Case Study: Turbulence

Julia Rosner, Emily Mittleman, Tracy , *Ashley*

2022-11-3

Introduction

Understanding and predicting turbulence in fluid motion is incredibly important to a vast range of problems, such as air pollution, population dynamics, and weather. However, turbulence looks random, irregular, unpredictable, making it difficult to understand. Thus, our goals are as follows: For a new parameter setting of (Re, Fr, St), predict its particle cluster volume distribution in terms of its four raw moments. Inference: Investigate and interpret how each parameter affects the probability distribution for particle cluster volumes.

Methodology

Linear Modeling Our univariate exploratory data analysis of Re, Fr, St, and each moment revealed that the R moments are heavily right skewed, which poses a problem to linear regression. We applied log transformations on each moment to obtain more normally distributed variables. The log-transformed R moments are approximately normal, so we will use these values for inference and prediction. It appears that the relationship between St and R moments are linear. From what we can see, treating the Fr variable as categorical, there is also a linear and decreasing relationship between Fr (gravitational acceleration) and R moments. Lastly, there is also a linear and decreasing relationship between Re (Reynold's number) and R moments.

Accordingly, we fit a basic linear model onto each log-transformed response variable. Linear models were evaluated based on adjusted R^2 values and $\Pr(>|\text{t-value}|)$ for coefficient estimates. We chose to treat Re and Fr as factors or categorical variables, as Re only takes on the values of 90, 224, and 398; Fr only takes on the values 0.052, 0.3, and infinity. While the adjusted R^2 value for the first raw moment was very high at 0.9949, subsequent moments exhibited decreasing adjusted R^2 values, with the fourth raw moment having an adjusted R^2 value of 0.6518. We explored multicollinearity through VIFs for each model, which were very low. We also explored the addition of interaction terms to the model. The only interaction term which was significant for all raw moments was the interaction between Re and Fr. St and Re only have significant interaction for the first moment. Therefore, we constructed the linear models with an interaction between Re and Fr:

$$log(R_{moment1}) = -2.73 + 0.25(st) - 3.816(Re_{224}) - 5.988(Re_{398}) - 0.263(Fr_{0.3}) - .329(Fr_\infty) + 0.221(Re_{224} * Fr_{0.3})$$
$$+ 0.402(Re_{224} * Fr_\infty) + 0.502(Re_{398} * Fr_\infty)$$

$$log(R_{moment2}) = 5.187 + 0.834(st) - 7.434(Re_{224}) - 11.384(Re_{398}) - 6.416(Fr_{0.3})$$

Adding the interaction term between Re and Fr improved the fit of the model according to the adjusted R^2 values, which are much higher for every moment. With this new interaction term included, the adjusted R^2 value for R_moment_1 was slightly higher than before at 0.9966, and increased for moment 2 at 0.8909, moment 3 at 0.8770, and moment 4 0.8809 respectively.

To analyze predictive performance of our models, we split data into training and testing sets to evaluate the predictive ability of the models we explored. The linear models with the interaction term for Re and Fr outperformed any other linear model, producing lower test MSEs for every moment of R. Having this interaction term significantly improved the test MSEs of the linear model.

- interpret interaction between Re and Fr physically
- Are the effects identified above similar over all central moments (i.e., over all response variables), or are there effects which differ between, say, the mean and the variance? Try to interpret the latter effects using the three parameters mean physically
- central vs raw moments?

Polynomial Regression

Results

- base linear model: St doesn't even seem to be a significant predictor in some of these models. Linear regression might not work super well here.
- interaction terms: Re and g(Fr) appear to be significant, but St doesn't have any significant interactions.
- overall linear regression both with and without interaction terms: MSE comparable for each moment, Really bad results, interaction terms aren't helping much. Let's try some other methods other than simple linear regression.
- splines without interaction: better performance than linear model, but still room for improvement
- splines with interaction: Adding the interaction term to the splines makes the performance of this model superior to anything we've used previously (when considering numerical datapoints) based on Adj R^2 and MSE in test validation.

Conclusion