Turbulence Analysis

Tingnan Hu, Peter Liu, Islina Shan, Ken Ye, Nancy Zhang

2023-10-28

Introduction

Turbulence is one of the fascinating topics in the research in fluid dynamics. It is characterized by its chaotic motion, rapid fluctuations and lack of predictable patterns. Yet, there have been numerous attempts in scientific literature trying to model the behavior of turbulent flows, as turbulent flows are prevalent in our world and are the underlying forces that drive plenty of the physical processes, from wisps of smoking swirling up from the cigarette to mixing of chemicals in industrial processes. A better understanding and prediction of turbulent flow will help us gain a deeper insight into a wide range of applications, such as improved aerodynamics in airplane designs and better climatic modelling.

A subdomain in turbulent flow research deals with particle clustering in turbulent flow focusing on small particles" behavior in turbulent fluids. For our project, we are provided with a set of simulation results on small particle probability distribution. The outcome variable was originally a probability distribution for particle cluster volumes, but it was converted into its first four raw moments, $E[X]$ to $E[X^4]$, to facilitate analysis. The predictor set contains three variables:

- Reynolds number, `Re`, which provides information on the type of flow a fluid is experiencing. A low `Re` corresponds with laminar flow (smooth and orderly), while a high `Re` corresponds with turbulent flow.
- Gravitational acceleration, `Fr`, which measures the gravitational forces particles are experiencing.
- Stokes number, `St`, where larger value corresponds with larger particle size.

The main research objective of our project will be to build a viable statistical model to predict the response variable (first four raw moments of particle probability distribution) using the three predictors at hand and the provided training set. Specifically, we are interested in the following:

- Does there exist a significant linear relationship between the predictors and the raw four moments?
- Is there any significant interaction effects between predictors on the response variables?
- Does a linear regression model suffice? Do we need a more complex model to better explain the relationship between the predictors and responses?
- Do the identified effects of the predictors vary for the four moments?

Ultimately, we aim for our model to capture adequate trends in our training data, so that for a new parameter setting of (`Re`, `Fr`, `St`), we can accurately predict its particle cluster volume distribution in terms of its four raw moments, as well as make inference on how each parameter affects the probability distribution for particle cluster volumes.

Methodology

First, we examine the predictor and response variables and perform adequate transformations. For predictor variables, we first noticed that `Fr` only takes on 0.052, 0.3, and Inf in both our training and testing data set, and directly using these values as they are is not viable as they contain infinity. Therefore, we create a new categorical variable called `gravity` using the following categorization:

Fr	Gravity
$\mathbf{Fr < 0.1}$	low gravity
$\mathbf{0.1 < Fr < 1}$	moderate gravity
$\mathbf{Fr > 1}$	high gravity

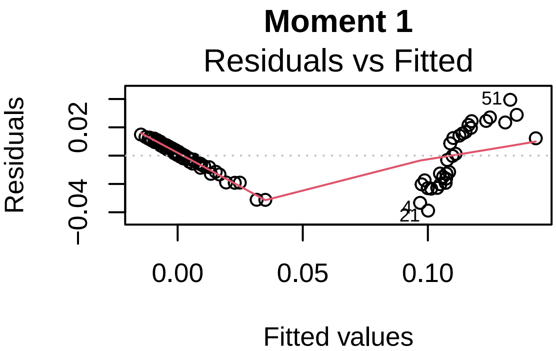
We also noticed that the predictor variable `Re` only takes on 90, 224, and 398 in both our training and testing data set. We thus create a new categorical variable called `flow` using the following categorization:

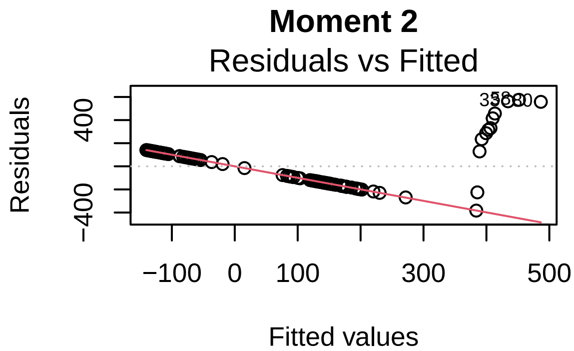
Re	Flow
$\mathbf{Re < 100}$	low flow
$\mathbf{100 < Re < 300}$	moderate flow
$\mathbf{Re > 300}$	high flow

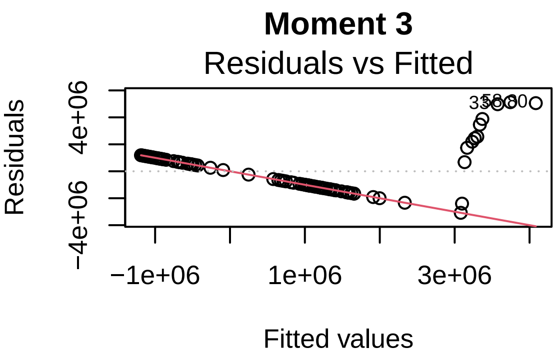
Simple Linear Regression

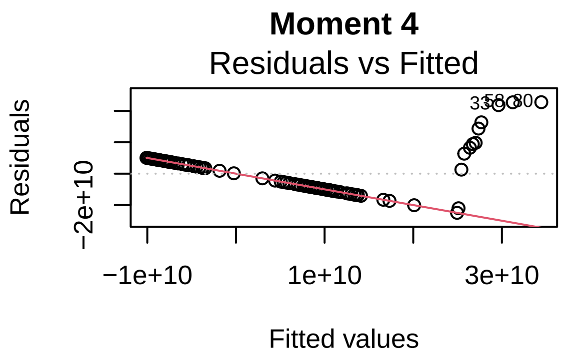
We begin with simple linear regression, yielding adjusted R-squared values of 0.9251, 0.4175, 0.4023, and 0.3906 for moments 1 to 4. In all four models, the p-values are close to zero, indicating significant linear relationships between the predictors and the raw moments.

