A methodological framework to assess multi-pollutant personal air quality exposure for improved health associations

Elizabeth L. Martin

Ph.D.

Supervisor: Prof R.L. Jones

Yusuf Hamied Department of Chemistry

University of Cambridge

Jesus College

September 2023

Abstract

A methodological framework to assess multi-pollutant personal air quality exposure for improved health associations

Elizabeth L. Martin

Current assessments link poor air quality to around seven million premature deaths worldwide annually. However, exposure studies, often utilising measurements from stationary outdoor instruments from sparse monitoring networks, cannot capture spatial heterogeneity or the fact that people spend significant fractions of their time indoors. This failure to assess the actual pollution exposure individuals receive leads to inaccuracies in pollution-health associations, potentially masking the factors that drive the observed health responses, resulting in misinformed policies.

To address these limitations, a portable personal air quality monitor (PAM) was developed, allowing for the assessment of actual personal exposure to key pollutants: CO, NO, NO_2, O_3, and $PM_{2.5}$, as well as providing location (GPS) and other parameters for time-activity assessment.

The work in this thesis develops a framework, which, when applied to large scale fieldwork studies, is capable of disaggregating personal exposure by source and linking it to health parameters for hundreds of participants. At the core of the framework is a methodology for apportioning personal exposure into pollution generated by indoor sources and pollution generated by outdoor sources.

This apportionment is achieved by employing a mass-balance model and estimating values of ventilation rates, indoor loss rates and indoor source characteristics, collectively referred to as "exposure determinants".

The framework was applied to data from the AIRLESS project, which involved the deployment of PAMs to 250 residents of Beijing and the surrounding area. Personal exposure to NO_2, O_3 and $PM_{2.5}$ was found to be lower than that inferred from measurements from stationary outdoor reference instruments, suggestive of indoor

losses for these pollutants.

The results show differences between indoor-generated and outdoor-generated exposures, for example, 55% of participants' exposure to CO was from indoor sources, compared with 30% of $PM_{2.5}$. Apportioned exposure metrics, for example indoor- and outdoor- generated CO, while the same molecule, may be proxies for different mixtures of pollutants, which may have different health impacts.

As expected, home ventilation rates were higher in the summer than in the winter, and the overall mean ventilation rate was estimated to be 3.12 hr^{-1}, which is comparable to values found in the literature. Knowledge of the seasonal and demographic variability of exposure determinants will be crucial in the future modelling of total personal exposure at the population scale.

This thesis concludes with the construction of a Linear Mixed Effects Model (LMEM), linking the novel exposure metrics and estimated exposure determinants to a health marker, in this case Peak Expiratory Flow (PEF). While the associations with personal exposure and PEF appear minimal in this study (concerns about the accuracy of self-reported PEF as an indicator are raised), it is expected that this framework will be of significant value when extended to directly examine the effects of the novel exposure metrics and estimated exposure determinants on other health parameters. This will provide insights into the source-related health effects of air pollution to drive more effective environmental policy.

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee

Acknowledgements

Firstly, I would like to thank my PhD supervisor, Professor Rod Jones, for providing me with the opportunity to conduct this research, and for his support and guidance throughout the PhD. I would like to thank Dr Lia Chatzidiakou for her mentorship throughout the last four years, and for always helping me when I was stuck.

I would like to acknowledge other members of the Jones group: Jess Fleming for her camaraderie and a great conference trip to the USA (and research trip to a shed in Birmingham); Ray Freshwater and Dr Lekan Popoola for their general support on the fourth floor and in the (literal) field; Dr Anika Krause for welcoming me and for helping me learn R during the first few months; Yilin Li for struggling through LMEMs with me, and Dr Andrea Di Antonio, whose in-depth investigations into the Vajont Dam made me feel comparatively stable during my thesis write-up phase. Extended thanks go to Dr Shiwei Fan, Siriel Saladin, Mushtari Saidikova, Rachel Georgiou, Dr Chiara Giorio, Prof Anja Schmidt, Prof Alex Archibald and the rest of the 4th floor and CAS. Thanks to Libby, Kieran, Mady, Ben and the rest of the Chemgals for making my time in the chemistry department so enjoyable.

The framework presented in this thesis was developed using the AIRLESS dataset. I am very grateful to those involved in the project, especially the volunteers who carried around the PAMs. I would like to thank EPSRC for funding my PhD.

I would like to thank everyone at Jesus College for making my time in Cambridge so special, including the Porters and Roost staff for supporting my well-being, and the boatman, JC, who undid all of their efforts by teaching me to row.

Last but not least, I am immensely grateful to my friends and family, particularly my parents, for their encouragement and support. I've had an amazing four years.

Contents

List of Tables

List of Figures

Abbreviations

AIRLESS	AIR pollution on cardiopuLmonary disEaSe in urban and peri-urban reSidents in Beijing
APHH	Atmospheric Pollution and Human Health in a Chinese Megacity
APPCAP	Air Pollution Prevention and Control Action Plan
ATM	Standard Atmosphere (unit)
BP	Blood Pressure
CFC's	ChoroFluoroCarbons
COMEAP	Committee on the Medical Effects of Air Pollutants
COPD	Chronic Obstructive Pulmonary Disease
COPE	Characterisation of COPD Exacerbations Using Environmental Exposure Modelling
COVID-19	Coronavirus 2 (SARS-CoV-2), a contagious disease caused by the virus severe acute respiratory syndrome
EC	Electrochemical
EPA	Environmental Protection Agency
FeNO	Fractional Exhaled Nitric Oxide
FTP	File Transfer Protocol
FVC%	Forced Vital Capacity
GPRS	General Packet Radio Service
GPS	Global Positioning System

HIC	High Income Country
IC	Integrated Circuit
INCHEM-Py	INdoor CHEMical model in Python
IQR	Interquartile range
I/O	Indoor/Outdoor Ratio
LMEM	Linear Mixed Effects Model
LMIC	Lower Middle Income Country
LPG	Liquefied Petroleum Gas
NO_x	Collective term for Nitrogen Oxides
OPC	Optical Particle Counter
PAM	Personal Air Monitor
PEF	Peak Expiratory Flow
PKU	Peking University
PM	Particulate Matter
$PM_{2.5}$	Particulate Matter with diameter of less than 2.5 micrometers
QA/QC	Quality Assurance/ Quality Control
RH	Relative Humidity
SD	Secure Digital
SOAs	Secondary Organic Aerosols
TSNAs	Tobacco-Specific Nitrosamines
UMIC	Upper Middle Income Country
VOCs	Volatile Organic Compounds
WHO	World Health Organisation
ZTD	Zenith Tropospheric Delay

Chapter 1

Introduction

1.1 Air pollution epidemiology

1.1.1 A brief history of air pollution epidemiology

A link between poor air quality and health is not a recent discovery. The Greek physician Hippocrates attributed illness to the quality of the air over 2400 years ago: "They are likely to have deep, hoarse voices, because of the atmosphere, since it is usually impure and unhealthy in such places."[56]. Similar associations were made within other early societies. The harmful effects of soot/dust on the skin were described in the Book of Exodus, Old Testament, during the 13th century BC in Egypt[86].

However it was not understood how the inhaled air interacted with the body, for example, Aristotle believed that the role of breathing was to cool the heart[4]. Following Priestley's discovery of oxygen in 1774[108] and Lavoisier's investigations into respiration, there was a large increase in research into air quality and epidemiology in the late 1700s and early 1800s. The most prominent example was in 1775 when Percivall Pott made the link between air pollution and cancer by observing a very high incidence of testicular cancer in chimney sweeps resulting in short life expectancy[106].

There was a significant change in public awareness of the health impacts of pollution after the Great Smog event of 1952 when a thick layer of smog covered London for four days. This is thought to be the worst air pollution event in the history of the United Kingdom and had a major effect on environmental research and government regulation (the Clean Air Act was introduced in 1956). In the weeks following the smog event, government medical reports estimated that up to 4,000 people had died as a direct result of the event[152]. Figure 1.1[152] shows a graph, plotted in 1954, of how the number of deaths significantly increased with the concentrations of SO_2 and smoke. Figure 1.1 also shows that the number of deaths remained elevated beyond the period of elevated smoke and SO_2 levels, indicating a lag effect of pollution on health. Recent re-analysis of the data suggests that the total number of fatalities may have been considerably greater, with estimates of between 10,000 and 12,000 deaths[9].

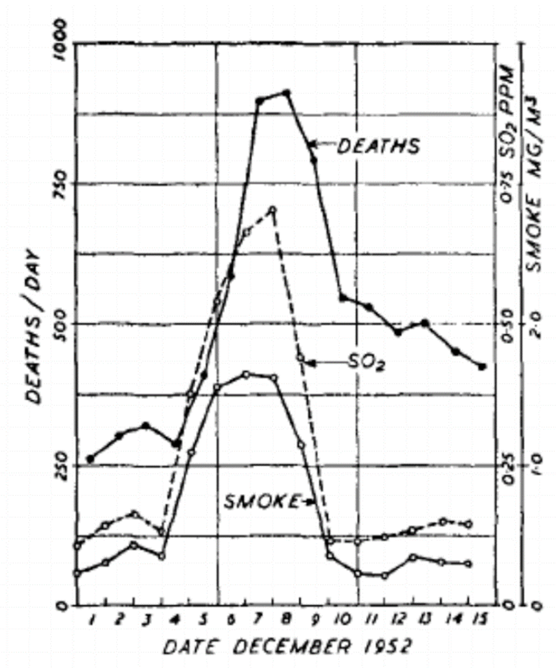

Figure 1.1: The great smog of London: Daily air pollution and death rates during the Great Smog of London, taken from Wilkins et al.[152]

Researchers now derive the risks of pollution on health using three main methods:

- **Epidemiological approaches** including time series analyses from cohort and panel studies
- **Clinical studies** involving controlled exposures of normal and susceptible people
- **Toxicology/ laboratory research** including animal exposure, in vitro approaches and genetic approaches

Epidemiological studies examine the health outcomes of large populations within their natural environment and often involve large sample sizes, allowing researchers to analyse data from a diverse range of individuals with different characteristics, demographics, and susceptibilities. Clinical and laboratory experiments can be used to investigate mechanisms and establish if there is a causal link.

1.1.2 Statistical methods: LMEMs

The development of Linear Mixed Effect Models (LMEMs) in the late 1900s provided researchers with a valuable statistical method to draw pollution health associations from epidemiological data. By incorporating so called fixed and random effects, LMEMs provide more accurate associations and improve the handling of dependencies within the data.

Fixed effects in LMEMs represent systematic and population-level factors (e.g., age, sex, pollution levels) that are expected to have an impact on health outcomes. On the other hand, random effects capture the impact of variables that specific predictions are not available for, such as genetic differences between participants.

The benefits of LMEMs were identified in an early air pollution study in 1984[127], which applied an LMEM to data from a panel study with a binary outcome: Asthma attack (yes/no), and compared the results with a simpler statistical approach[72]. One major advantage was the ability to use all of the data, including that from subjects with incomplete data; LMEMs handle missing data by using all available data to estimate the fixed effects and random effects. The only recorded drawback of LMEMs in this study was their computationally intensive nature, which is largely addressed by modern advances in computing. LMEMs are increasingly becoming the standard tool for investigating associations between pollution exposure and health.

1.1.3 Summary of air pollution epidemiology

Air pollution is known to increase rates of mortality and morbidity globally. Each year, an estimated 4.5 million deaths[a] are attributed to the effect of outdoor air pollution globally[43]. Emerging evidence shows that breathing polluted air has an adverse effect on the respiratory system and recent studies link it to every major organ system, including the central nervous, cardiovascular and pulmonary systems[132].

This PhD aims to further advance the understanding of air pollution on health. It focuses on assessing air pollution exposure of epidemiological cohorts, for five

[a] 6.7 million deaths are attributed to the combined effects of indoor and outdoor exposure to air pollution globally[43]

pollutants: CO, NO, NO_2, O_3 and $PM_{2.5}$ ($PM_{2.5}$ are particles with a diameter of less than 2.5 micrometers). It also demonstrates how an LMEM can be designed for pollutant-health associations. Ambient concentrations of these five pollutants have been shown to affect health, and Appendix A.1.3 reviews their previous health associations. Factors contributing to error in current pollution-health associations are explored below.

1.2 Factors contributing to error in pollution-health associations

1.2.1 Misclassification of exposure

The difference between the exposure metrics used when making the pollution-health association, and the 'true' exposure of the population at risk is referred to as **misclassification** of exposure.

Studies restricted to stationary outdoor measurements will not capture total personal exposure, which in turn results in an inaccurate assessment of the impact that air pollution has on human health. This is because these studies are not able to capture the effects of an individual moving between different microenvironments, and the different concentrations, loss processes and sources within the different microenvironments. The different pollutant levels in a range of microenvironments were measured by Chatzidiakou et al. using a UK cohort of 35 participants and the results are shown in Figure 1.2. Pollutant levels were found to vary between the microenvironments. The participants kept an activity log, and the study involved development of an automated model to detect the microenvironment of the participant.

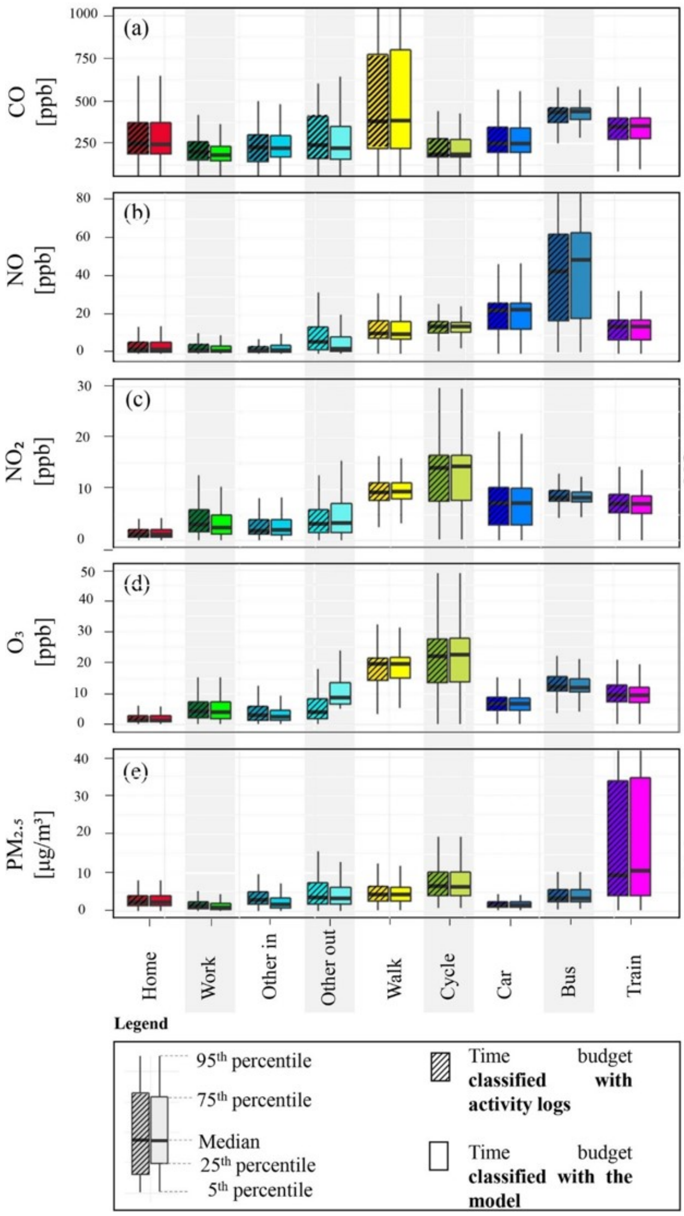

Figure 1.2: Pollutant levels in a range of microenvironments: Box plots of personal exposure of 35 UK participants to multiple pollutants in different microenvironments. For each microenvironment, the left hatched box plot shows the microenvironment classified using participants' activity logs and the right solid-colour box plot using the automated model. This figure was adapted from Chatzidiakou et al.[21]. The full paper can be found in Appendix A.1.2. Pollutant concentrations were found to vary across microenvironments.

1.2.2 Indoor/outdoor ratios to infer indoor exposure for large populations

People spend most of their time indoors. This was demonstrated with the same UK cohort from Figure 1.2, and is shown in Figure 1.3.

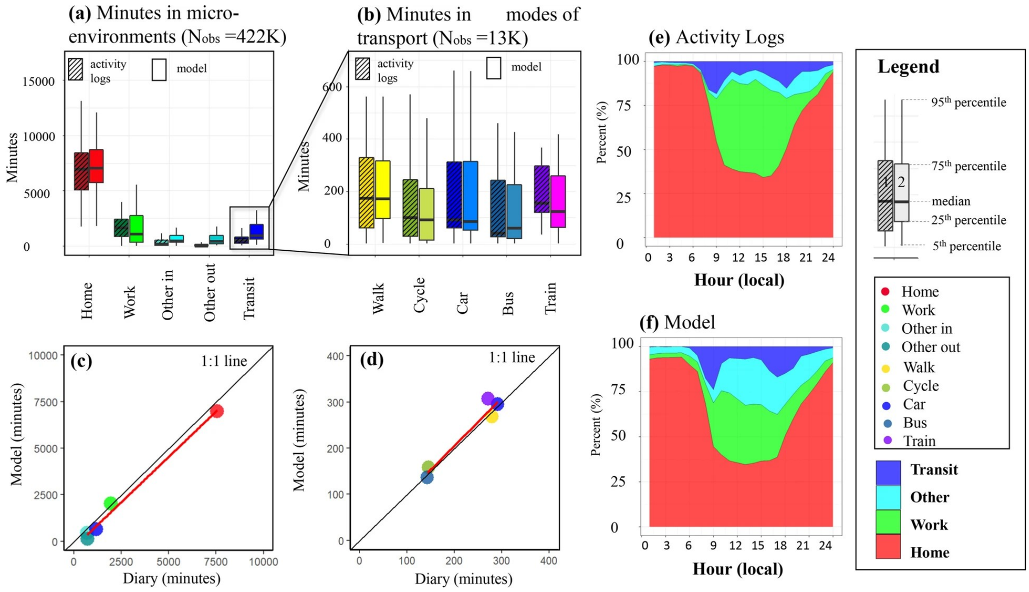

Figure 1.3: Time budget of microenvironments visited: Time budgets of 37 participants residing in London, UK and Cambridge, UK. (a and b) box plots of participants' time budgets in different static microenvironments and modes of transport classified with activity logs (left, shaded box plot) and a developed model for activity classification (right, solid-colour box plot). (c and d) Corresponding scatterplots of mean time (in minutes) spent in visited microenvironments are shown in a colour scale at the bottom. (e and f) Average diurnal time budget profile of all participants classified with the activity logs and with the model. This has been taken from Chatzidiakou et al.[21]. The cohort spent the majority of their time indoors.

A simple method to estimate indoor pollution exposure is to assume that indoor pollutant concentrations and outdoor pollutant concentrations are related by a fixed ratio (indoor-outdoor air quality ratio, abbreviated to I/O ratio). Often studies take this ratio from literature or calculate it from a subset of their participants[126]. The advantage of using a ratio is that it can be applied to accessible outdoor data from already installed, stationary air pollution monitoring stations. Estimations for large populations can be made without the requirement of measuring inside every home or indoor environment. It has been used directly in health studies[117;12].

Figure 1.4 shows how the I/O ratio for NO_2 and $PM_{2.5}$ measured in a range of indoor microenvironments varied over a 24-hour period, evaluated over 6-9 month periods[126].

The I/O ratio for NO_2 is seen to increase by a factor of two during core operation

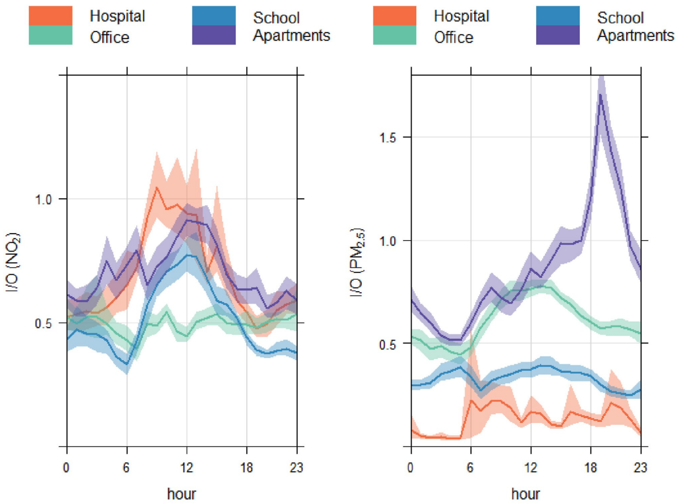

Figure 1.4: Dynamic variability of I/O ratios: Aggregated or 'typical' I/O ratios for NO_2 and $PM_{2.5}$. Evaluated over 6–9 month periods in a school, hospital, office (3 indoor microenvironments from Mon-Fri only) and 18 apartments (living rooms) in the UK. Both species show significant I/O variation over the 24-hour period, for a range of microenvironments. This figure has been taken from Stamp et al.[126]

hours for the school and hospital microenvironments. These two microenvironments were mechanically ventilated. The I/O ratio for $PM_{2.5}$ measured in the apartment microenvironment shows a strong peak around 19:00–20:00 where the I/O ratio reaches above 1.5. The I/O ratio is the result of the building as a pollutant modifier but also of the activity within a building[126]. The I/O value can exceed 1 when indoor sources are dominant. In this case, the strong peak around 19:00-20:00 is attributed to cooking.

This study showed that using a single I/O ratio to infer the indoor concentration of pollutants likely leads to exposure misclassification due the large variations of I/O ratios over a 24-hour period as a result of different building operation and occupant behaviours. Studies have attempted to minimise the effect of occupant behaviour on the I/O ratio by limiting the participant sample to non-smokers[89].

1.2.3 Correlations between individual pollutant species

In the urban environment, humans are exposed to a range of pollutants. These pollutant concentrations are often highly correlated due to common sources.

Ignoring pollutant correlations in pollutant-health models can lead to error. For example, NO_2 and $PM_{2.5}$ have both been identified as key pollutants with respect to health (see Appendix A.1.3). The Committee On the Medical Effects of Air Pollution (COMEAP) have released multiple reports detailing the scientific and methodological challenges in interpreting the extent of the independence of the associations of mortality with concentrations of NO_2 and $PM_{2.5}$ [25;26;27]. They have carried out systematic reviews and a meta-analysis of epidemiological studies of long-term average concentrations of ambient pollutant levels in the outdoor environment. In the outdoor environment, traffic-related pollution is dominant and NO_2 and $PM_{2.5}$ are strongly correlated in traffic emissions. COMEAP report that these correlations (Pearson's correlation coefficient was found to be as high as 0.85 in one of their meta-analysis studies[17]) made it impossible to estimate reliably the effects of the explanatory variables individually.

These correlations must be considered when making health associations; the effect of ambient CO concentrations on stroke mortality decreased and became non-significant when controlling for co-emitted species ($PM_{2.5}$, NO_2, and SO_2)[79].

O_3 has been found to be anti-correlated with CO, NO_2 and $PM_{2.5}$ in some cities[17;133;46;146]. Additionally, the output of the ADMS-Urban model predicted anti-correlation between NO_2 and O_3 in London, UK as shown in Figure 1.5.

This anti-correlation of O_3 with pollutants such as NO_2, as well as positive correlation with temperature, can result in misleading observed O_3-health associations[17;10]. In Europe, an apparent protective effect of O_3 on health has been observed; the Office for National Statistics in the UK found a negative correlation with COVID-19 mortality and O_3 exposure. In this study, exposure to higher ozone was suggested to act as proxy for living in rural environments[36].

1.2.4 Exposure to indoor- and outdoor-generated pollution

Section 1.2.3 showed that correlations between pollutants can lead to error, however, the co-emission of pollutants from a single source can be used to opportunistically associate pollution sources and health. Key pollutants can be used as a marker

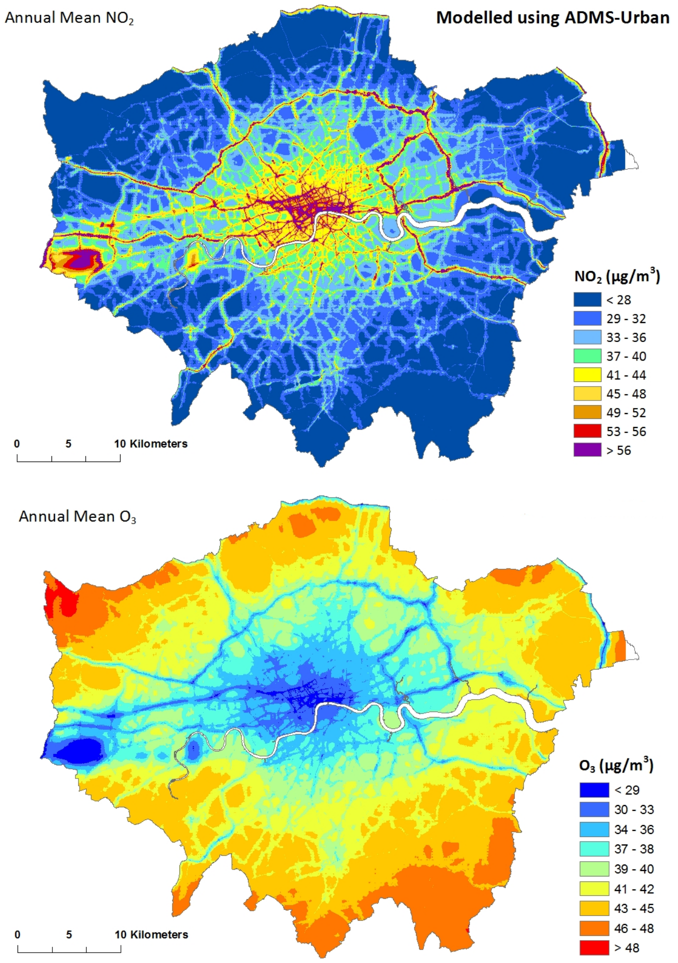

Figure 1.5: Ozone and nitrogen dioxide urban concentrations: Contour plot of London, UK showing the annual average NO_2 and O_3 concentrations predicted by ADMS-Urban model for 2008. This figure has been taken from the ADMS-Urban website: `http://www.cerc.co.uk/environmental-software/ADMS-Urban-model.html`. The NO_2 and O_3 concentrations are anti-correlated.

(also referred to as proxy) of the co-emitted mixture of pollutants from a common source. For example, in the urban ambient outdoor environment, CO is mainly emitted from vehicle exhaust and therefore has been considered a proxy for all traffic emitted pollution[13;11;114].

Conversely, in the indoor environment, a major source of CO is from cooking. CO concentrations have been used as a marker of cooking emissions[48]. Therefore, while the same molecule, indoor-generated CO is a proxy for a different mixture of air than outdoor-generated CO.

Indoor and outdoor pollution are not independent due to exchange through ventilation (the relationship between indoor and outdoor air is explored further in Chapter 2). Therefore, the air in the indoor environment (where people spend most of their time), is made of two "components": outdoor-generated pollution that has been ventilated indoors, and indoor-generated pollution that has been produced by indoor sources. As indoor- and outdoor-generated air mixtures are likely different due to different sources, these two components might be expected to have different health effects. The measured indoor concentrations of proxy species, such as CO, should therefore be apportioned into the two components, which should be considered separately in health models. The results will inform targeted interventions that impact indoor and outdoor sources of pollution[163].

Source-apportionment of $PM_{2.5}$ is of particular interest to health. Differences in the elemental composition of $PM_{2.5}$ between indoor and outdoor environments have been recorded and have been attributed to factors such as the indoor presence of tobacco smoke[77]. Outdoor particles have also been found to contain more crustal elements, such as Si[7]. As the elemental composition of indoor- and outdoor-generated $PM_{2.5}$ are not identical, they are expected to have different toxicities.

Two studies have developed methodologies to source-apportion indoor exposure: Vu et al.[141] source-apportion exposure to NO_2 and $PM_{2.5}$, while Zhang et al.[163] have focused specifically on apportioning exposure to $PM_{2.5}$.

Vu et al.[141] conducted their study in homes in London, UK and found that the contribution of indoor sources to indoor concentrations to both NO_2 and $PM_{2.5}$ in London homes was 26–37%. Contribution of indoor sources to indoor concentrations of both $PM_{2.5}$ and NO_2 was highest around typical cooking times (18:00–19:00).

Zhang et al.[163] apportioned indoor- and outdoor-generated $PM_{2.5}$ measured in homes in Beijing and Pinggu, China. The contributions of indoor-generated $PM_{2.5}$ to indoor concentrations measured in the homes of Pinggu (rural) participants were found to be 19% and 18% during the winter and summer respectively. In Beijing (urban) the contributions in winter and summer were found to be 7% and 6% respectively. Stronger indoor $PM_{2.5}$ sources were observed in the rural cohort. They commented that using portable instruments for indoor monitoring would add flexibility to field campaigns for multiple homes.

Both studies only apportioned indoor air pollution measured in the home environment, and didn't consider any other microenvironments. Apportioned total personal exposure would be more suitable when associating indoor and outdoor sources with health outcomes. Additionally, both studies involved the deployment of stationary air quality sensors both inside and outside of participant homes which is burdensome to researchers.

1.2.5 Low-temporal resolution of pollution measurements

Health studies often use coarse averages of air pollution exposure, normally averaged over 24-hours, or instruments with low temporal resolution, for example only sampling daily, or less often. In 2009, the EPA published a report calling for analysis on the health effects occurring from exposures at sub-daily averaging times for particulate matter (PM). They noted that at the time there was insufficient exposure data for any PM size fraction with health effects to make a causality determination[136]. Since publication of the EPA report, sub-daily exposure studies have been conducted for $PM_{2.5}$ and have reported respiratory and cardiovascular outcomes within 1-hour of exposure to a 10 $\mu g/m^3$ increase in $PM_{2.5}$[160;125].

Figure 1.6 shows short-lived high-pollution events, referred to as emission events, in personal CO exposure, likely due to sources such as cooking or smoking. The black dashed line shows the levels averaged over 24-hour time periods, which would traditionally be associated with health outcomes.

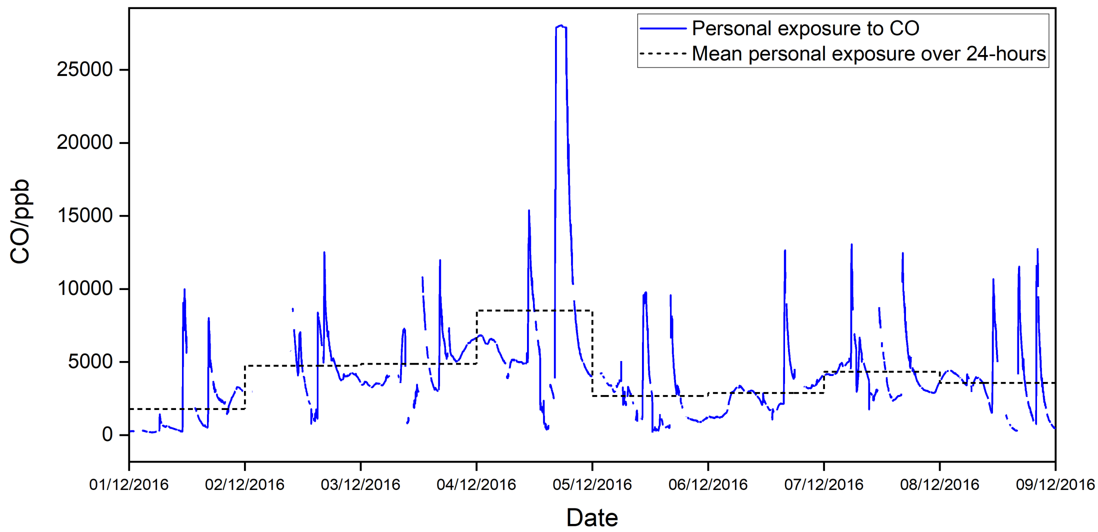

Figure 1.6: Emission events in personal exposure: Time series of personal CO (blue) and the mean personal CO level calculated over 24-hour periods (black dashed). This data was collected as part of the AIRLESS project. Details of the data collection during this project can be found in Chapter 3. Short-lived high-pollution events are observed in the personal exposure data.

Intense exposures of short duration are likely of particular concern, as elevated levels penetrate into the body and target tissue, which may alter metabolism, overload protective or repair mechanisms, and amplify tissue responses[124]. Thresholds, for example the 90^{th} or 95^{th} percentile over a fixed period, have been used to define emission events[68;102;107] and emission events have been characterised by their frequency or magnitude[107]. Associations of these emission event metrics, as well as other sub-daily exposure metrics, with a range of health endpoints, may give insights into the factors driving the observed health effects of air pollution. It has been suggested that different exposure metrics may be better suited to different health endpoints[125].

1.3 Aim and objectives of this PhD thesis

The overarching aim of this work is:

- **To develop a framework to improve understanding of the effects of air pollution on health**

This thesis develops a methodological framework with the following primary objectives:

- **To generate novel exposure metrics from personal air quality monitoring data for use as explanatory variables in pollution-health models**
- **To estimate the effect of the factors that determine exposure to air pollution, specifically home ventilation rates, indoor loss rates and indoor sources**
- **To develop a methodology to link the novel exposure metrics, and exposure determinants, with health markers in future studies**

Chapter 2

A mass-balance model to understand indoor exposure

The work in this thesis develops a framework capable of generating novel exposure metrics and estimating the factors that influence exposure. In developed countries, people spend around 80%-90% of their time indoors[85;21]. Less data are available in lower-middle income countries (LMICs), although they indicate a similar trend, with the majority of time spent indoors[45]. Therefore it is important to understand indoor pollution, and how it relates to more routinely measured outdoor pollution. Figure 2.1 shows indoor and outdoor time series for a participant from the AIRLESS dataset for three pollutants. The AIRLESS dataset was collected in China, and will be described in Chapter 3.

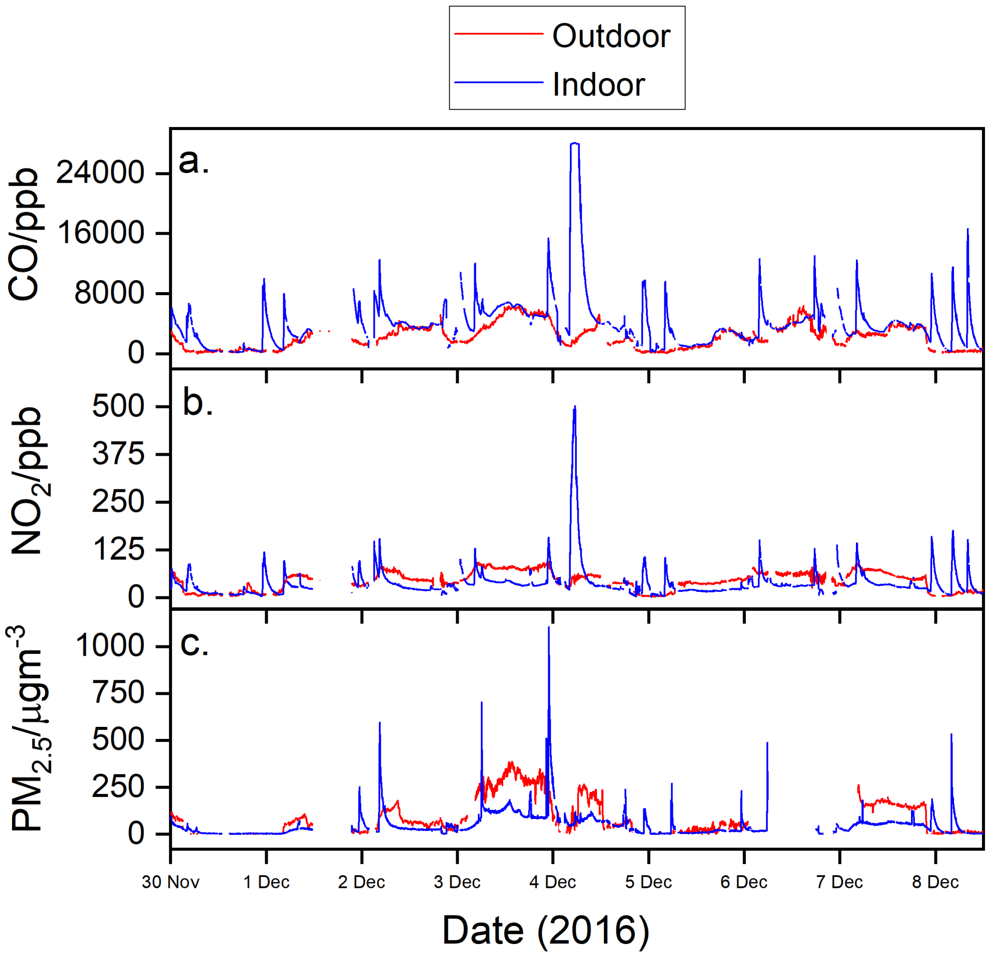

Figure 2.1: Example indoor and outdoor time series: Time series of outdoor (red) and indoor (blue) for three species: a.) CO, b.) NO_2 and c.) $PM_{2.5}$.

This illustrative example highlights some frequently observed features in indoor pollutant time series:

- The indoor data for all pollutants feature spikes of high pollution levels that decay. These will be referred to as indoor emission events
- For CO, excluding the indoor emission events, the indoor data generally follow closely outdoor concentrations, as seen in Figure 2.1a
- For NO_2 and $PM_{2.5}$, excluding the indoor emission events, the indoor data are generally lower than the outdoor concentrations but still appear to follow the overall trend of the outdoor concentrations, as seen most clearly in Figure 2.1b

- All three pollutants are often emitted simultaneously during these indoor emission events, however not always, for example in the early hours of the 1st December, emission events are observed in the CO and NO_2 time series (Figures 2.1a and 2.1b) but not in the $PM_{2.5}$ time series (Figure 2.1c)
- There are inter-pollutant differences in the emission features, for example, the $PM_{2.5}$ decays the fastest to the concentrations before the emission event (Figure 2.1c)

These features in Figure 2.1 can be modelled with a mass-balance equation, often called the continuity equation. The equation can simulate indoor concentrations of a targeted pollutant using outdoor concentrations, ventilation rates, indoor loss rates and indoor sources. These factors are collectively referred to as exposure determinants throughout this thesis. This chapter introduces the continuity equation, followed by a description of the exposure determinants.

Using the solution to the continuity equation as a starting point, the work in this chapter uses simulated case studies to explore how the application of the continuity equation to indoor and outdoor measurements can form the basis of a methodology. This methodology aims to source-apportion indoor exposure into pollution generated from indoor sources and outdoor sources, and aims to estimate the effects of the factors that determine exposure to pollution while indoors.

2.1 The continuity equation

If indoor environments are considered to be a single zone in which the air pollutants' concentrations are assumed homogeneous (uniformly mixed), then a non-steady state mass-balance equation can be used to represent the rate of change in indoor pollutant concentration (Equation 2.1).

$$\frac{dI_t}{dt} = (O_t - I_t)k_{\text{vent}} - I_t k_{\text{sink}} + F_t \tag{2.1}$$

where:

- O_t is the outdoor concentration of pollutant at time t
- I_t is the indoor concentration of pollutant at time t
- k_{vent} is the rate coefficient of indoor zone ventilation

- k_{sink} is the rate coefficient of indoor pollutant removal (indoor losses)
- F_t is the production of the pollutant within the indoor zone, per volume of the zone, at time t

The continuity equation (Equation 2.1) shows that the rate of change in indoor pollutant concentration can be expressed using three terms.

The first term is the rate of change of the indoor concentration of a pollutant due to ventilation.

The second term is the rate of change of the indoor concentration of a pollutant due to indoor loss processes, including chemical reactions and surface deposition. This term is zero for inert species as they do not react in the time frames considered.

The final term is the emission rate which accounts for the production of the pollutant within the zone (indoor sources). The rate of production of the pollutant (mass/time) is divided by the volume of the zone, assuming the air in the zone is uniformly mixed. This results in emission events in indoor time series; see Figure 2.1. For this thesis, peak will refer to the maximum concentration reached during an indoor emission event.

2.2 The major determinants of indoor pollutant concentrations

Equation 2.1 showed that the indoor levels of a pollutant are determined by exposure determinants (outdoor pollutant concentrations, indoor sources, indoor loss processes and ventilation). This section introduces the exposure determinants and explores how values of these exposure determinants have been estimated in previous studies. Later in this thesis (Chapters 4 and 5), exposure determinant values will be estimated for the home microenvironment in China and compared with specific values measured in homes in China from the literature.

2.2.1 Outdoor pollution concentrations

Outdoor pollutants in the lower part of the troposphere enter indoor environments via ventilation. The troposphere is dominated by N_2, O_2 and Ar, with the remaining species constituting less than 1% of the troposphere. Although they are in very small concentrations, these trace constituents play a vital role in tropospheric chemistry. The concentration of pollutants depends on sources chemistry, and other loss processes.

In urban outdoor environments, a principal source of air pollutants are road traffic emissions, and, depending on location and prevailing winds, additional contributions from power plants, industrial boilers, incinerators, ships and agriculture[164].

Figure 2.2 shows the main constituents in diesel and gasoline exhaust.

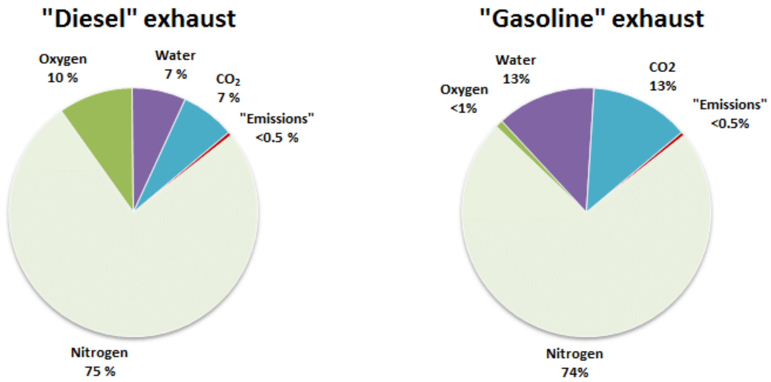

Figure 2.2: Composition of exhaust: Example pie charts of the main constituents in exhaust gas from "diesel" combustion using excess air and stoichiometric spark-ignited "gasoline" combustion. Concentrations are shown as %(V/V). The exact composition of exhaust is fuel- and engine-dependent. "Emissions" are described in the main text. This figure is taken directly from Aakko-Saksa et al.[1]

The "Emissions" sector in this figure includes:

- Carbon monoxide (CO)
- Nitrogen oxides and other nitrogen containing compounds, such as ammonia (NH_3) and nitrous oxide (N_2O)
- Particulate matter consisting of elemental carbon, organic compounds, anions (sulphates, nitrates), and metals
- Hundreds of hydrocarbons, for example benzene and 1,3-butadiene, or greenhouse gases, such as methane
- Carbonyl compounds, such as formaldehyde, acetaldehyde, and acrolein
- Polycyclic aromatic compounds, for example, polyaromatic hydrocarbons (PAHs), nitro-PAHs, and oxy-PAHs

The last three bullet points are categorised as volatile organic compounds (VOCs). Road traffic emissions also contain non-exhaust particle emissions from brakes, tyre wear and re-suspension of road dust[84].

O_3 is a secondary pollutant in the troposphere. The main source of tropospheric O_3 is via photochemical reactions of NO_x, VOCs and CO in the presence of sunlight, however, O_3 can then oxidise NO to regenerate NO_2. There are many complex non-linear photochemical reactions which determine the balance between O_3, NO_x and VOCs, which are well described in literature[41]. Factors such as NO_x (NO_x-limited regime) or VOCs (NO_x-saturated regime) sensitivity, as well as physical processes such as O_3 deposition and stratosphere-troposphere exchange, influence whether O_3 is net produced or net destroyed. The correlation between ambient O_3 and NO_2 has been found to be anti-correlated in some Chinese cities[46;146], and correlated in others[146].

Photolysis is a key chemical removal process in the troposphere. Species that are photolysed include O_3, formaldehyde (CH_2O), methyl iodide (CH_3I), hydrogen peroxide (H_2O_2), NO_2 and nitrate (NO_3). O_3 photolyses to produce $O(^1D)$ which can then react with H_2O[75] as shown in the Equation 2.3 to produce OH radicals:

$$O_3 \xrightarrow{h\nu(290nm \leq \lambda \leq 336nm)} O_2 + O(^1D) \tag{2.2}$$

$$O(^1D) + H_2O \longrightarrow 2\,OH \tag{2.3}$$

OH is a minor product from the photolysis of O_3 - more than 97% of the $O(^1D)$ react back again to O_3 - however OH is extremely reactive, acting as the dominant daytime oxidant, and has the ability to remove the majority of trace gases emitted into the atmosphere. Therefore, OH radicals are nicknamed the “detergents of the atmosphere”.

Tropospheric pollution concentrations are also affected by outdoor physical pollutant removal processes, such as dry deposition (the direct removal of gases and aerosol at the Earth’s surface) and wet deposition (the washout of both vapour phase and particulate-bound chemicals via rain, fog or snow).

2.2.2 Indoor pollution sources

There are many sources of indoor air pollution. These can include:

- Fuel-burning combustion appliances
- Tobacco products
- Building materials and furnishings, including asbestos
- Products for household cleaning and maintenance or personal care
- Central heating and cooling systems and humidification devices

Emissions from cooking constitute a significant source of some key pollutants indoors, specifically CO, NO_2 and particulates. Homes with poorly maintained or poorly ventilated cooking appliances that burn fossil fuels, and homes that rely on the burning of biomass fuels are particularly affected[154]. Cooking fuels have been classified as polluting and clean, as shown in Figure 2.3.

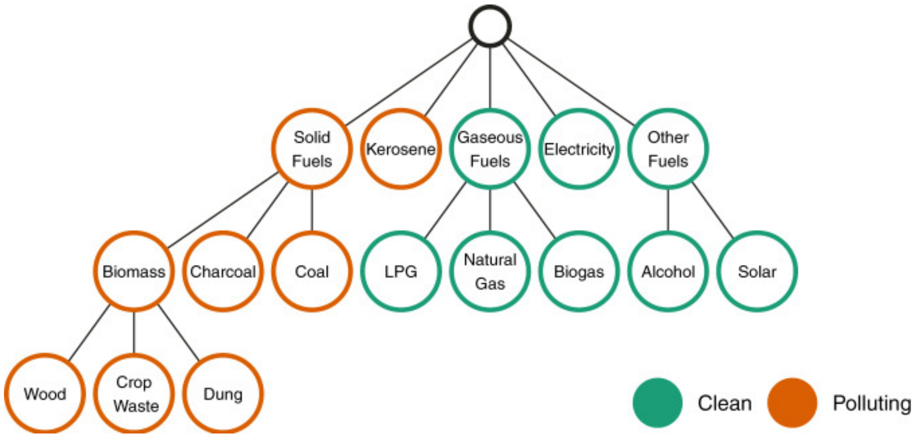

Figure 2.3: Classification of cooking fuels as clean or polluting, taken from Stoner et al.[128]

O_3 can be generated indoors by electronic devices (such as photocopiers and printers), and, ironically, from some air purification devices[61]. Fadeyi[38] proposed that the contribution of ozone air purification equipment with a high ozone emission rate to indoor concentration could easily exceed that of outdoor ozone.

There is increasing evidence that a range of complex products from indoor chemical reactions (of reaction rates fast enough to compete with ventilation rates) may be harmful to health[18]. There is less light indoors, so photolysis reactions are thought to be less important in the indoor environment compared to the outdoor environment. The major gaseous indoor reactions are summarised in Figure 2.4.

Much of the research into indoor chemistry to date has been focused on O_3 as it is highly reactive. Indoor O_3 reacts with unsaturated organic compounds such as terpenes, principally on indoor surfaces. These reaction rates are often fast

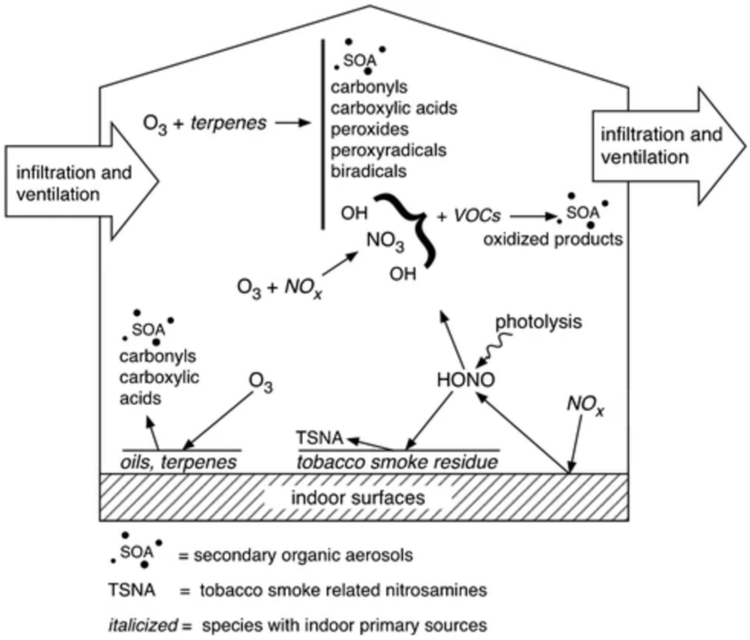

Figure 2.4: Indoor reactions: Major reactants, products, and pathways of indoor chemistry, adapted from Morrison et al.[92]

enough to compete with typical air exchange rates so can influence indoor air chemistry[151]. This produces a large number of products including stable carbonyls such as formaldehyde, short-lived oxidised organic species, free radicals, and even secondary organic aerosols (SOAs)[92]. O_3 also reacts with NO_x forming OH which can initiate oxidation reactions as in the outdoor environment, leading to more oxidative chemistry indoors[18].

As with O_3, NO_2 reacts on indoor surfaces. Surface conversion of NO_2 forms nitrous acid (HONO)[42;88]. HONO has been shown to exhibit short and long term health effects[110;67]. HONO can react with residual nicotine from tobacco smoke sorbed to indoor surfaces, forming carcinogenic tobacco-specific nitrosamines (TSNAs)[123]. HONO can also be photolysed to form OH and NO_3[28]. This is included in the diagram in Figure 2.4.

Particles are produced indoors but are also produced from outdoor sources and ventilate indoors. Particles indoors are impacted by the transport and transformation processes shown in Figure 2.5.

Larger indoor particles can deposit onto surfaces. Deposited particles can resuspend and pollute the indoor air during occupant activities. Suspended particles can also agglomerate forming larger particles via a process called coagulation. This has been observed when burning four gas rings. The emitted particles agglomerate, causing a shift in the peak value of particle size distribution over time[32].

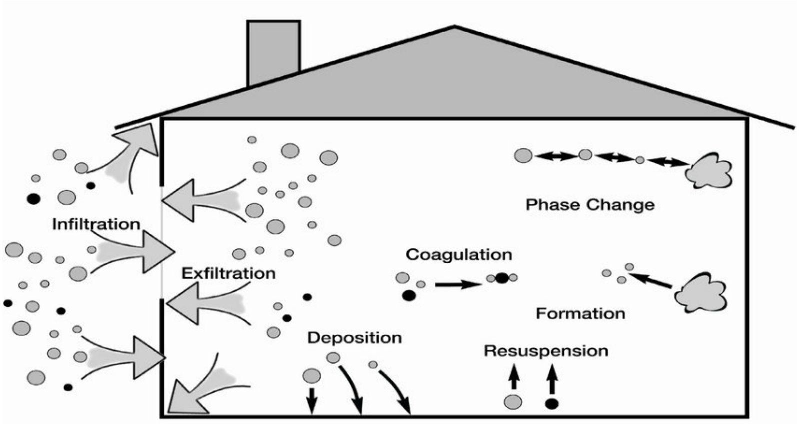

Figure 2.5: Particle processes: Transport and transformation processes which impact the indoor concentration of particulate matter, taken from Thatcher et al. [131]

Phase changes of PM can be observed through the adsorption of organic substances or water [131]. The hygroscopicity[a] of PM varies with the chemical composition and particle size [91].

2.2.3 Indoor pollutant loss processes

Removal processes such as heterogeneous reactions on surfaces and surface deposition in the indoor environment have relatively more impact indoors due to much higher surface area to volume ratios when considering the wall and floor coverings and furniture. The loss rate of a reactive pollutant (normally h^{-1}) can be evaluated as the sum of two terms: the surface-to-volume ratio of the indoor space and the deposition velocity of the species [95]. Experimentally, indoor loss rates have been estimated by installing indoor air quality monitors and either recording the loss rate after artificial elevation of the pollutant [166], or by using the I/O ratio for time periods or environments where there are no indoor sources [166;80].

2.2.4 Ventilation

In naturally ventilated buildings there are two main forces that drive air exchange: wind-driven and buoyancy-driven ventilation. Wind-driven ventilation arises from the different pressures created by wind around the building and the openings and cracks that permit flow through the building. Research has been focused on how these pressure differences vary with building shape, wind direction, and the presence

[a] Hygroscopicity is the tendency of a solid substance to absorb moisture from the surrounding atmosphere

of nearby buildings. Buoyancy-driven ventilation occurs as a result of the temperature difference between the interior and exterior[78]. For example, if the interior is warmer than the exterior, ground-level openings allow colder denser air to enter the building. This air is warmed and rises and escapes the building at higher openings, therefore creating upflow displacement ventilation.

Tracer gas techniques are widely used to measure ventilation rates in buildings. A tracer gas is a substance used to tag volumes of air and so can be used to infer their bulk movement[121]. Ventilation rates in buildings are often measured using tracer gas techniques by measuring the decrease rate of a tracer gas released and dispersed uniformly in the space. The technique requires: (a) the space to be open-plan so airflow occurs only between the specific room and the ambient environment; (b) a perfect mixing of the tracer gas in the space; (c) the driving forces determining the air flow rate remaining constant during the experiment.

There are specific requirements for the tracer gas. It needs to be easily measured, unique or significantly higher than background levels, safe and inert and it must not affect airflow. Commonly, studies have opportunistically used metabolic CO_2 produced from breathing in occupied buildings as a suitable tracer gas because it is inexpensive, easily measured and is only present in comparatively low concentrations in the ambient atmosphere[31;47]. SF_6 is another common choice as it is easy to detect at low concentrations.

Tracer gas methods have recently been extended to public transport to understand the risk of infection of COVID-19 while travelling. This paper can be found in Appendix A.1.1. The effect of COVID-19 on ventilation research is explored in Appendix A.2.1.

2.2.5 Summary of major determinants of indoor pollutant concentrations

Specific concentrations and rates for the exposure determinants are often difficult to obtain experimentally outside of the laboratory due to the high variability between buildings, occupancy behavioural patterns and instrumentation limitations. Additionally, studies involve time-intensive isolated experimental setups. This thesis develops a methodological framework that can estimate these values for hundreds of participants without the requirement of setting up indoor and outdoor stationary pollution monitors.

2.3 Analytical solution to the continuity equation

The continuity equation is a first-order differential equation that can be solved analytically for the indoor concentration of the pollutant at time t, if O and F are considered constants between time=0 and time=t. The analytical solution to the continuity equation (derived in Appendix A.2.2) is shown below.

$$I_t = \frac{Ok_{\text{vent}}}{(k_{\text{vent}}+k_{\text{sink}})} + \frac{F}{(k_{\text{vent}}+k_{\text{sink}})} + (I_0 - \frac{Ok_{\text{vent}}}{(k_{\text{vent}}+k_{\text{sink}})} - \frac{F}{(k_{\text{vent}}+k_{\text{sink}})})e^{-(k_{\text{vent}}+k_{\text{sink}})t} \tag{2.4}$$

Equation 2.4 can be interpreted by considering the steady-state value (I_{SS}). The analytical solution to the continuity equation contains 3 terms. When in steady-state, the last term is zero giving:

$$I_{SS} = \frac{Ok_{\text{vent}}}{(k_{\text{vent}} + k_{\text{sink}})} + \frac{F}{(k_{\text{vent}} + k_{\text{sink}})} \tag{2.5}$$

By substituting I_{SS} in, Equation 2.4 becomes:

$$I_t = I_{SS} + (I_0 - I_{SS})e^{-(k_{\text{vent}}+k_{\text{sink}})t} \tag{2.6}$$

If there was an instantaneous spike of indoor pollution at t=0, I_t would decay to the steady-state value, I_{SS}, with the time constant $k_{\text{vent}} + k_{\text{sink}}$.

2.4 Simulated case studies to conceptualise exposure

This section illustrates indoor concentrations of pollutants considering four different simple scenarios of inert and reactive pollutants in the presence and absence of indoor sources. In this work, inert is used to refer to a species that is not affected by indoor loss processes.

The input values are as follows:

- O is the value of the outdoor concentration within each discrete 1-minute time interval. The outdoor values for each minute are taken from a simple representation of an outdoor time series, created by the addition of two sine functions with a vertical offset. The starting O value is 10 concentration units. In practice, the outdoor time series is assumed to not be a defined function of t.

- k_{vent} is given the constant value of 1 hour^{-1}
- k_{sink} is given the constant value of 2 hour^{-1}
- I_0 for an inert pollutant is given the starting value of 10 concentration units. The start value of I_0 for a reactive pollutant is 3 concentration units.
- F represents indoor sources. Three instantaneous (minute-long) indoor emission events at random points in the time series were generated with varying strength. For all other 1-minute time periods, F is 0.

While it is acknowledged that both ventilation rates (k_{vent}) and loss rate coefficients (k_{sink}) may vary over time in homes, they are set constant in this simulation for simplicity. The axis scales of the simulated graphs have been chosen for consistency across case studies and show a 24-hour period.

2.4.1 Case Study 1: Indoor concentrations of an inert pollutant without indoor sources

Case Study 1 models the indoor concentration, I_t, of an inert pollutant which is not affected by chemical loss processes ($k_{\text{sink}} = 0$) without any indoor sources ($F = 0$). Therefore, Equation 2.4 becomes:

$$I_t = O + (I_0 - O)e^{-k_{\text{vent}}t} \tag{2.7}$$

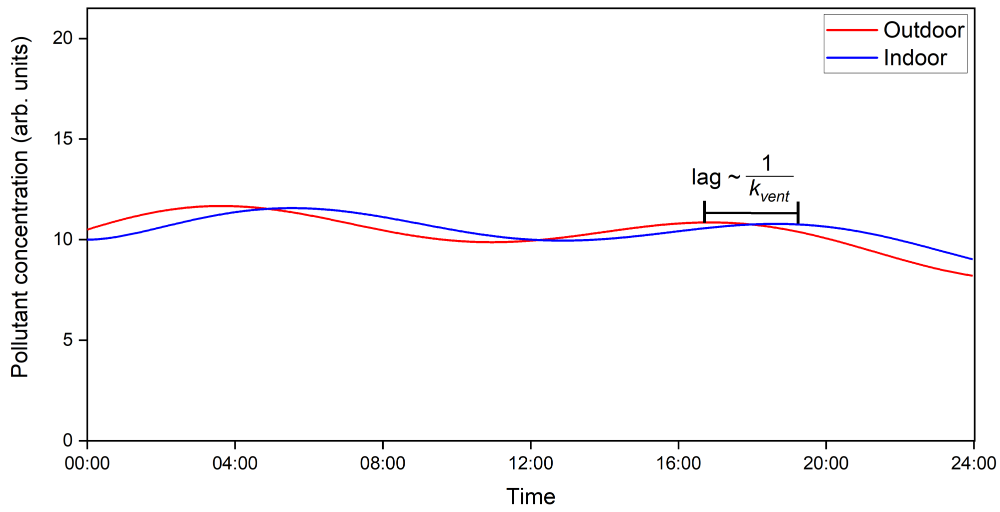

Figure 2.6: Case Study 1: Idealised outdoor time series (red) and generated indoor time series of an inert pollutant (blue) without indoor emission events. The y-axis scale has been chosen for consistency across case studies.

Figure 2.6 shows that indoor levels follow outdoor pollutant levels closely, however, indoor levels are time delayed (lagged) compared with the outdoor levels. This lag can be approximated as the reciprocal of the ventilation rate, as proved mathematically in Appendix A.2.3.

This case study assumes that all pollution measured indoors (blue) was produced by outdoor sources (red) and has ventilated indoors.

2.4.2 Case Study 2: Indoor concentrations of an inert pollutant with indoor sources

Case Study 2 extends Case Study 1 by adding three 1-minute sources (F) at irregular intervals of varying magnitude. Equation 2.4 becomes:

$$I_t = O + \frac{F}{k_{\text{vent}}} + (I_0 - O - \frac{F}{k_{\text{vent}}})e^{-k_{\text{vent}}t} \tag{2.8}$$

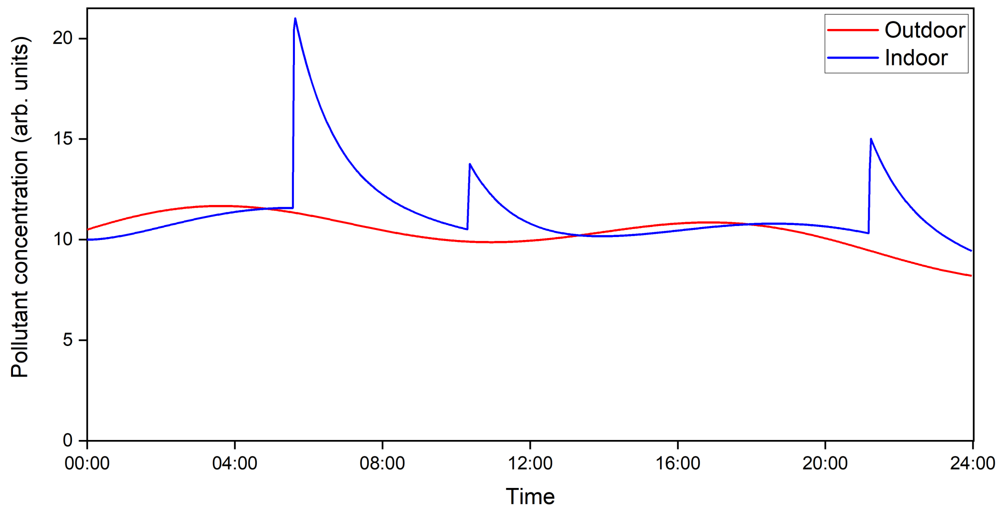

Figure 2.7: Case Study 2: Idealised outdoor time series (red) and generated indoor time series (blue) of an inert pollutant with indoor sources

Figure 2.7 shows that the indoor levels of an inert gas follow the outdoor levels, as in Figure 2.6, however, in this case study, the inert species has instantaneous indoor sources which elevate the indoor concentration. The pollutant level then decays at a rate that is influenced by the ventilation of the indoor space. This will be shown in Figures 2.11 and 2.12.

For this case study, the indoor level of the pollutant is an accumulation of pollution that was generated by outdoor sources that ventilated indoors ($I_{t(outgen)}$), and pollution that was generated by indoor sources ($I_{t(ingen)}$). These two components are shown in Figure 2.8.

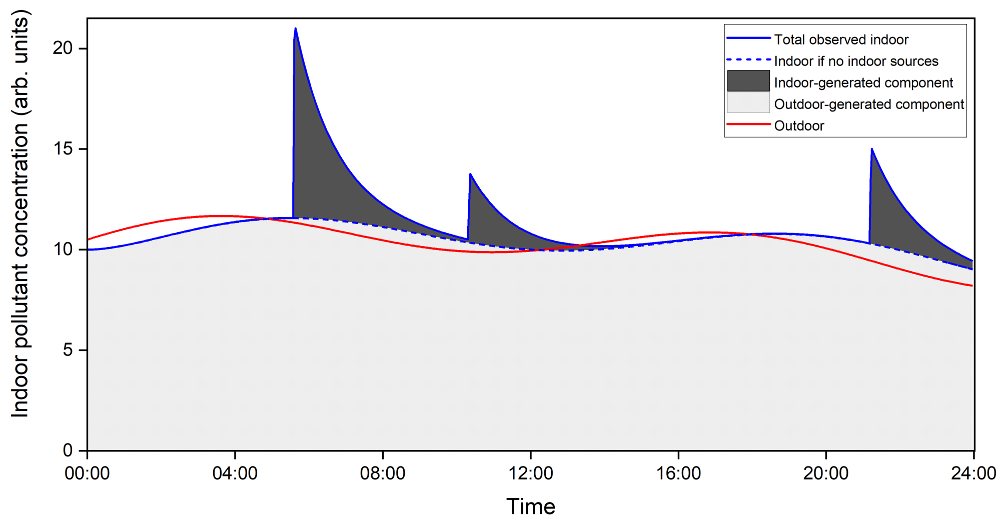

Figure 2.8: Indoor- and outdoor-generated components of total observed indoor (inert pollutant): Time series showing the indoor concentration of a pollutant (I_t) and its components: the pollution generated outdoors ($I_{t(outgen)}$) (light grey) and the pollution generated indoors ($I_{t(ingen)}$) (dark grey). The solid blue line represents the total observed indoor concentration of an inert pollutant. The dashed blue line represents the indoor concentration of an inert pollutant with no indoor sources.

The equations for the two components are as follows:

$$I_{t(ingen)} = \frac{F}{k_{vent}} + (I_{0(ingen)} - \frac{F}{k_{vent}})e^{-k_{vent}t} \tag{2.9}$$

$$I_{t(outgen)} = O + (I_{0(outgen)} - O)e^{-k_{vent}t} \tag{2.10}$$

The components sum to give Equation 2.8.

2.4.3 Case Study 3: Indoor concentrations of a reactive pollutant without indoor sources

Case Study 3 models a reactive pollutant ($k_{\text{sink}} > 0$) without sources ($F = 0$). Equation 2.4 becomes:

$$I_t = \frac{Ok_{\text{vent}}}{(k_{\text{vent}} + k_{\text{sink}})} + (I_0 - \frac{Ok_{\text{vent}}}{(k_{\text{vent}} + k_{\text{sink}})})e^{-(k_{\text{vent}}+k_{\text{sink}})t} \tag{2.11}$$

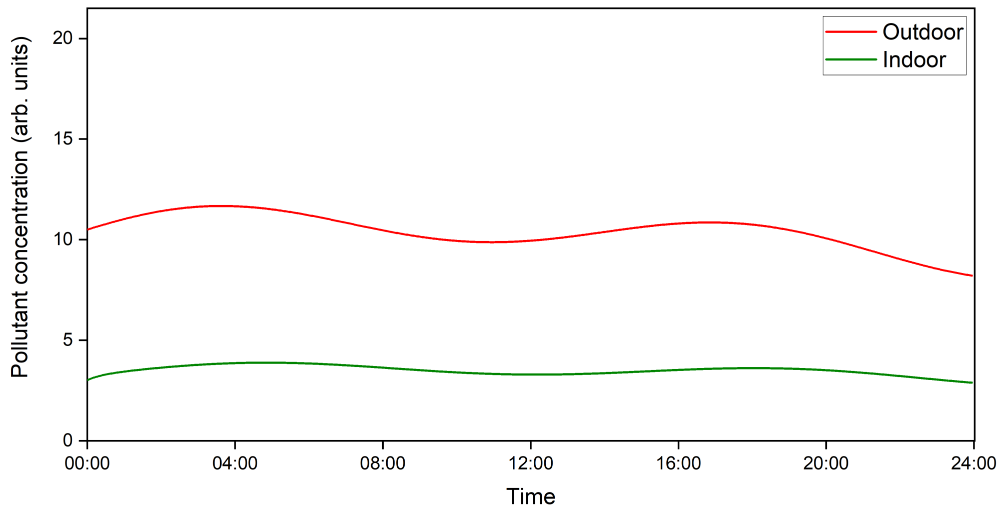

Figure 2.9: Case Study 3: Idealised outdoor time series (red) and generated indoor time series of a reactive pollutant without indoor sources (green)

The indoor level of a reactive pollutant is lower than the outdoor level due to indoor chemical losses primarily from surface reactions or deposition (for larger particles).

Case Study 3 considers a reactive species with no indoor sources and so it can be assumed that all pollution measured indoors was produced by outdoor sources and has been ventilated indoors. I_t can therefore be renamed $I_{t(outgen)}$:

$$I_{t(outgen)} = \frac{O_t k_{\text{vent}}}{(k_{\text{vent}} + k_{\text{sink}})} + (I_{0(outgen)} - \frac{O_t k_{\text{vent}}}{(k_{\text{vent}} + k_{\text{sink}})})e^{-(k_{\text{vent}}+k_{\text{sink}})t} \tag{2.12}$$

When considering long time periods where $t \to \infty$, the last term in Equation 2.12 tends to 0. The equation becomes:

$$I_{(outgen)} = \frac{O_t k_{\text{vent}}}{(k_{\text{vent}} + k_{\text{sink}})} \tag{2.13}$$

2.4.4 Case Study 4: Indoor concentrations of a reactive pollutant with indoor sources

Finally, in Case Study 4, the same three indoor sources are introduced for the reactive pollutant. The equation that describes this case study is the full analytical solution to the continuity equation (Equation 2.4):

$$I_t = \frac{Ok_{\text{vent}}}{(k_{\text{vent}}+k_{\text{sink}})} + \frac{F}{(k_{\text{vent}}+k_{\text{sink}})} + (I_0 - \frac{Ok_{\text{vent}}}{(k_{\text{vent}}+k_{\text{sink}})} - \frac{F}{(k_{\text{vent}}+k_{\text{sink}})})e^{-(k_{\text{vent}}+k_{\text{sink}})t} \tag{2.4}$$

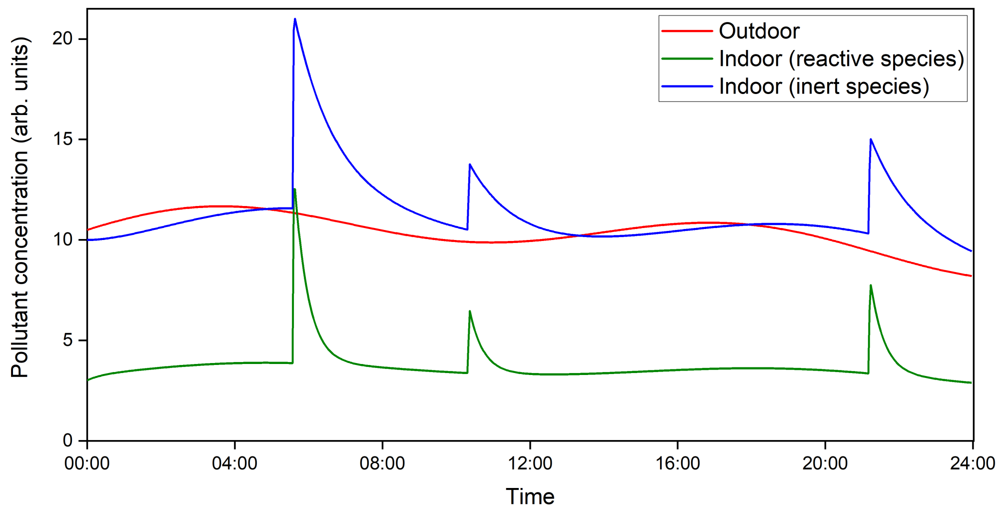

Figure 2.10: Case Study 4: Idealised outdoor time series (red) and generated indoor time series of a reactive pollutant with indoor sources (green). Case Study 2 is included (blue) (indoor time series of an inert pollutant) as a visual comparison between a reactive and inert pollutant.

The indoor concentration of the reactive species is lower than outdoor levels (except during strong indoor emission events), and decays faster than that of an inert species (the green decays after a peak are steeper than the blue in Figure 2.10). This is because the total decay rate for the reactive species is the cumulative influence of ventilation rates and loss rate mechanisms $k_{\text{vent}} + k_{\text{sink}}$, as shown in Figure 2.11.

To visualise the difference in the decay rates of inert and reactive pollutants, the natural logarithms of the decays are plotted in Figure 2.12. The gradient of the line is $-k_{\text{vent}}$ for the inert species and $-(k_{\text{vent}} + k_{\text{sink}})$ for the reactive species. For this case study, k_{vent} is 1 hour^{-1} and k_{sink} is 2 hour^{-1}.

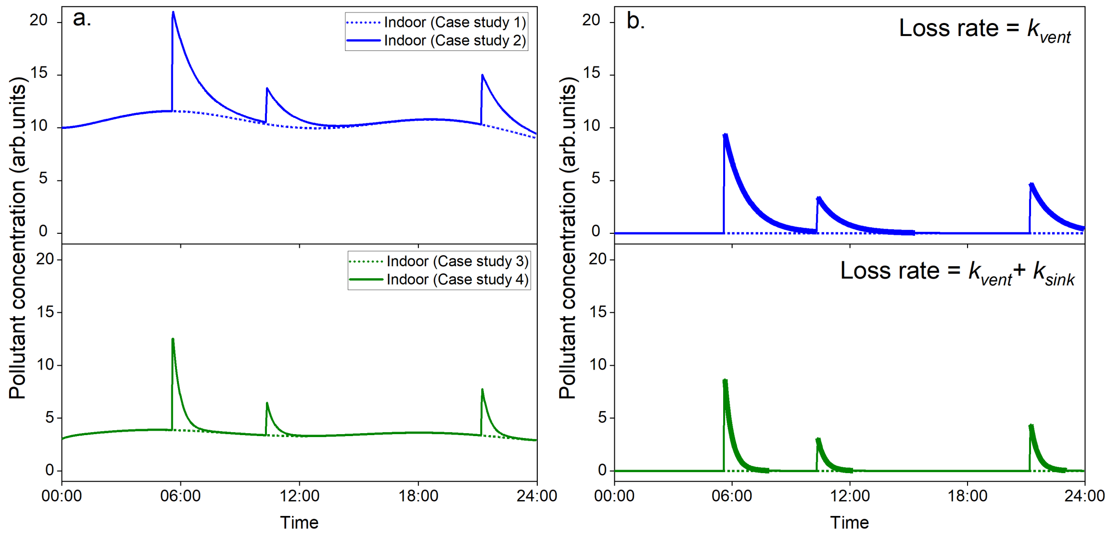

Figure 2.11: Illustration of how indoor emission events can be isolated for an inert and reactive species: a.) Indoor concentrations of pollutants for all four case studies. The inert case studies are shown on the top plot and the reactive case studies are shown on the bottom plot. b.) Time series from case studies without indoor sources have been subtracted from time series with sources. This leaves just the indoor-generated source events. Thick lines have been used to identify the decaying regions of the emission events.

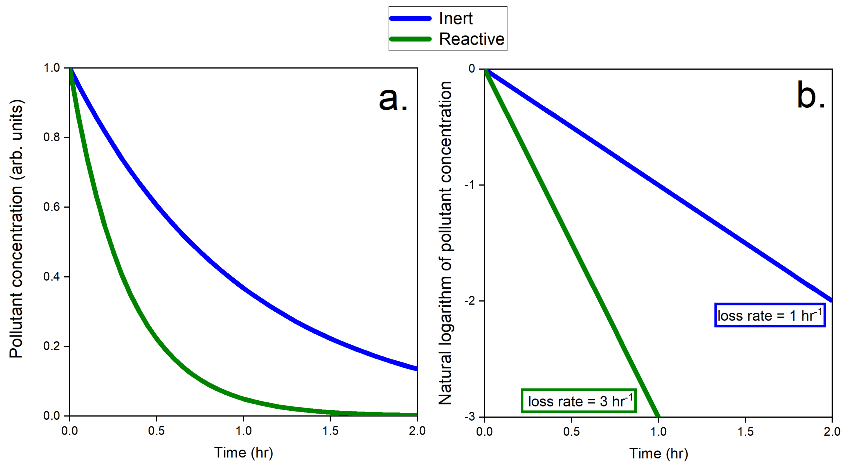

Figure 2.12: Comparison of the decay rates of indoor emission events for inert and reactive species: As in Figure 2.11, blue represents an inert pollutant and green represents a reactive pollutant. a.) Graph of y=e^{-cx} decays where c is the decay rate of the exponential curves in Figure 2.11b and x takes values between 0-2 hours. This is to allow for easy visual comparison of the decay rates. b.) The natural logarithm of the data Figure 2.12a. The gradient of the lines equals -($k_{\text{vent}} + k_{\text{sink}}$). $k_{\text{sink}} = 0$ for the inert pollutant.

