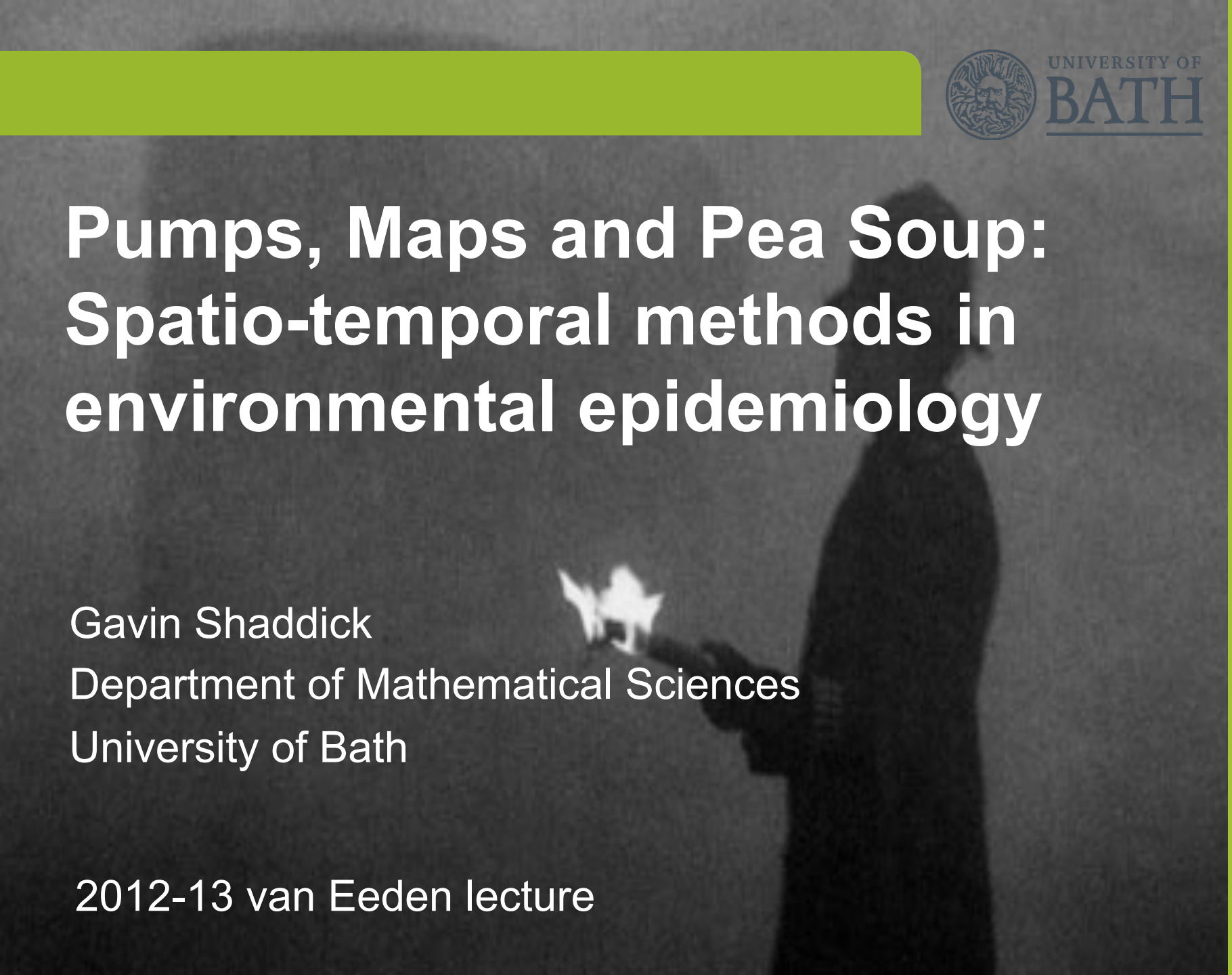
Spatiotemporal models in environmental epidemiology

Lecture 1: Introduction: Based on a guest lecture given by Professor Gavin Shaddick

Reference: Shaddick, G., Zidek, J. V., & Schmidt, A. M. (2023). Spatio–Temporal Methods in Environmental Epidemiology with R. CRC Press.



Pumps, Maps and Pea Soup: Spatio-temporal methods in environmental epidemiology

Gavin Shaddick
Department of Mathematical Sciences
University of Bath

2012-13 van Eeden lecture

Thanks

- Constance van Eeden Fund
- Department of Statistics, University of British Columbia
- Prof. Jim Zidek

- This lecture inaugurates a one term special topics graduate course in statistics, in the Department of Statistics (Stat547L)

Outline

- Introduction
- Spatial-temporal epidemiology
- Spatial misalignment
- Example: spatio-temporal modelling of air pollution
- Preferential sampling of exposures
- Course overview
- Current research topics

What is epidemiology?

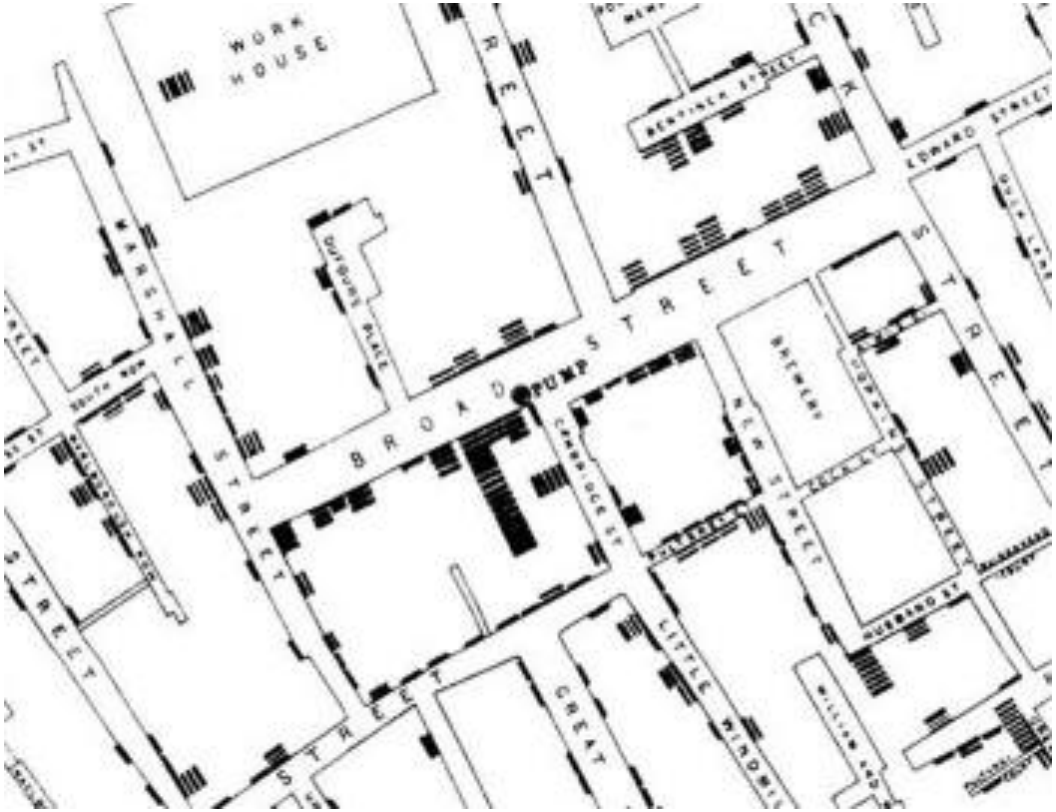
- “The study of skin diseases?”

