TITLE GOES HERE AND WILL BE AUTOMATICALLY SET IN ALL CAPS

Draft of March 22, 2021 at 23 : 24

BY

MIN-HSIU HSU

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Mechanical Science and Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2021

Urbana, Illinois

Adviser:

Assistant Professor Chenhui Shao

Abstract

To my parents, for their love and support.

Acknowledgments

Table of Contents

List of Tables

List of Figures

LIST OF ABBREVIATIONS

Chapter 1

Introduction

Remanufacturing has been considered an important role in reducing energy consumption and environmental pollution, and

Chapter 2

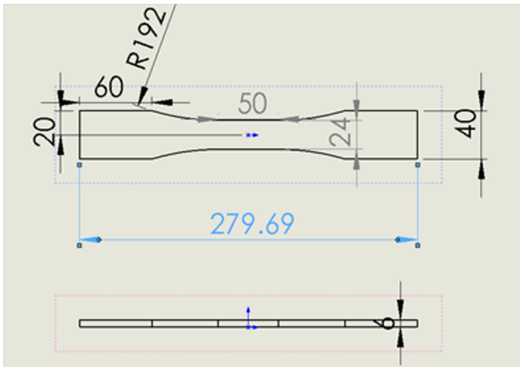
Experimental Dataset

This chapter describes the dataset and the experimental procedure used in this research for developing machine learning applications. First, the cyclic fatigue testing was conducted till the fracture of a specimen to acquire the fatigue characteristics of a material. Second, to mimic the scenarios in the remanufacturing industry, interrupted fatigue testing was utilized to produce specimens at different fatigue levels as a representation of end-of-life products. Then, linear and nonlinear ultrasound measurements are used to evaluate the fatigue damage of those specimens stopped at the predetermined number of cycles in the interrupted fatigue test. Besides, the residual stress and full width at half maximum data from X-ray diffraction are also presented.

2.1 Life cycle fatigue testing

The life cycle fatigue testing aims to collect fatigue life data to understand the fatigue behavior of our targeted material. The fatigue life of a material is defined as the total number of cycles that a material can sustain under a specified loading condition. In order to develop the S-N curve of a material, the material is tested at different loading stress amplitudes, and the fatigue test is repeated multiple times for each loading stress amplitude to account for the variance of fatigue life.

The fatigue testing in this research is led by Prof. Li's group at the Penn State University. The targeted material is 5052-H32 aluminum alloy which is widely used for car body construction in the automotive industry. Figure 2.1 shows the dimension of the specimen and the test machine. Three loading amplitudes, 11.7, 12.7, and 14.7 kN for the cyclic fatigue testing are selected to develop the S-N curve which is shown in Figure 2.2.



