PREDICT FUEL FLOW RATES OF AIRPLANES

A PROJECT PAPER

Submitted in partial fulfilment of the requirement for the award of the degree of

MASTERS IN DATA SCIENCE

In course

TOPICS IN OPTIMIZATION

by

MEHREET SINGH BAJAJ

1274698

New York Institute of Technology

DECLARATION

I hereby declare that the project paper entitled " **PREDICT FUEL FLOW RATES OF AIRPLANES** " submitted as part of the partial course requirements for the course ***Topics in Optimization*** , for the award of the degree of Masters in Data Science at New York Institute of Technology during the ***Ist*** semester, has been carried out by me. I declare that the project has not formed the basis for the award of any degree, associate ship, fellowship or any other similar titles elsewhere.

Further, I declare that I will not share, re-submit or publish the code, idea, framework and/or any publication that may arise out of this work for academic or profit purposes without obtaining the prior written consent of the Course Faculty Mentor and Course Instructor.

Signature of the Student:

Place: New York Institute of Technology

Date: 16.05.2020

16.05.2020

CERTIFICATE

This is to certify that the project entitled " **PREDICT FUEL FLOW RATES OF AIRPLANES**" is a record of bonafide work carried out as part of the course ***Topics in Optimization (DTSC-615)*** , under my guidance by ***Mehreet Singh Bajaj (1274698)*** , during the academic semester ***I st*** , in partial fulfillment of the requirements for the award of the Degree of Masters in Data Science , at New York Institute of Technology during academic year 2020.

PROF. JERRY CHENG
Project paper guide
New York Institute of Technology

ACKNOWLEDGEMENT

I wish to express my deep sense of gratitude to my guide Professor Jerry Cheng, Assistant Professor , Computer and Data Science department , New York Institute of Technology for his guidance and useful suggestions, for encouraging me to complete my work on time, without which it would've been impossible to complete this project paper, that too in time. I'm really thankful to him for believing in me and my project idea and helping me in shaping it up even better than I thought I could.

I feel blessed to have parents Dr Satinderjit Singh Bajaj and Dr Jagminder Kaur Bajaj . It was due to the intellectual discussion that they had with me all the time that always motivated me and inspired me to do something , and I could come up with the willingness to do this project.

I would also like to thank my cousin Preet Kanwal Singh who would keep me updated with the project requirements and help me out whenever I was stuck in a problem.

I would also like to mention my friend Hellen Viales and younger brother Prabh Arjun Singh Bajaj's little inputs about suggesting good modifications to the documents to make it look more appealing.

MEHREET SINGH BAJAJ (1274698)

Masters in Data Science

Topics in Optimization (1st semester)

ABSTRACT

Fuel is of utmost importance as it is one of the non renewable resources and using it in an efficient and smart way is highly required in today's date. This is the most common problem faced by the airline industry these days .Fuel constitutes around 30% of the operating cost of airlines due to which we have a higher price of tickets . Developing optimization strategies especially on fuel is of prime importance to airlines and reducing the emission of staggering amounts of greenhouse gases along with reducing fuel intake can have a significant positive impact on the environment .Given an idea of the utility of the resources used by my work so that no extra resources are used and may go in vain

Now the problem statement is that

The ability to predict the Fuel Flow (FF) rate of airplanes during different phases of a flight (Taxi, Takeoff, Climb, Cruise, Approach, and Rollout) will help understand

- The significant drivers of FF rate for each of these phases and also help understand the factors that make the airplanes perform at higher levels of fuel efficiency during the different phases of a flight.
- Insights from the exercise can help derive the best practices, which make flights more fuel efficient under different conditions.

After the project, using optimization techniques it was observed that climb, cruise and approach are the most important phases for optimizing fuel consumption. The most important features consisted of rate of change of altitude, longitudinal acceleration and ground speed. The clearest visible trend between predictors and fuel flow rate is in the climb phase. In other phases, some of the predictors have a weakly visible trend, but since the root mean squared error is small, it is assumed that the features have strong nonlinear interactions which are not clearly visible in simple plots.

All of the work was done in Jupyter Notebook using Python and with the concepts of supervised Machine Learning with various data cleaning techniques from tidying up the data to applying different models like Random Forests and eXtreme Gradient Boosting using various optimization techniques to get the best suitable results and then validation and checking performance using Root Mean Square error (rmse) method.

TABLE OF CONTENTS

Page No.

INTRODUCTION

1.1 Basic Introduction

In today's fast growing world in which we are surrounded by problems which can be solved easily using the knowledge of Artificial Intelligence and optimization techniques , we discovered one of the common problems which has been neglected for a long time and it really needs a solution. Using optimization strategies we can enhance the productivity and the output for the real life problems, while saving a lot of resources and using them in a very efficient and productive way .

