Air pollution analysis with R

Dr James Tate
E: j.e.tate@its.leeds.ac.uk
drjamestate

Air quality

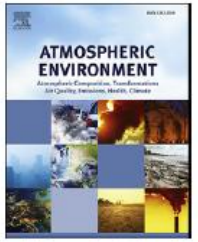
Sources & dispersion of pollutants

Atmospheric Environment 127 (2016) 163–175

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv



Understanding how roadside concentrations of NO_X are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees

Arwa Sayegh*, James E. Tate, Karl Ropkins

Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK

HIGHLIGHTS

- Air quality at urban, motorway, and tunnel sites has been studied and compared.
- New method has been developed for splitting traffic data to four traffic states.
- Deriving traffic influence on roadside NO_x depends on the quality of background NO_x.
- Different traffic states have been shown to have different influence on roadside NO_x.
- Roadside NO_x appears to reach a minimum at around 22 °C of ambient air temperature.

Sayegh, A., Tate, J., Ropkins, K. 2016. Understanding how roadside concentrations of NOx are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees. Atmospheric Environment, 127 (2016), 163-175, http://dx.doi.org/10.1016/j.atmosenv.2015.12.024

Air quality

Sources & dispersion of pollutants

Predictors used to model hourly roadside NO_x concentrations [$\mu g\ m^{-3}$].

Predictors	Description	Variable type
$[NO_x]_{bkgd}$	Background NO_x concentrations in $\mu g\ m^{-3}$	Continuous
$K_{near\text{-}bound}$	Hourly traffic densities on the bound nearer to the monitoring station in veh km^{-1}	Continuous
$K_{distant\text{-}bound}$	Hourly traffic densities on the opposite bound to the monitoring station in veh km^{-1}	Continuous
$\bar{U}$	Hourly wind speed in m s^{-1}	Continuous
θ	Hourly wind direction in degree (°)	Continuous
RH	Hourly relative humidity in percentage (%)	Continuous
T	Hourly air temperature in degree Celcius (°C)	Continuous
JD	Day of the year	Discrete (1–365/366)
Day	Day of the week	Discrete (1–7)
H	Hour of the day	Discrete (1–24)

Relative influence (%) of individual predictors on roadside NO_x concentrations [$\mu g\ m^{-3}$].

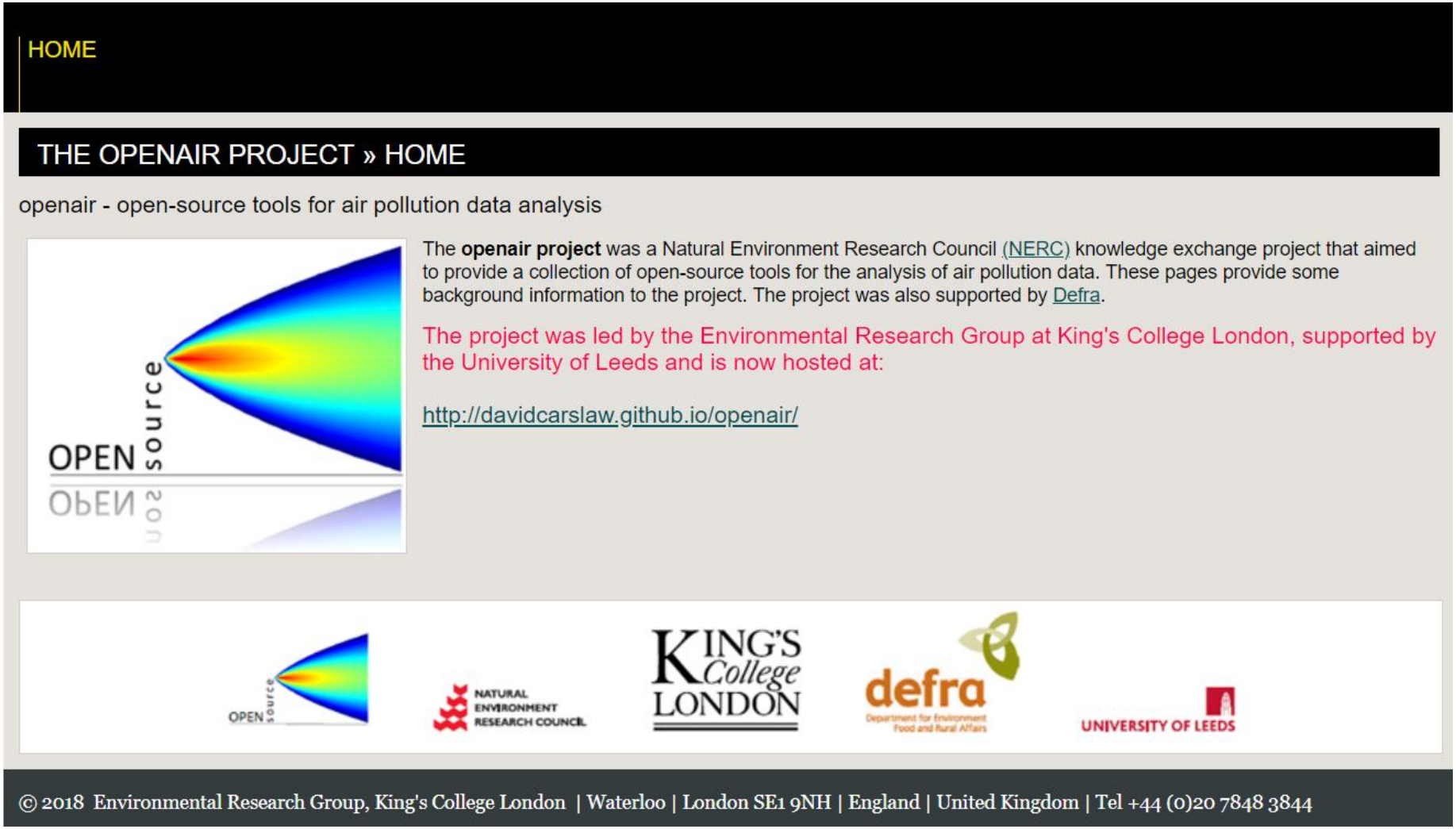
BRT model predictors	Urban		Open motorway	Motorway tunnel	Mean values[a]
	Monitored $[NO_x]_{bkgd}$	Modelled $[NO_x]_{bkgd}$			
$[NO_x]_{bkgd}$	37%	4%	3%	6%	4%
$K_{near\text{-}bound}$	19%	20%	10%	25%	18%
$K_{distant\text{-}bound}$	6%	10%	11%	–	11%
$\bar{U}$	12%	23%	9%	3%	12%
θ	9%	8%	31%	5%	15%
RH	4%	7%	8%	6%	7%
T	5%	10%	11%	18%	13%
JD	4%	12%	8%	17%	12%
Day	3%	4%	5%	10%	6%
H	1%	2%	4%	10%	5%

