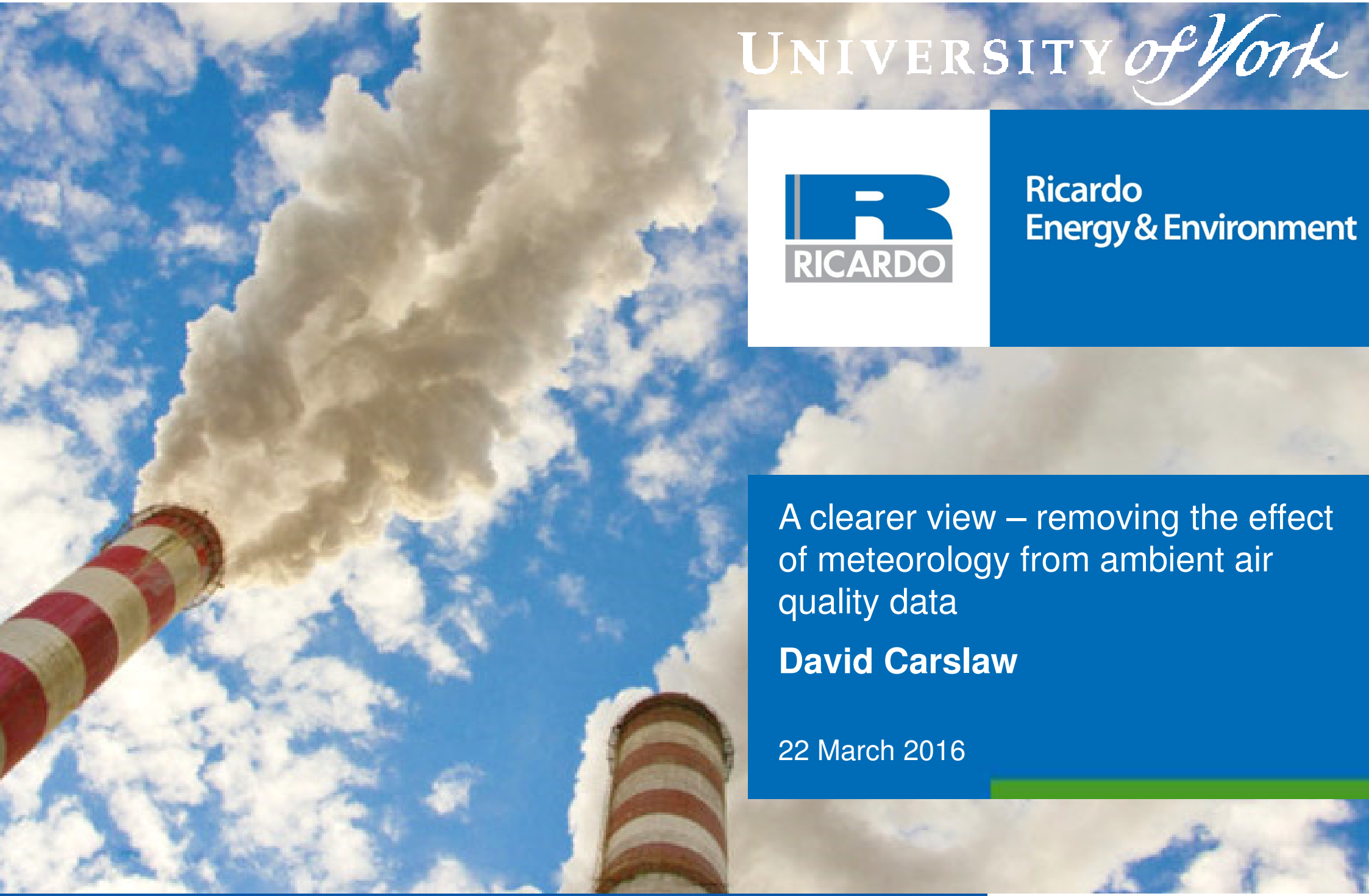


UNIVERSITY of York

Ricardo Energy & Environment

A clearer view – removing the effect of meteorology from ambient air quality data

David Carslaw

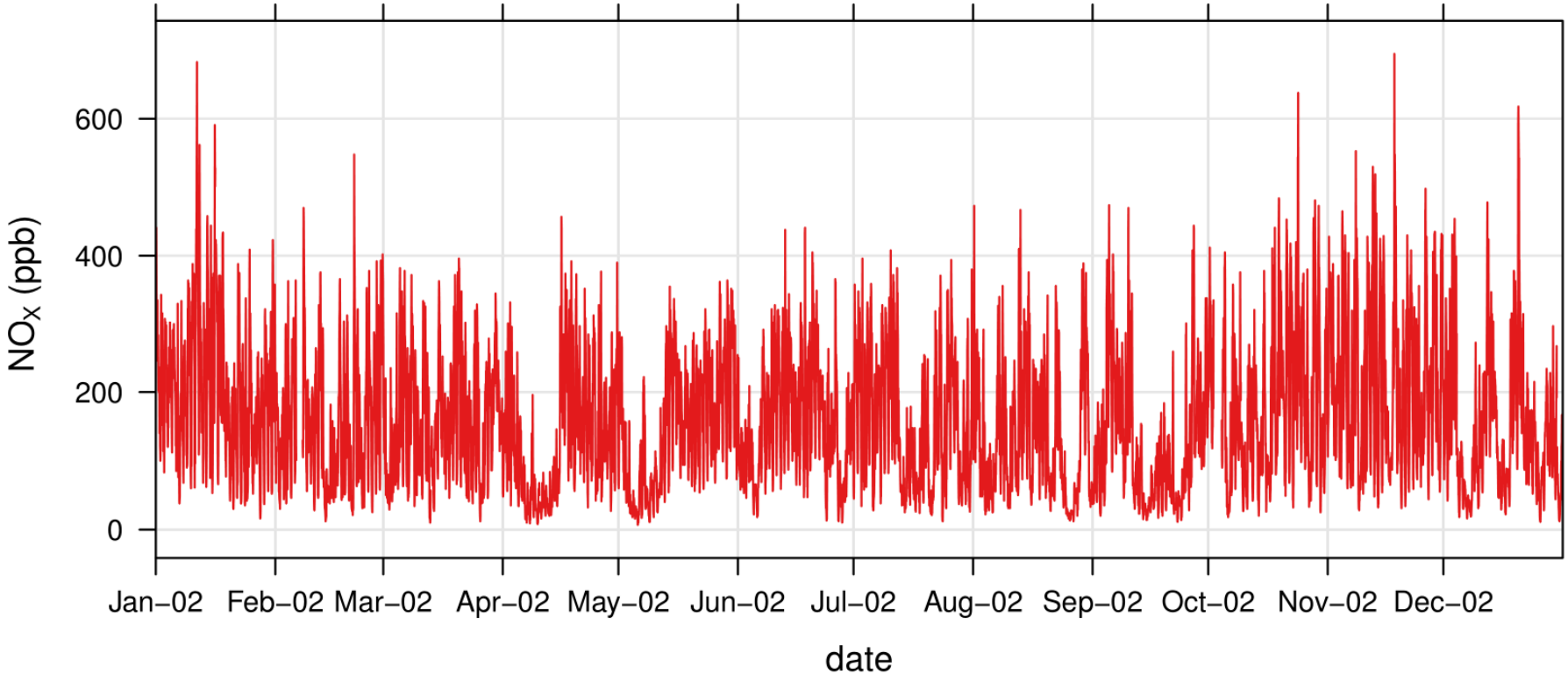
22 March 2016

This talk

1. Brief overview of **openair** and its status – topic for today is an extension of the work in this area
2. How meteorology affects air pollution and what can be done to 'remove' it
3. Examples of removing meteorological variation
   - Shipping emissions of SO_2
   - Particle number counts – fuel effects and after-treatment technologies
   - Trends in NO_x and NO_2 at selected Scottish roadside sites
4. Concluding remarks

