis should therefore be obtained by referencing the aforementioned (mandatory) documents for calculating risk.

The treatments detailed in Section 3 and Section 4 of this document are very powerful; usage in many practical applications is usually very limited, but in certain applications may prove highly beneficial.  It must be borne in mind however, that data input requirements are considerably more onerous than semi-quantitative or qualitative approaches; the analytical techniques can be no less onerous either (as is evidenced in certain of the narrative above), however a large part of this is covered in the detail above where the example calculations show how the formulae may be solved in practice.

## 5.1   Quantitative versus Deterministic Methods

The deterministic approach (i.e. semi-quantitative or qualitative methods) has the distinct advantage of simplicity and the capability of being applied; this approach therefore lends itself to ease of application.  The disadvantages of the deterministic approach may often, but not entirely be linked to inaccuracies in the input data, but notably it is the inability to deal with uncertainties in the input data that is the primary weakness.  This therefore may lead to an underestimation of the likelihood of failure or overestimating the true 'risk' associated with a given item of equipment; in other words, the outputs from deterministic assessments can be rather uncertain.  In certain circumstances more conservative approaches may be employed which make use of more extreme values of the basic input variables and this may make the output assessments of risk overly conservative.

The primary advantage of probabilistic assessment is that it facilitates quantification of failure probability on the basis of uncertain data.  The probabilistic approach overcomes the seemingly arbitrariness of selecting contingencies in the deterministic method and causes attention to be focused on the degree of 'risk' associated with the operation of a given equipment item.  The probabilistic approach reduces the weakness in the deterministic method concerning the assumptions made with regard to the input variables, but this does not necessarily remove the possibility that important parameters are omitted, or perhaps even misjudged.  Nevertheless, the use of probabilistic methods should allow better management decisions to be made based on evaluation of the primary

threats to a given piece of equipment. But there are disadvantages in that it is intensive, time consuming and can be very complex. In addition, the inherent strength of probabilistic methods is often compromised in two areas which can limit application. The first of these is the data available to support the calculation of 'risks' and the calculation methods themselves. When performing a probabilistic assessment certain assumptions regarding the input data may have to be made; using engineering judgement often provides reasonable results, which can be improved upon in the future. In certain instances however, the lack of available data may actually preclude the use of probabilistic assessments. This problem is generally masked in the deterministic approach by the broad assumptions of what constitutes a credible threat. The second area is the choice of the target level of 'risk', which in its simplest form in this case was based on pre-defined target levels from the Safety Case or COMAH report. Clearly, a choice has to be made as to what constitutes an 'acceptable level of risk' and a strategy of risk management developed to ensure that this target level is not over-stepped. Nevertheless in view of such limitations, it may be concluded that a probabilistic assessment approach which effectively mirrors equipment operations, provides a superior basis upon which to manage risk and would therefore likely maximize both safety and business performance.