A solution to the problem of calibration of low-cost air quality measurement sensors in networks

Georgia Miskell, Jennifer Salmond, and David Edward Williams

ACS Sens., **Just Accepted Manuscript** • DOI: 10.1021/acssensors.8b00074 • Publication Date (Web): 06 Mar 2018

A solution to the problem of calibration of low-cost air quality measurement sensors in networks

Georgia Miskell [a,b], Jennifer A. Salmond [b], and David E. Williams [a,c,*]

[a] School of Chemical Sciences, University of Auckland, New Zealand

[b] School of Environment, University of Auckland, New Zealand

[c] MacDiarmid Institute for Advanced Materials and Nanotechnology, Wellington, New Zealand

david.williams@auckland.ac.nz

Abstract

We provide a simple, remote, continuous calibration technique suitable for application in a hierarchical network featuring a few well-maintained, high-quality instruments ("proxies") and a larger number of low-cost devices. The ideas are grounded in a clear definition of the purpose of a low-cost network, defined here as providing reliable information on air quality at small spatiotemporal scales. The technique assumes linearity of the sensor signal. It derives running slope and offset estimates by matching mean and standard deviations of the sensor data to values derived from proxies over the same time. The idea is extremely simple: choose an appropriate proxy and an averaging-time that is sufficiently long to remove the influence of short-term fluctuations but sufficiently short that it preserves the regular diurnal variations. The use of running statistical measures rather than cross-correlation of sites means that the method is robust against periods of missing data. Ideas are first developed using simulated data and then demonstrated using field data, at hourly and 1 min time-scales, from a real network of low-cost semiconductor-based sensors. Despite the almost naïve simplicity of the method, it was robust for both drift detection and calibration correction applications. We discuss the use of generally available geographic and environmental data as well as microscale land-use regression as means to enhance the proxy estimates and to generalize the ideas to other pollutants with high spatial variability, such as nitrogen dioxide and particulates. These improvements can also be used to minimize the required number of proxy sites.

Keywords

Air quality, low-cost sensor network, blind calibration, semiconducting oxide, electrochemical sensor, ozone, nitrogen dioxide

Recently, interest has developed in supplementing the use of sparse networks of precision regulatory instruments in air quality management with measurement using high density networks of lower cost instruments which aim to provide reliable information to citizens at local-scales[1-5]. Field studies have previously illustrated the potential of such high-density networks to reveal pollutant patterns not evident in the more sparsely distributed regulatory networks[6-13]. In part, the change is being driven by the falling costs of computation and communication. A weak link is instrumentation and sensors, but in recent years these have also significantly advanced[12,14-19]. As a consequence, calibration costs and data verification tools are areas that now demand attention in order properly to exploit the potential of low-cost air quality sensor networks[3,5,14,20-23]. Currently, air quality instruments require frequent, rigorous calibration to provide reliable data meeting regulatory-standards. Calibration is a significant operational cost, which may be as important as instrument cost in limiting the number of instruments practically-achievable in a dense network deployment[6,24,25].

"Factory calibration" is used here to describe the calibration by the manufacturer to align sensor signals to the variable of interest. Factory calibration can often be inappropriate for environmental monitoring because varying meteorological and other important controls (such as the influence of other pollutants) are not accounted for by such calibration[6,8,23]. Thus, a second field or in-situ calibration outside of factory settings is usually required before the sensor can provide meaningful environmental results[7,13,18]. This is referred to here as "field calibration". Regular calibration is required to ensure ongoing accuracy, especially when using sensors that are prone to drift[26].

A popular field calibration technique for low cost sensors is periodic co-location of the instrument with a regulatory standard instrument. This technique is appealing because of its simplicity and accuracy. Yet, it has some disadvantages. These include periods of missing data where the sensor has been removed for calibration, low network scalability due to intensive labor requirements to regularly move sensors for co-location, and an inability to adapt or correct for in-situ changes over time such as calibration drift. To overcome these

issues remote calibration techniques, where the sensor is corrected without co-location, have been devised[27,28]. One method is where the network data are expressed as a sum of spatial frequencies, which are oversampled; in other words, the spatial frequency of the measurement locations is greater than the lowest spatial frequency in the data so the data are cross-correlated between the different measurement locations[29,30]. Linearity of sensor response is assumed. Essentially, this method adjusts slope and offset of the individual devices to maximize cross-correlations between network locations and has been referred to as 'blind' calibration: 'blind' because this calibration method does not rely on ground truth data to guide expectations, although accurately known data from at least one location are required. A 'semi-blind' calibration was also introduced , where some further ground truth about the network was available[29,30]. This method is not however robust enough to deal with missing data and correlated interferences between different atmospheric constituents could be a significant issue. Another approach is to use verified data to correct for extraneous influences[23,31-33]. However, a practical limitation can be providing a sufficiently large training set of sufficiently good data. Data fusion, which is combining sensor measurements with model results, is an emerging and promising calibration technique[34]. However, the stability of the sensor response is assumed.

In previous work we developed a drift detection algorithm that exploited network cross-correlations[21]. The idea of a "proxy" dataset was introduced, and a starting point was a clear specification of the low-cost network purpose. Proxies were chosen based on land-use similarity. We also introduced a method in other work wherein the signal is split into a regular component (the general diurnal variations) and a fluctuating local-scale component. Autoregression and multiple linear regression with proxy sites, based on the fluctuating component, were able to distinguish regional from local influences and further identify drift[31]. Both of these proposed methods worked in real-time, averaging over a short (three-day) time-scale, and were reasonably robust against periods of missing data.

In the present paper, we extend this work and demonstrate a simple method for continuous remote calibration of individual devices. We exploit the idea of a hierarchical network, in which a few well-calibrated instruments act as 'proxies' to which other low-cost devices are referred. We also exploit the fact that air quality shows regular diurnal patterns and general correlations with land-use. Figure 1 illustrates the proposed network architecture, which features a 'compliant' layer of regulatory-standard instruments, a 'managed' layer of low-

cost instruments which are subject to the automated procedures of reliability checking previously described[21] and which are replaced as necessary, and the third layer of instruments, which are designed to have a minimum cost, minimal pre-checking before installation and whose calibration is updated using methods described here.

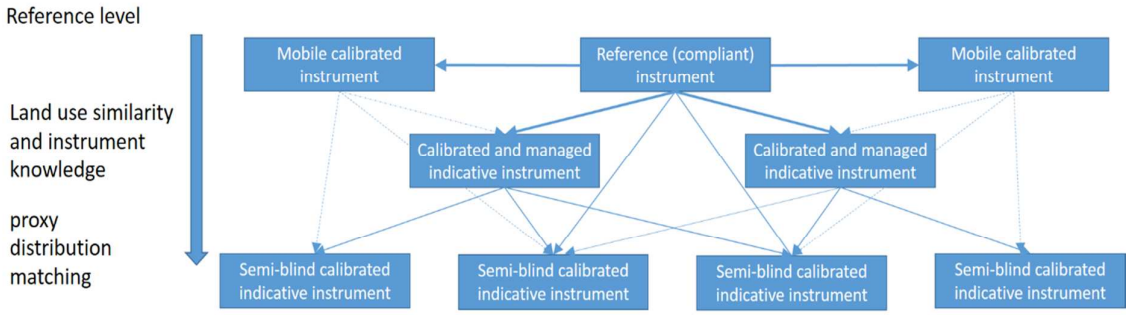


Figure 1: A network architecture for a minimum-maintenance, high spatial density air quality measurement network that exploits network cross-correlations. The proxy distribution matching method for 'semi-blind' calibration is described in the present paper.

The spatial density will normally increase down through the layers as costs become less for the lower layers. Mobile instruments would provide a further level of checking if a network was operating such a system[35]. As before, a pre-condition is defining a clear specification of the network purpose. A second requirement is that the calibration method should need minimal training data and be able to update routinely in near real-time. The method is demonstrated here using a dense network of devices, that are not pre-calibrated, on both short-term (1 min) and hourly-averaged ozone (O_3) data. Finally, we discuss the potential for land-use regression to extend the methods for pollutants such as nitrogen dioxide (NO_2) and particulates, that may be more spatially variable than O_3, and to minimize the required number of reference measurement sites. We discuss the important problem of cross-interferences between pollutants in the sensor response.

Theory

We define that the purpose of a high spatial density network is to provide reliable information on the occurrence of high concentrations at the local-scale with short time resolution since total exposure is often determined as the result of repeated exposure to episodic short-time-scale events. Here, the time resolution and acceptable uncertainty in

the reported values will determine the spatial density of proxies. What uncertainty thresholds are appropriate for local-scale low-cost sensor data is a subtle question that requires a clear understanding how the data might be used. This is a decision problem – accounting for uncertainty in the light of decisions to be made based on the data – that moves well beyond the scope of the present work but which is nevertheless important in setting the context[22]. We do not address it directly but focus on the reliability of reporting of occurrences above an arbitrary threshold.

