Atmospheric Environment 228 (2020) 117428

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: http://www.elsevier.com/locate/atmosenv

Hierarchical network design for nitrogen dioxide measurement in urban environments

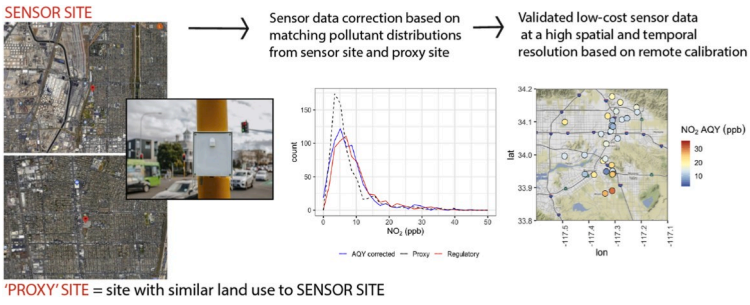
Lena Weissert [a,c,1], Elaine Miles [b], Georgia Miskell [a,2], Kyle Alberti [b], Brandon Feenstra [d], Geoff S. Henshaw [b], Vasileios Papapostolou [d], Hamesh Patel [b], Andrea Polidori [d], Jennifer A. Salmond [c], David E. Williams [a,*]

[a] School of Chemical Sciences and MacDiarmid Institute for Advanced Materials and Nanotechnology, University of Auckland, Private Bag 92019, Auckland, 1142, New Zealand
[b] Aeroqual Ltd, 460 Rosebank Road, Avondale, Auckland, 1026, New Zealand
[c] School of Environment, University of Auckland, Private Bag 92019, Auckland, 1142, New Zealand
[d] South Coast Air Quality Management District, 21865 Copley Drive, Diamond Bar, CA, 91765, USA

HIGHLIGHTS

- Shows how to get reliable data from an extensive low-cost network of NO_2 sensors.
- Shows drift correction methods for electrochemical sensors in hierarchical network.
- Demonstrates use of statistical comparison to trusted proxy.
- Land-use similarity used to select reference sites as proxies, which can be distant.
- Uncertainties low enough to resolve spatio-temporal variations over short distances.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Air quality sensors
Nitrogen dioxide
Low-cost sensor network
Calibration

ABSTRACT

We present a management and data correction framework for low-cost electrochemical sensors for nitrogen dioxide (NO_2) deployed within a hierarchical network of low-cost and regulatory-grade instruments. The framework is founded on the idea that it is possible in a suitably configured network to identify a source of reliable 'proxy' data for each sensor site that has a similar probability distribution of measurement values over a suitable time period, and that sensor data can be checked and corrected by comparison of the sensor data distribution with that of the proxy. The framework is rule-based and easily modified. We use the reference network to choose proxies and check proxy reliability. We demonstrate the application of this methodology to low-cost instruments that use an electrochemical NO_2 sensor together with a semiconducting oxide-based sensor for ozone (O_3). The three NO_2 sensor response parameters (offset, O_3 response slope, and NO_2 response slope) which are known to vary significantly as a consequence of ambient humidity and temperature variations, we show can be estimated by minimising statistical measures of divergence between sensor-estimated and proxy NO_2 distributions over a 3-day window. We show how the parameter variations and statistical divergence measures with

* Corresponding author.
E-mail address: david.williams@auckland.ac.nz (D.E. Williams).
[1] Present address: Aeroqual Ltd, 460 Rosebank Road, Avondale, Auckland 1026, New Zealand.
[2] Present address: Trustpower, 108 Durham St, Tauranga, New Zealand.

https://doi.org/10.1016/j.atmosenv.2020.117428
Received 7 November 2019; Received in revised form 14 February 2020; Accepted 21 March 2020
Available online 25 March 2020

respect to the proxy can be used to indicate error conditions. The major error is due to a diurnally-varying, spatially-correlated offset term that is large for extremes of temperature, which we show can be estimated through its spatial correlation, using sensors co-located at reference sites. With these procedures, we demonstrate measurement at nine different locations across two regions of Southern California over seven months with average root mean square error ±7.2 ppb (range over locations 4–11 ppb) without calibration other than the remote proxy comparison. We apply the procedures to a network of 56 sensors distributed across the Inland Empire and Los Angeles County regions. The results show large variations in NO_2 concentration taking place on short time- and distance scales across the region. These spatiotemporal NO_2 variations were not captured by the more sparsely distributed regulatory network of air monitoring stations demonstrating the need for reliable data from dense networks of monitors to supplement the existing regulatory networks.

1. Introduction

The question of reliability of data from low-cost sensors is contentious and difficult to address (Williams, 2019). An approach that uses independent information to support sensor data is promising. We present one such approach here, applied to measurement of nitrogen dioxide with electrochemical cells, that extends previously described methods for O_3 (Miskell et al., 2016, 2018). The ideas are grounded in the concept of believability of results based on general experience (Williams, 2019). Central to the procedure is the concept of a trusted proxy. In the present work, proxy results are from a device that is trusted (for example, because it is regularly calibrated) which it is believed has a similar probability distribution of data over a suitably chosen time to the device to be checked and corrected. The belief on data distribution similarity is based on general experience: for example land-use regression of pollutant concentration which shows that land-use variables can be chosen that are similar at sites with similar pollutant time-averaged concentrations.

Advancement in technology has resulted in the availability of low-cost sensors that can be used to collect real-time NO_2 data at a high spatial and temporal resolution (Snyder et al., 2013). When deployed in dense hierarchal networks, low-cost sensors offer an opportunity to collect neighbourhood-level air pollution data. They have been used to detect small scale variations (Mead et al., 2013) and discriminate emissions due to different activities and emission sources (Popoola et al., 2018). Thus, they have become a popular choice for community-based air quality networks and community science projects (Clements et al., 2017; Hubbell et al., 2018). However, uncertainties remain about the data reliability of low-cost NO_2 sensors largely due to drift and interferences with other pollutant gases and variations associated with changes in temperature and relative humidity (Isiugo et al., 2018; Lewis et al., 2016; Mead et al., 2013; Weissert et al., 2019). In an attempt to calibrate the sensors and assess their accuracy, sensors are typically co-located against a well-maintained regulatory reference instrument for a period of time before and after deploying them in the field (Isiugo et al., 2018; Sadighi et al., 2018; Weissert et al., 2019). This appears a suitable approach only for short term deployments, while long-term deployments would require ongoing re-calibration (van Zoest et al., 2019) leading to calibration and maintenance costs that may quickly exceed the costs of the instruments (Clements et al., 2017). In addition, this approach assumes that the calibration parameters obtained from the co-location of the low-cost sensors at a reference site are transferable to other locations in the sensor network. A recent study from a network of NO_2 sensors in Eindhoven, Netherlands has shown that the calibration coefficients could not easily be transferred from one location to another within a city likely due to drift and interference effects being different for individual sensors (van Zoest et al., 2019) and to significant time-variation of the individual sensor response parameters. One suggestion to overcome this problem is the use of a mobile reference sensor that is moved from one location to another for calibration, which would account for the spatial and temporal differences in the calibration parameters (Williams, 2019; van Zoest et al., 2019). However, the costs associated with this approach may quickly outweigh the benefits of the low-cost sensors particularly if they are deployed in dense networks.

In our previous work, we developed a semi-blind management framework to verify the reliability of low-cost sensor data using general knowledge of the sensor and pollutant. Consequently, we were able to demonstrate remote correction of low-cost sensors that are deployed in dense networks (Alavi-Shoshtari et al., 2013; Miskell et al., 2016, 2018, 2019). The management framework was tested using hierarchical networks, consisting of well-maintained regulatory-grade instruments and low-cost O_3 sensors deployed around the Lower Fraser Valley (LFV) in Canada (Miskell et al., 2018) and Southern California (Miskell et al., 2019). Data from the well-maintained regulatory-grade instruments were first used to determine suitable proxies across the region, and then to provide suitable proxy data to check for drift and if necessary apply a correction. We defined a proxy as a reliable source of data within the network but at a different location to the site of interest, whose data has a similar probability distribution (Miskell et al., 2018). A proxy site can be selected based on proximity or similar land-use (Miskell et al., 2016, 2019). Testing this approach in these two distinct regions, which differ considerably in terms of geography, traffic patterns, climate and population density, suggested that the approach is transferable.

The purpose of this paper is to extend the management framework to electrochemical sensors for NO_2, where the measurement model for the sensor is more complex than a simple 2-parameter model, and where interfering effects of climate variables are also complex. The proxy selection problem also seems more complex, because NO_2 is generally considered to be highly variable spatially and temporally.

2. Methods

In this section, after describing the study sites (2.1) and the low-cost sensors (2.2), we give the formal statement of the proxy model (2.3) and describe the proxy selection based on land-use criteria (2.4). We give a method to address the problem of assessing the reliability of proxy-corrected data using only the results from the sensors and the proxies in (2.5). In (2.6) we set out in detail the correction method for the electrochemical NO_2 sensor, addressing the particular problems that this presents.

2.1. Study sites

The study sites were distributed across the Los Angeles region (Fig. 1). There are five regulatory sites in the Los Angeles city ('LA') and four sites in the Inland Empire ('IE') which includes Riverside and San Bernardino Counties in Southern California (Fig. 1). The sites are equipped with continuous reference method Nitrogen Oxides (NOx) analyzers, which are regularly maintained and serviced by the South Coast Air Quality Management District (South Coast AQMD). Eight sites are equipped with a model 42i NOx analyzer by Thermo Fischer Scientific (Franklin, MA), while the Fontana site is equipped with a model 200E NOx analyzer by Teledyne Advanced Pollution Instrumentation (San Diego, CA). At each site, we had a low-cost instrument that measures O_3 and NO_2 (details below: model AQY, Aeroqual, Auckland, NZ). We used data from January–July 2018 for the IE network and from March–July for the LA network. Vehicle emissions, particularly from heavy-duty vehicles, are the main source of NO_2 in the LA region

(AQMP, 2016). Nitrogen oxides (NOx) are precursors to both O_3 and particulate matter (PM) and therefore of major concern for air quality management (AQMP, 2016). Measurements are mixing ratios: parts-per-billion (10^9) by volume (ppb).

2.2. Low-cost sensors

The low-cost sensors deployed in the Los Angeles network are the AQY v0.5 sensors from Aeroqual Ltd, Auckland, New Zealand. We use the term 'sensor' here to refer both to the instrument package (O_3, NO_2, T, RH and $PM_{2.5}$) and to the detection element. Ozone was measured using a gas-sensitive semiconducting (GSS) oxide, WO_3, as the detection element (Aliwell et al., 2001; Hansford et al., 2005; Utembe et al., 2006; Williams et al., 2002). Air flow-rate modulation and temperature modulation are used to cancel interferences due to water vapour, and to continually reset and re-zero the sensor. This device has been shown to be robust, reliable and accurate for ambient monitoring (Bart et al., 2014; Miskell et al., 2018; Williams et al., 2013). NO_2 was measured using an electrochemical sensor, whose response has been characterised in detail (Weissert et al., 2019). O_3 and NO_2 measurements were collected with 1 min time resolution and then hourly-averaged. The instrument has been described in detail in Weissert et al. (2019). The electrochemical NO_2 sensor element was supplied by Membrapor.

2.3. Proxy model

A critical element of the framework is the proxy model, which is described in detail in Miskell et al. (2018, 2019). To summarise, if $X_{j,t}$ denotes the true concentration at site j and time t, $Y_{j,t}$ denotes the sensor result, and $\widehat{X}_{j,t}$ the estimate of X derived from the measurement model, then the proxy model proposes that, over some time t_d that is sufficiently long to average short-term fluctuations, a proxy site, k, can be identified, with data $Z_{k,t}$ such that the empirical cumulative probability distribution of Z_k is a reliable estimate of the distribution of X_j, evaluated over t_d. Then the parameters of the measurement model can be estimated by adjusting them such that the distribution of $\widehat{X}_j$ approximates the distribution of Z_k. In the previous work, the measurement model parameters were adjusted to match moments of $\widehat{X}_{j,t}$ and Z_k. We assume a linear measurement model:

$$\widehat{X}_{j,t} = \widehat{a}_0 + \widehat{a}_1 Y_{j,t} + \varepsilon_{j,t} \quad (1)$$

with parameters obtained by matching the mean, $E\{\}$ and variance, $var\{\}$ of $\widehat{X}_{j,t}$ and Z_k

$$\widehat{a}_1 = \sqrt{var\{Z_{k,t-t_d:t}\}/var\{Y_{j,t-t_d:t}\}} \quad (2)$$

$$\widehat{a}_0 = E\{Z_{k,t-t_d:t}\} - \widehat{a}_1 E\{Y_{j,t-t_d:t}\} \quad (3)$$

Then, given the linearity assumptions,

$$X_{j,t} = \alpha_0 + \alpha_1 \widehat{X}_{j,t} + e_{j,t} \quad (4)$$

with (unknown) parameters

$$\alpha_1 = \sqrt{var\{X_j\}/var\{Z_k\}}; \alpha_0 = E\{X_j\} - E\{Y_j\}\sqrt{var\{X_j\}/var\{Y_j\}} - \widehat{a}_0\alpha_1 \quad (5)$$

Proxy selection aims to achieve $\alpha_0 \approx 0$ and $\alpha_1 \approx 1$. An obvious issue is that a precise estimate of uncertainty is not available since the true X is by definition unknown. The reference network is used to develop the criteria for choosing proxies and to indicate the resultant uncertainty.