Explore Response Variable Transformation







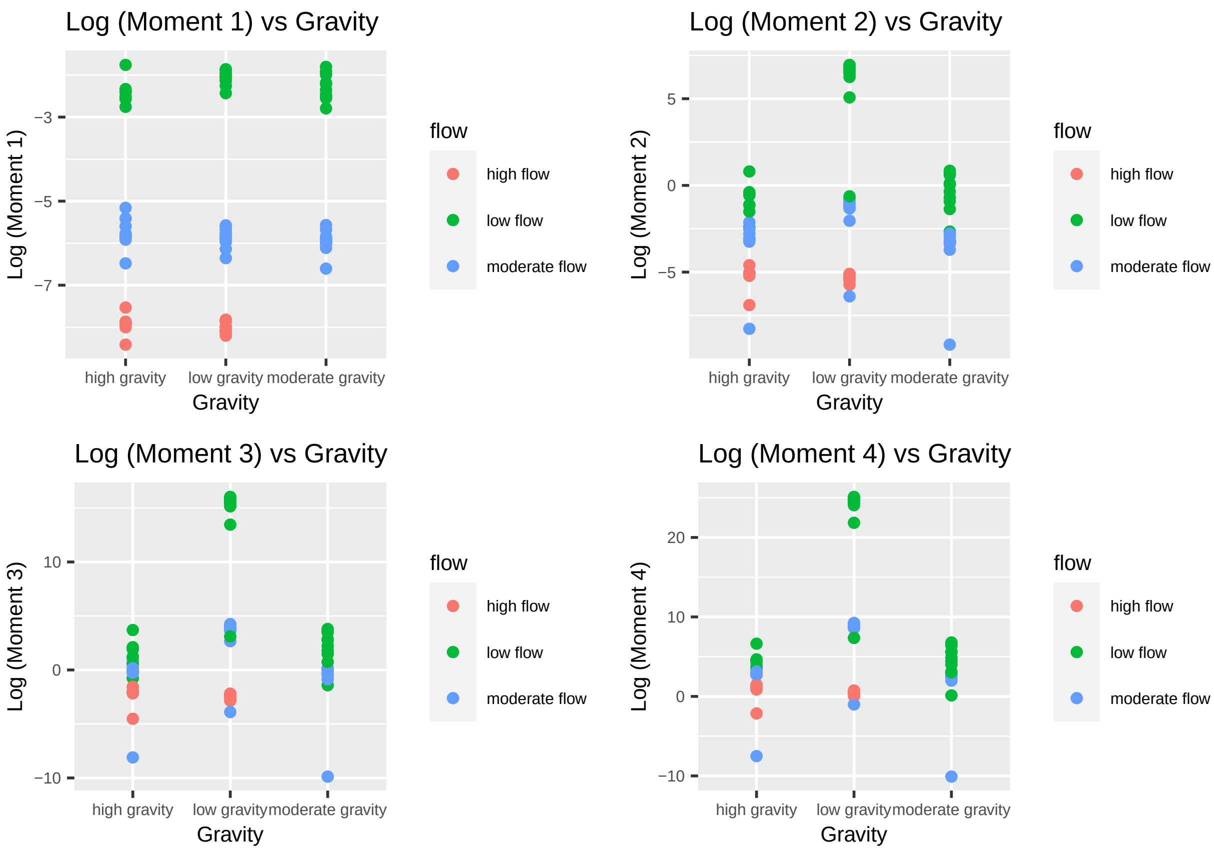


Looking at the Residuals vs Fitted plots, we observe clear patterns, which suggests that the linearity assumption is violated.

However, log-transforming the response variables not only results in more favorable Residuals vs. Fitted plots but also leads to improved adjusted R-squared values of 0.9949, 0.7633, 0.6802, and 0.6518 for moments 1 to 4.

Explore Interaction Effect

We then investigate the possible interaction effects between the predictor variables. The plot below suggests a possible interaction effect between `gravity` and `flow`.