- “The study of the distribution and determinants of health-related states in specified populations, and the application of this study to control health problems."

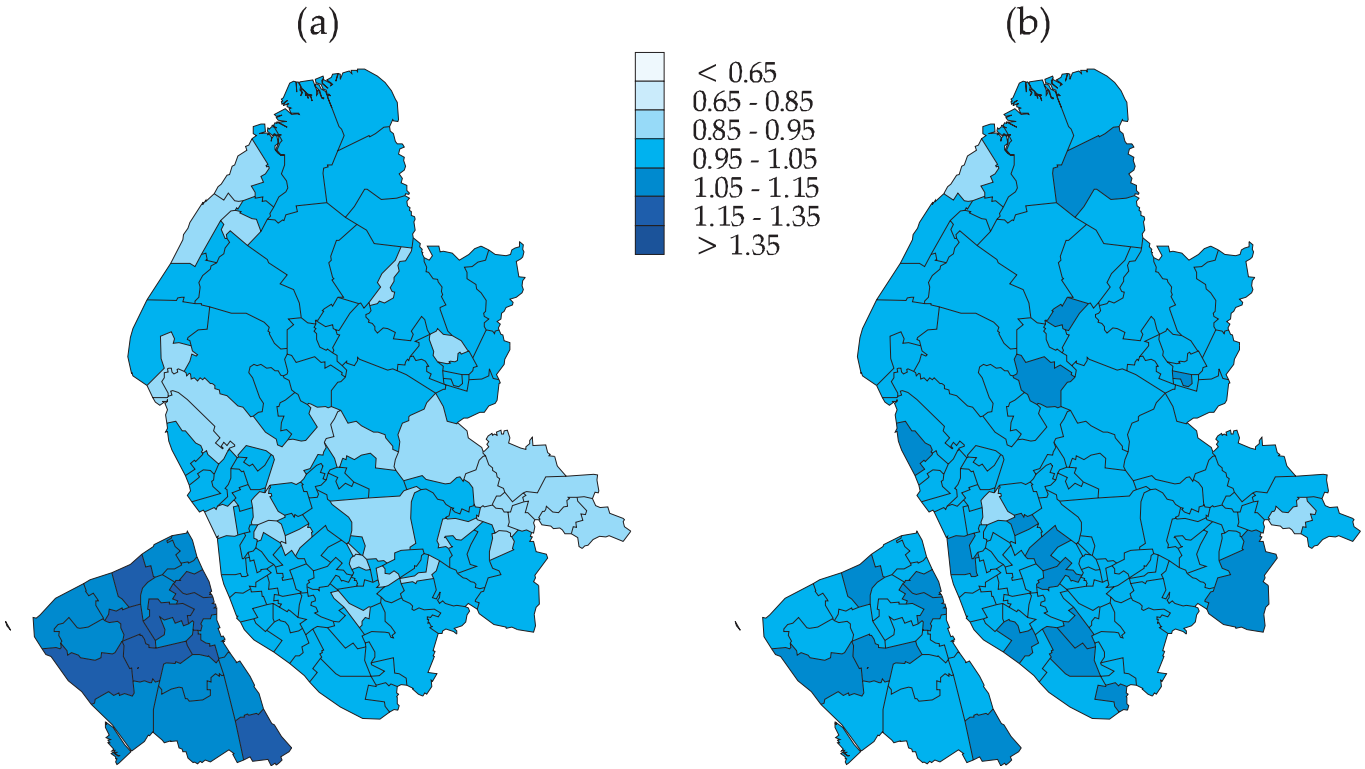
The early days…
John Snow and the Broad Street pub

The early days…
John Snow and the Broad Street pump

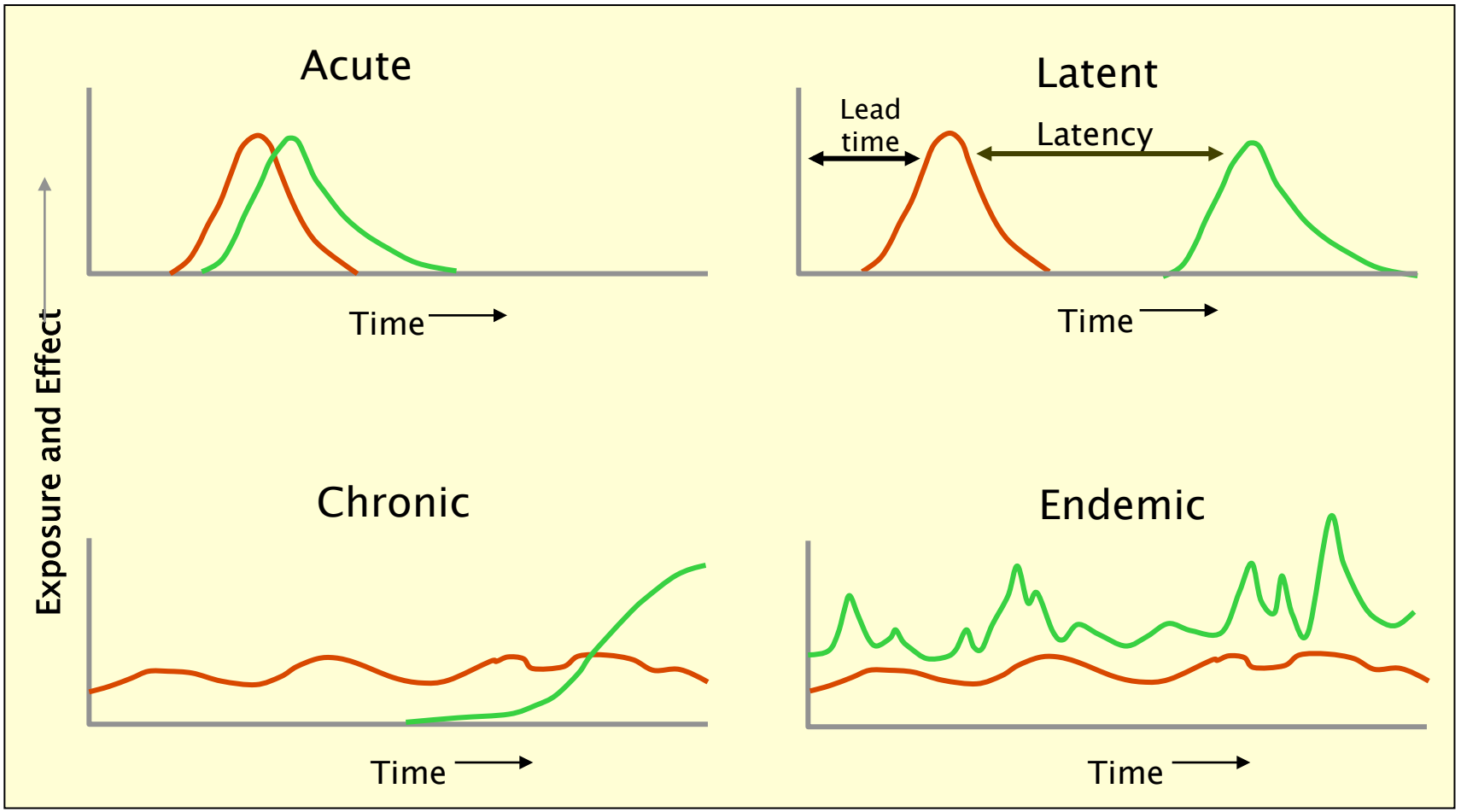
Number of cholera cases in proximity to water pump, Soho, London 1854



