

# INTPIX4NAを用いたX線残留応力測定と 機械学習による転がり疲労の定量評価

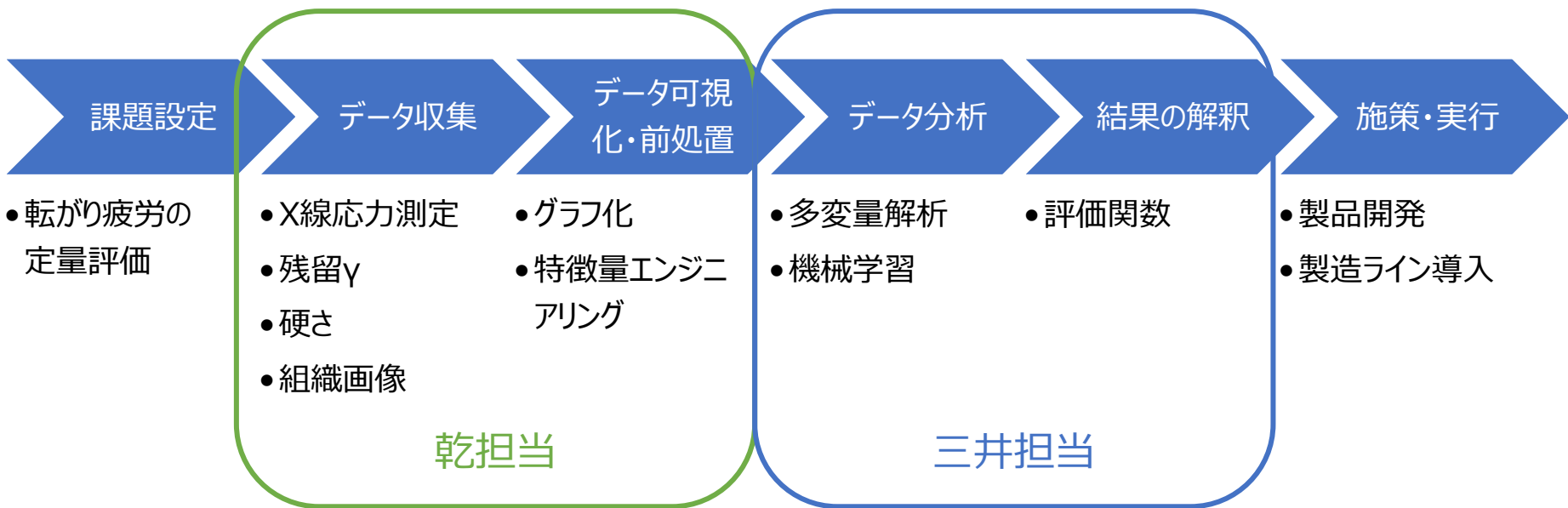
滋賀大学  
データサイエンス・AIイノベーション研究推進センター  
三井真吾

金沢大学 佐々木敏彦  
(株)不二越 乾典規



滋賀大学

- スラスト荷重試験
- X線残留応力測定
- 残留応力・回折環半価幅の位置依存性・時間変化
- 回帰分析
  - 線形回帰
    - 重回帰分析
  - 非線形回帰
    - ニューラルネットワーク回帰分析
    - ランダムフォレスト回帰分析
- まとめ



**Machine Learning Algorithm Cheat Sheet**

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the goal you want to achieve with your data.

Microsoft Azure

**What do you want to do?**

- Extract information from text**
  - Test Analytics**
    - Derives high-quality information from text
    - Answers questions like: What info is in this text?
    - Latent Dirichlet Allocation: Unsupervised topic modeling, group texts that are similar from a column of free text
    - Extract N-gram Features from Text: Creates a dictionary of n-grams from a column of free text
    - Feature Hashing: Converts text data to integer encoded features using the Vowpal Wabbit library
    - Progressive Text: Performs cleaning operations on text, like removal of stop-words, case normalization
    - WordVector: Converts words to values for use in NLP tasks, like recommended named entity recognition, machine translation
- Predict between several categories**
  - Multiclass Classification**
    - Answers complex questions with multiple possible answers
    - Answers questions like: Is this A, B or C or D?
    - Multiclass Logistic Regression: Fast training times, linear model
    - Multiclass Neural Network: Accuracy, long training times
    - Multiclass Decision Forest: Accuracy, fast training times
    - One-vs-All Multiclass: Depends on the two-class classifier
    - One-vs-One Multiclass: Depends on binary classifier, less sensitive to imbalanced dataset with larger complexity
    - Multiclass Boosted Decision Tree: Non-parametric, fast training times and scalable
- Predict between two categories**
  - Two-Class Classification**
    - Answers simple two-choice questions, like yes or no, true or false
    - Answers questions like: Is this A or B?
    - Two-Class Support Vector Machine: Under 100 features, linear model
    - Two-Class Averaged Perceptron: Fast training, linear model
    - Two-Class Decision Forest: Accurate, fast training
    - Two-Class Logistic Regression: Fast training, linear model
    - Two-Class Boosted Decision Tree: Accurate, fast training, large memory footprint
    - Two-Class Neural Network: Accurate, long training times
- Generate recommendations**
  - Recommenders**
    - Predicts what someone will be interested in
    - Answers the question: What will they be interested in?
    - Use the Train Wide & Deep Recommender module: Hybrid recommender, both collaborative filtering and content-based approach
    - SVD Recommender: Collaborative filtering, better performance with lower cost by reducing dimensionality
- Discover structure**
  - Clustering**
    - Separates similar data points into intuitive groups
    - Answers questions like: How is this organized?
    - K-Means: Unsupervised learning
- Find unusual occurrences**
  - Anomaly Detection**
    - Identifies and predicts rare or unusual data points
    - Answers the question: Is this weird?
    - One Class SVM: Under 100 features, aggressive boundary
    - PCA Based Anomaly Detection: Fast training times
- Predict values**
  - Regression**
    - Makes forecasts by estimating the relationship between values
    - Answers questions like: How much or how many?
    - Fast Forest Quantile Regression: Predicts a distribution
    - Poisson Regression: Predicts event counts
    - Linear Regression: Fast training, linear model
    - Bayesian Linear Regression: Linear model, small data sets
    - Decision Forest Regression: Accurate, fast training times
    - Neural Network Regression: Accurate, long training times
    - Boosted Decision Tree Regression: Accurate, fast training times, large memory footprint
- Classify images**
  - Image Classification**
    - Classifies images with popular networks
    - Answers questions like: What kind of image represent?
    - ResNet: Modern deep learning neural network
    - DenseNet

