Case Report

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

2021-10-13

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.

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, 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, a closer examination of the data reveals that Fr and Re only contains three levels ($Fr \in \{0.052, 0.3, \inf\}$ and $Re \in \{90, 224, 398\}$) in the data, despite their numerical nature. 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 Selection

2.3 Final Models

The forms of the statistical models, one for each moment, are:

$$R\_moment_1 \sim St * (Fr, Re)$$

$$central_2 \sim poly(St^{1/4}, 2) * (Fr, Re)$$
$$central_3 \sim poly(St^{1/3}, 2) * (Fr, Re)$$
$$central_4 \sim poly(\sqrt{St}, 2) * (Fr, Re)$$

where $Fr \in \{0.052, 0.3, \inf\}$ and $Re \in \{90, 224, 398\}$ are categorical variables and (Fr, Re) is an interaction term with 9 levels representing all combinations of Fr and Re. For the fitted coefficients, see the results section.

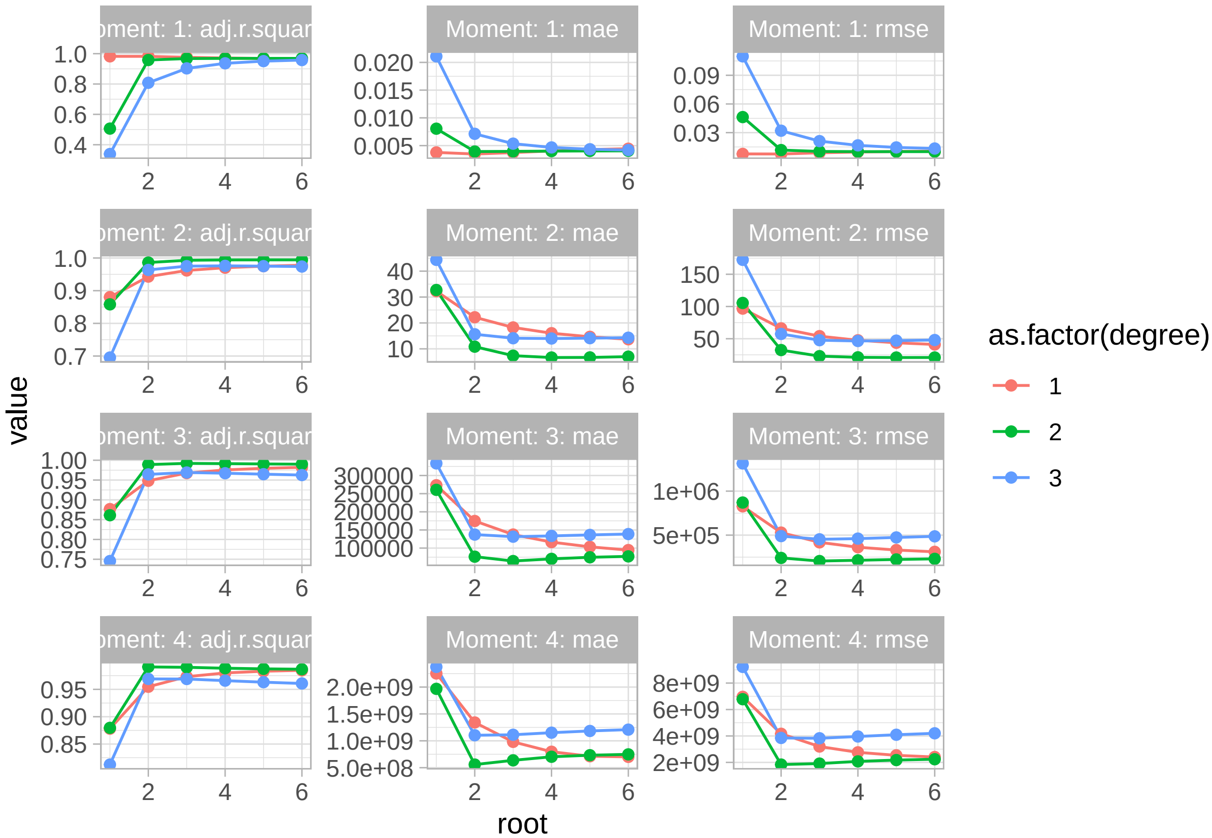
2.2 Justification

For other key properties, the three-way interaction between the predictors arises from exploratory data analysis where the relationship between St and each moment appears highly influenced by the combined levels of Re and Fr (refer to Figures S1-S4 in the Appendix). In particular, this relationship varies between roughly linear to noticeably curved, potentially quadratic or cubic, and appears to be very strong. Additionally, we notice from the curved shape of the relationship between `St` and the response that we may benefit from taking certain roots of `St` to make that more linear.

2.2 Fitting Process

We used K-fold cross validation with $K = 5$ to choose the model features, exploring linear regression and then nonlinear extensions into polynomial terms and interactions based on the EDA above. Since we only have 89 training observation, we do this in an effort to reduce the likelihood of overfitting.

Specifically, for each moment, we train candidate models to predict the moment with the general formula `~ poly(St^(1/root), degree)*(Fr, Re)`, varying the `degree` parameter from 1 to 3, varyind `root` from 1 to 6, and testing a log transformation of the response. We then look at the estimated test errors (see Figure x) and chose the parameters that give lower errors, 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. (@Jing where to mention diagnostics)



