

# Pre-Processing O3 Data

```
library(zoo)
```

```
## Warning: package 'zoo' was built under R version 3.6.3
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

```
library(stats)
```

## Summary

```
##           London   Berlin   Madrid   Rome
## 2018-01-01 46.84612 57.41639 50.70971 26.781250
## 2018-01-02 33.60102 31.25111 42.76274 24.630208
## 2018-01-03 56.26416 47.44597 18.82032  8.119792
## 2018-01-04 42.79985 48.48153 20.97162  6.697917
## 2018-01-05 33.86294 43.12917 36.54381  9.507473
## 2018-01-06 15.74904 26.68944 53.63199 19.630208
```

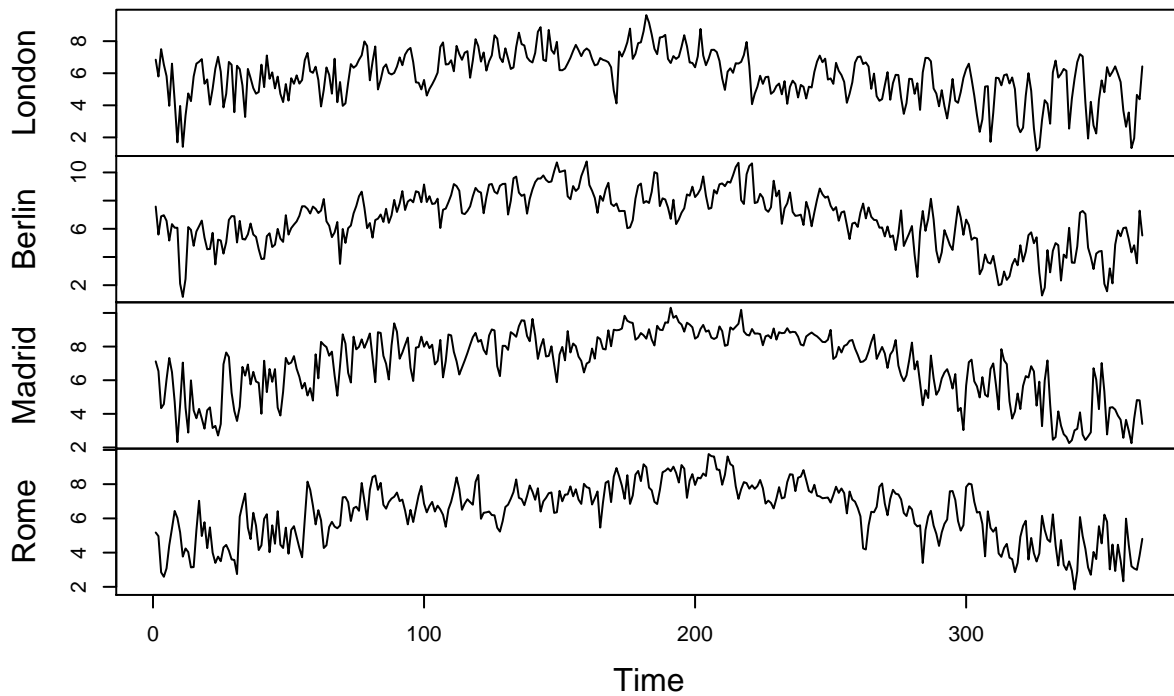
## Common trends

Explain stylized facts from the literature.

Taking the square-root transform to bring the data back to the light tailed domain as suggested in Cho & Fryzlewicz (2021) and imputing any missing values by linear interpolation, the common trends from the literature are somewhat apparent.

