STA 325 Case Study

Jeff Du, Sahil Tilak

Introduction

Understanding fluid motion and turbulence is one of most challenging problems that physicists face. This case study will examine the relationship between three key properties of particles (Reynolds Number, Stokes Number, and Froud Number) and the probability distribution of particle cluster volumes. Through exploratory data analysis and nonlinear regression models, we are able to make predictions about the distribution of particle cluster volumes given parameter settings in terms of their four moments as well as better understand the way in which each parameter affects distribution.

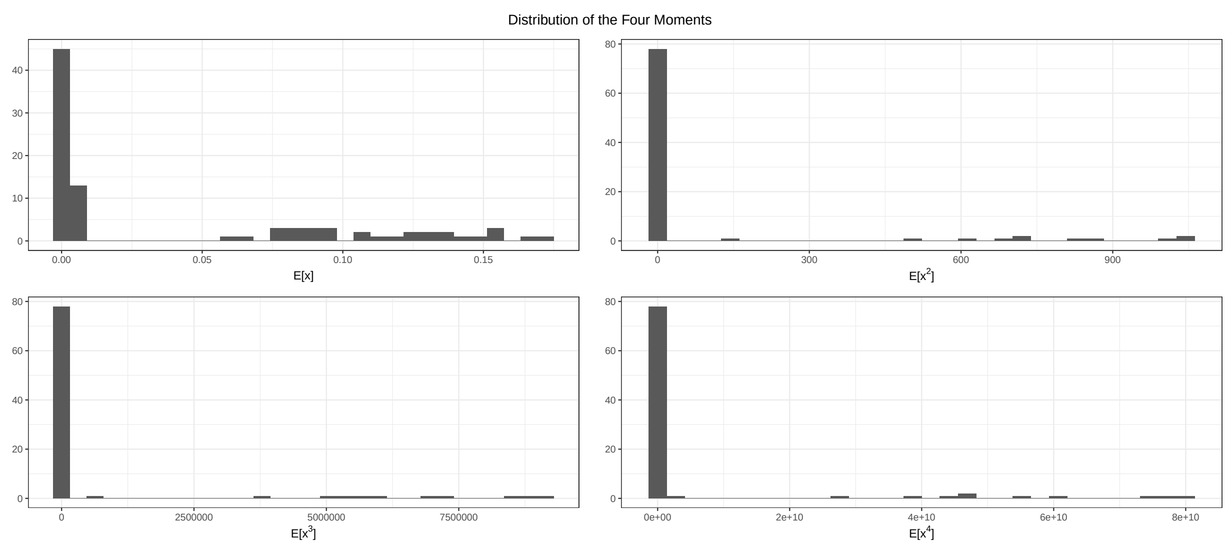
The dataset consists of 7 columns and 89 rows. Each row contains data from a simulation with different particle parameters. There are three predictor variables, or parameters:

Reynolds Number: a measure of the intensity of fluid turbulence of the particle. There are three values of Reynolds Number in the dataset: 90 (baseline value in regression), 224, and 338.

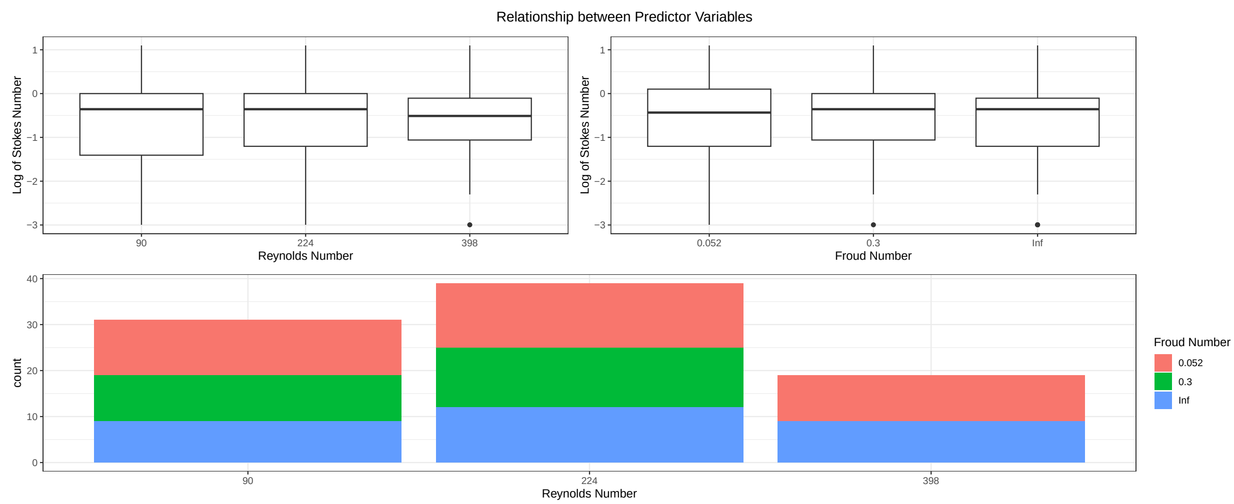
Stokes Number: a measure of the particle size and density. Values lie on the interval $[0, 3]$ in the dataset.

