

# Relatório Prova Prática 1 - Séries Temporais

---

## 1) Análise da concentração de partículas inaláveis finas (MP<sub>2.5</sub>) da cidade de London, província de Ontário, Canadá, no ano de 2013

---

### Contextualização do problema:

O arquivo Matter2013-London.csv, obtido em [www.airqualityontario.com](http://www.airqualityontario.com), contém informações sobre as concentrações de partículas inaláveis finas (MP<sub>2.5</sub>, em  $\mu g/m^3$ ), registradas no ano de 2013 em London, cidade situada na província canadense de Ontário. Os valores são registrados a cada hora. Para determinado dia D, as variáveis  $H_1$  a  $H_{24}$  representam as medições feitas na primeira até a 24<sup>a</sup> hora do dia, respectivamente.

O objetivo desta análise é fazer uma análise exploratória desta série temporal dada por:

$$Y_t = \min\{H_{j,t}\}$$

---

### item i)

Dados brutos antes do tratamento:

```
# ----- leitura e carregamento
library(tibbletime)
library(dplyr)
library(ggplot2)
library(readr)
library(lubridate)
library(reshape2)

matter.loc <- "/home/allan/Documents/1S2018/A_SERIES_TEMPORAIS/aulas/dados/Matter2013-London.csv"

matter <- read_csv(matter.loc)
head(matter)
```

```
## # A tibble: 6 x 28
##   `Station ID` Pollutant   Date      H01  H02  H03  H04  H05  H06
##   <int> <chr>      <date>    <int> <int> <int> <int> <int> <int>
## 1    15026 Fine Partic~ 2013-01-01     9     7     5     4     3     4
## 2    15026 Fine Partic~ 2013-01-02     5     6     7     7     6     5
## 3    15026 Fine Partic~ 2013-01-03    22    14    12     9     7     7
## 4    15026 Fine Partic~ 2013-01-04    17    16    14    13    15    18
## 5    15026 Fine Partic~ 2013-01-05    20    22    25    26    27    27
## 6    15026 Fine Partic~ 2013-01-06    24    26    21    18    17    17
## # ... with 19 more variables: H07 <int>, H08 <int>, H09 <int>, H10 <int>,
## #   H11 <int>, H12 <int>, H13 <int>, H14 <int>, H15 <int>, H16 <int>,
## #   H17 <int>, H18 <int>, H19 <int>, H20 <int>, H21 <int>, H22 <int>,
## #   H23 <int>, H24 <int>, X28 <chr>
```

Dados após preparação:

```
# ----- data preparation

# passando para formato long;
# eliminando as duas primeiras colunas;
# eliminando os NA's e as entradas com 9999 e -999;
# agrupando por dia e calculando o minimo
matter2 <- matter %>%
  dplyr::select(-(1:2)) %>% # eliminando duas primeiras colunas
  melt(id.vars = "Date") %>% # to long
  magrittr::set_colnames(c("Date", "hora", "MP")) %>%
  dplyr::arrange(Date, hora) %>% # ordenar por dia/hora
  mutate(MP = replace(MP, MP == 9999 | MP == -999, NA)) %>% # sol: https://stackoverflow.com/questions/
  na.omit() %>%
  group_by(Date) %>%
  summarise(MP = as.numeric(min(MP, na.rm=TRUE))) # a serie eh baseada no minimo do dia
# nao precisamos das horas

# passando para tibble e tibblertime:
matter3 <- matter2 %>%
  #tibble::as_tibble() %>%
  #na.omit() %>% # eliminando os NA's
  mutate(Date = lubridate::ymd(Date)) %>% # usando ymd do lubridate
  as_tbl_time(index = Date)

head(matter3)
```

```
## # A time tibble: 6 x 2
## # Index: Date
##   Date      MP
##   <date>    <dbl>
## 1 2013-01-01 10.0
## 2 2013-01-02 10.0
## 3 2013-01-03 12.0
## 4 2013-01-04 11.0
## 5 2013-01-05 13.0
## 6 2013-01-06 10.0
```