The grand challenge…

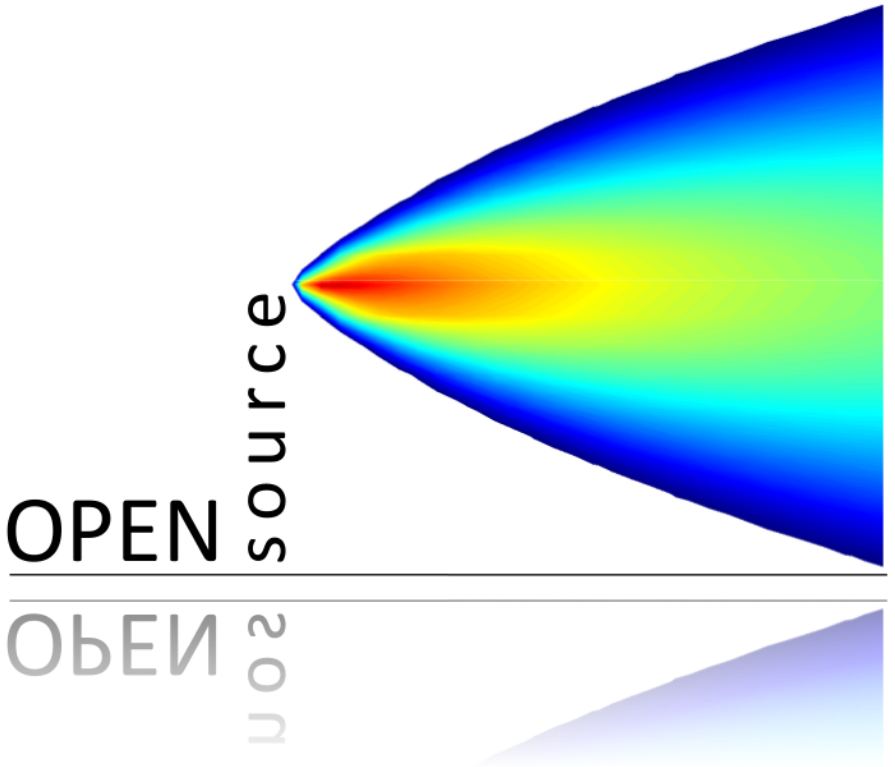
How to extract meaning from this…?



… and how do we know what changes have occurred and when they happened?

The openair project

- Started in 2008 with 3-year NERC funding with additional support from Defra
- **Aim: to make innovative open-source data analysis tools freely available to the air quality community**
  - *Sub aim*: As much as possible, no programming knowledge required by users
- Use software called R
  - Often thought of as statistical software
  - But it is really a programming language ***specifically designed for data analysis***
  - Usage and capabilities continue to grow at a rapid rate (> 8,000 'packages')
- **openair** is one of these R packages dedicated to the analysis of air quality data



