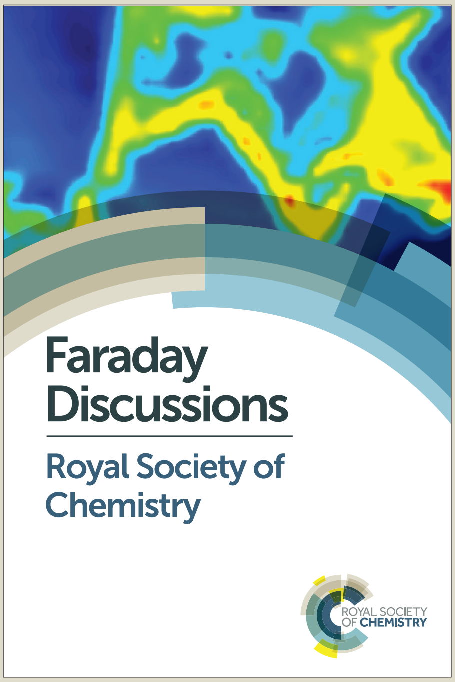
Faraday Discussions

Accepted Manuscript

This manuscript will be presented and discussed at a forthcoming Faraday Discussion meeting. All delegates can contribute to the discussion which will be included in the final volume.

Register now to attend! Full details of all upcoming meetings: http://rsc.li/fd-upcoming-meetings



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

Faraday Discussions ‘Chemistry in the Urban Atmosphere’ 6-8 April 2016. Invited paper.

Evaluating the performance of low cost chemical sensors for air pollution research.

Alastair C Lewis[1,2], James D Lee[1,2], Peter M Edwards[1], Marvin D Shaw[1], Mat J Evans[1,2], Sarah J Moller[1,2], Katie Smith[1], Jack W Buckley[1], Matthew Ellis[1], Stefan Gillott[3], Andrew White[4]

[1] Wolfson Atmospheric Chemistry Laboratories, University of York, Heslington, York, YO10 5DD, UK.
[2] National Centre for Atmospheric Science, University of York, Heslington, York, YO10 5DD, UK.
[3] Environment Department, University of York, Heslington, York, YO10 5DD, UK.
[4] Department of Electronics, University of York, Heslington, York, YO10 5DD, UK.

Abstract

Low cost pollution sensors have been widely publicized, in principle offering increased information on the distribution of air pollution and a democratization of air quality measurements to amateur users. We report a laboratory study of commonly-used electrochemical sensors and quantify a number of cross-interferences with other atmospheric chemicals, some of which become significant at typical suburban air pollution concentrations. We highlight that artefact signals from co-sampled pollutants such as CO_2 can be greater than the electrochemical sensor signal generated by the measurand. We subsequently tested in ambient air over a period of three weeks, twenty identical commercial sensor packages alongside standard measurements and report on the degree of agreement between references and sensors. We then explore potential experimental approaches to improve sensor performance, enhancing outputs from qualitative to quantitative, focusing on low cost VOC photoionization sensors. Careful signal handling for example was seen to improve limits of detection by one order of magnitude. The quantity, magnitude and complexity of analytical interferences that must be characterised to convert a signal into a quantitative observation, with known uncertainties, makes standard individual parameter regression inappropriate. We show that one potential solution to this problem is the application of supervised machine learning approaches such as Boosted regression trees and Gaussian processes emulation.

Correspondence to A.C Lewis (ally.lewis@york.ac.uk)

Introduction.

Long-term atmospheric monitoring of classical pollutants such as NO_x, O_3, SO_2, CO, Volatile Organic Compounds (VOCs) and Particulate Matter (PM) has typically been conducted to support air quality compliance and national frameworks for pollution control. Such data has also made vital contributions that have supported research in more general atmospheric chemistry, for emissions assessment and in many epidemiological studies of health impacts[1]. Historically the users of air pollution instrumentation, and the resulting data, have been technical specialists in the field, supported by established protocols for gas and particle metrology. The most common observational strategy for urban atmospheric chemistry has been to make high accuracy air pollution measurements at small numbers of notionally representative urban/suburban locations, with modeling used to support interpolation to unmonitored locations[2]. The constituency and user-base for surface air pollution measurements has been largely static for perhaps fifty years, and changes in the methodologies and technologies employed for measurement have been incremental, most notably an evolution in methods used for regulatory monitoring of particulate matter. The situation for technologies that measure air pollution from space is rather different, but is outside of this discussion.

A fundamental change to the surface monitoring status quo, enabled by low cost air pollution sensors, has been mooted for more than a decade[345] and gained much media attention in recent years [in the last year, for example: The Times, 2014; *New York Times*, 2015; *The Guardian*, 2015][678]. Such a technology shift has enormous attractions for research users; dense or portable networks of measurement would give unprecedented insight into human exposure[9], into CFD and high-resolution model performance, into emissions and much more. Possibly more significant than the research value, such technologies could result in the democratization of air quality observations to the general public[10], making local air pollution measurement as common, or more so than amateur meteorological stations.

Whilst there is no single agreed definition of what constitutes a low cost sensor (the phrase naturally being rather subjective), in this paper we are referring to devices that purport to make autonomous observations of multiple pollutant parameters at a capital cost in the range of 100-10000 USD per observing location. This contrasts with traditional multi-pollutant analytical capability, such that

---

[1] Review of the UK Air Quality Index, 2011, ISBN 978-0-85951-699-0.

[2] http://uk-air.defra.gov.uk/reports/cat13/1309250915_130923_Review_of_air_quality_monitoring_station_classifications.pdf

[3] M.C. Carotta, G. Martinelli, L. Crema, C. Malagù, M. Merli, G. Ghiotti, E. Traversa. *Sensors and Actuators B: Chemical*, 2001, **76** ,336–343.

[4] Bourgeois, W., Romain, A. C., Nicolas, J., and Stuetz, R. M. *J. Environ. Monitor.*, 2003, **5**, 852–860.

[5] W. Tsujita, A. Yoshino, H. Ishida, T. Moriizumi, *Sensors and Actuators B: Chemical*, 2005, **110** (2), 304–311.

[6] http://www.thetimes.co.uk/tto/environment/article4219453.ece

[7] http://www.nytimes.com/2015/04/16/business/experimenting-at-home-with-air-quality-monitors.html?_r=0

[8] http://www.theguardian.com/environment/2015/aug/10/lord-drayson-takes-on-uk-air-pollution-crisis-with-new-smart-sensor

[9] B. Predic, Y. Zhixian, J. Eberle, D. Stojanovic, D and K. Aberer, IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013, 303-305.

[10] D. Hasenfratz, O. Saukh,S. Sturzenegger, L. Thiele, *2nd International Conference on Mobile Sensing*, 2012.

would meet prescribed US or European requirements for air quality compliance assessment, that have costs of the order >100,000 USD per observing location.

If one considers low cost air pollution sensors as just another type of atmospheric instrumentation, their development and user adoption has differed fundamentally from most other new analytical technologies. The traditional pathway for new analytical capability is for instruments to emerge from either technology-intensive private companies or research laboratories, and be used first by a small number of high-skill early adopters. The atmospheric early adopters typically test, refine and often improve the early versions of instruments, and publish peer-reviewed papers that give the early adopter scientific advantage, by the very fact they have measured something in a novel way. If successful, instruments then propagate into more widespread use, the extent of which depends on the parameter being measured. For classical air pollutants, instruments may ultimately meet requirements for legal compliance or regulatory emission measurements.

Sensors however have followed a somewhat different pathway. Many of the basic sensing technologies used in current commercial devices were created for other applications, often combustion or occupational health measurement, and have been co-opted into ambient measurements at a later date. There have been a range of developers including university labs[11][12], private companies and even devices created via crowd-sourced funding[13] [e.g. AirQuality Egg,]. There is a small body of literature that address issues around interferences, notably water vapour[14], data usability[15] and field calibration[16][17][18][19] , and the conceptual advantages for developing countries[20] . We have struggled however to find examples in the peer-reviewed literature where a research user has unlocked new atmospheric understanding through the deployment of an air pollution sensor network, and this seems unusual when compared with historical technological precedent. Whilst academic useage has been modest, public uptake of air pollution sensors has been growing, for example in London, Breathe Heathrow[21] and Change London[22] are private initiatives that propose the use of large numbers of air pollution sensors, the latter suggesting up to 10,000 may be installed. A consequence is that whilst the concepts and potential for such devices are well-established in the community – from researchers through to legislators - the peer reviewed literature on quantitatively successful deployments is limited.

