Juan Pablo Orjuela Mendoza

Using personal monitors to assess BC exposures and inhaled doses

This chapter in a nutshell

- Geolocation, physical activity levels, and personal exposure to BC were measured for 122 participants in three European cities, and their inhaled doses were estimated.
- Daily averages of BC exposure were found to be between 1,720ng/m^3 for Barcelona during the mid-season average (spring and autumn) and 3,100ng/m^3 in the London winter.
- Transport microenvironments were shown to contribute with 10.5% of daily BC exposures and 18.1% of the daily inhaled dose despite people spending 7.5% on average of their day in these, showing similar results as those from the model for NO_2 and $PM_{2.5}$ in the previous chapter.
- Cyclists were shown to have average inhaled doses while in transport that were twice as high as those from pedestrians and more than five times more than those from vehicle users due to the combination of high exposures with high respiration rates.
- Differences in exposure levels and inhaled doses while in transport were large enough to show statistically significant differences between daily values of different mode users.
- No correlation was found between measured BC geometric means of participants and the modelled NO_2 concentrations at home locations despite these two pollutants showing good correlation in an urban background station.

Introduction

In Chapter 4 I presented how to use GPS and accelerometer data from a free-living population to model their exposure and inhaled doses as a way to solve the spatial resolution issues discussed in Chapter 3. I showed how time spent in transport was particularly relevant as it accounted for more than 35% of inhaled doses despite taking on average not more than 10% of the participants' time. However, one of the main questions that remains is whether or not these models do reflect in any way real conditions. In the model presented previously, ratios derived from the literature for indoors and outdoors, as well as for the various travel modes, represented a very simplified approach to real conditions. Although uncertainty analyses of travel mode ratios and stochastic variables of respiration rates showed encouraging results with percentage differences of ~+/-10% for daily geometric mean exposures and inhaled doses, a gold standard to evaluate the accuracy and precision of modelled exposure is to compare it to actual measurements of personal exposures (Physick *et al.*, 2011; Dons, Kochan, *et al.*, 2014; Bekö *et al.*, 2015).

Black carbon (BC) is an air pollutant that has been associated with multiple health effects such as increase in all-cause and cardiovascular mortality and cardiopulmonary hospital admissions in short and long term cohort studies (Jansen *et al.*, 2005; Janssen *et al.*, 2011, 2012). It is a solid carbon-based fraction of particulates and its main sources are diesel combustion engines and some residential burning of wood and coal. Portable devices such as micro aethalometers have been increasingly used for personal exposure of black carbon (Dons *et al.*, 2011; Hankey and Marshall, 2015; Williams and Knibbs, 2016; Laeremans *et al.*, 2018) as they are reliable and easy to use. The differences between BC and other carbon fractions commonly cited in the literature such as elemental carbon (EC) and organic carbon (OC) are discussed in more detail in Chapter 2. For the purposes of this chapter, the most relevant characteristics of

BC are its clear associations with health outcomes, its association with diesel engines, and the relative ease to measure personal exposure using portable devices.

Coupling micro aethalometers with GPS and physical activity wearable sensors helps identify where do the episodes of high BC concentrations happen and estimate inhaled doses of air pollution at the same time. However, with current technologies, deploying large measuring campaigns is impractical mainly due to the elevated costs this would imply (Gulliver and Briggs, 2011; Physick *et al.*, 2011; Gurram, Stuart and Pinjari, 2014). Nevertheless, the data collected is of great value and helps understand how different activity patterns and physical activity levels affect daily exposure averages and inhaled doses of air pollutants. Getting a sense of the variability in daily exposures between individuals, the microenvironments that contribute the most to daily inhaled doses, and their associations with traditional variables such as home-based concentration levels is key in creating healthier urban environments.

In this chapter I will present the results from my involvement in the Physical Activity through Sustainable Transport Approaches (PASTA) project, an international consortium of thirteen European institutions studying cycling and walking policies in seven case-study cities. The project proposed a mixed-method and multilevel approach to investigate the determinants of active mobility and evaluate measures meant to increase it (Dons *et al.*, 2015; Gerike *et al.*, 2016). It included a longitudinal survey and the systematic collection of data in each case-study city that could give contextual differences between them and aid in finding generalizable trends. The PASTA project had other areas in which I was not actively involved such as stakeholder meetings to identify gaps between research and practitioners, and a summary of good practices in terms of promoting active travel around the world. The core of my involvement was associated to two additional modules aimed at evaluating the health impacts of physical activity in polluted urban contexts and understanding people’s mobility patters from personal phone data. I was also part of the data analysis working group, involved in the cleaning and analysing of the main survey data. The first of the modules mentioned, the so-called "health add-on" module is the topic of this chapter and is described next. The second, the "tracking add-on" module, collected smart phone-based geo-tracking data from a subset of PASTA survey participants: these data are still under study and the results will not be part of this document.

The "health add-on" module included health and air pollution measurements; descriptions of the sample and results from the health analyses are also described elsewhere (Dons *et al.*, 2017; Laeremans *et al.*, 2018). This chapter will be focused on my work in these in relation to my PhD thesis. However, it is also worth mentioning that as part of the data analysis working group in PASTA, I was involved in several journal publications that use data I collected and that my role in each is explicitly mentioned in the associated papers. Since these deviate from the main argument in this document, I will not make explicit mention to them unless deemed relevant.

In Chapter 3, I showed how population-activity-based exposures could be estimated using call detail records to identify air pollution hotspots based on people's activity patterns and not only home or work location. I also showed the implications of temporal patterns of air pollution in exposure estimates. The use of call records allowed for the use of a large sample of 8 million different callers during two months in seven Italian cities, but this came with the cost of aggregated data that gave no information on the microenvironments in which these people were and the results were particularly sensitive to the level of spatial aggregation. In Chapter 4, a different approach was taken in which detailed data of individuals' exact location and physical activity data from smartphones was used to model daily exposures and inhaled doses as well as the contribution to these of the different microenvironments and transport modes used. However, in that case I had no access to actual measurements of air pollution data and thus the validity of such results remained unanswered. Following that line of more detailed data on an individual level at the expense of using smaller samples and higher costs, in this chapter I will show the results of *measured* daily geometric mean exposures of BC and estimated inhaled doses based on physical activity levels from a free-living population of 120 participants in three European cities (40 participants in each). The contribution to daily BC exposures and inhaled doses of different microenvironments will be calculated based on an activity and mode detection algorithm that combines GPS and accelerometer data instead of the detailed activity diaries used in Chapter 4. As in previous chapters, I will compare the daily activity-based exposure averages with home-based estimates and comment on their differences. As I use more detailed information the question on how to extrapolate from these results to a scale useful for epidemiological studies will remain. However, in Chapter 6 I will address this question by suggesting that linear regression models based on information from surveys, activity diaries, geolocation and a limited amount of personal exposure measures have the potential of becoming an important and affordable tool to estimate daily means of personal exposure and total inhaled doses at larger scales.

Methods

Primary data collection

Exposure concentrations of BC, physical activity levels and geo-position data were collected in three PASTA cities, namely Antwerp, Barcelona and London. In each city close to 40 participants were recruited from the general PASTA survey. I was in charge of the recruiting process for London while two colleagues (also PhD students), Michelle Laeremans and Ione Avila-Palencia, were in charge of recruitment in Antwerp and Barcelona respectively. The first step in the recruiting process was to select people form the PASTA survey (the reader is referred to Gerike et al. 2016 for details on the survey) who had answered "Yes" to the question in the baseline questionnaire that read: "*As part of the PASTA project we also conduct research using new innovative measurement devices. These devices study your mobility, your physical activity, or measure health indicators of air pollution. Would you be interested in participating in one of these studies?*". As our attention was to recruit healthy adults so as to avoid any interference with our health measurements, we filtered for those who according to their self-reported weight and height had a body mass index (BMI) of less than 30 (Delfino et al (Delfino *et al.*, 2014) showed that the effect of an increase in BC on blood pressure is higher on subjects with higher BMI), were non-smokers (Cole-Hunter *et al.*, 2013) and were between 18 and 65 years of age (Jacobs *et al.*, 2010). Through an email sent to the selected subgroup we requested additional health information to further select participants who complied with the following criteria:

- No diseases or medication that could influence health parameters measured (Jacobs *et al.*, 2010; Cole-Hunter *et al.*, 2013; Sarnat *et al.*, 2014), such as:
    - Non-asthma pulmonary disease
    - Asthma medication
    - Previous myocardial infarct, heart failure or other cardiovascular, pulmonary, neurological and endocrine diseases
    - Subjects who used digoxin or beta blockers
    - Subjects on anti-platelet therapy
    - Diabetes
    - Respiratory infections two week prior to the start of the study
- Not pregnant or planning on being pregnant in the next year (Sarnat *et al.*, 2014)
- Preferably, not exposed to tobacco smoke at home (Weichenthal *et al.*, 2011)

In addition, subjects were chosen so as to have a mix of physically active and inactive participants, as defined by the WHO's recommendations - i.e. above or below 150 min of moderate physical activity or 75 min of vigorous physical activity per week (World Health Organization, 2010).

After checking for these, participants who fulfilled all the criteria were contacted via email and invited to take part in the health add-on. To explain in greater detail what the project implied and answer any questions, participants were offered a call or a personal meeting. If they agreed to take part in the study, a starting date would be scheduled. An economic incentive of £150 was offered for the London participants and €150 for the Antwerp and Barcelona participants.

Every participant was monitored for one week and asked to follow the steps illustrated in Figure 5.1. The process would then be replicated in two more weeks in a way that each participant would be monitored for one week in winter, one week in summer, and one week in either spring or autumn (this last one referred to in this study as the “mid-season” week). The measurement campaign was conducted between February 2015 and March 2016, with about six participants per city starting every other week.

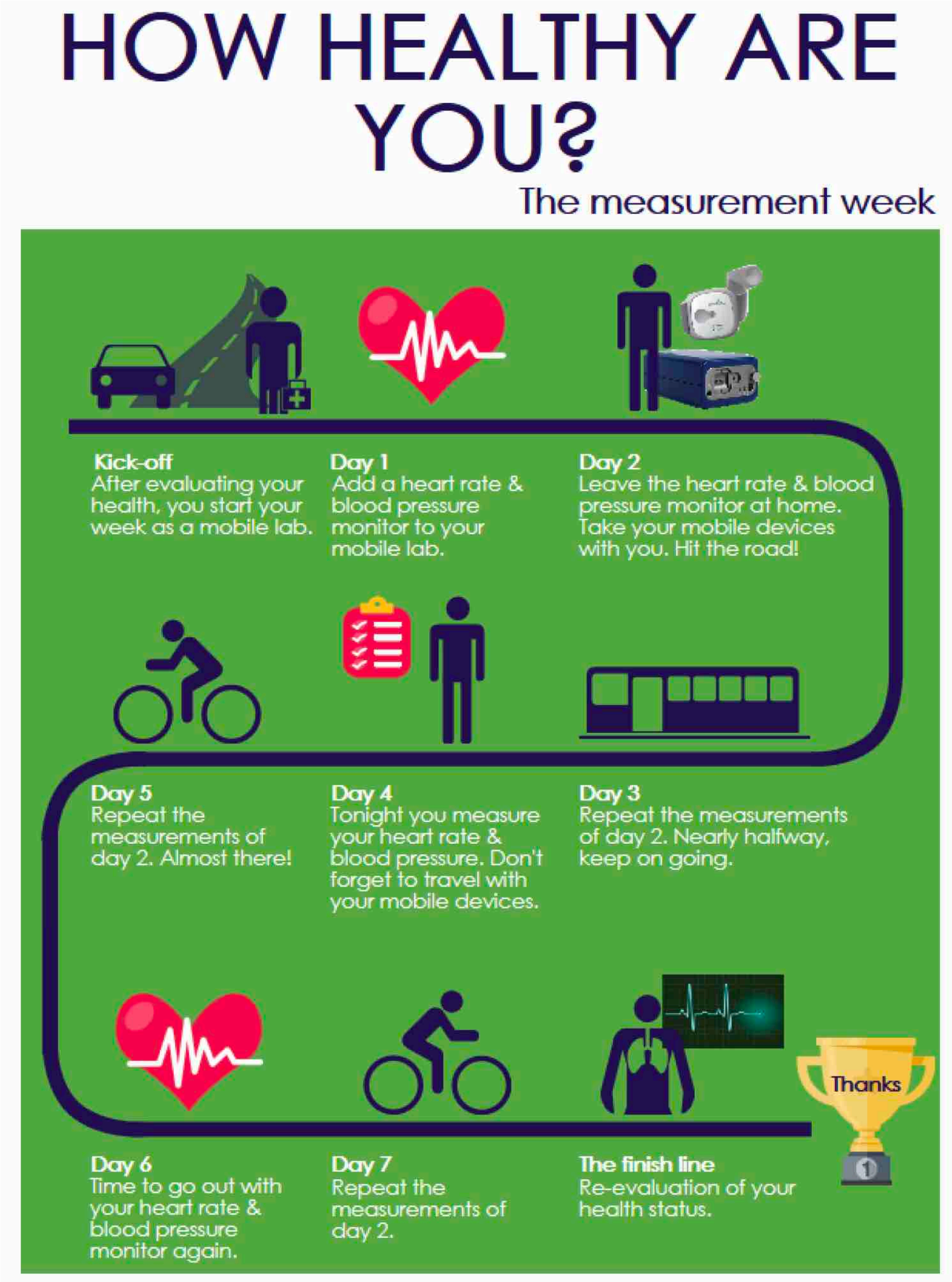


Figure 5.1. Illustration of the measuring week given to every participant at the beginning of each measuring week with further instructions and a daily check-list of things to do, such as charging the devices or changing filters if necessary.

Participants were given a microaethalometer (AE51,AethLabs, San Francisco, CA), a GPS (i-gotU GT-600), a smartphone (Samsung Galaxy SII), a physical activity monitor (SenseWear armband, model MF-SW, BodyMedia, Pittsburgh, PA), a heart-rate monitor (Zephyr Bioharness, Zephyr Technology Corp.,Annapolis, MD) and a blood pressure monitor (Omron M3) for each week. To collect the devices participants were asked to go to the PASTA local offices (i.e. Imperial College in London, ISGlobal in Barcelona and VITO in Antwerp) where the participants were given instructions on how to use them, asked to sign the informed consent forms and additional health measurements such as retinal images, exhaled NO and spirometry were taken. The purpose of these health measurements including the use of the heart-rate and blood pressure monitors fall outside the scope of this document and the reader is referred to other publications for more information (Dons *et al.*, 2017; Laeremans *et al.*, 2018). For this study, the relevant devices are the micro aethalometer, the GPS, the smartphone and the physical activity monitor. Details on these are shown in Table 5.1.

Table 5.1. General description of key devices used

Device	Main variables measured	Frequency of measurement	Frequency of charge	Exported files info
Micro aethalometer	BC concentrations	Set to 5 minutes with 100ml/min flow	Participants were asked to charge this every night	BC concentrations in ng/m3 and other status variables (i.e. filter attenuation, flow status, battery left)
GPS	Latitude and longitude	Variable depending on	Participants were asked	.csv and .kml files with latlong

		satellite signal (close to every 10 seconds)	to charge this every night	information and time
Smartphone	Latitude and longitude; tri-axial accelerometry	GPS: Variable depending on satellite signal (close to every 10 seconds) Accelerometer: 10Hz	Participants were asked to charge every night and all phones were equipped with batteries with double the capacity than the default*.	.csv and .kml files with latlong as well as accelerometer data.
SenseWear	METS and activity type	30Hz	Participants were asked to charge this twice throughout the week	File with information on METS, physical activity level assignment, galvanic skin response, activity classification, body and near-body temperature

*Despite this, the ExpoApp used in this case was too battery demanding and participants continuously reported the phone not making it through the day without additional charging

Participants were also asked to fill out three additional questionnaires during their measurement week: one before starting the measurements, one half-way through their measurements and one on their last day. These included an activity diary where participants would give a detail account of the trips done the day before filling out the questionnaire, including origin, destination, trip purpose, trip length, mode(s) used and details about any stages in their trip (e.g. if a participant cycled to a tube station, took the tube, and then walked to the office, this trip would have three stages: cycling, on the tube and walking).

With 40 participants per city, each being monitored for three weeks in three different seasons, the campaign was a coordinated effort that implied a fair amount of work from all people involved, that included the scheduling of appointments with participants, charging the devices, entering the settings on each device before giving them to participants, printing the instructions and check-lists, downloading the data after each measuring week and saving the files in the PASTA servers through a secure file transfer protocol (SFTP).

Urban background concentrations

Antwerp and Barcelona do not have urban background results available for download online but these were requested from local authorities by our PASTA partners in those cities. Both cities measure BC background concentrations with a Multi-angle Absorption Photometer (MAAP) and the data given corresponded to daily averages for the entire data collection period (February 2015 to March 2016). London data can be downloaded from the London Air website (London Air and King's College London, 2018) and is collected in North Kensington Station with a Magee Aethalometer model AE22. London data can be downloaded at an hourly resolution and thus, all temporal analysis requiring an hourly resolution where done using only the London data.

Data cleaning and pre-processing

The micro aethalometer data was first cleaned by running the United States Environmental Protection Agency's (EPA) optimized noise-reduction algorithm (ONA) for micro aethalometer results (Hagler *et al.*, 2011). This algorithm developed by the EPA is recommended for users of real-time black carbon data using micro aethalometers to reduce noise and negative values in the data. Although participants were instructed to change the device's filter every two days, sometimes they would fail to do so and some high attenuation values, indicating high filter loading, were present in the data. Data collected during error codes associated to the flow being out of range or with signal out of range due to lack of replacing the filters, and data with extremely high (above 250,000 ng/m^3) or low (-50,000 ng/m^3) 5-minute averages were all excluded (Dons *et al.*, 2013). These limit values were set by the PASTA data analysis working group based on previous experiences but will be reviewed in greater detail before these results are submitted for publication.

The GPS data was run through the Physical Activity Location Measurement System - PALMS algorithm (Tandon *et al.*, 2013; Bekö *et al.*, 2015) developed and maintained by the University of California, San Diego, to identify trip start and end times, repeated locations, and a first estimate of mode used. To guarantee participants' anonymity when using the web-based PALMS tool, latitude, longitude and time results were shifted before using PALMS and then shifted back offline. This way, no exact coordinates were shared online in compliance with European ethical commitments and regulations applied in PASTA.

The smartphone data was collected using the ExpoApp and processed using the SensorLab software developed for this purpose. SensorLab uses the accelerometer and assisted GPS data collected by the smartphone and outputs one csv file and one KML file per measurement period with accelerometer results, GPS coordinates, and origin of the coordinate data (GPS or network data).

All devices were checked for time zone consistencies and corrected when clock changes happened during the measuring week. Data collected by the SenseWear, GPS and micro aethalometer was merged in one single dataframe in R with a minute-by-minute list starting at the first micro aethalometer measurement of the week and finishing at the last. Since the micro aethalometer output data had a 5-minute resolution, each BC concentration result was repeated for the five minutes of measurements, while GPS coordinates were adjusted to match the closest minute and a missing data indicator ("NA") assigned where no data was available (due to loss of signal, for example). SenseWear data had a minute resolution so these needed no adjustment when matching the results in one single dataframe.

Identifying the mode used

The mode assignment categorized each trip into identified by PALMS into four modes: cycle (including electric cycles), walking, vehicle, and other. Category "vehicle" included private car, van or motorcycle as well as buses, taxis and rail. The category "other" is assumed to have some unclassified rail points from the underground metro systems due to loss of GPS signal, as well as scooters, or other non-motorized modes if used.

The PALMS algorithm used estimates the mode used while in transport based only on GPS velocity. Although PALMS is programmed to deal with accelerometer data also, the format needed by PALMS for this was not compatible with the SenseWear output files. The SenseWear itself has an independent activity detection function based on the accelerometer data that classifies every reading in one of the following: walking, running, stationary biking, resting, resistance, motoring, road biking, sleeping and elliptical training. The algorithms to derive these activities are part of the software provided by the SenseWear and are not open code so the categories mentioned did not exactly match the desired ones for this specific study. These two mode assignment methods were not always consistent, and the GPS output files had some data with misclassified time stamps due to software failure, so I developed a “combined method” as explained in detail below and sketched in Figure 5.2.

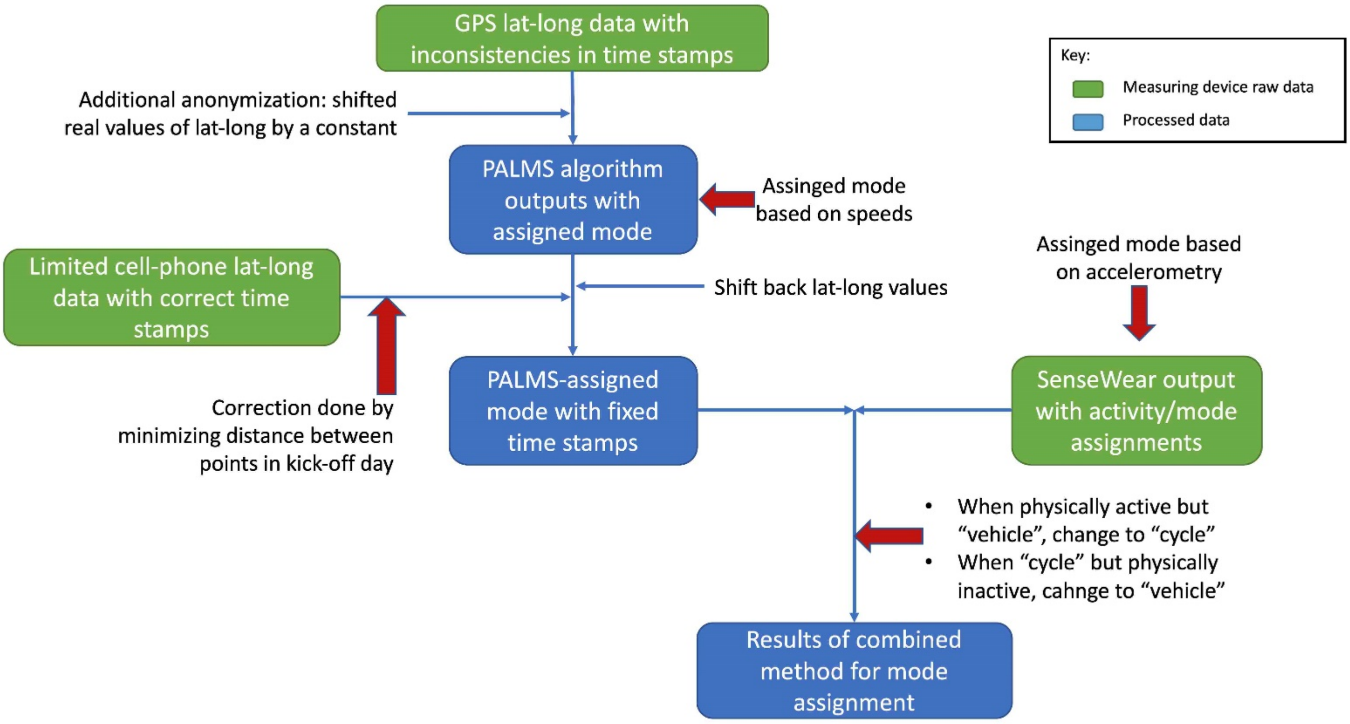


Figure 5.2. The combined method for mode assignment using the information available.

The GPS data was first given as an input to the PALMS algorithm that gave an output with trips start and end points defined and a minute by minute latitude and longitude reading with an assigned transport mode when in a trip. The GPS temporal inconsistencies mentioned before were adjusted by comparing coordinate values of the first few hours of the measuring week (during the kick-off day) between the smartphone and the GPS and shifting the timestamp on the GPS to minimize the distance between the coordinates given by the two devices at the same time. The smartphone data was only used for this purpose since the short battery life of the smartphone made the GPS readings more useful for all other purposes (see footnote on Table 5.1).

The mode identified by PALMS and with the corrected time stamps was then corrected using the physical activity levels of the trip measured by the SenseWear: Trips identified as “vehicle” in the PALMS output but with high physical activity levels were assigned “cycle” and trips identified as “cycling” but sedentary in the SenseWear were assigned “vehicle”. This was done under the assumption that since PALMS had only speed data, it could assign “vehicle” to particularly fast cyclists or “cycle” to low-speed motorists, but that the SenseWear would still adequately measure their physical activity data.

To evaluate the performance of the three mode detection algorithms (i.e. PALMS-assigned with fixed time stamps, SenseWear output, and the combined method), their results were compared to the self-reported data from the activity diary, using the following six metrics based on (Donaire-Gonzalez *et al.*, 2016): i) sensitivity (sens): i.e. the percentage of recognised trip minutes over all trip minutes reported in the diary; ii) specificity (spec): i.e. the percentage of non-trip minutes recognised over all non-trip minutes reported in the diary; iii) the positive predictive value (PPV): i.e. the percentage of true trips identified over all data identified as a trip; iv) the negative predictive value (NPV): i.e. the percentage of true non-trips over all data identified as non-trips; v) the accuracy (acc): i.e. the percentage of all correctly identified trips or non-trips over all data collected; and vi) the F1 score: i.e. the harmonic mean of precision and sensitivity. Equations for each of these metrics are shown below. Note that these results are done only for the two days in the week for which an activity diary is filled out.

$$Sens = TP/(TP + FN) \tag{5.1}$$

$$Spec = TN/(TN + FP) \tag{5.2}$$

$$PPV = TP/(TP + FP) \tag{5.3}$$

$$NPV = TN/(TN + FN) \tag{5.4}$$

$$acc = (TP + TN)/(TP + FP + FN + TN) \tag{5.5}$$

$$F1 = 2TP/(2TP + FP + FN) \tag{5.6}$$

Where TP are “True positives”; FN are “false negatives”; TN are “True negatives; FP are “False positives”.

Given our particular interest in active modes, to assess the performance of the three mode-detection methods the values of these six metrics were compared for three types of trips: “All” (identifying trips in general); “Active” (identifying cycling and walking trip stages) and; “Cycling” (identifying cycling trip stages). A desired value of 0.3 was arbitrarily defined beforehand[1] as the minimum for all these metrics and the method with a higher number of metrics scoring 0.3 or higher was chosen as the most adequate one for mode assignment.

Inhaled doses

Daily inhaled doses were calculated using the same method proposed in chapter 4 defined by equation 4.3 and based on (de Nazelle, Rodríguez and Crawford-Brown, 2009). The only difference worth noting in this case is that the air pollution data is the *measured* BC instead of the modelled NO_2 or $PM_{2.5}$ from equation 4.1.

Data analyses

Descriptive statistics such as geometric means of BC exposures and physical activities were performed for each city and for the different seasons. The measured BC and physical activity data were tested for normality and lognormality using a Kolmogorov-Smirnov test ($\alpha = 0.05$) but these failed to reject the null

[1] The importance of defining a permissible value beforehand rests in the fact that it will avoid just using the “least bad” of all. However, as can be seen from the results below, the method selected made the cut by a very tight margin and if a higher value had been chosen, none of the methods would have approved this criterion.

hypothesis and thus, neither distribution can be assumed (Figure 5.5). From the results shown in chapter 4, exposures and inhaled doses in different transport modes are hypothesised to be significantly different so paired Mann-Whitney tests (α = 0.05) were done for exposures while in transport to test for this. Temporal trends of physical activity, BC exposures and inhaled doses were explored through the computation of geometric means for the same hours of the day for the entire data set.

Daily geometric mean BC exposures, daily geometric mean physical activity, and daily total inhaled doses of BC were calculated per individual for every day of measured days with more than 90% of recorded micro aethalometer data (i.e. with at least 1,296 minutes of data). Paired Mann-Whitney tests (α = 0.05) were also done for daily geometric mean exposures, daily geometric mean physical activity data, and total daily inhaled doses depending on transport mode most commonly used per day (defined as the mode with the longest time spent in it per day regardless of number of stages or trips).

Microenvironmental contributions to BC exposures and BC inhaled doses were computed based on time spent in each (i.e. “home”, “work/study”, “transport” and “other – place”), following the procedure explained in Chapter 4. These contributions were also computed for the different transport modes used. Locations for “home” and “work/study” were obtained directly from the PASTA survey that explicitly asks for these and were assumed to remain constant through the entire data collection period. All non-trip points within a 100m-radius of “home” or “work/study” locations were assigned these location tags and the rest were tagged with “other - place”, which would include, for example, recreation and social activities, both indoors and outdoors.