For this case study, the indoor level of the reactive pollutant is an accumulation of pollution that was generated by outdoor sources and ventilated indoors ($I_{t(outgen)}$), and pollution that was generated by indoor sources ($I_{t(ingen)}$), as shown in Figure 2.13.

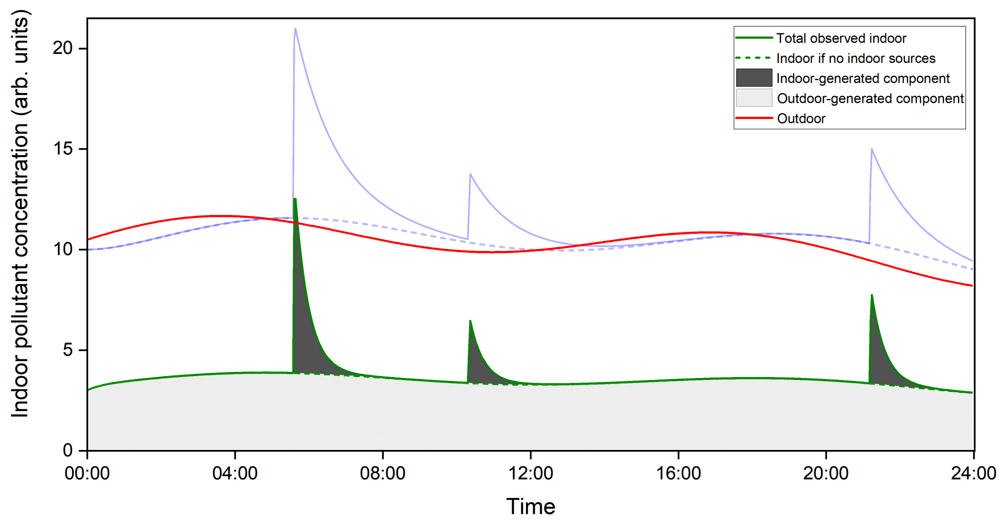

Figure 2.13: Indoor- and outdoor-generated components of total observed indoor (reactive pollutant): Time series showing the indoor concentration of a reactive pollutant (I_t) and its components: the pollution that was generated outdoors ($I_{t(outgen)}$) (light grey) and the pollution generated indoors ($I_{t(ingen)}$) (dark grey). The solid green line represents the total observed indoor concentration of an inert pollutant. The dashed green line represents the indoor concentration of an inert pollutant with no indoor sources. The equivalent components are shown in faint blue lines for an inert species.

The equations for the two components are as follows:

$$I_{t(ingen)} = \frac{F}{k_{vent} + k_{sink}} + (I_{0(ingen)} - \frac{F}{k_{vent} + k_{sink}})e^{-(k_{vent}+k_{sink})t} \tag{2.14}$$

$$I_{t(outgen)} = \frac{Ok_{vent}}{k_{vent} + k_{sink}} + (I_{0(outgen)} - \frac{Ok_{vent}}{k_{vent} + k_{sink}})e^{-(k_{vent}+k_{sink})t} \tag{2.15}$$

2.5 Methods to estimate exposure metrics and exposure determinants from the application of the continuity equation to indoor and outdoor data

2.5.1 The ventilation rate of an indoor space

Time-lag between indoor and outdoor levels to estimate k_{vent} (Lag method): In a scenario where there are no indoor sources of an inert species (Case Study 1), the ventilation rate of an indoor space can be estimated as the reciprocal of the lag between the indoor and outdoor concentrations of the inert pollutant. This is proved mathematically in Appendix A.2.3.

Constant decay method to estimate k_{vent}: In a scenario where there are indoor sources of an inert species (Case Study 2), the ventilation rate of the indoor space can be estimated from the decay after the indoor emission event (the constant decay method). This method assumes that the indoor sources are instantaneous events with no additional indoor sources of pollution during the decay, and that the outdoor level of the pollutant is constant during the decay. Although technically this method only requires the data points at the beginning and end of the decay, using all data points in the decay can improve the precision of the k_{vent} value.

2.5.2 The indoor loss rate of a reactive species

Indoor/outdoor (I/O) ratio to estimate k_{sink}: In a scenario where there are no indoor sources of a reactive species (Case Study 3), k_{sink} can be estimated from the I/O ratio of the reactive pollutant. If k_{vent} and k_{sink} are considered constants over a time period, $I_{t(outgen)}$ is just a scaled version of O_t.

$$\frac{I_{t(outgen)}}{O_t} = \frac{k_{\text{vent}}}{(k_{\text{vent}} + k_{\text{sink}})} \tag{2.16}$$

If k_{vent} is known, k_{sink} can be estimated from the I/O ratio of a reactive species in the absence of sources of the reactive species.

Constant decay method to extract k_{sink}: In a scenario where there are indoor sources of a reactive pollutant (Case Study 4), the indoor loss rate of that pollutant can be estimated by fitting exponential curves to decaying regions of the indoor emission events (the constant decay method) in the indoor data of a reactive species. As the exponential decay rate after an indoor reactive pollution spike is the cumulative effect of ventilation and the indoor loss processes, a known ventilation rate can be applied to estimate k_{sink}. This method assumes that the indoor sources are instantaneous events with no additional indoor sources of pollution during the decay, and that the outdoor-generated component of indoor air of the pollutant is constant during the decay.

2.5.3 Concentrations of indoor- and outdoor-generated components of indoor air

Indoor- and outdoor-generated components of indoor pollution for inert species: For an inert species, the ventilation rate can be applied to outdoor data to generate the outdoor-generated component of indoor air. This time series is then subtracted from indoor data to isolate the indoor-generated component of indoor air. The components were shown in Figure 2.8.

Indoor- and outdoor-generated components of indoor pollution for reactive species: The ventilation rate and indoor loss rate can be applied to outdoor data to estimate the outdoor-generated component of indoor air. This time series is then subtracted from indoor data to isolate the indoor-generated component of indoor air. The components were shown in Figure 2.13.

2.5.4 Characteristics of indoor emission events

Strong indoor emission events (sources) in the indoor-generated component of indoor air are of particular interest with respect to health (in this example, "strong indoor emission events" are defined as over the 90^{th} percentile of the indoor-generated component). Figure 2.14 shows different characteristics of indoor emission events which may act as metrics in health models.

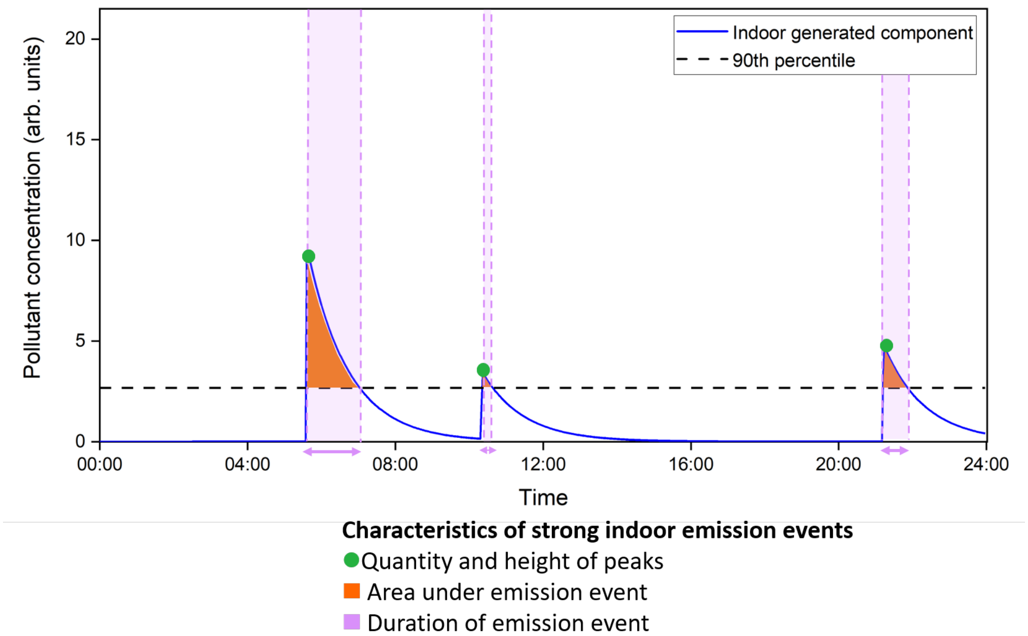

Figure 2.14: Characterising indoor emission events: A time series of the indoor-generated component of indoor air of an inert species is shown in blue. The 90^{th} percentile of the indoor-generated component is shown with a black dashed line. Quantifiable characteristics of emission events are indicated.

2.6 Chapter summary

This chapter introduced the continuity equation, and included a description of the exposure determinants. It provided simulated case studies to explore how the application of the continuity equation to indoor and outdoor measurements can form the basis of a methodology to source-apportion indoor exposure into pollution generated from indoor sources and outdoor sources.

Chapter 3

AIRLESS dataset

This chapter presents the data and previous work performed as part of the **"AIR pollution on cardiopuLmonary disEaSe in urban and peri-urban reSidents in Beijing"** (AIRLESS) project, which took place in China. The methodological framework outlined in Chapter 2 has been applied to the data from the AIRLESS project.

3.1 Air quality in China

The majority of air pollution epidemiology studies used cohorts living in Europe and North America. The work in this thesis develops a methodological framework using a dataset from China. It is important that epidemiological studies are conducted in China as inconsistencies between pollution-health associations made in China compared with developed countries may arise. They may be a result of overlooking differences in industrial activities, behavioural and socioeconomic factors, as well as topography. For example, size distributions of outdoor airborne particles have been shown to vary by country[103] and it is likely that their composition may be different too. Furthermore, studies have found that genetics can influence the health response to particulate matter[158;115;113].

China has experienced rapid industrialisation, urbanisation, and transportation development[83] which has led to outdoor pollution levels that are consistently well above the upper limits indicated by the WHO guidelines[83;6]. Figure 3.1 shows the differing particulate matter levels in select cities around the world.

The framework developed in this thesis was only applied to data collected in China,

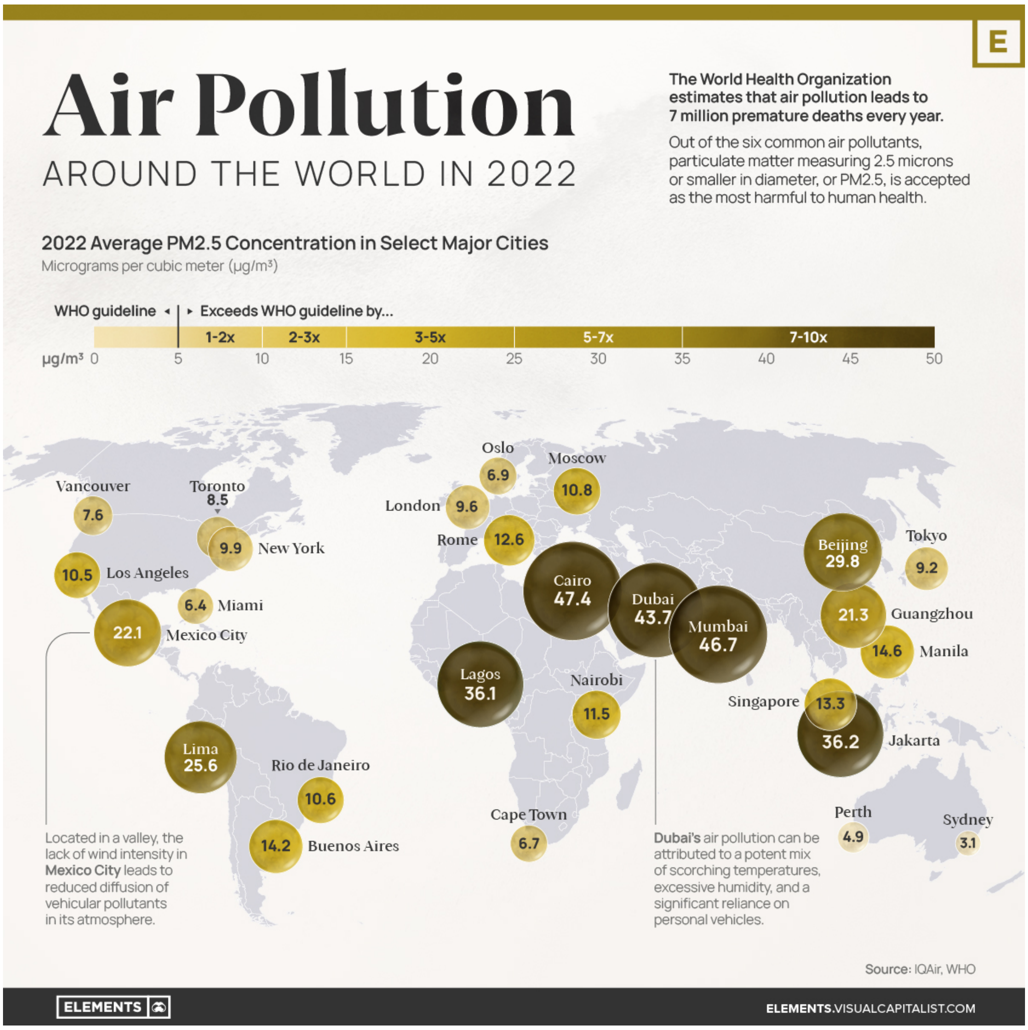

Figure 3.1: Air pollution around the world: This figure is a visualisation of data from the 2022 World Air Quality Report [43], created by VC Elements [140]. The $PM_{2.5}$ levels in selected cities are compared to the WHO ambient guidelines [155]. The majority of air pollution epidemiology studies used cohorts living in Europe and North America. The selected cities in these continents have lower levels of $PM_{2.5}$ than the Chinese cities.

which restricts the interpretation of results to the context of China. Nevertheless, the framework has been designed to be applicable to studies with similar designs in different countries.

3.2 AIRLESS study overview

In 2016, over 150 UK and Chinese scientists joined forces to understand the causes and impacts of air pollution in Beijing. The research programme"Atmospheric Pollution and Human Health in a Chinese Megacity" (APHH) had the ultimate aim of informing air pollution solutions and thus improving public health. AIRLESS was an APHH project between King's College London, Imperial College London, the University of Cambridge and Peking University[65]. The aim of the project was to investigate the associations between exposure to multiple air pollutants and changes in health outcomes with a focus on cardiopulmonary biomarkers in urban and rural residents in China[52]. Members of the Jones group focused on developing and deploying a lower-cost portable sensor platform to measure the concentrations of pollutants that participants were exposed to. In addition to the personal exposure measurements, comprehensive outdoor reference measurements were collected as part of the project.

The full protocol for the AIRLESS study can be found at `https://acp.copernicus.org/articles/20/15775/2020/`.

3.3 Participant overview

The study was organised as a panel study[a], repeated in two seasons, winter (Nov 2016- Jan 2017) and summer (May- June 2017). The participants were recruited from two locations. The participants of the urban cohort lived close to the Peking University (PKU) Hospital, Beijing. The rural cohort lived in the Pinggu district (formerly Pinggu county), an area around 80km east of Beijing[52]. In total, 251 individuals participated in this study:

- Winter Beijing: 123 participants
- Winter Pinggu: 128 participants
- Summer Beijing: 102 participants
- Summer Pinggu: 116 participants

The following information was collected through a baseline questionnaire after the participants were enrolled[51]:

- Demographic information (e.g. gender, age, education, income)
- Current and past domestic energy use patterns (e.g. types of fuels and stoves, frequency of cooking and heating stove use)

[a] A panel study is a type of longitudinal research where data is collected from the same individuals repeatedly over a period of time and changes are detected

- Active and second-hand smoking history (only current non-smokers were recruited)
- Dietary habits (e.g. consumption of alcohol, coffee/tea, sugar beverage drinking, fried food, vegetables)
- Sleep quality
- Daily activity patterns (transportation, exercise, and potential exposure sources)
- Major health conditions, events, and diagnoses of non-cardiovascular outcomes since the original enrolment
- Regular medication or supplement usage
- Characteristics of their homes (window and door features, floor of building where they reside)

Figures 3.2 and 3.3 show selected participant statistics for the AIRLESS cohort, collected using this questionnaire.

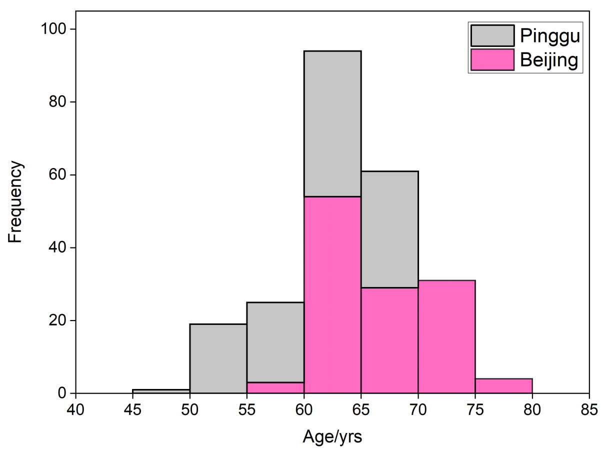

Figure 3.2: Age distribution: Stacked histogram of the ages of the participants in the AIRLESS cohort, split by residence site, N=251. The mean age for the Beijing cohort was 65.7 and the mean age for the Pinggu cohort was 60.7.

The Beijing cohort were mainly retired (108/123), with a few participants working in education (11/123)(principally employees of Peking University), whereas the Pinggu cohort mainly worked in the agricultural sector (76/128) and in housekeeping (16/128). The Beijing cohort had a similar number of male and female participants, whereas the Pinggu cohort had a higher ratio of females (77/123). Gas was found to

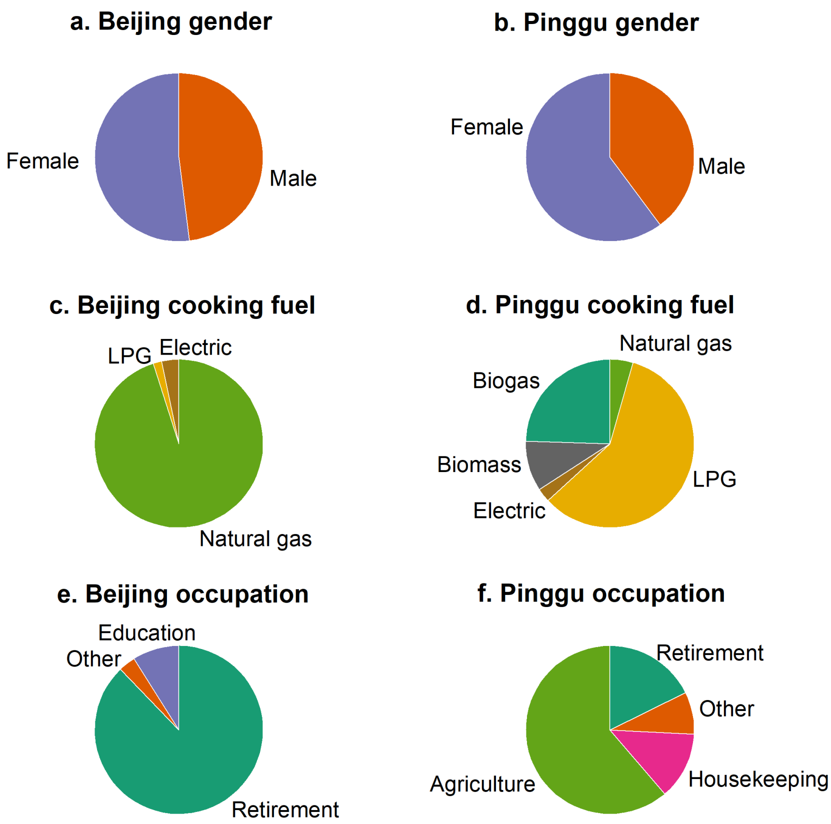

Figure 3.3: Participant statistics of the AIRLESS cohorts: a.) and b.) Pie charts of gender breakdown for Beijing and Pinggu respectively. c.) and d.) Pie charts of cooking fuel type for Beijing and Pinggu respectively. e.) and f.) Pie charts of occupations for Beijing and Pinggu respectively. Data is for 251 participants, N=123 in Beijing (left) and N=128 in Pinggu (right). LPG = Liquefied Petroleum Gas. All information displayed was collected using a questionnaire.

be the most common cooking fuel in both cohorts (natural gas for the Beijing cohort and LPG for the Pinggu cohort); however, there was also a significant fraction of the Pinggu cohort who relied on biogas (28/123) or biomass (11/123) burning as their primary cooking fuel.

3.4 Outdoor reference instruments

Two air quality monitoring stations recorded outdoor air pollution levels in proximity to most participants' residential addresses: one for the urban cohort (Beijing) and one for the rural cohort (Pinggu). These stations will be referred to as the "reference instruments". The Pinggu residents all live within 5 km of the reference instrument. Most of the Beijing residents live within 5 km of the reference instrument, as shown in Figure 3.4.

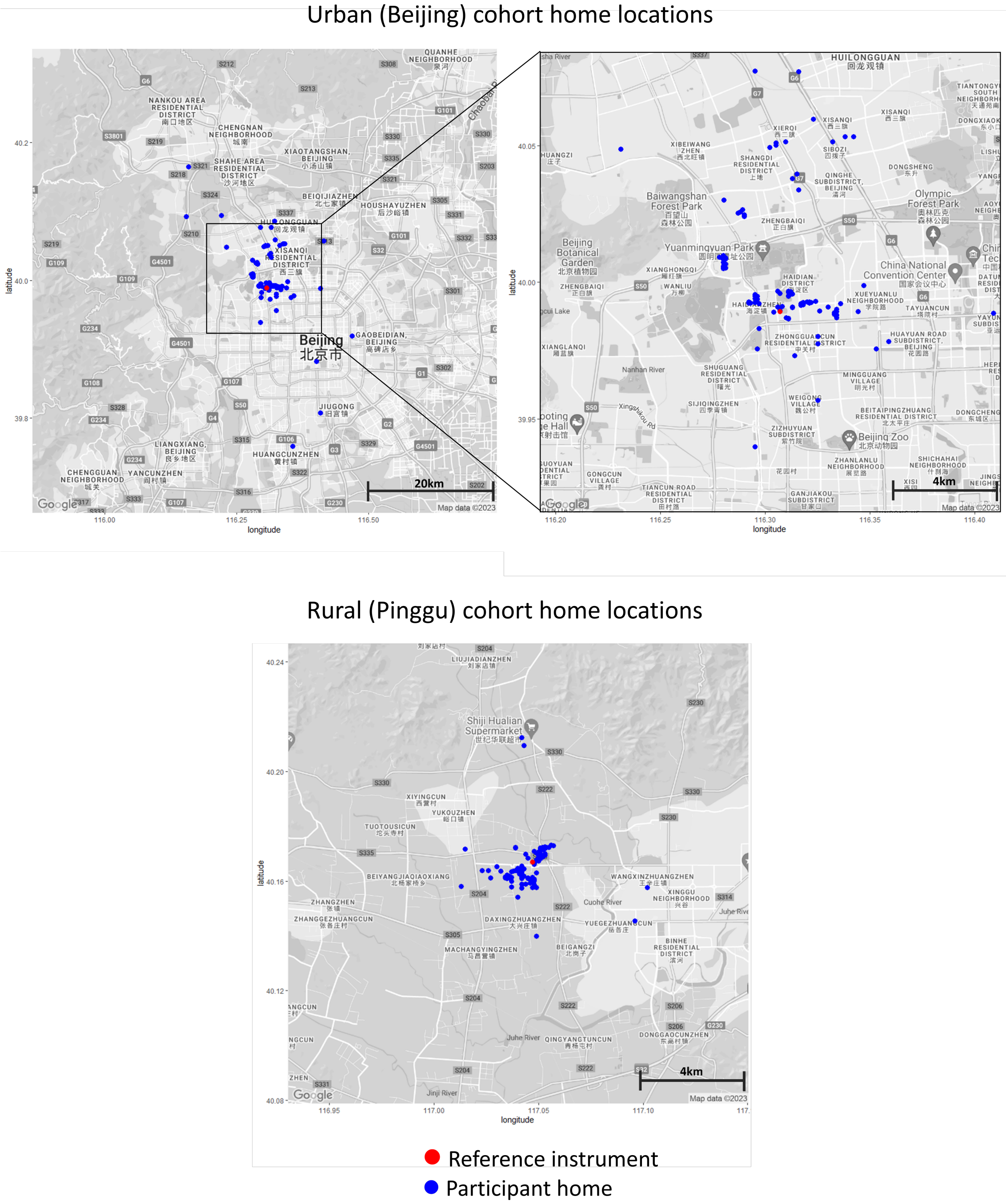

Figure 3.4: Home and reference instrument locations: Maps showing the locations of the participants' homes (blue) and the locations of the reference instruments (red). The top left map shows the homes of all Beijing participants. The top right is zoomed in to show that the proximity of the majority of the Beijing participants' homes to the reference instrument. The bottom map shows the homes of the Pinggu cohort. Maps have been generated from Google Maps using an API key.

The Beijing reference instrument was located on the PKU campus in the Haidian district of Beijing. It was on the roof of a six-storey building and was away from direct emission sources (limited vehicles allowed on the campus). The Pinggu reference instrument was deployed on the roof of a one-storey building. The data collected at the reference instruments will be referred to as "outdoor data" for the purposes of this report. Measurements from these reference instruments in Beijing and Pinggu have been shown to be a suitable alternative to measurements directly outside peoples' homes[163] when performing source-apportionment. The sampling interval of the reference instruments was 1 minute. The measurements from these reference instruments are typical of the data previously used to derive personal human exposure metrics.

Appendix A.3.1 contains the specific technologies used in the reference instruments.

Collection rate of reference air quality measurements
The percentage of minute-resolution outdoor data recorded during the AIRLESS campaign is shown in Table 3.1.

Percentage of minute-resolution reference data recorded during the AIRLESS Campaign				
Pollutant	Winter (7th Nov- 21st Dec 2016)		Summer (22nd May- 21st June 2017)	
	Beijing	Pinggu	Beijing	Pinggu
CO	85.90	85.26	84.72	96.40
NO	95.82	94.91	94.49	95.50
NO_2	99.94	94.85	94.49	95.50
O_3	91.98	94.43	94.56	73.97
$PM_{2.5}$	95.83	91.07	94.87	1.66 (99.68*)

Table 3.1: Recorded reference data: A table containing the percentage of minute-resolution reference data during the AIRLESS field campaign for the five key species. *This value represents the percentage of reference data after interpolation between hour-resolution data.

The reference instrument in Pinggu only measured $PM_{2.5}$ at hour-resolution (as opposed to minute-resolution) during the summer campaign, resulting in a low percentage of recorded data. In this thesis, the $PM_{2.5}$ data collected in Pinggu in the summer has been estimated between the hourly measurements, for each minute, using linear interpolation.

3.5 Personal exposure measurements

The assessment of the pollution exposure that individuals receive requires a portable device, capable of accurately measuring pollutant concentrations.

The Personal Air Monitor (PAM) was used in the AIRLESS study to measure the personal exposure of the participants. Each participant was given a PAM for one continuous week in the winter and one continuous week in the summer. The device was developed at the Department of Chemistry, University of Cambridge, in collaboration with Atmospheric Sensors Ltd[22]. The PAM is an autonomous sensor platform of multiple sensors for physical and chemical parameters. The combined cost of the sensors alone is less than £600 and the total cost of the PAM is less than £2000, making it a "lower-cost" system[29] . The sampling interval of the gaseous pollutants and particulate matter for the AIRLESS project was set to 1 minute. At this time resolution, a single charge lasts for 20 hours. The data is stored on an SD card inside the monitor and uploaded through a General Packet Radio Service (GPRS) to a secure access FTP server.

The air pollution sensors in the PAM are miniaturised, allowing the platform to be highly portable (around 400g), and small enough to be worn by a participant during daily life. Electrochemical sensors (ECs) are used to measure gaseous pollutant concentrations and an Optical Particle Counter (OPC) measures Particulate Matter (PM) concentrations. A table describing the specific electrochemical sensors and OPC can be found in Appendix A.3.1. Technical descriptions of the operating principles of the EC sensors and OPC are found in Appendix A.3.2. The PAM collects temperature and relative humidity (RH) measurements to assess the thermal environment of the participant, as well as auxiliary parameters, such as noise, acceleration and Geo-coordinated Position System (GPS) data which can be used for activity assessment.

As the PAM was developed to capture personal exposure of participants, it is important that it can accurately measure pollutant levels. RMSE values summarise the mean difference between measurements from the PAM sensors and certified instruments from a co-location period. Indoor and in-transit co-location deployments in the UK, China, Germany, and Kenya across seasons have been organised to evaluate the performance of the PAM under different conditions[73]. The PAM parameters are shown in Table 3.2.

It has been demonstrated that the performance of the PAM's components can be comparable with the performance of reference instrumentation across a wide range of conditions, including in indoor environments, diverse outdoor environments and in static and non-static deployments[22].

PAM calibration and validation

To calibrate the PAM for the AIRLESS project, four co-location periods of the PAM and reference instrument took place in Beijing for 19 days before and after the winter

Parameter	Method	RMSE
Spatial coordinates	Global positioning system (GPS)	
Background noise	Microphone	
Physical activity	Triaxial accelerometer	
Temperature	Band-gap IC	
Relative humidity	Capacitive	
Particulate matter	Optical particle counter (OPC)	9 μg/m^3
CO	Electrochemical sensors (EC)	31.6 ppb
NO		3 ppb
NO_2		3 ppb
O_3		2.7 ppb

Table 3.2: PAM parameters: A table summarising the parameters measured by the PAM, including the method used and the error associated (for pollutant parameters)[22]. These RMSE values were calculated during a deployment in the UK. Due to unavailable measurements, the PM measurements in the UK could not be corrected for RH effects, which resulted in only a moderate correlation with the reference instrument[22;73]. A table containing the names of the specific electrochemical sensors and OPC can be found in Appendix A.3.1.

and summer campaigns of the AIRLESS project. The calibration parameters were extracted from these periods with similar environmental conditions and in the same geographical area in which the monitors had been or were to be deployed.

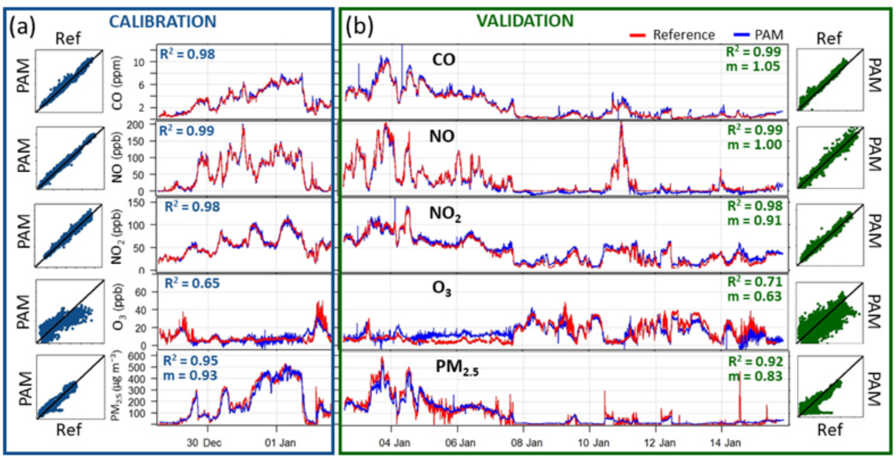

Figure 3.5: PAM calibration: The time series of the pollutants measured by the PAM (blue) closely follow the reference instruments (red) in both the calibration (Figure 3.5a) and validation (Figure 3.5b) periods during the outdoor co-location after the winter campaign. These figures have been taken from Chatzidiakou et al.[22].

The time series of the pollutants measured by the PAM closely follow the reference instruments in both a calibration (Figure 3.5a) and validation (Figure 3.5b) period.

Collection rate of personal air quality measurements

The percentages of minute-resolution data collected by the participants during the campaign are shown in Table 3.3.

Percentage of minute-resolution PAM data recorded by the AIRLESS cohort				
Pollutant	Winter		Summer	
	Beijing	Pinggu	Beijing	Pinggu
CO	96.32	95.05	97.77	95.89
NO	93.98	98.70	88.45	94.89
NO_2	80.26	73.51	84.27	71.83
O_3	80.87	81.70	82.71	84.43
$PM_{2.5}$	97.98	92.64	94.39	86.92

Table 3.3: Recorded PAM data: A table containing the percentage of minute-resolution PAM pollutant data recorded by participants during the AIRLESS field campaign for the 5 key species. This is after post-processing and calibration.

3.6 Determination of the microenvironment of the participant

An automated time-activity model is used in this thesis to classify the main microenvironments visited by the participants[20] while carrying the PAM. The model computes the space-time utilisation distributions of the GPS coordinates for each participant and classifies the microenvironment and activity using metrics such as time spent in each location, re-visitation rate and metrics of directional movement (Figure 3.6). The detected microenvironments are: home, work and transit. A further step was taken to split the "work" category into "work indoors" and "work outdoors", as described in Appendix A.3.3.

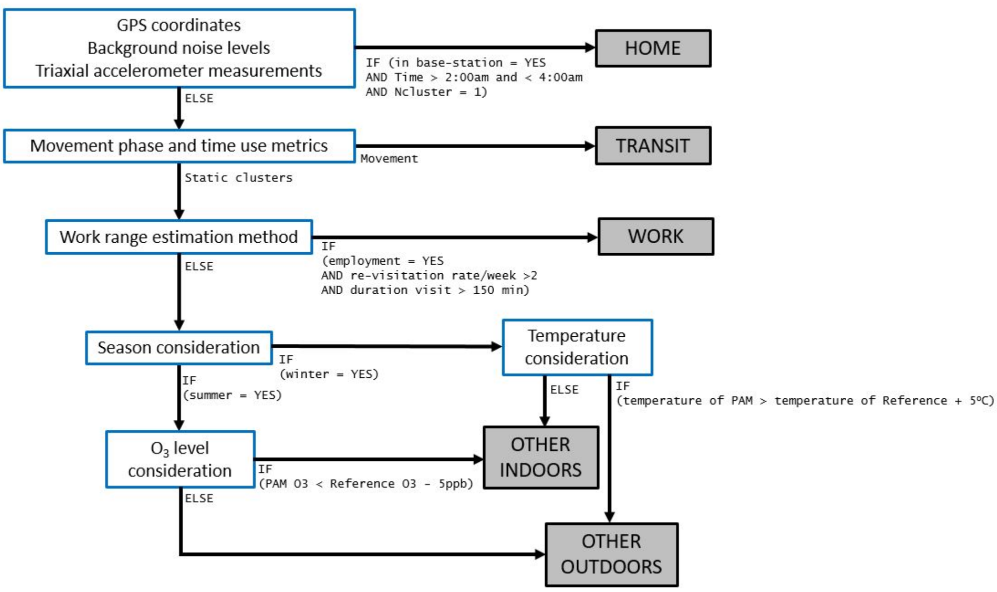

Figure 3.6: Time-activity model: A flowchart showing the steps involved in determining the microenvironment of a participant in the time-activity model.

Results from this time activity model, and a recent extension to this model (not applied in this work) have been evaluated against manual time-activity logs kept by participants[20;21].

The time-activity breakdown for the AIRLESS cohort is shown in Figure 3.7 and shows that the participants across both locations and seasons spend the majority of their time at home.

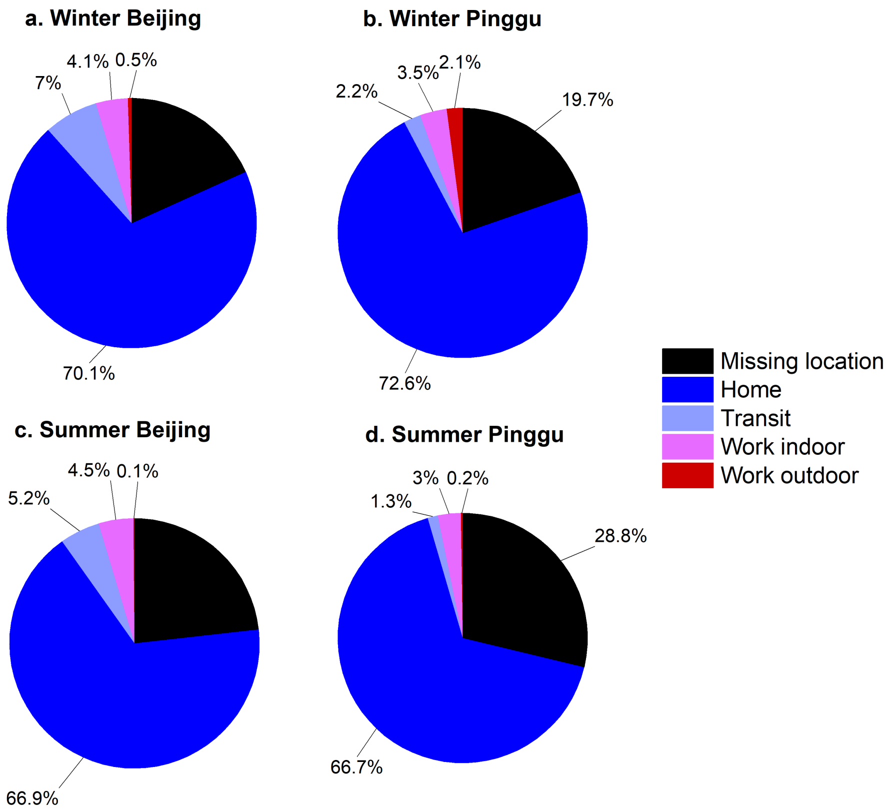

Figure 3.7: Results from the time-activity model: Pie charts displaying the proportion of time that the participants spent in different microenvironments for both seasons and locations. Participants across all four categories appear to spend most of their time at home.

There are missing location data points in the Pinggu cohort (Figure 3.7). Missing time-activity assignments were found to be more prevalent during haze events. Haze events in China have been shown to affect the zenith tropospheric delay (ZTD), which refers to the amount of time it takes for radio waves to pass through the Earth's atmosphere from the GPS satellite to the GPS receiver on the ground. High atmospheric particle concentrations, such as during haze events, can lead to scattering and absorption of GPS signals, leading to an increase in ZTD, resulting in error and missing values[150].

3.7 Health data collection

Biomarkers were collected for each participant on days 0, 3 and 7 between 8:30am-9am, except for Peak Expiratory Flow (PEF) measurements, which were measured every morning by each participant and written down in a diary. A full list of the health measurements is shown in Table 3.4.

Biological pathways	Sample/device	Health endpoints
Blood pressure and heart rate	Omron HEM 907	Systolic pressure, diastolic pressure, heart rate
Endothelial function	Pulse wave analyser	AP/AIx/ED/SEVR
Respiratory inflammation	Peak flow meter	PEF
	Exhaled breath	FE_{NO}
	EBC	pH
		Cytokines, e.g. IL-1α, IL-1β, IL-2, IL-6, IL-8, TNFα, IFN-γ
Cardiovascular inflammation	Serum	CRP
		Cytokines, e.g. IL-1α, IL-1β, IL-2, IL-6, IL-8, TNFα, IFN-γ
	Plasma	WBCs, neutrophils, monocytes, lymphocytes
Metabolic	Serum	TG, HDL, LDL, cholesterol
		Glucose, insulin, HOMA-IR
	Serum, urine	Untargeted/targeted metabolomic signatures
Oxidative stress	Urine	MDA, creatinine
	Plasma	DNA repair enzymes
Genetic-related pathways	Blood	Genetic and epigenomic profiles

Table 3.4: Health measurements: A table containing the measurement plans for health markers in the AIRLESS study. This table has been taken directly from Han et al.[51]

Although many health markers were collected during the AIRLESS project, Article 28 of the 2021 Chinese Personal Information Protection Law states that personal health information is considered sensitive data. At the time of writing this thesis, only the PEF measurements were therefore available for analysis and presentation.

3.8 Chapter summary

In summary, an automated time-activity model has been applied to the data recorded by high-performance personal air monitors (PAMs). This has captured the daily personal exposure of hundreds of participants in China, in a range of microenvironments for five pollutants, as part of the AIRLESS project. Additionally, health and demographic information have been collected from these participants.

Chapter 4

Application of the methodological framework to a large scale fieldwork study: an example participant

This chapter demonstrates the specific computational steps involved in applying the methods outlined at the end of Chapter 2 to the data recorded in the field with low-cost portable sensor platforms, described in Chapter 3. The methodology is applied to the data from a single participant from the AIRLESS dataset as a demonstration.

4.1 Assumptions of the methodological framework

In applying the continuity equation to air pollution data, assumptions are being made (presented in Section 2.1). It is assumed that air pollutant concentrations are homogeneous (uniformly mixed) within single zones. Below is an assessment as to whether assumptions are expected to be upheld or broken when applying the methodology to the AIRLESS dataset.

Firstly, the homes of the AIRLESS participants are unlikely to be single homogeneous indoor zones as they are made of different rooms, connected by doors that can be open to varying degrees. Several studies of ventilation rates have been undertaken in thousands of households[49;54;63]. These studies employ a single-zone approach when conducting the tracer gas method. However, one study by Van Ryswyk et al.[139] estimated the bias with the single-zone assumption in measurement of residential air exchange using the tracer gas method, compared with a “two-zone” set-up. They concluded that the assumption of a single well-mixed air zone very likely results in an under prediction of the ventilation rate of around 16%, compared to a two-zone set-up. This Van Ryswyk study was conducted in Canadian homes with the two zones on different floors of the houses (often one of the zones was in the basement). It is likely that there are less differences between different rooms (zones) in the homes of the AIRLESS cohort as the size of homes in China are smaller than in Canada (on average 646 square footage compared with 1948 squared footage[156]), and many of the AIRLESS population live in single-story flats, instead of in houses of multiple floors.

Secondly, within rooms, air pollution sources are often localised. The AIRLESS study did not involve investigations into the mixing or spatial distribution of pollutants. In this work, CO is used as the tracer gas due to the strong indoor sources as most AIRLESS households rely on LPG, natural gas and bio-gas for domestic energy use (see Figure 3.3), however cooking stoves are a localised source in the kitchen, which may not uphold the assumption of a homogeneous single zone. Standard protocols developed for measuring pollution for household energy projects call for kitchen concentrations to be measured one horizontal meter from the centre of stove and at a height of 1.5m[112]. One study[70] evaluated how representative the CO concentrations at 1.5m are of the kitchen’s average CO concentration. This study was performed across 5 kitchens in India, and for a total of 70 cooking events. The kitchen CO concentration was measured by eight CO monitors and, for each monitor, was weighted by the relative volume of kitchen air represented by that monitor. The median ratio of the eight weighted concentrations to the 1.5m concentration was 0.95, suggesting that the 1.5m location is representative of overall CO concentration in the kitchen. Although CO measurements at a height of 1.5m may be a good estimate of the average CO concentration in the kitchen, it has been speculated that spatial distribution of CO is sensitive to stove type, for example, a charcoal stove’s plume likely has greater upward convection whereas the kerosene stove’s plume may be more likely to be mixed[70]. On the other hand, CO has a slow removal rate from the lower atmosphere (the mean residence time of CO in the lower atmosphere has been estimated to be between 0.3 years and 5 years[111]) and no known indoor loss processes, it can be considered inert in these timescales, strengthening the assumption of homogeneous mixing of CO over longer time periods.

Further studies into the spatial distribution of air pollutants in homes and specific rooms within homes is limited. It is common for the assumption of homogeneity to be made in tracer gas and modelling studies[49;54;63;70]. Some studies use a fan to achieve homogeneity[66].

4.2 Source-apportionment of personal exposure of an inert species, including the estimation of ventilation using the tracer gas method

4.2.1 Estimation of ventilation using the constant decay method

As mentioned in the previous section, in this work, CO is used as the tracer gas.

Two possible methods to estimate k_{vent} were outlined in Section 2.4. The constant decay method has been selected as the preferred method of estimating k_{vent} as there were strong CO peaks identified in the data. Additionally, the lag is hard to detect in the data as the variability in the background levels is on the same time scale as the lag.

CO data from an urban participant is used to demonstrate the individual steps when estimating the ventilation rate.

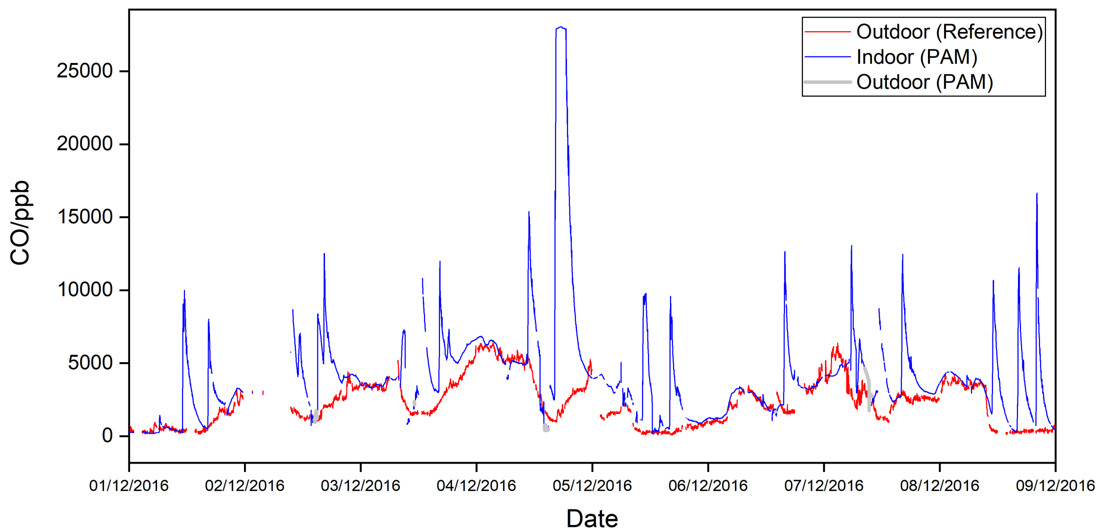

Figure 4.1: Real data to illustrate time series features of an inert species: A time series of outdoor PAM data (grey), indoor PAM data (blue) and outdoor reference data (red). PAM data was recorded by Participant U143.

The inert property of CO can be seen in the time series of this example participant; the outdoor reference closely follows the data recorded indoors by the PAM, with the exception of the indoor emission events which elevate the indoor CO concentration to above the outdoor concentration.

Automated code has been developed that can estimate ventilation rates retrospectively from the dataset. Figure 4.2 is a flowchart of the constant decay method, which is followed by specific computational steps.

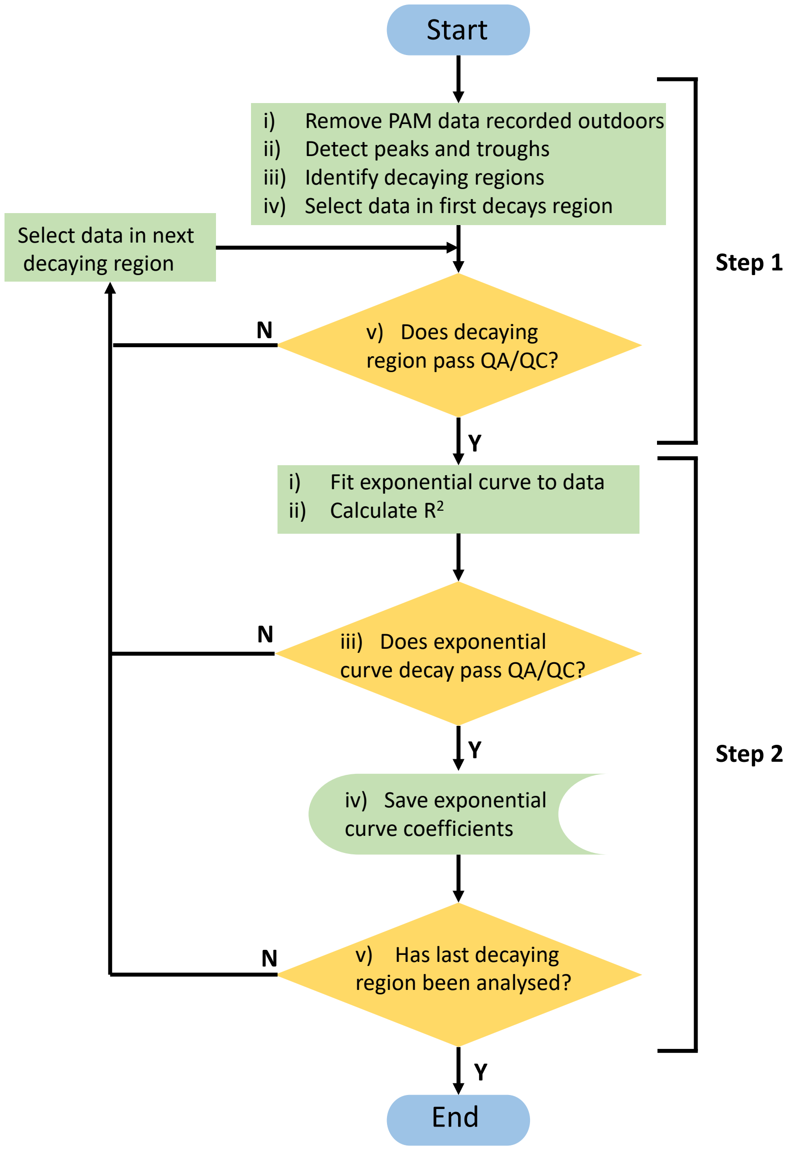

Figure 4.2: Constant decay method steps: Flowchart of the automated code to identify and characterise the exponential decays of indoor emission events in indoor time series. QA/QC = quality assurance/quality control.

Step 1: Identifying decaying regions

1i: The time-activity model, when applied to PAM data, allows for the identification of periods when the participant was in an indoor environment. Only these data are retained. If the ventilation of a singular microenvironment is of interest, the specific indoor microenvironment can be selected at this stage. For this example, participant data recorded in all identified indoor environments are retained.[a]

1ii: detects peaks and troughs in the indoor (PAM) time series with an automated algorithm. Peaks were identified as the rolling maximum within 30 data points and troughs were identified as the rolling minimum within 15 data points. These detected peaks and troughs for the example participant are shown in Figure 4.3.

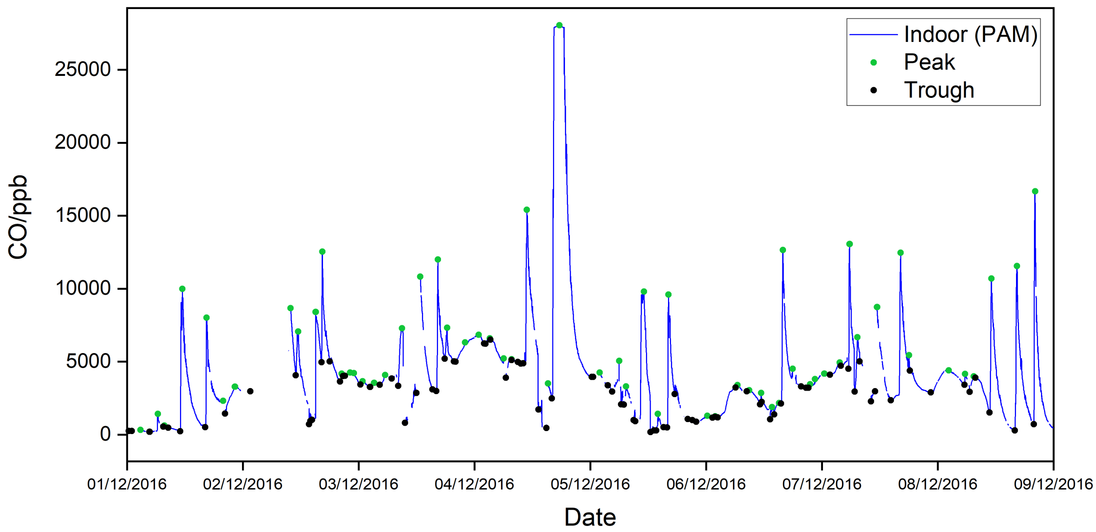

Figure 4.3: Peaks and troughs in indoor CO data: A time series of the indoor (PAM) CO (blue) recorded by Participant U143. Detected peaks (green) and troughs (black) are flagged.

1iii: identifies decaying regions as periods between a peak and the following trough.

1iv: selects the data points during the first identified decaying region.

1v: applies a QA/QC approach to the decaying region. The number of data points in the decaying region must be over 5 and the vertical range of the decay must be larger than the (85th percentile)-(15th percentile) range of the time series. This aims to ensure that no small decays (due to the noise of the data) are analysed.

Additionally, this algorithm assumes a constant O value over the whole decay (see Step 2i). The drawback of this is that it does not take into account changes in

[a] For this dataset, "home", "work indoor", "transit" were assumed to be indoor environments.

outdoor pollution that would affect the rate of the decay. A previous study that used CO_2 as a tracer gas set a limit of 40ppm variation in outdoor CO_2 during the decay[40]. If the outdoor CO_2 varied more than this during the decay, the decay was rejected. Similar considerations were made in this work. This QA/QC step additionally rejects decaying regions where the variation in outdoor CO levels (during and for the previous hour) was more than 20% of the vertical range of the decay.

All decaying regions that passed Step 1 are indicated in Figure 4.4 by green shading. Appendix A.4.1 provides a demonstrative case (a 12-hour period), explaining why certain decays after peaks pass this QA/QC step, and why other do not.

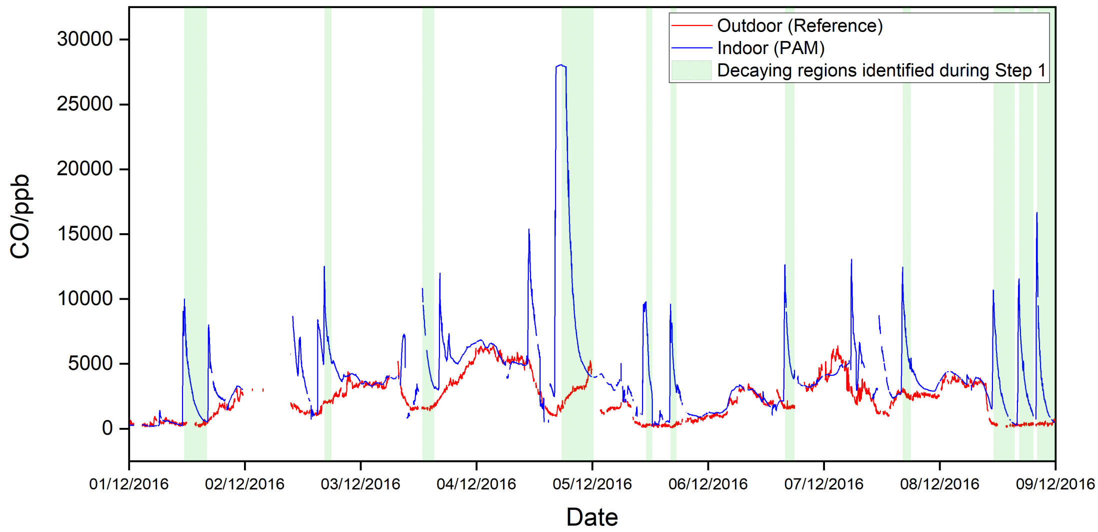

Figure 4.4: Identified decaying regions detected during Step 1: A time series of the indoor (PAM) CO (blue) recorded by Participant U143, the outdoor (reference) data (red). The decaying regions that passed through Step 1 are indicated by green shading.

Step 2: Fitting exponential curves

2i: assumes that the indoor (PAM) data points within the decaying region are decaying exponentially and fits an exponential curve by the general formula:

$$y = a + be^{-cx} \tag{4.1}$$

Where:

- y is I_t
- a is O, which is assumed to be the mean of the outdoor time series during the

decay*

- b is I_0 (the peak height)
- c is k_{vent}
- x is t

giving Equation 2.7.

*The mean outdoor values are calculated using reference data recorded during the decay and during the hour before the start of the decay. These mean values are plotted for all decaying regions that were accepted through Step 1 in Figure 4.5.

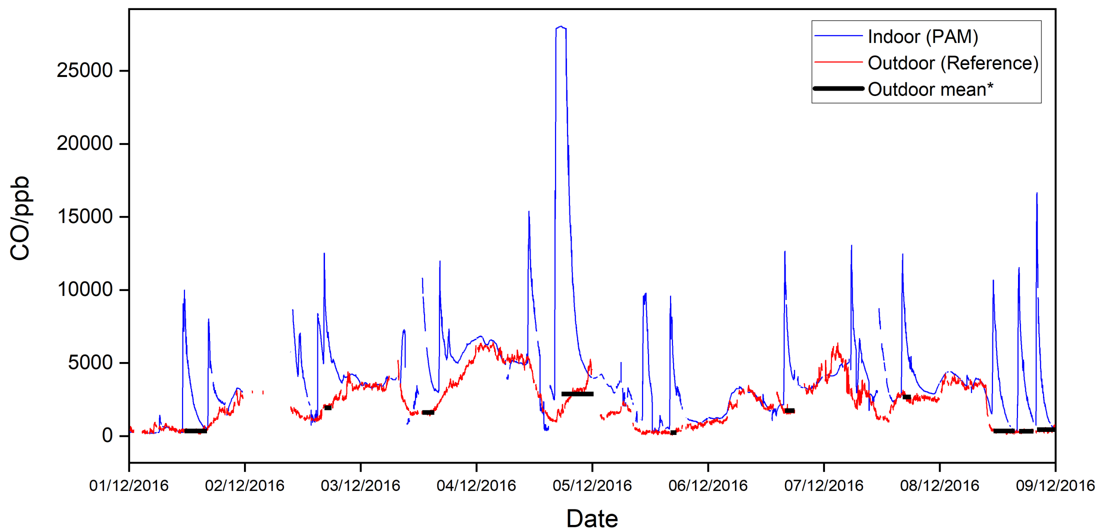

Figure 4.5: Mean outdoor levels during decaying regions: A time series of the indoor (PAM) CO (blue) recorded by Participant U143, the outdoor (reference) data (red) and the mean* outdoor values (black). *The mean outdoor values are calculated during the decay and during the hour before the start of the decay and are plotted for the decaying regions.

Two possible fitting methods have been considered to estimate the exponential coefficients from the data. The first method fits an exponential curve to the raw data using a least squares fit. The second method performs a log transformation of the data, and then linear regression is used to estimate the coefficients. These two methods produce slightly different coefficients, but the second method is selected as a weighted fit towards smaller y values is not desirable. This is explored further in Appendix A.4.2.

Figure 4.6 shows the fit of an exponential curve to the first decaying region in the example participants' data.

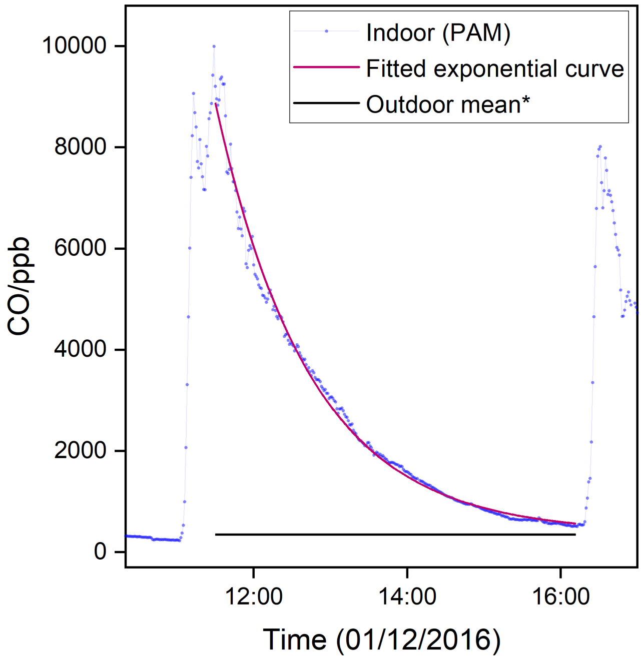

Figure 4.6: Fit of exponential decay to data within a single decaying region: Indoor CO data measured by the PAM (blue) from 10:20 to 17:00 on 01/12/2016. An exponential curve has been fitted to the data between the peak and trough. *The mean outdoor value is calculated during the decay and during the hour before the start of the decay and is plotted for the decaying region.

2ii: To assess whether the decay can be described by an exponential fit, the natural logarithm of the decay was compared against the natural logarithm of a modelled exponential curve using linear regression. The goodness-of-fit of the logged exponential curves was assessed with the r^2 value of the fit.

2iii: applies a QA/QC approach to the exponential curve. The estimated ventilation rates (c) were retained only if they were positive (so that e^{-cx} decays) and the r^2 was larger than 0.75 indicating that the log of the decay is almost linear, and that there are no additional strong sources of the pollutant during the decay (it is assumed that the indoor source is instantaneous (minute-long) as mentioned in Section 2.4). The resultant curves are plotted in Figure 4.7, and the decay rates are compared in Figure 4.8.

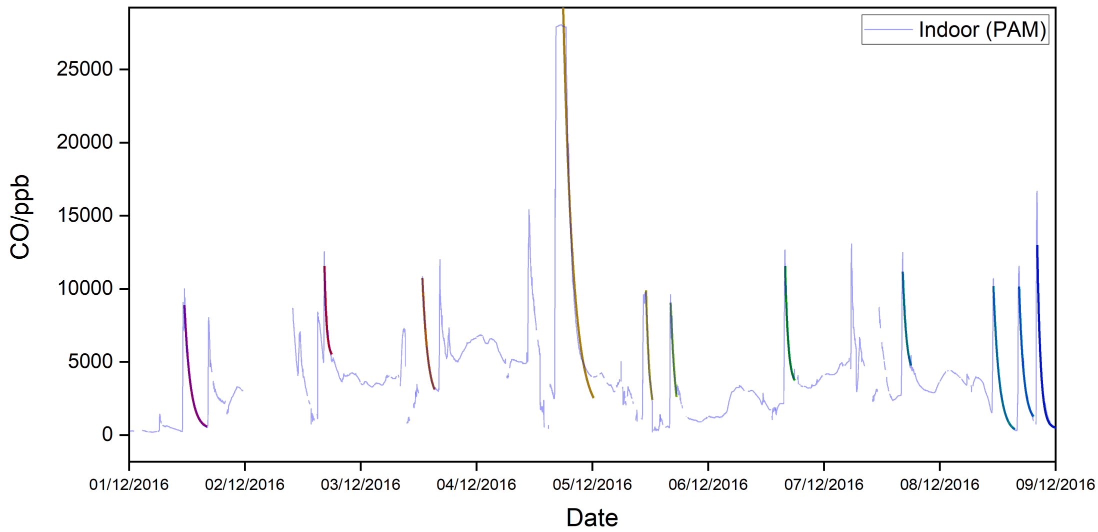

Figure 4.7: Fit of exponential decays to data within identified decaying regions: A time series of the indoor PAM CO concentration (blue) recorded by Participant U143 with exponential curves that passed QA/QC Step 2iii, fitted to the decaying regions that passed the QA/QC Step 1v. The indoor PAM data have been plotted at 70% transparency so that the fitted exponential decays are more visible.

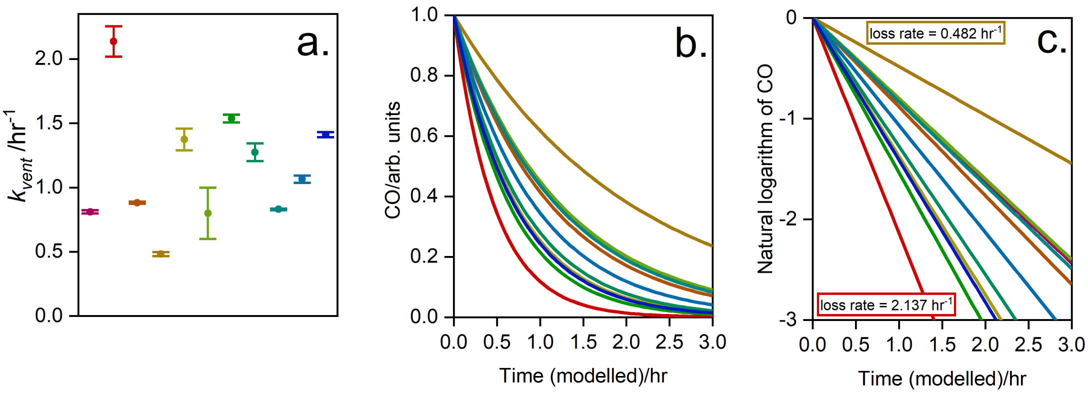

Figure 4.8: Visualisation of estimated ventilation rates: Three plots for visualisation of the ventilation rates estimated from Figure 4.7. The colours selected in these three plots correspond to the colours of the exponential decays in Figure 4.7. a.) Simple scatter plot of the estimated ventilation rates, along with the standard errors. b.) Graph of $y = e^{-cx}$ decays where c are exponential decay rates and x takes values between 0-3 hours. This is to allow for easy visual comparison of the decay rates. c.) Natural logarithms of the exponential curves from b. The loss rates for the fastest and slowest decays have been included in the figure.

2iv: The coefficients of the exponential decays were saved in a table of a relational database together with other contextual information. This is shown for the example participant in Table 4.1. For this example, the time section length during which the decay rate was analysed was set to be 12 hours. Letters of the alphabet are used to denote 12-hour time sections in chronological order starting at 6am before data was recorded, with the letter a. 12-hour time sections starting at 06:00 and 18:00 were selected to capture differences between night and day.

Participant ID	Season	Field campaign	Start time of decaying region	End time of decaying region	Time section	Micro-environment	a	b	c (k_{vent})	r^2
U143	winter	Beijing	01/12/2016 11:30	01/12/2016 16:12	c	home	28.018	8508.3	0.811	0.997
U143	winter	Beijing	02/12/2016 16:30	02/12/2016 18:00	e	home	3343.6	6284.1	2.137	0.981
U143	winter	Beijing	03/12/2016 12:48	03/12/2016 15:18	g	home	547.54	8588.6	0.882	0.999
U143	winter	Beijing	04/12/2016 17:39	05/12/2016 00:17	i	home	-1690.5	32663	0.482	0.909
U143	winter	Beijing	05/12/2016 11:08	05/12/2016 12:27	k	home	682.06	8946.9	1.374	0.986
U143	winter	Beijing	05/12/2016 16:14	05/12/2016 17:27	k	home	-1549.0	10353	0.800	0.981
U143	winter	Beijing	06/12/2016 15:57	06/12/2016 17:55	m	home	1607.1	8219.4	1.537	0.997
U143	winter	Beijing	07/12/2016 16:20	07/12/2016 18:02	o	home	1243.9	7257.2	1.274	0.985
U143	winter	Beijing	08/12/2016 11:07	08/12/2016 15:31	q	home	-209.33	10025	0.830	0.991
U143	winter	Beijing	08/12/2016 16:26	08/12/2016 19:24	q	home	557.71	9261.8	1.065	0.992
U143	winter	Beijing	08/12/2016 20:11	09/12/2016 00:02	r	home	-8.2138	12559	1.412	0.977

Table 4.1: Table of coefficients: Table of the coefficients of the exponential curves that were saved for Participant U143 in Step 2iv, along with additional contextual information. r^2 is the value calculated in Step 2ii. Microenvironment here refers to the mode microenvironment of the participant during that specific decaying region.

Coefficients and contextual information for all participants were appended as rows as the algorithm iterated.

4.2.2 Estimation of the outdoor-generated component of indoor levels of an inert species

The outdoor-generated component $I_{t(outgen)}$ for an inert species is the outdoor time series but with the ventilation rate applied. Ventilation rates may have large temporal variation driven by changing wind velocities and temperature gradients between the indoor and outdoor microenvironments. Time sections where no k_{vent} value was estimated using the constant decay method do not have a 12-hour constant value. Time sections with one k_{vent} value use this value as their 12-hour constant value. Time sections with more than one k_{vent} value will use a mean value, calculated using inverse-variance weighting. This is justified further in Appendix A.4.3. The resultant 12-hour constant k_{vent} values for this participant are in the lower section of Table 4.2.

Time section	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
k_{vent}			0.811		2.137		0.882		0.482		1.374		1.537		1.274		0.830	1.412
											0.800						1.065	

⇩

k_{vent}			0.811		2.137		0.882		0.482		1.200		1.537		1.274		1.050	1.412

Table 4.2: Calculating ventilation rate for each time section: The top table is the "time section" and "c" columns from the table in Figure 4.1. The bottom table shows the calculated k_{vent} values for each 12-hour time section. All values are for Participant U143.

These ventilation rates can then be applied to their respective 12-hour time sections of the data through the relationship in Equation 2.7 to produce a time series of $I_{t(outgen)}$. This is plotted in Figure 4.9.

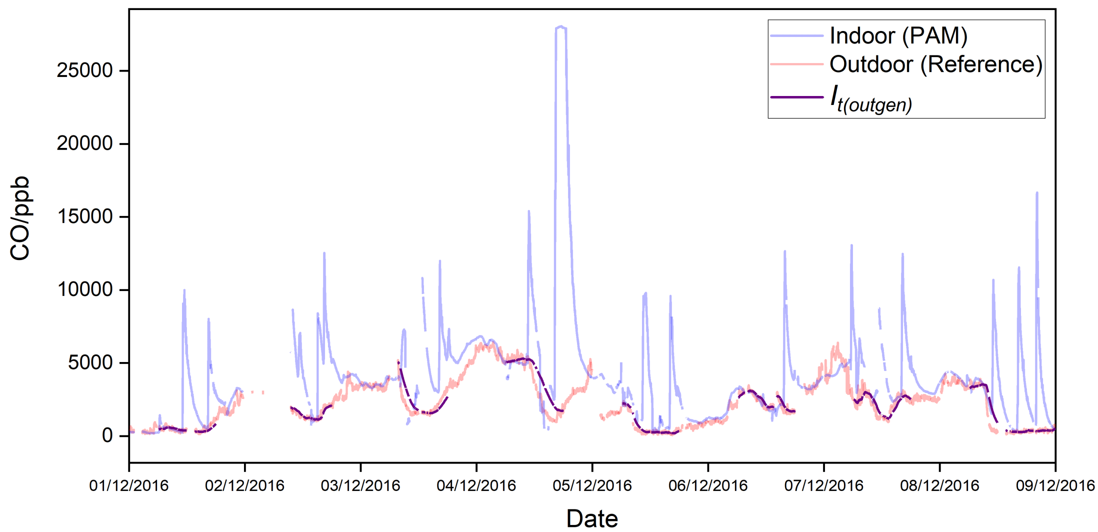

Figure 4.9: Applied ventilation rates to outdoor data: A time series of the indoor (PAM) CO (blue) recorded by Participant U143 and the CO reference data (red). The outdoor-generated component of indoor air ($I_{t(ingen)}$) is plotted (purple). This has been generated by applying the ventilation rates shown in Figure 4.2 to their respective 12-hour time sections. The indoor (PAM) data and outdoor (reference) data have been plotted at 70% transparency so that $I_{t(outgen)}$ is more visible.

To find the indoor-generated component ($I_{t(ingen)}$), the outdoor-generated component ($I_{t(outgen)}$) is subtracted from the indoor (PAM) time series.

4.2.3 Source-apportionment for all time sections for an inert species

The number of k_{vent} values estimated from the AIRLESS dataset, broken down by location and season, can be found in Chapter 5, Figure 5.8. 160 out of the 469 time series from the AIRLESS project had no assigned k_{vent} values due to a lack of indoor CO indoor emission events in the data. Therefore many time sections had no assigned k_{vent} value. For time sections with insufficient CO indoor emission events, the $I_{t(outgen)}$ (and therefore $I_{t(ingen)}$) time series cannot be generated using this method. An alternative method can estimate values of $I_{t(outgen)}$ and $I_{t(ingen)}$ over coarse time intervals if it is assumed that the last term in Equation 2.7 tends to 0, and so the time lag between indoor and outdoor data, due to ventilation, is assumed insignificant. Having made this approximation, the indoor-generated and outdoor-generated components of CO are plotted in Figure 4.10. The outdoor-generated time series is assumed to be the same as the outdoor level. The outdoor time series was subtracted from the indoor PAM data to produce the indoor-generated time series.

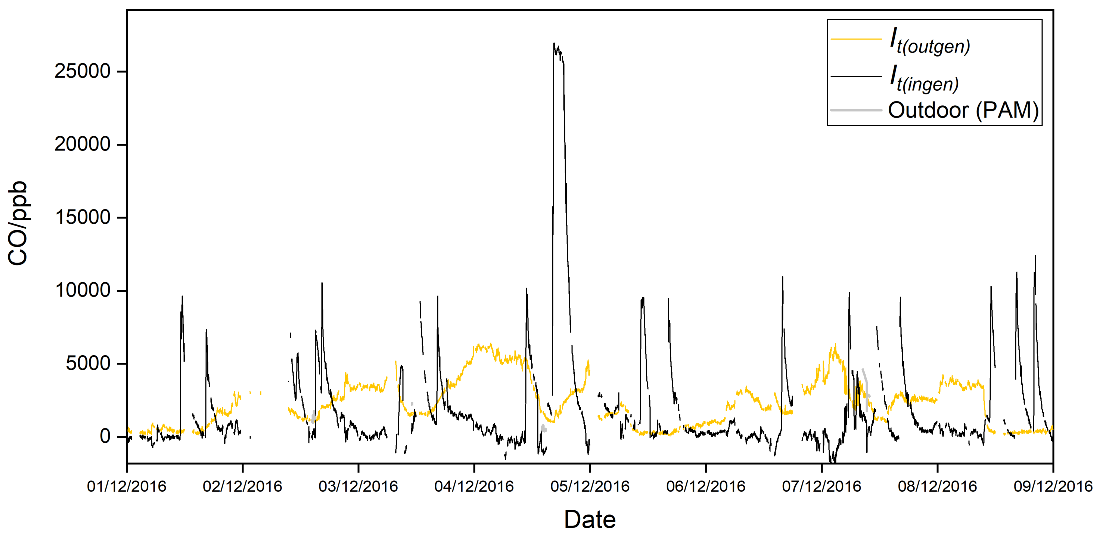

Figure 4.10: Indoor- and outdoor-generated components of CO time series: A time series of outdoor (PAM) data (grey), the indoor-generated component of the CO ($I_{t(ingen)}$) (black) and the outdoor-generated component of the CO ($I_{t(outgen)}$) (orange). PAM data was recorded by Participant U143. This participant spent very little time outdoors.

12-hour averages of indoor- and outdoor-generated components of personal exposure are calculated to be used in pollution-health models. As participants do not spend all of their time indoors, the following considerations are made:

For the outdoor-generated component: participants breathe outdoor-generated pollution when outdoors (grey) and when indoors (outdoor-generated component of indoor air)(orange). Participants are always exposed to outdoor-generated exposure and so the mean of the orange and grey lines represents the outdoor-generated component of total personal exposure.

For the indoor-generated component: participants are assumed to only breathe indoor-generated pollution when in an indoor environment. Participants are not always indoors, so to account for that, the average of the indoor-generated component measured indoors (black) must be time-weighted (multiplied by the proportion of time that the participant is indoors).

The results for this participant for all 12-hour sections are shown in Figure 4.11. The results for the whole dataset are shown in Chapter 5.

