

# Compare Results

Audrey Smith

6/4/2021

## Prepare Data

```
params_aq <- read.csv("../aeroqual/2021-06/pse_calvals_2021-06.csv")  
  
param_paths_pse_all_03 <- list.files('results', 'OZONE*', full.names = T)  
param_path_pse_current_03 <- param_paths_pse_all_03[length(param_paths_pse_all_03)]  
params_pse_03 <- read.csv(param_path_pse_current_03)  
  
param_paths_pse_all_NO2 <- list.files('results', 'NO2*', full.names = T)  
param_path_pse_current_NO2 <- param_paths_pse_all_NO2[length(param_paths_pse_all_NO2)]  
params_pse_NO2 <- read.csv(param_path_pse_current_NO2)  
  
raw_data <- read.csv('raw_data/raw_60_min.csv') %>%  
  filter(ID %in% c('AQY BB-633', 'AQY BB-642'))
```

```
proxy_03 <- get_current_proxy_data('OZONE')
```

```
## [1] "Reading existing proxy data from results/proxy/OZONE_proxy_data.csv"  
## [1] "Requesting data starting with 2021-07-01 11:00:00"  
## [1] "Proxy data for OZONE now current - writing to results/proxy"
```

```
proxy_NO2 <- get_current_proxy_data('NO2')
```

```
## [1] "Reading existing proxy data from results/proxy/NO2_proxy_data.csv"  
## [1] "Requesting data starting with 2021-07-01 10:00:00"  
## [1] "Requesting proxy data from Air District for NO2 from 2021-07-01 10 to 2021-07-01 11"  
## [1] "Proxy data for NO2 now current - writing to results/proxy"
```

```

03_data_aq <- inner_join(params_aq, raw_data, by = 'ID') %>%
  filter(ID %in% c('AQY BB-633', 'AQY BB-642')) %>%
  filter(Time >= deployment_date & Time >= start_date & Time <= end_date & Time < 2021) %>%
  transmute(ID=ID, TIME=Time, O3.RAW = O3_raw,
            O3.AQ = round(O3.gain*(O3_raw - O3.offset), 2),
            O3.GAIN.AQ=O3.gain, O3.OFFSET.AQ=O3.offset)

03_data_pse <- inner_join(params_pse_03, raw_data, by = 'ID') %>%
  filter(ID %in% c('AQY BB-633', 'AQY BB-642')) %>%
  filter(Time >= deployment_date & Time >= start_date & Time <= end_date & Time < 2021) %>%
  transmute(ID=ID, TIME=Time,
            O3.PSE = round(O3.offset + O3.gain*O3_raw, 2),
            O3.GAIN.PSE=O3.gain, O3.OFFSET.PSE=O3.offset)

03_data <- inner_join(03_data_aq, 03_data_pse, by = c('ID', 'TIME')) %>%
  inner_join(., proxy_03, by = c('TIME'='timestamp_pacific')) %>%
  mutate(TIMESTAMP = ymd_hms(TIME), O3.PROXY = round(proxy_rand, 2)) %>%
  dplyr::select(ID, TIME, TIMESTAMP, O3.RAW, O3.PROXY, O3.AQ, O3.PSE, O3.GAIN.AQ, O3.GAIN.PSE, O3.OFFSET.AQ, O3.OFFSET.PSE) %>%
  filter(O3.GAIN.PSE != 1)

```

```

NO2_data_aq <- inner_join(params_aq, raw_data, by = 'ID') %>%
  filter(ID %in% c('AQY BB-633', 'AQY BB-642')) %>%
  filter(Time >= deployment_date & Time >= start_date & Time <= end_date & Time < 2021) %>%
  mutate(O3.AQ = O3.gain*(O3_raw - O3.offset), OX.AQ = Ox.gain*(Ox_raw - Ox.offset), NO2.INT = OX.AQ-NO2.a.value*O3.AQ) %>%
  transmute(ID=ID, TIME=Time, NO2.GAIN.AQ=NO2.gain, NO2.OFFSET.AQ=NO2.offset, NO2.AQ = round(NO2.gain*(NO2.INT - NO2.offset), 2))

NO2_data_pse <- inner_join(params_pse_03, raw_data, by = 'ID') %>%
  filter(ID %in% c('AQY BB-633', 'AQY BB-642')) %>%
  filter(Time >= deployment_date & Time >= start_date & Time <= end_date & Time < 2021) %>%
  transmute(ID=ID, TIME=Time, NO2.RAW = NO2_raw, Ox.RAW = Ox_raw, O3.PSE = round(O3.offset + O3.gain*O3_raw, 2)) %>%
  inner_join(params_pse_NO2, by = 'ID') %>%
  filter(TIME >= deployment_date & TIME >= start_date & TIME <= end_date & TIME < 2021) %>%
  transmute(ID=ID, TIME=TIME, NO2.RAW=NO2.RAW,
            NO2.PSE = round(NO2.offset + NO2.gain.Ox*Ox.RAW - NO2.gain.O3*O3.PSE, 2),
            NO2.GAIN.PSE=NO2.gain.Ox, NO2.OFFSET.PSE=NO2.offset)

NO2_data <- inner_join(NO2_data_pse, proxy_NO2, by = c('TIME'='timestamp_pacific')) %>%
  inner_join(NO2_data_aq, by = c('ID', 'TIME')) %>%
  mutate(TIMESTAMP = ymd_hms(TIME), NO2.PROXY = round(proxy_rand, 2)) %>%
  dplyr::select(ID, TIME, TIMESTAMP, NO2.RAW, NO2.PROXY, NO2.PSE, NO2.AQ, NO2.GAIN.PSE, NO2.GAIN.AQ, NO2.OFFSET.PSE, NO2.OFFSET.AQ) %>%
  filter(NO2.GAIN.PSE != 1)

```

## Write Functions to Examine Performance

```
get_timeseries <- function(input_data, aqy_id, month.i, pollutant){

  timeseries_plot <- ggplot(input_data) +
    geom_line(aes(x = TIMESTAMP, y = POL.RAW, color = 'raw')) +
    geom_line(aes(x = TIMESTAMP, y = POL.PROXY, color = 'proxy')) +
    geom_line(aes(x = TIMESTAMP, y = POL.AQ, color = 'aeroqual')) +
    geom_line(aes(x = TIMESTAMP, y = POL.PSE, color = 'pse')) +
    labs(title = paste(pollutant, 'Timeseries', '-', aqy_id, 'Month of', month.i),
        x = 'Timestamp', y = paste('Pollutant Concentration (ppb)')) +
    theme(plot.title = element_text(hjust = .5))

  print(timeseries_plot)

}
```

```
get_error_timeseries <- function(input_data, aqy_id, month.i, pollutant){

  error_timeseries <- input_data %>%
    mutate(AE.RAW = POL.RAW - POL.PROXY, AE.AQ = POL.AQ - POL.PROXY,
          AE.PSE = POL.PSE - POL.PROXY) %>%
    ggplot() +
    geom_line(aes(x = TIMESTAMP, y = AE.RAW, color = 'raw')) +
    geom_line(aes(x = TIMESTAMP, y = AE.AQ, color = 'aeroqual')) +
    geom_line(aes(x = TIMESTAMP, y = AE.PSE, color = 'pse')) +
    labs(title = paste('Absolute Error for', pollutant, '-', aqy_id, 'month of', month.i),
        x = 'Timestamp', y = 'Absolute Error (ppb)') +
    theme(plot.title = element_text(hjust = .5))

  print(error_timeseries)

}
```

```
get_scatterplot <- function(input_data, aqy_id, month.i, pollutant){

  scatterplot <- ggplot(input_data) +
    geom_line(aes(x = POL.PROXY, y = POL.PROXY, color = '1-to-1 fit')) +
    geom_point(aes(x = POL.PROXY, y = POL.RAW, color = 'raw'), alpha = .5) +
    geom_point(aes(x = POL.PROXY, y = POL.PSE, color = 'pse'), alpha = .5) +
    geom_point(aes(x = POL.PROXY, y = POL.AQ, color = 'aeroqual'), alpha = .5) +
    labs(title = paste(pollutant, '[Proxy] vs [Measured] for', aqy_id, 'month of', month.i),
        x = 'Timestamp', y = 'Pollutant Concentration - (ppb)') +
    theme(plot.title = element_text(hjust = .5))

  print(scatterplot)

}
```

```
get_lms <- function(input_data, aqy_id, month.i){

  model_raw <- summary(lm('POL.PROXY~POL.RAW', data = input_data))
  model_pse <- summary(lm('POL.PROXY~POL.PSE', data = input_data))
  model_aq <- summary(lm('POL.PROXY~POL.AQ', data = input_data))

  r_sq <- data.frame(cbind(month.i, round(model_raw$r.squared, 5), round(model_pse$r.squared, 5),
  ), round(model_aq$r.squared, 5)))
  colnames(r_sq) <- c('month', 'raw', 'PSE', 'AQ')

  return(r_sq)
}
```

```
get_offset_timeseries <- function(input_data, aqy_id, pollutant){

  offset_timeseries <- ggplot(input_data) +
    geom_line(aes(x = TIMESTAMP, y = POL.OFFSET.PSE, color = 'pse')) +
    geom_line(aes(x = TIMESTAMP, y = POL.OFFSET.AQ, color = 'aeroqual')) +
    ggtitle(paste(pollutant, 'Offset Over Time for', aqy_id)) +
    theme(plot.title = element_text(hjust = .5))

  print(offset_timeseries)
}
```

```
get_gain_timeseries <- function(input_data, aqy_id, pollutant){

  gain_timeseries <- ggplot(input_data) +
    geom_line(aes(x = TIMESTAMP, y = POL.GAIN.PSE, color = 'pse')) +
    geom_line(aes(x = TIMESTAMP, y = POL.GAIN.AQ, color = 'aeroqual')) +
    ggtitle(paste(pollutant, 'Gain Over Time for', aqy_id)) +
    theme(plot.title = element_text(hjust = .5))

  print(gain_timeseries)
}
```

```
get_r2_plot <- function(input_data, aqy_id, pollutant){

  r2_plot <- ggplot(input_data) +
    geom_line(aes(x = as.numeric(month), y = as.numeric(AQ), color = 'aeroqual')) +
    geom_line(aes(x = as.numeric(month), y = as.numeric(PSE), color = 'pse')) +
    labs(x = 'Month', y = 'R^2',
         title = paste(pollutant, 'Proxy ~ Measured R^2 Over Time for', aqy_id)) +
    theme(plot.title = element_text(hjust = .5))

  print(r2_plot)
}
```

## Run Performance Tests

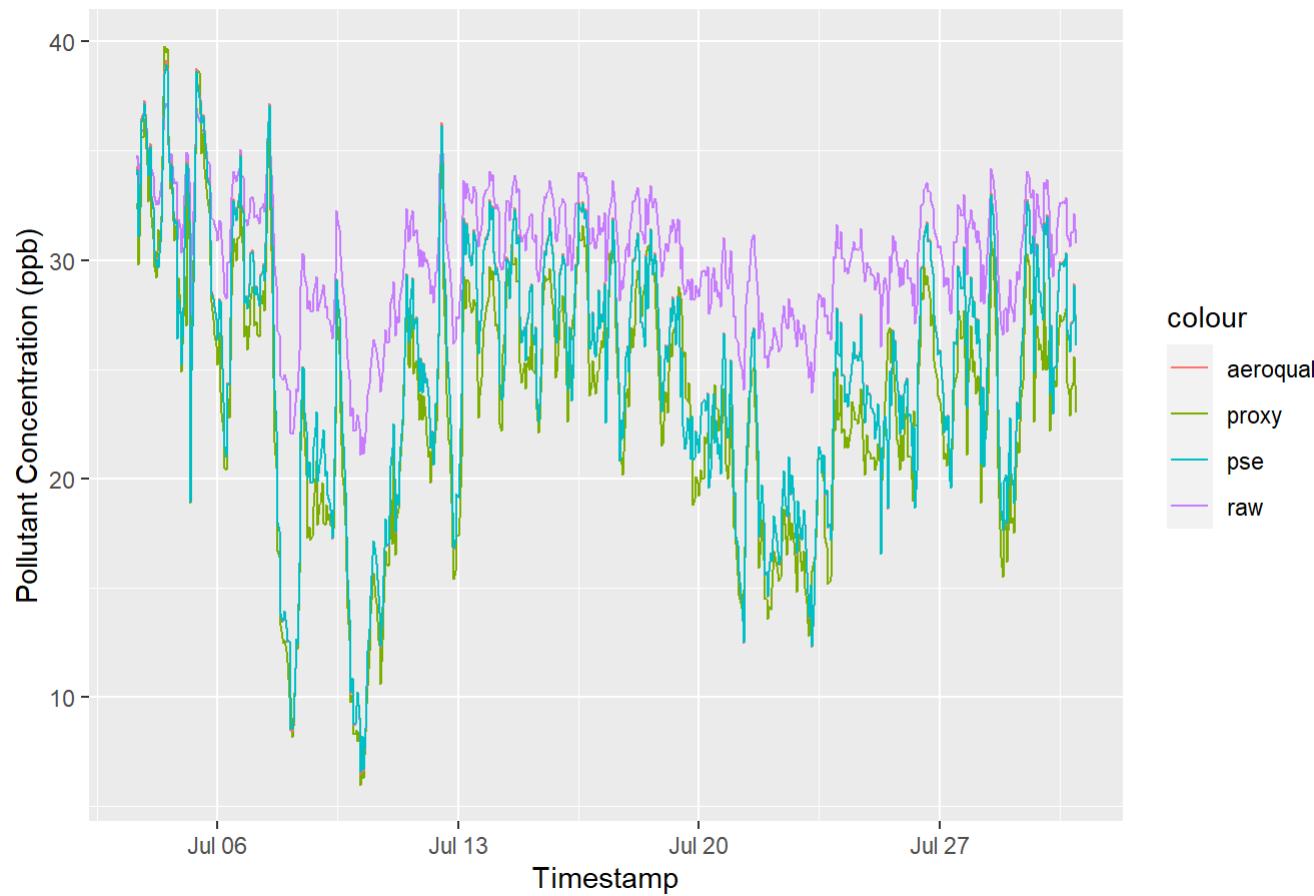
```
months <- c('07', '08', '09', '10', '11', '12')
```

```
make_all_plots <- function(input_data, aqy_id, pollutant){  
  
  colnames(input_data) <- str_replace_all(colnames(input_data), pollutant, 'POL')  
  
  monthly_models <- data.frame()  
  
  for(month.i in months){  
  
    data_for_month <- filter(input_data, ID == aqy_id & TIME >= paste0('2020-', month.i, '-01 0  
0:00:00') & TIME <= paste0('2020-', month.i, '-31 23:59:59'))  
  
    get_timeseries(data_for_month, aqy_id, month.i, pollutant)  
  
    get_error_timeseries(data_for_month, aqy_id, month.i, pollutant)  
  
    get_scatterplot(data_for_month, aqy_id, month.i, pollutant)  
  
    month.i_model <- get_lms(data_for_month, aqy_id, month.i)  
    monthly_models <- rbind(monthly_models, month.i_model)  
  
  }  
  
  data_for_deployment <- filter(input_data, ID == aqy_id)  
  
  overall_model <- get_lms(data_for_deployment, aqy_id, 'Full Deployment')  
  all_models <- rbind(overall_model, monthly_models)  
  print(all_models)  
  
  get_r2_plot(monthly_models, aqy_id, pollutant)  
  
  get_offset_timeseries(data_for_deployment, aqy_id, pollutant)  
  
  get_gain_timeseries(data_for_deployment, aqy_id, pollutant)  
  
}
```

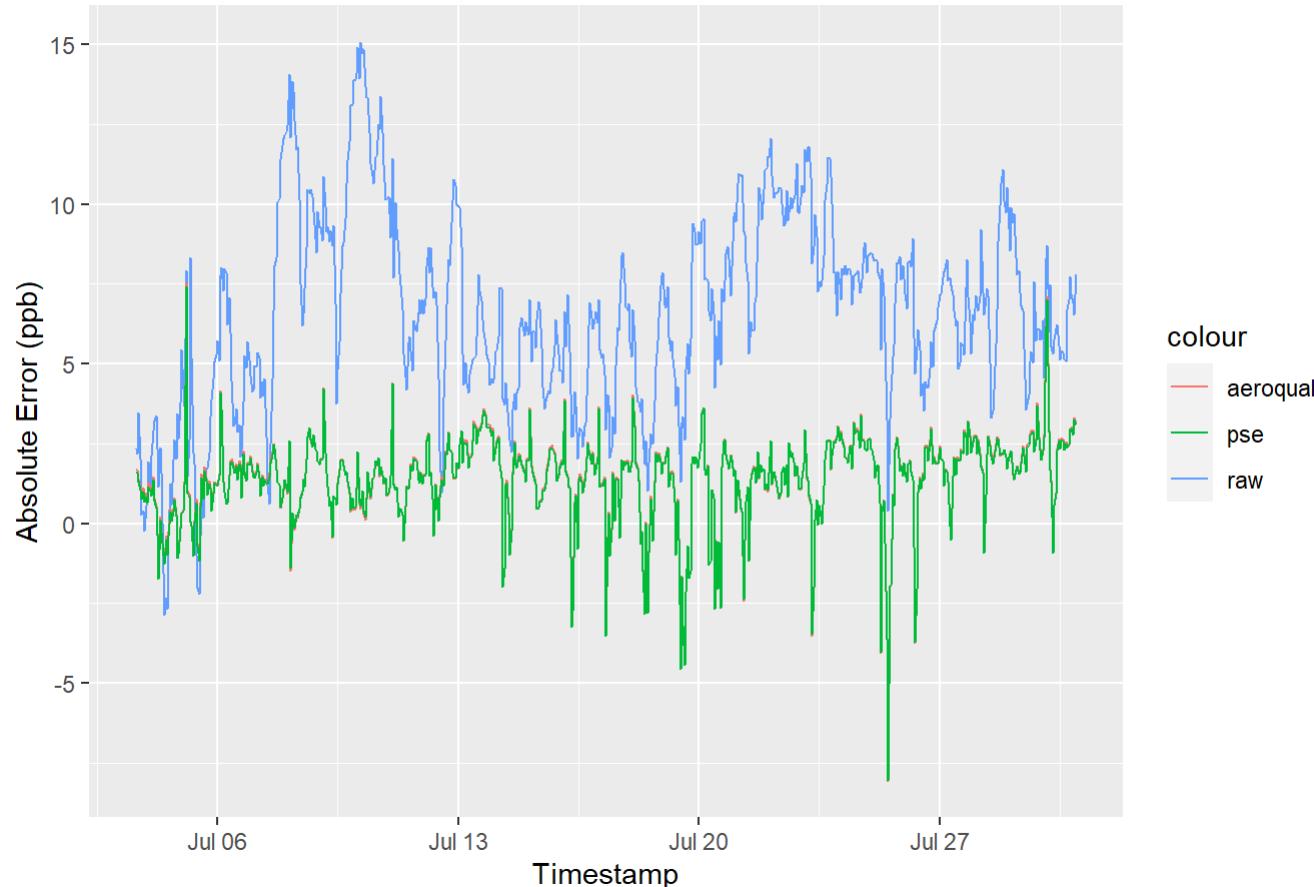
```
make_all_plots(O3_data, 'AQY BB-633', '03')
```



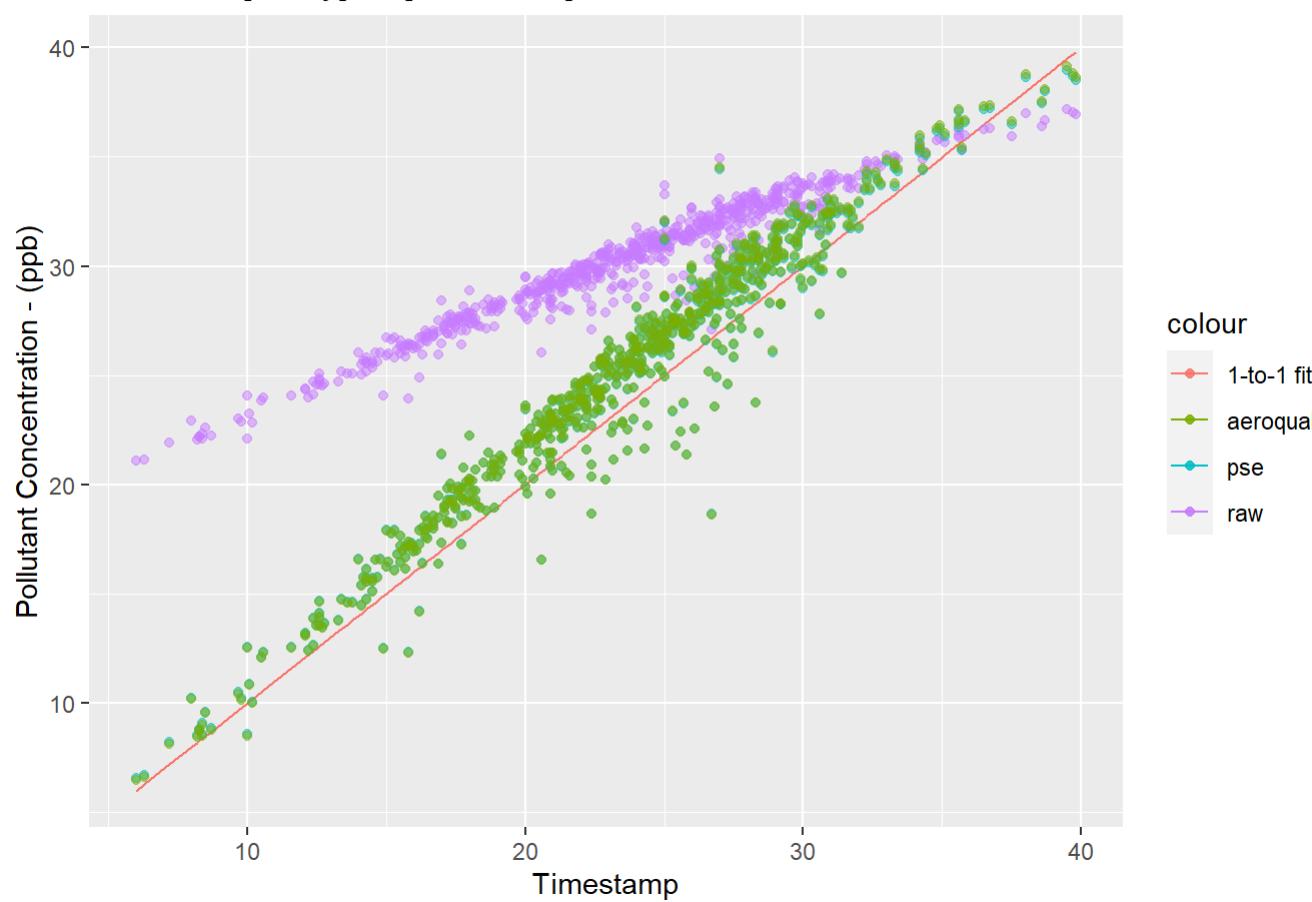
### O3 Timeseries - AQY BB-633 Month of 07



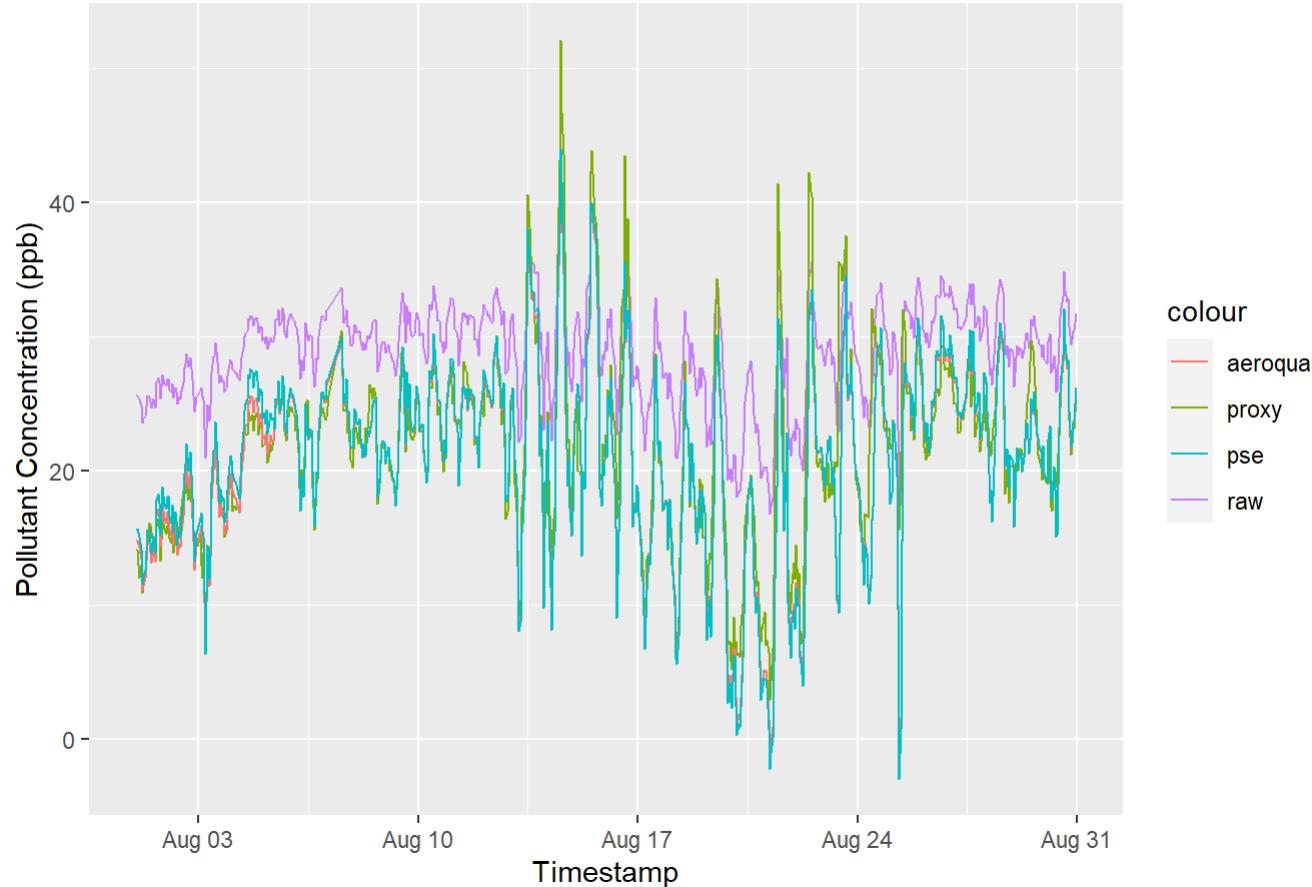
### Absolute Error for O3 - AQY BB-633 month of 07



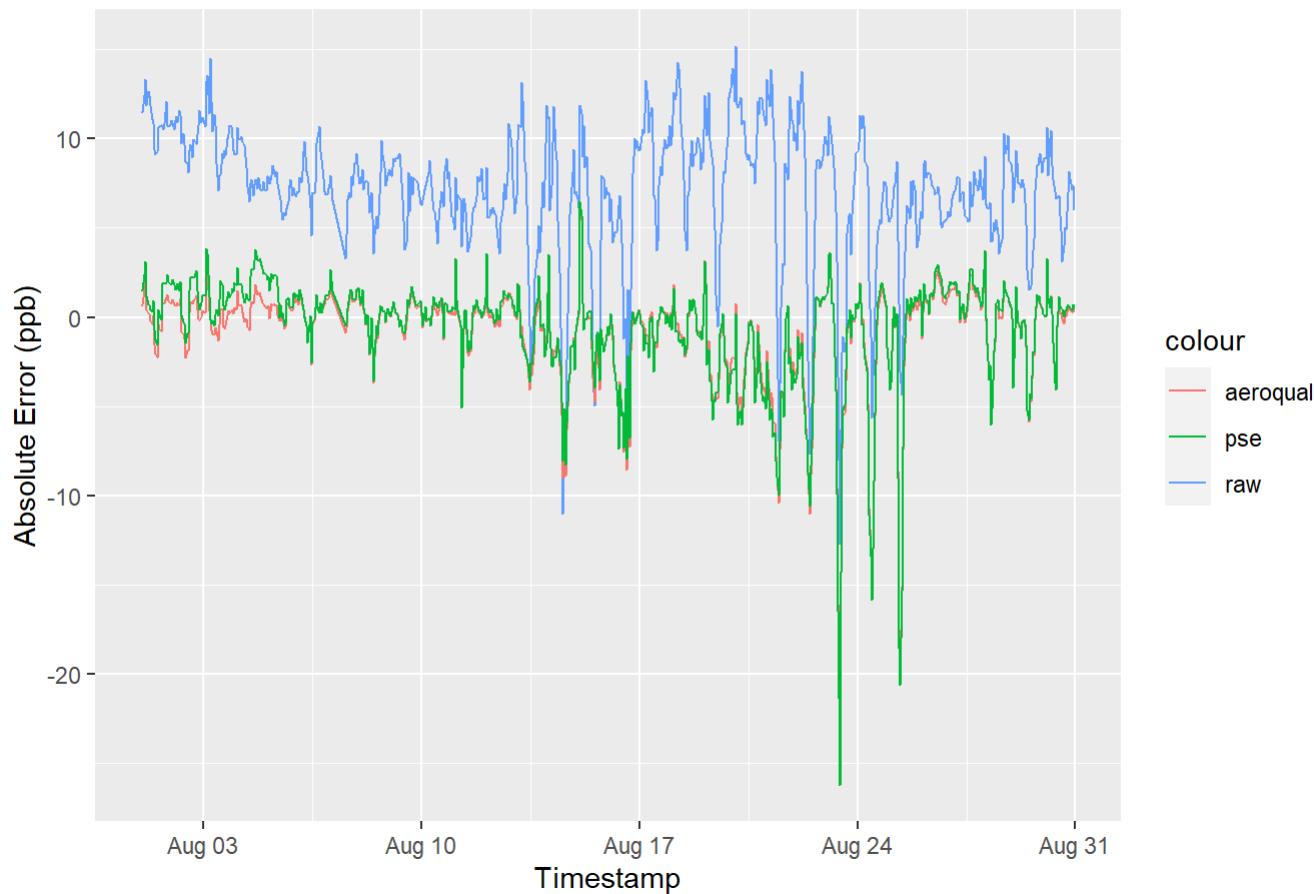
### O3 [Proxy] vs [Measured] for AQY BB-633 month of 07



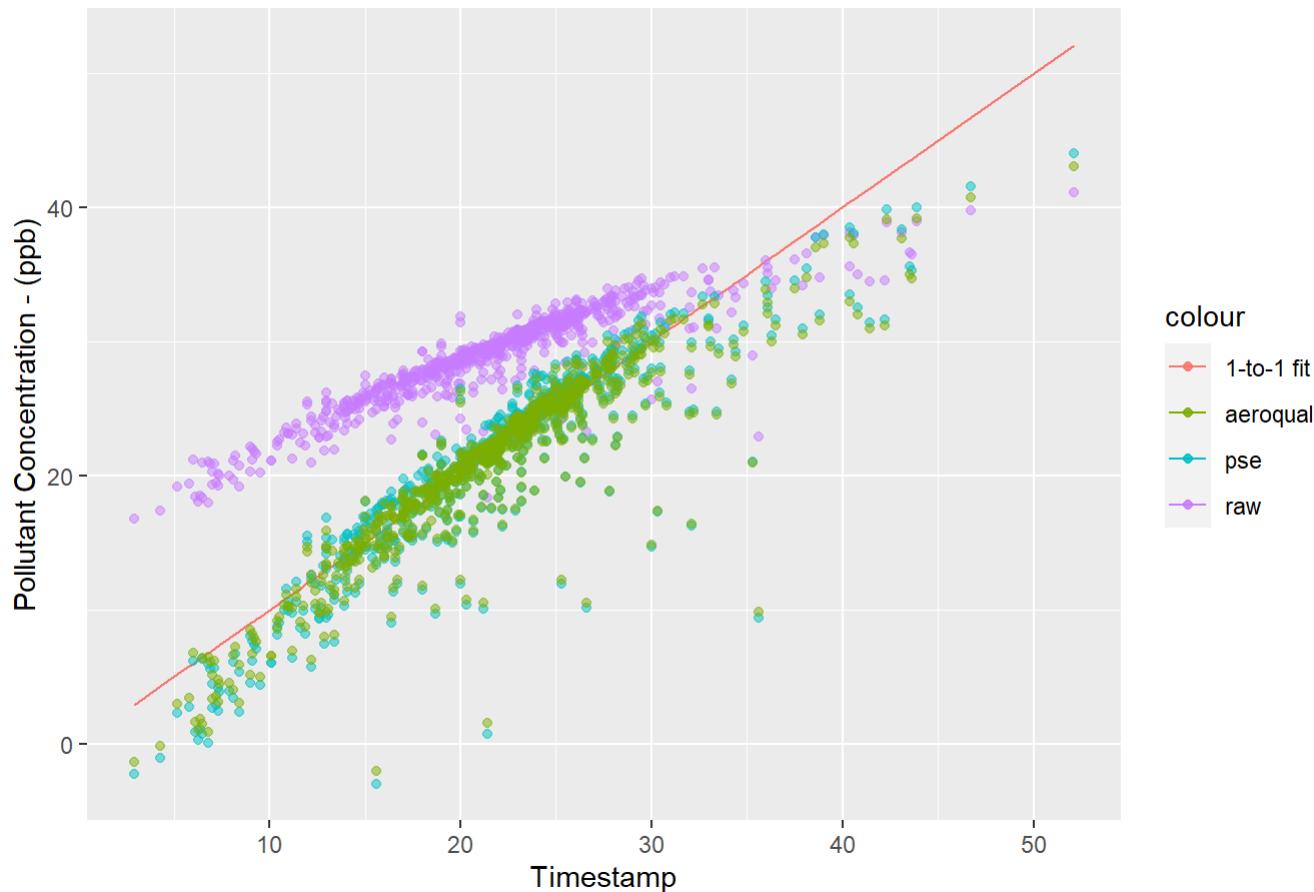
O3 Timeseries - AQY BB-633 Month of 08



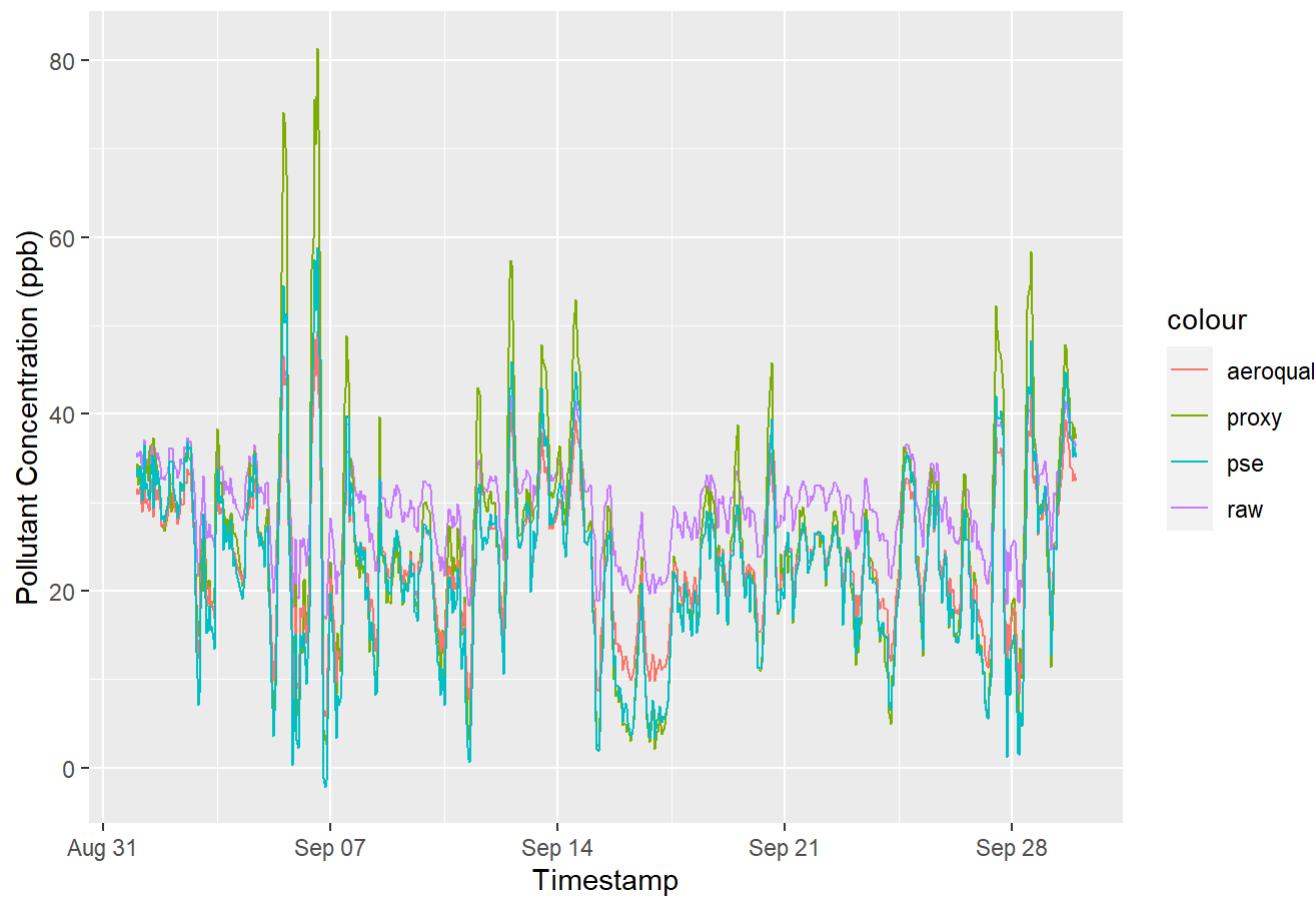
## Absolute Error for O3 - AQY BB-633 month of 08



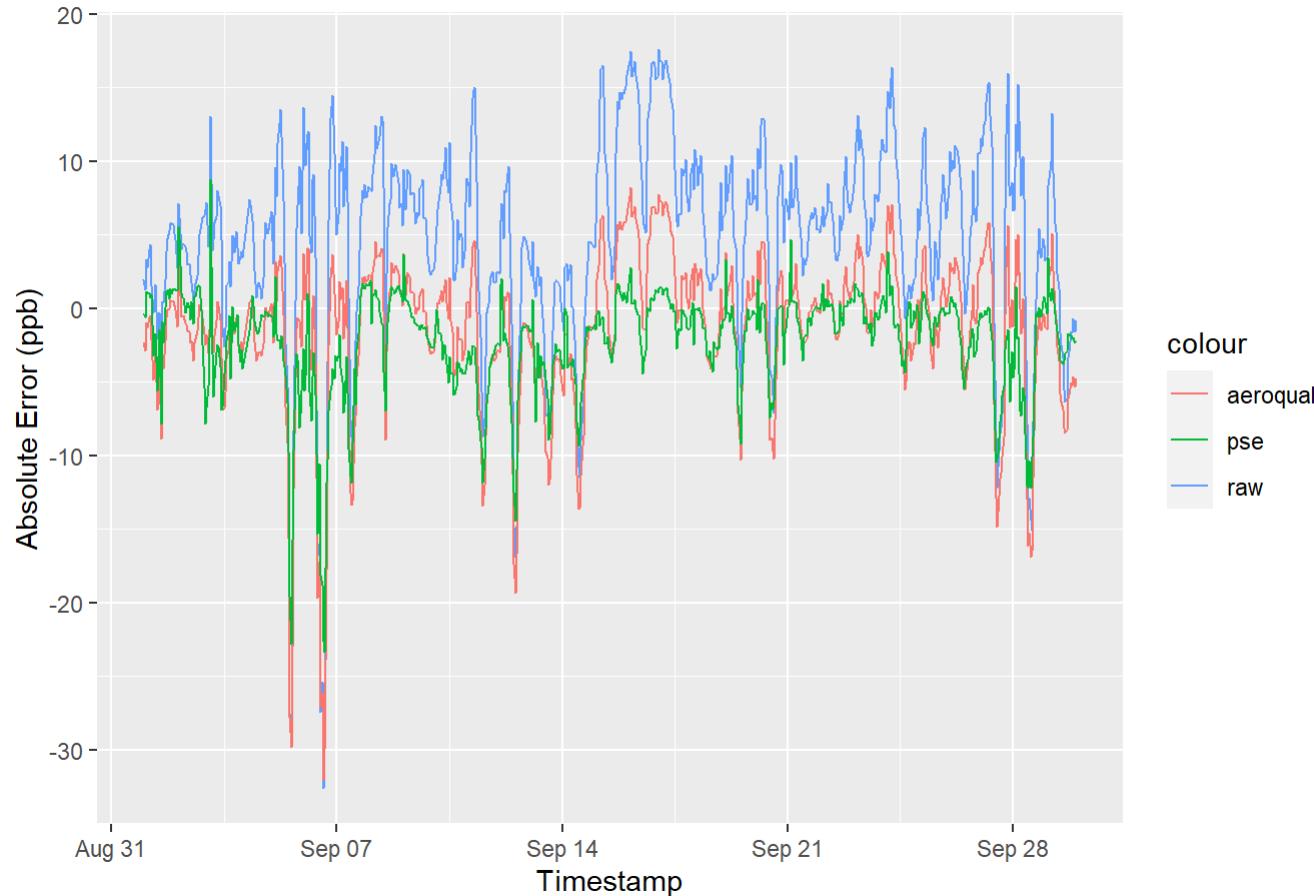
## O3 [Proxy] vs [Measured] for AQY BB-633 month of 08



