Case Report

Jack Lichtenstein, Abbey List, Jingxuan Liu, Linda Tang, Mary Wang, and Justin Zhao

2021-10-14

1 Introduction

Turbulence is a fundamental concept in fluid mechanics. Irregular, unpredictable, and energy-dissipating, turbulent flow enhances mixing, which leads to non-uniform distribution of particles and cluster formation. Understanding and predicting turbulence has great practical significance in many scientific areas including aeronautical engineering and environmental science.

Existing research suggests that the clustering of particles subject toturbulent flow is mainly determined by three main parameters: Reynolds (Re) number, Froude (Fr) number and Stokes (St) number, which correspond to the intentsity of turbulence, impact of gravitational acceleration, and particle properties (Chanson, 2009; Ireland et al., 2016). These parameters may influence clustering individually; a large St, for example, correlates with large particle size, which tends to form relatively loose clusters (Ireland et al., 2016). The parameters may also interact with each other to impact cluser formation.

Despite the importance of turbulence, current understanding of *how* Re, Fr, and St contribute to clustering remains rudimentary. Similation methods like the Direct Numerical Simulation (DNS) of Navier-Stokes equations have been applied, but progress is limited by the time-consuming and computationally-expensive nature of such methodologies (Moin & Mahesh, 1998). Leveraging on generated data, the present study *aims* to build a statistical model that investigates how Re, Fr and St influence particle cluster volume distribution. We hope that our model will 1) enhance our understanding of the relative influence of these parameters in turbulent flow, and 2) enable a quick and efficient prediction of a particle cluster volume distribution without the needs for simulation.

2 Methods

2.1 Data

The data ($n = 89$) for the present study comes from Direct Numerical Simulation. Voronoi Tessellation, a technique that examines general features of individual clusters in the underlying turbulence, was applied to generate a distribution of cluster volumes. The original data contains information regarding the generated distributions in the form of the first four raw moments $E(X)$, $E(X^2)$, $E(X^3)$, and $E(X^4)$, as well as Re, Fr, and St values.

For better interpretability, we transformed the 2nd, 3rd, and 4th raw moments into central moments (Moment 1 is untouched as it already signifies the mean), which describe the variance (how flow varies over time), skewness (indication of symmetric properties of the flow), and kurtosis of the distribution (how particular cluster volumes deviate). In addition, despite their numerical nature, Fr and Re only contains three levels ($Fr \in \{0.052, 0.3, \text{inf}\}$ and $Re \in \{90, 224, 398\}$, see Figure S1 for distribution) in the data. Given such, the variables were converted into factors (See Section 2.4 for an alternative modeling approach that may overcome this limitation).

2.2 Model Building

A closer examination of the data revealed interesting interactive patterns among the independent and dependent variables. Specifically, St appears to assume a strong, non-linear relationship with each of the moments (Figures S1-S4 in the Appendix). These relationships vary between roughly linear to noticeably curved, potentially quadratic or cubic. In addition, such relationships diverge across different moments, and appears highly influenced by the combined levels of Re and Fr. From the curved shape of the relationship between `St` and the responses, we also noticed that we may benefit from taking certain roots of `St` to make that more linear.

Given the paucity of theoretical background in related field to guide model building, we decided to adopt a k-fold validation approach. This approach would allow us to explore different combinations of polynomial patterns and the interactions observed in our EDA and select the best fit model. Specifically, we used K = 5 given the small sample size at hand to avoid overfitting.

To implement the k-fold cross validation, for each moment, we trained candidate models to predict the moment with the general formula $Moment_i \sim poly(St_i^{(1/root)}, degree) * (Fr, Re)$, varying the *degree* parameter from 1 to 3, *root* from 1 to 6, and testing a log transformation of the response. We combined Fr and Re to form a new interaction variable (Fr, Re) with 9 levels representing all combinations of Fr and Re. We tested the model on each fold, and chose the parameters that give lower root mean squared errors (See Figure 1), with preference for less complexity if the error is similar. After selecting the features for each moment's model, we then fit the final models on the full training dataset using standard least-squares for interpretation.

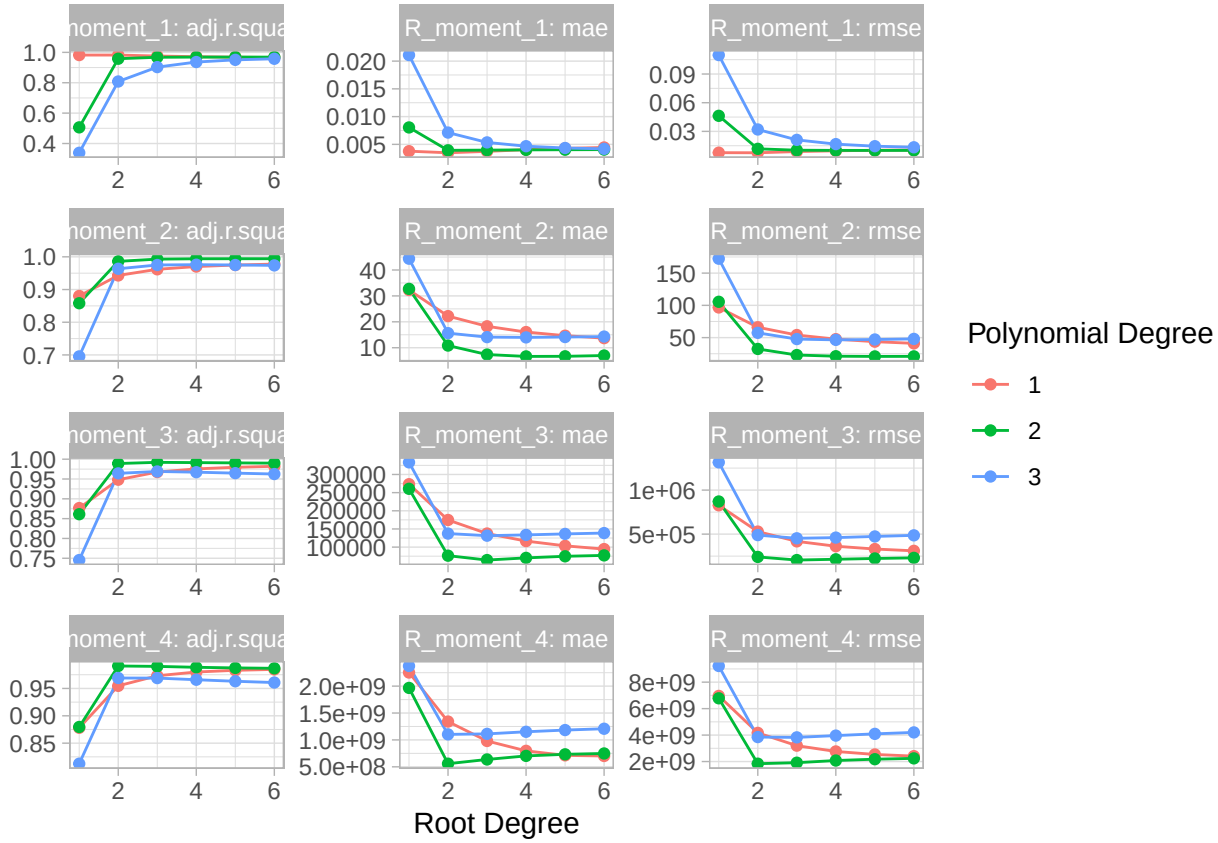


Figure 1. Mean adjusted r-squared value, mean absolute error, and root mean squared error of models for each moment with varying root and polynomial degrees on St. The combination that lead to the lowest root mean squared error was selected as the final model for the moment.

