INTERNATIONAL INSTITUTE OF WELDING

UNITED KINGDOM DELEGATION

Commission XIII - WG1

Doc: IIW-XIII-WG1-114 - 03

FEBRUARY 2003

BEST PRACTICE GUIDE ON STATISTICAL ANALYSIS OF FATIGUE DATA

C R A Schneider and S J Maddox
TWI,
Granta Park, Great Abington, Cambridge, UK

CONTENTS

GLOSSARY OF TERMS

This glossary describes statistical terms as they are used in this guide, as well as those that are commonly used in other guidance documents/standards on this subject. Kendall and Buckland give more general definitions of these terms.

Alternative hypothesis
In decision theory, any admissible hypothesis that is distinct from the *null hypothesis*.

Censored data
Response data, such as fatigue endurance, is described as censored when its exact value is unknown but, for instance, it is known to fall within a certain range of values. Censored fatigue data is generally 'right' censored, which means that the endurance is known to be greater than a particular value (typically because the test stops before failure actually occurs).

Characteristic curve/value
A fatigue design (or characteristic) curve is established by adopting characteristic values that lie a certain number of standard deviations below the mean S-N curve (see Section 6).

Chi-square distribution
The chi-square, or χ^2, distribution is the statistical distribution followed by the sum of squares of ν independent normal variates in standard form (i.e. having zero mean and standard deviation of one). It is useful in determining *confidence limits* for the standard deviation of a sample drawn from a *normal distribution* (see Section 8.3).

Confidence interval/level /limits
Confidence limits are statistics derived from sample values, between which a population parameter under estimation will lie with some fixed probability P% (called the confidence level). The interval between the upper and lower confidence limits is called a confidence interval.

Degrees of freedom
In regression analysis, the number of degrees of freedom f is equal to the sample size n minus the number of coefficients estimated by the *regression*. It is also used as a parameter of a number of distributions, including χ^2, F and *Student's t*.

Design curve
See *Characteristic curve*.

Extreme value statistic
The statistic given by the smallest (or largest) observation in a sample. An extreme value statistic is a particular type of *order statistic* (so the terms are often used interchangeably).

Gaussian distribution
An alternative name for the *normal distribution*.

Hypothesis
Conjecture to be tested by some statistical analysis.

Least squares method
In *regression* analysis, a method of estimation in which the regression coefficients are estimated by minimising the sum of the squares of the deviations of the data points from the fitted regression line. In certain cases, the method is equivalent to the *maximum likelihood method* (see Section 5.3.1).

Linear regression
See *regression*.

Log-normal distribution
The distribution pertaining to the variate X when $\log X$ follows a *normal distribution*.

Maximum likelihood method
A method of estimating parameters of a population (e.g. *regression* coefficients) as those values for which the likelihood of obtaining the observed data is maximised (see Section 5.3.1).

Normal distribution
A symmetrical distribution that commonly arises as the sum of a large number of variates (e.g. measurement errors) having similar distributions to one another. For this reason, data is often assumed to follow a normal distribution in the absence of information to the contrary.

Null hypothesis
In decision theory, the hypothesis under test.

Order statistics
When a sample is arranged in ascending order of magnitude, the ordered values are called order statistics. The term can also refer, more specifically, to the extreme values of the sample.

Population
The complete set from which a random sample is taken, e.g. the set of S-N data from all components of a given type, in the context of fatigue testing.

Prediction interval/limits
Prediction limits are the limits between which a given proportion (typically 95%) of the population lies. The interval between the upper and lower prediction limits is called a prediction interval.

Random variable
See *variate*.

Regression
Process of estimating the coefficients of an equation for predicting a response y (such as $\log N$) in terms of certain independent variates (such as $\log S$). In the case of *linear regression*, the fitted equation is of the form $y = mx + c$.

Significance level
In decision theory, the probability (typically set to 5%) that the null hypothesis will be incorrectly rejected when it is, in fact, true.

Standard deviation
The most widely used measure of dispersion of a variate, equal to the square root of the *variance*.

Student's t distribution
The t distribution is the statistical distribution of the ratio of a sample mean to a sample variance for samples from a *normal distribution* in standard form (i.e. having zero mean and standard deviation of one). It is useful in determining *confidence limits* for the mean of a small sample drawn from a *normal distribution*.

Tolerance limits
Tolerance limits are values of a variate, following a given type of distribution, between which it is stated with confidence γ% that at least a proportion P% of the population will lie. This statement is made on the basis of a sample of n independent observations. A tolerance limit can thus regarded to be a confidence limit on a confidence limit.

Variance
The mean of the squared deviations of a variate from its arithmetic mean.

Variate
A quantity (also called *random variable*) that may take any of the values of a specified set with a specified relative frequency or probability.

1. INTRODUCTION

Fatigue testing is the main basis of the relationship between the fatigue resistance of a given material, component or structural detail and cyclic loading. The results of such fatigue endurance tests are plotted on graphs relating applied loading (force, stress, strain, etc) and the number of cycles to failure. Since test specimens and testing conditions are never identical, the resulting data are invariably scattered. Consequently, some judgement is required when using them to establish the required relationship. Statistical methods are available to assist in this analysis of fatigue test data, and indeed some recommendations on their use for analysing fatigue data are available.[1,2] However, they do not deal with all the statistical analyses that may be required to utilise fatigue test results and none of them offers specific guidelines for analysing fatigue data obtained from tests on welded specimens. With the increasing use of fatigue testing to supplement design rules, an approach that is now encouraged in some Standards[3-5], there is a need for comprehensive guidance on the statistical analysis of fatigue test results.

This is the subject of the present Best Practice Guide. At this stage, the focus is on fatigue endurance test results obtained under constant amplitude loading, as used to produce S-N curves. Thus, the loading is expressed as a stress range, S, and the fatigue resistance is expressed as the number of cycles, N, that can be endured by the test specimen. In general, however, the same methods can be applied to fatigue endurance test results expressed using any measure of the loading (e.g. force, strain) and results obtained under variable amplitude loading. They can also be used to analyse fatigue crack propagation data, where the loading is expressed as the stress intensity factor range, ΔK, and the fatigue resistance is expressed as the rate of crack propagation da/dN. Since the analyses are concerned purely with the experimental data, they are independent of the material tested.

2. OBJECTIVE

- To establish best practice for the statistical analysis of fatigue data obtained from welded specimens.

3. ASSUMPTIONS

3.1. FORM OF S-N CURVE

a) There is an underlying linear relationship between $\log S$ and $\log N$ of the form:

$$\log N = \log A - m \log S \qquad [1]$$

where m is the slope and $\log A$ is the intercept. This can be re-written in a form that is commonly used to describe S-N curves in design rules:

$$S^m N = A \qquad [2]$$

Note that, in practice, this assumption will only hold true between certain limits on S, as illustrated in Fig. 1. The lower limit on S is determined by the fatigue endurance limit (or just 'fatigue limit'), the stress range below which fatigue failure will not occur. In practice this is usually chosen on the basis of the endurance that can be achieved without any evidence of fatigue cracking, typically between $N = 2 \times 10^6$ and 10^7 cycles. The upper limit on S is dependent on the static strength of the test specimen but is commonly taken to be the maximum allowable static design stress[6]. However, the linear relationship between applied strain range and fatigue life for data obtained under strain control extends to much higher pseudo-elastic stress (i.e. strain × elastic modulus) levels.

b) The fatigue life N for a given stress range S is log-normally distributed.

c) The standard deviation of log N does not vary with S.

d) Each test result is statistically independent of the others.

These assumptions are rarely challenged in practice. But, if there is any reason to doubt their validity, there are statistical tests available that can help to identify departures from these assumptions. Some of these tests are listed below.

