



A DATA DRIVEN REPORT ON MATERIAL PERFORMANCE AND FAILURE 2026

JANUARY 16

Authored by: Sesethu M. Bango



Logo
Name

INVESTIGATING AND PREDICTING FATIGUE BEHAVIOUR OF STRUCTURAL ALLOYS ACROSS DIFFERENT CYCLE LOADING REGIMES

SECTION 1 — BACKGROUND

Fatigue behaviour plays a central role in the performance and reliability of structural alloys. In the context of a materials engineering company, the ability to understand how and why materials fail under cyclic loading directly influences product safety, service life, and design optimisation. Components in sectors such as automotive, aerospace, energy systems, and industrial machinery regularly experience repeated stresses and strains. Over time, these loading cycles can initiate microstructural damage that ultimately leads to fatigue failure.

High-entropy alloys (HEAs), being a class of materials with unique mechanical and microstructural properties, offer promising performance under such demanding conditions. However, the variability and complexity of fatigue behaviour across different HEA compositions require structured, data-driven analysis. Both Low-Cycle Fatigue (LCF) and High-Cycle Fatigue (HCF) must be understood to evaluate how alloys behave across the full range of operational conditions—from high-strain, plastic-deformation regimes to high-frequency, low-amplitude elastic regimes.

This report focuses on analysing and modelling fatigue behaviour using experimental datasets for LCF and HCF conditions. The aim is to extract actionable insights and generate predictive models that assist engineers, researchers, and management in assessing material suitability and identifying the factors most strongly influencing fatigue performance.

The scope of this work includes:

- Cleaning and preparing LCF and HCF datasets extracted from controlled fatigue testing.
- Exploring the dominant trends, relationships, and key mechanical drivers behind fatigue life.
- Applying established fatigue models—specifically the Coffin–Manson relationship for LCF and the Basquin law for HCF.
- Interpreting model behaviour to understand how strain amplitudes, stresses, and mechanical properties influence cycles-to-failure.

The analysis does **not** include metallurgical microscopy, microstructure evolution modelling, or environmental effects such as corrosion or thermal aging. Instead, it focuses strictly on the mechanical fatigue behaviour captured in the datasets and the engineering value this behaviour provides.

This background sets the context for the subsequent sections, which detail the cleaned data, exploratory analysis, and fitted fatigue-life models.

SECTION 2 — Summary of Findings

This study investigated material fatigue behaviour across Low-Cycle Fatigue (LCF) and High-Cycle Fatigue (HCF) regimes using a combination of classical materials engineering theory and data-driven modelling techniques. The analysis yielded several key findings relevant to both technical understanding and practical application. For the LCF regime, fatigue life was found to be strongly governed by strain-based parameters. Total strain amplitude, plastic strain amplitude, and related strain-derived features exhibited strong negative correlations

with fatigue life. Classical strain–life behaviour was clearly observed, and regression-based modelling aligned well with Coffin–Manson theory. The combined elastic–plastic strain formulation provided a more accurate and physically meaningful representation of fatigue behaviour than individual strain components considered in isolation.

In contrast, HCF behaviour was characterised by significantly greater scatter. Stress-based variables showed only moderate correlations with fatigue life, and no single parameter was found to dominate fatigue performance. Normalised stress measures and frequency-related variables provided some explanatory power, but overall predictability remained limited. This reflects the inherently multi-factor and variable nature of high-cycle fatigue mechanisms.

Model building and evaluation further highlighted the contrast between regimes. Linear regression models performed well for LCF, achieving relatively strong explanatory power and closely matching experimental trends. For HCF, linear regression with polynomial features provided marginal improvements over baseline models, but overall model performance remained modest. Increasing model complexity through ensemble methods and extensive hyperparameter tuning did not result in meaningful gains, indicating model saturation. From a broader perspective, the findings demonstrate that data-driven models are most effective when applied within a strong theoretical framework. Where fatigue behaviour is well-structured and governed by known physical mechanisms, statistical models can provide valuable insight and support engineering decision-making. Conversely, in regimes dominated by variability and interacting factors, such models are better suited for trend identification rather than precise life prediction.

Overall, the study confirms the complementary role of data-driven analysis in materials engineering: enhancing understanding, supporting interpretation, and guiding further investigation, while remaining grounded in established fatigue theory.

SECTION 3 — REFLECTION

This project highlighted both the potential and the limitations of applying data-driven methods to fatigue analysis within a materials engineering context. While statistical and machine learning techniques offer powerful tools for identifying trends and relationships in experimental data, their effectiveness is strongly dependent on the underlying physical structure of the problem.

A key outcome of the work was the demonstration that data science methods can be successfully applied to engineering problems when guided by appropriate domain knowledge. In the Low-Cycle Fatigue regime, where behaviour is governed by well-understood strain-controlled mechanisms, relatively simple regression models were able to capture meaningful relationships that aligned closely with classical fatigue theory. This reinforced the value of combining theoretical understanding with data-driven approaches rather than treating modelling as a purely computational exercise.

In contrast, the High-Cycle Fatigue analysis underscored the limitations of machine learning in physical systems characterised by inherent variability and multi-factor dependence. Despite testing more complex models and feature engineering strategies, predictive performance remained limited. This highlighted that increasing model complexity does not necessarily lead to better results when the underlying data is highly scattered or when key influencing variables are not fully represented.