2.3 Final Models

The final models for each moment, based on the k-fold cross-validation approach, are as below:

$$Raw\ Moment_1 \sim St * (Fr, Re)$$

$$Central\ Moment_2 \sim poly(St^{1/4}, 2) * (Fr, Re)$$

$$Central\ Moment_3 \sim poly(St^{1/3}, 2) * (Fr, Re)$$

$$Central\ Moment_4 \sim poly(\sqrt{St}, 2) * (Fr, Re)$$

2.4 Model Extension

As noted above, extrapolating beyond the three levels of Re and three levels of Fr given may be useful for more general predictions and interpretation, since the actual variables have wide domains (@Jing how to word this). To deal with the limitations of our current models, we also provide the following related models that can be used for extrapolation:

$$R\_moment_1 \sim ns(St, df = 1) * Re * Fr'$$

$$R\_moment_2, R\_moment_3, R\_moment_3 \sim ns(\sqrt{St}, df = 1) * Re * Fr'$$

In these models, Re is numeric and Fr is transformed to a numeric variable on $[0, 1]$ using $Fr' = \frac{2}{\pi} * arctan(Fr)$. The form is similar in keeping the significant three-way interaction, except with natural splines to address the poor fits of polynomials at the tails, a location that is especially important in learning about particle behavior in high turbulence. Again, the root and degree are chosen through 5-fold cross validation. See appendix for fitted coefficients.

Results

Selected outputs from the final models of each moment are displayed in Table 1. Please refer to Table S1 for the error values for the final models and Tables S2-S5 for full outputs.

Term	β	SE	t	p value
Moment 1				
interaction0.052: 90	0.127	0.002	57.356	<0.001
interaction0.3: 90	0.098	0.002	41.71	<0.001
interactionInf: 90	0.093	0.002	38.808	<0.001
poly(St, 1):interaction0.052: 90	0.137	0.02	6.86	<0.001
poly(St, 1):interaction0.3: 90	0.26	0.021	12.542	<0.001
poly(St, 1):interactionInf: 90	0.247	0.021	11.536	<0.001
Moment 2				
interaction0.052: 90	698.156	7.492	93.183	<0.001
poly(St, 2)1:interaction0.052: 90	2603.304	67.525	38.553	<0.001
poly(St, 2)2:interaction0.052: 90	-792.335	67.664	-11.71	<0.001
Moment 3				
interaction0.052: 90	5693362.973	71324.394	79.824	<0.001
poly(St, 2)1:interaction0.052: 90	23598512.694	643161.093	36.691	<0.001
poly(St, 2)2:interaction0.052: 90	-6676295.814	644193.322	-10.364	<0.001
Moment 4				
interaction0.052: 90	46674298604.957	650861350.994	71.712	<0.001
poly(St, 2)1:interaction0.052: 90	208144184726.544	5873317691.538	35.439	<0.001

poly(St, 2)2:interaction	0.052:	90	-66729237684.868	5891030956.742	-11.327	<0.001

Table 1. Outputs from final models of each moment. Only the significant terms are displayed here given limited space. Please refer to Tables XX-XX in Appendix for full output.

From Table S1 in the Appendix, we can observe that all models possess an adjusted R^2 of at least 0.98, indicating that at least 98% of the variation in each centralized moment is explained by the predictors in the models. The test error (modeled by root mean squared error) is only 0.00764 and 20.86 for models of centralized moments 1 and 2, respectively. It is larger for models of centralized moments 3 and 4 (above 200,000 and above 1 million, respectively), but this is expected given the extremely large values in these moments.

For centralized moment 1, we see that interactions between all settings of Fr and the lowest setting of Re, along with their interactions with St, yield an expected increase in mean particle cluster volume size. This indicates that smoother flow at any level of gravitational acceleration and particle size may allow larger clusters to form, as they are not broken up by chaotic turbulence. These coefficients are statistically significant at the $\alpha = 0.01$ level with small standard errors.

For centralized moments 2, 3, and 4, it appears that the interaction between lowest Fr and lowest Re, along with its interaction with St, has a positive association with each moment. The interaction with squared St has a negative association with each moment (note that the exact value of this squared transformation differs based on the root transformation applied to St for each moment) (?????? ask simon about interpreting positive non-squared, negative squared). This indicates that smoother flow and low gravitational acceleration are associated with greater variance, skew, and kurtosis (distributional "tailedness") of particle cluster size. The effect of particle size at these levels of Re and Fr may differ from the effect at other levels, as it is only statistically significant at these low settings of flow intensity and gravitational acceleration. These coefficients are statistically significant at the $\alpha = 0.01$ level.

Discussion

One limitation of our analysis is that Fr and Re are categorical variables in our models, while they are truly numeric. This may have a negative effect on predictive performance for extrapolation on values outside our categories; however, Section 2.4 describes a numerical model that can be used for extrapolation (did we ever provide the results for that model?).

Also, our model was fit on Reynolds numbers up to 398, although numbers in the thousands are common in practice. Large Reynolds numbers are highly relevant to real life situations as a measure of turbulence intensity in atmospheric, oceanic turbulence flow (J. den Toonder et al., 1997), and although our modeling approach avoids the high computational cost of simulation at large Reynolds numbers, it may not be suitable for extrapolation at these settings due to the small numbers in the dataset. Using numerical variables and higher Reynolds numbers may be topics for further investigation.

In this study, we have utilized four linear models with polynomial terms and interaction terms between Reynolds' number, Froude's number, and Stokes' number to predict the first four centralized moments (mean, variance, skew, and kurtosis) of particle cluster distribution. These variables explained a large percentage of variation in each of the four moments, with reasonable mean squared error for each model. We found that smoother flow at any level of gravitational acceleration and particle size may be associated with larger mean cluster size, and smoother flow with low gravitational acceleration seemed to have a positive association with variance, skew, and kurtosis. Also, the effects of particle size on each moment seemed to be significantly different at lower flow intensity and lower gravitational acceleration than at other settings. Finally, it appears that for centralized moments 2, 3, and 4, the main effect of St (at low Fr and Re) has a positive association with the moments, while the squared effect has a negative association (how to explain this ?). In summary, we hope these results can be used towards both predicting the distribution of particle cluster size as well as understanding the effects of Reynolds, Froude, and Stokes numbers on particle cluster size.

References

- J. M. J. den Toonder, & Nieuwstadt, F. T. M. (1997, November 1). Reynolds number effects in a turbulent pipe flow for low to moderate re. AIP Publishing. Retrieved October 12, 2021, from https://aip.scitation.org/doi/pdf/10.1063/1.869451.
- Schlichting, H., Gersten, K., Krause, E., & Oertel, H. (2017). Boundary-layer theory. Springer.
- Chanson, Hubert (2009). "Development of the Bélanger Equation and Backwater Equation by Jean-Baptiste Bélanger (1828)" (PDF). Journal of Hydraulic Engineering. 135 (3): 159–63. doi:10.1061/(ASCE)0733-9429(2009)135:3(159).
- Ireland, P. J., Bragg, A. D., & Collins, L. R. (2016, May 11). The effect of Reynolds number on inertial particle dynamics in isotropic turbulence. part 2. simulations with gravitational effects: Journal of Fluid Mechanics. Cambridge Core. Retrieved October 12, 2021, from https://www.cambridge.org/core/journals/journal-of-fluid-mechanics/article/effect-of-reynolds-number-on-inertial-particle-dynamics-in-isotropic-turbulence-part-2-simulations-with-gravitational-effects/67C55CDC28B1B1C7868B7A402E279AF9.
- Moin, P., & Mahesh, K. (n.d.). Direct numerical simulation: A tool in turbulence research. Annual Reviews. Retrieved October 12, 2021, from https://www.annualreviews.org/doi/full/10.1146/annurev.fluid.30.1.539.
- slides: https://sakai.duke.edu/access/content/group/e1e1b166-17bd-4efc-bdfb-f3909d696910/Case%20Study/Data_Expedition_F2020_Reza_Jon.pdf

