

# Initial Testing & Setup

*September 5, 2018*

## Intro

- This report summarizes some of the developments of the methods / scripts needed for modeling, as well as to look at some basic data visualisations and simple linear regression methods.
- The examples are from Enniscrone, by looking at the bivariate relationship between water quality and the Met Eirann gauged precipitation data nearby.

## Data Setup

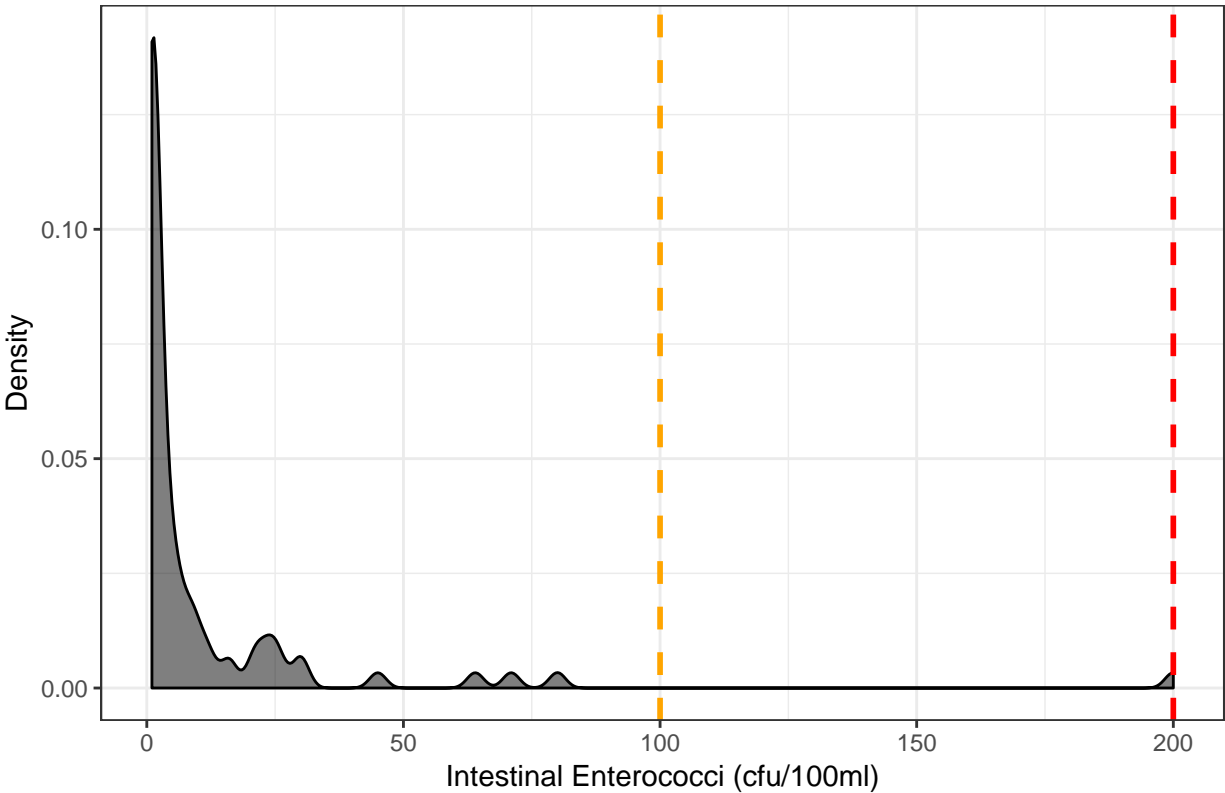
- Water Quality Data
  - Intestinal Enterococci (cfu/100ml)
  - E-Coli (cfu/100ml)
- Precip Data from Gauges (mm)
  - Rain at Day 0, -1, -2, -3
  - Aggregate over 4 Days