Fuel Management is one of such problems which an individual faces in every day to day life. In the busy hustle bustle of life one does not notice that fuel plays an important role , majorly for travelling and as we know the most expensive means of travelling is via Airplanes. So we will try to understand why flight tickets are so expensive and how fuel is the major factor directly related to the cost . Next, from this article we will also come to know how fuel emission through airplanes emits greenhouse gases and how we can use the fuel efficiently (using optimization techniques) to produce less of those harmful gases and in a way providing a good solution to the problem which is environment friendly.

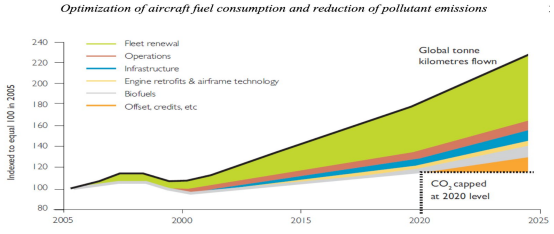


Fig. 1 Greenhouse gas emissions of the global aviation and development technology aiming to achieve carbon neutral growth by 2020 (IATA 2010a, b)

Fuel flow rate of an airplane is the key component in deciding the aircraft's engine performance. Also, the amount of gas it emits , decides the fuel used and directly relates to the cost . That means of the total cost which the airline handles, 30 % of it is just the cost of fuel. We can imagine how much fuel the plane consumes for one round , which can be reduced if used efficiently. So developing cost saving strategies especially on fuel is of prime importance to airlines and reducing the emission of staggering amounts of greenhouse gases along with reducing fuel intake can have a significant positive impact on the environment.

we will solve this problem with **optimization strategies along with the concepts of Data Science and Machine Learning** and train our machine with the flight record data to see how the target value (Fuel Flow (FF)) rate varies during different phases of flight, from where we will come to know about the main drivers which have a significant effect on fuel flow during the flight . This later on can help us derive the best practices which makes flight more fuel efficient during different phases of flight. So our main work is to derive the top features from these phases and later on we can do Exploratory Analysis on it.

We will build a statistical model using the machine learning techniques for aircraft engine data . All the variables used in our model are continuous and metric , our machine learning problem is basically a regression problem. We are using two models to solve our issue :

A. **RANDOM FOREST MODEL (using extra Tree Regressor method)**
B. **BOOSTING(XGboost - eXtreme Gradient Boosting)**

These algorithms have many widespread applications and help to conclude to a better result than others. XGboosting has been used in many competitions as it gives better results and very less error values, when compared to other models.

1.2 Scope of the Work

There are relatively few studies which have taken into account operational flight data to model fuel emission in planes, also applying such optimization strategies will improve the overall value and prove to be beneficial for the airlines industries and people in general as a cost saving strategy.. The integration of various types of operational data has been shown to improve the estimates of aircraft fuel consumption . Uncertainties in fuel flow rate values can be estimated by the Data driven models of engine fuel flow rate . With all this , our project Fuel Flow rate of airplanes is modelled with the help of operational data from our Flight Data Records (FDR). As all the readings including the different parameters for aircraft and engine are recorded with the help of sensors during the flight, FDR provides reliable real flight data . Due to random variations in the parameters and phases of flight our model is **stochastic** and not deterministic as a result of the manufacturing , flow turbulence, component deterioration and other internal errors and other external ambient disturbances .

"Predicting Fuel Flow of Airplanes" focuses mainly on finding the **top features** in different phases of flight responsible for high consumption of fuel during those phases and predicting the Fuel Flow rate in it . These top features will help us understand the nature of the flight . As most of the flight flies in between the same phases . In this project, to find the top features we have used various optimization techniques along with machine learning principles to airplane flight recorder data to model the fuel flow rate (FF) as a function of the aircraft altitude, ground speed, vertical speed, and takeoff mass in the airborne phases of flight. So overall if we come to know about these top features we can later do an exploratory analysis over it and derive best practices saving the fuel and benefitting airline companies in many ways .

REQUIREMENT ANALYSIS

2.1 Dataset

Our dataset consists of uncompressed data of about 14.4 GB of Flight record data (FDR) from 1005 different flight instances which represents readings of numerous sensors on aircraft and other engineering parameters. Every second measurements were done each of which contained 227 parameters. Flight data consisted of measurements during all the **8 phases** (unknown, preflight, taxi, takeoff, climb, cruise, descent and roll out) .

