Modelling exposure and health impacts of air pollution: the GEMS project

John Gulliver

University of the West of Scotland

& Imperial College London

David Briggs & Anna Hansell

Imperial College London

Global and regional Earth-system Monitoring using Satellite and in-situ data

An EC FP6 Integrated Project, developing:

- global modelling and data assimilation for greenhouse gases, reactive gases and aerosols
- a validated global production system, including surface-flux estimation
- collaborative regional modelling, analysis and forecasting of air quality for Europe

GEMS health studies

1. Demonstrate the use of GEMS data for retrospective health assessment via epidemiological analysis of air pollution episodes
2. As a basis for health risk assessment for policy support (using forecasting from GEMS models)
3. GEMS can also potentially be used for health risk management and intervention (EU-wide emission legislation; local sources - traffic management; climate change adaptation and mitigation)

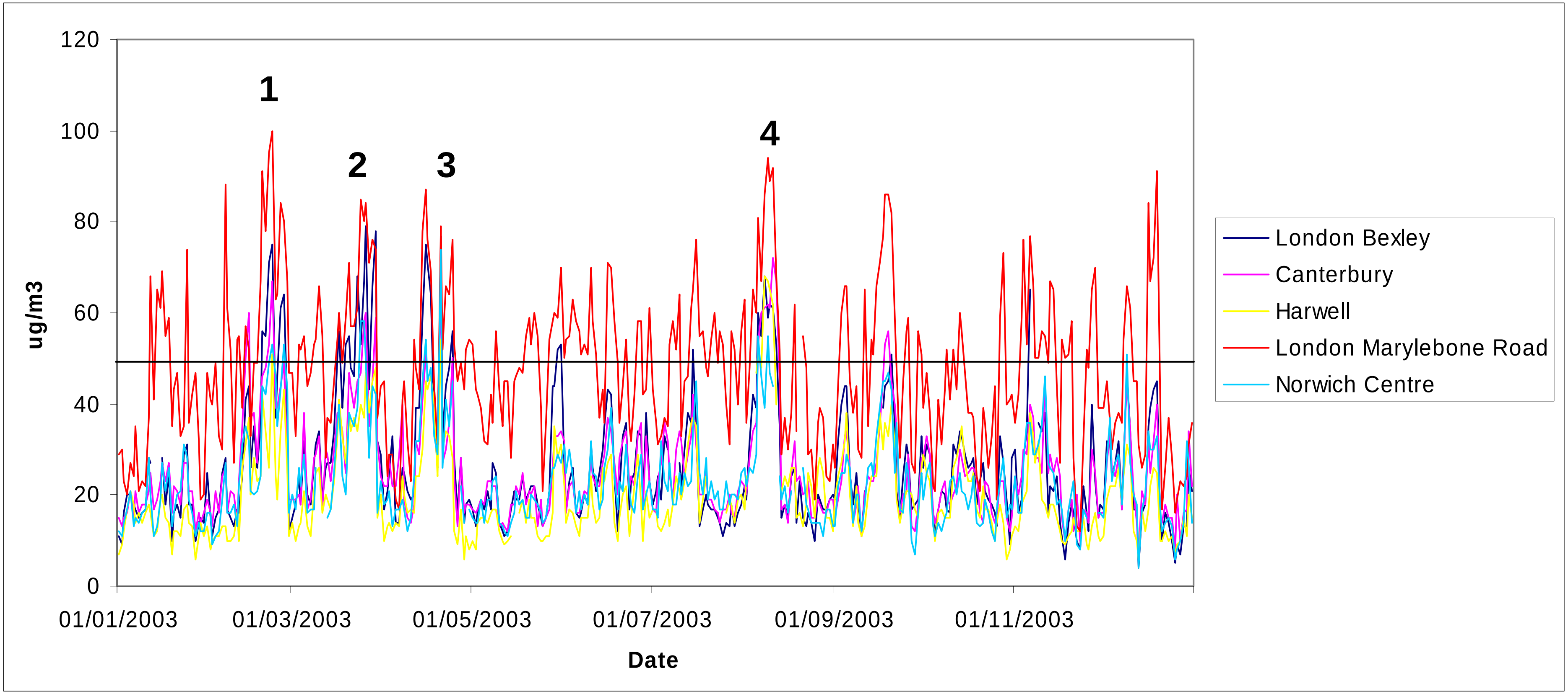
Health study: research questions

1. Are there spatial variations in health risk across London?
2. Do risks vary between air pollution episodes and non-episodes:
    - in terms of rates;
    - in terms of geographic pattern
3. To what extent are the spatial variations explained by variations in:
    - socio-economic status
    - air pollution?
4. To what extent can these risks be predicted through the use of GEMS data (when data become available)

Retrospective analysis

1. UK study focussing on:
   - Long-range air pollution + local source contributions
   - Surface temperature
   - Acute effects (i.e. daily)
2. Selection of major trans-boundary air pollution episodes:
   - Use of autoregressive statistical techniques to define episodes
   - Duration around Ca. 10 days +
   - Long-range events tracking across UK and Ireland
3. Selection of matched reference period
   - Same season
   - Same days of week
   - Same duration
   - ‘No (significant) difference’ temperature, windspeed /wind direction
   - >1 month before/after any episode

