

Mobile monitoring for mapping spatial variation in urban air quality: development and validation of a methodology based on an extensive dataset

Joris Van den Bossche^{a,b,*}, Jan Peters^a, Jan Verwaeren^b, Dick Botteldooren^c, Jan Theunis^a, Bernard De Baets^b

^aVITO - Flemish Institute for Technological Research, 2400 Mol, Belgium

^bKERMIT, Dept. of Mathematical Modelling, Statistics and Bioinformatics, Faculty of Bioscience Engineering, Ghent University, 9000 Ghent, Belgium

^cFaculty of Engineering and Architecture, Ghent University, 9000 Ghent, Belgium

Abstract

Mobile monitoring is increasingly used as an additional tool to acquire air quality data at a high spatial resolution. However, given the high temporal variability of urban air quality, a limited number of mobile measurements may only represent a snapshot and not be representative. In this study, the impact of this temporal variability on the representativeness is investigated and a methodology to map urban air quality using mobile monitoring is developed and evaluated.

A large set of black carbon (BC) measurements was collected in Antwerp, Belgium, using a bicycle equipped with a portable BC monitor (micro-aethalometer). The campaign consisted of 256 and 96 runs along two fixed routes (2 and 5 km long). Large gradients over short distances and differences up to a factor of 10 in mean BC concentrations aggregated at a resolution of 20 m are observed. Mapping at such a high resolution is possible, but a lot of repeated measurements are required. After computing a trimmed mean and applying background normalisation, depending on the location 24 to 94 repeated measurement runs (median of 41) are required to map the BC concentrations at a 50 m resolution with an uncertainty of 25 %. When relaxing the uncertainty to 50 %, these numbers reduce to 5 to 11 (median of 8) runs. We conclude that mobile monitoring is a suitable approach for mapping the urban air quality at a high spatial resolution, and can provide insight into the spatial variability that would not be possible with stationary monitors. A careful set-up is needed with a sufficient number of repetitions in relation to the desired reliability and spatial resolution. Specific data processing methods such as background normalisation and event detection have to be applied.

Keywords:

Mobile monitoring, Urban air quality, Black carbon, High resolution, Mapping, Spatial variation

1. Introduction

The urban environment shows a high variability in air pollutant concentrations. The dynamics of primary emissions and secondary formation induce important differences in pollution levels in space and time. Especially for traffic-related air pollutants such as NO_x and fine particulate matter (PM), these differences can occur on a small scale (Seinfeld and Pandis, 2012; Kaur et al., 2007) and are important to take into account for mapping the pollution levels and accurate exposure assessment (Fruin et al., 2014). In contrast to traditional stationary monitoring stations, mobile platforms are able to acquire air quality data at a high spatial resolution (Wallace et al., 2009; Zwack et al., 2011b). But at the same time, due to the high temporal variability of the urban air quality and the mobile nature of the measurements, the representativeness of the mobile measurements is a major issue. These considerations are very relevant given the increasing use of mobile air quality monitoring as a solution to measure micro-scale variability at a high spatial and temporal resolution. Kuhlbusch et al. (2014), for example, mention mobile monitoring to collect highly spatially and

temporally resolved data in their recommendations for future European air quality monitoring.

The objective of this study is to develop and validate a method to map urban air quality at high spatial resolution using mobile monitoring. Based on a large experimental dataset of mobile air quality measurements, the impact of the high spatio-temporal variability is investigated, leading to practical considerations and proposed guidelines with regard to mobile monitoring campaigns to map the urban air quality. This study focuses on black carbon (BC), a constituent of fine particles and a primary particle that is emitted from incomplete combustion (EPA, 2011). BC is a good indicator of combustion-related air pollution, and exposure to BC is recognized by the World Health Organisation to be associated with cardiovascular mortality (Janssen et al., 2012).

There is a rapidly growing literature on mobile air quality measurements. A non-exhaustive overview of studies is given below. Mobile measurements are performed with different platforms, e.g. pedestrians (Zwack et al., 2011a), bicycles (Berghmans et al., 2009; Boogaard et al., 2009; Peters et al., 2013; Sullivan and Pryor, 2014), trams (Hagemann et al., 2014; Hasenfratz et al., 2014) and cars (Westerdahl et al., 2005; Hu et al., 2012; Hudda et al., 2014). Mobile measurements are used for a range of different purposes, e.g. to assess personal exposure by

*Corresponding author

Email address: jorisvandenbossche@gmail.com (Joris Van den Bossche)

equipping the study object with a portable monitor (Berghmans et al., 2009; Boogaard et al., 2009; Dons et al., 2011; Spira-Cohen et al., 2010), to assess the exposure in different modes of transport (Kaur et al., 2005; Int Panis et al., 2010; Kingham et al., 2013), to study spatial variation in air pollution (Weijers et al., 2004; Zwack et al., 2011c; MacNaughton et al., 2014), to investigate seasonal and regional variation (Bukowiecki et al., 2003), to study spatio-temporal correlation with noise (Weber, 2009), or to develop and validate air quality models (Zwack et al., 2011a; Merbitz et al., 2012). Other studies address the potential of using mobile measurements to construct air pollution maps at a high spatial resolution (e.g. Hagler et al., 2010; Choi et al., 2012; Hudda et al., 2014; Peters et al., 2014). Mobile monitoring techniques also receive an increasing attention for participatory sensing and crowd-sourcing methods. Already in 2006, Burke et al. mentioned the theoretical potential of participatory sensing to investigate the relationship between air quality and public health. Volunteers can systematically collect targeted data sets, or data are collected opportunistically during (repeated) daily activities or trips. This potential is increasingly acknowledged, e.g. by scientists of the US Environmental Protection Agency (Snyder et al., 2013). Some participatory or community-based sensing projects designed new measurement devices with low cost sensors for large-scale and mobile deployment (see e.g. Dutta et al., 2009; Elen et al., 2012).

Some critical issues related to mobile monitoring of air quality have to be acknowledged. These issues arise from the combination of the high temporal variability of air quality and the mobile nature of the measurements. The spatial variation in (local) concentrations in a city is mainly linked to differences in traffic density, street topology (e.g. street canyons) and the distance to sources. There is also a strong temporal variability in pollutant concentrations. This variability and its cause differ from pollutant to pollutant, but we can distinguish different levels of temporal variability relevant for the analysis in this paper: day-to-day, within-day and micro-scale variability. Day-to-day variability is mainly caused by the meteorological conditions and urban background fluctuations. Within-day variability is additionally related to the traffic dynamics, with significant peak concentrations at the morning and evening rush hours. Finally, at a small temporal scale (i.e. of the order of seconds), there are also large fluctuations in the measured pollutant concentration due to emission events, typically the passing of a vehicle. Mobile measurements, which can be seen as time-space series of data, result in only a short snapshot at a certain location in time. Therefore, given the temporal variability, the exact timing of the measurement does not necessarily allow to generalize the measurements as a typical value for that location.

When mobile monitoring is used to derive high resolution maps that are representative in time, large amounts of data are required to represent the range of possible meteorological and traffic conditions (Padró-Martínez et al., 2012) and data aggregation has to be performed. Peters et al. (2013) indicate the need for repeated measurements to map local pollution patterns. In the study of Peters et al. (2014) it is demonstrated that the variability in urban BC and ultra-fine particles (UFP) concen-

trations can be mapped at a spatial resolution of 10 m. However, they used up to 256 repeated measurement runs, which is not always practically feasible. Van Poppel et al. (2013) applied background correction to reduce the number of repetitions required to obtain representative results of the spatial variability of pollutants at different micro-environments in a city. The results are also sensitive to the chosen data processing approaches, as stressed by Brantley et al. (2014), who explored the influence of different techniques for local emission event detection, background estimation and averaging.

Therefore, suitable monitoring strategies and analysis methods have to be defined that take spatio-temporal representativeness into account. A monitoring campaign was set up to collect an extensive dataset with high spatio-temporal resolution along predefined routes in a period of 4 weeks. For more details about the measurements, see Peters et al. (2014), who used the same dataset to describe the spatial and temporal variation in detail. The large number of repeated measurement runs performed in this study allows to address the following research questions about obtaining a representative and consistent image of the spatial variability of the urban air quality: (1) What is the impact of events and the used spatial resolution? (2) How representative is this image in time? (3) How many runs are required to obtain this image of the spatial variation? (4) Can we reduce the required number of runs by reducing the variability through filtering for events or by applying background normalisation?

2. Material and methods

2.1. Mobile monitoring campaign

An extensive monitoring campaign¹ using a mobile monitoring platform, the Aeroflex (Elen et al., 2013), was set up in Antwerp (51°12' N, 4°26' E, medium-sized city of 480,000 inhabitants, 985 inhabitants km²), Belgium, during 4 weeks in February-March 2012. The Aeroflex is a bicycle equipped with compact air quality measurement devices to monitor at a high temporal resolution (up to one second). Each measurement is automatically linked to its geographical location and time of acquisition using GPS (GlobalSat BU-353) and internet time. The instrumentation used was the micro-aethalometer (MicroAeth Model AE51, Aethlabs) to monitor BC. This instrument measures the concentration of optically absorbing aerosol particles (equivalent black carbon (EBC, in $\mu\text{g m}^{-3}$) using a mass-specific absorption cross-section (MAC) of 12.47 m²g⁻¹ at 880 nm (Petzold et al., 2013)) by the rate of change in light attenuation of a particle spot on a filter through which air is pumped. The filter ticket was changed at the start of each measurement day, the inlet flow rate was set at 150 mL min⁻¹ and measurements were made at a temporal resolution of one second. The response time is negligible. The optical measurement of the filter is instantaneous and the lag in time due to pumping of the air through the tube is less than a second for a flow rate of 150 mL min⁻¹. Measuring BC with the micro-aethalometer at

¹The dataset is available upon request.