[a] Mean values exclude the urban site model with monitored $[NO_x]_{bkgd}$.

Sayegh, A., Tate, J., Ropkins, K. 2016. Understanding how roadside concentrations of NOx are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees. Atmospheric Environment, 127 (2016), pp163-175, DOI:10.1016/j.atmosenv.2015.12.024

www.openair-project.org/ R package

Manual & worked examples / data included



Carslaw, D., Ropkins, K. 2012. Openair - An r package for air quality data analysis. Environmental modelling & software, vol 27-28, pp. 52-61. DOI: 10.1016/j.envsoft.2011.09.008

DATA PREPARATION

Time stamp and prevailing meteorology (wind speed, direction)

- Handling *date* and *time* formats are a pain in most packages
- Use built-in *POSIXct* function to define. "*ct*" stands for calendar time
- Define *date* and *time* field in data frame as:
  - dd/mm/yyyy hh:mm GMT

- Many openair functions require co-ordinated wind speed & direction information with fields labelled as:
  - "ws" -> wind speed
  - "wd" -> wind direction

Beck, C. 2012. Handling date-times in R. http://biostat.mc.vanderbilt.edu/wiki/pub/Main/ColeBeck/datestimes.pdf

Pollution

London breaches annual air pollution limit for 2017 in just five days

Brixton Road in Lambeth has already broken legal limits for toxic air for the entire year, with many other sites across the capital set to follow

Damian Carrington

@dpcarrington

Friday 6 January 2017 10.03 GMT

Air pollution and traffic on Putney High Street on 3 January 2017, one of London's worst pollution hotspots. Photograph: Elizabeth Dalziel/Greenpeace

London has breached its annual air pollution limits just five days into 2017, a "shameful reminder of the severity of London's air pollution", according to campaigners.

CASE STUDY
Putney High Street, London

Putney monitoring sites: https://www.londonair.org.uk/

- North: Putney (urban background)
- Central: Putney High Street Façade (roadside)
- South: Putney High Street (kerbside)

DEFRA AURN:

https://uk-air.defra.gov.uk/

- Alternative background site at Teddington with meteorological data (prevailing wind speed & direction)

CASE STUDY
Putney High Street 2016, London

- Resources:
  - Hourly data for 2016 "Putney_2016.csv". Fields:
    - Datetext [Date and Time field GMT]
    - no2_Putney [Nitrogen dioxide (NO_2) concentrations (ratified) at Putney station]
    - nox_Putney [Oxides of Nitrogen (NO_x) concentrations (ratified) at Putney station]
    - no2_PutneyHS [Nitrogen dioxide (NO_2) concentrations (ratified) at Putney High Street station]
    - nox_PutneyHS [Oxides of Nitrogen (NO_x) concentrations (ratified) at Putney High Street station]
    - no2_PutneyHSF [Nitrogen dioxide (NO_2) concentrations (ratified) at Putney High Street Facade]
    - nox_PutneyHSF [Oxides of Nitrogen (NO_x) concentrations (ratified) at Putney High Street Facade]
    - wd [prevailing wind direction from nearby Teddington station]
    - ws [prevailing wind speed from nearby Teddington station]
    - temp [ambient temperature from nearby Teddington station]
  - Script file "airquality.r"

NOTE: all provided data is "ratified" i.e. quality checked and adjusted to calibrations. Initially "raw" data is available on websites but is termed "provisional"

Beck, C. 2012. Handling date-times in R. http://biostat.mc.vanderbilt.edu/wiki/pub/Main/ColeBeck/datestimes.pdf