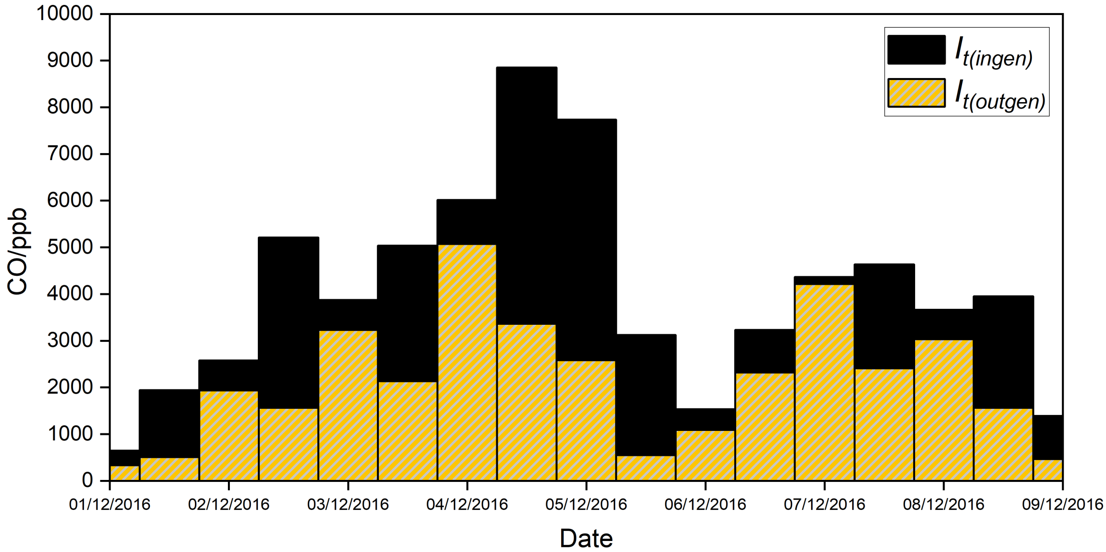

Figure 4.11: Averages of indoor- and outdoor-generated components of total personal exposure to CO for the example participant: Stacked bar chart showing 12-hour estimates of the exposure of Participant U143 to indoor-generated and outdoor-generated CO. Indoor-generated exposure has been time weight adjusted. When a participant is outdoors, it is assumed that participant is only exposed to outdoor-generated CO.

For the example participant shown in Figure 4.11, the proportion of indoor-generated to outdoor-generated is generally higher in the day (06:00-18:00) than in the night (18:00-06:00). Effects like these are quantified for the whole population in Figure A.8.

4.3 Source-apportionment of personal exposure of a reactive species

Figure 4.12 shows that there are evident indoor loss processes for NO_2 for Participant U143 as the indoor level recorded by the PAM is predominantly lower than the outdoor (reference) level.

The outdoor-generated component of indoor concentrations for a reactive species is estimated by applying the ventilation rate and indoor loss rate to outdoor data. As mentioned in Section 4.2.2, many time sections do not have an assigned k_{vent} value. Therefore the I/O ratio is used for apportionment of a reactive species instead of using the loss rate and ventilation rate.

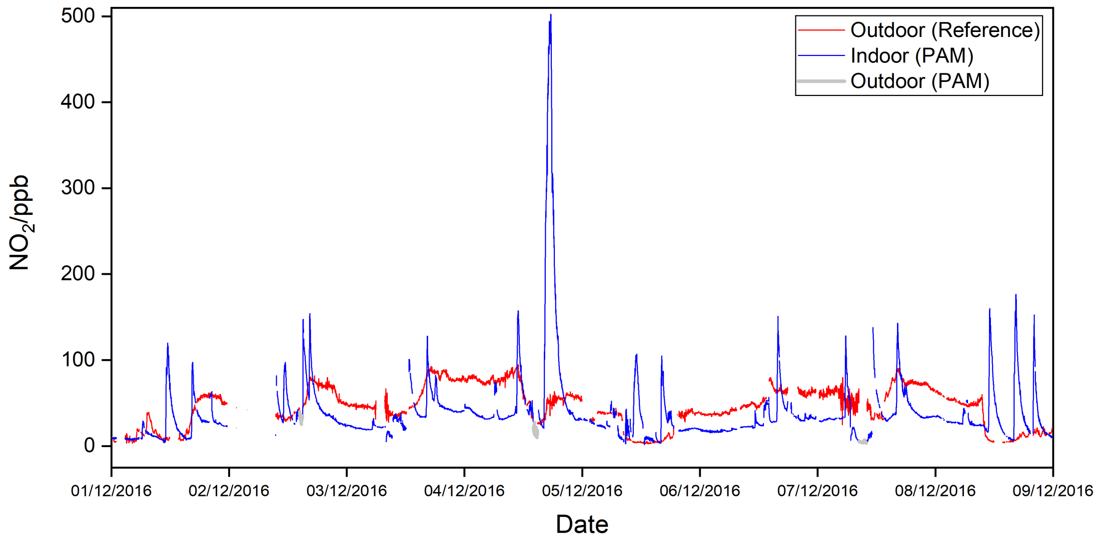

Figure 4.12: Real data to illustrate time series features of a reactive species: A time series of outdoor (PAM) NO_2 data (grey), indoor (PAM) NO_2 data (blue) and the NO_2 outdoor (reference) data (red). PAM data was recorded by Participant U143.

It is assumed that the I/O ratio remains constant for the 12-hour time sections, although in reality, it is likely more dynamic. The estimation of the I/O ratio for a single time section is shown in Figure 4.13 for the example participant.

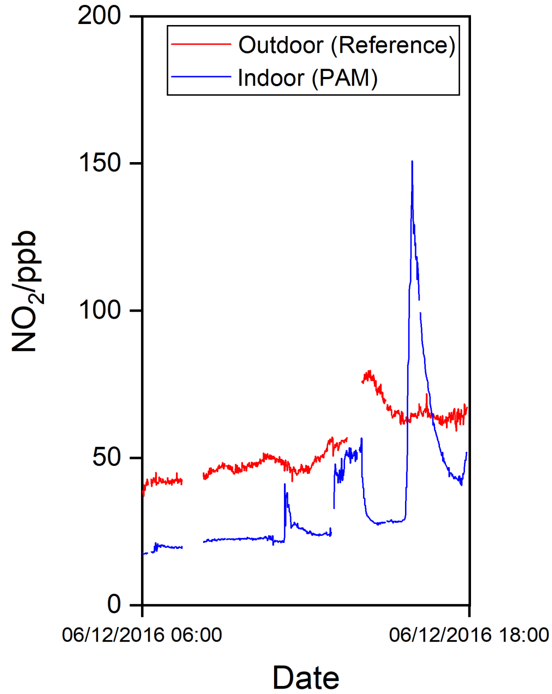

Figure 4.13: Single time section of NO_2 indoor and outdoor levels: The 13th 12-hour time section of Figure 4.12 (the time section labelled "m"). The indoor (PAM) NO_2 data is blue and the NO_2 outdoor (reference) data is red. PAM data was recorded by Participant U143.

There are three steps to source-apportion the indoor level of a reactive species:

Step 1 The I/O ratio is estimated using time points where there is an absence of indoor sources. For each time point, the indoor (PAM) data is divided by the outdoor (reference) data. Only the time points where this results in a number less than 1 are retained. This removes the strong indoor emission events in the indoor PAM data (blue) in Figure 4.13. Note, if there is a constant indoor source of the pollutant (strong enough that that the indoor level is higher than the outdoor level), time points are not retained past this step. This methodology is not appropriate for pollutants with strong constant indoor sources of pollution.

Step 2 The remaining values of the indoor (PAM)/outdoor (reference) for this 12-hour period are plotted in a histogram as shown below (Figure 4.14). The selected I/O ratio is the mode of the distribution. For this example time section, the I/O ratio is 0.475.

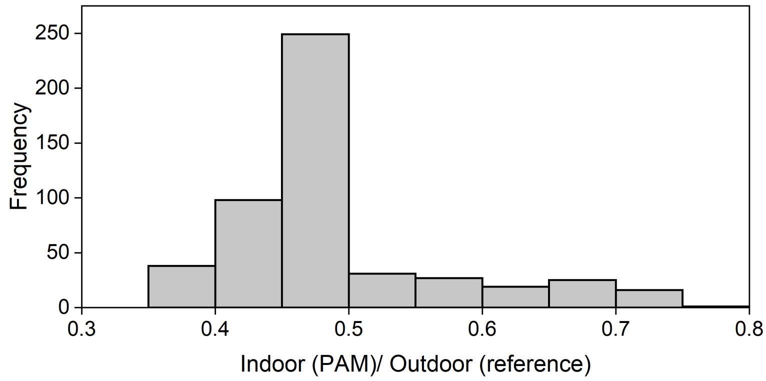

Figure 4.14: Indoor/outdoor histogram for single time section: A histogram of every indoor (PAM)/outdoor (reference) data point larger than 1 for the 12 hours of data being analysed. Sturges' breaks are used to bin the data. PAM data was recorded by Participant U143.

Step 3 The outdoor-generated time series is produced by multiplying the outdoor (reference) data by the optimum I/O ratio as shown in Figure 4.12 (the red time series is multiplied by the optimum I/O ratio (0.475) to produce the orange time series).

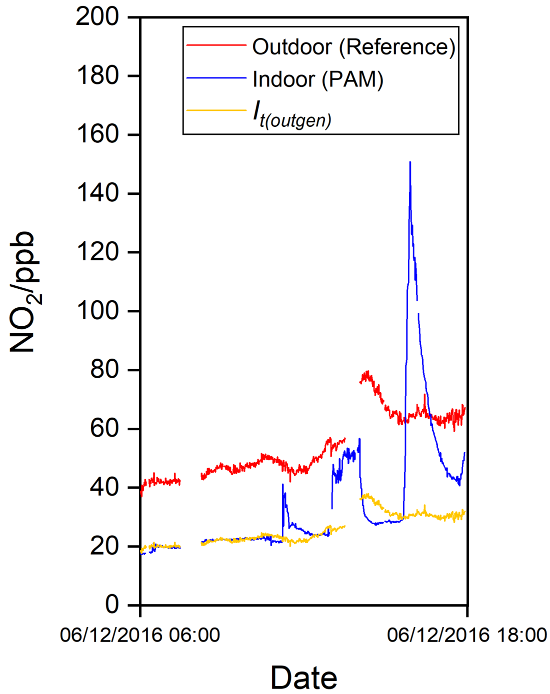

Figure 4.15: Outdoor-generated component of indoor air for single time section: A 12-hour time series of the indoor (PAM) data (blue), the outdoor (reference) data (red) and the outdoor-generated time series (orange). PAM data was recorded by Participant U143.

Figure 4.15 shows the outdoor-generated pollution (orange). The blue spikes above the orange correspond to indoor-generated emission events. This methodology is applied to every 12-hour time section for this participant and the results are shown in Figure 4.16.

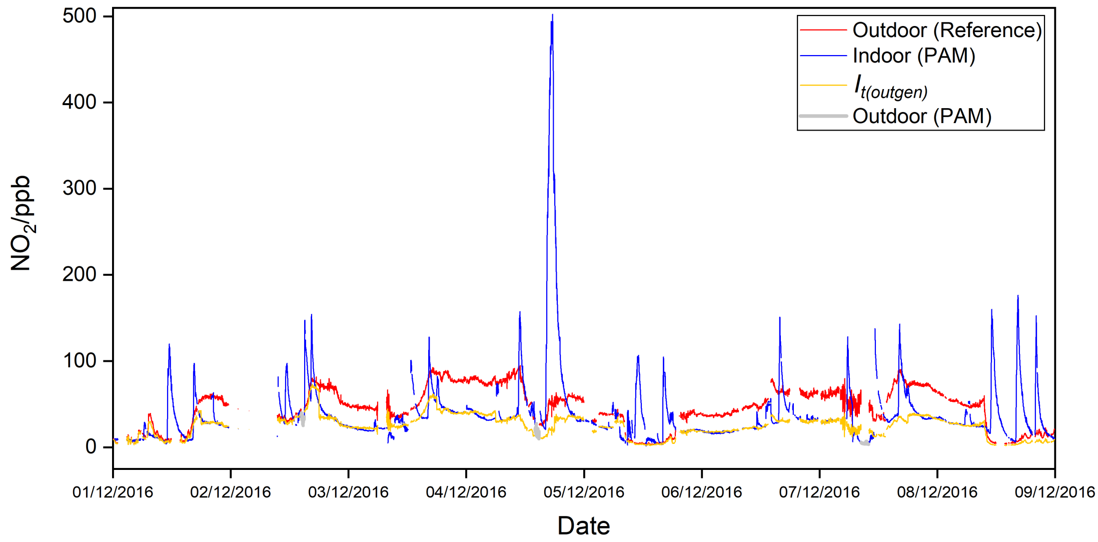

Figure 4.16: Outdoor-generated component of NO_2 time series: A time series of outdoor (PAM) NO_2 data (grey), indoor (PAM) NO_2 data (blue), NO_2 outdoor (reference) data (red), and outdoor-generated NO_2 (orange). PAM data was recorded by Participant U143.

The emission events that correspond to indoor-generated pollution ($I_{t(ingen)}$) can be isolated by subtracting the outdoor-generated pollution ($I_{t(outgen)}$) (orange) from the PAM indoor measurements (blue). The indoor-generated (black) and outdoor-generated (orange) components of the indoor PAM measurements are plotted together in Figure 4.17.

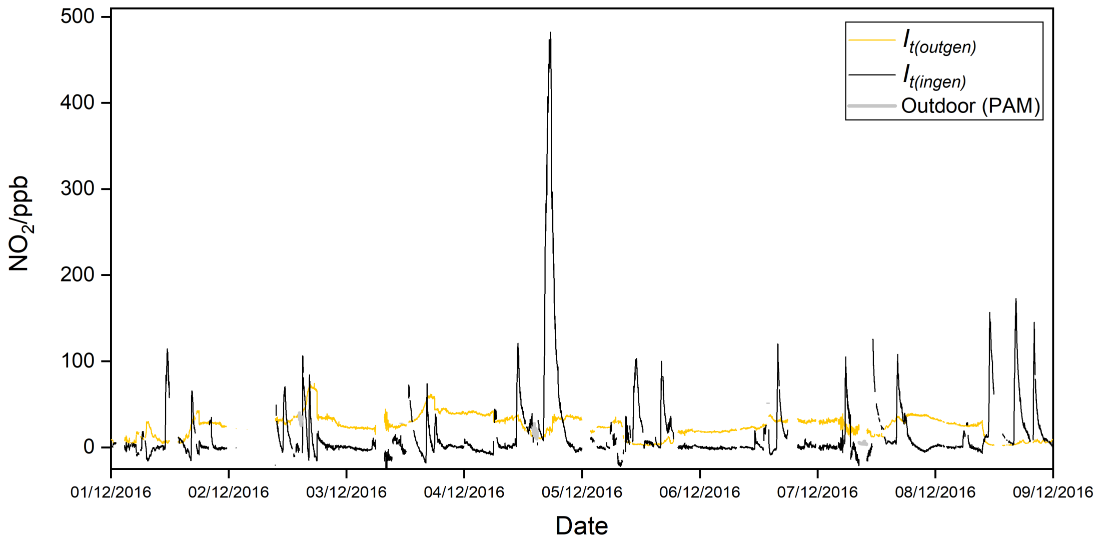

Figure 4.17: Indoor- and outdoor-generated components NO_2 exposure for the whole of the example participant's time series: A time series of the indoor-generated NO_2 time series (black), the outdoor-generated NO_2 time series (orange), and the outdoor (PAM) NO_2 data (grey). PAM data was recorded by Participant U143.

12-hour averages of indoor- and outdoor-generated components of total personal exposure are calculated in the same way as for an inert species, as in Section 4.2.3.

The results for this participant for all 12-hour sections are shown in Figure 4.18.

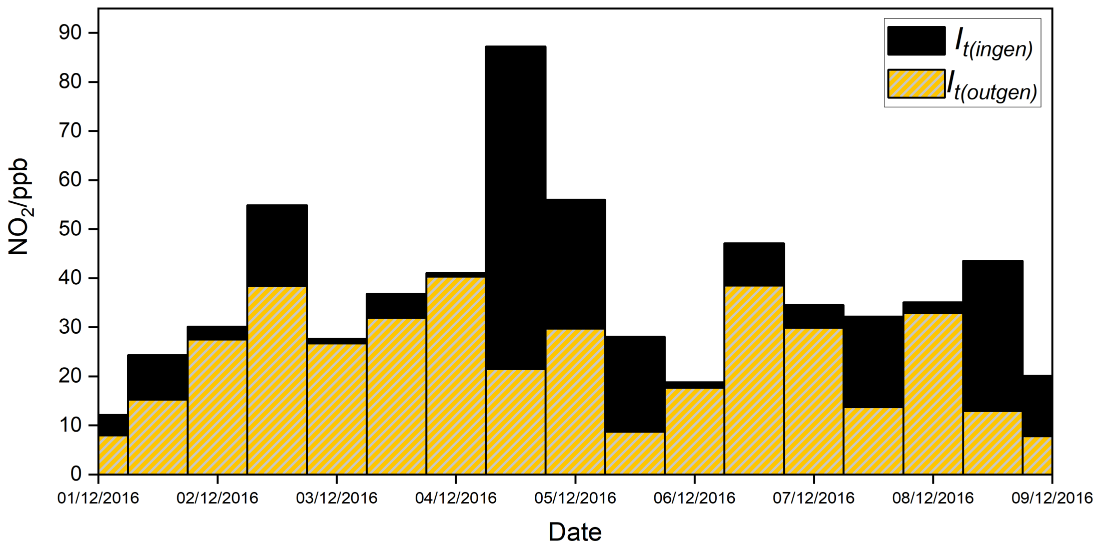

Figure 4.18: Averages of indoor- and outdoor-generated components of total personal exposure to NO_2 for the example participant: Stacked bar chart showing 12-hour estimates of the exposure of Participant U143 to indoor-generated and outdoor-generated NO_2. Indoor-generated exposure has been time weight adjusted as it is assumed that participants, when in an outdoor environment, are only exposed to outdoor-generated NO_2.

For the example participant shown in Figure 4.18, the proportion of indoor-generated to outdoor-generated is generally higher in the day (06:00-18:00) than in the night (18:00-06:00).

4.4 Estimation of indoor pollutant loss rates

As mentioned in Section 2.2, indoor reactive and particulate species can have many fates, including reacting, depositing onto surfaces and transforming into many products. Quantities of indoor loss rates are key exposure determinants and can be used to model pollution exposure in population-scale studies. The computational steps to estimate the indoor loss rate, via two methods, are detailed below.

4.4.1 Estimation of the indoor loss rate using the I/O ratio

Using the relationship in Equation 2.16 (repeated below), a constant value of k_{sink} can be estimated from each 12-hour time section if the corresponding I/O ratio and k_{vent} for the 12-hour period are known. Methods of estimating the I/O ratio and k_{vent} for 12-hour time sections were detailed in Sections 4.3 and 4.2.1 respectively.

$$\frac{I_{t(outgen)}}{O_t} = \frac{k_{\text{vent}}}{(k_{\text{vent}} + k_{\text{sink}})} \tag{2.16}$$

The results of this are shown in Table 4.3 for each 12-hour time section for the example participant.

Time section	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
k_{vent}			0.811		2.137		0.882		0.482		1.200		1.537		1.274		1.050	1.412
I/O ratio	0.763	0.907	0.635	0.485	0.882	0.495	0.555	0.533	0.425	0.585	0.255	0.475	0.475	0.495	0.470	0.535	0.480	0.475
k_{sink}			0.243		0.214		0.441		0.724		0.300		1.690		1.912		0.945	1.765

Table 4.3: Estimated indoor loss rates using the I/O ratio method: A table showing the constant values of k_{vent} and I/O ratio for each 12-hour time section, and the resultant calculated constant k_{sink} values. All data shown is for Participant U143 and the reactive pollutant being analysed is NO_2.

4.4.2 Estimation of the indoor loss rate using the constant decay method

Alternatively, the indoor loss rate can be estimated from the decay rates of indoor emission events using the constant decay method. The automated steps to identify and characterise exponential decays were outlined in Figure 4.2. For a reactive species, the asymptote of the decays is the mean outdoor-generated component during the decay. Figure 4.19 shows the indoor PAM data (I_t), the outdoor-generated component of indoor air ($I_{t(outgen)}$) and the mean $I_{t(outgen)}$ values during identified decaying regions (calculated from data recorded during the decay and during the hour before the start of the decay).

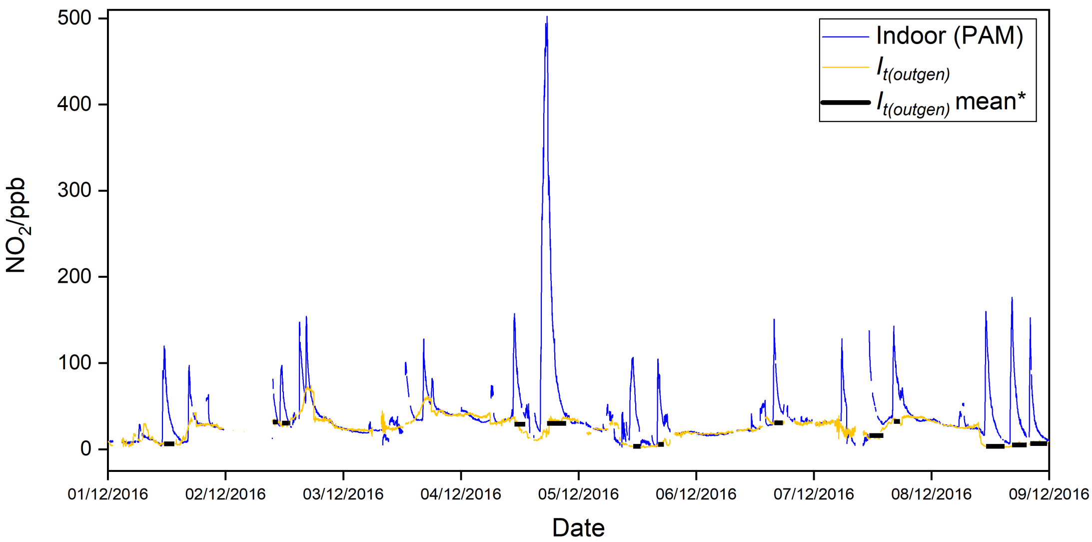

Figure 4.19: Mean outdoor-generated component levels during decaying regions: A time series of the indoor (PAM) NO_2 (blue) recorded by Participant U143, the outdoor-generated component of indoor NO_2 ($I_{t(outgen)}$) (orange) and the mean* $I_{t(outgen)}$values (black). *The mean $I_{t(outgen)}$values were calculated from data recorded during the decay and during the hour before the start of the decay and are plotted for the decaying regions.

The fitted exponential curves are plotted in Figure 4.20.

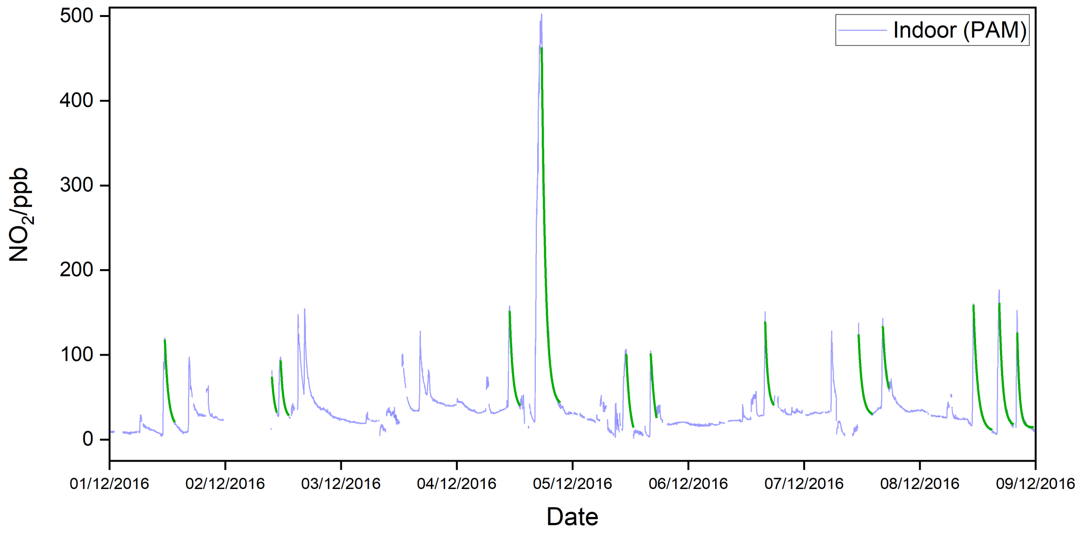

Figure 4.20: Fit of exponential decays to data within identified decaying regions: A time series of the indoor (PAM) NO_2 (blue) recorded by Participant U143 with fitted exponential curves (red). The exponential curves and decaying regions passed the same QA/QC requirements as in Section 4.2.1 with fitting exponential curves to inert data to estimate ventilation. The indoor (PAM) data have been plotted at 70% transparency so that the fitted exponential decays more visible.

As the decay rates are the cumulative influence of ventilation and indoor loss processes, to calculate a value of k_{sink} from these decay rates, a value of k_{vent} must be subtracted. The value of k_{vent} is assumed constant for each 12-hour section. A constant k_{vent} value for each 12-hour section was calculated using the methodology detailed in Section 4.2.2 and the values were tabulated in Figure 4.2. Figure 4.21a shows the CO exponential decays that were used to infer ventilation rates for 12-hour time sections. These ventilation rates are plotted for their respective 12-hour time section in Figure 4.21b. Figure 4.21c is a repeat of Figure 4.20 and allows for visualisation of the timing of the NO_2 decays with respect to the 12-hour time sections. Figure 4.22 directly compares the CO and NO_2 decay rates measured within the same time section using graphs of y=e^{-cx}. c are exponential decay rates of both CO and NO_2 decays and these plots have been produced for a 2-hour period, for each time section with at least one NO_2 decay.

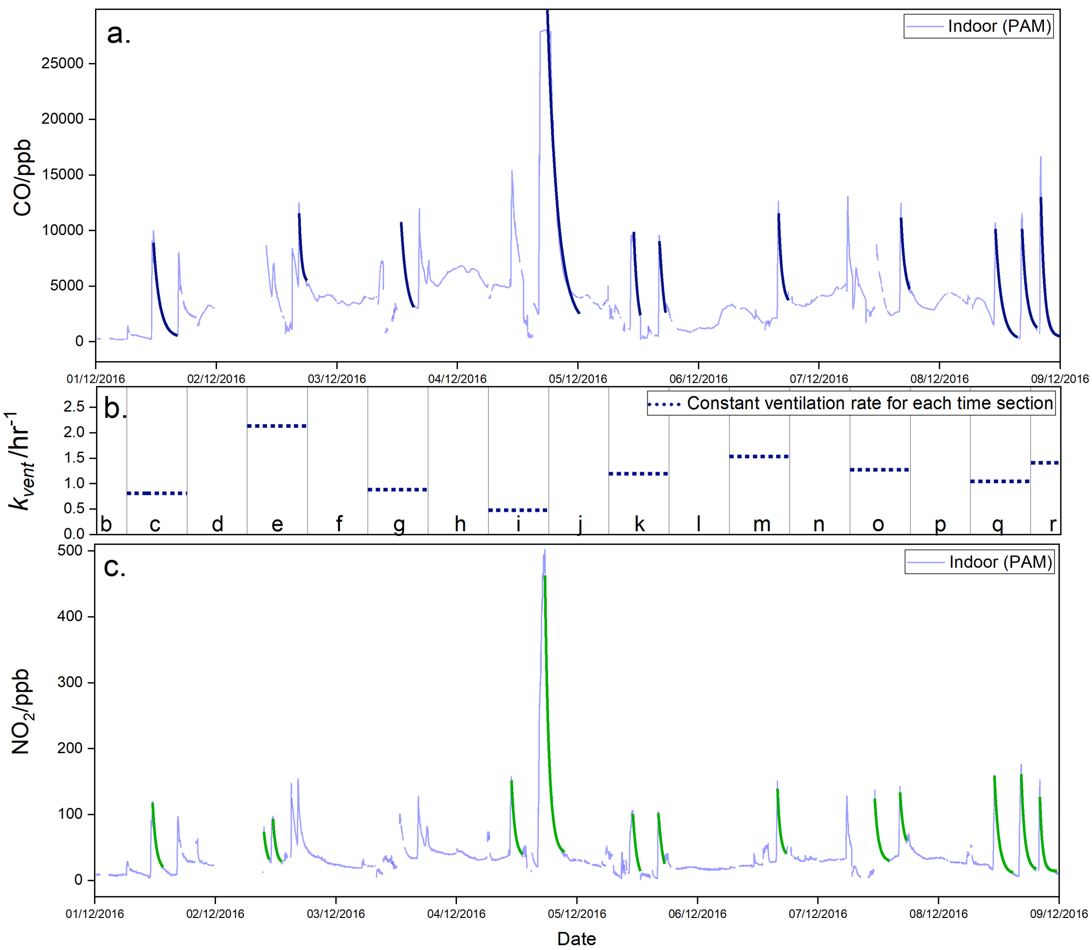

Figure 4.21: Comparison of exponential decays in CO and NO_2 data: a.) Indoor (PAM) CO (transparent blue) with exponential decays fitted (dark blue). Ventilation rates are inferred from the decay rates of the dark blue exponential curves. This plot contains identical data to Figure 4.7. b.) The 12-hour constant k_{vent} values calculated in Figure 4.2, inferred from a. c.) Indoor (PAM) NO_2 (transparent blue) with exponential decays fitted (green). This plot contains identical data to Figure 4.20.

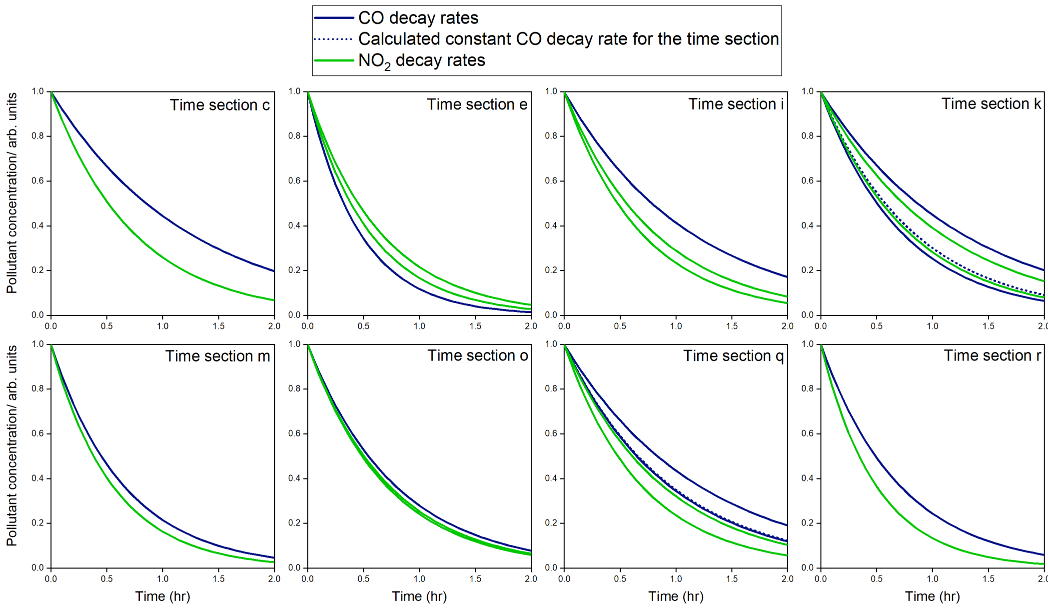

Figure 4.22: Comparison of decay rates in CO and NO_2 data: Plots of y=e^{-cx} decays where c are exponential decay rates, estimated from CO (dark blue) and NO_2 (green) time series. x takes values between 0-2 hours and have been produced for each time section with at least one NO_2 decay. Dotted decays have used the calculated constant k_{vent} value for the time section and are plotted for time sections where more than one k_{vent} was recorded. The green more frequently decays faster than the dark blue however there are cases when this is reversed, which result in negative k_{sink} values.

Occasionally the exponential decays from the CO time series are faster than the exponential decays in the NO_2 series. This results in a negative outputted k_{sink} value. This may be due to increases in ventilation, for example from opening a window, which aren't captured in the assumed constant ventilation rate for the 12-hour section.

4.4.3 Comparison of the indoor loss rate from the two estimation methods

The estimated k_{sink} values for this example participant using the two estimation methods are plotted in a histogram in Figure 4.23 and are compared for each time section in Figure 4.24. Section 5.2.2 explores how the differences between the methods could be investigated.

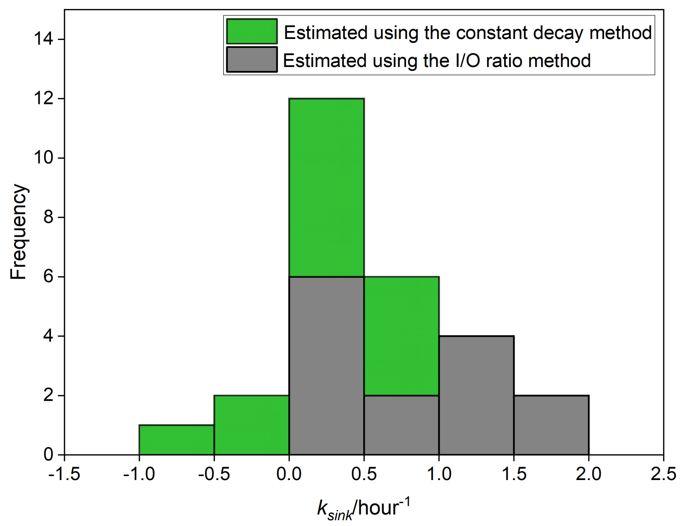

Figure 4.23: Distribution of $k_{\mathbf{sink}}$ values: A stacked histogram of the values of NO_2 k_{sink} estimated using two different methods: using the I/O ratio method (grey) and the constant decay method (green) for Participant U143.

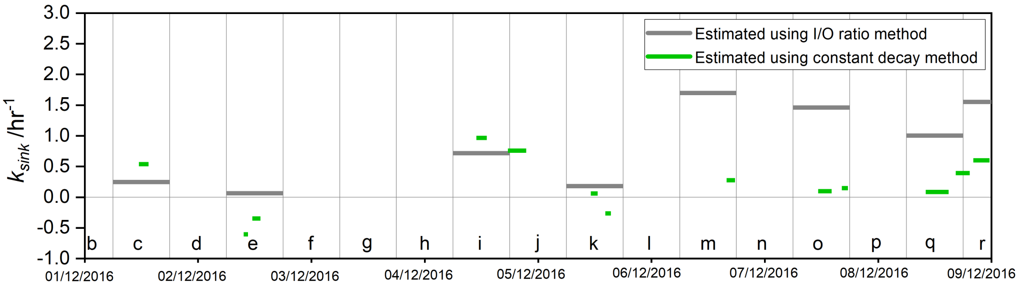

Figure 4.24: Comparison of the $k_{\mathbf{sink}}$ values estimated for individual time sections: Time plot comparing the estimated k_{sink} values for NO_2 estimated for Participant U143 using 2 methods. One method uses the I/O ratio and k_{vent} value, producing a constant k_{sink} value (grey). The other method uses the exponential decay rates found in the NO_2 data and subtracts the constant k_{vent} value for that time section (green).

Both of these methods involve application of the constant ventilation rate within a 12-hour time section. An alternative to would be to subtract the ventilation rate inferred from a synchronised CO decay from the NO_2 exponential decay rates in Figure 4.21. Practically, this synchronised approach has been achieved by selecting the CO exponential decays and the NO_2 exponential decays that have some overlap in time. This method has pros and cons:

Pros: With this alternative approach, the data being analysed for each k_{sink} value fall within smaller time ranges compared with subtracting a constant value of k_{vent} over 12 hours. The time activity model (Section 3.6) computes the space-time utilisation distributions of the GPS coordinates (and therefore microenvironment) for each participant for each minute of their data. Whilst the constant k_{vent} value approach takes the mode microenvironment computed over the 12-hour period, the synchronised approach takes the mode microenvironment computed only during the synchronised exponential decays. Therefore the synchronised approach is more useful if wanting to compare k_{sink} for a range of microenvironments.

Cons: This alternative method requires synchronised CO NO_2 decays. For the example participant, this would not appear to be a problem as the CO NO_2 decays are often synchronised. However, most participants have fewer indoor emission events and other reactive pollutants are not as often co-emitted with CO in the indoor environment, therefore this approach is not suitable for this dataset.

4.5 Characterising indoor emission events

Sections 4.2 and 4.3 detailed how indoor-generated pollution can be isolated from data recorded indoors by the PAM. These indoor-generated pollution time series feature indoor emission events where the level of indoor pollutants increased sharply as a result of an indoor source of pollution eg. cooking or smoking.

The characteristics of these events could provide insights into the sources which can be used for indoor air quality modelling or health associations. Often the pollution exposure metrics that are inputted into pollutant-health models are concentrations averaged over coarse time periods and so the health effect of emission events is lost: long-term exposure to low pollution concentrations may have a different health effect than short-term exposure to high pollution levels.

This section demonstrates the methodology to extract some quantifiable characteristics of indoor emission events. CO indoor-generated data from Participant U143 is used as an example case.

4.5.1 Detection of indoor emission events

In Step 1ii of Section 4.2.1, peaks and troughs were found in the indoor CO time series for Participant U143. Using the same algorithm, peaks in the indoor-generated component of indoor air can be found (peaks are defined as the rolling maximum within 30 data points).

Strong indoor emission events are of principal interest, so a threshold limit is used. In the example case the 90^{th} percentile of the indoor-generated CO time series, calculated across the whole cohort (3495 ppb) is used.

4.5.2 Extraction of the quantity and height of peaks of indoor emission events

Figure 4.25 plots the events in the indoor-generated data for this example participant.

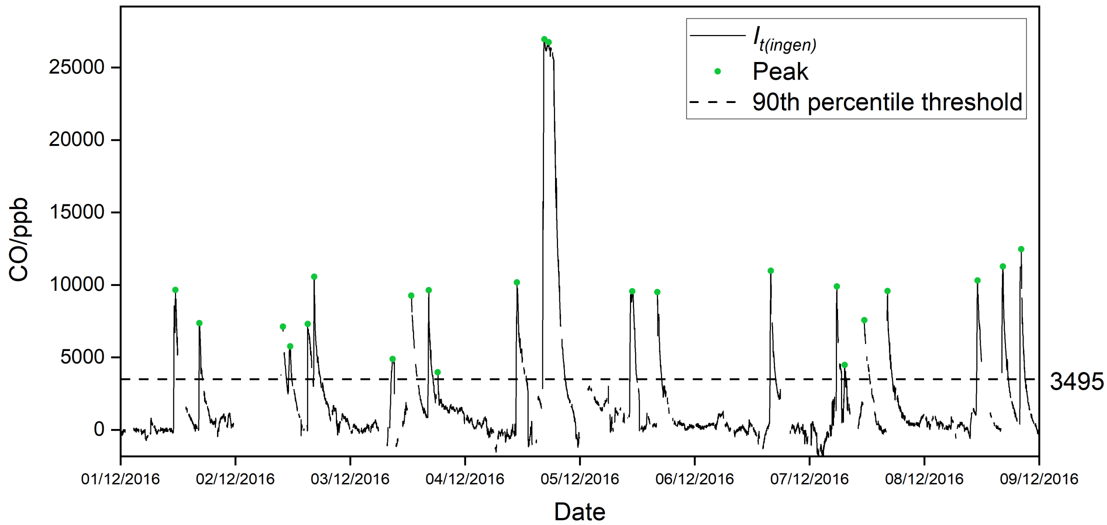

Figure 4.25: Peaks of indoor emission events: A time series of the indoor-generated CO (black) with the peak of indoor emission events (green). The 90^{th} percentile threshold is plotted as a dashed line. The indoor-generated CO was produced using indoor PAM data recorded by Participant U143.

For this example, 22 indoor emission events were identified over a week of measurements and the average height was 9653 ppb. The quantity and height of the peaks of emission events can be obtained for all 12-hour time sections, pollutants and participants in the AIRLESS cohort, and then linked with health markers using an LMEM.

4.5.3 Extraction of area under indoor emission events

The areas of indoor emission events have been estimated as the region under the graph between each data point of the time series as a trapezoid. The dataset contains periods of missing data and periods where the participant was not indoors. To ensure that the algorithm does not approximate trapezoids over large time periods of missing data or outdoor data, data gaps of over 2 hours are assumed to begin and end with troughs. This avoids the algorithm overestimating the area of the events. Figure 4.26 shows the area under the time series, but above the threshold, for the example participant. The area is estimated in units of (ppb)x(hour).

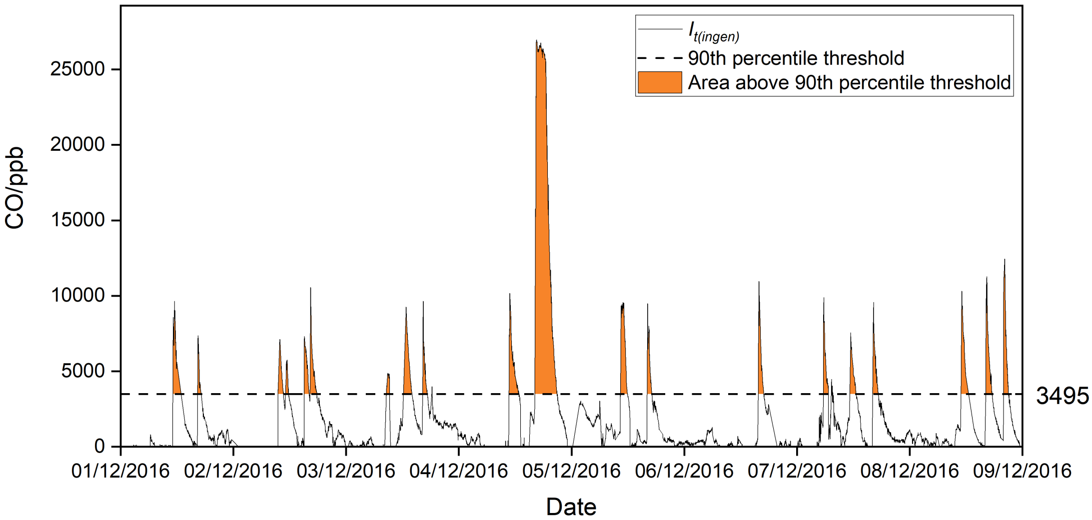

Figure 4.26: Area under indoor emission events: A time series of the indoor-generated CO (black). The 90^{th} percentile threshold is plotted as a dashed line. The area under the time series (estimated using the trapezoid rule) above the 90^{th} percentile threshold shown in orange. The indoor-generated CO was produced using indoor PAM data recorded by Participant U143.

The total orange area in Figure 4.26 is estimated to be 15684 ppb hour, however this metric can be obtained for all 12-hour time sections, pollutants and participants in the AIRLESS cohort, and then linked with health markers using an LMEM.

4.5.4 Extraction of the duration of indoor emission events

A metric of the duration of indoor emission events has been calculated for the data above the threshold. The duration is given as a percentage and was calculated by dividing the number of data points recorded by the participant while their indoor-generated pollutant level exceed the threshold, by the total number of data points recorded by the participant while they were indoors. This is demonstrated for the example participant's indoor-generated CO in Figure 4.27 for the 90^{th} percentile threshold.

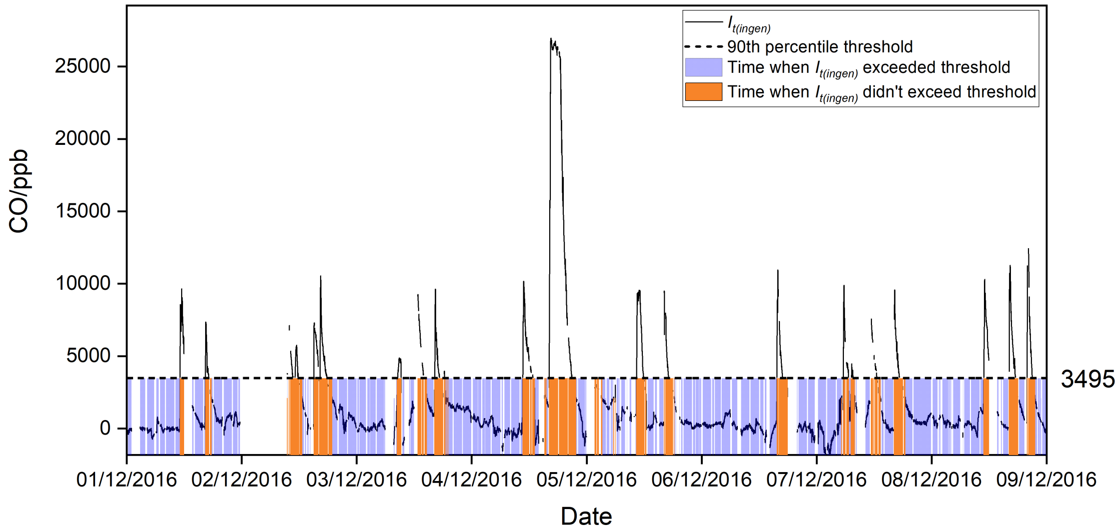

Figure 4.27: Duration of indoor emission events: A time series of the indoor-generated CO (black). The 90^{th} percentile threshold is plotted as a dashed line. Along the x-axis are orange and purple lines indicating whether the data point falls above the threshold or below the threshold. For this example participant, 24% of the recorded data points exceed the threshold. The indoor-generated CO was produced using indoor PAM data recorded by Participant U143.

The duration of the indoor emission events above the threshold in Figure 4.27 is calculated as 24%:

$$\frac{\text{number of data points above threshold}}{\text{number of data points below threshold}} = \frac{2208}{9040} = 24\% \tag{4.2}$$

4.6 Chapter summary

This chapter provided an example case of how the continuity equation and methodology outlined in Chapter 2 can be applied to PAM and reference data from the AIRLESS dataset, as described in Chapter 3. The same methodology is applied to the whole AIRLESS cohort. Key results and insights are presented in Chapter 5.

Chapter 5

Assessment of personal exposure of the AIRLESS cohort

Chapter 2 introduced a methodological framework to assess total personal exposure, which was applied to the data from a single participant from the AIRLESS dataset in Chapter 4.

This chapter applies this framework to the whole AIRLESS dataset, generating novel exposure metrics and estimating values of exposure determinants. Key insights and results are compared with literature values, focusing on results from studies in China where available.

5.1 Novel exposure metrics from personal air quality monitoring data

Exposure refers to the extent to which an individual is in contact with air pollutants, encompassing the duration, frequency, and concentration of their exposure. It is a measure of the direct interaction between humans and air pollutants.

On the other hand, risk involves the probability and magnitude of adverse health effects resulting from exposure to these pollutants. Risk integrates not only the level and duration of exposure but also dose (influenced by the rate of breathing) and factors in the susceptibility of individuals, considering variations in age, pre-existing health conditions, and genetic predispositions. This thesis considers exposure. Chapter 7 explains how this work could be extended to consider dose.

Many pollution-health assessments use exposure metrics from stationary outdoor instruments, as explored in Appendix A.1.3. Comparison of the novel metrics (generated by the application of the methodological framework demonstrated in Chapter 4 to the AIRLESS dataset) with the metrics from stationary outdoor instruments allows for assessment of exposure misclassification.

5.1.1 Personal exposure metrics

Personal monitoring captures the effects of an individual moving between different microenvironments. Figure 5.1 shows how pollutant concentrations measured by the PAM varied between different microenvironments. These concentrations have been broken down by location and season and can be found in Appendix A.5.1.

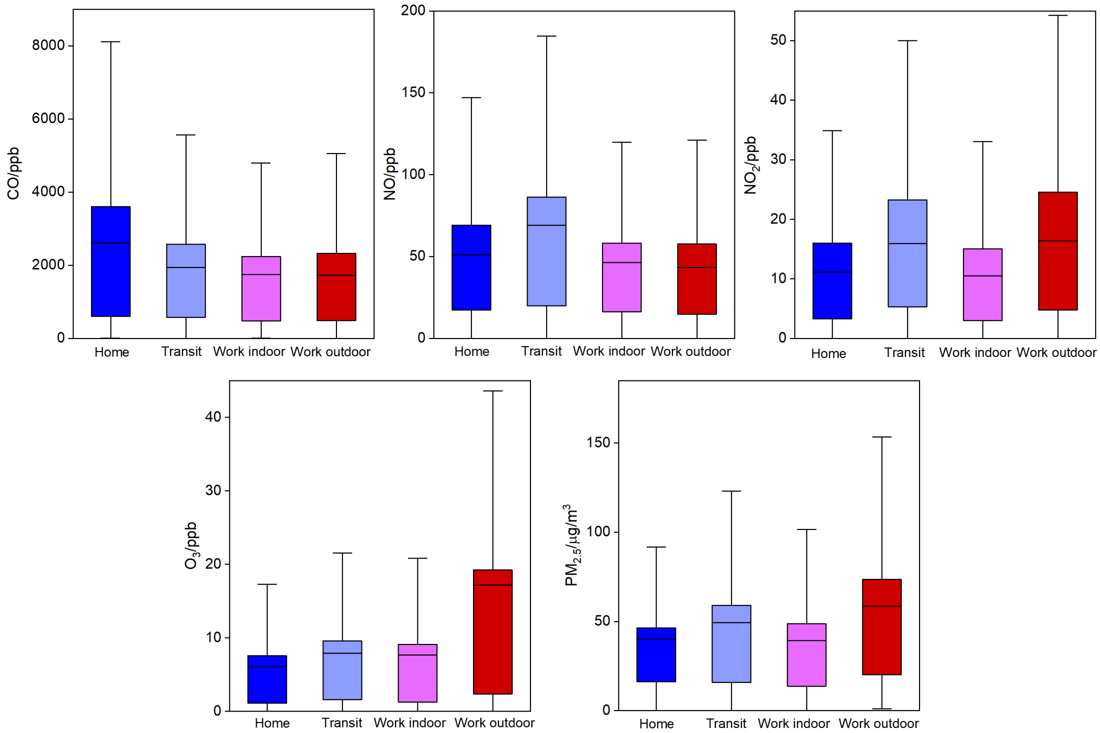

Figure 5.1: Mean concentration in different microenvironments: Box plots of averages over 12 hour periods for CO, NO, NO_2, O_3 and $PM_{2.5}$ exposure in four different microenvironments, calculated from personal exposure data recorded by the PAM. The boxes indicate the quartiles, and the whiskers indicate the minimum and maximum values.

China has published indoor and outdoor air quality standards, which can be found in Appendices A.5.3 and A.5.2. Due to the different averaging times of the exposure in this work and the averaging times of the standards, no direct comparisons are made.

For NO_2, O_3 and $PM_{2.5}$, the levels in the "work outdoor" microenvironments were generally higher than the other microenvironments. This suggests that the I/O ratios for these species would be less than one; the indoor pollutant loss processes dominate over the effects of indoor sources in these microenvironments. I/O ratios measured in China for these species have been found to be less than one, as shown in Appendix A.5.4.

The majority of individuals' time was spent at home (see Figure 3.7), and as a result, they inhale the greatest portion of all of the key pollutants while being at home (see Figure 5.2).

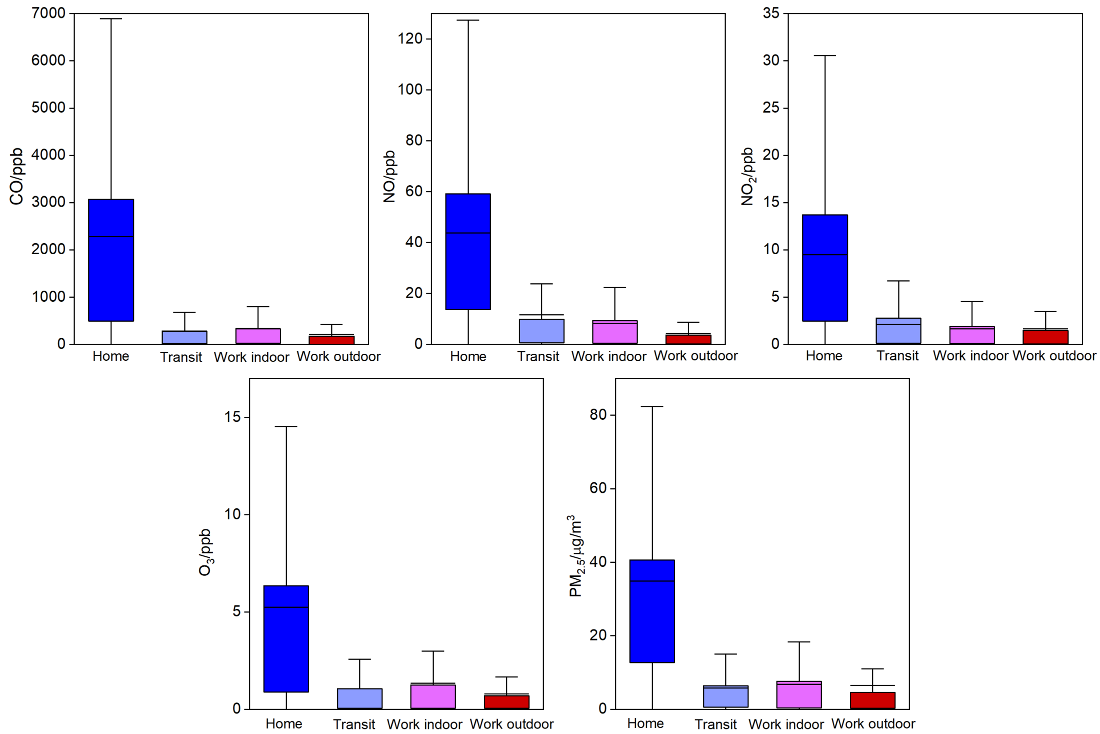

Figure 5.2: Time-weighted exposure in different microenvironments: Box plots of weighted averages over 12 hour periods for CO, NO, NO_2, O_3 and $PM_{2.5}$ exposure in four different microenvironments, calculated from personal exposure data recorded by the PAM. Averages are weighted using the time proportion spent in each microenvironment. The participants inhale the greatest portion of all of the key pollutants while being at home

Figure 5.3 compares the ambient levels measured by the reference instrument with the total personal exposure measured by the participants carrying the PAMs. Figure 5.3 confirms that using ambient pollutant levels as a proxy for personal exposure can result in exposure misclassification.

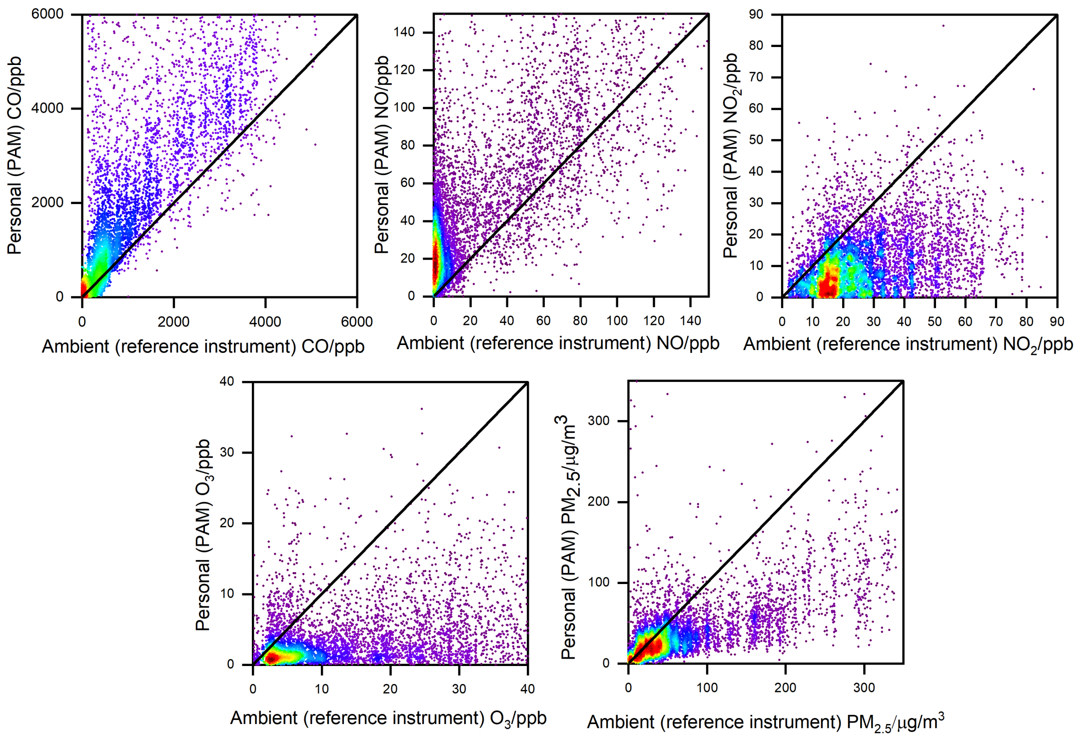

Figure 5.3: Comparison of ambient levels and personal exposure: Density scatter plots, where each point is the mean value over the time section (time sections in this report are defined as 12-hour periods, from 06:00-18:00 and then 18:00-06:00). On the y-axis are the values recorded by the PAM (the personal measurements) and on the x-axis are the values recorded by the reference instruments. The 1:1 line is shown in black. The association between ambient and personal exposure deviate from the 1:1 line in different ways for the different pollutants, as explained in the main text.

In Figure 5.3, for CO and NO, the majority of points are above the 1:1 line. The average personal exposure to ambient exposure ratios for these species are 1:0.51 and 1:0.65, respectively. As the participants spend most of their time at home (Figure 3.7), the difference between ambient reference instruments and personal measurements suggests that there are sources of CO and NO in the home environment. The participants breathe more CO and NO than assumed in the pollution-health models. This suggests that the toxicity of these pollutants may be currently being overestimated.

In Figure 5.3, for NO_2, O_3 and $PM_{2.5}$, the majority of points fall below the 1:1 line. The average personal exposure to ambient exposure ratios for these species are 1:182, 1:3.22, and 1:1.181, respectively. As the participants spend most of their time at home (Figure 3.7), the difference between ambient reference instruments and personal measurements suggests that there are indoor losses for these pollutants in the home environment. The indoor loss rates for these pollutants are quantified in

Section 5.2.2. The participants breathe less NO_2, O_3 and $PM_{2.5}$ than assumed in pollution-health models. This suggests that the toxicity of these pollutants may be currently being underestimated.

5.1.2 Pollutant correlations in personal exposure

As explained in Section 1.2.3, correlations between pollutants can lead to error in pollutant-health associations. The ambient measurements made by the reference instruments during the AIRLESS project show correlations and anti-correlations between pollutants, as shown in Figure 5.4.

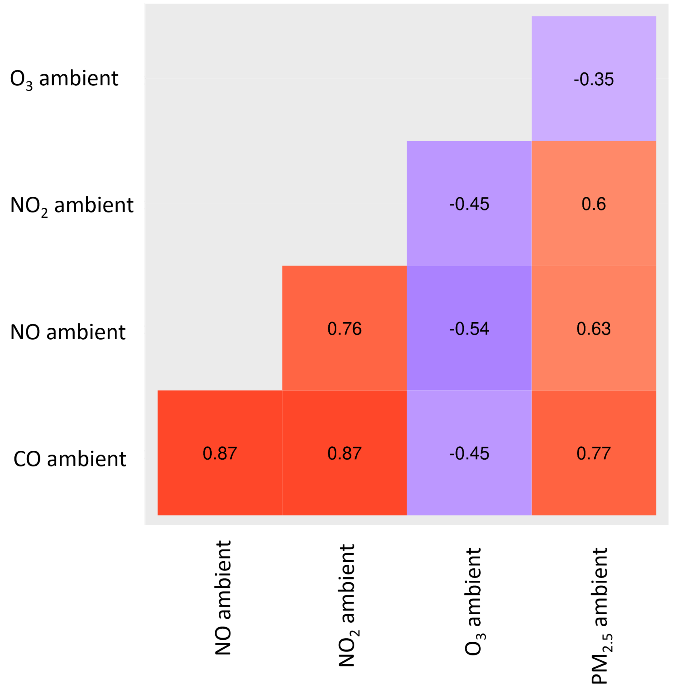

Figure 5.4: Ambient correlation matrix: Pearson's correlation coefficient values for 5 key species. Coefficients were calculated using 12-hour mean values of the reference data.

However, in the correlation matrix for personal exposure, there are much weaker correlations between pollutants. This may be a result of indoor sources and indoor loss processes for these pollutants.

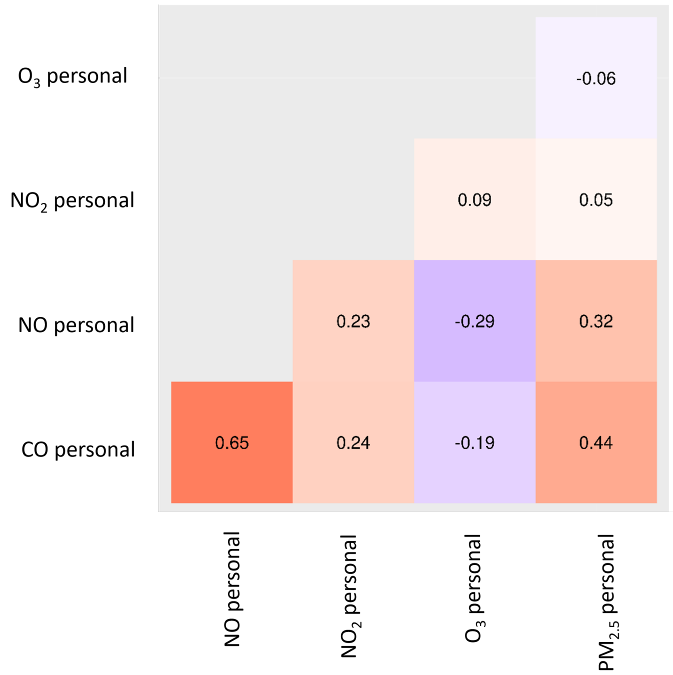

Figure 5.5: Personal exposure correlation matrix: Pearson's correlation coefficient values for 5 key species. Coefficients were calculated using 12-hour PAM mean values.

When considering ambient exposure, using a single pollutant coefficient for NO_2 and a single-pollutant coefficient for $PM_{2.5}$ and adding the results has been shown to overestimate the combined effects of the two pollutants[27 64]. As pollutant correlations are lower in personal exposure (Figure 5.5), personal exposure metrics would produce more reliable links between specific pollutants and their health outcomes.

5.1.3 Source-apportioned total personal exposure metrics

Indoor- and outdoor-generated exposure should be inputted separately into health models, as explained in Section 1.2.4. Apportioned exposure metrics, for example indoor-generated CO, while the same molecule, can act as proxy for a different mixture of air than outdoor-generated CO.

Box plots of weighted means show the results of the apportionment for the key species. The breakdown for season, location and time of day can be found in Appendix A.5.7.

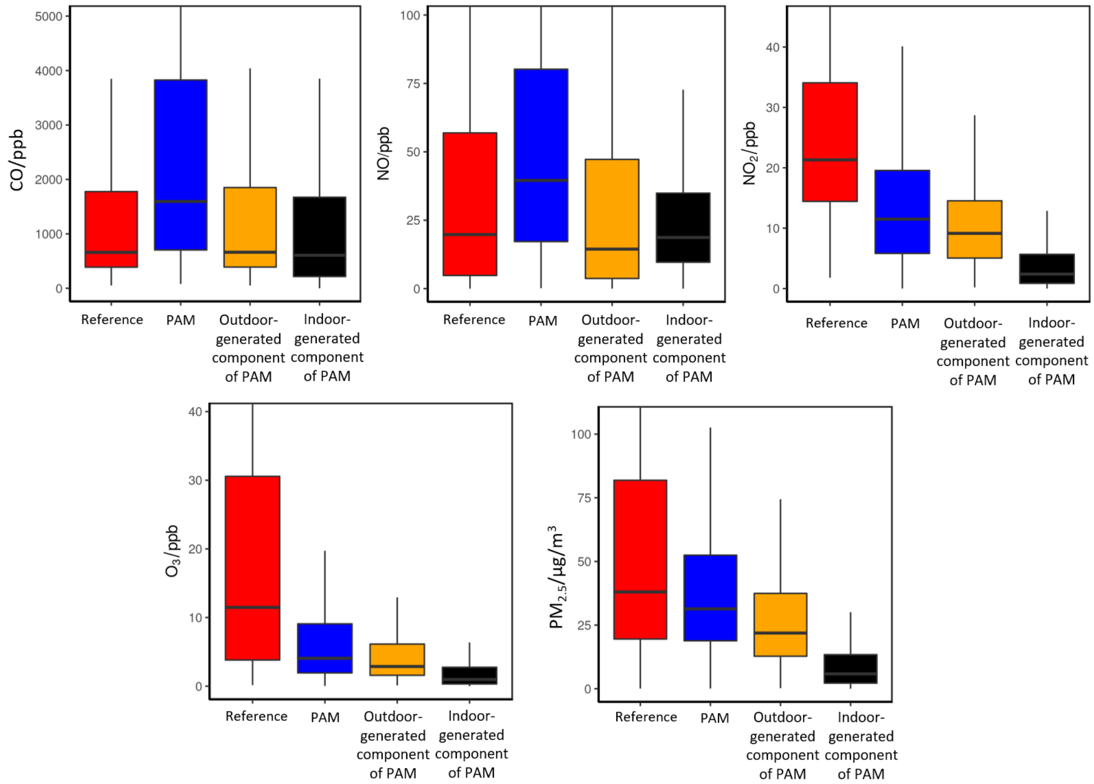

Figure 5.6: Exposure box plots: Reference (red), PAM (blue) exposure box plots, generated from 12-hour mean values. PAM exposure has been source-apportioned into outdoor-generated (orange) and indoor-generated (black) box plots.

Participants were generally exposed to more NO that was generated by indoor sources than outdoor sources, however for the other pollutants, the reverse is true. The exposure of participants to indoor-generated O_3 is very low, which is expected due to limited indoor O_3 sources.

The weighted exposure, by percentage, of the AIRLESS participants to the two components of exposure is shown in Table 5.1. A stacked column of this data can be found in Appendix can be found in Appendix A.5.5.

Percentage of indoor-generated component to total personal exposure	CO	NO	NO_2	O_3	$PM_{2.5}$
Overall	55.3	49.2	29.7	32.3	30
Winter	56.1	44.8	32.3	46.6	30.4
Summer	49.7	76.7	26.9	21.2	29.5
Pinggu	66.5	52.8	29.2	26	32
Beijing	41.3	46.2	30	38	24.9
Day	54.7	55.6	33.9	30.5	31.8
Night	55.8	43.1	24.8	35	28.1

Table 5.1: Percentages of indoor-generated exposure: A table of the percentages of total personal exposure that was generated by indoor sources. The percentages for specific seasons, locations and time of data are also included. This data is plotted above in Figure A.8.

The exposure percentages in Table 5.1 appear to show a relatively large percentage of O_3 exposure originating from indoor sources. This effect may arise due to the very low levels of O_3 recorded by the PAM, especially in the winter, as shown in Appendix A.12. Therefore absolute values of apportioned exposure may be more useful than percentages as metrics for O_3. Additionally, the RMSE of the O_3 sensor in the PAM is 2.7 ppb (see Table 3.2) which is larger than the mean (2.5 ppb), median (1.0 ppb) and IQR (2.4 ppb) of the estimated indoor-generated exposure. Therefore the observed large percentage of indoor-generated O_3 exposure may be a result of sensor error.

5.1.4 Source-apportioned $PM_{2.5}$

Source-apportionment is valuable when applied to $PM_{2.5}$, as explored in Section 1.2.4. Figure 5.7 shows the diurnal patterns of the reference measurements, the PAM exposure, and the two components of the PAM exposure. Diurnal plots for all species are found in Appendix A.5.6.

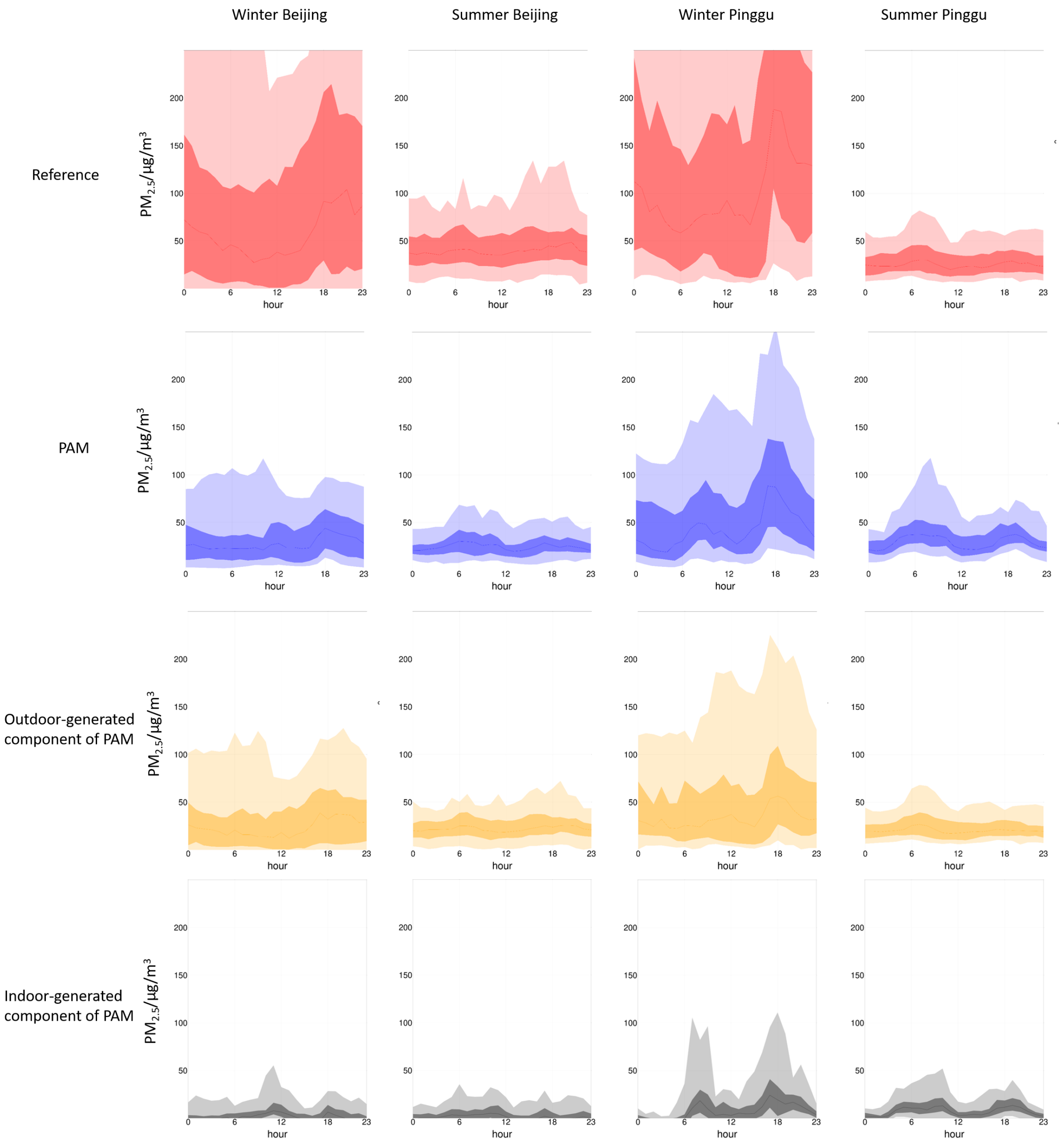

Figure 5.7: Diurnal plot of apportioned $PM_{2.5}$: Diurnal plots of the PAM data (blue) and its apportioned indoor-generated component (black) and outdoor-generated component (orange) of total exposure to $PM_{2.5}$, recorded by the PAMs, for the AIRLESS cohort. The diurnal plot of the outdoor levels recorded by the reference instrument (red) is also included. The plots display the median, 25th and 75th percentiles and 5th and 90^{th} percentiles. Negative indoor-generated values were removed before producing these plots.

From the $PM_{2.5}$ diurnal plots, for both seasons and locations, there appears to be an indoor source of $PM_{2.5}$ around 18:00, which may indicate the cooking of an evening meal. This is an expected feature in the indoor-generated component and is a positive sign that apportionment has been achieved. Additionally, for both seasons and locations, the outdoor-generated component appears to be a scaled version of the reference. This scaling is due to the indoor loss processes which are quantified in Section 5.2.2.

Although, to our knowledge, this is the first study to source-apportion total personal exposure, some of these same AIRLESS participants have had their $PM_{2.5}$ exposure in their bedrooms apportioned in a new study conducted by Zhang et. al[163]. This subset of the AIRLESS cohort were selected (N=71) using the following criteria:

- A subset of the Beijing residents who lived within 100m of a main road (N=39)
- A subset of the Pinggu residents depending on their primary cooking and heating methods for heterogeneity (N=32)

Their set-up involved placing $PM_{2.5}$ monitors inside the bedroom and outside the house of the participants for 72 hours. 14 houses were fitted with both indoor and outdoor monitors, however only the data from 12 of the houses were used to train the model due to malfunction of two of the instruments. The model was then tested on the other participants' data, using indoor measurements and reference instruments (acting as the outdoor measurements). The Zhang et al. methodology takes a different approach to isolating the indoor-generated portion of the indoor air. They classify points within sharp rising-edges and sharp falling-decays, where similar changes are not observed in the outdoor time series, as having indoor origin. Furthermore, an "absolute threshold" criteria of $4\mu g/m^3$ was necessary for classification of peaks of smaller magnitude which were not classified as indoor origin by the method above. Absolute thresholds have not been used in the methodological framework developed in this thesis as they may cause issues when applying this method to a dataset/ in countries with different ambient and indoor levels, and source characteristics.

They found that indoor-generated $PM_{2.5}$ contributed less to the levels measured inside the bedrooms than outdoor-generated $PM_{2.5}$ (between 6% and 19%). The method used in this thesis found higher contributions of indoor-generated to personal exposure (between 24.9% and 32%). The Zhang study averaged their data over 15 minute intervals which may mask short term indoor sources, and the indoor levels were exclusively measured in the bedroom. Both of these factors would be expected to give lower percentages of indoor-generated pollution and are likely responsible for the discrepancies between the values reported in the Zhang paper and the values reported in this thesis. However, both studies found higher contribution of indoor-generated $PM_{2.5}$ in the rural location, and both studies found a stronger contribution of indoor-generated $PM_{2.5}$ around traditional meal times.

5.2 Estimated values of exposure determinants

The estimation of the effect of the factors that determine exposure to air pollution, specifically ventilation rates, indoor loss rates and indoor emission events, and variations of these determinants with season, location (urban vs rural) and time of day, will be crucial in the future modelling of total personal exposure at the population scale. Applying these determinants to outdoor reference measurements (using the relationship described by the Continuity equation) will produce improved estimates of indoor exposure, and therefore total exposure of the five key pollutants. Additionally, the resultant modelled indoor levels could be used to infer the concentrations of products from indoor reactions.

As people spend most of their time at home (Figure 3.7), this section presents the estimation of these exposure determinants for the home microenvironment.

5.2.1 Ventilation in the home

Ventilation rates are estimated from the AIRLESS dataset using CO as a tracer gas via the constant decay method. The number of indoor emission events of CO, and therefore estimated ventilation rates, may be dependent on season or location. Figure 5.8 displays the number CO decays in the time series of the participant's whole week-long deployment from which a ventilation rate was estimated.

