Prediction of remaining useful time in turbofan engines

David Rais

Description of the assignment

- Four datasets, divided into *train* and *test* sets.

- *Training* sets: multiple run-to-failure trajectories
  - equivalent engines
  - starting always in *normal* condition
  - developing fault at random point of time

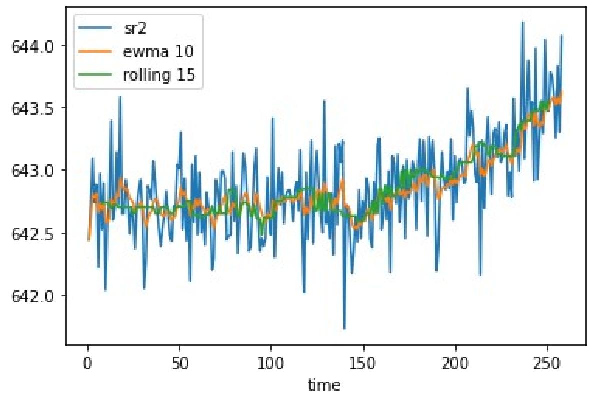
- *Test* sets: multiple truncated trajectories
  - The same parameters as train data
  - Truncated some at some time point before the failure
  - The true value of RUL for each trajectory is revealed
    - can be used for evaluation of the models

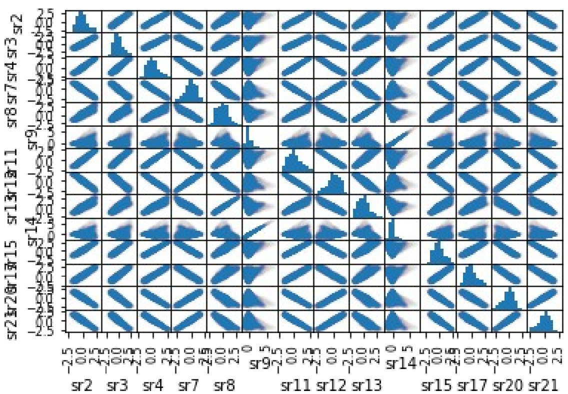
- Tasks:
  - Perform an exploratory analysis of the data
  - Develop a predictive model for the remaining useful life of the engine and report its accuracy in terms of R-squared and RMSE error.
  - Explain your process and chosen model.

Data exploration

Reveals the data are simulation of degradation of turbofan engines created in NASA laboratory

- Each trajectory consists of time series of 21 noisy sensor readings (SRs) and 3 operation conditions (OPs). Some of the SRs are steady and useless.
- The noise elimination by moving average showed an **exponential** development of a fail and high correlation