The only BC monitor that participants had was their personal exposure one so comparison with home-based exposures had to be done through modelling of the home-based BC values. However, since I had no access to an ambient air pollution model for BC for the three cities, the NO_2 LUR model used in Chapter 3

(Vienneau *et al.*, 2013) was used as a proxy and NO_2 values were extracted from there at home locations. A Spearman correlation test was done between NO_2 concentrations at home from the LUR model and the geometric mean of daily BC exposures. To check for the validity of this approach, hourly NO_2 and BC data was downloaded from an urban background monitoring station in London (North Kensington station) during the data collection period and the Spearman correlation coefficient between these two pollutants was calculated. This verification was only done for London as it is the only city with online access to the BC data for download (London Air and King's College London, 2018). Also, since the data was downloaded only for a background station, evaluation of other stations associated to specific sources (e.g. industrial or road-side stations) could help in further validation of this assumption in space and time but this was left for future studies.

Results

Sample

A total of 122 people were recruited and monitored in the three PASTA cities of the health add-on (40 in London, 41 in Antwerp and 41 in Barcelona). All participants were monitored for three weeks (with three exceptions that collected data for only two) recording physical activity levels, BC exposure levels, and geoposition. Our sample had 55% women and the median age was 34 (Figure 5.3). Despite our efforts to fulfil the inclusion and exclusion criteria, two participants had a self-reported BMI higher than the established limit of 30 (30.4 and 31.1) as the recruitment phase was coming to an end and we were struggling to find participants. There is no evidence to suggest that this should affect the results presented here and thus, these have been included in all the analyses.

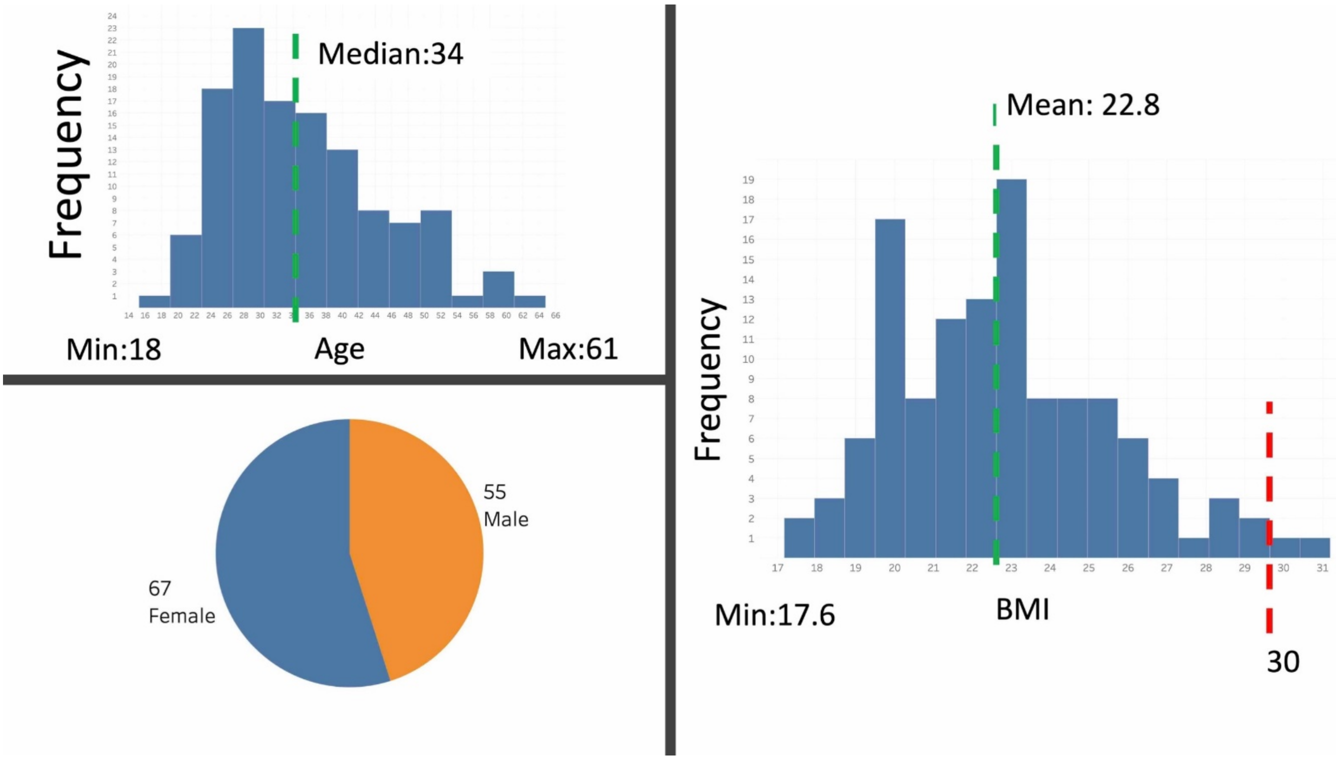


Figure 5.3. Age, sex, and BMI distribution of the sample. As a useful comparison, median ages in Belgium, Spain and the UK in 2015 were 41.4, 42.4 and 40.0 respectively. In terms of BMI, in 2014 13.2%, 14.9% and 20.5% of adults between 18 and 64 in those same countries in the same order had a BMI higher that 30 (obese)[2].

Our sample was particularly biased towards active travel based on their answers to the PASTA survey but again, this was difficult to control for as the recruitment phase came to an end. Figure 5.4 shows the answers to the question: *How often do you currently use each of the following methods of travel to get to and from places?* In our sample, out of the 121, 111 said they walked daily or between 1 and 3 days a week and 69 said they cycled daily or between 1 and 3 days a week, while only 23 (19% of the sample) said they never cycled. At a city level, for contrast, in London and Barcelona, only 2% and 1.7%[3] of the trips respectively were done cycling in 2015. Furthermore, in London, only 1 in 6 people cycle at least once a year and out of those who do cycle, close to 80% cycle only occasionally (less than once a week) according to Transport for London (Transport for London (TfL),

[2] Data from *Eurostat* available at https://ec.europa.eu/eurostat/data/database
[3] Data from *European City Ranking 2015, Best practices for clean air in urban transport* available at http://www.sootfreecities.eu/sootfreecities.eu/public/city

2018). The use of public transport and car is more evenly distributed as shown in Figure 5.4.

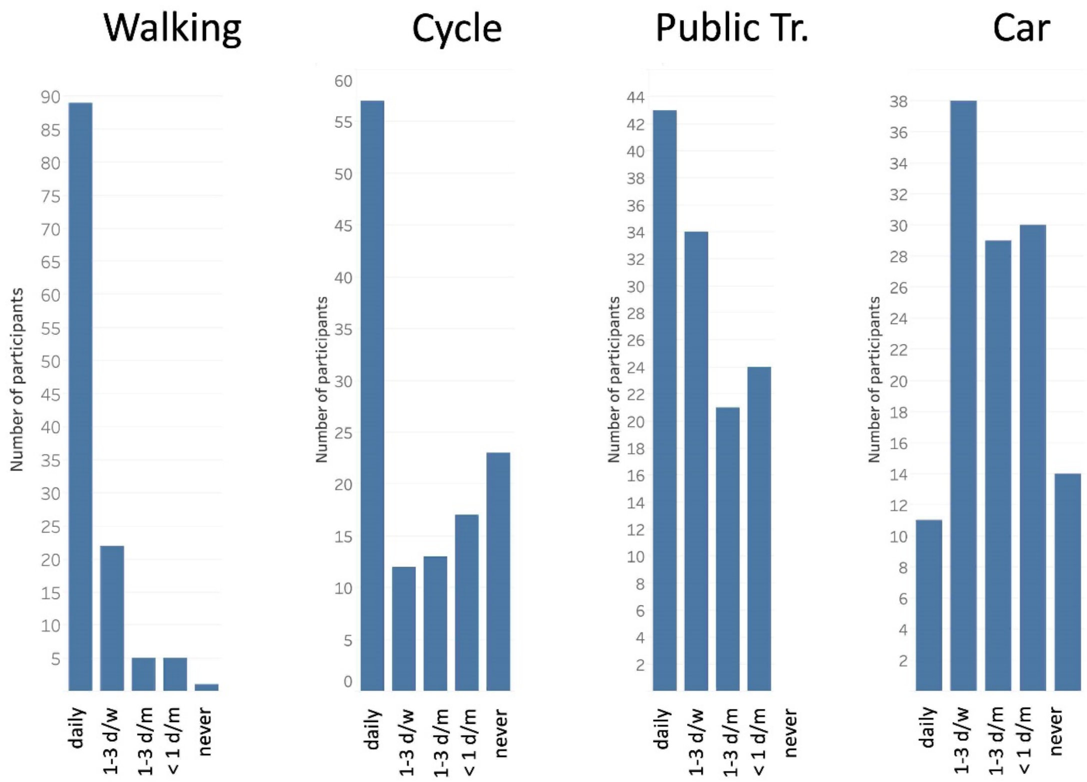


Figure 5.4. Distribution of frequency of mode use based on PASTA survey answers (d/w stands for days per week and d/m stands for days per month). Total sample size = 121.

As mentioned before, Kolmogorov-Smirnov tests were performed in order to assess if the distributions of the BC exposures, physical activity (METS), and inhaled doses were normally or lognormally distributed. However, these failed to reject the null-hypothesis and thus, neither of these distributions can be assumed, as can be suspected from the histograms of the logarithms of these shown in Figure 5.5. The statistical tests performed on the results are therefore all non-parametric tests.

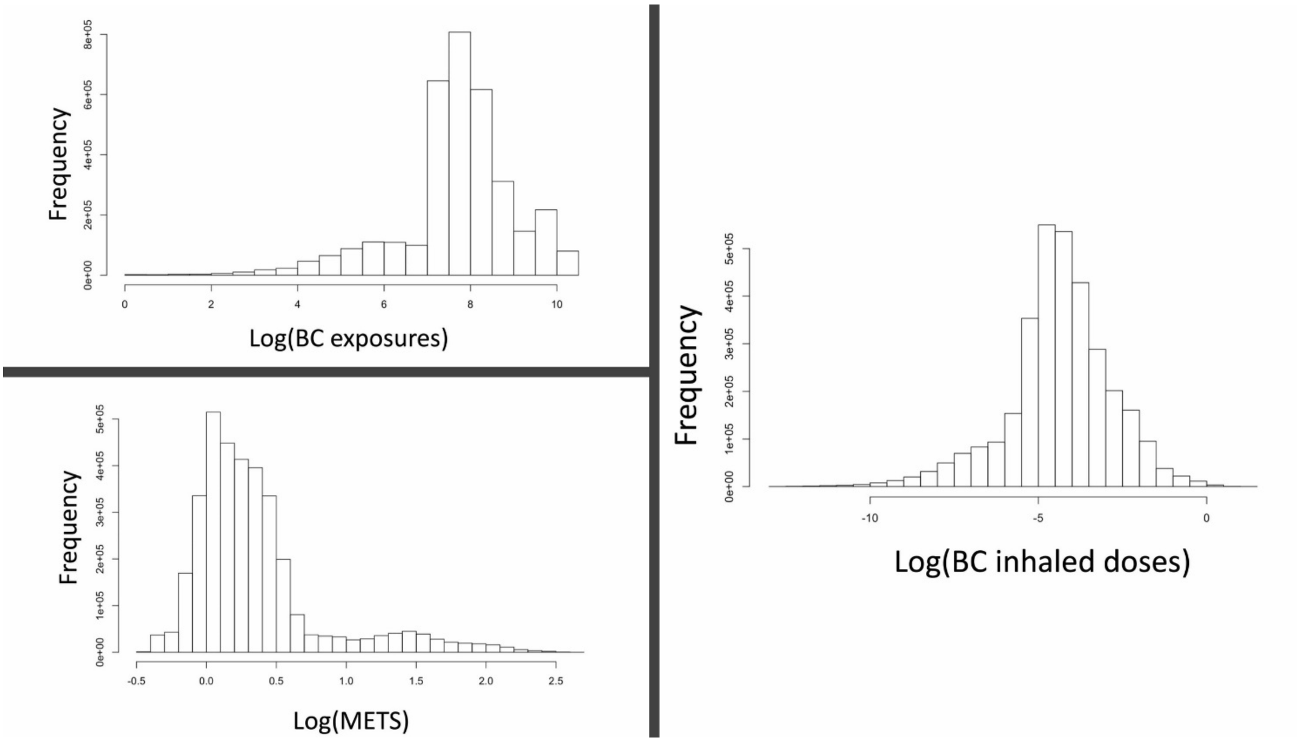


Figure 5.5. Histograms of the logarithms of BC exposures, METS, and inhaled doses. The distributions are not lognormal (Kolmogorov-Smirnov test, p-value > 0.05 in all three cases)

Identifying the mode used

Based on the results of the six metrics defined to evaluate the best mode detection algorithm, I chose the "combined" method based on the data from PALMS corrected by physical activity levels from the SenseWear (Figure 5.2). As shown in Table 5.2, the combined method performs better than the other two methods in all six indicators and for the three types of trips (all trips, active trips and cycling trips). For example, the combined method correctly identifies 38% (Sens = 0.3801) of all trips mentioned in the activity diary, while the SenseWear only identifies 20% (Sense =0.1983) and PALMS 27% (Sense = 0.2740) of them. In all cases, most of non-trips are correctly identified as shown by the high values of indicators based on these (Spec and NPV), since most of the time people are not moving and the analysis is on a minute by minute basis. These values are much lower than what would be ideal and highlight the importance of developing more robust mode detection algorithms. This is one of the main shortcomings of this study and is discussed further at the end of the chapter. However, values do not fall far from a similar exercise presented by Donaire and collaborators (Donaire-Gonzalez *et al.*, 2016). In the end, trips where divided into four modes: walk, cycle, vehicle (which includes motorized vehicles such as car, bus, and overground trains), and unknown (which will cover any other unidentified modes).

Table 5.2. Comparison of sensitivity and specificity in identifying trips across three mode identification methods (based on: SenseWear activity recognition; the PALMS mode identification software; a combined approach (see Figure 5.1) for three categories of trips: all trips, walking and cycling trips only, cycling trips only.

Method	Type of trip	Sens[a]	Spec[b]	PPV[c]	NPV[d]	ACC[e]	F1 score[f]
SenseWear	All	0.1983	0.9257	0.1937	0.9277	0.8656	0.1959
	Active	0.1709	0.9497	0.1410	0.9595	0.9138	0.1545
	Cycling	0.1037	0.9905	0.1620	0.9842	0.9750	0.1264
PALMS	All	0.2740	0.9483	0.3137	0.9381	0.8948	0.2925
	Active	0.1742	0.9720	0.2198	0.9630	0.9375	0.1943
	Cycling	0.1696	0.9879	0.1900	0.9862	0.9746	0.1792
Combined	All	0.3801	0.9593	0.4426	0.9480	0.9140	0.4090
	Active	0.2448	0.9758	0.3127	0.9663	0.9442	0.2746
	Cycling	0.2338	0.9892	0.2657	0.9872	0.9768	0.2487

[a]Sens: Sensitivity. [b]Sepc: Specifity. [c]PPV: Positive predictive value. [d]NPV: Negative predictive value. [e]ACC: Accuracy. [f]F1 score. As defined by equations 5.1 – 5.6

Minute-by-minute geometric means and temporal trends

Of all cities and seasons combinations, winter exposures in London are the highest, and mid-season (spring and autumn) in the Barcelona are the lowest (Table 5.3). In London and Antwerp, the winter season yields highest exposures of the three seasons, while in Barcelona summer exposures are highest. In fact, the highest summer exposures across the 3 cities are found in Barcelona, where the summer exposure also yields the highest variability across all seasons and city combinations, as indicated by the geometric standard deviations.

Table 5.3. Geometric means of BC concentrations per city per season

City	Season	BC geometric mean (ng/m^{-3})	BC geometric SD
Antwerp	Summer	2,038.0	3.73
	Winter	2,227.2	3.90
	Mid (spring and autumn)	1,930.2	3.79
Barcelona	Summer	2,310.1	4.33
	Winter	2,015.0	3.91
	Mid (spring and autumn)	1,720.7	4.00
London	Summer	2,178.8	3.77
	Winter	3,098.9	3.64
	Mid (spring and autumn)	2,173.2	3.84

When the data from the different transport modes and microenvironments is analysed on a minute by minute basis, BC concentrations are highest in vehicles with average exposure concentrations of 3,483 ng/m^3 followed by cyclists that are on average exposed to 3,033 ng/m^3 while in transport. Vehicle users are in average exposed to BC concentrations 36% higher than those from pedestrians. As expected, the lowest average concentrations were found when stationary as shown in the violin plot of Figure 5.6. The violin plots (Figures 5.6 to 5.8) show the median of the results (white dot) and the 95% confidence interval (black bar), while showing a probability density curve indicating the distribution of the data.

As expected, physical activity levels are substantially different while in different travel modes, with cyclists having average METS 16% higher than those of pedestrians and more than twice from people in vehicles. Note that a change from, for example 1.5 METS (resting in a vehicle) to 3 METS (light to moderate cycling) could imply an increase of 360% in oxygen intake and a 170% increase in heart rate (Ohashi, Kamioka and Matsuoka, 2002). The cycling and walking modes display two clear local maxima, which may indicate how those modes often entail periods of rest, or that two distinct distributions of active travellers exist.

The combination of high BC exposures and higher physical activity results in cyclists inhaling 46% more BC per minute than pedestrians and 2.5 times more than vehicle users. Pairwise Kolmogorov-Smirnov tests showed that differences between all modes for BC exposures, physical activity levels, and minute inhaled doses were all statistically significant ($p \ll 0.001$).

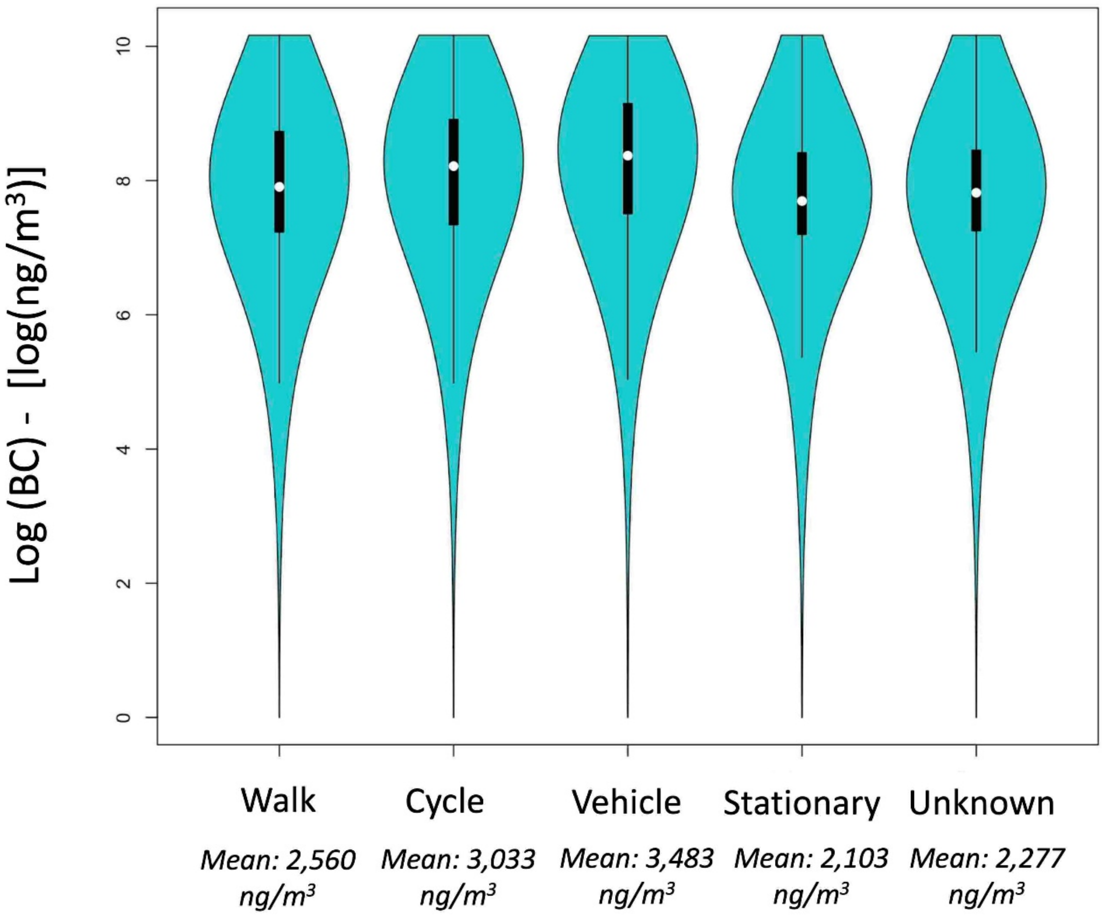


Figure 5.6. BC exposures while in transport or stationary. The y-axis is in a logarithmic scale so that differences are clear to see, and means are given for each mode to give a sense of the real values.

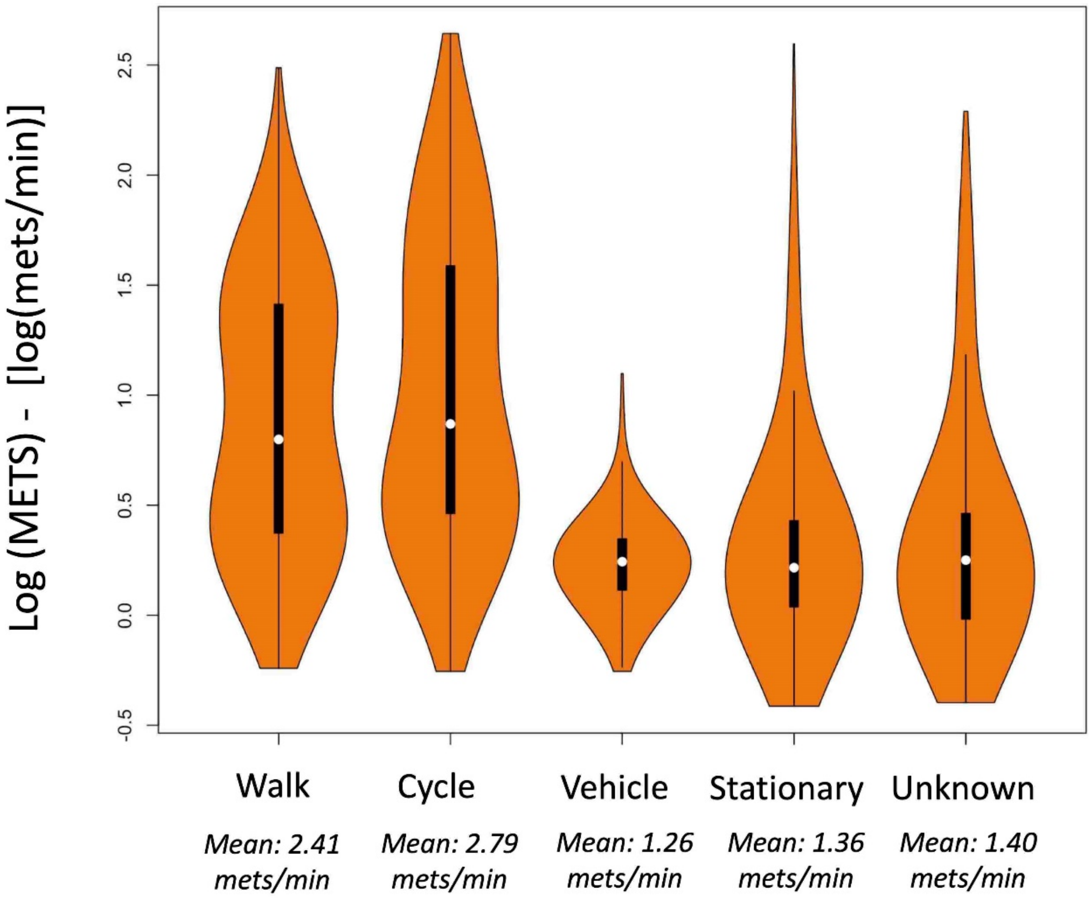


Figure 5.7. Physical activity (METS) while in transport or stationary. The y-axis is in a logarithmic scale so that differences are clear to see, and means are given for each mode to give a sense of the real values.

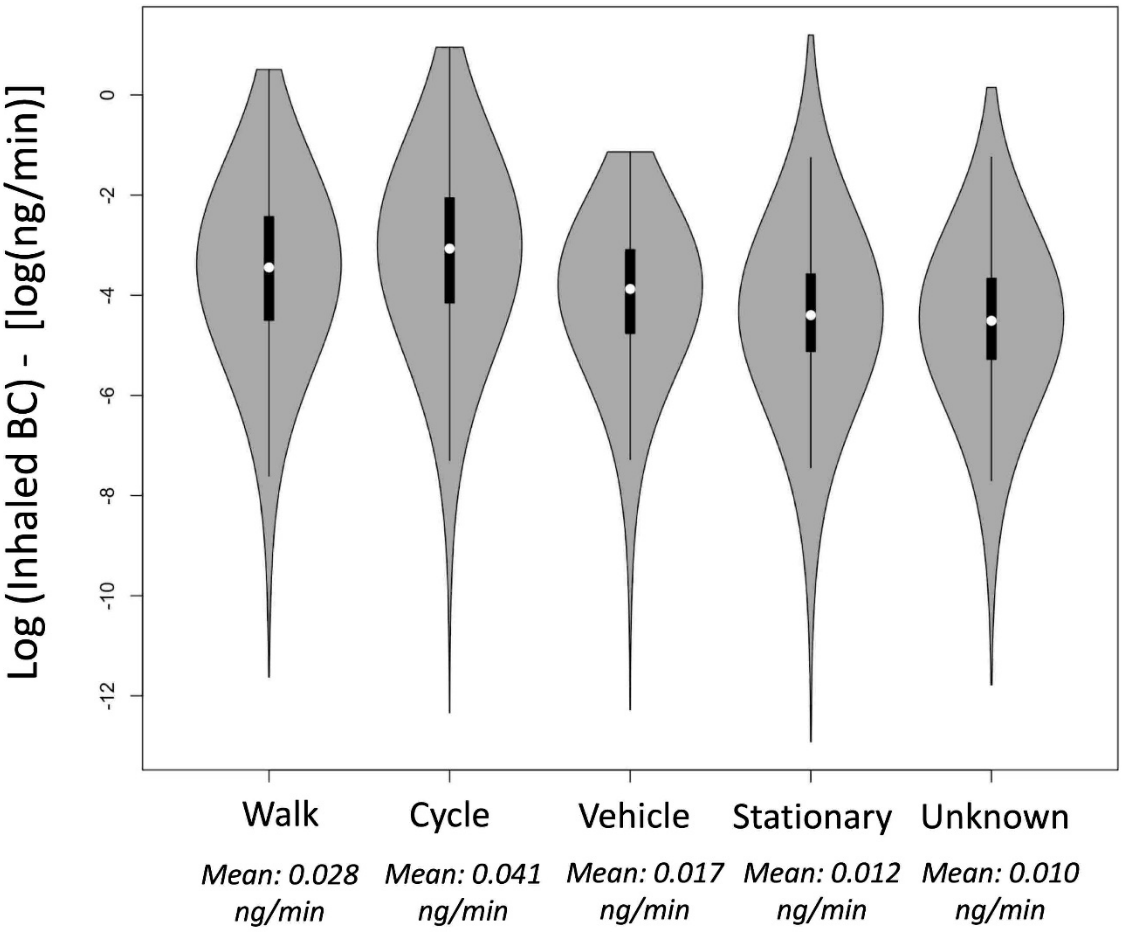


Figure 5.8. BC inhaled doses per minute while in transport or stationary. The y-axis is in a logarithmic scale so that differences are clear to see, and means are given for each mode to give a sense of the real values.

Geometric means of the minute-by-minute BC exposures, METS and BC inhaled doses were aggregated by the hour of the day and the results are shown in Figure 5.9. These results are also compared to BC ambient air concentration data from London, the only city for which I had hourly information of the urban background station. The results show that BC exposures have a temporal pattern with two local maxima, but the second peak exposures of the day happen earlier than background concentrations, closer to 1:00pm, in contrast to background concentrations that have the second peak close to 8:00pm. The early peak in exposures could be associated to activities taking place in the outdoors in places of high concentration levels at lunch time, which would be supported by the relatively high physical activity levels seen throughout the day. As I have mentioned before, the sample is biased towards particularly active participants that may use the early afternoon for their exercise. However, peak physical activity happens at around 5:00pm, time at which background concentrations are also higher than at 1:00pm, while the BC exposures between 4:00pm and 7:00pm show a steep decline which is not easily explainable. One possible explanation is that the effect of the afternoon commute from highly polluted city centres to the residential areas result in an overall decline in exposures in the afternoon greater than the effect observed in the background station. Another possible reason for the apparent early fall of personal exposures compared to background concentrations, is the natural delay in the background station with respect to what is happening in the city centre and closer to congested streets. A map of London's NO_2 spatial distribution in 2013 with the location of the urban background station based on the information provided by London Air (London Air - Kings College London, 2018) is shown in Figure 5.10 to further illustrate this point. High concentrations associated to a rise in emissions from in-city sources will naturally take longer to reach background stations than personal monitors on the road for example. Finally, it would be desirable to compare BC background station behaviour in London with data from the other two cities to confirm they have similar behaviour. For now, the discrepancies between personal exposures and background readings will be considered to be the result of a combination of the arguments above and further exploration of this will be needed in future work.

