Low-cost sensors for the measurement of atmospheric composition.

Prof. Alastair C Lewis

Wolfson Atmospheric Chemistry Laboratories, University of York, Heslington, York, UK.

ally.lewis@york.ac.uk
@AtmosChemYork

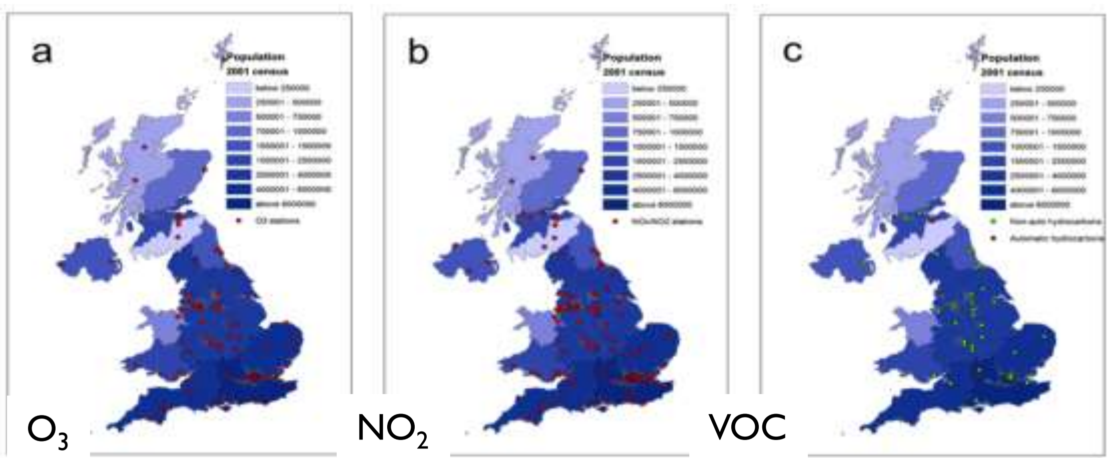
Based on:
Lewis, Peltier and von Schneidmesser, WMO, 2018
Lewis *et al*., *Faraday Discussions*. **189**, 85-103, 2016
Smith *et al. Faraday Discussions*. **200, 621-637,** 2017.
Pang *et al. Sensor and Actuators B*, **240**, 753-766, 2017.
Lewis and Edwards, *Nature*, **535**, 29-31, 2016

UNIVERSITY of York

WACL Wolfson Atmospheric Chemistry Laboratories

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Sensors – improving on the observational gaps?



- UK as an example: approximately 140 static measurement sites.
- One static monitoring site per 250,000 population

- Many new commerical products aimed at personal exposure monitoring.
- Professional applications include exposure / health science
- Amateur user applications include behaviour change, public interest, campaigning groups.

UNIVERSITY *of York*

Rapidly changing commercial landscape

The challenge of component and system diversity

- The rate of change of sub-components can be rapid
- Past studies don't necessarily represent current capability
- The data quality from one sensor may differ from a network of sensors
- *There is far more to 'cost' than just buying the equipment*

Metal oxide
~ £5
~ since 1960s

Electrochemical / voltammetric
~ £50
~ since 1980's

Photochemical
~ £200
~ since 1990's

Micro-optical
> £100
~ since 2000's

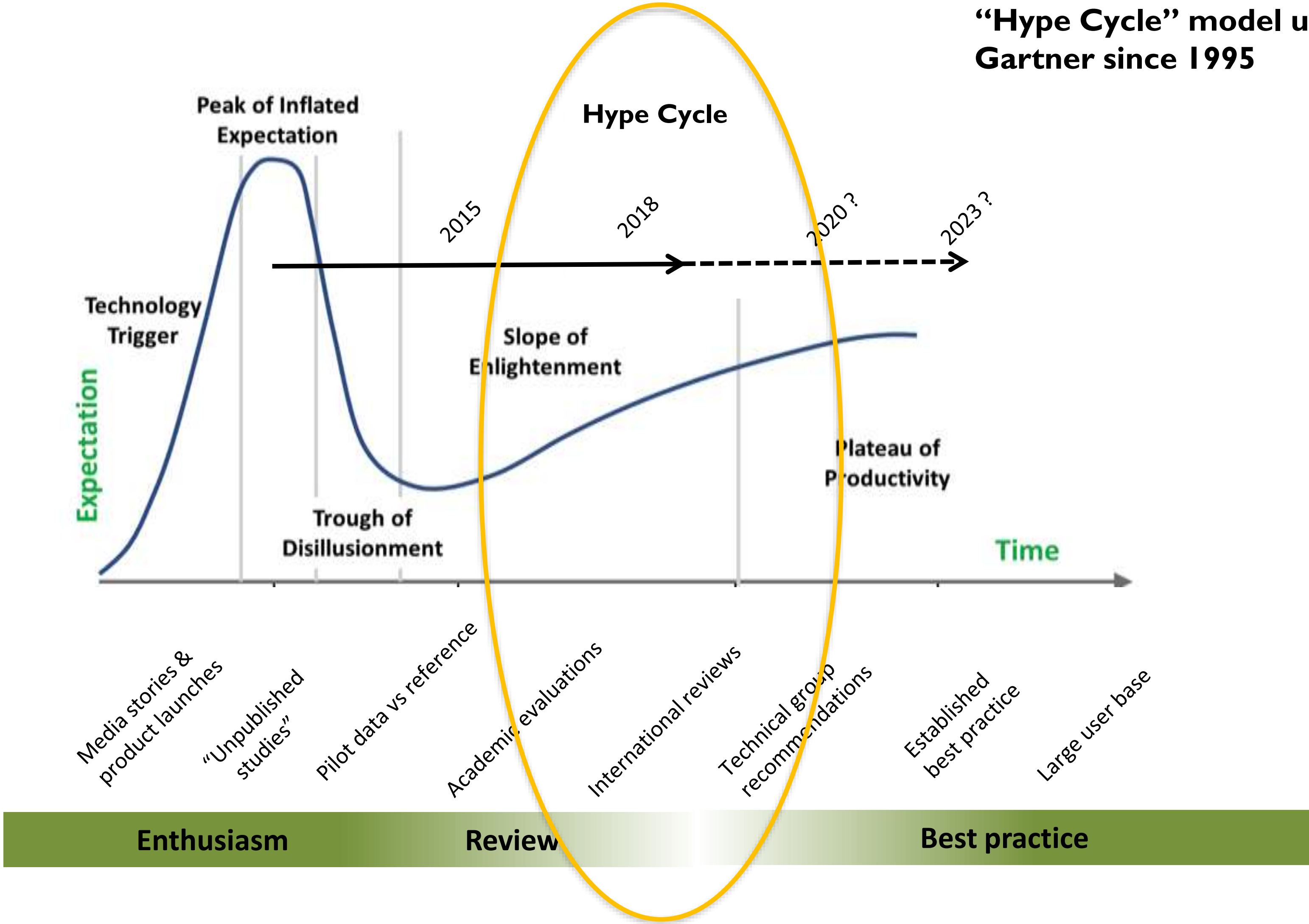
Sensor → Micro-electro-mechanical (MEMS) device

sensor
/ˈsɛnsə/
noun
noun: **sensor**; plural noun: **sensors**
a device which detects or measures a physical property and records, indicates, or otherwise responds to it.