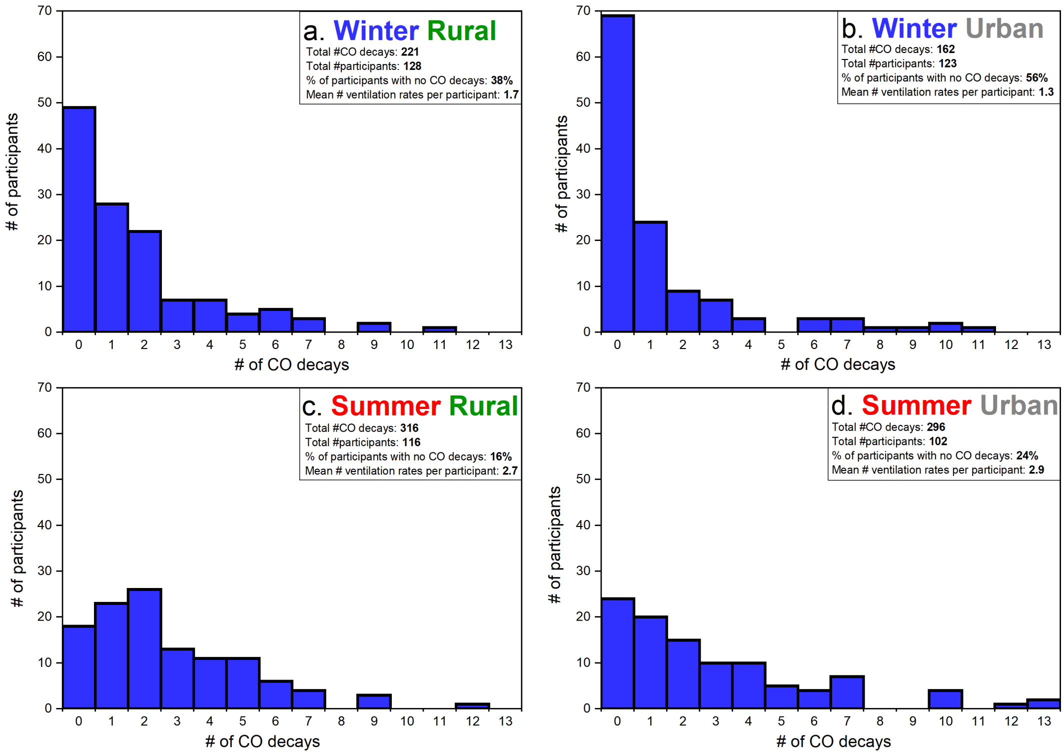

Figure 5.8: Quantity of estimated ventilation rate values: Histograms of the number CO decays in the time series of the participant's whole week-long deployment from which a ventilation rate was estimated. Only the ventilation rates estimated from CO decays while the participant was in home environment are included. The histograms are split by season and location. Key statistics are included on the plots. There are many participant time series from which no ventilation rates were estimated.

In winter there was a higher proportion of CO time series where no ventilation rates could be estimated. This may be due to the higher indoor level of CO in the winter due to heating of the homes, therefore leading to less defined peaks in the indoor CO time series which pass the QA/QC steps of constant decay method algorithm (Section 4.2.1).

Box plots containing the estimated ventilation rates for the home microenvironment can be found in Figure 5.9.

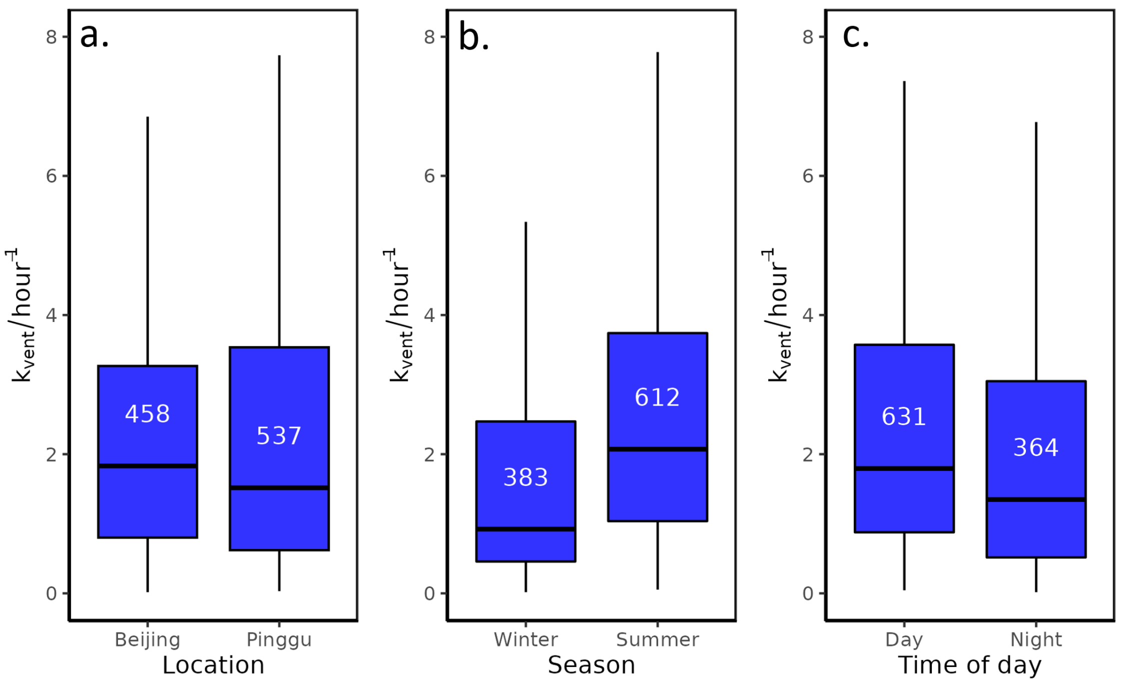

Figure 5.9: Home ventilation: Box plots of the estimated ventilation rates for the whole AIRLESS cohort, estimated when the participants were in the home. a.) split by location, b.) split by season, c.) split by time of day, with Day=06:00-18:00 and Night=18:00-06:00. The number of ventilation rates used to produce each box plot is shown on the plot in white writing.

The average ventilation rate estimated for the home environment for the AIRLESS cohort was 3.12 hr^{-1}. The median value was 1.68 hr^{-1}. The AIRLESS cohort adjusted doors and windows as they liked. The ventilation rates estimated sit between the window-open and window-closed measured ventilation rates measured in homes in China found in literature (Table 5.2), as would be expected. Ventilation rate data for indoor environments (particularly homes) in China are scarce[58;59].

A t-test indicated that the ventilation rates estimated in the summer months were significantly higher than those estimated in winter ($t = 2.662$, $p = 0.00396$). This is likely due to more opening of doors and windows during these periods. This is observed in the literature, as shown in Table 5.2.

A t-test indicated that the ventilation rates estimated in Beijing and Pinggu are not statistically significantly different ($t = 0.9618$, $p = 0.1682$), which is in agreement with the Hou et al.[59] who found no significant differences between ventilation in rural and urban homes in China.

In the questionnaire (see Section 3.3) the participants were asked about some of the characteristics of their homes. Appendix A.5.10 contains box plots of the estimated ventilation rates, separated by door and window features, and the floor of

Author and date	Number of residences	Season	Environment	Room type	Ventilation inference method	Mean or median air exchange rates (h^{-1})	Window opening
Hou (2019)	294	Spring, Summer, Autumn, Winter	Urban (cities across all 5 climate zones)	Bedroom, living room	Tracer gas (CO_2)	0.34 Report includes breakdown by season and climate zone.	No attempt to change occupant behaviour, window opening was recorded windows were closed 52%-82% of the time (depending on the season and climate zone).
Hou (2018)	399	Spring, Summer, Autumn, Winter	Urban (Tianjin) Rural (Cangzhou)	Bedroom, living room, children' bedroom	Tracer gas (CO_2)	Spring: 0.27 Summer: 1.11 Autumn: 0.29 Winter: 0.30 Report includes breakdown by room, urban/rural.	No attempt to change occupant behaviour, window opening was recorded. Study includes monthly breakdown of window opening frequencies.
Sun (2011)	348	Summer, Winter	Tianjin University campus	Student dorm room	Tracer gas (CO_2)	Summer: 4.7 Winter: 0.7	In summer, doors/windows of dorm rooms were fully opened. In winter, doors and windows were closed at night (1:00–8:00 am).
Huang (2017)	1	Summer	Rural (Zhaji village of Jingxian county)	Huizhou traditional dwellings, hall, bedroom	By measurement of wind speed, room air change rates were based on the airflow rates and airflow direction of different building openings.	Open windows doors hall: 15.54 Open windows doors bedroom: 10.88 Closed windows doors hall: 1.00 Closed windows hall bedroom: 0.65	Both windows open and windows closed conditions were investigated.
Shi (2015)	34	Autumn	Urban (Beijing)	Main bedroom, living room	Tracer gas (CO_2)	0.17	Window closed conditions.
Cheng (2018)	202	Autumn	Urban (Guangzhou)	Bedroom	Tracer gas (CO_2)	0.41	Window closed conditions.

Table 5.2: Ventilation values in literature from studies in China: A summary table of studies which have measured the ventilation rates of home environments in China. Studies included are Hou (2019)[59], Hou (2018)[58], Sun (2011)[129], Huang (2017)[62], Shi (2015)[122] and Cheng (2018)[24].

the building where the resident resides.

The air-tightness of buildings has significantly increased since the energy crisis in the early 1970's[90] and many studies report that it has increased further since the 1990's because air-tight buildings are more energy efficient and therefore economically favourable[101;87;19]. A ramification of this is that indoor air quality may deteriorate, compromising the health and comfort of building occupants[57]. In 2012, Wargocki[148] attempted to find out how much ventilation is needed in existing homes to reduce health risks by reviewing the published scientific literature investigating the association between measured ventilation rates and observed health problems. It was concluded that it is likely that health risks occur when ventilation rates are below 0.4 air changes per hour (h^{-1}) in existing homes, although it was noted that there are very few studies on this issue and many of them suffer from deficient experimental design, as well as a lack of proper characterisation of actual exposures occurring indoors.

The minimum required air change rates for residential buildings in China were published in 2012 by the National Standard of the People's Republic of China and are shown in Table 5.3.

Floor area range (m^2)	Air change rate (h^{-1})
Under 10	0.7
Between 10 and 20	0.6
Between 20 and 50	0.5
Over 50	0.45

Table 5.3: Chinese ventilation standards: Table of minimum required residential air change rates in China, published by the National Standard of the People's Republic of China[94]

The estimated ventilation rates for the AIRLESS cohort are generally above the minimum required rates for residential buildings in China. The majority of the residential building stock in China (around 50 billion m^2) do not use mechanical systems for ventilation, with opening windows and infiltration being the most common ventilation methods. The extent of self-reported window opening has even been used as a proxy for ventilation in a health study[144].

5.2.2 Indoor loss rates in the home

Indoor loss rates have been estimated from the I/O ratio method and are plotted in Figures 5.10 and 5.11.

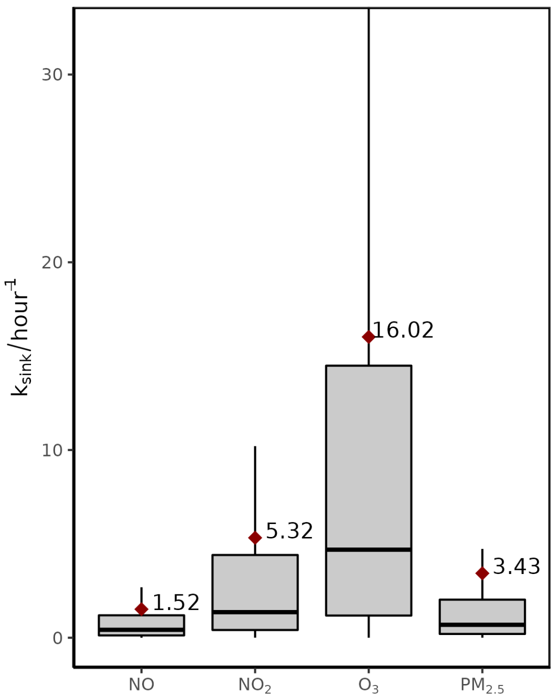

Figure 5.10: Indoor loss rates for reactive key pollutants: Box plots for the loss rates in the home environments of 4 key species: NO, NO_2, O_3 and $PM_{2.5}$. The red dot indicated the mean pollutant loss rate.

The estimated indoor pollutant loss rate values show that O_3 has the strongest indoor loss rate in the home environment, followed by NO_2, $PM_{2.5}$ and then NO.

There are limited loss rates recorded for these pollutants in homes in literature, and even less for homes in China. Rates recorded outside of China may not be applicable to homes in China, for example, about 70% of home floors in American residences are covered in carpet[143], compared with only 10% in Chinese homes[120] which would be expected to have a significant effect on loss rates via surfaces reactions and deposition.

The literature values of O_3 loss rates are also higher than those of the other species. Indoor O_3 loss rates in homes in western countries (mainly the USA) have been found to range between 2.8 h^{-1} to 7.2 h^{-1}[151]. A study in China recorded slightly lower loss rates between 1.3 h^{-1} to 6.0 h^{-1} across 14 residences[80]. O_3 is principally removed by surface reactions in the indoor environment, and a literature review has found that the surface-treated materials may have more impact than the underlying materials on ozone deposition[119].The indoor loss rate of NO_2 has been recorded in a small number of studies. In Chicago, IL, a closed window study estimated the loss rate of NO_2 to range from 0.06 h^{-1} – 0.47 h^{-1}, with an average value of 0.27 ± 0.12 h^{-1}[166]. The NO loss rate was recorded to be 0.04 h^{-1} ± 0.03 h^{-1}[166].

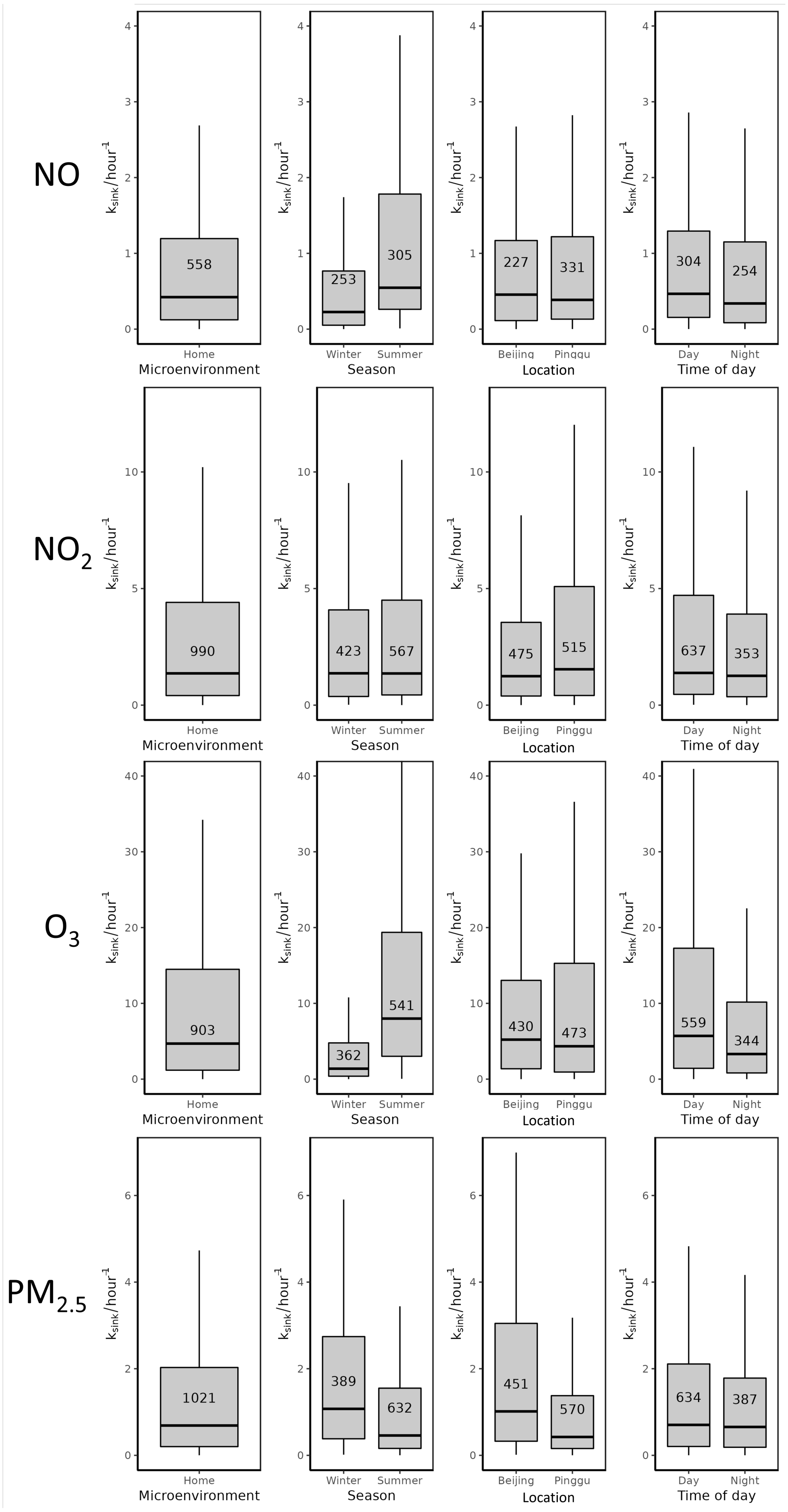

Figure 5.11: Indoor loss rates for reactive key pollutants: Box plots for the loss rates in the home environments of 4 key species: NO, NO_2, O_3 and $PM_{2.5}$. The variation between seasons, location and time of day are included.

When comparing the median and quartile values, the loss rates of the three gases are higher in the summer than in the winter. This is generally expected as most reactions that occur indoors proceed faster as the temperature increases[137].

When applying unpaired two-sample t-tests to investigate the differences in the means of the loss rates during the summer and winter, it was found that the increased loss rate for NO and O_3 in the summer is statistically significant (t=3.337, p-value=0.0009012 and t=6.4366, p-value=2.08×10^{-10} respectively). Conversely, when comparing the means of the loss rates of NO_2 between the summer and winter, the loss rate in the winter is higher and this is statistically significant (t=-2.6476, p-value=0.008332). The difference between the means for $PM_{2.5}$ was found to be statistically insignificant (t=-1.757, p-value=0.07923) at the 0.05 significance level.

The indoor pollutant loss rate values recorded in the summer for O_3 are notably higher than in the winter, when comparing both the median and mean. A higher ambient temperature (T = 30 °C) has been shown to increase the ozone removal rate in a chamber study by 3 times compared with a cooler ambient temperature (T = 20 °C)[137].

Conversely, the k_{sink} value for $PM_{2.5}$ is higher in the winter. This was also observed in homes in London, UK[141]. It is suggested that this is a result of decreased ventilation and airflow during the winter (faster deposition and less resuspension). Additionally, deposition rates of $PM_{2.5}$ have been reported to be higher at lower temperatures in a study of airborne biomass particles[162].

A second method to estimate k_{sink} was also outlined in Chapter 2. The constant decay method analysed the loss rate of an indoor emission event in the reactive species time series. The results of this method, compared to the I/O ratio method, for the home microenvironment, are shown in Figure 5.12.

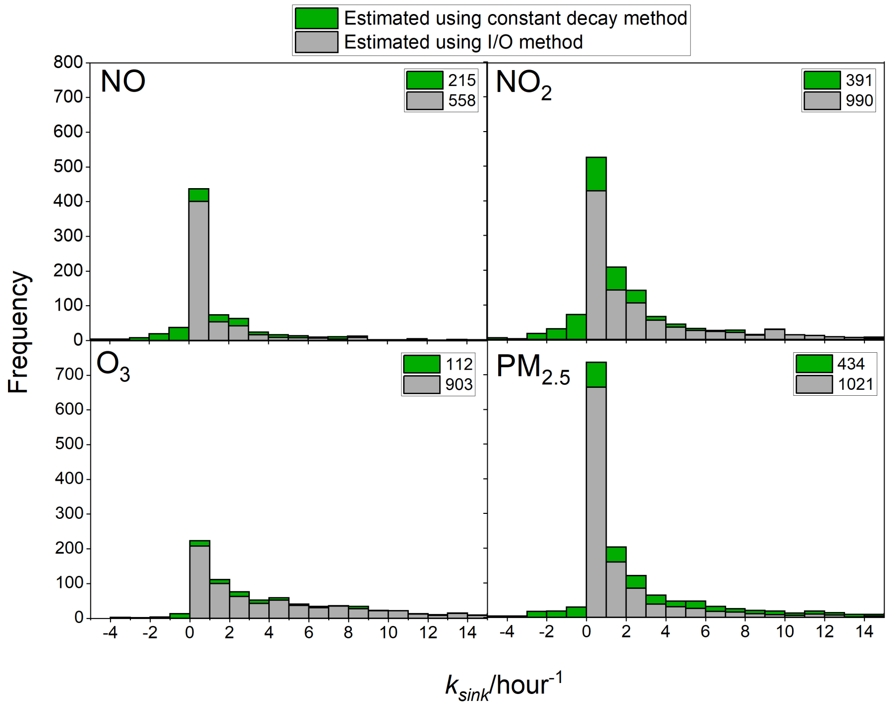

Figure 5.12: Comparison of methods of estimation of indoor loss rates: Stacked histograms showing the number of k_{sink} values, estimated for the home environment, from the AIRLESS dataset. The I/O ratio method is shown in grey and the constant decay method is shown in green. More k_{sink} values were estimated using the I/O ratio method.

The constant decay method produced less values of k_{sink}, although both methods agree on the most populous bin. The constant decay method for estimation of k_{sink} results in some negative loss rates. This occurs when the average CO decay rate of indoor emission events over the 12 hour time-section is larger than the decay rate of the indoor emission event in the reactive pollutant time series. This was demonstrated for the example participant in Section 4.4.2. The means of the indoor loss rates using the two methods are not significantly similar (p = 0.0036). As part of future work, the statistical analysis of differences in the k_{sink} values from the two methods, within the same time section, could be conducted using a Bland-Altman plot. However, as both methods use the same k_{vent} value, care should be taken to avoid false similarities between the two methods.

Automated estimation of the pollutant loss rates inside homes is key in the modelling total personal exposure for populations. These loss rates result in significant differences between indoor and outdoor concentrations of the reactive pollutants.

5.2.3 Characterisation of indoor emission events in the home

Section 4.5 introduced a methodology to characterise indoor emission events in indoor-generated time series. Indoor emission events in the home can be used to make inferences about participant behavioural patterns and lifestyles. This valuable information will be crucial in the future modelling of total personal exposure at the population scale. Additionally, characteristics of the emission events may be linked to health outcomes.

Metrics of characteristics (quantity and heights of peaks, duration and area under) of the indoor emission events were estimated from the AIRLESS dataset. The correlations between these three metrics are shown in Appendix A.5.8.

Lifestyle inferences can be made from the characteristics of indoor emissions. Below is a demonstration of how the height of the peaks of indoor emission events in the home may be able to indicate the cooking fuel that participants use.

In Section 4.5.2, methodology to extract the mean peak height of indoor emission events was introduced. This methodology has been applied to the whole AIRLESS cohort. Presented here are the mean peak heights of indoor emission events of CO, NO_2 and $PM_{2.5}$ (defined using their 90^{th} percentile values, calculated from the whole AIRLESS cohort, as thresholds) whilst the participant was in the home, for each 12-hour time section. The 90^{th} percentile thresholds for CO, NO_2 and $PM_{2.5}$ are 3485 ppb, 17.61 ppb and 39.99 $\mu g/m^3$ respectively.

scatter plots to explore associations between the mean peak heights of indoor emission events of these 3 pollutants are shown in Figure 5.13, split by location (Pinggu and Beijing).

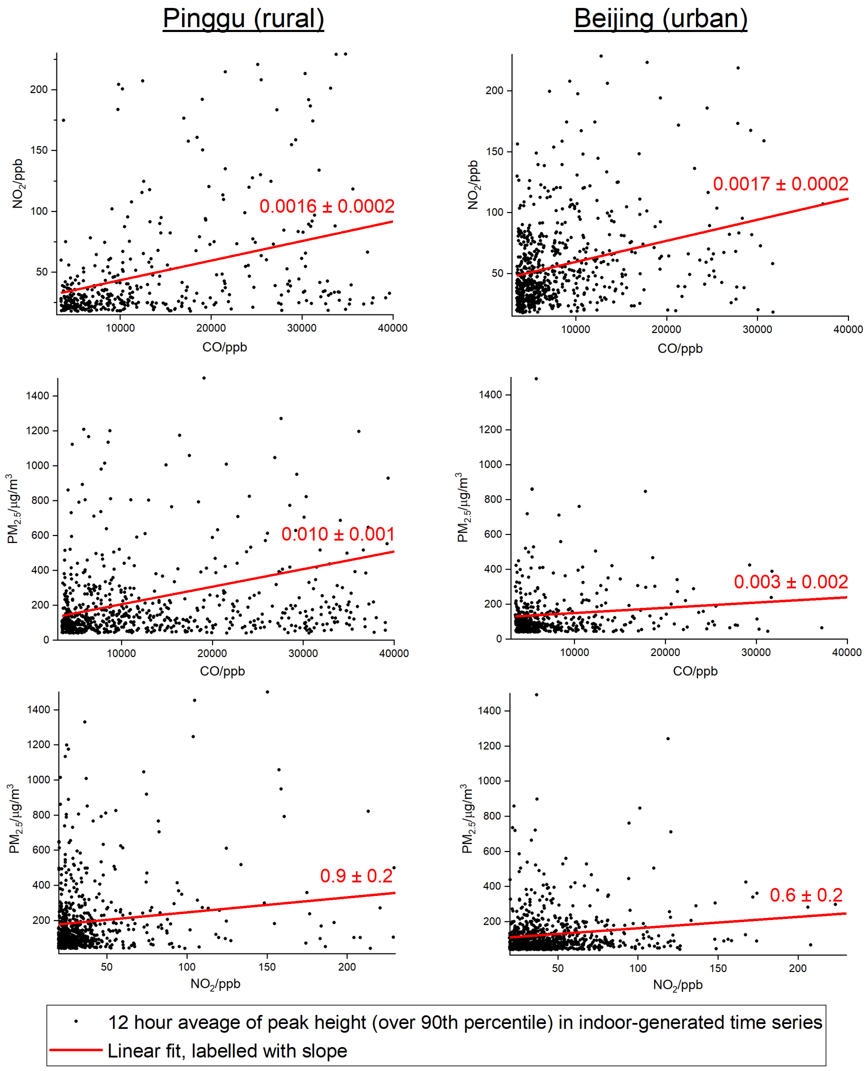

Figure 5.13: Peak heights of indoor emission events: scatter plots, where each point is the mean peak height of indoor emission event for a 12-hour time section. Plots show associations between 3 pollutants: CO, NO_2 and $PM_{2.5}$. The lines of best fit is shown in red, along with the gradient and standard error of the gradient. Data is plotted for the Pinggu and Beijing cohorts separately.

The CO:$PM_{2.5}$ ratio of the mean peak heights of indoor emission events in the indoor-generated data differs between locations; the slope between CO and $PM_{2.5}$ is around 3 times steeper for the Pinggu cohort.

A major short-lived source of $PM_{2.5}$ pollution in the home is from cooking. Cooking fuel, and the high temperatures in excess of 200°C that are required for some traditional Chinese cooking methods, such as stir-, pan-, and deepfrying, generate over 300 reaction products[167].

The differences in this ratio suggests that cooking produces a different mixture of pollutants between the urban and rural residents, possibly due to differences in the cooking fuel used between cohorts. As Figure 3.3 showed, the majority of Beijing residents used natural gas as their cooking fuel, however, over half of Pinggu residents used LPG. The same data from Figure 5.13 is now split by the fuel used by the participants when cooking, and the results are shown in Figure 5.14.

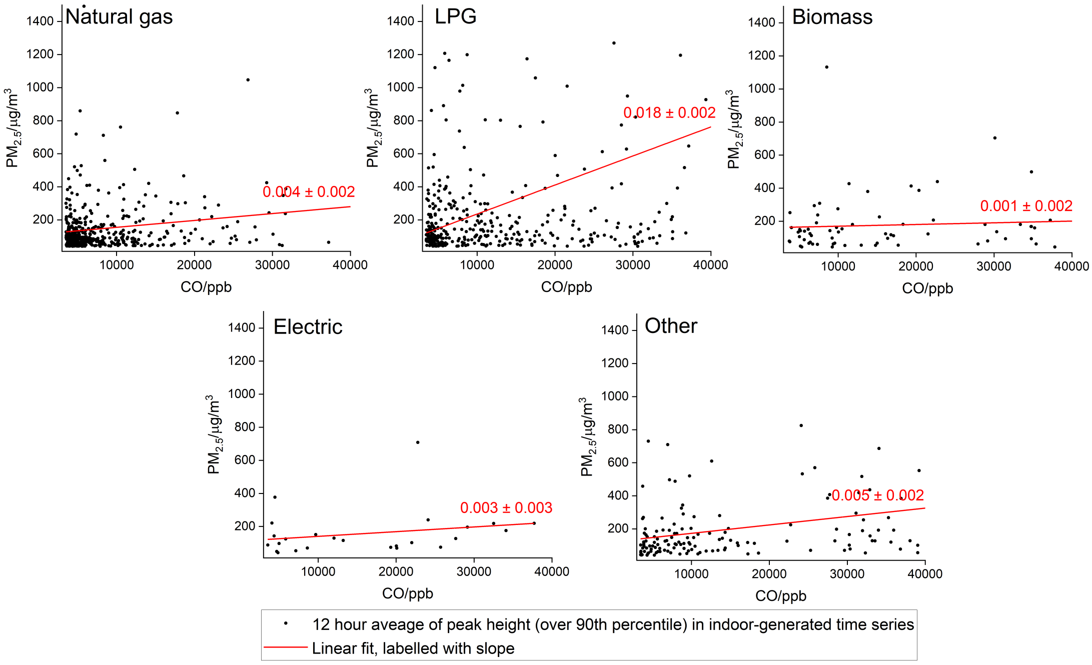

Figure 5.14: CO and $PM_{2.5}$ peak heights of indoor emission events for different cooking fuel types: scatter plots, where each point is the mean peak height of indoor emission event for a 12-hour time section. Plots show associations between CO and $PM_{2.5}$ for all AIRLESS participants, separated by the cooking fuel type used by the participant. The lines of best fit is shown in red, along with the gradient and standard error of the gradient.

The strongest linear relationship between the CO and $PM_{2.5}$ indoor emission event heights was found in the LPG users, possibly because LPG burns at a higher temperature than natural gas. A recent systematic review has found that the temperature used for cooking is positively correlated with the PM concentration[74]. In China, the population has been shifting (and is projected to continue shifting) to clean cooking fuels[128], as explored in Appendix A.5.9.

5.3 Chapter Summary

This chapter displayed the capabilities of the developed automated framework to source-apportion personal exposure and estimate values of the exposure determinants that affect personal exposure in the AIRLESS dataset. Estimated values are comparable with those in literature. The methodology can be applied to data recorded by hundreds of participants during PAM deployments worldwide, to estimate these values without the need for deploying stationary monitors inside and outside of buildings.

Chapter 6

Linear mixed-effects modelling for health associations

Chapter 5 derived the estimated values of novel metrics (source-apportioned personal exposure and exposure determinants) from the application of the methodology outlined in Chapter 2 to the AIRLESS dataset. This chapter will demonstrate how linear mixed effects-models (LMEMs) can be constructed to make associations between the metrics estimated in Chapter 5 and health endpoints. It is expected that the results from such models can provide insight into the factors which drive the observed health responses.

The effect of these novel metrics on Peak Expiratory Flow (PEF) will be investigated. PEF is a widely used technique to measure lung function as it can be measured by the participant in their home. Although other health endpoints were collected during the AIRLESS project (Table 3.4), Article 28 of the 2021 Chinese Personal Information Protection Law states that personal health information is considered sensitive data. Therefore, the presentation and analysis of the other health parameters collected during the AIRLESS project are restricted.

6.1 LMEMs in epidemiological research

Linear mixed-effects models (LMEMs) have become a standard tool for investigating associations between air pollution and health, as they can account for dependencies in data[23;168;98;44].

It is expected that there are dependencies within the AIRLESS dataset. The AIR-

LESS study was a longitudinal study, where the individual's PEF values were measured over time (the morning of every day that they carried a PAM). Detailed information on the AIRLESS dataset can be found in Chapter 3. In this dataset, the PEF values from the same participant may not be independent; a PEF measurement from one participant is expected to be more similar to another measurement from the same participant, than to one from a different participant. Ignoring dependencies results in overestimating sample size and, therefore, misleadingly small standard errors. This artificially increases confidence in the outcome coefficients, which can result in a Type I error (incorrectly rejecting the null hypothesis)[100].

In LMEMs, fixed effects are used to capture the systematic and population-level factors that are expected to have an impact on health markers. In this case, PEF is expected to be impacted by factors such as age, sex and pollution levels.

Random effects can capture the variability in PEF data that is not explained by fixed effects but is specific to the participant.

6.2 AIRLESS dataset structure

If the PEF values are dependant on the participant, then the AIRLESS dataset can be described as a two-level data structure, with the participants at the higher level (level 2) and the PEF measurements at the lower level (level 1). The two levels of this dataset are shown in the unit diagram in Figure 6.1, along with some of the covariates that were measured at each level of the dataset. The dependence of PEF on the individual participants is mathematically confirmed later in Section 6.4.

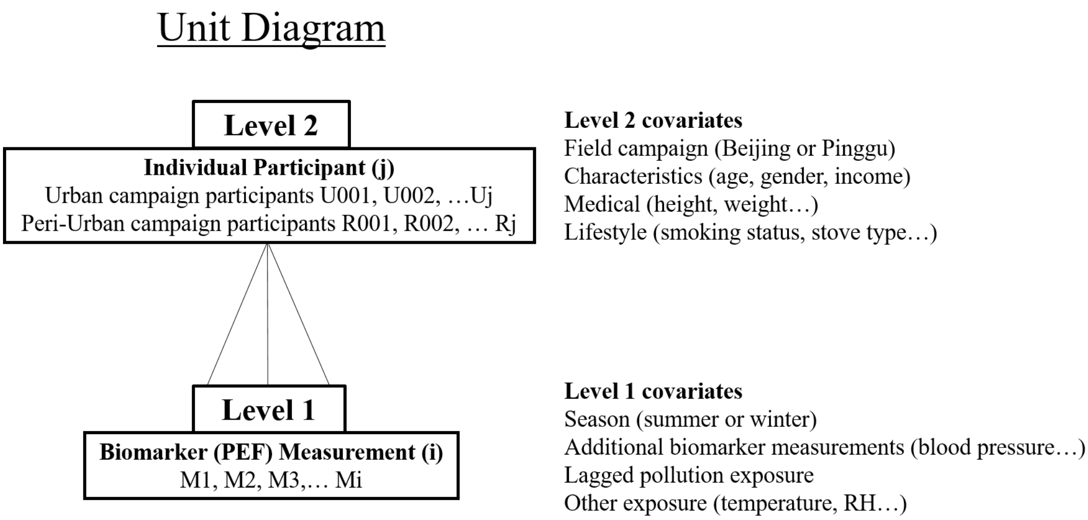

Figure 6.1: AIRLESS data structure: Unit diagram for the AIRLESS dataset, including some of the covariates recorded at the two levels. This is the proposed unit diagram, assuming that PEF measurements are dependent on participant, which is mathematically confirmed later in this chapter.

6.3 Model 1: Empty (linear regression) model

Model 1 fits the PEF data from the AIRLESS dataset as a single-level "empty" (linear regression) model.

$$PEF_{ij} = \beta_0 + r_{ij} \tag{6.1}$$

Where:

- PEF_{ij} refers to a uniquely identified PEF measurement
- β_0 is the overall intercept. For the empty model, this is the mean of all PEF measurements
- r_{ij} is the residual

For this empty model, β_0 is the mean of the PEF measurements. The deviance was calculated using the formula:

$$deviance = -2 \times \; log - likelihood \tag{6.2}$$

Where log-likelihood is a measure of the model's fit.

The mean PEF value recorded was 385 L/min, as shown in Table 6.1.

	Parameter	Model 1
β_0	Intercept	385.055 ± 1.971
σ_e^2	Measurement variance	10784.72
	Deviance	33642.44

Table 6.1: Model 1 results: A table containing the model outcome statistics from Model 1. PEF is measured in units of L/min.

6.4 Model 2: Variance-component model

To test whether the PEF values are dependent on the participant, i.e. this dataset has two levels (with PEF measurement at level 1 and participant at level 2 as proposed in Figure 6.1), a variance-component model is constructed:

$$PEF_{ij} = \beta_0 + \mu_j + e_{ij} \tag{6.3}$$

Where:

- β_0 is the overall intercept, although this is not necessarily the same as the mean PEF value calculated at the offset. The overall intercept in the variance component model averages the participant means proportional to the group size (to the number of measurements per participant).
- μ_j is the between-participant residual, accounting for variation in PEF between participants
- e_{ij} is the within-participant residual, accounting for variation in PEF within a single participants measurements

Table 6.2 shows the results of the empty model (Model 1) and the variance-component model (Model 2) testing for clustering at the participant level (level 2) for 250 participants in Pinggu and Beijing in winter and summer.

	Parameter	Model 1	Model 2
β_0	Intercept	385.055 ± 1.971	385.366 ± 5.615
σ_μ^2	Participant variance	-	7052
σ_e^2	Measurement variance	10784.72	3860
	Deviance	33642.44	31516.52

Table 6.2: Model 2 results: Table containing the results of Model 2 (the variance-component model). Model 1 is included for comparison. PEF is measured in units of L/min.

This model shows that around two thirds of the overall PEF variation lies between participants. Model 2 is compared to Model 1 (an empty linear regression model that does not take into account two levels of the dataset) to see if it is significantly better at explaining the dataset. This can be assessed by looking at the reduction in deviance (D) by calculating the likelihood ratio (LR):

$$LR = D1 - D2 \tag{6.4}$$

In this case, the LR is calculated to be 2125.92. The resultant LR value is compared to a chi-squared distribution. The critical value for testing is at the 5% level with one degree of freedom, taking a value of 3.84 ($p>0.001$). The LR statistic greatly exceeds this value and so there is strong evidence for between-participant differences, therefore, a two-level mixed-effects model is necessary with participant ID as a random effect.

6.5 Model 3: Random-intercept model with gender as a fixed effect

The variance-component model (Model 2) can explain much of the variation in the response variable, however, inserting covariates (explanatory variables) in the model may explain the variation further.

The random-intercept model has two parts. The **fixed part** includes the overall model intercept (β_0), as well as the slope between the explanatory variable and the response variable (β_1) multiplied by the explanatory variable. The **random part** ($\mu_j + e_{ij}$) is random in the sense that these residuals are able to vary to find the optimum intercept for each participant. In other words, intercepts for each participant are allowed to vary; however, the gradient of the slopes (the effect of air pollution on health) is the same across all participants. The parameters that are estimated in the random intercept model are coefficients (β_0 and β_1) for the fixed part and variances (σ^2_μ and σ^2_e) for the random part.

The variance-component model (Model 2) showed that 64% of the variation in PEF is explained by differences between participants. However, it is well known that the explanatory variable gender has a large effect on the response variable PEF[81]. The 64% variance at the participant level could partly be explained by the participants' gender. As gender was measured as a binary variable, "female" will act as the reference category (baseline) and will be given a value of 0 (whereas male will be given a value of 1).

Model 3 is:

$$PEF_{ij} = \underbrace{\beta_0 + \beta_1 \mathbf{male}_{ij}}_{fixed\,part} \underbrace{+\mu_j + e_{ij}}_{random\,part} \tag{6.5}$$

Where:

- β_0 is the overall intercept. Each group (participant) has a line with the same gradient (β_1) fitted to their data points but the overall intercept varies for each of the lines. β_0 is the mean of the intercepts when all other covariates (in this case gender) are set to their reference categories. Since "female" is the reference category, the overall intercept corresponds to the average baseline value of the response variable when considering only females.
- β_1 is the overall slope coefficient for gender (the gradient between PEF and male)
- μ_j is the between-participant residual, accounting for variation in PEF between participants (the distance between the group (participant) line and the overall line)
- e_{ij} is the within-participant residual, accounting for variation in PEF within a single participants measurements due to changes over time (the distance between the individual data-point and the line for that participant's data points)

The results are shown in Table 6.3.

	Parameter	Model 1	Model 2	Model 3
β_0	Intercept	385.055 ± 1.971	385.366 ± 5.615	344.296 ± 6.260
β_1	Gender_Male	-	-	94.293 ± 9.486
σ_μ^2	Participant variance	-	7052	4843
σ_e^2	Measurement variance	10784.72	3860	3862
	Deviance	33642.44	31516.52	31434.3

Table 6.3: Model 3 results: A table containing the results of Model 3 (random-intercept model with gender as a fixed effect). Results from Models 2 and 3 are included. PEF is measured in units of L/min.

As expected, on average men's PEF value is higher than women's; the intercept on gender is found to be 94.3 L/min.

To determine whether the effect of gender is statistically significant, the z-ratio for gender is compared to a standard normal distribution:

$$z = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{94.293}{9.486} = 9.940 \tag{6.6}$$

The z-test has a critical value for testing at the 5% level of ±1.96. The calculated ratio exceeds this value and so participant gender has a significant effect on PEF. Additionally, now that gender has been accounted for as a fixed effect in the model, the overall intercept has decreased from 385.366 L/min to 344.296 L/min. After introducing the fixed effect of gender, the baseline level of the response variable (which corresponds to the reference category, female) decreases. This means that the average value of the response variable for females is lower when considering the influence of gender compared to the baseline model without gender as a fixed effect.

6.6 Model 4: Random-intercept model with multiple fixed effects

Multiple fixed effects can be added simultaneously to the fixed part of the model (Model 4). The overall slope coefficient between the additional covariates (n) and the response variable is denoted as β_n. The overall intercept (β_0) is now the mean of the intercepts when ALL covariates are set to their reference categories. Level 1 covariates (PEF measurement level, e.g. daily temperature) and level 2 covariates (participant level, e.g. gender) are included as fixed effects in a random-intercept model in the same way. The random part of the model remains the same.

For this illustrative case, the selection of the covariates was driven by theory and included factors known to affect PEF in "Reference Values and Related Factors for Peak Expiratory Flow in Middle-Aged and Elderly Chinese" by C Ji[69].

Temperature and relative humidity were entered as continuous variables, with lags matching the lags of the pollutant exposure metrics (see Section 6.8).

Fixed covariate name	Level	Data type	Reference	Notes
FieldCampaign_Pinggu	2	Dummy	Beijing	
Age	2	Continuous	0 years	
Gender_Male	2	Dummy	Female	
Income	2	Ordinal	<2500RMB	Participant's total household income over the past year
Education	2	Ordinal	No schooling	Highest level of education
PastSmoking_PastSmoker	2	Dummy	Never smoked	Evaluation of participant's past circumstances in terms of tobacco smoking (all participants were current non-smokers at time of study)
SecondHandSmoke	2	Categorical	Never	Whether participant has resided with a smoker for over 6 months
CookingStoveType	2	Categorical	Gas	Type of cooking stove primarily used in participant's household
Season_Summer	1	Dummy	Winter	
Temperature_2dayav	1	Continuous	0	Average of the temperature recorded by the PAM, 2 days prior to the PEF measurement
Relative_humidity_2dayav	1	Continuous	0	Average of the relative humidity recorded by the PAM over 2 days prior to the PEF measurement

Table 6.4: Fixed effects: A table of the fixed effects included in the random-intercept model.

The results from Model 4 are shown in Table 6.5.

	Parameter	Model 1	Model 2	Model 4		
β_0	Intercept	385.055 ± 1.971	385.366 ± 5.615	395.251	±	80.378
β_1	FieldCampaign_Pinggu	-	-	17.156	±	21.114
β_2	Age	-	-	-1.356	±	1.013
β_3	Gender_Male	-	-	90.250***	±	12.596
β_6	Income (2500-4999RMB)	-	-	34.467	±	29.516
β_7	Income (5000-9999RMB)	-	-	-5.490	±	25.196
β_8	Income (10000-19999RMB)	-	-	-0.512	±	25.181
β_9	Income (20000-49999RMB)	-	-	7.050	±	23.731
β_{10}	Income (50000-99999RMB)	-	-	26.137	±	27.460
β_{11}	Income (>100000RMB)	-	-	18.330	±	27.118
β_{12}	Education (Primary schooling)	-	-	14.949	±	24.676
β_{13}	Education (Middle/high schooling)	-	-	11.834	±	24.399
β_{14}	Education (College and beyond)	-	-	16.708	±	30.458
β_{15}	PastSmoking_PastSmoker	-	-	-6.699	±	14.472
β_{16}	SecondHandSmoke (in the past)	-	-	-24.858*	±	12.050
β_{17}	SecondHandSmoke (Currently)	-	-	-5.541	±	13.359
β_{18}	CookingStoveType (Electric)	-	-	55.251*	±	-17.788
β_{19}	CookingStoveType (Chimney-fixed stove (using biomass or coal))	-	-	-17.788	±	25.194
β_{20}	CookingStoveType (other)	-	-	-39.871	±	36.643
β_{21}	Season_Summer	-	-	-9.206	±	4.785
β_{22}	Temperature_2dayav	-	-	0.579	±	0.361
β_{23}	Relative_humidity_2dayav	-	-	0.128	±	0.165
σ_μ^2	Participant variance	-	7052	4306		
σ_e^2	Measurement variance	10784.72	3860	3956		
	Deviance	36135.06	33642.44	28968.9		

Table 6.5: Model 4 results: A table of the results of Model 4 (the random-intercept model with multiple fixed effects). Results of Models 1 and 2 are included. PEF is measured in units of L/min. * denotes $p<0.05$. *** denotes $p<0.001$

The results from Model 4 are shown graphically in Figure 6.2. The estimated changes in Figure 6.2 are plotted with 95% confidence intervals. This is to show the uncertainty in the estimates. A 95% confidence interval is the interval commonly used by statisticians and indicates that there is a 95% chance that the true effect lies within the interval. If the interval includes zero, then it suggests that the impact of the fixed effect on PEF is not different enough from zero to be confident that it reflects a real effect (the effect on PEF is not statistically significant).

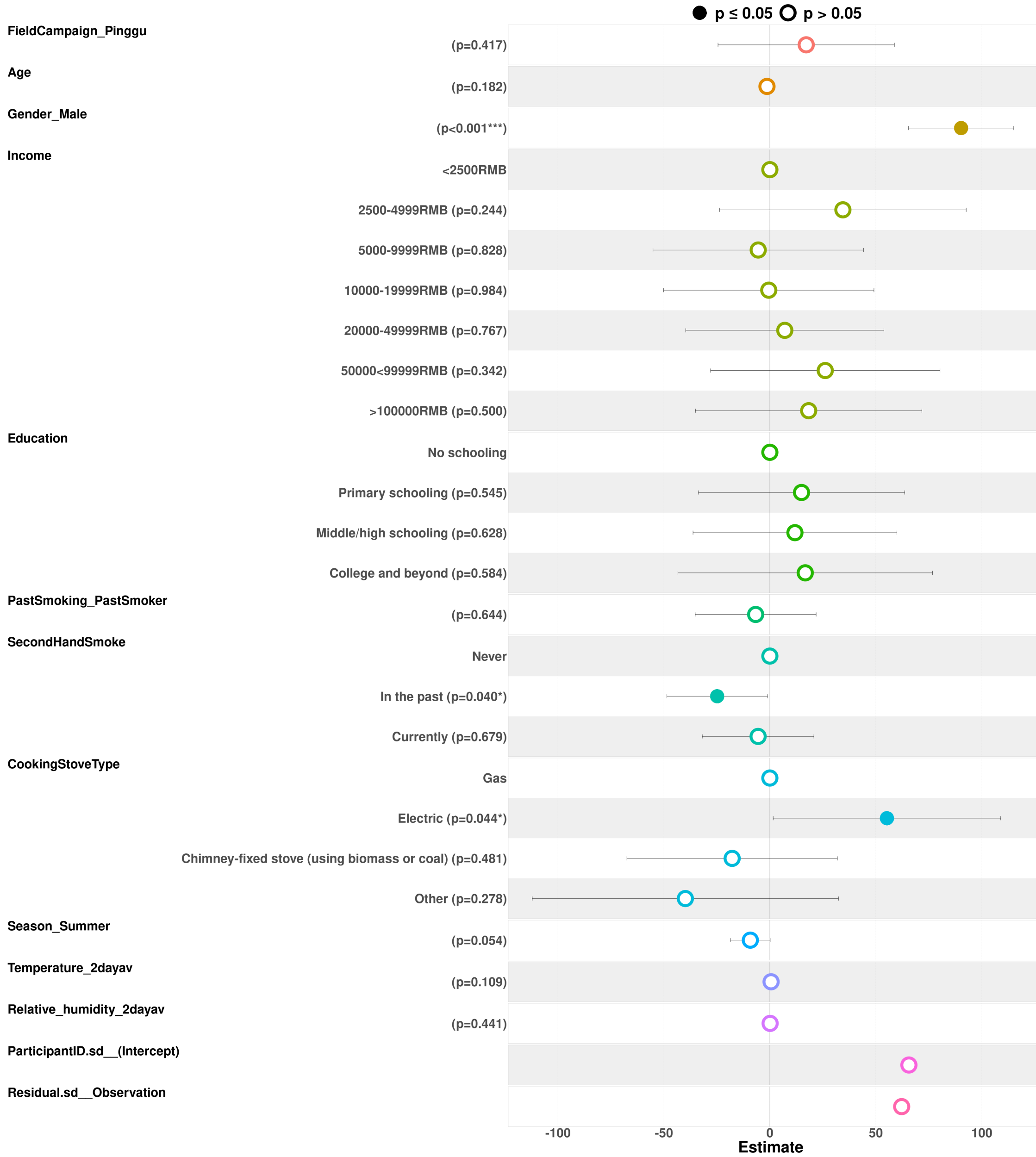

Figure 6.2: Model 4 fixed effects: Estimated fixed-effect covariates, plotted with 95% confidence intervals.

Gender, past second-hand smoke exposure (reducing lung function), and using an electric stove (improving lung function) were found to have statistically significant effects on PEF. These associations have also been found in the literature[165;69]. However, this does not mean that the fixed effects included in the LMEM should be limited to these fixed effects. Selection criteria of fixed effects for mixed-effects modelling is an active research topic amongst statisticians. Ultimately, the decision of which fixed effects to include should be based on a combination of statistical significance, model complexity, and subject-matter expertise. As all of these covariates have demonstrated an effect on PEF in previous research[69], and the deviance of Model 4 is lower than that of Model 2, all covariates will be retained in for the single-pollutant model in Section 6.8.

6.7 Random-slope model

A random-slope model could be used to capture the trajectories of a participant's PEF scores over time (i.e. over the week). This is not used in this case because random-slope mixed-effects models are more prone to over-fitting compared to random-intercept models when the sample size is limited as in this case. This study is limited to daily PEF measurements over two week deployments in a sample of 250.

6.8 Model 5: Linear mixed-effects model for pollution metric-health associations

A range of pollution metrics will be associated with PEF. These metrics are incorporated one at a time in the same way as fixed effects in Model 4. The methodology to derive these metrics can be found in Chapters 2 and 4, and values of these metrics for the AIRLESS cohort can be found in Chapter 5.

Exposure metrics (for CO, NO, NO_2, O_3 and $PM_{2.5}$) :

- Ambient exposure
- Personal exposure
- Indoor-generated exposure
- Outdoor-generated exposure

Exposure determinant metrics:

- Ventilation rate in the home
- Indoor pollutant loss rate in the home (NO, NO_2, O_3 and $PM_{2.5}$)

- Number of emission events in the home (CO, NO, NO_2, O_3 and $PM_{2.5}$)
- Area under emission events in the home (CO, NO, NO_2, O_3 and $PM_{2.5}$)
- Duration of emission events in the home (CO, NO, NO_2, O_3 and $PM_{2.5}$)

Lag

Pollution-health associations are normally evaluated using pollution exposure on the same day or those within a few previous days. The pollution exposure metrics are inputted into the single-pollutant model as different time averages to explore the delayed effects of pollution on health. The lags explored here are as shown by the timeline below. The same lag is used for temperature and humidity. As values of the exposure determinant metrics were not measured every day, the delayed effect of these metrics is not investigated in this thesis.

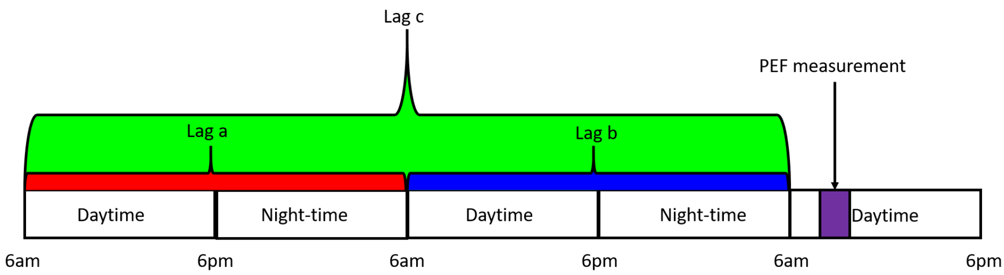

Figure 6.3: Lag: Timeline showing how the explanatory variable averages were calculated for this demonstrative case.

Presentation of results

Two ways of presenting the results of Model 5 have been considered.

The first presents the estimated change in PEF (L/min) associated with an interquartile range (IRQ) increase in the exposure metric. Changes in PEF with IQR increases are presented for all metrics.

The second presents the estimated PEF change (L/min) associated with a specified unit increase in the metric, which will only be used to present the influence on PEF of the exposure determinant metrics (Figures 6.4 and 6.6). For this work, the specified unit increases of the exposure metrics are 1000 ppb increase for CO, a 10 ppb increase for NO, NO_2 and O_3, and a 10 $\mu g/m^3$ increase for $PM_{2.5}$. The specified unit increase for CO is larger, so the results can be compared on the same axis as the other pollutants. The benefit of using specified unit increases is that it provides a straightforward interpretation of how changing pollutant exposure can affect PEF, particularly useful for policymakers.

Associations between ambient and personal exposure and PEF

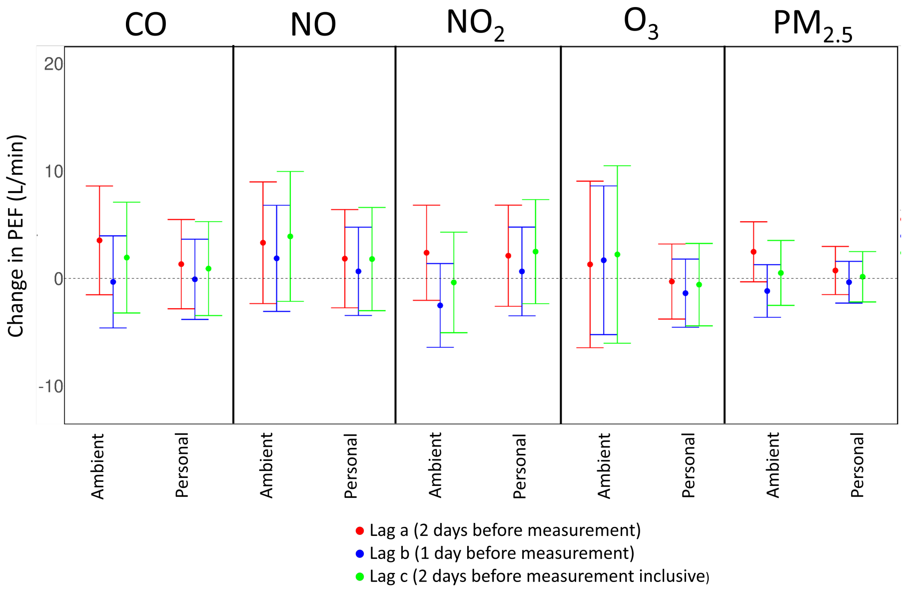

Figure 6.4: Effect of ambient and personal exposure (IQR) on PEF: Estimated changes with 95% confidence intervals in PEF (L/min) associated with a IQR increases in ambient and personal exposure to the key air pollutants. All fixed effects shown in Figure 6.2 are adjusted for. The scale of the y-axis is kept constant across all associations of increases in IQR with PEF.

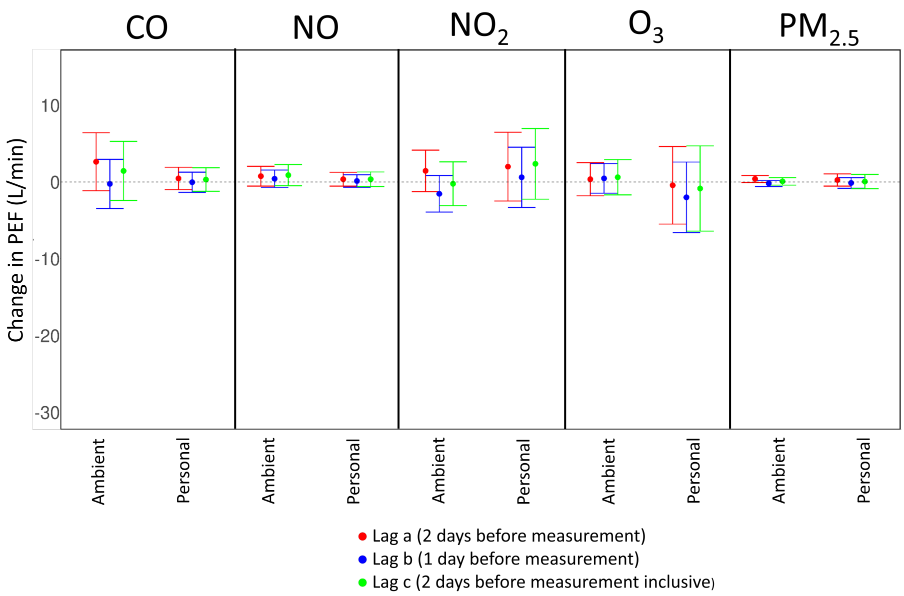

Figure 6.5: Effect of ambient and personal exposure (unit) on PEF: Estimated changes with 95% confidence intervals in PEF (L/min) associated with a 1000 ppb increase for CO, a 10 ppb increase for NO, NO_2 and O_3, and a 10 $\mu g/m^3$ increase for $PM_{2.5}$ in ambient and personal exposure. Ambient data was recorded at reference instruments and personal measurements were recorded by PAMs (see Chapter 3). All fixed effects shown in Figure 6.2 are adjusted for.

Associations between indoor-generated and outdoor-generated exposure and PEF

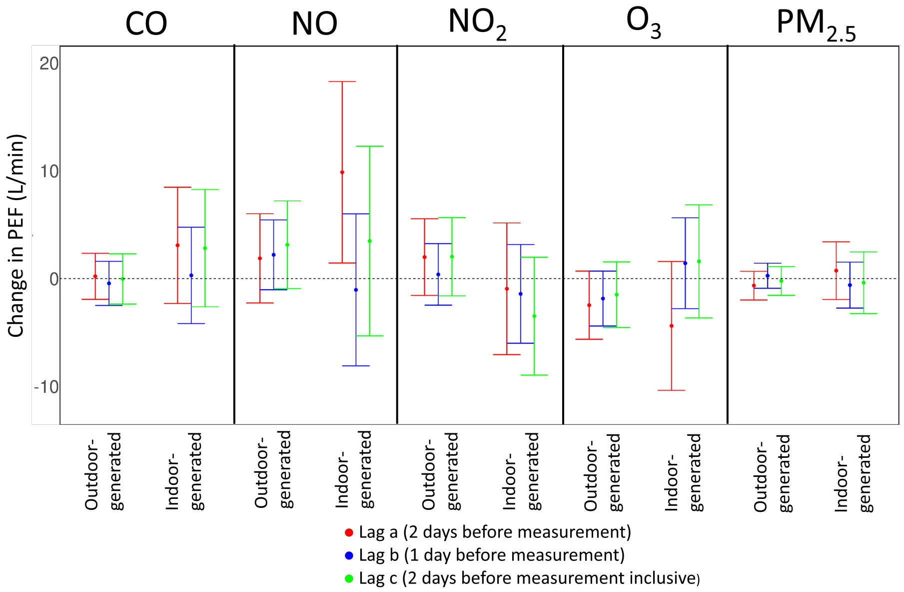

Figure 6.6: Effect of indoor-generated and outdoor-generated exposure (IQR) on PEF: Estimated changes with 95% confidence intervals in PEF (L/min) associated with a IQR increases in indoor-generated and outdoor-generated exposure to the key air pollutants. All fixed effects shown in Figure 6.2 are adjusted for.

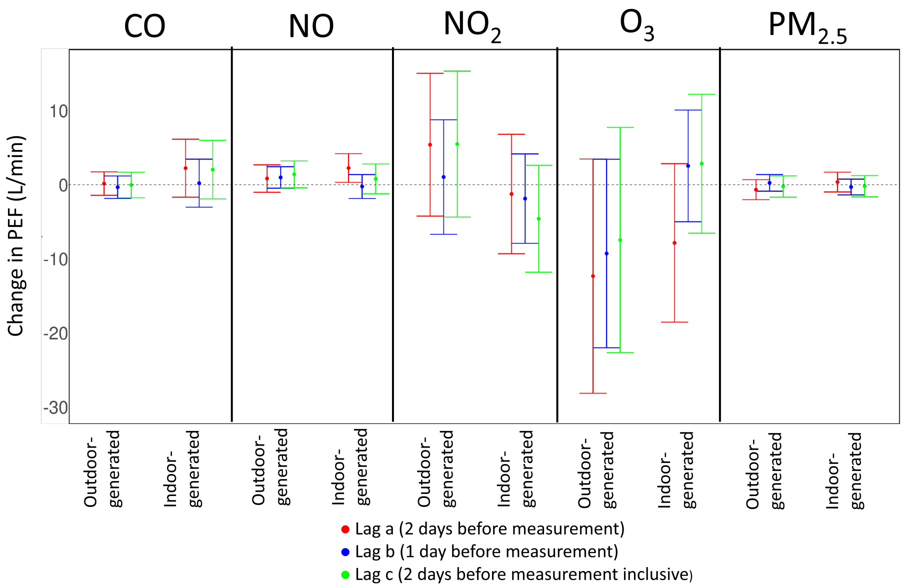

Figure 6.7: Effect of indoor-generated and outdoor-generated exposure (unit) on PEF: Estimated changes with 95% confidence intervals in PEF (L/min) associated with a 1000 ppb increase for CO, a 10 ppb increase for NO, NO_2 and O_3, and a 10 $\mu g/m^3$ increase for $PM_{2.5}$ in indoor-generated and outdoor-generated exposure. All fixed effects shown in Figure 6.2 are adjusted for.

Associations between exposure determinants and PEF

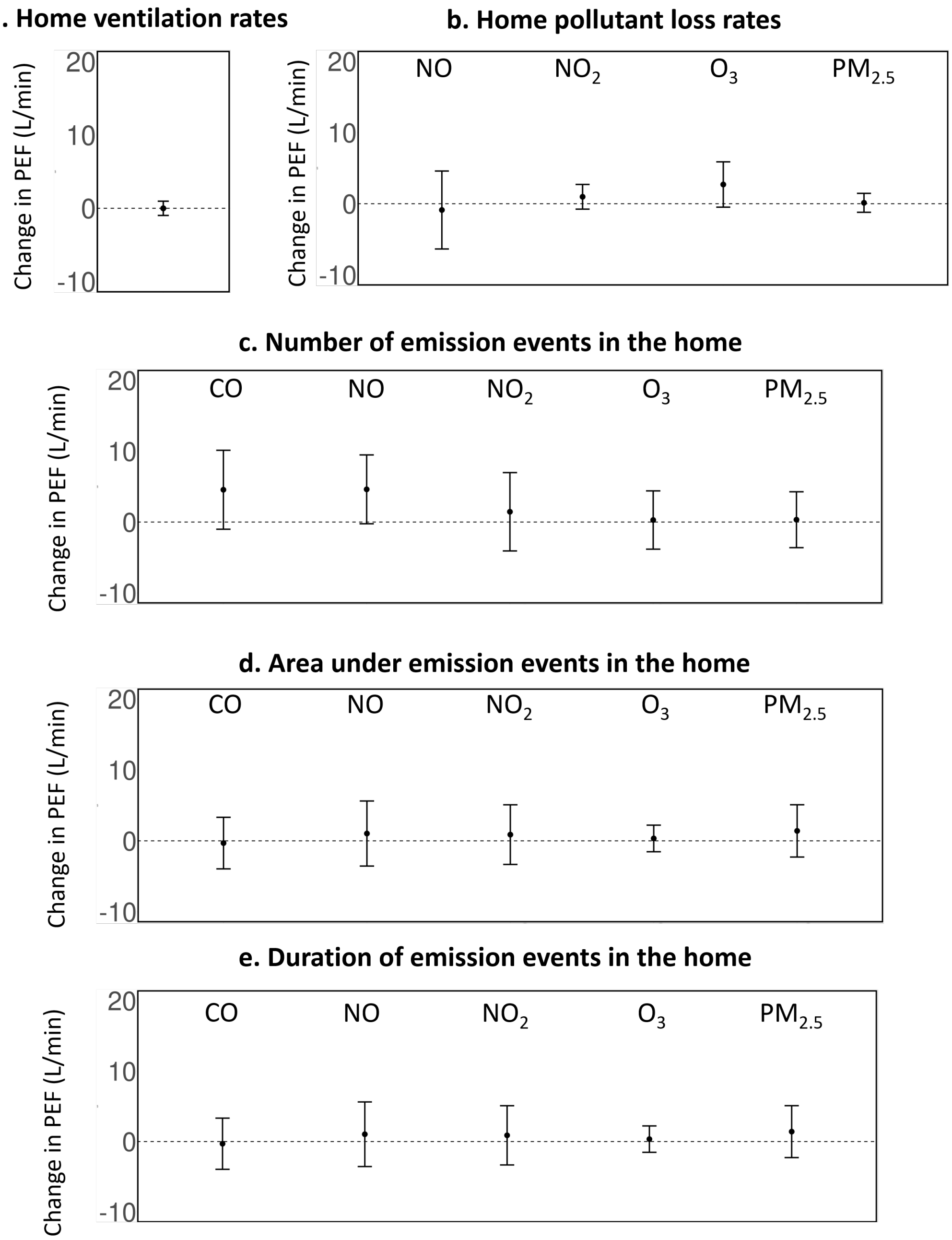

Figure 6.8: Effect of exposure determinants (IQR) on PEF: Estimated changes with 95% confidence intervals in PEF (L/min) associated with a IQR increases in exposure determinants: a) Home ventilation rates; b) Home pollutant loss rates; c) Number of emission events in the home; d) Area under emission events in the home; e) Duration of emission events in the home. Methodology to extract these metrics was shown in Section 2.5.4 and was demonstrated for the example participant in Section 4.5. As for the example participant, the threshold was selected to be the 90^{th} percentile. All fixed effects shown in Figure 6.2 are adjusted for. None of the exposure determinant metrics have a significant effect on PEF. The loss rates were estimated using the I/O ratio method.

Discussion of the results

Figures 6.6 and 6.7 indicate that increases in indoor-generated NO two days before a PEF measurement results in a higher PEF score. This is the only metric found to have a statistically significant effect on PEF. The other metrics have been shown to have no significant effect on PEF. However, tentative insights may still be inferred from the output values, which vary between metrics.

Most of the exposure metrics averaged at lag b (just the day before the PEF measurement) had a less positive (or more negative) association with PEF compared with the other lags, suggesting that the short term effects on PEF should be explored; lags of less than one day before the PEF measurement should be considered in future studies.

Although not significant, personal and ambient O_3 appear to have contrasting effects on health. Figures 6.4 and 6.5 show that ambient O_3 has a protective effect. Ambient ozone has been found to have a protective effect on health in the literature[17;10]. This observed effect has been attributed to the anti-correlation of O_3 with other pollutants such as NO_2 and $PM_{2.5}$. In this study, personal O_3 is shown to have a negative effect on PEF. This contradiction supports the requirement for personal monitoring of O_3 for lung function associations.

6.9 PEF as the response variable

Many of the metrics associated with health in this work are novel (for example, source-apportioned exposure) or are less commonly associated directly with health (such as ventilation or indoor pollutant loss rates). So statistically insignificant associations with PEF are not unexpected given the lack of prior testing or established expectations. However, strong associations of ambient levels of $PM_{2.5}$, NO_2 and O_3 in China with reduced lung function have been established in the literature. These associations are explored further in Appendix A.1.3. The lack of statistically significant associations for these pollutants is surprising.

For example, ambient O_3 has been shown to decrease lung function in the literature (Appendix A.1.3). As ambient O_3 is known to decrease lung function, it would be expected that high ambient O_3 would result in lower PEF values. Figures A.13 and A.12 in the Appendix show that ambient O_3 is significantly higher in the summer than in the winter. However, the PEF measurements collected as part of the AIRLESS study do not change with season, as shown in Figure 6.9. Therefore, the reliability of the PEF data is questioned.

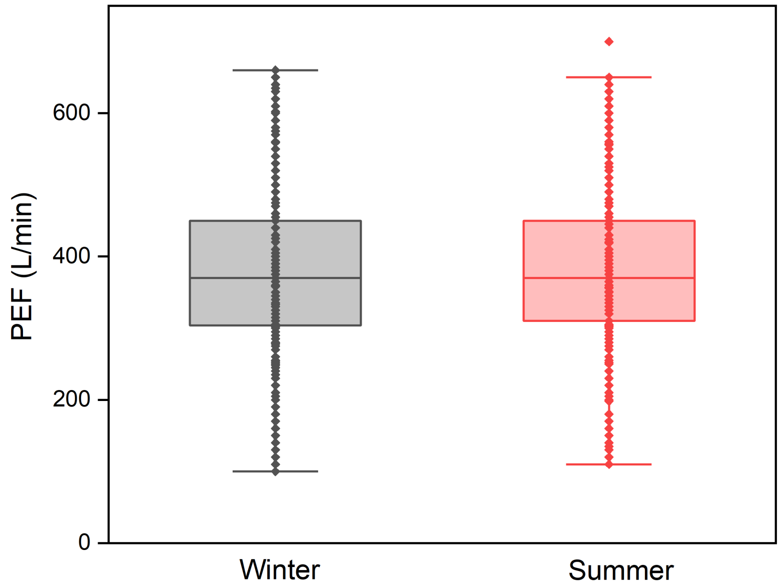

Figure 6.9: Seasonal variation of PEF All PEF measurements recorded during the AIRLESS project, separated by season. PEF does not appear to change between seasons.

During the AIRLESS project, only the first of the participant's measurements were supervised by a healthcare professional. As a result, proper lack of technique and inconsistent effort of the participant may result in unreliable data. This is supported by the data. On average, the PEF values measured by a participant ranged 27% from the normal peak flow rate for that participant. The American Lung Association report that, in general, a normal peak flow rate can vary as much as 20%[3]. The high variability of PEF measurements from the same participant supports that the reliability of the PEF data collected in AIRLESS study should be questioned. Other studies have found PEF data unreliable because participants have incorrectly recorded or even fabricated the PEF score[82;71].

A previous study explored associations between ambient O_3 and health. They also found no significant effect of O_3 on PEF; however, they did find a significant effect on other measures of lung function[23]. It is suggested that a spirometry test could be considered an alternative method to measure lung function, as these tests are performed by healthcare professionals in clinical settings.

6.10 Chapter summary

Despite concerns with using PEF measurements as the response variable, this chapter has demonstrated the construction of an LMEM to derive associations between exposure and health. Although not significant, ambient O_3 levels were found to have a protective effect on health, whereas personal O_3 exposure was found to harm health, showing that misclassification in exposure can impact health associations. It is expected that LMEMs, such as the one constructed here, will be of significant value when extended to directly examine the effects of the novel exposure metrics and estimated exposure determinants described in Chapter 5 on other health parameters. This chapter highlighted areas of consideration when constructing pollutant-health models, including the selection of fixed effects and exploration into lag.

Chapter 7

Conclusions

7.1 Interpretation of key findings

The work in this thesis has developed a methodological framework to assess and source-apportion personal air quality exposure to better understand the health impacts of air pollutants. Thanks to recent technological and computational advancements (particularly the progress in personal air quality monitoring and the development of the time-activity model), personal exposure in a range of microenvironments can be captured. However, other factors, such as the source of pollution, or the ventilation of the home environment may be driving health responses directly or indirectly. This framework allows for exploration into the associations of such additional metrics and characteristics with health.

Generation of novel exposure metrics

The first objective of the methodological framework was to generate novel exposure metrics from personal air quality measurements, for five pollutants. The framework generated personal exposure metrics. Comparison of personal exposure metrics to those inferred from measurements from stationary outdoor reference instruments showed differences; personal NO_2, O_3 and $PM_{2.5}$ exposure was found to be lower than that inferred from the stationary outdoor reference instruments, and for CO and NO, it was higher. Misclassification of exposure has been demonstrated through these comparisons, suggesting that personal measurements should be used in pollution-health models to avoid error in associations. The misclassification of exposure was attributed to the effects of loss processes and sources of pollution when the participants were indoors. Additionally, the five pollutants were less correlated in the personal exposure metrics than in the data measured by the reference instrument.

Breaking these pollutant correlations will improve the reliability of the estimated effects of individual pollutants on health.

Furthermore, the framework source-apportioned total personal exposure into pollution generated by indoor and outdoor sources. The methodology involved applying estimated ventilation rates and indoor pollutant loss rates to outdoor data for periods where the participants were detected to be indoors. This produced estimates of levels of pollutants which had been generated outdoors but had then ventilated indoors. These levels were subtracted from the personal measurements for these periods, retaining the pollution generated by indoor sources. When applied to the AIRLESS dataset, indoor-generated pollution was found to contribute between 29.7 % and 55.3 % to total exposure, depending on the pollutant. This apportionment allows for distinct associations between health outcomes of air mixtures from indoor and outdoor sources. For $PM_{2.5}$, a higher proportion of indoor-generated pollution is observed around 18:00 across both summer and winter and in Beijing and Pinggu, which has been attributed to emissions from cooking an evening meal. These results indicate that interventions solely targeting ambient air pollutant concentrations will only solve part of the air pollution problem.

To our knowledge, this is the first automated method developed to source-apportion total personal exposure.

Estimation of factors that determine exposure to air pollution
The second objective of the methodological framework was to estimate the factors that determine exposure to air pollution, specifically when the participant is in their home. People spend most of their time indoors, so these exposure determinants will be crucial in modelling total personal exposure at the population scale.

Home ventilation rates were estimated using the constant decay method. The average ventilation rate was found to be 3.18 hr^{-1} and was higher during the summer and the day, attributed to more frequent window opening. Most estimated home ventilation rates were higher than those recommended by the Chinese government (Table 5.3).

Indoor loss rates were inferred using the I/O ratio. The I/O ratio was calculated for each participant and each 12-hour time period separately to capture the effects of day and night. Additionally, the ratio was calculated in the absence of indoor sources. O_3 was found to have the strongest indoor loss rate in the home environment, followed by NO_2, $PM_{2.5}$ and then NO. Seasonal effects were found, particularly for the loss rate of O_3. Indoor loss rates were also estimated using the constant decay method, although this method produced less estimated values.

A method of characterising indoor emission events has been developed. The magnitude, frequency and area under the event are captured using this framework. This work showed that not only are quantities of these characteristics expected to be crucial when modelling exposure at the population levels, but they can also be used to infer information about the participants, for example, the type of cooking fuel used.

