MANCHESTER 1824
The University of Manchester

A Data Integration Approach to Estimate Personal Exposures to Air Pollution

Dr. Matthew Thomas

Lecturer in Data Science & Analytics
Department of Earth & Environmental Sciences
Email: matthew.l.thomas@manchester.ac.uk

With thanks to…

- **University of Manchester**
  - Matthew L Thomas
  - David Topping
  - Thomas J. Brannan
  - Hugh Coe
  - James Evans
- **Royal Holloway**
  - Gavin Shaddick
- **Technical University of Denmark**
  - Karyn Morrissey (Technical University of Denmark)
- **Newcastle University**
  - Mike Diessner
- **Alan Turing Institute**
  - Ruth Bowyer
  - Fernando Benitez-Paez
- **University of Exeter**
  - Stefan Siegert
- **University of British Columbia**
  - James Zidek

Air Pollution and Health

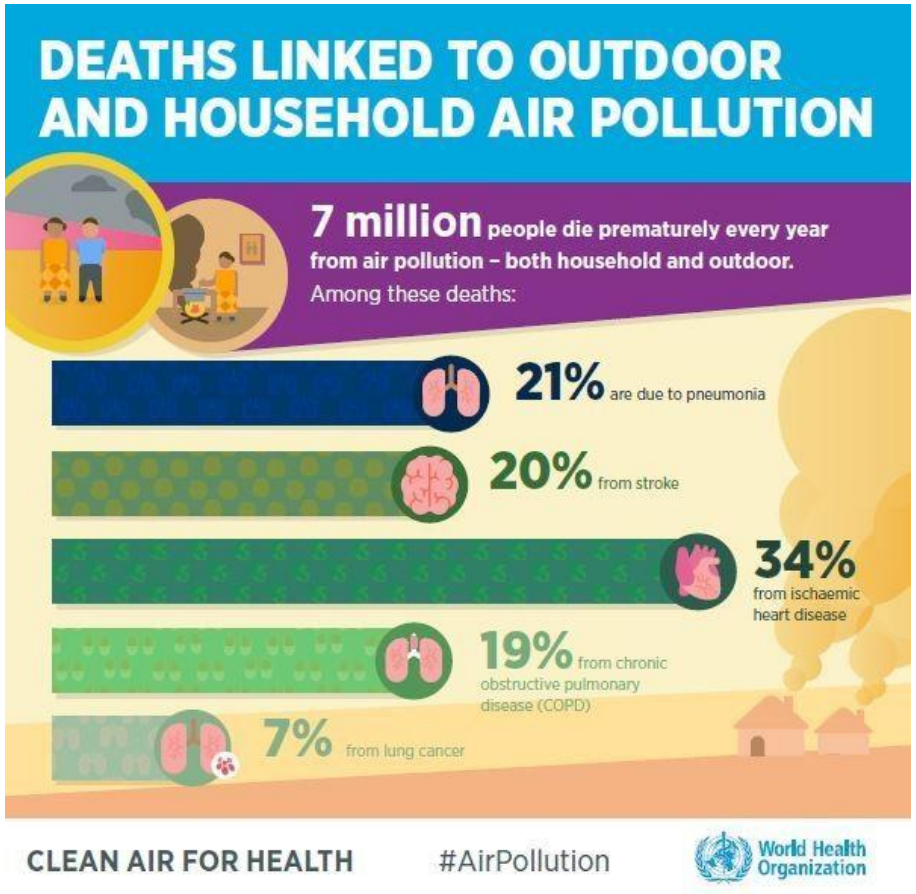
- Globally, air pollution is the largest environmental risk to health
- WHO estimate that 4.2 million premature deaths every year can be attributed to ambient (outdoor) air pollution
- Over 99% of people worldwide are exposed to levels of $PM_{2.5}$ above the WHO Air Quality Guidelines
- Understanding the links between air pollution and health require detailed information of exposures to air pollution

Air Pollution and Health

- Accurate estimates of exposure to air pollution are required
  - Global, national and local levels
  - Measures of uncertainty
- How do we estimate exposures to air pollution?
  - Multiple data sources and products
  - Data is often out of date for decision making
  - Lack of disaggregated/detailed information
- What type of exposure
  - Population level
  - Personal level

A realistic estimate of exposure?