(a) Schematic of the 5052-H32 aluminum alloy specimen

(b) MTS 100KN Landmark fatigue testing system at Prof. Jingjing li's lab

Figure 2.1: Life cycle fatigue testing setup

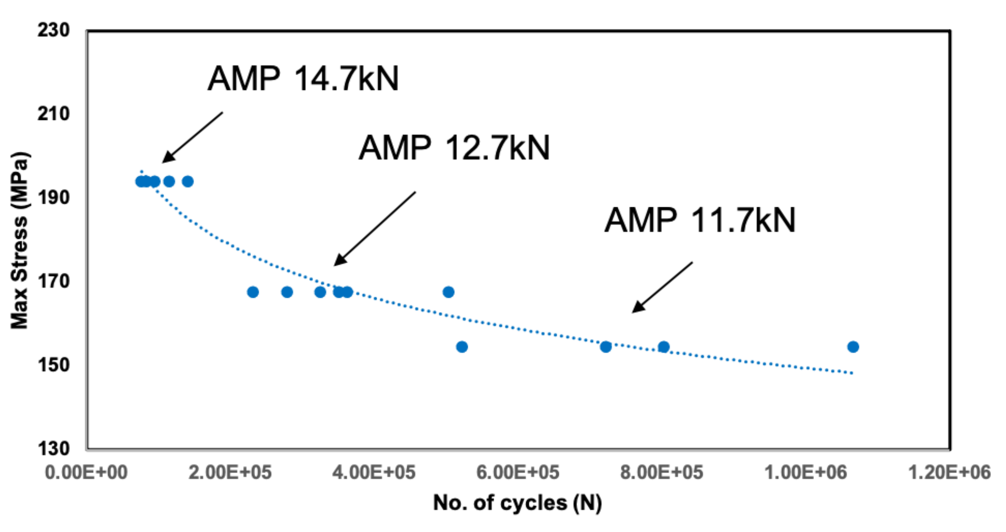


Figure 2.2: S-N curve for 5052-H32 aluminum alloy

2.2 Interrupted fatigue testing

The purpose of performing interrupted fatigue testing is to produce specimens at various fatigue levels by stopping the testing at several predetermined number of cycles. Considering the material cost and the time spent, the number of cycles applied to the specimens is set to be two levels, 33% and 67% fatigue life corresponding to the three loading amplitudes, 11.7, 12.7, and 14.7 kN. These specimens are used to represent the end-of-life products having different fatigue damage levels from the remanufacturing industry. Besides, three specimens without going through fatigue testing, i.e., 0% fatigue life, are included as specimens at the healthy state. The summary of

Table 2.1: Summary of the interrupted fatigue testing specimens

Specimen ID	Loading Amplitude (kN)	Percentage of Fatigue Life (%)	Max Stress Applied (MPa)
1	11.7	33	176
2	11.7	33	176
3	11.7	67	176
4	11.7	67	176
5	12.7	33	195
6	12.7	33	195
7	12.7	67	195
8	12.7	67	195
9	14.7	33	221
10	14.7	33	221
11	14.7	67	221
12	14.7	67	221
13	–	0	–
14	–	0	–
15	–	0	–

the interrupted fatigue testing specimens is presented in Table 2.1

2.3 Linear and nonlinear ultrasound measurements

In this research, linear ultrasonic (LU) and nonlinear ultrasonic (NLU) testing serve as the two main NDE methods for measuring the accumulated fatigue damage in the specimens. The ultrasonic testing is led by Prof. Matlack's group, and the testing system is shown in Figure 2.3. The LU and NLU measurements are both 1-D time domain signals, but the two approaches differ based on different theories and parameters, e.g., excitation wave shape, frequency, amplitude. Examples of LU and NLU signals are presented in Figure 2.4.

LU and NLU measurements were collected at nine locations in a specimen as illustrated in Figure 2.5, and each location was measured three times to understand the measurement repeatability. As a result, for each specimen, there are $9 \times 3 = 27$ signal profiles produced.

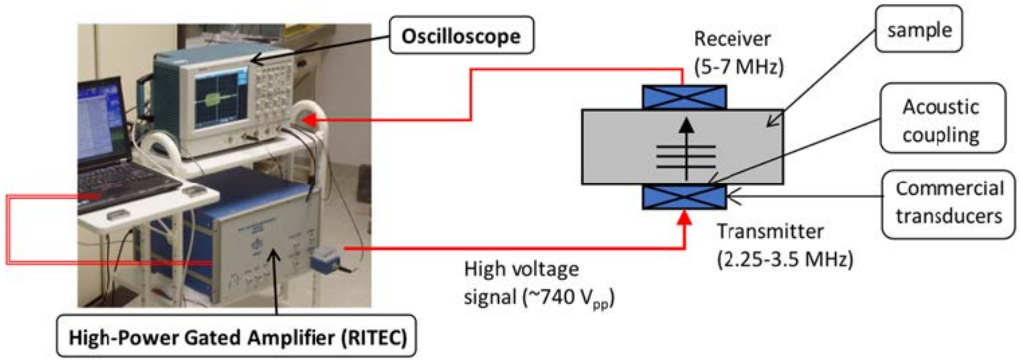
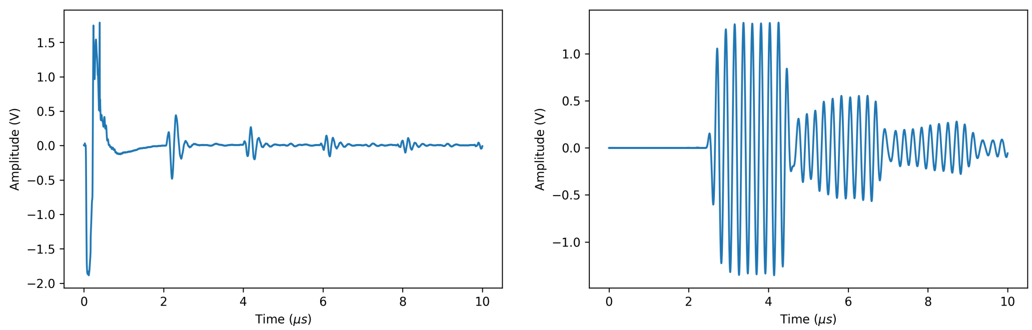


