The Data Incubator

The Data Incubator Capstone Project

Nikhil Ajgaonkar
Department of Physics & Astronomy
University of Kentucky

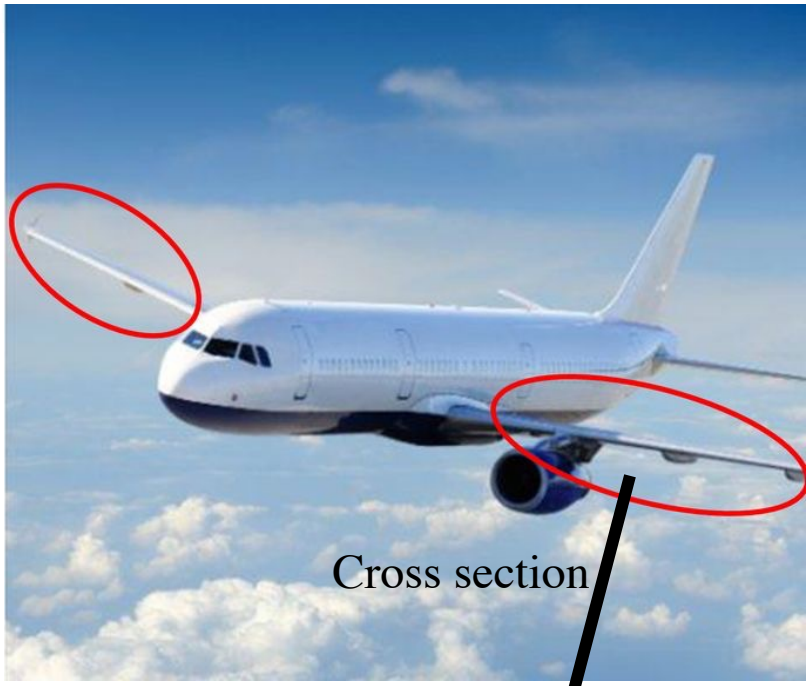
Introduction

1. Airfoil self noise is caused due to interaction of smooth non-turbulent flow with airfoil edges near wake.

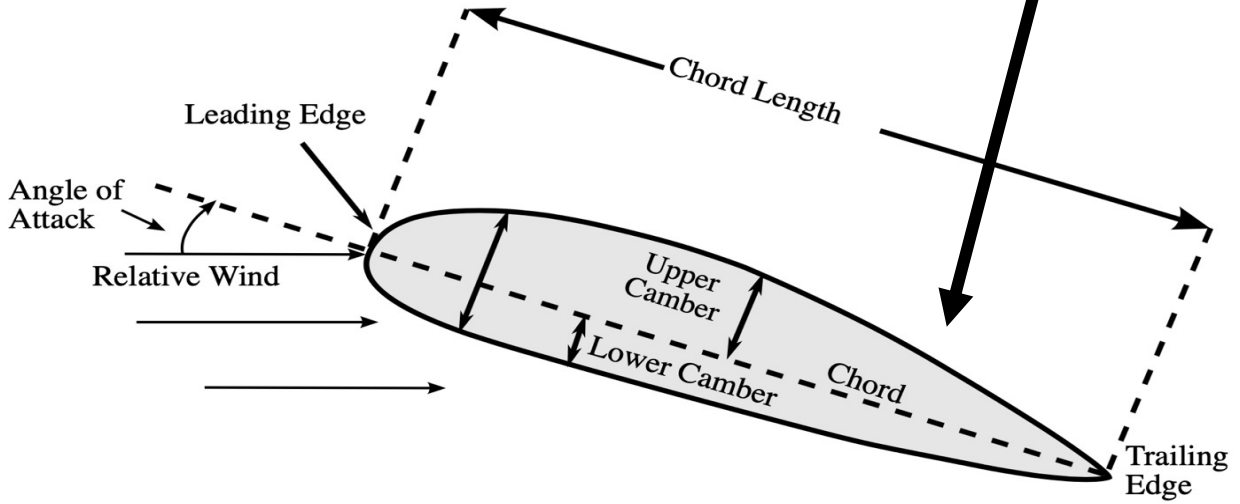
2. At high speeds, noise can increase vibrations on the airplane wing and can cause the wing to break.