Overall, the work demonstrated that data-driven modelling is most effective when used as a complementary tool to established engineering methods. By integrating materials science principles with statistical analysis,

the project provided insight into fatigue behaviour while maintaining an appropriate level of confidence in the conclusions drawn.

SECTION 4 — DATA COLLECTION AND PREPARATION

This section describes how the fatigue datasets were assembled, prepared, and enriched to support exploratory analysis and fatigue-life modelling for a materials engineering context.

4.1 Data Sources and Structure

The analysis is based on an experimental fatigue database of high-entropy alloys, provided in spreadsheet format. The dataset contains separate records for Low-Cycle Fatigue (LCF) and High-Cycle Fatigue (HCF) testing, with each fatigue regime represented by both summary-level information and individual test results. For the LCF analysis, the summary dataset was merged with the corresponding individual test dataset using a common material identifier. The resulting merged LCF dataset contained **46 rows and 28 columns**, capturing strain amplitudes, fatigue life, mechanical properties, grain size, temperature, and derived features. Similarly, for the HCF analysis, the summary and individual datasets were merged using the same identifier approach. After filtering for valid stress and fatigue-life records, the final merged HCF dataset contained **256 rows and 29 columns**, including stress amplitudes, ultimate tensile strength, testing frequency, stress ratios, and fatigue life.

4.2 Data Preparation Approach

Rather than applying aggressive filtering, the data preparation strategy focused on preserving physically meaningful fatigue data while ensuring consistency across variables used for analysis and modelling. This approach ensured that the resulting datasets remained representative of experimental fatigue behaviour, while still being suitable for statistical exploration and regression-based modelling.

Key preparation steps included:

- Standardising column names to ensure consistency across merged datasets.
- Retaining only records with valid fatigue life and loading information required for analysis.
- Ensuring numerical fields used in modelling were finite and positive, particularly for log-transformed variables.

This preparation strategy balanced data quality with data retention, which is particularly important when working with experimental fatigue datasets that may be limited in size.

4.3 Feature Engineering

To enhance the analytical value of the datasets, additional features were engineered for both LCF and HCF regimes. These features were selected to reflect known fatigue mechanisms and to support interpretable modelling.

Low-Cycle Fatigue (LCF) Features

For the LCF dataset, the following derived features were created:

- **Elastic-to-total strain ratio**, representing the proportion of elastic deformation in the applied strain.
- **Plastic-to-total strain ratio**, capturing the contribution of plastic deformation to overall strain.
- **Estimated stress amplitude**, calculated using elastic strain amplitude and elastic modulus, providing an approximate measure of stress under cyclic loading.

-
- **Fatigue strength coefficient**, approximated from available tensile strength data to support fatigue model fitting.
 - **Strain energy density**, estimated as a combined measure of stress and total strain, reflecting the energy input per cycle.
 - **Log-transformed fatigue life**, used to linearise relationships for regression-based analysis.

An elastic modulus value was manually assigned based on material grouping to introduce an additional physically meaningful variable into the dataset. This was done to enable further insight into stress–strain behaviour and fatigue response, particularly in cases where direct stress measurements were not available.

High-Cycle Fatigue (HCF) Features

For the HCF dataset, the following features were engineered:

- **Normalized stress**, defined as stress amplitude divided by ultimate tensile strength, allowing comparison across materials.
- **Log-transformed fatigue life**, supporting linear stress–life modelling.
- **Stress ratio**, included to capture mean stress effects.
- **Goodman-adjusted stress**, introduced as an optional correction for mean stress influence.
- **Frequency bins**, grouping tests by loading frequency to explore potential rate effects on fatigue life.

These features enabled both descriptive exploration and physics-informed modelling of stress–life behaviour under high-cycle fatigue conditions.

4.4 Prepared Dataset for Analysis

Following data preparation and feature engineering, both the LCF and HCF datasets were ready for exploratory data analysis and fatigue-life modelling. The resulting datasets provide a consistent and enriched foundation for identifying dominant fatigue drivers, comparing fatigue regimes, and fitting established fatigue models.

