Turbulence Analysis

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

2023-10-27

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 $E[X^4]$ 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, utilizing the data in a training set provided. Specifically, we are interested in the following:

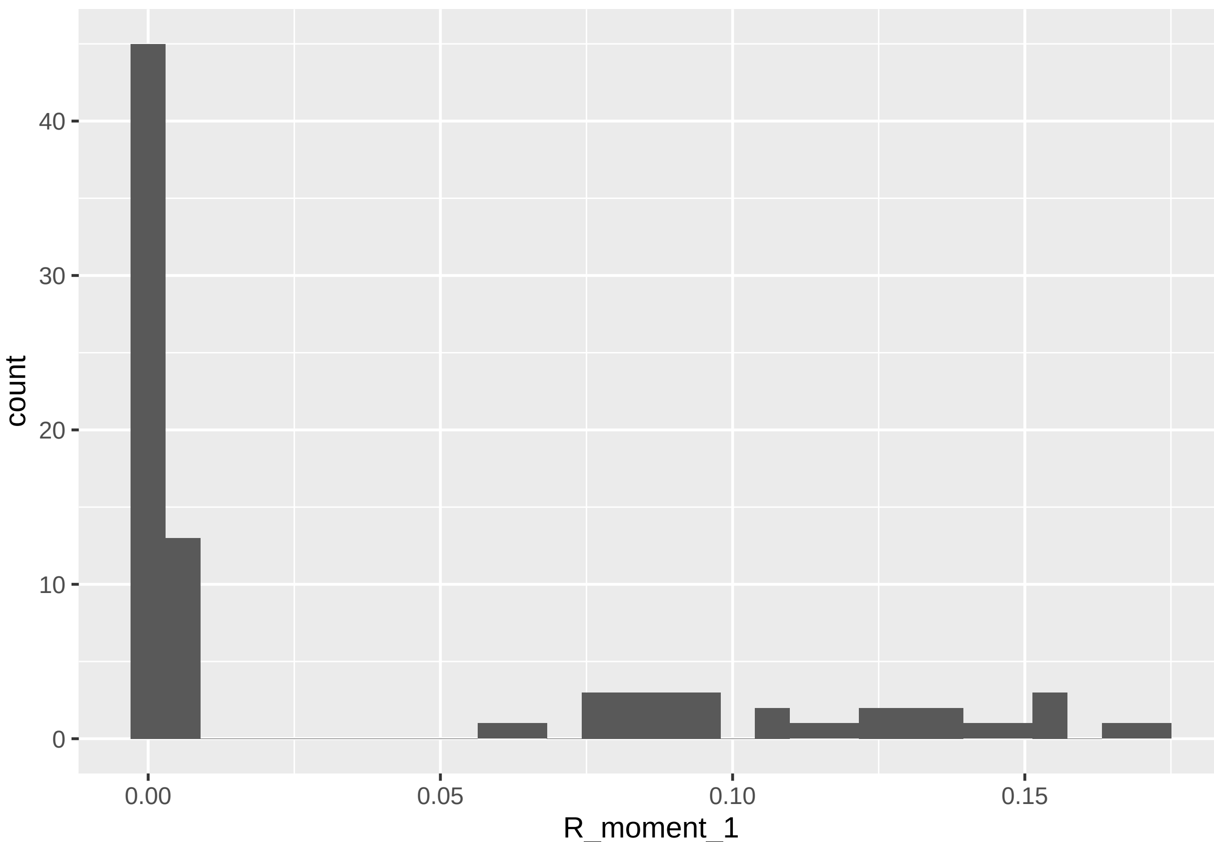
- Does there exist a significant linear relationship between the predictors and the raw four moments?
- Is there any significant interaction effects between predictors on the response variables?
- Does a linear regression model suffice? Or a more complex model is needed to better explain the relationship between the predictor and response
- Are identified effects for predictors the same for all moments, or they differ for each different moment?

Ultimately, we wish our model to capture sufficient trends in our training data, so that we can predict the four moments in our test set data as accurately as possible.

Methodology

We begin by some transformations on both predictor and response variables. For predictor variable, we first noticed that `Fr` only takes on 0.052, 0.3 and Inf in our training and testing data set, and directly using it is not viable since it contains infinity. Since `Fr`<1 corresponds with a subcritical flow while `Fr` > 1 corresponds with a super critical flow, we create a new categorical variable `flow` by the following:

Flow	Fr
super subcritical	Fr < 0.1
subcritical	0.1 < Fr < 1
supercritical	Fr > 1



Results

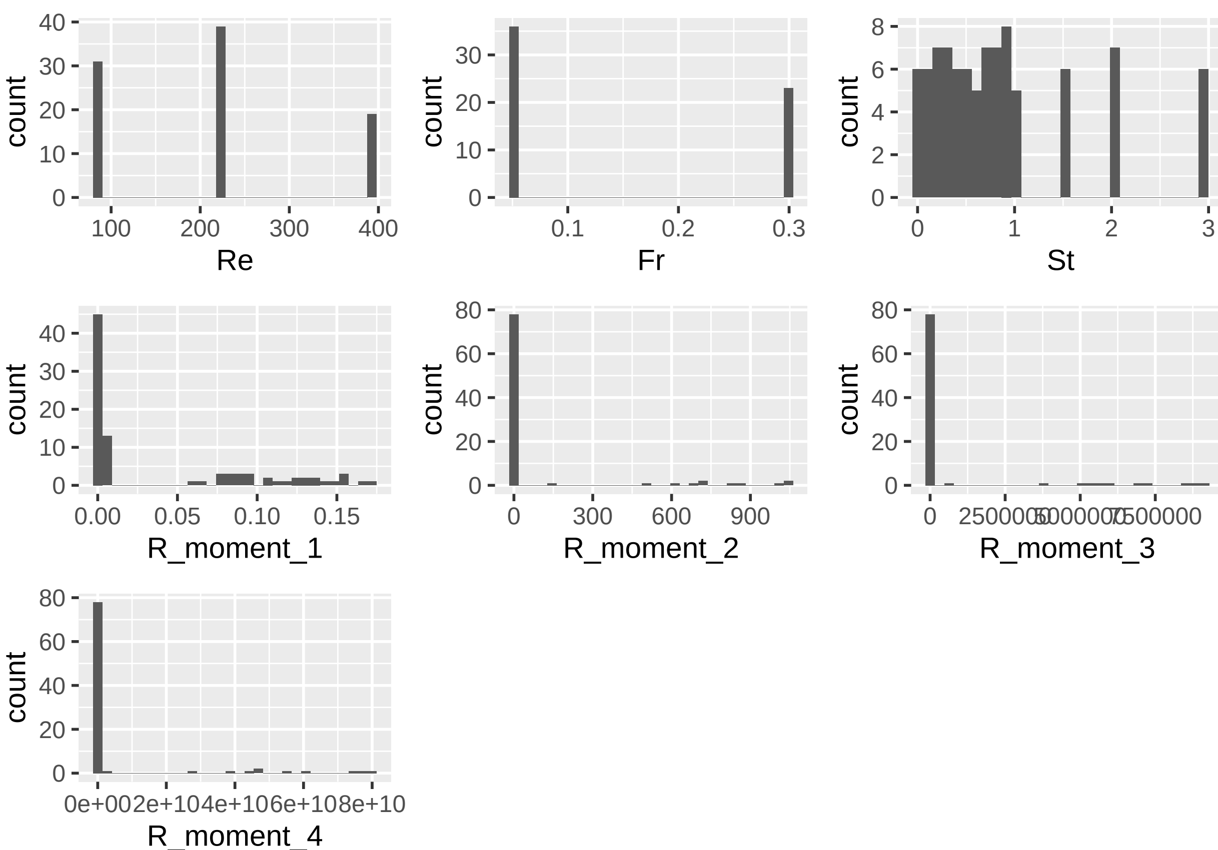
Conclusion

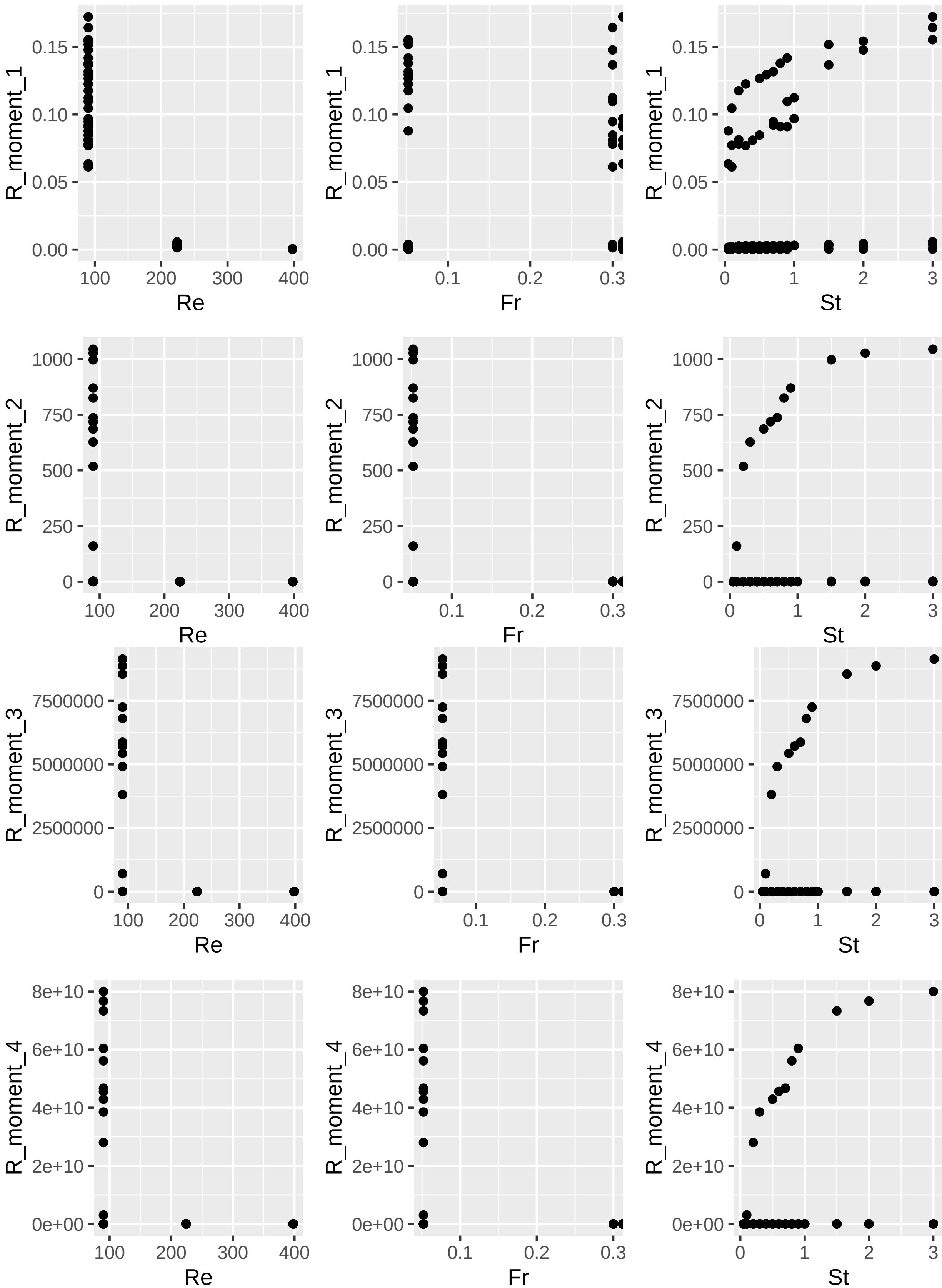
Appendix

EDA

```
##        St               Re              Fr           R_moment_1
##  Min.   :0.0500   Min.   : 90.0   Min.   :0.052   Min.   :0.000222
```

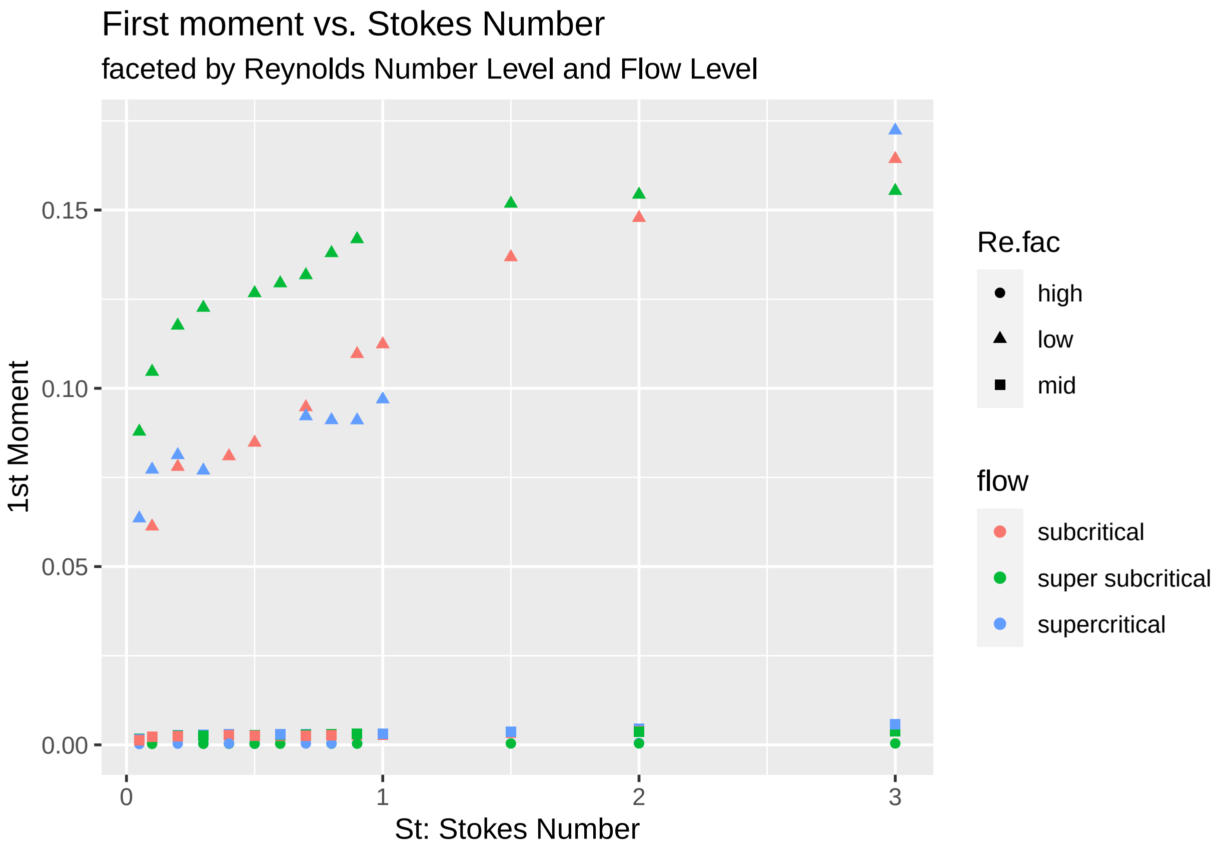
```
##  1st Qu.:0.3000   1st Qu.: 90.0   1st Qu.:0.052   1st Qu.:0.002157
##  Median :0.7000   Median :224.0   Median :0.300   Median :0.002958
##  Mean   :0.8596   Mean   :214.5   Mean   :  Inf   Mean   :0.040394
##  3rd Qu.:1.0000   3rd Qu.:224.0   3rd Qu.:  Inf   3rd Qu.:0.087868
##  Max.   :3.0000   Max.   :398.0   Max.   :  Inf   Max.   :0.172340
##    R_moment_2          R_moment_3        R_moment_4            flow
##  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
##  log.moment.2 <- log(R_moment_2) log.moment.3 <- log(R_moment_3)
##  Min.   :-9.1805                 Min.   :-9.8759
##  1st Qu.:-3.7101                 1st Qu.:-1.4131
##  Median :-2.5157                 Median : 0.1692
##  Mean   :-1.6941                 Mean   : 2.1070
##  3rd Qu.:-0.6264                 3rd Qu.: 3.7002
##  Max.   : 6.9511                 Max.   :16.0282
##  log.moment.4 <- log(R_moment_4)    Re.fac
##  Min.   :-10.087                 Length:89
##  1st Qu.:  1.185                 Class :character
##  Median :  3.037                 Mode  :character
##  Mean   :  5.954
##  3rd Qu.:  8.584
##  Max.   : 25.105
```