Assumptions

a) Sensor data are a linear transformation of the true data:

If $X_i(t)$ denotes the true data value at location *i* and time *t*, then the sensor data, $Y_i(t)$ over a range of expected measurements, will satisfy:

$$X_i(t) = a_0 + a_1 Y_i(t) \quad (1)$$

We use the term "sensor data" to refer to the results delivered by the low-cost device: that is, to the device comprising the air sampling system, gas sensor and associated electronics and controls that delivers a signal according to (1). In practice, we have found this to be a robust assumption, for devices measuring NO_2 and O_3[7,18,21,36]. The validity in practice depends on the configuration of the sensor and the device that converts the sensor signal into the supplied measurement result, and a proper understanding of the effect of interferences from other atmospheric constituents. The O_3 device used is based on a semiconducting oxide where the output is linearized based on a factory calibration that is loaded onto the device, and which has been extensively validated in measurements both in the laboratory and the urban atmosphere[7,18]. Previous results have shown sensors to retain a linear response even though they may drift[7,27].

b) Frequency distribution of site and proxy data have similar functional form over a suitable time period:

By this statement we mean, for example, that both distributions should be lognormal with some limitations, that are to be explored, on the range of mean and variance. Alternatively, the distributions could be similar around the mean but differ in the tail. Below, we explore

empirically this statement. Slope and offset estimates can then be derived simply by matching mean and variance from (1), where the proxy distribution, *Z*, is taken as an estimate for the true data distribution, *X*.

If $Z_i(t\text{-}t_d\text{: }t)$ denotes the proxy data and $Y_i(t\text{-}t_d\text{:}t)$ denotes the sensor device data for site *i* over the time interval $(t - t_d : t)$ then we define the slope, $\widehat{a_1}$ and offset, $\widehat{a_0}$, estimates for the corrected data. That is, for the estimate of X_i at time *t*, $\hat{X}_{i,t}$:

$$\widehat{\alpha_1} = \sqrt{\frac{\text{var}\langle Z(t-t_d:t)\rangle}{\text{var}\langle Y(t-t_d:t)\rangle}} \tag{2}$$

$$\widehat{\alpha_0} = \text{E}\langle Z(t - t_d\text{: }t)\rangle - \hat{a}_1\text{E}\langle Y(t - t_d\text{: }t)\rangle \tag{3}$$

Where E< > denotes the arithmetic mean evaluated over the time period ($t\text{-}t_d\text{:}t$) and var< > denotes the arithmetic variance about the mean. Then:

$$\hat{X}_{i,t} = \hat{a}_0 + \hat{a}_1 Y_{i,t} \tag{4}$$

(2-4) can be contrasted with results from linear regression. Least squares linear regression takes the data as pairs, assumed correlated, with the result:

$$\frac{Z-\text{E}\langle Z\rangle}{\sqrt{\text{var}\langle Z\rangle}} = r_{Z,Y}\frac{Y-\text{E}\langle Y\rangle}{\sqrt{\text{var}\langle Y\rangle}} \tag{5}$$

Where $r_{Z,Y}$ denotes the pairwise correlation coefficient. In our method, we do not assume pairwise correlation but match the probability distributions, which is equivalent to setting $r_{Z,Y} = 1$ in (5). We will show that land-use correlation can be used to estimate $r_{Z,Y}$ and hence refine the estimates $\widehat{a_1}$ and $\widehat{a_0}$. The idea of matching distribution parameters averaged over time distinguishes our method from those that use correlations of the data.

We explored, using simulated data, by how much $\widehat{a_1}$ and $\widehat{a_0}$ estimates change when the proxy distribution differs from the true data distribution. Lognormal distributions[37] were used since this distribution, in general, gives a reasonable representation of air quality data[38,39]. Results are given in the Supporting Information (SI) illustrating that, even if the

proxy distribution is significantly different from the true data distribution, estimates that satisfy criteria for indicative air quality data are obtained.

According to (2-4), $\mathrm{E}\langle\hat{X}(t-t_d\colon t)\rangle = \mathrm{E}\langle Z(t-t_d\colon t)\rangle$ and $\mathrm{var}\langle\hat{X}(t-t_d\colon t)\rangle = \mathrm{var}\langle Z(t-t_d\colon t)\rangle$. Thus, if the distributions are characterized by only two parameters, then the site distribution over sample time t_d is constrained to be the same as the proxy distribution. Thus, there is a question of whether the corrected data deliver any useful information over and above that delivered by the proxy. The prior time period, t_d, imposes an averaging filter on the data. It is empirically chosen to obtain a reasonable estimate of the distributions of *Z* and *Y* but is also chosen to average the short-term fluctuations and emphasize the longer-term, regular component of the concentration variation within the distribution of values. In the framework of our previous study in which we decomposed the variation into regular and fluctuating parts[31] , here we are tending to match the diurnal component of the site and proxy in order to estimate the slope and offset of the examined sensor device. We show later that, with a suitable choice of t_d, the distribution of *X*, including the marginal distribution of exceedances of a threshold, is reliably estimated.

An alternative method to consider to decouple the estimate of the marginal distribution of *X* from the marginal distribution of the proxy, *Z*, is the use of data truncation. This is where the mean and standard deviation estimates are derived from a subset of the data, e.g. where the data above ("right truncation") or below ("left truncation") a threshold are removed. So, for example, the distributions of *Z* and *Y* could be matched only for values below the arithmetic mean. The use of data truncation to improve estimates using simulated data is demonstrated in the SI.

Methods

Data: Data were from a subset of nine sensors of a network of low-cost semiconducting oxide-based devices measuring ambient ground-level O_3 around the Lower Fraser Valley (LFV), British Columbia, Canada, including the city of Vancouver, in 2012[7]. The selected subset of sensors was co-located with regulatory analyzers: thus, the true data, *X*, were available for comparison with the estimate from the sensor data, $\hat{X}$. The station locations, together with the mean and standard deviation of O_3 concentration over the study period,

are shown in the SI. A continuous six-week run of data was available. Measurements were made every minute. Analyzer station mean values over the study period ranged from 12 – 24 ppb and the standard deviation about the mean was between 9 – 14 ppb. Analysis was done using R (version 3.3.3), with plots made using packages 'ggplot2'[40] and 'ggmap'[41].

Proxy selection and adjustment:

For the central band of the latitude of the LFV, which is the location for the present demonstration, a very simple assignment of proxies by broad-brush land-use classification proved effective. In the SI, we provide a more detailed discussion of proxy selection.

As in previous work[21], sites were grouped into three broad land-use categories within 1 km scale: urban (commercial and industrial), residential (residential and park) and rural (rural and agricultural). Table 1 lists the sensor co-location sites and their chosen proxy site.

Table 1: List of sites and proxies, with their land-use classification and the status of the sensor at the end of the deployment, independently determined[9]

Site	Sensor status	Land-use (< 1 km)	Selected proxy	Selected proxy land-use (< 1 km)	Selected proxy distance (km)
Abbotsford	Intact	Residential	Langley	Residential	19.7
Langley	Drift	Residential	Maple Ridge	Residential	13.3
Maple Ridge	Drift	Residential	Langley	Residential	13.3
North Delta	Drift	Residential	Langley	Residential	25.3
Pitt Meadows	Drift	Rural	Chilliwack	Residential	56.7
Port Moody	Drift	Residential	Second Narrows	Commercial	12.6
Richmond	Intact	Residential	Langley	Residential	39.7
Second Narrows	Drift	Urban	Burnaby Kensington Park	Residential	4.4
Surrey	Intact	Residential	Langley	Residential	10.1

Most co-located sites were in residential land-uses, and so selecting a suitable proxy with similar land-use was straightforward. There were three sensors in co-located locations that were somewhat difficult to find suitable independent proxies having similar land-uses within 1 km as described in[42]. For these three sites, we used proxies from analyzer stations that had the most similar land-use based on satellite imagery. These three measurement locations – Pitt Meadows, Port Moody, and Second Narrows – were checked for any calibration correction issues that may be linked back to this weakness in the proxy selection.

For sites within the same land-use category, we took r_{ZY} = 1 (5). If sites in a different land-use category were to be used as proxies, we estimated slope correction factors (r_{ZY} in (5)) in the following way (SI: S5). First, we regressed the hourly O_3 concentrations between regulatory stations in the different land-use categories, fixing the intercept at zero. The median value of the regression slope between stations in two different land-use categories was taken as the slope correction factor for comparison of data from a site in one of the land-use categories with a proxy in the other land-use category. The full data set of regression slopes is given in the SI (S5). Table 2 gives the derived slope corrections.

Table 2: Slope correction factors for proxy data when compared to locations in different land-uses.

Sensor land-use	Proxy land-use			# locations
	Urban	Residential	Rural	
Urban	1	0.76	0.68	3
Residential	1.23	1	0.87	11
Rural	1.28	1.03	1	3

Choice of averaging-time, t_d

Figure 2 shows the pairwise data correlation between two reference stations in the 'residential' land-use category for different running averaging-times, t_d. The data were collected at 1 min sampling frequency and any data pairs for which one or both values were missing were ignored. The correlation of both running mean and running standard deviation improves as t_d increases. We used a running averaging-time t_d = 72 hours. This value was chosen because it was sufficiently long both to smooth the effects of short-term fluctuations and to yield values of $\widehat{a_1}$, and $\widehat{a_0}$ that remained reasonably constant with an increase of t_d, but sufficiently short that it retained the diurnal variation and kept the initialization time to a minimum.

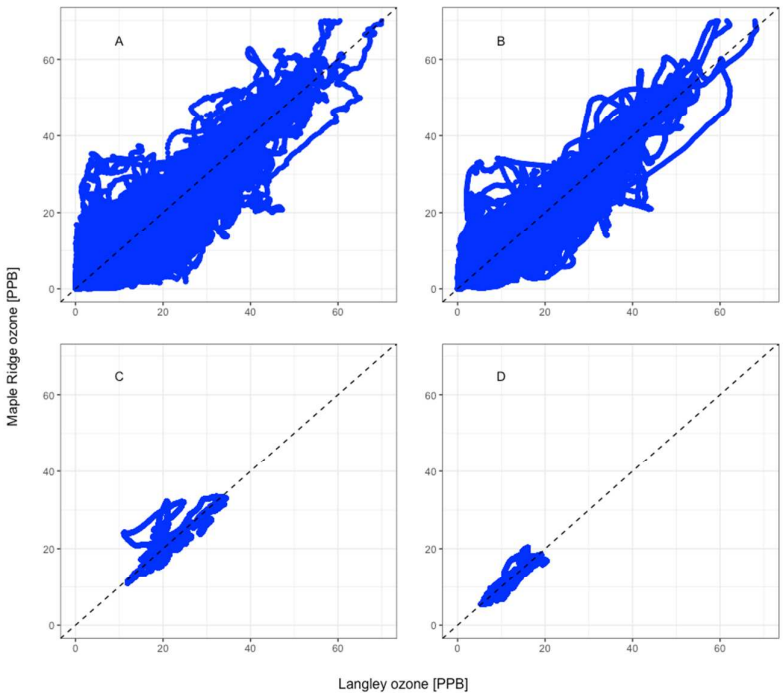


Figure 2: illustration of the effect of a running average on the correlation of ozone concentrations from two different stations in the Lower Fraser Valley, with measurements made at 1 min time resolution. A: t_d = 10-min mean ozone; B: 1-hour mean ozone; C: 72-hour mean ozone; D: 72-hour standard deviation ozone. The dotted line is the 1:1 fit.

Evaluated using a running t_d = 72-hour sample, both the mean and standard deviation are essentially identical: r_{ZY} = 1. Thus, the diurnal and regional variation dominates the signal at this time-scale. However, using a running 10-min time-scale illustrates the effect of local variations dominating the signal. Cross-correlations of hourly data for each LFV regulatory station measuring O_3 are in the SI.

Evaluation:

For hourly measurements, we evaluated the mean absolute error (MAE) of the corrected O_3 sensor signal from the actual concentration measured by the analyzer with which it was co-located. For both hourly and 1 min measurements, we compared the number of exceedances of a threshold concentration recorded by the corrected sensor signal with the number recorded by the analyzer. We arbitrarily chose a threshold that defined the top 5% of all regulatory station readings. For hourly data, this was for concentrations above 41.9 ppb; and for 1 min data, this was 46.2 ppb. We made a simple count of the fraction of the recorded exceedances that were “false negatives” (FN) - the corrected sensor signal was

below the threshold whilst the analyzer signal was above - and which were "false positives" (FP) - the sensor signal was above the threshold whilst the analyzer signal was below.

Results

The hourly raw data from the analyzers and their co-located sensors is given in the SI (S5).

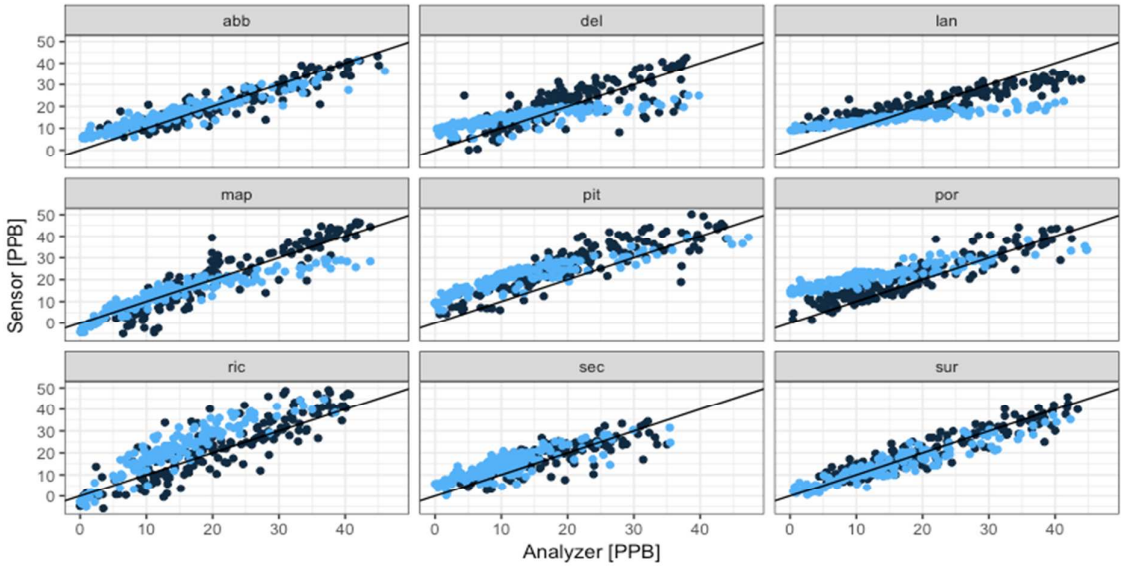


Figure 3: scatterplots with 1:1 line of the first (dark) and last (light) week of co-located measurements. Data is hourly factory-calibrated sensor data.

Figure 3 shows the correlation between sensor and analyzer for the first and last week of the 6-week deployment. Some devices had calibration faults from the start of monitoring, some showed a calibration change and drift part-way through deployment, and some remained stable for the entire examined period. Despite the drifts and mis-calibration, all co-located sensors remained linearly responsive to O_3, which is a requirement of our method. Similar behavior has been reported by others[27]. Failure modes for the devices have been discussed extensively[18]. A major cause of the drift was particle deposition within the sensor, linked to forest fires in Siberia and to a large fireworks display[7]. Previous work has shown how to detect these drifts and mis-calibrations[4,7,21,43,44]. Co-location ("field calibration") gives the true values of a_1 and a_0, (1), which we evaluated using the first week and the last week of hourly data for all nine devices. Then, with the selected proxy, $\widehat{a_1}$ and

$\widehat{a_0}$ were calculated for the same time periods ("semi-blind calibration"). Figure 4 compares the two sets of values. The semi-blind calibration according to (2 – 4) gave excellent results, even for very large drifts. In the SI, we give weekly scatterplots for all the sensors, corrected according to the methods of (2 – 4) further illustrating the excellent agreement obtained. Even in the cases where proxies were in different land-use, results were acceptable provided data accuracy expectations were relatively low.

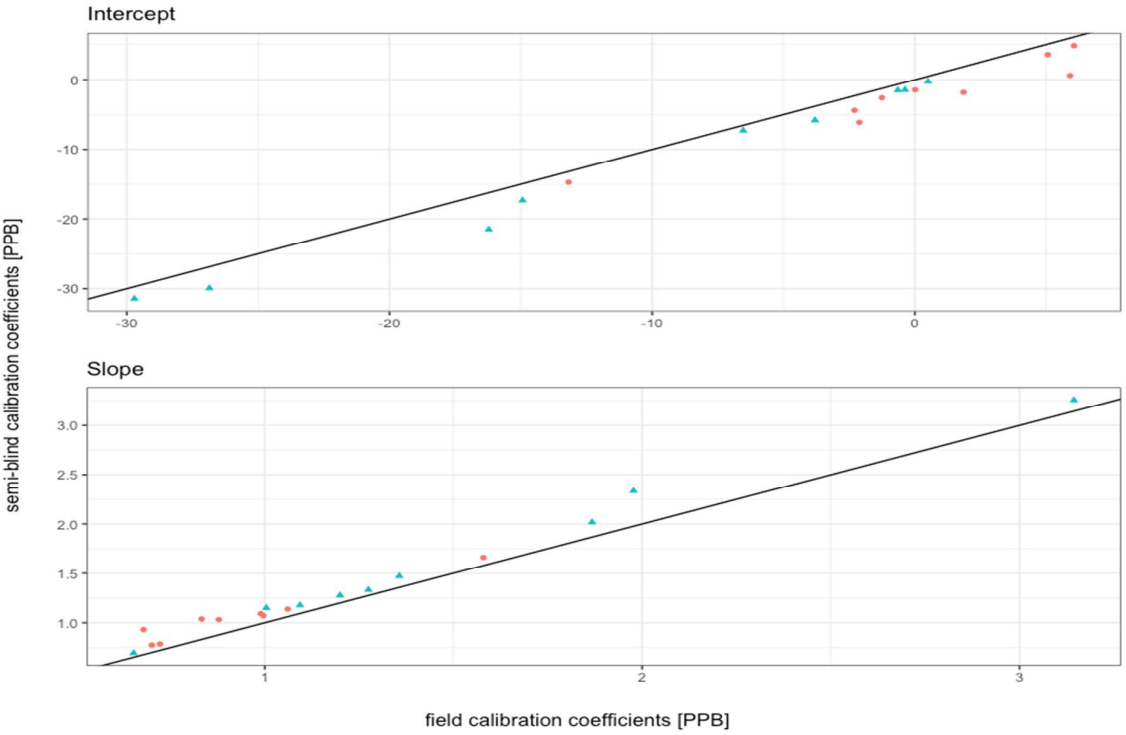


Figure 4: first (red) and last (blue) calibration coefficients for all co-located sites using t_d = 72 hr: semi-blind calibration values $\widehat{a_1}$, (slope) and $\widehat{a_0}$ *(intercept) against field calibration values a_1 and a_0. Line is 1:1 fit.*

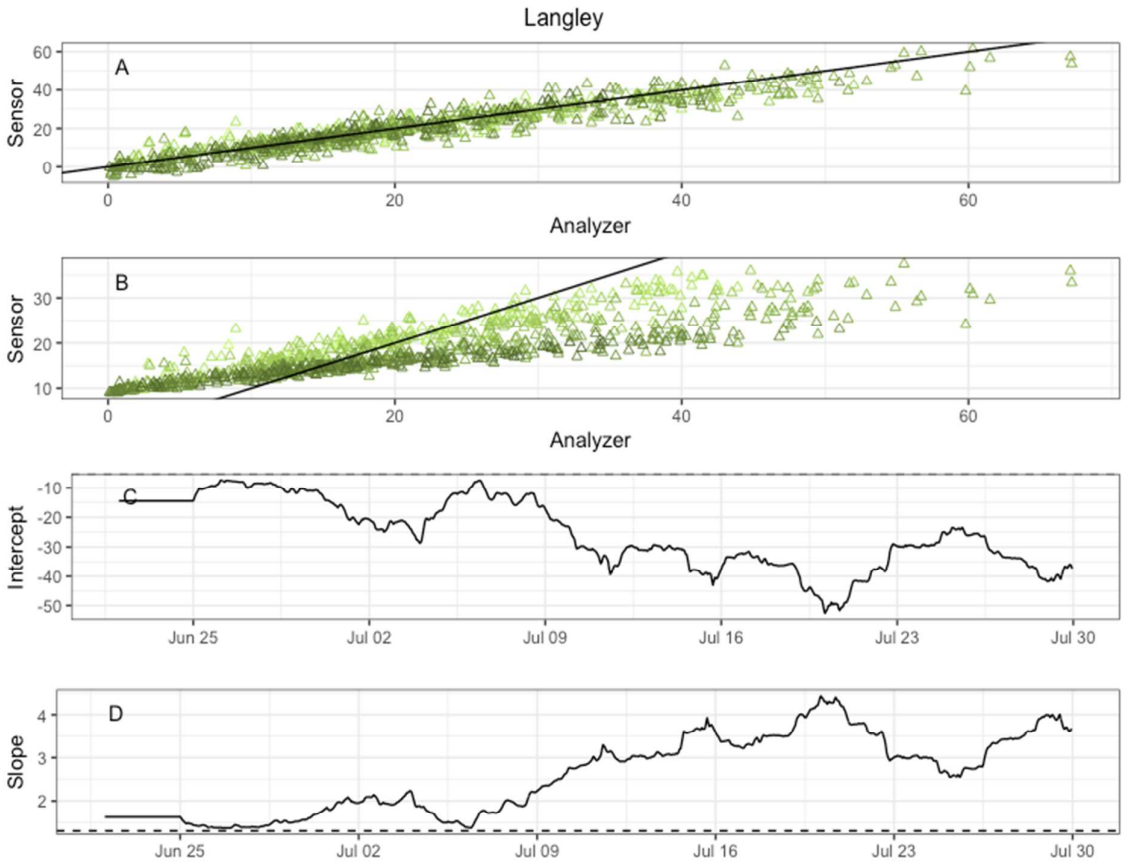


Figure 5: A – scatterplot of analyzer and semi-blind corrected sensor hourly data (earlier weeks = lighter colors and later weeks = darker colors). B – similar to A using raw sensor data. C – stepping intercept estimate, $\widehat{a_0}$, and D – stepping slope estimate, $\widehat{a_1}$ with dashed lines the drift detection thresholds used in [22].

Figure 5 shows the week-by-week performance of the method in the case of a sensor that was impacted by the mid-July fires, and also had an initial factory calibration that was in error. The agreement between the corrected sensor and the analyzer is excellent (figure 5 A). The semi-blind method was successful in calibrating and routinely updating correct coefficient values for a poor sensor without any initial co-location requirement, despite the corrections being large. The proxy comparison also clearly showed that the factory calibration was in error.

Table 3 summarizes results using the MAE between the sensor signal and the co-located analyzer.

Table 3: MAE scores (ppb) for the nine co-located sites, using all six weeks of data and the final week of data. Comparison of the raw sensor data with the semi-blind calibrated result; hourly data

Site	All data		Final week	
Calibration	Factory	Semi-blind	Factory	Semi-blind
Abbotsford	3.3	3.2	3.7	2.8

Langley	6.4	3	5.8	2.3
Maple Ridge	3.8	3.5	3.4	3
North Delta	4.8	5.6	5.5	5
Pitt Meadows	6.5	5.1	8.1	5
Port Moody	6	4.1	9.3	3.8
Richmond	5.9	4.2	7.5	3.2
Second Narrows	4.2	4.7	4.7	4.3
Surrey	3.1	3.5	2.6	2.8

Devices that drifted showed MAE scores that were larger in the final week. The semi-blind calibration method significantly improved these scores. Results were satisfactory without using any refinement with a more detailed application of land-use regression to improve the mean and variance estimates (see later discussion) with some MAE scores impressive.

The efficacy of using slope correction factors (table 2) to allow the use of proxies in different land-use classifications was assessed. We used both a heavily urban (Robson Square) and rural mountain location (Hope Airport) as examples. Figure 6 shows the MAE scores for all the sensor sites using these two proxies. The simple slope correction factors improved the accuracy of the predictions.

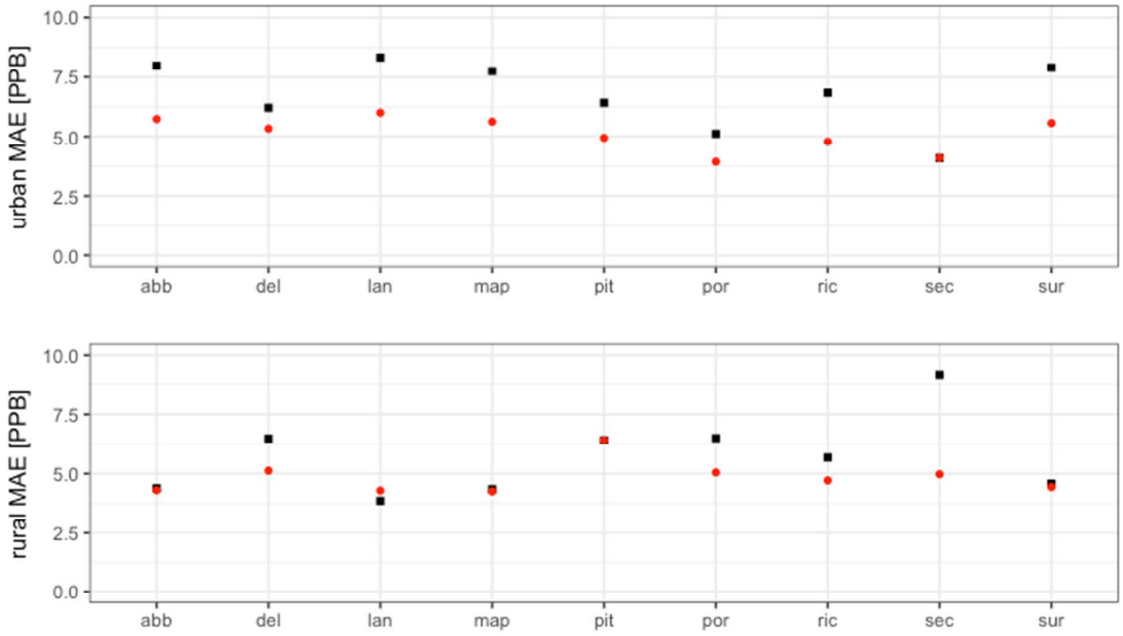


Figure 6: MAE scores between each co-located analyzer and semi-blind corrected sensor hourly data using Robson Square (urban) and Hope Airport (rural) as proxies. Black squares = unadjusted proxy data; red circles = slope correction coefficients applied. Those where only a red circle is present are where proxy land-uses were the same.

In the Theory section, we noted that, if the distributions are characterized by only two parameters, then the site distribution over t_d is constrained to be the same as the proxy distribution. Thus, there is a question of whether the corrected data deliver any useful information over and above that delivered by the proxy. Particularly, there is the question whether the frequency of exceedances of a threshold is constrained to be the same as the proxy. Two ideas were offered: first was that the use of a sufficiently long averaging-time, t_d, meant that short time-scale fluctuations of the site and proxy data were decoupled; and second was that data truncation could be used to match site and proxy only over a part of the range. One co-located site was selected as an example from the network where the sensor O_3 signal significantly drifted and the co-located regulatory station observed periods of particularly high concentration (Langley: figure 5), and the method was applied to 1 min time resolution data. The semi-blind calibration coefficients used results from the hourly data using t_d = 72 hr. The corrected sensor results followed the co-located regulatory data well, not the proxy (figure 7). Thus, using another site as a proxy in the semi-blind correction method did not mask those short-time-scale site-specific details that are important in providing information beyond that given by the proxy.

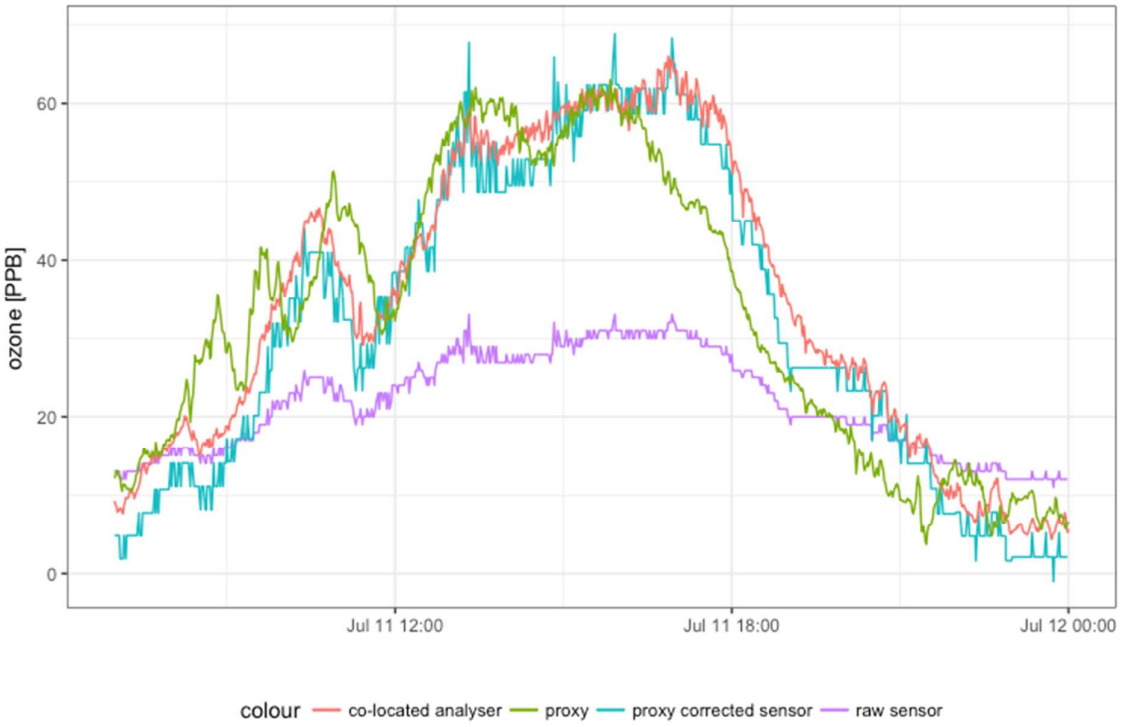


Figure 7: section of the 1 min data for a co-located site (Langley), comparing the regulatory analyzer and proxy data with the raw and proxy-corrected sensor data.

A scatterplot of the high 1 min concentrations from Langley shows that both the left truncation and the complete dataset gave similar results. Right truncation gave values that were consistently lower (figure 8). One may deduce that the proxy and sensor data were similar in distribution above the mean and somewhat dissimilar below the mean. This is consistent with direct observation of the raw data (SI: S5): the sensor signal had become compressed over the concentration range and particularly the minimum concentration reported (found at night) had been elevated.

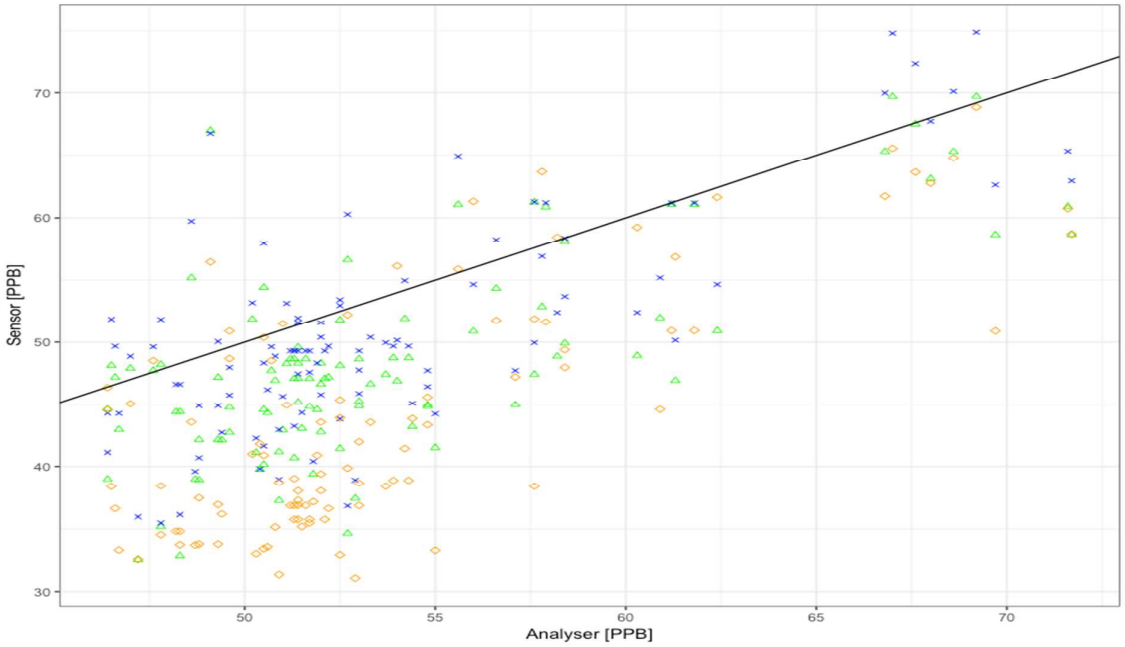


Figure 8: scatterplot of the 1 min data from Langley location using calibration coefficients derived from the hourly data results (t_d = three days). A random sample of 1000 observations is used for plot clarity. Green triangles = full dataset. Blue crosses = left truncation at the mean. Orange diamonds = right truncation at the mean.

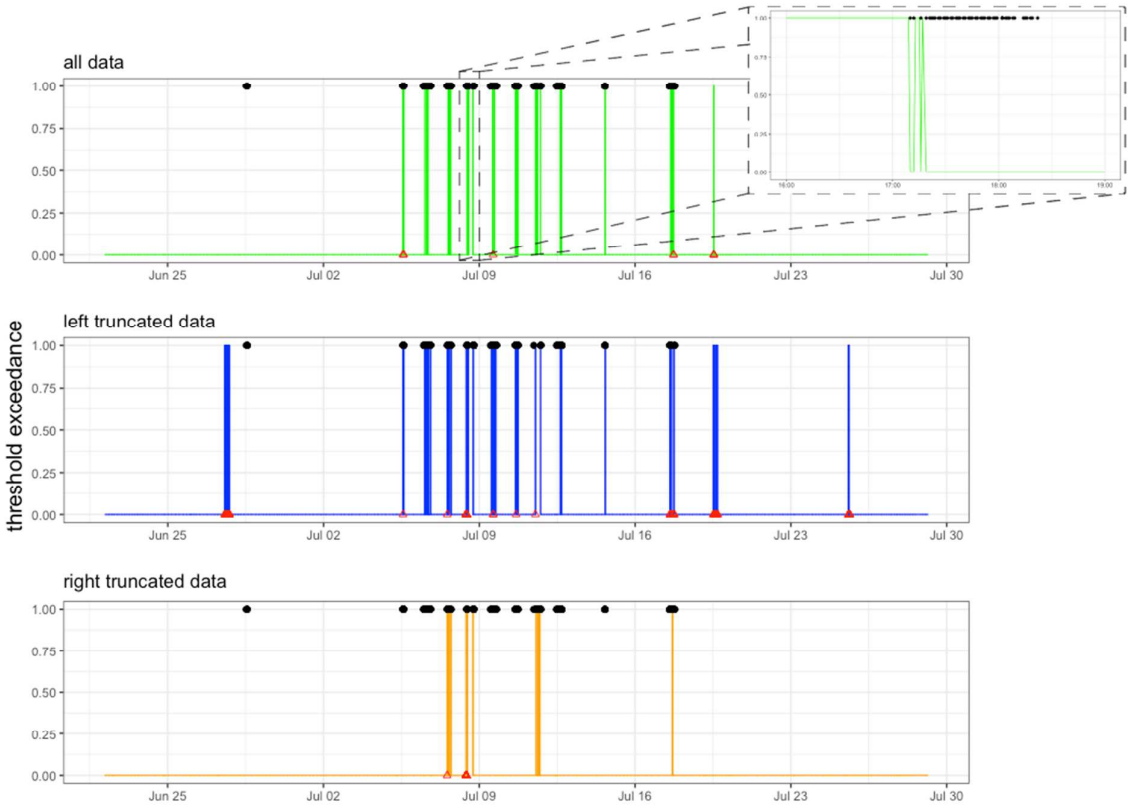


Figure 9: 1 min high concentration data as a binary signal for Langley location using the three different calibration results (all data or a truncation). Colored lines show the corrected sensor results in reporting high concentrations (>46.2ppb), where 1 = where a high value has been found, and 0 otherwise. Black points at 1 are false negatives, where a high observation is recorded by the co-located analyzer but not the corrected sensor. Red triangles at 0 are false positives, where a high observation is recorded by the corrected sensor but not the co-located analyzer. The inset shows a close-up where errors may be due to the two instruments recording different air samples causing slight lags.

Figure 9 shows a simple time series of exceedances of the threshold, for the corrected sensor signal using all data and left- and right-truncations at the mean. Marked on the diagram are where the co-located regulatory analyzer recorded exceedances. It is clear that the corrected sensor data using all data or left-truncation reliably recorded the exceedances, with a very small number of false negatives of short duration. The false negatives were removed by right-truncation but some exceedances were missed. The best accuracy and reliability fit for reporting short-term high concentrations when using the threshold of 46.2 ppb were where all t_d recent data were used to derive calibration coefficients. For these data, 86% of high regulatory concentrations were also recorded as high by the sensor, with 24% of corrected sensor results signaling high values being false positives. Left truncation results also showed 86% of corrected data were over the threshold

when regulatory results were also high. A higher false positive rate – 34% - was recorded, however. The right truncation calibration result showed 82% of sensor values correctly identifying high episodes, and 56% of reported high values were false positives. Thus, the scheme was successful in reliably calibrating sensors and on reporting high concentration episodes at short time-scales. Comparison of the reporting of exceedances using different truncation schemes could be developed to provide an estimate of uncertainty for the reporting. The reliability of reporting of exceedances was less if hourly data were used (SI: S7). The improvement in results in comparison to those found for hourly data is simply because averaging tends to bias the corrected data towards the proxy.

Discussion

The method of deriving calibration coefficients by matching the mean and standard deviation of the data to that of a proxy has been shown to be a robust means of correcting data from drifting or mis-calibrated low-cost O_3 sensors. The idea is extremely simple: to choose an appropriate proxy and running averaging-time t_d that is sufficiently long to remove the influence of short-term fluctuations but sufficiently short that it preserves the regular diurnal variations. Proxy choice based on land-use similarity has been demonstrated to be effective. The use of a simple correction factor for relating a device to a proxy in a different land-use was also shown to be effective. The method gave a reliable running recalibration of devices that had significantly drifted, and of devices that had initially been mis-calibrated. Sensor data corrected using our method reliably reported high concentration exceedances for both hourly and 1 min data. Use of data truncation in the proxy matching identified where the proxy and sensor data distributions differed, and could be developed into a method for blind assessment of the reliability of the results.

There are a number of questions. First: how accurate do estimates of the mean and the standard deviation need to be? This question is a decision problem: namely how much uncertainty is permitted for a specific dataset or network in relation to the way the data will be used. Under a network where accuracy or precision requirements are relatively low, the presented calibration method and recommended corrections may be all that is necessary. Discussion on the purposes of low-cost networks and the expectations for data accuracy is beyond the context of work presented here, but is critical for future development[22].

Second: how limiting is the assumption of sensor linearity? Sensor linearity from a number of studies is reported in a recent electrochemical sensor review[14]. Further, the linearity of the sensor response need only cover the range of the expected measurements. Semiconducting oxides generally show a nonlinear response[45] but over the range of environmental pollutant concentrations, the response is essentially linear[18,46 20] and is in any case easily corrected with a simple factory calibration. Hence, linearity of response over the required range is unlikely to be an issue. However, the effect of cross-sensitivity to other atmospheric species, which may be correlated with the pollutant of interest, could be a serious limitation. Both semiconducting oxides and electrochemical cells show significant cross-sensitivity effects, including effects of variable humidity and ambient temperature. The instrument needs to be designed to cancel these. Thus, the semiconducting oxide-based O_3 instrument used in the present work uses a gas flow-rate stepping method that cancels the effects of other atmospheric constituents[18,47]. The NO_2 electrochemical cell has, in practice in the atmosphere, a significant interference only to O_3 though effects of variation of ambient temperature and humidity can be important. However, measurement of O_3 using a semiconducting oxide instrument provides a very simple and reliable correction[36].

Third: the data that we have analyzed here was only for O_3, and so an argument may be made that these methods apply exclusively to this pollutant, which, at least on a regional scale, has a relatively smooth spatial variation. Would the semi-blind calibration technique translate onto other common urban air pollutants, such as NO_2 or particles, where the spatial variability may be significantly greater? The key is to estimate from proxy measurements the mean and standard deviation of concentrations over the averaging-time and the question is how to make these estimates for a pollutant with significant small-scale spatial variation without requiring an impractically large number of reference-standard proxies.

Powerful methods are available for estimating the spatial variation of the mean. Pollutant patterns are determined by dispersion from sources, deposition on land and buildings, and transport and reaction in the atmosphere, and a combination of diffusion tube measurement, land-use regression, chemical transport models and satellite data, has proved very effective in the estimation of the spatial variation of mean values of the major pollutants[48,49]. These methods have been extended to sub-km scale[50]. Recently we have demonstrated microscale land-use regression for NO_2, based both on diffusion tube

measurements[44] and on measurements using handheld instruments[36], and have further developed these ideas to include meteorological variables as well as urban design[36]. These studies, which are straightforward to perform at a neighborhood scale, provide a simple way of correcting a proxy's long-term mean value to account for local features as well as provide an estimate about the uncertainty associated with the correction. Furthermore, mobile devices can deliver high spatial resolution information on long-term mean concentrations[51]. Both local[51] and regional-scale[52,53] spatial patterns seem reasonably stable over time, further supporting the idea of using a proxy for estimating the mean, with appropriate correction factors.

Refining if necessary the estimate of the standard deviation of the signal that is appropriate to the site is of particular importance since the ratio of the standard deviations determines the slope estimate in (2). There are simple, empirical correlations that provide the means for this refinement. Air pollutant concentrations rather generally show a lognormal frequency distribution[39], for which it further appears that the arithmetic standard deviation is proportional to the mean. This result was noted some time ago for dust concentrations[54]. In the SI, we show data for O_3. In recent work using mobile measurement of particles and NO_2 in a long-term survey of a city, it was noted that the pollutant concentrations were lognormally distributed with the median well correlated with the arithmetic mean (provided a few extreme local 'hotspots' were removed from the correlation)[51]. For a lognormal distribution, the linear correlation of the median with the arithmetic mean implies linear correlation of the arithmetic standard deviation with the arithmetic mean. Thus, if there are sufficient proxy sites to estimate the correlation of standard deviation with mean, then correction factors for local variations of standard deviation can be derived from a land-use regression model for the local mean. In this way, one could take into account local land-use features that have a significant impact, such as traffic emissions and bus stops[36,44,55]. The uncertainties in these estimates should be straightforward to incorporate into the method we have described, deriving uncertainties in the slope and offset estimates independently and hence ultimately in the concentration.

Conclusion

The proposed data correction method is a viable, simple, remote, self-updating calibration technique for a hierarchical network of low-cost air quality sensors where a network of regulatory level instruments also operates. The proposed hierarchical network has a number of regulatory instruments that provide the top-level of proxy, and a denser network of low-cost instruments that are managed using these proxies, where their data are confirmed as reliable using the methods that we have described previously, but where their data are not corrected. This second level of the network provides a second level of proxy for the final level, of high spatial density, that has data corrected as we have described. The network of managed and reference devices would provide the necessary data for the correlation of mean and standard deviation. Such a network, supplemented if necessary with mobile devices, could be configured for an optimal tradeoff between cost, maintenance and coverage, though further work is required to determine the minimal number of regulatory-standard instruments needed and to explore methods for estimating uncertainties. Further research may look to confirm the methods in other networks, for different pollutants, with different land-use and geographical patterns and with different network configuration.

Acknowledgements

This work was supported by the NZ Ministry of Business, Innovation and Employment (contract UOAX1413) and by Callaghan Innovation, New Zealand (studentship to GM). We would like to thank Mark Bart, Bruce Ainslie and Stuart Grange who installed and operated the low-cost instrument network, and Aeroqual Ltd, NZ, for the supply of the instruments and for the data communication infrastructure. We would like to thank the many individuals and facilities who generously hosted instruments. We are also grateful to MetroVancouver staff, particularly Ken Reid and Julie Saxton, for the provision of data and access to sites. The University of British Columbia, Department of Geography provided important logistical support.

Supporting Information Available. The following files are available free of charge: miskell_blindcalibration_supportinginformation.pdf; contents: Modelling the effect of different proxies using simulated data; Site descriptions and locations for the nine co-located sensors; Overall mean and standard deviation of ozone concentration at different sites; Land-use correlations for ozone concentration in the Lower Fraser Valley; Reference

site correlation and slope correction coefficients for different land-use categories; Stepping semi-blind calibration results for different locations; Assessment of the reliability of exceedance determination using hourly data, and the effect of data truncation

References

(1) 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. The changing paradigm of air pollution monitoring. *Environ Sci Technol* **2013**, *47*, 11369-11377.

(2) Commodore, A.; Wilson, S.; Muhammad, O.; Svendsen, E.; Pearce, J. Community-based participatory research for the study of air pollution: a review of motivations, approaches, and outcomes. *Environmental monitoring and assessment* **2017**, *189*, 378.

(3) Hu, K.; Rahman, A.; Bhrugubanda, H.; Sivaraman, V. HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation From Fixed and Mobile Sensors. *IEEE Sensors Journal* **2017**, *17*, 3517-3525.

(4) Miskell, G.; Salmond, J.; Williams, D. E. Low-cost sensors and crowd-sourced data: Observations of siting impacts on a network of air-quality instruments. *Science of the Total Environment* **2017**, *575*, 1119-1129.

(5) Rai, A. C.; Kumar, P.; Pilla, F.; Skouloudis, A. N.; Di Sabatino, S.; Ratti, C.; Yasar, A.; Rickerby, D. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Science of The Total Environment* **2017**, *607*, 691-705.

(6) Cross, E. S.; Williams, L. R.; Lewis, D. K.; Magoon, G. R.; Onasch, T. B.; Kaminsky, M. L.; Worsnop, D. R.; Jayne, J. T. Use of electrochemical sensors for measurement of air pollution: Correcting interference response and validating measurements. *Atmospheric Measurement Techniques* **2017**, *10*, 3575-3588.

(7) Bart, M.; Williams, D. E.; Ainslie, B.; McKendry, I.; Salmond, J.; Grange, S. K.; Alavi-Shoshtari, M.; Steyn, D.; Henshaw, G. S. High density ozone monitoring using gas sensitive semi-conductor sensors in the Lower Fraser Valley, British Columbia. *Environmental science & technology* **2014**, *48*, 3970-3977.

(8) Brienza, S.; Galli, A.; Anastasi, G.; Bruschi, P. A low-cost sensing system for cooperative air quality monitoring in urban areas. *Sensors* **2015**, *15*, 12242-12259.

(9) Capezzuto, I.; De Vito, S.; Fattoruso, G.; Massera, E.; Buonanno, A.; Formisano, F.; Di Francia, G. An app based air quality social sensing system built on open source Hw/Sw tools. *Sensors* **2015**, 309-313.

(10) Cheadle, L.; Deanes, L.; Sadighi, K.; Gordon Casey, J.; Collier-Oxandale, A.; Hannigan, M. Quantifying Neighborhood-Scale Spatial Variations of Ozone at Open Space and Urban Sites in Boulder, Colorado Using Low-Cost Sensor Technology. *Sensors* **2017**, *17*, 2072.

(11) Heimann, I.; Bright, V.; McLeod, M.; Mead, M.; Popoola, O.; Stewart, G.; Jones, R. Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. *Atmospheric Environment* **2015**, *113*, 10-19.

(12) Mead, M.; Popoola, O.; Stewart, G.; Landshoff, P.; Calleja, M.; Hayes, M.; Baldovi, J.; McLeod, M.; Hodgson, T.; Dicks, J. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment* **2013**, *70*, 186-203.

(13) Weissert, L.; Salmond, J.; Miskell, G.; Alavi-Shoshtari, M.; Grange, S.; Henshaw, G.; Williams, D. Use of a dense monitoring network of low-cost instruments to observe local changes in the diurnal ozone cycles as marine air passes over a geographically isolated urban centre. *Science of the Total Environment* **2017**, *575*, 67-78.

(14) Baron, R.; Saffell, J. Amperometric Gas Sensors as a Low Cost Emerging Technology Platform for Air Quality Monitoring Applications: A Review. *ACS sensors* **2017**, *2*, 1553-1566.

(15) Choi, S.; Kim, N.; Cha, H.; Ha, R. Micro sensor node for air pollutant monitoring: Hardware and software issues. *Sensors* **2009**, *9*, 7970-7987.

(16) Jovašević-Stojanović, M.; Bartonova, A.; Topalović, D.; Lazović, I.; Pokrić, B.; Ristovski, Z. On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter. *Environmental Pollution* **2015**, *206*, 696-704.

(17) Magno, M.; Boyle, D.; Brunelli, D.; Popovici, E.; Benini, L. Ensuring survivability of resource-intensive sensor networks through ultra-low power overlays. *IEEE Transactions on Industrial Informatics* **2014**, *10*, 946-956.

(18) Williams, D. E.; Henshaw, G. S.; Bart, M.; Laing, G.; Wagner, J.; Naisbitt, S.; Salmond, J. A. Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network. *Measurement Science and Technology* **2013**, *24*, 065803.

(19) Piedrahita, R.; Xiang, Y.; Masson, N.; Ortega, J.; Collier, A.; Jiang, Y.; Li, K.; Dick, R.; Lv, Q.; Hannigan, M. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques* **2014**, *7*, 3325-3336.

(20) Fishbain, B.; Lerner, U.; Castell, N.; Cole-Hunter, T.; Popoola, O.; Broday, D. M.; Iñiguez, T. M.; Nieuwenhuijsen, M.; Jovasevic-Stojanovic, M.; Topalovic, D. An evaluation tool kit of air quality micro-sensing units. *Science of The Total Environment* **2017**, *575*, 639-648.

(21) Miskell, G.; Salmond, J.; Alavi-Shoshtari, M.; Bart, M.; Ainslie, B.; Grange, S.; McKendry, I. G.; Henshaw, G. S.; Williams, D. E. Data verification tools for minimizing management costs of dense air-quality monitoring networks. *Environmental science & technology* **2015**, *50*, 835-846.

(22) Lewis, A.; Edwards, P. Validate personal air-pollution sensors. *Nature* **2016**, *535*, 29 - 31.

(23) 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. Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday Discussions* **2016**, *189*, 85-103.

(24) Fonollosa, J.; Fernández, L.; Gutiérrez-Gálvez, A.; Huerta, R.; Marco, S. Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization. *Sensors and Actuators B: Chemical* **2016**, *236*, 1044-1053.

(25) Rodriguez-Lujan, I.; Fonollosa, J.; Vergara, A.; Homer, M.; Huerta, R. On the calibration of sensor arrays for pattern recognition using the minimal number of experiments. *Chemometrics and Intelligent Laboratory Systems* **2014**, *130*, 123-134.

(26) Peterson, P. J.; Aujla, A.; Grant, K. H.; Brundle, A. G.; Thompson, M. R.; Vande Hey, J.; Leigh, R. J. Practical Use of Metal Oxide Semiconductor Gas Sensors for Measuring Nitrogen Dioxide and Ozone in Urban Environments. *Sensors* **2017**, *17*, 1653.