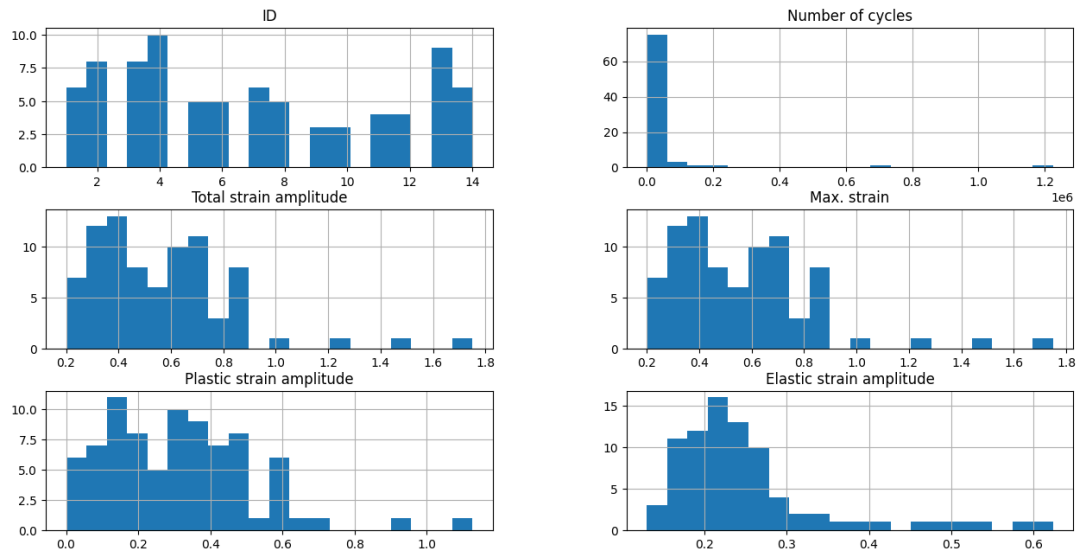
5.1 Univariate Analysis

The LCF dataset exhibits a wide range of fatigue responses. The number of cycles to failure spans **several orders of magnitude**, ranging from a few hundred cycles to values exceeding one million cycles. This wide spread is characteristic of low-cycle fatigue behaviour, where small changes in strain amplitude can result in large variations in fatigue life.

The strain-related variables—**plastic strain amplitude, maximum strain, and total strain amplitude**—show **high variability** and non-Gaussian distributions. These variables are heavily skewed, reflecting the presence of both low-strain, long-life tests and high-strain, short-life tests within the dataset. This variability reinforces the need for log-transformed fatigue life and non-linear modelling approaches.

In contrast, the **elastic strain amplitude** exhibits a more regular distribution, resembling a **left-skewed Gaussian shape**. This suggests that elastic strain values are more tightly clustered and less variable than plastic or total strain measures, particularly in the LCF regime where plastic deformation is significant.

Histograms of Numeric Columns



5.2 Bivariate Analysis: Strain Amplitudes vs Fatigue Life

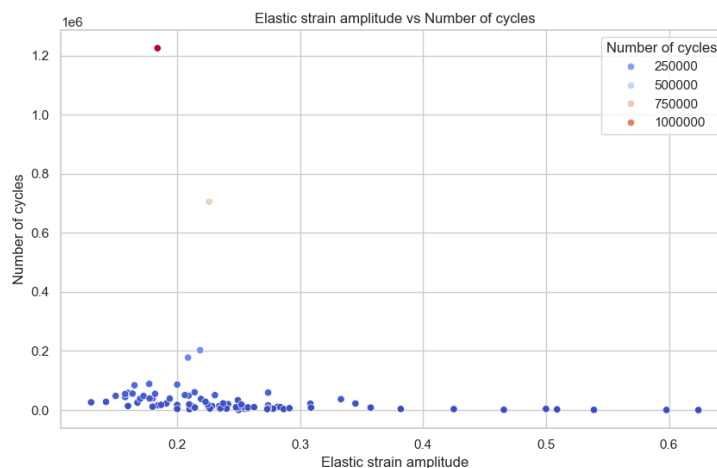
The relationship between strain amplitudes and fatigue life is illustrated through scatter plots of elastic, plastic, and total strain amplitudes against the number of cycles to failure.

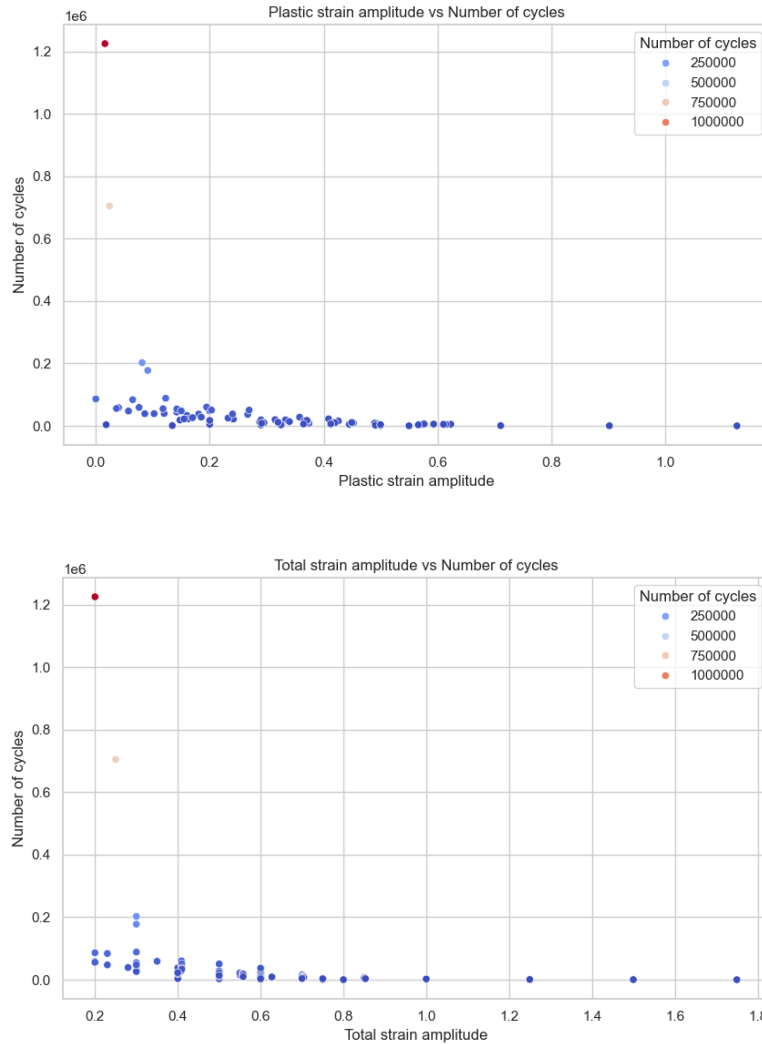
The **plastic strain amplitude** shows a **clear and strong negative relationship** with fatigue life. As plastic strain amplitude increases, the number of cycles to failure decreases sharply. This trend highlights the dominant role of plastic deformation in driving damage accumulation under low-cycle fatigue conditions.