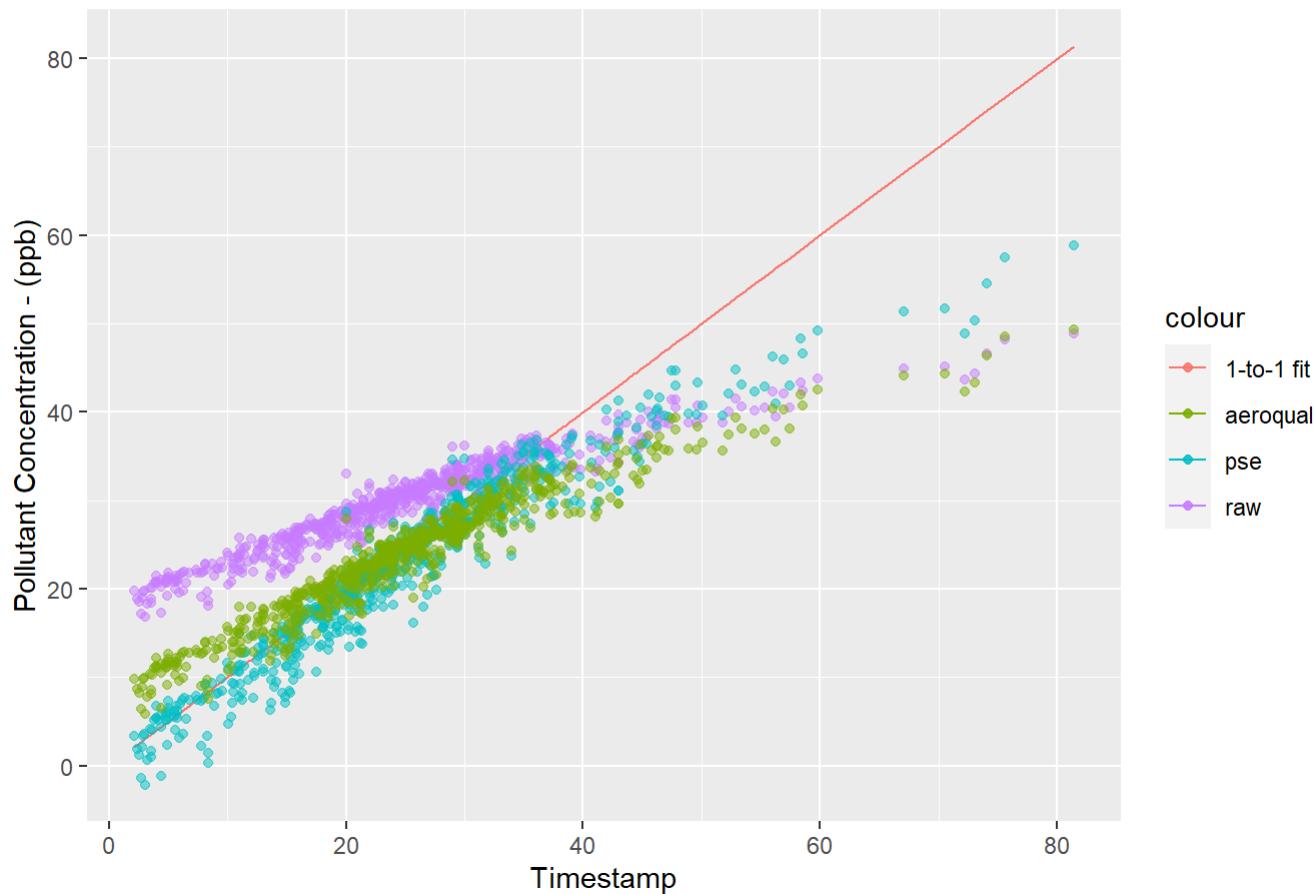
## O3 Timeseries - AQY BB-633 Month of 09



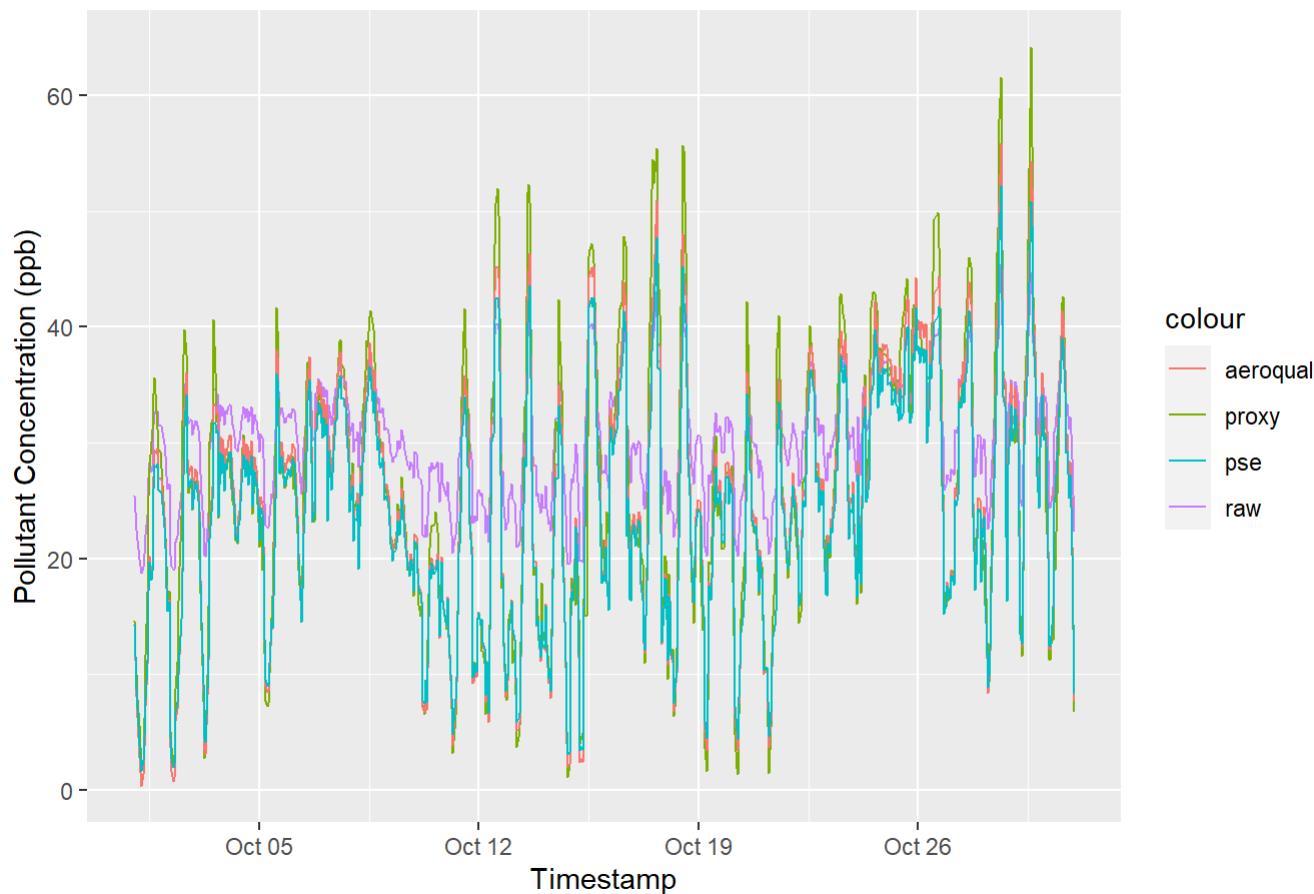
## Absolute Error for O3 - AQY BB-633 month of 09



## O3 [Proxy] vs [Measured] for AQY BB-633 month of 09



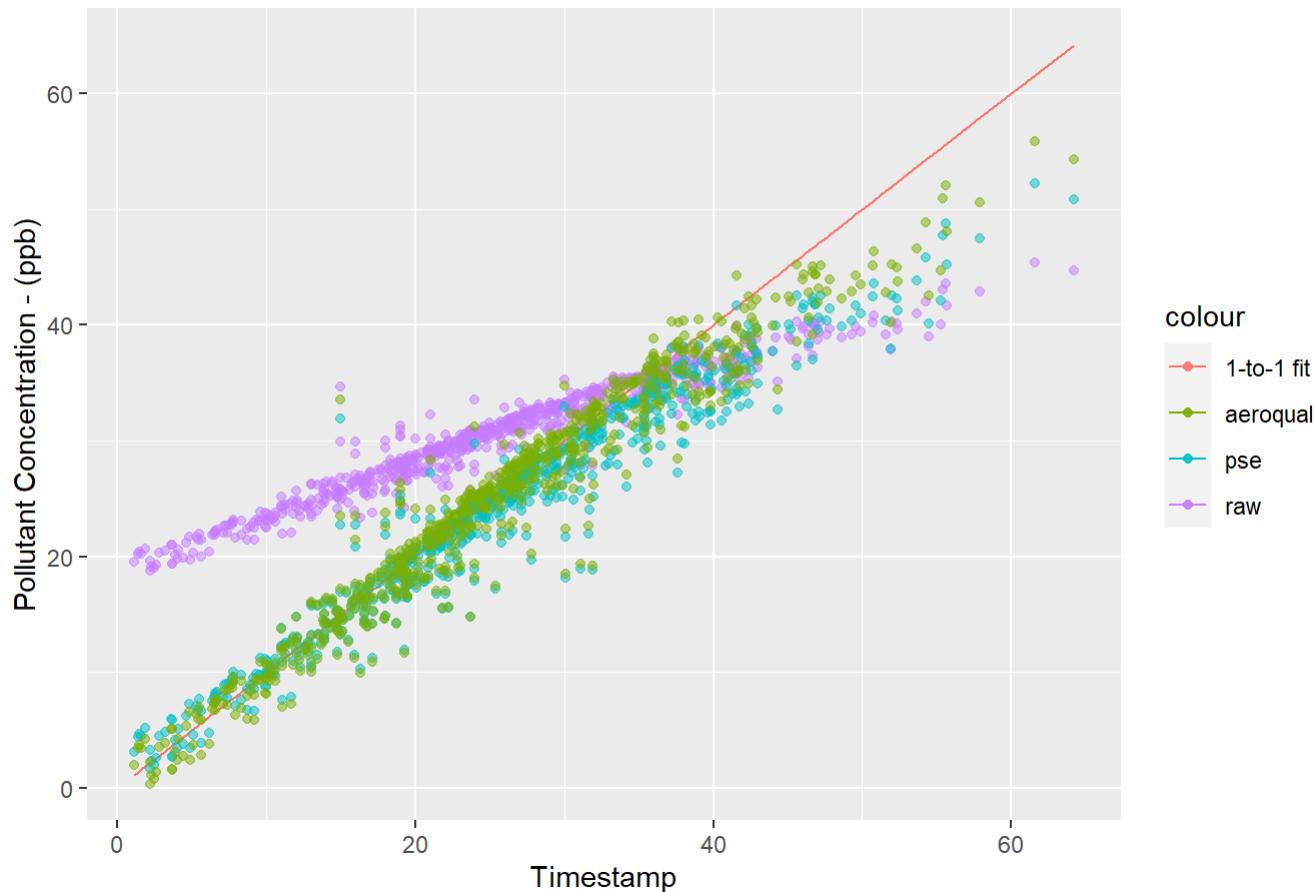
## O3 Timeseries - AQY BB-633 Month of 10



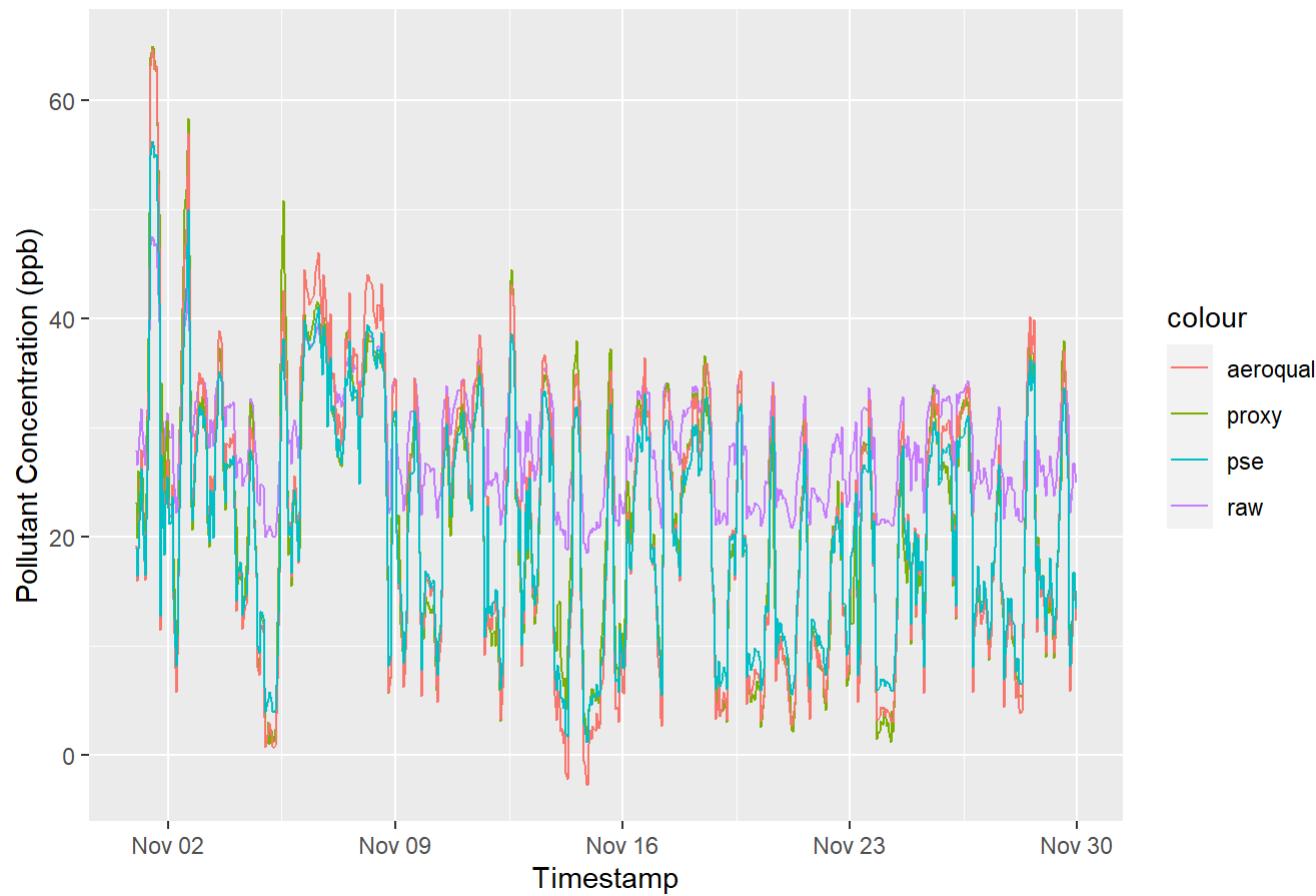
## Absolute Error for O3 - AQY BB-633 month of 10



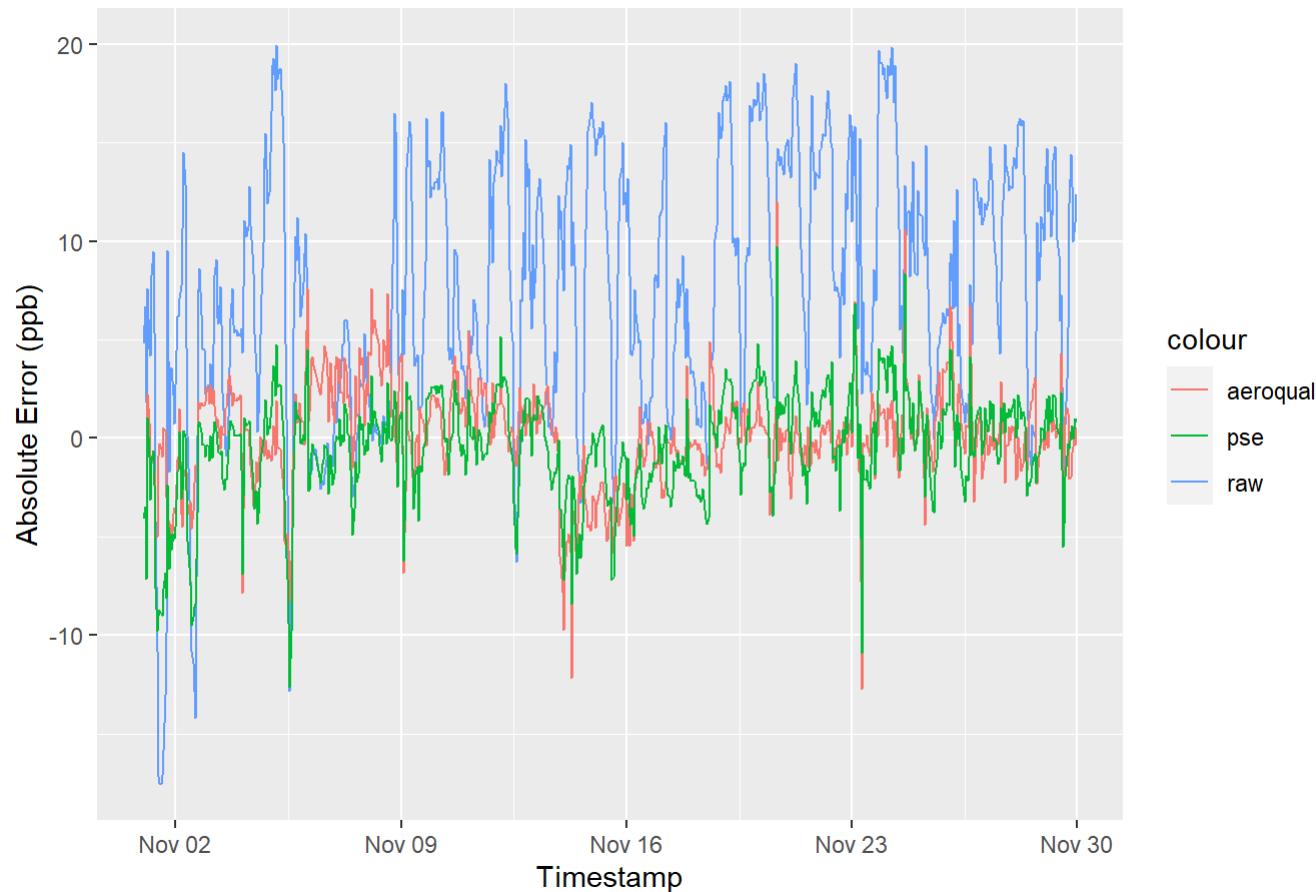
## O3 [Proxy] vs [Measured] for AQY BB-633 month of 10



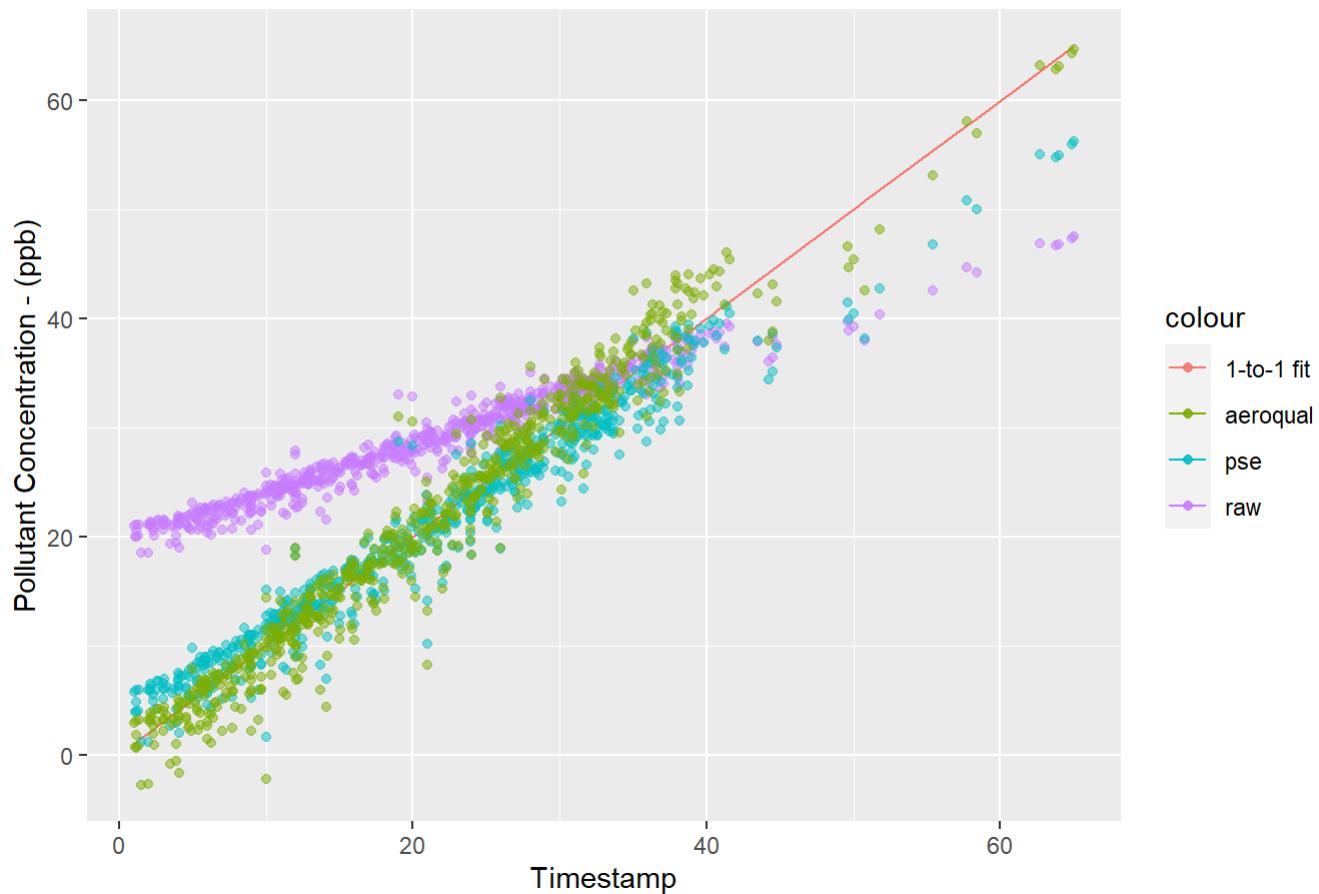
## O3 Timeseries - AQY BB-633 Month of 11



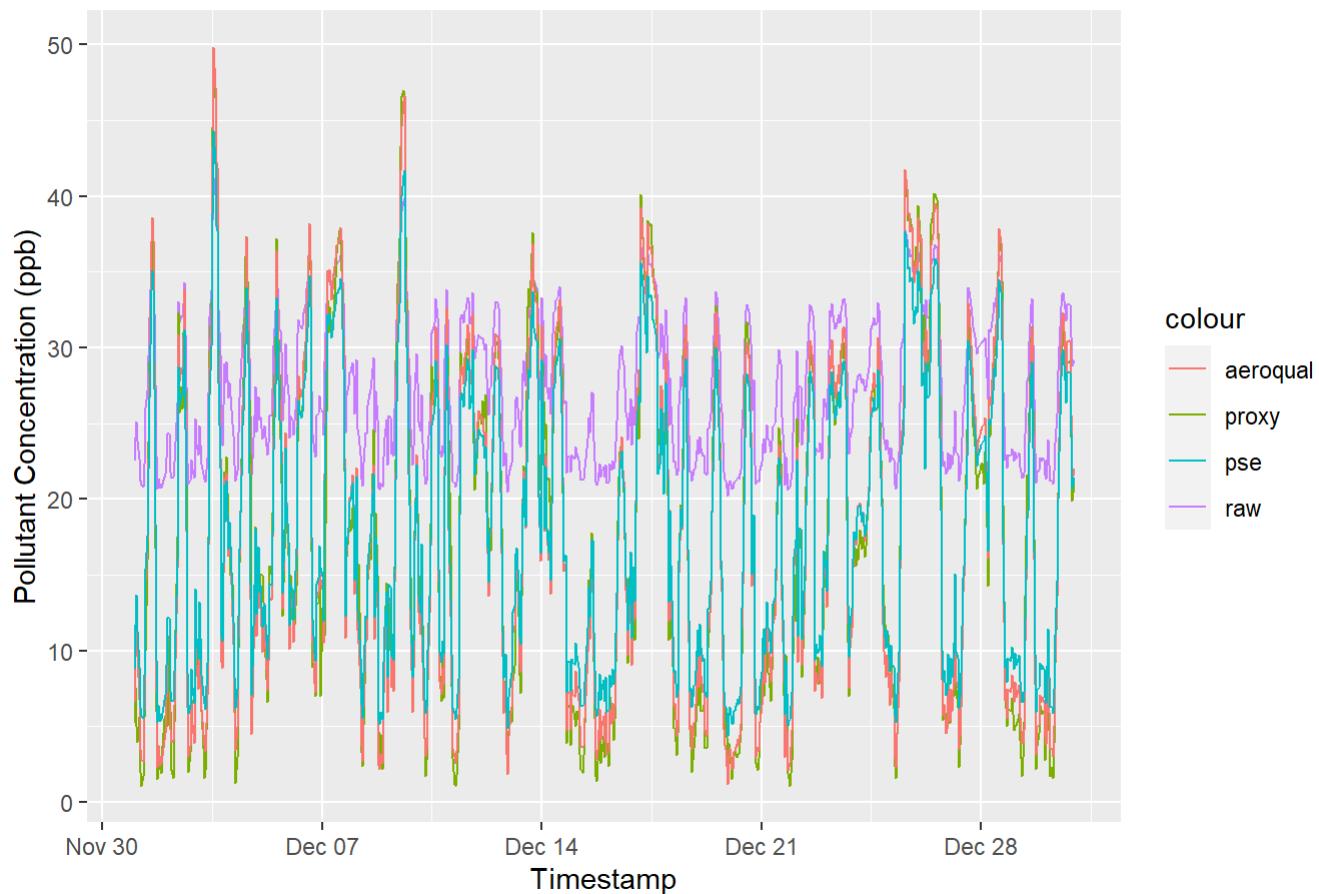
## Absolute Error for O3 - AQY BB-633 month of 11



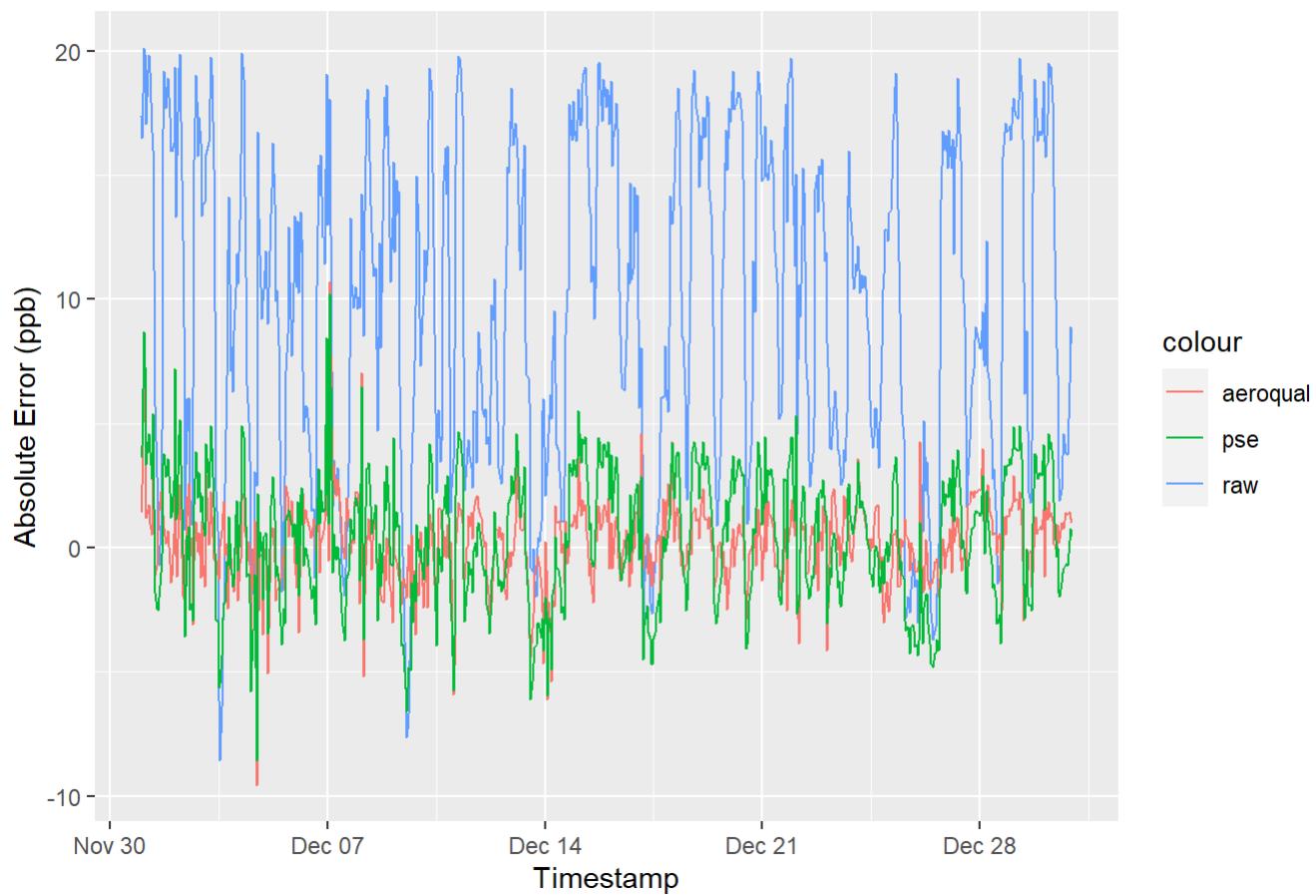
## O3 [Proxy] vs [Measured] for AQY BB-633 month of 11



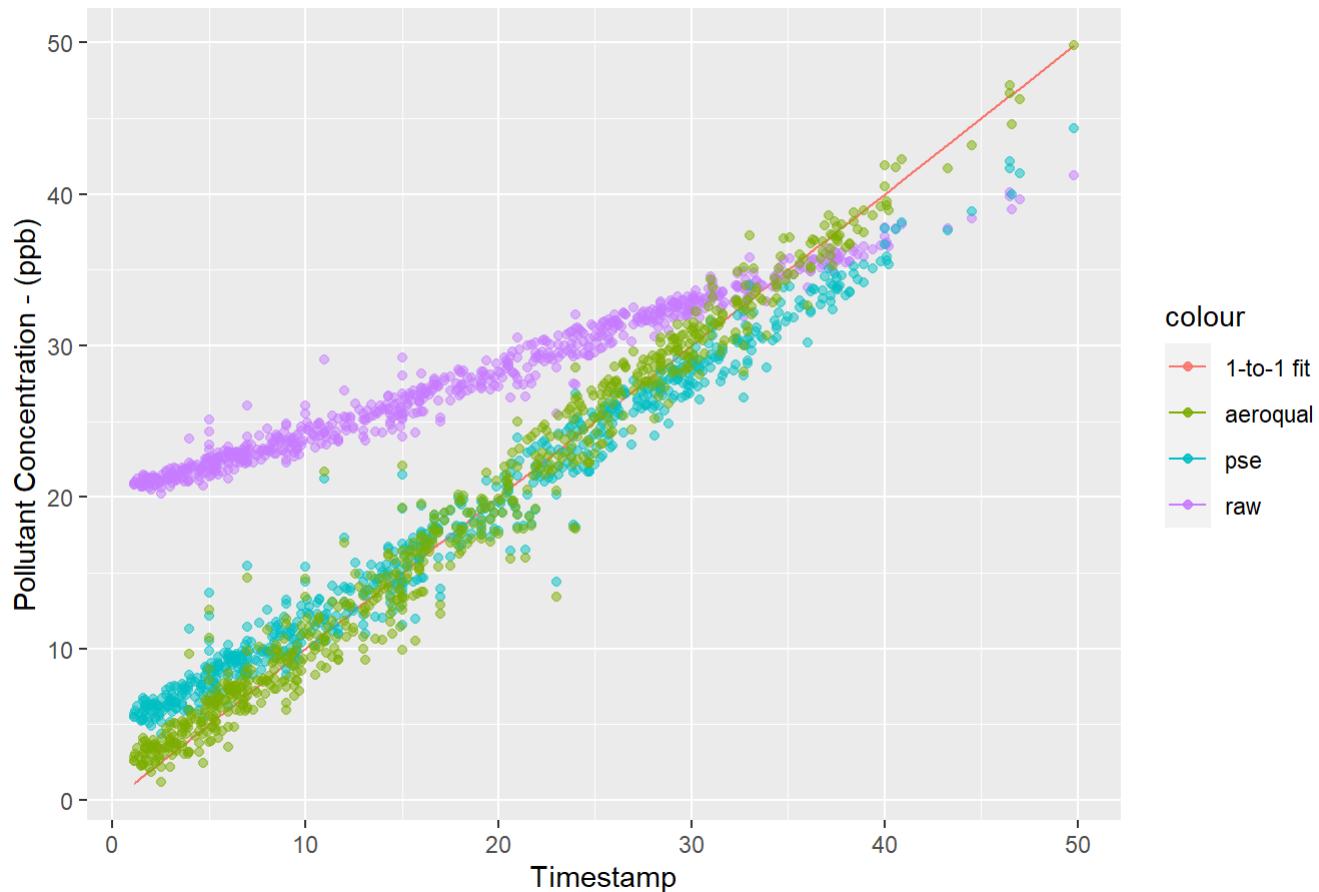
## O3 Timeseries - AQY BB-633 Month of 12



## Absolute Error for O3 - AQY BB-633 month of 12

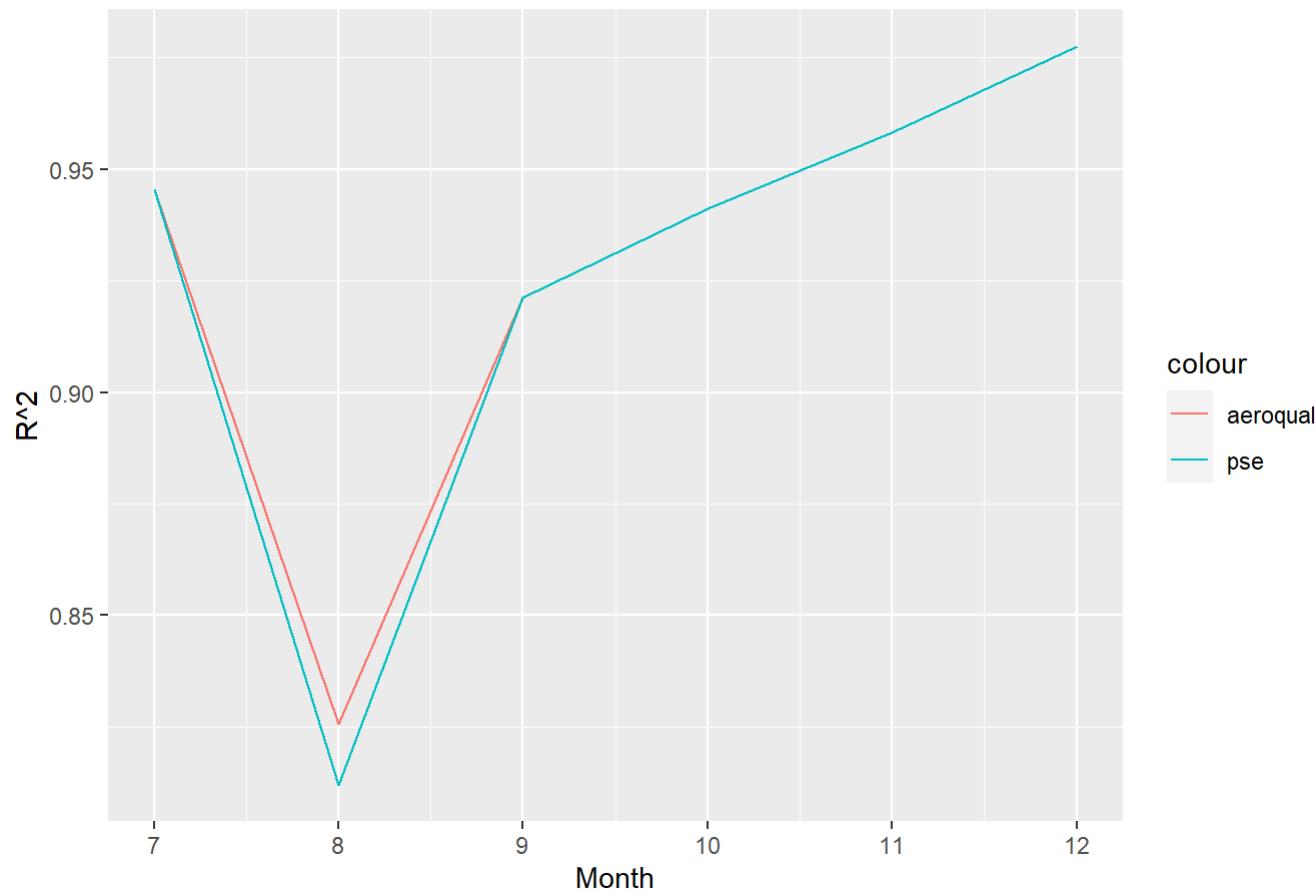


## O3 [Proxy] vs [Measured] for AQY BB-633 month of 12

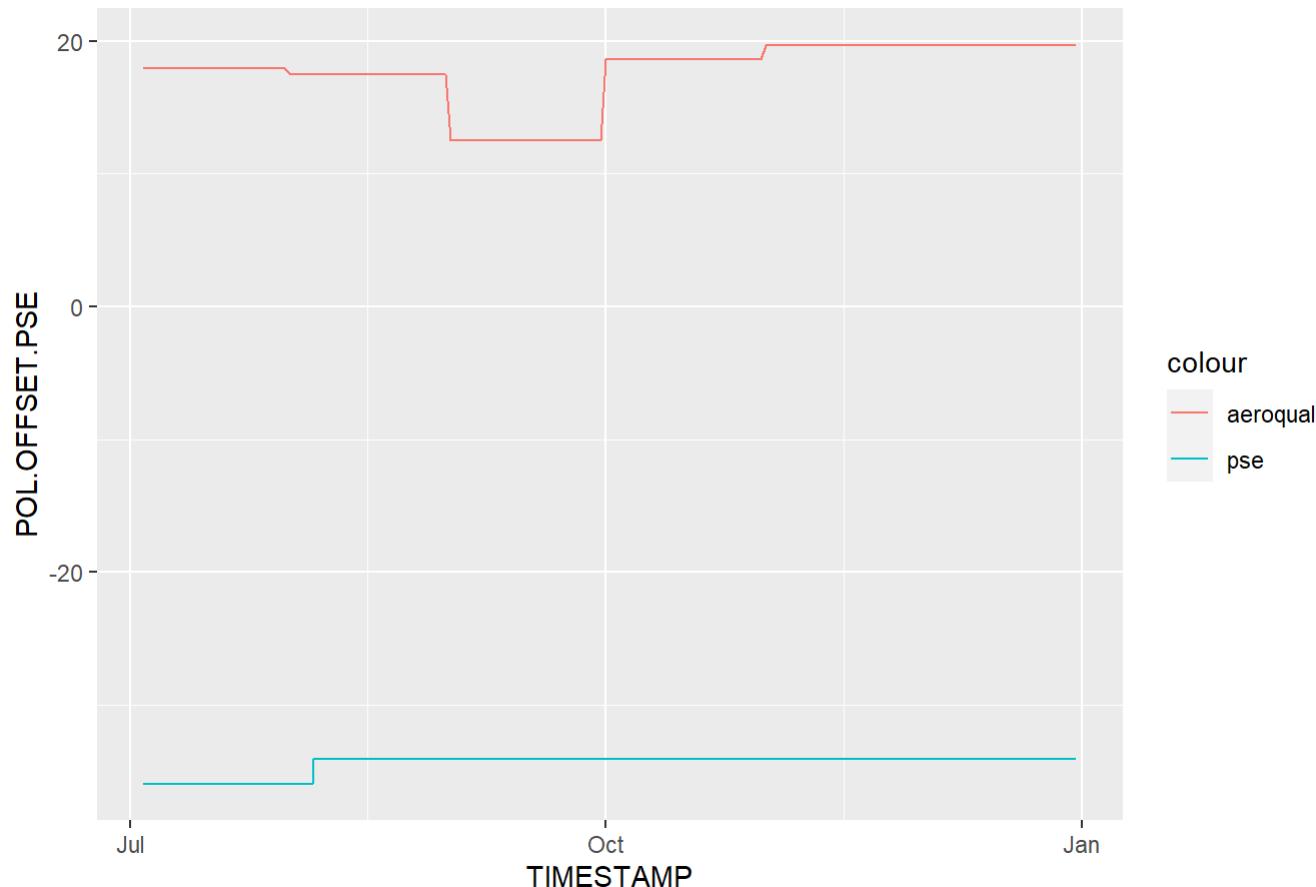


```
##          month     raw      PSE      AQ
## 1 Full Deployment 0.93177 0.92231 0.9088
## 2             07 0.94545 0.94546 0.94545
## 3             08 0.82566 0.81203 0.82563
## 4             09 0.92125 0.92125 0.92125
## 5             10 0.94121 0.94122 0.94121
## 6             11 0.95829 0.95829 0.95829
## 7             12 0.97763 0.97762 0.97763
```