Ultimately low cost sensors are attempting something exceptionally challenging. To make a trace gas measurement to a usable degree of accuracy and precision, and with stability over time is very hard.

---

[11] S.R. Utembe, GM. Hansford, M.G Sanderson, R.A. Freshwater, K.F.E. Pratt, D.E. Williams, R.A. Cox and R.L. Jones, *Sensors and Actuators, B: Chemical*, 2006, **114**, 507-512.

[12] R. Piedrahita, Y. Xiang, N. Masson, J. Ortega, A. Collier, Y. Jiang, K. Li, R.P. Dick, Q. Lu, M. Hannigan and L. Shang, *Atmos. Meas. Tech*., 2014, **7**, 3325-3336.

[13] https://airqualityegg.wikispaces.com/AirQualityEgg

[14] J.H. Sohn, M. Atzeni, L. Zeller and G. Pioggia, *Sensors and Actuators B: Chemical*, 2008, **131**, 230–235.

[15]

[16] M. Kamionka, P. Breuil and C. Pijolat, *Sensors and Actuators B: Chemical*, 2006, **118**, 323–327,

[17] D. E. Williams, G.S. Henshaw, M. Bart, G. Laing, J. Wagner, S. Naisbitt and J.A Salmond, *Meas. Sci. Technol*., 2013, **24**, 065803, 12 pages.

[18] D. Hasenfratz, O. Saukh, L. Thiele, *Wireless Sensor Networks*, 2012, **7158,** 228-244.

[19] S. De Vito, P. Delli Veneri, E. Esposito, M. Salvato, V. Bright, R.L. Jones, O. Popoola, *Proceedings of the 2015 18th AISEM Annual Conference*.

[20] J. Ikram, A. Tahir, H. Kazmi, Z. Khan, R. Javed and U. Masood. *Environmental Systems Research*, 2012, **1**, 1-8.

[21] http://theodi.org/summer-showcase-breathe-heathrow-methodology

[22] http://www.changelondon.org/airpollution

However since exposure to air pollution is a legally defined quantity in most countries it seems surprising that there are relatively few detailed published measures of performance, particularly for analytical technologies that already have the general public as a user.

In this paper we report i) a detailed laboratory analytical study of common sensor sensitivities and a comprehensive analysis of cross interferences to other atmospheric constituents including other air pollutants and also variable species such as CO_2, H_2 and H_2O. ii) report on initial performance of a commercial low cost air pollution sensing system when compared alongside to more traditional measurement technologies and iii) we describe the development of a scientific concepts that may exploit total VOC sensors and establish analytical performances and interferences. Our intention is to provide a snapshot of performance using current state of the art; we do not speculate on future advances, or attempt to make a valuation of non-research aspects of existing of low cost sensor devices, for example in prompting public engagement with the topic area or related policy issues.

Results and Discussion

Electrochemical sensor interferences.

Air pollution sensors are typically not as analytically specific as current air pollution reference methods (e.g. UV, chemiluminescence, gas chromatography, mass spectrometry etc), and there is potential that sensor measurements may suffer from interferences. Changes in ambient water vapour and temperature have long been known to affect sensor performance[23], but there is also potential interference due to exposure and response to other co-pollutants[24,25]. A small number of previous studies have quantified some of these interferences and then applied corrections to ambient observations[26]. The problem is potentially exacerbated by high concentration co-pollutants, for example CO_2, that are in excess of the measurand often by many orders of magnitude. In an urban setting many gaseous pollutants are correlated in their temporal behaviour to some degree, and interferences may be difficult to distinguish from examination of ambient data alone.

The performance of five commercially available electrochemical sensors designed for the measurement of CO (CO-B4, Alphasense Ltd), O_3 (OX-B421, Alphasense Ltd), NO (NO-B4, Alphasense Ltd), NO_2 (NO_2-B4, Alphasense Ltd) and SO_2 (SO_2-B4, Alphasenase Ltd) in ambient air were evaluated in the laboratory. [Datasheets are available from http://www.alphasense.com/]. Our aim was to establish the selectivity of these sensors to their target compounds, and quantitatively characterise chemical interference to other pollutants. We then evaluate the scale of impacts of co-pollutants through an inter-comparison exercise alongside reference measurements of the same pollutants in ambient air.

All sensors were housed within a single 3D printed PLA flow cell (Makerbot) with calibration or ambient air introduced to the sensor heads simultaneously under controlled conditions. The data acquisition rate was 1 Hz (average to 5 minute intervals) using a Labjack U6 (Labjack) measurement and automation device and Labview software (Labview 2012). All gas lines used were ¼” PTFE with stainless steel fittings (Swagelok). Since previously studies have indicated that pressure, humidity, temperature and flowrate can effect sensor response, we recorded these using an LM35 temperature

---

[23] M.L. Hitchman, N.J. Cade, K.T. Gibbs, N.J.M. Hedley, *Analyst*, 1997, **122** , 1411–1418

[24] C. Austin, B. Roberge, N. Goyer. *J. Environ. Monit.*, 2006, **8** , 161–166.

[25] C. Hamann, A. Hamnett, W. Vielstich, *Electrochemistry*, (second ed.), 978-3-527-31069-2, Wiley-VCH , 2007.

[26] M.I. Mead, O.A.M. Popoola, G.B. Stewart, P. Landshoff, M. Calleja, M. Hayes, J.J. Baldovi, M.W. McLeod, T.F. Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J.R. Saffell, R.L. Jones, *Atmospheric Environment*, 2013, **70**, 186-203.

sensor (Texas Instruments), MPX4200A absolute pressure sensor (Freescale Ltd) and an HIH-4000-001 humidity probe (Honeywell). During the analysis periods, in line gas temperature and pressure were consistent, 20.2 ± 0.7 °C, 1.0003 ± 0.0009 Bar respectively, minimising their effects upon the sensors.

Chemical interferences of co-pollutants

The experimental set-up used a pure air generator (PAG003, Eco-physics) to create the balance gas. Zero air from the generator during these experiments was measured and contained < 10 ppt, < 5 ppt, < 10 ppt, <10 ppt and < 50 ppt of NO, NO_2, O_3, SO_2 and CO respectively. Before work commenced each sensor's working electrode voltage (WE) and auxiliary electrode voltage (AE) offset was determined using the zero gas. Zero air was then humidity controlled using deionised water with a dew point generator (DG-3, Michell Instruments). Finally known gas phase mixtures of NO, NO_2, SO_2, CO_2, H_2 and CO were blended into air from calibration standards in N_2 (BOC) using a gas dilution device (Multi-gas calibrator, S6100, Monitor Europe). O_3 concentrations were generated *in situ* using a gas dilution device and mercury UV lamp, the output concentrations from which were confirmed using a separate online monitor (Thermo, model 49i). Humidity dependent WE and AE sensitivities toward the listed compounds were evaluated using five concentrations and linear regression analysis each performed at 4 different humidities (Table 1). Mole fraction test ranges used were 0 - 200 ppb for CO, 0 - 350 ppb for O_3, 1 – 160 ppb for NO, 0 – 160 ppb for NO_2, 0 – 400 ppm for CO_2 and 0 - 40 ppb for SO_2, with each measurement point performed at 15, 30, 45, 60 %RH.

No observable change in AE voltage was observed during the experiments, and only the effect upon the WE voltage is shown here. We summarise the mV interference induced on a sensor for each co-pollutant in Table 1. There is a complex set of interconnections between co-pollutant and measurand. In some cases the *absolute* interferent signal induced by a co-pollutant is very small, e.g. see CO_2 impacts on the SO_2 or NOx sensors. However the relative amounts of certain co-pollutants to the measurand is such that the overall response of a given sensor could be dominated by the co-pollutant response.

Sensor	**Compound**							
	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

Table 1. Working electrode responses (in mV ppb^{-1} of co-pollutant) induced by the presentation of co-pollutants in zero air across five electrochemical sensors, with a four point humidity correction calibration applied via working electrode offset correction (mV / %RH).

