Using sensor data and inversion techniques to systematically reduce dispersion model error

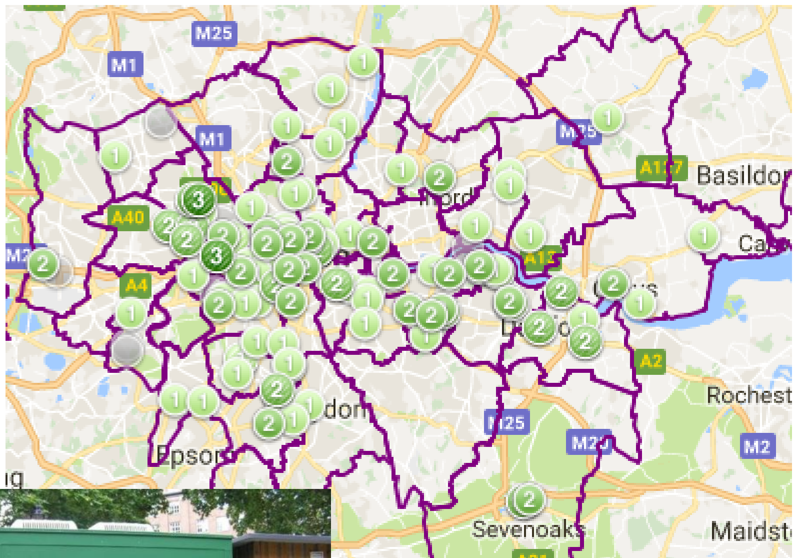
D. J. Carruthers[1], *A. L. Stidworthy*[1], *D. Clarke*[2], *K.J. Dicks*[3], *R. L. Jones*[4], *I. Leslie*[5], *O. A. M. Popoola*[4], *A. Billingsley*[6] *and M. Seaton*[1]

[1]Cambridge Environmental Research Consultants
[2]Cambridgeshire County Council, Cambridge
[3]Cambridge City Council, Cambridge
[4]Department of Chemistry, University of Cambridge
[5]Computer Laboratory, University of Cambridge
[6]AQMesh Environmental Instruments, Stratford-upon-Avon

Harmo18, Bologna, October 2017

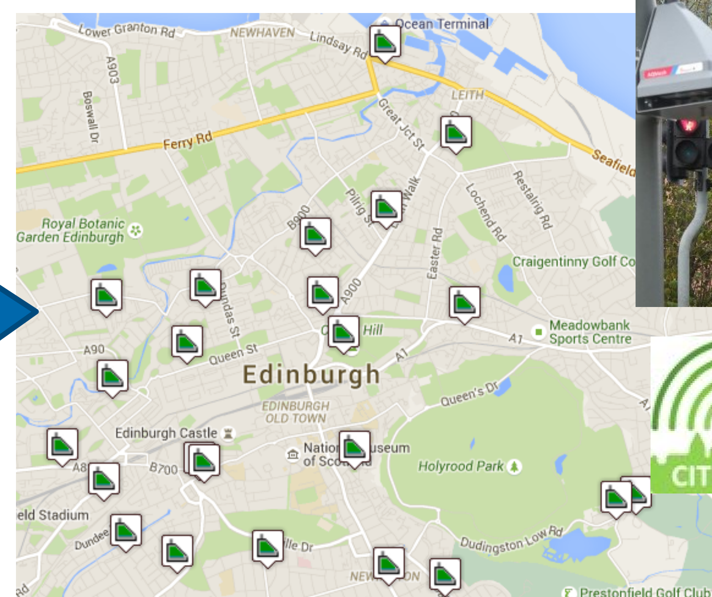
Motivation

London Air



Traditional reference-standard air quality monitoring networks are high quality, but difficult to site and expensive to maintain, so the number of monitors is limited.

Could low-cost sensors, which are less accurate but easier to site and cheaper to buy and maintain, help to improve modelling?



CITI-SENSE

CERC

Harmo18, Bologna, October 2017

Motivation

- Emissions errors in urban areas account for a significant proportion of dispersion model error
- Traditionally, dispersion models such as CERC's ADMS-Urban model are validated against data from reference monitors:
  - Modellers either use the validation to improve model setup; or
  - Calculate and apply a model adjustment factor to model results
- New low cost air pollution sensors allow large networks of sensors to be installed across a city
- Accuracy and reliability is generally lower than reference monitors, but larger spatial coverage is possible
- How can we best use these sensor data in modelling?
- If the data are not accurate and reliable enough for model validation, maybe we can use the data in a different way...

Overview

Methodology

Case Study: Cambridge

Future development

Methodology: Introduction

- The aim was to develop an inversion technique to use monitoring data from a network of sensors to automatically adjust emissions to improve model predictions
- Basic idea:
  - Run ADMS-Urban to obtain modelled concentrations at monitor locations in the normal way
  - Take these modelled concentrations and their associated emissions as a 'first guess', together with
    a) monitored concentration data
    b) information about the error in the monitored data and the proportion of that error that is systematic across all monitors
    c) information about the error in the emissions data and the proportion of that error that is systematic across all sources
  - Use an inversion technique to calculate an adjusted set of emissions that reduces error in the modelled concentrations

Methodology: Introduction

- There are some conditions that have to be satisfied for such a scheme to work:
  a) The modelled concentration must be proportional to the emissions, which means that complex effects like chemistry have to be ignored
  b) Each modelled source must contribute to the concentration at least one receptor (monitor)
  c) Each receptor included must have non-zero modelled and monitored concentration

- The technique developed uses a Bayesian inversion approach following work by others, for example as used by the Met Office for estimating volcanic ash source parameters using satellite retrievals [Webster *et al*, 2016]

Methodology: Cost function

We define a cost function J(x) with two terms: one that describes the error in the modelled concentration (left-hand term) and one that describes the error in the emissions (right-hand term):

$$J(\mathbf{x}) = (\mathbf{Mx} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{Mx} - \mathbf{y}) + (\mathbf{x} - \mathbf{e})^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{e})$$

Quantity	**Definition**	**Dimensions**
x	*Vector of emissions (result)*	*n*
M	Transport matrix relating the source term to the observations	n by k
y	Vector of observations	k
R	Error covariance matrix for the observations	k by k
e	Vector of first guess emissions	n
B	Error covariance matrix for the first guess emissions	n by n

The aim is to minimise J to obtain **x**, a vector of adjusted emissions.

Methodology: Cost function input

Quantity	**Definition**	**Dimensions**
x	*Vector of emissions (result)*	*n*
M	Transport matrix relating the source term to the observations	n by k
y	Vector of observations	k
R	Error covariance matrix for the observations	k by k
e	Vector of first guess emissions	n
B	Error covariance matrix for the first guess emissions	n by n

- The **y** and **e** vectors are straightforward to form
- To find the transport matrix **M** we run ADMS-Urban with unit emission rates for all sources and obtain the concentration at the receptors due to each source; the concentration results give **M**

Methodology: Estimating error

Estimating emissions error (**B**)

- For **emissions**, we need to estimate for each source (pair):
  - Emission error

    = Uncertainty Factor x Emission Rate
  - Co-varying emission error

    = Covariance Factor x Emission Error
- Example causes of co-varying error: common emissions factors, proximity of sources to each other
- Total Emission error includes both co-varying emission error and independent emission error

Estimating sensor error (**R**)

- For **sensors**, we need to estimate:
  - Sensor error

    = Uncertainty Factor x Monitored Concentration
  - Co-varying sensor error

    = Covariance Factor x Sensor Error
- Example causes of co-varying error: same sensor type, ambient temperature, humidity
- Total Sensor error includes both co-varying sensor error and independent sensor error

Methodology: Summary

Step 1: Run ADMS-Urban to obtain hourly modelled concentrations at monitoring site locations

Step 2: Form the transport matrix, error covariance matrices, emissions vector and monitored data vector for each hour

Step 3: Run the optimisation scheme independently for each hour

Step 4: Create an hourly factors (.hfc) file from the adjusted hourly emissions data

Step 5: Re-run ADMS-Urban using the adjusted emissions .hfc file

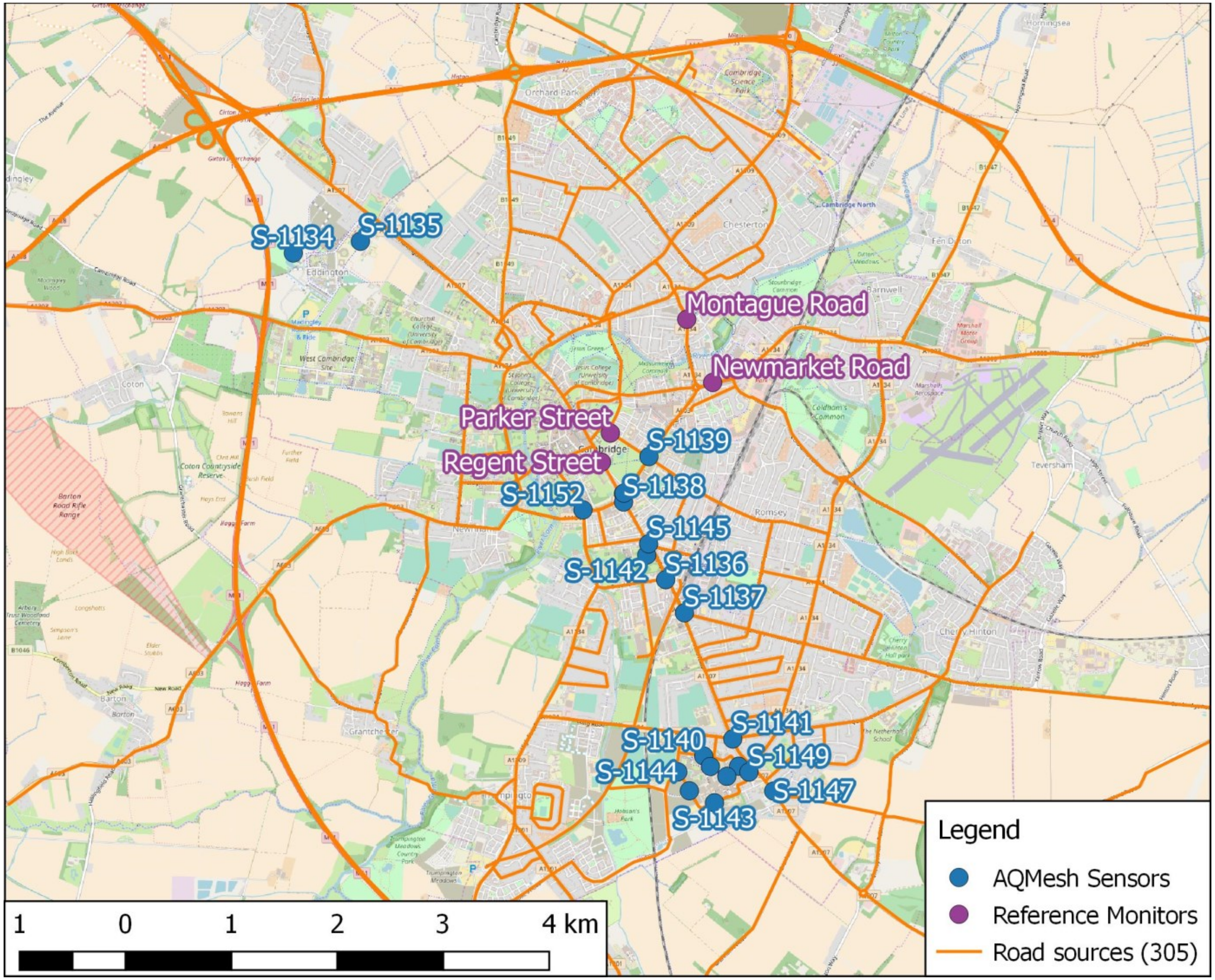
Cambridge Case Study: Background

20 AQMesh sensor pods

4 Reference monitors

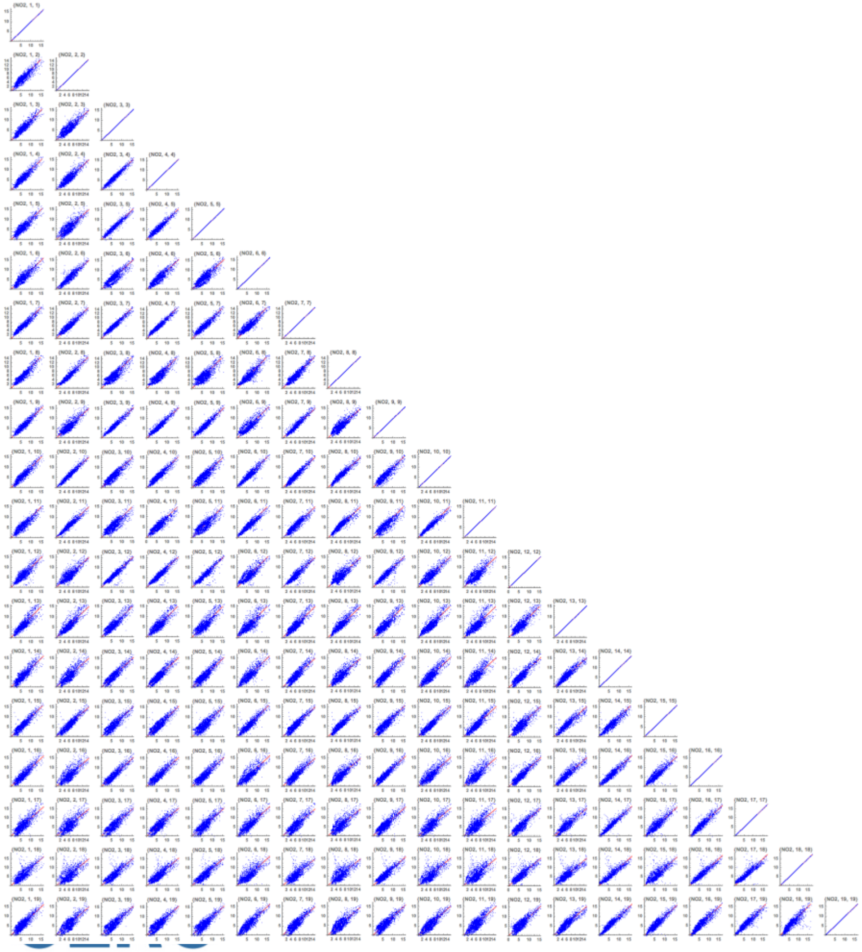
3-month analysis period, July-Sept 2016

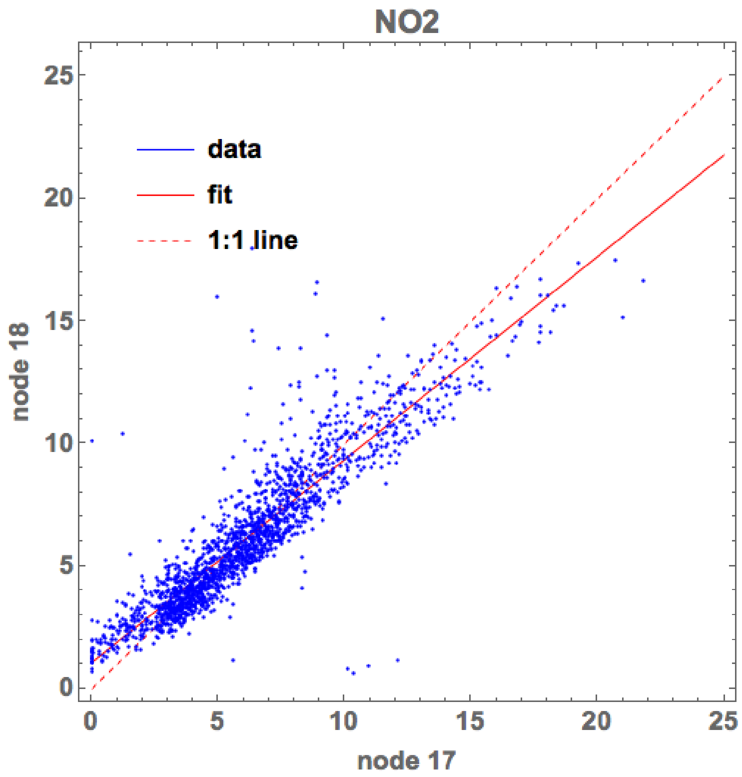
305 road sources



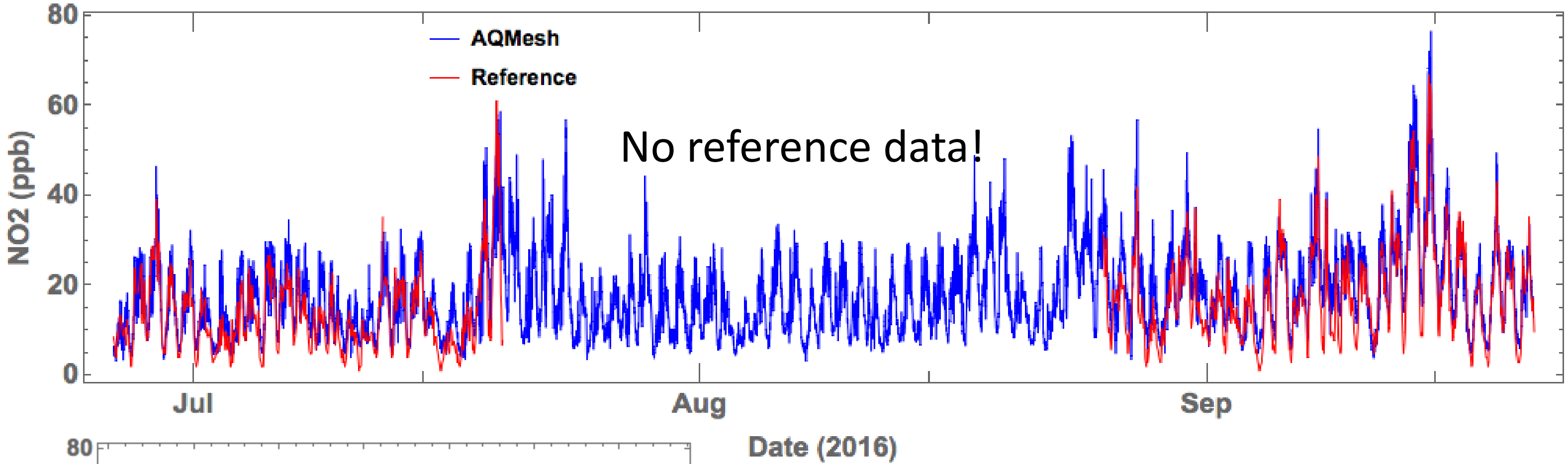
AQMesh Sensors

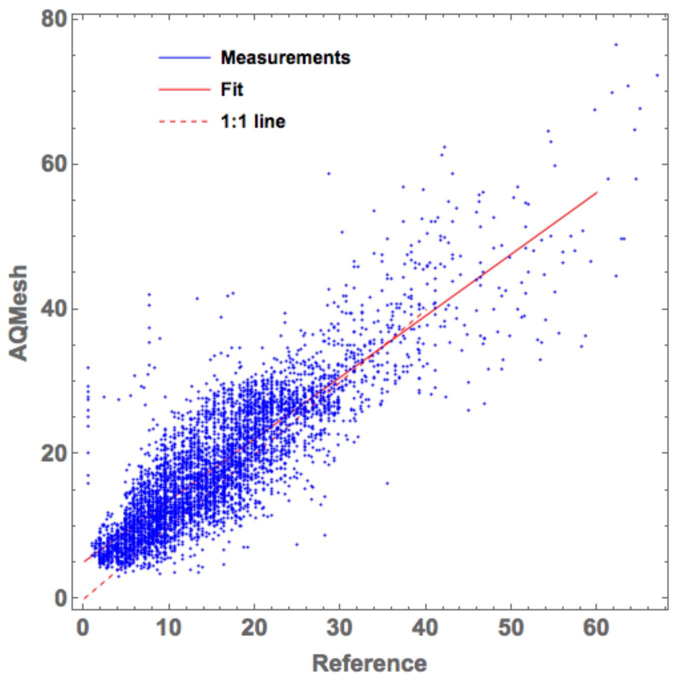
- Used out of the box – no local calibration; pre-calibrated at AQMesh test facility
- Example of performance: NO_2 sensor-sensor comparison





NO_2 Gonville Place comparison

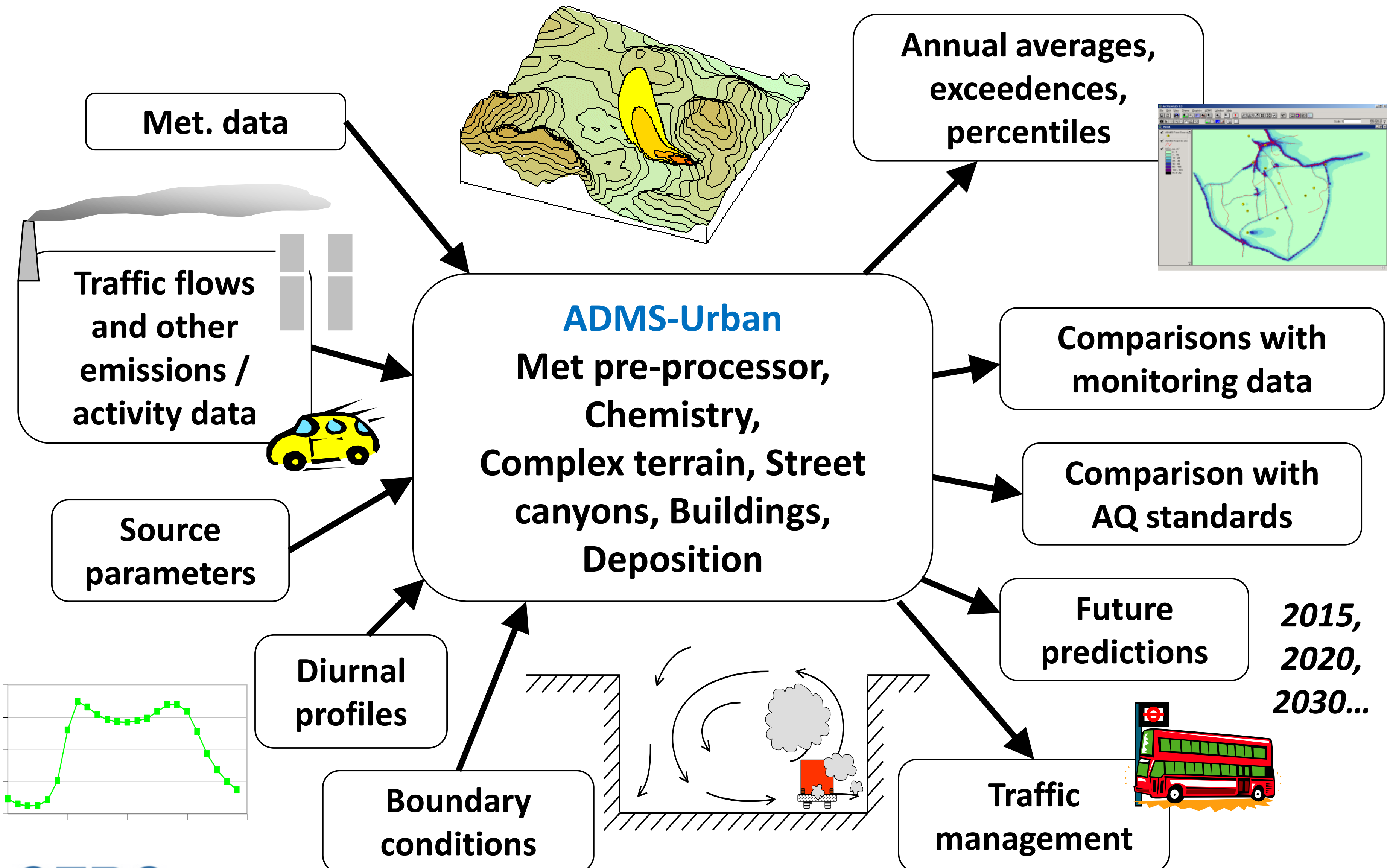




	Gradient	Intercept	R^2
pre	1.07 (0.01)	10.0 (0.1)	0.50
post	0.82 (0.01)	5.1 (0.13)	0.74

- Outliers removed (from reference)
- Significantly improved R^2
- AQMesh ~ 0.82 of reference (unscaled)

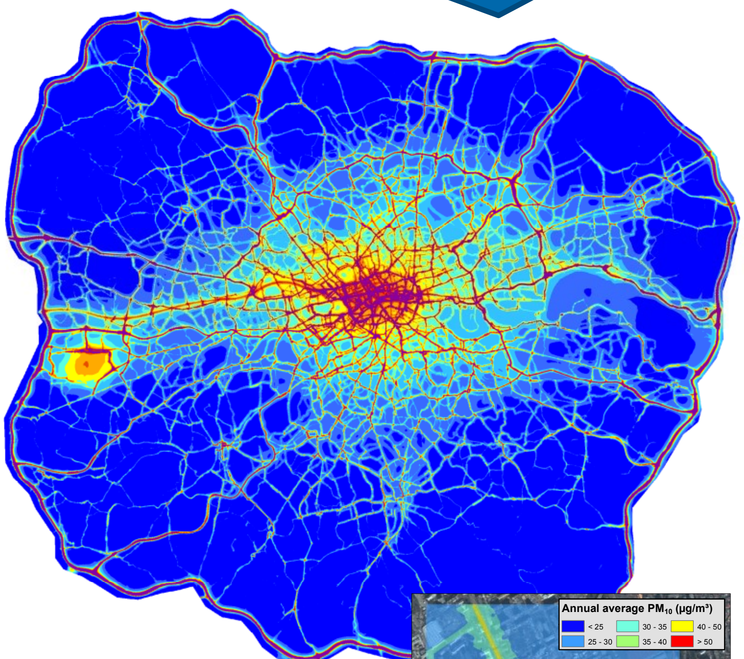
CERC's ADMS-Urban Model



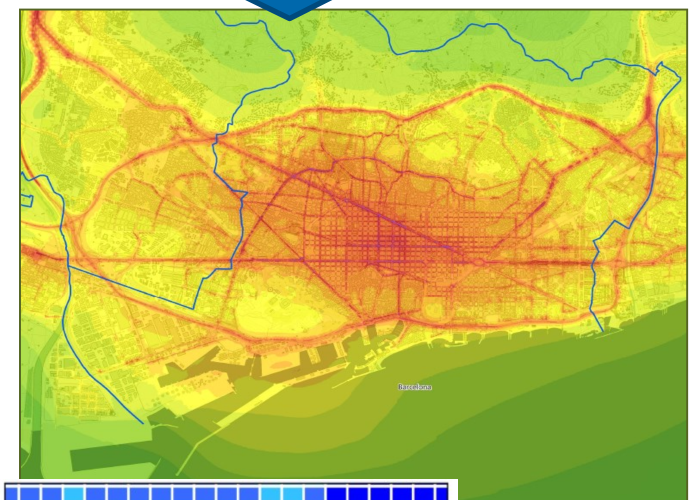
CERC

CERC's ADMS-Urban Model

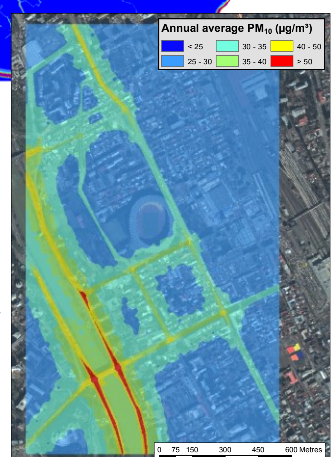
Annual average NO_2 concentrations in Greater London calculated using ADMS-Urban

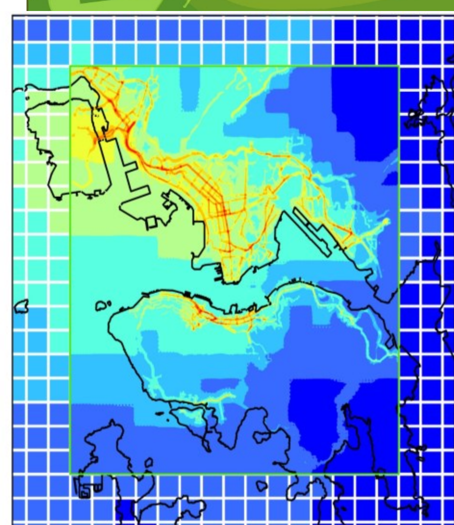


Annual average NO_2 concentration map of Barcelona calculated using ADMS-Urban. Modelling by Barcelona Regional.



Annual average PM_{10} for Tblisi in Georgia calculated using ADMS-Urban





Annual average NO_2 over Hong Kong Island calculated using ADMS-Urban linked with CAMx regional model

CERC

Cambridge Case Study: Aims

- Two aims:
  - Sense-check optimisation results, find and correct errors
  - Test this hypothesis: *Using inversion techniques, we can use sensor data to improve emissions and thereby improve model performance, judged at independent reference monitors.*
- Initial study -simple implementation
  - Only one source type: roads
  - Only one pollutant: NO_X
  - Only 20 sensors – relatively small network
  - Simple representation of error covariance

Methodology: Summary

Step 1: Run ADMS-Urban to obtain hourly modelled concentrations at monitoring site locations

Step 2: Form the transport matrix, error covariance matrices, emissions vector and monitored data vector for each hour

Step 3: Run the optimisation scheme independently for each hour

Step 4: Create an hourly factors (.hfc) file from the adjusted hourly emissions data

Step 5: Re-run ADMS-Urban using the adjusted emissions .hfc file

Steps 2 to 5 completed three times, for three different scenarios:

1. AQMesh sensor and reference monitors included in the optimisation.

- Including the reference monitors helps us to sense-check the results and identify any errors

2. Only reference monitors included in the optimisation

- This scenario is also included to sense-check results and identify errors

3. Only AQMesh sensors included in the optimisation.

- In this scenario, the reference monitor data is kept as an independent dataset for model validation.