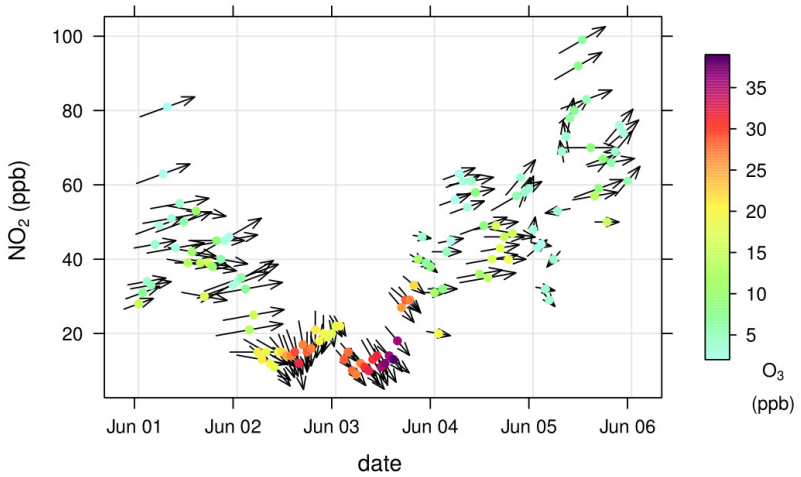
R – software that can do many things…

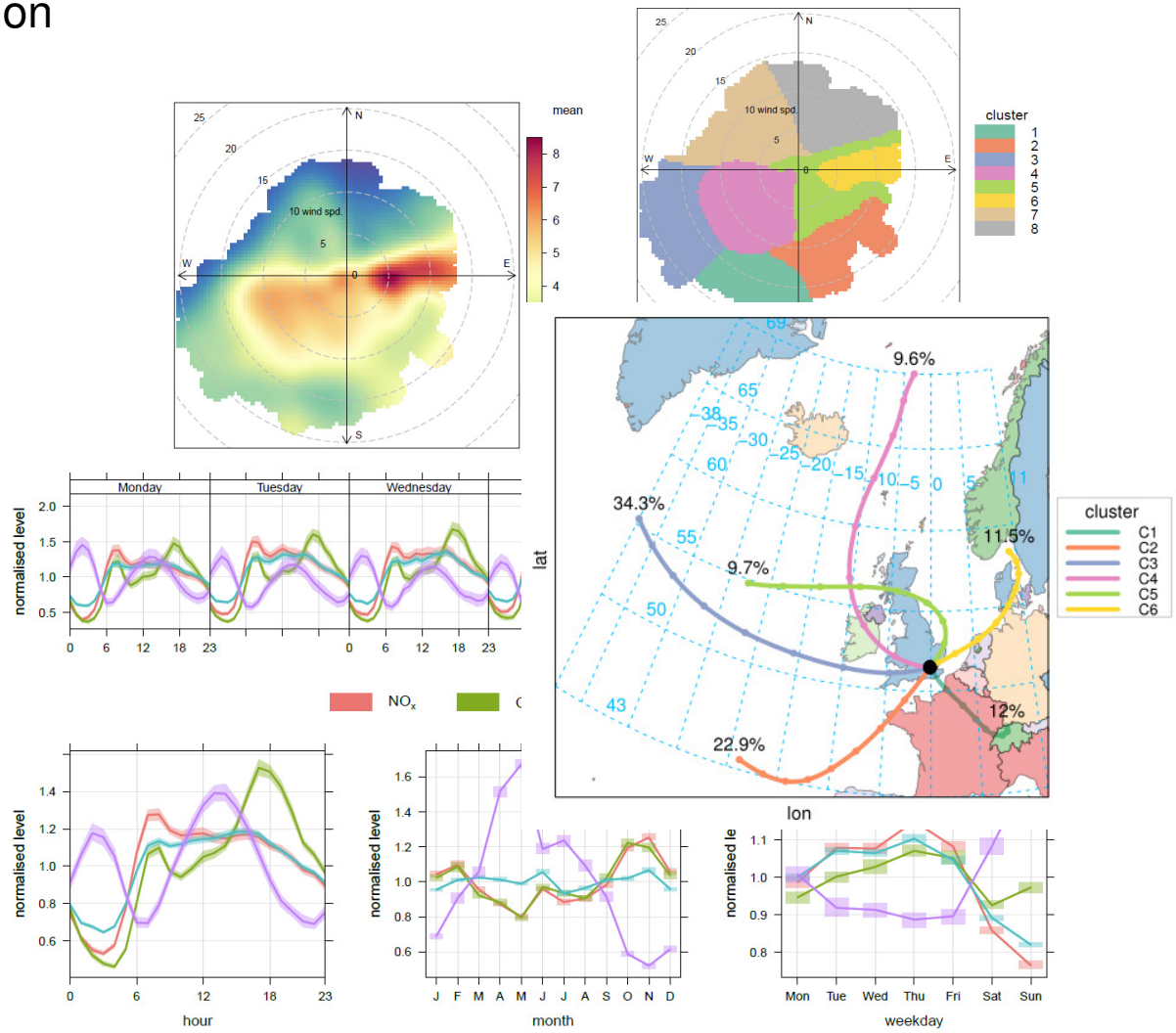
- Excellent for access to many data formats
  - csv, txt, Excel, binary files, databases (e.g. SQL Server, MySQL, Postgres…), XML, JSON, web scraping, NetCDF, …
- Did I mention > 8000 packages?
  - Almost endless possibilities
  - Excellent code sharing on *GitHub* (think Facebook for computer code)
- Growing capability for reproducible reporting / research and interactive web-based 'rich content' documents

Overview of openair

Many capabilities

- Source detection and characterisation
- Robust trend analysis
- Local and regional cluster analysis
- Back trajectory analysis
- Model evaluation
- Training courses
- Widely used*

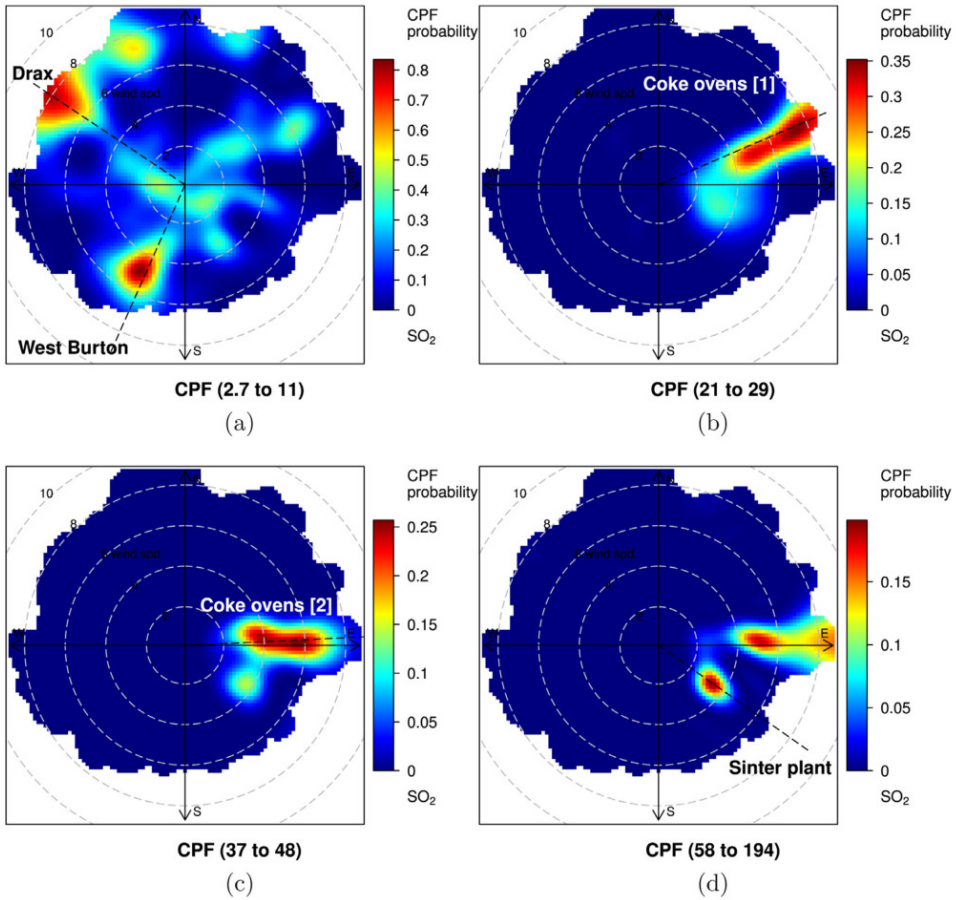




*Downloaded >40,000 times via RStudio, ~3,000 times a month, top 5% of all R packages

Data analysis needs – getting more out of data

- Lots of tools for source characterisation and identification
- Recently more focus on interactive tools, mapping, back trajectory analysis
- Persistent problem though: ***can be difficult to know what is going on because of the strong influence due to meteorology***
- This would be much easier if we had the same weather every day!
- Instead we say that certain years were ‘good’ or ‘bad’…