Figure 2.3: Experimental setup for LU and NLU at Prof. Matlack's Lab



(a) Linear ultrasonic signal

(b) Nonlinear ultrasonic signal

Figure 2.4: Examples of linear and nonlinear ultrasonic signals

2.4 X-ray diffraction measurement

Another quantity of interest, residual stress, is measured by X-ray diffraction (XRD) in this research. Residual stress is known to be associated with fatigue behaviors such as crack initiation and propagation. Besides, the full width at half maximum height (FWHM) of the diffraction peak in XRD is also extracted. Prof. Li's group performed the XRD measurements for a subset of specimens in the interrupted fatigue testing. The XRD data is used in the regression tasks in Chapter 5 as target variables.

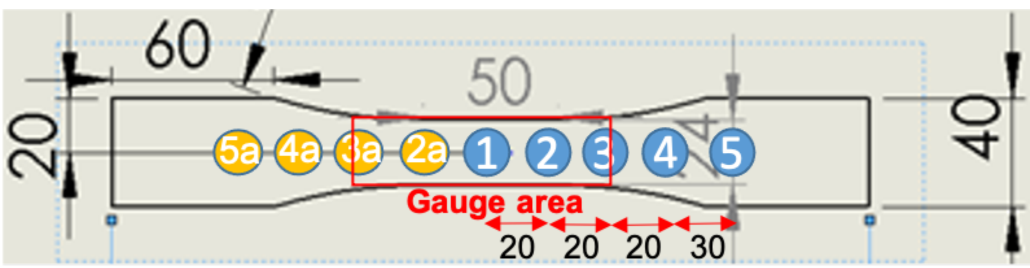


Figure 2.5: Schematic of the measurement locations for LU and NLU measurements. (The unit of length is in mm)

Chapter 3

Model Development

This chapter introduces a model development procedure used in both classification tasks in Chapter 4 and regression tasks in Chapter 5. The procedure involves: *a*) signal pre-processing, *b*) feature generation, *c*) feature selection, *d*) model training, *e*) model validation, and *f*) hyperparameter tuning, as shown in Figure .

3.1 Signal pre-processing

It is essential to reduce noises and extract regions of interest in signals by signal processing before we perform other analyses. Figure presents this process. First, DC bias was removed by subtracting the mean amplitude of a signal to prevent models from fitting on bias. Second, considering the computational cost from the high resolution data, we choose to downsample the ultrasonic signals. Third, we define the region or interest as the interval which contains the ultrasonic signal responses, and thus the other parts of a signal are discarded so that redundant information is not included.

3.2 Feature generation

Since ultrasonic sensor signals are unstructured, which is difficult to process, feature extraction methods are needed to create a representative set of values, i.e., features that aggregate the information from an entire signal. In this stage, physics-based and data-driven features are generated. The hybrid feature pool enables us to incorporate both physics knowledge and data-driven information into models.

3.2.1 Physics-based features

Given that physics modeling is built on theories or comprehensive experiment studies, physics-based features are robust, explainable, and suitable for applications having limited amounts of data such as the fatigue testing data in this research. Therefore, features from traditional LU and NLU testings become potential candidates for the model.