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

Model Output:

Table 1: Moment 1 Results

term	estimate	std.error	statistic	p.value
(Intercept)	0.003	0.002	1.907	0.060
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	0.000
interaction0.3: 224	0.000	0.002	-0.027	0.978
interaction0.3: 90	0.098	0.002	41.710	0.000
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	0.000
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	0.000
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	0.000
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	0.000

Table 2: Moment 2 Results

term	estimate	std.error	statistic	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	0.970
interaction0.052: 90	698.156	7.492	93.183	0.000
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	0.990
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	0.000
poly(St, 2)2:interaction0.052: 90	-792.335	67.664	-11.710	0.000
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	0.960
poly(St, 2)2:interactionInf: 90	2.465	70.111	0.035	0.972

Table 3: Moment 3 Results

term	estimate	std.error	statistic	p.value
(Intercept)	43.840	48275.00	0.001	0.999
poly(St, 2)1	154.844	451277.02	0.000	1.000
poly(St, 2)2	-41.039	447051.06	0.000	1.000
interaction0.052: 398	-43.755	76246.91	-0.001	1.000
interaction0.052: 90	5693362.973	71324.39	79.824	0.000
interaction0.3: 224	-43.116	72158.25	-0.001	1.000
interaction0.3: 90	-30.311	75921.38	0.000	1.000
interactionInf: 224	-42.878	71405.69	-0.001	1.000
interactionInf: 398	-43.706	81412.62	-0.001	1.000
interactionInf: 90	-34.657	78554.65	0.000	1.000
poly(St, 2)1:interaction0.052: 398	-154.697	758861.53	0.000	1.000
poly(St, 2)2:interaction0.052: 398	40.973	760359.59	0.000	1.000
poly(St, 2)1:interaction0.052: 90	23598512.694	643161.09	36.691	0.000
poly(St, 2)2:interaction0.052: 90	-6676295.814	644193.32	-10.364	0.000
poly(St, 2)1:interaction0.3: 224	-152.405	783966.20	0.000	1.000
poly(St, 2)2:interaction0.3: 224	40.812	799650.13	0.000	1.000
poly(St, 2)1:interaction0.3: 90	-33.168	725477.07	0.000	1.000
poly(St, 2)2:interaction0.3: 90	70.091	746165.90	0.000	1.000
poly(St, 2)1:interactionInf: 224	-149.857	669641.33	0.000	1.000
poly(St, 2)2:interactionInf: 224	42.715	653634.76	0.000	1.000
poly(St, 2)1:interactionInf: 398	-154.426	878331.44	0.000	1.000
poly(St, 2)2:interactionInf: 398	40.944	851269.57	0.000	1.000
poly(St, 2)1:interactionInf: 90	-68.988	689000.80	0.000	1.000
poly(St, 2)2:interactionInf: 90	85.963	661742.70	0.000	1.000