**scikit-learn algorithm cheat sheet**

**classification**

- START
- more data?
  - > 50 samples: SGD Classifier
  - < 50 samples: Kernel Approximation
- number of samples?
  - > 1000 samples: SVC, Ensemble Classifiers, Staircase
  - < 1000 samples: KNeighbors Classifier, Linear SVC
- labeled data?
  - no: SVM, Staircase
  - yes: SGD Classifier, Linear SVC
- predicting a category?
  - no: SVM, Staircase
  - yes: SGD Classifier, Linear SVC

**regression**

- START
- more data?
  - > 50 samples: SGD Regressor
  - < 50 samples: Kernel Approximation
- number of samples?
  - > 1000 samples: SVM, Ensemble Regressors
  - < 1000 samples: KNeighbors Classifier, Linear SVC
- labeled data?
  - no: SVM, Staircase
  - yes: SGD Regressor, Linear SVC
- predicting a quantity?
  - no: SVM, Staircase
  - yes: SGD Regressor, Linear SVC

**clustering**

- START
- number of samples?
  - > 1000 samples: MiniBatch KMeans, MeanShift, VBGM
  - < 1000 samples: KMeans
- labeled data?
  - no: SVM, Staircase
  - yes: SGD Regressor, Linear SVC
- predicting structure?
  - no: SVM, Staircase
  - yes: SGD Regressor, Linear SVC

**dimensionality reduction**

- START
- number of samples?
  - > 1000 samples: Randomized PCA, IsoMap, Spectral Embedding, LLE
  - < 1000 samples: RidgeRegression, SVR, ElasticNet
- labeled data?
  - no: SVM, Staircase
  - yes: SGD Regressor, Linear SVC
- predicting a quantity?
  - no: SVM, Staircase
  - yes: SGD Regressor, Linear SVC

## 目的

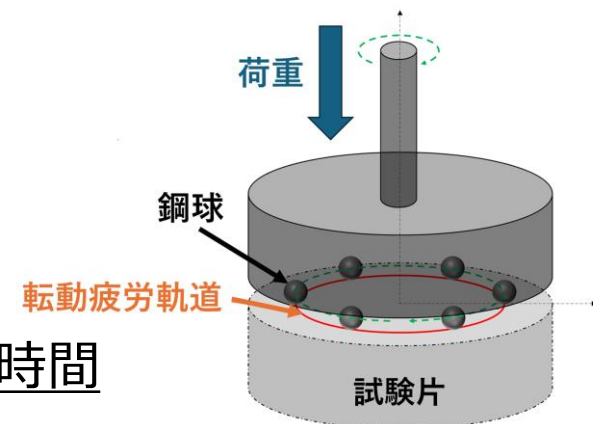
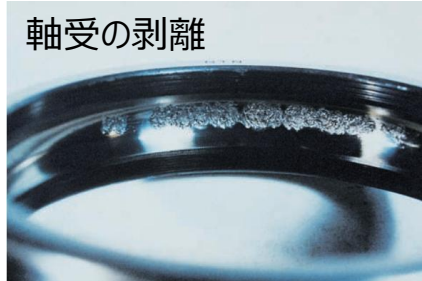
- 軸受やレールの損傷状態の定量的評価法を開発する。
- 転がり疲労の定量的な評価のために、スラスト荷重試験を行い、X線を用いた残留応力・回折環半価幅測定を行った。
- 測定データの**統計的解析・機械学習**により、試験時間の回帰分析を行った。

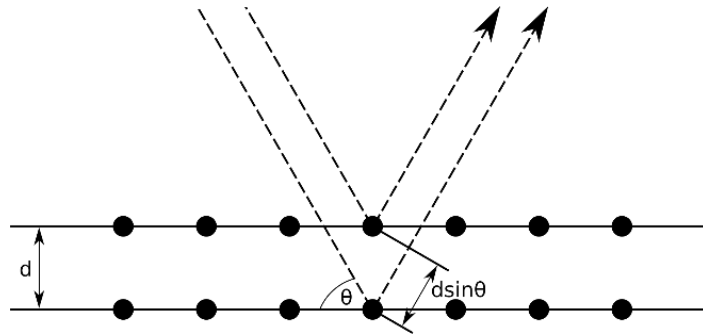
## 試験条件（詳細は乾さんと同様）

- SUJ2試験片
- 鋼球通過位置：半径19.25mm
- 鋼球通過回数： $10^8$ 回(556時間)