The plot below suggests a very possible interaction effect between Stokes number and Reynolds number on

1st Moment:



Simple Linear Regression

- We made `Fr` a categorical variable when fitting a linear regression model, as `Fr` only has three unique values both in the training and testing dataset; one of these values is `Inf`, which should not be used in a linear regression analysis.

First Moment

```
##
## Call:
## lm(formula = R_moment_1 ~ Re.fac + St + flow + Re.fac:St, data = train)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.024358 -0.006729  0.003284  0.004195  0.023329
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -0.0037676  0.0043378  -0.869 0.387662
## Re.faclow                 0.0854744  0.0044191  19.342  < 2e-16 ***
## Re.facmid                 0.0023091  0.0043881   0.526 0.600175
## St                       -0.0016501  0.0031049  -0.531 0.596560
## flowsuper subcritical     0.0105639  0.0028057   3.765 0.000314 ***
## flowsupercritical        -0.0001185  0.0029338  -0.040 0.967876
```

```
## Re.faclow:St           0.0308060  0.0037499   8.215 2.83e-12 ***
## Re.facmid:St           0.0023811  0.0038207   0.623 0.534893
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01005 on 81 degrees of freedom
## Multiple R-squared:  0.9702, Adjusted R-squared:  0.9676
## F-statistic: 376.3 on 7 and 81 DF,  p-value: < 2.2e-16
```

Using 5-fold cross-validation to estimate the test set error

```
## Linear Regression
##
## 89 samples
##  3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 73, 72, 69, 72, 70
## Resampling results:
##
##   RMSE        Rsquared   MAE
##   0.01415704  0.9398071  0.009977374
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

Trying using polynomial terms up to degree of 5 for stokes number:

```
## Analysis of Variance Table
##
## Model 1: response ~ St + flow + Re.fac
## Model 2: response ~ poly(St, 2) + flow + Re.fac
## Model 3: response ~ poly(St, 3) + flow + Re.fac
## Model 4: response ~ poly(St, 4) + flow + Re.fac
## Model 5: response ~ poly(St, 5) + flow + Re.fac
##   Res.Df      RSS Df  Sum of Sq      F Pr(>F)
## 1     83 0.019399
## 2     82 0.019352  1 4.7187e-05 0.1959 0.6593
## 3     81 0.019180  1 1.7206e-04 0.7142 0.4006
## 4     80 0.019134  1 4.5704e-05 0.1897 0.6643
## 5     79 0.019031  1 1.0305e-04 0.4278 0.5150
```

Judging from the p value for the associated F-statistics, only the first order term is necessary.

Moments 2-4

```
##
## Call:
## lm(formula = log(R_moment_2) ~ Re + St + flow, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -5.7541 -1.0168 -0.3029  0.8348  3.4238
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.433205   0.550793   2.602  0.01095 *
## Re                    -0.025963   0.001856 -13.989  < 2e-16 ***
## St                     0.733752   0.258191   2.842  0.00563 **
## flowsuper subcritical  3.700934   0.520442   7.111 3.52e-10 ***
## flowsupercritical      0.929358   0.542426   1.713  0.09034 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.9 on 84 degrees of freedom
## Multiple R-squared:  0.7499, Adjusted R-squared:  0.738
## F-statistic: 62.98 on 4 and 84 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = log(R_moment_3) ~ Re + St + flow, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.5905 -1.9037 -0.4285  1.7964  5.9281
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            5.139254   0.948688   5.417 5.66e-07 ***
## Re                    -0.033938   0.003197 -10.616  < 2e-16 ***
## St                     0.964896   0.444709   2.170   0.0329 *
## flowsuper subcritical  7.104356   0.896411   7.925 8.56e-12 ***
## flowsupercritical      1.611925   0.934277   1.725   0.0881 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.272 on 84 degrees of freedom
## Multiple R-squared:  0.683,  Adjusted R-squared:  0.6679
## F-statistic: 45.25 on 4 and 84 DF,  p-value: < 2.2e-16
```

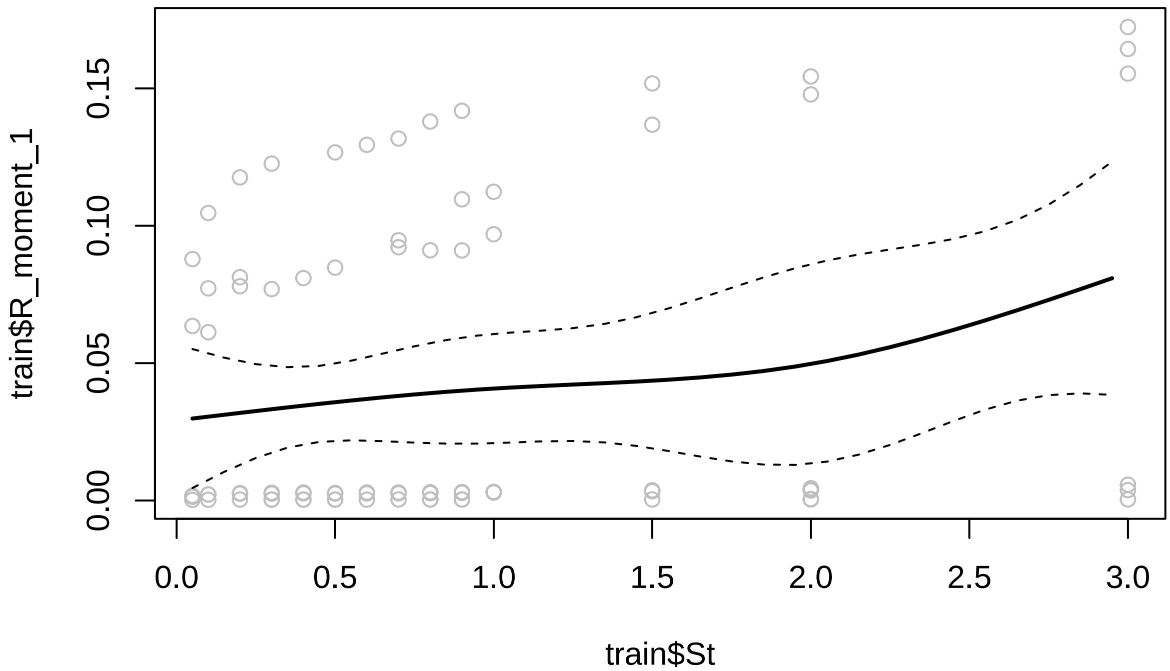
```
##
## Call:
## lm(formula = log(R_moment_4) ~ Re + St + flow, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0985  -2.8732  -0.7093   2.6849   8.3406
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            8.999300   1.334956   6.741 1.86e-09 ***
## Re                    -0.042210   0.004498  -9.383 1.00e-14 ***
## St                     1.152984   0.625777   1.842   0.0689 .
## flowsuper subcritical 10.487017   1.261394   8.314 1.42e-12 ***
## flowsupercritical      2.299173   1.314678   1.749   0.0840 .
## ---
```

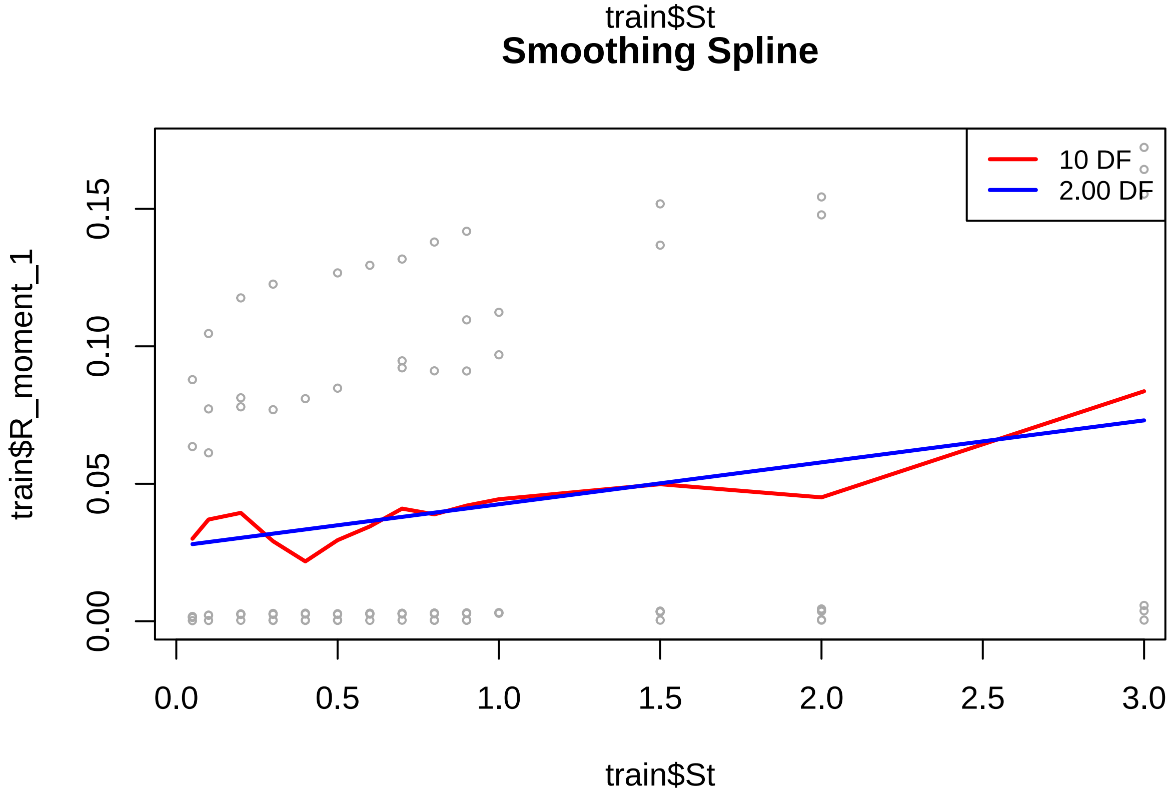
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.605 on 84 degrees of freedom
## Multiple R-squared:  0.6607, Adjusted R-squared:  0.6445
## F-statistic: 40.89 on 4 and 84 DF,  p-value: < 2.2e-16
```

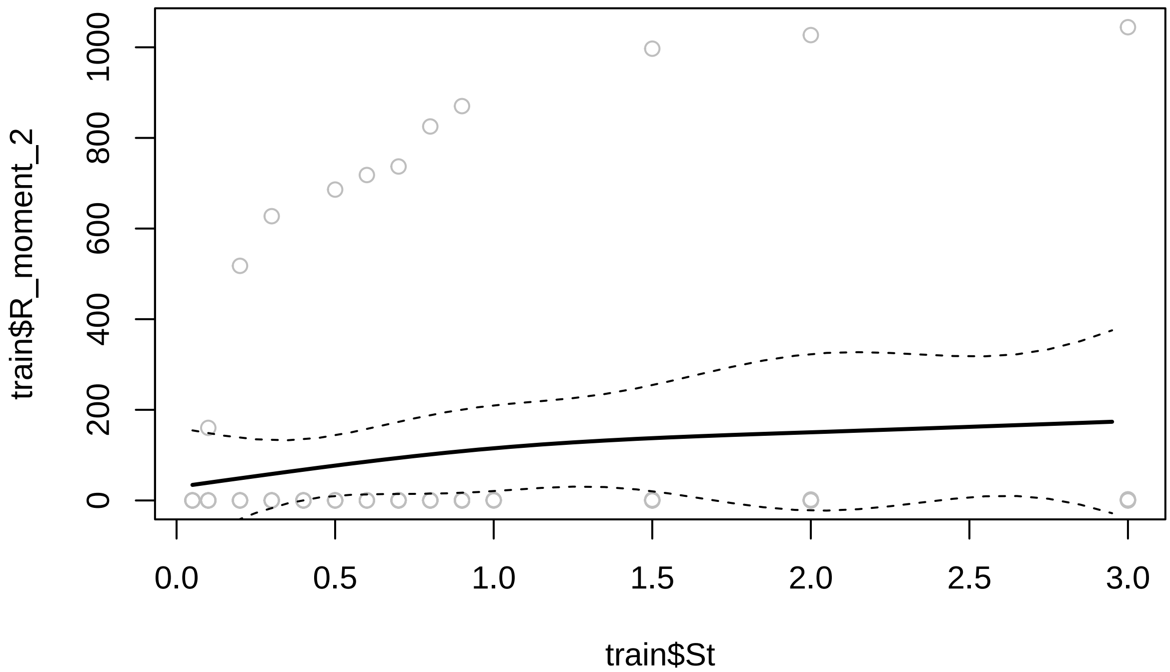
Considering a simple linear regression on the first moment: we have a 0.97 adjusted R squared value and significant F-statistics; however, the residual vs fitted values plot indicates a obvious non-linear trend, which suggests that the linearity assumption is violated.

