ARTICLE IN PRESS

STOTEN-21003; No of Pages 11

Science of the Total Environment xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Low-cost sensors and crowd-sourced data: Observations of siting impacts on a network of air-quality instruments

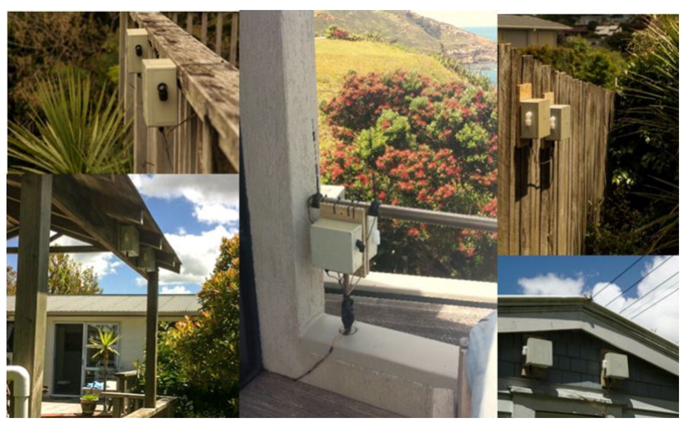
Georgia Miskell [a,b,*], Jennifer Salmond [b], David E. Williams [a]

[a] MacDiarmid Institute for Advanced Materials and Nanotechnology, School of Chemical Sciences, 23 Symonds Street, University of Auckland, New Zealand
[b] School of Environment, 10 Symonds Street, University of Auckland, New Zealand

HIGHLIGHTS

- Intra-site sensor variability was small indicating low sensitivity to siting type.
- Short-term local activities were identified but did not significantly impact reporting scales.
- Crowd-sourced sites in proximity to regulatory analyzers measured similar trends.
- With quality control checks, crowd-sourced networks provided useful additional data.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 15 August 2016
Received in revised form 14 September 2016
Accepted 22 September 2016
Available online xxxx

Editor: D. Barcelo

Keywords:
Ozone
Air quality network
Siting effect
Regional variation
Local variability
Citizen science

ABSTRACT

Low-cost sensors offer the possibility of gathering high temporal and spatial resolution crowd-sourced data-sets that have the potential to revolutionize the ways in which we understand individual and population exposure to air pollution. However, one of the challenges associated with crowd-sourced data ('citizen science'), often from low-cost sensors, is that citizens may use sites strongly affected by local conditions, limiting the wider significance of the data. This paper examines results from a low-cost network measuring ground-level ozone to evaluate the impact of siting on data quality. Locations at both reference stations and at private homes or research centers were used, and thought of as a typical 'crowd-sourced' network. Two instruments were co-located at each site to determine intra-site variability and evaluated by standard performance statistics and local-scale activity logs. The wider application of the data for both regional Inter-site variability was evaluated to show-case the wider value and usefulness of crowd-sourced data. Analysis of intra-site variability showed little differences at most sites (<5 ppb). Large differences in intra-site variability were detected when sensors were exposed to direct sunlight (causing thermal variations within the instrument) and proximity to large emission sources. Short-term local activities, such as lawn-mowing, were identifiable in the data, but had minimal impact on standard reporting time-scales, and so did not pose as being significant limitations or errors. Inter-site evaluation demonstrated that dense networks of low-cost sensors can add value to existing networks, with minimal impact on the

* Corresponding author at: MacDiarmid Institute for Advanced Materials and Nanotechnology, School of Chemical Sciences, 23 Symonds Street, University of Auckland, New Zealand.
E-mail address: gmis011@aucklanduni.ac.nz (G. Miskell).

http://dx.doi.org/10.1016/j.scitotenv.2016.09.177

Please cite this article as: Miskell, G., et al., Low-cost sensors and crowd-sourced data: Observations of siting impacts on a network of air-quality instruments, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.177

overall data-set quality. Sensors located in crowd-sourced locations nearby to regulatory analyzers were able to capture similar trends and concentrations, supporting their ability to report on wider conditions. Thus crowd-sourced approaches to monitoring (with suitable calibration and data quality control checks) may be an appropriate method for increasing the temporal and spatial resolution of air quality networks.

1. Introduction

Recent air quality research has focused on different approaches in using data from low-cost instruments to supplement data provided by official regulatory bodies (Snyder et al., 2013). Whilst low-cost instruments have the potential to make a significant contribution to our understanding of the temporal and spatial variation of air pollutant concentrations in urban areas, concerns over their accuracy and precision have limited their widespread use (Ottinger, 2010; Snyder et al., 2013; Tregidgo et al., 2013). However, recent innovations in techniques to detect sensor error and improve accuracy (e.g. Alavi-Shoshtari et al., 2013; Miskell et al., 2016) are proving increasingly successful, and attention is now moving away from assessing their reliability towards developing best-practice guidelines for the use of this new technology (Nieuwenhuijsen et al., 2015; Xiang et al., 2016, U.S. Environmental Protection Agency, n.d).

One area that has been given little attention so far is the impact of local siting on determining the spatial and temporal representativeness of the data. If low-cost, crowd-sourced data is to be adopted in air quality research, then the impact upon measured concentrations of siting instruments on private homes or education centers needs to be understood. Traditionally, strict regulations surround the siting of regulatory monitoring locations to ensure that datasets are representative of a given area or land-use type and local-scale effects are controlled for (Ministry for the Environment, 2009; U.S. Environmental Protection Agency, 2013). For example, recommendations typically include that the instrument is not adjacent to any walls, avoidance of large trees, certain facades (e.g. wood), and chemical interferences (e.g. vehicle emissions), and above the urban canopy layer (Ministry for the Environment, 2009; Moosavi et al., 2015).

Citizen science approaches which may see instruments located on education centers or private homes or in gardens could provide complementary information to regulatory datasets about the effects of different land-use and settings in previously unmonitored locations (Brienza et al., 2015; Ho et al., 2014). However, they can be expected to violate a number of siting recommendations because of power requirements, aesthetics, and household surroundings (e.g. building material). Data from instruments at poorly selected locations (which may occur in crowd-sourced data due to the siting often being outside of the data users control) may not be representative of wider conditions due to dominant effects of extremely local conditions or events specific to that site. This has the potential to make data from these sites unsuitable for reporting from a network perspective, and any temporal or spatial averaging of the data could be misleading from air quality management perspectives. There is therefore a need to assess the impact of different types of siting and to develop quality assurance techniques to allow the citizen scientist (and those using that data) to know how to interpret, and what value to place on, the data from their instrument (Bonney et al., 2016; Ho et al., 2014; Wolters et al., 2016).

This paper examines the effect of local siting on data quality to address the overall enquiry on the usefulness of low-cost data. Data from a network of instruments (mounted on a variety of siting options, such as on regulatory stations or on walls of private houses) were analyzed for their intra- (within a site) and inter- (between sites) variability. Differences within a site were compared to their surroundings using regression and standard statistical diagnostics to ascertain whether certain factors were related to large intra-site differences. Factors with large differences could then be recommended to the citizen scientist to avoid when mounting an instrument, or to the data user in deciding whether to include the site within network analysis. Inter-site analysis examined how a crowd-sourced network can assist in developing and improving our understanding of the temporal and spatial variability of urban O_3 by using standard statistical diagnostics. Finally, differences between crowd-sourced sites to nearby reference stations were analyzed for their ability to capture the wider pattern and to give support for providing data representative of an area.

2. Materials and methods

2.1. Data

The data used here were collected from a network of low-cost instruments measuring ground-level ozone (O_3) around Auckland, New Zealand, over a twelve-month period (November 2014–November 2015) with two instruments operating per site (<2 m distance apart). The data were validated by using methods described previously, with good quality data capture for over 75% of all observations (Bart et al., 2014; Williams et al., 2013). Auckland has a subtropical oceanic climate, with humid summers and mild winters and prevailing wind direction from the Southwest (Adeeb and Shooter, 2004). O_3 is a secondary pollutant formed from the photochemical reaction of NO_X or VOCs with UV, which causes regular spatial profiles and so regional patterning can be expected from synoptic weather patterns up and downwind of urban centers or central business districts (CBD) where precursors are produced (often traffic-related) (Bart et al., 2014). O_3 concentrations in Auckland are typically low compared to other urban centers due to titration from nitrous gases along with the geographic setting (Jiang et al., 2014), with a peak of O_3 in the winter to spring months (July–October), believed to be from greater stratospheric intrusion rather than local sources (Adeeb and Shooter, 2004). High O_3 days occur at different times at different locations across Auckland, suggesting the significance of local-scale controls (Adeeb and Shooter, 2004). Auckland has three reference stations measuring O_3 (Fig. 1), with two (Musick Point, MP, and Whangaparaoa, WHA) operating only during the summer months. Therefore, our understanding of O_3 throughout the year in Auckland is determined from one site (Patumahoe, PAT).

The low-cost sensors used were Aeroqual gas-sensitive semiconducting oxide (GSS), which have been successfully used in a number of field studies (Bart et al., 2014; Deville Cavellin et al., 2016; Lin et al., 2015) and were found to have good performance when compared against other commercial low-cost instruments (SCAMD, 2015). Sensor specifications state a level of accuracy to 5 ppb, which has been used here as a benchmark threshold as a true, or real, difference between co-located measurements. Previous work (Bart et al., 2014) found a standard error of 6 ppb when devices were compared to co-located analyzer stations for over 6000 measurements, giving support for this level of precision of the device. All low-cost instruments were field linearized and adjusted data reported, with detailed information on corrections presented in the Supporting material. The methods described by Bart et al., 2014 were used to check instrument performance; the sensor assembly in the instrument was replaced when a signal or baseline drift was detected (typically every 2–5 months; median 3 months). Overall, the response was good, with high linearity and no significant differences between co-located concentrations following calibration. Site locations are illustrated in Fig. 1, and include both reference ($n = 3$)

Please cite this article as: Miskell, G., et al., Low-cost sensors and crowd-sourced data: Observations of siting impacts on a network of air-quality instruments, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.177