Sensitivity analysis of the staircase method to determine the fatigue limit

C. Blanco[1,a], A. Martín Meizoso[1,2,b]

[1] School of Engineering (TECNUN) University of Navarra, Pº Manuel Lardizábal, 13, 20018 Donostia-San Sebastián. Spain.

[2] Materials Department, CEIT, Pº Manuel Lardizábal, 15, 20018 Donostia-San Sebastián. Spain.

[a] cblanco@tecnun.es, [b] ameizoso@ceit.es

Keywords: fatigue, fatigue limit, staircase method, maximum likelihood, numerical simulation.

Abstract. A sensitivity analysis is carried out modelling the staircase method to determine the fatigue limit of a material. Using the statistical method of maximum likelihood and assuming a normal distribution for the fatigue limit, the mean and standard deviation are estimated. The effect of initial load, the number of experiments and the load step are analyzed, as well as the resulting error obtained by a numerical simulation procedure. It is concluded that 30 tests is probably too short for most applications because the expected error is 30% of the population standard deviation in the best possible scenario (whenever the step size actually matches the scatter of the population). On the other hand, more than 300 tests are not worthy, because the obtained precision is not appreciably increased. The best range for the step is between 0.2 and 1.5 times the standard deviation of the population. In any case, al least two load levels with a mixture of runouts and failures must turned out.

Introduction

For most nonferrous metals such as aluminum, copper and magnesium, S-N curves gradually drop off and failure will occur eventually if units are tested or are in service long enough. Fatigue data on ferrous and titanium alloys indicate that experimental units tested below a particular stress level are unlikely to fail. The S-N curve for these materials exhibits a strong curvature and an asymptotic behavior. This limiting stress level is called the fatigue limit which is defined as the maximum load (or stress) that a component may withstand without failure before a specified life (intended to be infinite in theory or -more practically- a given number of cycles ranging from 3 million to 10 million).

Fatigue limit is of crucial importance for most engineering applications subjected to alternating loads: railways, automotive, airplane structures and engines… It is determined by carrying out experiments with samples tested to different alternating loads. In last decades various statistical methods to evaluate fatigue limit have been developed [1,2]. The usual procedure, described in BS ISO 12107:2003 norm [3], is the staircase method whose sensibility is analyzed in this work. One alternating load is applied to the test-piece (or component) until a premature failure or runout is obtained. If a failure is produced before the specified number of cycles, then the following test-piece is tested to by-one-step reduced alternating load. If a runout is obtained the next experiment is run with one-step increased load. After a specific number of experiments the results are statistically analysed.

In order to obtain the required results, parameters like first load level, load step and number of tests have to be chosen properly. For example, if the first load level is too high (low) and the load step very small, the initial data points will likely be a string of failures (runouts). If the load step is too big, only two testing load levels will be obtained, producing the bigger one always failures and runouts the smaller. Concerning the number of experiments, it is obvious that reliability of experiment increases with the number of tests, but the same does the cost.

In the next section the procedure to analyze the sensitivity of the stair method respect of these three parameters is described. The model estimates the mean and the typical deviation of the fatigue limit. In any planned fatigue experiment, there is always some amount of scatter in the dada due to a variety of random factors. Some of them can be mitigated preparing the specimen carefully; there are others like the slight differences in the microstructure which are incontrollable. Since crack initiation is a microstructural phenomenon and, in high cycle regime, fatigue life is dominated by the crack initiation phase, for statistical design purpose the estimation of the deviation is strictly necessary. There are methods that evaluate the mean of fatigue limit by thermography [4, 5, 6], with the advantage of testing, theoretically, only one specimen. The disadvantage of these type of accelerated methods is the lack of scattering information. It could be useful take the result of one of these methods as a first load level.

Simulation procedure

A normal distribution of the fatigue limit for a given life (10^x cycles, where x might be 6 or 7, depending on the actual application) will be assumed. Other distributions are possible, but it is simpler and very frequent according to the limit central theorem: whenever the result depends on many independent variables, the obtained distribution tends to be normal. So, it is the usual case for the fatigue limit that depends on the material composition, homogeneity, grain size, defect sizes, inclusion sizes, environment, etc. In any case, a very large number of experiments would be required to be able to discern among different distribution functions.