2.4. Proxy selection

The proxy sites for the O_3 and NO_2 correction were established using data from the well-maintained South Coast AQMD regulatory network deployed in the LA and Inland Empire region (Fig. 1) with the procedures described in Miskell et al. (2019), Miskell et al. (2016) and Miskell et al. (2018). In previous work on O_3 measurement (Miskell et al., 2018, 2019), we used the closest proximity location regulatory site as a proxy, since ozone concentrations are reasonably well correlated across a region except in close proximity to heavily-trafficked roads – a source of nitric oxide, NO (Guttorp et al., 1994; Coyle et al., 2002; McConnell et al., 2006). However, given the high spatial variability of NO_2, here we explored the applicability of proxies chosen based on land-use similarities, because land-use has proven to be capable of explaining a significant fraction of the total variance of concentration across a region (Hoek et al., 2008) and hence has a logical basis for a choice based on trust derived from experience. Variables were chosen based on a systematic literature review on land-use regression (LUR) models developed for the North American Region. In total, significant covariates from 21 published NO_2 LUR studies (Supporting Information, SI Table S1) were ranked to identify the most commonly reported land-use variables explaining NO_2 variability in urban areas (Fig. 2). As expected, the most commonly used predictors for NO_2 concentrations are related to traffic, length of major roads and distance to major road, followed by land-use (commercial/industrial) and population density. The most used covariate was traffic, followed by major road length. However, local traffic (within 1 km) was not available for each site, thus we decided not to use traffic estimates for the proxy selection. We evaluated the land-use model against the reference data as described below. The adopted proxy model used only three parameters, because these were easily derivable from open-source data: distance to freeway (major multi-lane

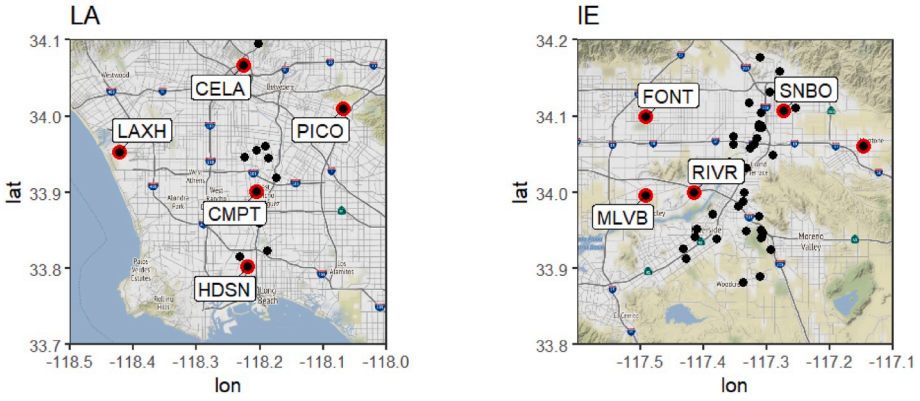


Fig. 1. Map of the regulatory sites (red points) and the low-cost instruments (black points) in the Los Angeles (LA) and Inland Empire (IE) region. A low-cost instrument was co-located at each regulatory site where both O_3 and NO_2 were measured. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

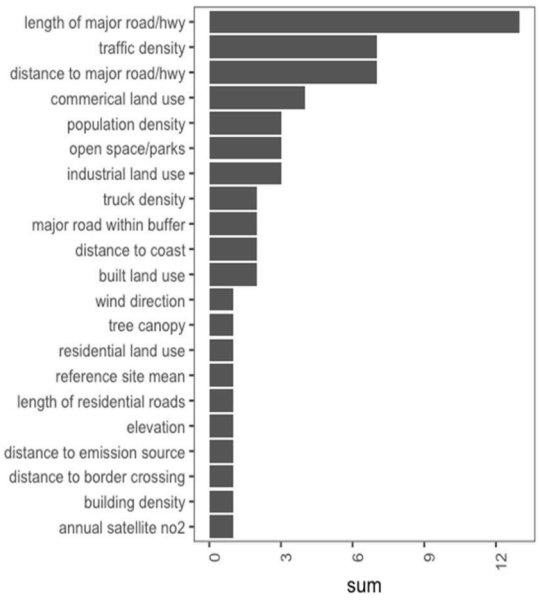


Fig. 2. Commonly used covariates in LUR models to predict NO_2 concentrations in the North American region. 'Sum' is the total number of times the particular covariate has been used across the studies cited.

highway) primary road length within 1 km and elevation above sea level. The parameters for the different reference sites are given in Table 1. Each variable was scaled so that they all had a similar range and were therefore comparable.

To identify the proxy site with the greatest similarity to the test site, we used the *k*-Nearest Neighbour classification (*k*NN). *K*NN is a supervised statistical learning technique aiming to classify the data to a given category based on a similarity using a test and a training set. The algorithm finds the k training samples that are closest to the regulatory data of interest and assigns the most suitable proxy among the k training examples to the regulatory site. Here, we use $k = 2$ (the second closest neighbour relative to a point is used). The proxy selection is shown in Table 1. Some of these sites were also those in closest proximity. A notable exception is the site RIVR where the land-use proxy is some 75 km distant, in a different physical region (Fig. 1).

2.5. Estimating the reliability of proxy-corrected data

Some means is required, given only the proxy results and the measurement results of the sensor network, to indicate the reliability of the proxy-corrected sensor results. We used two approaches, tested using the regulatory network and sensors co-located with regulatory instruments.

First, we applied to the regulatory network data the framework approach introduced by Miskell et al. (2016) to detect drift. It is based on three different tests comparing the distribution of the measurement result, Y, with that of a proxy, Z, evaluated running over a time t_d: the Kolmogorov-Smirnov (K–S) test for significant difference between the distributions Y and Z, and the estimates from moment-matching of apparent slope, $\hat{a}_1$, and offset, $\hat{a}_0$ (eqs (2) and (3)). When predetermined thresholds are exceeded for any of the parameters for a duration of five days (t_f), an alarm is triggered and is used to indicate potential sensor drift. The probability distributions are determined over a window of three days, t_d. Data are corrected if one or more than one alarm is triggered. Each regulatory site was tested against its chosen proxy. Given that regulatory data are used, the proxy signalling drift will in fact be an indicator of periods when the proxy site is not representative ("false alarm") and therefore this measure tests the performance of the proxy. The thresholds were: $p^*_{KS} = 0.05$ (for $n = 72$ hourly-averaged measurements) $\hat{a}_1 = 1 \pm 0.25$, $\hat{a}_0 = 0 \pm 5$ ppb. The thresholds and time windows are arbitrary and can be varied to make the test more or less rigorous.

Secondly, we considered measures of statistical divergence between the distributions of the corrected sensor data, $\widehat{X}_j$, and the proxy, Z_k. The statistical divergence between $\widehat{X}_j$ and Z_k should be smaller than that between Z_k and the unknown X_j. If the proxy distribution is a good estimator of the true data distribution, then the divergence between $\widehat{X}_j$ and Z_k should be small. With increasing dissimilarity between the distributions of Z_k and the unknown X_j the observed divergence between $\widehat{X}_j$ and Z_k should increase. Fig. 3 illustrates the idea, by using log-normal distributions as a model. The difference between the distributions of X_j and $\widehat{X}_j$ at the mean of $\widehat{X}_j$ is shown as a function of the KS and Hellinger divergences (see SI, for definition) between the distributions $\widehat{X}_j$ and Z_k for different model distributions. Fig. 3 shows that the two divergences are, as expected, correlated, but that they are differently sensitive to differences in the shape of the distributions. Fig. 3 illustrates that just using a threshold value for one divergence to signal that the likely error between true and estimated result has exceeded an acceptable value might incur a significant false negative rate – indicating an issue when the error was acceptable – depending on the shape of the proxy distribution in comparison with the (unknown) true data distribution. However, Fig. 3 also shows that in principle a threshold curve for the two divergences between corrected sensor and proxy data can be set, that reduces the false negative rate. We checked these expectations in practice by comparing results across the well-calibrated reference instruments using the choice of proxy for these, and by evaluating the root-mean-square error for sensors co-located at reference sites, where the sensor was corrected by the proxy. Proxy-corrected data for sensors co-located at reference sites were used to evaluate the dependence of measurement error on the distribution divergence between corrected sensor and proxy data. Thus, the regulatory network data are used to establish appropriate proxies and error estimates for the low-cost network, which in turn would be used to extend the scope of the regulatory network to neighbourhood scale.

Table 1
Descriptions for the nine regulatory locations, the NO_2 proxy chosen by land-use similarity and the physical distance between each location and its NO_2 proxy. Land-uses are based on publicly available data. Proxy sites in **bold** are also the sites in closest proximity. O_3 proxy sites are those in closest proximity.

Name	ID	Dist. to freeway/ m	Elevation above sea-level/ m	Primary road length within 1 km/ m	NO_2 Proxy site	NO_2 Proxy distance/ km	O_3 proxy
Rubidoux	RIVR	685	248	6708	CELA	75	MLVB
Mira Loma	MLVB	2480	220	0	SNBO	24	RIVR
San Bernardino	SNBO	2620	316	2408	MLVB	24	RIVR
Fontana	FONT	3210	363	5889	SNBO	20	MLVB
Pico Rivera	PICO	803	58	6563	**CELA**	16	CELA
Compton	CMPT	1660	22	7040	**HDSN**	12	HDSN
LAX Hastings	LAXH	4450	37	4270	**CMPT**	21	CMPT
Long Beach (Hudson)	HDSN	1150	10	5566	**CMPT**	11	CMPT
Central LA	CELA	917	89	5168	HDSN	30	PICO

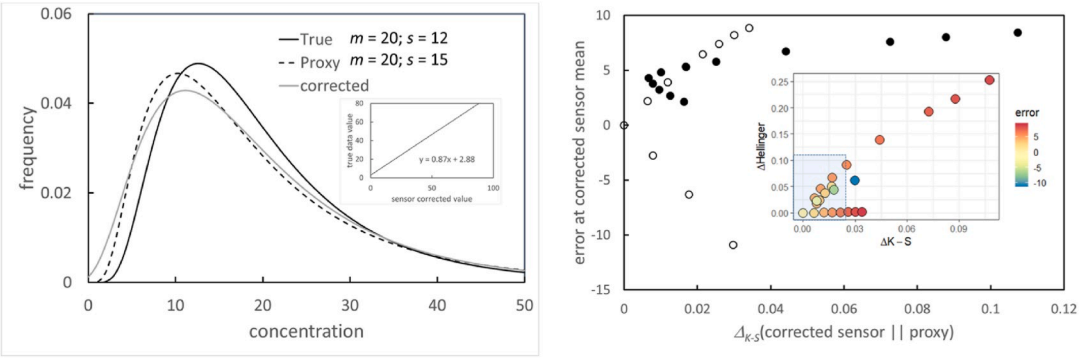


Fig. 3. Log-normal model distributions used to model the proxy distribution fitting procedure. Left: example distributions of true, proxy and sensor data corrected using eqs (1)–(3); and linearity of corrected sensor predicting true data, according to eqs (4) and (5). Right: illustration of the variation of the estimation error (difference between corrected sensor and true data at the arithmetic mean of the corrected sensor data). with the Kolmogorov-Smirnov divergence, $\Delta_{K\text{-}S}$, and Hellinger divergence between corrected sensor and proxy. Shaded area illustrates the possibility of determining whether the correction is acceptable by using a combination of the two divergences. Open symbols: fixed arithmetic mean, *m*, and varying arithmetic standard deviation, *s*; filled symbols: fixed proxy *s*, varying *m*. The dimensionless parameters for the model distributions have been chosen to match approximately the observed NO_2 data in ppb.