- Majority of research related to the health effects of air pollution has been at a population level
  - Measured or modelled concentrations of ambient pollution
  - Matched to residential address
- This does not necessarily reflect individual's exposures to different levels of air pollution throughout the day
- People move through a series of micro-environments with different levels of pollution
  - Work, home, school, outdoor, car, etc…

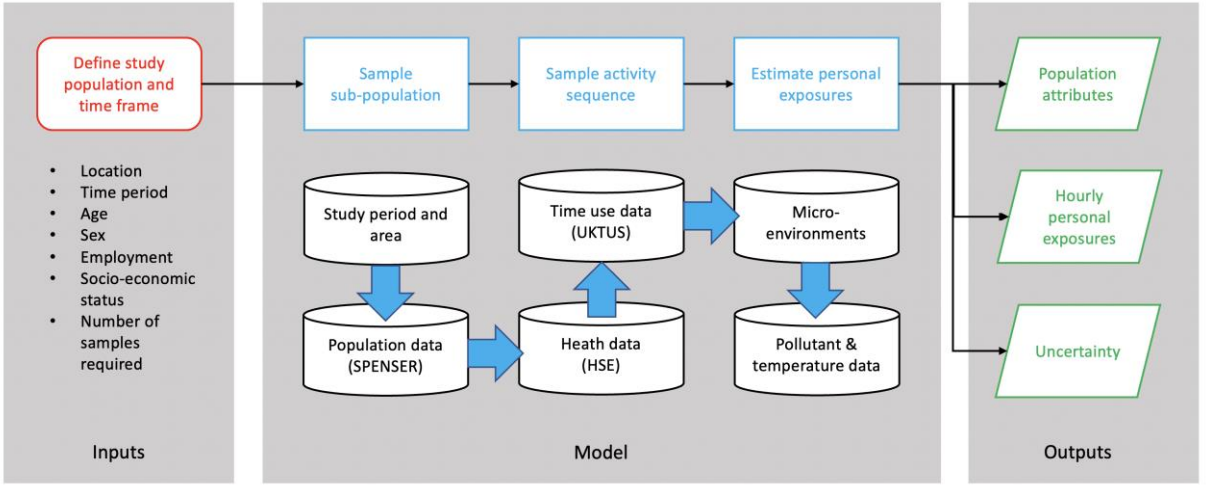


Data Integration Model for Exposures (DIMEX)

- The aim is to estimate personal exposures to air pollution
- Framework for integrating data on air pollution concentrations with population demographics, activities, locations and other factors affecting individuals' exposures
- Simulate the daily exposure of different population groups using agent-based modelling
- Differences between personal exposures and concentrations
- Consider all information relating to individual's exposure to pollutant in question
  - I: individual factors, e.g. age and sex
  - E: external factors, e.g. pollution, temperature
  - B: human behaviour

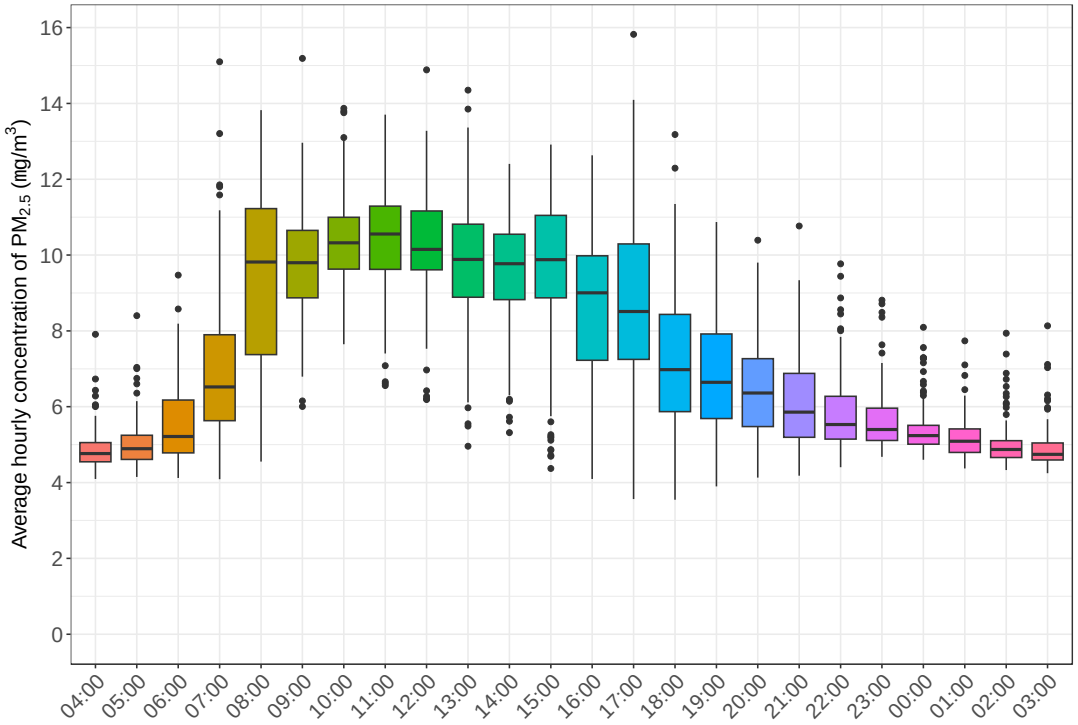
Framework for Personal Exposures Estimation