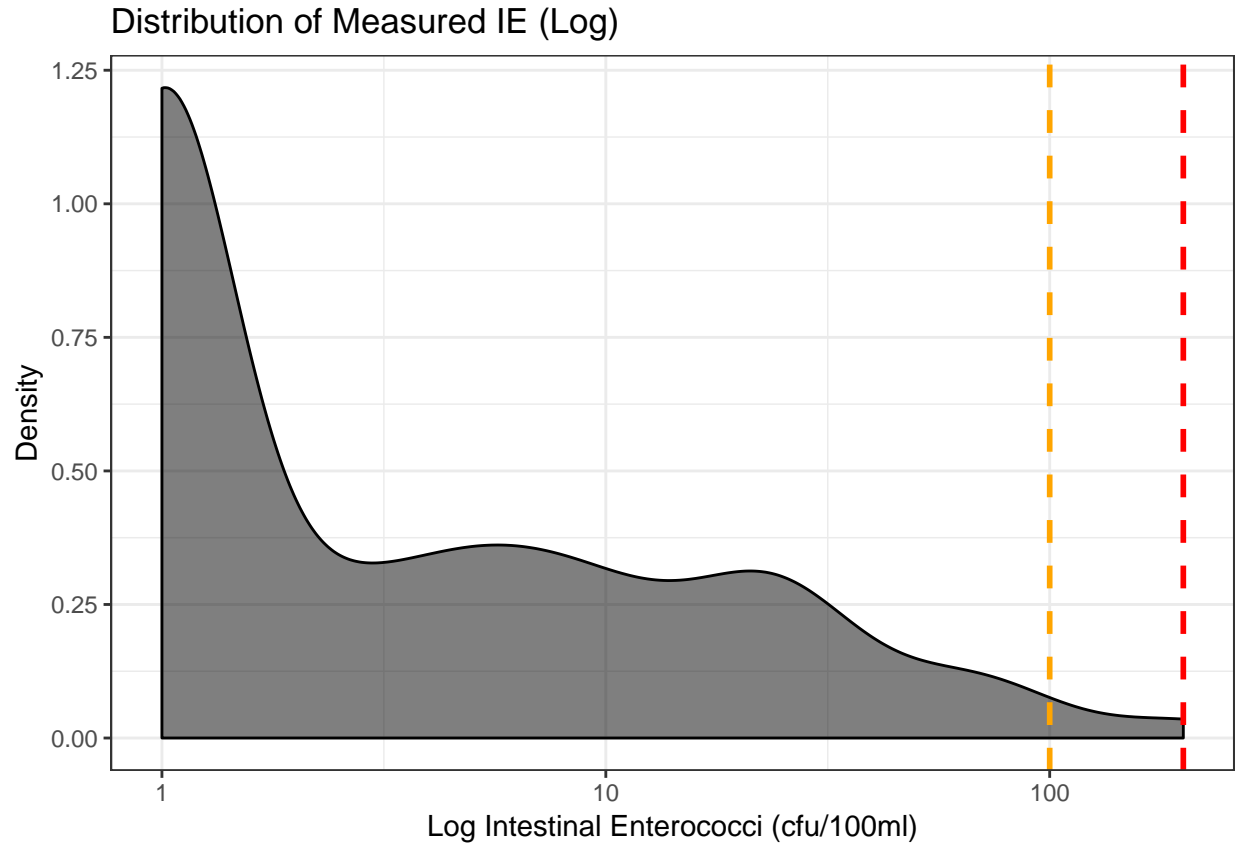
## Intestinal Enterococci

### Density Plots

- These plots show the distribution of the observed IE values.
- Vertical lines
  - Red = limit for “Sufficient” water quality (200 cfu/100ml)
  - Orange = limit for “Good” water quality (100 cfu/100ml)

