Environmental Science & Technology

Article

pubs.acs.org/est

Data Verification Tools for Minimizing Management Costs of Dense Air-Quality Monitoring Networks

Georgia Miskell,[†,‡] Jennifer Salmond,[†] Maryam Alavi-Shoshtari,[‡,§] Mark Bart,[∥] Bruce Ainslie,[⊥] Stuart Grange,[†,‡] Ian G. McKendry,[#] Geoff S. Henshaw,[∇] and David E. Williams*[,‡]

[†]School of Environment, Faculty of Science, University of Auckland, Auckland 1142, New Zealand

[‡]MacDiarmid Institute for Advanced Materials and Nanotechnology, School of Chemical Sciences, Faculty of Science, University of Auckland, Auckland 1142, New Zealand

[§]Department of Mathematics, Faculty of Science, University of Auckland 1142, Auckland, New Zealand

[∥]Air Quality Ltd, 40A George Street, Mt Eden, Auckland 1024, New Zealand

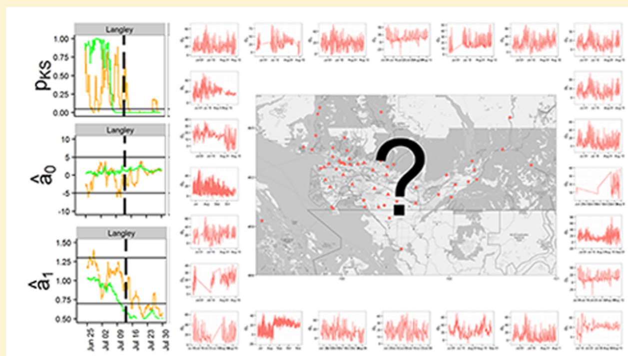
[⊥]Meteorological Services of Canada, Environment Canada, Vancouver V1V 1V7, Canada

[#]Department of Geography, The University of British Columbia, Vancouver V6T 1ZU, Canada

[∇]Aeroqual Ltd, 109 Valley Road, Mt Eden, Auckland 1024, New Zealand

Supporting Information

ABSTRACT: Aiming at minimizing the costs, both of capital expenditure and maintenance, of an extensive air-quality measurement network, we present simple statistical methods that do not require extensive training data sets for automated real-time verification of the reliability of data delivered by a spatially dense hybrid network of both low-cost and reference ozone measurement instruments. Ozone is a pollutant that has a relatively smooth spatial spread over a large scale although there can be significant small-scale variations. We take advantage of these characteristics and demonstrate detection of instrument calibration drift within a few days using a rolling 72 h comparison of hourly averaged data from the test instrument with that from suitably defined proxies. We define the required characteristics of the proxy measurements by working from a definition of the network purpose and specification, in this case reliable determination of the proportion of hourly averaged ozone measurements that are above a threshold in any given day, and detection of calibration drift of greater than ±30% in slope or ±5 parts-per-billion in offset. By analyzing results of a study of an extensive deployment of low-cost instruments in the Lower Fraser Valley, we demonstrate that proxies can be established using land-use criteria and that simple statistical comparisons can identify low-cost instruments that are not stable and therefore need replacing. We propose that a minimal set of compliant reference instruments can be used to verify the reliability of data from a much more extensive network of low-cost devices.



INTRODUCTION

This paper addresses the confidence with which data delivered by a network of low-cost air quality measurement instruments can be used to represent reliably local pollutant concentrations, and presents ideas on the trade-off of minimum-cost design for reliable high-density networks. Traditionally, financial and logistical constraints have meant that air quality scientists and managers have had to choose between very accurate high temporal resolution measurements made at a limited number of sites and low resolution low accuracy measurements made at a much larger number of sites.[1] This makes it difficult to resolve accurately the complex patterns in urban air quality in time and space, limiting the ability to identify, understand, predict, and mitigate air pollution episodes.[2] However, recent developments of low-cost, easy to use sensors, often portable with minimal power and environmental housing requirements, together with advances in data management, processing, and communications has made it financially and logistically conceivable to operate a spatially dense network of monitors with high temporal resolution.[3−5] Such networks can have the potential to resolve the complex spatial, and temporal heterogeneity of air pollution concentrations in urban centers in near-real time[6,7] and would make it possible to answer new questions about the underlying causes of poor air quality (ensuring more accurate modeling and prediction at local scales), improve the ability to identify

Received: September 17, 2015
Revised: December 9, 2015
Accepted: December 10, 2015

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

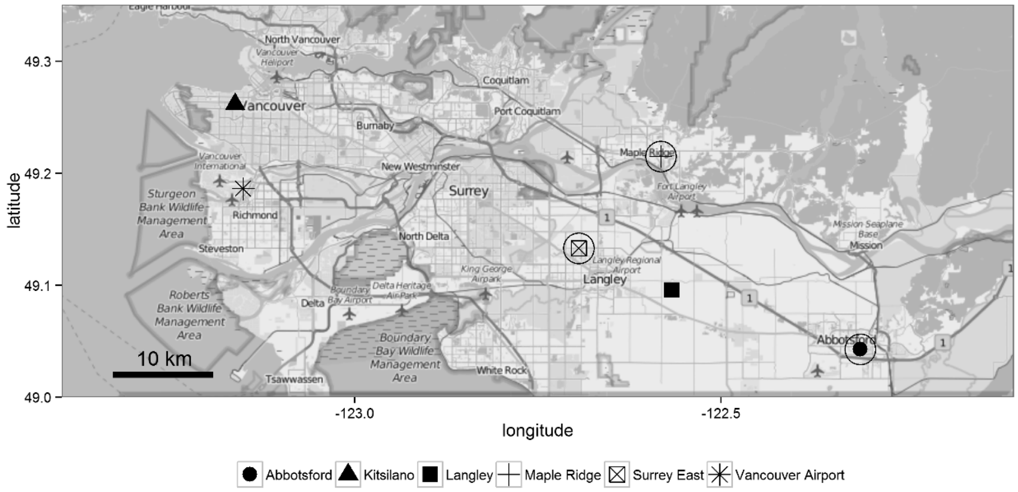


Figure 1. Location of the reference sites discussed in the text. Information on land-use can be found in ref 36. Circled locations are the sites used for method performance, with others serving as proxies.[41] The map was created using using the R package "ggplot" and "ggmap".[48]

the links between air quality and human health (or environmental degradation), identify potential air pollution "hot spots" and enhance the ability to quantify the impacts of pollutant mitigation techniques.[5,8] These advances have been hailed as a new paradigm in air quality monitoring, and welcomed as the way forward by both air quality managers and scientists alike.[2] However, the need for increased temporal and spatial resolution of air quality monitoring data has made it very tempting to proceed with such technology before its performance has been thoroughly assessed.[5]

One of the biggest challenges facing the successful adoption of dense networks of low-cost instrumentation is the ability to determine the performance of these instruments in urban environments and the ongoing quality of the data during deployment, in real-time and at minimal cost.[9,10] To be effective, users must have confidence that the data from any instrument (including low-cost instruments) must be sufficiently accurate and precise to meet the design criteria of the network, and the limitations of the instruments need to be clearly stated.[1,11] To date there have been very few published studies describing large-scale deployments of low-cost instruments[9,10,12] which also provide information or methodologies for assessing the reliability of the data sets provided from these instruments. Conventionally, confidence in the data from an instrument is ensured by a regular program of calibration traceable to reference standards.[13−15] However, as the number of devices in a network grows, the costs associated with such a program of calibration and maintenance can become very large. For low-cost instruments it is therefore useful to introduce a new measure, that of instrument "reliability", in order to describe instrument performance.[8,16,17] This measure is less restrictive than compliance, but it does require clarity in its definition such that users have confidence in the data within known and defined constraints.

Reliability can be assessed using temporary or permanent colocation of one or more instruments. There have been studies using low-cost devices where the random colocation of devices allows spot-checking of one against another.[14] However, there remains the possibility that unusual trends in a monitor which is not currently colocated against another may be misinterpreted erroneously as either fluctuations in environmental processes or as instrument error. Reliability can also be assessed using computational techniques to detect and compensate for changes from an expected pattern,[18] and for specifically defined instrument conditions.[19,20] These methods generally assume either a model for the phenomenon being sensed, typically exploiting correlations across the network, or a model for the behavior of the sensor within the instrument. Simple multivariate time series, principal component analysis, or "soft sensor" methods that we[21] and others[13,22−25] have used before require a long time run of data to establish the model from which drifts and malfunctions can be detected. General limitations of such approaches include the accuracy and reliability of models (and the data they are built upon) and their stability over time, which all methods suffer from to some degree.

