

# Residuals plot

*Alex Shtein*

*June 01, 2017*

## General

The plots bellow present the residuals from the calibration stage (MOD1- as we call it). The MOD1 or calibration stage dataset includes PM monitors data (PM2.5 concentration) and with the collocated satellite data (AOD variable in the nearest range of 1.5 km) and additional spatial temporal predictors.

```
## Load libraries
library(lme4)
library(ggplot2)
library(data.table)
library(mgcv)
library(dplyr)
library(rgdal)

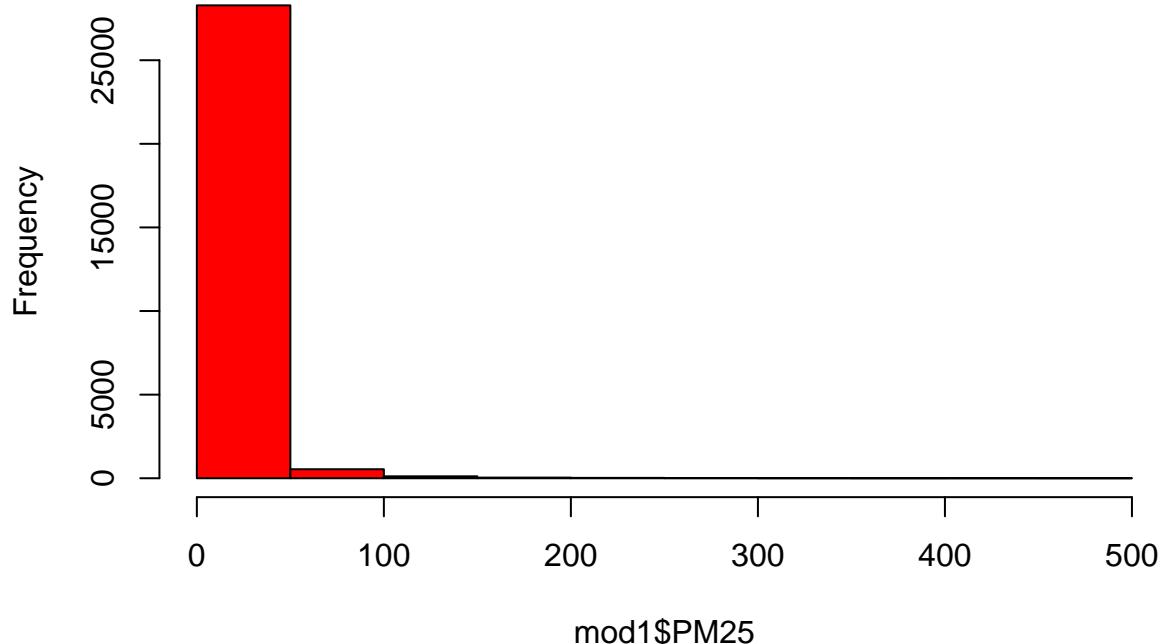
## Load Mod1
mod1 <-readRDS("N:/Projects/P028.IL.Israel.MAIAC.PM.V2/work/RDS_files/mod1/mod1.AQ.2003_2015.PM25_Daily")

## This is the variable that we would like to predict (PM2.5)
summary(mod1$PM25)

##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##      0.10   14.40  19.04   21.67  24.66  459.90

hist(mod1$PM25,col="red")
```

