Airfoil Self-Noise Prediction

Nikhil Ajgaonkar

Overview

The noise generated by an aircraft is an environmental issue for the aerospace industry. It is one of the major sources of noise pollution. Reducing the aerodynamic noise in an aircraft remains by far the most important topic in the aerospace industry. Aircraft noise mainly can be classified into mechanical noise and aerodynamic noise. Mechanical noise arises due to moving parts of the engine while aerodynamic noise occurs due to interactions of the airflow with the surfaces of the aircrafts at high speeds. In this project we will be discussing the noise caused due to aerodynamics of an aircraft. A main component of this noise is the airfoil self-noise. When an airfoil interacts with a smooth non-turbulent inflow, a turbulence is produced at its boundaries due to the interaction between the airfoil blades and incoming inflow or wakes. This turbulence is called self-noise as it depends on the geometry of the airfoil itself. As airfoil self-noise makes most of the aerodynamic noise, it is therefore important to model the airfoil self-noise to understand the factors it depends on, thereby providing measures to reduce it.

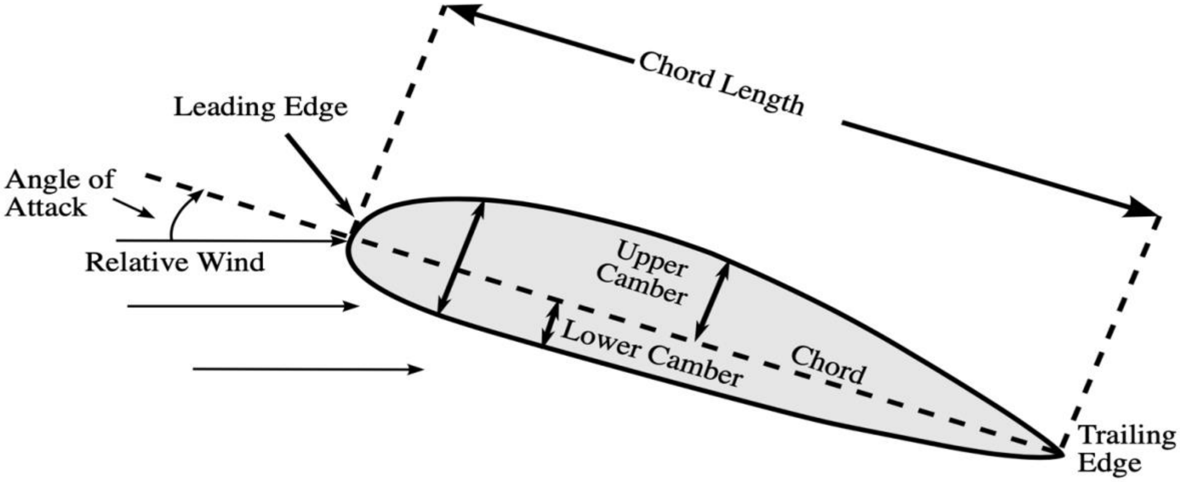


Figure 1: Cross section of an airfoil showing chord length, angle of attack and both the airfoil edges. Image downloaded from en-academic.com

Data Set

We will be using the NASA data set downloaded from the UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise#). The data set consists of various noise measurements from different sizes of NACA 0012 airfoils when placed at various wind tunnel speeds and angles of attack. The data set consists of 6 variables with 1503 measurements. Out of the 6 variables, 5 are the independent variables (geometry and external factors of the airfoil) and the 6th variable is the dependent variable which is the airfoil self-noise measured in decibels. The parameters are mentioned below:

Independent Variables:

1. F: Frequency (Hertz)
2. theta: Angle of attack (Degrees)
3. L: Chord length (meters)
4. v: Free-stream velocity (meters/sec)
5. T: Suction side displacement thickness(meters)

Dependent Variables:

1. P: Scaled sound pressure level (decibels)

Motivation

Figure 1 shows the cross section of an airfoil with the non-turbulent inflow hitting the airfoil from the left. This non-turbulent inflow from the left can interact with the airfoil geometry and cause turbulence depending upon the angle of attack, chord length and other geometric factors of the airfoil. The noise generated is therefore dependent on these factors. Our main aim is to predict the self-noise generated by the airfoil in the form of a scaled sound pressure level. We plan to use regression analysis (linear /polynomial) to model the sound pressure level of the airfoil as a function of the independent variables shown above. In our analysis we find that some of the independent variables are correlated with respect to each other and can introduce multicollinearity. Multicollinearity does not affect the prediction, precision of prediction and goodness of fit. It only affects the coefficients and biases them depending upon the severity of correlation between the independent variable. Our main goal here is to predict the noise pressure as a function of independent variables so that we know which of these variables contribute more towards the noise pressure. Therefore, we must accurately know the regression coefficients of the independent variables and must apply LASSO/RIDGE regularization techniques to deal with multicollinearity. The NASA data set comes with the exact values of the independent and dependent variables with no uncertainties on them. We plan to simulate the uncertainties on the dependent and independent variables and try to see the effect on regression models for various levels of uncertainties on the variables.

Exploratory Data Analysis (EDA)

The first step of the EDA is to see if the data consists of any missing or null values. We do not have any missing values in our data set but if there were any missing values, we would exclude them form our data set rather than interpolating them. Interpolation techniques adds unnecessary variance, and it is best to just impute the missing values if they are present in small quantities. The next step of the EDA is to see if we have any outliers in the variables. The existence of outliers can affect the accuracy of the regression models. It is therefore important to detect and remove the outliers in the data set to improve the accuracy and performance of the regression models. Figure 2 shows box plots of all the variables in the NASA data set. It can be seen from figure 2 that the variables like F, theta, T and P contain outliers. We detect and remove an outlier in a variable if it falls below Q1 – (1.5IQR) and is above Q3 + (1.5*IQR). Here Q1 and Q3 are the first and the third quartiles and the IQR is the interquartile range of the variables. They are shown in the boxplots for each variable in figure 2.