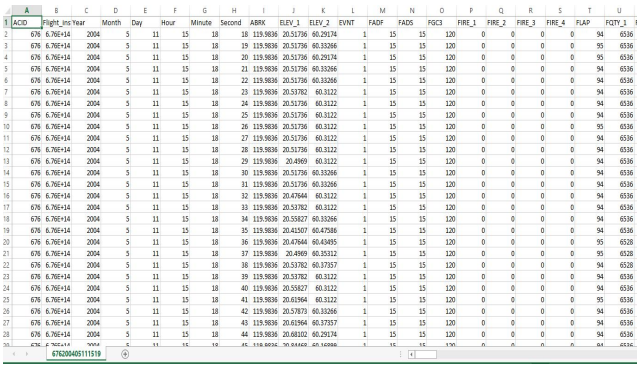
The dataset has been divided into training and testing set .The data distribution across training and test data sets is as follows:

- **Training dataset: 60%**
- **Test set: 40%**

There are 5 training data files each containing 200 csv files with record data for training and 1 testing file for validation of our model.

2.2 A Look At Data

This is what our dataset looks like containing **227 variables(columns)** and **4365 rows** . This is just one file. There are **200 CSV** files in one training file and there are **5 training files** each containing 200 files.



2.3 Methodology

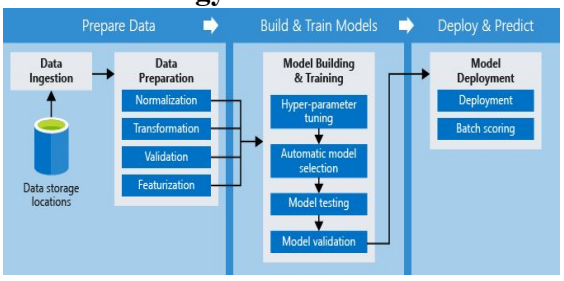


Fig 12 : The methodology diagram

This is the methodology which we will follow:

- **Data storage locations** : The recorded data or the raw data that is stored and will be used as the base for our whole project. More heavy the data is , the more trained our machine will be and more accurate the results will be.

- **Data ingestion** : Importing our set of data from the stored location to our main code which will further be processed.This includes steps like understanding our data, collecting it , describing the data, exploring and verifying the quality of it.

- **Data Preparation** : Consists of steps like selecting the data, cleaning it, constructing , integrating and formatting the data using technique like Normalization , Transformation, Validation and Featurization

- **Model Building & Training** : In this stage we will apply hyper parameter tuning (choose a set of optimal hyperparameters for a learning algorithm . It's value controls the learning process). Model selection(selecting various supervised algorithmic models which gives a better results to the given model), Model Testing(Testing the models and checking the errors with the rmse methods), scoring(applying those algorithmic models from historical data set to new dataset)

- **Model Deployment:** scoring (applying those algorithmic models from historical data set to new dataset in order to uncover the practical insights that help solving a problem) and finally deploying the model.

BACKGROUND OVERVIEW

3.1 Conceptual overview

Driving fuel efficiency involves developing strategies that touch upon various aspects of airplanes - broadly some of which are highlighted below:

- **Aspects related to Aircraft's actions on the ground** - e.g. include reducing taxiing times to reduce engine running times which translate into reduced fuel intake.
- **Aspects related to route planning** – e.g. taking shorter routes when inflight to destination taking in to consideration any altitude restrictions that exist.
- **Aspects related to aircraft design** – e.g. improving aerodynamics, redesigning aircraft components to conserve fuel or reducing the weight on board like installation of lighter seats.

There are different phases in flight during a flight instance. All the readings from the sensors fit into the airplane are taken from these sensors every second and saved in flight record data .

There are mainly 7 phases of flight and one unknown phase which holds the data which was not recognised to be from any phase .Here is the description of the phases of the flight from 0 to 7 .

Variable PH in our dataset refers to the phase of flight .

3.2 PH TABLE

PH	*PHASE NAME*	*DESCRIPTION*
0	**Unknown**	Consists of those readings in which phase cannot be determined
1	**Preflight**	Phase before the flight includes some checkups of airplane and ensure everything is working fine
2	**Taxi**	This phase refers to the estimated time an aircraft spends taxing between it's parking stand and runway or vice versa
3	**Takeoff**	Fuel required to takeoff the plane
4	**Climb**	phase pointing to increasing altitude of an airplane
5	**Cruise**	Phase when an aircraft levels after a climb to set altitude before it sets to descend
6	**Approach**	A final approach is the last leg in an aircraft is lined up with the runway and Descending for landing
7	**Rollout**	The phase of an aircraft's landing during which it travels along the runway while losing speed

3.3 HOW TO OVERCOME THE PROBLEM ?

The Flight Record Data (FDR) can be used to

- Understand which features of dataset are correlated with Target(FF) we are trying to predict during different phases of a flight (Taxi, Takeoff, Climb, Cruise, Approach, and Rollout)
- Derive best practices that make flights fuel efficient under different conditions?
- We can later on do Exploratory Analysis on these top features

APPROACH

This is just a short overview of what strategies have been used to solve the problem . The detailed description will be in the code and explanation section.

4.1 Importing libraries

Firstly we imported some of the basic important libraries used in python like numpy, pandas, glob, os, sklearn, matplotlib and seaborn for enhanced visualization of our results using plots.

4.2 Loading into Pandas

Some pre-processing steps included downloading the data and loading it. , pre-processed to make it smaller and faster to load. All the CSV(comma separated) files were loaded into pandas dataframe and concatenated. After this the 64-bit data-types (int64,float64) were converted to 32-bit data-types (int32, float32). Finally, the combined Data Frame was saved as **pickle** files. The total size of train and test data pickle files on disk was about 6.5 GB.

4.3 Tidying the Data

Now , to check the the basic statistics of the data and separate the unwanted data from the important data DataFrame.describe() method was used and some variables showing no variance were observed and dumped using VarianceThreshold method and all N/A if were present were also thrown out.

4.4 Visualization

We compared fuel flow across various flight phases (There are 7 phases) excluding Unknown Phase. As different phases have very different fuel flows.We also segregated why few flight instances are different from others in terms of Fuel Flow.

We have 600 flight instances . Data visualization was made clear using distplot and boxplot.

Then using FacetGrid function we obtained fuel flow plot for every phase on a separate axis .

Violin Plot and swarmplot showed the spread of fuel flow in the cruise phase was very wide .

Then compared and visualised Flight_instance with FF using visualization through boxplots. As the

instances were very large in number it became difficult to visualise.

4.5 Splitting the data

Now the sklearn train_test_split , cross_validation K Fold strategy was used here to split data into pairs into groups .

4.6 Checking the Errors

Using the Root Mean Square Error method the more validation error that comes up the poorer the model is .

4.7 Models

Like data visualization techniques can be applied to present data , optimization techniques can be implemented to plan for actionable conclusions . We have used two models here for better results .Optimized Machine Learning Algorithms like ExtraTreesRegressor (using Decision tree regression strategies) and eXtreme Gradient Boosting have shown amazing results in many competitive coding contests and are known for their less error in rmse and fast computation. In the end one with less validation error was considered a better model .

4.8 Finding the Top Features

Now top features were found using both the models using the **plot_importance method** we created which was using **.feature_importances method**. Then the feature ranking was done, from where we come to know that PH(phase of flight) is one of the most important variable

4.9 Finding the correlation using Heatmaps

TO find the correlation between top features including the FF variable with each other we get the correlated features and remove highly correlated ones.

From this we will also get other top features which are important for minimizing the Fuel Flow rate.

DIFFERENCE IN VARIOUS PHASES OF FLIGHT

There are seven different phases of the flight in this dataset, marked by the values of 'PH' 0 to '7' a 'PH' value of 0 indicates that the phase is unknown.

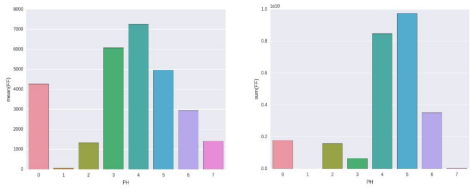


Fig 1 The mean fuel flow rate (left) and total fuel consumption (right) in each phase of flight. From these graphs, phase 4 and phase 5 appear to be the most important phases that drive the fuel consumption in an aircraft.

The plots in fig 1 show mean and total fuel consumption in different phases of flight.

The mean and standard deviation of the subset of data with unknown phase (mean = 4266, std = 2440) is similar to those of the entire dataset(mean = 4219, std = 2356).

This observation gives a strong indication that the **subset of data with PH = 0 is representative of the entire dataset and thus has a distribution of flight-phases similar to the entire dataset**.

The preflight phase is small and the fuel consumption is zero for most of the time (80% of the instances in this phase have FF = 0). The total fuel consumption is **highest** in the **climb (PH=4) and cruise (PH=5) phases**, which is obvious because these two phases are among the l**ongest phases of flights**.

The **approach (PH=6)** phase is also of a significant duration, but since the aircraft is **not accelerating anymore**, the fuel consumption is **smaller** in this phase.

The **takeoff (PH=3**) has a high rate of fuel consumption, but since this phase **does not last too long,** the total fuel consumption is rather small in this phase.

The difference between total fuel consumed in different flight instances is driven mainly by the **distance between source and destination**. Since the data does not contain the distance information, we should derive it from the **ground speed**. The GS_Mean column has ground speed as a function of time, and thus we can obtain distance by simply integrating this column. The following plot shows the total fuel consumption as a function of total distance for each of the flight instances given in the data.