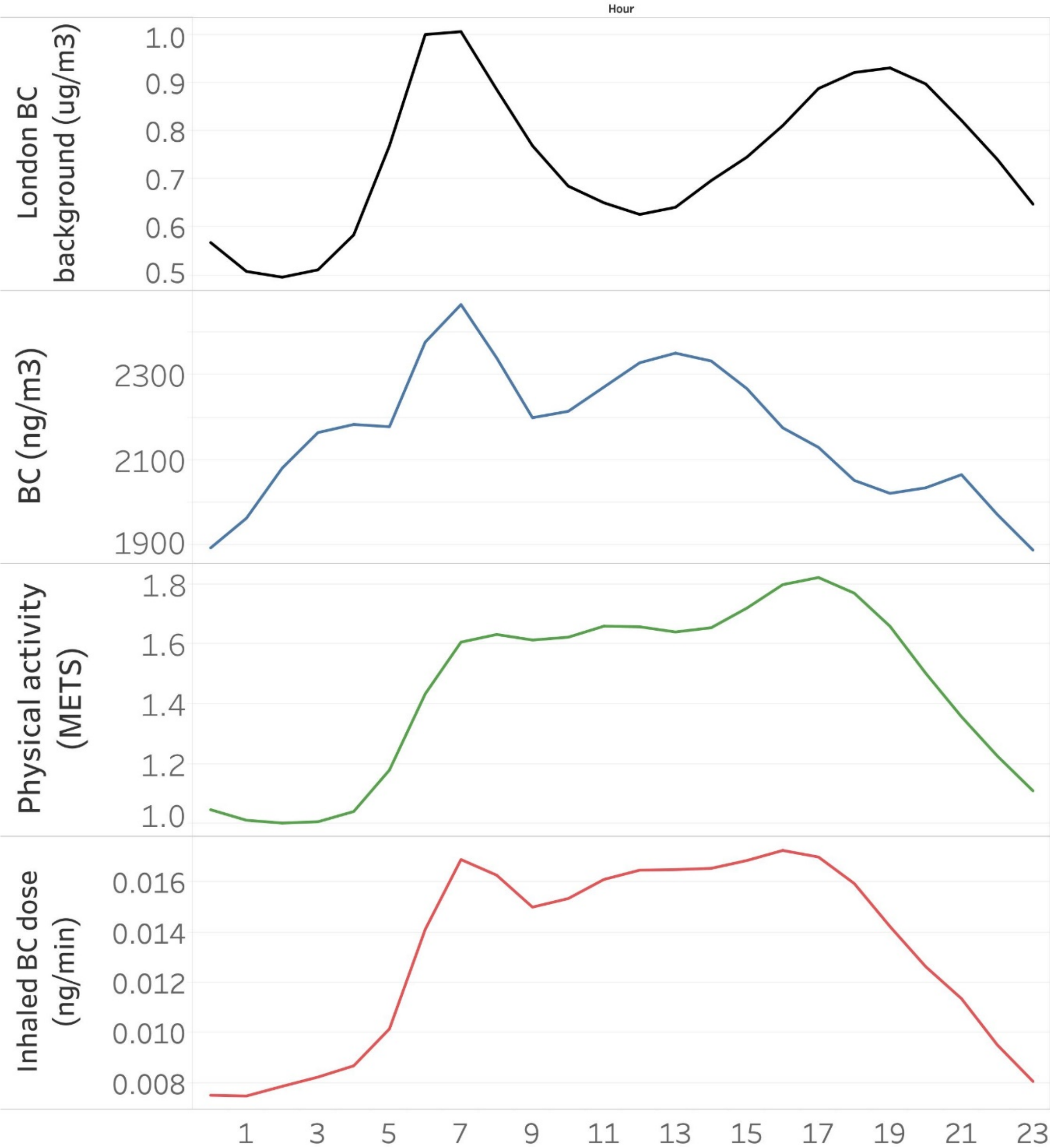


Figure 5.9. Temporal trends of BC background concentrations (only for London), BC exposure, physical activity and BC inhaled doses. Geometric means aggregated by hour of the day for all days in the data collection period.

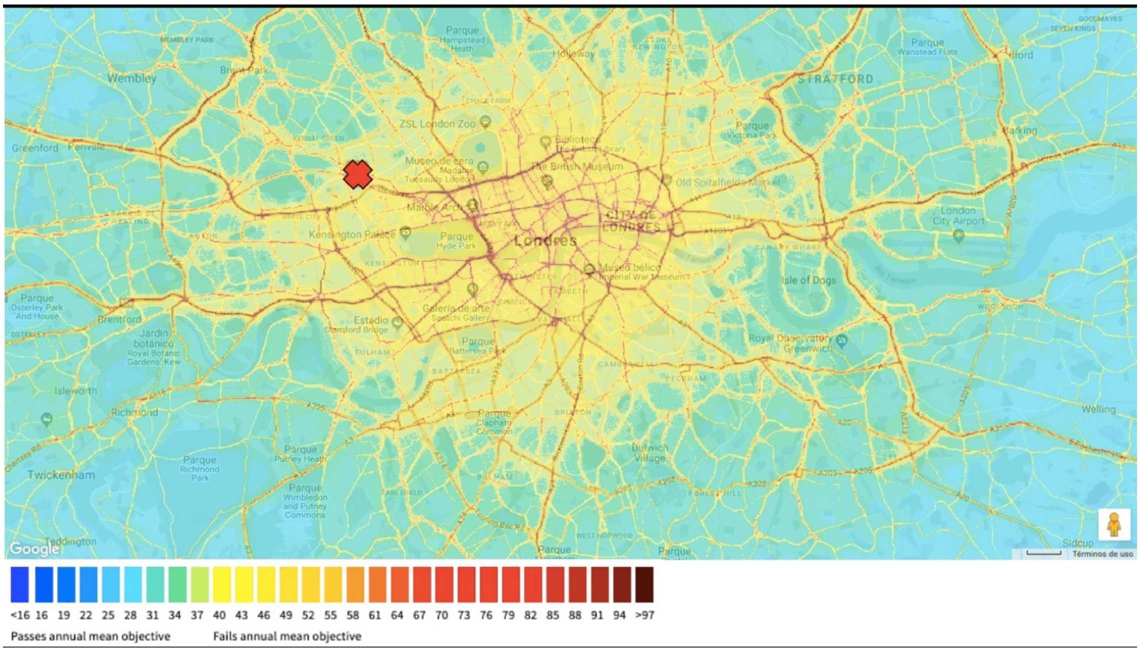


Figure 5.10. Location of the urban background station used (denoted by the red X) in the London annual NO_2 concentration map provided by London Air (London Air - Kings College London, 2018)based on 2013 data. The colour scale is in $\mu g/m^3$ of NO_2.

Daily results and microenvironmental contributions

The differences in BC exposures, physical activity and BC inhaled doses can also be observed for different mode users when daily geometric means (exposures and physical activity) and totals (inhaled doses) are computed. The results show that vehicle users were exposed to daily geometric mean exposures of BC that were 3.7% and 3.0% higher than those experienced by pedestrians and cyclists respectively (Figure 5.11). Participants who used the cycle more than any other mode, had daily inhaled doses that were 10.7% and 12.2% higher than those calculated for pedestrians and vehicle users respectively (Figure 5.13).

When aggregated to daily geometric means and totals, the differences are not always significant (Figures 5.11 – 5.13). The distributions of the data in “vehicle” and “unknown” stand out as having less variance for all three outcomes (BC exposure, physical activity, and BC daily inhaled doses) which could be related to less route choices for example, and lower impact of background concentrations. However, it is important to remind the reader that the sample was shown to be bias towards walking and cycling (Figure 5.4) and these tighter dispersions could just be the product of smaller samples.

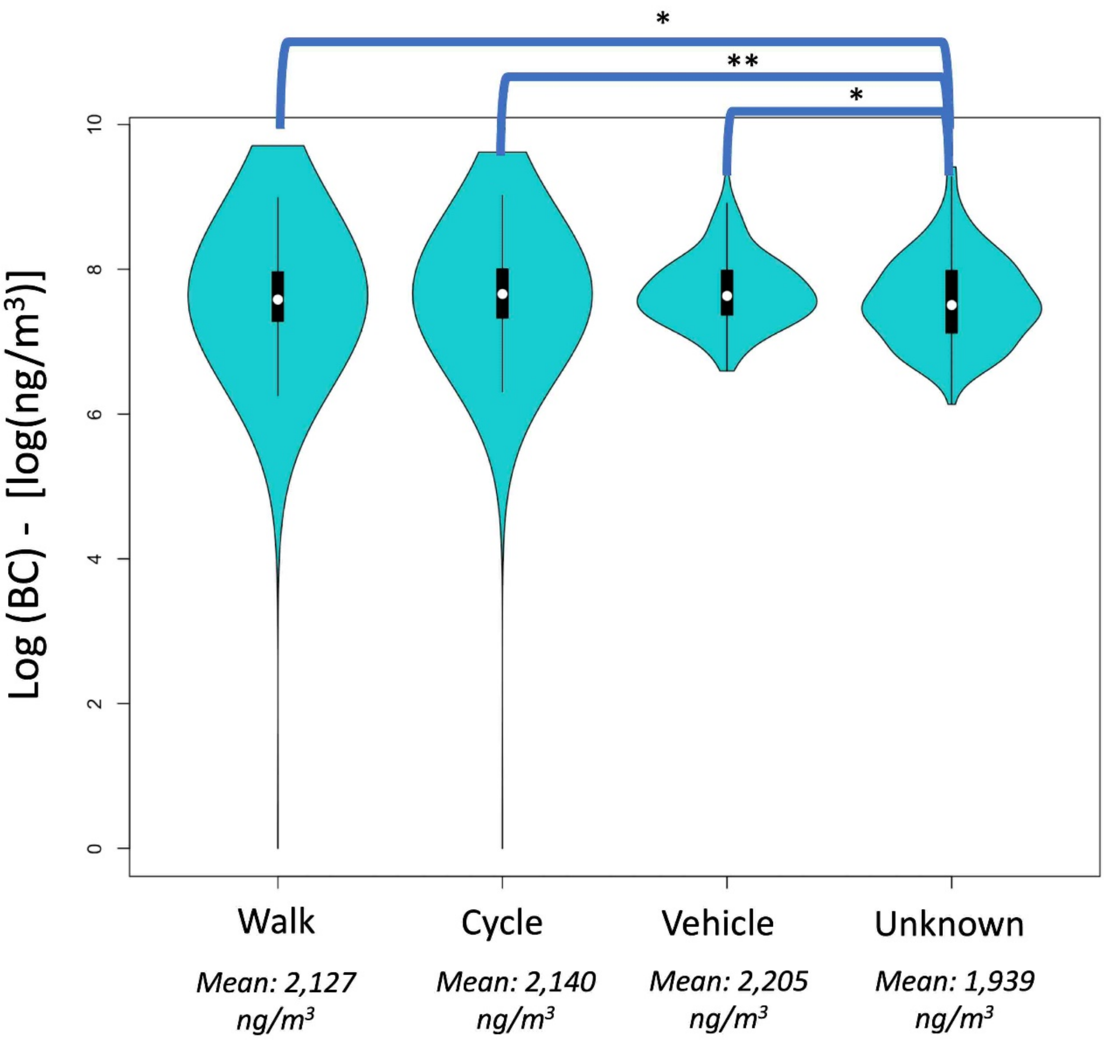


Figure 5.11. Daily geometric mean BC exposures divided by mode most commonly used per day. The y-axis is in a logarithmic scale. Means are given for each mode and the top square brackets indicate the level of significance of the differences between modes using a Kolmogorov-Smirnov test (*: p-value ≤ 0.05; **: p-value ≤ 0.01; ***: p-value ≤ 0.001; ****: p-value ≤ 0.0001)

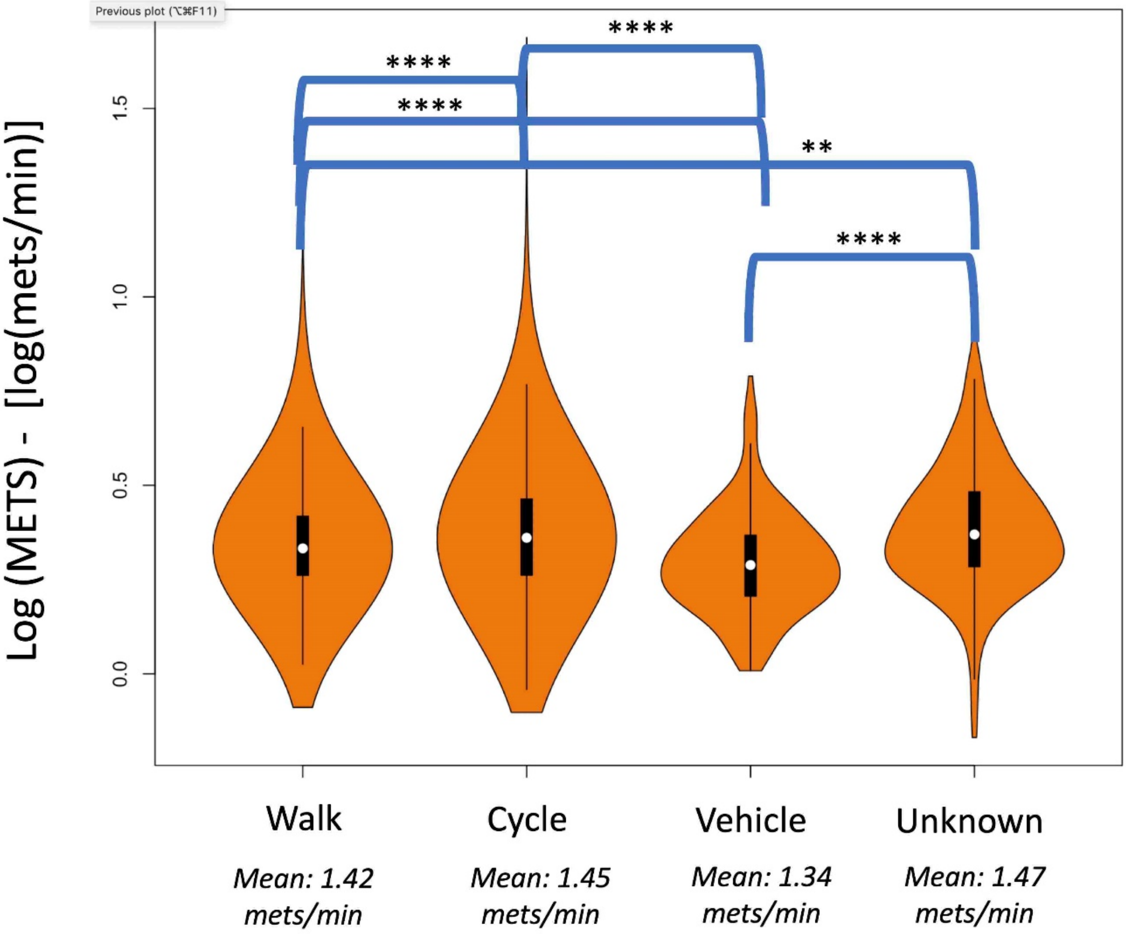


Figure 5.12. Daily geometric mean of physical activity (METS) divided by mode most commonly used per day. The y-axis is in a logarithmic scale. Means are given for each mode and the top square brackets indicate the level of significance of the differences between modes using a Kolmogorov-Smirnov test (*: p-value ≤ 0.05; **: p-value ≤ 0.01; ***: p-value ≤ 0.001; ****: p-value ≤ 0.0001)

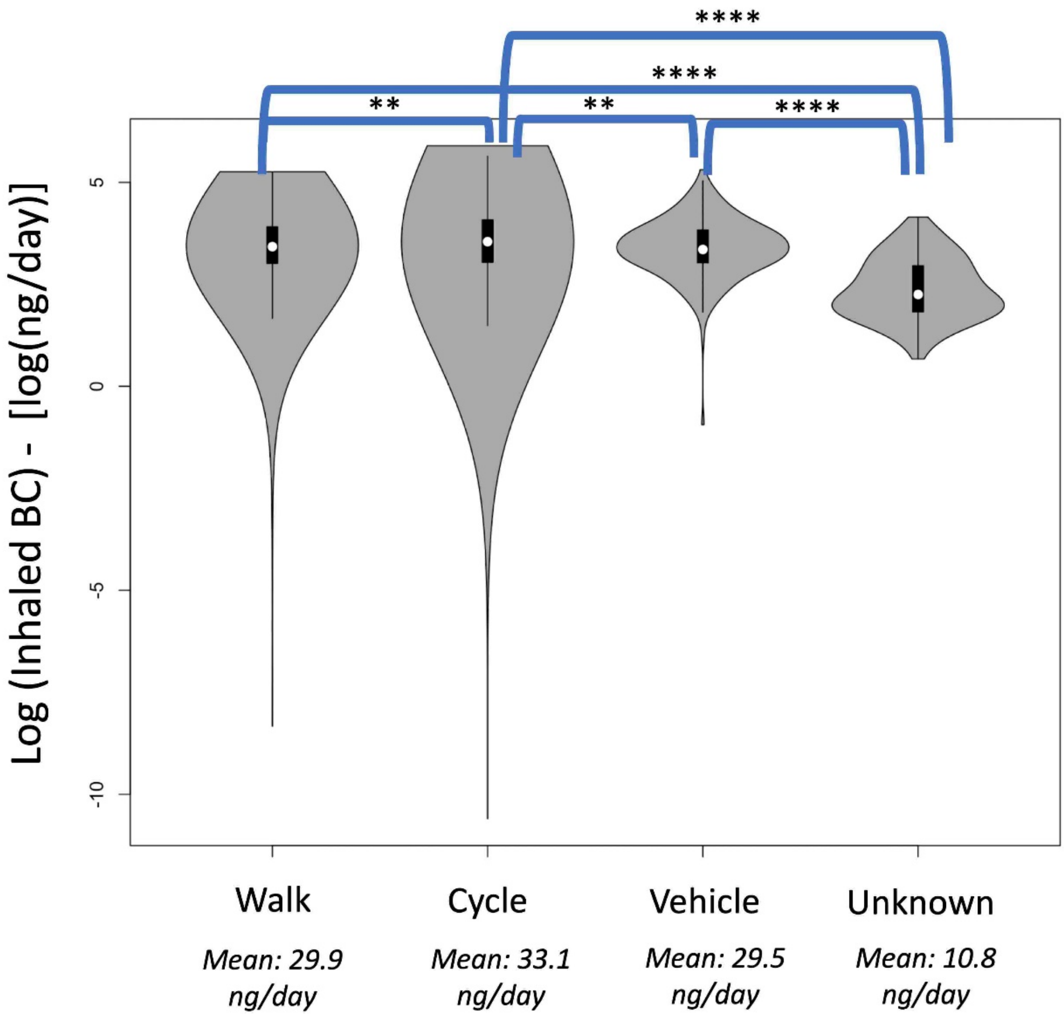


Figure 5.13. Total daily BC inhaled doses divided by mode most commonly used per day. The y-axis is in a logarithmic scale. Means are given for each mode and the top square brackets indicate the level of significance of the differences between modes using a Kolmogorov-Smirnov test (*: p-value ≤ 0.05; **: p-value ≤ 0.01; ***: p-value ≤ 0.001; ****: p-value ≤ 0.0001)

As in chapter 4, these differences in transport modes hint towards transport microenvironments playing an important role in daily results so percentages contributions of “home”, “work”, “transport” and “other” were calculated. In this case, I found that although on average people spend 7.5% of their time in transport, this microenvironment contributes with 10.5% to daily BC exposures and 18.1% to daily inhaled doses (Figure 5.14). These results fall more in line with previous studies mentioned before (Dons *et al.*, 2011; de Nazelle *et al.*, 2012) than the ones calculated for NO_2 and $PM_{2.5}$ in chapter 4. However, the biggest change is observed for environments classified as “other” that although they account for only 7.5% of the time spent, their contribution to BC exposures is as much as

50.2% on average and 48.2% to daily inhaled doses. Without more detailed feedback from participants on those other activities, it is difficult to hypothesise on the reasons for this. One reasonable explanation could be that transport on the underground metro systems is not classified as “transport” by the algorithms but as “other-place”. In the underground, BC exposures are overestimated due to resuspended iron from the rails and thus, contributions would be artificially higher. Despite efforts made to determine underground trips, this was not possible for this study. This in part would also explain the comparatively low results for sensitivity (sens) and positive predictive value (PPV) shown in Table 5.2. Another hint pointing towards the classification “unknown” including underground trips is the reduction in average contributions from 50.2% for BC exposures to 48.2% in inhaled doses. This suggests low levels of activity while in other microenvironments so leisure sports can be ruled out.

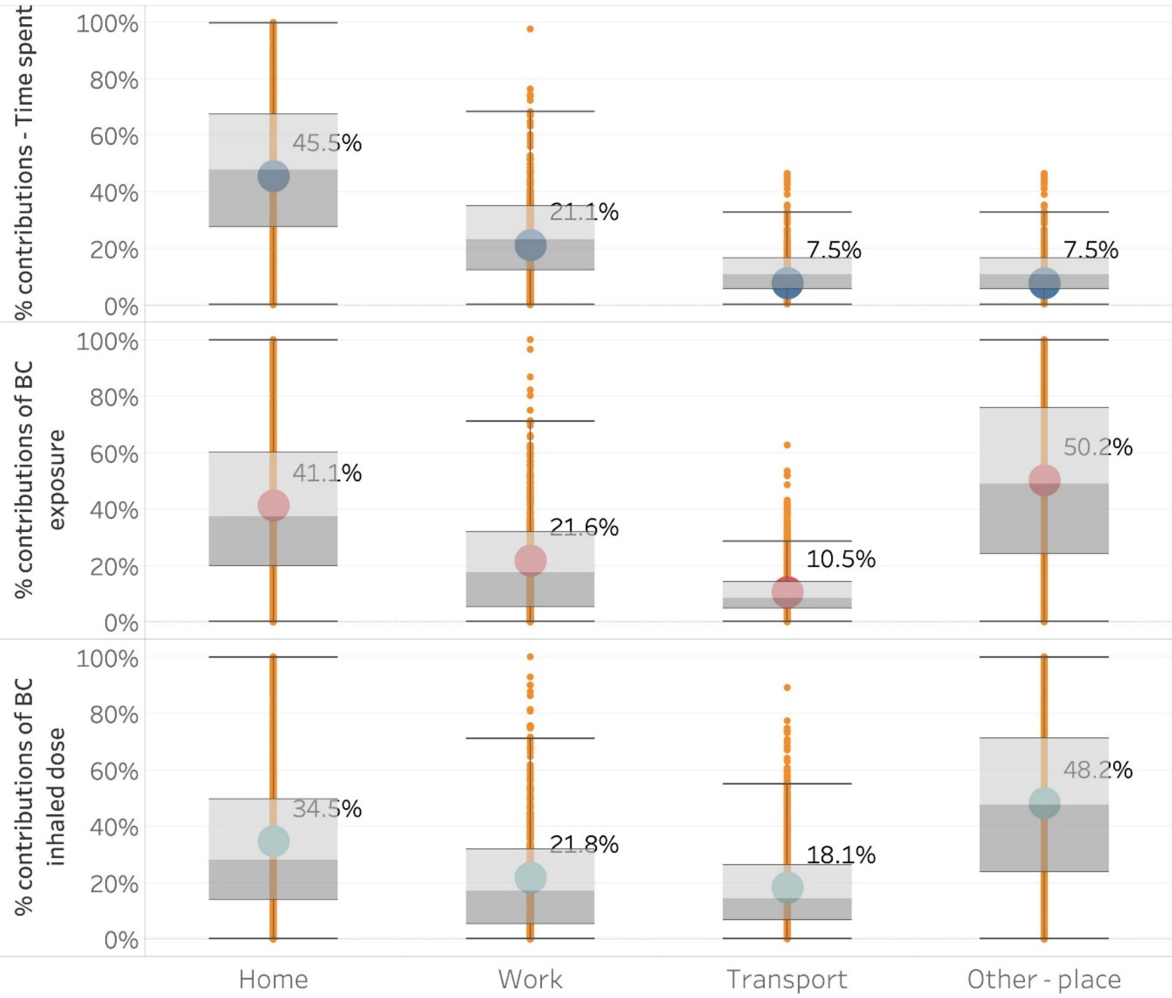


Figure 5.14. Percentage contributions of BC exposures and inhaled doses of different microenvironments compared to the percentage time of the day spent in each. Dots and labels indicate the averages.

Separating the above results for transport into the different modes highlights the role of cycling in the daily contributions. On average, cyclists spend 4.4% of their day cycling, and that time accounts for 6.5% of their daily exposures and 13.3% of their daily inhaled doses (Figure 5.15). This contribution to inhaled BC doses is almost twice as pedestrians (6.9%) and more than five times as vehicle users (2.5%). The activity classified as “unknown” is clearly a case of misclassification with participants spending 32.3% of their time in it, which is more than 7 hours per day, and yet, contributing with only 12.9% to total inhaled doses. However, it is important to note here the number of data points associated to this mode, shown by the colour scale in Figure 5.15: while for walking and cycling averages are calculated from more than 1,000 points (trip stages), the mode “unknown” only has 6. This means on the one hand that the influence of this mode on overall averages is very limited, and on the other, seems unlikely that these are trips from the underground systems as the self-reported frequencies of use of public transport in Figure 5.4 would suggest many more trip stages done in this mode.

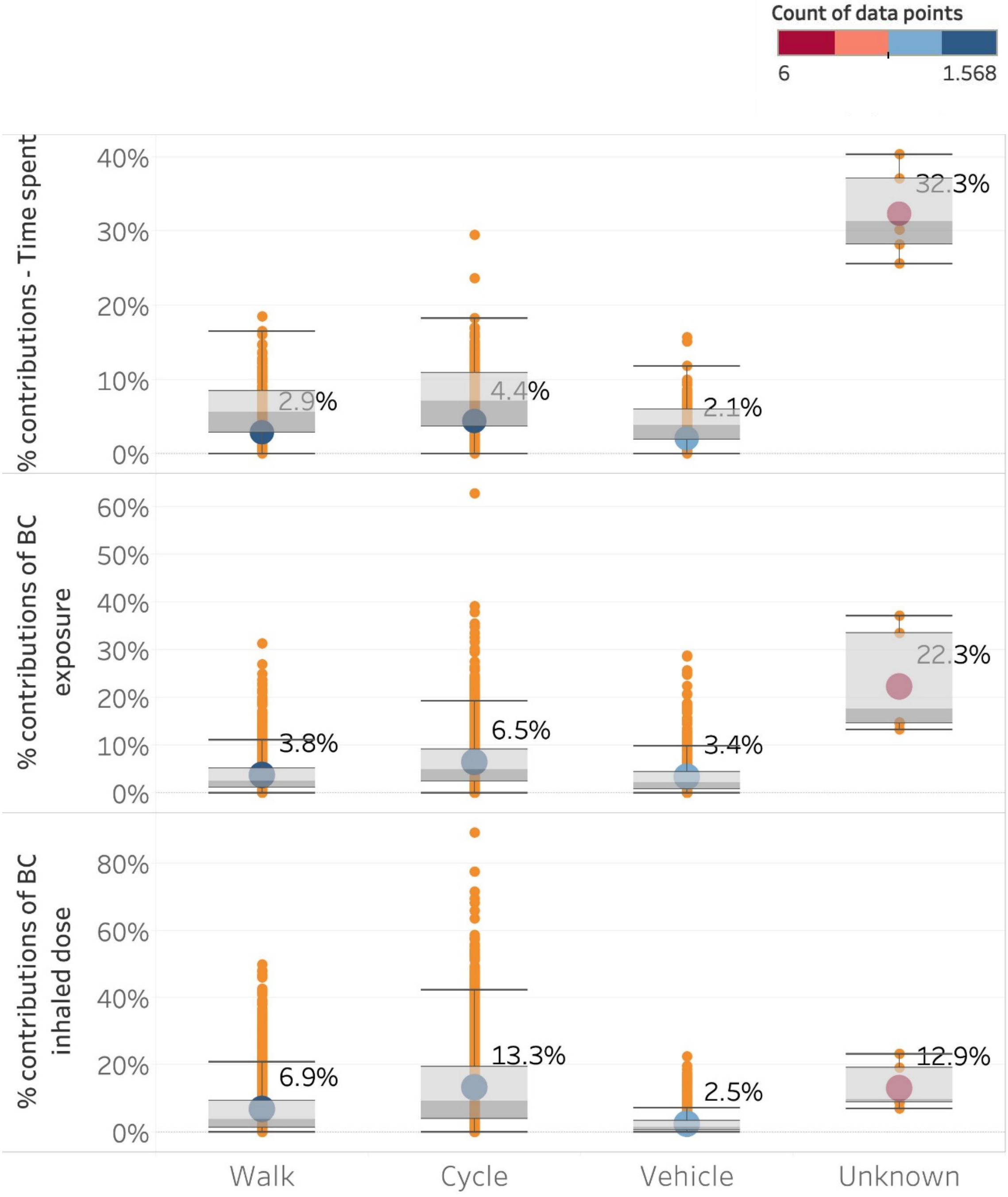


Figure 5.15. Percentage contributions while in transport of BC exposures and inhaled doses of different modes compared to the percentage time of the day spent in each. Dots and labels indicate the averages. The colour of the dots indicates the amount of data available for each average as shown in the colour scale.