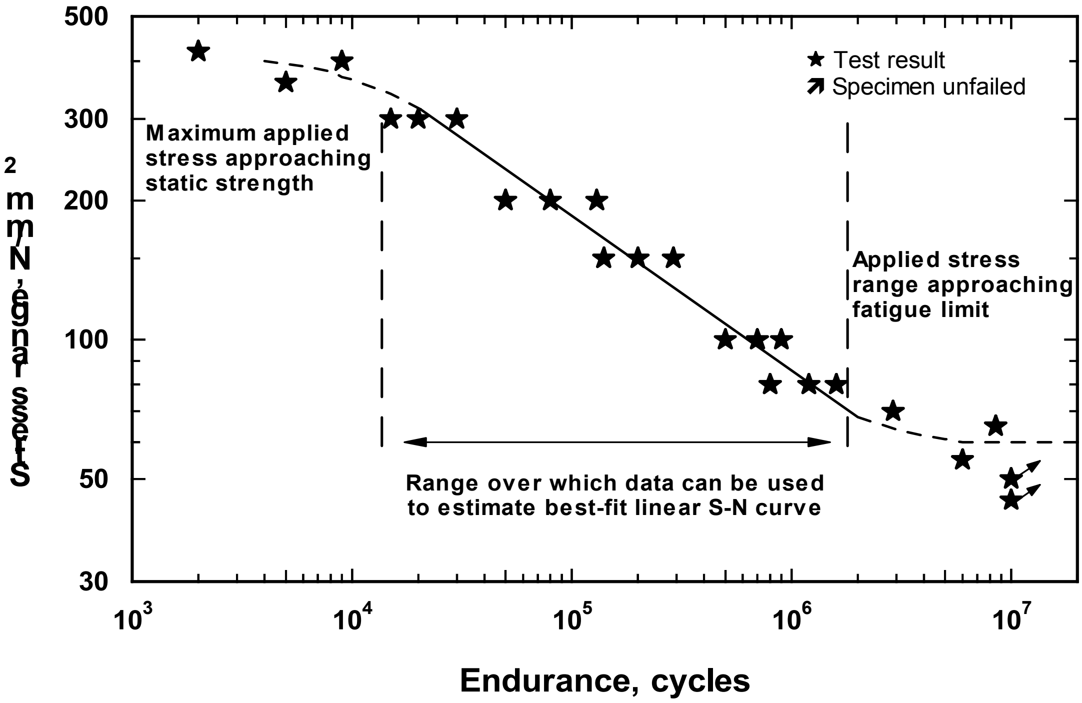

Fig.1 Typical fatigue endurance test data illustrating deviations from linear S-N curve.

3.2. TESTS FOR LINEARITY OF RELATIONSHIP BETWEEN LOGS AND LOGN

One common method of testing for non-linearity in a relationship is to fit a polynomial (typically a quadratic or cubic) in log$S$ to the data. Polynomial regression is available within most statistical software packages (e.g. MINITAB[7]). Analysis of Variance (ANOVA) can then be used to test whether the quadratic/cubic regression components are statistically significant (see Ferguson[8] for a worked example).

Box and Tidwell[9] describe a less well-known approach, which is to add a term of the form (log*S*)ln(log*S*) to the usual linear regression model. If the coefficient in this variable is significant, then this can be taken as evidence of non-linearity.

3.3. TESTS THAT *N* IS LOG-NORMALLY DISTRIBUTED

The simplest check is an 'eyeball' assessment of whether a normal probability plot of the departures (or 'residuals') of log*N* from the regression line of log*S* versus log*N* follows a linear trend. This can be done either by using standard statistical software, or by plotting the residuals on normal probability paper.

There is also a wide variety of formal statistic-based tests of normality, many of which are implemented in statistical software packages.[10-14]

3.4. TESTS OF HOMOGENEITY OF STANDARD DEVIATION OF LOG*N* WITH RESPECT TO *S*

This assumption is most easily checked by simply examining a plot of the 'residuals' from the regression versus log*S*. The assessment can be backed up by partitioning the residuals into appropriate groups and applying either Bartlett's test,[15] if log*N* is believed to be normally distributed, or otherwise Levene's test.[7]

3.5. TESTS OF STATISTICAL INDEPENDENCE OF TEST RESULTS

This assumption is difficult to check, in practice. A good starting point is to examine plots of the 'residuals' from the regression against both log*S* and against the order in which the results were collected (in case there is some time-dependence). There should not be any recognisable patterns in the residuals in either of these plots. If there are, the data can be grouped accordingly, and variations in the mean level can then be tested using Analysis of Variance. Any inhomogeneity in the standard deviations of the groups can be tested as in Section 3.4.

4. FITTING AN S-N CURVE

In their simplest form, S-N data comprise *n* data points (logS_i, logN_i), where S_i is the stress range and N_i is the endurance in cycles. This endurance is either the number of cycles to failure (or some pre-determined criterion, such as the attainment of a particular size of fatigue crack) or the number of cycles endured without failure. Fig.2 shows an example of such data, together with some fitted S-N curves.[16]

Special attention is drawn to the fact that fatigue test results are traditionally plotted with log*S* as the y-axis and log*N* as the x-axis. The standard approach in curve fitting is to assume that the parameter plotted on the x-axis is the independent variable and that plotted on the y-axis is the dependent variable. However, the opposite is the case with fatigue data presented in the traditional way. Consequently, care is needed to ensure that log*N* is treated as the **dependent** variable.

Considering only the results from specimens that failed, the intercept log*A* and slope *m* of the 'best fit' line through the data (called the 'mean' line in Fig.2) are estimated by ordinary linear regression, as described by Gurney and Maddox.[6]. The method usually used to estimate the slope and intercept coefficients is called 'least squares estimation'. This method is based on choosing those values of the coefficients that minimise the sum of the squared deviations (or 'residuals') of the observed values of log N_i from those predicted by the fitted line. TWI originally used such data and this method of analysis to derive the fatigue design rules for welded steel structures that have since formed the basis of most fatigue design rules in the world[6,17].

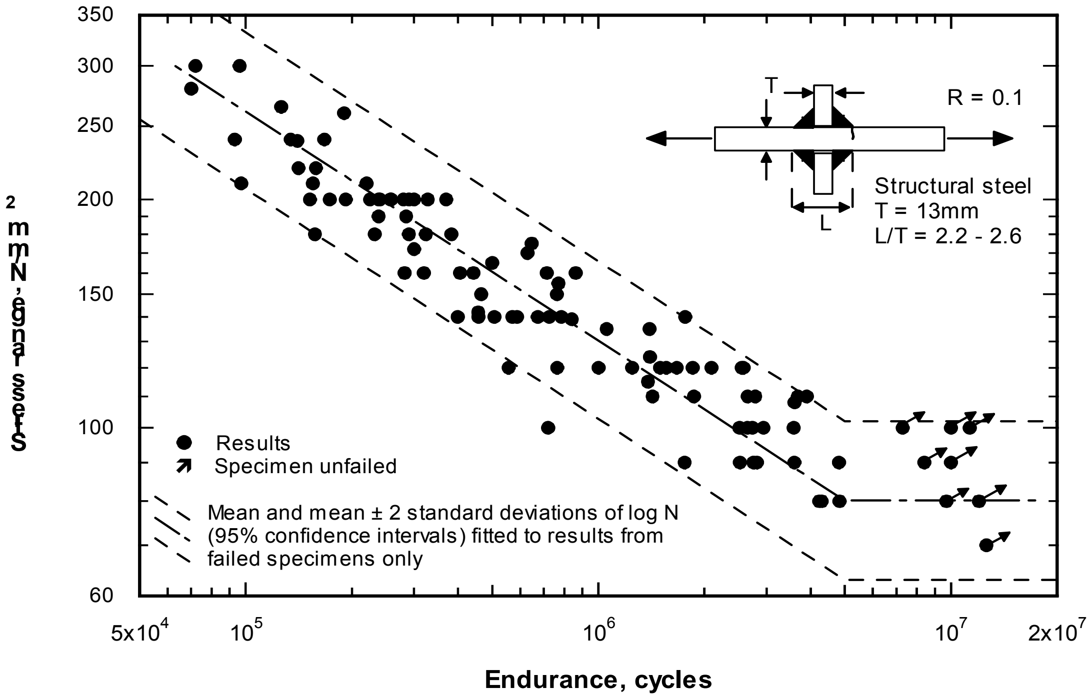

Fig.2 Example of S-N data (Maddox)[16]

