



Segment B (Predictive Modelling) – CA

EE5110 Special Topics in Automation and Control

Predictive Maintenance for Flight Engines with Python

Report Submitted By

SARAH REDDY

(Metric No. - A0146476W)

Mail ID - E0010686@u.nus.edu

**MSc Electrical Engineering, Dept of Electrical and Computer
Engineering**

National University of Singapore

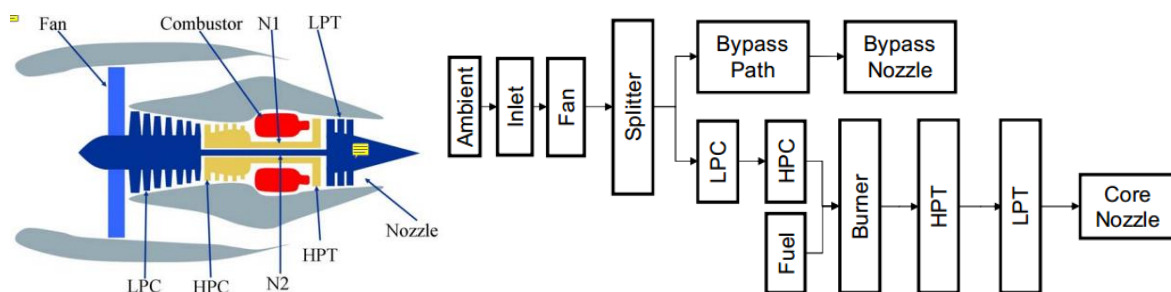
1. Introduction

Predictive Modelling – Damage Propagation

Predictive modelling makes use of statistics to predict the outcomes. Often the event that is to be predicted occurs in the future, but regardless of when an event occurred, predictive modelling can be applied to any sort of unknown event. Predictive modelling can be used to predict the failure of a machine based on the past data. Most often the model is chosen based on detection theory, to try to guess the probability of an outcome given a set of input data. Any regression model can be used for the purpose of event prediction.

Airlines are looking for means to reduce the cost for maintenance, and also to minimize unexpected maintenance issues. Maintenance delays cause a lot of inconvenience and glitches cost the airlines billions of dollars. Hence sensors data from flights which has experienced unexpected maintenance issues are collected and analysed to find any pattern that could indicate a problem or a failure approaching, so that the issues can be fixed before they lead to a failure and a delay.

Predictive modelling is used for damage propagation modelling for aircraft engine run-to-failure simulation. The data from the sensors is used to predict the remaining useful life of an aero-propulsion engine and also to predict if the system will fail within a specific time frame. From the time evolution a target signal is generated and given to Lasso model and Random Forest model, which helps them to predict the remaining time to failure. The diagram below shows the main elements of the engine and the flow chart shows how the various elements are assembled.



Objective

The objective of this assignment is to learn how to process the signal data collected from flight engines and to build a predictive model to predict the remaining timespan of the engine. The task given is to predict how long it will be before the engine fails, based on data from multiple sensors that predict its status. A sample data processing and analytics code is provided in Ipython. The code is to be run for one engine fleet FD-001 to understand smoothing technique and 2 analytical models: Lasso and Random Forest. The knowledge is then supposed to be applied to a new tool fleet FD-003.

2. Question 1

We have training and testing data from 4 machine fleets (one fleet is one group of machines), FD001 – FD004. From your opinion which scenarios should we build a model for; all machine fleet or build individual models for each fleet?

The training and testing data from 4 machines fleets have been provided. The machine fleets FD001-FD004 are of the same type. From my opinion the scenarios when we should build a model for; all machine fleets or build individual models for each fleet is explained below

Scenarios when we should build a model for all machine fleets:

1. In this scenario a particular engine fleet may undergo different flight conditions from one flight to another. Depending on a lot of factors the amount type and rate of accumulation of damage will vary for each engine. In the paper “Damage Propagation Modelling for Aircraft Engine Run-to-Failure Simulation Abhinav Saxena, Kai Goebel, Don Simon, Neil Eklund” it is assumed that the amount of damage that is accumulated during a particular flight will not be directly quantifiable solely based on the flight conditions and the duration of the flight, and hence we must rely only on the information extracted from the sensor data that is collected during each flight. One measurement snapshot per flight is made use of in order to characterize the engine health during and after a flight, while treating the data collected during the flight as noise (during a flight a 787 can generate up to 1Terabyte of data). Since we are relying solely on the sensor data and not on any other parameters (such as noise and initial wear) to build a dynamic model, this same model can be used for all the engine fleets. The number of data collected for each fleet may be different, but it’s the quality of the data that is of importance.
2. We can use the same predictive model for flights which are used for the same purposes; for example, we can use the same model for all commercial passenger flight with the same capacity for number of people.