3. Reducing airfoil self-noise is an important problem in the aerospace industry.

How to predict airfoil self noise?



Cross section



The Data Set

Noise measurements from different sizes of NACA 0012 airfoils placed at various wind tunnel speeds(v) and angles of attack(θ).

(http://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise#)

Number of measurements = 1503, Number of variables = 6 (Continuous)

Independent Variables (X)	**Dependent Variable (Y)**
Frequency (f) [Hz]	**Scaled sound pressure level (P) [dB]** **Aim: Predict P**
Angle of Attack (θ) [deg]	
Chord Length (L) [m]	
Free stream velocity (v) [m/s]	
Suction side displacement Thickness (t) [m]	

Noise measurements from different sizes of NACA 0012 airfoils placed at various wind tunnel speeds(v) and angles of attack(θ).

(http://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise#)

Number of measurements = 1503, Number of variables = 6 (Continuous)

Exploratory Data Analysis

1. **Check for null values**

2. **Outlier Removal** => For a variable q:

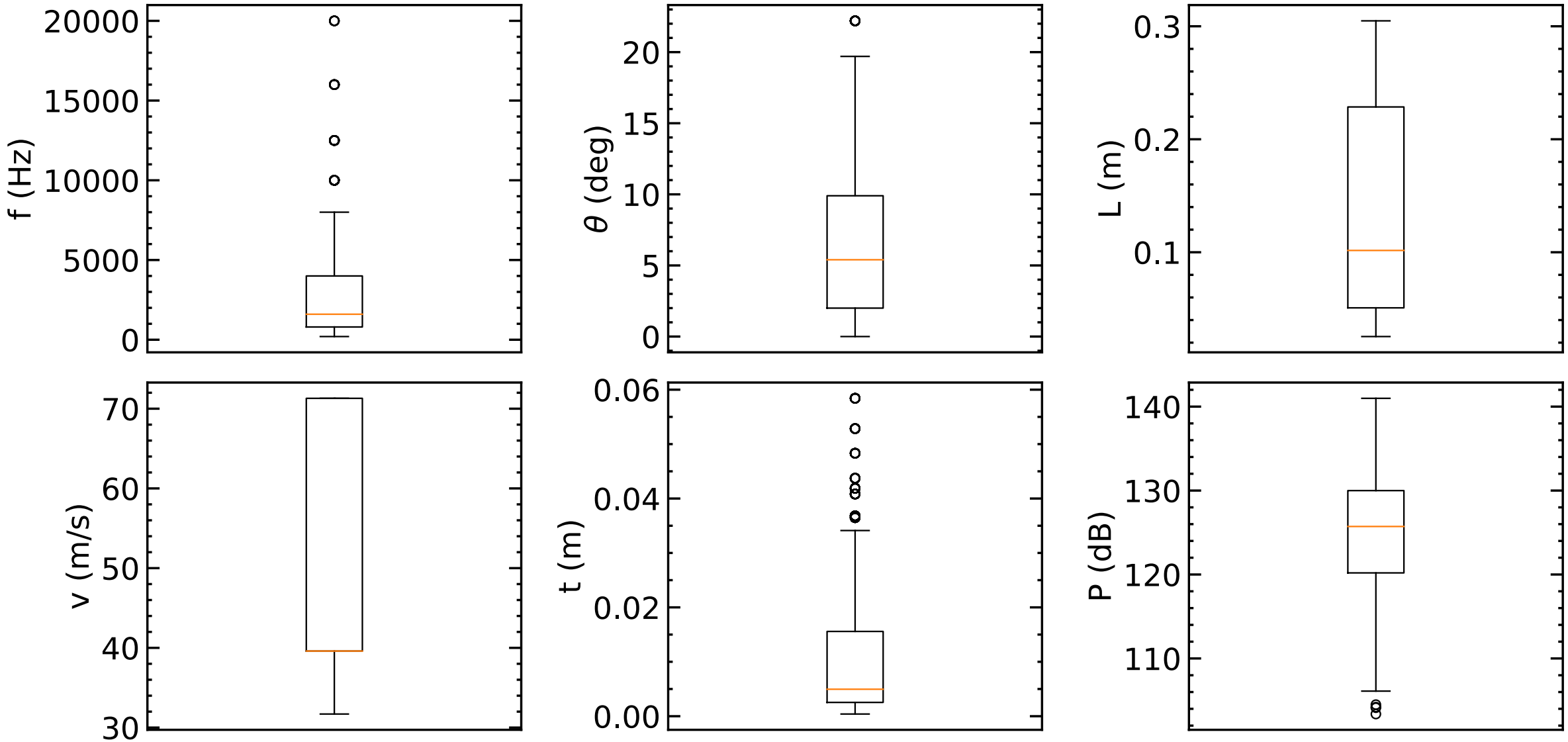
$$Outlier(q)_{Lower}, Outlier(q)_{Upper} = q^{25th} - 1.5IQR,\ q^{75th} + 1.5IQR$$