## Histogram of mod1\$PM25



```
# Define calibration formula - mixed effects model
m1.formula <- as.formula(PM25 ~ aod_047
                           ## Spatial predictors
                           +Elev.s +ndvi.s+Dis_Rd1_2012.s+P_In.s+P_Ur.s+P_Ag.s+P_OS.s+
                           ## Temporal predictors
                           daily_hpbl.s+Temp_D.s+N02_D.s
                           ## Random slope and intercept for per day nested in metregion
                           +(1+aod_047|day/metreg))

mod1fit <- lmer(m1.formula,data=mod1)
summary(mod1fit)

## Linear mixed model fit by REML ['lmerMod']
## Formula: PM25 ~ aod_047 + Elev.s + ndvi.s + Dis_Rd1_2012.s + P_In.s +
##           P_Ur.s + P_Ag.s + P_OS.s + daily_hpbl.s + Temp_D.s + N02_D.s +
##           (1 + aod_047 | day/metreg)
## Data: mod1
##
## REML criterion at convergence: 203715
##
## Scaled residuals:
##     Min      1Q   Median      3Q     Max 
## -29.2047 -0.3456 -0.0378  0.2885 25.2517 
## 
## Random effects:
## Groups   Name        Variance Std.Dev. Corr
##
```

```

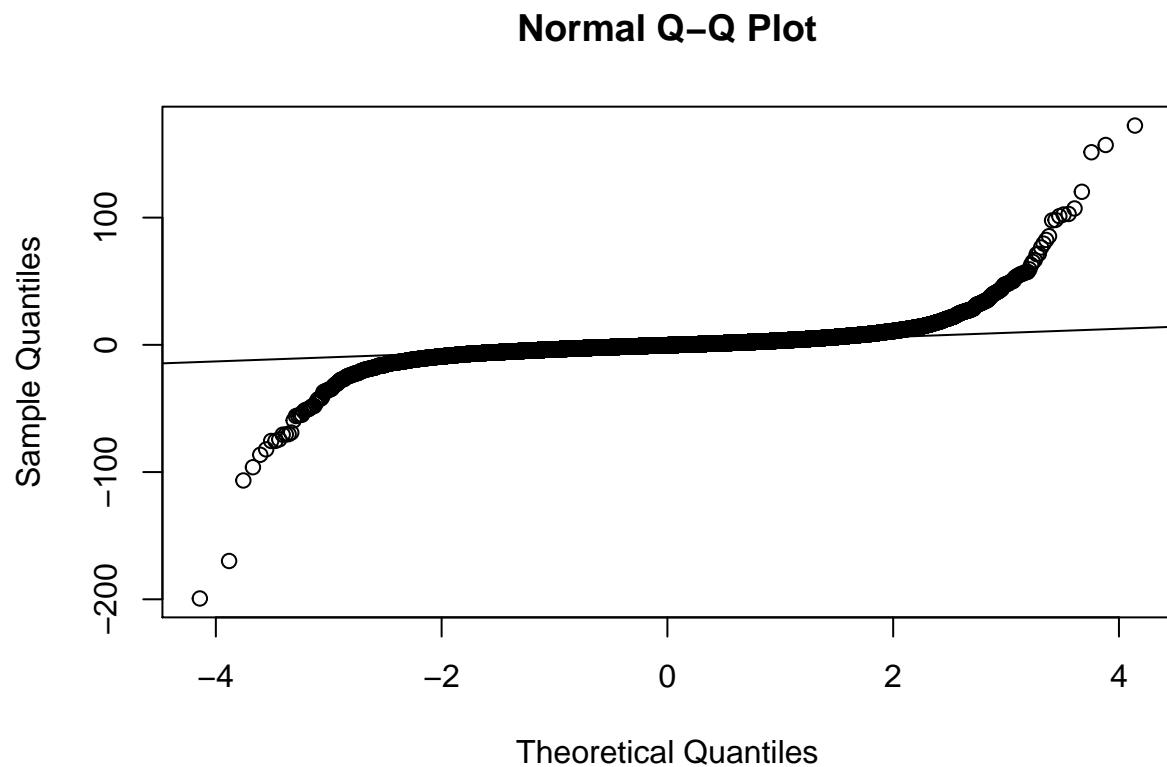
##  metreg:day (Intercept) 9.93 3.151
##                aod_047 488.51 22.102 -1.00
##  day          (Intercept) 22.43 4.736
##                aod_047 719.00 26.814 -0.06
##  Residual           46.59 6.825
## Number of obs: 29019, groups: metreg:day, 13302; day, 3340
##
## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)      14.45995  0.24344 59.40
## aod_047         20.51935  1.03538 19.82
## Elev.s          -0.56520  0.07004 -8.07
## ndvi.s          -0.15047  0.04409 -3.41
## Dis_Rd1_2012.s -0.29694  0.13974 -2.12
## P_In.s          0.17112  0.02226  7.69
## P_Ur.s          1.14264  0.23292  4.91
## P_Ag.s          2.54252  0.46181  5.51
## P_OS.s          1.76536  0.49535  3.56
## daily_hpbl.s   -0.73586  0.07818 -9.41
## Temp_D.s        0.20698  0.11400  1.82
## NO2_D.s         0.85693  0.04017 21.33
##
## Correlation of Fixed Effects:
##              (Intr) ad_047 Elev.s ndvi.s D_R1_2 P_In.s P_Ur.s P_Ag.s P_OS.s
## aod_047     -0.659
## Elev.s      0.135  0.064
## ndvi.s     -0.051 -0.033  0.021
## Ds_R1_2012. 0.369 -0.037  0.097  0.061
## P_In.s      0.064  0.002  0.144  0.043  0.009
## P_Ur.s      0.072 -0.024  0.015 -0.051  0.202 -0.276
## P_Ag.s      0.108 -0.017  0.040 -0.038  0.222 -0.256  0.986
## P_OS.s      0.125 -0.020 -0.017 -0.088  0.163 -0.236  0.976  0.974
## daily_hpbl.s 0.083 -0.054 -0.069  0.131 -0.041 -0.021  0.030  0.025  0.050
## Temp_D.s    0.202 -0.031  0.204  0.032  0.087  0.031  0.029  0.046  0.028
## NO2_D.s     0.179 -0.042 -0.101  0.058  0.245 -0.113  0.148  0.207  0.216
## dly_h. Tmp_D.
## aod_047
## Elev.s
## ndvi.s
## Ds_R1_2012.
## P_In.s
## P_Ur.s
## P_Ag.s
## P_OS.s
## daily_hpbl.s
## Temp_D.s    -0.127
## NO2_D.s     0.027  0.069
mod1$pred.m1 <- predict(mod1fit)
mod1$resid=residuals(mod1fit)
print(summary(lm(PM25~pred.m1,data=mod1))$r.squared)

## [1] 0.8581068

```

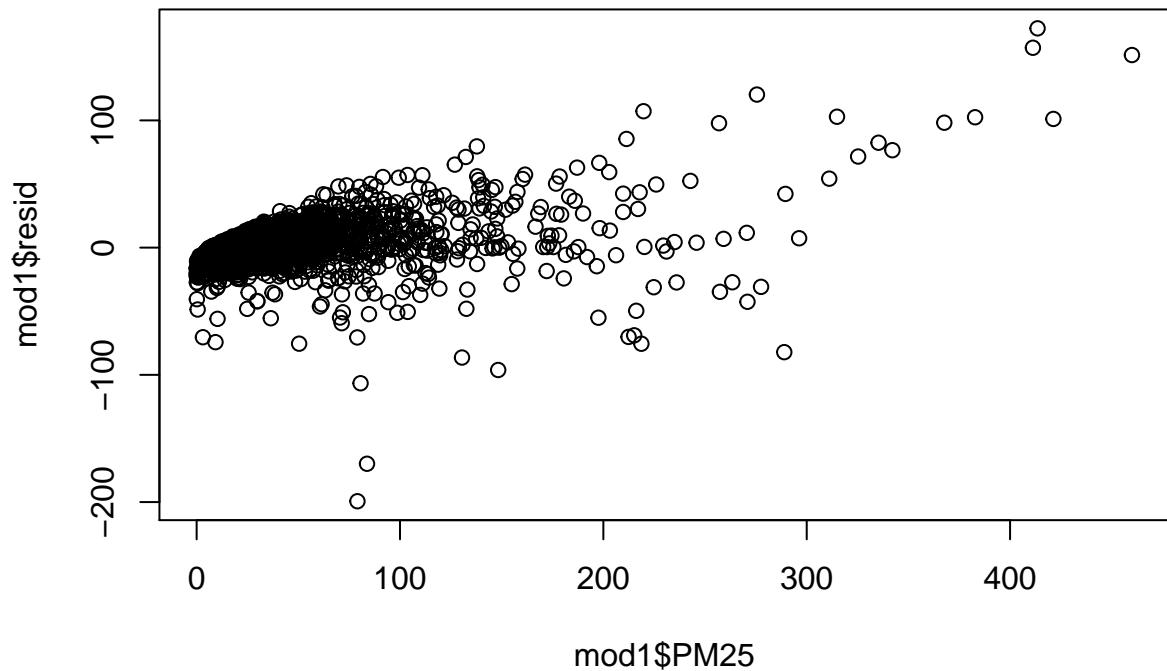
## Plots of residuals

```
qqnorm(mod1$resid)
qqline(mod1$resid)
```