Scenarios when we should build individual models for each fleet:

1. In order to make the model more realistic we will need to consider the initial wear and the noise in addition to the sensor data. The initial wear will differ from machine to machine, and can occur due to manufacturing inefficiencies. This initial wear can make a difference in the operational lifespan of the component. Noise is another factor that can make a difference in the useful life of a component. There are many sources of noise which contribute at different stages of the process, that can be during a flight, and it affects the sensor measurements. Performance parameters like efficiency and flow may not change monotonically to model process noise that also incorporates between flight maintenance operation, which can even lead to a better performance in the following flights. When we take all this into consideration to build a model, we will need to build individual models for each machine fleet, since they will all have different initial wear and noise sources.

2. Another scenario when we should consider individual models is when the purpose of the flight is different. For example, we cannot use the same model for a cargo flight and a commercial passenger flight.

Since we know that the engine fleets FD001-FD004 are of the same type, we can use the same predictive model for all the engine fleets based solely on sensor data.

3. Question 2

We should smooth out the sensor signals to remove measurement noises. Should we do that for operational settings too?

In signal smoothing the original data points of a signal are modified and the individual points are reduced and replaced by related approximate points to form a smoother plot. Smoothing the original data set helps to create an approximate function that helps to detect important patterns in the data while removing the noise.

The method used for smoothing is seasonal decomposition. Seasonal decomposition is an algorithm that is used to decompose a time series data into 3 components i.e., trend, seasonal, remainder. Each of these components represent one of the underlying categories of patterns that for the original data. These can be used to reconstruct back the original data. Each of these components have a specific behavior or characteristics.

1. Trend- The trend component at time t represents the long term progression of the data series. This trend is significant when there is an increasing or decreasing pattern in the data.
2. Seasonal- This component indicates the seasonality, for example the seasonal changes that may affect the data. It is always of a fixed and known period.
3. Remainder or irregular component- This represent the noise present in the data. It describes the random and irregular influences on the original data. This consists of the remainder of the time series data after the seasonal and trend components have been removed.

The main objective of using seasonal decomposition is to estimate the seasonal effects. Removing the seasonal effects from the sensor data can show the trend of the data more clearly. This trend is indicative of direction the sensor values are going. As discussed in the section above, one measurement snapshot per flight is taken to characterize the engine health during or immediately after the flight. The effects of the during-flight-maintenance is not modeled and is incorporated as process noise. It is necessary for this noise to be removed in order to understand the trend of the sensor data.

Based on the same reasoning and analysis, it is therefore necessary that the data representing the operational settings have to be smoothened too, in order to observe the trend and get rid of the during-flight maintenance noise. Operational setting have a huge influence on the measurements or the feature extracted from the system, so that the time series plot of the

features will show large variations which overwhelm the degradation which causes the trend. This will create a difficulty in predicting the degradation trend. Hence the data that has been collected with varying operational settings have to be smoothed and preprocessed.