- DIMEX consists of structural linkages between the model elements and builds upon previous exposure simulators, including pCNEM, SHEDS, APEX
- Incorporates new modelling techniques and increasing availability of data
  - Demographic information
  - Activity patterns
  - Micro-environments
- Generates a sequence of pollutant concentrations to which a randomly selected individual is exposed over time



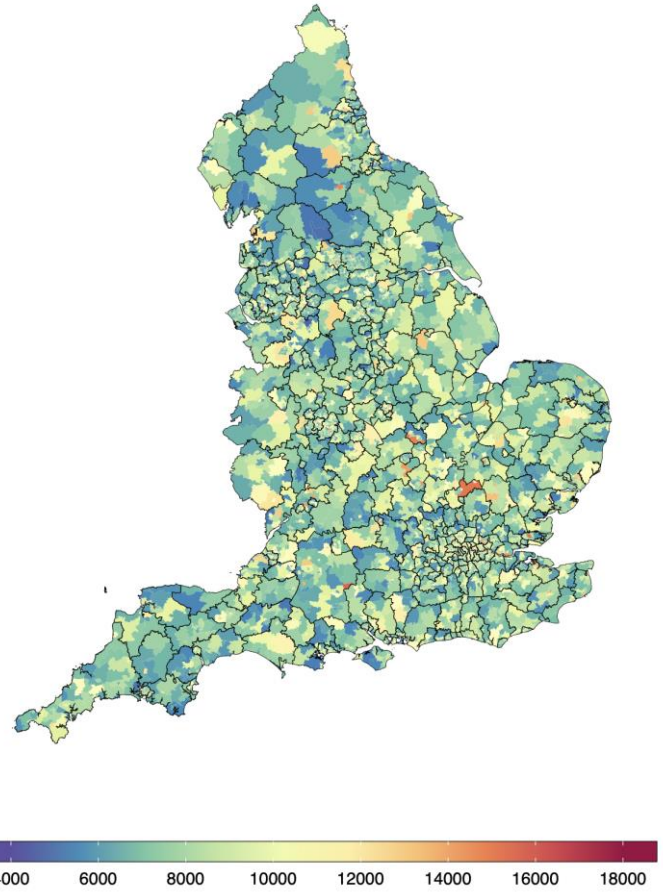
Framework for Personal Exposures Estimation

- Estimates of personal exposures aggregated to populations
  - Measures of uncertainty
  - Individual's personal trajectory maps
  - Map differences between personal exposures and concentrations
- Can be used as inputs for health impact analyses and epidemiological risk models



Data Integration Model for Exposures (DIMEX)