Comparisons with home-based exposures

As explained in the methods, comparisons between home-based exposure estimates and the measured BC exposures was done using the NO_2 assigned concentrations from the LUR model used in chapter 3 (Vienneau *et al.*, 2013) as I know of no BC model for the three cities and the participants only had a personal monitor with them and no monitor at home. In order to check the validity of NO_2 as a proxy to BC, daily means from urban background stations for these two pollutants in the three cities were obtained and a Spearman correlation coefficient between NO_2 and BC background readings was found to be 0.80 (Figure 5.16).

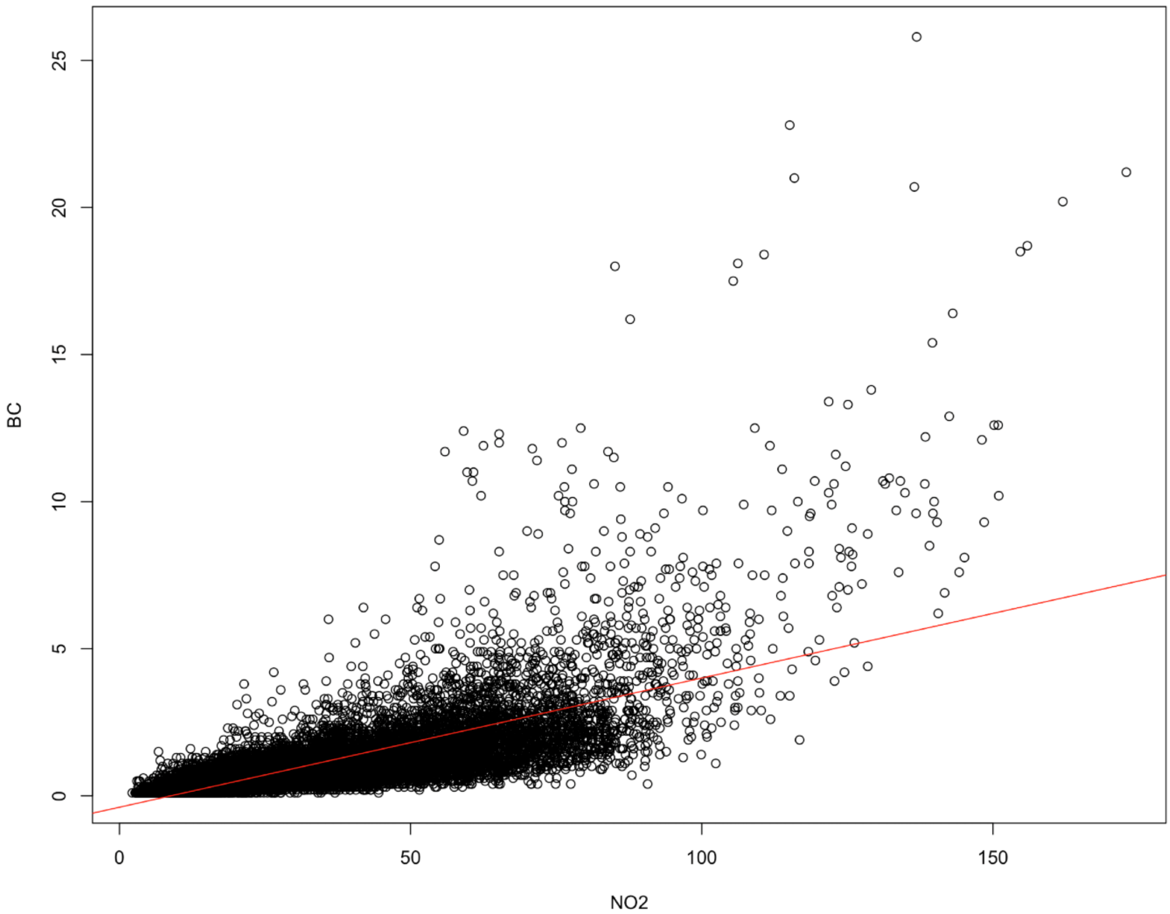


Figure 5.16. Daily BC and NO_2 concentration means from urban background stations in the three cities. The trend line in the correlation plot suggests a systematic error in NO_2 concentrations higher than 100ug/m^3 but this bias has not been explored further.

Despite this, there was no correlation between measured daily geometric mean exposures of BC and modelled NO_2 concentrations at home location (Spearman coefficient = 0.03, p-value = 0.76). There are of course many reasons that could explain this low correlation. First, NO_2 and BC are only partially correlated at the background station (R^2 0.8); second, the NO_2 model itself is also only partially correlated with NO_2 measurements themselves (R^2 between 0.5 and 0.7 depending on the source, see chapter 4 for details); third, and the main argument being made here, because of a lack of correlation between personal monitoring and ambient air concentrations at home locations. At this stage it is difficult to say how much of the lack of correlation observed can be associated to each source of uncertainty. In part, this argues in favour of developing ambient air concentration models for BC as it is a widely used pollutant for personal exposures, but the interpretation of such results will be limited without a reliable model for larger scales.

These results also contrast with the high correlation shown in Figure 4.4 in the previous Chapter, where modelled ambient air NO_2 concentrations at home correlated well (Pearson 0.84) with modelled personal exposure to NO_2. The question would then be, which one is right, and the best answer is neither. Figure 4.4 is probably an overestimation of correlations between personal exposures and ambient air concentration as in that case, both are modelled concentrations, and one is an endogenous variable of the model that estimates the other. However, Figure 5.16 is an underestimation due to the first two uncertainties mentioned before.

In an attempt to explore further any patterns between personal exposures and home locations, Figures 5.17 – 5.20 were developed. Figure 5.17 shows a dispersion plot comparing the geometric mean of the measured BC personal exposures by each participant throughout the entirety of their measuring campaign and the modelled NO_2 ambient air concentrations at their home on those same days. From this plot, no clear patterns emerge besides some clear clusters associated to the city. This further supports the idea that general background trends have a strong influence on personal exposures, stronger than intra-city variability measured at home location. The maps shown in Figures 5.18 to 5.20 show the geometric mean of all data points collected by participants at their home location. The maps seem to confirm that there is no obvious pattern linking BC exposures (bubble colours) or inhaled doses (bubble size) with home location in any of the three cities. The maps show no visually evident spatial patterns as the ones typically seen in ambient air concentration models. Note for example how the dark red points in Barcelona and London are not in the city centre, as dispersion and LUR models suggest, but in the periphery of the city centre, probably due to these participants having longer commutes for example. In Antwerp, the participant with the highest exposure mean is surrounded by participants with some of the lowest ones.

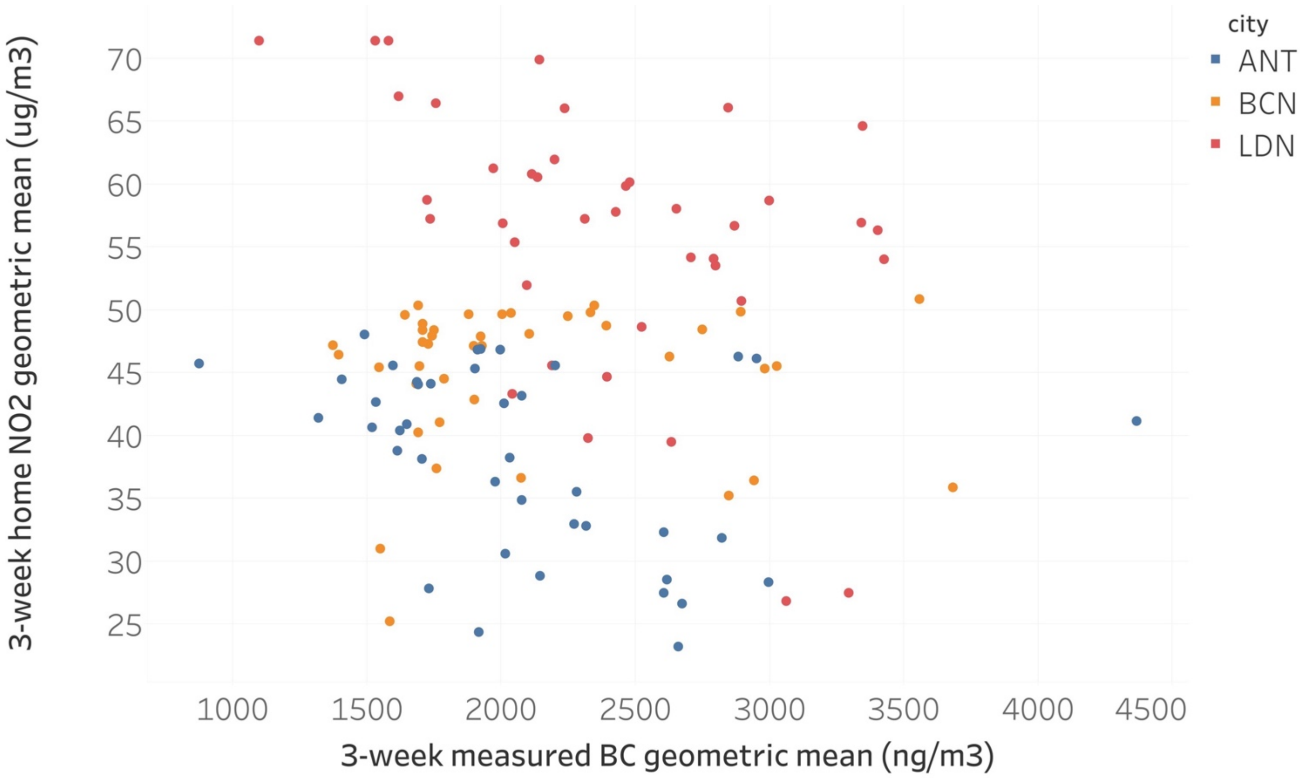


Figure 5.17. Dispersion plot of three-week geometric means of home NO_2 concentrations from LUR model corrected by temporal variations and geometric means of measured BC per participant. Both home NO_2 and measured BC means are computed for the same dates of the three weeks of measuring. Every dot represents one participant, and these are colour-coded by city.

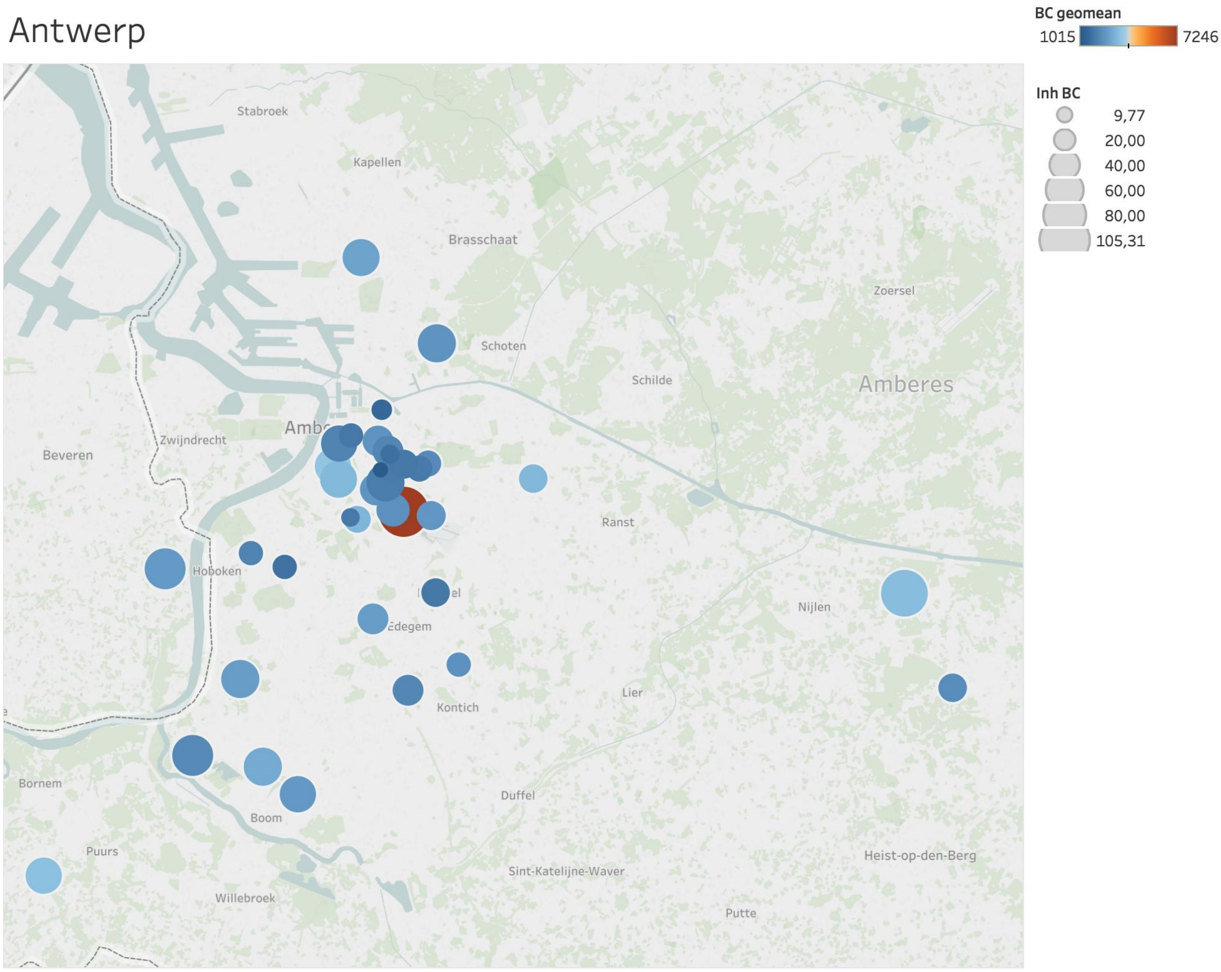


Figure 5.18 Home address distribution of participants in Antwerp. The colour represents 3-week geometric averages of daily BC exposures (ng/m^3) and the bubble size represents 3-week geometric averages of daily BC inhaled doses (ng/day).

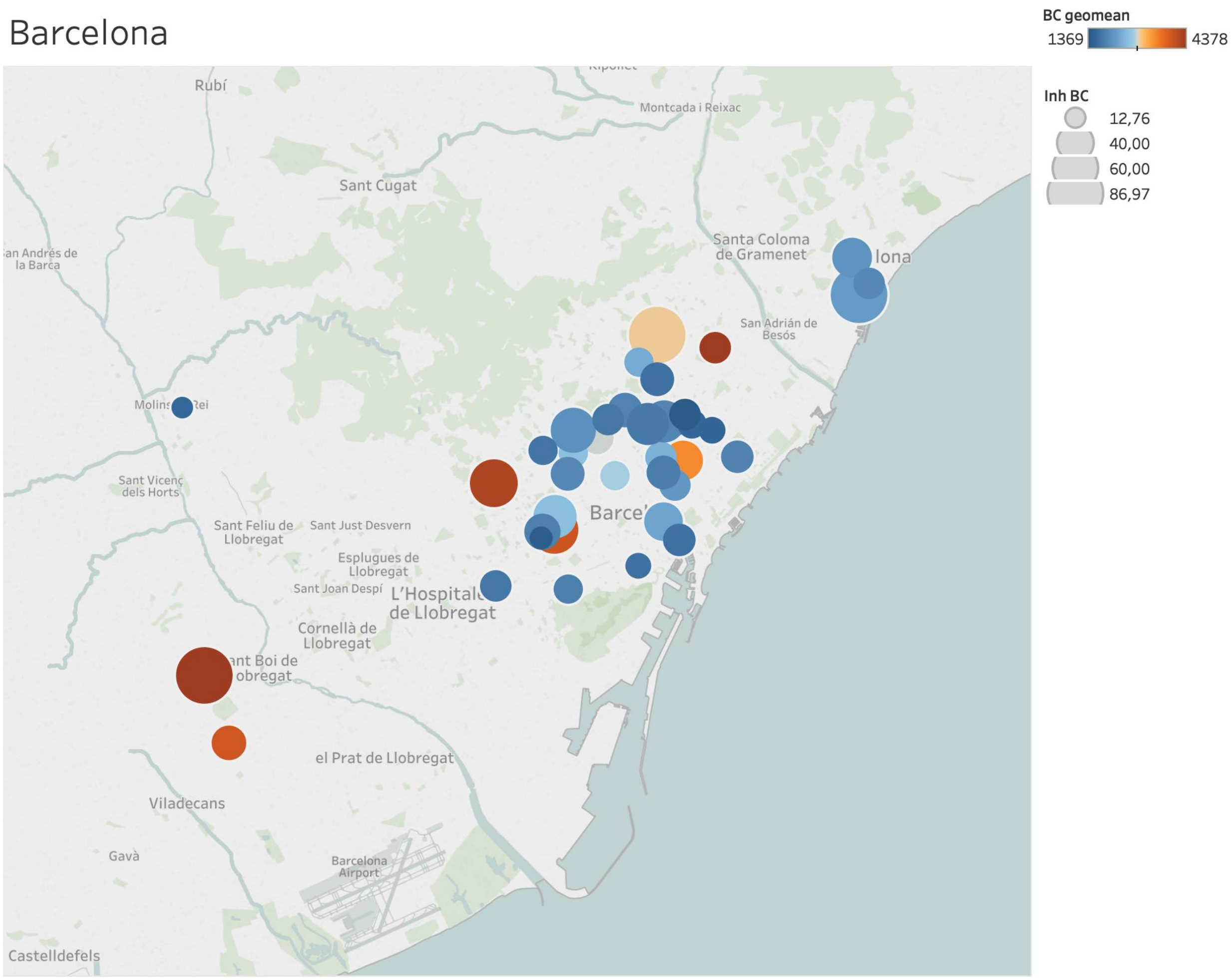


Figure 5.19 Home address distribution of participants in Barcelona. The colour represents geometric daily averages of BC exposures (ng/m^3) and the bubble size represents BC daily inhaled doses (ng/day).

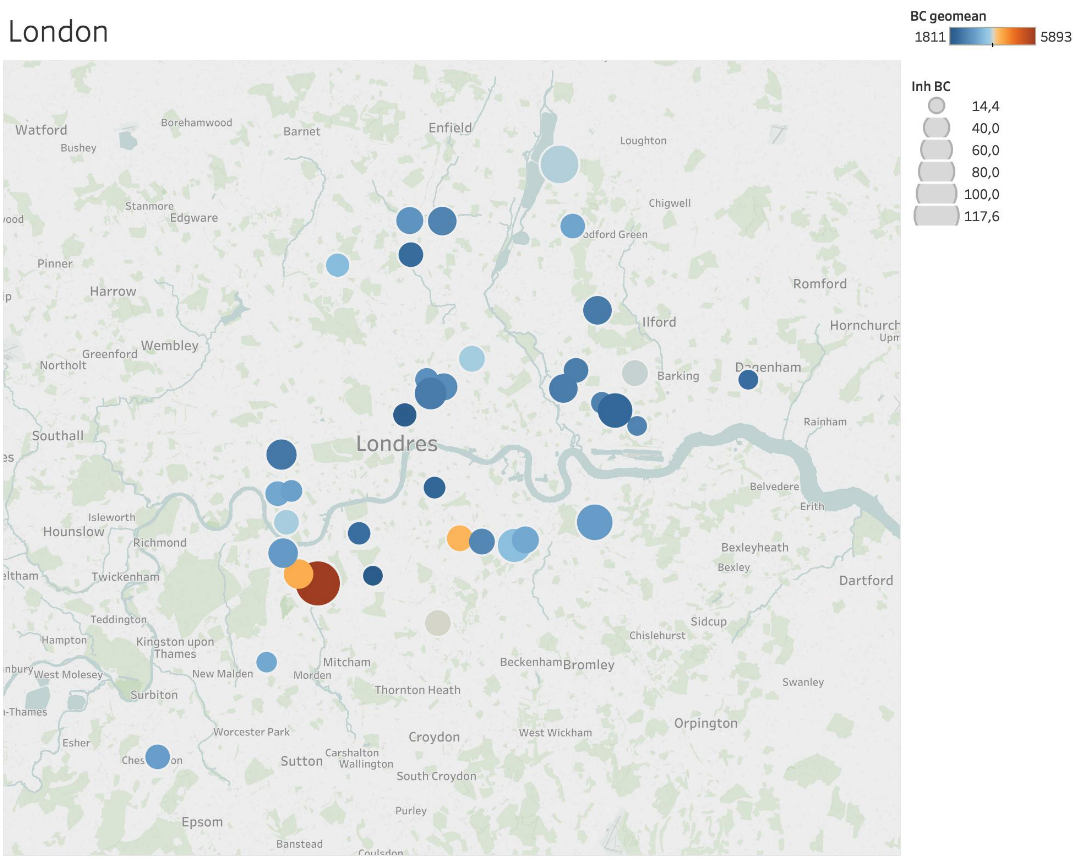


Figure 5.20 Home address distribution of participants in London. The colour represents geometric daily averages of BC exposures (ng/m^3) and the bubble size represents BC daily inhaled doses (ng/day).

Discussion

In this chapter I've presented the results of measured BC exposures in three European cities part of the PASTA project (Antwerp, Barcelona and London). A total of 122 participants monitored their BC exposures with a micro aethalometer, their geolocation with a GPS, and their physical activity levels with a SenseWear armband for three weeks each, distributed in three different seasons between February 2015 and March 2016. Results show average BC exposures between 1,720ng/m^3 in Barcelona during spring and autumn, and 3,100ng/m^3 in London during winter. Using an algorithm that combines GPS and SenseWear data I defined trips and assigned one of four modes to each trip stage: pedestrian, cycling, vehicle (including public transport) and other. With information on home and work/study location from the PASTA baseline questionnaire I also calculated percentage contributions to daily geometric mean BC exposures and to daily BC inhaled doses from each microenvironment and compared them with the time spent at each. Transport microenvironments were shown to contribute with 18.1% of the daily inhaled dose despite people spending 7.5% on average of their day in these. Cyclists were shown to have average inhaled doses twice as high as those from pedestrians and more than five times more than those from vehicle users. Finally, I compared NO_2 concentrations from a LUR model with geometric means of daily BC exposures and found no correlation between them.

The results divided by seasons show that there is no consistency between cities in terms of the season of highest exposure. While in London and Antwerp winter exposures where higher (3,100ng/m^3 and 2,040ng/m^3 respectively), in Barcelona the highest BC exposures were found in the summer (2,310ng/m^3), which were also the highest summer exposures measured in the three cities. The results of these average exposure values are a combination of seasonal meteorological changes, and seasonal activity changes that are associated typically with more time outdoors during warmer weathers and probably also changes in locations where these activities take place. Further research could be focused on how such changes of activity occur throughout the year and their implications for personal exposure levels. The results from the temporal trends from Figure 5.9 are additional evidence of the importance of activity patterns on people's exposures as shown by the early peak in exposure compared to the background station. The figure shows background levels only for London so still some caution is in place when interpreting the results.

Significant differences were found for geometric mean BC exposures and inhaled doses for when participants were in different microenvironments and transport modes. Cyclists were exposed to the second highest mean BC exposures after vehicle users and the highest inhaled doses per minute of travel. Although when aggregated to daily averages of BC exposures significant differences were only found between "unknown" mode and the others, the differences were statistically significant for total daily inhaled doses between cyclists and all other modes.

Percentage contributions of the different microenvironments show further support to the relevance of transport. I found an increase from 7.5% of time spent in transport to 10.5% contribution to daily BC exposures and a further increase to 18.1% to BC inhaled doses, values that are in line with those shown by Dons and her collaborators (Dons *et al.*, 2012) for BC measurements from 62 individuals in Flanders for seven days. As mentioned in chapter 4, they found that the 6% of the time spent in transport accounted for 21% of personal exposure and 30% of the inhaled dose. In a modelling exercise done in Barcelona by de Nazelle and her collaborators (de Nazelle, Seto, *et al.*, 2013), again 6% of the time in transport accounted for 24% of daily NO_2 inhaled doses. The results I showed in chapter 4 following the same methodology as de Nazelle showed higher proportion of participant's time in transport (9%) and higher contributions to inhaled doses (45%) for both NO_2 and $PM_{2.5}$.

It is worth noting that in all three studies mentioned, and in the results shown in this chapter, the contribution of transport to personal exposures is between 1.4 (here) and 3.5 (Dons *et al.*, 2012) times the estimated average time spent in transport microenvironments. This proportion could be underestimated in the results shown here due to misclassification of underground metro trips. In another study done for ultrafine particles (Bekö *et al.*, 2015), the authors suggest exposure contributions roughly equal to the fraction of time spent in each microenvironment. However, although this seems to hold true for their active transport results, passive transport shows a proportion of 1.6, again in line with the results discussed here. Bekö and collaborators did not calculate inhaled doses. The results shown here highlight the importance of including these when understanding microenvironmental contributions to air pollution.

This chapter shows no correlation between NO_2 concentrations for a LUR model with BC daily exposures (spearman = 0.03), despite these two pollutants being highly correlated based on the results from the urban background stations (spearman = 0.8). It is worth noting that correlations between home locations and modelled personal exposures shown in Chapter 4 were strong with Pearson correlation coefficients of 0.84 and 0.76 for $PM_{2.5}$ and NO_2 respectively. However, maps of home locations and average BC exposure measures shown here don't suggest specific patterns between these either. One option for the big drop in correlation between personal exposure and home locations is associated to influence of other sources not captured in the model proposed in Chapter 4, such as indoor sources at home or at work, or that leisure time activities should be explored in more detail. The data collected for this study was collected throughout a year with participants monitoring personal exposures, geolocation, and physical activity in three different seasons, while the data from Chapter 4 was collected during a shorter time window (three months) and for less than a week by each participant. Therefore, seasonality in activity patterns could also add to lower correlation in these results, with people changing their routines in summer and winter compared to spring when the data from Chapter 4 was collected. Furthermore, maps of geometric means of measured BC exposures and inhaled doses plotted at the participants' home locations show how in Barcelona for example, some people living in the suburbs have the highest BC exposures and inhaled doses of the entire sample. These results therefore question the validity of using home-location for air pollution epidemiological studies.

The results from both this chapter and Chapter 4 highlight how cyclists are disproportionally affected by air pollution. In both cases of modelled (NO_2 and $PM_{2.5}$) and measured (BC) inhaled doses, active travellers and specially cyclists, had a higher air pollution intake rate due to a combination of higher physical activity and personal exposure concentrations, despite these being "clean" modes with zero emissions. In other words, those commuters contributing the least to the problem, are at higher risk. As cities promote active travel as a way of healthier living, safer environments need to be offered to these citizens deciding to do the mode shift. This is not to say that cycling in a polluted city is unhealthy. According to Tainio and collaborators (Tainio *et al.*, 2015), health benefits due to physical activity outweigh the risks due to increased exposures in a wide variety of simulated air pollution levels that can be found around the world. Furthermore, Mueller and her collaborators (Mueller *et al.*, 2015)did a systematic review of publications in the developed world on the health impact assessments of active transport and also found that the benefits of walking and cycling clearly outweigh the detrimental effects of air pollution and traffic incidents. However, questions in terms of what happens to children and vulnerable populations remain and should be addressed soon. In an effort to promote healthy and sustainable cities, both air pollution and sedentarism need to be resolved instead of making citizen choose the lesser of two evils. Until now, cycling infrastructure design has been dominated by traditional transport indicators of connectivity, efficiency and road safety. The results presented in these chapters are a strong argument in favour of including air pollution exposures as an additional criterion.

One of the limitations of this study is associated to the mode detection algorithm. Despite efforts to improve on it, sensitivity (trips correctly identified as a fraction of all trips in the activity diary) and positive predictive value (trips correctly identified as a fraction of all times marked as trips) for all the options tried was never above 0.45; in other words, less than half of the minutes assigned as trips in the algorithm were also assigned as trips in the activity diaries. The activity diaries are not a gold standard either and due to time approximations doing the analysis on a minute-by-minute bases is already a high bar, so the results are still useful. In another study Donaire and collaborators (Donaire-Gonzalez *et al.*, 2016) compared data from activity diaries and a map-matching algorithm based on GPS data also, and found sensitivities and positive predictive values of 0.61 and 0.55 for trip detection, which is not far from the results presented here.

In addition to possible misclassification from the algorithm, the division in only four modes is very limited. The mode “vehicle” includes both public and private transport which could have very different distributions. Using GPS-based approaches is limited when trying to identify underground trips as there is no GPS reception there. The activity diaries and questionnaires defined by PASTA ask for use of public transport without distinguishing between, for example, bus or underground. This made it even more difficult to identify underground trips, as there was no information on when participants used it.

As mentioned before, the correlation results for home-location and personal exposure concentrations presented in this and the previous chapter should be understood as an upper and lower bound of possible correlation coefficients, and are thus merely indicative of then uncertainties associated to using home location as the preferred location for chronic exposures to air pollution in epidemiological studies. As daily activities and a dynamic location of citizens are included in future studies, biases will be reduced, and results will gain statistical power, which in turn will turn in results easier to communicate to key stakeholders such as the general public and decision makers, and probably many apparent inconsistencies will be resolved.

Detailed data as the one presented here is in practical terms impossible at large scales, is there a way to use a small sample of personal monitors and model personal exposures and inhaled doses based on variables that are easier to collect for larger populations? In chapter 6, I will use the results from the data collection campaign presented here to explore generalized linear regression models as a tool to estimate people's exposures and inhaled doses of BC. I will use four different levels of explanatory variables with varying degrees of detail and show the difference in explanatory power of each based on the coefficients of determination (R^2) of the results. The applicability of these approaches in future studies will be discussed before giving way to a general discussion and conclusions from this entire thesis.

Generalized linear regression models for black carbon exposure and inhaled dose

This chapter in a nutshell

- Linear regression models are explored as a way to estimate BC daily exposures and inhaled doses (response variables) using four levels of explanatory variables with varying degrees of detail.
- The results show a wide range of explanatory power, from some models based on activity diaries with R^2 of less than 0.05 while others based only on background data explain up to 30% of the variability in the data.
- The best results were obtained when all different explanatory variable levels were used in one integrated model which explained 40% and 35% (with $\alpha = 0.05$) of the variability in average daily BC exposures and average daily inhaled doses respectively.

Introduction

I've made the case until now of how despite being intuitively obvious that health impacts of air pollution depend on actual personal exposures and inhaled doses, these variables have been systematically ignored in air pollution modelling. Chapters 3 and 4 show ways on how to include people's daily activities using cell-phone data, and other authors have proposed similar methodologies based on GPS and smartphones (Su *et al.*, 2015), activity diaries (Physick *et al.*, 2011), or with activity-based transport modelling (Beckx *et al.*, 2009; Dons, Kochan, *et al.*, 2014). Another reason for not incorporating personal exposures is the lack of reliable and affordable data; measuring personal exposures may be expensive so collecting data in big samples with reliable devices seems impossible in practical terms. However, with ratios derived from the literature or simplified models that account for air pollution concentrations in different microenvironments, some authors have published modelling solutions to this problem. Using such ratios in combination with GPS and activity diaries, I modelled in chapter 4 personal exposures of more than 170 participants in Barcelona, following a similar methodology to the one shown in (de Nazelle, Seto, *et al.*, 2013). I showed with Monte Carlo simulations that although there is big uncertainty associated to the ratios used, that uncertainty translated in only +/- 10% of the daily estimated exposures. Other authors such as (Physick *et al.*, 2011; Beevers *et al.*, 2013; Dons, Kochan, *et al.*, 2014; Sun, Moshfeghi and Liu, 2017) have also modelled personal exposure using similar ratios or simple models to correct for indoor activities, local sources, and travelling modes, but these are still not common practice in epidemiological studies and their application has been limited to samples of tens or hundreds in most cases with few exceptions (de Nazelle, Rodríguez and Crawford-Brown, 2009; Smith *et al.*, 2016).

Inhaled doses are even more neglected in epidemiological studies and examples of literature that incorporates this variable are limited. In a study with measurements between 2007 and 2008, Zuurbier and collaborators included an estimate of inhaled doses in 34 adults using heart-rate monitors (Zuurbier *et al.*, 2011). In 2009 de Nazelle and collaborators also used inhaled doses to show the importance of incorporating this variable in policy and scenario modelling (de Nazelle, Rodríguez and Crawford-Brown, 2009) although they mention high uncertainty in their estimates. As mentioned before in chapter 4 and 5, when inhaled doses are incorporated, relative contribution of transport microenvironments is much higher than when using personal exposure estimates (chapter 4) and measurements (chapter 5).

Land Use Regression models have shown to be a good tool to associate ambient air concentrations to land uses via simple statistical regressions; there is no reason to think that this idea could not be extrapolated to air pollution exposures and inhaled doses. Despite how difficult it is to model people's behaviours and activities, these can show regular patterns frequently visiting the same places, in the same modes (González, Hidalgo and Barabási, 2008) that could lead to regular exposure levels as well. If we managed to estimate personal exposures and inhaled doses of large populations using regression models based on limited measured data, perhaps we could incorporate inhaled doses in epidemiological studies within a reasonable cost. This chapter introduces some of these ideas and explores linear regressions for different levels of detail of collected data. More than offering a "one-size-fits-all" regression that can explain urban inhaled doses of black carbon, I present a series of regressions using different sources of information and comment on their results and implications for future research.

In this chapter I use the data from the PASTA questionnaires, activity diaries, and geolocation data presented in the previous chapter to construct linear regression models that associate these variables to the measured daily average black carbon (BC) exposures and estimated daily inhaled doses. The aim of this part of the study is to quantify how much of the variability in the individual results of BC exposures and inhaled doses can be explained through the use of a combination of variables that may be easier to collect in larger samples. I construct linear regression models using four different levels of explanatory variables of varying degree of detail: general information on home location and background levels from monitoring stations, answers from the PASTA online survey, information from activity diaries, and the geolocation and mode detection data explained in chapter 5. The results from the linear regressions are evaluated based on their coefficient of determination (R^2) when including variables that are significant to the 90% and 95% significance levels (i.e. α = 0.1 and α = 0.05). Since the data was collected in three different cities and this was considered to be main source of variation in the data, the regression models were explored twice: once for the entire data set including the city as a covariate, and again using only the data from London. After presenting the results of the various regression models, I discuss on the implications of these for further research and their applicability in epidemiological studies and urban environmental policy.

Methods

Data

Monitoring data

The monitored data came from the PASTA health add-on explained in detail in chapter 5 and included BC exposure data collected with a micro aethalometer, geolocation data collected with a GPS and physical activity data from a SenseWear armband. The data was collected by 122 participants for three weeks in three different seasons each. The data cleaning and processing to calculate trip start and end points as well as detecting mode type is also explained in chapter 5.

PASTA survey

In chapter 5 I briefly mentioned the PASTA survey, a longitudinal survey and the core module of the PASTA project. It is described elsewhere (Gerike *et al.*, 2016; Avila-palencia *et al.*, 2018; Dons *et al.*, 2018) and the reader is referred to those references for more detailed explanations. The survey was rolled out in seven European cities, with more than 1,000 participants each, that were recruited from November 2014 to December of 2016. For this study however, only the answers from the 120 health add-on participants were selected. The PASTA survey started with a baseline questionnaire that took around 30 minutes to fill out and asked participants about their mobility habits, home and work/study location, typical physical activity levels for leisure, work or travel, general health, attitudes towards air pollution, noise and active travel, and general household composition. Participants would then receive a new questionnaire thirteen days after the last questionnaire was submitted in a sequence of two short follow-ups and one long. The baseline questionnaire and the long follow-up had activity diaries that asked about trips and stages done the day before filling the questionnaire out. Participants could stop filling questionnaires any time and they would stop receiving new questionnaires, although two reminders were sent if a participant had a new questionnaire unfinished.

Additionally, all health add-on participants (those monitoring their BC exposures and other variables explained in detail in chapter 5), filled out additional questionnaires during their monitoring weeks related to home characteristics such as window glazing, if they lived in a house or a flat (and in what floor if applicable), if the room where they kept the micro aethalometer while at home faced a main street or not; and habits such as use of gas cooking stoves or candles while with the micro aethalometer. Health add-on participants also had to fill out two additional activity diaries during their monitoring weeks, one half-way through and one at the end of each week.

Urban background concentrations

Temporal variability of air pollution was approximated from daily averages of three urban background stations, one in each city using the same data explained in Chapter 5. It is worth noting again that background data for Antwerp and Barcelona were only available as daily averages so any analysis that required hourly background data was only possible using the London data.

Ambient air pollution model

As a measure of spatial distribution of air pollution I used the LUR model mentioned in chapter 3 and described elsewhere (Vienneau *et al.*, 2013). This model has NO_2 concentrations for the entire region where this study took place and although it does not include BC concentrations, as mentioned in Chapter 5, NO_2 and BC concentrations from the North Kensington station were found to correlate well (Spearman = 0.8) so this was considered a good proxy. The use of NO_2 as a proxy for BC has been discussed in the previous Chapter, but again it is worth mentioning that this is only a broad approximation of the spatial gradients of ambient air pollution.

Explanatory variables

The general methodology followed to construct the linear regressions is shown in Figure 6.1. The figure shows four different levels of explanatory variables for BC exposure and inhaled BC doses (i.e. the response variables). Next, I will present in increasing level of detail each level of explanatory variables used.

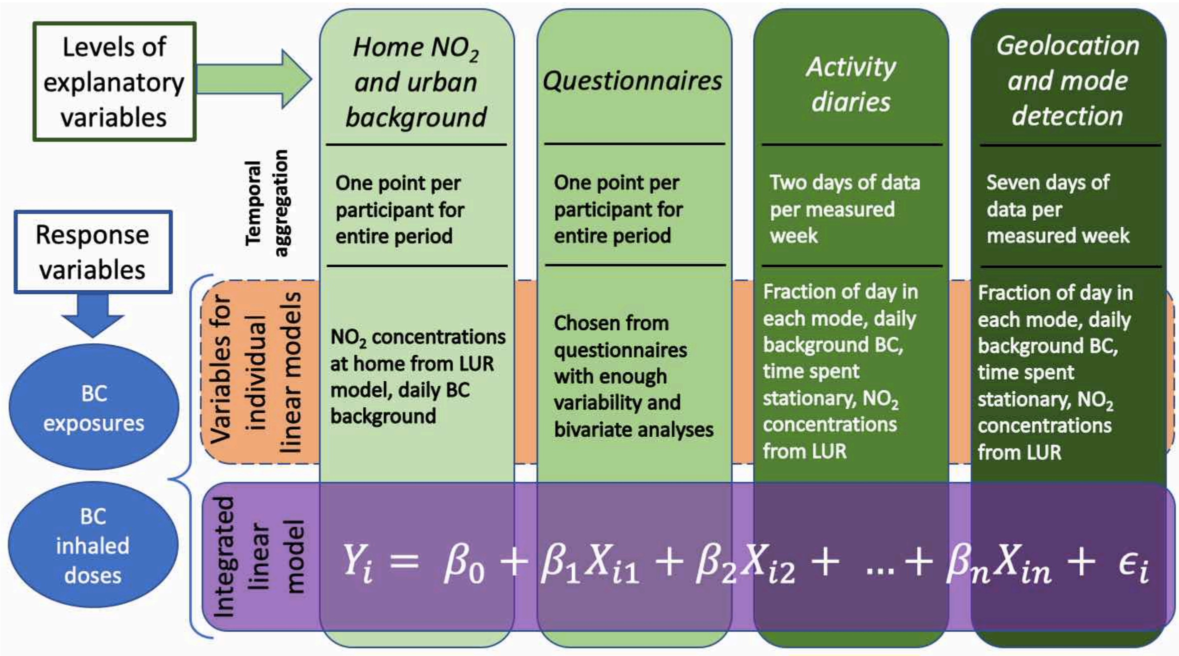


Figure 6.1. Schematic description of the general method for linear regressions

Home and urban background BC concentrations

Ambient air concentrations of NO_2 at participant's home location was shown to have no correlation on its own with average BC exposures measured in this dataset as also discussed before. However, this does not exclude the possibility of this variable being important when used in combination with other explanatory variables, so it was included in this part of the study. As explained before, participants were asked for their home location in the baseline questionnaire of the PASTA survey. The locations set by participant were then overlaid onto the LUR NO_2 model (Vienneau *et al.*, 2013) using R (Team and R Development Core Team, 2016) and R packages developed for spatial analysis (Lewin-Koh and Bivand, 2011; Hijmans and van Etten, 2014; Bivand *et al.*, 2016; Richard A. Becker *et al.*, 2016).

The corresponding NO_2 concentration for each home location was extracted and assigned to each participant. Note that for the linear regressions only the relative difference in NO_2 concentrations will be important and thus, there is no need for temporal corrections as those shown in chapter 3. Home location was considered to be constant for the entire data collection period.

A background BC average was first assigned to every participant for every day of collected data. However, in order for both variables (i.e. home location and background data) to have the same temporal bases, daily background averages per participant were aggregated to one single personal average background BC value for the entire set of days that each participant measured. Note that most participants measured on different days as we only had six or seven participants per city per week.

General questionnaires

A second level of explanatory variables used are the answers to the PASTA questionnaires, excluding the activity diaries, as those will be used in the third level explained bellow. The pasta survey was first divided by questionnaire type: Baseline questionnaire (BLQ), Follow-ups (FUP), and Health add-on specific questionnaires (HAO). Questions from each questionnaire type were classified between those that resulted in data that was continuous (e.g. hours typically spent at work), categorical (e.g. level of concern with air pollution), or indicative (from which relevant variable could be deduced, e.g. body mass index deduced from height and weight). This process is illustrated in Figure 6.2 and Table 6.1 shows the questions initially selected. However, as noted in the same table, some of them were excluded due to lack of enough data or lack of variability in the data in the participants' answers; when more than 80% of the answers were the same or missing (NA's), the question was excluded.

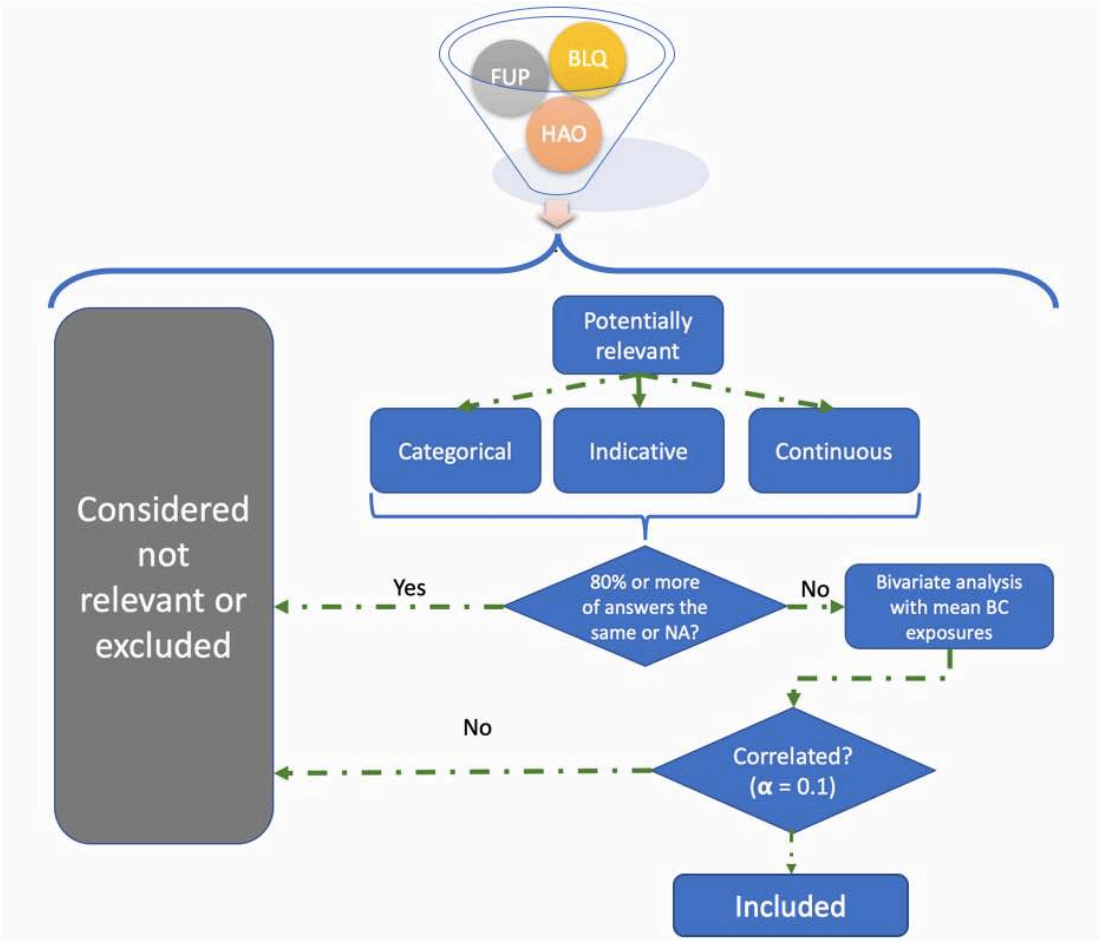


Figure 6.2. Selection of variables for the linear regressions

Table 6.1. Questions used for bivariate analysis

Question	Type	Levels	Result	Included	Comments
Sex	Categorical	Female/Male	Categorical	Yes	
Age	Continuous	18 - 65	Continuous	Yes	
Do you have access to a car or van?	Categorical	Always/sometimes/Never	Categorical	Yes	
Have you used public transport at least once in the last year?	Categorical	Yes/No	Categorical	No	More than 80% said "yes"
What type of public transport ticket do you use most?	Categorical	One-way ticket and top-up/Seasonal/Student/Other	Categorical	Yes	
Do you have access to a bicycle (private, or through a bike sharing system)?	Categorical	Yes/No	Categorical	Yes	
How often do you currently use each of the following methods of travel to get to and from places?					
-Walk	Categorical	Daily or almost daily/ 1-3 days per week/ 1-3 days per month /Less than once per month /Never	Categorical	Yes	
-Bicycle				Yes	
-Electric bicycle				No	More than 80% said "never"

-Motorcycle or moped				No	More than 80% said "never"
-Public Transport				Yes	
-Car or van				Yes	
Listed below are potential activities that you may engage in in your daily life. Which method of travel would you be most likely to use?					
-Visiting friends or family in your city	Categorical	Walk/Bicycle/Motorcycle or moped/Public Transport/Car or van/Other	Categorical	Yes	
-Shopping for groceries				Yes	
-Going to a restaurant				Yes	
-Taking a weekend excursion				Yes	
-Engaging in sports				Yes	
In a typical week, on how many days do you do a total of 30 min or more of physical activity, which is enough to raise your breathing rate?	Continuous	1 - 7	Continuous	Yes	
On average, approx. how many hours do you work per week?	Continuous	0 - 24	Continuous	Yes	
In a typical week, on how many days do you walk for at least 10 minutes	Indicative	0 - 7	Continuous	Yes	These two questions were

continuously to get to and from places?					combined into one variable: Total time in typical week
Typically, how much time do you spend walking on such a day?	Indicative	>=0			
In a typical week, on how many days do you cycle for at least 10 minutes continuously to get to and from places?	Indicative	0 - 7	Continuous	Yes	These two questions were combined into one variable: Total time in typical week
Typically, how much time do you spend cycling on such a day?	Indicative	>=0			
In a typical week, on how many days do you use an electric bike for at least 10 minutes continuously to get to and from places?	Indicative	0 - 7	Continuous	No	These two questions were combined into one variable: Total time in typical week. More than 80% had zero (0)
Typically, how much time do you spend cycling with an electric cycle on such a day?	Indicative	>=0			

How much time do you usually spend sitting or reclining on a typical day?	Continuous	>=0	Continuous	Yes	
How tall are you?	Indicative	>=0	Continuous	Yes	These two variables were combined to calculate BMI
How much do you weigh?	Indicative	>=0			
Are you worried that air pollution in the neighbourhood of either your home or work can lead to health problems?	Categorical	Not worried at all/Not worried/Neither worried not worried/Worried/Extremely worried	Categorical	Yes	
How much are you disturbed, bothered or annoyed by noise from road traffic in the neighbourhood of either your home or work?	Categorical	Not Annoyed at all/Not Annoyed/Neither Annoyed not Annoyed/Annoyed/Extremely Annoyed	Categorical	Yes	

What is your highest level of completed education?	Categorical	No degree/Primary education/Secondary education/University education	Categorical	No	More than 80% answered "University education"
Please tick one box to describe your ethnic background.	Categorical	Asian/Black/White/Mixed/Arab/Latin-American/Other	Categorical	No	More than 80% answered "White"
Which type of windows do you have?					
-Triple glazed windows in all rooms	Indicative	No/Yes	Categorical	No	These questions were combined into one associated to "window type". More than 80% had double glazing
-Triple glazed windows in a few rooms	Indicative	No/Yes			
-Double glazed windows in all rooms	Indicative	No/Yes			
-Double glazed windows in a few rooms	Indicative	No/Yes			
-Single glazed windows in all rooms	Indicative	No/Yes			
-Single glazed windows in a few rooms	Indicative	No/Yes			
How frequently do you use your gas cooker?	Categorical	More than 5 times per week / 3-5 times per week/ Less	Categorical	Yes	

Do you have a heating stove or a fireplace?	Categorical	No/Yes	Categorical	Yes	
Do you have carpet in your living room?	Categorical	No/Yes	Categorical	No	More than 80% answered "no"
Do you live in a house or a flat?	Categorical	House/Flat/Other	Categorical	Yes	
On which floor do you live?	Categorical	Ground/First/Second/Third/Above	Categorical	Yes	
Where is the living room located?	Categorical	Facing the street/away from the street/Other	Categorical	Yes	
Does your work involve vigorous-intensity activities for at least 10 minutes continuously?	Categorical	No/Yes	Categorical	No	More than 80% said "no"
Does your work involve moderate-intensity activities for at least 10 minutes continuously?	Categorical	No/Yes	Categorical	No	More than 80% said "no"
In the last seven days on how many days did you walk for at least 10 minutes continuously to get to and from places	Indicative	0 - 7	Continuous	Yes	These two questions were combined into one variable:

Typically, how much time do you spend cycling on such a day?	Indicative	>=0			Total time in last week
In the last seven days on how many days did you cycle for at least 10 minutes continuously to get to and from places	Indicative	0 - 7	Continuous	Yes	These two questions were combined into one variable: Total time in last week
Typically, how much time do you spend cycling on such a day?	Indicative	>=0			
Do you do any vigorous-intensity sports	Categorical	Yes/No	Categorical	Yes	
In the last seven days on how many days did you do vigorous-intensity sports	Indicative	0 - 7	Continuous	Yes	These two questions were combined into one variable: Total time in last week
Typically, how much time do you spend doing vigorous-intensity sports on such a day?	Indicative	>=0			
Do you do any moderate-intensity sports	Categorical	Yes/No	Categorical	Yes	

In the last seven days on how many days did you do moderate-intensity sports	Indicative	0 - 7	Continuous	Yes	These two questions were combined into one variable: Total time in last week
Typically, how much time do you spend doing moderate-intensity sports on such a day?	Indicative	>=0			
In the last 7 days, how much time did you typically spend sitting or reclining?	Continuous	>=0	Continuous	Yes	
During the last 7 days, did you drive your car with open windows?	Categorical	Daily more than an hour/Daily less than an hour/A few times/Never	Categorical	Yes	
During the last 7 days, did you drive your car with the ventilation on?	Categorical	Daily more than an hour/Daily less than an hour/A few times/Never	Categorical	Yes	
During the last 7 days, how many times did you open the windows in the room where you will place the µ-aethalometer?	Categorical	Daily more than an hour/Daily less than an hour/A few times/Never	Categorical	Yes	

During the last 7 days, did you light candles in the room where you will place the μ-aethalometer?	Continuous	0 - 7	Continuous	Yes	

To choose only those variables with some level of statistical significance for estimating daily BC exposures averages and inhaled doses, I did bivariate analyses between the answers from the selected questions and daily BC averages using an ANOVA test (α = 0.1) for categorical variables and spearman correlation test (α = 0.1) for continuous data. Indicative variables were transformed to either categorical or continuous as shown in Table 6.1 before testing for significance. The level of significance was kept at 0.1 given that as a first exploratory exercise I wanted to avoid excluding variables that could gain significance when used with other covariates in the linear regressions (i.e. the Type II error of the linear models). However, some level of significance required of the variables to be used was necessary to reduce the probability of finding false associations (i.e. Type I error). Answers to the questions selected were assumed to be constant for the entire data collection period as there is only one answer per participant.

Activity diaries

The third level of explanatory variables are the activity diaries. Although the baseline questionnaire and the long follow-ups have activity diaries, the ones used here were only those coming from the health add-on additional questionnaires as those coincide with the periods of measurement. This means that there are two activity diaries per participant per week of measurement.

Activity diaries asked for a detailed description of trips done the day before to filling these out. They gave trip start and end times, number of stages per trip if using different modes, and trip purpose. The available modes were walking, cycling, car or van, motorcycle, and public transport. Public transport included both buses and rail systems[4]. Stage duration was asked except when public transport was selected, in which case the question asked was only the time waiting at the station. The time assigned to public transport was then the difference between the trip total duration and

[4] The questionnaire was ambiguous in terms of taxis and other personal public transport systems such as Uber etc. These trips could be in either “car” or “public transport” depending on the participant’s interpretation.

the sum of all other stages. In some cases, the sum of stage duration was higher than the total time of the trip, so these were truncated when total trip time was reached.

Activity diaries also asked for location of origin and destination but not for any details on the route followed. This meant that location throughout the day could be derived for stationary times but for travel times a straight line between origin and destination was traced and a constant speed assumed. Using the LUR model, NO_2 concentrations at each location were then extracted. Concentrations while in transport are possibly an underestimate since they are not following specific transport corridors were concentrations will be comparatively higher.

From this information the following variables were calculated and used in the linear regressions:

i) Fraction of the day spent in each transport mode,
ii) total time spent stationary
iii) daily geometric mean average of assigned NO_2 concentrations, and
iv) mean daily BC concentrations from urban background stations.

Geolocations and mode detection

The last level of explanatory variables was based on the GPS and mode definition algorithm described in chapter 5. From these data I calculated the same variables as for the activity diary data:

i) Fraction of the day spent in each transport mode,
ii) total time spent stationary
iii) daily geometric mean average of assigned NO_2 concentrations, and
iv) mean daily BC concentrations from urban background stations.

The difference between these data is the origin of the data (based on the GPS and SenseWear instead of participants' travel diaries) and that the data was available for all seven days of each monitoring week.

Response variables

The response variables were individuals' daily geometric means of measured BC exposures from the micro aethalometer, , where *j* and *k* refer to the participant and day respectively; and resulting daily total inhaled doses, , estimated using measured physical activity levels in the same way explained in chapters 4 and 5:

$$BC_{GM,j,k} = \left(\prod_{l} bc_{l,j,k}\right)^{1/l} \tag{6.1}$$

Where $bc_{l,j,k}$ indicates the BC micro aethalometer reading *l* (typically every five minutes) for participant *j* in day *k*.

$$InhBC_{j,k} = \sum_{i} bc_{i,j,k} * VR(METS_{i,j,k}, sex_j, age_j, BM_j, rnorm1, rnorm2, rnorm3) \tag{6.2}$$

Where $bc_{i,j,k}$ indicates the BC reading in minute *i* from participant *j* on day *k*, and was defined in chapter 4 by equation 4.3 and Tables 4.2 to 4.4.

Since there are three different levels of temporal aggregation (Figure 6.1), these explanatory variables were averaged and matched as shown in Table 6.2. The integrated linear model (explained in detail below) uses data from all levels so it was aggregated to one point per participant.

Table 6.2. Response and explanatory variables

Level of explanatory variables	Temporal aggregation	Response variables used	Total points per explanatory variable (n)
Home NO_2 and urban background	One point per participant for the entire period	Geometric mean of and for all measured days (k)	122 (one per participant)
Questionnaires	One point per participant for the entire period	Geometric mean of and for all measured days (k)	122 (one per participant)
Activity diaries	Ideally, two days of data per measured week per participant	and values for days corresponding to activity diaries	565[5] days of activity diaries
Geolocation and mode detection	Typically, seven days of data per measured week per participant	and values for measured days	2,537 days of valid data
Integrated linear model	One point per participant for the entire period	Geometric mean of and for all measured days (k)	122 (one per participant)

Linear regression models

Linear regression models were developed using the explanatory variables of each level and associated to the corresponding response variables. The general equation of the linear regression used is shown in equation 6.3.

$$Y_p = \beta_0 + \beta_1 X_{p1} + \beta_2 X_{p2} + \cdots + \beta_q X_{pq} + \varepsilon_p$$

[5] Not all participants remembered to fill out their activity diaries, so some days are missing.

(6.3)

Where Y_p is response variable p, X_q is one of the q variables included in each level, β_q is the coefficient for each variable q, and ε_p is the error term or residual coefficient for response variable p.

Every model started with the inclusion of all variables in each level and calculating their individual p-value to assess the level of significance. At first a 90% significance level (α = 0.1) was considered. A supervised stepwise backward elimination process was used as follows: all variables with a p-value higher than 0.1 were then ordered in increasing p-value and the one with the highest one was discarded. The process was repeated discarding one variable at a time until no variables with a p-value higher than 0.1 were left. The result was recorded with its corresponding adjusted R^2 and the p-value for the entire model. Since the variables left were not necessarily independent, the model was tested for interactions between them, but these were in all cases found to be not significant. The process was then continued to the 95% (α = 0.05) significance level and both results are presented.

Finally, data from all four levels of explanatory variables was aggregated to one point per participant for the entire period by calculating the mean of those with higher temporal resolution. The process to define one lineal regression using these was then repeated for and the R^2 and p-value recorded and compared to the ones from the individual processes.

In every model the city from which the data was collected from (Antwerp, Barcelona or London) was included as a covariate and in various cases, this variable was shown to explain an important part of the variability, as shown in the results below. In order to assess the value of each model to estimate the variability within one of the cities, the process was also repeated only for London. This meant of course having around a third of the data but illustrates the process when no intercity variability is involved.

Results

Home NO_2 and urban background

By running the simplest and most aggregated linear model, I found that by correcting for city and daily averages of BC background concentrations from one monitoring station from each city, I could explain 35% of the variability of the average BC exposure data (Table 6.3). Results from the NO_2 concentrations at home location were not found significant at the 90% or 95% level. However, the explanatory variables were found not significant enough when modelling average inhaled doses. The only exception was when the model was run using only the London data, that home location NO_2 concentrations were found significant at the 95% level (p-value = 0.045), but this explained only 7.8% of the variability in the data and the β coefficient was negative, suggesting a slightly inverse relation between the geometric mean of measured BC exposures and NO_2 concentrations at home location.

Table 6.3. Results for the first level of explanatory variables - background concentrations and home location (α = 0.05)

Response variable	Explanatory variables used (p-values)	Coefficient sign for continuous data (+ or -)	Adjusted R^2	Model p-value
BC exposures	-Daily Background (<<0.001) -City (<<0.001)	Daily background (+)	0.347	<<0.001
Mean daily BC Inhaled doses	None	NA	NA	NA
BC exposures in London	None	NA	NA	NA
Mean daily BC Inhaled doses in London	NO_2 concentration at home (0.0448)	NO_2 concentration at home (-)	0.078	0.0448

Questionnaires

The questions selected from the questionnaires were found to have no explanatory value of the variability of the BC average of daily exposures. All of the covariates where eliminated until only "city" was left, which on its own did explain 24% of the variability in the data. When run only for London, no significant correlation was found between the selected questions and the mean of daily BC exposures.

The results were nevertheless very different for the average BC inhaled doses. At the 90% significance level, it was found that sex (with women inhaling lower levels of pollution than men) and three question answers were included in the model:

- Total typical time walking per week declared by participants. Continuous variable resulting from the multiplication of two baseline questionnaire answers:
    - "In a typical week, on how many days do you walk for at least 10 minutes continuously to get to and from places?" (days)
    - "Typically, how much time do you spend walking on such a day?" (min)
- Total typical time cycling per week declared by participants. Continuous variable resulting from the multiplication of two baseline questionnaire answers:
    - "In a typical week, on how many days do you cycle for at least 10 minutes continuously to get to and from places?" (days)
    - "Typically, how much time do you spend cycling on such a day?" (min)
- Level of worry about air pollution. Categorical variable answering to the question: "*Are you worried that air pollution in the neighbourhood of either your home or work can lead to health problems?*" with level of worry ranging from 1 (not worried at all) to five (Extremely worried).

The resulting model had an R^2 of 0.308 and a p-value of 0.0123. When done to the 95% significance level, time walking was excluded, and the model still explained 29.4% of the variability and had a p-value <<0.001. People answering that they were "worried" about air pollution were found to be statistically significantly more exposed to BC than those answering "not worried at all". Only those "worried" (level 4) and not worried (level 2) where found to be statistically different from "not worried at" while differences with "neither worried or not worried" (level 3) and "extremely worried" (level 5) were not significant. When the model was ran for London, total typical time cycling, air pollution worry, and sex were again selected at the 90% significance level, but only sex was significant at the 95% level. Figure 6.3 shows the coefficient plot for the linear regressions for mean BC daily inhaled doses for London and the entire sample at the 95% significance level when time cycling, sex, and all levels of worry about air pollution are included. Table 6.4 summarizes these findings.

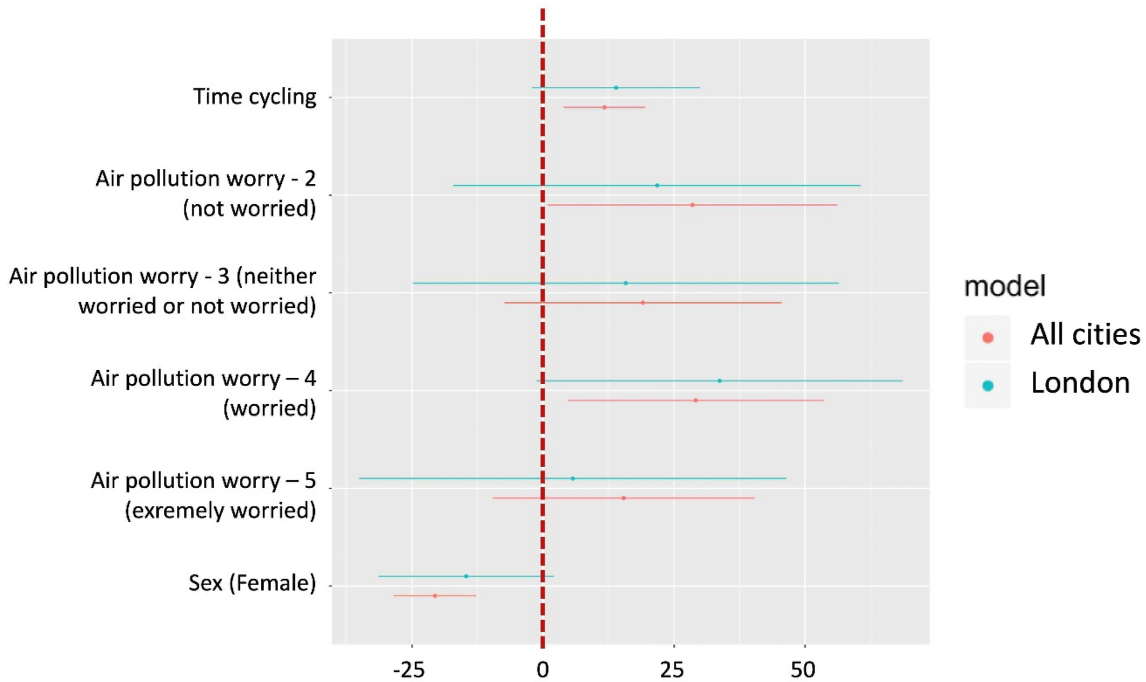


Figure 6.3. Coefficient plot (95% significance level) for average BC daily inhaled doses linear regression based on questionnaires. Categorical variables are compared to: (i) worry about air pollution: level 1 – not worried at all, and (ii) Sex: male.

Table 6.4. Results for the second level of explanatory variables – Questionnaires

Response variable	Explanatory variables used (p-values)	Coefficient sign for continuous data (+ or -)	Adjusted R^2	Model p-value
BC exposures (α = 0.1)	None	NA	NA	NA
Mean daily BC Inhaled doses (α = 0.05)	-Time cycling in a typical week (0.004) -Level of worry about air pollution (0.019 – 0.22) -Sex (<<0.001)	-Time cycling in a typical week (+) -Level of worry about air pollution (+) -Sex (-)	0.294	<<0.001
BC exposures in London (α = 0.1)	None	NA	NA	NA
Mean daily BC Inhaled doses in London	-Time cycling in a typical week (0.0863) -Level of worry about air pollution (0.058 – 0.779) -Sex (0.084)	-Time cycling in a typical week (+) -Level of worry about air pollution (+) -Sex (-)	0.281	0.008

Activity diaries

When using the activity diaries, the results show that the variables selected can explain 20% and 19% of the variability of the daily BC exposure geometric means for the day of the activity diary at the 90% and 95% significance levels respectively. The variables selected when using a 90% significance level were:

- fraction of the time walking based on trips from the activity diary
- Daily BC data from the background stations
- NO_2 geometric mean concentrations from the LUR map based on approximate locations from activity diaries

When the model was done to the 95% significance level, only the variable associated to background concentrations was left.

The linear regressions for inhaled doses did result in some variables to be included in the 95% significance level but the result explained less than 5% of the variability of the data. A similar thing happened for the data in London for both the BC daily average exposures and inhaled doses. Table 6.5 shows the results for the 95% significance level for both response variables in both the entire sample and the London data.

Table 6.5. Results for the third level of explanatory variables – Activity diaries (α = 0.05)

Response variable	Explanatory variables used (p-values)	Coefficient sign for continuous data (+ or -)	Adjusted R^2	Model p-value
BC exposures	-BC data from background stations (<< 0.001)	+	0.190	<< 0.001
Mean daily BC Inhaled doses	-NO_2 geometric	-NO_2 geometric	0.026	< 0.001

	means from activity diaries locations (0.004) -Fraction of the time spent indoors (0.009)	means from activity diaries locations (-) -Fraction of the time spent indoors (-)		
BC exposures in London	BC data from background stations (0.014)	+	0.027	0.014
Mean daily BC Inhaled doses in London	NO_2 geometric means from activity diaries locations (0.004)	-	0.060	< 0.001

Note that the geometric mean of NO_2 concentrations from the LUR and approximate locations is selected as a significant variable. However, the coefficient sign is again negative suggesting an inverse relation and with a very low explanatory power. One of the reasons for this could be how location when in transport is derived: as explained in the methods, this comes from straight lines connecting activity diary origin and destinations and thus, do not coincide with specific transport corridors where concentrations are expected to be higher. However, it could also mean that the LUR model used for NO_2 is not a good proxy for the spatial distribution of BC exposures as discussed before. To my knowledge there is no other air pollution model that could offer a better correlation with BC spatial distribution in the three cities studied.

Geolocation and mode detection

The fourth and last level of explanatory variables is based on GPS data and the mode detection algorithm described in chapter 5. Despite being data with a higher degree of detail, the results are not far from those found for the previous level, the activity diaries. One reason for these last two levels to perform so poorly could be the fact that daily variability is more influenced by other sources and activities not accounted for, so estimating daily averages becomes harder than estimating averages for the whole period as done for the first two levels. None of the resulting linear regression models managed to explain more than a 10% of the variability of the data as shown in Table 6.6 for the 95% significance level with the models for only London performing slightly better based on their R^2. It is worth noting though that fraction of the day spent walking and cycling do show up in almost all models with very high p-values and positive coefficients (i.e. the more the time spent cycling, the higher the daily BC exposures and inhaled doses).

Table 6.6. Results for the fourth level of explanatory variables – Geolocation and mode detection (α = 0.05)

Response variable	Explanatory variables used (p-values)	Coefficient sign for continuous data (+ or -)	Adjusted R^2	Model p-value
BC exposures	- Fraction of the day spent walking (<<0.001) -Fraction of the day spent in other modes (0.001)	- Fraction of the day spent walking (+) -Fraction of the day spent in other modes (+)	0.015	<<0.001

Mean daily BC Inhaled doses	- Fraction of the day spent walking (<0.001) -Fraction of the day spent cycling (<<0.001) - Fraction of the day spent in other modes (0.013) - NO_2 geometric means from GPS locations (<<0.001)	- Fraction of the day spent walking (+) -Fraction of the day spent cycling (+) - Fraction of the day spent in other modes (-) - NO_2 geometric means from GPS locations (-)	0.033	<<0.001
BC exposures in London	- Fraction of the day spent walking (<<0.001) - NO_2 geometric means from GPS locations (<<0.001)	- Fraction of the day spent walking (+) - NO_2 geometric means from GPS locations (+)	0.080	<<0.001

Mean daily BC Inhaled doses in London	- Fraction of the day spent walking (0.041) -Fraction of the day spent cycling (<<0.001)	- Fraction of the day spent walking (+) -Fraction of the day spent cycling (+)	0.083	<<0.001

Integrated linear model

The last model evaluated was one which includes all the variables from all four levels. Variables from activity diaries and GPS were aggregated to the entire measuring period by computing their mean. The results show that a combination of these variables can explain 40% of the variability in the average daily BC exposures and a 35% of the variability in the average daily inhaled doses in all three cities (Table 6.7). Even more, these results are consistent when only the London data is used. In all cases, two of the levels are involved and GPS data is always one of them. In the models for Inhaled doses, answers from the questionnaires are also found to be significant to the 95%. The four models have mostly different variables with only a few shared between some of them and not one being present in all the models (Figure 6.4). However, they all show a consistent trend of including variables associated to one of the modes of transport, with negative coefficients for vehicle and positive coefficients for active modes. Note that the model for BC exposures for London includes both the fraction of the day spent walking from the GPS and that from the activity diary. These variables should not be independent, but when interactions between them were included, these were not significant.

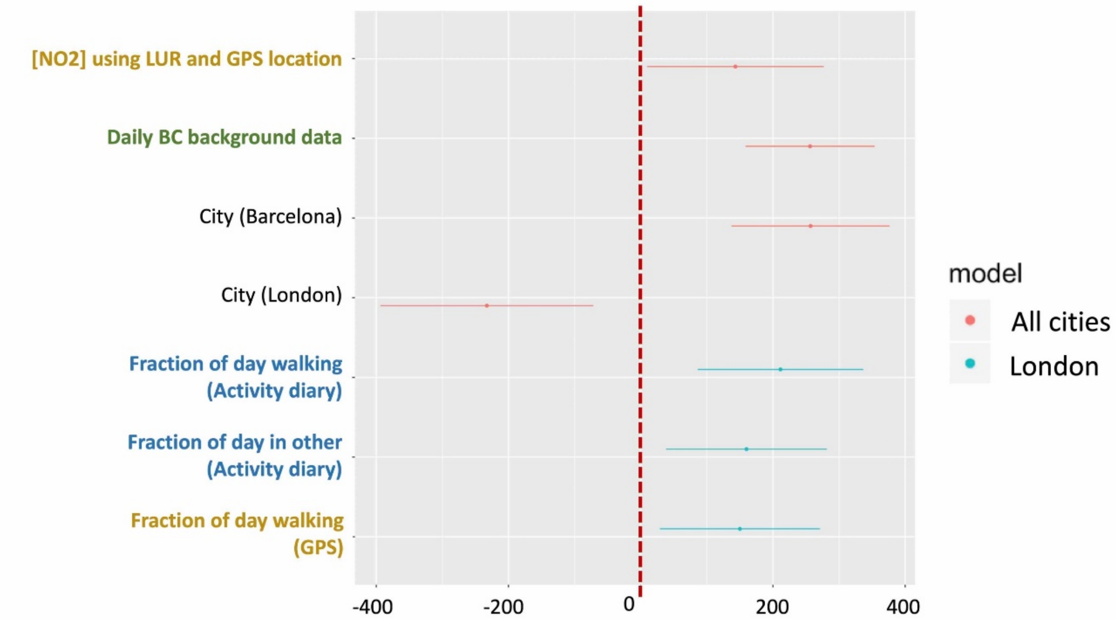
Table 6.7. Results for the integrated lineal models (α = 0.05)

Response variable	Explanatory variables used (p-values)	Coefficient sign for continuous data (+ or -)	Adjusted R^2	Model p-value
BC exposures	-NO_2 geometric means from GPS locations (0.036) -BC data from background stations (<<0.001) -City (0.005)	-NO_2 geometric means from GPS locations (+) -BC data from background stations (+)	0.408	<<0.001
Mean daily BC Inhaled doses	-Time walking in a typical week from questionnaires (0.006) -Time cycling in a typical week (0.004) -Sex (<<0.001)	-Time walking in a typical week from questionnaires (+) -Time cycling in a typical week (+)	0.358	<<0.001

	-Fraction of the day spent in vehicle based on GPS (0.002) -Fraction of the day spent stationary based on GPS (<<0.001)	-Fraction of the day spent in vehicle based on GPS (-) -Fraction of the day spent stationary based on GPS (-)		
BC exposures in London	-Fraction of the day spent walking based on activity diary (0.002) -Fraction of the day spent in other mode based on activity diary (0.012) -Fraction of the day spent walking based on GPS (0.017)	-Fraction of the day spent walking based on activity diary (+) -Fraction of the day spent in other mode based on activity diary (+) -Fraction of the day spent walking based on GPS (+)	0.385	<0.001
Mean daily BC Inhaled doses in London	-Sex (0.019) -NO_2 geometric means from GPS locations (0.004) -Fraction of the day spent in vehicle based on GPS (0.012)	-NO_2 geometric means from GPS locations (-) -Fraction of the day spent in vehicle based on GPS (-)	0.346	<0.001

	-Fraction of the day spent stationary based on GPS (0.014)	-Fraction of the day spent in stationary based on GPS (-)

(A) average BC daily exposures



(B) average BC daily inhaled doses

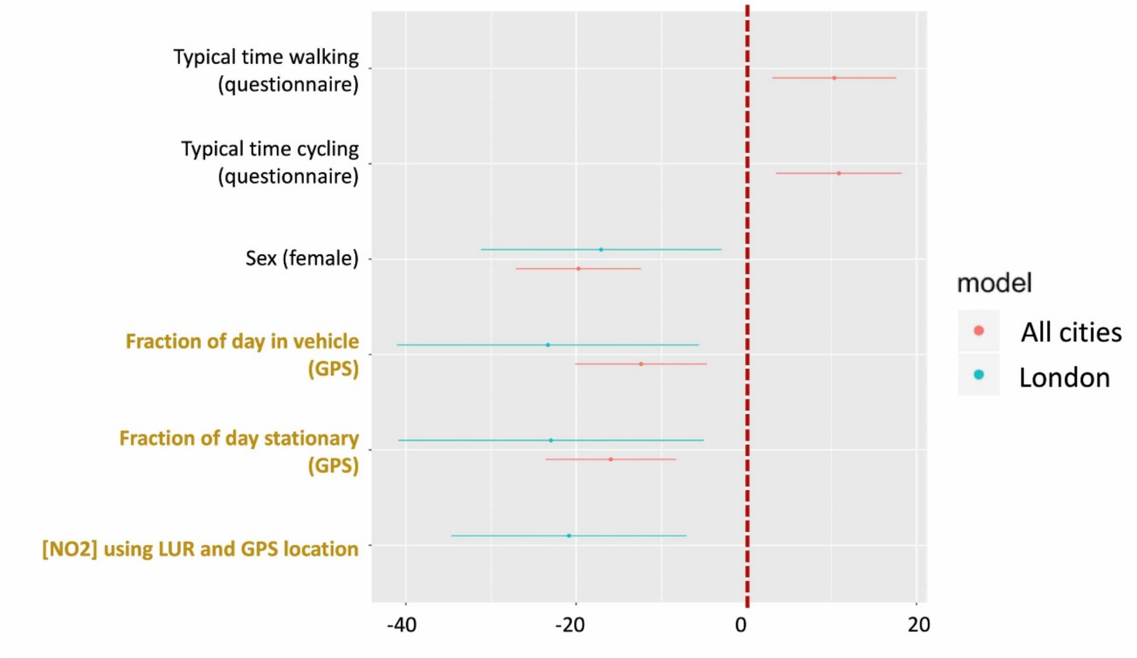


Figure 6.4. Coefficient plot (95% significance level) for average BC daily exposures (A) and average BC daily inhaled doses (B) linear regressions based on all available explanatory variables. Categorical variables are compared to: (i) City: Antwerp and (ii) Sex: male. Colours of the variable labels refer to the origin of the data: **dark gold** = GPS; **green** = urban background stations; **blue** = Activity diaries; black = questionnaires.

Discussion

In this chapter I have presented an exploration of linear regression models from four different types of explanatory variables with varying levels of detail, to find associations between these variables and two response variables, namely daily averages of measured BC exposure and estimated daily BC inhaled doses. Linear regression variables have been found by gradually and systematically eliminating variables that show no significant association with the response variables to the 90% and 95% significance levels (i.e. $\alpha = 0.1$ and $\alpha = 0.05$). The models have been presented for the data collected in three different cities (Antwerp, Barcelona and London) and then only for one of them (London) to eliminate intercity variability. The results show a wide range of explanatory power, from some models based on activity diaries with R^2 of less than 0.05 while others based only on background data can explain up to 30% of the variability in the data. However, the best results were obtained when all variables were used in one integrated model, explaining between 35% and 40% (with $\alpha = 0.05$) of the variability of the response variables in the three cities and in London.

Daily urban background BC data was shown to correlate reasonably well with measured BC exposures explaining 34% of the variability of average daily BC exposures when used with “city” as a covariate. However, this variable was not found to be significant for average daily inhaled doses of the participants in this study. A similar proportion of explained variability but for BC inhaled doses was found using data from the PASTA baseline questionnaire that reached an R^2 of 0.294. This was achieved by including three variables: typical time estimated by participants of cycling in a week, level of worry about air pollution, and their sex. The correlation with sex comes from the way inhaled doses are calculated that have lower respiration rates for women than for men so this was expected. Maybe a more interesting result is the level of association with level of worry about air pollution, that suggests the possible inclusion of variables based on participants attitude and perceptions to environmental problems to explain variability in inhaled doses. It is important to mention that this exercise in no way intends to find any causal link between the explanatory and response variables. In this particular case of worry about air pollution level for example, it could be that people are worried precisely because of their perceived high inhalation rates, but there is no evidence to establish this.

The best results were obtained when a combination of variables was used. Variability of averages of daily BC exposures was better explained ($R^2 = 0.4$) with a combination of urban background concentrations, geolocation and fraction of the day in different microenvironments, as determined by a GPS. BC inhaled doses used questionnaire data of sex and active travel in combination with GPS data to explain as much as 38% of the variability in the data. These results highlight the importance of using geolocation and mode assignment methods when modelling personal exposures and inhaled doses. Furthermore, it highlights the relevance of transport and urban mobility in air pollution health evaluations.

As mentioned here and in earlier chapters, personal exposures and inhaled doses of air pollution have been traditionally neglected in epidemiological and health impact assessment studies. The argument for this is commonly elevated costs associated to exposure and inhaled doses assessments. However, the results presented here suggest that some uncertainty associated to these variables could be dealt with through the use of simple questionnaires and some complementary geolocation data. Incorporating these variables into future research and policy evaluations could help identify populations particularly at risk such as active travellers. As cities promote physical activity as a way of dealing with non-communicable diseases associated to sedentarism, it will be more relevant to evaluate air pollution with these variables in mind.

The main purpose of this last chapter was to show an alternative method to estimate personal exposures inhaled doses using statistical tools and a limited amount of information. The relevance of these results relies more on offering an alternative and promising methodology to be explored further as an option to include inhaled doses in future health impact assessments, than in the identification or exclusion of variables themselves. Just as LUR models have offered a simple, though less robust, option to estimate spatial gradients of pollutants via statistical regressions when the resources for complex dispersion and chemical models are not available, this chapter extrapolates this concept to personal exposures and inhaled doses. The epidemiological study of air pollution impacts will benefit from accounting for activity patterns and physical activity levels and these regression methodologies offer a way of doing so without needing astronomical budgets.

More specifically, from the linear regression models presented here there are three important observations that can be extracted: i) time spent in transport and mode of choice are two key variables that consistently appear as relevant to explain variability in personal exposure and inhaled doses. This is in line with the observations from Chapters 4 and 5 where the importance of transport has been highlighted; ii) background concentration correlates well with personal exposures in London (the only city for which hourly BC data were available) offering evidence of how general temporal trends from a single monitoring station do go a long way in explaining changes in citizens' exposures, claim that is also supported by the consistent appearance of the variable "city" as a covariate in the models for all three cities; iii) although there are some common variables between exposures and inhaled doses, changes in each are not necessarily explained by the same variables and regression models should be built independently for each. This emphasises the importance of including physical activity levels in future epidemiological studies to reduce bias in the results.

Some results that are not as conclusive concern the method chosen to identify the time spent and mode while in transport. In the final combined method, all three methods (general questionnaire, activity diary, and GPS) appeared as relevant in at least one model without being conclusive on which one would be a best measure. On the one hand, the ambiguity of which method is best to account for people's time spent in transport and preferred mode is associated to the poor mode detection algorithms used here despite efforts to improve them as mentioned before. As is, it would seem that these algorithms need to be dramatically improved if the investment in collecting these data is going to show its real value for money. The other side of the same coin would be considering that with very simple questions about typical mode use, it is already sufficient to explain an important part of the observed variability. As said before, time and mode of transport is a key variable for both response variables, but it remains uncertain which method is best to achieve this. Still, questionnaires could be an interesting avenue for future research as some results did point towards the relevance of other perception variables such as level of worry of air pollution (Figure 6.3) which can only be captured via perception questionnaires or personal interviews. Level of worry is the only example that came out from this exercise, but other personal perception variables could prove to be relevant as questionnaires designed specifically for the purpose of finding statistical relationships between behaviours and personal exposures and inhaled doses are deployed in future studies.

Results associated to how the spatial variability of air pollution within each city described by the NO2 LUR model affects the BC personal exposures and inhaled doses are non-conclusive. The uncertainties associated with the use of an NO_2 LUR model as a proxy for BC spatial variation start to build up and make it difficult to conclude much: i) the validation results with NO_2 measured data quoted by the authors shows the model only explained between 48% and 58% of the variability in the data, ii) the model was done for 2013 and the BC data was collected between 2015 and 2016 with the assumption that changes in space would be homogeneous and thus correction with one urban background station per city would suffice, iii) correlations between BC and NO_2 data from urban background stations shown in the previous chapter still had a 20% of unexplained variability between the two pollutants (Figure 5.16), and iv) there was nothing to account for different spatial gradients between BC and NO_2. Despite geolocation playing some role in the linear regressions proposed, its importance could be underestimated just because of an inadequate tool to evaluate it. Out of all the results of the linear models presented there is one particularly odd one worth highlighting: some of the resulting models included the NO_2 geometric means from activity diaries locations but the correlation coefficient between these and BC inhaled doses was negative, which would suggest an inverse relation between NO_2 ambient air concentrations and BC inhaled doses. This would contradict the enormous amount of the evidence that positively correlate NO_2 and BC (mentioned in Chapter 2 in detail) besides going against basic common sense. However, the method used to estimate geolocation based on activity diary data was simply a direct path between origin and destination so these results should therefore be understood only as evidence of how precise GPS data will be needed to adequately account for the relevance of spatial distribution of ambient air concentrations in determining personal exposures and inhaled doses. In future studies other ways to evaluate the spatial distribution of air pollution should be explored. Since BC personal exposure devices such as the micro aethalometer used here are being more widely used, an effort to develop BC concentrations maps seems to be justified. It is worth highlighting that this study is an exploratory exercise of statistical methods to address some of the gaps in air pollution exposure and health studies but developing a robust set of explanatory variables will take time.

An important limitation of this research is the absence of a validation dataset. The linear regressions shown here indicate some interesting statistical trends in measured BC exposures and estimated inhaled doses based on physical activity levels, but there is no evidence to show that these trends are applicable in other contexts. Furthermore, the lack of a common variable in the models makes it difficult to generalize and gives no clues on weather these trends will hold true in other samples. As commented in the previous chapter, the sample used for this study has a bias towards active travel that could explain the importance of walking- and cycling-related variables in the results.

The method presented here has been a first step towards understanding statistical relations between variables that can potentially be collected at large scales and exposures and inhaled doses of air pollution. Future research could work on designing questionnaires specifically for this purpose and could target specific vulnerable populations such as children or active commuters. This could help to identify certain trends better and offer policy makers evidence for action to protect them.