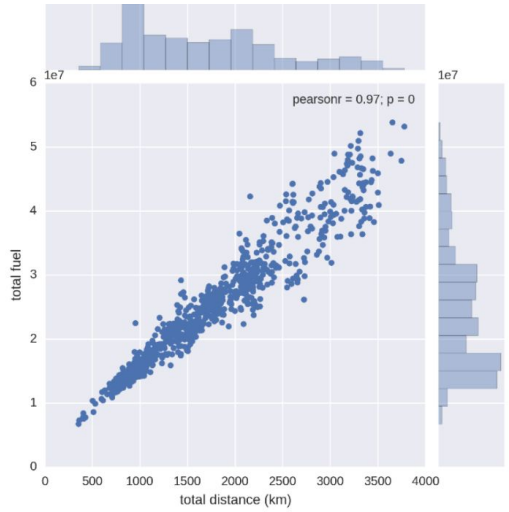
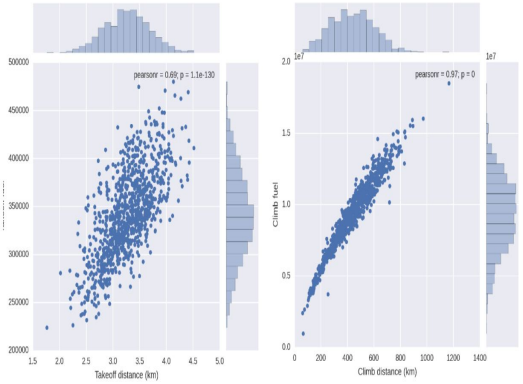


Fig 2 Joint distribution of fuel consumption and distance covered by the aircraft. A very strong linear correlation between distance and fuel consumed (pearson coefficient = 0.97) is found.

Since the taxi phase is mostly determined by the traffic and distance between aircraft bay and runway, it is not really in the direct control of the flight operators. The rollout phase is very small and hence not particularly interesting. The other phases, namely takeoff, climb, cruise and approach are the ones which correspond to travel between source and destination and are deemed to be more important.



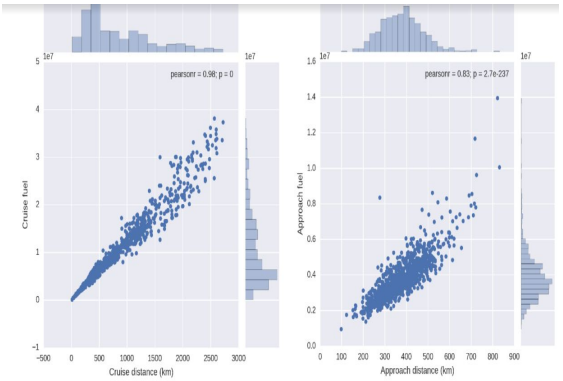
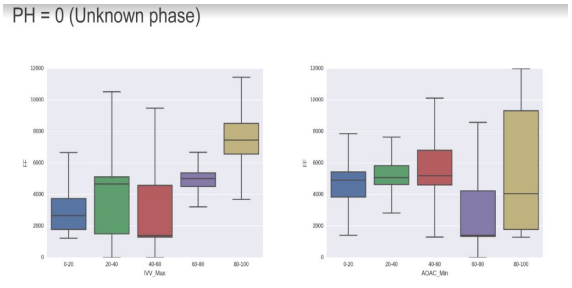


Fig 3 The joint distribution of fuel consumption and distance for takeoff, climb, cruise and approach phases. **The Climb and cruise phases have the largest fuel consumption**, but the relationship of fuel consumed with distance has a very strong correlation. On the other hand, the takeoff and approach phases have significant deviation from a linear relationship which hints at possible improvements.

A breakdown of the analysis of fuel vs distance into different phases gives an idea about **which phases need to be and can be optimized for low fuel consumption**. In fig 3, we show joint distribution of fuel and distance for different phases. This analysis has an implicit assumption that the flight **path is straight between source and destination**; which is a reasonable but not completely accurate assumption.

In the following subsections, the report will focus on **key differences in the predictors in different phases of flight**. As mentioned above, the **takeoff, climb, cruise and the approach** phases are the important phases as far as optimization of fuel consumption is concerned.



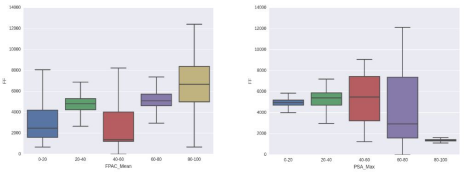


Fig 4 The relationship of fuel flow in the unknown phase with top individual predictors - *IVV_Max, AOAC_Min, FPAC_Mean* and *PSA_Max*. Here the x-axis is five equally sized groups of samples such that the first box corresponds to values between 0 and 20th percentile of predictor, the second box corresponds to values between 20th and 40th percentile and so on. Clearly, there is a significant relationship between fuel flow rate and these predictors but none of these four relationships are linear.