2.6. Correction method for the NO_2 sensor

The measurement model for the NO_2 sensor, relating the measured current in the electrochemical cell, i_{meas}, to the indicated concentration of NO_2, C_{NO_2}, is (Weissert et al., 2019):

$$C_{NO_2} = \beta_0 - \beta_1 i_{meas} - \beta_2 C_{O_3} \quad (6)$$

Factory calibration of the assembled instrument before field deployment determines a number, $C_{ox} = \beta_1 i_{meas} - \beta_0$, which is linearly related to the raw current measurement. The instrument reports C_{ox} as well as the NO_2 concentration derived from the factory calibration and the uncorrected O_3 concentration determined with the O_3 sensor. Now, the offset, β_0, and the response slopes, β_1 and β_2, can be time-varying, for example in response to changes in atmospheric humidity or temperature. The objective of the procedure is to estimate and correct for this variation. The measurement model to be used, therefore, given the results reported by the instrument, is written

$$\widehat{C}_{NO_2} = \widehat{b}_0 + \widehat{b}_1 C_{ox} - \widehat{b}_2 C_{O_3} + e \quad (7)$$

where *e* denotes any signal not accounted for by the principal variables assumed to drive the response and which also includes any measurement noise. Following the concepts described earlier, the correction method estimates values of the parameters $\widehat{b}_j$ to match the probability distribution over time t_d of the estimate $\widehat{C}_{NO_2}$ to that of a proxy, Z_{NO_2}, by minimising a suitably chosen objective function. The proxy site is chosen based on land-use similarity. We explored two methods which gave similar results. First, we evaluated minimisation of the sum of squared differences of the first three moments of the distributions. Second, we evaluated minimisation of the Kullback-Leibler divergence (see SI, for definition) of the two distributions, $D_{KL}(\mathbb{P}(\widehat{C}_{NO_2})||\mathbb{P}(Z_{NO_2}))$. The moment matching method emphasises the tails of the distributions. The Kullback-Leibler method, on the other hand, emphasises the most probable values, and its minimisation is equivalent to maximising the mutual information or minimising the relative information entropy of the two distributions and we chose it for that reason. In the following sections, we present the results from the minimisation of D_{KL}. Thus, we aim to find best estimates $\widehat{b}_j$ such that:

$$D_{KL}\left(\mathbb{P}\left(\widehat{C}_{NO_2}\middle|C_{ox}, C_{O_3}, \widehat{b}_j\right)\middle|\middle|\mathbb{P}(Z_{NO_2})\right) = min \quad (8)$$

In this calculation, the value of C_{O_3} used is that delivered by the O_3 sensor which is checked and corrected if necessary according to the management framework as previously described (Miskell et al., 2019). The parameters are re-estimated only when the comparison of the (previously) estimated $\widehat{C}_{NO_2}$ with the proxy gives an alarm, thus minimising the computational overhead. The process is initiated using the concentration values given by the pre-deployment factory calibration, denoted here $C_{NO_2}(raw)$. The probability distribution of the estimate should be a sum of three distributions corresponding to the three terms. The variability of C_{ox} would be determined by the noise in the electrochemical sensor (Weissert et al., 2019) and the averaging approach used to reduce this. O_3 and NO_2 measurements were collected with 1 min time resolution and then were hourly-averaged. Based on the results in (Weissert et al., 2019) we expect the standard deviation of this number to be less than 1 ppb. The RMSE of C_{O_3}, corrected according to the management framework, is 5.4 ppb for all reference sites combined and the entire study period (January–August), with a maximum RMSE of 7 ppb for individual sites (Miskell et al., 2019).

Two issues could affect the reliability of the parameters in equation (7) obtained through minimisation of the difference between the probability distributions. First, if the distributions approximate simple 2-parameter distributions (e.g. log-normal) then deriving three parameters from the comparison over-fits the data and would raise issues of correlation between the parameter estimates. A second potential issue, under circumstances where O_3 and NO_2 report similar concentration levels, is that an unconstrained minimisation could easily lead to physically unreasonable estimates with the parameters changing sign.

Indeed, we noted that minimisation of D_{KL} with $\widehat{C}_{NO_2}$ calculated with eq (7) without physically realistic initial estimates of the parameters, could easily lead to false minima with physically unrealistic parameter values (e.g. inverted sign). Physically realistic initial estimates for the minimisation were obtained as follows:

a) the measurement model is approximated by setting $b_2 = b_1$ as observed and also theoretically expected for an electrochemical sensor of this type without O_3 decomposition catalyst applied (Weissert et al., 2019).
b) the initial estimates of b_0 and b_1 (=b_2) are obtained by moment matching to the proxy:

$$\widehat{b}_{2,init} = \widehat{b}_{1,init} = \sqrt{var\{Z_{NO_2}\}/var\{C_{O_x} - C_{O_3}\}} \quad (9)$$

$$\widehat{b}_{0,init} = E\{Z_{NO_2}\} - E\{C_{O_x} - C_{O_3}\} \quad (10)$$

following which the b_j are iterated in eq (7) to minimise D_{KL} (eq (8)). The value C_{ox} is the raw signal from the electrochemical sensor using the internal offset and slope values as above, hourly averaged, and (as noted above) the value of C_{O_3} used is that delivered by the O_3 sensor, hourly averaged, checked and corrected if necessary according to the

management framework as previously described (Miskell et al., 2019). The management framework is schematically illustrated in Fig. 4.

3. Results and discussion

In the following, in (3.1) we describe the general characteristics of the data and in (3.2) the results of the proxy assessment. To test the sensor correction method described in (2.6) in the absence of noise introduced by the proxy and by the ozone sensor, in (3.3) we show the application of the sensor correction method to co-location data, using the reference O_3 and NO_2 data from the co-location site. This study identified an important offset error term, that varied on a timescale less than the framework error detection timescale and therefore prevented the framework compensating. The offset error was climate-related (mostly but not entirely ambient temperature) and spatially correlated. The correlation with ambient temperature was non-linear and the effects of extremes were persistent, hence although extremes of temperature could in principle be used to signal an issue, we chose not to attempt correction by use of a correlation with temperature. Instead, we used the knowledge that this offset error term was spatially correlated to apply an additional correction, derived using the closest proximity proxy site. Then, in (3.4) we show application of the framework to sensors that were co-located at reference sites, but using proxy data and the sensor ozone data. By comparison with the reference data from the site of co-location, we evaluate the error in the whole procedure. Table 2 gives the comparison between corrected data and the reference instrument data at each stage of the development, at each site, averaged over the time of the study. We show how tracking of the derived sensor parameters over time identifies occasions when errors might be excessive, either due to issues with the proxy, or when the sensor drift was excessive indicating the need to replace the sensor. We show how statistical divergence measures between sensor and proxy identify circumstances with potentially large error, following the ideas developed in (2.5). We present a detailed analysis of the errors and show that proxies chosen by land-use criteria that are a significant distance away from the sensor site can give satisfactory results. In (3.5) we show the large-scale local variations revealed by the low-cost sensor network.

3.1. General characteristics of the data

The temporal and spatial variability of NO_2 concentrations, as shown by the reference instrument network, was, as expected, large: Fig. 5. The distribution of values varied month-by-month, from broad and bimodal in winter (mean ± sd temperature/RH: 16 ±7 °C/46 ± 26%) to narrower and monomodal in summer (mean ± sd temperature/RH: 27 ±6 °C/55 ± 21%), and displayed different patterns at different sites (Fig. 5a). Diurnal variations were irregular: in winter, the variation was typically small fluctuations upon a large and variable background; in summer, values were frequently low and hardly varying (Fig. 5b). The diurnal variation showed patterns that were frequently similar across a number of sites whilst being very different at others (Fig. 5c). At some sites and months of the year, the concentration distribution averaged over the month approximated a simple log-normal (hence 2-parameter) distribution; at others, it did not, being very strongly tailed to high values (Fig. S1).

3.2. Proxy evaluation using the reference network

The wide variation in NO_2 distribution across space and time, exemplified by the results given in Fig. 5, illustrates the challenge in determining appropriate proxies for a region as varied as that of Southern California. The framework that we have proposed uses comparison of the statistical distribution of concentration over a rolling timescale of 3 days to signal an 'alarm' and of 5 days to signal a 'failure'. We applied this framework to the reference network, using the proxies for each site determined by land-use criteria (Table 1). Fig. 6 is an overview of the number of times the framework approach signalled an alarm due to a threshold being exceeded ("false alarm"). These are "false alarms" because regulatory stations are being compared. Typically, the alarm was raised due to differences in the distribution of NO_2 concentrations between the site and its proxy ('KS-test'), followed by a change in the slope ('MV-slope'). The intercept ('MV-intercept') remained mostly stable between the regulatory site and its proxy site. The frequency of false alarms over the several months of the study was small: < 1% of the total of hourly average values were falsely indicated as alarms, even at the RIVR site, where the proxy was 75 km distance in a different region. Therefore, approaching the problem through a comparison of probability distributions over an appropriately chosen timescale indeed provides a way of defining suitable proxies. Fig. 6 shows that the sites could be classified into two classes: those where the proxy false alarm rate was very small, < 0.2%, and others where it was somewhat higher but still <1%. For the first group of sites, the proxy-corrected reference data almost exactly matched the uncorrected data. Conditions at the latter group of sites that might give rise to the higher rate of false alarms

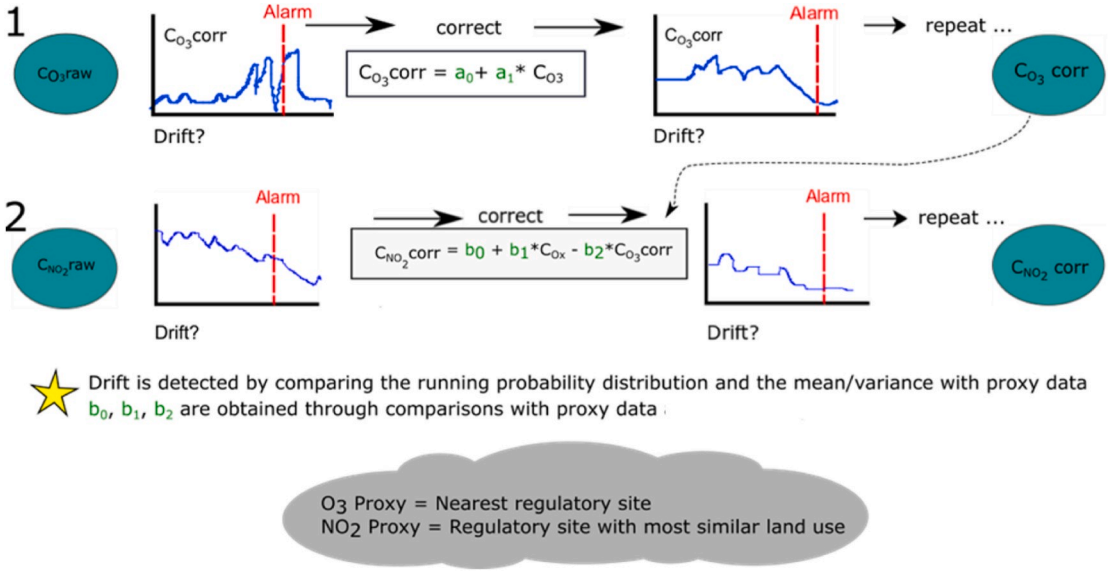


Fig. 4. Summary of the O_3 and NO_2 management framework and correction process.

Table 2
Summary statistics comparing the uncorrected, framework corrected and framework corrected plus *es* AQY NO_2 data against the regulatory, for correction using co-located instrument data (described in section 3.3), and for correction using proxies and the sensor ozone data (described in section 3.4).