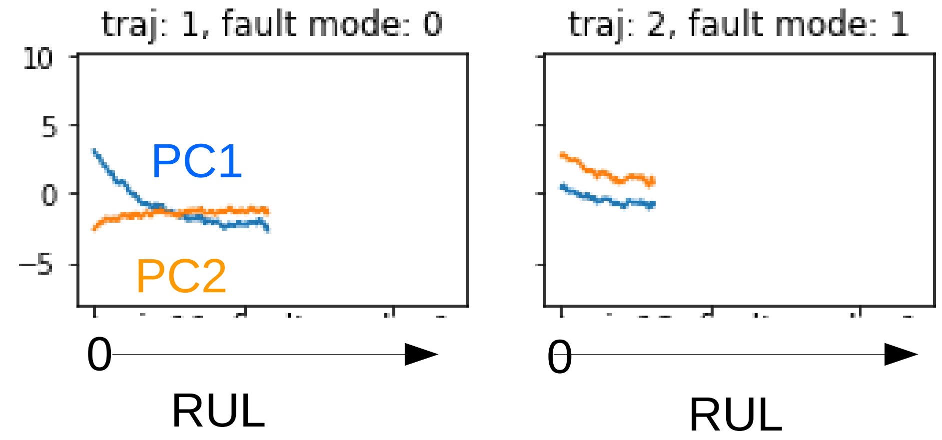




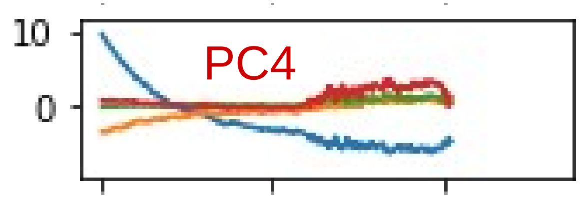
- Further data formating steps of the SRs:
  - **normalization**
  - **reduction** with Principal Component Analysis (PCA), transforms 21 SRs --> 4 PCs
  - **time reversal**: the time transforms in RUL, i.e. at the time of failure (RUL=0) higher.

Prediction of RUL in the dataset FD003

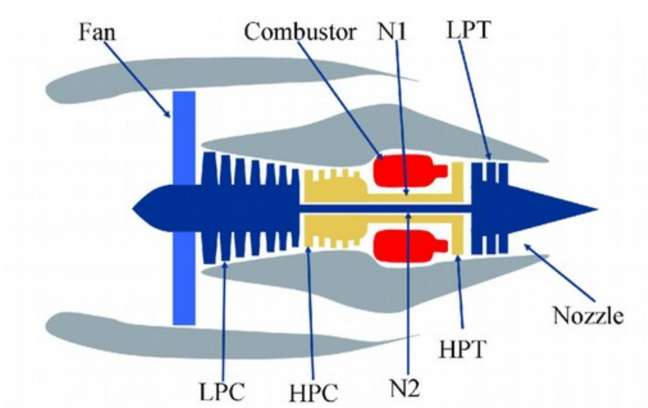
- 100 train trajectories, 100 test trajectories
- 4 PCs explain 97% variance
- 2 fault modes (FMs)
  - categorized by sign of PC2 slope (linear regression)



- PC1 and PC2: prediction of RUL in the test set
- high values of PC4 indicate “burn-in” period of normal operation this part was removed



- The RUL prediction was performed separately for the two fault modes (HPC degradation and Fan degradation)