As mentioned before, PH=0 is actually a mini version of the entire dataset. The top individual predictors for this part of data were IVV_Max, AOAC_Min, FPAC_Mean and PSA_Max. The top combined feature for this phase was a combination of IVV_Mean, PT_Min and CAS_Max. The relationship of the important individual predictors with the fuel consumption is shown in the fig 4.

PH = 1 (Preflight)

The preflight phase is a small phase and **80% of the time in this phase does not consume any fuel.** The average fuel consumption for the total time in this phase is just **61** which is very small and insignificant compared to the average of all flight phases (4219). The top four individual predictors of fuel consumption for this phase are **VIB_1_Mean, OIT_4, OIPL and N1T_Max** and the top combined predictor was a combination of **SAT, PAC_Mean, TAS_Mean and IVV_Min.**

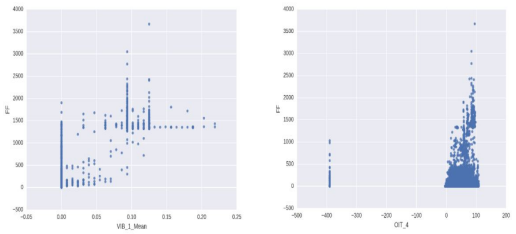


Fig 5 The relationship of fuel flow in the preflight phase with top individual predictors - **VIB_1_Mean, OIT_4, OIPL andN1T_Max**. Because of a skewed distribution of fuel flow (most of the samples having zero fuel flow), the distribution of fuel flow as a function of percentiles of predictors was **not informative**

PH = 2 (Taxi)

There is a significant amount of time spent in this phase (**20%** of the time in flights data given in this competition). The average fuel consumption (**1334**) is just above half of the overall average,and thus this phase accounts for about **6% of the total fuel consumed**. The top four individual predictors of fuel consumption for this phase are **LONG_Max, OIPL, VIB_1_Mean and FQTY**_4and the top combined predictor is a combination of **PT_Mean, OIPL and FPAC_Min.** The following graph shows the relationship of fuel flow rate with respect to these four individual predictors.

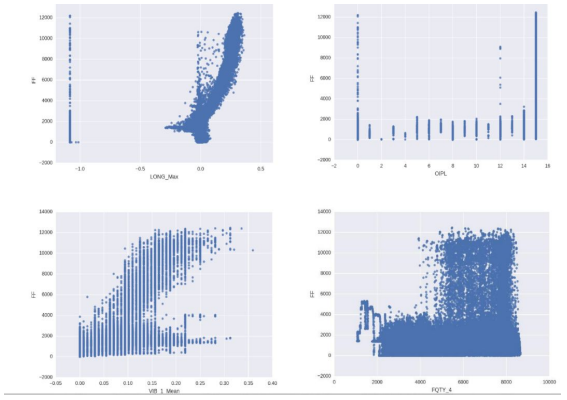


Fig 6 The relationship of fuel flow in the taxi phase with top individual predictors - **LONG_Max, OIPL, VIB_1_Mean and FQTY_4.Only** LONG_Max shows a clear relationship, for which the fuel flow rate increases with increase in the value of LONG_Max. The fuel flow was VIB_1_Mean also has a faint visible pattern, but the other two graphs (OIPL and FQTY_4) are very uninformative. The validation root mean squared error in this phase was about 120, so the absence of a clear visual relationship between fuel flow and top predictors is indicative of strong non-linear interaction between predictors.

PH = 3 (Takeoff)

The average fuel flow rate in this stage is very large **(6077)** but since this phase lasts for a **small time**, it accounts for only about **2.5%** of the total fuel consumption. The top individual predictors are **LONG_Max, FLAP, IVV_Max and AOAC_Min**. The best combination of features for prediction of fuel consumption in this phase was **GS_Min, N1T_Max, IVV_Min and FPAC_Min.**

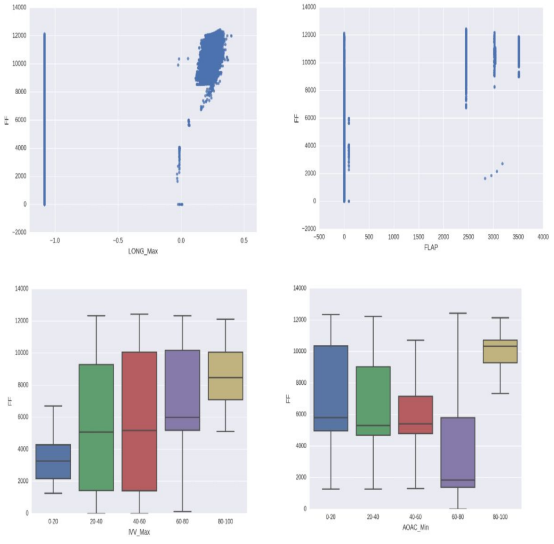


Fig 7 The relationship of fuel flow in the takeoff phase with top individual predictors - **LONG_Max, FLAP, IVV_Max and AOAC_Min**.

