file sent:

Using low-cost sensor technologies and advanced computational methods to improve dose estimations in health panel studies: Results of the AIRLESS project

Journal:	*Journal Of Exposure Science And Environmental Epidemiology*
Manuscript ID	JESEE-20-3188
Manuscript Type:	Original Article
Date Submitted by the Author:	14-Jan-2020
Complete List of Authors:	CHATZIDIAKOU, LIA; University of Cambridge, Department of Chemistry Krause, Anika; University of Cambridge, Department of Chemistry Han, Yiqun; Imperial College London, and King's College London, MRC-PHE Centre for Environment & Health; Peking University, College of Environmental Sciences and Engineering; King's College London, Department of Analytical, Environmental and Forensic Sciences Chen, Wu; Peking University, College of Environmental Sciences and Engineering Yan, Li; Imperial College London, and King's College London, MRC-PHE Centre for Environment & Health; King's College London, Department of Analytical, Environmental and Forensic Sciences Popoola, Olalekan; University of Cambridge, Department of Chemistry Wu, Yangfeng; Peking University, Clinical Research Institute Liu, Jing; Capital Medical University, Institute of Heart, Lung and Blood Vessel Diseases, Beijing Anzhen Hospital Hu, Min; Peking University, College of Environmental Sciences and Engineering; Peking University, The Beijing Innovation Center for Engineering Science and Advanced Technology Barratt, Benjamin; King's College London, Department of Analytical, Environmental and Forensic Sciences; Imperial College London, and King's COllege London, MRC-PHE Centre for Environment & Health; King's College London, NIHR Health Protection Research Unit in Health Impact of Environmental Hazards Kelly, Frank; Imperial College London, and King's College London, MRC-PHE Centre for Environment and Health; King's College London, Department of Analytical, Environmental and Forensic Sciences; King's College London, NIHR Health Protection Research Unit in Health Impact of Environmental Hazards Zhu, Tong; Peking University, College of Environmental Sciences and Engineering; Peking University, The Beijing Innovation Center for Engineering Science and Advanced Technology Jones, Roderic; University of Cambridge, Department of Chemistry
Keywords:	Air pollution (or Air quality), Exposure sensors, Spatial and computational methods, Exposure assessment, Monitoring methods

SCHOLARONE™
Manuscripts

Using low-cost sensor technologies and advanced computational methods to improve dose estimations in health panel studies: Results of the AIRLESS project

Lia Chatzidiakou[1,+,*], Anika Krause[1,+], Yiqun Han[2,3,4], Wu Chen[3], Li Yan[2,4], Olalekan A.M. Popoola[1], Yangfeng Wu[5], Jing Liu[6], Min Hu[3,7], Ben Barratt[2,4,8], Frank J. Kelly[2,4,8], Tong Zhu[3,7], and Roderic L. Jones[1]

[1]Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK
[2]MRC-PHE Centre for Environment & Health, Imperial College London and King's College London, London, W2 1PG, UK
[3]College of Environmental Sciences and Engineering, Peking University, Beijing, 100871, China
[4]Department of Analytical, Environmental and Forensic Sciences, King's College London, London, SE1 9NH, UK
[5]Peking University Clinical Research Institute, Beijing, 100191, China
[6]Institute of Heart, Lung and Blood Vessel Diseases, Beijing Anzhen Hospital, Capital Medical University, 100029, Beijing
[7]The Beijing Innovation Center for Engineering Science and Advanced Technology, Peking University, Beijing, 100871, China
[8]NIHR Health Protection Research Unit in Health Impact of Environmental Hazards, King's College London, London, SE1 9NH, UK
[*]ec571@cam.ac.uk
[+]these authors contributed equally to this work

ABSTRACT

For Peer Review Only

Abstract

Air pollution epidemiology has primarily relied on fixed outdoor air quality monitoring networks and static populations. Taking advantage of recent advancements in sensor technologies and computational techniques, this paper presents a novel methodological approach that improves dose estimations of multiple air pollutants in large-scale health studies. We show the results of an intensive field campaign that measured personal exposures to gaseous pollutants and particulate matter of a health panel of 251 participants residing in urban and peri-urban Beijing with 60 personal air quality monitors (PAMs). Outdoor air pollution measurements were collected in monitoring stations close to the participants' residential addresses. Based on parameters collected with the PAMs, we developed an advanced computational model that automatically classified time-activity-location patterns of each individual during daily life at high spatial and temporal resolution. Applying this methodological approach in two established cohorts, we found substantial differences between doses estimated from outdoor and personal air quality measurements. The PAM measurements also significantly reduced the correlation between pollutant species often observed in static outdoor measurements, reducing confounding effects. Future work will utilise these improved dose estimations to investigate the underlying mechanisms of air pollution on cardio-pulmonary health outcomes using detailed medical biomarkers in a way that has not been possible before.

keywords: novel sensor technologies, gaseous pollutants, particulate matter, time-activity-location patterns, exposure misclassification, dose estimation, health outcomes

For Peer Review Only

1 Introduction

Over the last decades, in rapidly developing countries, such as China, the disease burden has shifted from a profile dominated by infectious diseases to one increasingly characterised by non-communicable diseases (NCDs)[1]. Air pollution is now the leading environmental risk factor for NCDs resulting in millions of premature deaths and accelerating rates of chronic disease worldwide[2]. Epidemiological studies have had significant impact in the setting of national and international air quality standards to protect global populations from the detrimental effects of air pollution. However, most of these studies commonly derive metrics of short-term exposure from static outdoor monitoring networks with low spatial and temporal resolution[3]. Such measurements are generally highly correlated at these coarser scales and cannot separate the individual health effects of pollutants[4]. Failure to capture the high granularity of total personal exposure introduces exposure misclassification that can lead to bias in health estimations[5,6].

A range of complex interacting factors drive the high ambient air pollution heterogeneity, while individual variability in personal exposure includes a behavioural component[7] as a person moves between different microenvironments with varying emission sources. During daily life, peak exposure events often occur during commuting[8] while the indoor environment is a significant site for exposure in part because people spend substantial fractions (often as much as 90%) of their time indoors[9]. Indoor air is affected by outdoor pollutants penetrating building envelopes with additional indoor sinks, sources and emissions from building materials which cannot be captured by static outdoor monitoring networks[10].

Several studies have identified large discrepancies between personal exposure measurements and outdoor concentrations[7]. These exposure uncertainties may introduce prediction errors and bias with substantial implications for interpreting epidemiological studies on air pollution, particularly the time-series analyses[5]. The between-subject variability is large because air pollutants concentrations vary significantly by both location and activity. Therefore, a comprehensive personal exposure assessment requires two components: (1) the pollutant concentrations the person is exposed to; and, (2) the recording of a person's time-activity patterns which may vary with age, gender, occupation and socio-economic

For Peer Review Only

status[11]. Physical activity levels affect the potential *dose* of inhaled air pollution[12].

In light of this challenge, "Effects of AIR pollution on cardiopuLmonary disEaSe in urban and peri-urban reSidents in Beijing" (AIRLESS)[13] nested within the within the "Air pollution and human health in a Chinese megacity" (APHH) research programme was initiated[14]. The aim of the AIRLESS project was to address the complex issue of multipollutant exposures on cardiopulmonary outcomes. This paper presents the results of the field deployment of 60 portable air pollution sensor platforms for one week during the winter and summer season in 251 participants of the AIRLESS panel study residing in urban and peri-urban China. The expectation is that similar effects would be evident in the general population hence this paper has wider significance than just these cohorts. The main objective of this work is to demonstrate that novel sensor technologies and computational methods offer a paradigm shift in collecting highly resolved measurements of individualised air pollutants improving dose estimations in large-scale health studies.

2 Materials and Methods

This section briefly describes the methodology employed for creating a comprehensive database of validated personal concentrations and time-activity location patterns of 251 participants of a health panel study matched with intensive monitoring of outdoor air pollution levels. In the last subsection, methodologies for estimating dose with traditional and highly resolved exposure metrics are outlined. The AIRLESS project will integrate these results of detailed doses and exposures to multiple air pollutants with changes in cardio-pulmonary health outcomes to ensure the biggest scientific and policy impact.