- Wave velocity

  In LU testing, ultrasonic wave velocity is a stiffness based measure which is associated with macroscopic damage such as crack/void coalescence and propagation. The wave speed is the distance divided by the time-of-flight (TOF) that a ultrasonic wave transverses in the material, as shown by Equation (3.1)

  $$v = \frac{2D}{\Delta t} \tag{3.1}$$

  where wave velocity is denoted by v, and D is the thickness of the specimen. Δt is the time difference between the actuation pulse and the response signal. Notice that, in our LU testing setup, one transducer severs as both the transmitter and receiver. Thus, the excitation signal travels $2D$ and the phase is changed $180°$ when received.

- Nonlinear acoustic parameter β

  While wave velocity from LU testing is able to detect fatigue damage at macro-scale, it is limited because it cannot detect defects much smaller than the probing wavelength, e.g., 1mm. In contrast, NLU techniques are based on a different physical principle: nonlinear elasticity from nano- and micro-scale defects induce harmonic generation. The nonlinear acoustic parameter is related to the amplitude of generated harmonics. This nonlinear parameter changes due to defects such as dislocations, local plastic strain, precipitates, and micro-cracks, all of which are orders of magnitude smaller than the probing wavelength. Here, we apply fast Fourier transform (FFT) to a NLU measurement and simply calculate the nonlinear parameter by using the ratio between the amplitudes of the fundamental and the harmonic waves given by Equation (3.2)

$$\beta = \frac{A_2}{A_1} \tag{3.2}$$

where A_1, A_2 is the amplitude of the fundamental wave and the second-order harmonic wave, respectively.

3.2.2 Data-driven features

The physics-based features alone, however, are not enough to capture all of the information from the LU and NLU signals. As a result, a large number of features engineered from the time domain, frequency domain, and time-frequency domain of ultrasonic measurements are added to the feature pool.

- Time domain features

  Time domain features are peak amplitudes, ratios between peak amplitudes, and components from Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Statistics in time domain such as median, quantiles, variance, skewness, and kurtosis are also included. Besides, from the envelope analysis of a NLU signal, wave duration, wave energy, and the ratios between these quantities are calculated.

- Frequency domain features

  Frequency domain analysis offers some of the information that is not presented in the time domain. This information is especially valuable for periodic signals such as ultrasonic measurements. Thus, after applying fast Fourier transform (FFT), peak amplitudes, ratios between peak amplitudes, peak frequencies, frequency centroid and variance in FFT spectrum are extracted as the frequency domain features.

- Time-frequency domain features

  Ultrasonic signals are usually not stationary, i.e., frequency changes in time, because the interaction between ultrasonic waves and discontinuities within the material. Therefore, time-frequency analysis is needed to describe the phenomena. Discrete wavelet transform (DWT) is adopted to decompose ultrasonic measurements into several frequency bands. Then, statistics such as mean, median, kurtosis, and skewness are recorded for each frequency band.

We concatenate features from LU and NLU testing together, and thus the feature pool contains XXX features in total. A list of candidate features for LU and NLU measurements is displayed in Table

3.3 Feature selection

Feature selection aims to remove features that are redundant. Irrelevant features are common to see when we construct features without fully understanding a physical process. For example, the relationship between fatigue mechanism and ultrasonic responses. By including only the best subset of features for a prediction task, feature selection helps develop robust models against overfitting and improve model generalizability. There exists various feature selection techniques which can be mainly classified into three categories: filter methods, wrapper methods, and embedded methods. Each of these methods has its advantages, disadvantages, and suitable application scenarios.