Appendix

Figures

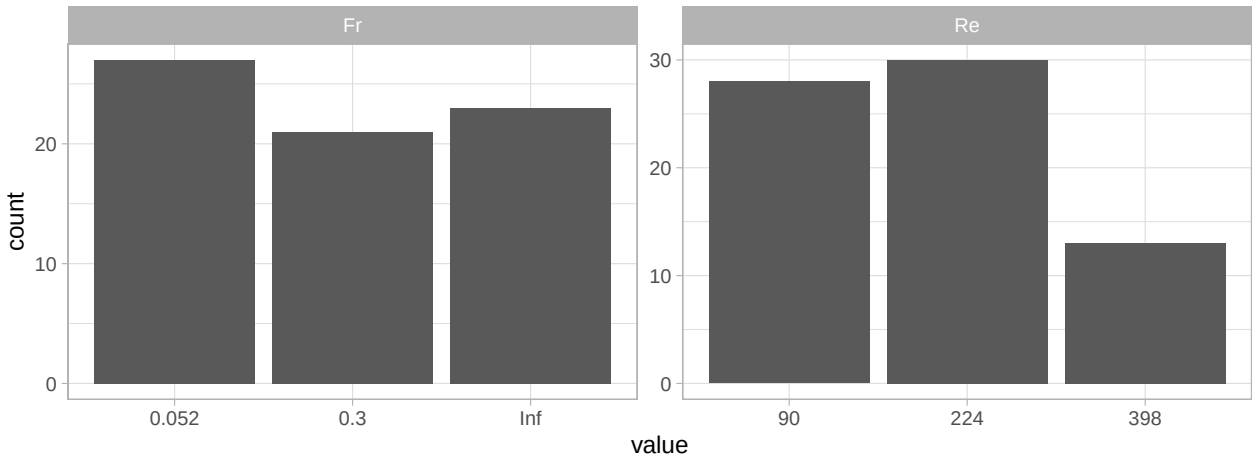


Figure S1. Distribution of the three levels of Re and Fr. The values are treated as factor levels in our main analysis.

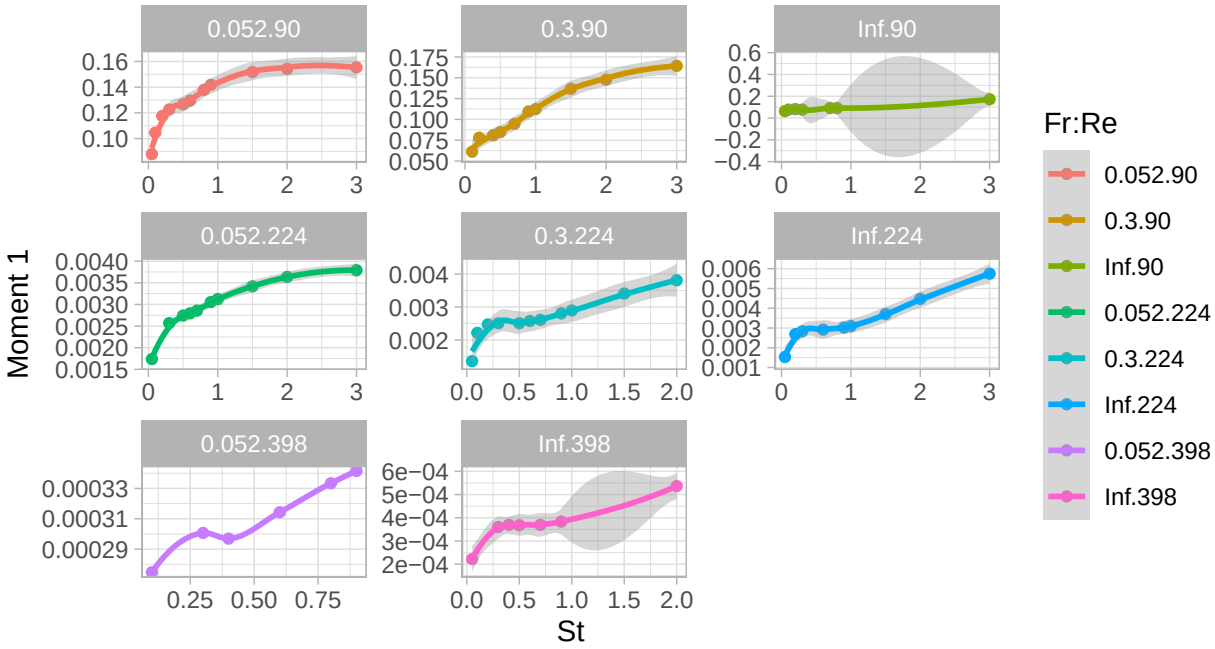


Figure S2. Moment 1 values as a function of St at different levels of interaction between Fr and Re.

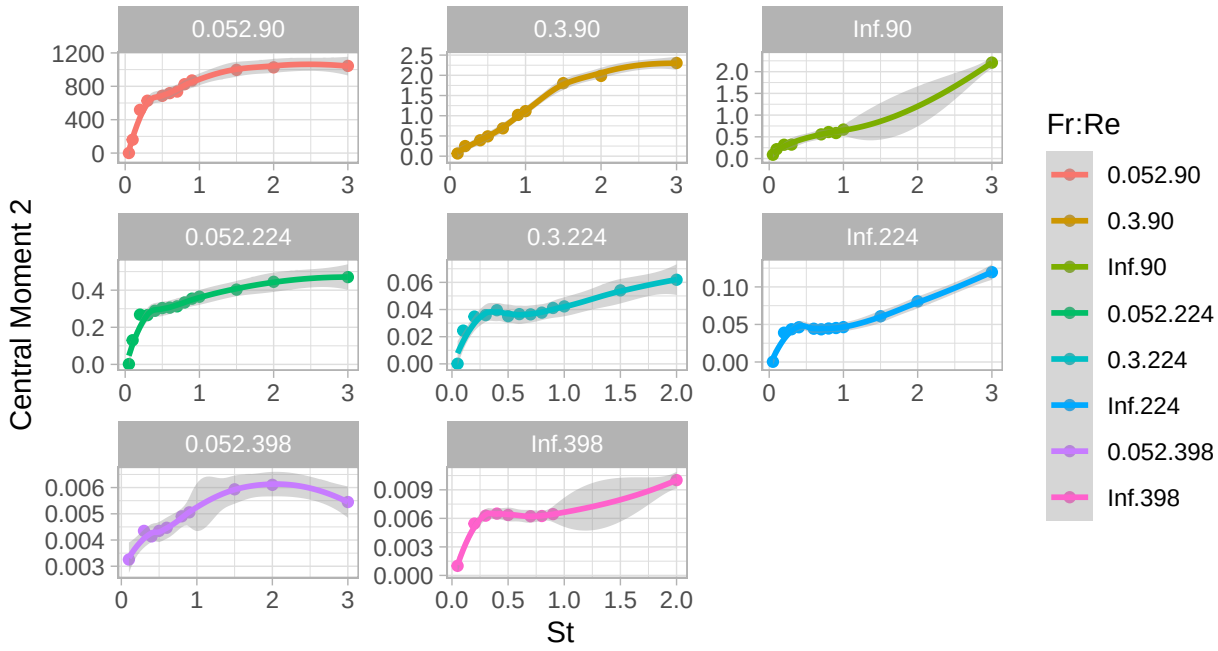


Figure S3. Central Moment 2 values as a function of St at different levels of interaction between Fr and Re.