SIRs for (a) lung and (b) brain cancer in North-West England, 1991-96



Temporal relationships between exposure and effect



Environmental space-time field: smog in 1950s London

Great smog of 1952 – a four day ‘pea souper’

- Early winter with snow in November
- Extra burning of coal
- Started 5th December
- Area of high pressure trapping the smog
- Light winds
- 4000 excess deaths in next two weeks compared with previous two weeks

Ensuing developments

- 1956 UK clean air act
- 1960s UK National survey monitoring network
- 1970 US clean air act
  - to protect human health (mortality / morbidity)
    - without regard to cost
  - to protect human welfare (crops, forests)
- 1971 EPA formed
- Present day guidelines at both national and international level

Spatio-temporal epidemiology

- Disease risk depends on the classic epidemiological triad of person (genetics/behaviour), place and time

- Place is a surrogate for exposures present at that location
  - environmental exposures in water/air/soil, or the lifestyle characteristics of those living in particular areas.

- Time is a surrogate for exposures present at that moment in time
  - environmental exposures in air, or the lifestyle characteristics that might influence exposures over time

Need for spatio-temporal methods

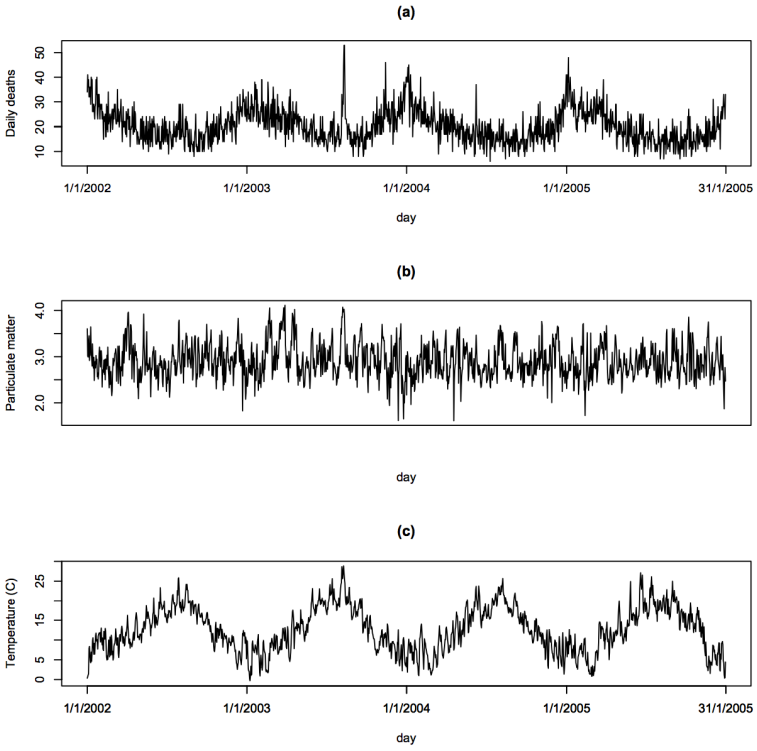
- Epidemiological studies are very often both spatial and temporal
- When do we need to ‘worry’, i.e. acknowledge the spatial and temporal components?
  - are we explicitly interested in the spatio-temporal pattern of disease incidence?
    - e.g. disease mapping, cluster detection
  - is the clustering a nuisance quantity that we wish to acknowledge, but are not explicitly interested in?
    - e.g. spatio-temporal regression

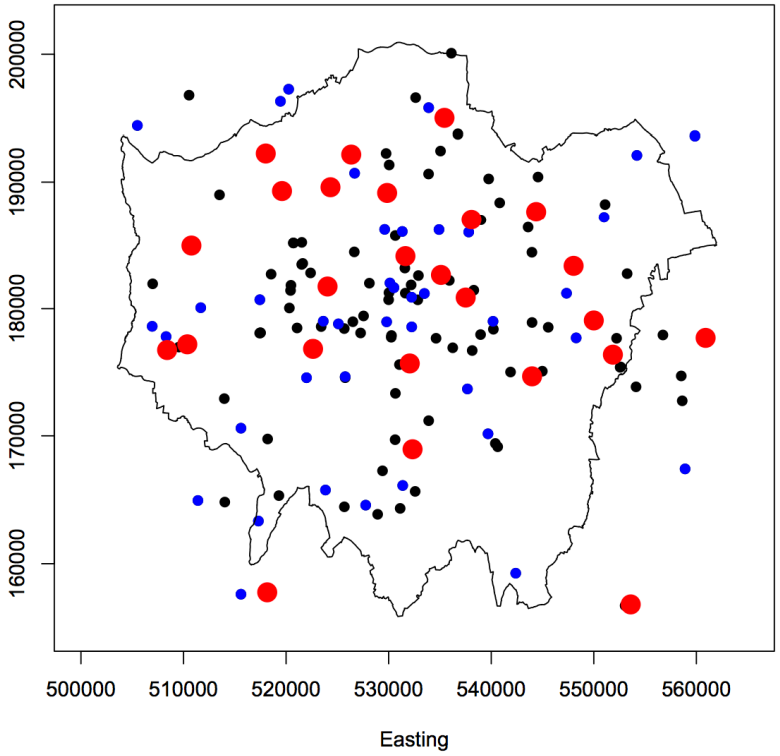
Growing interest in spatio-temporal epidemiology due to:

- Public interest in effects of environmental ‘pollution’
- Development of statistical/epidemiological methods for investigating disease ‘clusters’
- Epidemiological interest in the existence of large/ medium spread in chronic disease rates over time across different areas
- Data availability: collection of health data over time at different geographical scales
- Modelling exposures over space and time
- Increase in computing power and methods (Geographical Informations Systems)

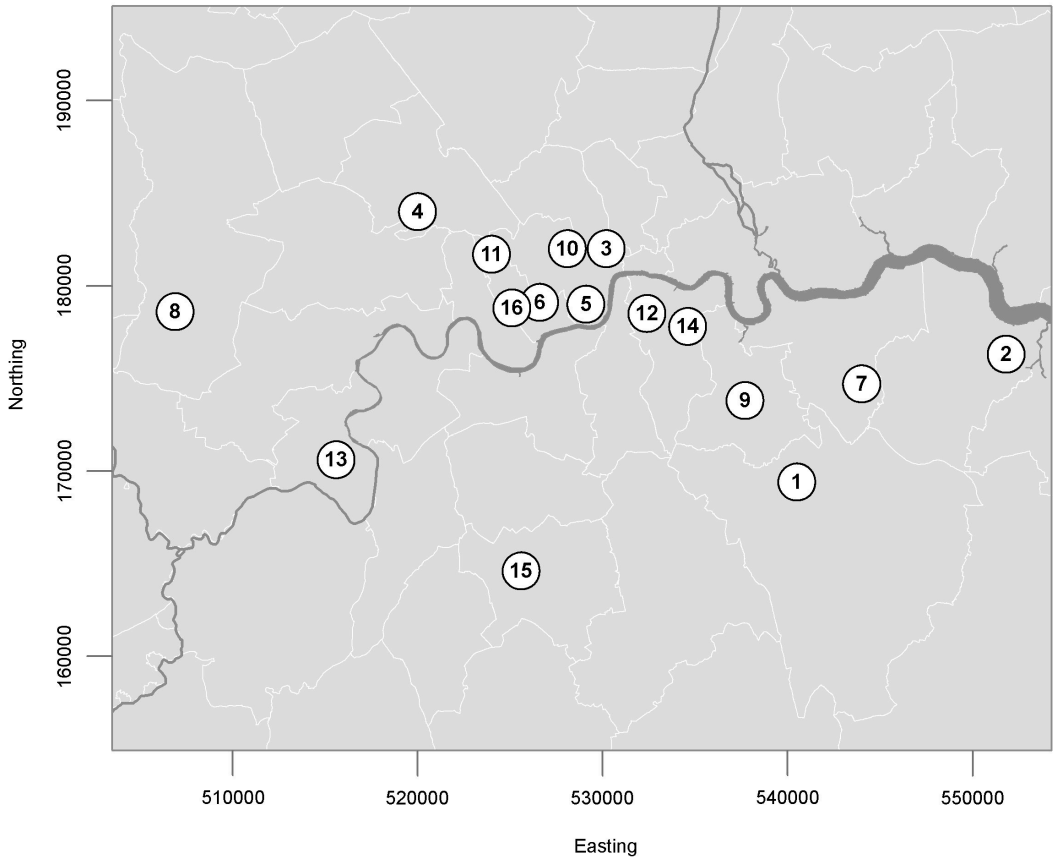
Performing spatio-temporal analyses

Link health outcomes to exposures in time and space





Linking health and exposure data: spatial misalignment



Spatial misalignment

- Case 1: Health data may be available in a number of areas where exposure data is not available.
  - Spatial modelling can be used in order to estimate exposures in unmeasured areas.
  - Are there any issues with this approach?
- Case 2: Health data may relate to the entire study region whereas the pollution data are measured at a number of distinct (point) locations across the study region
  - Within an area, e.g. a city, there may be a number of monitoring sites.
  - What is the best estimate of exposure to use?

Summaries of exposure

- The exposure within an area is often represented by the mean of several measurements
  - e.g. average of concentrations of air pollution from monitors within the area
- Potential for bias will depend on:
  - spatial variation
  - monitor placement
  - measurement error
- Statistical methods should acknowledge exposure variability
  - ecological bias

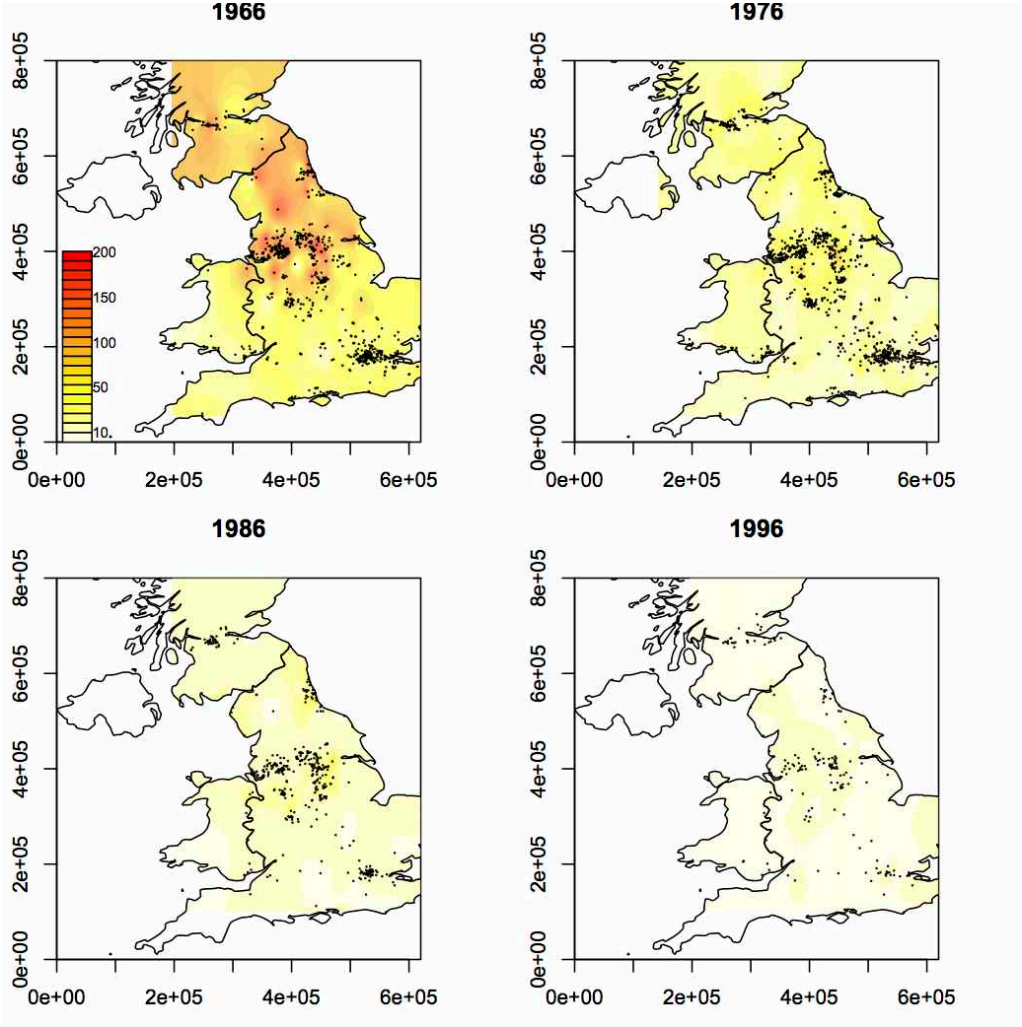
Spatio-temporal modelling of air pollution