	Proxies									Co-located					
	Uncorrected			Framework corrected			Framework corrected + *es*			Framework corrected			Framework corrected + *es*		
REF Site	R^2	MAB	RMSE	R^2	MAB	RMSE	R^2	MAB	RMSE	R^2	MAB	RMSE	R^2	MAB	RMSE
RIVR	0.72	6.34	8.13	0.71	6.04	7.71	0.74	5.35	6.81	0.76	3.92	5.65	0.81	3.78	5.32
MLVB	0.53	7.86	9.72	0.68	5.01	6.79	0.71	4.97	6.70	0.69	7.11	8.80	0.75	7.14	8.71
SNBO	0.26	9.12	11.18	0.52	8.02	10.43	0.56	8.14	10.80	0.69	4.14	5.49	0.75	3.68	5.05
FONT	0.47	10.13	11.92	0.63	5.25	6.95	0.67	5.42	6.96	0.61	5.72	7.71	0.72	4.64	6.30
PICO	0.07	9.89	11.99	0.52	4.36	5.56	0.61	4.02	5.15	0.56	3.5	5.67	0.69	2.69	4.69
CMPT	0.70	6.36	7.45	0.77	3.72	5.10	0.76	3.84	5.17	0.86	2.97	3.96	0.87	3.00	3.93
LAXH	0.60	4.90	6.62	0.63	5.81	7.21	0.60	5.92	7.35	0.84	2.24	3.11	0.85	2.22	2.95
HDSN	0.46	10.06	11.82	0.68	3.37	4.35	0.72	3.19	4.19	0.76	2.96	3.98	0.82	2.67	3.58
CELA	0.6	14.71	15.88	0.60	4.53	6.03	0.55	4.58	6.55	0.78	2.99	4.24	0.79	2.53	4.09

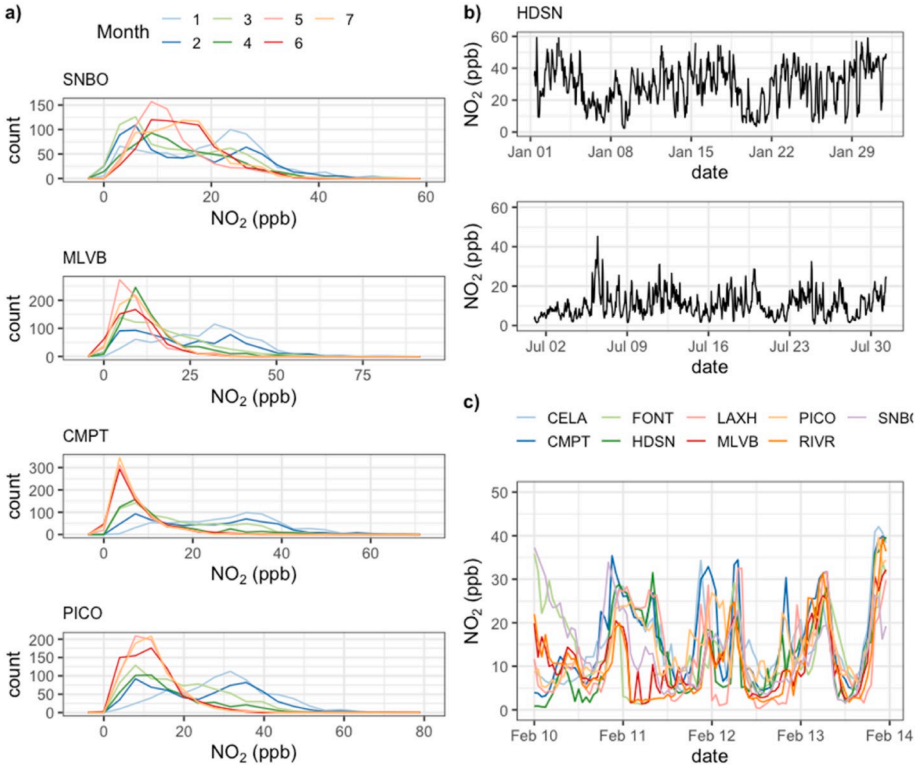


Fig. 5. Examples of data for the spatio-temporal variation of NO_2, as shown by the reference instrument network. a) Frequency distributions of concentration at exemplar different sites month-by-month (month 1: January 2018; month 7: July 2018). b) Time series at the Hudson (HDSN) site, exemplifying variation in winter (January) and summer (July). (c) Time series over a few days at all nine sites, showing both similarities and differences across the study area.

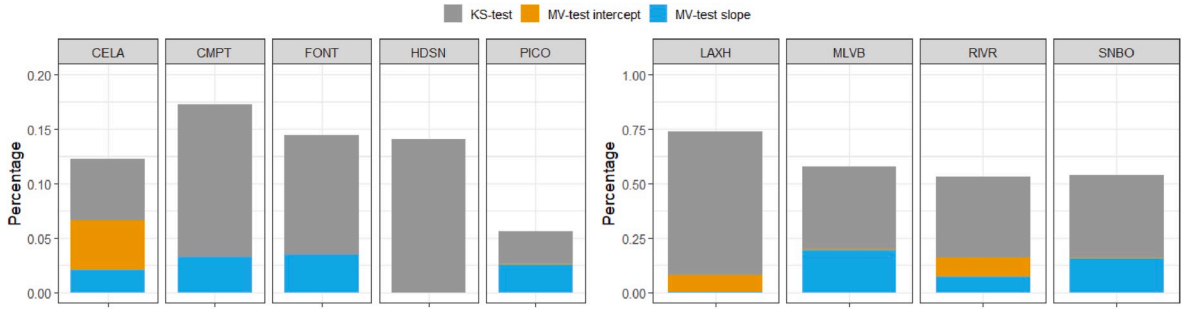


Fig. 6. Percentage of hourly average measurements at the different reference sites that are indicated as alarms by the proxy site (sites given in Table 1), for the three different tests. The two panels show on different scales the two groups of site: those with very low alarm rate (left) and those with a higher alarm rate (right).

and consequent errors in correction are next examined.

Fig. 7 shows that the dominant factor driving false alarms and hence over-correction was wind speed and direction. The error at MLVB, RIVR and SNBO is a slope error – *var*{SNBO}/*var*{MLVB} > 1 when the wind was from the W, with the effect greater at SNBO when the wind speed was low; and *var*{CELA}/*var*{RIVR} >1 when the wind was from NNE and the wind speed was high. The error at LAXH is dominantly an offset error: *E*{CMPT} > *E*{LAXH} when the wind speed was low. The sites at RIVR and CELA are very distant and in different regions. The false alarms and over-correction with high wind speeds from the NNE are likely related to the mountains NE of the site and their effect on ozone transport and hence titration of vehicle-emitted NO. Further, false alarms at RIVR were also more common when wind speed was $<5\ m\ s^{-1}$, a condition under which observed NO_2 concentrations would be expected to be dominated by local emissions and possibly therefore different from those at a proxy site: again, possibly related to ozone transport. The site at SNBO is 3 km east of the San Bernadino Santa Fe Train Depot, a major transport hub and presumably source of NO_2 emissions (Weissert et al., 2020) and thus potentially significantly increasing the variations at SNBO when the wind is from the west. LAXH is the regulatory site at the Los Angeles International (LAX) airport and it is the least similar to any other regulatory site in terms of land-use. It is in effect a large open space close to the ocean. At LAXH, the KS test trigger was more frequent than at any other site: the concentration distribution at LAXH was the least similar between sites, as was also determined by other measures of divergence (eg Kullback-Leibler).

Effects of local geography and meteorology are not easily captured in land-use models. However, the effects are clear and easily captured in simple rules. For example, it would be possible to introduce conditional statements in the framework approach for situations when the proxy site is not suitable. The reference instrument network could be used to identify whether there was a more suitable proxy to use under these circumstances: for example, under error conditions to switch the proxy for RIVR to one within the same valley and with similar land-use, which would be MLVB. The use of MLVB as proxy for RIVR indeed resulted in fewer false alarms. Whilst land-use similarity is arguably a more logical basis for proxy choice than simple proximity, and such a choice provides a logical basis for network design to minimise the number of proxy sites, Fig. S2 shows that indeed for some sites the closest proximity other site was a better proxy. The data that we give in this paper have not, however, been additionally corrected in this way.

3.3. Using co-location data to evaluate the sensor correction method, sensor parameter variation and error terms

Most co-location studies use regression methods. In contrast, our proxy comparison is based on similarity of probability distributions over a time interval. Therefore, we used comparison of probability distributions on the co-location data to evaluate the performance of this method. For this part of the work we used the co-located reference O_3 data to avoid noise associated with the sensor O_3 correction. Fig. 8 shows hexbin scatter plots of the sensor NO_2 corrected using the co-located reference O_3 data against the co-location reference NO_2 over the 7 months of the study. The derived sensor parameter variations over time are given in Fig. S2. Parameter variation over time, within bounds, is expected. However, the sensors at MLVB and RIVR showed a downward drift of the slope parameters from July onward, very marked at MLVB, which the method compensated by an increase in the offset parameter. This behaviour should be taken as an indicator of sensor failure. All other sensors appeared stable. The hexbin plots show a significant scatter of the results. However, Fig. 9 shows that the difference between sensor-indicated NO_2 and reference NO_2 had a part that showed a diurnal variation as well as a part that showed apparently random variation. The difference term was to a degree spatially correlated, and particularly correlated between locations within the Inland Empire region. The variations tended to become larger at inland locations compared to those close to the sea. The correlation matrix is given in Table S2, the correlations between sites in closest proximity are shown in Fig. S4, and the dependence of the correlation on inter-site distance and region is shown in Fig. S5. Fig. 9 also shows the joint probability distribution of the difference term and ambient temperature, measured by the sensor. Extreme values of the difference were associated with high (~50 °C) or low temperature (<12 °C), and there was a slow drift following an extreme (Fig. 9) but otherwise there was not a strong

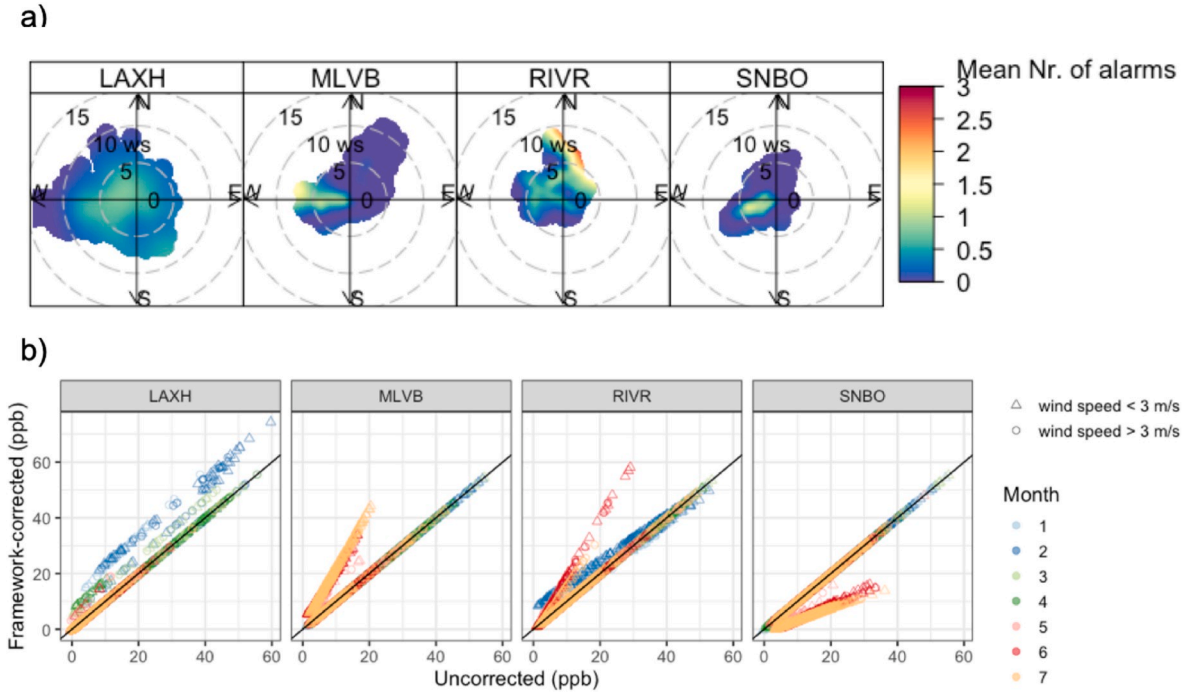


Fig. 7. Analysis of false alarms triggered by the proxy, at sites where these were more significant. (a) mean number of alarms per hour segmented by wind speed and direction. (b) Proxy-corrected reference data compared with uncorrected data, segmented by month and wind speed.