Table 4: Moment 4 Results

term	estimate	std.error	statistic	p.value
(Intercept)	6.329674e+03	440616039	0.000	1
poly(St, 2)1	2.281739e+04	4120937650	0.000	1
poly(St, 2)2	-7.314889e+03	4083468150	0.000	1
interaction0.052: 398	-6.328116e+03	693398850	0.000	1
interaction0.052: 90	4.667430e+10	650861351	71.712	0
interaction0.3: 224	-6.316081e+03	669059950	0.000	1
interaction0.3: 90	-6.095794e+03	692279873	0.000	1
interactionInf: 224	-6.310199e+03	651907226	0.000	1
interactionInf: 398	-6.326896e+03	750617248	0.000	1
interactionInf: 90	-6.188656e+03	717973151	0.000	1
poly(St, 2)1:interaction0.052: 398	-2.281494e+04	6642921904	0.000	1
poly(St, 2)2:interaction0.052: 398	7.313046e+03	6791496681	0.000	1
poly(St, 2)1:interaction0.052: 90	2.081442e+11	5873317692	35.439	0
poly(St, 2)2:interaction0.052: 90	-6.672924e+10	5891030957	-11.327	0
poly(St, 2)1:interaction0.3: 224	-2.276938e+04	7806409239	0.000	1
poly(St, 2)2:interaction0.3: 224	7.311320e+03	7756410915	0.000	1
poly(St, 2)1:interaction0.3: 90	-2.040424e+04	6414839820	0.000	1
poly(St, 2)2:interaction0.3: 90	7.752787e+03	6574014389	0.000	1
poly(St, 2)1:interactionInf: 224	-2.270400e+04	6085571696	0.000	1
poly(St, 2)2:interactionInf: 224	7.361339e+03	5990771690	0.000	1
poly(St, 2)1:interactionInf: 398	-2.280884e+04	8562989532	0.000	1

term	estimate	std.error	statistic	p.value
poly(St, 2)2:interactionInf: 398	7.312457e+03	8377443486	0.000	1
poly(St, 2)1:interactionInf: 90	-2.118780e+04	6312036025	0.000	1
poly(St, 2)2:interactionInf: 90	8.131388e+03	5982558250	0.000	1

Discussion

Broadly speaking, a Reynolds number > 4000 generally represents a relatively chaotic and turbulent flow, a Reynolds number < 2300 generally represents a smooth laminar flow and any number in between typically represents transient flow (Schlichting et al., 2017).

However, large Reynolds numbers is highly relevant to real life situations (atmospheric, oceanic turbulence flow).

The Reynolds number is a measure of the intensity of turbulence, with a higher Reynolds number corresponding to a higher intensity of turbulence (J. den Toonder et al., 1997). The Froude number is a measure of the impact of gravitational acceleration on fluid motion;

For instance, a cumulonimbus cloud at a high level above the ground will have a smaller Froude number (compared to lower hanging stratus clouds) because it experiences a lower intensity of gravitational acceleration relative to other clouds (Chanson, 2009). Stokes number is a description of particle properties; a large Stokes number correlates with large particle size which tends to form relatively loose clusters (Ireland et al., 2016).

but it is extremely time consuming and computationally expensive . In addition, the DNS method cannot be practically applied to simulate flows with large Reynolds numbers which requires high resolution, leading to long computation time.

