
Article:
Smith, Katie R., Edwards, Peter M. orcid.org/0000-0002-1076-6793, Evans, Mathew J. orcid.org/0000-0003-4775-032X et al. (5 more authors) (2017) Clustering approaches to improve the performance of low cost air pollution sensors. FARADAY DISCUSSIONS. pp. 621-637. ISSN 1364-5498

https://doi.org/10.1039/c7fd00020k

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Clustering approaches to improve the performance of low cost air pollution sensors†

Katie R. Smith[a], Peter M. Edwards[ab], Mathew J. Evans[b], James D. Lee[a], Marvin D. Shaw[a], Freya Squires[a], Shona Wilde[a], and Alastair C. Lewis[b]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Low cost air pollution sensors have substantial potential for atmospheric research and for the applied control of pollution in the urban environment, including more localized warnings to the public. The current generation of single-chemical gas sensors experience degrees of interference from other co-pollutants and have sensitivity to environmental factors such as temperature, wind speed and supply voltage. There are uncertainties introduced also because of sensor-to-sensor response variability, although this is less well reported. The sensitivity of Metal Oxide Sensors (MOS) to volatile organic compounds (VOCs) changed with relative humidity (RH) by up to a factor of five over the range 19-90%RH and with an uncertainty in the correction of a factor two at any given RH. The short-term (second to minute) stabilities of MOS and electrochemical CO sensor responses were reasonable. During more extended use inter-sensor quantitative comparability was degraded due to unpredictable variability in individual sensor responses (to either measurand or interference or both) drifting over timescales of several hours to days. For timescales longer than a week identical sensors showed slow, often downwards, drifts in their responses which diverged across six CO sensors by up to 30% after two weeks. The measurement derived from the median sensor within clusters of 6, 8 and up to 21 sensors was evaluated against individual sensor performance and external reference values. The clustered approach maintained the cost competitiveness of a sensor device, but the median concentration from the ensemble of sensor signals largely eliminated the randomised hour-to-day response drift seen in individual sensors and excluded the effects of small numbers of poorly performing sensors that drifted significantly over longer time periods. The results demonstrate that for individual sensors to be optimally comparable to one another, and to reference instruments, they would likely require frequent calibration. The use of a cluster median value eliminates unpredictable medium term response changes, and other longer term outlier behaviours, extending the likely period needed between calibration and making a linear interpolation between calibrations more appropriate. Through the use of sensor clusters rather than individual sensors existing low cost technologies could deliver significantly improved quality of observations.

1 Introduction

Low cost sensor networks are an appealing prospect for use in atmospheric chemistry research in particular offering the potential to greatly increase the spatial resolution of observations. Improved spatial observations can support improved source apportionment, improved validation of emission and transport models and give better estimates of human exposure[1]. Arrays of air pollution sensors are now being deployed in both indoor[2,3,4,5] and outdoor[6,7,8] environments and there is increasing confidence in the quality of observations generated[4,7,8].

[a] Wolfson Atmospheric Chemistry Labs, Dept. of Chemistry, University of York, York, YO10 5DD UK. Tel: 01904 324178; E-mail: ks826@york.ac.uk
[b] National Centre for Atmospheric Science, University of York, Heslington, YO10 5DD.

Technical approaches to using low-cost sensors tend to focus on incorporating different types of individual sensors into compact self-contained packages[3,6,7,9] in an attempt to monitor multiple pollutants simultaneously in a manner analogous to reference air quality monitoring stations. The inclusion of multiple different sensors in a single package is also used to support the self-correction of any cross-interferences between sensors. The most advanced air quality sensor packages have shown good capability to recreate general patterns of pollution behaviour when co-located next to, and compared with, reference instruments[1,6,7] and they have notable skill in detecting individual pollution events[7,10].

Determining absolute concentrations is considerably more challenging when sensors are deployed outside the laboratory environment, since they have substantial sensitivity to surrounding environmental conditions and do not normally have access to in-service reference materials for calibration[1]. The assumption used in most sensor deployments is that once initially calibrated, a sensor will maintain its response characteristics for long periods, much like a thermocouple. Sensors typically have unique sensitivities towards both the target gas and towards cross interferences and raw signal data is often processed using multivariate regression models[6] and pattern recognition analysis[2,11] to correct for the multiple variables which impact sensor signals[7,10]. Experiments have shown that air flow[12], temperature[10,12], humidity[4,9,12], exposure to other atmospheric gases[7,8], drift over time[6,8] and sensor arrangements[7] cause signal variability[6,13], lowering sensor reproducibility and impacting on data quality[11]. Approaches to improving the stability of sensor response include using temperature steps throughout a sample period[4] and noise reduction using well-designed circuit boards[7]. However for an atmospheric instrument that will not have access to calibration materials during its period of operation, an understanding of how sensor responses change over the full range of timescales for which data may be collected and used is fundamental.

Metal oxide sensors (MOS) provide a continuous measure of total volatile organic compounds (VOCs) in air and have sufficiently fast responses to identify pollution events on the second timescale[3]. Used in the laboratory, in combination with calibration models and multivariate regression it is possible to differentiate and semi-quantify VOCs[5,14,15]. Several studies have shown that MOS sensors often exhibit nonlinear responses towards VOCs[3,5,8], although their response does become linear at VOC concentrations below 100 ppb, a value reasonably representative of ambient air[1,5]. Nonlinear relationships exist between MOS VOC sensitivity and other variables, for example RH and temperature, further complicating calibration and ambient use. Sensitivities and response times can be different for notionally identical sensors[3,14,16] therefore transferring a single calibration model from the laboratory to deployment, and indeed from sensor to sensor is difficult[5]. There is an emerging literature showing that corrections for both chemical and environmental factors can be improved using more complex statistical models[13] that go beyond simple linear regressions, for example, using Partial Least Squares[14], neural networks[17] or Gaussian process emulation[1]. These techniques have shown improvements in the extraction of VOC concentration data from sensor signals with response drift, cross interference and sensor to sensor variations[7,10]. Training data for these processes is improved by including laboratory calibrations as well as real world ambient data[17].

This paper establishes the variability of response characteristics in a VOC MOS and CO electrochemical sensor, both applied for ambient measurements. Both are relatively inexpensive technologies[18] suitable for use in high-density networks. We evaluate inter-sensor variability, whether this is systematic or randomised across a population of sensors, and the time constants for change. From this a "clustered" approached is developed and a minimal number of sensors required to generate an improved median concentration is established. By using fundamentally low cost components, even in clusters of >20 identical sensors the conceptual capital cost advantage is maintained, but data quality improved.