$q^{25th} = 25^{\text{th}}$ percentile of q distribution

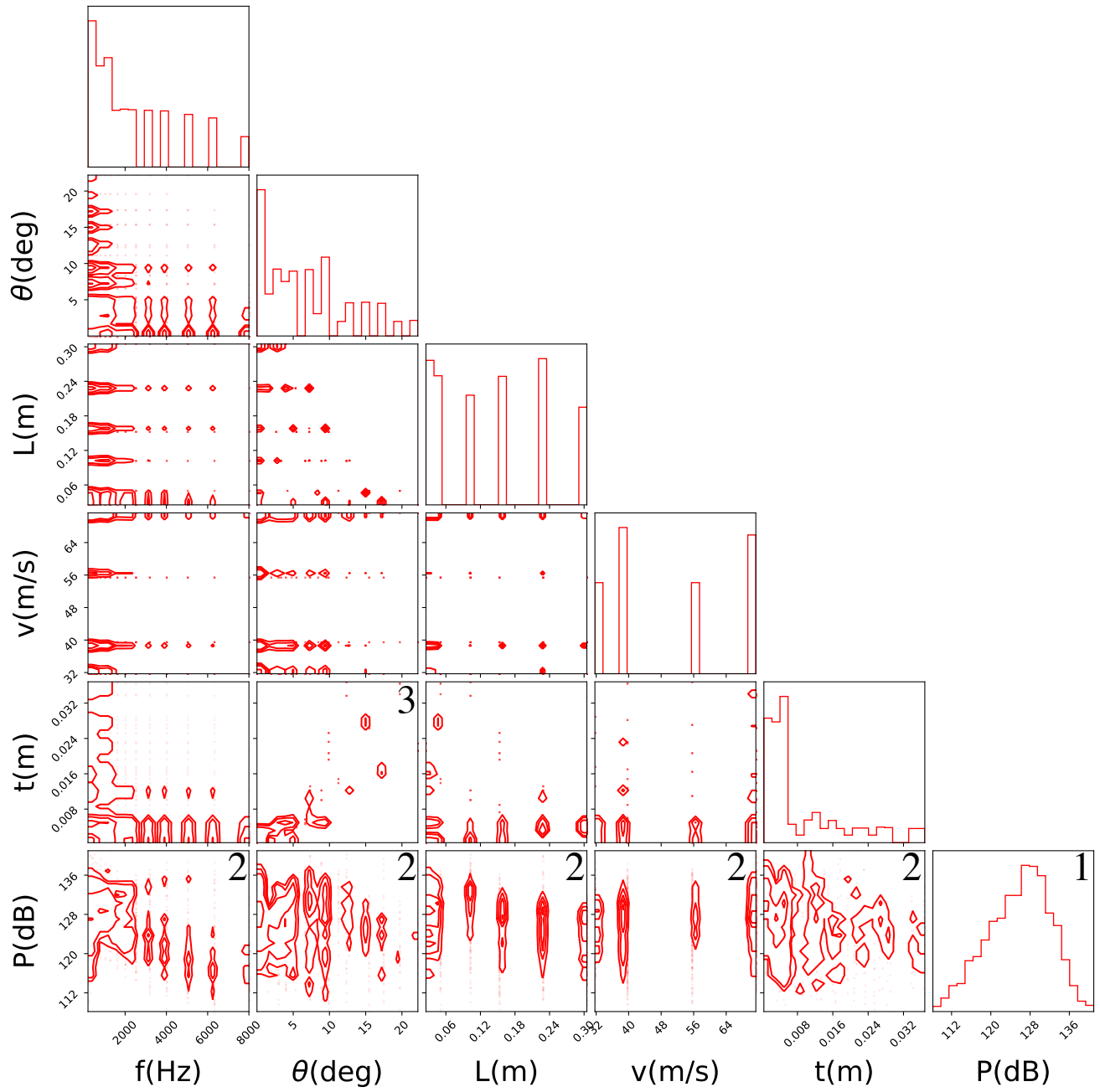
$q^{25th} = 75^{\text{th}}$ percentile of q distribution

