

S-N Curve Modeling for Finite Life Range under the Assumption of Linearly Changing Scatter

Zhou Yingjie

Schaeffler Holding (China) Co., Ltd.

Tang Xuezhi

Schaeffler Trading (Shanghai) Co., Ltd.

ABSTRACT

Fatigue test data in finite life region are normally analyzed separately for each stress level because of varying scatters, while another method assuming equal scatters on all levels and thus unifying all data into one model is also welcome for its efficient usage of data. Combining the idea of the two methods, a new method that assumes linear change of scatter according to stress levels is developed in this paper. The algorithm derives from maximum likelihood estimation and general Newton's method. Monte Carlo method is used for theoretical validation and the new method is proved to be correct with the estimations. Finally a set of test data which shows a tendency of trapezoidal data distribution is analyzed and an S-N curve for finite life range is created. The new method has its advantage in certain situations so it's meaningful to add it to the alternatives of fatigue data analysis methods.

CITATION: Yingjie, Z. and Xuezhi, T., "S-N Curve Modeling for Finite Life Range under the Assumption of Linearly Changing Scatter," *SAE Int. J. Mater. Manf.* 7(2):2014, doi:10.4271/2014-01-0970.

INTRODUCTION

A large portion of industrial products work under alternating force or random vibration, e.g. most parts on an automobile. It's necessary to do fatigue test on these parts to analyze their degradation property and thus to predict their life. Fatigue test is a kind of durability test which requires lots of time and testing resources. Since new development or product modification comes up frequently in current days, it's not possible to do thorough fatigue life researches for all of them. In order to get as much information as possible from limited test data, numerical methods are necessary.

S-N curve is the most popular model to describe fatigue life for metal parts. For model simplification, an S-N curve is usually divided into 3 regions: long life range where no failure occurs, finite life range where all samples run until failure, and transition area that contains both types[1, 2, 3, 4], see Figure 1. The analysis for transition area is relatively simple, e.g. staircase method ignores the cycle number and considers the data either as failure or non-failure. But the model in finite life region describes the failure probability as a function of both stress(or force) amplitude and cycle number. The complexity brought up different analysis methods for finite life range modeling.

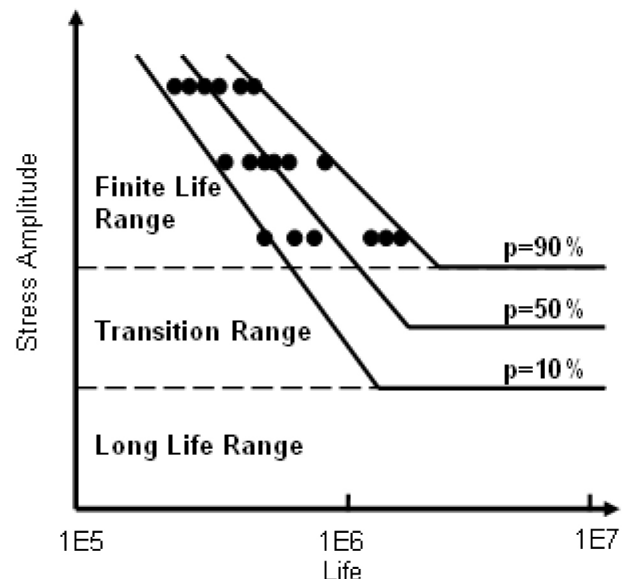


Figure 1. Simplification of S-N Curve

SEPARATED AND COMBINED METHOD

The commonly used separated and combined method are both suggested by various international test standards.

According to the separated Method, fatigue tests are carried out on several stress levels. Data on each level are assumed to follow Gaussian distribution. The failure probability of each point is estimated with Rossow-estimator. Then mean and standard deviation of the distribution are calculated through linear regression. For detail please see related references [1, 2, 3, 5, 6].

The separated method analyses each stress level separately, and requires for each level at least 6 samples. The reliability improves with more samples.

Combined method assumes

1. The center points(mean of the failure distribution) of each level lie on a straight line.
2. The failure probability on each level has the same standard deviation. The confidence bounds, e.g. 10% and 90%, of different levels lie parallel to the center line(under logarithmic scale).

First the slope is calculated through linear regression with all data points. Then they are transformed to a reference level (any stress level is OK) according to the slope, see Figure 2. Since the scatter is assumed constant, transformed data follow the same distribution. Then the same regression method used in the separate method is used to calculate mean and standard deviation. For detail please see related references[2,3,7, 8, 9].