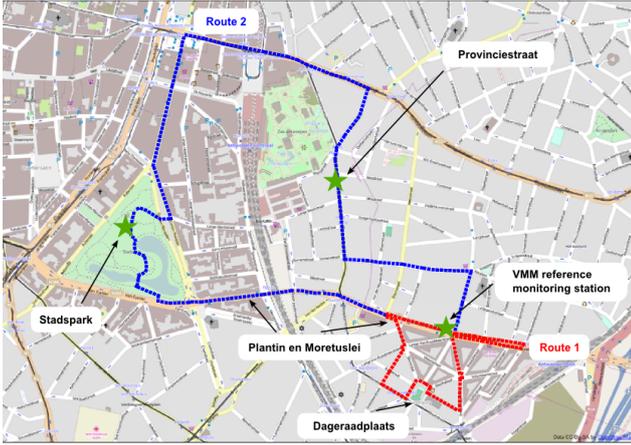


Figure 1: Overview of the bicycle routes in the city of Antwerp, Belgium: route 1 (red) and route 2 (blue). The green stars indicate the locations of the stationary measurements.

one second time resolution is challenging due to the occurrence of signal noise, especially at low concentrations. Therefore, a noise reduction algorithm was used (Hagler et al., 2011) (see Section 2.3).

A controlled monitoring set-up with fixed routes and time slots was used to guarantee sufficient temporal and spatial coverage to address the research questions. The mobile measurements were made along two fixed routes (about 2 and 5 km long, see Figure 1). The routes were designed before the start of the monitoring campaign in order to include streets and places with different configuration and traffic intensity. The starting point of both routes was a reference air quality monitoring station of the Flemish Environmental Agency (VMM, station 42R801) in the Plantin en Moretuslei (a busy arterial road). Route 1 was an approximately 2 km long loop at the Southern side of the Plantin en Moretuslei, route 2 was a loop of 5 km at the Northern side of this road. The average duration to complete one run was 10 and 26 min for route 1 and 2, respectively. The major part of the routes was located in residential and commercial areas. Route 1 passed by a low-traffic square (Dageraadplaats), whereas an urban green background area (Stadspark) was crossed by route 2. An overview of the most important streets of both routes is given in Table 1.

The campaign consisted of 256 runs along route 1 and 96 runs along route 2, spread over 11 days in a 4-week period between 2012-02-13 and 2012-03-08. The runs always occurred between 7 am and 1 pm on weekdays, covering the morning rush, and were performed with two equipped bikes, monitoring the two routes simultaneously. This resulted in the collection of 92 hours of mobile measurements.

2.2. Complementary stationary measurements

Stationary measurements at 3 locations accompanied the mobile campaign. BC was measured continuously during the 4 week period of the mobile campaign at the Stadspark (urban green), Provinciestraat (street canyon) and Plantin en Moretuslei (two times two lanes entrance road). See Figure 1 for the exact locations. These stationary BC measurements

were performed with micro-aethalometers with a flow rate of 50 mL min^{-1} and a 5 min measurement rate.

Furthermore, measurements from the VMM reference monitoring station at the Plantin en Moretuslei were also available. At this station, BC is measured with Multi-Angle Absorption Photometry (MAAP, Petzold and Schonlinner, 2004; Petzold et al., 2005) at 1 min resolution.

2.3. Data validation and quality control

Additional stationary measurements were made with all the micro-aethalometers (from mobile and stationary monitoring) at the VMM reference monitoring station at the Plantin en Moretuslei (Figure 1) for data quality control. The individual micro-aethalometers were put together for in total 20 hours to identify possible bias between instruments. The micro-aethalometer measurements were compared between the different instruments, and also to the measurements of the MAAP instrument at the monitoring station. Noise reduction was realised using the Optimized Noise-reduction Averaging algorithm (ONA algorithm from Hagler et al., 2011), with an attenuation threshold of 0.05. Rescaling parameters were derived from the comparison of the individual micro-aethalometers with the MAAP instrument from the monitoring station. Before determining the rescaling parameters, filter loading effects were accounted for by an additional correction algorithm according to Virkkula et al. (2007):

$$BC_{corrected} = (1 + k \cdot ATN)BC_{uncorrected} \quad (1)$$

where ATN is the attenuation. The parameter k was estimated by fitting Eq.(1) using non-linear least squares, for each micro-aethalometer independently. As the values did not differ significantly, an average value for k of 0.009 for all micro-aethalometers was used. For the fitting, the MAAP data were used as $BC_{corrected}$, as this device is considered to be less influenced by the loading effect compared with the micro-aethalometer (Petzold et al., 2005). Based on 30 min means, regressions between the corrected BC concentrations of the micro-aethalometers and the MAAP with slopes between 0.83 and 1.13 (R^2 of 0.97-0.98) were obtained. These slopes were used together with Eq.(1) to correct the mobile micro-aethalometer data. After the correction, the micro-aethalometer data showed a mean absolute error of $0.3 \mu\text{g m}^{-3}$ compared to the MAAP based on 30 min means (leading to a relative error percentage of 7 % for the calibration period). When looking at the agreement between the micro-aethalometers themselves (comparing all others to one), the comparison results in an average error of $0.6 \mu\text{g m}^{-3}$ at 1 second resolution and $0.2 \mu\text{g m}^{-3}$ at 1 min resolution (15 % and 5 % error, respectively).

The spatio-temporal nature of mobile data requires several data cleansing and processing operations involving the interpolation and correction of GPS data. The GPS signal is sometimes lost or inaccurate during the monitoring. Because the GPS data were mostly missing for only short periods (2.1 s on average), a linear interpolation was used to insert these data points. When visualized on a map, GPS data were often slightly off track. Therefore, the GPS data were projected on the streets based

Table 1: Description of some street characteristics of a selection of streets along monitoring route 1 (R1) and route 2 (R2).

	Street name	Abbrev.	Description	Speed limit (km/h)	Configuration		Traffic density ^a (vehicles per day)	Length ^b (m)
					Nr lanes	Biking lane		
R1	Plantin en Moretuslei	PM	Entrance road	70	2	yes	43,381	890
	Kleinebeerstraat	KB	Residential	50	1	no	1,269	105
	Lange Altaarstraat	LA	Residential	50	1	no	5,585	120
	Wolfstraat	WO	Residential	30	1	no	5,680	145
	Dageraadplaats	DP	Public square, very low traffic	/	/	no	NA (~0)	170
R2	Plantin en Moretuslei	PM	Entrance road	70	2	yes	43,381	995
	Stadspark	SP	Urban green	/	/	yes	NA (~0)	565
	Quellinstraat	QU	Tertiary	50	1	no	9,590	315
	F. Rooseveltplaats	FR	Square, bus stops	50		partly	21,422	70
	Carnotstraat	CA	Entrance road	50	1	yes	22,963	430
	Provinciestraat	PR	Tertiary (street canyon)	50	1	no	12,174	690
	Bleekhofstraat	BL	Residential	30, 50	1	no	4081	490

^a Modelled traffic density from the Traffic Centre Flanders based on traffic count data (includes both light and heavy traffic).

^b Length of part of the street that is included in the route.

on shortest distance. Only the streets that are included in the routes were used for this projection. Contextual information about travel direction and street orientation were further used to select the correct street to project the data upon.

The accuracy of the GPS location of the measurements is related to the quality of the GPS signal (GPS accuracy and projection accuracy) and the synchronization of the measurement device. The uncertainty of the GPS location is difficult to assess, as we have no reference to compare with. The GPS device (GlobalSat BU-353) specifies an accuracy of the horizontal position of 10 m. At certain locations in urban environments, this can be larger due to reflection of signals by high buildings (observed from plotting the geographical data). The accuracy of the perpendicular projection on the streets could not be quantified exactly. The overall accuracy is probably around 10 m. The synchronization of the micro-aethalometer was checked and adjusted to the GPS time every day and the deviation was never more than 1 s. Given the average driving speed of 3.2 m s^{-1} , this will not have an important impact on the spatial accuracy.

2.4. Data analysis and processing methods

The term *run* is used to denote the completion of an entire route once, and this is repeated at different days and/or different times of one day. During each run, a dataset of point measurements is collected. When the mobile platform passes by at a certain location during a run, this is called a *passage*. In this way, the monitoring campaign consists of a number of passages at the different locations of a route.

The analyses in this study are performed at different spatial scales by spatially aggregating the original data at different levels. The spatial entities varied from the entire routes, the different individual streets of a route, to segments of a certain length within a street. The street segments were constructed at different spatial resolutions (segment length varying from 5, 10, 20, ... to 100 m). The mobile data were spatially aggregated by allocating the measurements to one of the spatial entities based on the projected GPS data. The measurements allocated to a specific spatial entity are then aggregated by computing the arith-

metic mean. For all analyses, this aggregation was performed per passage. As such, for each individual run, an average BC concentration was calculated for each spatial entity, resulting in *passage means*.

For the highest spatial resolution, i.e. the segments (represented by their midpoints), concentration profiles were constructed to explore the spatial variability at this level. On these figures, the mean concentration and standard error on the mean, calculated for each segment based on the passage means, are plotted.

Different methods are explored to reduce the impact of extreme values on the spatially aggregated result. Firstly, two other averaging statistics are used besides the arithmetic mean: the median and the α -trimmed mean. The α -trimmed mean is computed by removing the α largest and α smallest values and computing the arithmetic mean of the remaining values. This is a statistical measure of central tendency that is less influenced by the effects of the tails of a distribution. As the air quality data in this study are positively skewed, this will result in a biased estimator of the population mean. Indeed, this estimator will systematically underestimate the population mean. However, due to the reduced tail effects we believe this is an appropriate statistic. In this paper, a trimmed mean resulting from the removal of the lowest and highest value is adopted. This corresponds to $\alpha = 0.5\%$ for samples of up to 200 runs. Unless specified otherwise, trimming was always performed on the passage means and for each spatial entity separately (the passage mean itself is not calculated with trimming). Secondly, an emission event filtering method is used: the running coefficient of variation (COV) method from Hagler et al. (2012). For this method, a running 5 s standard deviation of the BC concentrations is calculated and divided by the mean concentration of the entire sampling period. The 99th percentile of the calculated COV is used as a threshold and all data points with a COV above this threshold are removed along with the data points 2 s before and after (Brantley et al., 2014).

Background concentrations are obtained from the stationary

measurements at the centre of an urban green (the Stadspark) without sources within a direct vicinity of 50 m. These measurements can therefore be considered as urban background concentrations (Hoek et al., 2002). These measurements are available at 5 min resolution and are processed in two ways: the hourly median value and the moving mean (using a Gaussian window with standard deviation of 30 min). The mobile measurement data are normalised with a background normalisation method based on these values to minimize the influence of meteorological day-to-day variations in the urban air quality. This normalisation is performed through a combination of the additive and multiplicative method as used in Dons et al. (2012). The normalised values are calculated as:

$$BC_{norm,i} = BC_i - BC_{bg,i} + \overline{BC_{bg}} \quad (\text{additive}) \quad (2a)$$

$$BC_{norm,i} = BC_i / BC_{bg,i} \cdot \overline{BC_{bg}} \quad (\text{multiplicative}) \quad (2b)$$

with BC_i the original BC measurement at time i , $BC_{bg,i}$ the background measurement at time i , and $\overline{BC_{bg}}$ the mean background concentration for the full period. The additive method is applied to high concentrations (measurement is greater than the background value) and the multiplicative method to lower concentrations (measurement is smaller than the background value), see Dons et al. (2012) for more details.