CASE STUDY
Putney High Street 2016, London

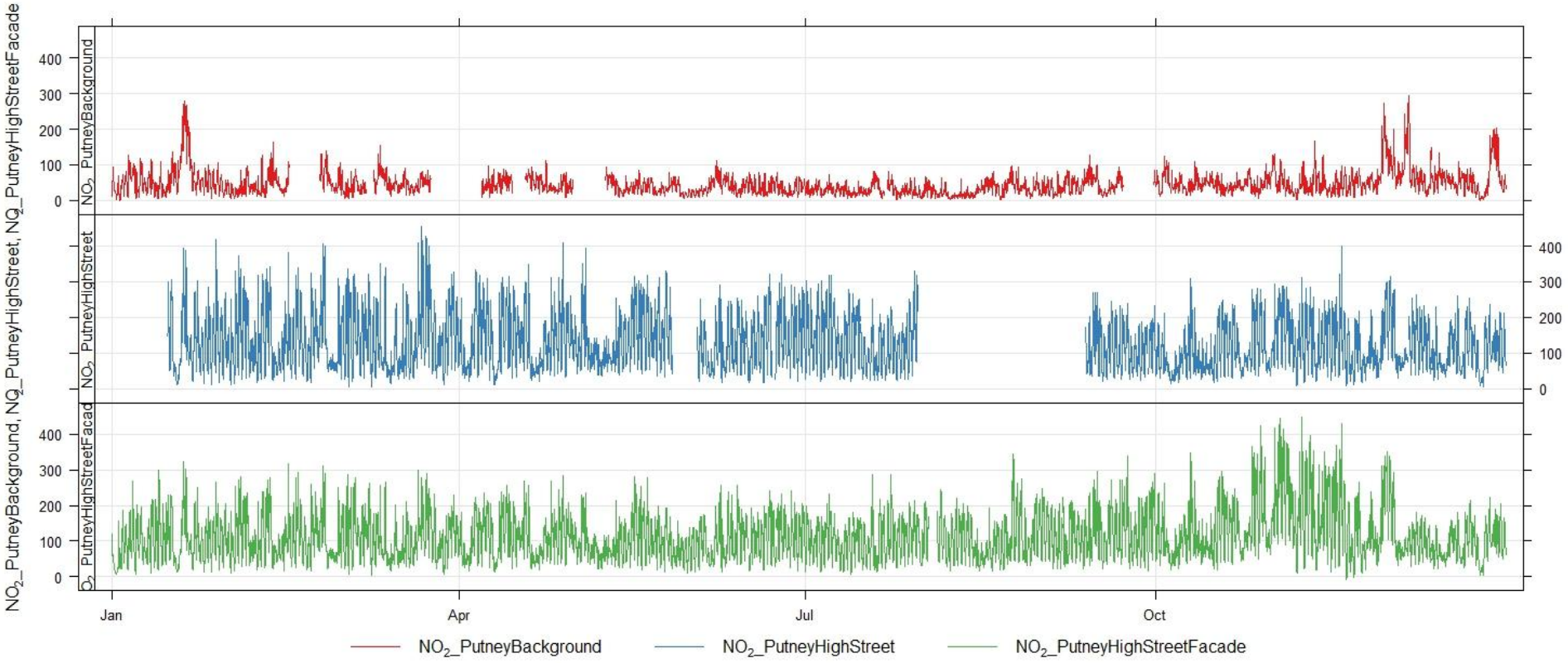
- Install and load openair
- Import data.frame

```
#install and load package | OPENAIR
#note: openair now relies on several packages, which if not available need to be installed
install.packages("openair")
library(openair)

# import data.frame i.e. read .csv file
aq16 <- read.csv("Putney_2016.csv", header=T, na.strings="NA")

# useful function to check data.frame import
summary(aq16)

# define date & time format, add to data.frame, rename headers
datetime <- as.POSIXct(strptime(aq16$datetext, format = "%d/%m/%Y %H:%M", "GMT"))
aq16 <- cbind(aq16, datetime)
names(aq16)[1:11] <- c("datetext","no2_Putney","nox_Putney","no2_PutneyHS","nox_PutneyHS",
                       "no2_PutneyHSF","nox_putneyHSF","wd","ws","temp","date")
```