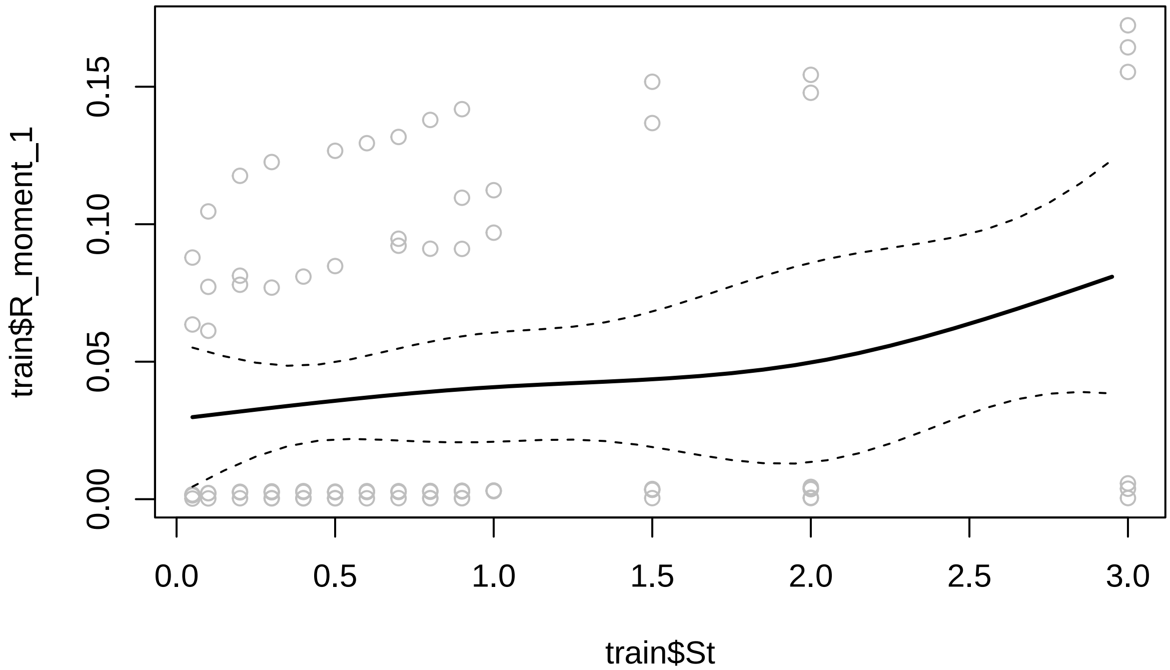
The plot reveals that fitting a linear regression line between `gravity` and the response variables for low, moderate, and high `flow` would result in different slopes. These varying slopes serve as indicators of an interaction between these two variables. This interpretation is further validated by incorporating the interaction term into our simple linear regression models, which, in turn, yields significant p-values for the interaction terms.

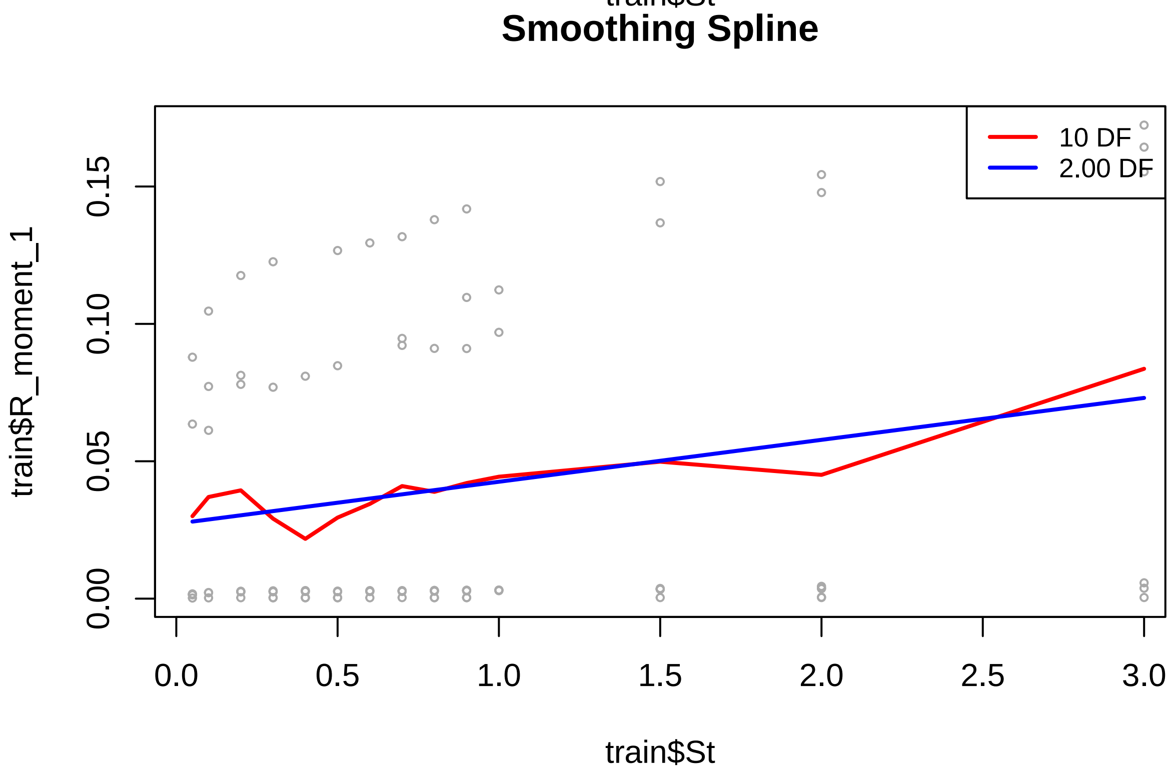
Incorporating this insight into our simple linear regression models by including the interaction term `gravity:flow` results in adjusted R-squared values of 0.9966, 0.8909, 0.8770, and 0.8809 for moments 1 through 4.