A similar, though slightly less pronounced, negative trend is observed for **total strain amplitude**. Since total strain combines both elastic and plastic components, it captures the overall fatigue response while still reflecting the strong influence of plastic strain.

The **elastic strain amplitude** also displays a negative relationship with fatigue life, but the scatter is more dispersed. This indicates that elastic strain alone is a weaker predictor of fatigue life in the LCF regime compared to plastic or total strain measures.

Across all three plots, high-cycle data points are clustered at low strain amplitudes, while high strain amplitudes consistently correspond to low fatigue lives, reinforcing the strain-controlled nature of LCF behaviour.





5.3 Correlation Analysis

Correlation analysis further quantifies the observed trends between strain-related variables and fatigue life. The strongest correlations with fatigue life are observed for:

- **Total strain amplitude**
- **Maximum strain**
- **Estimated stress amplitude**
- **Strain energy density**

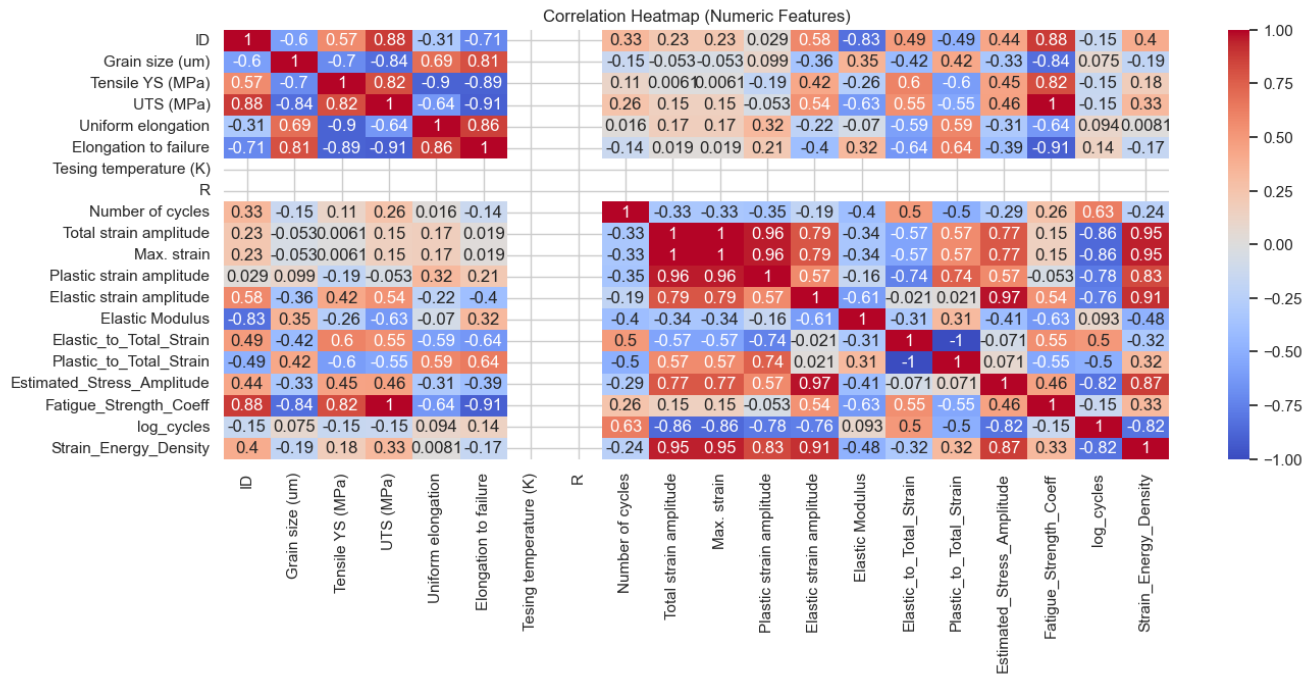
All of these variables exhibit **strong negative correlations** with fatigue life, with correlation coefficients **greater than |0.80|**. This confirms that variables representing the overall deformation and energy input per cycle are the most influential drivers of fatigue damage in the LCF regime.