With a generalization purpose, all the variables with be normalized by the population mean value and its standard deviation, i.e. normalized variable

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

will be used throughout this paper. Note that, in a simulation procedure, the actual distribution is N(0,1) and the computers run the simulated and random experiments. So, the answer is known and the results are analysed, and their ability to obtain the right answer examined.

The degrees of freedom are the total number of experiments, N (as few as possible compatible with the required precision and/or safety concerns or needs), the size of the load step Δz (usually in kN or MPa) and the initial load or stress z_0. A sensitivity analysis is performed on these variables.

Whether a failure or a runout is simulated depends on the cumulative failure probability for the assumed (true) distribution at this load (or stress) level, z_i

$$F(z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} \exp\left(-\frac{s^2}{2}\right) ds . \tag{2}$$

Note that the cumulative probability is uniformly distributed between 0 and 1. Now, a pseudo-random number (uniformly distributed between 0 and 1) a, is generated and compared with (2). If $a < F(z_i)$ failure is assumed, otherwise a runout is anticipated.

The simulation runs again increasing or decreasing the load by the given load step Δz, until the total number of desired experiments, N, are simulated.

Fig. 1 shows the results of a simulation. It begins at a load level $z_0 = 2$. Starting with large loads saves time: it is faster to produce a failure than a runout, so usually a large load level is chosen to start with. The load step chosen is $\Delta z = 1$.

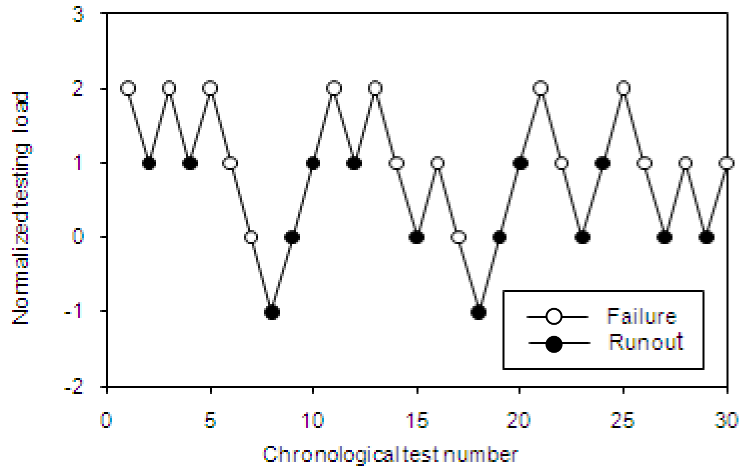


Fig.1. Typical simulation sequence of 30 tests in accordance with the staircase method ($z_0 = 2$, $\Delta z = 1$).

Eventually, a number n of different load levels z_i ($i = 1, \ldots n$) have been tested, each one with a different numbers of failures $f(i)$, and runouts $r(i)$. Table 1 summarizes these results for the simulation corresponding to Fig. 1.

Table 1. Results obtained in the simulation corresponding to Fig. 1.

Tested load level, *z*	**Failures, *f***	**Runouts, *r***
-1	0	2
0	2	6
1	7	6
2	7	0
	Total = 16	**Total = 14**

Maximum likelihood method [7, 2] is used to derive the estimated population mean μ, and standard deviation, σ. The likelihood for the previous sequence is given by

$$L(\mu,\sigma) = \prod_{j=1}^{n} \left[F\left(z_j;\mu,\sigma\right)\right]^{f(j)} \left[1 - F\left(z_j;\mu,\sigma\right)\right]^{r(j)} \tag{3}$$

where

$$F\left(z_j;\mu,\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z_j} \exp\left[-\frac{(s-\mu)^2}{2\sigma^2}\right] ds\,. \tag{4}$$

