University of Potsdam
Institute of Environmental Science and Geography
Research Institute for Sustainability, Helmholtz Centre Potsdam

Using low-cost sensors to gather high resolution measurements of air quality in urban environments and inform mobility policy

by

Seán Schmitz

Cumulative Dissertation

for the degree of
"doctor rerum naturalium" (Dr. rer. nat.)
in Geoecology

submitted to the

Faculty of Science
Institute of Environmental Science and Geography
at the
University of Potsdam, Germany

Date of Disputation: 31.05.2023

Supervisors:
1st Supervisor: Prof. Dr. Mark Lawrence; RIFS, Helmholtz Centre Potsdam; University of Potsdam, Germany
2nd Supervisor: PD Dr. Ariane Walz; MLUK Brandenburg; University of Potsdam, Germany

Mentor:
Dr. Erika von Schneidemesser, RIFS, Helmholtz Centre Potsdam

Reviewers:
Prof. Dr. Mark Lawrence; RIFS, Helmholtz Centre Potsdam; University of Potsdam, Germany
Prof. Dr. Mira Pöhlker; TROPOS Leipzig; University of Leipzig, Germany
Prof. Dr. Otto Klemm; Universität Münster

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-60105
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-601053

Acknowledgements

Without the contributions of a great many people, this work would have never come to fruition. While the success of this work will be reflected only in my name as a certificate and a title, I stand on the shoulders of many who helped me achieve this goal.

First and foremost, I would like to thank my family for supporting me throughout the years and helping me get to this stage; without all of you crazy Schmitzles and Morrisseys, I wouldn't be who I am. Of course, this extends to my amazing Whirlpool-WG roommates, Jochen, Juliette, and Lou, who make every day brighter and full of life.

I would also like to thank Dr. Mark Lawrence and Dr. Ariane Walz, as well as all the members of my PhD Advisory Committee for supervising this thesis. Your support was crucial for allowing me the space to develop this work. Thank you both for giving me this opportunity. I am also deeply grateful to you Mark for supporting not only this work, but also my continued research and engagement with the field of sustainability at RIFS.

Speaking of, RIFS was and has been an incredible work environment, which fostered the success of this work. Most specifically, it's been my colleagues at RIFS that have made working there so enjoyable. Whether it be during chats over coffee or lunch, in-depth discussions on the global challenges facing humanity, targeted debates on our research, or spontaneous meetings in the hallways, it was always a with a helpful sense of collegiality and in a positive atmosphere. With that I want to especially thank all my RIFS colleagues; I will forever cherish my time at RIFS, as without the environment we fostered together, I would never have been able to achieve this work.

I'd also like to thank the PhD cohort at RIFS for all the great moments we shared together and for your support along the way. Together Docs strong!

A huge thanks also to Lea, a truly special Person, for her amazing support in helping to format this thesis. You are the absolute best!

All my love and thanks go out to my girlfriend Lou, without whom I would definitely have gone stark raving mad while working on this thesis from home during the pandemic. I love you with all my heart and am so glad you were there for me throughout when I needed you most.

I would be remiss to not especially thank my colleague Dr. Alex Caseiro, who joined our team just at the start of this project. Without your support, Alex, much of this work would have not been possible; your problem-solving skills, brilliant knowledge of atmospheric chemistry, and kind heart are an important foundation of this work. Thank you for everything.

Last, but by no means least, this work would have never been achieved, nor would it have even begun, were it not for the support and trust placed in me by Dr. Erika von Schneidemesser. From the second I arrived at RIFS as an intern during my master's program, she recognized my potential and fought to keep me in her research group. Without her confidence in me, I likely never would have even embarked upon this journey. Beyond that, regardless of the importance of my questions, or how often they came, she was always there to answer them. It is with her example that I hope to continue as a researcher in the years to come. While a PhD candidate must do the majority of the work for their thesis themselves, they cannot do it all alone; group support, especially from supervisors, is key. Erika fostered the best possible environment for me to thrive, and I am convinced the work would not be of this quality without these positive surroundings. For your support I will be forever grateful. Cheers to a few more good years together Erika!

Table of Contents

Abbreviations

Abbreviation	Full meaning
AIC	Akaike Information Criterion
ANN	Artificial neural networks
AQG	Air quality guideline
AR	Aspect ratio
AR5/6	Fifth/Sixth Assessment Report
BLH	Boundary layer height
BLUME	Berlin air quality measurement network
BMBF	Federal Ministry of Education and Research of Germany
CAMS	Copernicus Atmospheric Monitoring Service
CFD	Computational fluid dynamics
CH_4	Methane
CO	Carbon monoxide
DWD	German weather service
EC	Electrochemical cells
EEA	European Environment Agency
ERF	Effective radiative forcing
EU	European Union
GAMs	Generalized additive models
GHG	Greenhouse gas
GPS	Global positioning system
GSM	Global system for mobile communications
HDV	Heavy-duty vehicle
HNO_3	Nitric acid
HO_2	Hydroperoxyl radical
IPCC	International Panel on Climate Change
KD	Kottbusser Damm
LCS	Low-cost sensors
LDV	Light-duty vehicle
LEZ	Low emission zones
LR	Linear regression
MAE	Mean average error
ML	Machine-learning
MLH	Mixing layer height
MLR	Multiple linear regression
MOS	Metal oxide semiconductors
MWFK	Ministry for Science, Research and Culture of the Federal State of Brandenburg
MWU	Mann-Whitney-Wilcoxon U-test
N_2O_5	Dinitrogen pentoxide
NMVOCs	Non-methane volatile organic compounds

NO	Nitric oxide
NO_2	Nitrogen dioxide
NO_3	Nitrate
NO_x	Nitrogen oxides
O_2	Oxygen
O_3	Ozone
OH	Hydroxyl radical
$PM_{2.5}$ / PM_{10}	Particulate matter 2.5 and 10 microns in diameter
ppb	Parts-per-billion
ppm	Parts-per-million
R^2	Coefficient of determination
RF	Random forest
RH	Relative humidity
RIFS	Research Institute for Sustainability, Helmholtz Centre Potsdam
RMSE	Root mean squared error
SC	Summer campaign
SenUMVK	Berlin Senate Department for the Environment, Urban Mobility, Consumer Protection and Climate Action
SO_2	Sulphur dioxide
SUMP	Sustainable urban mobility plan
SVR	Support vector regression
T	Temperature
UB	Urban background
VI	Variable importance
VOCs	Volatile organic compounds
WC	Winter campaign
WD	Wind direction
WHO	World Health Organization
WS	Wind speed

Executive Summary

Air pollution has been a persistent global problem in the past several hundred years. While some industrialized nations have shown improvements in their air quality through stricter regulation, others have experienced declines as they rapidly industrialize. The WHO's 2021 update of their recommended air pollution limit values reflects the substantial impacts on human health of pollutants such as NO_2 and O_3, as recent epidemiological evidence suggests substantial long-term health impacts of air pollution even at low concentrations. Alongside developments in our understanding of air pollution's health impacts, the new technology of low-cost sensors (LCS) has been taken up by both academia and industry as a new method for measuring air pollution. Due primarily to their lower cost and smaller size, they can be used in a variety of different applications, including in the development of higher resolution measurement networks, in source identification, and in measurements of air pollution exposure. While significant efforts have been made to accurately calibrate LCS with reference instrumentation and various statistical models, accuracy and precision remain limited by variable sensor sensitivity. Furthermore, standard procedures for calibration still do not exist and most proprietary calibration algorithms are black-box, inaccessible to the public. This work seeks to expand the knowledge base on LCS in several different ways: 1) by developing an open-source calibration methodology; 2) by deploying LCS at high spatial resolution in urban environments to test their capability in measuring microscale changes in urban air pollution; 3) by connecting LCS deployments with the implementation of local mobility policies to provide policy advice on resultant changes in air quality.

In a first step, it was found that LCS can be consistently calibrated with good performance against reference instrumentation using seven general steps: 1) assessing raw data distribution, 2) cleaning data, 3) flagging data, 4) model selection and tuning, 5) model validation, 6) exporting final predictions, and 7) calculating associated uncertainty. By emphasizing the need for consistent reporting of details at each step, most crucially on model selection, validation, and performance, this work pushed forward with the effort towards standardization of calibration methodologies. In addition, with the open-source publication of code and data for the seven-step methodology, advances were made towards reforming the largely black-box nature of LCS calibrations.

With a transparent and reliable calibration methodology established, LCS were then deployed in various street canyons between 2017 and 2020. Using two types of LCS, metal oxide (MOS) and electrochemical (EC), their performance in capturing expected patterns of urban NO_2 and O_3 pollution was evaluated. Results showed that calibrated concentrations from MOS and EC sensors matched general diurnal patterns in NO_2 and O_3 pollution measured using reference instruments. While MOS proved to be unreliable for discerning differences among measured locations within the urban environment, the

concentrations measured with calibrated EC sensors matched expectations from modelling studies on NO_2 and O_3 pollution distribution in street canyons. As such, it was concluded that LCS are appropriate for measuring urban air quality, including for assisting urban-scale air pollution model development, and can reveal new insights into air pollution in urban environments.

To achieve the last goal of this work, two measurement campaigns were conducted in connection with the implementation of three mobility policies in Berlin. The first involved the construction of a pop-up bike lane on Kottbusser Damm in response to the COVID-19 pandemic, the second surrounded the temporary implementation of a community space on Böckhstrasse, and the last was focused on the closure of a portion of Friedrichstrasse to all motorized traffic. In all cases, measurements of NO_2 were collected before and after the measure was implemented to assess changes in air quality resultant from these policies. Results from the Kottbusser Damm experiment showed that the bike-lane reduced NO_2 concentrations that cyclists were exposed to by 22 ± 19%. On Friedrichstrasse, the street closure reduced NO_2 concentrations to the level of the urban background without worsening the air quality on side streets. These valuable results were communicated swiftly to partners in the city administration responsible for evaluating the policies' success and future, highlighting the ability of LCS to provide policy-relevant results.

As a new technology, much is still to be learned about LCS and their value to academic research in the atmospheric sciences. Nevertheless, this work has advanced the state of the art in several ways. First, it contributed a novel open-source calibration methodology that can be used by a LCS end-users for various air pollutants. Second, it strengthened the evidence base on the reliability of LCS for measuring urban air quality, finding through novel deployments in street canyons that LCS can be used at high spatial resolution to understand microscale air pollution dynamics. Last, it is the first of its kind to connect LCS measurements directly with mobility policies to understand their influences on local air quality, resulting in policy-relevant findings valuable for decisionmakers. It serves as an example of the potential for LCS to expand our understanding of air pollution at various scales, as well as their ability to serve as valuable tools in transdisciplinary research.

Zusammenfassung

Luftverschmutzung ist seit hundert Jahren ein anhaltendes globales Problem. Während sich die Luftqualität in einigen Industrieländern durch strengere Vorschriften verbessert hat, hat sie sich in anderen Ländern im Zuge der schnell fortschreitenden Industrialisierung verschlechtert. Die Aktualisierung der von der WHO für das Jahr 2021 empfohlenen Grenzwerte für die Luftverschmutzung spiegelt die erheblichen Auswirkungen von Schadstoffen wie Stickstoffdioxid (NO_2) und Ozon (O_3) auf die menschliche Gesundheit wider, da neuere epidemiologische Erkenntnisse darauf hindeuten, dass Luftverschmutzung selbst bei niedrigen Konzentrationen erhebliche langfristige gesundheitliche Auswirkungen hat. Parallel zu den Entwicklungen in unserem Verständnis der gesundheitlichen Auswirkungen von Luftverschmutzung wurde die neue Technologie der Low-Cost-Sensoren (LCS) sowohl von der Wissenschaft als auch von der Industrie als neue Methode zur Messung der Luftverschmutzung aufgegriffen. Vor allem aufgrund ihrer geringeren Kosten und kleineren Größe können sie in einer Vielzahl von Anwendungen eingesetzt werden, u. a. bei der Entwicklung von Messnetzen mit höherer räumlicher Auflösung, bei der Identifizierung von Quellen und bei der Messung der Luftverschmutzung. Es wurden zwar erhebliche Anstrengungen unternommen, um LCS mit Hilfe von Referenzinstrumenten und verschiedenen statistischen Modellen genau zu kalibrieren, aber die Genauigkeit und Präzision bleiben durch die variable Sensorempfindlichkeit begrenzt. Darüber hinaus gibt es immer noch keine Standardverfahren für die Kalibrierung, und die meisten proprietären Kalibrierungsalgorithmen sind Blackboxen, die für die Öffentlichkeit nicht zugänglich sind. Mit dieser Arbeit soll die Wissensbasis über LCS auf verschiedene Weise erweitert werden: 1) durch die Entwicklung einer Open-Source-Kalibrierungsmethodik; 2) durch den Einsatz von LCS mit hoher räumlicher Auflösung in städtischen Umgebungen, um ihre Fähigkeit zur Messung kleinräumlicher Veränderungen der städtischen Luftverschmutzung zu testen; 3) durch die Verknüpfung von LCS-Einsätzen mit der Umsetzung lokaler Verkehrsmaßnahmen, um politische Empfehlungen zu den daraus resultierenden Veränderungen der Luftqualität geben zu können.

In einem ersten Schritt wurde festgestellt, dass LCS mit Hilfe von sieben allgemeinen Schritten konsistent und mit guter Leistung gegenüber Referenzinstrumenten kalibriert werden können: 1) Bewertung der Rohdatenverteilung, 2) Datenbereinigung, 3) Kennzeichnung von Daten, 4) Modellauswahl und -abstimmung, 5) Modellvalidierung, 6) Export der endgültigen Vorhersagen und 7) Berechnung der damit verbundenen Unsicherheit. Durch die Betonung der Notwendigkeit einer konsistenten Berichterstattung über Details bei jedem Schritt, insbesondere bei der Modellauswahl, -validierung und -leistung, hat diese Arbeit die Bemühungen um eine Standardisierung der Kalibrierungs-

methoden vorangetrieben. Darüber hinaus wurden mit der Open-Source-Veröffentlichung von Code und Daten für die siebenstufige Methodik Fortschritte bei der Reformierung der weitgehenden Blackbox-Natur von LCS-Kalibrierungen erzielt.

Nach der Einführung einer transparenten und zuverlässigen Kalibrierungsmethode wurden die LCS zwischen 2017 und 2020 an verschiedenen Straßen eingesetzt. Unter Verwendung von zwei Arten von LCS, Metalloxid (MOS) und elektrochemisch (EC), wurde ihre Leistung bei der Erfassung der erwarteten Muster der NO_2- und O_3-Belastung in Städten bewertet. Die Ergebnisse zeigten, dass die kalibrierten Konzentrationen der MOS- und EC-Sensoren mit den allgemeinen Tagesmustern der NO_2- und O_3-Belastung übereinstimmten, die mit Referenzgeräten gemessen wurden. Während sich MOS als unzuverlässig erwies, wenn es darum ging, Unterschiede zwischen den gemessenen Orten innerhalb der städtischen Umgebung zu erkennen, entsprachen die mit kalibrierten EC-Sensoren gemessenen Konzentrationen den Erwartungen aus Modellierungsstudien zur Verteilung der NO_2- und O_3-Belastung in Straßenschluchten. Daraus wurde der Schluss gezogen, dass LCS für die Messung der Luftqualität in Städten geeignet sind, auch zur Unterstützung der Entwicklung von Luftverschmutzungsmodellen auf städtischer Ebene, und dass sie neue Erkenntnisse über die Luftverschmutzung in städtischen Umgebungen liefern können.

Um das letzte Ziel dieser Arbeit zu erreichen, wurden zwei Messkampagnen im Zusammenhang mit der Umsetzung von drei verkehrspolitischen Maßnahmen in Berlin durchgeführt. Bei der ersten handelte es sich um den Bau einer Pop-up-Radweg auf dem Kottbusser Damm als Reaktion auf die COVID-19-Pandemie, bei der zweiten um die vorübergehende Einrichtung eines Gemeinschaftsraums in der Böckhstraße und bei der letzten um die Sperrung eines Teils der Friedrichstraße für den gesamten motorisierten Verkehr. In allen Fällen wurden NO_2-Messungen vor und nach der Durchführung der Maßnahme durchgeführt, um die Veränderungen der Luftqualität infolge dieser Maßnahmen zu bewerten. Die Ergebnisse des Experiments am Kottbusser Damm zeigten, dass die NO_2-Konzentrationen, denen die Radfahrer ausgesetzt waren, durch den Radweg um 22 ± 19 % gesenkt wurden. In der Friedrichstraße sank die NO_2-Konzentration durch die Straßensperrung auf das Niveau des städtischen Hintergrunds, ohne dass sich die Luftqualität in den Seitenstraßen verschlechterte. Diese wertvollen Ergebnisse wurden den verantwortlichen Ansprechpersonen in der Stadtverwaltung, die für die Bewertung des Erfolgs und der Zukunft der Maßnahmen verantwortlich sind, schnell mitgeteilt, was die Fähigkeit von LCS unterstreicht, politisch relevante Ergebnisse zu liefern.

Da es sich um eine neue Technologie handelt, muss noch viel über LCS und ihren Wert für die akademische Forschung im Bereich der Atmosphärenwissenschaften gelernt werden. Dennoch hat diese Arbeit den Stand der Technik in mehrfacher Hinsicht verbessert. Erstens wurde eine neuartige Open-Source-Kalibrierungsmethode entwickelt, die von LCS-Anwender*innen für verschiedene Luftschadstoffe verwendet werden kann. Zweitens wurde die Beweisgrundlage für die Zuverlässigkeit von LCS zur Messung der Luftqualität in Städten gestärkt, indem durch neuartige Einsätze in Straßenschluchten festgestellt wurde, dass LCS mit hoher räumlicher Auflösung zum Verständnis der Dynamik der Luftverschmutzung auf kleinräumlicher Ebene eingesetzt werden kann. Schließlich ist es die erste Studie dieser Art, die LCS-Messungen direkt mit verkehrspolitischen Maßnahmen verknüpft, um deren Einfluss auf die lokale Luftqualität zu verstehen, was zu politisch relevanten Erkenntnissen führt, die für Entscheidungsträger*innen wertvoll sind.

Die Studie ist ein Beispiel für das Potenzial von LCS, unser Verständnis von Luftverschmutzung in verschiedenen Maßstäben zu erweitern, sowie für ihre Fähigkeit, als wertvolle Werkzeuge in der transdisziplinären Forschung zu dienen.

1 Introduction

Throughout the past several hundred years, air pollution has been one of the most extensive, persistent, and challenging environmental problems on Earth (Fowler et al., 2020). One cannot discuss the Anthropocene without mentioning the impacts of air pollution on the biosphere, human health, and the climate. Despite successes in some parts of the world in addressing acute air pollution in the past few decades, it remains a significant stressor of human and environmental health in both developed and developing countries.

There are myriad biogenic and anthropogenic air pollutants that influence the environment and affect human health in various ways. Among them are gases such as nitrogen dioxide (NO_2), ozone (O_3), and volatile organic compounds (VOCs), and particulate matter (PM). While all of them are relevant when investigating the impacts of air pollution, this work focuses explicitly on two gaseous pollutants and their measurement: NO_2 and O_3. These species were selected for this study as their chemistry is interlinked, they have negative impacts on both the environment and human health, and are especially relevant to policymakers.

More recently, various state actors, scientists, and private industries have developed and implemented a wide range of technologies to effectively monitor air pollution and manage its impacts. With expanding innovation, new technologies and approaches to measuring air pollution have appeared in the past 10-15 years. Of particular importance are recent developments in low-cost sensing technology, henceforth referred to as low-cost sensors (LCS). The construction, calibration, and application of LCS has brought about rapid growth in the spatiotemporal measurement of air pollution around the world and enabled a new generation of end-users to engage with air pollution. Furthermore, as will be explored in this work, unique characteristics of LCS have encouraged innovation in the application of air pollution measurements to citizen-science and policy-relevant contexts.

In the following chapter, scientific literature will be explored and summarized regarding the connections between air pollution and the environment (1.1), including linkages to climate change, atmospheric chemistry, and meteorology; human health (1.2); and mobility (1.3). Thereafter, technical details of measurements of air pollution using reference-grade and low-cost instruments (1.4) will be described. Last, the important role of the Research Institute for Sustainability's (RIFS) transdisciplinary approach to scientific research on this work will be briefly explained.

1.1 Air pollution and the environment

Air pollution engages with the environment in complex ways; it both influences and is influenced by physical and chemical interactions with the biosphere, lithosphere, and atmosphere. These relationships define how air pollutants are emitted, transformed, transported, and deposited, as well as the magnitude of their impact on the surrounding environment. This is particularly evident in the troposphere, where anthropogenic and biogenic air pollutants react with other chemical species and are influenced by meteorology. This section highlights typical NO_x and O_3 chemical pathways and processes, further connecting them with meteorology and climate change.

1.1.1 Chemical pathways of NO_x and O_3

NO_x and O_3 are both among the most important molecules in atmospheric chemistry (Seinfeld & Pandis, 2006). NO_x refers to the total combination of nitric oxide (NO) and nitrogen dioxide (NO_2) and is short-lived, with an atmospheric lifetime of hours to a few days. Globally, 122 teragrams (Tg) of NO_x (as NO_2) were emitted in 2017 from anthropogenic activities (McDuffie et al., 2020), coming primarily from the shipping, land transportation, and energy production sectors. Transport emissions taken together, including shipping, land transport, and aviation, account for roughly 50% of the global total (Szopa et al., 2021). Over the last several decades, however, emissions-reducing technology has led to a consistent decline in annual anthropogenic NO_x emissions (McDuffie et al., 2020; Szopa et al., 2021).

Tropospheric O_3, on the other hand, is not emitted directly from anthropogenic or biogenic sources. Instead, it is a secondary pollutant and greenhouse gas (GHG) formed primarily through in situ chemical production and transport from the stratosphere, with an atmospheric lifetime of days to weeks (25 days on average, globally) (Seinfeld & Pandis, 2006). The global O_3 burden based on multi-model and observational estimates was calculated to be 347 ± 28 Tg in the year 2010 (Szopa et al., 2021), which is an increase from previous estimates of 337 ± 23 Tg and 310 Tg reported in AR5 and Seinfeld and Pandis, (2006), respectively. From 1850 to present day, it is estimated that the tropospheric O_3 burden has increased by 109 ± 25 Tg, largely due to increases in emissions of precursors such as NO_x and VOCs.

In the troposphere, NO_x and O_3 interact with one another in a chemical cycle that leads to the constant formation and destruction of O_3. The photolysis of NO_2 at wavelengths <424 nm leads to the production of an oxygen atom (eq. 1), which quickly reacts with oxygen (O_2) in a termolecular reaction with a third body (M) to produce O_3 (eq. 2). Once formed, O_3 can react with NO to regenerate the initial NO_2 (eq. 3), thereby maintaining a steady-state cycle (Seinfeld & Pandis, 2006).

$$NO_2 + h\nu \rightarrow NO + O \quad (1)$$

$$O + O_2 + M \rightarrow O_3 + M \quad (2)$$

$$O_3 + NO \rightarrow NO_2 + O_2 \quad (3)$$

However, in the presence of non-methane VOCs (NMVOCs), carbon monoxide (CO), and methane (CH_4), a different set of chemical reactions occurs to produce O_3 (Monks et al., 2015; Seinfeld & Pandis, 2006; Sillman, 1999; E. von Schneidemesser et al., 2015), seen below (eqs. 4-8):

$$VOC + OH \xrightarrow{O_2} RO_2 + H_2O \quad (4)$$

$$CO + OH \xrightarrow{O_2} HO_2 + CO_2 \quad (5)$$

$$RO_2 + NO \xrightarrow{O_2} VOC + HO_2 + NO_2 \quad (6)$$

$$HO_2 + NO \rightarrow OH + NO_2 \quad (7)$$

$$NO_2 + hv \rightarrow NO + O \quad (8)$$

$$O + O_2 + M \rightarrow O_3 + M \quad (9)$$

In brief, VOCs are oxidized by the hydroxyl radical (OH) to form peroxy radicals (RO_2), which then replace O_3 in the steady-state cycle and react with NO to produce NO_2. Similarly, CO can react with OH to form the hydroperoxyl radical (HO_2), which also replaces O_3 in the formation of NO_2. The net result of these reactions is the regeneration of the OH radical, allowing for the continuous oxidation of VOCs and a net accumulation of O_3 in the troposphere (R. Atkinson, 2000; E. von Schneidemesser et al., 2015).

Given these reactions, O_3 production tends to fall into two distinct regimes: VOC-limited and NO_x-limited. As VOCs and NO_x compete for reaction with the OH radical, the ratio of VOC/NO_x determines which regime dominates (Seinfeld & Pandis, 2006), but other factors such as VOC reactivity, biogenic emissions, and meteorology also impact the O_3 production regime (E. von Schneidemesser et al., 2015). In areas of high NO_x concentrations, the rate of removal of NO_x is limited by the availability of OH radicals, so further emissions do not increase the chemical processing of NO_x. In the rest of the troposphere, NO_x concentrations are far exceeded by OH concentrations, allowing for sufficient chemical removal of NO_x and a NO_x-limited regime (Sillman, 1999). Urban environments, on the other hand, tend to be NO_x-saturated or VOC-limited due to substantial local NO_x emissions (E. von Schneidemesser et al., 2015), whereas the opposite is true in the rest of the troposphere (Sillman, 1999).

1.1.2 Role of meteorology

The chemistry of NO_x and O_3 in the troposphere is closely linked with regional meteorology. These interactions occur at macro-, meso-, and microscale. At macroscale (horizontal distances > 1000 km), air parcel trajectories are determined largely by synoptic patterns and fronts. Features such as high- or low-pressure systems and cold fronts influence meteorological factors such as wind speed, wind direction, and temperature, and can have large impacts on local air pollution. For example, high-pressure systems tend to lead to stagnant meteorological conditions with low wind speeds, reducing pollutant dispersal and increasing local concentrations (Jacob & Winner, 2009; Perez, Garcia, Sanchez, Pardo, & Fernandez-Duque, 2020; Petetin et al., 2020; Erika von Schneidemesser et al., 2021). Synoptic scale meteorology also leads to the long-range, transboundary transport of air pollutants with lifetimes of days to weeks, as is the case with O_3 (Derwent, 2004; Monks et al., 2009). At mesoscale, the circulation of air parcels is dominated by influences from orographic features and land-sea interactions (Perez et al., 2020). The former, such as mountain-valley breeze circulation can influence the regional transport of precursors to polluted, VOC-limited areas, leading to enhanced O_3 production (Beaver & Palazoglu, 2009; Pusede et al., 2014). The latter, mixed with synoptic scale influences such as pressure, can affect wind speed and direction at the land-sea interface, thereby leading to pollutant dispersal or stagnation (Perez et al., 2020).

While macro- and mesoscale meteorological patterns shape the flow, type, and composition of air parcels that move through a region, microscale meteorological influences play an important role in local concentrations of air pollutants such as NO_x and O_3, particularly in urban environments. Microscale in this case refers to phenomena occurring at scales of 1 km or less (Seinfeld & Pandis, 2006). These include the mixing layer height (MLH), temperature (T), relative humidity (RH), and wind speed (WS), as well as urban morphology (Perez et al., 2020). While T, RH, WS, and MLH also fall under the umbrella of mesoscale meteorological patterns, their interactions with e.g., urban morphology and urban heat island effects result in unique microscale patterns relevant in the context of air pollution. For O_3, significant positive and negative correlations exist with T and RH, respectively (Melkonyan & Kuttler, 2012; Noelia Otero et al., 2018; N. Otero, Sillmann, Schnell, Rust, & Butler, 2016) in central Europe. Higher temperatures usually correlate with an increase in solar radiation, which implies more available energy for photochemical activity. As such, this relationship is particularly strong in summer, where longer days lead to enhanced O_3 production following photolysis of NO_2 (Ordóñez et al., 2005; Pusede et al., 2014).

Microscale meteorological influences on NO_x are quite different than for O_3, largely due to the greater role played by local emissions in urban areas. Studies have found NO_x to be minorly correlated with temperature (Pearce, Beringer, Nicholls, Hyndman, & Tapper, 2011) or not at all (Voiculescu et al., 2020); others identified a weak, positive relationship with relative humidity (Aldrin & Haff, 2005; Pearce et al., 2011; Voiculescu et al., 2020); and a few find minor evidence of a relationship with MLH (Geiß et al., 2017; Wagner & Schäfer, 2017). Instead, the most important meteorological factor for NO_x in urban environments is wind speed (Carslaw, Beevers, & Tate, 2007; Grundström, Hak, Chen, Hallquist, & Pleijel, 2015; Harkey, Holloway, Oberman, & Scotty, 2015; Pearce et al., 2011). This is most evident on days with low wind speeds, as NO_x emissions accumulate in stagnated air, leading to increases in concentrations of NO_2, as emitted NO is quickly converted to NO_2 (Carslaw et al., 2007; Elminir, 2006; Grundström et al., 2015; Pearce et al., 2011). Given the increased role of local NO_x emissions, wind direction has also been found to play an important role, with areas downwind of highways, urban centres, and industrial areas subject to increased NO_2 concentrations (Arain, Blair, Finkelstein, Brook, & Jerrett, 2009; Venkatram, Snyder, Isakov, & Kimbrough, 2013).

1.1.3 Connection with climate change

Through physical and chemical processes in the atmosphere, air pollutants also have an influence on and are influenced by climate change. O_3 is of particular importance in this regard, as it is a GHG in the troposphere. The Sixth Assessment Report (AR6) from the International Panel on Climate Change (IPCC) reported a contribution to effective radiative forcing (ERF) from 1750-2019 of 0.47 W m^{-2} (90% confidence range: 0.24 – 0.70 W m^{-2}) for O_3 (Szopa et al., 2021). For context, the ERF attributed to CO_2 over the same period is 2.16 W m^{-2} (Gulev et al., 2021). Due to the secondary nature of O_3, much of the increase in radiative forcing is attributed to rising anthropogenic emissions of precursors such as CH_4, NMVOCs, CO, and NO_x (Monks et al., 2015; Stevenson et al., 2013; E. von Schneidemesser et al., 2015). There are various further indirect effects of tropospheric O_3 on climate, including increased warming through the inhibition of plant uptake of CO_2 from the degradation of chlorophyll, as well as cooling from the repeated oxidation of VOCs to produce OH radicals that oxidize sulphur dioxide (SO_2) to sulphuric acid (Monks et al., 2015; E. von Schneidemesser et al., 2015).

For NO_x, impacts on radiative forcing occur largely through secondary pathways. Through the production of tropospheric O_3 as described in Section 1.1.1, NO_x has a positive ERF, whereas through enhanced production of the OH radical, which reduces the atmospheric lifetime of CH_4, and nitrate aerosols it has a negative ERF. Overall, these indirect effects of anthropogenic NO_x emissions from 1750-2019 were estimated to lead to a net ERF of –0.27 W m^{-2} (90% confidence range: -0.55 – 0.01 W m^{-2}), which is more negative than the AR5 estimate (Szopa et al., 2021). Due to spatiotemporal differences in NO_x effects on O_3 formation, anthropogenic NO_x emissions lead to warming in the short-term (10-20 years), but in the long-term the resultant removal of CH_4 has an overall cooling effect (E. von Schneidemesser et al., 2015).

1.2 Air pollution and health

As laid out in the previous section, NO_x and O_3 are important, prevalent air pollutants in the troposphere. These chemical species also impact human health in a variety of different ways. The Lancet recently published a comprehensive assessment of the health impacts of pollution in which it was estimated that 6.7 (5.9-7.5) million global deaths in 2015 were due to air pollution risk factors (Landrigan et al., 2018). A recent reassessment confirmed these values, finding the majority of deaths (4.1 million) were attributed to exposure to ambient $PM_{2.5}$ pollution, with 370,000 attributed to O_3 pollution (Fuller et al., 2022). A more recent study evaluated the impacts of ambient air pollution exposure in Europe using an updated evidence base to re-evaluate this burden of disease, estimating the number of premature deaths to be more than a factor of two higher than the Global Burden of Disease study in 2015 (Lelieveld, Klingmuller, et al., 2019).

As the majority of premature deaths from air pollution globally are associated with $PM_{2.5}$, most studies focus on quantifying the impact of particulates on human health. There are, however, a significant number of studies assessing the long-term impacts of NO_2 and O_3 exposure on human health. Several systematic reviews on long-term exposure to NO_2 (> 1 year) have found effects of NO_2 on mortality independent of other pollutants in multi-pollutant models (S. Huang et al., 2021; Huangfu & Atkinson, 2020). Similarly, longitudinal studies have found statistically significant positive associations between peak O_3 concentrations and respiratory disease, cardiovascular disease, lung cancer, and chronic obstructive pulmonary disease (COPD), independent of other pollutants (Huangfu & Atkinson, 2020; Kazemiparkouhi, Eum, Wang, Manjourides, & Suh, 2020; Lim, Hayes, et al., 2019). Other studies have identified statistically significant increased risks of mortality (all-cause, cardiovascular, and respiratory) even at low concentrations of air pollution (Brunekreef et al., 2021; Hanigan et al., 2019; Stafoggia et al., 2022), further highlighting the significant health effects of long-term exposure to any amount of air pollution.

1.2.1 Updated WHO air quality guidelines

Using the evidence base available in 2005, the World Health Organization (WHO) created a series of recommended air quality limit values for pollutants such as PM, NO_2, and O_3. These were partially used by the European Union (EU) to inform its air quality directive in 2008 (European Parliament, 2008). In light of new evidence presented in the previous section, especially with regards to new findings of the impacts of low concentrations of air pollutants on health, the WHO updated its recommended limit values in 2021 (WHO, 2021), which have been summarized in Table 1-1.

Table 1-1 Comparison of the WHO air quality guideline (AQG) levels for various pollutants between their original 2005 and the 2021 updated values, as well as with the current EU air quality limit values.

Pollutant	Averaging Time	2005 AQG level	2021 AQG level	Current EU limit-value
$PM_{2.5}$ ($\mu g.m^{-3}$)	Annual	10	5	25
	24-hour	25	15	*NA*
PM_{10} ($\mu g.m^{-3}$)	Annual	20	15	40
	24-hour	50	45	50
O_3 ($\mu g.m^{-3}$)	Peak Season	*NA*	60	*NA*
	8-hour	100	100	120
NO_2 ($\mu g.m^{-3}$)	Annual	40	10	40
	24-hour	*NA*	25	*NA*

Of the various changes made to the AQG levels, most notable is the reduction in the recommended annual average NO_2 level from 40 $\mu g.m^{-3}$ to 10 $\mu g.m^{-3}$. As noted in their report, the previous value was informed by morbidity effects observed in children exposed to elevated indoor NO_2 concentrations, whereas the updated value reflects the aforementioned long-term exposure studies on all-cause and respiratory mortality (WHO, 2021). The inclusion of a 24-hour recommended AQG value for NO_2 was similarly based on available epidemiological evidence. For O_3, the evidence base was not strong enough to justify a reduction in the 8-hour AQG level, but a peak seasonal value of 60 $\mu g.m^{-3}$ over the 6 months with the highest moving average was included to reflect the strong seasonality of this species.

As reported by the European Environment Agency (EEA), almost all Europeans are exposed to $PM_{2.5}$ (96%), NO_2 (89%), and O_3 (95%) concentrations above the new WHO recommended AQG levels (EEA, 2022). This continued exposure has enormous costs on society and any action taken to reduce air pollution will benefit the surrounding population (Hoffmann et al., 2021). With these changes, it is expected that the EU will aim to update its air quality directive to reflect the new targets (EEA, 2021a), ensuring the heightened health and policy-relevance of air pollution going forward.

1.3 Air pollution and the mobility sector

There are many sources of air pollution, both anthropogenic and biogenic, with the some of the largest sources coming from energy generation, industrial processes, transportation, and natural sources such as dust. In urban environments, one of the largest contributors to local concentrations of primary pollutants such as NO_x and secondary pollutants such as O_3 is the transport sector. In 2018, 39% of the total NO_x emissions across the EU-28 countries were contributed by the road transport sector, with non-road transport contributing a further 8% (EEA, 2020b). In the same year in Germany, road and non-road transport accounted for 40% and 4% of all NO_x emissions, respectively (Umweltbundesamt, 2022). In large urban agglomerations with large vehicle fleets such as Berlin, the relative contribution from transport to total emissions is higher. In 2015, car transport alone contributed 39% of the total NO_x emissions in Berlin, with a further 9% coming from other forms of vehicle transport (e.g., trucks, busses) (SenUMVK, 2022b).

While the transport sector still contributes a substantial portion of total NO_x emissions, the overall total in 2018 decreased by roughly 50% in the EU-28 with respect to 2000, despite increases in passenger and freight transport (EEA, 2020b). In Germany in 2018, total NO_x emissions were 58% of their 1990 levels, with a 64% reduction in road and non-

road transport emissions (Umweltbundesamt, 2022). These reductions from road transport are in large part due to improvements in vehicle technology that has reduced the total NO_x emitted over the lifetime of light-duty vehicles (Carslaw et al., 2019). However, despite substantial improvements in the past several decades, NO_2 concentrations are still well above the new recommended WHO AQG levels in all European cities and more needs to be done to further reduce the contribution from the mobility sector.

1.3.1 Impact of mobility infrastructure on air pollution

With such a large proportion of NO_x emissions coming from vehicle traffic, the extent and type of mobility infrastructure plays an important role in the dispersion and concentration of urban NO_2 and O_3 pollution. For example, highways, artery roads, and intersections typically exhibit higher NO_x emissions due to elevated traffic levels (Apte et al., 2017; Borge et al., 2016; Wallace & Kanaroglou, 2008), leading to high NO_2 concentrations, depending on meteorological conditions and site-specific characteristics affecting pollutant distribution. This in turn impacts the local formation and destruction of O_3, as explored in Section 1.1.1. Overall, changes to the built environment that impact the number of vehicles on a particular street or across a city can impact local NO_2 and O_3 concentrations.

Various studies have assessed changes in air quality associated with transportation infrastructure. Atmospheric models are often used to this effect, for example to assess the impact of large-scale transport policies such as traffic management strategies on air pollution (B. Degraeuwe, Pisoni, Christidis, Christodoulou, & Thunis, 2021; Tang, McNabola, & Misstear, 2020), project changes in air quality in association with Sustainable Urban Mobility Plans (SUMPs) (Pisoni, Christidis, Thunis, & Trombetti, 2019), or to calculate impacts on traffic-related air pollution based on various future emissions scenarios (Holnicki, Nahorski, & Kałuszko, 2021; Roustan, Pausader, & Seigneur, 2011). Other studies rely on city-wide air pollution monitoring networks to assess changes in air quality in connection with transport policies such as low emission zones (LEZ) (Boogaard et al., 2012; J. Gu et al., 2022; Lebrusán & Toutouh, 2020) and speed limits (Folgerø, Harding, & Westby, 2020). Fewer studies use targeted measurement campaigns to assess changes, such as local impacts of traffic management strategies (Krecl et al., 2020). While these studies are effective at assessing large-scale impacts of centralized policies on air quality, they are inadequate for assessing local impacts at street-level.

1.3.2 The mobility transition in Berlin

Cities across Europe are beginning to transform their approach to mobility infrastructure to mitigate climate change and air pollution. In Germany, this has been dubbed the “Verkehrswende” or “mobility transition” (Herberg, Haas, Oppold, & von Schneidemesser, 2020; Kallenbach, 2020). This movement exemplifies a shift in the underlying mobility paradigm, away from the dominance of individual motorized transport and towards the greater uptake of active mobility, particularly cycling. An important step in the development of this transition occurred in Berlin in 2018 with the passing of the Berlin Mobility Act, which was largely influenced by a citizen-led referendum on increasing the availability and safety of cycling infrastructure (Becker & Renn, 2019; Kunst, 2018; D. von Schneidemesser, Herberg, & Stasiak, 2020).