2 Experimental

2.1 VOC detection with metal oxide sensors (MOS)

Sensor systems for VOCs have a particular attraction since existing measurements are very sparse due to the expense and practicalities of using gas chromatography or mass spectrometry in the field. Total VOC as measured by a MOS is an operationally defined value representing a bulk or "total VOC" concentration and is not easy to directly compare against existing reference measurements or standards. The value for research of a total VOC sensor measurement is likely to be associated with the mapping of geographic distributions and comparison of temporal behaviours, but this can only usefully be derived if sensor devices are highly reproducible amongst themselves. Figaro TGS2602 MOS are applicable for use in air pollution sensing systems because they are sensitive towards VOCs (at the ppb level), they are small (8 mm diameter), commercially available, inexpensive ($\sim$£10) and require low power and simple electronics to function[19]. The sensor requires a circuit voltage of 5 $\pm$ 0.2 V and a separate heater power supply that provides 5 $\pm$ 0.2 V to the integrated sensor heater (Figaro TGS2602 datasheet). Reducing compounds such as VOCs adsorb onto the sensing surface, which is a tin dioxide n-type semiconductor mounted on an alumina substrate[20]. The reducing compounds are oxidized by oxyanions on the surface of the sensing material, with electrons that were previously drawn to the oxygen then available to conduct. The change in surface conductance, and hence resistance - which is measured as the signal - is proportional to the concentration of VOCs in the sur-

rounding environment[20]. As is the case with other MOS devices they were initially used in applications such as leak detection for refrigerators and industrial safety, with use for ambient air quality monitoring a later, and more challenging, application (Figaro datasheet). In this study sensors are operated and used in clusters rather than each sensor having a separate set of supply circuits and data capture. Eight identical Figaro TGS2602 MOS from a common batch are mounted in a circular pattern on a single printed circuit board (custom designed). The overall dimensions of each sensor cluster are 90 mm (width) x 120 mm (length) x 30 mm (height). Change in individual sensor resistance is measured as a voltage across a load resistor in series with the sensing surface, and then an analogue to digital conversion prior to signal communication to a microprocessor (Arduino Uno) device. This load resistor is variable to enable all sensor signals to be "zeroed" to a common voltage at a baseline concentration reading. Data acquisition occurs at 1 Hz in all experiments described here.

2.2 Carbon monoxide detection with electrochemical sensors

Carbon monoxide has declined as a pollutant in many countries but it remains a very useful observation for air pollution research since it is an excellent tracer of combustion processes and often valued in particular for testing of model performance. Classical instrumentation for CO detection is based on IR absorption, GC, or cavity enhanced methods, but all are relatively costly and so the attraction of a sensor based approach is clear. The carbon monoxide (CO) electrochemical sensors used here were supplied by Alphasense Ltd (part no. CO-B4). This type of sensor is currently found in several commercial air pollution sensor packages. Each CO sensor contains three electrodes; the working (sensing) electrode (WE) is located closest to the surrounding environment and contains a high surface area electro-catalyst to optimize the oxidation of CO, the counter electrode (CE) lies underneath the WE and a wetting filter and the opposing redox reaction occurs here to generate an equivalent current[7], and a reference electrode is also included to ensure the WE potential remains in a suitable range (Alphasense CO-B4 specification sheet). The CO electrochemical sensors come with an Individual Sensor Board (ISB) which has been specifically calibrated for that sensor. The preset correction factors to convert the voltage output signal from the two electrodes to a concentration is therefore suitable for that sensor, on that ISB. The sensors are 32 mm in diameter and the mounting board is 39.1 mm x 44.6 mm x height 30 mm with the sensor mounted onto the board. The CO electrochemical sensors require a low-noise 3.5 to 6.4 V power supply, and the signal is in the form of voltage output from the auxiliary and the working electrodes.

2.3 Sensitivity and variability of MOS sensors to environmental parameters

Previous work has shown that some low-cost sensor technologies are prone to multiple cross sensitivities from both chemical and physical parameters[6,7,8,9,10,12,13]. These cross-sensitivities have the potential to produce significant signal interferences when measuring in ambient air, particularly when the target measurand is at low concentration and the interference is abundant or highly variable. Lewis et al. 2016 showed an example of NO_2 / CO_2 interferences highlighting that although the absolute sensitivity of a NO_2 sensor to CO_2 was low, the high abundance of CO_2 in air made this an important consideration at NO_2 mixing ratios below 20 ppb[1]. Ambient humidity is one of the more straightforward environmental interferences that can be evaluated and potentially corrected for, since it can be measured alongside the sensor at relatively little extra cost or complexity. Whilst collecting the data is easy, using this to create a correction is less so, particularly if the effects are nonlinear or not reproducible. As an example, water vapour has the effect of both changing MOS sensor sensitivity to VOCs and also generates an artefact signal in its own right. Figure 1 shows the impact of relative humidity on the sensitivity to VOCs of a set of 16 identical MOS (Figaro TGS2602) - expressed as mV $ppb_{[VOC]}$. A controlled mixture of gas phase VOCs (gas cylinder mixture of pentane, heptane, toluene, ethyl benzene, nonane and m+o-xylene at 5 ppm in N_2 gas) and humid air are presented to two clusters. The water vapour content in the gas stream is controlled by a DG3 Dewpoint Generator, and the experiments run using a computer controlled set of mass flow controllers to blend and dilute the gases.

Figure 1 shows only the nonlinear effect of relative humidity on VOC sensor sensitivity; in addition there is a direct sensor signal response to water itself (typical value -7 mV $\%^{-1}$). Figure 1 is also annotated with the humidity range encountered on a recent field deployment to Beijing, an experiment referred to later in more detail, showing the importance of accounting for these environmental dependencies. If an instrument is to be deployed to multiple locations across the globe it must be able to operate over the full range of relative humidity shown in this figure. For each humidity value tested in the lab the range of observed VOC sensitivity factors across the 16 sensors (housed in two clusters) is shown with bars on the y axis, varying considerably between individual sensors at any single humidity. Repeatedly testing this response at varying humidities yields a reproducible shape of curve for the population of 16 sensors as whole, but at an individual sensor level repeated tests do not give individually reproducible values. This has the consequence that no canonical sensor correction value can be applied, but only that a range of possible sensor response factors can be constrained, perhaps to a factor of two. We examine this further in the next section. In practical application for Beijing, and for a period of

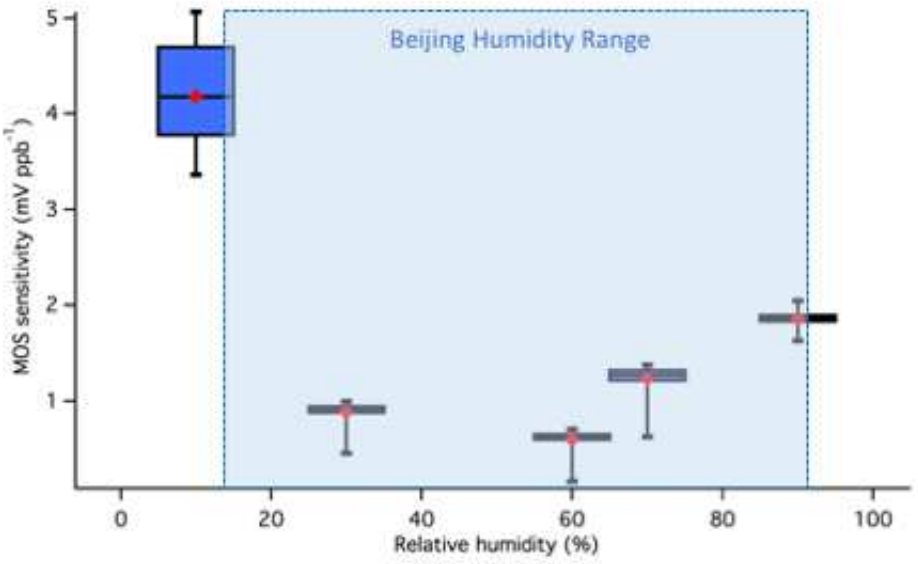

Fig. 1 The range of observed sensor sensitivities to gas phase VOCs as a function of changing ambient relative humidity; tests conducted on 16 identical MOS in two clusters under controlled laboratory conditions. The blue shaded area indicates the humidity range observed during a recent campaign to Beijing, China and this shows the importance of understanding how the surrounding environment impacts sensor behaviour.