Linking of the novel exposure metrics, and exposure determinants, with health markers
The final objective of the framework was to link the novel exposure metrics and exposure determinants with health markers. This work demonstrated the construction of a statistical method (LMEM) to link pollution exposure with a health marker (in this case, PEF). Concerns about relying on PEF as a response variable were raised (Section 6.9), and most associations were not found to be significant. However, such an LMEM is expected to give insights into the effects of exposure metrics and estimated exposure determinants when extended to other health endpoints.

7.2 Limitations

The methodologies developed and applied as part of the framework are based on the continuity equation (Equation 2.1). This equation assumes that the indoor spaces inhabited by participants are single zones where the air pollutant concentrations are homogeneous (uniformly mixed). This is not a true representation of indoor air.

The developed source-apportionment methodologies require the estimation of ventilation rates. In this study, ventilation rates are estimated opportunistically from large influxes in CO concentration from cooking. For the AIRLESS cohort, this is suitable as the participants mainly relied on LPG and natural gas as their cooking fuels. This methodology may not be suitable for a cohort primarily using an electric cooker. Appendix A.5.9 shows the shift in primary cooking fuel in China and globally. Additionally, this methodology does not consider constant indoor sources of pollution.

7.3 Suggested future work

This work estimates exposure metrics, not dose. Dose is a measure of the amount of pollution that enters the body. Inhalation rates vary between people and can vary with factors such as age and gender. Additionally, an individual's inhalation rates may drastically change with physical activity.

The AIRLESS cohort only represents a subset of the population (including only current non-smokers over the age of 45). Additionally, although very large for a personal monitoring study, the sample size remains relatively small, which may introduce error into pollution health associations. Future work would apply this framework to other populations.

Only one reference instrument was used in each location to collect outdoor data. Although measurements from reference instruments in Beijing and Pinggu have been shown to be a suitable alternative to measurements directly outside peoples' homes[163], future work could explore the use of the ADMS-Urban model to model outdoor levels outside peoples' homes. Additionally, the reference instrument used in Pinggu during the campaign only measured hourly $PM_{2.5}$. Future work would investigate the impact of hourly reference data in place of minute reference data on the results from the framework developed in this thesis.

The exposure metrics estimated by the framework are estimated for each 12-hour time period. 06:00-18:00 and 18:00-06:00 periods were chosen. Using 12-hour time periods is an improvement on most studies, which use averages calculated over days or months. Future work may involve selecting time periods so that daytime included most daily activities, such as participants cooking their evening meal after 18:00. This would allow for further insights into the effects of participant activities on exposure.

A threshold was used to define the events when characterising indoor emission events (Section 2.5.4). This threshold is arbitrary and should be varied to see if it affects health associations. The threshold could be specific for a cohort or field campaign or a concentration previously identified in policy, for example, the indoor air quality standards for the respective country (for China, these are shown in Appendix A.5.3). It could also be a level identified as harmful from clinical or laboratory studies.

7.4 Concluding remark

The developed framework represents an important step in advancing our understanding of the effects of pollutants on health by providing the capability of generating novel exposure metrics and exposure determinants from personal air quality measurements.

Academic Output

Air flow experiments on a train carriage - Towards understanding the risk of airborne transmission

Huw Woodward, Shiwei Fan, Rajesh K. Bhagat, Maksim Dadonau, Megan Davies Wykes, Elizabeth Martin, Sarkawt Hama, Arvind Tiwari, Stuart B. Dalziel, Roderic L. Jones, Prashant Kumar and Paul F. Linden

Atmosphere, 2021

Abstract: A series of experiments was undertaken on an intercity train carriage aimed at providing a “proof of concept” for three methods in improving our understanding of airflow behaviour and the accompanied dispersion of exhaled droplets. The methods used included the following: measuring CO_2 concentrations as a proxy for exhaled breath, measuring the concentrations of different size fractions of aerosol particles released from a nebuliser, and visualising the flow patterns at cross-sections of the carriage by using a fog machine and lasers. Each experiment succeeded in providing practical insights into the risk of airborne transmission. For example, it was shown that the carriage is not well mixed over its length, however, it is likely to be well mixed along its height and width. A discussion of the suitability of the fresh air supply rates on UK train carriages is also provided, drawing on the CO_2 concentrations measured during these experiments.

Contribution: I spent a week working with the team to develop and conduct a number of experiments in a train carriage. I contributed in a number of ways: 1) Contributed to the development of methodology and experimental setup 2) Refurbished instruments and uploaded and shared data from the experiments 3) Contributed to the writing-review and editing process.

Relevance: This work employed tracer gas techniques to infer the ventilation of an indoor space. Aerosol concentrations were measured.

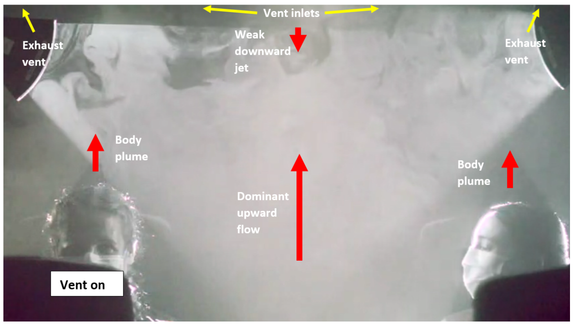

Figure 7.1: Involvement in air flow study: Snapshot of flow visualisation experiment at end of carriage during ventilated conditions, as part of the "Air flow experiments on a train carriage - Towards understanding the risk of airborne transmission" study, taken from Woodward et al.[153]

The experiments were performed at short notice and during somewhat challenging conditions during the COVID-19 pandemic (during the August of 2020). The full published article can be found in Appendix A.1.1.

Automated classification of time-activity-location patterns for improved estimation of personal exposure to air pollution

Lia Chatzidiakou, Anika Krause, Mike Kellaway, Yiqun Han, Yilin Li, Elizabeth Martin, Frank J. Kelly, Tong Zhu, Benjamin Barratt and Roderic L. Jones

Environmental Health, 2022

Abstract: Air pollution epidemiology has primarily relied on measurements from fixed outdoor air quality monitoring stations to derive population-scale exposure. Characterisation of individual time-activity-location patterns is critical for accurate estimations of personal exposure and dose because pollutant concentrations and inhalation rates vary significantly by location and activity. We developed and evaluated an automated model to classify major exposure-related microenvironments (home, work, other static, in-transit) and separated them into indoor and outdoor locations, sleeping activity and five modes of transport (walking, cycling, car, bus, metro/train) with multidisciplinary methods from the fields of movement ecology and artificial intelligence. As input parameters, we used GPS coordinates, accelerometry, and noise, collected at 1 min intervals with a validated Personal Air quality Monitor (PAM) carried by 35 volunteers for one week each. The model classifications were then evaluated against manual time-activity logs kept by participants. Overall, the model performed reliably in classifying home, work, and other indoor microenvironments (F1-score>0.70) but only moderately well for sleeping and visits to outdoor microenvironments (F1-score=0.57 and 0.3 respectively).

Random forest approaches performed very well in classifying modes of transport (F1-score>0.91). We found that the performance of the automated methods significantly surpassed those of manual logs. Automated models for time-activity classification can markedly improve exposure metrics. Such models can be developed in many programming languages, and if well formulated can have general applicability in large-scale health studies, providing a comprehensive picture of environmental health risks during daily life with readily gathered parameters from smartphone technologies.

Contribution: I contributed to the writing-review and editing process.

Relevance: The time-activity model developed in this paper computes the space-time utilisation distributions of the GPS coordinates for participants and classifies the microenvironment using metrics such as time spent in each location, re-visitation rate and metrics of directional movement. The model classifications are evaluated against manual time-activity logs kept by participants. This thesis uses a simplified version of the model developed in this paper to classify the individual time-activity-location patterns of the participants. The model is a key component of the framework developed in this thesis. The full published article can be found in Appendix A.1.2

Regeneration of the Cambridgeshire Fenlands fieldwork

The Cambridgeshire Fens project was set up by the Cambridge Landscape Regeneration Centre. The Centre's researchers are working to find the best ways of protecting both the Fen's ecosystem and its farmers by developing an integrated framework for the Fens to reconcile food production, reduce carbon emissions, secure water resources, manage flood risk, enrich biodiversity and improve resilience.

Contribution: I contributed to the fieldwork, making decisions on sensor spatial distribution for carbon emission measurements, setting up equipment, recording the details of the setup and replacing monitors.

Relevance: This work measured concentrations of ambient species in the rural environment.

Market research for Open-Seneca

Open-Seneca is a university based project that deploys citizen science air pollution monitoring networks across developing cities.

Contribution: I worked in a Development i-Team of seven researchers from across the University of Cambridge. Our team advised Open-Seneca on sustainable organisational structures that could be adopted in the future. We interviewed governmental bodies, academics and citizen scientists, and evaluated the pros and cons of different possible legal entities.

Relevance: The results of this work enable Open-Seneca to make more informed choices regarding their organisational structure, resulting in efficient and organised deployment of personal air quality sensors.

References

[1] Aakko-Saksa, P., Roslund, P., and Koponen, P. (2016). Development and validation of comprehensive emission measurement methods for alternative fuels at VTT. *VTT Technical Research Centre of Finland*.

[2] Amatngalim, G. and Hiemstra, P. (2018). Airway Epithelial Cell Function and Respiratory Host Defense in Chronic Obstructive Pulmonary Disease. *Chinese Medical Journal*, 131(9):1099.

[3] American Lung Association (2023). Measuring Your Peak Flow Rate.

[4] Aristotle and Reeve, C. D. C. (2017). *De Anima*. Hackett Publishing Company, Inc (US).

[5] Badman, D. G. and Jaffe, E. R. (1996). Blood and air pollution: State of knowledge and research needs. *Otolaryngology - Head and Neck Surgery*, 114(2):205–208.

[6] Banerjee, R., Dutta, M., and Roy, S. (2012). Evaluating the health cost of transport pollution. *International Journal of Physical and Social Sciences*.

[7] Baxter, L. K., Clougherty, J. E., Laden, F., and Levy, J. I. (2007). Predictors of concentrations of nitrogen dioxide, fine particulate matter, and particle constituents inside of lower socioeconomic status urban homes. *Journal of Exposure Science and Environmental Epidemiology*, 17:433–444.

[8] Becker, S., Madden, M. C., Newman, S. L., Devlin, R. B., and Koren, H. S. (1991). Modulation of human alveolar macrophage properties by ozone exposure in vitro. *Toxicology and Applied Pharmacology*, 110(3):403–415.

[9] Bell, M. L., Ebisu, K., Leaderer, B. P., Gent, J. F., Lee, H. J., Koutrakis, P., Wang, Y., Dominici, F., and Peng, R. D. (2014). Associations of PM2.5 constituents and sources with hospital admissions: Analysis of four counties in

connecticut and Massachusetts (USA) for persons less of equal to 65 years of age. *Environmental Health Perspectives*, 122(2):138–144.

[10] Bentayeb, M., Wagner, V., Stempfelet, M., Zins, M., Goldberg, M., and Pascal, M. (2015). Association between long-term exposure to air pollution and mortality in France: A 25-year follow-up study. *Environment International*.

[11] Bertrand, L., Dawkins, L., Jayaratne, R., and Morawska, L. (2020). How to choose healthier urban biking routes: CO as a proxy of traffic pollution. *Heliyon*, 6(6):e04195.

[12] Borrego, C., Tchepel, O., Costa, A. M., Martins, H., Ferreira, J., and Miranda, A. I. (2006). Traffic-related particulate air pollution exposure in urban areas. *Atmospheric Environment*, 40(37):7205–7214.

[13] Brook, R. D., Franklin, B., Cascio, W., Hong, Y., Howard, G., Lipsett, M., Luepker, R., Mittleman, M., Samet, J., Smith, S. C., and Tager, I. (2004). Air Pollution and Cardiovascular Disease. *Circulation*, 109(21):2655–2671.

[14] Calfapietra, C., Fares, S., and Loreto, F. (2009). Volatile organic compounds from Italian vegetation and their interaction with ozone. *Environmental Pollution*, 157(5):1478–1486.

[15] Calfapietra, C., Fares, S., Manes, F., Morani, A., Sgrigna, G., and Loreto, F. (2013). Role of Biogenic Volatile Organic Compounds (BVOC) emitted by urban trees on ozone concentration in cities: A review. *Environmental Pollution*, 183:71–80.

[16] Cao, J., Yang, C., Li, J., Chen, R., Chen, B., Gu, D., and Kan, H. (2011). Association between long-term exposure to outdoor air pollution and mortality in China: A cohort study. *Journal of Hazardous Materials*, 186(2-3):1594–1600.

[17] Carey, I. M., Atkinson, R. W., Kent, A. J., Van Staa, T., Cook, D. G., and Anderson, H. R. (2013). Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. *American journal of respiratory and critical care medicine*, 187(11):1226–1233.

[18] Carslaw, N. (2019). Chemical Reactions in the Indoor Atmosphere. In *Issues in Environmental Science and Technology*, number 48, pages 105–126. Royal Society of Chemistry.

[19] Chappells, H. (2012). *The Future of Sustainability*, volume 10. Berkshire Publishing Group, Great Barrington, United States.

[20] Chatzidiakou, L., Krause, A., Han, Y., Chen, W., Yan, L., Popoola, O. A., Kellaway, M., Wu, Y., Liu, J., Hu, M., Barratt, B., Kelly, F. J., Zhu, T., and Jones, R. L. (2020). Using low-cost sensor technologies and advanced computational methods to improve dose estimations in health panel studies: Results of the AIRLESS project. *Journal of Exposure Analysis and Environmental Epidemiology.*

[21] Chatzidiakou, L., Krause, A., Kellaway, M., Han, Y., Li, Y., Martin, E., Kelly, F. J., Zhu, T., Barratt, B., and Jones, R. L. (2022). Automated classification of time-activity-location patterns for improved estimation of personal exposure to air pollution. *Environmental Health: A Global Access Science Source*, 21(1):1–21.

[22] Chatzidiakou, L., Krause, A., Popoola, O. A. M., Di Antonio, A., Kellaway, M., Han, Y., Squires, F. A., Wang, T., Zhang, H., Wang, Q., Fan, Y., Chen, S., Hu, M., Quint, J. K., Barratt, B., Kelly, F. J., Zhu, T., and Jones, R. L. (2019). Characterising low-cost sensors in highly portable platforms to quantify personal exposure in diverse environments. *Atmospheric Measurement Techniques*, 12(8):4643–4657.

[23] Chen, J. H., Hu, D. T., Jia, X., Niu, W., Deng, F. R., and Guo, X. B. (2020). Monitoring metrics for short-term exposure to ambient ozone and pulmonary function and airway inflammation in healthy young adults. *Journal of Peking University (Health Sciences)*, 52(3):492–499.

[24] Cheng, P. L. and Li, X. (2018). Air infiltration rates in the bedrooms of 202 residences and estimated parametric infiltration rate distribution in Guangzhou, China. *Energy and Buildings*, 164:219–225.

[25] COMEAP (2015a). 2015a Statement: The evidence for the effects of nitrogen dioxide on health . Technical report, GOV.UK.

[26] COMEAP (2015b). 2015b NO2 working group. Working group on the quantification of mortality associated with long-term average concentrations of nitrogen dioxide (NO2) Note of the Meeting held on Monday 6 July 2015.

[27] COMEAP (2015c). Interim statement on quantifying the association of long-term average concentrations of nitrogen dioxide and mortality.

[28] Cox, R. A., Derwent, R. G., and Holt, P. M. (1976). Relative rate constants for the reactions of OH radicals with H2, CH4, CO, NO and HONO at atmospheric pressure and 296 K. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 72(0):2031–2043.

[29] Cross, E. S., Williams, L. R., Lewis, D. K., Magoon, G. R., Onasch, T. B., Kaminsky, M. L., Worsnop, D. R., and Jayne, J. T. (2017). Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. *Atmospheric Measurement Techniques*, 10(9):3575–3588.

[30] Cserbik, D., Chen, J. C., McConnell, R., Berhane, K., Sowell, E. R., Schwartz, J., Hackman, D. A., Kan, E., Fan, C. C., and Herting, M. M. (2020). Fine particulate matter exposure during childhood relates to hemispheric-specific differences in brain structure. *Environment international*, 143:105933.

[31] Cui, S., Cohen, M., Stabat, P., and Marchio, D. (2015). CO2 tracer gas concentration decay method for measuring air change rate. *Building and Environment*, 84:162–169.

[32] Dennekamp, M., Howarth, S., Dick, C. A. J., Cherrie, J. W., Donaldson, K., and Seaton, A. (2001). Ultrafine particles and nitrogen oxides generated by gas and electric cooking. *Occupational & Environmental Medicine*.

[33] Devlin, R. B., McKinnon, K. P., Noah, T., Becker, S., and Koren, H. S. (1994). Ozone-induced release of cytokines and fibronectin by alveolar macrophages and airway epithelial cells. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 266(6 10-6).

[34] Ding, S., Teo, Z. W., Wan, M. P., and Ng, B. F. (2021). Aerosols from speaking can linger in the air for up to nine hours. *Building and Environment*, 205.

[35] Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G., and Speizer, F. E. (1993). An Association between Air Pollution and Mortality in Six U.S. Cities. *New England Journal of Medicine*, 329(24):1753–1759.

[36] Dutton, A. (2020). Coronavirus (COVID-19) related mortality rates and the effects of air pollution in England. Technical report, Office for National Statistics.

[37] Else, H. (2020). How a torrent of COVID science changed research publishing - in seven charts. *Nature*, 588(7839):553.

[38] Fadeyi, M. O. (2015). Ozone in indoor environments: Research progress in the past 15 years. *Sustainable Cities and Society*, 18:78–94.

[39] Fehsenfeld, F., Calvert, J., Fall, R., Goldan, P., Guenther, A. B., Hewitt, C. N., Lamb, B., Liu, S., Trainer, M., Westberg, H., and Zimmerman, P. (1992). Emissions of volatile organic compounds from vegetation and the implications for atmospheric chemistry. *Global Biogeochemical Cycles*, 6(4):389–430.

[40] Few, J. and Elwell, C. A. (2021). Applying the CO2 concentration decay tracer gas method in long-term monitoring campaigns in occupied homes: identifying appropriate unoccupied periods and decay periods. *International Journal of Building Pathology and Adaptation (ahead of print).*

[41] Finlayson-Pitts, B., Waste, J. P. J. A. ., and 1993, u. (1993). Atmospheric chemistry of tropospheric ozone formation: scientific and regulatory implications. *Taylor & Francis*, 43(8):1091–1100.

[42] Finlayson-Pitts, B. J., Wingen, L. M., Sumner, A. L., Syomin, D., and Ramazan, K. A. (2003). The heterogeneous hydrolysis of NO2 in laboratory systems and in outdoor and indoor atmospheres: An integrated mechanism. *Physical Chemistry Chemical Physics*, 5(2):223–242.

[43] Fuller, R. and Landrigan, P. J. (2022). Pollution and health: a progress update. *The Lancet Planetary Health*, 6(6):e535–e547.

[44] Gao, N., Xu, W., Ji, J., Yang, Y., Wang, S. T., Wang, J., Chen, X., Meng, S., Tian, X., Xu, K. F., and Xu, K. F. (2020). Lung function and systemic inflammation associated with short-Term air pollution exposure in chronic obstructive pulmonary disease patients in Beijing, China. *Environmental Health: A Global Access Science Source*, 19(1):1–10.

[45] Gao, Q., Wang, F., Hu, L., Yu, J., Liu, R., Wang, Y., and Liu, Y. (2017a). Changes in Time Spent Outdoors During the Daytime in Rural Populations in Four Geographically Distinct Regions in China: A Retrospective Study. *Photochemistry and Photobiology*, 93(2):619–625.

[46] Gao, W., Tie, X., Xu, J., Huang, R., Mao, X., Zhou, G., and Chang, L. (2017b). Long-term trend of O3 in a mega City (Shanghai), China: Characteristics, causes, and interactions with precursors. *Science of The Total Environment*, 603-604:425–433.

[47] Ghazi, C. J. and Marshall, J. S. (2014). A CO2 tracer-gas method for local air leakage detection and characterization. *Flow Measurement and Instrumentation*, 38:72–81.

[48] Gibbs-Flournoy, E. A., Gilmour, M. I., Higuchi, M., Jetter, J., George, I., Copeland, L., Harrison, R., Moser, V. C., and Dye, J. A. (2018). Differential exposure and acute health impacts of inhaled solid-fuel emissions from rudimentary and advanced cookstoves in female CD-1 mice. *Environmental Research*, 161:35–48.

[49] Gilbert, N. L., Gauvin, D., Guay, M., Héroux, M.-, Dupuis, G., Legris, M., Chan, C. C., Dietz, R. N., and Lévesque, B. (2006). Housing characteristics and indoor concentrations of nitrogen dioxide and formaldehyde in Quebec City, Canada. *Environmental Research*, 102(1):1–8.

[50] Goldman, E. (2020). Exaggerated risk of transmission of COVID-19 by fomites. *The Lancet Infectious Diseases*, 20(8):892–893.

[51] Han, Y., Chen, W., Chatzidiakou, L., Krause, A., Yan, L., Zhang, H., Chan, Q., Barratt, B., Jones, R., Liu, J., Wu, Y., Zhao, M., Zhang, J., Kelly, F. J., and Zhu, T. (2020a). Effects of AIR pollution on cardiopuLmonary disEaSe in urban and peri-urban reSidents in Beijing: Protocol for the AIRLESS study. *Atmospheric Chemistry and Physics*, 20(24):15775–15792.

[52] Han, Y., Chen, W., Chatzidiakou, L., Yan, L., Zhang, H., Chan, Q., Barratt, B., Jones, R., Liu, J., Wu, Y., Zhao, M., Zhang, J., Kelly, F. J., Tong, Z., and Team, A. (2020b). Effects of AIR pollution on cardiopuLmonary disEaSe in urban and peri-urban reSidents in Beijing: protocol for the AIRLESS study. *Atmospheric Chemistry and Physics*.

[53] He, M. Z., Kinney, P. L., Li, T., Chen, C., Sun, Q., Ban, J., Wang, J., Liu, S., Goldsmith, J., and Kioumourtzoglou, M. A. (2020). Short- and intermediate-term exposure to NO2 and mortality: A multi-county analysis in China. *Environmental Pollution*, 261.

[54] Heroux, M.-E., Clark, N., Ryswyk, K. V., Mallick, R., Gilbert, N. L., Harrison, I., Rispler, K., Wang, D., Anastassopoulos, A., Guay, M., MacNeill, M., and Wheeler, A. J. (2010). Predictors of Indoor Air Concentrations in Smoking and Non-Smoking Residences. *International Journal of Environmental Research and Public Health*, 7(8):3080–3099.

[55] Hilburn, M. (2014). *Encyclopedia of Toxicology*. Elsevier.

[56] Hippocrates, Hayes, E., and Nimis, S. A. (2013). *Hippocrates' On airs, waters, and places, and the Hippocratic oath : an intermediate Greek reader : Greek text with running vocabulary and commentary*. Faenum Publishing, Ltd.

[57] Hollowell, C. D., Berk, J. V., Boegel, M. L., Miksch, R. R., Nazaroff, W. W., and Traynor, G. W. (1980). Building ventilation and indoor air quality. *Studies in Environmental Science*, 8(C):387–396.

[58] Hou, J., Sun, Y., Chen, Q., Cheng, R., Liu, J., Shen, X., Tan, H., Yin, H., Huang, K., Gao, Y., Dai, X., Zhang, L., Liu, B., and Sundell, J. (2019). Air change rates in urban Chinese bedrooms. *Indoor Air*, 29(5):828–839.

[59] Hou, J., Zhang, Y., Sun, Y., Wang, P., Zhang, Q., Kong, X., and Sundell, J. (2018). Air change rates at night in northeast Chinese homes. *Building and Environment*, 132:273–281.

[60] Hu, Y. and Zhao, B. (2020). Relationship between indoor and outdoor NO2: A review. *Building and Environment*, 180.

[61] Huang, Y., Yang, Z., and Gao, Z. (2019). Contributions of Indoor and Outdoor Sources to Ozone in Residential Buildings in Nanjing. *International Journal of Environmental Research and Public Health 2019, Vol. 16, Page 2587*, 16(14):2587.

[62] Huang, Z., Wu, Z., Yu, M., and Dong, Y. (2017). The Measurement of Natural Ventilation in Huizhou Traditional Dwelling in Summer. *Procedia Engineering*, 205:1439–1445.

[63] Hun, D. E., Corsi, R. L., Morandi, M. T., and Siegel, J. A. (2010). Formaldehyde in Residences: Long-Term Indoor Concentrations and Influencing Factors. *Indoor Air*.

[64] Hurley, J. F. (2018). Working Paper for COMEAP Report: “Associations of long-term average concentrations of nitrogen dioxide with mortality” and Working Paper 2: ”A viewpoint on using adjusted coefficients for NO2 and PM2.5”.

[65] Imperial College London (2015). AIRLESS: The Effects of AIR pollution on cardiopulmonary disease in urban & peri-urban residents in Beijing. *http://erg.ic.ac.uk/research/home/projects/AIRLESS-The-Effects-of-AIR-pollution-on-cardiopulmonary-disease-in-urban-peri-urban-residents-in-Beijing.html*.

[66] Jankovic, A., Gennaro, G., Chaudhary, G., Goia, F., and Favoino, F. (2022). Tracer gas techniques for airflow characterization in double skin facades. *Building and Environment*, 212:108803.

[67] Jarvis, D. L., Leaderer, B. P., Chinn, S., and Burney, P. G. (2005). Indoor nitrous acid and respiratory symptoms and lung function in adults. *Thorax*, 60(6):474–479.

[68] Jeong, H. and Park, D. (2018). Characteristics of peak concentrations of black carbon encountered by elementary school children. *Science of the Total Environment*, 637-638:418–430.

[69] Ji, C., Xia, Y., Dai, H., Zhao, Z., Liu, T., Tong, S., Zhang, X., and Zhao, Y. (2021). Reference Values and Related Factors for Peak Expiratory Flow in Middle-Aged and Elderly Chinese. *Frontiers in Public Health*, 9.

[70] Johnson, M., Lam, N., Brant, S., Gray, C., and Pennise, D. (2011). Modeling indoor air pollution from cookstove emissions in developing countries using a Monte Carlo single-box model. *Atmospheric Environment*, 45(19):3237–3243.

[71] Kamps, A., Roorda, R., Thorax, P. B., and 2001, u. (2001). Peak flow diaries in childhood asthma are unreliable. *Thorax*, 56:180–182.

[72] Korn, E. L. and Whittemore, A. S. (1979). Methods for Analyzing Panel Studies of Acute Health Effects of Air Pollution. *Biometrics*, 35(4):795.

[73] Krause, A. (2020). *Using novel portable air quality monitors to improve personal exposure and dose estimations for health studies*. PhD thesis, University of Cambridge, Cambridge.

[74] Lachowicz, J., Milia, S., Jaremko, M., Oddone, E., Cannizzaro, E., Cirrincione, L., Malta, G., Campagna, M., and Lecca, L. (2023). Cooking Particulate Matter: A Systematic Review on Nanoparticle Exposure in the Indoor Cooking Environment. *Atmosphere*, 14(12).

[75] Levy, H. (1971). Normal Atmosphere: Large Radical and Formaldehyde Concentrations Predicted. *Science*, 173(3992):141–143.

[76] Li, G., Xue, M., Zeng, Q., Cai, Y., Pan, X., Meng, Q., and Li, G. (2017). Association between fine ambient particulate matter and daily total mortality: An analysis from 160 communities of China. *Science of the Total Environment*, 599:108–113.

[77] Lim, J. M., Jeong, J. H., Lee, J. H., Moon, J. H., Chung, Y. S., and Kim, K. H. (2011). The analysis of PM2.5 and associated elements and their indoor/outdoor pollution status in an urban area. *Indoor Air*, 21.

[78] Linden, P. F. (1999). The Fluid Mechanics of Natural Ventilation. *Annual Review of Fluid Mechanics*, 31(1):201–238.

[79] Liu, C., Yin, P., Chen, R., Meng, X., Wang, L., Niu, Y., Lin, Z., Liu, Y., Liu, J., Qi, J., You, J., Kan, H., and Zhou, M. (2018). Ambient carbon monoxide and cardiovascular mortality: a nationwide time-series analysis in 272 cities in China. *The Lancet Planetary Health*, 2(1):e2–e3.

[80] Liu, Y., Misztal, P. K., Arata, C., Weschler, C. J., Nazaroff, W. W., and Goldstein, A. H. (2021). Observing ozone chemistry in an occupied residence. *Proceedings of the National Academy of Sciences of the United States of America*, 118(6).

[81] Lomauro, A. and Aliverti, A. (2018). Sex differences in respiratory function. *Breathe*, 14(2):131–140.

[82] Malo, J. L., Trudeau, C., Ghezzo, H., L'Archevêque, J., and Cartier, A. (1995). Do subjects investigated for occupational asthma through serial peak expiratory flow measurements falsify their results? *Journal of Allergy and Clinical Immunology*, 96(5):601–607.

[83] Mannucci, P. M. and Franchini, M. (2017). Health Effects of Ambient Air Pollution in Developing Countries. *International Journal of Environmental Research and Public Health*, 14(9).

[84] Matz, C. J., Egyed, M., Hocking, R., Seenundun, S., Charman, N., and Edmonds, N. (2019). Human health effects of traffic-related air pollution (TRAP): a scoping review protocol. *Systematic reviews*, 8(1).

[85] Matz, C. J., Stieb, D. M., Davis, K., Egyed, M., Rose, A., Chou, B., and Brion, O. (2014). Effects of Age, Season, Gender and Urban-Rural Status on Time-Activity: Canadian Human Activity Pattern Survey 2 (CHAPS 2). *International Journal of Environmental Research and Public Health 2014, Vol. 11, Pages 2108-2124*, 11(2):2108–2124.

[86] Mazokopakis, E. E. and Karagiannis, C. G. (2019). Environmental and Medical Aspects Related to the Sixth Plague of Egypt. *Maedica (Bucur)*, 14(3).

[87] Mcneil, S., Quaglia, L., Bassett, M., Overton, G., and Plagmann, M. (2012). A survey of airtightness and ventilation rates in post 1994 NZ homes. *Energy in Buildings and Communities Programme.*

[88] Mertes, S. and Wahner, A. (1995). Uptake of nitrogen dioxide and nitrous acid on aqueous surfaces. *Journal of Physical Chemistry*, 99(38):14000–14006.

[89] Miller, K. A., Spalt, E. W., Gassett, A. J., Curl, C. L., Larson, T. V., Avol, E., Allen, R. W., Sverre Vedal, Szpiro, A. A., and Kaufman, J. D. (2019). Estimating ambient-origin PM 2.5 exposure for epidemiology: observations, prediction, and validation using personal sampling in the Multi-Ethnic Study of Atherosclerosis. *Journal of Exposure Science & Environmental Epidemiology*, 29:227–237.

[90] Mohammed, M. O., Song, W., Ma, W., Li, W., Ambuchi, O. J., Thabit, M., and Li, Y. (2015). Trends in indoor–outdoor PM2.5 research: A systematic review of studies conducted during the last decade (2003–2013). *Atmospheric Pollution Research*, 6(5):893–903.

[91] Molnár, A., Imre, K., Ferenczi, Z., Kiss, G., Gelencsér, A., Molnár, A., Imre, K., Ferenczi, Z., Kiss, G., and Gelencsér, A. (2020). Aerosol hygroscopicity: Hygroscopic growth proxy based on visibility for low-cost PM monitoring. *AtmRe*, 236:104815.

[92] Morrison, G. (2015). Recent Advances in Indoor Chemistry. *Current Sustainable/Renewable Energy Reports*, 2(2):33–40.

[93] Muller, C. O., Yu, H., and Zhu, B. (2015). Ambient Air Quality in China: The Impact of Particulate and Gaseous Pollutants on IAQ. *Procedia Engineering*, 121:582–589.

[94] National Standard of the People's Republic of China (2012). Design Standard for Energy Efficiency of Residential Buildings in Hot Summer and Warm Winter Zone. *Beijing: China Architecture & Building Press*, JGJ75.

[95] Nazaroff, W., Gadgil, A., and Weschler, C. (1993). Critique of the Use of Deposition Velocity in Modeling Indoor Air Quality. In *Modeling of Indoor Air Quality and Exposure*, page 81. ASTM International.

[96] Nemmar, A., Hoet, P. H., Vanquickenborne, B., Dinsdale, D., Thomeer, M., Hoylaerts, M. F., Vanbilloen, H., Mortelmans, L., and Nemery, B. (2002). Passage of Inhaled Particles Into the Blood Circulation in Humans. *Circulation*, 105(4):411–414.

[97] Niu, Y., Zhou, Y., Chen, R., Yin, P., Meng, X., Wang, W., Liu, C., Ji, J. S., Qiu, Y., Kan, H., and Zhou, M. (2022). Long-term exposure to ozone and car-

diovascular mortality in China: a nationwide cohort study. *The Lancet Planetary Health*, 6(6):e496–e503.

[98] Nori-Sarma, A., Thimmulappa, R., Venkataraman, G. V., Warren, J. L., Berman, J. D., Whittaker, S. D., Kulick, E. R., Wellenius, G. A., Mahesh, P. A., and Bell, M. L. (2021). NO2 exposure and lung function decline in a cohort of adults in Mysore, India. *Environmental Research Communications*, 3(5):055001.

[99] Oberdörster, G., Oberdörster, E., and Oberdörster, J. (2005). Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles. *Environmental Health Perspectives*, 113(7):823–839.

[100] O'Connell, A. A. and McCoach, D. B. (2008). *Multilevel modeling of educational data.* IAP.

[101] Offermann, F. J. (2010). IAQ in airtight homes. *ASHRAE Journal*, 52(11):58–61.

[102] Peters, J., Van den Bossche, J., Reggente, M., Van Poppel, M., De Baets, B., and Theunis, J. (2014). Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. *Atmospheric Environment*, 92:31–43.

[103] Philip, S., Martin, R. V., Van Donkelaar, A., Lo, J. W. H., Wang, Y., Chen, D., Zhang, L., Kasibhatla, P. S., Wang, S., Zhang, Q., Lu, Z., Streets, D. G., Bittman, S., and Macdonald, D. J. (2014). Global chemical composition of ambient fine particulate matter for exposure assessment. *Environmental Science and Technology*, 48(22):13060–13068.

[104] Pica, N. and Bouvier, N. M. (2012). Environmental factors affecting the transmission of respiratory viruses. *Current Opinion in Virology*, 2(1):90.

[105] Popoola, O. A., Stewart, G. B., Mead, M. I., and Jones, R. L. (2016). Development of a baseline-temperature correction methodology for electrochemical sensors and its implications for long-term stability. *Atmospheric Environment*, 147:330–343.

[106] Pott, P. (1775). Chirurgical observations relative to the cataract, the polypus of the nose, the cancer of the scrotum, the different kinds of ruptures, and the mortification of the toes and feet. *Chirurgical Observations*, pages 63–68.

[107] Preller, L., Burstyn, I., De Pater, N., and Kromhout, H. (2004). Characteristics of peaks of inhalation exposure to organic solvents. *The Annals of occupational hygiene*, 48(7):643–652.

[108] Priestley, J. (1774). *Experiments and observations on different kinds of air. Second edition, Volume 2.* Cambridge University Press.

[109] Pun, V. C., Manjourides, J., and Suh, H. (2017). Association of Ambient Air Pollution with Depressive and Anxiety Symptoms in Older Adults: Results from the NSHAP Study. *Environmental health perspectives*, 125(3):342–348.

[110] Rasmussen, T. R., Brauer, M., and Kjærgaard, S. (1995). Effects of nitrous acid exposure on human mucous membranes. *American Journal of Respiratory and Critical Care Medicine*, 151(5):1504–1511.

[111] Robbins, R. C., Borg, K. M., and Robinson, E. (1968). Carbon monoxide in the atmosphere. *Journal of the Air Pollution Control Association*, 18(2):106–110.

[112] Rouse Jonathan, Büttner, L., and Rehfuess, E. (2008). Evaluating household energy and health interventions: a catalogue of methods.

[113] Ryan, P. H., Bernstein, D. I., Lockey, J., Reponen, T., Levin, L., Grinshpun, S., Villareal, M., Khurana Hershey, G. K., Burkle, J., and LeMasters, G. (2012). Exposure to Traffic-related Particles and Endotoxin during Infancy Is Associated with Wheezing at Age 3 Years. *Am J Respir Crit Care Med.*, 180(11):1068–1075.

[114] Samoli, E., Atkinson, R. W., Analitis, A., Fuller, G. W., Green, D. C., Mudway, I., Anderson, H. R., and Kelly, F. J. (2016). Associations of short-term exposure to traffic-related air pollution with cardiovascular and respiratory hospital admissions in London, UK. *Occupational and Environmental Medicine*, 73(5):300–307.

[115] Schroer, K. T., Biagini Myers, J. M., Ryan, P. H., LeMasters, G. K., Bernstein, D. I., Villareal, M., Lockey, J. E., Reponen, T., Grinshpun, S., and Khurana Hershey, G. K. (2009). Associations between Multiple Environmental Exposures and Glutathione S-transferase P1 on Persistent Wheezing in a Birth Cohort. *Journal of Pediatrics*, 154(3).

[116] Settimo, G., Yu, Y., Gola, M., Buffoli, M., and Capolongo, S. (2023). Challenges in IAQ for Indoor Spaces: A Comparison of the Reference Guideline Values of Indoor Air Pollutants from the Governments and International Institutions. *ATMOSPHERE*, 14(4):1–37.

[117] Setton, E. M., Peter, C. P., Cloutier-Fisher, D., and Hystad, P. W. (2008). Spatial variations in estimated chronic exposure to traffic-related air pollution in

working populations: A simulation. *International Journal of Health Geographics*, 7.

[118] Sharma, S., Chandra, M., and Kota, S. H. (2020). Health Effects Associated with PM2.5: a Systematic Review. *Current Pollution Reports*, 6(4):345–367.

[119] Shen, J. and Gao, Z. (2018). Ozone removal on building material surface: A literature review. *Building and Environment*, 134:205–217.

[120] Shen, Y. and Wang, J. (2004). The applications and market outlooks of a couple of decoration materials of floors. *Build. Decor.*, pages 40–42.

[121] Sherman, M. H. (1990). Tracer-Gas Techniques For Measuring Ventilation in a Single Zone. *Building and Environment*, 25(4):365–374.

[122] Shi, S., Chen, C., and Zhao, B. (2015). Air infiltration rate distributions of residences in Beijing. *Building and Environment*, 92:528–537.

[123] Sleiman, M., Gundel, L. A., Pankow, J. F., Jacob, P., Singer, B. C., and Destaillats, H. (2010). Formation of carcinogens indoors by surface-mediated reactions of nicotine with nitrous acid, leading to potential thirdhand smokehazards. *Proceedings of the National Academy of Sciences of the United States of America*, 107(15):6576–6581.

[124] Smith, T. (2001). Studying peak exposure – toxicology and exposure statistics. *Exposure Assessment in Epidemiology and Practice*, pages 207–209.

[125] Son, J. Y. and Bell, M. L. (2013). The relationships between short-term exposure to particulate matter and mortality in Korea: Impact of particulate matter exposure metrics for sub-daily exposures. *Environmental research letters*, 8(1):014015.

[126] Stamp, S., Burman, E., Chatzidiakou, L., Cooper, E., Wang, Y., and Mumovic, D. (2022). A critical evaluation of the dynamic nature of indoor-outdoor air quality ratios. *Atmospheric Environment*, 273:118955.

[127] Stiratelli, R., Laird, N., and Ware, J. H. (1984). Random-Effects Models for Serial Observations with Binary Response. *Biometrics*, pages 961–971.

[128] Stoner, O., Lewis, J., Martínez, I. L., Gumy, S., Economou, T., and Adair-Rohani, H. (2021). Household cooking fuel estimates at global and country level for 1990 to 2030. *Nature Communications*, 12(1).

[129] Sun, Y., Zhang, Y., Bao, L., Fan, Z., and Sundell, J. (2011). Ventilation and dampness in dorms and their associations with allergy among college students in China: a case-control study. *Indoor Air*, 21(4):277–283.

[130] Suryadhi, M. A. H., Abudureyimu, K., Kashima, S., and Yorifuji, T. (2019). Nitrogen dioxide and acute respiratory tract infections in children in Indonesia. *Archives of Environmental and Occupational Health*.

[131] Thatcher, T. L., Lai, A. C. K., Moreno-Jackson, R., Sextro, R. G., and Nazaroff, W. W. (2002). Effects of room furnishings and air speed on particle deposition rates indoors. *Atmospheric Environment*, 36:1811–1819.

[132] Thurston, G. D., Kipen, H., and Brunekreef, B. (2017). A joint ERS/ATS policy statement: What constitutes an adverse health effect of air pollution? An analytical framework. *European Respiratory Journal*, 49(1).

[133] Tian, J., Fang, C., Qiu, J., and Wang, J. (2021). Analysis of ozone pollution characteristics and influencing factors in northeast economic cooperation region, China. *Atmosphere*, 12(7):843.

[134] Tian, L., Zeng, Q., Dong, W., Guo, Q., Wu, Z., Pan, X., Li, G., and Liu, Y. (2017). Addressing the source contribution of PM2.5 on mortality: An evaluation study of its impacts on excess mortality in China. *Environmental Research Letters*, 12(10).

[135] United States Environmental Protection Agency (2023). Criteria Air Pollutants. *https://www.epa.gov/criteria-air-pollutants*.

[136] U.S. EPA (2009). Integrated Science Assessment for Particulate Matter. Technical report.

[137] Valuntaite, V., Šerevičiene, V., Girgždiene, R., and Paliulis, D. (2012). Relative humidity and temperature impact to ozone and nitrogen oxides removal rate in the experimental chamber. *Journal of Environmental Engineering and Landscape Management*, 20(1):35–41.

[138] Van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., Tamin, A., Harcourt, J. L., Thornburg, N. J., Gerber, S. I., Lloyd-Smith, J. O., De Wit, E., and Munster, V. J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *The New England journal of medicine*, 382(16):1564–1567.

[139] Van Ryswyk, K., Wallace, L., Fugler, D., MacNeill, M., Héroux, M. , Gibson, M. D., Guernsey, J. R., Kindzierski, W., and Wheeler, A. J. (2015). Estimation of bias with the single-zone assumption in measurement of residential air exchange using the perfluorocarbon tracer gas method. *Indoor Air*, 25(6):610–619.

[140] VC Elements (2023). VC Elements. *https://elements.visualcapitalist.com/.*

[141] Vu, T. V., Stewart, G. B., Kitwiroon, N., Lim, S., Barratt, B., Kelly, F. J., Thompson, R., Smith, R. B., Toledano, M. B., and Beevers, S. D. (2022). Assessing the contributions of outdoor and indoor sources to air quality in London homes of the SCAMP cohort. *Building and Environment*, 222.

[142] Wang, F., Meng, D., Li, X., and Tan, J. (2016). Indoor-outdoor relationships of PM2.5 in four residential dwellings in winter in the Yangtze River Delta, China. *Environmental Pollution*.

[143] Wang, S. (2003). The Current Conditions and Developing Tendency of Building Decoration Materials of America. *Development Guide to Building Materials*, pages 59–66.

[144] Wang, W., Chen, J., Jin, X., Ping, Y., and Wu, C. (2022). Association between indoor ventilation frequency and cognitive function among community-dwelling older adults in China: results from the Chinese longitudinal healthy longevity survey. *BMC Geriatrics*, 22(1):1–9.

[145] Wang, Y., Wild, O., Chen, X., Wu, Q., Gao, M., Chen, H., Qi, Y., and Wang, Z. (2020). Health impacts of long-term ozone exposure in China over 2013–2017. *Environment International*, 144.

[146] Wang, Z., Lv, J., Tan, Y., Guo, M., Gu, Y., Xu, S., and Zhou, Y. (2019). Temporospatial variations and Spearman correlation analysis of ozone concentrations to nitrogen dioxide, sulfur dioxide, particulate matters and carbon monoxide in ambient air, China. *Atmospheric Pollution Research*, 10(4):1203–1210.

[147] Wang, Z., Tan, Y., Guo, M., Cheng, M., Gu, Y., Chen, S., Wu, X., and Chai, F. (2023). Prospect of China's ambient air quality standards. *Journal of Environmental Sciences (China)*, 123:255–269.

[148] Wargocki, P. (2017). The Effects of Ventilation in Homes on Health. *International Journal of Ventilation*, 12(2):101–118.

[149] Weber, T. P. and Stilianakis, N. I. (2008). Inactivation of influenza A viruses in the environment and modes of transmission: a critical review. *The Journal of infection*, 57(5):361–373.

[150] Wei, L., Yang, E., Liu, Z., and Sun, X. (2017). Effects of Fog and Haze Weather on GPS Zenith Tropospheric Delay and Precise Point Positioning. *Journal of Geodesy and Geodynamics*, 37(11):1146–1149.

[151] Weschler, C. J. (2000). Ozone in indoor environments: Concentration and chemistry. *Indoor Air*, 10(4):269–288.

[152] Wilkins, E. T. (1954). Air pollution aspects of the London fog of December 1952. *Quarterly Journal of the Royal Meteorological Society*, 80(344):267–271.

[153] Woodward, H., Fan, S., Bhagat, R. K., Dadonau, M., Wykes, M. D., Martin, E., Hama, S., Tiwari, A., Dalziel, S. B., Jones, R. L., Kumar, P., and Linden, P. F. (2021). Air flow experiments on a train carriage—Towards understanding the risk of airborne transmission. *Atmosphere*, 12(10):1–19.

[154] World Health Organisation (2010). *WHO Guidelines for Indoor Air Quality: Selected Pollutants*. Geneva.

[155] World Health Organization (2021). WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Technical report.

[156] World Population Review (2024). House Size by Country 2024.

[157] Yang, B. Y., Qian, Z., Howard, S. W., Vaughn, M. G., Fan, S. J., Liu, K. K., and Dong, G. H. (2018a). Global association between ambient air pollution and blood pressure: A systematic review and meta-analysis. *Environmental Pollution*, 235:576–588.

[158] Yang, I. A., Fong, K. M., Zimmerman, P. V., Holgate, S. T., and Holloway, J. W. (2009). Genetic susceptibility to the respiratory effects of air pollution. *Postgraduate Medical Journal*, 85(1006):428–436.

[159] Yang, Z., Shen, J., and Gao, Z. (2018b). Ventilation and Air Quality in Student Dormitories in China: A Case Study during Summer in Nanjing. *International Journal of Environmental Research and Public Health*.

[160] Yao, J., Brauer, M., Wei, J., McGrail, K. M., Johnston, F. H., and Henderson, S. B. (2020). Sub-daily exposure to fine particulate matter and ambulance

dispatches during wildfire seasons: A case-crossover study in British Columbia, Canada. *Environmental Health Perspectives*, 128(6):1–10.

[161] Yorifuji, T., Kashima, S., Tsuda, T., Ishikawa-Takata, K., Ohta, T., Tsuruta, K. I., and Doi, H. (2013). Long-term exposure to traffic-related air pollution and the risk of death from hemorrhagic stroke and lung cancer in Shizuoka, Japan. *Science of the Total Environment*, 443:397–402.

[162] Yuan, Y., Li, S., Chen, T., and Ren, J. (2023). Effects of Ambient Temperature and Humidity on Natural Deposition Characteristics of Airborne Biomass Particles. *IJERPH*, 20(3):1–14.

[163] Zhang, H., Fan, Y., Han, Y., Yan, L., Zhou, B., Chen, W., Cai, Y., Chan, Q., Zhu, T., Kelly, F. J., and Barratt, B. (2022). Partitioning indoor-generated and outdoor-generated PM2.5 from real-time residential measurements in urban and peri-urban Beijing. *Science of The Total Environment*, 845:157249.

[164] Zhang, K. and Batterman, S. (2013). Air pollution and health risks due to vehicle traffic. *The Science of the total environment*, 0:307.

[165] Zhao, B., Bai, L., Wan, R., Wang, Y., and Qin, L. (2022). Exposure to second-hand smoke is an independent risk factor of small airway dysfunction in non-smokers with chronic cough: A retrospective case-control study. *Front Public Health*.

[166] Zhao, H., Gall, E. T., and Stephens, B. (2019). Measuring the Building Envelope Penetration Factor for Ambient Nitrogen Oxides. *Environmental Science and Technology*, 53(16):9695–9704.

[167] Zhao, Y. and Zhao, B. (2018). Emissions of air pollutants from Chinese cooking: A literature review. *Building Simulation*, 11(5):977–995.

[168] Zhao, Z., Chen, R., Lin, Z., Cai, J., Yang, Y., Yang, D., Norbackd, D., and Kan, H. (2016). Ambient carbon monoxide associated with alleviated respiratory inflammation in healthy young adults. *Environmental Pollution*, 208:294–298.

[169] Zhou, Z., Liu, Y., Yuan, J., Zuo, J., Chen, G., Xu, L., and Rameezdeen, R. (2016). Indoor PM2.5 concentrations in residential buildings during a severely polluted winter: A case study in Tianjin, China. *Renewable and Sustainable Energy Reviews*, 64:372–381.

Appendix A

Supplementary materials

A.1 Chapter 1 supplementary information

A.1.1 Publication: Air flow experiments on a train carriage - Towards understanding the risk of airborne transmission

atmosphere

Article

Air Flow Experiments on a Train Carriage—Towards Understanding the Risk of Airborne Transmission

Huw Woodward [1,*], Shiwei Fan [2], Rajesh K. Bhagat [3], Maksim Dadonau [3], Megan Davies Wykes [4], Elizabeth Martin [2], Sarkawt Hama [5], Arvind Tiwari [5], Stuart B. Dalziel [3], Roderic L. Jones [2], Prashant Kumar [5] and Paul F. Linden [3]

1 Centre for Environmental Policy, Imperial College London, London SW7 1NE, UK
2 Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK; sf588@cam.ac.uk (S.F.); em783@cam.ac.uk (E.M.); rlj1001@cam.ac.uk (R.L.J.)
3 Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, UK; rkb29@cam.ac.uk (R.K.B.); md732@cantab.net (M.D.); s.dalziel@damtp.cam.ac.uk (S.B.D.); p.f.linden@damtp.cam.ac.uk (P.F.L.)
4 Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK; msd38@cam.ac.uk
5 Global Centre for Clean Air Research (GCARE), Department of Civil & Environmental Engineering, University of Surrey, Guildford GU2 7XH, UK; s.m.hama@surrey.ac.uk (S.H.); a.tiwari@surrey.ac.uk (A.T.); p.kumar@surrey.ac.uk (P.K.)
* Correspondence: huw.woodward@imperial.ac.uk

Abstract: A series of experiments was undertaken on an intercity train carriage aimed at providing a "proof of concept" for three methods in improving our understanding of airflow behaviour and the accompanied dispersion of exhaled droplets. The methods used included the following: measuring CO_2 concentrations as a proxy for exhaled breath, measuring the concentrations of different size fractions of aerosol particles released from a nebuliser, and visualising the flow patterns at cross-sections of the carriage by using a fog machine and lasers. Each experiment succeeded in providing practical insights into the risk of airborne transmission. For example, it was shown that the carriage is not well mixed over its length, however, it is likely to be well mixed along its height and width. A discussion of the suitability of the fresh air supply rates on UK train carriages is also provided, drawing on the CO_2 concentrations measured during these experiments.

Keywords: airborne transmission; COVID-19; public transport; ventilation; aerosol dispersion

Citation: Woodward, H.; Fan, S.; Bhagat, R.K.; Dadonau, M.; Wykes, M.D.; Martin, E.; Hama, S.; Tiwari, A.; Dalziel, S.B.; Jones, R.L.; et al. Air Flow Experiments on a Train Carriage—Towards Understanding the Risk of Airborne Transmission. *Atmosphere* **2021**, *12*, 1267. https://doi.org/10.3390/atmos12101267

Academic Editor: Amir Aliabadi

Received: 1 September 2021
Accepted: 22 September 2021
Published: 29 September 2021

1. Introduction

The COVID-19 pandemic has resulted in a much decreased capacity on UK rail services, with physical distancing rules applied for much of the pandemic that force trains to operate at half capacity or less. According to Department for Transport (DfT) statistics, following the first UK lockdown, passenger numbers have remained well below capacity [1]. This is likely due to the large increase in the number of people working from home resulting in a reduction in passenger numbers during peak hours and people's tendency to avoid non-essential travel due to concerns regarding the risk of infection by the SARS-CoV-2 virus while travelling. Travel by public transport such as rail is perceived by some commentators as potentially high risk due to the potential of interacting with a large number of people at the station, the possibility of being in close proximity to other people during the journey, and the requirement of spending extended periods of time within a confined space with others.

These concerns may not be entirely unfounded and any train journey inevitably carries a degree of risk of infection by the SARS-CoV-2 virus, particularly as exposure to asymptomatic individuals is seen as a major mode of transmission [2]. Infection can occur via three routes: by contact with infected surfaces, droplet transmission, and airborne transmission [3–6]. Transmission via surface contact is mitigated by regular cleaning of touch points, the use of antimicrobial products that can provide a degree of residual protection, provision of hand sanitiser in stations, and promotion of regular hand washing. By "droplet transmission", we refer to transmission via large droplets exhaled while coughing or sneezing and also, to a lesser degree, while breathing or talking. Larger droplets fall to the ground within seconds [7,8]; therefore, enforcing physical distancing and, with sufficient levels of compliance, wearing masks are effective mitigation strategies against this transmission route. By "airborne transmission", we refer to the transmission of the virus via smaller particles that tend to follow the dominant airflow patterns within a space and can remain in the air for extended periods of time (minutes to hours). There exists considerable evidence to suggest that airborne transmission of SARS-CoV-2 is viable [9]; therefore, appropriate mitigation measures should be taken against this transmission route [3–6,10]. However, mitigation strategies continue to be mainly focused on surface contact and droplet transmission only.

The risk of airborne transmission is likely to be higher in small spaces such as a train carriages, especially if ventilation rates are low. In these cases, virus-laden particles exhaled by an infected person can accumulate and result in high doses for uninfected persons within the space. In addition to the provision of fresh outdoor air, the risk of airborne transmission is likely to depend on the airflow patterns within the space. For example, areas within a space with low airflow velocities could result in an accumulation of infected air with virus-laden aerosols. Thermal stratification within a space can restrict vertical mixing of air and result in high concentrations of virus-laden particles in certain zones, potentially including the breathing zone. Therefore, when assessing the relative risk of airborne transmission indoors, it is important to have an understanding of both the outdoor air supply and the airflow patterns within the space. Alternatively, CO_2 concentrations can be measured to provide an estimate of the concentrations of exhaled breath within the space, and the risk of transmission can be estimated in turn [11–13].

Public transport poses a unique challenge when evaluating the risk of airborne transmission and when seeking to establish suitable mitigation strategies. Firstly, the ventilation rates for vehicles with some degree of natural ventilation (i.e., windows that can be opened) can vary significantly depending on the speed of travel and the number of open windows. The exchange of air when doors are opened to allow passengers to board and alight should also be considered and also applies to vehicles that are otherwise entirely mechanically ventilated. Estimating the ventilation rate on buses or trains is, therefore, often difficult. The problem is compounded by the wide range of carriage types in operation, which all have different dimensions and ventilation configurations. Secondly, plumes driven by body heat can have a significant impact on the airflow patterns within vehicles [14–18], which in turn can vary depending on the number and location of occupants. Of course, the distribution of the airborne pathogen within the vehicle can depend on the location of the infectious individual. The heating of surfaces by solar radiation can also affect airflow within the vehicle on sunny days [14,15,19]. Furthermore, the movement of people has been shown to have a considerable impact on the mixing of airborne contaminants indoors [20] and is likely to play a significant role in vehicles, particularly during busy periods. Finally, buses and trains are long and narrow with few access points. Therefore, it is often difficult to avoid being in close proximity with other passengers. Furthermore, the long, narrow shape is likely to cause the airflow patterns within the space to be particularly sensitive to the location of the air inlet and extract vents.

An understanding of the outdoor air provision in addition to the airflow patterns within public transport vehicles is, therefore, essential in order to effectively mitigate the risk of airborne transmission. The literature on airflow patterns within buses and trains is focused mainly on assessing the thermal comfort of passengers, e.g., [14–19]. In [21,22], respiratory droplet transmission within a train carriage is modelled using Reynolds-Averaged Navier–Stokes (RANS) computational fluid dynamics (CFD). These consider different ventilation configurations for carriages of high-speed trains in China. The range of dispersion of droplets, in addition to their residence time within the carriage, is shown to vary significantly for the different ventilation outlet positions considered. CFD studies of airborne transmission on a bus include [23,24], who demonstrated the sensitivity of the likelihood of transmission on a bus to the location of the infected person relative to the extract vent. Considerable effort has been made in understanding and modelling the airborne transmission of pathogens within aircraft cabins, e.g., [25–28]. Mazumdar and Chen [27] used a one-dimensional diffusion model to predict the concentrations of a gaseous contaminant along the length of an airliner cabin, while [29,30] used a zonal model. In [31], aerosol droplets were released within the cabin of aircraft, and concentrations were measured at various locations. It was shown that the very high air recirculation rates within commercial aircraft are very effective in diluting aerosol particle concentrations. In [32–34], it was shown that acceleration-induced body forces occurring during both the climb and descent stages of flight can affect the dispersion of a contaminant and the resulting exposure of passengers.

While examples of CFD studies of airflow behaviour in vehicles can be found in the literature, there are few examples of full-scale experiments in public transport vehicles. In this paper, we outline the experimental procedure for three experiments implemented on an inter-city train carriage in the UK. These experiments included measuring CO_2 generated by volunteer "passengers", flow visualisations of artificial smoke released within the carriage, and the measurement of the concentrations of aerosols released from a nebuliser. The aim in each case was to improve our understanding of the airflow patterns and aerosol dispersion within the carriage and to determine the utility of each method. Time for planning and executing these experiments was limited; therefore, some of the experimental standards normally expected were not met. For example, we could not perform the desired number of repeat runs. However, sufficient data were gathered to provide a demonstration of "proof of concept", while also providing insights into the airflow behaviour within the carriage.

2. Materials and Methods

2.1. Carriage Layout and Ventilation

The carriage used for the experiments consists of a passenger saloon and two vestibules at either end. This carriage layout is similar to other types of intercity carriages running on the GB rail network, but it is different from a typical carriage on a commuter/regional network, which are more common and more heavily used. The saloon takes the majority of the space inside the carriage (Figure 1) and includes 88 seats. The volume of the saloon is approximately 113 m^3, while the volume of the entire carriage is approximately 140 m^3. The majority of seats are in an "airline" configuration, however, some seats face each other across a table. There is a door at both ends of the saloon, which leads to a vestibule. These internal doors open automatically when approached but are otherwise shut. Each vestibule has a door on either side for passenger boarding and alighting, along with a third door for access to the next carriage.

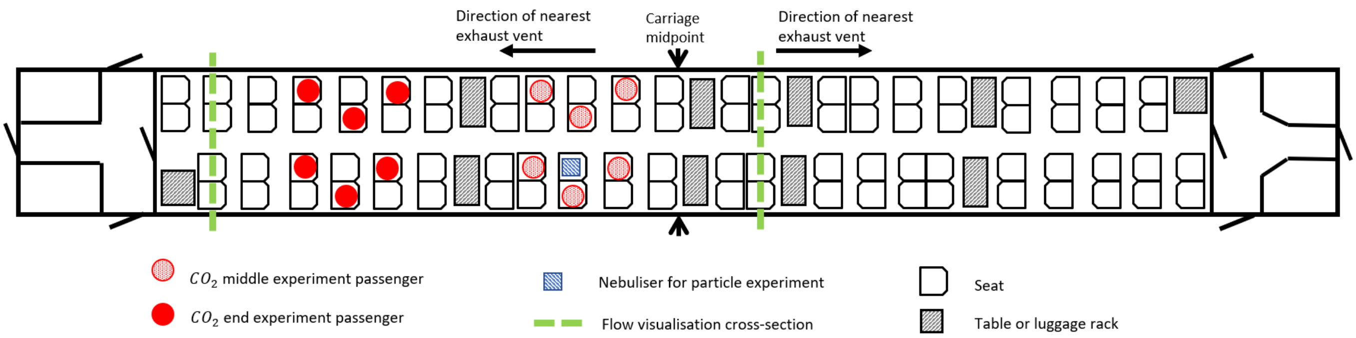

Figure 1. Schematic of carriage seat layout and experimental layouts for this study. Red circles indicate the position of passengers for the end experiments (full) and middle experiments (shaded).

The carriage is mechanically ventilated, and the windows within the carriage cannot be opened. Two Heating, Ventilation, and Air Conditioning (HVAC) units are located on the roof, one at each end of the carriage. These units provide a supply of outdoor air to the saloon and vestibules, drive the flow of conditioned air within the carriage, and can either heat or cool the air as required. The conditioned air, which is a mixture of outdoor air and recirculated air, is vented into the saloon from the ceiling along the entire length of the carriage (Figure 2). The extract vent for each ventilation unit is located on the ceiling at each end of the saloon. Heaters are located near the floor along each wall of the saloon to provide additional heating capacity to supplement that the ventilation system.

The carriage ventilation system can be set to operate in several different modes: automatic heating or cooling and forced cooling or forced heating. The outdoor temperature was above 30 °C during the days of these experiments; therefore, forced cooling was chosen as the ventilation setting used for each experiment. Forced cooling can be run at several different cooling rates, from 0% to 100%. A 75% cooling capacity was used, and the desired temperature within the saloon was set to 21 °C at the HVAC control unit. The exact flow rates of outdoor air provided to the carriage and the flow rates of recirculated air were not known. However, the design specification of the carriage HVAC system specified a fresh air supply in the range of 22.5–30 m^3 min^{-1} and a recirculation flow rate of 30–60 m^3 min^{-1}. The heaters remained switched off at all times.

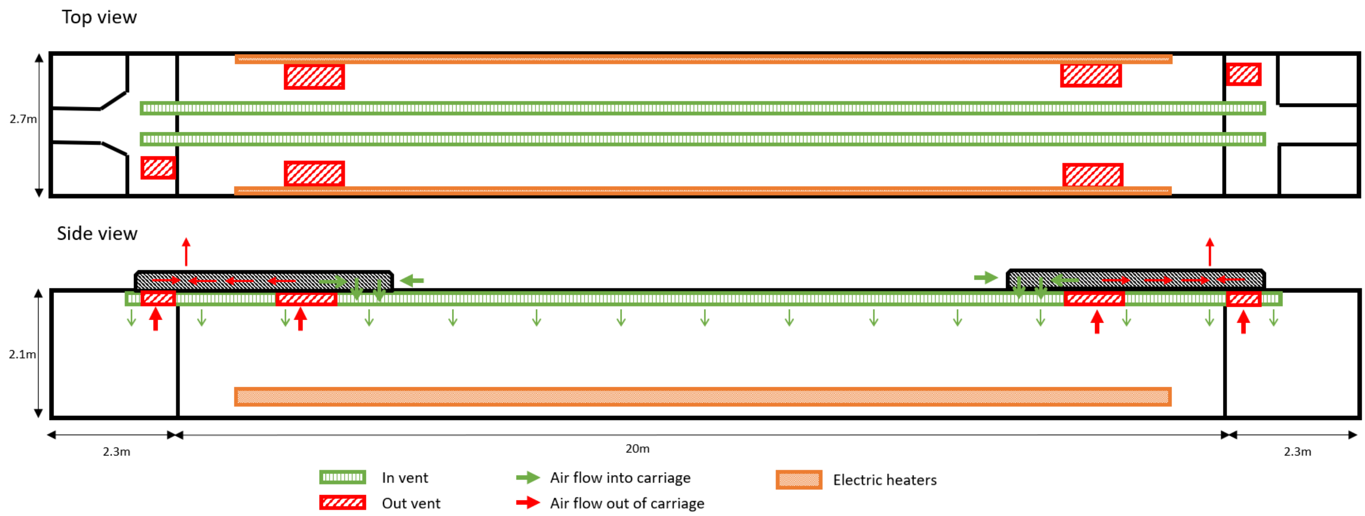

Figure 2. Schematic of carriage ventilation. Air is supplied from the ceiling along the entire length of the carriage. Extract vents are located at either end of the saloon and in each vestibule. A large proportion of the air is recirculated back into the carriage.

2.2. Outline of Experiments

2.2.1. CO_2 Experiment

The objective of the CO_2 experiments was to explore the feasibility of using CO_2 generated by exhaled breath and CO_2 sensors to resolve concentration differences within the saloon and to observe whether ventilation removed any stratification.

The CO_2 experiments involved the generation of CO_2 by members of the research team representing passengers and sitting in the carriage while the ventilation was switched off, allowing CO_2 concentrations to rise before switching the ventilation on and measuring the decay rate of CO_2 at several locations. Six members of the research team acted as passengers and sat in the carriage for a total of 35 min at a time. The passengers were arranged to maximise the distance between each passenger while occupying three rows of seats. This resulted in a staggered formation as shown in Figures 3 and 4, as well as Figure 1. The ventilation was switched off at the start of the experiment. After 15 min, the ventilation was switched on and used the 75% forced cooling setting. The passengers remained in their seats for an additional 20 min before the experiment was stopped. The experiments were carried out at two locations within the saloon: the first near the midpoint of the saloon and the second near the end of the saloon. Both experiments were run twice.

Seven CO_2 sensors (K33-LP T, SenseAir AB, Delsbo, Sweden) were placed at various locations within the carriage (Figures 3 and 4). The sensors were calibrated with a reference analyser (G2201-i, Picarro Inc., Santa Clara, USA). The percentage error of reading was within 3% in the range of 0–3000 ppm. These sensors are labelled M1 to M7. Sensors M3 and M5 were placed on the backs of seats at the height of the typical breathing zone for sitting passengers and in close proximity to the six passengers. M6 and M7 were also placed at the back of a seat but at a greater distance from the passengers. M4 was placed on the luggage rack, while M1 was attached to the ceiling. Sensors M1, M4, and M5 were all located at the same distance along the length of the carriage. For the first experiment near the midpoint of the carriage, M2 was placed on the back of a seat on the opposite side to M6, closer to the extract vent. For the experiment at the end of the carriage, M2 was placed on the luggage rack above M7 and directly below the extract vent.

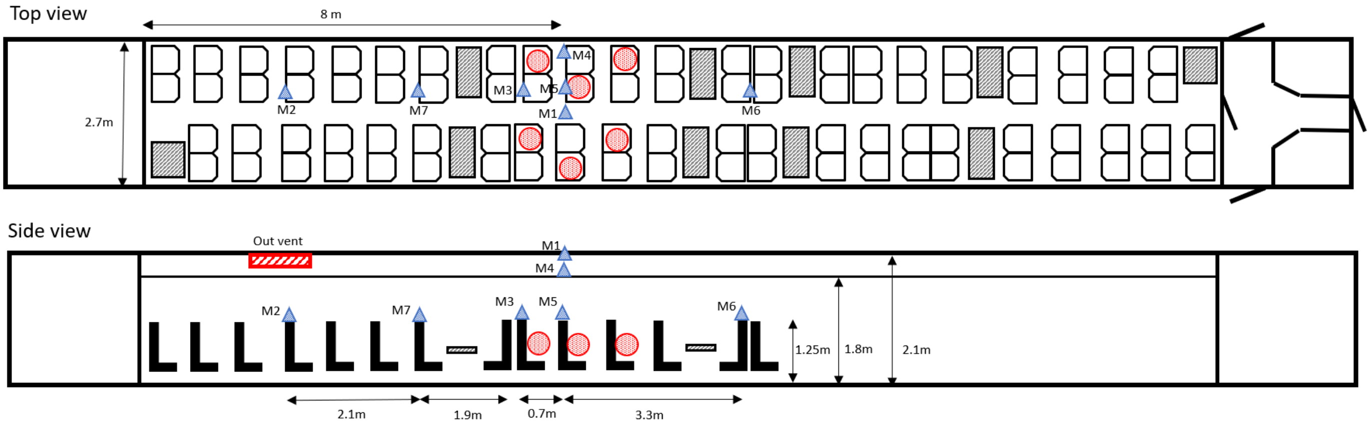

Figure 3. CO_2 experiment set up for middle of carriage. Red circles indicate position of passengers, and blue triangles indicate position of CO_2 sensors.

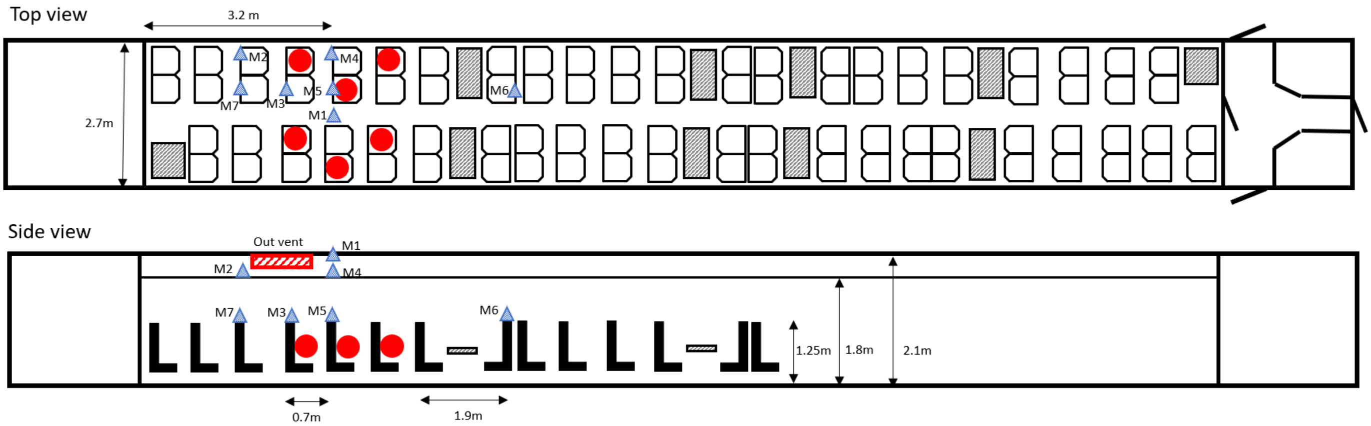

Figure 4. CO_2 experiment set up for end of carriage. Red circles indicate position of passengers, and blue triangles indicate position of CO_2 sensors.

The experiments reported here were performed during the ongoing COVID-19 pandemic. The experimental procedure was, therefore, complicated by the necessity to mitigate the risk of infection from any potentially asymptomatic participating researchers. With this in mind, the time period during which the passengers were asked to sit in close proximity with the ventilation off was limited to 15 min. An additional 20 min was allowed with the ventilation switched on. Due to these relative short time periods, a steady state in CO_2 concentrations was never reached, either during the ventilation off or ventilation on period. Furthermore, due to the limited time available on the carriage, only two runs of each experiment were performed.

2.2.2. Aerosol Dispersion Experiment

The aim of the particle dispersion experiment was to map the aerosol distribution over adjacent seats during a continuous release of aerosols under the ventilation off and on conditions. The size of exhaled droplets ranges between 0.01 and 1000 µm [35]; therefore, it is important to consider the dispersion of different aerosol size fractions. A nebuliser was used as a source to generate continuous aerosols made up of sodium chloride solution (salt; 1% by weight) at a flow rate of 6 L per minute, which is within the range of the human breathing rate, typically 5–7 L/min^{-1} while resting [36]. The use of a nebuliser allowed us to investigate the significance of droplet mass on dispersion. The aerosol particles released by the nebuliser had a size range of 0.25 µm to 16.5 µm. Six laser particle counters (Dylos1700) were used to measure concentrations of fine ($PM_{2.5}$; aerodynamic diameter $\leq$ 2.5 µm) and coarse (PM_{10}; aerodynamic diameter $\leq$ 10 µm) aerosol particles at different adjacent seats (Figure 5). These aerosol monitors have been successfully deployed in previous work [37,38]. As for the CO_2 experiments, a 75% forced cooling setting was used for the "ventilation on" period of these experiments. There were no passengers present during this experiment; therefore, the effects of body plumes and people movement were not considered.

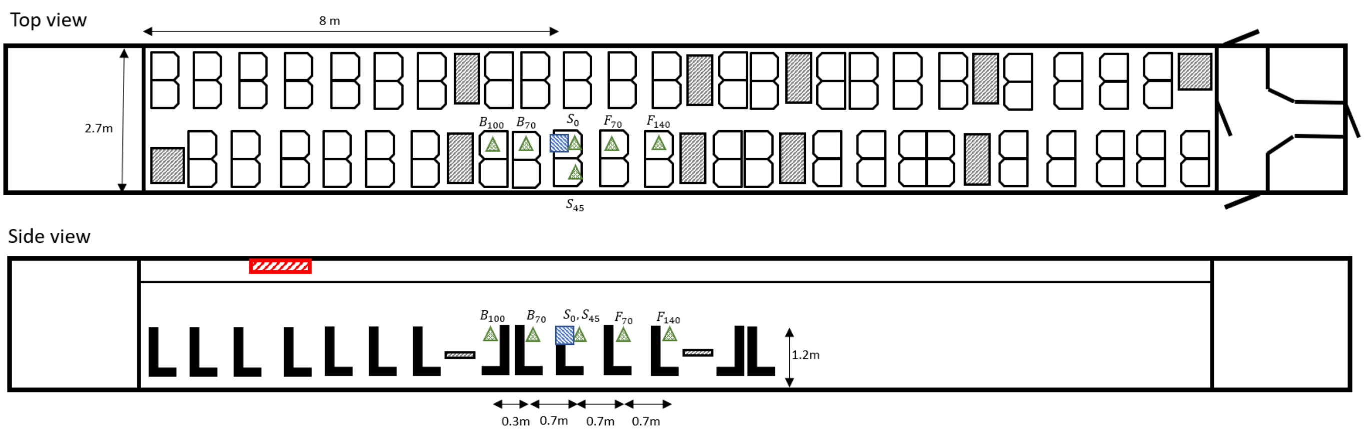

Figure 5. Aerosol dispersion experiment setup. The location of the nebuliser is shown by the blue square and the PM sensors are shown by green triangles.

As part of quality control and assurance process, we carried out co-location measurements over a period of 8 h prior to the experiments in order to assess relative accuracy. Pearson correlation coefficients (r) greater than 0.93 and 0.87 were observed for $PM_{2.5}$ and PM_{10} , respectively, as observed in Figure A1.

The aerosol monitors were mounted on the back of seats at the typical breathing height of a sitting passenger (1.2 m above the floor) at various locations near the nebuliser (Figure 5). A total of six sets of experiments, each for 25 min, were conducted under both "ventilation off" (indoor cabin temperature, T = 32.3 $\pm$ 1.2 °C; relative humidity, RH = 39.5 $\pm$ 2.9%) and on (T = 30.8 $\pm$ 0.2 °C; RH = 40.3 $\pm$ 2.5%) conditions. Before each experiment, the ventilation was switched on for 15 min to clear the accumulated concentrations over the measurement duration and to reach a stabilised background aerosol concentration level. Another 5 min was allowed before the start of each experiment after switching the ventilation on or off in order to allow the carriage flow to reach a quasi-steady-state condition.

The six aerosol monitors located at different locations within a cabin were marked as B_{100}, B_{70}, S_0, S_{45}, F_{70}, and F_{140}; B, F, and S refer to behind, in front, and the same row as the seat on which the nebuliser was placed, respectively. The nebuliser release faced towards the "front" direction. The subscript indicates the horizontal distance in cm with respect to the source (Figure 5). For example, F_{70} and B_{70} indicate that these aerosol monitors were placed in front of and behind the source seat, respectively, at a distance of 70 cm from the source (S_0). S_{45} refers to the seat in the same row as S_0 at a distance of 45 cm. The monitor at S_0 was placed directly below the nebuliser outlet, within a few centimetres. The nearest ventilation extract was located behind the nebuliser's seat.