2.1 The participant sample

The measurements were collected as part of the AIRLESS project which was designed as a panel study with repeated personal exposure and clinical measurements of 123 urban and 128 peri-urban participants during the winter (14th Nov–21st Dec 2016) and summer seasons (22nd May–21st Jun 2017). Each participant carried a PAM for one week in each season. Thirty PAM devices were deployed at both the urban and peri-urban clinic sites, which enabled the recruitment of 30 subjects each week at each location. This paper will mainly focus on the analysis of the winter campaign data. The participants residing in the

For Peer Review Only

urban site were 59-75 years old, and were primarily retired (88%). In the peri-urban site the age range was between 50-73 years old, and their primary occupation was agriculture (53%) followed by retirement 17% and housekeeping 13%[13].

2.2 Outdoor air pollution measurements

Intensive ambient air pollution monitoring campaigns were launched simultaneously next to the urban and peri-urban clinics, which were in close proximity to most subjects' residential addresses. Both of these stations measured background air pollution levels, as they were located away from direct sources. These measurements had the same time resolution as the personal measurements (1 min). A detailed description of the ambient air pollution monitoring campaign is presented in Shi et al., 2019[14].

2.3 The personal air quality monitor (PAM)

The PAM is an autonomous unit that incorporates multiple sensors for activity, and for physical and chemical parameters[15]. The compact and lightweight design of the PAM ($\sim$ 400g) makes the unit suitable for personal exposure assessment. The time resolution of the measurements was set at 1 min time intervals resulting in a battery life on a single charge of $\sim$ 24 hours. The PAM collects multiple timestamped geo-coordinated measurements of gaseous pollutants, particulate matter, temperature, relative humidity, background noise levels and accelerometry. Measurements are transmitted to a secure server through GPRS for further processing.

Previous work[15] characterised the performance of the PAM that integrates multiple miniaturised sensors for nitrogen oxides (NO_x), carbon monoxide (CO), ozone (O_3) and particulate matter (PM) measurements. Overall, the sensors of the PAM captured air pollutants concentrations in unprecedented detail in outdoor and indoor microenvironments across seasons and in different geographical settings making it suitable for collecting highly resolved exposure metrics at large scale. Following the methodology described in that paper, the raw measurements of gaseous pollutants and particulate matter were converted to physical units.

Data capture rate of personal measurements after cleaning was 81% demonstrating the deployment feasibility and participant acceptability of novel sensor technologies.

For Peer Review Only

2.4 The time-activity model

This subsection provides a brief overview of the progressive composite model (Figure 1) that classifies time-activity-location patterns automatically using: (1) auxiliary parameters collected with the PAM (geo-coordinates, background noise and acceleration levels) as input; and (2) machine learning techniques of spatio-temporal clustering, movement analysis methods, geographical information systems (GIS)[16] and rule-based algorithms.

The participants carried the PAMs during their daily life (Figure 1 a). In the first step, the model computes the space-time utilisation distributions of the GPS coordinates for each participant (Figure 1 b)[17]. The resulting metrics (time spent in each location, re-visitation rate and metrics of directional movement) were used to classify each point in one of three core location categories (Figure 1 c): *home, other static locations* and *in transit*. *Sleeping* was further classified in the *home* category using local time, background noise levels and deposition of coarse particulate matter. *In transit* category was classified into five modes of transport (*walking, cycling, motorbike, car/bus and train/tube*) (Figure 1 d) to capture distinct air pollution micro-environments and different inhalation rates.

The time-activity classification of the data collected during the winter campaign showed that, in line with previous research[9], the participants spent as much as 90% of their time at home (Figure 1 e) partly due to socio-economic factors (e.g. little agricultural activity in winter and large percentage of retired participants). Travel behaviour is a complex issue affected by a multitude of factors, for instance supply and costs of transportation alternatives, incomes as well as urban size and spread. In line with previous studies[18], the urban participants spent 5% of their time budget travelling, and covered a larger spatial area than the peri-urban group that were relatively sedentary (Figure 1 a).

2.5 Personal concentrations, exposure and dose estimates

The basic concepts used in exposure assessments were developed in the early 80s[19]. Personal exposure is defined as the contact of a person with a pollutant of concentration c, at a particular time t. We refer to the mean personal exposure as the average pollutant concentration in the visited microenvironment over the corresponding time period[20]. In the following, *exposure misclassification* is defined as the difference between exposure estimated from measurements at a fixed monitoring station and personal exposure measured

For Peer Review Only

by portable sensors.

The air pollution *dose* describes the amount of pollutant that is actually received by the organism by inhalation. As a first approximation, the *potential dose* can be defined as the inhaled amount of a pollutant, assuming a total absorption of the pollutant by the body. In the following, the term "dose" will refer to this potential dose. The dose $D(t)$ per time unit [μg min^{-1}] is the product of the air pollution concentration c [μg m^{-3}] with the inhalation rate f [m^3 min^{-1}][20] (Equation 1). The total dose is the integral of D(t) over a defined period of time, in this study over the participation time of each individual (7 days).

$$D(t) = f(t)\,c(t) \tag{1}$$

In epidemiological studies, ambient monitoring data are typically averaged for the study area and short-term exposure on any given day is assumed to be the same for the entire population[21]. To understand the differences that arise from the spatial resolution of air pollution measurements employed and the varying time-activity-location patterns of individuals, three approaches were adopted to estimate *dose*:

- **Method A** uses air pollution measurements from the static monitoring station closest to the participant's residential address c_{stat}(t) representing the method employed by the majority of epidemiological studies to estimate exposure. Although this approximation is generally poor, the relevant parameter for interpretation is the extent to which actual personal exposures are correlated to the area-average exposure over time[21]. This method uses generic inhalation rates f_{gen} = 9 L m^{-1}[12] as the level of physical activity may not be available.

- **Method B** assumes the same generic inhalation rate f_{gen}, but utilises highly resolved air pollution measurements in the immediate proximity of the participant collected with the PAM c_{PAM}.

- **Method C** estimates intake in an optimal way by using air pollution concentrations measured in the immediate microenvironment of each participant at high temporal

For Peer Review Only

resolution (c_{PAM}), and inhalation rates derived from the physical activity intensity (f_{act}) estimated with the time-activity model (subsection 2.4).

3 Results

3.1 Seasonal variation of outdoor and personal air pollution concentrations

In summary, outdoor air pollution was poor during the winter campaign with all pollutants, except ozone, being significantly higher than during the summer (Figure 2). Synoptic-scale meteorological analysis suggests that the degraded winter outdoor air quality was due to the greater stagnation and weak southerly circulation[14]. More specifically, winter outdoor air pollution was characterised by several high $PM_{2.5}$ pollution events, with peak hourly concentrations ranging up to 617 μg m^{-3}; whereas, during the summer there were events of high ozone concentrations with the highest hourly average of 168 ppb.

Concentrations measured with the PAMs carried by participants showed two distinct profiles (Figure 2) consistent between seasons:

- **Personal CO and NO concentrations**. Partially driven by the outdoor concentrations, levels of CO and NO, measured with the PAMs showed a strong seasonal variation with higher levels measured during the winter season. The difference between personal and outdoor concentrations was much higher during winter indicating stronger sources in proximity to the participants compared with the summer.
- **Personal NO_2, $PM_{2.5}$, and O_3 concentrations**. Contrary to personal CO and NO levels, which broadly followed the outdoor trends, NO_2, $PM_{2.5}$, and O_3 levels (Figure 2) were significantly lower than outdoor levels in both seasons and showed little ($PM_{2.5}$, and O_3) or no (NO_2) seasonal variation.

The personal measurements show that there is a substantial exposure misclassification that could be introduced when using outdoor measurements as exposure metrics. Apart from the substantial difference in the magnitude of personal and outdoor measurements, there is also a poor correlation ($R^2 < 0.2$ across all pollutants; see Figure S1 in Supplementary Material) indicating that in that environment exposure metrics derived from

For Peer Review Only

the outdoor monitoring stations explained an insignificant amount of the variability in personal exposure.

3.2 Correlation between individual pollutants

The previous subsection highlighted that measurements from static outdoor monitoring sites are poor surrogates for personal exposure levels stressing the need for measurements as close as possible to the individual to capture the high granularity of personal exposure.