CASE STUDY
Putney High Street 2016, London

- Basic time series check using openair timePlot function

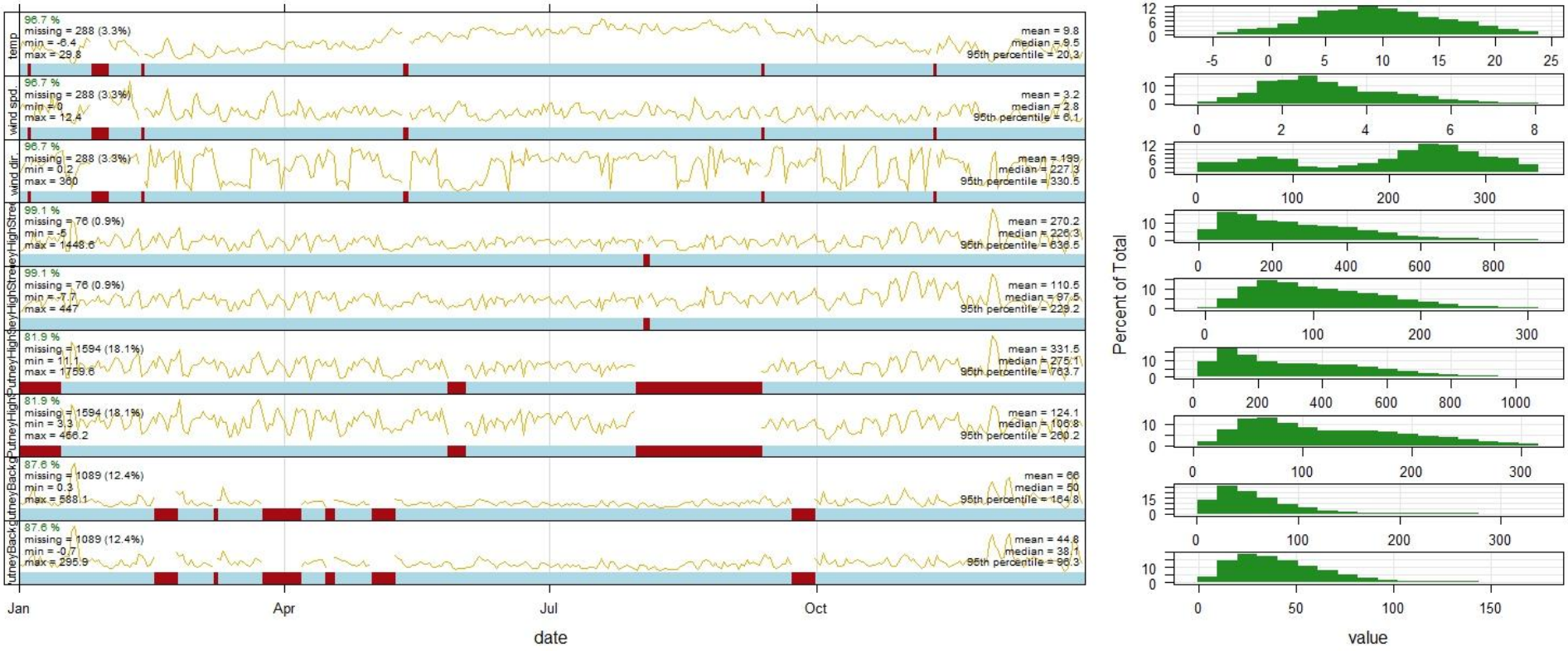
```
#timePlot function - time series plot of NO2 measurements
timePlot(aq16, pollutant =
c("no2_PutneyBackground","no2_PutneyHighStreet","no2_PutneyHighStreetFacade"))
```



CASE STUDY
Putney High Street 2016, London

- Rapid summary of time series, distributions and summary stats etc using openair summaryPlot function

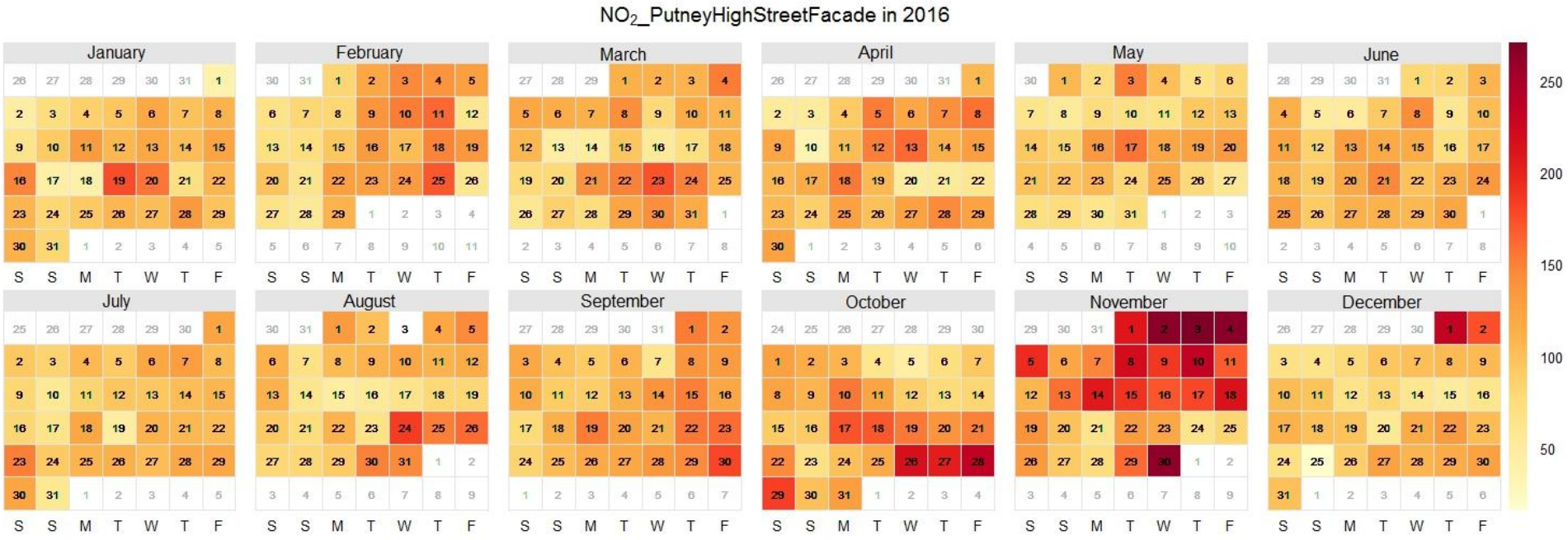
```
#summaryPlot - rapid summary of time series, distributions and summary stats
summaryPlot(aq16)
```



CASE STUDY
Putney High Street 2016, London

- Rapid summary of time series, distributions and summary stats etc using openair summaryPlot function