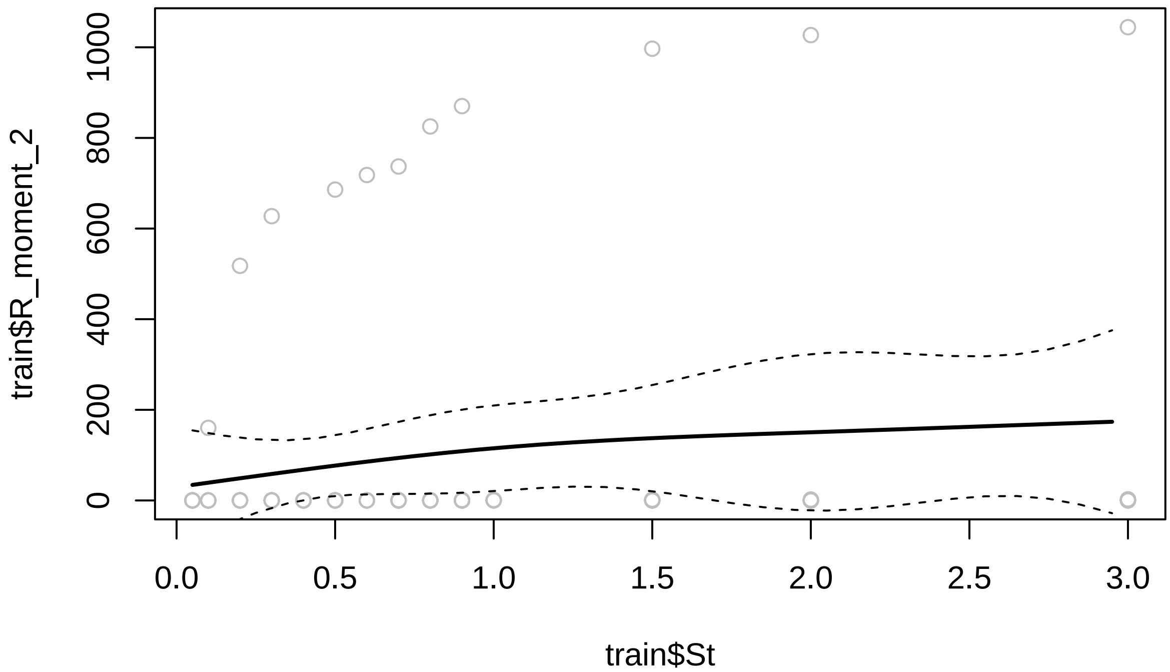
Explore Polynomial Term

Ridge Regression

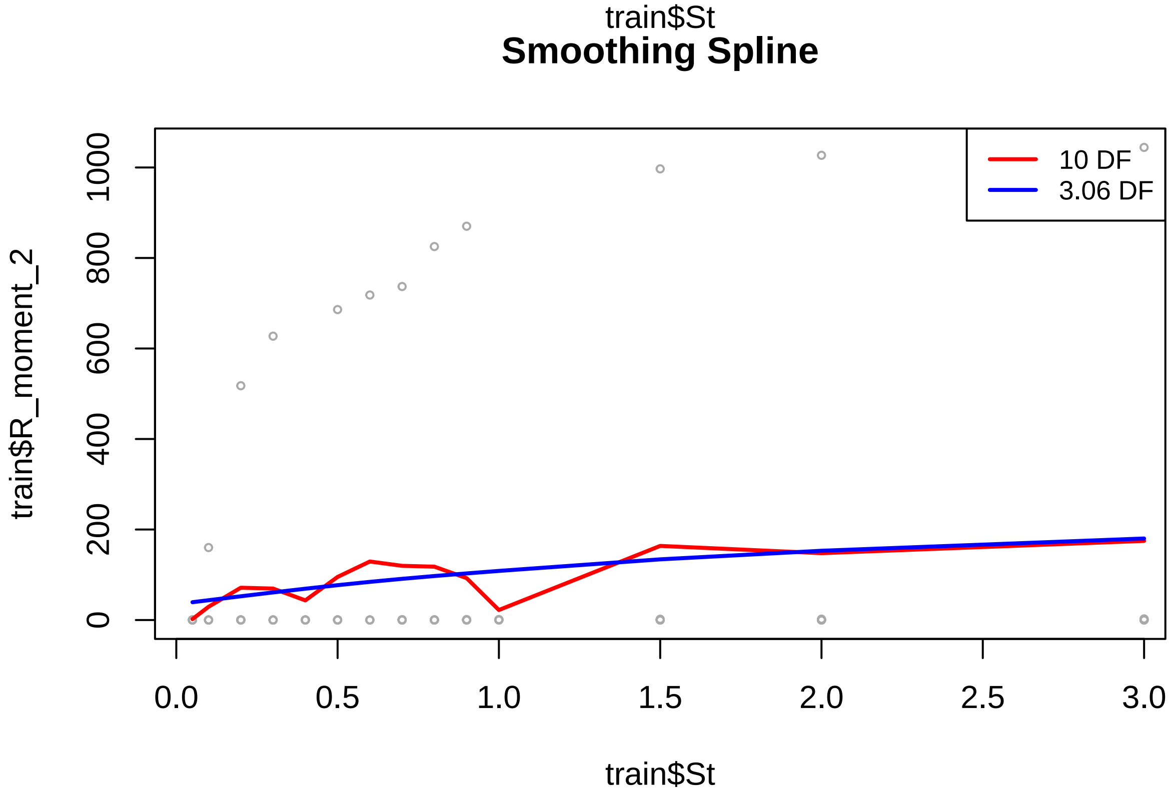
Splines



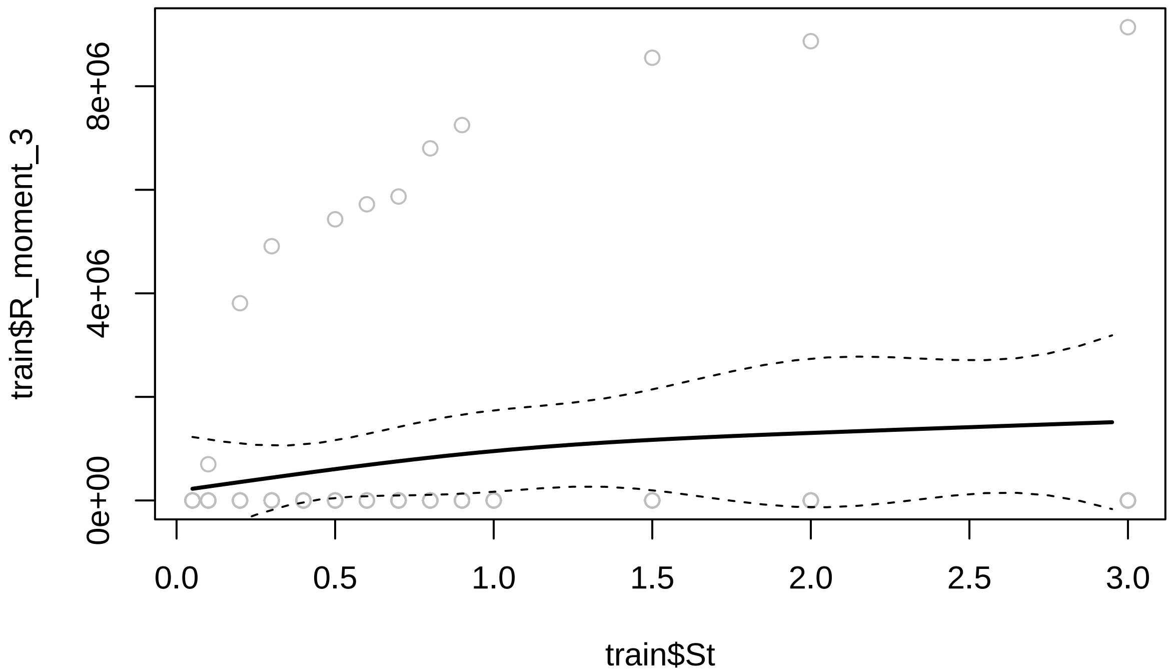


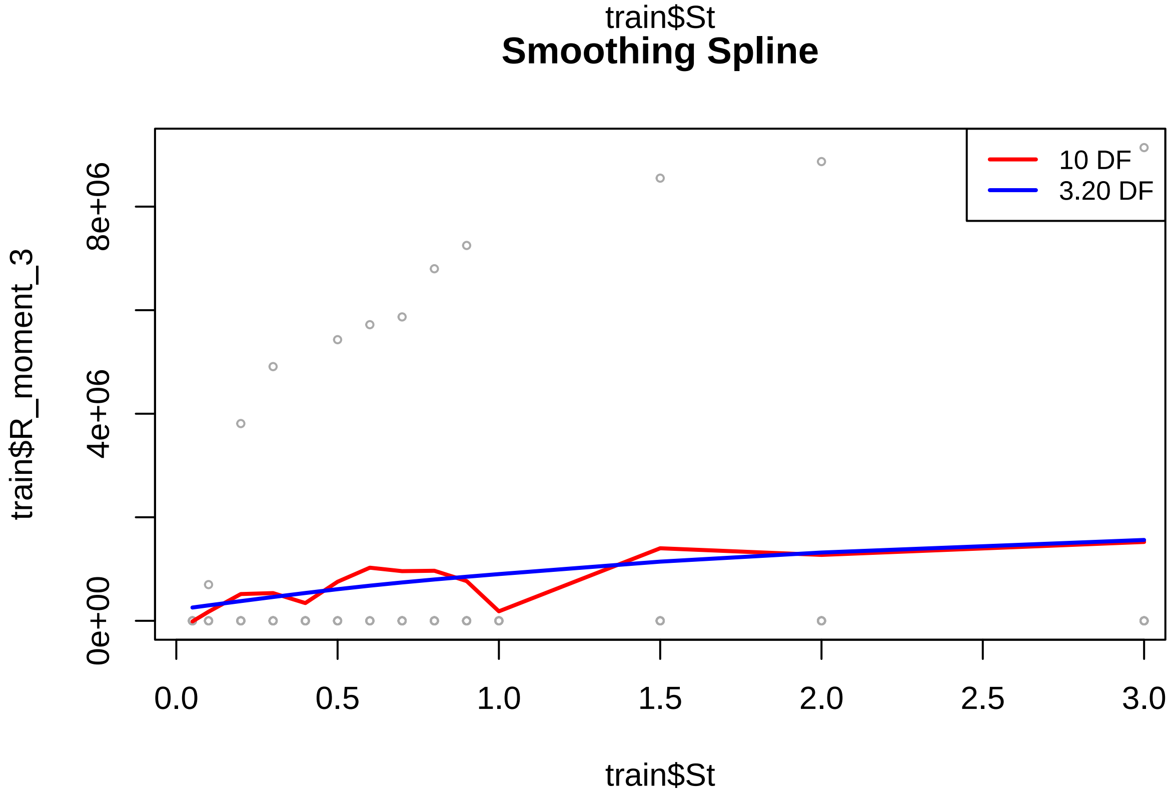


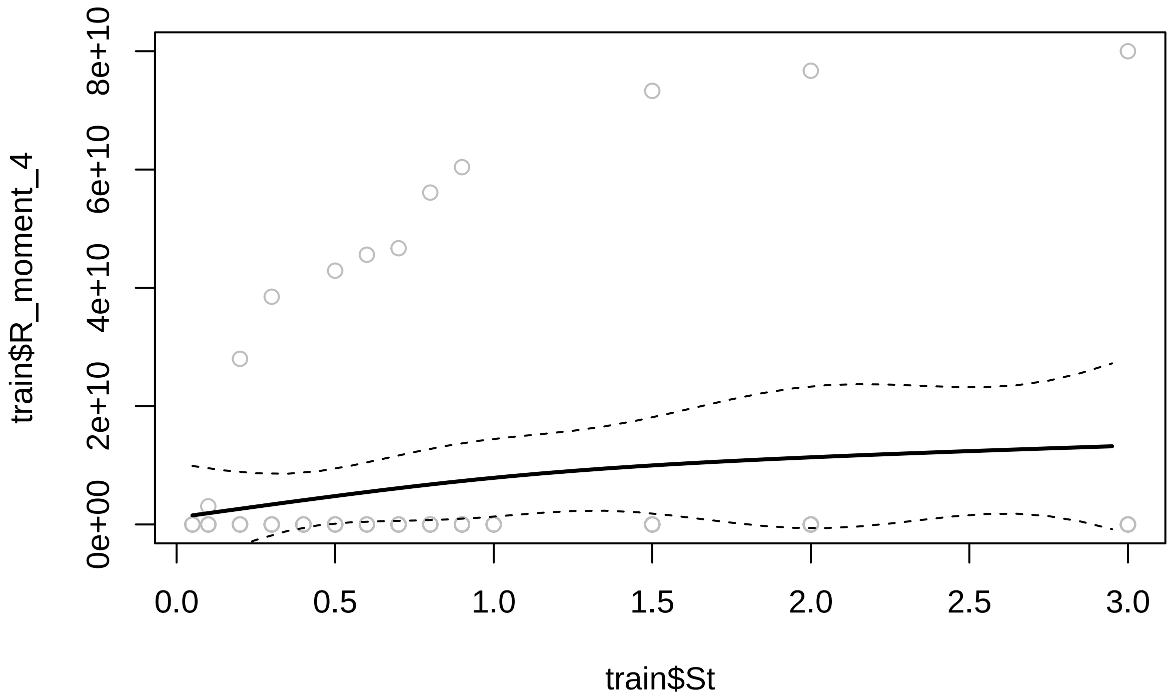
train$R_moment_2
0
200
400
600
800
1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St



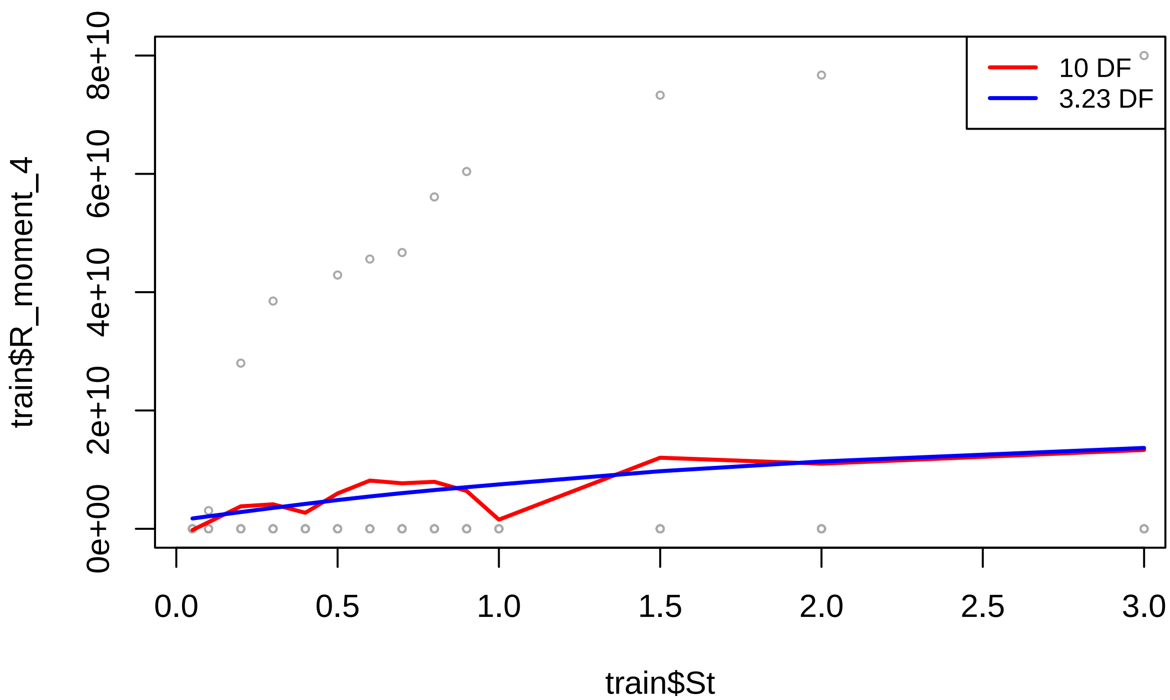
Smoothing Spline
10 DF
3.06 DF
train$R_moment_2
0
200
400
600
800
1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St


train$R_moment_3
0e+00
4e+06
8e+06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St


Smoothing Spline
10 DF
3.20 DF
train$R_moment_3
0e+00
4e+06
8e+06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St