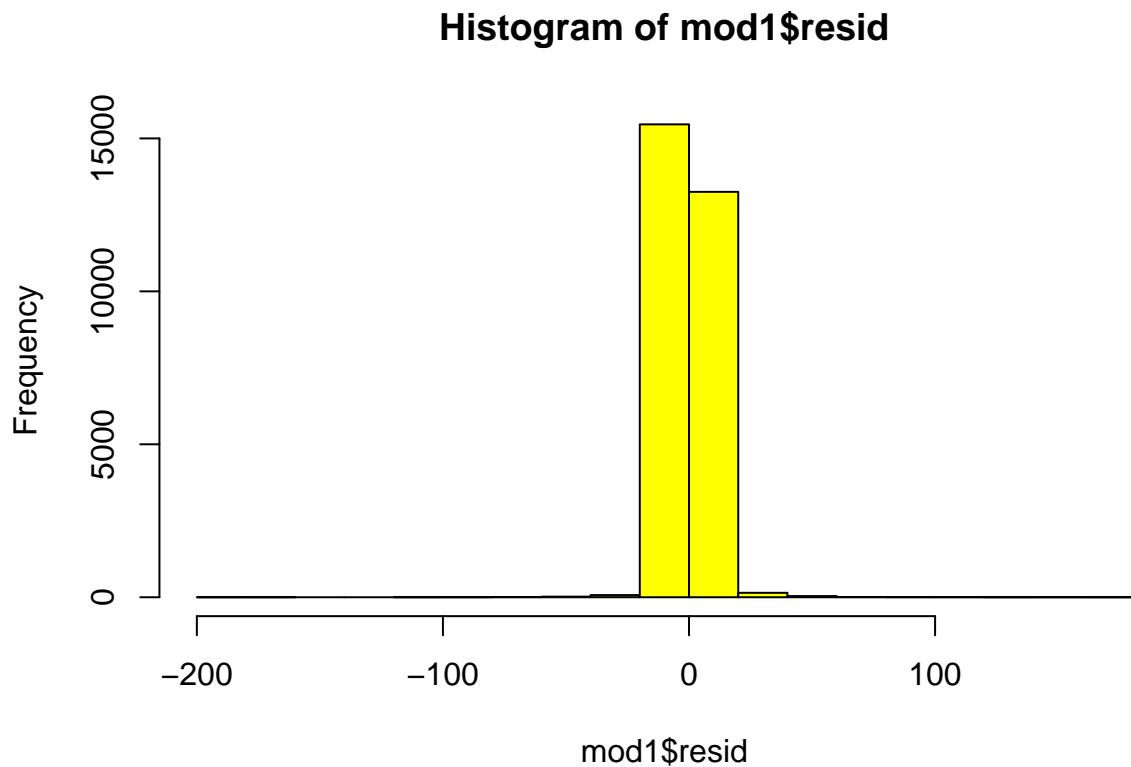


```
plot(mod1$resid~mod1$PM25,main="residuals ~ observed PM2.5")
```

### residuals ~ observed PM2.5



```
hist(mod1$resid,col="yellow")
```



```
summary(mod1$resid)
```

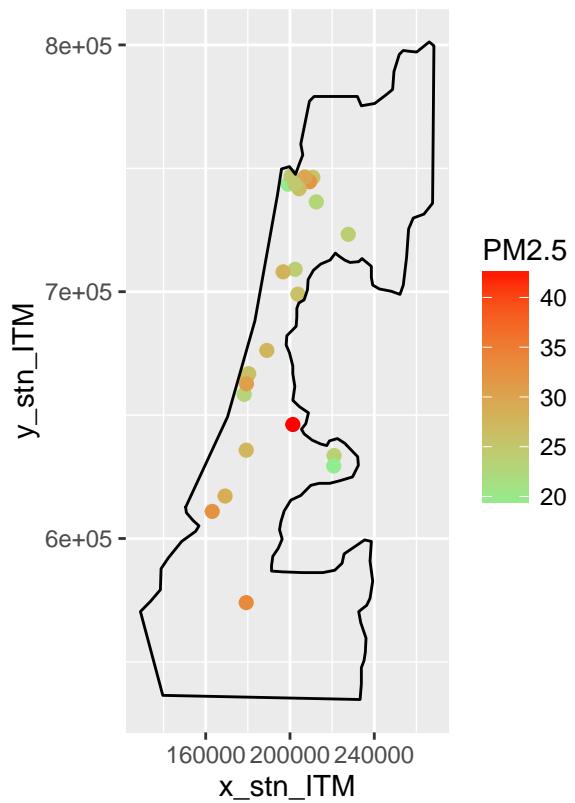
```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## -199.3000 -2.3590  -0.2581   0.0000   1.9690  172.4000
```