Analyses were performed using R (R Core Team, 2014), pandas (PyData Development Team, 2014) and QGIS (QGIS Development Team, 2014).

2.5. Data experiment

A data experiment was conducted to investigate the representativeness of the measurements, i.e. to address the research question of how many mobile runs are required to obtain a representative estimate of the air quality in a certain area. In this study, a representative estimate means that the average BC concentration of a *subsample* of runs for a spatial entity is within a certain percentage, the *deviation* (e.g. $\pm 25\%$), from the overall average value obtained from the entire mobile measurement campaign for that spatial entity. In the absence of an integrated average of continuous measurements at each location, this overall average is considered as a good approximation of the pollutant concentration to be measured. The representativeness is thus constrained by the space and time of the data collection, which was biased towards daytime hours during one month of the year in an urban environment, which is not necessarily representative for other periods or locations.

The data experiment consisted of repeatedly taking subsamples through the generation of random combinations of all runs. Starting from the measurements of a single run and then cumulatively adding measurements from randomly selected runs (random selection with replacement), the average concentration was calculated for each combination of increasing size for each spatial entity. This way, the average pollutant concentration could be evaluated in function of the increasing number of runs included in the sample and compared with the overall average value to see after how many sampled runs the average converged to the overall average. Convergence is obtained when the average of the sampled runs deviates less than the deviation

percentage from the overall average, and does so consistently when adding new runs. For one experiment, the procedure explained above is repeated a high number of times (1000 iterations, where each iteration is one random combination of the runs) to minimize the influence of random effects. This data experiment is based on the work of Peters et al. (2013), but differs from it in the application of a random selection with replacement instead of without replacement. When using random selection without replacement (each run can only be selected once), the average would converge exactly to the overall average when more runs are added and convergence would always be obtained. To prevent this, random selection with replacement is used. The data experiment was repeated for different set-ups: at different levels of spatial aggregation, with the arithmetic and trimmed mean, and with and without background normalisation.

3. Results

3.1. Occurrence and impact of events

Mobile air quality measurements as described in this study show typical characteristics due to the high temporal resolution of the data collection and the varying spatial context. A large variability is observed in the measurements at one second resolution of one run (Figure 2a). Measurements with a low variability (e.g. Stadspark) are followed by measurements with a higher variability in other streets. In the regions with a high variability, sharp peaks are seen due to the proximity to sources, leading to large differences in BC levels within a couple of seconds. Events, typically a closely passing car or bus, cycling in the emission plume right behind a vehicle, idling vehicles or local traffic congestion, are causing these high BC levels. These events can occur systematically at certain places or can be more accidental. We will refer to these last events as random events. The occurrence of events leads to a distribution of the data as shown in Figure 2b. A part of the BC range is cut off on this figure: around 0.5% of the data at one second resolution fall outside the range of the figure ($0-50 \mu\text{g m}^{-3}$). This highest 0.5% of the measurement data has a moderate impact on the overall mean of all runs: discarding these values decreases the overall mean with 7%. However, the following two examples will illustrate that extreme values can have a much larger impact at a smaller spatial resolution or when having fewer repetitions.

The first example is the distribution of the mean BC concentration per passage at the street level in the Grotebeerstraat (Figure 3). Some passages have a clearly higher mean than most others, with a maximal value of $33.1 \mu\text{g m}^{-3}$. Such an extreme value can have a big impact on the resulting average street level concentration of the Grotebeerstraat. On all 256 runs, this influence is limited (a mean of $3.6 \mu\text{g m}^{-3}$ for all runs vs. $3.5 \mu\text{g m}^{-3}$ without the extreme). However, in the case of a much smaller sample size, e.g. 20 randomly selected runs, this impact is very significant: 5.2 vs $3.5 \mu\text{g m}^{-3}$ with and without the extreme, which is an increase of around 50% due to a single passage.

The second example is shown in Figure 4, which depicts the concentration profile in the Provinciestraat at 20 m resolution.

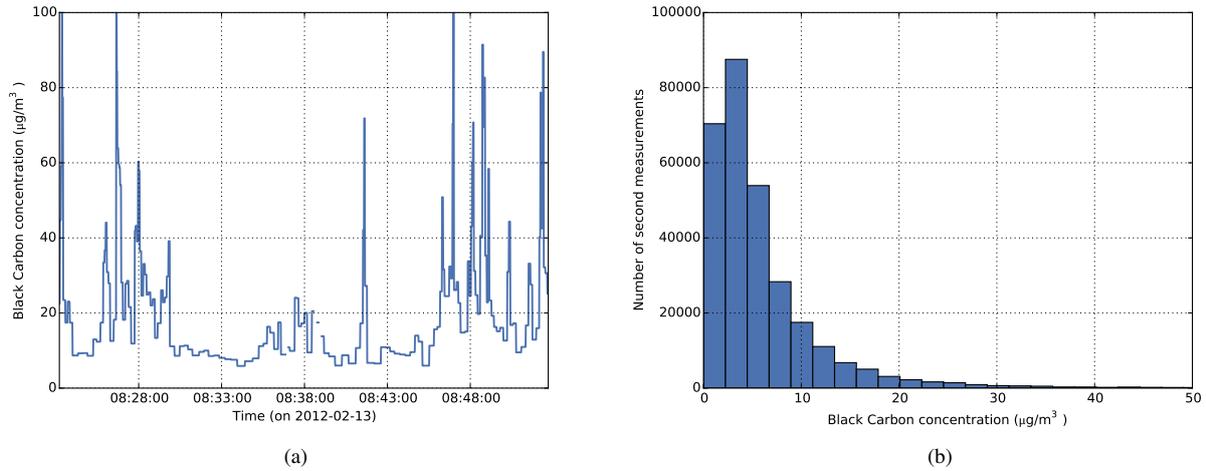


Figure 2: (a) Example run of mobile measurements at one second resolution for route 2 on 2012-02-13 and (b) the distribution of all mobile measurement data for routes 1 and 2 combined.

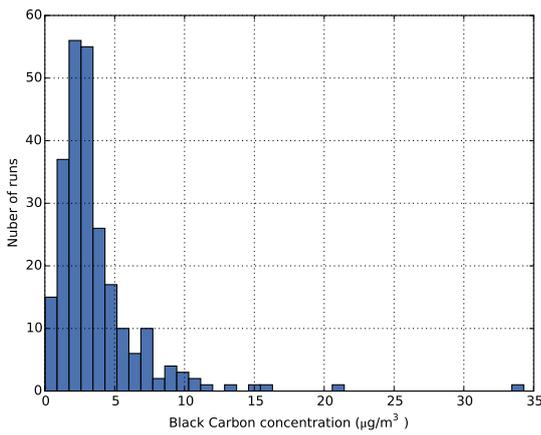


Figure 3: Distribution of the passage means for the Grotebeerstraat (route 1, a total of 249 runs included).

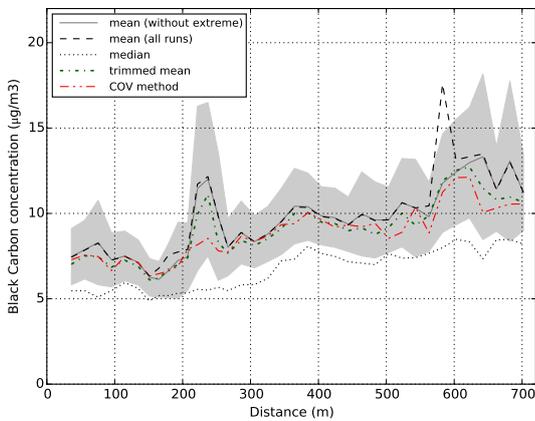


Figure 4: Concentration profile of the Provinciestraat: BC concentration levels at 20m resolution using different processing methods: mean, median, trimmed mean and COV method.

The mean concentration profile is shown for all 96 runs (dashed line), and there is a peak in the BC concentration at 225 and at 580 m. Looking at the peak at 580 m, it appears that this peak can largely be attributed to a single measurement run that includes extreme measurements. Excluding this particular run smooths the BC profile and lowers the average BC concentration at that place from $17.6 \mu\text{g m}^{-3}$ to $11.8 \mu\text{g m}^{-3}$ (a reduction of 49%). If the total number of runs is 20 instead of 96, the impact is even larger: 43.3 vs. $12.1 \mu\text{g m}^{-3}$ for 20 randomly selected runs with and without the extreme, which is an increase of more than 250% due to a single passage. The previous examples show that a single run can have a large impact on the result, especially at a smaller spatial resolution or when having less repetitions.

Next, different aggregation functions besides the arithmetic mean are explored to decrease the impact of these extreme measurements, while retaining the characteristic concentration profile: median, trimmed mean and the rolling COV method (see Section 2.4). The results of these methods can be seen in Figure 4 for the Provinciestraat. With all the different statistics, the concentration profile no longer exhibits the peak at 580 m, while following roughly the same spatial trend. However, the median is much lower than the arithmetic mean. This large difference between the mean and the median is due to the high positive skewness of the distribution. Both the trimmed mean and the COV method result in only slightly lower values than the profile based on the arithmetic mean. But the COV method misses the peak at 220 m, while with the trimmed mean this peak is still present. The application of the trimmed mean is further examined in Sections 3.3 and 3.5.

3.2. Spatial variation: concentration profiles

Observations on the spatial variability within the urban environment based on this dataset of mobile measurements are extensively reported in Peters et al. (2014). At the street level, significant differences in mean levels were found (up to a factor

of 2) for the aggregated data of all passage means per street. A clear correspondence with traffic and street topology was observed. Lowest concentrations were found at places with low traffic intensities as Dageraadplaats and Stadspark. Highest concentrations were found in the Plantin en Moretuslei and Provinciestraat. The Plantin en Moretuslei has the highest traffic intensities, while the Provinciestraat has lower traffic intensities, but is a street canyon and has no separate biking lane.