Gráficos da série temporal  $Y_t$ :

```
# ----- plots serie:

# --- serie:
# ggplot:
# matter3 %>%
# ggplot() +
#   geom_line(aes(x = Date, y = MP), colour="orange") +
#   theme_minimal()+
#   labs(title="Série diária de MP")

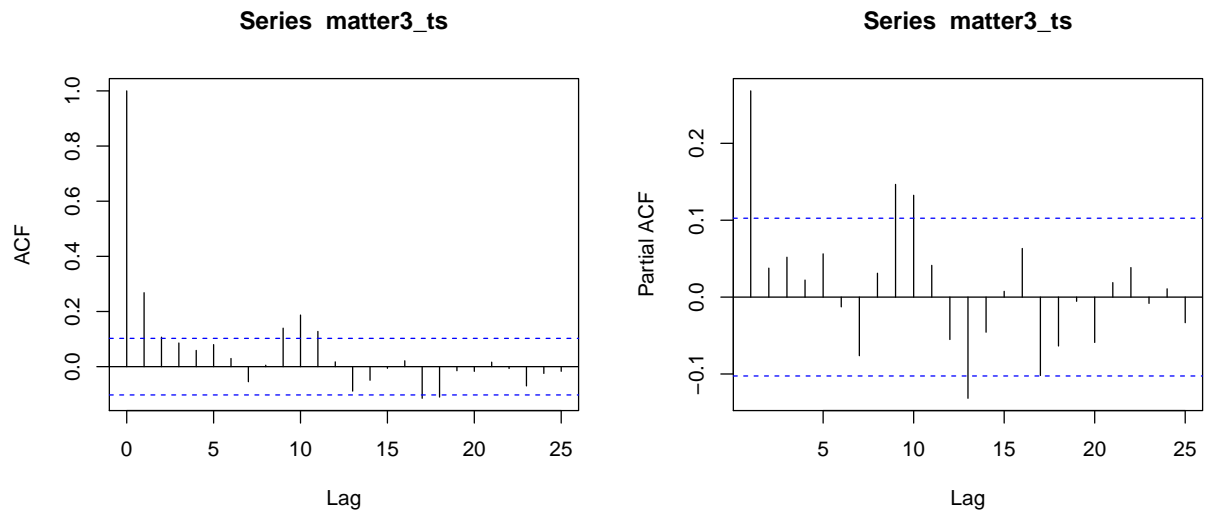
# dygraphs
library(dygraphs)
```

```
library(xts)
matter3_ts <- xts(matter3 %>% dplyr::select(MP), order.by=matter3$Date)
# dygraph(matter3_ts, elementId='matter3') %>% dyRangeSelector()# %>% dyUnzoom()
dygraph(matter3_ts, elementId='matter3', main="Série diária de MP" ) %>% dyRangeSelector() %>% dyUnzoom()
  dySeries("MP", color = "orange")

# rCharts:
# library(rCharts)
# library(rjson)
# dPlot(MP ~ Date, data = matter3, type="line")
```

Plots da ACF e PACF para analisar estacionariedade:

```
# ----- plots:
par(mfrow=c(1,2))
acf(matter3_ts)
pacf(matter3_ts)
```



### item ii)

Com base no gráfico da Função de Autocorrelação (FAC ou ACF em inglês), é possível inferir que a série é **estacionária**. Há um decaimento rápido da autocorrelação nos primeiros *lags* e, mesmo com uma certa variabilidade neste padrão, a correlação parece tender rapidamente a zero conforme  $h$  aumenta. Este comportamento indica a presença de estacionariedade na série. Se houvesse um decaimento lento da FAC, isto poderia indicar o efeito de uma memória de mais longo prazo nos dados. Consequentemente, teríamos a dependência (entre os valores de  $Y_t$ ) persistente ao longo do tempo e a não-estacionariedade da série, o que não é o caso.

### item iii)

Com base na análise descritiva apresentada no *item i*), o decaimento geométrico nos primeiros *lags* da FAC, sugere a existência de um modelo Autoregressivo  $A.R.(p)$ . Para identificar a ordem deste modelo, podemos

olhar para o gráfico da Função de Autocorrelação Parcial (FACP ou PACF em inglês). Este gráfico apresenta *spike* significativa (ordenado) apenas para  $h = 1$ . Por isso, o modelo de filtro linear adequado para o processo em questão seria um Modelo Autoregressivo de Ordem Um ( $A.R.(1)$ ):

$$Y_t = \mu_0 + \rho Y_{t-1} + \epsilon_t$$