### Plot the residuals spatially

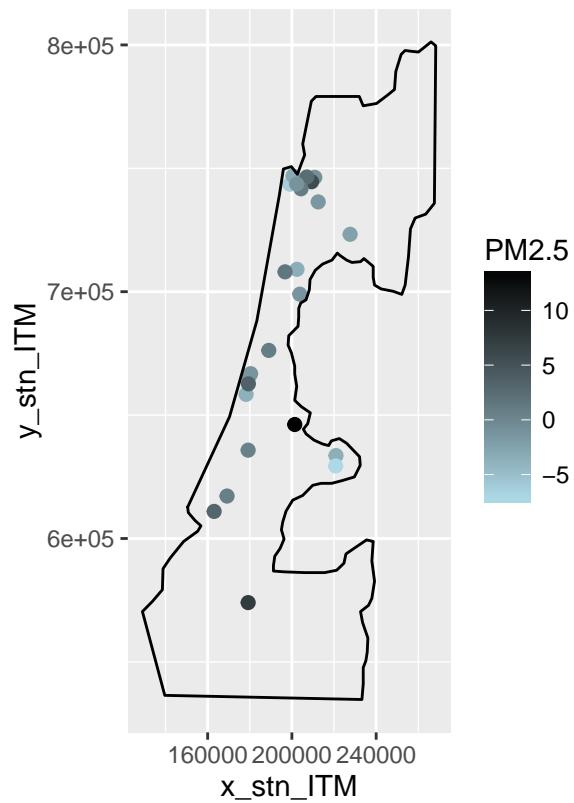
Note that in each day we have different amount of stations. This happens because in some days we do not have co-located AOD measurement in part of the stations.

```
## OGR data source with driver: ESRI Shapefile
## Source: "N:/Projects/P028.IL.Israel.MAIAC.PM.V2/work/Qgis/General/Project_border/Project_aoi", layer
## with 1 features
## It has 1 fields
## Integer64 fields read as strings:  Id
```