Though not a primary focus of the Berlin Mobility Act, as the city’s infrastructure shifts away from motorized transport, both vehicles and their emissions are displaced, resulting

in impacts on local air quality. As specific measures implemented under this act are decentralized, planned at city district level (Rode, 2019), and enacted irregularly over time, the aforementioned methods for assessing their impacts on air quality are inappropriate for determining air quality effects of individual measures, largely due to issues with resolution and uncertainty. Instead, targeted measurements associated with the implementation of each measure are needed, as the city-wide monitoring network is too sparse to capture local effects. Such local measurements can disentangle specific effects on local exposure to air pollution in association with each measure. The next section will highlight a new technology that meets these requirements and can fill an important gap in our understanding of the impacts of the mobility transition on air quality in cities.

1.4 Measuring air pollution

Currently, there are a wide range of instruments available for measuring air pollution. For in-situ, ground-based measurement of air pollution, these include, but are not limited to, reference-grade instruments, typically used in monitoring compliance to air quality regulations; passive samplers, a cost-effective method with high spatial but low temporal resolution; and low-cost sensors (LCS), an industrial technology that has found new applications in air pollution research. The latter is the focal point of this work and therefore will be the primary subject of this section, particularly with regards to the measurement of NO_2 and O_3.

1.4.1 Low-cost sensors

LCS are part of a changing paradigm in air pollution monitoring of the past 10-15 years, in which instruments are becoming cheaper, more portable, and easier to use (Snyder et al., 2013). These sensing elements are typically combined with other technological components for sampling (e.g., inlets or fans), operational hardware (e.g., microcontrollers), power systems (e.g., batteries, power supply units), storage and transmission capability (e.g., SIM cards, Wi-Fi, GSM), and weatherproof physical enclosures to form what are known as *sensor systems* (Peltier, 2020). While LCS have opened up a range of new applications for air pollution monitoring, they are still subject to greater uncertainty than reference instruments, as they are less sensitive, less precise, and are subject to various cross-sensitivities between chemical species (Peltier, 2020). In general, LCS can be separated into two broad categories: 1) those that measure concentrations of gas-phase species such as NO_2, O_3, or CO; and 2) those that measure PM or other properties of particles and aerosols (Snyder et al., 2013). This work focuses on gas-phase LCS for measurement of NO_2 and O_3, therefore this section will describe the technology, their calibration, and potential applications for research.

1.4.1.1 Technology

Available gas-phase LCS typically operate under a simple unifying principle of measurement, in which the sensing material interacts with the desired gas-phase species, producing a measurement with units of voltage or resistance. The two main technologies used to this end for measuring NO_2 and O_3 are metal oxide semiconductors (MOS) and electrochemical cells (EC) (Snyder et al., 2013). In the former, the semiconductor is composed of three main parts: the surface, the bulk, and the particle boundary (Figure 1-1). When heated to several hundred degrees centigrade, oxygen atoms in the air parcel will bond to the semiconductor, thereby extracting electrons from the surface. These bonded oxygen atoms will then either react with ambient gases or adsorb to the

semiconductor surface, changing the sensor resistance proportionally to the amount of gas reacting with the semiconductor (Peterson et al., 2017). This resistance is the raw sensor signal that is then calibrated against reference measurements of the target species.

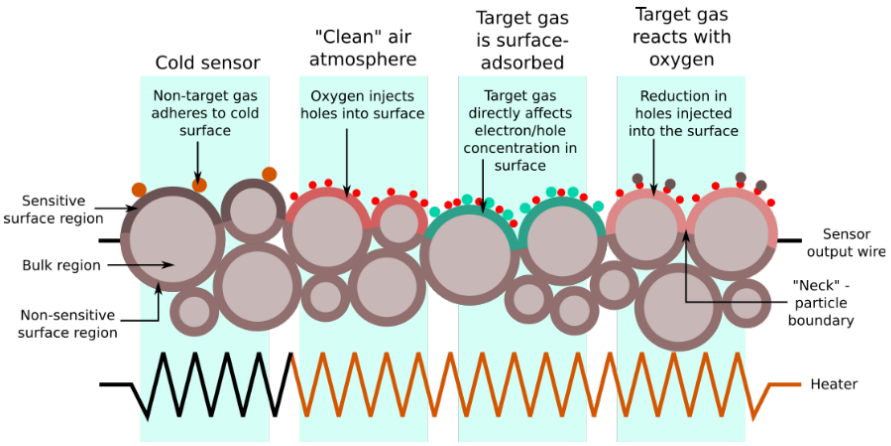

Figure 1-1 Diagram of the typical functioning of a metal oxide sensor describing the various components as well as the four types of reaction that can occur with gas-phase species. Reproduced under the CC BY 4.0 license from Peterson et al. (2017).

As a result of this mechanism MOS are subject to several limitations. The relationship between change in sensor resistance and species concentrations is typically non-linear, the semiconductors are sensitive to changing environmental conditions, particularly high temperature and relative humidity, and they are subject to numerous cross-sensitivities (Peltier, 2020; Peterson et al., 2017; Rai et al., 2017; Laurent Spinelle, Gerboles, Villani, Aleixandre, & Bonavitacola, 2015). Their lifetime is roughly 1-2 years but can also degrade sooner as continuous adsorption of chemicals to the surface of the semiconductor reduces the sensor's sensitivity. This causes a drift in the baseline sensor resistance over just a few months of deployment, therefore requiring MOS to be regularly calibrated with reference measurements to ensure suitable performance (Peterson et al., 2017). Last, as they must be heated to high temperatures, they have notable warm-up times (Peterson et al., 2017), including if the sensor turns off temporarily, and can give off excess heat inside the sensor system.

Electrochemical cell sensors operate via a different principle of measurement. These are composed of electrochemical cells in which a current is generated when a reduction-oxidation (redox) reaction at the interface between electrode and electrolyte produces a flow of electrons (Popoola et al., 2016). There are typically three or four electrodes in these sensors, depending on the model, including the worker, reference, counter, and auxiliary nodes, the latter of which is unique to four-electrode EC sensors (Figure 1.2). The target gas undergoes this redox reaction at the worker electrode at the gas-electrolyte interface, the current of which is counterbalanced by the reaction at the counter electrode. The reference electrode measures only in the electrolyte to keep the worker electrode at a stable potential. Similarly, the auxiliary electrode is never in contact with the gas species and is used to provide information on environmental effects on the worker electrode, particularly from temperature (Popoola et al., 2016; Rai et al., 2017). The resultant current is a measure of voltage and is calibrated against reference concentrations of the target species.

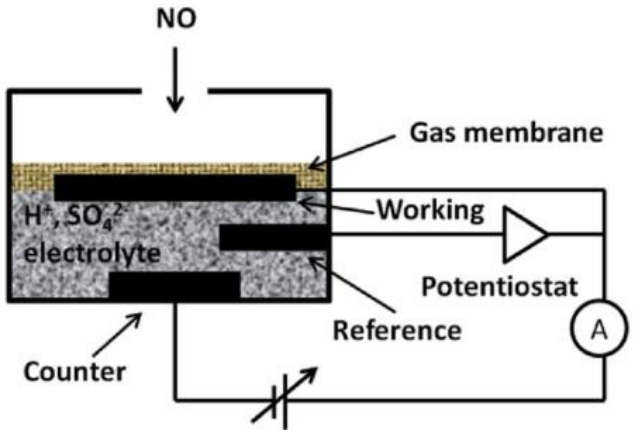

Figure 1-2 Diagram of the components of an electrochemical sensor for the measurement of gas-phase species, in this case with nitrogen oxide (NO). Reproduced under the CC BY 4.0 license from Popoola, Stewart, Mead, and Jones (2016)

EC sensors are also known to be susceptible to various sources of uncertainty and interference (David E. Williams, 2020). In particular, they are sensitive to temperature (particularly above 30°C) and relative humidity and require correction (Castell et al., 2017; Peltier, 2020; Popoola et al., 2016). As the electrolyte degrades through continuous use, they also experience drift over the course of 6 months to a year and require repeated calibration with reference instrumentation (Mead et al., 2013), though this drift is notably less substantial than for MOS. While still susceptible to cross-sensitivities with other gaseous species, EC sensors are more sensitive, stable, and accurate at parts-per-billion (ppb) concentrations for specific gas-phase species than MOS (Peltier, 2020; Russell et al., 2022).

1.4.1.2 Calibration

Given that LCS do not directly produce a measure of the ambient concentrations of gas-phase species, they must be properly calibrated before use in the field. To effectively calibrate this raw data against reference concentrations, models using various statistical and machine-learning techniques have been explored in various studies. Among these are linear regression (LR), multiple linear regression (MLR), random forests (RF), artificial neural networks (ANN), and support vector regression (SVR), which – with the exception of LR – have all been found to be appropriate for calibrating LCS (Bigi, Mueller, Grange, Ghermandi, & Hueglin, 2018; Cordero, Borge, & Narros, 2018; Hagan et al., 2018; Karagulian et al., 2019; Lewis et al., 2016; Malings et al., 2019; Laurent Spinelle et al., 2015; Laurent Spinelle, Gerboles, Villani, Aleixandre, & Bonavitacola, 2017; Zimmerman et al., 2018). However, there is tremendous variability among these results, as the environmental conditions of these deployments are quite different, leading to variable influences of cross-sensitive species, temperature, and relative humidity on sensor performance. Moreover, results vary across sensor systems, as performance is manufacturer, model, and firmware specific, and even if the sensor components themselves are the same, performance can vary (Peltier, 2020).

In contrast to regulatory monitoring of air pollution with reference instruments, there is still an absence of standardized procedures for LCS calibration. Some work to this end is underway in the United States (US) and Europe (EU) such as at the US environmental protection agency (US EPA) (R. Williams et al., 2019) or the EU Joint Research Centre (JRC) (Schneider et al., 2019), including some joint and individual efforts such as the EuNetAir sensor comparison exercise (Borrego et al., 2016; Borrego et al., 2018), or systematic

reviews on calibration methodologies (Jose M. Barcelo-Ordinas, Doudoub, Garcia-Vidala, & Badache, 2019; Karagulian et al., 2019). Despite these efforts, gaps still exist in the application of methodologies, including in the reporting of performance metrics and details on model selection, tuning, and validation. Further challenging standardization efforts are private sector companies, which sell LCS and sensor systems that use proprietary black-box algorithms to provide calibrated data to their users, and do not publish their methodologies (Peltier, 2020). As such, there are currently no standards for the calibration of LCS.

1.4.1.3 Applications

Given their small size and lower cost, LCS were from the beginning praised as having great potential for new applications in air pollution research (Snyder et al., 2013). In the past decade, many of these applications have been explored, including, but not limited to: 1) using LCS for improved personal exposure measurements in epidemiological studies (Jerrett et al., 2017; Mahajan & Kumar, 2020; Morawska et al., 2018); 2) increasing the spatial resolution of regulatory measurement networks with large numbers of LCS (J. M. Barcelo-Ordinas, Ferrer-Cid, Garcia-Vidal, Ripoll, & Viana, 2019; Mead et al., 2013; Popoola et al., 2018; Spandonidis et al., 2020; D. E. Williams, 2019); 3) citizen science studies that empower individuals to engage with the measurement of air pollution (Bosello, Delnevo, & Mirri, 2020; Languille et al., 2020; Mahajan, Luo, Wu, & Chen, 2021; Mijling, Jiang, de Jonge, & Bocconi, 2018; Ripoll et al., 2019); and 4) source apportionment studies (Bousiotis et al., 2022; Bousiotis et al., 2021). While LCS are still too inaccurate to be used for regulatory purposes or for personal exposure measurement (Castell et al., 2017; Jerrett et al., 2017) they can provide useful indicative measurements on air pollution sources and trends, especially where real-time information is necessary (McKercher, Salmond, & Vanos, 2017).

Despite the rapid expansion in research on applications of LCS, there remain still many unexplored areas of research. For example, beyond their deployment in higher spatial resolution measurement networks, LCS could also be deployed in targeted campaigns in various urban environments to understand more about the micro-dynamics of air pollution. This would have implications for improving understanding of pollution exposure and could even be used to validate and improve urban-scale models. Alternatively, such deployments could also be connected with physical changes to urban environments, such as in coordination with policies that modify transportation infrastructure to understand their impacts on air pollution and exposure.

1.5 Transdisciplinary research and air pollution

Many of the daunting environmental, social, and political problems of the 21st century are challenging, if not impossible, to solve due to their complex and interconnected characteristics. Arguably, air pollution belongs to the category of ‘wicked’ problems (Lawrence, Williams, Nanz, & Renn, 2022); emissions of air pollutants come from every aspect of human civilization, their impacts on society range from the economic to the epidemiological, and due to complex transboundary impacts, no single nation can mitigate air pollution alone. As such, while traditional scientific methods are effective at understanding fundamental processes underlying air pollutants and their impacts, they are less capable of providing the required information to decision makers. It is here where transdisciplinary research can be a valuable tool for tackling these ‘wicked’ problems.

There are three phases that generally describe the process by which transdisciplinary research proceeds (Lawrence et al., 2022). During Phase 1, a societal problem is transformed into a boundary object by a collaborative research team, which often results in an overarching idea or concept linking the societal and scientific aspects of the problem, thereby collectively identifying knowledge gaps relevant to all actors. Once this framework is established, in Phase 2 integrative methods are developed and applied to co-create the knowledge needed, with research typically proceeding in a disciplinary fashion and then becoming harmonized with interdisciplinary methods. In other words, Phase 2 is the actual "doing" of the research identified as necessary during Phase 1. In the final Phase, the different types of knowledge generated are re-integrated into scientific and societal discussions, through typical forms such as journal articles and presentations in the former, and in negotiations, debates, and news media in the latter. In turn, these discussions lead to a reformulation of the societal problem, by which the transdisciplinary research process begins anew.

1.6 Objectives and Research Questions

As previously outlined, with the growing inclusion of LCS into research in atmospheric chemistry, there are several gaps that remain to be investigated. In particular, more research is needed in standardizing LCS calibration methodologies, as well as in making them more transparent for users; in understanding and expanding their potential applications; and in their deployment for the production of policy-relevant results. With this in mind, this thesis investigates several clear objectives with three main guiding research questions:

(RQ 1) How can the calibration of low-cost sensors be standardized, while also made more transparent and accessible for end-users?

Although many researchers are expanding the knowledge base on methodologies for the calibration of LCS, there are still no standards to which they are beholden. Furthermore, the majority of LCS are marketed by private companies using calibration methodologies that are proprietary secrets; in other words, they are black-box and inaccessible to end-users. This work will develop an open-source calibration methodology for gas-phase sensors, with the goal of establishing a clear calibration protocol for standardization purposes, while also providing the relevant resources for end-users to implement themselves. In addition, it lays a foundation for the rest of this work by establishing a protocol for routine production of measurements of NO_2 and O_3 with gas-phase LCS.

(RQ 2) How suitable are LCS for the measurement of NO_2 and O_3 in urban environments? What new insights into urban NO_2 and O_3 pollution can be identified with high spatial resolution deployments?

Many studies seek to deploy LCS in urban environments, often with a focus on citizen science, building new measurement networks, or complementing existing networks. However, few have assessed precisely to what extent LCS capture expected patterns of urban pollution. Fewer still, if any, have measured air pollution with targeted deployments of LCS at high spatial resolution in urban environments such as street canyons. To expand upon this knowledge gap, several targeted deployments of LCS will be carried out in Berlin to assess their performance in measuring NO_2 and O_3 pollution, as well as to uncover their potential for application to this line of research.

(RQ 3) Can LCS be deployed in urban environments in connection to mobility policy and generate policy-relevant results for decision makers?

One line of research that has not yet been investigated is the direct application of LCS deployments to assess the impact of (mobility) policy. Due to their greater flexibility and lower cost, LCS can be deployed rapidly and at high spatial resolution, allowing them to be deployed alongside specific mobility policies to assess their impact on air quality. This work will be the first of its kind to use LCS to measure local air quality before and after physical changes to mobility infrastructure, assessing not only the ability of LCS to be used in this way, but also their potential to inform policymakers of their policies' impact on air quality. To this end, the three phases of transdisciplinary research will be utilized to engage with relevant societal actors, co-create the relevant knowledge, and disseminate the results back into public and political discourse.

1.7 Outline of thesis

The research described in this dissertation was conducted entirely in the Urban Air Quality, Mobility, and Health Research Group at the Research Institute for Sustainability, Helmholtz Centre Potsdam (RIFS). This thesis is structured around a core of four chapters, comprised of two published articles and two manuscripts submitted for peer-review.

Chapter 2: Unravelling a black box: an open-source methodology for the field calibration of small air quality sensors

In this chapter, a seven-step methodology for the field calibration of gas-phase low-cost sensors is introduced, with the goal of fulfilling objectives under RQ 1. It provides guidelines for end-users in predicting concentrations of NO_2 and O_3 using input data from reference instruments with statistical models. The RF machine-learning technique and MLR are used as examples, though other techniques could also be used with this methodology. This paper was published in *Atmospheric Measurement Techniques* as:

Schmitz, Seán, Sherry Towers, Guillermo Villena, Alexandre Caseiro, Robert Wegener, Dieter Klemp, Ines Langer, Fred Meier, and Erika von Schneidemesser. 2021. 'Unravelling a black box: an open-source methodology for the field calibration of small air quality sensors', *Atmospheric Measurement Techniques*, 14: 7221-41.

Personal contributions: development of the methodology and concept; data analysis; development of code and functions for online publication; production of figures and tables; primary author of the manuscript with contributions from co-authors.

Chapter 3: Using low-cost sensors to measure vertical and horizontal gradients of NO_2 and O_3 pollution in three street canyons in Berlin

This chapter serves as a proof-of-concept that LCS can be used to reliably measure NO_2 and O_3 pollution in urban environments and seeks to provide answers to RQ 2. For this study, measurement campaigns were conducted in three street canyons in Berlin between 2017 and 2020 using MOS and EC LCS. Using the seven-step methodology outlined in Chapter 2, all LCS used in these experiments were calibrated against reference instrumentation to provide concentrations of NO_2 and O_3. Results support the utility of LCS for reliable measurement of urban air pollution and indicate that LCS in targeted, high spatial resolution deployments can reveal new insights into the impacts of urban

morphology on pollutant concentrations. This manuscript was published in *Atmospheric Environment* as:

Schmitz, S., Villena, G., Caseiro, A., Meier, F., Kerschbaumer, A., & von Schneidemesser, E. (2023). Calibrating low-cost sensors to measure vertical and horizontal gradients of NO2 and O3 pollution in three street canyons in Berlin. *Atmospheric Environment*, 307(2), 119830.

Personal contributions: planning of measurement campaigns; installation of sensors; sensor co-location and calibration; data collection; data analysis; production of figures and tables; primary author of manuscript with contributions from co-authors.

Chapter 4: Do new bike lanes impact air pollution exposure for cyclists? — a case study from Berlin

Chapter 4 builds upon the foundations established in Chapters 2 and 3 by directly connecting measurements of NO_2 pollution with transport policy in Berlin, largely focusing on answering RQ 3. In a case-study from 2020, a measurement campaign was conducted on Kottbusser Damm in Berlin alongside the construction of a new bike-lane and the implementation of a car-free community space on a side-street. Using before-after stationary and mobile measurements, the direct impacts of these policies on local air quality and on cyclists' exposure to NO_2 pollution was quantified. These results were then directly communicated to the relevant policymakers, who were also partners in the development of the project. This paper was published in *Environmental Research Letters* as:

Schmitz, Seán, Alexandre Caseiro, Andreas Kerschbaumer, and Erika von Schneidemesser. 2021. 'Do new bike lanes impact air pollution exposure for cyclists?—a case study from Berlin', *Environmental Research Letters*, 16.

Personal contributions: concept design; planning and execution of measurement campaign; coordination with project partners; data collection; sensor co-location and calibration; data analysis; production of figures and tables; primary author of manuscript with contributions from co-authors.

Chapter 5: Low-cost system application for policy assessment: a case study from Berlin

In the final chapter of this work, another transport policy is assessed using targeted deployments of LCS, also addressing RQ 3. Beginning in 2020, a section of the Friedrichstrasse street in Berlin was closed to car traffic and opened to pedestrians and cyclists. In connection with this mobility policy and in partnership with responsible decision makers, a measurement campaign was conducted to quantify the impact of the street closure on local air quality, both on the Friedrichstrasse and on side streets. Results indicate that NO_2 concentrations sank to the level of the urban background on the Friedrichstrasse and did not increase on side streets. It serves as another case-study of the potential for LCS to provide policy-relevant advice in targeted campaigns alongside mobility policies. This manuscript was submitted to *Atmospheric Environment: X* as:

Caseiro, Alexandre, Seán Schmitz, Andreas Kerschbaumer, and Erika von Schneidemesser. 'Low-cost system application for policy assessment: a case study from Berlin'. Atmospheric Environment: X. Submitted on 26.08.2022.

Personal contributions: design and planning of measurement campaign; coordination with project partners; sensor co-location and calibration; data collection; minor contributions to data analysis; minor contribution to production of figures and tables; second author of manuscript, primary author of methods and discussion sections.

2 Unravelling a black box: An open-source methodology for the field calibration of small air quality sensors

Abstract

The last two decades have seen substantial technological advances in the development of low-cost air pollution instruments using small sensors. While their use continues to spread across the field of atmospheric chemistry, the air quality monitoring community, as well as for commercial and private use, challenges remain in ensuring data quality and comparability of calibration methods. This study introduces a seven-step methodology for the field calibration of low-cost sensor systems using reference instrumentation with user-friendly guidelines, open access code, and a discussion of common barriers to such an approach. The methodology has been developed and is applicable for gas-phase pollutants, such as for the measurement of nitrogen dioxide (NO_2) or ozone (O_3). A full example of the application of this methodology to a case study in an urban environment using both Multiple Linear Regression (MLR) and the Random Forest (RF) machine-learning technique is presented with relevant R code provided, including error estimation. In this case, we have applied it to the calibration of metal oxide gas-phase sensors (MOS). Results reiterate previous findings that MLR and RF are similarly accurate, though with differing limitations. The methodology presented here goes a step further than most studies by including explicit, transparent steps for addressing model selection, validation, and tuning, as well as addressing the common issues of autocorrelation and multicollinearity. We also highlight the need for standardized reporting of methods for data cleaning and flagging, model selection and tuning, and model metrics. In the absence of a standardized methodology for the calibration of low-cost sensor systems, we suggest a number of best practices for future studies using low-cost sensor systems to ensure greater comparability of research.

Published as:

Schmitz, S., Towers, S., Villena, G., Caseiro, A., Wegener, R., Klemp, D., Langer, I., Meier, F., & Von Schneidemesser, E. (2021). Unravelling a black box: An open-source methodology for the field calibration of small air quality sensors. *Atmospheric Measurement Techniques*, 14(11), 7221–7241. https://doi.org/10.5194/amt-14-7221-2021

2.1 Introduction

Air pollution remains a leading cause of premature death globally (Landrigan et al., 2018). The recent trend in air pollution research of using low-cost sensors (LCS) to measure common gas-phase and particulate air pollutants (e.g. CO, NOx, O_3, PM) is an attempt to close gaps in our understanding of air pollution and make its measurement cheaper, widespread, and more accessible (Kumar et al., 2015; Lewis et al., 2016; Lewis, von Schneidemesser, & Peltier, 2018). The development of these new technologies represents a paradigm shift that has opened up air pollution monitoring to a much wider audience (Morawska et al., 2018; Snyder et al., 2013). In recent years, LCS have been used to develop or supplement existing air pollution monitoring networks to provide higher spatial resolution (e.g. CitiSense, U.S. EPA Village Green), as well as in a citizen science contexts to report on and share information about air quality (e.g. AirVisual, Purple Air) (Morawska et al., 2018; Muller et al., 2015). Projects like these are a promising step towards empowering citizens with greater knowledge of their local air quality.

However, as there are myriad commercially available LCSs that use a variety of sensors and have substantial differences in quality, standardizing their application remains challenging and urgent (Karagulian et al., 2019). In measuring gas-phase pollutants, for example, metal oxide sensors (MOS) and electrochemical sensors (EC) are often used which have different limits of detection and cross-sensitivities that need to be taken into account (Lewis et al., 2016; Lewis et al., 2018; Rai et al., 2017). Under ambient conditions, the performance of these two sensor types varies substantially, with some studies reporting moderate to good agreement with concentrations measured by reference instrumentation, whereas others find very poor agreement (Lewis et al., 2018). A further challenge is that many LCS are in the form of small sensor systems[1] sold as ready-to-use products to customers, most often using a “black box” proprietary calibration algorithm for producing concentrations which, along with raw data, is not publicly available (Karagulian et al., 2019). Furthermore, a wide range of calibration techniques have been applied to LCS in field studies, but lack uniformity in metrics used, experimental setup, reference equipment, and environmental conditions, making it difficult to draw conclusions about their overall performance (Karagulian et al., 2019; Rai et al., 2017).

In general, pairwise reference calibration has been done on an individual sensor system basis as well as a sensor system cluster basis, also known as “sensor fusion” (Jose M. Barcelo-Ordinas et al., 2019). The former tends to be more accurate but becomes logistically and computationally intensive for large numbers of LCS and is more sensitive to sensor decay and medium-scale drift. The latter has been shown to be effective at calibrating groups of sensors when using the median sensor signal of a co-located cluster of sensors to develop a single calibration model applicable to all sensors (K. R. Smith et al., 2017; Kate R. Smith et al., 2019). Using a cluster-based approach has been shown to produce calibration factors that may be more robust over longer time frames but have higher margins of error for individual sensors. Both methods have their advantages and disadvantages that must be balanced based on the desired application for the sensor systems. Further methods for calibration beyond pairwise reference calibration include

[1] In this case “sensor” and “LCS” refer to the sensor components which react chemically with various air pollutants, whereas “sensor system” refers to the complete device, including sensors, housing unit, data storage, etc.

node-to-node calibration (Kizel et al., 2018) or proxy calibration (Miskell, Salmond, & Williams, 2018).

Previous research has used linear regression, multiple linear regression (MLR), and machine-learning techniques such as random forest (RF), artificial neural networks (ANN), and support vector regression (SVR) to calibrate LCS with reference instrumentation for gas-phase pollutants. Here too, there is a lack of standardization, as MLR, RF, ANN, and SVR have all been found to be the most accurate method across various studies (Bigi et al., 2018; Cordero et al., 2018; Hagan et al., 2018; Karagulian et al., 2019; Lewis et al., 2016; Malings et al., 2019; Kate R. Smith et al., 2019; Zimmerman et al., 2018). Only linear regression has been consistently identified as an unsuitable model, largely because it fails to take into account cross-sensitivities and environmental influences on sensor functioning and because sensors responses are often non-linear. For this same reason, nonparametric methods such as the aforementioned machine-learning techniques tend to be more accurate, as they are better at modelling non-linear sensor responses while being able to better take into account interferences in sensor functioning (Jose M. Barcelo-Ordinas et al., 2019; Karagulian et al., 2019). However, it must be said that any of these statistical methods can be applied as long as they properly account for autocorrelation, multicollinearity, and non-linearity in the data with relevant transformations.

There are several key issues with previous work on calibrating LCS that must be acknowledged. First, the metrics used to report model suitability vary substantially. Karagulian et al. (2019) found in their comprehensive review of the LCS literature that only the coefficient of determination (R^2) was applicable for cross-comparison of all studies. While this metric can be useful in measuring the agreement between LCS data and reference measurements, it does not give a sense of the model error. Future studies should, at a minimum, report R^2, root mean square error (RMSE), and mean average error (MAE), when discussing calibration performance (Jose M. Barcelo-Ordinas et al., 2019; Karagulian et al., 2019). Second, while there are many studies that calibrate LCS with MLR or machine-learning techniques, the associated model selection, validation, and tuning methods are rarely reported. The latter of these is especially important for machine-learning (ML) techniques with many tuning parameters, where the problem of over-fitting is more common. Some studies do report steps for model validation (Hagan et al., 2018; Laurent Spinelle et al., 2015; Zimmerman et al., 2018) or model tuning (Bigi et al., 2018; Laurent Spinelle et al., 2015), but they do not go into depth as to how these were determined or optimized. Especially with “black box” techniques such as ANN, SVR, or RF, reporting steps taken to validate the model and optimize parameters is crucial to ensuring consistency among studies. Last, the issues of multicollinearity and autocorrelation, which are common among LCS time series data and of substantial importance when using MLR, are rarely addressed. If at all mentioned, they are referred to as being better handled by non-linear ML techniques such as SVR or RF (Bigi et al., 2018) or as potentially obscuring the statistical significance of models (Masiol, Squizzato, Chalupa, Rich, & Hopke, 2018). This study seeks to take a step forward in ensuring these issues are addressed in future LCS calibration studies.

In the absence of a standardized calibration methodology, the ever-growing body of LCS literature will continue to be largely incomparable, with research running in parallel using varied methods. Though several comprehensive reviews of LCS have been completed which establish helpful guidelines for their use (Lewis et al., 2018; R. Williams et al., 2014),

best practices for calibration with reference instruments that should be undertaken in any field deployment were not specifically reported. More recently, Barcelo-Ordinas et al. (2019) published an extensive study on the calibration of LCS, including some general calibration guidelines. While these are a helpful guide for calibration methodologies, they lack important details on the post-processing of data during the model-building process. This study seeks to expand upon this work and specifically address the standardization of individual pairwise calibration of LCS housed in sensor systems with reference instrumentation by presenting user-friendly guidelines, open access code, and a discussion of common barriers to field calibration. With the publication of this step-by-step methodology for the statistical calibration of low-cost sensor systems, we hope to establish a framework from which calibration methods can be better compared.

2.2 Methods

The following section outlines a methodology for the deployment and field calibration of LCS for the measurement of gas-phase pollutants. First, some key considerations for the experimental deployment of small sensor systems will be discussed. Second, a 7-step statistical calibration methodology for the post-processing of data will be described. Last, an example of the use of this methodology, both for deployment and calibration, using data collected during a measurement campaign in 2017 and 2018, is provided (Section 2.3).

For this methodology, it is important to first establish under which circumstances the following steps would apply. This is a reference-based pairwise method for the individual calibration of small sensor systems and therefore the user will need to have access to reference instrumentation with which the small sensor systems can be co-located, whether their own or in collaboration with e.g. a city monitoring network. This makes the methodology inapplicable for individual users in a citizen science context that may not have access to reference instrumentation. These reference instruments should adhere to standardized guidelines on accuracy (i.e. EU Air Quality Directive (2008/50/EC), U.S. National Ambient Air Quality Standards (NAAQS)). A co-location in this sense refers to the installation of the small sensor systems in the close vicinity (ca. 1-3 meters) of the reference instruments, so that they receive the same parcels of air. This paper focuses on the usage of field (i.e. in-situ) co-locations in calibrating small sensor systems. If access to reference data or the raw small sensor data is not possible, then this methodology cannot be applied. Furthermore, while it was applied here to sensor systems containing metal oxide LCS, this methodology is also equally as applicable to electrochemical (EC) LCS or photoionization detectors (PID), as these produce a similar measure of voltage that varies in response to changing concentrations of gas-phase species and have similar cross-sensitivities to temperature and relative humidity. It is not directly applicable for optical particle counters (OPC) for the measurement of particulate matter, as the transformation of the raw data into concentrations during calibration functions differently, though some of the principles discussed here are still relevant. For an application of this methodology to EC sensors, please see Schmitz et al., (2021).

2.2.1 Key considerations for the experimental deployment of small sensor systems

When calibrating small sensor systems, the experimental deployment and co-location of devices is a key step with several important considerations that must be accounted for. First, the co-location with reference instrumentation should ideally occur at the same test site where the small sensor systems are to be deployed. If unfeasible for logistical reasons,

an analogue site should be selected. Criteria for analogue selection entail similar characteristics as those for test site selection. The analogue site should: 1) have similar sources and ranges of concentrations of air pollutants as the test site; 2) experience comparable meteorological conditions and similar circulation dynamics; and 3) be physically located in the same region as the test site. While it is unlikely that there will be a perfect analogue site, any field calibration should take these criteria into consideration in order to enhance validity of experimental results.

Second, the frequency and timing of co-locations should reflect site-specific variations in meteorological conditions. Generally, these should be done often enough so that co-location data cover similar ranges of meteorological conditions and concentrations of air pollutants as the experimental data, but not so often that there is a concomitant loss of experimental data. A rule-of-thumb for long-term experiments (>6 months) in temperate seasonal environments is a 2-week co-location every 2-3 months. For short- to medium-term experiments, a 2-week co-location before and after and perhaps one in-between, depending on changes in meteorological conditions, should suffice. Regular co-location allows for the establishment of datasets that cover not only changes in meteorology, but also sensor functioning and interactions of potentially cross-sensitive species. If these considerations are taken into account during the experimental deployment, the likelihood that these datasets will be of good quality will be higher. In this study, we focus primarily on stationary field deployment of low-cost sensor systems. There are, however, other forms of deployment, including indoor and mobile, for which these criteria also apply. It is important to mention that there may be other considerations required in such alternative forms of deployment, e.g., more scrutinous data cleaning in mobile deployments due to impacts of rapidly changing environments on sensor performance.

2.2.2 The 7-step statistical calibration method

Raw data from small sensor systems, if treated and transformed properly, can provide informative air pollutant concentrations. This treatment must, however, be rigorous if the resultant concentrations are to be used in further analysis. This section provides a general description of a seven-step methodology for the post-processing and calibration of LCS data gathered with small sensor systems. Multiple Linear Regression (MLR) and Random Forest (RF) were selected as calibration methods to be used in this methodology, although it can be generally applied to other regression or machine learning methods. Information on the functions and packages from the open-source R statistical software program R Core Team (2019) used in this methodology is provided for each step. This information and the code can be found in the open-source repository Zenodo (see below for DOI).

2.2.2.1 Step 1: Analyse and understand raw data distribution

The first step is to gain a general understanding of the data. Specifically, establishing an overview of data distributions and potential data quality issues (data gaps, presence of outliers, changes in baselines, etc.) is helpful for identifying problems and solutions during calibration. It should also be checked that all associated metadata are available for all datasets.

In this study, all variables that were to be used in model selection were assessed in this step. For example, the distributions of the reference concentrations, small sensor system raw data, and meteorological variables from the co-location and experimental datasets were analysed. Meteorological variables including temperature, relative humidity, and wind speed and direction across the co-location and experimental datasets were

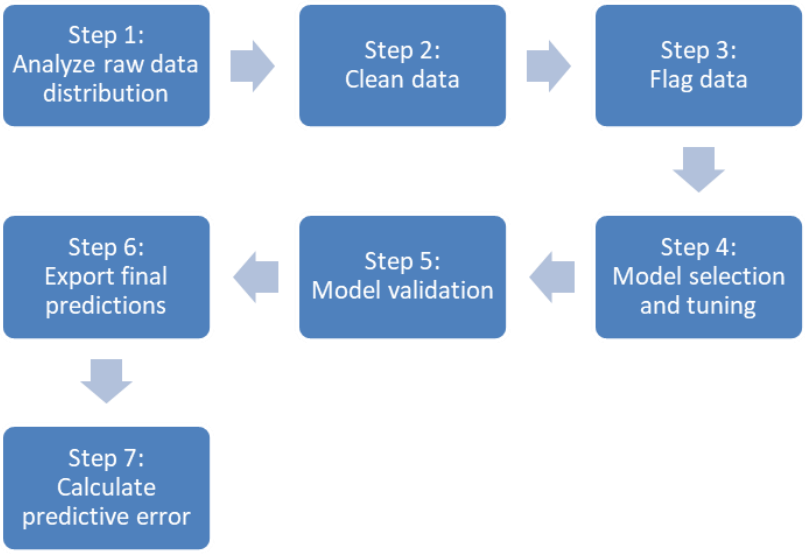

Figure 2-1 Schematic representation of the seven-step calibration method for processing small sensor system data.

compared. Additional variables that could be considered but were not analysed here include precipitation, boundary layer height, and insolation, among others. A visual assessment of these data using histograms, violin plots, and time series plots was conducted. This step provided information about the structure of each available co-location dataset and the experimental dataset crucial to decision-making in later steps.

2.2.2.2 *Step 2: Data cleaning*

Next, the datasets should be cleaned of erroneous outliers and unreliable data. This step is crucial, as outliers can have a particularly strong effect on calibration models and especially so on linear regression models.

To accomplish this, the time series plots generated in step 1 were first used to visually evaluate the data. Sequence outliers resultant from sensor warm-up time or sensor malfunctioning were identified and removed using an automated function. Next, an algorithm was tested, trained, and implemented that uses a simple z-test with a running mean and standard deviation to detect point outliers resultant from instrument measurement error. Tests of normality with datasets greater than 50 points are irrelevant in determining whether parametric tests can be used or not (Ghasemi & Zahediasl, 2012). Analysis of the data in this study revealed the same, as data segments of less than 30 points consistently passed the Shapiro-Wilk test, but with progressively larger data segments, more and more of the data failed the test. Therefore, it was assumed that the data aligned enough with the normal distribution for this test to apply. The size of the moving time frame from which the running mean and standard deviation were calculated and the z-score threshold used to designate ‘outlierness’ were tested and optimized. Durations of 1, 2, 5, 10, 30, 60, 120, and 300 minutes were considered for the moving window and thresholds of 3, 4, 5, and 6 were tested. This was done for each variable individually. The points flagged as outliers with this method were then graphically assessed against neighbouring datapoints to prevent inadvertent removal of peak emission events. In other cases where assessing all outliers is impractical, it is recommended to do so with a random subset of outliers. Furthermore, if substantial short-term events are expected due to the deployment environment, such as during mobile measurements, a more thorough check of potential outliers should be done. Other

outlier detection functions using Autoregressive Integrated Moving Average (ARIMA) and Median Absolute Deviation (MAD) were tested and were found to be inappropriate for this data.

2.2.2.3 *Step 3: Flag data for further scrutiny*

Experimental data outside the range of the co-location data (i.e., beyond the minimum and maximum values) should be flagged as they may be less reliably predicted than those which are in-range and should be given a higher level of uncertainty (Kate R. Smith et al., 2019). Flagging such data points strikes a balance between removing them from the analysis and highlighting their associated uncertainty.

Once flagged, these data points were treated differently in later analysis (Section 2.3.5). Similarly, co-location data outside the range of the experimental data received a flag. During the model selection process, these flags were used to remove data that may serve to bias the model. While this may seem unnecessary, if the experimental range of environmental conditions is much smaller than those of the co-locations, it could be that using a smaller, more comparable range of co-location data is more suitable for model selection. This is data and model dependent, however, and was therefore tested in Step 6.

2.2.2.4 *Step 4: Model selection and tuning*

Model selection and tuning is a seldom-reported step that is vital in ensuring the calibration model is suitable for use. Rigorously scrutinizing a variety of potential models and optimizing their parameters provides reproducible justification for the final model selected. This is particularly important for machine-learning techniques which can have a wide array of parameters for tuning model performance. Furthermore, appropriate methods used in model selection ensure that problems of multicollinearity and autocorrelation can be corrected for, as superfluous predictors suffering from these issues will be identified and removed. Before building and selecting potential models, the relationships between predictors and response variable, including potential transformations, must be determined. This is important for linear regression models but is not relevant for ML techniques which do not take these transformations into account. Often the sensor specifications indicate what type of transformation (exponential, log-linear, etc.) may be necessary.

The co-location data were used in this step to train various models and determine the best fitting MLR and RF models. In this case, log transformations were recommended for the MOS sensors used, but were cross-checked with other common transformations including: log-log, square-root, and inverse. Model selection proceeded through backwards selection using the coefficient of determination (R^2), root mean squared error (RMSE), and the Akaike Information Criterion (AIC) (Akaike, 1973) for MLR or Variable Importance (VI) (Breiman, 2001) for RF as criteria. To determine the best models, the training data set was broken up into smaller sets by using a moving window of four days to train the models and the fifth day to test. The models with the best average RMSE over the various fifth day predictions were selected.

For RF the model parameters of mtry (the number of randomly selected variables at each node), min.node.size (the minimum number of data points in the final node), and splitrule (the method by which data are split at each node) were optimized by testing various combinations and selecting the most accurate in terms of RMSE, with data split in the same manner as for MLR. Subsequently, measures of AIC for the regression model and VI

for the random forest model were assessed to determine which predictors should remain in the model. For MLR, this involved the repeated bootstrapping of the training set combined with stepwise selection, using the AIC to robustly determine predictor inclusion. The models were then finally tested on the test subset and assessed using RMSE and R^2. The most accurate MLR and RF models were then sent to the next step for validation.

2.2.2.5 Step 5: Model validation

Model validation is often overlooked but is necessary to ensure that the most accurate model selected is reliable (i.e. has good predictive power for independent data). While a singular instance of splitting the dataset during the model selection process into training and testing subsets is one method of validating the model, an additional step ensures more rigorous validation.

In this case, to validate the MLR and RF models selected in Step 4, the co-location data was repeatedly split into training and testing subsets at a ratio of 75/25. This was done by splitting the co-location training set into continuous blocks representing 25% of the training data (in this case 6 days) as test subsets and using the rest of the co-location data to train the model. A robustness cross-check with various splitting ratios was conducted and found that changing the splitting ratio did not significantly impact the results. Using continuous blocks instead of random sampling is necessary to account for the autocorrelation in the data (Carslaw & Taylor, 2009). The accuracy of the final models was then assessed on the continuous blocks using R^2, RMSE, and Variable Importance. These metrics were then graphed across all continuous blocks to assess model stability. In this case, instability refers to major differences in R^2 and RMSE between folds likely caused by differing field conditions among the training and test folds. If this is seen, it indicates that the model may be too sensitive to changes in field conditions. If the graphs showed instability across the various folds, Step 4 was repeated and a new model was selected for validation.

2.2.2.6 Step 6: Export final predictions

Once the selected model has been validated, the next step in the process is to export predictions of the experimental data as concentrations. Only co-location data deemed relevant from the Steps 1-3 should be used to train the model, which is then used to predict experimental concentrations.

In this case, the co-location data were used to train the best MLR and RF models identified in Steps 4 and 5. These models were then applied to the raw experimental data in order to predict final concentrations. The final predictions were then graphed and compared using time plots and histograms.

2.2.2.7 Step 7: Calculate total predictive error

Last, it is vital that overall error and confidence intervals for the predictions are reported in this step. Most models have associated methods for reporting metrics such as standard error which can be used to establish confidence intervals around the predictions. Compounded to this must be the technical error associated with measurements from the reference instruments. Thus, the overall error should combine technical and statistical error.

In this study, to test the impact of the precision of the reference measurements on model accuracy, the reference NO_2 and O_3 data were smeared using a normal distribution with

each point as the mean and each instrument's measure of imprecision as the standard deviation. Smearing refers to transforming the data by shifting the actual value within the range of uncertainty. This test therefore determined whether the imprecision given by each instrument's specifications should be factored into the overall predictive error. This was done over 50 iterations to see how model accuracy responded to shifts in reference concentrations within the margins of error. Co-location data were split 75/25 into a training set and testing set, respectively. In each iteration, separate MLR and RF models for NO_2 and O_3 were trained; each was trained once with the reference measurements and once with smeared reference measurements. All models were then tested for predictive accuracy on the testing subset, to compare the impact of smeared versus measured reference data on model performance.

Last, the overall uncertainty was calculated. For the reference instruments, the technical measurement was taken from their specifications. This was added to the overall statistical error, for which the median MAE across all blocks from the model validation step was used. Both the MLR and RF model calculated a measure of standard error, which was compared with the combined uncertainty measure. The more appropriate of the two was then added to the final predictions from Step 6.

2.3 Example application of the methodology

2.3.1 Small sensor systems used

The small sensor systems used in this example are EarthSense Systems, Ltd. "Zephyr" prototypes[2], henceforth referred to as "Zephyrs". This term refers to the whole small sensor system including housing, sensors, GPS, etc. Installed within the Zephyr prototypes were a number of Metal Oxide Sensors (MOS) that measure reducing gases, oxidizing gases (used here for detection of nitrogen dioxide), ozone, and ammonia, as well as a meteorological sensor for temperature and relative humidity; see Table 2-1 for more on these sensor specifications. These MOS sensors typically experience significant amounts of drift four months after initial calibration, which is why in this study co-locations were conducted at high frequency, before and after each experiment. For greater detail on the development, functioning, and operation of the sensors housed within these prototypes see Peterson et al. (2017).

2.3.2 Reference instruments

The reference instrumentation included a Teledyne Model T-200 NO/ NO_2 /NO_x Analyser and a 2B Technologies, Inc. Ozone Monitor. These instruments were intercompared with reference instruments – CAPS (Aerodyne, U.S.A.), CLD 770 AL ppt (ECOPHYSICS, Switzerland) and O242M (Environnement S.A., France) – from the Forschungszentrum Jülich as part of the measurement campaign and showed decent agreement (R^2 = 0.70) for NO_2 and good agreement (R^2 = 0.88) for O_3 (see Figures 2.S-1-2.S-3 in the supplemental information). Ambient air temperature and relative humidity (Lambrecht, PT100) data one block away from the experimental site were provided by the Free University Berlin for two measurement campaigns (more information in Section 2.3.3). Wind speed and direction (Campbell Scientific, IRGASON) were measured 10m above the roof of the main building of the Technical University Berlin (TUB) at Campus Charlottenburg, which is

[2] The EarthSense Zephyrs have since evolved substantially and, as such, this study does not represent current performance or configuration.

Table 2-1 Sensors installed within the EarthSense Zephyr prototypes. Table reproduced from Peterson et al. (2017).

Gases Measured	Sensor Model	Method of detection	Gas detected and detection limits
Reducing gases	SGX Sensortech MICS-4514	Redox reaction	CO: 1-1000 ppm NH_3: 1-500 ppm C_2H_5OH: 10-500 ppm H_2: 1-1000 ppm CH_4: >1000 ppm
Oxidising gases	SGX Sensortech MICS-4514	Redox reaction	NO_2: 0.05-10 ppm H2: 1-1000 ppm
Ozone	SGX Sensortech MICS-2614	Redox reaction	10-1000 ppb
Ammonia	SGX Sensortech MICS-5914	Redox reaction	NH_3: 1-500 ppm C_2H_5OH: 10-500 ppm H_2: 1-1000 ppm C_3H_8: >1000 ppm $C_2H_8(CH_4)_2$: >1000 ppm
Temperature and relative humidity	GE Measurement and Control CC2D25	Polyamide capacitance	Temp.: -40 – 125 °C RH: 0 – 100%

located across the street from the experimental site. This site is part of the Urban Climate Observatory (UCO) Berlin operated by the TUB for long-term observations of atmospheric processes in cities (Scherer, Ament, et al., 2019).

2.3.3 Experimental deployment

Measurements were conducted in a street canyon on the Charlottenburg Campus of the TUB, on the façade of the Mathematics Building (52° 30' 49.7" N, 13° 19' 34.5" E) as a part of several measurement campaigns of the joint project 'Three-dimensional observation of atmospheric processes in cities' (3DO) (Scherer, Ament, et al., 2019), which was part of the larger research program Urban Climate Under Change $[UC]^2$ (Scherer, Antretter, et al., 2019). The area directly around the measurement site consists of university buildings with a wide main thoroughfare (Strasse des 17. Juni) that runs from East to West through Berlin (see Figure 2.S-6 in the supplementary information). These occurred during two measurement campaigns which are henceforth referred to as the Summer Campaign (SC), which includes all 2017 measurements, and the Winter Campaign (WC) which includes the 2018 measurements, respectively (Figure 2-2).

For the field calibration the Zephyrs were co-located with the aforementioned reference instruments at the deployment site. The reference station for co-location was set up in an office on the 6th floor of the Mathematics building on the south facing façade that provided constant power for reference instrumentation and the Zephyrs, as well as space for air inlet tubing to be passed through the windows to the reference instrumentation (Figure 2-3). The Zephyrs and the air inlets were attached next to each other on the same railing outside the office. This ensured that all instruments were receiving the same parcels of air throughout the co-location. One Zephyr was co-located with reference instrumentation throughout the summer campaign (s71) and one throughout the winter campaign (s72). The reference station measurements were continuous throughout the co-locations and the experiments. The experiments took place from July 29th – August 28th and from September 20th – October 12th in 2017, and from January 27th – February 23rd in 2018. Five co-locations were conducted in total across the two campaigns. These took

place from July 18th – July 27th, August 29th – September 7th and October 14th – October 27th (all in 2017) during the SC and from January 13th – January 24th and February 23rd – March 8th (both in 2018) during the WC. All dates refer to time frames of the data presented, as the first and last days of deployment or co-location were not used owing to different start and end times of installation, as well as sensor warm up times. To compare sensor performance between s71 and s72, an intercomparison of available co-location raw data was conducted for the oxidizing MOS (Oxa) and ozone MOS (O3a). During all co-locations, the sensors had a linear relationship and an $R^2 > 0.95$ (Figures 2.S-4 and 2.S-5). In only one instance was this not the case (co-location 2, O3a), where the R^2 was 0.59 and a deviation from linearity was detected. This relationship was linear and normal in all other co-locations.

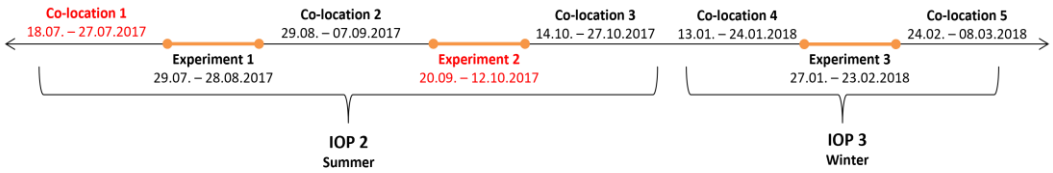

Figure 2-2 Timeline of SC and WC depicting the relationship between co-locations and experiments. Due to technical issues of individual instruments, data were unavailable for the segments marked in red.

This example focuses on Zephyr s71 during the SC and Zephyr s72 during the WC. For the sake of brevity, all graphs and tables included in this section pertain only to the former. Those relevant for the latter can be found in the Supplementary Information. Due to continuous co-location of these two sensors, the statistical models established using the 7-step method could be trained with co-location data and, atypically, assessed for their accuracy using reference concentrations during the entire experimental window. What follows is a thorough description of the application of the seven-step method for calibration.

In order to calibrate the Zephyrs, reference NO_2 and O_3 data, meteorological data, and raw data from the Zephyr sensors were used. Concentrations of NO_2 from the Teledyne T200 NOx Analyser and O_3 from the O3-2B Technologies instruments were used as response variables in the models. Ambient temperature (T_{amb}) and relative humidity (RH_{amb}) data as well as wind speed (WS) and direction (WD) data were tested as predictors in the statistical models. Four variables from the Zephyrs themselves were also tested in the statistical models as predictors: 1) Oxa, a measure of resistance from one MOS sensor used to detect oxidizing substances (in this case NO_2); 2) O3a, another measure of resistance from a MOS sensor that detects O_3; 3) a measure of temperature collected by the Zephyr (T_{int}); and 4) a measure of relative humidity collected by the Zephyr (RH_{int}). Finally, the binary time-of-day (ToD) variable was created to distinguish between night and day, as the chemistry of the analysed species changes significantly. Further reference data on other species would have been beneficial to this calibration, as the MOS do exhibit cross-sensitivities to other species, but resources were insufficient, and these data were not collected.

Figure 2-3 Set-up of the co-location of the prototype Zephyrs with reference instruments on the 6th floor of the Mathematics building. The grey units are the Zephyrs, and the two inlet tubes connect to the reference devices located inside the office.

2.3.4 Seven-step calibration of the Zephyrs

The temperature and relative humidity from the Zephyrs (T_{int} and RH_{int}) reflect the conditions within the sensor system and typically parallel ambient data, however, with an offset. These data are henceforth referred to as "internal" temperature and relative humidity. Throughout the example, both internal and ambient T and RH are used to assess their comparative influence on model accuracy. This was tested as ambient T and RH from reference instruments are not always available at experimental sites, whereas the internal T and RH of the Zephyrs are always available. The reference and meteorological data had an original time resolution of 1 minute whereas the Zephyr data was collected at a time resolution of 10 seconds. Analysis during the seven-step process was conducted using 5-minute averages except for outlier detection, which was done at original time resolution.

2.3.4.1 Step 1: Analyse raw data distribution

The distributions of the reference, meteorological, and Zephyr data were first compared between each co-location individually, both co-locations together, and the experimental deployment data of Experiment 1. The violin plots of ambient RH and T, NO_2, and O_3 for co-location 2 (Figure 2-4) show that the meteorological conditions and pollutant concentrations experienced were quite similar to those of the experiment. The ranges, median values, and the interquartile ranges are quite similar. This is further reflected by the similarity in distributions of both the Zephyr MOS sensor data (Oxa and O3a) and the reference instrument data between the 2nd co-location and the experiment.

By contrast, the distributions of the same variables for the 3rd co-location (Figure 2-4) are demonstrably different from the other co-location and the experimental data. The ambient temperature and relative humidity conditions were significantly cooler and wetter in the 3rd co-location than during the experiment and the NO_2 and O_3 concentrations were much higher and lower, respectively. Furthermore, the MOS sensor data in this co-location have a much different median and IQR than the experiment although the overall range is similar.

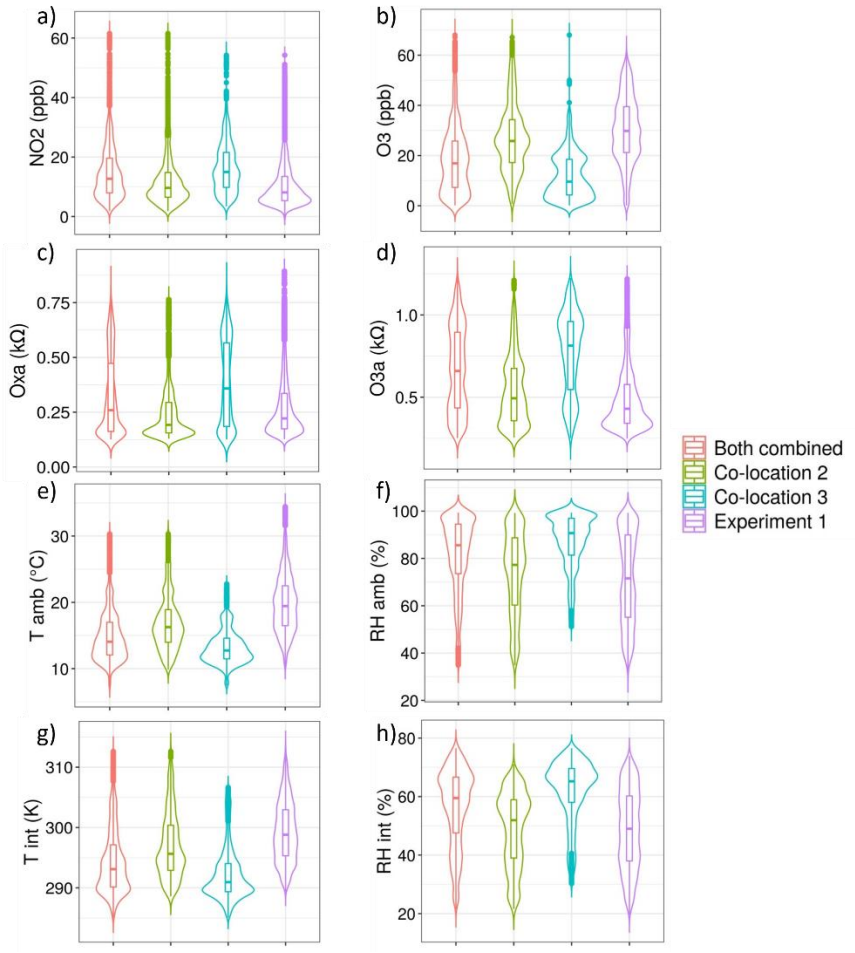

Figure 2-4 Violin plots of a) reference NO_2, b) reference O_3, c) Oxa, d) O3a, e) T_{amb}, f) RH_{amb}, g) T_{int}, and h) RH_{int} for co-location 2, co-location 3, both co-locations combined, and the experimental data.

With both locations combined (Figure 2-4), the distributions of all variables are representative of the experimental data, but with worse agreement than with co-location 2 alone. These results suggested that the 2nd co-location alone could be the best training set for the model building process. In order to further assess this hypothesis, co-location 2, co-location 3, and a combination of both were used in exporting final model predictions and evaluated using the atypical co-located experimental data as a "comparison" dataset.

2.3.4.2 *Step 2: Data cleaning*

Point outliers were determined using the developed outlier detection function. The threshold and running window parameters were optimized individually for each variable. This was done through visual assessment of points identified as outliers under various parameters, in order to determine if the designation was appropriate. For the reference NO_2 and O_3 data, using a z-score threshold of five and a running mean calculated with 120 data points (equivalent to two hours of data) was optimal for identifying true outliers. Using a lower threshold often falsely identified the extremes of normal data spikes as outliers. Figure 2-5 shows example outliers that were identified using the function described above for the reference data. The reference relative humidity and temperature data provided by the Free University had been pre-processed and as such no outliers were identified in those data.

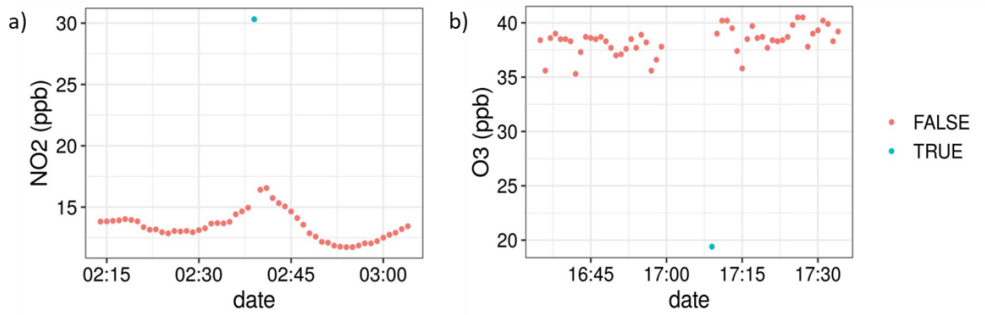

Figure 2-5 Examples of outliers detected on reference data using a z-test with running mean for the SC. A value of "TRUE" means the point was deemed an outlier by the z-test.

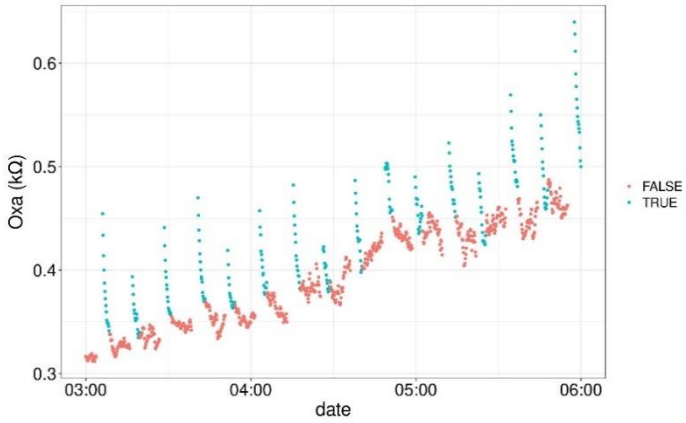

Figure 2-6 Example of outliers due to MOS sensor warm-up following a GSM connection of the Zephyrs. A value of "TRUE" indicates the point was included in the 2.5-minute MOS warm-up period.

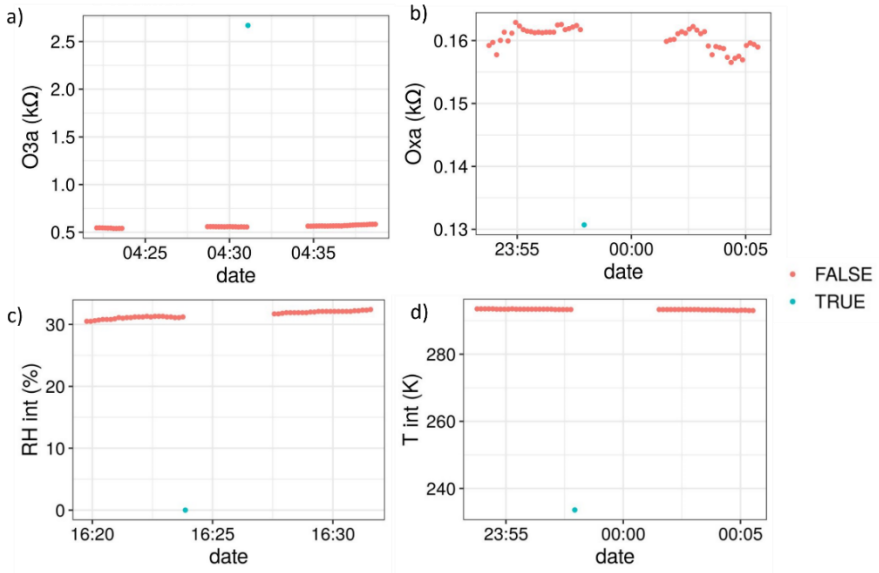

Figure 2-7 Examples of outliers detected on Zephyr s71 data using a z-test with running mean for the SC. A value of "TRUE" means the point was deemed an outlier.

As part of normal operation, the Zephyrs send logged data via GSM connection every 15 minutes to a database maintained by EarthSense. When this occurs, all metal oxide sensors in the device turn off. The MOS sensors by design, however, run quite hot and require a constant input of power to maintain their temperature. As can be seen in Figure 2-6, each time the MOS sensors turn off, they need to warm-up again before stabilizing.

The time series plots developed in Step 1 were key to identifying and addressing this issue. By developing a function in R that analyses the MOS sensor data patterns following time-gaps due to GSM connection, we developed a rule-of-thumb for identifying and removing these data. Analysis of this issue showed that the sensors required two and a half minutes to warm-up and return to normal functionality.

Once the time-gap anomalies were removed from the Zephyr data, the outlier detection function was applied to the four Zephyr variables in original time resolution. As can be seen in Figure 2-7, outliers were detected for the four Zephyr variables with a z-score threshold of five and a running mean of 360 data points (equivalent to one hour of data). It is likely that these anomalous data points all result from brief technical failures within the instrument.

2.3.4.3 *Step 3: Flagging the data*

Given that the data coverage from the 2nd co-location encompassed most of the experimental data, only a few points during the experiment were flagged for being out-of-bounds of the 2nd co-location set. As can be seen in Figure 2-8 a, only low NO_2 concentrations from the experimental set were flagged. The 3rd co-location experienced a narrower range of NO_2 concentrations, as can be seen in Figure 2-4 from Step 1. As such, more experimental data points of lower concentrations and some of high concentrations were flagged for this co-location (Figure 2-8 b). This shows the utility of comparing the results of Step 1 with the flags generated in Step 3.

Similarly, the 2nd co-location dataset received few flags, as most variables had comparable ranges to those of the experimental dataset. For example, only a few data points in which the internal Zephyr temperature dipped below ~289K were flagged (Figure 2-8 c). For the 3rd co-location, which was conducted in colder conditions in October, far more data points were flagged (Figure 2-8 d). This indicated that a larger portion of the 3rd co-location could be unsuitable for use in calibration. It also proved valuable for later analysis when analysing the final predicted concentrations of the model in Step 6.

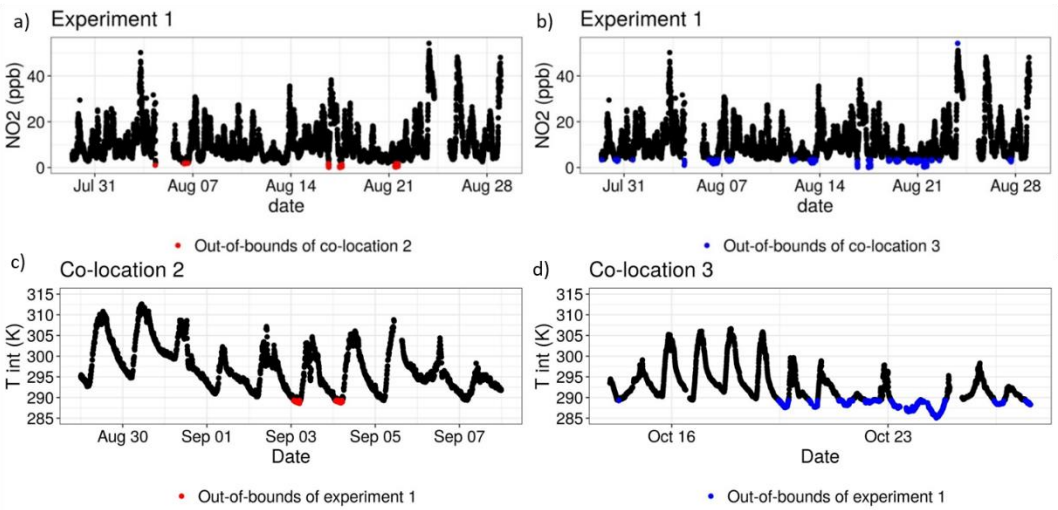

Figure 2-8 a) and b) Example time series plots of the experimental data with points out-of-bounds of the 2nd and 3rd co-location flagged, respectively. c) Time series plot of the 2nd co-location with points flagged for being out-of-bounds of the experimental data set. d) Time series plot of the 3rd co-location with points flagged for being out-of-bounds of the experimental data set.

2.3.4.4 Step 4: Model selection

The results of the model selection process can be seen in Tables 2-2 – 2-5. For readability, these tables reflect a later stage in the process, after which a wide range of other models had already been tested and excluded on the basis of AIC and accuracy metrics. A combination of these metrics was used to designate the "best" models in which RMSE and R^2 received a higher priority than AIC. The most accurate MLR model for predicting NO_2 was determined to be one in which Oxa, O3a, RH, and T were included as single terms with interactions between all variables. The relationship between NO_2 and Oxa was determined to be logarithmic, whereas the relationship to T was determined to be inverse. This is in line with what would be expected in urban environments, as T can be seen as a proxy for insolation, which causes the photolysis of NO_2. For O_3 the most accurate MLR model had Oxa, O3a, RH, and T included as single terms with interactions. The relationship between Oxa and O_3 was also determined to be logarithmic. For both NO_2 and O_3, MLR models using ambient T and RH were consistently more accurate than those using internal T and RH.

Table 2-2 Results of the MLR model selection process for NO_2. The most accurate model is in bold font. RMSE and MAE are in units of ppb.

Formula	R^2	RMSE	MAE	AIC
NO_2~ log(Oxa) * O3a * RH_{amb} * (1/T_{amb})	**0.82**	**3.90**	**3.02**	**26527.04**
NO_2~ log(Oxa) + O3a + RH_{amb} + (1/T_{amb})	0.83	4.11	3.20	27027.58
NO_2~ log(Oxa) * O3a * RH_{amb} * T_{amb}	0.77	4.26	3.57	25896.72
NO_2~ log(Oxa) + O3a + RH_{amb} + T_{amb}	0.84	4.15	3.33	26602.66
NO_2~ log(Oxa) * O3a * RH_{int} * (1/ T_{int})	0.77	4.61	3.65	27552.32
NO_2~ log(Oxa) + O3a + RH_{int} + (1/ T_{int})	0.81	4.59	3.89	28174.42
NO_2~ log(Oxa) * O3a * RH_{int} * T_{int}	0.78	5.78	4.49	26621.64
NO_2~ log(Oxa) + O3a + RH_{int} + T_{int}	0.80	4.62	3.92	27936.49

Table 2-3 Results of the MLR model selection process for O_3. The most accurate model is in bold font. RMSE and MAE are in units of ppb.

Formula	R^2	RMSE	MAE	AIC
O_3 ~ log(Oxa) * O3a * RH_{amb} * T_{amb}	0.91	3.38	2.59	23842.89
O_3 ~ log(Oxa) + O3a + RH_{amb} + T_{amb}	0.94	3.05	2.46	25023.60
O_3 ~ log(Oxa) * O3a * (1/RH_{amb}) * T_{amb}	**0.92**	**2.91**	**2.29**	**24088.18**
O_3 ~ log(Oxa) + O3a + (1/RH_{amb}) + T_{amb}	0.94	3.20	2.44	25077.30
O_3 ~ log(Oxa) * O3a * RH_{int} * T_{int}	0.81	4.06	2.80	26173.68
O_3 ~ log(Oxa) + O3a + RH_{int} + T_{int}	0.92	3.67	2.69	28054.23
O_3 ~ log(Oxa) * O3a * (1/ RH_{int}) * T_{int}	0.82	4.30	3.03	26374.23
O_3 ~ log(Oxa) + O3a + (1/ RH_{int}) + T_{int}	0.91	3.67	2.78	28178.23

Table 2-4 Results of the RF model selection process for NO_2. Min.node.size and split rule were optimized in a previous step not shown here for brevity and are therefore constant. RMSE and MAE are in units of ppb.

Formula	mtry	min.node.size	Split rule	R^2	RMSE	MAE
NO_2 ~ Oxa + O3a + RH_{amb} + T_{amb} + ToD + WD + WS	7	5	extratrees	0.70	4.49	3.62
NO_2 ~ Oxa + O3a + RH_{amb} + T_{amb} + ToD + WD	6	5	extratrees	0.71	4.58	3.49
NO_2 ~ Oxa + O3a + RH_{amb} + T_{amb} + ToD	5	5	extratrees	0.75	4.43	3.51
NO_2 ~ Oxa + O3a + RH_{amb} + T_{amb}	**4**	**5**	**extratrees**	**0.76**	**4.01**	**3.25**
NO_2 ~ Oxa + O3a + RH_{amb}	3	5	extratrees	0.74	4.08	3.26
NO_2 ~ Oxa + O3a + T_{amb}	2	5	extratrees	0.76	4.64	3.92
NO_2 ~ Oxa + O3a	2	5	extratrees	0.70	4.44	3.38
NO_2 ~ Oxa + O3a + RH_{int} + T_{int} + ToD + WD + WS	7	5	extratrees	0.60	5.06	3.97
NO_2 ~ Oxa + O3a + RH_{int} + T_{int} + ToD + WD	6	5	extratrees	0.58	5.20	4.09
NO_2 ~ Oxa + O3a + RH_{int} + T_{int} + ToD	5	5	extratrees	0.58	5.34	4.00
NO_2 ~ Oxa + O3a + RH_{int} + T_{int}	3	5	extratrees	0.63	5.12	3.93
NO_2 ~ Oxa + O3a + RH_{int}	2	5	extratrees	0.65	5.59	4.37
NO_2 ~ Oxa + O3a + T_{int}	2	5	extratrees	0.70	4.86	3.79

Table 2-5 Results of the RF model selection process for O_3. Min.node.size and split rule were optimized in a previous step not shown here for brevity and are therefore constant. RMSE and MAE are in units of ppb.

Formula	mtry	min.node.size	Split rule	R^2	RMSE	MAE
O_3 ~ Oxa + O3a + RH_{amb} + T_{amb} + ToD + WD + WS	4	5	extratrees	0.92	3.37	2.42
O_3 ~ Oxa + O3a + RH_{amb} + T_{amb} + ToD + WD	4	5	extratrees	0.90	3.20	2.53
O_3 ~ Oxa + O3a + RH_{amb} + T_{amb} + ToD	2	5	extratrees	0.90	3.10	2.45
O_3 ~ Oxa + O3a + RH_{amb} + T_{amb}	2	5	extratrees	0.92	3.39	2.52
O_3 ~ Oxa + O3a + RH_{amb}	2	5	extratrees	0.93	3.71	2.62
O_3 ~ Oxa + O3a + T_{amb}	**2**	**5**	**extratrees**	**0.93**	**2.95**	**2.36**
O_3 ~ Oxa + O3a	2	5	extratrees	0.90	4.09	2.87
O_3 ~ Oxa + O3a + RH_{int} + T_{int} + ToD + WD + WS	4	5	extratrees	0.90	3.44	2.46
O_3 ~ Oxa + O3a + RH_{int} + T_{int} + ToD + WD	4	5	extratrees	0.90	3.60	2.46
O_3 ~ Oxa + O3a + RH_{int} + T_{int} + ToD	2	5	extratrees	0.91	3.64	2.42
O_3 ~ Oxa + O3a + RH_{int} + T_{int}	2	5	extratrees	0.87	3.72	2.65
O_3 ~ Oxa + O3a + RH_{int}	2	5	extratrees	0.87	3.92	2.77
O_3 ~ Oxa + O3a + T_{int}	2	5	extratrees	0.85	3.78	2.66

For random forest, the most accurate NO_2 model was determined to be one that included Oxa, O3a, ambient RH, and ambient R. The optimal mtry parameter was determined to be 4, with a minimum node size of 5. For predicting O_3 the results were similar to those of NO_2, except that ambient T replaced ambient RH. For both NO_2 and O_3 the use of ambient T and RH produced more accurate models. Overall, the random forest models performed very similarly to the MLR models, with only slight differences in R^2 and RMSE.

2.3.4.5 Step 5: Model validation

For MLR and RF, the R^2 and RMSE for each block were saved and plotted (Figure 2-9 a-d). As can be seen, the models using ambient T and RH for both O_3 and NO_2 remained relatively stable across all blocks. They consistently have a higher R^2 and a lower RMSE than the models trained with internal T and RH, for both NO_2 and for O_3. Conversely, the models trained with internal T and RH are much more volatile in terms of R^2 and RMSE, for both NO_2 and O_3. In addition, blocks 11, 12, and 13 show a marked decrease in R^2 and increase in RMSE across all models with internal T and RH. This trend was true for several models tested at this step, indicating that the internal T and RH were less stable for these blocks. Generally, the differences in RMSE between ambient and internal T and RH were more pronounced for NO_2 than for O_3. This is true across most blocks and indicates that the final concentrations should be predicted using ambient T and RH data instead of internal. Tables 2-6 and 2-7 show the median R^2 and RMSE for all selected models for NO_2 and O_3, respectively. They reveal MLR and RF using ambient T and RH are similarly accurate at predicting NO_2 and O_3. The differences in accuracy are more pronounced for the models using internal T and RH.

Of all predictors included in the RF models, the MOS variable O3a had the highest VI for predicting both O_3 and NO_2 (Figure 2-10 a-d). The MOS variable Oxa was also of relative importance, usually as the 2nd most important variable, with the exception of the O_3 models for which temperature (internal or ambient) was sometimes the 2nd most important variable. Results from these graphs indicate that all variables should remain in the RF models.

Table 2-6 Median R^2 and RMSE across all test blocks of the best MLR and RF models using internal and ambient T and RH for NO_2. RMSE and MAE are reported in units of ppb.

	NO_2	**Median R^2**	**Median RMSE**	**Median MAE**	
MLR	$NO_2 \sim \log(Oxa) * O3a * RH_{amb} * (1/T_{amb})$	0.82	4.35	3.54	*Model 1a*
	$NO_2 \sim \log(Oxa) * O3a * RH_{int} * (1/T_{int})$	0.67	6.12	4.10	*Model 1b*
RF	$NO_2 \sim Oxa + O3a + RH_{amb} + T_{amb}$	0.75	4.88	3.90	*Model 2a*
	$NO_2 \sim Oxa + O3a + T_{int}$	0.72	5.29	3.89	*Model 2b*

Table 2-7 Median R^2 and RMSE across all test blocks of the best MLR and RF models using internal and ambient T and RH for O_3. RMSE and MAE are reported in units of ppb.

	O_3	**Median R^2**	**Median RMSE**	**Median MAE**	
MLR	$O_3 \sim \log(Oxa) * O3a * (1/RH_{amb}) * T_{amb}$	0.91	3.83	2.86	*Model 3a*
	$O_3 \sim \log(Oxa) * O3a * (1/RH_{int}) * T_{int}$	0.82	4.81	3.79	*Model 3b*
RF	$O_3 \sim Oxa + O3a + T_{amb}$	0.90	3.77	3.00	*Model 4a*
	$O_3 \sim Oxa + O3a + T_{int} + RH_{int} + ToD$	0.86	5.20	4.10	*Model 4b*

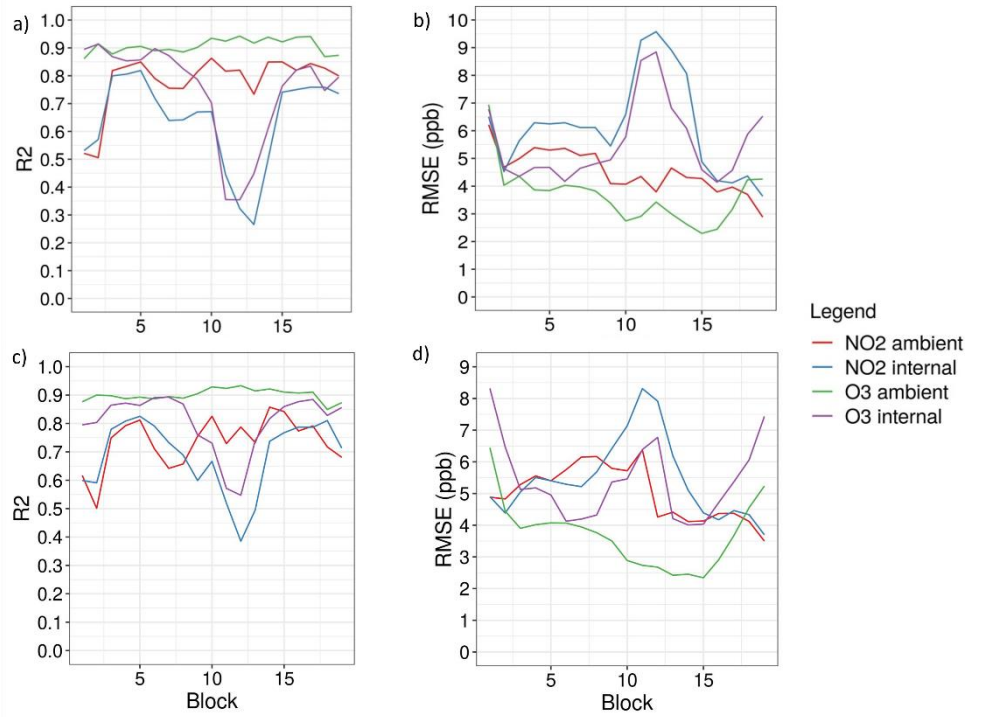

Figure 2-9 a) R^2 and b) RMSE over the 19 test blocks for the MLR models (1a, 1b, 3a, 3b), respectively. c) R^2 and d) RMSE over the 19 blocks for the RF models (2a, 2b, 4a, 4b), respectively.

