Case Report

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

2021-10-12

1 Introduction

Turbulence is a fundamental concept in fluid mechanics. Understanding how fluids move has enormous importance to a vast range of problems, with practical applications in aeronautical engineering, environmental science, astronomy, and more. In contrast to laminar flow, turbulent flow is irregular, unpredictable, and energy-dissipating. Turbulence enhances mixing which leads to non-uniform distribution of particles and cluster formation. In general, the final state of particles subject to turbulent flow is determined by three main parameters: Reynolds (Re) number, Froude (Fr) number and Stokes (St) number. 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). 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). 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).

Despite the importance of turbulence and particle clustering, we do not understand this process well very. Researchers has primarily used Direct Numerical Simulation (DNS) of Navier-Stokes equations to study this problem, but it is extremely time consuming and computationally expensive (Moin & Mahesh, 1998). 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. However, large Reynolds numbers is highly relevant to real life situations (atmospheric, oceanic turbulence flow). Thus, the primary research **objective** of this study is to build a statistical model to investigate how Reynolds (Re) number, Froude (Fr) number and Stokes (St) number influence particle cluster volume distribution and allows prediction without the need for simulation.

The data for this study comes from Direct Numerical Simulation using a range of parameters. Then, Voronoi Tessellation, a technique that looks at general features of individual clusters in the underlying turbulence, has been applied to generate a distribution of cluster volumes. 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. We hope that our research model will enable a quick and efficient prediction of a particle cluster volume distribution and enhance our understanding of the relative influence of these parameters in turbulent flow.

2 Methods

2.1 Model

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

While Fr and Re are continuous values in physics, both variables contain only three levels in our training and test data (Appendix). Since we see from the test data that we are specifically interested in these levels, we treat both as categorical variables in modeling. For Fr in particular, 0.3 represents cumulonimbus clouds and 0.052 represents cumulus clouds. See 2.4 for a separate model set that addresses these limitations.

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)
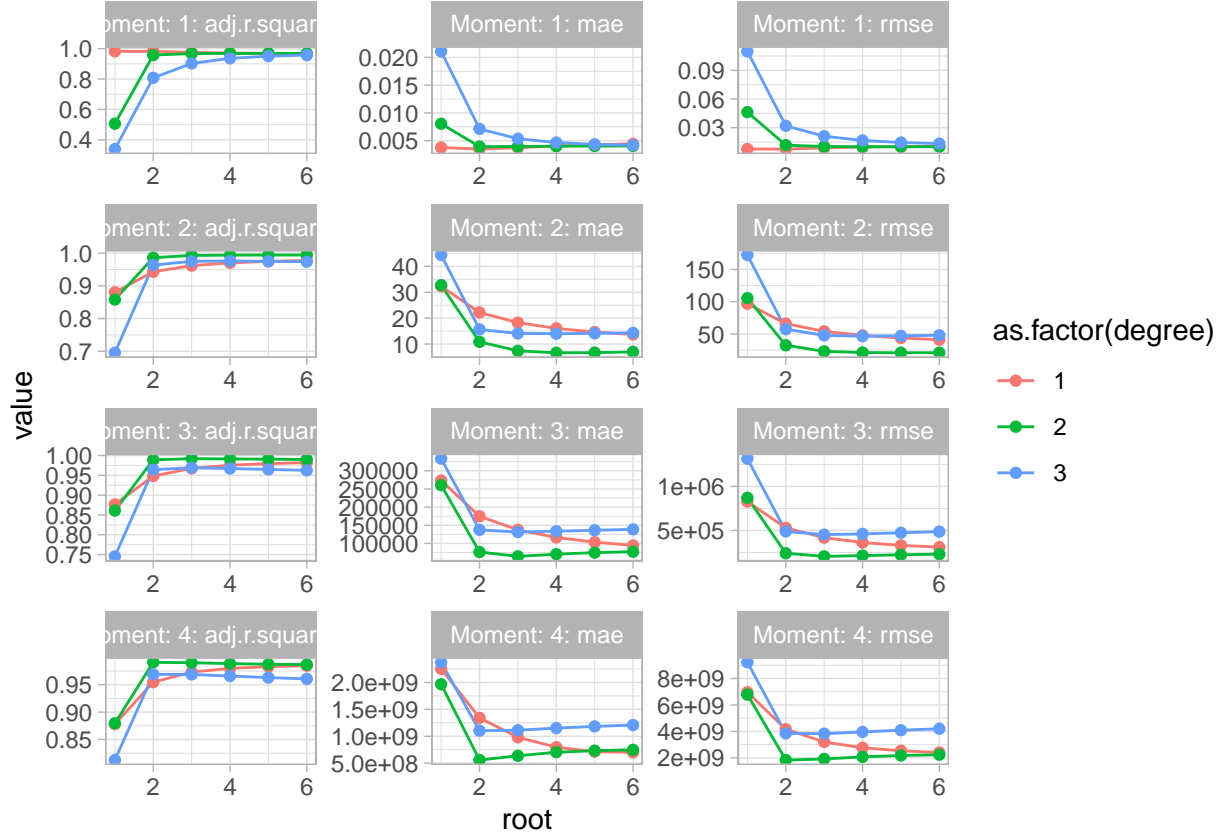
2.4 Model Extension