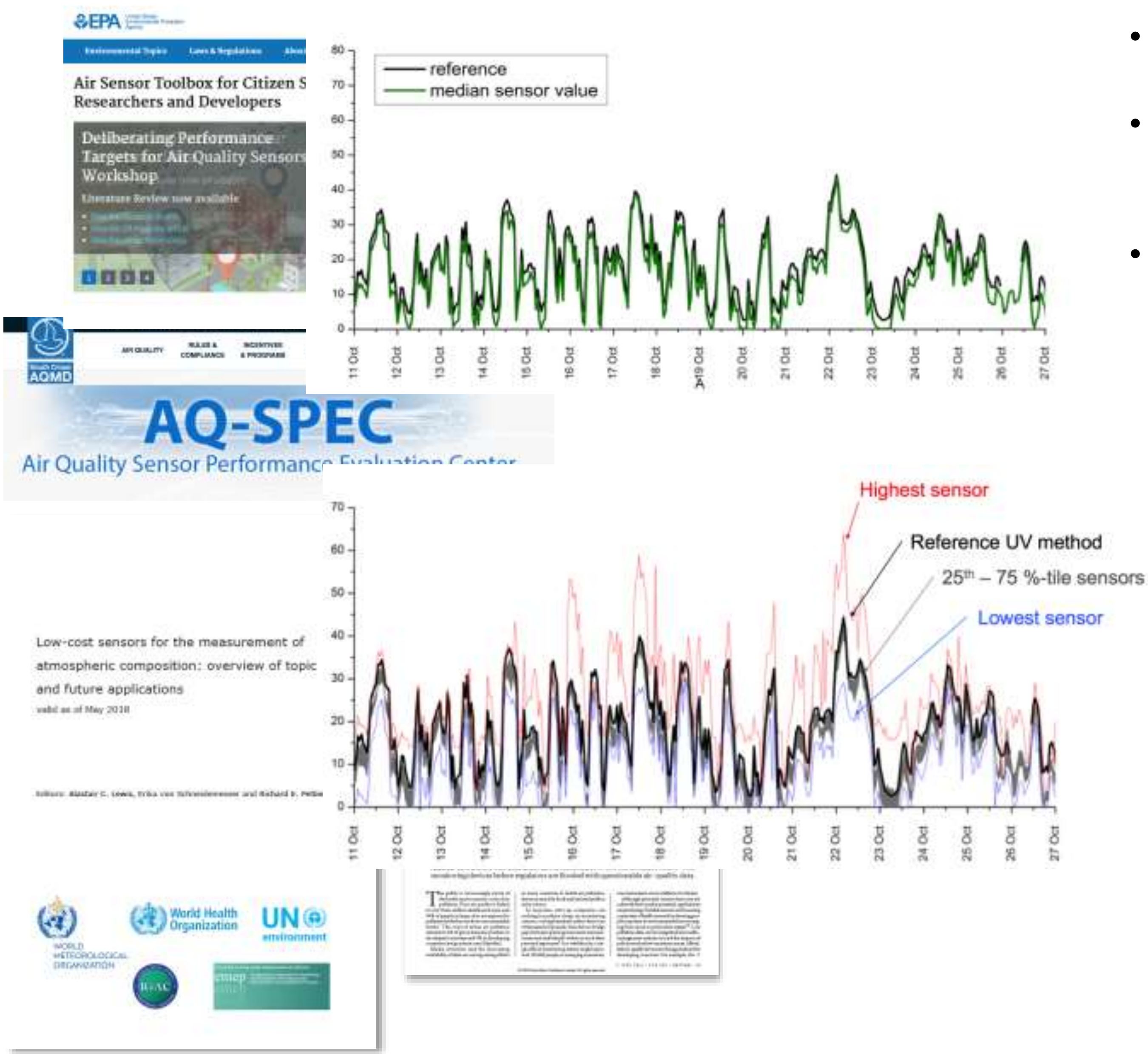
Sensors have many other positive attributes beyond unit cost

Where are we now?

"Hype Cycle" model used by Gartner since 1995



2014-2018 –increase in "evaluation and advice"



- Side by side comparison has been the main metric.
- Many positive examples of correlations next to reference monitors.
- Increasing use of training data and machine learning against reference

- True 'blind' inter-comparisons can be less good.
- Inter-sensor variability is less well defined – heavy tuning to the 'best sensor'
- Very few annual or longer studies or performance

UNIVERSITY of York

Applications and data requirements

WORLD METEOROLOGICAL ORGANIZATION

- Not an exclusive list of applications, but these are some that have been proposed by WMO
- General requirements in terms of sensor performance differ by application

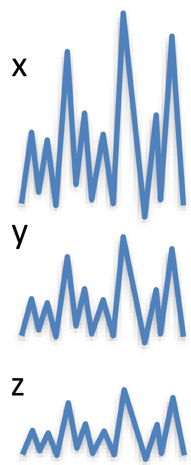
Sensor data

Temporal variability

e.g. 'Pollution is highest in the morning'

Minimum requirements:

1. Sensors are stable over the period of interest
2. Sensors respond broadly to the pollution parameter

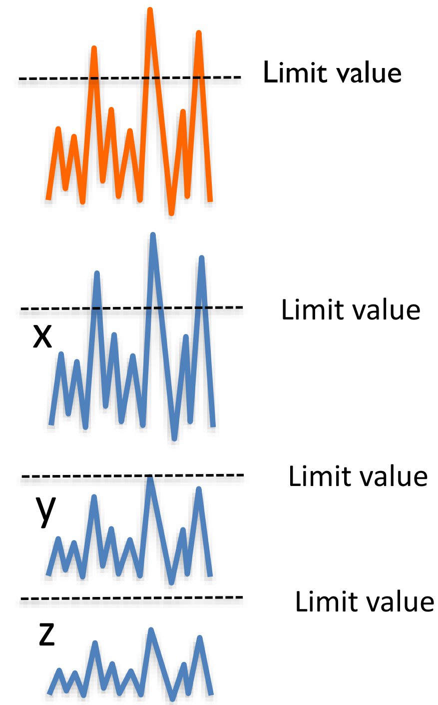


Spatial variability

e.g. 'location x has higher air pollution than locations y and z'

1. Stable over the period of interest
2. Responds broadly to pollution parameter
3. Sensors are internally reproducible

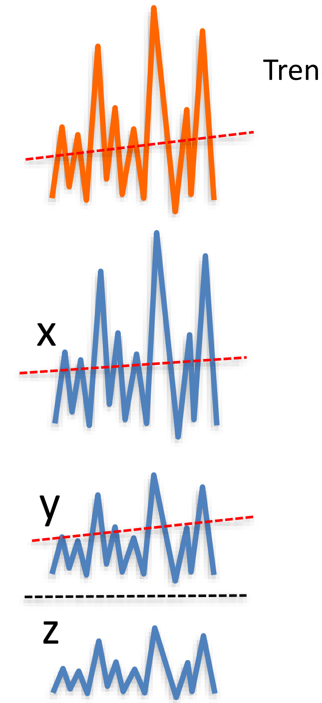
Reference site data



Concentration dependence

e.g. 'location x exceeds the AQ limits but y and z do not'

1. Stable over the period of interest
2. Sensors are compound specific
3. Sensors are externally reproducible

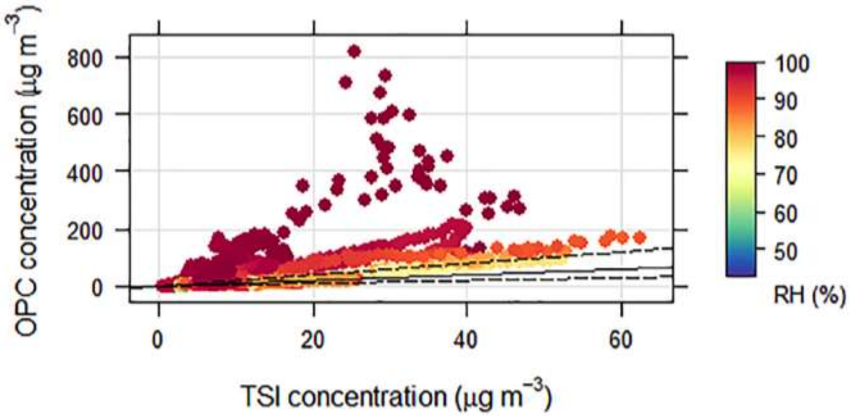


Long-term trends

e.g. 'species at location x is increasing at 3% yr'

1. Stable over the period of interest
2. Sensors are compound specific
3. Sensors are globally intercomparable

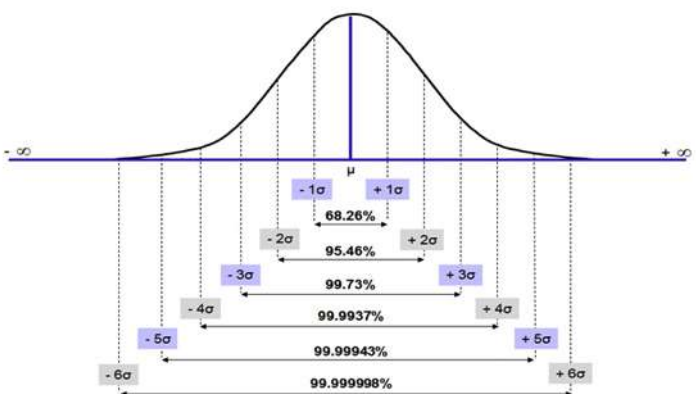
Some of the key issues identified



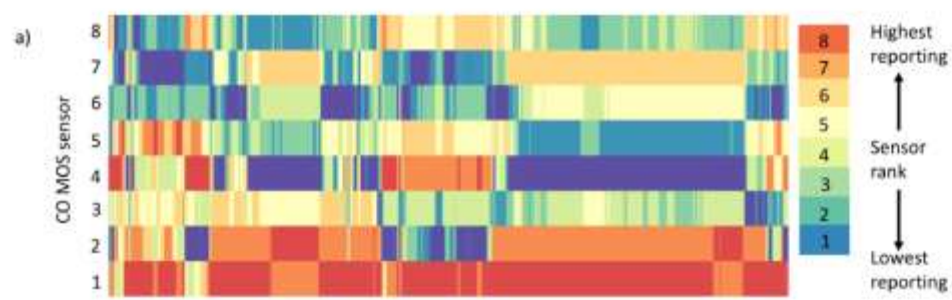
Sensitivity to meteorology and environment

Sensor	Compound				Co-pollutants			
	CO	SO_2	NO	O_3	NO_2	CO_2	H_2	%RH[a]
CO - B4	0.378	-0.013	0.000	0.0200	0.032	0.000	-0.032	0.201
OX-B421	0.000	-0.016	-0.110	0.439	0.44	9.5×10^{-5}		0.560
SO_2-B4	0.013	0.210	0.023	-0.014	-0.32	9.8×10^{-6}		0.000
NO-B4	0	0.007	0.558	-0.011	-0.590	1.8×10^{-5}		-0.303
NO_2-B4	0	0.004	-0.008	0	0.148	2.3×10^{-5}		0.000

Sensitivity to other air pollutants (interferences)