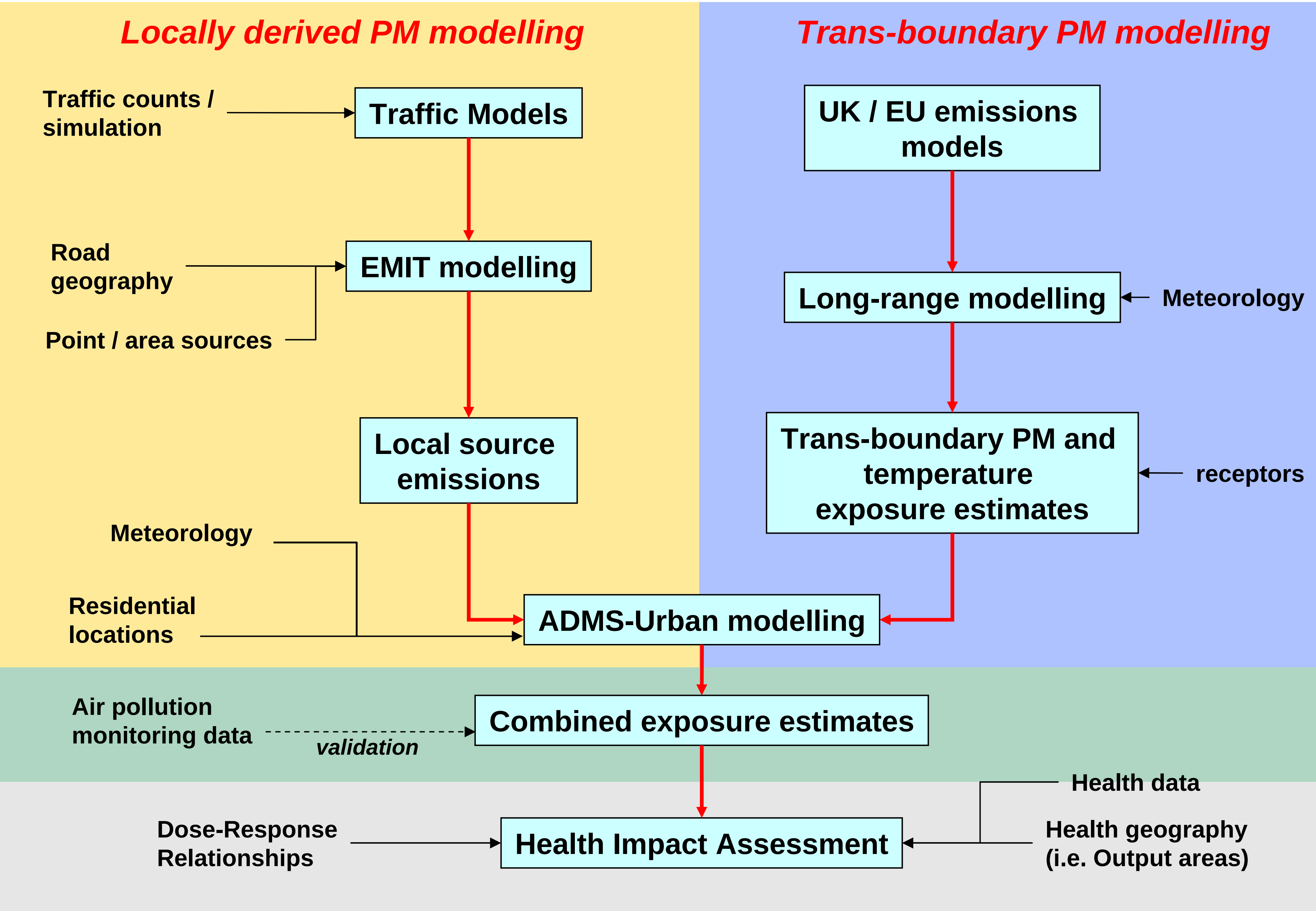
PM10 (daily mean) episodes in 2003



1. 13th to 26th February (14 days)
2. 16th to 29th March (14 days)
3. 14th to 27th April (14 days)
4. 1st to 15th August (15 days)

AQS for daily mean PM10 = 50 ug/m3

UK health study – modelling option ‘1’



Modelling of Locally derived PM10 (ADMS-Urban):

- Based on LAEI emissions inventories for 2002 and 2003
- Road geography (OS Land-line) and data on traffic composition integrated in GIS for c. 60, 000 road links
- Emissions rates for each road and grid source (1km) calculated using EMIT and transferred to ADMS-Urban
- Hourly meteorological data (wind parameters, temperature, cloud cover) obtained for 2002-2004 for London
- Circa 100 ADMS-Urban runs for each day x 57 days x 2 years

Modelling of long-range and sub-national transport of PM10 (NAME):

- Trans-boundary and sub-national transport of fine particulates
- Production of secondary PM10 over UK and Europe

Approaches to modelling air pollution

1. PM10 = **f**(Local traffic and non-traffic sources,
long-range and regional air pollution from the NAME model)

2. PM10 = **f**(local traffic and non-traffic sources,
rural air pollution monitoring data)

Approach 2. is possible for only hind-casts and subsequent retrospective health impact assessment….but is a good marker for what should be possible in terms of model performance!

How much of the explained variation in monitored concentration can we expect to be able to model?

Site effects: we are not directly modelling localised particles from soil, vegetation, construction work etc!

Correlation between modelled far-travelled PM_{10} (NAME) and monitoring sites

Site	*type*	*r*	*r^2*	*SEE (µg/m^3)*
A3	Roadside	0.62 *[0.39]*	0.38 *[0.16]*	11.77 *[9.47]*
Bexley	Urban Background	0.64 *[0.50]*	0.41 *[0.25]*	11.82 *[8.13]*
Bloomsbury	Urban Centre	0.68 *[0.50]*	0.46 *[0.25]*	9.47 *[8.01]*
Brent	Urban Background	0.66 *[0.48]*	0.43 *[0.23]*	10.17 *[7.64]*
Camden	Kerbside	0.63 *[0.43]*	0.39 *[0.19]*	10.69 *[9.44]*
Eltham	Suburban	0.64 *[0.51]*	0.40 *[0.20]*	11.40 *[8.38]*
Haringey	Roadside	0.66 *[0.51]*	0.44 *[0.26]*	10.20 *[7.74]*
Hillingdon	Suburban	0.64 *[0.50]*	0.41 *[0.25]*	11.14 *[9.09]*
Kensington & Chelsea	Urban Background	0.67 *[0.53]*	0.45 *[0.28]*	10.40 *[8.62]*

[non-episode periods, only]

Rural sites r^2 = 0.45 - 0.60

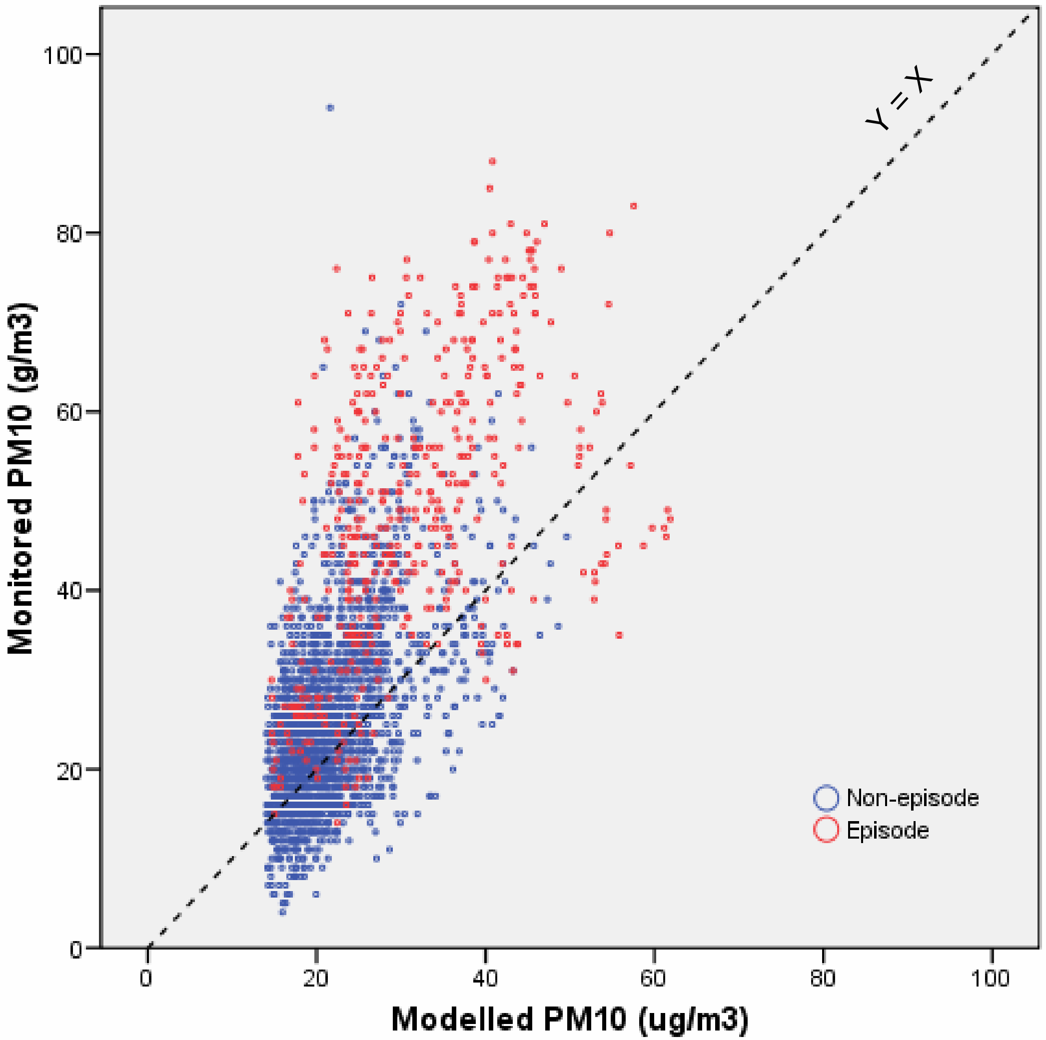
N = 250 days (Bloomsbury = 147 days)