(27) Mueller, M.; Meyer, J.; Hueglin, C. Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich. *Atmospheric Measurement Techniques* **2017**, *10*, 3783-3799.

(28) Talampas, M. C. R.; Low, K.-S. Maximum likelihood estimation of ground truth for air quality monitoring using vehicular sensor networks. *TENCON 2012-2012 IEEE Region 10 Conference, IEEE* **2012**, 1-6.

(29) Balzano, L.; Nowak, R. Blind calibration of sensor networks. *Proceedings of the 6th international conference on Information processing in sensor networks, ACM* **2007**, 79-88.

(30) Balzano, L.; Nowak, R. Blind calibration of networks of sensors: Theory and algorithms. *Networked Sensing Information and Control. Springer* **2008**, 9-37.

(31) Alavi-Shoshtari, M.; Salmond, J. A.; Giurcăneanu, C. D.; Miskell, G.; Weissert, L.; Williams, D. E. Automated data scanning for dense networks of low-cost air quality instruments: Detection and differentiation of instrumental error and local to regional scale environmental abnormalities. *Environmental Modelling & Software* **2018**, *101*, 34-50.

(32) Alavi–Shoshtari, M.; Williams, D. E.; Salmond, J. A.; Kaipio, J. P. Detection of malfunctions in sensor networks. *Environmetrics* **2013**, *24*, 227-236.

(33) Joseph, J.; Sharif, H. O.; Sunil, T.; Alamgir, H. Application of validation data for assessing spatial interpolation methods for 8-h ozone or other sparsely monitored constituents. *Environmental pollution* **2013**, *178*, 411-418.

(34) Schneider, P.; Castell, N.; Vogt, M.; Dauge, F. R.; Lahoz, W. A.; Bartonova, A. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environment international* **2017**, *106*, 234-247.

(35) Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; Thiele, L. Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing* **2015**, *16*, 268-285.

(36) Miskell, G.; Salmond, J. A.; Williams, D. E. Use of a handheld low-cost sensor to explore the effect of urban design features on local-scale spatial and temporal air quality variability. *Science of The Total Environment* **2018**, *619*, 480-490.

(37) Johnson, N.; Kotz, S.; Balakrishnan, N.: *Continuous Univariate Distributions* John Wiley & Sons: NY, 1994; Vol. 1.

(38) Sharma, S.; Sharma, P.; Khare, M.; Kwatra, S. Statistical behavior of ozone in urban environment. *Sustainable Environment Research* **2016**, *26*, 142-148.

(39) Bencala, K. E.; Seinfeld, J. H. On frequency distributions of air pollutant concentrations. *Atmospheric Environment (1967)* **1976**, *10*, 941-950.

(40) Wickham, H.: *ggplot2: elegant graphics for data analysis*; Springer, 2016.

(41) Kahle, D.; Wickham, H. ggmap: Spatial Visualization with ggplot2. *R Journal* **2013**, *5*.

(42) Station information: Lower Fraser Valley air quality monitoring network. http://www.metrovancouver.org/services/air-quality/AirQualityPublications/2012LowerFraserValleyAirQualityMonitoringReport.pdf.

(43) Miskell, G.; Salmond, J.; Grange, S.; Weissert, L.; Henshaw, G.; Williams, D. Reliable Long-Term Data from Low-Cost Gas Sensor Networks in the Environment. *Proceedings* **2017**, *1*, 400.

(44) Weissert, L.; Salmond, J.; Miskell, G.; Alavi-Shoshtari, M.; Williams, D. Development of a microscale land use regression model for predicting NO_2 concentrations at a heavy trafficked suburban area in Auckland, NZ. *Science of The Total Environment* **2018**, *619*, 112-119.

(45) Williams, D. E. Semiconducting oxides as gas-sensitive resistors. *Sensors and Actuators B: Chemical* **1999**, *57*, 1-16.

(46) Williams, D. E.; Salmond, J.; Yung, Y. F.; Akaji, J.; Wright, B.; Wilson, J.; Henshaw, G. S.; Wells, D. B.; Ding, G.; Wagner, J. Development of low-cost ozone and nitrogen dioxide measurement instruments suitable for use in an air quality monitoring network. *Sensors, 2009 IEEE* **2009**, 1099-1104.

(47) Aliwell, S.; Halsall, J.; Pratt, K.; O'Sullivan, J.; Jones, R.; Cox, R.; Utembe, S.; Hansford, G.; Williams, D. Ozone sensors based on WO3: a model for sensor drift and a measurement correction method. *Measurement Science and Technology* **2001**, *12*, 684 - 690.

(48) Knibbs, L. D.; Coorey, C. P.; Bechle, M. J.; Cowie, C. T.; Dirgawati, M.; Heyworth, J. S.; Marks, G. B.; Marshall, J. D.; Morawska, L.; Pereira, G. Independent validation of national satellite-based land-use regression models for nitrogen dioxide using passive samplers. *Environmental science & technology* **2016**, *50*, 12331-12338.

(49) Wang, M.; Sampson, P. D.; Hu, J.; Kleeman, M.; Keller, J. P.; Olives, C.; Szpiro, A. A.; Vedal, S.; Kaufman, J. D. Combining land-use regression and chemical transport modeling in a spatiotemporal geostatistical model for ozone and PM2. 5. *Environmental science & technology* **2016**, *50*, 5111-5118.

(50) Hanigan, I. C.; Williamson, G. J.; Knibbs, L. D.; Horsley, J.; Rolfe, M. I.; Cope, M.; Barnett, A. G.; Cowie, C. T.; Heyworth, J. S.; Serre, M. L. Blending multiple nitrogen dioxide data sources for neighborhood estimates of long-term exposure for health research. *Environmental science & technology* **2017**, *51*, 12473-12480.

(51) Apte, J. S.; Messier, K. P.; Gani, S.; Brauer, M.; Kirchstetter, T. W.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Vermeulen, R. C.; Hamburg, S. P. High-Resolution Air Pollution Mapping

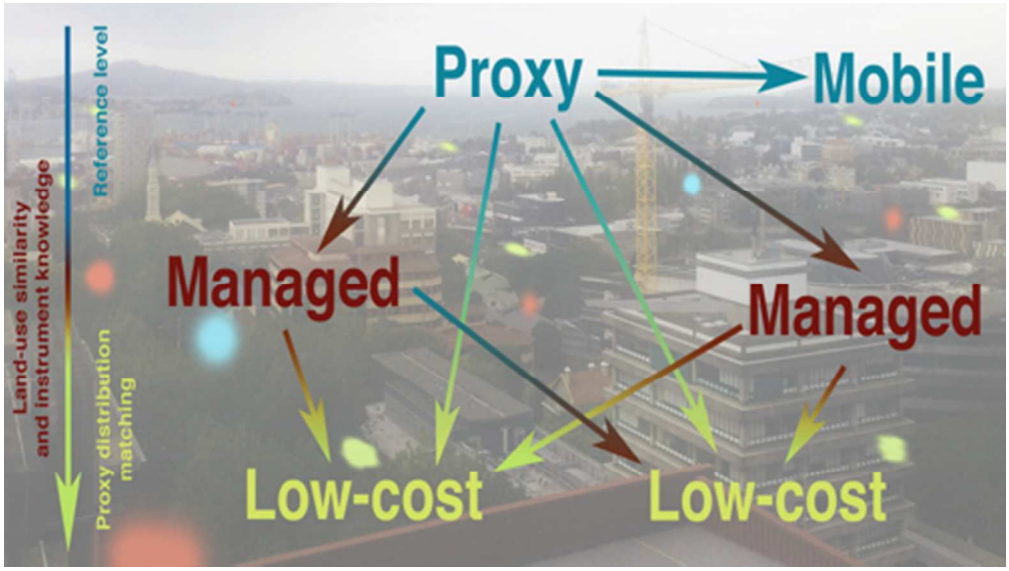
with Google Street View Cars: Exploiting Big Data. *Environmental Science & Technology* **2017**, 6999 - 7008.

(52) Tong, N. Y.; Leung, D. Y. Ozone diurnal characteristics in areas with different urbanisations. *International Journal of Environment and Pollution* **2012**, *49*, 100-124.

(53) Ito, K.; Thurston, G. D.; Nádas, A.; Lippmann, M. Monitor-to-monitor temporal correlation of air pollution and weather variables in the North-Central US. *Journal of Exposure Science and Environmental Epidemiology* **2001**, *11*, 21-32.

(54) Oldham, P. The nature of the variability of dust concentrations at the coal face. *British journal of industrial medicine* **1953**, *10*, 227 - 234.

(55) Miskell, G.; Salmond, J.; Longley, I.; Dirks, K. N. A novel approach in quantifying the effect of urban design features on local-scale air pollution in central urban areas. *Environmental science & technology* **2015**, *49*, 9004-9011.



TOC image

84x47mm (150 x 150 DPI)