- Concentrations of black smoke measured in the UK from 1960s to 1990s
  - Beaver report (1954) and clean air act (1956) stressed importance of fine airborne smoke and sulphur dioxide
  - National survey
    - 1952: 66 towns and 5 London boroughs
    - mid-1960s: 1000+ sites
    - mid-1990s: 200 sites
- Examine changes over time and variations over space
- Effects of reduction in network over time

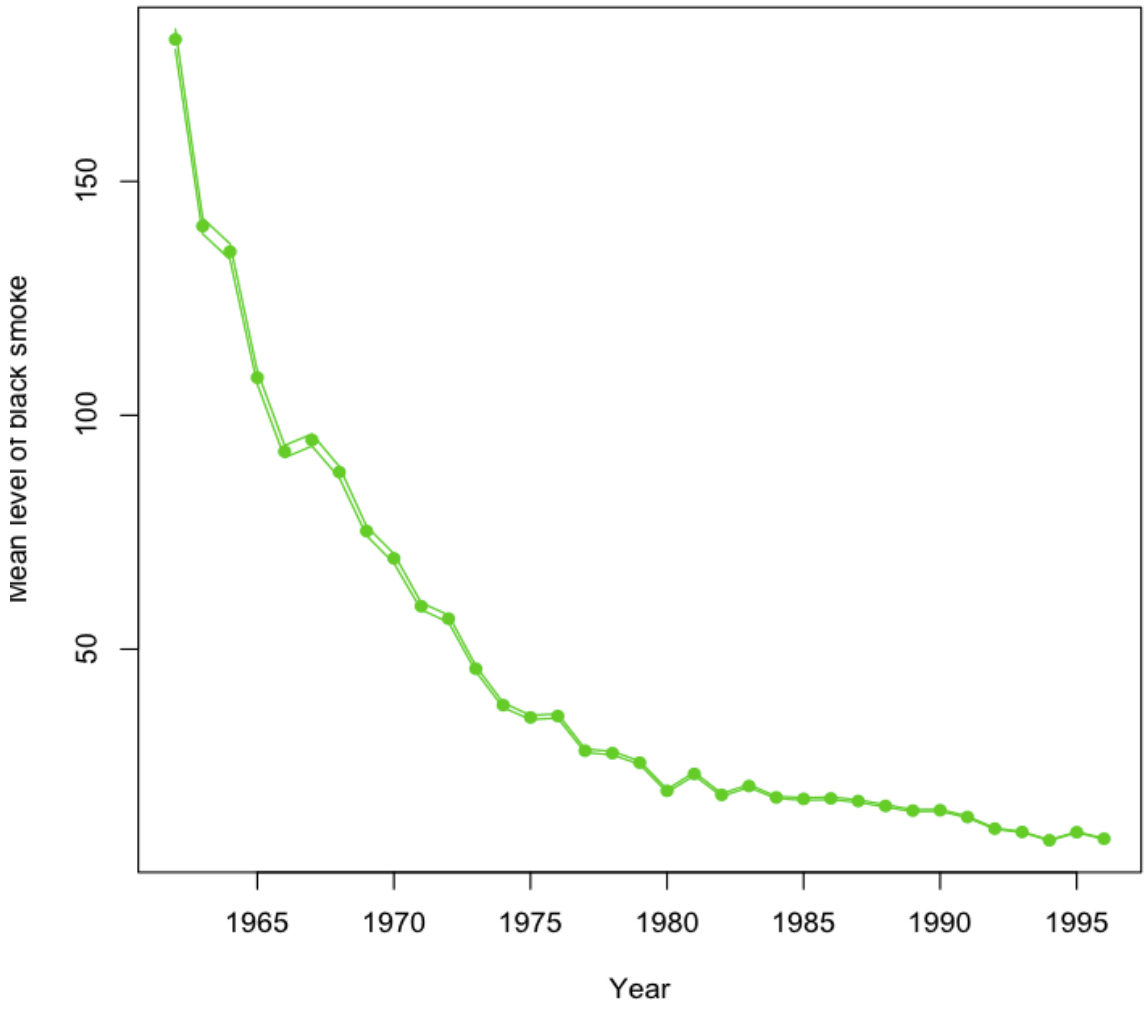
Black smoke

- consists of fine particulate matter
- is emitted mainly from fuel combustion
- following the large reductions in domestic coal use, the main source is diesel-engined vehicles
- measured by its blackening effect on filters

Decrease in concentrations over time



Decrease in annual averages over time



Modelling the field over space and time

- Bayesian hierarchical model
- Annual average (log) for each site modelled as a function of time and space
- $\log(Y_{st}) = \beta_0+\beta_s+\beta_t+\varepsilon_{st}$
  - s = location, t = year
- Linear effect of time (after taking logs)
- Site random effects are assumed MVN
  - $\beta_s \sim MVN(0, \sigma^2 I)$ - independent
  - $\beta_s \sim MVN(0, \sigma^2 \Sigma)$ – spatial