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:

```
##
## Call:
## lm(formula = central_1 ~ poly(St, 1) * interaction, data = .)
```

3

```
##
## Residuals:
##        Min          1Q     Median         3Q        Max
## -0.0262390 -0.0000713  0.0000154  0.0002443  0.0129019
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      2.863e-03  1.501e-03   1.907   0.0604 .
## poly(St, 1)                      4.396e-03  1.402e-02   0.314   0.7547
## interaction0.052: 398           -2.535e-03  2.346e-03  -1.080   0.2835
## interaction0.052: 90             1.268e-01  2.210e-03  57.356  < 2e-16 ***
## interaction0.3: 224             -6.075e-05  2.214e-03  -0.027   0.9782
## interaction0.3: 90               9.809e-02  2.352e-03  41.710  < 2e-16 ***
## interactionInf: 224              2.620e-04  2.217e-03   0.118   0.9063
## interactionInf: 398             -2.468e-03  2.505e-03  -0.985   0.3278
## interactionInf: 90               9.335e-02  2.405e-03  38.808  < 2e-16 ***
## poly(St, 1):interaction0.052: 398 -4.053e-03  2.072e-02  -0.196   0.8454
## poly(St, 1):interaction0.052: 90   1.369e-01  1.996e-02   6.860 1.89e-09 ***
## poly(St, 1):interaction0.3: 224    2.437e-03  2.545e-02   0.096   0.9240
## poly(St, 1):interaction0.3: 90     2.595e-01  2.069e-02  12.542  < 2e-16 ***
## poly(St, 1):interactionInf: 224    4.275e-03  2.032e-02   0.210   0.8340
## poly(St, 1):interactionInf: 398   -3.469e-03  2.892e-02  -0.120   0.9049
## poly(St, 1):interactionInf: 90     2.469e-01  2.140e-02  11.536  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.005616 on 73 degrees of freedom
## Multiple R-squared:  0.9916, Adjusted R-squared:  0.9899
## F-statistic: 575.2 on 15 and 73 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = central_2 ~ poly(St, 2) * interaction, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.577  -0.010   0.000   0.011  89.044
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.3044     5.0710   0.060    0.952
## poly(St, 2)1                        1.0255    47.3808   0.022    0.983
## poly(St, 2)2                       -0.2487    46.9662  -0.005    0.996
## interaction0.052: 398              -0.2998     8.0281  -0.037    0.970
## interaction0.052: 90              698.1556     7.4923  93.183   <2e-16 ***
## interaction0.3: 224                -0.2656     7.5338  -0.035    0.972
## interaction0.3: 90                  0.5539     7.9809   0.069    0.945
## interactionInf: 224                -0.2564     7.5000  -0.034    0.973
## interactionInf: 398                -0.2979     8.5133  -0.035    0.972
## interactionInf: 90                  0.3634     8.2468   0.044    0.965
## poly(St, 2)1:interaction0.052: 398  -1.0172    81.7578  -0.012    0.990
## poly(St, 2)2:interaction0.052: 398   0.2468    80.9380   0.003    0.998
## poly(St, 2)1:interaction0.052: 90 2603.3041    67.5249  38.553   <2e-16 ***
## poly(St, 2)2:interaction0.052: 90 -792.3349    67.6642 -11.710   <2e-16 ***
```

```
## poly(St, 2)1:interaction0.3: 224      -0.8988    79.4812  -0.011    0.991
## poly(St, 2)2:interaction0.3: 224       0.2351    81.7941   0.003    0.998
## poly(St, 2)1:interaction0.3: 90        5.3420    77.8781   0.069    0.946
## poly(St, 2)2:interaction0.3: 90        1.6001    80.1465   0.020    0.984
## poly(St, 2)1:interactionInf: 224      -0.8010    70.4519  -0.011    0.991
## poly(St, 2)2:interactionInf: 224       0.3012    68.4500   0.004    0.997
## poly(St, 2)1:interactionInf: 398      -1.0055    89.8253  -0.011    0.991
## poly(St, 2)2:interactionInf: 398       0.2456    86.5481   0.003    0.998
## poly(St, 2)1:interactionInf: 90        3.6604    72.2151   0.051    0.960
## poly(St, 2)2:interactionInf: 90        2.4652    70.1108   0.035    0.972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.97 on 65 degrees of freedom
## Multiple R-squared:  0.9961, Adjusted R-squared:  0.9947
## F-statistic: 725.9 on 23 and 65 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = central_3 ~ poly(St, 2) * interaction, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -720016       0       0       0  712324
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    4.384e+01  4.828e+04   0.001    0.999
