The open-source air pollution project

openair

David Carslaw

The UK Air Quality Forecasting Seminar
16th July 2009

Outline

Outline

Opportunities and barriers

Analysis of measurement and model output data

Opportunities

- The analysis of air quality data can provide important insights into air pollution
- There is a huge amount of data available
- Insightful analysis provides the *evidence* to support air quality management decisions

Barriers

- No consistent set of tools available to carry out analysis
- Tools can be spread across many different software applications
- Many useful approaches are simply unavailable
- Lack of time, money or ideas about what can be done

The openair project

Summary of project

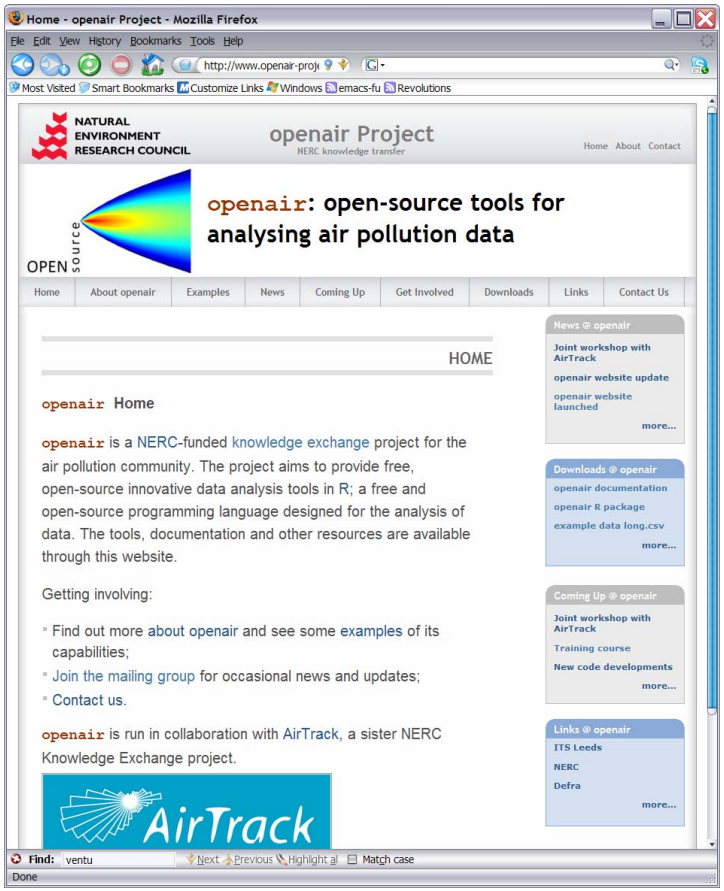
Key points

- 3-year NERC project to October 2011
- Develop and make available open-source data analysis tools to AQ community
- Use **R** statistical software as the platform
  - Highly capable software for "programming with data"
  - Develop a "package" of tools and progressively include advanced methods not widely available
  - Transparency — all code open to scrutiny

openair website

Central resource for the project

- Available at www.openair-project.org
- **openair** package – development version
- All documentation, data sets etc.
- Mailing list and newsletter
- Sister NERC project *AirTrack* at the University of Lancaster with complimentary aims
- Joint **openair**/AirTrack workshop, London, 1st October 2009



Data analysis

How best to analyse data?

- Data analysis is most useful when built around specific questions, however...
- Exploratory data analysis can be very insightful and is under-used — but time consuming

John Tukey sums it up:

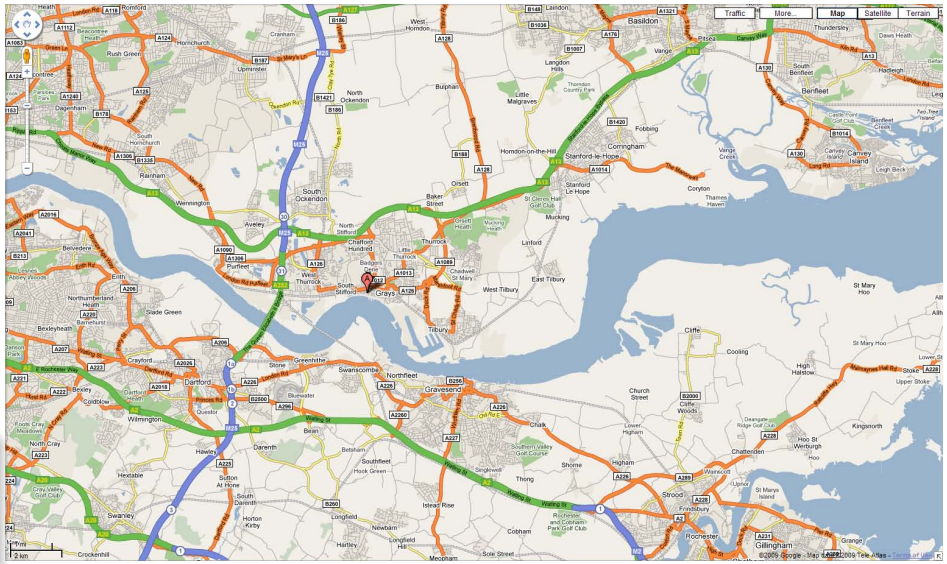
> *"The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data."*