Here, we develop a simple framework within which to address these issues, from consideration of data delivered by a three-month installation of a network of low-cost ozone monitoring instruments deployed over the Lower Fraser Valley (LFV), British Columbia, including the Vancouver urban area (Figure 1). We develop the framework so that general qualitative knowledge can be used, both about the spatiotemporal behavior of the pollutant and about the performance characteristics of the measurement system, which encompasses the instrument and its sampling system and the location characteristics. To do this, we reflect on the stated purpose of a dense low-cost network. Following Snyder et al.[5] we specify this network as supplementing a compliant ambient air monitoring network, extending coverage and providing reliable information for communities, including improved local coverage for exposure assessment and enhancing source compliance monitoring. For these purposes, reliable determination of the short-term statistics of the ozone concentration is required: for example, the proportion of hourly average values that are above a threshold. Here, the instrument response time can be assumed fast enough that each measurement is not affected by the preceding measurement. So although ozone generally follows a smoothly varying time series, for verifying instrument performance against the stated network purpose we can also treat it as a probability distribution, ignoring time variation. We argue that reliability can be assessed using an appropriate proxy, which is not a prediction of the air pollution concentration field, but instead is a model for the statistics of the pollution field. Simple statistical tests can then be used to answer the binary question: "does the instrument need maintenance or

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

not?" We emphasize that, although we base our proxy on the statistics of the pollution field, the proxy specifically is not a model of pollutant concentrations in time or space. Based on the proxy comparison, we reduce "reliability" to a binary decision: the data are considered reliable, or they are not and some action is required. Action can include further inspection, to determine whether it is the instrument or its local environment that has changed, or whether recalibration or replacement of the instrument is required. If the data are considered reliable then they are taken as representative of a "ground truth" measurement of the concentration field. Thus, this approach is different in principle from ideas of "blind" or "semiblind" calibration which assume that a set of devices is measuring a common field and then attempt to adjust individual calibrations in order to achieve consistency of the estimated field.[26] It is also different from ideas which utilize a large number of low-accuracy instruments to generate an approximate estimate of a field or to detect and locate an unusual perturbation.[27] The aim here is to make minimal or no assumptions about the behavior of the device and as few assumptions as possible about the phenomenon being measured, consistent with the need to establish confidence in the data. Prior knowledge, such as the known general dependence of air quality on land-use or location or weather patterns, is incorporated in our approach through the appropriate choice of proxies. This approach can be interpreted as assessment of the conditional probability distribution of the data from any given instrument given by the proxy. Formally, this would require the modeling of conditional distributions in a Bayesian framework. Bayesian analysis has been widely used in fault detection algorithms of sensor networks[27,28] as well as air quality models.[29] However, full Bayesian analysis focuses on modeling the posterior distribution (here, the conditional probability of data given by the proxy) by training the model and by hierarchical learning from the historic data,[27,29] making assumptions about the statistical characteristics of the data (for example, the covariance structure of the data[29]) or data discretization.[28] This approach generally requires large training sets of data. Although the aim of this paper can be interpreted as a Bayesian approach, in contrast to a full Bayesian analysis we aim to make minimal assumptions about data characteristics, to avoid the need for assessment of conditional probability distributions based on large data sets accumulated over long periods of time, and to develop a method that can be applied essentially in real time.

In this paper, we first formally define the required characteristics of a proxy in relation to the statistical tests to be used to determine whether, by comparison with the proxy, data from any given instrument are considered to be reliable. Then we develop the idea that suitable proxies can be established using land-use and location information. We compare data with a proxy using an approach qualitatively similar to the use of control charts in industrial process control.[30,31] We first test the speed and reliability with which the methods detect instrument failures by using data from low-cost instruments colocated with reference stations. Then we extend the evaluation using reference stations with which instruments are not colocated, comparing the results with those for colocation. Finally, we evaluate using only the network of low-cost instruments. In this way, we demonstrate the possibility of having a minimal set of compliant reference instruments, along with data from a much more extensive network of low-cost devices, in order to capture an area's air quality with the best outcome on trade-off between coverage, cost and reliability.

■ MATERIALS AND METHODS

1. Data. The data analyzed were from a spatially dense network of low-cost ozone instruments using gas-sensitive semiconducting oxide sensor (GSS) technology deployed in the Lower Fraser Valley (LFV), British Columbia, from May-November 2012. The LFV has a well-established and well-maintained network of reference stations against which the performance of the GSS instruments was assessed.[9] The GSS network in total involved 50 different sites and 58 different instruments, a subset of which was used for the present paper. In the previous work, we established acceptance criteria for the data based on qualitative inspection of the derived ozone data and on the variation of the raw resistance data from the GSS instruments utilizing knowledge of the effects on the signal of known failure modes of the devices (microstructure change in the sensor element; air flow rate fluctuations and failures of the air intake pump; variations in the sensor temperature control; deposition of dirt causing ozone decomposition in the inlet filter, in the sampling lines and on the internal surfaces of the sensor) as described in detail in Williams et al.[32,33] Some examples are given in the SI. The determination of failure (or not) by these methods is referred to in the present work as "single sensor assessment". Approximately 50% of the devices deployed were recorded in this way as "failing" at some time during deployment: some from sensor failures and many from the effects of dirt and insects either in the air inlet or deposited onto the sensor. Nine GSS instruments were colocated with reference stations and deployed for long periods (locations in the SI). These data were used in the present work in order to compare the colocated reference results with failures recorded with respect to different choices of proxy. The open-source statistical software package 'R' was used for all calculations (version 3.2.2).

2. Formal Definition of a "Proxy" and Definition of "Failure" of an Instrument, Evaluated with Respect to the "Proxy". Let $Y_k(t)$ denote the ozone concentration measured by the instrument and $X_k(t)$ be the true ozone concentration at the instrument inlet, at location k and time t. Formally, we can consider that $Y_k(t)$ is a predictor of $X_k(t)$, or we can consider that the conditional probability distribution of Y_k given X_k is stable and time-invariant; in other words, although the ozone concentration (actual or measured) is highly time varying, the measured process conditional on the underlying actual process is stationary, provided that the data Y_k are measuring reliably. Of course, the reliability of $Y_k(t)$ cannot be assessed directly, since $X_k(t)$ is by definition unknown. Hence we seek some proxy, $Z_k(t)$ against which $Y_k(t)$ can be assessed. As noted in the Introduction, we've defined the purpose of the network as supplementing a compliant ambient air monitoring network, specifically to provide reliable local determination of the short-term statistics of the ozone concentration. Hence first we develop a proxy comparison based on these short-term statistics.

a. Assessment Based on Running Probability Distribution of Data. Let $F_{Xk}(x;\ t_1,...,t_1 + t_d)$ denote the empirical cumulative distribution (ECD: see SI) of $X_k(t)$ obtained over a time index t_d (an integer that counts the time steps of measurement) running over the data stepping by one time step. We define the purpose of the network to be to deliver a reliable local estimate of $F_{Xk}(x;\ t_1,...,t_1 + t_d)$. The data, $Y_k(t)$ and the

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

proxy, $Z_k(t)$ provide two different estimates, $F_{Yk}(x;\, t_1,\ldots,t_1 + t_d)$ and $F_{Zk}(x;\, t_1,\ldots,t_1 + t_d)$, of this distribution, and we compute the running two-sample Kolmogorov–Smirnov (K–S) probability with which the two estimates reflect the same underlying distribution. The number of time steps, t_d, used for the determination of the distribution is empirically chosen to obtain a reasonably representative estimate of the variation of the ozone concentration that is sufficiently large that the distribution is well estimated and missing values have a small effect, but sufficiently small that action can be taken in a reasonable time depending on the result of the test. In the present work, we have used hourly averaged data for $Y_k(t)$ and $Z_k(t)$, and so the time step is 1 h.

Let $p_{KS}{}^*(Z_k,Y_k)$ denote the critical value for the K–S statistic. If $p_{KS}(Z_k,Y_k) > p_{KS}{}^*(Z_k,Y_k)$ then $F_{Y_k}(x;,\, t_1,\ldots,t_1 + t_d)$ and $F_{Zk}(x;\, t_1,\ldots,t_1 + t_d)$ can be considered to be estimates of the same distribution. In this case, the instrument is then considered to be functioning as expected (termed "intact") and we take $X_k(t) = Y_k(t)$ over the interval $(t_1,\ldots, t_1 + t_d)$. However, if $p_{KS} \le p_{KS}{}^*$, then an alarm is signaled. Either the instrument is not intact (a "true alarm") or the local environment has changed either with respect to the proxy or to the measurement location, or the proxy is not suitable for ozone concentration variation at the measurement location (a "false alarm"). A criterion has to be defined to discriminate in a practically acceptable way between these two alternatives. Site-specific, weather-specific, and large-scale event-specific phenomena can be expected when analyzing air quality time-series, and so periodic alarms may occur even for intact measurements, where the proxy signal is not a suitable estimator for $F_{Xk}(x;\, t_1,\ldots,t_1 + t_d)$. For such variations, if the low-cost instrument were intact, then while an alarm might be signaled, after some time p_{KS} could be expected to once more increase above the threshold. However, for instrument failure—a drift or sustained change in calibration parameters—then a clear and sustained pattern of change in $F_{Yk}(x;\, t_1,\ldots,t_1 + t_d)$ with respect to $F_{Zk}(x;\, t_1,\ldots,t_1 + t_d)$ would be observed and hence a sustained pattern of change in p_{KS}. We therefore define:

$$X_k(t) \neq Y_k(t) \text{ if } p_{KS} < p_{KS}^* \text{ for } t > t_f \tag{1}$$

where the time index, t_f, is considered sufficient for confidence that the two distributions remain different, and is empirically determined. The threshold probability, $p_{KS}{}^*(Z_k,Y_k)$ and the averaging time, t_d, are also to be specified.

b. Assessment Based on Mean and Variance. Instrument calibration during manufacture or installation establishes $Y_k(t)$ as a linear predictor of $X_k(t)$:

$$X_k(t) = a_0 + a_1 Y_k(t) + e_{Y,k}(t) \tag{2}$$

where, immediately following the calibration within some acceptable specification and with some certainty, the offset, $a_0 \cong 0$ and the slope, $a_1 \cong 1$, and the error term, $e_{Y,k}$ is a white-noise process with no nonlinear ozone concentration dependence, mean = 0 and variance that satisfies some specification on the required accuracy of the measurement. The assessment of whether the data are reliable then becomes an assessment of whether any drift in the parameters a_0, a_1, and $e_{Y,k}$ remains within bounds defined by the network specification. We suggest that the absence of long-term drifts in a_0, a_1, and $e_{Y,k}$ can be detected by choosing a proxy whose mean $\mu(Z_k(t_1,\ldots,t_1 + t_d))$ and variance $\mathrm{var.}(Z_k(t_1,\ldots,t_1 + t_d))$ evaluated over the interval $(t_1, \ldots, t_1+t_d)$ satisfy:

$$\mu(X_k(t_1, ..., t_1 + t_d)) \cong b_0 + b_1\mu(Z_k(t_1, ..., t_1 + t_d)) \tag{3}$$

$$\mathrm{var}(X_k(t_1, ..., t_1 + t_d)) \cong b_1^2\mathrm{var}(Z_k(t_1, ..., t_1 + t_d)) + \mathrm{var}(e_{Z,k}(t_1, ..., t_1 + t_d)) \tag{4}$$

where a good proxy implies $\mathrm{var.}(e_{Z,k}) \ll \mathrm{var.}(Z_k)$ and the parameters b_0, b_1, and $\mathrm{var.}(e_{Z,k})$ at most fluctuate within defined bounds over the observation period. An example of such a proxy would be a signal that was a linear transformation of $X_k(t)$ with some transformation of the time scale such as a phase delay. From eq 2, $\mu(X_k) = a_0 + a_1\mu(Y_k)$ and $\mathrm{var.}(X_k) = a_1^2\mathrm{var.}(Y_k) + \mathrm{var.}(e_k)$ so given the definition of the proxy (eqs 3 and 4) we define estimators for the slope $\hat{a}_1$ and offset $\hat{a}_0$ as

$$\hat{a}_1 = \left(\frac{\mathrm{var}(Z_k(t_1, ..., t_1 + t_d))}{\mathrm{var}(Y_k(t_1, ..., t_1 + t_d))}\right)^{1/2} \cong \frac{a_1}{b_1}\left(1 - \frac{1}{2b_1^2}\frac{\mathrm{var}(e_{Z,k}(t_1, ..., t_1 + t_d))}{\mathrm{var}(Z_k(t_1, ..., t_1 + t_d))}\right) \cong \frac{a_1}{b_1} \tag{5}$$

$$\begin{aligned}\hat{a}_0 &= \mu(Y_k(t_1, ..., t_1 + t_d)) - \mu(Z_k(t_1, ..., t_1 + t_d))\left(\frac{\mathrm{var}(Z_k(t_1, ..., t_1 + t_d))}{\mathrm{var}(Y_k(t_1, ..., t_1 + t_d))}\right)^{1/2} \\ &= \frac{a_0 - b_0}{b_1} + \frac{1}{2b_1^2}\frac{\mathrm{var}(e_{Z,k}(t_1, ..., t_1 + t_d))}{\mathrm{var}(Z_k(t_1, ..., t_1 + t_d))}\left(\frac{b_0 - a_0}{b_1} + \mu(Z_k(t_1, ..., t_1 + t_d))\right)\end{aligned} \tag{6}$$

where we have assumed that in normal operation the instrument variance is much less than the proxy variance. We track the variation of $\hat{a}_0$ and $\hat{a}_1$, determined over time, t_d, where an alarm is signaled if these quantities move out of defined bounds, which are now specified based on an acceptable instrument specification for the error in a_0 and a_1. As discussed above, the assumption is that the parameters associated with the proxy will in general remain bounded and not systematically drift. Hence if $\hat{a}_1$ and $\hat{a}_0$ remain outside the bounds for time $t > t_f$ then either $X_k(t) \neq Y_k(t)$ (i.e., $a_0 \neq 0$ and/or $a_1 \neq 1$ within the bounds of the instrument specification), or $Z_k(t)$ has ceased to be a good proxy.

3. Choices for the "Proxy", $Z_k(t)$. One approach to obtaining $Z(t)$ is to develop a spatiotemporal model for the ozone field. Here we develop a different and simpler idea. A type of remote proxy is required in order to evaluate each sensor site and so variables with associations with ozone were assessed. If a general variable could assist in explaining ozone,

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX–XXX

Table 1. MetroVancouver Reference Monitoring Station Characteristics for the Co-Located Sites and the Selected Representative Land-Use Sites[a]

		co-located reference monitoring stations									other reference stations	
station		Abbotsford	Langley	Maple Ridge	North Delta	Pitt Meadows	Port Moody	Richmond South	Second Narrows	Surrey East	Hope Air	Richmond Air
station number		T33	T27	T30	T13	T20	T9	T17	T6	T15	T29	T31
latitude (N)		49.043	49.096	49.215	49.158	49.245	49.281	49.141	49.302	49.133	49.370	49.186
longitude (W)		122.310	122.567	122.582	122.902	122.709	122.849	123.108	123.020	122.694	121.499	123.152
elevation (m amsl)		58	88	50	113	4	6	4	4	79	40	1
population (within 1 km)		8894	1089	8501	11229	203	2882	11247	1125	8534	204	850
land-use in this study		"res"	"res"	"res"	"res"	"agr"	"com"	"res"	"com"	"res"	"rur"	"com"
land-use (% within 1 km)	agr.	NA	30	5	0	85	0	10	0	2.5	NA	0
	com.	NA	0	10	10	13	12.5	0	40	0	NA	85
	res.	NA	60	60	75	0	12.5	75	10	90	NA	0
	other	NA	10	15	15	2	75	15	50	7.5	NA	15

[a]"agr.", agricultural; "com.", commercial; "res.", residential; "rur.", rural. The "other" land-use category is for surroundings such as water and park.

then this could be used to guide proxy selection. Ozone concentration is primarily driven by sunlight, motor vehicle traffic density and location with respect to prevailing wind direction over the urban center.[34,35] Because ozone concentration is usually low at night and has a diurnal variation reflecting sunlight intensity, the daily mean value also reflects the daily range of values. Land-use regression (LUR) is a statistical technique that has traditionally been used to estimate local pollutant concentrations using surrounding land-use, traffic characteristics, and observed concentrations. Typically in these models, high-resolution land-use data is used as a proxy for estimating the strength of local emission sources. As such, LUR techniques have been used to estimate the spatial distribution of primary pollutants[36] and in the context of the LFV for ultrafine particles,[37] oxides of nitrogen and fine particulate matter[38] and wood smoke.[39] Because ozone is a secondary pollutant, whose peak daytime concentration can be dependent on nonlocal emissions, meteorology (including wind speed, direction, mixed layer depth), and day-to-day hand over processes, LURs have traditionally not been used in ozone exposure assessments. While Kerchoffs et al.[40] used LUR for their ozone exposure assessment over The Netherlands, their analysis was used to estimate the spatial distribution of long-term (two-week average) concentrations, and was not able to resolve the high temporal variability seen in urban ozone concentrations. Nevertheless, their analysis showed that long-term mean ozone concentrations are correlated to site descriptions, such as land-use and location[36] and in the present analysis, we exploit this association to develop improved proxies for the statistics for short-term (e.g., hourly resolved) ozone concentrations within the LFV. Specifically, we propose that a proxy based on similar land-use to the location to be assessed could satisfy on average both the condition of a stable conditional distribution required by method (a) above and the condition of a stable relationship of mean and variance required by method (b). Thus, we evaluate the choice of $Z_k(t)$ as a reference station signal or a network median having similar land-use to the location, k. Furthermore, simple general land-use descriptors appeared to be sufficient: ozone measurements from different instruments in the same general land-use appeared to be highly correlated and those in different land-uses to be distinct (SI).[9] Therefore, we explored the use of simple land-use descriptors to establish the proxies.

Site descriptions for a regression analysis were based on a detailed report of the local air quality monitoring station characteristics[41] and had variables inlet height, elevation, land-use (three buffers), population density (two buffers), and orientation toward the downtown Robson Square station, which is situated close to the center of the region's precursor emissions footprint. Ozone data was from 20 MetroVancouver sites over the three-month GSS instrument deployment, with both daily (91 days × 20 sites, n = 1840) and total (n = 20) concentrations considered. The variable with the highest R^2 for both daily and total was land-use category within 1 km radius of each station (0.18 and 0.62 respectively). Addition of other possible variables did not improve the regression, which may be due to small variability in values (e.g., inlet height mostly around 4 m), multicollinearity issues (e.g., population density Variance Inflation Factor > 3), or uneven distribution of categories (e.g., 12 sites were South-East or East of Robson Square). The layout of the LFV could be confounding as highly urbanized/commercial areas were close to the sea, and residential areas were predominantly inland and downwind from the prevailing wind. From this, four broad land-use categories were derived from the MetroVancouver report to represent the locations of the GSS instruments at a 1 km buffer (agricultural, commercial, residential, and rural). The category "rural" is not an official category; however, as the network objective was to extend the current scale, many GSS instruments were located on the outskirts of the LFV. To establish the appropriate category for the GSS instrument locations which did not have land-use designations, aerial imagery was compared with aerial imagery against the selected reference sites (Table 1). Four different proxies were evaluated using hourly averaged data:

(1) The colocated reference data: this should give the most reliable indicator of instrument performance, against which the other proxies and the performance of the data assessment methods can be evaluated;

(2) A nonco-located reference station in the same land-use classification as the instrument under assessment;

(3) A nonco-located reference station in a different land-use classification;

(4) The hourly median value of all intact low-cost instruments in the same land-use classification as the instrument under assessment.

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

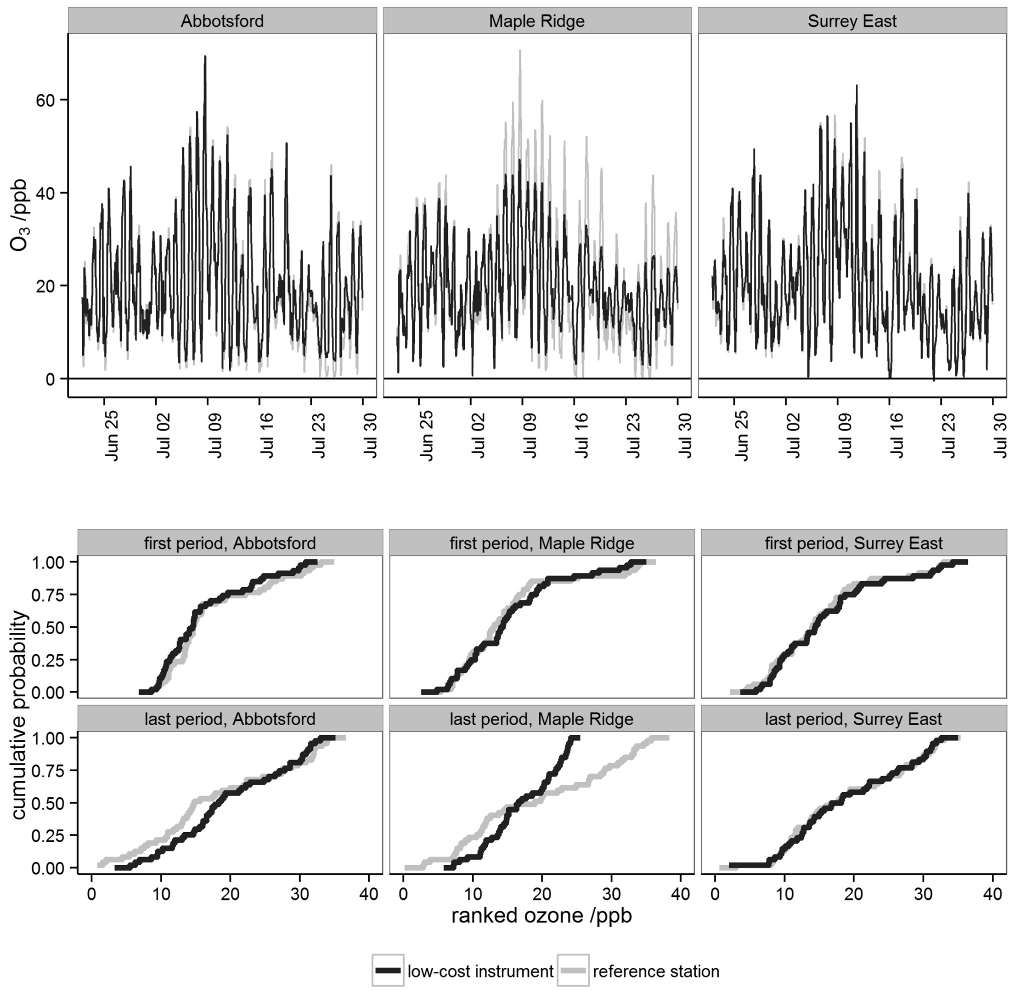


Figure 2. Top: Time-series for the three selected sites discussed in the text where low-cost GSS instruments were colocated with reference stations. Bottom: Comparison of the 72 h empirical cumulative probability distribution of ozone measurement from the reference instruments and the colocated GSS instruments at the beginning and end of the assessment period (first period: 29 June to 1 July 2012; last period: 28 to 30 July 2012).

For each proxy $Z_k(t)$ control charts of p_{KS} and of $\hat{a}_1$ and $\hat{a}_0$ were constructed. A sensitivity analysis was carried out to establish the values for the averaging time index, t_d, where t_d = 72 h was selected (SI). The failure time, t_f, was evaluated by consideration of the fluctuations associated with different proxies and is discussed later, with t_f = 120 h selected. The selected exceedance thresholds that were required to be defined for sensor drift detection in (a) and (b) were p_{KS}^* = 0.05, $\hat{a}_1$ 1 ± 0.3 and $\hat{a}_0$ 0 ± 5 ppb based on the maximum variability for transfer standard accuracy ($\hat{a}_0$) and for indicative measurement ($\hat{a}_1$) suggested by the U.S. Environmental Protection Agency.[16,42] The results for data acceptance were compared with the single sensor assessment.

■ RESULTS

Reference Station As "Proxy". Inspection of the colocated GSS instrument time-series showed all the GSS instruments having excellent agreement with the reference data at the beginning of the monitoring period (SI). Four maintained this agreement throughout and five showed a change relative to the reference data. Three colocated sites (Abbotsford, "abb"; Maple Ridge, "map"; and Surrey East, "sur") were selected to illustrate relatively good, poor, and excellent performance of the GSS instruments (Figures 1 and 2). Figure 3 presents the respective control charts for p_{KS}, $\hat{a}_1$ and $\hat{a}_0$ using the three different reference station proxies. For Surrey, a stable GSS device is indicated by the comparison with the local reference station and by the comparison with the remote proxy in the same land-use. However, the remote proxy in different land-use showed oscillations in the parameters that crossed the thresholds, particularly for the time around July 9th when forest fires in Siberia caused elevated ozone over Vancouver.[43] At Maple Ridge there was a clear change in GSS behavior, consistent with a change in a_1 which was signaled by all three proxies, and which was detected at a date earlier than that indicated by the single sensor assessment. For Abbotsford, a subtle drift was flagged using the local reference station proxy but was not flagged by the remote reference station proxies. The single sensor assessment also did not detect the drift. Close inspection of the Abbotsford time series (Figure 2) shows the apparent drift to be caused by the GSS instrument late in the month reporting lower nocturnal ozone readings than those recorded by the local reference instrument. Table 2 compares the failure time determined by each of the three measures at each of the nine reference locations along with the failure time determined by single sensor assessment.

Network Median As "Proxy". Figure 4 shows p_{KS}, $\hat{a}_1$ and $\hat{a}_0$ for four different residential reference stations as the signal $Y_k(t)$ against the median signal for that part of the GSS network in the residential land-use (n = 18 instruments, locations in S.I.)

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

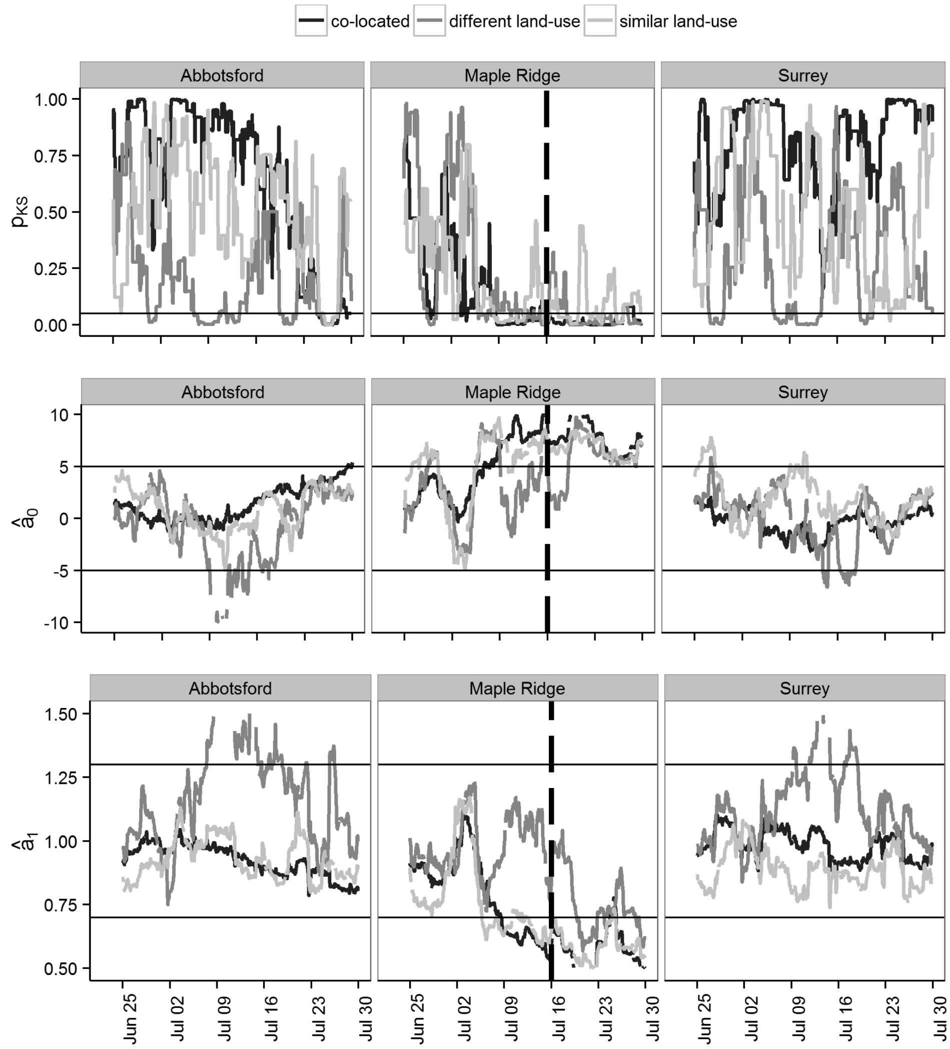


Figure 3. Control charts for the three selected sites using the proxies: colocated reference station, remote reference station with a similar surrounding land-use (residential: Langley), and remote reference station with a different surrounding land-use (commercial: Vancouver Airport). Horizontal lines represent designated thresholds for p_{KS}, $\hat{a}_0$, and $\hat{a}_1$ and the vertical dashed line for Maple Ridge is where single sensor assessment identified sensor drift.

as a proxy, $Z_k(t)$. Some of the GSS devices failed during the period, and were removed from the median calculation when that happened. This assessment both tests the variability of signals within a given land-use classification and the usefulness of the network median as a proxy. The parameters for the different stations showed oscillations, which exceeded the thresholds but not for a sustained time, most notably around the time of the Siberian fire impacts, which was also a period of high ozone concentrations.

An alternative way to evaluate "failure" would be simply to count the fraction of the total number of estimators at any given time and take this as an empirical probability of failure, p_f (reliable data is then $1 - p_f$). The residential GSS instruments and those colocated at reference stations were evaluated using this technique for the month of July (Figure 6). For this evaluation, the single sensor assessment method has been included as a measure, and the failure data for this method is marked on the plot using a large symbol. Clear failures were signaled by a rapid fall with time in $(1 - p_f)$ where all measures indicated failure within less than 1 week. This approach resolved some of the inconsistencies in Table 2 and so seemed more robust than simply taking a single measure. For some devices, however, the fall with time in $(1 - p_f)$ was rather slower and did not immediately correspond with indication of failure by the single sensor assessment method. This may indicate either a slow and subtle drift in the device characteristics or a slow change in the environmental conditions at the specific locality with respect to the proxy.

■ DISCUSSION

Three major questions are investigated in this paper: can the use of simple statistical comparisons with proxies provide a reliable framework for continuous, automated, quantitative assessment of data reliability from dense networks of low-cost ozone measurement instruments; can such comparisons work reliably without extensive training data and preferably with a rolling time window that is relatively short; and can suitable locations for reference stations to provide proxy data for a high-density network of low-cost instruments be specified based on land-use criteria? Our results show that instrument failure can be identified reliably using simple statistical comparisons between low-cost instruments and suitably chosen proxies, specifically through the use of a remote reference station having the same land-use as the low-cost network instruments. The

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Table 2. Drifting Dates for the Nine Co-Located GSS Instruments[a]

verification method	verification threshold	Abbotsford	Langley	Maple Ridge	North Delta	Pitt Meadows	Port Moody	Richmond	Second Narrows	Surrey East
		sites								
single sensor	S.S.A.	OK	13-Jul	16-Jul	9-Jul	19-Jul	17-Jul	26-Jul	OK	OK
co-locate	$\hat{a}_0$	OK	OK	12-Jul	5-Jul	24-Jul	20-Jul	OK	OK	OK
	$\hat{a}_1$	OK	15-Jul	13-Jul	6-Jul	22-Jul	21-Jul	OK	OK	OK
	p_{KS}	OK	12-Jul	13-Jul	18-Jul	12-Jul	24-Jul	27-Jul	26-Jul	OK
different land-use	$\hat{a}_0$	OK	OK	24-Jul	8-Jul	13-Jul	23-Jul	OK	14-Jul	OK
	$\hat{a}_1$	12-Jul	24-Jul	OK	8-Jul	OK	31-Jul	OK	OK	OK
	p_{KS}	OK	20-Jul	23-Jul	18-Jul	28-Jul	5-Jul	27-Jul	5-Jul	OK
similar land-use	$\hat{a}_0$	OK	OK	15-Jul	1-Jul	NA	NA	OK	NA	OK
	$\hat{a}_1$	OK	15-Jul	17-Jul	1-Jul	NA	NA	OK	NA	OK
	p_{KS}	OK	12-Jul	OK	8-Jul	NA	NA	OK	NA	OK
network	$\hat{a}_0$	OK	OK	OK	8-Jul	NA	NA	OK	NA	OK
	$\hat{a}_1$	OK	OK	OK	8-Jul	NA	NA	OK	NA	OK
	p_{KS}	OK	19-Jul	18-Jul	13-Jul	NA	NA	27-Jul	NA	OK

[a]OK is where instruments were labelled as intact and NA is where instruments were not in a residential land-use. Verification thresholds are for single sensor assessment (S.S.A.), K−S probability (p_{KS}), and assessments using the mean and variance ($\hat{a}_0$, $\hat{a}_1$).

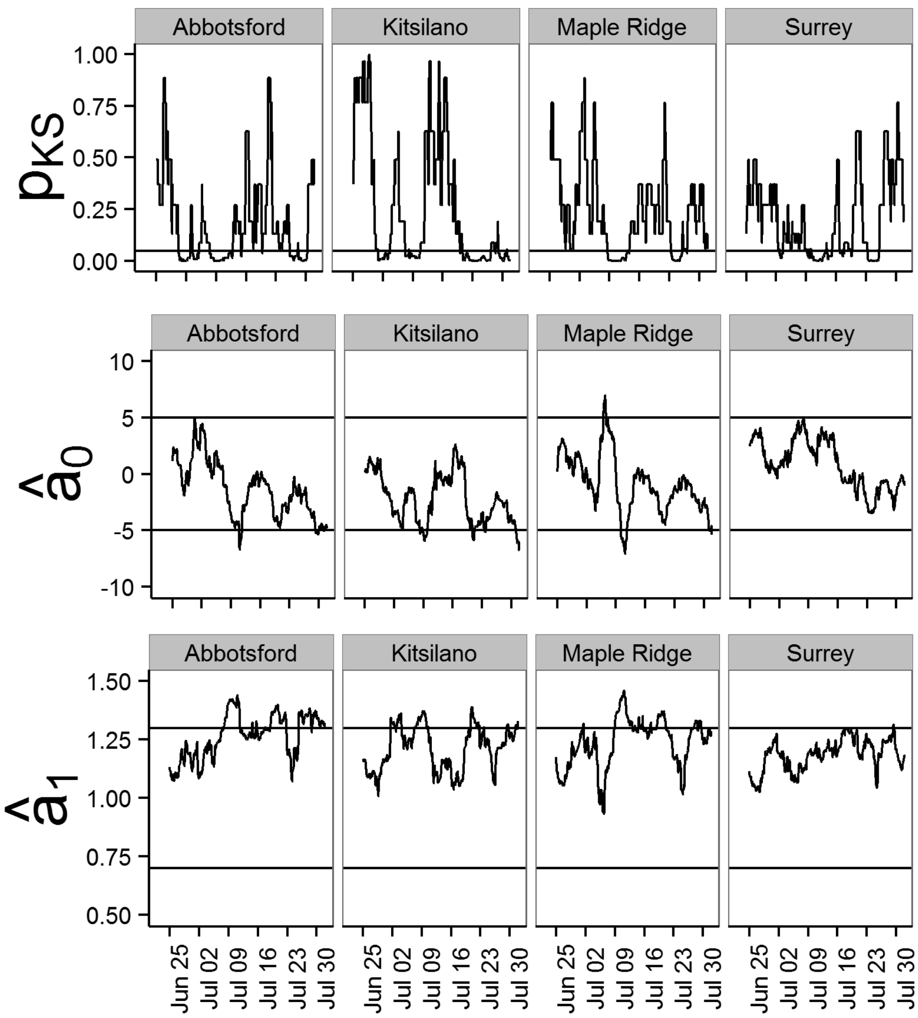


Figure 4. Control charts for four different reference stations (locations in Figure 1) in the "residential" land-use category using the residential GSS network median as a proxy.

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

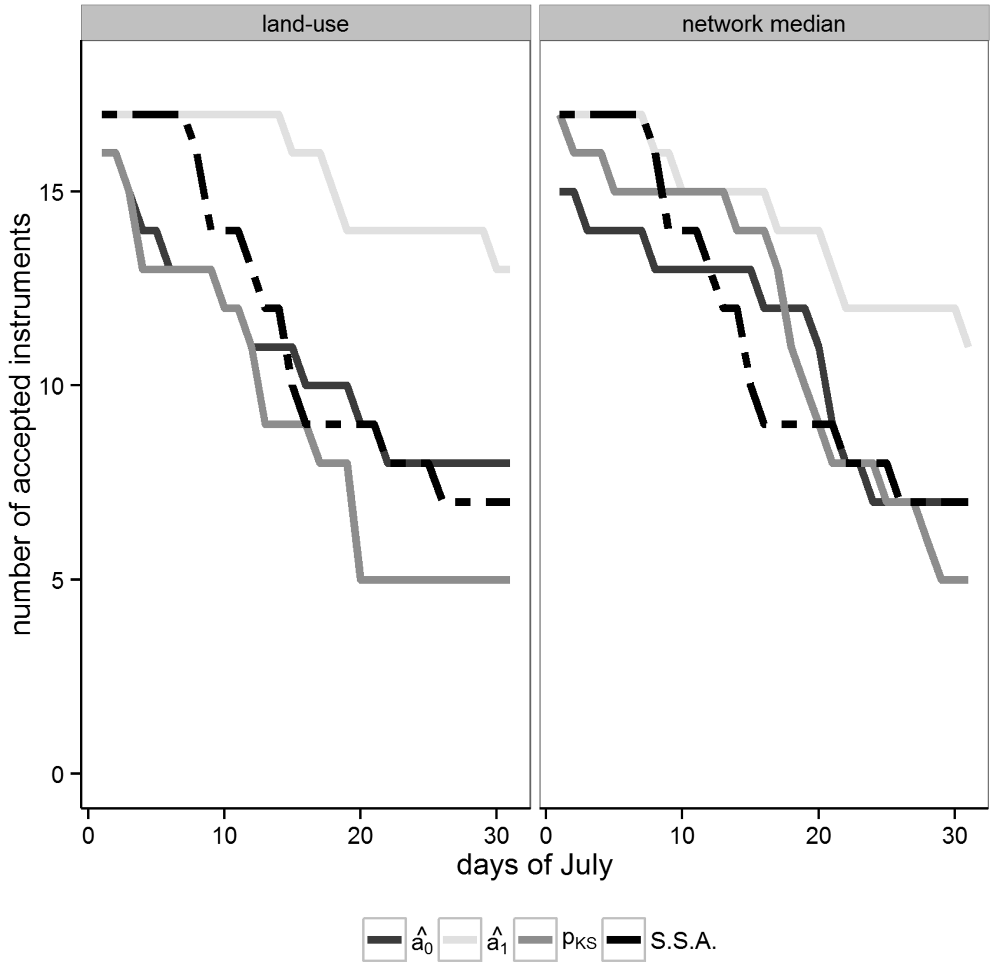


Figure 5. Number of accepted GSS devices in the "residential" land-use category over time, evaluated using the different measures, with either reference station in the same land use as proxy or median of accepted instruments in the same land use as proxy. Single sensor assessment is included as the black dashed line.

results show that, using these methods, instrument calibration drift can be discovered on a time scale of less than 1 week through the use of a 72 h comparison rolling hourly. The hypothesis that periodic natural variation in p_{KS}, $\hat{a}_1$ and $\hat{a}_0$ from site-specific variability would be present in stable instruments was confirmed. This expected behavior could be distinguished from systematic drift of a measurement instrument by the length of time for which the thresholds were exceeded. This means that one can reasonably define a failure time, t_f, to discriminate instrument drift from local variations (see SI). Table 2 shows that this choice led to reasonable consistency of the time of failure indication between the colocated references, the proxies based on the same land-use and the single sensor assessment. Variation was observed to increase, although it remained periodic, if the proxy was not in the same land-use category as the sensor and so false alarms could result if t_f was not sufficiently long (Figure 3). However, if t_f was sufficiently long then, even in this case a stable device such as that at Surrey would be correctly classified and failures correctly identified.

The use of the residential low-cost network median as a proxy against reference stations (Figure 4) showed that the low-cost network median was a stable proxy, and second that, although there was variability between the signals from the reference stations within the "residential" land use classification, the variability was periodic and only occasionally triggered an alarm which was in any case generally not sustained for longer than t_f. That is, the false alarm rate for the methods seems acceptably low given a reliable land-use classification upon which to base the proxy.

Figure 5 shows, for the 18 GSS instruments in the residential category that were not colocated with reference stations, a comparison of failure detection by single sensor assessment with that determined by each of the three methods with two different proxies: reference station in the same land-use, and residential GSS median. For each measure, as devices failed they were removed from the median calculation. The results indicate that p_{KS} and $\hat{a}_1$ for both proxies tended to overestimate drift and $\hat{a}_0$ underestimated. Such effects would however be sensitive to the arbitrary choice of thresholds. Some of the differences between single sensor assessment and the signaling of failure were tracked to an incorrect GSS calibration (one case), the GSS instrument was at an elevated site (one case), or that the GSS instrument was near the edge of the LFV air-shed (one case).

A layered approach to failure identification like the one proposed here could exploit the strengths of each test, and provide better support for drift detection than just one method. A combination of a network median in the same land-use and a reference station in a different land-use as proxies might be a way of using these ideas to work with a network with a lower density of reference stations. These ideas could be extended

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

into a Bayesian formulation that better weighted the different estimators.

The GSS network data also illustrated the potential difficulties associated with siting that could affect results. An example is the subtle drift of the GSS instrument at Abbotsford when it was evaluated against the colocated proxy. Data were accepted by all the other proxy methods and by the single sensor assessment. Examination of the ECD showed this was due to an apparent under-indication of low ozone concentrations by the GSS instrument relative to the local reference station for some nights. The GSS instrument, however, was mounted behind the roof lip of the reference station, which was near a wooded area and beside a large car park, whereas the reference inlet was mounted on a mast above the roof. It is quite possible, therefore, that the GSS instrument was indeed sampling lower concentrations on some nights when the air was probably still.

The ideas developed in this paper point to the feasibility of a monitoring network that combines a small number of compliant reference stations with a large number of low-cost instruments. It would seem reasonable to propose that a few well-maintained reference stations could be used to check network average proxies, and that these proxies could be used to check individual instruments. A network may then not necessarily require a well-maintained reference station in each land-use category. Instead, a network average could be used for each individual category and checked against a reference instrument in other land-use categories. Indications from different proxies could be combined as illustrated in Figure 6.

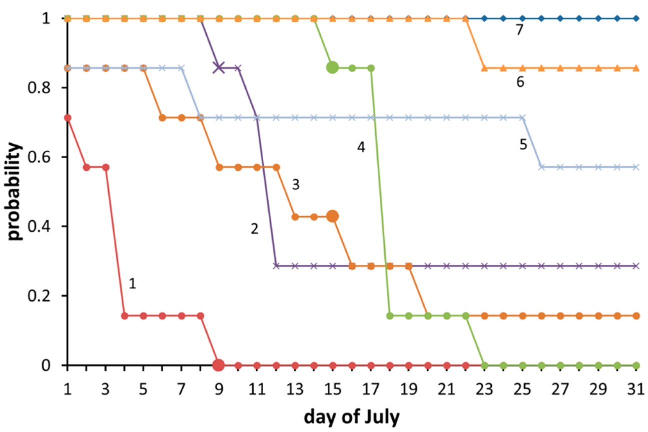


Figure 6. Empirical probability of reliable data, (1 − (number of measures indicating failure/total numbers of measures)) for selected "residential" low-cost GSS instruments over July. The large symbols mark where single sensor assessment labeled failure. A full set of failure times for each measure for all 18 "residential" instruments is in the SI (Table S3). Sites (SI Figure S4 for locations) are 1: North Delta; 2: Horseshoe Bay; 3: Maple Ridge; 4: North Tsawassen; 5:32nd Ave; 6: Surrey East; 7: Abbotsford.

In this way, the network costs, both of capital expenditure and maintenance, to achieve the purpose of providing adequately reliable data could be minimized and wide coverage secured. As the methods here use simple thresholds, they can easily be automated.

While LUR models can explain a substantial fraction (typically 50−80%)[36,40,44] of concentration variability for primary pollutants, it is likely that some of the unexplained variability stems from the lack of meteorological processes included in LUR modeling. As such, simple improvements to LUR models have been made through incorporating meteorological processes. For example, Larson et al.[39] use noncircular buffers in their wood smoke analysis in an attempt to capture nocturnal drainage effects; Ainslie et al.[45] used buffers whose size is governed by local wind speed and atmospheric stability to capture the source-receptor nature of atmospheric dispersion; and Su et al.[46] included wind speed, direction and cloud cover in the regression analysis in order to improve the temporal and spatial variability of the modeled concentration fields. We speculate the ability of our method to assess network reliability might similarly be improved through some simple incorporation of meteorological processes in our proxy development. Ozone is a pollutant that has a relatively smooth spatial spread over a large scale although there can be significant small-scale variations. The use of remote proxies that are rather widely spaced has clearly worked well in this case. Pollutants such as nitrogen oxides have a very large spatial variability so it remains an open question as to the utility of the methods we have described for these cases, particularly the spatial density that would be required of the network. Recent work, however, points to the utility of LUR on a local scale for elucidating explanatory urban design variables that can account for a significant part of the local variability[47] and that therefore in principle can be used to develop proxies.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.5b04421.

Examples of single sensor assessment of failures; ECD and KS definition and formula; Figures S1−S6; Tables S1−S3 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: + 649 373 7599, ext 89877; email: david.williams@auckland.ac.nz.

Notes

The authors declare the following competing financial interest(s): Aeroqual Ltd. are the manufacturers of the air quality sensors used here. AirQuality Ltd. provide services for air quality network management. D.E.W., G.S.H., and B.W. are shareholders in Aeroqual Ltd. D.E.W., M.B., and B.W. are shareholders in Air Quality Ltd. Otherwise, the authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the many individuals and facilities who generously hosted instruments as well as Adrian Jones for his work on decommissioning sites. We are also grateful to the staff of MetroVancouver, particularly Ken Reid and Julie Saxton, for the provision of data and access to sites as well as Roxanne Vingarzan of Environment Canada and Jari Kaipio of University of Auckland for critical discussions. Funding was provided by the NZ Ministry of Business, Innovation and Employment, contract UOAX0709 and UOAX1413, as well as the Canadian Natural Sciences and Engineering Research Council. The University of British Columbia, Department of Geography provided important logistical support.

REFERENCES

(1) Williams, M. Monitoring of exposure to air pollution. *Sci. Total Environ.* **1995**, *168* (2), 169−174.
(2) Knox, A.; Mykhaylova, N.; Evans, G.; Lee, C.; Karney, B.; Brook, J. The expanding scope of air pollution monitoring can facilitate sustainable development. *Sci. Total Environ.* **2013**, *448*, 189−196.
(3) Snyder, E.; Watkins, T.; Solomon, P.; Thoma, E.; Williams, R.; Hagler, G.; Shelow, D.; Hindin, D.; Kilaru, V.; Preuss, P. The changing paradigm of air pollution monitoring. *Environ. Sci. Technol.* **2013**, *47* (20), 11369−11377.
(4) Piedrahita, R.; Xiang, Y.; Masson, N.; Ortega, J.; Collier, A.; Jiang, Y.; Li, K.; Dick, R.; Lv, Q.; Hannigan, M.; et al. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmos. Meas. Tech.* **2014**, *7* (10), 3325−3336.
(5) Kumar, P.; Morawska, L.; Martani, C.; Biskos, G.; Neophytou, M.; Di Sabatino, S.; Bell, M.; Norford, L.; Britter, R. The rise of low-cost sensing for managing air pollution in cities. *Environ. Int.* **2015**, *75*, 199−205.
(6) Ma, Y.; Richards, M.; Ghanem, M.; Guo, Y.; Hassard, J. Air pollution monitoring and mining based on sensor grid in London. *Sensors* **2008**, *8* (6), 3601−3623.
(7) Jiao, W.; Hagler, G.; Williams, R.; Sharpe, R.; Weinstock, L.; Rice, J. Field assessment of the village green project: An autonomous community air quality monitoring system. *Environ. Sci. Technol.* **2015**, *49* (10), 6085−6092.
(8) Wang, A.; Brauer, M. *Review of Next Generation Air Monitors for Air Pollution*, Report for Environment Canada, December 2014.
(9) Mead, M.; Popoola, O.; Stewart, G.; Landshoff, P.; Calleja, M.; Hayes, M.; Baldovi, J.; McLeod, M.; Hodgson, T.; Dicks, J.; et al. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmos. Environ.* **2013**, *70*, 186−203.
(10) Bart, M.; Williams, D.; Ainslie, B.; McKendry, I.; Salmond, J.; Grange, S.; Alavi-Shoshtari, M.; Steyn, D.; Henshaw, G. High density ozone monitoring using gas sensitive semi-conductor sensors in the Lower Fraser Valley, British Columbia. *Environ. Sci. Technol.* **2014**, *48* (7), 3970−3977.
(11) Qin, S.; Badgwell, T. A survey of industrial model predictive control technology. *Control Engineering Practice* **2003**, *11* (7), 733−764.
(12) Moltchanov, S.; Levy, I.; Etzion, Y.; Lerner, U.; Broday, D. M.; Fishbain, B. On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Sci. Total Environ.* **2015**, *502*, 537−547.
(13) De Vito, S.; Piga, M.; Martinotto, L.; Di Francia, G. CO, NO_2 and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic Bayesian regularization. *Sens. Actuators, B* **2009**, *143* (1), 182−191.
(14) Hasenfratz, D., Saukh, O.; Thiele, L. On-the-fly calibration of low-cost gas sensors. In *9th European Wireless Sensor Networks*, Trento, Italy, 2012, *7158*228−244.10.1007/978-3-642-28169-3_15
(15) Holstius, D.; Pillarisetti, A.; Smith, K.; Seto, E. Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. *Atmos. Meas. Tech.* **2014**, *7* (4), 1121−1131.
(16) Environmental Protection Agency. *Air Sensor Guidebook*, EPA/600/R-14/159, 2014.
(17) Muller, C.; Chapman, L.; Grimmond, C.; Young, D.; Cai, X. Sensors and the city: A review of urban meteorological networks. *International Journal of Climatology* **2013**, *33* (7), 1585−1600.
(18) Liu, J.; Chen, D.; Shen, J. Development of self-validating soft sensors using fast moving window partial least squares. *Ind. Eng. Chem. Res.* **2010**, *49* (22), 11530−11546.
(19) Talampas, M., & Low, K. Maximum likelihood estimation of ground truth for air quality monitoring using vehicular sensor networks. In *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2012.
(20) Tsujita, W.; Yoshino, A.; Ishida, H.; Moriizumi, T. Gas sensor network for air-pollution monitoring. *Sens. Actuators, B* **2005**, *110* (2), 304−311.
(21) Alavi-Shoshtari, M.; Williams, D.; Salmond, J.; Kaipio, J. Detection of malfunctions in sensor networks. *Environmetrics* **2013**, *24* (4), 227−236.
(22) Harkat, M.; Mourot, G.; Ragot, J. An improved PCA scheme for sensor FDI: Application to an air quality monitoring network. *J. Process Control* **2006**, *16* (6), 625−634.
(23) Joseph, J.; Sharif, H.; Sunil, T.; Alamgir, H. Application of validation data for assessing spatial interpolation methods for 8-h ozone or other sparsely monitored constituents. *Environ. Pollut.* **2013**, *178*, 411−418.
(24) Hasenfratz, D.; Saukh, O.; Thiele, L. Model-driven accuracy bounds for noisy sensor readings. *9th IEEE International Conference on Distributed Computing in Sensor Systems* **2013**, 165−174.
(25) Zhang, L.; Tian, F.; Peng, X.; Yin, X. A rapid discreteness correction scheme for reproducibility enhancement among a batch of MOS gas sensors. *Sens. Actuators, A* **2014**, *205*, 170−176.
(26) Balzano, L.; Nowak, R. Blind calibration of sensor networks. *IPSN 2007: Proceedings of the Sixth International Symposium on Information Processing in Sensor Networks* **2007**, 79−88.
(27) Chin, G., Jr.; Choudhury, S.; Kangas, L.; McFarlane, S.; Marquez, A. Fault detection in distributed climate sensor networks using dynamic Bayesian networks. *Proceedings - 2010 6th IEEE International Conference on e-Science, eScience 2010* **2010**, 121−128.
(28) Liu, Z.; Mrad, N. Validation of strain gauges for structural health monitoring with Bayesian belief networks. *IEEE Sens. J.* **2013**, *13* (1), 400−407.
(29) Foley, K.; Reich, B.; Napelenok, S. Bayesian analysis of a reduced-form air quality model. *Environ. Sci. Technol.* **2012**, *46* (14), 7604−7611.
(30) Ainslie, B.; Reuten, C.; Steyn, D.; Le, N.; Zidek, J. Application of an entropy-based Bayesian optimization technique to the redesign of an existing monitoring network for single air pollutants. *J. Environ. Manage.* **2009**, *90* (8), 2715−2729.
(31) Shewhart, W.; Deming, W. *Statistical Method from the Viewpoint of Quality Control*; Dover Publications: New York, 1986.
(32) Williams, D.; Henshaw, G.; Wells, B.; Ding, G.; Wagner, J.; Wright, B.; Yung, Y.; Salmond, J. Development of Low-Cost Ozone Measurement Instruments Suitable for Use in an Air Quality Monitoring Network. *Chemistry in New Zealand* **2009**, *73* (1), 27−33.
(33) Williams, D.; Henshaw, G.; Bart, M.; Laing, G.; Wagner, J.; Naisbitt, S.; Salmond, J. Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network. *Meas. Sci. Technol.* **2013**, *24*, 065803.
(34) Adeeb, F.; Shooter, D. Variation of surface ozone in the ambient air of Auckland, New Zealand. *Environ. Monit. Assess.* **2004**, *95* (1−3), 201−220.
(35) Chameides, W. Ozone precursor relationships in the ambient atmosphere. *Journal of Geophysical Research* **1992**, *97* (D5), 6037−6055.
(36) Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42* (33), 7561−7578.
(37) Abernethy, R.; Allen, R.; McKendry, I.; Brauer, M. A land use regression model for ultrafine particles in Vancouver, Canada. *Environ. Sci. Technol.* **2013**, *47* (10), 5217−5225.
(38) Henderson, S.; Beckerman, B.; Jerrett, M.; Brauer, M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol.* **2007**, *41* (7), 2422−2428.
(39) Larson, T.; Su, J.; Baribeau, A.; Buzzelli, M.; Setton, E.; Brauer, M. A spatial model of urban winter wood smoke concentrations. *Environ. Sci. Technol.* **2007**, *41* (7), 2429−2436.
(40) Kerckhoffs, J.; Wang, M.; Meliefste, K.; Malmqvist, E.; Fischer, P.; Janssen, N.; Beelen, R.; Hoek, G. A national fine spatial scale land-use regression model for ozone. *Environ. Res.* **2015**, *140*, 440−448.
(41) MetroVancouver. *Station Information: Lower Fraser Valley Air Quality Monitoring Network*, December 2012.

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(42) Environmental Protection Agency. *Quality Assurance Handbook for Air Pollution Measurement Systems—Volume Two*, EPA-454/B-13-003, 2013.

(43) Cottle, P.; Strawbridge, K.; McKendry, I. Long-range transport of Siberian wildfire smoke to British Columbia: Lidar observations and air quality impacts. *Atmos. Environ.* **2014**, *90*, 71−77.

(44) Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.; Kunzli, N.; Schikowski, T.; Marcon, A.; et al. Development of NO_2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe - the ESCAPE project. *Atmos. Environ.* **2013**, *72*, 10−23.

(45) Ainslie, B.; Steyn, D.; Su, J.; Buzzelli, M.; Brauer, M.; Larson, T.; Rucker, M. A source area model incorporating simplified atmospheric dispersion and advection at fine scale for population air pollutant exposure assessment. *Atmos. Environ.* **2008**, *42* (10), 2394−2404.

(46) Su, J.; Brauer, M.; Ainslie, B.; Steyn, D.; Larson, T.; Buzzelli, M. An innovative land use regression model incorporating meteorology for exposure analysis. *Sci. Total Environ.* **2008**, *390* (2−3), 520−529.

(47) 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. *Environ. Sci. Technol.* **2015**, *49* (15), 9004−9011.

(48) Kahle, D.; Wickham, H. ggmap: Spatial visualization with ggplot2. *R Journal* **2013**, *5* (1), 144−161.

DOI: 10.1021/acs.est.5b04421
Environ. Sci. Technol. XXXX, XXX, XXX−XXX