Air quality project : Time Series Analysis

Yousra Cherkaoui Tangi - Lou Peltier - Guilhem Bois

08 janvier, 2020

Contents

Preliminary

Our project consists of time series analysis using R programming language. Our data set contains the responses of a gas mutisensor device deployed on the field in an italian city, in a significantly polluted area, at road level. The dataset contains 9358 instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device. Data were recorded from March 2004 to February 2005.

```r
dim(data)
```

```
## [1] 9471   17
```

```r
str(data)
```

```
## 'data.frame':    9471 obs. of  17 variables:
##  $ Date         : chr  "10/03/2004" "10/03/2004" "10/03/2004" "10/03/2004" ...
##  $ Time         : chr  "18.00.00" "19.00.00" "20.00.00" "21.00.00" ...
##  $ CO.GT.       : chr  "2,6" "2" "2,2" "2,2" ...
##  $ PT08.S1.CO.  : int  1360 1292 1402 1376 1272 1197 1185 1136 1094 1010 ...
##  $ NMHC.GT.     : int  150 112 88 80 51 38 31 31 24 19 ...
##  $ C6H6.GT.     : chr  "11,9" "9,4" "9,0" "9,2" ...
##  $ PT08.S2.NMHC.: int  1046 955 939 948 836 750 690 672 609 561 ...
##  $ NOx.GT.      : int  166 103 131 172 131 89 62 62 45 -200 ...
##  $ PT08.S3.NOx. : int  1056 1174 1140 1092 1205 1337 1462 1453 1579 1705 ...
##  $ NO2.GT.      : int  113 92 114 122 116 96 77 76 60 -200 ...
##  $ PT08.S4.NO2. : int  1692 1559 1555 1584 1490 1393 1333 1333 1276 1235 ...
##  $ PT08.S5.O3.  : int  1268 972 1074 1203 1110 949 733 730 620 501 ...
##  $ T            : chr  "13,6" "13,3" "11,9" "11,0" ...
##  $ RH           : chr  "48,9" "47,7" "54,0" "60,0" ...
##  $ AH           : chr  "0,7578" "0,7255" "0,7502" "0,7867" ...
##  $ X            : logi  NA NA NA NA NA NA ...
##  $ X.1          : logi  NA NA NA NA NA NA ...
```

Attribute	Description
Date	Date (DD/MM/YYYY)
Time	Time (HH.MM.SS)
CO(GT)	True hourly averaged concentration CO in mg/m^3 (reference analyzer)
PT08.S1(CO)	PT08.S1 (tin oxide) hourly averaged sensor response (nominally CO targeted)
NMHC(GT)	True hourly averaged overall Non Metanic HydroCarbons concentration in $microg/m^3$(reference analyzer
C6H6(GT)	True hourly averaged Benzene concentration in $microg/m^3$ (reference analyzer)
PT08.S2(NMHC)	PT08.S2 (titania) hourly averaged sensor response (nominally NMHC targeted)
NOx(GT)	True hourly averaged NOx concentration in ppb (reference analyzer)
PT08.S3(NOx)	PT08.S3 (tungsten oxide) hourly averaged sensor response (nominally NOx targeted)
NO2(GT)	True hourly averaged NO2 concentration in $microg/m^3$ (reference analyzer)
PT08.S4(NO2)	PT08.S4 (tungsten oxide) hourly averaged sensor response (nominally NO2 targeted)
PT08.S5(O3)	PT08.S5 (indium oxide) hourly averaged sensor response (nominally O3 targeted)
T	Temperature in Â°C
RH	Relative Humidity (%)
AH	AH Absolute Humidity

I)-Cleaning and prepocessing

1) Data cleaning

2) Data preprocessing

II)-Model fitting on the time series of interest

1) MA model

2) AR model

3) ARMA model

4) Residuals

5) GARCH model

6) Prediction intervals for the 10 most recent data

III)-Training on the times series of interest using explanatory times series

1) Preprocessing

2) Time varying coefficients

3) QLIK

4) Prediction

IV) Conclusion