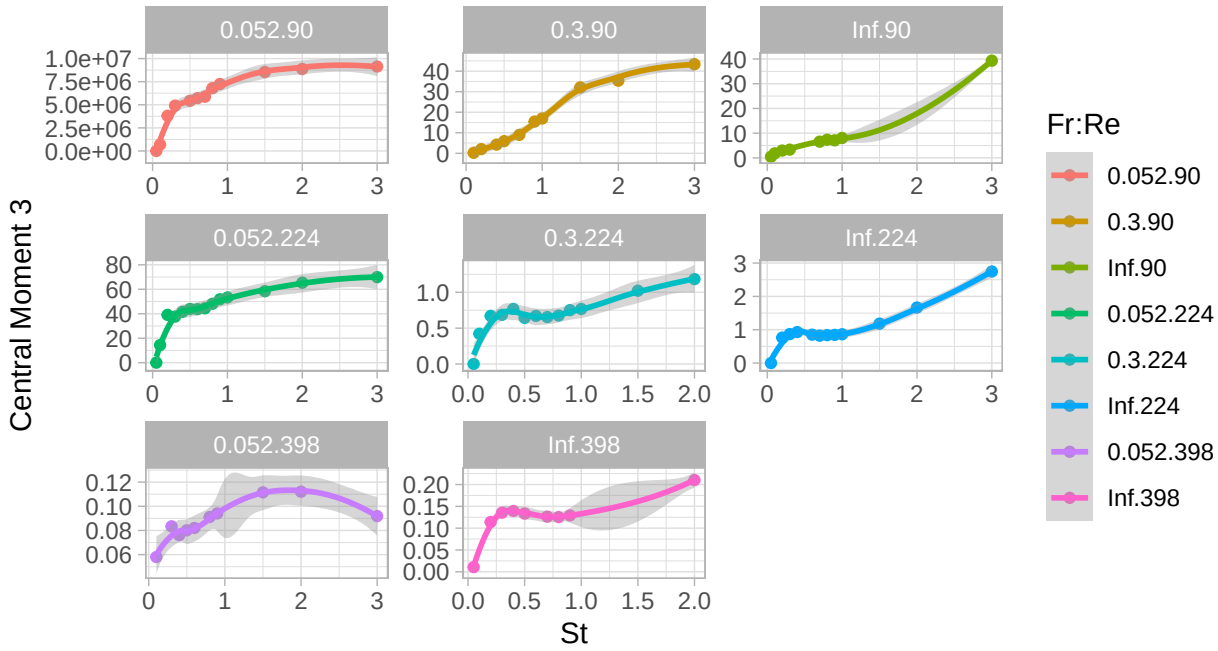


Figure S4. Central Moment 3 values as a function of particle size at different levels of interaction between Fr and Re.

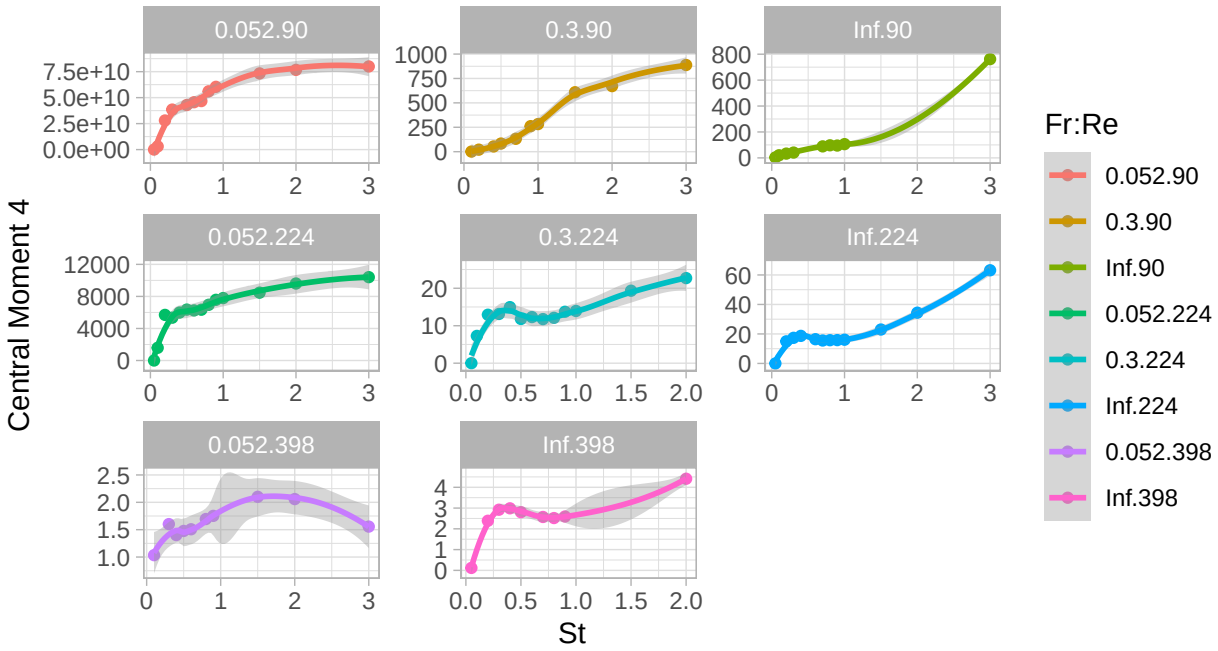


Figure S5. Central Moment 4 values as a function of particle size at different levels of interaction between Fr and Re.

Model Diagnoistics

Modeling assumptions for linear regression:

- **Linearity**: Linearity is satisfied for all 4 models since in the predicted vs. standardized residuals plots (Panel A), there's no obvious pattern in the residuals are the value of the predictors increase. The residuals are randomly scattered around 0, supporting that there's a linear relationship between the predictors and the response (after variable transformation).
- **Constant variance**: In the residuals vs. predicted plot for all 4 models, the vertical spread of the residual remain relatively constant as the predicted values increases, suggesting that the variance of the error is constant along all predicted values (Panel A).
- **Independence**: The data was generated from Direct Numerical Stimulation (DNS) of the Navier-Stokes equations, where each trial was conducted independently using different values of the parameter (Re, Fr, St). Based on the information about data collection, we believe the independence assumption is satisfied.
- **Normality**: Normality may be violated since the distribution of the residuals doesn't follow a normal distribution and the points do not fall along a straight diagonal line on the normal quantile plot. The flat region in the normal quantile plot indicates there's a lot of nearly identical values, and the curved shape suggests that the distribution may be heavy-tailed. This violation makes sense in the context of the dataset since we observed that both the raw moments and central moments are highly skewed to the right with most values close to 0 with some extreme outliers (Panel B).

Influential points and outliers: We assessed influential points in our dataset using Cook's distance (Panel C). Besides having a few influential points for model 1, all observations had Cook's distance less than 0.5 for the rest of the models. In addition, from the predicted vs. standardized residuals plots (Panel A), there are only a few outliers with standardized residual greater than 5.

Multicolinearity: We assessed multicolinearity using variance inflation factor (VIF), which was below 10 for all predictors in all 4 models, suggesting no serious issue of multicolinearity.