Most frequently, instead of the likelihood, its logarithm is maximized (note that the logarithm is a monotonic function on the argument). In this way, numeric computational problems tend to reduce, mostly when the data and exponents in (3) are very large. The new function to optimize is

$$\log L(\mu,\sigma) = \sum_{j=1}^{n} \left\{ f(j) \log\left[F\left(z_j;\mu,\sigma\right)\right] + r(j) \log\left[1 - F\left(z_j;\mu,\sigma\right)\right] \right\} \tag{5}$$

where "log" holds for natural or decimal logarithms. In this paper natural logarithms are used.
Fig. 2 shows the summary of the simulation failure frequencies as a function of the applied loads (black circles) and their confidence interval (corresponding to a binomial distribution) as gray segments. The estimated population distribution in thick red line passes as close as possible to the experimentally observed frequencies (maximizes its likelihood). In thin dashed line the true population used for the simulation is also shown.

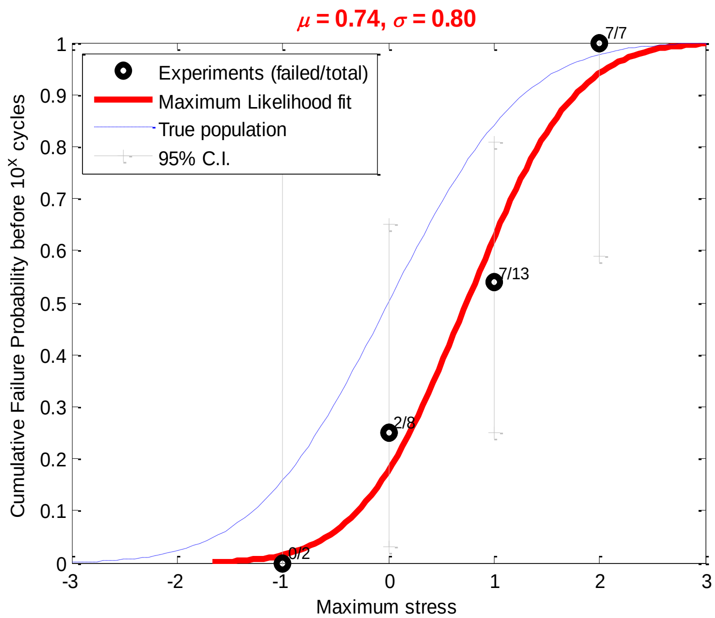


Fig. 2. Cumulative distribution function estimated from simulation reported in Fig. 1 in thick red line. The true distribution is shown as a thin dashed line.

In this particular example the final result overestimates the population mean by a 74% (of the standard deviation, our standard units) and the population scatter is slightly underestimated (0.8 vs. the actual 1.0). The mean shift might be due to the fact of beginning with large loads. It is clear that a better precision is desired and, in a lot of cases, required.

Results and discussion

Effect of the number of tests. The BS ISO 12107 norm [3] states that a minimum of 30 tests should be carried out, as it was done in the example shown in Fig. 1 and Fig. 2. If the number of test is increased the results shown in Table 2 are obtained.

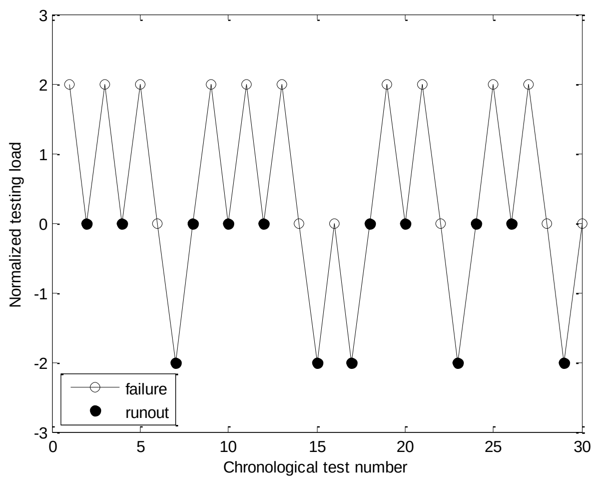
Table 2. Results for simulation with different number of tests (z_0 = 2, Δz = 1).