$IQR =$ Interquartile range

Exploratory Data Analysis



Exploratory Data Analysis

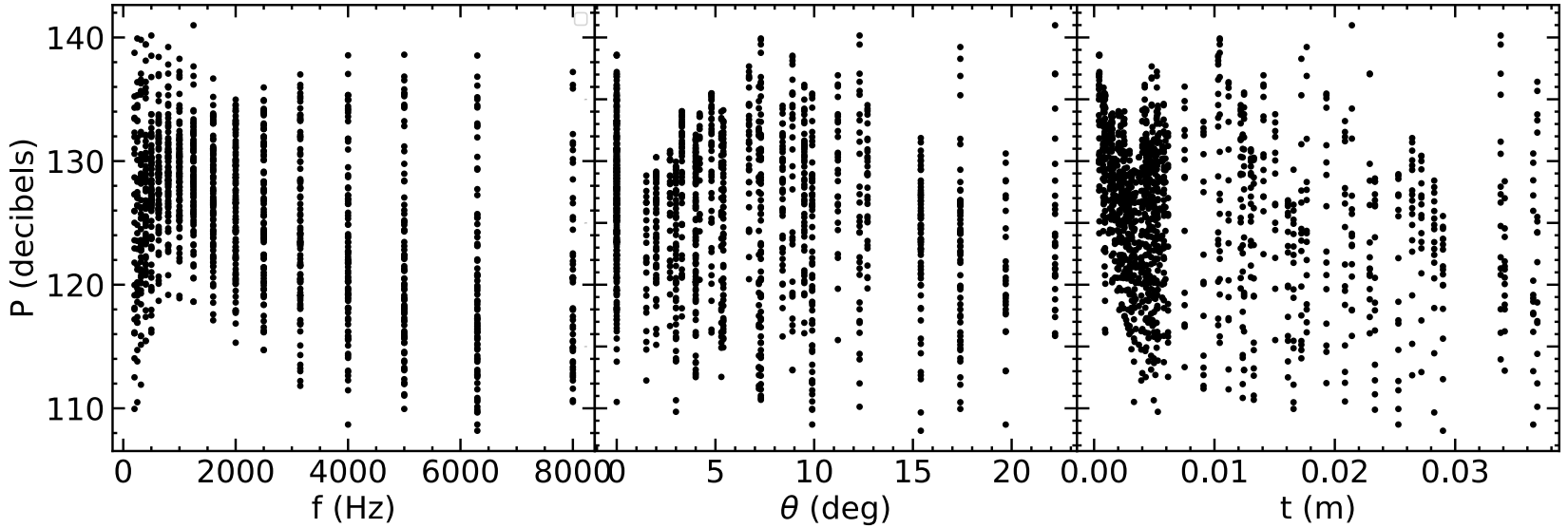


1. Distribution of P is approximately Gaussian.