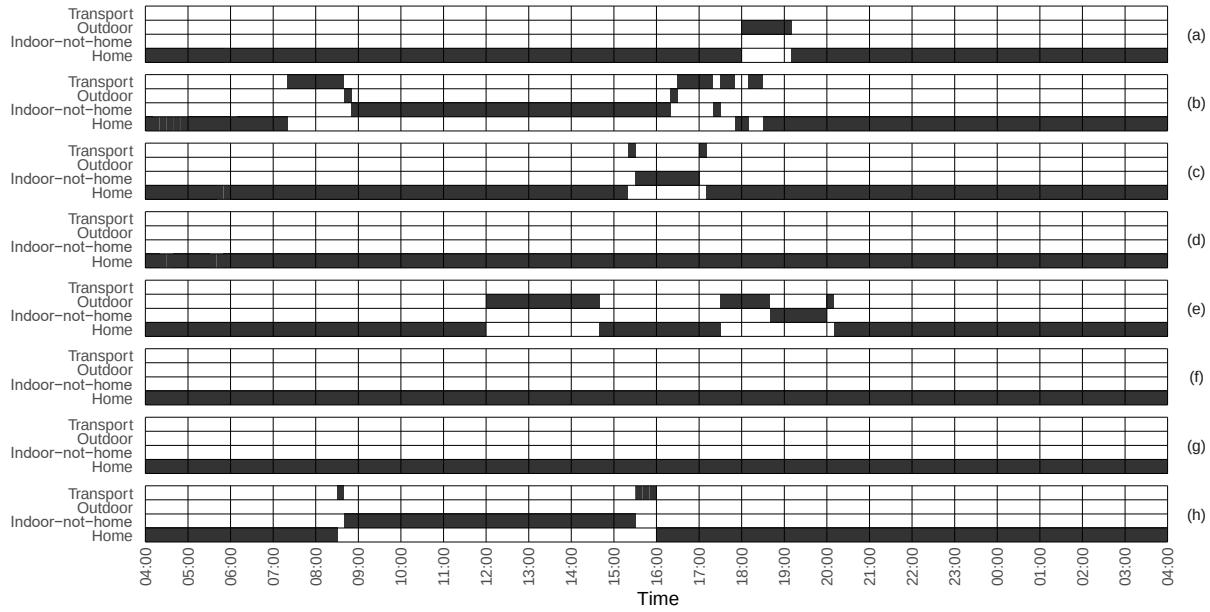
- **Population Data**
  - Sampled from an underlying synthetic population (SPENSER) from the area of interest.
  - SPENSER combines census data with other surveys and datasets to create a geo-referenced synthetic population forecast.



Data Integration Model for Exposures (DIMEX)

- **Population Data**
  - Sampled from an underlying synthetic population (SPENSER) from the area of interest.
  - SPENSER combines census data with other surveys and datasets to create a geo-referenced synthetic population forecast.

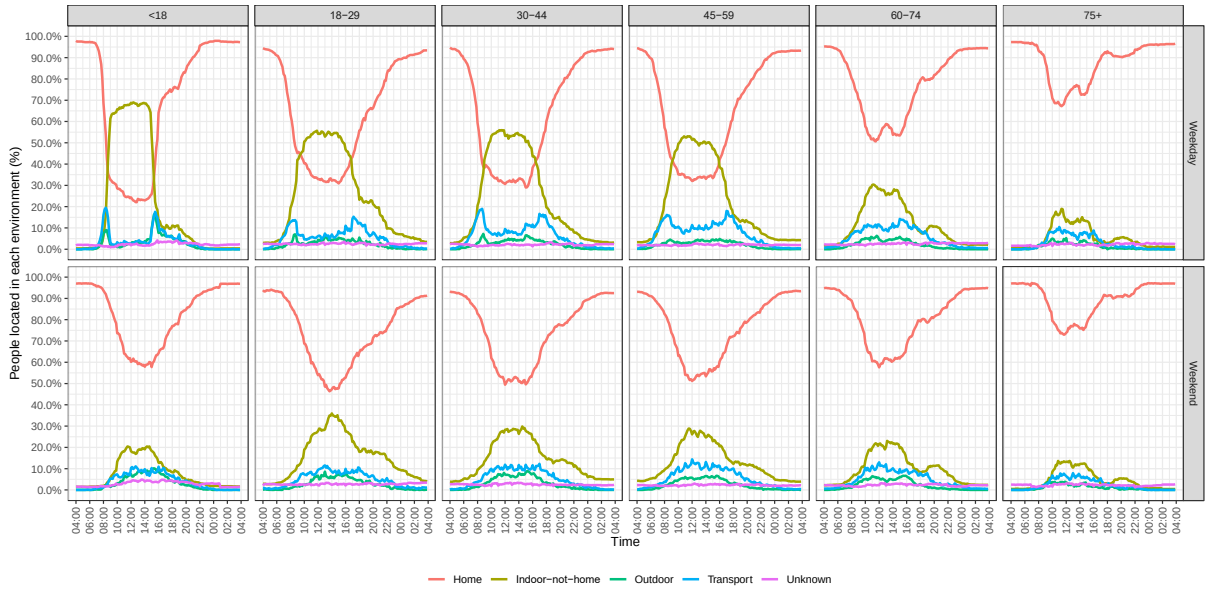
- **Activity Sampler**
  - UK Time Use Survey provides information on how people in the UK spend their time.
  - Diaries consists of sequences of activities and the locations that they take place in.