Froud Number: a measure of the gravitational acceleration of the particle. There are three values in the dataset: ∞, 0.3, and 0.052 (baseline value in regression).

There are four response variables, which are each of the four moments ($\mathbb{E}[x]$, $\mathbb{E}[x^2]$, $\mathbb{E}[x^3]$, $\mathbb{E}[x^4]$) of the probability distribution function of Voronoï volumes. We can examine the distributions through histograms.



We can see that all four moments are right skewed and not normally distributed. We also wanted to examine the relationship between the predictors to determine if there is value in including interaction effects. To do so, we plot boxplots to measure the relationship between the continuous and categorical variables (i.e. the relationship between Stokes Number and Reynolds/Froud Number.



We notice potential relationships between all three combinations of predictor variables.

Methodology

To examine the relationship between a particle's fluid turbulence, gravitational acceleration, and density on the four moments, we fit four nonlinear regression models with each moment as the response. We considered a number of transformations on each moment, such as a Box-Cox transformation, but ultimately decided to log transform each moment to have a clear interpretation of our subsequent regression coefficients. We also log transformed each particle's Stokes number as this predictor was far from normally distributed as well.

Regarding our predictors, we first converted both the Reynolds number and the Froud number of each particle to categorical predictors, because there only existed three unique values for each parameter, and because the numerical differences between such values of Reynolds and Froud numbers was not easily interpretable. We posited from our background research that fitting three interaction, effects for all three predictors would better model the relationship between such predictors. For example, it is well established in existing research that fluid particle acceleration (Fr) is innately related to the turbulence of a flow (Re), in line with the Kolmogorov microscales.Therefore, we included all three potential interactions in our model (Stokes number interactions are log-transformed).

We also considered adding high order polynomials of the log of the Stokes number. We ran analysis of variance tests fitting models of varying degrees of log of the Stokes number and found that the quartic fit appeared to be reasonable for all moments. Therefore, our general model for each moment is as follows:

$$
\begin{aligned}
Y = \beta_0 + \beta_1 log(Stokes) + \beta_2 log(Stokes)^2 + \beta_3 log(Stokes)^3 + \beta_4 log(Stokes)^4 + \beta_5 Reynolds \\
+ \beta_6 Froud + \beta_7(log(Stokes) * Froud) + \beta_8(log(Stokes) * Reynolds) + \beta_9(Froud * Reynolds)
\end{aligned}
$$

Finally, we ran both stepwise forward and backward variable selections on each of our four models using AIC as our criteria. We personally care less about penalizing more complex models since we are given so few predictors to begin with anyways. Forward and backward selection did not remove or add any variables to the second, third, and fourth moment models, but did remove the interaction term between the Reynolds number and the log of the Stokes number for the first moment model. However, we still decided to include this interaction because we believe it is scientifically grounded.

Results

Here we display the results for our four regression models.