A further significant limitation introduced when using measurements from static monitoring stations as metrics of exposure is usually the high correlation between different species. For example, COMEAP report[4] concluded that *insufficient evidence on the health impacts of NO_2 due to the high correlation of this pollutant with other traffic-related pollutants such as primary combustion particles, particle number concentration or carbon monoxide*. As a result, the statistical associations of each individual pollutant with a health effect will, to some extent, also reflect the effects of other pollutants in the group.

The difficulty of interpreting the results of highly correlated pollutants persists even when multi-pollutant models are applied in the statistical analysis, as the multicollinearity introduced from highly correlated species prohibits multipollutant health models from assigning specific health effects to individual pollutants. In addition, if one correlated pollutant has a larger exposure misclassification than another, this may result in the effects associated with a causal relationship being under-estimated whilst non-causal associations are spuriously overestimated[4].

The correlation between outdoor pollutants will vary by location, in this case, Figure 3 (a) shows scatter plots between the outdoor NO_2 and $PM_{2.5}$ concentrations measured at the static outdoor monitoring station at the primary urban site used, and (b) measured with mobile sensors carried by the urban participants. While the two pollutants were strongly correlated at the outdoor monitoring station ($R^2 = 0.64$), low correlations were observed in the personal measurements ($R^2 = 0.05$). This demonstrates clearly that personal monitoring can break the correlation between outdoor correlated pollutants because they capture variable emissions from sources in the direct environment of a person with changing compositions.

Figure 3 (b) uses the personal exposure measurements and breaks them down by

For Peer Review Only

location as classified by the time-activity model (subsection 2.4). A low correlation between the two pollutants was observed at home and other static locations ($R^2 < 0.1$), whereas the two pollutants were moderately correlated in transit environments ($R^2 = 0.25$). This is due to similar emission sources in traffic outdoor environments where pollutants are generally more correlated, also reflected in the measurements at the outdoor monitoring site. Although not shown here equivalent arguments apply to CO and O_3.

3.3 Exposure and dose estimations

The large differences between outdoor and personal concentrations highlighted in the previous subsection were driven by time-activity-location patterns of individual participants as well as infiltration rates of outdoor pollutants in visited indoor microenvironments.

The bar plots in Figure 4 show the average inhaled dose of urban and peri-urban participants calculated with the three dose estimation methods described in Section 2.5. A detailed description of the calculations used to create Figures 4, 5 and 6 is given in Section S1.

As methods A and B are integrating a constant inhalation rate and the measurements were taken over the same time period, the average dose depends only on the measured pollutant concentrations in the surrounding microenvironment. Therefore, the doses are directly proportional to the exposures, and the difference between the two methods is a measure of the exposure misclassification between personal and outdoor estimates which was substantial in all cases. For example, the outdoor stations overpredicted $PM_{2.5}$ exposure by up to 4-fold, while exposure to CO was under-predicted by up to 5-fold.

The difference between method B and method C, both derived from personal measurement, was marginal (Figure 4) despite integrating activity-dependent inhalation rates in method C. This was mostly due to the low physical activity levels of the participants which resulted in an average inhalation rate similar to the generic one used for method A and B. The home microenvironment was the most important modifier of personal dose, partly because participants spent most of their time there (Figure 1 e). Additionally, strong indoor sources of CO and NO operated in the home microenvironment elevating personal dose. On the other hand, indoor doses of NO_2, O_3 and $PM_{2.5}$ were lower indicating the presence of strong chemical sinks.

For Peer Review Only

Using $PM_{2.5}$ and NO_2 as an example, Figure 5 (a) shows the total pollutant dose of each participant over their participation week (calculated with dose estimation method C). The contributions from different microenvironments to the total dose are colour-coded. While generally the urban participants receive a lower PM dose than the peri-urban group, the urban participants received higher doses of NO_2 making, therefore, the exposure profiles of these two groups distinct. However, the variation between individual participants was larger than the variation between the two groups (Figure 5 b). Although the participants spent little time in transportation, it was a significant site of exposure to both pollutants particularly for the urban participants.

Figure 6 shows the average pollutant dose per minute the participants inhaled during different activities, calculated with the three methods A, B and C. As method B integrates one generic inhalation rate for all activities, the average dose is proportional to the pollutant concentrations the participants were exposed to during the different activities. The home environment had the biggest impact on CO dose which was probably caused by indoor emission sources such as cooking and heating. The average NO_x dose was highest during street-level transportation possibly due to strong sources in the traffic environment. When inhalation rates were taken into account (method C), the maximum dose was received during active modes of transport (walking, cycling) due to the increased physical activity levels and inhalation rates. It is therefore likely that the inhaled dose would be significantly underestimated in more active subgroups of the population.

4 Discussion

Exposure misclassification of air pollution remains one of the biggest limitations of epidemiological research on the health impacts of air pollution, preventing the discipline to move from general associations to specific ones. In the absence of personal / indoor measurements, health studies have mainly relied on available outdoor monitoring network data to assess short-term exposures[21]. This paper demonstrates a new methodological framework where novel sensor technologies and advanced computational methods offer a paradigm shift to estimate activity-weighted air pollution exposure in large-scale health studies.

In total, 60 validated personal air quality platforms with miniaturised novel sensors

For Peer Review Only

that measure physical parameters, gaseous pollutants and particulate matter were deployed to 251 participants of two established cohorts residing in urban and peri-urban Beijing, China[13]. Time-activity-location classifications were derived automatically using GPS coordinates, accelerometry and background noise levels collected with the personal monitors. Because such auxiliary data can be collected with widely used smartphones from a large number of the population[22], this technique potentially provides unobtrusive means of enhancing epidemiological exposure data at low cost minimising participant burden.

The relatively sedentary elderly participants spent approximately 90% of their time at home and as little as 2% outdoors, which is in line with previous research in developed countries[9]. The home environment was, therefore, the major contributor to overall exposure, and an important modifier of personal concentrations for all investigated air pollutant species. Exposure differences between the two participant groups were attributed partly to the variation in domestic energy use e.g. in winter the urban building stock in China relies on centralised gas heating system, while traditional biomass and coal stoves remain the key emission source for heating and cooking in peri-urban areas. However, the exposure variability between participants was larger than the variability between the two groups, stressing the need to go beyond current methodologies to estimate population exposures.

We found low correlations and substantial differences in the magnitude estimated from outdoor and personal air quality measurements. An important implication for health studies is that relying on outdoor measurements could introduce significant error and bias in health models depending on individual pollutants' chemical reactivity and strength of local emission sources. The magnitude of the health effects derived from improved exposure estimates is likely to be different than previous estimations using outdoor measurements as metrics of exposure[3]. The extend of misclassification is hard to quantify as it varies significantly with season and location.

Traffic-related pollutants, such as NO_2 and $PM_{2.5}$, are generally highly correlated when measured at coarse spatial and temporal scales, making it challenging to distinguish a causal link between a single pollutant and a specific health outcome[4]. A major advantage

For Peer Review Only

of the proposed methodology is that novel sensor technologies enable the collection of personal measurements at high spatial resolution, and therefore significantly reduce the correlation between individual air pollutants observed at monitoring stations.

Further work matches these estimations with detailed medical biomarkers to draw more reliable associations between air pollution exposure and health impacts[13]. This paper focused on a participant sample with specific personal and socio-economic characteristics that resulted in generally low physical activity. However, the demonstrated approach is applicable in diverse geographical settings and subgroups. The expectation is that the dose misclassification will likely be larger in subpopulations that are more physically active, such as children, increasing the importance of the proposed methodology in disentangling the complex mechanisms of health risks and individual susceptibility to air pollution.

References

1. Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The Lancet. 2017 5;389(10082):1907–1918.

2. World Health Organisation. Air pollution levels rising in many of the world's poorest cities. http://www.who.int/mediacentre/news/releases/2016/air-pollution-rising/en/#.WPn5eqjj5qk.mendeley&title=Air pollution levels rising in many of the world's poorest cities&description=More than 80%25 of people living in urban areas that monitor air pollution; 2016.

