

# The Open-Source Air Pollution Project

Community tools for analysing air pollution data

David Carslaw

Institute for Transport Studies, University of Leeds

14 May 2008

# Summary

- Background to the project
- Progress and examples
- Future directions and developments

# The benefits of openness

- Transparency should be at the heart of environmental regulation
  - Those affected by environmental decisions should be able to scrutinise the tools that lead to those decisions

# The benefits of openness

- Transparency should be at the heart of environmental regulation
  - Those affected by environmental decisions should be able to scrutinise the tools that lead to those decisions
  - Many examples from the USA

# The benefits of openness

- Transparency should be at the heart of environmental regulation
  - Those affected by environmental decisions should be able to scrutinise the tools that lead to those decisions
  - Many examples from the USA
- Open-source software

# The benefits of openness

- Transparency should be at the heart of environmental regulation
  - Those affected by environmental decisions should be able to scrutinise the tools that lead to those decisions
  - Many examples from the USA
- Open-source software
  - All source code made available, free and can be modified by anyone

# The benefits of openness

- Transparency should be at the heart of environmental regulation
  - Those affected by environmental decisions should be able to scrutinise the tools that lead to those decisions
  - Many examples from the USA
- Open-source software
  - All source code made available, free and can be modified by anyone
  - No longer in the realm of enthusiasts e.g. Linux, MySQL, R

# The benefits of openness

- Transparency should be at the heart of environmental regulation
  - Those affected by environmental decisions should be able to scrutinise the tools that lead to those decisions
  - Many examples from the USA
- Open-source software
  - All source code made available, free and can be modified by anyone
  - No longer in the realm of enthusiasts e.g. Linux, MySQL, R
- Both promote participation and ownership



# The benefits of openness

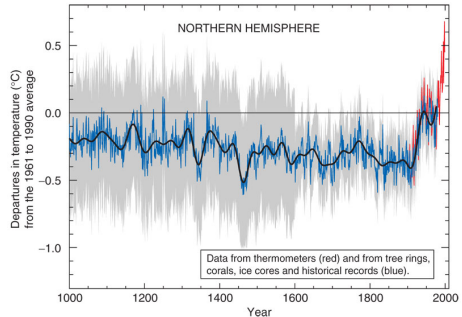
- Transparency should be at the heart of environmental regulation
  - Those affected by environmental decisions should be able to scrutinise the tools that lead to those decisions
  - Many examples from the USA
- Open-source software
  - All source code made available, free and can be modified by anyone
  - No longer in the realm of enthusiasts e.g. Linux, MySQL, **R**
- Both promote participation and ownership

# Interesting aside – climate change and the “hockey stick”

- Controversy over IPCC “hockey stick” temperature graph<sup>a</sup>
- Statistical methods shown to be flawed (McIntyre and McKittrick)
- Code and data made available – some in R

---

<sup>a</sup>Mann, M.E. et al. (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, Vol. 392, pp. 779787.



# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'

# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'
  - To make it easier to analyse data and to gain insights from it

# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'
  - To make it easier to analyse data and to gain insights from it
  - Exploit the enormous (and growing) amount of air pollution data available

# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'
  - To make it easier to analyse data and to gain insights from it
  - Exploit the enormous (and growing) amount of air pollution data available
  - Progressively include advanced approaches generally not widely available

# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'
  - To make it easier to analyse data and to gain insights from it
  - Exploit the enormous (and growing) amount of air pollution data available
  - Progressively include advanced approaches generally not widely available
- Outputs

# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'
  - To make it easier to analyse data and to gain insights from it
  - Exploit the enormous (and growing) amount of air pollution data available
  - Progressively include advanced approaches generally not widely available
- Outputs
  - An R 'package' dedicated to air pollution analysis



# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'
  - To make it easier to analyse data and to gain insights from it
  - Exploit the enormous (and growing) amount of air pollution data available
  - Progressively include advanced approaches generally not widely available
- Outputs
  - An R 'package' dedicated to air pollution analysis
  - A web site to act as a central resource

# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'
  - To make it easier to analyse data and to gain insights from it
  - Exploit the enormous (and growing) amount of air pollution data available
  - Progressively include advanced approaches generally not widely available
- Outputs
  - An R 'package' dedicated to air pollution analysis
  - A web site to act as a central resource
  - Comprehensive documentation

# Project aims

- To build a set of **free**, open-source tools for the analysis of air pollution data
  - Use highly developed open-source statistical software 'R'
  - To make it easier to analyse data and to gain insights from it
  - Exploit the enormous (and growing) amount of air pollution data available
  - Progressively include advanced approaches generally not widely available
- Outputs
  - An R 'package' dedicated to air pollution analysis
  - A web site to act as a central resource
  - Comprehensive documentation

# Progress – case studies and user involvement

- Sefton Council
  - Support as part of their 'Beacon' status

# Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities

## Progress – case studies and user involvement

- Sefton Council
  - Support as part of their 'Beacon' status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA

## Progress – case studies and user involvement

- Sefton Council
  - Support as part of their 'Beacon' status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works

# Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works
  - Highly complex mixture of sources



## Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works
  - Highly complex mixture of sources
  - Many non-road traffic sources

## Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works
  - Highly complex mixture of sources
  - Many non-road traffic sources
- Defra/AEA

# Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works
  - Highly complex mixture of sources
  - Many non-road traffic sources
- Defra/AEA
  - Developments to enhance informatics value of AURN data

# Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works
  - Highly complex mixture of sources
  - Many non-road traffic sources
- Defra/AEA
  - Developments to enhance informatics value of AURN data
- NERC Knowledge Transfer bid

## Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works
  - Highly complex mixture of sources
  - Many non-road traffic sources
- Defra/AEA
  - Developments to enhance informatics value of AURN data
- NERC Knowledge Transfer bid
  - Would provide significant funding

# Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works
  - Highly complex mixture of sources
  - Many non-road traffic sources
- Defra/AEA
  - Developments to enhance informatics value of AURN data
- NERC Knowledge Transfer bid
  - Would provide significant funding
  - Widely supported

# Progress – case studies and user involvement

- Sefton Council
  - Support as part of their ‘Beacon’ status
  - Help inform continuing air quality management activities
- North Lincolnshire/AEA
  - Scunthorpe steel works
  - Highly complex mixture of sources
  - Many non-road traffic sources
- Defra/AEA
  - Developments to enhance informatics value of AURN data
- NERC Knowledge Transfer bid
  - Would provide significant funding
  - Widely supported

# What effort is required to use these tools?

- The project aims to overcome barriers to analysis
  - Lack of time, money, specialist software, know-how or all four!



# What effort is required to use these tools?

- The project aims to overcome barriers to analysis
  - Lack of time, money, specialist software, know-how or all four!
- Not required to learn R

# What effort is required to use these tools?

- The project aims to overcome barriers to analysis
  - Lack of time, money, specialist software, know-how or all four!
- Not required to learn R
  - Learning R can be hard work

# What effort is required to use these tools?

- The project aims to overcome barriers to analysis
  - Lack of time, money, specialist software, know-how or all four!
- Not required to learn R
  - Learning R can be hard work
  - Make things as simple as possible for the user

# What effort is required to use these tools?

- The project aims to overcome barriers to analysis
  - Lack of time, money, specialist software, know-how or all four!
- Not required to learn R
  - Learning R can be hard work
  - Make things as simple as possible for the user

Example of code required to make a polar plot

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

Some examples of current capabilities to follow ...

## What effort is required to use these tools?

- The project aims to overcome barriers to analysis
  - Lack of time, money, specialist software, know-how or all four!
- Not required to learn R
  - Learning R can be hard work
  - Make things as simple as possible for the user

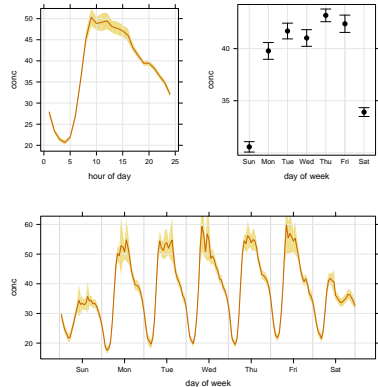
### Example of code required to make a polar plot

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

Some examples of current capabilities to follow . . .

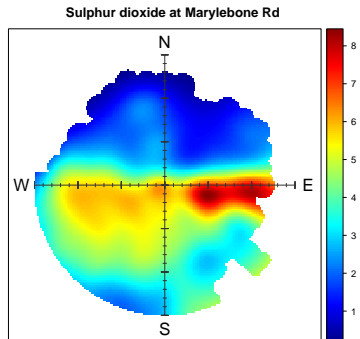
# Diurnal variation in concentrations

- Diurnal and day of week variations can provide clues as to the source
- Function `diurnal.error` produces three plots
- Uncertainty bands can help determine whether one source is different from another

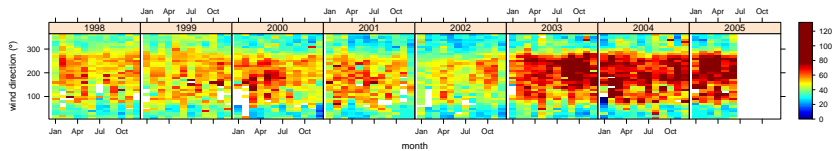


# Bivariate polar plots

- Useful for source detection
- Methods have been extended to 'model' surface concentrations
- Can usefully be combined with other methods

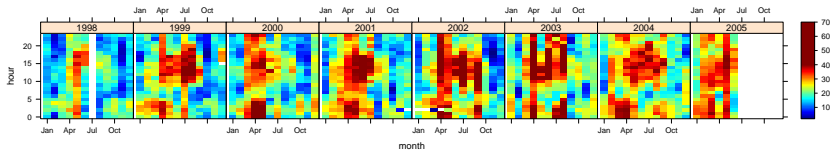


# Concentrations by wind direction, year and month

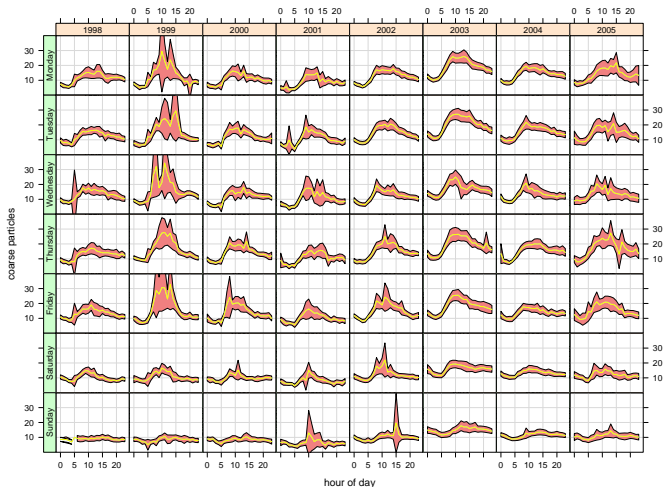




# Maximum hourly ozone concentrations by hour of day, year and month



# Diurnal and day of week variation – $PM_{coarse}$

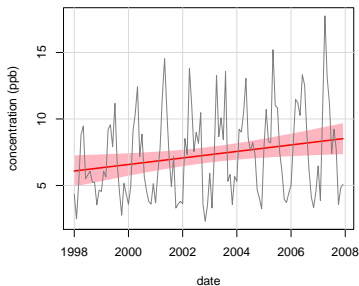


## Flexible treatment of trends – ozone at Marylebone Road

- Trend with smooth fit and 95 % confidence intervals

## Flexible treatment of trends – ozone at Marylebone Road

- Trend with smooth fit and 95 % confidence intervals
- Trend with smooth fit and 95 % confidence intervals – **deseasonalise** first