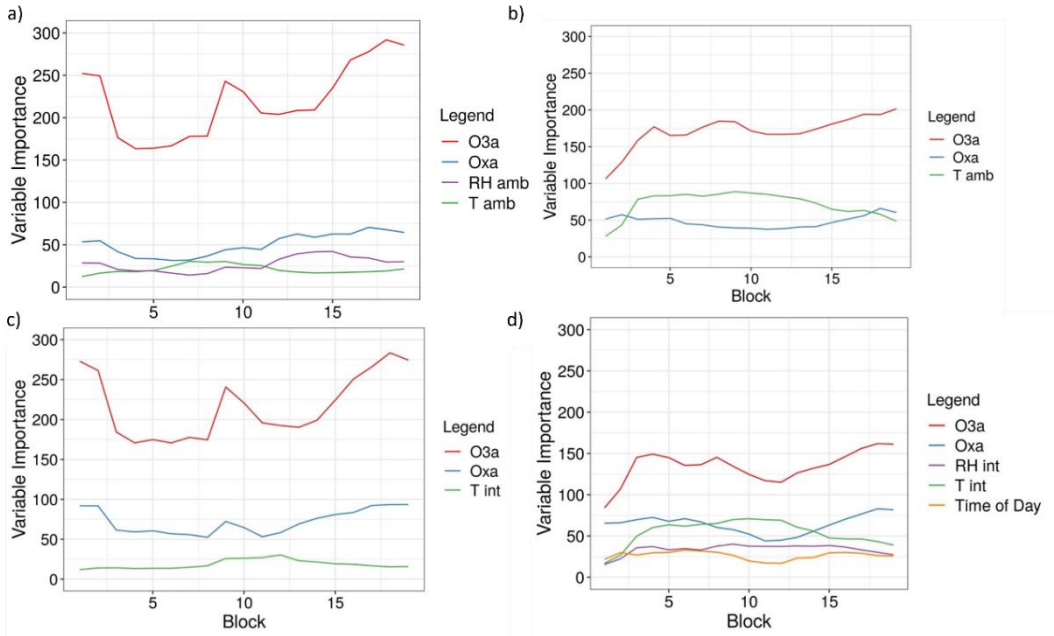

Figure 2-10 Variable importance over the 19 test blocks of a) model 2a, b) model 4a, c) model 2b, and d) model 4b.

2.3.4.6 *Step 6: Predicting final concentrations*

Final concentrations predicted for NO_2 and O_3 using the MLR and RF models with both ambient and internal T and RH can be seen in Figure 2-11. While the results indicated that ambient T and RH should be used, both are included here for further analysis beyond the seven-step methodology. For NO_2, the MLR models predict a much narrower range of concentrations and occasionally predict negative concentrations (Figure 2-11 a). The RF models tend to predict higher concentrations than MLR, have a wider range, and don't

predict negative concentrations (Figure 2-11 b). For O_3, the differences between MLR and RF are less pronounced, with both capturing the diurnal cycle well (Figure 2-11 c-d). In all figures it can be seen that models using ambient T and RH consistently predict higher concentrations than those using internal T and RH. This indicates that there is a difference between predictions using Zephyr internal versus reference temperature and relative humidity sensors.

2.3.4.7 *Step 7: Calculating predictive error*

As can be seen in Figure 2-12, smearing the reference data had minimal impact on the predictive accuracy of all models. This indicates that the uncertainty of the reference instruments did not impact the predictive accuracy of the models and can therefore in this case be ignored as a potential interference. Overall predictive error was then calculated as the reference error plus median MAE of each model across all blocks from the model validation step. The T-200 NOx instrument has a measurement uncertainty of 0.5% of the measurement above 50 ppb or an uncertainty of 0.2 ppb below 50 ppb. For the Tech 2B Ozone Monitor, the uncertainty was the larger between 2% of the measurement or 1 ppb. This can be seen in Figure 2-13, which depicts the MLR and RF predicted concentrations for Experiment 1 with shaded regions representing the uncertainty. The uncertainty of the RF and MLR models was fairly similar, but was higher for NO_2 than for O_3. This reflects the findings from Steps 4-6 in which O_3 was predicted more accurately than NO_2 by both models. The standard error for MLR models was found to not reflect the realistic accuracy of the predicted concentrations in relation to actual concentrations, as it was found to be very low. The RF models calculated a more appropriate measure of standard error using the infinitesimal jackknife method (Wager, Hastie, & Efron, 2014), but for consistency with the MLR models, this measure was not used. The accuracy of the final models in predicting on experimental data for which reference concentrations are not available for comparison is then best reflected by combining the uncertainty of the reference instruments with the median MAE of the test blocks during Step 5 (Tables 2-6 and 2-7).

2.3.5 Extra validation step

To further test the impact of using more representative training datasets, the final models identified in Steps 4 and 5 were trained with each co-location individually as well as with both combined. The predictive accuracy of these separate models was then compared using the experimental dataset for which reference NO_2 and O_3 measurements were available, as Zephyr s71 was co-located throughout the experiment. Additionally, these datasets were also tested with data points flagged in Step 3 removed to understand further influences on model accuracy. This extra validation allowed for better evaluation of the performance in predicting experimental concentrations of the MLR and RF models selected with the seven-step method. This is, however, atypical for field studies, as these sensor systems are intended to be deployed independently of reference instrumentation.

Table 2-8 shows the results of training these various models for NO_2. The most accurate model at predicting experimental concentrations was the RF model using internal T and trained with data only from co-location 2. The same model trained with all available co-location data was slightly more inaccurate. Co-location 3 was the least accurate of the training subsets, reiterating findings from Step 1. For the MLR models, this dip in accuracy when using exclusively co-location 3 as the training set was most pronounced, as can be seen in Table 2-8 and Figures 2-14 g-h. When filtering out flagged data points, most NO_2 models improved slightly in predictive accuracy. This was most pronounced for those

using co-location 3 as a training set, which improved substantially in terms of R^2. This alludes largely to the impact of seasonal changes on co-location 3, which experienced different meteorological and pollution conditions than were present during experiment 1. While results show that this co-location was not useful for accurate prediction, it is likely that it would have been more relevant for prediction on experiment 2, during which the environmental conditions were more comparable. Similarly, co-location 1 would likely have been more valuable for prediction with experiment 1 than with experiment 2. However, due to the loss of data from s71, this could not be assessed more closely in this study.

For O_3, the most accurate model was the RF model using internal T and RH and trained exclusively using data from co-location 2, though the MLR internal model for the same co-location was of comparable accuracy. The RMSE for this model was substantially lower than the one trained using ambient T and RH. With the MLR models, this difference in predictive accuracy between models trained with internal and ambient T and RH was much greater, again favouring the internal models. Co-location 3 was highly inaccurate at predicting experimental data, further reiterating findings from Step 1 that indicated the unsuitability of this co-location for use in predicting final concentrations. Figures 2-15 e-f clearly depict the boundaries for predictions with RF models when the training data are unsuitable, as is the case with co-location 3. This is a fundamental flaw of RF models as they cannot predict outside the bounds of the co-location data they are trained with. Filtering out the points flagged in Step 3 did not improve the predictive accuracy of models trained exclusively with co-location 2, but it substantially improved those trained with co-location 3, especially those using internal T and RH.

Table 2-8 Results of RF and MLR models for NO_2 trained with co-location 2, co-location 3, or a combination of both when tested on the comparison experimental dataset. In the lower half of the table, the models are trained with the same datasets but are tested on the experimental dataset with data points outside the ranges of each training dataset filtered out.

	NO_2	**Co-location 2**		**Co-location 3**		**Both co-locations**	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
MLR	NO_2 ~ log(Oxa) * O3a * RH_{amb} *(1/T_{amb})	0.66	5.49	0.22	12.08	0.61	5.66
	NO_2 ~ log(Oxa) * O3a * RH_{int} * (1/T_{int})	0.66	6.41	0.57	10.99	0.67	5.55
RF	NO_2 ~ Oxa + O3a + RH_{amb} + T_{amb}	0.67	5.23	0.41	6.64	0.66	4.97
	NO_2 ~ Oxa + O3a + T_{int}	**0.73**	**4.44**	0.61	5.60	0.68	4.87
	NO_2 – filtered						
MLR	NO_2 ~ log(Oxa) * O3a * RH_{amb} *(1/T_{amb})	0.63	5.53	0.45	12.88	0.62	5.62
	NO_2 ~ log(Oxa) * O3a * RH_{int} * (1/T_{int})	0.63	6.49	0.64	11.17	0.69	5.54
RF	NO_2 ~ Oxa + O3a + RH_{amb} + T_{amb}	0.65	5.09	0.66	5.64	0.56	4.90
	NO_2 ~ Oxa + O3a + T_{int}	0.71	4.38	0.65	6.01	0.68	4.85

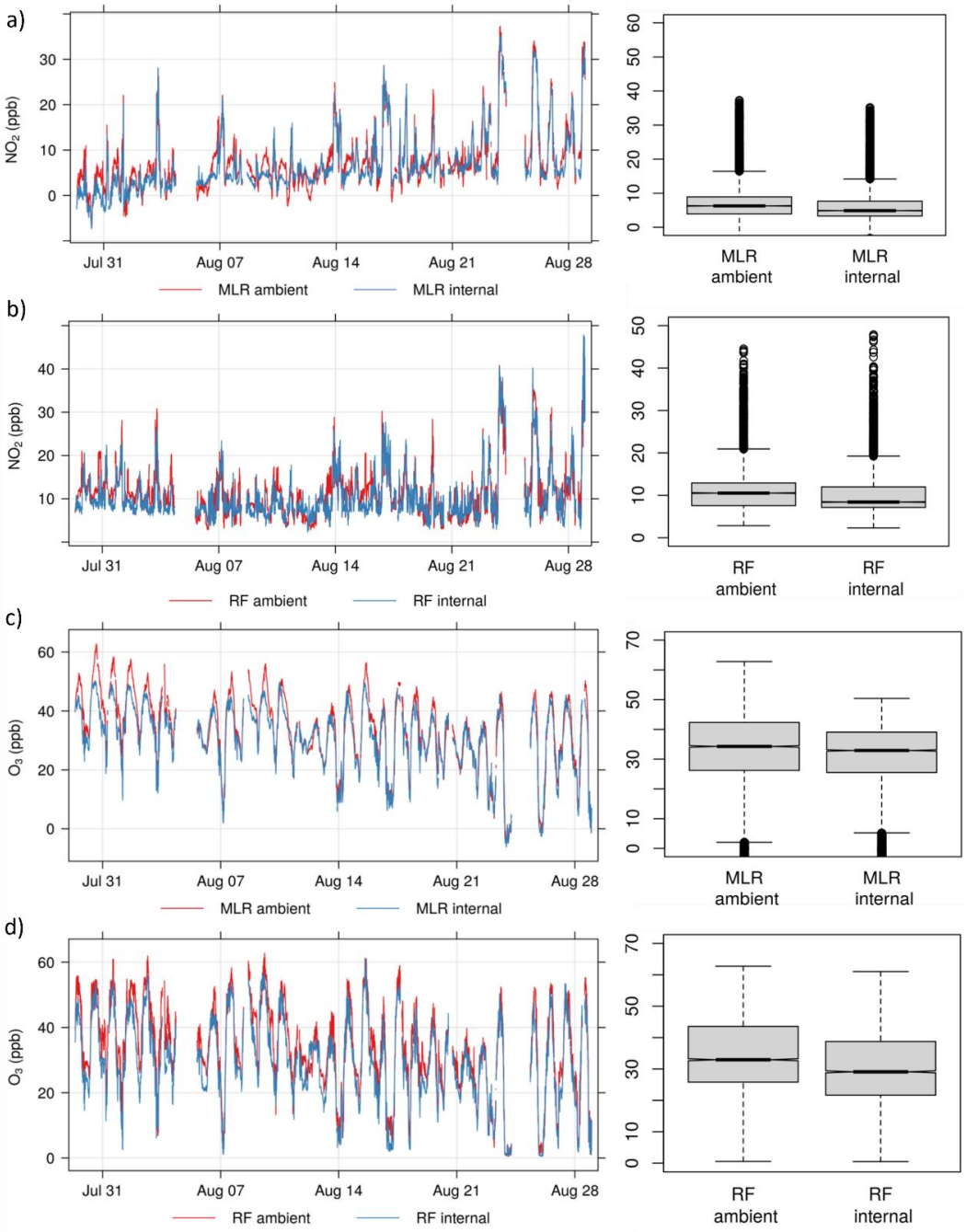

Figure 2-11 Time series plots and boxplots for Experiment 1 of a) predicted NO_2 concentrations using the MLR model, b) predicted NO_2 concentrations using the RF model, c) predicted O_3 concentrations using the MLR model, d) predicted O_3 concentrations using the RF model. 'Ambient' and 'internal' refer to the use of ambient or internal T and RH data in each model.

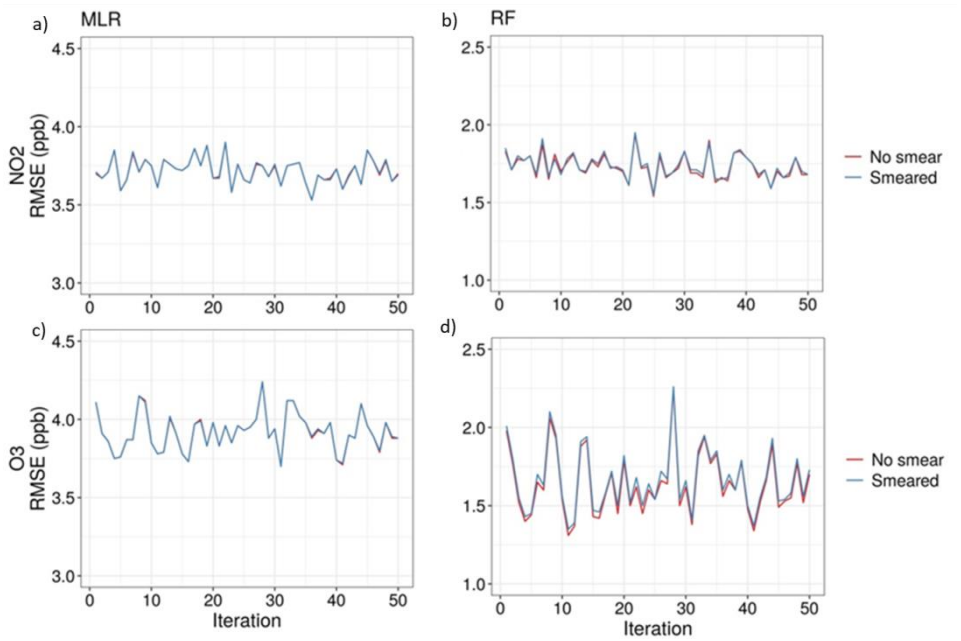

Figure 2-12 RMSE of models trained using smeared reference measurements versus actual reference measurements for a) NO_2 with MLR, b) NO_2 with RF, c) O_3 with MLR, and d) O_3 with RF.

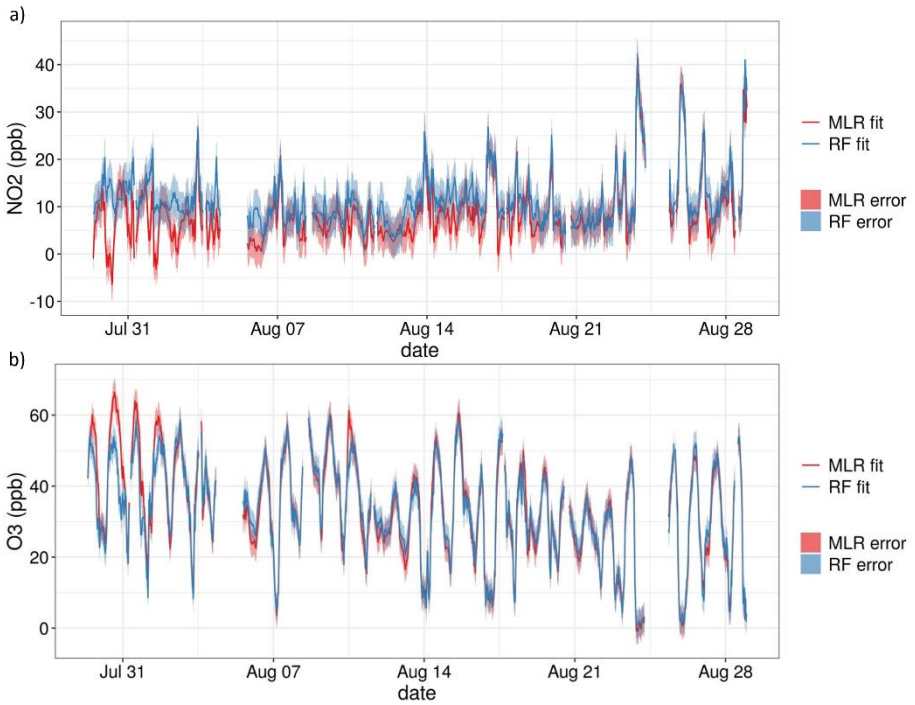

Figure 2-13 Time series plots of both MLR and RF predictions for Experiment 1 including the measurement uncertainty as shaded regions for a) NO_2 and b) O_3. Data were averaged to 30-minute resolution.

Table 2-9 Results of RF and MLR models for O_3 trained with co-location 2, co-location 3, or a combination of both when tested on the comparison experimental dataset. In the lower half of the table, the models are trained with the same datasets but are tested on the experimental dataset with data points outside the ranges of each training dataset filtered out.

	O_3	**Co-location 2**		**Co-location 3**		**Both co-locations**	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
MLR	O_3 ~ log(Oxa) * O3a * $(1/RH_{amb})$ * T_{amb}	0.86	7.00	0.86	5.12	0.88	6.06
	O_3 ~ log(Oxa) * O3a * $(1/RH_{int})$ * T_{int}	0.94	3.37	0.16	17.20	0.91	3.94
RF	O_3 ~ Oxa + O3a + T_{amb}	0.91	5.14	0.73	7.77	0.91	5.14
	O_3 ~ Oxa + O3a + T_{int} + RH_{int} + ToD	**0.94**	**3.31**	0.67	9.95	0.92	3.80
	O_3 – filtered						
MLR	O_3 ~ log(Oxa) * O3a * $(1/RH_{amb})$ * T_{amb}	0.85	6.78	0.85	4.13	0.87	5.97
	O_3 ~ log(Oxa) * O3a * $(1/RH_{int})$ * T_{int}	0.93	3.33	0.52	9.24	0.91	3.90
RF	O_3 ~ Oxa + O3a + T_{amb}	0.91	5.18	0.77	5.13	0.90	5.15
	O_3 ~ Oxa + O3a + T_{int} + RH_{int} + ToD	0.93	3.30	0.65	7.53	0.91	3.82

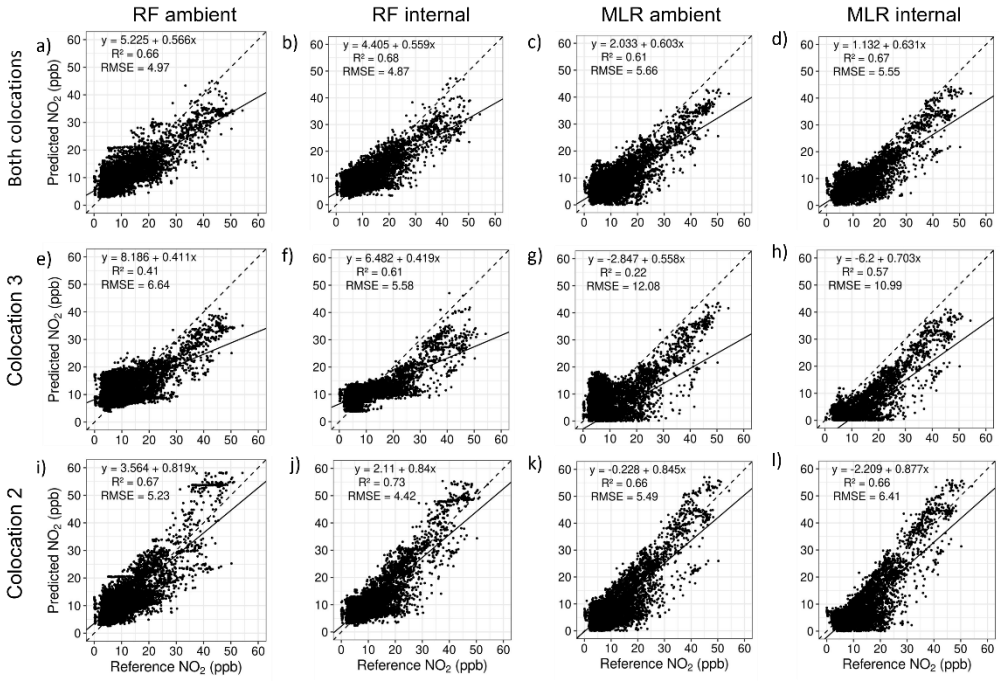

Figure 2-14 Scatter plots of predicted NO_2 versus reference NO_2 concentrations for the experimental data using MLR and RF models trained with co-location 2 (i-l), co-location 3 (e-h), and both combined (a-d). All concentrations are reported in ppb.

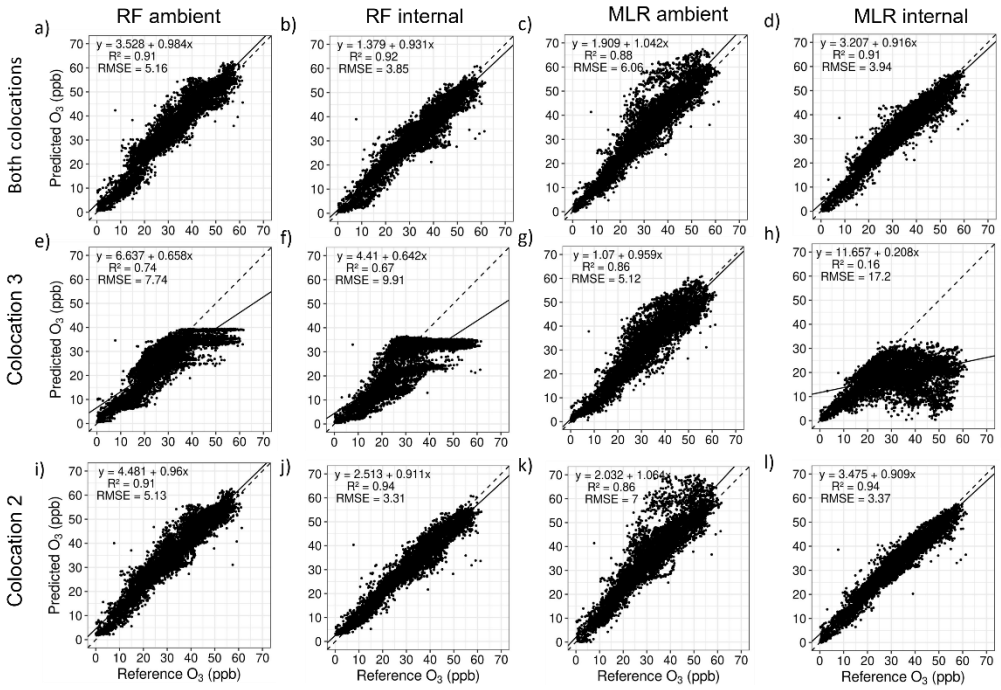

Figure 2-15 Scatter plots of predicted O_3 versus reference O_3 concentrations for the experimental data using MLR and RF models trained with co-location 2 (i-l), co-location 3 (e-h), and both combined (a-d). All concentrations are reported in ppb.

2.4 Discussion

The results of this study have several implications for the field of low-cost sensors. In line with other research, this study found that MLR and RF were similarly accurate in predicting experimental concentrations of NO_2 and O_3 (Karagulian et al., 2019), though the differences in accuracy between MLR and RF were more pronounced for O_3 than for NO_2. In fact, it was found that RF was the better predictor of both O_3 and NO_2 concentrations when evaluated with the longer experimental data set, albeit only slightly. This contrasts with findings from the model selection and validation process, as the MLR models were consistently more accurate at predicting on subsets of the co-location data. What this indicates is that models found to be more accurate during "calibration" may have differing model performance when assessed with a "comparison" dataset, in this case the experimental dataset that was co-located throughout for one sensor. This is a result that has been found previously, where the R^2 is lower for comparison datasets than for calibration (Karagulian et al., 2019). If RF, MLR, or other ML techniques are selected for their accuracy when predicting on calibration data and are not tested on comparison data, it may well be that the performance does not hold for new experimental data. Given the similarity between RF and MLR in predicting NO_2 and O_3 found in this study, as well as in the literature, either method can be used. However, as MLR is simpler to implement than most ML techniques, has fewer parameters that need to be optimized, and the model calculations are well understood, unlike the black-box calculations of RF and most ML techniques, this should be the preferred option to achieve greater model transparency and control.

Further important to the proper evaluation of model accuracy is the reporting of multiple metrics such as RMSE and MAE, in addition to R^2. It is quite clear from Tables 2-8 and 2-9 that R^2 is not the best metric with which to measure predictive accuracy of calibration models. Models trained with co-location 3 exclusively to predict O_3, for example, had an

R^2 greater than 0.70, which is acceptable agreement. Those same models, however, had an RMSE of >7 ppb, which is much more inaccurate than an R^2 of 0.70 alone would reveal. As another example, the same models trained exclusively with co-location 2 for O_3 (Table 2-9) had an R^2 between 0.86 and 0.94, but had a wide range of RMSE between 3.30-7.00 ppb. It is therefore crucial that multiple performance metrics are used to evaluate calibration models before final decisions are made on their suitability. At a minimum, R^2 and RMSE should be reported.

Multicollinearity is an issue common not only to MLR, but also to small sensor systems, which often have multiple LCSs measuring the same or similar species with heavily auto-correlated data. While uncommonly addressed in the literature, except for a few studies mentioning its influence on MLR models (Bigi et al., 2018; Hagan et al., 2018; Masiol et al., 2018), the solution, as presented in Steps 4 and 5, is relatively straightforward. To ensure that the predictor variables included in the final model are, in fact, explanatory, the model should be repeatedly validated using bootstrapped samples. To deal with autocorrelation, this validation should be done using continuous blocks and not with random sampling. Including these steps in the model-building process is simple and should be considered best practice.

Further underlining the importance of repeated validation is the variation in results when using ambient or internal T and RH. While the inclusion of ambient meteorological data led to more accurate models during calibration, this did not hold for the comparison dataset. Instead, for the prediction of both NO_2 and O_3, it was internal T and RH data that led to more accurate prediction. This indicates that for the prediction of NO_2 and O_3 concentrations with EarthSense Zephyrs, not only are the internal T and RH sensors acceptable for use in predictive models, but they are likely more representative of normal operating conditions. Given that the MOS sensors radiate large amounts of heat, the conditions inside the Zephyrs are significantly different than ambient conditions. As such, the internal T and RH sensors likely better represent the exact environmental conditions under which species are adsorbing to the MOS sensors. However, given that models using ambient data were more accurate during the validation step and significant differences between predictions of models trained with internal vs ambient T and RH were identified, these results require closer inspection, which should be the subject of future research.

The final results also reveal the value of pre-processing the data in Steps 1-3. It became clear by looking at the distribution of the co-location datasets in Step 1 that co-location 3 might be unsuitable for use in predicting the experimental concentrations. These data were then flagged in Step 3. While the models trained exclusively with co-location 3 were substantially less accurate than those using data from co-location 2, their accuracy increased when flagged experimental data points outside the range were removed. In essence, the 3rd co-location was useful for predicting on experimental data within its range of conditions, but very inaccurate for those outside of that range. Co-location 2, on the other hand, was identified as being well-suited for prediction in Step 1 and received few flagged points in Step 3. Final results indicate that MLR and RF models trained with co-location 2 perform better than those trained with co-location 3, for both NO_2 and O_3. Combining the two co-locations did not improve the predictive accuracy for NO_2 or O_3, when compared with the more-suitable 2nd co-location (Tables 2-8 and 2-9). As such, training calibration models with co-location 2 exclusively would have been correctly justified using evidence from Step 1. What is evident from this analysis is that ensuring

quality of training data used in calibration is crucial to accurate prediction. Incorporating quality control into the calibration methodology is therefore an important best practice.

Finally, LCS data should be reported with associated error values. While we discussed RMSE in the context of model fit and validation, as well as a method for evaluating whether reference instrument accuracy affects the model output, error values should be reported not just in the assessment of the LCSs themselves, but also with some form of representative error associated to the reported concentration data. Our recommendation is to combine the uncertainty of the reference instruments with the median MAE across blocks from the model validation step. As can be seen in Tables 2-8 and 2-9, the RMSE of predictions tested with the comparison experimental dataset are quite similar to the median RMSE values in Tables 2-6 and 2-7. This indicates that using median error from the model validation step is quite representative of the LCS uncertainty. However, over longer measurement campaigns, this should be repeatedly tested and validated with additional co-location training sets, so as to account for sensor drift, deteriorating functionality, and varying meteorological conditions.

2.5 Conclusions

While many details of this methodology are already well-known, they are often overlooked or go unreported in published literature. In most cases not all aspects are included. As a result, many studies assessing pairwise calibration methodologies for low-cost sensors cannot be compared. In the absence of calibration standards for these technologies in a field that continues to diversify and grow, researchers must start to consolidate around an agreed-upon set of best practices. This study has highlighted several of them. First, details on model selection, validation, and tuning must be reported if researchers are to be able to effectively compare results across studies. If models are not rigorously tested for suitability using standardized methods, especially with “black-box” machine-learning techniques, then their comparison will remain challenging at best. Second, models should be validated not only on the calibration dataset, but also on a separate comparison dataset, if possible. All validation should be done using R^2 and RMSE, at a minimum. This will provide greater insight into the suitability of selected models for prediction on experimental data as well as better comparability across studies. Third, pre-processing the data, including visual inspection, outlier removal, and data-flagging are an integral part of an effective calibration methodology. Understanding the quality and distribution of available data is important to identifying problems and solutions encountered during calibration.

Last, it is clear that a standardized methodology for the calibration of low-cost sensors is needed if they are to be incorporated into air quality monitoring programs and contribute new insights to the field of atmospheric chemistry. This seven-step methodology seeks to fill a gap in the literature up until now left largely unreported. In addition, this methodology, complete with relevant R code, is the first to be completely transparent and open-access. This is a valuable contribution to a young, but rapidly growing body of literature surrounding low-cost sensors. With this work, we hope to begin pulling back the curtains on the black box of sensor calibration.

Code availability

All relevant code for this study can be found in this open-access Zenodo repository: https://doi.org/10.5281/zenodo.4317521

Data availability

All relevant data for this study can be found in this open-access Zenodo repository: https://doi.org/10.5281/zenodo.4309853

Acknowledgements

The authors would like to thank Mark Lawrence for his support of the research work, Martin Schultz, Clara Betancourt, and Phil Peterson for their discussions on the subject, Achim Holtmann for his support of the meteorological data collection, and Roland Leigh, Jordan White and the EarthSense team for their collaboration and support. The research of EvS, SS, AC, GV is supported by RIFS, with financial support provided by the Federal Ministry of Education and Research of Germany (BMBF) and the Ministry for Science, Research and Culture of the State of Brandenburg (MWFK). Much of the research was carried out as part of the research program 'Urban Climate Under Change' [UC]2, contributing to 'Research for Sustainable Development' (FONA; www.fona.de), within the joint-research project 'Three-dimensional observation of atmospheric processes in cities (3DO)', funded by German Federal Ministry of Education and Research (BMBF) under grant 01LP1602E.

Author Contributions

SS developed the method, conducted the analysis, and wrote the paper. ST supported the method development and contributed to revisions and editing. GV conducted the measurements, contributed to the concept development for the measurement campaigns, and contributed to editing. AC contributed to the method development and data analysis, as well as writing and editing. IL, FM collected and provided processed meteorological data in the context of the measurement campaigns, as well as writing and editing. DK and RW did the comparison of the reference instruments. EvS developed the concept and contributed to the measurements, method development, writing, editing and overall coordination. All authors gave final approval for publication.

Competing interests

The authors declare that they have no conflict of interest.

2.S Supplemental information

2.S.1 Comparison of reference instruments with FZ Jülich

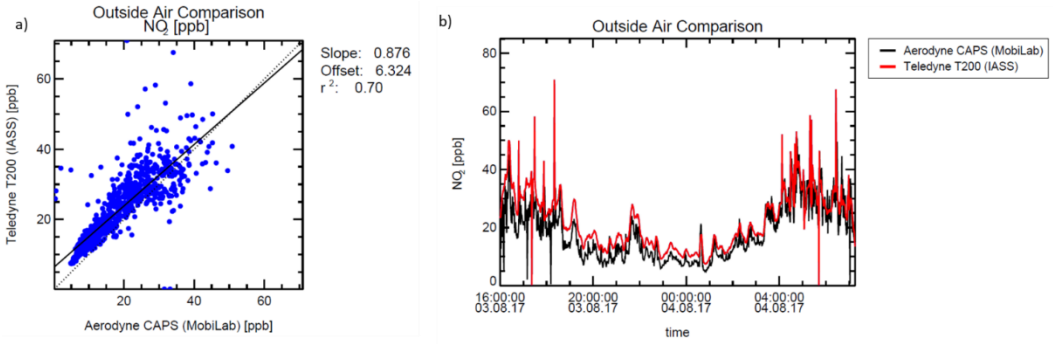

Figure 2.S-1 Intercomparison of the T-200 NO_2 concentrations with the CAPS (Aerodyne, U.S.A.) from Forschungszentrum Jülich. The data are presented as: a) a scatter plot to establish a correction factor, and b) a time series plot showing the comparison between the two instruments and the range selected for the correction factor.

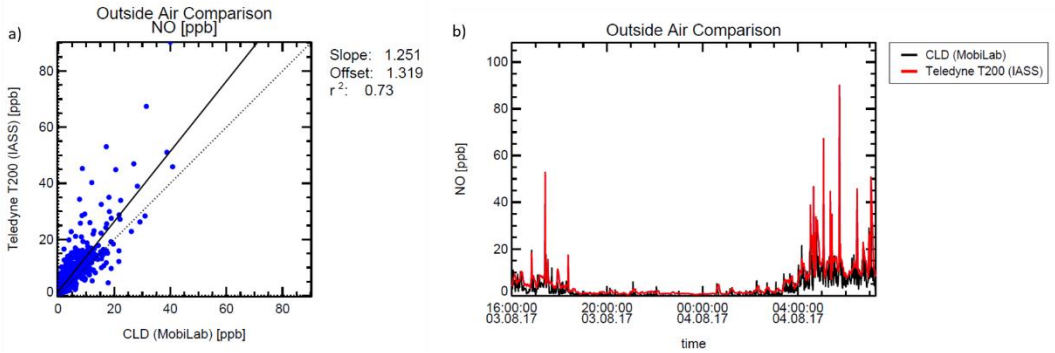

Figure 2.S-2 Intercomparison of the T-200 NO concentrations with the CLD 770 AL ppt (ECO Physics, Switzerland) from Forschungszentrum Jülich. The data are presented as: a) a scatter plot with selected data to establish a correction factor, and b) a time series plot showing the comparison between the two instruments and the range selected for the correction factor.

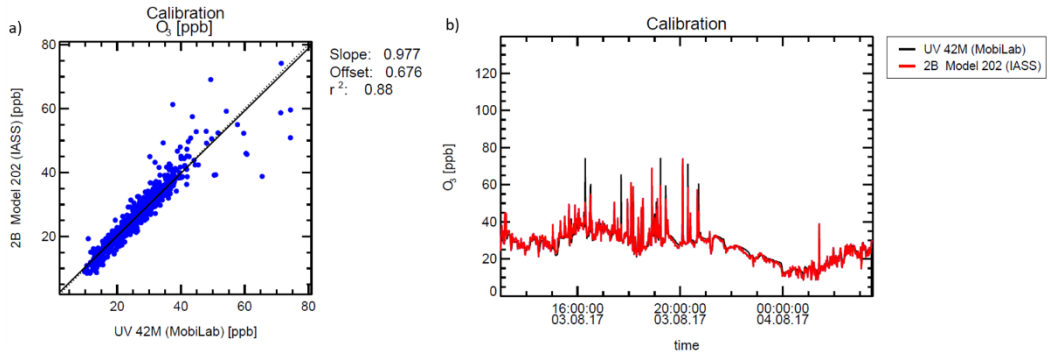

Figure 2.S-3 Intercomparison of the Tech 2B Ozone Monitor O_3 concentrations with the O242M (Environnement S.A., France) from Forschungszentrum Jülich. The data are presented as: a) a scatter plot of all data to establish a correction factor, and b) a time series plot showing the comparison between the two instruments.

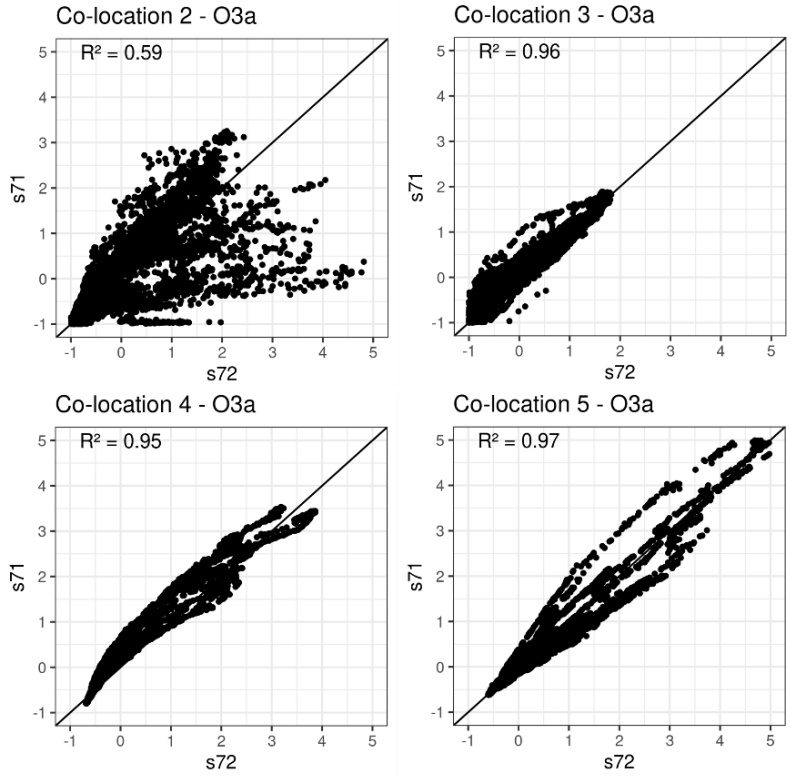

Figure 2.S-4 Intercomparison of standardized raw MOS O3a data from sensors s71 and s72 during all co-locations in the summer and winter campaigns.