The following equation was used to normalise the measured aerosol concentrations in order to understand the aerosol concentrations in relative terms so that the concentration ranged between 0 and 1.

$$C_{norm} = \frac{\text{Average aerosol concentration at a location}}{\text{Average aerosol concentration at source}}. \quad (1)$$

The R statistical software (R Core Team, 2019) in the Open-air software package [39] was used to carry out data processing and statistical analyses.

2.2.3. Airflow Visualisation

The dominant flow patterns across the width and height of the carriage were visualised both near the middle and the end of the saloon. The locations of the visualised cross-sections are indicated in Figure 1 by the green dashed lines. Flow pattern visualisation was conducted by tracing the motion of a neutrally buoyant, inert fog under the ventilation flow.

Lasers (30 mW; 520 nm) were fitted with Powell lenses in order to form a diverging laser sheet and then mounted and aligned in order to illuminate the fog across a carriage cross-section. In order to enhance visualisation, the lights in the carriage were switched off, and plastic blackout sheets were used to cover the windows. Initially, the ventilation within the carriage was switched off. A section of the carriage approximately 2 m in length was isolated by using curtains before being filled with non-toxic, artificial, and theatrical smoke consisting of 70% water and 30% glycol droplets of size ranges between 5 and 10 µm. Once the carriage section was suitably filled, the plastic sheets were removed, and two passengers sat on either side of the aisle such that the illuminated cross-section of the carriage passed over their shoulders and heads (see Figure 6). Once in position, the flow generated by the body plumes of the passengers was allowed to develop before the ventilation was switched on (75% forced cooling). The flow was visualised until the fog was dispersed to a degree that visualisation was no longer effective. Typically, this allowed a minute or two of visualisation.

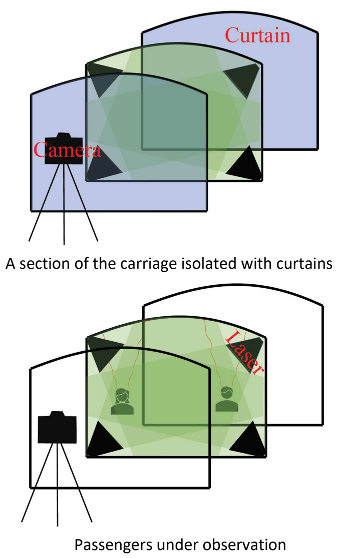

Figure 6. Schematic of the flow visualisation setup. A 2 m long section of the carriage was isolated using curtains and filled with inert smoke. Once the section was filled with smoke, the curtains were removed, allowing passengers to enter the section and sit across the aisle from each other on seats illuminated with laser lights. The body heat of the passengers produce thermal plumes, andventilation was subsequently switched on. Cameras were used to record the movement of the tracer smoke along with the dominant flows.

3. Results

3.1. CO_2 Experiment

Figure 7 shows the CO_2 measured by all sensors over the period beginning at the start of the first experiment and ending at the end of the last experiment. The figure includes uncontrolled periods in between experiments. The dashed lines indicate the beginning of each experiment when the ventilation was switched off, the time at which the ventilation was switched on, and the end of the experiment when the passengers were permitted to move from their seated positions. The first two experiments, conducted before 1330, were at the middle of the saloon (Figure 3), while the second two experiments, after 1430, were at the end of the saloon (Figure 4).

The period of each experiment is evident from the increase in CO_2 concentrations when the ventilation was switched off, followed by the rapid decrease in concentrations when the ventilation was switched on. These periods are also indicated by the dashed vertical lines. It is also evident that a steady-state was not reached at any point.

It is useful to see the data shortly before the beginning of each experiment as they highlight a limitation, which is that concentrations varied significantly between the start point of each experiment. No effort was made to control the period up to the beginning of each experiment. Therefore, the number of people present prior to each experimental run could vary in addition to the time period and ventilation setting used between each experiment, resulting in a variation in the initial CO_2 concentrations at the beginning of each run. This variation is evident when considering the concentrations at the start of the first and second run for both the middle and end experiments. Concentrations were lower at the start of the second run in both cases as these began shortly after the period of forced ventilation from the first run, which resulted in a significant decrease in concentrations. These differences in initial conditions are reflected in the concentration trends observed during the experiments. For example, the rate of increase in CO_2 was greater for the second run in both cases as the initial concentrations were lower. Despite the higher rate of increase, concentrations were generally still lower after 15 min, with no ventilation for the second runs.

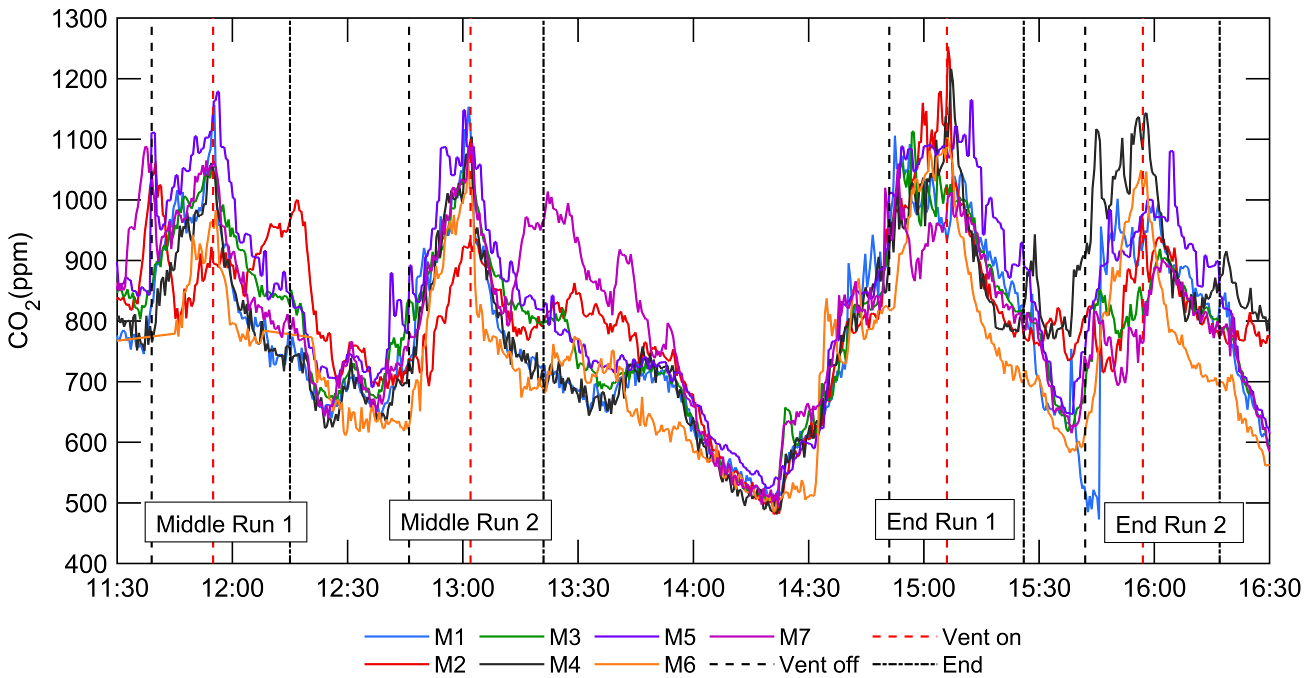

Figure 7. CO_2 measurements for all sensors between the start of the first experiment and the end of the last experiment. The first two experimental runs were conducted near the middle of the carriage, while the last two were conducted near the end of the carriage and an extract vent. Following "Middle Run 2", the occupancy of the carriage reduced to zero, resulting in a large drop in CO_2. Passengers returned to the carriage shortly before 14:30, resulting in an increase in CO_2.

A large decrease in concentrations was observed over the lunch period, between 1330 and 1420, during which time the saloon was empty, and the ventilation was set to automatic mode. At 1420, some people returned to the carriage and sat near the senors, from which point there was a sharp increase in concentrations. The maximum number of people in the saloon at any given time was six.

Figure 8 shows the concentrations for the sensors placed at breathing zone height for sitting passengers. The highest concentrations were generally observed for M5 located between two rows of passengers, with the concentrations here consistently above those at M7 and M3. The distance between M3 and M5 is 0.8 m. Once ventilation was switched on at 900 s (15 min), the concentrations began to decrease at each location. The rate of decrease was highest for M6 (the sensor furthest from the nearest extract vent) for all experiment runs. For the experiment at the middle of the saloon, a much lower rate of decay was observed for M2 (the sensor closest to the nearest extract vent) despite being located further from the CO_2 source. In fact, for the first run, an increase in concentrations was observed for M2. This suggests that the CO_2 generated by the passengers travelled towards the nearest extract vent, shown in Figures 3 and 4; when the ventilation was switched on, the elevated CO_2 at M6 quickly diluted, while the reduction due to dilution at M2 was countered by elevated concentrations advected by air flow from the direction of the passengers. The rates of decay for the M3, M5, and M7 sensors are higher at the middle of the carriage than compared to the end. This suggests that the effective ventilation rate is higher at the middle of the carriage, at least initially.

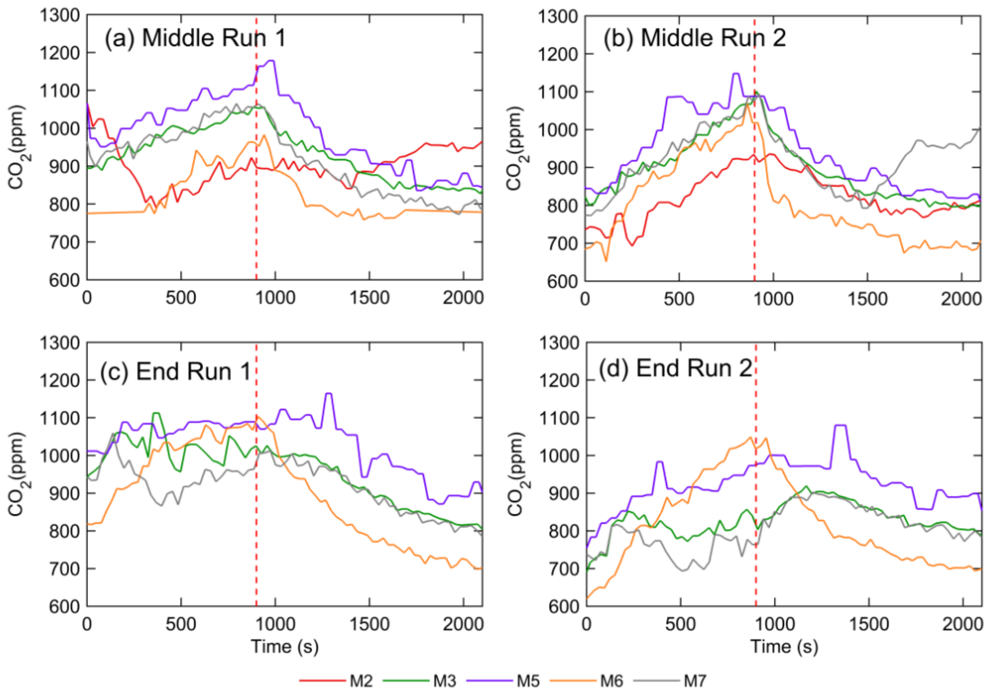

Figure 8. CO_2 concentrations for sensors placed within breathing zone for (**a**) middle run 1, (**b**) middle run 2, (**c**) end run 1 and (**d**) end run 2. Red dashed lines show time at which ventilation is switched on.

Figure 9 shows the concentrations for sensors placed at different heights at the same location along the length of the saloon. For the experiment at the middle of the carriage, this constituted M1 (ceiling), M4 (luggage rack), and M5 (breathing zone) only. Here, no significant difference was observed between M1 and M4, however, M5 showed consistently higher concentrations. A similar picture was observed for M1, M4, and M5 at the end of the saloon. At the end of the saloon, M2 was located directly above M7. In this case, M2 showed higher concentrations than M7, indicating that there may be stratification during the unventilated period. Once the ventilation was switched on, the concentrations at the two locations quickly converged, suggesting that the ventilation was effective at mixing the air vertically.

The steady-state was not reached for these experiments; threrefore, we were unable to draw firm conclusions based on the absolute differences in concentrations between locations as it was unclear to what extent these differences would have converged given sufficient time. However, it seems likely that the concentration at certain locations would have remained higher than others. For example, it is likely that the steady-state concentration for M6 would have been significantly lower than for all other sensors based on the trends observed in Figure 8. This suggests that the saloon was not well mixed along its length despite the supply of recirculated air throughout the saloon, but it was well mixed over its height.

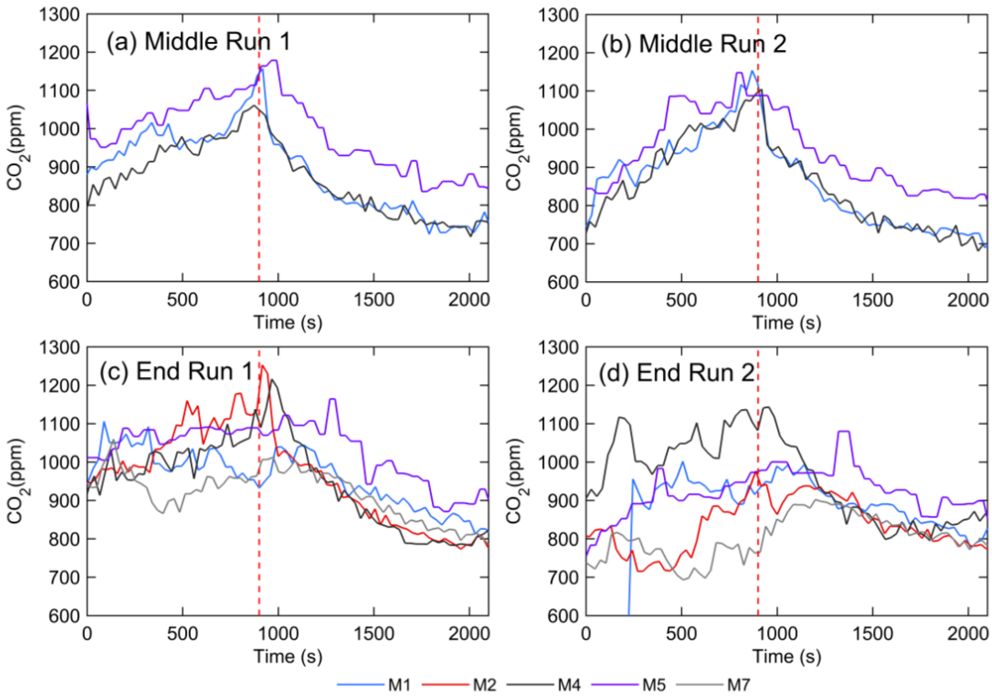

Figure 9. CO_2 concentrations for sensors placed at different heights for (**a**) middle run 1, (**b**) middle run 2, (**c**) end run 1 and (**d**) end run 2. Red dashed lines show time point at which ventilation is switched on.

3.2. Aerosol Dispersion

Figure 10 shows the normalised mean concentrations measured for the aerosol released from the nebuliser during the ventilation off and on periods at each location. During the unventilated period, the PM_{10} concentration dropped off very quickly from the source location by a factor of nearly nine between S_0 and the next nearest monitor, S_{45}. The relative decrease was much greater than that observed for $PM_{2.5}$ concentrations, which decreased by 40% between these two locations. This highlights the difference in the dispersion of the smaller and larger particles. During this unventilated period, there were no advective flows present within the carriage, and the dilution of the aerosol occurs due to diffusion and small scale turbulent mixing. Due to their greater mass, the larger aerosol particles are not dispersed as effectively as the smaller particles under these conditions. The low relative humidity in the carriage and the use of salt solution to generate the aerosols will have resulted in evaporation, resulting in a decrease in droplet size, which will also have contributed to the difference observed between the size fractions as some of the initially larger droplets reduced in size and became attributed to the smaller size fraction.

Switing on the ventilation resulted in a significant reduction in concentrations at all locations (see Table A1); mixing was increased due to advection and increased turbulence within the carriage. The largest relative decrease of 72% was observed at S_0 for the coarse aerosol, highlighting the effectiveness of the ventilation in driving the dilution of these larger particles. However, the normalised concentrations of the fine particles remained greater than those of the coarse particles at each location away from the source, indicating the more effective mixing of the finer fraction.

When the ventilation was switched on, higher concentrations were observed for both the coarse and fine fraction in the "backward" direction than compared to the "forward" direction. This suggests that the prevailing flow was directed in the "backwards" direction. This was the direction towards the nearest extract vent in the saloon, as also observed from the CO_2 measurements.

Finally, the concentrations at S_{45}, which was located at the seat next to the source, were only marginally greater than those at B_{70} located on the next row. This suggests that, over these short length scales, the degree of mixing across the width of the saloon was similar to that along its length.

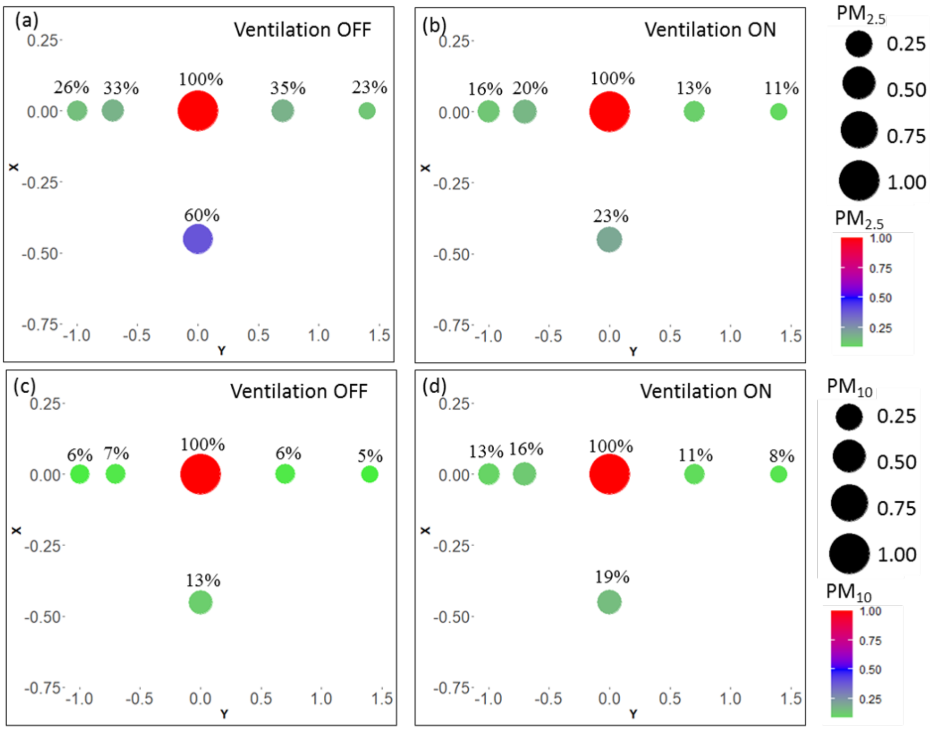

Figure 10. Spatial distribution of averaged normalised (**a**,**b**) fine and (**c**,**d**) coarse aerosol particle concentrations at each location under unventilated and ventilated conditions.

3.3. Flow Visualisation

Figures 11–14 show still images of the flow visualisation experiment at the middle and end of the carriage, respectively. Red arrows are used to indicate the direction of persistent air flow. Videos of these flow visualisations are available in the Supplementary Materials for which the airflow patterns are clearer.

During the unventilated period, the body plumes rising from the two passengers are clearly visible in the video footage for both the middle and end cases. In the middle of the carriage, there were no other persistent flows present other than some turbulent mixing. At the end of the carriage, there was a weak but persistent downward flow from the ceiling above the passenger on the right. This flow in turn forced the body plume from the passenger on the right to rise at an angle towards the centre of the carriage. This downward flow may have been due to an asymmetry in the body plumes generated by the two passengers.

In the middle of the carriage, when the ventilation was switched on, a strong downward jet was observed to flow from the ceiling inlet vents (Figure 2). This downward flow was sufficiently strong to extend between the two passengers and beyond the lower edge of the image, acting as an air curtain between the passengers. The body plumes rising from the two passengers continued to drive an upward flow while the ventilation was switched on, causing significant upward acceleration.

At the end of the carriage, when the ventilation was switched on, the flow patterns were very different compared to those observed in the middle of the carriage. In this case, only a very weak downward jet was observed to flow from the inlet vents, extending only a few centimetres into the space, while a persistent upward flow was observed across the remaining cross section of the carriage. In this case, the passengers were sat directly below the extract vents. The dominant upward flow is driven by the suction of these vents.

Figure 11. Snapshot of flow visualisation experiment at the middle of carriage during unventilated conditions. Red arrows indicate direction of persistent air flow.

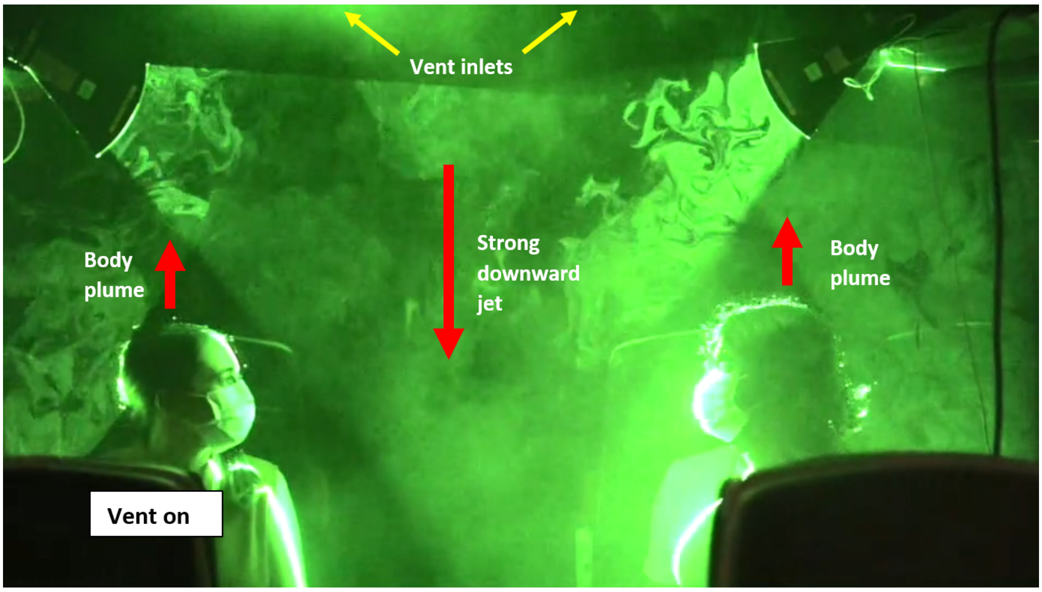

Figure 12. Snapshot of flow visualisation experiment at middle of carriage during ventilated conditions. Red arrows indicate direction of persistent air flow. Yellow lines indicate location of inlet and extract vents.

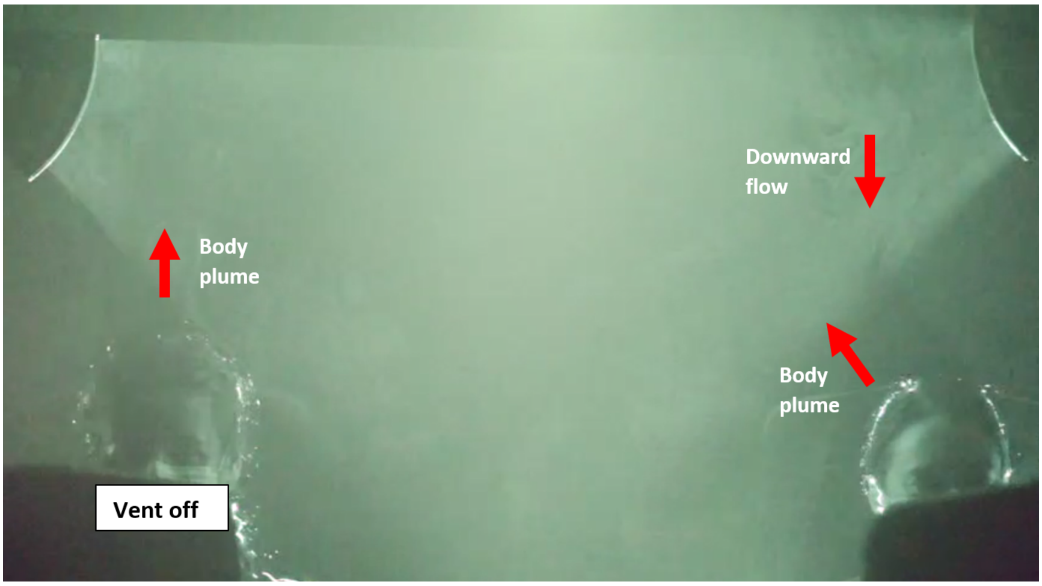

Figure 13. Snapshot of flow visualisation experiment at end of carriage during unventilated conditions. Red arrows indicate direction of persistent air flow.

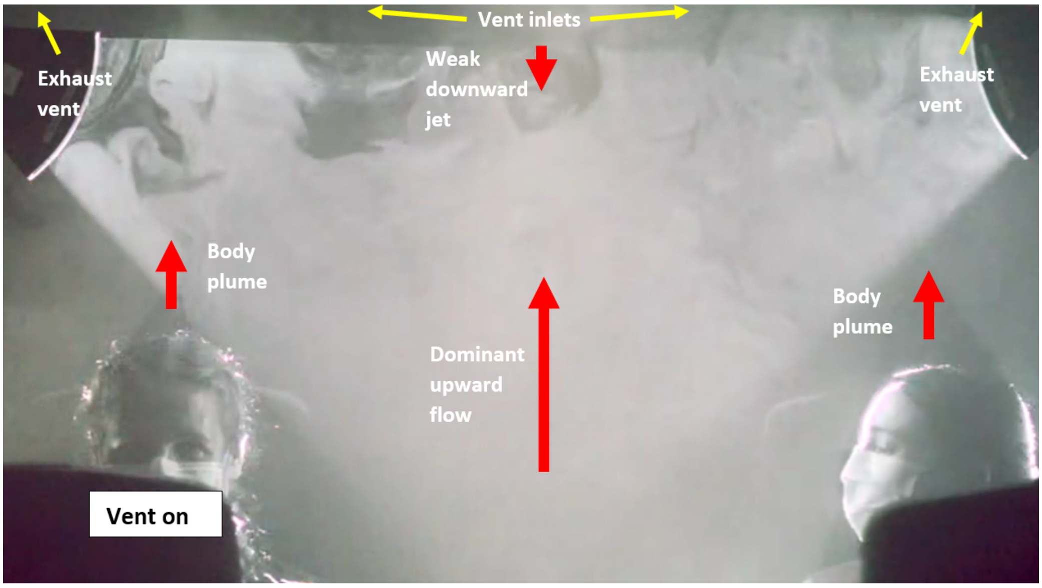

Figure 14. Snapshot of flow visualisation experiment at end of carriage during ventilated conditions. Red arrows indicate direction of persistent air flow. Yellow lines indicate location of inlet and extract vents.

4. Discussion

Given that only six people were used for the CO_2 experiments, relatively high concentrations were measured in the carriage. While the steady-state was not reached during the CO_2 experiments, it is clear from Figures 8 and 9 that concentrations near the passengers (M3, M5, and M7) converge towards a value of around 800 ppm. Given that the carriage has a seated occupancy of 88, the concentrations are likely to be considerably higher in a busy carriage. CIBSE recommends an outdoor air flow rate for buildings of 10 L s^{-1} person^{-1} (Ls^{-1}p^{-1}) [40]. When achieved, this ensures that CO_2 concentrations are unlikely to exceed 1000 pmmin a well-mixed space (concentrations above which have been related to adverse health impacts [41]). The ventilation flow rate was not known for these experiments, and it is not clear how the provision of fresh air provided by the 75% forced cooling setting used for these experiments compares with that provided by the automatic function of the ventilation system during normal service. While ventilation rates can be estimated from CO_2 decay curves or steady-state concentrations (e.g., [42]), these estimates depend on a well mixed assumption, which is not the case for the CO_2 experiments here. However, the design specification of the HVAC system placed the minimum and maximum ventilation rates, that is, the rate of supply of fresh air, at 22.5–30 m^3 min^{-1}. During normal operation, the exact rate varies within these limits in response to temperature senors; however, the system does not react directly to the occupancy of the carriage. Taking the carriage volume of 140 m^3, this equates to 9.6–12.9 air changes per hour (ACH). If we assume a carriage at half seating occupancy, holding 44 passengers, this works out as 8.5 to 11.4 $Ls^{-1}p^{-1}$. For a carriage at full seating occupancy, these values will be halved. In a recent review by the National Engineering Policy Centre (NEPC), values as low as 4 to 6 ACH were given for the provision of outdoor air to certain UK rail carriages [43]. Assuming the same carriage volume of 140 m^3 and an occupancy of 44 passengers, this works out as 3.5 to 5.3 $Ls^{-1}p^{-1}$, which is significantly lower than that recommended by CIBSE for buildings, but comparable to the ASHRAE recommended flow rates for commercial aircraft of 3.5 $Ls^{-1}p^{-1}$ (ASHRAE Standard 161). However, for a busy carriage, the air flow rate per person will be significantly lower.

Train carriages are not required to meet the same ventilation standards as indoor spaces in buildings. Ventilation systems on train carriages tend to be optimised for energy efficiency and passenger comfort rather than air quality. To minimise the energy consumption of HVAC systems on trains, much of the air supply is recirculated air rather than outdoor air that usually requires a higher degree of heating or cooling in order to maintain passenger comfort. The provision of outdoor air by these HVAC systems can, therefore, be low, and the recirculation of air could result in the dispersion of virus-laden particles throughout the space. While the recirculated air will be passed through a filter within the HVAC unit, most filters are too coarse to remove smaller viral particles [44]. Unlike aircraft, which are fitted with High Efficiency Particle Arrestance (HEPA) filters [28], this is not a requirement for train carriages. It is not clear from the experiments performed in this study what effect the recirculation of air has on the risk of transmission. For trains, European Union (EU) regulations and those adopted by the UK's Rail Safety and Standards Board specify that CO_2 concentrations should not exceed 5000 ppm (EU regulation No 1302/2014); however, there are no further requirements regarding indoor air quality. Given that CO_2 concentrations, together with the occupancy and HVAC filter efficiency, can be directly related to the risk of airborne transmission [11–13], the absence of more stringent regulations may be a cause for concern in terms of mitigating airborne transmission in addition to general air quality considerations. Further investigation is required to determine the efficacy of the HVAC filter in removing viral-laden particles from the air.

Within the context of the current COVID-19 pandemic, it should be noted that the risk of airborne transmission relative to that via droplets or contaminated surfaces is still not well understood. However, it is by now clear that airborne transmission is a significant component, as is now acknowledged by the World Health Organization [45]. The degree to which increasing the fresh air supply rate within a space reduces the risk of airborne transmission depends on the airflow structures within the space [46], the main factor being what proportion of the additional fresh air supplied reaches the breathing zone. However, given the experiments presented here suggesting that the carriage is well mixed along the vertical direction, it is likely that an increase in fresh air supply will result in a reduction in transmission risk. To what extent and whether adjusting the ventilation rates is a sensible measure remain outstanding questions that require further research. Train operators in the UK have taken practical measures currently available towards minimising the risk to passengers while travelling during the pandemic. These measures include the use of antimicrobial surface treatment, encouraging passengers to sit as far as possible from others, and enforcing mask wearing at all times. Furthermore, the risk of airborne transmission is limited by the short time periods typically spent in train carriages relative to other environments, for example, in buildings. It is also worth noting that public transport will, on the whole, have lower viral emission risk factors as most people tend to be passive while travelling rather than talking or exercising, which increase viral emissions considerably [47]. For these reasons, while we have compared the fresh air supply rates on train carriages to those recommended for buildings to provide context, we are not necessarily suggesting that equivalent rates are necessary or practical for train carriages. It is also for these reasons that the risk of infection on an individual basis on a train carriage is likely to be low. However, given the large number of passengers who travel by rail every day, the contribution to the population level "R" rate may be significant and justifies further investigation.

The experiments revealed the complexity of the airflow patterns and, therefore, the dispersion of particles within the carriage saloon. Significant differences were observed in the CO_2 concentrations within the saloon along its length. Therefore, we can conclude that the air within the saloon is not well mixed along its length, at least not while the train is stationary or while travelling at steady speed. Maximising the physical distance between passengers along the length of the carriage is, therefore, likely to be an effective strategy at reducing the risk of airborne transmission. A downward jet observed in the flow visualisation at the middle of the saloon may act as an air curtain along the aisle; however, the aerosol concentrations measured at S_{45} were similar to those measured on the row behind the source. This suggests a similar degree of mixing in both directions in the absence of passengers. therefore, it seems that, in the case of a busy carriage where physical distancing is not possible, there is not much of advantage to either sitting across the aisle on the same row or sitting one row ahead of or behind another passenger.

The ventilation seemed effective at removing any stratification of CO_2 concentrations; therefore, it may be appropriate to consider the saloon as well mixed throughout its height. The airflow visualisations also demonstrated the importance of considering the convective plumes generated by the body heat of the passengers. These were clearly visible both when ventilation was on and off and may have a significant effect on the initial trajectory of exhaled droplets in addition to the general flow patterns within the saloon, particularly when occupancy is high. The experiments also demonstrate the sensitivity of the airflow to the location of the extract vents. Both the CO_2 and aerosol particle dispersion experiments showed a strong bias in dispersion towards the nearest extract vent. Furthermore, significantly different flow patterns were observed at the end and middle of the saloon. The dominant upward flow observed at the end of the saloon is due to the suction of the extract vents that were positioned directly above. The different flow behaviour between the middle and end of the saloon, along with the large differences in CO_2 measured along its length, suggests that the risk of airborne transmission may vary depending on the seating positions of the passengers and the location of any infected passenger.

The aerosol dispersion experiments demonstrated the importance of considering particle size or mass. Measurements suggested a slightly higher degree of dilution for the fine fraction of particles than for the coarse fraction; however, both size fractions were dispersed effectively. This, along with the large decrease in CO_2 measured with distance from passengers, suggests that physical distancing, where possible, is likely to be an effective strategy for reducing the risk of airborne transmission, particularly from larger droplets. The size and mass of viral-laden droplets can cover a wide range [35]. Therefore, it is important to understand the dispersion behaviour for the full range of exhaled droplet sizes (0.01–1000 μm). In these experiments, only the difference between two size ranges was considered. Ideally, a more advanced particle counter would be used to achieve insight into a broader range of particle sizes. Furthermore, while CO_2 is a useful indicator of exhaled air, its measurements do not provide insight into dispersion of larger droplets.

There are several limitations to the experiments presented here. First, only a limited number of runs were performed for each experimental method. Second, it was not always possible to allow sufficient time to reach a steady-state during and in between experiments. These limitations were due to the short time period available on the train. Finally, as the experiments took place during the COVID-19 pandemic, the time spent on the carriage was limited in order to mitigate the risk of transmission between those undertaking the experiments.

Despite these factors, the utility of the methods used has been demonstrated for full scale experiments. They have also shown the complexity of the airflow within an intercity train carriage and have provided some useful insights into flow behaviour within the saloon. It is clear that simple approaches such as using the Wells–Riley equation [48], which assumes a well-mixed space, are unlikely to provide accurate estimates of the probability of infection. It is also worth noting that an intercity carriage is likely to represent the simplest case in relation to airflow and droplet dispersion. In this case, journey times tend be longer; therefore, passengers are more likely to remain seated for longer periods of time, there are fewer occurrences of boarding and alighting, and the carriage doors do not open directly into the saloon. This is not the case for regional trains in which the increase in people's movement, increase in the frequency of stops, carriage accelerations, and decelerations in addition to the exchange of air when doors are opened are likely to significantly increase the complexity of the problem. In order to fully understand the implications of these insights to the risk of airborne transmission in addition to their relevance for different ventilation settings and carriage occupancy, a high fidelity model such as CFD may be required. The data gathered here will prove useful for comparison with and provide confidence in future CFD simulations. Furthermore, the experiments have provided useful insights for the development of a 1D advection–diffusion model which is currently work in progress. An alternative approach for understanding the risk of airborne transmission on the carriage is to deploy CO_2 sensors within the carriage while in service; the measurements can be used to estimate transmission risk [13].

5. Conclusions

The experiments presented in this paper were performed at short notice and during somewhat challenging conditions during the COVID-19 pandemic (during the August of 2020). Therefore, they do not represent a comprehensive analysis of the airborne transmission on the carriage; nevertheless, they are a rare example of experiments conducted at full-scale on an operational train carriage. Three experiments were performed on a stationary intercity train carriage using a single ventilation setting (75% forced cooling). The data obtained consisted of CO_2 measurements of exhaled air, measurements of aerosol particles from a nebuliser, and flow visualisations of fog illuminated using lasers. All three experiments were successful in providing useful insights into the flow and dispersion behaviour on the carriage and also demonstrated the "proof of concept" for these methods for full-scale experiments. For example, it was found that the carriage saloon is not well mixed along its length; however, it is likely to be well mixed along its height and width. This is useful information for the rail operator when considering suitable seating restrictions to enforce physical distancing. Based on the findings reported here, it is recommended that in order to mitigate the risk of airborne transmission, mask wearing should be encouraged on intercity train carriages, and any practical measures available to encourage physical distancing between passengers during periods of low occupancy should be implemented. While some of the insights may seem intuitive, there is value in their verification. These findings will inform further experiments that are planned.

The suitability of the fresh air supply rates on UK train carriages is also discussed by drawing on the CO_2 concentrations measured during these experiments.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/atmos12101267/s1, Video S1: center of carriage; Video S2: end of carriage.

Author Contributions: Conceptualization, H.W., S.F., R.K.B. and P.F.L.; methodology, H.W., S.F., R.K.B., M.D., M.D.W., E.M., A.T., S.H., S.B.D., R.L.J., P.K. and P.F.L.; validation, S.F., S.H., A.T. and P.K.; formal analysis, H.W., S.F., R.K.B., S.H. and A.T.; resources, S.B.D., R.L.J., P.K. and P.F.L.; data curation, S.F., R.K.B., M.D., E.M., S.H. and A.T.; writing—original draft preparation, H.W., R.K.B., S.H. and A.T.; writing—review and editing, S.F., M.D.W., E.M., S.B.D., R.L.J., P.K. and P.F.L.; visualization and supervision, S.B.D., R.L.J., P.K. and P.F.L.; funding acquisition, R.L.J. and P.F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) Grand Challenge grant "Managing Air for Green Inner Cities" (MAGIC) (grant number EP/N010221/1). P.F.L. and P.K. thank the support from the EPSRC-funded project "CO-TRACE (COvid-19 Transmission Risk Assessment Case studies-Education Establishments; EP/W001411/1)", and P.K. acknowledges the EPSRC/University of Surrey supported Official Development Assistance project "Knowledge transfer and practical application of research on indoor air quality (KTP-IAQ)".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This work was carried out with the support of First Group Rail for which the authors are tremendously grateful. We would particularly like to thank Jonathan Bridgewood who made these experiments possible.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

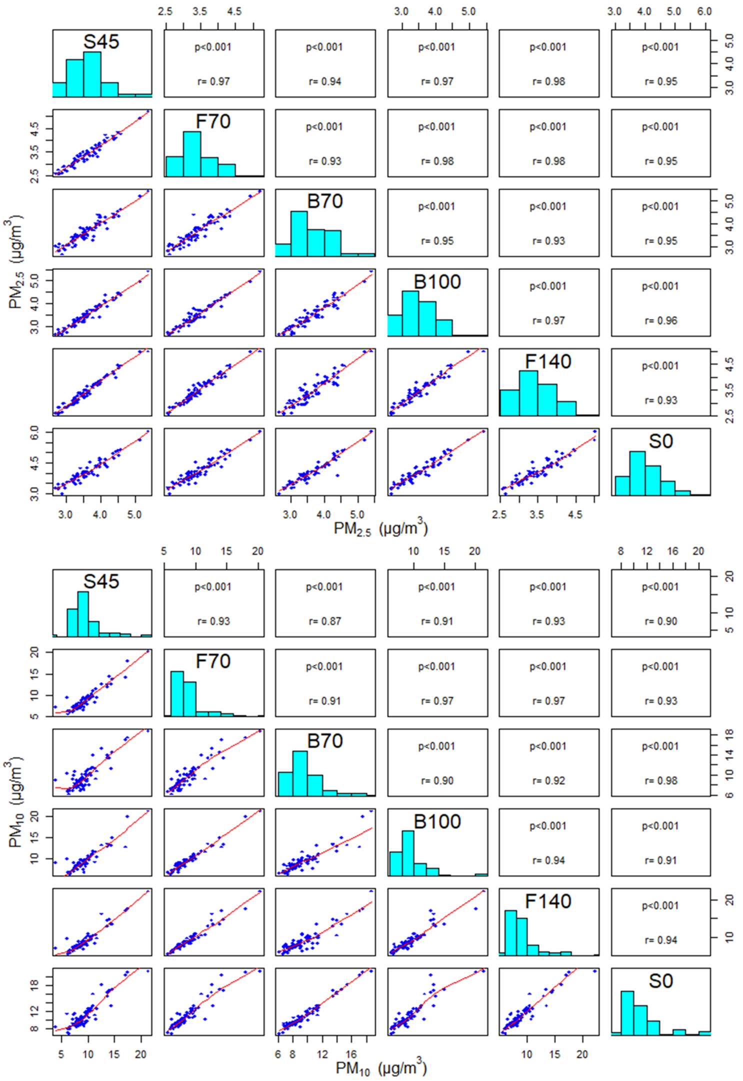

Figure A1. Correlation matrix of aerosol monitors during the co-location campaign.

Table A1. Mean and standard deviation concentrations (μ g m^{-3}) of fine and coarse aerosol concentrations and normalised concentrations at different distances from the source under ventilation off and on conditions.

	S_0	B_{100}	B_{70}	S_{45}	F_{70}	F_{140}
PM_{10} Off	1701 ± 511.9	77 ± 16	91 ± 19	194 ± 45	88 ± 18	66 ± 18
C_{norm}(%)	100	6	7	13	6	5
PM_{10} On	483 ± 134	48 ± 7	61 ± 10	74 ± 18	39 ± 6	31 ± 3
C_{norm}(%)	100	13	16	19	11	8
$PM_{2.5}$ Off	369 ± 216	55 ± 12	65 ± 14	143 ± 34	67 ± 14	49 ± 13
C_{norm}(%)	100	26	33	60	35	23
$PM_{2.5}$ On	283 ± 67	38 ± 6	48 ± 8	58 ± 14	32 ± 5	25 ± 3
C_{norm}(%)	100	16	20	23	13	11

References

1. DfT. Transport Use by Mode: Great Britain, Since 1 March 2020. 2020. Available online: https://www.gov.uk/government/statistics/transport-use-during-the-coronavirus-covid-19-pandemic (accessed on 23 September 2020).
2. The Royal Society. *Face Masks for the General Public | Royal Society DELVE Initiative*; The Royal Society: London, UK, 2020.
3. Burridge, H.C.; Bhagat, R.K.; Stettler, M.E.J.; Kumar, P.; De Mel, I.; Demis, P.; Hart, A.; Johnson-Llambias, Y.; Felipe King, M.; Klymenko, O.; et al. The ventilation of buildings and other mitigating measures for COVID-19: A focus on wintertime. *Proc. R. Soc. A* **2021**, *477*, 20200855. [CrossRef]
4. Morawska, L.; Tang, J.W.; Bahnfleth, W.; Bluyssen, P.M.; Boerstra, A.; Buonanno, G.; Cao, J.; Dancer, S.; Floto, A.; Franchimon, F.; et al. How can airborne transmission of COVID-19 indoors be minimised? *Environ. Int.* **2020**, *142*, 105832. [CrossRef]
5. Morawska, L.; Cao, J. Airborne transmission of SARS-CoV-2: The world should face the reality. *Environ. Int.* **2020**, *139*, 105730. [CrossRef] [PubMed]
6. Mittal, R.; Ni, R.; Seo, J.H. The flow physics of COVID-19. *J. Fluid Mech.* **2020**, *894*. [CrossRef]
7. Bourouiba, L.; Dehandschoewercker, E.; Bush, J. Violent expiratory events: On coughing and sneezing. *J. Fluid Mech.* **2014**, *745*, 537–563. [CrossRef]
8. Lindsley, W.G.; Noti, J.D.; Blachere, F.M.; Szalajda, J.V.; Beezhold, D.H. Efficacy of face shields against cough aerosol droplets from a cough simulator. *J. Occup. Environ. Hyg.* **2014**, *11*, 509–518. [CrossRef] [PubMed]
9. Van Doremalen, N.; Bushmaker, T.; Morris, D.H.; Holbrook, M.G.; Gamble, A.; Williamson, B.N.; Tamin, A.; Harcourt, J.L.; Thornburg, N.J.; Gerber, S.I.; et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* **2020**, *382*, 1564–1567. [CrossRef] [PubMed]
10. Kumar, P.; Morawska, L. Could fighting airborne transmission be the next line of defence against COVID-19 spread? *City Environ. Interact.* **2019**, *4*, 100033. [CrossRef]
11. Rudnick, S.N.; Milton, D.K. Risk of indoor airborne infection transmission estimated from carbon dioxide concentration. *Indoor Air* **2003**, *13*, 237–245. [CrossRef]
12. Vouriot, C.V.M.; Burridge, H.C.; Noakes, C.J.; Linden, P.F. Seasonal variation in airborne infection risk in schools due to changes in ventilation inferred from monitored carbon dioxide. *Indoor Air* **2021**. [CrossRef]
13. Burridge, H.C.; Fan, S.; Jones, R.L.; Noakes, C.J.; Linden, P.F. Predictive and retrospective modelling of airborne infection risk using monitored carbon dioxide. *Indoor Built Environ.* **2021**, accept. [CrossRef]
14. Conceição, E.Z.; Silva, M.C.; Viegas, D.X. Airflow Around a Passenger Seated in a Bus. *HVAC R Res.* **1997**, *3*, 311–323. [CrossRef]
15. Conceicao, E.; Silva, M.; Viegas, D. Evaluation of the thermal comfort level perceived by a passenger in a bus compartment. In Proceedings of the Fifth International Conference on Air Distribution in Rooms, ROOMVENT, Yokohama, Japan, 17–19 July 1996; Volume 96, pp. 403–413.
16. Li, J.; Cao, X.; Liu, J.; Mohanarangam, K.; Yang, W. PIV measurement of human thermal convection flow in a simplified vehicle cabin. *Build. Environ.* **2018**, *144*, 305–315. [CrossRef]
17. Pawar, S.; Gade, U.R.; Dixit, A.; Tadigadapa, S.B.; Jaybhay, S. Evaluation of Cabin Comfort in Air Conditioned Buses Using CFD. In Proceedings of the SAE 2014 World Congress & Exhibition, Detroit, MI, USA, 8–10 April 2014, [CrossRef]
18. Konstantinov, M.; Wagner, C. Numerical Simulation of the Thermal Comfort in a Train Cabin. *Int. J. Railw. Technol.* **2015**, *4*, 69–88. [CrossRef]
19. Yang, L.; Li, X.; Tu, J. Thermal comfort analysis of a high-speed train cabin considering the solar radiation effects. *Indoor Built Environ.* **2020**, *29*, 1101–1117. [CrossRef]
20. Mingotti, N.; Wood, R.; Noakes, C.; Woods, A.W. The mixing of airborne contaminants by the repeated passage of people along a corridor. *J. Fluid Mech.* **2020**, *903*, A52. [CrossRef]

21. Wang, H.; Lin, M.; Chen, Y. Performance evaluation of air distribution systems in three different China railway high-speed train cabins using numerical simulations. *Build. Simul.* **2019**, *7*, 629–638. [CrossRef]
22. Zhang, L.; Li, Y. Dispersion of coughed droplets in a fully-occupied high-speed rail cabin. *Build. Environ.* **2012**, *47*, 58–66. [CrossRef]
23. Yang, X.; Ou, C.; Yang, H.; Liu, L.; Song, T.; Kang, M.; Lin, H.; Hang, J. Transmission of pathogen-laden expiratory droplets in a coach bus. *J. Hazard. Mater.* **2020**, *397*, 122609. [CrossRef]
24. Zhu, S.; Srebric, J.; Spengler, J.D.; Demokritou, P. An advanced numerical model for the assessment of airborne transmission of influenza in bus microenvironments. *Build. Environ.* **2012**, *47*, 67–75. [CrossRef]
25. Olsen, S.J.; Chang, H.L.; Cheung, T.Y.Y.; Tang, A.F.Y.; Fisk, T.L.; Ooi, S.P.L.; Kuo, H.W.; Jiang, D.D.S.; Chen, K.T.; Lando, J.; et al. Transmission of the Severe Acute Respiratory Syndrome on Aircraft. *N. Engl. J. Med.* **2003**, *349*, 2416–2422. [CrossRef] [PubMed]
26. Yan, W.; Zhang, Y.; Sun, Y.; Li, D. Experimental and CFD study of unsteady airborne pollutant transport within an aircraft cabin mock-up. *Build. Environ.* **2009**, *44*, 34–43. [CrossRef]
27. Mazumdar, S.; Chen, Q. A one-dimensional analytical model for airborne contaminant transport in airliner cabins. *Indoor Air* **2009**, *19*, 3–13. [CrossRef]
28. Elmaghraby, H.A.; Chiang, Y.W.; Aliabadi, A.A. Ventilation strategies and air quality management in passenger aircraft cabins: A review of experimental approaches and numerical simulations. *Sci. Technol. Built Environ.* **2018**, *24*, 160–175. [CrossRef]
29. Ryan, P.; Spengler, J.D.; Halfpenny, P.F. Sequential box models for indoor air quality: Application to airliner cabin air quality. *Atmos. Environ.* **1988**, *22*, 1031–1038. [CrossRef]
30. Ko, G.; Thompson, K.M.; Nardell, E.A. Estimation of Tuberculosis Risk on a Commercial Airliner. *Risk Anal.* **2004**, *24*, 379–388. [CrossRef] [PubMed]
31. Silcott, D.; Kinahan, S.; Santarpia, J.; Silcott, R.; Silcott, P.; Silcott, B.; Distelhorst, S.; Herrera, V.; Rivera, D.; Crown, K.; et al. *TRANSCOM/AMC Commercial Aircraft Cabin Aerosol Dispersion Tests*; United States Transportation Command (USTRANSCOM): Belleville, IL, USA; Air Mobility Command (AMC): Scott Air Force Base, IL, USA, 2020.
32. Elmaghraby, H.A.; Chiang, Y.W.; Aliabadi, A.A. Are aircraft acceleration-induced body forces effective on contaminant dispersion in passenger aircraft cabins? *Sci. Technol. Built Environ.* **2019**, *25*, 858–872. [CrossRef]
33. Elmaghraby, H.A.; Chiang, Y.W.; Aliabadi, A.A. Airflow design and source control strategies for reducing airborne contaminant exposure in passenger aircraft cabins during the climb leg. *Sci. Technol. Built Environ.* **2020**, *26*, 901–923. [CrossRef]
34. Elmaghraby, H.A.; Chiang, Y.W.; Aliabadi, A.A. Normal and extreme aircraft accelerations and the effects on exposure to expiratory airborne contaminant inside commercial aircraft cabins. *Sci. Technol. Built Environ.* **2020**, *26*, 924–927. [CrossRef]
35. Bake, B.; Larsson, P.; Ljungkvist, G.; Ljungström, E.; Olin, A. Exhaled particles and small airways. *Respir. Res.* **2019**, *8*. [CrossRef]
36. Kannan, R. Breathing Rates and Breathing Air Constituents. *Int. J. Innov. Sci. Eng. Technol.* **2015**, *2*, 557–558.
37. Kumar, P.; Rivas, I.; Sachdeva, L. Exposure of in-pram babies to airborne particles during morning drop-in and afternoon pick-up of school children. *Environ. Pollut.* **2017**, *224*, 407–420. [CrossRef]
38. Kumar, P.; Hama, S.; Nogueira, T.; Abbass, R.A.; Brand, V.S.; de Fatima Andrade, M.; Asfaw, A.; Aziz, K.H.; Cao, S.J.; El-Gendy, A.; et al. In-car particulate matter exposure across ten global cities. *Sci. Total Environ.* **2021**, *750*, 141395. [CrossRef]
39. Carslaw, D.C.; Ropkins, K. Openair—An R package for air quality data analysis. *Environ. Model. Softw.* **2012**, *27–28*, 52–61. [CrossRef]
40. CIBSE. Chartered Institution of Building Services Engineers (CIBSE) Guide A: Environmental Design. 2015. Available online: https://cibse.org/Knowledge-by-publication-type/Guides (accessed on 31 August 2021).
41. Wargocki, P. The Effects of Ventilation in Homes on Health. *Int. J. Vent.* **2013**, *12*, 101–118. [CrossRef]
42. Batterman, S. Review and Extension of CO_2 Based Methods to Determine Ventilation Rates with Application to School Classrooms. *Int. J. Environ. Res. Public Health* **2004**, *14*, 145. [CrossRef]
43. National Engineering Policy Centre. A Rapid Review of the Engineering Approaches to Mitigate the Risk of COVID-19 Transmission on Public Transport. 2020. Available online: httpswww.raeng.org.ukpublicationsreportsa-rapid-review-of-the-engineering-approaches-to-mi (accessed on 31 August 2021).
44. REHVA. How to Operate HVAC and Other Building Service Systems to Prevent the Spread of the Coronavirus (SARS-CoV-2) Disease (COVID-19) in Workplaces. REHVA COVID-19 Guidance Document. 2020. Available online: httpswww.rehva.eufileadminuser_uploadREHVA_COVID-19_guidance_document_V4_23112020_V2.pdf (accessed on 31 August 2021) .
45. World Health Organisation. Available online: https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted (accessed on 20 September 2021).
46. Aliabadi, A.A.; Rogak, S.N.; Bartlett, K.H.; Green, S.I. Preventing Airborne Disease Transmission: Review of Methods for Ventilation Design in Health Care Facilities. *Adv. Prev. Med.* **2011**, *2011*, 124064. [CrossRef] [PubMed]
47. Buonanno, G.; Stabile, L.; Morawska, L. Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment. *Environ. Int.* **2020**, *141*, 105794. [CrossRef] [PubMed]
48. Riley, E.C.; Murphy, G.; Riley, R.L. Airborne spread of measles in a suburban elementary school. *Am. J. Epidemiol.* **1978**, *107*, 421–432. [CrossRef]

A.1.2 Publication: Automated classification of time-activity-location patterns for improved estimation of personal exposure to air pollution

Chatzidiakou *et al. Environmental Health* (2022) 21:125
https://doi.org/10.1186/s12940-022-00939-8

Environmental Health

RESEARCH Open Access

Automated classification of time-activity-location patterns for improved estimation of personal exposure to air pollution

Lia Chatzidiakou[1*], Anika Krause[1,2], Mike Kellaway[3], Yiqun Han[4], Yilin Li[1], Elizabeth Martin[1], Frank J. Kelly[4], Tong Zhu[5], Benjamin Barratt[4] and Roderic L. Jones[1]

Abstract

Background: Air pollution epidemiology has primarily relied on measurements from fixed outdoor air quality monitoring stations to derive population-scale exposure. Characterisation of individual time-activity-location patterns is critical for accurate estimations of personal exposure and dose because pollutant concentrations and inhalation rates vary significantly by location and activity.

Methods: We developed and evaluated an automated model to classify major exposure-related microenvironments (*home, work, other static, in-transit*) and separated them into indoor and outdoor locations, *sleeping activity* and five modes of transport (*walking, cycling, car, bus, metro/train*) with multidisciplinary methods from the fields of movement ecology and artificial intelligence. As input parameters, we used GPS coordinates, accelerometry, and noise, collected at 1 min intervals with a validated Personal Air quality Monitor (PAM) carried by 35 volunteers for one week each. The model classifications were then evaluated against manual time-activity logs kept by participants.

Results: Overall, the model performed reliably in classifying home, work, and other indoor microenvironments (F1-score>0.70) but only moderately well for sleeping and visits to outdoor microenvironments (F1-score=0.57 and 0.3 respectively). Random forest approaches performed very well in classifying modes of transport (F1-score>0.91). We found that the performance of the automated methods significantly surpassed those of manual logs.

Conclusions: Automated models for time-activity classification can markedly improve exposure metrics. Such models can be developed in many programming languages, and if well formulated can have general applicability in large-scale health studies, providing a comprehensive picture of environmental health risks during daily life with readily gathered parameters from smartphone technologies.

Keywords: Portable sensor technologies, Multi-pollutant personal exposure, Automated time-activity classification

Background

Ambient air pollution is a leading environmental risk factor for chronic disease and millions of premature deaths every year worldwide [1]. Much of this evidence comes from epidemiological studies conducted in western countries where networks of outdoor reference monitoring stations have been used to provide indications of the effects of ambient air pollution on population health [2]. Recent studies focused on a global analysis of estimated source contributions to outdoor air pollution and related health effects using updated emissions inventories, satellite and air quality modelling, and relationships between

*Correspondence: ec571@cam.ac.uk

[1] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Rd, CB2 1EW Cambridge, UK
Full list of author information is available at the end of the article

air quality and health at global, regional, country, and metropolitan-area scales [3].

However, as individuals move between different, highly heterogeneous microenvironments that are mainly situated indoors, outdoor static measurements become potentially poor metrics of actual personal exposure [4], leading in many cases to bias and error in health estimations [5]. Adding to the complexity of measuring personal pollutant concentrations, physical activity levels, in turn, affect the dose of inhaled air pollution. For example, while a comprehensive review of the literature found the highest exposure to particulate matter when travelling by car compared with cycling [6], the highest whole trip doses were in fact experienced by cyclists [7] because their higher physical activity levels resulted in greater amounts of pollutant received by the body through larger volumes of inhaled air [8].

Accounting for individual mobility and activity patterns is therefore critical for improved exposure and dose estimations. Such information has been commonly collected with different self-reported questionnaires [9] which often introduce participant error and missing data [10, 11] and increase the participation burden (i.e. time and effort required to complete) [12]. A growing number of studies have taken advantage of increasingly widespread sensor technologies, such as geographical positioning system (GPS) sensors in smartphones, to improve the accuracy of indirect air pollution exposure assessment in large-scale health studies by tracking people's time-location patterns [13–16].

Time-activity patterns and modes of transport cannot be derived from the GPS raw data directly without further data processing. Only a few studies aim to classify time-activity patterns during daily life using GPS tracking data (smartphone-based or handheld devices), in some cases combined with temperature, light or motion sensors [17–24] to develop primarily rule-based models and/or random forest (RF) learning techniques for a small number of participants over a few days.

In a previous paper [25], we developed, deployed and comprehensively evaluated the performance of a highly portable air pollution sensor platform (PAM) for personal exposure assessments in health studies. We now aim to present a methodological framework as the basis of an approach that automatically classifies and integrates time-activity patterns in personal exposure assessments. This work is toward an overarching aim of capturing total personal multi-pollutant dose in unprecedented detail and, together with medical outcomes, identifying underlying mechanisms of the detrimental effects of specific air pollutants on health. While we use auxiliary parameters collected with a custom-made sensor platform as inputs, such parameters can be readily collected with smartphone technologies, making this method transferable to large-scale health studies.

Conceptual structure of the time activity model

We developed a model to classify major exposure-relevant microenvironments (*home, work, other static, in transit*) and subclassified them into *indoor* and *outdoor* locations, *sleeping* activities and five modes of transport (*walking, cycling, car, bus, train/metro*) using two open-source software components, `R` [26, 27] and `PostgreSQL` [28, 29]. The input parameters for this model (GPS coordinates, noise and accelerometry) were collected with the PAM [25] (S1). Information on data management, post-processing and sensor performance can be found in Chatzidiakou et al., 2019 [25] and in S1.

The PAM has been previously deployed in a number of health studies to monitor the thermal parameters (temperature and RH) and personal exposure of participants to multiple pollutants at high spatial and temporal resolution [30, 31] including carbon monoxide (CO), nitric oxide (NO), nitrogen dioxide (NO_2), ozone (O_3) and size segregated particulate matter (PM). However, pollutant measurements[1] and thermal parameters were not used as predictors in this model in order to make this methodology generally applicable to other studies and also transferable to different geographical settings and varying seasons.

The model can be conceptualised as a series of six consecutive steps, as shown in Fig. 1, to classify major microenvironments, activities and modes of transport (shown in red font), combining rule-based algorithms (blue) and artificial intelligence (AI) methods (purple) summarised in Table 1.

Step 1 aims to identify the home location with a simple rule-based algorithm to effectively reduce the volume of the data that will be processed with a Lagrangian home-range estimation method [32, 33] in *Steps 2* and *3*. In that way we effectively reduce the volume of data because such methods generally require higher computation power to implement more complex *geometric* or *probabilistic* models[2]. We adopt an existing technique [34] developed in the field of ecology and extend its use to human mobility studies. It combines the robustness of geometric estimators with the simplicity of

[1] with the exception of the larger fraction of PM for sleeping activity

[2] Geometric estimators aim to delineate the spatial extent of an individual's movement by constructing polygons (called *hulls*) of all visited places. Probabilistic estimators create the probability density (called *utilisation distribution*) that an individual is found at a given point in space and represent the density of use of space. Widely used geometric methods are convex hull methods while the most common probabilistic methods are kernel density methods to analyse animal territory and movement [32].

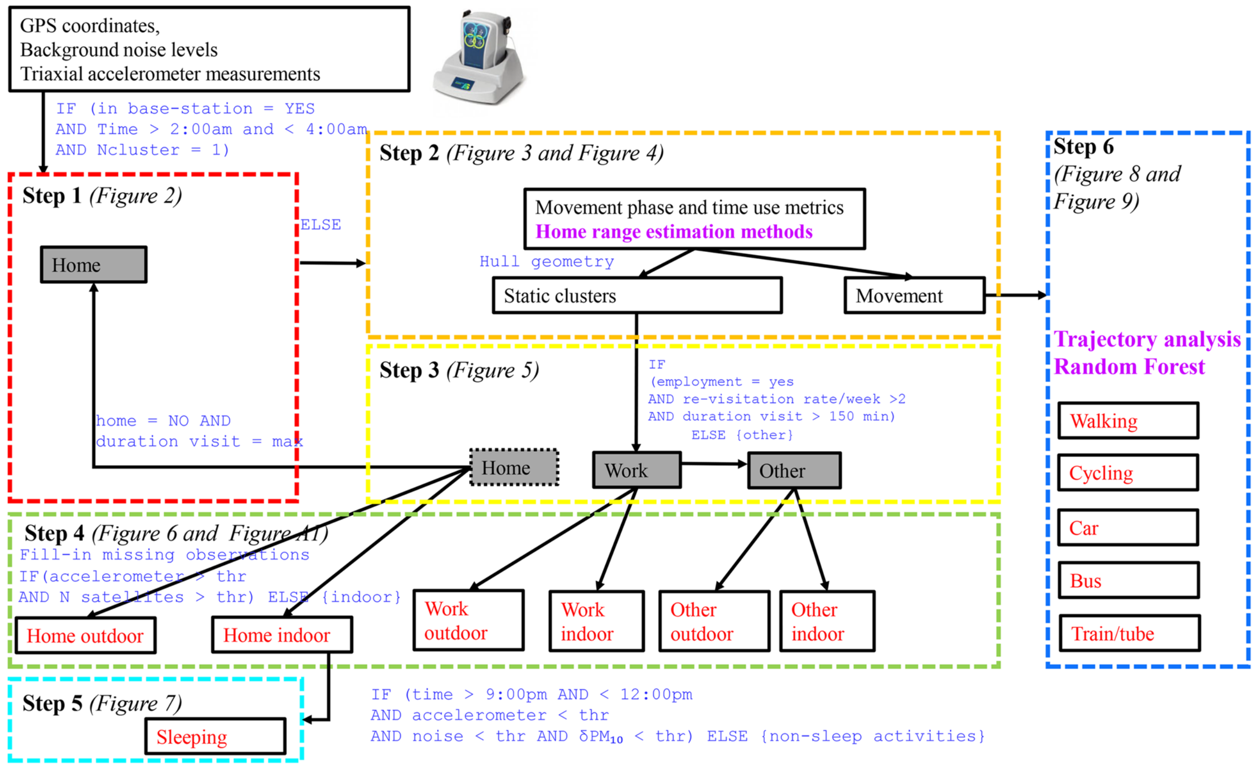

Fig. 1 Flow chart of the time activity model

Table 1 Summary of AI methods integrated into the time-activity model

AI method	R implementation	Outputs
Home-range method that combines geometric and probabilistic estimators	Time Local Convex Hull (T-LoCoH) [34]	**Polygon (hull) geometry** gives information on directional movement vs. static clusters (*Step 2*). **Visitation rate and duration of visit** enable classifications based on behavioural patterns of the individual (*Step 3*).
Trajectory analysis	Adehabitat LT [37]	Segmentation of movement with the Lavielle method [57]
Predictor selection for Random Forest (RF) classification with three-step elimination process based on data-driven thresholds for high dimensional datasets	VSURF [38]	Predictor variables for RF models collected with the PAM (movement, noise, GPS information) and baseline questionnaire (common modes of transport), and extracted from spatial analysis
RF classification of the mode of transport with the 10-fold evaluation method	RandomForest [65]	Probabilistic classification for each mode of transport

probabilistic methods to identify important place-marks and fully characterise exposure-relevant behavioural patterns of how the individual uses their activity space.

Step 4 and *Step 5* employ rule-based algorithms to interpolate missing observations, separate indoor from outdoor static microenvironments and classify sleeping activity. Finally, in *Step 6* we classify modes of transport observations with RF [35], the use of which is considered best practice in travel mode classification [36]. To assist the classification, we perform trajectory analysis [37] to extract useful metrics of movement. Important predictor variables for RF model development were selected with an automated method [38] suitable for high-dimensional data (see Table 1).

Additional to the above main R software environment packages that form the backbone of the model, we used for spatial analysis and visualisation: sp [39, 40], rgdal [41], raster [42], gpclib [43], OpenStreetMap [44], ggplot2 [45] and ggmap [46], rayshader [47]; for time-series analysis, data manipulation and visualisation:

openair [48], dplyr [49], plot3D [50]; and for clustering and classification: caret [51], dbscan [52].

The model development steps are described in detail below and illustrated using information from one representative participant over a period of one week.

Step 1: Rule-based algorithm for home location identification to reduce computational demand of the time-activity model

The rationale of this simple algorithm relies on common behavioural patterns of most people in western settings, who tend to spend most of their nighttime at home (Fig. 2b). This assumption holds particularly in this study but it can be readily adjusted to shift workers who may be at home at different times. We identified periods when the PAM was in the base-station - the dock used by participants to charge the PAM at home - (as indicated by the input voltage of the unit) and when the local time was between 02:00-04:00 AM; therefore, making it more likely that the participant was at home. Due to GPS errors, these points tended to be displaced around the home location as illustrated in Fig. 2c, often falling outside the GIS building boundaries.

A clustering algorithm (in this case k-means in R) was applied to this data subset to determine whether the scattered points formed a single cluster for each participant. For a few participants, multiple clusters were detected hence *home* could not be determined in this step (for example, due to sleeping in multiple locations or lack of satellite reception during the selected period) and for these participants *home* was subsequently classified in *Step 2* as the location where the participant spent most of their time.

If a single cluster was identified, a spatial elliptical zone (*"buffer zone"*) was created around each home microenvironment by extracting the centroid coordinates and the individual spread distances (δLon and δLat) (Fig. 2c). Any spread is expected to depend on contextual factors (such as building construction characteristics and GPS signal quality) and was typically found to range from 60m to 500m([23, 24]. Data points within that spatial zone (Fig. 2c) were classified as *home* and were separated into *indoor* and *outdoor* in *Step 4.*

Step 2: Stationary locations and movement patterns from space-use metrics

The remaining observations (i.e. those not belonging within the home spatial zone) were analysed with the R package T-LoCoH [34] (Table 1) to distinguish between movement and static activities. The strength of this technique is that it models space-use (*Step 2*) and time-use (*Step 3*) simultaneously. It does that by employing a scaling that relates distance and time in reference to an individual's characteristic velocity (time-scaled distance). Previous studies have found that such estimators that incorporate a temporal component with individual-specific parameters generally perform better than traditional estimators [53]. We first used the extracted geometric features to classify static clusters and directional

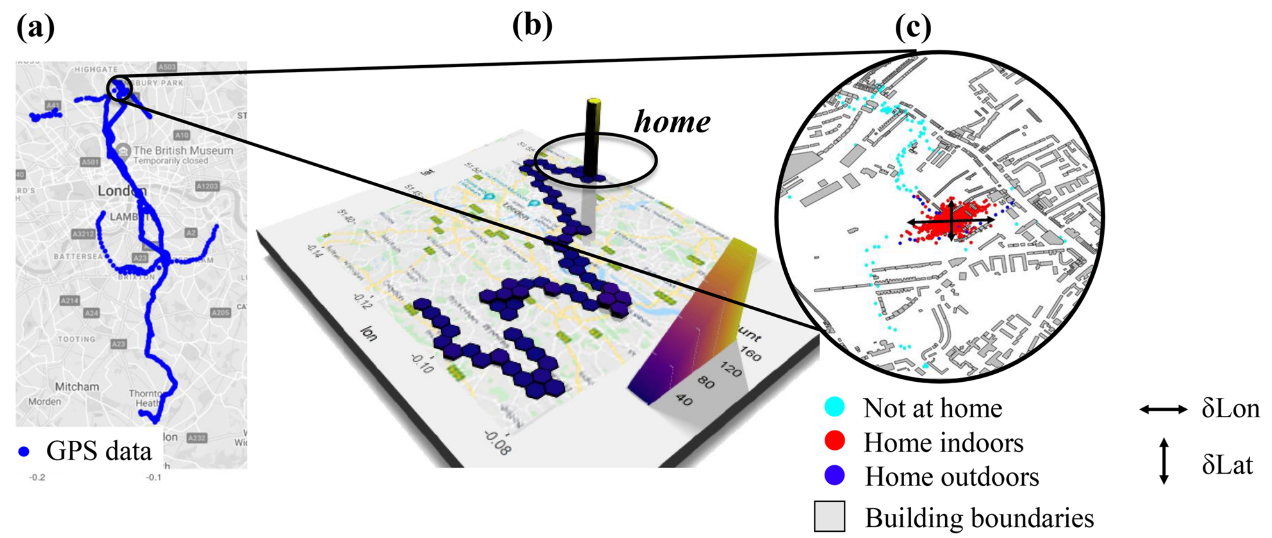

Fig. 2 Graphical flow chart of home identification of the time-activity model. (a) Map of the raw GPS data (blue) collected from a representative participant carrying a personal air quality monitor over a week. (b) 3D density plot of participant's time budget projected on a map. "Home" location has the highest point density (i.e. most time spent). (c) A spatial elliptical zone created with a rule-based model to identify "home" that included indoor (red) and outdoor (blue) micro-environments (separated in Step 4). The spread distances (δLon and δLat) around the centroid are often larger than the GIS footprints of the buildings (grey) and depend on multiple factors. Map data from Google Maps 2021 (a and b) and OpenStreetMap(c)

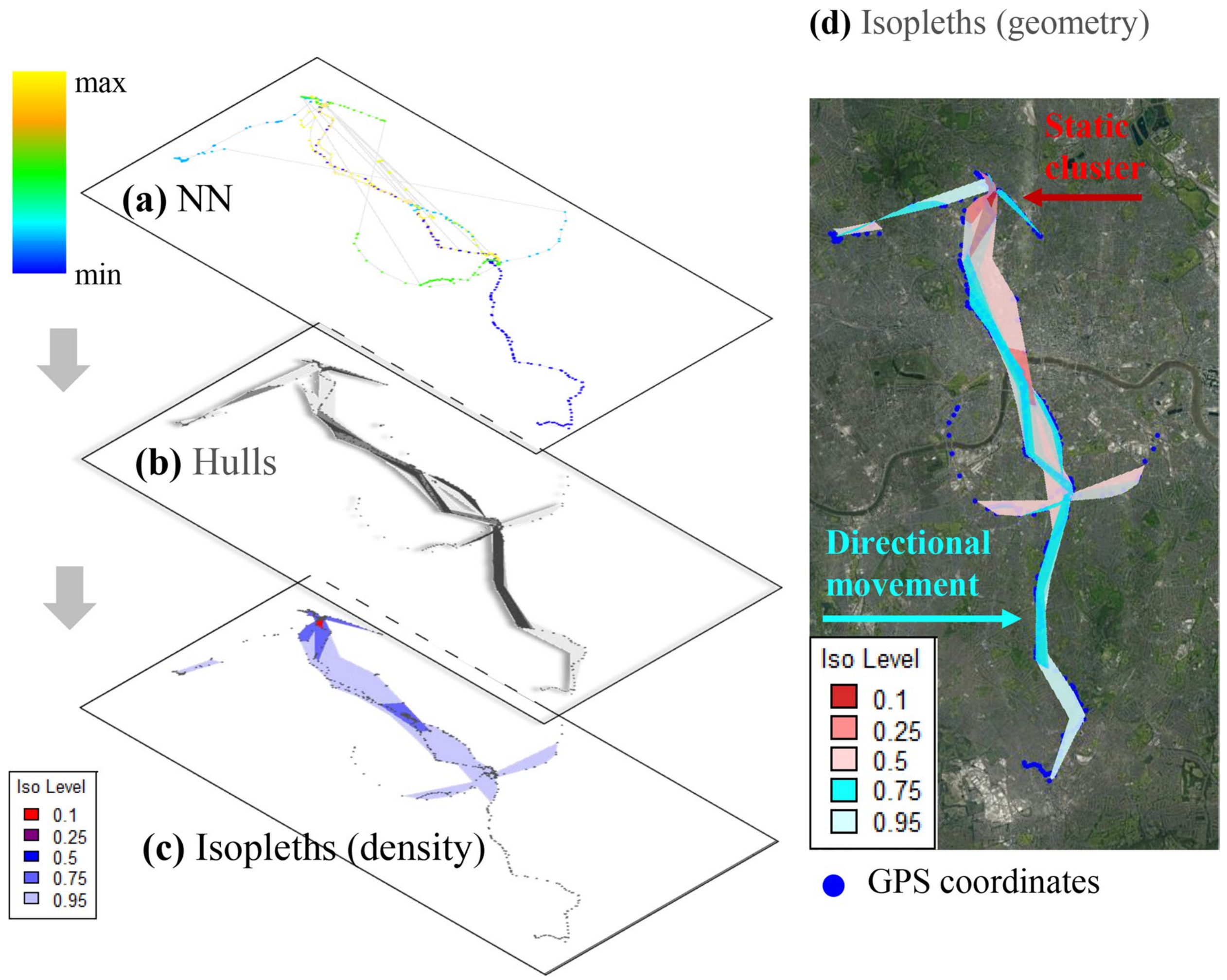

Fig. 3 Example graphical flow of space-time utilisation distribution analysis (step 2) implemented with the T-LocoH package in R. (a) First, nearest neighbours were identified with the adaptive method (α-NN) (b) Minimum convex polygons (hulls) were then produced from these α-NN (c) Hulls were merged by point density to create density isopleths (utilisation distributions) to characterise space intensity use. (d) Hulls were merged by the eccentricity of the bounding ellipse to create elongation isopleths to characterise movement and were projected on a map (Google Maps 2021)

movement following the workflow illustrated in Fig. 3 and described below:

- **Figure 3a: Defining nearest neighbours with the adaptive method**. GPS data were first converted to a conformal (Universal Transverse Mercator) projection because it preserves local angles and represents shapes accurately and without distortion for small areas. The algorithm begins by identifying a set of nearest neighbours around each point (Fig. 3a) based on their time-scaled distance. Participants did not utilise areas in a uniform pattern, but rather selected areas based on their individual activities, resulting in heterogeneous coverage of both dense and sparse areas. To account for these patterns, the selection of nearest neighbours [34] was performed with the adaptive method (α-NN).[3]
- **Figure 3b: Geometry of the enclosing polygons**. Each parent point and its nearest neighbours were bound together with a minimum convex polygon or a hull (Fig. 3b). Hulls are the building blocks of the subsequent analysis and have different properties (point density and shape) which in turn provide important

[3] The adaptive method specifies that the sum of the distances of all nearby points around each parent point is less than or equal to α. Essentially, this method adjusts the size of the circles that enclose nearest neighbours based on the frequency of use of each area. In regions with more data, smaller circles can be constructed resulting in a higher resolution of space-use metrics. Because α is defined empirically, we used an automated method to find a suitable value for each participant [34].

information on the use of space. The eccentricity of the ellipse bounding a hull is a good approximation of its shape, which specifies whether an individual is in movement or stationary. For example, a bounding ellipse with an eccentricity value close to zero resembles a circle and indicates areas where the individual was stationary for an extended period, resulting in a dense cluster of points similar to the red cluster presented earlier in Fig. 2c. In contrast, elongated bounding ellipses have an eccentricity value close to one because they enclose nearest neighbours that form linear segments indicating areas of directional movement.