Both **plastic strain amplitude** and **elastic strain amplitude** also show strong negative correlations with fatigue life, though plastic strain remains the more influential of the two. This supports the interpretation that plastic deformation governs damage accumulation under low-cycle loading.

Interestingly, the strain ratio features show moderate correlations:

- **Elastic-to-total strain ratio** exhibits a **positive correlation** with fatigue life ($r \approx +0.5$), indicating improved fatigue life when a larger fraction of the total strain is elastic.

- **Plastic-to-total strain ratio** shows a **negative correlation** ($r \approx -0.5$), further reinforcing the detrimental effect of plastic deformation on fatigue performance.



5.4 Deeper Insights from Strain–Life and Mechanical Relationships

A closer examination of strain–life relationships confirms that **total strain amplitude** and **plastic strain amplitude** have higher correlation magnitudes with fatigue life than elastic strain amplitude. This hierarchy aligns with established fatigue theory, where plastic strain dominates damage processes in low-cycle fatigue. When comparing mechanical properties with fatigue life, **ultimate tensile strength (UTS)** shows a stronger correlation with the number of cycles to failure than yield strength (YS). This suggests that overall material strength capacity plays a more significant role in resisting fatigue damage under cyclic plastic deformation than the initial yield point alone.

All tests were conducted at the same temperature, and therefore no temperature-dependent effects on fatigue life are observed or discussed in this analysis.

This exploratory analysis establishes that LCF behaviour in the dataset is strongly strain-controlled, with plastic deformation and energy-related measures serving as the primary indicators of fatigue life. These findings provide a clear foundation for the fatigue modelling presented in the next section.

SECTION 6 — DATA EXPLORATION (HIGH-CYCLE FATIGUE)

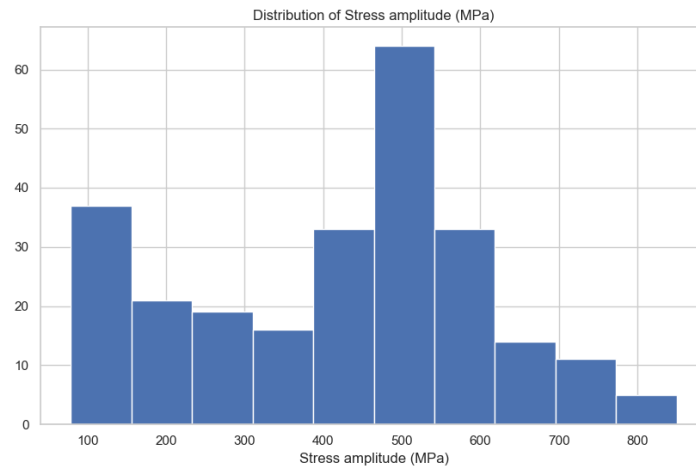
This section explores the High-Cycle Fatigue (HCF) dataset, focusing on stress–life behaviour, statistical trends, correlation patterns, and the influence of loading frequency on fatigue life.

6.1 Univariate Analysis

The HCF dataset spans the high-cycle fatigue regime, with the number of cycles to failure ranging from **tens of thousands to approximately ten million cycles**. This range is consistent with elastic-dominated fatigue behaviour, where failure occurs after a large number of loading cycles at relatively lower strain levels.

The final prepared HCF dataset contains **253 rows and 29 columns**, providing a substantially larger sample size than the LCF dataset and enabling broader statistical exploration.

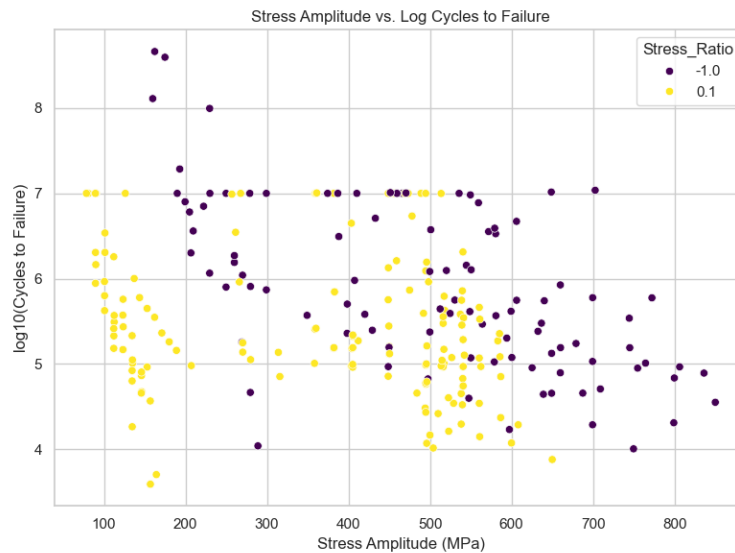
The **stress amplitude** distribution is approximately **normally distributed**, as shown in the histogram, with the exception of the higher-stress region represented by the first few bins. This deviation likely reflects the presence of specific test conditions or material groups operating at low stress levels rather than random experimental noise.