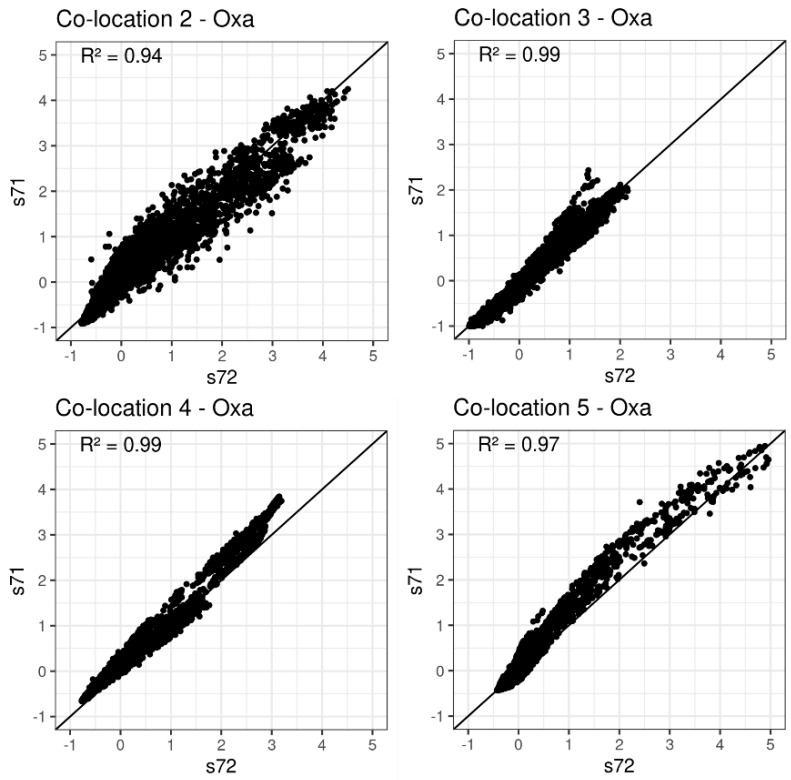

Figure 2.S-5 Intercomparison of standardized raw MOS Oxa data from sensors s71 and s72 during all co-locations in the summer and winter campaigns.

2.S.2 Map of the measurement site

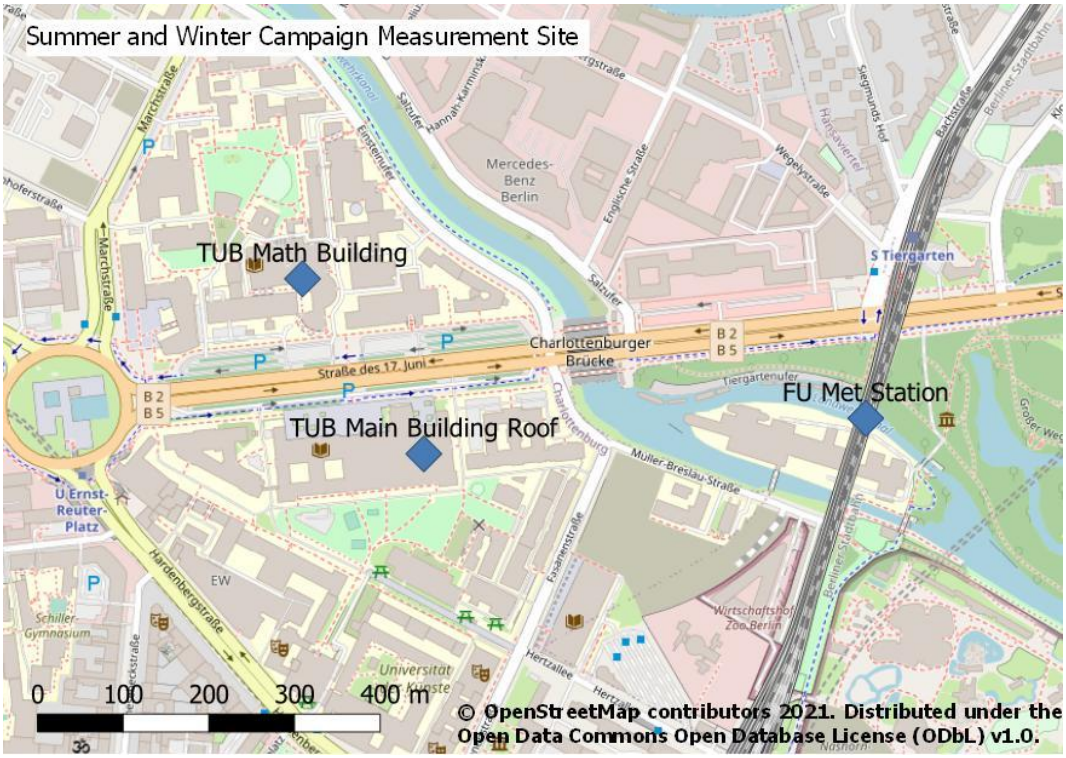

Figure 2.S-6 Map of the experimental deployment on and near the campus of the Technical University Berlin (TUB) during the Summer Campaign of 2017 and the Winter Campaign of 2018. The meteorological data provided by the Free University (FU) were collected at the site labeled on the map.

2.S.3 Final results from using s72 in the Winter Campaign, 2018

Table 2-10 Median R^2 and RMSE across all test blocks of the best MLR and RF models using internal and ambient T and RH for NO_2. RMSE and MAE are reported in units of ppb.

	NO_2	Median R^2	Median RMSE	Median MAE
MLR	NO_2 ~ log(Oxa) + log(O3a) + RH_{amb} + $1/T_{amb}$	0.55	4.13	3.05
	NO_2 ~ log(Oxa) + log(O3a) + RH_{int} + $1/T_{int}$	0.55	4.13	3.08
RF	NO_2 ~ Oxa + O3a + T_{amb}	0.53	4.35	3.36
	NO_2 ~ Oxa + O3a + T_{int}	0.53	4.17	3.25

Table 2-11 Median R^2 and RMSE across all test blocks of the best MLR and RF models using internal and ambient T and RH for O_3. RMSE and MAE are reported in units of ppb.

	O_3	Median R^2	Median RMSE	Median MAE
MLR	O_3 ~ Oxa + 1/O3a + RH_{amb} + T_{amb}	0.74	4.62	3.75
	O_3 ~ Oxa + 1/O3a + RH_{int} + T_{int}	0.72	5.16	4.17
RF	O_3 ~ Oxa + O3a + T_{amb}	0.79	4.21	3.13
	O_3 ~ Oxa + O3a + T_{int}	0.94	2.60	2.04

Table 2-12 Results of RF and MLR models for NO_2 trained with co-location 4, co-location 5, or a combination of both when tested on the Experiment 3 for IOP 3. In the lower half of the table, the models are trained with the same datasets but are tested on Experiment 3 with data points outside the ranges of each training dataset filtered out.

Formula	Co-location 4		Co-location 5		Both co-locations	
NO_2	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
NO_2 ~ log(Oxa) + log(O3a) + RH_{amb} + $1/T_{amb}$	0.53	5.77	**0.69**	**4.52**	0.66	4.65
NO_2 ~ log(Oxa) + log(O3a) + RH_{int} + $1/T_{int}$	0.52	5.87	0.68	4.59	0.66	4.70
NO_2 ~ Oxa + O3a + RH_{amb} + T_{amb}	0.37	6.52	0.64	5.13	0.63	4.82
NO_2 ~ Oxa + O3a + RH_{int} + T_{int}	0.45	6.28	0.45	6.28	0.64	4.77
NO_2 – filtered						
NO_2 ~ log(Oxa) + log(O3a) + RH_{amb} + $1/T_{amb}$	0.46	4.67	**0.65**	**4.53**	0.63	4.54
NO_2 ~ log(Oxa) + log(O3a) + RH_{int} + $1/T_{int}$	0.45	4.70	0.65	4.61	0.63	4.60
NO_2 ~ Oxa + O3a + RH_{amb} + T_{amb}	0.34	5.10	0.62	4.95	0.61	4.70
NO_2 ~ Oxa + O3a + RH_{int} + T_{int}	0.36	5.01	0.64	4.81	0.62	4.64

Table 2-13 Results of RF and MLR models for NO_2 trained with co-location 4, co-location 5, or a combination of both when tested on Experiment 3 for IOP 3. In the lower half of the table, the models are trained with the same datasets but are tested on Experiment 3 with data points outside the ranges of each training dataset filtered out.

Formula	Co-location 4		Co-location 5		Both co-locations	
O_3	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
O_3 ~ Oxa + 1/O3a + RH_{amb} + T_{amb}	0.76	6.35	0.81	8.57	0.73	4.72
O_3 ~ Oxa + 1/O3a + RH_{int} + T_{int}	0.80	5.94	0.82	7.94	0.73	4.75
O_3 ~ Oxa + O3a + T_{amb}	0.82	4.48	0.83	8.35	0.81	4.73
O_3 ~ Oxa + O3a + T_{int}	0.86	3.85	0.90	4.95	**0.95**	**2.30**
O_3 – filtered						
O_3 ~ Oxa + 1/O3a + RH_{amb} + T_{amb}	0.80	3.86	0.81	8.53	0.73	4.65
O_3 ~ Oxa + 1/O3a + RH_{int} + T_{int}	0.81	3.88	0.82	7.92	0.73	4.61
O_3 ~ Oxa + O3a + T_{amb}	0.88	2.33	0.82	8.30	0.80	4.80
O_3 ~ Oxa + O3a + T_{int}	0.91	2.06	0.91	4.52	**0.95**	**2.30**

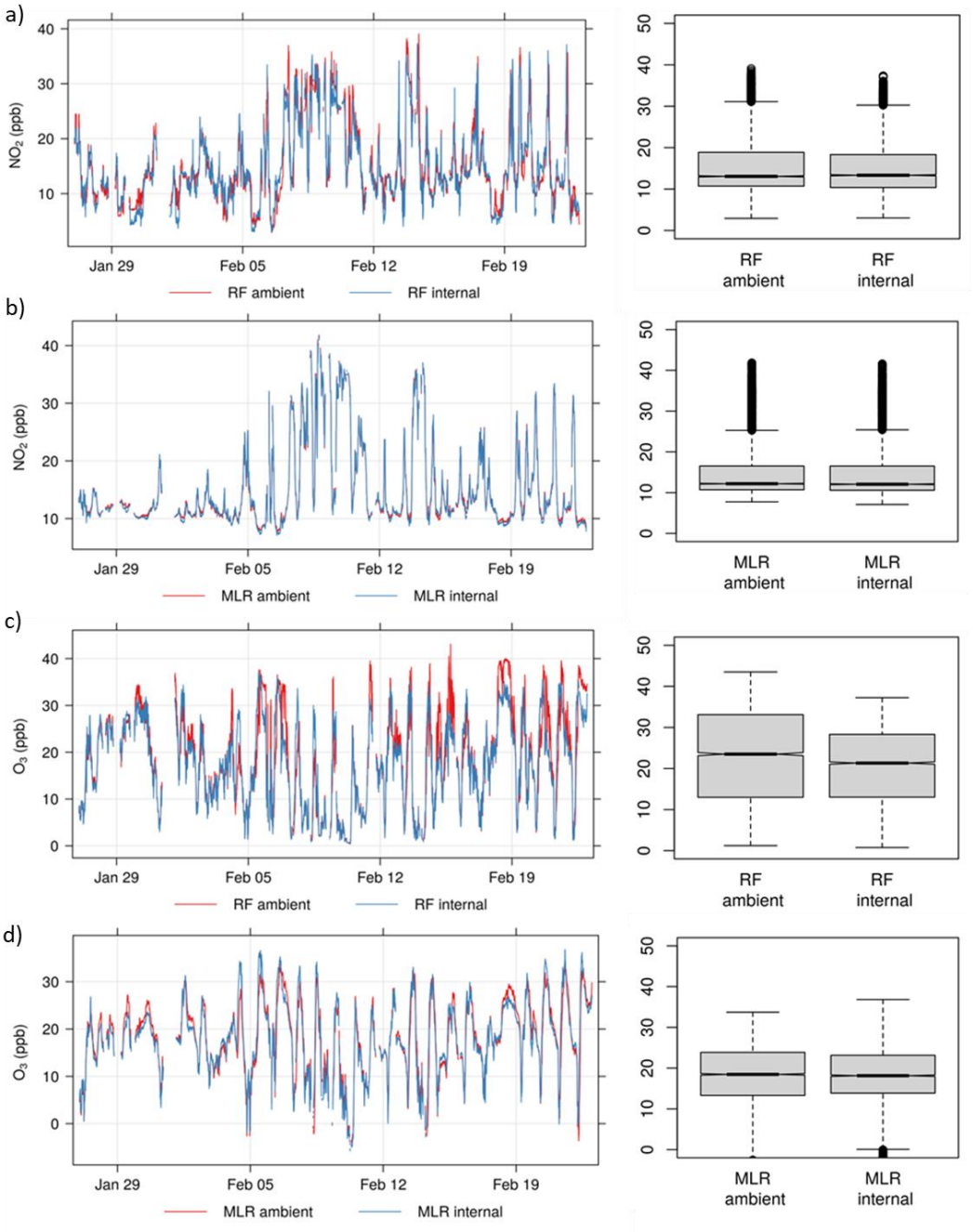

Figure 2.S-7 Time plots and histograms for IOP 3 Experiment 3 of a) predicted vs. reference NO_2 concentrations using the RF model, b) predicted vs. reference NO_2 concentrations using the MLR model, c) predicted vs. reference O_3 concentrations using the RF model, d) predicted vs. reference O_3 concentrations using the MLR model. 'Ambient' and 'internal' refer to the use of ambient or internal T and RH data in each model.

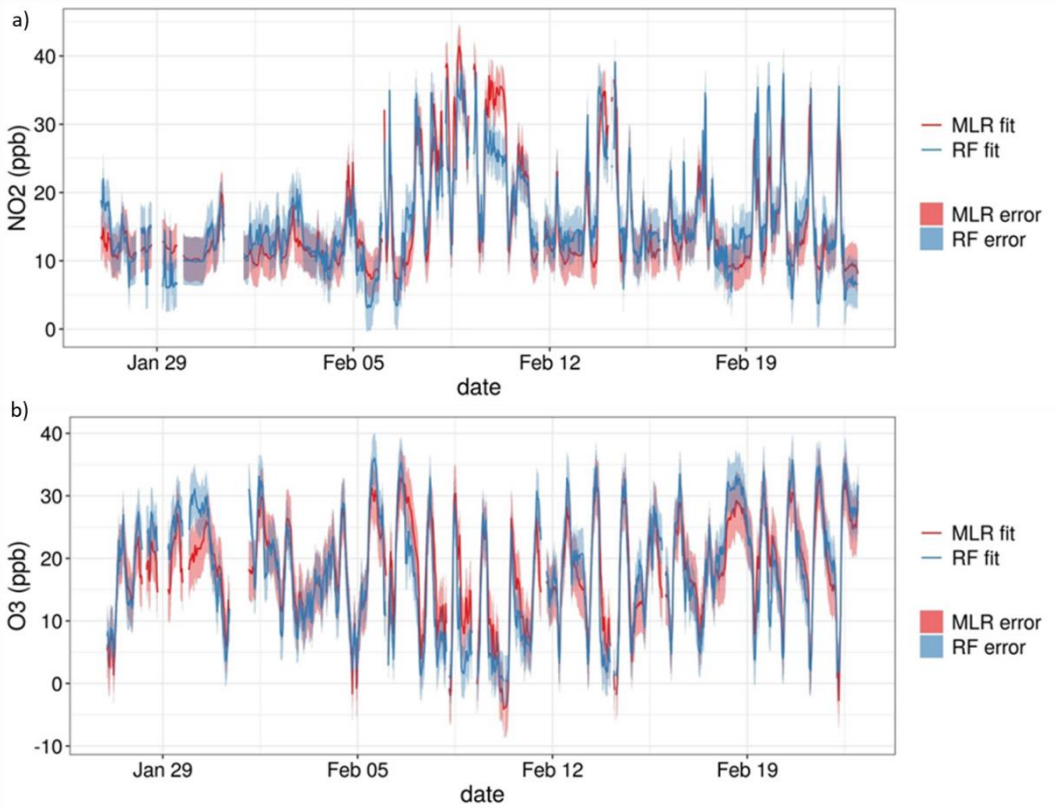

Figure 2.S-8 Time plots of both MLR and RF predictions for IOP 3 Experiment 3 including the 95% confidence intervals as shaded regions for a) NO_2 and b) O_3. Data were averaged to 30-minute resolution.

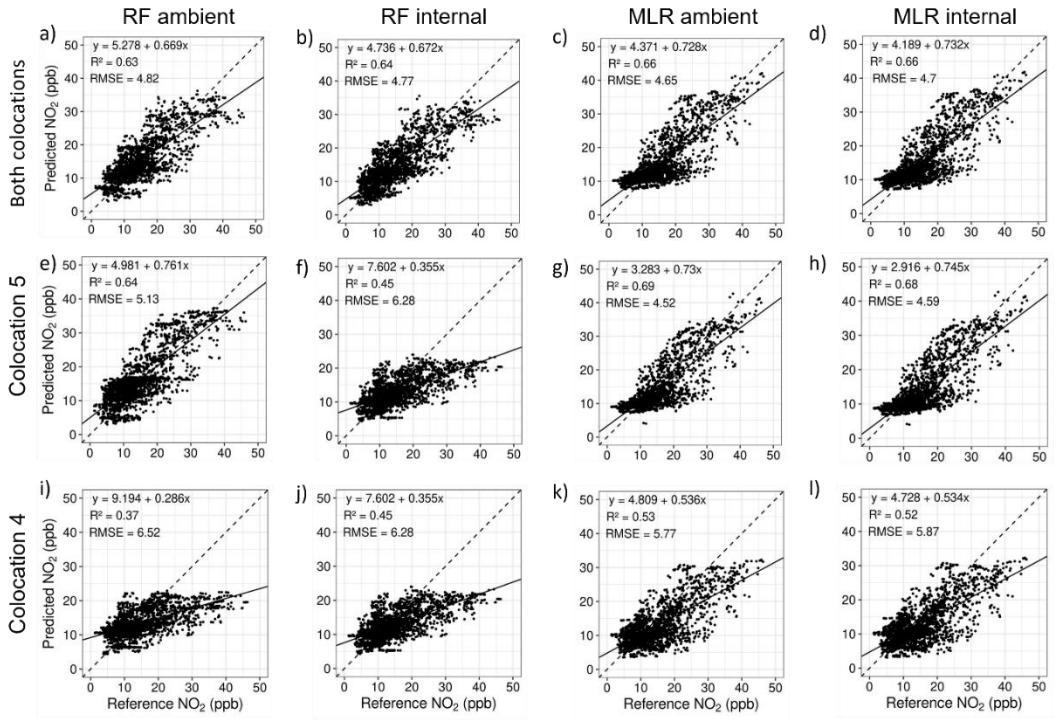

Figure 2.S-9 Scatter plots of predicted NO_2 versus reference NO_2 concentrations for Experiment 3 in the Winter Campaign using MLR and RF models trained with co-location 4 (i-l), co-location 5 (e-h), and both combined (a-d). All concentrations are reported in ppb.

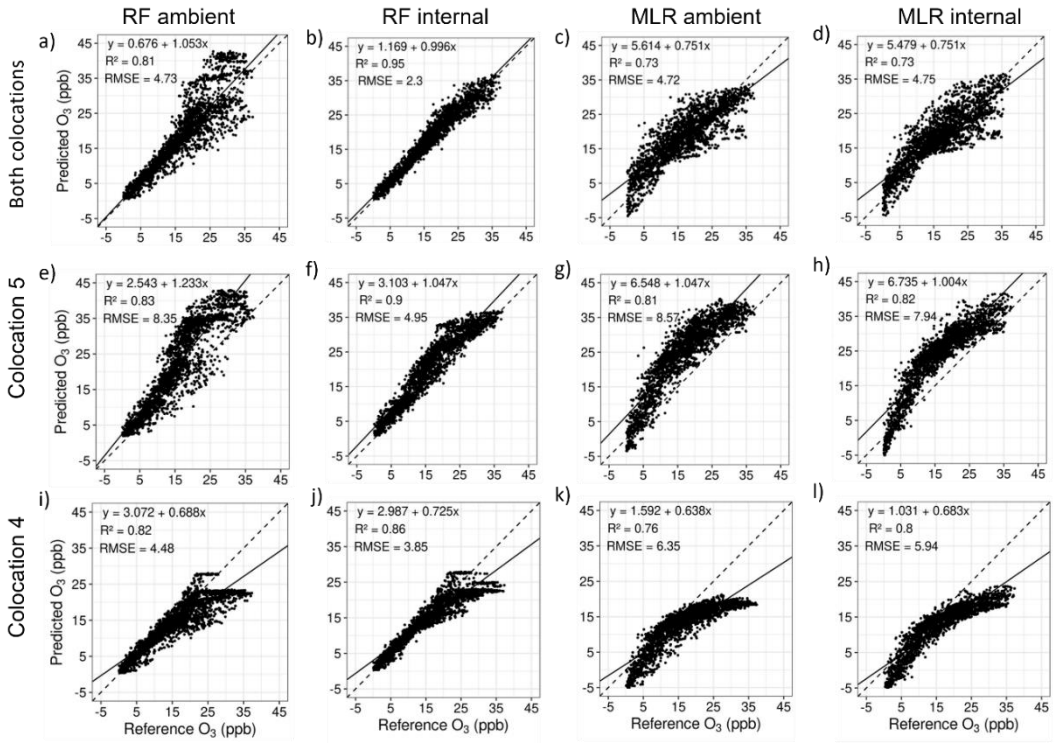

Figure 2.S-10 Scatter plots of predicted O_3 versus reference O_3 concentrations for Experiment 3 in the Winter Campaign using MLR and RF models trained with co-location 4 (i-l), co-location 5 (e-h), and both combined (a-d). All concentrations are reported in ppb.

3 Using low-cost sensors to measure vertical and horizontal gradients of NO_2 and O_3 pollution in three street canyons in Berlin

Abstract

Despite improvements in air quality over the last several decades, air pollution will continue to be a leading cause of harmful health effects in European cities as urban populations continue to grow. In recent years, the technology of low-cost sensors (LCS) has been adapted for use in expanding air pollution measurements at higher spatial resolution in cities across the globe. In a novel application, this exploratory study deploys metal oxide (MOS) and electrochemical (EC) low-cost sensors housed in EarthSense Zephyr sensor systems to measure NO_2 and O_3 concentrations in three street canyons in Berlin in winter, spring, and summer from 2017-2020. After calibration with reference instrumentation using the seven-step methodology outlined by Schmitz et al. (2021), we compare the measured concentrations with reference and urban background concentrations and investigate relationships with meteorology. We find that, following proper calibration, LCS capture expected patterns of urban pollution in association with diurnal chemistry and meteorology well. Additionally, EC sensors outperform MOS and allow for greater insights into local patterns of pollution. Furthermore, we measure concentrations of NO_2 and O_3 in street canyons that match expectations from modelling studies, indicating that high spatial resolution deployment of LCS could successfully yield new insights in urban microenvironments and inform model development. While LCS have a wider range of uncertainty than reference instruments, these results suggest that they can be reliably used for several new applications, such as validating urban street canyon models or measuring air pollution alongside changes to urban infrastructure.

Keywords: low-cost; electrochemical; metal oxide sensor; street canyon; air pollution; urban;

Published as:

Schmitz, S., Villena, G., Caseiro, A., Meier, F., Kerschbaumer, A., & von Schneidemesser, E. (2023). Calibrating low-cost sensors to measure vertical and horizontal gradients of NO2 and O3 pollution in three street canyons in Berlin. *Atmospheric Environment*, 307(2), 119830. https://doi.org/10.1016/j.atmosenv.2023.119830.

3.1 Introduction

Air pollution has remained a persistent problem in cities worldwide for centuries (Fowler et al., 2020). Over 55% of the world's population were living in cities in 2018, with this number projected to rise to 68% by 2050. In Europe, 74% of the population was already living in urban areas in 2018 (*World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*, 2019). With such a dense concentration of the population living in cities, urban air pollution negatively affects human health on both the individual and societal scale. Systematic reviews of epidemiological studies on the influence of air pollution on human health are clear; particulate matter (PM), nitrogen dioxide (NO_2), and ozone (O_3) are leading causes of premature mortality and various respiratory and cardiovascular diseases (Chen & Hoek, 2020; S. Huang et al., 2021; Huangfu & Atkinson, 2020; Landrigan et al., 2018; Mills, Atkinson, Anderson, Maynard, & Strachan, 2016; Ohlwein, Kappeler, Kutlar Joss, Kunzli, & Hoffmann, 2019; Orellano, Reynoso, Quaranta, Bardach, & Ciapponi, 2020; Tainio et al., 2021; Zheng, Orellano, Lin, Jiang, & Guan, 2021). These substantial health effects translate to an average reduction of life expectancy of 2.2 years in Europe (Lelieveld, Klingmuller, et al., 2019).

There are many factors influencing urban air pollution, including local emissions, chemical reactions, long-range transport, and regional meteorological patterns at macro-, meso-, and microscale (Perez et al., 2020). At macroscale, synoptic patterns and fronts determine air parcel trajectories over large distances; at mesoscale, orographic features predominantly change the way in which air parcels circulate, recirculate, and stagnate in a given region; and at microscale, the atmospheric boundary layer height, local temperature and relative humidity, and urban morphology are of significance (Perez et al., 2020).

Previous studies in Central Europe and Germany have found significant positive correlations between O_3 concentrations and temperature, as well as negative relationships with relative humidity (Melkonyan & Kuttler, 2012; Noelia Otero et al., 2018; N. Otero et al., 2016), with some evidence of a negative influence of wind speed in some parts of Europe (Melkonyan & Kuttler, 2012; N. Otero et al., 2016). The relationship with ambient temperature is particularly strong in summer, as higher temperatures correlate with increased solar radiation, which cause an increase of both biogenic volatile organic compounds (BVOCs) emissions and evaporative emissions of anthropogenic VOCs (Ordóñez et al., 2005; Pusede et al., 2014), as well as an increase in photochemical activity. One study in Berlin found a clear, positive relationship between mixing layer height (MLH) and O_3 concentrations, but attributed this to other physical processes such as the photochemical production of O_3 and downward mixing from the residual layer (Geiß et al., 2017).

NO_2 concentrations are affected differently by local meteorology than O_3 as they are highly correlated with traffic NO_x emissions in urban areas and vary substantially due to complex dispersion patterns. Temperature has been found to play a minor role (Pearce et al., 2011), with some studies finding temperature to be the least relevant meteorological parameter in predicting NO_2 (Voiculescu et al., 2020). In rural areas, there is a larger influence of temperature through increased soil NO_x emissions (Romer et al., 2018). Some studies have found no strong relationship between NO_2 and relative humidity (Aldrin & Haff, 2005; Pearce et al., 2011), though others do see a consistent positive correlation (Voiculescu et al., 2020). The aforementioned study in Berlin found relationships between

MLH and NO_x concentrations that were negative at rural and urban background sites, but positive at traffic sites (Geiß et al., 2017). Another study from Germany, however, found NO_2 to be the pollutant least affected by MLH (Wagner & Schäfer, 2017). The most dominant meteorological factor for NO_2 in urban environments appears to be wind speed, especially as pertains to dispersion of traffic-related emissions (Carslaw et al., 2007; Bart Degraeuwe et al., 2017; Pearce et al., 2011). Particularly on days with very low wind speeds, NO_2 emissions do not disperse in the stagnated air and concentrations tend to increase (Carslaw et al., 2007; Elminir, 2006; Pearce et al., 2011).

In addition to meteorology, urban morphology affects the distribution of air pollutants as it changes how air flows across the city. The flow is affected predominantly by wind speed and direction, and architectural form, influencing concentration and location of pollutant accumulation. These relationships have been modelled often using computational fluid dynamics (CFD) models and are typically well-understood (Bright, Bloss, & Cai, 2011; Gonzalez Olivardia, Zhang, Matsuo, Shimadera, & Kondo, 2019; Y.-D. Huang et al., 2019; Park et al., 2019; Park, Kim, Kim, Park, & Cheong, 2015; Voordeckers, Lauriks, et al., 2021; Voordeckers, Meysman, Billen, Tytgat, & Van Acker, 2021). In general, wind that travels parallel to a street canyon and flows through it will lead to elevated pollutant concentrations downwind, whereas wind travelling perpendicularly to the street will lead to elevated concentrations on the leeward side of the street with respect to the windward side and the rooftop (Bright et al., 2011; Y.-D. Huang et al., 2019; Voordeckers, Lauriks, et al., 2021), particularly of pollutants emitted from traffic such as PM or NO_x. There are, however, few studies that validate these patterns of street canyon pollution with observations (Chew, Glicksman, & Norford, 2018; Kwak, Lee, Seo, Park, & Baik, 2016), as such measurement campaigns are logistically intensive to organize and expensive if reference-grade instruments are used.

The relationship between concentrations at street-level versus the building rooftops of street canyons is also well-understood, with several studies calculating road-to-roof concentration ratios (Kukkonen et al., 2001; Kwak et al., 2016; S.-J. Park et al., 2015; Xie, Zhang, Qi, & Tang, 2003). For O_3, concentrations are typically lower in street canyons, with road-to-roof concentration ratios between 0.53 and 0.85. NO_2 concentrations, on the other hand, are higher at street-level with road-to-roof ratios between 1.2 and 3.2 across various studies. These relationships are most affected by street morphology and wind direction, though wind speed does have an impact on the road-to-roof ratio of O_3 and NO_2 (Kwak et al., 2016).

In the past decade, low-cost sensors (LCS) have become a popular new technology in the field of atmospheric chemistry. The atmospheric measurement community has embraced research investigating the accuracy (Cross et al., 2017; Karagulian et al., 2019; Malings et al., 2019; Rai et al., 2017; Russell et al., 2022; Zimmerman et al., 2018), applicability (Bigi et al., 2018; Castell et al., 2017; McKercher et al., 2017; Morawska et al., 2018), and technology (Fishbain et al., 2017; Peterson et al., 2017; Spandonidis et al., 2020; David E. Williams, 2020) behind LCS to further understand the extent to which they can improve our understanding of air pollution, particularly in urban environments. Many studies assess the extent to which these sensors can complement traditional monitoring networks by filling measurement gaps in cities (J. M. Barcelo-Ordinas et al., 2019; Kim, Shusterman, Lieschke, Newman, & Cohen, 2018; Popoola et al., 2018; D. E. Williams, 2019), with some others conducting measurements in a citizen science context (Bosello et al., 2020; Ripoll et al., 2019). However, few studies have looked at measuring air

pollution in specific urban environments to understand micro-scale patterns of pollution (Hofman et al., 2022), with none that have measured air quality in street canyons.

While it is crucial that research on the calibration and development of LCS continues, more research is required regarding the evaluation of LCS performance in field experiments. In contrast to reference-grade instrumentation, the versatility of LCS enables their deployment at high spatiotemporal resolution to measure complex patterns of air pollution in micro-scale urban environments. Some recent work has highlighted the importance of field calibration at the experimental site or an analogue, as well as the need for regular co-location with reference instrumentation to appropriately capture changes in sensor sensitivity and meteorological conditions (Schmitz, Towers, et al., 2021b). However, this has thus far not been thoroughly investigated. Therefore, this exploratory study seeks to fill a gap in the literature by evaluating field calibrated LCS measurements of NO_2 and O_3 in urban environments, particularly in street canyons. The three main research questions guiding this work are:

1. Do field calibrated LCS capture expected changes in urban NO_2 and O_3 pollution in accordance with changes meteorology?
2. Are field-calibrated LCS able to capture expected changes in NO_2 and O_3 concentrations in urban street canyons and what new insights can they bring?
3. What implications do these results have for the applicability of LCS in the measurement of urban NO_2 and O_3 pollution?

In the following, Section 3.2 elaborates on the sensors used, the measurement sites selected, and the choice of statistical analysis. Section 3.3 highlights the key results ascertained from this work and Sections 3.4 and 3.5 discuss their general and policy-relevant implications, as well as limitations to this research. Further figures and tables relevant for this work can be found in the supplementary information.

3.2 Methods

3.2.1 Sensor specifications

The air quality monitors used in this study are EarthSense Zephyrs©, henceforth referred to as “Zephyrs”, which refers to the whole small sensor system including low-cost sensors, housing, battery, etc. Two versions of Zephyrs were used throughout this study as they were being continually developed by EarthSense: Zephyr “prototypes” and “Fully operational Zephyrs”. Installed within the Zephyr prototypes were an assortment of metal oxide sensors (MOS) that measure reducing gases, oxidizing gases, ozone, and ammonia, as well as environmental sensors for temperature and relative humidity; see Table 3-1 for more on these sensor specifications. For greater detail on the development, functioning, and operation of the sensors housed within these prototypes, see Peterson et al. (2017). Installed within the fully operational EarthSense Zephyrs are primarily electrochemical cell (EC) sensors that measure NO_2 and O_3, a micro-optical sensor that measures PM, as well as further environmental sensors for internal temperature (T), relative humidity (RH), and pressure (Table 3-2). Included in both versions of the Zephyrs are: 1) a rechargeable lithium-ion battery; 2) an internal fan for air intake and expulsion; 3) a global positioning system (GPS) unit; and 4) a Global System for Mobile Communications unit for sending logged data to an external database. The air intake is located on the bottom of the Zephyrs

Table 3-1 Sensors installed within the EarthSense Zephyr prototypes. Table reproduced from Peterson et al. (2017).

Variables Measured	Sensor Model	Sensor Type	Gas detected and detection limits
Reducing gases	SGX Sensortech MICS-4514	Metal Oxide	CO: 1-1000 ppm NH_3: 1-500 ppm C_2H_5OH: 10-500 ppm H_2: 1-1000 ppm CH_4: >1000 ppm
Oxidizing gases	SGX Sensortech MICS-4514	Metal Oxide	NO_2: 0.05-10 ppm H2: 1-1000 ppm
Ozone	SGX Sensortech MICS-2614	Metal Oxide	.01-1 ppm
Ammonia	SGX Sensortech MICS-5914	Metal Oxide	NH_3: 1-500 ppm C_2H_5OH: 10-500 ppm H_2: 1-1000 ppm C_3H_8: >1000 ppm $C_2H_8(CH_4)_2$: >1000 ppm
Temperature and relative humidity	GE Measurement and Control CC2D25	Polyamide capacitance	Temp.: -40 – 125 °C RH: 0 – 100%

Table 3-2 Sensors installed within the fully operational EarthSense Zephyrs.

Variables Measured	Sensor Model	Sensor Type	Gas detected and detection limits
NO_2 O_3	Alphasense NO2A43F Alphasense OXA431	Electrochemical	NO_2: 0-10 ppm O_3: 0-7.5 ppm
PM_1 / $PM_{2.5}$ / PM_{10}	Plantower PMS5003	Micro-optical	PM_1: 0-20,000 µg.m^{-3} $PM_{2.5}$: 0-20,000 µg.m^{-3} PM_{10}: 0-20,000 µg.m^{-3}
Temperature, pressure, and relative humidity	Sensirion SHT31 Bosch BME680	Environmental	Temp.: -40 – 125 °C Pres.: 300 – 1,100 hPa RH: 0 – 100%

3.2.2 Measurement sites

Between 2017 – 2020, measurements were conducted in Berlin, Germany, at three separate sites with different street morphologies, as is depicted below in Figure 3-1. These sites (Strasse des. 17 Juni, Frankfurter Allee, and Kottbusser Damm) are briefly described in the following sections. For all dates of deployments and general characteristics of each measurement campaign, see Table 3-3.

3.2.2.1 Strasse des 17. Juni

Measurements were conducted in a main thoroughfare that cuts through the campus of the Technical University (TU) Berlin, at Strasse des 17. Juni 135 (TU Mathematics Building) (52° 30' 49.7" N, 13° 19' 34.5" E) and Strasse des 17. Juni 13 (52° 30’ 44.2” N, 13° 19’ 41.4” E) (Main TU Building) on the roof. The area directly around the measurement site is mainly university buildings with a main thoroughfare (Strasse des 17. Juni) that runs West-East through Berlin (see Figure 3-1 a). At the experimental site, the Strasse des. 17 Juni has six lanes of car traffic, three in each direction; there is a median strip used for parking in the middle of the street, as well as available parking on each side of the street; and there is a slip road that allows cars to access additional parking areas that are off the main street, in some cases also integrating a bike lane. The total distance between the buildings

including the street, median, sidewalk, and parking area is ca. 115 m, with an aspect ratio (AR) of 0.37, thus classifying it as a wide asymmetric avenue canyon, according to Vardoulakis, Fisher, Pericleous, and Gonzalez-Flesca (2003). A 2019 study by the Berlin Senate Department for the Environment, Urban Mobility, Consumer Protection and Climate Action estimated that 42,700 cars and 920 trucks traverse this segment of the street on workdays (SenUMVK, 2019).

During this experiment, five Zephyrs were deployed on the façade of the mathematics building (south-facing façade - north side of the street), one on the roof of the Main TU Building (north-facing - south side of the street), and one on a parked van in the median strip of the street (Figure 3-1). The Zephyrs were field calibrated during co-locations with reference instruments at the deployment site in an office on the 6th floor of the mathematics building (Figure 3.S-1). These reference measurements were continuous throughout the co-locations and the experiments. Reference instrumentation used for co-location included a Teledyne Model T-200 NOx Analyser and a 2B Technologies Ozone Analyser. Five co-locations were conducted in connection with the experimental deployment (see Table 3-3).

3.2.2.2 *Frankfurter Allee*

In 2018, measurements were conducted on Frankfurter Allee in front of the Friedrichshain-Kreuzberg district administrative office building at Frankfurter Allee 35/37 (52° 30’ 54.9” N, 13° 27’ 41.3” E). This street is a main thoroughfare with a West-East orientation connecting central Berlin with the eastern part of the city, high volumes of traffic (48,500 cars, 1,110 trucks on workdays) (SenUMVK, 2019), and a typical street canyon morphology. Along the length of the street, there are three lanes of traffic in each direction, but directly at this site in the westward direction is a turning lane, representing a 4th lane of traffic. On each side of the road are five-story buildings, roughly 22 m tall. The street canyon itself is roughly 40 m wide, including traffic lanes, median, and sidewalk (see Figure 3-1 b). It has an AR of 0.55 and should be classified as a long, symmetric avenue canyon (Vardoulakis et al., 2003).

During this experiment, two Zephyrs were deployed at street level on lampposts at roughly 3 m height, one south-facing on the northern side of the street, and the other north-facing on the southern side of the street. A third Zephyr was placed on the roof of the administrative office building (south-facing), but for logistical reasons, it could not be placed at the highest point, but rather 2m lower on a terraced portion of the roof. See Table 3-3 for dates of deployment and co-location. In this experiment, no field calibration station could be set-up directly at the site; instead, the Zephyrs were co-located at a roadside monitoring station (MC 174) of the Berlin air pollution measurement network – Berliner Luftgüte Messnetz (BLUME), located in the same street canyon several blocks east of the experimental site at Frankfurter Allee 86b (52° 30’ 50.7” N, 13° 28’ 11.7” E). The BLUME network measures a variety of air pollutants (including NOx, PM, and O_3) and adhere to strict standards to determine compliance with air quality in accordance with EU regulations (European Parliament, 2008).

The instruments located in this measurement station are an APNA-370 NOx Monitor to measure NO_2 concentrations in ambient air in accordance with DIN EN 14211 (DIN, 2012) and a HORIBA APOA 370 for O_3. Figure 3.S-2 (see supplementary information) shows the general set-up of this co-location, with the Zephyrs installed on the railing of the roof of the station, near the air intake used for the reference instruments inside. The units face away from the street to prevent potential damage from passing trucks or busses.

3.2.2.3 Kottbusser Damm

The final experiment took place on Kottbusser Damm in 2020 as a part of a larger measurement campaign associated with the construction of a new bike-lane (Schmitz, Caseiro, Kerschbaumer, & von Schneidemesser, 2021). Kottbusser Damm is a Northwest-Southeast oriented thoroughfare between the Berlin city districts of Friedrichshain-Kreuzberg and Neukölln but does not experience traffic volume as large as on Frankfurter Allee or Strasse des 17. Juni (23,400 cars, 660 trucks on workdays) (SenUMVK, 2019). As with Frankfurter Allee, Kottbusser Damm has a typical urban street canyon morphology, but with fewer traffic lanes in either direction it has a comparably higher aspect ratio. The experiment took place at Kottbusser Damm 25 and involved an intersecting side-street (Lenau Strasse) (Figure 3-1 c). The AR of this site is 0.63 and should be classified as a long, symmetric regular canyon (Vardoulakis et al., 2003).