	N = 30	**N = 100**	**N = 300**	**N = 1000**
μ	0.74	0.27	0.02	-0.03
σ	0.80	0.91	0.93	1.06

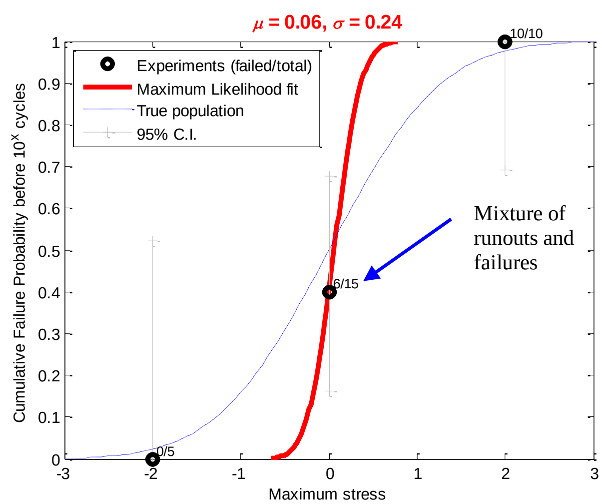
With 100 tests, the mean value for the population is overestimated by a 27% and the standard deviation is underestimated by a 9%. Probably it is the right number of experiments for most applications. The results for N = 1000 are not better than those for only N = 300 (errors are in the opposite directions and of comparable amounts), but 300 tests is indeed a number too large and the campaign expensive.

Analysis of load (or stress) step. Table 2 shows the evolution of one particular prediction by increasing the number of experiments, for a constant and optimized load step (identical to the standard deviation of the sampled population, as recommended by BS ISO 12107 norm [3]). Now, let us analyze the influence of the load step in the estimated results.

Fig. 3(a) shows the simulated sequence with a larger load step (twice the recommended), and Fig. 3(b) the estimated normal distribution. Again a larger load is chosen to start with and N = 30.



(a)



(b)

Fig. 3. Simulation with N = 30, z_0 = 2 and Δz = 2. (a) Sequence of tests. (b) Estimated normal distribution.

With a very large load step, we only test at three load levels and the results at the extreme loads are all runouts and all failures. There is only one load level with a mixture of runouts and failures. This fact is the reason of the wrong estimation of the population scatter.

Fig. 4 shows the huge improvement obtained for N = 100 tests.

The main difference is because now two load levels with failure frequencies different of 0 or 1 are obtained: z = -2 with 1 failure and 27 runouts and z = 0 with 28 failure and 11 runouts. Even when the failure frequency observed for the minimum stress range is very close to zero (1/28) but not zero.

Table 3 summarizes the results obtained for different initial loads (+2σ, 0 and -2σ), step sizes and number of tests. Light gray cells show when there are two or more significant load levels (those with a mixture of failures and runouts). In darker gray when the relative error in mean and standard deviation estimated are less than 20%.

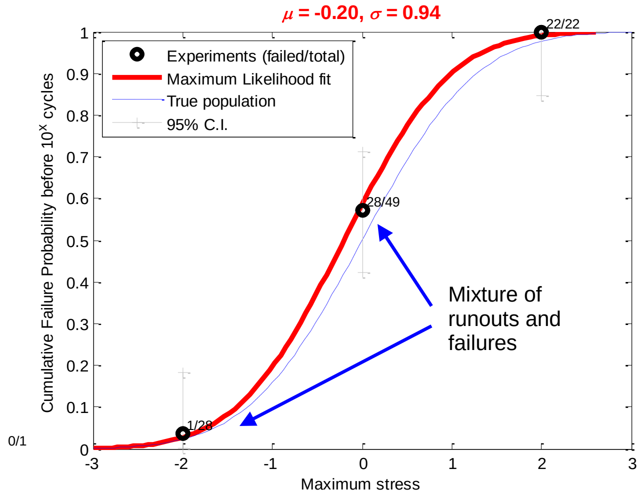


Fig. 4. Estimated normal distribution for a simulation. (N = 100, z_0 = 2 and Δz = 2).