3. Zkaynak HK, Baxter LK, Dionisio KL, Burke J, Özkaynak H, Baxter LK, et al. Air pollution exposure prediction approaches used in air pollution epidemiology studies. Journal of Exposure Science and Environmental Epidemiology. 2013;23(6):566–572.

4. Committee on the Medical Effects of Air Pollutants. Associations of long-term average concentrations of nitrogen dioxide with mortality. Public Health England; 2018.

5. Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, et al. Exposure measurement error in time-series studies of air pollution: Concepts and consequences.

For Peer Review Only

Environmental Health Perspectives. 2000;108(5):419–426.

6. Sarnat SE, Sarnat JA, Mulholland J, Isakov V, Özkaynak H, Chang HH, et al. Application of alternative spatiotemporal metrics of ambient air pollution exposure in a time-series epidemiological study in Atlanta. Journal of Exposure Science and Environmental Epidemiology. 2013;23(6):593–605.

7. Steinle S, Reis S, Sabel CE. Quantifying human exposure to air pollution-Moving from static monitoring to spatio-temporally resolved personal exposure assessment. Science of The Total Environment. 2013;443:184–193.

8. Karanasiou A, Viana M, Querol X, Moreno T, de Leeuw F. Assessment of personal exposure to particulate air pollution during commuting in European cities—Recommendations and policy implications. Science of The Total Environment. 2014 8;490:785–797.

9. Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, et al. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. Journal of Exposure Analysis and Environmental Epidemiology. 2001;11(3):231–252.

10. Sundell J, Levin H, Nazaroff WW, Cain WS, Fisk WJ, Grimsrud DT, et al. Ventilation rates and health: Multidisciplinary review of the scientific literature. Indoor Air. 2011 6;21(3):191–204.

11. Schweizer C, Edwards RD, Bayer-Oglesby L, Gauderman WJ, Ilacqua V, Juhani Jantunen M, et al. Indoor time-microenvironment-activity patterns in seven regions of Europe. Journal of Exposure Science and Environmental Epidemiology. 2007 3;17(2):170–181.

12. US Environmental Protection Agency (EPA). Exposure Factors Handbook: 2011 Edition. National Center for Environmental Assessment, Washington, DC. 2011;p. 1–1466.

13. Han Y, Chen W, Chatzidiakou L, Yan L, Zhang H, Wang Y, et al.. Effects of AIR pollution on cardiopuLmonary disEaSe in urban and peri-urban reSidents in Beijing: protocol for the AIRLESS study (in press);.

For Peer Review Only

14. Shi Z, Vu T, Kotthaus S, Harrison RM, Grimmond S, Yue S, et al. Introduction to the special issue "in-depth study of air pollution sources and processes within Beijing and its surrounding region (APHH-Beijing)". Atmospheric Chemistry and Physics. 2019;19(11):7519–7546.

15. Chatzidiakou L, Krause A, Popoola OAM, Di Antonio A, Kellaway M, Han Y, et al. Characterising low-cost sensors in highly portable platforms to quantify personal exposure in diverse environments. Atmospheric Measurement Techniques. 2019 5;12:4643–4657.

16. OpenStreetMap contributors. OpenStreetMap. Planet dump, retrieved from https://planet.osm.org; 2017.

17. Lyons AJ, Turner WC, Getz WM. Home range plus: A space-time characterization of movement over real landscapes. Movement Ecology. 2013 7;1(1):2.

18. Spalt EW, Curl CL, Allen RW, Cohen M, Williams K, Hirsch Ja, et al. Factors influencing time-location patterns and their impact on estimates of exposure: the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). Journal of Exposure Science and Environmental Epidemiology. 2016;26(4):341–348.

19. Ott WR. Concepts of human exposure to air pollution. Environment International. 1982 1;7(3):179–196.

20. Monn C. Exposure assessment of air pollutants: a review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. Atmospheric Environment. 2001;35(1):1–32.

21. Brauer M. How much, how long, what, and where: Air pollution exposure assessment for epidemiologic studies of respiratory disease. Proceedings of the American Thoracic Society. 2010;7(2):111–115.

22. De Nazelle A, Seto E, Donaire-Gonzalez D, Mendez M, Matamala J, Nieuwenhuijsen MJ, et al. Improving estimates of air pollution exposure through ubiquitous sensing technologies. Environmental Pollution. 2013 5;176:92–99.

23. Calenge C. Analysis of animal movements in R: the adehabitatLT package. Office national de la chasse et de la faune sauvage; 2019.

For Peer Review Only

Acknowledgments

We would like to thank all members of the AIRLESS team who helped during fieldwork in urban Beijing and peri-urban sites. We would also like to thank the AIRPOLL and AIRPRO study team for the collected ambient air pollution data for reference calibration and further health analysis.

Conflict of interest

None

Author contributions statement

LC and AK have contributed equally to this paper. It was conceptualised by LC, AK and RLJ. The field deployment was performed by LC, AK, YH, LY and WC. The PIs of the two cohorts were YW and JL. MH was in charge of the Peking University monitoring station. The data curation was performed by LC and AK. LC, AK, OAMP and RLJ contributed to the formal data analysis. Data were visualised by AK and LC. Resources were provided by BB, FJK, TZ and RLJ. The original draft was written by LC and AK and reviewed and edited by YH, BB and RLJ.

Funding

This project is jointly supported by the National Natural Science Foundation of China (NSFC Grant 81571130100) as well as the Natural Environment Research Council (NERC Grant NE/N007018/1) of the UK. The NSFC funding is mainly used to support the field work in China, and NERC funding is mainly used for coordination and the further analysis.

For Peer Review Only

List of Figures

Figure 1. Flow chart of time-activity model. (a) Raw GPS data of 251 participants carrying 60 PAMs during the winter fieldwork campaign (14th Nov–21st Dec 2016) plotted on urban and peri-urban maps. Map data Google 2019. (b) 3D map of a representative participant illustrating the relative amount of time spent in visited locations. The space-time utilisation distribution was constructed using advanced spatio-temporal analysis of the GPS data[17]. (c) Separation of static clusters from clusters with directional movement using derived parameters from step b. (d) Classification of mode of transport using movement analysis methods[23], GIS[16], PAM data (e.g. speed, acceleration) and questionnaire responses collected in the panel study. (e) Time spent in different locations by the two cohorts.

Figure 2. The white whisker box plots illustrate outdoor air pollution levels measured at the reference monitoring stations at the urban and suburban sites during the summer (May-June 2017) and winter (Nov-Dec 2016) campaigns. The blue boxplots show the levels measured with 60 PAMs (blue) deployed to 251 participants at the urban and peri-urban site during the same periods.

Figure 3. Correlation between the concentration measurements of two different pollution species (NO_2 and $PM_{2.5}$). (a) Measurements of a static monitoring station (PKU Beijing, grey) compared to measurements of portable monitors (urban site, black). Measurements were taken over the same time period. (b) Portable monitor measurements (inset of graph a) separated by location (home, other static and transit).

Figure 4. Dose estimations of various air pollutants using methods A, B and C (Section 2.5) in the urban (left) and peri-urban (right) participants.

Figure 5. (a) Total pollutant dose of each participant over one week for $PM_{2.5}$ (top) and NO_2 (bottom), estimated using method C. Colours mark the contributions of each activity to the total dose. (b): Density plots of the individual doses in the urban (blue) and peri-urban (black) cohort, the vertical dotted lines mark the average weekly dose of the two cohorts.

Figure 6. Mean pollutant dose per minute during different activities using the three estimation methods A, B and C (averaged over the urban and peri-urban cohort).