Correlation between modelled ‘TOTAL’ PM_{10} and monitoring sites

Site	*type*	*r*	r^2	*SEE (µg/m³)*
A3	Roadside	0.64	0.41	11.52
Bexley	Urban Background	0.64	0.41	11.84
Bloomsbury	Urban Centre	0.69	0.47	9.36
Brent	Urban Background	0.66	0.43	10.17
Camden	Kerbside	0.65	0.42	10.42
Eltham	Suburban	0.64	0.40	11.36
Haringey	Roadside	0.69	0.47	9.88
Hillingdon	Suburban	0.67	0.44	10.84
Kensington & Chelsea	Urban Background	0.67	0.44	10.39

N = 250 days (Bloomsbury = 147 days)

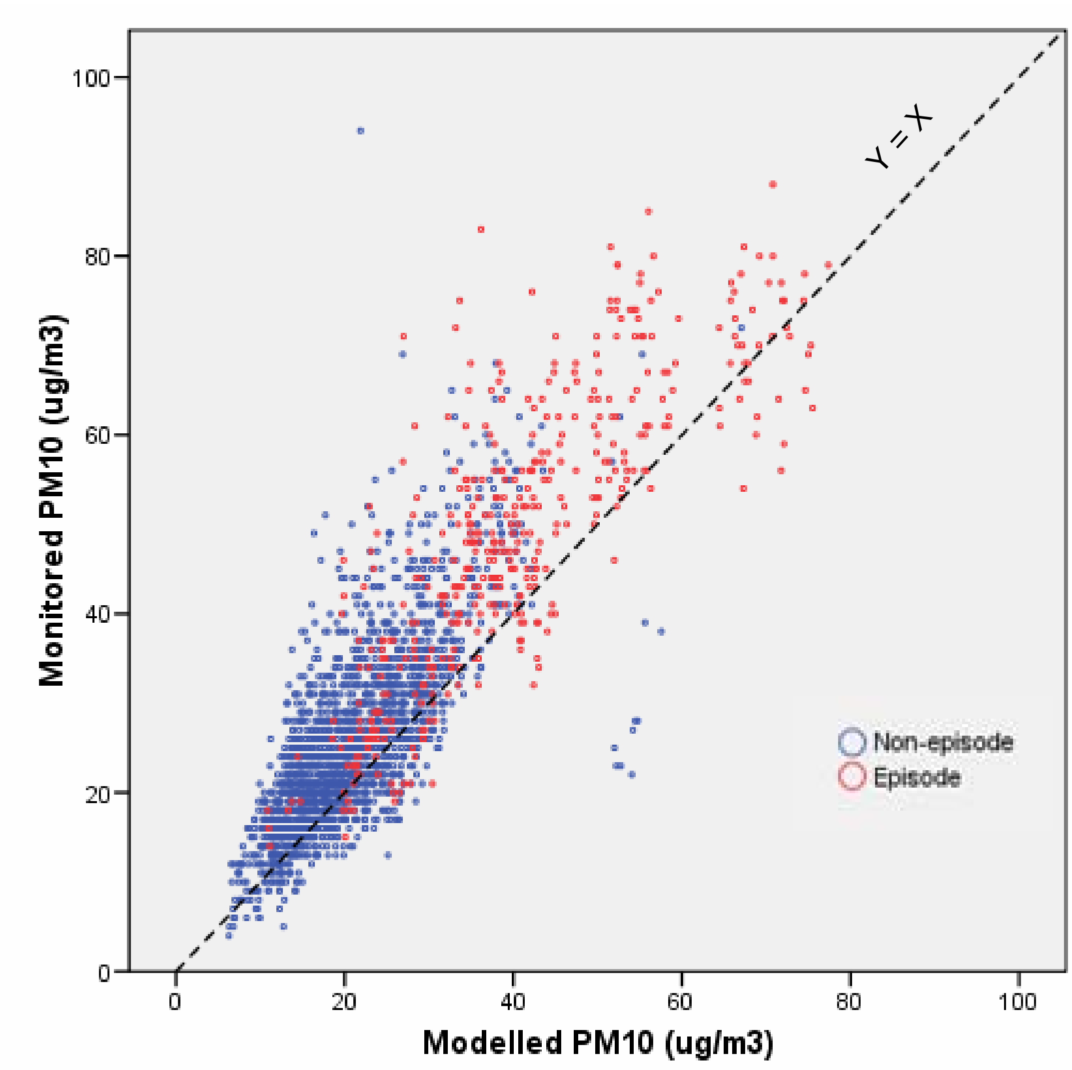
Modelled (NAME + ADMS) versus monitored PM_{10} across nine London monitoring sites for all ‘matched’ days in 2003



r	r2	SEE	FA2	N (days)
.663	.439	10.83	93%	250

FA2 = 81% during episode periods

Modelled (Rural monitored pollutants + ADMS) versus monitored PM_{10} across nine London monitoring sites for all 'matched' days in 2003