Outline

Example analysis at a background site

Thurrock — east of the M25

- Import a few years of data for a range of pollutants
- Examples of how some of the **openair** functions can be applied



Example

```
tk1 = import("d:/mydata/thurrock.csv")
```

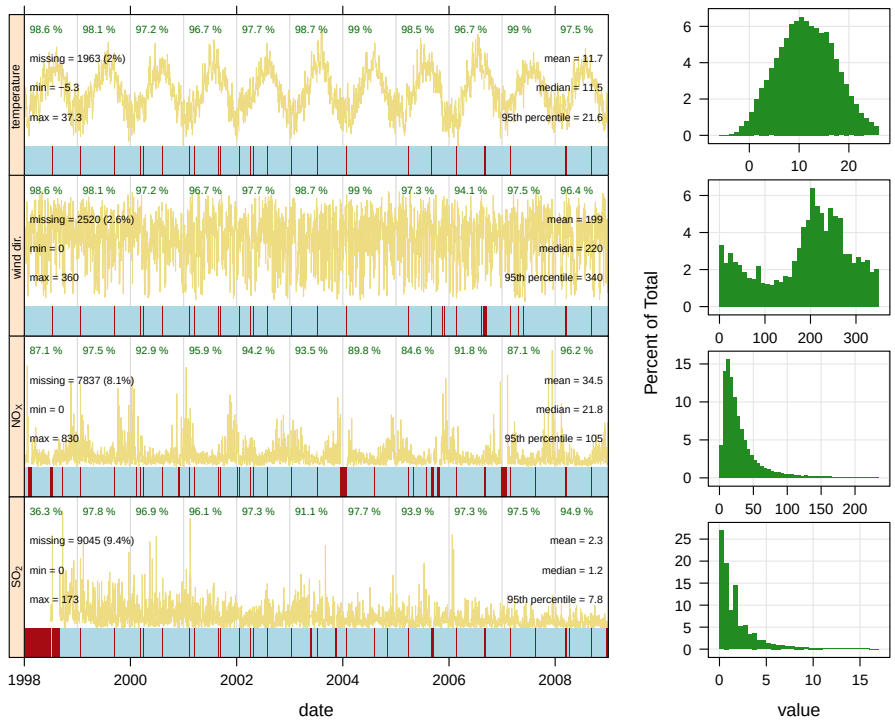
Quickly summarise data

The summarise function

- Always a good idea to look at data first before doing anything more serious
- The `summarise` function provides a way to do this rapidly

Example

```
summarise(tk1)
```



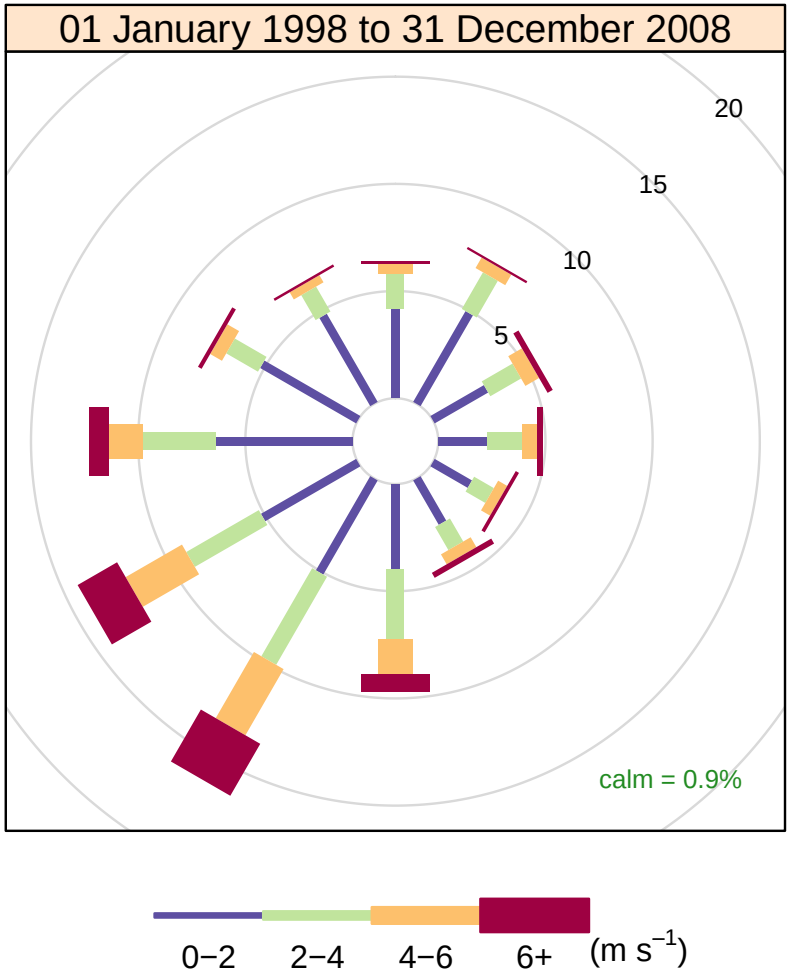
What do the met conditions look like?