Distribtion of Measured IE

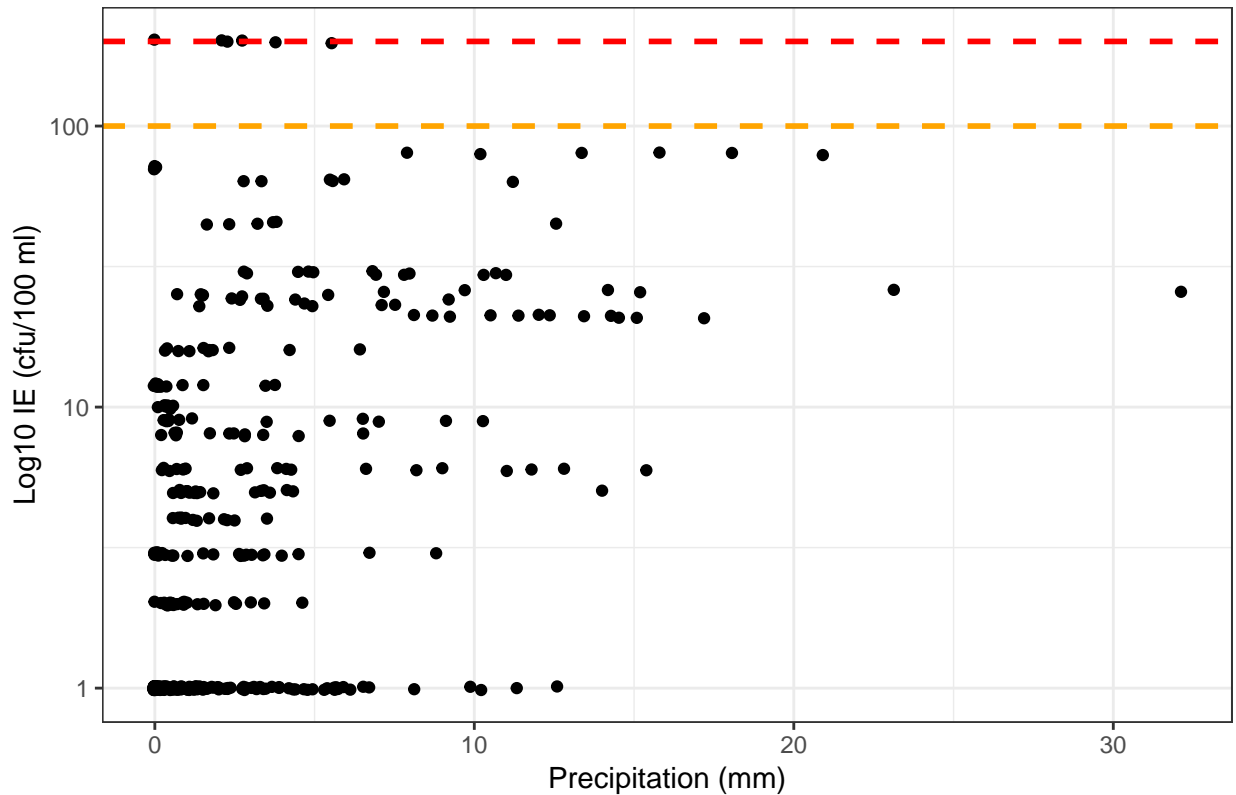




### Scatterplots

- This plot shows the relationship between measured IE & same-day precipitation
- Horizontal lines
  - Red = limit for “Sufficient” water quality (200 cfu/100ml)
  - Orange = limit for “Good” water quality (100 cfu/100ml)

## Intestinal Enterococci ~ Day-1 Precipitation



## Bi-Variate Linear Regression

- This matrix shows the R-Squared value of the Intestinal Enterococci ~ Rainfall
  - ie\_rain = same-day
  - ie\_rain1 = day-1
  - ie\_rain2 = day-2
  - ie\_rain3 = day-3
  - ie\_rain\_sum = 4-day sum
- Each row shows the results from a different Met Eierann gauge

```
##          IE_rain IE_rain1 IE_rain2 IE_rain3 IE_rain_sum
## dly3135 0.0067   0.026   4.8e-03 0.00073 0.00234
## dly3436 0.0014   0.049   2.0e-03 0.02260 0.01154
## dly3735 0.0047   0.090   5.8e-03 0.00840 0.00052
## dly4235 0.0044   0.050   8.8e-03 0.00230 0.01237
## dly4635 0.0101   0.107   8.5e-05 0.00532 0.01011
## dly5735 0.0012   0.109   9.5e-03 0.00116 0.00680
```

- From this test, there were no strong relationships

```
##          IE_log_rain IE_log_rain1 IE_log_rain2 IE_log_rain3 IE_log_rain_sum
## dly3135 0.00922     0.22     0.0278     1.3e-02     0.043
## dly3436 0.00229     0.19     0.0017     4.5e-02     0.067
```

```

## dly3735      0.01170      0.26      0.0014      4.2e-06      0.022
## dly4235      0.00158      0.19      0.0459      3.6e-03      0.097
## dly4635      0.02008      0.25      0.0100      1.4e-03      0.064
## dly5735      0.00076      0.23      0.0017      1.3e-02      0.041

##          IE_log10_rain IE_log10_rain1 IE_log10_rain2 IE_log10_rain3
## dly3135      0.00922      0.22      0.0278      1.3e-02
## dly3436      0.00229      0.19      0.0017      4.5e-02
## dly3735      0.01170      0.26      0.0014      4.2e-06
## dly4235      0.00158      0.19      0.0459      3.6e-03
## dly4635      0.02008      0.25      0.0100      1.4e-03
## dly5735      0.00076      0.23      0.0017      1.3e-02
##          IE_log10_rain_sum
## dly3135      0.043
## dly3436      0.067
## dly3735      0.022
## dly4235      0.097
## dly4635      0.064
## dly5735      0.041