The scale of interferences in ambient air.

We next performed a set of ambient measurements using the five sensors, plus separate reference measurements of the interferent co-pollutants shown in Table 1. We use the ambient measurements to provide us with a representative range of co-pollutant values for which possible sensor interferences may be calculated.

The reference instruments used in this study, and for the later section 2, were as follows: For O_3, the reference instrument was a Thermo Environmental Instruments (TEI) 49C UV absorption analyser which is a United States Environmental Protection Agency (USEPA) equivalent method. Calibration of the instrument was carried out using a TEI Primary Ozone standard, which itself is certified yearly by the UK National Physical Laboratory (NPL). The instrument provided minute averaged data. NOx was measured using a custom built, high sensitivity instrument (Air Quality Design Inc). The instrument uses two channels, with NO measured using the well known chemiluminescence technique. NO_2 was quantified using the second channel, with NO_2 being converted to NO using a blue light LED converter centred at 395 nm. The NO_2 mixing ratio is derived from the difference between total NOx and NO mixing ratios. This method provides an accurate and largely interference free method of NO_2 detection, in contrast to commercially available NOx analysers which typically use heated Molybdenum catalysts to convert NO_2 to NO, which are known to be subject to significant interferences from other reactive nitrogen species[27] . The instrument is calibrated via addition of 5 sccm of known NO concentration (from a 5 ppm NO in Nitrogen standard cylinder (BOC - tied to the NPL scale)) to the ambient sample, which is scrubbed of NOx during the calibration procedure. The conversion efficiency of the LED converter is measured in each calibration using gas phase titration of the NO to NO_2 on addition of O_3. A more detailed description of a similar system can be found in Lee et al.[28]. the instrument produces data at 1 second average. The O_3, NO and NO_2 reference data were subsequently averaged to 15 minutes for comparison with the sensors. Carbon dioxide concentrations

---

[27] M. Steinbacher, C. Zellweger, B. Schwarzenbach, S. Bugmann, B. Buchmann, C. Ordóñez, A. S. H. Prevot, and C. Hueglin. *J. Geophys. Res*., 2007, **112**, D11307.

[28] J.D. Lee, S.J Moller, K.A. Read, A.C. Lewis, L. Mendes and L.J. Carpenter, L. *J. Geophys. Res.*, 2009, **114**, D21302.

were determined every 5 minutes using a Dual Column SRI 8610C GC. The dual channel GC has both a flame ionisation detector (FID) and electron capture detector (ECD); CO_2 measurements were made on the FID channel. Ambient air was flowed through the sample loop at a rate of 100 ml min^{-1} to flush and fill the sample loop. The sample is then injected onto the GC column (3 m Haysep D packed column) held at a temperature of 60 ^{o}C for the duration of the GC run. Hydrogen is used as the carrier gas. After exiting the GC column the effluent is passed through a methanizer packed with a nickel catalyst and held at 320 ^{o}C to convert the CO_2 into methane for detection by the FID. Ambient air was introduced to the flow cell from the main sample inlet using a stainless steel diaphragm metal bellows pump (Senior Aerospace, MB302) flow regulated to 1 slpm using a needle valve (Swagelok). Sensor and reference measurement data was averaged to 5 minute intervals and evaluated over an 18 day period (7/8/2015 – 25/8/2015).

Average mixing ratios in ambient air determined from the reference measurements were 23.6 ± 12.3, 1.3 ± 7.2, 5.1 ± 0.2, 0.2± 0.1, 106 ± 24, 676 ± 161 ppb and 389 ± 24 ppm for O_3, NO, NO_2, SO_2, CO, H_2 and CO_2 respectively. Ambient humidity was 59.1 % ± 12.1. Using the ambient measurements over a 18 day period we then calculate the interference induced on each sensor. We take the mean mixing ratio of each co-pollutant and then express the effect of this in inducing an artefact sensor signal. We show the scale of this interference signal as a percentage of the measurand, which was determined independently using a reference method.

Sensor	**CO**	**SO_2**	**NO**	**O_3**	**NO_2**	**CO_2**	**H_2**
Observed ppb	106 ± 24,	0.2± 0.1	1.3 ± 7.2	23.6 ± 12.3	5.1 ± 0.2	389 ± 24 (ppm)	676 ± 161
CO - B4	-	-0.01	0.00	1.40	0.40	0.00	-53.98
SO_2 - B4	4270.57	-	1.25	-9967.40	5194.90	12063.45	
NO - B4	0	0.14	-	-34.12	-415.71	-985.32	
OX- B421	0.00	0.00	20.60	-	22.49	365.79	
NO_2 - B4	0	0	-20.61	0	-	118.94	

Table 2. First row: Observed mean ambient pollution mixing ratio and one sigma range over 18 days. Subsequent rows show the impact of the signal induced by a co-pollutant expressed as a percentage of the mean ambient mixing ratio of the measurand.

What is significant in our test study is the relative scale of impact that co-pollutants have on the induced sensor signal when the analyte pollutant is at typical European suburban values. In our study the ambient mean NO_2 mixing ratio is around 5 ppb which generates a mV sensor signal that is approximately the same as the artefact mV signal induced by the presence of ambient CO_2. SO_2 is a more extreme example: here our ambient measurements are very low, typically 0.2 ppb, such that the interference signals from the more abundant co-pollutants CO, O_3, NO_2 and CO_2 swamp any direct analytical signal from SO_2 itself. The impacts of co-pollutants in our test location are severe for certain sensors although we would stress that the scale of impact depends very much on the ratio of measurand to co-pollutant.

We show in Figure 1 the relative responses that are induced in a NO_2 sensor arising from co-exposure to a range of different CO_2 and NO_2 mole fractions. At polluted roadside locations, for example central London, kerbside NO_2 can be as high as 500 ppb, with corresponding CO_2 around 750 ppm . With this pollutant mixture the sensor response would be overwhelmingly (>98%) due to the signal induced by NO_2 gas and only a small contribution from CO_2. The fractional contribution to this particular sensor signal then changes markedly as one moves from a polluted roadside scenario to suburban or background values. At typical suburban values of 450 ppm CO_2 and 10 ppb NO_2, the signal induced from the sensor is apportioned approximately 50:50 between the gases. We conclude that the influence of co-pollutants on sensor response requires very careful evaluation since sensors within a heterogenous network are likely to suffer from substantially differing degrees of analyte interference depending on the ratio of co-pollutant to measurand.

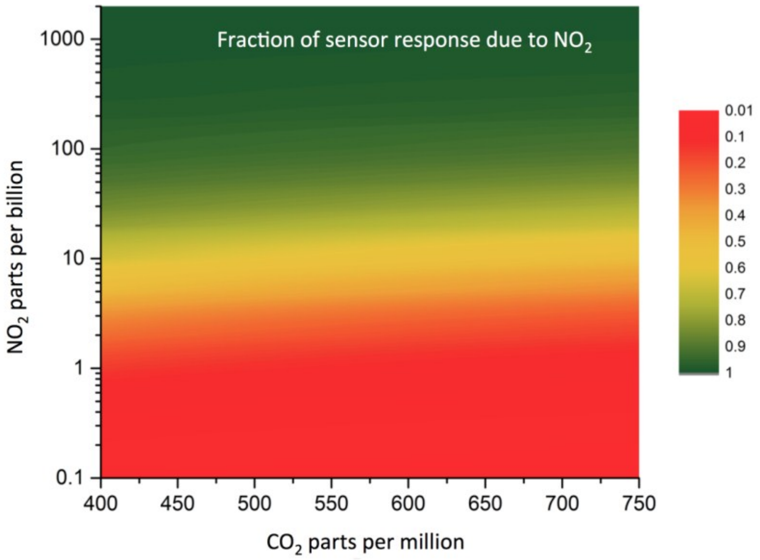


Figure 1. Fraction of electrochemical sensor response due to [NO_2] for a given co-exposure to CO_2 over the CO_2 range 400 to 750 ppm.

