Response Surface DOE
Airfoil Noise

Run Using Minitab for the DOE and Python code using a Random Forest model for the Simulation

Kevin Maher
Vettejeep365@gmail.com

October 1, 2017

Introduction

- Running a DOE on a Machine Learning model of a regression data set
  - A Method of Optimizing the feature of interest in the regression data set when that data set is complex – for example it does not fit first order linear regression well
- The Machine Learning Model, here a Random Forest, learns the characteristic function of the underlying data
  - The Random Forest is a “Black Box” model – hard to interpret
- The DOE Allows for:
  - Optimization, maximizing or minimizing an output of interest
  - Identification of important variables, squared terms and interactions
  - Exploration of a data set using DOE methodology when the actual physical system is not available or in a real-world problem this might allow for zeroing in on the optimal operating region without running excessive expensive physical experiments

Overview

- Goal is to minimize the noise of an airfoil with a DOE
- Data
  - Data: UCI Airfoil Self-Noise Data Set
  - Regression data, not DOE data
  - Does not fit first order multi-variate linear regression well
- Model: a Machine Learning model will be fit to the data – the DOE can then be run on the fitted model
  - A set of Random Forest models will be used to fit the data set – these models tend to capture non-linearity well
  - Multiple models (4) will be utilized since averaging models has been shown to produce better results and it gives the DOE replicates
- A Response Surface DOE will be demonstrated here
  - This is a good choice because of curvature, noting that first order linear regression did not fit well

Data

- 5 Input Variables
  - Frequency (Hz) – probably the 1st mechanical resonance of the airfoil but this information is not given in the data set
  - Angle of Attack of the Airfoil (degrees)
  - Chord Length of the Airfoil (meters)
  - Free stream air velocity (meters per second) - wind tunnel air velocity
  - Airfoil Thickness (Meters)
- A Plackett-Burman screening DOE confirmed curvature in the response so all 5 input variables were kept in the model
  - The Box-Benhken response surface DOE also had a different view of feature significance from the Plackett-Burman experiment, not surprising considering the effects of curvature

DOE Design

- Box-Behnken Response Surface Design Chosen for Model
  - Because of the ease with which alternate models may be tried in simulation a Central Composite Design was also tried – in this case the quality of the residuals indicated that the Box-Behnken design might be a better model so that it what is presented here
  - DOE factors generally centered about the minimum noise point in the input data, but since this minimum in the input could be an outlier affected by noise or uncontrolled factors not provided in the given data set, the DOE will be used to predict the best operating point to provide the least airfoil noise