Smoothing Spline



Results

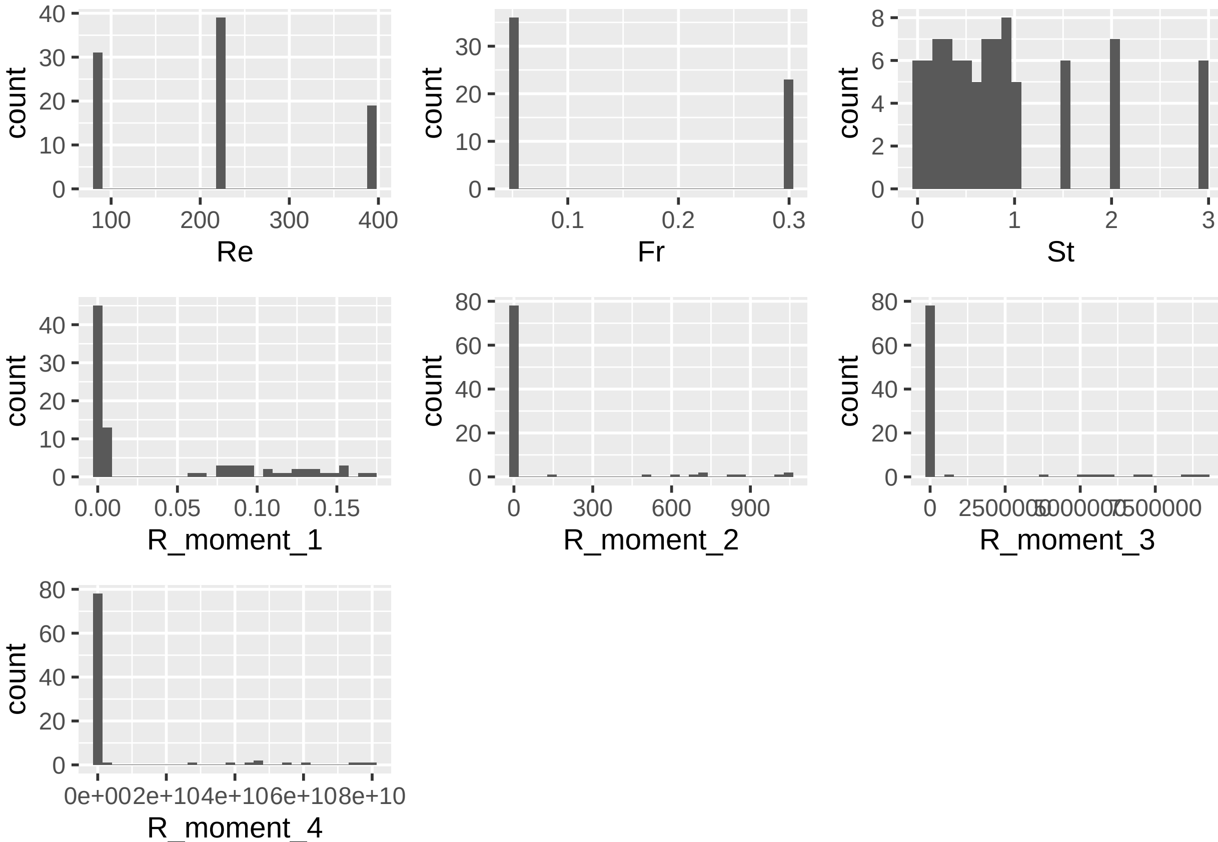
(Final Model ?)