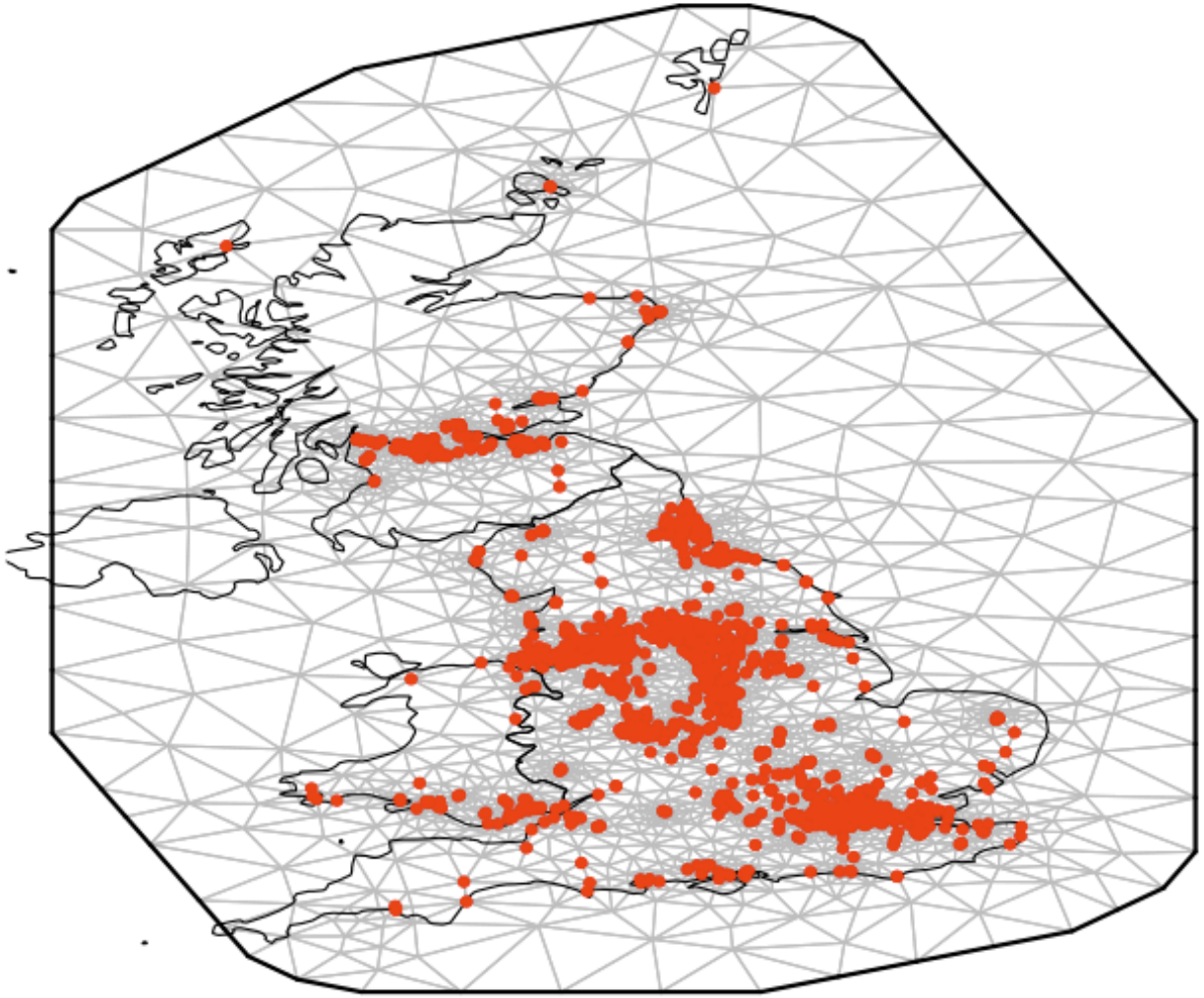
Spatial component

- If there is spatial correlation between sites (after allowing for the effect of time) then the Σ will be determined by the form of the relationship between correlation and distance.
- assume that the spatial effects represent a stationary spatial process
  - correlation between the sites dependent only on the distance between sites and not their actual location.
  - common class of models used to model such relationships is the Matern Class.
    - exponential model is a special case

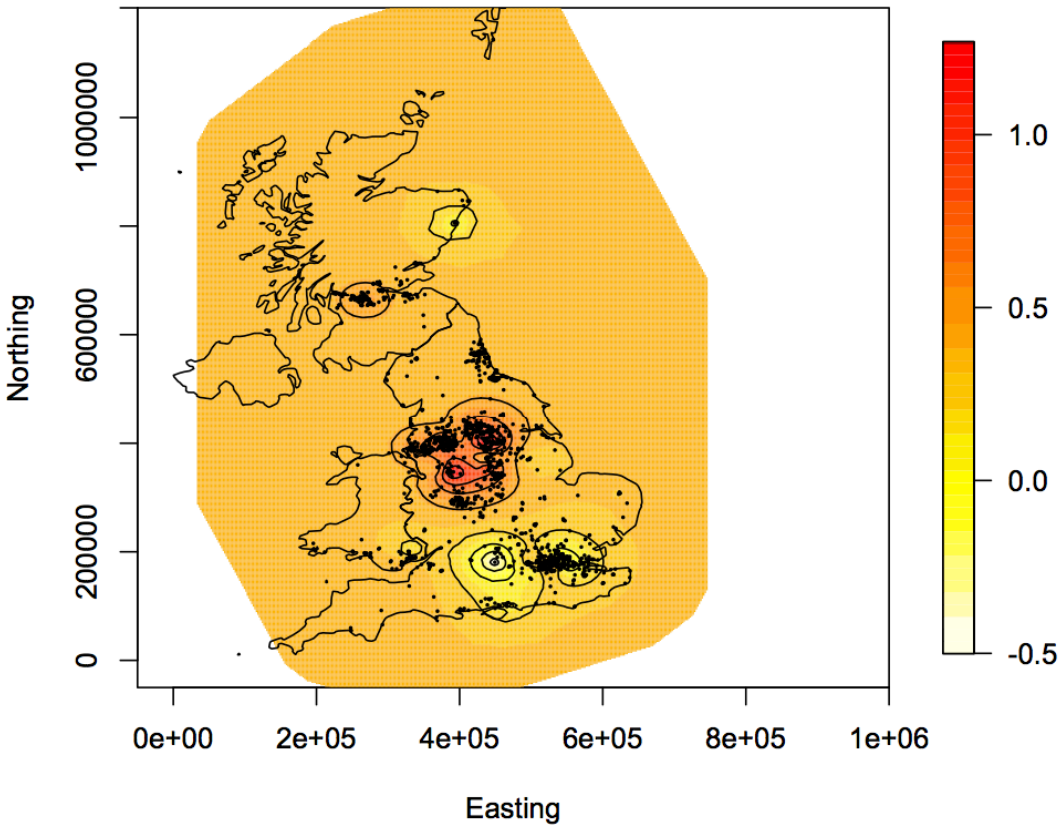
Computation

- MCMC is computationally demanding with large number of sites (1466)
- INLA uses Laplace approximations to obtain posterior marginals
  - for the latent field
  - hyperparameters
- SPDE approach
  - Gaussian field with Matern spatial covariance
  - Solution to a SPDE
  - Approximate solution to SPDE using finite element approach (Delauney triangulation)