Sensor to sensor variability



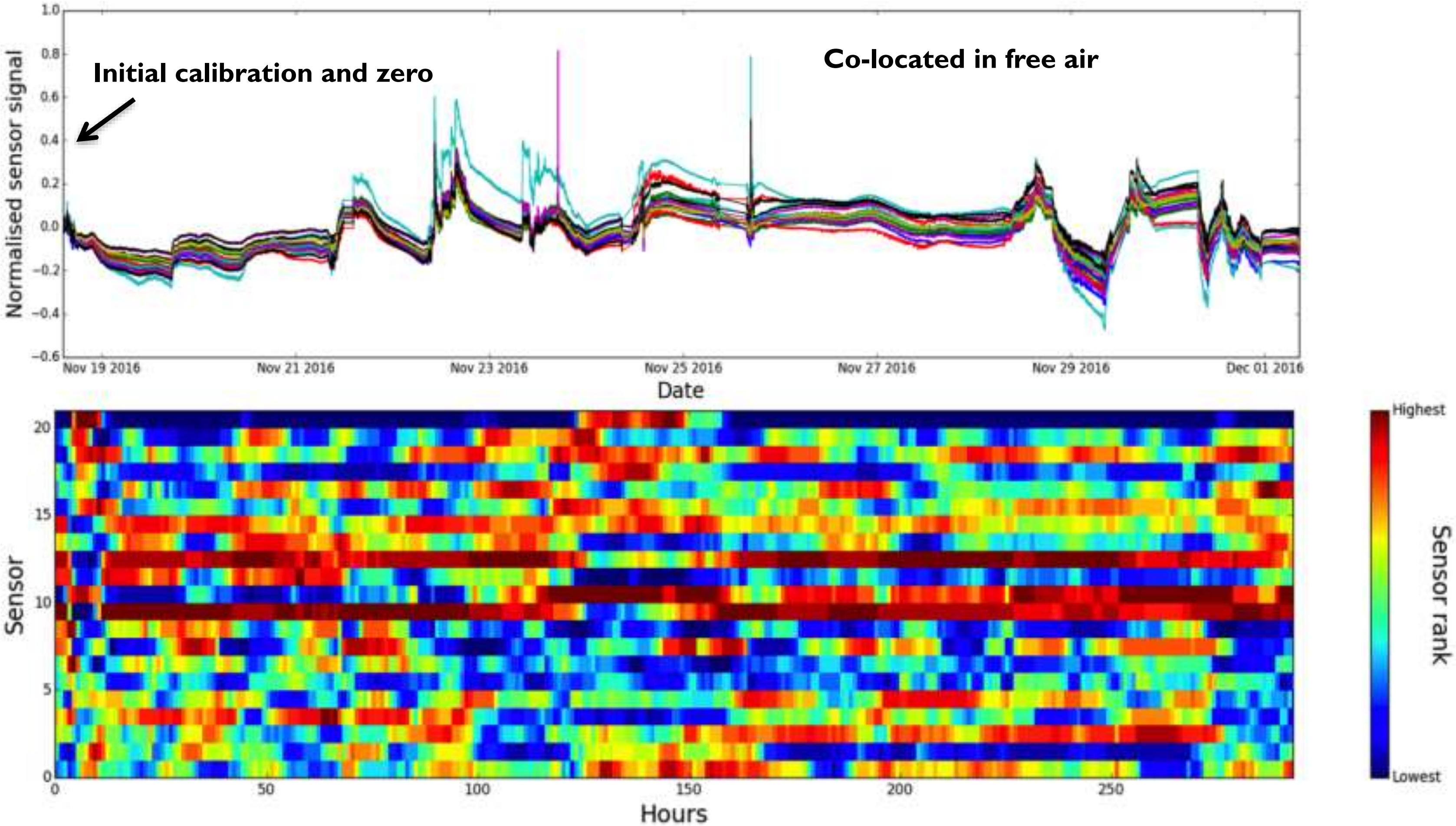
Long-term performance and change

- Data processing strategies are proposed to potentially correct for some or all of these data quality factors notably "Machine Learning" in it many different forms.

University of York

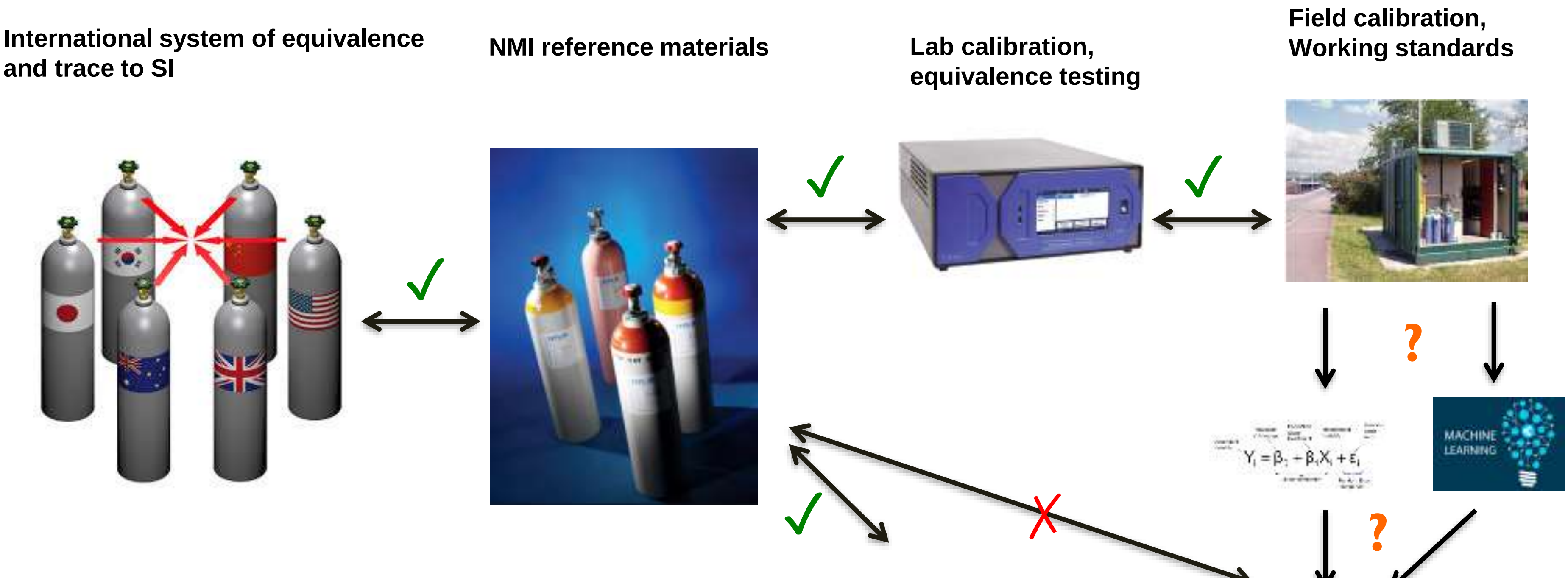
Understanding sensor variability

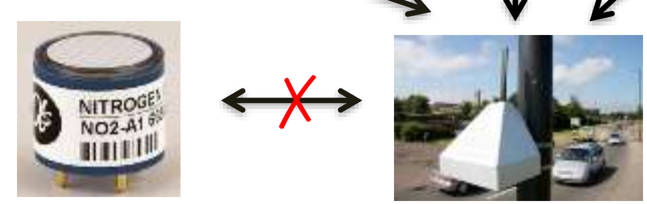
- Sensors are 'predictably unpredictable', but there is often collective skill
- 20 identical MOS sensors, with temperature and humidity controlled
- Ranking the sensors from the highest responding to lowest