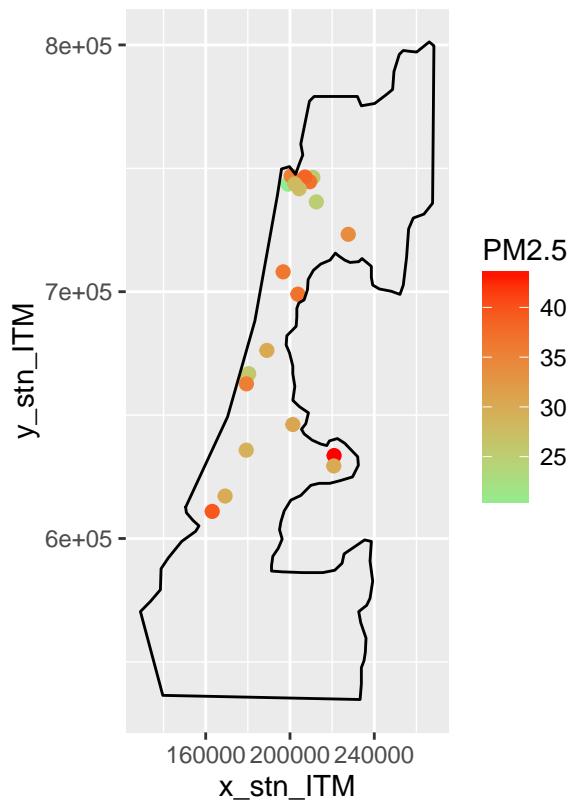
PM2.5 concentration for 1.8.15



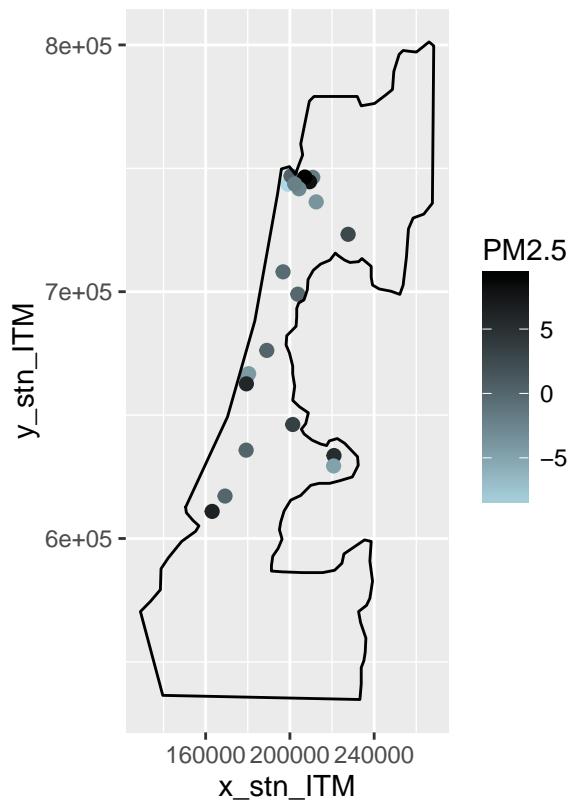
PM2.5 residuals for 1.8.15

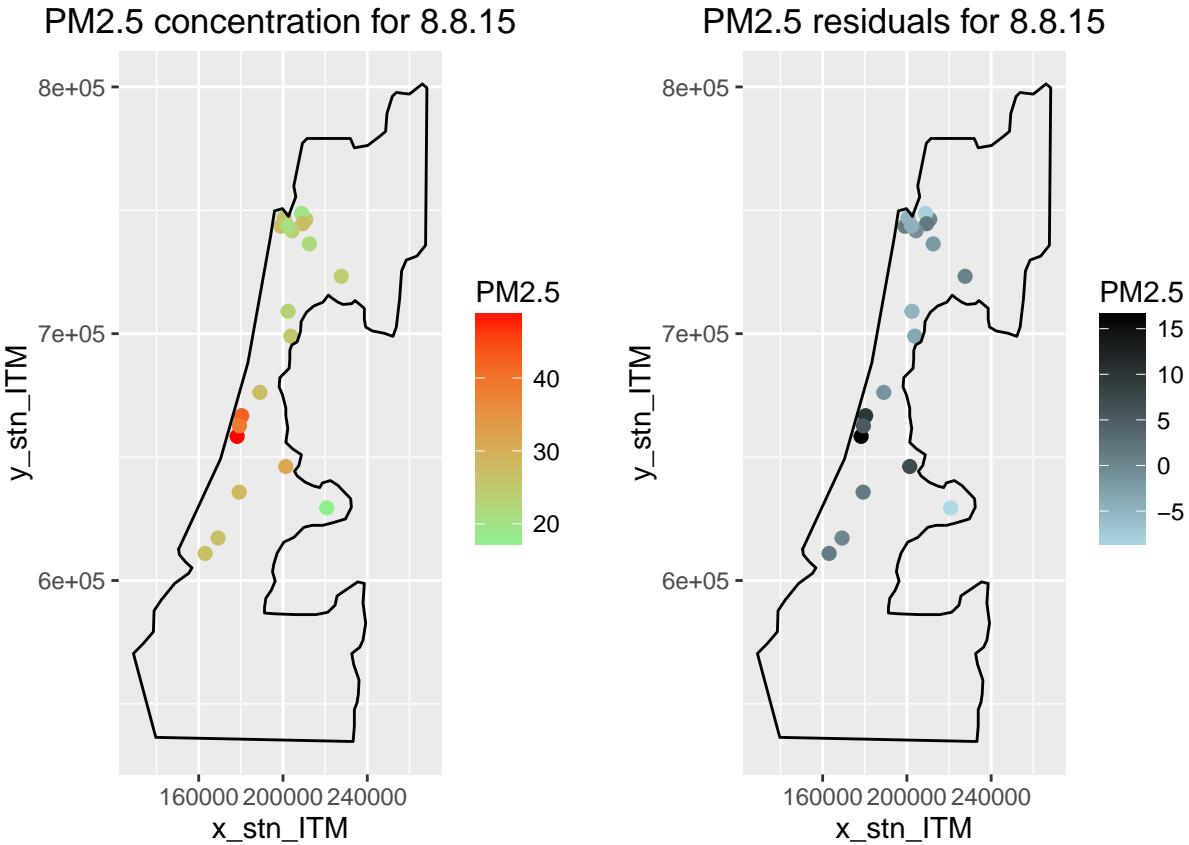


PM2.5 concentration for 3.8.15



PM2.5 residuals for 3.8.15





### Transforming PM2.5 and evaluating the residuals

```
# Define calibration formula - mixed effects model- with transformation of PM
mod1$PM25.sq=sqrt(mod1$PM25)
m2.formula <- as.formula(PM25.sq ~ aod_047
                           ## Spatial predictors
                           +Elev.s +ndvi.s+Dis_Rd1_2012.s+P_In.s+P_Ur.s+P_Ag.s+P_OS.s+
                           ## Temporal predictors
                           daily_hpbl.s+Temp_D.s+N02_D.s
                           ## Random slope and intercept for per day nested in metregion
                           +(1+aod_047|day/metreg))

mod2fit <- lmer(m2.formula,data=mod1)
summary(mod2fit)

## Linear mixed model fit by REML ['lmerMod']
## Formula: PM25.sq ~ aod_047 + Elev.s + ndvi.s + Dis_Rd1_2012.s + P_In.s +
##           P_Ur.s + P_Ag.s + P_OS.s + daily_hpbl.s + Temp_D.s + N02_D.s +
##           (1 + aod_047 | day/metreg)
## Data: mod1
##
## REML criterion at convergence: 60071.9
##
## Scaled residuals:
##       Min        1Q    Median        3Q       Max
## -1.500000 -0.500000  0.000000  0.500000  1.500000
```

```

## -13.5903 -0.4311 -0.0024 0.4343 9.5916
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## metreg:day (Intercept) 0.02548 0.1596
##           aod_047    1.83260 1.3537 -0.94
## day       (Intercept) 0.37255 0.6104
##           aod_047    6.15905 2.4817 -0.25
## Residual          0.31814 0.5640
## Number of obs: 29019, groups: metreg:day, 13302; day, 3340
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 3.839128 0.022667 169.37