Each data point is used not only for the estimation of the level it belongs to, but also its statistical information is contributed to other levels. With increased efficiency on data usage, it's possible to further reduce the sample number.

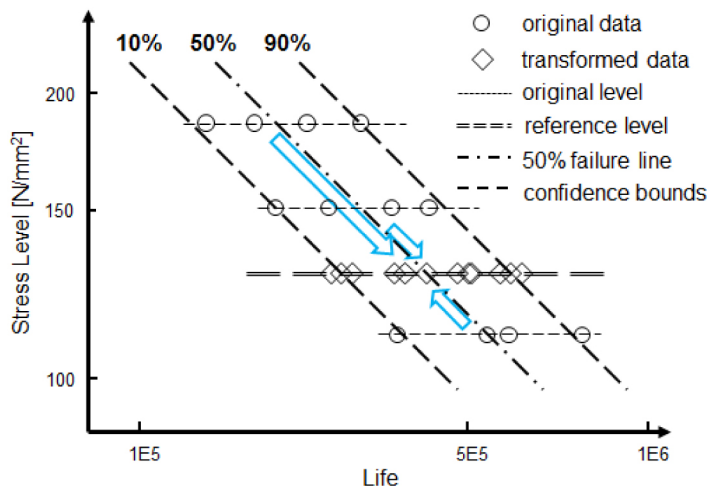


Figure 2. Combined Method

LINEAR-SCATTER METHOD

The separated method ignores the usually valid relationship among different levels, e.g. means of all levels lie on a straight line. The statistical information of data is not effectively used in modeling. The combined method has a higher efficiency but it

can be further optimized because there's actually no evidence to support a always-equal-scatter assumption. In fact, failure probability distribution often has larger scatters on lower levels [4,5,9,10]. So it's meaningful to develop a more flexible model.

Assumption of Linearly Changing Scatter

First, the straight center line assumption is still regarded as correct. Even in the analysis with the separated method, means on different levels are often fitted to a straight line for finite life range. This assumption is the basic way to raise the efficiency and should be kept.

The new model differs from the current ones in the assumption of scatter. It assumes that the standard deviations changes linearly according to the stress level, see Figure 3.

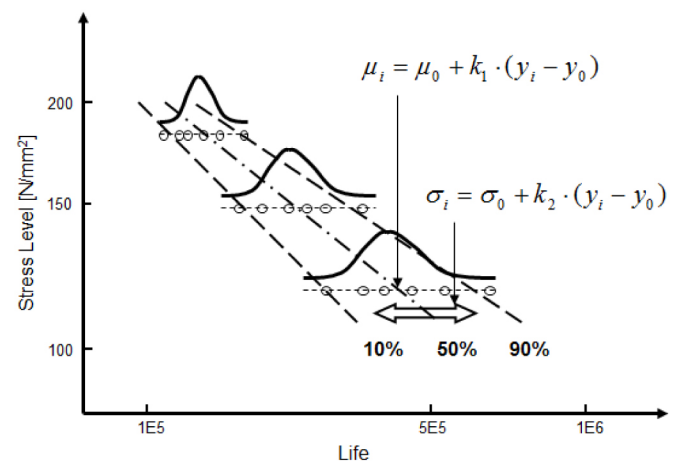


Figure 3. Assumption of Linearly Changing Scatters

Derived from the basic function of Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ a general failure probability function for the new model is formulated as}$$

$$f(x_i, y_i) = \frac{1}{\sqrt{2\pi}(\sigma_0 + k_2(y_i - y_0))} \cdot \exp\left(-\frac{1}{2} \left(\frac{x_i - \mu_0 - k_1(y_i - y_0)}{\sigma_0 + k_2(y_i - y_0)}\right)^2\right)$$

(1)

μ_0 mean of the reference level

σ_0 standard deviation of the reference level

k_1 inversed slope, $\Delta\mu/\Delta y$ in Figure 3

k_2 linear change speed of scatters, $\Delta\sigma/\Delta y$ in Figure 3

x_i logarithmic value of failure cycle

y_i logarithmic value of stress

In order to describe a multi-level model, a reference level is necessary. Like the one used in the combined method, any level is OK.