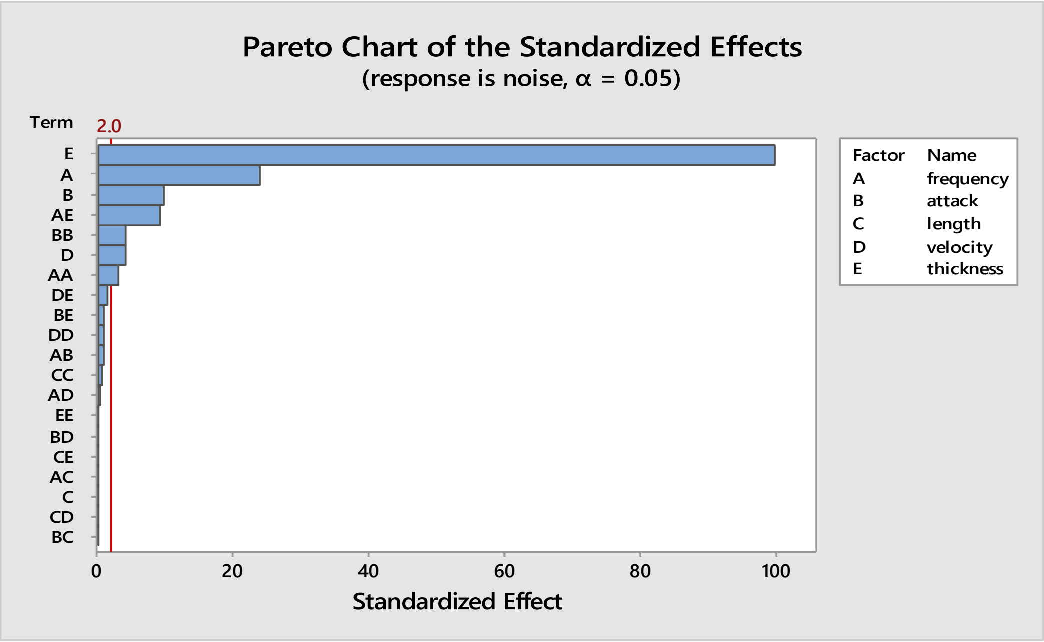
Create Response Surface Design: Factors

Factor	Name	Low	High
A	frequency	4000	7000
B	angle	9	15
C	length	0.2	0.26
D	velocity	34	44
E	thickness	0.03	0.06

Help OK Cancel

Box-Benhken DOE Variable Importance

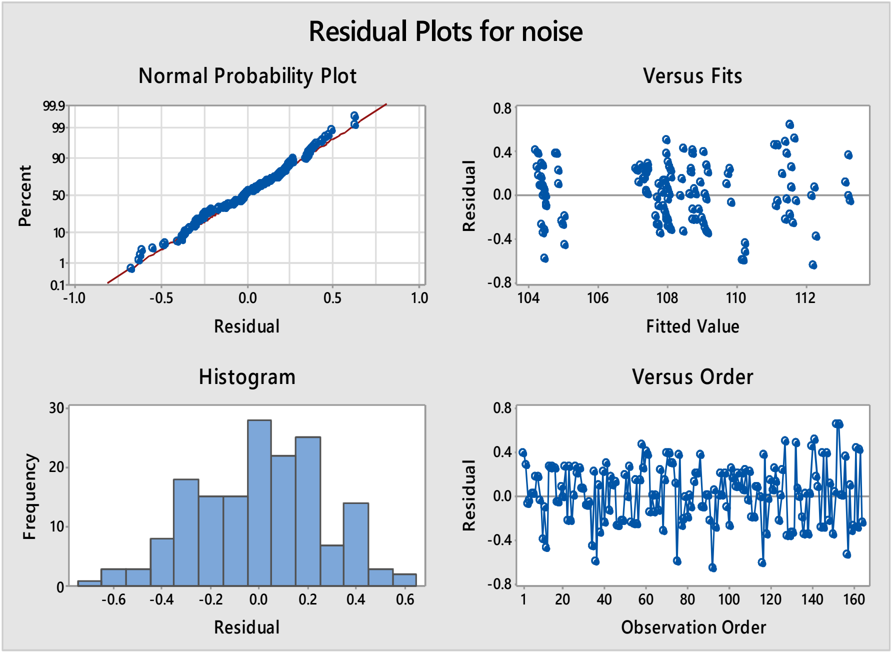
- Only Length among individual features is not considered statistically important
  - The four significant features and the interactions that are above the significance threshold will be kept in the model – other terms will be discarded
  - Length and Velocity had been rejected as unimportant by the screening DOE, but the screening DOE is misleading in this case due to model curvature