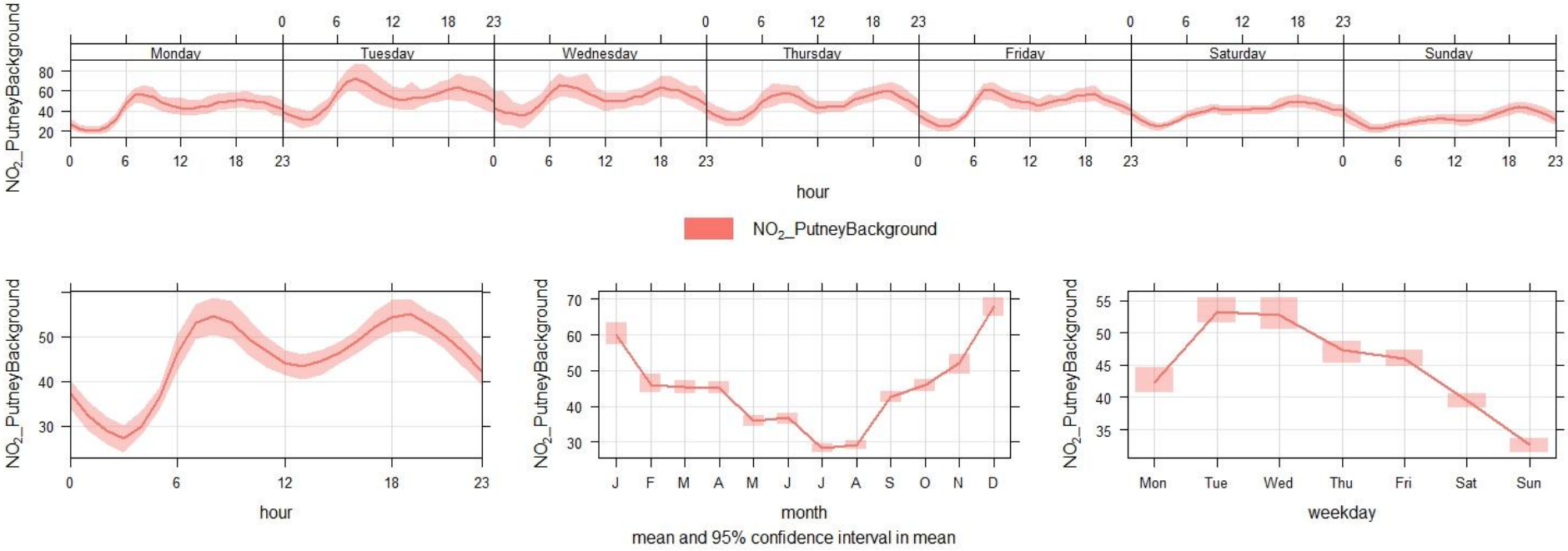
```
#calendarPlot function of NO2 concentrations and then wind speed
calendarPlot(aq16, pollutant = "no2_PutneyBackground", year = 2016)
calendarPlot(aq16, pollutant = "no2_PutneyHighStreet", year = 2016)
calendarPlot(aq16, pollutant = "no2_PutneyHighStreetFacade", year = 2016)
calendarPlot(aq16, pollutant = "ws", year = 2016)
```



CASE STUDY
Putney High Street 2016, London

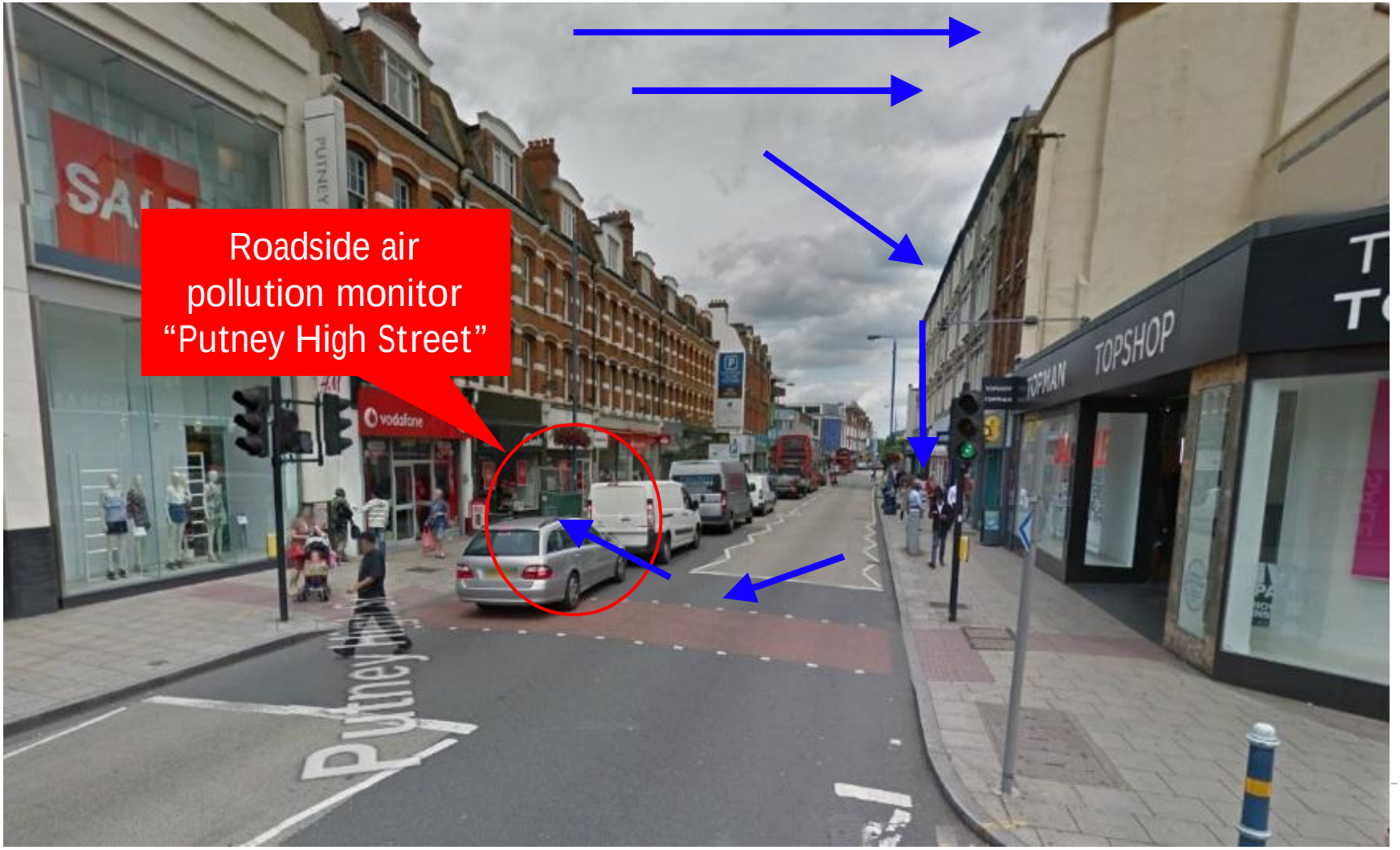
- Average variation by hour, day of week and month of year with openair timeVariation function

```
#timeVariation in NO2 concentrations
timeVariation(aq16, pollutant = "no2_PutneyBackground")
timeVariation(aq16, pollutant = "no2_PutneyHighStreet")
timeVariation(aq16, pollutant = "no2_PutneyHighStreetFacade")
```