5. TREATMENT OF RESULTS WHERE NO FAILURE HAS OCCURRED

5.1. INTRODUCTION

In Section 4, it was assumed that each specimen under test yielded an exact failure endurance. However, there are circumstances in which results are obtained from specimens, or parts of specimens, that have not failed. Such results, which are plotted in Fig.1 and 2 as 'Specimen unfailed', are often termed 'run-outs'. Depending on the circumstances, it may or may not be possible to use the results from unfailed specimens in the statistical analysis of the data. Indeed, even nearby results from specimens that did fail may need to be excluded from that analysis.

5.2. RESULTS ASSOCIATED WITH THE FATIGUE LIMIT

Most components exhibit a fatigue endurance limit under constant amplitude loading, defined as the stress range below which failure will not occur. In order to establish the constant amplitude fatigue limit (CAFL) experimentally, it is generally assumed to be the highest applied stress range for a given applied mean stress or stress ratio (minimum/maximum applied stress) at and below which the test specimen endures a particular number of cycles without showing any evidence of fatigue cracking. In smooth specimens, with no obvious stress concentration features, that endurance is usually around 2 x 10^6 cycles. However, in severely notched components, including most weld details, 10^7 cycles is commonly chosen.

For design purposes, it is usually assumed that the S-N curve extends down to the CAFL and then turns sharply to become a horizontal line. The data in Fig.2 have been treated in this way, with the assumption that the CAFL coincides with N = 5 x 10^6 cycles in this case, as assumed in some fatigue design rules[17]. However, in practice fatigue test results usually follow an S-N curve that gradually changes slope in the region of the CAFL, as illustrated in Fig.1. Clearly, test results **from either failed or unfailed** specimens that lie in this transition region approaching the CAFL should **not** be combined with those obtained at higher stresses when estimating the best-fit linear S-N curve. Some judgement is needed when deciding which results fall into this category but, as a guide, any from notched or welded specimens that give N < 2 x 10^6 cycles could be included, or N<10^6 cycles in the case of results obtained from smooth specimens. A test for linearity (see Section 3.2) could be used to confirm the choice.

Depending on the circumstances, it may be necessary to model the S-N curve more precisely and include the transition regions at high and low applied stresses shown in Fig.1. In such a case, the data are no longer assumed to fit a linear log *S* vs. log *N* relationship, but one that corresponds to an S-shaped curve, such as[18]:

$$N = \frac{B.\exp\left[-\left(\frac{S-E}{C}\right)^{D}\right]}{S-E} \qquad [3]$$

where *B*, *C*, *D* and *E* are constants.

5.3. FATIGUE TESTS STOPPED BEFORE FAILURE

Two other situations in which fatigue test results refer to unfailed specimens provide information that can be used in the estimation of the best-fit S-N curve. In both cases, they are situations in which fatigue failure would have occurred eventually if testing had continued. Thus, results that lie in the transition region approaching the CAFL discussed in Section 5.2 are excluded. The two situations are:

a) the test is stopped deliberately, perhaps because of time constraints;

b) the test specimen contains more than one site for potential fatigue failure and fails from just one of them. At this stage, the remaining sites are only partly through their fatigue lives.

This second case was the situation in the welded specimens that gave the results in Fig.2. Failure occurred as a result of fatigue cracking from the weld toes on one side of the attachment leaving the other side intact. Thus, their test results could have been included in Fig. 2 as 'Specimen unfailed'. In other circumstances, a test will generally be stopped even if failure occurs at a completely unrelated location. In all these cases, it is clearly desirable to infer as much as possible from the locations where failure has not occurred (so-called 'censored' data), as well as those where it has (called 'exact' data).

5.3.1. Maximum Likelihood Method

The maximum likelihood method provides an appropriate tool for solving the general problem of estimating the 'best fit' line through censored test data.[19,20] In a quite fundamental sense, the maximum likelihood method provides estimates for the slope and gradient coefficients that maximise the likelihood of obtaining the observed data. Thus, the resulting estimates are those that agree most closely with the observed data. In the special case of exact data, the maximum likelihood method leads to the least squares function on which linear regression is generally based.[19] In the general case, numerical iteration is required to derive maximum likelihood estimates. Fortunately, linear regression of censored data has general application in the field of reliability analysis, and so many statistical software packages can perform the required calculations within a few seconds on a modern PC.

5.3.2. Alternative Method based on Extreme Value Statistics

A special case of censored data can arise when testing a number of specimens, each of which contains the same number M of nominally identical welds and any of which might fail first. Maddox[21] (appended to this report for convenience) shows that, if each specimen is tested until it fails at exactly one of the M potential locations, then the S-N curve for a single weld can be established using least squares estimation, together with the tabulated extreme value statistics for the normal distribution. Although this approach is less flexible than maximum likelihood estimation, it can be performed and/or verified by hand calculation. It should also be noted that, in this case, least squares estimation is no longer equivalent to maximum likelihood estimation (because extreme value statistics for the normal distribution are not themselves distributed normally).

6. ESTABLISHING A DESIGN (OR CHARACTERISTIC) CURVE

6.1. PREDICTION LIMITS

For design purposes, it is necessary to establish limits between which a given proportion (typically 95%) of the data lie. These bounds are often termed 'confidence limits'.[6] In this Guide, the term 'prediction limits' is used instead, to avoid confusion with the confidence limits on the coefficients of the regression line.

In the case of exact data, considered in Section 4, the prediction limits at stress range S can be expressed explicitly, in the form:[15]

$$\log N_{P\%}^{\pm} = (\log A + m \log S) \pm t\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{\left(\log S - \overline{\log S}\right)^2}{\sum_{i=1}^{n}\left(\log S_i - \overline{\log S}\right)^2}} \qquad [4]$$

where: $\log A$ and m are the coefficients of the regression line through the n data points ($\log S_i$, $\log N_i$), as in Section 4,

$\overline{\log S}$ is the mean of the n values of $\log S_i$,

t is the appropriate percentage point of Student's t distribution, with f degrees of freedom,

$\hat{\sigma}^2$ is the best estimate of the variance of the data about the regression line, which is equal to the sum of squared residuals divided by the number of degrees of freedom f, and

f is equal to $n - 2$ in the case where the two coefficients of the regression line have both been estimated from the data.

In general, these prediction limits are hyperbolae, rather than straight lines, which are closest to the regression line at the mean log-stress value $\overline{\log S}$. However, it is often assumed that design curves will only be applied to values of log S that are not far removed from the mean value $\overline{\log S}$ (see Gurney and Maddox[6] for a justification of this assumption). In this case, the third (final) term under the square root of equation [4] can be ignored, and the resulting prediction limits are parallel to the regression line.