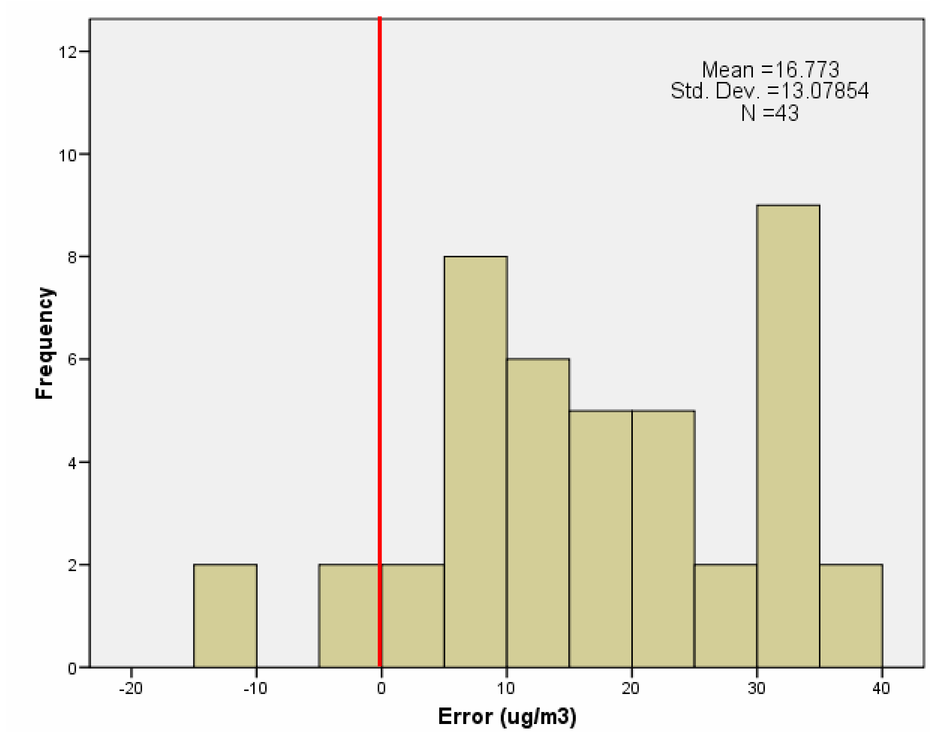


r	r2	SEE	FA2	N (days)
.863	.745	7.31	97%	250

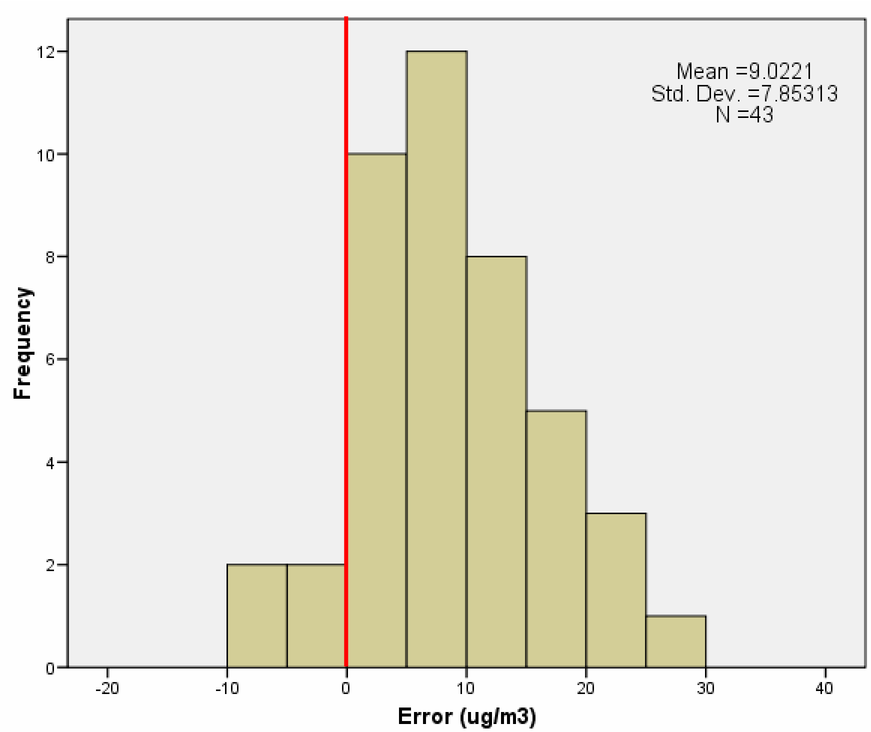
FA2 = 97% during episode periods

Comparison of performance between modelling approaches during April 2003 and August 2003 episode periods

Model including NAME estimates



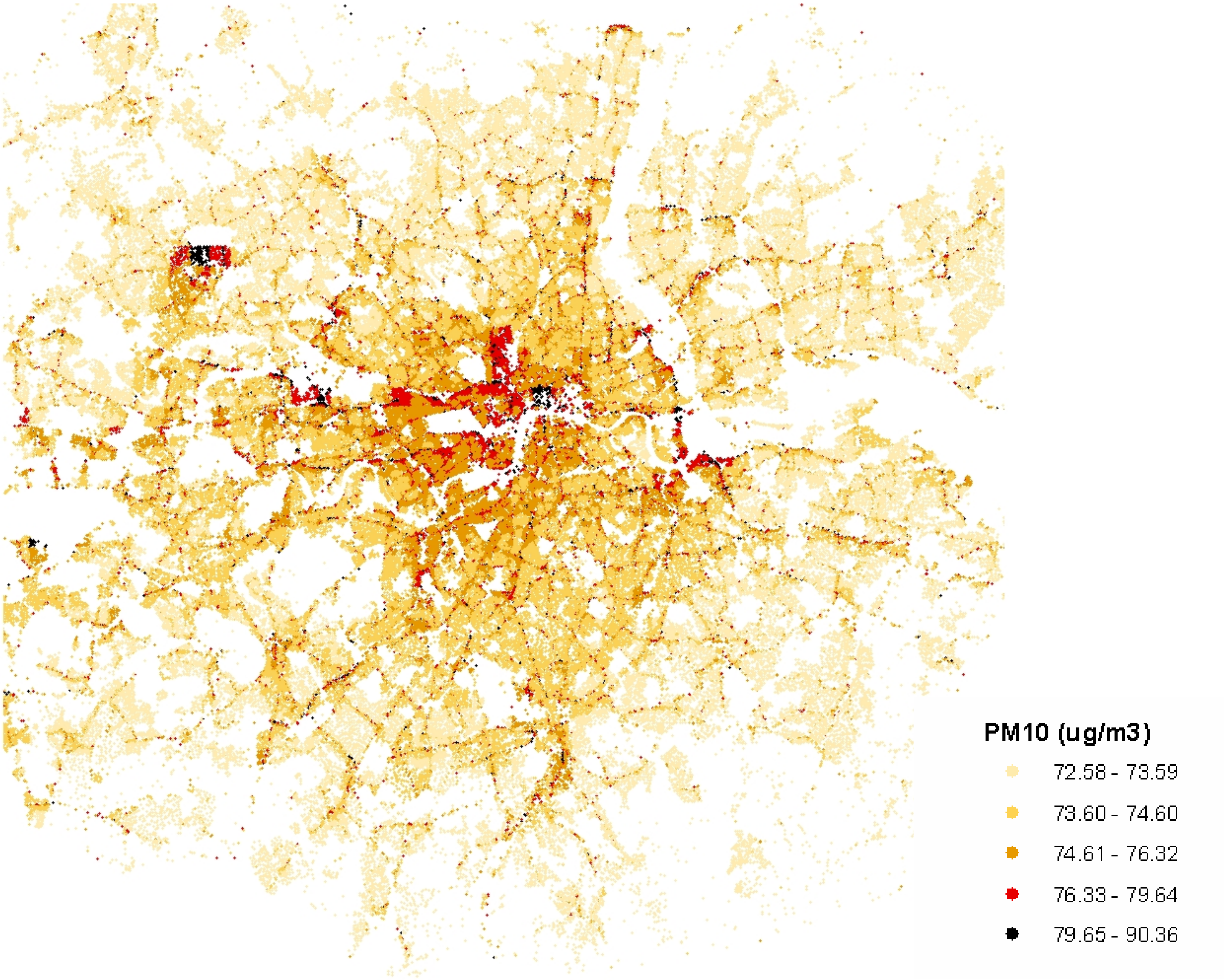
Model including data from rural monitoring