As with Frankfurter Allee, two Zephyrs were placed diagonally from one another at street-level facing the street, with the one on the eastern side of the street placed at the intersection with Lenau Strasse. A third Zephyr was placed on the 4th floor windowsill of the western building outside. No access to the roof at this site was obtained, therefore the Zephyr was placed as high on the building as possible. See Table 3-3 for dates of deployment. As with Frankfurter Allee, no space for on-site co-locations was available and therefore the Zephyrs were co-located at BLUME station MC 117 at Schildhornstrasse 76. While this station is in a different part of the city, it has a similar morphology (2 traffic lanes each direction) and traffic volume (31,900 cars, 450 trucks on workdays) as Kottbusser Damm and was considered suitable as an analogue site (Schmitz, Towers, et al., 2021b).

During the Kottbusser Damm campaign, the COVID-19 global pandemic caused strict lockdowns to be enforced in Berlin. To prevent any influence of these policies on the results presented here, data collected during the lockdown between March 22nd and April 23rd, 2020, were removed from the analysis. There were still measures in place after April 23rd, 2020, but as activity in the city returned almost to pre-lockdown levels, these data were not removed. Closer analysis of this lockdown data was considered out of the scope of this study.

Table 3-3 Details about each experimental deployment, including morphology, dates, and sensor type.

Site (Season)	Aspect Ratio	LCS Type (No. of Zephyrs)	Pollutants Measured	Co-locations	Experimental Deployment
Strasse des 17. Juni (Summer)	0.37	MOS (5)	NO_2 O_3	18.07.2017 - 27.07.2017; 29.08.2017 - 07.09.2017; 14.10.2017 - 27.10.2017	29.07.2017 - 28.08.2017
Strasse des 17. Juni (Winter)	0.37	MOS (5)	NO_2 O_3	13.01.2018 - 25.01.2018; 23.02.2018 - 08.03.2018	27.01.2018 - 23.02.2018
Frankfurter Allee (Summer)	0.55	EC (3)	NO_2 O_3	26.06.2018 - 07.07.2018; 01.09.2018 - 19.09.2018	20.07.2018 - 30.08.2018
Kottbusser Damm (Winter/Spring)	0.63	EC (3)	NO_2	02.02.2020 - 18.02.2020; 15.05.2020 - 02.06.2020	21.02.2020 - 11.06.2020

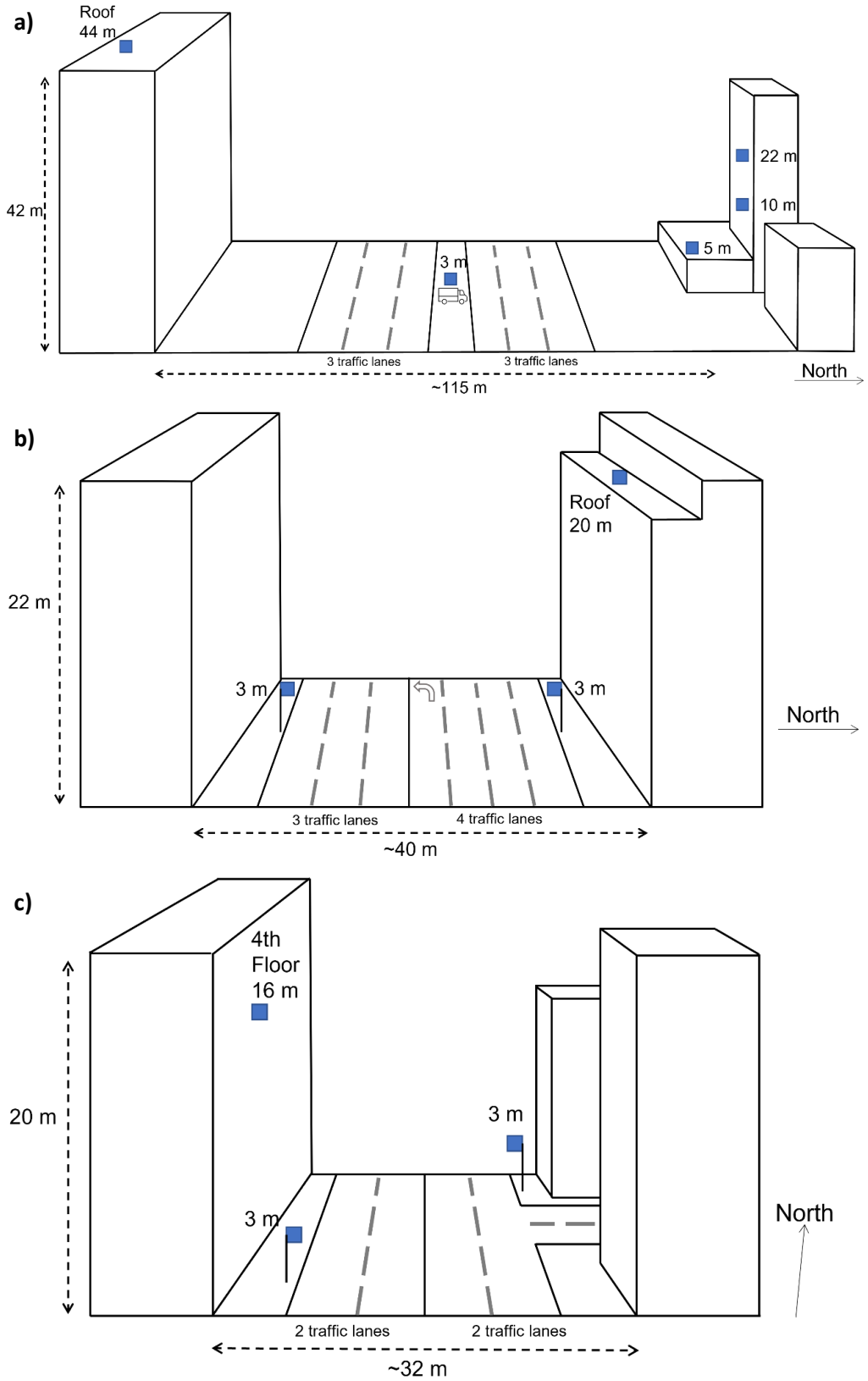

Figure 3-1 Street morphology and locations of the Zephyrs during the experimental deployment on a) Strasse des 17. Juni, b) Frankfurter Allee, and c) Kottbusser Damm in Berlin, Germany. Diagrams not to scale.

3.2.3 Other data sources

At each experimental site, the meteorological parameters of temperature (T), relative humidity (RH), wind direction (WD), wind speed (WS), and mixing layer height (MLH) were collected from available sources. These were used both in the calibration of the Zephyrs and in subsequent analysis. For the campaign on Strasse des 17. Juni, T and RH data were available from a Free University (FU) monitoring site (52° 30’ 45” N, 13° 20’ 8.16” E) roughly 400 m away. Wind (WD and WS) and MLH data were available from a monitoring site on top of the TU main building. MLH was calculated by implementing the COBOLT method on ceilometer data collected with a Lufft CHM15k (Geiß, 2016).

For the experiments on Frankfurter Allee and Kottbusser Damm, meteorological data were gathered from the German Weather Service (Deutscher Wetterdienst; DWD) station located at the former Tempelhof Airport (52° 28’ 57” N, 13° 24’ 00” E). From this station, T, RH, WD, and WS data were collected for use with these two experimental sites, as no on-site meteorological data were available. Ceilometer data at these two sites were also unavailable, therefore data from the top of the TU main building were used. This is a limitation to this work, as on-site meteorological conditions, especially with regards to WD and WS, were likely different from those at Tempelhof Airport, located several kilometres away. However, given the exploratory nature of this study, the inclusion of these meteorological parameters did still provide helpful indications of their importance to local patterns of air pollution. Henceforth, this data will be referred to as ‘prevailing’ WD and WS.

Each field campaign was also compared with urban background NO_2 and O_3 concentrations measured by the BLUME network. In this case, averages of several urban background stations surrounding the experimental sites were calculated for each campaign. For NO_2, stations MC 010, MC 018, MC 042, and MC 171 were used, whereas for O_3 data were only available from stations MC 010 and MC 042 (Figure 3-2). Data from these stations are reported in $\mu g.m^{-3}$ and were converted to ppb for this study using standard temperature (298 K) and pressure (1 atm).

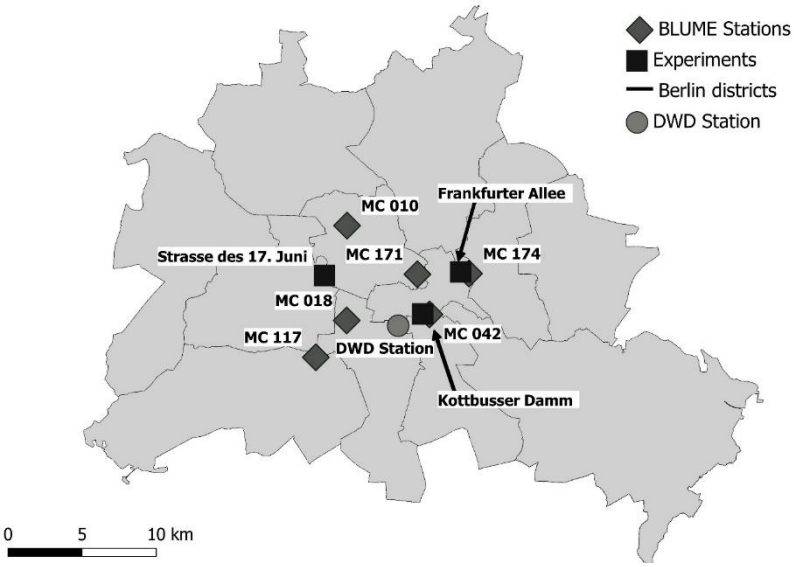

Figure 3-2 Locations of BLUME measurement stations and experimental sites in Berlin. MC 010, MC 171, MC 018, and MC 042 are urban background sites, whereas MC 117 and MC 174 are roadside monitoring stations.

3.2.4 Statistical methods

3.2.4.1 Zephyr calibration

The raw signal, which is typically a measure of voltage or resistance for gas-phase sensors, must first be properly calibrated with reference instrumentation. Using the seven-step methodology developed in Schmitz, Towers, et al. (2021b), the Zephyr prototypes and fully-operational Zephyrs were calibrated with the aforementioned reference instruments during co-locations conducted before and after each field deployment (Table 3-3). The Zephyr prototypes were calibrated using raw signals from the MICS-4514 and MICS-2614 metal oxide sensors, which measure oxidizing gases and O_3, respectively. Additional measures of internal Zephyr temperature and relative humidity, which reflect sensor operating conditions, were also included in the model building process. The fully operational Zephyrs were calibrated using raw signals from the working electrodes of the Alphasense electrochemical cell sensors (NO2A43F and OXA431), as well as with internal Zephyr temperature and relative humidity. Correction of the raw signal using the auxiliary and reference electrodes of the Alphasense sensors was found to not be necessary for calibration. All available field co-location data were subsequently used to train models of best-fit, which in turn were used to produce concentrations of NO_2 and O_3. All concentrations reported in this manuscript were produced from MLR models optimized using the seven-step methodology with the aforementioned input data. For greater detail on this methodology, please refer to Schmitz, Towers, et al. (2021b). All statistical analysis was conducted using R statistical software (R Core Team, 2022).

To assess calibration model performance during each experimental period, tests were conducted using validation sets from each co-location for each individual Zephyr. For co-locations that occurred prior to each campaign, models were trained with the initial 75% of the data and tested on the remaining 25%, closest to the experimental window. For co-locations that occurred after each campaign, the same principle was used with the initial 25% and remaining 75% of the data. Results of these tests can be found in the supplementary information.

3.2.4.2 Statistical tests

To compare distributions and means of NO_2 and O_3 concentrations between Zephyrs during each experiment, Mann-Whitney-Wilcoxon U-tests (MWU) and student's t-tests, respectively, were conducted. The null hypothesis of the MWU test states that the distributions of both populations are identical. In this case, a p-value of 0.05 was used to determine significance. If the value of the test was lower than this threshold, the alternative hypothesis was accepted and the two distributions were considered significantly different from one another. Results of the student's t-tests can be found in the supplementary information.

3.2.4.3 Generalized additive models

To assess the influence of meteorology on calibrated NO_2 and O_3 concentrations, generalized additive models (GAMs) were built with T, RH, WS, WD, and MLH as predictors for every Zephyr, reference instrument, and set of urban background data across all experiments. In short, GAMs are regression models that replace linear coefficients with smoothing splines for each covariate, which allows for better modelling of complex non-linear relationships, as is often the case with air pollution data (Carslaw et al., 2007; Hastie & Tibshirani, 1990). The formula can be given by:

$$\log(y_i) = \beta_0 + \sum_{j=1}^{n} s_j(x_{ij}) + \varepsilon_i$$

where y_i is the NO_2 or O_3 concentration at time *i*, β_0 is the mean of the response, $s_j(x_{ij})$ is the smoothing function of covariate j at time i, the total number of covariates is given by *n*, and ε_i is the normally distributed residual at time residual at time *i* (Hastie & Tibshirani, 1990; Pearce et al., 2011). The *mgcv* package using R statistical software and the *gam()* function were used to this end. Smoothing parameter estimation occurs automatically, using a mix of generalized cross validation (GCV) and the un-biased risk estimator (UBRE) criterion to optimize the smoothing function based on penalized regression splines (S. N. Wood, 2017).

These models were then used to calculate the total explained variance of NO_2 and O_3 concentrations by the five meteorological variables independently of source emissions. This limits the explanatory power of the models, especially for NO_2 given the importance of local emissions, but is suitable for exploring general relationships with meteorology. Partial effects plots were made to understand influences of each meteorological variable on NO_2 and O_3. GAMs were favoured over linear models due to the existence of various non-linearities in the relationships between NO_2 and O_3 and meteorological parameters. Linear models do not model these complex relationships effectively and were therefore not used (Carslaw et al., 2007; Pearce et al., 2011; Thompson, Reynolds, Cox, Guttorp, & Sampson, 2001).

3.3 Results

In the following section, analysis of the calibrated NO_2 and O_3 concentrations from the Zephyrs is presented, focusing first on general distributions of pollution, then looking more closely at relationships with meteorology and street canyon morphology. Strasse des 17. Juni was excluded from the assessment of impacts of street canyon morphology due to its low aspect ratio, heterogeneous building morphology, and the short extent of the canyon. The dynamics of the site are different than would be expected in a typical urban street canyon, particularly with regards to wind flow and pollutant dispersion and the uncertainty and variability between MOS installed in the prototype Zephyrs is too high to identify gradients at the site.

3.3.1 Model performance

Models built for each Zephyr during each experiment were validated with a test set in each co-location where data were available to test model performance. Results are presented in Figures 3.S-3 - 3.S-6 in the supplementary information. On Strasse des 17. Juni in the summer (Figure 3.S-3), MLR models for MOS performed poor to well, with large variability among sensors for NO_2 (R^2: 0.25 – 0.72; MAE: 1.85 – 3.84 ppb) and well for O_3 with less variability (R^2: 0.79 – 0.95; MAE: 2.75 – 4.29 ppb) across both co-locations. In the winter (Figure 3.S-4), performance was poorer for NO_2 (R^2: 0.18 – 0.79; MAE: 1.69 – 7.47 ppb) and better for O_3 albeit with greater variability (R^2: 0.88 – 0.97; MAE: 0.91 – 4.03 ppb). In all cases, models tended to overpredict both NO_2 and O_3 when compared to reference instrumentation, with evidence of nonlinearities present in some of the modelled relationships. Figures 3.S-7 and 3.S-8 show diurnal profiles of concentrations of all units versus reference concentrations during each test set for the summer and winter campaigns, respectively. In general, intra-sensor agreement of calibrations and agreement with reference concentrations was acceptable for NO_2 and very good for O_3 in

the summer, but with greater disagreement at night. In the winter, there was greater intra-sensor disagreement for NO_2 and poor performance with respect to reference concentrations, whereas for O_3, agreement between sensors and with reference concentrations remained good.

On Frankfurter Allee (Figure 3.S-5), most MLR models for EC sensors performed acceptably, with one sensor performing very poorly, for NO_2 (R^2: 0.02 – 0.58; MAE: 3.50 – 8.28 ppb), and well for O_3 with the exception of one sensor (R^2: 0.36 – 0.95; MAE: 3.84 – 13.05 ppb) during the first co-location. This is corroborated by diurnal profiles of the calibrations depicted in Figure 3.S-9. As a result, issues with data quality and consistency were identified, and therefore training data were not included from this co-location for any Zephyrs during final calibration. For this same reason, data from 20.07. – 05.08.2018 during the experiment in later analysis were also excluded. During the second co-location, models performed in all but one case well for NO_2 (R^2: 0.48 – 0.87; MAE: 2.19 – 5.44 ppb) and in all cases very well for O_3 (R^2: 0.93 – 0.98; MAE: 1.20 – 2.57 ppb). Models tended to overpredict NO_2 at lower concentrations (0 – 10 ppb). Intra-sensor agreement and agreement with reference concentrations during co-location 2 was very good for O_3 and poor for NO_2 (Figure 3.S-9), especially under lower night-time concentrations. On Kottbusser Damm (Figure 3.S-6) NO_2 models performed well across both co-locations (R^2: 0.61 – 0.91; MAE: 2.12 – 3.85 ppb), while also showing a tendency to overpredict at lower concentrations. Diurnal profiles show good intra-sensor agreement and improved performance versus reference concentrations as compared to the co-locations conducted on Frankfurter Allee (Figure 3.S-10), highlighting the importance of training data quality for calibration.

3.3.2 Diurnal profiles of each experiment

Figures 3-3 and 3-4 show the diurnal profiles of the NO_2 and O_3 concentrations captured over the course of each experiment, including data from the Zephyrs, reference instruments, and the urban background. Results of Wilcoxon-Mann-Whitney U-tests for statistically significant differences between the distributions can be found in Tables 3-4 and 3-5, with results of students t-tests of the means in Tables 3.S-1 and 3.S-2.

On Strasse des 17. Juni, both in the winter and the summer, the Zephyrs capture expected daily fluctuations in NO_2 and O_3 concentrations (Figure 3-3 a-d). NO_2 summer concentrations at street-level were slightly higher than those at all other heights, the 6th floor reference site, and the urban background. No clear evidence exists for a vertical gradient in NO_2 concentrations across the various heights, though all sites are statistically different from one another via the u-test (Table 3-4). A student's t-test reveals that the means at the 1st floor, rooftop, and the urban background are not statistically different (Table 3.S-1). The relationship for O_3 is reversed, with rooftop, 3rd floor, and 6th floor concentrations higher than at street level and on the 1st floor, but also higher than 6th floor reference and ground-level urban background concentrations. These relationships are within the range of uncertainty calculated during calibration and therefore are not definitive. Measurements at the 3rd floor, 6th floor, and rooftop are not statistically different from one another via a u-test, nor are measurements at street level, the 1st floor, and the reference 6th floor (Table 3-4). Results of the student's t-test display the same relationships, except at street-level versus the reference 6th floor and at roof top versus the 6th floor, where mean concentrations are statistically different.

In winter, there is limited evidence for a vertical gradient in NO_2 concentrations, as 1st floor, 3rd floor, and rooftop concentrations are all lower than at street-level. However, 6th floor Zephyr and reference concentrations, as well as urban background concentrations, are similar to those at street level. 6th floor concentrations are not statistically different than street level or urban background concentrations via both a u-test (Table 3-4) and student's t-test (Table 3.S-1). O_3 concentrations similarly show some evidence of a vertical gradient, opposite in direction to NO_2. All sites have statistically different means and distributions via the t-test and u-test, respectively, except for street level and the urban background, and the 3rd floor and roof top.

On both Frankfurter Allee and Kottbusser Damm, diurnal patterns of NO_2 match those typical of an urban street canyon, except for at the southern lamppost on Frankfurter Allee (Figure 3-4). This Zephyr consistently overpredicted NO_2 at lower concentrations during calibration (Figure 3.S-5) and therefore has a higher uncertainty range than the other units on Frankfurter Allee. All NO_2 concentrations during both experiments are statistically different from one another, except for the 4th floor and the western side of the street on Kottbusser Damm, where the distributions are not statistically different (Table 3-5). Diurnal concentrations on Frankfurter Allee also match expected patterns of O_3 pollution, with a clear early morning dip and late afternoon peak in accordance with known photochemical processes. Concentrations are much higher on the roof and the northern side of the street than the southern, with roof concentrations not significantly different than the urban background (Table 3-5).

3.3.3 Relationship to meteorology

Contained in Table 3-6 are the adjusted R-squared values of generalized additive models (GAMs) modelling measured NO_2 and O_3 concentrations versus five meteorological variables (T, RH, WS, WD, MLH) for each experiment. On Strasse des 17. Juni in summer, these meteorological variables account for 57 – 72% and 83 – 89% of the variance in NO_2 and O_3 concentrations, respectively. The explained variance in winter is slightly less for NO_2 (37 – 51%) and substantially less for O_3 (46 – 73%). On Frankfurter Allee (summer), meteorology explains 33 – 41% and 72 – 87% of the variance in NO_2 and NO_2 concentrations, respectively. On Kottbusser Damm (winter/spring), meteorology explains 37 – 48% of the variance in NO_2 concentrations. Overall, this matches expectations and agrees between LCS and reference measurements. All missing explained variance can likely be attributed to the missing data on source emissions and activity.

Figures 3-5 and 3-6 display partial effects plots of the GAMs for each site and meteorological variable across all the experiments for NO_2 and O_3, respectively. These capture the non-linear effects of each meteorological variable on NO_2 and O_3 across their measured range in each experiment. For NO_2, WD and WS are the most relevant predictors, highlighting the importance of dilution and transport. More stationary air parcels (WS < 2 m s-1) correlate with higher NO_2 concentrations whereas higher WS (> 4 m s-1) correlate with lower NO_2 concentrations. On Strasse des 17. Juni all measurements exhibit similar relationships with WD in both summer and winter, indicating that wind effects due to street morphology affect all Zephyrs similarly. This differs at Frankfurter Allee and Kottbusser Damm, where different sides of the street exhibit different relationships with prevailing background WD, suggesting the local relevance of street morphology on NO_2 and O_3 concentrations.

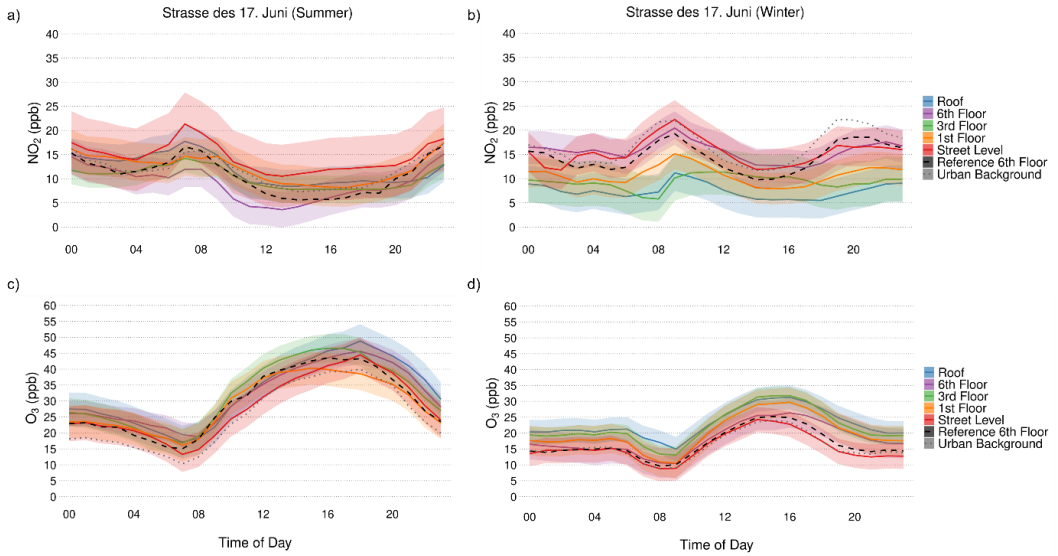

Figure 3-3 Diurnal profiles of NO_2 (a, b) and O_3 (c, d) concentrations at the Strasse des 17. Juni experimental site in Summer (a, c) and Winter (b, d). Dashed lines represent reference NO_2 and O_3 concentrations measured on the 6th Floor of the mathematics building. Dotted lines represent the average urban background concentrations. Shaded areas represent the uncertainty range calculated for each Zephyr using the seven-step methodology.

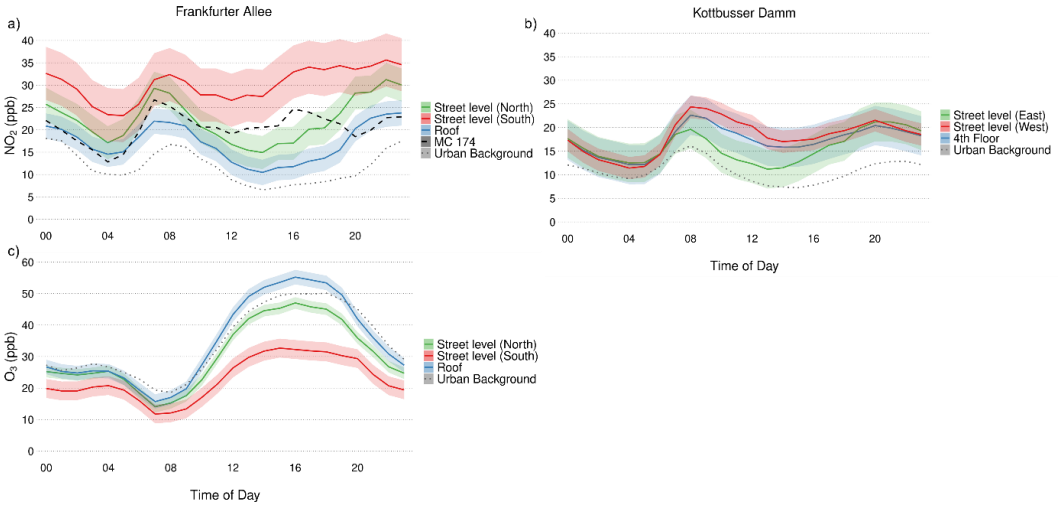

Figure 3-4 Diurnal profiles of NO_2 (a, b) and O_3 (c) concentrations at the Frankfurter Allee (a, c) and Kottbusser Damm (b) experimental sites. Dashed lines represent reference NO_2 and O_3 concentrations measured on the 6th Floor of the mathematics building. Dotted lines represent the average urban background concentrations. Shaded areas represent the uncertainty range calculated for each Zephyr using the seven-step methodology.

Table 3-4 Wilcoxon-Mann-Whitney U-tests of NO_2 and O_3 concentrations at the Strasse des 17. Juni experimental site in Summer and Winter. Extremely small p-values are rounded down to 0.

Strasse des 17. Juni (Summer)

NO_2	Street	1st Floor	3rd Floor	6th Floor	Roof	Reference 6th Floor	Urban Background
Street	1						
1st Floor	0	1					
3rd Floor	0	0	1				
6th Floor	0	0	0	1			
Roof	0	0.020	0	0	1		
Reference 6th Floor	0	0	0.004	0	0	1	
Urban Background	0	0	0.010	0	0.007	0	1

O_3	Street	1st Floor	3rd Floor	6th Floor	Roof	Reference 6th Floor	Urban Background
Street	1						
1st Floor	0.091	1					
3rd Floor	0	0	1				
6th Floor	0	0	0.210	1			
Roof	0	0	0.673	0.087	1		
Reference 6th Floor	0.084	0.886	0	0.007	0	1	
Urban Background	0	0	0	0	0	0	1

Strasse des 17. Juni (Winter)

NO_2	Street	1st Floor	3rd Floor	6th Floor	Roof	Reference 6th Floor	Urban Background
Street	1						
1st Floor	0	1					
3rd Floor	0	0.005	1				
6th Floor	0.063	0	0	1			
Roof	0	0	0	0	1		
Reference 6th Floor	0.008	0	0	0	0	1	
Urban Background	0.001	0	0	0.052	0	0	1

O_3	Street	1st Floor	3rd Floor	6th Floor	Roof	Reference 6th Floor	Urban Background
Street	1						
1st Floor	0	1					
3rd Floor	0	0	1				
6th Floor	0	0	0	1			
Roof	0	0	0.818	0	1		
Reference 6th Floor	0.024	0	0	0.001	0	1	
Urban Background	0.615	0	0	0	0	0.054	1

Table 3-5 Wilcoxon-Mann-Whitney U-tests of NO_2 and O_3 concentrations at Frankfurter Allee and Kottbusser Damm. No O_3 data could be captured for the Kottbusser Damm campaign. Extremely small p-values are rounded down to 0.

	Frankfurter Allee				
NO_2	**Street (North)**	**Street (South)**	**Roof**	**MC 174**	**Urban Background**
Street (North)	1				
Street (South)	0	1			
Roof	0	0	1		
MC 174	0.042	0	0	1	
Urban Background	0	0	0	0	1

	Kottbusser Damm			
NO_2	**Street (West)**	**Street (East)**	**4th Floor**	**Urban Background**
Street (West)	1			
Street (East)	0	1		
4th Floor	0.163	0	1	
Urban Background	0	0	0	1

O_3	**Street (North)**	**Street (South)**	**Roof**	**Urban Background**
Street (North)	1			
Street (South)	0	1		
Roof	0	0	1	
Urban Background	0	0	0.891	1

Table 3-6 Adjusted R-squared values for the generalized additive models (GAM) built using measured NO_2 and O_3 concentrations and five explanatory meteorological variables (T, RH, WS, WD, MLH) for each site and experiment, where data were available.

	Strasse des 17. Juni (Summer)						
	Street	**1st Floor**	**3rd Floor**	**6th Floor**	**Roof**	**Reference**	**UB**
NO_2 Adj.-R^2	0.57	0.63	0.72	0.58	0.64	0.63	0.59
O_3 Adj.-R^2	0.89	0.84	0.83	0.89	0.89	0.80	0.81
	Strasse des 17. Juni (Winter)						
	Street	**1st Floor**	**3rd Floor**	**6th Floor**	**Roof**	**Reference**	**UB**
NO_2 Adj.-R^2	0.37	0.48	0.53	0.45	0.51	0.48	0.44
O_3 Adj.-R^2	0.46	0.62	0.59	0.73	0.58	0.58	0.55
	Frankfurter Allee						
	Street (North)	**Street (South)**	**Roof**	**MC 174**	**UB**		
NO_2 Adj.-R^2	0.33	0.41	0.41	0.51	0.63		
O_3 Adj.-R^2	0.80	0.72	0.87		0.85		
	Kottbusser Damm						
	Street (West)	**Street (East)**	**4th Floor**	**UB**			
NO_2 Adj.-R^2	0.37	0.45	0.48	0.56			

Some ambient temperature effects on NO_2 concentrations are seen across the experiments, particularly at low temperatures reflecting night-time conditions, but are strongest at Frankfurter Allee, the sole experiment with temperatures exceeding 30°C. The partial effects plot in Figure 3-5, panel k indicates a stronger relationship with temperature on the southern side of the street by both the Zephyr and the reference instrument (MC 117) than on the northern side of the street. No strong influences of MLH are seen across the experiments, except on Frankfurter Allee (Figure 3-5, panel o), the only experiment with a recorded MLH above 3 km. Under elevated MLH conditions NO_2 concentrations increase slightly at all sites. However, confidence intervals for this relationship are very large due to the small sample size of datapoints with MLH > 3 km.

For O_3 in summer (Strasse des 17. Juni; Frankfurter Allee), the majority of the variance in concentrations is explained by ambient temperature, whereas the other meteorological variables are significant but only marginally important. In winter (Strasse des 17. Juni), RH, WS, and WD are of similar importance in explaining the variance in concentrations. The predictor of lowest significance for explaining O_3 at all sites in all experiments is MLH.

Partial effects for ambient temperature are clear in both summer experiments (Figure 3-6, panels a, k) at all sites, exhibiting a strong correlation with O_3 concentrations. In all three experiments, this relationship is reversed for ambient RH; values < 60% are associated with higher concentrations and values > 90% associated with lower concentrations, though this is strongest in winter. Wind speeds below 2 m s-1 and above 5 m s-1 are associated with lower and higher O_3 concentrations, respectively, across the three experiments. While there are no differences between sites with regards to WD on Strasse des 17. Juni, these are evident on Frankfurter Allee. No clear relationship on the northern side of the street can be discerned, but O_3 concentrations on the southern side of the street exhibit a strong dependence on wind direction.

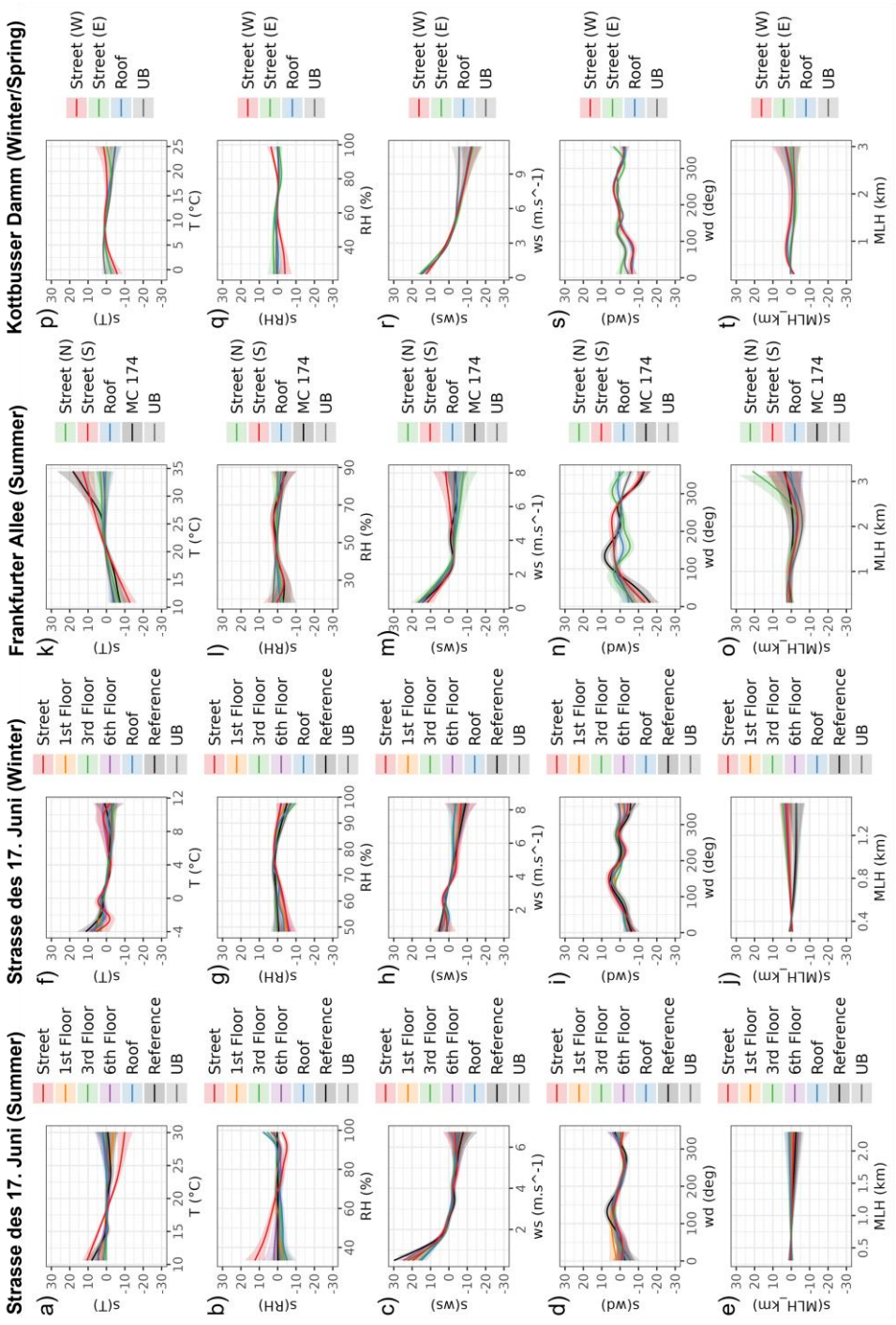

Figure 3-5 Partial effects plots of each meteorological variable as calculated by the generalized additive models (GAMs) for NO_2 for each site at Strasse des 17. Juni in Summer (a-e) and Winter (f-j), Frankfurter Allee (k-o), and Kottbusser Damm (p-t).

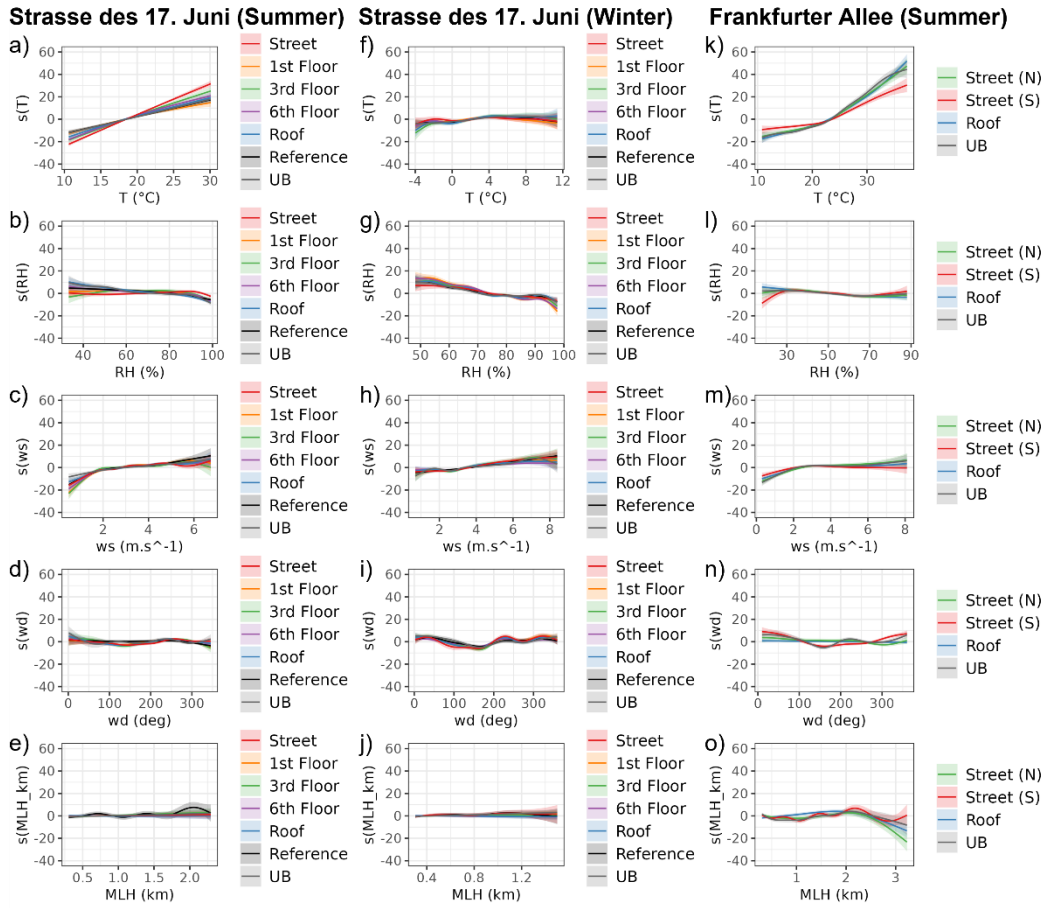

Figure 3-6 Partial effects plots of each meteorological variable as calculated by the generalized additive models (GAMs) for O_3 for each site at Strasse des 17. Juni in Summer (a-e) and Winter (f-j), and Frankfurter Allee (k-o).

3.3.4 Relationship to urban morphology

Road-to-roof ratios of hourly NO_2 and O_3 concentrations across the experiments can be seen in Table 3-7. The ratio for NO_2 is greater than 1 in each case, indicating that concentrations are diluted at elevated heights compared to street-level, except for Kottbusser Damm. At this site, the 'roof' concentrations are from the 4th floor and therefore are still within the street canyon. The NO_2 ratios reflect this, indicating that the concentrations at the 4th floor are roughly equal to those at street-level. On Strasse des 17. Juni the ratio is greater in the winter than in the summer. On Frankfurter Allee the NO_2 ratios are greater on the southern side of the street than on the northern. For O_3, ratios are always less than 1, indicating higher concentrations on the roof compared to street level. At Strasse des 17. Juni, the winter ratio is smaller than the summer ratio. On Frankfurter Allee, the O_3 ratio on the southern side of the street is much lower than on the northern side.

Table 3-7 Road-to-roof ratios of NO_2 and O_3 concentrations at each experimental site. *On Kottbusser Damm the "roof" concentrations used are those from the 4th floor and therefore still within the street canyon.

	Strasse des 17. Juni (Summer)	Strasse des 17. Juni (Winter)	Frankfurter Allee (N)	Frankfurter Allee (S)	Kottbusser Damm (W)*	Kottbusser Damm (E)*
NO_2	1.34	2.53	1.33	2.03	1.05	0.95
O_3	0.87	0.66	0.88	0.69		

While these ratios indicate influences of street canyon morphology on overall NO_2 and O_3 concentrations, closer analysis reveals more about the nature of these relationships. Figure 3-7 shows boxplots of NO_2 and O_3 hourly concentrations for Frankfurter Allee and Kottbusser Damm broken down by daily average cardinal wind direction (e. g. NESW) and various wind speeds. As can be seen in panels a, c, and e, the relative relationships between concentrations at the various Zephyr locations in each street canyon change with prevailing wind direction, suggesting the influence of local urban morphology on street canyon pollution circulation patterns. On Frankfurter Allee, NO_2 concentrations on the southern side of the street are highest with wind blowing from the South and lowest with wind coming from the North. In the latter case, NO_2 concentrations are not significantly different on opposite sides of the street. All other sites are statistically significant ($p < 0.05$) from one another using a Wilcoxon-Mann-Whitney test (Tables 3.S-3 – 3.S-8), regardless of wind direction. NO_2 concentrations at roof level do not vary with wind direction and are consistently higher than the urban background.

O_3 concentrations on Frankfurter Allee (Figure 3-7, panel c) follow similar but inverse patterns with regards to prevailing wind direction. With wind from the North, street-level O_3 concentrations are not significantly different from one another, whereas with wind from the South, concentrations at street-level and on the roof on the northern side of Frankfurter Allee are not significantly different from one another or from the urban background. When wind blows parallel to the street from the East, there is no statistically significant difference between concentrations at the roof and the northern side of the street. Roof concentrations are lower than those of the urban background with wind blowing from all directions except from the South and West, where there is no statistically significant difference between them. At all sensor locations and in the urban background, O_3 concentrations were lowest with wind blowing from the West.

On Kottbusser Damm (Figure 3-7, panel e), clear differences between NO_2 concentrations on the eastern and western sides of the street are seen with varying wind direction. With wind from the North and East, concentrations on the eastern side are higher, while the reverse is true when wind blows from the South and West. However, the relationships from the North and East are not statistically significant ($p < 0.05$). Concentrations at the 4^{th} floor are not statistically different from those at street-level on the western side of the street, regardless of wind direction, and are consistently higher than the urban background.

On Frankfurter Allee and Kottbusser Damm, prevailing wind speed does not affect the relative relationships of NO_2 and O_3 concentrations between sensor locations. At both sites, NO_2 concentrations decrease with increasing wind speed, as evidenced by smaller interquartile ranges of concentrations at higher wind speeds, lower median concentrations, and less extreme peaks. These patterns are matched in the urban background and agree with results from the GAMs (Section 3.3.3). At wind speeds < 2 m s-1, there are no statistically significant differences between all street-level NO_2 and O_3 concentrations and the elevated concentrations (rooftop on Frankfurter Allee and 4^{th} floor on Kottbusser Damm). Only at wind speeds from 4-6 m s-1 are rooftop O_3 concentrations significantly different from the urban background.

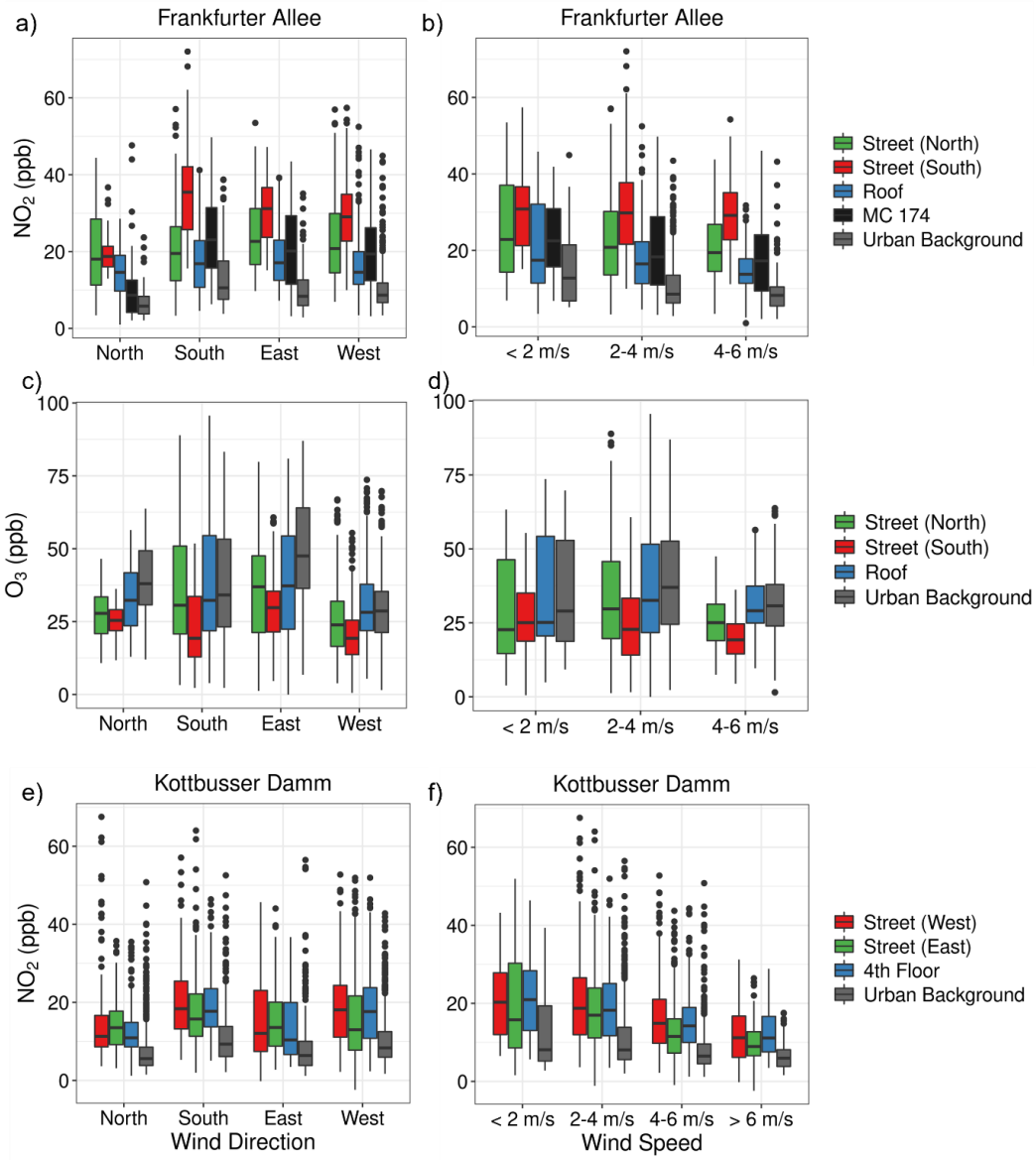

Figure 3-7 Boxplots of Zephyr, reference, and urban background NO2 and O3 hourly concentrations for the Frankfurter Allee (a-d) and Kottbusser Damm (e, f) campaigns, subset by cardinal wind direction and various wind speeds.

3.3.5 Insolation and O_3 / NO_2 concentrations

A closer analysis of the internal Zephyr temperature (Figure 3-8 for co-locations and experiments) profiles on Frankfurter Allee and Kottbusser Damm was completed to further understand differences in overall NO_2 and O_3 concentrations. As can be seen in Figure 3-8, the internal temperature profiles of the units match up during co-locations and differ substantially during experimental deployment. During the Frankfurter Allee experiment, the unit on the southern side of the street experienced temperatures up to 5-7 K cooler than the units on the northern side of the street and the roof during the day, with temperatures equalizing at night. Therefore, these profiles suggest that internal Zephyr temperature can be evaluated as a proxy for direct insolation. As such, we hypothesize that less photolysis would occur on the southern side of Frankfurter Allee, which experiences far more shading than on the northern side, leading to lower total O_3 production and elevated NO_2 concentrations. While this could also occur as a result of high temperatures (>30 °C) influencing EC performance, no strong influence of high temperature was seen on NO_2 concentrations by Zephyrs located on the northern side of

the street (Figure 3-5). To account for the impact of temperature on sensor functioning, a sensitivity analysis was conducted in which internal Zephyr T was removed from calibration models and new concentrations were predicted. This revealed no differences in NO_2 concentrations at this site, and slight reductions in O_3 concentrations on the northern side of the street, but with no change to the significant differences between Zephyr locations (Figure 3.S-11). However, the Zephyr on the southern side of the street exhibited poorer performance for predicting NO_2 (Figures 3.S-5 and 3.S-9) and therefore this relationship for NO_2 is not conclusive.

On Kottbusser Damm, the same impact of photolysis can be found for the lower NO_2 concentrations on the eastern side of the street and is more suggestive due to improved sensor performance (Figures 3.S-6 and 3.S-10). The internal temperature of this Zephyr peaks three to four hours later than the western street-level and 4th floor units, respectively, more NO_2 can be photolyzed over the course of the day. As can be seen in Figure 3-4, this peak in insolation matches with a respective reduction in NO_2 concentrations, which equalize at night with the concentrations on the western side of the street and the 4th floor. No sensitivity analysis was conducted here, as internal Zephyr temperature was not included in the original model of best-fit, as it was removed during model selection.

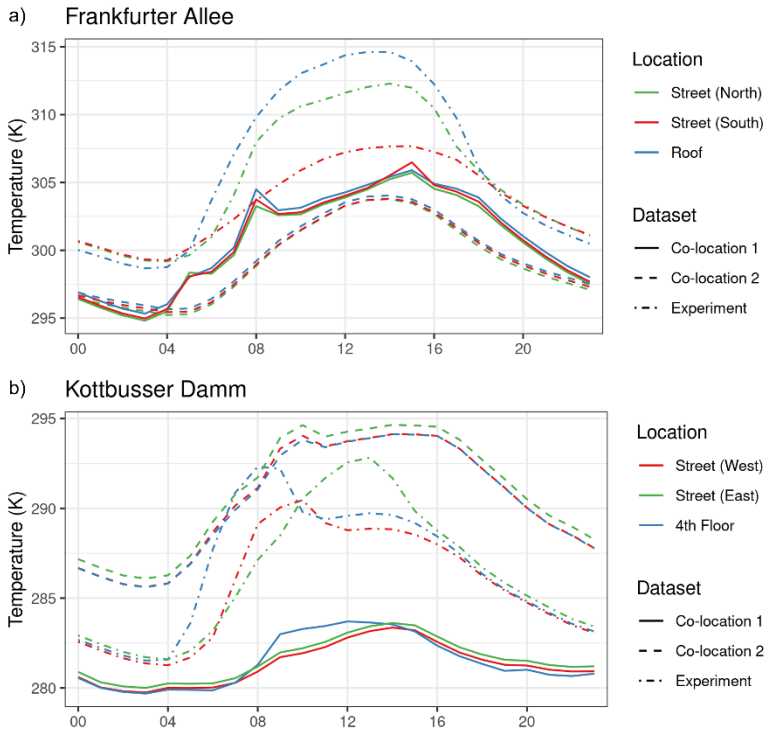

Figure 3-8 Diurnal profiles of the Zephyrs' internal temperature deployed at Frankfurter Allee (a) and Kottbusser Damm (b). Shown are profiles for the co-locations and for the experiments. During co-location 1 for the Kottbusser Damm campaign, the Zephyr deployed on the 4th floor was at a different BLUME station than the other 2 units. In the 2nd co-location all three were at the same BLUME station.

3.4 Discussion

There are several key results of this study of value to the expanding field of LCS literature. First, while many studies have evaluated the calibration of LCS and tested their accuracy in field measurement campaigns (J. M. Barcelo-Ordinas et al., 2019; De Vito et al., 2018;

Mahajan et al., 2021; Ripoll et al., 2019), few have assessed to what extent LCS capture expected patterns of urban pollution or how their measurements are influenced by meteorology, and none have deployed them in urban street canyons. Our findings presented here show that both MOS and EC LCS accurately measure typical diurnal patterns of NO_2 and O_3 pollution in urban environments following proper calibration with reference instrumentation. However, MOS tend to be less accurate and more variable in performance than EC, which matches recent research (Russell et al., 2022), therefore we deem the former inappropriate for discerning high spatial resolution differences in air pollution.

Second, these general patterns in pollution can to a large extent be explained by micro-scale meteorology, as evidenced by the GAMs built in this study. The total variance explained by meteorology for the LCS agrees well with the amount explained for reference measurements at the experimental sites and in the urban background. Furthermore, the confluence of insolation and wind direction appear to play an important role in the variability of NO_2 and O_3 concentrations in street canyons. On Frankfurter Allee, regardless of prevailing wind direction, NO_2 and O_3 concentrations are consistently higher and lower, respectively, on the southern side of the street than on the northern side. On Kottbusser Damm, while concentrations are significantly different on opposite sides of the street, they equalize at night and vary substantially during the day. This suggests that photochemical processes play an important role in the variability of NO_2 and O_3 concentrations in street canyons (Zhang et al., 2020), but stronger evidence is needed for this to be conclusive.

For O_3, previous studies found strong positive and negative relationships between O_3 concentrations and T and RH, respectively (Melkonyan & Kuttler, 2012; Noelia Otero et al., 2018; N. Otero et al., 2016). This relationship is particularly strong in summer when T is strongly correlated with insolation, leading to photochemical O_3 production, and weaker in winter, which our study confirms. Therefore, local production contributes much more to O_3 concentrations in summer and less so in winter. We did not find a negative relationship between O_3 and WS or a positive relationship with MLH, as found in other studies (Geiß et al., 2017; Melkonyan & Kuttler, 2012; N. Otero et al., 2016).

For NO_2, our results agree with other studies that the most dominant meteorological factor is wind speed (Carslaw et al., 2007; Elminir, 2006; Pearce et al., 2011). Also consistent with the literature are the lack of strong relationships with T, RH, and MLH. In the latter case, our results agree with Wagner and Schäfer (2017) that MLH does not have an impact on NO_2 concentrations. We find minor evidence of a relationship with NO_2 when MLH is greater than 3 km in height but lack enough data to say this conclusively. In general, less variance is explained by meteorology for NO_2 than for O_3, which accurately reflects the greater role of production from local emissions and short-range transport for NO_x and of photochemical production for O_3. Furthermore, less variance is explained by meteorology for O_3 in the winter than in the summer, further highlighting the role of photochemical production on local concentrations. The strong relationship with wind speed indicates that dispersal and dilution are most important in explaining NO_2 concentrations from a local meteorological perspective.

Third, our measurements show that LCS can capture micro-scale differences in NO_2 and O_3 concentrations in street canyons. Prevailing wind direction and speed have statistically significant impacts on the concentrations measured by each Zephyr on both Frankfurter Allee and Kottbusser Damm. In both cases, NO_2 and O_3 concentrations are higher and

lower, respectively, on the leeward side of the street when wind blows perpendicularly, which matches expectations from previous studies modelling street canyon flow (Bright et al., 2011; Gonzalez Olivardia et al., 2019; Y.-D. Huang et al., 2019; Park et al., 2019; S.-J. Park et al., 2015; Voordeckers, Lauriks, et al., 2021; Voordeckers, Meysman, Billen, Tytgat, & Van Acker, 2021). Measured differences between Zephyrs on these two streets are reduced or statistically insignificant with parallel wind flow.

We also find indications of a strong influence of insolation on O_3 and NO_2 concentrations using the Zephyrs' measure of internal temperature as a proxy. On Frankfurter Allee, daytime O_3 concentrations are consistently lower on the southern side of the street than on the northern side and on the roof. Zephyr internal temperature profiles reveal that this side of the street is exposed to less, if any, direct sunlight than the northern side. As such, less NO_2 is photolyzed and thereby less O_3 is produced. Similarly, on Kottbusser Damm, daytime NO_2 concentrations on the eastern side of the street are lower than on the western side, with virtually equal night-time concentrations, indicating greater daytime photolysis on the eastern side. However, this should be considered as preliminary evidence, as Zephyr internal temperature is only a proxy for insolation, not a direct measurement of it.

3.4.1 Implications for future research with LCS

These results of this exploratory study are valuable to the growing field of LCS research for several reasons. Primarily, end-users of MOS and EC LCS, including actors outside of the academic arena, can be confident that LCS accurately capture general patterns of urban NO_2 and O_3 pollution, matching diurnal trends and meteorological influences. While LCS cannot be used to determine exact magnitudes of concentrations and associated changes within the necessary uncertainty requirements for monitoring compliance with limit-values, they are appropriate for general measurement of urban air pollution. Two caveats to this are the need for rigorous field calibration with reference instrumentation, as suggested in Schmitz, Towers, et al. (2021b), and the relatively short lifetime of LCS, which can increase long-term operational costs (Peltier, 2020).

The ability of LCS to measure expected changes and differences in micro-scale urban environments, such as in street canyons in this study, brings further added value to this field of research. We find that LCS can be deployed at high spatial resolution to understand differences in air pollution in varying urban environments. Few studies have measured air pollution at this spatial resolution using LCS, highlighting the novelty of these results and underscoring the versatility of LCS. Furthermore, it underscores the potential for LCS to be deployed in many different contexts by various end-users, without the need for complex and computationally intensive modelling. In fact, LCS could potentially be used to validate urban-scale models and inform their further development. While this exploratory study does not combine modelling exercises with LCS measurements to avoid further enlarging its scope, we recommend future work investigate this line of research. Additionally, LCS could be deployed to measure changes in air pollution in connection with changes to physical and transport infrastructure in urban environments.

3.4.2 Limitations

There are several limitations to this study. While the calibration of the LCS has been optimized using the seven-step method outlined in Schmitz, Towers, et al. (2021b), there is still a range of uncertainty associated with these measurements that limits the scope of

our conclusions. The uncertainty varies between experiments and sensors and is subject to the quality and quantity of available training data during calibration. This is particularly relevant for MOS sensors, which show much greater inter-sensor variability in performance and sensitivity. EC sensors are more consistent and perform better overall when compared with reference instruments. As such, the consistency in measured concentrations across Zephyrs and agreement with reference and urban background concentrations, particularly for EC sensors and less so for MOS, gives us confidence that the results presented here accurately reflect changes in local pollution.

Due to a combination of financial and logistical reasons, the number of deployable Zephyrs in this study was constrained. As such, this study sought to identify indications of LCS performance in measuring NO_2 and O_3 gradients in the following three street canyons, but future studies should include greater numbers of deployed sensor systems to clarify the preliminary results provided here. Moreover, deploying instrumentation for the measurement of local meteorology (e.g., T, RH, WS, WD) alongside each sensor system is highly recommended. In this study, for similar reasons, this was not feasible.

For the assessment of impacts of meteorology on NO_2 and O_3 concentrations, there were two key limitations to this research. First, on-site meteorological data were only available at Strasse des 17. Juni and not at Frankfurter Allee or Kottbusser Damm. While data from the weather station several kilometres away from each site proved sufficient to glean indications into street canyon pollution dispersal dynamics, future research should seek to measure meteorological parameters alongside LCS. Second, no real-time emissions or traffic data were available for implementation in this work. Without such data, further insights into LCS performance in measuring street canyon pollution (e.g., for validating urban-scale models) will remain limited. Future research must also include emissions data and traffic counts to complete the picture outlined in this study.

During the measurement campaign on Kottbusser Damm, there was a technical failure with the reference O_3 instrument during co-location that prevented the proper calibration of O_3 concentrations with the Zephyrs. Without O_3 data for this campaign, a more complete understanding of air pollution in this street canyon was not possible, nor was a full comparison to the campaign on Frankfurter Allee. Furthermore, during this campaign health and safety measures in response to the COVID-19 pandemic such as lockdowns were enacted that impacted the levels of air pollution on the street. Data collected during the strictest lockdowns were removed from this analysis, but it is possible that other indirect effects on air pollution, such as changing mobility habits and the construction of a pop-up bike lane, impacted the results presented here.

3.5 Conclusions

In this exploratory study, two different types of LCS housed in EarthSense Zephyrs were deployed in a novel application to measure NO_2 and O_3 pollution in three different urban environments in Berlin between 2017 – 2020. We find that properly calibrated LCS measure expected levels of urban NO_2 and O_3 pollution consistently and agree with reference measurements on site and in the urban background well. A substantial amount of variance in measured NO_2 and O_3 concentrations is explained by local meteorology across all experiments, with good agreement between LCS and reference measurements. As found in other studies, wind speed is the most important predictor of NO_2 concentrations, whereas temperature is most important for explaining O_3 in summer. In

winter, less total variance in O_3 concentrations is explained by meteorology, with RH, wind speed, and wind direction of greater importance.

This study also assessed measurements of NO_2 and O_3 in two street canyons in Berlin, finding good agreement with known patterns of street canyon pollution described in various studies modelling wind flow and pollution distribution. Specific preliminary results include: 1) identifying significant differences in NO_2 and O_3 concentrations on opposite sides of the street; 2) varying relationships between NO_2 concentrations at street-level and on the roof or the 4th floor; and 3) measurable impacts of insolation on O_3 concentrations on different sides of the street. These results appear to be attributable to influences of urban morphology and wind flow within the street canyons, but are not conclusive due to the several key limitations to the study. However, these indications still highlight the ability of LCS to reliably measure urban air pollution at high spatial resolution, which should continue to be taken advantage of in future research.

As the academic and commercial usage of LCS continues to expand, it is crucial to continue to better understand their potential. This exploratory study shows that measurements of NO_2 and O_3 pollution using MOS and EC LCS match expected diurnal patterns in urban environments, align with known relationships to micro-scale meteorological parameters, and can capture high spatial resolution details of street canyon pollution. Future research should expand on this study by including calibrated measurements of nitrogen monoxide (NO) and particulate matter (PM), as well as include higher spatial resolution in-situ meteorological data in the assessment of street canyon pollution. Future efforts should also consider connecting LCS measurements and urban-scale models. End-users, including researchers and citizens, can be confident that measurements they take with LCS are reliably capturing general patterns and trends in air pollution, provided they are properly calibrated with reference instrumentation. As a result, potential new applications for LCS arise, such as validating urban pollution models, understanding air pollution patterns in diverse micro-scale urban environments, and connecting measurements to changes in the physical urban environment.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to thank Dr. Katja Grunow, Philipp Tödter, and Marcel Kyrsiak from the Senate Department for the Environment, Urban Mobility, Consumer Protection and Climate Action for their assistance during co-location at the BLUME stations; Dr. Roland Leigh, Dr. Jordan White, and the whole EarthSense team for their continued support of this work; Dr. Edward C. Chan for his input and discussions on the subject; Tarik Mustafa and Martin Wittau of the German Federal Association for Sustainability for use of their 4th floor window on Kottbusser Damm;

Funding

The research of Seán Schmitz, Guillermo Villena, Alexandre Caseiro, and Erika von Schneidemesser is supported by RIFS, with financial support provided by the Federal Ministry of Education and Research of Germany (BMBF) and the Ministry for Science, Research and Culture of the State of Brandenburg (MWFK). Much of the research was carried out as part of the research program “Urban Climate Under Change” [UC]^2,

contributing to “Research for Sustainable Development” (FONA; https://www.fona.de/de/, last access: October 4th, 2022), within the joint-research project “Three-dimensional observation of atmospheric processes in cities (3DO)”, funded by the German Federal Ministry of Education and Research (BMBF) under grant 01LP1602E.

3.S. Supplemental information

3.S.1 Section 1: Overview of experiments and co-locations.

Figure 3.S-1 Set-up of the co-location of the Zephyrs with reference instruments on the 6th floor of the mathematics building. The grey units are the Zephyrs, and the two inlet tubes bring air to the reference devices located inside the office.

Figure 3.S-2 Set-up of the co-location of the Zephyrs at BLUME station MC 117 (Schildhornstrasse). A similar set-up was used at MC 174 (Frankfurter Allee) during the co-locations on Frankfurter Allee. An inlet to the right of frame draws air into the station for the reference instruments to use.

3.S.2 Section 2: Assessment of model performance across experiments

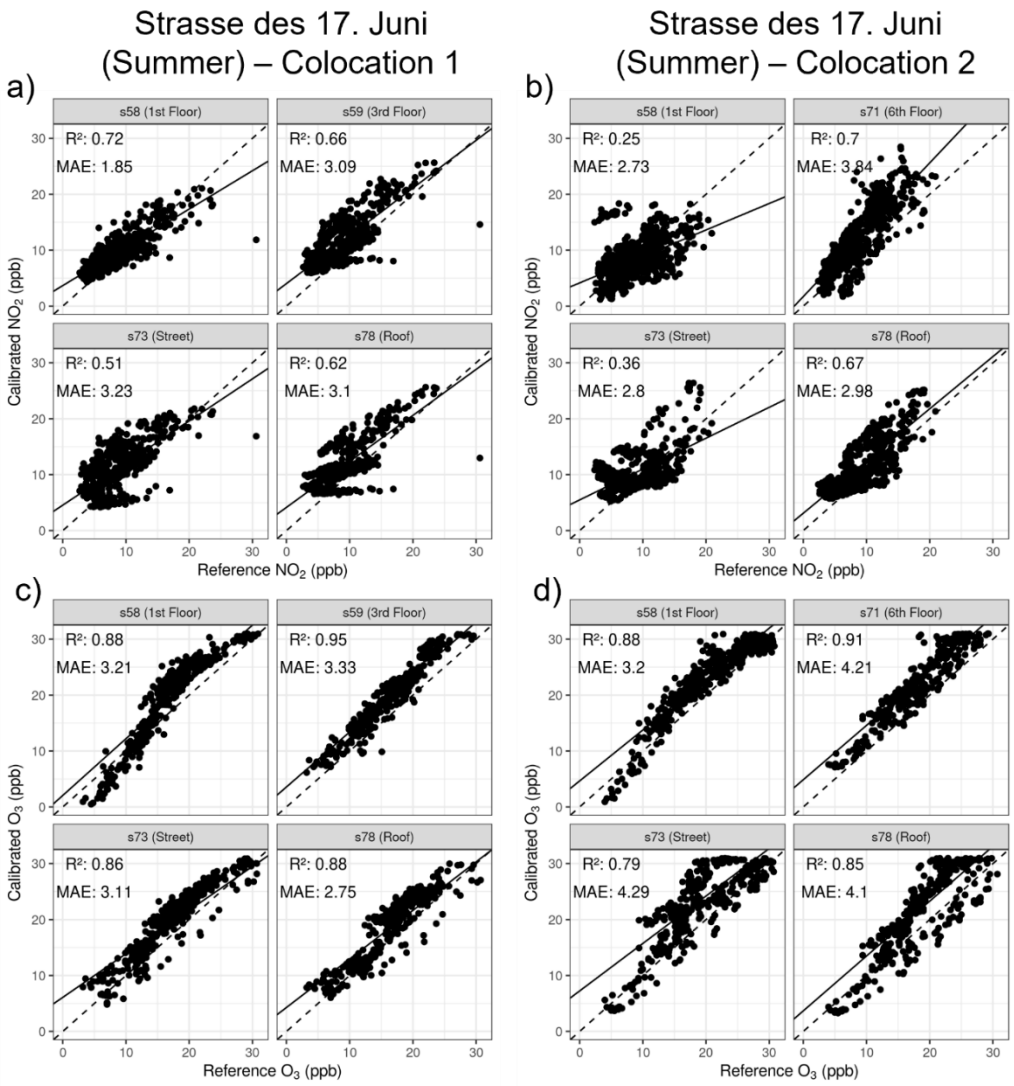

Figure 3.S-3 Scatter plots of calibrated Zephyr NO_2 (a,b) and O_3 (c,d) concentrations versus reference concentrations for a validation set in each co-location of the Summer, 2017 experiment on Strasse des 17. Junis. Model performance metrics of R^2 and mean average error (MAE) are given for each Zephyr. The validation set is composed of the final 25% of the data for co-location 1, which occurred prior to the experiment, and the initial 25% of the data for co-location 2, which occurred following the experiment. The units are labeled by their id number and their location during the experimental window.

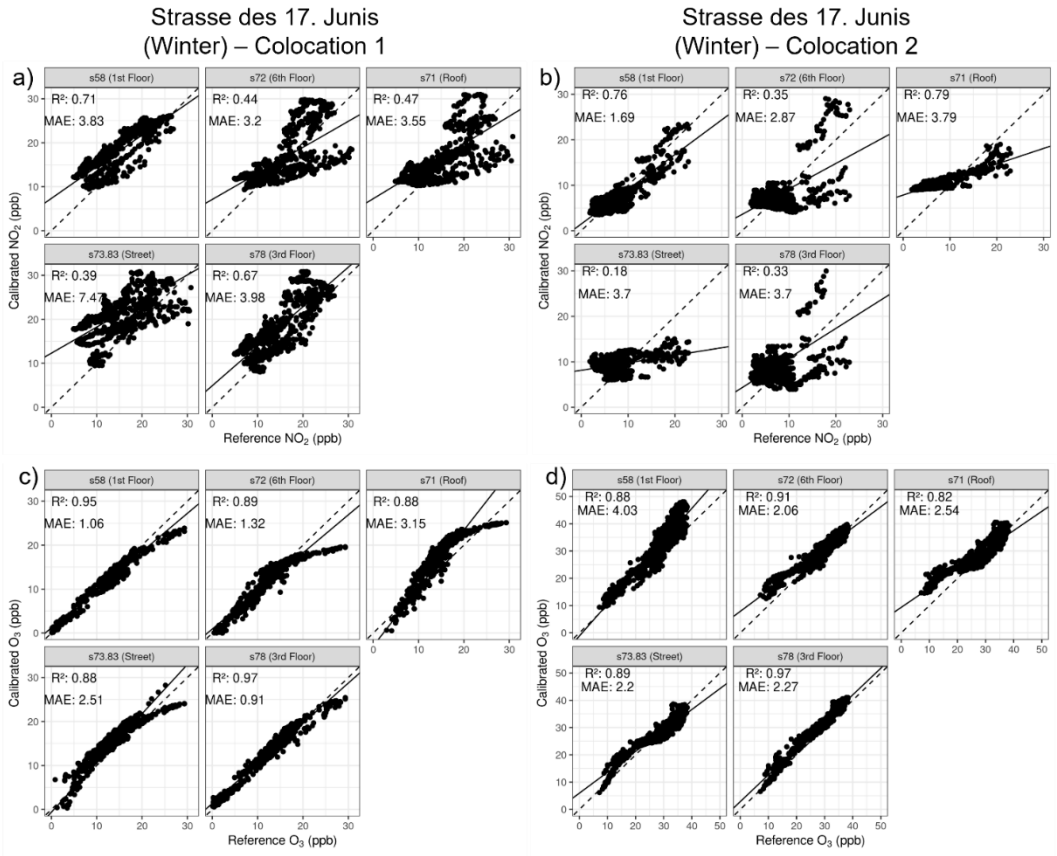

Figure 3.S-4 . As with Figure S3 for NO_2 (a,b) and O_3 (c,d) concentrations on Strasse des 17. Juni in Winter, 2018.

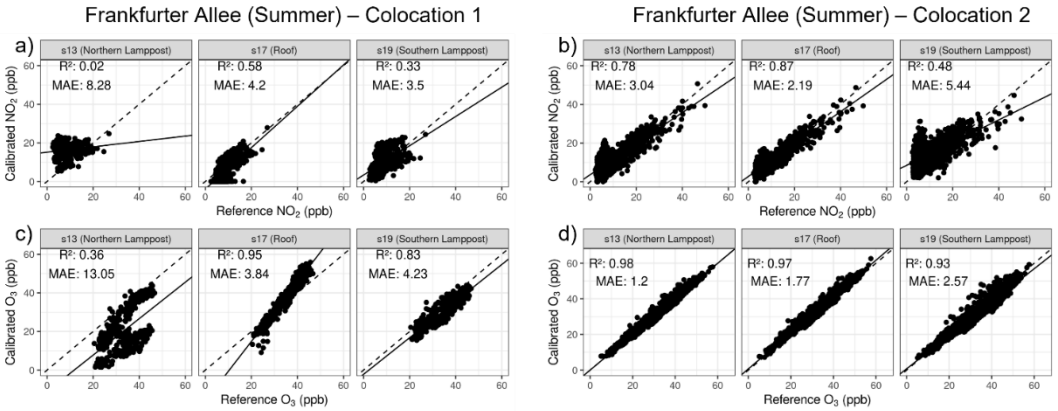

Figure 3.S-5 As with Figure S3 for NO_2 (a,b) and O_3 (c,d) concentrations on Frankfurter Allee in Summer, 2018.

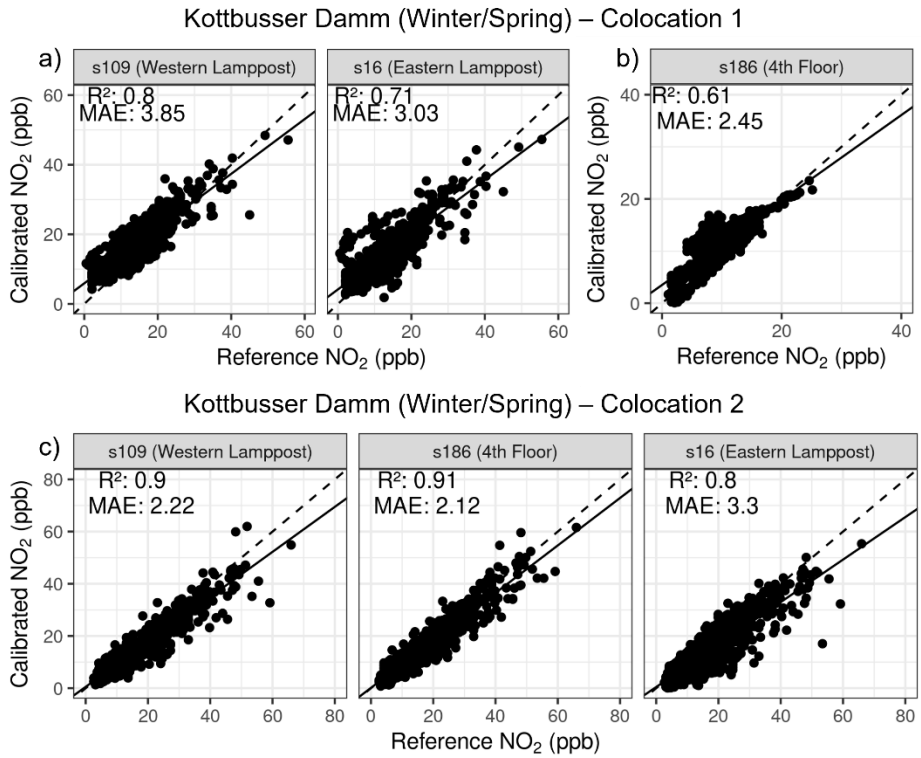

Figure 3.S-6 As with Figure S3 for NO_2 concentrations on Kottbusser Damm in Winter and Spring, 2020.

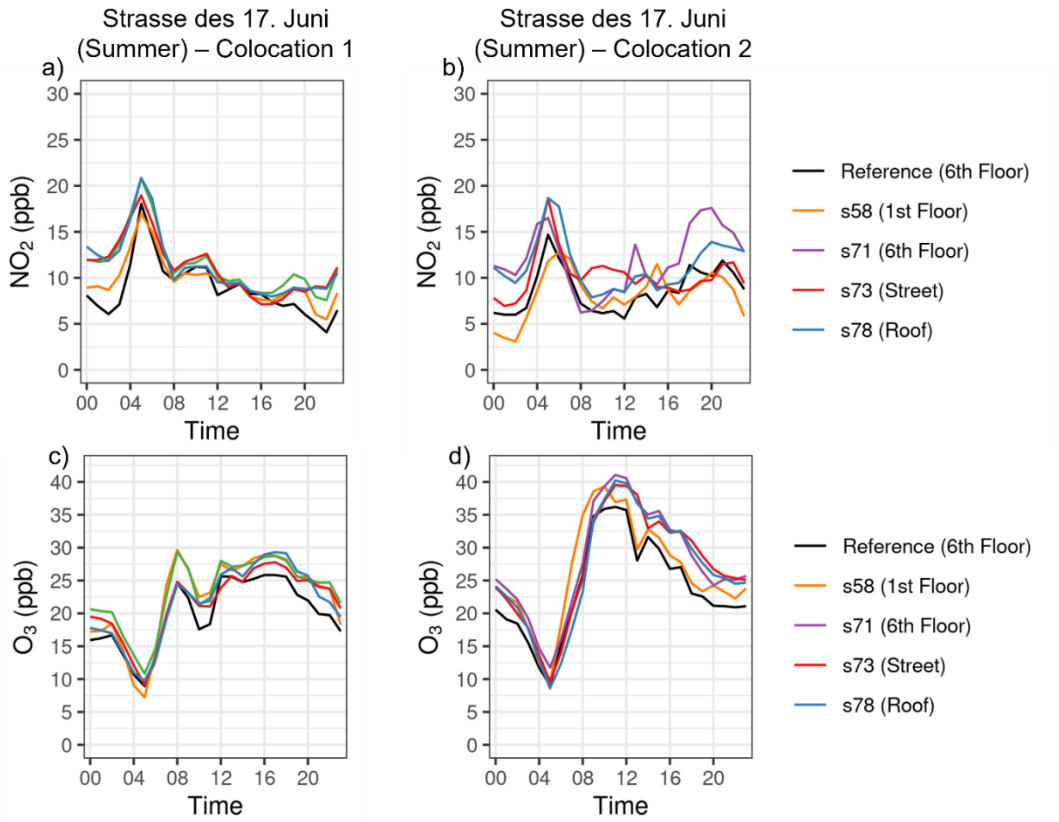

Figure 3.S-7 Diurnal plots of calibrated Zephyr NO_2 (a,b) and O_3 (c,d) concentrations versus reference concentrations for a validation set in each co-location of the Summer, 2017 experiment on Strasse des 17. Junis. The validation set is composed of the final 25% of the data for co-location 1, which occurred prior to the experiment, and the initial 25% of the data for co-location 2, which occurred following the experiment. The units are labeled by their id number and their location during the experimental window.

