Sustainable Cities and Society 60 (2020) 102239

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned

Priyanka deSouza[a,*], Amin Anjomshoaa[a], Fabio Duarte[a,b], Ralph Kahn[c], Prashant Kumar[d], Carlo Ratti[a]

[a] Senseable City Lab, Massachusetts Institute of Technology, Cambridge MA, United States
[b] Pontifícia Universidade Católica do Paraná, Brazil
[c] NASA Goddard Space Flight Center, Greenbelt, MA, 20771, United States
[d] Global Centre for Clean Air Research (GCARE), Department of Civil and Environmental Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, GU2 7XH, United Kingdom

ARTICLE INFO

Keywords:
Drive-by sensing
Low-cost sensors
Particulate matter
Routine sampling

ABSTRACT

Air quality monitoring (AQM) is crucial for cities to develop management plans supporting population health. However, there is a dearth of measurements due to the high cost of standard reference instruments. Mobile AQM using low-cost sensors deployed on routine fleets of vehicles can enable the continuous detection of fine-scale pollutant variations in cities at a lower cost. New methods need to be developed to interpret these measurements. This paper presents three such methods. First, we propose a technique to identify aerosol hotspots. Second, we employ techniques published previously to assess the generalizable map of fine and coarse particle number concentrations, to understand qualitatively the contribution of local and regional sources across the region sampled. By using the raw number concentration of differently sized particles from the Optical Particle Counters (OPCs) instead of the noisier mass concentrations, we obtain more robust results. Third, in order to evaluate source signatures in cities, we propose another technique, in which we cluster the entire range of aerosol size-distribution measurements acquired. The properties of each cluster provide insight into the aerosol source characteristics in the sampling environment. We test these methods using a dataset we collected by mounting OPCs on two trash-trucks in Cambridge, Massachusetts.

1. Introduction

Poor air quality is a major environmental health risk in cities. Air quality monitoring is crucial for developing informed air quality management plans. However, setting up and maintaining air quality monitors is expensive. Even in the United States, resource constraints dictate that the regulatory air quality monitoring network is sparse, with only 2–5 regulatory monitors per 1 million people or 1000 km^2 in 60 % of US census urban areas (Apte et al., 2017).

Furthermore, air pollutant concentrations in complex urban environments display high variability and sharp gradients over distances as small as 10 m (Brantley et al., 2013; Van den Bossche et al., 2015). To characterize the reactive-pollutant spatial variation *in situ*, even a dense (but realistic) network of fixed air quality monitors cannot capture this variability. Mobile air quality monitoring can be used to obtain air pollution concentrations at high spatial resolution with a smaller number of monitors over a fixed period of time.

The difficulty of working with mobile air quality monitoring data arises from the combination of complex spatiotemporal sampling and temporal air quality variability in different locations, related to traffic dynamics, street topology, meteorology, background source strength, etc. (Goel & Kumar, 2015; Van den Bossche et al., 2015). As mobile sensors capture only a snapshot of air pollution at a given location and time, this temporal variability makes it difficult to characterize the air pollution at a given location based on these measurements alone. In order to produce detailed, representative aggregate air quality maps, large amounts of data over different meteorological and traffic conditions would be required (Apte et al., 2017; Brantley et al., 2013).

'Purpose-built' mobile air quality monitoring labs often require dedicated vehicles and trained research staff as drivers. For this reason, most mobile air quality monitoring studies to date have been relatively short-term campaigns and provide insufficient repetitive frequency to reveal long-term spatial air quality trends in a city (Bukowiecki et al., 2002; Kolb et al., 2004; Pirjola et al., 2012, 2004). Apte et al. (2017)

* Corresponding author.
E-mail address: desouzap@mit.edu (P. deSouza).

https://doi.org/10.1016/j.scs.2020.102239
Received 29 September 2019; Received in revised form 10 May 2020; Accepted 11 May 2020
Available online 18 May 2020

conducted one of the first long-term mobile air quality monitoring studies with a routine fleet of vehicles. Their study used high-quality reference PM air quality monitors on Google Street View cars to repeatedly sample every street in Oakland, CA, over the course of a year.

On the other end of the spectrum of air quality monitors, the use of low-cost monitors (costing less than USD $3000) is increasing (Kumar et al., 2015; Morawska et al., 2018; Snyder et al., 2013), and several community-based mobile monitoring studies have used these low-cost instruments (Dutta et al., 2009; Elen et al., 2013). Mahajan and Kumar (2020) have evaluated the use of low-cost sensors for quantifying personal exposure. Low-cost monitors offer the possibility of systematic air quality monitoring even for resource-limited cities.

We add to this literature by developing new techniques to analyse measurements from mobile low-cost sensors deployed on another routine fleet of vehicles: trash-trucks, which are deployed in nearly all major cities globally. Specifically, we show how data from mobile, low-cost sensors can be used: 1) to detect pollution hot spots from major, fixed, possibly intermittent sources in the built environment, 2) to develop a qualitative understanding of where local versus regional sources dominate in a city, and 3) to identify pollutant source signatures on different street segments. Such insights about air pollution can help city managers develop effective air quality management plans.

Anjomshoaa et al. (2018) compared the utility scheduled vehicles, such as trash-trucks, with non-scheduled vehicles, such as taxis, as urban air quality sensing platforms in cities. Although taxis operate 24 h a day and potentially sample at a higher rate the busiest streets in a city, there is no guarantee that they cover all streets. On the other hand, although trash-trucks operate for short periods during the day, and have a lower sampling frequency of most streets, they provide complete spatial coverage of city streets. Therefore, using trash-trucks as a scalable sampling platform is worth investigating.

We test our techniques using measurements made in Cambridge, Massachusetts, between April 21, 2017 and August 14, 2017, where we deployed low-cost Optical Particle Counters (OPC-N2s) on two trash-trucks as a pilot experiment.