In the model development pipeline, we adopted a wrapper method called Recursive Feature Elimination with Cross-validation (RFECV) to obtain the optimal feature subset that achieves the best predictive performance in multiple training/test data splits for a single model. Figure X shows the RFECV algorithm. First, recursive feature elimination (RFE) starts from a set with all available features and eliminate k features step by step based on the feature ranking with regressors/classifiers until the predetermined number of features n is reached. Nevertheless, the best number of features to select n^* is not determined. To find out n^* while alleviating the problem of overfitting, cross-validation (CV), a statistical model validation technique, is used along with RFE. CV partitions a dataset into training set and validation set in each fold. A model is evaluated multiple times with different partitions, and n^* is determined by the overall validation results. Then, RFE selects the optimal n^* features from the feature pool. We choose 5-fold classification in this feature selection procedure to avoid adding too much computation cost due to the fact that RFE is already computationally expensive.

3.4 Model training and validation

Model training and validation involve another CV loop. However, the CV here is not for finding the best feature subset but for providing a generalized estimate of a model's performance. Specifically, leave-one-group-out CV (LOGOCV) is applied, where each group contains three repeated measurements at one measurement location in one specimen. Here, we make an assumption that each group, i.e., each measurement location in a specimen, is an independent sample because of the differences in microstructure. In LOGOCV, each group is tested once and received validation scores by a predictor trained on the other groups, which efficiently utilizes the dataset and assesses the generalization capacity of a model.

3.5 Hyperparameter tuning

Searching an optimal set of hyperparameters is another crucial stage that significantly influences model performance in ML model development. We use grid search accompanied with the validation scores from the LOGOCV to tune hyperparameter in a simple and faster manner. Grid search exhaustively considers all candidates from predefined hyperparameter combinations. The number of hyperparameters and the range of each hyperparameter vary in different learning algorithms. The detailed settings will be discussed in Chapter 4 and 5.

Chapter 4

Remaining Useful Life Prediction

In this chapter, we propose a framework for predicting RUL of EoL products based on the ultrasonic testing. The framework has two parts: a) a ML classification task and b) a RUL inference procedure based on a S-N curve. First, ultrasonic signals are fed into ML classifiers to predict the loading condition and the number of fatigue cycles that a sample has gone through. Second, we estimate RUL from a S-N curve with the predicted loading condition and fatigue cycles.

4.1 Problem formulation

Given the goal of predicting RUL on EoL products, we need to formulate this as a ML problem first. In this section, We discuss possible formulations by considering the characteristics of the fatigue dataset and the impact on the ML system in practice.

4.1.1 Dataset

In this RUL prediction task, the dataset is constructed on the ultrasonic measurements on the interrupted fatigue testing specimens in Table 2.1. There are 15 specimens and each of these were measured at 9 locations alongside 3 repeated measurements, producing 405 observations in total. Notice that we treat one measurement location in one specimen as a sample in the model training and validation procedure with LOGOCV, as described in Section 3.4. Besides, each specimen is tested by a combination from 4 loading amplitudes and 3 fatigue levels (the percentage of fatigue life), which forms the labels of a specimen.

Table 4.1: Summary of the RUL prediction dataset

Specimen ID	Measurement Locations	Number of Repeated Measurements	Label (amplitude, percent of fatigue life)
1 2	$1 \sim 9$	3	Class 1 (11.7 kN, 33%)
3 4			Class 2 (11.7 kN, 67%)
5 6			Class 3 (12.7 kN, 33%)
7 8			Class 4 (12.7 kN, 67%)
9 10			Class 5 (14.7 kN, 33%)
11 12			Class 6 (14.7 kN, 67%)
13 14 15			Class 0 (0 kN, 0%)

4.1.2 Target variables

Obviously, RUL is directly translated by the percentage of fatigue life that a sample has gone through. The percentage of fatigue life as a continuous target variable is normally treated as a regression task. However, since we only have 3 different percentage of fatigue life in the dataset, which is not ideal for regression modeling, we decided to view the percentage of fatigue life as a discrete variable and the problem becomes a classification task.

Loading amplitude is another target variable to be considered because loading condition affects the mechanism of fatigue damage in a material. For instance, at 33% fatigue life, a sample undergoes 11.7 kN loading and a sample undergoes 14.7 kN loading could exist different fatigue damages. Hence, we place a label that is a combination of loading amplitude and the percentage of fatigue life on each sample. Table 4.1 presents the labeled RUL prediction dataset.