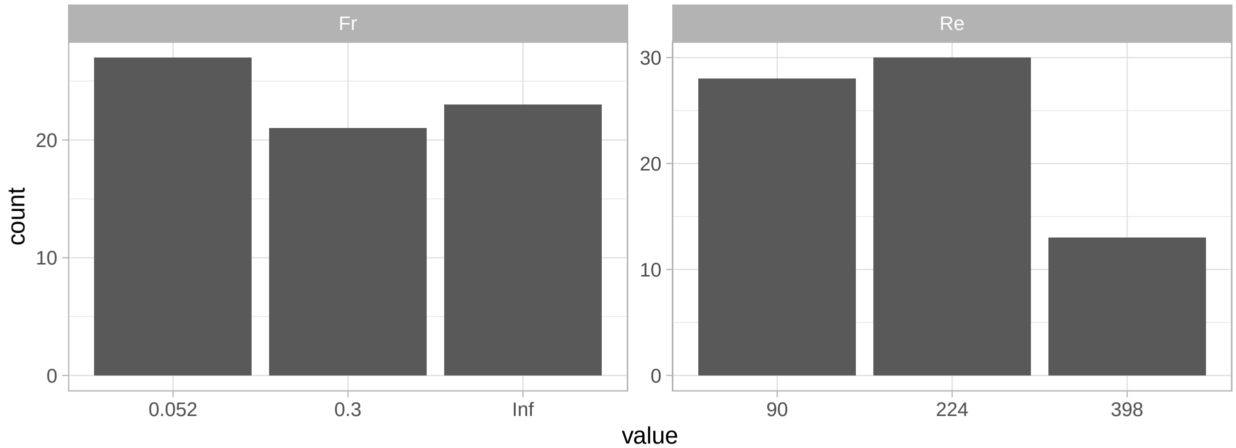
Since this probability distribution of cluster volumes is harder for statistical learning methods to work with, we will summarize this distribution by its first four raw moments $E(X)$, $E(X^2)$, $E(X^3)$, and $E(X^4)$, the latter three which we transform to central moments as response variables for inference and back for prediction. Theoretically, we are interested in the insights for each of these four moments - the mean of the distribution could be a good indicator of how flow behaves on average, the variance of the distribution could dictate how flow varies over time, the skew of the distribution could illustrate asymmetric properties of the flow, and the kurtosis of the distribution could indicate how particular cluster volumes deviate.

For Fr in particular, 0.3 represents cumulonimbus clouds and 0.052 represents cumulus clouds.