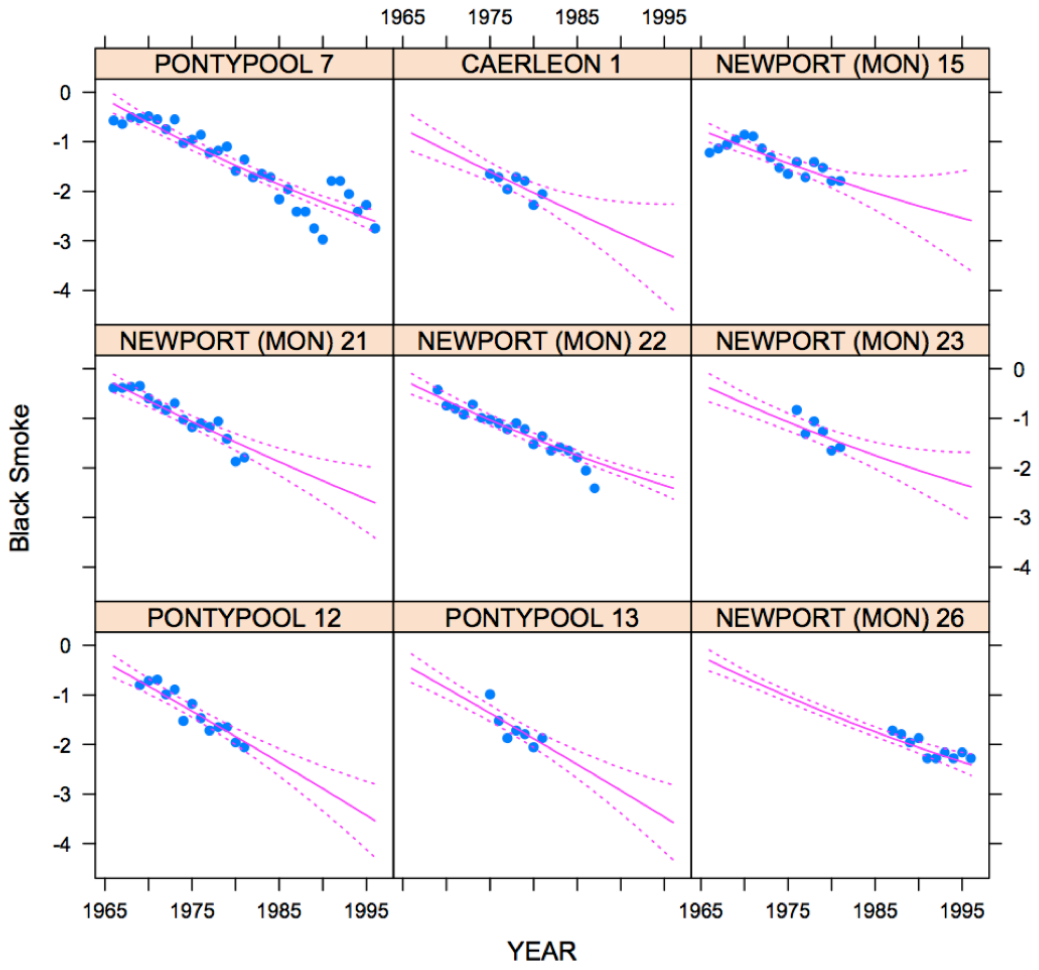
Creating a mesh using triangulation



Spatial predictions

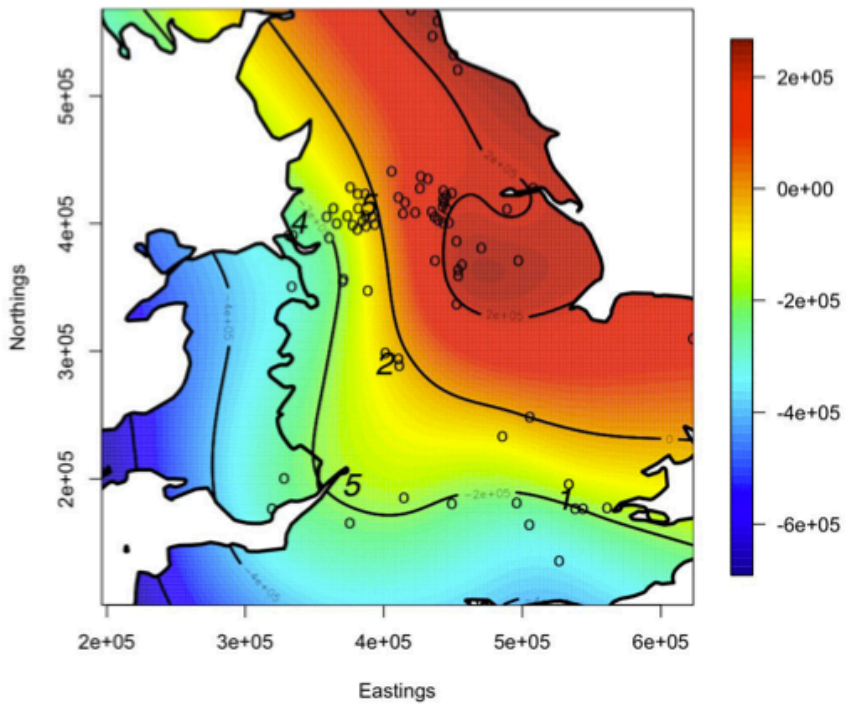


Predicted values over time



Modelling assumptions

- Is it reasonable to:
  - expect the spatial component of the model to be constant over time?
  - to assume a stationary spatial model?
- Evidence of non-stationarity
  - Incorporate geographical covariates (trend)
  - e.g. urban-rural indicator