4.2 Design of classifiers

In this section, classifiers are designed for predicting the loading condition (amplitude) and the percent of fatigue life that a sample had experienced. Several classifiers are developed and the performance of each of those methods is evaluated based on the model development procedure in Chapter 3.

4.2.1 Multi-class classifier

A multi-class classifier is trained to classify a sample into one of the 7 classes. Figure XXX shows the inference process of a multi-class classifier. In multi-class problems, the classes are mutually exclusive. For example, class 1 and class 3 have nothing related. Despite we claimed that various loading conditions result in different fatigue behaviors in material, however, some similarities are still existed, e.g., at 33% fatigue life, samples undergone 11.7 kN and 12.7 kN are expected to be on a similar damage level. This idea can be applied to the commonality in samples at different percents of fatigue life as well. With the assumption about mutual exclusivity, multi-class formulation does not capture these characteristics of the data.

4.2.2 Multi-output classifier

On the other hand, multi-output classification is capable of dealing with mutually non-exclusive classes by predicting the loading amplitude and percentage of fatigue life separately. A multi-output classier outputs multiple labels, where each label is considered a multi-class classification problem. In this design, we build one classifier for predicting the loading amplitude from input signals; another one for classifying a signal into one of the percentages of fatigue life. Then, the two predicted labels are combined based on a rule-based algorithm to output a class label that is consistent with the label in the dataset, as depicted in Figure XXX. We call these two classifiers loading amplitude classifier (LAC) and fatigue cycle classifier (FCC), respectively. Each of the two classifiers are trained separately with the same model development procedure but different target variables. As a result, unlike the single multi-class classifier trying to learn a way to separate 7 classes, the multi-output classification builds LAC for a 4-class problem for loading amplitude and FCC for a 3-class problem for the percentage of fatigue life, which makes the problems easier to learn.

4.2.3 Two-stage classifier

Table XXX shows the design of a two-stage classifier. Extended from the idea of multi-output classifier in Subsection 4.2.2, a two-stage classifier is a classifier chains which predicts the loading amplitude and percentage of fatigue life in an order. The classification starts with a LAC and the predicted loading amplitude is added to the feature space of a $\mathrm{FCC}_{\text{two-stage}}$. By utilizing the predicted label as a feature for next classifiers, the label dependence is preserved, i.e., the prediction of the percentage of fatigue life sample has undergone is associated with its loading amplitude. In this case, we put the LAC before the $\mathrm{FCC}_{\text{two-stage}}$ because sometimes the loading condition is given in real life scenarios. Note that, in the training phase, the true loading amplitude is used to train the $\mathrm{FCC}_{\text{two-stage}}$, but the $\mathrm{FCC}_{\text{two-stage}}$ takes the predicted loading amplitude as one of the input features to do inference.

4.2.4 Hierarchical classifier

We further transform the multi-output task into a hierarchical classification scheme which is composed of multiple local classifiers based on a tree structure, shown in Figure XXX. One advantage of the hierarchical classification is to exploit parent-child class relationships present in the class hierarchy. Here, we train local classifiers per parent node in the taxonomy of the hierarchical classification problem. Specifically, one LAC and three FCCs for each loading amplitude are built, where each of the FCCs are trained by samples with the corresponding loading amplitude only. A prediction is inferred by the following manner: *a*) the LAC first output the loading amplitude for input LU and NLU signals. *b*) the predicted loading amplitude is also used to choose one of the FCCs for predicting the percentage of fatigue life. *c*) the FCC predicts the percentage of fatigue life. For example, 11.7 kN loading amplitude is predicted by the LAC. Therefore, $\mathrm{FCC}_{11.7\ \mathrm{kN}}$ where the subscript stands for the loading amplitude that the classier corresponds to, is selected and outputs a 33% fatigue life prediction. Finally, the predicted result is Class 1 (11.7 kN, 33%).