Concentration profiles were constructed for certain sections of the routes to study the spatial variability at a high spatial resolution. The result for the 20 m resolution can be seen in Figures 5 and 6. These figures indicate that large differences in air quality exist within a single street. For example, in the Plantin en Moretuslei (Figure 5a) there is a very large peak at the tunnel (up to $19 \mu\text{g m}^{-3}$), and a second large peak at the intersection with the Provinciestraat (up to $17 \mu\text{g m}^{-3}$), where the values are two to three times higher than the BC concentrations in the rest of the street (5 to $10 \mu\text{g m}^{-3}$). Also the intersection of the Plantin en Moretuslei and the Montensstraat shows an increased concentration, which is located near the VMM monitoring station (Figure 5b). The locations with the lowest BC concentrations are the Stadspark and Dageraadplaats (Figure 6, 2 - $3 \mu\text{g m}^{-3}$). BC concentrations also show a low variability at these locations. Figure 6a shows a gradual increase throughout the Korte Altaarstraat to Wolfstraat from one end to the other, which is a part of route 1 from the Dageraadplaats to Plantin en Moretuslei that exhibits increasing traffic and proximity to a major traffic road. The Provinciestraat also shows a gentle gradient between two parts of the street (Figure 4). In contrast, a very steep gradient is observed between the Stadspark and the Quinten Matsijslei (a 2 - 3 times increase over a distance of 50 m, Figure 6b). In conclusion, large gradients over short distances between different urban micro-environments can be observed from the mobile data aggregated at resolution of 20 m. Small-scale differences up to a factor of 10 are found (in comparison to the factor 2 for street level averages). Urban greens and traffic-free squares are clearly distinguished by low BC concentrations from the surrounding streets.

3.3. Effect of spatial resolution on reproducibility

To investigate the reproducibility of the observed spatial variation in the concentration profiles and the influence of the spatial resolution, we look at the consistency of the spatial patterns between independent subsamples of different sizes. Two random but non-overlapping subsamples of all runs are taken. Based on each subsample a concentration profile is calculated for the whole route as the average concentration for each segment using the arithmetic or trimmed mean. The correlation between the two profiles is determined as the R^2 of a linear regression. This analysis is done for different resolutions (segment lengths between 5 and 100 m) and for different sample sizes (10 , 20 , 50 and 100 runs), and for each of these combinations this is repeated 1000 times to obtain an average R^2 (Figure 7). The average R^2 increases with increasing sample size; the concentration profiles are more similar when comparing larger samples. The R^2 also increases with increasing segment length, although this increase is large for small segments

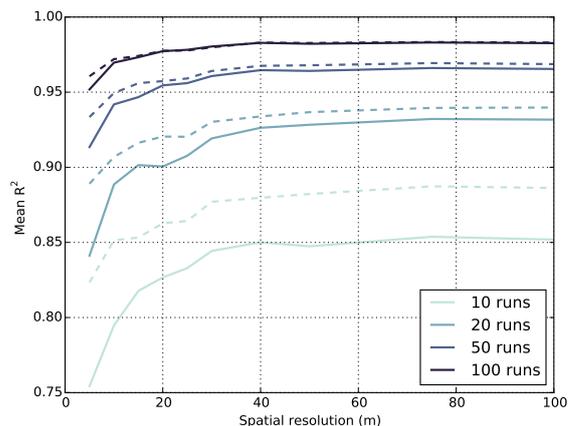


Figure 7: The R^2 of the comparison of two random but non-overlapping subsamples in function of the spatial resolution and for different sizes of the subsamples. The line indicates the mean R^2 of 1000 repetitions for each combination of resolution and sample size. Full lines: arithmetic mean, dashed lines: trimmed mean.

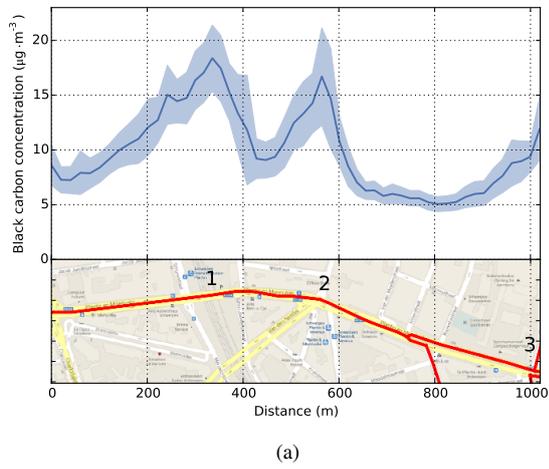
up to 20 m but almost completely levels off for segments of 40 to 50 m and larger. Using a trimmed mean instead of the arithmetic mean in the construction of the concentration profiles increases the R^2 , and thus makes the profiles more reproducible. This increase is most clear for higher resolutions and for smaller sample sizes, while for a sample size of 100 runs it makes almost no difference.

3.4. Comparison of stationary and mobile measurements

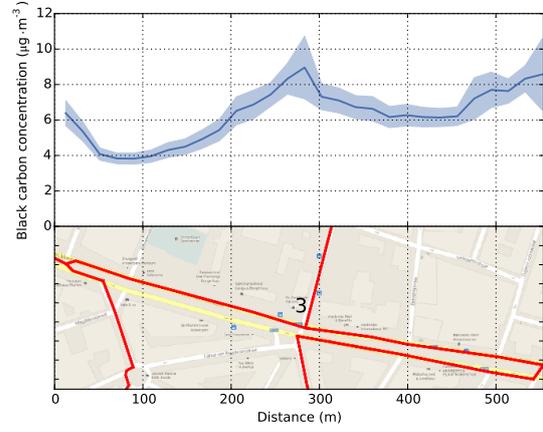
Stationary and mobile measurements of BC are compared at three locations (Provinciestraat, Stadspark and Plantin en Moretuslei). For the Plantin en Moretuslei, only measurements within a range of 100 m around the stationary monitor are used because the intra-street variability is much higher than at the other two locations. Stationary measurements are available as 5 min means for each location. Mobile measurements are averaged per passage in the particular street (passage mean).

First, mobile measurements were compared directly with the stationary measurements by matching each passage mean with the nearest 5 min stationary measurement. The exact time period of the measurements will not coincide using this method, and the results also indicate that there is too much variation on the individual passages to have a clear relationship between the individual mobile passages and stationary measurements ($R^2 = 0.21$, 0.40 and 0.14 for Provinciestraat, Stadspark and Plantin en Moretuslei, respectively).

When the measurements are averaged per day using the arithmetic mean, the relationship between the mobile and stationary measurements becomes stronger for all locations ($R^2 = 0.58$, 0.96 and 0.63 , respectively, see Figure 8). For the three locations combined, this yields an R^2 of 0.69 . But when we remove the one extreme value for the Provinciestraat as discussed in the second example in Section 3.1, the R^2 becomes 0.82 . The averages of the stationary measurements are calculated based on only those measurements made during the same time frame in

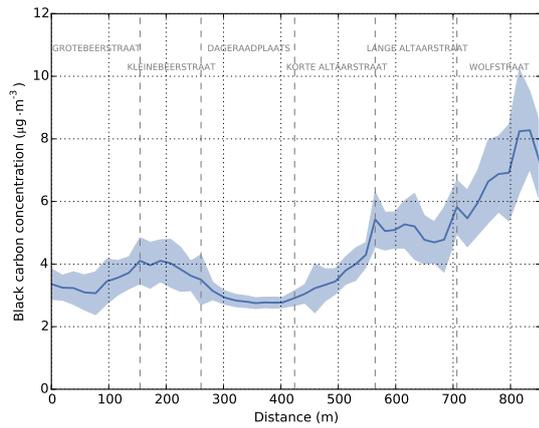


(a)

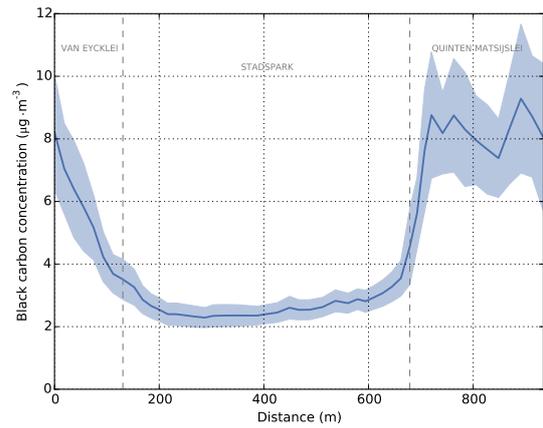


(b)

Figure 5: BC concentration profiles for the Plantin en Moretuslei (a) based on the measurements of route 2 and (b) of route 1 (at 20 m resolution, value is the mean of passage means). Shaded area depicts two times the standard error on the mean. The annotations indicate (1) the tunnel, (2) the intersection with the Provinciestraat and (3) the intersection with the Montensstraat and Wolfstraat.



(a)



(b)

Figure 6: BC concentration profiles for (a) the Dageraadplaats and (b) Stadspark (at 20 m resolution, value is the mean of passage means). Shaded area depicts two times the standard error on the mean.

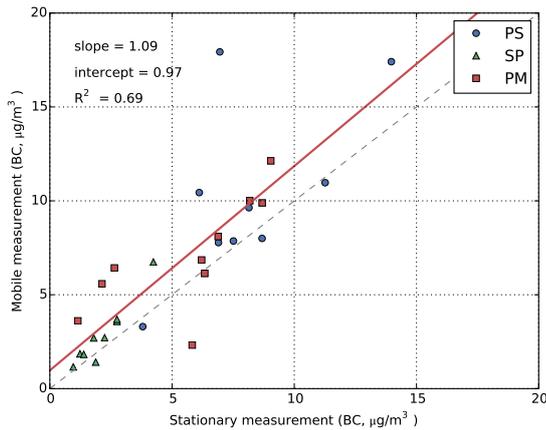


Figure 8: Comparison between mobile and stationary BC measurements for the three locations (for abbreviations, see Table 1): comparison of averages per day of passage means and stationary measurements (during the mobile time frame, i.e. 7 am to 1 pm). Grey dashed line: 1:1 line, red line: regression line.

which mobile measurements were conducted (7 am to 1 pm). In this way, the influence of measuring during other time periods, e.g. lower night-time concentrations, which are not measured in the mobile campaign, on the mean value of the stationary measurements is cancelled out. The best correlation in both cases is found for the Stadspark, where there is less variation caused by direct sources. The combined R^2 of 0.82 indicates that the variations in time (day-to-day variations) are also captured by the mobile measurements, although the mobile measurements result in higher concentrations (positive intercept).