## poly(St, 2)1                   1.548e+02  4.513e+05   0.000    1.000
## poly(St, 2)2                  -4.104e+01  4.471e+05   0.000    1.000
## interaction0.052: 398         -4.376e+01  7.625e+04  -0.001    1.000
## interaction0.052: 90           5.693e+06  7.132e+04  79.824  < 2e-16 ***
## interaction0.3: 224           -4.312e+01  7.216e+04  -0.001    1.000
## interaction0.3: 90            -3.031e+01  7.592e+04   0.000    1.000
## interactionInf: 224           -4.288e+01  7.141e+04  -0.001    1.000
## interactionInf: 398           -4.371e+01  8.141e+04  -0.001    1.000
## interactionInf: 90            -3.466e+01  7.855e+04   0.000    1.000
## poly(St, 2)1:interaction0.052: 398 -1.547e+02  7.589e+05   0.000    1.000
## poly(St, 2)2:interaction0.052: 398  4.097e+01  7.604e+05   0.000    1.000
## poly(St, 2)1:interaction0.052: 90   2.360e+07  6.432e+05  36.691  < 2e-16 ***
## poly(St, 2)2:interaction0.052: 90  -6.676e+06  6.442e+05 -10.364 2.11e-15 ***
## poly(St, 2)1:interaction0.3: 224   -1.524e+02  7.840e+05   0.000    1.000
## poly(St, 2)2:interaction0.3: 224    4.081e+01  7.997e+05   0.000    1.000
## poly(St, 2)1:interaction0.3: 90    -3.317e+01  7.255e+05   0.000    1.000
## poly(St, 2)2:interaction0.3: 90     7.009e+01  7.462e+05   0.000    1.000
## poly(St, 2)1:interactionInf: 224   -1.499e+02  6.696e+05   0.000    1.000
## poly(St, 2)2:interactionInf: 224    4.271e+01  6.536e+05   0.000    1.000
## poly(St, 2)1:interactionInf: 398   -1.544e+02  8.783e+05   0.000    1.000
## poly(St, 2)2:interactionInf: 398    4.094e+01  8.513e+05   0.000    1.000
## poly(St, 2)1:interactionInf: 90    -6.899e+01  6.890e+05   0.000    1.000
## poly(St, 2)2:interactionInf: 90     8.596e+01  6.617e+05   0.000    1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 180600 on 65 degrees of freedom
## Multiple R-squared:  0.9949, Adjusted R-squared:  0.9932
## F-statistic: 556.1 on 23 and 65 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = central_4 ~ poly(St, 2) * interaction, data = .)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -6.532e+09 -1.000e+01  0.000e+00  5.000e+00  7.352e+09
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      6.330e+03  4.406e+08    0.00        1
## poly(St, 2)1                     2.282e+04  4.121e+09    0.00        1
## poly(St, 2)2                    -7.315e+03  4.083e+09    0.00        1
## interaction0.052: 398           -6.328e+03  6.934e+08    0.00        1
## interaction0.052: 90             4.667e+10  6.509e+08   71.71   <2e-16 ***
## interaction0.3: 224             -6.316e+03  6.691e+08    0.00        1
## interaction0.3: 90              -6.096e+03  6.923e+08    0.00        1
## interactionInf: 224             -6.310e+03  6.519e+08    0.00        1
## interactionInf: 398             -6.327e+03  7.506e+08    0.00        1
## interactionInf: 90              -6.189e+03  7.180e+08    0.00        1
## poly(St, 2)1:interaction0.052: 398 -2.281e+04  6.643e+09    0.00        1
## poly(St, 2)2:interaction0.052: 398  7.313e+03  6.791e+09    0.00        1
## poly(St, 2)1:interaction0.052: 90   2.081e+11  5.873e+09   35.44   <2e-16 ***
## poly(St, 2)2:interaction0.052: 90  -6.673e+10  5.891e+09  -11.33   <2e-16 ***
## poly(St, 2)1:interaction0.3: 224   -2.277e+04  7.806e+09    0.00        1
## poly(St, 2)2:interaction0.3: 224    7.311e+03  7.756e+09    0.00        1
## poly(St, 2)1:interaction0.3: 90    -2.040e+04  6.415e+09    0.00        1
## poly(St, 2)2:interaction0.3: 90     7.753e+03  6.574e+09    0.00        1
## poly(St, 2)1:interactionInf: 224   -2.270e+04  6.086e+09    0.00        1
## poly(St, 2)2:interactionInf: 224    7.361e+03  5.991e+09    0.00        1
## poly(St, 2)1:interactionInf: 398   -2.281e+04  8.563e+09    0.00        1
## poly(St, 2)2:interactionInf: 398    7.312e+03  8.377e+09    0.00        1
## poly(St, 2)1:interactionInf: 90    -2.119e+04  6.312e+09    0.00        1
## poly(St, 2)2:interactionInf: 90     8.131e+03  5.983e+09    0.00        1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.648e+09 on 65 degrees of freedom
## Multiple R-squared:  0.994,  Adjusted R-squared:  0.9919
## F-statistic: 467.7 on 23 and 65 DF,  p-value: < 2.2e-16
```

Discussion

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
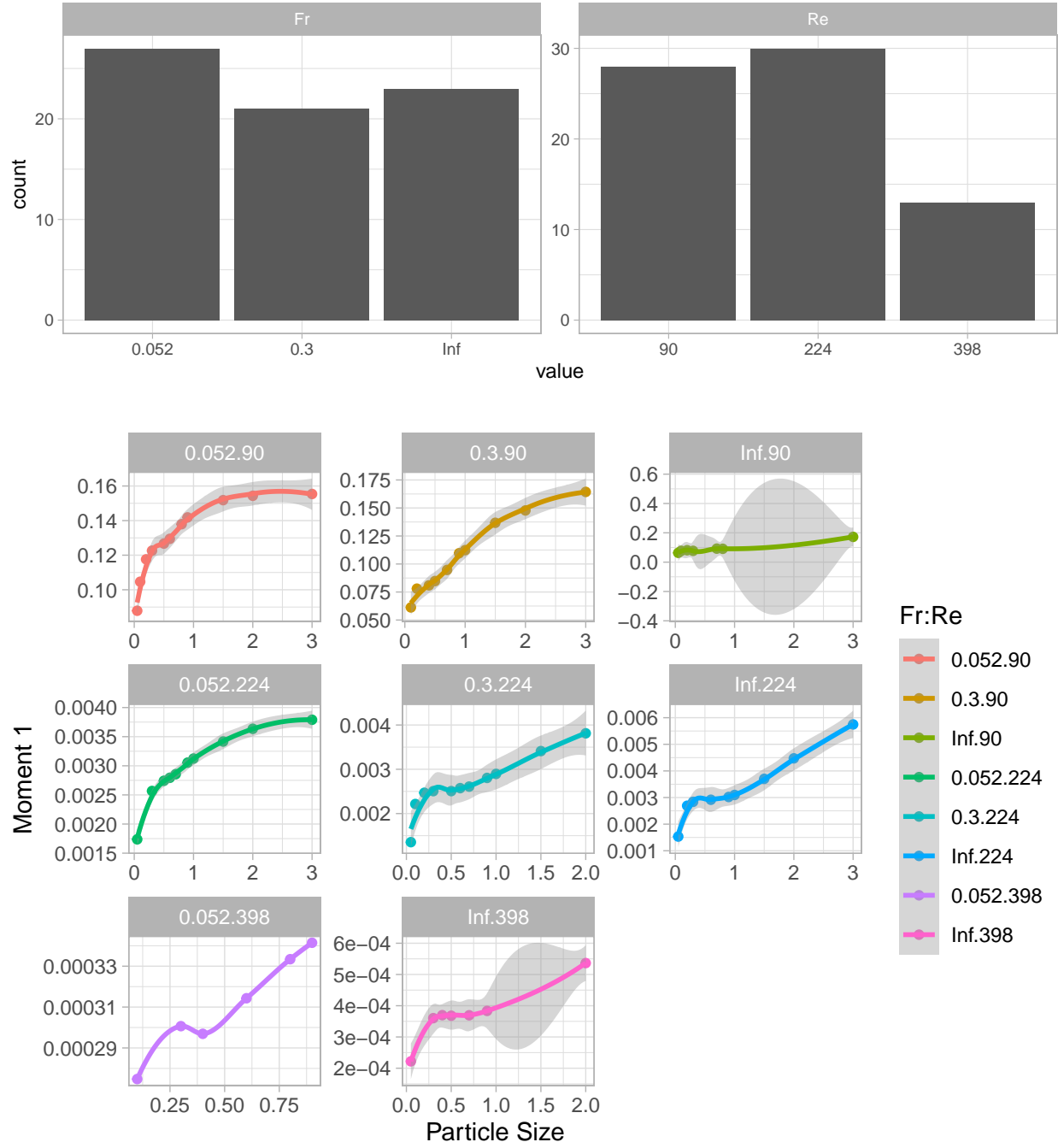
Appendix