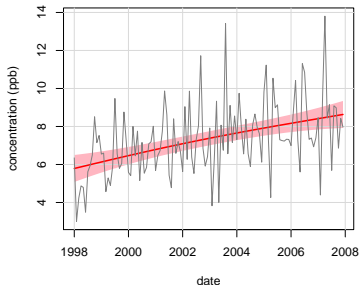


## Flexible treatment of trends – ozone at Marylebone Road

- Trend with smooth fit and 95 % confidence intervals
- Trend with smooth fit and 95 % confidence intervals – **deseasonalise** first

## Flexible treatment of trends – ozone at Marylebone Road

- Trend with smooth fit and 95 % confidence intervals
- Trend with smooth fit and 95 % confidence intervals – **deseasonalise** first
- Consider “no trend” hypothesis through bootstrap resampling

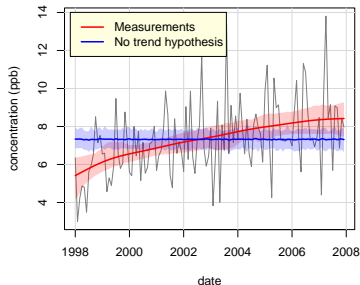


## Flexible treatment of trends – ozone at Marylebone Road

- Trend with smooth fit and 95 % confidence intervals
- Trend with smooth fit and 95 % confidence intervals – **deseasonalise** first
- Consider “no trend” hypothesis through bootstrap resampling

## Flexible treatment of trends – ozone at Marylebone Road

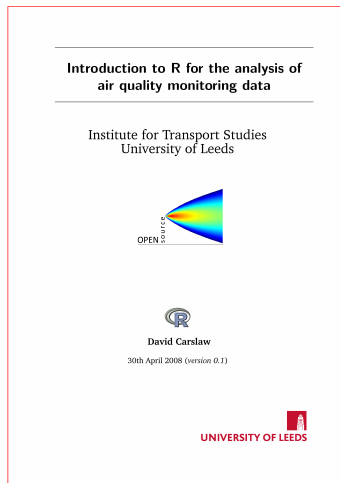
- Trend with smooth fit and 95 % confidence intervals
- Trend with smooth fit and 95 % confidence intervals – **deseasonalise** first
- Consider “no trend” hypothesis through bootstrap resampling





# Documentation

- Documentation has been started
  - Part 1: Introduction to using R to analyse monitoring data
  - Part II: Dedicated functions to analyse monitoring data
- Longer term aims
  - Develop a *Framework* for analysing data
  - Case studies spanning a range of contemporary problems



## Longer-term aims

- The big picture
  - Change the way we do things

## Longer-term aims

- The big picture
  - Change the way we do things
  - Environmental models?

## Longer-term aims

- The big picture
  - Change the way we do things
  - Environmental models?
  - Work that is truly reproducible

# Longer-term aims

- The big picture
  - Change the way we do things
  - Environmental models?
  - Work that is truly reproducible
- Wider issues

# Longer-term aims

- The big picture
  - Change the way we do things
  - Environmental models?
  - Work that is truly reproducible
- Wider issues
  - Developing countries

# Longer-term aims

- The big picture
  - Change the way we do things
  - Environmental models?
  - Work that is truly reproducible
- Wider issues
  - Developing countries
  - Actively seek participation of researchers elsewhere in the world

## Longer-term aims

- The big picture
  - Change the way we do things
  - Environmental models?
  - Work that is truly reproducible
- Wider issues
  - Developing countries
  - Actively seek participation of researchers elsewhere in the world



# Longer-term aims

- The big picture
  - Change the way we do things
  - Environmental models?
  - Work that is truly reproducible
- Wider issues
  - Developing countries
  - Actively seek participation of researchers elsewhere in the world

Thank you for your attention!

