

STA 325 Case Study

Jeff Du, Sahil Tilak

Introduction

Understanding is one of most challenging problems that physicists face today. This case study will attempt to understand how the relationship between three key properties of particles (Reynolds Number, Stokes Number, and Froud Number) and the probability distribution of particle cluster volumes.

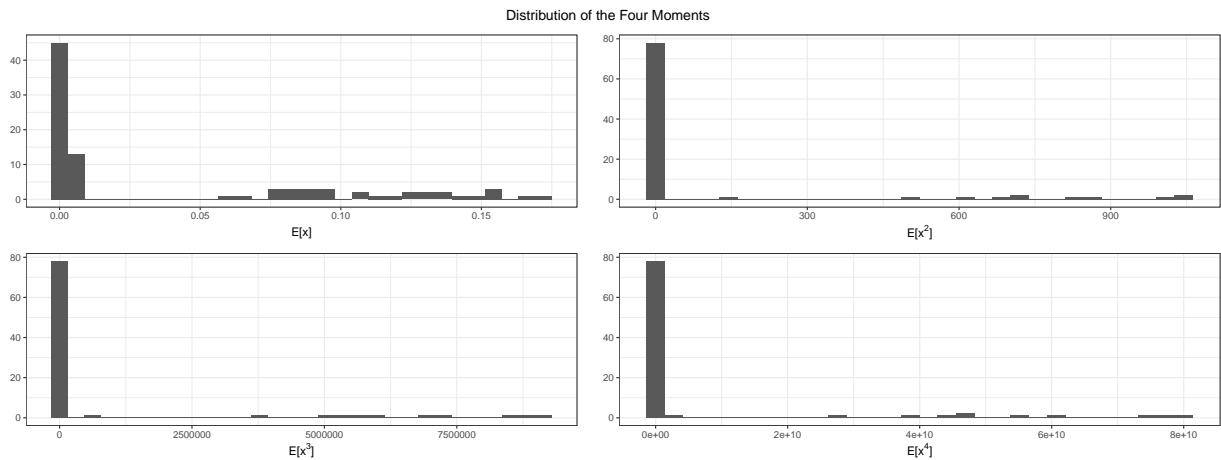
Through exploratory data analysis and nonlinear regression models, we are able to make predictions about the distribution of particle cluster volumes given parameter settings, as well as understand the way in which each parameter affects the distribution.

The dataset consists of 7 columns and 89 rows. Each row contains data from a simulation with different There are three predictor variables, or parameters: *Re*: Reynolds Number, which measures the intensity of fluid turbulence of the particle. There are three values of Reynolds Number in the dataset: 90, 224, and 338 *St*: Stokes Number, which measures the size and density of the particle. Stokes Number lies on the interval $[0, 3]$ in the dataset *Fr*: Froud Number, which measures the gravitational acceleration of the particle. There are three values in the dataset: ∞ , 0.3, and 0.052

There are four predictor variables, which are each of the four moments ($E[x]$, $E[x^2]$, $E[x^3]$, $E[x^4]$) of the probability distribution function of Voronoï volumes, which serve as a measure of the clustering of particles

EDA

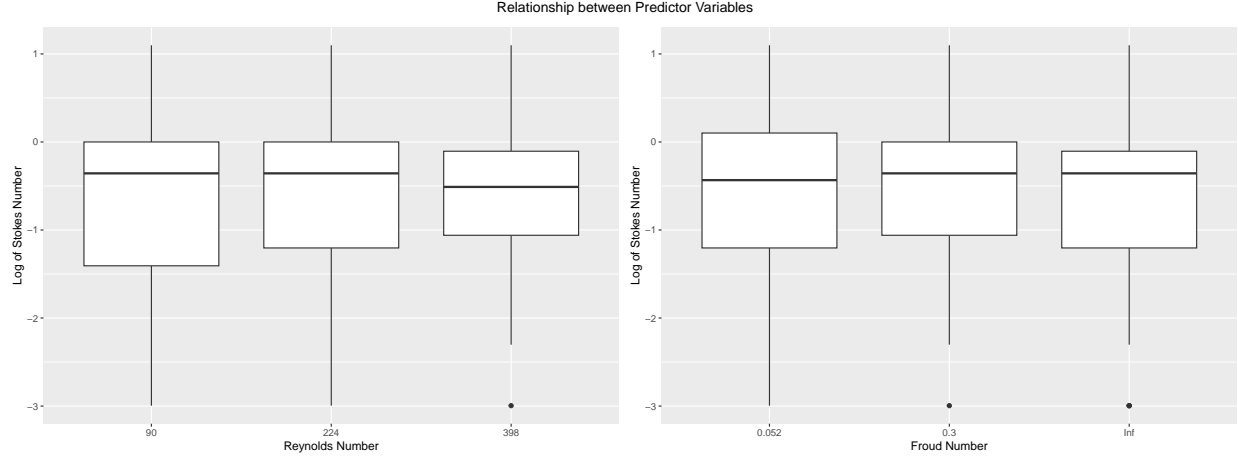
First, we chose to look at the distribution of the four response variables



We can see from an histogram of the response variables that all four moments are right skewed and not normally distributed. To account for this, we chose to log-transform the response variables to ensure they are 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 a continuous and categorical

variable (i.e. the relationship between Stokes Number and Reynolds/Froud Number) and a barplot to show the interaction between Reynold's Number and Froud Number



We notice that there are may be relationships all three combinations of predictor variables, so we will include the interaction effects in our intial model.