## 測定条件(INTPIX4NA搭載X線残留応力測定装置)

- $\sigma_x \cdot \tau_{xy}$ ・回折環半価幅、Cr-K $\alpha$ 、入射角 $\psi_0 = 35^\circ$
- 測定した疲労試験時間：0、24、72、120、240、556時間
- 半径方向：18.75-20.75mmを0.5mm間隔で5点
- 円周方向：3度ごとに120点を20点(60度分)を平均して6点

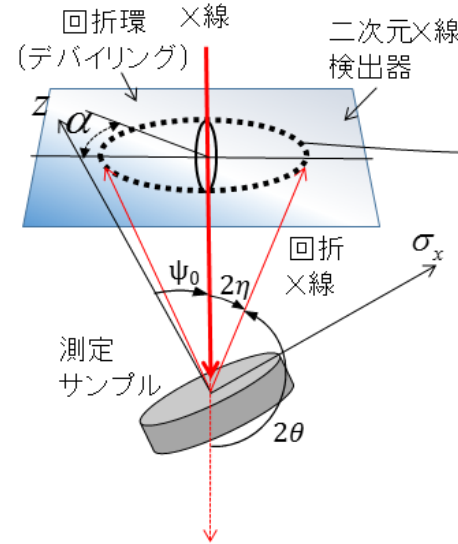




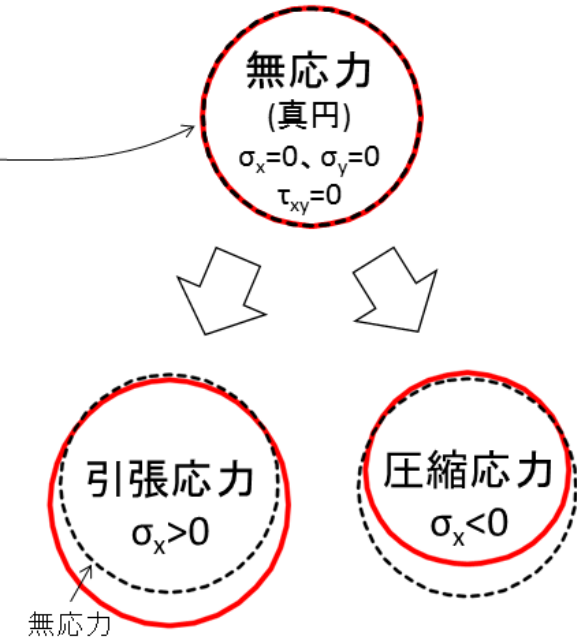
ブラッグの法則:  $2d \sin \theta = n\lambda$

$$\frac{d - d_0}{d} \equiv \varepsilon = (\theta_0 - \theta) \cot \theta_0$$

## 【回折環の測定】

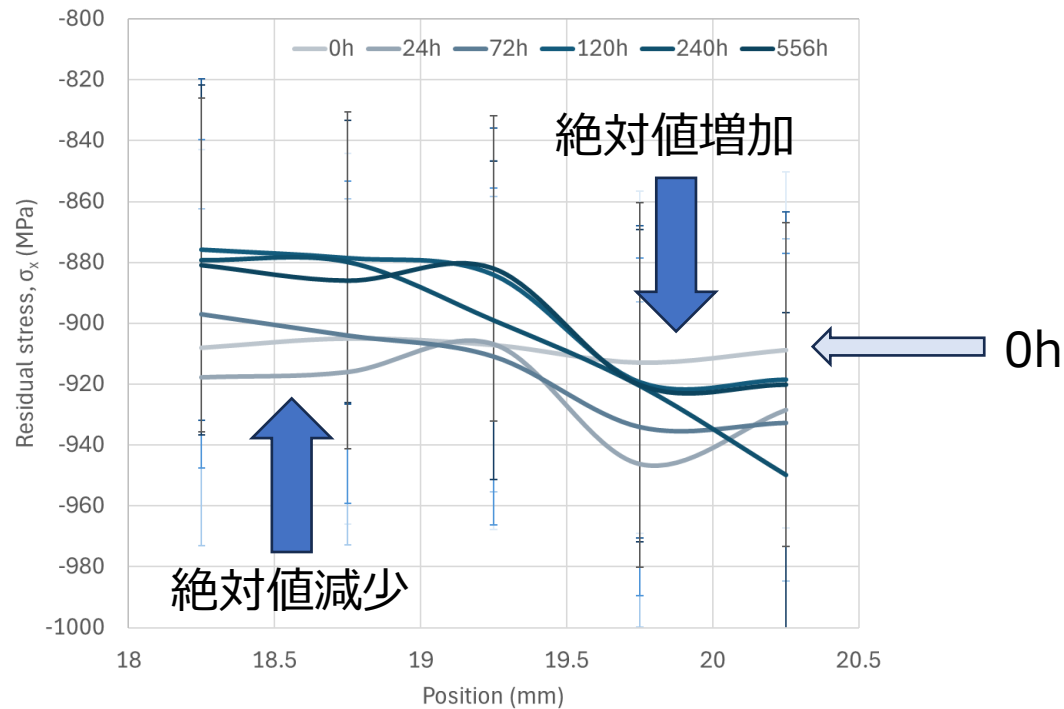


## 【応力による回折環の変化】



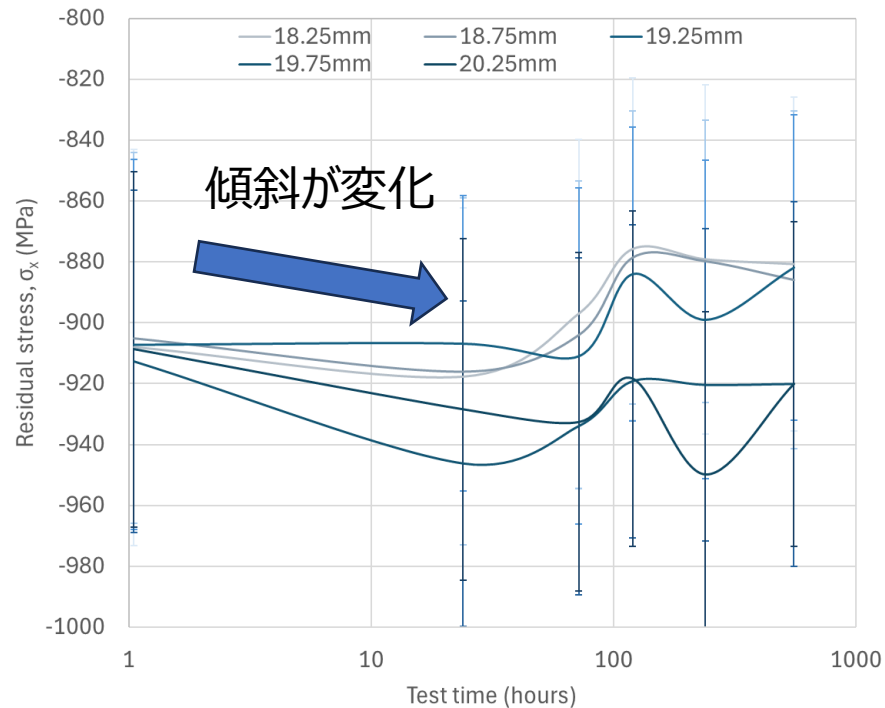
- 金属表面にX線を入射すると、ブラッグの法則により回折環が形成される。
- 残留応力によるひずみが回折環の変化として観測される。
- 回折環のゆがみ：残留応力、回折環半価幅：転位密度(硬さ)