2. P varies **non-linearly** w.r.t f, θ, L, v & t => **non-linear regression required**

3. Signs of multi-collinearity => **regularization**

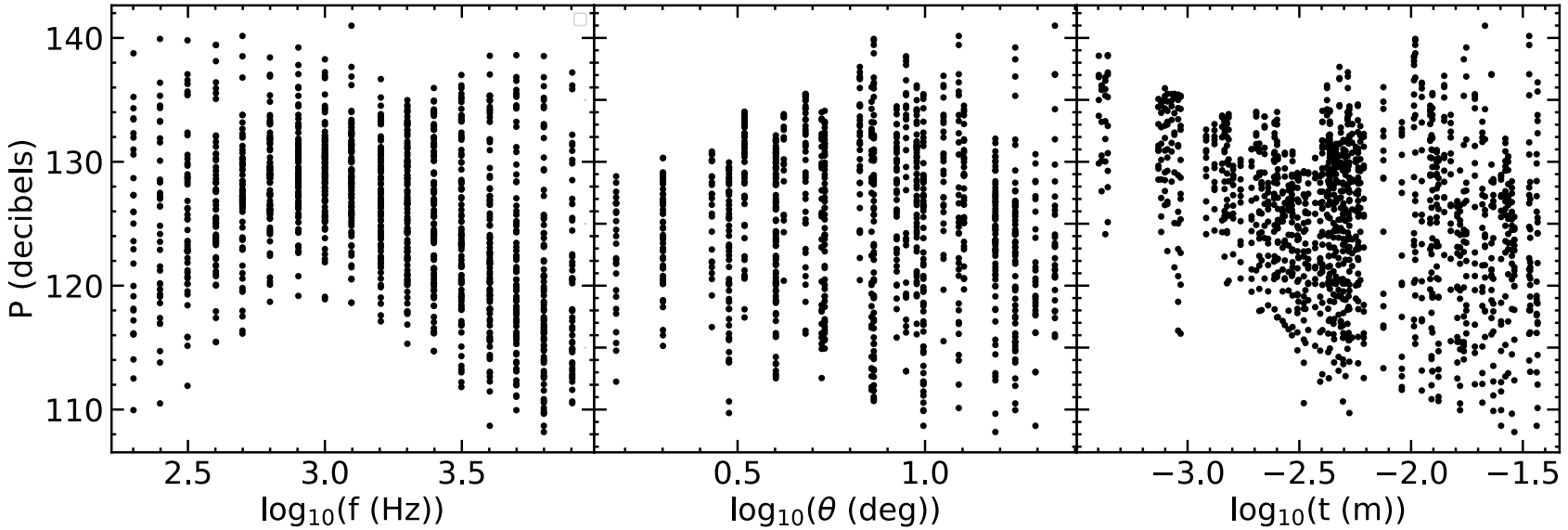
Change of variables (f,θ,t)