The rest of this paper is organized as follows: Section 2 describes the data used, the collection strategy, and the analysis methods. Section 3 presents the results obtained from applying the three methods sketched out above to our Cambridge dataset. Lessons learned and practical implications for future deployments are given in Section 4.

2. Materials

2.1. Low-cost particulate matter monitors

We use the data collected by Alphasense OPC-N2 monitors[1] deployed on two trash-trucks in the City of Cambridge for a total of 27 days between April and August 2019, to gain qualitative insights into potential sources of PM in this urban environment. The Alphasense OPC-N2 sensor measures particle counts ($N_{0.38\text{-}17.5}$) in 16 size bins ranging from 0.38 to 17.5 μm (Table S1, Supplementary Information). The OPC works by illuminating one particle at a time with focused laser light and measuring the intensity of light scattered. The amount of scattering from a particle is a function of the particle size. The instrument is calibrated using monodisperse particles of known size to derive counts for particles of different sizes.

The number and volume concentrations of particles can be obtained by dividing the particle counts by the flow rate and sampling time. The log-normal size distribution of particles at the midpoint of each diameter bin can be calculated using Eq. (1):

$$\frac{dN}{dln(D)}\ per\ volume\ of\ air\ at\ D_{midpoint} = \frac{\Delta N}{ln(D_{upper}) - ln(D_{lower})} \times \frac{1}{flow\ rate \times sampling\ time} \quad (1)$$

where N is the number concentration of PM within a size bin (#/mL), D is the diameter of the particles, D_{upper}, D_{lower} and $D_{midpoint}$ are the upper, lower and midpoint diameters of the OPC-N2 bins in units of (μm), ΔN is the number of particle-counts in each bin.

The particle counts agree well with reference instrument measurements for coarser particles (> 0.78 μm), providing detection efficiency ranging from 83 %–108 %, but the particle counts for finer particles (< 0.78 μm) are underestimated (detection efficiency ~78 % for monodisperse polystyrene spheres (Sousan, Koehler, Hallett, & Peters, 2016). Despite the greater noise in the detection of particles in the lower bins, the OPC measurements still provide useful information of the amount of finer aerosol in the atmosphere, and we retain these bins in this analysis.

A partly proprietary Alphasense data reduction algorithm makes assumptions about particle density and the number of particles with diameters smaller than 0.38 μm, to report PM_1, $PM_{2.5}$ and PM_{10}. These assumptions create uncertainty in the PM values that vary based on the ambient aerosol size distribution and density at the time of measurement.

Although the Alphasense OPC-N2 monitors are thus of lower quality than reference instruments, the particle number concentration measurements in 16 size bins represent much better constraints on the true values than the derived, size-resolved particle mass, as exemplified by the work of Sousan et al. (2016). Such information can constrain the aerosol size distribution over space, which can indicate local/regional pollution sources, as we demonstrate in the current study. It must be noted that If the aerosol is hygroscopic, under conditions of high humidity (RH > 85 %) the OPC interprets the hydrated particles as larger "dry" particles, and the reported number concentrations will have errors (Crilley et al., 2018). The RH during times of measurement in Cambridge was between 60–70 %, and therefore particle hydration is unlikely to be a major concern during our experiment.

In addition to characterizing the variation of PM_1, $PM_{2.5}$ and PM_{10} across our sampling route, we also aggregate the spatial variation of particle number concentrations in different size bins derived from the raw OPC-N2 measurements: N1 ($N_{0.38\text{-}1}$, comprising particles with diameters between 0.38 μm and 1 μm), and coarser particles: N12 ($N_{1\text{-}12}$, covering particles with diameters between 1 μm and 12 μm).

For more information about the experiment design, the study area and sampling protocol and the days on which sampling runs occurred in Cambridge, MA, please refer to Section S1 in Supplementary Information.

3. Methods

In this section we present three techniques designed to identify and characterize $PM_{2.5}$ hotspots, to estimate the generalised air pollution over the sampled routes, and to analyze aerosol size distribution from the OPC-N2s, yielding estimates of PM source signatures in different parts of the sampling route.

3.1. Identification and characterization of $PM_{2.5}$ hotspots

We identified all measurements where $PM_{2.5} > 100\ \mu g/m^3$, that we arbitrarily selected to be much higher than the EPA daily average standard of 35 μg/m^3. Such high measurements are substantially above the background values in the study region, as presented subsequently, and could either be 1) noise from the OPC-N2, or 2) an indicator of a strong local source of pollution.

In order to identify measurements that were products of local sources rather than noise, we used hierarchical clustering (Johnson,

[1] http://www.alphasense.com/index.php/products/optical-particle-counter/ ".

1967; Langfelder & Horvath, 2012) to cluster measurements made in the same spatial area using a distance cut off of 100 m.

We calculated the number of measurements in each cluster, as well as the number of unique days over which the measurements in each cluster were made. Most clusters contained only a few measurements, indicating that they could represent measurement noise. We thus only focused on clusters that contained > 30 measurements, or those for which the number of unique measurement days was greater than 1. We characterized each cluster by calculating the average properties (PM_1, $PM_{2.5}$, PM_{10}, N1, N12) of each cluster. The temporal nature of these hotspots and their average properties allow us to deduce potentially important local fixed sources that contribute to the hotspot formation. This information is useful for planners in developing air pollution management plans.

Even in locations where we see pollution hotspots, it is possible that the source involved only operates for short periods of time, so typical pollutant concentrations might be low. In the next subsection, we describe how we obtain the relative distribution of typical pollutant concentrations across the trash-truck sampling route.

3.2. Methodology for estimating generalized air pollution over the sampled routes

3.2.1. Pre-processing: background correction

To compare measurements made at the same location but on different days and at different times, we need to account for possible bias created by diurnal variation in background aerosol number concentration and PM concentrations over the study region. We assume that the background value varies temporally but not spatially over the region.

We assessed the background contribution using three different methods. The first method involves applying an hourly multiplicative factor derived from concentrations reported by a reference air quality monitor at a designated background site (Hagler et al., 2012; Van Poppel, Peters, & Bleux, 2013). In our case, this requires using the regulatory monitor at Boston's North End (N:42.363, E:-71.055, 4 km southeast of the center of the study region). Unfortunately, here is no reference air quality monitor in Cambridge, the site of our experiment. This technique involves uncertainties, in part because the OPC optical measurements are not directly comparable to the reference monitor's gravimetric ones.

The second method, following (Bukowiecki et al., 2002), takes the lowest 10th percentile of the pollutant concentrations for a given hour during the run as the fixed background value for that run.

The third method uses a time-series, spline-of-minimums approach, presented by Brantley et al. (2013), to estimate the background number concentrations of finer particles: N1, N12, as well as $PM_{2.5}$ for each day. We did this by (a) applying a rolling 30-second mean to smooth the measurements, (b) dividing the time series into discrete 10-minute segments and locating the minimum concentration in each segment, and (c) fitting a smooth, thin-plate regression spline through the minimum concentrations. Note that on nine days the two OPCs were operating simultaneously. We consider the total observations made for a given day in this methodology, consistent with our assumption that the background is temporally varying but spatially uniform.

We compared each of the three proposed methods to choose a background pollution value, and found that they produced similar corrected values. Specifically, the mean differences in the corrected N1, N12 and $PM_{2.5}$ values using the different methods were less than 5%. The differences between the corrected values and the raw N1, N12 and $PM_{2.5}$ measurements were also less than 5% (Table S2 in Supplementary Information). Given the minimal differences in background-corrected number concentrations and $PM_{2.5}$ values vis a vis the raw measurements using the different methods, we chose the splines-of-minimum approach to obtain background concentrations for all pollutants. This is supported by Brantley et al. (2013), who found this approach to be an effective way to account for background concentrations for a range of pollutants in their North Carolina study over a variety of meteorological conditions and sampling routes.

Once the method for evaluating background air pollution was selected, we performed a background time-of-day correction using Equations 2 and 3 to account for the period during which the trucks operated:

$$PM_{2.5,norm\ i} = PM_{2.5,OPC,\ i} \times PM_{2.5,\ bkg,median} / PM_{2.5,bkg,i} \quad (2)$$

where $PM_{2.5,OPC,i}$ is the OPC measurement for event i, $PM_{2.5,bkg,\ i}$ is the contemporaneous background value of pollution over the entire region, and $PM_{2.5,\ bkg,median}$ is the median of the $PM_{2.5,bkg}$ values on the day of measurement for the time period 07:00 to 14:00 h (local time).

However, Cambridge is a city with relatively clean air, so the background $PM_{2.5}$ is often very low. As a result, we apply an additive rather than a multiplicative background-correction factor:

$$PM_{2.5,norm\ i} = PM_{2.5,OPC\ i} - PM_{2.5,\ bkg,i} + PM_{2.5,\ bkg,median} \quad (3)$$

By subtracting the time-of-day-resolved regional background from the pollution measurement, we can now compare local air pollution over space. Note that, conversely, $PM_{2.5}$ in Equations 2 and 3 can be replaced with particle number concentrations (N1 or N12) to estimate background-corrected aerosol number concentrations.

3.2.2. Estimating generalizable pollutant values across the sampling route

Given our large dataset (> 500,000 observations), we applied a series of steps to convert the data into estimates of median concentrations for individual road segments over all sampling runs. We constructed these road segment estimates by dividing the Cambridge street network into segments of fixed lengths, using the 'Locate Points Along Lines' QGIS Python Plugin (https://plugins.qgis.org/plugins/LocatePoints/). We adopted a process for aggregating these data and deriving sampling error from previous work (e.g., Apte et al. (2017)).

First, we spatially-aggregated all our mobile (1) background-adjusted $PM_{2.5}$, and (2) background-adjusted number-concentration measurements for 0.38 μm–17.5 μm particles from the OPC-N2s, by snapping them to the road segment on which they were acquired. This allows measurements made in the same segment to be analysed as a group. This distance is small enough to capture pollutant-concentration gradients, but is not so finely sliced that GPS errors overwhelm the results. Therefore, we also use segment lengths of 30 m.

Second, we selected the median as an outlier-resistant metric of PM and number concentration central tendency, as others have done (Apte et al., 2017; Hankey & Marshall, 2015). We chose not to remove peak concentrations caused by encounters with vehicle exhaust plumes, as such plumes contribute to the particle concentration at a given location. For comparison, we also calculated the mean values, with the understanding that individual outliers can significantly skew those results.

Third, we used a set of bootstrap resampling procedures to quantify the effect of sample-to-sample variability and of sampling error on the median concentrations. As a metric of precision, we used the ratio of standard error of the median (mean) concentration to the median (mean) concentration itself. In general, the average skew of the median $PM_{2.5}$ concentrations for the 30 m segments is ~0.9. This indicates that the distributions are close to central tendency, with the mean slightly greater than the median.

3.3. Working with the aerosol size distribution from the OPC-N2s

It is challenging to analyse the aerosol size distribution at each point in time, because for each measurement Eq. (1) allows us to calculate the size distribution at the midpoint of each of the 16 bins. We thus have 16 data points for each time-step. Pey et al. (2008) showed that aerosol number and mass concentrations can be affected by multiple sources and atmospheric processes.

To simplify the analysis of the OPC size distributions systematically, we clustered the size-bin observations (without background-correction) using the k-means technique. The final cluster centres reflect particle number size distributions representative of each cluster, thereby reducing the complexity of the dataset (Beddows, Dall'Osto, & Harrison, 2009). This technique allows us to identify a small number of typical aerosol size distributions that can be compared across space and time, which can give us insights into the kinds of sources responsible for measurements within a cluster.

Without access to the size distribution of the background aerosol, it is impossible to perform a background correction on the aerosol size distribution. However, by applying the k-means clustering technique on the complete range of raw size distribution data, we are able to gain a better understanding of the source attribution of aerosols in the sampling route environment.