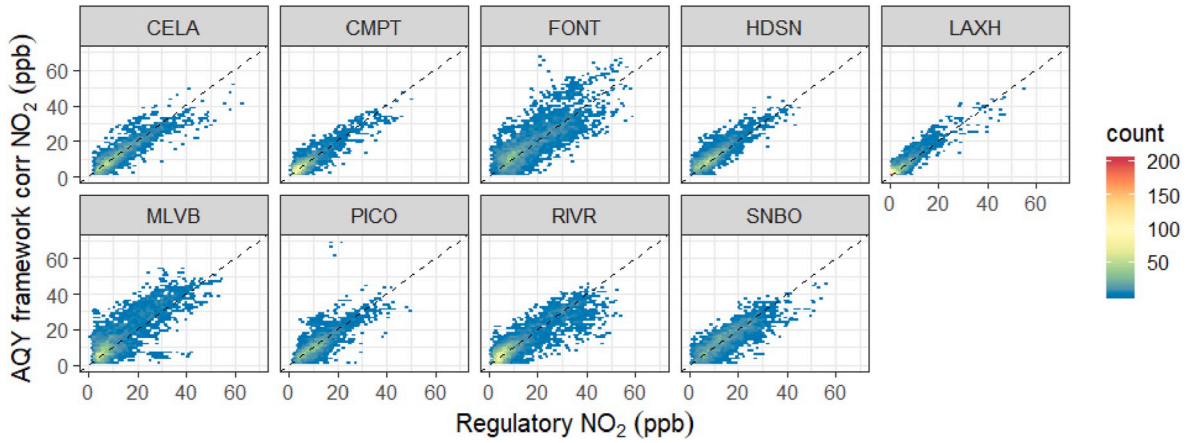


Fig. 8. Hexbin scatter plots showing the correlation of sensor NO_2 with co-location reference NO_2, where the sensor NO_2 is derived using the distribution matching method (Fig. 4 and eqs (7)–(10)) and the co-location reference NO_2 and O_3 data.

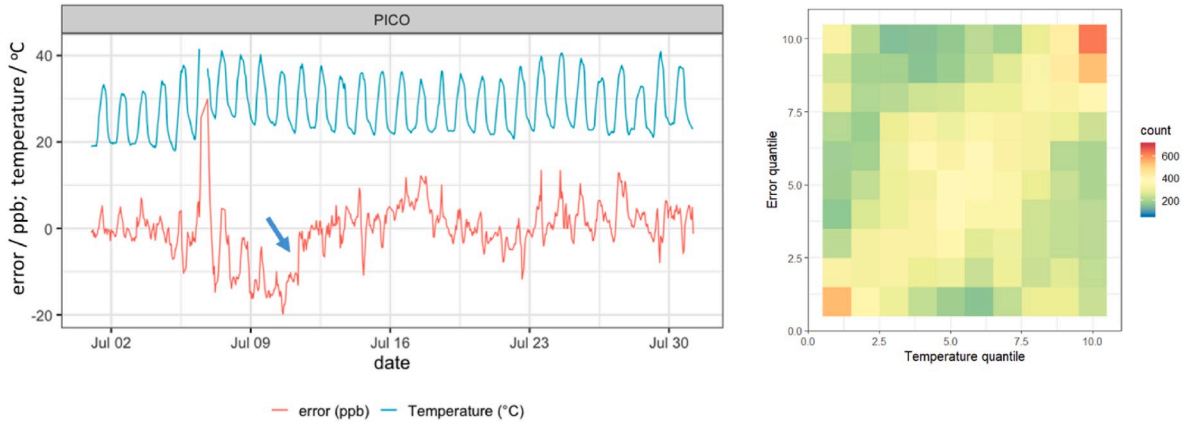


Fig. 9. (Left) Example time series of the difference term between framework-derived sensor result and the co-location reference result, illustrating the offset error. The arrow marks where the management framework detected and corrected the drift. (Right) Joint probability distribution of difference term ('error') and ambient temperature, for the whole data set. Temperature Quantiles (°C): 0: 1, 1: 12, 2: 14, 3: 16, 4: 18, 5:19, 6: 21,7: 23, 8: 26, 9: 30, 10: 50.

correlation.

The difference term can be attributed to large variations of the offset, $\widehat{b}_0$. The variations are related to the electrochemical sensor, indicating fluctuations with a time scale between 1 h and 3 days, these being the timescales of averaging and of comparison with the proxy distribution. The existence of a dependence of electrochemical NO_2 sensor signal on temperature, humidity and their rapid changes is known, but there is no simple relationship. Without a definitive model for the variations, it is difficult to provide a rigorously-based correction method. Below, we present an empirical method based on the observed spatial correlation.

Given these results, we rewrite the measurement model (eq (7)) as:

$$\widehat{C}_{NO_2} = \widehat{b}_0 + \widehat{b}_1 C_{ox} - \widehat{b}_2 C_{O_3} + e_S + \varepsilon \tag{11}$$

where e_S denotes a spatially correlated error term and ε the residual. Now, we propose a proxy method for estimating e_S. Since we have electrochemical sensors co-located at reference sites, and the term is spatially correlated, an estimate of e_S at some other site would be that value determined at the closest proximity reference site, at the required time. Fig. 10 illustrates the issues with this idea. Firstly, if proxy data are unavailable at any particular period, then obviously no correction can be made; secondly, although the error term is spatially correlated on average, at any particular time the difference between the values at the

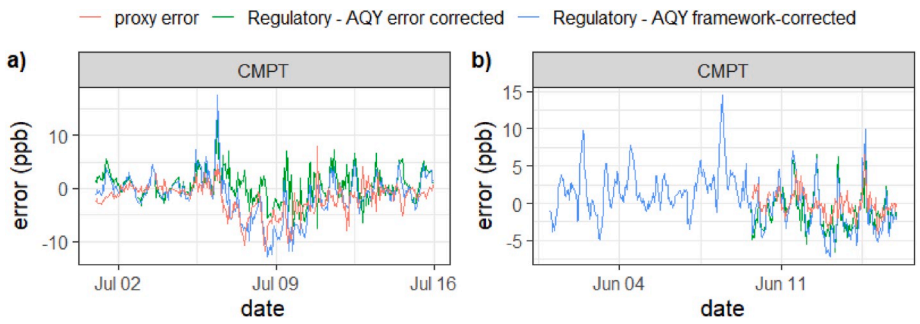


Fig. 10. Examples of the uncompensated error term, and its partial correction using the error determined at the closest proximity site. red: error term determined at the closest proximity site; blue: actual error ($e_S + \varepsilon$, eq (11)) determined at the measurement site following correction of the sensor using K-L method; green: actual error determined at the measurement site following correction of the sensor using K-L method and determination of e_S using the closest proximity site (damped and smoothed as described in the SI). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

measurement site and the proxy site can be large. Given these issues, we found that this method could however compensate a useful fraction of the difference term, provided the correction was limited: we used a sigmoid function to damp the error correction and a rolling average to smooth fluctuations; details are in the SI. Fig. 11 shows hexbin scatter plots for the co-location data where the error term e_S has been estimated from the closest proximity other site. The scatter is diminished at most sites. The overall RMSE improved and is 5 ppb (statistics for individual sites in Table 2). Given that the estimated error due to sensor noise is less than 1 ppb, the major contributor to this error would be uncompensated sensor responses, such as are reflected in the uncompensated offset error term shown in Figs. 9 and 10.

3.4. Using sensor ozone data and proxy sites to check and correct the NO_2 sensors

Here, the framework was applied to sensors located at reference sites, using the O_3 sensor, the proxy sites for NO_2 (land-use) and O_3 (proximity) to derive the sensor parameters using the K-L method according to eqs (7)–(10), and the closest proximity proxy also to determine the spatially-correlated error, e_S (eq. (11) and eq. (S1)). The resulting corrected sensor result is compared with the reference result at the same site.

Overall, the framework produced satisfactory results. Fig. 12 shows examples of the time variation of the uncorrected and corrected rolling mean absolute bias (MAB) in relation to the co-located regulatory NO_2 and the alarm signals triggered over time at the regulatory sites. Data for all sites is in Fig. S7. Fig. 13 shows the monthly average MAB at the different sites for the framework-corrected data. The management framework was able to detect and correct the drift resulting in a MAB within 2 and 10 ppb at most times and sites, which was a clear improvement to the uncorrected NO_2 MAB (up to 20 ppb) and considered satisfactory for an indicative air quality measurement (Snyder et al., 2013). Table 2 gives the statistics for each site averaged over the whole time of the study, and shows how the statistics improve at each step of the correction procedure.

To illustrate the operation of the framework, Fig. 14 shows three examples: a site where the correction was satisfactory, though with a slight slope error (FONT); one where the sensor failed (MLVB); and one where the proxy selection was inappropriate for a particular time (CMPT, July: see also Fig. 13). Full data for all sites are given in Figs. S7–S9. For the sensor at FONT (correction and sensor satisfactory), Fig. 14 shows that the major variation over time was in the offset: indeed reflected in the MAB of the uncorrected data shown in Fig. 13. The two slope parameters were essentially constant and close to unity, with small

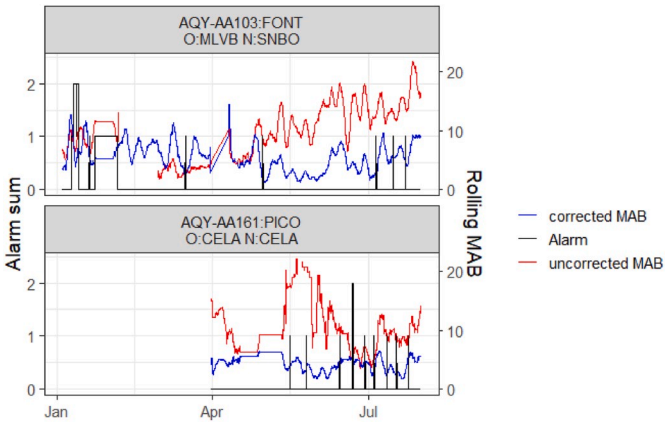


Fig. 12. Illustrative examples showing the number of alarm signals generated by the low-cost sensor data in comparison with the proxy data (left-axis), and uncorrected vs. corrected mean absolute bias (MAB) running over 72 h of the low-cost sensor data with respect to the co-located regulatory station (right-axis). One sensor from each region; top: FONT with O_3 and e_S proxy MLVB and NO_2 proxy SNBO; bottom: PICO with O_3 and e_S proxy CELA and NO_2 proxy CELA (shown in the label at the top of each chart).

fluctuations. The site and proxy distributions could be made almost coincident with a small alteration of response slope. The procedure resulted in the sensor data distribution being essentially coincident with the proxy distribution (evaluated over a month). For the sensor at MLVB, Fig. 14 shows a sudden and large jump in the offset during April which was associated with the start of a steady decrease in both the slope parameters. Inspection of the data showed the daily signal variations gradually decreasing towards zero. Although the correction in fact operated reasonably, clearly the sensor was failing, and there was an unknown event in mid-April that resulted in sensor failure. Monotonic change over time of the sensor parameters could be taken as indicative of sensor failure. For the sensor at CMPT, comparison of the data distributions given in Fig. 14 shows that, in July, the proxy and reference site distributions were very different. The procedure caused a bias in the sensor result towards the proxy with consequent over-estimation of the concentration. The response slope parameters both rose to values significantly greater than unity while the offset remained close to zero. The iterated minimum value of the objective function, D_{KL}, between corrected sensor data and the proxy became significantly larger. A strong variation of both slope parameters without corresponding variation of the offset, together with an increase in the iterated minimum

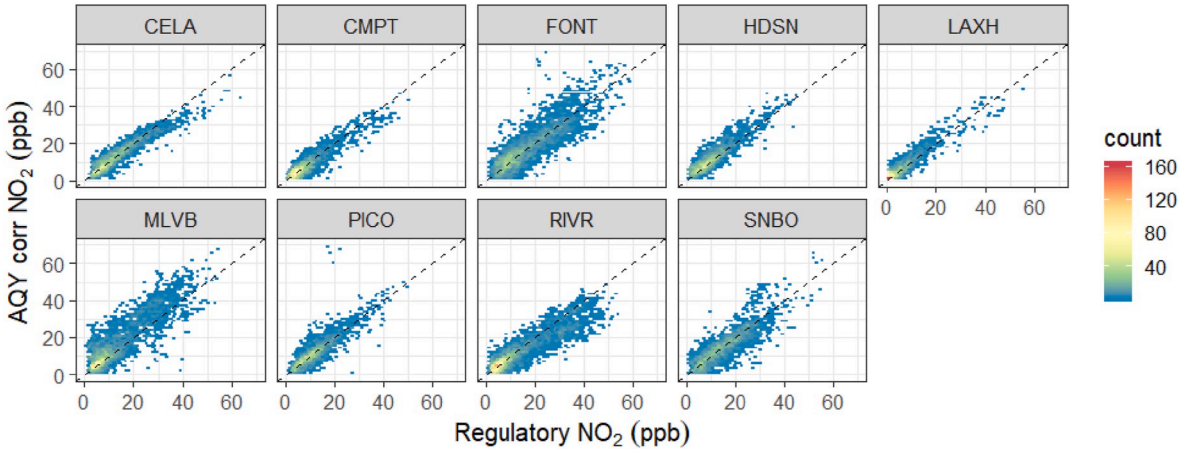


Fig. 11. Hexbin scatter plots showing the correlation of sensor NO_2 with co-location reference NO_2, where the sensor NO_2 is derived first using the framework distribution matching method (Fig. 4 and eqs (7)–(10)) and the co-location reference NO_2 and O_3 data, then by correcting using the closest proximity other co-location reference site to estimate e_S (eq (11)).