General discussion, future work and concluding remarks

General discussion

This document presents an exploration of different methodologies that model personal exposure concentrations and daily inhaled doses of air pollutants accounting for people's activity patters. In Chapter 3 I used call detail records (CDRs) from 8 million callers in seven Italian cities to calculate population-activity-based exposures to NO_2 and I compared the results with population-home-based exposures and through the use of difference maps I pointed out how this methodology highlights relevant areas in population-weighted means that traditional home-based assessments would miss. The temporal and spatial aggregation levels of the data used allowed for a large sample since no confidential or traceable information was used, but this came with the price of averaging out areas of high concentrations. In Chapter 4, an approach based on more detailed geolocation data was used to estimate daily geometric mean exposures and total inhaled doses of NO_2 and $PM_{2.5}$ in a free-living population of 180 participants in Barcelona. The level of detail of these data allowed to calculate percentage contributions to daily exposures and inhaled doses of the different microenvironments showing that transport contributed with an important fraction of daily exposures (15%) and even more significantly to inhaled doses (27%). In Chapter 5 I present results from measured BC in three European cities (Antwerp, Barcelona and London) and again calculate microenvironmental contributions to daily average exposures and inhaled doses. In that sample, participants spent 7.5% of their day in transport microenvironments and these contributed with 10.5% of daily BC exposures and 18.1% of the daily inhaled doses. Differences while in transport of exposure levels and inhaled doses reported in Chapter 5 were large enough to show statistically significant differences between daily values of different mode users. Using the data collected and processed as explained in Chapter 5, I present in the last chapter linear regression models to estimate BC daily exposures and inhaled doses using four levels of explanatory variables with varying degrees of detail. The results show a wide range of explanatory power, but the best results were obtained when all variables were used in one integrated model which explained 40% and 35% of the data variability (with α = 0.05) in average daily BC exposures and average daily inhaled doses respectively.

The different methodologies presented in this document highlight the differences between them and the traditional home-based approach. By presenting different data sources I intended to illustrate how activity-based exposure estimates can be achieved in different ways depending on the resources available and the objective of the studies. For example, using aggregated CDRs is relatively simple and useful to identify specific air pollution hotspots in a way that considers people's mobility and not just residential address. Policy makers can use methods as these to justify the prioritization of investments in certain areas or focused intervention such as low-emission zones or defining areas for pedestrians and cyclists only. However, such methodologies would only prove useful for some epidemiological studies as census tracts will typically have a higher spatial resolution than grids based on cell-phone networks.

The results of the Monte Carlo analysis in Chapter 4 have an important implication for the future of health and policy analyses of air pollution exposures: including stochastic variables to estimate exposures while in transport (i.e. mode ratios) and ventilation rates to compute inhaled doses will account for key variables consistently shown by the existing literature that influence the impacts of air pollution (i.e. personal exposure changes depending on the microenvironment and respiration rates), without adding much more than 15% to the overall uncertainty of the results. Although there are other sources of uncertainty that have not been accounted for (see Chapter 4 for details), these are also common to more traditional practices in epidemiological studies, and thus using this methodology could help improve the results of current air pollution and health studies.

One of the benefits of including mode ratios as stochastic variables that follow probability distributions based on the literature, is that it helps account for people's activities while implicitly including the variability associated to other factors such as ventilation conditions in vehicles and route choice from cyclists. The results also help identify particularly vulnerable populations (as defined in Chapter 2) that are not captured by home-based health studies, such as active commuters. Inhaled doses have not been included in many air pollution studies, but the results shown here point towards important research gaps created by this omission. In all three chapters where I calculate inhaled doses I use a similar methodology based on accelerometer data but other options are available to researchers as summarized in our publication with Dons and others (Dons *et al.*, 2017).

The results in Chapter 5 show how pedestrians and vehicle users were the most exposed to air pollution while in transport in line with other published studies (Bekö *et al.*, 2015; de Nazelle, Bode and Orjuela, 2017). When inhaled doses were computed, cyclists where shown to have the highest inhalation rates of BC per minute and these differences translated to 3.6 ng/day (11% of average daily inhaled doses) more in cyclists than in vehicle users. This difference may seem small, but this is an average over all users despite them using different routes, living and working in different locations etc, and the differences were shown to be statistically significant. Nevertheless, these results should not be interpreted as advocacy against cycling in urban settings. The analyses presented here have not considered the benefits of physical activity but other authors have (Tainio *et al.*, 2015), showing that those benefits outweigh the costs associated to air pollution. It should come as no surprise that average daily physical activity levels of cyclists were also statistically significantly higher than those from vehicle users. The results are rather a call to safer and cleaner cycling infrastructure in urban environments, providing, for example, exclusive lanes and green routes (Jarjour *et al.*, 2013; MacNaughton *et al.*, 2014).

The linear regression models presented in Chapter 6 are an application of a widely used concept in air pollution modelling (linear regressions as used in LUR models) to a widely overlooked variable (inhaled doses). The study had important limitations in terms of the selection of variables from questionnaires and activity diaries as these came from the PASTA survey designed with other purposes. Mode detection algorithms where also shown to have important misclassifications as suggested in Chapter 5. The choice of BC as the monitored pollutant had the advantage of having a well-established in the literature personal monitor (Dons *et al.*, 2011; Jarjour *et al.*, 2013; Williams and Knibbs, 2016), but also meant limitations in the spatial distribution models used. Regardless of all of these limitations, the integrated model had a model performance measured by its adjusted R^2 (0.40 for BC exposures and 0.35 BC inhaled doses) not far from those reported for LUR models widely used today (Jerrett *et al.*, 2005; de Nazelle, Aguilera, *et al.*, 2013; Vienneau *et al.*, 2013). Epidemiological studies already use LUR models in combination with home location (Atkinson *et al.*, 2018) to evaluate health implications of air pollutants despite home location having very little correlation with inhaled doses as shown here. Generating linear regression models for inhaled doses based on easy-to-collect survey and geolocation data could be a way of bypassing the complexity of all the physicochemical and human behavioural processes air pollution goes through before inhaled by citizens but help reduce biases in associated health studies.

Applicability and relevance of these results, and future work

The methodologies and results presented in this document have important implications not only in the general understanding of personal exposures and inhaled doses of air pollutants in West European cities but for policy makers, epidemiologists and air pollution researchers. In this section I will briefly present the most important findings for each of these groups, and some important questions that remain to be solved in future studies.

Policy makers

In chapter 3 I showed how to use simple methods to identify key areas in cities that should be given priority for air quality interventions. Despite many cities in Europe having some kind of air quality map that shows places of higher pollutant concentrations, if these are not weighted by people's location, these areas are not necessarily the most problematic ones. Cities can have areas of typically higher pollutant concentrations because of meteorological conditions for example, but with very low population densities and thus, represent a lower health burden than areas with lower concentrations but with thousands or hundreds of thousands of people exposed to them.

Some authors have resolved this issue using home locations from census data (Oxley *et al.*, 2013; Molnár *et al.*, 2015) but as shown in Chapter 3, the resulting areas using home location do not typically match the areas identified when accounting for people's daily activities. Using only residential address overestimates the importance of residential areas and underestimate the importance of central business districts or other areas of low residential density but where the working population in a city can spend a significant amount of their time. This is why accounting for people's approximate location can be of significant importance. In Chapter 3 I show how this can be done at a large scale and using a methodology easily replicable in various areas, as in my example with the seven Italian cities. The use of aggregate data implies no confidential data is needed and thus, using it at a governmental or local authority level should pose no big challenges.

By using these methodologies, governments could allocate air quality budgets in a more cost-effective way without having to make great investments in models as the one presented here. At a country, continental, or economic area level, a homogeneous methodology could be applied by policy makers to identify those areas where the combination of high population density and higher air pollution concentrations make them particularly problematic.

Another implication of this work of great relevance for policy makers is the identification of population at higher risk of suffering the health consequences of air pollution. Typically, air pollution priorities have been defined on either those living in areas of higher ambient air concentrations, or the most vulnerable such as children under five years of age, pregnant women, the elderly, or those with pre-existing respiratory and cardiovascular conditions. Although this approach helps prioritize certain relevant groups in the air quality discussion, it fails to account for other population at risk due to their activities and behaviours. One of these groups are cyclists and pedestrians. As shown here in Chapters 4 and 5, inhaled doses of cyclists and pedestrians are typically higher than those of other mode users due to a combination of high exposure concentrations and high respiration rates. This is particularly unfair considering they are using zero-emission transport modes that theoretically contribute to the solution of the air quality issues. The results presented here point towards a population at higher risk of suffering the health consequences of air pollution in West European urban settings and quantify the magnitude of the difference with other mode users. Just as policy makers have considered risk of traffic injuries and network connectivity in the design of cycle lanes and sidewalks, they should consider exposure to and inhalation of air pollutants in future developments also.

Finally, some of the results presented here add to other urban planning discussions such as the impacts of having people living in city peripheries or suburbs and working in city centres or central business districts (CBD) as referred to in some of the literature. The effect of urban sprawl has been widely studied and the environmental impacts are often discussed, as evidenced by studies such as the review presented by Wilson and Chakraborty (Wilson and Chakraborty, 2013) or the analysis of 45 large US metropolitan areas done by Stone (Stone, 2008) among other. A detailed discussion on urban sprawl and its environmental impacts is far beyond the scope of this document. However, it is worth noting how although increases in journey times and distances are often cited as the results of urban sprawl and people living in suburbs, the impact these may on air pollution exposure have mainly been studied in terms of an increase in air pollutant emissions or ambient air concentrations. The effect that these have on average personal exposures and daily inhaled doses of air pollutants has not been, to the best of my knowledge, incorporated in such analyses. The maps shown in Chapter 5 (Figures 5.17-5.19) offer some preliminary evidence on how this may have an important effect, with some of the people with highest average exposures and inhaled doses in the cities studied living outside of the city centres, in areas typically associated with lower ambient air pollution. If, as shown in Chapters 4 and 5 here, transport can account for close to 20% of daily inhaled doses despite accounting for less than 10% of the time in a day, an increase in daily commuting times will result as a general rule in higher inhaled doses.

Epidemiologists

This document presents different ways to deal with people's daily activities in order to estimate their personal exposure and inhaled doses of air pollutants. Finding different ways of accounting for these activities contributes to reducing biases in epidemiological studies by identifying key variables to include in future studies. Traditionally, ambient air concentrations at home have been used as the air pollution indicator but this document contributes to showing that using that single variable is insufficient. The systematic error introduced by using home address instead of people's change in location implies an information bias in epidemiological studies that could be moving risk ratios either towards or away from the null, depending on home locations and activities of those being classified as having the health outcome being studied. If, the bias is towards the null, accounting for people's activities should not dramatically change estimated values for risk ratios or odds ratios but should help in finding stronger correlations by improving the studies' precision reducing the cited confidence intervals. On the other hand, if the bias is away from the null, current risk ratio estimates for specific health outcomes would, not only have smaller confidence intervals, but the mean value would change considerably. More studies looking into the kind of biases associated to ignoring daily activities will help clarify what these biases mean for health effects estimates and it is my intuition that the results will show that including estimates of daily exposures and inhaled doses will help have risk ratio estimates for various health outcomes that are both more precise (smaller confidence intervals) and accurate (better estimate of average).

A first step towards evaluating these mentioned systematic errors should be the evaluation of bias associated to home-based approaches when compared to some of the methodologies presented here. In a 2014 study for example, Ragettli and her collaborators (Ragettli *et al.*, 2014) showed how neglecting to include commuter and work/school results in a 12% underestimation of health effects. An equivalent analysis should be developed for the samples shown here and quantify the biases induced.

Another area that needs further study from a public health perspective and that these methods will help clarify is the effects of repeated short-term peak exposures on chronic health impacts. Traditionally air pollution health impacts have been divided into acute effects associated with specific episodes when specific meteorological or uncommon peak emissions (e.g. in wild fires) raise total exposure concentrations, and chronic effects due to consistent exposure to air pollution through the years. However, there is little research on the cumulative effect of repeated exposure peaks on chronic air pollution exposures and associated health effects. Results shown in Chapter 3 suggest that even if average concentrations don't change much, accounting for people's activity patterns result in high exposure peaks that are not accounted for when using home locations. Chapters 4 and 5 highlight how daily commute trips can account for more than 20% of inhaled doses for different pollutants despite participants not spending more than 10% of their time in transport. The modelling tools that have been presented here may help epidemiologists research the impacts of these repeated peak exposures in long-term air pollution exposure.

Finally, although the studies presented here have all focused on healthy adults, the results presented here are useful for other more vulnerable groups such as children, people with cardiovascular and respiratory diseases and the elderly. The impacts of transport in children's daily exposures have been explored for example by Buonanno and collaborators (Buonanno *et al.*, 2013) and they also found that BC exposures where highest while in transport. Using the methodologies here will also help explore how different microenvironments influence vulnerable groups beyond the traditional average concentrations used in large cohort studies.

Air pollution exposure researchers

Methodologies presented here range from the relatively simple use of CDRs in Chapter 3 to account for people's activity patters in lack of sophisticated hybrid models, to the more detailed approach of monitoring personal exposures and physical activity levels with wearable devices as those of Chapter 5. The importance of each methodology and the results shown have been pointed out in their individual discussion sections. However, in a more general context, they all open a world of opportunities for more personal exposure studies. The use of wearable devices will offer new monitoring opportunities when studying vulnerable populations, in the design of cycling infrastructure, and in the impacts of outdoor activities in people's general wellbeing. Policy makers will need more evidence to support some of the observations mentioned here before and it will be up to air pollution exposure researchers to provide such evidence. It is my hope that the results presented here serve not only as information on inhaled doses and the impact of daily activities on daily averages but will motivate other researchers to apply these methods in studies of vulnerable populations and to inform policy makers in the development of sustainable futures.

Chapter 6 presents a new avenue of research that deserves some additional notes here. Air pollution science has reached levels of sophistication and conceptual understanding that have allowed for complex modelling tools such as the hybrid and dispersion models mentioned in Chapter 2. However, even as these tools get more and more sophisticated and offer new insights of air pollution exposure levels, many cities in the world will struggle to find the resources to create them. As affordable portable devices and citizen science gains importance in the air pollution arena around the world, statistical analysis and regressions with big data sets will also gain relevance particularly in places with limited resources. In Chapter 6 I present an exploratory exercise on how to use such tools to estimate personal exposures and inhaled doses of BC. Despite the uncertainty around the results themselves limits the importance of the results, they are indicative of how regression modelling can be used in cases of limited resources to explore detailed impacts of air pollution in various populations. There is still a lot to learn from this, but the results presented offer good evidence of how to incorporate inhaled doses in air pollution analysis in different contexts with limited resources.

The results presented here point towards new research gaps that should be addressed by air pollution researchers in the future. One important limitation that has been mentioned throughout this document is the lack of adequate tools that account for city-wide concentration gradients of BC. Given the increasing popularity of BC personal exposure monitors, good dispersion and LUR models that cover spatial and temporal changes of this pollutant at a regional and city level are without a doubt a necessity. Air quality standards and WHO recommended maximum concentration levels do not include BC, and thus, air quality networks are not as complete for BC as they are for other criteria pollutants such as NO_2 and $PM_{2.5}$. This has also resulted in fewer tools available to create such needed models. However, as shown here, the adequate modelling and interpretation of personal exposures is currently limited due to gaps between personal measures and mesoscale modelling tools that offer context to the exposure measurements being explored.

Finally, it is worth noting that the inclusion of personal exposures and inhaled doses in epidemiological studies has been suggested by authors for years. It has been a decade now since de Nazelle and collaborators published their research on the built environment and health (de Nazelle, Rodríguez and Crawford-Brown, 2009). Referring to the inclusion of inhaled doses the authors conclude their abstract saying, "this innovative risk assessment uncovers critical gaps in the literature that must be further researched to allow essential comprehensive analyses of planning decisions" (de Nazelle, Rodríguez and Crawford-Brown, 2009). Our understanding of inhaled doses in air pollution health studies and its implications in planning decisions have improved since with more published papers on the matter, but these are still far from being seriously considered in main stream health and policy assessments. This suggests that part of the future work associated to these studies should be done hand in hand with epidemiologists, health specialists and policy makers to try to understand how these studies can be used to improve current practices.

Uncertainties, limitations, and what I would have done differently

Throughout this document different uncertainties and limitations have been mentioned. The previous section is intended to highlight the main messages for specific groups and suggested future work. Depending on personal interests, the reader is referred to the specific chapters that should be evaluated in more detail, and where more specific discussions take place. However, let me finish with some general notes on uncertainties, limitations and lessons learnt from this process.

The two main uncertainties mentioned along this document have been the use of NO_2 concentration maps as a proxy for BC, and the mode detection algorithms. The first one has been discussed in several sections and it has been justified in the fact that there are no BC models at the city level in the cities studied. The relationships and differences between these pollutants have been discussed also in Chapter 2. The main impact of this proxy has been the somehow contradictory results between personal exposures and ambient air concentrations at home, since these where highly correlated in Chapter 4 while not correlated at all in Chapter 5. The main message however remains the same: ambient concentrations at home location is insufficient to account for the complexities and variables associated to personal exposures and studies that do consider daily activities will help to reduce bias in epidemiological studies. Other microenvironments such as those associated to active transport play a crucial role in understanding differences in inhaled doses and their health impacts.

The second major uncertainty is the mode detection algorithms used. The data used in Chapter 3 is insufficient to determine any microenvironmental condition, including if people are indoors or outdoors, and if travelling, in which mode. The mode detection from Chapter 4 was done from paper-based activity diaries which made the specific method impractical for much larger populations. The mode assigned to participants collecting the data used in Chapter 5 and 6 was based in a modified algorithm that when compared to travel diaries from the main survey adequately assigned less than 50% of the time spent in transport. Developing better mode detection algorithms will be key in future studies. Other authors have used Google history to track participants locations and mode choice (Su *et al.*, 2015), but confidentiality issues make this a questionable option for larger populations – in their paper, Su and collaborators mention having only one participant. Research tools for travel mode detection based on GPS and accelerometer data still need to be improved. The PALMS algorithm used here and in other several other studies (Tandon *et al.*, 2013; Bekö *et al.*, 2015; Dewulf *et al.*, 2016) was deactivated at the end of 2018. However, this should not question the main results associated to the importance of transport microenvironments and the particular attention needed in active travel. The uncertainty analysis from Chapter 4 for example, suggest that although uncertainties around transport mode contrasts and inhaled doses do exist, these do not add more than 15% to the overall uncertainty of the results.

Finally, these studies have all been carried out in cities of four Western European countries, namely Italy, Spain, Belgium and the United Kingdom and the results presented here do not necessarily hold true for other regions of the world. In some cities in the United States where typical private car trips can be longer and in a higher proportion, these results can be dramatically different. Differences in urban development and housing policies will also influence the differences between home-based exposures and activity-based exposures. In the developing world with higher air pollution levels, some of the differences shown here may be enhanced (e.g. inhaled dose differences between active and passive commuters) while others may be reduced due to higher background levels (e.g. differences between microenvironmental contributions to daily exposures).

The decisions on the data used here had opportunistic origins. The data for Chapter 3 was provided in a limited time window that offered an opportunity to explore it but no data collection was possible and if the seven Italian cities where to be included, one single map for the entire region would have to be used, thus the decision of using only NO_2 data. The data for Chapter 4 had already been collected as mentioned in that chapter and thus, little opportunity to discuss the inclusion of personal monitoring was left. The data used in Chapters 5 and 6 came with the invaluable support of the PASTA project, but as is normal in big consortiums, trade-offs have to be made in order to serve the overall project above personal interests. For future studies, some of the most important changes I would include are listed below:

1. Be consistent in the pollutants used. The most precise and easily accessible maps for Chapter 3 were those for NO_2, the mode ratios for Chapter 4 with a complete analysis of probability distributions are those for $PM_{2.5}$, and the most reliable personal monitor for the methods described in Chapter 5 measures BC. This made the complete study jump between pollutants and offer conclusions sometimes hard to follow. Deciding on what pollutant to study or if all three should be studied is something that will depend on future interests and main objectives. However, if possible, that consistency should be maintained.
2. Dedicate more time to work on mode detection algorithms. At the beginning of this research project, I did not give enough thought to how to deal with mode detection, based on the naïve thought that this would be solved through the use of the ExpoApp (explained in Chapter 5). However, mode detection based on the ExpoApp had not been developed fully at the time yet and the collected data proved in the end to be very limited due to important battery issues with the smartphones used. Although a great deal of effort was put into improving mode detection, the results were insufficient.
3. Develop questionnaires for the specific purposes of this study. As mentioned before, working in multidisciplinary groups comes with immense benefits that I consider outweigh the possible costs. However, in the case of the general PASTA survey, many questions had to be summarized or formulated in a way that did not serve best my personal interests. One clear example is the public transport questions in the activity diary that do not distinguish between buses, rail and underground. For many practical purposes in sustainable transport studies, these two modes serve similar purposes and grouping them makes sense. However, specifically for personal exposure analyses, they are very different and being unable to distinguish them had a big impact in the results of Chapter 4 (not part of PASTA but also grouped rail and metro in Barcelona), and Chapters 5 and

6. Beyond mode detection, questions associated to other activities, and other air pollution perceptions and attitudes also limited the analysis.

Concluding remarks

Different methodologies to model personal exposures and inhaled doses of NO_2, $PM_{2.5}$ and BC that account for people's activity patters have been explored and their results compared to traditional home-based approaches. The different data sources and methods used provide different insights relevant to policy and health analyses such as the identification of hotspots to prioritise interventions, and the quantification of exposure and inhaled doses differences between commuters in Western European cities. The results offer new evidence on the importance of including activity patterns and ventilation rates in future epidemiological studies to avoid biases that emerge when based only on home location. Linear regression models are proposed as a method that, although needs further work, can associate easy-to-collect data with personal exposures and inhaled doses.

References

Adams, H. S. *et al.* (2001) 'Fine particle (PM2.5) personal exposure levels in transport microenvironments, London, UK', *Science of the Total Environment*, 279(1–3), pp. 29–44. doi: 10.1016/S0048-9697(01)00723-9.

Adams, H. S. *et al.* (2002) 'Assessment of road users' elemental carbon personal exposure levels, London, UK', *Atmospheric Environment*, 36(34), pp. 5335–5342. doi: 10.1016/S1352-2310(02)00637-4.

Alexander, L. K. *et al.* (2015) 'Sources of Systematic Error or Bias: Information Bias', *ERIC Notebook*. UNC Gillings school of global public health, p. 5. Available at: https://sph.unc.edu/files/2015/07/nciph_ERIC14.pdf.

Area de Barcelona - Autoritat del Transport Matropolita (2018) *ENQUESTA DE MOBILITAT EN DIA FEINER 2017 (EMEF 2017) Resum executiu de resultats principals*.

Arnold, S. J. *et al.* (2004) 'Introduction to the DAPPLE Air Pollution Project', *Science of the Total Environment*, 332(1–3), pp. 139–153. doi: 10.1016/j.scitotenv.2004.04.020.

Atkinson, R. W. *et al.* (2018) 'Long-term concentrations of nitrogen dioxide and mortality', *Epidemiology*, 29(4), pp. 460–472. doi: 10.1097/EDE.0000000000000847.

Avila-palencia, I. *et al.* (2018) 'The e ff ects of transport mode use on self-perceived health , mental health , and social contact measures : A cross-sectional and longitudinal study', *Environment International*. Elsevier, 120(August), pp. 199–206. doi: 10.1016/j.envint.2018.08.002
.

Awang, M. B. *et al.* (2005) 'Air quality in Malaysia : Impacts , management issues and future challenges', (2000), pp. 183–196.

Barlow, J. F. *et al.* (2014) 'Meteorology, Air Quality, and Health in London: The ClearfLo Project', *Bulletin of the American Meteorological Society*, 96(5), pp. 779–804. doi: 10.1175/bams-d-12-00245.1.

Beckx, C. *et al.* (2009) 'A dynamic activity-based population modelling approach to evaluate exposure to air pollution: Methods and application to a Dutch urban area', *Environmental Impact Assessment Review*. Elsevier Inc., 29(3), pp. 179–185. doi: 10.1016/j.eiar.2008.10.001.

Beevers, S. D. *et al.* (2012) 'One way coupling of CMAQ and a road source dispersion model for fine scale air pollution predictions', *Atmospheric Environment*. Elsevier Ltd, 59, pp. 47–58. doi: 10.1016/j.atmosenv.2012.05.034.

Beevers, S. D. *et al.* (2013) 'Air pollution dispersion models for human exposure predictions in London', *Journal of Exposure Science and Environmental Epidemiology*. Nature Publishing Group, 23(6), pp. 647–653. doi: 10.1038/jes.2013.6.

Bekö, G. *et al.* (2015) 'Contribution of various microenvironments to the daily personal exposure to ultrafine particles: Personal monitoring coupled with GPS tracking', *Atmospheric Environment*, 110, pp. 122–129. doi: 10.1016/j.atmosenv.2015.03.053.

Bell, M. L., Zanobetti, A. and Dominici, F. (2013) 'Evidence on vulnerability and susceptibility to health risks associated with short-term exposure to particulate matter: a systematic review and meta-analysis.', *American journal of epidemiology*, 178(6), pp. 865–876. doi: 10.1093/aje/kwt090.

Bencala, K. E. and Seinfeld, J. H. (1976) 'On frequency distributions of air pollutant concentrations', *Atmospheric Environment (1967)*, 10(11), pp. 941–950. doi: 10.1016/0004-6981(76)90200-6.

Beverland, I. J. *et al.* (2012) 'A Comparison of Short-term and Long-term Air Pollution Exposure Associations with Mortality in Two Cohorts in Scotland', *Environmental Health Perspectives*, 120(9), pp. 1280–1285. doi: 10.1289/ehp.1104509.

Bigazzi, A. Y. and Figliozzi, M. A. (2014) 'Review of Urban Bicyclists ' Intake and Uptake of Traffic-Related Air Pollution Review of Urban Bicyclists ' Intake and Uptake', *Transport Reviews*. Taylor & Francis, 34(2), pp. 221–245. doi: 10.1080/01441647.2014.897772.

Bivand, R. *et al.* (2016) 'rgeos: Interface to Geometry Engine', *Cran*. Available at: https://cran.r-project.org/package=rgeos.

Bond, T. C. and Sun, K. (2005) 'Can reducing black carbon emissions counteract global warming?', *Environmental Science and Technology*, 39(16), pp. 5921–5926. doi: 10.1021/es0480421.

Brand, C. (2016) 'Beyond "Dieselgate": Implications of unaccounted and future air pollutant emissions and energy use for cars in the United Kingdom', *Energy Policy*. Elsevier, 97(X), pp. 1–12. doi: 10.1016/j.enpol.2016.06.036.

Briggs, D. J. *et al.* (2008) 'Effects of travel mode on exposures to particulate air pollution', *Environment International*, 34(1), pp. 12–22. doi: 10.1016/j.envint.2007.06.011.

Buonanno, G. *et al.* (2013) 'Children exposure assessment to ultrafine particles and black carbon: The role of transport and cooking activities', *Atmospheric Environment*. Elsevier Ltd, 79, pp. 53–58. doi: 10.1016/j.atmosenv.2013.06.041.

Chen, C., Lin, Y. and Hsu, C. (2017) 'Does air pollution drive away tourists ? A case study of the Sun Moon Lake National Scenic Area , Taiwan', *Transportation Research Part D*. Elsevier Ltd, 53(December 2011), pp. 398–402. doi: 10.1016/j.trd.2017.04.028.

Clougherty, J. E. *et al.* (2013) 'Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: The New York City Community Air Survey (NYCCAS)', *Journal of Exposure Science and Environmental Epidemiology*. Nature Publishing Group, 23(3), pp. 232–240. doi: 10.1038/jes.2012.125.

Cole-Hunter, T. *et al.* (2013) 'Utility of an alternative bicycle commute route of lower proximity to motorised traffic in decreasing exposure to ultra-fine particles, respiratory symptoms and airway inflammation -- a structured exposure experiment.', *Environmental health : a global access science source*, 12(1), p. 29. doi: 10.1186/1476-069X-12-29.

COMEAP (2009) *Long-Term Exposure to Air Pollution: Effect on Mortality*, *A report by the Committee on the Medical Effects of Air Pollutants*. A report by the Committee on the Medical Effects of Air Pollutants.

COMEAP (2015) *Interim Statement on Quantifying the Association of Long-Term Average Concentrations of Nitrogen Dioxide and Mortality*, *Committee On the Medical Effects of Air Pollutants*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/485373/COMEAP_NO2_Mortality_Interim_Statement.pdf.

COMEAP (2016) *Long-term Exposure to Air Pollution and Chronic Bronchitis*, *A report by the Committee on the Medical Effects of Air Pollutants*. A report by the Committee on the Medical Effects of Air Pollutants.