In order to choose the number of clusters, we examined the within-group sum-of-squares error for cluster sizes ranging from 2 to 30 to determine an optimum number of clusters. Fig. S3 in Supplementary Information shows that the error in representing the full dataset decreases sharply between 1 and 4 or 5 clusters. When we applied k-means clustering with more than 5 groups, the average size distributions of the newly created clusters had a similar shape to that of previously identified clusters, albeit with different total number concentrations. Normalizing the size distribution by the number concentration of aerosols per measurement might have led to better results. Unfortunately, because the OPC doesn't detect particles with diameters < 0.38 μm, the total aerosol number concentration at each measurement is unknown, and we were thus unable to normalize the measured size distribution.

To avoid over-interpreting the data, we cluster the data into five groups. We evaluated the average size distribution and the spatial and temporal variation of each cluster to infer source characteristics and assess the aerosol dynamics at work in our dataset.

4. Results and Discussion of the application of these techniques in Cambridge

4.1. Hotspot identification

Fig. 1 shows the results of performing hierarchical clustering on $PM_{2.5}$ measurements > 100 μg/m^3 as described in Section 3.1. Forty-four distinct clusters were identified. Thirty-seven of the clusters contained fewer than 10 measurements made over the course of a unique day, indicating that these spikes could be artefacts. We highlight in Fig. 1 four of the seven other clusters of $PM_{2.5}$ values recorded by the OPC, where the number of measurements is > 30, or the number of unique days over which the measurements are made are > 1.

Table S3 in Supplementary Information reports the number of measurements that comprise each cluster and the number of unique days over which measurements in each cluster were made. Table S3 also provides a google maps image of the location at which each cluster was made. All of these clusters were only detected on a handful of days, indicating the temporal intermittency of the local sources responsible for these hotspots that we now explore. In addition, Fig. S4 in the Supplementary Information depicts the average particle size distribution for each cluster. The Table associated with Fig. S4 reports the average PM_1, $PM_{2.5}$, N1, N12 for each cluster and the time at which the measurements in each cluster were observed.

We now explore each of the four robust hotspots in detail:

1) Cluster 1 comprises 30 measurements of high $PM_{2.5}$ values that were made at the site marked '1' in Fig. 1, which is an organic farm in Rocky Hill Farm, in the city of Saugus, Massachusetts, also a site of waste disposal. A smaller cluster comprising 11 measurements is nearby. However, it is worth noting that on only one of the 27 days of sampling a trash-truck travelled north to the Saugus dumping zone. (See Fig. S2 in Supplemental Information, which displays the number of unique days over which each street along the trash-truck routes was sampled.)
2) The location of cluster 2 is at the Department of Public Works, where the trash-trucks are housed. This is the largest hotspot, with > 1,800 $PM_{2.5}$ measurements exceeding 100 μg/m^3. Although the trash-trucks travelled to and from this location on every day of their operation, such high $PM_{2.5}$ values were seen on only three days during the experiment.
3) The third cluster is located at a waste collection site in Roxbury, Boston, MA. The number of unique days over which measurements comprising this cluster were made is ten. This indicates that there is likely a major, fairly consistent source of particulate matter at this site.
4) The last $PM_{2.5}$ cluster is on Hamlin Street, in Cambridge, close to a large parking area and a park. It consists of 34 measurements made over a span of two unique days.

The modal diameter of the aerosol distribution on Hamlin Street and one of the clusters at Saugus (the smaller cluster, comprising 11 measurements) is > 1 μm, as seen in Table associated with Fig. S4. As most combustion pollution particles tend to be well below 0.5 μm in diameter, this suggests that local soil or dust particles make a large contribution to pollution at these sites. From the Table associated with Fig. S4, the high ratios of $PM_{10}/PM_{2.5}$ for these clusters bolsters this hypothesis. The shape of the average aerosol size distribution of the other cluster at Saugus indicates a massive number concentration of fine particles at this site, which is borne out by the low $PM_{10}/PM_{2.5}$, high N1 and low N12 concentrations. At Roxbury, the shape of the aerosol distribution indicates a complex environment, with high concentrations of fine as well as coarser particles. A massive number concentration of coarse particles is found at the Cambridge Public Works Department.

Three of the four hotspot locations identified above are at waste disposal sites, and the Cambridge Public Works Department concentration is likely due to the indoor housing of trash-trucks. In both cases, it indicates that personnel, such as the trash-truck drivers, are exposed to high pollution values at these locations. Stationary monitoring would be required to measure air quality at the sites when the trash-trucks are not present, to interpret these air pollution values generally.

In the next subsection, we highlight the general values of pollution across the sampling routes, taking into consideration all measurements made over the period of study. This gives us information about the 'typical' sources that contribute to pollution at each location.

4.2. Spatial patterns

The median background-corrected concentration of $PM_{2.5}$, fine particles (N1), and coarse particles (N12) are depicted in Fig. 2a, b and c respectively. Fig. S5a (Supplementary Information) shows the median $PM_{2.5}$ on road segments where the normalised error in the median $PM_{2.5}$ derived from bootstrapping is ≤ 20 %, and the number of unique days on which a road segment was sampled exceeds unity. Fig. S5b and c are similar plots for N1, and N12. Fig. S5 thus shows us pollutant values at locations along the sampling route for which we are reasonably confident to have estimated the 'typical' value of pollution during the period of study.

Fig. 2a indicates that on average, $PM_{2.5}$ in Cambridge is likely low and uniform across the city for weekdays between 07:00 to 14:00 local time, when and where the trash-trucks operate. Some of the locations where high $PM_{2.5}$ values are observed coincide with previously identified hotspots, such as at the Roxbury waste disposal/transfer site, whereas at the Cambridge Public Depot and Hamlin Street, we see $PM_{2.5}$ observed are only moderately high, indicating that hotspots of pollution in the latter locations are sporadic or atypical.

High values of $PM_{2.5}$ over the period of study were also observed on

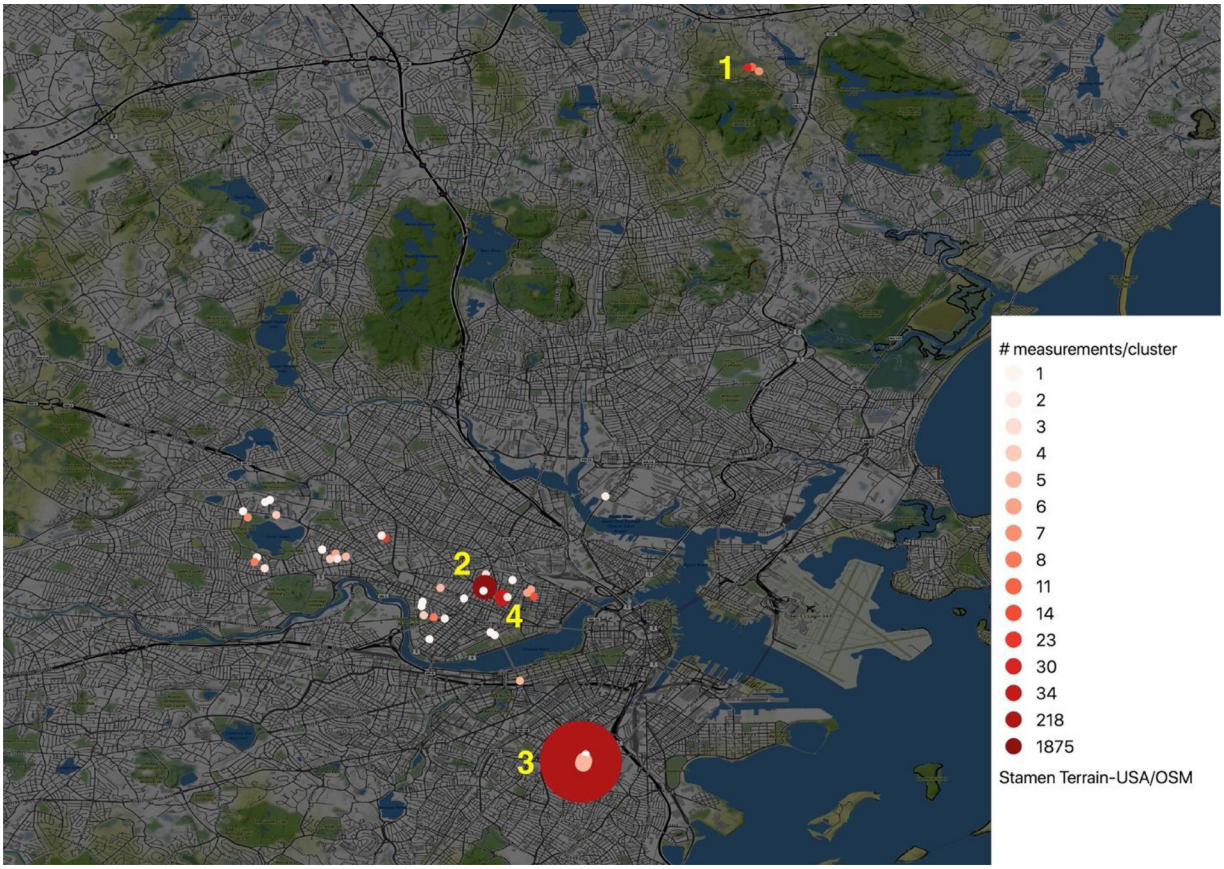

Fig. 1. Locations of the centroid of clusters produced via hierarchical clustering described in Section 3.1. The color of the cluster represents the number of data points in each cluster. The size of the cluster corresponds to the number of unique days of measurement corresponding to each cluster. Four large hotspots/clusters (number of points in the cluster are > 10 and number of unique days of measurement > 1) are identified and numbered.

Prospect Street (close to the trash-truck depot), as well as on Broadway across the Malden Bridge, near the Everett Casino construction site. As the normalised error in the median is > 20 % (Fig. S5a), further generalizable statements about levels of pollution at these locations would require more measurements. The sites with the highest $PM_{2.5}$ values in Fig. 2a also appear as hot spots for N1 (Fig. 2b) and N12 (Fig. 2c), indicating high concentrations of both sub-micron and super-micron particles. However, from Fig. S5b and c, as for $PM_{2.5}$, only the calculated 'generalizable' N1 and N12 concentrations at the Roxbury waste transfer/disposal site is stable. More measurements need to be made at the other locations to gain confidence in the 'typical' pollutant concentration levels at these locations. Other than these sites, the $PM_{2.5}$ values observed during the sampling are much lower than the EPA daily averaged standard of 35 μg/m^3 overall, as well as the EPA annual standard of 12 μg/m^3.

There are at least two reasons why the distribution of larger particles is likely to be more localized than that of fine particles: 1) Larger particles tend to travel shorter distances than finer particles under similar wind conditions (Wilson, Kingham, Pearce, & Sturman, 2005), and 2) There are additional sources of fine particles. This is indeed the case: Fig. S5c shows high local concentrations of N12 along Cambridge Street, where we observed many construction projects going on during the period of sampling. N1 is more dispersed along Cambridge Street and its surrounding environs (Fig. S5b). In addition, high fine particle concentrations on main roads, such as Brattle Street and Cambridge Street, indicate that vehicular traffic in these areas are additional sources of fine particulate matter.

4.3. Analysis of the size distribution of particulates monitored

As discussed in Section 3.3, we used K-means clustering to interpret the OPC particle size-concentration data, and identified five clusters in the optimal grouping. Unlike the hierarchical clustering method presented in Section 3.1, where similar measurements located within a radius of 100 m of each other were grouped, we use the k-means analysis to identify signatures of similar sources across time *and* space. The average size distribution of each cluster is shown in Fig. 3. Table 1 gives the average pollutant concentrations and trash-truck velocity corresponding to each cluster, as well as the number of days on which measurements corresponding to each cluster were made.

Fig. 3 shows that the aerosol concentration mode values for clusters 3 and 5 occur at ~0.78 μm diameter. Modes for clusters 1, 2, and 4 occur at diameters < 0.38 μm. This indicates that the sources contributing to the measurements in the different clusters are likely distinct. Fig. S6 in Supplementary Information is a map of the most frequent cluster present on each 30-meter road segment in Cambridge.

Clusters 1 and 4, having the lowest contributions from particles larger than 0.38 μm, dominate in most parts of Cambridge. Cluster 4 is dominant on main roads/major intersections on the sampling routes, with high concentrations of background-corrected N1 (Fig. 2b). Although from Table 1 the background concentration of $PM_{2.5}$ makes up a large fraction of the $PM_{2.5}$ measured, it appears that measurements corresponding cluster 1 are almost entirely due to background/regional PM sources. Cluster 4, on the other hand, is composed of measurements where vehicular traffic contributes noticeably to the aerosol load.

After clusters 1 and 4, cluster 2 is the most prevalent, with high values of pollutants (though lower than that in clusters 3 and 5). Observations corresponding to this cluster are observed on all 27 days.

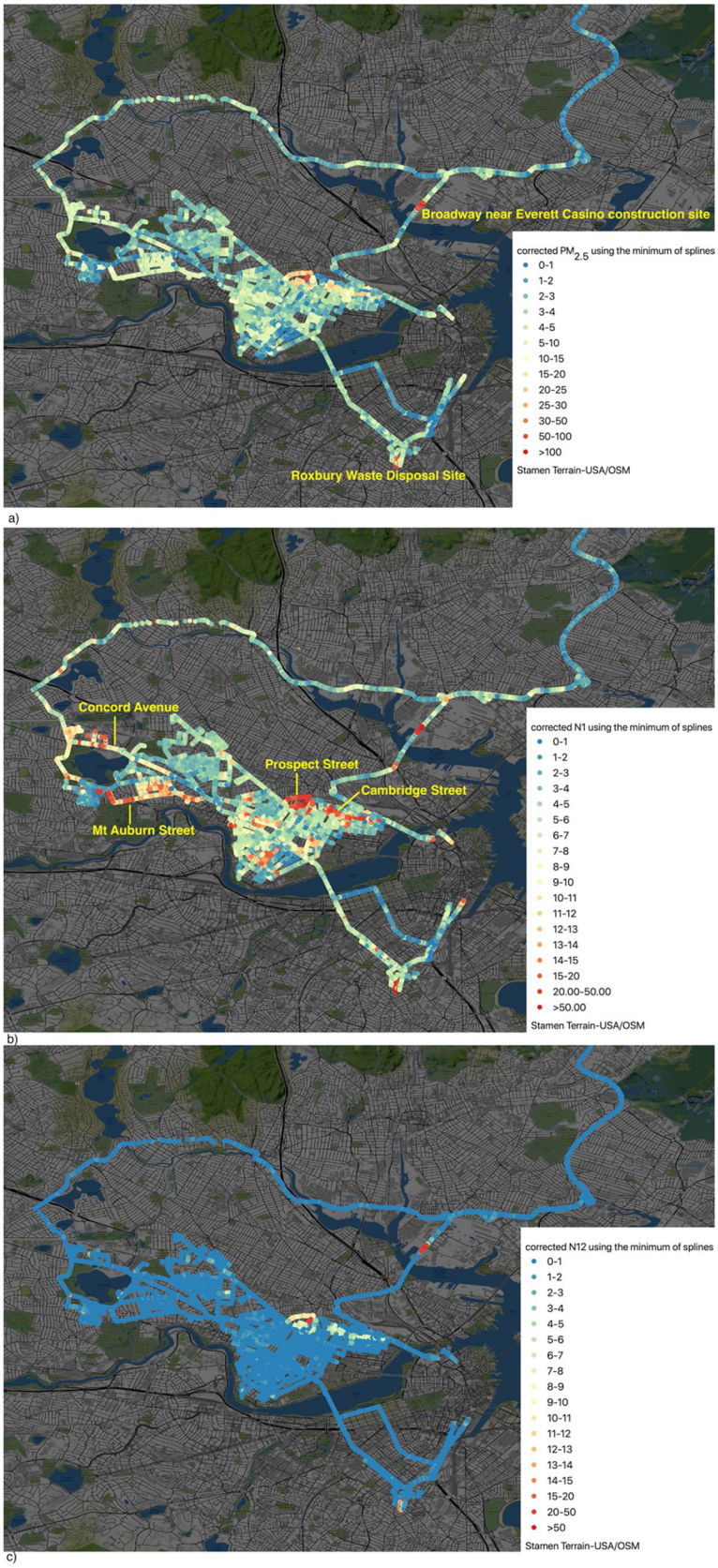

Fig. 2. a) Map of median $PM_{2.5}$ (μg/m^3) for each 30-meter road segment that the trash-trucks travelled, after the background correction had been made, b) Map of the median background-corrected number concentration (#/mL) of particles having diameters between 0.38 μm and 1 μm (N1), c) Map of the median background-corrected number concentration of particles having diameters between 1 μm and 12 μm (N12).

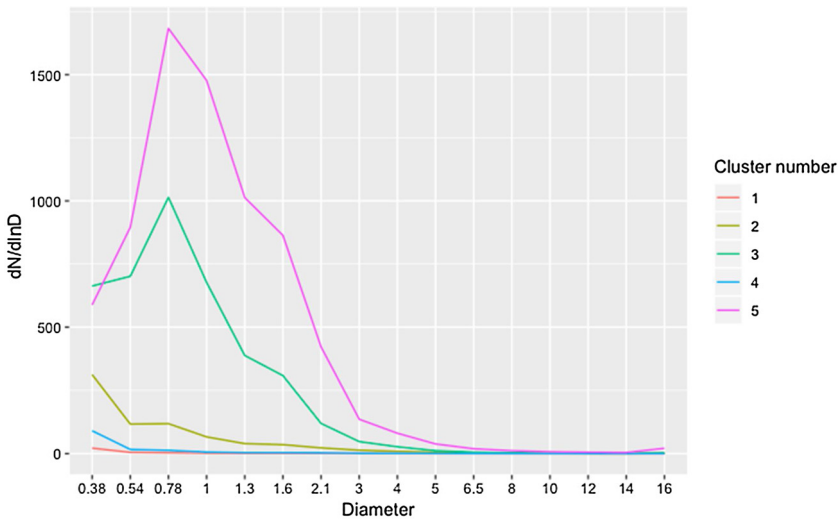

Fig. 3. Average size distribution of each cluster.

This could indicate that these observations were likely due to vehicles. The number concentration of fine particles corresponding to cluster 2 are higher than for cluster 4, suggesting that these measurements might be due to passing vehicular emission plumes.

From Fig. S7c and e, we see that clusters 3 and 5 correspond to a small number of measurements made along the sampling route. Measurements within cluster 3 were made in six different locations on seven different days (Table 1). These locations correspond to the smaller clusters of $PM_{2.5}$ hotspots depicted in Fig. 1. Cluster 5 corresponds only to observations at two locations made on two different days (Table 1).

There is a spatial overlap between measurements in cluster 3 and 5. Although the aerosol size distribution of clusters 3 and 5 appear to be similar, the number concentration of measurements corresponding to cluster 5 are higher. This could indicate that the sources contributing to both clusters are the same, but due to either temporal variations of the source characteristics, or via the mediation of the built environment, different aerosol number concentrations were observed. From Table 1, we see that both clusters 3 and 5 correspond to very low trash-truck velocities. This could indicate that the trash-trucks were stationary or idling when observations corresponding to these clusters were made.

5. Conclusions and practical implications

Our results indicate that the city of Cambridge air is relatively clean and spatially uniform ($PM_{2.5}$ is $< 10\,\mu g/m^3$). Using low-cost OPCs, we found that fine particles tended to concentrate along heavily trafficked roads, and we identified several coarse-mode particle hotspots in close proximity to likely sources, such as a waste transfer site and the Cambridge Public Works depot. We recommend a future experiment to validate these results, by co-locating the mobile low-cost monitors with at least one high-quality instrument to calibrate and/or validate the OPC measurements.

As background pollution appears to comprise a major fraction of the aerosol concentrations measured by the trash-trucks, in future deployments it is important to ensure that background pollution concentration is well characterized, probably using measurements from nearby fixed monitors located in areas away from local sources. We also note the need to record when a stopped truck is idling and when it is at a halt with the engine off, to better characterize self-emissions. We further propose the development of a standard protocol that can be used by different mobile air quality monitoring studies for other cities.

Our insights result from the deployment of low-cost monitors on trash-trucks, which run from 07:00 to 14:00 on weekdays. Thus, in future studies these measurements need to be supplemented by other scheduled or non-scheduled vehicles that operate at different hours to obtain truly representative pollution values over the region. Scheduled vehicles, such as buses, have the advantage of traversing the same street segments several times per day, whereas with unscheduled vehicles, such as taxis, we can still use a relatively small fleet (if compared with the total fleet of the city) to collect data in more randomly distributed street segments not covered by buses.

Despite the limitations of the case study in Cambridge, Massachusetts, this paper demonstrates that insights into the spatial and temporal nature of sources and their impact in the urban environment can be obtained via low-cost monitors. Importantly, this paper argues that the oft-discarded aerosol size distribution data from the Alphasense OPC-N2 within the range of detection can yield information about air pollution in urban areas that have important implications for air pollution management plans. Combining the deployment and analytical tools, we believe that mobile air quality monitoring using existing urban vehicles can be done more extensively and relatively inexpensively.

Declaration of Competing Interest

The authors declare that they have no known competing financial

Table 1
The average background-corrected PM_1, $PM_{2.5}$, PM_{10}, N1 and N12 and contribution of background aerosol to each $PM_{2.5}$ measurement for each cluster type.

Cluster number	PM_1 μg/m^3	$PM_{2.5}$ μg/m^3	PM_{10} μg/m^3	mean background $PM_{2.5}$ μg/m^3	N1 #/cm^3	N12 #/cm^3	Average Velocity (m/s)	Number of unique days	Number of observations corresponding to each cluster
1	2	3	13	3	4	0.74	1.2	27	440,988
2	41	71	316	5	79	23	1.0	27	5,039
3	215	419	817	9	323	178	0.2	7	682
4	8	11	38	5	18	2	1.2	27	128,205
5	321	804	2586	9	414	460	0.1	2	886

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Priyanka deSouza and Prashant Kumar thank the University of Surrey's 'Santander PostGraduate Mobility Award' for making this work possible. We also thank the members of the Senseable City Lab Consortium for supporting this work. Prashant Kumar acknowledges the support received through the iSCAPE (Improving Smart Control of Air Pollution in Europe) project, funded by the European Community's H2020 Programme (H2020-SC5-04-2015) under the Grant Agreement No. 689954. The work of R. Kahn is supported in part by NASA's Climate

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.scs.2020.102239.

References

Anjomshoaa, A., Duarte, F., Rennings, D., Matarazzo, T. J., deSouza, P., & Ratti, C. (2018). City scanner: Building and scheduling a mobile sensing platform for smart city services. *IEEE Internet of Things Journal, 5*, 4567–4579. https://doi.org/10.1109/JIOT.2018.2839058.

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., et al. (2017). High-resolution air pollution mapping with google street view cars: Exploiting big data. *Environmental Science & Technology, 51*, 6999–7008. https://doi.org/10.1021/acs.est.7b00891.

Beddows, D. C. S., Dall'Osto, M., & Harrison, R. M. (2009). Cluster analysis of rural, urban, and curbside atmospheric particle size data. *Environmental Science & Technology, 43*, 4694–4700. https://doi.org/10.1021/es803121t.

Brantley, H. L., Hagler, G. S. W., Kimbrough, S., Williams, R. W., Mukerjee, S., & Neas, L. M. (2013). Mobile air monitoring data processing strategies and effects on spatial air pollution trends. *Atmospheric Measurement Techniques Discussions, 6*, 10443–10480. https://doi.org/10.5194/amtd-6-10443-2013.

Bukowiecki, N., Dommen, J., Prévôt, A. S. H., Richter, R., Weingartner, E., & Baltensperger, U. (2002). A mobile pollutant measurement laboratory—Measuring gas phase and aerosol ambient concentrations with high spatial and temporal resolution. *Atmospheric Environment, 36*, 5569–5579. https://doi.org/10.1016/S1352-2310(02)00694-5.

Crilley, L. R., Shaw, M., Pound, R., Kramer, L. J., Price, R., Young, S., et al. (2018). Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmospheric Measurement Techniques,* 709–720.

Dutta, P., Aoki, P. M., Kumar, N., Mainwaring, A., Myers, C., Willett, W., et al. (2009). Common sense: Participatory urban sensing using a network of handheld air quality monitors. *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, SenSys' 09* (pp. 349–350). . https://doi.org/10.1145/1644038.1644095.

Elen, B., Peters, J., Poppel, M. V., Bleux, N., Theunis, J., Reggente, M., et al. (2013). The aeroflex: A bicycle for mobile air quality measurements. *Sensors, 13*, 221–240. https://doi.org/10.3390/s130100221.

Goel, A., & Kumar, P. (2015). Characterisation of nanoparticle emissions and exposure at traffic intersections through fast–response mobile and sequential measurements. *Atmospheric Environment, 107*, 374–390. https://doi.org/10.1016/j.atmosenv.2015.02.002.

Hagler, G. S. W., Lin, M.-Y., Khlystov, A., Baldauf, R. W., Isakov, V., Faircloth, J., et al. (2012). Field investigation of roadside vegetative and structural barrier impact on near-road ultrafine particle concentrations under a variety of wind conditions. *Science of The Total Environment, 419*, 7–15. https://doi.org/10.1016/j.scitotenv.2011.12.002.

Hankey, S., & Marshall, J. D. (2015). Land use regression models of on-road particulate air pollution (particle number, black carbon, PM2.5, particle size) using mobile monitoring. *Environmental Science & Technology, 49*, 9194–9202. https://doi.org/10.1021/acs.est.5b01209.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*, 241–254. https://doi.org/10.1007/BF02289588.

Kolb, C. E., Herndon, S. C., McManus, J. B., Shorter, J. H., Zahniser, M. S., Nelson, D. D., et al. (2004). Mobile laboratory with rapid response instruments for real-time measurements of urban and regional trace gas and particulate distributions and emission source characteristics. *Environmental Science & Technology, 38*, 5694–5703. https://doi.org/10.1021/es030718p.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., et al. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environment International, 75*, 199–205. https://doi.org/10.1016/j.envint.2014.11.019.

Langfelder, P., & Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software,* 46.

Mahajan, S., & Kumar, P. (2020). Evaluation of low-cost sensors for quantitative personal exposure monitoring. *Sustainable Cities and Society* p. 102076.

Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., et al. (2018). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment International, 116*, 286–299. https://doi.org/10.1016/j.envint.2018.04.018.

Pey, J., Rodríguez, S., Querol, X., Alastuey, A., Moreno, T., Putaud, J., et al. (2008). Events and cycles of urban aerosols in the western Mediterranean. *Atmospheric Environment, 42*, 9052–9062.

Pirjola, L., Lähde, T., Niemi, J. V., Kousa, A., Rönkkö, T., Karjalainen, P., et al. (2012). Spatial and temporal characterization of traffic emissions in urban microenvironments with a mobile laboratory. *Atmospheric Environment, 63*, 156–167. https://doi.org/10.1016/j.atmosenv.2012.09.022.

Pirjola, L., Parviainen, H., Hussein, T., Valli, A., Hämeri, K., Aaalto, P., et al. (2004). "Sniffer"—A novel tool for chasing vehicles and measuring traffic pollutants. *Atmospheric Environment, 38*, 3625–3635. https://doi.org/10.1016/j.atmosenv.2004.03.047.

Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S. W., et al. (2013). The changing paradigm of air pollution monitoring. *Environmental Science & Technology, 47*, 11369–11377. https://doi.org/10.1021/es4022602.

Sousan, S., Koehler, K., Hallett, L., & Peters, T. M. (2016). Evaluation of the Alphasense optical particle counter (OPC-N2) and the Grimm portable aerosol spectrometer (PAS-1.108). *Aerosol Science and Technology, 50*, 1352–1365. https://doi.org/10.1080/02786826.2016.1232859.

Van den Bossche, J., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J., & De Baets, B. (2015). Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset. *Atmospheric Environment, 105*, 148–161. https://doi.org/10.1016/j.atmosenv.2015.01.017.

Van Poppel, M., Peters, J., & Bleux, N. (2013). Methodology for setup and data processing of mobile air quality measurements to assess the spatial variability of concentrations in urban environments. *Environmental Pollution, Selected Papers from Urban Environmental Pollution, 2012*(183), 224–233. https://doi.org/10.1016/j.envpol.2013.02.020.

Wilson, J. G., Kingham, S., Pearce, J., & Sturman, A. P. (2005). A review of intraurban variations in particulate air pollution: Implications for epidemiological research. *Atmospheric Environment, 39*, 6444–6462. https://doi.org/10.1016/j.atmosenv.2005.07.030.