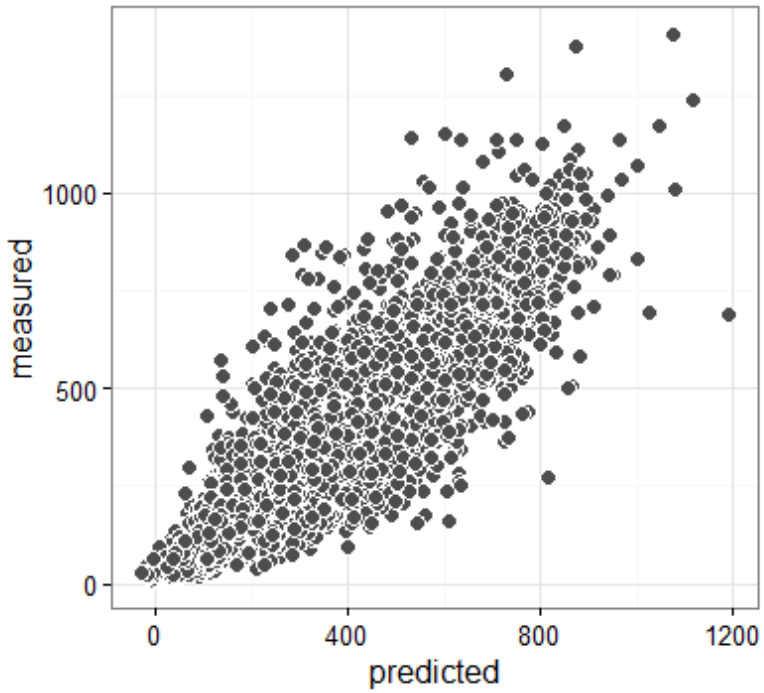


Removing meteorological variation in air pollution trends

- At a simple level, we can build a linear model to describe concentrations in terms of meteorological (and other) variables
- However, relationships between variables are rarely linear and interact with one another – the underlying variation is highly complex for air pollution
- Use 'modern' statistical technique called *boosted regression trees*
  - Very powerful, flexible modelling framework
  - Can account for non-linear relationships, variable interactions, categorical and continuous variables, missing data, …
  - Not black boxes – models can be understood in terms of variable importance and relationships between variables
- Once we have a model developed we can randomly sample the meteorological input 100s or 1000s times, run the model and average the results

Model building

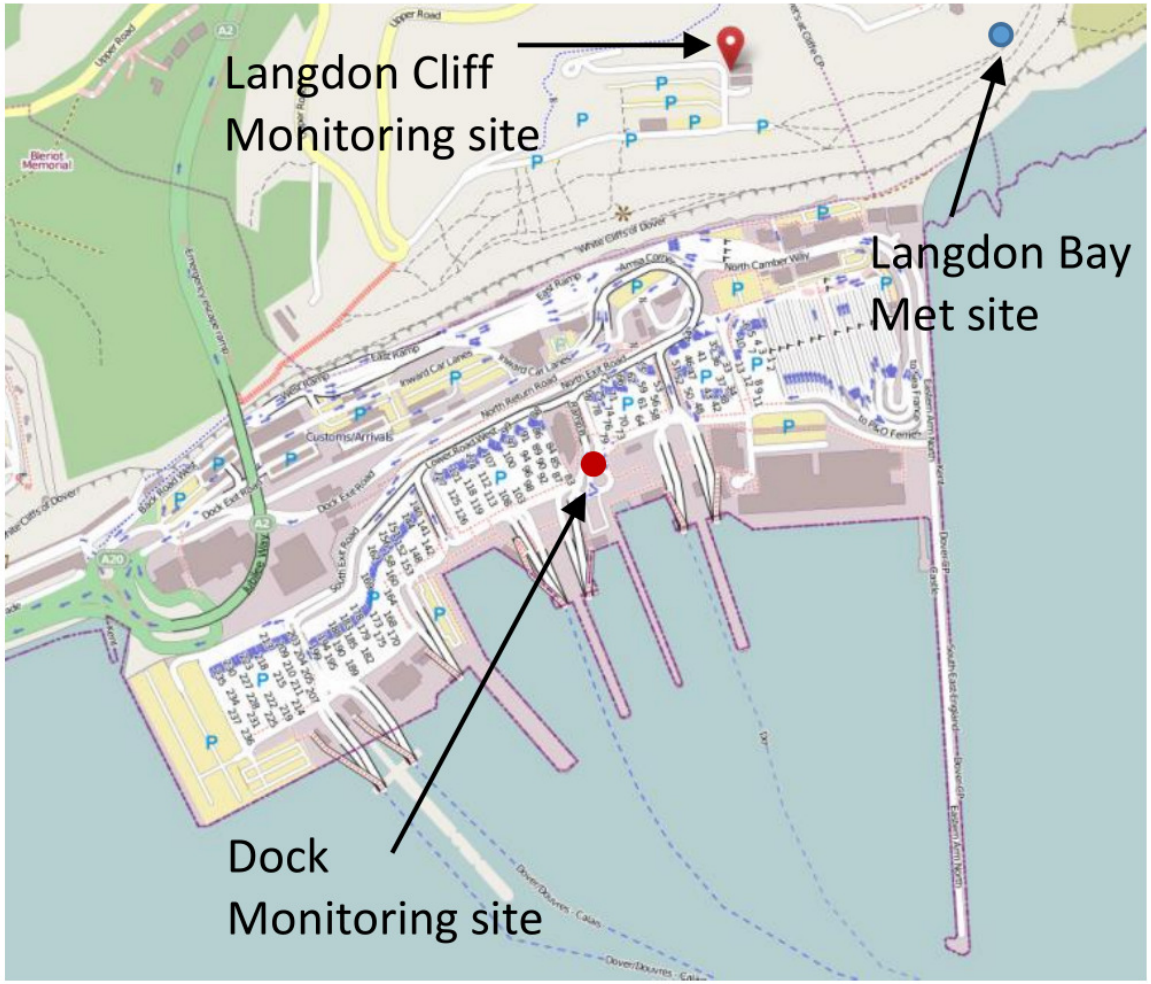
- Build models using random 80% of the data – withhold 20% to evaluate the model
  - Model evaluated on data not used for model development
- Can generally build very good models for *hourly data* – especially when compared with hourly dispersion modelling