The same measurements are used to create a boxplot as an overall comparison for all days for the three locations (Figure 9). The street level averages from the mobile measurements are systematically higher than the average of stationary measurements in the same street, but this figure also indicates that the relative spatial differences between the three locations are comparable.

3.5. Number of repetitions based on a data experiment

A data experiment was performed as explained in the methods Section 2.5 in order to investigate how many mobile runs are required to obtain a representative estimate of the air quality. Firstly, the entire route is taken as the spatial level; later on the analysis is refined for smaller spatial resolutions. Figure 10 shows the result of one typical data experiment (based on all data of route 1 with a total of 256 runs). The evolution of the mean for all 1000 iterations is depicted on the left. For each of these iterations, the number of runs required for convergence was determined and plotted in a density plot (Figure 10, right), using a deviation of 25 % in this example. This number is at most 17 in 95% of the iterations. When narrowing the deviation to 20 %, this number increases to 29 runs, and it drops to 12 runs when allowing a deviation of 30 %.

This data experiment is repeated for each street of route 1 separately. Due to different pollution dynamics, different results are expected in the different streets. The results are sum-

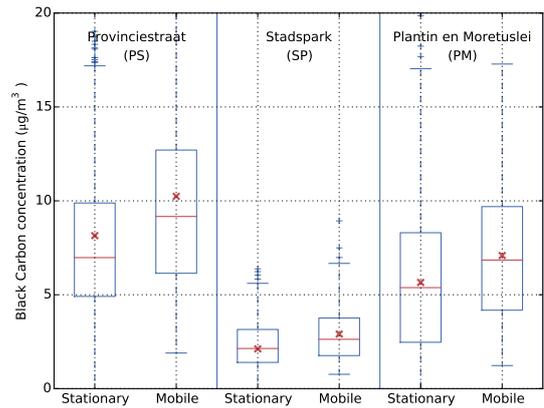


Figure 9: Boxplot of the stationary and mobile measurements for the three locations: comparison of the passage means (mobile data) and 5 min means (stationary data) of all days. For the stationary measurements, only the measurements during the mobile time frame (7 am to 1 pm, and also on the same days as the mobile measurements were conducted) are used.

marised in Table 2 for some streets of route 1 for a deviation of 25 and 50 %. In general, the required number of repetitions for the separate streets is higher than for the overall mean. This is caused by the increasing variation between the runs when going to a higher spatial resolution because the sample becomes smaller per spatial entity. The mean values at Dageraadplaats and Plantin en Moretuslei converge more quickly. At the traffic-free square of Dageraadplaats, there is less variation between different runs, which explains the lower number of repetitions required to obtain convergence. The Plantin en Moretuslei is a busy street and a large variability is expected, but it is also the longest street, levelling off this variability by a larger temporal smoothing. On the other hand, the mean value at the Wolfstraat and Kleinebeerstraat converges slowly. These streets are both shorter and the Wolfstraat ends at the crossing with a busy street. This illustrates a shortcoming of using street sections with varying length as the spatial aggregation level. Therefore, the same analysis is performed with segments at 20 m and 50 m resolution. There is a large variation in the number of repetitions required for the different segments. In Table 2, this variation is expressed as the 10th, 50th and 90th percentile of the required number of repetitions for the different segments. In 10 % of the segments convergence occurs after less than 33 repetitions, while in 10 % of the segments this is only after more than 141 runs at a spatial resolution of 50 m. In general the required number of repetitions is only slightly lower at 50 m resolution compared with the higher resolution of 20 m.

Next, a similar data experiment is carried out with a trimmed mean instead of the arithmetic mean (Table 2). On the whole dataset, this does not have a positive effect, but when looking at specific streets, this can have an impact. For example, for the Grotebeerstraat (street with an outlier), the required number of repetitions decreases from 66 to 50 and for the Kleinebeerstraat from 82 to 68 for a deviation of 25 %. Using a deviation of 50 % the relative decreases in the required number of repetitions are

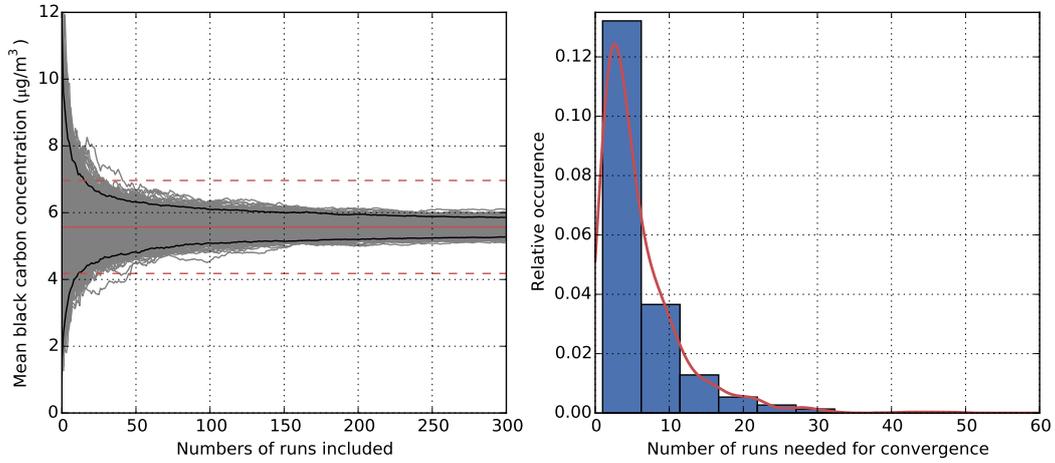


Figure 10: Results of a typical data experiment for all data of route 1 (total of 256 runs). Left: evolution of the mean value in function of the included number of runs for 1000 iterations. The overall mean and the deviation of 25 % around this mean is indicated by the horizontal red lines. The black lines indicate the 2.5 % and 97.5 % percentiles of the mean value. Right: density plot of the required number of runs to obtain convergence for these iterations.

Table 2: Results of the data experiments: required number of repetitions using a deviation of 25 and 50 % at different levels of spatial aggregation: the entire route (route 1), different streets (for abbreviations see Table 1) and 20 and 50 m segments (10th, 50th and 90th percentiles are given), and using different processing methods: arithmetic mean ('standard'), trimmed mean ('trimmed'), with a background normalisation ('background') and the combination of both ('both'). The minimum for each spatial level and for 25 and 50 % is shown in *italics*.

	25 %				50 %			
	standard	trimmed	background	both	standard	trimmed	background	both
Route	17	18	<i>11</i>	12	4	5	3	4
Street level								
PM	18	20	<i>13</i>	14	5	6	3	4
DP	18	19	<i>14</i>	<i>14</i>	5	5	4	4
WO	63	61	60	57	16	11	15	9
KB	82	68	75	<i>61</i>	21	10	20	9
GB	66	50	62	<i>42</i>	18	8	18	7
KA	44	34	40	25	12	7	12	5
20 m segments								
10 %	31	30	26	24	8	6	7	5
50 %	59	54	52	48	15	10	13	8
90 %	164	112	156	<i>108</i>	44	13	43	12
50m segments								
10 %	33	31	29	24	8	7	6	5
50 %	57	52	50	42	14	10	13	8
90 %	141	102	141	<i>94</i>	39	13	40	11

larger. In some cases, however, there is no or even a negative effect. For the Plantin en Moretuslei and Dageraadplaats, the required number of runs increases.

A third experiment is carried out to investigate the use of background normalisation. Background normalisation decreases the variability between runs by accounting for temporal variations in the background concentration, and consequently also the required number of repetitions. The background normalisation clearly decreases the required number of repetitions for all streets compared with the basic experiment, up to 30 % for Dageraadplaats and Plantin en Moretuslei. The background normalisation was performed with both types of background concentrations (hourly median and moving mean), but no significant difference was noticed in the resulting required number of repetitions. Therefore, only the results obtained with the hourly median background concentration are shown in Table 2.

The lowest values for the required number of repetitions are found when combining the use of the trimmed mean and performing the background normalisation. Using this combination reduces the required number of repetitions with 25-33 % (35 - 70 % for 50 % deviation) compared with the initial data experiment. This leads to a number of 24 to 94 repeated measurements runs (10th and 90th percentiles, median of 41) for the different 50 m segments with a deviation of 25 %. When relaxing the deviation to 50 %, these numbers reduce to 5 to 11 (median of 8) runs.

4. Discussion

4.1. Mobile monitoring to assess spatial variation

The BC concentrations in the urban environment show a high variability due to differences in, amongst other, traffic intensity, street topology and proximity of other sources. The results of this case study confirm this variability, showing large variations even over short distances within a single street (see concentration profiles in Figures 5-6). The measured average concentrations at 20 m resolution range from 2-3 $\mu\text{g m}^{-3}$ in the Stadspark to values of 19 $\mu\text{g m}^{-3}$ at the tunnel in Plantin en Moretuslei, while an average concentration of around 6 $\mu\text{g m}^{-3}$ is measured at the monitoring station.

It is clear that stationary monitors are not very well suited to study this level of variability, and that the result can vary greatly depending on where in the street the monitor is placed. This strongly confirms that stationary monitoring stations are not always representative for the population exposure (Kaur et al., 2007). Mobile monitoring, on the other hand, is a suitable monitoring approach for mapping the air quality at a high spatial resolution, as shown with the results above. It can provide insight into the spatial variability that would not be possible with stationary monitors. Previous case studies already showed the potential of repeated mobile measurements for the assessment of spatial variability of pollutants at different micro-environments in a city at street level (Van Poppel et al., 2013; Peters et al., 2013) and within street-level (Zwack et al., 2011b; Peters et al., 2014).

4.2. Influence of events on spatial concentration patterns

The concentration profiles discussed in the previous section show a large variability of the BC concentrations in space, but from the time series on Figure 2a it is clear that there is an even larger variation in time when looking at the one second resolution measurements. With mobile monitoring, measurements at different locations at different times are obtained. As a consequence of this moving monitor, the short-term variability in time can cause an apparent variation in space. However, this variation should not necessarily be attributed to that specific location. The magnitude of this impact and possible solutions to deal with it is investigated.

The results in Section 3.1 clearly show a potentially large influence of extreme events on the average measured BC concentration both at the street level as for segments, and thus on the resulting spatial pattern. This impact is relatively small for all 256 or 96 runs (for route 1 and 2, respectively), but for samples of more realistic size it is much more prominent. For the examples illustrated in Section 3.1 and with a sample size of 20 runs, the influence of one run with an extreme event can go up to 50 to 250 % for a street and 20 m segment, respectively.

To minimize the influence of the short-term temporal variation, data are smoothed, i.e. several one second measurements are averaged over time. In this paper, a spatial smoothing is applied by using passage means at different spatial levels. But even at the street level, events can still have a large impact. The initial spatial smoothing is thus not sufficient to deal with all of the short-term temporal variation at this spatial resolution. Therefore, in a second step, filtering methods or alternative statistics for the average are applied.

Before looking into the processing and possible removal of peaks, it is important to ask the question whether they are relevant to the aim of the measurements. Depending on the application, it can be sensible or not to remove some of the extreme events. For example, in the study of Hagler et al. (2012), an event detection algorithm is used to remove these extreme values from the dataset. In their study, the purpose was to measure the concentration profile in a near-road environment to investigate the influence of a highway. Therefore, a passing car could distort the result heavily as they are less interested in the local emissions of that car. Hudda et al. (2014) and Choi et al. (2012) used a rolling percentile to obtain a baseline concentration to discard the local peaks, as they wanted to map the larger scale variations due to the impact of an airport and a highway, respectively. However, in an urban environment with busy traffic, peak concentrations are an integral part of the air quality that people are typically exposed to. If the goal of the measurement campaign is to map the local concentrations to which cyclists are exposed, it is important to include the peak concentrations. On the other hand, as we have seen, the rare occurrence of large peaks can also distort the concentration profile to a large degree.

To deal with these events and to reduce their impact, we looked at filtering methods and alternative statistics for the arithmetic mean. The median can be used as a more representative central tendency measure as opposed to the mean since air quality measurements are not normally distributed and skewed,

and so the median tends to be more robust to bias caused by emission events (Peters et al., 2013; Brantley et al., 2014). However, the results show that the median also leads to much lower BC concentrations than the average concentration. Being interested in the average concentration, the median will not be used further in this paper. A trimmed mean, with trimming performed for each spatial entity separately, was used instead of the median. The trimmed mean appears to be less sensitive to peaks than the arithmetic mean, but without the large underestimation of the average concentration compared to the median. The COV method as used in Hagler et al. (2012) and Brantley et al. (2014) to detect events also shows good results. However, this method has the risk that it can mask hotspots where peaks occur systematically. This is also reflected in the results, where the peak at 225 m is masked when using the COV method. Recurring peaks at the same place should not be removed but accounted for because they are relevant. Therefore the COV method is not appropriate for the goal in this study and is not used further, although in cases such as Hagler et al. (2012) where the goal is not to measure the sources in direct vicinity, this can certainly be a valuable method. The trimmed mean does not have this issue: when a peak systematically occurs at a certain location, the trimmed mean will only remove part of it and the peak will still be present in the concentration profile. When having smaller sample sizes, the risk of removing an extreme event that was not random but systematic becomes larger.

The results discussed above show that the arithmetic mean can potentially exhibit a large bias due to the occurrence of extreme events. The trimmed mean can be used to counter this, but it is delicate to draw a line between the detrimental impact of an event and typical peak exposure at a location. The trimmed mean seems to hold a good trade-off between these two: it still measures systematically occurring peaks leading to a typical concentration profile, but reduces the detrimental impact of random events on the result. But most important is to be aware of this issue when performing and analysing this kind of measurement campaigns, and taking appropriate actions such as looking at different statistics and investigating locations where large differences between those statistics appear, in more detail. The application of the trimmed mean will further be validated in the following sections on spatial resolution and number of repetitions.

4.3. Spatial resolution

This paper deals with mapping the air quality at different levels of spatial data aggregation (route, streets, segments). But which resolution is relevant and achievable? The entire route can represent the mean exposure of a cyclist on that trajectory, or an estimation of the average concentration in a neighbourhood. The street level gives more detail. This level can be a somewhat artificial delineation, as there are streets with different length and large variations within a single street exist (e.g. the Plantin en Moretuslei, Figure 5), but can be easy to communicate. The segments of certain length provide a more systematic and higher spatial resolution. From Figures 5 and 6 it is clear that the urban air quality (BC levels) can exhibit large variations over short distances. Differences up to an order of 3

over distances of around 50 m are seen. To be able to map the urban air quality and identify hotspots, higher resolutions up to 20 - 50 m are useful. Dependent on the application, such a high resolution can be relevant.

Apart from the relevance to measure air quality at such a high resolution, it is also important to see whether it is achievable with this type of monitoring. The spatial accuracy of the GPS is estimated at 10 m, so a resolution below 10 m is not feasible. The large influence of extreme events on the concentration profile has been discussed before. The question can be raised if it is not the case that a large part of the spatial variation in the profile is caused by such random events. The analysis in Section 3.3 (Figure 7), however, indicates that mapping the air quality at a resolution of 40-50 m leads to consistent profiles and is reproducible. Using higher spatial resolutions up to 20 m is also possible with a small increase in the uncertainty. Of course, this depends on the sample size as the concentration profile will be more accurate or a higher spatial resolution can be used with more repeated runs. Enlarging the grid size from 50 to 100 m does not improve the reproducibility of the profile significantly. When using larger segments, data are smoothed over a longer time period and a reduced variability from events is expected. But at the same time, this will also potentially combine street sections that can be very different, and as such increase the heterogeneity within a single segment. For smaller sample sizes, it is more important to deal with events (in this case using the trimmed mean).

4.4. Temporal representativeness

For a particular location, mobile monitoring gives only a very sparse coverage in time compared with a continuously measuring stationary monitor. This raises the question whether a limited mobile dataset, sparse in time, can be used to estimate the average concentration for the full time period during which the mobile measurements are carried out (in this case 7 am - 1 pm during 4 weeks). To validate the mobile measurements, they are compared with stationary measurements. Because the measurements are not performed at exactly the same place and with the same time resolution, and only three locations are available, the strength of the comparison is limited. But when it is assumed that the stationary monitor is representative for the neighbouring, relatively homogeneous street section, comparing the mobile and stationary measurements can give a valuable indication of the representativeness in time of the mobile measurements.

The results of this comparison indicate that with a limited set of mobile measurements it is possible to find indicative results similar to stationary measurements. The comparison of the average values per day for the same time period (Figure 8) shows that the variations in time (day-to-day variations) are also captured by the mobile measurements. From Figure 9 it can be concluded that the mobile measurements are able to map similar spatial differences as a stationary monitor at each location, although the mobile measurements are consistently slightly higher than the stationary measurements. This can be explained by the shorter distance to the traffic compared with the stationary monitoring locations.

Despite the limited temporal resolution at one point in space, it can be concluded that mobile measurements are able to assess the air quality at a fixed location as could be performed with a stationary monitor. The comparison is performed for three locations, but it is assumed this conclusion can be extrapolated to other locations where mobile measurements were done. Mobile measurements give a representative image of the local concentrations (with certain spatial aggregation level, see previous section), and in this way mobile monitoring can provide for a spatial coverage that is not possible with stationary monitors.

The analysis here is limited to the time period (hours and days) of the mobile measurement campaign, but the question on representativeness could be broadened to longer periods. This could possibly be done with a background normalisation to adjust the measurements for temporal variation, as is also done for short-time stationary measurement campaigns (e.g. Hoek et al., 2002; Eeftens et al., 2012). However, this falls outside the scope of this paper.

4.5. Number of repetitions

The case study presented in this paper, consisting of 256 runs of a specific route, is a very extensive monitoring campaign. In the previous section, the analysis was based on all 256 or 96 runs (for route 1 and 2, respectively). Here, we will discuss whether it is possible to obtain this representative image of the local concentrations with a smaller number of runs. The overall mean of all runs is used as a good estimate of the ‘true’ average concentration at that place.

The results of the data experiment summarised in Table 2 indicate clearly that it is indeed possible to achieve similar results with less repetitions. The required number of repetitions depends on the desired confidence of the result. When decreasing the deviation from the overall mean from 50 % to 25 %, the required number of repetitions increases with a factor of 3 to 9 depending on the location. The required number of repetitions does also strongly depend on the location, as the variation of the measurements at each location is different.

Using the trimmed mean decreases the number of runs up to 33 % for certain locations (for a deviation of 25 %). So by excluding part of the extreme measurements, the trimmed mean converges more rapidly to the overall mean based on all measurements (computed without trimming) compared with the arithmetic mean. As such, we get in general a better estimation of the average BC concentration with the trimmed mean. However, for other streets, it leads to a slight increase in the required number of repetitions. This is due to the fact that using the trimmed mean introduces a systematic bias from the ‘real’ mean, and for locations with a limited number of extreme values this will lead to an increase of the required number of repetitions. But in general, it has a positive effect and the combination of the trimmed mean and a background normalisation does lead to a decrease in the required number of repetitions for all locations. So the trimmed mean and background normalisation are valuable approaches to reduce the required number of repetitions or improve the reliability for a given number of repetitions. The results indicate that it does not matter which

of both types of background concentration values is used. Including temporal variation in the background concentration at below-hourly scale by using a smoothed moving mean did not further reduce the required number of repetitions compared to using hourly medians in the background normalisation.

The numbers of 25 and 50 % used for the deviation give an indication of the uncertainty of the resulting average concentration. They denote that, if a mobile campaign is conducted, the resulting concentration will be within 25 or 50 % from the ‘true’ average. The European directive on ambient air quality (2008/50/EC) uses these numbers of 25 and 50 % as the data quality objectives for particulate matter for fixed and indicative measurements, respectively. However, note that this deviation only includes the uncertainty coming from the variability in the measurand noticed by repeated measurements, and does not include other sources that contribute to the data quality (e.g. the uncertainty on the instrument).