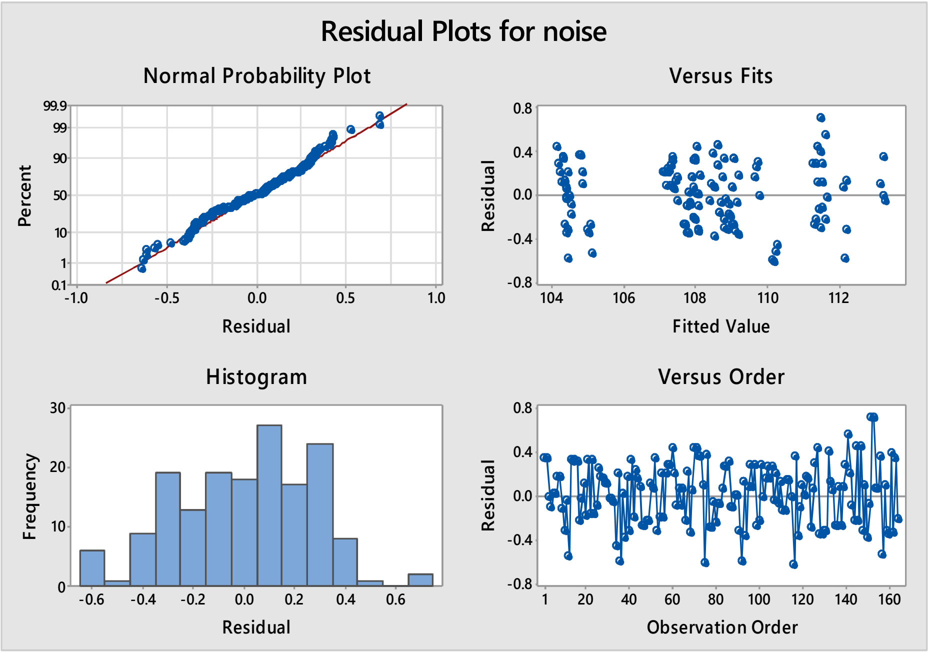
Residual Plots from the DOE

- Residuals are generally normally distributed and show no significant problems
  - Gaps for fitted value in the “Versus Fits” plot seem to be an artifact of the discreet levels of the DOE factors