just a few weeks of measurement, a relative humidity range from 17% to 90% was observed over which an individual sensor sensitivity towards VOCs would vary over a range of roughly a factor of five, and with an uncertainty of around a factor of two at any given humidity. These experiments do not identify why the response of a given sensor appears to vary in identical humidity conditions, but they highlight that interference effects on individual sensors are not necessarily single fixed values, but at best can be constrained to some relatively broader range of values. The scale of the impact of an environmental factor such as RH means that in-service calibration of interference responses is at least as important as calibration of the measurand itself, and definitely cannot be assumed to be a constant. A lack of reproducibility in how each individual sensor responds to an interference is clearly a crucial limiting factor when these sensors are then translated into the field. The timescales for any changes in the effects of interferences are therefore central to defining how often a sensor must be calibrated, which in turn has large implications for how sensor networks are delivered. We next examine the extent to which environmental interferences on MOS sensor responses can be removed, and test this over a range of timescales.

2.4 Variability of MOS sensor response in zero air

Figure 1 showed the impacts of a single environmental variable on the sensor response. By placing clusters of sensors in a single flow cell flushed with zero air the effect of the measurand (or lack of) can be held constant whilst multiple other factors such as temperature, supply voltage and humidity are allowed to vary within a certain range. By testing multiple sensors at the same time some insight can be gained into the degree of reproducibility between sensors to a commonly experienced interference, and whether these effects are stable over a given timescale.

Figure 2 shows a MOS cluster of eight TGS2602 sensors measuring zero air, free from VOCs, generated using a Pure Air Generator, a device tested previously for VOC content using GC methods. The response from the cluster of sensors was monitored over a period of c.a. 6 hours in an environmentally controlled laboratory, but where ambient humidity and temperature did vary to some small degree. A gas calibration sensitivity of 1 mV ppb$^{-1}_{[VOC]}$ (Figure 1) was used to convert the sensor signal into a VOC mixing ratio. Figure 2a shows the raw sensor signals, offset to zero at t = 0, as a time series and probability density function (pdf) for each of the 8 sensors; this shows that in this experiment sensor signals varied over a range that would equate to approximately 10 ppb [VOC]. Classically a plot of this kind would be used to help infer some measure of the sensitivity of the instrument. However, the observed timeseries variability is not random, with all sensors following some similar trends over the six hours, shown by the solid black median value line. This component of the observed variability that is common to all 8 sensors is not due to random signal noise, but instead likely due to small variations in temperature and humidity during the experiment - environmental interferences. Subtraction of the median sensor signal from all sensors considerably reduces the observed pdf spread (Figure 2b), with

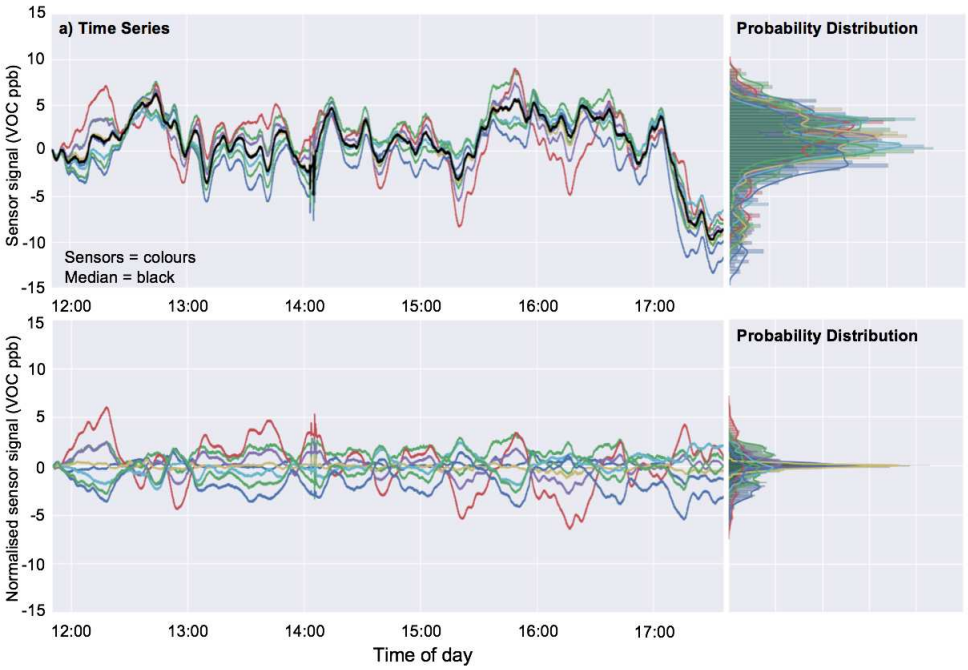

Fig. 2 a) Normalized time series for 8 identical MOS sensors in zero air over a six hour period (median = black line), and b) the same observations but with the median signal subtracted to eliminate contributions to individual sensors signals from external interferences such as temperature and humidity.

the remaining variability consisting of an individual sensor noise component and a residual environmental interference component due to differences in sensor sensitivities to the environmental parameters.

Plots of the probability density functions for each of the sensor signals in Figure 2 centered on the individual sensor mean are shown in Figure 3, both with (a) and without (b) the subtraction of the median sensor signal, showing that all sensors show a similar variability around the mean and that the major variability is driven by factors common across all sensors (as per Figure 2). Figure 3c shows a Fourier transform power spectrum of a single example sensor signal from Figure 2a, showing characteristics of a pink noise spectrum, with the most power in the lower frequencies (slope of approximately -1). Pink noise is more troublesome that white noise, which would show a flat power spectrum, as the application of smoothing and low-pass filtering to reduce noise is less effective for pink noise than for white noise. A given standard deviation of pink noise will have a greater effect on the accuracy of a measurement than the same standard deviation of white noise. If the observed pink noise is a characteristic of the sensor, then there should be no correlation between the different sensor signal noise components. Figure 2a, however, shows that much of the low frequency variability (minutes timescale) does show correlation across multiple sensors. This suggests that this low frequency variability is due to variations in environmental factors, such as laboratory temperature that typically vary on these timescales, which all sensors have a response to. Removal of the median sensor signal from each individual sensor signal should thus reduce the low frequency power in the power spectrum. Figure 3d shows the power spectrum of the same sensor signal in Figure 3c after removal of the sensor median, and shows a reduction in the slope and a flattening of the spectrum above 0.01 Hz, resulting from a reduction in the low frequency power.

Autocorrelation plots are a commonly-used tool for checking randomness in a data set. This randomness is ascertained by computing autocorrelations for data values at varying time lags. If random, such autocorrelations should be near zero for any and all time-lag separations. If non-random, then one or more of the autocorrelations will be significantly non-zero. The effect of the sensor response to common factors can also be seen in the autocorrelation plots of the sensor signals (Figures 3e and f). These show an increase in the sensor-to-sensor variability between the raw sensor signal autocorrelations (Figure 3e) and the autocorrelation of the signals after subtraction of the median sensor signal (Figure 3f). The average time taken for a sensor autocorrelation to become uncorrelated (within 95% confidence of 0) is approximately 15 minutes prior to median subtraction, but reducing to approximately 10 minutes afterwards.