For Peer Review Only

Supplementary Material

S1. Detailed description of the data analysis for Figures 4 - 6

Figure 5 (a) shows the total weekly doses $D(p)$ of each participant p split by activity act. For this graph, the total dose per activity $D_{act}(p)$ was first calculated for each participant by integrating the time-dependent dose $D(t,p)$ over each point of time in which the activity was performed (Equation 2). For instance, if participant 1 slept every night from 22:00 pm to 07:00 am, then their total dose while sleeping D_{sleep} would be determined as the integral of their personal dose $D(t)$ over all 7 nights between these hours (p = 1, act = sleep, $t_{act} = t_{sleep}$ = daily 22:00 pm - 07:00 am).

$$D_{act}(p) = \int_{t_{act}} D(t,p)dt \tag{2}$$

$D(t)$ was calculated as the product of pollutant concentration and inhalation rate following Equation 1 (in this case using Method C). The total dose $D(p)$ of each participant was calculated as the sum of all activity doses $D_{act}(p)$ (Equation 3). Missing data were corrected for by normalising each participation period to exactly 7 days.
The contributions of each activity to the total dose for each participant $\chi_{act}(p)$ and averaged over all participants $\bar{\chi}_{act}$ were calculated using Equations 4 and 5 (used in Figure 4).

$$D(p) = \sum_{act} D_{act}(p) \tag{3}$$

$$\chi_{act}(p) = D_{act}(p)/D(p) \tag{4}$$

$$\bar{\chi}_{act} = \frac{\chi_{act}(p_1) + \chi_{act}(p_2) + \ldots + \chi_{act}(p_i)}{i} \tag{5}$$

Figure 6 shows the mean pollutant dose per minute during different activities averaged over all participants of the two cohorts $\bar{D}_{act}(cohort)$. Firstly, the mean dose per activity $\bar{D}_{act}$ was determined for each participant by dividing each activity dose $D_{act}(p)$ by the total time the participant spent on that activity t_{act} (Equation 6). For instance, if participant 1 spent 50 hours sleeping during his participation week and inhaled 75 mg CO over this time, their mean CO dose while sleeping would be 75 mg / 3000 min = 25 μg per min. The individual results of each participant were then averaged over both cohorts (Equation 7).

$$\bar{D}_{act}(p) = D_{act}(p)/t_{act}(p) \tag{6}$$

$$\bar{D}_{act}(cohort) = \frac{\bar{D}_{act}(p_1) + \bar{D}_{act}(p_2) + \ldots + \bar{D}_{act}(p_i)}{i} \tag{7}$$

Figure 4 shows the mean dose per minute averaged over the two cohorts (urban and peri-urban) calculated with different dose estimation approaches. The mean dose per minute was determined by dividing the total dose of a participant $D(p)$ (calculated with

For Peer Review Only

the three different dose estimation methods A, B and C) by the total participation time $t_{tot}(p)$ (cf. Equation 6). The average contributions from different locations shown in the bar of method C (colours) were determined as the average contributions to the total dose as shown in Equation 5.

Note, that the results in Figure 4 are not equal to the average of the mean dose per minute by activity of Figure 6 because the activities were performed over different time periods (e.g. the contribution of the mean dose while sleeping contributes more to the total dose than the mean dose while cycling as people spend much more time sleeping - unless we are at the Tour de France).

S2. Correlation between personal and outdoor measurements

Epidemiological studies assume an association between outdoor and personal exposure[21], and therefore that outdoor measurements can be used as a surrogate for personal exposure. Graph S1 shows the correlation between measurements from portable monitors and the nearest monitoring station during the winter (grey) and summer (green) campaign of the AIRLESS project. The graphs show poor correlation ($R^2 < 0.2$) between the two exposure estimation approaches across all pollutants and seasons demonstrating that outdoor concentrations are a poor predictor for personal exposures.