### O3 Proxy ~ Measured R^2 Over Time for AQY BB-633

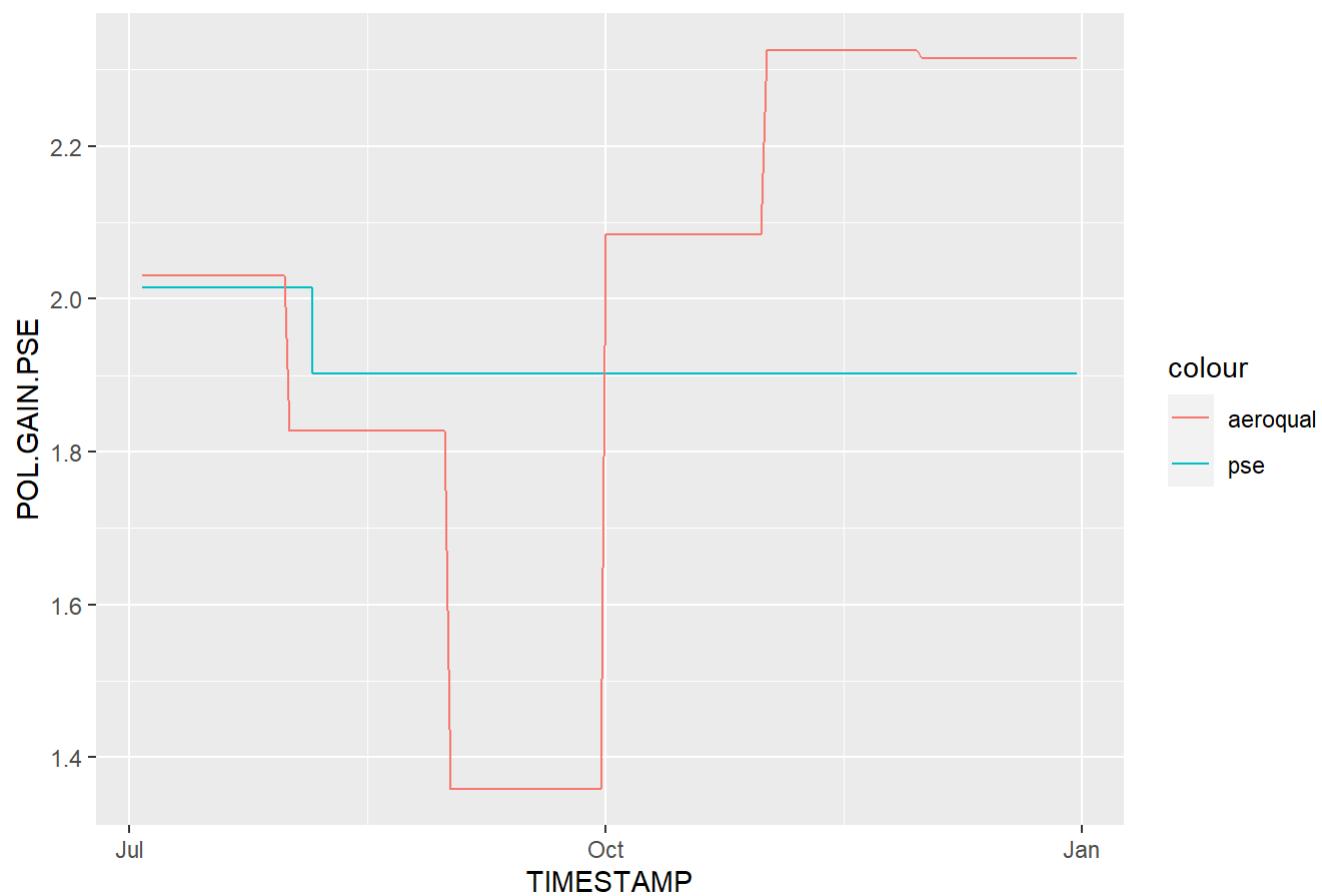


### O3 Offset Over Time for AQY BB-633



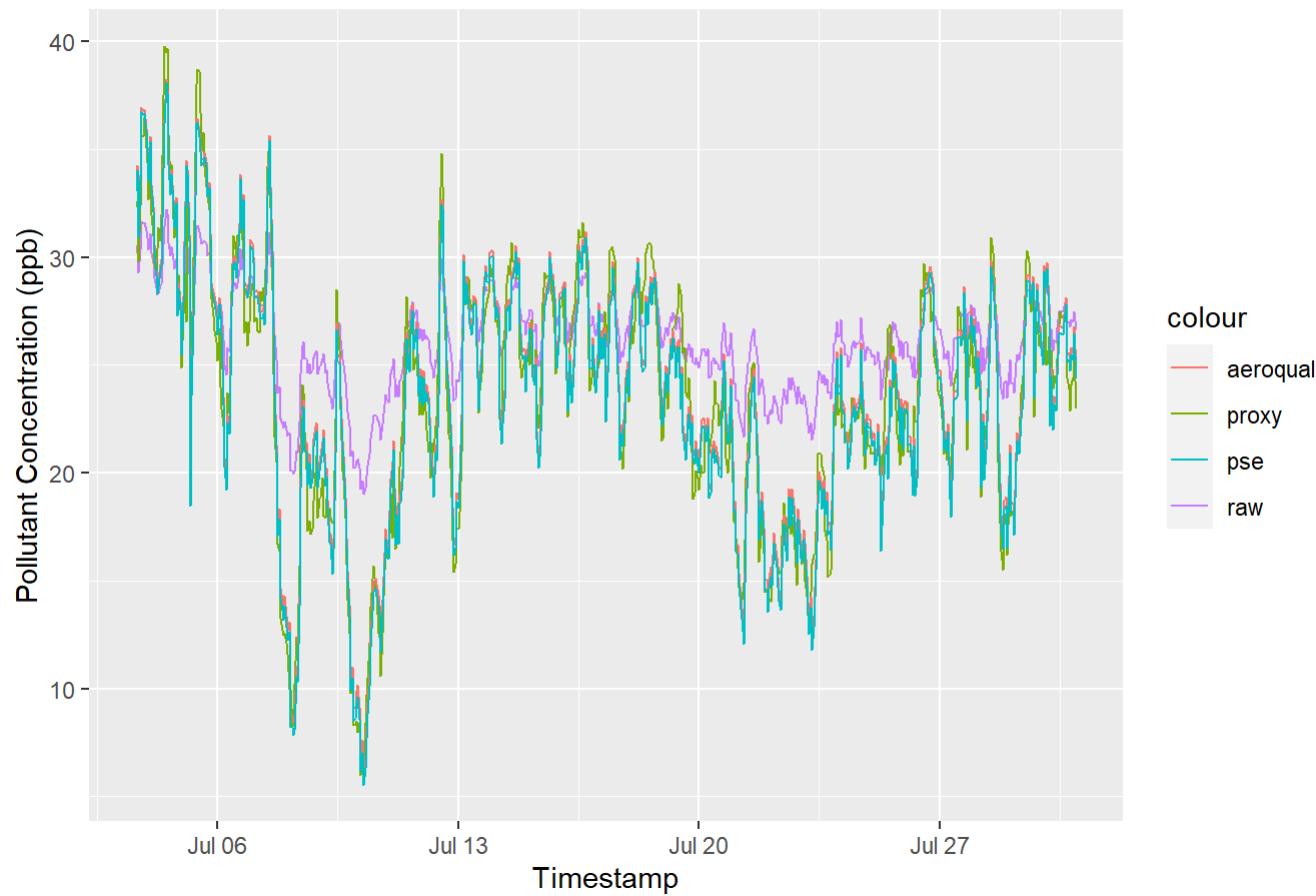
### O3 Gain Over Time for AQY BB-633

## Compare Results

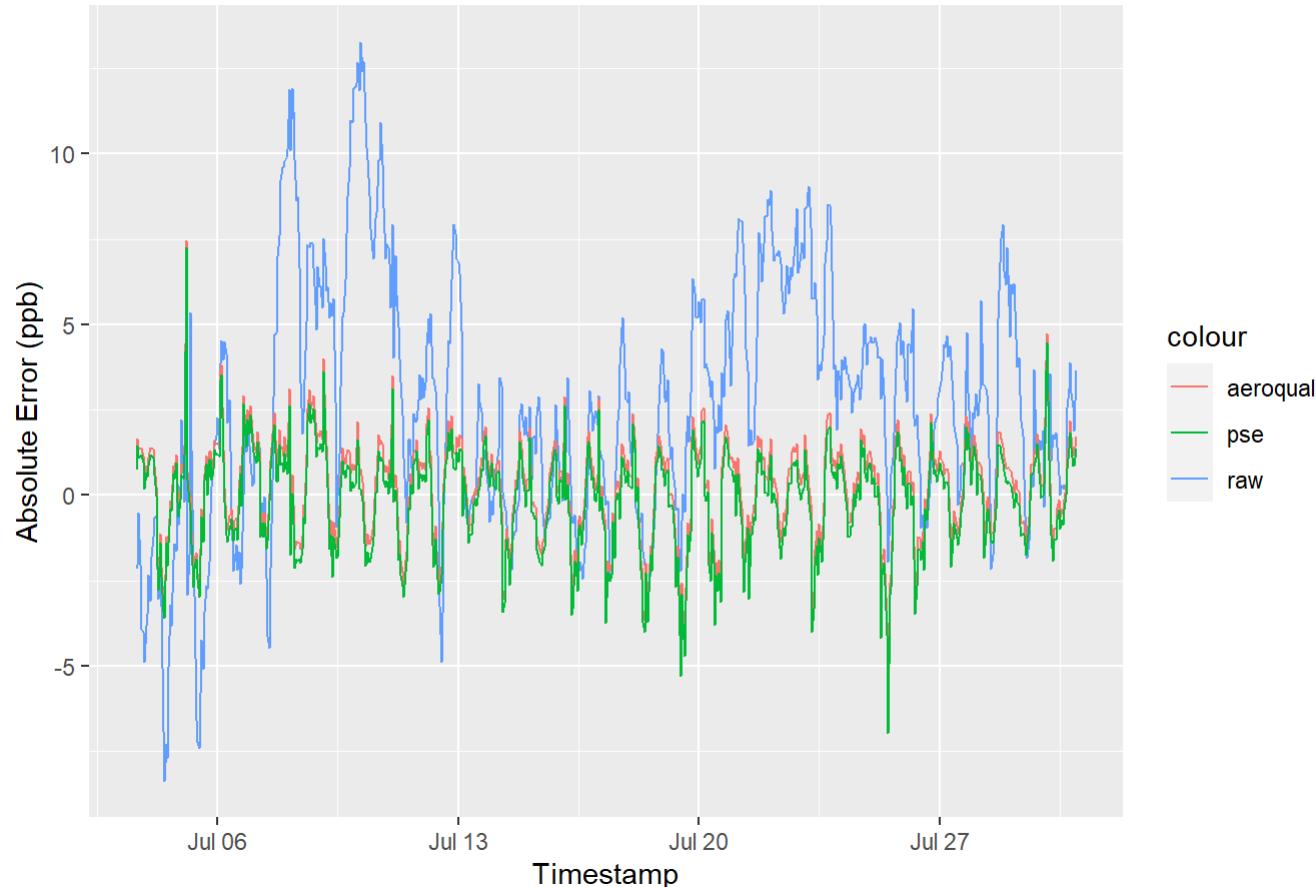


```
make_all_plots(03_data, 'AQY BB-642', '03')
```

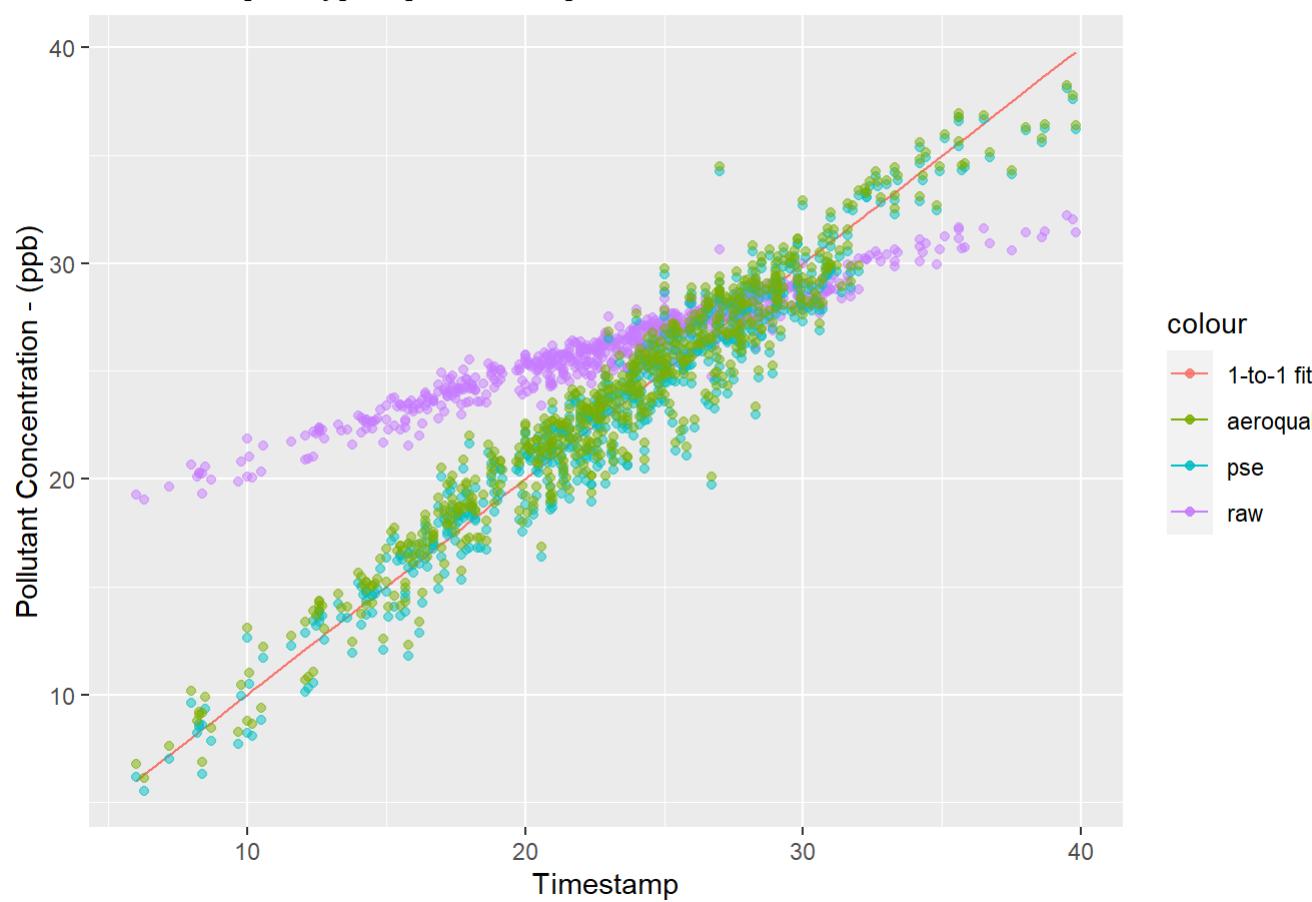
## O3 Timeseries - AQY BB-642 Month of 07



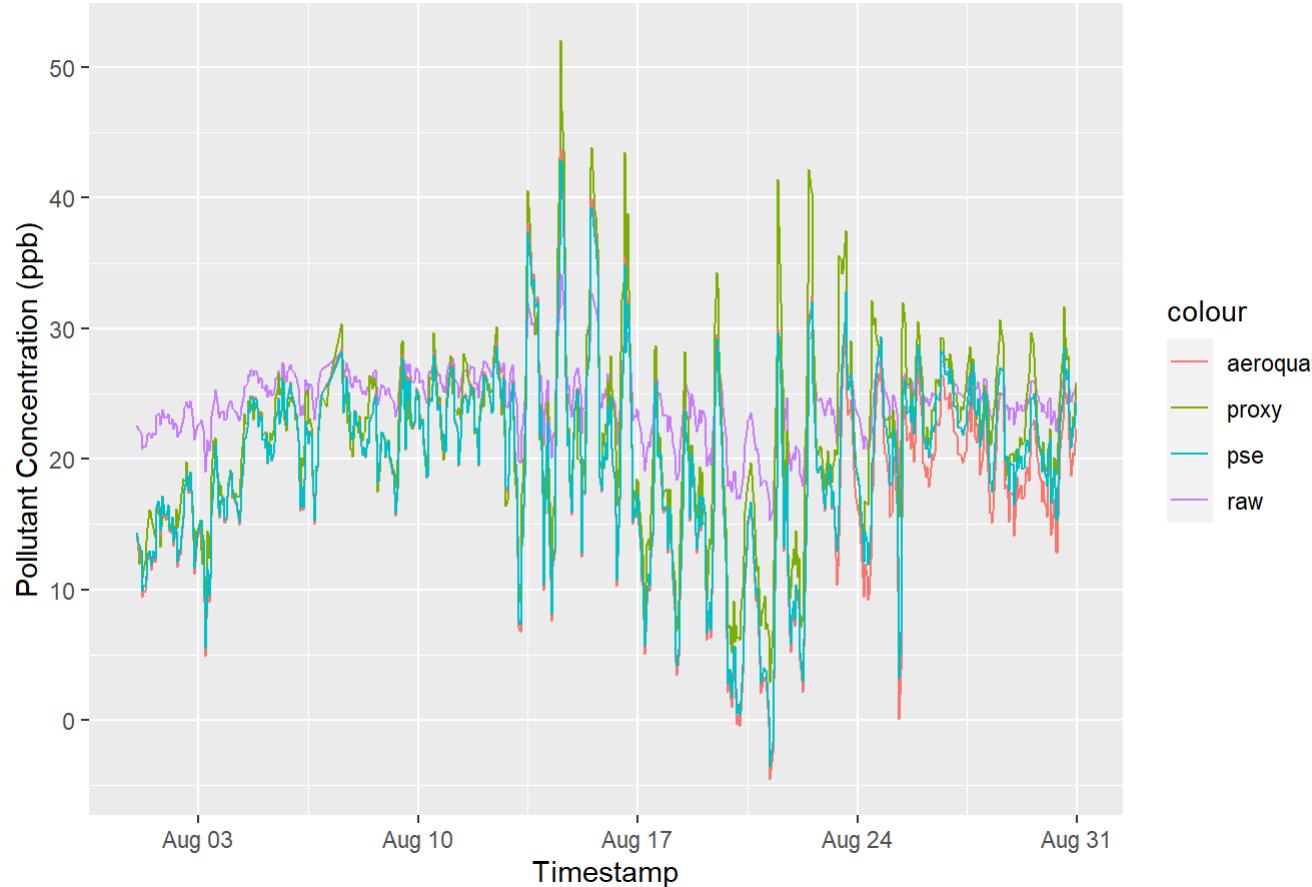
## Absolute Error for O3 - AQY BB-642 month of 07



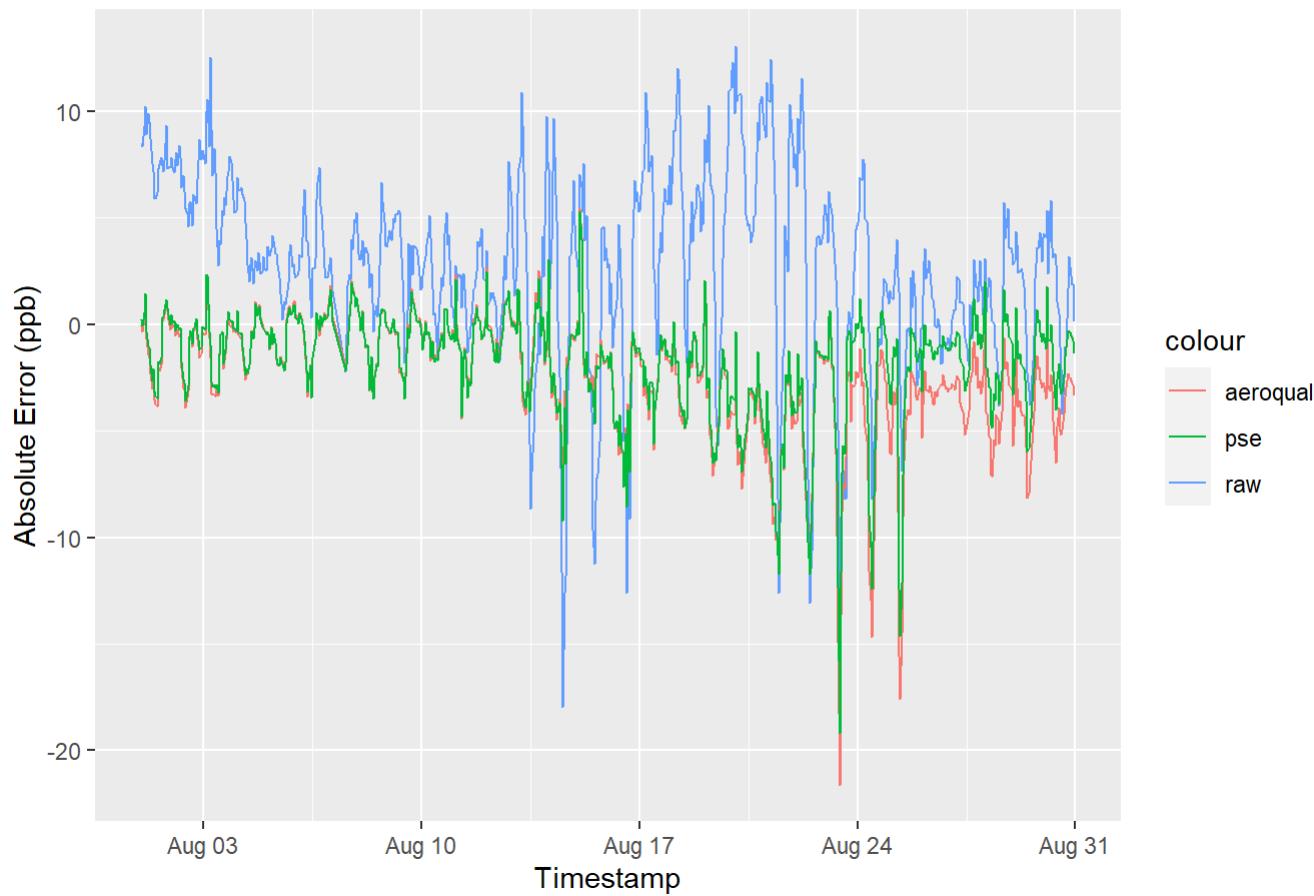
## O3 [Proxy] vs [Measured] for AQY BB-642 month of 07



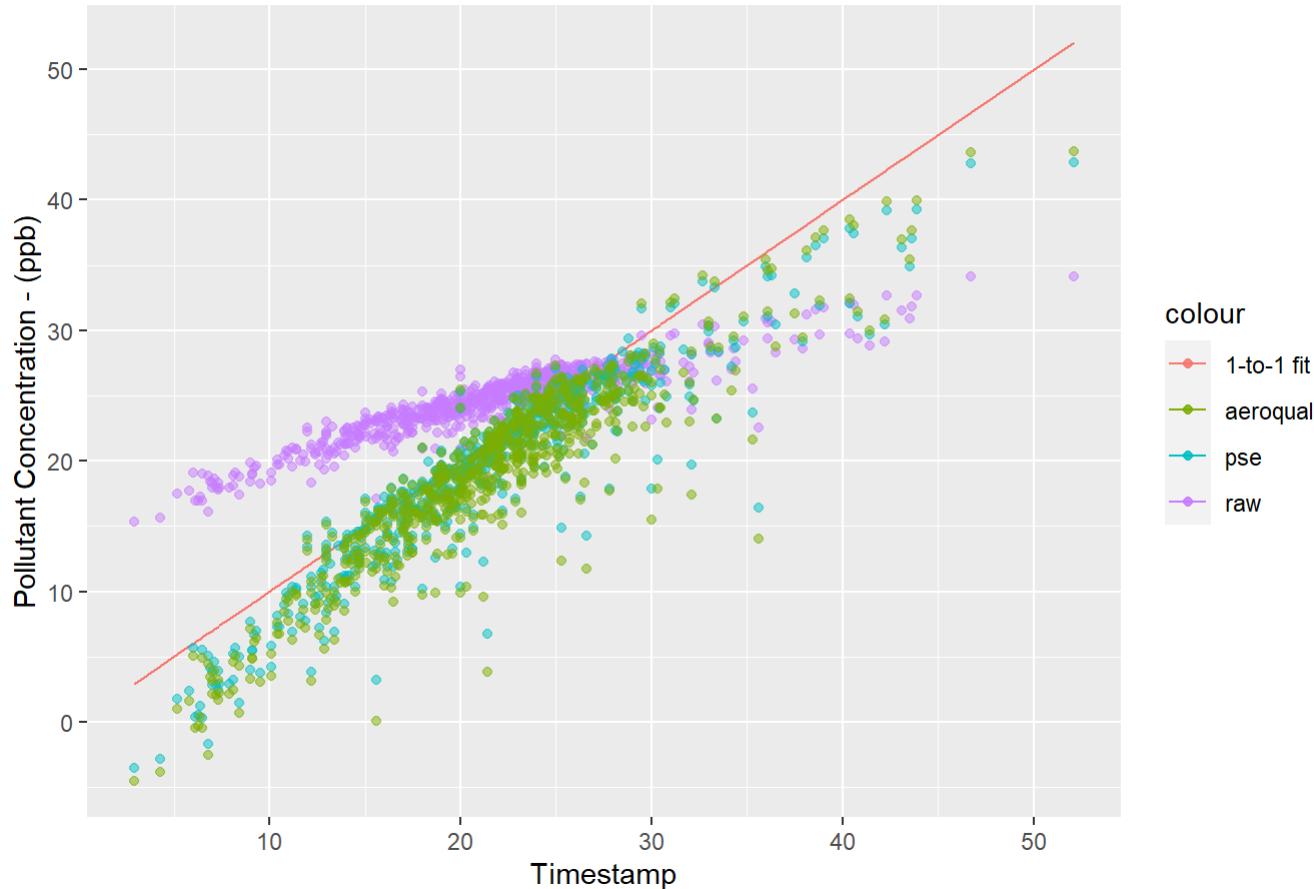
O3 Timeseries - AQY BB-642 Month of 08



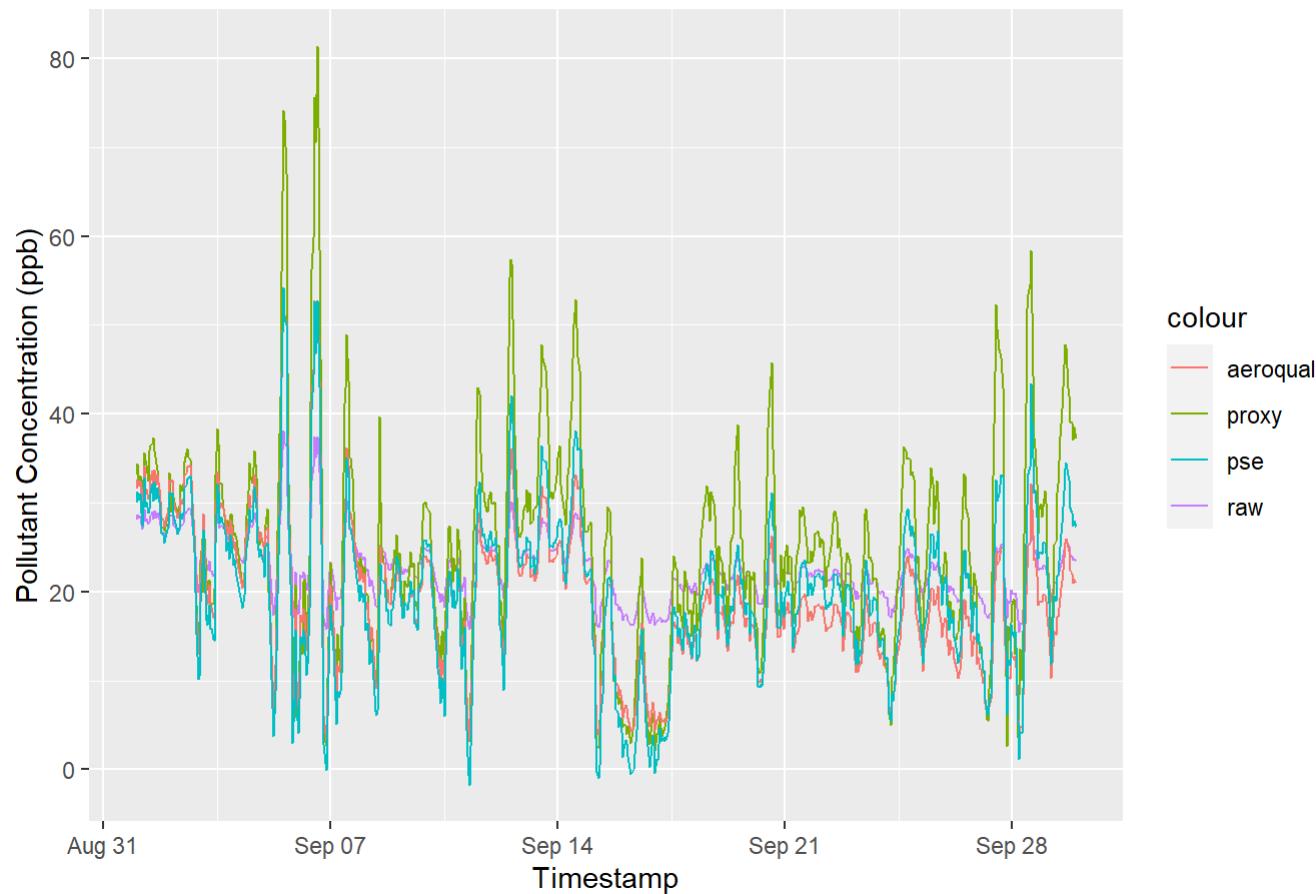
## Absolute Error for O3 - AQY BB-642 month of 08



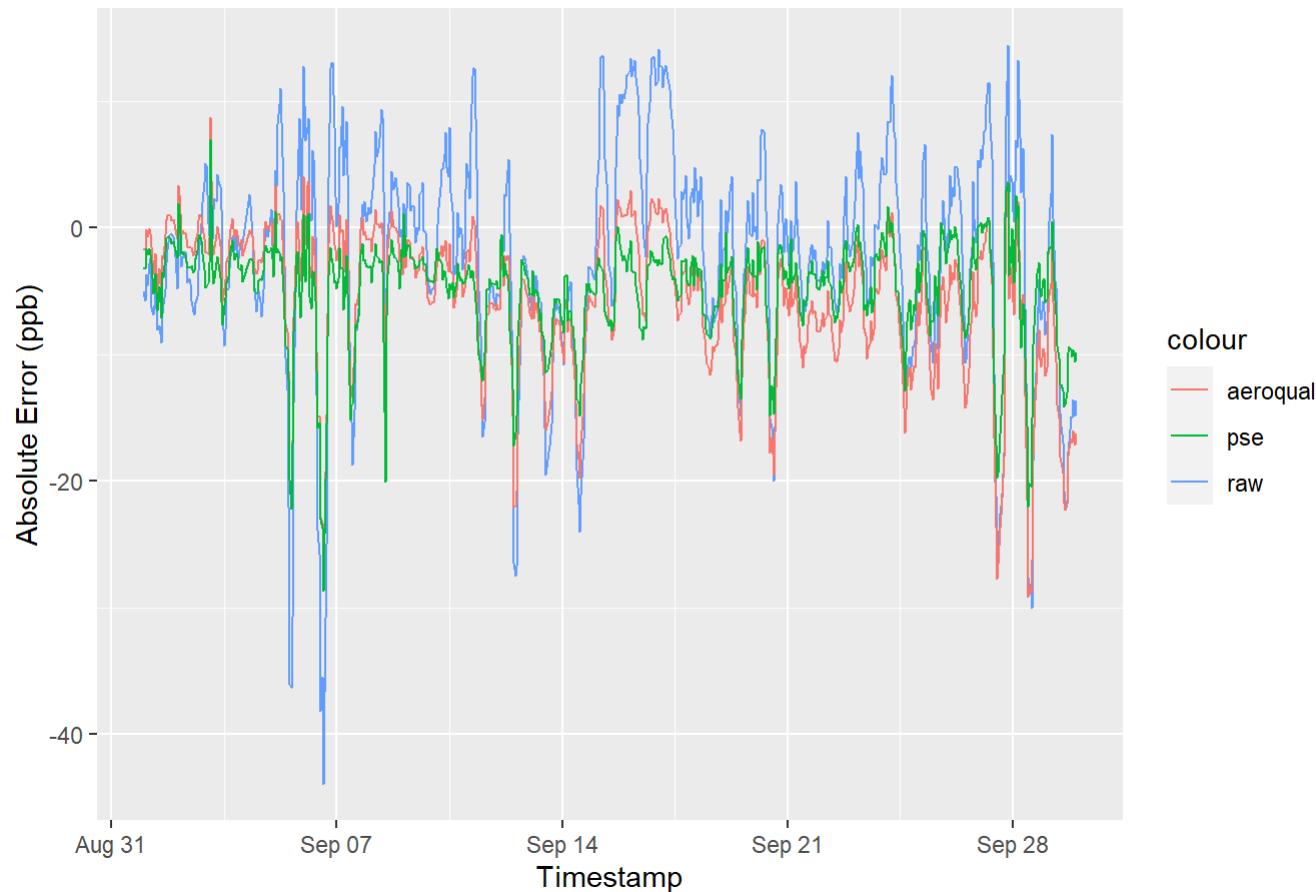
## O3 [Proxy] vs [Measured] for AQY BB-642 month of 08



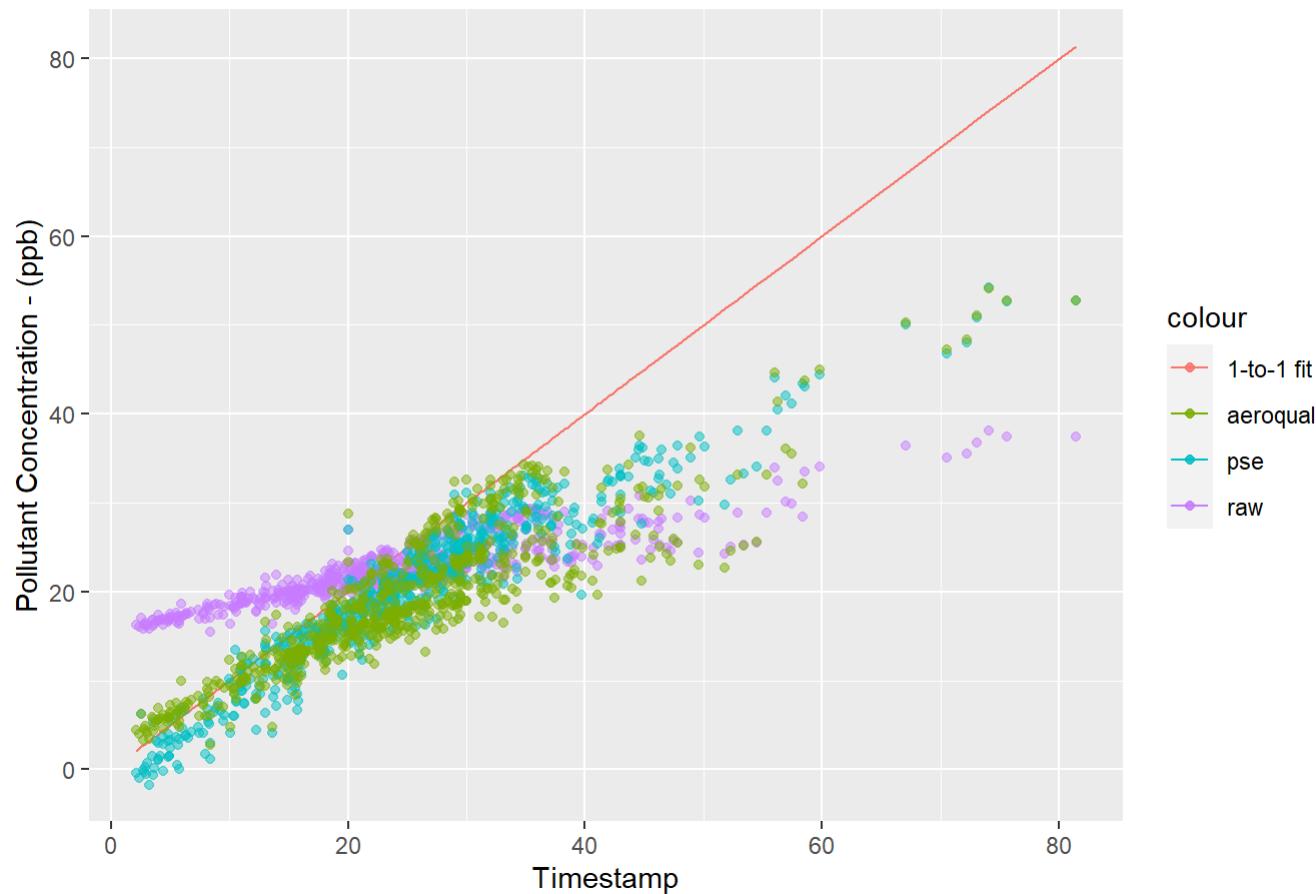
## O3 Timeseries - AQY BB-642 Month of 09



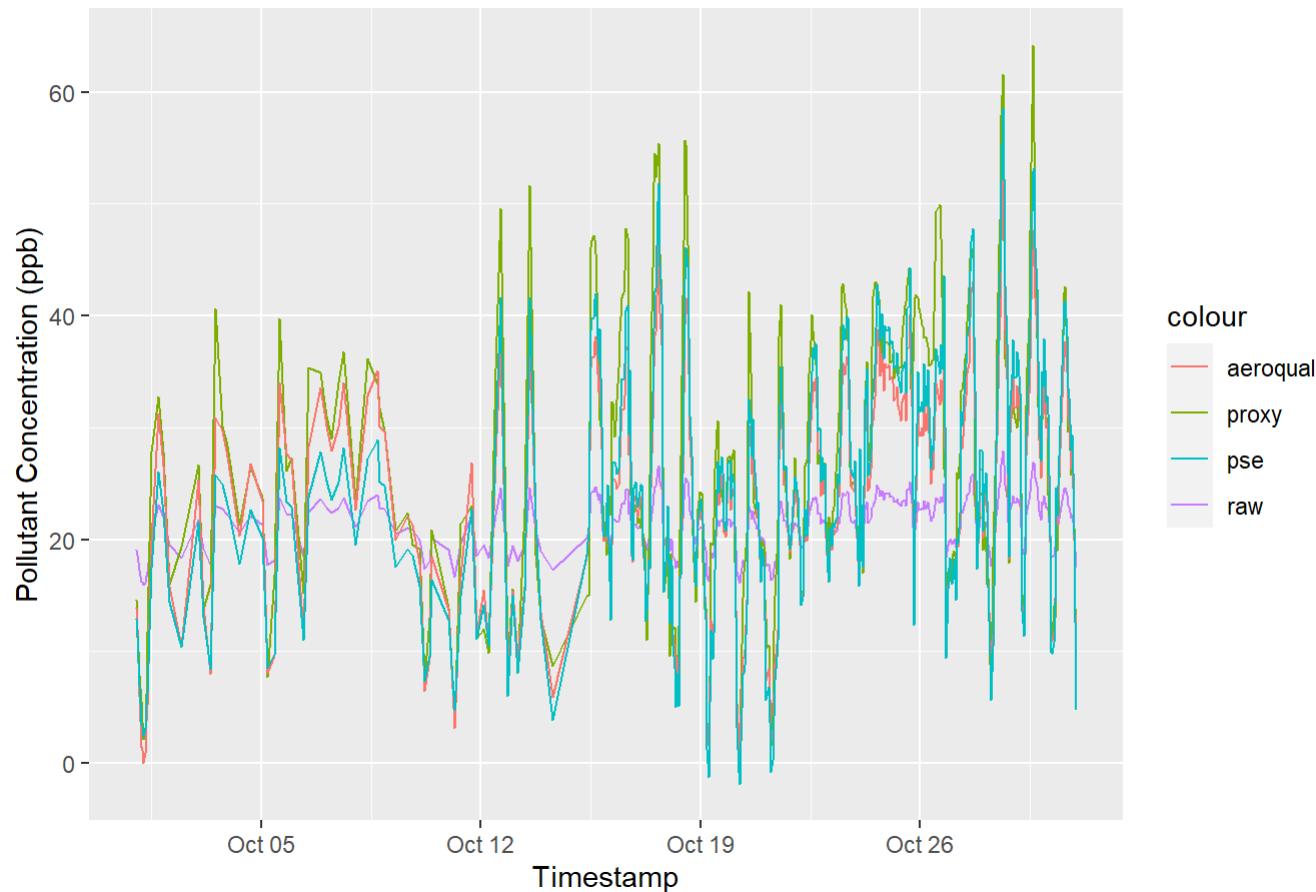
## Absolute Error for O3 - AQY BB-642 month of 09