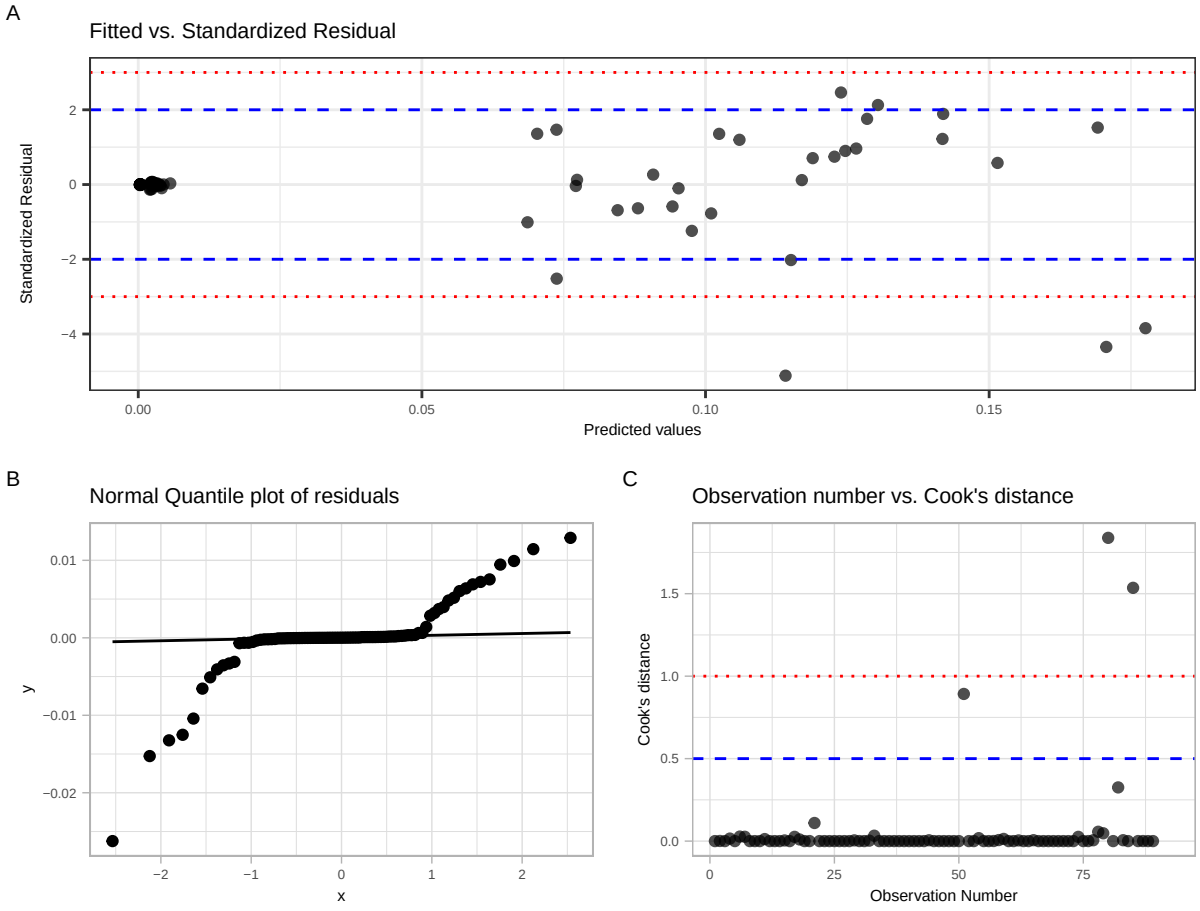


Figure S?. Diagnoistic plots for Model 1.

	GVIF	Df	GVIF^(1/(2*Df))
poly(St, 1)	6.228387	1	2.495674
interaction	1.323103	7	1.020200
poly(St, 1):interaction	8.024337	7	1.160381

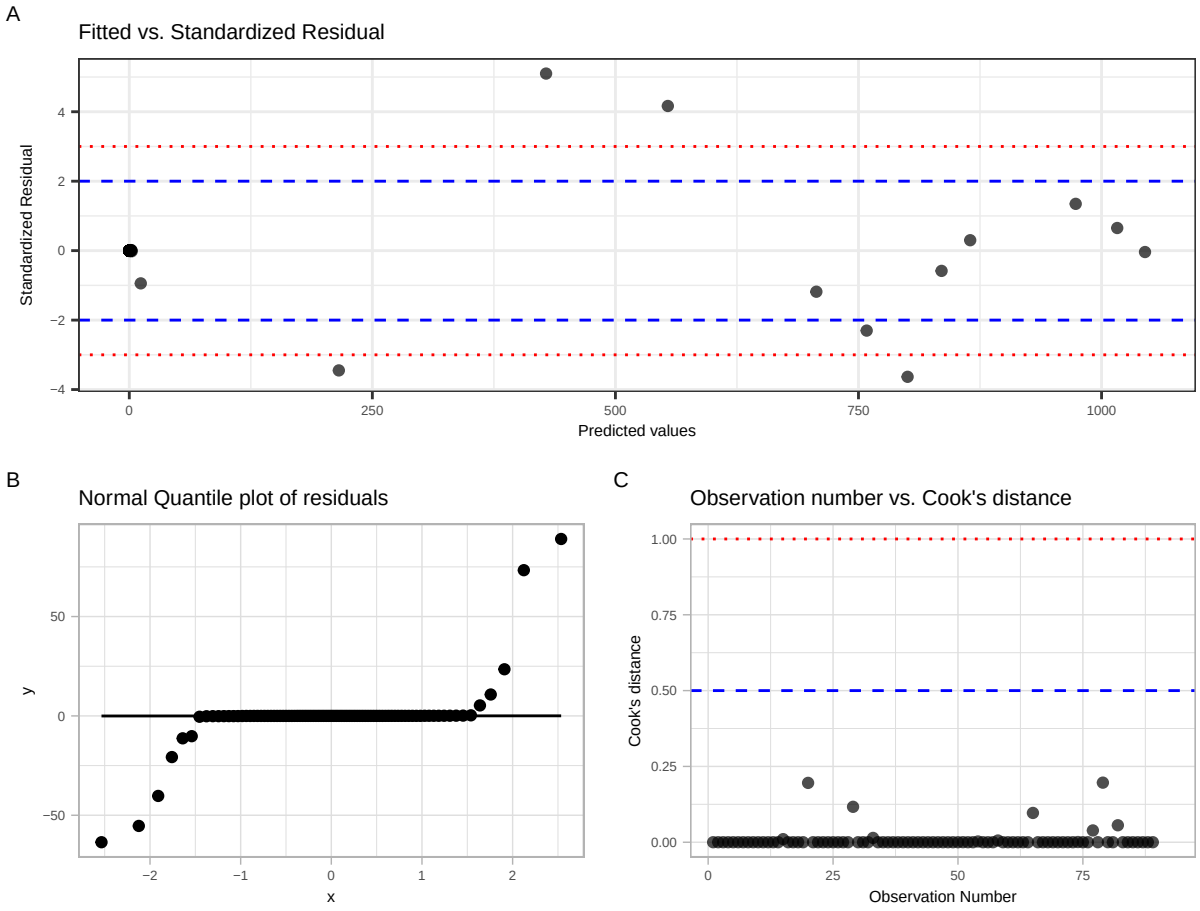


Figure S?. Diagnoistic plots for Model 2.

	GVIF	Df	GVIF^(1/(2*Df))
poly(St, 2)	38.225845	2	2.486505
interaction	1.538693	7	1.031260
poly(St, 2):interaction	56.570018	14	1.155028

	GVIF	Df	GVIF^(1/(2*Df))
poly(St, 2)	38.253169	2	2.486949
interaction	1.576793	7	1.033063
poly(St, 2):interaction	57.982268	14	1.156045

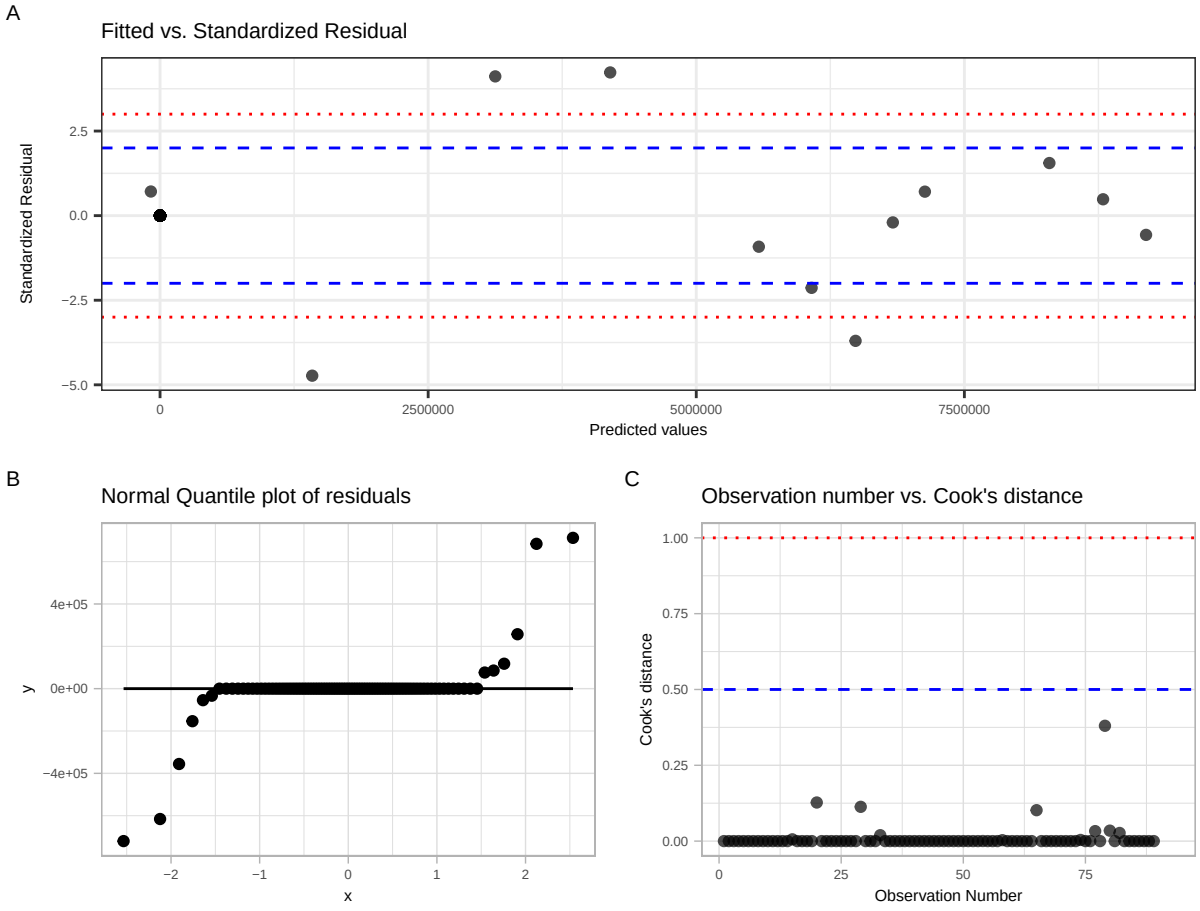


Figure S?. Diagnoistic plots for Model 3.