Furthermore, as the sample size increases, the second term under the square root (i.e. $1/n$) becomes negligible; Gurney and Maddox[6] ignored this term for sample sizes larger than 20, which incurs an error of at most 2% in the width of the resulting prediction interval. A closely related situation is where the slope m is chosen to take a fixed value, e.g. $m = 3$ is usually chosen for welded steel joints for which the fatigue life is dominated by crack growth[6,17]. In this case, the expression under the square root is exactly equal to one, and the number of degrees of freedom f should be increased by one (from $n - 2$ to $n - 1$). Note that this latter approach is generally recommended whenever the sample size is less than ten.[5]

As the sample size becomes even larger, the percentage points of the t distribution approach those of the normal distribution, so that approximate two-sided 95% prediction limits are, for example, given by substituting $t = 2$. IIW recommendations[5] indicate that this approximation can be used for sample sizes larger than 40 (again incurring an error of at most 2% in the width of the resulting prediction interval).

The two-sided 95% prediction limits are symmetrical, so there is a confidence of 97.5% of exceeding the lower limit $\log N_{95\%}^{-}$. This lower one-sided 97.5% prediction limit forms the basis of the most widely used fatigue design curves[17] (e.g. BS 5400, BS 7608, BS PD 5500, HSE Offshore Guidance, DNV rules).

6.2. TOLERANCE LIMITS

For sample sizes smaller than 40, the IIW document[5] suggests an alternative methodology for establishing a characteristic curve, based on estimating confidence limits on the prediction limits. Such limits are called 'tolerance limits'[22] and they are further discussed in Section 11. The use of tolerance limits rather than prediction limits would yield a more conservative design curve, and tolerance limits have the advantage that they explicitly allow for uncertainty in estimates of population statistics (e.g. standard deviation) from a small sample. However, their use for design purposes would have the following disadvantages:

- They are inherently more complicated, and therefore there is an increased risk of misinterpretation.
- They are harder to implement (within a spreadsheet, for instance).
- The extra conservatism may not be warranted for all applications.
- The validity of the approach depends critically on the assumption of normality.
- Their use represents a fundamental change to previous practice, so there is a risk of incompatibility with current design rules.

Therefore, it is not always appropriate to base a design curve on a tolerance limit. Tolerance limits are, nevertheless, a valuable tool for studying the sensitivity of a design curve to the size of the sample on which it is based (see Section 11). It is therefore recommended that tolerance limits be used as a means of justifying design curves that are based on small samples, especially for critical applications.

6.3. RESULTS WHERE NO FAILURE HAS OCCURRED

For censored data, statistical software can be used to estimate prediction limits (or tolerance limits). Alternatively, where the approach of Section 5.3.2 applies, approximate prediction limits can be derived from the set of failed specimens, in which case they take the same form as equation [4].

7. PREDICTING FATIGUE LIFE

7.1. INDIVIDUAL WELD

The mean fatigue life of an individual weld/sample is given by the 'best fit' S-N curve (Sections 4 and 5). The lower one-sided *P*% prediction limit N_{PL}^{-} is the best estimate of the fatigue life that will be exceeded by a given proportion *P*% of such weld details (Section 6).

7.2. STRUCTURE CONTAINING MANY WELDS

Assuming there is no redundancy in the structure, the fatigue life of a structure containing M identical welds, any of which might fail, is given by the minimum of the M individual fatigue lives, thus:

$$N_{\min} = \min\{N_1, N_2, \ldots, N_M\} \qquad [5]$$

where the individual fatigue lives $N_1, N_2, \ldots, N_M$ are identically distributed.

This fatigue life will, on average, be lower than that of an individual weld. The mean fatigue life of the overall structure can be estimated (at least approximately) from the extreme value statistics of the normal distribution, as described by Maddox (see Appendix).[21] However, prediction limits on the fatigue life of the structure are best obtained from the relationship:

$$\Pr(N_{\min} > N_{PL}^{-}) = \left[\Pr(N_1 > N_{PL}^{-})\right]^M \qquad [6]$$

Thus, if N_{PL}^{-} is the lower one-sided *P*% prediction limit on the fatigue life of a single weld, then it is also equal to the lower one-sided *Q*% prediction limit on the fatigue life of the overall structure as long as $Q\% = [P\%]^M$. Hence, the lower one-sided *Q*% prediction limit on the fatigue life of the structure is equal to the $[Q\%]^{1/M}$ one-sided prediction limit for a single weld (which can be evaluated as in Section 6).

8. JUSTIFYING THE USE OF A GIVEN DESIGN CURVE FROM A NEW DATA SET

8.1. PROBLEM

To validate the use of a particular design Class on the basis of a limited number n of new fatigue tests. It is assumed that the mean S-N curve for the particular design category (e.g. BS 7608 Class D) being validated, $S^m N = A_D$, and the corresponding standard deviation of $\log N$, σ, are known.

8.2. APPROACH

Initially the assumption is made that the new test results form part of the same population as that used to determine the design S-N curve (this is called the null hypothesis). Then, hypothesis testing is used to show that, under this assumption, it is very unlikely (at a specified significance level) that the new results would give such long fatigue lives. This is the basis for regarding the null hypothesis as implausible, and for accepting the alternative hypothesis that the new results actually belong to a population having longer fatigue lives than the main database.

8.3. ASSUMPTIONS

In addition to the assumptions of Section 3, the analysis of this Section depends on one or two extra assumptions concerning the compatibility of the new data set with the design curve. As for earlier assumptions, there are standard statistical tests available that can help to identify departures from the assumptions; these tests are also identified below.

a) The slope of the mean S-N curve for the new test results is the same as the slope m of the design curve. This assumption is needed where an S-N curve is assumed for the new test results, i.e. only in Sections 8.5.3. In these cases, TWI recommends the following test be routinely applied.

   This assumption can be tested, at a given significance level α% (e.g. the 5% level), by checking that m falls within the two-sided $(100 - \alpha)$ % confidence interval on the slope m_{test} of a regression line fitted through the new results. In the case of exact data, considered in Section 4, this confidence interval is given by:[15]

$$\log m^{\pm}_{test,P\%} = m_{test} \pm t\hat{\sigma}\sqrt{\frac{1}{\sum_{i=1}^{n}\left(\log S_i - \overline{\log S}\right)^2}} \qquad [7]$$

   where: the new data points are given by ($\log S_i$, $\log N_i$), for $i = 1,\ldots, n$,
   $\overline{\log S}$ is the mean of the n values of $\log S_i$,
   t is the two-sided P% percentage point (where $P = 100 - \alpha$) of Student's t distribution, with $n - 2$ degrees of freedom, and
   $\hat{\sigma}^2$ is the best estimate of the variance of the data about the regression line (as in Section 6.1).

b) The standard deviation $\hat{\sigma}$ of $\log N$ about the mean S-N curve (having the assumed fixed slope m) is the same for the new tests as for the main database.
   Under this assumption, the ratio $\nu\hat{\sigma}^s/\sigma^2$ follows a χ^2 distribution, with f degrees of freedom, where f is equal to n minus the number of coefficients estimated from the new data[23] ($f = n - 1$, usually). The assumption can thus be tested by reference to tabulated percentage points of the χ^2 distribution[22] (these are also commonly available within spreadsheet packages). This test is recommended in cases (which rarely arise in practice) where the standard deviation $\hat{\sigma}$ for the new tests is larger than the standard deviation σ for the main database.

8.4. METHOD

The null hypothesis is that the new results belong to the same population as the main database. The alternative hypothesis, that the new results belong to a population having longer fatigue lives than the main database, is accepted at the 5% level of significance if:

$$\overline{\log N_{test}} \geq \overline{\log N_D} + \frac{1.645\sigma}{\sqrt{n}} \qquad [8]$$

where $\overline{\log N_{test}}$ is the mean logarithm of the fatigue life from the tests at a particular stress,

$\overline{\log N_D}$ is the logarithm of the life from the mean S-N curve for the design Class, and the value 1.645 is obtained from standard normal probability tables for a probability of 0.95.