The wind.rose function

- Plot a traditional wind rose using the `wind.rose` function
- lots of options to control how the data are plotted

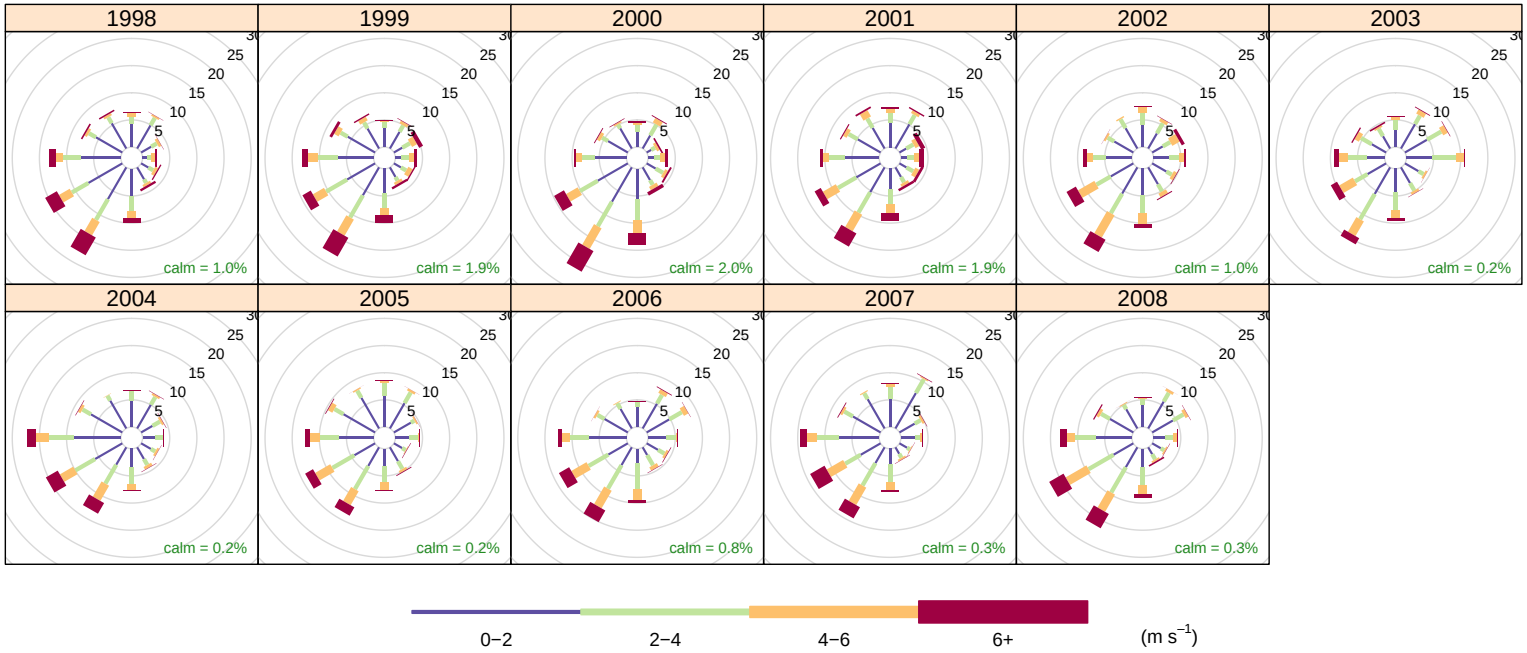
Example

```
wind.rose(tk1)
```



Wind roses by year

Plot by year, month, hour of the day...



Example

wind.rose(tk1, type = "year")

How do concentrations vary in time?