Residuals with All Variables

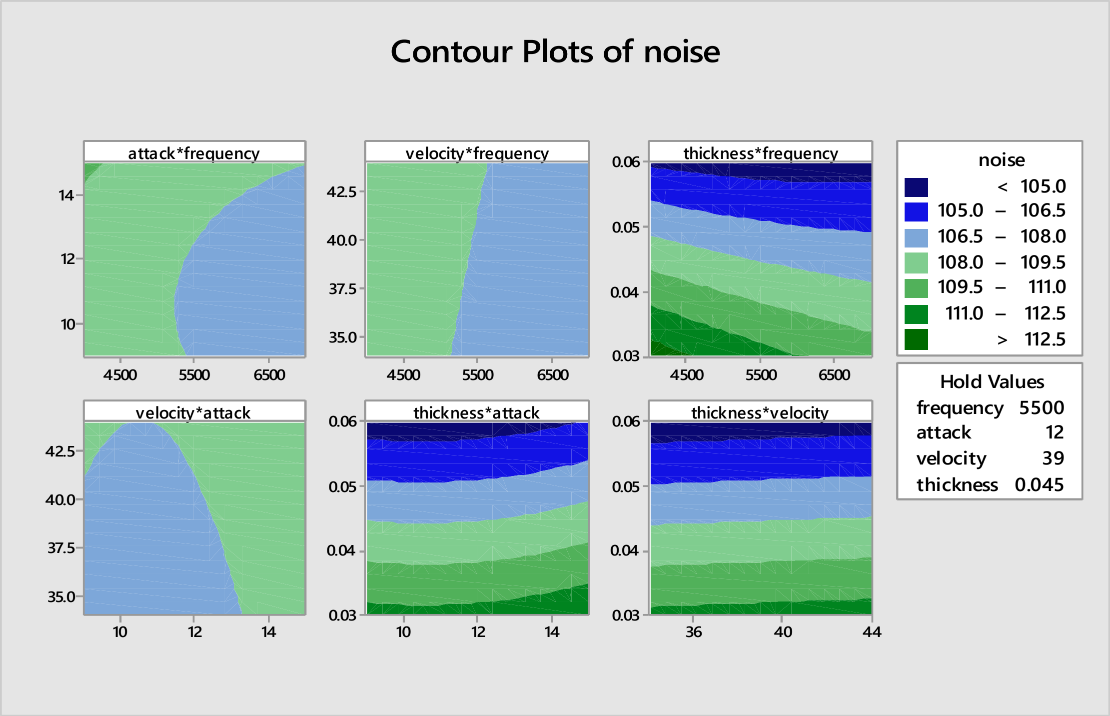


Residuals with Unimportant Variables Removed



Contour Plots from the DOE

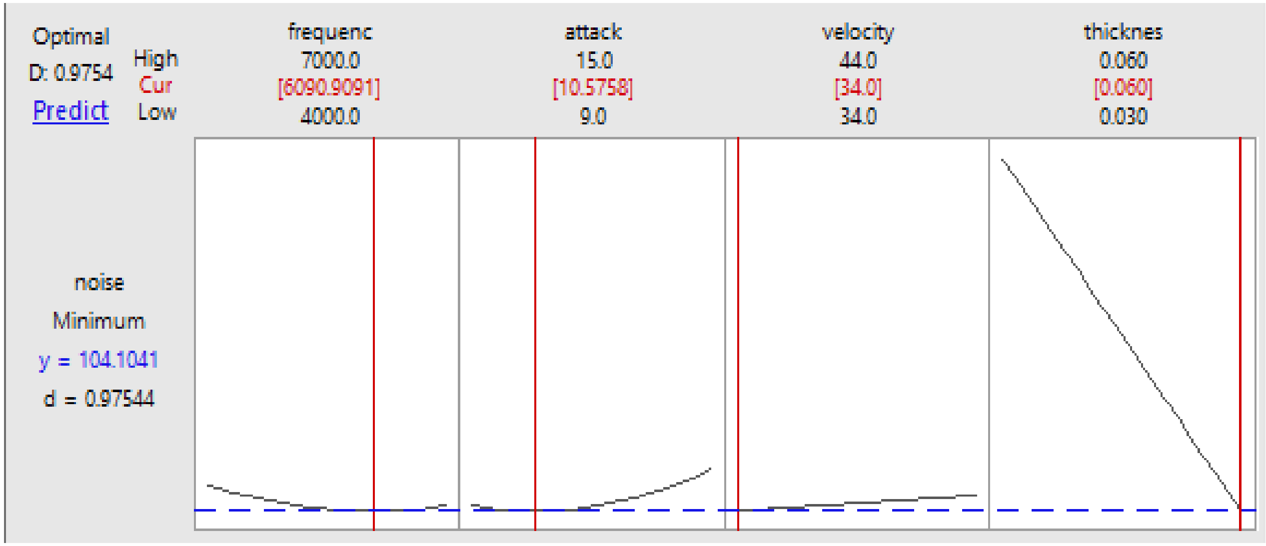
- Contour plots do not show clear local minimums
  - This DOE was constrained to a feature level region where the Random Forest model could be considered valid
  - Further study of the input data and Random Forest model would be needed to determine if the DOE could be modified to concentrate better on finding a better optimum noise level



It is possible that local minimums have been found for frequency and angle of attack – see the next page

Minimizing Noise Using Minitab

- Minitab's Response Optimizer was used to determine settings for the variables to produce minimum noise within the design space of the DOE
- Predicts a minimum noise of 104.1 dB within DOE range
  - Frequency and Angle of Attack show effects of curvature and the model may have found optimal operating points for these two variables
  - Velocity and Thickness have linear response and could perhaps be further optimized



Comparing Minitab Result with the Model

- Minitab's Response Optimizer was used to determine settings for the variables to produce minimum noise within the design space of the DOE
- Predicts a DOE minimum noise of 104.1 dB
- Random Forest minimum noise: 104.0 dB
  - Using the optimal settings predicted by the DOE and inputted into the Random Forest Model
- The is amazing agreement between the DOE prediction and the machine learning simulation result

Citation

Citation: Airfoil Self Noise Data Set (2014). UCI Machine Learning Repository. Retrieved from: https://archive.ics.uci.edu/ml/datasets/airfoil+self-noise.