Note that the corresponding level of significance of 5% is commonly considered to give a sufficiently low probability of concluding that the populations are different in the case where they are actually the same. For other significance levels, different values are obtained from the tables, e.g.

10% level of significance: 1.285
5% level of significance: 1.645
2.5% level of significance: 1.960
1% level of significance: 2.330

Another way to express Eq.[8] is in the form:

$$g(N_{test}) \geq g(N_D).10^{\left(\frac{1.645\sigma}{\sqrt{n}}\right)} \qquad [9]$$

where $g(N)$ is the geometric mean of the appropriate fatigue lives.

Depending on the form of the information obtained from the fatigue tests, Eq.[5] and Eq.[6] can be applied in a number of ways, some of which are detailed below.

8.5. PRACTICAL APPLICATIONS

8.5.1. Tests Performed at the Same Stress Level

If all the fatigue tests are performed at the same stress level, $\overline{\log N_{test}}$ is the mean fatigue life obtained and n is the total number of tests. Unless the new results lie on an S-N curve with the same slope as the design curve, this approach only validates the Class at the stress level used for the new tests.

8.5.2. Repeat Tests at Selected Stress Levels

If a number of tests are repeated at a few selected stress levels, Eq.[8] is applied in turn for each stress level, $\overline{\log N_{test}}$ being the mean life for each stress and n being the number of tests performed at the particular stress level considered. These tests validate the design curve over the selected range of stress levels, even if the S-N curve for the new test results does not have the same slope m as the main database.

8.5.3. Tests Performed to Produce an S-N Curve

If tests are performed at various stress levels and an S-N curve is fitted, Eq.[8] can be modified to compare this curve and the mean S-N curve for the design Class being validated. A condition is that the curve fitted to the new results is assumed to have the same slope *m* as the design curve (see Section 8.3(a)), giving an equation of the form log$N$ + $m$log$S$ = $\overline{\log A_{test}}$. Then, Eq.[8] can be rewritten:

$$\overline{\log A_{test}} \geq \overline{\log A_D} + \frac{1.645\sigma}{\sqrt{n}} \qquad [10]$$

or

$$g(A_{test}) \geq g(A_D).10^{\left(\frac{1.645\sigma}{\sqrt{n}}\right)} \qquad [11]$$

where *n* is the total number of tests. As a result, this approach is less demanding than that in Section 8.5.2 above because it relies on the S-N curves having the same slope.

As an example, consider the situation in which 9 specimens are fatigue tested to failure to validate the use of Class D at the 5% level of significance:

Mean S-N curve for Class D:	$S^3N = 3.99 \times 10^{12}$
Standard deviation of log N:	$\sigma = 0.2097$
	$n = 9$

Thus, the S-N curve fitted to the test results, assuming $m = 3$, must lie on or above the target S-N curve

$$S^3N = A_{target}$$

where, from Eq.[11],

$$A_{target} = 3.99 \times 10^{12}.10^{\left(\frac{1.645 \times 0.2097}{3}\right)}$$

$$= 5.2 \times 10^{12}$$

In terms of the required endurances, this corresponds to achieving a mean S-N curve that is at least 1.3 times higher than the mean Class D S-N curve, or 3.42 times higher than the Class D design curve, which lies 2σ below the mean.

Note that the specimens fatigue tested do not necessarily need to fail for inequality [10] to be satisfied; they simply need to last longer (on average) than the lives obtained from the target S-N curve. This may be a more convenient approach than one in which the fatigue test conditions need to be chosen in order to establish an exact S-N curve. If, for instance, all the tests are designed to stop when a fixed value of $\log A_{target}$ = log N + m log S is reached, and none of the specimens fail, then the use of the design Class can be justified using the target curve given by:

$$\log A_{target} = \overline{\log A_D} + \frac{1.645\sigma}{\sqrt{n}} \qquad [12]$$

Note that, in this case, it would not be possible to test statistically the assumption of Section 8.3(a) that the new results have the same slope *m* as the design curve.

Using the above example, but assuming that none of the nine specimens fails, the requirement would be that the endurance of each specimen must be at least 1.3 times higher than the corresponding mean Class D life.

9. TESTING WHETHER TWO DATA SETS ARE CONSISTENT

9.1. PROBLEM

It is often required to decide, on statistical grounds, whether two sets of S-N data can be regarded as forming part of the same population. For example, it may be necessary to test if a manufacturing change or the application of a post-weld fatigue life improvement technique really produces a significant change in fatigue performance. Similarly, the problem is likely to be of interest where the two data sets have been collected under conditions that are different (e.g. different research workers), but are expected to give comparable fatigue performance. The methods below can then be used to justify the amalgamation of the two data sets into one larger data set, or to justify the conclusion that there is a significant difference between them.

9.2. APPROACH

Initially, it is assumed that the two sets of test results follow the same S-N curve and have the same residual standard deviations about this curve (this so-called 'composite' null hypothesis is, in effect, a combination of hypotheses). The observed differences are calculated between:

(a) the coefficients of the two regression lines through the two data sets, and
(b) the standard deviations $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ of the residuals about these two regression lines.

Hypothesis testing is then used to check that, under these assumptions, it is not particularly unlikely (at a specified significance level) that the above statistics could have arisen by chance. This is the basis for regarding the null hypotheses as plausible, and for rejecting the alternative hypothesis that the two sets of test results belong to different populations.

Thus, the overall approach is similar to that in Section 8, the main difference being that here the 'desired' outcome will probably be to confirm, rather than reject, the null hypotheses. For simplicity, it is assumed here that both sets of S-N data are exact data, although it is believed that the methods can, in principle, be extended to the case of censored data.

9.3. TESTS PERFORMED AT THE SAME STRESS LEVEL

9.3.1. General

In this case, the coefficients of the 'regression lines' referred to above simply reduce to the point estimates $\overline{\log N_1}$ and $\overline{\log N_2}$ of the mean logarithms of the fatigue lives. Also, the 'residual standard deviations' reduce to simple standard deviations.

9.3.2. Test that Standard Deviations are Consistent

The null hypothesis is that the two sets of results belong to populations having the same standard deviation. This hypothesis is accepted at a given significance level α% (e.g. the 5% level) if: [23]

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \le F_{\nu_1}^{\nu_2} \qquad [13]$$

where $F_{\nu_1}^{\nu_2}$ is the P% percentage point (where $P = 100 - \alpha$) of the F distribution (as obtained either from standard probability tables or software), which is a function of the numbers of degrees of freedom f_1 and f_2 used to estimate $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$. As in Section 8, the numbers of degrees of freedom are one less than the corresponding sample sizes when the tests are all performed at the same stress level.

By convention, $\hat{\sigma}_1^2$ is taken to be the larger of the two standard deviations $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, while $\hat{\sigma}_2^2$ is taken to be the smaller of the two.

9.3.3. Test that the Mean Fatigue Lives are Consistent

The null hypothesis is that the two sets of results belong to populations having the same mean fatigue life. This hypothesis is accepted at a given significance level α% (e.g. the 5% level) if: [23]

$$\left|\overline{\log N_1} - \overline{\log N_2}\right| \le t\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma_e^2} \qquad [14]$$

where $\overline{\log N_1}$, $\overline{\log N_2}$ are the mean logarithms of the fatigue lives from the two sets of tests,
n_1 and n_2 are the sample sizes for the two sets of tests,
t is the two-sided P% percentage point (where $P = 100 - \alpha$) of Student's t distribution, with $(f_1 + f_2)$ degrees of freedom,
f_1 and f_2 are the numbers of degrees of freedom used to estimate $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, which are equal to $(n_1 - 1)$ and $(n_2 - 1)$ respectively when the tests are all performed at the same stress level (as in Section 9.3.2), and
σ_e is an estimate of the common variance of the two samples, given by:

$$\sigma_e^2 = \frac{\nu_1\sigma_1^2 + \nu_2\sigma_2^2}{\nu_1 + \nu_2} \qquad [15]$$

9.4. TESTS PERFORMED TO PRODUCE AN S-N CURVE

9.4.1. Test that Residual Standard Deviations are Consistent

Equation [13] still applies in this case. But, as in Section 8, the numbers of degrees of freedom f_1 and f_2 (also used to estimate $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$) are now equal to $(n_1 - 2)$ and $(n_2 - 2)$ respectively, because two coefficients have been estimated to obtain the S-N curves.

9.4.2. Test that the Intercepts of the Two S-N Curves are Consistent

In this case, the equivalent formula to equation [14] is:

$$\left|\overline{\log A_1} - \overline{\log A_2}\right| \le t\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{\left(\overline{\log S_1}\right)^2}{\sum_{i=1}^{n_1}\left(\log S_{1,i} - \overline{\log S_1}\right)^2} + \frac{\left(\overline{\log S_2}\right)^2}{\sum_{j=1}^{n_2}\left(\log S_{2,j} - \overline{\log S_2}\right)^2}\right)\sigma_e^2} \qquad [16]$$

where $\overline{\log A_1}$, $\overline{\log A_2}$ are the estimated intercepts of the regression lines through the two data sets,

n_1 and n_2 are the sample sizes for the two sets of tests,

t is the appropriate two-sided percentage point of Student's t distribution, with $(n_1 + n_2 - 4)$ degrees of freedom,

σ_e is an estimate of the common variance of the two samples (given by equation [15])

$\overline{\log S_1}$ is the mean of the n_1 values of $\log S_i$ (i.e. the first data set), and

$\overline{\log S_2}$ is the mean of the n_2 values of $\log S_j$ (i.e. the second data set).

9.4.3. Testing that the Slopes of the Two S-N curves are Consistent (*t*-test)

In this case, the equivalent formula to equation [14] is:

$$\left|m_1 - m_2\right| \le t\sqrt{\left(\frac{1}{\sum_{i=1}^{n_1}\left(\log S_{1,i} - \overline{\log S_1}\right)^2} + \frac{1}{\sum_{j=1}^{n_2}\left(\log S_{2,j} - \overline{\log S_2}\right)^2}\right)\sigma_e^2} \qquad [17]$$

where m_1 and m_2 are the estimated slopes of the regression lines through the two data sets, and all other notation is as in Section 9.4.2.

9.5. COMPOSITE HYPOTHESES

Note that if each of the three null hypotheses in Sections 9.4.1, 9.4.2 and 9.4.3 is tested at the 5% significance level, then there will be a probability of almost 15% that one of the three will be rejected even if all three are actually correct. For this reason, a lower (less demanding) significance level is often used for each individual test when a 'composite' hypothesis such as this is tested, e.g. a significance level of 1.7% for each individual test would roughly correspond to a 5% significance level for the 'composite' hypothesis. By a similar logic, it might be considered appropriate to choose a significance level of 2.5% for each of the two individual tests described in Section 9.3.

10. TESTING WHETHER MORE THAN TWO DATA SETS ARE CONSISTENT

10.1. PROBLEM

This section considers the extent to which the methods of Section 9 can be generalised to more than two sets of S-N data, i.e. the problem of deciding, on statistical grounds, whether the data sets can be regarded as forming part of the same population.

10.2. APPROACH

Initially the assumption is made that the *M* sets of test results follow the same S-N curve and have the same residual standard deviations about this curve (this combination of null hypotheses is another example of a 'composite' hypothesis). The observed differences between the following are then assessed:

(a) the coefficients of the *M* regression lines through the two data sets, and
(b) the standard deviations $\hat{\sigma}_k^2$ ($k = 1,\ldots, M$) of the residuals about the *M* regression lines.

Finally, hypothesis testing is used to check that, under these assumptions, it is not particularly unlikely (at a specified significance level) that the observed differences could have arisen by chance. This is the basis for regarding the null hypotheses as plausible, and for rejecting the alternative hypothesis that the *M* sets of test results belong to different populations.

Thus, the overall approach is analogous to that of Section 9. In particular, the observations of Section 9.5 concerning composite hypotheses also apply here. However, unless specifically stated otherwise, the methods of this section apply to exact data only.

10.3. TESTS PERFORMED AT THE SAME STRESS LEVEL

10.3.1. General

In this case, the coefficients of the 'regression lines' referred to above simply reduce to the point estimates $\overline{\log N_k}$ ($k = 1,\ldots, M$) of the mean logarithms of the fatigue lives, and the 'residual standard deviations' reduce to simple standard deviations.

10.3.2. Test that Standard Deviations are Consistent

The null hypothesis is that the *M* sets of results belong to populations having the same standard deviation. This hypothesis can be tested by direct application of Bartlett's test.[15] Note, however, that Bartlett's test is not robust to departures from normality. An alternative test, which is valid for any continuous distribution, is Levene's test, which is reported to be more robust for small samples.[7]

10.3.3. Test that Mean Fatigue Lives are Consistent

The null hypothesis is that the *M* sets of results belong to populations having the same mean logarithms for fatigue life. This hypothesis can be tested using a method known as single-factor Analysis of Variance (ANOVA)[23]. The ANOVA method is available within a wide range of software and (for exact data) from spreadsheet packages. It can also be performed (somewhat laboriously) by hand calculation, with reference to statistical tables of the percentage points of the *F* distribution. For censored data, the ANOVA method can be applied as part of a 'general linear model' (GLM) numerical procedure.[7]

The ANOVA method identifies whether there are overall differences between the mean logarithms of the fatigue lives of the *M* data sets, but it does not, in itself, identify which data sets can be regarded as consistent and which ones are inconsistent with one another. Differences between pairs of data sets can be investigated using the *t*-test of Section 9.3.3. Also, there is a graphical analogue to the ANOVA method known as 'Analysis of Means' (ANOM), which is particularly valuable for identifying if the mean logarithm of the fatigue life of one data set is significantly different from the other mean logarithms.[7]

10.4. TESTS PERFORMED TO PRODUCE AN S-N CURVE

10.4.1. Test that Residual Standard Deviations are Consistent

The same tests as in Section 10.3.2 apply here. However (as in Section 9.4.1), the number of degrees of freedom for each data set must be reduced by one accordingly. Note that this option was not available under the implementations of these tests in Release 12 of the MINITAB software package.[7]

10.4.2. Test that the Intercepts and Slopes of the S-N Curves are Consistent

The ANOVA method applies here, as in Section 10.3.3. Again, the main difference is that the numbers of degrees of freedom for each data set must be reduced accordingly. In this case, the use of hand calculations would be so laborious as to be rendered virtually impractical. However, GLM procedures are widely available that offer implementations of the ANOVA method, and are also able to handle censored data. Differences between pairs of data sets can be investigated using the *t*-tests of Section 9.4.2 or 9.4.3, as appropriate. However, the authors are not aware of any implementations of the ANOM method in this particular case.