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 *
```
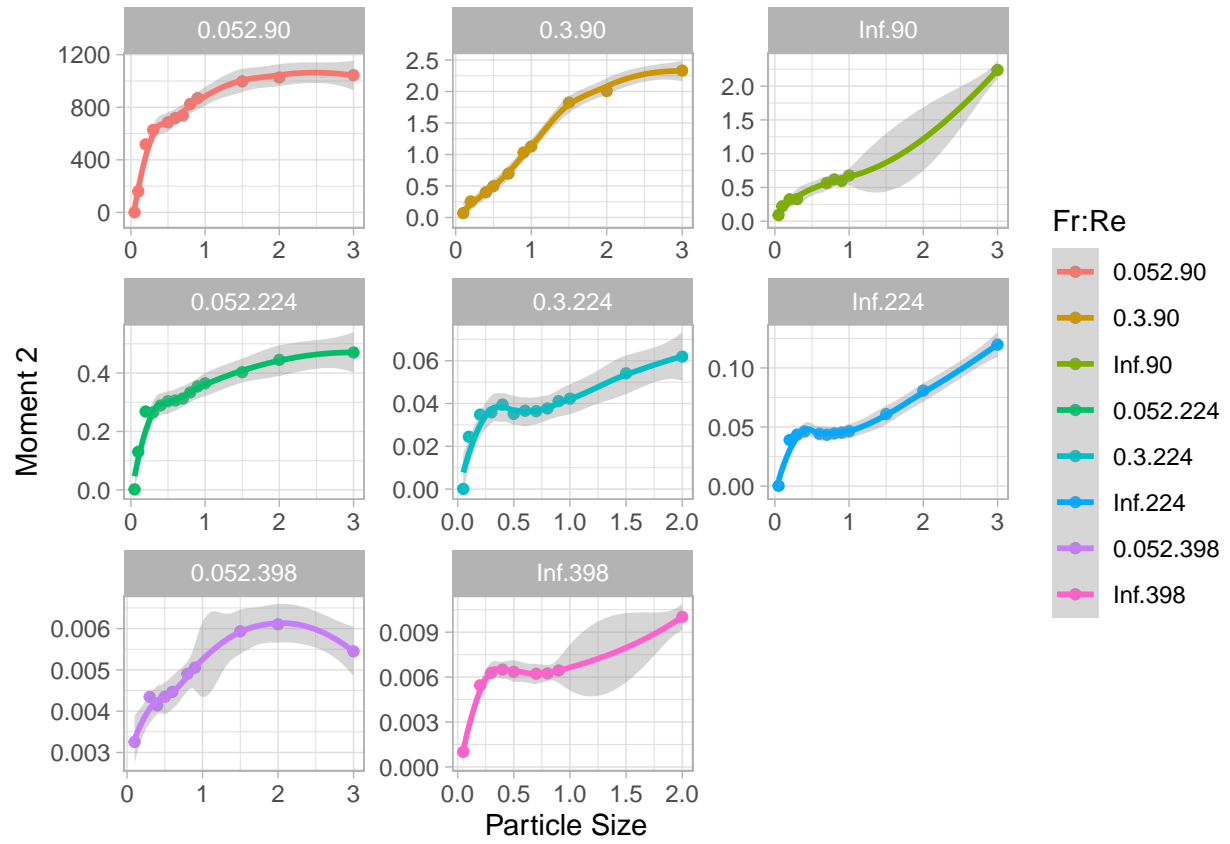
Figure 2: Moment 2 values as a function of particle size at different levels of interaction between Fr and Re.