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



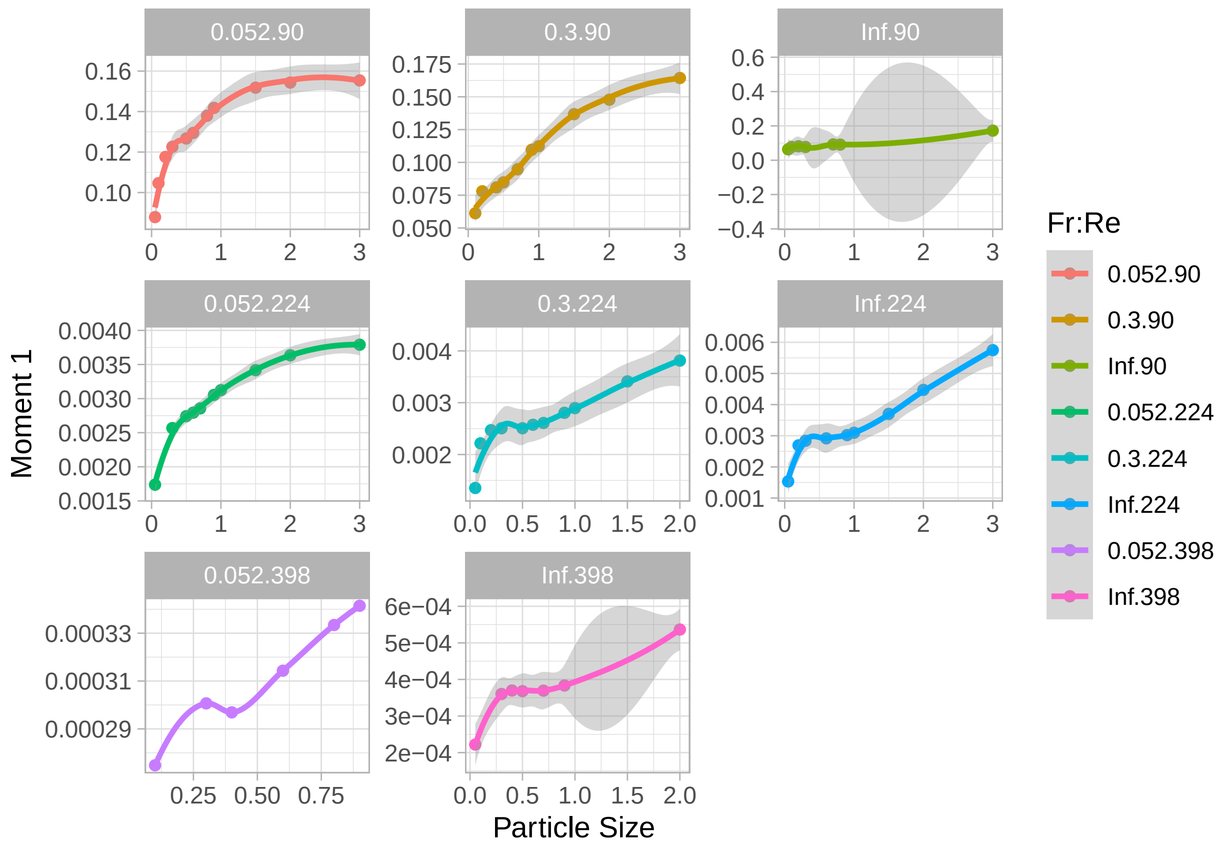


Figure 1: Moment 1 values as a function of particle size at different levels of interaction between Fr and Re.

moment	degree	root	log	name	value
1	1	2	FALSE	rmse	0.00764
1	1	2	FALSE	mae	0.00348
1	1	2	FALSE	adj.r.squared	0.98223

moment	degree	root	log	name	value
2	2	4	FALSE	mae	6.64653
2	2	5	FALSE	rmse	20.85953
2	2	5	FALSE	adj.r.squared	0.99428
3	2	3	FALSE	rmse	205,400.52488
3	2	3	FALSE	mae	64,334.51603
3	2	3	FALSE	adj.r.squared	0.99206
4	2	2	FALSE	rmse	1,837,747,421.90126
4	2	2	FALSE	mae	558,162,034.52986
4	2	2	FALSE	adj.r.squared	0.99101

```
##
## Call:
## lm(formula = R_moment_1 ~ ns(St, df = 1) * Re.numeric * Fr.numeric,
##     data = .)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.065019 -0.028210  0.005928  0.027642  0.051737
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          1.126e-01  1.450e-02   7.764 2.19e-11 ***
## ns(St, df = 1)                       1.096e-01  4.342e-02   2.525   0.0135 *
## Re.numeric                          -3.823e-04  6.211e-05  -6.155 2.72e-08 ***
## Fr.numeric                          -4.745e-02  2.396e-02  -1.980   0.0511 .
## ns(St, df = 1):Re.numeric           -2.513e-04  1.814e-04  -1.385   0.1699
## ns(St, df = 1):Fr.numeric            5.174e-02  7.948e-02   0.651   0.5169
## Re.numeric:Fr.numeric                2.070e-04  9.559e-05   2.165   0.0333 *
## ns(St, df = 1):Re.numeric:Fr.numeric -3.748e-04  3.397e-04  -1.103   0.2732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03225 on 81 degrees of freedom
## Multiple R-squared:  0.6929, Adjusted R-squared:  0.6664
## F-statistic: 26.11 on 7 and 81 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = R_moment_3 ~ ns(St, df = 1) * Re.numeric * Fr.numeric,
##     data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4261151 -1013391   -17650   469763  4738201
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          1579395    1002500   1.575  0.11905
## ns(St, df = 1)                       7342110    2429982   3.021  0.00337 **
## Re.numeric                             -5849       4369  -1.339  0.18439
## Fr.numeric                          -1641884    1629457  -1.008  0.31663
## ns(St, df = 1):Re.numeric             -20807      10392  -2.002  0.04862 *
```