The first two graphs are shown as simple scatterplots, while the latter two are shown as distribution of fuel flow rate in different percentile ranges of predictors. The graph format is chosen for best possible clarity.
The relationship of fuel flow rate with LONG_Max and IVV_Max (somewhat linearly increasing) and AOAC_Min (nonlinear relationship) are visible from the plots in fig 7.

PH = 4 (Climb)

This is the phase with the **highest average fuel flow rate (7254)** and it accounts for about **33%** of the total fuel consumption during a flight. The most important predictors for this phase are **PT_Mean, LONG_Max, FLAP and ALTR_Mean**. The fuel flow rate increases with increase in PT_Mean, LONG_Max, and ALTR_Mean as is clear from boxplots in fig 8 below. However, the relationship of fuel flow rate with FLAP is not clear from a plot, which implies a nonlinear relation or a strong interaction of FLAP with other features.

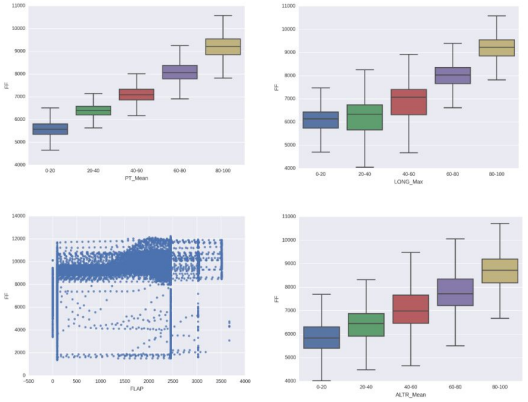


Fig 8 The relationship of fuel flow in the climb phase with top individual predictors - **PT_Mean, LONG_Max, FLAP and ALTR_Mean**. The third graph is shown as simple scatterplots, while the other three are shown as distribution of fuel flow rate in different percentile ranges of predictors. The graph format is chosen for best possible clarity.**The best combination of features as a predictor for climb phase was LONG_Max, OIT_4 and PT_Max.**

PH = 5 (Cruise)

This is the **longest phase of flights**, and constitutes on an average **32%** of total flight time, and so the total fuel consumption in this phase (~ **37.8%**) is more than any other phase. The average fuel consumption **(4966) i**s smaller than the **takeoff and climb phase** because the average altitude is very high, and the air is thin. Thus the work done against the drag of the atmosphere is small in the cruise phase, even if the ground speed is very high. The important individual predictors of fuel consumption in this phase are **CASS, N1T_Max, LONG_Max and FPAC_Max**, whereas the combined features which could best predict the fuel flow is a group of **PT_Min, CAS_Min, PI_Min and LONG_Min.**

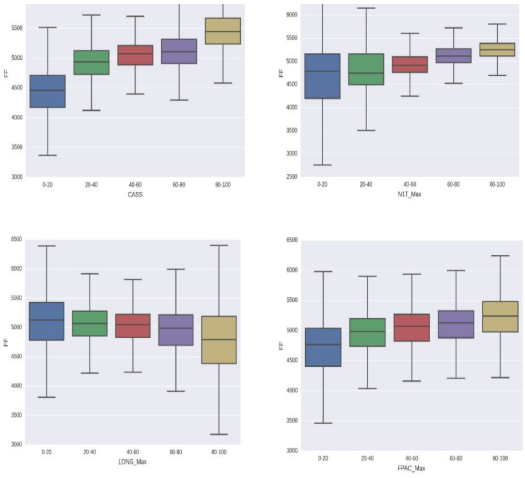


Fig 9 The distribution of fuel flow rate for cruise phase in different percentile ranges of top individual predictors - **CASS, N1T_Max, LONG_Max and FPAC_Max**. The fuel flow rate slightly increases with increase in **CASS, N1T_Max and FPAC_Max**, whereas it decreases slightly with increase in LONG_Max.