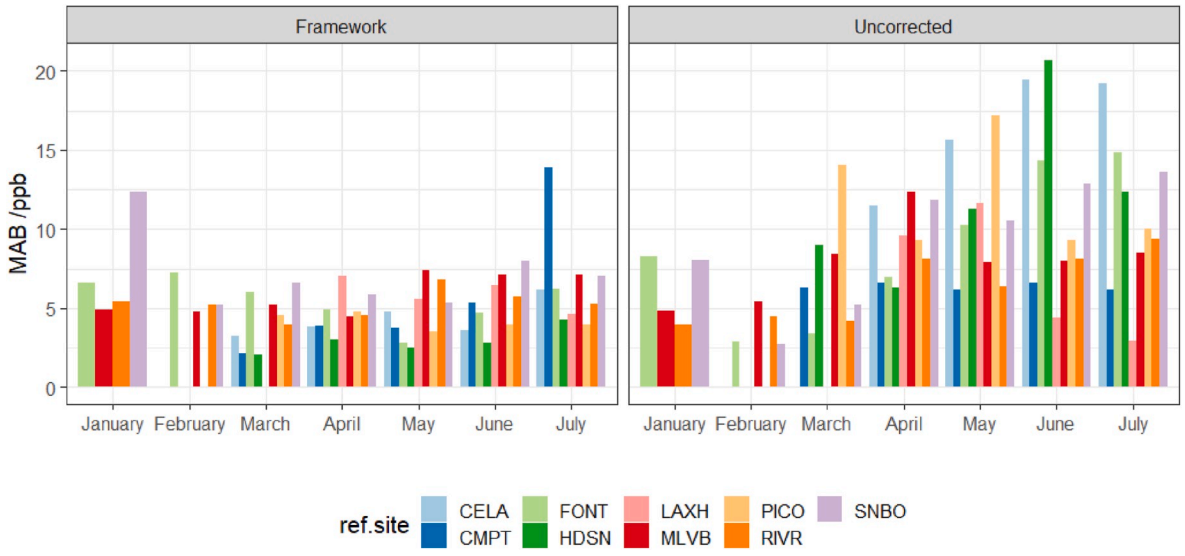


Fig. 13. Mean Absolute Bias (MAB) compared to uncorrected MAB per month for proxy-corrected data ('Framework') at the different sites.

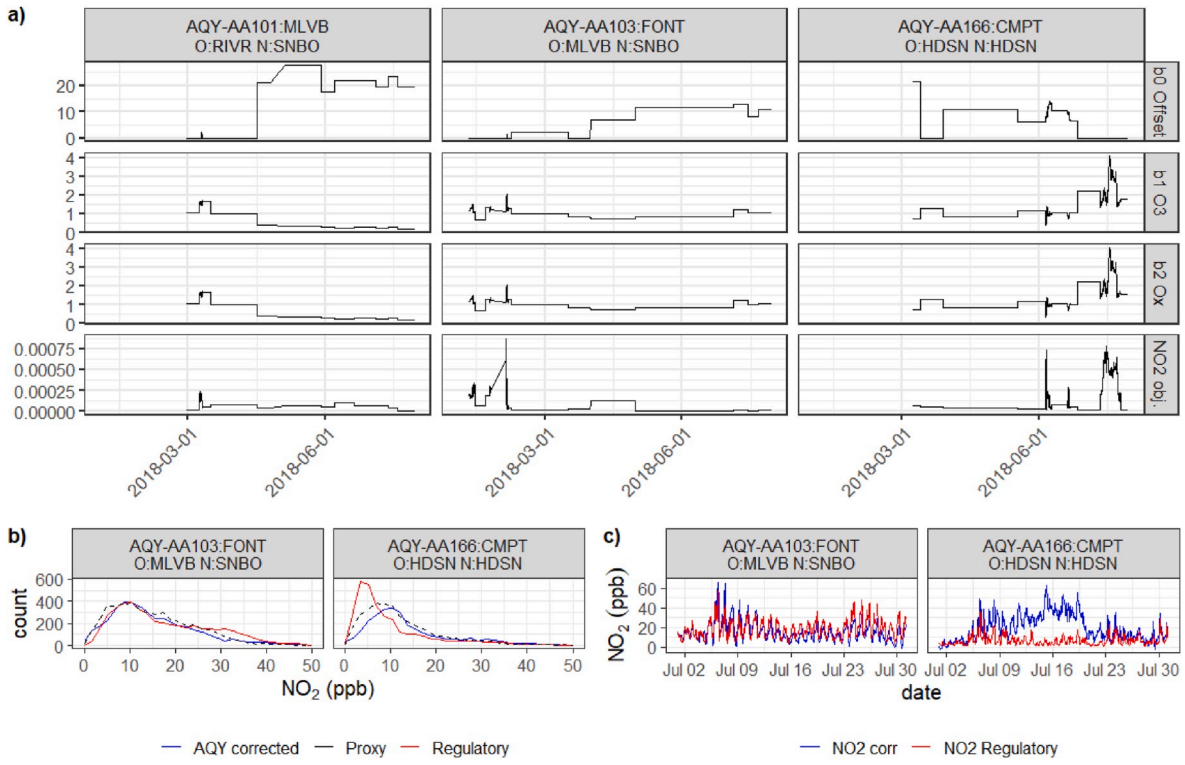


Fig. 14. a) Variation over time for three example sites of the fitted parameters: left, MLVB; middle, FONT; right, CMPT; top: offset, $\widehat{b}_0$, upper middle: slope parameter $\widehat{b}_1$; lower middle: slope parameter $\widehat{b}_2$; and bottom: the minimum obtained for the objective function, D_{KL} between sensor data according to equation (2) and the NO_2 proxy. At the top of each panel is shown the site designation and the proxies for ozone (O) and NO_2 (N). b) Distributions for the month of July of the regulatory station data, the proxy station data and the fitted sensor data (eq (6)), for sites at FONT (left) and CMPT (right). c) Time series for July comparing the fitted sensor data, $NO_{2,corr}$, and the regulatory data at the site with which the sensor was co-located; FONT: left; CMPT: right.

value of the objective function could be assumed as indicative of an issue with the proxy. Correlated variation of the slope parameters could be taken as indicative of an issue with over-fitting of the data. Rules could be added to the framework such that, if such occurrences were signalled, then the result using a different proxy could be obtained and compared.

Fig. 15 shows hexbin scatter plots of the correlation between the corrected sensor data and the co-located reference station. A hexbin scatter plot for the entire set of corrected sensor data is also presented. The majority of measured NO_2 concentrations were low, making the measurement task challenging. The hexbin plots show that the framework correction was generally successful, though clearly less so at the MLVB and SNBO sites. As noted above, the sensor at MLVB failed during April. The variation of the derived parameters for SNBO (Fig. S8) indicated issues with the proxy, which was confirmed by inspection of the frequency distribution of the NO_2 concentrations at the proxy site and at the SNBO regulatory site during June and July, partly explaining the

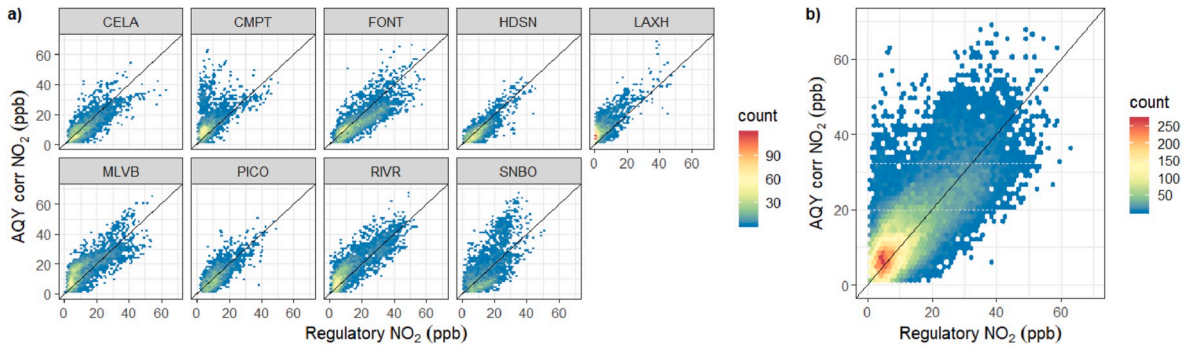


Fig. 15. a) Hexbin scatter plots for the framework-corrected data for the individual sites of co-location, for the entire study period. b) Hexbin scatter plot for the set of framework-corrected data (all sensors, all sites).

lower success of the management framework for these months. NO_2 concentrations can vary considerably at the sub-kilometre scale and the success of the management framework strongly depends on the representativeness of the land-use surrounding the reference sites for the low-cost sensor site that is calibrated (Li et al., 2019; van Zoest et al., 2019; Weissert et al., 2019). LAXH is the regulatory site at the Los Angeles airport and its proxy site (CMPT) is in central Los Angeles and may therefore not be a representative site for the local emissions at LAXH. Otherwise, Fig. 15 shows that the deviations about the 1:1 line were similar at all the sites. The RMSE for the individual sites varied between 4 and 11 ppb (Table 2). In particular, the results obtained at the RIVR site with CELA as proxy show that a very distant proxy with similar land-use can indeed give satisfactory results. For all sensors and sites, the framework-corrected data had a RMSE of 7.2 ppb. Fig. S10 shows the error distribution segmented by concentration quartile, for the entire dataset. There was a small concentration-dependent bias and the error distribution was broader for the highest concentration quartile, as is expected as a consequence of the contributions of errors in estimation of offset and slope. A further analysis of the offset and slope errors is given in the SI (Figs. S11 and S12, and Table S3). The mean offset error for the different sites varied from 0.1 to 6.5 ppb (average across all sites 1.4 ppb) with standard deviation from 2 to 6 ppb (average across all sites 4 ppb). The mean slope error for the different sites varied from 0 to −0.16 (average across all sites −0.05) with standard deviation from 0.03 to 0.12 (average across all sites 0.09). The largest slope error was at the RIVR site (−0.16) consistent with the slope error noted when the proxies were evaluated using reference site data (Fig. 7).