The conclusion that can be drawn is that the responses induced on each sensor by individual interferences do not change substan-

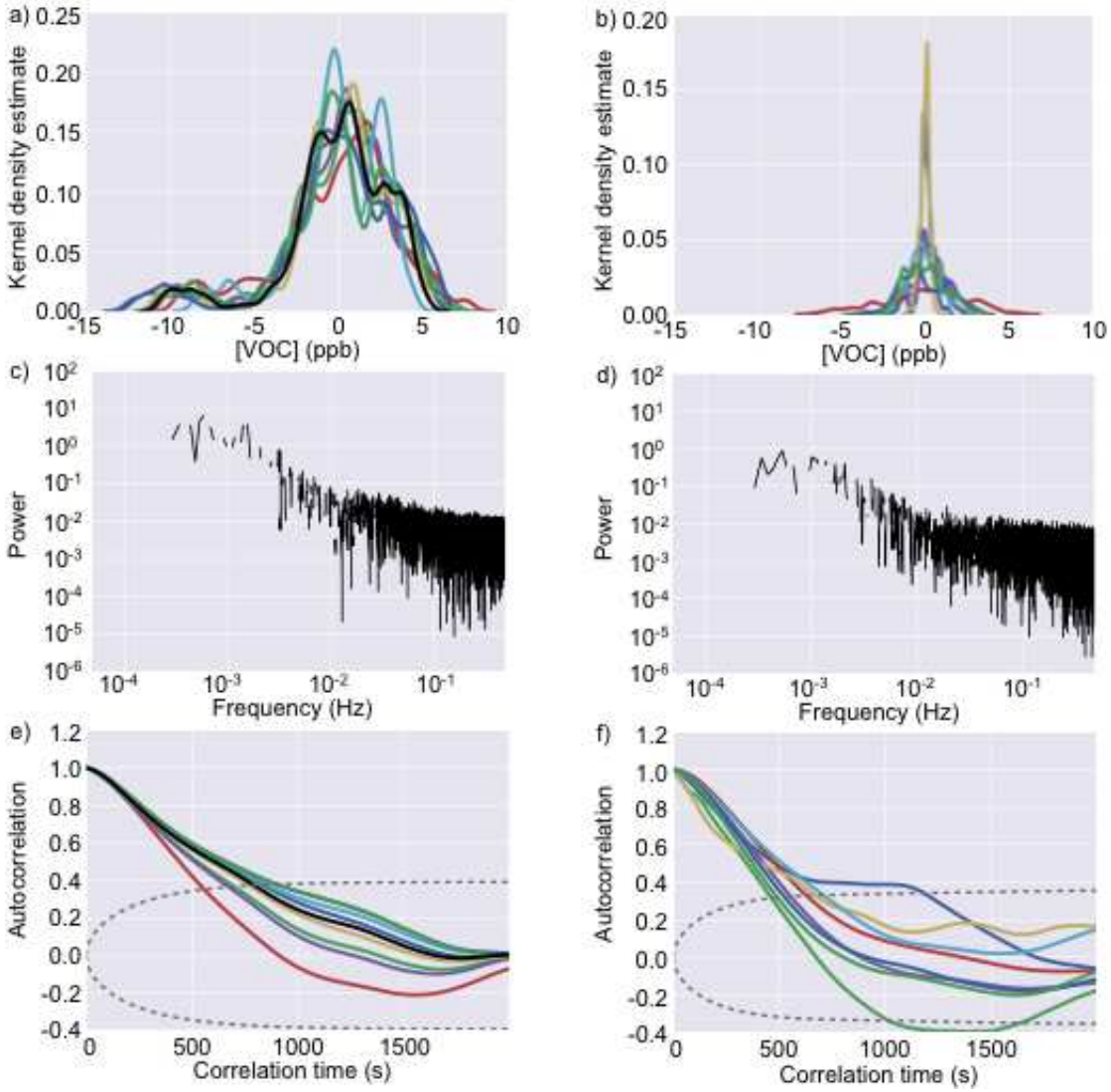

Fig. 3 The probability density, Fourier transform and autocorrelation plots for the MOS sensors in zero air. (a) shows the pdf of all 8 sensors (normalized to their individual sensor mean) showing overall distribution is driven by similar factors. (b) shows the pdf of all 8 sensors after subtraction of the median sensor signal (normalized to their individual sensor mean), showing differences in the individual sensor noise characteristics. Example Fast Fourier Transform of an individual sensor signal before (c) and after (d) subtraction of the median sensor signal. All sensors show red/pink noise characteristics but correlation timescale varies between sensors. (e) shows the autocorrelation of all individual sensor signals (colours) and the median sensor signal (black), and (f) shows the autocorrelation of the sensor signals after subtraction of the median signal. Grey dashed lines represent 95% confidence around 0. These plots show that all sensors show similar autocorrelation before subtraction of the median signal and become uncorrelated after 8-15 mins, after subtraction of the median signal the sensor signals show less autocorrelation, and become uncorrelated after 6-20 mins.

tially over timescales of seconds to a few hours. The subtraction of the median signal does not entirely remove the differences between individual sensor signals but the remaining differences can be considered as the noise for each sensor and the pdf is narrowed. Since repeated tests on different days for humidity effects (of the kind shown in Figure 1) did not support the use of a single response factor for individual sensors, it might be hypothesized that the MOS sensors are stable in terms of their response characteristics over a few hours, but possibly not over days or longer periods. Understanding this timescale is crucial, since this ultimately determines the required frequency of calibration of devices (both measurand and interferences), and the uncertainties associated with different data averaging periods. In the case of the sensor network deployment concept this time period between calibration will ideally be as long as possible.

2.5 The timescale MOS variability in ambient indoor air

The long term variability and drift of MOS clusters were next tested in ambient air with variable VOC content and over much longer periods than the laboratory experiments shown in Figures 2 and 3. Multiple clusters of eight TGS2602 sensors (making a total of 21 operational sensors) were used to passively sample air in a modern climate controlled indoor environment (~. 20°C) for a period of 20 days. An indoor environment was chosen since it provided variable atmospheric concentrations of VOCs, but a reasonably well constrained range of ambient temperatures and humidities as environmental interferences. The sensor signal voltages (normalised at time t = 0) are shown in Figure 4a. All sensors showed a good correlation (average inter sensor $R^2 = 0.923$) with one another through the period of test. There were clear daily cycles on weekdays and comparatively flat unchanging periods of VOC abundance on weekends, a result of low building occupancy and activity. In this regard the individual sensors worked well, all showing the same qualitative trends of changing VOCs or interferents indoors. The inter-sensor spread of observed values increased throughout the test period (standard deviation increasing from ~20 mV (~20 ppb VOC) on day 2 to ~100 mV (~100 ppb VOC) by day 19) as the sensor signals drift apart over the week+ timescale. The changes in responses of individual sensors cannot however be corrected assuming only a long-term linear drift (either upwards or downwards) trend for each individual sensor - an approach that would allow for a correction by linear interpolation between two calibration points. This becomes apparent in the rank-order plot shown in Figure 4b. This figure shows the ordering of the observations from the sensors, from the highest reporting value to the lowest. Early in the time series the sensor ranking of the sensors shows significant variability and changes regularly over time periods of around 6 hours to 1-2 days, particularly during periods where the mean sensor signal shows large changes. In simple terms, the highest reporting sensor is not always the highest and lowest not always the lowest, rather they change from day to day. On these day-long timescales the unpredictable changes in sensor sensitivity are possibly due to VOC sensitivity change or changes in sensitivity to interferent chemicals or physical parameters. This we refer to as medium term drift. Significant longer-term (more than one week) MOS drift means that towards the end of the time series less change in sensor rank is observed since the responses have begun to separate. Over several weeks there is substantial drift apart of the individual sensor outputs, to a value (100 ppb VOC) likely to be greater than the quantity itself being measured in air. In a practical deployment the trends shown by each sensor could have some value, but little quantitative comparison could be made between them. Medium term drift is superimposed on the long term drift and this also degrades the quantitative comparisons between sensors. On short timescales (6 hours) sensors appear to hold their response characteristics quite well, implying that over this timescale relative values between sensors could be directly compared.