CASE STUDY
Putney High Street 2016, London

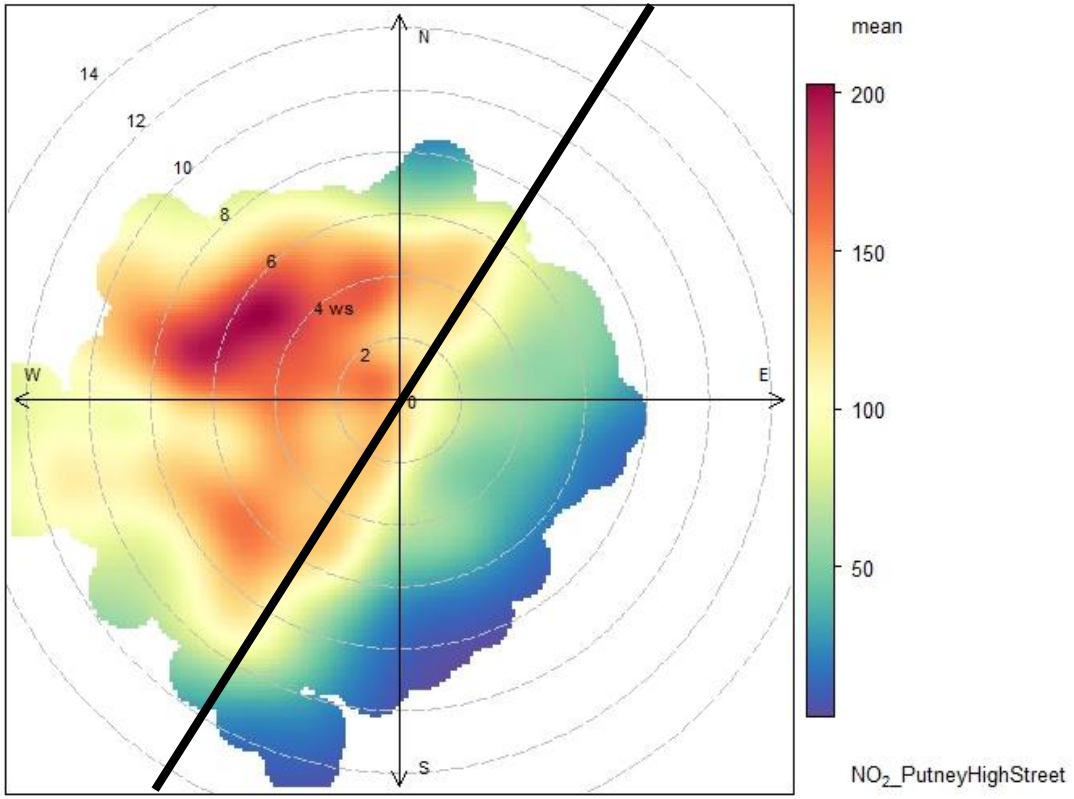
- Air pollution dispersion in “street canyons”



CASE STUDY
Putney High Street 2016, London

- Exploring the variation in concentrations with wind speed and direction with the openair polarPlot function (surface fitting by a Generalized Additive Models (GAM))

```
#polar plot for the Putney High Street monitor
polarPlot(aq16, pollutant = "no2_PutneyHighStreet")
```