The data allow us to address the question whether the error in the sensor-estimated concentration can be itself estimated knowing only the distributions of proxy concentration and sensor-estimated concentration over some time period, following the ideas set out in section 2.5. We have done this using data over each individual month. Fig. 16 gives the variation of the RMSE for each different month by site with the divergence between proxy distribution and sensor estimate for that month. Fig. 16 shows that a large value of KS divergence between sensor and proxy signals a larger error in the estimated concentration. Use of both KS and Hellinger divergence may make the discrimination more clear. Comparison of Figs. 16 and 14 and Fig. S8 also confirms that a large value for the converged objective function when sensor correction is initiated (D_{KL}, running over 3 days) signals a large error in the estimated concentration. In section 3.2, we showed that the proxy assumptions may not be valid at low wind speed when measured NO_2 concentrations are mostly a result of local emissions that are likely different from those at the proxy site. We compared the fit between the corrected sensor NO_2 concentrations and the regulatory concentrations for different wind directions and low versus high wind speed, but did not find any distinct patterns (Figures S13 – S16: hexbin scatter plots and error distribution across different wind directions/wind speed). The error distributions across the wind speeds and directions are close to Gaussian with a standard deviation not significantly different from the overall RMSE. Fig. 16 also shows that meteorological conditions signalled by the reference network as indicating that the proxy is not satisfactory did not necessarily lead to errors which were significant with respect to the other errors in the estimation. Finally, a very small divergence could mean that the model has been over-fitted such that the corrected sensor distribution agrees well with the proxy despite the true data distribution being different. The data distributions shown in Fig. S9 suggest an issue in some cases, but the dependence of the RMSE on divergence shown in

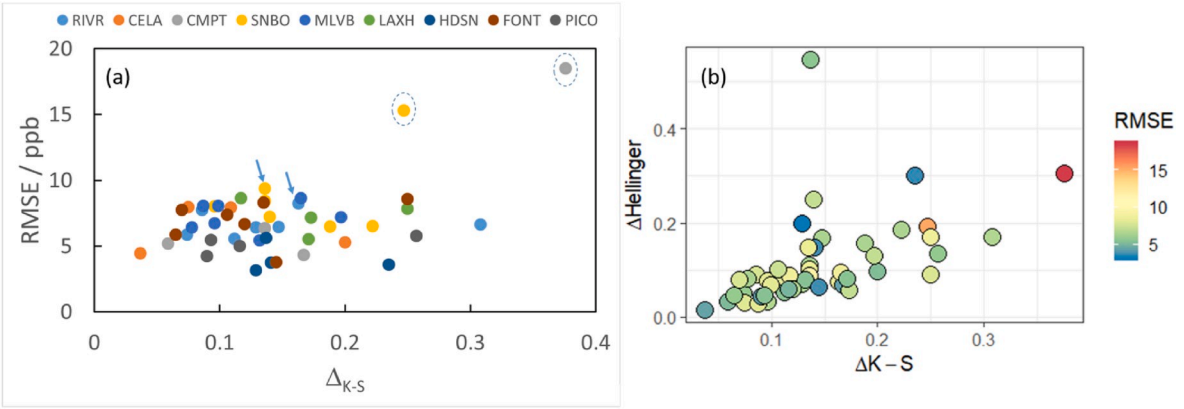


Fig. 16. Use of divergence measures between proxy and sensor-estimated concentrations to indicate the likelihood of larger errors in the estimated concentration; RMSE evaluated for each month at each site. Circled points are occurrences also marked by an unusually large value of the objective function used for the parameter determination (Kullback-Liebler divergence) Figs. 14 and S8: CMPT, July and SNBO, January). Points marked with arrows are occurrences where the proxy evaluation using the reference network indicated issues (RIVR and SNBO, June).

Fig. 16 does not indicate a problem.

Variation of water vapour pressure is known to have a significant effect on electrochemical sensors – particularly changes of offset, β_o (eq (6)), following rapid changes of humidity (Lewis et al., 2016). In Fig. S18, we show the distribution of the difference term between the framework-corrected sensor NO_2 and the regulatory NO_2 across different relative humidity quartiles. No distinct differences can be observed across different relative humidity quartiles, except at the highest, 71–100% RH, where the distribution may be bimodal, although there was no significant effect on the correlation with reference data (Figs. S17 and S18: hexbin scatter plots and error distribution across different RH bands). The error distributions are close to Gaussian with standard deviation not significantly different from the overall RMSE. Thus, the framework and offset error correction compensated for any effect of relative humidity variations.

The higher RMSE shown in Figs. 15 and 16, compared to the RMSE using the co-located regulatory O_3 and NO_2 to correct the data, is mostly related to issues with the fitting procedure applied using the proxy data (e.g. at CMPT in July, Fig. 14b) or to missing data from the proxy site. If proxy data are not available then the method simply uses the latest determined parameters. Specifically, the correction for e_S is not made. Inspection of the data showed indeed that the major error was due to incorrect estimation of e_S.

The potential of low-cost sensors to capture reliably episodes of high concentrations is of great importance for air quality measurements. Fig. 17 compares the number of times the low-cost sensor and the regulatory instruments recorded values > 75th percentile (20 ppb) per day and indicates that, in general, exceedances will be reliably indicated by the low-cost sensors managed as we have described (Spearman's rank correlation coefficient: 0.81). Comparison with Fig. 15 shows that the 'false positives' were associated with the site at CMPT, where, as noted above, the proxy comparison failed in July.

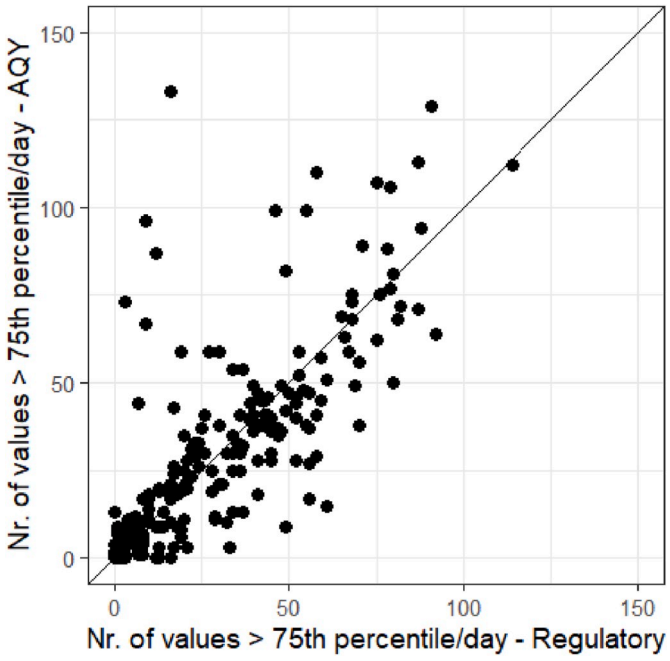


Fig. 17. Comparison between number of hourly AQY and regulatory measurements that exceeded the 75th percentile (20 ppb) per day across the whole study period (January to July) for all sites (Spearman's rank correlation coefficient: 0.81). There were up to 9 regulatory sites operating, hence a maximum of 216 hourly-averaged measurements per day.

3.5. Large local-scale spatial variations in nitrogen dioxide concentration revealed by the low-cost sensor network

The purpose of the low-cost network has been stated as the supplementary extension of a regulatory network to capture neighbourhood-scale variations. The method that we have described uses the regulatory network both to determine and validate the choice of proxy, and then to use the proxy distribution matching to check and re-calibrate if necessary the low-cost sensor network. Indeed, the low-cost sensor network revealed significant NO_2 concentration variations that were not captured by the regulatory network, as illustrated in Fig. 18a for one particular day, in Fig. 18d averaged over a month for specific hours of the day, and in Fig. S19 averaged over an entire month. The significance of the variations in relation to the measurement errors is assessed by comparison with the measurement error at particular reference sites at the same time (Fig. 18b) and by the mean and standard deviation of the error for each hour of the month determined at each reference station (Fig. 18c). There is no significant concentration-dependence of the error, which is indeed consistent with the analysis of the concentration-dependence of the errors given above and in the SI, and close to the RMSE determined for the entire study (7 ppb). Given that the error distribution is approximately Gaussian, as shown in Fig. S10, local differences of 10 ppb could be discriminated with ~70% confidence and of 20 ppb with ~90% confidence. Both high and low concentrations of NO_2 were very localized and transient, varying between extremes close to the highway network, varying significantly over relatively short distances and also tending to be higher near the mountains at the sides of the valleys. Given the errors, the magnitude of the variations is such that the network could nevertheless reveal the variations with confidence. We have in other work shown how to use data from the low-cost instrument network in this region alongside land-use correlations and wind speed-direction information to understand the spatio-temporal variation and identify specific, unusual features (Weissert et al., 2020) following ideas developed originally in Weissert et al., (2019).

4. Conclusion

In this paper, we have shown that useful and reliable data can be obtained from low-cost electrochemical NO_2 sensors operated in a dense hierarchical network which has a limited number of high-quality, regulatory-grade measurement sites. We achieved a RMSE of ~7 ppb, which was adequate to resolve reliably the high spatial and temporal variation in NO_2 concentration in the complex urban environment in Southern California. We used the regulatory sites as proxies, matching the probability distributions of sensor and proxy data running over a time window, to correct the sensor data. A satisfactory proxy could be identified by land-use similarity and could be a significant distance away from the sensor site. This is an extension of a management framework previously developed to detect and correct for drift in O_3 concentrations measured by low-cost air quality sensors.

When the management framework triggered an alarm, we minimised the Kullback-Leibler divergence between the distribution of the proxy data and the low-cost sensor data by adjusting of the sensor measurement model parameters. A critical element was the use of a semiconducting oxide-based ozone sensor to provide a robust and reliable ozone signal to correct for ozone interferences on the NO_2 sensor. The most significant effect on the error was an uncompensated variation of the baseline current of the electrochemical sensor on a timescale shorter than the distribution averaging timescale. This error was in part spatially correlated and had diurnal variations similar to the variations of ambient temperature, which allowed the error to be partially determined by using the closest proximity reference station with a co-located NO_2 sensor as a proxy. Sensor failure could be distinguished through a characteristic time variation of the derived parameters of the sensor measurement model. The results also indicated that failures of this approach, likely due to issues with compensation of the temperature-

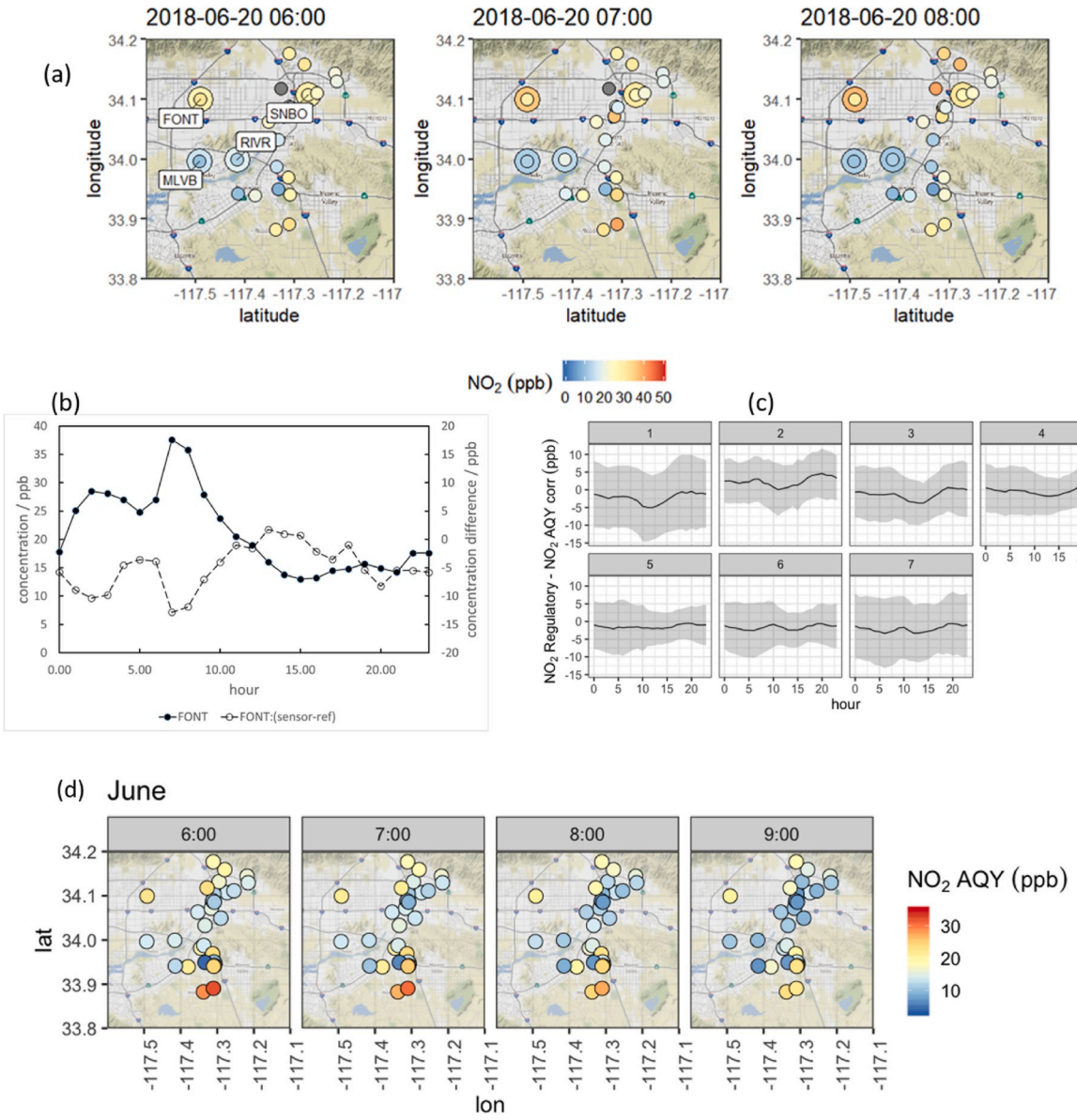


Fig. 18. (a) Example of neighbourhood-scale variation in NO_2 concentration, in the Riverside-San Bernadino region of Southern California, revealed by the low-cost sensor network, for three successive hours of a particular day. The reference network instruments and the concentration that they report are marked by the large circles. (b) Site concentration measured by the reference instrument at FONT, and the difference between the framework-corrected AQY result and the reference result at the same site, over the same day (20 June 2018). (c) Mean diurnal variation (line) and its standard deviation (shading) at all sites for each month (numbered 1, January – 7, July 2018). (d) Map of the mean diurnal variation for June 2018 in the Riverside-San Bernadino region of Southern California, revealed by the low-cost sensor network.