Table 1: Model 1

	Estimate	Std. Error	Pr(>|t|)
Intercept	-2.05	0.0199	7.05e-81
Log(St)	1.5	0.141	1.84e-16
Log(St)^2	0.196	0.0698	0.0064
Log(St)^3	0.471	0.0702	3.71e-09
Log(St)^4	-0.153	0.0706	0.0334
Re 224	-3.81	0.0288	9.75e-89
Re 398	-6.01	0.0311	1.22e-100
Fr 0.3	-0.237	0.031	5.94e-11
Fr Inf	-0.272	0.0327	3.64e-12
Fr 0.3 * Log(St)	0.102	0.0184	4.7e-07
Fr Inf * Log(St)	0.0647	0.016	0.000123
Re 224 * Log(St)	0.0136	0.0155	0.382
Re 398 * Log(St)	-0.0159	0.0203	0.436
Re 224 * Fr 0.3	0.261	0.0401	8.57e-09
Re 244 * Fr Inf	0.381	0.0409	4.68e-14
Re 398 * Fr Inf	0.488	0.0446	4.81e-17

Table 3: Model 3

	Estimate	Std. Error	Pr(>|t|)
Intercept	14.6	0.252	9.39e-63
Log(St)	19.3	1.79	1.07e-16
Log(St)^2	-8.55	0.886	1.14e-14
Log(St)^3	6.72	0.891	1.01e-10
Log(St)^4	-3.61	0.895	0.000133
Re 224	-11.5	0.365	3.31e-44
Re 398	-18.1	0.395	1.76e-55
Fr 0.3	-13	0.393	8.41e-46
Fr Inf	-13	0.415	4.19e-44
Fr 0.3 * Log(St)	0.14	0.234	0.55
Fr Inf * Log(St)	-0.622	0.202	0.00298
Re 224 * Log(St)	-0.312	0.196	0.116
Re 398 * Log(St)	-1.08	0.257	7.65e-05
Re 224 * Fr 0.3	8.97	0.508	4.9e-28
Re 244 * Fr Inf	8.74	0.519	7.14e-27
Re 398 * Fr Inf	13.2	0.566	1.21e-35

Table 2: Model 2

	Estimate	Std. Error	Pr(>|t|)
Intercept	6.01	0.146	3.15e-52
Log(St)	12.1	1.04	2.4e-18
Log(St)^2	-5.01	0.514	7.86e-15
Log(St)^3	4.12	0.517	1.58e-11
Log(St)^4	-2.16	0.52	8.74e-05
Re 224	-7.61	0.212	4.05e-48
Re 398	-12	0.229	1.28e-59
Fr 0.3	-6.74	0.228	2.24e-42
Fr Inf	-6.76	0.241	7.56e-41
Fr 0.3 * Log(St)	0.108	0.136	0.428
Fr Inf * Log(St)	-0.327	0.118	0.00694
Re 224 * Log(St)	-0.136	0.114	0.236
Re 398 * Log(St)	-0.627	0.149	7.48e-05
Re 224 * Fr 0.3	4.75	0.295	1.01e-25
Re 244 * Fr Inf	4.68	0.301	7.22e-25
Re 398 * Fr Inf	6.99	0.329	5.15e-33

Table 4: Model 4

	Estimate	Std. Error	Pr(>|t|)
Intercept	23.2	0.348	3.92e-67
Log(St)	25.6	2.48	6.41e-16
Log(St)^2	-11.6	1.22	2.03e-14
Log(St)^3	9.01	1.23	2.59e-10
Log(St)^4	-4.87	1.24	0.000187
Re 224	-15.4	0.504	2.92e-43
Re 398	-24.1	0.545	1.97e-54
Fr 0.3	-19.2	0.543	8.88e-48
Fr Inf	-19.2	0.573	4.96e-46
Fr 0.3 * Log(St)	0.153	0.323	0.637
Fr Inf * Log(St)	-0.901	0.28	0.0019
Re 224 * Log(St)	-0.468	0.271	0.0888
Re 398 * Log(St)	-1.46	0.355	0.000103
Re 224 * Fr 0.3	13.1	0.702	1.85e-29
Re 244 * Fr Inf	12.7	0.717	3.12e-28
Re 398 * Fr Inf	19.4	0.782	2.6e-37

The adjusted R-Squared values for each model were .9991, .9814, .9765, and .9757 respectively, indicating that essentially all of the variation in the four moments can be explained by the combination of the predictor variables that we chose. Using an alpha of .05 as a threshold for statistical significance, we find that every individual predictor is statistically significant in each of the four models, though both the sign and magnitude of the coefficients vary by moment.