```

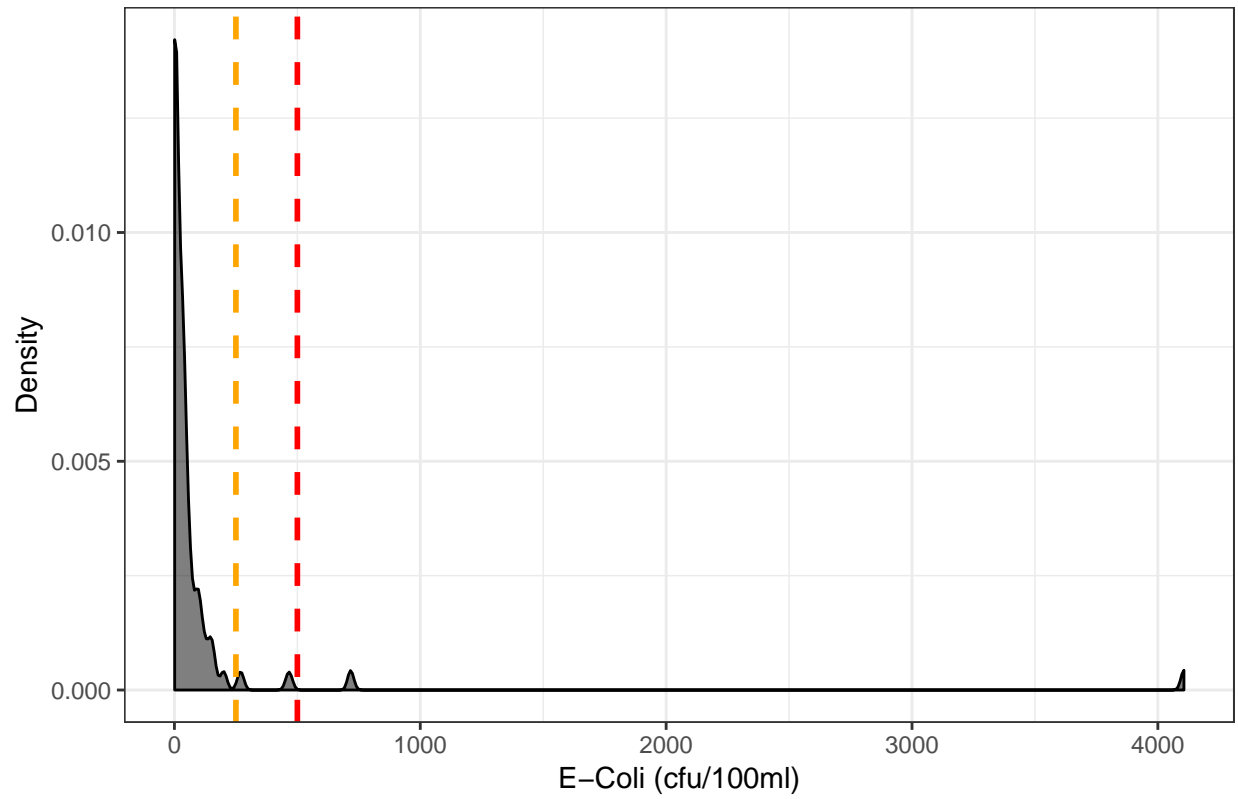
- The relationship is a bit better when considering the dependent variable with a log transformation (natural or base10)