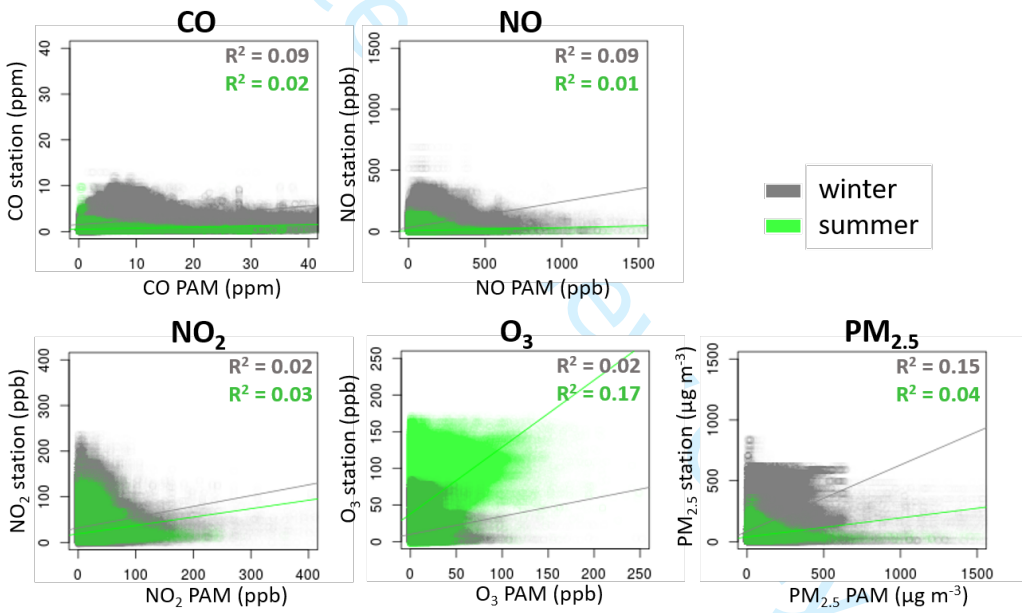
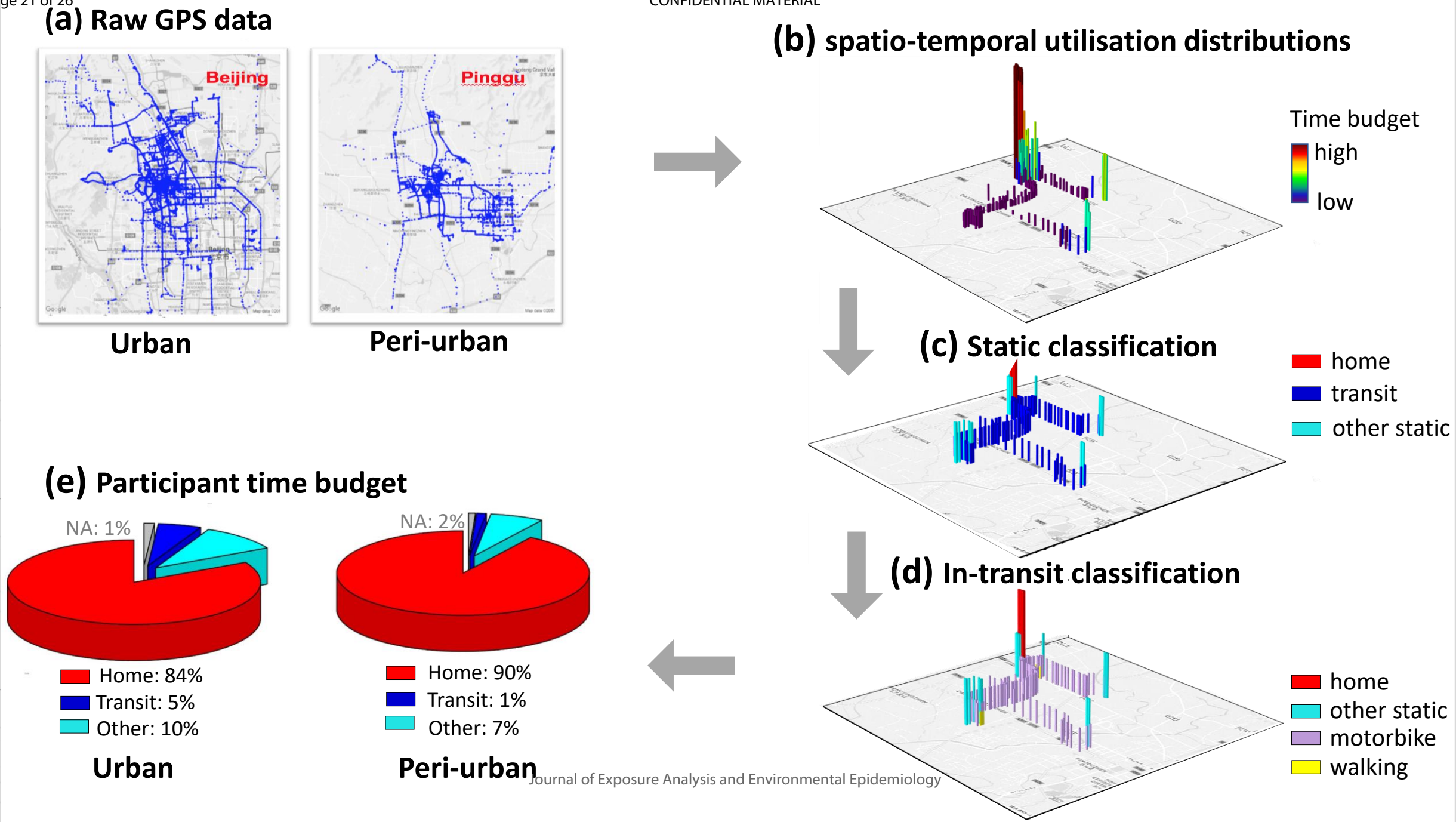
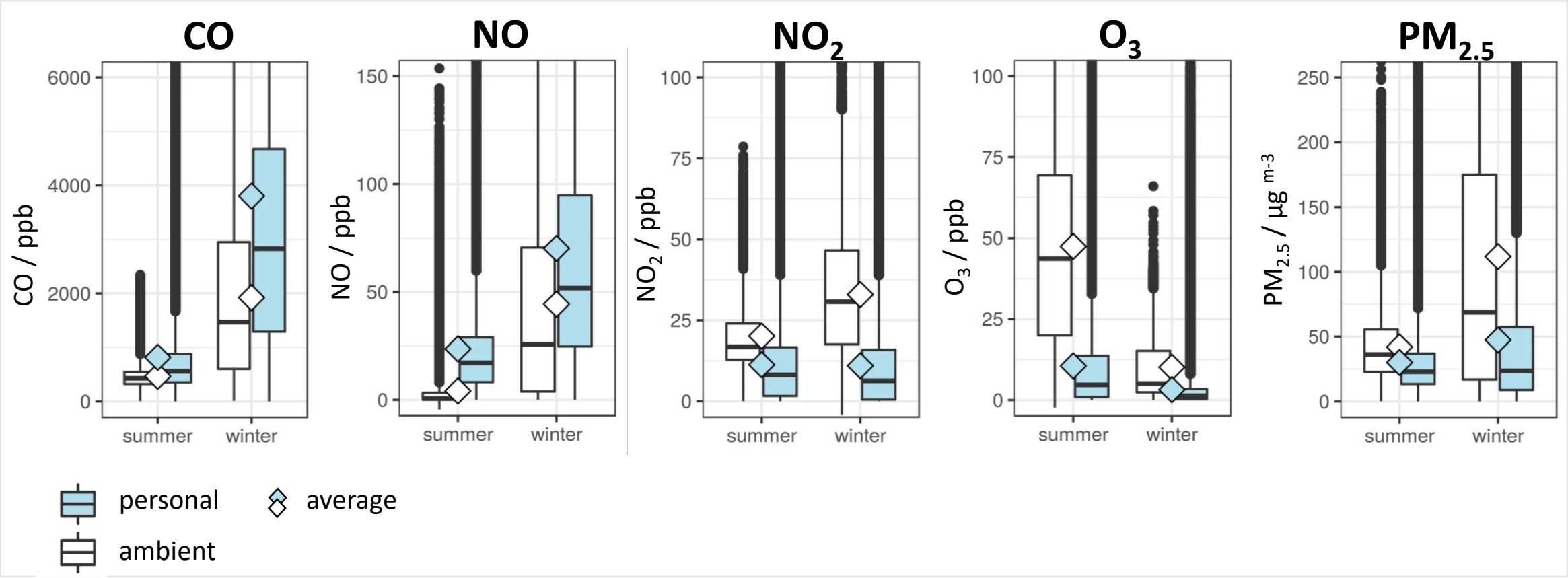
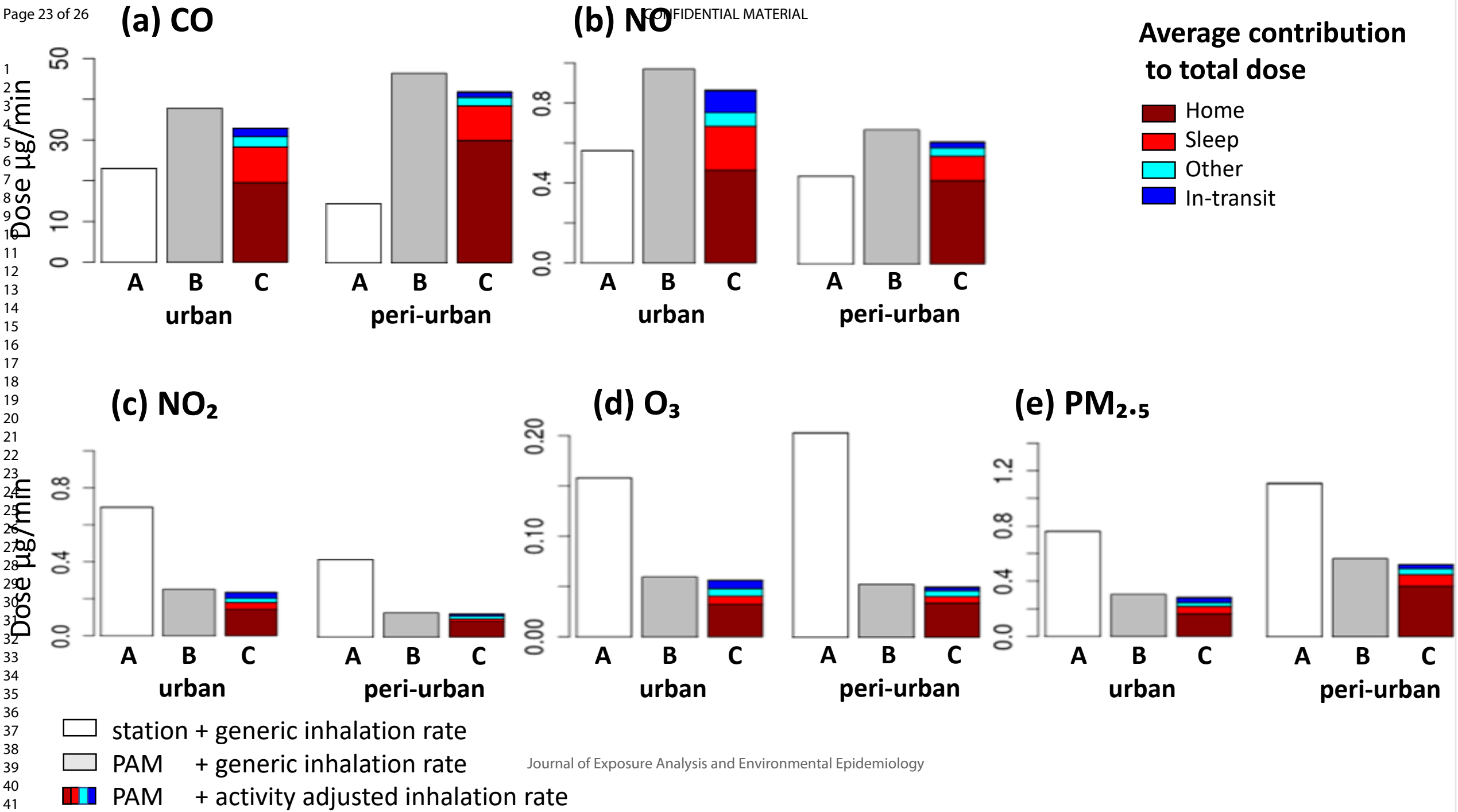


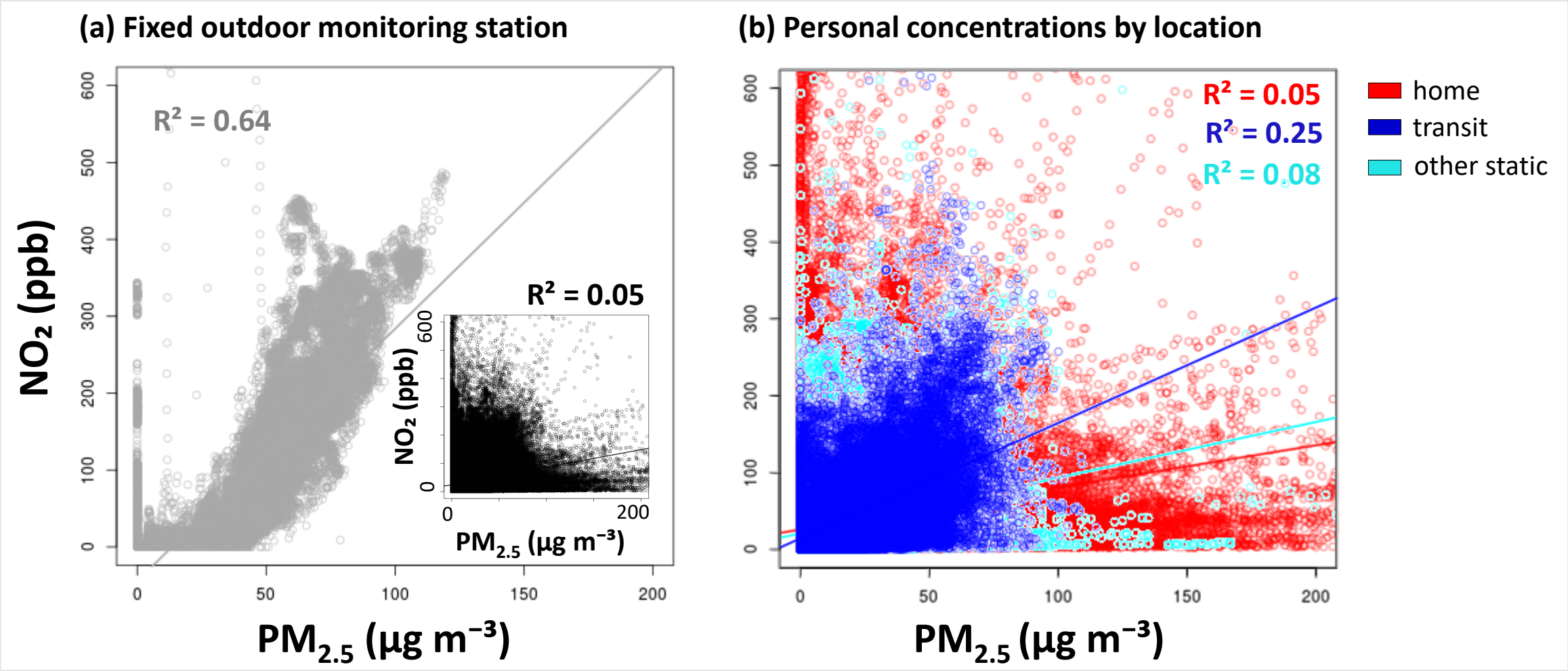
Figure S1. Correlation between the concentration measurements of personal and outdoor pollutants during the winter (grey) and summer (green) seasons. Personal and outdoor measurements are collected at 1 min sampling interval.


(a) Raw GPS data
Beijing
Pinggu
Urban
Peri-urban
(b) spatio-temporal utilisation distributions
Time budget
high
low
(c) Static classification
home
transit
other static
(d) In-transit classification
home
other static
motorbike
walking
(e) Participant time budget
NA: 1%
Home: 84%
Transit: 5%
Other: 10%
Urban
NA: 2%
Home: 90%
Transit: 1%
Other: 7%
Peri-urban

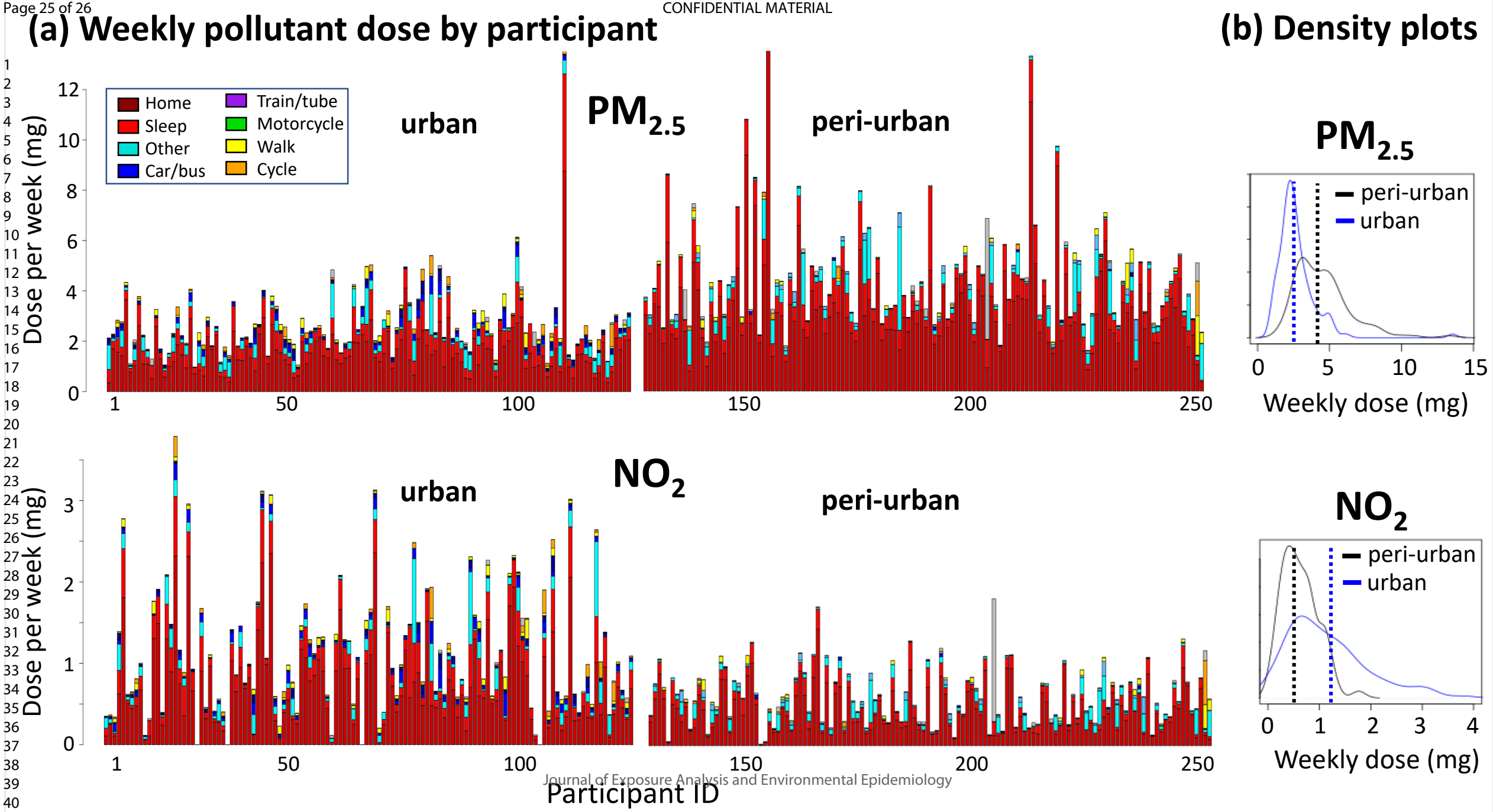

CO
NO
NO2
O3
PM2.5
CO / ppb
NO / ppb
NO2 / ppb
O3 / ppb
PM2.5 / µg m-3
summer
winter
personal
ambient
average