- 垂直応力のX軸方向成分 $\sigma_x$ の位置依存性を調査
- グラフは試験片1周の平均値と標準偏差



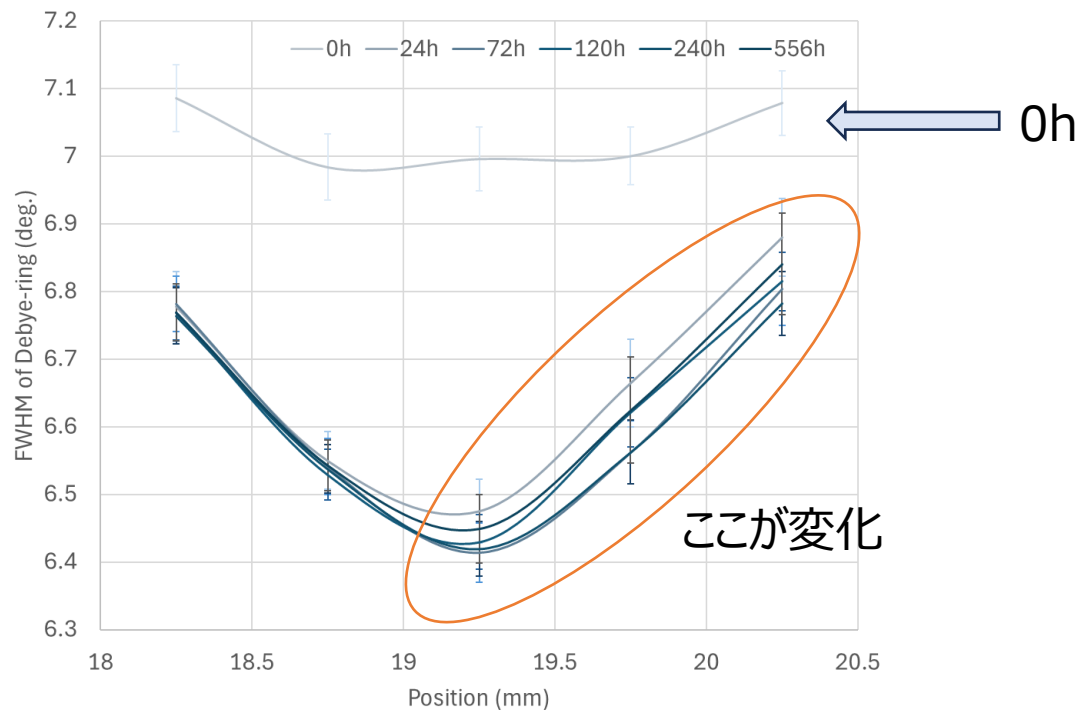
- 接触位置(19.25mm)の周辺でも変化している。

## ■ 垂直応力のX軸方向成分 $\sigma_x$ の時間変化を調査



- 24h以降は位置により変化量が異なるため、前の時間からの変化量 $\Delta\sigma_x$ を新たな特徴量として追加して回帰分析を行った。

## ■ 回折環半価幅の位置依存性を調査

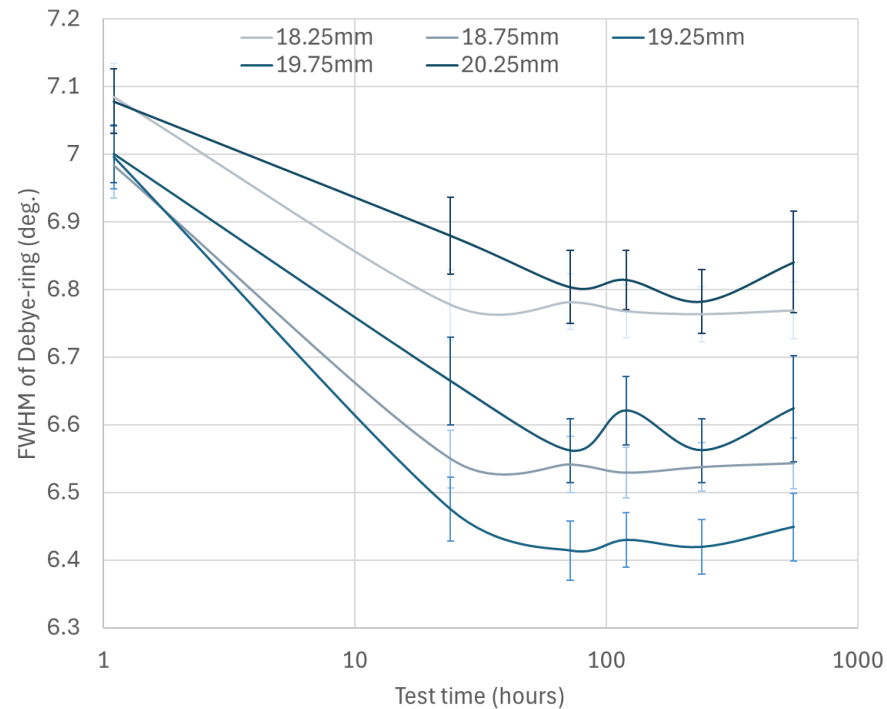


■ 回折環半価幅は24hまで大きく変化する。

■ 24h以降は、19.25mmより大きい半径で変化している。