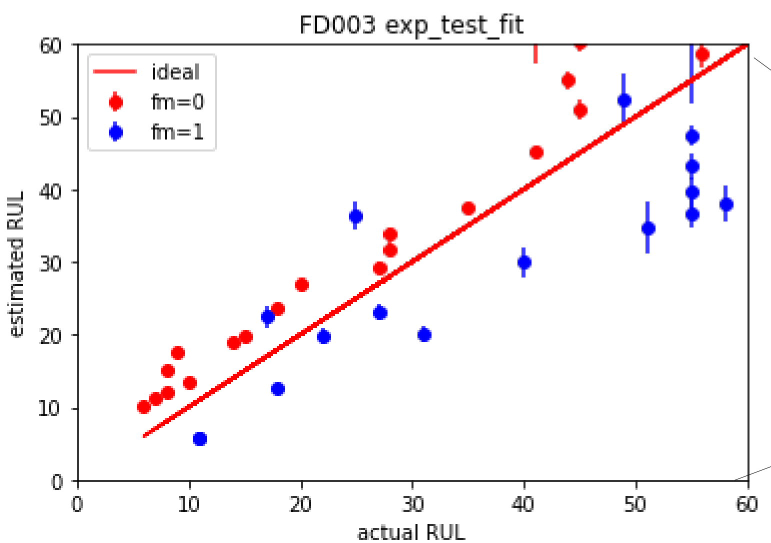


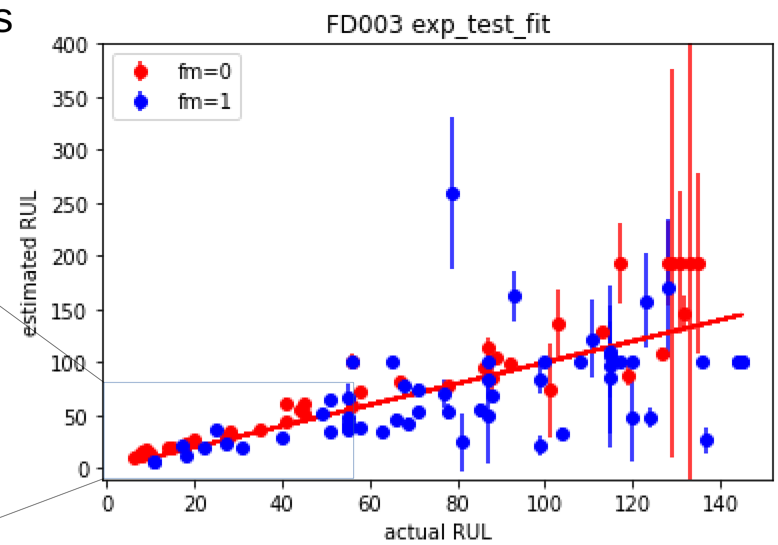
- Two models for RUL prediction were compared
  - **Non-linear regression** by an exponential function
  - **Artificial Neural Network**

Prediction by the non-linear regression

- Learning phase:
  - mean value of PC1 at the point of the failure (per FM) using the *train* set
- Prediction phase:
  - non-linear regression of trajectories in *test* set (per FM)

    PC1(*tr*)= A exp(-*k* (*tr*-RUL)) + *y*
- Result:
  - the model yields **very good predictions** for engines near failure, with RUL > 60 cycles
  - large uncertainty in the fit of RUL parameter indicates imprecise prediction

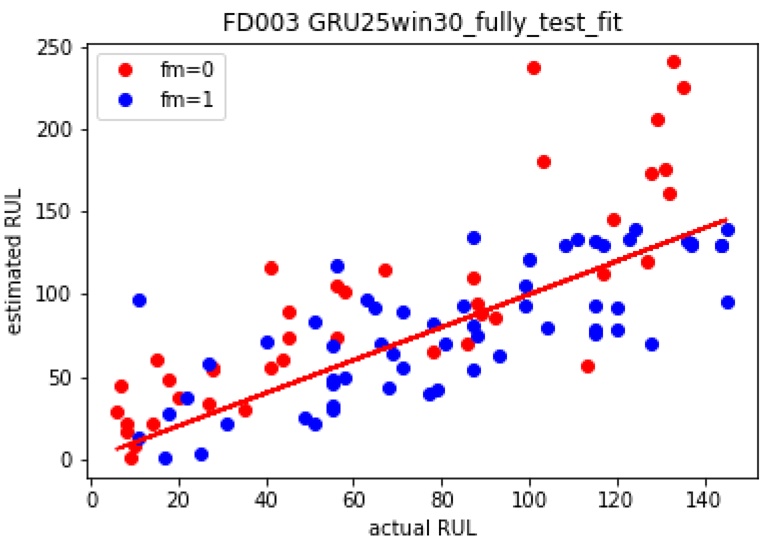




RMSE = 46.8 cycles

Artificial Neural Network (ANN) prediction

- Training phase per FM:
  - the *train* set trajectories (PC1 and PC2 only) chopped to short segments shown to the network, alongside with their value of RUL
- Evaluation phase:
  - the trained network was used to predict the RUL values for the *test* set:
- Result:
  - the prediction errors are showing rather uniform distribution, with low dependence on the value of actual RUL
  - The predictions are well correlated with the actual RUL.
  - the training and evaluation require decent computation resources
  - the precision is limited by the task budget



RMSE = 35.7 cycles,

Summary

- The non-linear regression model is very accurate for engines near the system failure.
- The ANN is better in engines that are mostly healthy. The model can be improved if the project budget is increased.
- Using each model in its own region of seems to be good compromise, when precision is crucial, with the current budget limitations.

model	RMSE	Notes
NL regression	46.8 cycles	fast computation, excellent accuracy for RUL<60 cycles
ANN	35.7 cycles	demands more computational resources for training,