## aod_047     1.856763 0.091117 20.38
## Elev.s      -0.074185 0.005865 -12.65
## ndvi.s      -0.015723 0.003717 -4.23
## Dis_Rd1_2012.s -0.023834 0.011586 -2.06
## P_In.s      0.019075 0.001846 10.34
## P_Ur.s      0.117849 0.019722 5.98
## P_Ag.s      0.260217 0.039077 6.66
## P_OS.s      0.179017 0.041989 4.26
## daily_hpbl.s -0.062318 0.006796 -9.17
## Temp_D.s     0.022346 0.010539 2.12
## NO2_D.s     0.091790 0.003355 27.36
##
## Correlation of Fixed Effects:
## (Intr) ad_047 Elev.s ndvi.s D_R1_2 P_In.s P_Ur.s P_Ag.s P_OS.s
## aod_047 -0.668
## Elev.s   0.118  0.068
## ndvi.s   -0.046 -0.032  0.020
## Ds_R1_2012. 0.332 -0.038  0.097  0.061
## P_In.s    0.056  0.003  0.145  0.043  0.007
## P_Ur.s    0.064 -0.020  0.014 -0.057  0.205 -0.275
## P_Ag.s    0.096 -0.013  0.038 -0.044  0.225 -0.256  0.986
## P_OS.s    0.110 -0.015 -0.018 -0.093  0.167 -0.237  0.977  0.975
## daily_hpbl.s 0.062 -0.040 -0.071  0.133 -0.046 -0.019  0.031  0.025  0.049
## Temp_D.s   0.184 -0.016  0.227  0.031  0.087  0.031  0.031  0.048  0.030
## NO2_D.s    0.162 -0.045 -0.102  0.058  0.247 -0.115  0.150  0.208  0.217
##           dly_h. Tmp_D.
## aod_047
## Elev.s
## ndvi.s
## Ds_R1_2012.
## P_In.s
## P_Ur.s
## P_Ag.s
## P_OS.s
## daily_hpbl.s
## Temp_D.s    -0.110
## NO2_D.s     0.011  0.051
mod1$pred.m1_sq <- predict(mod2fit)
mod1$pred.m1_sq2 <- mod1$pred.m1_sq*mod1$pred.m1_sq

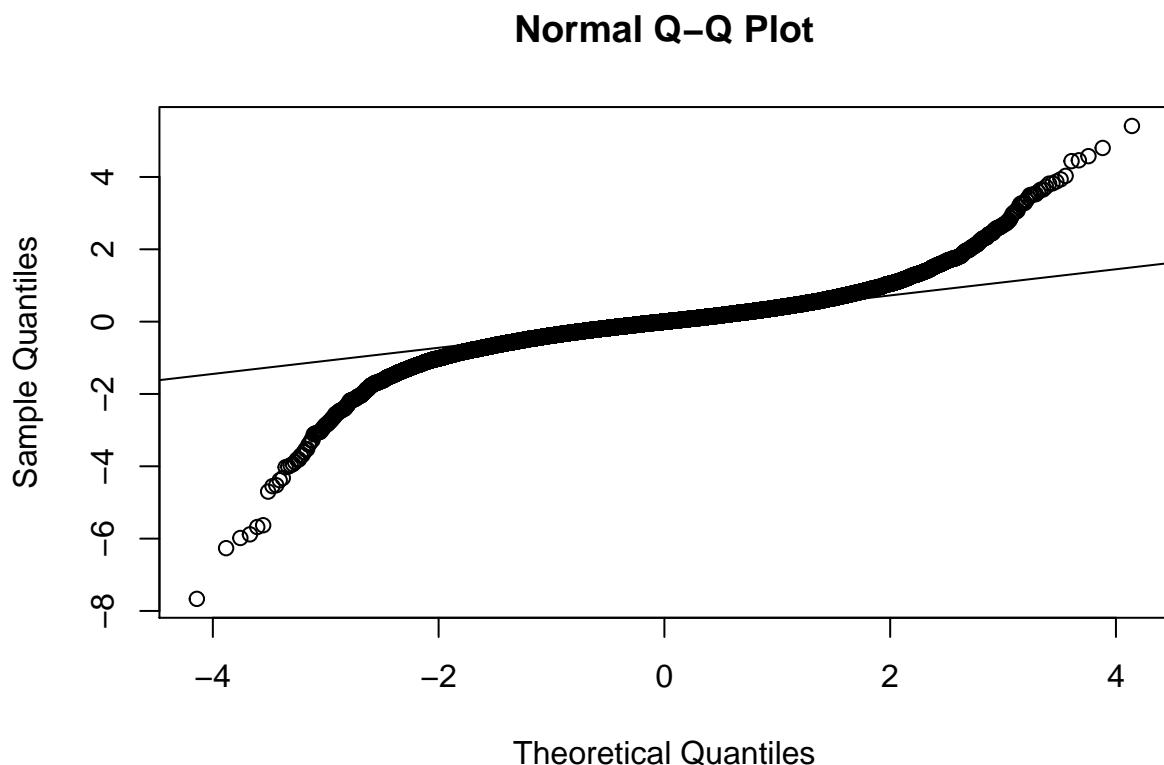
```

```
mod1$resid_sq=residuals(mod2fit)
mod1$resid_sq2=(mod1$resid_sq)^2
print(summary(lm(PM25~pred.m1_sq2,data=mod1))$r.squared)

## [1] 0.8562886
```

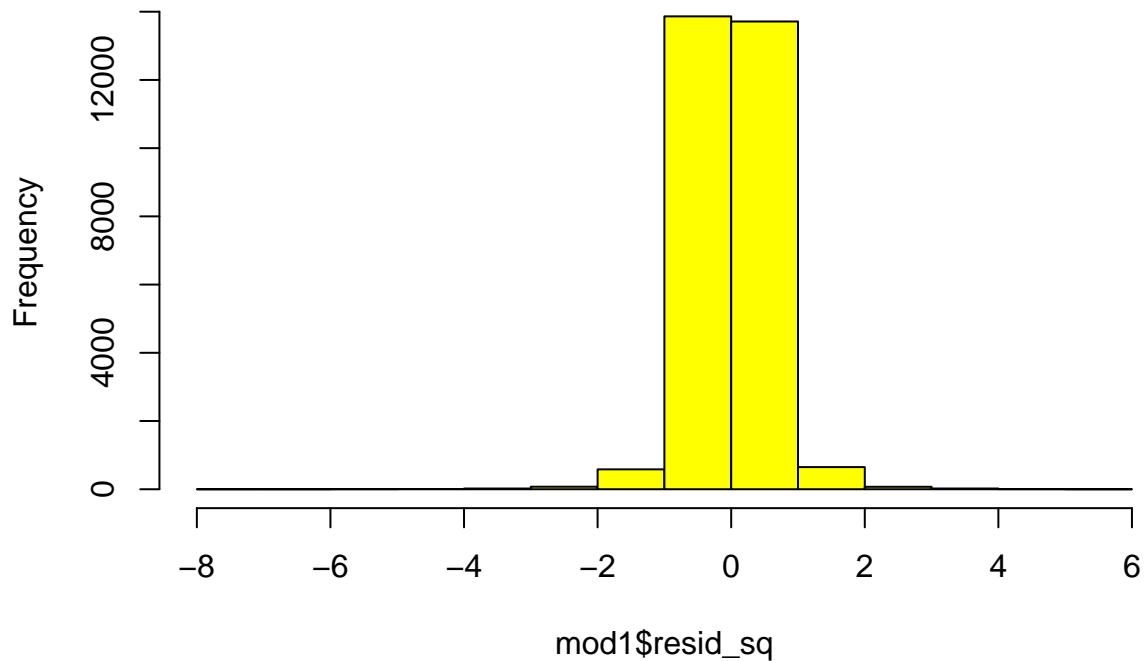
Plots of residuals (after transforming PM2.5)

```
qqnorm(mod1$resid_sq)
qqline(mod1$resid_sq)
```