$f \Rightarrow \log_{10}(f)$

$\theta \Rightarrow \log_{10}(\theta)$

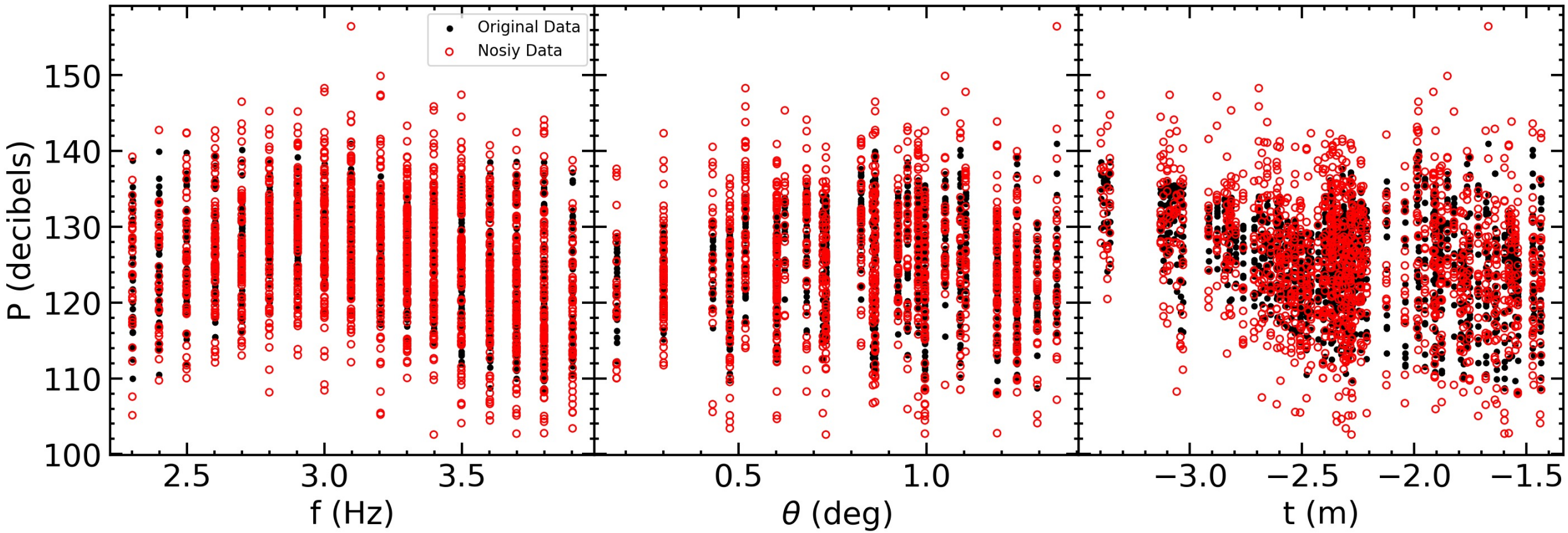
$t \Rightarrow \log_{10}(t)$



Adding Gaussian Noise to dependent variable, P

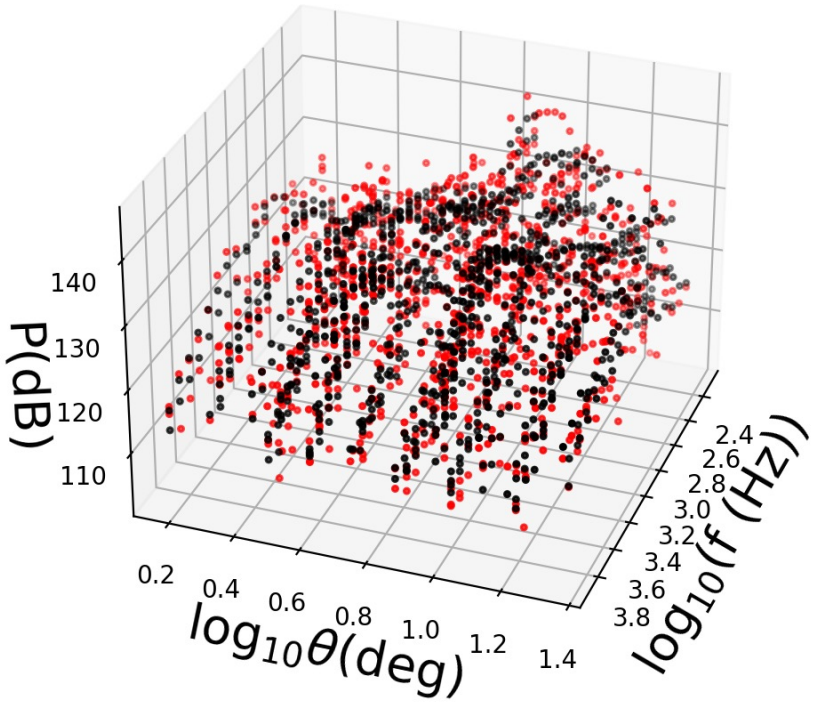
$P_{noisy} = P + N(0, \sigma)$; $\sigma = Noise\ in\ pressure\ in$ dB