Time averaging sensor signals is a way to address short term random noise, but this does not handle medium term semi-randomised drift in either sensor sensitivities to measurand or interferences. Correcting for changing response characteristics in an individual sensor can of course be achieved through regular and extensive multivariate calibrations in the field, as is applied to traditional instruments. The practicality of such complex calibrations appears rather at odds however with the conceptual model of low cost sensor science and the potential applications of low cost sensors (e.g. air quality monitors in developing nations, or for the general public). Based on these experiments indoors an approximate calibration frequency of perhaps once per day would appear reasonable if a quantitative comparison was required between individual sensors operating over multiple days or longer. Once the timescales for variability and drift are known, potential approaches to address and minimise this start to emerge. As is alluded to by the experimental design of the MOS clusters, one possible method is the use of multiple sensors whose signals are used as an ensemble in a single device. The use of larger numbers of sensors together still exploits the overall low cost of components, but minimises the impacts of medium term variability in any single individual sensor. Figure 5 shows the impact of averaging the individual sensor signals from Figure 4 on the correlation slopes calculated for unique permutations of MOS sensors against the median signal of all 21 sensors. Although inter-sensor MOS signals show a large degree of variability, the standard deviation of this is found to decrease with approximately 1/N, and providing sufficient sensors are used, the median sensor signal should be less sensitive to individual randomised drift. Thus, calibration factors applied to the cluster as a whole may be more robust than any individual sensor calibration, and applicable for longer periods of time between calibrations. This is because the necessity

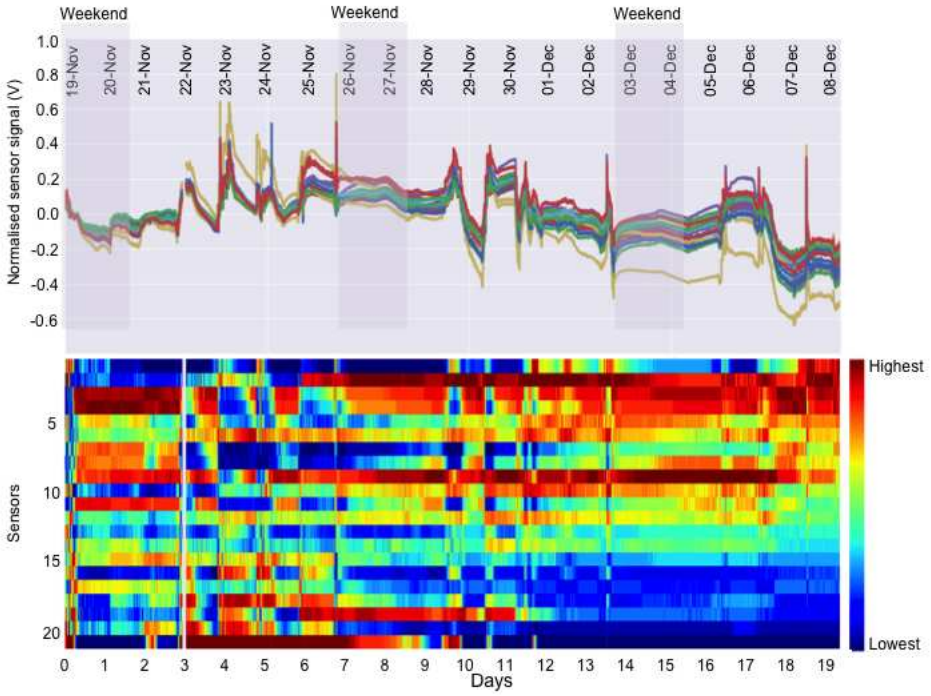

Fig. 4 a) The time series of 21 MOS sensor signals, normalised at t = 0, monitoring indoor air in a climate controlled building. b) Rank order plot of MOS sensor observations, y-axis sorted according to sensor ranking in the final hour of time series.

of frequent calibration when individual sensors are used alone is to accommodate and correct for outlying sensors, which are of course automatically excluded through selection of the median. Over week+ periods inevitably sensor response to the measurand begins to decrease (as with most instruments), but through the use of clustering the variability of individual sensors on the day timescale is removed and poor performing (or rapidly degrading sensors) over longer timescales do not influence the result. The remaining systematic decline in response can then more reasonably be corrected for via linear interpolation between relatively infrequent calibrations.

2.6 A clustered approach to CO sensor measurements

The previous section shows the potential improvements in quality of observation that may be achieved through taking an ensemble value from clusters of MOS sensors rather than a reliance on individual sensor outputs. However the real-world performance of the clustered approach is difficult to benchmark for MOS / VOCs, since there is no obvious equivalent reference measurement to compare against. In this section the clustered sensor concept is applied using Alphasense CO-B4 electrochemical sensors. By moving to CO as the measurand there is then an opportunity to compare individual sensor and cluster performance against a robust reference measurement. The observations are made during an experimental deployment at ground level (5 m) in central Beijing, China, for a period of 1 month. The choice of Beijing as a the test location exposes the sensors to a very wide range of both ambient concentrations and relative humidity. The cluster of CO sensors are housed in a 2 x 3 formation and mounted in a single machined flow cell. Ambient air is supplied to the sensor flow cell via a metal bellows pump, with flow rate throttled to 1.5 L min^{-1} using a KNF vacuum pump and an 1/4" needle valve. The sensor and flow cell are then housed inside a further enclosure and the device located in an air conditioned mobile laboratory, with dc supply provided via transformer and mains power. In this regard the CO sensors are placed in a far less challenging environment than is commonly applied to air pollution sensors - e.g. located outdoors, unregulated temperature, unregulated airflow, battery power supply etc. An Aero-Laser AL5002 vacuum UV instrument calibrated and zeroed every 9 hours against a BOC (1 ppm CO in air) standard provided reference measurements for two weeks of this test period.