Sensor traceability challenges

- There is a well established global system for equivalence and traceability based on binary and multi-component gas standards.
- The high sensitivity of sensors to environmental conditions, water vapour, chemical cross-interferences makes existing system incompatible

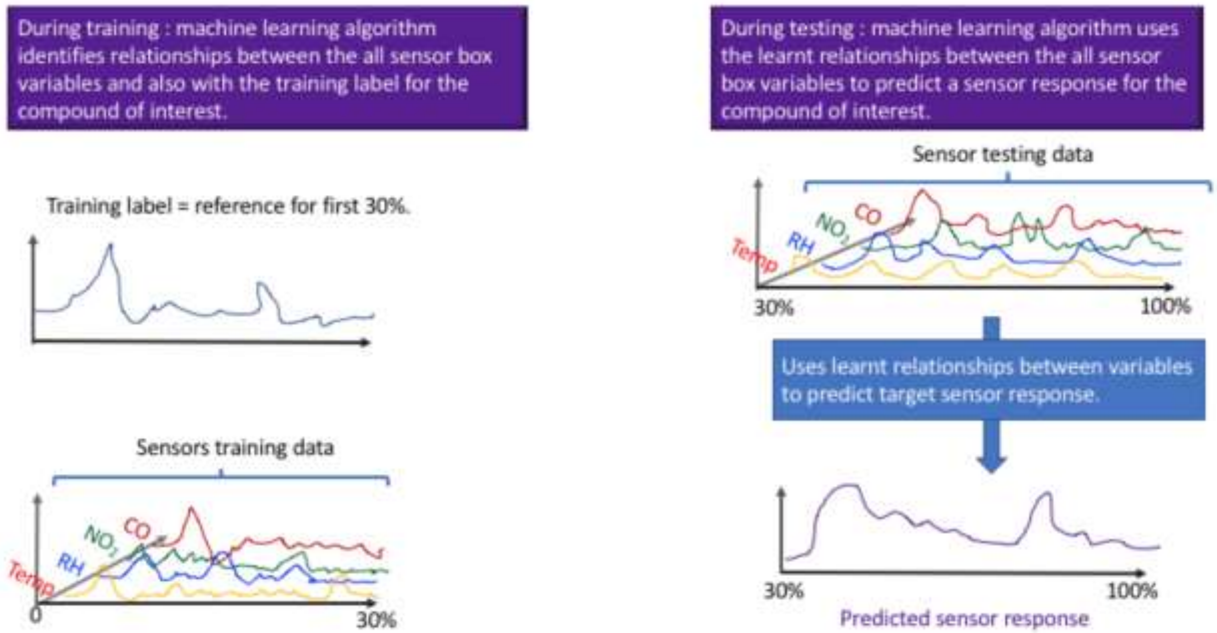




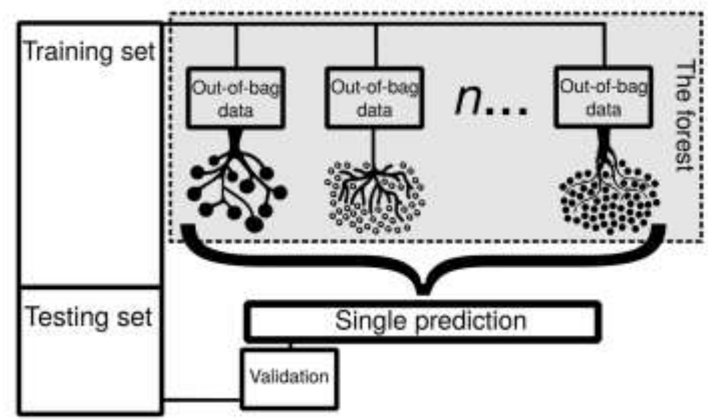
- Linear regression approach is at least ‘reversible’, but may not work very well when multiple corrections are needed.
- Machine learning methods can be a black box and not reversible.