The time.variation function

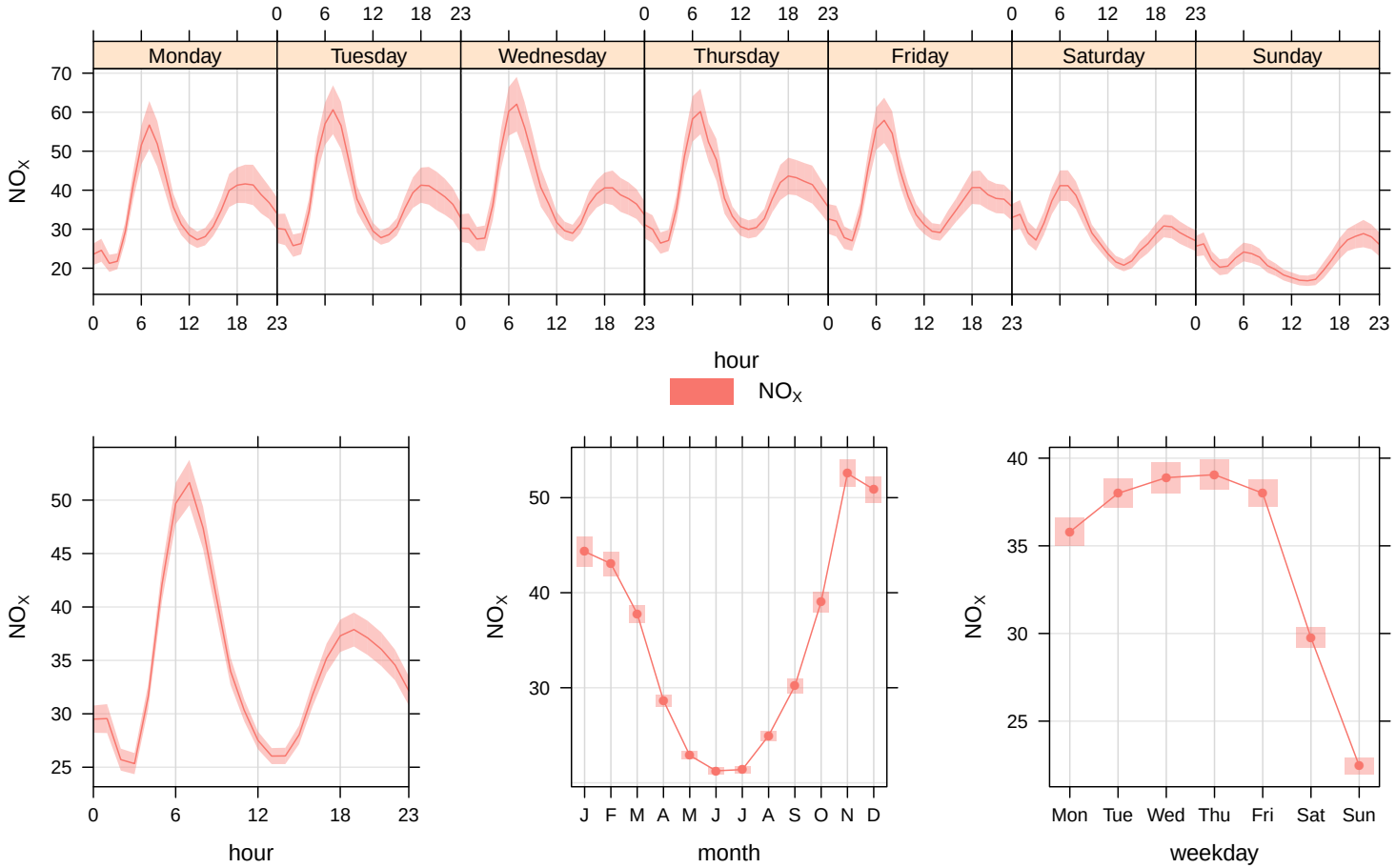
- The temporal variation in concentrations can provide important clues as to the source
- Sources can vary differently by hour of the day, day of the week and season
- Enhanced with further information
  - Traffic data
  - Meteorological data e.g. boundary layer height, atmospheric stability
- Excellent for model evaluation

Example

time.variation(tk1, pollutant = "nox")