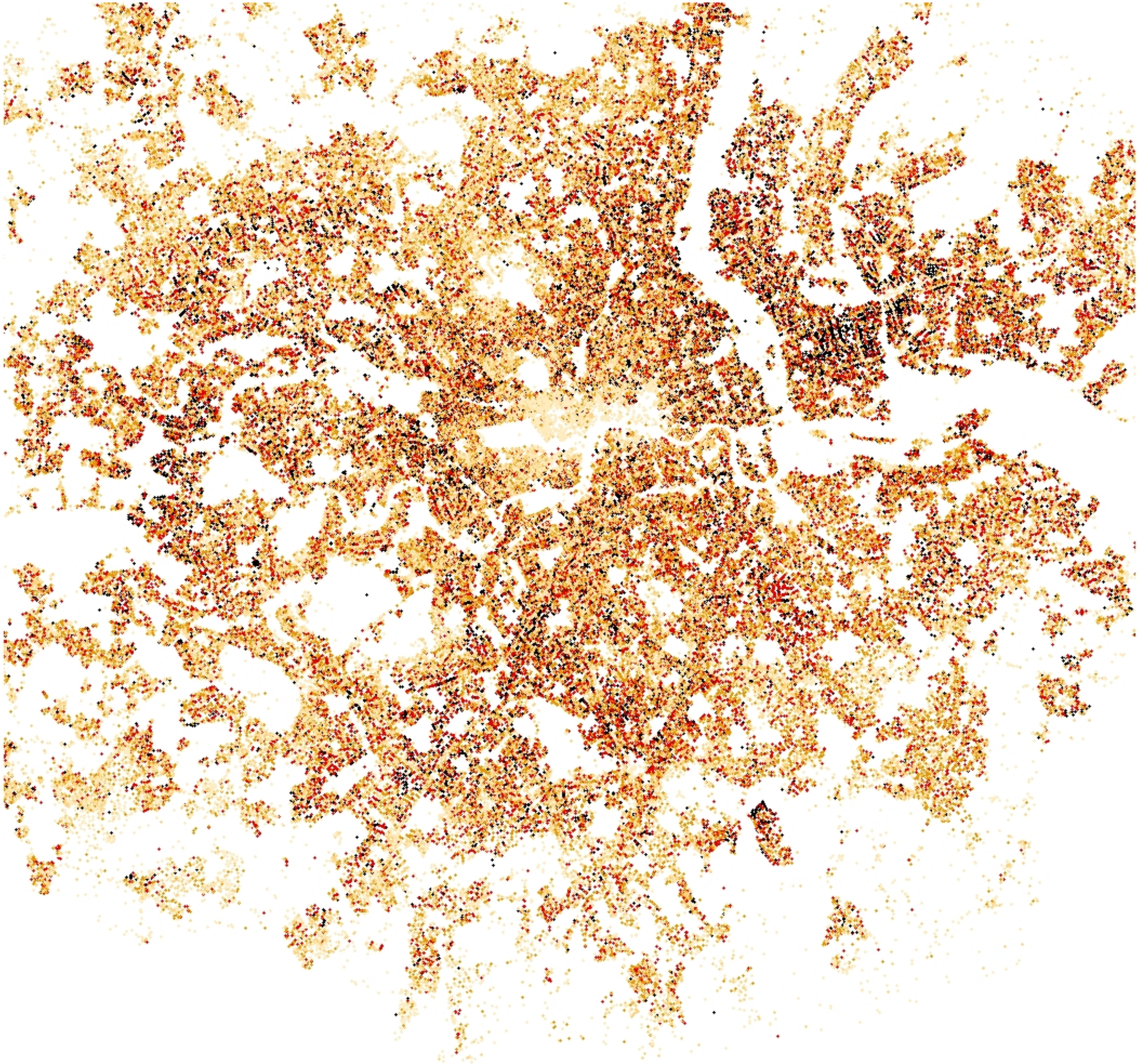
10 µg/m^3 error in PM_{10} corresponds to an error of 1.4* deaths / day in London

** Based on a baseline daily mortality rate of 166*

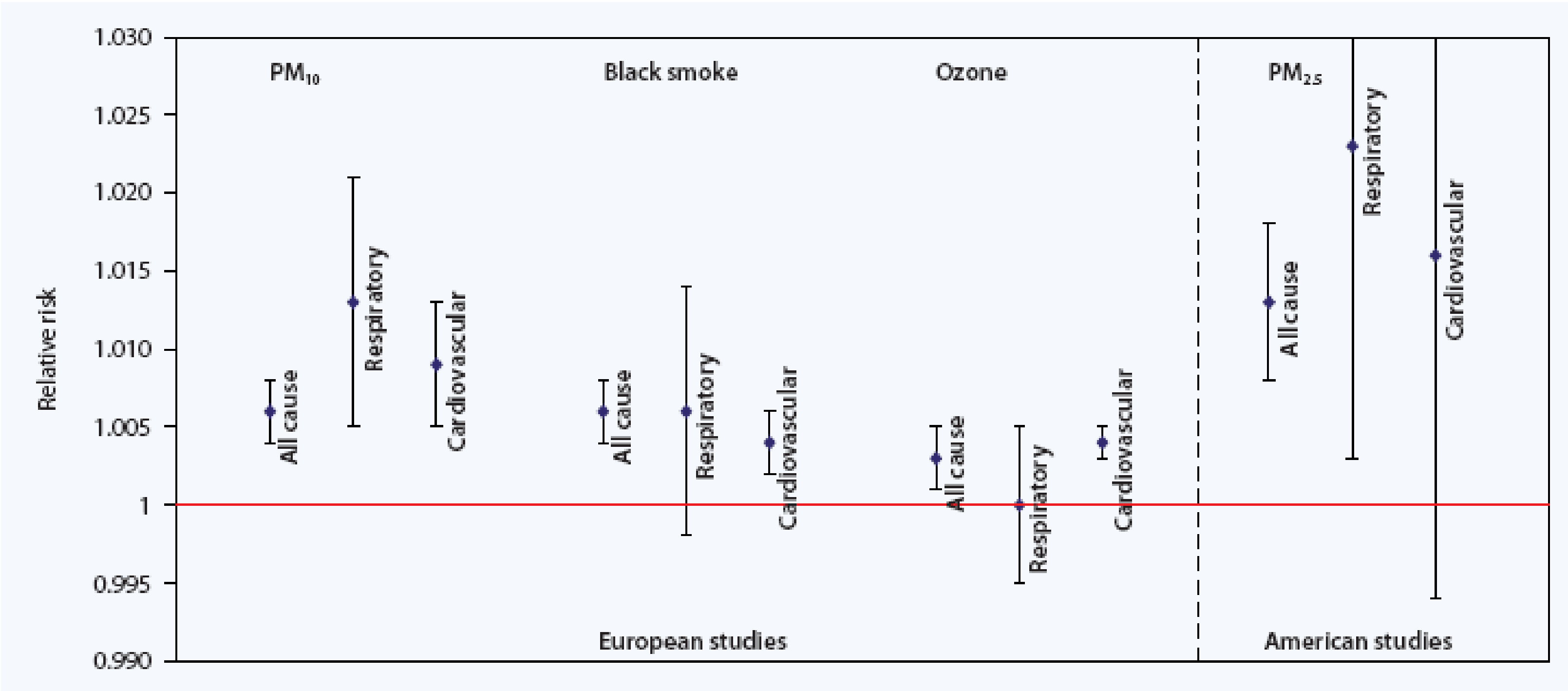
Modelled 'TOTAL' exposures by postcode in London: 8th August 2003



Population-weighted exposures by postcode in London: 8th August 2003



Meta-analysis of relative risks for mortality and different pollutants



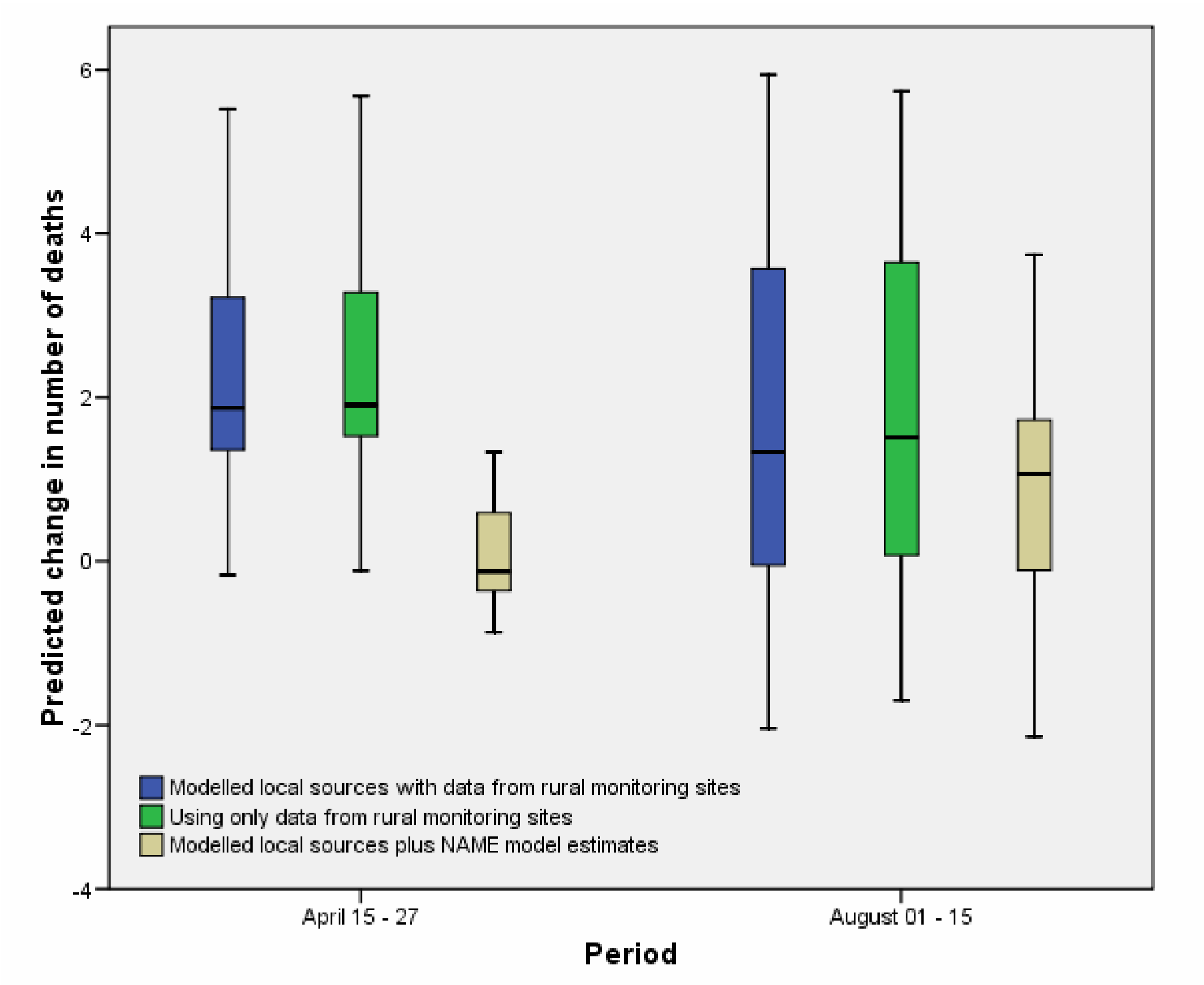
Source: WHO

Adopted estimates (increase in mortality per 10 µg/m^3 increase in PM_{10}) :
All-cause: 0.8%
Cardio / respiratory: 0.11%

Expected excess all-cause daily deaths in London (7.8 million resident population) for a 50 µg/m^3 increase in PM_{10} is **6.8**

Daily mortality estimates for London during two episode periods: comparison of different approaches

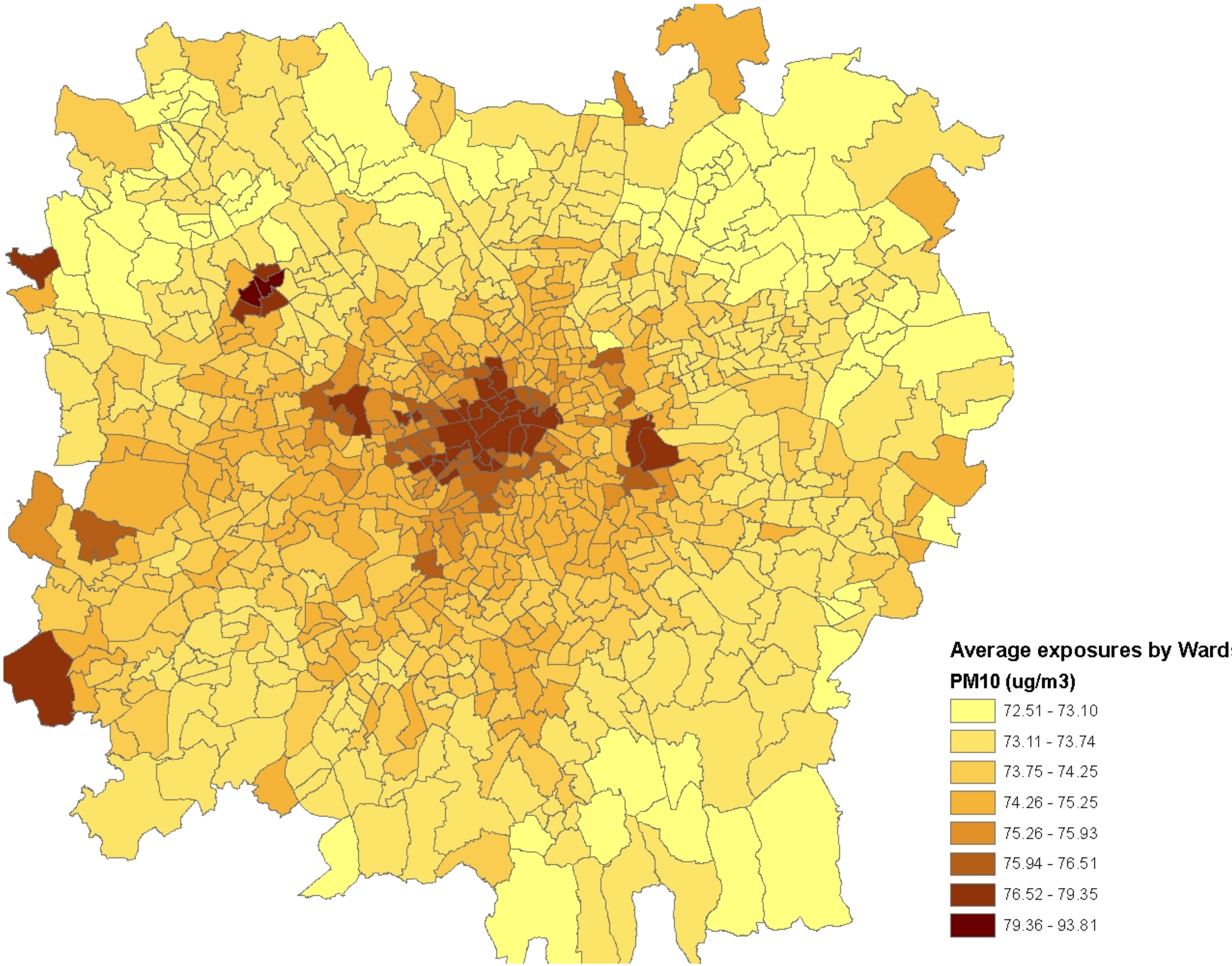
All-cause mortality



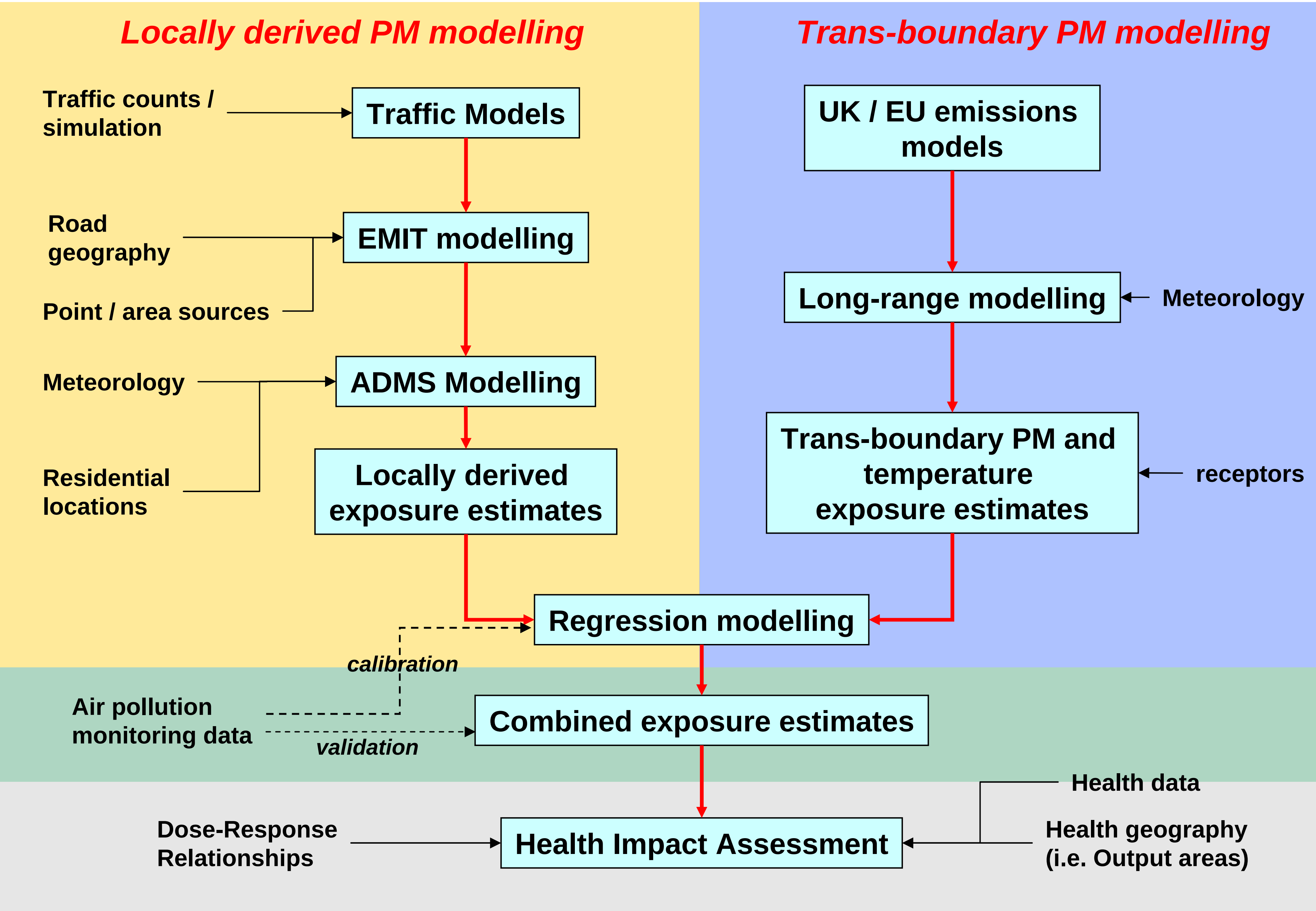
Sensitivity of risks estimates to using different models: April and August 2003

		April Count	April % of total	August Count	August % of total
All cause	*Actual*	*109*	*100*	*750*	*100*
	Modelled local + rural monitoring	30.1	28.0	28.2	3.8
	Modelled local + NAME	1.4	1.3	11.8	1.6
	Rural monitoring only	31.1	29.0	29.3	3.9
	Average of urban monitoring	33.3	30.5	31.1	4.1
Cardio / Respiratory are ‘first’ cause of death	*Actual*	*46*	*100*	*483*	*100*
	Modelled local + rural monitoring	16.4	35.7	14.3	3.0
	Modelled local + NAME	0.8	1.7	8.5	1.8
	Rural monitoring only	17.7	38.5	14.8	3.1
	Average of urban monitoring	18.8	41.0	15.9	3.2

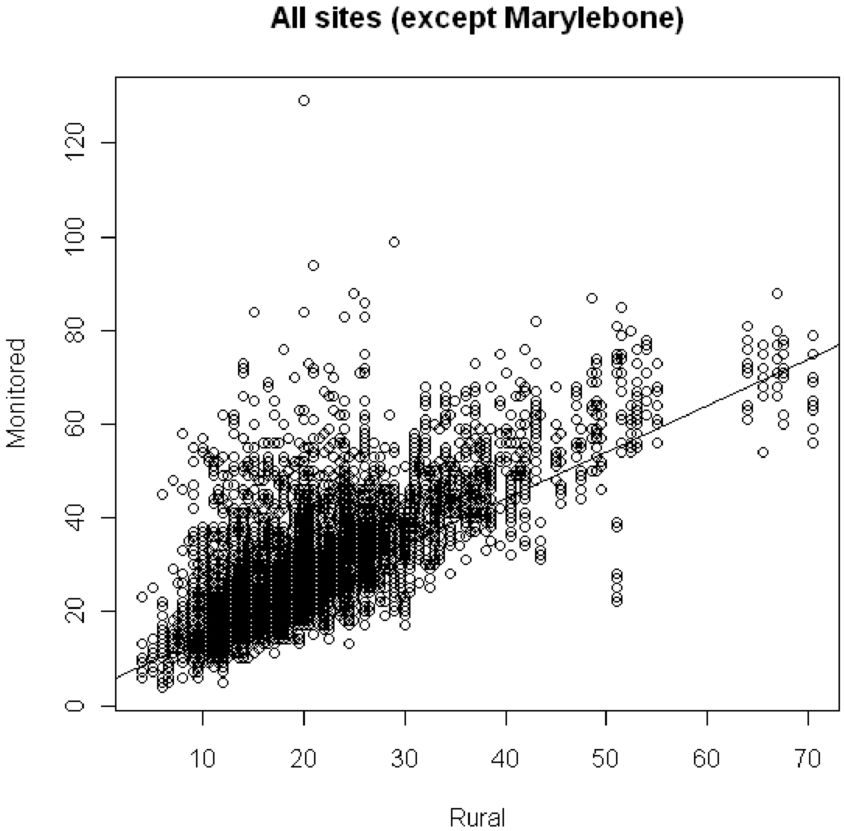
Average population-weighted exposures by Wards: 8th August 2003

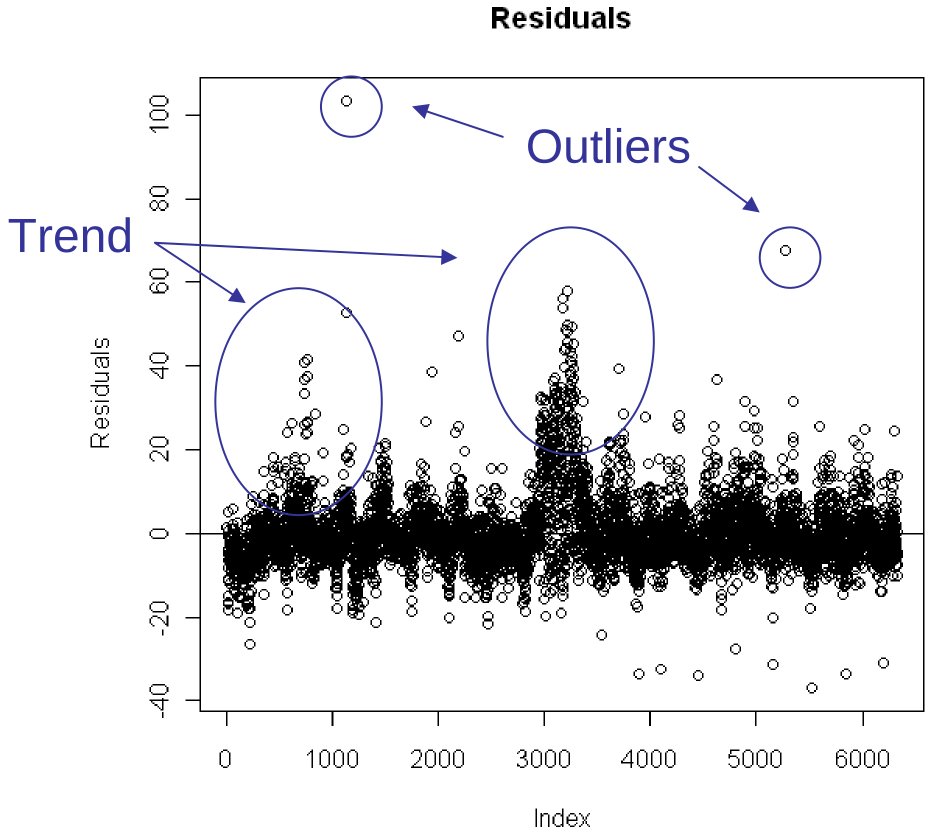


UK health study - modelling option ‘2’



One regression for all the sites?





Initial regression model formation

From the exploratory analysis:

- A trend in the residuals: a possible month/season effect should be taken into account
- A different behaviour in the monitoring sites regarding Local traffic and Local non-traffic sources
- Similar behaviour of the monitoring sites regarding data from rural monitoring

The model to start with might be the following:

$$PM10_i(t) = \alpha_i + \beta_{1i} X_{1i}(t) + \beta_{2i} X_{2i}(t) + \beta_3 X_{3i}(t) + \phi(t) + \varepsilon_i(t)$$

intercept

Coefficients for “LocalTraffic” and “LocalNonTraffic”, specific to each site

Coefficients for “RuralMon”, common to all the sites

Season (or month) effect

Error

Model [2]

To gain flexibility we might use a Generalized Additive model instead of a linear regression model:

$$PM10_i(t) = \alpha_i + f(X_{1i}(t)) + f(X_{2i}(t)) + f(X_{3i}(t)) + f(\phi(t)) + \varepsilon_i(t)$$

Non linear smoothed function of X

e.g. $$f(X_i(t)) = \beta_1 X_i(t) + \beta_2 X_i(t)^2 + \beta_3 X_i(t)^3 + \sum_j b_j (X_i(t) - k_j)_+^3$$

Non linear coefficients

Cubic piecewise regression

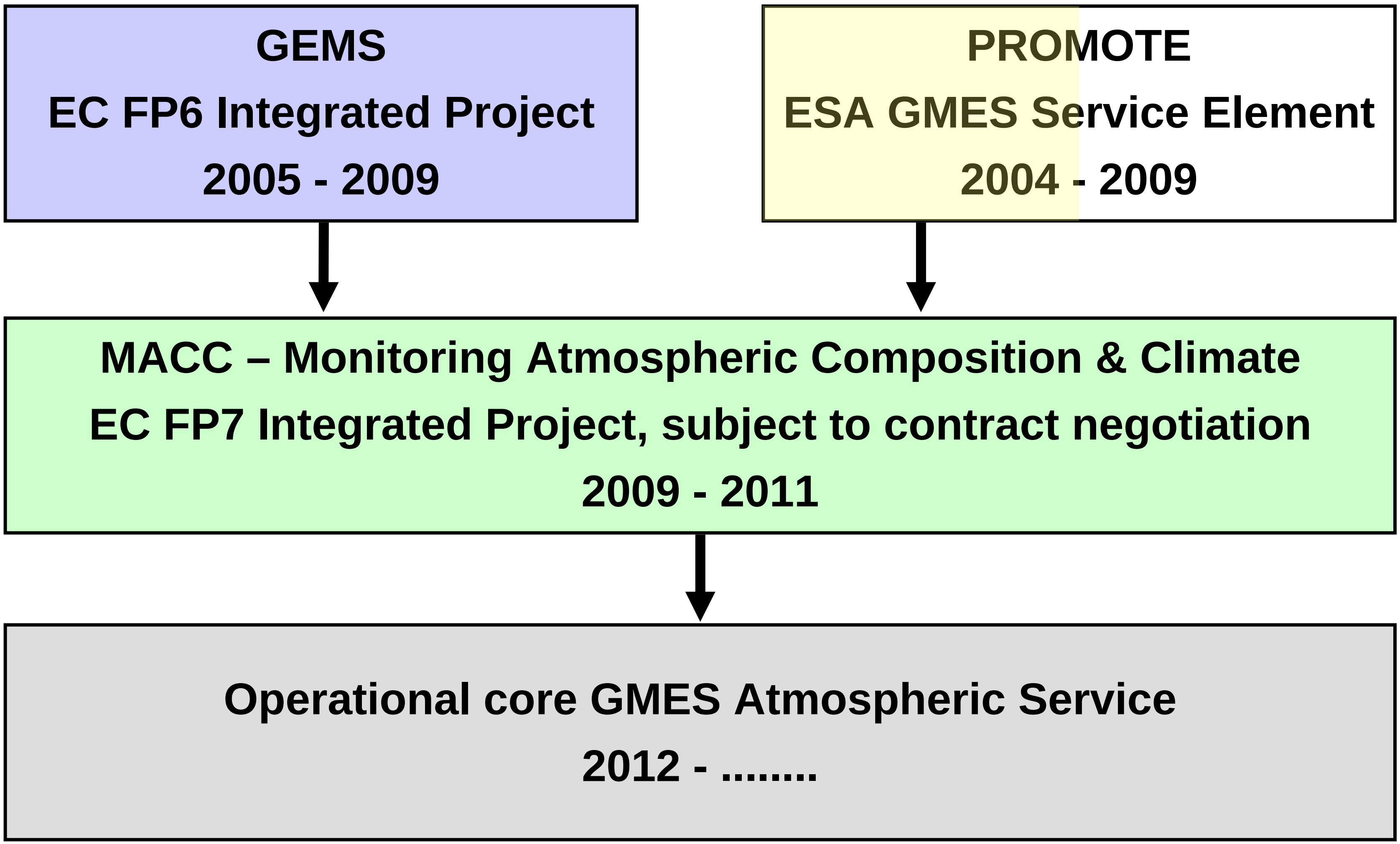
Again we can decide which parameters are "site specific" and which ones are common for all the sites

Early indications are that models may explain up to 80% of variation in monitored daily average PM_{10}

Next steps

- Determine optimal models using rural monitoring data
- Derive population-weighted exposure estimates across London and run area-based health risk assessment
- Repeat analysis incorporating GEMS ‘ensemble’ modelling (end summer 2008)
- Run epidemiological study to see if local and long-range PM_{10} have different dose-response functions (2008-2009)

Envisaged evolution of the core GMES Atmospheric Service



Regional Ensemble

- CAC
- EURAD
- REMO
- CHIMERE
- MOCAGE
- MM5-CAMx
- BOLCHEM
- EMEP
- SILAM
- MATCH
- NAME-AQ

3-day forecasts of surface NO_2 from 00UTC 11/01/2008