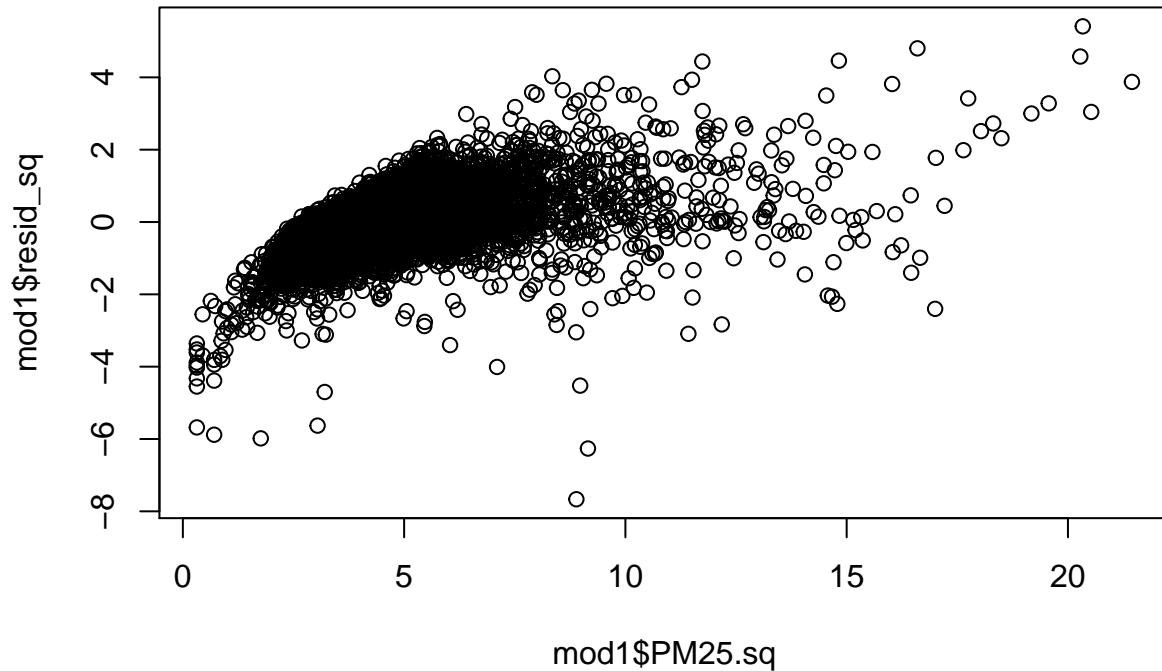
```
hist(mod1$resid_sq,col="yellow")
```

### Histogram of mod1\$resid\_sq



```
plot(mod1$resid_sq-mod1$PM25.sq,main="sqrt residuals ~ sqrt observed PM2.5")
```

## **sqrt residuals ~ sqrt observed PM2.5**



```
summary(mod1$resid_sq)
```

```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## -7.666000 -0.243200 -0.001376  0.000000  0.245000  5.410000
```

```
summary(mod1$resid_sq2)
```

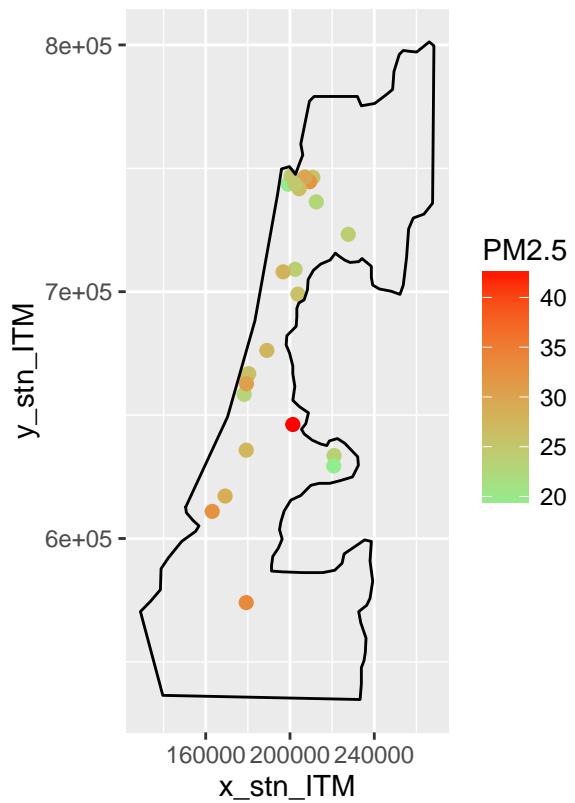
```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
##  0.00000  0.01206  0.05958  0.26500  0.20960  58.76000
```