A model that supports linear variation of scatters has a better chance to fit actual situations than a constant-scatter model which is used by the combined method. In fact, the former can be regarded as an extended version of the latter. More complicated functions may also be suitable to describe the scatter, but they lead to a huge complexity and are not considered here.

Equation (1) has 4 variables and is nonlinear. Linear regression is no longer effective, so maximum likelihood estimation (MLE for short) is used. According to MLE, the log-likelihood function is defined as

$$L = -\frac{n}{2} \ln 2\pi + \sum_{i=1}^n (-\ln(\sigma_0 + k_2(y_i - y_0))) - \frac{1}{2} \left(\frac{x_i - \mu_0 - k_1(y_i - y_0)}{\sigma_0 + k_2(y_i - y_0)} \right)^2 \quad (2)$$

Iterative Calculation

The solution satisfying MLE standard should make the partial derivatives of Equation (2) with respect to four variables equal zero.

$$\frac{\partial L}{\partial \mu_0} = \sum_{i=1}^n \frac{x_i - \mu_0 - k_1(y_i - y_0)}{(\sigma_0 + k_2(y_i - y_0))^2} = 0 \quad (3)$$

$$\frac{\partial L}{\partial \sigma_0} = \sum_{i=1}^n \left(-\frac{1}{\sigma_0 + k_2(y_i - y_0)} + \frac{(x_i - \mu_0 - k_1(y_i - y_0))^2}{(\sigma_0 + k_2(y_i - y_0))^3} \right) = 0 \quad (4)$$

$$\frac{\partial L}{\partial k_1} = \sum_{i=1}^n \frac{x_i - \mu_0 - k_1(y_i - y_0)}{(\sigma_0 + k_2(y_i - y_0))^2} (y_i - y_0) = 0 \quad (5)$$

$$\frac{\partial L}{\partial k_2} = \sum_{i=1}^n \left(-\frac{1}{\sigma_0 + k_2(y_i - y_0)} + \frac{(x_i - \mu_0 - k_1(y_i - y_0))^2}{(\sigma_0 + k_2(y_i - y_0))^3} \right) (y_i - y_0) = 0 \quad (6)$$

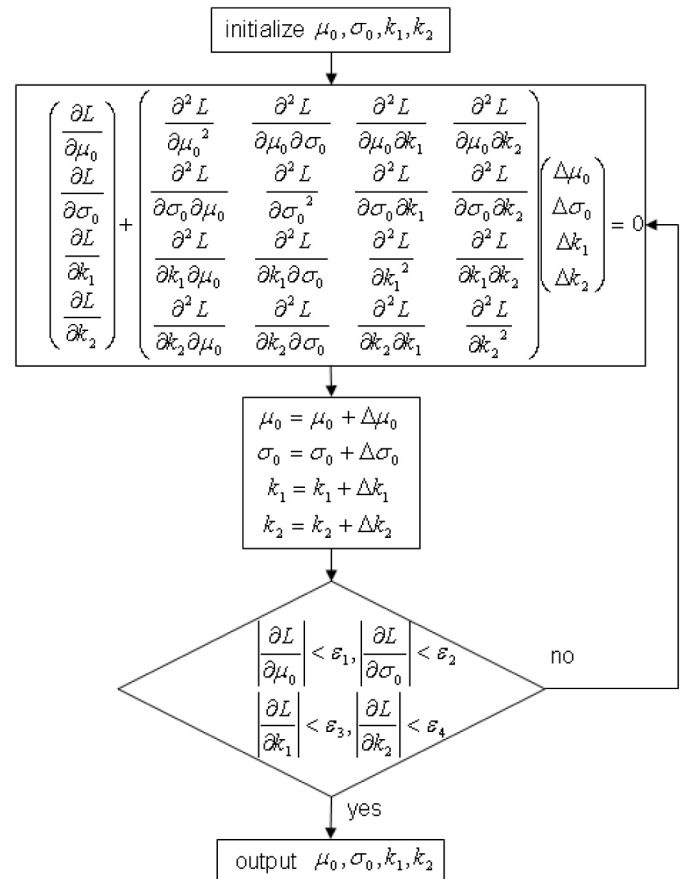


Figure 4. Solving MLE-Equations with General Newton's Method

The equations include summation and the solution is implicit. Newton's iterative method is used to solve the equations, see Figure 4.

The solution from MLE has smaller standard deviation, especially when the sample number is small. So a correction is necessary.