11. SENSITIVITY OF DESIGN CURVE TO SAMPLE SIZE

For a given stress range *S*, the lower one-sided *P*% prediction limit (as given by Eq.[4]) takes the general form:

$$\log N^{-}_{P\%} = \hat{\mu} - ts \qquad [18]$$

where: $\hat{\mu}$ is an estimate of the mean endurance at stress *S*, based on *n* observations,

s is an estimate of the standard deviation of the endurance at stress *S*, based on *f* degrees of freedom, and

t is the appropriate percentage point of Student's *t* distribution with *f* degrees of freedom,

f is equal to *n* minus the number of estimated coefficients (as previously).

Both $\hat{\mu}$ and *s* are subject to sampling uncertainties (especially when the sample size is small), which, in turn, can affect the accuracy of Eq.[18]. These sampling errors can be assessed by estimating a lower confidence limit of the form $\hat{\mu} - ks$ on the prediction limit $\log N^{-}_{P\%}$. This means a statement can be made of the form: 'At least a proportion *P*% of the population is greater than $\hat{\mu} - ks$ with confidence γ%'. The statistic *k* is called a one-sided tolerance limit factor.[22]

In the general case where the slope of the regression line is estimated from the data (i.e. $f = n - 2$), tolerance limit factors for the normal distribution are not readily available, either from standard statistical tables or from spreadsheet software. However, p117 of Owen[22] gives a formula for determining k, which requires the evaluation of both the P% percentage points of the normal distribution (which is readily available) and the γ% percentage points of the non-central t distribution. The 90%, 95% and 99% percentage points of the non-central t distribution can, in turn, be evaluated using the formulae and associated tables on p109-112 of Owen.[22] The required calculations are somewhat laborious, and are outside the scope of this Best Practice Guide. Refs. 24 and 25 contain further tables of the non-central t distribution. Owen[22] also tabulates k directly for sample sizes $n \leq 4$. Some statistical software packages also provide estimates of tolerance limits for the case $f = n - 2$ (for any given values of n, γ% and P%).[7]

When a fixed value is assumed for the slope of the regression line estimated from the data, then the number of degrees of freedom $f = n - 1$. In this case, Owen[22] tabulates k directly, over a wide range of sample sizes, for γ% = 90% and 95%, and P% = 90%, 95%, 97.5%, 99% and 99.9%. The cases likely to be of most interest in connection with fatigue analysis are γ% = 90%, with P% = 97.5% or possibly P% = 95%; for convenience, the corresponding values of k are reproduced in Table 1.

12. REFERENCES

1. BS 2864: 'Guide to statistical interpretation of data', BSI, London, 1976.

2. 'ASTM Practice for statistical analysis of linear or linearized stress-life (S-N) and strain-life (ε-N) fatigue data', ASTM E-739-91, Annual Book of ASTM Standards, ASTM, West Conshohocken, PA, 1998.

3. BS 8118: 'Structural use of aluminium, Part 1 Code of practice for design', BSI, London, 1991.

4. BS EN 13445: 'Unfired pressure vessels, Part 3', BSI, London, 2002.

5. Hobbacher A: 'Fatigue design of welded joints and components'. International Institute of Welding, Abington Publishing, Cambridge, 1996, Section 3.7.

6. Gurney T R and Maddox S J: 'A re-analysis of fatigue data for welded joints in steel'. *Welding Research International* 1973 **3** (4).

7. Minitab reference manual – Release 12 for Windows. Minitab Inc. (USA), February 1998. ISBN 0-925636-40-1.

8. Ferguson G A: 'Statistical analysis in psychology and education'. McGraw-Hill, New York, 1966, Table 21.4.

9. Box G E P and Tidwell P W: 'Transformation of the independent variables'. *Technometrics* 1962 **4** pp531-550.

10. D'Augostino R B and Stevens M A (eds.): 'Goodness-of-fit techniques'. Marcel Dekker, 1986.

11. Filliben J J: 'The probability plot correlation coefficient test for normality'. *Technometrics* 1975 **17** p111.

12. Ryan T A and Joiner B L: 'Normal probability plots and tests for normality'. Technical Report, Statistics Department, The Pennsylvania State University, 1976 (available from Minitab Inc.).

13. Shapiro S S and Francia R S: 'An approximate analysis of variance test for normality'. *Journal of the American Statistical Association* 1972 **67** p215.

14. Shapiro S S and Wilk M B: 'An analysis of variance test for normality (complete samples)'. *Biometrika* **52** p591.

15. Cooper B E: 'Statistics for experimentalists'. Pergamon, Oxford, 1969, pp100-102, 217. ISBN 0 08 012600 6.

16. Maddox S J: 'Fatigue assessment of welded structures', Proc. Int. Conf. 'Extending the life of welded structures', Pergamon, Press, Oxford, 1993, pp33-42.

17. Maddox S J: Fatigue design rules for welded structures'. *Prog. Struct. Engng. Mater.* **2** (1), Jan-March 2000, pp102-109.

18. Bastenaire F: 'New method for the statistical evaluation of constant stress amplitude fatigue test results' in 'Probabilistic Aspects of Fatigue', ASTM STP 511, ASTM, Philadelphia, 1972, pp 3-28.

19. Marquis G and Mikkola T: 'Analysis of welded structures with failed and non-failed welds based on maximum likelihood'. IIW document XIII-1822-2000, July 2000.

20. Wallin K: 'The probability of success using deterministic reliability' in 'Fatigue design and Reliability: ESIS Publication 23', Marquis G and Solin J (eds.), Elsevier Science Ltd, Amsterdam, 1999, pp39-50.

21. Maddox S J: 'Statistical analysis of fatigue data obtained from specimens containing many welds'. IIW draft document XIII-XV-122-94, April 1994.

22. Owen D B: 'Handbook of statistical tables'. Addison-Wesley, Reading, Massachusetts, 1962.

23. Erricker B C: 'Advanced general statistics'. ISBN 0 340 15178 1.

24. Johnson N L and Welch B L: 'Applications of the non-central *t*-distribution' *Biometrika* 1940 **31** pp362-389.

25. Resnikoff G J and Lieberman G J: 'Tables of the non-central *t*-distribution'. Stanford University Press, Stanford, California, 1957.

26. Kendall M G and Buckland W R: 'A dictionary of statistical terms'. Longman, New York, 1982.

Table 1 One-sided tolerance limit factors k for a normal distribution for $\gamma\%$ = 90%

Sample size n	Value of k for $P\%$ = 95%	Value of k for $P\%$ = 97.5%
2	13.090	15.586
3	5.311	6.244
4	3.957	4.637
5	3.401	3.983
6	3.093	3.621
7	2.893	3.389
8	2.754	3.227
9	2.650	3.106
10	2.568	3.011
11	2.503	2.936
12	2.448	2.872
13	2.403	2.820
14	2.363	2.774
15	2.329	2.735
16	2.299	2.700
17	2.272	2.670
18	2.249	2.643
19	2.228	2.618
20	2.208	2.597
21	2.190	2.575
22	2.174	2.557
23	2.159	2.540
24	2.145	2.525
25	2.132	2.510
30	2.080	2.450
35	2.041	2.406
40	2.010	2.371
45	1.986	2.344
50	1.965	2.320
60	1.933	2.284
70	1.909	2.257
80	1.890	2.235
90	1.874	2.217
100	1.861	2.203
120	1.841	2.179
145	1.821	2.158
300	1.765	2.094
500	1.736	2.062
∞	1.645	1.960

Appendix

STATISTICAL ANALYSIS OF FATIGUE DATA OBTAINED FROM SPECIMENS CONTAINING MANY WELDS

by S J Maddox

Welded specimens used to obtain fatigue data invariably contain more than one potential site for fatigue crack initiation. For example, in a simple butt weld there are four weld toes from which fatigue cracks could propagate. Specimens used to investigate the fatigue performance of attachments normally include more than one. In view of this situation, the fatigue life obtained from a test on a specimen containing *n* nominally identical welds, all of which might fail, is the lowest of *n* possible fatigue lives. Thus, this life is less than the average life which would have been obtained if each individual weld had been tested to failure. Similarly, the mean S-N curve obtained from regression analysis of the fatigue test results obtained from several welded specimen:

$$S^m N = A \tag{A1}$$

lies below that corresponding to failure in all the welds. Assuming that the fatigue lives are log normally distributed (as is usually found to be the case for welded joints) an estimate of the average life from this higher S-N curve can be obtained using extreme value statistics.