UNIVERSITY of York

Data correction techniques and AQ sensors



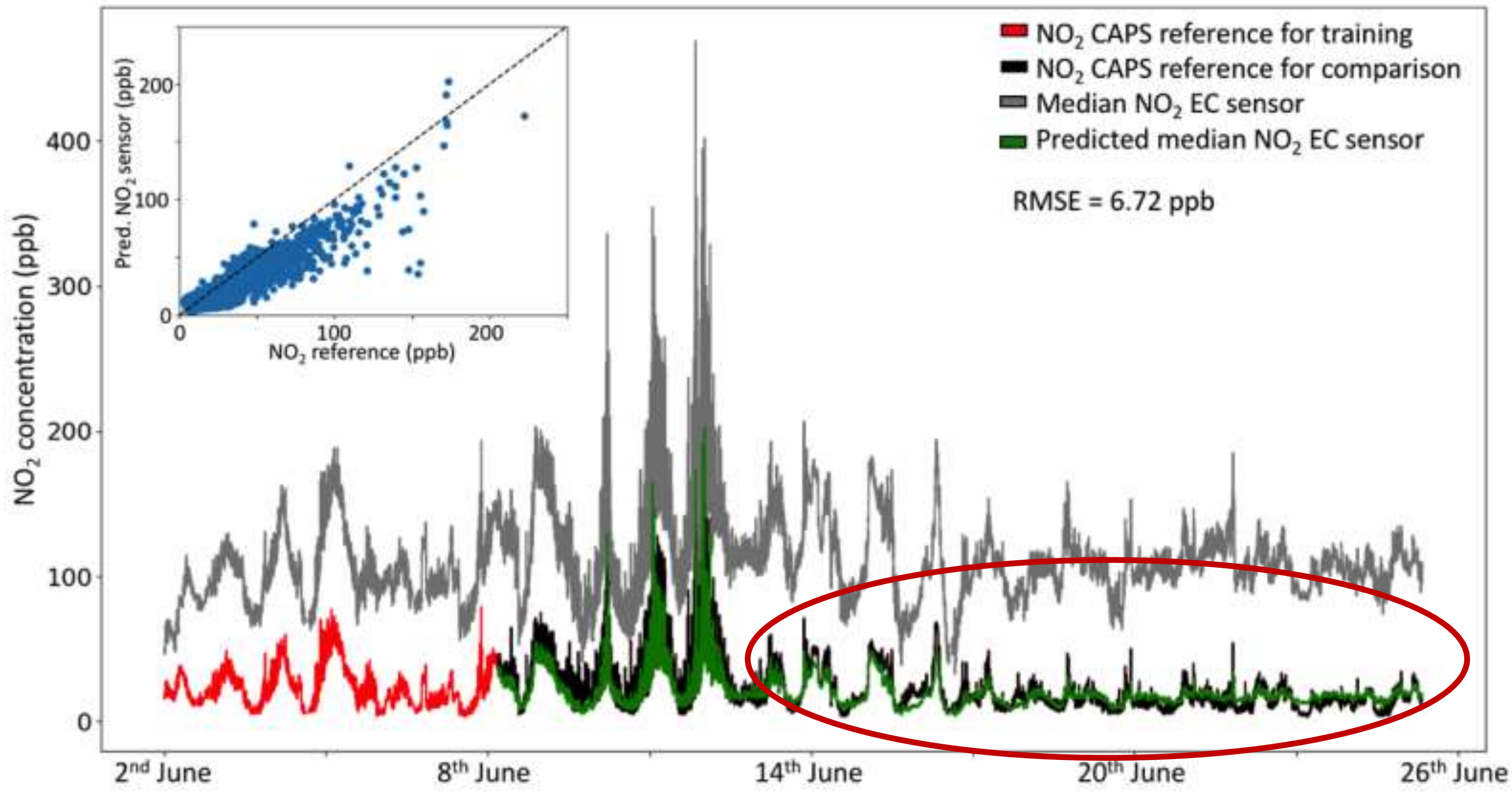
- ML can learns how an AQ sensor responds compared to a reference instrument, and then uses this to then improve the prediction for an unknown period.
- Needs to learn from co-measured data on the key interference parameters eg Temp, RH, windspeed, CO_2, other pollutants
- Boosted regression trees are one ML method that is ‘transparent’.



Boosted regression trees

University of York

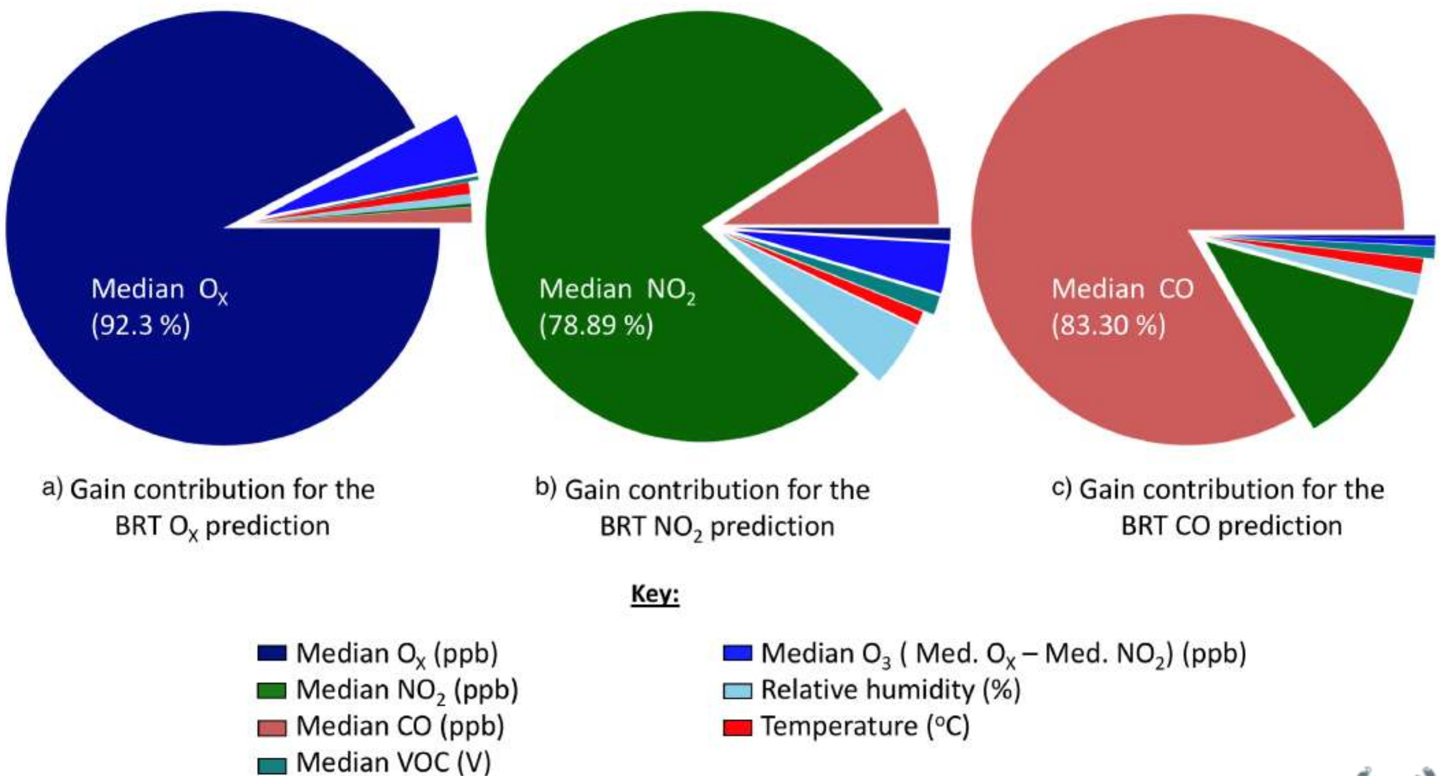
Boosted linear regression



- Using a week of reference data (in red) for NO_2, plus all other parameters, eg T, RH, CO, O_3, M etc for 'learning', then apply boosted linear regression to the green period.
- The ML method then produces sensor data that agrees better for NO_2 under more 'normal conditions'.

Measurement vs model?