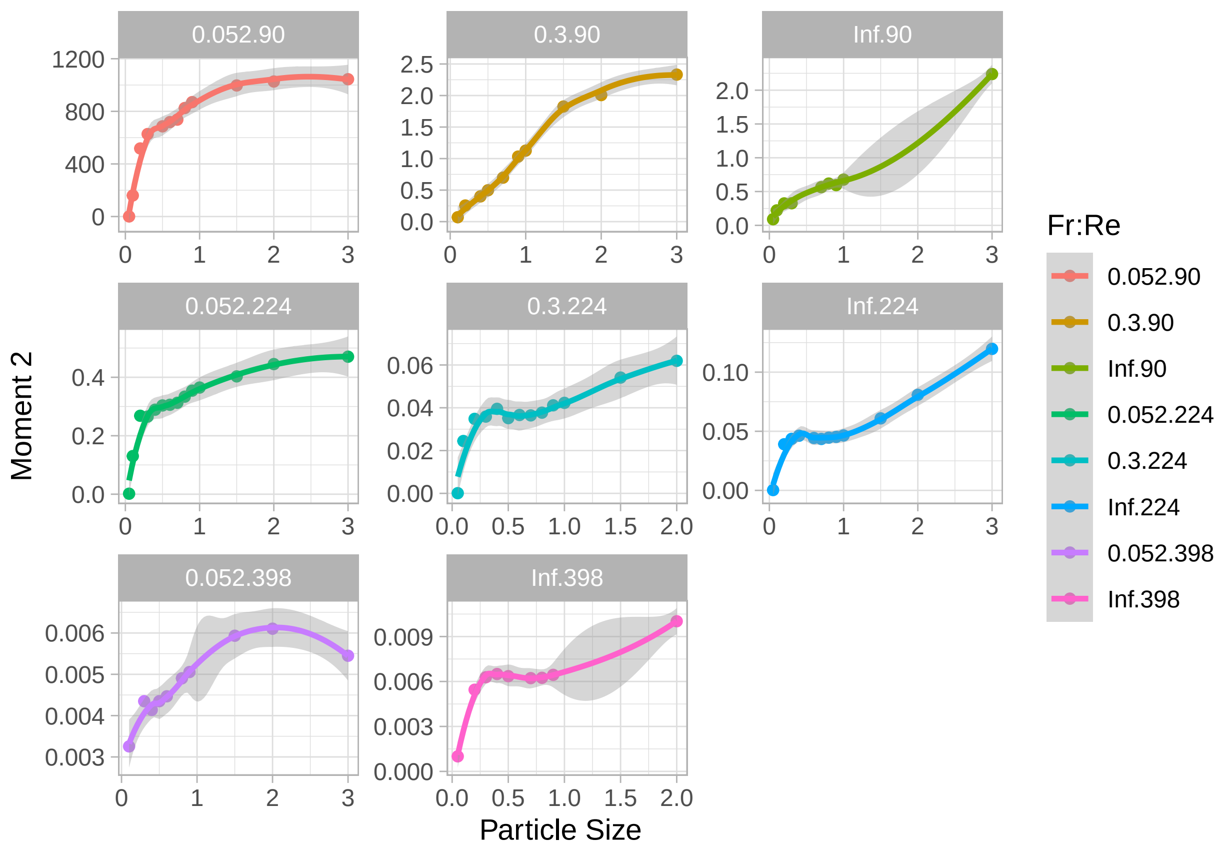


Figure 2: Moment 2 values as a function of particle size at different levels of interaction between Fr and Re.

