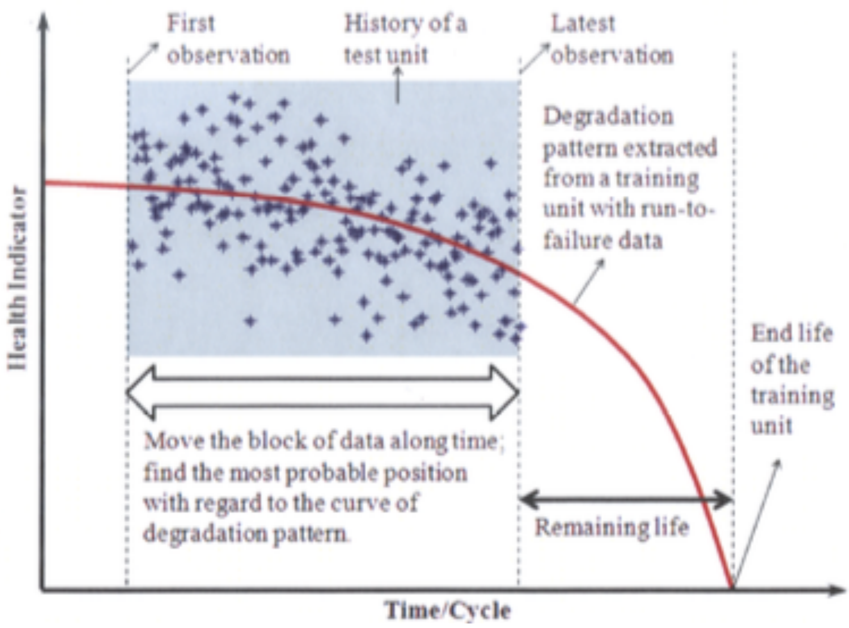
TITLE : RUL estimation of an aircraft turbine engine

INTRODUCTION

i) **Run-to-failure historical data from multiple units** of a system/component are recorded
ii) The historical data covers a representative set of units of the system/component;
iii) The history of each unit ends when it reaches a failure condition, or a preset threshold of undesirable conditions, after which no more runs will be possible or desirable.

Then, a **library of degradation patterns can be created** from these units with complete run-to-failure data (called training units). A unit whose remaining life will be predicted (called a test unit) also has its historical data recorded continuously. Instead of fitting a curve for a test unit and extrapolating it, the data will be matched to a certain life period of certain training units with the best matching scores. Finally, the RUL of the test unit can be estimated by using the real life of the matched training units minus the current life position of the test unit.



The remaining life of a test unit is estimated based on the actual life of a training unit that has the **most similar degradation pattern.**

METHODOLOGY

PERFORMANCE ESTIMATION

The multi-dimensional sensor readings, as well the features extracted from the raw sensor data, are first fused to produce a single **Health Indicator** (HI). The process to achieve this is called performance assessment.
We use linear multivariate regression model for estimating the HI.

$$y = \alpha + \boldsymbol{\beta}^T \cdot \mathbf{x} + \varepsilon = \alpha + \sum_{i=1}^{N} \beta_i x_i + \varepsilon$$

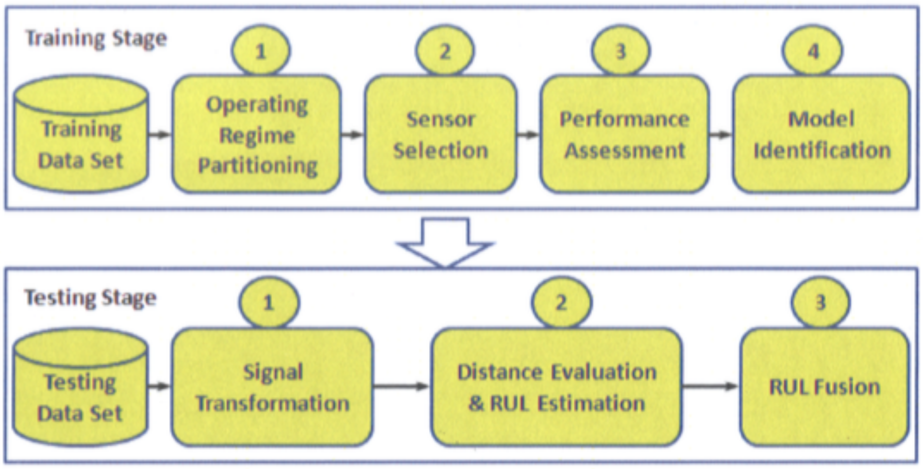
RUL ESTIMATION

First derive multiple representative **degradation models** from the training units, find the models with **similar degradation patterns as the test unit** and use them as the basis for RUL estimation. The model with the most similar degradation pattern is given the highest preference

EXPERIMENTAL DATA

The data set, provided by the 2008 PHM Data Challenge Competition, consists of multivariate time series that are collected from multiple units of an unspecified component. Each time series is from a different instance of the same complex engineered system, e.g., the data might be from a fleet of ships of the same type. There are three operational settings that have a substantial effect on unit performance. The data for each cycle of each unit include the unit ID, cycle index, 3 values for the operational settings and 21 values for 21 sensor measurements. The sensor data are contaminated with noise.