Figure 6a shows the observations from each sensor in the cluster and the reference CO time series data. All reported CO sensor readings are calculated using the individual calibrations provided by the manufacturer, but are normalised to each other at the start of the time series in Figure 6a, after an initial 12 hour warm up period. An offset of 324 ppb was then calculated between the

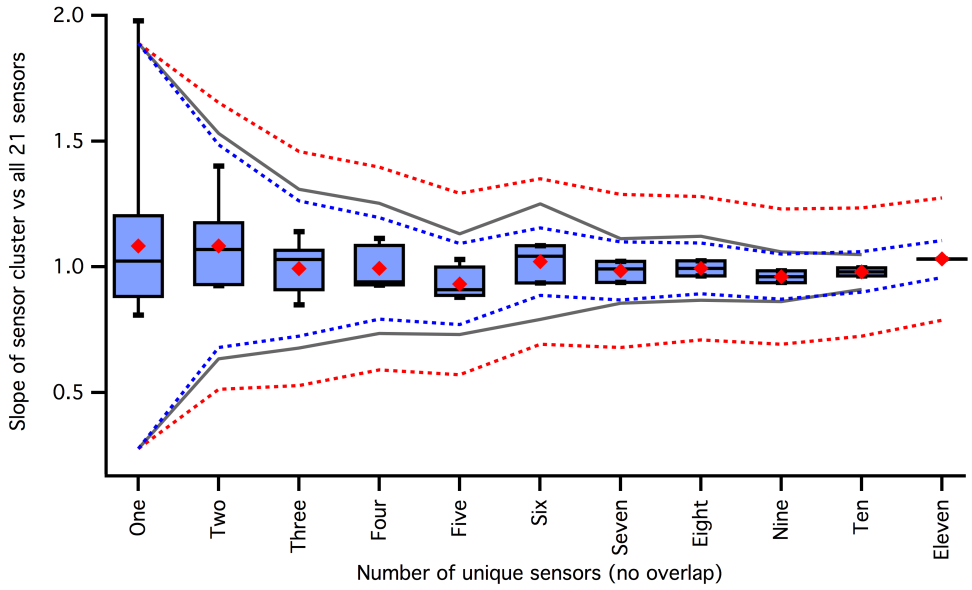

Fig. 5 The slope of sets of non-overlapping permutations of sensors against the median signal from all 21 sensors (i.e one = 21 slopes of individual sensors vs 21 sensor median, 2 = 10 slopes of averages of 2 sensors vs median of 21, 3 = 7 slopes of the median of 3 sensors vs median of 21) The red dots show the mean and the grey lines are $\pm$ 3 standard deviations on the mean. The dashed lines show calculated $\pm$ 3 standard deviations on the mean using a 1/square root of N decrease (red) and a 1/N decrease (blue) from the 1 sensor observation.

Aerolaser CO and the median sensor CO on the 22nd November and applied to all sensors, a value that was in line with an external zero air (PAG air) value from the sensors during the experiment.

There are several notable features in Figure 6a. The first is that the sensors qualitatively perform well against the reference and are very well-correlated both internally and with the reference instrument ($R^2 > 0.95$); it is clear that in terms of tracking temporal trends the CO sensors very perform well. There is however considerable divergence in the response of each sensor over the period of measurement. A rank ordering of the relative sensor signals is shown (Figure 6b) with the highest reporting sensor at any time point in red and the lowest in blue. Throughout the earlier part of the time series the ranking of the sensors changes frequently, particularly during periods where the reference CO shows large atmospheric changes. This is due to the sensors each responding to the different environmental conditions (for example temperature and humidity), with slightly different sensitivities to each parameter. Over time the individual sensors change their responses to different degrees, resulting in a significant spread in CO values across the 6 sensors ~370 ppb (or ~32%) between highest and lowest in Figure 6c. This long-term signal drift is clearer in Figure 6c which shows only the portion of the time series with reference CO measurements and all CO sensor offsets corrected to the reference at the beginning of the time period. Figure 6d shows the difference between each sensor and the CO reference measurement throughout the time period. The scale of drift would clearly be very significant if the individual sensors were spatially distributed and then a concentration gradient inferred between locations based on observations. Similar to the MOS sensors this longer-term drift between sensors results in less change in sensor rank towards the end of the period. Using a single sensor in isolation would imply that frequent calibration - perhaps of the order of daily would be the minimum needed to maintain optimal compatibility between sensors. The frequency of calibration would in practice need to adapt to the performance of the worst performing (largest drift) sensor, but without prior knowledge of which that was, all sensors would need to be subjected to the same regime.

If all sensors of a particular type drift with a symmetric distribution, then the use of all six CO sensors to derive a single CO mixing ratio offers the potential for a much more robust observation than any single sensor. If in practice there is both a random element (seen on the day timescale through rank order changes) and systematic change in response (likely downwards over time) then use of a cluster-derived value provides a means to extend the period between overall instrument calibration through exclusion of the outlier sensors. For the data shown in Figure 6, the R^2 value for the median sensor value versus the reference instrument (0.98) is higher than the average R^2 of each of the individual sensors. The impact of averaging over all permutations of multiple sensors on the calculated mean squared error (Figure 7a) and slope (Figure 7b) between the CO sensor cluster and the CO ref-

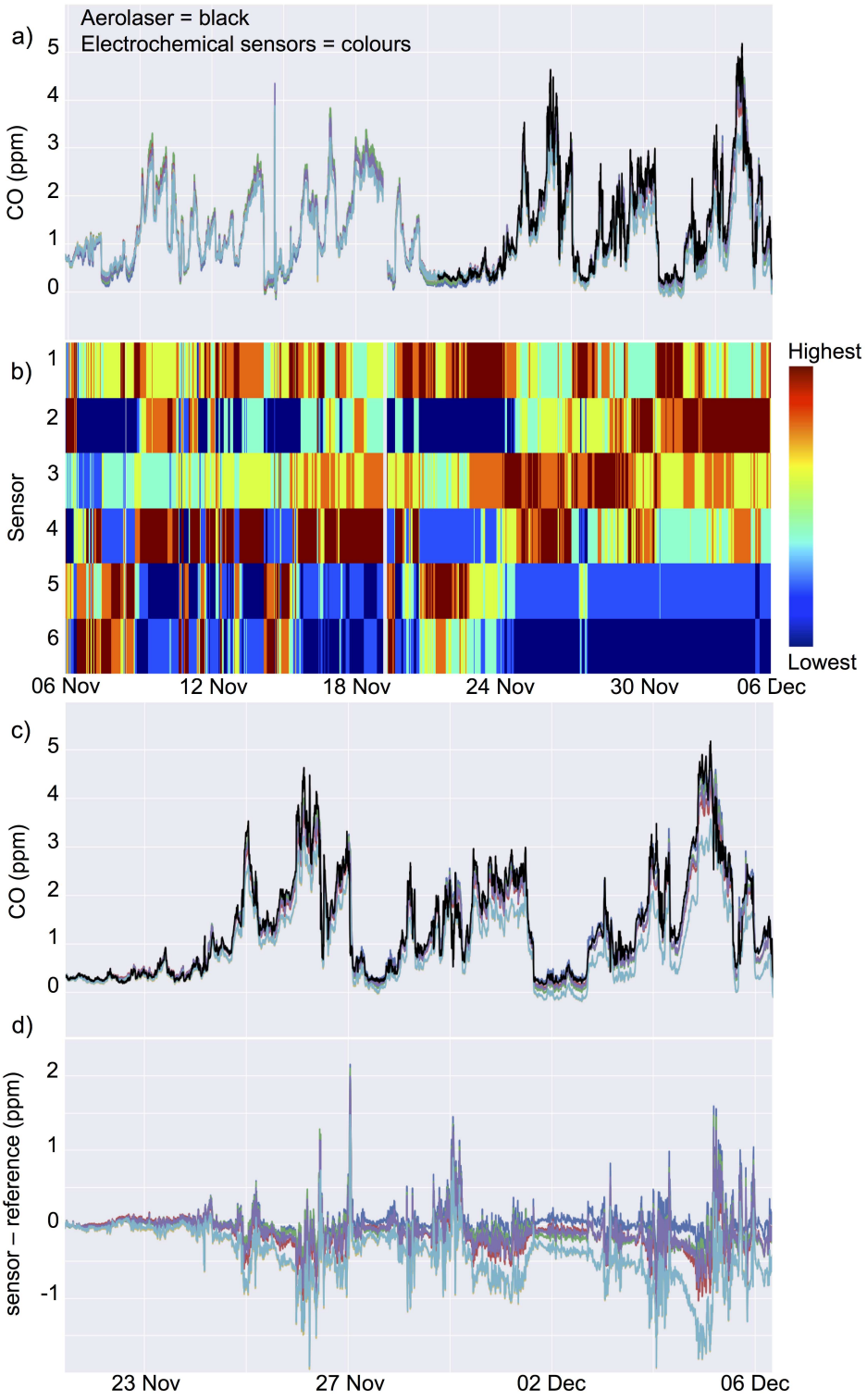

Fig. 6 a) Time series for the CO electrochemical sensors (coloured lines) and the reference CO measurement (black). b) A sensor rank plot for the 6 CO electrochemical sensors. The sensors are ranked according to their output, with the highest reporting sensor shown in red and the lowest in blue. The sensors frequently change ranking positions. c) Time series of period with both CO sensor and reference measurements. Sensors offset to match reference at t = 0 d) Individual sensor CO - reference CO.

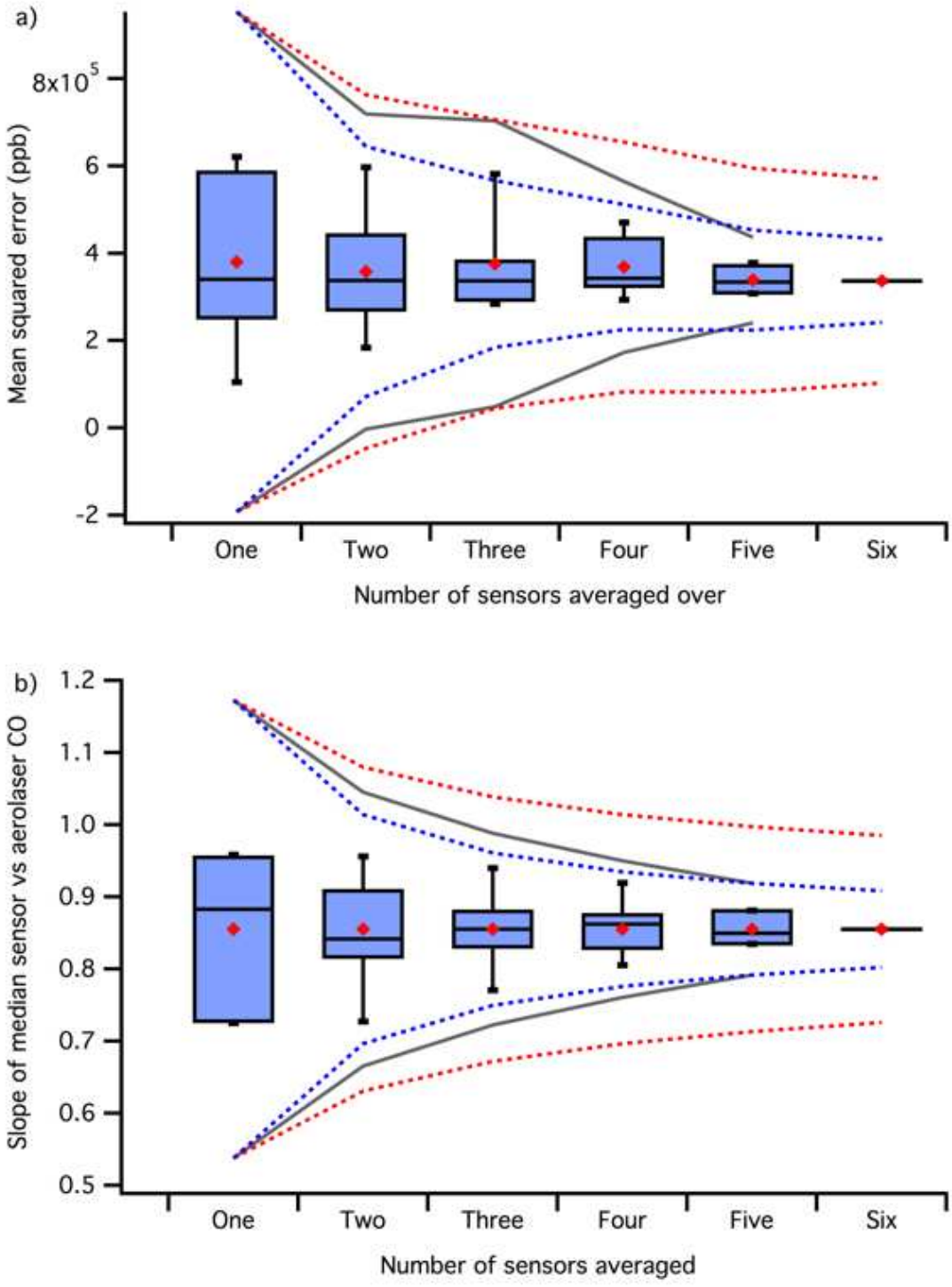

Fig. 7 Mean squared error (a) and slope (b) statistics for the correlation of the CO reference fluorescence instrument data shown in Figure 6c with sensor data calculated as the average of different combinations of sensor signals shown in the same figure, for every possible permutation of sensor combination. For example, the "Two" sensors averaged box plot shows the distribution of calculated mean squared error/slope of the average of two sensors against the reference instrument for all 15 possible combinations of two sensors from the set of six. The dashed lines show calculated $\pm$ 3 standard deviations on the mean using a 1/square root of N decrease (red) and 1/N decrease (blue) from the individual sensor versus reference instrument statistic.

erence measurement was calculated. Increasing the number of sensors that are averaged reduces the standard deviation in the calculated correlation metric, and with sufficient sensors should enable the characteristic sensor distribution to be determined and calibrated so as to minimise variability. The sharp decrease in correlation metric spread above 3 sensors averaged is due significant overlap between the permutations due to the small sample size (6 sensors). In this real-world test over time there is a gradual reduction in the sensor cluster's response to CO - at the start of the experiment (first two days 21^{st} to 22^{nd} Nov 2016) the median slope against reference is 0.997 and ten days later (3^{rd} to 4^{th} Dec 2016) it is 0.917, but the change is modest and linear for the median value. Whilst individual CO sensors would indicate the need for almost daily calibration (if they are to all individually match the performance of the reference) it might be reasonable for a cluster of 6 to be calibrated only perhaps once every 2 weeks. Since the overall cost of components in a 6-clustered CO device are around $500, we consider that this still falls with the general boundaries of low cost sensing, but delivers observations in an urban environment that are highly competitive with a reference instrument that was more than a factor of 100 more expensive.