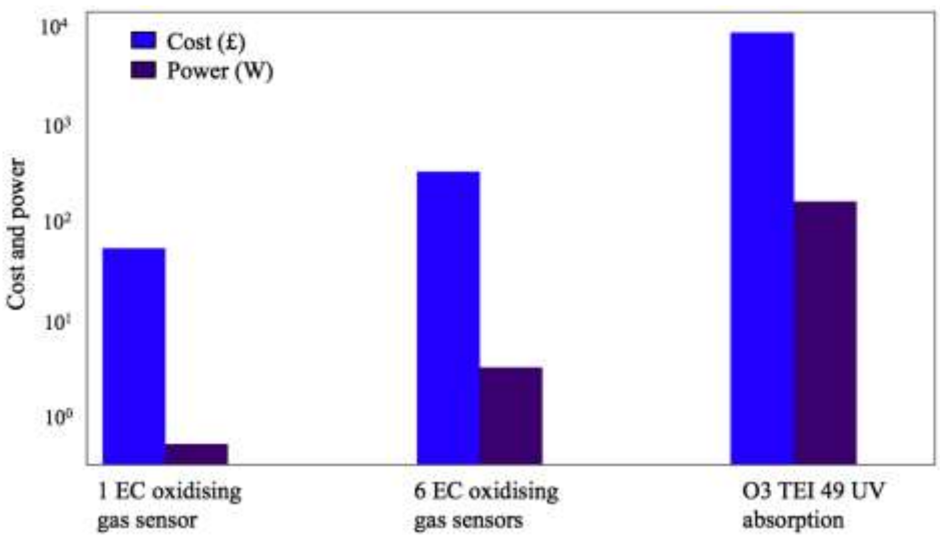
- Needs very careful supervision. It is possible to 'predict' a sensor value, even with no sensor present.
- Machine learning can make a very good guess of concentrations, just by learning how reference data responses to: time of day, days of week, weather and some other pollutants concentrations. **This is not a measurement!**
- Restricted to T, RH, VOC, CO, O_x etc for 'learning',

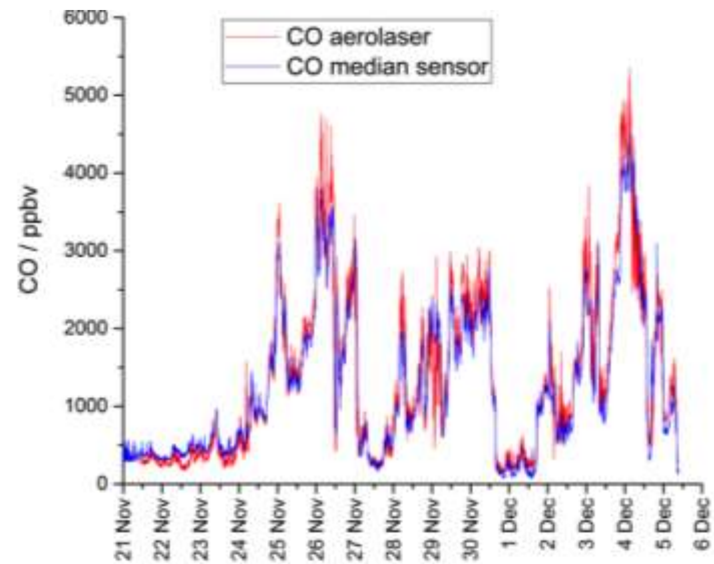


Machine Learning data source contributions used to predict final sensor value

Have we missed the point of sensors?

- We are trying to make a huge technology jump in one step:
    1. Massively reducing initial hardware cost, **AND**
    2. Making devices portable / externally deployable) **AND**
    3. Inventing a new calibration paradigm, **AND**
    4. Reducing operational burden of measurement
- A next step *may* be to tackle only one to two of these problems at a time:





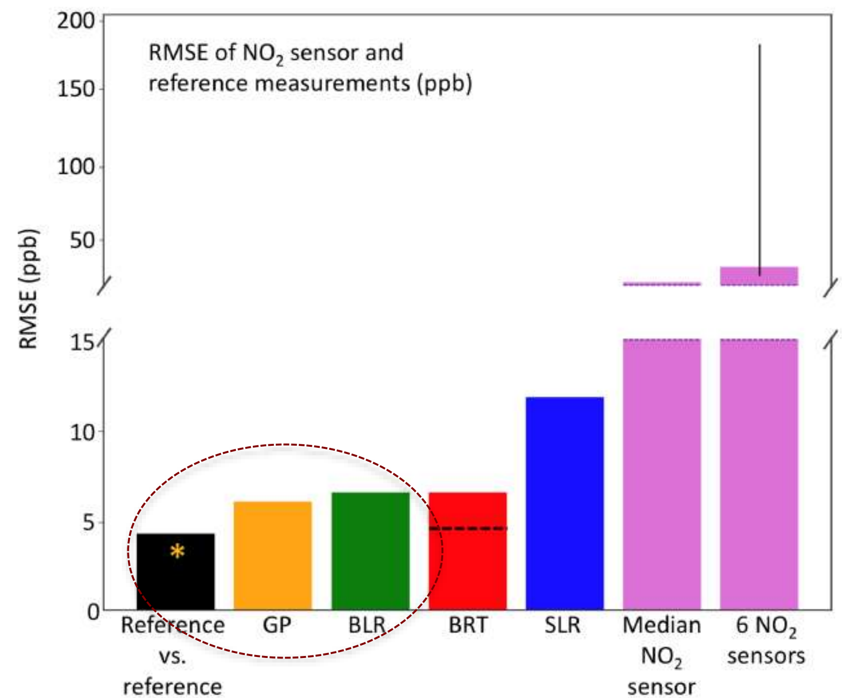
- Hardware cost is still low whether you buy one sensor, or 6, or 20.
- The biggest gain, from a operational perspective, is **low electrical power and fewer expensive high energy parts**.
- One sensor is rather poor, but the median of a cluster is much better.

UNIVERSITY of York

Using AQ sensors to make reference grade measurements

- Clusters of six identical sensors for each pollutant (48 sensors total)
- Use median value from the cluster for each parameter
- Make corrections using simultaneous T, RH, flow, CO_2 data.