Evaluating a multi-sensor package

There are few published studies that compare emerging commercial air pollution sensor packages against existing reference methods in the field. We aim here to establish the utility of data provided by on-the-market devices, the unit to unit precision of measurement, and the comparability of this against reference instruments to establish accuracy. We co-located 20 identical sensor units with stated capabilities to measure urban concentrations of O_3, NO, NO_2, total VOC, and PM. For this paper we anonymise the commercial supplier of the devices, since we do not wish the reader to infer any specific advantage or disadvantage of these particular devices relative to any other current commercial products. We note that some of the core internal component sensing technologies in the system we report on are also found in other commercially available packages. The sensor units were co-located on a flat roofed building (53°56'52.08"N, 1° 2'46.17"W) in York, UK approximately 12 m above ground level and with unobstructed airflow. The sensor inlets were around 30 m from a moderately trafficked single carriageway road. Sensors were placed next to a trace gas sample inlet manifold used as a supply of air to the standard reference instruments measuring O_3, NO and NO_2 housed in laboratories below.

Ozone: The calibration slope factors for the O_3 sensors were modified from initial factory settings following comparison against the reference photometric data and these values were applied on 11th Oct 2015. For the remaining three weeks of study the calibration constants for each sensor remained unchanged. Figure 2 shows the time-series obtained from the reference instrument and the median O_3 sensor observation at any given point in time (n = 20 sensors). The agreement between median sensor and reference is clearly very good. Beyond the median value, a R^2 value in excess of 0.9 was found for virtually all sensors when examined one-by-one. There is a small negative intercept potentially due to an interference at higher NO. We note that the median O_3 sensor value does not derive from a single sensor, but is contributed to from a number of different units over the period.

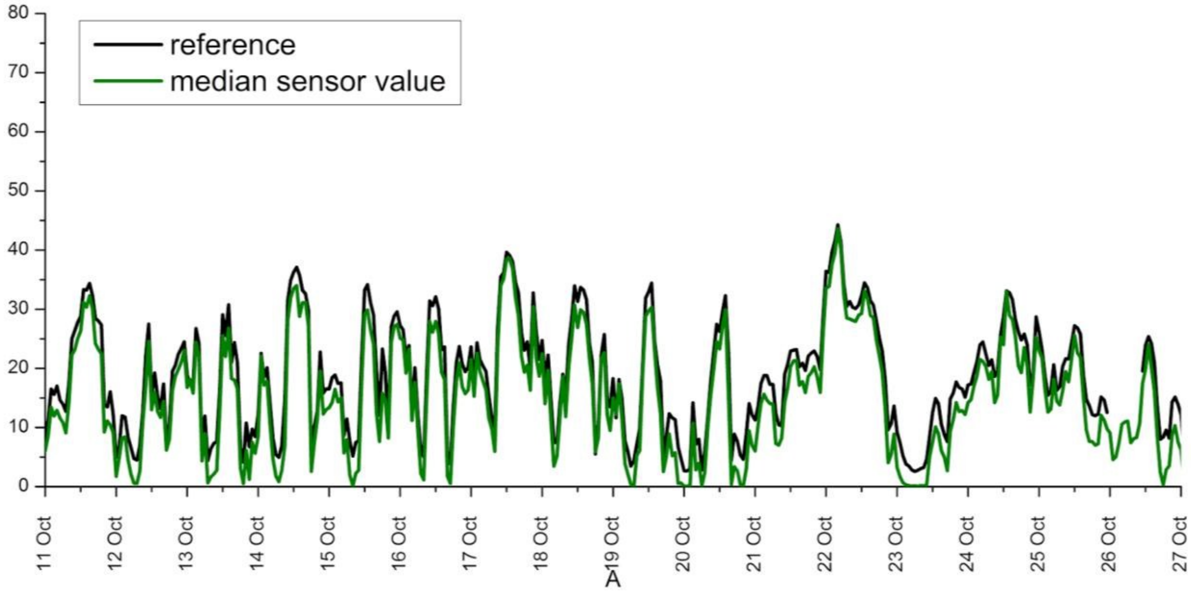


Figure 2. Time series comparison in the field of photometric ozone (black line) and median O_3 sensor (green line) (*n* = 20), R^2 0.904, intercept - 4.6 ppb.

We next examine the simultaneous performance of all sensors relative to the reference O_3 value and show this in Figure 3. The highest and lowest sensor values (red and blue lines) deviate significantly from the sensor median value and reference values, although the agreement of trends remains good through the time period, where R^2 is around 0.9 ± 0.06 (1 sigma) for all individual sensors. To avoid negatively skewing any comparison with a small number of poor performing outlying sensors, we show on the figure a grey shaded region which covers those sensors whose data falls within the 25 to 75th percentile group of sensors. The agreement of this reduced dataset with the reference is also very good and the internal spread of sensor values would be compatible with meaningful measurement given expected atmospheric variability. Taking those sensors in the interquartile range, more than 70% of observations fall within 5 ppb and 95% of observations fall within 10 ppb of the reference values.

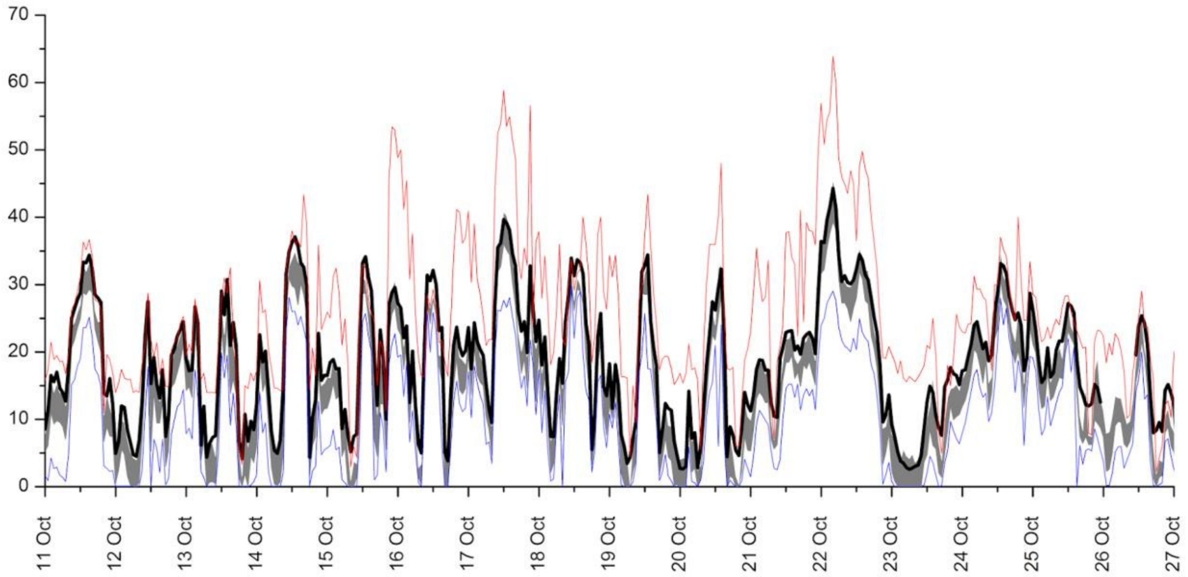


Figure 3. A time-series comparison of reference photometric O_3 instrument (**black line**), highest O_3 sensor (red line), lowest O_3 sensor (blue line). Grey shaded area shows those sensor lying in the 25th to 75th percentile range.

Our conclusion is that given an *in situ* initial normalization of calibration slope factors against a reference instrument, then the resulting O_3 sensor observations would be significant, e.g. in combination have precision that is better than typical daily atmospheric variability over our test period of three weeks, although numerical strategies to handle outlying sensors require development. An evaluation of longer-term stability, and issues such as frequency of field calibration would require further experiments. The analysis raises a wider implementation issue, suggesting that 'managed' deployments of sensors, where local calibration is coordinated and referenced to a single source, may be considerably more precise than that achieved from unstructured deployments of sensors without reference normalisation.

Nitrogen oxide: Our next analysis evaluates performance of the NO sensors against reference chemiluminescence observations and we apply a similar methodology that compares reference against minimum, maximum and those sensors in the 25th to 75th percentile range (Figure 4). In general terms the NO sensors show good temporal agreement against reference measurement, but with poorer linear regression statistics than the O_3 sensors. We note that NO is often at low concentrations in the background York atmosphere and this challenges the detection limit of the sensors in many periods. NO sensor observations compared against chemiluminescence gave R^2 values of 0.73 ± 0.21, although with an overall sensor bias underestimating NO by a factor 2.3 ± 1.1. It is possible that this underreporting bias for $[NO]_{sensor}$ may be correctable, although the inter-sensor spread would remain high.