6.2 Bivariate Analysis: Stress and Normalized Stress vs Fatigue Life

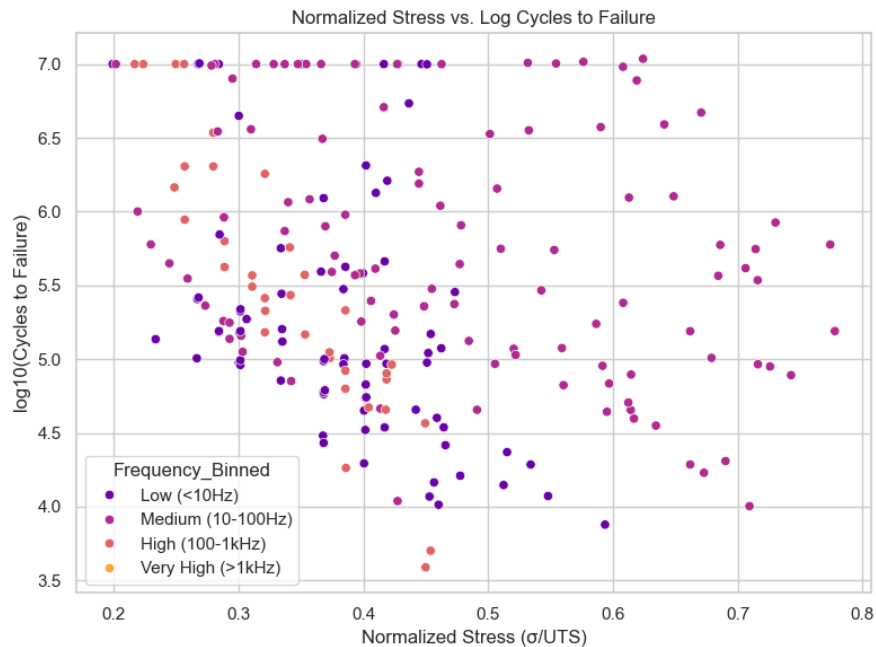
The relationship between stress amplitude and fatigue life is illustrated through scatter plots of **stress amplitude vs log(cycles to failure)** and **normalized stress vs log(cycles to failure)**.

In the stress amplitude plot, a **general downward trend** is visible: higher stress amplitudes tend to correspond to lower fatigue life. However, the data shows **considerable scatter**, particularly in the mid-stress range. This indicates that stress amplitude alone does not fully capture the variability in fatigue life under high-cycle conditions.



The normalized stress plot, which scales stress amplitude by ultimate tensile strength, shows a **slightly clearer structure**, but still exhibits significant dispersion. While normalized stress helps partially collapse data across different materials, it does not fully eliminate variability, suggesting that additional factors—such as loading frequency and stress ratio—also influence fatigue performance.

Colour-coding by frequency bins highlights that data points from different frequency ranges overlap substantially. This overlap explains why simple stress–life relationships are weaker in HCF compared to LCF, and why more scatter is expected in elastic-dominated fatigue regimes.



6.3 Correlation Analysis

Correlation analysis confirms that **no single variable exhibits a strong linear correlation** with fatigue life in the HCF dataset.

The strongest correlations with $\log(\text{cycles to failure})$ are observed for:

- **Maximum stress**, with a moderate negative correlation ($r \approx -0.35$).
- **Loading frequency**, with a moderate positive correlation ($r \approx +0.38$).

Other commonly used fatigue variables—including stress amplitude, normalized stress, stress ratio, and R-ratio—show weaker negative correlations, generally in the **−0.2 to −0.3 range**. These values indicate that while stress-related variables influence fatigue life, their individual predictive power is limited when considered in isolation.

This behaviour is typical of high-cycle fatigue, where fatigue life is sensitive to multiple interacting factors rather than a single dominant driver.

7. Model Building and Evaluation

This section presents the development, selection, and evaluation of predictive models for both Low-Cycle Fatigue (LCF) and High-Cycle Fatigue (HCF). The objective is to assess how well fatigue life can be explained and predicted using regression-based approaches, while maintaining interpretability and alignment with established fatigue theory.

7.1 Modelling Approach and Evaluation Strategy

For both fatigue regimes, supervised regression models were developed using an 80/20 train–test split. Model performance was evaluated primarily using the coefficient of determination (R^2) to assess variance explained, with Root Mean Squared Error (RMSE) used as a secondary metric to quantify prediction error magnitude in engineering terms.

Multiple modelling approaches were explored, including linear regression, polynomial regression, and ensemble-based methods. Final model selection prioritised a balance between predictive performance, physical interpretability, and robustness.

7.2 Low-Cycle Fatigue (LCF) Modelling

7.2.1 Model Selection

For the LCF dataset, **linear regression** was selected as the final modelling approach. Compared to more complex alternatives, it demonstrated significantly better performance and strong consistency with the strain-controlled nature of low-cycle fatigue.

The linear model achieved higher R^2 values and lower error metrics relative to other tested models, indicating that LCF behaviour in this dataset is well captured by linear relationships in transformed feature space.

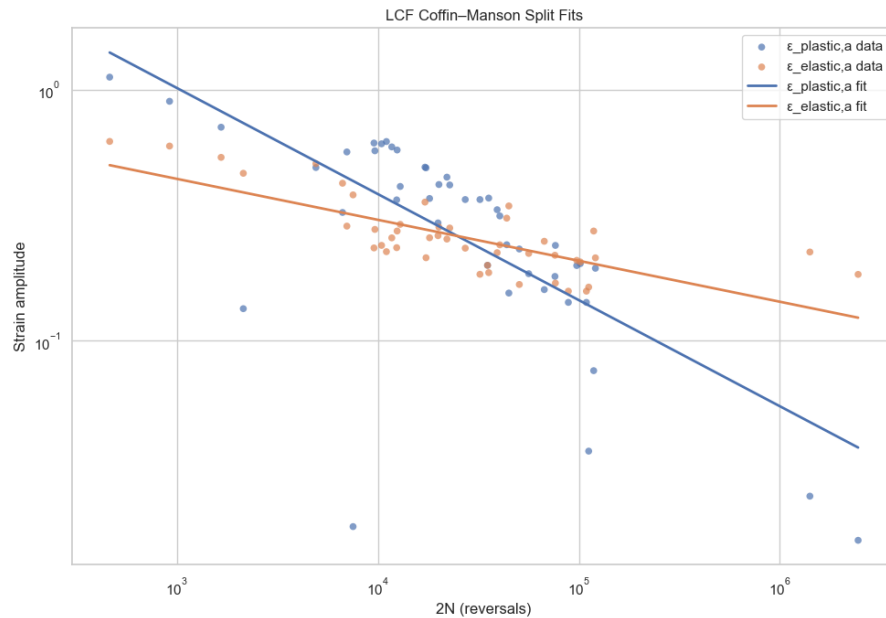
Model	R^2	MAE	RMSE
Linear Regression	0.946	0.123	0.180
Random Forest	0.754	0.267	0.383

7.2.2 Coffin–Manson Strain–Life Behaviour

The classical Coffin–Manson framework was used to analyse the elastic and plastic strain contributions to fatigue life. Separate log–log regressions were fitted for the elastic and plastic strain components, as well as for the combined total strain amplitude.

The plastic strain component exhibited a moderate fit ($R^2 \approx 0.49$), while the elastic strain component showed slightly improved explanatory power ($R^2 \approx 0.57$). Despite these moderate individual fits, the **combined total strain response closely followed experimental trends**, indicating that the superposition of elastic and plastic terms provides a more physically meaningful description of fatigue behaviour.

Both elastic and plastic strain fits aligned well with the experimental data on a log–log scale, and the composite Coffin–Manson curve captured the overall strain–life relationship effectively.



These results confirm that:

- Plastic strain dominates fatigue life in the LCF regime
- Elastic strain contributes meaningfully but is secondary
- The combined Coffin–Manson formulation best represents observed behaviour

7.3 High-Cycle Fatigue (HCF) Modelling

7.3.1 Model Selection

For the HCF dataset, **linear regression with polynomial features** was selected as the final model. While ensemble methods such as Random Forests achieved comparable performance, linear regression with polynomial expansion provided marginally better results while preserving interpretability.

Given the inherently scattered nature of high-cycle fatigue data, the performance difference between models was small. The selected model was therefore chosen based on its simplicity and transparency rather than raw predictive gain alone.

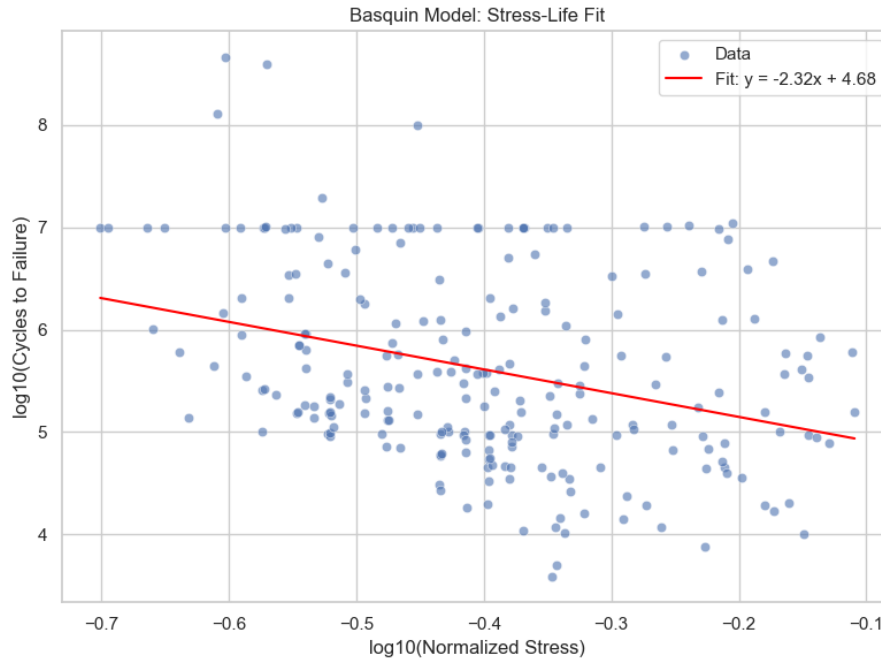
Model	R^2	MAE	RMSE
Linear Regression with Polynomial Features	0.6932	0.4145	0.5632
Random Forest	0.6929	0.4574	0.5635

7.3.2 Basquin-Type Stress–Life Behaviour

A Basquin-type regression was used to model the relationship between stress-related features and fatigue life in the high-cycle regime. The fitted model exhibited a low overall R^2 (≈ 0.10), reflecting the substantial scatter present in HCF data.