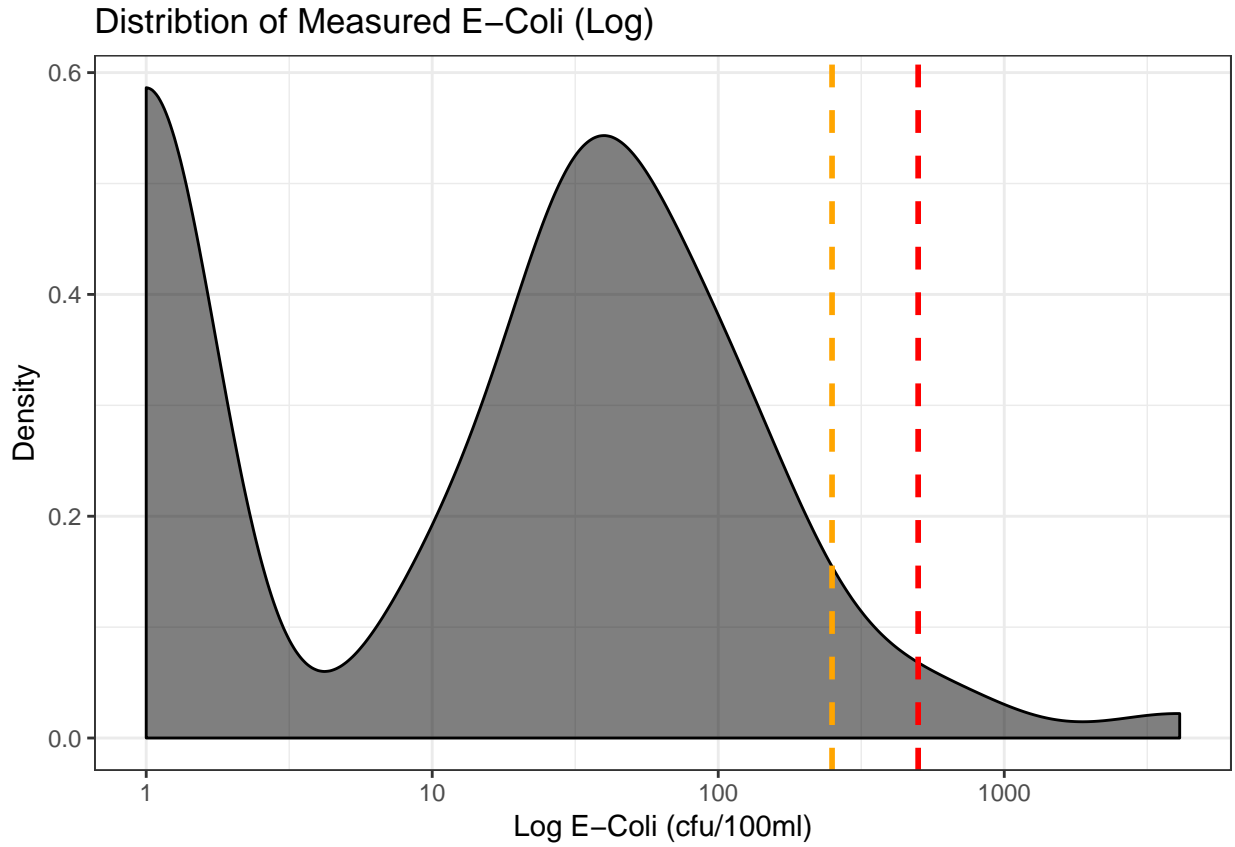
## E-Coli

### Density Plots

- These plots show the distribution of the observed E-Coli values.
- Vertical lines
  - Red = limit for “SufficIEnt” water quality (500 cfu/100ml)
  - Orange = limit for “Good” water quality (250 cfu/100ml)