- **Figure 3c and d: Defining areas with similar polygon geometry**. Depending on the research question, hulls can be sorted by a selected property, and then merged together to form isopleths that connect areas with the same numerical value of that property. In the example of Fig. 3c, areas that are used by the participant with the same intensity were merged to produce traditional utilisation distributions. When hulls with similar eccentricity values are merged as shown in Fig. 3d, similar movement patterns are connected in a single isopleth ranging from the highest elongation hull value close to 1 (cyan) capturing points in movement to the lowest elongation value close to 0 (red) indicating dense clusters of GPS points. In this way, similar movement patterns are grouped into a single isopleth. Isopleths typically contain 95% of the total points excluding outliers that occur frequently and could skew the results [34].

Figure 4 illustrates these extracted geometric features in 3D (top) and 2D (bottom) maps. The graphs show that both the eccentricity of the enclosing ellipses (Fig. 4a) and the number of nearest neighbours (Fig. 4b) provide strong discriminatory power to separate directional movement from static locations (Fig. 4c) with suitable thresholds.

Step 3: Behavioural patterns from time-use metrics

In the previous step, we constructed hulls using the time-scaled distance between GPS points. The time-scaled distance distinguishes points that are far away in time even though they may be close in Euclidean space. Therefore,

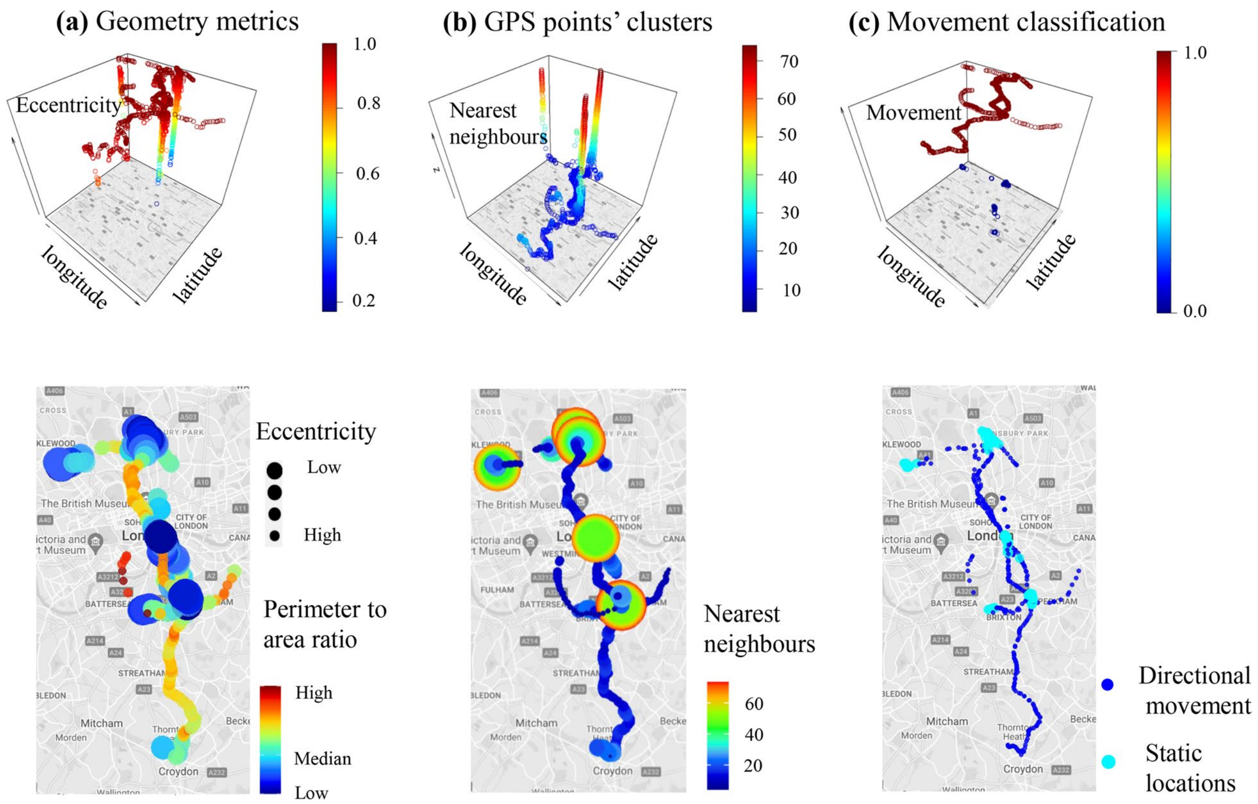

Fig. 4 Selected features for the classification of static clusters and directional movement are shown in 3D (top) and projected on maps (bottom) in a colour and size-scale. (a) The eccentricity and the perimeter-to-area ratio of the enclosing ellipse provide information on hull geometry and directional movement. (b) Dense clusters of nearest neighbours were constructed in areas used more frequently. (c) Final classification of static clusters and in-movement location based on thresholds of these features

the hulls are local not only in space but also in time enabling the characterisation of behavioural patterns with two important temporal features: the duration of visit and the revisitation rate over 12 hours to capture diurnal patterns of human behaviour.

The scatterplot of Fig. 5b shows that, based on the revisitation rate and duration of visit, seven distinct clusters were identified and projected on a map in Fig. 5a. Overall, three main categories can be identified: clusters which were visited often and for extended time periods (*Clusters 1* and *2*), clusters where the participant spent limited time (*Clusters 3* and *4*), and finally clusters visited once during the week but for longer time (i.e. more than an hour as in *Clusters 4, 5, 6* and *7*).

These extracted time-use metrics assisted the automated classification. *Cluster 1* (Fig. 5b) could be classified as *home* (if it had not been classified as such in *Step 1*) as shown in Fig. 5d. The cluster visited frequently and for extended time periods and was classified as *work* (in this example *Cluster 2*).

Cluster 4 was classified as in-movement, not only based on the hull metrics in *Step 2*, but also based on the low duration of visit as shown in Fig. 5b. Within *Cluster 4*, differences in revisitation rates (as illustrated by the size of points in Fig. 5c) can be used to distinguish daily commuting routes. For example, points between *home* and *work* have been revisited 3 times compared with points south of *work* that have only been visited once.

Finally, details on locations visited for extended periods but less often, (*Clusters 3,5,6* and *7*) could be retrieved from GIS maps and common behavioural patterns. For example, *Cluster 3* in proximity to home had short but frequent visits within the spatial zone of the overground station and could be classified as *waiting for the train* (Fig. 5e). Contrary, *Cluster 7* was only visited once but had a high duration of visit and together with the GIS information could have been classified as a secondary workplace location (Fig. 5f, KCL Waterloo Campus) .Both subclassifications were confirmed by the manual diary entries. Although this approach shows the capabilities of the model, it is beyond the scope of this work to subclassify each microenvironment and they were, therefore, all grouped as *other* but with a unique identifier (Fig. 5d). Currently, services such as Google Places API have the ability to return information on places of interest.

Overall, the technique illustrated here provides a simultaneous analysis of spatial and temporal patterns to separate static locations from directional movement and infer behavioural patterns on the use of space of the individual.

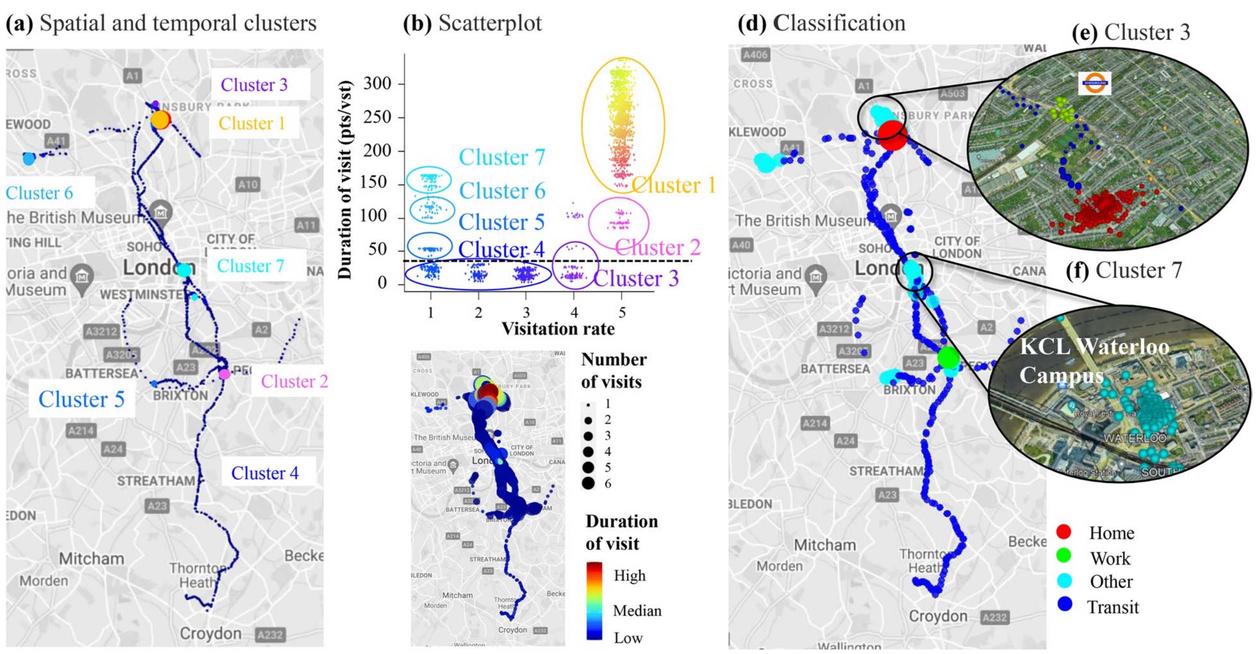

Fig. 5 Flow chart of the time activity model (a) Map of seven distinct clusters identified based on temporal information contained in the isopleths. (b) Scatterplot of the visitation rate (over 12h) vs the duration of visit (average points per visit). The dashed black line indicates the threshold in the duration of visit that discriminates between static locations from directional movement. (c) Map of time-use metrics during the participation week. The colour scale indicates the total minutes spent in each location while the size of the points corresponds to the number of visits. (d) Final classification of static locations into three microenvironments ("home", "work", "other") and in movement based on spatiotemporal behavioural patterns of the individual. (e and f) Subclassifications of "other" visited microenvironments derived from GIS information and behavioural patterns

Step 4: Separating indoor from outdoor microenvironments

GPS signal loss is common in indoor microenvironments, such as in the underground metro system, in urban areas with tall buildings and structures, or when the monitor is static in an indoor microenvironment for extended periods. In such cases, a large percentage of geo-coordinated observations may be missing. While this percentage will vary between deployments, in our sample it was found to be ~ 40%. A rule-based algorithm was developed to interpolate the missing locations using previous- and last-known locations and PAM auxiliary parameters as inputs (S2, Fig. A1), and in this way classify indoor microenvironments with limited GPS satellite reception.

Once missing observations were largely accounted for, each static microenvironment (*home, work, other*) was classified as *indoor* or *outdoor* with a rule-based algorithm (Fig. 1) formulated on the hypothesis that abrupt changes in acceleration and GPS signal quality are indicative of transitions between microenvironments. The algorithm used participant-specific thresholds of these two parameters to classify indoor and outdoor microenvironments and is visualised in Fig. 6 using data from a single participant-day.

Figure 6 presents the time-series of selected parameters (acceleration, number of satellites) to develop the indoor-outdoor separation algorithm (Fig. 6b and c), the corresponding map (Fig. 6f) with indoor (red) and outdoor (blue) classifications, as well as a 3D map of the number of satellites transmitting to the PAM receiver (Fig. 6g). Higher numbers of satellites are typically seen outdoors due to signal blockage in indoor environments (Fig. 6c and g).

We have included the manual diary logs, ozone levels measured with the PAM (Fig. 6e and h) and the time-derivative of RH as indirect ways to confirm the performance of the algorithm. During daytime, ozone levels are consistently very low indoors as shown in the 3D map in Fig. 6h (for example, locations A, B and C) due to the high reactivity and depletion on indoor surfaces, the limited solar radiation and the lack of indoor sources [54]. They are also significantly reduced during certain modes of transport (for example, B to C) for similar reasons. Finally, we have previously shown in a controlled experiment that fast changes in RH can flag rapid environmental changes as a person moves between different microenvironments [25]. Therefore, the time-derivative of RH could be used to flag

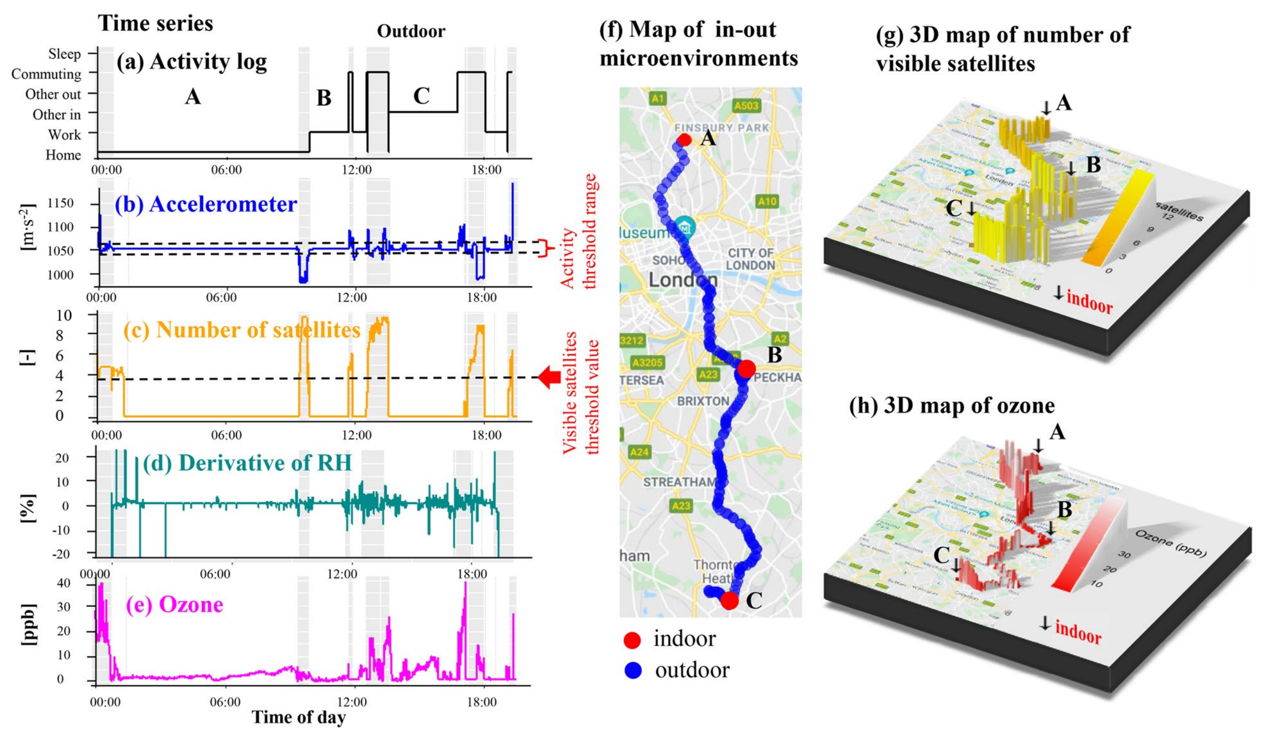

Fig. 6 Identifying transitions between indoor and outdoor microenvironments. (a) Time series of manual activity logs. Grey shaded areas indicate periods flagged as outdoor microenvironments with the rule-based algorithm. (b and c) Participant-specific thresholds (black dashed lines) of two parameters collected with the PAM (acceleration and number of visible satellites) were used to flag transitions between microenvironments. (d and e) In addition to manual logs, sudden changes in RH and ozone levels were used to evaluate the performance of the algorithm indirectly (f) Corresponding map of indoor (red) and outdoor (blue) microenvironments classified with the rule-based algorithm (g) 3D map visualising the number of satellites transmitting to the PAM GPS receiver. (h) 3D map of PAM ozone levels

the indoor-outdoor transition with high time precision (Fig. 6d).

The evaluation of the model with a single participant-day so far shows a high level of agreement between the algorithm predictions (grey shaded areas) and the manual activity logs (black line) shown in Fig. 6a. Additionally, the sharp spikes in the derivative of RH (Fig. 6d), and the rapid changes in ozone concentrations (Fig. 6e) further support that the rule-based model can discriminate between indoor and outdoor microenvironments well. Full evaluation is presented in Section 3.

Step 5: Characterisation of sleeping activity

The indoor home microenvironment was subdivided into *sleep* and non-sleep periods with a rule-based model (Fig. 1) based on the hypothesis that participants sleep when background noise levels and movement are the lowest. Additionally to the accelerometer showing that the PAM was stationary (Fig. 7), relative changes in the larger fractions of particulate matter were used as an indicator of movement in the room because larger particles would be expected to resuspend during periods of physical activity of the occupants [55]. The time derivative of PM_{10} was used to detect these changes of concentrations (Fig. 7). While in this case we use a specialised optical particle counter, such information on participant movement could have been collected with widely used wearable sensors (such as smartwatches). Participant-specific statistical thresholds were set for these three parameters to detect sleep activities followed by a smoothing filter over a 10 min rolling window applied on the binary classification to remove small disruptions.

Figure 7 shows that in this example there is an excellent agreement between manual activity logs (grey shaded area projected from time series) and algorithm-based classification (line segments highlighted in red) with a marginal overprediction of sleep because the algorithm cannot separate downtime before sleep from actual sleeping activity as recorded in the diary. This rule-based model for sleep is evaluated using the whole dataset in Section 3.

Step 6: Classification of transit modes

The periods classified as *in transit* were classified into, in this case, five modes of transportation. First, we created and selected predictor variables for the RF models which were trained and evaluated with a k-fold method as described below:

Trajectory analysis and segmentation

In-transit observations for each participant were grouped into individual commuting events (journeys). Stops were

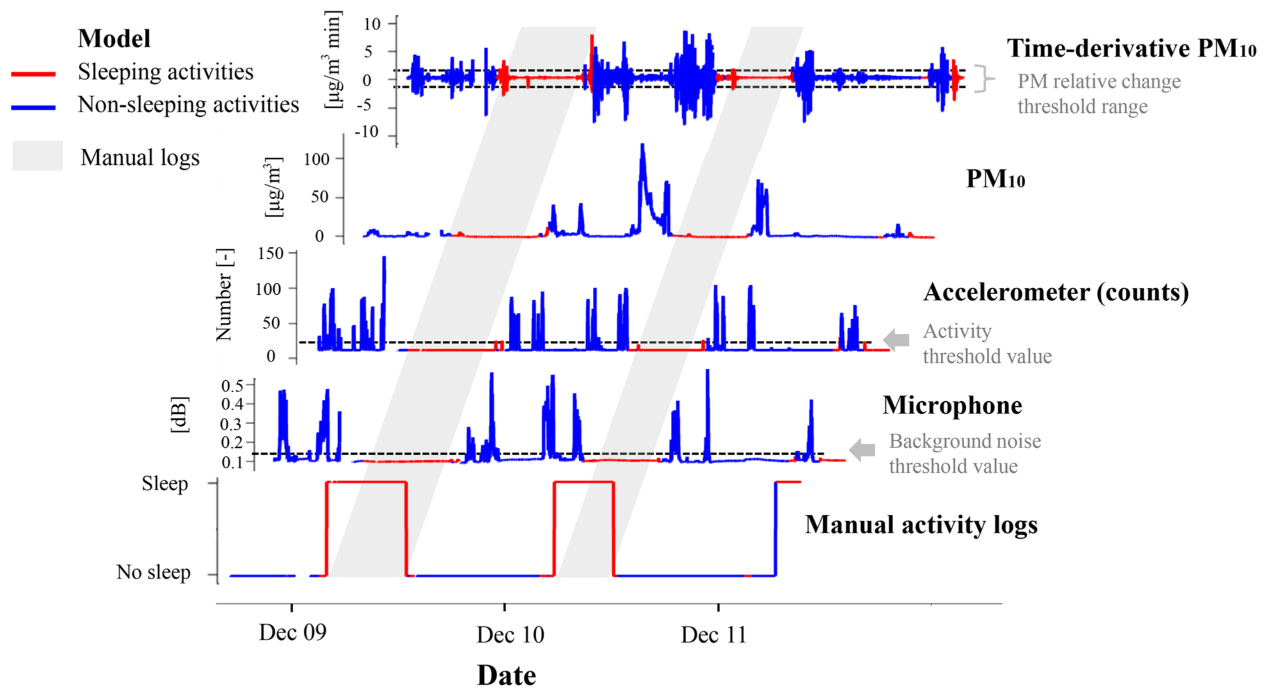

Fig. 7 Illustrative time series waterfall plot of selected PAM parameters used to classify sleep activity with a rule-based algorithm. Participant-specific thresholds (black dashed lines) were set for microphone and accelerometer levels and for the time-derivative of PM_{10}. Red line segments show time periods that the model classified as "sleep" while the blue line segments indicate non-sleep activities. Manual activity logs are presented for comparison as a time-series and as a grey shaded area

part of a journey if the participant stayed in a static location for less than 20 min (see Fig. 8a, otherwise a new journey was defined). Each journey was assigned to a "regular trajectory" [56] i.e., a continuous curve connecting successive locations of an individual recorded at regular intervals.

During a single journey, people are likely to change their mode of transport (for example, walking to the metro and then taking the train). To account for that, each trajectory was partitioned into smaller segments based on changes in patterns of movement data with the Lavielle method [57] implemented in the adehabitat LT package in R [37]. To illustrate this method, one journey is selected as a case study, partitioned automatically into two segments (Fig. 8b). These two segments of the trajectory are plotted on a map (Fig. 8c) by colour and projected on GIS (Fig. 8d) to retrieve information on public transport infrastructure and road networks. Because the points of the second segment fall on the railway network (magenta line in Fig. 8d), Segment 2 corresponds to a train ride. Manual activity logs of the participant are presented in Fig. 8e where a timing error in the activity entry in the transition between *walking* and *train* is indicated by both the GIS information and the speed derived from the distance between successive points.

Variable selection for RF

After all participant trajectories were segmented and projected on the GIS system, we had 60 variables that could be potentially used as predictors for the classification:

- 31 variables collected with the PAM: hour of the day, GPS coordinates and GPS diagnostic information (i.e., visible satellites), and extracted features from the

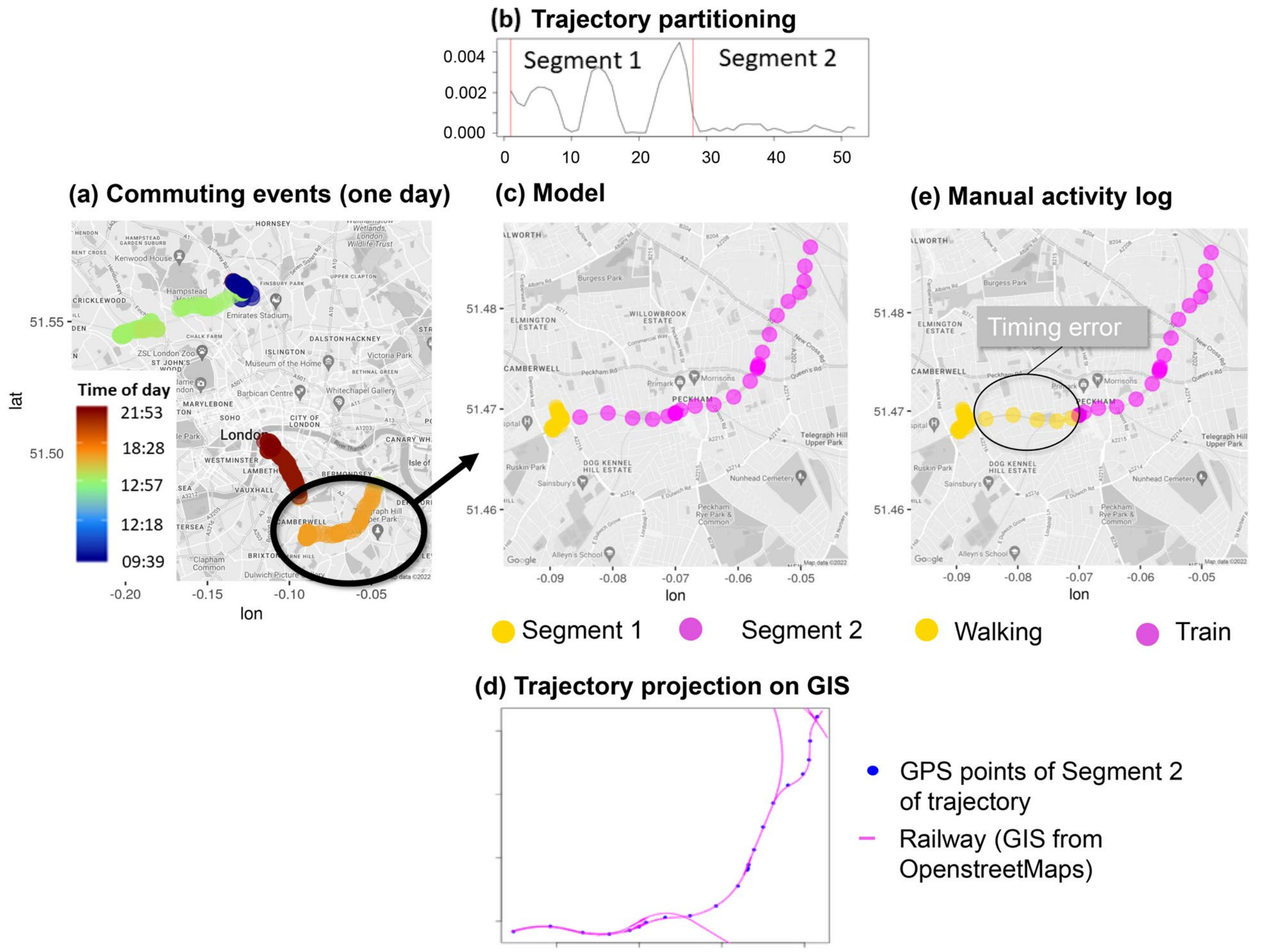

Fig. 8 Flow diagram of movement analysis implemented in adehabitat LT package in R. (a) Map of commuting events (journeys) of one participant during a typical day. The colour scheme indicates the time of day. (b) Segmentation of one trajectory (journey 18:28 in orange in a) using the Lavielle method identified two segments in the data. (c) The corresponding map of the trajectory in colour scale to differentiate the two segments. (d) Projection of segment 2 on the GIS system retrieved from Openstreetmap. The GPS points (blue) overlap with the railway infrastructure shown in magenta. (e) Corresponding map of the participant manual diary logs of that journey (see subsection 3.1). Visual inspection shows a delay in diary input that would result in small errors in model evaluation

accelerometer and microphone measurements which could have been collected with a smartphone (See full list Additional files, Table A1).

- 3 variables collected with the questionnaire: car and bicycle ownership and frequency of public transport use.
- 19 movement-phase metrics: Extracted with spatio-temporal clustering and trajectory analysis including absolute and relative angle of movement, Euclidean distance between consecutive points (speed), PAR of hulls etc. (See full list Additional files, Table A2)
- 7 variables retrieved from projecting the data on GIS: highway, railway, sidewalk, cycleway, busway and bus and train stops.

Variable selection for the classification was implemented using RF in the VSURF package [38] in R which is suitable for high dimensional datasets. This strategy does not depend on specific model hypotheses but is based on data-driven thresholds to make decisions. VSURF successively eliminates predictor variables in three steps: (1) starting with the preliminary elimination and ranking where all 60 variables were ranked by sorting the score of Variable Importance (VI) averaged over 50 RF runs. (2) In the second step, a nested collection of RF was constructed to select variables that led to the smallest out-of-the-bag (OOB) error. (3) Among those retained in the previous step, final variables for prediction were selected by constructing an ascending sequence of RF models and testing the variables in a stepwise manner. A variable was retained only if the decreased OOB error was significantly greater than the average variation obtained by adding noisy variables (Fig. 9)(calculated threshold here = 0.01).

The most important predictor variables retained with this method make intuitive sense: for *walking* and *train* the most important predictor was distance travelled, for *cycling* and driving it was the ownership of a bike and a car respectively, while for the *bus* it was the use of public transport (Fig. 9). This indicates that an equally valid approach would be to manually select and evaluate predictor variables based both on data-driven thresholds and hypothesis testing. Finally, we found that parameters extracted from GPS data with spatial and movement analysis methods (T-LOCOH and adehabitat LT) were more important predictors than raw PAM variables stressing the importance of appropriate feature extraction to optimise machine learning techniques.

RF development

Sensitivity tests were conducted for determining the maximum tree depth and number of trees. The RF was evaluated with a k-fold cross-validation method [58], which is a robust method for estimating the accuracy of a model. The dataset was split randomly into 10 mutually exclusive datasets of equal size. Then, on each iteration a new RF was trained independently on 9 subsets and evaluated on the remaining 1 subset of data, and this procedure was repeated 10 times. The final prediction error rate was calculated as the average performance metric of the 10 models. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once.

Evaluation of the time activity model

This section firstly describes the participant sample and recruitment procedures before comparing manual activity logs with model classifications.

Collection of activity logs for time-activity model evaluation

A convenience sample of 37 participants (office workers) were recruited (Additional Files, Fig. A2) via email lists and other methods. Participants were recruited from London, a megacity population ~9M and Cambridge, a relatively small UK city population ~125K, to allow evaluation of the model in different urban settings. One London and one Cambridge participant were excluded from the analysis due to incomplete diary entries (< 24h).

Upon enrolment, participants were briefed on the aims of the study, gave informed consent and filled in a standardised questionnaire of baseline information on exposure-relevant lifestyle (including e.g. car ownership), personal and demographic factors. The age distribution of the 35 participants ranged from 18 to 65 years, and were all in employment (Additional Files, Table A3).

Each participant was provided with a PAM [25] and was asked to carry it for at least one week typical of their normal activities.The average deployment time was 9 days with a minimum of 3 and a maximum of 20 days. Participants were informed that the monitors utilised GPS technology and were reassured that this information would not be accessed in real-time, but only used at the end of the study to analyse overall spatial and temporal relationships of anonymised data. No action was required by the participants to operate the PAM, other than to place it in its base-station overnight for charging and data transmission [25].

While carrying the PAM, they were asked to keep activity diaries using commercial smartphone apps [59, 60]. Smartphones were provided on request. The time-activity diary was semi-structured with some initial activities inserted in the diary as an example (e.g. *"sleeping"*). Participants were encouraged to fill in additional activities according to their lifestyles. At the end of the study, diary entries of the time-activity-location patterns were retrieved from their smartphones. Other than a personalised report of their own exposure profiles as feedback (see example Additional Files, Fig. A3), they did not receive compensation for their participation.

Overall, the participants reported 665 time-activity entries. These entries were assigned to two core categories: *location* and *activity*. Classifications were derived from the diaries by grouping similar entries together (e.g. supermarket, grocery, food shopping). Three exposure-related classifications were developed for the category location and eight classifications for activity (Additional Files, Table A4). These were integrated into the measurement dataset by labelling each data point of the time series with a numerical classifier. Activity logs were checked manually to identify periods of obviously erroneous entries, such as (a) being at two locations simultaneously; or (b) contradictory activities (e.g., *sleeping* and *cycling*) which were removed (~ 5% of the activity logs).

Aggregated participants' time budgets

Over 1.26M observations of PAM measurements at 20 sec time resolution were retained for the analysis (data capture rate 85%) and were averaged over 1-minute, resulting in N_{obs} ~422K of which ~91% had an associated manual log.

The aggregated time budgets and diurnal time-activity patterns of the participants are shown in Fig. 10. Average minutes per day spent in different microenvironments and modes of transport classified with the model show an excellent agreement with the activity logs (Fig. 10a-b), with strong linear correlation (Fig. 10c-d). In this study, the participants spent most of their time indoors at home (59.2%, min-max: 29.1%- 89.4%) or at work (16.2%, min-max: 0.0%- 41.2%), together accounting on average 75.4% of the total time budget. Time spent in other indoor static locations accounted for 9.3% (min-max: 0.0%-31.3%). Visits to outdoor microenvironments occupied only a small portion of the participants' time budget at 0.4% (min-max: 0.0%-3.9%). Travelling accounted for 5.2%, (min: 0.1% - 11.8%).

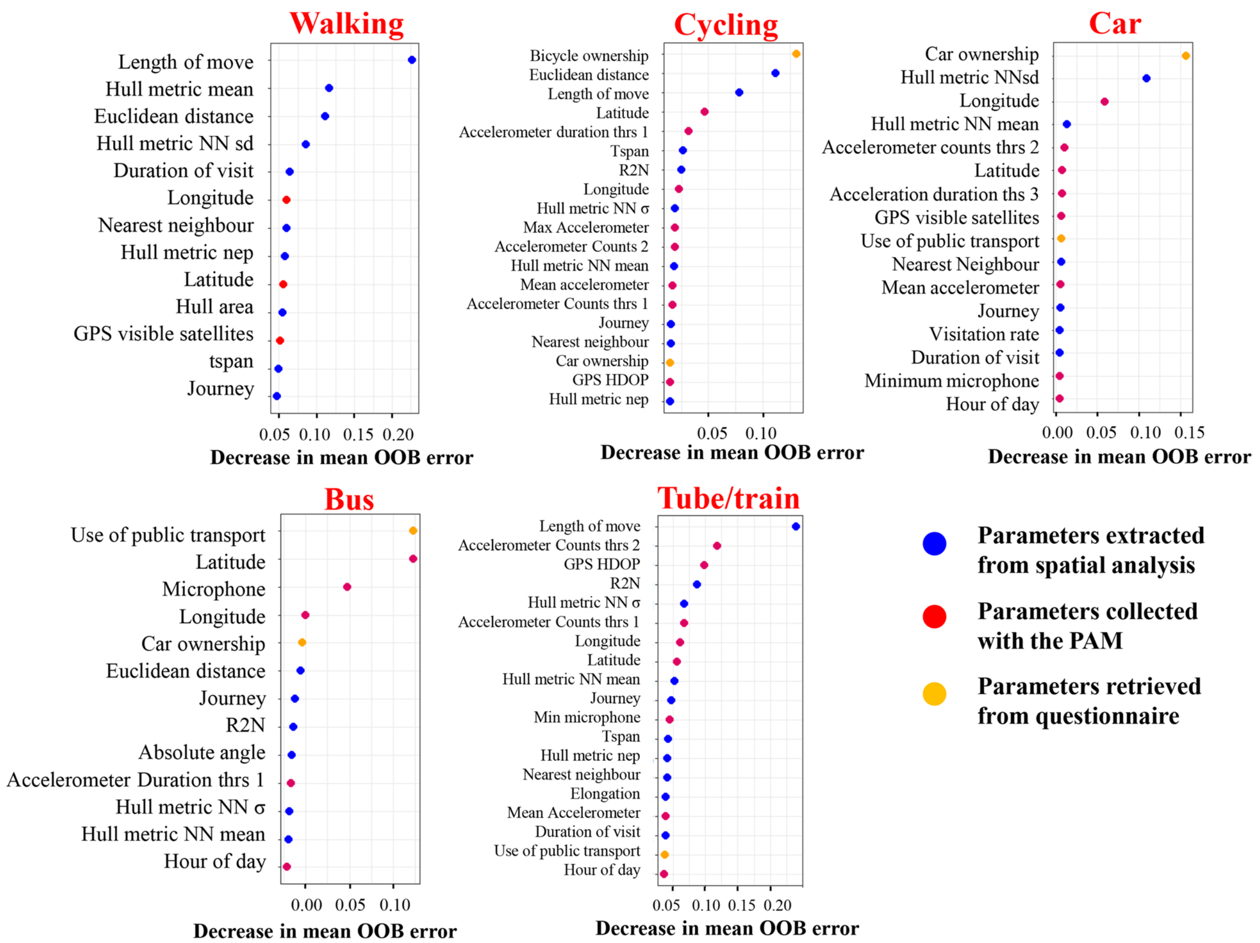

Fig. 9 Variable importance plots selected with the VSURF package in R for each mode of transport

The diurnal time budget aggregated among all participants captured by the model (Fig. 10f) agreed with the manual activity logs (Fig. 10e). The model overpredicted *other static* but underpredicted *work* possibly because participants had multiple work microenvironments but the model classified only the primary cluster as work (visited often and for extended time periods) as shown in *Step 3*. Regardless, the model managed to capture the participants' time-activity patterns well. Their patterns followed wider socio-economic patterns of adults in employment with distinctive commuting events during "rush hour" at 9:00 am and after 5:00 pm when participants returned home and stayed there until 6:00 am (Fig. 10f).

Evaluation of the time-activity model with confusion matrices

The model performance was evaluated against the manual classifications. Figure 11 visualises the confusion matrices for the binary classifications of different visited microenvironments and modes of transport.

Confusion matrices represent counts from predicted and actual values. The True Negative (TN) (blue, bottom right) shows the number of negative examples classified accurately. Similarly, True Positive (TP) (blue, top left) indicates the number of positive examples classified accurately. A False Positive (FP) (orange, top right) value corresponds to the number of actual negative examples classified as positive; and a False Negative (FN) (orange, bottom left) value is the number of actual positive examples classified as negative. We examined the accuracy (the overall effectiveness of the classifier), the sensitivity (the ability of the model to identify positive labels), the specificity (the ability of the model to identify negative labels) and the precision (the proportion of positive labels that are correctly classified) of the model. We included the F1 score, which is an overall good measure that combines precision and sensitivity and is a particularly useful indicator of model performance when there is a large number of actual negatives. The range of these metrics is 0 to 1 (or 0 to 100%). The greater the value, the better is the performance of the model.

The model performed well in classifying home (Fig. 11a) with balanced FP and FN classifications (*home*: sensitivity: 96%, specificity: 85%, precision: 90%, F1: 93%, accuracy: 91%). *Other indoor static locations* (Fig. 11d) were reliably identified with a small percentage of FP (indoor: sensitivity: 95%, specificity: 99%, precision: 86%, F1: 90%, accuracy: 98%). *Sleep* and the *work* microenvironment (Fig. 11c) were classified reasonably well though

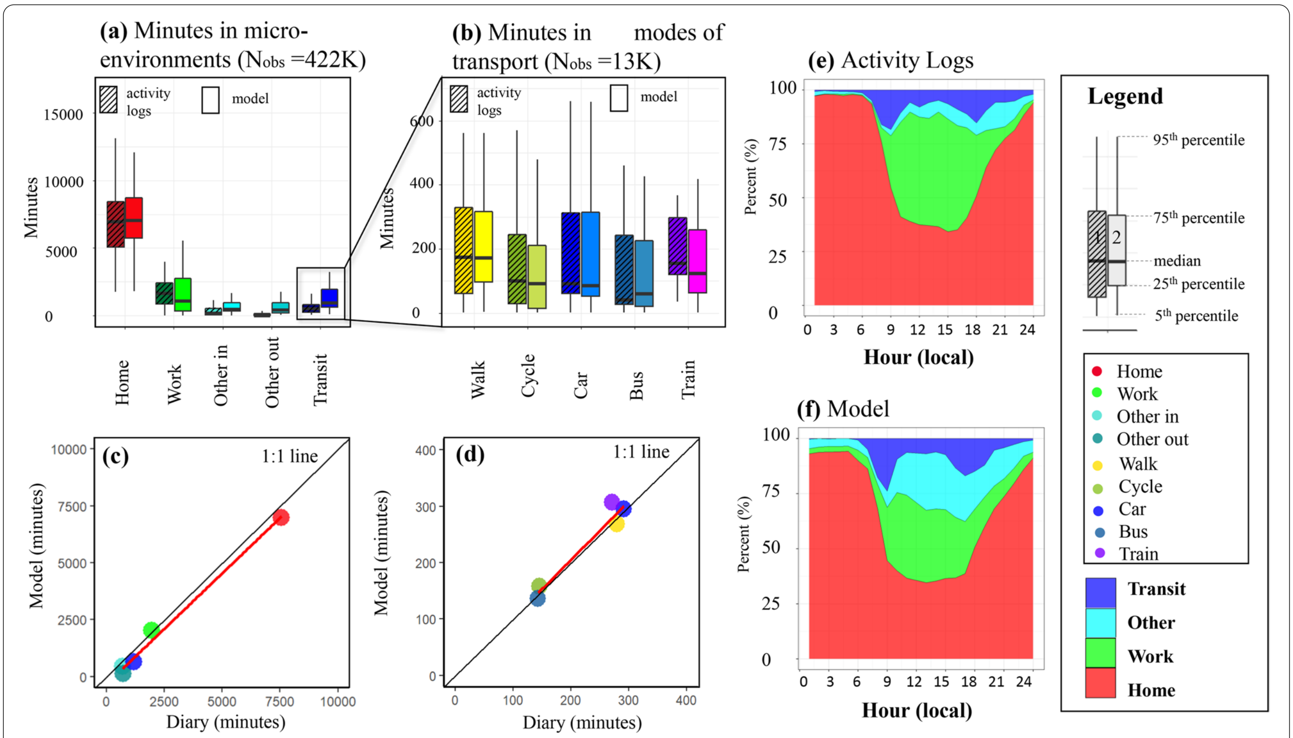

Fig. 10 Participants' time budgets. (a and b) Boxplots of participants' time budgets in different static microenvironments and modes of transport classified with activity logs (left, shaded boxplot) and the model (right, solid-colour boxplot). (c and d) Corresponding scatterplots of mean time (in minutes) spent in visited microenvironments are shown in a colour scale at the bottom. (e and f) Average diurnal time budget profile of all participants classified with the activity logs and with the model

only 26 out of 35 participants reported going to work (*sleep*: sensitivity: 79%, specificity: 80%, precision: 57%, F1: 66%, accuracy: 80%, *work*: sensitivity: 70%, specificity: 95%, precision: 72%, F1: 71%, accuracy: 90%).

The model overpredicted travel behaviour (Fig. 11b) and visits to outdoor static microenvironments (Fig. 11c) as shown by the relatively large number of observations classified as FP. Only 10 participants out of 35 reported a small fraction of time spent in outdoor static locations. As a result, while the accuracy and specificity for these activities were high (>96%), the precision and F1 score were lower (F1 *travel*: 66% and F1 *outdoor static*: 30%). A possible explanation is that logging short-duration trips and visits to outdoor locations might interfere with the ongoing activity and were therefore not recorded but were nevertheless detected by the model.

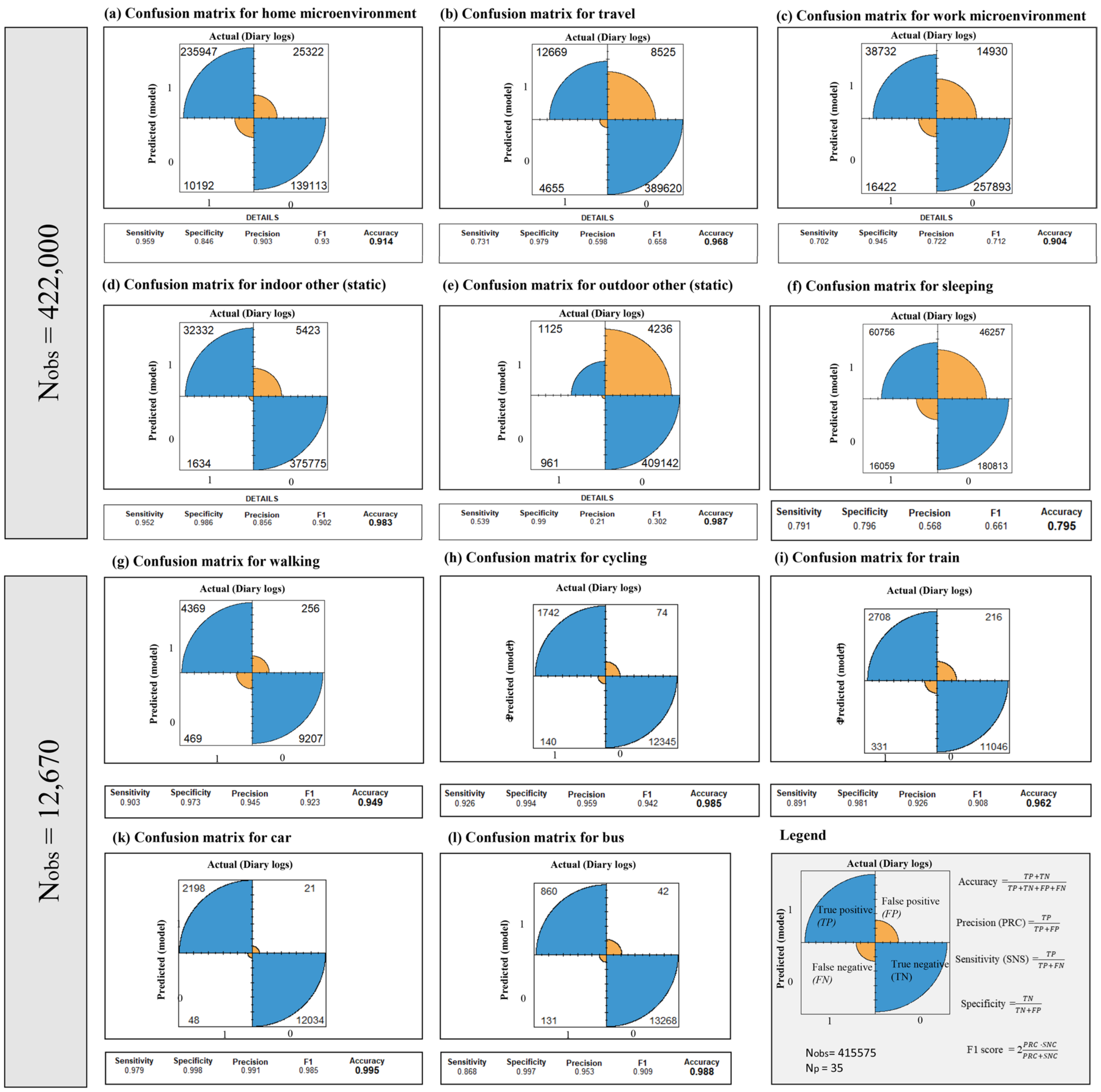

Fig. 11 Fourfold displays of confusion matrices to visualise the performance of the space-use model. Model predictions were compared against participant logs and assigned to one of four classes represented by a quarter of a circle as shown in the legend. The size of each quarter is proportional to the counts of observations belonging to that class. Blue quarters indicate correctly classified positive and negative labels while orange quarters correspond to erroneous classifications. Quantitative evaluation metrics are displayed under each fourfold plot for each visited micro-environment. (a-f) Microenvironments and activities identified with a composite model of rule-based algorithms and spatio-temporal movement analysis. (g-l) Modes of transport classified with an RF model applied to True Positive and True Negative transit observations

For this reason, periods where both the spatiotemporal-use estimator and the participant diary logs reported *travel* were retained to create a good training dataset amounting to a total of 790 trips (N_{obs}= 12670). The RF models had an excellent performance with sensitivity> 87%, specificity> 96%, precision>91%, accuracy>95% and F1 >91% (Fig. 11g-l).

Qualitative evaluation of the time-activity model

Despite the overall good performance of the model in classifying static microenvironments and modes of transport, we nevertheless detected inconsistencies between manual logs and model classifications. The first part uses a representative case-study participant to illustrate such inconsistencies originating either from limitations of the model itself or errors in the manual activity logs. The second part aims to understand the implications of these inconsistencies for the overall personal exposure estimations by comparing the resulting personal concentrations in different microenvironments classified with either one of the two methods for all participants and in doing so to demonstrate how automated models such as the one presented here can enhance air pollution health studies by providing a comprehensive picture of air pollution health risks in daily life.

Proof-of-concept for an example case-study participant

The case study shows a representative largely sedentary office worker who commuted via cycling and walking to work and visited other indoor and outdoor microenvironments (Fig. 12). The visual inspection of the maps in Fig. 12a and b indicates that the model performance surpasses manual classification mostly due to small timing errors as the participant may have had difficulty documenting the precise time of microenvironment transitions. For example, a walking trip through the park is erroneously classified as *work* microenvironment (timing error 2, Fig. 12a). The diary was less likely to specify visits to outdoor microenvironments compared with the model (misclassified *other outdoor static*, Fig. 12a).

Figure 12c presents the time series of one typical day. The participant commuted to work on foot at around 09:00 am, stayed there until 19:00 pm and walked back home choosing a different route this time. While both methods adequately captured the participant's time-activity patterns, the manual activity model had some missing observations and timing errors. In both trips a clear spike in all pollutants' levels was noticed: $PM_{2.5}$ reached maximum daily concentrations during the morning walk while NO_2 reached maximum daily concentrations during the evening walk (Fig. 12c). The participant spent the rest of the evening cooking, resting and visiting a nearby indoor environment on foot before returning home for the night. Indoor $PM_{2.5}$ levels at home were higher than in the work environment consistent with indoor emission sources during evening cooking activities.

Personal concentrations in visited microenvironments

Figure 13 visualises the concentrations in different microenvironments visited by all 35 participants (N_{obs} ~ 422K) classified both with the manual logs and the model. The distribution of concentrations of individual pollutants in each microenvironment was visualised with boxplots (Fig. 13a). On the left-hand side, the hatched boxplot shows observations classified with the manual activity logs while the solid-colour boxplot shows observations classified with the automated model.

The corresponding scatterplots of the mean concentrations in each microenvironment are shown in Fig. 13f-k in a colour scale. Most points fall on the one-to-one line indicating that classifying microenvironments with either one of the two methods resulted in insignificant differences between estimated concentrations. *Other out* was the most poorly classified microenvironment (Fig. 11e) possibly because the whole dataset contained less than 20 participant-hours reported to be spent outside (Fig. 10a). Figure 13f-k shows that mean concentrations estimated for *other out* microenvironments had the highest deviation from the one-to-one line particularly for ozone and particulate matter ($PM_{2.5}$). The model overpredicted mean ozone concentrations compared with the activity logs. Because higher ozone levels are generally expected to be seen outdoors (Fig. 6e) due to higher levels of photochemistry, the model classifications likely outperformed the manual activity logs.

Travelling in particular occupied only a small fraction of the total time budget (on average 5.2% of the participants' time, Fig. 10a), but is a significant site of exposure (Fig. 13). Because the sample of this study is small, some caution must be applied to the interpretation and the generalisability of that finding. Participants in both cities covered large spatial distances (Fig. 14). Cambridge participants covered a smaller spatial area compared with the London participants and primarily used active modes of transport (walking, cycling). In line with previous research [61], it seems that vehicle users (car and bus) are exposed to significantly

(See figure on next page.)

Fig. 12 Comparison of manual logs and automated time activity model for one case study participant. Colour-coded maps illustrating visited microenvironments and modes of transport during a week of a representative participant. (a) Classifications according to the activity log. (b) Classifications according to the automated activity model. Google maps 2021. (c) Time series of the manual activity log, model classifications and selected PAM parameters for one typical day

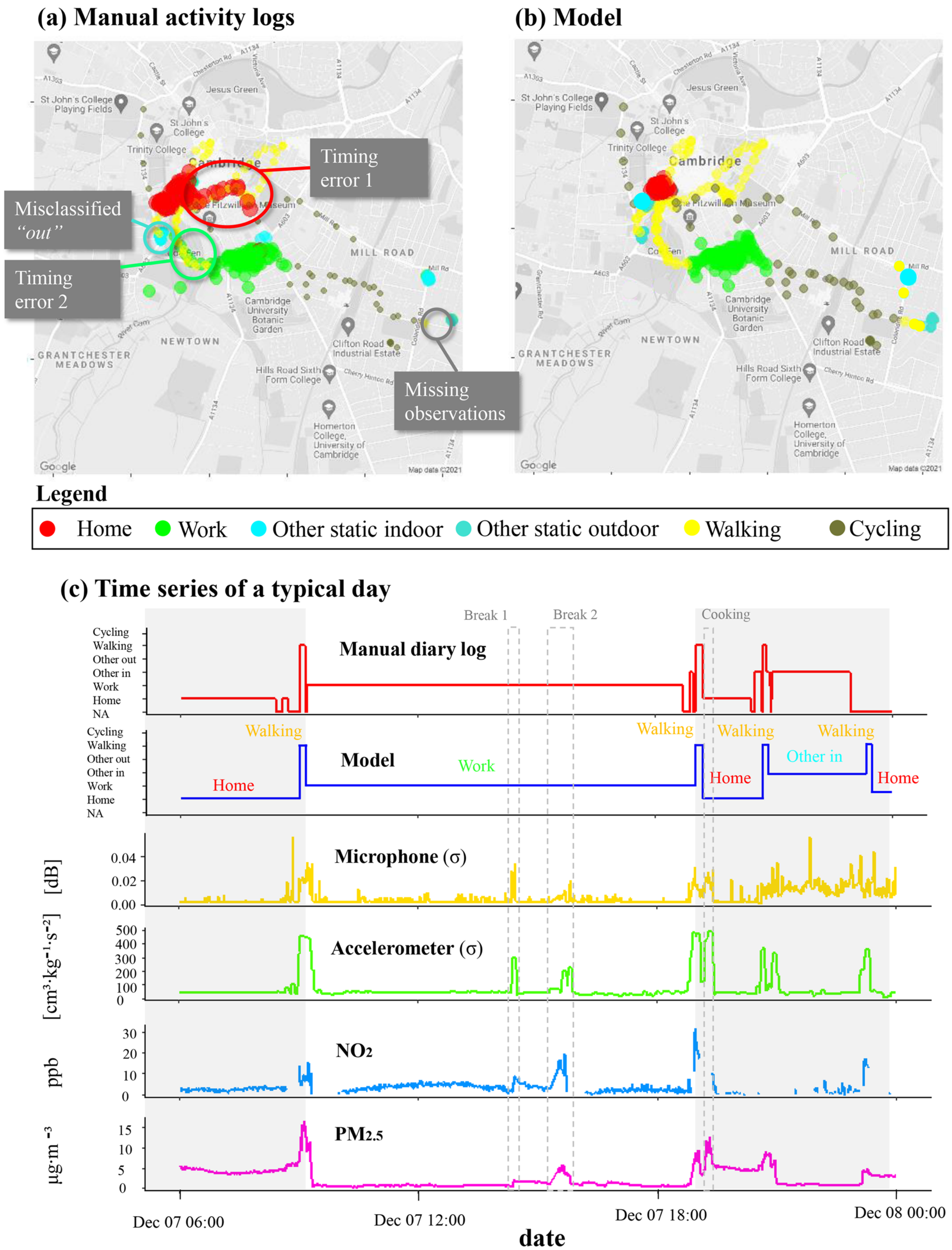

Fig. 12 (See legend on previous page.)

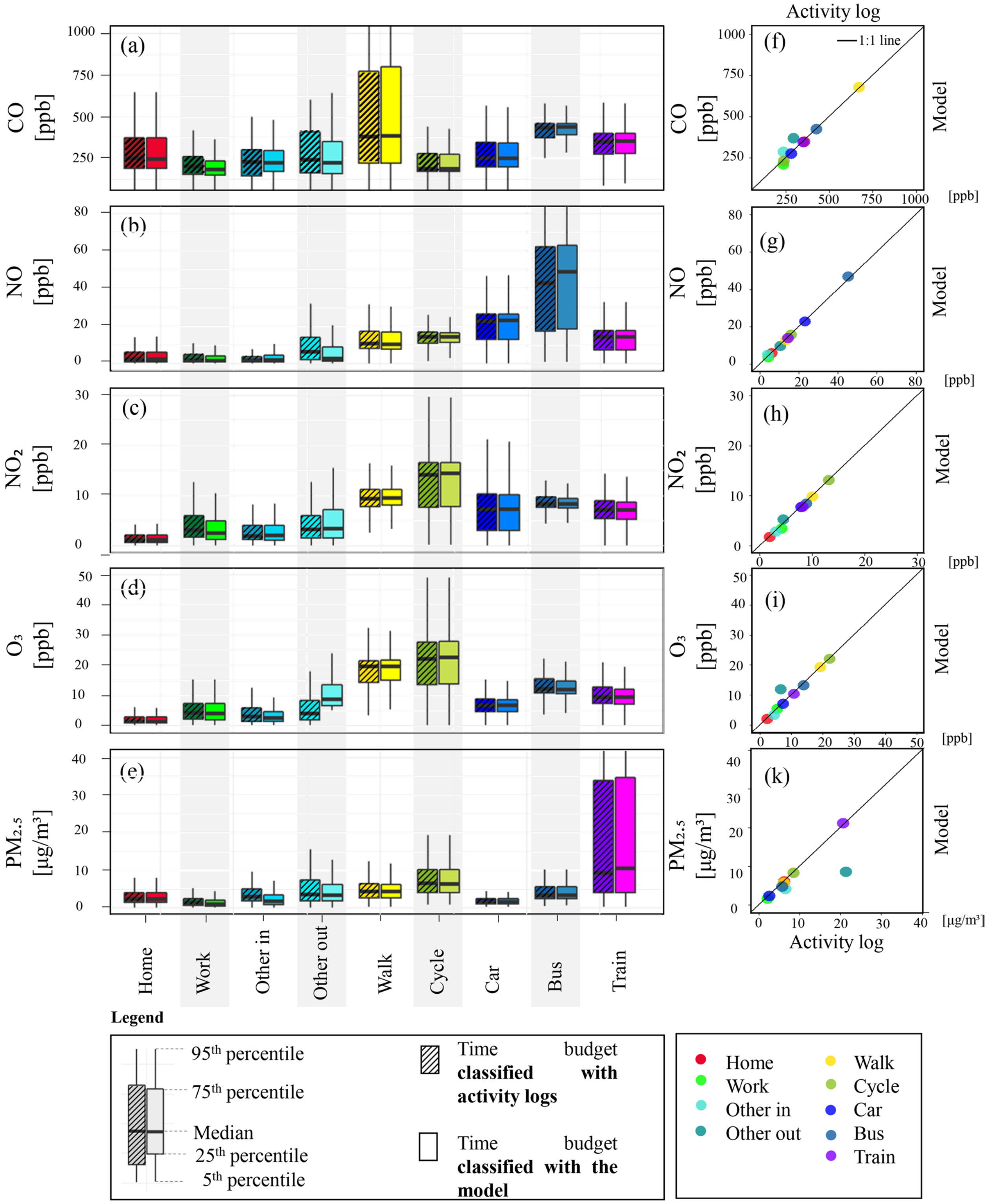

Fig. 13 Boxplots and scatter plots of personal exposure of 35 UK participants to multiple pollutants in different microenvironments. (a-e) For each activity, the left hatched boxplot shows entries classified with participants' activity logs and the right solid-colour boxplot with the automated model. (f-k) Mean concentrations of individual pollutants in visited microenvironments are shown in a colour-scale in scatter plots. The 1:1 line is in black

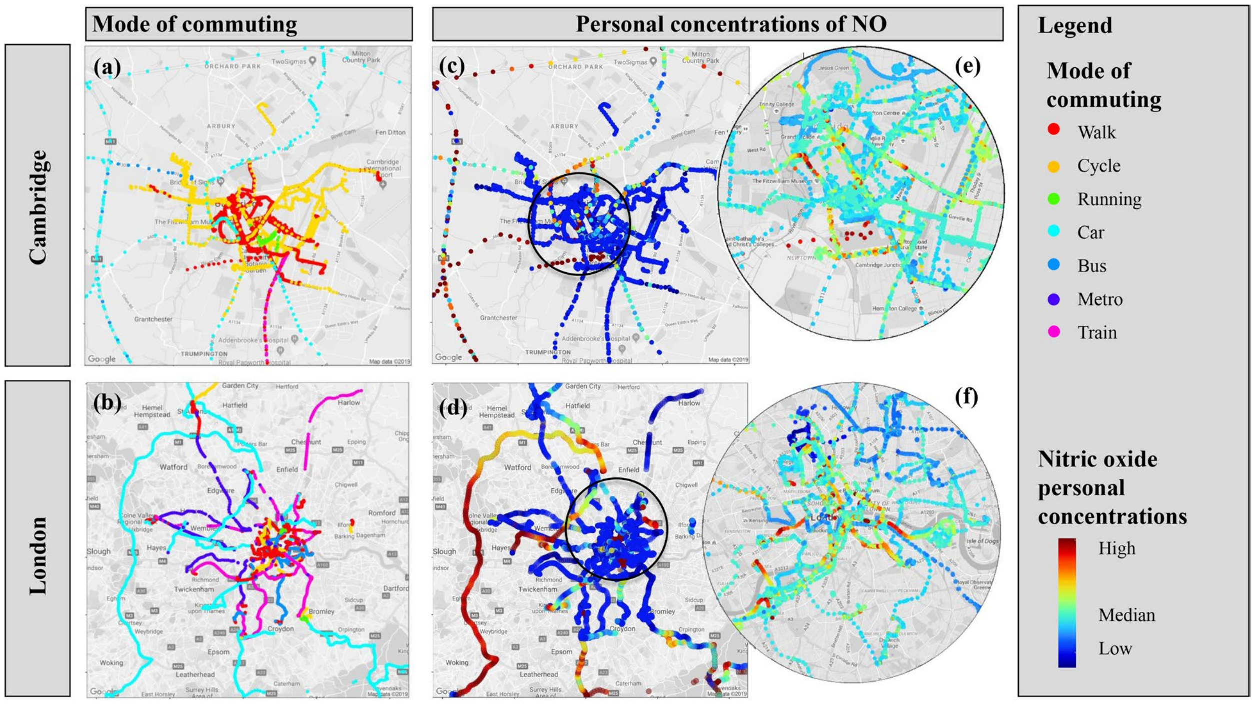

Fig. 14 Transportation modes and relative exposure to air pollution of 35 participants plotted on maps. (a) Cambridge and (b) London visualising modes of transport (c -f) Relative exposure to pollution (in this case NO) in Cambridge and London respectively shown in a colour-scale. Map data Google 2021

higher NO concentrations than cyclists or pedestrians (Fig. 13b), who appear to be exposed to higher NO_2 and O_3 levels(Fig. 13c-d). While this study is only a snapshot of exposure in transit, it seems that maximum air pollution levels (in this case NO) were encountered when travelling in major traffic arteries (for example M25 in the greater London area Fig. 14d) or the central bus station (Fig. 14e) and in areas where traffic is routinely static (i.e. bridges in London, Fig. 14f). Confirming previous research [62], the highest exposure to particulate matter ($PM_{2.5}$) was encountered by commuters using the train/metro system(Fig. 13e).

Discussion

Mobile sensor deployments can provide a picture of the rapidly changing and highly granular personal concentrations in a way that has not been possible before. This paper demonstrated a methodological framework that expands the capabilities of validated sensor platforms [25] with advanced computational methods to integrate time-activity patterns in personal exposure estimations.

Implementation of the model in different ways and programming languages

The parameters used in the time-activity model as predictors can be collected with smartphones making the method applicable more widely than with the specific sensor platforms. The model is readily extendable to include outputs from wearable biosensors in smartphones, such as heart and respiratory rate.

We employed multidisciplinary tools from the fields of movement ecology and AI and extended their use in human mobility studies to build a composite model that automatically classifies major time-activity location patterns of static spatial clusters and five modes of transport. We developed the model in R, an open-source free software environment, but equivalent algorithms can be developed in other programming languages that have similar capabilities for spatial and statistical analysis, such as Python.

Limitations

There are certain caveats with the methodology employed to develop and evaluate the time-activity model. First, a high rate of false positives was detected for outdoor and in-transit microenvironments, although these activities generally take up a small percentage of participants' time. We hypothesise that this is not due to limitations in the model's accuracy, but a limitation of manual activity logs employed in the evaluation. Even the most compliant participants may have difficulty correctly documenting the precise time of microenvironment transitions, as it might interfere with the ongoing activity. Secondly, due to the increased participation burden, the sample size of 35 participants was relatively small; however, previous research on time-activity patterns and transportation mode classification has reported that a sample size of around 30 participants is adequate to provide robust estimations of activity patterns [24, 63].

Main findings

The model had an overall good performance: the classification for static microenvironments had an F1-score for *home* of 0.93; for *work* of 0.71; for *other indoor static* of 0.9. The RF model for transportation mode classification had an excellent performance (F1 > 0.88). We found that the difference in concentrations of multiple pollutants in the nine microenvironments classified with either model or activity log was insignificant compared with the large spatial and temporal variation of personal exposure concentrations during daily life.

In line with previous research, street-level modes of commuting were associated with the highest levels of NO_2 and O_3 concentrations [61], in-vehicle trips (car and bus) were associated with marked exposure to NO [61] while the metro was associated with the highest exposure to PM [62]. These noticeable variations in concentrations between different microenvironments result in diverse personal exposures emphasising the potential for exposure misclassification when purely ecological (home location-based) exposure estimations are used in epidemiological research.

Future work

The next step involves the application of the model on larger health panel studies [30, 31] of hundreds of participants to characterise the exposure of vulnerable subgroups of the population in diverse geographical settings. As physical activity may lead to differing doses for similar exposures, future work aims to capture total personal multi-pollutant dose in unprecedented detail addressing a major gap in air pollution epidemiology. We will further investigate whether physical activity levels may be reliable physical, psychological, social, and cognitive health indicators for elderly and chronically ill cohort participants.

More importantly, as the pollution mixture inhaled during different activities likely originates from different emission sources, it may contain different chemicals with varying potential toxicity [64]. Therefore, neglecting the activity component in air pollution dose-health relationships might lead to erroneous conclusions regarding the toxicity of air pollutants. The time activity model enables the dissagregation of total personal exposure into different microenvironment-specific exposures from diverse emission sources and chemical sinks. Together with advanced source apportionment methods of personal exposure, future work aims to explore source-specific health effects.

Conclusions

Novel sensor technologies and computational techniques such as those demonstrated here have advantages over traditional time-activity-location diaries, which are laborious, prone to error and involve a limited number of participants. Collecting a wealth of time-activity information in unprecedented detail can increase our understanding of air pollution exposures and exposure-related behaviours that may be harmful to human health. Because individuals may have different susceptibilities to environmental exposures, together with the advancing field of *"-omics"*, this work builds towards providing comprehensive personalised advice to the individual to reduce their environmental health risks based on their unique health requirements and lifestyle.

Abbreviations

AI: Artificial Intelligence; α-NN: adaptive Nearest Neighbours; CART: Classification And Regression Trees; CO: Carbon monoxide; GIS: Geographic Information System; GPS: Global Positioning System; NO: Nitric oxide; NO_2: Nitrogen dioxide; O_3: Ozone; OOB: Out-Of-Bag (error); PAM: Personal Air quality Monitor; PM: Particulate Matter; QA/QC: Quality Assurance/Quality Control; RF: Random Forest; RH: Relative Humidity; VI: Variable Importance; PAR: Perimeter-to-Area Ratio.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12940-022-00939-8.

Additional file 1: The Personal air pollution monitor and data procedures. A brief description of the PAM sensor platform and data cleaning/feature extraction for the GPS coordinates, accelerometer and microphone readings. **Dealing with missing GPS observations.** Satellite signal loss in indoor environments is common. This section describes the rule-based algorithm developed to interpolate missing locations. **Variables evaluated for mode of transport classification.** Description of all PAM variables and extracted variables from spatio-temporal movement analysis used for RF model development. **Participant recruitment and feedback.** Descriptive summary of participants' characteristics, recruitment timeline, example of personal exposure feedback and grouping of manual logs into main categories.

Acknowledgements

We would like to thank all volunteers for participating in the study.

Authors' contributions

The paper was conceptualised by LC, BB and RLJ. The sensor platform was developed by LC and MK. The data curation was performed by LC and AK. LC, AK and RLJ contributed to the formal data analysis and data visualisation. Resources were provided by BB, FJK and RLJ. The original draft was written by LC and AK and reviewed and edited by all authors. The author(s) read and approved the final manuscript.

Funding

This research has been supported by the Medical Research Council of the UK (MR\L019744\1, COPE project) and MRC-PHE Centre for Environment and Health (MR\L01341X\1).

Availability of data and materials
The datasets generated and\or analysed during the current study are not publicly available due to sensitive information but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The pilot study received ethical approval by the King's College London ethics committee (Study Reference LRS15\162000).

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Rd, CB2 1EW Cambridge, UK. [2]Institute for Chemistry, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany. [3]Atmospheric Sensors Ltd, SG19 3SH Bedfordshire, UK. [4]Environmental Research Group, MRC Centre for Environment and Health, Imperial College London, W12 0BZ London, UK. [5]BIC-ESAT and SKL-ESPC, College of Environmental Sciences and Engineering, Center for Environment and Health, Peking University, 100871 Beijing, China.

Received: 1 March 2022 Accepted: 7 November 2022
Published online: 09 December 2022

References
1. Murray CJ, Aravkin AY, Zheng P, Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1223–49.
2. Özkaynak H, Baxter LK, Dionisio KL, Burke J. Air pollution exposure prediction approaches used in air pollution epidemiology studies. J Expo Sci Environ Epidemiol. 2013;23(6):566–72.
3. McDuffie E, Martin R, Yin H, Brauer M. Global Burden of Disease from Major Air Pollution Sources (GBD MAPS): A Global Approach. Res Rep Health Eff Inst. 2021;2021:210.
4. Chatzidiakou L, Krause A, Han Y, Chen W, Yan L, Popoola OA, et al. Using low-cost sensor technologies and advanced computational methods to improve dose estimations in health panel studies: Results of the AIRLESS project. J Expo Sci Environ Epidemiol. 2020;30(6):981–9.
5. Dionisio KL, Baxter LK, Burke J, Özkaynak H. The importance of the exposure metric in air pollution epidemiology studies: when does it matter, and why? Air Qual Atmos Health. 2016;9(5):495–502.
6. 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. Sci Total Environ. 2014;490:785–97.
7. Huang J, Deng F, Wu S, Guo X. Comparisons of personal exposure to PM2. 5 and CO by different commuting modes in Beijing, China. Sci Total Environ. 2012;425:52–59.
8. Moya J, Phillips L, Schuda L, Wood P, Diaz A, Lee R, et al. Exposure factors handbook: 2011 edition. US Environmental Protection Agency; 2011.
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. J Expo Sci Environ Epidemiol. 2001;11(3):231–52.
10. Elgethun K, Yost MG, Fitzpatrick CT, Nyerges TL, Fenske RA. Comparison of global positioning system (GPS) tracking and parent-report diaries to characterize children's time-location patterns. J Expo Sci Environ Epidemiol. 2007;17(2):196–206.
11. Kelly P, Krenn P, Titze S, Stopher P, Foster C. Quantifying the difference between self-reported and global positioning systems-measured journey durations: a systematic review. Transp Rev. 2013;33(4):443–59.
12. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. A practical guide to measuring physical activity. J Acad Nutr Diet. 2014;114(2):199.
13. 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. Environ Pollut. 2013;176:92–9.
14. Nyhan M, Kloog I, Britter R, Ratti C, Koutrakis P. Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data. J Expo Sci Environ Epidemiol. 2019;29(2):238–47.
15. Tang R, Tian L, Thach TQ, Tsui TH, Brauer M, Lee M, et al. Integrating travel behavior with land use regression to estimate dynamic air pollution exposure in Hong Kong. Environ Int. 2018;113:100–8.
16. Yu H, Russell A, Mulholland J, Huang Z. Using cell phone location to assess misclassification errors in air pollution exposure estimation. Environ Pollut. 2018;233:261–6.
17. Hu M, Li W, Li L, Houston D, Wu J. Refining time-activity classification of human subjects using the global positioning system. PLoS ONE. 2016;11(2): e0148875.
18. Stamatelopoulou A, Chapizanis D, Karakitsios S, Kontoroupis P, Asimakopoulos D, Maggos T, et al. Assessing and enhancing the utility of low-cost activity and location sensors for exposure studies. Environ Monit Assess. 2018;190(3):1–12.
19. Adams C, Riggs P, Volckens J. Development of a method for personal, spatiotemporal exposure assessment. J Environ Monitor. 2009;11(7):1331–9.
20. Breen MS, Long TC, Schultz BD, Crooks J, Breen M, Langstaff JE, et al. GPS-based microenvironment tracker (MicroTrac) model to estimate time-location of individuals for air pollution exposure assessments: Model evaluation in central North Carolina. J Expo Sci Environ Epidemiol. 2014;24(4):412–20.
21. Kim T, Lee K, Yang W, Do YuS. A new analytical method for the classification of time-location data obtained from the global positioning system (GPS). J Environ Monit. 2012;14(8):2270–4.
22. Glasgow ML, Rudra CB, Yoo EH, Demirbas M, Merriman J, Nayak P, et al. Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. J Expo Sci Environ Epidemiol. 2016;26(4):356–64.
23. Quinn C, Anderson GB, Magzamen S, Henry CS, Volckens J. Dynamic classification of personal microenvironments using a suite of wearable, low-cost sensors. J Expo Sci Environ Epidemiol. 2020;30(6):962–70.
24. Wu J, Jiang C, Houston D, Baker D, Delfino R. Automated time activity classification based on global positioning system (GPS) tracking data. Environ Health. 2011;10(1):1–13.
25. Chatzidiakou L, Krause A, Popoola OA, Di Antonio A, Kellaway M, Han Y, et al. Characterising low-cost sensors in highly portable platforms to quantify personal exposure in diverse environments. Atmos Meas Tech. 2019;12(8):4643–57.
26. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. https://www.R-project.org/.
27. RStudio Team. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston; 2021. http://www.rstudio.com/.
28. PostgreSQL B. PostgreSQL. 1996. http://www.PostgreSQL.org/about. Accessed 1 Feb 2022.
29. Conway J, Eddelbuettel D, Nishiyama T, Prayaga SK, Tiffin N. RPostgreSQL: R Interface to the 'PostgreSQL' Database System. 2021. R package version 0.7-3. https://CRAN.R-project.org/package=RPostgreSQL. Accessed 1 Feb 2022.
30. Moore E, Chatzidiakou L, Jones RL, Smeeth L, Beevers S, Kelly FJ, et al. Linking e-health records, patient-reported symptoms and environmental exposure data to characterise and model COPD exacerbations: protocol for the COPE study. BMJ Open. 2016;6(7): e011330.
31. Han Y, Chen W, Chatzidiakou L, Krause A, Yan L, Zhang H, et al. Effects of AIR pollution on cardiopuLmonary disEaSe in urban and peri-urban reSidents in Beijing: protocol for the AIRLESS study. Atmos Chem Phys. 2020;20(24):15775–92.
32. Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, et al. A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci. 2008;105(49):19052–9.
33. Fleming CH, Fagan WF, Mueller T, Olson KA, Leimgruber P, Calabrese JM. Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator. Ecology. 2015;96(5):1182–8.