dependent offset, parameter correlation, or to differences in local emission sources, or the lack of suitable proxy sites, could be signalled through consideration of the time variation of the corrected sensor parameters, of the value of the Kullback-Leibler objective function, and of other measures of statistical divergence between proxy data and corrected sensor data. The simple rule-based framework is easily modified to change proxy or signal uncertainty when such conditions occur, and also when other factors arise, such as particular wind direction or speed conditions, where the proxy is known (from other assessment using the reference network) to be unreliable. While the method is robust, it does require a network of reference-grade instruments that is sufficiently diverse to sample all the environments within the zone to be measured. Since the method compares probability distributions running over time, it is robust against short stretches of missing data. However, the method does require data availability, not only from the low-cost network but also of ozone and nitrogen dioxide measurements from the reference network.

Supplementary information

Kullback-Leibler, Hellinger and Kolmogorov-Smirnov divergence: definition; LUR studies published for the North American region; Comparison of site monthly concentration distribution with a log-normal model; Comparison of land-use proxy result with closest proximity proxy result, using reference station data; Fitted parameter variation over time for all sites; Spatially-dependent offset error analysis; Framework-and offset error-corrected results for all sites; Concentration-dependence of the error; Correlation of framework- and offset error-corrected sensor data with reference data grouped according to different wind direction, wind speed and humidity; Example maps of variation of mean concentration of NO_2 across the region; References to land-use studies of NO_2 concentration in North America.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: LW, EM, KA and GSH are employees of Aeroqual Ltd, manufacturer of the sensor nodes used in these studies. GSH and DEW are founders and shareholders in Aeroqual Ltd.

CRediT authorship contribution statement

Lena Weissert: Writing - original draft, Formal analysis, Validation, Software. **Elaine Miles:** Software, Data curation, Validation. **Georgia Miskell:** Software, Data curation. **Kyle Alberti:** Investigation, Data curation. **Brandon Feenstra:** Data curation, Supervision, Writing - review & editing. **Geoff S. Henshaw:** Validation, Data curation, Supervision. **Vasileios Papapostolou:** Supervision, Writing - review & editing. **Hamesh Patel:** Investigation. **Andrea Polidori:** Supervision, Writing - review & editing. **David E. Williams:** Conceptualization, Funding acquisition, Formal analysis, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgements

This work was funded by the New Zealand Ministry for Business, Innovation and Employment, contract UOAX1413. This work was performed in collaboration with the Air Quality Sensor Performance Evaluation Center (AQ-SPEC) at the South Coast Air Quality Management District (South Coast AQMD). The authors would like to acknowledge the work of Mr. Berj Der Boghossian for his technical assistance with deploying AQY sensor nodes. The authors would like to acknowledge the work of the South Coast AQMD Atmospheric Measurements group of dedicated instrument specialists that operate, maintain, calibrate, and repair air monitoring instrumentation to produce regulatory-grade air monitoring data. DEW acknowledges the support of a fellowship program at the Institute of Advanced Studies, Durham University, UK.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2020.117428.

References

Alavi-Shoshtari, M., Williams, D.E., Salmond, J.A., Kaipio, J.P., 2013. Detection of malfunctions in sensor networks. Environmetrics 24, 227–236. https://doi.org/10.1002/env.2206.

Aliwell, S.R., Halsall, J.F., Pratt, K.F., O'Sullivan, J., Jones, R.L., Cox, R.A., Utembe, S.R., Hansford, G.M., Williams, D.E., 2001. Ozone sensors based on WO_3: a model for sensor drift and a measurement correction method. Meas. Sci. Technol. 12, 684–690.

Air Quality Management Plan (Aqmp), 2016. South Coast AQMD Planning. Rule Development & Area Sources Division, Diamond Bar, CA.

Bart, M., Williams, D.E., Ainslie, B., McKendry, I., Salmond, J., Grange, S.K., Alavi-Shoshtari, M., Steyn, D., Henshaw, G.S., 2014. High density ozone monitoring using gas sensitive semi-conductor sensors in the Lower Fraser Valley, British Columbia. Environ. Sci. Technol. 48, 3970–3977. https://doi.org/10.1021/es404610t.

Clements, A.L., Griswold, W.G., Rs, A., Johnston, J.E., Herting, M.M., Thorson, J., Collier- Oxandale, A., Hannigan, M., 2017. Low-cost air quality monitoring tools: from research to practice (a workshop summary). Sensors 17. https://doi.org/10.3390/s17112478.

Coyle, M., Smith, R.I., Stedman, J.R., Weston, K.J., Fowler, D., 2002. s. Atmos. Environ. 36, 1013–1024.

Guttorp, P., Meiring, W., Sampson, P.D., 1994. A space-time analysis of ground-level ozone data. Environmetrics 5, 241–254.

Hansford, G.M., Freshwater, R.A., Bosch, R.A., Cox, R.A., Jones, R.L., Pratt, K.F., Williams, D.E., 2005. A low cost instrument based on a solid state sensor for balloon-borne atmospheric O_3 profile sounding. J. Environ. Monit. 7, 158–162. https://doi.org/10.1039/b412184h.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42, 7561 – 7478.

Hubbell, B.J., Kaufman, A., Rivers, L., Schulte, K., Hagler, G., Clougherty, J., Cascio, W., Costa, D., 2018. Understanding social and behavioral drivers and impacts of air quality sensor use. Sci. Total Environ. 621, 886–894. https://doi.org/10.1016/j.scitotenv.2017.11.275.

Isiugo, K., Newman, N., Jandarov, R., Grinshpun, S.A., Reponen, T., 2018. Assessing the accuracy of commercially available gas sensors for the measurement of ambient ozone and nitrogen dioxide. J. Occup. Environ. Hyg. 15, 782–791. https://doi.org/10.1080/15459624.2018.1513135.

Lewis, A.C., Lee, J.D., Edwards, P.M., Shaw, M.D., Evans, M.J., Moller, S.J., Smith, K.R., Buckley, J.W., Ellis, M., Gillot, S.R., White, A., 2016. Evaluating the performance of low cost chemical sensors for air pollution research. Faraday Discuss 189, 85–103.

Li, H.Z., Gu, P., Ye, Q., Zimmerman, N., Robinson, E.S., Subramanian, R., Apte, J.S., Robinson, A.L., Presto, A.A., 2019. Spatially dense air pollutant sampling: implications of spatial variability on the representativeness of stationary air pollutant monitors. Atmos. Environ. X 2. https://doi.org/10.1016/j.aeaoa.2019.100012, 100012.

McConnell, R., Berhane, K., Yao, L., Lurmann, F.W., Avol, E., Peters, J.M., 2006. Predicting residential ozone deficits from nearby traffic. Sci. Total Environ. 363, 166–174.

Mead, M.I., Popoola, O.A.M., Stewart, G.B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J.J., McLeod, M.W., Hodgson, T.F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J.R., Jones, R.L., 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos. Environ. 70, 186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060.

Miskell, G., Alberti, K., Feenstra, B., Henshaw, G., Papapostolou, V., Patel, H., Polidori, A., Salmond, J.A., Weissert, L.F., Williams, D.E., 2019. Reliable data from low-cost ozone sensors in a hierarchical network. Atmos. Environ. 214, 116870.

Miskell, G., Salmond, J., Alavi-Shoshtari, M., Bart, M., Ainslie, B., Grange, S., McKendry, I.G., Henshaw, G.S., Williams, D.E., 2016. Data verification tools for minimizing management costs of dense air-quality monitoring networks. Environ. Sci. Technol. 50, 835–846. https://doi.org/10.1021/acs.est.5b04421.

Miskell, G., Salmond, J.A., Williams, D.E., 2018. Solution to the problem of calibration of low-cost air quality measurement sensors in networks. ACS Sens. 3, 832–843. https://doi.org/10.1021/acssensors.8b00074.

Popoola, O.A.M., Carruthers, D., Lad, C., Bright, V.B., Mead, M.I., Stettler, M.E.J., Jones, R.L., 2018. Use of networks of low cost air quality sensors to quantify air quality in urban settings. Atmos. Environ. 194, 58–70. https://doi.org/10.1016/j.atmosenv.2018.09.030.

Sadighi, K., Coffey, E., Polidori, A., Feenstra, B., Lv, Q., Henze, D.K., Hannigan, M., 2018. Intra-urban spatial variability of surface ozone in Riverside, CA: viability and validation of low-cost sensors. Atmos. Meas. Tech. 11, 1777–1792. https://doi.org/10.5194/amt-11-1777-2018.

Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S., Shelow, D., Hindin, D.A., Kilaru, V.J., Preuss, P.W., 2013. The changing paradigm of air pollution monitoring. Environ. Sci. Technol. 47, 11369–11377. https://doi.org/10.1021/es4022602.

Utembe, S.R., Hansford, G.M., Sanderson, M.G., Freshwater, R.A., Pratt, K.F.E., Williams, D.E., Cox, R.A., Jones, R.L., 2006. An ozone monitoring instrument based on the tungsten trioxide (WO_3) semiconductor. Sensor. Actuator. B Chem. 114, 507–512. https://doi.org/10.1016/j.snb.2005.04.049.

van Zoest, V., Osei, F.B., Stein, A., Hoek, G., 2019. Calibration of low-cost NO_2 sensors in an urban air quality network. Atmos. Environ. 210, 66–75. https://doi.org/10.1016/j.atmosenv.2019.04.048.

Weissert, L.F., Alberti, K., Miskell, G., Pattinson, W., Salmond, J.A., Henshaw, G., Williams, D.E., 2019. Low-cost sensors and microscale land use regression: data fusion to resolve air quality variations with high spatial and temporal resolution. Atmos. Environ. 213, 285–295. https://doi.org/10.1016/j.atmosenv.2019.06.019.

Weissert, L., Alberti, K., Miles, E., Miskell, G., Feenstra, B., Henshaw, G.S., Papapostolou, V., Hamesh, P., Polidori, A., Salmond, J.A., Williams, D.E., 2020. Low-cost sensor networks and land-use regression: interpolating nitrogen dioxide concentration at high temporal and spatial resolution in Southern California. Atmos. Environ. 223, 117287. https://doi.org/10.1016/j.atmosenv.2020.117287.

Williams, D.E., 2019. Low cost sensor networks: how do we know the data are reliable? ACS Sens. 4, 2558–2565.

Williams, D.E., Aliwell, S.R., Pratt, K.F., Caruana, D.J., Jones, R.L., Cox, R.A., Hansford, G.M., Halsall, J., 2002. Modelling the response of a tungsten oxide semiconductor as a gas sensor for the measurement of ozone. Meas. Sci. Technol. 13, 923–931.

Williams, D.E., Henshaw, G.S., Bart, M., Laing, G., Wagner, J., Naisbitt, S., Salmond, J.A., 2013. Validation of low-cost ozone measurement instruments suitable for use in an air- quality monitoring network. Meas. Sci. Technol. 24, 065803 https://doi.org/10.1088/0957-0233/24/6/065803.