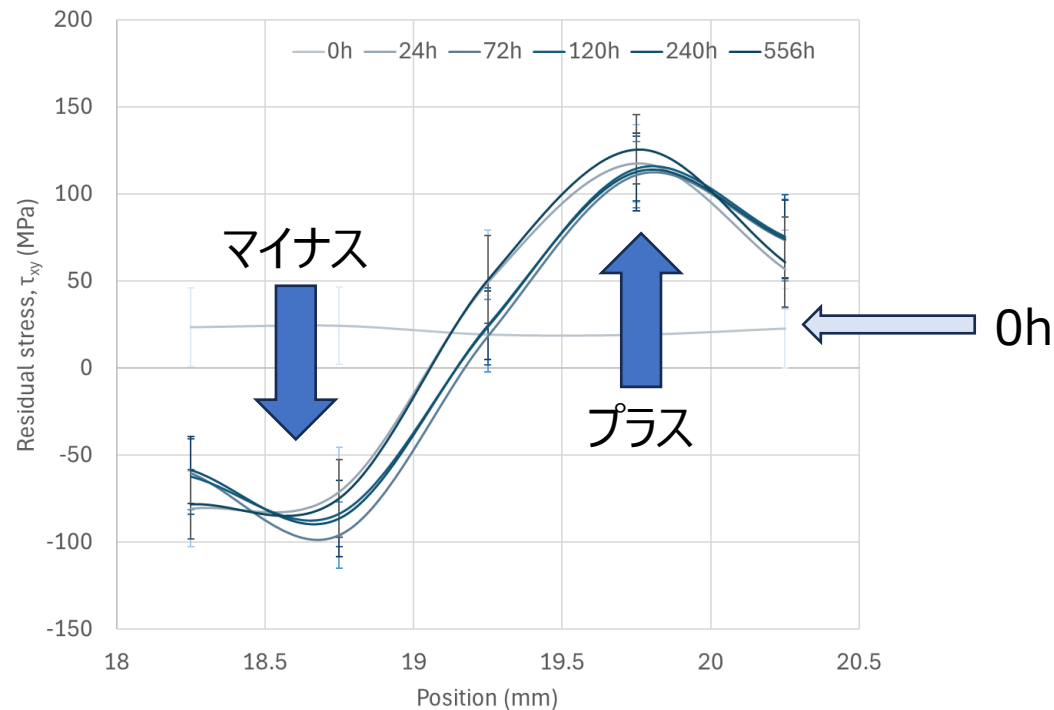


## ■ 回折環半価幅の時間変化を調査



■ 回折環半価幅は72hまで減少している。

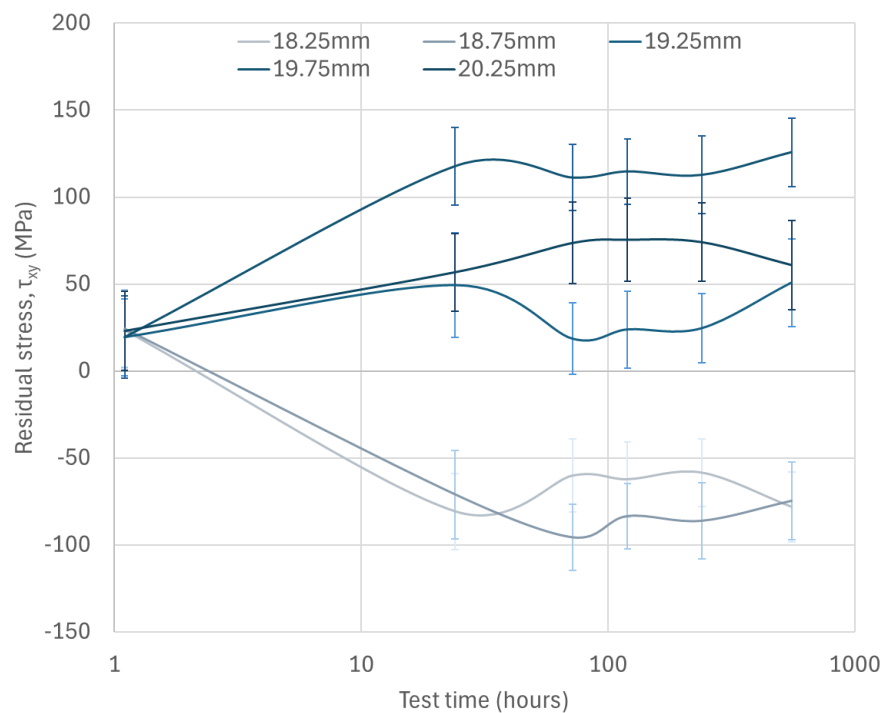
## ■ せん断応力 $\tau_{xy}$ の位置依存性を調査



■ せん断応力 $\tau_{xy}$ は24hまで大きく変化する。

■ 24h以降は、接触位置(19.25mm)の周辺で変化している。

## ■ せん断応力 $\tau_{xy}$ の時間変化を調査



■ せん断応力 $\tau_{xy}$ は24~72hまで変化する。

回帰分析とは

■結果となる数値(目的変数：試験時間)と要因となる数値(説明変数：応力値( $\sigma_x, \Delta\sigma_x, \tau_{xy}$ )、回折環半価幅(FWHM))の関係性を求めること。

線形回帰：変数の線形結合で表される。

■~~単回帰分析~~： ~~$y = a + bx$~~

■重回帰分析： $y = a + b_1x_1 + b_2x_2 \cdots + b_nx_n$

非線形回帰：変数の非線形結合で表される。

■ニューラルネットワーク回帰分析

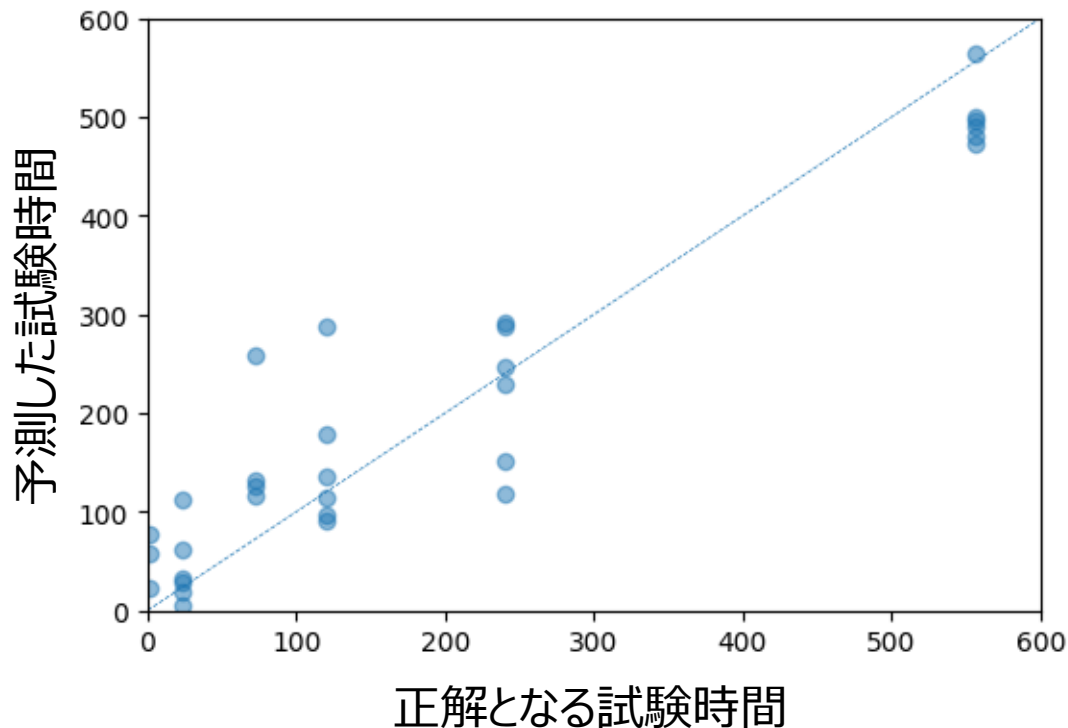
■ランダムフォレスト回帰分析

評価関数

■RMSE(2乗平均平方根誤差)

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{Predicted}_i - \text{Actual}_i)^2}{N}}$$

- 予測値した試験時間と正解となる試験時間の関係を確認した。
- 完全に予測できていれば比例関係になる。(斜めの線に乗る。)
- **RMSE=4897**



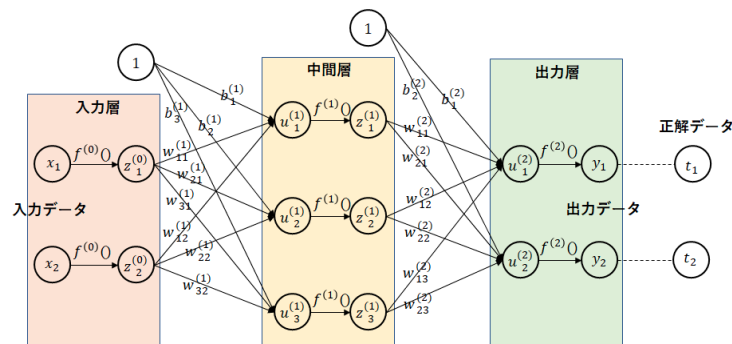
- 回折環半価幅の寄与が大きい

ニューラルネットワークとは

- 入力データ(行列 $z$ )に、重み $w$ を掛け、バイアス $b$ を足して、活性化関数 $f$ を掛け、次の層に出力。

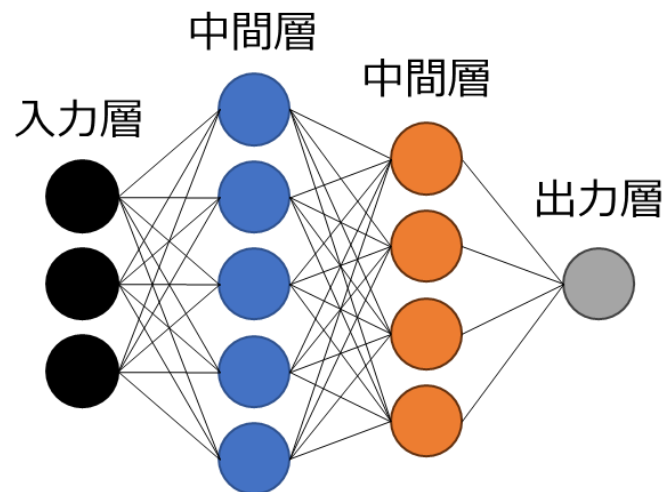
$$u = f\left(\sum (z \cdot w) + b\right)$$

- ハイパーパラメータ( $w$ や $b$ の学習率、 $f$ など)を調整して、損失関数を最小化する。



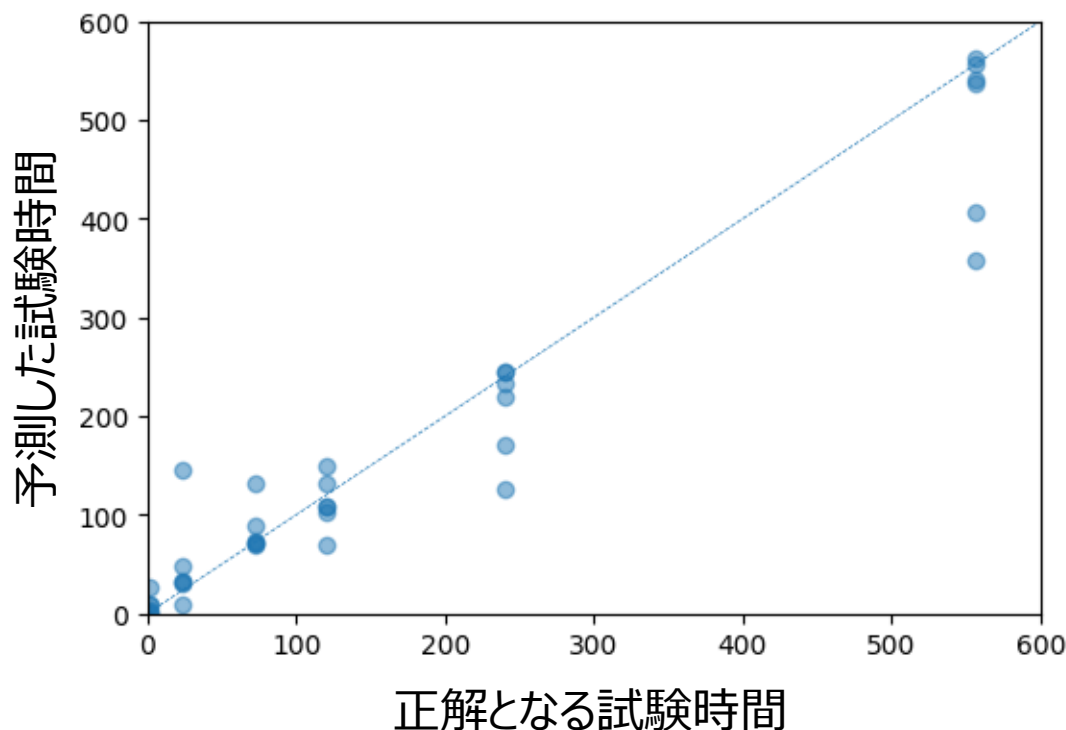
モデル

- 入力：20変数( $\sigma_x, \Delta\sigma_x, \tau_{xy}$ , FWHM×半径5点)
- 出力：1変数(試験時間)
- 中間層：2層(1層目32ノード、2層目16ノード)
- サンプル数：30(各試験時間あたり6データ)



■ 訓練データとテストデータに分け、訓練データで回帰分析を行った後、テストデータを用いて予測値を確認した。

■ **RMSE = 2916(10回平均)**



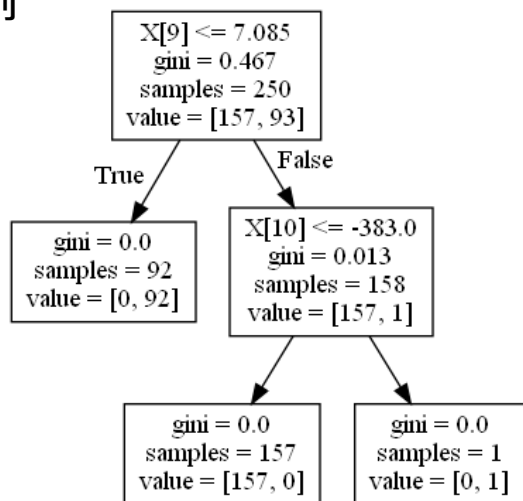
■ 重回帰分析よりRMSEが小さく精度が良い。

## ■ 決定木

- 最も精度が良くなるように条件分岐の条件を学習する。
- 6つの試験時間を上手く分ける条件を見つける。

上手く回帰できない。

例

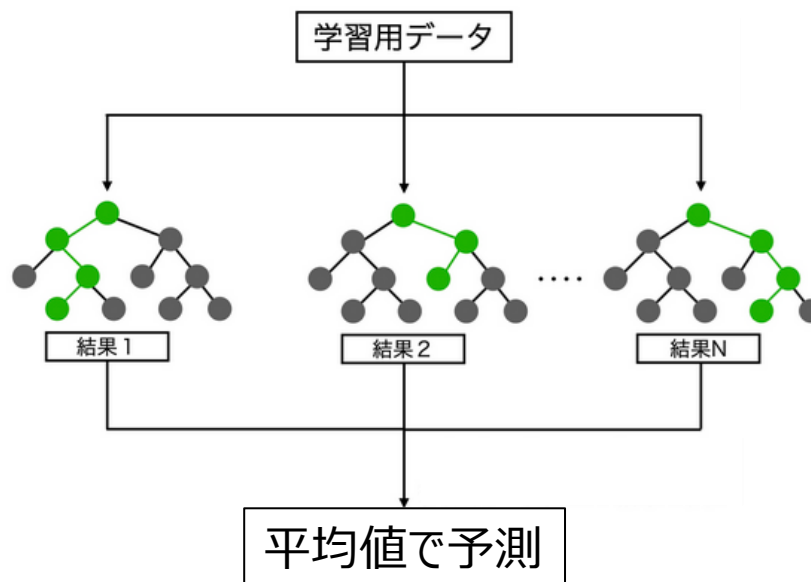


X[9] : 2点目の回折環半価幅  
X[10] : 3点目の残留応力 $\sigma_x$

## ■ ランダムフォレスト

- 複数の決定木の平均値を予測値とすることで精度を向上。

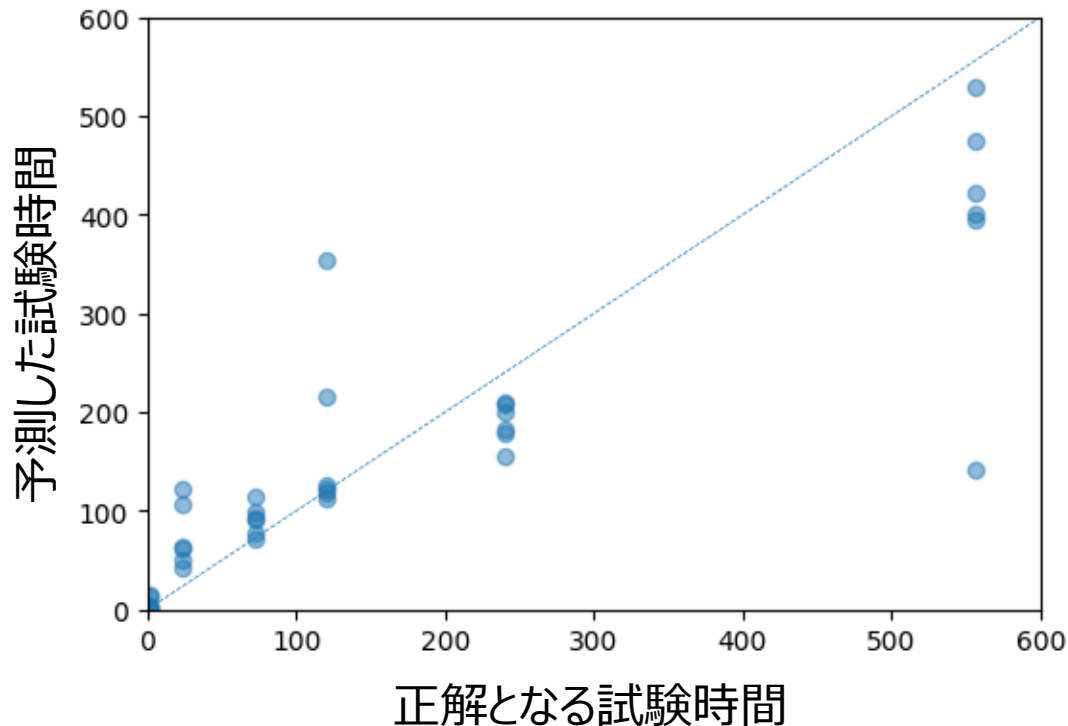
平均化することで連続値の予測が可能となり回帰分析できる。





■ 訓練データとテストデータに分け、訓練データで回帰分析を行った後、テストデータを用いて予測値を確認した。

■ **RMSE=7742(10回平均)**



■ 試験時間の長い領域の予測精度が悪い。

転がり疲労の定量的な評価のために、スラスト荷重試験を行い、残留応力・回折環半価幅測定を実施し、測定データの統計的解析・機械学習により、試験時間の回帰分析を行った。

接触位置の周りの測定データを用いて、重回帰分析、ニューラルネットワーク回帰分析、ランダムフォレスト回帰分析を行った結果、**ニューラルネットワーク回帰分析が最も精度が良かった。**

## 今後

- 新たな特徴量、3軸応力解析、残留オーステナイト測定、硬さ試験、回折環画像、金属組織画像を用いたマルチモーダル解析を行う。
- 抜き取り測定ではなく、多数の試験時間ごとに複数の試験片を用いて試験を実施して、統計的解析を行う。