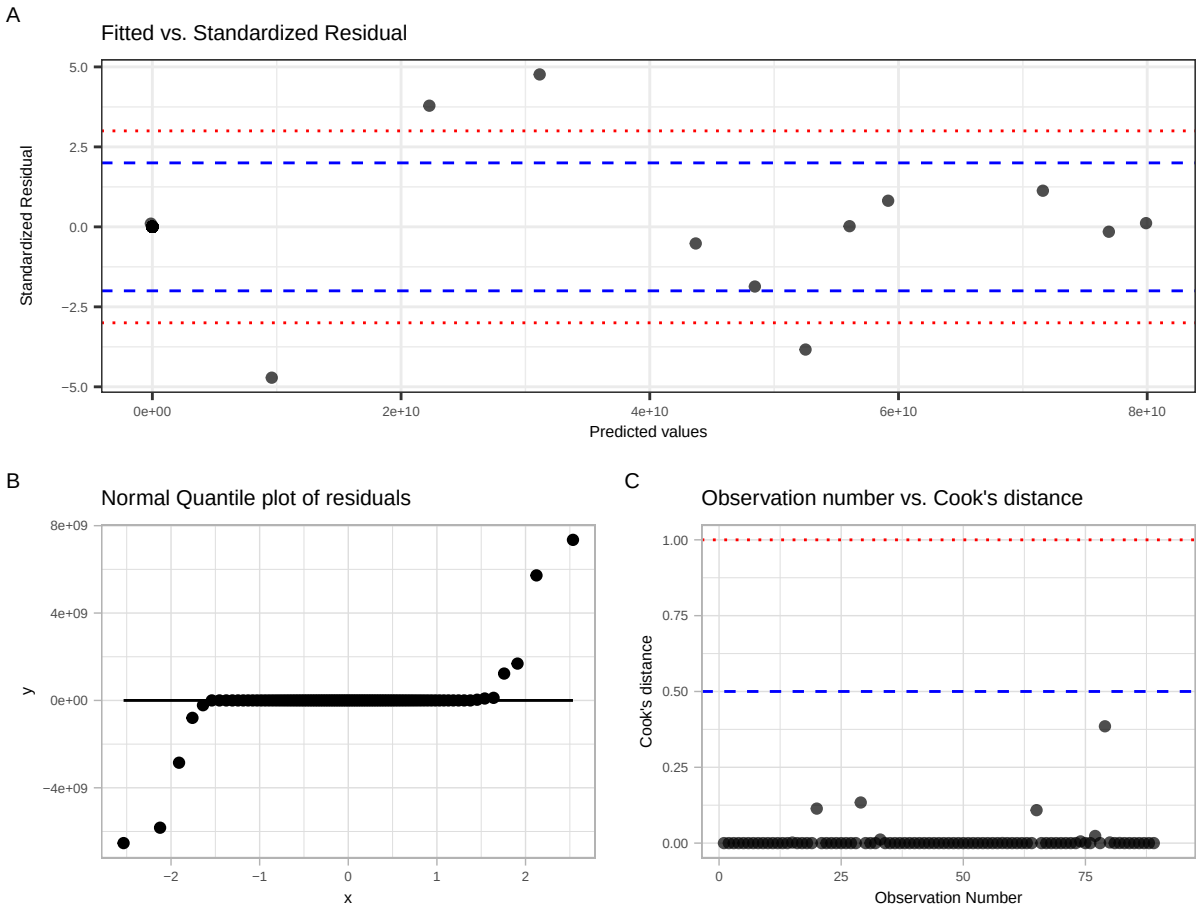


Figure S?. Diagnoistic plots for Model 4.

	GVIF	Df	GVIF^(1/(2*Df))
poly(St, 2)	38.348359	2	2.488495
interaction	1.686911	7	1.038056
poly(St, 2):interaction	62.142079	14	1.158909

moment	degree	root	log	name	value
R_moment_1	1	2	FALSE	rmse	0.00764
R_moment_1	1	2	FALSE	mae	0.00348
R_moment_1	1	2	FALSE	adj.r.squared	0.98223
R_moment_2	2	4	FALSE	mae	6.64671
R_moment_2	2	5	FALSE	rmse	20.85942
R_moment_2	2	5	FALSE	adj.r.squared	0.99428
R_moment_3	2	3	FALSE	rmse	205,406.99660
R_moment_3	2	3	FALSE	mae	64,341.96016
R_moment_3	2	3	FALSE	adj.r.squared	0.99206
R_moment_4	2	2	FALSE	rmse	1,837,891,935.00114
R_moment_4	2	2	FALSE	mae	558,101,834.24670
R_moment_4	2	2	FALSE	adj.r.squared	0.99101

Table S1. Errors from final models for each moment

Term	β	SE	t	p value
(Intercept)	0.003	0.002	1.907	0.06
poly(St, 1)	0.004	0.014	0.314	0.755
interaction0.052: 398	-0.003	0.002	-1.080	0.283
interaction0.052: 90	0.127	0.002	57.356	<0.001
interaction0.3: 224	0.000	0.002	-0.027	0.978
interaction0.3: 90	0.098	0.002	41.710	<0.001
interactionInf: 224	0.000	0.002	0.118	0.906
interactionInf: 398	-0.002	0.003	-0.985	0.328
interactionInf: 90	0.093	0.002	38.808	<0.001
poly(St, 1):interaction0.052: 398	-0.004	0.021	-0.196	0.845
poly(St, 1):interaction0.052: 90	0.137	0.020	6.860	<0.001
poly(St, 1):interaction0.3: 224	0.002	0.025	0.096	0.924
poly(St, 1):interaction0.3: 90	0.260	0.021	12.542	<0.001
poly(St, 1):interactionInf: 224	0.004	0.020	0.210	0.834
poly(St, 1):interactionInf: 398	-0.003	0.029	-0.120	0.905
poly(St, 1):interactionInf: 90	0.247	0.021	11.536	<0.001