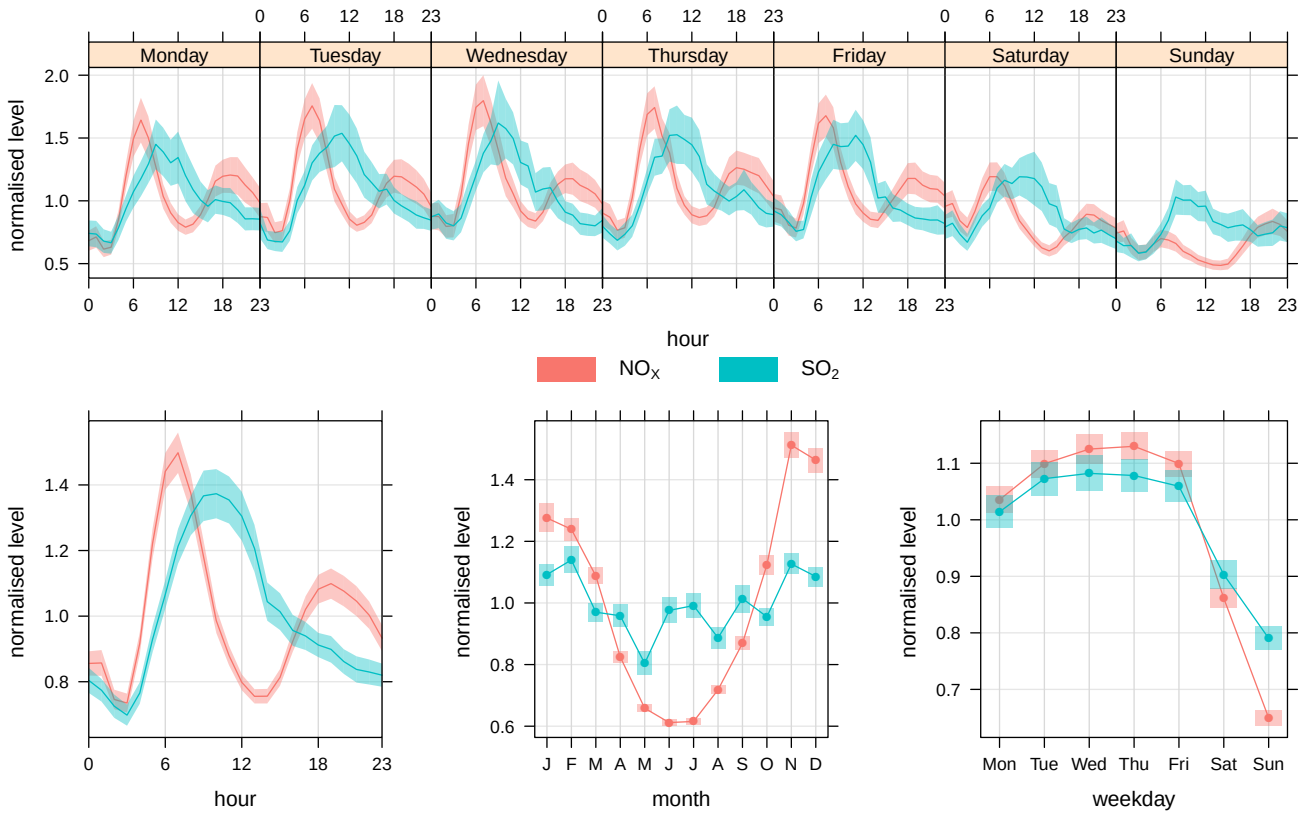
How do concentrations vary in time?

The time.variation function



How do concentrations vary in time?

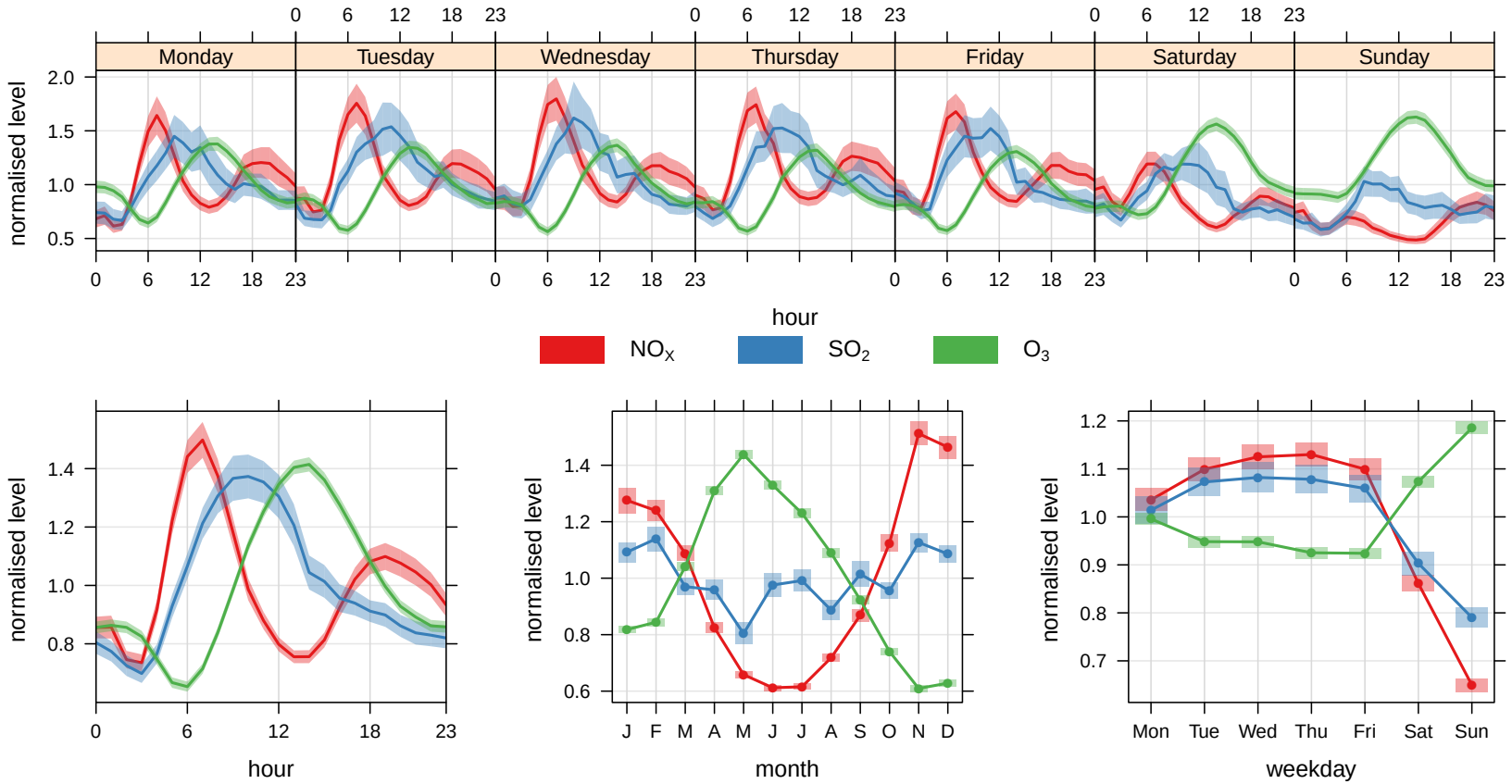
Two or more pollutants at once (SO_2 and NO_X)