## **Plot the residuals spatially (for the model with the transformed PM2.5)**

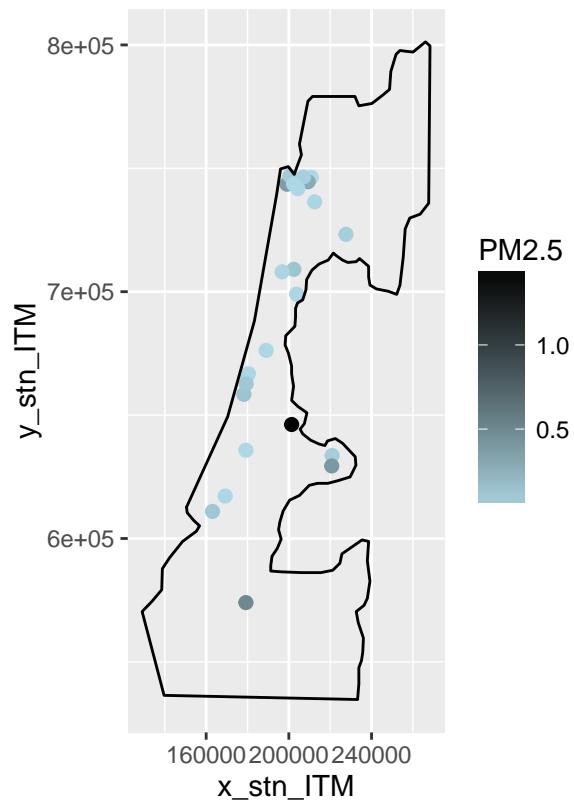
Note that in each day we have different amount of stations. This happens because in some days we do not have co-located AOD measurement in part of the stations.

```
## OGR data source with driver: ESRI Shapefile
## Source: "N:/Projects/P028.IL.Israel.MAIAC.PM.V2/work/Qgis/General/Project_border/Project_aoi", layer
## with 1 features
## It has 1 fields
## Integer64 fields read as strings:  Id
```

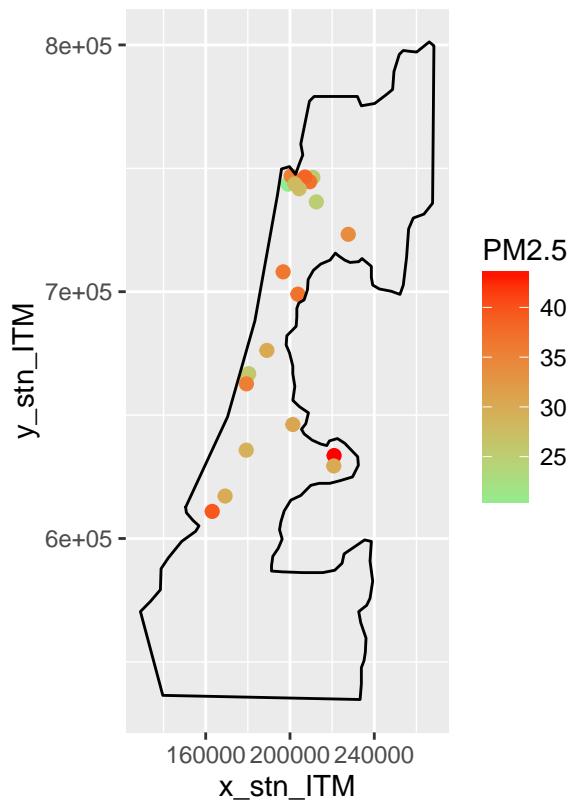
PM2.5 concentration for 1.8.15



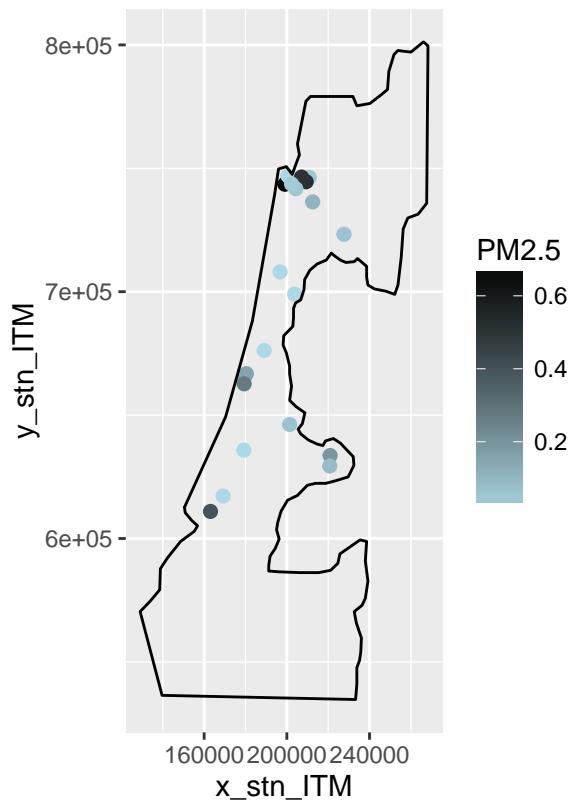
PM2.5 residuals for 1.8.15



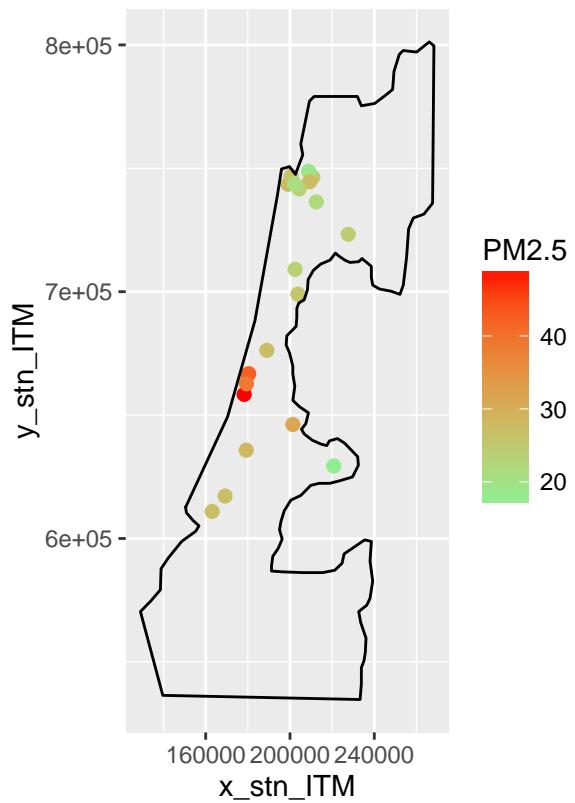
PM2.5 concentration for 3.8.15



PM2.5 residuals for 3.8.15



PM2.5 concentration for 8.8.15



PM2.5 residuals for 8.8.15