34. Lyons AJ, Turner WC, Getz WM. Home range plus: a space-time characterization of movement over real landscapes. Mov Ecol. 2013;1(1):1–14.
35. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
36. Martin BD, Addona V, Wolfson J, Adomavicius G, Fan Y. Methods for real-time prediction of the mode of travel using smartphone-based GPS and accelerometer data. Sensors. 2017;17(9):2058.
37. Calenge C. Analysis of animal movements in R: the adehabitatLT package. Vienna: R Foundation for Statistical Computing; 2011.
38. Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. Pattern Recogn. 2010;31(14):2225–36.
39. Pebesma E, Bivand RS. Classes and methods for spatial data in R: the sp package. R News. 2005;5(2):9–13. https://CRAN.R-project.org/doc/Rnews/.
40. Bivand RS, Pebesma E, Gomez-Rubio V. Applied spatial data analysis with R, Second edition. New YorK: Springer; 2013. https://asdar-book.org/.
41. Bivand R, Keitt T, Rowlingson B, Pebesma E, Sumner M, Hijmans R, et al. rgdal: Bindings for the geospatial data abstraction library. R package version 15-27. 2019;1:4–8. https://CRAN.R-project.org/package=rgdal.
42. Hijmans RJ. raster: Geographic Data Analysis and Modeling. 2021. R package version 3.5-2. https://CRAN.R-project.org/package=raster. Accessed 1 Feb 2022.
43. Peng R, Murdoch D, Rowlingson B, Alan M. gpclib: General Polygon Clipping Library for R. 2020. R package version 1.5-6. https://CRAN.R-project.org/package=gpclib. Accessed 1 Feb 2022.
44. Fellows I, Stotz J. OpenStreetMap: Access to Open Street Map Raster Images. 2019. R package version 0.3.4 3. https://CRAN.R-project.org/package=OpenStreetMap. Accessed 1 Feb 2022.
45. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Verlag New YorK: Springer; 2016. https://ggplot2.tidyverse.org.
46. Kahle D, Wickham H. ggmap: Spatial Visualization with ggplot2. R J. 2013;5(1):144–161. https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.
47. Morgan-Wall T. rayshader: Create Maps and Visualize Data in 2D and 3D. 2021. R package version 0.24.10. https://CRAN.R-project.org/package=rayshader. Accessed 1 Feb 2022.
48. Carslaw DC, Ropkins K. openair – An R package for air quality data analysis. Environ Model Softw. 2012;27–28:52–61.
49. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. 2021. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr. Accessed 1 Feb 2022.
50. Soetaert K. plot3D: Plotting Multi-Dimensional Data. 2021. R package version 1.4. https://CRAN.R-project.org/package=plot3D. Accessed 1 Feb 2022.
51. Kuhn M. caret: Classification and Regression Training. 2021. R package version 6.0-90. https://CRAN.R-project.org/package=caret. Accessed 1 Feb 2022.
52. Hahsler M, Piekenbrock M, Doran D. dbscan: Fast Density-Based Clustering with R. J Stat Softw. 2019;91(1):1–30.
53. Walter WD, Onorato DP, Fischer JW. Is there a single best estimator? Selection of home range estimators using area-under-the-curve. Mov Ecol. 2015;3(1):1–11.
54. Nazaroff W, Gadgil AJ, Weschler CJ. Critique of the use of deposition velocity in modeling indoor air quality. ASTM Spec Tech Publ. 1993;1205:81–81.
55. Chatzidiakou L, Mumovic D, Summerfield AJ. What do we know about indoor air quality in school classrooms? A critical review of the literature. Intell Build Int. 2012;4(4):228–59.
56. Calenge C, Dray S, Royer-Carenzi M. The concept of animals' trajectories from a data analysis perspective. Ecol Inform. 2009;4(1):34–41.
57. Lavielle M. Detection of multiple changes in a sequence of dependent variables. Stoch Process Appl. 1999;83(1):79–102.
58. Kim JH. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Comput Stat Data Anal. 2009;53(11):3735–45.
59. ATracker – Daily Task and Time Tracking. https://atracker.pro/home.html. Accessed 1 Feb 2022.
60. aTimeLogger – The beautiful way to track your time. http://www.atimelogger.com/. Accessed 1 Feb 2022.
61. Mead M, Popoola O, Stewart G, Landshoff P, Calleja M, Hayes M, et al. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos Environ. 2013;70:186–203.
62. Smith J, Barratt B, Fuller G, Kelly F, Loxham M, Nicolosi E, et al. $PM_{2.5}$ on the London Underground. Environ Int. 2020;134:105188.
63. Brondeel R, Pannier B, Chaix B. Using GPS, GIS, and Accelerometer Data to Predict Transportation Modes. Med Sci Sports Exerc. 2015;47(12):2669–75.
64. Kelly FJ, Fussell JC. Linking ambient particulate matter pollution effects with oxidative biology and immune responses. Ann N Y Acad Sci. 2015;1340(1):84–94.
65. Liaw A, Wiener M, et al. Classification and regression by randomForest. R News. 2002;2(3):18–22.

Publisher's Note

A.1.3 The health effects of outdoor air pollution

This appendix gives an overview of the literature including major studies and previous systematic meta-analyses. The appendix focuses on the associations that have been made between outdoor pollution and health. For the reasons mentioned in Section 3.1, studies of Chinese cohorts based in China, where possible, will be a focus. Four pollutants ($PM_{2.5}$, NO_x, O_3 and CO) are introduced and their health effects are reviewed. These pollutants have been identified by the United States Environmental Protection Agency (EPA) as "pollutants that harm your health and the environment"[135].

Particulate Matter

Particulate matter (PM) are tiny particles of solid or liquid suspended in the air. They are classified by their diameter, with $PM_{2.5}$ referring to PM with a diameter of less than 2.5 micrometers. Figure A.1 provides a comparison of the size of $PM_{2.5}$ with the cross-section of a human hair.

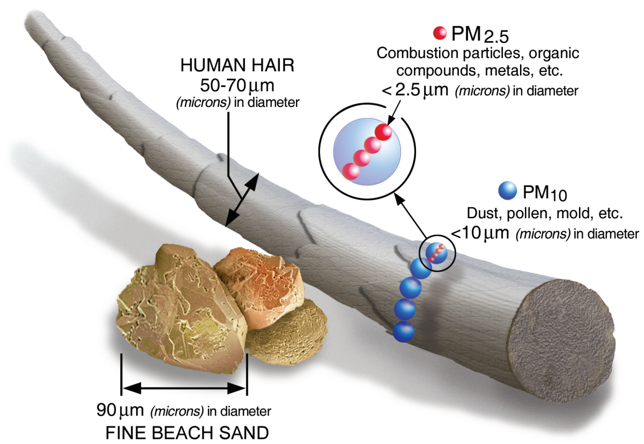

Figure A.1: The size of $PM_{2.5}$: Visual comparison of the diameter of $PM_{2.5}$ with respect to the diameter of a human hair. It has been created by the EPA and can be found on the website: `https://www.epa.gov/pm-pollution/particulate-matter-pm-basics`

Small particles are of particular interest with respect to health as they have the ability to penetrate deep into the lungs, enter the alveoli on the bronchioles, and even pass through the air-blood barrier into the blood stream[96], as shown in Figure A.2.

A strong connection between $PM_{2.5}$ and mortality has been established. The landmark "Harvard Six Cities" published in 1993 found that the risk of death in high polluted areas, across six US cities was 26% higher in high polluted areas than in low polluted areas[35], and that mortality was strongly associated with the levels of

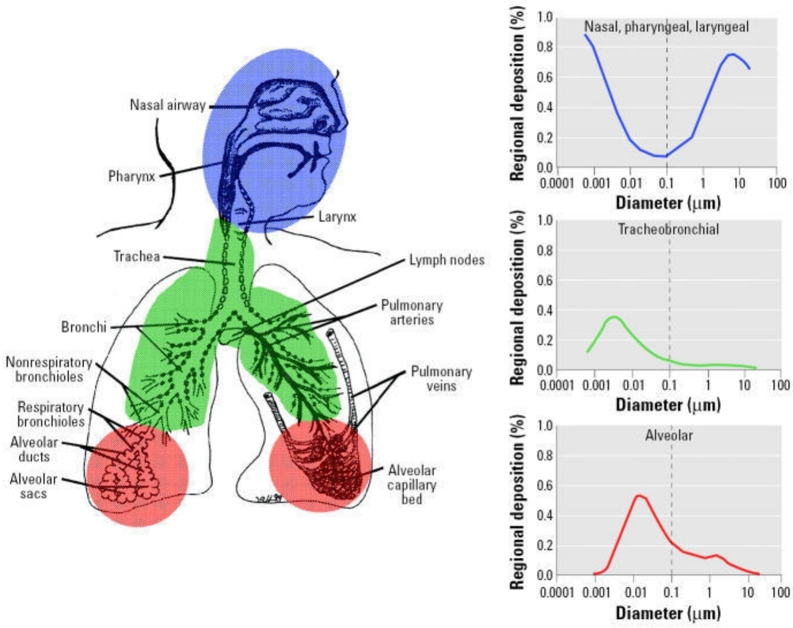

Figure A.2: Deposition of particles in the lungs: The predicted fractional deposition of atmospheric particles in the respiratory system as a function of particle diameter. Fine particles exhibit a tendency to penetrate deeper into the tracheobronchial (green) and alveolar (red) regions of the respiratory tract. Reproduced by Dr Andrea Di Antonio, from Oberdorster et al.[99]

fine particles, as shown in Figure A.3. In 2006, Laden et al. conducted an eight-year follow-up study during a period of decreased $PM_{2.5}$ levels (1990-1998) and observed that the decline in air pollution was linked to reduced mortality risk.

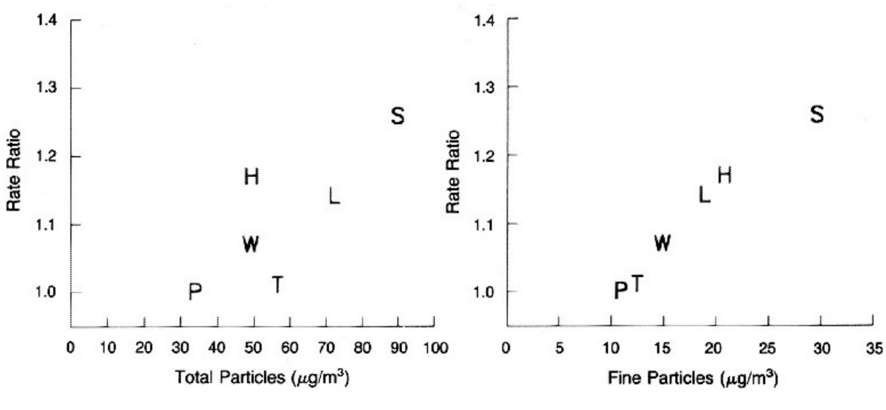

Figure A.3: Harvard six cities study: Mean values are shown for the measures total particles (diameter of less than 10 micrometers) and fine particles ($PM_{2.5}$). Mortality was more strongly associated with the levels of fine particles than with the levels of total particles. P denotes Portage, Wisconsin; T Topeka, Kansas; W Watertown, Massachusetts; L St. Louis; H Harriman, Tennessee; and S Steubenville, Ohio. This figure is taken from Dockery et al.[35]

More recently, Sharma et al. published a systematic review on the health effects associated with $PM_{2.5}$[118], made between 2015–2019. The review focuses on mortality and morbidity in turn.

The mortality section shows that exposure to $PM_{2.5}$ in China is consistently linked to deaths of various medical classifications including lung cancer, cardiovascular and respiratory deaths. A 10 $\mu g/m^3$ increase in 2-day moving average $PM_{2.5}$ concentration on total mortality corresponded to a 0.17% (95% CI 0.10%-0.23%) increase at a national level in 160 Chinese communities between 2013 and 2014[76]. If $PM_{2.5}$ concentrations in China had met the WHO interim target in 2013, the avoidable excess deaths in 2013 in China have been modelled to have been between 279,000 and 898,000[134].

The morbidity assessment in the Sharma et al. review showed that short-term exposure to $PM_{2.5}$ in China is consistently linked to respiratory disease, derived by analysis of hospital admission figures. In Beijing, a 10 $\mu g/m^3$ increase in $PM_{2.5}$ concentration was associated with a 0.82% (95% CI: 0.38%–1.26%) increase in COPD-related hospital admission during the period of 2013-2017[76].

Outside of China, emerging stand-alone studies conducted in the United States are revealing the diverse health effects of $PM_{2.5}$. Exposure to $PM_{2.5}$ has recently been linked to neuropsychiatric disorders, such as anxiety[109], and has also been suspected to influence structural brain development in childhood[30].

Nitrogen Oxides

Vehicle emissions are a major outdoor source of Nitric Oxide (NO) and Nitrogen Dioxide (NO_2). High-temperature fuel combustion results in N_2 in the air oxidising to NO, which further oxidises to NO_2. NO_x is a description of the sum of NO and NO_2. NO is not considered to be hazardous to health at typical ambient conditions, however, it plays a key role in the formation of tropospheric ozone and so is of interest with regard to health.

In the lungs NO_2 reacts with water forming both nitric (HNO_3) and nitrous (HNO_2) acids, which damage the lung cells[55]. NO_2 is understood to detrimentally affect people's respiratory function globally[161;130;17] and within China[16;44]. He et al.[53] linked NO_2 exposure to respiratory deaths from data from across China from 2013-2015. Per 10 $\mu g/m^3$ increase in NO_2, they estimated a 0.57% (95% CI: -0.04%-1.18%) increase in non-accidental mortality for the previous day NO_2.

However, the cardiovascular effects of ambient NO_2 exposure have been found to be both insignificant and sensitive to modelling choices[53;157].

Ozone

In the troposphere, ozone (O_3) is a secondary pollutant, produced by the reaction between NO_x and VOCs in the presence of solar radiation. Outdoor sources of VOCs include vehicle emissions and industrial activities. Natural vegetation emissions of certain VOCs (e.g. isoprene) also contribute to O_3 formation, especially on the regional scale[39;14;15].

O_3 is a powerful oxidant. When inhaled, it oxidises the first layer of cells in the airway surface, including the airway epithelial cells[a]. This releases reactive oxygen species which cause further oxidative damage to the airway[8;33].

Exposure to ambient O_3 in China has been shown to have respiratory effects. It decreased lung function[b] (an IQR increase in 5-day moving average of O_3 was associated with a 3.7% (95% CI: -7.1%- -0.2%) decrease in FEV_1) and increased airway inflammation[c] (an IQR increase in 5-day moving average of O_3 was associated with a 25.3% (95% CI: 3.6% - 51.6%) increase in FeNO) among healthy young adults[23]. Wang et al. estimated 186,000 (95% CI: 129,000-237,000) respiratory deaths attributable to O_3 exposure in China on a 5-year average[145].

Additionally, cardiovascular health effects have been associated with ambient O_3 exposure in China for example increases in blood pressure levels of middle-aged and older adults[97]. Wang et al. attributed 125,000 (95% CI: 42,000-204,000) cardiovascular deaths in China to O_3 exposure on a 5-year average basis[145].

Carbon Monoxide

Carbon monoxide (CO) is a product of incomplete combustion. The greatest outdoor sources of CO to outdoor air are cars, lorries, buses and other vehicles and machinery that burn fossil fuels.

CO binds strongly to haemoglobin, modifying its conformation and reducing its capacity to transfer oxygen, affecting the function of different organs which consume high levels of oxygen especially the heart[5]. A nationwide time-series analysis of cities in China from 2013 to 2015 found robust evidence linking short-term exposure to ambient CO and increased cardiovascular disease mortality, especially from coronary heart disease. This study is currently the largest study done in a low- or middle-income country (LMIC)[79].

[a]The airway epithelial plays a critical role in maintaining the conduit for air and plays a key role in the removal and neutralisation of potential harmful substances in inhaled air[2]

[b]Lung function was measured as the forced expiratory volume in a second (FEV_1)

[c]Airway inflammation was measured as the fractional exhaled nitric oxide (FeNO)

A.2 Chapter 2 supplementary information

A.2.1 Advancements in understanding of ventilation as a result of the COVID-19 pandemic

At the start of the pandemic, it was assumed that the SARS-CoV-2 virus was principally spread via contaminated surfaces. This assumption seemed validated by research, published in March 2020, which reported that coronavirus SARS-CoV-2 can remain viable and infectious on surfaces for days[138]. However, Goldman accused studies of using virus concentrations which exceeded those typically found in droplets[50]. Other researchers were making similar conclusions to Goldman and the scientific understanding about the virus transmission changed. It is now understood that contact and airborne transmission is possible, however, the respiratory (droplet) route is likely to be the principal method of transmission; COVID-19 aerosol droplets from speaking can remain suspended in stagnant air for up to 9 hours[34]. As demonstrated mostly for influenza viruses, environmental factors that may affect airborne virus survival include ventilation, temperature, humidity, pH and ultraviolet radiation[104 149].

Ventilation of indoor spaces has therefore become an important research topic which, coupled with shorter peer review times for papers related to coronavirus[37], has resulted in many recent advancements and publications in this area. This includes the expansion of tracer gas methods from buildings to public transport. A "proof of concept" study showed that these methods can be applied to a train carriage, using CO_2 as the tracer gas and as a proxy for exhaled breath[153]. This paper can be found in Appendix A.1.1.

A.2.2 Deriving the Analytical Solution of the Continuity Equation

$$\frac{dI_t}{dt} = (O_t - I_t)k_{\text{vent}} - I_t k_{\text{sink}} + F_t$$
$$\frac{dI_t}{dt} = O_t k_{\text{vent}} - I_t(k_{\text{vent}} + k_{\text{sink}}) + F_t$$

As the AIRLESS data was recorded at discrete sampling intervals (1 minute), for each interval, O_t and F_t are assumed to be constant. They become O and F.

Therefore $P = Ok_{\text{vent}} + F$, where P is a constant,

$$\frac{dI_t}{dt} = P - I_t(k_{\text{vent}} + k_{\text{sink}})$$

To solve this first order differential equation the integrating factor $= e^{\int(k_{\text{vent}}+k_{\text{sink}})dt}$ is used giving:

$$\begin{aligned}
e^{(k_{\text{vent}}+k_{\text{sink}})t}\frac{dI_t}{dt} &= Pe^{(k_{\text{vent}}+k_{\text{sink}})t} - I_t(k_{\text{vent}} + k_{\text{sink}})e^{(k_{\text{vent}}+k_{\text{sink}})t} \\
e^{(k_{\text{vent}}+k_{\text{sink}})t}\frac{dI_t}{dt} + I_t(k_{\text{vent}} + k_{\text{sink}})e^{(k_{\text{vent}}+k_{\text{sink}})t} &= Pe^{(k_{\text{vent}}+k_{\text{sink}})t}
\end{aligned}$$

Integration step

$$\begin{aligned}
e^{(k_{\text{vent}}+k_{\text{sink}})t}I_t &= P\int e^{(k_{\text{vent}}+k_{\text{sink}})t} + c \\
e^{(k_{\text{vent}}+k_{\text{sink}})t}I_t &= \frac{P}{(k_{\text{vent}} + k_{\text{sink}})}e^{(k_{\text{vent}}+k_{\text{sink}})t} + c \\
I_t &= \frac{P}{(k_{\text{vent}} + k_{\text{sink}})} + ce^{-(k_{\text{vent}}+k_{\text{sink}})t}
\end{aligned}$$

At t=0, $I_0 = \frac{P}{k_{\text{vent}}+k_{\text{sink}}} + c$, which rearranges to $c = I_0 - \frac{P}{k_{\text{vent}}+k_{\text{sink}}}$ and therefore:

$$I_t = \frac{P}{(k_{\text{vent}} + k_{\text{sink}})} + (I_0 - \frac{P}{(k_{\text{vent}} + k_{\text{sink}})})e^{-(k_{\text{vent}}+k_{\text{sink}})t}$$

Substituting in P gives the analytical solution.

A.2.3 Proof that the lag between the outdoor pollution and the indoor pollution is approximated as the reciprocal of the ventilation rate

A simple simulated case will be used. It will be assumed that outdoor air pollution levels infiltrate indoors at a constant rate, where $O_t = t$.

The solution to the continuity equation for this case:

$$I_t = O_t + (I_{t-1} - O_t)e^{-k_{\text{vent}}} \tag{A.1}$$

The target expression, with the indoor pollutant concentration being identical to the outdoor concentration but with a time lag (λ), can be written as:

$$I_t = O_t + \lambda \tag{A.2}$$

If we consider two adjacent time points, these expressions can be combined, to give:

$$O_t + \lambda = O_t + (O_{t-1} + \lambda - O_t)e^{-k_{\text{vent}}} \tag{A.3}$$

Which can be simplified to:

$$\lambda = (\lambda - 1)e^{-k_{\text{vent}}} \tag{A.4}$$

and then rearranged to:

$$\lambda = \frac{-1}{e^{k_{\text{vent}}} - 1} \tag{A.5}$$

e^k can be approximated, for small values of k_{vent}, using the Taylor series expansion:

$$e^x = 1 + x - \frac{x^2}{2} + \frac{x^3}{3} - ... \simeq 1 + x \tag{A.6}$$

And therefore, the lag (λ) can be approximated as $\frac{-1}{k_{\text{vent}}}$.

To show this simulated case, a graph has been produced, where the starting value of the indoor concentration is 150, and the value of k_{vent} is 0.1 hr $^{-1}$. The lag can be seen to settle at around a value of 10.

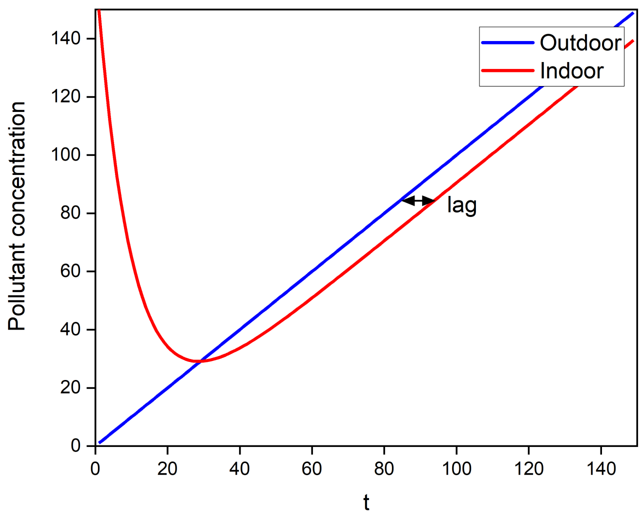

Figure A.4: Time series of simulated outdoor and indoor pollutant concentrations to show that the lag between these time series can be approximated as $\frac{1}{k_{\text{vent}}}$

A.3 Chapter 3 supplementary information

A.3.1 Specific sensors in the PAM and reference instrument

Instrument	CO Sensor	NO Sensor	NO_2 Sensor	O_3 Sensor	$PM_{2.5}$ Sensor
Reference Instrument	Nondispersive infrared, Thermo Fisher Scientific model 48i	Chemiluminescence, Thermo Fisher Scientific model 42i	Chemiluminescence, Thermo Fisher Scientific model 42i	UV absorption, Thermo Fisher Scientific model 49i	TEOM
PAM	Alphasense Ltd CO-A4	Alphasense Ltd NO-A4	Alphasense Ltd NO2-A43F	Alphasense Ltd Ox-A431	Alphasense OPC-N2

Table A.1: A table of the specific sensors in the PAM and reference instrument

A.3.2 Workings of EC senors and the OPC

Electrochemical Sensors measure gaseous pollutant concentrations. In most conventional electrochemical sensors, the are 3 electrodes: working, reference and counter electrodes. These sensors work using the amperometric principle of operation involving a working electrode made of a material that reacts with the target species in the sample. When a voltage is applied across the working electrode, a current flows between the electrode and the sample, which is proportional to the concentration of the target species. The current is then measured and used to determine the concentration of the species. The reference electrode is made of a stable material with known electrochemical potential that is used as a reference point for measuring the potential of the working electrode. The counter electrode does not participate in the sensing reaction, but facilitates the transfer of electrons to or from the working electrode; it completes the circuit. The electrochemical sensors in the PAM include a 4th electrode, an auxiliary electrode, which only measures temperature. It is used to correct for the temperature dependence of the cell potentials during post-processing[105].

Optical Particle Counters (OPC) measure Particulate Matter (PM) concentrations. The OPC used in the PAM is a miniaturised, commercially available particle counter (Alphasense OPC-N2). It works by first drawing in a sample of the air using a fan. The sample arrives into an optical volume, where it is illuminated by a laser, causing the particles to scatter light. The scattered light is detected by a photo-detector, which is used to infer the size. Particles are binned by diameter, and an algorithm is applied to calculate the PM_1, $PM_{2.5}$, and PM_{10} fractions. This work only focuses on the $PM_{2.5}$ fraction as the PM_1 measurements are less reliable (particles with diameter $< 0.38\mu$m are not able to be detected using this OPC) and the fraction between PM_{10} and $PM_{2.5}$ are expected to be less relevant for health outcomes ($PM_{2.5}$ penetrates deeper into the lungs[96]). Additionally, $PM_{2.5}$ is traditionally the fraction of focus in pollution health studies, allowing for comparisons with previous work.

A.3.3 Splitting “work” into “work indoors” and “work outdoors”

A different method was applied for the two seasons. For the winter season, the difference in temperature recorded by the PAM and reference instrument was used to determine whether the participant was indoors or outdoors, with a temperature of at least 5°C higher than the reference instrument indicating that the participant was indoors. The thermal conditions measured during the deployments can be found in Appendix A.3.4. For the summer campaigns, differences in reference and PAM O_3 levels were used, with an O_3 concentration of at least 5ppb lower than the outdoor levels indicating that the participant was in the indoor environment.

A.3.4 Thermal conditions during deployment

The outdoor ambient temperature during both campaigns was measured at the reference instrument monitoring sites. During the winter campaign, it ranged from -9°C to 16°C and during the summer, it ranged from 15°C to 41°C.

Temperature data collected by the PAM also aids when assessing the microenvironment of the participant, see Section 3.6. In the winter, the Beijing residents were exposed to an average temperature of 20°C and the Pinggu residents were exposed to an average temperature of 11°C. In the summer, both cohorts were exposed to an average temperature of 25°C.

The mean RH value recorded by the PAM 46% in the winter and 55% in the summer.

A.4 Chapter 4 supplementary information

A.4.1 Demonstrating Step 1v on a 12-hour period of data

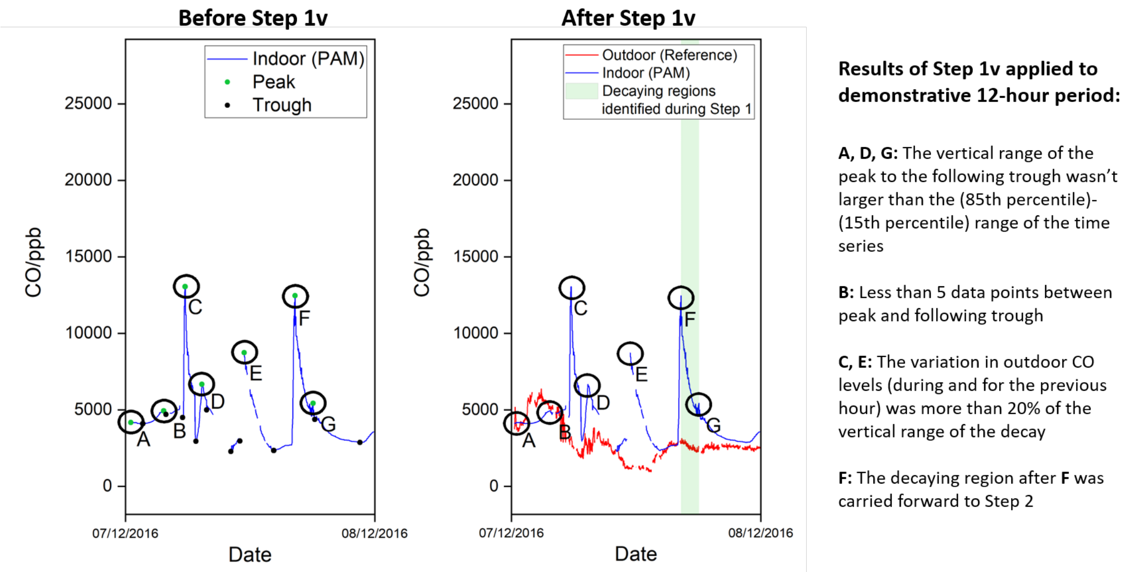

Figure A.5: Step 1v: The left-hand plot shows a 12-hour period of the CO data recorded by Participant U143. The 12-hour period selected is from 7th Dec-8th Dec 2016. The peaks and troughs are indicated. This plot is a subplot of Figure 4.3. Each peak has be labelled from A-G. The right-hand plot shows the decaying region identified during Step 1 (pale green region), and also includes CO data recorded by the Reference instrument during the same period (red). Only the decaying region after peak F passed Step 1v. On the right-hand side are the reasons why certain decaying regions following peaks passed Step 1v, and why others didn't.

A.4.2 Two methods of extracting coefficients of decays in data

The first method fits a curve to raw data, where it is assumed that the error in the data is a percentage error. This can be described as a multiplicative error, for example an error of 2% typically means 2/100 times the y value: larger values have larger associated absolute error. In contrast, the second method takes the natural logarithm of the raw data. The product rule can be used to show that the error now becomes an additive error (as opposed to a multiplicative error):

$$ln(y \times error) = ln(y) + ln(error) \tag{A.7}$$

As a result, the errors are now independent of the magnitude of y, and therefore the residuals for all the data points in the decay become uniform. In practice this results in less of a skewed weighting towards the smaller y values when applying the fit than in the first method.

In this work, the second method is used. This is because a weighted fit towards smaller y values is not desirable.

A.4.3 Inverse-variance weighting

Inverse-variance weighting is used to calculate the 12 hour mean. This method recognises that raw residuals for small groups are unreliable and therefore pulls them towards the overall average. The formula for this is shown below:

$$\text{Inverse-variance weighted mean} = \frac{\sum_{i=1}^{n} \left(\sigma_i^2\right)^{-1} x_i}{\sum_{i=1}^{n} \left(\sigma_i^2\right)^{-1}} \tag{A.8}$$

Where:

- σ_i is the variance ($= \sqrt{\frac{\sum(x_i - \hat{x})^2}{N}}$)
- x_i is a sequence of independent observations
- $\hat{x}$ is the ordinary mean of the observations ($= \frac{\sum x_i}{N}$)
- N is the number of observations

A.5 Chapter 5 supplementary information

A.5.1 Pollution concentrations in different microenvironments, separated by location and season

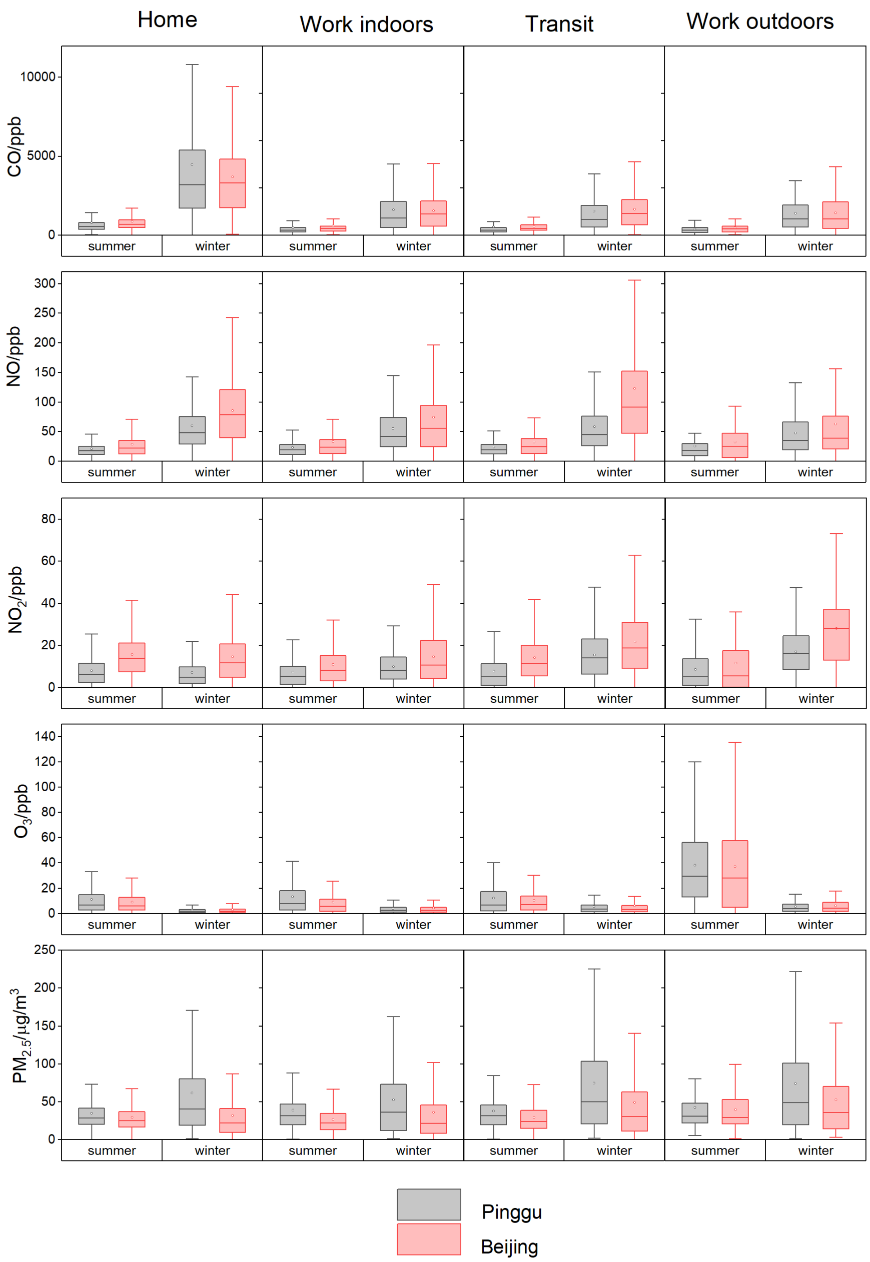

Figure A.6: Pollution concentrations in different microenvironments, separated by location and season: Box plots of averages over 12 hour periods for CO, NO, NO_2, O_3 and $PM_{2.5}$ exposure in four different microenvironments, calculated from personal exposure data recorded by the PAM, separated by location and season.

A.5.2 Outdoor air quality standards in China

National standards of ambient air quality were introduced in China in 1982 (GB 3095-82). Over the last 40 years, these standards have been revised and replaced, with a full second revision issued in 2012 (GB 3095-2012), shown in Table A.2.

Pollutant	Averaging time	Limit (Class 2) (μg/m^3)	Limit (Class 2) (ppb)
CO	annual	4,000	3430
	hourly	10,000	8590
NO_x	annual	50	26.2
	24 hours	100	52.4
	hourly	250	131
NO_2	annual	40	20.9
	24 hours	80	41.8
	hourly	200	105
O_3	daily (8 hour maximum)	160	80.2
	24 hours	200	100
$PM_{2.5}$	annual	35	-
	24 hours	75	-

Table A.2: Ambient air quality standards, China: A table of the current ambient air quality standards for CO, NO_x, NO_2, O_3 and $PM_{2.5}$ (GB 3095-2012). These are four of the key pollutants, although the standards extend to other pollutants. Only the standards for Class 2 are included. Class 1 is applied to special areas of environmental protection such as natural reserves and national parks. Class 2 includes all other areas, including residential areas and so it more applicable to health. The standards were published in units of μg/m^3. The gas data in this thesis are presented in ppb, and therefore the ambient standards have been converted into ppb, assuming 20°C and 1atm.

These current Chinese air quality standards are higher than the interim targets set by the World Health Organisation (WHO)[93 155].

The full revision was followed by China implementing its Air Pollution Prevention and Control Action Plan (APPCAP) from 2013 to 2017, which included tightening standards for industrial emissions, reducing coal consumption, and decreasing the number of heavily polluting vehicles on the roads. Some regions have a relatively high proportion of cities which are meeting the standards as a result. These regions are encouraged to aim for more stringent air quality targets to further improve local ambient air quality[147].

Still, the proportion of cities reaching the set standards remains low despite seeing a reduction in emissions as a result of the APPCAP, and O_3 and $PM_{2.5}$ have become key pollutants restricting the ambient air quality in China from reaching their standards[147].

A.5.3 Indoor air quality standards in China

In 2002, indoor air quality standards in China were published and they came into effect in 2003. They apply to residencies and office buildings. They are compared to the WHO indoor air quality standards in Table A.3.

Although these indoor air quality standards for China exist, there is no specific

	Indoor air quality standards, China			Indoor air quality standards, WHO		
Pollutant	Averaging time	Limit (Class 2) ($\mu g/m^3$)	Limit (Class 2) (ppb)	Averaging time	Limit ($\mu g/m^3$)	Limit (ppb)
CO	hourly	10,000	8590	hourly	35,000	30,100
NO_2	hourly	240	125	hourly	200	105
O_3	hourly	160	80.2	8-hour	100	50.1
PM_{10}	daily	150	-	daily	45	-

Table A.3: Indoor air quality standards, China: A table of the current indoor air quality standards for CO, NO_2, O_3 and PM_{10} (GB/T 18883-2002). These are four of the key pollutants. Standards have been set for other species and physical properties such as temperature and ventilation. The pollutant standards were published in units of $\mu g/m^3$. The gas data in this report are presented in ppb, and therefore the standards have been converted into ppb, assuming 20°C and 1 atm. The 2021 WHO indoor air quality standards are included for comparison.

benchmark or regulation about how indoor air quality should be evaluated and intervened for improvement[116].

A.5.4 I/O ratios in China

O_3 is reactive in the indoor environment. As a result, and due to limited sources in the indoor environment, the I/O ratio for O_3 is normally less than one[61]. O_3 I/O ratios varied between 0.21 – 1.00 in student dormitories in China, with the large range attributed to window opening conditions[159]. A study across residential buildings in Nanjing, China recorded O_3 I/O ratios O_3 in a room with no O_3 sources (no photocopying devices, air purifiers or kitchen disinfectant devices) of between 0.06 and 0.62 depending on the season and door opening conditions[61].

Although NO_2 also reacts indoors, its I/O ratio value can be higher than 1. Hu et al.[60] reviewed NO_2 I/O ratios from different countries around the world. Their results, for the home microenvironment are shown in Figure A.7. NO_2 ratio values were found to be higher in Pakistan, Egypt, and Bangladesh (median I/O values of 1.66, 1.54 and 1.54 respectively), particularly in rural kitchens during the winter, which was attributed to considerable indoor sources and low ventilation rates. 11 of the reviewed studies were conducted in homes in China. Over 750 I/O values were recorded in Chinese homes and the median was 0.98.

$PM_{2.5}$ I/O ratios measured in Chinese homes tend to be less than 1[169;142], suggesting that loss processes, such as deposition, dominate, however the strength of indoor $PM_{2.5}$ sources have been shown to significantly affect the I/O ratio: the range of $PM_{2.5}$ I/O ratio values from a study conducted in Chinese residential buildings were between 0.73–0.75 for a room with sources and between 0.41–0.46 for a room

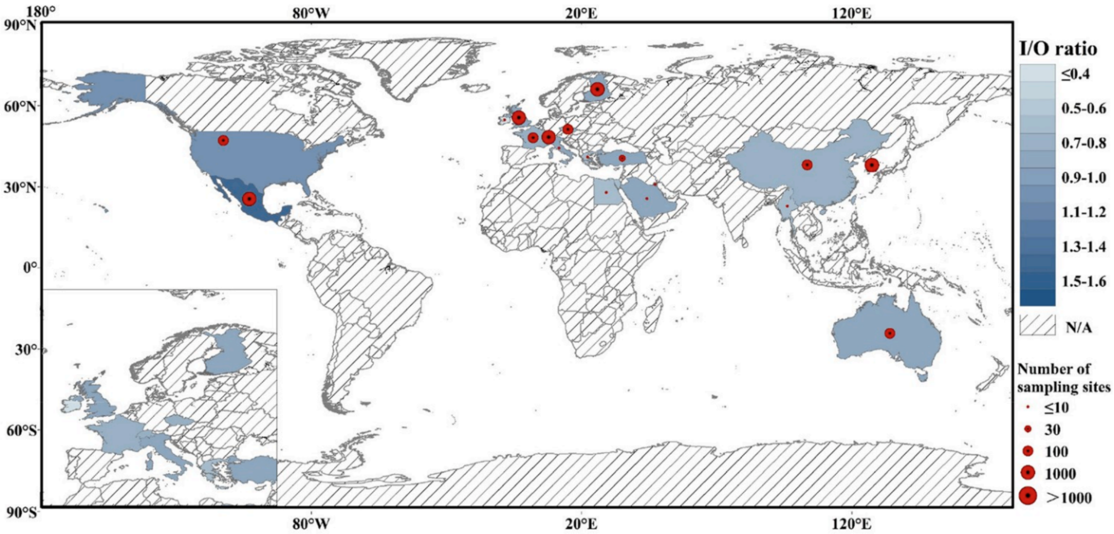

Figure A.7: Median NO_2 I/O ratio values in residential buildings in countries worldwide: A world map comparing the median I/O ratio values measured in studies performed worldwide. The median I/O values are larger than 1 for some countries, indicating dominant NO_2 sources in the home microenvironment. This figure has been taken from Hu et al.[60]

without[169]. Additionally, a study comparing four residential dwellings with different building envelope air tightness levels in China obtained a range of 0.167-0.867. The lowest I/O ratio was recorded for a home with high air-tightness and few indoor sources[142].

A.5.5 Exposure, by percentage, to the two components of exposure

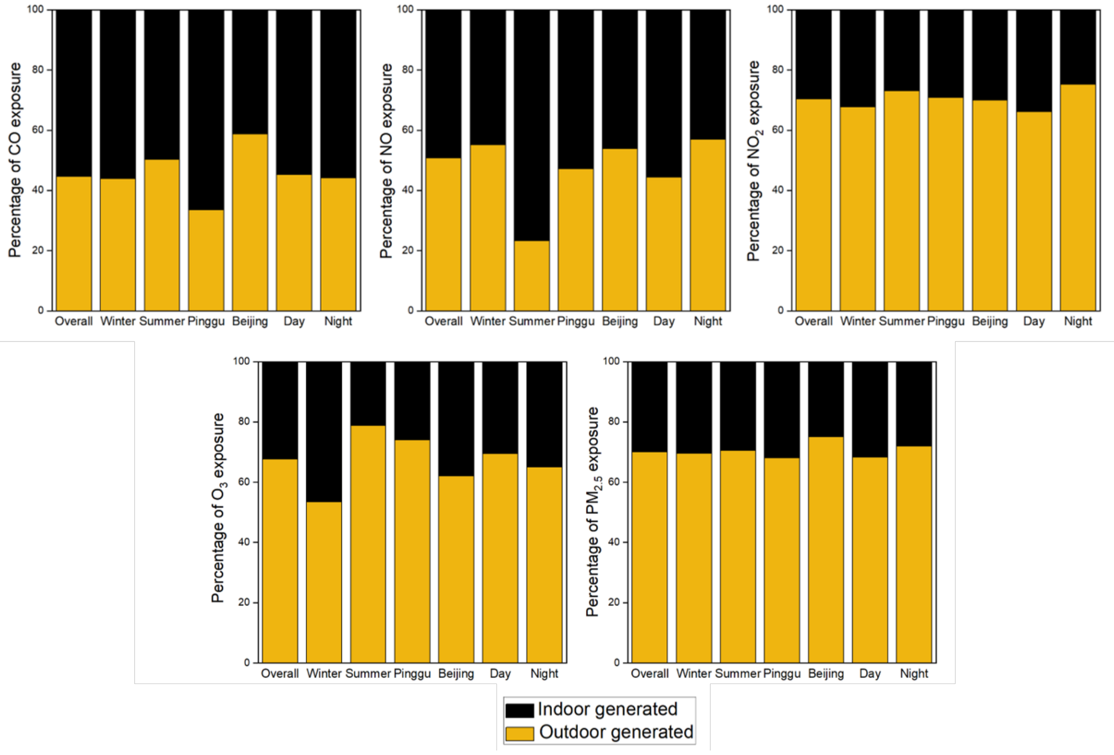

Figure A.8: Source-apportioned percentages of exposure: Stacked percentage column plots of the apportioned total personal exposure measured by the PAMs during the AIRLESS project. The percentages for specific seasons, locations and time of data are also included.

A.5.6 Exposure diurnal plots

The diurnal plots for the four exposure types (reference, personal, and the two components of personal exposure) for $PM_{2.5}$ is found in Chapter 5. The equivalent plots for the other key species are shown below.

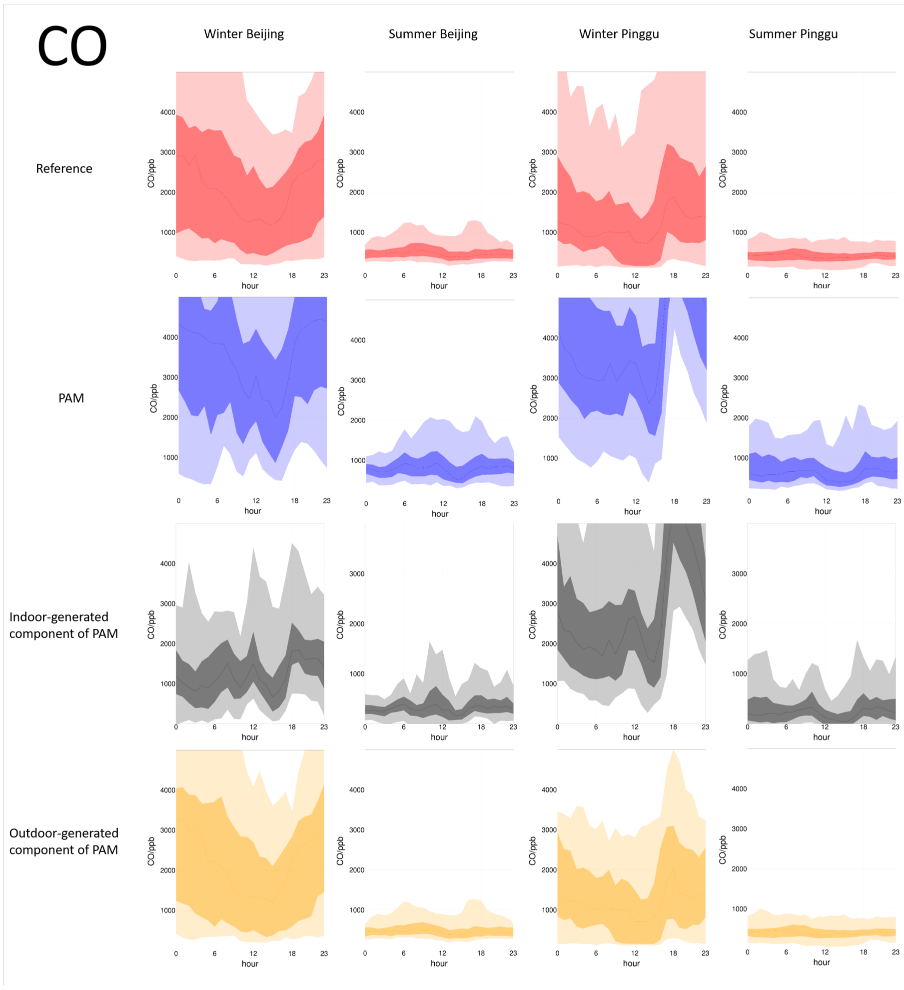

Figure A.9: Dirunal plots of source-apportioned CO Diurnal plots of the indoor-generated (black) and outdoor-generated (orange) portions of total exposure to CO, recorded by the PAMs, for the AIRLESS cohort. The diurnal plot of the outdoor levels recorded by the reference instrument (red) is included for comparison. The plots display the median, 25th and 75th percentiles and 5th and 95th percentiles. Negative indoor-generated values were removed before producing these plots.

Figure A.9 shows that for CO, the outdoor-generated component of indoor air has a very similar diurnal shape to that recorded by the reference instrument. The indoor-generated diurnal plot shows that indoor sources are stronger in the winter time. Additionally, across both seasons and locations, the indoor-generated component appears strongest around mealtimes, with a peak around or just before midday, and a peak in the evening around 6pm. Indoor-generated CO appears particularly strong for the Pinggu cohort during the winter.

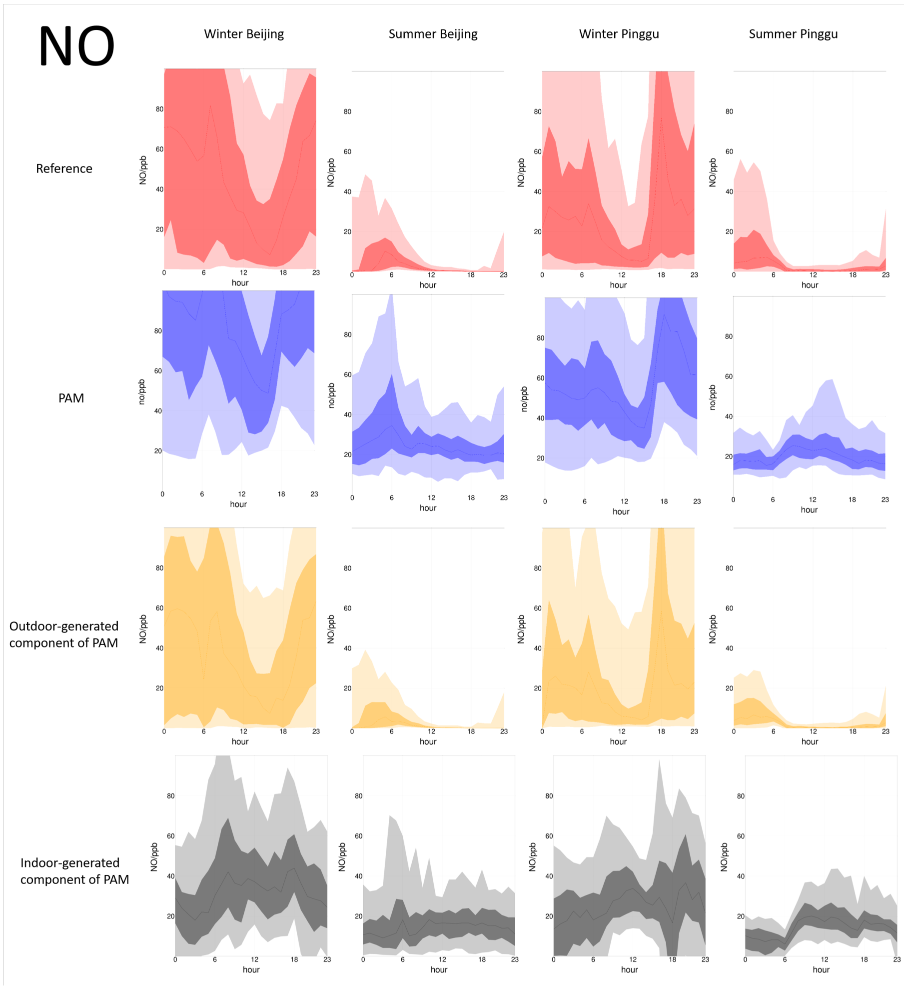

Figure A.10: Diurnal plots of source-apportioned NO Diurnal plots of the indoor-generated (black) and outdoor-generated (orange) portions of total exposure to NO, recorded by the PAMs, for the AIRLESS cohort. The diurnal plot of the outdoor levels recorded by the reference instrument (red) is included for comparison. The plots display the median, 25th and 75th percentiles and 5th and 90th percentiles. Negative indoor-generated values were removed before producing these plots.

As with CO, the outdoor-generated component of indoor air has a very similar diurnal shape to that recorded by the reference instrument. However for NO, the indoor-generated component is less defined around mealtimes. The outdoor-generated and outdoor (reference) components in the summer are stronger in the early hours of the morning. NO reacts rapidly with O_3, however, O_3 is depleted during the nighttime, allowing a build up of NO, until photolytic activity recommences at sunrise.

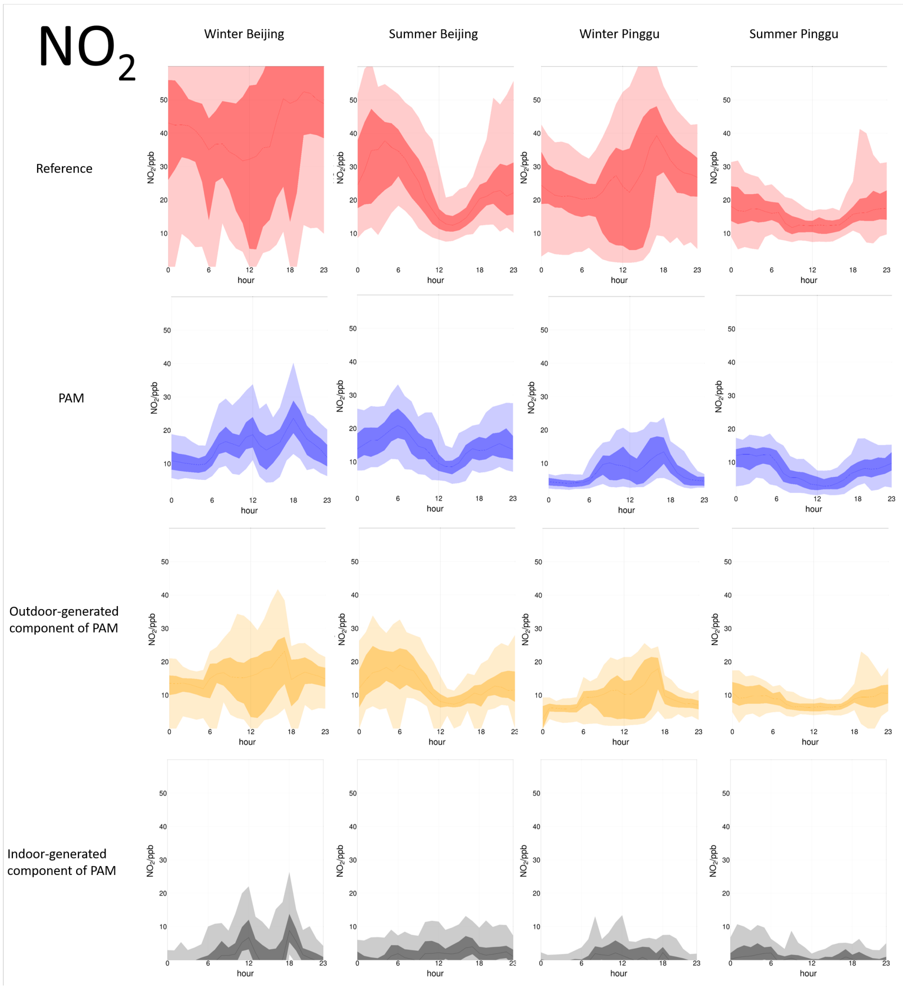

Figure A.11: Diurnal plots of source-apportioned NO_2 Diurnal plots of the indoor-generated (black) and outdoor-generated (orange) portions of total exposure to NO_2, recorded by the PAMs, for the AIRLESS cohort. The diurnal plot of the outdoor levels recorded by the reference instrument (red) is included for comparison. The plots display the median, 25th and 75th percentiles and 5th and 90th percentiles. Negative indoor-generated values were removed before producing these plots.

The outdoor-generated components of the previous two pollutants had very similar diurnal shape to that recorded by the reference instrument. Figure A.11 shows that NO_2 shares the general diurnal shape, however the outdoor-generated component has smaller magnitude. This is attributed to NO_2 being more reactive in the indoor environment. Additionally, strong indoor sources can be seen around mealtimes for the Winter Beijing cohort.

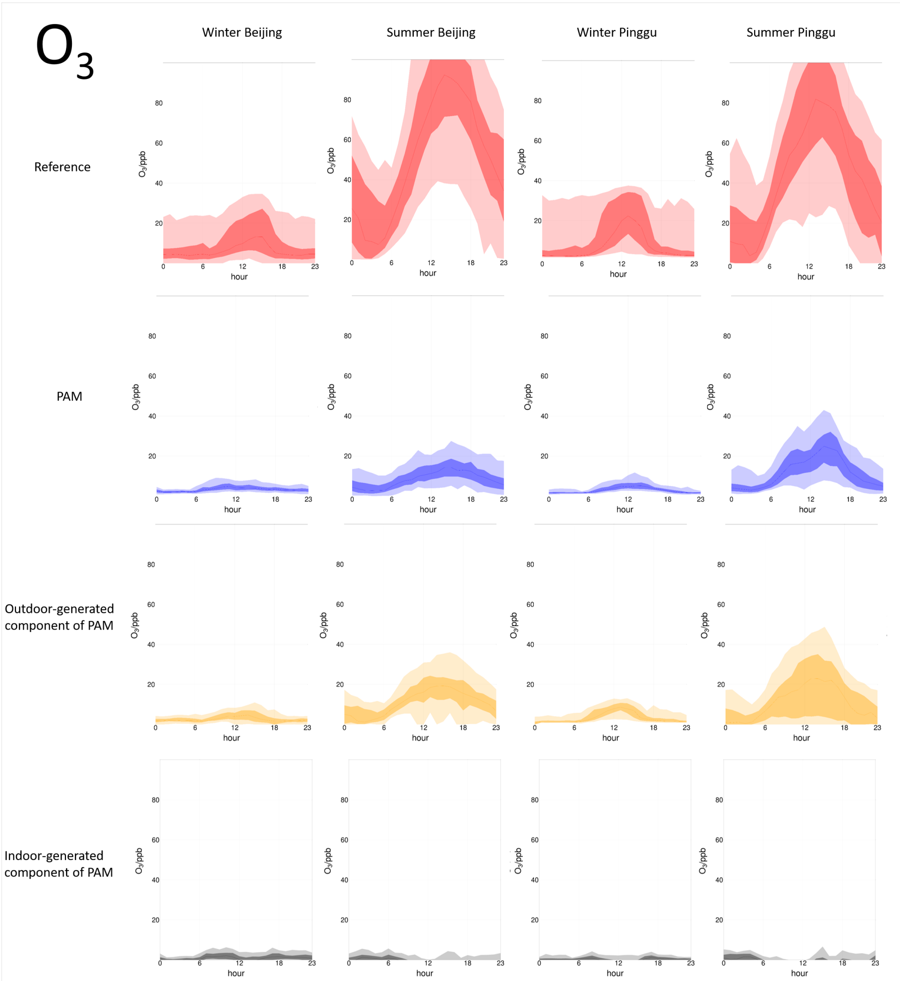

Figure A.12: Diurnal plots of source-apportioned O_3 Diurnal plots of the indoor-generated (black) and outdoor-generated (orange) portions of total exposure to O_3, recorded by the PAMs, for the AIRLESS cohort. The diurnal plot of the outdoor levels recorded by the reference instrument (red) is included for comparison. The plots display the median, 25th and 75th percentiles and 5th and 90th percentiles. Negative indoor-generated values were removed before producing these plots.

Figure A.12 shows very low exposure to indoor-generated O_3. The outdoor-generated follows the same diurnal shape as the outdoor reference data however is much lower, indicating that O_3 has a strong indoor sink.

A.5.7 Exposure box plots

Box plots for exposure to reference, personal and apportioned personal exposure can be found in Chapter 5 for all key species. Figure A.13 shows the exposures broken down by season, location and time of day.

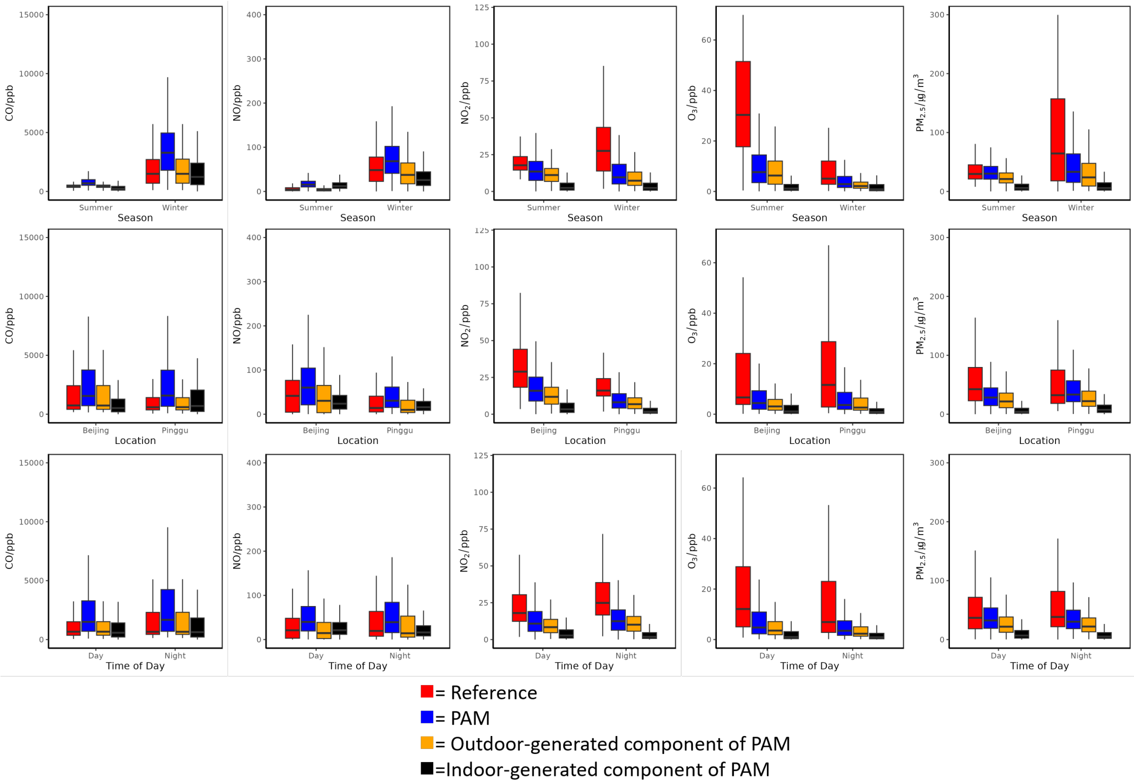

Figure A.13: Exposure box plots Reference (red), PAM (blue) exposure box plots, generated from 12-hour mean values. PAM exposure has been source-apportioned into outdoor-generated (orange) and indoor-generated (black) box plots. Box plots have been produced for both seasons, locations and times of day.

Season: For all pollutants except for O_3, higher exposures were recorded in the winter season. Generally across both seasons for all pollutants, participants were exposed to more outdoor-generated pollution than indoor-generated pollution (orange higher than black), except for NO. This may be due to the low levels of ambient NO during the summer. Ambient O_3 (measured by the reference instruments) is much higher in the summer than in the winter as expected, however, this difference is less pronounced in the other O_3 exposure metrics.

Location: The exposure metrics were generally across location, with NO and NO_2 exposure (by all exposure metrics) measuring slightly higher in Beijing than in Pinggu.

Time of day: Ambient O_3 (as measured by the reference instruments) is higher during the day than during the night, however, as with the summer winter breakdown, this difference is less pronounced in the other O_3 exposure metrics. Indoor- and outdoor-generated exposure for all species does not appear to change significantly between day and night. The selected day and night cut-off times do not coincide with participants being awake and asleep or with other participant behaviours; many participants may cook after 6pm for example. This may explain the absence of expected differences in the indoor-generated exposures between day and night.

A.5.8 Indoor emission event correlations

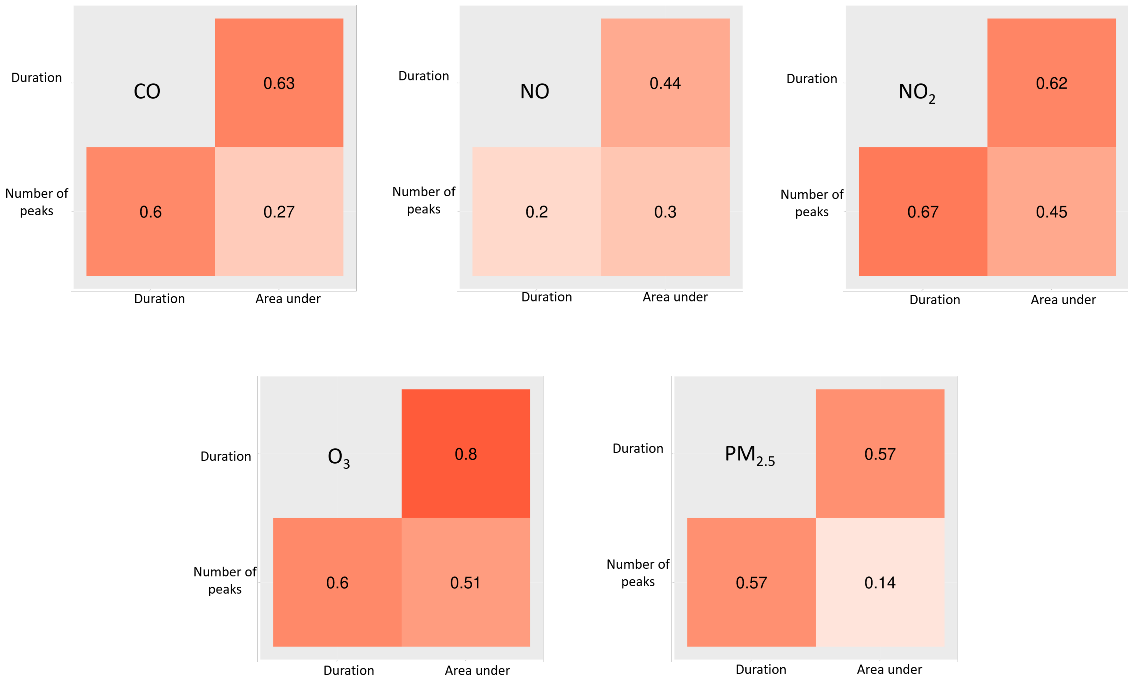

Figure A.14: Correlation matrix of the characteristics of indoor emission events Pearson correlation coefficients plots of the three characteristics (quantity and heights, duration and area under) of the indoor emission events, extracted for the whole AIRLESS dataset, for the five key pollutants.

A.5.9 A shift in primary cooking fuel

In China, the population has been shifting (and is projected to continue shifting) to clean cooking fuels in both urban and rural cohorts as shown in Figure A.15. The estimated trends for different global regions are shown in Figure A.16, showing that some other global regions are not expected to make such a dramatic transition.

The primary cooking fuel used by the AIRLESS cohort in 2016-2017 was shown in Chapter 3, Figure 3.3. The majority of urban participants used natural gas, where as LPG, biomass and biogas were the most common methods in the rural cohort. The primary fuel use of the AIRLESS urban and rural participants appear to be representative of urban and rural populations in China more generally.

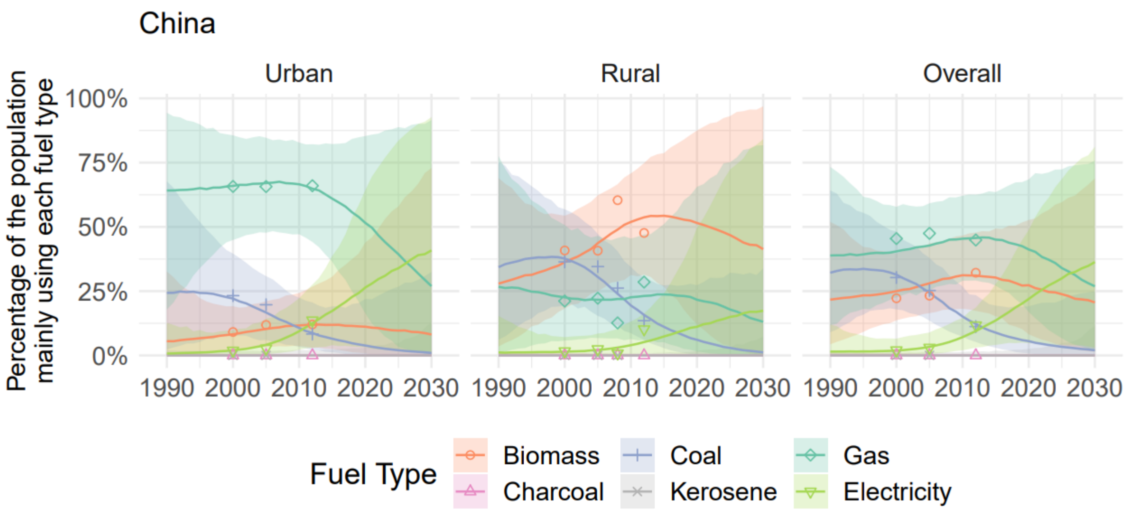

Figure A.15: Estimated (posterior median) percentage of the population using a polluting fuel as their primary cooking fuel in China, with 95% uncertainty intervals (shaded). Figure has been taken from Stoner et al.[128]

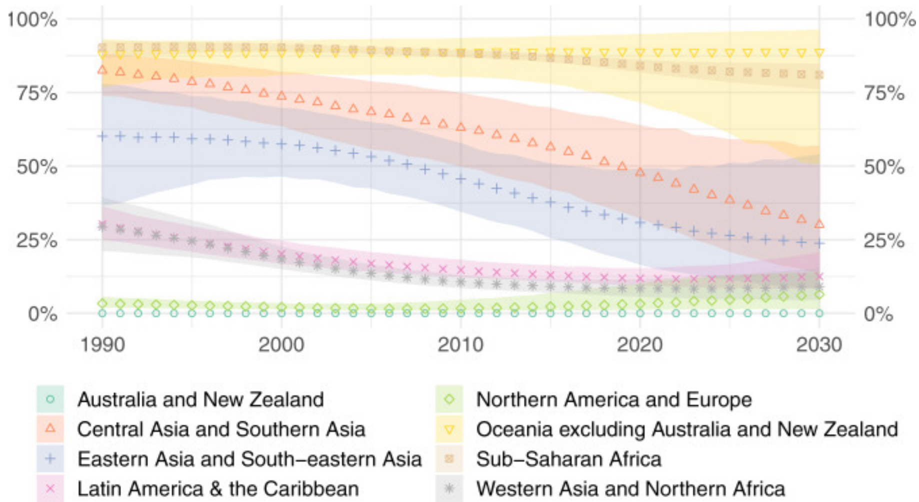

Figure A.16: Estimated (posterior median) percentage of the global population mainly cooking with polluting fuels in each region, with 95% uncertainty intervals (shaded). Figure has been taken from Stoner et al.[128]

A.5.10 Home ventilation and building characteristics

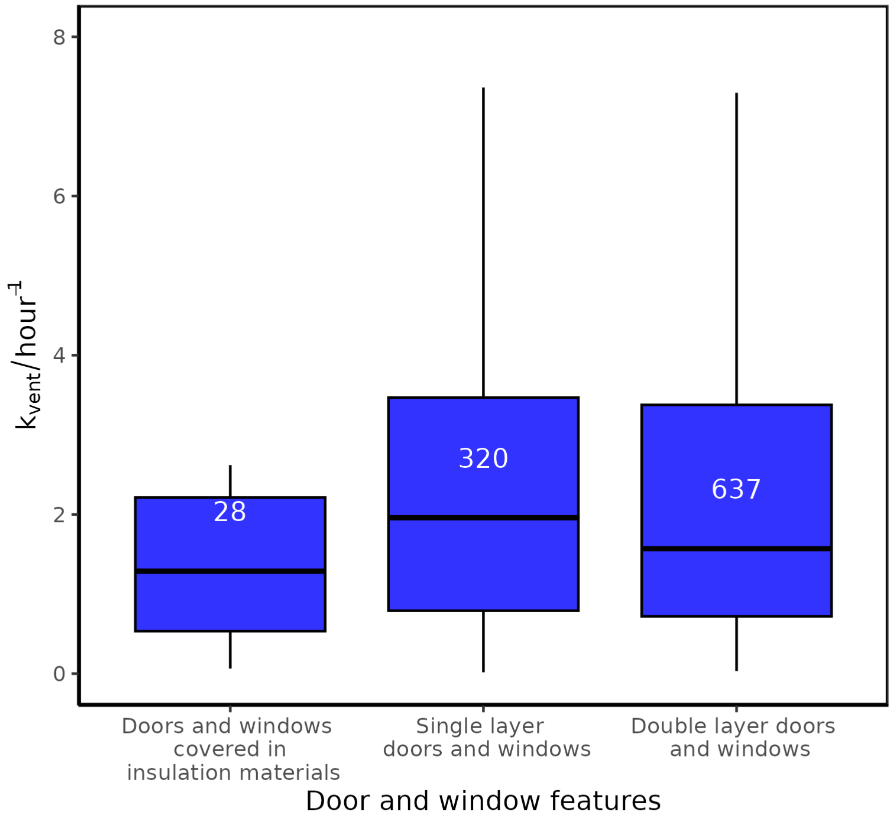

Figure A.17: Ventilation and window and door characteristics: Box plots of the ventilation rates estimated in participants' homes, separated by the window and door characteristics in the participants' homes.

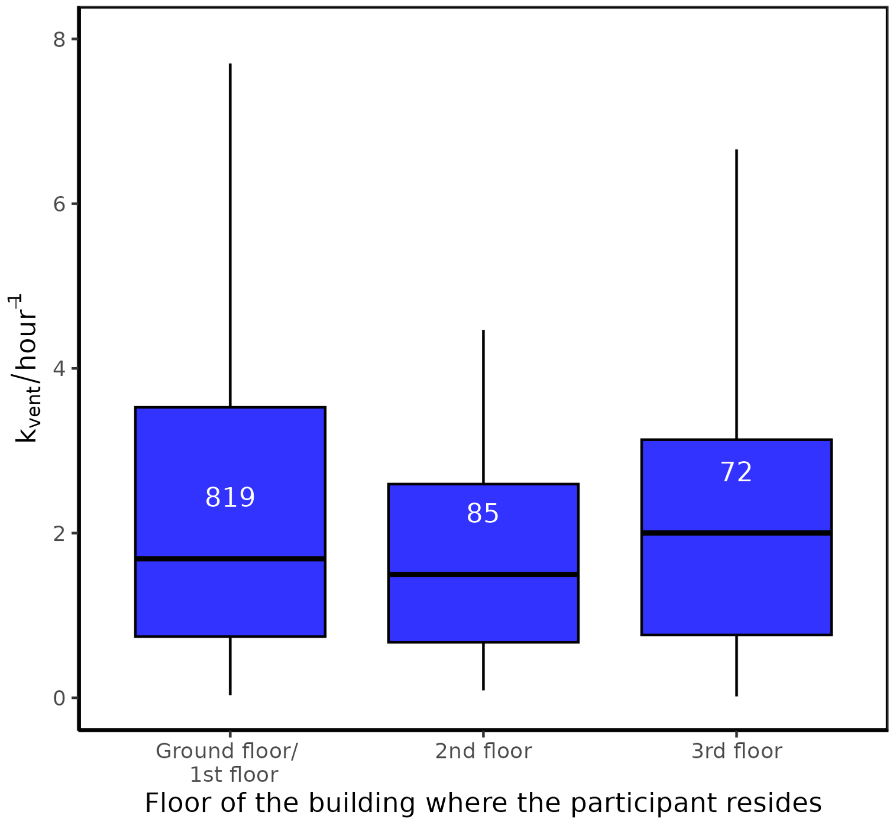

Figure A.18: Ventilation and floor level: Box plots of the ventilation rates estimated in participants' homes, separated by the floor of the building where the participant resides.