Normalising the concentrations helps greatly when comparing different pollutants

How do concentrations vary in time?

Two or more pollutants at once (SO_2, NO_X and O_3)



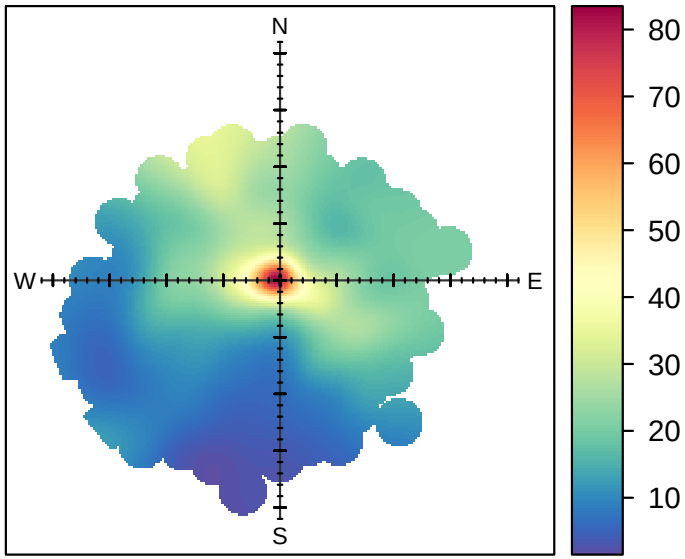
Very different behaviours and underlying reasons for differences

Polar plots and source identification

Concentrations by wind speed and direction

Variation with wind speed *and* direction can help identify sources and source characteristics[1]

- Tall stack emissions vs. ground-level sources
- Wind-blown sources e.g. particle re-suspension
- Hot buoyant plumes e.g. aircraft jets
- Local street canyon mixing



Example

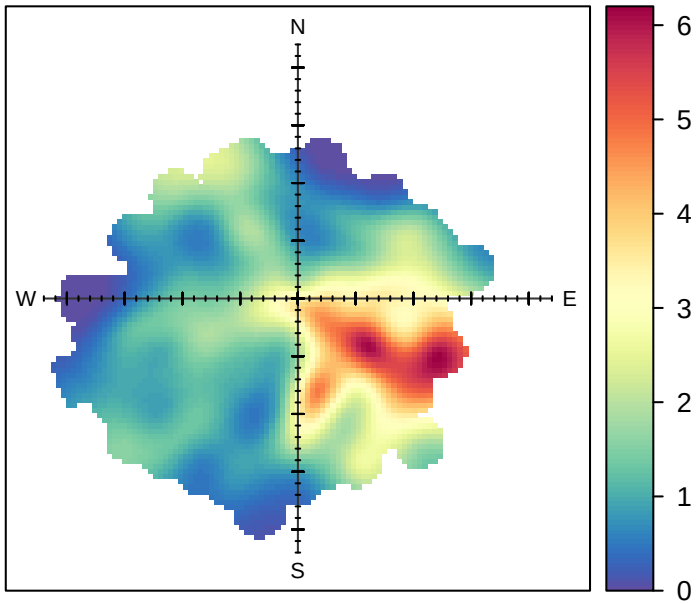
```
polar.plot(tk1, pollutant = "nox")
```

[1] Carslaw et al. (2006). Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport. *Atmos. Env.*, 40 (28), 5424-5434.