CASE STUDY
Putney High Street 2010-16, London

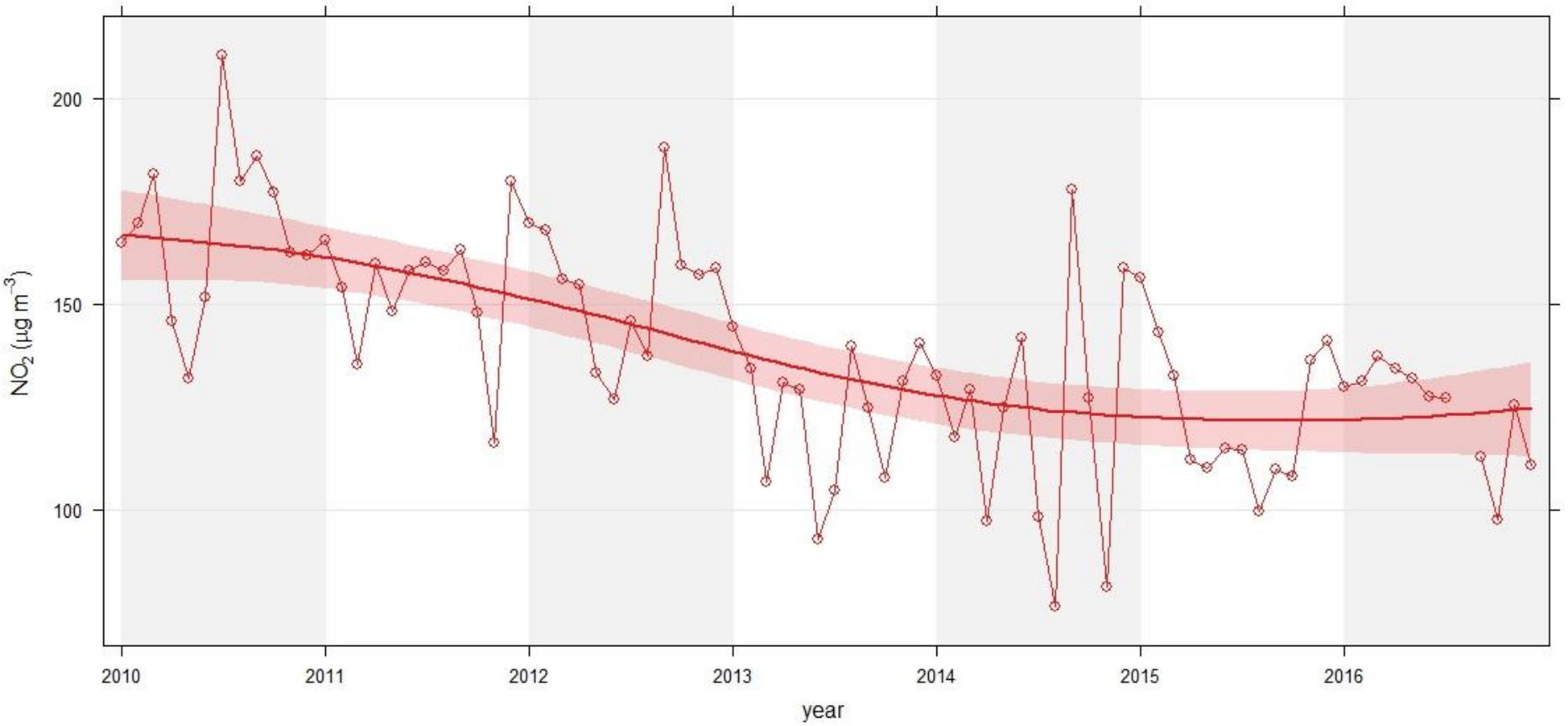
- Long term trend analysis using openair smoothTrend and TheilSen functions
- Import data.frame and define date format

```
# long term trend analysis
# import 2010-2016 data.frame
aqLONG <- read.csv("Putney_AQ_2010_2016.csv", header=T, na.strings="NA")

# useful data.frame check command
summary(aqLONG)

# define date & time format, add to data.frame, rename headers
datetime <- as.POSIXct(strptime(aqLONG$datetext, format = "%d/%m/%Y %H:%M", "GMT"))
aqLONG <- cbind(aqLONG, datetime)
names(aqLONG)[1:8] <-
c("datetext","no2_PutneyBackround","nox_PutneyBackround","no2_PutneyHighStreet","nox_PutneyHiSt
reet","no2_PutneyHighStreetFacade","nox_PutneyHighStreetFacade","date")
```