Ridge Regression

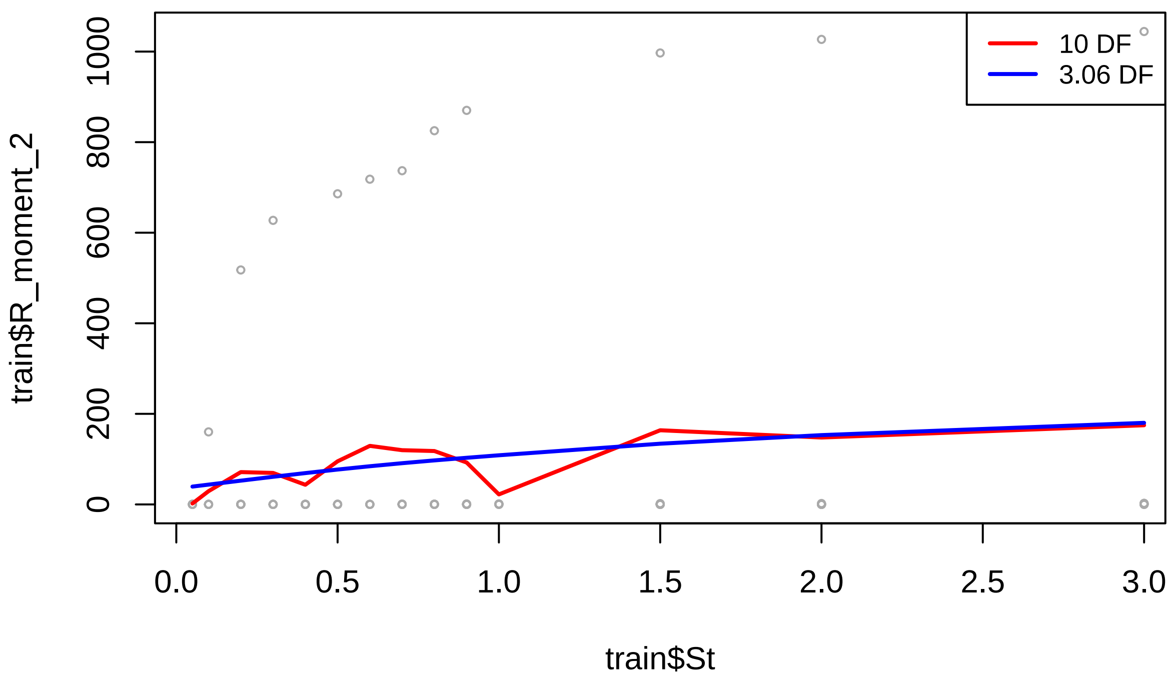
Natural Splines

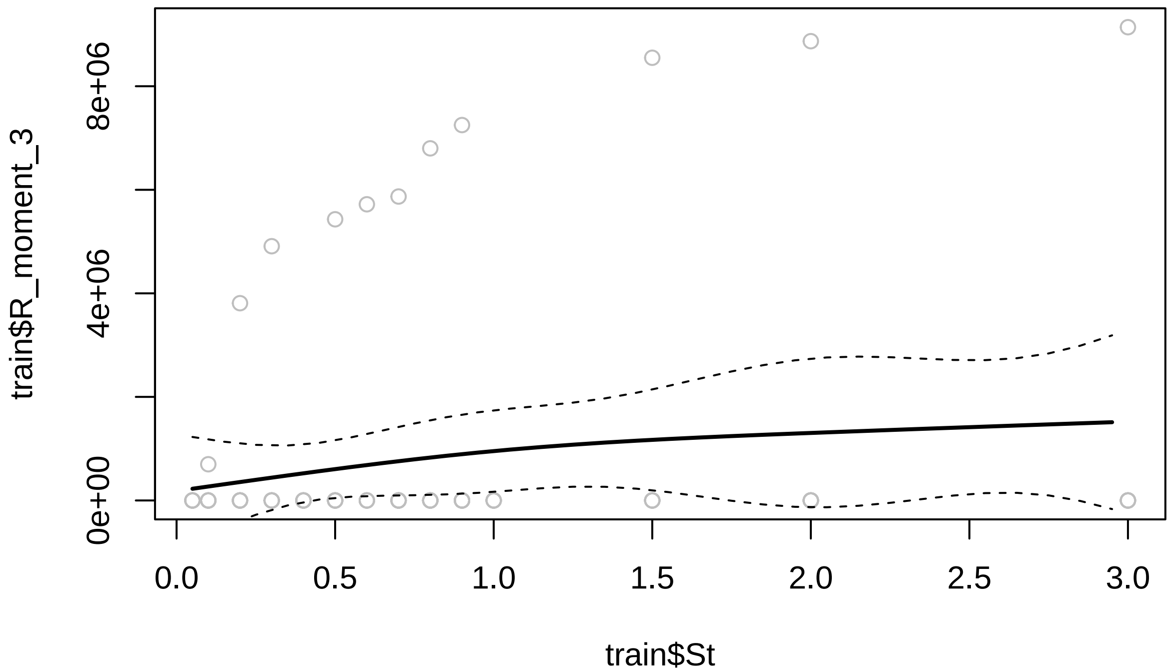




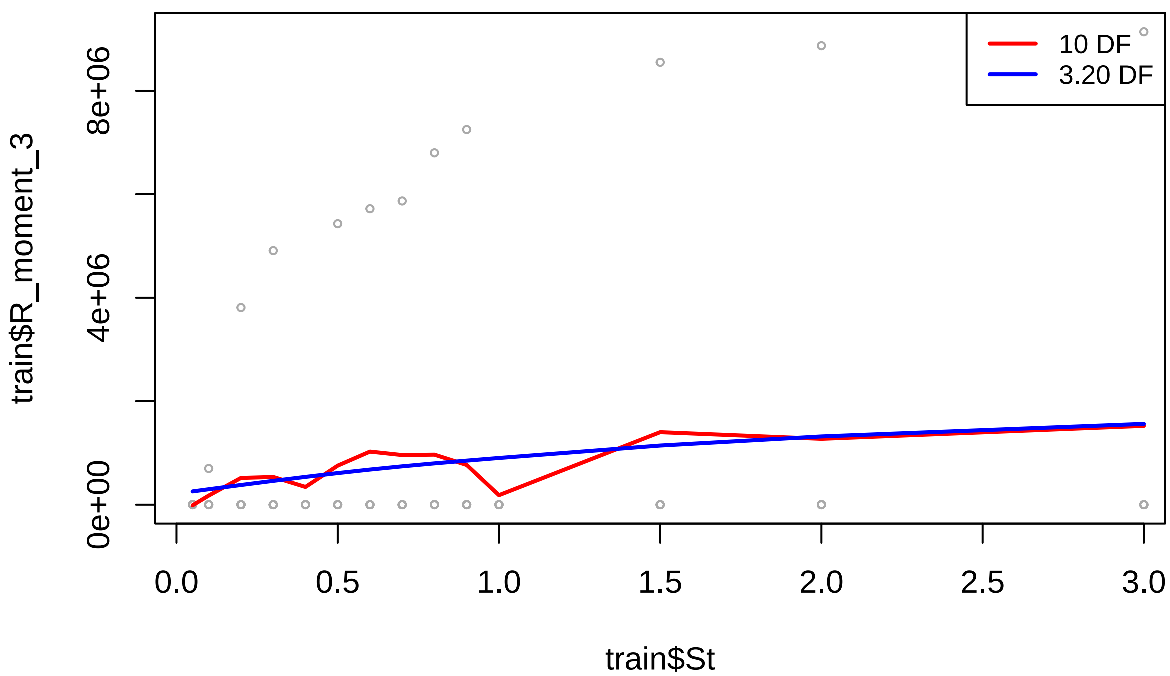


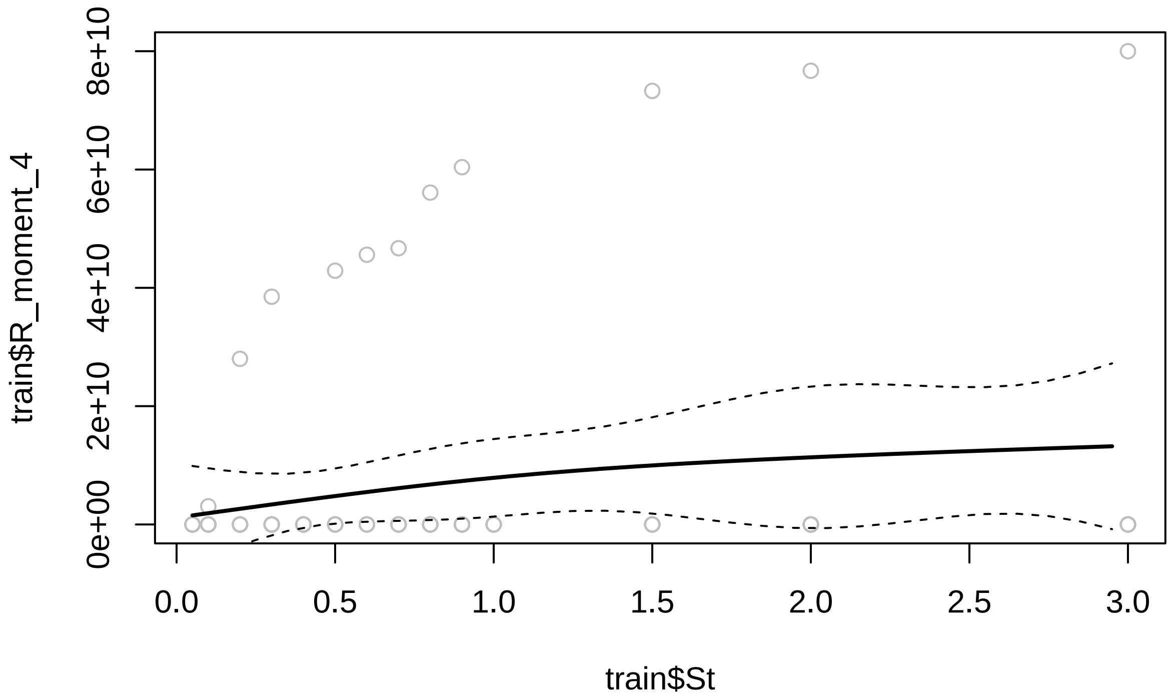
Smoothing Spline

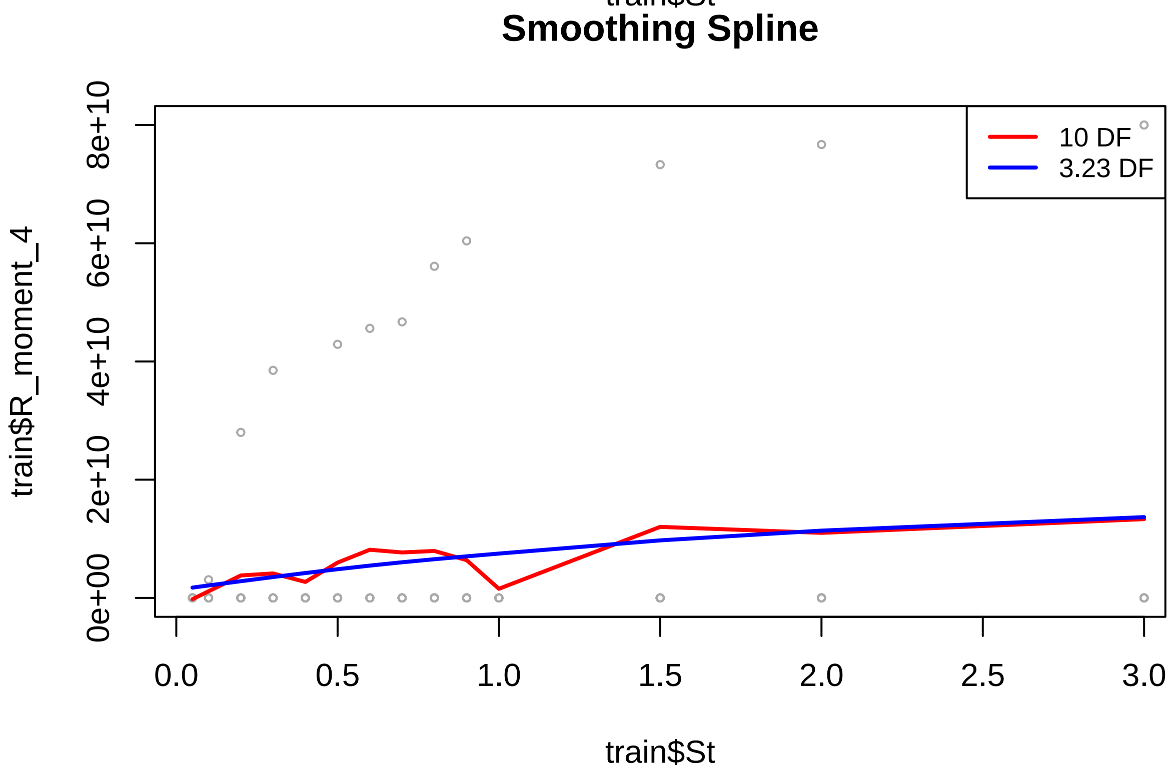




Smoothing Spline




train$R_moment_4
0e+00
2e+10
4e+10
6e+10
8e+10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St


Smoothing Spline
10 DF
3.23 DF
train$R_moment_4
0e+00
2e+10
4e+10
6e+10
8e+10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St