Order statistics enable an estimate to be made of the expected smallest value of a random sample of *n* observations drawn from a normally distributed parent population. Tabulated values are available to represent the mathematical expression used [A1,A3]. This actually relates to a parent normal distribution with zero mean and unit variance and gives the expected expressed in numbers of standard deviations (standard deviation = √ variance). An extract from the table is shown in Table A1; for a sample of, say, 5, the expected smallest value is -1.163, that is 1.163 standard deviations below the mean. The standard deviation of the smallest of '*n*' randomly selected observations from a distribution is less than that of a single observation. However, again, statistical tables are available[A2] to estimate it[1]; an extract is given in Table 2. Thus, for n = 5, the variance is 0.4475, whereas Table A1 relates to a variance of 1. Thus, the variance for the minimum of five samples is 0.4475 x 1 = 0.4475.

As an example, consider the fatigue test results obtained from tests on specimens incorporating three test welds. The minimum fatigue life from three observations is therefore known (log N_3) together with the standard deviation of log N for those observations, σ_3. Thus, the statistical method is used in reverse to infer the value corresponding to n = 1 (i.e. the mean life obtained from three times as many specimens each with a single weld, log N_1) and the corresponding standard deviation of log N, σ_1. Referring Tables A1 and A2 for n = 3, the expected deviation is 0.846 and $(\sigma_1)^2 = 0.5595$. Thus,

$$\log N_1 = \log N_3 + 0.846\sqrt{\frac{\sigma_3^2}{0.5595}} \tag{A2}$$

and the corresponding standard deviation of log N is

$$\sqrt{\frac{\sigma_3^2}{0.5595}}\,.$$

To illustrate the use of the technique, consider the fatigue test results obtained from 21 specimens each with three test welds given in Fig.A1. Regression analysis of all the data gave the equation:

$$S^{2.90}\,N = A \qquad [A3]$$

where A = 2.02 x 10^{12} as shown in Fig.A1, and the standard deviation of log N was 0.158. Since N is proportional to A, Eq.[A2] can be used directly to deduce the corresponding constant for the adjusted S-N curve for single welds:

$$\log A_1 = \log A_3 + 0.846\ \sqrt{\frac{(0.158)^2}{0.5595}} \qquad [A4]$$

$$= 12.306 + 0.846\ \sqrt{\frac{0.0251}{0.5595}}$$

$$= 12.485$$

or A_1 = 3.06 x 10^{12}

Thus, the equation of the adjusted mean S-N curve is $S^{2.90}$ N = 3.06 x 10^{12} as shown in Fig.A1, representing a 52% increase in fatigue endurance at a stress range of 100 N/mm^2. The standard deviation of log N has, however, now increased to

$$\sqrt{\frac{0.0251}{0.5595}} = 0.212.$$

Therefore, if an S-N curve some number of standard deviations below the mean was of interest the increase in fatigue endurance for single welds would be less. For example, for two standard deviations the increase is only 18%.

A further application of extreme value statistics is to deduce the average fatigue life of a structural member containing many welds any of which may fail. For example, for a member which incorporates 10 elements welded together in line, from Table A1 the **mean** fatigue life of such a member can be expected to be 1.539 standard deviations below the life obtained from any S-N curve deduced for single welds.

REFERENCES

A1 Fisher R A and Yates F: 'Statistical tables for biological, agricultural and medial research', Oliver & Boyd, Edinburgh, Sixth Edition, 1963.

A2 Pearson E S and Hartley H O: 'Biometrika tables for statisticians', Cambridge University Press, 1962.

A3 Beyer W H: ' Handbook of tables for probability and statistics', Second Edition, CRC Press, Florida, 1986.

Table A1 Extract from table of expected values of normal order statistics [A1]

n	2	3	4	5	10	20	30
Expected deviation	-0.564	-0.846	-1.029	-1.163	-1.539	-1.867	-2.04

Table A2 Extract from table of variances or order statistics [A2]

n	2	3	4	5	10	20
Variance	0.6817	0.5595	0.4917	0.4475	0.3433	0.2757

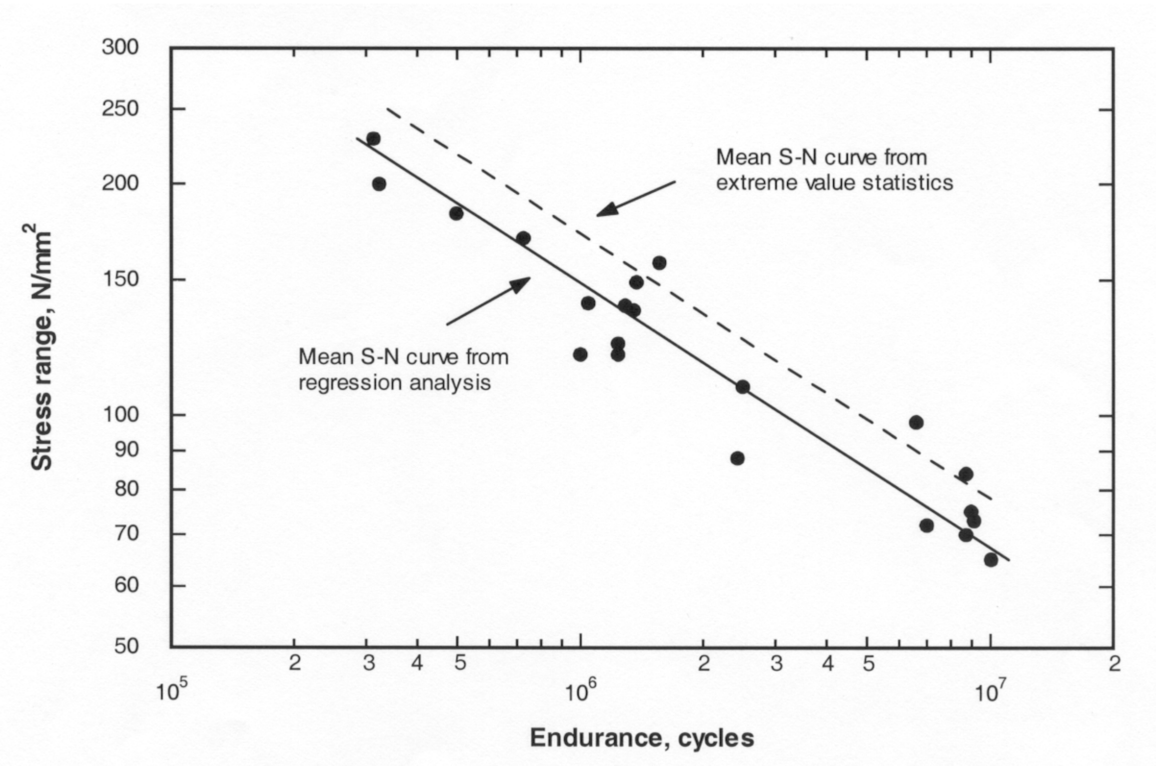

Figure A1 Fatigue data obtained from specimens containing three welds, only one of which failed, analysed using extreme value statistics