Polar plots and source identification

Concentrations by wind speed and direction

- The plot for SO_2 is markedly different to NO_X
- Evidence of at least three sources
  - Can be shown to be a refinery, power station and industrial sources



Example

```
polar.plot(tk1, pollutant = "so2")
```

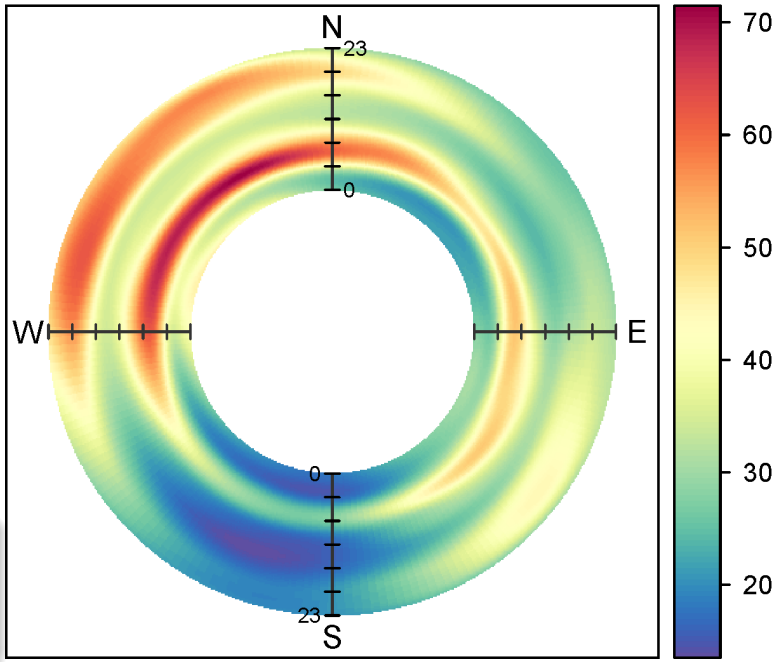
Temporal polar plots

Concentrations by wind direction and time

- Plot as an annulus
- Consider how concentrations vary by hour of the day, day of the week, season or trend by wind direction
- For NO_X highest concentrations at night from north-west

Example

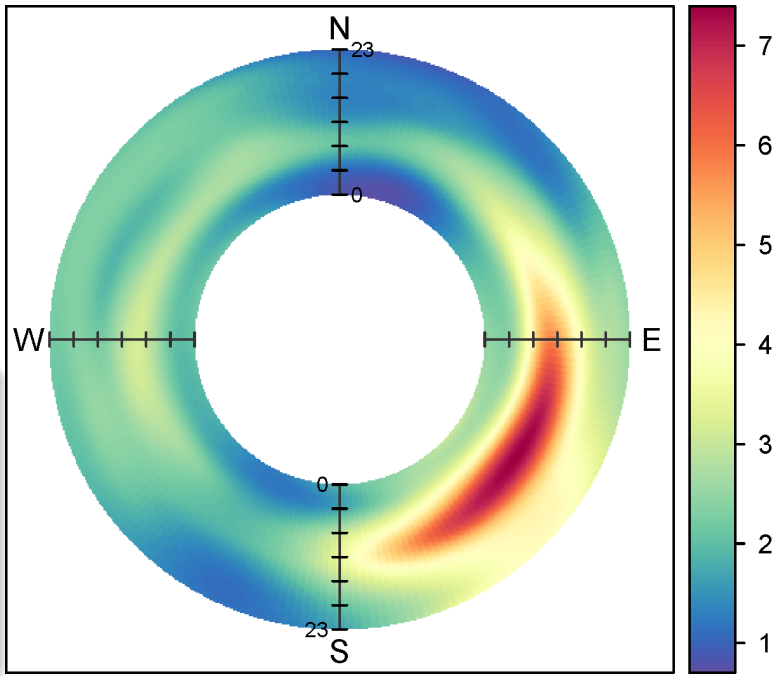
```
polar.annulus(tk1, pollutant = "nox")
```



Temporal polar plots

Concentrations by wind direction and time

- The SO_2 plot is markedly different to NO_X
- Concentrations highest during daytime and from south-east



Example

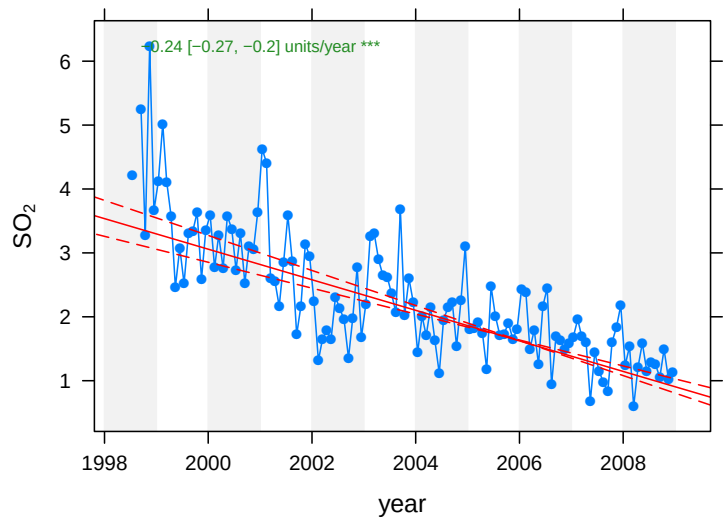
polar.annulus(tk1, pollutant = "so2")