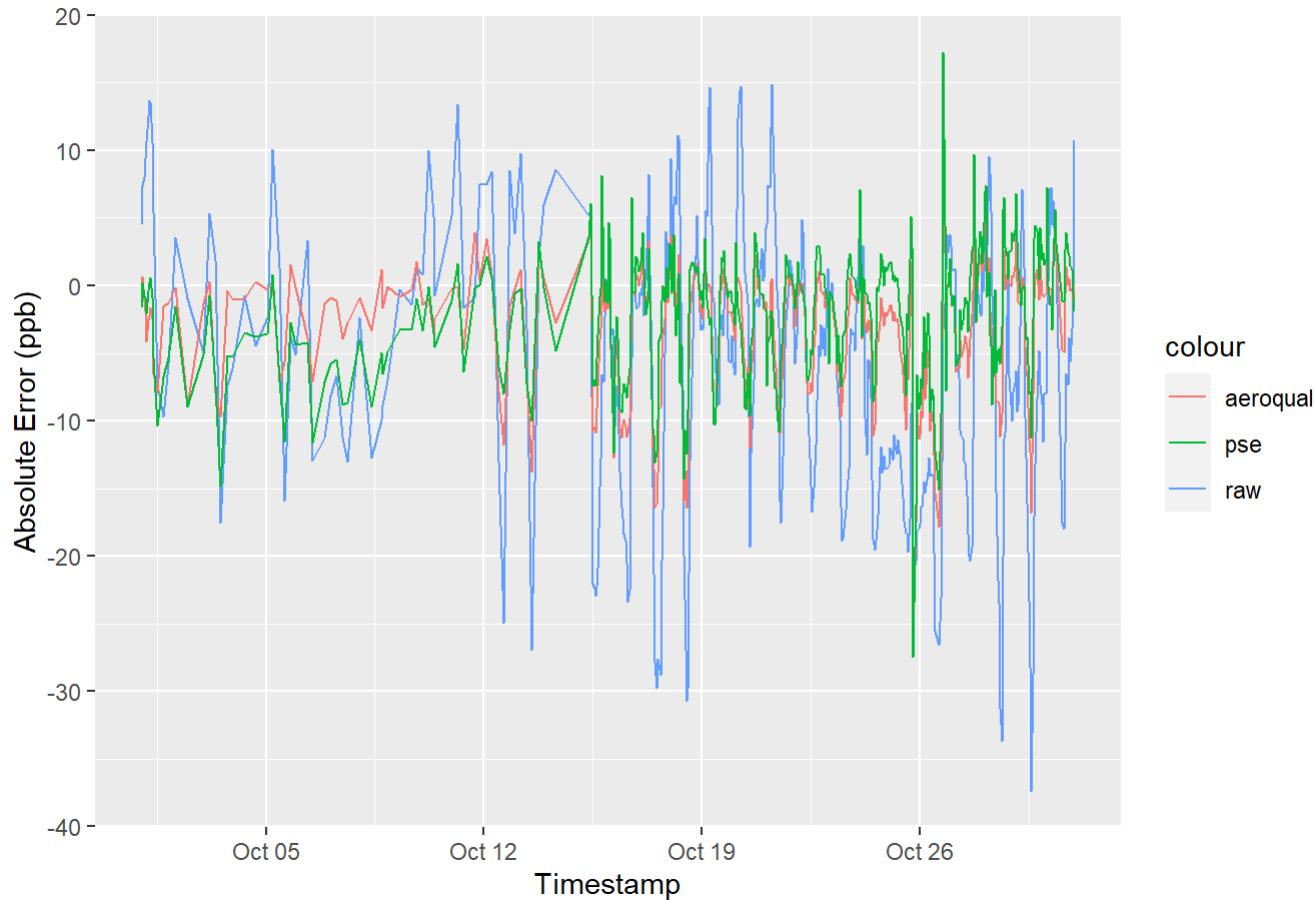
## O3 [Proxy] vs [Measured] for AQY BB-642 month of 09



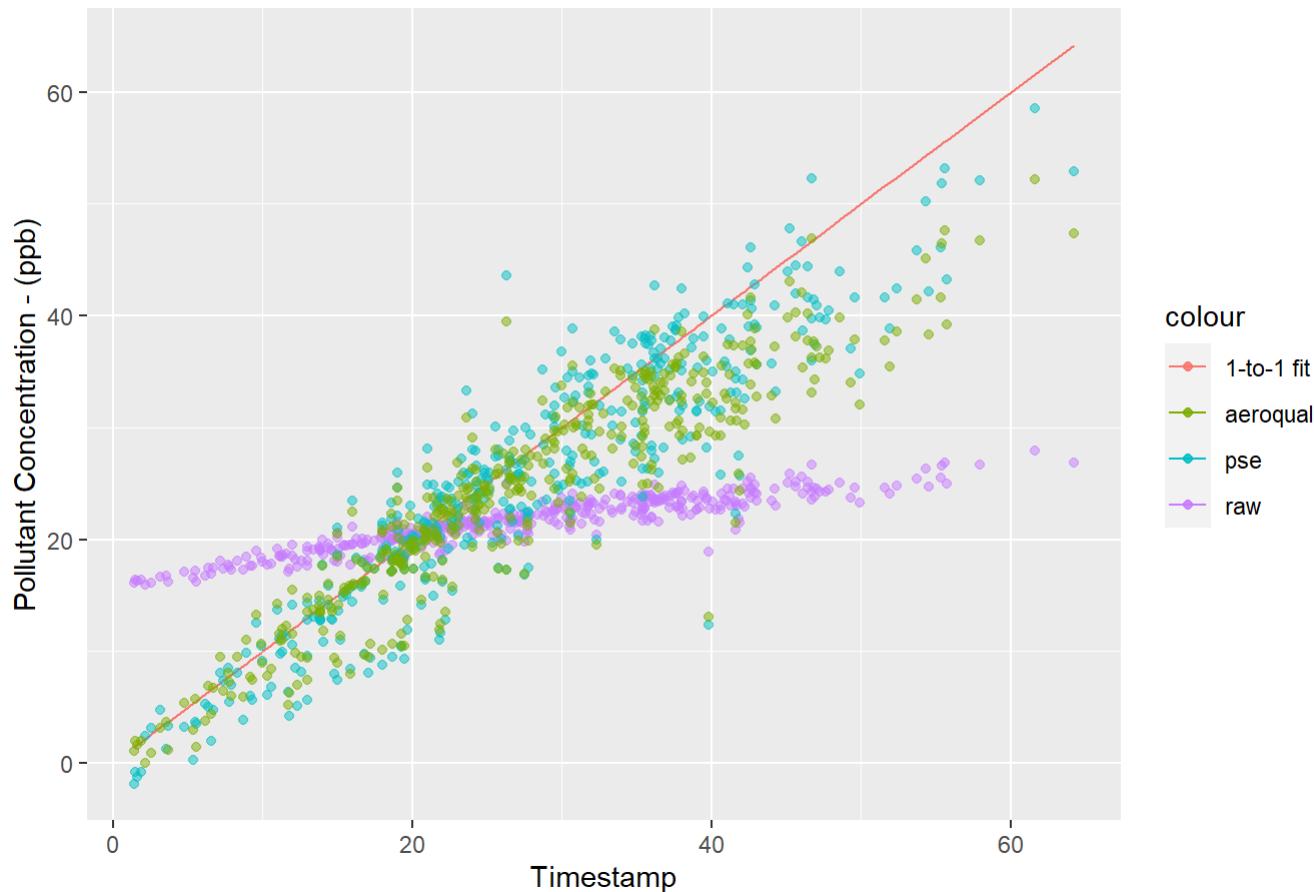
## O3 Timeseries - AQY BB-642 Month of 10



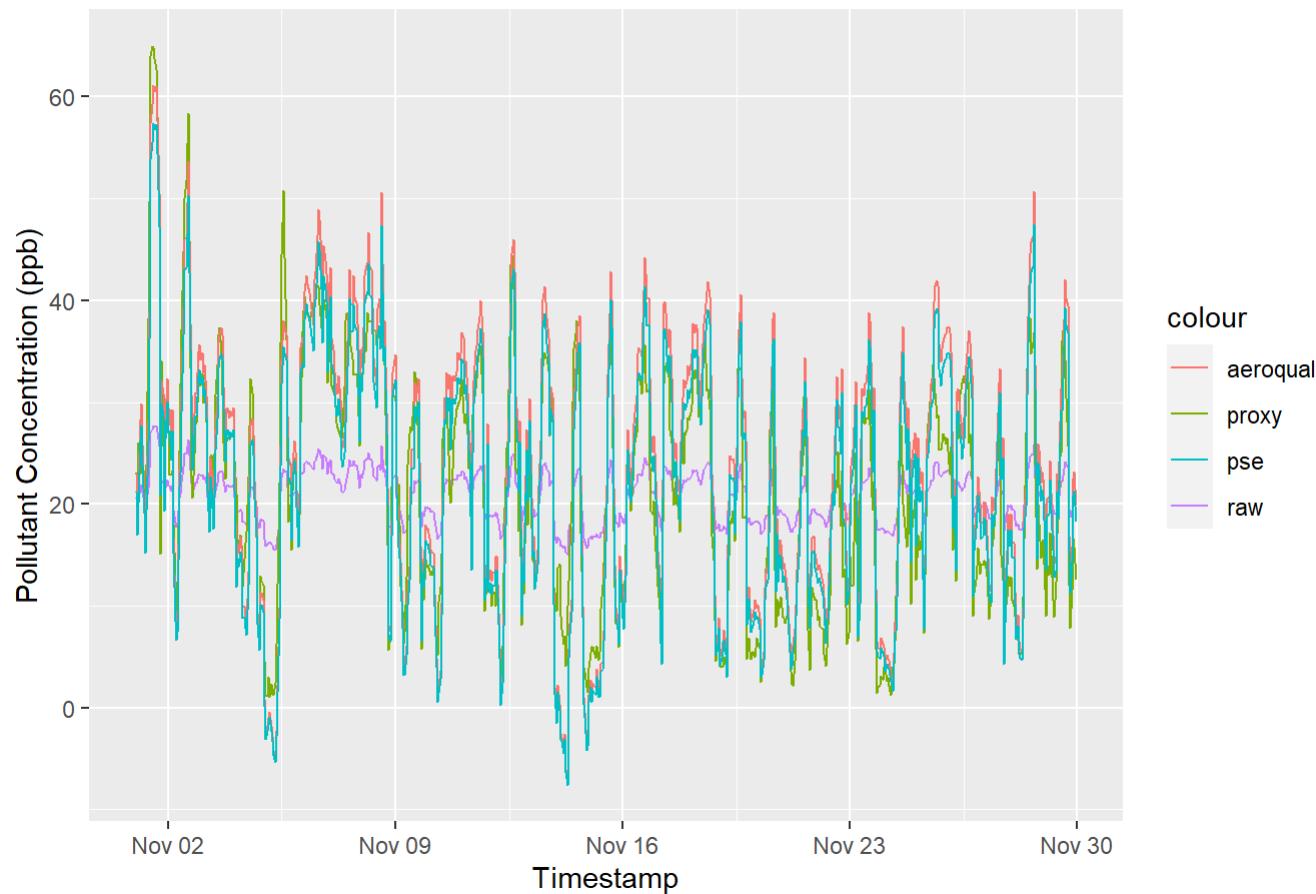
## Absolute Error for O3 - AQY BB-642 month of 10



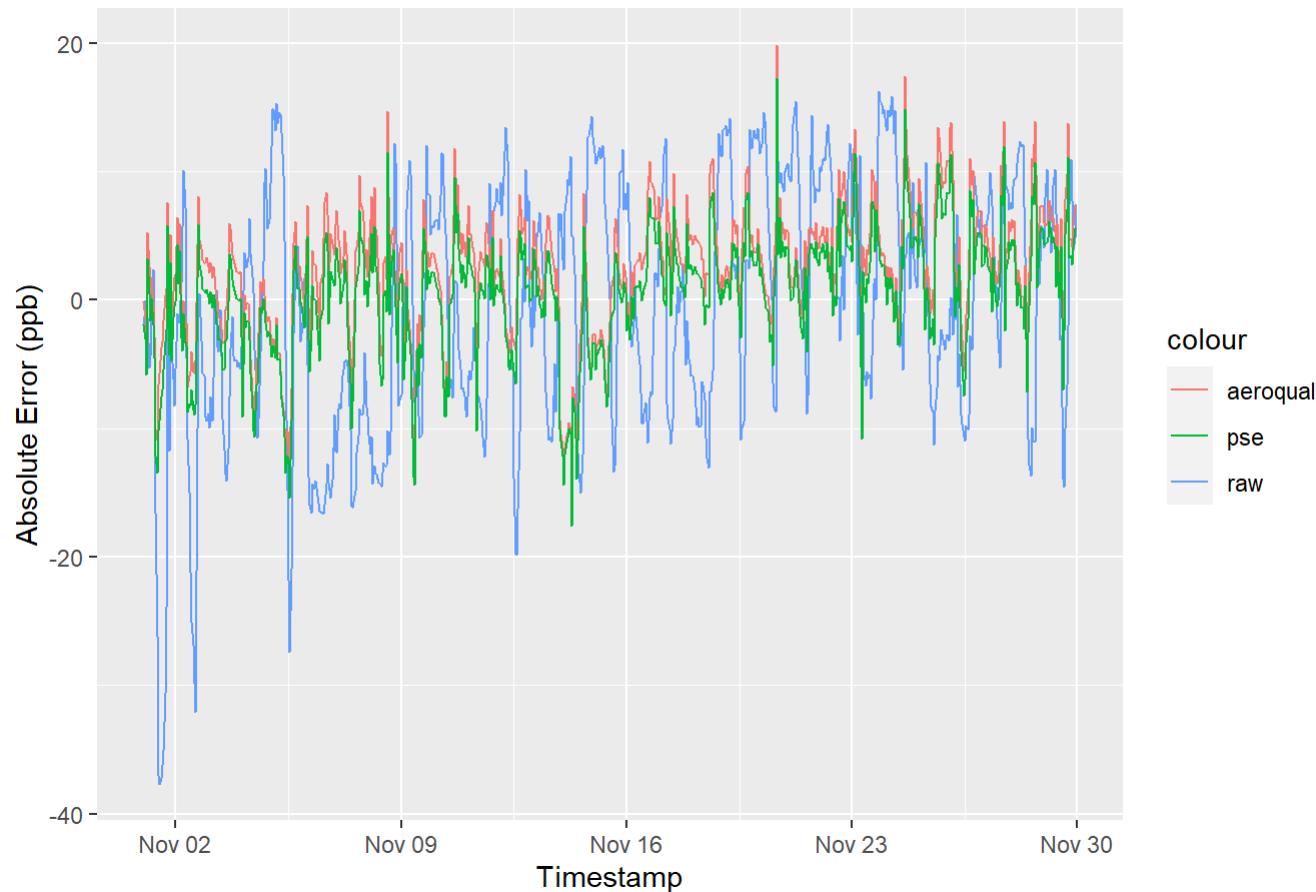
## O3 [Proxy] vs [Measured] for AQY BB-642 month of 10



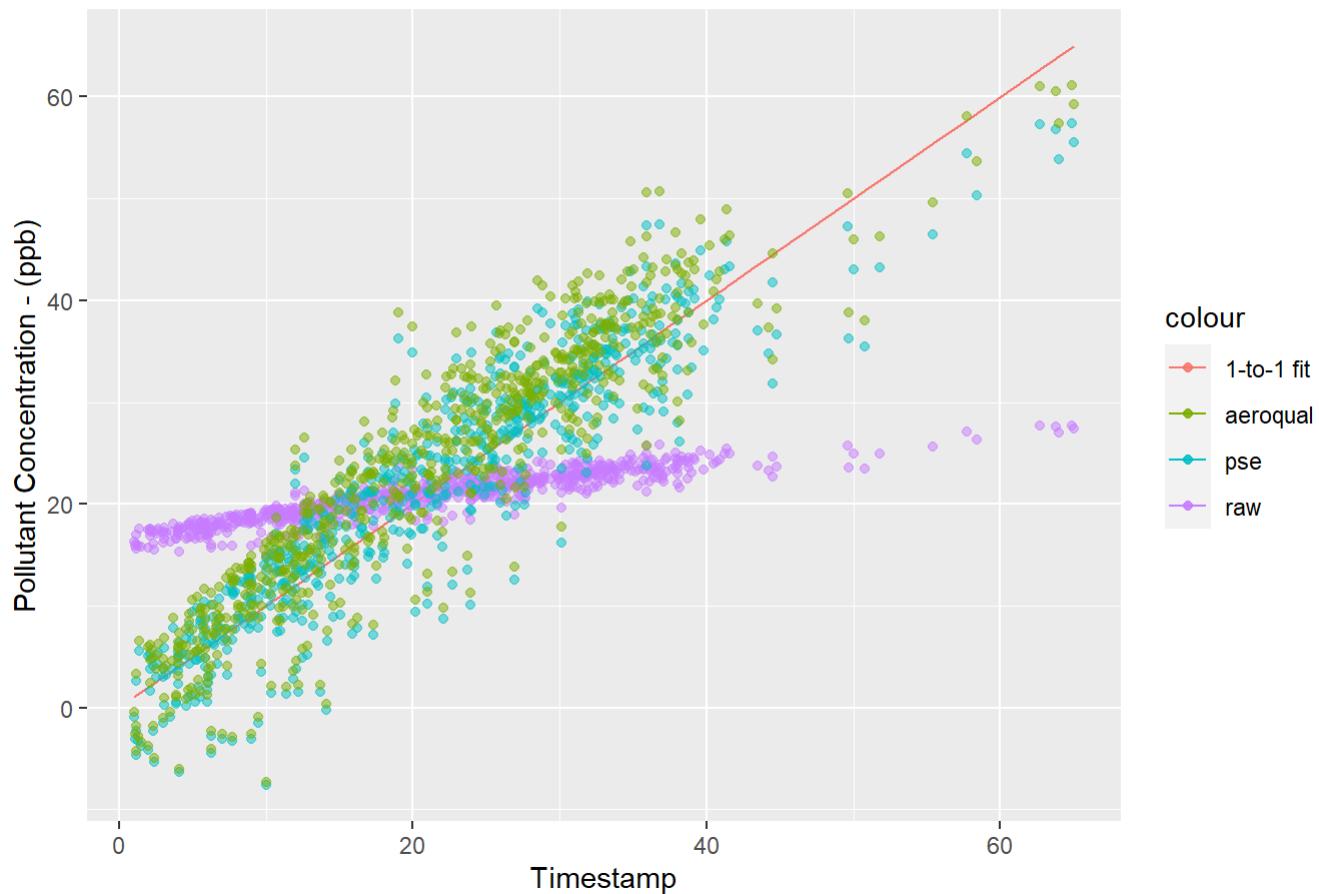
## O3 Timeseries - AQY BB-642 Month of 11



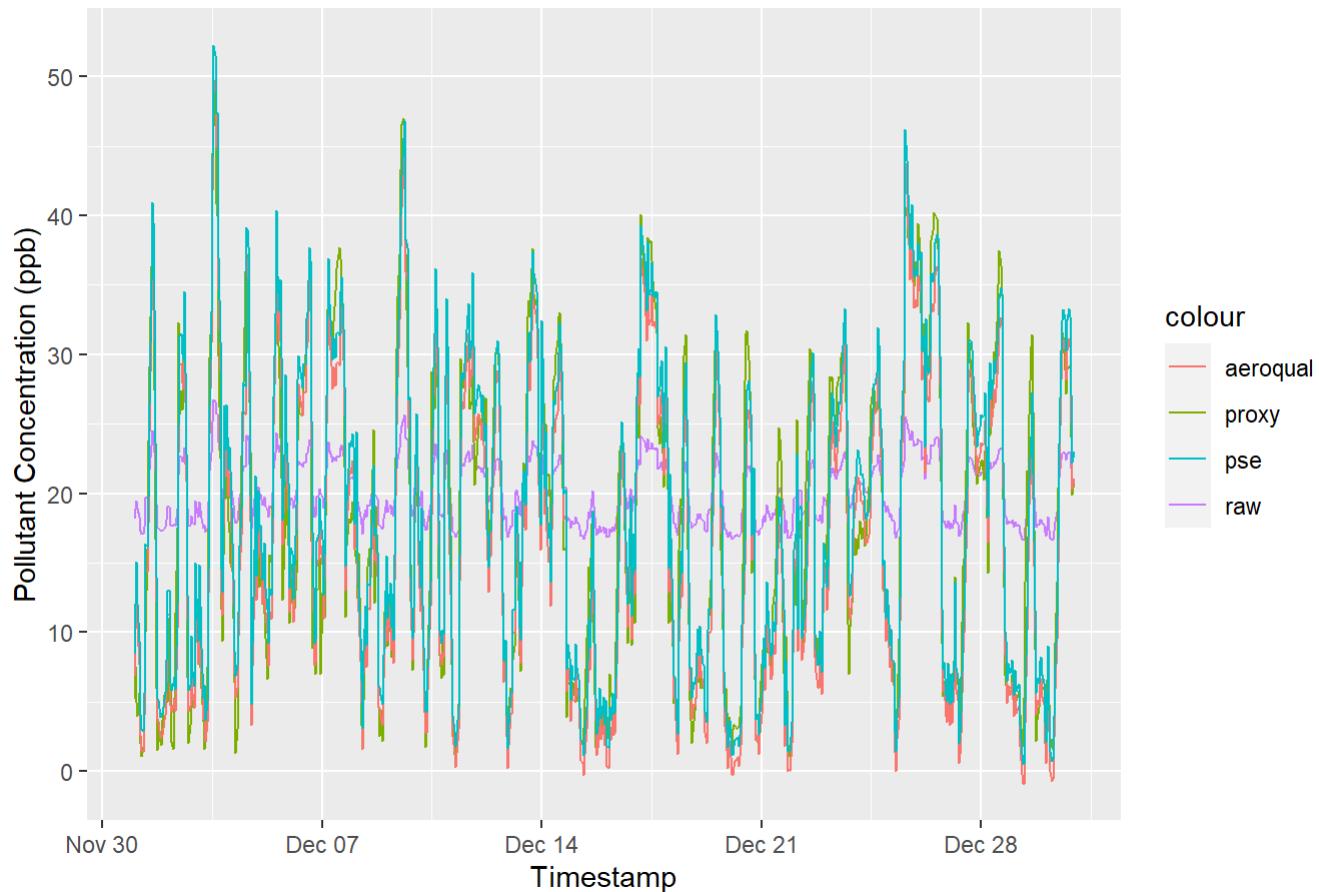
## Absolute Error for O3 - AQY BB-642 month of 11



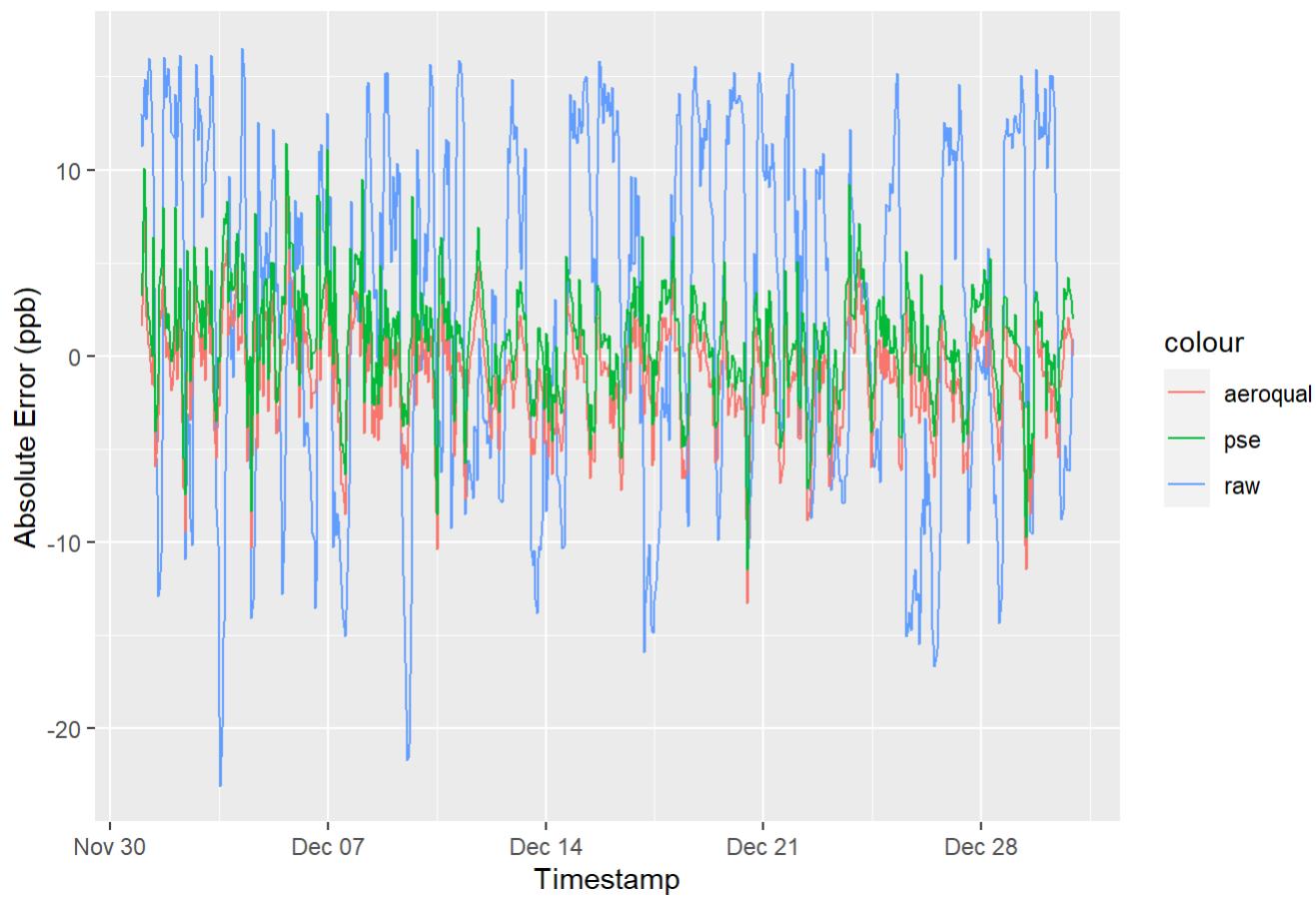
## O3 [Proxy] vs [Measured] for AQY BB-642 month of 11



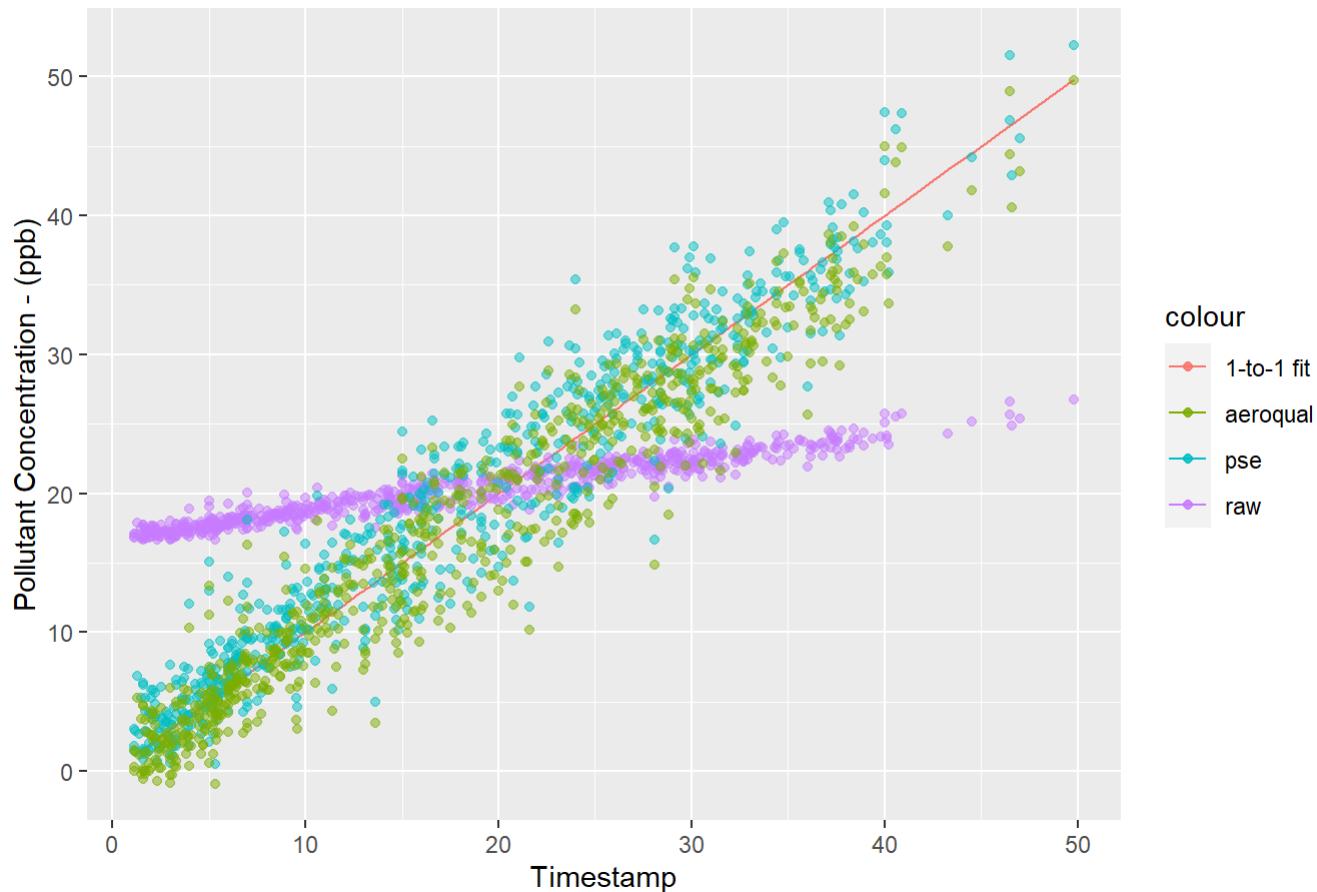
## O3 Timeseries - AQY BB-642 Month of 12



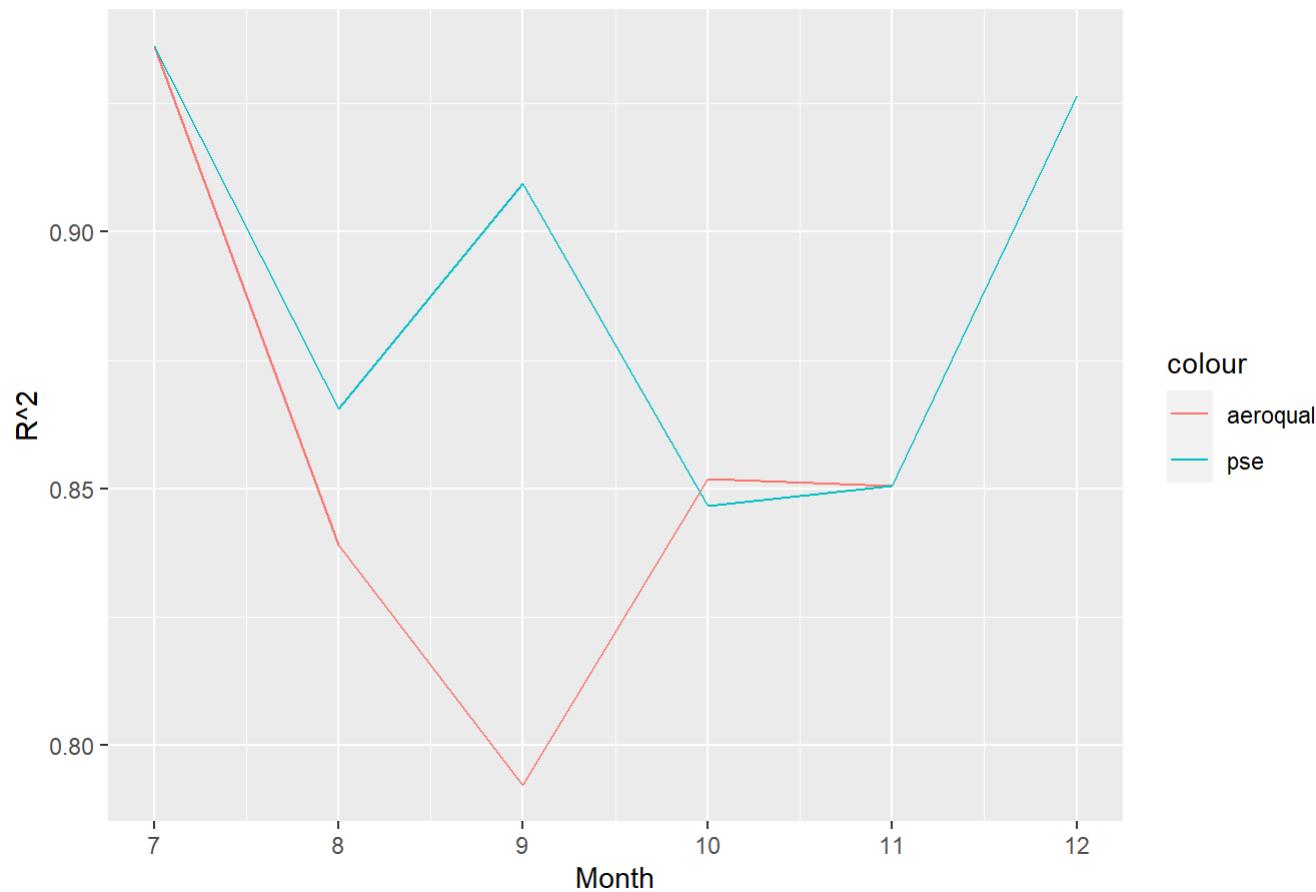
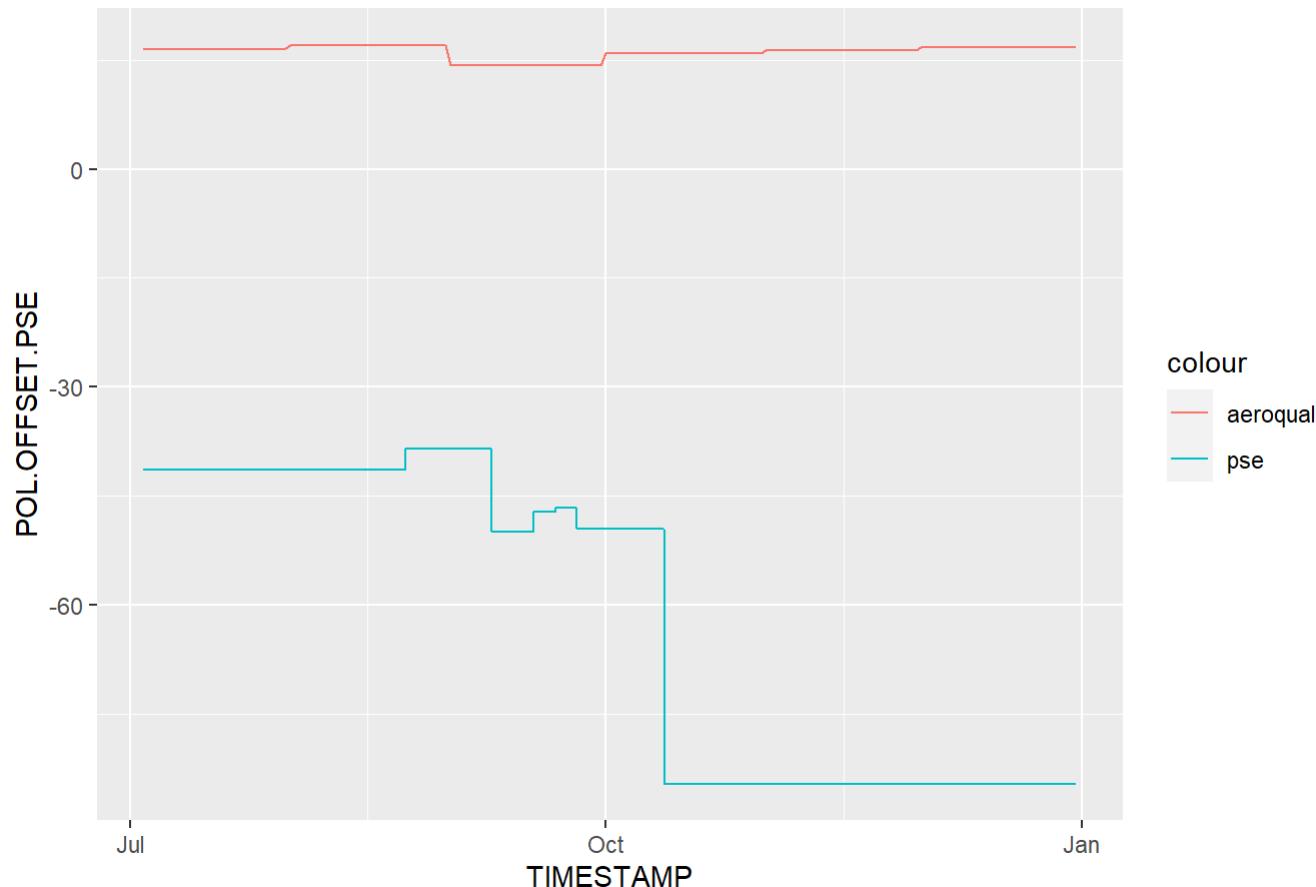
## Absolute Error for O3 - AQY BB-642 month of 12