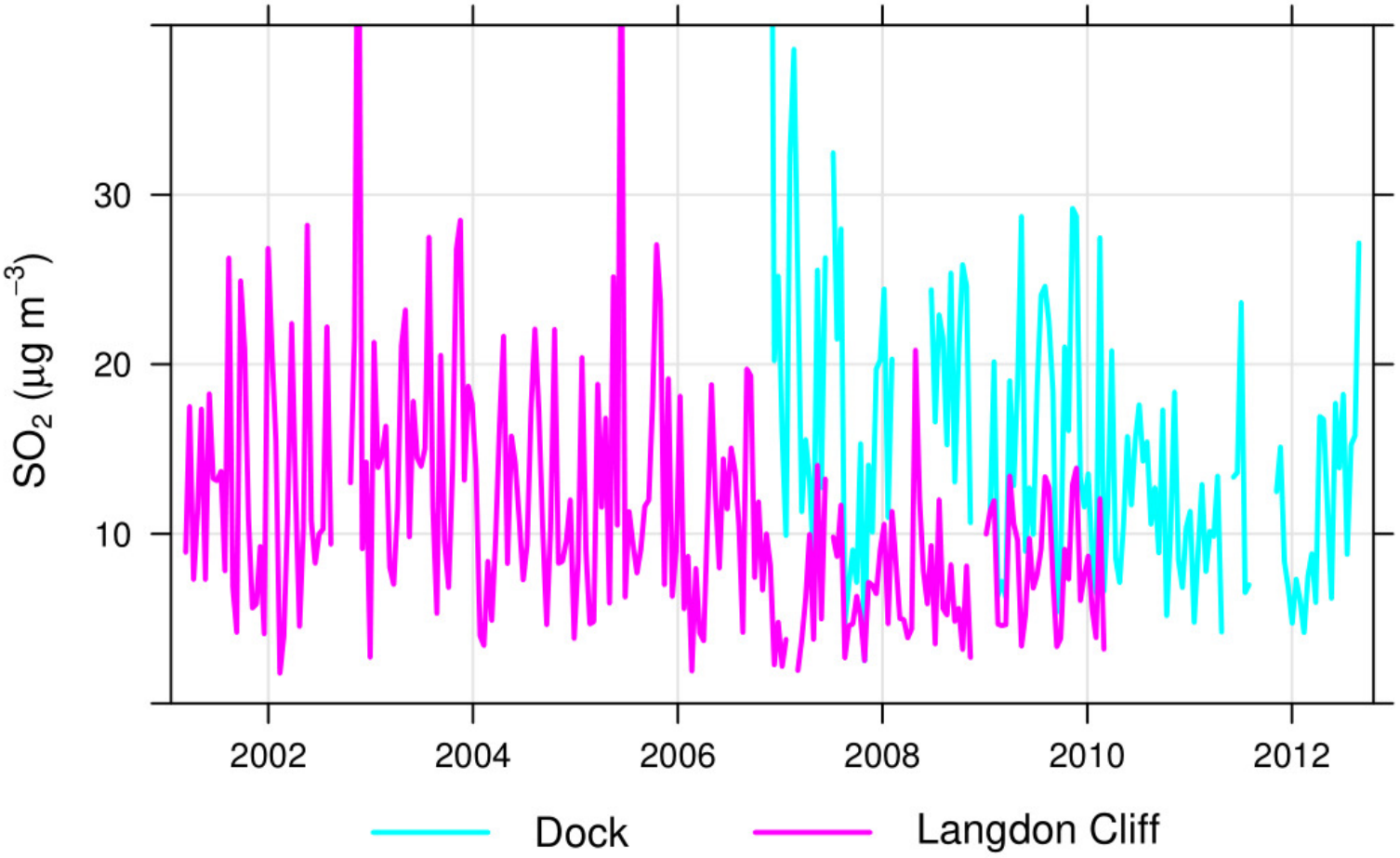
statistic	value
n	8522.00
FAC2	0.95
MB	0.21
MGE	57.63
NMGE	0.19
RMSE	85.49
r	0.92
COE	0.67
IOA	0.84

Example 1: SO_2 concentrations at Dover

- Interest related to fuel sulphur used in shipping
- SO_2 measured at two locations
- Concern over exceedances of 15-minute SO_2 EU Limit Value