Cambridge Case Study: ADMS-Urban Setup

Emissions

- Annual averages + diurnal profiles (weekdays, Saturdays, Sundays)
- Road traffic count data from UK Govt and County Council
- Guided bus flows
- Road traffic emission factors for 2016 from the UK National Atmospheric Emissions Inventory (NAEI), adjusted for real-world emissions

Met data

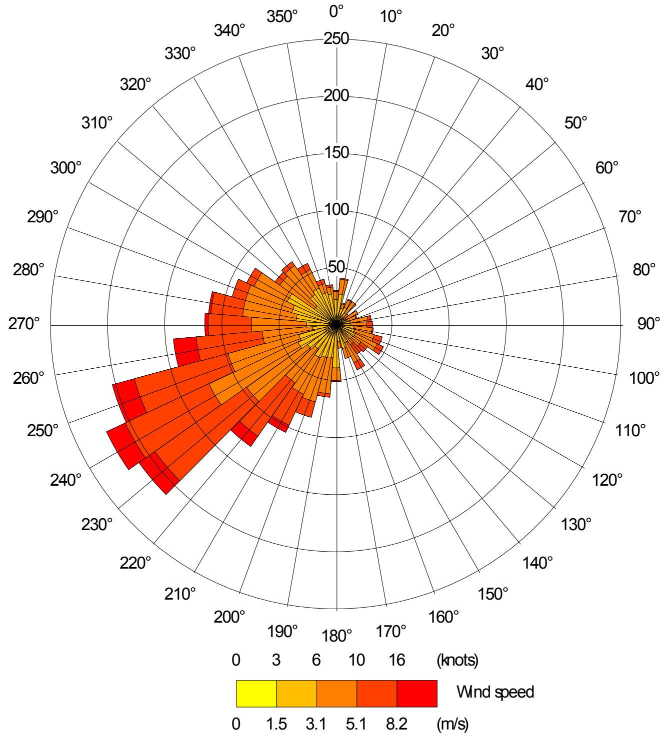
- Andrewsfield Met Office site, 21 June – 30 September

Background data

- Background NO_x from Defra AURN measurements at rural sites

Monitoring data used for validation

- All monitoring data are provisional apart from Gonville Place reference monitor; AQMesh data were obtained in real time.



CERC

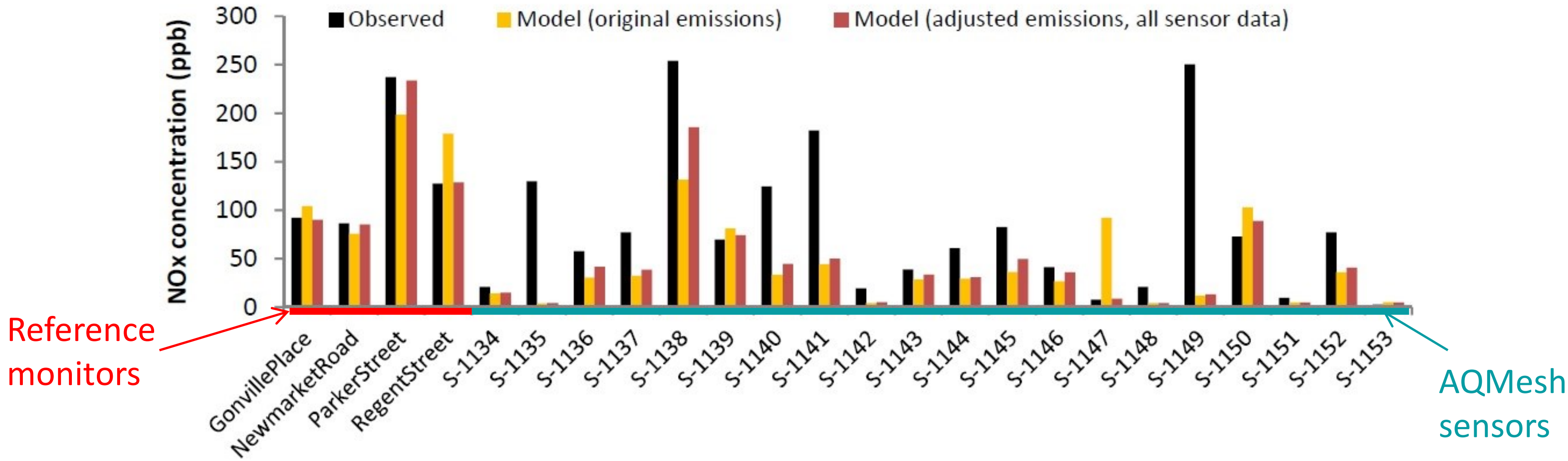
Cambridge Case Study: Error Estimation

Parameter name	Description	Value
U_{OR}	Observation uncertainty factor (reference monitors)	0.1
U_{OS}	Observation uncertainty factor (AQmesh sensors)	0.3
U_{ORF}	Observation error covariance factor (reference monitors)	0.05
U_{OSF}	Observation error covariance factor (AQmesh sensors)	0.1
U_{E}	Emissions uncertainty factor	0.5
U_{EF}	Emissions error covariance factor	0.4

- Plausible estimates - would need refinement in any further study
- Assumed error covariance factors for both the sensors and reference instruments were small
- Assumed error covariance factors are more significant for emissions since depend on road traffic emission factors common to all sources