- Can choose type = "weekday", "season" and "trend"

Trends

Trend analysis in openair

- Trends are an important component of air quality analysis
- **Mann-Kendall** analysis often used for environmental time series
- Consider monthly trends with option to de-seasonalise the data
- Use bootstrap simulation techniques to estimate uncertainties and block bootstrap to deal with autocorrelation

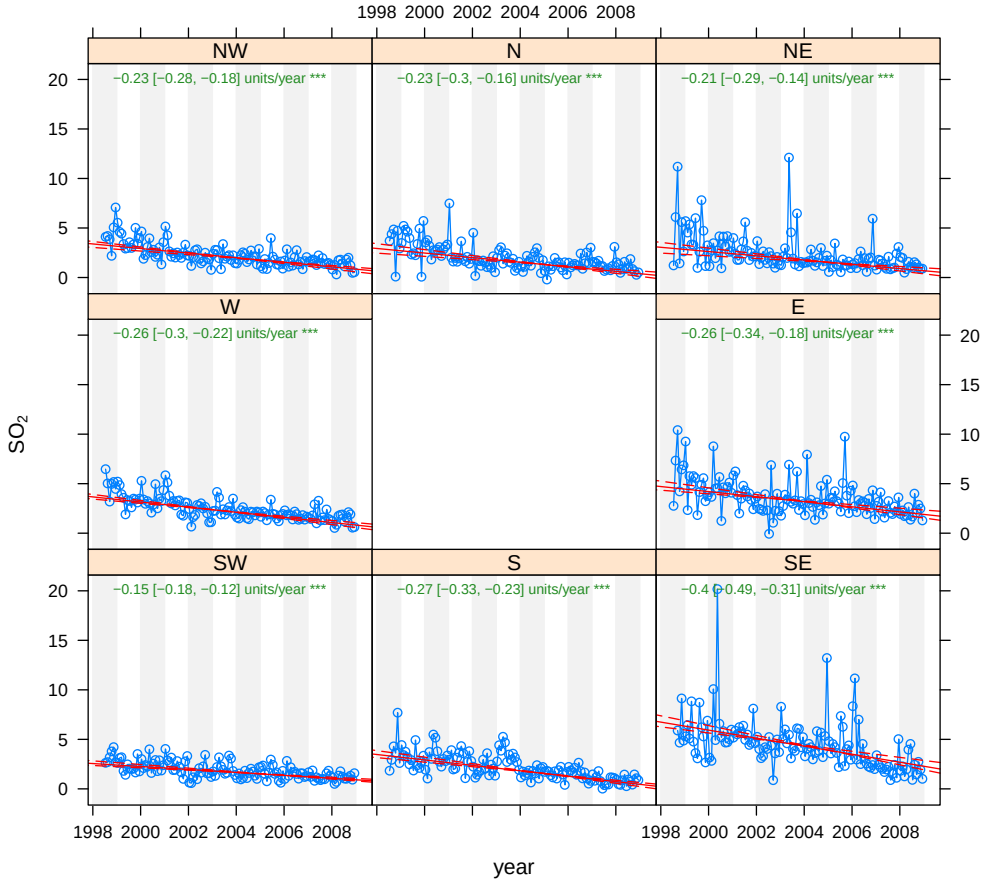


Example

MannKendall(tk1, pollutant = "so2")

Trends

Can consider trends in many different ways



Trends can be ‘conditioned’ by many different variables—here by wind direction

Outline

Developments

1. Reviewing scientific literature and will adopt promising approaches
   - Examples include, improved temporal characterisation e.g. Fourier analysis, change-point detection
   - Better quantitative analysis
2. Better support for model evaluation
   - Automate the process of evaluating models
   - Develop a range of metrics
3. The **openair** package
   - Graphical-user interface (GUI)?
   - Remote repository with full version control and easier installation
   - Develop documentation
   - Reproducible analyses using *Sweave*, R and LaTeX

Developments

1. Reviewing scientific literature and will adopt promising approaches
   - Examples include, improved temporal characterisation e.g. Fourier analysis, change-point detection
   - Better quantitative analysis
2. Better support for model evaluation
   - Automate the process of evaluating models
   - Develop a range of metrics
3. The **openair** package
   - Graphical-user interface (GUI)?
   - Remote repository with full version control and easier installation
   - Develop documentation
   - Reproducible analyses using *Sweave*, R and LaTeX

Developments

1. Reviewing scientific literature and will adopt promising approaches
   - Examples include, improved temporal characterisation e.g. Fourier analysis, change-point detection
   - Better quantitative analysis
2. Better support for model evaluation
   - Automate the process of evaluating models
   - Develop a range of metrics
3. The **openair** package
   - Graphical-user interface (GUI)?
   - Remote repository with full version control and easier installation
   - Develop documentation
   - Reproducible analyses using *Sweave*, R and LaTeX

Thank you for you attention!

Questions?

David Carslaw
d.c.carslaw@its.leeds.ac.uk