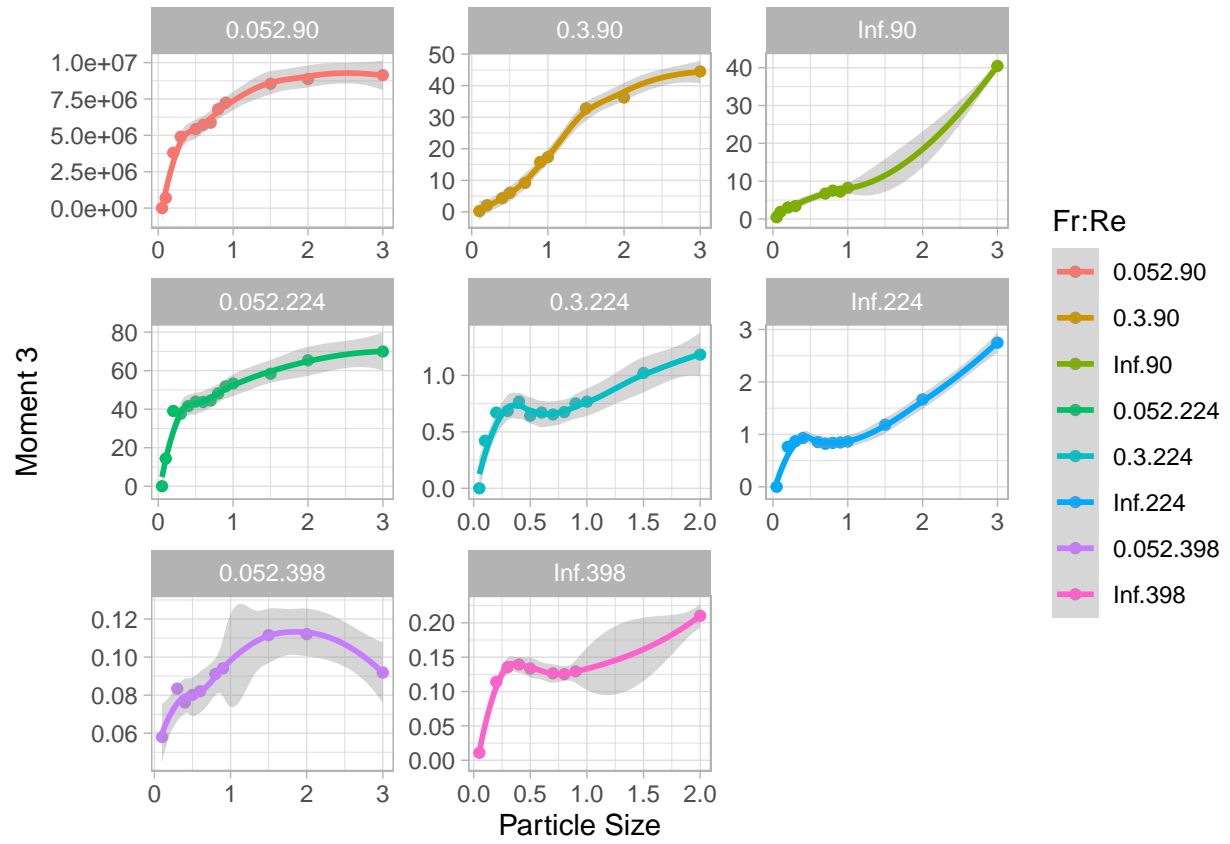
Figure 3: Moment 3 values as a function of particle size at different levels of interaction between Fr and Re.
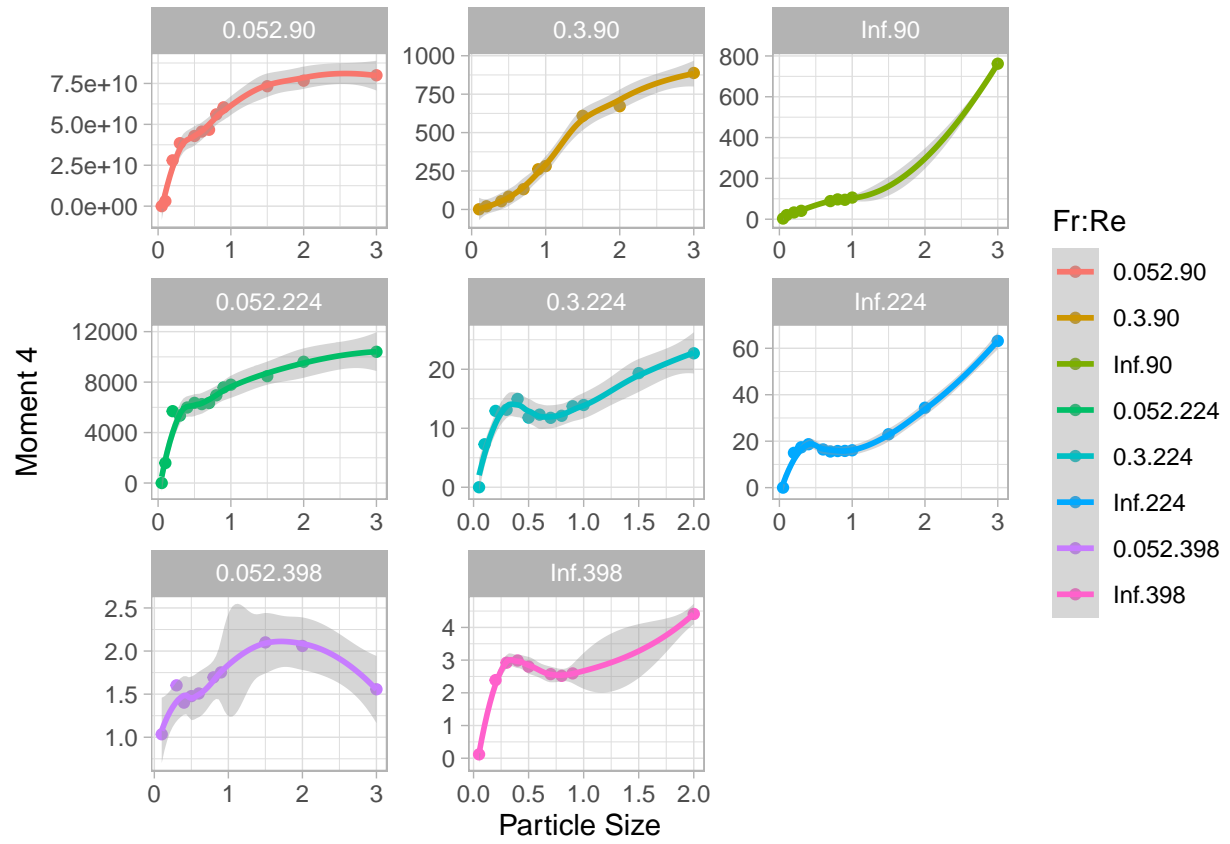
Figure 4: Moment 4 values as a function of particle size at different levels of interaction between Fr and Re.

```
## 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
```