Distribution of Measured E-Coli

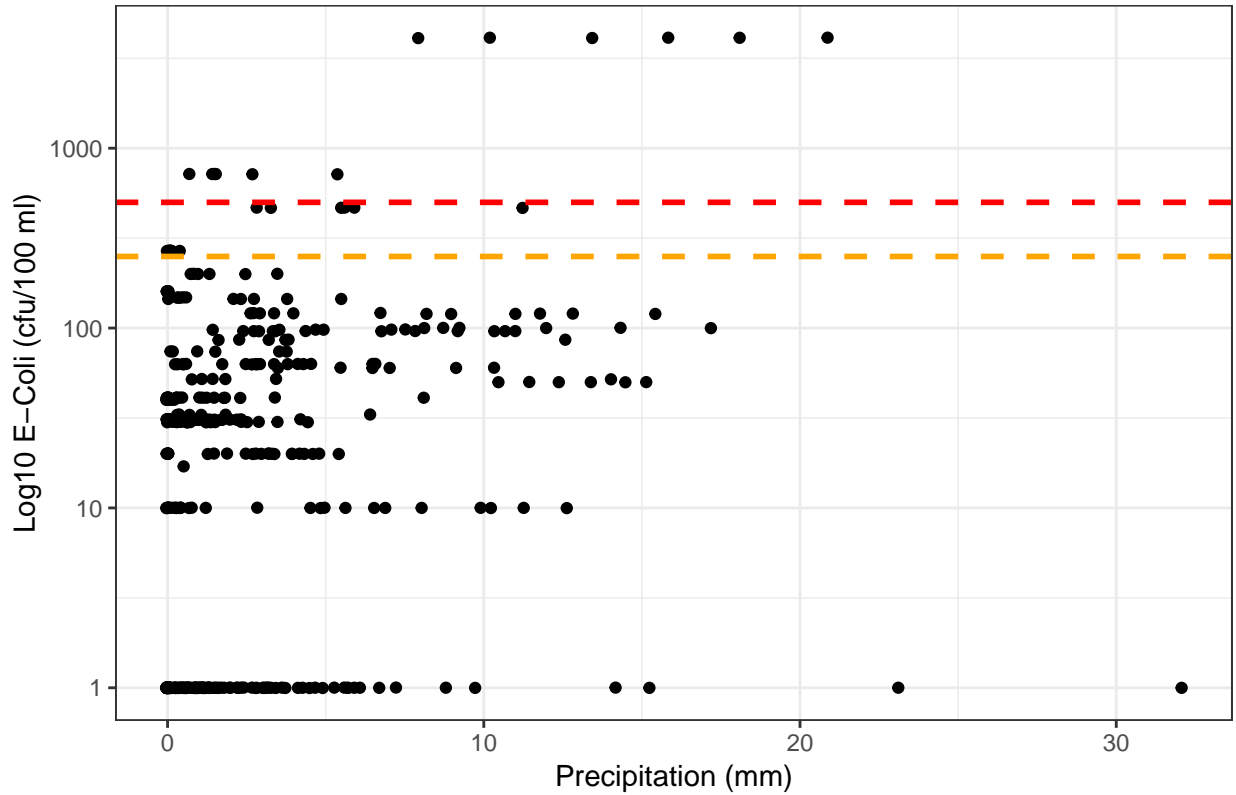




## Scatterplots

- This plot shows the relationship between measured IE & same-day precipitation
- Horizontal lines
  - Red = limit for “Sufficient” water quality (500 cfu/100ml)
  - Orange = limit for “Good” water quality (250 cfu/100ml)

## E-Coli ~ Day-1 Precipitation



## Bi-Variate Linear Regression

- This matrix shows the R-Squared values of E-Coli ~ Rainfall
  - ec\_rain = same-day
  - ec\_rain1 = day-1
  - ec\_rain2 = day-2
  - ec\_rain3 = day-3
  - ec\_rain\_sum = 4-day sum
- Each row shows the results from a different Met EIERann gauge

```
##          ec_rain ec_rain1 ec_rain2 ec_rain3 ec_rain_sum
## dly3135  7e-04   0.06   3e-02   1e-01   0.08
## dly3436  2e-03   0.05   3e-02   5e-01   0.17
## dly3735  3e-03   0.28   3e-03   1e-04   0.05
## dly4235  5e-04   0.10   2e-02   7e-03   0.07
## dly4635  2e-04   0.22   6e-03   2e-03   0.07
## dly5735  1e-03   0.15   1e-05   8e-02   0.05
```

- From this test, there were also no strong relationships.
  - Highest value is 0.28 from the day-1 rainfall of gauge #3735

```
##          ec_log_rain ec_log_rain1 ec_log_rain2 ec_log_rain3 ec_log_rain_sum
## dly3135      0.003      0.12      0.066      5e-02      0.11
```



## dly3436	0.008	0.10	0.044	1e-01	0.15
## dly3735	0.003	0.08	0.002	1e-02	0.04
## dly4235	0.011	0.04	0.074	9e-03	0.09
## dly4635	0.001	0.08	0.026	1e-06	0.06
## dly5735	0.009	0.09	0.004	2e-02	0.06