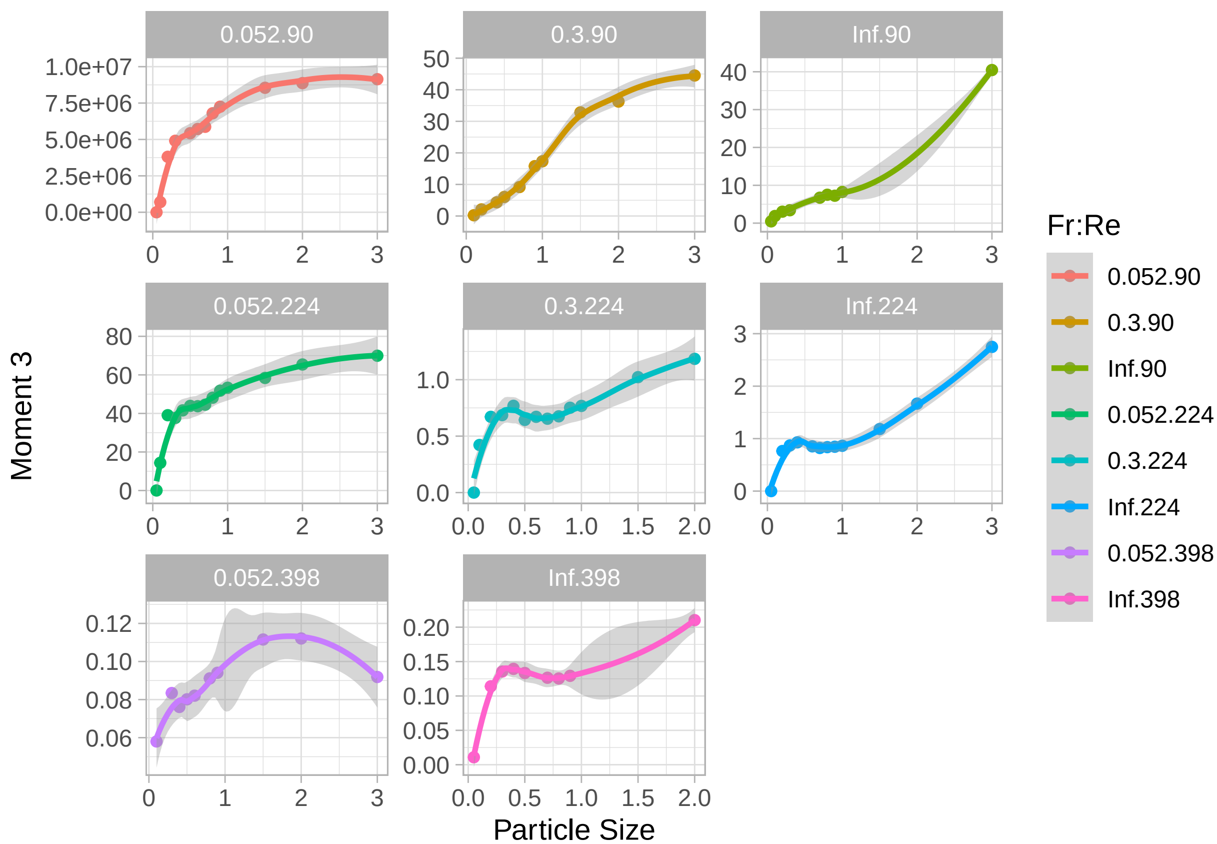


Figure 3: Moment 3 values as a function of particle size at different levels of interaction between Fr and Re.