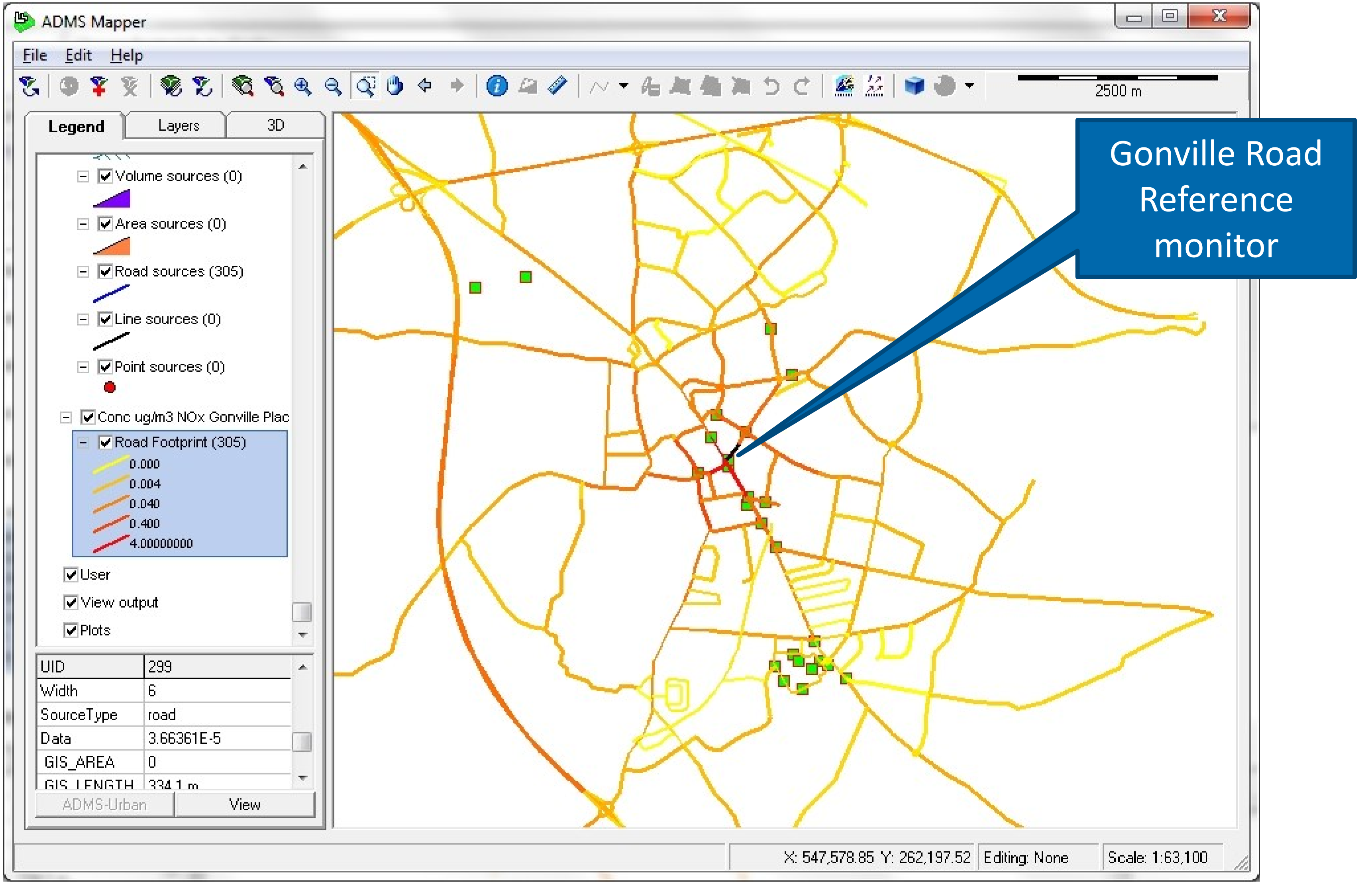
Cambridge Case Study: Optimisation results

Example of observed, modelled and adjusted NO_x concentration for one hour only



- The optimisation makes greater adjustment to the modelled concentration at the reference monitors than to the modelled concentration at the sensors - sensor uncertainty is higher than the reference monitor uncertainty
- The optimisation adjusts the modelled concentration at all sensors, not just at a selection of sensors - non-zero error covariance between sensors

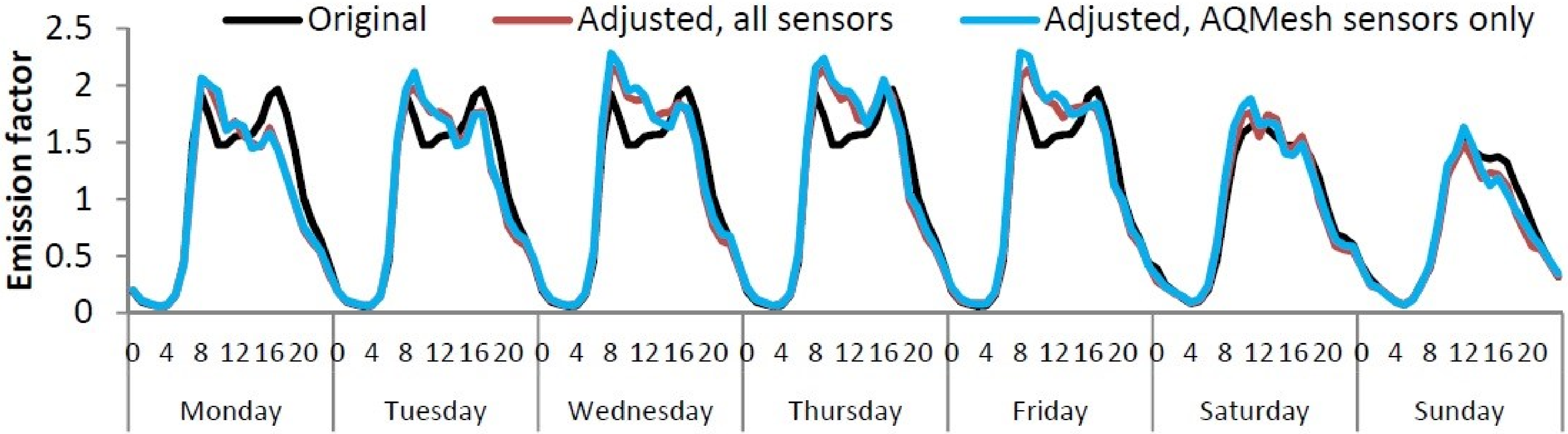
Footprint: Contribution of roads to receptor