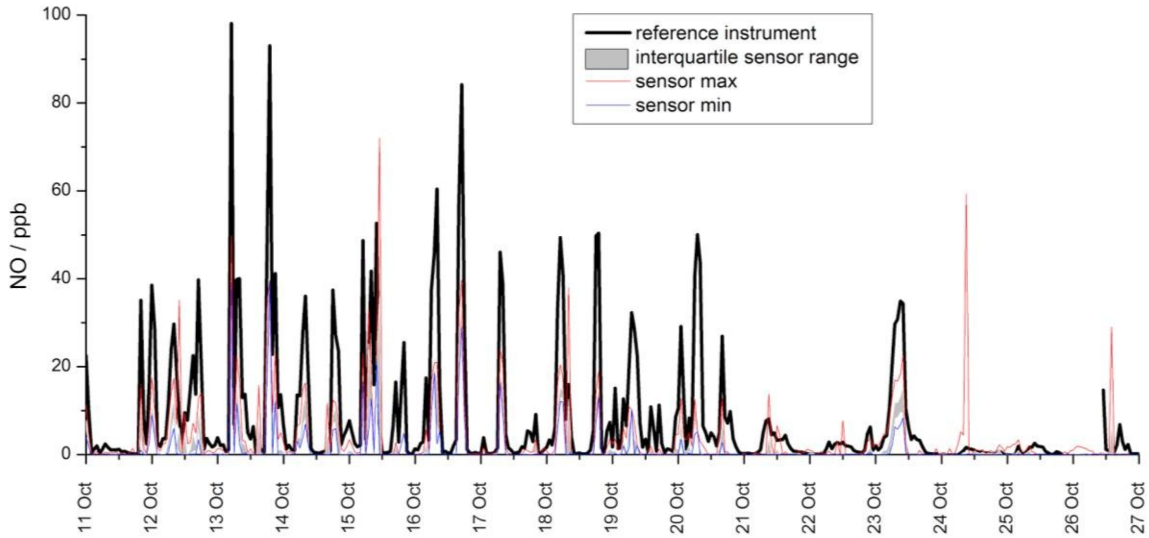


Figure 4. A time-series comparison of reference chemiluminescence instrument (**black line**), highest NO sensor (red line), lowest NO sensor (blue line). Grey shaded area shows those NO sensors lying in the 25th to 75th percentile range.

Nitrogen dioxide: A comparison of reference NO_2 against sensor observation showed poor statistical agreement, suggesting that the sensor was not measuring this chemical parameter exclusively. This contrasts with O_3 and NO from sections 2.1 and 2.2 where there was good evidence for an analytically selective measurement, even if calibration and sensor reproducibility were less than ideal. An analysis of correlation coefficients of each sensor against the reference gave a median $R_2 = 0.25 \pm 0.13$). The sensors typically reported high concentrations, over-reporting against the reference by a factor of 3.2 ± 1.7. Setting to one side the poor agreement in trends between reference and sensor, this substantial over-reporting implied considerable exceedence of air quality standards, where reference measurements showed values well inside regulatory limits. Whilst the direct comparison of NO_2 sensor against reference was poor, the 25th - 75th percentile range of sensors was relatively narrow. This implies that the sensors are broadly reproducible in their response, even if they are not solely measuring the reported parameter of NO_2. Further work is required to determine the exact response characteristics of this particular NO_2 sensor, potentially leading to an operationally defined air quality measure, similar perhaps to a ‘total’ VOC observation.

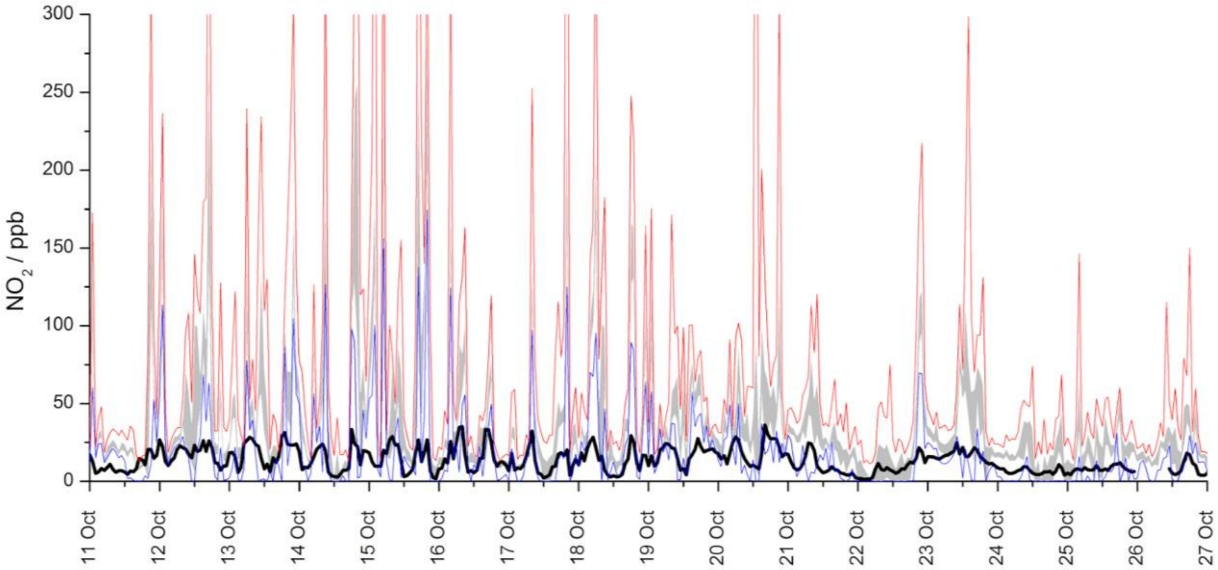


Figure 5. A time-series comparison of reference chemiluminescence NO_2 instrument (**black line**), highest NO_2 sensor (red line), lowest NO_2 sensor (blue line). Grey shaded area shows those NO_2 sensors lying in the 25th to 75th percentile range.

Particulate matter: The final two sensors evaluated (PM and VOC) do not have as precise a reference value to compare against. For NO, NO_2 and O_3 there is a single binary reference gas mixture that ties observations to amount of substance. For Particulate Matter a range of reporting metrics are available, eg $PM_{2.5}$, PM_{10}, and these in turn are somewhat operationally defined in terms of individual instrument responses. For PM, we compare the 20 sensors co-located against one another to determine inter-sensor variability, and then compare two sensors against an air quality reference station in York to establish accuracy against a current standard air quality method for PM reporting.

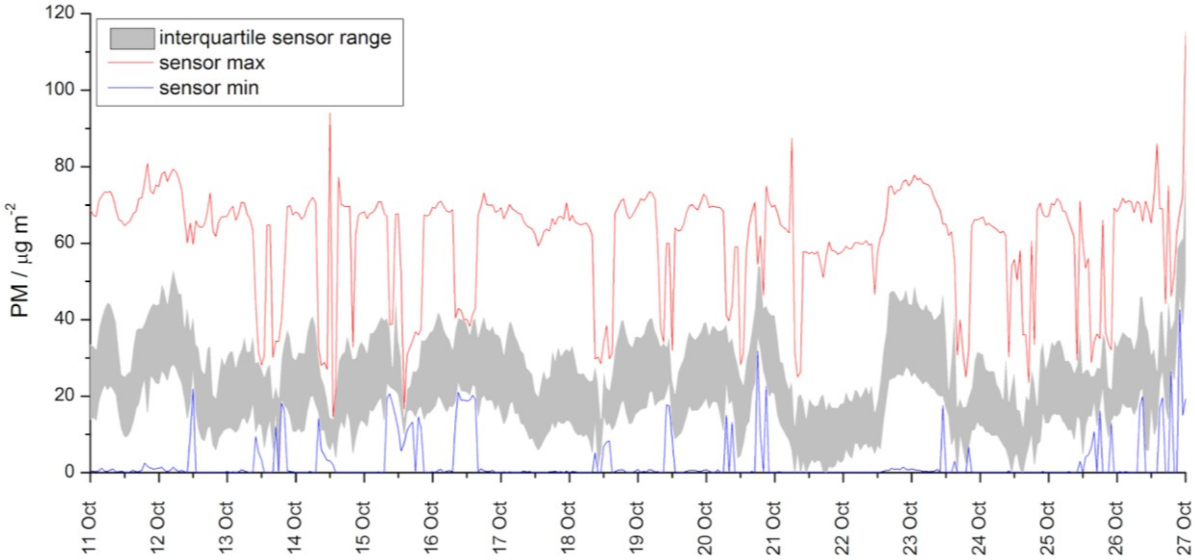


Figure 6. A time-series comparison of highest PM sensor (red line), lowest PM sensor (blue line). Grey shaded area shows those PM sensors lying in the 25th to 75th percentile range.