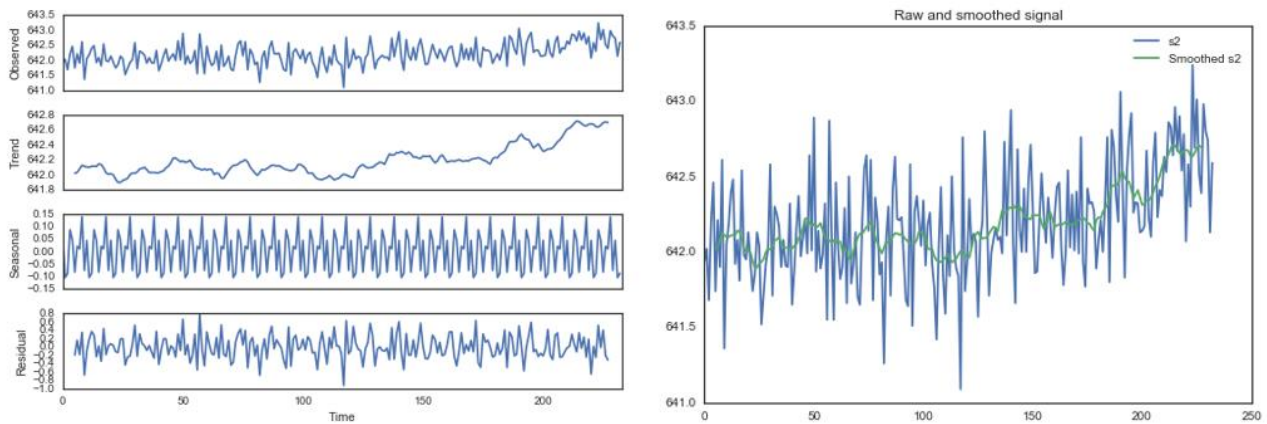


Figure: The above figure shown the decomposition of a sensor from the engine fleet FD003

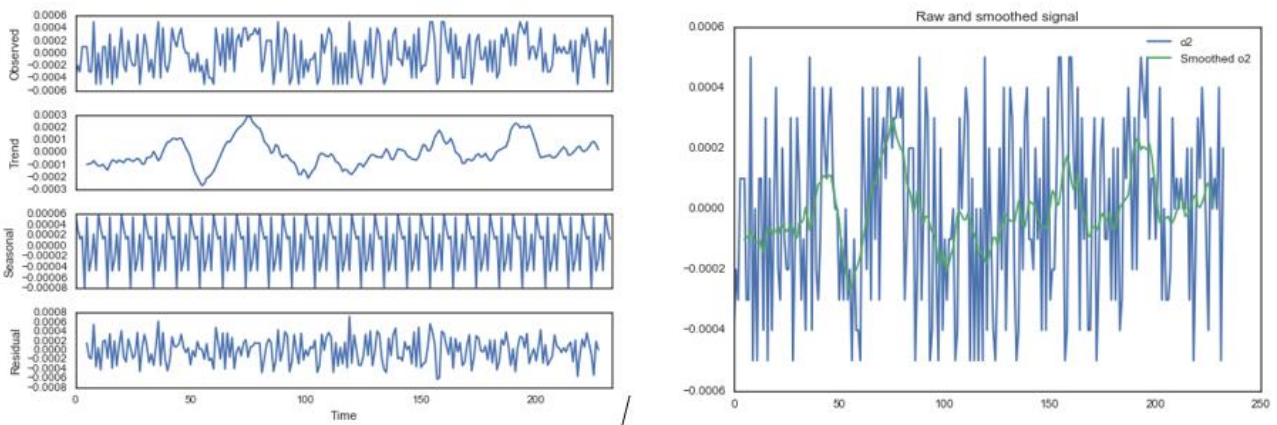


Figure: The above figure shown the decomposition of an operational setting from the engine fleet FD003

4. Question 3

From the time evolution, we need to construct a target signal for our model to predict. We have 2 options:

- 1. Remaining time, which decreases linearly.**
- 2. Decelerating health index, which decreases exponentially.**

Could you give a short discussion on the advantage and disadvantage of each?

Any phenomena in a system can be modeled using the health index. In this case. The health index of aircraft engine modules such as the compressor and the turbine, are described by

efficiency (e) and the flow (f). For various flaw modes the trajectories for flow and efficiency vary, and they are modeled as separate health indices. The health index model for efficiency and flow are as given below:

$$e(t) = 1 - d_e - \exp\{a_e(t)t^{b_e(t)}\}$$

$$f(t) = 1 - d_f - \exp\{a_f(t)t^{b_f(t)}\}$$

The overall health $H(t)$ is formed by aggregating $e(t)$ and $f(t)$. $H(t)$ represents that engine simulation response to the given input values. $H(t) = g((t), f(t))$

Where g is the minimum of all operative margins considered (HPC, HPT and EGT)

$$g(e(t), f(t)) = \min(m_{Fan}, m_{HPC}, m_{HPT}, m_{EGT}),$$

where m is a function of efficiency and flow.