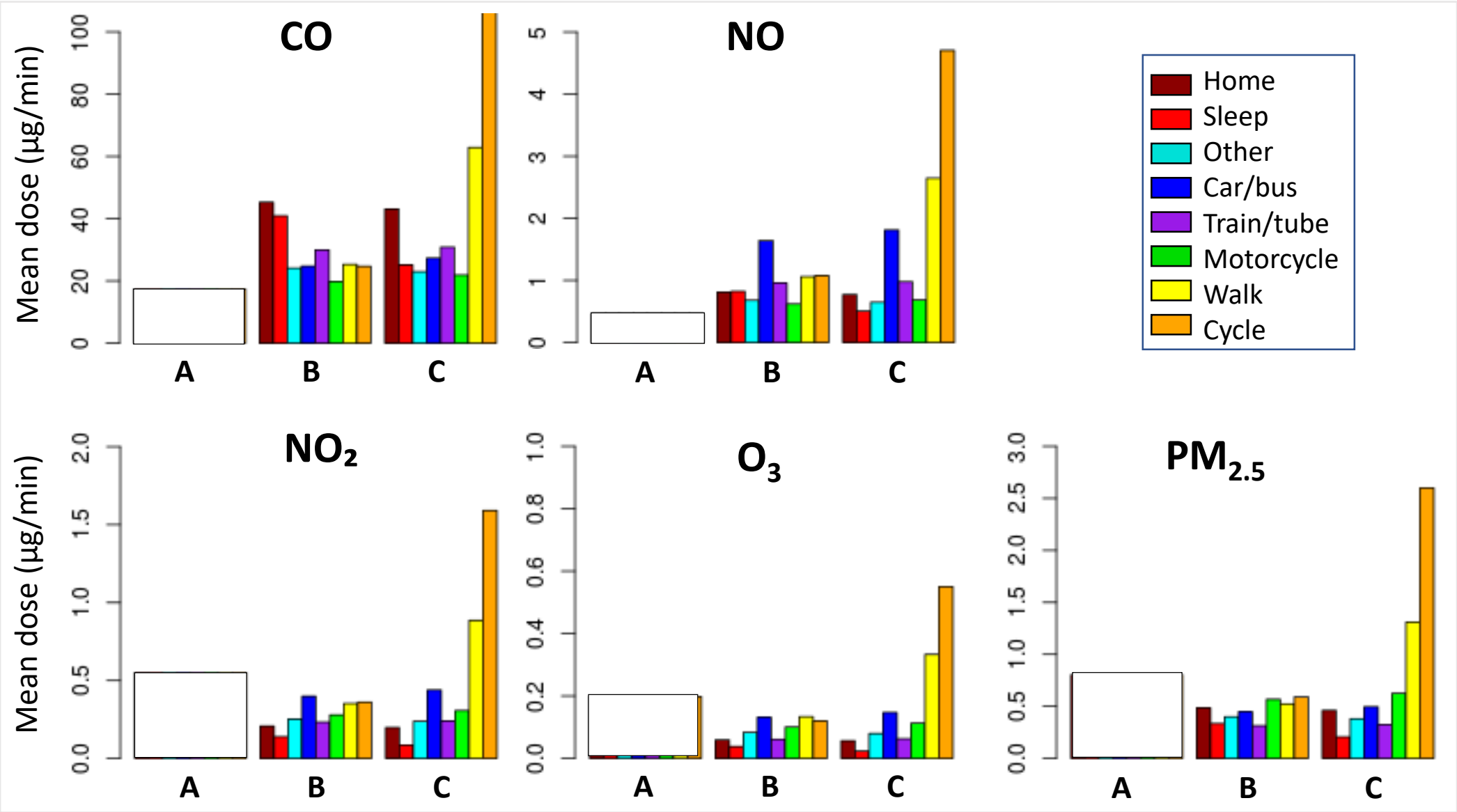
CONFIDENTIAL MATERIAL

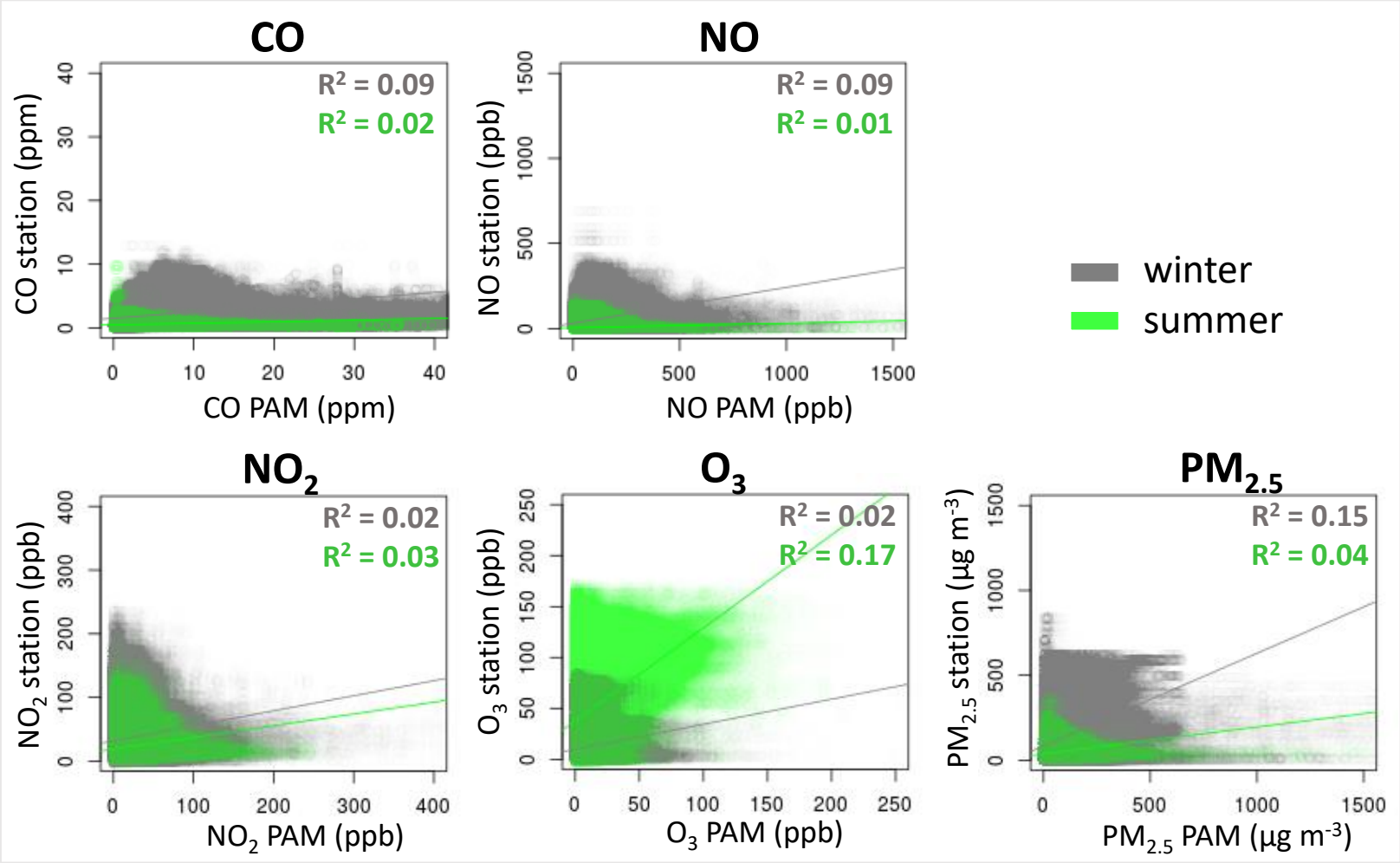



(a) Fixed outdoor monitoring station
R² = 0.64
R² = 0.05
NO₂ (ppb)
PM2.5 (µg m⁻³)
(b) Personal concentrations by location
R² = 0.05
R² = 0.25
R² = 0.08
home
transit
other static
NO₂ (ppb)
PM2.5 (µg m⁻³)

CONFIDENTIAL MATERIAL




CO
NO
NO2
O3
PM2.5
Mean dose (µg/min)
Mean dose (µg/min)
A
B
C
Home
Sleep
Other
Car/bus
Train/tube
Motorcycle
Walk
Cycle


CO
R² = 0.09
R² = 0.02
CO station (ppm)
CO PAM (ppm)
NO
R² = 0.09
R² = 0.01
NO station (ppb)
NO PAM (ppb)
winter
summer
NO2
R² = 0.02
R² = 0.03
NO2 station (ppb)
NO2 PAM (ppb)
O3
R² = 0.02
R² = 0.17
O3 station (ppb)
O3 PAM (ppb)
PM2.5
R² = 0.15
R² = 0.04
PM2.5 station (µg m-3)
PM2.5 PAM (µg m-3)