There is a generally good statistical agreement between the individual sensor PM measurement (R^2 = 0.8 ± 0.11), but with a significant spread in reported concentrations, even taking the reduced range of central sensor values. For PM the 25th to 75th percentile sensors (grey range in Figure 6) span a concentration range ~ 20 ug m^{-3}. This measurement uncertainty is significant in the context of current regulatory limits (eg European Union Directive values: $PM_{2.5}$ annual mean 25 ug m^{-3}, PM_{10} 40 ug m^{-3}). A comparison of sensor particulate matter measurements against a standard reference observations was made at a roadside location in York (Fishergate, TEOM-FDMS, following EN12341:1998). This co-located two sensor units alongside a City of York air quality monitoring site, affiliated to the Defra automatic urban networks for air quality. The data was taken from uk-air.defra.gov.uk. Figure 7 shows this comparison, with modest agreement between sensor and reference PM_{10} (R^2 = 0.35, intercept ~10 ug m^{-3}). A positive bias was observed in the sensor data.

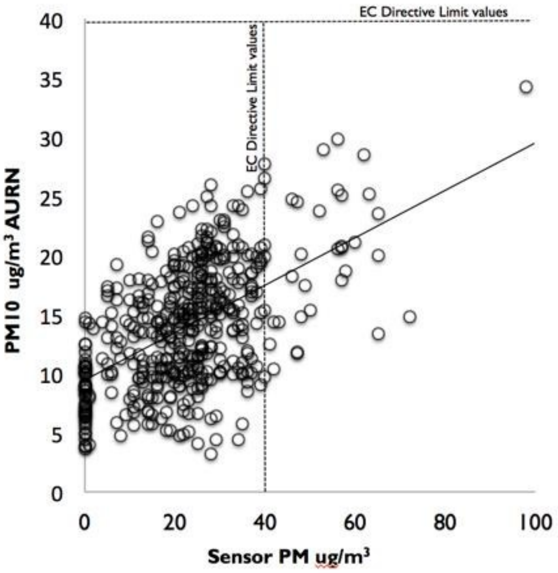


Figure 7. Three week comparison of hourly sensor PM observations against a UK standard PM reference method at the Fishergate air quality station - data from uk-air.defra.gov.uk. EU Directive PM_{10} limit values marked on the plot for information.

Volatile Organic Compounds (VOC): The final sensor we evaluated was a measurement of Volatile Organic Compounds (VOC) using a metal oxide sensing approach. This 'total VOC' definition is common in the sensor reporting of VOCs since metal oxide devices or photoionisation sensors do not speciate individual VOCs in the way that chromatographic or mass spectrometric methods do. The conceptual definition of 'total' is straightforward, but the conversion of a sensor signal to a quantitative value is difficult since sensors typically respond differently to individual organic compounds. Since there is no reference measurement for total VOC, in this section we evaluate only the inter-sensor variability in response, and then in Section 3 discuss in further detail compound specific responses and the potential application of total VOC sensors in the field.

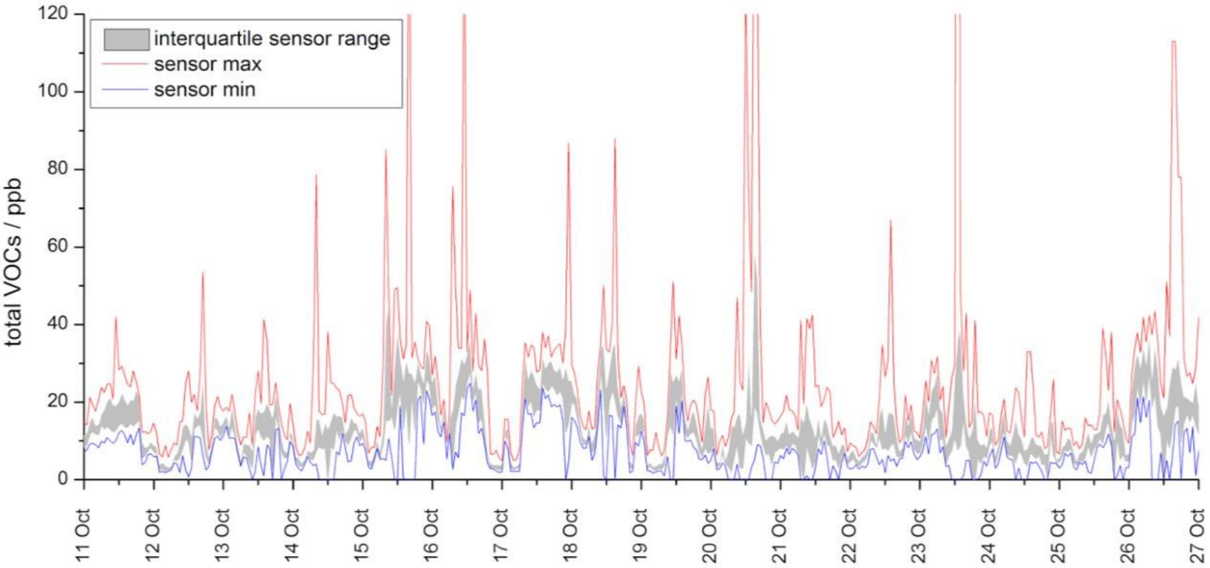


Figure 8. A time-series comparison of highest total VOC sensor (red line), lowest total VOC sensor (blue line). Grey shaded area shows those VOC sensors lying in the 25th to 75th percentile range.

Whilst there is some considerable spread in the values reported between highest to lowest VOC co-located sensor in the study, the sensors data within the 25th to 75th percentile are relatively tightly constrained, with the sensor to sensor variability being modest compared to the atmospheric variability. Since the sensor reproducibility is encouraging, it demands some further thinking on how a relatively non-specific but reproducible observation could be used to address particular scientific problems given VOCs provide one of the principal controls over the functioning of the Earth system and have a significant impact on air quality, climate change and the carbon budget.

New strategies for VOC sensors

Current methodologies for studying the impacts of VOCs on atmospheric chemistry and climate generally use process level or single point observations of concentrations or fluxes to provide constraints for regional and global chemistry models. The upscaling of emissions from process scale (typically that of cm to km scales) to that of the model grid scale (1,10,100 km^2 depending on model and application) are based on the very limited literature datasets since instrumental methods are highly complex. Of all the air pollution parameters considered in this paper, VOCs typically have the most limited observational datasets and by extension would benefit most from usable low cost sensors. In the UK for example there are in excess of 300 official online O_3 air quality measurements, 200 NOx measurements, but only four online VOC measurement stations. This limited information on spatial and temporal variability makes it difficult to quantify uncertainties in model constraints or link emissions to process level controls.

The use of sensors for the detection of VOCs has the potential to deliver something different but complementary to the existing experimental approaches – that is a direct measurement with substantially degraded level of chemical detail compared to the single site observations, but with well-resolved time information and spatial resolution. Recent advances in sensors for the detection of

VOCs, adopted from the hazard detection industry[29], provide a particular opportunity to answer questions inaccessible to conventional research methodologies.

The non-selective nature of current VOC sensor technologies is not as limiting in terms of potential science applications as might first be presumed, as existing literature shows that in many environments VOC speciation is significantly less variable than the changes in total concentration. Our study from section 2.5 shows that metal oxide sensor to sensor variability can be reasonable, when compared with atmospheric variability. Thus the use of a selective and sensitive VOC observation technique (e.g. GC) alongside a distributed network of sensors may enable a mapping of detailed speciated information onto an improved measure of overall temporal/spatial variability. For a VOC sensor to provide useful data on ambient variability, it must be able to detect small changes in VOC concentrations and require limits of detection (LOD) in the 10-100s of pptv range. Sensor response must also be stable over timescales significantly longer than that of the variability being observed (hours to days), and any interferences must be quantifiable and correctable.