A comparable analysis with this data experiment was performed in Peters et al. (2013) and Van Poppel et al. (2013). They concluded that a limited set of mobile measurements (20 to 24 runs) makes it possible to map the air quality, and for some streets this number could even be reduced to less than 10 (Peters et al., 2013). They used the median as the central tendency measure and a deviation of 15 %. Van Poppel et al. (2013) showed that the use of background normalisation leads to a faster convergence towards a representative concentration and reduces the number of runs required to obtain representative results significantly, which is in line with the results in this paper. In those papers, a more limited number of repetitions to perform the data experiment was used. While their conclusions on the variation between streets and background normalisation are still valid and are confirmed here, the results in this paper contradict their conclusion on absolute numbers of required repetitions. This is likely due to the limited number of runs in their initial dataset and the different sampling method in the data experiment (sampling without replacement instead of with replacement). Xu et al. (2007) also uses a similar approach but to develop another sampling strategy, namely repeated short-term stationary measurements with a mobile laboratory.

The results of the data experiments are limited to the constraints of the available dataset. Firstly, the analysis is constrained by the space and time of the data collection, which was biased in this study towards daytime hours during one month of the year in an urban environment, and hence is not necessarily representative for other periods or locations. The data collection mainly focused on the time of the day when the majority of the people are exposed to outdoor pollution, and did not include night-time. The measurements were equally spread over 11 days including a wide range of background concentrations (meteorological conditions). Secondly, one of the disadvantages of the approach of the data experiment is the dependence on the size of the dataset. The dataset has to be large enough to be able to draw justified conclusions. But given that the results for the required number of runs shown in Table 2 are well below the total number of 256 runs, we conclude that the number in this case study is enough. Lastly, in absence of continuous measurements, the overall mean of the dataset is used as

an approximation of the pollutant concentration to be measured to determine convergence, as the ‘real’ concentration we want to measure is not known. The large number of measurements and the preceding analyses in Section 4.4 on the comparison of stationary and mobile measurements indicate that this approach is justified and solid conclusions can be drawn.

5. Practical considerations and guidelines for mapping the urban air quality using mobile monitoring

In this paper, the use of mobile monitoring to map the local urban air quality at high resolution has been evaluated. Local air quality refers to the average air quality people are exposed to in urban environments and is not limited to urban baseline concentrations, but also includes very localized effects, which are measured in this paper at a resolution of 20 to 50 m (below 100 m). Such high-resolution mapping of the local air quality can be used to detect local hotspots, to derive activity-based exposure to air pollution, or to validate high-resolution models, amongst other applications.

Mobile monitoring is often presented as an additional monitoring tool to increase the spatial resolution of air quality maps. However, as discussed in this study, a limited number of mobile measurements may only represent snapshots of the air quality and is not necessarily representative. Therefore a careful set-up of the mobile monitoring campaign is needed, and the processing and interpretation of the mobile data deserve increased attention.

- The use of the arithmetic mean to aggregate mobile data potentially exhibits a bias due to the occurrence of events, certainly when dealing with a smaller sample size. Appropriate actions to investigate this bias are to assess the possible impact of events on the results in relation to the goal of the campaign and to investigate the results from different statistics (arithmetic mean, median, trimmed mean).
- The trimmed mean resulting from the removal of the lowest and highest measurement is proposed as an integrated outlier detection method to reduce the detrimental impact of random events on the calculated averages. This approach does not over-correct for systematically occurring peaks that can be typical at certain locations or streets and is a better estimator for the average concentration than the arithmetic mean.
- Mobile data are sparse in time when a limited number of monitoring platforms is used to cover a large area. Therefore, temporal aggregation is needed. Further, to account for the temporal variation in the background concentrations, a background normalisation should be used. Background normalization reduces the variability between different runs and leads to a more reliable mapping of the air quality at lower monitoring efforts.
- Spatial aggregation is applied to smooth the data at different spatial levels (routes, streets, segments). We showed that mapping of the air quality at a spatial resolution up to

50 m is feasible for BC. A higher spatial resolution of 20 m could be obtained with a slightly increased uncertainty depending on the sample size.

- To generate high-resolution maps over larger areas (several km²), large quantities of data are required to include the range of possible meteorological conditions and the range of local air quality conditions (depending on local sources, e.g. traffic intensity), and to counter occasional and exceptional events. It is important to assess whether enough repetitions are made in relation to the goals of the monitoring campaign.
- The data requirements can be reached by performing repeated monitoring runs. The exact requirements depend on the mapping resolution in space and time and the quality constraints of the results, but can be decreased by using a trimmed mean for averaging and by performing a background normalisation. When targeting an uncertainty of 50 %, 10 repeated measurement runs will satisfy to estimate the average concentration at almost all locations. When performing 40 repeated measurement runs, the average concentration could be estimated at half of the locations with an uncertainty of 25 % or below.

There are also some important limitations to the approach of mobile monitoring. Firstly, mobile data are representative for the time period in which they were collected, which is typically a confined time frame. A possibility is to use background normalization to extrapolate the results to other periods assuming that the local contributions are similar in time and that the main temporal variation is caused by background effects that are normalized. Secondly, it should also be stressed that the representativeness of the result depends on the type of mobile platform. In this study a bicycle was used as monitoring platform and hence the results are representative for the exposure of cyclists, but not necessarily for other transport modes such as pedestrians and car drivers, or residents. Thirdly, measurements of BC are used in this study. Pollutants such as BC, UFP and NO_x are strongly traffic-related and will therefore show a high variability in urban environments. Other types of pollutants such as PM_{2.5} or PM₁₀ have less small-scale dynamics and the guidelines will differ for these other pollutants. Lastly, despite of strategies to reduce the required number of repetitions, a large number of repeated measurement runs is still required. In the current set-up of this study, it required dedicated man-power to carry out the measurements, which can be very time consuming. Alternative approaches exist to reduce this workload taking advantage of existing infrastructure or using the normal daily routines. Monitors can be attached at e.g. trams, as is done in studies in Zurich, Switzerland (Hasenfrazz et al., 2014) and Karlsruhe, Germany (Hagemann et al., 2014). Another approach is to rely on participatory or community-based sensing, possibly with low-cost sensors, to collect a larger amount of data (Snyder et al., 2013).

6. Conclusion

Mobile monitoring is a suitable monitoring approach for mapping the urban air quality at a high spatial resolution. Mapping at a spatial resolution up to 20 m is relevant as large gradients over short distances are observed and differences up to a factor of 10 in BC concentrations are measured at this resolution. As we have shown in this paper, mapping at such a high resolution is also possible, but a lot of repeated measurements are required. This can, however, be reduced by dealing with extreme events using the trimmed mean and by performing a background normalisation. Based on a targeted measurement campaign with 256 runs and after computing a trimmed mean and applying background normalisation, we found that in this case study 24 to 94 repeated measurement runs (median of 41) for the different locations are required to map the BC concentrations at 50 m resolution with an uncertainty of 25 %. When relaxing the uncertainty to 50 %, these numbers reduce to 5 to 11 (median of 8) runs. Given the increasing use of mobile air quality monitoring and its potential for participatory sensing and crowd-sourcing projects, the representativeness of the data that are collected in this way is a key issue. Methods and guidelines to collect and process these data in a proper way have to be drafted. The considerations and guidelines raised in this study are very relevant in that context. A careful set-up is needed with a sufficient number of repetitions in relation to the desired reliability and spatial resolution, dependent on the aim of the campaign. Specific data processing methods such as background normalisation and event detection are applied. Given this careful set-up and processing, mobile monitoring can provide insight into the spatial variability that would not be possible with stationary monitors.

Acknowledgement

We are grateful to Jo Van Laer, Daniel Wilczek and Matteo Regente for their assistance in the collection of the mobile data. This research was partly funded by the IDEA (Intelligent, Distributed Environmental Assessment) project, a 4-year strategic basic research project, financially supported by IWT-Vlaanderen (Flemish Agency for Innovation by Science and Technology), partly by the EveryAware project, funded under the Seventh Framework Programme, IST - FET Open Scheme program, EU RD contract IST-265432 and partly by VITO research funding.