AQ cluster: 6x (CO, O_3, NO, NO_2, VOC, CO_2) + T, P RH



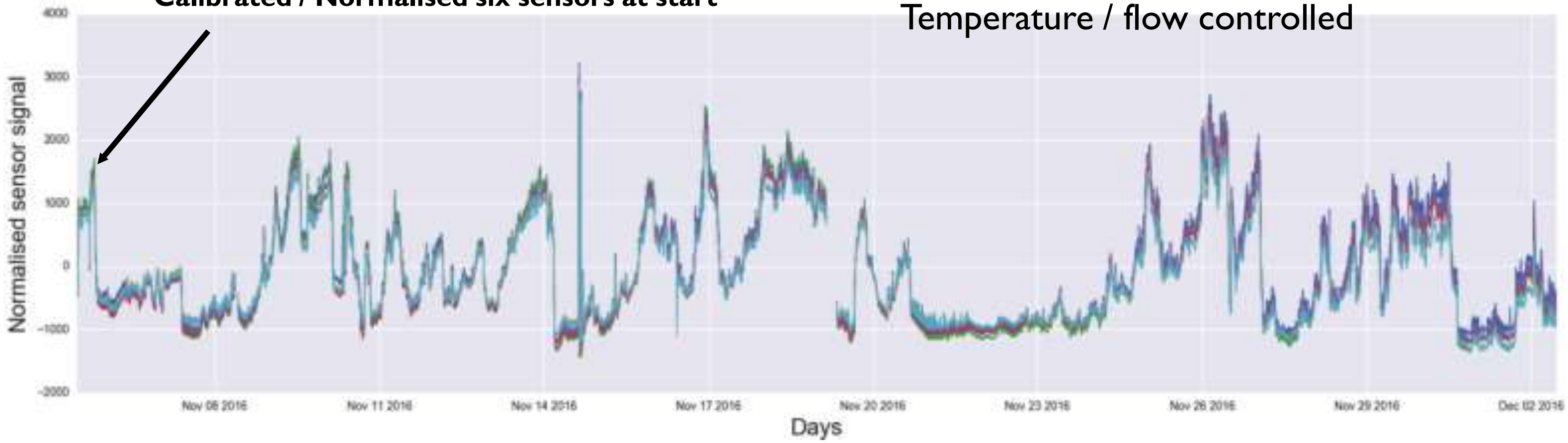
RMSE using clusters + ML methods vs reference instrument

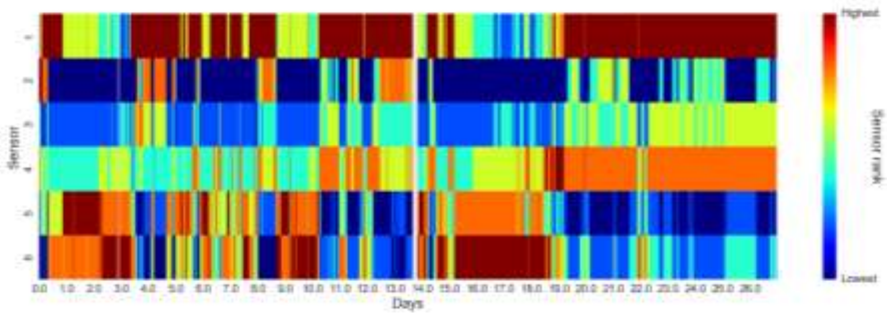
- Measuring multiple simultaneous parameters and using sensor cluster median values + ML, produces data very close regulatory standard.
- Power ~100W, cost ~10,000 USD, weight 10 kg.

How clusters of sensors help

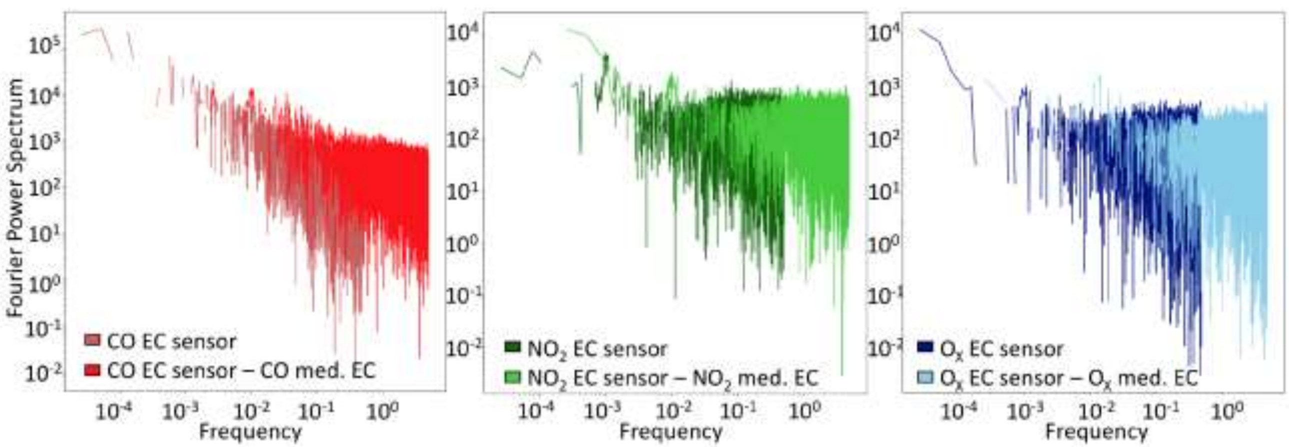
Cluster of 6 x CO electrochemical sensors used in Beijing
Temperature / flow controlled

Calibrated / Normalised six sensors at start





- Randomised day-to-day drift is the biggest source of error
- Using a the cluster median value resolves this over week-to month timescales.



Subtracting the median value generates white noise, vs pink noise for individual sensors

Conclusions

- Current off-the-shelf AQ sensor devices are **highly variable**.
- Some **market attrition** of the poorest quality already; survival of the fittest.
- **Publication bias** with limited reports on uncertainties in the real-world
- **Chemical cross-interferences do occur** at low concentrations,
- **Environmental interferences** can require very large corrections
- Randomised response drifts over the **hour to day timescale** are large compared to instantaneous sensitivity (which is often very good).

- Operating under **stable** 'lab' conditions gives better results than placing uncontrolled outdoors – obvious.
- **Clusters of sensors** can solve single-sensor drift problems, but maintain many important operational advantages, like power, size, costs.
- **Statistical methods** offer considerable promise for removing interferences
- But too much reliance on ML can mean it is simply a model prediction, not a measurement
- Think more about **incremental steps** with sensors, rather than immediately solving all cost/autonomy/power/calibration challenges in one step?

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
WACL Wolfson Atmospheric Chemistry Laboratories
UNIVERSITY of York