Photo Ionisation Detectors (PIDs) are one available sensor technology for Σ[VOC] measurements. These small units (~ 2 cm diameter) use ~ 100 mW and work via the ionisation of molecules whose ionisation energy is less than that of the photons from the UV lamp in the sensor (including most unsaturated VOCs). Current manufacturer quoted LOD of ~5 ppb are however inadequate for ambient observations and improving this has not been a requirement for the majority of industrial PID applications. Reducing the LOD by approximately an order of magnitude through thorough lab characterisation and signal optimisation, increases substantial the utility of the sensors for atmospheric chemistry applications.

Through detailed laboratory experiments we found the PID background signal to be highly sensitive to variables such as local electromagnetic interference, noise on the sensor supply voltage and ground, and noise introduced during the transmission of signal voltages to a suitable analogue to digital converter. Through a series of improvements to minimise the effects of these electronic interferences we were able to reduce the PID background signal standard deviation, for a measurement period of 60 hours, from approximately 10 mV to 3 e^{-2} mV. These improvements in electronics and signal handling enabled small signals from low VOC concentrations to be discernable from the background. Figure 9a shows PID response to isoprene binary mixtures in N_2 over the range 0 and 2.5 ppb. We show the corresponding 3 sigma estimated LOD in Figure 9b.

---

[29] G. Manes, G. Collodi, R. Fusco, L. Gelpi and A. Manes, *International Journal of Distributed Sensor Networks*, 2012, Article ID 820716, 13 pages.

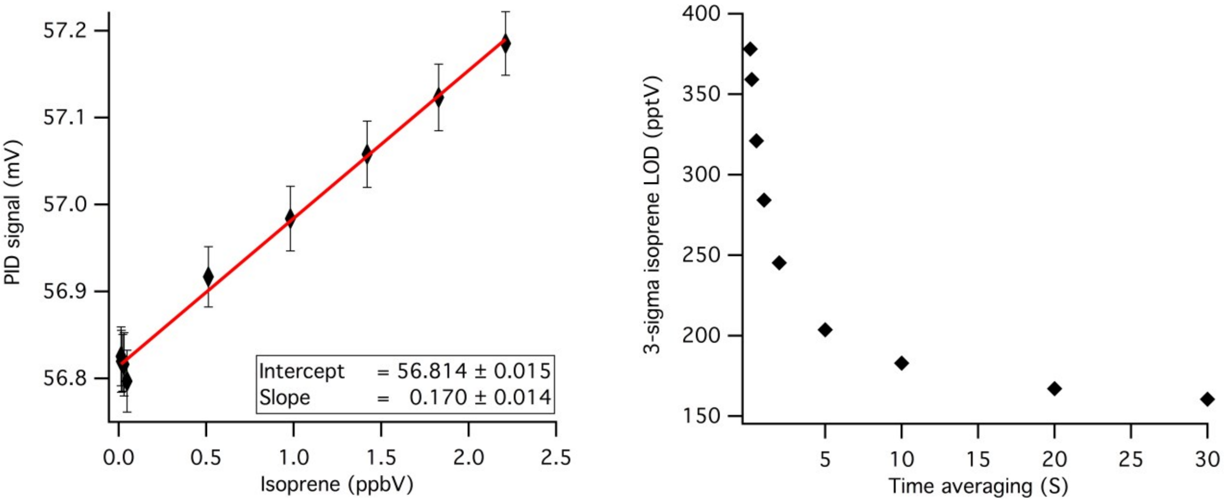


Figure 9: a) PID signal voltage against isoprene mixing ratio (in N_2 balance gas), mean PID signal response and calculated isoprene mixing ratio over a 1-minute sample time, with error bars representing ± 1 standard deviation. Red line shows linear fit to data. b) Estimated PID three-sigma LOD for a range of signal averaging times.

The sensor response shown in Figure 9 suggests that order of magnitude improvements in sensor detection limits are possible with careful signal conditioning. However translating this quantitative performance to ambient operating conditions, compared with the highly optimised laboratory conditions used in Figure 9, requires all possible interferences to be characterised over relevant ranges. Figure 10 shows for example the non-linear PID background signal response to changing relative humidity (RH), which is of a comparable magnitude to that of isoprene (Fig. 9). As well as the obvious complications of correcting for signals from multiple interferences, PID sensor sensitivity to VOCs have also been found to change with RH and will likely also be impacted by other interfering signals. Laboratory experiments have also shown PID response to be sensitive to operational parameters such as sensor supply voltage and temperature.

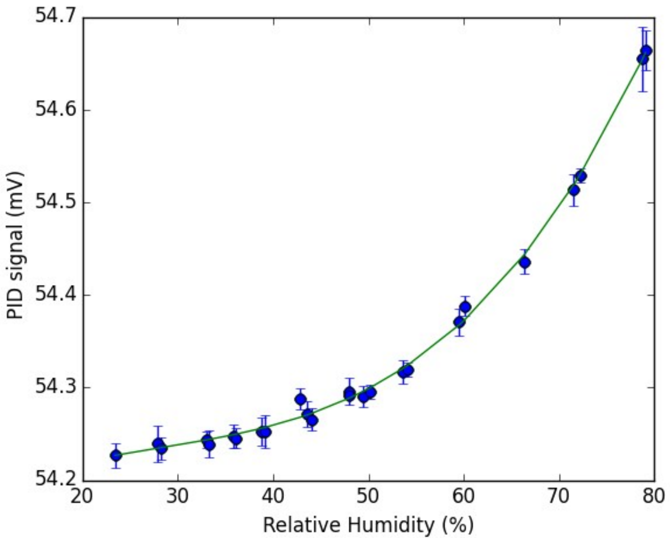


Figure 10: Baseline PID response against relative humidity, with error bars showing ± 1 standard deviation, highlighting potential artefact signal generated solely by a changing humidity. Green line is an exponential fit to the data.

The results shown in Section 2 show the potential performance of sensor devices to illustrate trends and temporal behaviours, successfully so for O_3, NO, PM and VOC - but also the limitations in trying to obtain quantitative information - successful in our study only for O_3. The quantity, magnitude and complexity of analytical interferences that must be characterised in order to convert a VOC (or indeed any sensor) signal into a meaningful quantitative observation, with known uncertainties, makes standard individual parameter regression approaches extremely complicated and uncertain. One potential solution to this problem is the application of statistical methods combined with supervised machine learning approaches, such as boosted regression trees (BRT)[30][31] or Gaussian processes emulation (GPE) or Kriging-Based optimisation[32]. These algorithms have been shown to have significant predicting power once deterministic functions have been created using a training data-set, and have the advantage of being able to deal with complex nonlinearities and covariances.

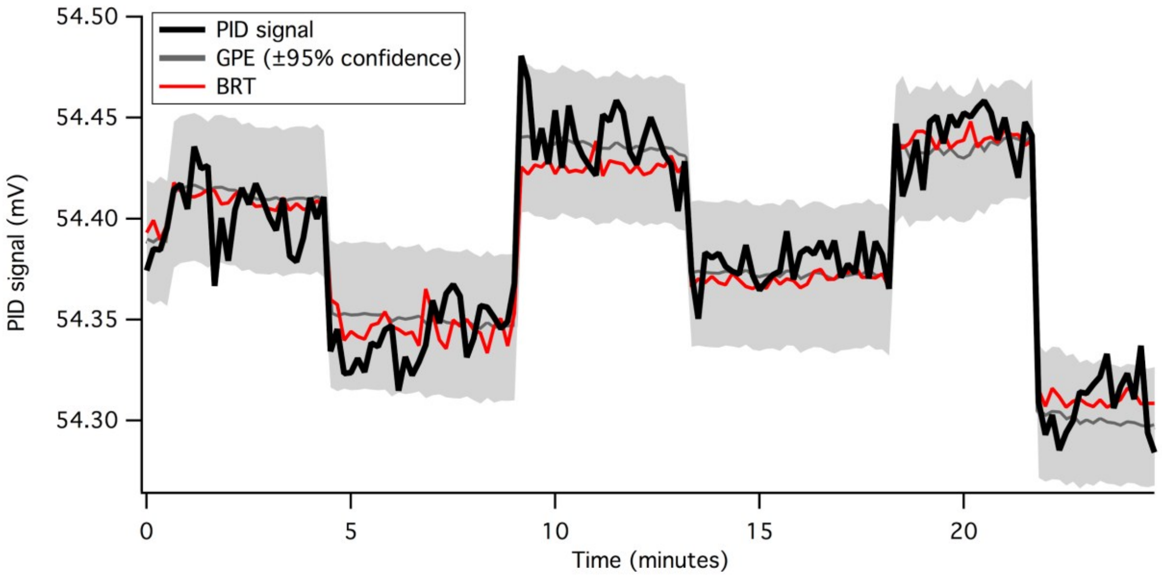


Figure 11. Raw PID and signals corrected for humidity, temperature and supply voltage using boosted regression tree (red) and Gaussian processes emulation (grey), for six different isoprene abundances in N_2 mixtures.