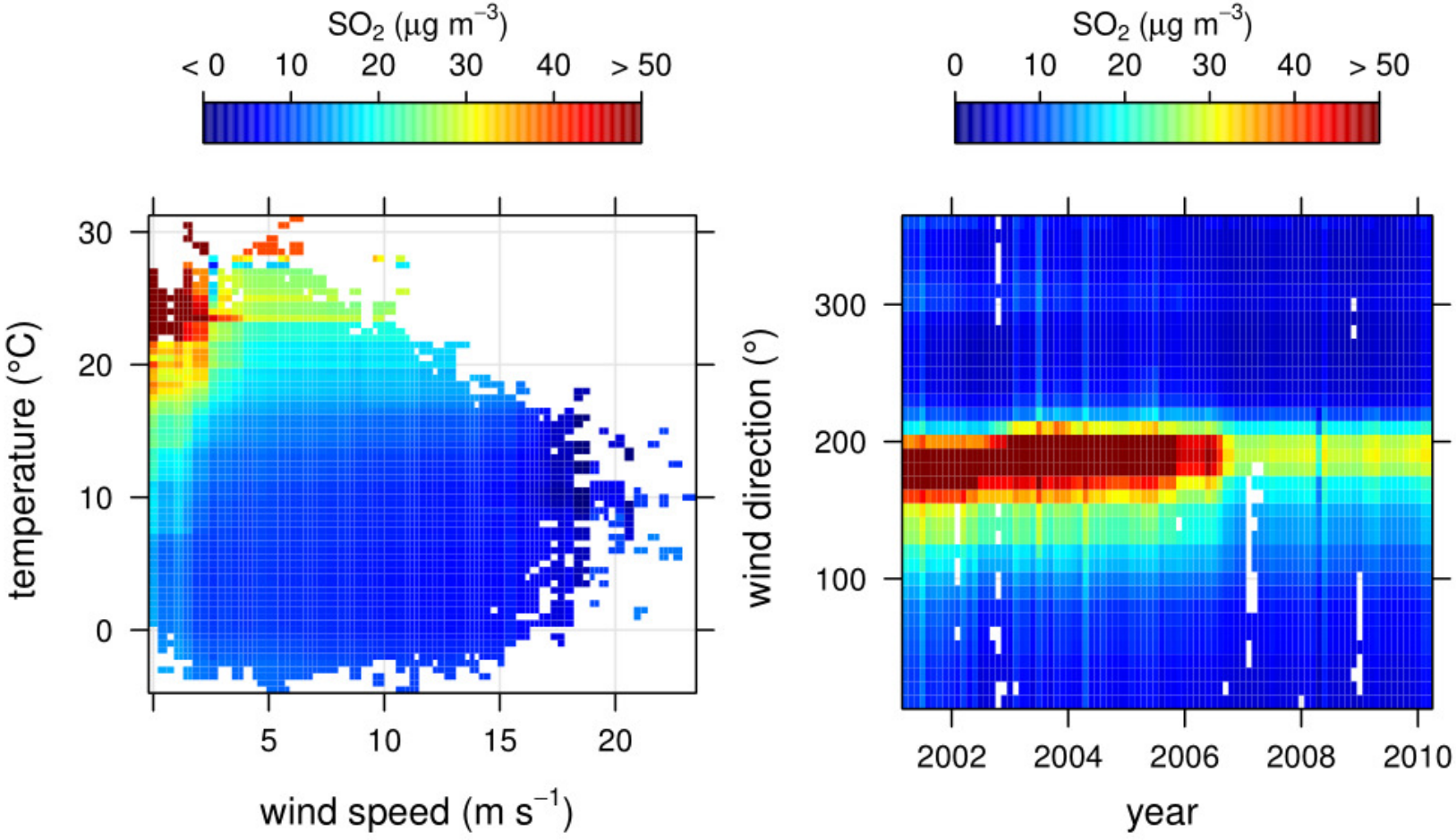
SO_2 concentrations at Dover



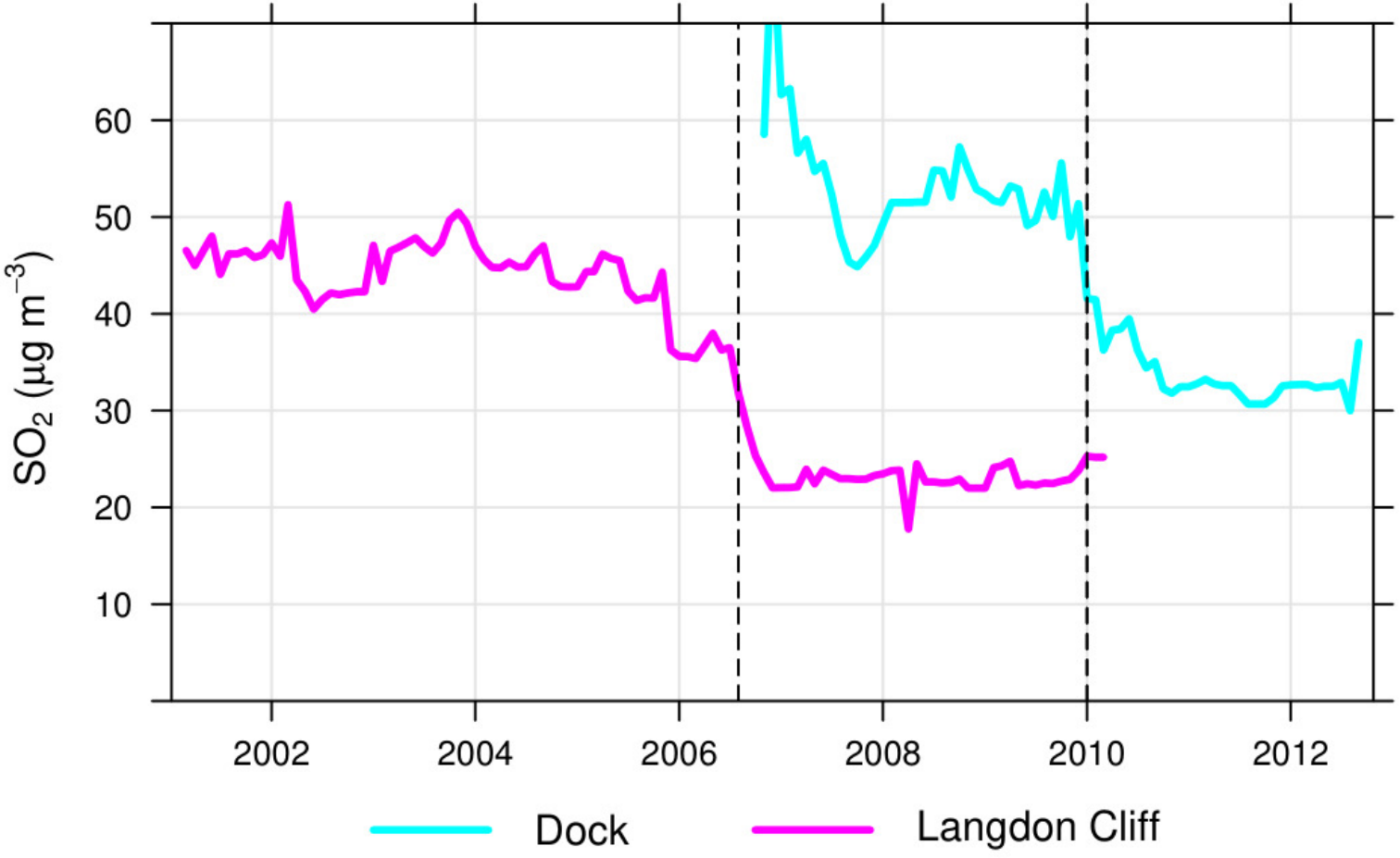
- Hard to see what changed and when – even with 2-week averages
- Not helped by having two sites over two different time periods

SO_2 concentrations at Dover

- Can explore important two-way interactions using the model
  - High ambient temperature, low wind speed conditions important i.e. unstable atmospheric conditions
  - Trend strongly influenced by wind direction



SO_2 concentrations at Dover

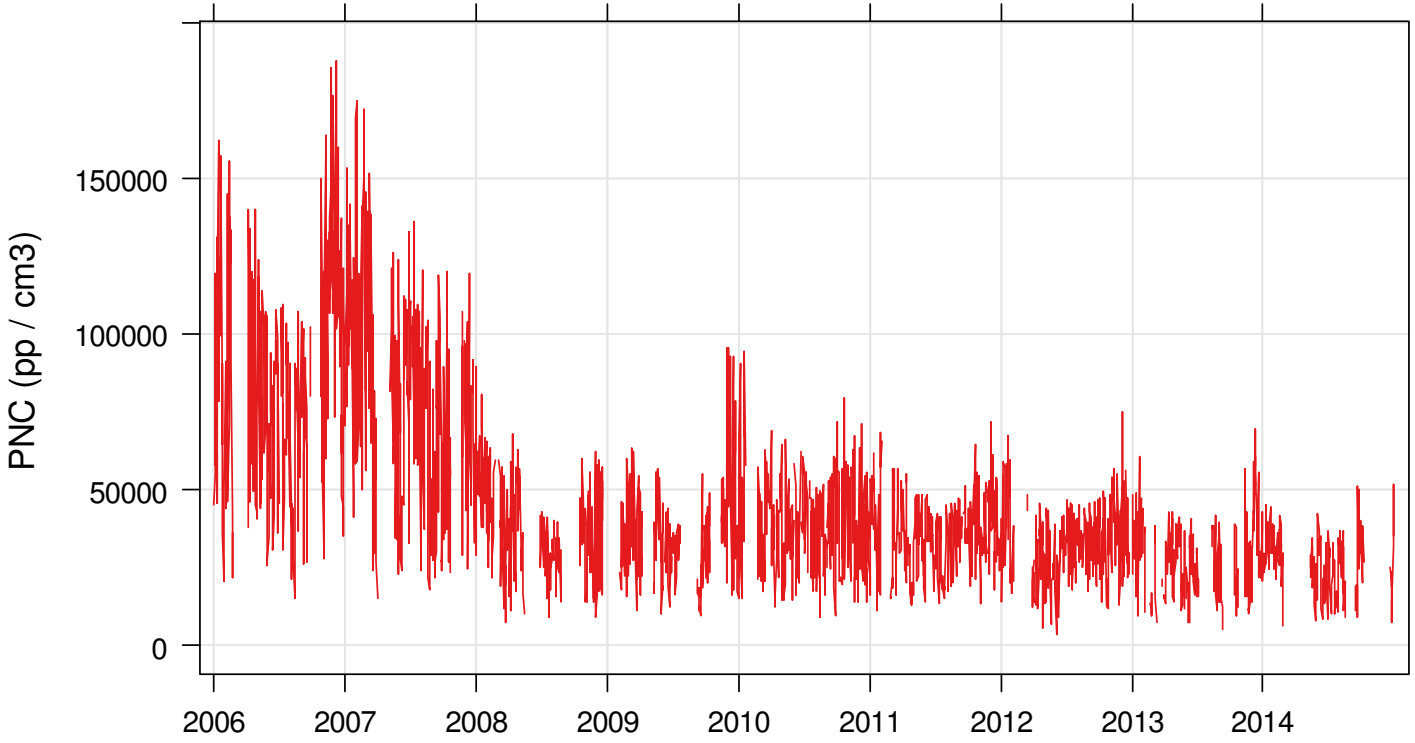


- Fuel sulphur emission limits changed from 2.7 to 1.5% in **August 2006**, then 1.5 to 1.0% in **January 2010** (MARPOL)
- Clear indication in the 'de-weathered' trend data that these changes did actually occur and were of the same magnitude as suggested in the emissions legislation

Example 2: Particle number (PN) counts at Marylebone Road

- Measured since the early 2000s using CPC (Condensation Particle Counter)
  - PN thought to be important from a health perspective – although no ambient air quality limit (there is a limit for diesel, GDI vehicle emissions)
  - Lots of factors that could affect particle numbers in the atmosphere including
    - after-treatment technologies such as diesel particulate filters
    - fuel sulphur content
    - London Low Emission Zone (February and July 2008)
  - Small particles also have highly complex behaviour in the atmosphere – they coagulate, nucleate, evaporate
- What is the evidence of changes for Marylebone Road?