$\sigma = 3$dB

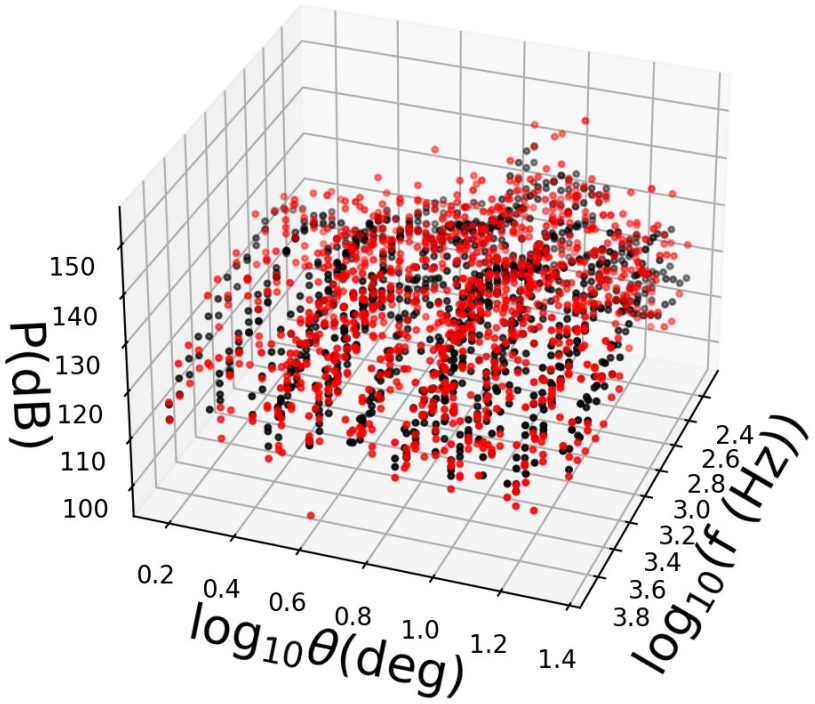


3D view of added noise

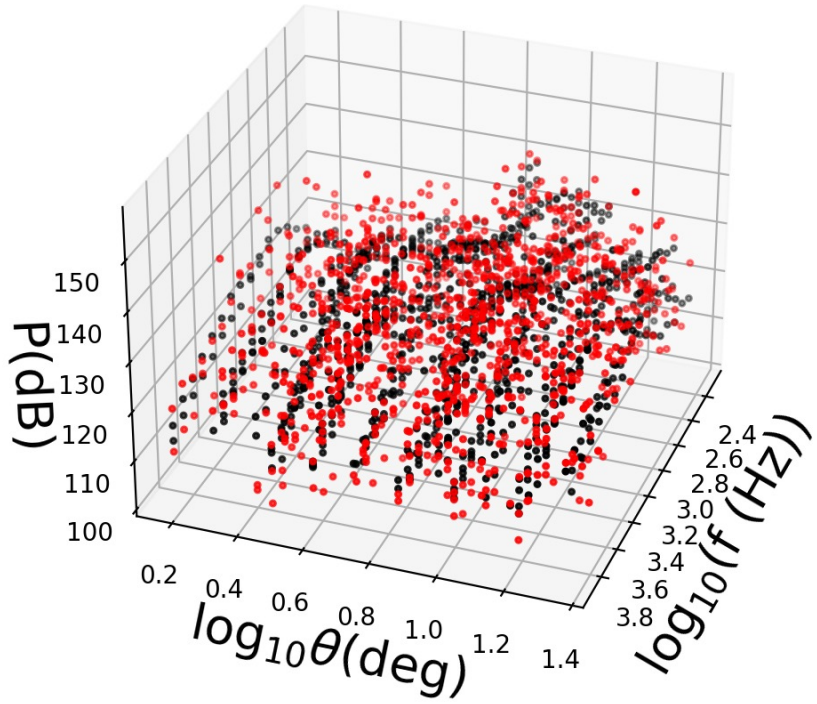
$\sigma = 3$dB



$\sigma = 5$dB



$\sigma = 8$dB



Future work

$\Rightarrow$ Predict P using different techniques like non-linear regression (OLS, SVR, Neural Networks etc) and compare the accuracies.

$\Rightarrow$ Use regularization techniques (Lasso/Ridge) to constrain the parameters of correlated independent variables

$\Rightarrow$ Perform the above two for different levels of added noise to P. Relate the accuracy of the selected model.

Thank you….

Questions?