Data Integration Model for Exposures (DIMEX)

- **Population Data**
  - Sampled from an underlying synthetic population (SPENSER) from the area of interest.
  - SPENSER combines census data with other surveys and datasets to create a geo-referenced synthetic population forecast.

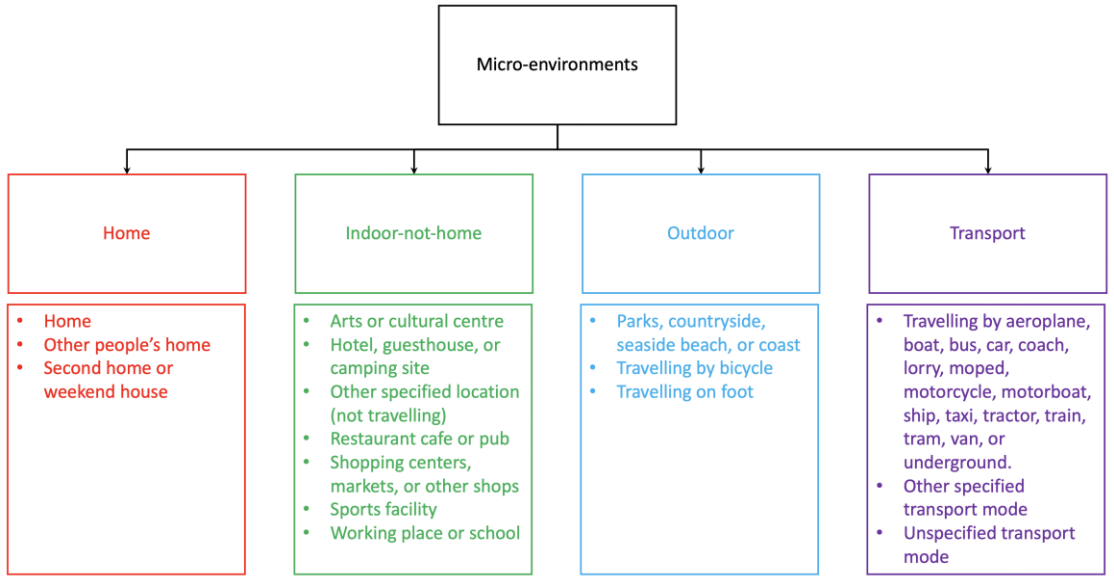
- **Activity Sampler**
  - UK Time Use Survey provides information on how people in the UK spend their time.
  - Diaries consists of sequences of activities and the locations that they take place in.



Data Integration Model for Exposures (DIMEX)