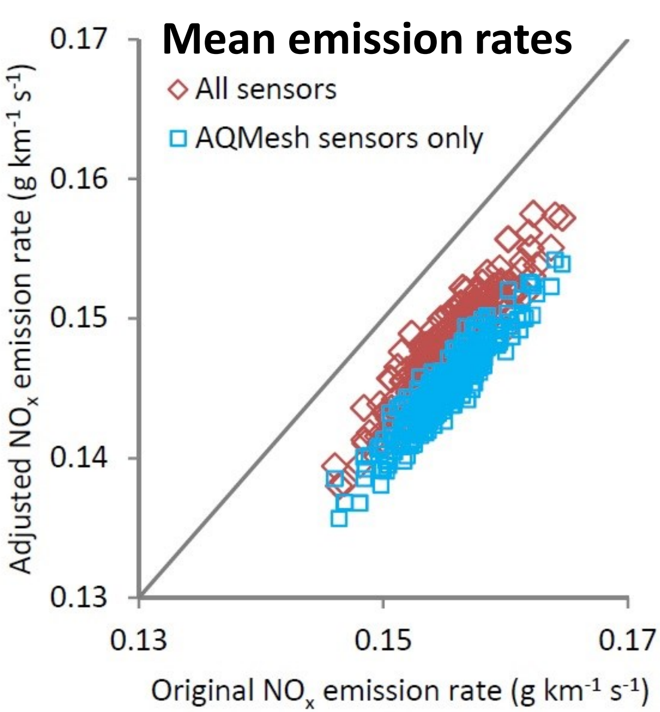
CERC

Cambridge Case Study: Optimisation results

Average diurnal profiles



Emission rate ($gkm^{-1}s^{-1}$)	Original	Adjusted	Change
All sensors	0.1552	0.1478	-4.8%
AQMesh only		0.1452	-6.5%



Optimisation increases AM peak, decreases PM peak

Optimisation reduces mean emission rate of all sources

Including reference monitors has only a small effect on overall emissions

CERC

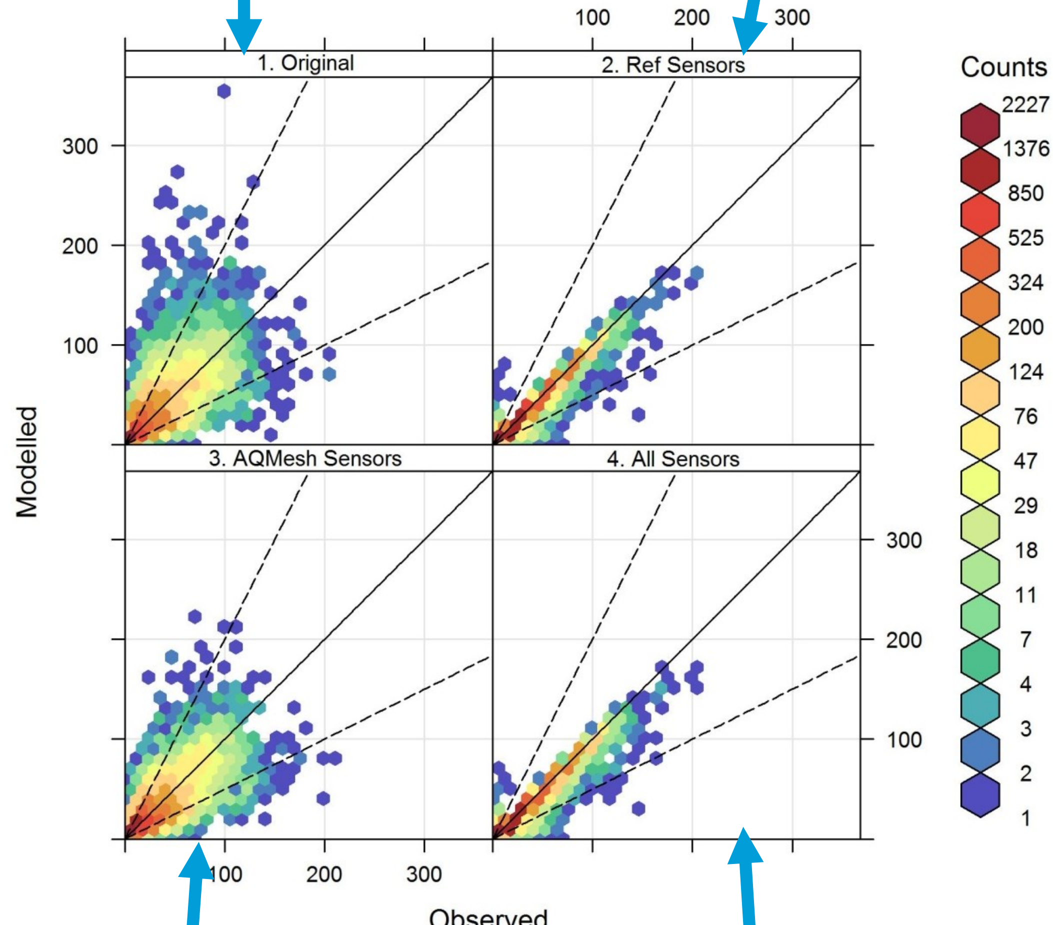
Cambridge Case Study: Model outcomes

Validation at Reference sites only

Statistics		1.	2.	3.	4.
Mean	Obs	31.2	31.2	31.2	31.2
	Mod	34.5	29.9	31.0	29.4
StDev	Obs	27.9	27.9	27.9	27.9
	Mod	31.0	26.6	27.0	26.1
	MB	3.30	-1.28	-0.23	-1.78
	NMSE	0.51	0.04	0.39	0.05
	R	0.70	0.98	0.75	0.97
	Fac2	0.71	0.94	0.73	0.94

Original model results

Only reference sensor data included in optimisation



Only AQMesh sensor data included in optimisation

All sensor data included in optimisation

CERC

Conclusions

- The optimisation scheme presented here, using inversion techniques to modify pollution emission rates based on sensor data, has been shown to improve the accuracy of modelled concentrations.
- This study used a relatively simple representation of error covariance. Indicators of emissions error covariance that are not yet accounted for include:
  - Distance between sources
  - Meteorological factors such as temperature
  - Multiple pollutants (only NO_x so far)
  - Different source types (only roads so far)
- Defining/refining the covariance in error between different pollutants and between difference source types presents a challenge
- These initial results suggest that this approach could make practical use of large networks of low-cost sensors to improve dispersion model results and emission inventories.