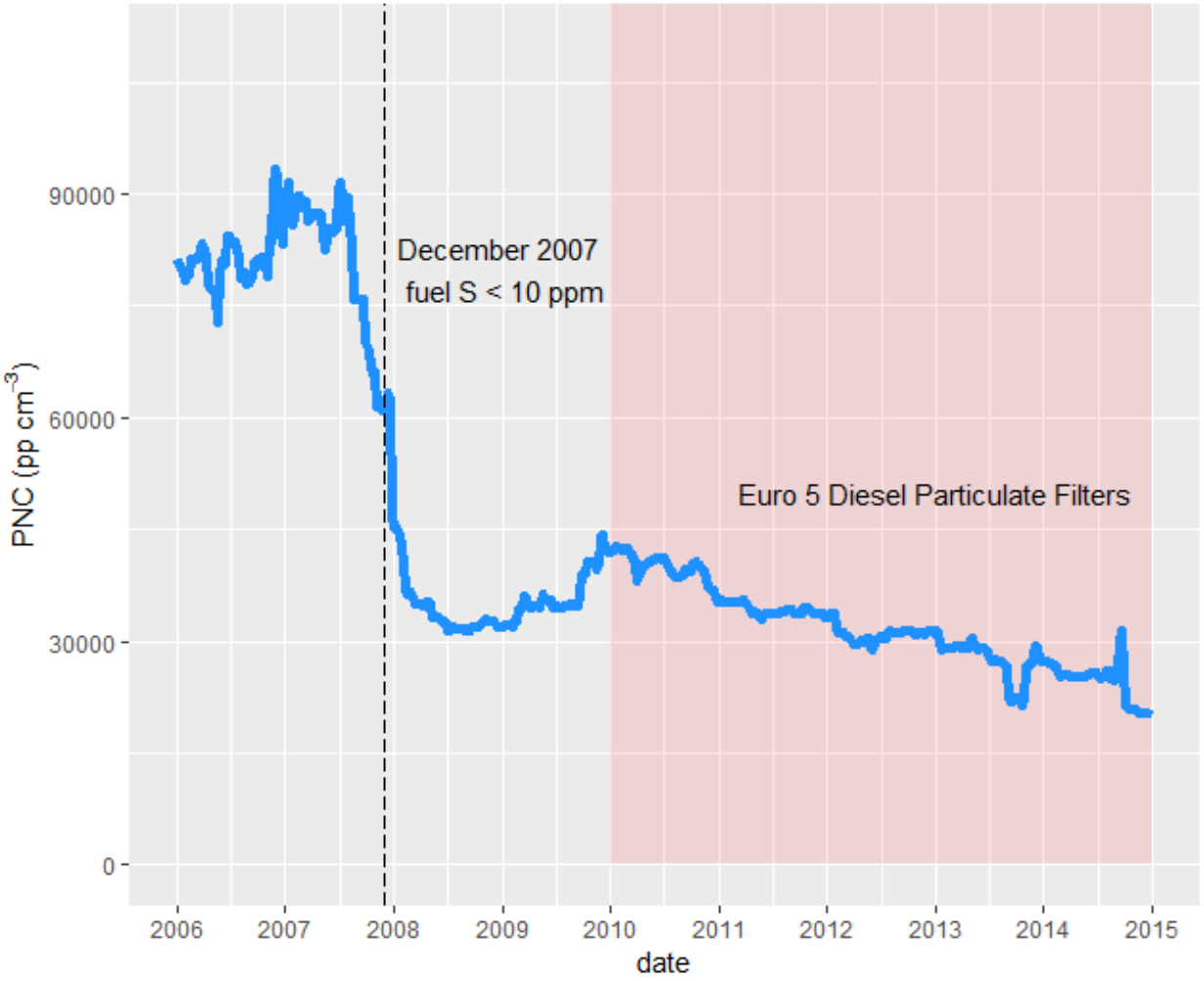
Particle number counts at Marylebone Road



- Plot shows daily mean values – quite a lot of missing data
- It is clear there have been decreases in concentration but not very clear *when* these occurred
- Build statistical model and remove meteorological variation

Particle number counts at Marylebone Road

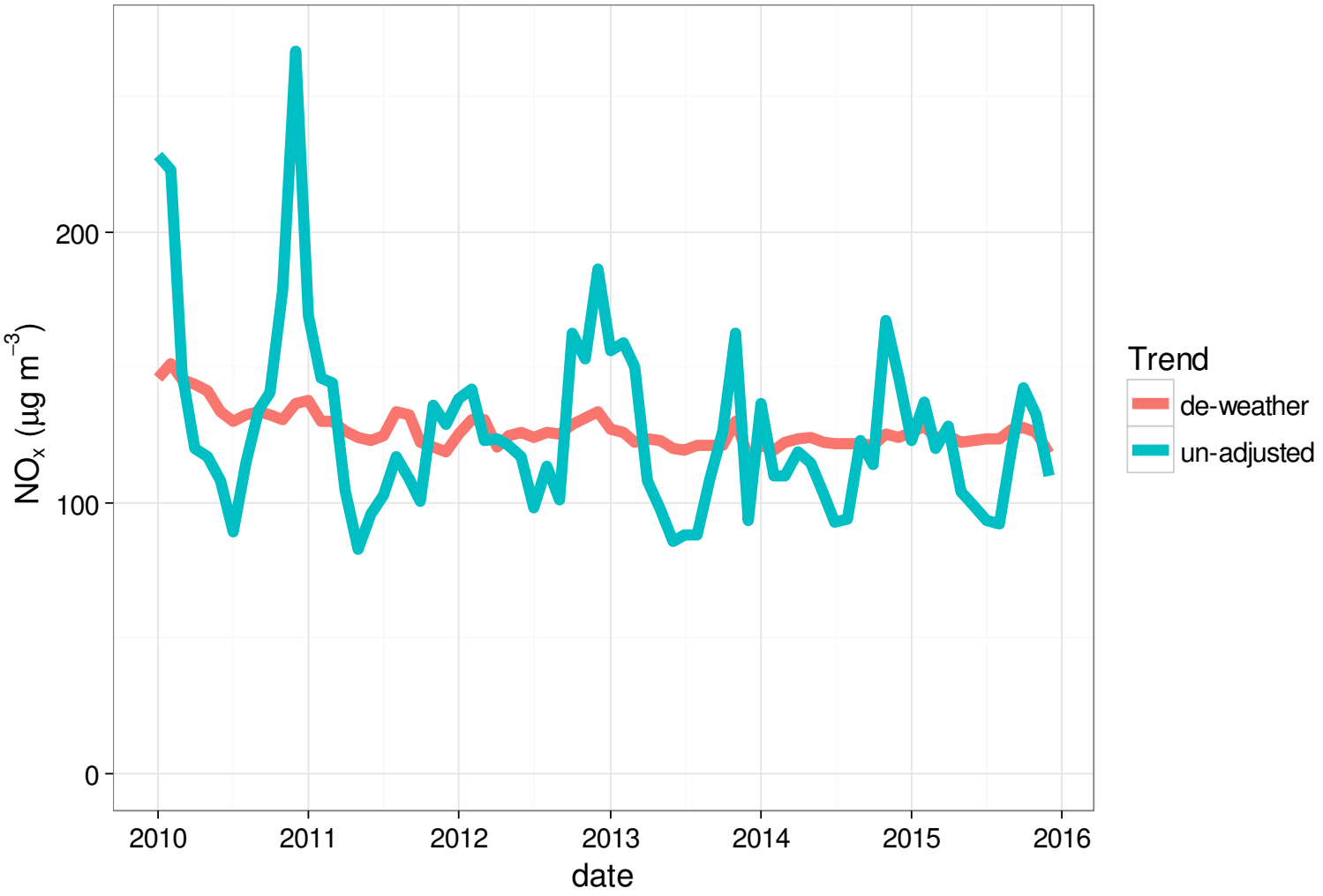
- Very clear that PN decreased sharply mid-2007
- UK Petroleum Industry Associated says 10 ppm sulphur in fuel met by **4 December 2007**
- Strongly suggests that reducing fuel sulphur has had the greatest effect on PN