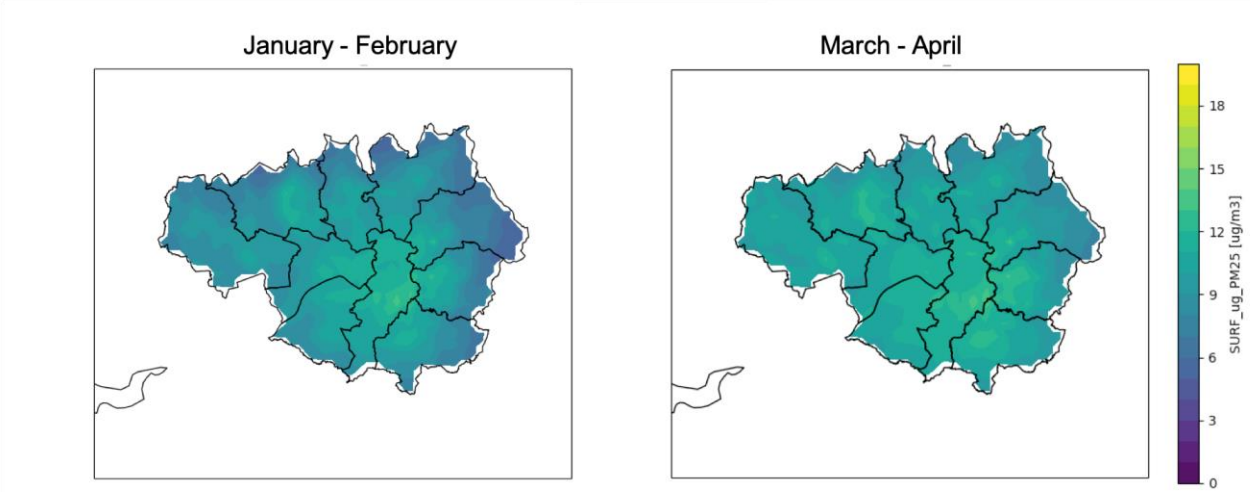
- **Population Data**
  - Sampled from an underlying synthetic population (SPENSER) from the area of interest.
  - SPENSER combines census data with other surveys and datasets to create a geo-referenced synthetic population forecast.

- **Activity Sampler**
  - UK Time Use Survey provides information on how people in the UK spend their time.
  - Diaries consists of sequences of activities and the locations that they take place in.
  - Locations were categorised into four "micro-environments": Home, Indoor-not-home, Outdoor and Transport.



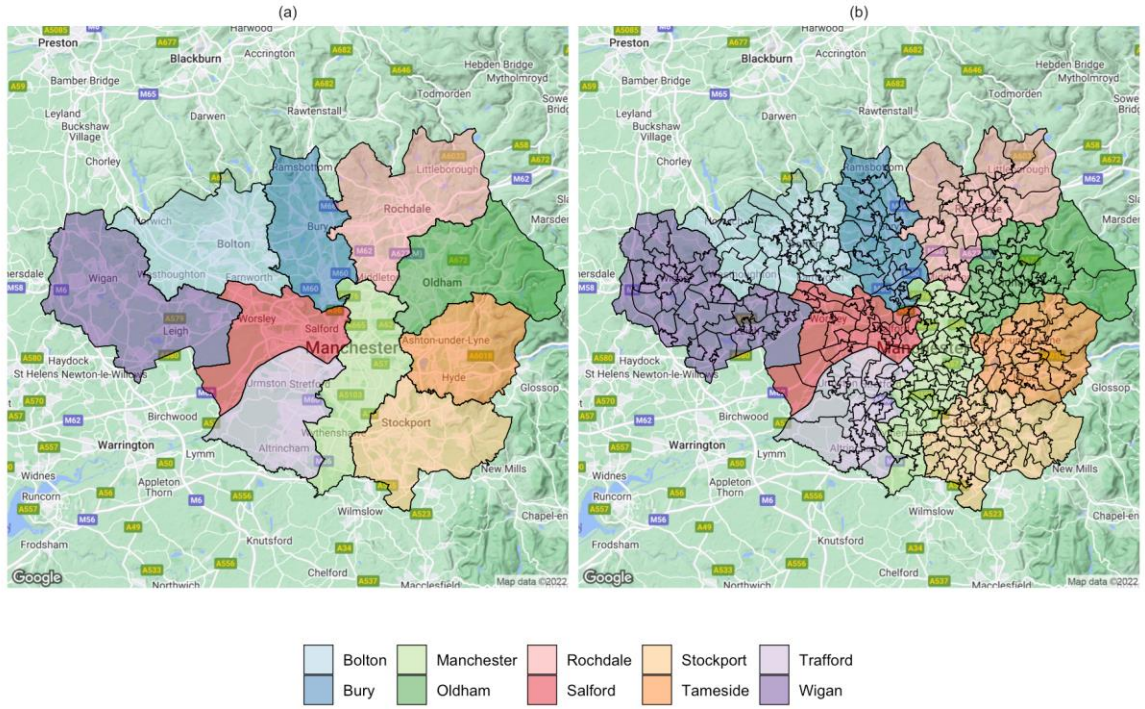
Data Integration Model for Exposures (DIMEX)

- **Exposures**
  - Activity sequences matched to concentrations of air pollutions in each micro-environment.
  - Modelled as a function of the ambient and/or non-ambient sources of air pollution.
  - Outdoor concentrations come from ground measurements or EMEP model.