3 Conclusions

Air pollution sensors are a high profile and rapidly growing field of technology with many attractive conceptual benefits that allow for increases the spatial resolution of monitoring networks and reduction in cost. As has been reported previously environmental interferences such as temperature and humidity on sensors can be significant and have nonlinear impacts. With MOS devices we show that the effects of an interference such as water vapour on a population of sensors can be large, changing the response of sensors to VOCs by a factor of five over a plausible range of atmospheric humidity conditions. The effects of the interference on each sensor was seen to vary significantly and were not constant when experiments were repeated over different days/weeks, with individual sensor responses varying by up to a factor of two for any given humidity. Over timescales of up to several hours the effects of environmental interferences could be removed from a data set through subtraction of a median sensor value when 8 identical MOS sensors were exposed to zero air. The mismatch between apparent 'correctability' of interference over minute to hour timescales compared to much more variable response factors over days and longer was tested using a 20 day controlled assessment of the performance of 21 identical MOS devices in ambient indoor air. The MOS showed a high degree of inter-sensor correlation throughout the whole period, but over timescales of >6 hours to around 1-2 days individual sensor response factors (possibly to VOCs or interferences, this could not be discerned) appeared to vary, superimposed on longer-term (week+) drifts.

The long term changes seen in individual sensors did not appear amenable to simple linear interpolation for correction, and if all sensors were to be made quantitatively comparable would imply frequent calibrations. This variability in each sensor's response was reflected in the distribution of correlation slopes calculated for unique permutations of MOS sensors against the median signal of all 21 sensors. Using an ensemble value from an increasing number of sensors, the range of slopes was significantly reduced, indicating advantages in using a VOC measurement based on a cluster of sensors rather than single sensors. To allow a comparison of sensors against a well characterised reference instrument, a cluster of 6 CO sensors was tested against a vacuum UV instrument in central Beijing. In general terms the individual sensors were very well correlated to the reference measurement through 4 weeks of measurement. The apparent response characteristics for each sensor appeared to vary however in a similar way to the MOS with noticeable response variability over the > 6 hour to 1-2 day timescale. This variability in day to day response was visible with frequent changes in rank ordering of the sensor derived concentrations. This variability was similarly superimposed on a longer-term slight downwards drift in sensor responses, but by the end of 4 weeks the differences between the highest and lowest responding CO sensor were very pronounced (more than 32% different / equivalent to 370 ppb), the spread impacted particularly by one outlier sensor. A comparison of the median CO sensor value against the reference was however very good, and although this value displayed some long-term downwards trend against the regularly calibrated reference instrument, this would be correctable with interpolation between calibrations on 2-4 week timescales.

These two commonly used air pollution sensors can provide very good qualitative measures of pollution trends. The performance of sensors, as measured by voltage output response to measurand, and other interferences, is stable over short time periods, but can vary unpredictably on timescales of >6 hours to 1-2 days. Over longer time periods the responses of individual sensors drift in more systematic ways that may be correctable through linear interpolation, although the extent of long term drift varies considerably between sensors. The observed variability in performance of individual sensors over hours to a few day suggests that to achieve high precision comparability between sensors very frequent multi parameter calibration would be required. Through the use of a median value from an ensemble or cluster of sensors, the effects of medium term drift in individual sensors can be largely eliminated. By extension, use only of a median sensor value from a cluster of sensors then considerably lengthens the period between calibrations and makes linear interpolation between those calibrations potentially more applicable. The number of sensors required in a cluster will depend both on the inherent sensor-to-sensor variability in both signal noise components and sensitivity, as well as the required level of stability and accuracy

for the chosen purpose. Ultimately low cost air pollution sensors are no different to all other atmospheric chemistry instruments, with inherent requirements for in-service calibration, zeroing and an understanding of possible interferences. Applying a clustering approach could bring the current generation of sensor technologies closer to existing reference instruments in terms of their performance, but ultimately regular in field calibrations are still likely to be required for optimal comparability.

4 Acknowledgements

AIRPRO grant NE/N007115/1, AIRPOLL grant NE/N006917/1, NCAS/NERC ODA Foundation grant, PE acknowledges a Marie Skłodowska-Curie individual fellowship, KR and FS acknowledge NERC SPHERES DTP PhDs.

References

1 A. C. Lewis, J. Lee, P. M. Edwards, M. D. Shaw, M. J. Evans, S. J. Moller, K. Smith, M. Ellis, S. Gillott, A. White and J. W. Buckley, *Faraday Discuss.*, 2015, **189**, 85–103.

2 S. Zampolli, I. Elmi, F. Ahmed, M. Passini, G. C. Cardinali, S. Nicoletti and L. Dori, *Sensors Actuators, B Chem.*, 2004, **101**, 39–46.

3 A. Caron, N. Redon, B. Hanoune and P. Coddeville, *Build. Environ.*, 2016, **107**, 19–28.

4 M. Leidinger, T. Sauerwald, T. Conrad, W. Reimringer, G. Ventura and A. Schütze, *Procedia Eng.*, 2014, **87**, 1449–1452.

5 E. J. Wolfrum, R. M. Meglen, D. Peterson and J. Sluiter, *IEEE Sensors*, 2006, **6**, 1638–1643.

6 W. Jiao, G. Hagler, R. Williams, R. Sharpe, R. Brown, D. Garver, R. Judge, M. Caudill, J. Rickard, M. Davis, L. Weinstock, S. Zimmer-Dauphinee and K. Buckley, *Atmos. Meas. Tech. Discuss.*, 2016, 1–24.

7 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 and R. L. Jones, *Atmos. Environ.*, 2013, **70**, 186–203.

8 R. Piedrahita, Y. Xiang, N. Masson, J. Ortega, A. Collier, Y. Jiang, K. Li, R. P. Dick, Q. Lv, M. Hannigan and L. Shang, *Atmos. Meas. Tech.*, 2014, **7**, 3325–3336.

9 O. Helli, M. Siadat and M. Lumbreras, *Sensors Actuators B*, 2004, **103**, 403–408.

10 I. Heimann, V. B. Bright, M. W. Mcleod, M. I. Mead, O. A. M. Popoola, G. B. Stewart and R. L. Jones, *Atmos. Environ.*, 2015, **113**, 10–19.

11 H.-K. Hong, H. W. Shin, H. S. Park, D. H. Yun, C. H. Kwon, K. Lee, S.-T. Kim and T. Moriizumi, *Sensors and Actuators*, 1996, **4005**, 68–71.

12 K. L. Dorsey and A. P. Pisano, *IEEE Sensors*, 2016, **16**, 699–705.

13 L. Capelli, S. Sironi and R. D. Rosso, *Sensors*, 2014, **14**, 19979–20007.

14 E. J. Wolfrum, R. M. Meglen, D. Peterson and J. Sluiter, *Sensors Actuators B Chem.*, 2006, **115**, 322–329.

15 S. Zhang, C. Xie, D. Zeng, H. Li, Z. Bai and S. Cai, *IEEE Sensors*, 2008, **8**, 1816–1823.

16 S. Zhang, C. Xie, D. Zeng, H. Li, Y. Liu and S. Cai, *Sensors Actuators B Chem.*, 2009, **142**, 243–252.

17 M. Kamionka, P. Breuil and C. Pijolat, *Sensors Actuators B Chem.*, 2006, **118**, 323–327.

18 G. F. Fine, L. M. Cavanagh, A. Afonja and R. Binions, *Sensors*, 2010, **10**, 5469–5502.

19 N. Barsan, D. Koziej and U. Weimar, *Sensors Actuators, B Chem.*, 2007, **121**, 18–35.

20 G. Korotcenkov, *Mater. Sci. Eng. B*, 2007, **139**, 1–23.