CASE STUDY
Putney High Street 2010-16, London

- Long term trend analysis using openair smoothTrend function

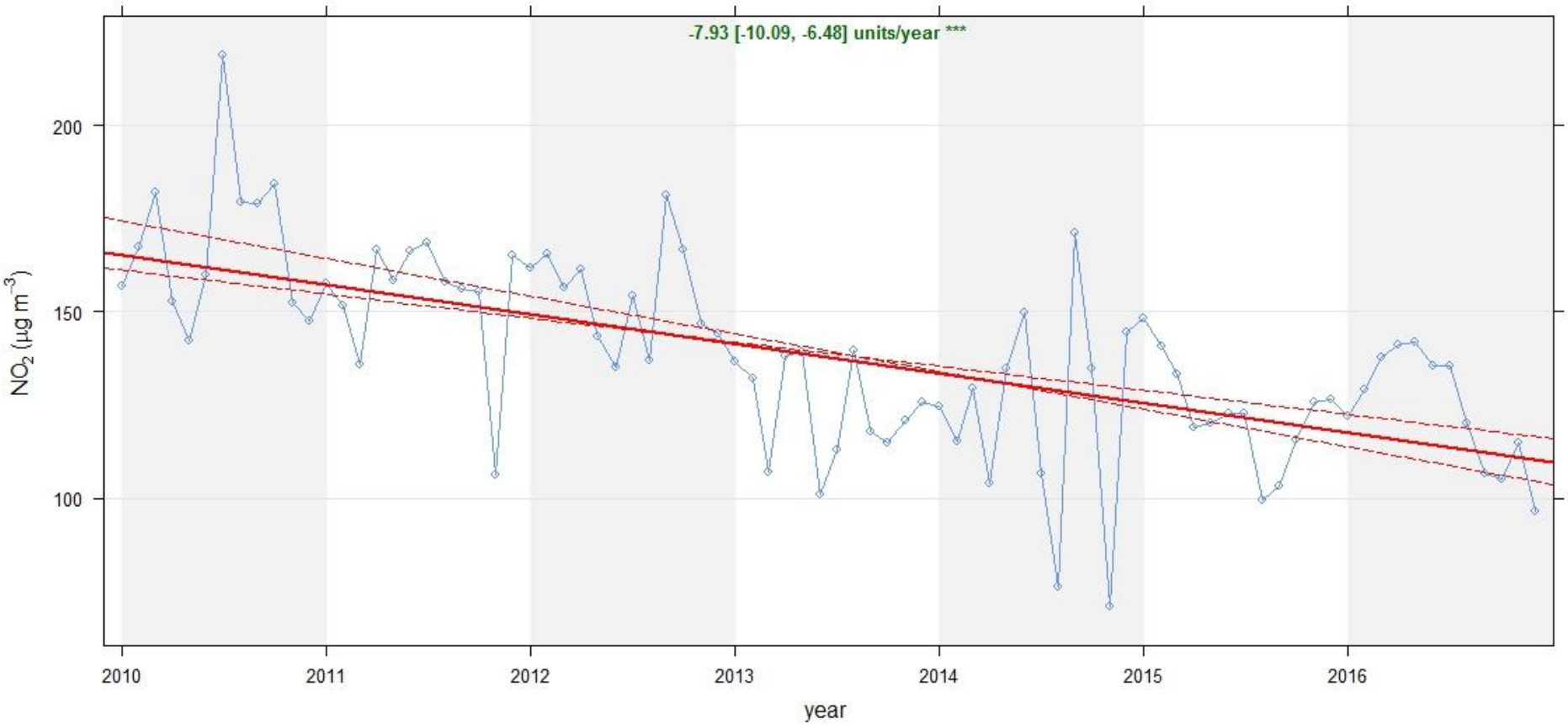
```
# the smoothTrend function calculates trends in the monthly mean concentrations
smoothTrend(aqLONG, pollutant = "no2_PutneyHighStreet", ylab = "NO2 (ug/m3)")
```



CASE STUDY
Putney High Street 2010-16, London

- Long term trend analysis using openair TheilSen function

```
# ascertain trends using the 'TheilSen function'
TheilSen(aqLONG, pollutant = "no2_PutneyHighStreet", ylab = "NO2 (ug/m3)", deseason = TRUE)
```



www.openair-project.org/ R package

Further functions

- Investigate pollution levels near your home and / or office
- Use data import functions:
  - National air pollution monitoring database / website [AURN]
  - Kings College London air pollution monitoring database / website [KCL]
- Find site:
  - DEFRA AURN interactive map: https://uk-air.defra.gov.uk/interactive-map?network=aurn
  - London Air: https://www.londonair.org.uk/london/asp/publicdetails.asp
  - And site code in document: https://uk-air.defra.gov.uk/assets/documents/reports/cat13/1011121246_Site_Classifications_Report_v2.pdf

e.g. Leeds Centre: Code = "LEED" Leeds Centre SE 29976 34268 429976 434268 Background Urban

```
# Leeds Centre AURN
leedsaq <- importAURN(site = "LEED", year = 2000:2017, pollutant = c("nox", "no2"))
TheilSen(leedsaq, pollutant = "no2", ylab = "NO2 (ug/m3)", deseason = TRUE)

# Headingley kerbside Leeds AURN
headaq <- importAURN(site = "LED6", year = 2000:2017, pollutant = c("nox", "no2"))
TheilSen(headaq, pollutant = "no2", ylab = "NO2 (ug/m3)", deseason = TRUE)
```

Leeds Centre AURN

Site Information for Leeds Centre (UKA00222)

Site Information | Pollutants Measured | Networks

Leeds Centre Information

The monitoring station is located within a self-contained, air-conditioned housing located approximately 30 metres from a busy 4-lane inner-city road, the A660 which is subject to periodic congestion during peak periods. The location is approximately 150 metres from an urban motorway, the A58(M). The surrounding area is generally open and comprises a busy urban setting road network.

UK-AIR ID: UKA00222

EU Site ID: GB0584A

Environment Type: Urban Background

Altitude (metres): 78

www.openair-project.org/ R package

Some further useful functions

Plot concentrations on a map with the GoogleMapsPlot function. It is based on the Google Static Maps API (http://code.google.com/apis/maps/documentation/staticmaps/)

Evaluate the statistical performance of models against observations. The modStats function provides numerous evaluation statistics for comparing models against measurements and models against other models.