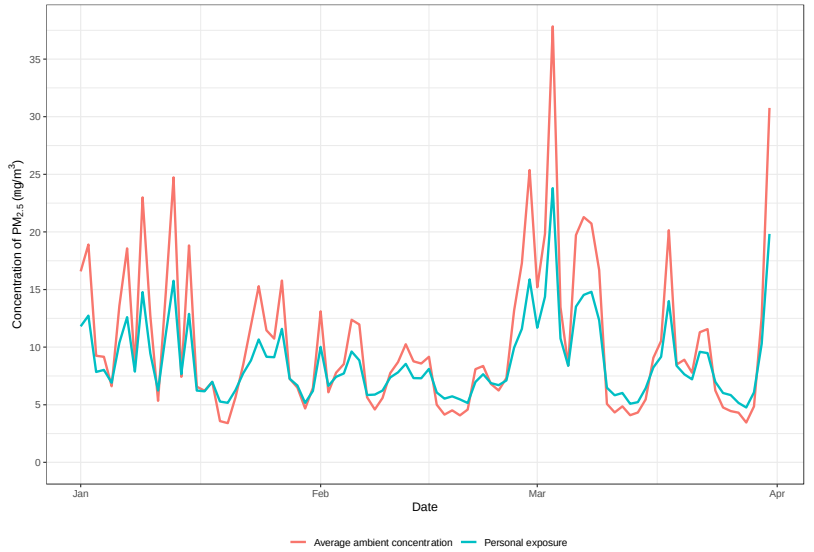
Case Study: Greater Manchester

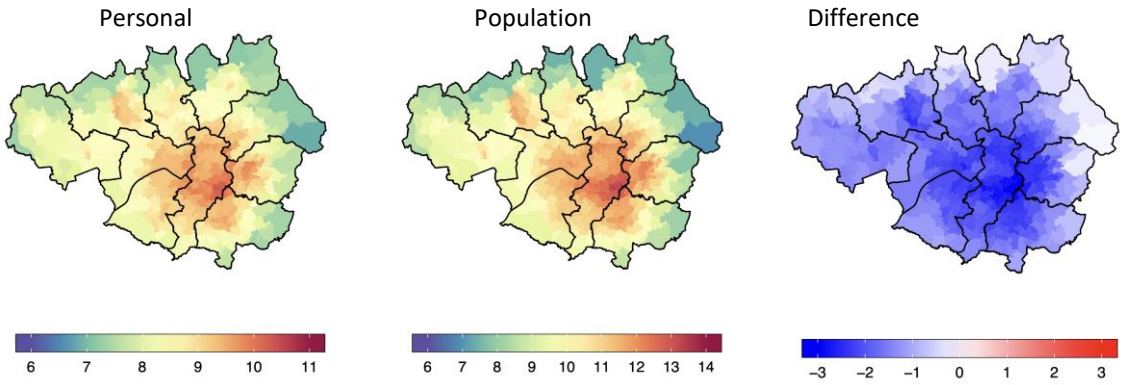
- DIMEX was used to estimate personal exposures for simulated individuals in Greater Manchester, UK
  - 2.8 million inhabitants
  - 10 metropolitan boroughs
  - 364 MSOAs
- Used modelled concentrations from EMEP for January – March 2021
- Sampled 100 individuals from the synthetic populations for each MSOA



Case Study: Greater Manchester

- DIMEX outputs hourly estimates of personal exposures that can be aggregated over **space** and **time**
- Compare personal exposures and ambient concentrations **temporally**
  - Differences of up to 15 μg/m3 for days with high concentrations
- Compare personal exposures and ambient concentrations **spatially**
  - Largest differences between the personal exposures and the ambient concentrations in urban areas
- Personal exposures are generally lower





Conclusions

- It’s important that we accurately estimate exposures to air pollution
- Many different data sources, types and products all telling the same thing
- Data Integration can be used to bring many of these sources together
  - Data Integration Model for Exposures (DIMEX)
- Many more things to do can be done
  - Linking directly with health
  - Short-term exposures
  - Other pollutants/exposures
- Requires collaboration

Thank you! Any questions

Get in touch: matthew.l.thomas@manchester.ac.uk