After attempts with various types of existing mathematical formula, the form of Bessel's correction is considered as suitable.

$$s = \sqrt{\frac{n}{n-p}} \cdot \sigma \quad (7)$$

s estimated standard deviation

σ standard deviation

n number of samples

n-p degrees of freedom

For linear models, p equals the number of independent variables. But it's hard to be determined for a nonlinear model, so an empirical value ($p=3$) is selected with the help of regression and Monte Carlo simulation. This correction is proved effective in the scatter analysis in the next section.

With a common computer and any programmable software, the calculation can be completed without perceivable delay to achieve a precision of 0.1-1% of absolute values. Initial values are necessary for iteration. In application, a rough subjective estimation from the data distribution is a proper way to determine them. In other cases, like the automatic process of huge amounts of virtual data in Monte Carlo simulation, the initial values of μ_0 , σ_0 and k_1 can be calculated through the combined method, while 0 is proved as a proper initial value for k_2 .

THEORETICAL VALIDATION

On the practical side, huge test data and long time of product tracking maybe necessary to check if a model is good. In contrast, validation with computer generated data based on known data distribution is efficient, cost-saving and data-resourceful, which makes it the main validation method in some statistical engineering works such as [11,12].

Monte Carlo Simulation

Monte Carlo method is used to check whether the newly developed algorithm gives correct estimations to the virtual data which is distributed according to the assumption of linearly changing scatter.

Random value generating functions (Randomize & Rnd from VBA in this paper) are used to create virtual data between 0 and 1. These values are then transformed through inverse cumulative Gaussian distribution to simulate the failure mode. The result forms a typical bell-shape, see Figure 5.

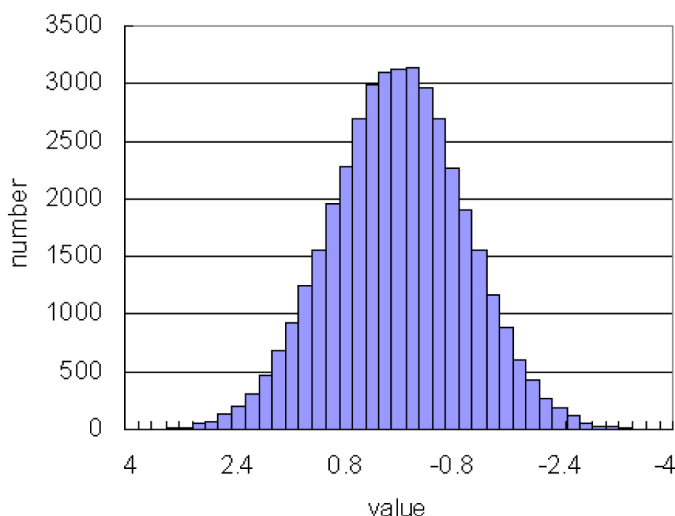


Figure 5. Monte Carlo Simulation for Gaussian Distribution

According to the linear-scatter assumption, data from the first step are distributed to several levels.

$$\mu_i = \mu_0 + k_1(y_i - y_0) \quad (8)$$

$$\sigma_i = \sigma_0 + k_2(y_i - y_0) \quad (9)$$

μ_i mean of the i -th level

σ_i standard deviation of the i -th level

Different sample numbers are simulated from 6 to 40. 6 to 8 samples per level are most common in application and larger numbers are used to check the convergence property. Also different level numbers are tried, but since their results show similar tendencies, only those of the 5-level simulation are shown due to limited space in figures.

Mean and standard deviation of distribution (m and s for short) are estimated and recorded. For each sample number, 60,000 simulations are run and the mean and standard deviation of each estimated variable (mean of m , s ; standard deviation of m , s) are calculated and compared.

Mean of the estimated variable shows whether the estimations lie around the actual value. If mean of m/μ or s/σ equals 1, the central limit theorem is satisfied and the method is statistically correct.

Standard deviation of the estimated variable shows how concentrated the estimations are around their mean. A smaller standard deviation means better concentration and thus more stable estimation for each single data set.

Since the new method contains more variables, only m , s are evaluated so the result is comparable.

Analysis of Mean

Figure 6, 7, 8 show that mean of m from three methods lie very close to the actual value, meaning all of them are correct with the estimation of m .

With mean of m verified, smaller standard deviation means better reliability of each estimation. The standard deviation of m from the linear-scatter method is nearly identical with the combined method, and both are smaller than the separated method, see Figure 9, 10, 11.