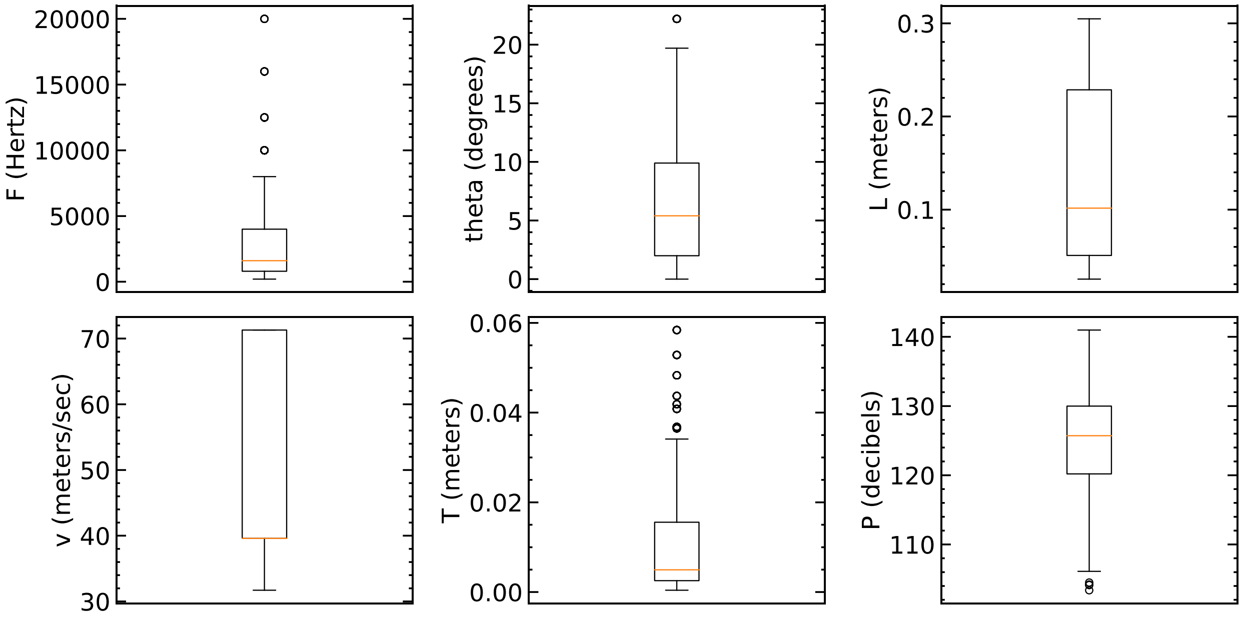
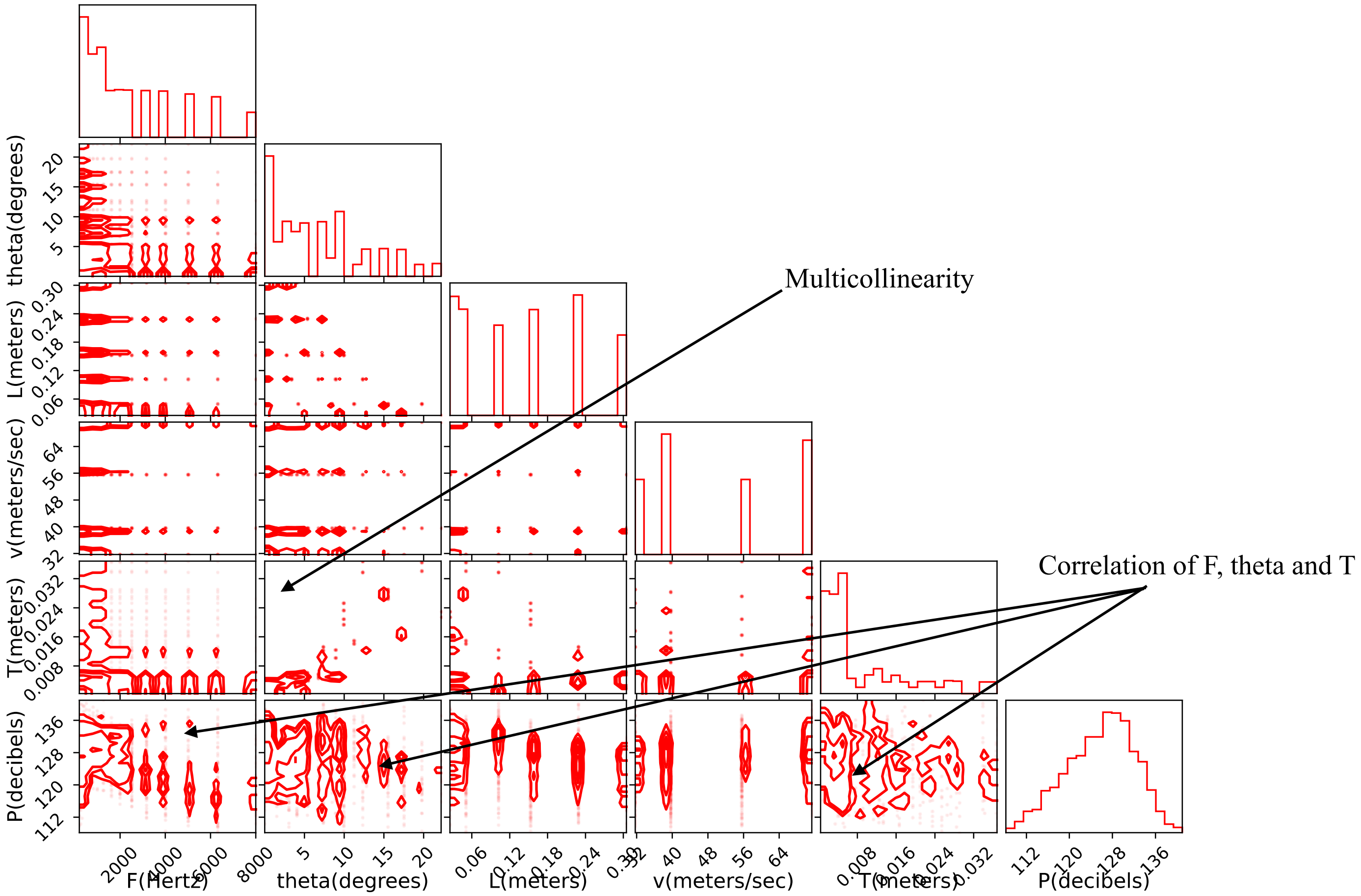


Figure 2: Box plot of all variables from the NASA dataset.

The last step of the EDA is to check the correlations between the independent and dependent variables after removing the outliers. Figure 3 shows a corner scatter plot between the independent variables and the dependent variable. The bottom most row of figure 3 shows that the noise pressure level P has weak correlations with frequency F, angle of attack theta and suction thickness T. There also seems to exist a correlation between independent variables like angle of attack and suction thickness (second to last row - second column). This shows that we do have a multicollinearity in our independent variables.