COMEAP (2017) *Quantification of the effects of long-term exposure to ambient air pollution on cardiovascular morbidity.*

COMEAP (2018a) *Associations of long-term average concentrations of nitrogen dioxide with mortality*, *A report by the Committee on the Medical Effects of Air Pollutants*.

COMEAP (2018b) *Statement on the Evidence for Health Effects in the Travelling Public Associated With Exposure To Particulate Matter in the London Underground*. Available at: https://www.healtheffects.org/publication/national-particle-component-toxicity-npact-initiative-%0Ahttps://www.gov.uk/government/publications/particulate-air-pollution-on-london-underground-health-effects.

Cyrys, J. *et al.* (2012) 'Variation of NO2 and NOx concentrations between and within 36 European study areas: Results from the ESCAPE study', *Atmospheric Environment*, 62(2), pp. 374–390. doi: 10.1016/j.atmosenv.2012.07.080.

Dekoninck, L., Botteldooren, D. and Int Panis, L. (2013) 'An instantaneous spatiotemporal model to predict a bicyclist's Black Carbon exposure based on mobile noise measurements', *Atmospheric Environment*, 79, pp. 623–631. doi: 10.1016/j.atmosenv.2013.06.054.

Delfino, R. J. *et al.* (2014) 'Asthma Morbidity and Ambient Air Pollution: Effect modification by residential traffic-related air pollution', *Epidemiology*, 25(1), pp. 48–57. doi: 10.1097/EDE.0000000000000016.

Dewulf, B. *et al.* (2016) ‘Dynamic assessment of inhaled air pollution using GPS and accelerometer data’, *Journal of Transport and Health*, 3(1), pp. 114–123. doi: 10.1016/j.jth.2015.10.004.

Di, Q. *et al.* (2017) ‘Association of short-term exposure to air pollution with mortality in older adults’, *JAMA - Journal of the American Medical Association*, 318(24), pp. 2446–2456. doi: 10.1001/jama.2017.17923.

Donaire-Gonzalez, D. *et al.* (2013) ‘Comparison of physical activity measures using mobile phone-based calfit and actigraph’, *Journal of Medical Internet Research*, 15(6), pp. 1–11. doi: 10.2196/jmir.2470.

Donaire-Gonzalez, D. *et al.* (2016) ‘Benefits of Mobile Phone Technology for Personal Environmental Monitoring’, *JMIR mHealth and uHealth*, 4(4), p. e126. doi: 10.2196/mhealth.5771.

Dons, E. *et al.* (2011) ‘Impact of time-activity patterns on personal exposure to black carbon’, *Atmospheric Environment*. Elsevier Ltd, 45(21), pp. 3594–3602. doi: 10.1016/j.atmosenv.2011.03.064.

Dons, E. *et al.* (2012) ‘Personal exposure to Black Carbon in transport microenvironments’, *Atmospheric Environment*, 55, pp. 392–398. doi: 10.1016/j.atmosenv.2012.03.020.

Dons, E. *et al.* (2013) ‘Street characteristics and traffic factors determining road users’ exposure to black carbon’, *Science of the Total Environment*. Elsevier B.V., 447, pp. 72–79. doi: 10.1016/j.scitotenv.2012.12.076.

Dons, E., Van Poppel, M., Kochan, B., *et al.* (2014) ‘Implementation and validation of a modeling framework to assess personal exposure to black carbon’, *Environment International*. Elsevier Ltd, 62, pp. 64–71. doi: 10.1016/j.envint.2013.10.003.

Dons, E., Van Poppel, M., Int Panis, L., *et al.* (2014) ‘Land use regression models as a tool for short, medium and long term exposure to traffic related air pollution’, *Science of the Total Environment*, 476–477, pp. 378–386. doi: 10.1016/j.scitotenv.2014.01.025.

Dons, E., Kochan, B., *et al.* (2014) ‘Modeling personal exposure to air pollution with AB^2C: Environmental inequality’, *Procedia Computer Science*. Elsevier Masson SAS, 32, pp. 269–276. doi: 10.1016/j.procs.2014.05.424.

Dons, E. *et al.* (2015) ‘Physical Activity through Sustainable Transport Approaches (PASTA): protocol for a multi-centre , longitudinal study’, *BMC Public Health*. BMC Public Health, pp. 1–11. doi: 10.1186/s12889-015-2453-3.

Dons, E. *et al.* (2017) ‘Wearable Sensors for Personal Monitoring and Estimation of Inhaled Traffic-Related Air Pollution: Evaluation of Methods’, *Environmental Science and Technology*, 51(3), pp. 1859–1867. doi: 10.1021/acs.est.6b05782.

Dons, E. *et al.* (2018) ‘Concern over health effects of air pollution is associated to NO 2 in seven European cities’. Air Quality, Atmosphere & Health, (2), pp. 591–599.

Ekström, M., Sjödin, Å. and Andreasson, K. (2004) ‘Evaluation of the COPERT III emission model with on-road optical remote sensing measurements’, *Atmospheric Environment*, 38(38), pp. 6631–6641. doi: 10.1016/j.atmosenv.2004.07.019.

Englert, N. (2014) ‘Nanoparticles and human health: A review of epidemiological studies’, *Biointeractions of Nanomaterials*, 149, pp. 175–208. doi: 10.1201/b17191.
European Commission (2004) *How Europeans spend their time*.

Faustini, A., Rapp, R. and Forastiere, F. (2014) ‘Nitrogen dioxide and mortality: Review and meta-analysis of long-term studies’, *European Respiratory Journal*, 44(3), pp. 744–753. doi: 10.1183/09031936.00114713.

Fuller, G. (2018) *The Invisible Killer*. London: Melville House UK.

Gariazzo, C., Pelliccioni, A. and Bolignano, A. (2016) ‘A dynamic urban air pollution population exposure assessment study using model and population density data derived by mobile phone traf fi c’, *Atmospheric Environment*. Elsevier Ltd, 131, pp. 289–300. doi: 10.1016/j.atmosenv.2016.02.011.

Gerharz, L. E. *et al.* (2013) ‘Spatio-temporal modelling of individual exposure to air pollution and its uncertainty’, *Atmospheric Environment*. Elsevier Ltd, 64(2), pp. 56–65. doi: 10.1016/j.atmosenv.2012.09.069.

Gerike, R. *et al.* (2016) ‘Physical Activity through Sustainable Transport Approaches (PASTA): a study protocol for a multicentre project’. doi: 10.1136/bmjopen-2015-009924.

Gillespie, J. *et al.* (2016) ‘Estimation of spatial patterns of urban air pollution over a 4-week period from repeated 5-min measurements’, *Atmospheric Environment*. Elsevier Ltd, 150(2), pp. 295–302. doi: 10.1016/j.atmosenv.2016.11.035.

González, M. C., Hidalgo, C. a and Barabási, A.-L. (2008) ‘Understanding individual human mobility patterns.’, *Nature*, 453(June), pp. 779–782. doi: 10.1038/nature06958.

Gulliver, J. and Briggs, D. (2011) ‘STEMS-Air: a simple GIS-based air pollution dispersion model for city-wide exposure assessment.’, *The Science of the total environment*, 409(12), pp. 2419–29. doi: 10.1016/j.scitotenv.2011.03.004.

Gurram, S., Stuart, A. L. and Pinjari, A. R. (2014) ‘Impacts of travel activity and urbanicity on exposures to ambient oxides of nitrogen and on exposure disparities’, *Air Quality, Atmosphere & Health*, 8, pp. 97–114. doi: 10.1007/s11869-014-0275-6.

Hagler, G. S. W. *et al.* (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(5), pp. 539–546. doi: 10.4209/aaqr.2011.05.0055.

Hankey, S. and Marshall, J. D. (2015) 'On-bicycle exposure to particulate air pollution: Particle number, black carbon, PM2.5, and particle size', *Atmospheric Environment*. Elsevier Ltd, 122, pp. 65–73. doi: 10.1016/j.atmosenv.2015.09.025.

Hertel, O. *et al.* (2008) 'A proper choice of route significantly reduces air pollution exposure - A study on bicycle and bus trips in urban streets', *Science of the Total Environment*, 389(1), pp. 58–70. doi: 10.1016/j.scitotenv.2007.08.058.

Hijmans, R. J. and van Etten, J. (2014) 'raster: Geographic data analysis and modeling', *R package, version 2.3-12. http://CRAN.R-project.org/package=raster*. doi: 10.1016/j.indcrop.2009.04.001.

Holman, C., Harrison, R. and Querol, X. (2015) 'Review of the efficacy of low emission zones to improve urban air quality in European cities', *Atmospheric Environment*. Elsevier Ltd, 111, pp. 161–169. doi: 10.1016/j.atmosenv.2015.04.009.

de Hoogh, K. *et al.* (2014) 'Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies', *Environment International*. Elsevier Ltd, 73, pp. 382–392. doi: 10.1016/j.envint.2014.08.011.

de Hoogh, K. *et al.* (2016) 'Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data', *Environmental Research*. Elsevier, 151(2), pp. 1–10. doi: 10.1016/j.envres.2016.07.005.

Hooper, L. G. and Kaufman, J. D. (2018) 'Ambient air pollution and clinical implications for susceptible populations', *Annals of the American Thoracic Society*, 15(April), pp. S64–S68. doi: 10.1513/AnnalsATS.201707-574MG.

Hwang, S. H., Park, J. B. and Park, W. M. (2018) 'Radon and NO2levels and related environmental factors in 100 underground subway platforms over two-year period', *Journal of Environmental Radioactivity*. Elsevier, 181(2), pp. 102–108. doi: 10.1016/j.jenvrad.2017.11.003.

Int Panis, L. *et al.* (2010) 'Exposure to particulate matter in traffic: A comparison of cyclists and car passengers', *Atmospheric Environment*, 44(19), pp. 2263–2270. doi: 10.1016/j.atmosenv.2010.04.028.

Ivy, D., Mulholland, J. A. and Russell, A. G. (2008) 'Development of ambient air quality population-weighted metrics for use in time-series health studies', *Journal of the Air and Waste Management Association*, 58(5), pp. 711–720. doi: 10.3155/1047-3289.58.5.711.

Jacobs, L. *et al.* (2010) 'Subclinical responses in healthy cyclists briefly exposed to traffic-related air pollution: An intervention study', *Environmental Health: A Global Access Science Source*, 9(1), pp. 1–8. doi: 10.1186/1476-069X-9-64.

Jansen, K. L. *et al.* (2005) 'Associations between health effects and particulate matter and black carbon in subjects with respiratory disease', *Environmental Health Perspectives*, 113(12), pp. 1741–1746. doi: 10.1289/ehp.8153.

Janssen, N. A. *et al.* (2012) 'Health effects of black carbon', *World Health …*. Available at: http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2012/health-effects-of-black-carbon.

Janssen, N. a. H. *et al.* (2011) 'Black Carbon as an Additional Indicator of the Adverse Health Effects of Airborne Particles Compared with PM10 and PM2.5', *Environmental Health Perspectives*, 119(12), pp. 1691–1699. doi: 10.1289/ehp.1003369.

Janssen, N. a H. *et al.* (2005) 'Associations between ambient, personal, and indoor exposure to fine particulate matter constituents in Dutch and Finnish panels of cardiovascular patients.', *Occupational and environmental medicine*, 62(12), pp. 868–877. doi: 10.1136/oem.2004.016618.

Jarjour, S. *et al.* (2013) 'Cyclist route choice, traffic-related air pollution, and lung function: a scripted exposure study', *Environmental Health: A Global Access Science Source*. Environmental Health, 12(1), pp. 1–12. doi: 10.1186/1476-069X-12-14.

Jerrett, M. *et al.* (2005) 'A review and evaluation of intraurban air pollution exposure models', *Journal of Exposure Analysis and Environmental Epidemiology*, 15(2), pp. 185–204. doi: 10.1038/sj.jea.7500388.
Johnson, T. (2002) 'A Guide to Selected Algorithms, Distribution, and Databases Used in Exposure Models Developed by the Office of Air Quality Planning and Standards', *Vasa*, pp. 1–370. doi: https://www.epa.gov/sites/production/files/2013-08/documents/report052202.pdf.

Kaur, S. and Nieuwenhuijsen, M. J. (2009) 'Determinants of personal exposure to PM 2.5, ultrafine particle counts, and CO in a transport microenvironment', *Environmental Science and Technology*, 43(13), pp. 4737–4743. doi: 10.1021/es803199z.

Laeremans, M. *et al.* (2017) 'Physical activity and sedentary behaviour in daily life : A comparative analysis of the Global Physical Activity Questionnaire (GPAQ) and the SenseWear armband', pp. 1–15.

Laeremans, M. *et al.* (2018) 'Short-term e ff ects of physical activity , air pollution and their interaction on the cardiovascular and respiratory system', *Environment International*. Elsevier, 117(December 2017), pp. 82–90. doi: 10.1016/j.envint.2018.04.040.

Lähde, T. *et al.* (2014) 'Mobile particle and NOx emission characterization at Helsinki Downtown: Comparison of different traffic flow areas', *Aerosol and Air Quality Research*, 14(5), pp. 1372–1382. doi: 10.4209/aaqr.2013.10.0311.

Lao, J. and Teixidó, O. (2011) 'Air quality model for Barcelona', 147(2), pp. 25–36. doi: 10.2495/AIR110031.

Latza, U., Gerdes, S. and Baur, X. (2009) 'Effects of nitrogen dioxide on human health: Systematic review of experimental and epidemiological studies conducted between 2002 and 2006', *International Journal of Hygiene and Environmental Health*, 212(3), pp. 271–287. doi: 10.1016/j.ijheh.2008.06.003.

Lewin-Koh, N. and Bivand, R. (2011) 'maptools: Tools for reading and handling spatial objects', *R package*, p. 74. Available at: http://cran.r-project.org/package=maptools.

London Air - Kings College London (2018) *Annual Pollution Maps*. Available at: https://www.londonair.org.uk/london/asp/annualmaps.asp (Accessed: 16 April 2019).

London Air and King's College London (2018) *London Air - Data Downloads*.
MacNaughton, P. *et al.* (2014) 'Impact of bicycle route type on exposure to traffic-related air pollution', *Science of the Total Environment*. Elsevier B.V., 490(2), pp. 37–43. doi: 10.1016/j.scitotenv.2014.04.111.

Mediavilla-Sahagún, A. and ApSimon, H. M. (2003) 'Urban scale integrated assessment of options to reduce PM10 in London towards attainment of air quality objectives', *Atmospheric Environment*, 37(33), pp. 4651–4665. doi: 10.1016/j.atmosenv.2003.07.004.

Michaels, R. A. and Kleinman, M. T. (2000) 'Incidence and apparent health significance of brief airborne particle excursions', *Aerosol Science and Technology*, 32(2), pp. 93–105. doi: 10.1080/027868200303803.

Millard, S. P. (2013) 'EnvStats : An R Package for Environmental Statistics'. New York: Springer, p. 291. doi: 10.1007/978-1-4614-8456-1.

Minguillón, M. C. *et al.* (2012) 'Source apportionment of indoor, outdoor and personal PM2.5 exposure of pregnant women in Barcelona, Spain', *Atmospheric Environment*, 59, pp. 426–436. doi: 10.1016/j.atmosenv.2012.04.052.

Minguillón, M. C. *et al.* (2012) 'Within-city contrasts in PM composition and sources and their relationship with nitrogen oxides', *Journal of Environmental Monitoring*, 14(10), pp. 2718–2728. doi: 10.1039/c2em30469d.

Miskell, G., Salmond, J. A. and Williams, D. E. (2018) 'Solution to the Problem of Calibration of Low-Cost Air Quality Measurement Sensors in Networks', *ACS Sensors*, 3(4), pp. 832–843. doi: 10.1021/acssensors.8b00074.

Molnár, P. *et al.* (2015) 'Residential NOx exposure in a 35-year cohort study. Changes of exposure, and comparison with back extrapolation for historical exposure assessment', *Atmospheric Environment*, 115(x), pp. 62–69. doi: 10.1016/j.atmosenv.2015.05.055.

Monn, C. (2002) 'Exposure assessment of air pollutants: a review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone', *Developments in Environmental Science*, 1(C), pp. 117–182. doi: 10.1016/S1474-

8177(02)80007-9.

Morabia, A. *et al.* (2010) 'Potential health impact of switching from car to public transportation when commuting to work', *American Journal of Public Health*, 100(12), pp. 2388–2391. doi: 10.2105/AJPH.2009.190132.

Mueller, N. *et al.* (2015) 'Health impact assessment of active transportation: A systematic review', *Preventive Medicine*. Elsevier Inc., 76, pp. 103–114. doi: 10.1016/j.ypmed.2015.04.010.

Müller, D. *et al.* (2011) 'Car indoor air pollution - Analysis of potential sources', *Journal of Occupational Medicine and Toxicology*. BioMed Central Ltd, 6(1), p. 33. doi: 10.1186/1745-6673-6-33.

de Nazelle, A. *et al.* (2012) 'A travel mode comparison of commuters' exposures to air pollutants in Barcelona', *Atmospheric Environment*. Elsevier Ltd, 59, pp. 151–159. doi: 10.1016/j.atmosenv.2012.05.013.

de Nazelle, A., Aguilera, I., *et al.* (2013) 'Comparison of performance of land use regression models derived for Catalunya, Spain', *Atmospheric Environment*, 77, pp. 598–606. doi: 10.1016/j.atmosenv.2013.05.054.

de Nazelle, A., Seto, E., *et al.* (2013) 'Improving estimates of air pollution exposure through ubiquitous sensing technologies.', *Environmental pollution (Barking, Essex : 1987)*. Elsevier Ltd, 176, pp. 92–9. doi: 10.1016/j.envpol.2012.12.032.

de Nazelle, A., Bode, O. and Orjuela, J. P. (2017) 'Comparison of air pollution exposures in active vs . passive travel modes in European cities : A quantitative review', *Environment International*. Elsevier Ltd, 99, pp. 151–160. doi: 10.1016/j.envint.2016.12.023.

de Nazelle, A., Rodríguez, D. a and Crawford-Brown, D. (2009) 'The built environment and health: impacts of pedestrian-friendly designs on air pollution exposure.', *The Science of the total environment*. Elsevier B.V., 407(8), pp. 2525–35. doi: 10.1016/j.scitotenv.2009.01.006.

Northover, G. (2017) *Exposure contrasts between travel modes: a comparison across continents*. Imperial College London.

O'Driscoll, R. *et al.* (2016) 'A Portable Emissions Measurement System (PEMS) study of NOxand primary NO2emissions from Euro 6 diesel passenger cars and comparison with COPERT emission factors', *Atmospheric Environment*. Elsevier Ltd, 145(2), pp. 81–91. doi: 10.1016/j.atmosenv.2016.09.021.

Ohashi, Y., Kamioka, M. and Matsuoka, K. (2002) 'Relationship between Respiratory Movements and Energy Efficiency in the Post-Exercise Recovery Phase.', *Journal of the Japanese Physical Therapy Association*, 4(1), pp. 7–11. doi: 10.1298/jjpta.4.7.

Oldenkamp, R., Van Zelm, R. and Huijbregts, M. A. J. (2016) 'Valuing the human health damage caused by the fraud of Volkswagen', *Environmental Pollution*. Elsevier Ltd, 212, pp. 121–127. doi: 10.1016/j.envpol.2016.01.053.

Orío Hernández, A. *et al.* (2013) *Análisis de la calidad del aire en España: Evolución 2001 - 2012*.

Oxley, T. *et al.* (2009) 'Background, Road and Urban Transport modelling of Air quality Limit values (The BRUTAL model)', *Environmental Modelling and Software*. Elsevier Ltd, 24(9), pp. 1036–1050. doi: 10.1016/j.envsoft.2009.02.011.

Oxley, T. *et al.* (2012) 'Pollution abatement from road transport: Cross-sectoral implications, climate co-benefits and behavioural change', *Environmental Science and Policy*. Elsevier Ltd, 19–20, pp. 16–32. doi: 10.1016/j.envsci.2012.01.004.

Oxley, T. *et al.* (2013) 'Modelling future impacts of air pollution using the multi-scale UK Integrated Assessment Model (UKIAM)', *Environment International*. Elsevier Ltd, 61, pp. 17–35. doi: 10.1016/j.envint.2013.09.009.

Pant, P., Huynh, W. and Peltier, R. E. (2018) 'ScienceDirect Exposure to air pollutants in Vietnam : Assessing potential risk for tourists', *Journal of Environmental Sciences*. Elsevier B.V., 73, pp. 147–154. doi: 10.1016/j.jes.2018.01.023.

Pérez, N. *et al.* (2010) 'Variability of particle number, black carbon, and PM10, PM 2.5, and PM1 Levels and Speciation: Influence of road traffic emissions on urban air quality', *Aerosol Science and Technology*, 44(7), pp. 487–499. doi: 10.1080/02786821003758286.

Physick, W. *et al.* (2011) 'Measurements of personal exposure to NO2 and modelling using ambient concentrations and activity data', *Atmospheric Environment*, 45(12), pp. 2095–2102. doi: 10.1016/j.atmosenv.2011.01.063.

Ragettli, M. S. *et al.* (2013) 'Commuter exposure to ultrafine particles in different urban locations, transportation modes and routes', *Atmospheric Environment*. Elsevier Ltd, 77, pp. 376–384. doi: 10.1016/j.atmosenv.2013.05.003.

Ragettli, M. S. *et al.* (2014) 'The relevance of commuter and work/school exposure in an epidemiological study on traffic-related air pollution.', *Journal of exposure science & environmental epidemiology*. Nature America, Inc. doi: 10.1038/jes.2014.83.

Richard A. Becker *et al.* (2016) 'maps: Draw Geographical Maps. R package'. Available at: https://cran.r-project.org/package=maps.

Russo, A. *et al.* (2014) 'NO2, PM10 and O3 urban concentrations and its association with circulation weather types in Portugal', *Atmospheric Environment*. Elsevier Ltd, 89(2), pp. 768–785. doi: 10.1016/j.atmosenv.2014.02.010.

Sahsuvaroglu, T. *et al.* (2006) 'A land use regression model for predicting ambient concentrations of nitrogen dioxide in hamilton, ontario, canada', *Journal of the Air and*

Waste Management Association, 56(8), pp. 1059–1069. doi: 10.1080/10473289.2006.10464542.

Samet, J. M., Mawbwry, M. C. and Spengler, J. D. (1987) 'Respiratory effects of indoor air pollution', *Allergy and clinical immunology*, 79(1).

Sarnat, J. A. *et al.* (2014) 'Exposure to traffic pollution, acute inflammation and autonomic response in a panel of car commuters', *Environmental Research*, 133, pp. 66–76. doi: 10.1016/j.envres.2014.05.004.

Schembari, A. *et al.* (2013) 'Personal, indoor and outdoor air pollution levels among pregnant women', *Atmospheric Environment*. Elsevier Ltd, 64(2), pp. 287–295. doi: 10.1016/j.atmosenv.2012.09.053.

Setton, E. *et al.* (2011) 'The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates', *Journal of Exposure Science and Environmental Epidemiology*. Nature Publishing Group, 21(1), pp. 42–48. doi: 10.1038/jes.2010.14.

Shah, A. S. V. *et al.* (2015) 'Short term exposure to air pollution and stroke: Systematic review and meta-analysis', *BMJ (Online)*, 350(January). doi: 10.1136/BMJ.h1295.

Slowik, J. G. *et al.* (2007) 'An Inter-Comparison of Instruments Measuring Black Carbon Content of Soot Particles', *Aerosol Science and Technology*, 41(3), pp. 295–314. doi: 10.1080/02786820701197078.

Smith, J. D. *et al.* (2016) 'London Hybrid Exposure Model: Improving Human Exposure Estimates to NO2 and PM2.5 in an Urban Setting', *Environmental Science and Technology*, 50(21), pp. 11760–11768. doi: 10.1021/acs.est.6b01817.

Stone, B. (2008) 'Urban sprawl and air quality in large US cities', *Journal of Environmental Management*, 86(4), pp. 688–698. doi: 10.1016/j.jenvman.2006.12.034.

Strak, M. *et al.* (2010) 'Respiratory health effects of ultrafine and fine particle exposure in cyclists.', *Occupational and environmental medicine*, 67(2), pp. 118–124. doi: 10.1136/oem.2009.046847.

Su, J. G. *et al.* (2015) 'Integrating smart-phone based momentary location tracking with fixed site air quality monitoring for personal exposure assessment', *Science of the Total Environment*. Elsevier B.V., 506–507, pp. 518–526. doi: 10.1016/j.scitotenv.2014.11.022.

Sun, Y., Moshfeghi, Y. and Liu, Z. (2017) 'Exploiting crowdsourced geographic information and GIS for assessment of air pollution exposure during active travel', *Journal of Transport and Health*. Elsevier Ltd, 6(March), pp. 93–104. doi: 10.1016/j.jth.2017.06.004.

Tainio, M. *et al.* (2015) 'Can air pollution negate the health benefits of cycling and walking?', *Preventive Medicine*. The Authors, 87, pp. 233–236. doi: 10.1016/j.ypmed.2016.02.002.

Tandon, P. S. *et al.* (2013) 'Indoor versus outdoor time in preschoolers at child care', *American Journal of Preventive Medicine*. Elsevier Inc., 44(1), pp. 85–88. doi: 10.1016/j.amepre.2012.09.052.

Team, R. D. C. and R Development Core Team, R. (2016) 'R: A Language and Environment for Statistical Computing', *R Foundation for Statistical Computing*. Vienna, Austria, p. 409. doi: 10.1007/978-3-540-74686-7.

Transport for London (TfL) (2018) *Travel in London, report 11*, *Report 11*. Available at: http://www.tfl.gov.uk/assets/downloads/corporate/travel-in-london-report-number-1.pdf.

United States Environmental Protection Agency (EPA) (2018) *About Risk Assessment*.

Valero, N. *et al.* (2009) 'Concentrations and determinants of outdoor, indoor and personal nitrogen dioxide in pregnant women from two Spanish birth cohorts.', *Environment international*, 35(8), pp. 1196–201. doi: 10.1016/j.envint.2009.08.002.

Vienneau, D. *et al.* (2013) 'Western european land use regression incorporating satellite- and ground-based measurements of NO2 and PM10', *Environmental Science and Technology*, 47(23), pp. 13555–13564. doi: 10.1021/es403089q.

Wallace, L. (2000) 'Correlations of personal exposure to particles with outdoor air measurements: A review of recent studies', *Aerosol Science and Technology*, 32(1), pp. 15–25. doi: 10.1080/027868200303894.

Wang, X. *et al.* (2013) 'Application of Statistical Distribution of PM10 Concentration in Air Quality', *Biomedical and Environmental Sciences*. The Editorial Board of Biomedical and Environmental Sciences, 26(8), pp. 638–646. doi: 10.3967/0895-3988.2013.08.002.

Weichenthal, S. *et al.* (2011) 'Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists', *Environmental Health Perspectives*, 119(10), pp. 1373–1378. doi: 10.1289/ehp.1003321.

Williams, R. D. and Knibbs, L. D. (2016) 'Daily personal exposure to black carbon: A pilot study', *Atmospheric Environment*. Elsevier Ltd, 132, pp. 296–299. doi: 10.1016/j.atmosenv.2016.03.023.

Wilson, B. and Chakraborty, A. (2013) 'The environmental impacts of sprawl: Emergent themes from the past decade of planning research', *Sustainability (Switzerland)*, 5(8), pp. 3302–3327. doi: 10.3390/su5083302.

World Health Organization (2010) *Global recommendations on physical activity for health*. Available at: https://www.who.int/dietphysicalactivity/publications/9789241599979/en/.

World Health Organization (2018a) *Ambient (outdoor) air quality and health*.

World Health Organization (2018b) *Ambient air pollution: health impacts*.

Xu, B. and Hao, J. (2017) 'Air quality inside subway metro indoor environment worldwide: A review', *Environment International*. Elsevier, 107(May), pp. 33–46. doi: 10.1016/j.envint.2017.06.016.

Zhang, Q. *et al.* (2015) 'Variations in PM2.5, TSP, BC, and trace gases (NO2, SO2, and O3) between haze and non-haze episodes in winter over Xi'an, China', *Atmospheric Environment*. Elsevier Ltd, 112(2), pp. 64–71. doi: 10.1016/j.atmosenv.2015.04.033.

Zuurbier, M. *et al.* (2010) 'Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route', *Environmental Health Perspectives*, 118(6), pp. 783–789. doi: 10.1289/ehp.0901622.

Zuurbier, M. *et al.* (2011) 'In-traffic air pollution exposure and CC16, blood coagulation, and inflammation markers in healthy adults', *Environmental Health Perspectives*, 119(10), pp. 1384–1389. doi: 10.1289/ehp.1003151.