Conclusion

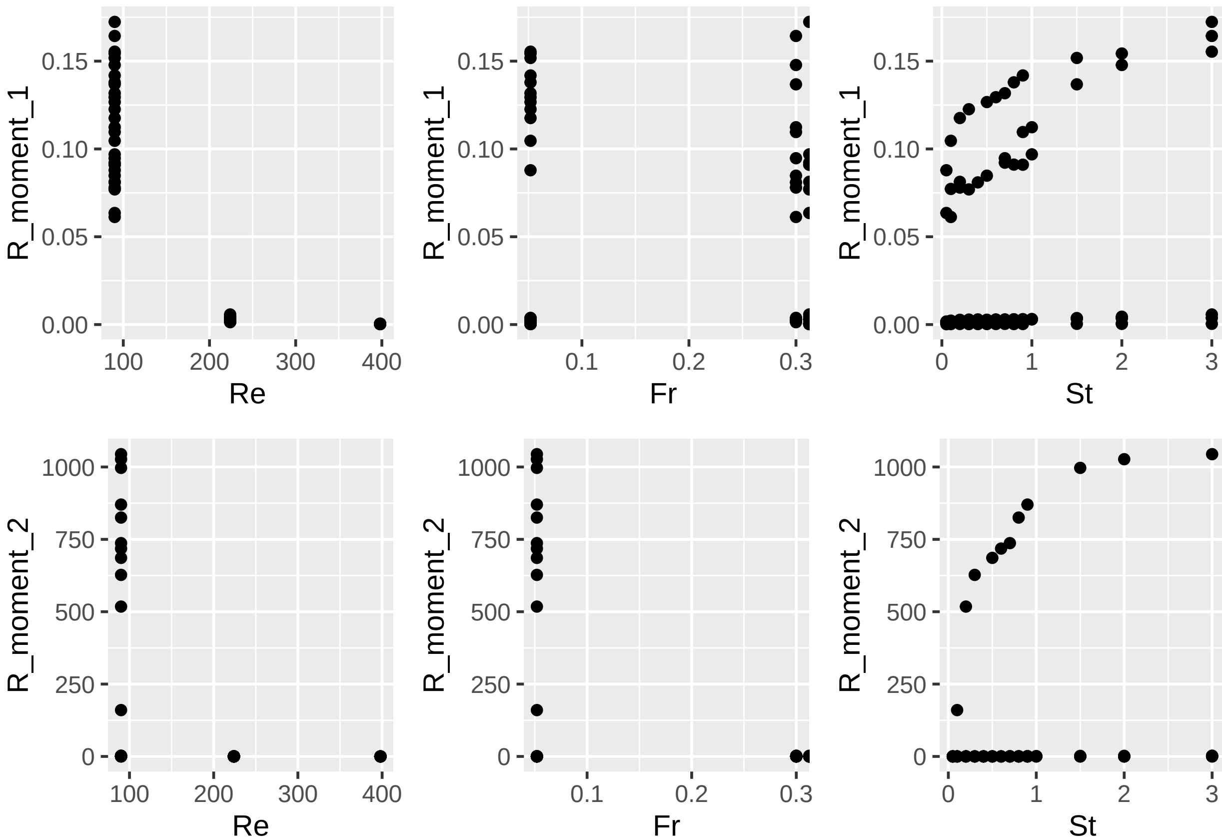
Appendix

EDA

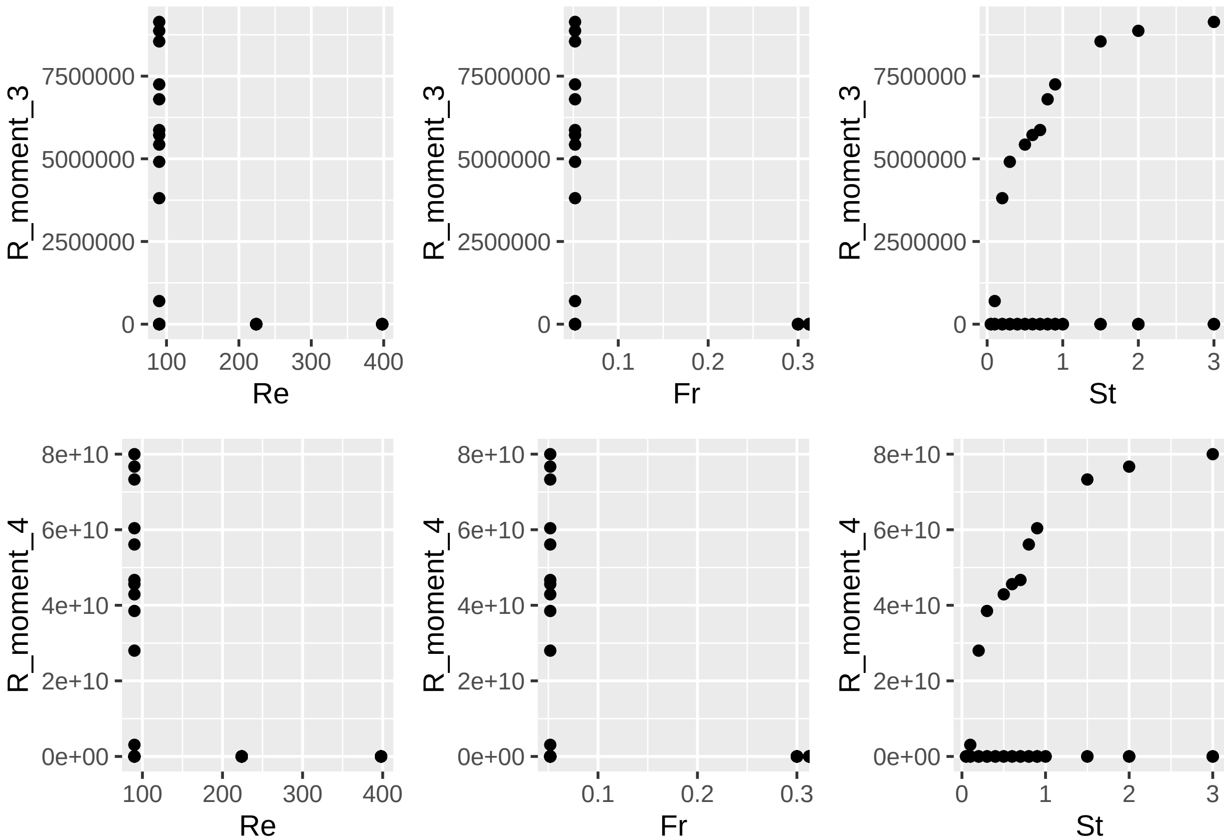
```
##        St               Re              Fr           R_moment_1
##  Min.   :0.0500   Min.   : 90.0   Min.   :0.052   Min.   :0.000222
##  1st Qu.:0.3000   1st Qu.: 90.0   1st Qu.:0.052   1st Qu.:0.002157
##  Median :0.7000   Median :224.0   Median :0.300   Median :0.002958
##  Mean   :0.8596   Mean   :214.5   Mean   :  Inf   Mean   :0.040394
##  3rd Qu.:1.0000   3rd Qu.:224.0   3rd Qu.:  Inf   3rd Qu.:0.087868
##  Max.   :3.0000   Max.   :398.0   Max.   :  Inf   Max.   :0.172340
##    R_moment_2         R_moment_3       R_moment_4          gravity
##  Min.   :   0.0001   Min.   :      0   Min.   :0.000e+00   Length:89
##  1st Qu.:   0.0245   1st Qu.:      0   1st Qu.:3.000e+00   Class :character
##  Median :   0.0808   Median :      1   Median :2.100e+01   Mode  :character
##  Mean   :  92.4902   Mean   : 753370   Mean   :6.194e+09
##  3rd Qu.:   0.5345   3rd Qu.:     40   3rd Qu.:5.345e+03
##  Max.   :1044.3000   Max.   :9140000   Max.   :8.000e+10
##      flow
##  Length:89
##  Class :character
##  Mode  :character
##
##
##
```




```
## TableGrob (3 x 3) "arrange": 7 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (1-1,3-3) arrange gtable[layout]
## 4 4 (2-2,1-1) arrange gtable[layout]
## 5 5 (2-2,2-2) arrange gtable[layout]
## 6 6 (2-2,3-3) arrange gtable[layout]
## 7 7 (3-3,1-1) arrange gtable[layout]
```




```
## TableGrob (2 x 3) "arrange": 6 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (1-1,3-3) arrange gtable[layout]
## 4 4 (2-2,1-1) arrange gtable[layout]
## 5 5 (2-2,2-2) arrange gtable[layout]
## 6 6 (2-2,3-3) arrange gtable[layout]
```




```
## TableGrob (2 x 3) "arrange": 6 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (1-1,3-3) arrange gtable[layout]
## 4 4 (2-2,1-1) arrange gtable[layout]
## 5 5 (2-2,2-2) arrange gtable[layout]
## 6 6 (2-2,3-3) arrange gtable[layout]
```