```
sample.cities <- sqrt(sample.cities)
for (city in names(sample.cities)) sample.cities[[city]] <- na.approx(sample.cities[[city]])
plot.ts(sample.cities, main = "O3 concentration 2018")
```

## O3 concentration 2018



### Removing trends

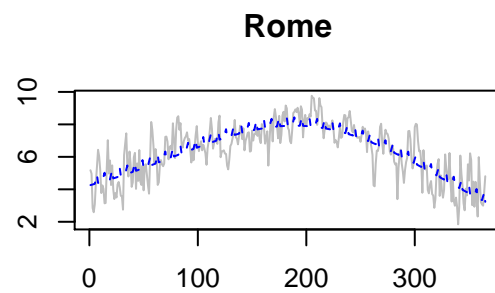
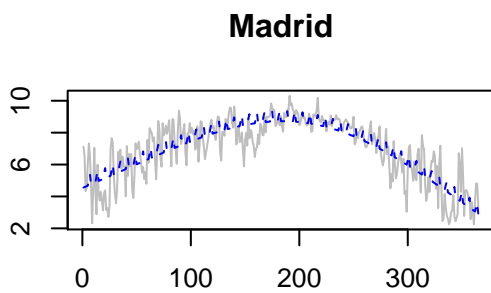
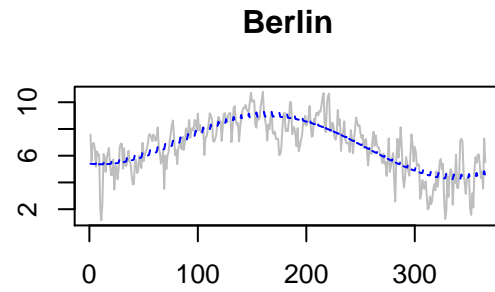
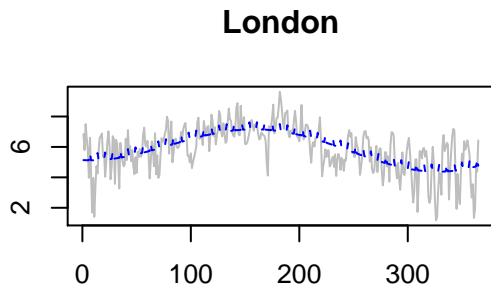
```
days <- as.Date(rownames(sample.cities), format = "%Y-%m-%d")
tt <- 1:length(days)

sat <- ifelse(weekdays(days) == "Saturday", 1, 0)
sun <- ifelse(weekdays(days) == "Sunday", 1, 0)

s.qt <- sin(2*pi*tt/365/4)
c.qt <- cos(2*pi*tt/365/4)
s.yr <- sin(2*pi*tt/365)
c.yr <- cos(2*pi*tt/365)

trend.models <- list()