Despite the low variance explained, the regression was statistically significant ($p \approx 5.0 \times 10^{-7}$), indicating that the identified relationships are unlikely to be due to random chance. This combination of low R^2 and high statistical significance is characteristic of fatigue datasets where multiple interacting variables influence life outcomes.

The fitted slope and intercept are physically reasonable and consistent with the expected inverse relationship between stress amplitude and fatigue life.



7.3.3 Feature Engineering and Model Saturation

Several domain-inspired feature engineering strategies were evaluated, including stress normalisation, interaction terms, and frequency-related features. While some polynomial transformations led to measurable improvements, further feature augmentation did not result in meaningful gains in predictive performance. Extensive Random Forest hyperparameter tuning—including large ensemble sizes and controlled tree depth—also failed to improve model accuracy and in some cases slightly degraded performance. This behaviour suggests **model saturation**, where additional complexity does not translate into improved generalisation. These results indicate that the predictive ceiling for the available HCF data has largely been reached using conventional regression and ensemble approaches.

7.4 Model Limitations and Interpretation

The modelling results highlight a clear contrast between fatigue regimes. LCF behaviour is comparatively well-structured and predictable due to strain-controlled mechanisms, whereas HCF behaviour is more dispersed and sensitive to multiple interacting factors such as stress amplitude, frequency, and test conditions. As a result, HCF models should be interpreted as providing trend-level insight rather than precise life prediction. The models remain valuable for understanding relative influences and guiding further analysis, but they are not intended to replace detailed experimental assessment.

7.5 Summary of Modelling Outcomes

- Linear regression provided the best balance of performance and interpretability for both LCF and HCF.
- Coffin–Manson analysis confirmed classical strain–life behaviour in the LCF regime.
- HCF models exhibited limited predictive strength due to inherent data scatter, despite statistically significant relationships.

-
- Increased model complexity yielded diminishing returns, indicating saturation.

These findings motivate cautious but informed use of data-driven models in fatigue analysis and support their role as complementary tools alongside established materials engineering methods.

8. Conclusion

This study demonstrates the value of combining established materials engineering theory with data-driven modelling techniques to analyse fatigue behaviour across different regimes. While predictive accuracy varies between Low-Cycle Fatigue (LCF) and High-Cycle Fatigue (HCF), the integrated approach provides meaningful insight into underlying mechanisms and practical trends.

A clear distinction emerged between fatigue regimes. LCF behaviour was well-structured and strain-dominated, allowing classical strain–life relationships, such as the Coffin–Manson formulation, to be captured effectively using relatively simple regression models. In contrast, HCF behaviour exhibited significantly greater scatter, reflecting the influence of multiple interacting factors and limiting the predictive strength of stress–life models.

From an engineering and organisational perspective, the project demonstrates a strong capability to extract insight from experimental fatigue data and to apply regression and machine learning techniques appropriately to materials engineering problems. The results highlight where data-driven models can be relied upon for interpretation and where caution is required, particularly when dealing with inherently variable high-cycle fatigue behaviour.

Although predictive performance in the HCF regime was limited, the models remained valuable as decision-support tools. They enabled identification of dominant trends, assessment of relative variable importance, and informed comparison between modelling approaches, even when precise life prediction was not feasible.

The main limitations of this study stem from experimental scatter, dataset heterogeneity, and the absence of additional controlling variables such as microstructural descriptors or temperature variation. These constraints are typical of fatigue datasets and do not detract from the validity of the observed trends.

Future work would benefit from more controlled experimental data, as well as expanded studies incorporating a wider range of loading frequencies and temperatures. Such enhancements would improve model generalisation and enable deeper exploration of fatigue mechanisms, particularly in the high-cycle regime.

Overall, this work supports the use of data-driven modelling as a complementary tool to classical fatigue theory, reinforcing engineering understanding while acknowledging the inherent complexity of fatigue behaviour in real materials systems.

9. References

- Coffin, L.F. (1954). A study of the effects of cyclic thermal stresses on a ductile metal. *Transactions of the ASME*, 76, pp. 931–950.
- Basquin, O.H. (1910). The exponential law of endurance tests. *Proceedings of the American Society for Testing Materials*, 10, pp. 625–630.
- Dowling, N.E. (2013). *Mechanical Behavior of Materials: Engineering Methods for Deformation, Fracture, and Fatigue*. 4th ed. Pearson Education.
- Suresh, S. (1998). *Fatigue of Materials*. 2nd ed. Cambridge University Press.
- Callister, W.D. and Rethwisch, D.G. (2020). *Materials Science and Engineering: An Introduction*. 10th ed. Wiley.

Schijve, J. (2009). *Fatigue of Structures and Materials*. 2nd ed. Springer.

Stephens, R.I., Fatemi, A., Stephens, R.R. and Fuchs, H.O. (2000). *Metal Fatigue in Engineering*. 2nd ed. Wiley.