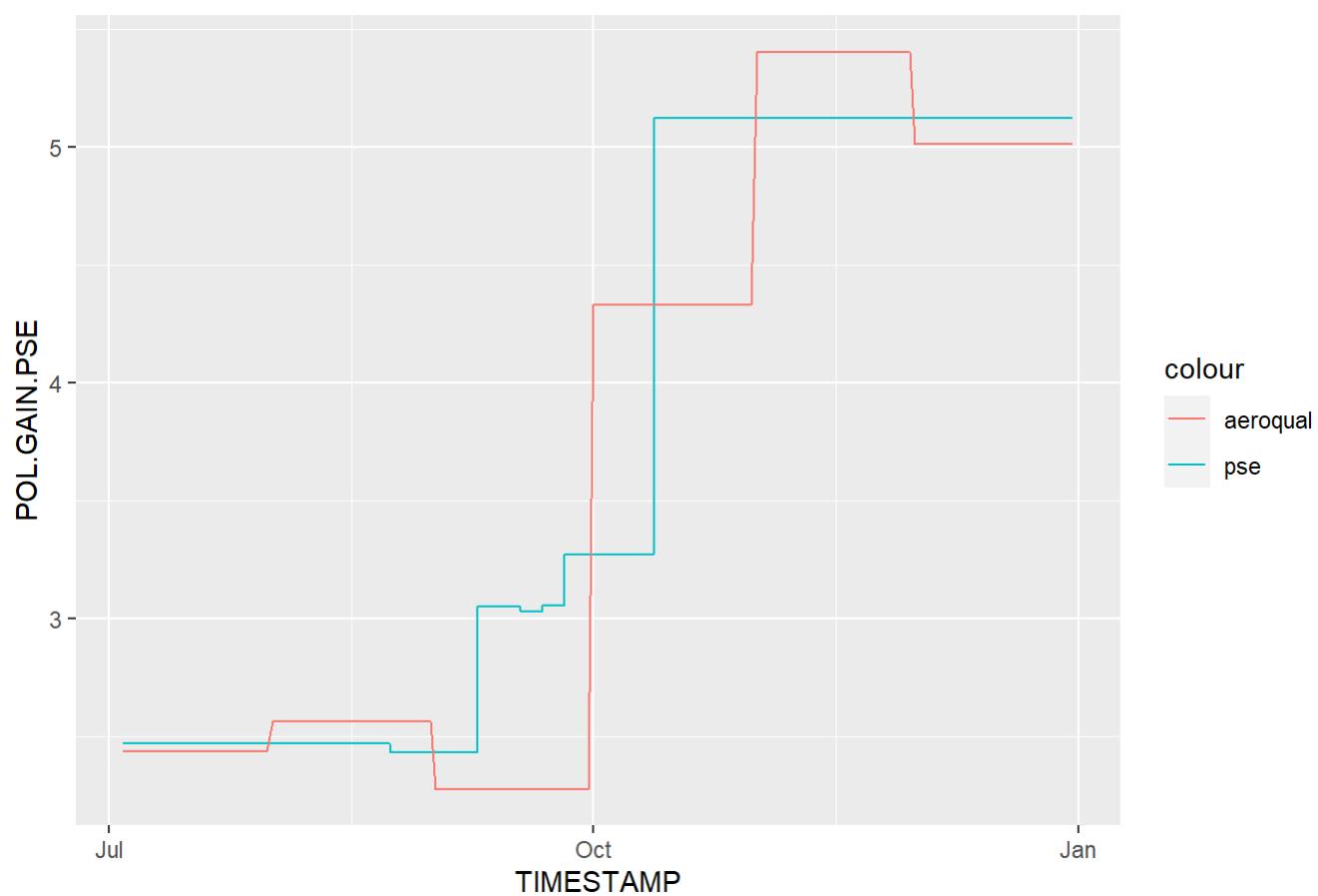
## O3 [Proxy] vs [Measured] for AQY BB-642 month of 12



```
##          month     raw      PSE      AQ
## 1 Full Deployment 0.51751 0.84718 0.79159
## 2             07 0.93618 0.93618 0.93618
## 3             08 0.83907 0.86549 0.83903
## 4             09 0.79235 0.90946 0.79235
## 5             10 0.85183 0.84658 0.85183
## 6             11 0.85054 0.85054 0.85053
## 7             12 0.92646 0.92646 0.92646
```

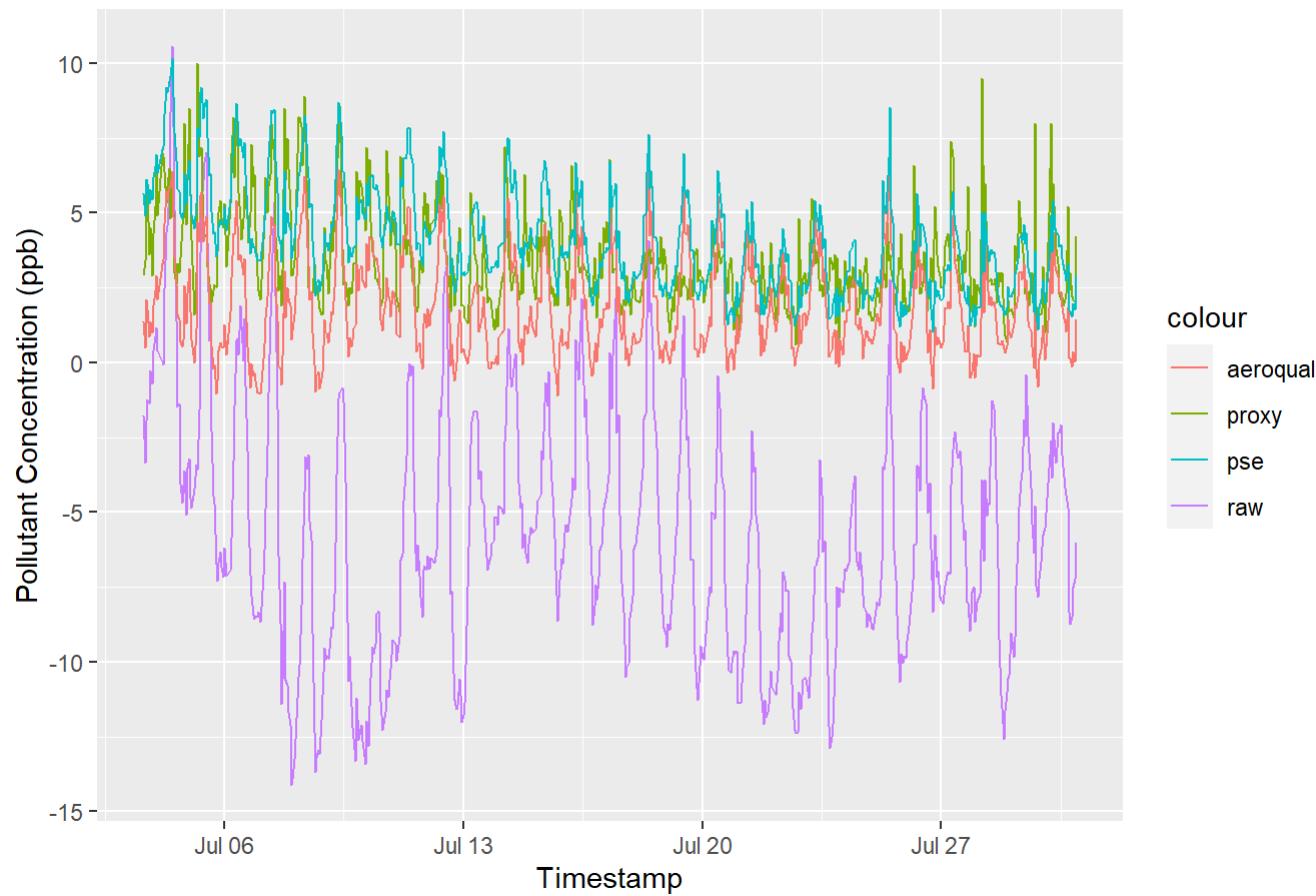
**O3 Proxy ~ Measured R^2 Over Time for AQY BB-642****O3 Offset Over Time for AQY BB-642****O3 Gain Over Time for AQY BB-642**

## Compare Results

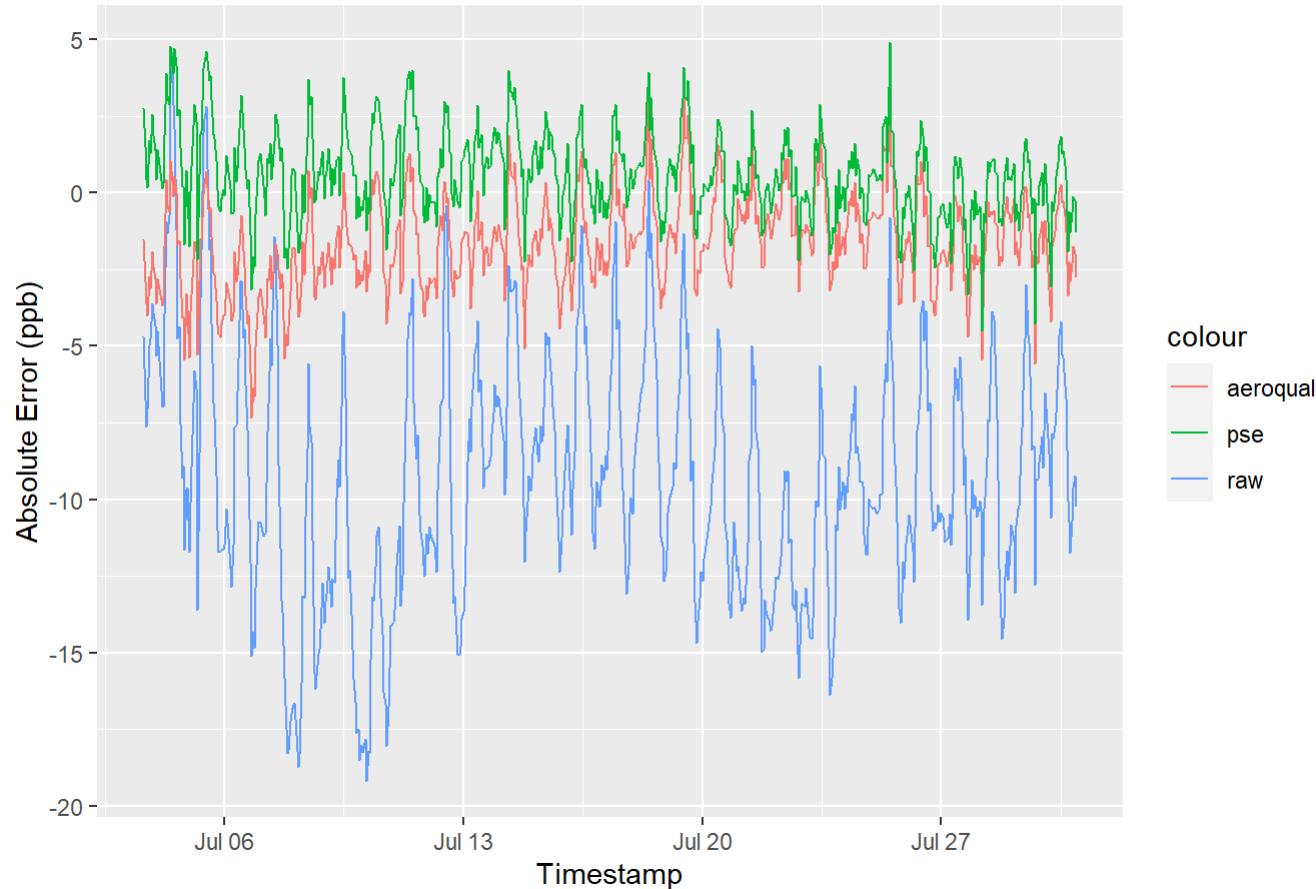


```
make_all_plots(NO2_data, 'AQY BB-633', 'NO2')
```

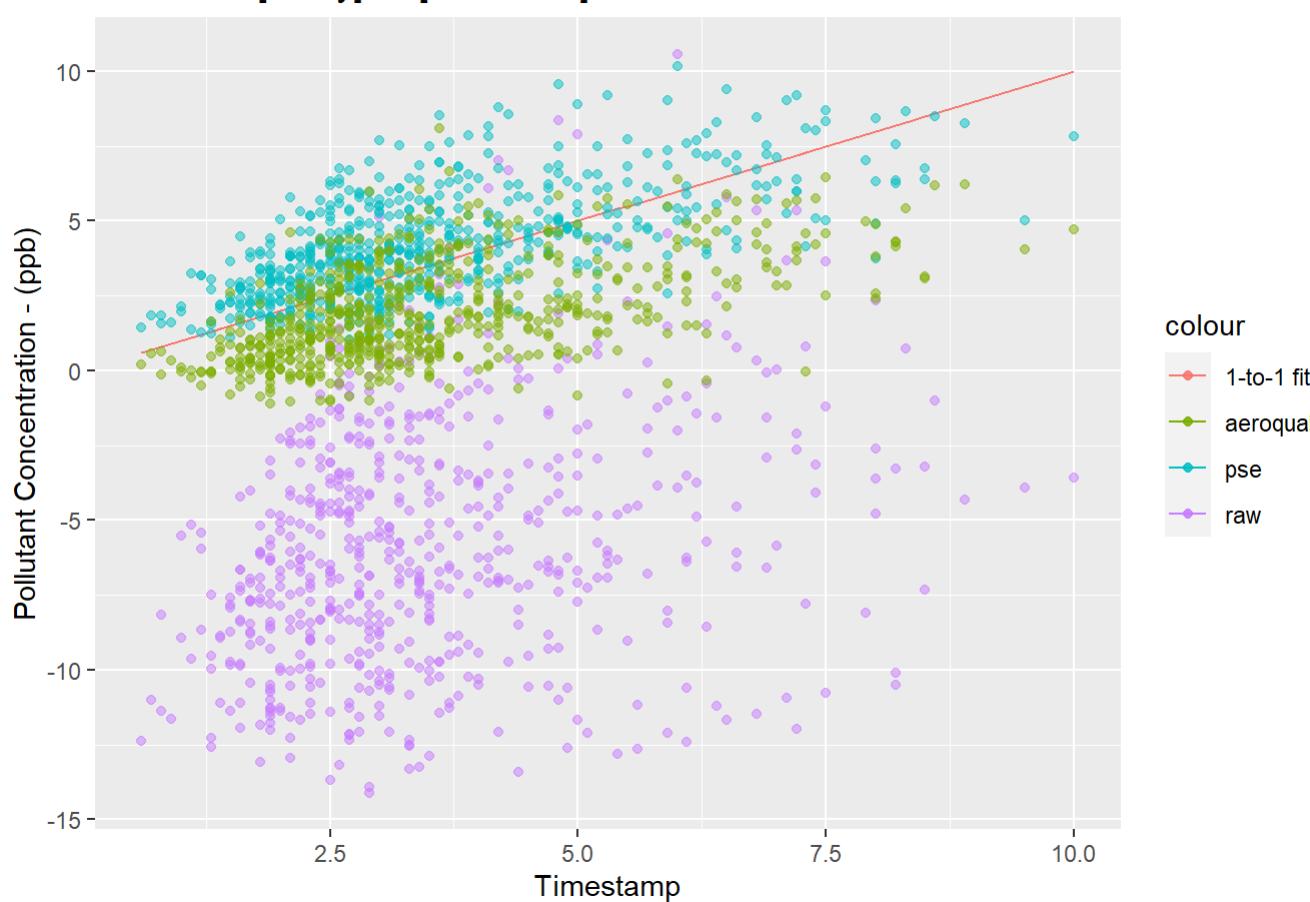
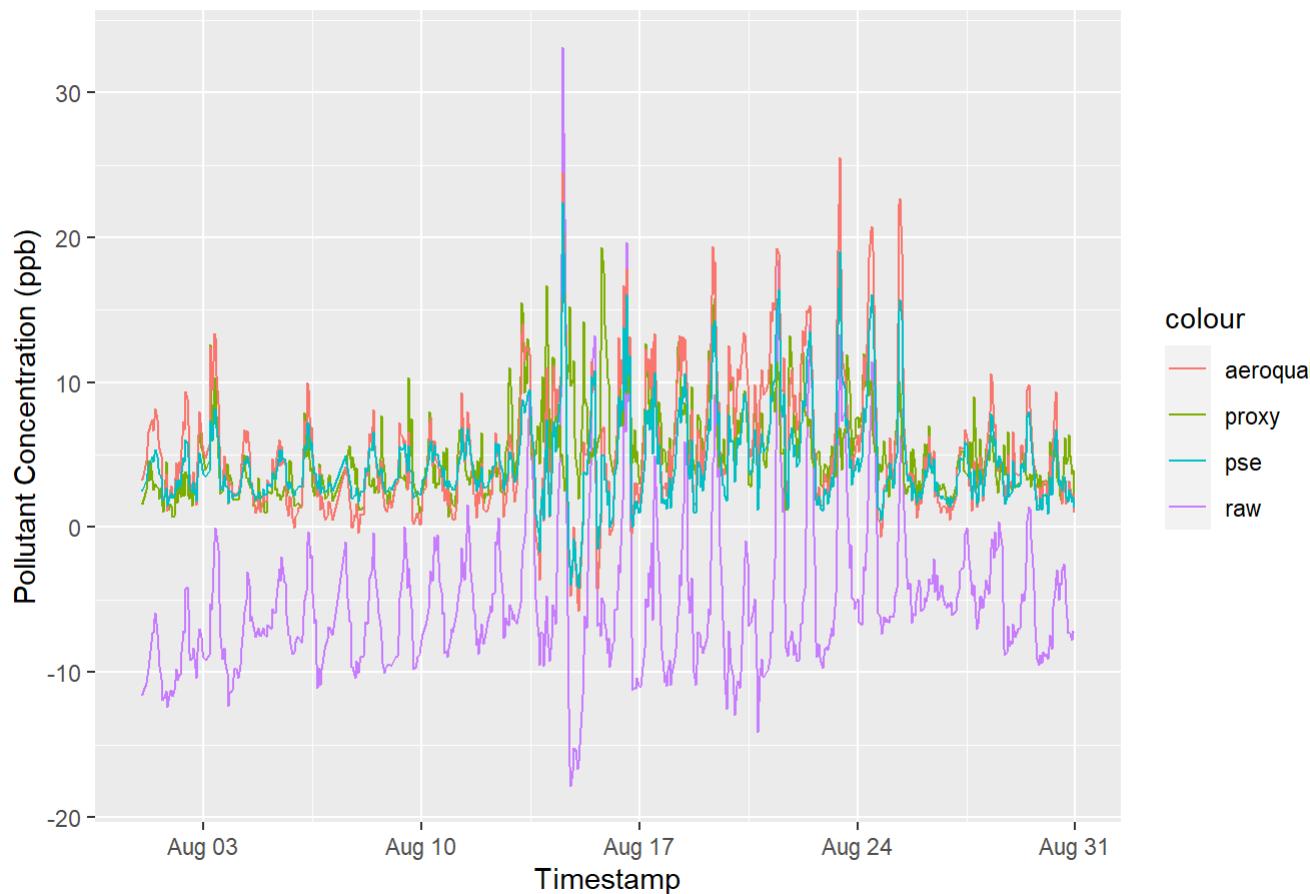
## NO2 Timeseries - AQY BB-633 Month of 07



## Absolute Error for NO2 - AQY BB-633 month of 07



## NO2 [Proxy] vs [Measured] for AQY BB-633 month of 07

NO<sub>2</sub> Timeseries - AQY BB-633 Month of 08

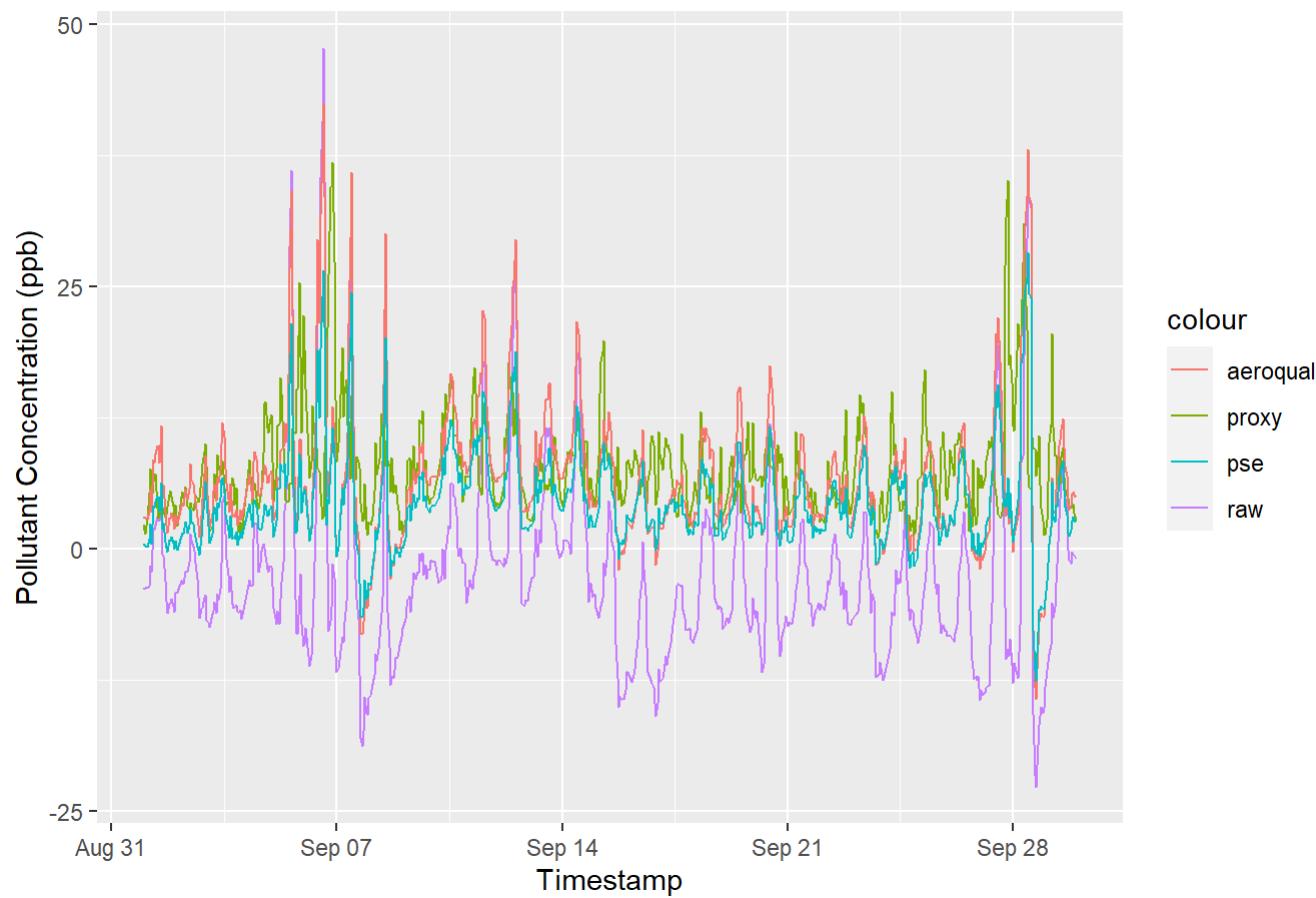
## Absolute Error for NO2 - AQY BB-633 month of 08



## NO2 [Proxy] vs [Measured] for AQY BB-633 month of 08



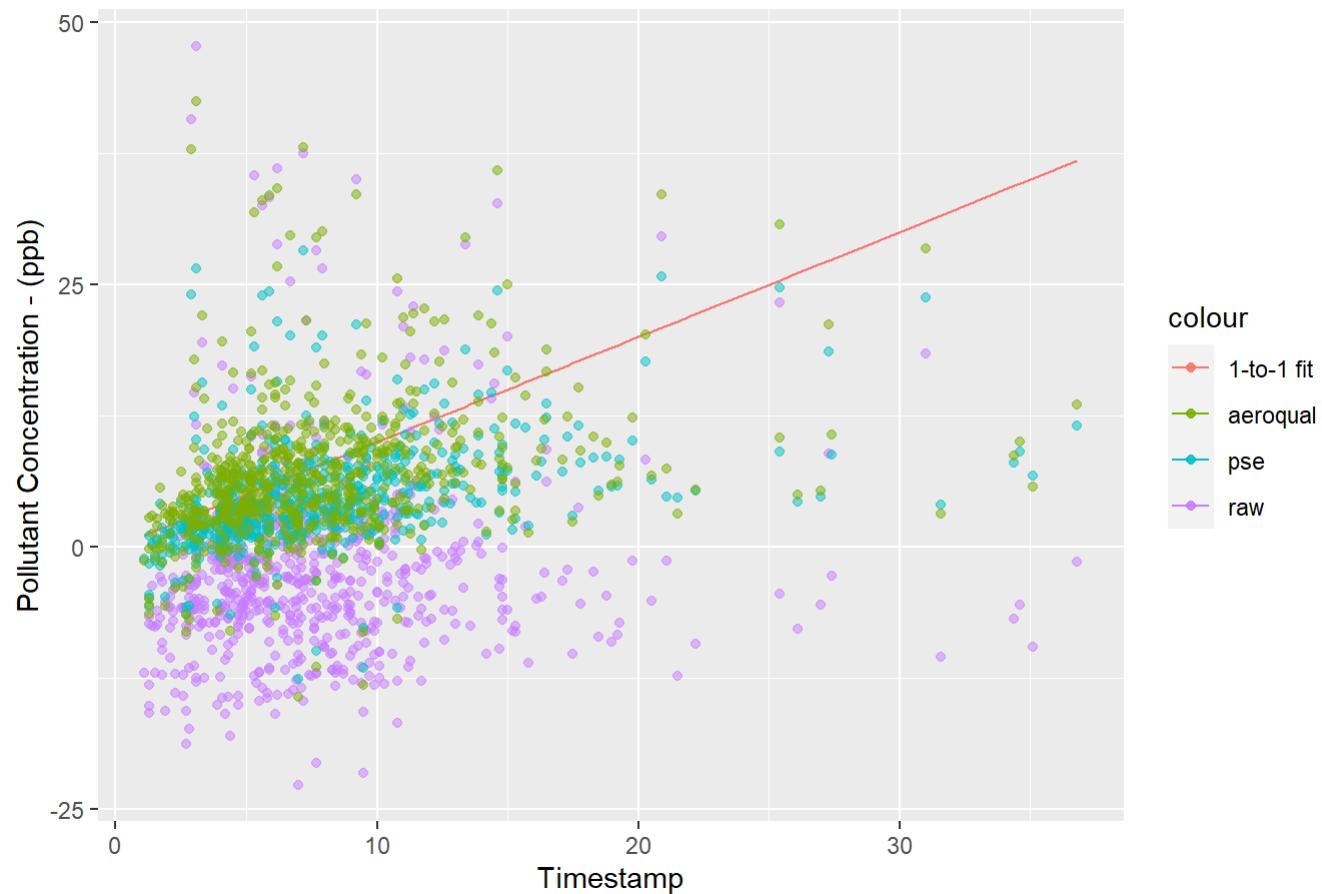
## NO2 Timeseries - AQY BB-633 Month of 09



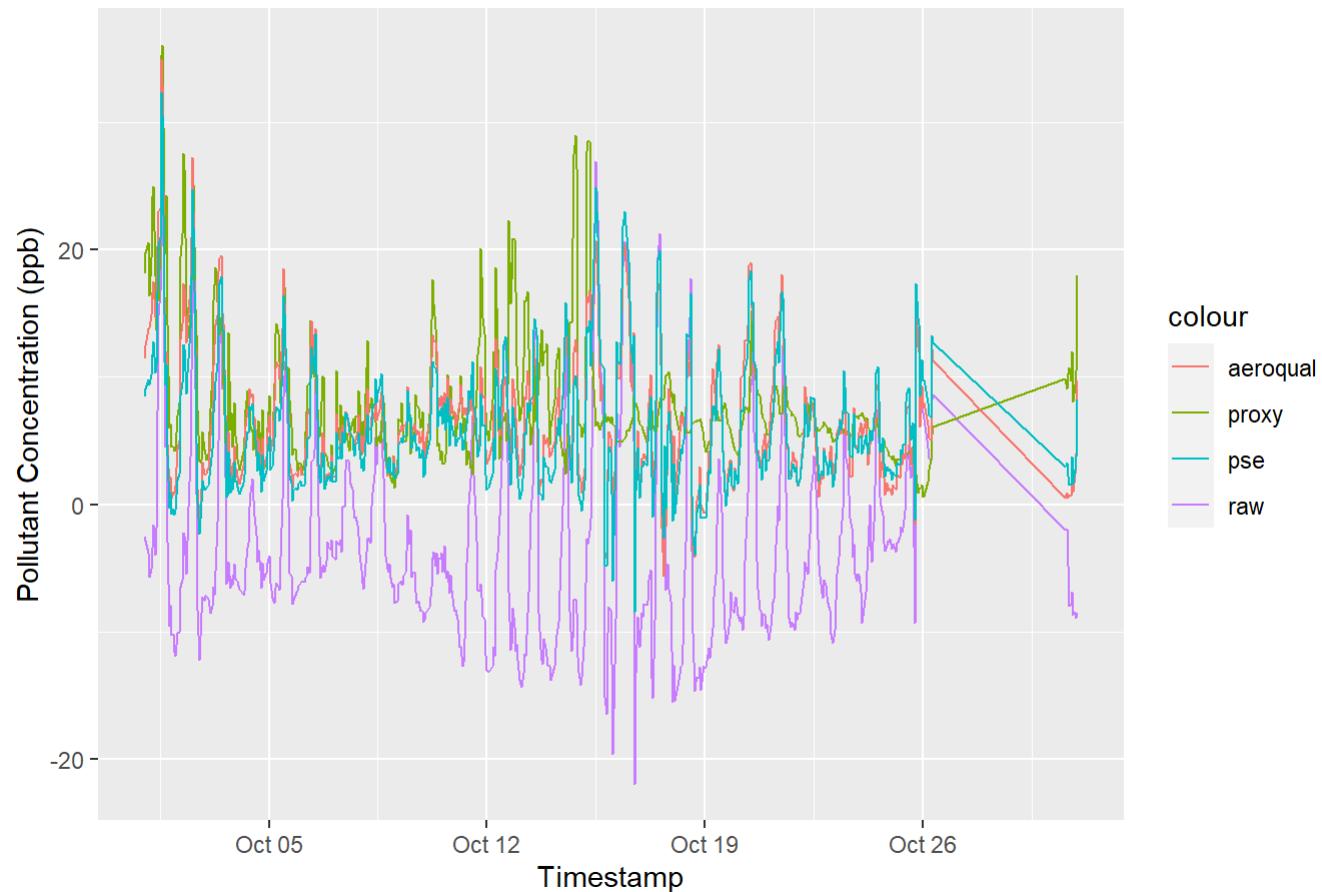
## Absolute Error for NO2 - AQY BB-633 month of 09



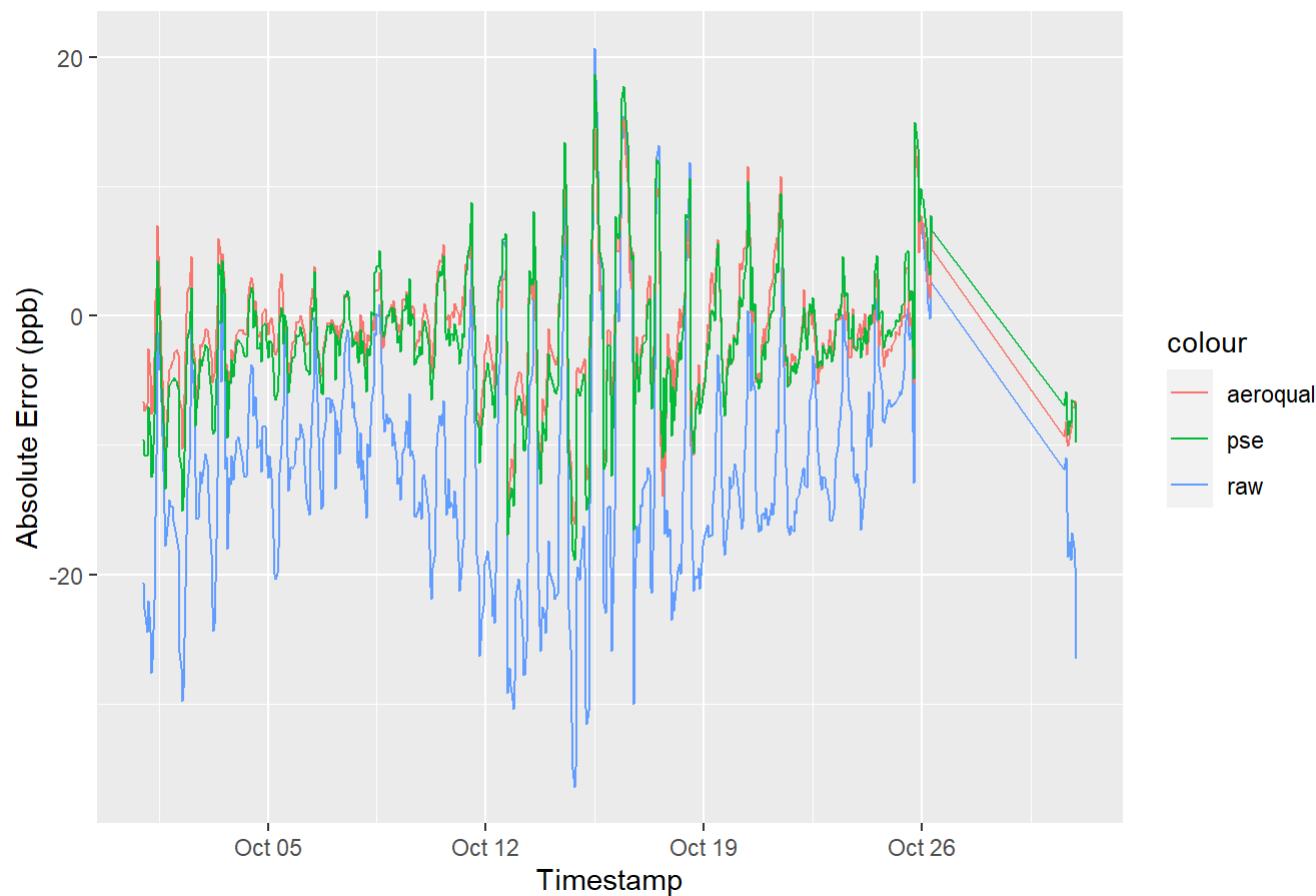
## NO2 [Proxy] vs [Measured] for AQY BB-633 month of 09



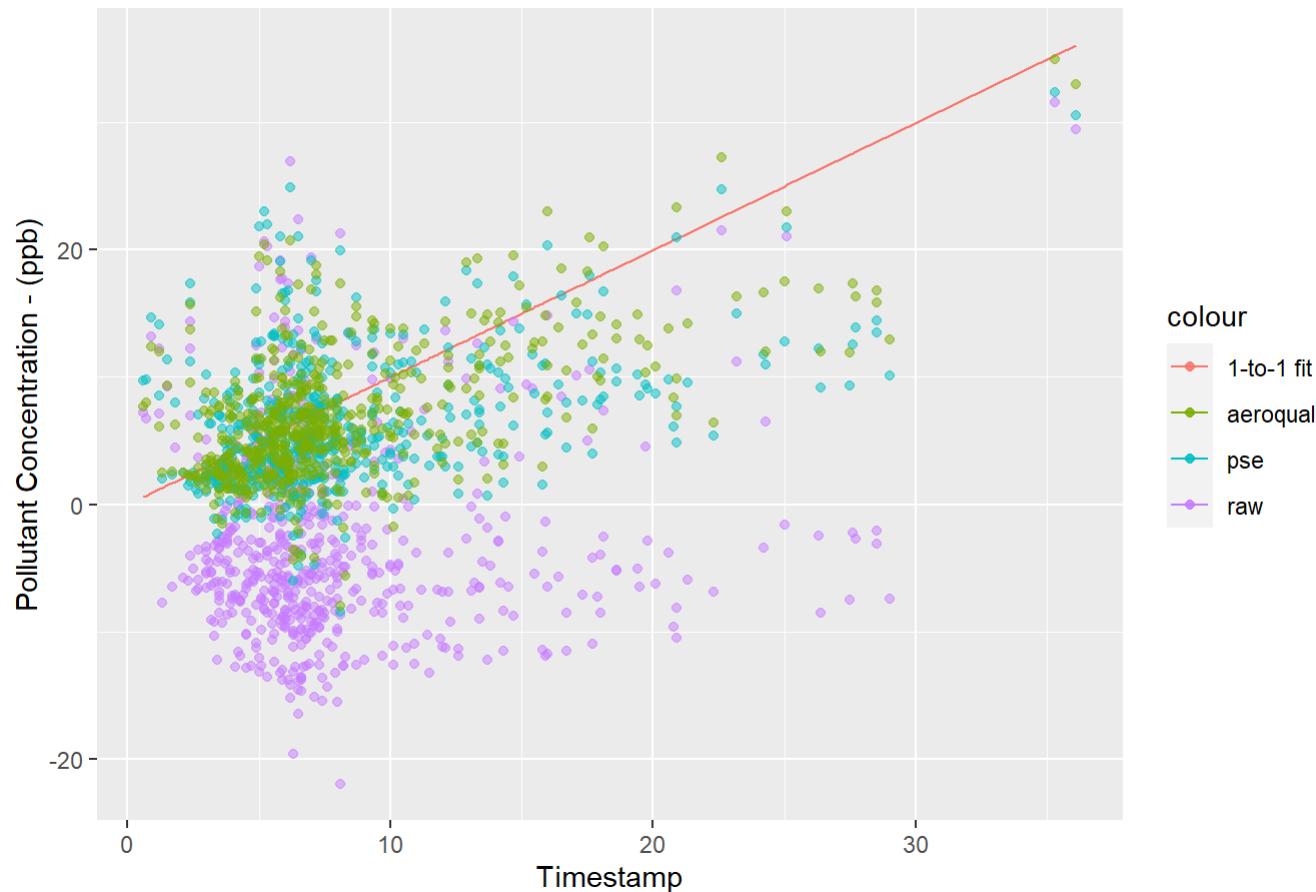
## NO2 Timeseries - AQY BB-633 Month of 10



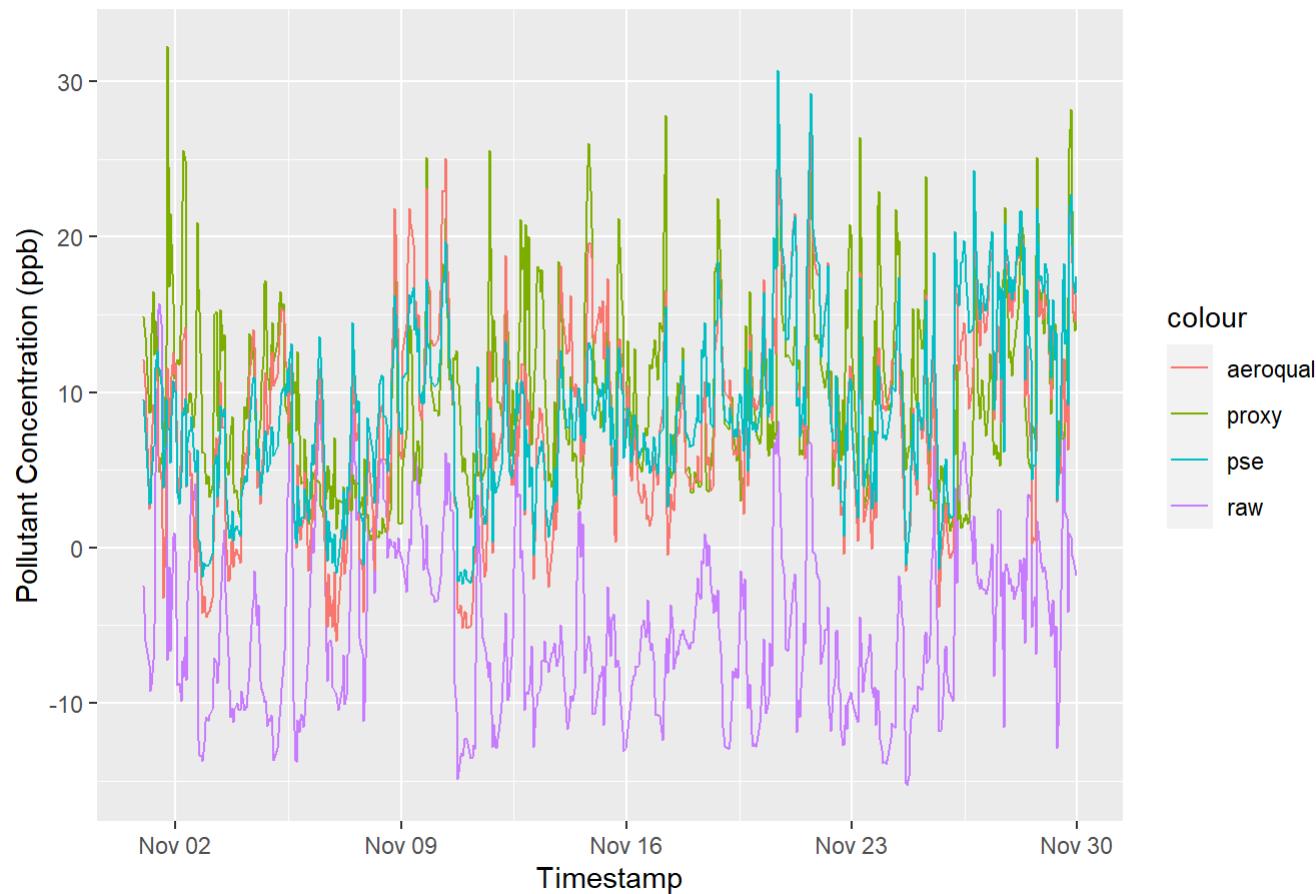
## Absolute Error for NO2 - AQY BB-633 month of 10



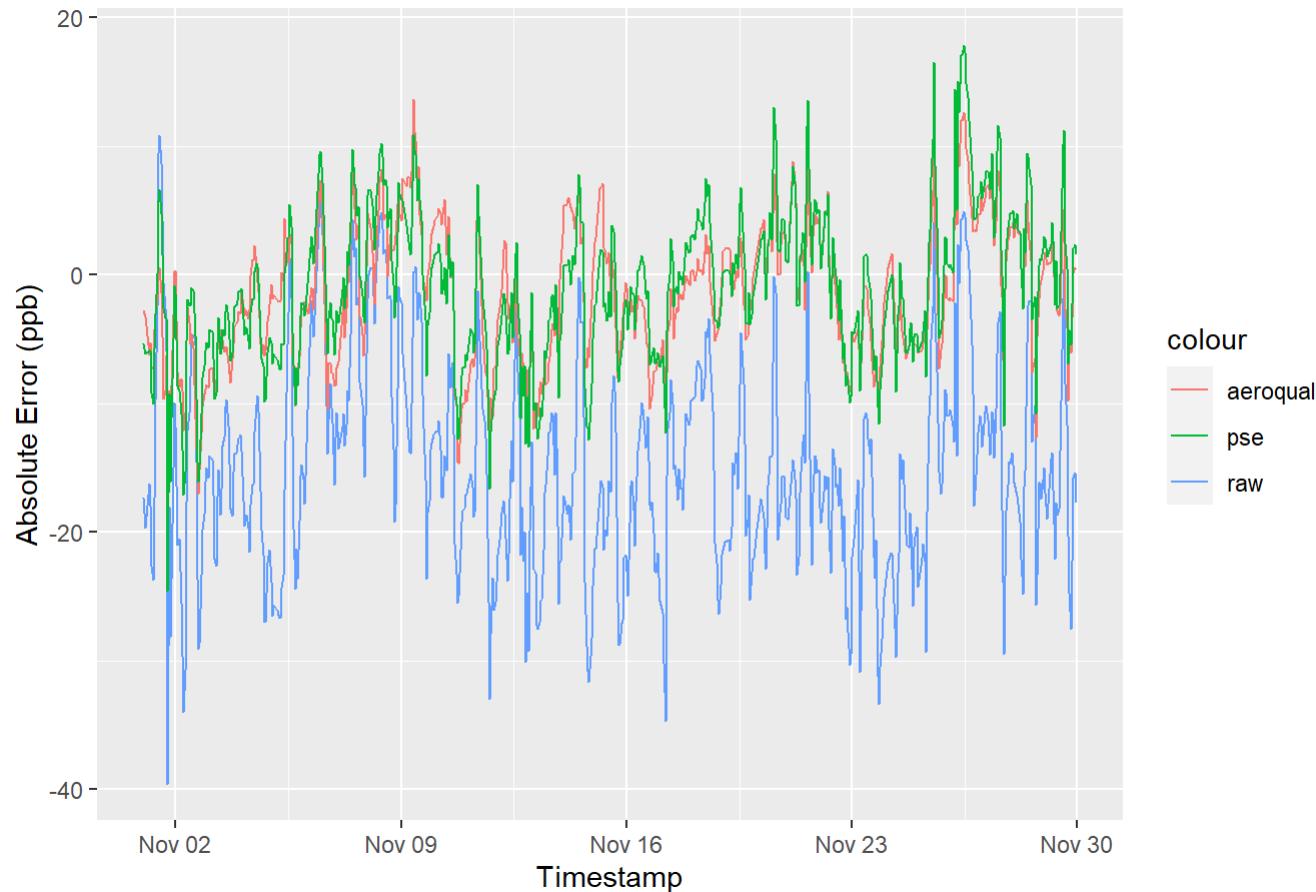
## NO2 [Proxy] vs [Measured] for AQY BB-633 month of 10



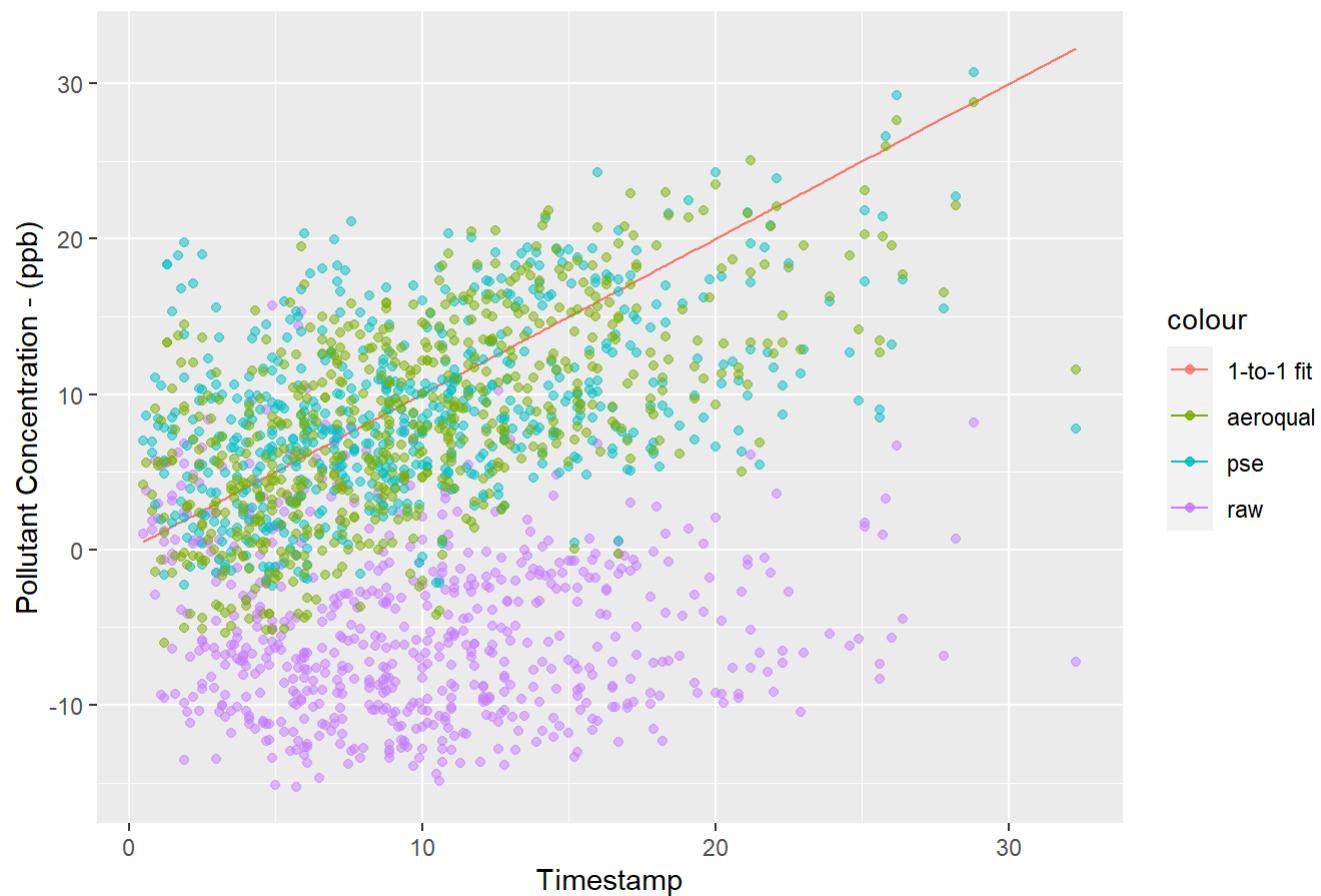
## NO2 Timeseries - AQY BB-633 Month of 11



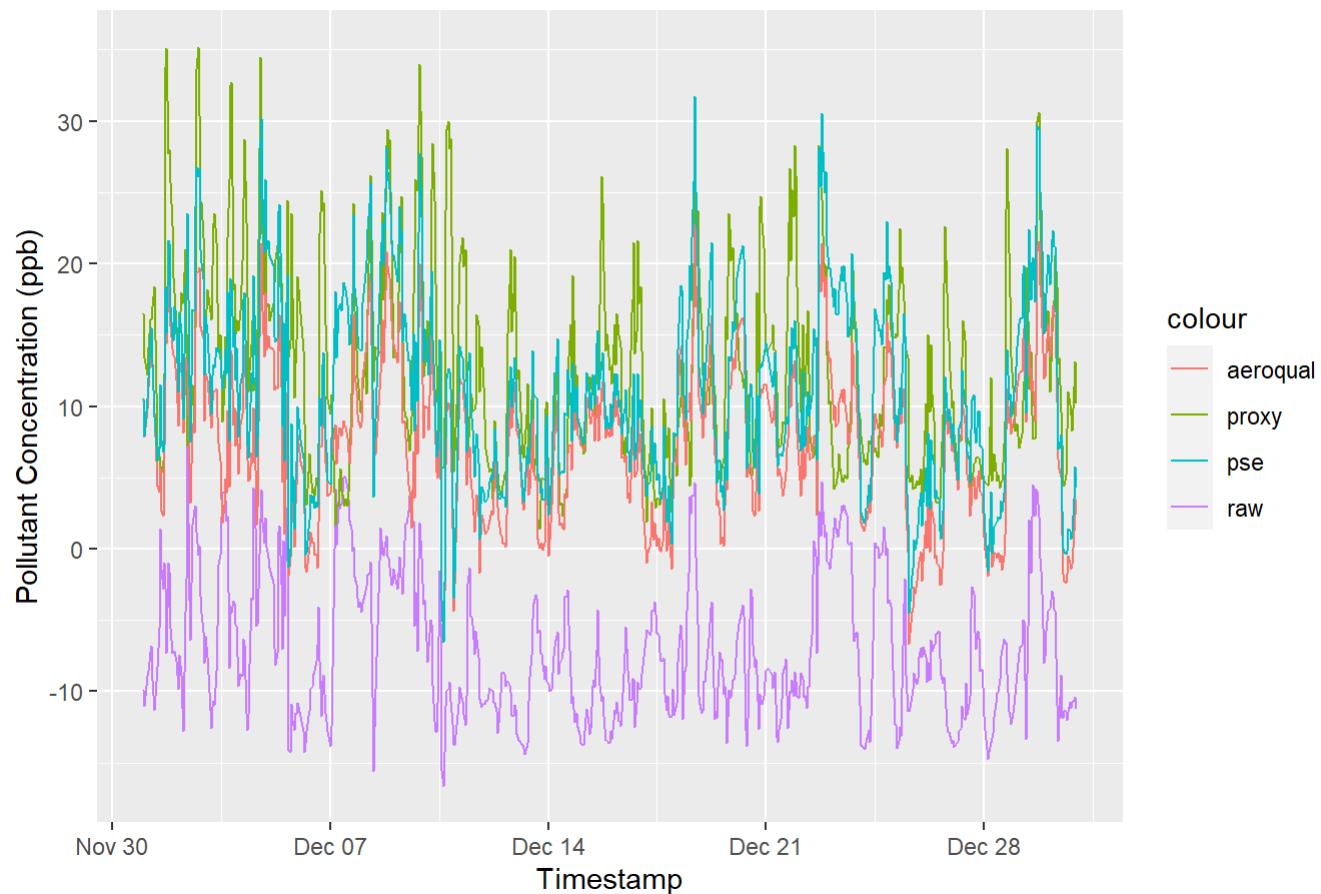
## Absolute Error for NO2 - AQY BB-633 month of 11



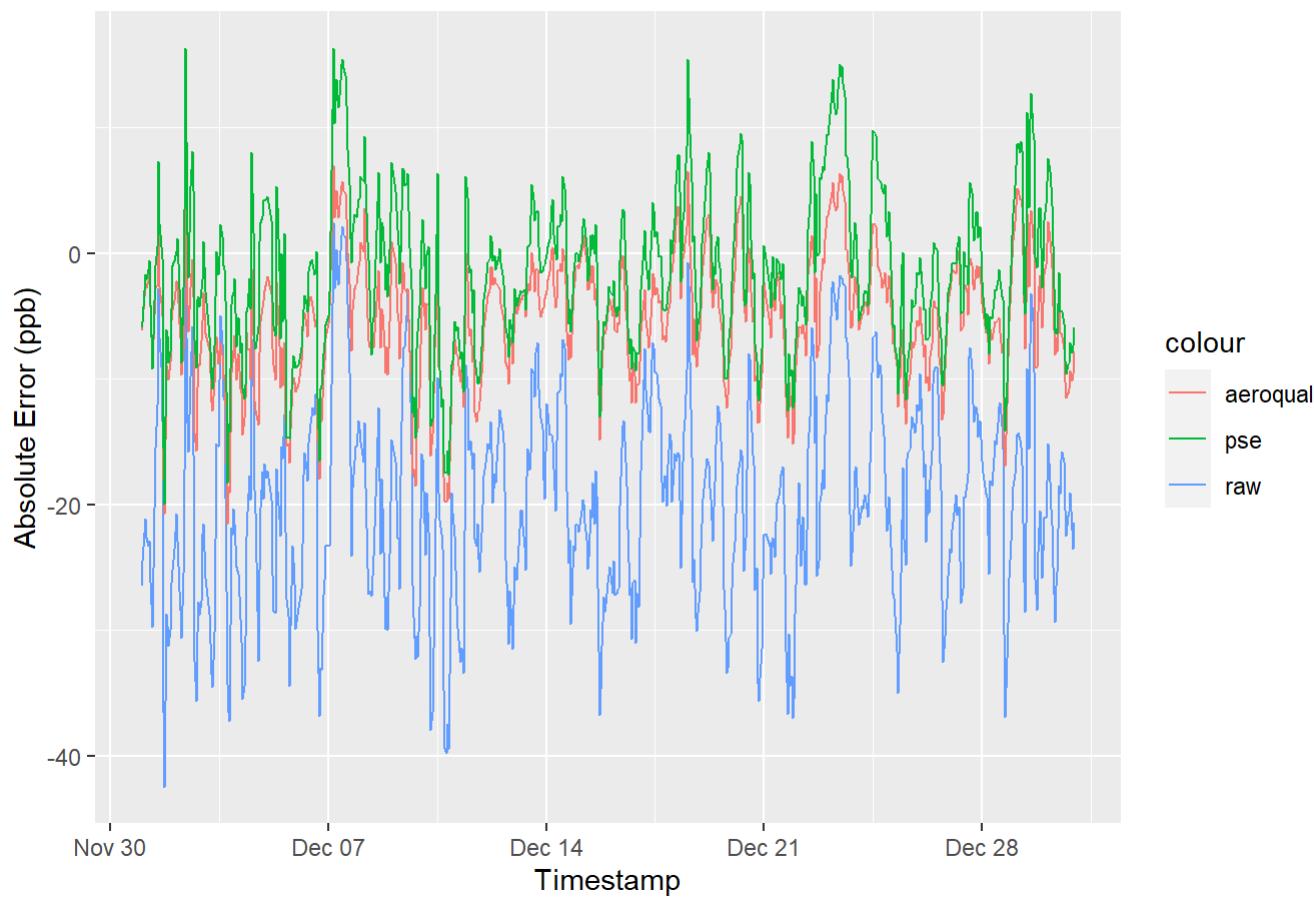
## NO2 [Proxy] vs [Measured] for AQY BB-633 month of 11



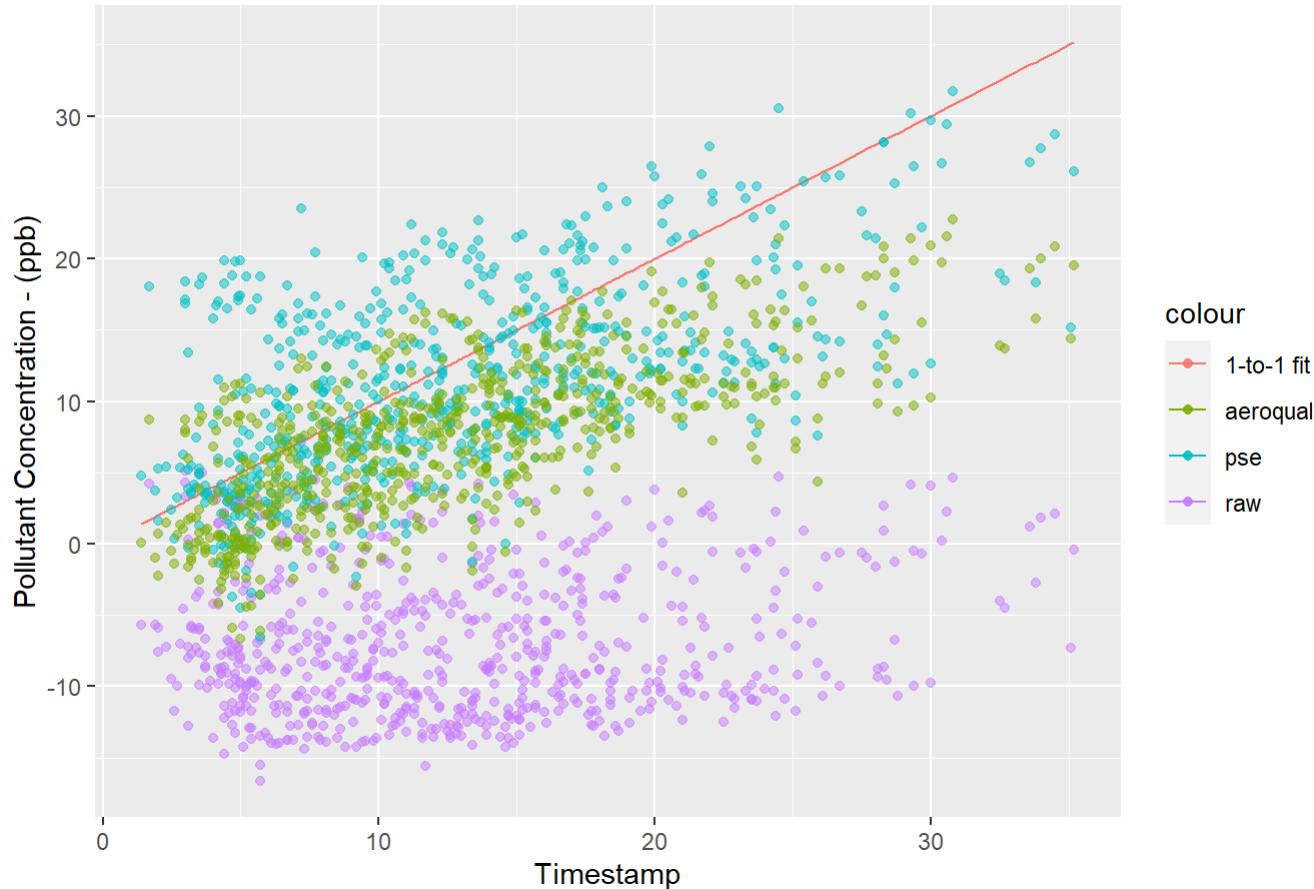
## NO2 Timeseries - AQY BB-633 Month of 12



## Absolute Error for NO2 - AQY BB-633 month of 12

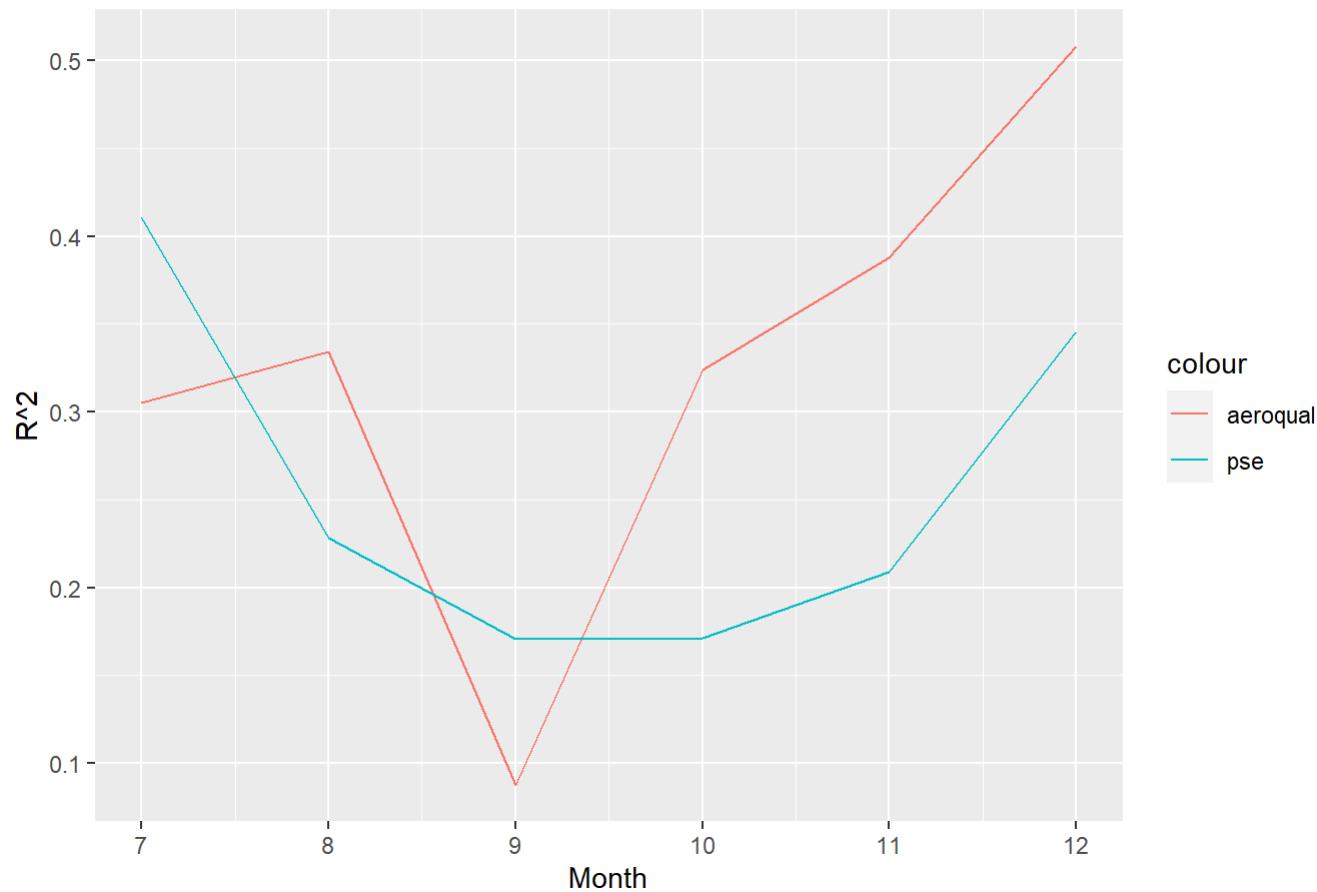


## NO2 [Proxy] vs [Measured] for AQY BB-633 month of 12

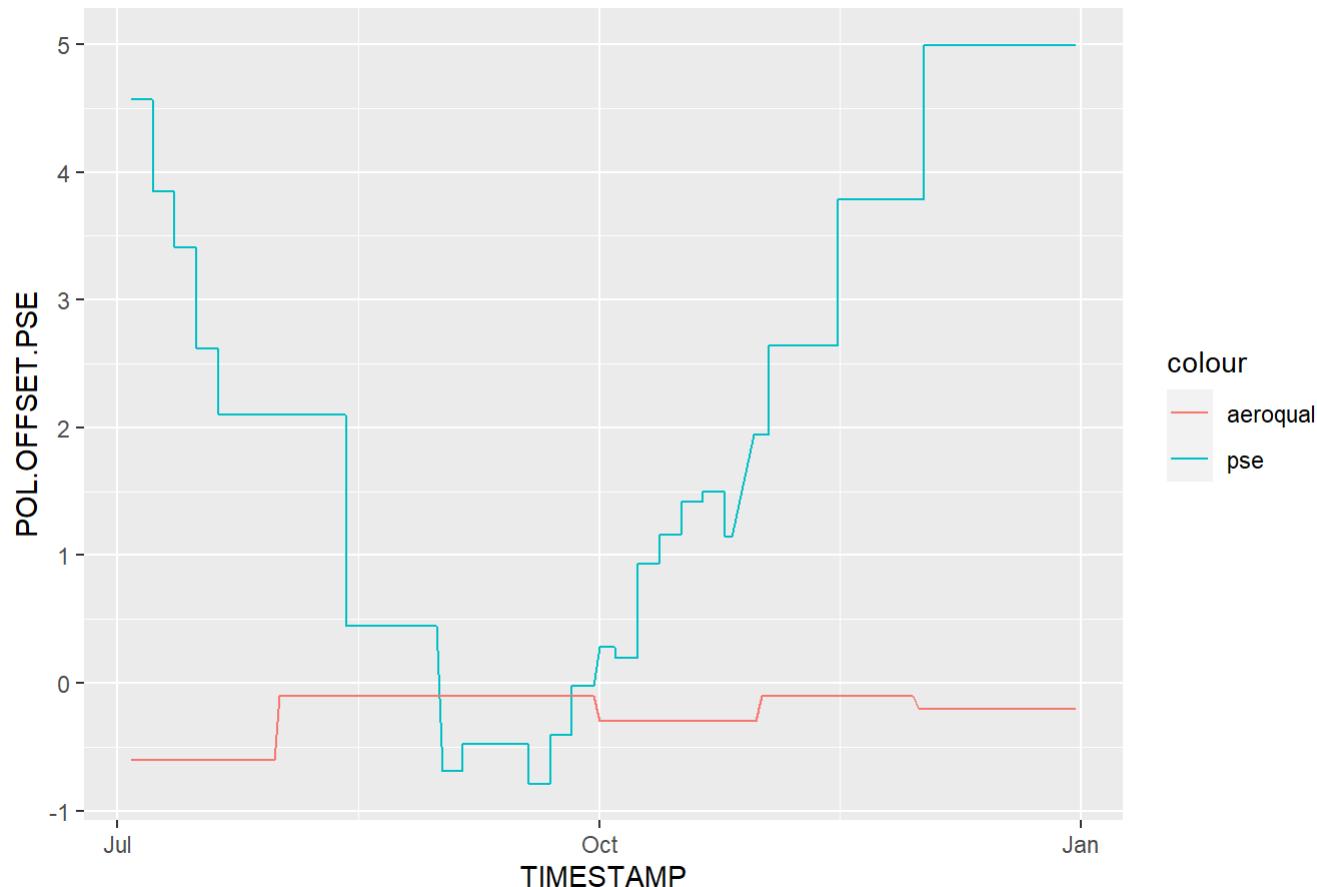


```
##          month      raw      PSE      AQ
## 1 Full Deployment 0.00209 0.36109 0.34982
## 2              07 0.08326 0.41048 0.30509
## 3              08 0.07389 0.2281  0.3344
## 4              09 0.01097 0.17064  0.0877
## 5              10 0.01211 0.17098 0.32383
## 6              11 0.00297 0.20912 0.38802
## 7              12 0.01048 0.34544 0.50841
```

## NO2 Proxy ~ Measured R^2 Over Time for AQY BB-633

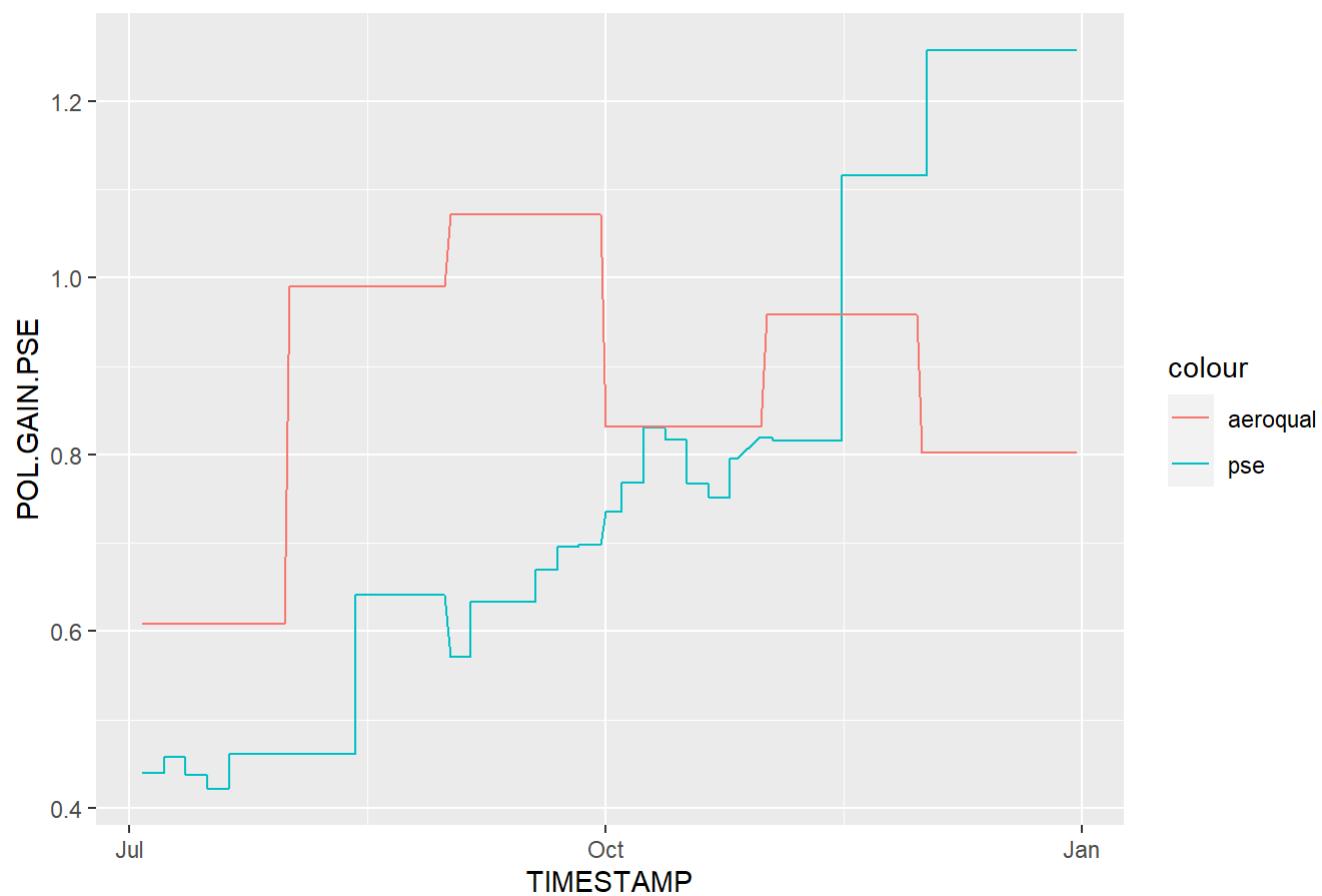


## NO2 Offset Over Time for AQY BB-633



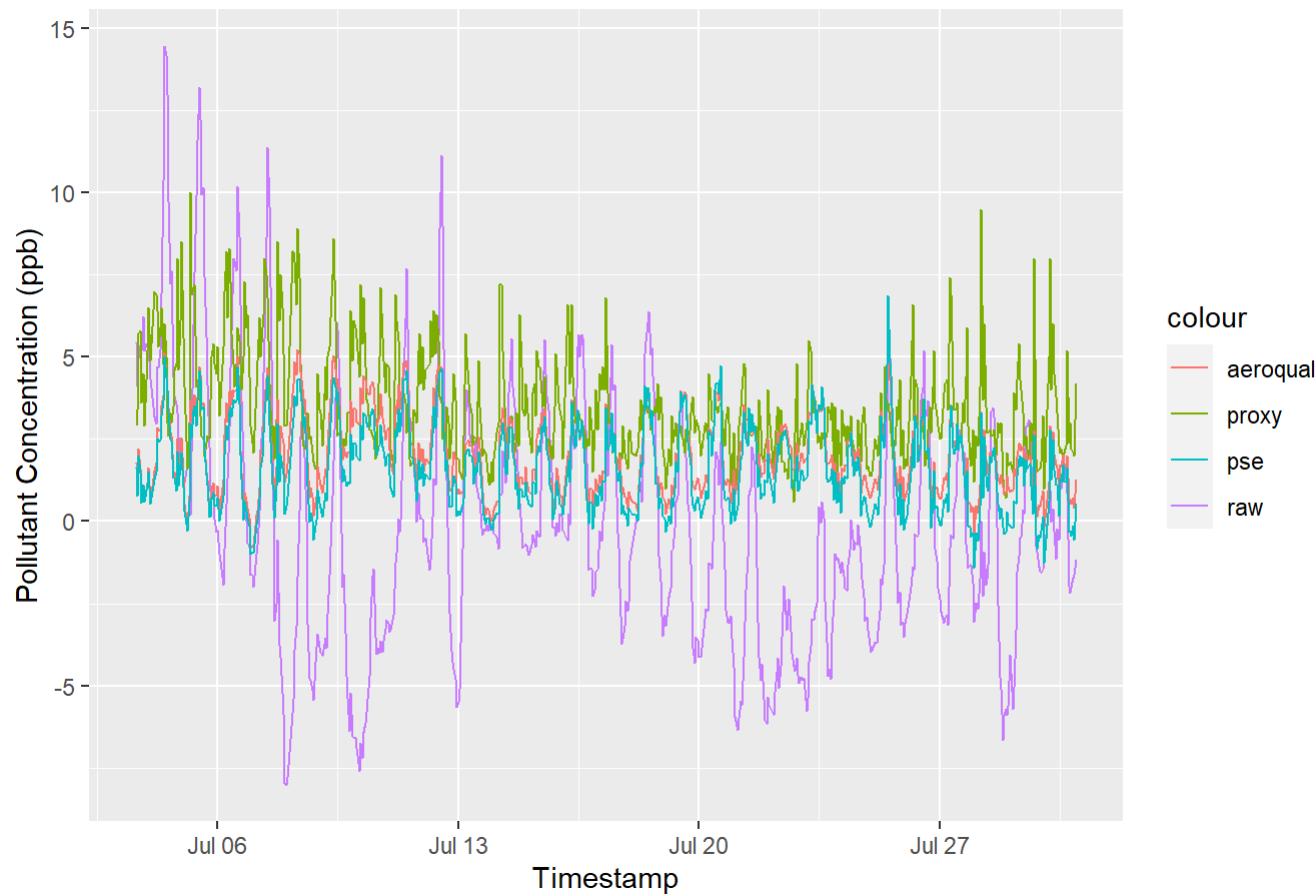
## NO2 Gain Over Time for AQY BB-633

## Compare Results

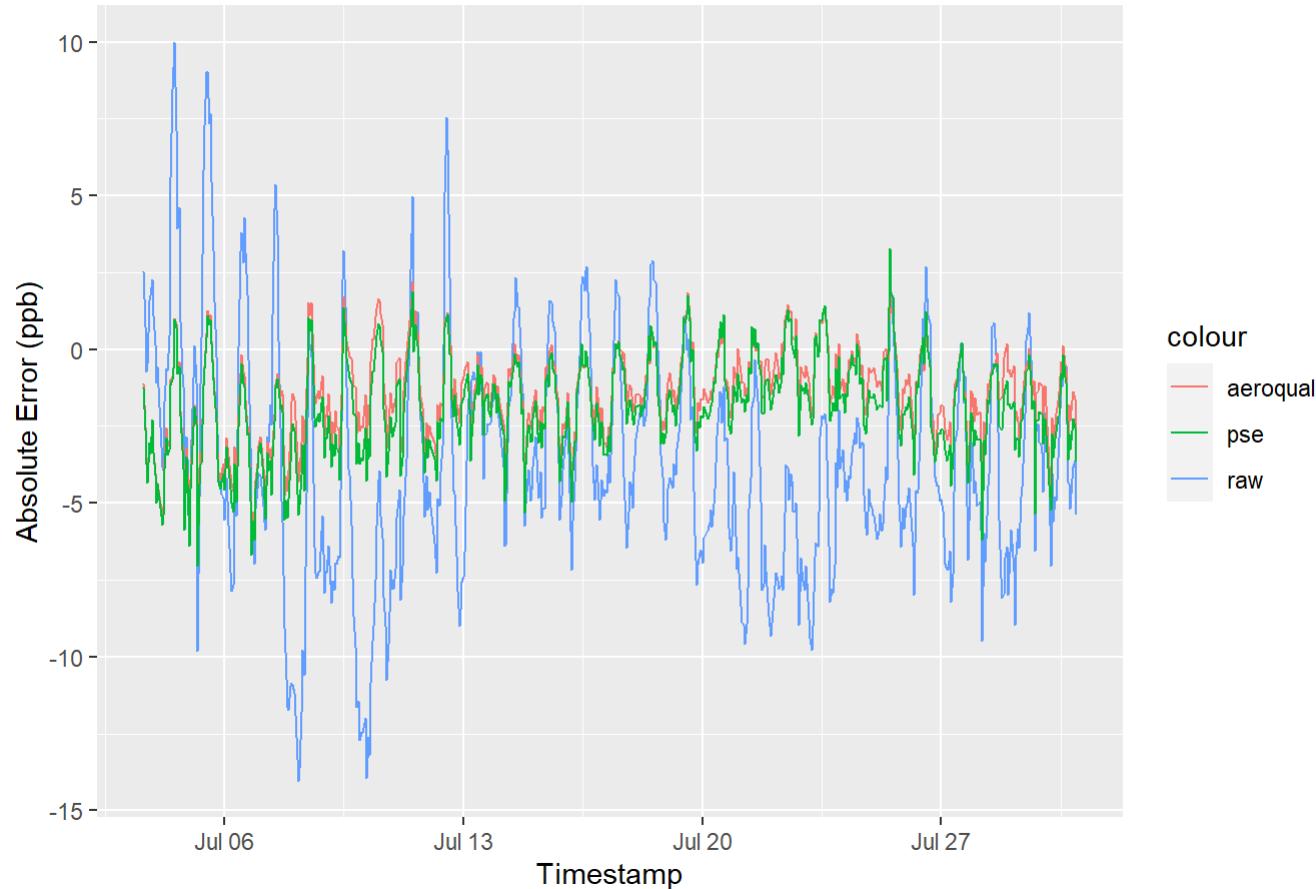


```
make_all_plots(NO2_data, 'AQY BB-642', 'NO2')
```

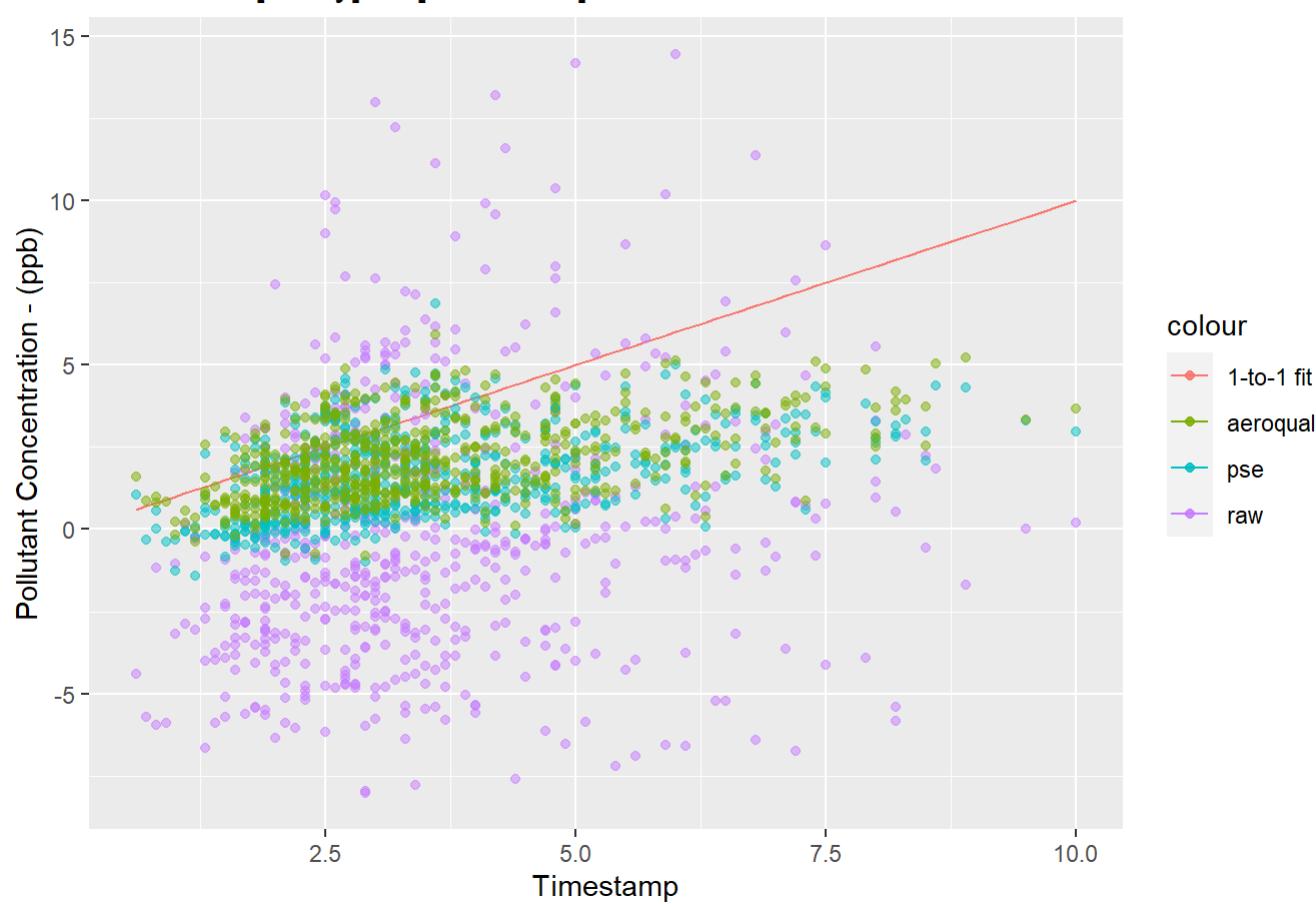
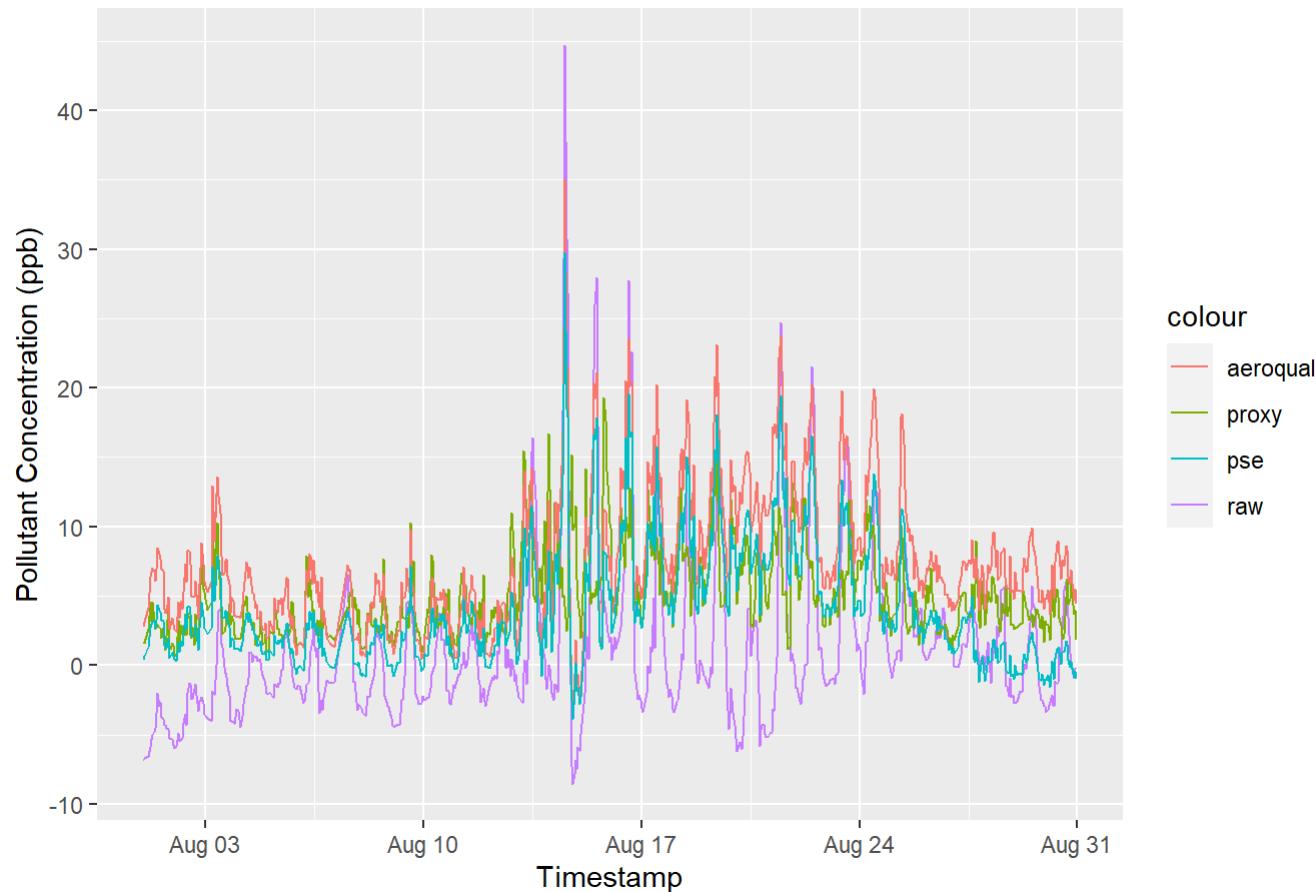
### NO2 Timeseries - AQY BB-642 Month of 07

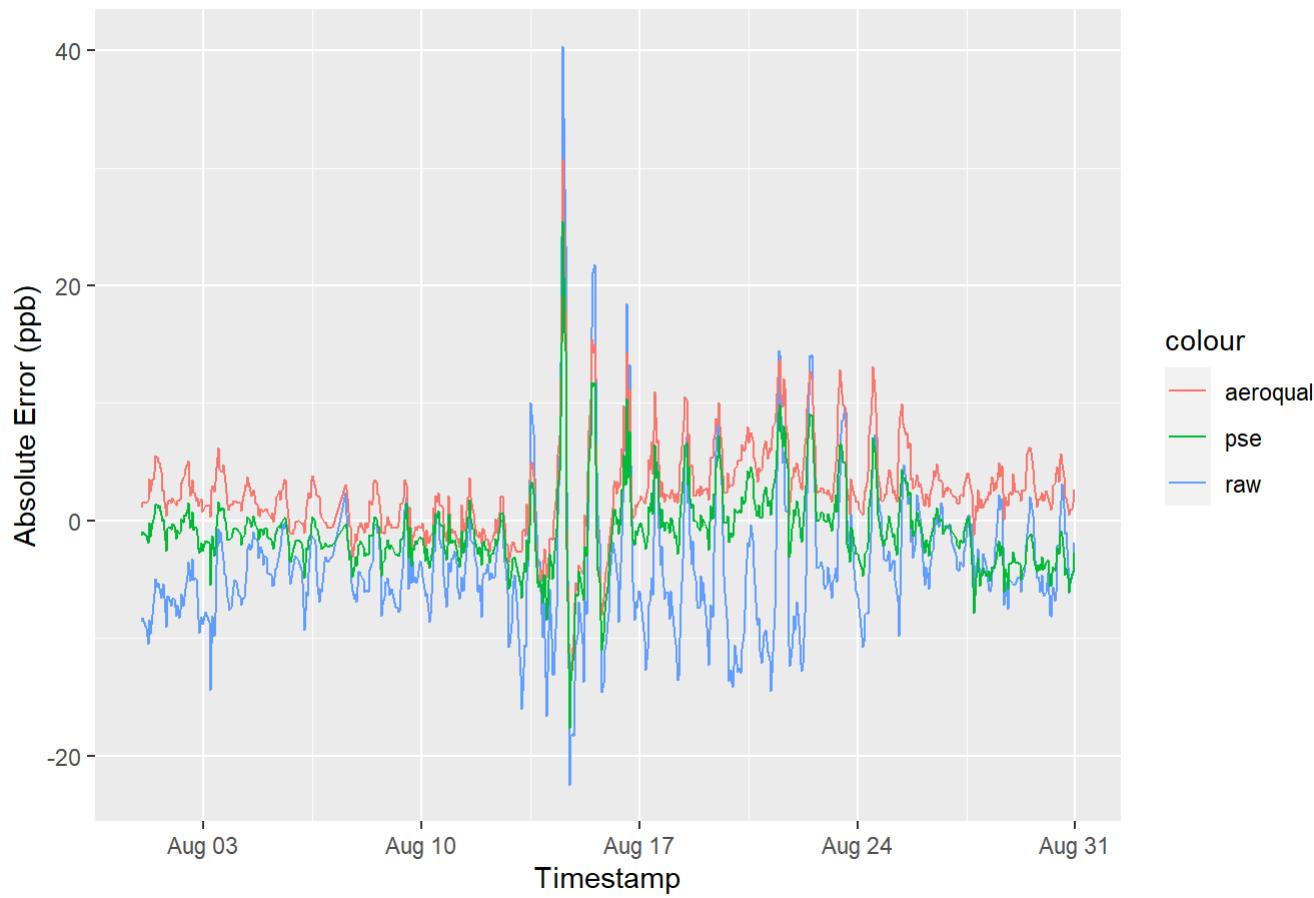
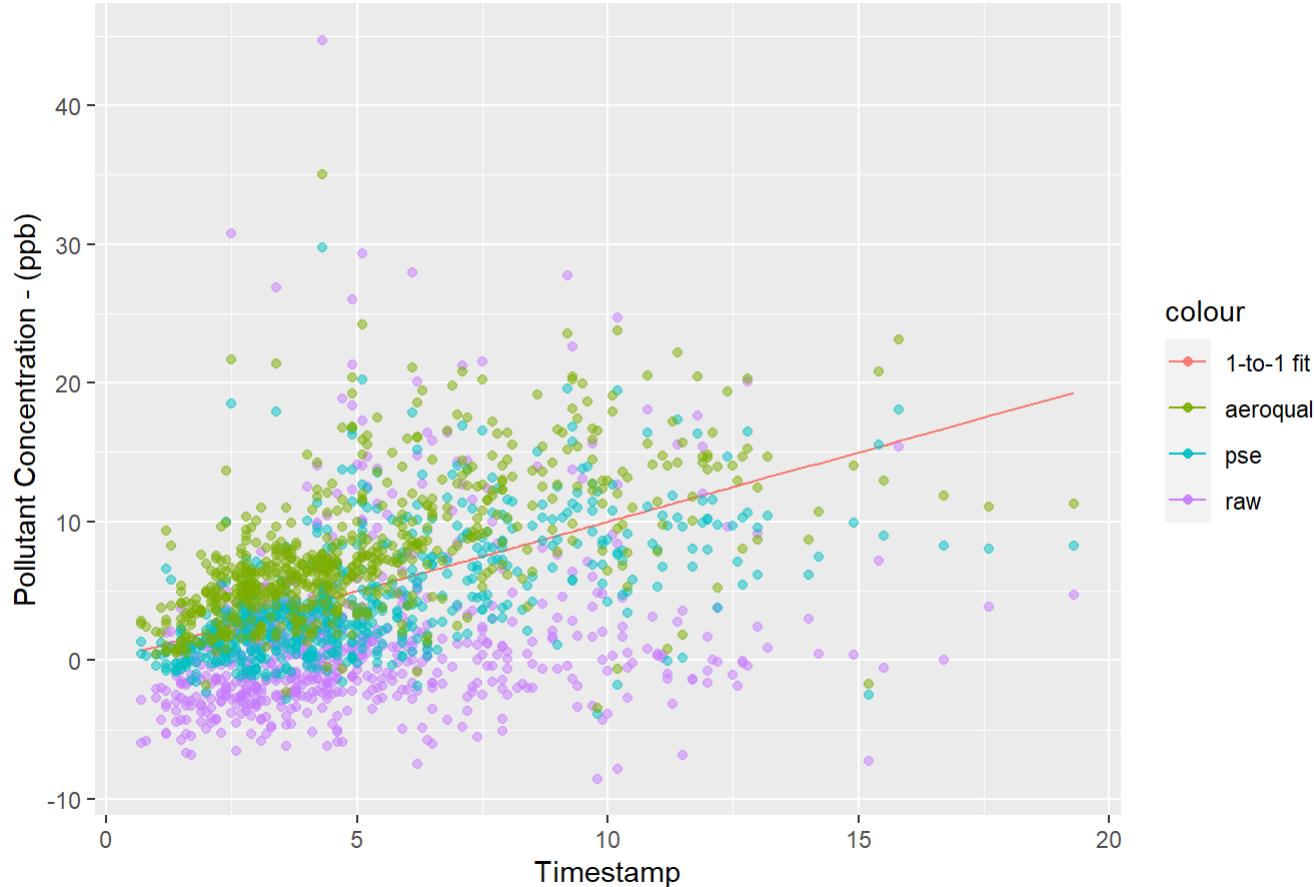


### Absolute Error for NO2 - AQY BB-642 month of 07

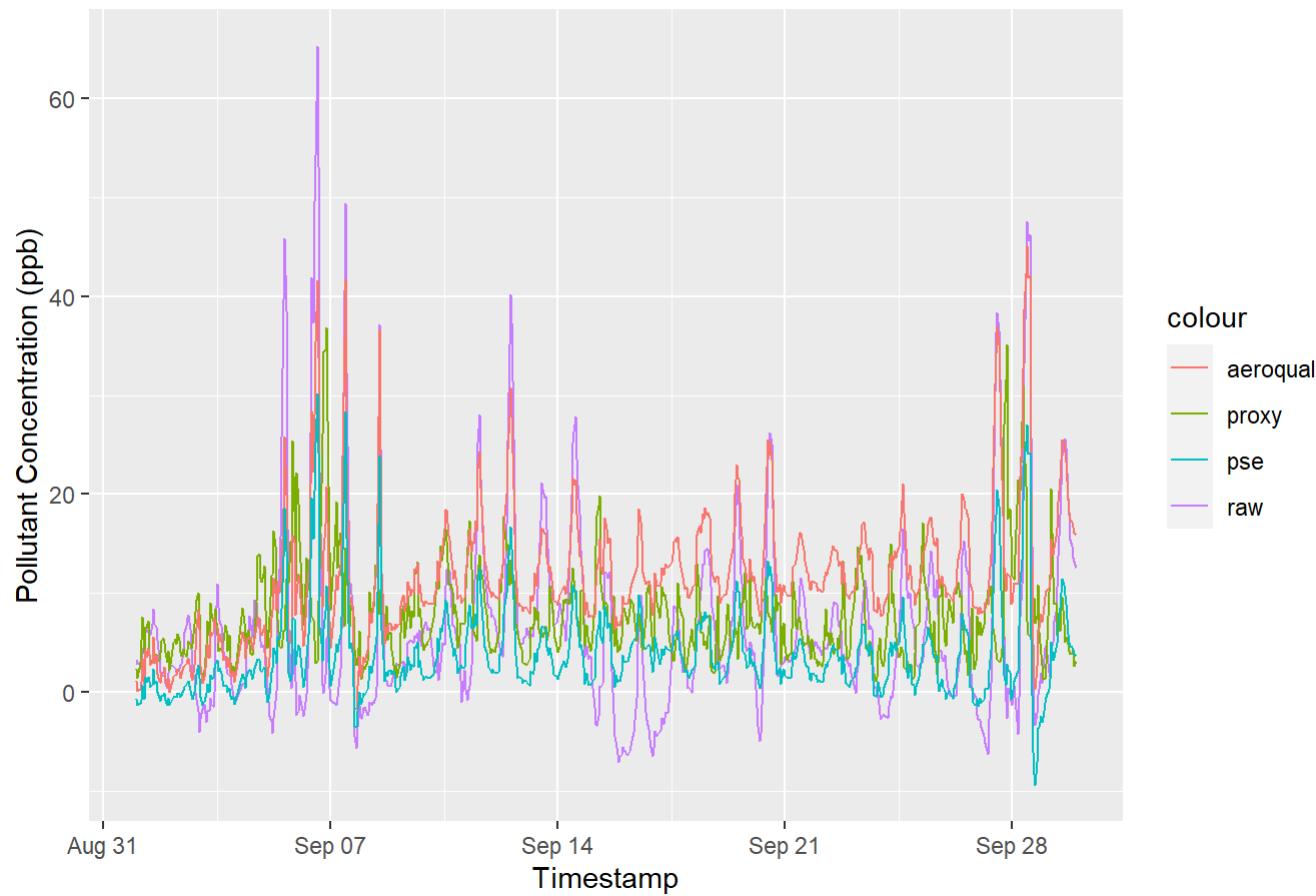


### NO2 [Proxy] vs [Measured] for AQY BB-642 month of 07

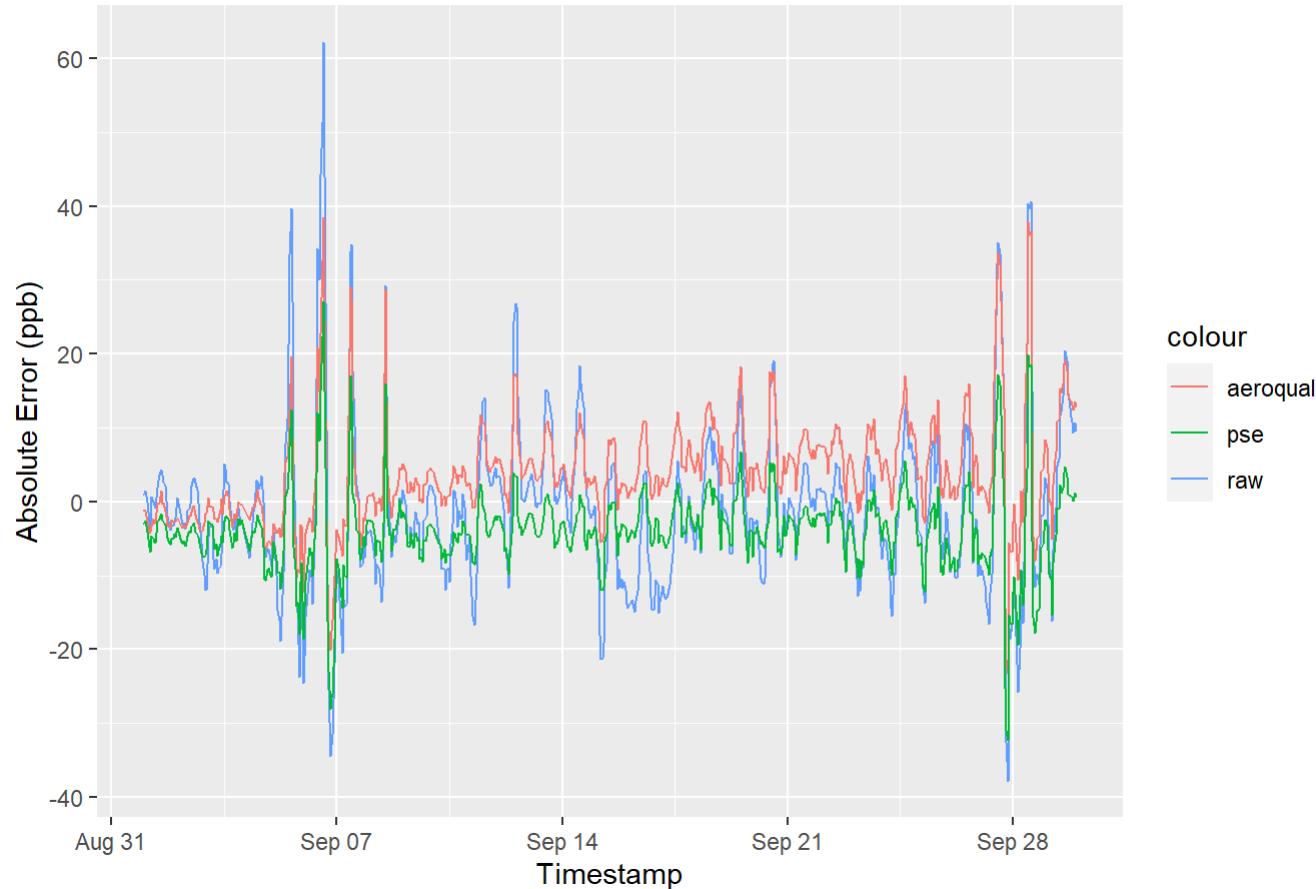
NO<sub>2</sub> Timeseries - AQY BB-642 Month of 08

Absolute Error for NO<sub>2</sub> - AQY BB-642 month of 08NO<sub>2</sub> [Proxy] vs [Measured] for AQY BB-642 month of 08

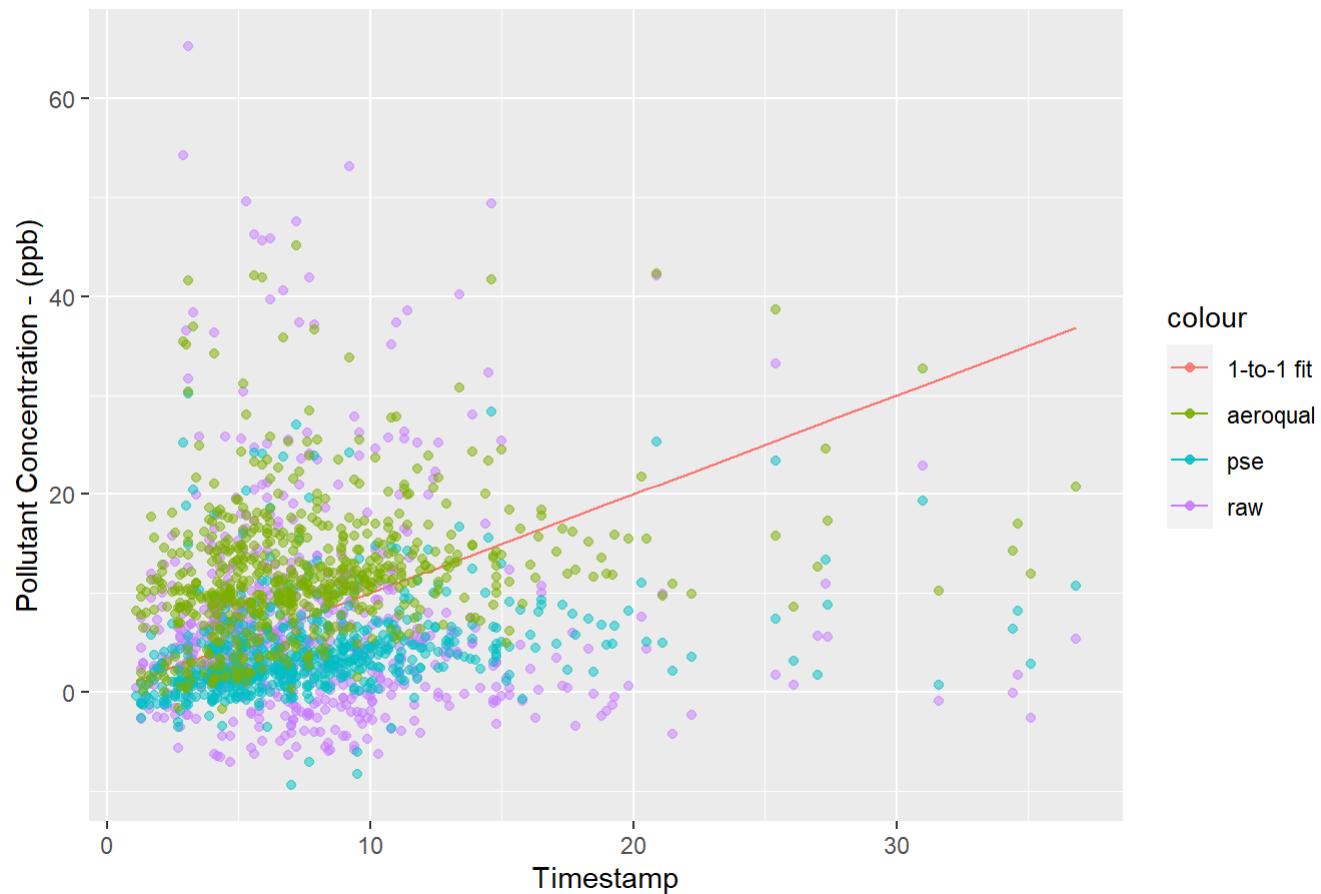
## NO2 Timeseries - AQY BB-642 Month of 09



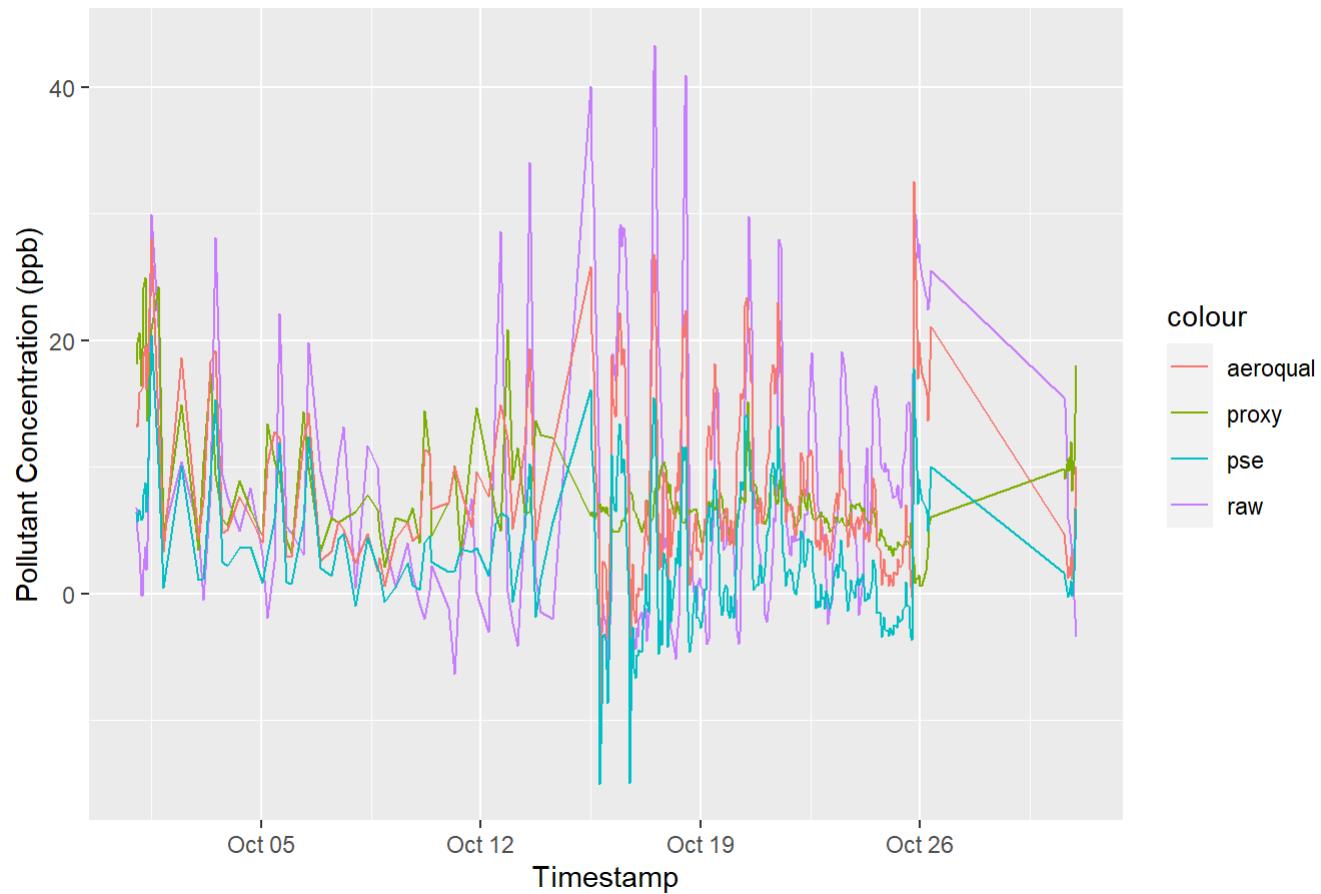
## Absolute Error for NO2 - AQY BB-642 month of 09



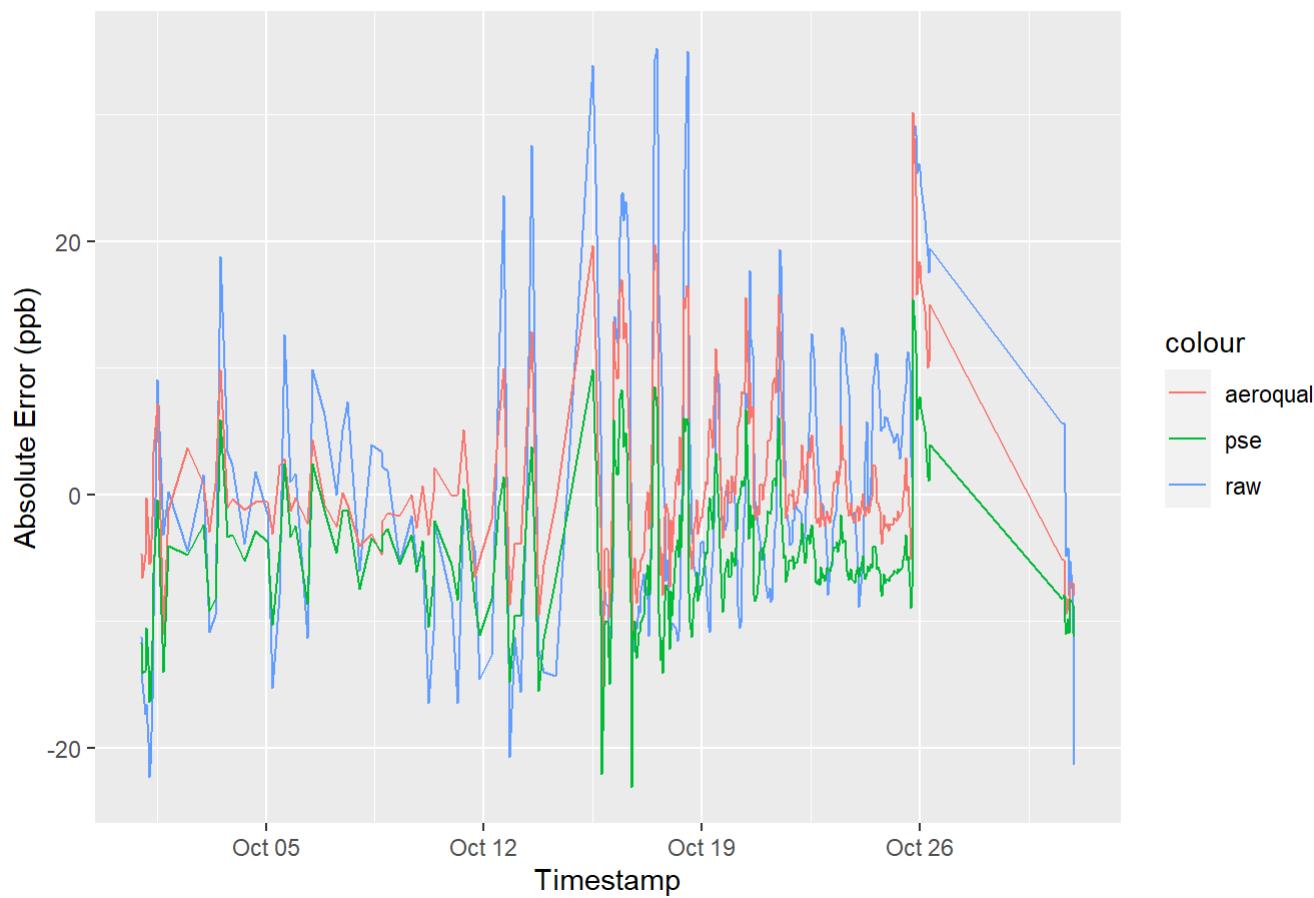
## NO2 [Proxy] vs [Measured] for AQY BB-642 month of 09



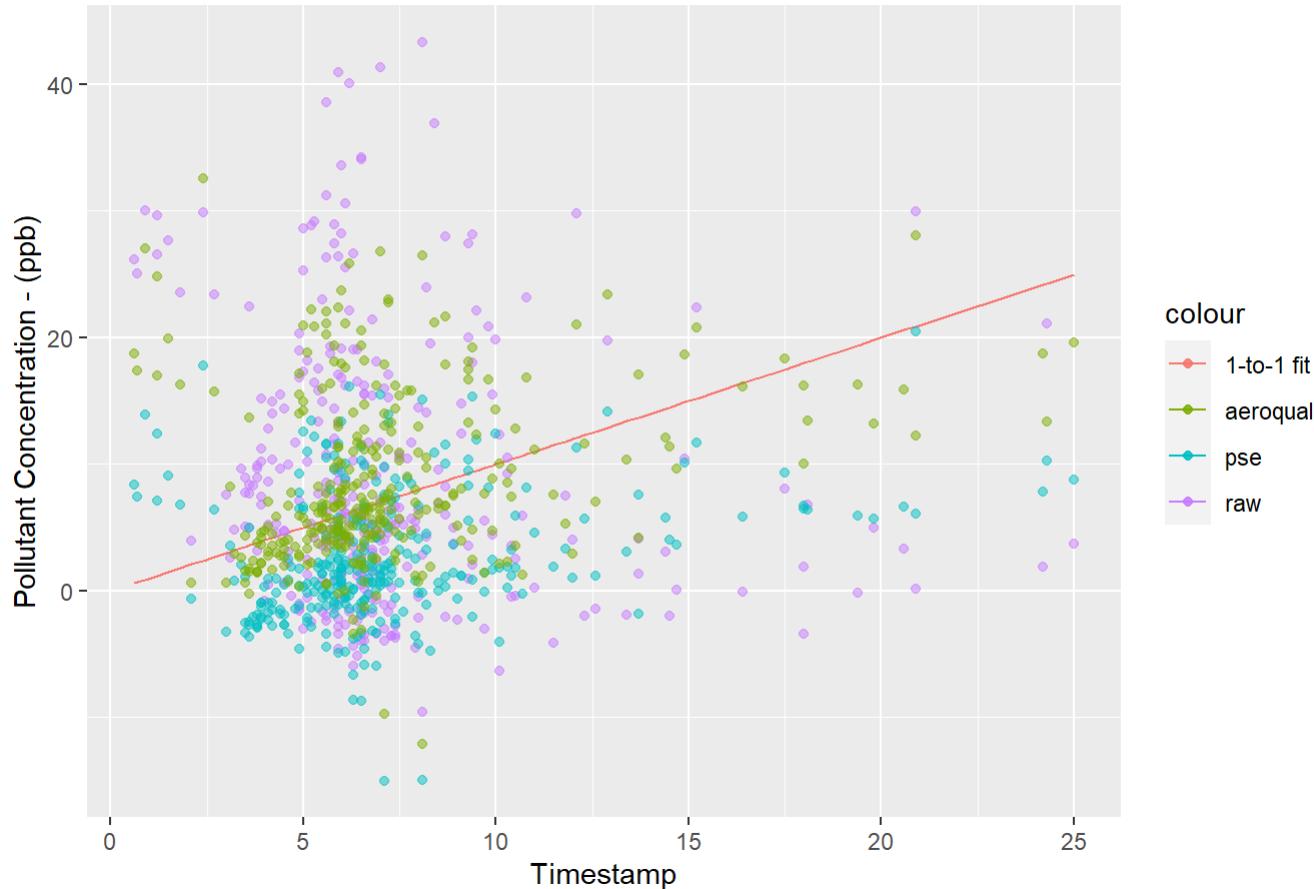
## NO2 Timeseries - AQY BB-642 Month of 10



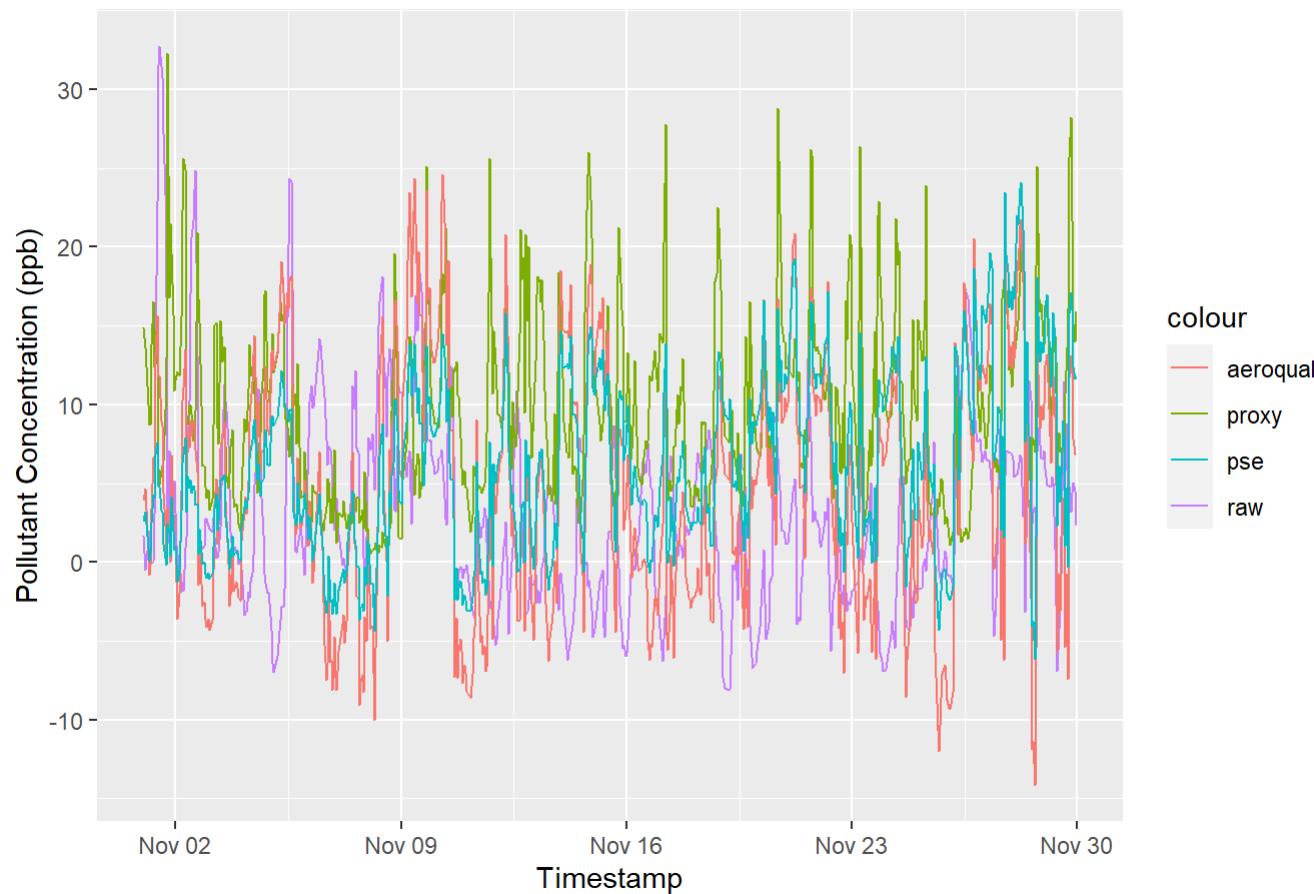
## Absolute Error for NO2 - AQY BB-642 month of 10



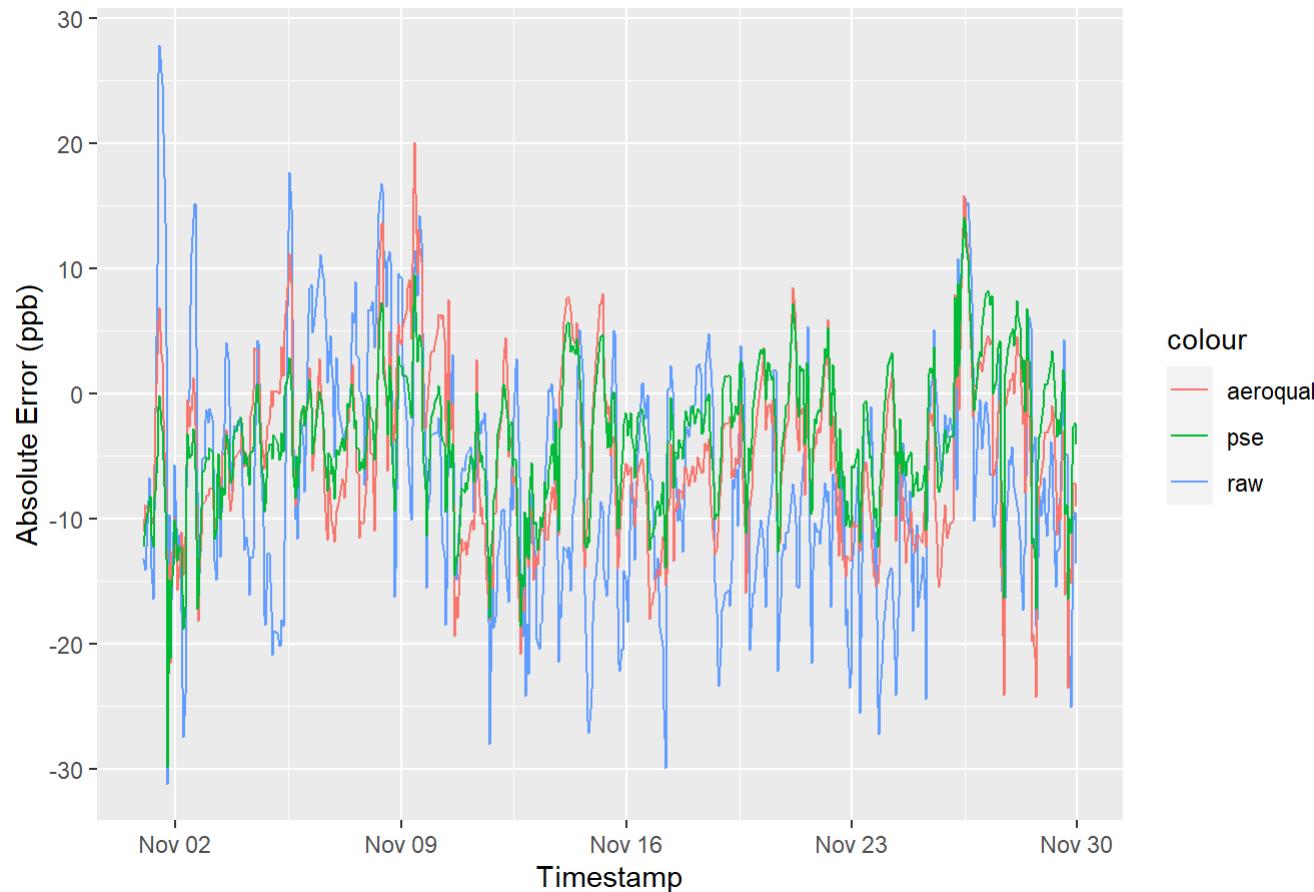
## NO2 [Proxy] vs [Measured] for AQY BB-642 month of 10