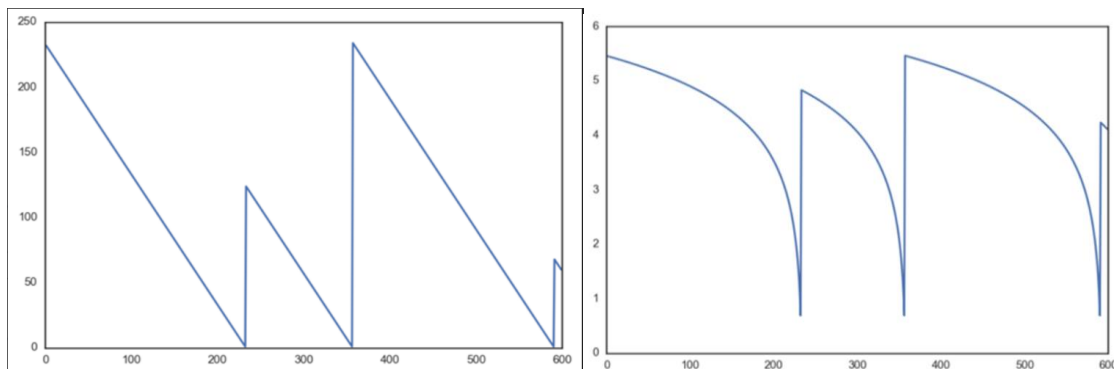
From the above discussion we can see that the health indices are being modeled exponentially.

Advantages and disadvantages of constructing a linear target signal:

We can see that the health indices can be modeled effectively using exponential model. Real time health indices are not linearly decreasing, there are many factors contributing to the health. Hence modeling a target signal linearly will not be a realistic representation of a real time scenarios, since real time scenarios will have a mixture of increasing or decreasing trend, and a linear model will be too much of an approximation. The advantage of using a linear target signal is that, it provides an early prediction of impending failure. An early prediction is always better than a late prediction. In cases where the data in-between is considered as noise, a linear target signal is more than enough to help with prediction.

Advantages and disadvantages of constructing an exponential target signal:

An exponential target signal is more realistic when considering real time data. We have seen that in this case, the health indices of aircraft engine modules such as the compressor and the turbine have been modeled exponentially. Hence with the help of an exponential target signal, the lasso or random forest model will be able to effectively predict. One disadvantage of using exponential target signal is it gives a comparatively late prediction when compared to using a linear one.



The 2 plots shown in the previous page are the linear and exponential target signals. These indicate the remaining useful life. The x-axis shows the time evolution, and the start of each new cycle indicates that once maintenance is done, the RUL is reset.

5. Question 4

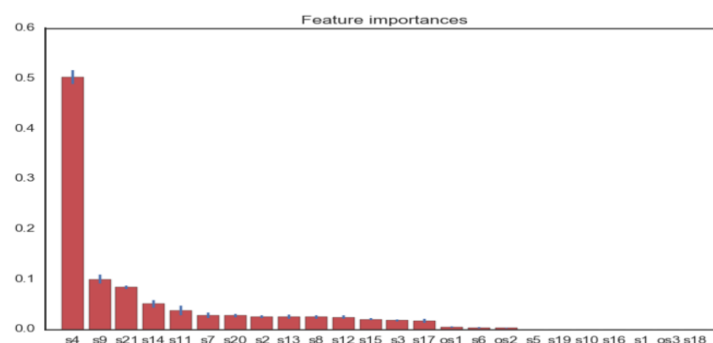
Which sensors can show the degradation of the propulsion engines? Any uncertainty? What is the physical meaning of these sensors?

The aircraft engine consists of built-in control system which comprises of a fan-speed controller, and a set of limiters and regulators. The limiters include 3 high-limit regulators that prevent the engine from exceeding its design limits for core-speed, engine-pressure ratio, exit temperature of high-pressure turbine(HPT); A regulator that prevents the static pressure at the high-pressure compressor (HPC) exit from going too low.