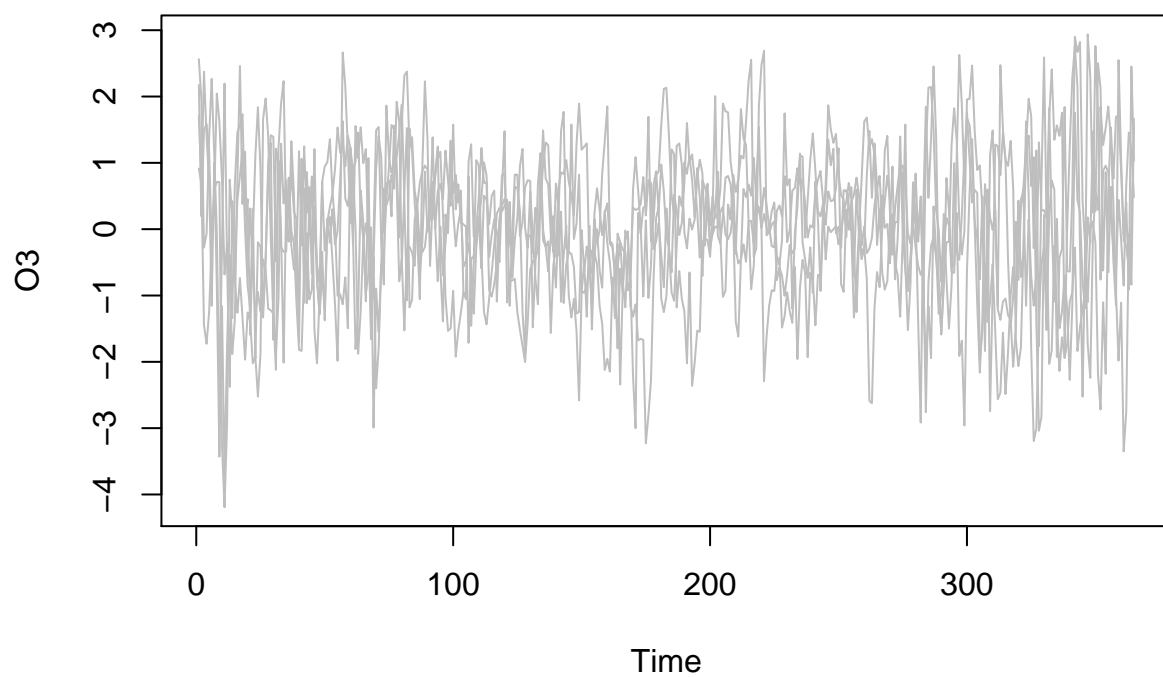
for (city in names(sample.cities)) trend.models[[city]] <- lm(
  sample.cities[[city]] ~ sat + sun + s.qt + c.qt + s.yr + c.yr
)
```



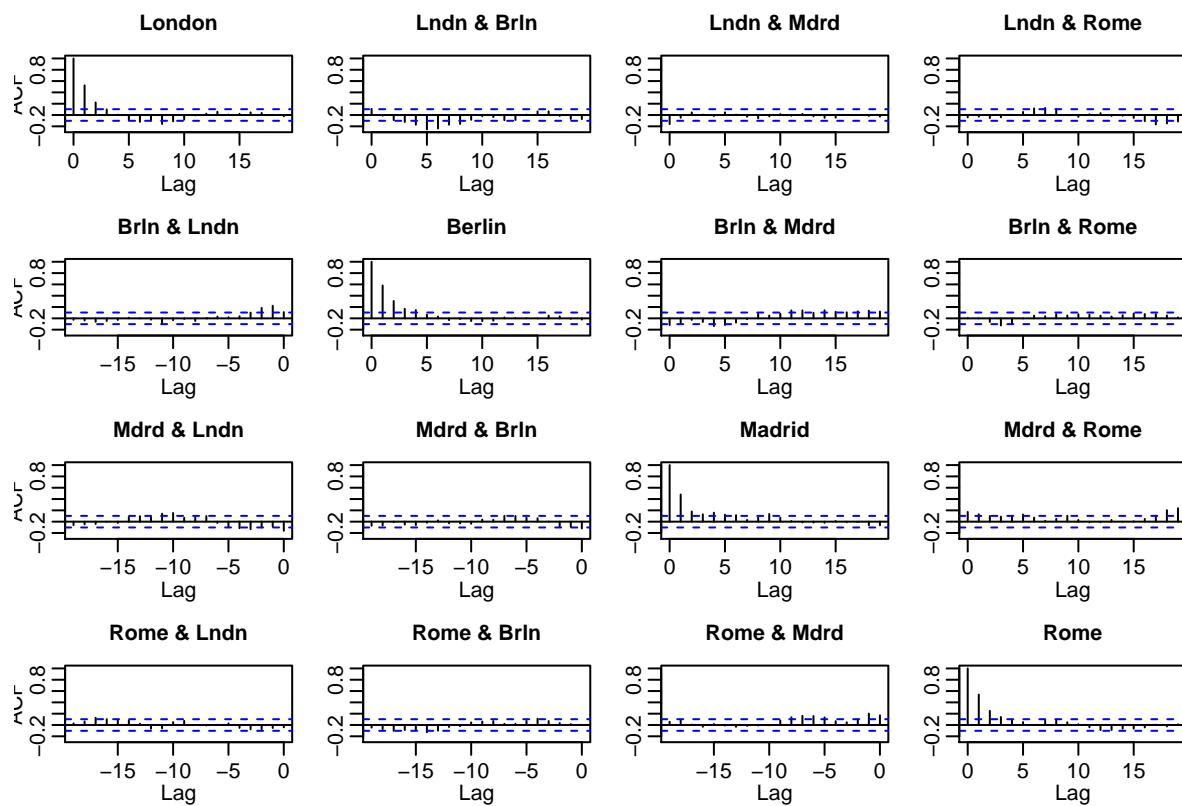
```
summary(trend.models[["London"]])