Table S2. Full output of the final model for predicting Moment One

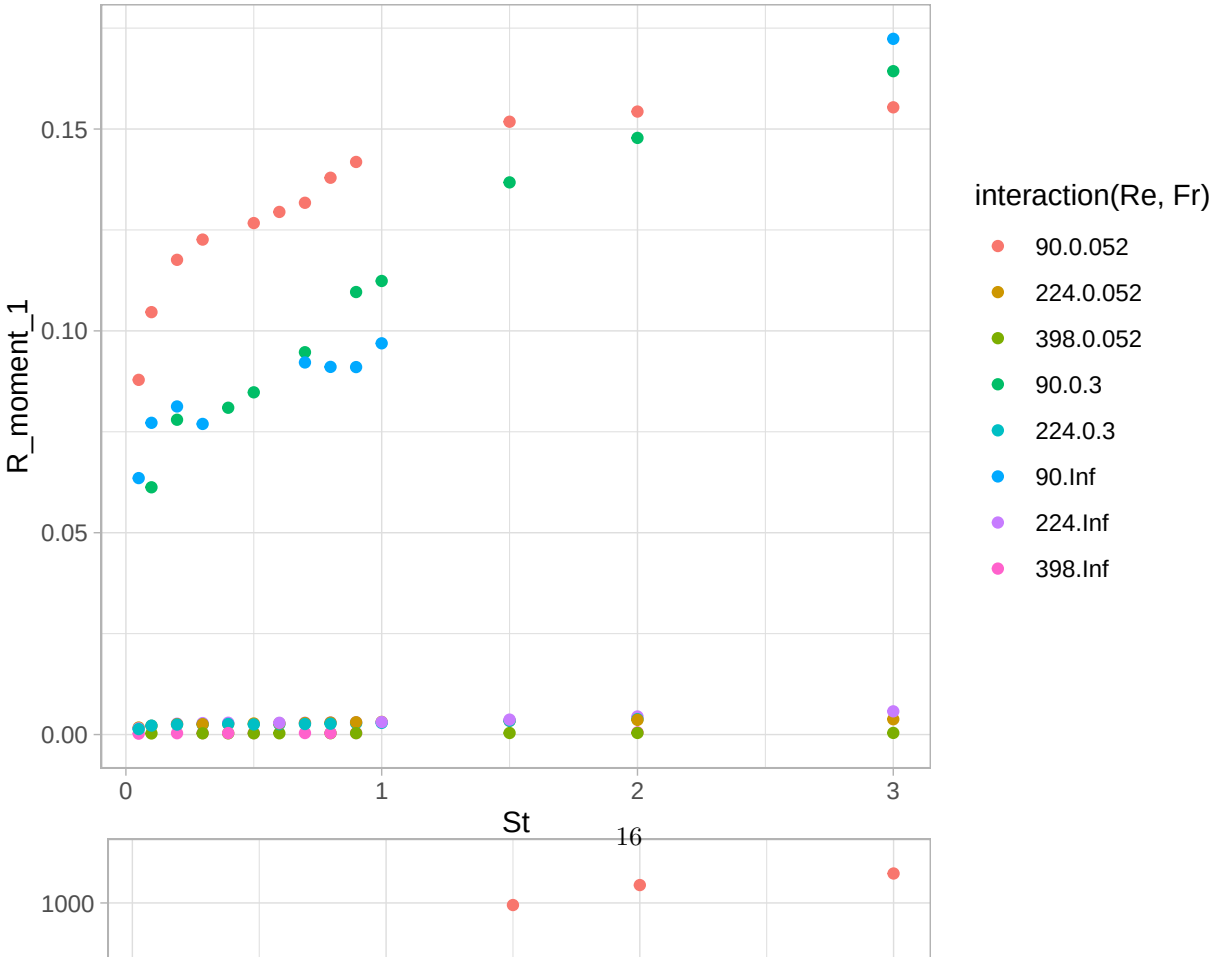
Term	β	SE	t	p value
(Intercept)	0.304	5.071	0.060	0.952
poly(St, 2)1	1.025	47.381	0.022	0.983
poly(St, 2)2	-0.249	46.966	-0.005	0.996
interaction0.052: 398	-0.300	8.028	-0.037	0.97
interaction0.052: 90	698.156	7.492	93.183	<0.001
interaction0.3: 224	-0.266	7.534	-0.035	0.972
interaction0.3: 90	0.554	7.981	0.069	0.945
interactionInf: 224	-0.256	7.500	-0.034	0.973
interactionInf: 398	-0.298	8.513	-0.035	0.972
interactionInf: 90	0.363	8.247	0.044	0.965
poly(St, 2)1:interaction0.052: 398	-1.017	81.758	-0.012	0.99
poly(St, 2)2:interaction0.052: 398	0.247	80.938	0.003	0.998
poly(St, 2)1:interaction0.052: 90	2603.304	67.525	38.553	<0.001
poly(St, 2)2:interaction0.052: 90	-792.335	67.664	-11.710	<0.001
poly(St, 2)1:interaction0.3: 224	-0.899	79.481	-0.011	0.991
poly(St, 2)2:interaction0.3: 224	0.235	81.794	0.003	0.998
poly(St, 2)1:interaction0.3: 90	5.342	77.878	0.069	0.946
poly(St, 2)2:interaction0.3: 90	1.600	80.146	0.020	0.984
poly(St, 2)1:interactionInf: 224	-0.801	70.452	-0.011	0.991
poly(St, 2)2:interactionInf: 224	0.301	68.450	0.004	0.997
poly(St, 2)1:interactionInf: 398	-1.006	89.825	-0.011	0.991
poly(St, 2)2:interactionInf: 398	0.246	86.548	0.003	0.998
poly(St, 2)1:interactionInf: 90	3.660	72.215	0.051	0.96
poly(St, 2)2:interactionInf: 90	2.465	70.111	0.035	0.972

Table S3. Full output of the final model for predicting Central Moment Two

Term	β	SE	t	p value
(Intercept)	43.840	48275.00	0.001	0.999
poly(St, 2)1	154.844	451277.02	0.000	>0.999
poly(St, 2)2	-41.039	447051.06	0.000	>0.999
interaction0.052: 398	-43.755	76246.91	-0.001	>0.999
interaction0.052: 90	5693362.973	71324.39	79.824	<0.001
interaction0.3: 224	-43.116	72158.25	-0.001	>0.999
interaction0.3: 90	-30.311	75921.38	0.000	>0.999
interactionInf: 224	-42.878	71405.69	-0.001	>0.999
interactionInf: 398	-43.706	81412.62	-0.001	>0.999
interactionInf: 90	-34.657	78554.65	0.000	>0.999
poly(St, 2)1:interaction0.052: 398	-154.697	758861.53	0.000	>0.999
poly(St, 2)2:interaction0.052: 398	40.973	760359.59	0.000	>0.999
poly(St, 2)1:interaction0.052: 90	23598512.694	643161.09	36.691	<0.001
poly(St, 2)2:interaction0.052: 90	-6676295.814	644193.32	-10.364	<0.001
poly(St, 2)1:interaction0.3: 224	-152.405	783966.20	0.000	>0.999
poly(St, 2)2:interaction0.3: 224	40.812	799650.13	0.000	>0.999
poly(St, 2)1:interaction0.3: 90	-33.168	725477.07	0.000	>0.999
poly(St, 2)2:interaction0.3: 90	70.091	746165.90	0.000	>0.999
poly(St, 2)1:interactionInf: 224	-149.857	669641.33	0.000	>0.999
poly(St, 2)2:interactionInf: 224	42.715	653634.76	0.000	>0.999
poly(St, 2)1:interactionInf: 398	-154.426	878331.44	0.000	>0.999
poly(St, 2)2:interactionInf: 398	40.944	851269.57	0.000	>0.999
poly(St, 2)1:interactionInf: 90	-68.988	689000.80	0.000	>0.999
poly(St, 2)2:interactionInf: 90	85.963	661742.70	0.000	>0.999