4.2.5 Evaluation metrics

We calculate accuracy, recall, precision, and F1-measure to evaluate a classifier's performance with the LOGOCV result, and confusion matrices are also presented. Although there exist other evaluation metrics for multi-output problems, we evaluate the aforementioned classifiers in a unified multi-class classification problem with the label defined in Table 4.1. Thus, these classifiers are comparable with each other. Since there is not much class imbalance in the dataset, accuracy is an overall indicator for a model's performance. Furthermore, the model's performance on each class is provided by recall, precision and F1-measure. Finally, confusion matrices serve as a visualization for summarizing the detailed result of the testing of a classifier. We let practitioners to decide the importance of each metric as it depends on different scenarios. For example, recall may be more important than precision in classifying highly-fatigued samples due to the consideration of safety.

4.2.6 Results

In this subsection, the classification performance of the proposed methods is presented and are compared with each other in Table XXX. A classifier's learning algorithm and the number of features determined by the model development procedure are also listed. It is noticed that the proposed hierarchical classifier outperforms other methods discussed in Section 4.2 effectively leveraging the characteristics of the RUL data. Therefore, the hierarchical classifier is chosen to demonstrate the RUL estimation algorithm in the following sections.

4.3 RUL estimation with a S-N curve

In the proposed framework, after obtaining the predicted loading amplitude and percentage of fatigue life, we estimate the RUL for a sample with the S-N curve of that material.

4.3.1 S-N curve with statistical distributions

Although a S-N curve is often referred to the best-fit line for fatigue data, fatigue life data can be modeled as statistical distributions to account for the variability in fatigue life. The randomness in the fatigue life comes from the stochastic behavior in fatigue process and the variance of microstructure in materials. As a result, the fatigue life has been widely modeled by statistical distributions including Gaussian (normal), log-normal, and Weibull distribution. Due to the limited amount of available fatigue life data in this research, we use a Gaussian normal distribution to model the fatigue life X_a at a loading amplitude a as

$$X_a \sim Normal(\bar{x_a}, \frac{s_a^2}{n_a}) \tag{4.1}$$

where $\bar{x_a}$ is the sample mean, s_a is the sample standard variance, and n_a is the sample size of fatigue data at loading amplitude a.

4.3.2 RUL inference procedure

A S-N curve serves as a look-up table for linking the classifier's predictions to the RUL. The inference procedure is detailed below:

1. Plotting the predicted loading amplitude and percentage of fatigue life. For example, the prediction, Class 1 (11.7 kN, 33% fatigue life), is plotted as the red dot in Figure XXX.

2. Determine the fatigue life with the corresponding loading amplitude from the S-N curve. The fatigue life can be in the format of a single value such as the median fatigue life or a statistical distribution.

3. If the fatigue life is represented by a value, the RUL can be estimated by Equation (4.2):

   $$RUL = N_{fl} - N_c \tag{4.2}$$

   where N_{fl} is the fatigue life from a S-N curve and N_c is the number of cycles translated from the predicted percentage of fatigue life a sample has undergone.

If the fatigue life at a given loading amplitude is modeled as a normal distribution, one way to estimate the RUL of a sample is to transform the fatigue life distribution by a factor of the predicted percentage of fatigue life, as in Equation (4.3) to (4.5):

$$X \sim Normal(\mu, \sigma^2) \tag{4.3}$$

$$Y = pX \tag{4.4}$$

$$Y \sim Normal(p\mu, p^2\sigma^2) \tag{4.5}$$

where Y is a random variable representing the RUL of a sample; X follows a normal distribution $Normal(\mu, \sigma^2)$ with μ as the mean and σ^2 as the variance of the fatigue life, and F is scaled by p which is the predicted percentage of fatigue life. Then, the point estimates, e.g., mean, median, or 25% quantile, as well as the interval estimate of RUL can be inferred. For instance, the 90% confidence interval of RUL is

$$CI = [\mu - z_{0.05}p\sigma,\ \mu + z_{0.95}p\sigma] \tag{4.6}$$

4.3.3 Example

4.4 Discussion

Chapter 5

Residual Stress & FWHM Prediction

Chapter 6

Conclusion

References