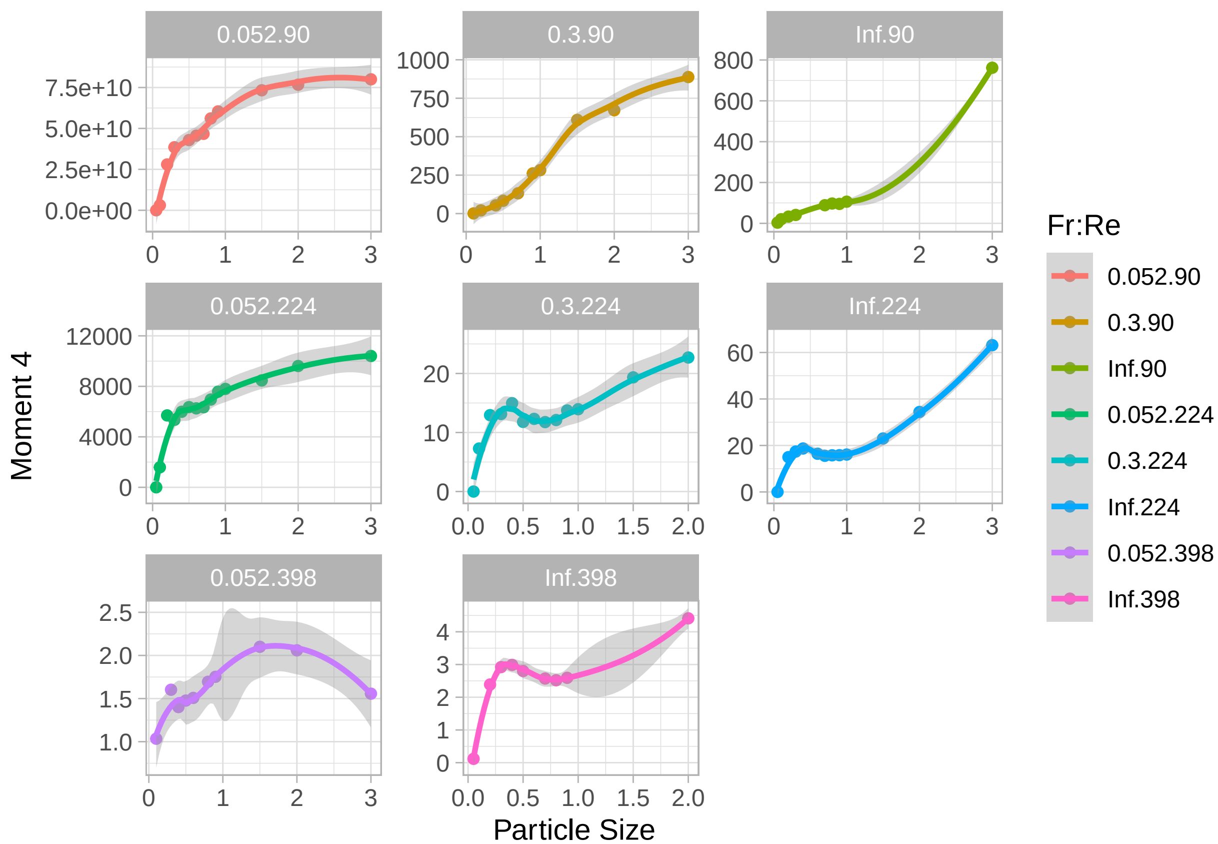


Figure 4: Moment 4 values as a function of particle size at different levels of interaction between Fr and Re.

```
## ns(St, df = 1):Fr.numeric           -8824714    4287321  -2.058  0.04277 *
## Re.numeric:Fr.numeric                   5809       6613   0.878  0.38231
## ns(St, df = 1):Re.numeric:Fr.numeric    25642      17867   1.435  0.15508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1754000 on 81 degrees of freedom
## Multiple R-squared:  0.4056, Adjusted R-squared:  0.3543
## F-statistic: 7.897 on 7 and 81 DF,  p-value: 2.825e-07
```

```
##
## Call:
## lm(formula = R_moment_4 ~ ns(St, df = 1) * Re.numeric * Fr.numeric,
##     data = .)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.696e+10 -7.454e+09 -1.524e+08  3.926e+09  4.098e+10
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          1.046e+10  8.367e+09   1.250  0.21494
## ns(St, df = 1)                       6.774e+10  2.028e+10   3.340  0.00127 **
## Re.numeric                          -3.992e+07  3.646e+07  -1.095  0.27687
## Fr.numeric                          -1.074e+10  1.360e+10  -0.789  0.43218
## ns(St, df = 1):Re.numeric           -1.943e+08  8.673e+07  -2.241  0.02779 *
## ns(St, df = 1):Fr.numeric           -8.077e+10  3.578e+10  -2.257  0.02668 *
## Re.numeric:Fr.numeric                3.895e+07  5.519e+07   0.706  0.48242
## ns(St, df = 1):Re.numeric:Fr.numeric  2.365e+08  1.491e+08   1.586  0.11658
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.464e+10 on 81 degrees of freedom
## Multiple R-squared:  0.4097, Adjusted R-squared:  0.3587
## F-statistic: 8.031 on 7 and 81 DF,  p-value: 2.19e-07
```