Multicollinearity

Correlation of F, theta and T

Figure 3: Corner plot showing correlation between dependent and independent variables.

Adding Gaussian noise to the data

In most of the data science projects related to science and technology, the uncertainty on variables is either missing or ignored. The data is never exact always has an uncertainty associated with it. The uncertainty arises due to the least count of the instruments used to measure the variables. The least count of an instrument determines the accuracy of the variable measured through it. As an example, let us take the frequency column F from our data set. If the least count of the instrument measuring the frequency is 100 Hz, then a reading of frequency F = 2000 Hz in the table can have values anywhere between 1900 Hz and 2100 Hz. If the least count is smaller say 20 Hz then the same value can lie between 1980 Hz and 2020 Hz, thus making the value more accurate. The addition of error in the variables can shift the data points to different values especially when data is high dimensional. This can hugely affect the regression models and accuracy. Most importantly, by introducing uncertainty in the data, we can make predictions with a certain level of accuracy. Due to this we can at least be certain that our predictions are correct within the margin of error.

To simulate the error, we consider the least counts on F, theta, T and P to be 50 Hz, 0.5 degrees, 0.0001 meters and 1 decibel as a start. The error is added in the form of gaussian uncertainties for each measurement. For example, to calculate noisy frequency F_n,

$$F_n = F + \eta[0,1]L_f$$

Where $\eta[0,1]$ is a random number generated from a gaussian distribution centered at zero with a spread of 1 sigma and L_f is the least count of the frequency measurement. The noisy frequency is calculated for each of the 1503 measurements. Similarly, we add noise to theta, T and P. The chord length L and Free stream velocity take discrete values and therefore we do not add any uncertainty to these variables. Figure 4 shows the correlation for P against F, theta and T for noisy (red points) and original (black) data after removing the outliers. The noisy data is slightly shifted as compared to the original data. This shift is more prominent when visualized in 3 dimensions.

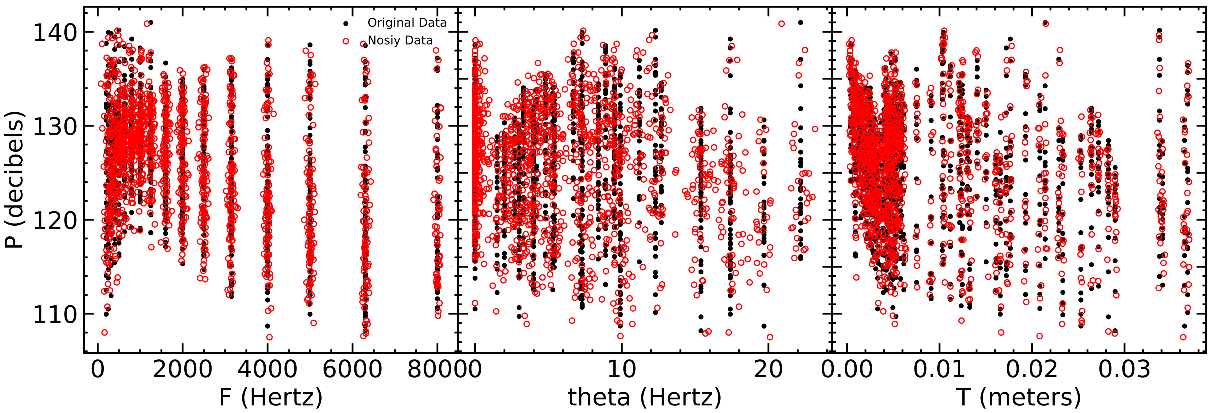


Figure 4: Comparison of noise pressure P against frequency F, angle of attack theta and suction side thickness T for noisy and original data.

Future Work

1. Perform regression analysis on the data set to predict the dependent variable. We must make sure to use not too simple or not too complicated models to take care of the bias-variance tradeoff.
2. Add regularization models to the regression models to pin down certain variable coefficients accurately and remove unnecessary variables from the data set.
3. Add different levels of noise (least counts) and perform regression analysis to see how it affects the predictions. We think that the predictions could improve for good quality data (having small least counts).