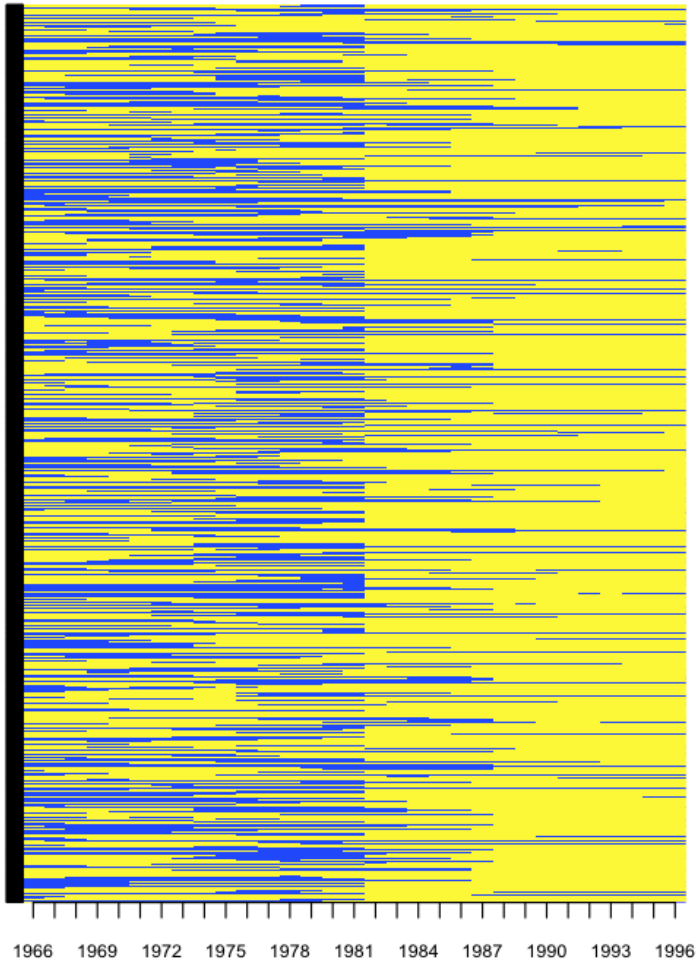
Is the data representative of ‘the truth’?

- Do monitoring networks provide information that represent underlying levels of pollution
  - for use in epidemiological studies
  - to inform policy
  - to check adherence to standards

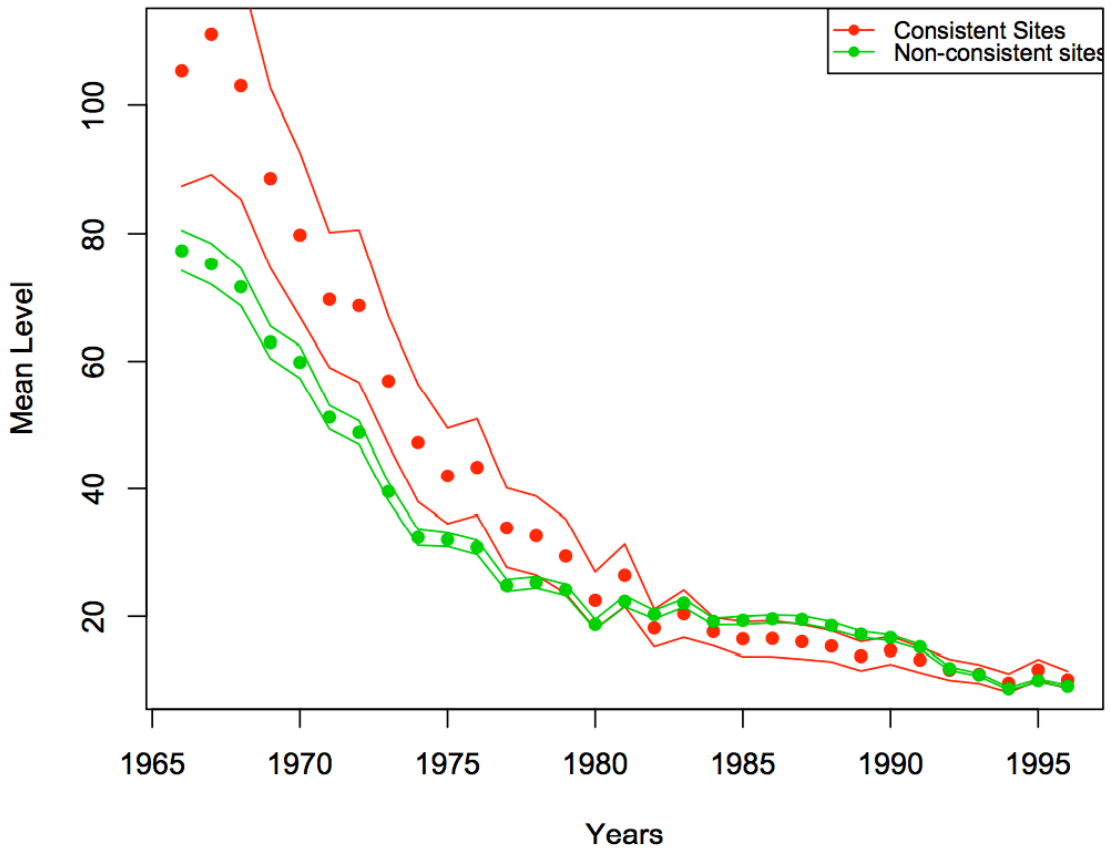
Preferential sampling

- Arises when the process that determines the locations of the monitoring sites and the process being modelled (concentrations) are in some ways dependent
- If monitoring sites are located in areas that are expected to have high (or low) concentrations
  - background levels outside of urban areas
  - levels in residential areas
  - levels near pollutant sources

Decrease in number of sites over time



Consistent v. non-consistent sites



Can we model the probability of staying in the network?

- EU directive now explicitly says that monitors can be withdrawn if measurements (yearly averages) are below guideline limits for three consecutive years
- Is there evidence that this type of reasoning (or other) has been in action over time?
- Use a logisitic regression model for the probability that a site is retained each year.
- Very strong effect of previous years measurements when reducing the network
- We are working on trying to use such probabilities to try and estimate sampling weights in a Horowitz-Thompson style correction (from survey sampling)

The network today

- In 2006 the Black Smoke/ SO2 network was replaced by the UK Black Carbon research monitoring programme
  - 20 monitoring sites
  - Locations chosen to aid health assessment
    - coal burning areas of the UK
    - general urban background exposure.

- The UK recently obtained more time to comply with EU limits for particulate pollution.
  - Limits set for 2010 may not be met in in London 25 years after these limits were passed into law.

Course on Spatio-temporal methods in environmental epidemiology

- Covers methods used in environmental epidemiology where the distribution of health outcomes and related exposures are measured over both space and time
- Strong emphasis on the implementation of models in practice
- Application of the methods will be demonstrated by using commonly available computer packages:
  - R, OpenBUGS and INLA

Current research topics

- Combine disease and exposure models
  - Bayesian hierarchical models
- Feed through variability in modelled exposures models to health models in a coherent fashion
- Multiple exposures and endpoints
- Spatial-temporal modelling
  - Non-stationarity
  - Non-separable models
  - Preferential sampling
- Efficient computation
  - Increased availability of data