Methodology

To examine the relationship between a particle's fluid turbulence, gravitational acceleration, and density on the four moments of the cluster probability distribution, we fit four nonlinear regression models with each moment as an individual response. We examined through our EDA that the four moments are far from normally distributed, and so we considered a number of transformations on each moment, such as a Box-Cox transformation. Ultimately, we 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 (St) 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 a categorical predictor, because there only existed three unique values for each predictor in the dataset, 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 each of the 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).

In terms of increasing our model complexity, we considered using high order polynomials of $\log(\text{St})$ for each of the four moments. We ran analysis of variance tests fitting different models of varying degrees of $\log(\text{St})$ and found that the quartic fit appeared to be reasonable for all of the moments.

Therefore, our general model for each moment is as follows:

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

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

Model 1:

Model 1:					Estimate	Std. Error	t value	Pr(> t)
				(Intercept)	14.6	0.252	57.9	9.39e-63
(Intercept)	Estimate	Std. Error	t value	poly(log(St), 4)1	19.3	1.79	10.7	1.07e-16
poly(log(St), 4)1	1.5	0.141	10.6	poly(log(St), 4)2	-8.55	0.886	-9.65	1.14e-14
poly(log(St), 4)2	0.196	0.0698	2.81	poly(log(St), 4)3	6.72	0.891	7.54	1.01e-10
poly(log(St), 4)3	0.471	0.0702	6.7	poly(log(St), 4)4	-3.61	0.895	-4.03	0.000133
poly(log(St), 4)4	-0.153	0.0706	-2.17	Re224	-11.5	0.365	-31.5	3.31e-44
Re224	-3.81	0.0288	-133	Re398	-18.1	0.395	-45.7	1.76e-55
Re398	-6.01	0.0311	-193	Fr0.3	-13	0.393	-33.2	8.41e-46
Fr0.3	-0.237	0.031	-7.67	FrInf	-13	0.415	-31.4	4.19e-44
FrInf	-0.272	0.0327	-8.31	Fr0.3:log(St)	0.14	0.234	0.6	0.55
Fr0.3:log(St)	0.102	0.0184	5.53	FrInf:log(St)	-0.622	0.202	-3.07	0.00298
FrInf:log(St)	0.0647	0.016	4.06	Re224:log(St)	-0.312	0.196	-1.59	0.116
Re224:log(St)	0.0136	0.0155	0.88	Re398:log(St)	-1.08	0.257	-4.19	7.65e-05
Re398:log(St)	-0.0159	0.0203	-0.784	Re224:Fr0.3	8.97	0.508	17.6	4.9e-28
Re224:Fr0.3	0.261	0.0401	6.5	Re224:FrInf	8.74	0.519	16.9	7.14e-27
Re224:FrInf	0.381	0.0409	9.32	Re398:FrInf	13.2	0.566	23.4	1.21e-35
Re398:FrInf	0.488	0.0446	10.9	Model 17				

Model 2:

Model 2:					Estimate	Std. Error	t value	Pr(> t)
	Estimate	Std. Error	t value	(Intercept)	23.2	0.348	66.6	3.92e-67
(Intercept)	6.01	0.146	41.1	poly(log(St), 4)1	25.6	2.48	10.3	6.41e-16
poly(log(St), 4)1	12.1	1.04	11.7	poly(log(St), 4)2	-11.6	1.22	-9.52	2.03e-14
poly(log(St), 4)2	-5.01	0.514	-9.74	poly(log(St), 4)3	9.01	1.23	7.32	2.59e-10
poly(log(St), 4)3	4.12	0.517	7.97	poly(log(St), 4)4	-4.87	1.24	-3.94	0.000187
poly(log(St), 4)4	-2.16	0.52	-4.15	Re224	-15.4	0.504	-30.5	2.92e-43
Re224	-7.61	0.212	-35.9	Re398	-24.1	0.545	-44.2	1.97e-54
Re398	-12	0.229	-52.3	Fr0.3	-19.2	0.543	-35.5	8.88e-48
Fr0.3	-6.74	0.228	-29.6	FrInf	-19.2	0.573	-33.5	4.96e-46
FrInf	-6.76	0.241	-28.1	Fr0.3:log(St)	0.153	0.323	0.474	0.637
Fr0.3:log(St)	0.108	0.136	0.797	FrInf:log(St)	-0.901	0.28	-3.22	0.0019
FrInf:log(St)	-0.327	0.118	-2.78	Re224:log(St)	-0.468	0.271	-1.73	0.0888
Re224:log(St)	-0.136	0.114	-1.2	Re398:log(St)	-1.46	0.355	-4.11	0.000103
Re398:log(St)	-0.627	0.149	-4.2	Re224:Fr0.3	13.1	0.702	18.6	1.85e-29
Re224:Fr0.3	4.75	0.295	16.1	Re224:FrInf	12.7	0.717	17.8	3.12e-28
Re224:FrInf	4.68	0.301	15.5	Re398:FrInf	19.4	0.782	24.8	2.6e-37
Re398:FrInf	6.99	0.329	21.3					

Model 3: