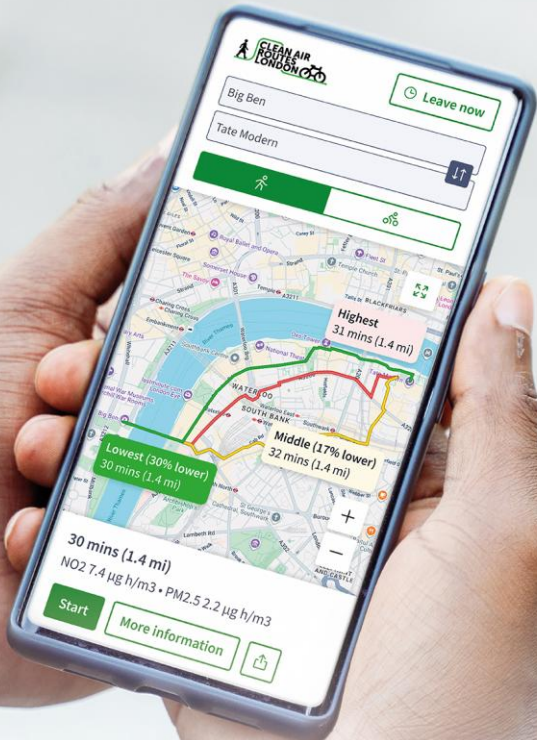


Plan your journey with
cleanairroutes.org



Find the cleanest route
for your journey with
Clean Air Routes London
and help protect yourself
from air pollution.



Scan me with
a smartphone!

cleanairroutes.org

City of Westminster



City of
Westminster

Challenges of applying data fusion to AQ forecasting

FAIRMODE Technical Meeting 2026
WG6 sensors and data fusion

Jenny Stocker

Copenhagen 30th September 2026

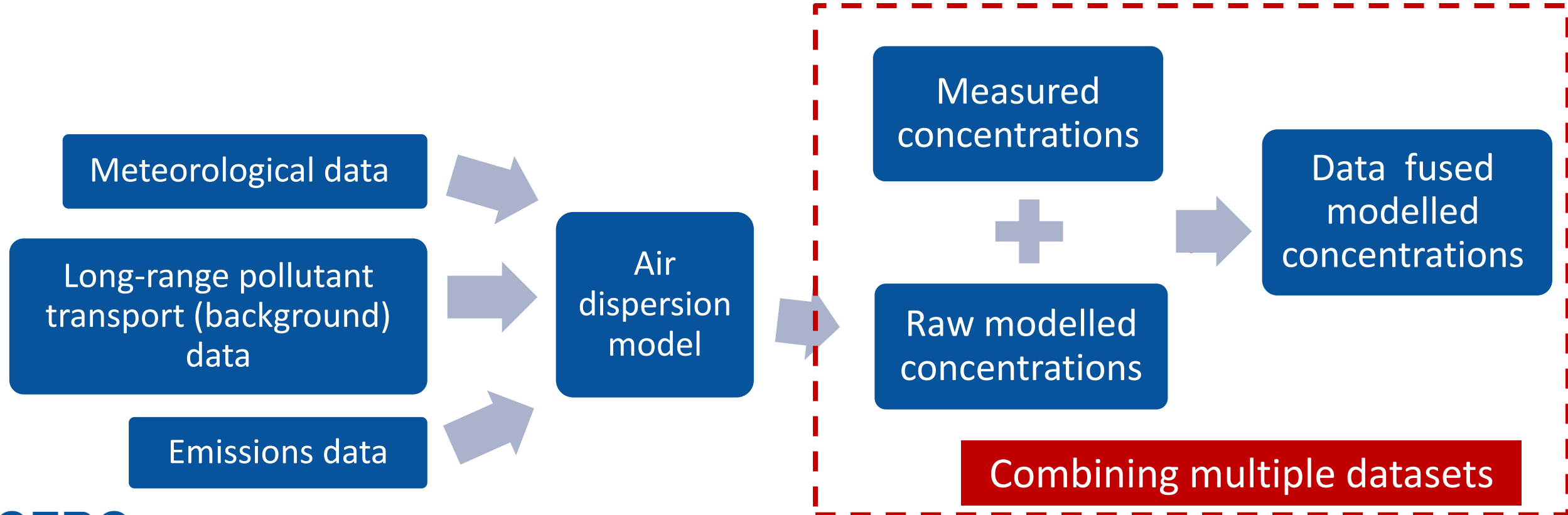
CERC

Cambridge Environmental Research Consultants
Environmental Software and Services

Challenge

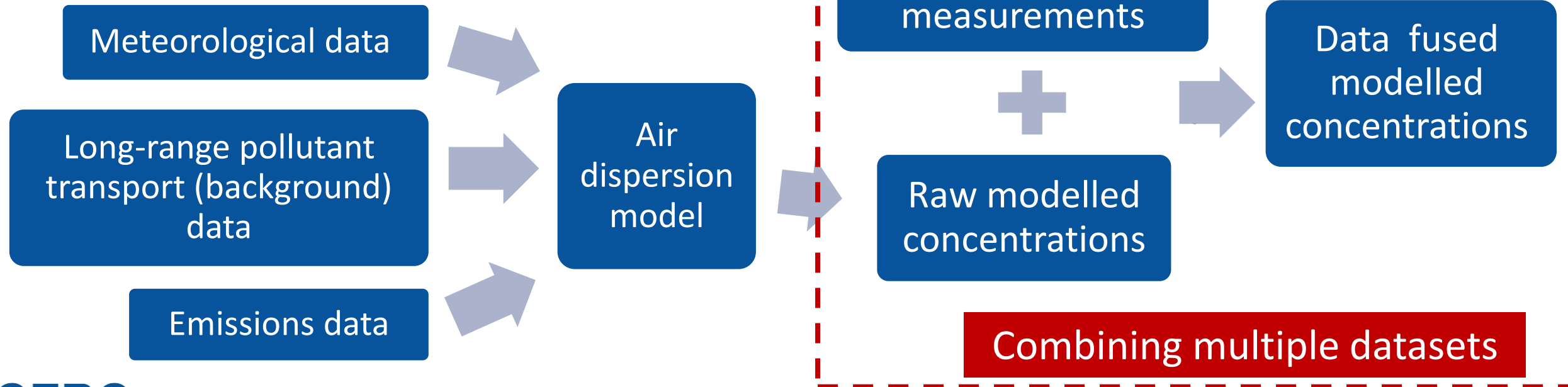
- ‘Standard’ data fusion involves using AQ measurements to improve modelled concentrations for historical periods
- For AQ forecasting systems, measurements aren’t available

‘Standard’ data fusion

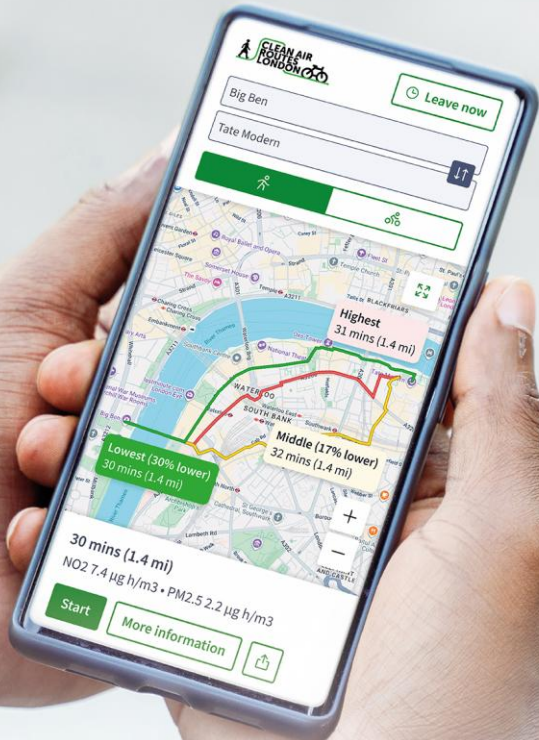


Challenge

- ‘Standard’ data fusion involves using AQ measurements to improve modelled concentrations for historical periods
- For AQ forecasting systems, measurements aren’t available
- How much can we learn from historical model performance to improve forecasts?



Plan your journey with
cleanairroutes.org



Find the cleanest route
for your journey with
Clean Air Routes London
and help protect yourself
from air pollution.



Scan me with
a smartphone!

 cleanairroutes.org

 City of Westminster



Example application

City of Westminster tool to reduce people's exposure to air pollution as they walk and cycle through London

- Derived from forecast air quality datasets generated within CERC's *airTEXT* air quality forecasting system for London and the southeast of England
- Google Routes data / data transfer via APIs
- Hourly forecasts for up to 3 days ahead
- Composite air pollution impact rating for each route calculated from PM_{2.5} and NO₂ data
- ***Incorporates hyperlocal AQ monitoring data***
- Tool use can lead to more healthy transport choices and, in some cases, may prevent people from making a journey.

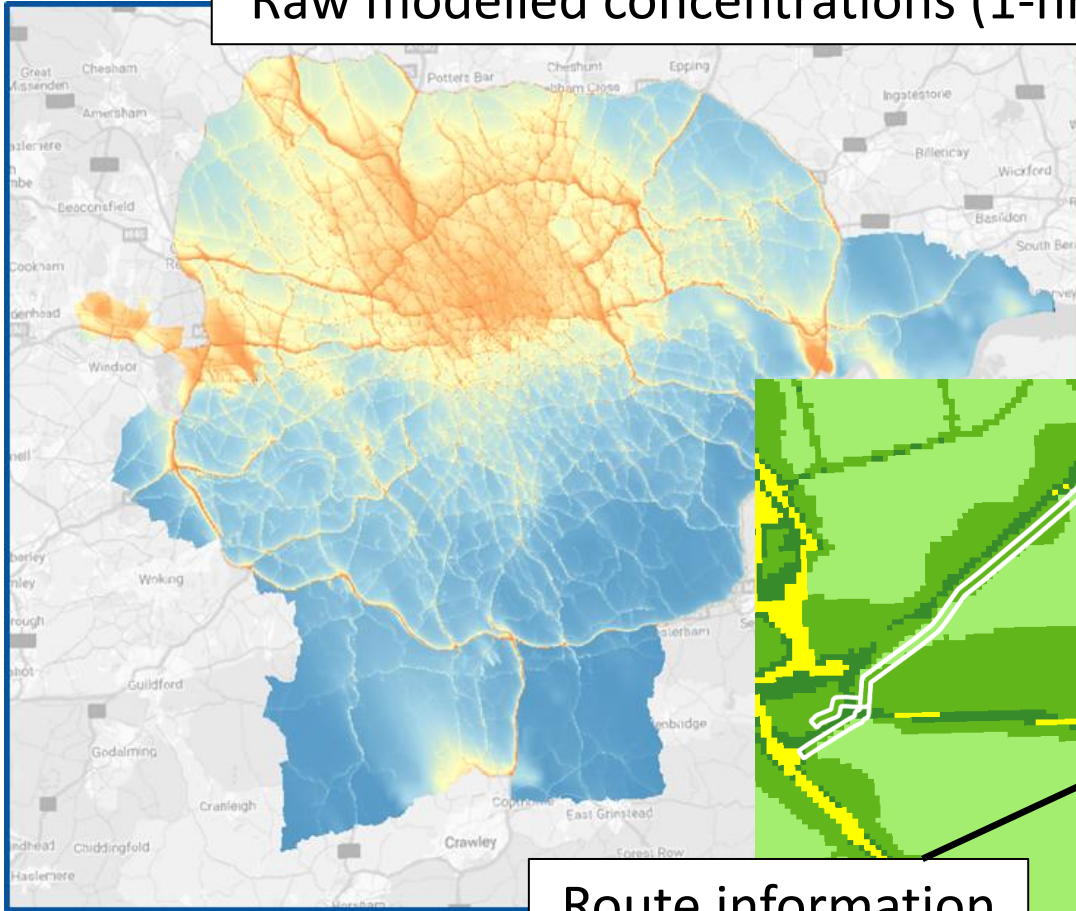
<https://cleanairroutes.org/>



City of
Westminster

System data

Raw modelled concentrations (1-hr)



Route information



Air quality measurements from reference monitor and small sensor networks



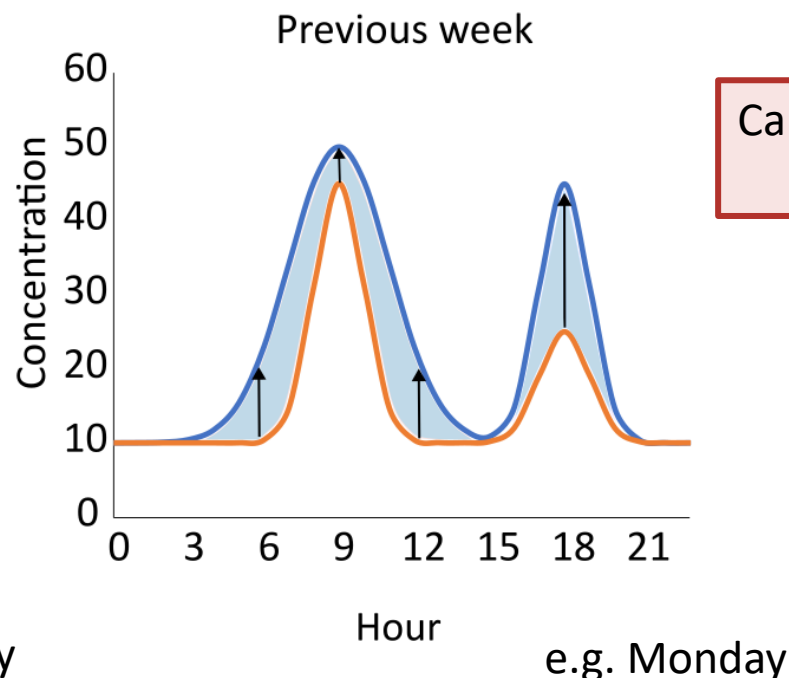
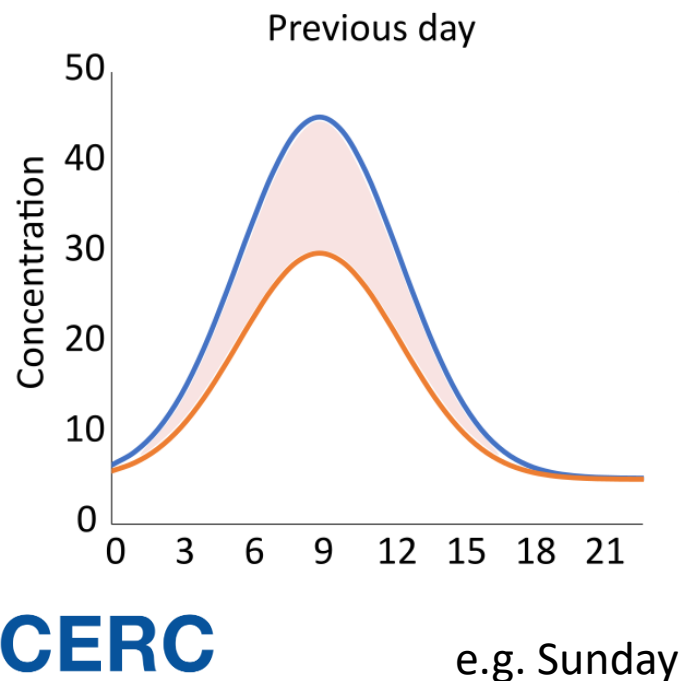
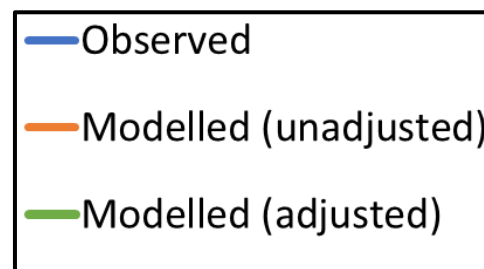
- All measurement data from the previous day available by ~ 04:00, for input to the 'data fusion' calculations
- Measurement data for the previous week used as input to the data fusion calculations

'Forecast' data fusion methodology

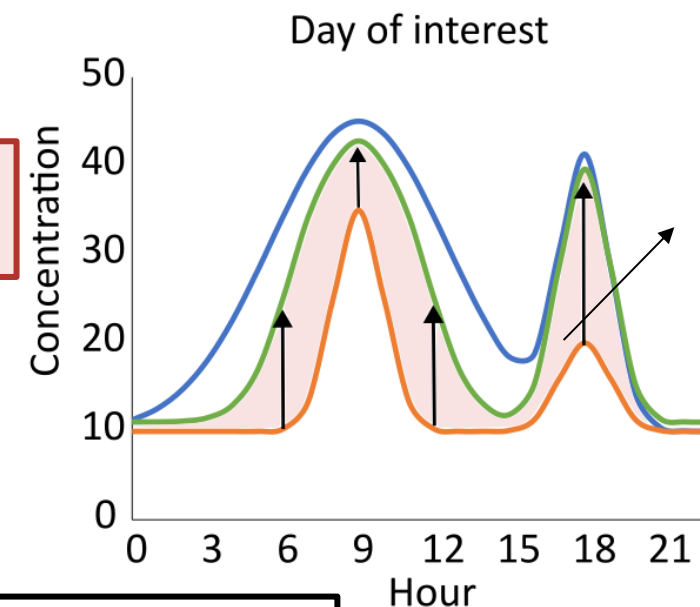
- Use historic differences between measured and modelled data to estimate the likely difference for current or future hours.
- Can't simply apply yesterday's hourly bias due to weekday-weekend differences

- Approach:

- Calculate concentration bias correction from previous day
- Apply temporal variation from last week



Calculate overall bias

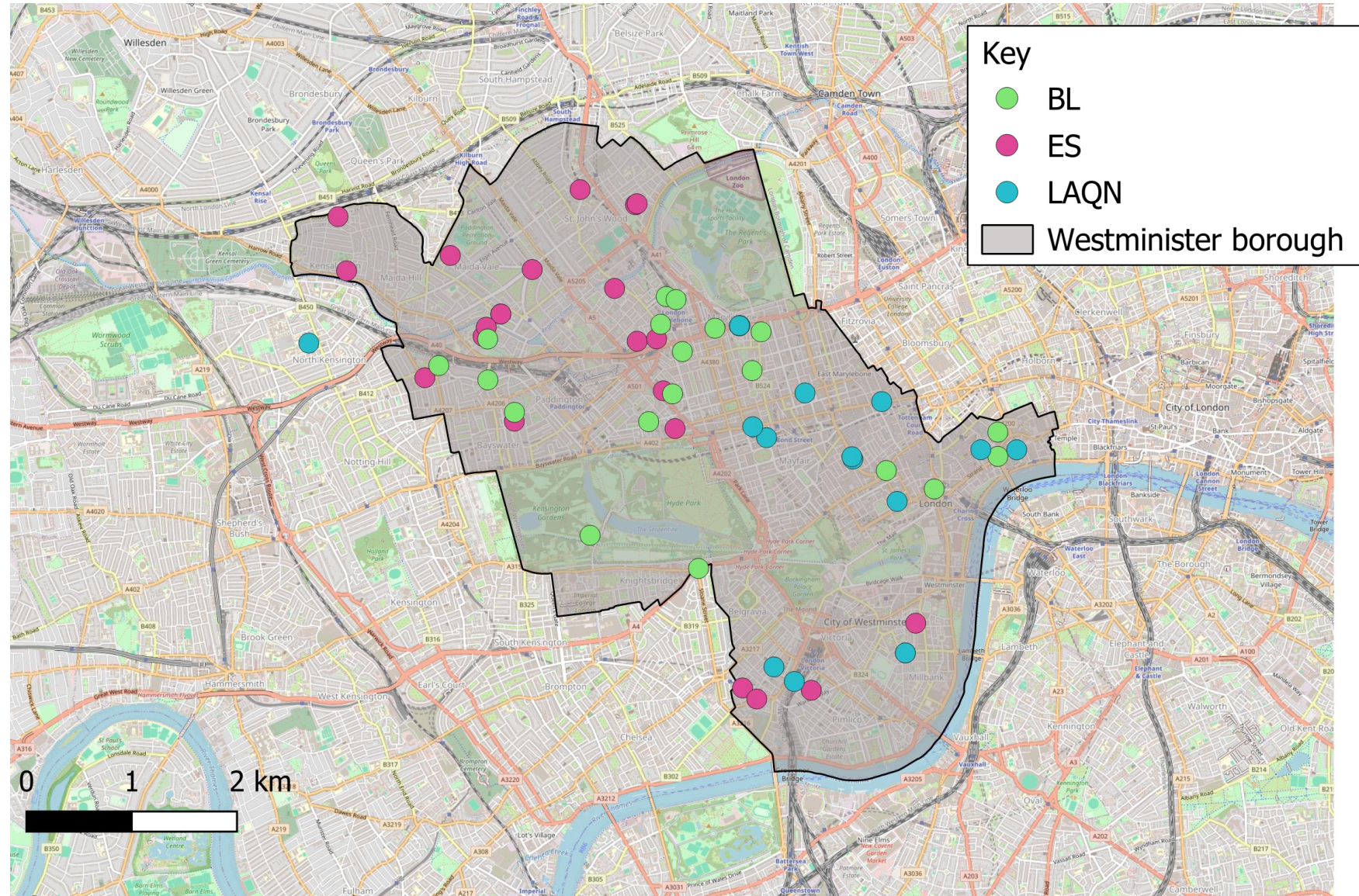


Calculate proportionate split of bias

Data Fusion 2024 case study: Westminster monitors

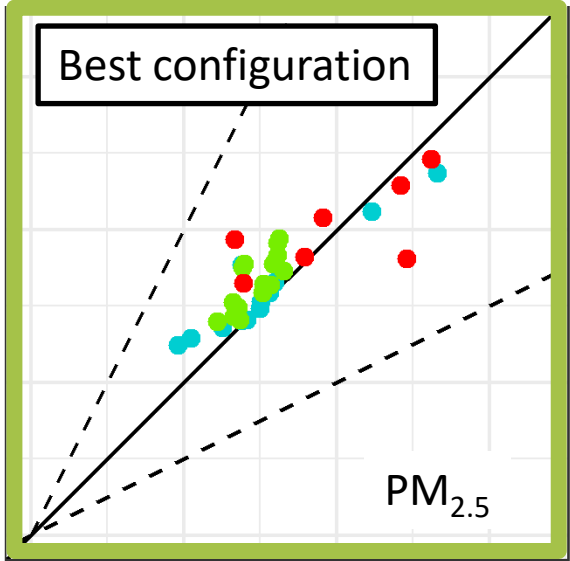
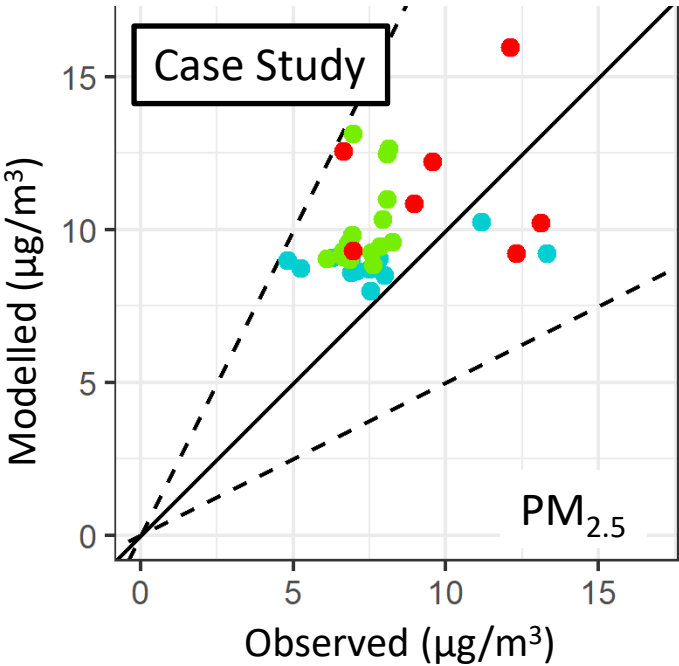
Monitoring networks active in 2024:

- **Breathe London (BL)**
 - Clarity sensors
 - 22 sensors
- **Earthsense (ES)**
 - Zephyr sensors
 - 22 sensors
- **London Air Quality Network (LAQN)**
 - reference monitor
 - 20 monitors



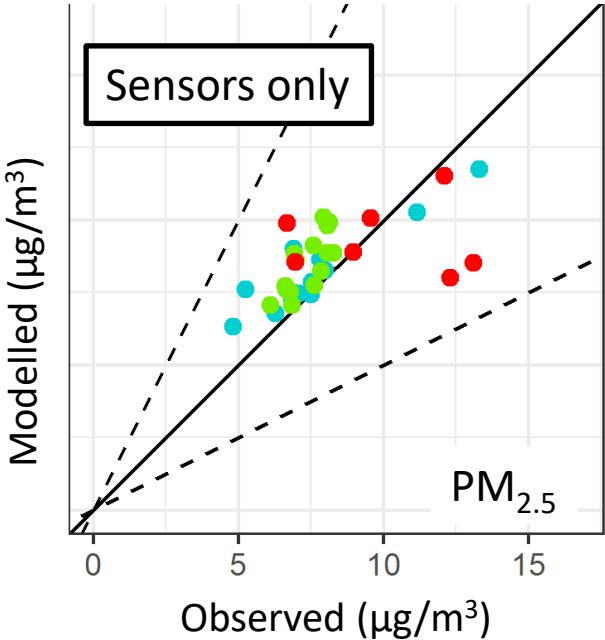
Data fusion 2024 case study: testing

Parameter	Variables	Testing
Network	- LAQN - ES - BL	Include, exclude for particular pollutants
Location	- Roadside - Background	Include / exclude roadside
Weighting by type	- Reference monitor weighting - Small sensor weighting	Weighting (reference:sensor) equal (1:1) / unequal (1:0.5)
Adjustment limits	- Min limit - Max limits	Limit adjustment of model more / less



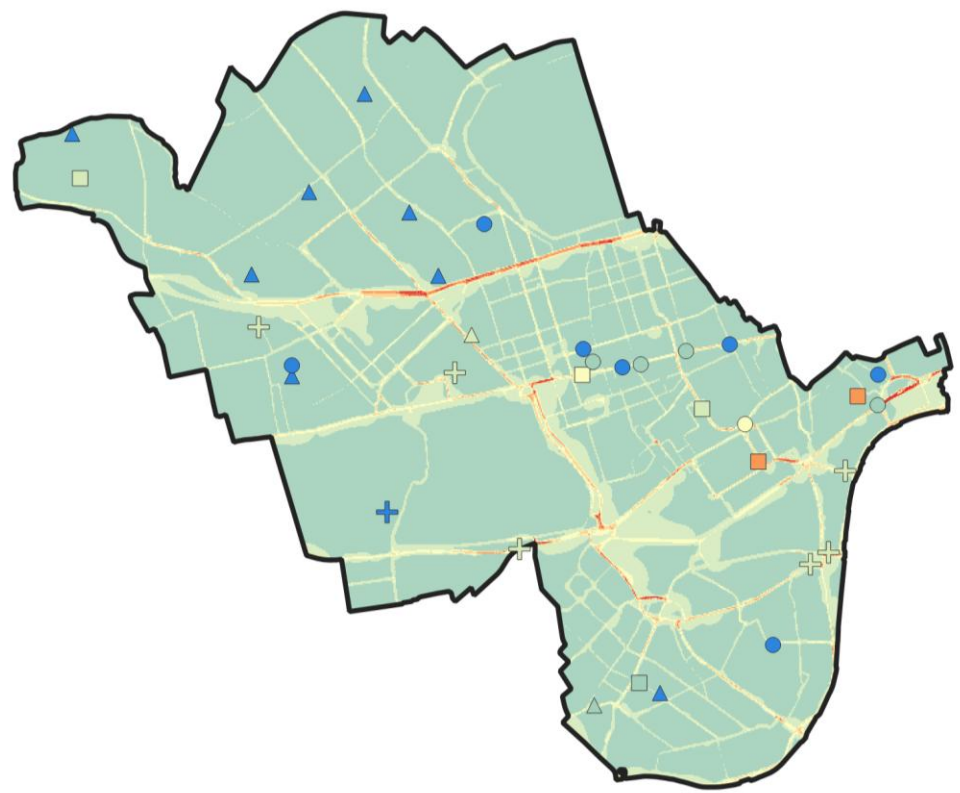
Conclusions:

- Best to use all available data
- Better to limit how much model can be adjusted

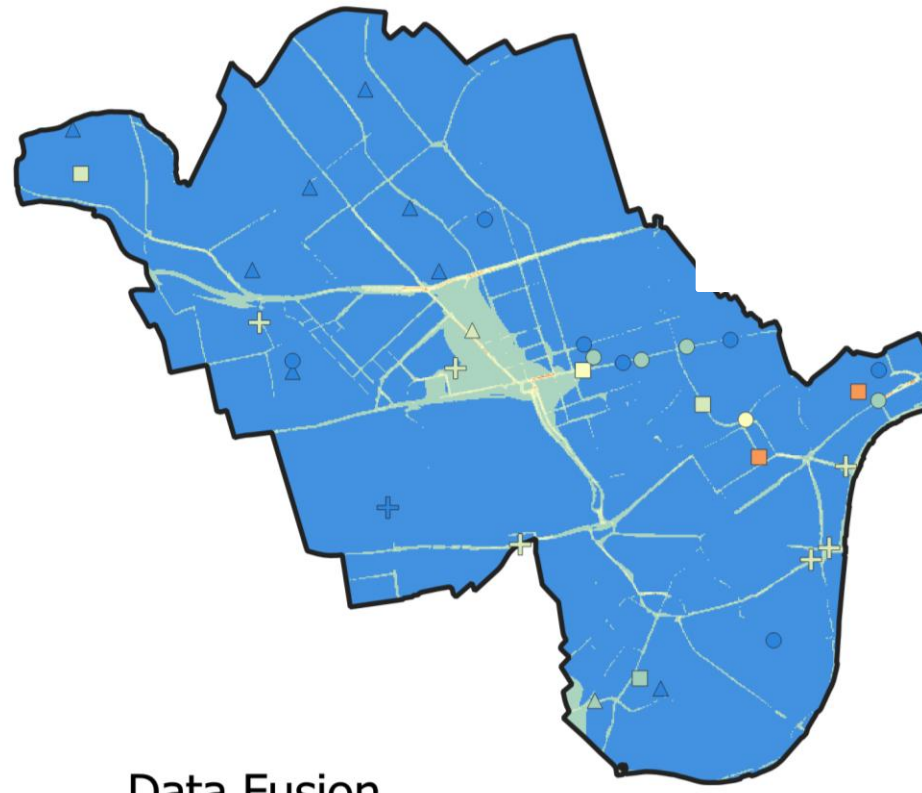


Data Fusion 2025 results: Westminster PM_{2.5}

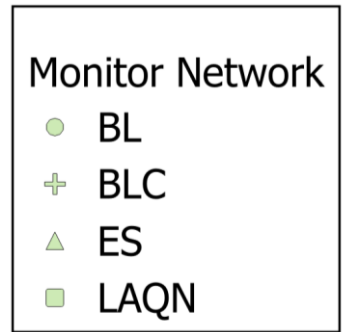
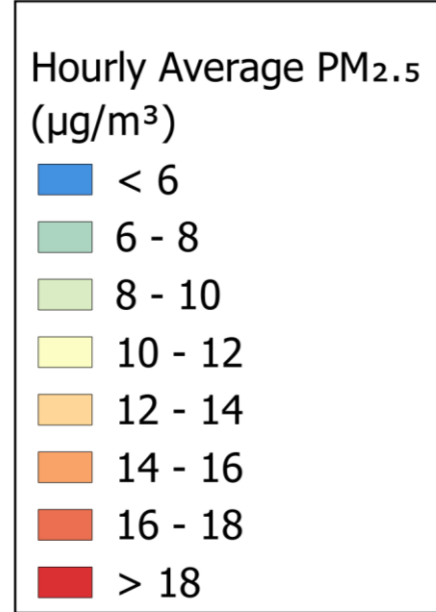
Data Fusion results selected for the
9th October 2025 at 11:00



Raw Model Output



Data Fusion

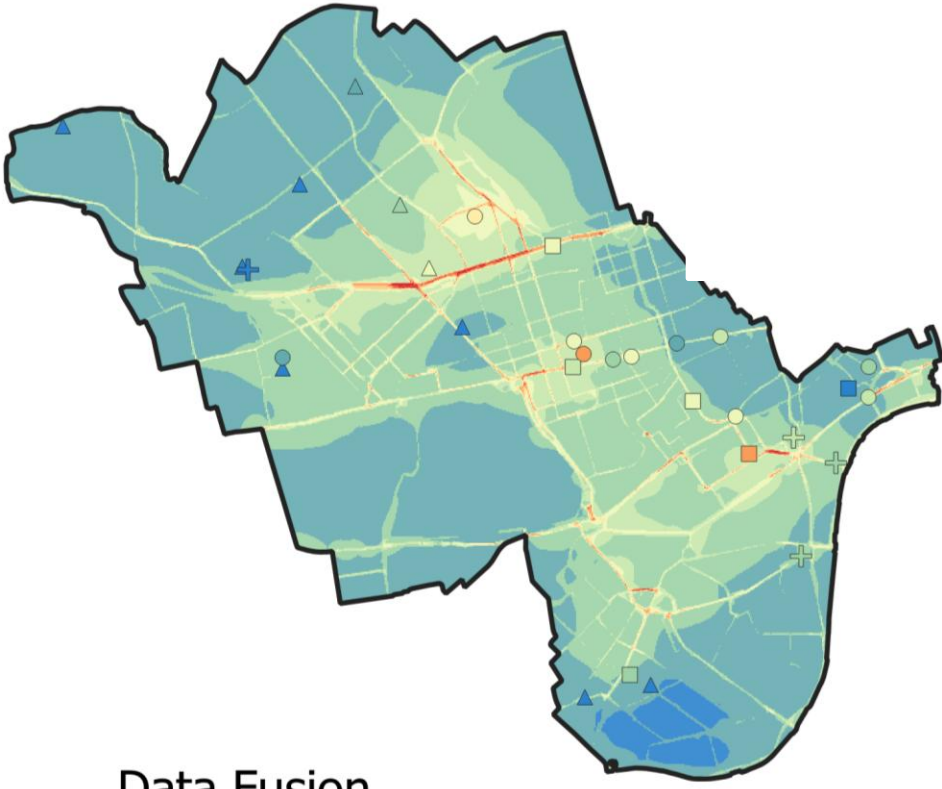


Data Fusion 2025 results: Westminster NO₂

Data Fusion results selected for the
9th October 2025 at 11:00

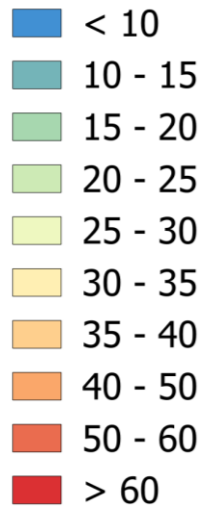


Raw Model Output

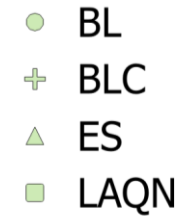


Data Fusion

Hourly Average NO₂
(µg/m³)

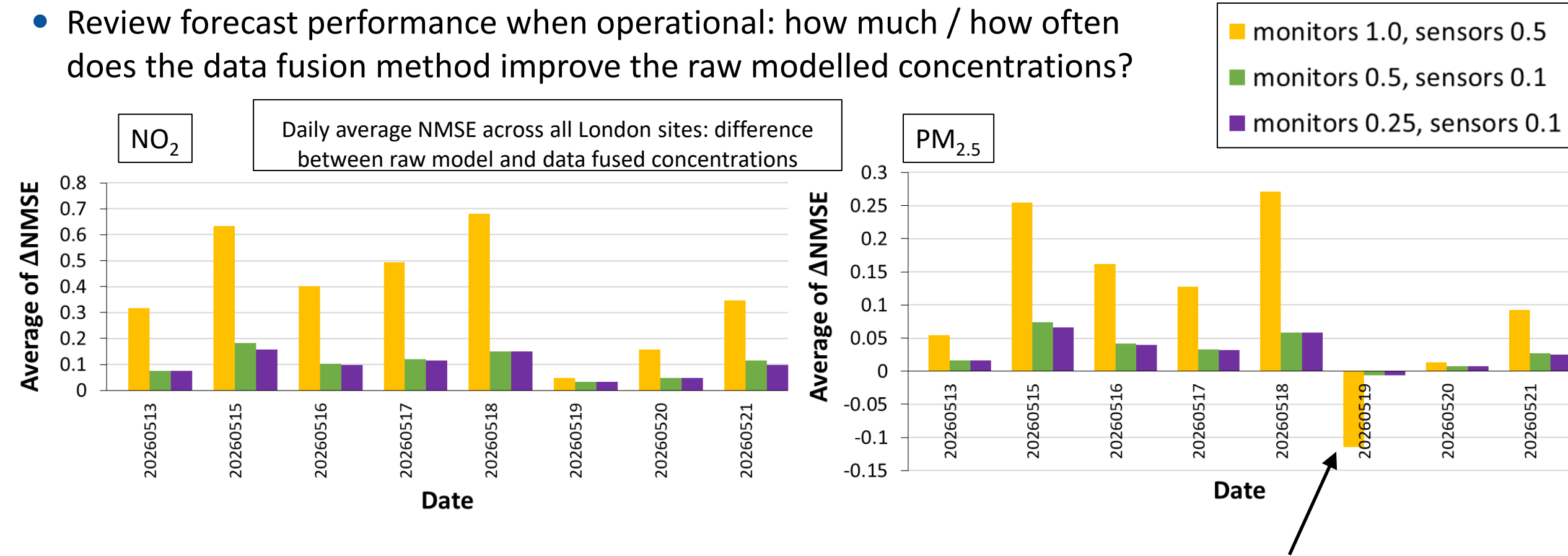


Monitor Network



Operational data fusion performance

- Review forecast performance when operational: how much / how often does the data fusion method improve the raw modelled concentrations?



- Operationally, there's a choice between improving most hours a lot and make a few quite a lot worse, and adjusting less (so the improvements are reduced, but so are the adjustments that make results worse)

For discussion

- Common requirement to improve air quality forecasts using historical measurement data:
 - Methods usually apply spatially homogeneous biases
 - Possibly less common to apply spatial data fusion?
- AQ forecast metrics are daily (e.g. daily average, max daily). More challenging to get good performance for hourly forecasts
- Data fusion methodology motivated by experience of model performance e.g. 'typical' flow hourly variations used as input to the model are only broadly representative on a road by road basis
- Methodology could be **pollutant dependent** because some pollutants are strongly influenced by long-range pollutant transport ($\text{PM}_{2.5}$), others by local emissions (NO_2)
- **Machine learning approaches** incorporating additional data sources (e.g. meteorological data, holiday periods) may improve 'forecast' data fusion outcomes

Questions?
Comments?

jenny.stocker@cerc.co.uk