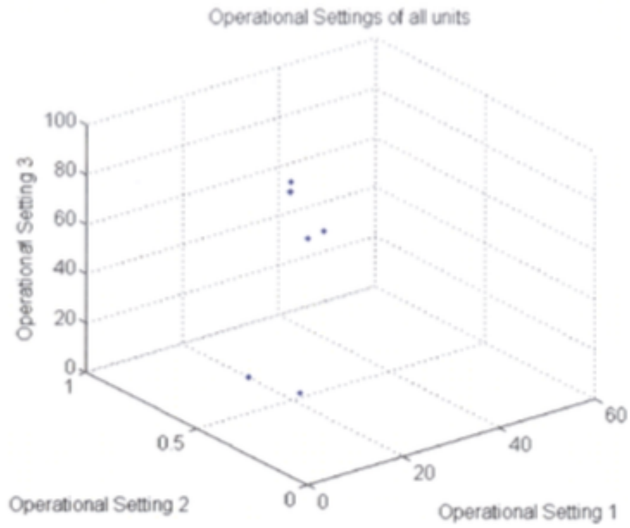
Each unit starts with different degrees of initial degradation and manufacturing variation which is unknown. This degradation and variation is considered normal. The unit is operating normally at the start of each time series, and develops a fault at some point during the series.



COMPLETE PROCESS MODEL

OPERATIONAL REGIME PARTIONING

Based on Operational settings the dataset can be divided into 6 modes as is clearly visible in the 3D plot without the use of any advanced clustering techniques



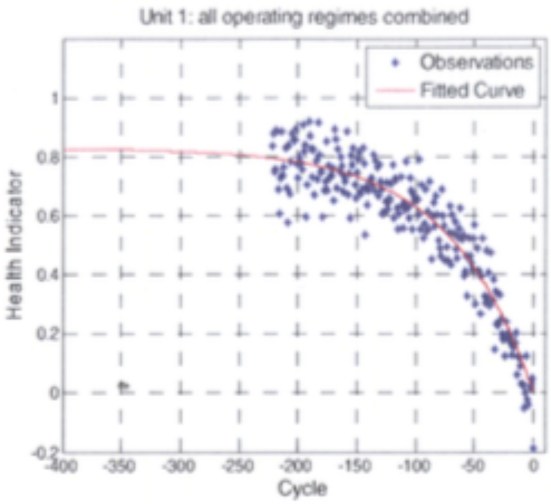
SENSOR SELECTION

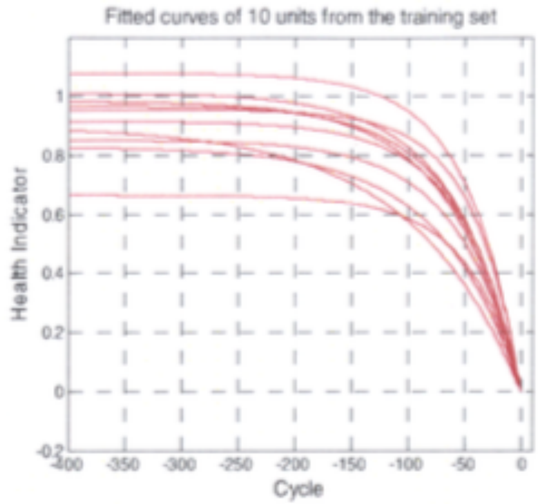
A few sensors have single or multiple discrete values, from which it is hard to find trace of system degradation. Most of sensors with continuous values exhibit a monotonic trend during the lifetime of the units. However, some of them show inconsistent end-life trends among the different units in the training data set which might indicate, for example, different failure modes of the system. It might be possible to first classify the units by failure modes based on these sensors and then process them using different prediction models; this strategy, however, will encounter two challenges. First, the end-life readings of these sensors spread out over a large range, which make it hard to quantize the failure modes without extra information. Second, the failure modes might not be unambiguously identifiable, if not completely indiscernible, at the early age of a unit, and thus might contribute little to RUL estimation when only early history of the unit is available. Therefore, only those continuous-value sensors with a consistent trend are selected for further processing. Theses sensors are indexed by 2, 3, 4, 7, 11, 12, 15, 20 and 21.

MODEL IDENTIFICATION

the exponential (nonlinear) regression models are used to describe the relationship between the adjusted cycle index and the HI *y:*

$$y = a(e^{bC^{adj}+c} - e^{c}) + \varepsilon$$





As can be seen, the fitted curves have exponential degradation pattern for the HI values, comprise a set of models that are used for finding the RUL of the test data set based on similarity measure calculated between the test data and the set of fitted models, a simple euclidean distance can be used as a similarity score

For the test data, first it is classified based on operating regime and then transformed to HI values using the performance assessment models trained on the training dataset, then the similarity is measured between the set of models and the test unit is calculated for estimating the RUL.