##
## Call:
## lm(formula = sample.cities[[city]] ~ sat + sun + s.qt + c.qt +
##     s.yr + c.yr)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.7132 -0.7451  0.0604  0.8147  2.7578
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.518953   2.358478   4.036 6.65e-05 ***
## sat           0.002467   0.175784   0.014 0.988812
## sun           0.471242   0.175797   2.681 0.007689 **
## s.qt          -3.111678   1.863048  -1.670 0.095752 .
## c.qt          -2.736223   1.855142  -1.475 0.141108
## s.yr           0.535525   0.141757   3.778 0.000185 ***
## c.yr          -1.644984   0.325798  -5.049 7.08e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.157 on 358 degrees of freedom
## Multiple R-squared:  0.4121, Adjusted R-squared:  0.4023
## F-statistic: 41.83 on 6 and 358 DF, p-value: < 2.2e-16
```

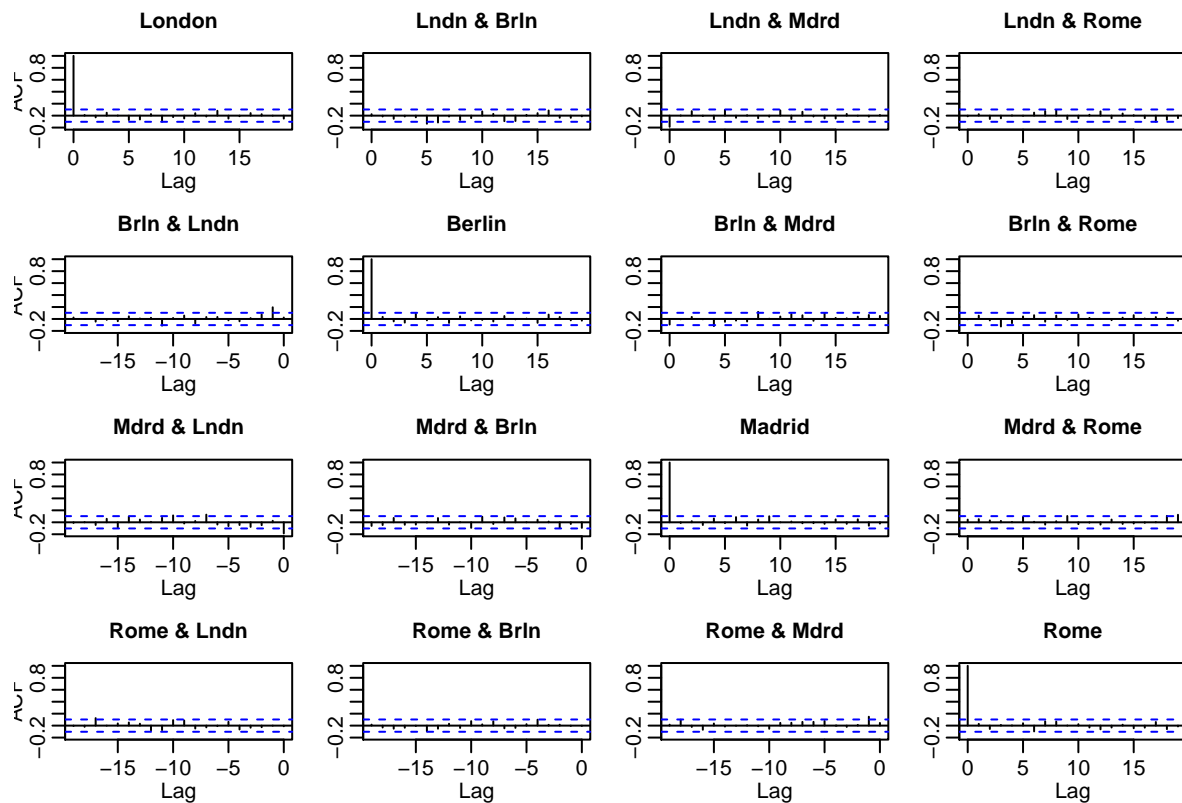
### De-trended data



## Removing serial correlation



```
ar.models <- list()
for (city in names(sample.cities)) ar.models[[city]] <- ar(detrended.data[[city]], order.max = 3)
```



Finally, the det-trended and whitened data is plotted below.

```
ts.plot(whitened.data, col = "grey", main = "Whitened and de-trended data", ylab = "03")
```

### Whitened and de-trended data