PH = 6 (Approach)

At approach, the average fuel consumption **(2940)** is much smaller than **climb, takeoff** or **cruise.**But this phase still accounts for about **13.6%** of total fuel consumed, as the average time duration of this phase is significant (**19.5% of total flight time**). The best individual predictors for this phase are **N1T_Max, MACH_Min, LONG_Max and IVV_Min**. There was no combination of four or less features which could explain even **40%** of the variance in fuel consumption for this phase. Fig 10 below shows a rather weak relationship between fuel flow rate and the top individual predictors.

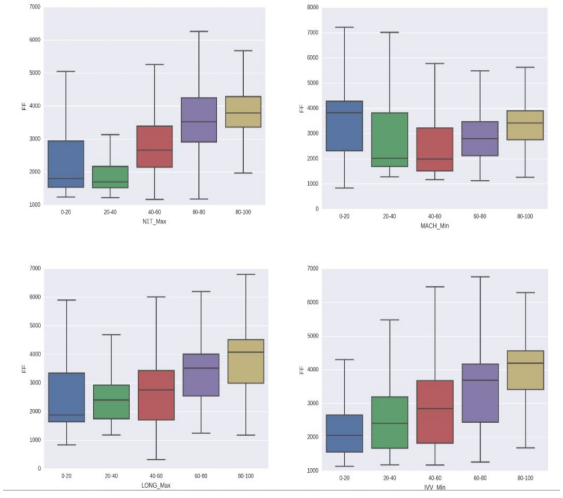


Fig 10 The distribution of fuel flow rate for approach phase in different percentile ranges of top

individual predictors -**N1T_Max, MACH_Min, LONG_Max and IVV_Min**.
The fuel flow rate slightly increases with increase in N1T_Max,LONG_Max and IVV_Min, whereas it first decreases and then slightly increases with increase in MACH_Min. The visual dependence in each of these graphs is very poor.

PH = 7 (Rollout)

This phase consists of the **smallest amount of time and smallest total fuel consumption**, which is just **0.12%** of the total fuel consumed. Thus this phase does **not** have any **meaningful influence** on the amount of fuel used. Nevertheless, we did modelling to predict fuel flow for this phase and found that the top predictors were **LONG_Max, FADS, FPAC_Mean and CAS_Mean**. The relationship of fuel flow rate with LONG_Max, and FPAC_Mean is shown as scatterplot in fig 11, and those with FADS and CAS_Mean are shown as boxplots with various values and various percentile ranges of predictors respectively.

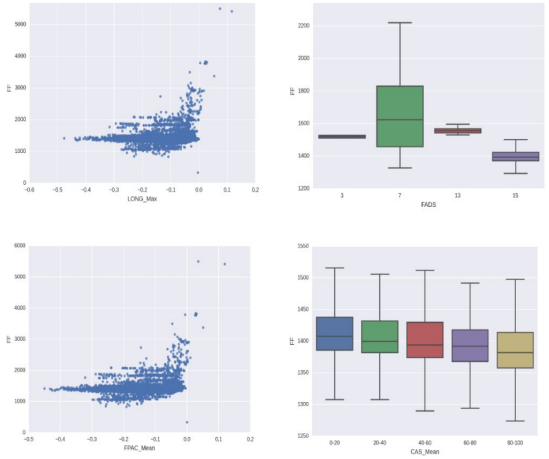


Fig 11 The distribution of fuel flow rate for rollout phase in different percentile ranges of top individual predictors -**LONG_Max, FADS, FPAC_Mean and CAS_Mean**. The dependence of fuel flow on these predictors is not visually clear from the graphs. The format of graph (scatter plot vs boxplot) is chosen for best possible clarity.

CONCLUSION

The most important phases for **optimizing fuel consumption** are **climb, cruise and approach**. The overall important features are **rate of change of altitude, longitudinal acceleration** and **ground speed**. The clearest visible trend between predictors and fuel flow rate is in the **climb phase**. In other phases, some of the predictors have a weakly visible trend, but since the root mean squared error is small, it is assumed that the features have strong nonlinear interactions which are not clearly visible in simple plots.

REFERENCES

- Yashovardhan S. Chati*, Hamsa Balakrishnan*, "**STATISTICAL MODELING OF AIRCRAFT ENGINE FUEL FLOW RATE**" , International Council of the Aeronautical Sciences, 2016 , 10 pages (Last Accessed 18th May 2019)
- **Predict fuel consumption of airplanes** , Magic Data , (Last Accessed - 18th may 2019)
- **Predict fuel consumption of airplanes during different phases of flight ,** Crowdanalytix, (Last Accessed - 18th May 2019)
- **Data Science With Python course** , Datacamp , (Last Accessed - 14th may 2019)
- **Machine Learning Nanodegree Program** , Udacity (Last Accessed - 28th April 2019)
- **Data Science for all** channel , Youtube (LAst Accessed 16th May 2019)