The C-MAPSS data has 14 inputs, these inputs include fuel flow and set of 13 health-parameters inputs. This allows the user to simulate and check the effects of the faults and deteriorations in any of the engine's five rotating components. The 5 rotating components are Fan, LPC, HPC, HPT and LPT. This list of parameters to measure the system response is given below

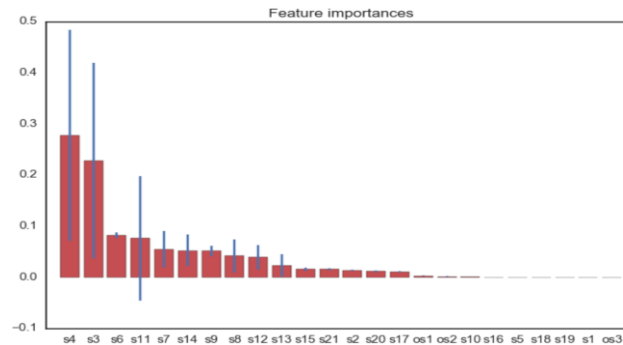
Symbols	Description
T2 (S2)	Total temperature at fan inlet
T24 (S3)	Total temperature at LPC outlet
T30 (S4)	Total temperature HPC outlet
T50 (S5)	Total temperature at LPT outlet
P2	Pressure at fan inlet
P15	Total pressure in bypass-duct
P30	Total pressure at HPC outlet
Nf	Physical fan speed
Nc	Physical core speed
epr	Engine pressure ratio (P50/P2)
Ps30	Static pressure at HPC outlet
phi	Ratio of fuel flow to Ps30
NRf	Corrected fan speed
NRc	Corrected core speed
BPR	Bypass Ratio
farB	Burner Fuel-air ratio
htBleed	Bleed Enthalpy

FD001 – Train: On running the experiment of the machine fleet FD001 training data we observe the following results.



From the above plot obtained from running the Random Forest model, we can see clearly that the sensor S4 is responsible for the degradation of the propulsion engines. We know that sensor S3 is responsible for measuring the total temperature at the HPC outlet, i.e., T30. This indicates a failure in HPC.

FD003 – Train: On running the experiment on the machine fleet FD003 training data we observe the following results.

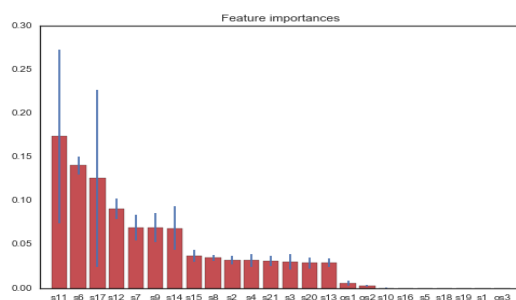


From the plot shown above which is obtained by running the Random Forest model on the FD003 training data set, we can clearly see 2 sensors majorly responsible for the degradation of the propulsion engine. These sensors are S3 and S4 which are responsible for the measurement of the total temperature at the LPC outlet and the HPC outlet respectively i.e., T24 and T30. This can indicate 2 types of failures, described as follows:

1. This can be an indication of both HPC and LPC failure, OR
2. From the flow chart shown under the introduction, we can see that a failure with the FAN can also cause both these sensors to fail.

All these predictions have some degree of certainty in them. We are just predicting what may happen in the future depending on the present and past data that has been collected and analyzed. As in the case of FD003 train data, we see 2 sensors which contribute a lot to the degradations, but we cannot predict for sure if it is the temperature sensors failure (mode failure) or if it is due to Fan failure.

FD003 – Test: On running the experiment on the machine fleet FD003 testing data we observe the following results. From the plot shown below which is obtained by using the random forest model, we see that the sensors S11, S6 and S17 have a significantly more influence on the degradation of the propulsion engine. These represent failures in Engine pressure ratio, Pressure at fan inlet and Burner fuel-air ratio. It is given that Engine pressure ratio is a function of the pressure at the fan inlet. Hence this may be a major cause of failure in the future.



6. References

1. <http://blog.revolutionanalytics.com/2016/05/predictive-maintenance-r-code.html>
2. https://etd.ohiolink.edu/!etd.send_file?accession=ucin1282574910&disposition=inline
3. <https://azure.microsoft.com/en-us/documentation/articles/cortana-analytics-playbook-predictive-maintenance/>
4. www.wikipedia.com
5. "Damage Propagation Modelling for Aircraft Engine Run-to-Failure Simulation
Abhinav Saxena, Kai Goebel, Don Simon, Neil Eklund"