Table 3. Properties of sequence tests and normal distribution for different first loads, load step and number of experiments.

$z_0 = 2$

N / **Δz**	**10**	**30**	**100**	**300**	**1000**
0,10					
0,15					
0,20					
0,25					
0,50					
0,75					
1,00					
1,25					
1,50					
2,00					
2,50					
3,00					
3,50					
4,00					

$z_0 = 0$

N / **Δz**	**10**	**30**	**100**	**300**	**1000**
0,10					
0,15					
0,20					
0,25					
0,50					
0,75					
1,00					
1,25					
1,50					
2,00					
2,50					
3,00					
3,50					
4,00					

$z_0 = -2$

N / **Δz**	**10**	**30**	**100**	**300**	**1000**
0,10					
0,15					
0,20					
0,25					
0,50					
0,75					
1,00					
1,25					
1,50					
2,00					
2,50					
3,00					
3,50					
4,00					

It is quite clear than without at least two significant load levels the chances to estimate the true normal distribution are very poor. It has been checked that more than two significant load levels do not improve the estimation in a significant way. It is due to the assumption about the distribution: three or more significant load-levels do not define better the normal fit; it might be useful is the kind of distribution –any not normal- were under question.
Observe that, with less than 100 experiments, only three estimations were obtained with errors less than 20%.

The lower bound of the fatigue limit
In above section the mean of fatigue limit has been estimated. For design purpose, a safe load (lower bound of fatigue limit) has to be taken. According to Benard [8,9], the load with a failure probability of 0,001 is considered as lower bound, i.e.

$$\tilde{z}_i = \mu_i - 3.0902\sigma_i \,. \qquad (6)$$

Fig. 5 shows the cumulative frequency for the lower bound after analysing 1000 runs each one with $N = 30$, $z_0 = 1$ and $\Delta z = 1$. The vertical, black, dashed line represents the true position of the lower bound. Note that the median of lower bound estimated for 1000 runs is bigger than the solution -3.0902 -which is not conservative in the practice-, and the range of distribution is very big (from -6σ to 0). As it could be expected, if the number of tests is increased, the values estimated for the lower bound improve (Cfr. Fig. 6).

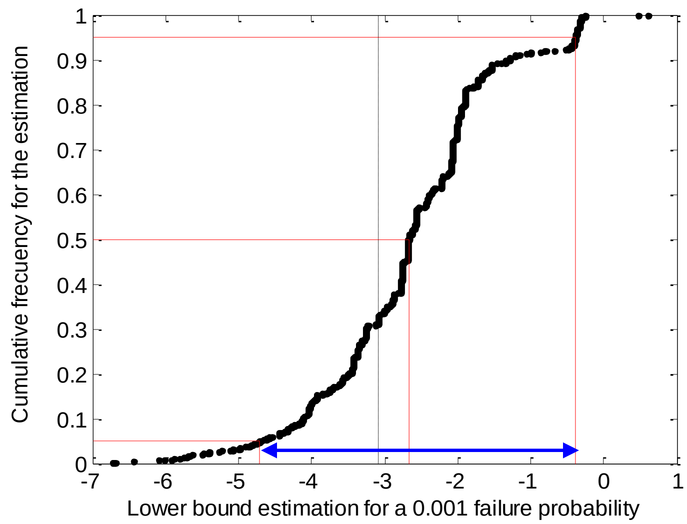


Fig. 5. Distribution of lower bound for 1000 runs.
($N = 30$, $z_0 = 1$ and $\Delta z = 1$).