References

- Berghmans, P., Bleux, N., Int Panis, L., Mishra, V.K., Torfs, R., Van Poppel, M., 2009. Exposure assessment of a cyclist to PM10 and ultrafine particles. *Science of The Total Environment* 407, 1286–98.
- Boogaard, H., Borgman, F., Kamminga, J., Hoek, G., 2009. Exposure to ultrafine and fine particles and noise during cycling and driving in 11 Dutch cities. *Atmospheric Environment* 43, 4234–4242.
- Brantley, H.L., Hagler, G.S.W., Kimbrough, E.S., Williams, R.W., Mukerjee, S., Neas, L.M., 2014. Mobile air monitoring data-processing strategies and effects on spatial air pollution trends. *Atmospheric Measurement Techniques* 7, 2169–2183.
- Bukowiecki, N., Dommen, J., Prévôt, A.S.H., Weingartner, E., Baltensperger, U., 2003. Fine and ultrafine particles in the Zürich (Switzerland) area measured with a mobile laboratory: an assessment of the seasonal and regional variation throughout a year. *Atmospheric Chemistry and Physics* 3, 1477–1494.
- Burke, J., Estrin, D., Hansen, M., Ramanathan, N., Reddy, S., Srivastava, M.B., 2006. Participatory sensing, in: *Workshop on World-Sensor-Web (WSW06): Mobile Device Centric Sensor Networks and Applications*, pp. 117–134.
- Choi, W., He, M., Barbesant, V., Kozawa, K.H., Mara, S., Winer, A.M., Paulson, S.E., 2012. Prevalence of wide area impacts downwind of freeways under pre-sunrise stable atmospheric conditions. *Atmospheric Environment* 62, 318–327.
- Dons, E., Int Panis, L., Van Poppel, M., Theunis, J., Wets, G., 2012. Personal exposure to Black Carbon in transport microenvironments. *Atmospheric Environment* 55, 392–398.
- Dons, E., Int Panis, L., Van Poppel, M., Theunis, J., Willems, H., Torfs, R., Wets, G., 2011. Impact of time-activity patterns on personal exposure to black carbon. *Atmospheric Environment* 45, 3594–3602.
- Dutta, P., Aoki, P., Kumar, A., Mainwaring, A., Myers, C., Willet, W., Woodruff, A., 2009. Common Sense: participatory urban sensing using a network of handheld air quality monitors, in: *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pp. 349–350.
- Eeftens, M., Tsai, M.Y., Ampe, C., Anwander, B., Beelen, R., Bellander, T., Cesaroni, G., Cirach, M., Cyrys, J., de Hoogh, K., De Nazelle, A., de Vocht, F., Declercq, C., Ddel, A., Eriksen, K., Galassi, C., Gražulevičien, R., Grivas, G., Heinrich, J., Hoffmann, B., Iakovides, M., Ineichen, A., Katsouyanni, K., Korek, M., Krämer, U., Kuhlbusch, T., Lanki, T., Madsen, C., Meliefste, K., Mölter, A., Mosler, G., Nieuwenhuijsen, M., Oldenwening, M., Penanen, A., Probst-Hensch, N., Quass, U., Raaschou-Nielsen, O., Ranzi, A., Stephanou, E., Sugiri, D., Udvardy, O., Vaskövi, E., Weinmayr, G., Brunekreef, B., Hoek, G., 2012. Spatial variation of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2 - Results of the ESCAPE project. *Atmospheric Environment* 62, 303–317.
- Elen, B., Peters, J., Van Poppel, M., Bleux, N., Theunis, J., Reggente, M., Standaert, A., 2013. The aeroflex: a bicycle for mobile air quality measurements. *Sensors* 13, 221–240.
- Elen, B., Theunis, J., Ingarra, S., Molino, A., Van den Bossche, J., Reggente, M., Loreto, V., 2012. The EveryAware SensorBox: a tool for community-based air quality monitoring, in: *Workshop Sensing a Changing World*, pp. 1–7.
- EPA, 2011. Report to Congress on Black Carbon - External Peer Review Draft. Technical Report EPA-450/D-11-001.
- Fruin, S., Urman, R., Lurmann, F., McConnell, R., Gauderman, J., Rappaport, E., Franklin, M., Gilliland, F.D., Shafer, M., Gorski, P., Avol, E., 2014. Spatial variation in particulate matter components over a large urban area. *Atmospheric Environment* 83, 211–219.
- Hagemann, R., Corsmeier, U., Kottmeier, C., Rinke, R., Wieser, A., Vogel, B., 2014. Spatial variability of particle number concentrations and NOx in the Karlsruhe (Germany) area obtained with the mobile laboratory AERO-TRAM. *Atmospheric Environment* 94, 341–352.
- Hagler, G.S., Lin, M.Y., Khlystov, A., Baldauf, R.W., Isakov, V., Faircloth, J., Jackson, L.E., 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.
- Hagler, G.S., Thoma, E.D., Baldauf, R.W., 2010. High-resolution mobile monitoring of carbon monoxide and ultrafine particle concentrations in a near-road environment. *Journal of the Air & Waste Management Association* 60, 328–336.
- Hagler, G.S., Yelverton, T.L., Vedantham, R., Hansen, A.D., Turner, J.R., 2011. Post-processing method to reduce noise while preserving high time resolution in Aethalometer real-time black carbon data. *Aerosol and Air Quality Research* 11, 539–546.
- Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Thiele, L., 2014. Pushing the spatio-temporal resolution limit of urban air pollution maps, in: *Pervasive Computing and Communications (PerCom)*, 2014 IEEE International Conference, pp. 69–77.
- Hoek, G., Meliefste, K., Cyrys, J., Lewné, M., Bellander, T., Brauer, M., Fischer, P., Gehring, U., Heinrich, J., Van Vliet, P., Brunekreef, B., 2002. Spatial variability of fine particle concentrations in three European areas. *At-*

- ospheric Environment 36, 4077–4088.
- Hu, S., Paulson, S.E., Fruin, S., Kozawa, K., Mara, S., Winer, A.M., 2012. Observation of elevated air pollutant concentrations in a residential neighborhood of Los Angeles California using a mobile platform. *Atmospheric Environment* 51, 311–319.
- Hudda, N., Gould, T., Hartin, K., Larson, T.V., Fruin, S.A., 2014. Emissions from an international airport increase particle number concentrations 4-fold at 10 km downwind. *Environmental Science & Technology* 48, 6628–35.
- Int Panis, L., de Geus, B., Vandenbulcke, G., Willems, H., Degraeuwe, B., Bleux, N., Mishra, V., Thomas, I., Meeusen, R., 2010. Exposure to particulate matter in traffic: A comparison of cyclists and car passengers. *Atmospheric Environment* 44, 2263–2270.
- Janssen, N.A., Gerlofs-Nijland, M.E., Lanki, T., Salonen, R.O., Cassee, F., Hoek, G., Fischer, P., Brunekreef, B., Krzyzanowski, M., 2012. Health effects of black carbon. Technical Report. World Health Organization.
- Kaur, S., Nieuwenhuijsen, M., Colvile, R., 2005. Personal exposure of street canyon intersection users to PM_{2.5}, ultrafine particle counts and carbon monoxide in Central London, UK. *Atmospheric Environment* 39, 3629–3641.
- Kaur, S., Nieuwenhuijsen, M., Colvile, R., 2007. Fine particulate matter and carbon monoxide exposure concentrations in urban street transport microenvironments. *Atmospheric Environment* 41, 4781–4810.
- Kingham, S., Longley, I., Salmond, J., Pattinson, W., Shrestha, K., 2013. Variations in exposure to traffic pollution while travelling by different modes in a low density, less congested city. *Environmental Pollution* 181, 211–218.
- Kuhlbusch, T.A., Quincey, P., Fuller, G.W., Kelly, F., Mudway, I., Viana, M., Querol, X., Alastuey, A., Katsouyanni, K., Weijers, E., Borowiak, A., Gehrig, R., Hueglin, C., Bruckmann, P., Favez, O., Sciare, J., Hoffmann, B., Espenyttri, K., Torseth, K., Sager, U., Asbach, C., Quass, U., 2014. New Directions: The future of European urban air quality monitoring. *Atmospheric Environment* 87, 258–260.
- MacNaughton, P., Melly, S., Vallarino, J., Adamkiewicz, G., Spengler, J.D., 2014. Impact of bicycle route type on exposure to traffic-related air pollution. *Science of the Total Environment* 490, 37–43.
- Merbitz, H., Fritz, S., Schneider, C., 2012. Mobile measurements and regression modeling of the spatial particulate matter variability in an urban area. *Science of The Total Environment* 438, 389–403.
- Padró-Martínez, L.T., Patton, A.P., Trull, J.B., Zamore, W., Brugge, D., Durant, J.L., 2012. Mobile monitoring of particle number concentration and other traffic-related air pollutants in a near-highway neighborhood over the course of a year. *Atmospheric Environment* 61, 253–264.
- Peters, J., Theunis, J., Van Poppel, M., Berghmans, P., 2013. Monitoring PM₁₀ and ultrafine particles in urban environments using mobile measurements. *Aerosol and Air Quality Research* 13, 509–522.
- Peters, J., Van den Bossche, J., Reggente, M., Van Poppel, M., De Baets, B., Theunis, J., 2014. Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. *Atmospheric Environment* 92, 31–43.
- Petzold, A., Ogren, J.A., Fiebig, M., Laj, P., Li, S.M., Baltensperger, U., Holzer-Popp, T., Kinne, S., Pappalardo, G., Sugimoto, N., Wehrli, C., Wiedensohler, A., Zhang, X.Y., 2013. Recommendations for reporting “black carbon” measurements. *Atmospheric Chemistry and Physics* 13, 8365–8379.
- Petzold, A., Schloesser, H., Sheridan, P.J., Arnott, W.P., Ogren, J.A., Virkkula, A., 2005. Evaluation of multiangle absorption photometry for measuring aerosol light absorption. *Aerosol Science and Technology* 39, 40–51.
- Petzold, A., Schonlinner, M., 2004. Multi-angle absorption photometry - a new method for the measurement of aerosol light absorption and atmospheric black carbon. *Journal of Aerosol Science* 35, 421–441.
- PyData Development Team, 2014. pandas: Python Data Analysis Library. URL: <http://pandas.pydata.org/>.
- QGIS Development Team, 2014. QGIS Geographic Information System. Open Source Geospatial Foundation. URL: <http://qgis.osgeo.org>.
- R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- Seinfeld, J.H., Pandis, S.N., 2012. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons.
- Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S.W., Shelow, D., Hindin, D.A., Kilaru, V.J., Preuss, P.W., 2013. The changing paradigm of air pollution monitoring. *Environmental Science & Technology* 47, 11369–11377.
- Spira-Cohen, A., Chen, L.C., Kendall, M., Sheesley, R., Thurston, G.D., 2010. Personal exposures to traffic-related particle pollution among children with asthma in the South Bronx, NY. *Journal of Exposure Science & Environmental Epidemiology* 20, 446–56.
- Sullivan, R., Pryor, S., 2014. Quantifying spatiotemporal variability of fine particles in an urban environment using combined fixed and mobile measurements. *Atmospheric Environment* 89, 664–671.
- 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* 183, 224–233.
- Virkkula, A., Makela, T., Hillamo, R., Yli-Tuomi, T., Hirsikko, A., Hameri, K., Koponen, I., 2007. A simple procedure for correcting loading effects of aethalometer data. *Journal of the Air & Waste Management Association* 57, 1214–1222.
- Wallace, J., Corr, D., Deluca, P., Kanaroglou, P., McCarry, B., 2009. Mobile monitoring of air pollution in cities: the case of Hamilton, Ontario, Canada. *Journal of Environmental Monitoring* 11, 998–1003.
- Weber, S., 2009. Spatio-temporal covariation of urban particle number concentration and ambient noise. *Atmospheric Environment* 43, 5518–5525.
- Weijers, E., Khlystov, A., Kos, G., Erisman, J., 2004. Variability of particulate matter concentrations along roads and motorways determined by a moving measurement unit. *Atmospheric Environment* 38, 2993–3002.
- Westerdahl, D., Fruin, S., Sax, T., Fine, P.M., Sioutas, C., 2005. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmospheric Environment* 39, 3597–3610.
- Xu, X., Brook, J., Guo, Y., 2007. A statistical assessment of saturation and mobile sampling strategies to estimate long-term average concentrations across urban areas. *Journal of the Air & Waste Management Association* 57, 1396–1406.
- Zwack, L.M., Hanna, S.R., Spengler, J.D., Levy, J.I., 2011a. Using advanced dispersion models and mobile monitoring to characterize spatial patterns of ultrafine particles in an urban area. *Atmospheric Environment* 45, 4822–4829.
- Zwack, L.M., Paciorek, C.J., Spengler, J.D., Levy, J.I., 2011b. Characterizing local traffic contributions to particulate air pollution in street canyons using mobile monitoring techniques. *Atmospheric Environment* 45, 2507–2514.
- Zwack, L.M., Paciorek, C.J., Spengler, J.D., Levy, J.I., 2011c. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. *Environmental Health Perspectives* 119, 852–9.