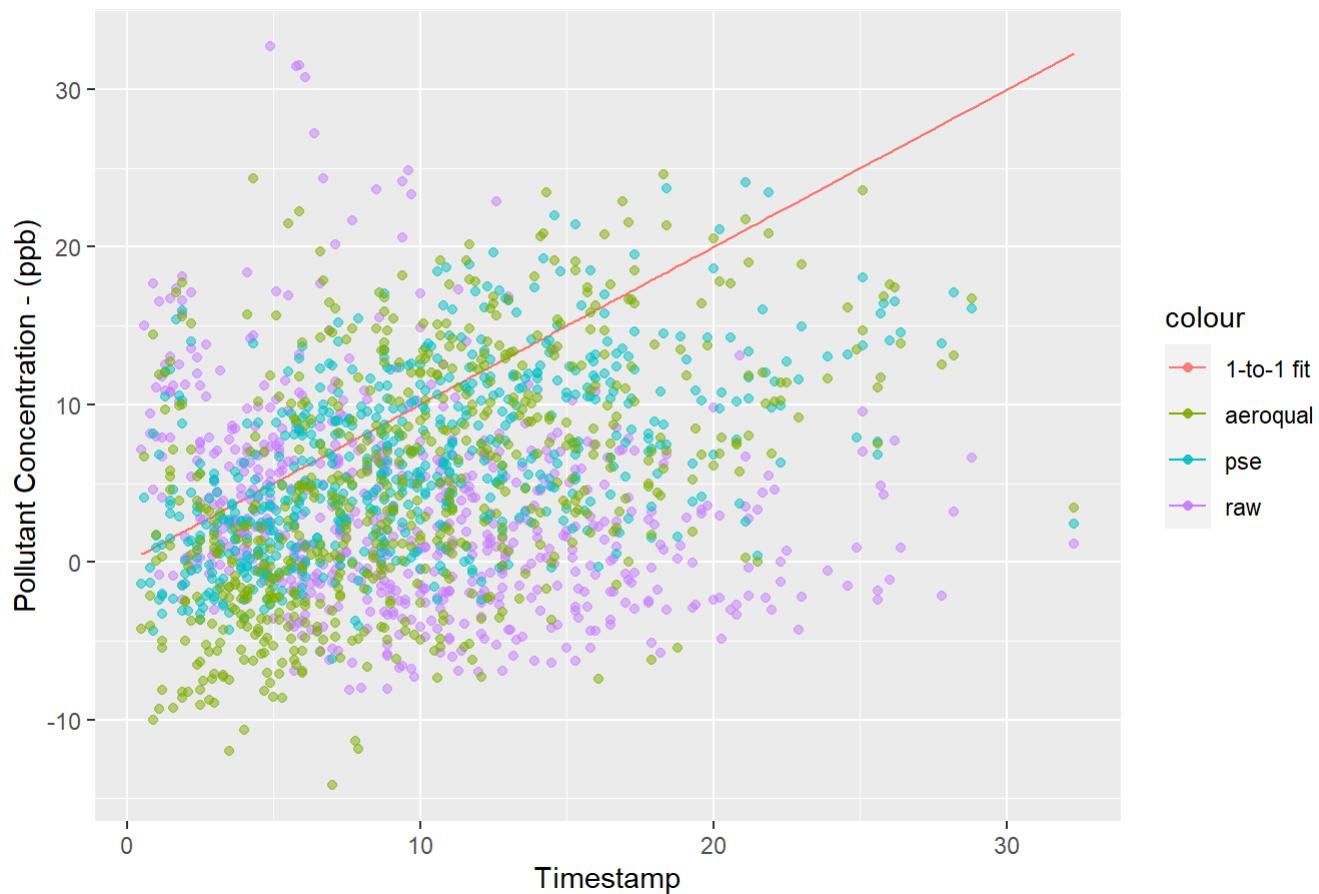
## NO2 Timeseries - AQY BB-642 Month of 11



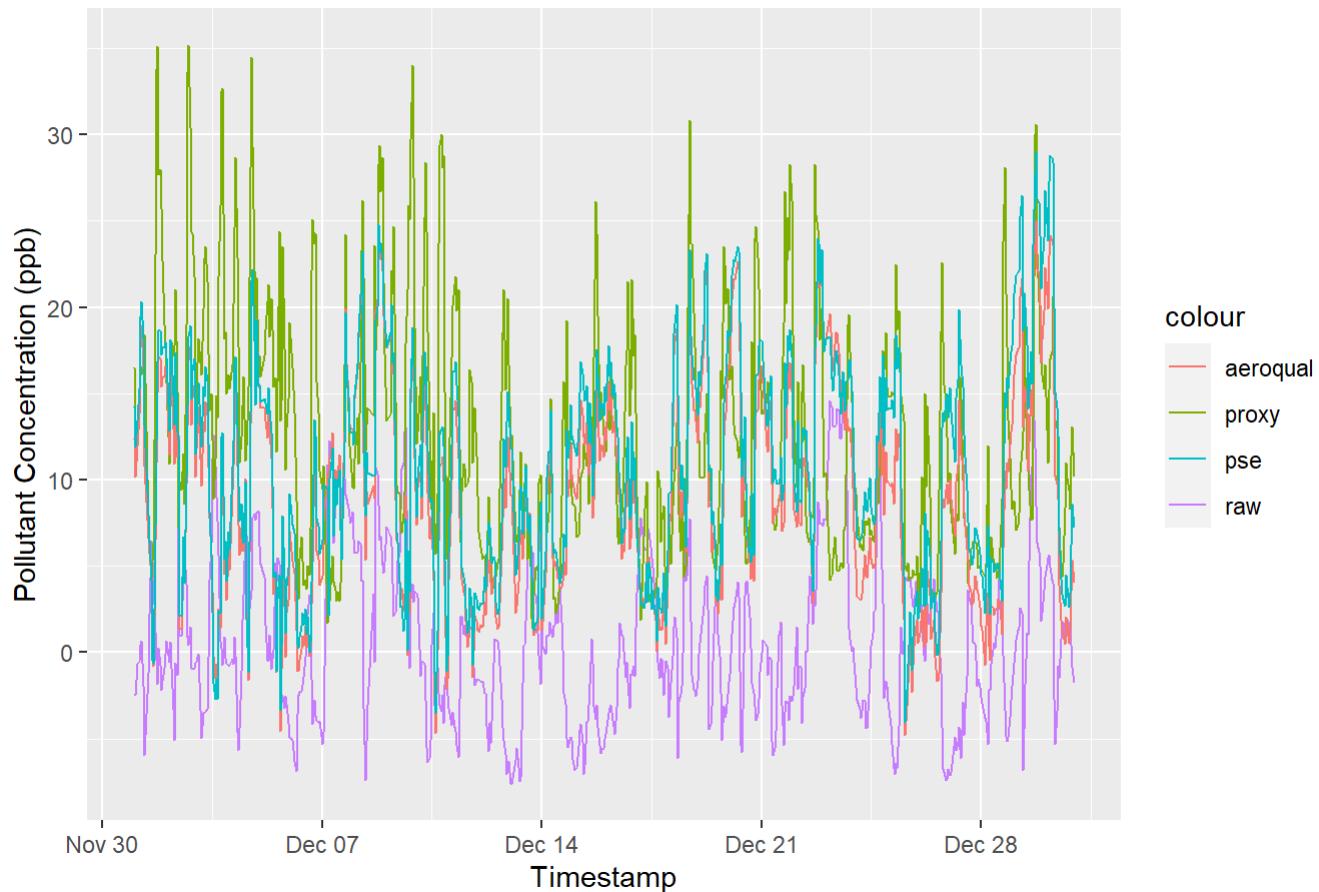
## Absolute Error for NO2 - AQY BB-642 month of 11



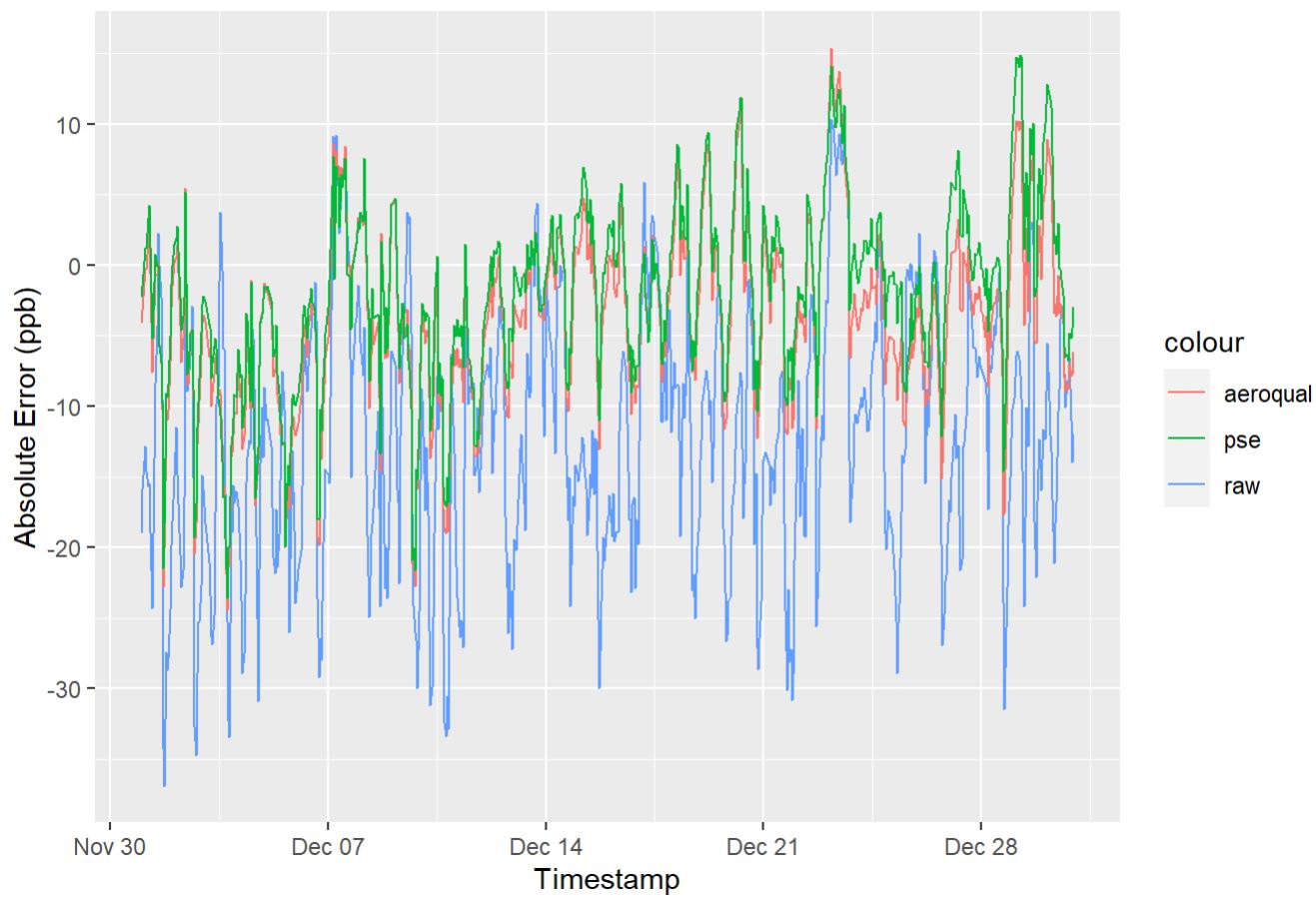
## NO2 [Proxy] vs [Measured] for AQY BB-642 month of 11



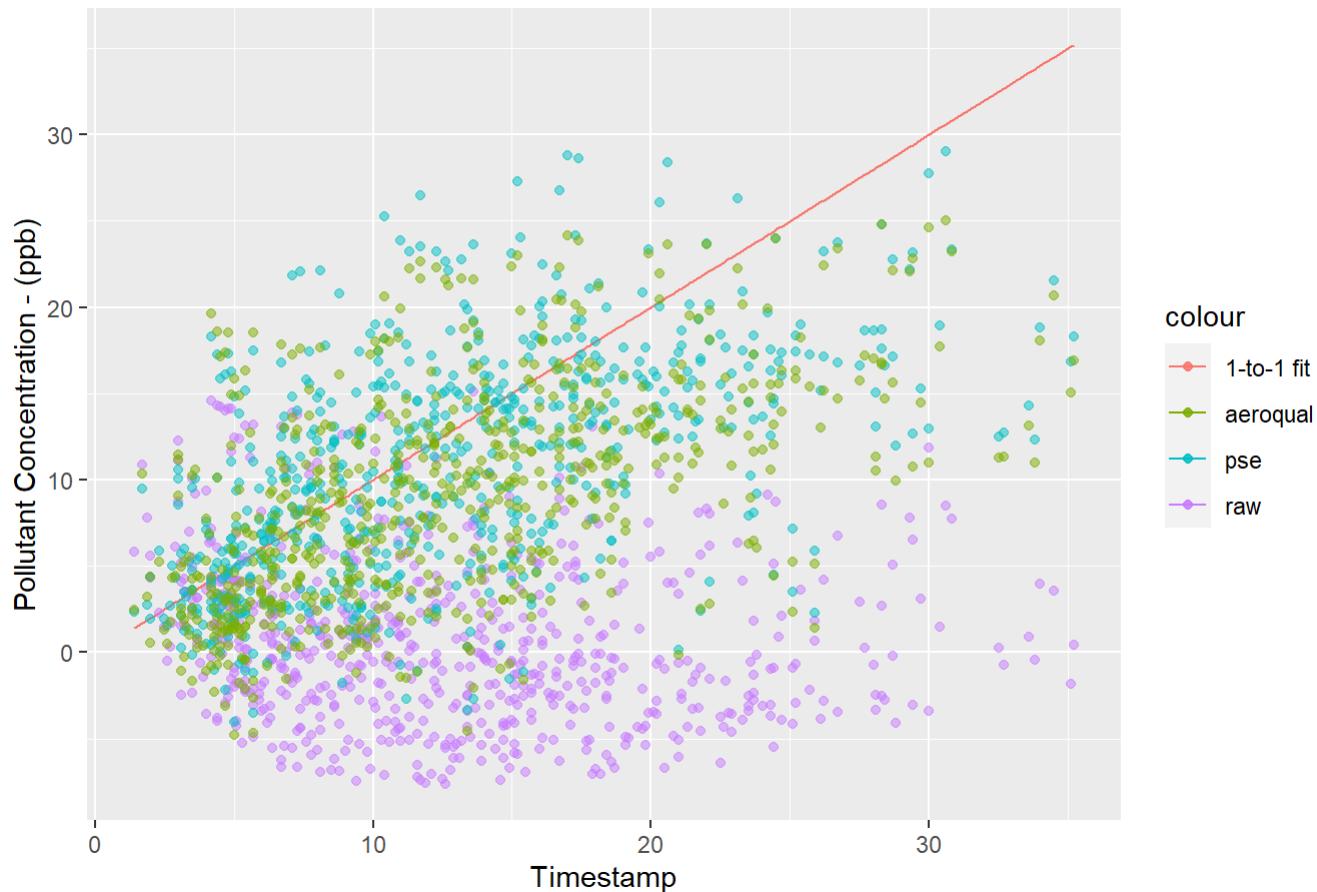
## NO2 Timeseries - AQY BB-642 Month of 12



## Absolute Error for NO2 - AQY BB-642 month of 12

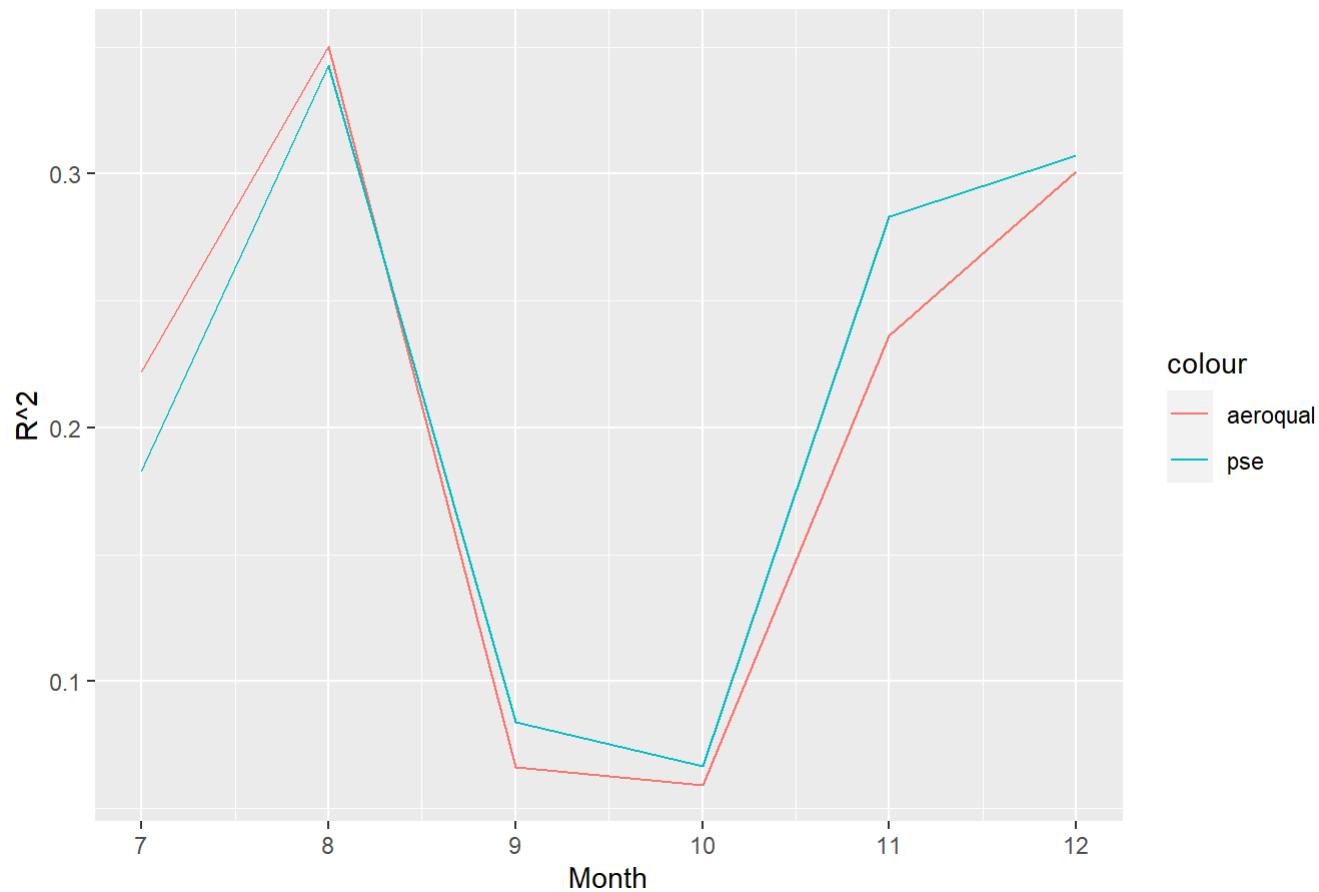


## NO2 [Proxy] vs [Measured] for AQY BB-642 month of 12

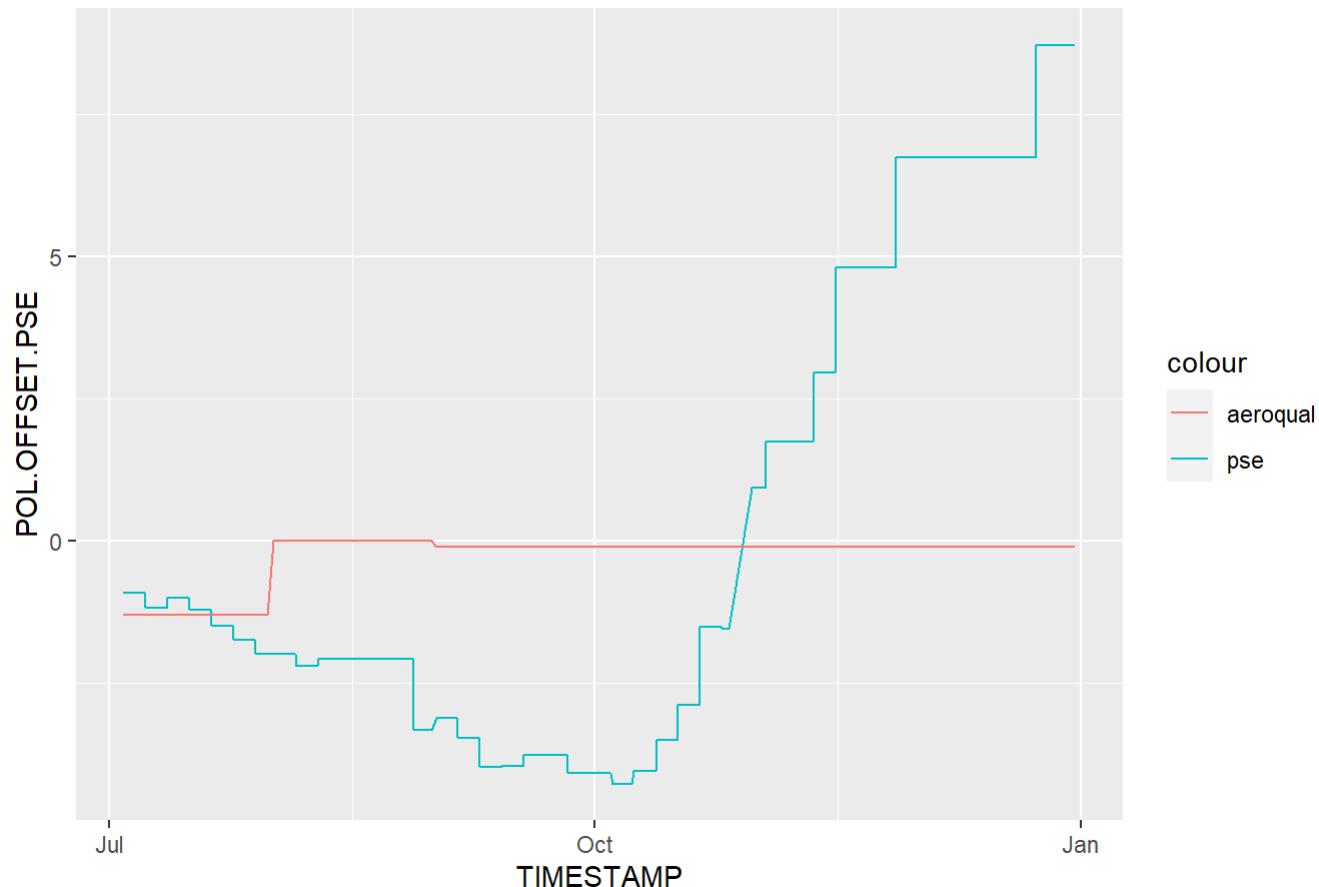


```
##          month      raw      PSE      AQ
## 1 Full Deployment 0.00131 0.36241 0.18356
## 2                  07 0.03021 0.18307 0.22217
## 3                  08 0.05849 0.34288 0.35047
## 4                  09 2e-04 0.08401 0.06613
## 5                  10 0.01849 0.06664 0.05924
## 6                  11 0.06217 0.2831  0.2362
## 7                  12 0.02605 0.30737 0.30099
```

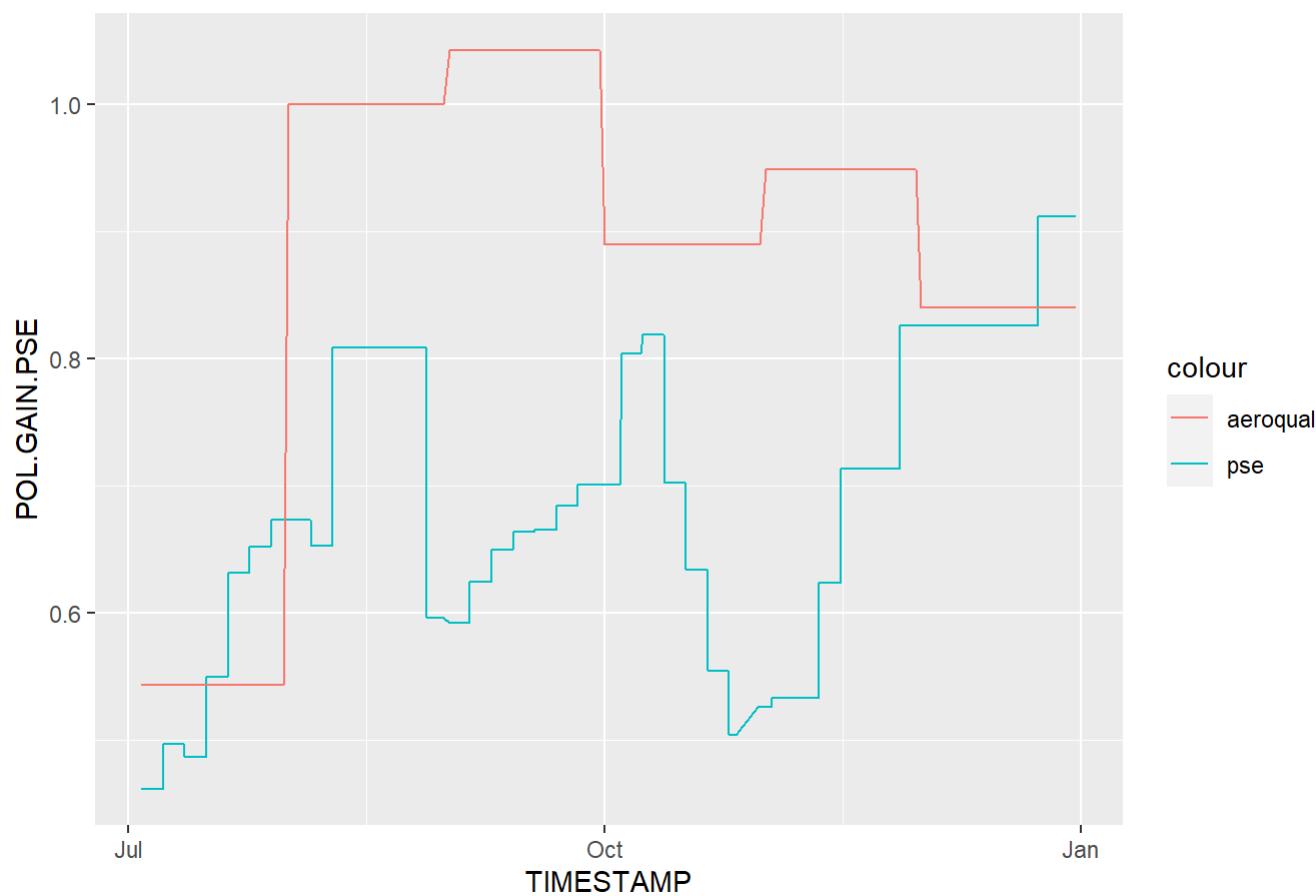
## NO2 Proxy ~ Measured R^2 Over Time for AQY BB-642



## NO2 Offset Over Time for AQY BB-642



## NO2 Gain Over Time for AQY BB-642



## MANUAL PLOTS

```
get_scatter_facet_aqy <- function(in_data, pollutant, aqy_id){

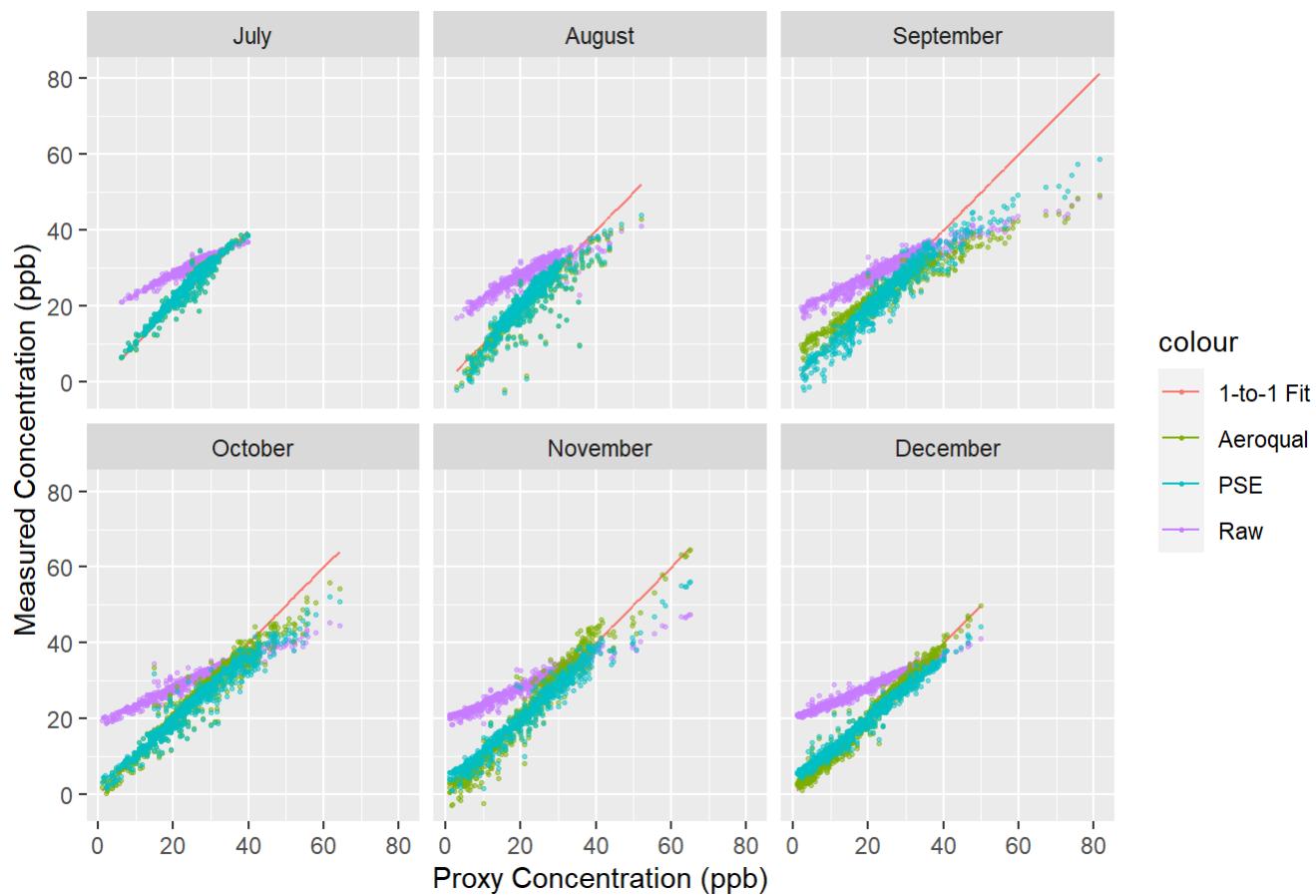
  colnames(in_data) <- str_replace_all(colnames(in_data), pollutant, 'POL')

  scatter_facet <- in_data %>%
    filter(ID == aqy_id) %>%
    mutate(MONTH = month(TIMESTAMP, label = T, abbr = F), .after = 'TIMESTAMP') %>%
    ggplot() +
    geom_line(aes(x = POL.PROXY, y = POL.PROXY, color = '1-to-1 Fit')) +
    geom_point(aes(x = POL.PROXY, y = POL.RAW, color = 'Raw'), size = .5, alpha = .5) +
    geom_point(aes(x = POL.PROXY, y = POL.AQ, color = 'Aeroqual'), size = .5, alpha = .5) +
    geom_point(aes(x = POL.PROXY, y = POL.PSE, color = 'PSE'), size = .5, alpha = .5) +
    labs(x = 'Proxy Concentration (ppb)', y = 'Measured Concentration (ppb)', title = paste('Calibrated', pollutant, '& Proxy Values - ', aqy_id)) +
    theme(plot.title = element_text(hjust = .5)) +
    facet_wrap(~MONTH)

  return(scatter_facet)
}

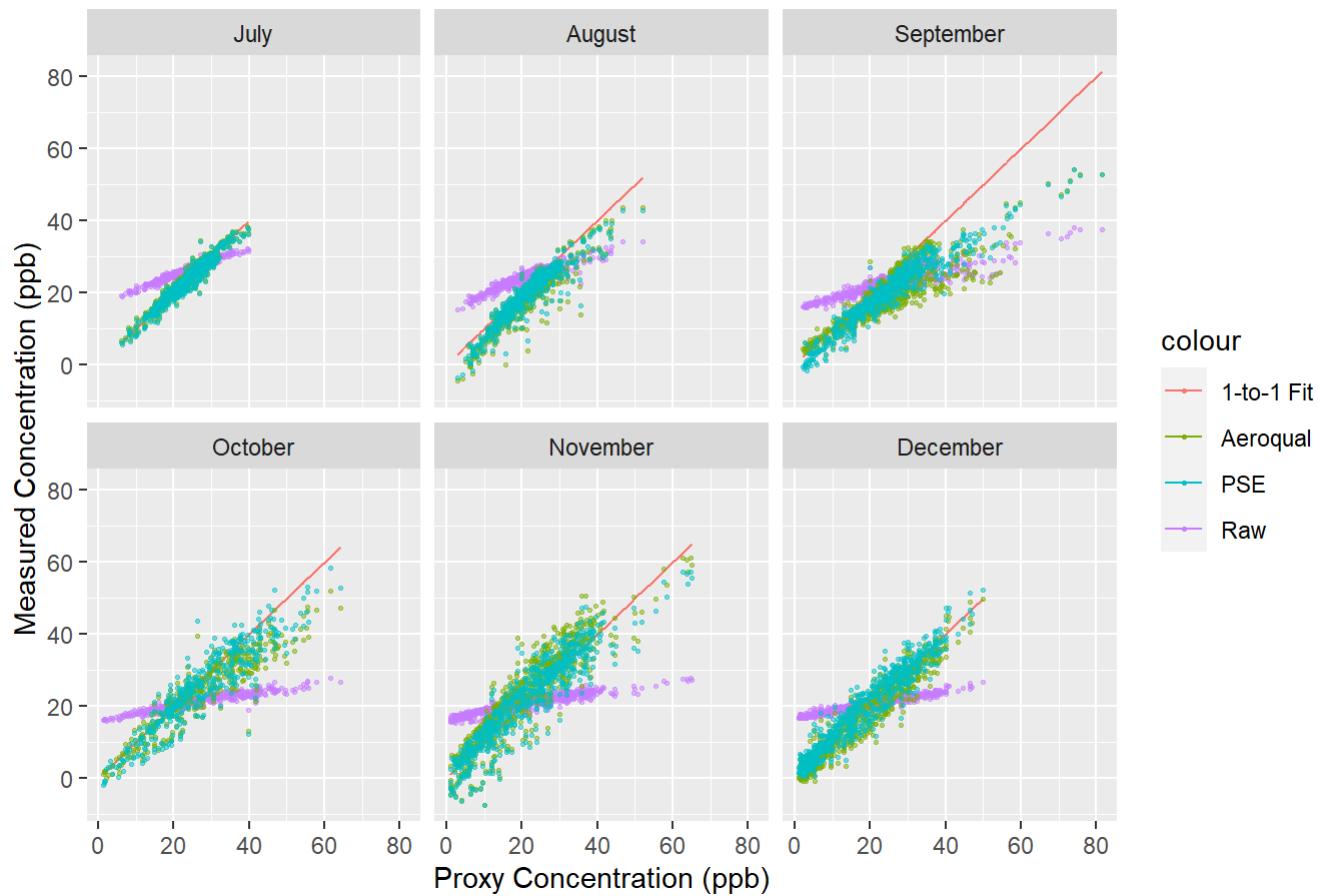
get_scatter_facet_aqy(O3_data, 'O3', 'AQY BB-633')
```

## Calibrated O3 &amp; Proxy Values - AQY BB-633

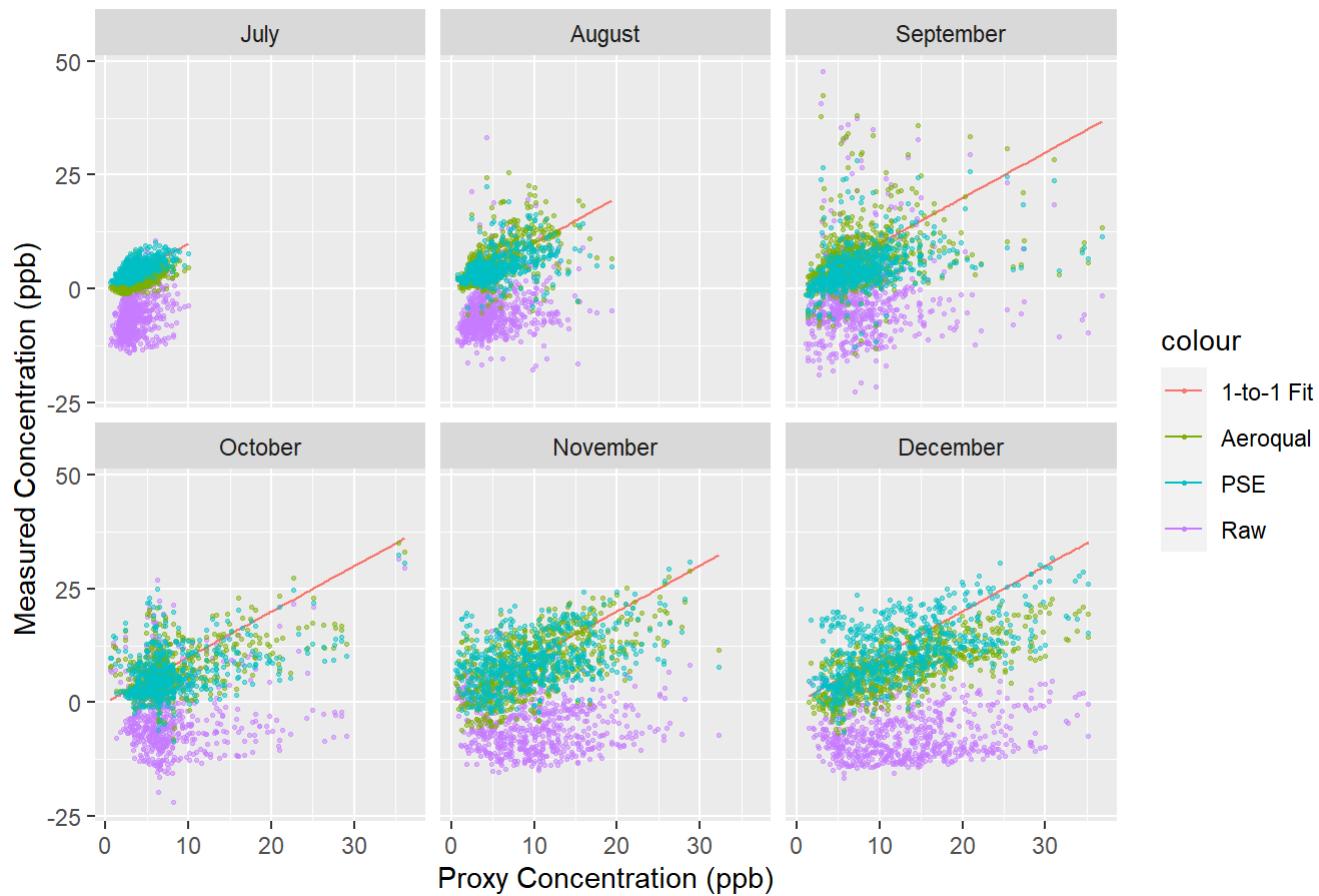


```
get_scatter_facet_aqy(03_data, '03', 'AQY BB-642')
```

## Calibrated O3 &amp; Proxy Values - AQY BB-642



```
get_scatter_facet_aqy(N02_data, 'N02', 'AQY BB-633')
```

Calibrated NO<sub>2</sub> & Proxy Values - AQY BB-633

```
get_scatter_facet_aqy(NO2_data, 'NO2', 'AQY BB-642')
```

Calibrated NO<sub>2</sub> & Proxy Values - AQY BB-642