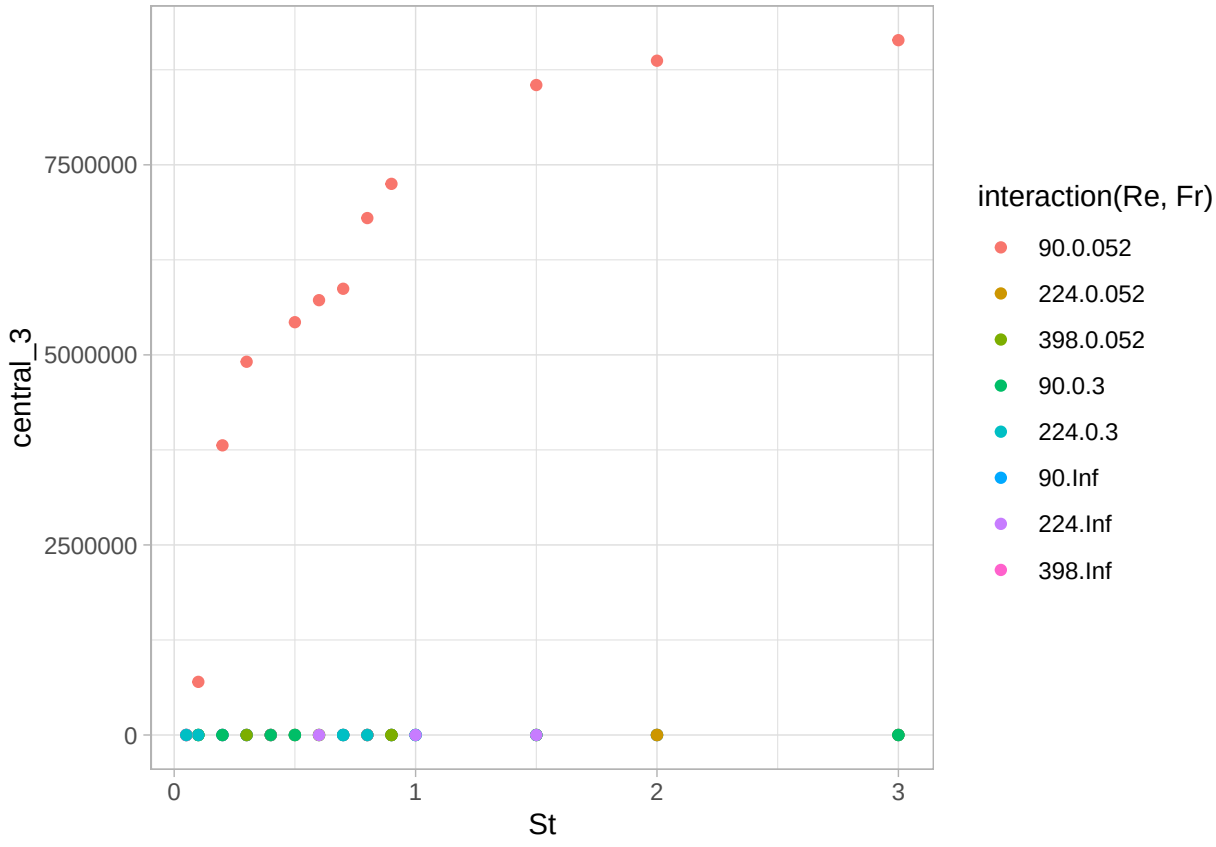
Table S4. Full output of the final model for predicting Central Moment Three

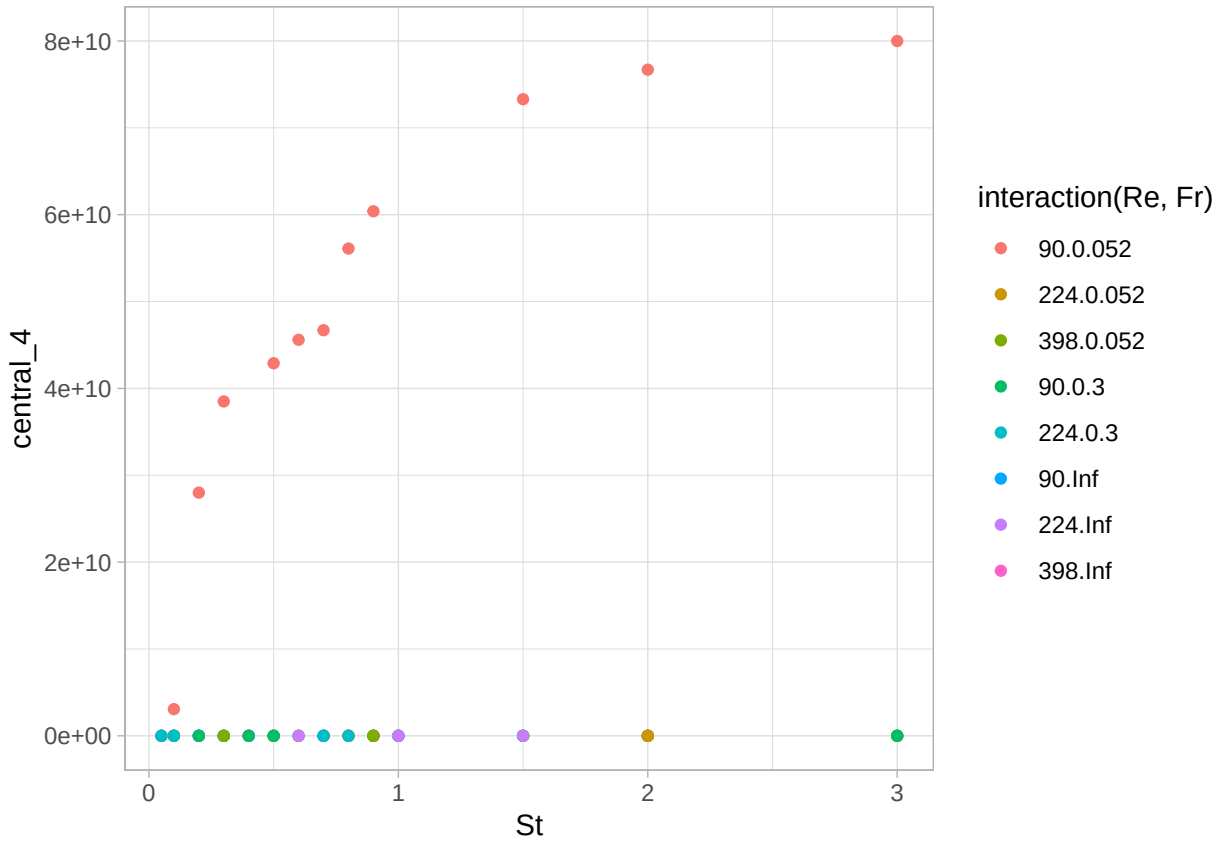
Tables



Term	β	SE	t	p value
(Intercept)	6.329674e+03	440616039	0.000	>0.999
poly(St, 2)1	2.281739e+04	4120937650	0.000	>0.999
poly(St, 2)2	-7.314889e+03	4083468150	0.000	>0.999
interaction0.052: 398	-6.328116e+03	693398850	0.000	>0.999
interaction0.052: 90	4.667430e+10	650861351	71.712	<0.001
interaction0.3: 224	-6.316081e+03	669059950	0.000	>0.999
interaction0.3: 90	-6.095794e+03	692279873	0.000	>0.999
interactionInf: 224	-6.310199e+03	651907226	0.000	>0.999
interactionInf: 398	-6.326896e+03	750617248	0.000	>0.999
interactionInf: 90	-6.188656e+03	717973151	0.000	>0.999
poly(St, 2)1:interaction0.052: 398	-2.281494e+04	6642921904	0.000	>0.999
poly(St, 2)2:interaction0.052: 398	7.313046e+03	6791496681	0.000	>0.999
poly(St, 2)1:interaction0.052: 90	2.081442e+11	5873317692	35.439	<0.001
poly(St, 2)2:interaction0.052: 90	-6.672924e+10	5891030957	-11.327	<0.001
poly(St, 2)1:interaction0.3: 224	-2.276938e+04	7806409239	0.000	>0.999
poly(St, 2)2:interaction0.3: 224	7.311320e+03	7756410915	0.000	>0.999
poly(St, 2)1:interaction0.3: 90	-2.040424e+04	6414839820	0.000	>0.999
poly(St, 2)2:interaction0.3: 90	7.752787e+03	6574014389	0.000	>0.999
poly(St, 2)1:interactionInf: 224	-2.270400e+04	6085571696	0.000	>0.999
poly(St, 2)2:interactionInf: 224	7.361339e+03	5990771690	0.000	>0.999
poly(St, 2)1:interactionInf: 398	-2.280884e+04	8562989532	0.000	>0.999
poly(St, 2)2:interactionInf: 398	7.312457e+03	8377443486	0.000	>0.999
poly(St, 2)1:interactionInf: 90	-2.118780e+04	6312036025	0.000	>0.999
poly(St, 2)2:interactionInf: 90	8.131388e+03	5982558250	0.000	>0.999

Table S5. Full output of the final model for predicting Central Moment Five




8e+10
6e+10
4e+10
2e+10
0e+00
central_4
0
1
2
3
St
interaction(Re, Fr)
90.0.052
224.0.052
398.0.052
90.0.3
224.0.3
90.Inf
224.Inf
398.Inf