One approach to characterising the multiple dependencies of a PID sensor signal could be to use these statistical methods to develop models of sensor response. For PIDs the four key variables that describe the majority of sensor output are [VOC], RH, temperature and supply voltage. Figure 11 shows the raw sensor signal output (solid black line) compared with that calculated using BRT (red) and GPE (grey, plus shaded uncertainties). Both models were constructed using approximately 3 hours of training data and then used to predict a single sensor response based on a new set of observations of step-changing isoprene mixing ratio, relative humidity, temperature and sensor supply voltage. The model skill in reproducing the average sensor signal implies the majority of the observed variability in

---

[30] D.C. Carslaw and P.J. Taylor, *Atmospheric Environment*, 2009, **43**, 3563-3570.
[31] J. Elith, J.R. Leathwick and T. Hastie, *Journal of Animal Ecology*, 2008, **77**, 802-813.
[32] O. Roustant, D. Ginsbourger and Y. Deville, *Journal of Statistical Software*, 2012, **51**, 1-55.

this experiment can be explained by the variability in the 4 predictor variables. At ambient atmospheric concentrations [VOC] can in practice be only a minor contributor to the total signal response, and smaller than the variability introduced by the other three variables. Once confidant that the model describes all factors controlling sensor response, these approaches can then be used to predict desired response variables. Figure 12 shows the known isoprene mixing ratio (black line) compared with that calculated *via* BRT (red), GPE (grey). Also shown is the calculated mixing ratio using individual parameter regression (blue), clearly poorer than the other two approaches. The short duration of the dataset meant that RH, temperature and supply voltage variability was small compared to isoprene, and thus all 3 methods agree reasonably well and all have a mean deviation from the true isoprene of $< \pm 1.5$ ppb. However, as the quantity and dynamic range of interference parameters increases, the potential of these approaches to diagnose desired variables becomes much more significant.

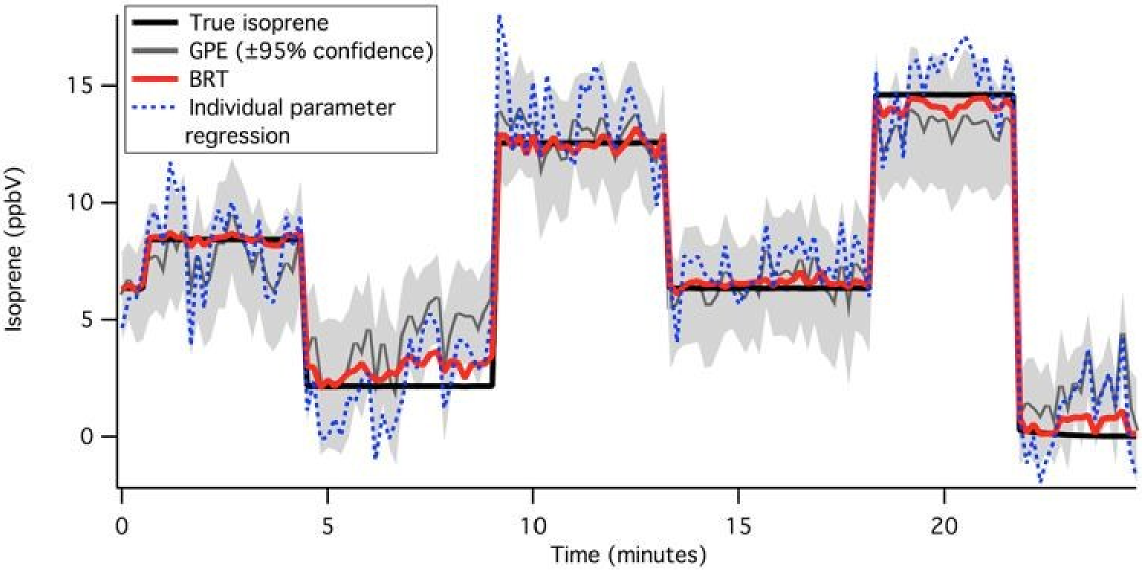


Figure 12. Isoprene mixing ratio based on calculated calibration gas output (solid black line), PID signal corrected using individual regression parameters (blue dashed), Boosted regression tree (red) and Gaussian processes emulation (grey, with shaded uncertainties)

Through the combination of these statistical models and the generation of comprehensive training data sets, using both laboratory and real-world experimental data, is possible that meaningful information can be retrieved from complex sensor signals, providing supporting observations of the major interferences are available. The success of these methods is, however, dependant on both the sensitivity and reproducibility of the sensor response as well as the quality and scope of the training data used.

Conclusions

We conclude that for all air pollution sensor systems a full analytical evaluation of sensor performance is required using both real-world pollutant abundances *and* in the presence of a full suite of realistic co-pollutants and interferences. Such evaluations are currently lacking in the literature. We highlight that in absolute terms the interference from stable trace gases (e.g. CO_2, H_2 etc), expressed as mV ppb^{-1}, are generally very small, but that a high ratio of such co-pollutants to the measurand can render the sensor observations prone to large artefact responses. A short-term evaluation of twenty commercial air pollution sensor packages provided some mixed results. For ozone a good overall agreement was found between reference measurements and sensors, once sensor calibration values had been linked to the *in situ* reference observation. Some challenges remain however in designing deployment and data strategies that can identify data from outlier sensors that deviate from the reference by more than the scale of expected atmospheric variability in urban environments. For other air pollution parameters, varying degrees performance were found. For NO and PM, general trends in atmospheric pollution were recreated by the sensors, although with notable negative and positive biases compared to reference chemiluminescence or TEOM-FDMS. NO_2 sensor measurements from the commercial package did not track trends in reference NO_2, and this may be responding to some different air pollution metric that requires better definition. VOC sensors showed good unit to unit variability, but present a further challenge in defining firstly, what exactly they are responding to, and secondly, how this non-specific data may be used scientifically. We show that for PID-based VOC measurement, detection limits can be improved through electronic handling of signals, but that multiple interfering factors affect the overall VOC response. In some cases the sensor response to [VOC] is smaller than that rising from variability in humidity, temperature and supply voltage. We identify that when multiple signal corrections are required for a sensor then application of individual regression parameters is less successful than the statistical methods of boosted regression tree or Gaussian process emulation. Overall we conclude that the state of the art can at best be considered a 'work in progress', with promising performance for some chemicals and rather poor performance for others. Very extensive trace-level laboratory testing of multiparameter sensor responses is still required along with alternative numerical strategies than can correct for sub-optimal measurement. Only then will the sensor concept yield data that is reliable, fit for purpose in the public domain and potential useful for research applications.

Acknowledgements.

ACL, JDL, SJM, and MJE thank the National Centre for Atmospheric Science for support. ACL also acknowledges support for this work through NERC grant NE/F015240/1. PE acknowledges the financial support of a Marie Sklodowska-Curie fellowship. KS is currently in receipt of a NERC SPHERES DTP PhD studentship, ME and SG received support through NERC and EPSRC summer research studentship programmes. We grateful for funding provided by the University of York for some sensor devices tested in this study in a project lead by Prof Alistair Boxall.