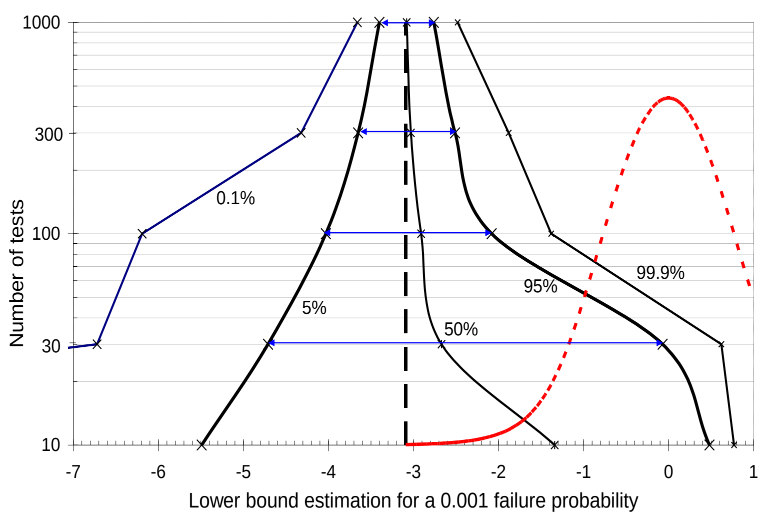


Fig. 6. Distribution of lower bound as a function of the number of tests.
($z_0 = 1$ and $\Delta z = 1$).

For illustrating purposes, the dashed red line in Fig. 6 represents the true normal distribution used in the simulations (in a different ordinate scale, not represented). Note the dispersion of lower bound for $N < 100$ (blue horizontal line between 5% and 95%), and the not conservative bias predicted for the median estimation (labelled 50%), if the number of experiments is very small.

Summary

A statistic model has been conducted to analyse the sensitivity of the staircase method to determine the fatigue limit. In particular, from the study of the influence of the number of tests, the initial load and the load step, the following conclusions can be drawn:

The number of 30 tests per campaign is probably too short for most applications because the expected error is 30% of the population standard deviation even when the load step actually matches the scatter of the population, as the norm recommends. Moreover, the lower bound in this case becomes not reliable as it is not conservative. On the other hand, more than 300 tests are not worthy, as the obtained precision is not appreciably increased.

The best range for the load step is between 0.2 and 1.5 times the standard deviation of the population. In any case, al least two load levels with a mixture of runouts and failures must be tested to get an accurate estimation. In most of other cases standard deviation of the population is grossly underestimated.

Acknowledgments

The authors acknowledge the financial support from the Spanish Ministry of Science and Technology (Ref: MAT2008-03735/MAT) and form the Basque Country (Ref: PI09-09).

References

[1] Pollak, R.D. “Analysis of Methods for Determining High Cycle Fatigue Strength of a Material with Investigation of Ti-6A1-4V Gigacycle Fatigue Behavior”. Air Force Institute of Technology, 2005.

[2] Castillo, E. and Fernández-Canteli, A. (2009) *A unified Statistical Methodology for Modeling Fatigue Damage*. Springer.

[3] BS ISO 12107:2003, “Metallic materials –Fatigue testing- Statistical planning and analysis of data”, 2003, British Standards.

[4] La Rosa, G. and Risitano, A., “Thermographic methodology for rapid determination of the fatigue limit of materials and mechanical components”. *International Journal of Fatigue* 22 (2000) 65-73.

[5] Gómez, C., Canales, M., Calvo, S. and Núñez, J.M. “Detección de daño en acero estructural sometido a diferentes estados de carga mediante termografía por infrarrojos”, *Anales de Mecánica de la Fractura* 25 Vol II, (2008), p. 660.

[6] Canales, M.,Calvo, S., Royo, J.M., Núñez, J.L., Iatza, I. and Landaberea, A. “Material constants of multiaxial models base don stresses obtained forma accelerated method”. *Anales de Mecánica de la Fractura* 29, Vol I, (2012) 313-318

[7] Spindel, J.E. and Heibach, E., “The method of maximum likelihood applied to the statistical analysis of fatigue data”, *International Journal of Fatigue* 1 (1979) 81-88.

[8] Benard, A. and E.C. Bosi-Levenbach, Statistica Neerlandica, Vol. 7 (1953) p. 163. As quoted by [4].

[9] Abernethy, R.B., “The New Weibull Handbook”, 4th edt. Robert B. Abernerthy, USA, 2000.