Example 3: trends in NO_x and NO_2

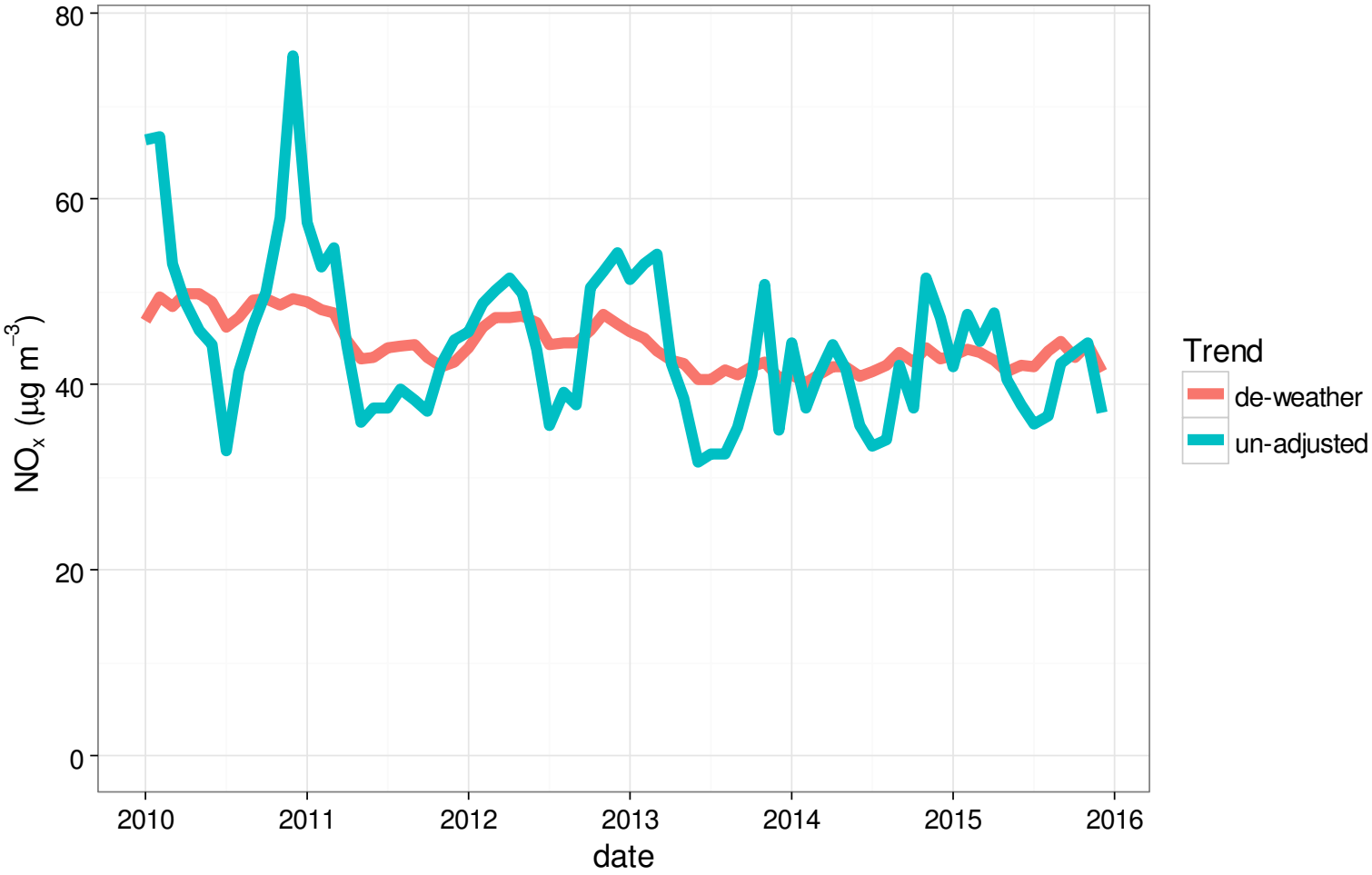
- Important issue currently related to diesel vehicles, VW scandal…
- More important than ever to understand how atmospheric concentrations vary
  - Ambient measurements are the 'ground truth' and provide the evidence we need regardless of what we think emissions are doing…
- Example of four Scottish Roadside sites (in Aberdeen, Dumfries, Inverness and Glasgow)
- What have the trends in NO_x and NO_2 looked like since 2010?

Trends in NO_x concentration



- Monthly means averaged across 4 sites
- The ‘de-weathered’ version shows only a small change in concentration over past 6 years

Trends in NO_2 concentration



- Monthly means averaged across 4 sites
- The ‘de-weathered’ version only shows a small change in concentration over past 6 years

Concluding remarks

- The weather exerts a strong and complex effect on ambient concentrations of pollutants
- These effects can falsely mask or emphasise trends
- Much of the meteorological variation can be ‘removed’ through modern statistical methods
- Removing this variation:
  - Significantly increases the chances of quantifying changes due to emissions changes, interventions etc.
  - Provides a much clearer understanding of trends in air pollutants

Thank you for your attention!

David Carslaw

david.carslaw@ricardo.com
david.carslaw@york.ac.uk