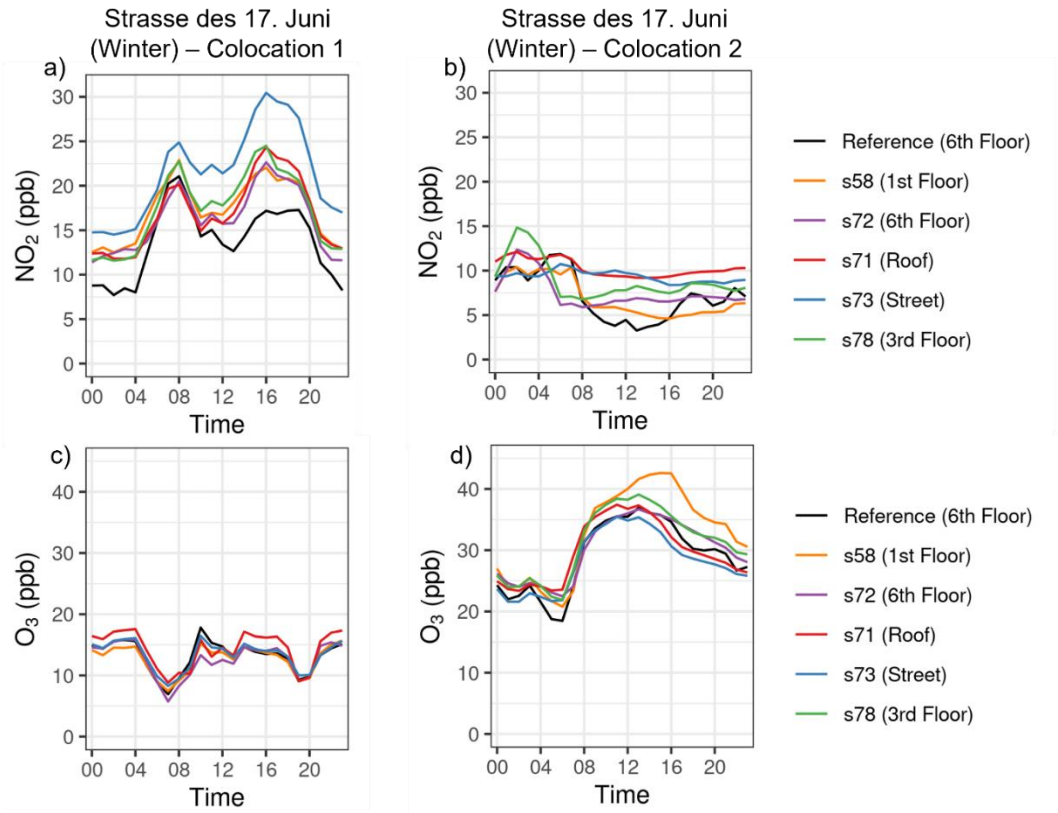

Figure 3.S-8 As with Figure 3.S-7 for NO_2 (a,b) and O_3 (c,d) concentrations on Strasse des 17. Juni in Winter.

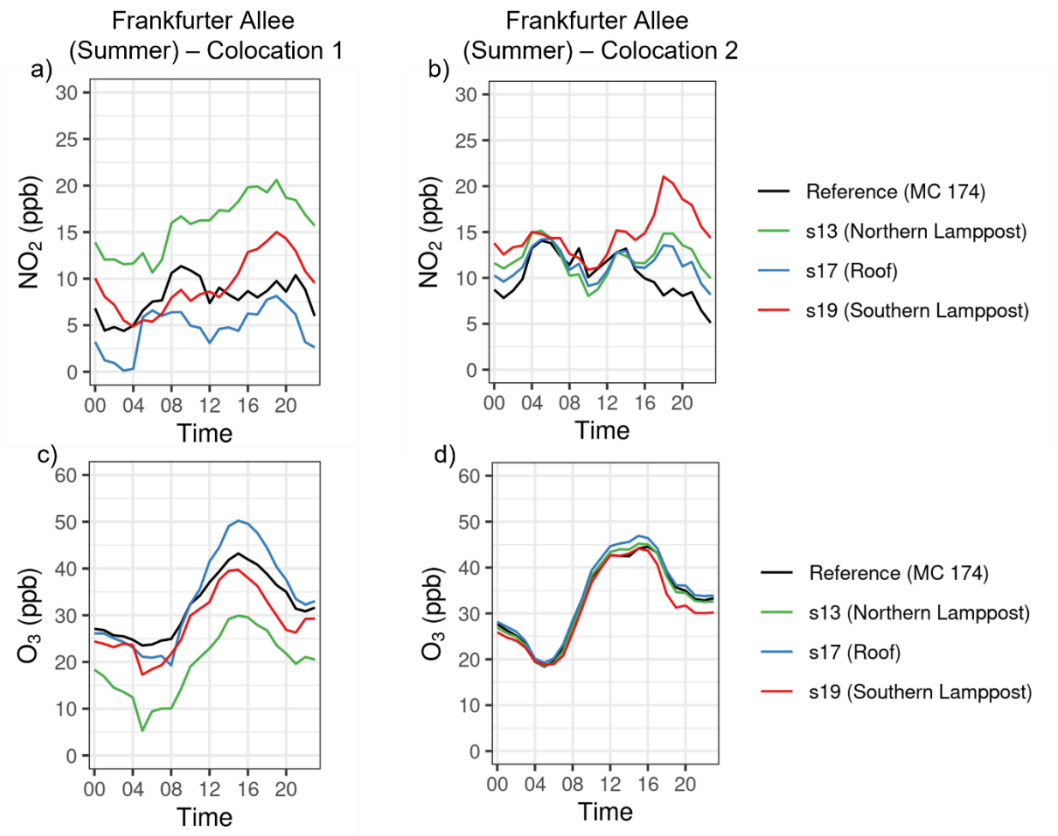

Figure 3.S-9 As with Figure 3.S-7 for NO_2 (a,b) and O3 (c,d) concentrations on Frankfurter Allee in Summer, 2018.

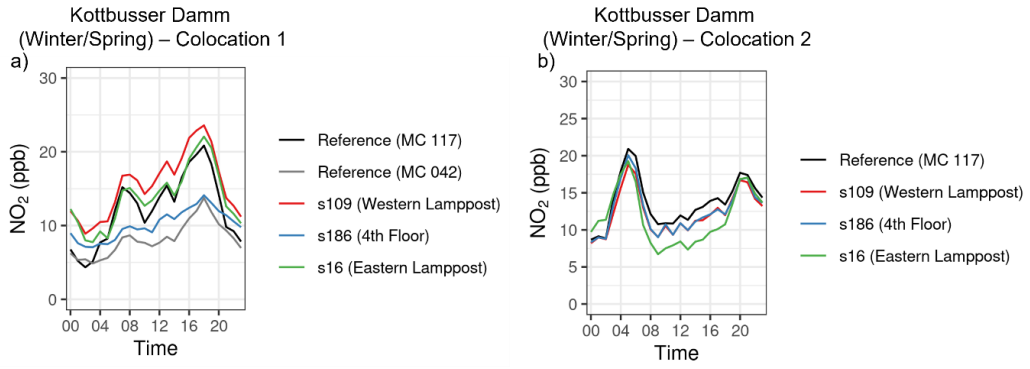

Figure 3.S-10 As with Figure 3.S-7 for NO_2 (a,b) concentrations on Kottbusser Damm in Winter/Spring, 2020. During the first co-location, unit s186 was co-located at a different BLUME station (MC 042) than the other two. All three were co-located at MC 117 during the second co-location.

Table 3.S-1 Student's t-tests of NO_2 and O_3 concentrations at the Strasse des 17. Juni experimental site in Summer and Winter. Extremely small p-values are rounded down to 0.

NO_2 Summer	Street	1st Floor	3rd Floor	6th Floor	Roof	Reference 6th Floor	Urban Background
Street	1						
1st Floor	0	1					
3rd Floor	0	0	1				
6th Floor	0	0	0.037	1			
Roof	0	0.232	0	0	1		
Reference 6th Floor	0	0.001	0.121	0.003	0.015	1	
Urban Background	0	0.08	0	0	0.541	0.068	1

NO_2 Winter	Street	1st Floor	3rd Floor	6th Floor	Roof	Reference 6th Floor	Urban Background
Street	1						
1st Floor	0	1					
3rd Floor	0	0	1				
6th Floor	0.907	0	0	1			
Roof	0	0	0.002	0	1		
Reference 6th Floor	0.05	0	0	0.013	0	1	
Urban Background	0.159	0	0	0.121	0	0	1

O_3 Summer	Street	1st Floor	3rd Floor	6th Floor	Roof	Reference 6th Floor	Urban Background
Street	1						
1st Floor	0.215	1					
3rd Floor	0	0	1				
6th Floor	0	0.004	0.185	1			
Roof	0	0	0.418	0.032	1		
Reference 6th Floor	0.024	0.237	0.001	0.084	0	1	
Urban Background	0.001	0	0	0	0	0	1

O_3 Winter	Street	1st Floor	3rd Floor	6th Floor	Roof	Reference 6th Floor	Urban Background
Street	1						
1st Floor	0	1					
3rd Floor	0	0	1				
6th Floor	0	0.001	0	1			
Roof	0	0	0.781	0	1		
Reference 6th Floor	0.002	0	0	0.005	0	1	
Urban Background	0.217	0	0	0	0	0.033	1

Table 3.S-2 Student's t-tests of NO_2 and O_3 concentrations at the Frankfurter Allee and Kottbusser Damm experimental sites. Extremely small p-values are rounded down to 0.

NO_2 Frankfurter Allee	**Street (N)**	**Street (S)**	**Roof**	**MC 174**	**Urban Background**
Street (N)	1				
Street (S)	0	1			
Roof	0	0	1		
MC 174	0.012	0	0	1	
Urban Background	0	0	0	0	1

NO_2 Kottbusser Damm	**Street (W)**	**Street (E)**	**4th Floor**	**Urban Background**
Street (W)	1			
Street (E)	0	1		
4th Floor	0.014	0	1	
Urban Background	0	0	0	1

O_3 Frankfurter Allee	**Street (N)**	**Street (S)**	**Roof**	**Urban Background**
Street (N)	1			
Street (S)	0	1		
Roof	0	0	1	
Urban Background	0	0	0.771	1

Table 3.S-3 Wilcoxon-Mann-Whitney U-tests of NO_2 concentrations at Frankfurter Allee broken down by wind direction.

NO_2		North					South				
		Street (N)	Street (S)	Roof	MC 174	Urban Background	Street (N)	Street (S)	Roof	MC 174	Urban Background
	Street (N)	1					1				
	Street (S)	0.686	1				0	1			
	Roof	0.009	0	1			0.038	0	1		
	MC 174	0	0	0.008	1		0.009	0	0	1	
	Urban Background	0	0	0	0	1	0	0	0	0	1
Frankfurter Allee		West					East				
		Street (N)	Street (S)	Roof	MC 174	Urban Background	Street (N)	Street (S)	Roof	MC 174	Urban Background
	Street (N)	1					1				
	Street (S)	0	1				0	1			
	Roof	0	0	1			0	0	1		
	MC 174	0	0	0.003	1		0	0	0.524	1	
	Urban Background	0	0	0	0	1	0	0	0	0	1

Table 3.S-4 Wilcoxon-Mann-Whitney U-tests of NO_2 concentrations at Kottbusser Damm broken down by wind direction.

NO_2		North				South			
		Street (W)	Street (E)	4th Floor	Urban Background	Street (W)	Street (E)	4th Floor	Urban Background
	Street (W)	1				1			
	Street (E)	0.075	1			0	1		
	4th Floor	0.245	0.002	1		0.418	0	1	
	Urban Background	0	0	0	1	0	0	0	1
Kottbusser Damm		West				East			
		Street (W)	Street (E)	4th Floor	Urban Background	Street (W)	Street (E)	4th Floor	Urban Background
	Street (W)	1				1			
	Street (E)	0	1			0.438	1		
	4th Floor	0.626	0	1		0.5	0.1	1	
	Urban Background	0	0	0	1	0	0	0	1

Table 3.S-5 Wilcoxon-Mann-Whitney U-tests of O_3 concentrations at Frankfurter Allee broken down by wind direction.

O_3		North				South			
		Street (N)	Street (S)	Roof	Urban Background	Street (N)	Street (S)	Roof	Urban Background
	Street (N)	1				1			
	Street (S)	0.205	1			0	1		
	Roof	0.018	0	1		0.234	0	1	
	Urban Background	0	0	0.005	1	0.149	0	0.807	1
Frankfurter Alle		West				East			
		Street (N)	Street (S)	Roof	Urban Background	Street (N)	Street (S)	Roof	Urban Background
	Street (N)	1				1			
	Street (S)	0	1			0.023	1		
	Roof	0	0	1		0.269	0.001	1	
	Urban Background	0	0	0.566	1	0	0	0.008	1

Table 3.S-6 . Wilcoxon-Mann-Whitney U-tests of NO_2 concentrations at Frankfurter Allee broken down by wind speed.

NO_2		< 2 m.s^{-1}					2-4 m.s^{-1}				
		Street (N)	**Street (S)**	**Roof**	**MC 174**	**Urban Background**	**Street (N)**	**Street (S)**	**Roof**	**MC 174**	**Urban Background**
	Street (N)	1					1				
	Street (S)	0.096	1				0	1			
	Roof	0.046	0	1			0	0	1		
	MC 174	0.414	0.002	0.206	1		0	0	0.04	1	
	Urban Background	0	0	0.024	0	1	0	0	0	0	1

Frankfurter Allee		4-6 m.s^{-1}				
		Street (N)	**Street (S)**	**Roof**	**MC 174**	**Urban Background**
	Street (N)	1				
	Street (S)	0	1			
	Roof	0	0	1		
	MC 174	0.005	0	0.045	1	
	Urban Background	0	0	0	0	1

Table 3.S-7 Wilcoxon-Mann-Whitney U-tests of NO_2 concentrations at Kottbusser Damm broken down by wind speed.

NO_2		< 2 m.s^{-1}				2-4 m.s^{-1}			
		Street (W)	**Street (E)**	**4th Floor**	**Urban Background**	**Street (W)**	**Street (E)**	**4th Floor**	**Urban Background**
Kottbusser Damm	**Street (W)**	1				1			
	Street (E)	0.481	1			0.089	1		
	4th Floor	0.622	0.284	1		0.45	0.007	1	
	Urban Background	0	0.004	0	1	0	0	0	1
		4-6 m.s^{-1}				**> 6 m.s^{-1}**			
		Street (W)	**Street (E)**	**4th Floor**	**Urban Background**	**Street (W)**	**Street (E)**	**4th Floor**	**Urban Background**
	Street (W)	1				1			
	Street (E)	0.003	1			0	1		
	4th Floor	0.185	0.06	1		0.164	0	1	
	Urban Background	0	0	0	1	0	0	0	1

Table 3.S-8 Wilcoxon-Mann-Whitney U-tests of O_3 concentrations at Frankfurter Allee broken down by wind speed.

O_3		< 2 m.s^{-1}				2-4 m.s^{-1}			
		Street (N)	**Street (S)**	**Roof**	**Urban Background**	**Street (N)**	**Street (S)**	**Roof**	**Urban Background**
Frankfurter Alle	**Street (N)**	1				1			
	Street (S)	0.87	1			0	1		
	Roof	0.166	0.177	1		0	0	1	
	Urban Background	0.1	0.073	0.772	1	0	0	0.449	1

		4-6 m.s^{-1}			
		Street (N)	**Street (S)**	**Roof**	**Urban Background**
Frankfurter Alle	**Street (N)**	1			
	Street (S)	0	1		
	Roof	0.017	0	1	
	Urban Background	0	0	0.009	1

3.S.4 Section 4: Sensitivity analysis of T on model performance on Frankfurter Allee

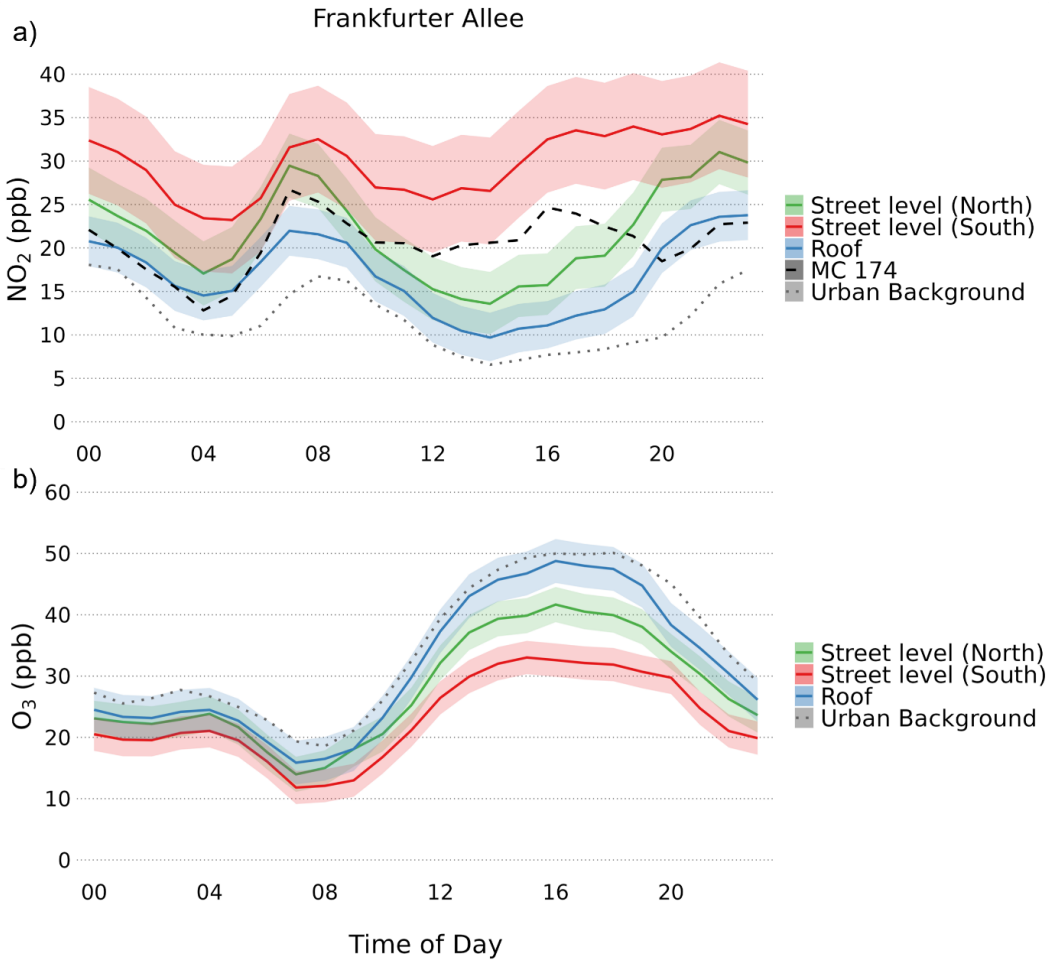

Figure 3.S-11 Diurnal profiles of NO_2 (a) and O_3 (b) concentrations at the Frankfurter Allee experimental site. Dashed lines represent reference NO_2 and O_3 concentrations measured on the 6th Floor of the mathematics building. Dotted lines represent the average urban background concentrations. Shaded areas represent the uncertainty range calculated for each Zephyr using the seven-step methodology. Modelled NO_2 and O_3 concentrations were produced without the input of internal Zephyr temperature to account for potential impacts on sensor performance.

Table 3.S-9 Wilcoxon-Mann-Whitney U-tests of NO_2 and O_3 concentrations at Frankfurter Allee. Extremely small p-values are rounded down to 0. Modelled NO_2 and O_3 concentrations were produced without the input of internal Zephyr temperature to account for potential impacts on sensor performance.

NO_2 Frankfurter Allee	**Street (N)**	**Street (S)**	**Roof**	**MC 174**	**Urban Background**
Street (N)	1				
Street (S)	0	1			
Roof	0	0	1		
MC 174	0.256	0	0	1	
Urban Background	0	0	0	0	1

O_3 Frankfurter Allee	**Street (N)**	**Street (S)**	**Roof**	**Urban Background**
Street (N)	1			
Street (S)	0	1		
Roof	0	0	1	
Urban Background	0	0	0.004	1

Table 3.S-10 Student's t-tests of NO_2 and O_3 concentrations at Frankfurter Allee. Extremely small p-values are rounded down to 0. Modelled NO_2 and O_3 concentrations were produced without the input of internal Zephyr temperature to account for potential impacts on sensor performance.

NO_2 Frankfurter Allee	**Street (N)**	**Street (S)**	**Roof**	**MC 174**	**Urban Background**
Street (N)	1				
Street (S)	0	1			
Roof	0	0	1		
MC 174	0.110	0	0	1	
Urban Background	0	0	0	0	1

O_3 Frankfurter Allee	**Street (N)**	**Street (S)**	**Roof**	**Urban Background**
Street (N)	1			
Street (S)	0	1		
Roof	0	0	1	
Urban Background	0	0	0.003	1

4 Do new bike lanes impact air pollution exposure for cyclists? – A case study from Berlin

Abstract

Cities in the 21st century are dynamically changing in response to environmental and societal pressures, not least among which are climate change and air pollution. In some of these metropoles, such as Berlin, a transformation of mobility systems has already begun. Along a mid-sized street in Berlin, a measurement campaign was conducted in 2020 to accompany the construction of a bike lane and the implementation of a community space along one of the side-streets. Using the new technology of low-cost sensors, higher resolution measurements of local air quality were enabled. Stationary and mobile measurements were taken using EarthSense Zephyr sensor systems before and after the construction of the bike lane and during the timeframe when the community space was in place. It was found that the implementation of the bike lane led to a reduction in NO_2 exposure for cyclists. During periods when the community space was in place, a reduction in NO_2 concentrations was also measured. This study highlights not only the utility of low-cost sensors for the measurement of urban air quality, but also their value in a science-policy context. Measuring local air quality changes in response to traffic interventions will enhance understanding of the associated health benefits, especially in connection with measures promoting more sustainable modes of active travel. More research of this nature is needed to gain a clear understanding of the impacts of traffic interventions on local air quality for better protection of human health.

Keywords: air pollution exposure; low-cost sensors; sustainable mobility; transport policy; science-policy

Published as:

Schmitz, S., Caseiro, A., Kerschbaumer, A., & von Schneidemesser, E. (2021). Do new bike lanes impact air pollution exposure for cyclists? – A case study from Berlin. *Environmental Research Letters*, 16(8). https://doi.org/10.1088/1748-9326/ac1379.

4.1 Introduction

Ambient air pollution was estimated to contribute to around 4.2 million deaths globally in 2015 (Landrigan et al., 2018). Other studies using updated hazard risk ratios (Lelieveld, Klingmuller, et al., 2019) and alternative risk and exposure assumptions (Burnett et al., 2018; Vohra et al., 2021) suggest that this number might be twofold larger. The health effects of ambient air pollution are significant in Europe, with 74% and 99% of its urban population exposed to particulate matter less than 2.5 µm in diameter ($PM_{2.5}$) and ozone (O_3) concentrations, respectively, above WHO recommended limit values in 2018 (EEA, 2020b).

Urban air pollution is a major human health problem with substantial emissions from the transport sector. Consequently, participants in urban transportation (car-drivers, cyclists, pedestrians, etc.) are exposed to high levels of air pollution. The level of exposure is driven by many factors, including, but not limited to, transport emissions, city and transportation infrastructure, time spent commuting, and climate and meteorology. In Europe, car-drivers are exposed to the largest amount of air pollution, followed by cyclists and public transportation users, with pedestrians typically exposed to the least amount (A. de Nazelle, Bode, & Orjuela, 2017; Rank, Folke, & Jespersen, 2001; Raza, Forsberg, Johansson, & Sommar, 2018). A systematic review found that commuters using motorized transport had increased exposure to air pollution due to their proximity to traffic and high air interchange whereas the increased inhalation rates and commuting time of active commuters caused them to have a higher inhaled dose (Cepeda et al., 2017). The calculation of exposure varies across these studies, but an intercomparison of these methods reveals that there is no single best method, many are appropriate, and they should be selected based on the size and objectives of the study (E. Dons et al., 2017)

The direct health impacts of this air pollution exposure have been studied extensively, though most studies use particulate matter and black carbon as proxies for all air pollutants, with less assessing the impact of exposure to high levels of nitrogen dioxide (NO_2). One study found a significant relationship between exposure to NO_2 and heart rate variability in healthy adults (Weichenthal et al., 2011), whereas previous studies identified this relationship only in elderly populations or subjects with pre-existing cardiovascular disease. A cohort study found similar results, with long-term exposure to NO_2 pollution associated with higher risk of heart failure (Sorensen et al., 2017). Moreover, systematic reviews of studies assessing NO_2 exposure and mortality have consistently found evidence of NO_2 exposure associated with a higher risk of all-cause, cardiovascular, and respiratory mortality that might be independent of other common air pollutants (R. W. Atkinson, Butland, Anderson, & Maynard, 2018; Faustini, Rapp, & Forastiere, 2014; S. Huang et al., 2021; Huangfu & Atkinson, 2020). This connection between NO_2 and health effects was also made by participants in a study across seven European cities, in which levels of pollution at their home addresses was significantly linked to their concern over the health effects of air pollution (Evi Dons et al., 2018).

While studies show that the benefits of active travel outweigh the negative health effects of air pollution exposure (Cepeda et al., 2017; Giallouros, Kouis, Papatheodorou, Woodcock, & Tainio, 2020; Tainio et al., 2021), further reductions in exposure by choosing low-traffic routes can additionally reduce associated health effects (Jarjour et al., 2013; Shrestha, Mullins, Zhao, Selvey, & Rumchev, 2020). Research from Montréal and Bogotá revealed that cyclists' exposure to particulate matter and BC was reduced when riding on

separated, protected bike lanes as opposed to in-street facilities (Farrell, Weichenthal, Goldberg, & Hatzopoulou, 2015; Hernández, Ramírez, Benavides, & Franco, 2021). In addition to the reduction in air pollution exposure gained from implementing dedicated cycling infrastructure in cities, the primary increase in health benefits at city-scale comes from the consequent increase in physical activity as more citizens switch to active transport (Schepers et al., 2015).

Personal exposure measurements are limited, and regulatory monitoring stations are sparsely distributed throughout cities. This means that exposure to microenvironments, such as the transport environment, are poorly understood in the context of overall daily exposure. While urban background monitoring locations are often used in population exposure assessment, studies have shown poor agreement between measurements at these monitoring stations and personal exposure in transport environments (Audrey de Nazelle et al., 2012; Gulliver & Briggs, 2004; Ragettli et al., 2013; Xu, Jiang, Zhao, & Stephens, 2017). To overcome these issues, new technologies such as low-cost sensors are being used to increase the spatial resolution of monitoring networks (J. M. Barcelo-Ordinas et al., 2019; Kumar et al., 2015; Mead et al., 2013; Morawska et al., 2018; Popoola et al., 2018), to assess personal exposure (Mahajan & Kumar, 2020; Morawska et al., 2018; Piedrahita et al., 2014), and in mobile monitoring (Gao et al., 2016; Genikomsakis et al., 2018; Lim, Kim, et al., 2019; Lin et al., 2017a; E. von Schneidemesser et al., 2019b). While they are less accurate than reference-grade instruments, their low-cost and relatively small size make them more suitable for these applications. Furthermore, they can serve to increase our understanding of air pollution in urban environments, especially with regards to human health and exposure, providing valuable information not only for scientists, but also for citizens and policymakers.

For various reasons, European cities are starting to shift towards sustainable modes of transport. Some are focused on the win-win of achieving climate goals by reducing emissions of greenhouse gases and reducing the health impacts of air pollution, while others are focused on making these modes safer and more attractive for their citizens. To achieve these goals, many of these cities need to enact policies to encourage greater uptake of cycling (Brand et al., 2021; Nieuwenhuijsen, 2020), among other sustainable transport options. In cities like London, Barcelona, and Berlin, the shift towards more active transport has already begun (Aldred & Goodman, 2020; López, Ortega, & Pardo, 2020; D. von Schneidemesser et al., 2020). Berlin became the first federal state in Germany to enact a Mobility Act in 2018, which was driven largely by a citizen-led bicycle referendum called the 'Volksentscheid Fahrrad,' (D. von Schneidemesser et al., 2020). As a result, new cycling infrastructure, such as protected bike lanes, has been built in the city, with more planned in the coming years. On one street, Kottbusser Damm (KD) in the district of Friedrichshain-Kreuzberg, a protected bike lane was initially planned to be built in the late summer of 2020. With the onset of the COVID-19 pandemic, plans for this were accelerated and it was built as a temporary pop-up bike lane to provide safer infrastructure for citizens seeking to switch away from public transport and towards cycling. A separate measure that was planned and executed according to plan involved the transformation of a portion of a side-street (Böckhstrasse) that is directly linked to KD into a *Spielstrasse* or 'community space', for which the street was closed to through-traffic, allowing for greater outdoor space for public use. In this study, we conducted a measurement campaign using low-cost sensors to assess changes in air quality on Kottbusser Damm and on the Böckhstrasse in connection with the bike lane and the community space, respectively.

4.2 Methods

4.2.1 Small air quality sensors

This data used in this study were collected with small air quality sensors housed in the EarthSense Zephyr sensor system. Included in this sensor system are: 1) Electrochemical (EC) sensors that provide a measure of NO_2 and O_3; 2) micro-optical sensors that count particles to provide a measure of PM; 3) a global positioning system (GPS) unit; 4) internal temperature and relative humidity monitors; 5) an internal fan for air intake and expulsion; 6) a lithium-ion battery; and 7) a Global System for Mobile Communications (GSM) unit for sending logged data to an external database (EarthSense). These sensor systems are part of a new generation of air quality measurement devices that are lower-cost, smaller, and easier to use in comparison to standard reference instrumentation. As a result, they are easily deployed for both stationary and mobile measurements with a potential for high spatial and temporal resolution in various environments.

The EC sensors housed within the Zephyrs react in the presence of atmospheric gas-phase pollutants such as NO_2 and O_3, as the molecules chemically interacting with the measurement nodes of the sensor. To transform this raw voltage signal into concentrations of each pollutant, the sensor system must be co-located and calibrated with reference-grade air quality sensors. Co-location in this context refers to physically installing the sensor systems at a location where they will receive samples from the same parcel of air as the reference instruments.

4.2.2 Co-locations and reference instrumentation

To co-locate the Zephyrs, measurement stations of the Berlin Air Quality Measurement Network (BLUME) were used. When co-locating small sensors, it is crucial that the calibration site experience environmental conditions (i.e., pollution levels, meteorological conditions) as similar to those of the experimental site as possible (Peltier, 2020). In this case, the Zephyrs that were installed on lampposts along KD in Neukölln, Berlin were co-located at a roadside-traffic station, MC117 in Steglitz, Berlin. While it is situated several kilometres away from the experimental site in a different part of the city, the shape of the street canyon and traffic levels are comparable to those of KD. The Zephyr that was installed on the 1st floor of the primary school on the Böckhstrasse was co-located at an urban background station, MC042 in Neukölln, Berlin. This station is only a few blocks away from KD and since the side streets do not experience heavy traffic, it was selected as a more appropriate co-location site than MC117.

4.2.3 Sampling site and sampling strategies

As can be seen in Figure 4-1, four Zephyrs were deployed on lampposts on KD and one was deployed on the 1st floor façade of a primary school on the side-street Böckhstrasse. The latter location was selected as the school lies along the stretch of Böckhstrasse that was converted to a community space. For the mobile measurements, two separate routes were designed to capture various changes in air pollution associated with the new bike-lane; the primary route covered the length of KD, whereas the second route covered side-streets, including along Böckhstrasse. A timeline of the measurement campaign can be seen in Figure 4-2. In total, 9 sets and 11 sets of mobile measurements were conducted before and after the implementation of the bike lane, respectively, and were composed of three continuous loops along each route. These measurements were conducted during the morning, afternoon, and evening to capture the range of intra-day variability in NO_2 concentrations.

4.2.4 Statistical calibration and transformation

To calibrate the Zephyrs, data from the reference instruments was used to train statistical models that included as independent variables the raw sensor signal and meteorological conditions, such as temperature and relative humidity. To this end, the seven-step method was used to clean and flag the raw data; build, train, and optimize the parameters of a multiple linear regression (MLR) model; and predict the final concentrations with an associated measure of uncertainty for each measurement (Schmitz, Towers, et al., 2021a). A Random Forest (RF) model was also built, but this was determined to have no significant increase in accuracy over the MLR model and had more limitations. Further details on the seven-step methodology can be found in (Schmitz, Towers, et al., 2021a). The seven-step methodology has been developed and applied to gas-phase concentrations only at this point. For this reason, as well as the predominance of traffic emissions to total NO_x emissions in urban areas in Europe and Berlin, the analysis presented here is limited to NO_2.

To account for changes in meteorological conditions, traffic patterns, and other external forces such as the COVID-19 lockdowns that may have influenced variations in NO_2 concentrations on KD, all calibrated concentrations were normalized. In this case, hourly averaged concentrations from the six stationary Zephyrs along KD and on the side-streets were normalized to the hourly average of 4 urban background (UB) stations by subtracting the UB hourly averaged concentrations from the Zephyr hourly averaged concentrations. Similarly, the mobile measurements along KD and the side-streets were normalized to the five-minute averages of the four lamppost Zephyrs on KD and the Nansenstrasse urban background station, respectively. This higher resolution was necessary as each set of mobile measurements took roughly ~40 minutes for each route.

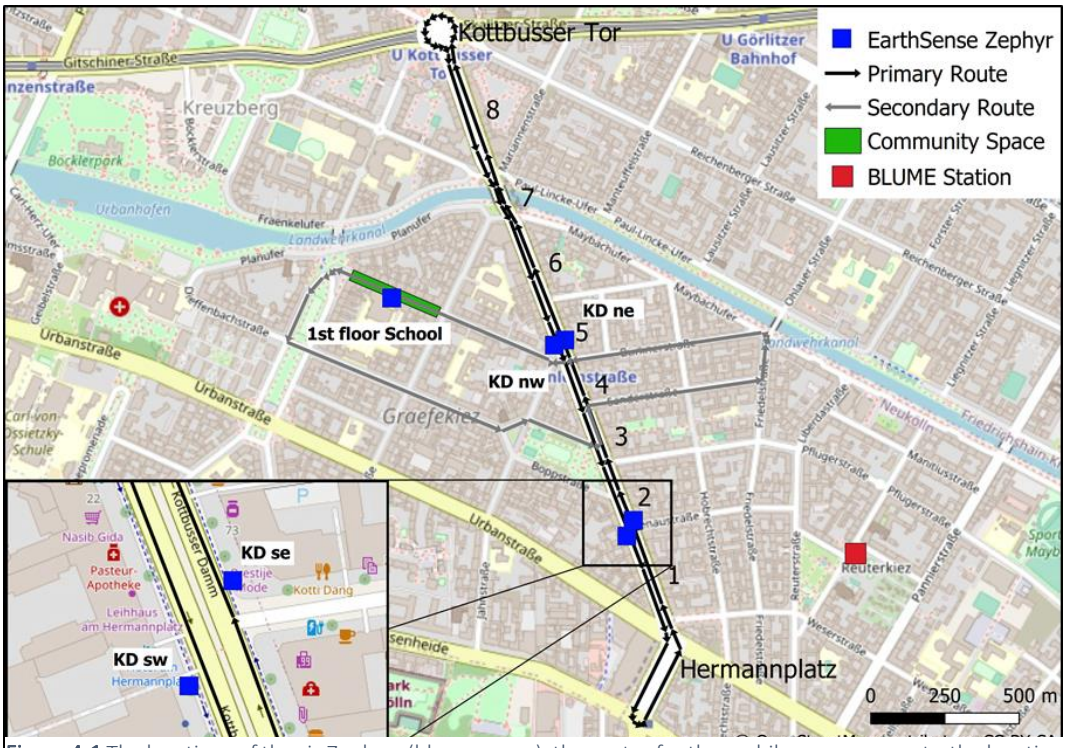

Figure 4-1 The locations of the six Zephyrs (blue squares), the routes for the mobile measurements, the location of the Nansenstrasse BLUME station (red square), and the location of the temporary community space. The new bike lane was implemented along the entire length of KD between Urbanstrasse and Kottbusser Tor. Labeled along the primary route are the individual street segments.

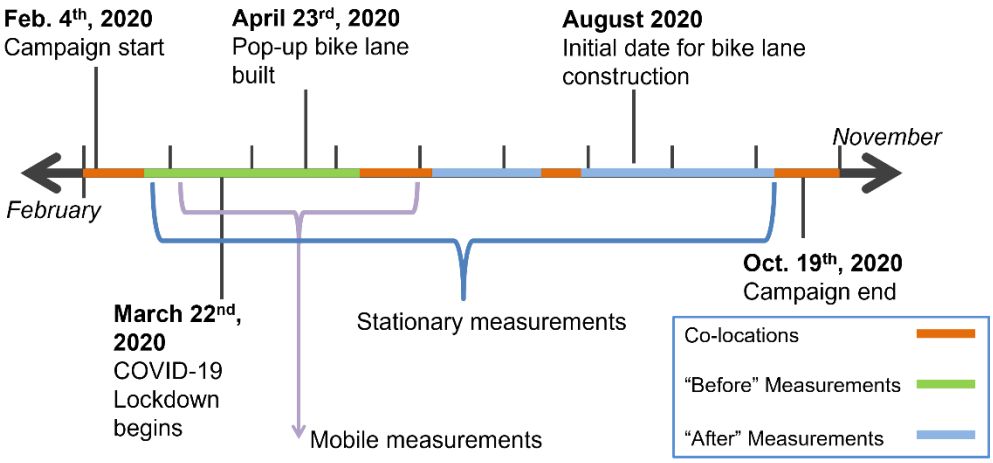

Figure 4-2 Timeline of the measurement campaign depicting the temporal coverage of the mobile and stationary measurements, as well as key events. "Before" refers to the measurements captured prior to bike lane construction and "after" refers to those following its implementation. Each tick represents the beginning of a new month, starting in February, 2020.

4.3 Results

4.3.1 Stationary measurements

As can be seen in Figure 4-3, the stationary measurements along KD exhibit a similar diurnal pattern with good agreement. The side-street Zephyr on the primary school agrees more closely with the average urban background NO_2 across four stations. In all cases, the morning and evening peaks in NO_2 align, with the KD Zephyrs showing higher concentrations and a lower mid-day dip as a result of higher local emissions from the street. Using a Mann-Whitney U-test, the distributions of all stationary Zephyrs were compared (Table 4-1). All KD Zephyrs were found to have significantly different distributions than the primary school Zephyr Among the lamppost Zephyrs, the northern and southern Zephyrs form two distinct groups. Given these results, it was concluded that normalizing the mobile measurements along the primary route to the nearest pair of sensors was appropriate. For measurements along segments between the two pairs, a weighted average was used. However, using the simple average of all four Zephyrs produced similar results.

Analysis was conducted to assess the relative impact of the bike lane on the stationary measurements, which reflect the local air pollution conditions on KD. However, due to the significant impact of the COVID-19 pandemic and subsequent stringent measures in Berlin, potential effects from the bike lane on NO_2 concentrations for the general KD area, as represented by the stationary measurements on KD, could not be isolated. More data would be needed to establish this connection.

Table 4-1 Wilcoxon-Mann-Whitney U-tests of the difference in means between each of the stationary Zephyrs across the entire experiment. Reported in the table are p-values. Italicized and bolded are values below the Bonferroni-corrected p-value of 0.01, which indicate acceptance of the alternative hypothesis, that the distributions of the 2 sensors are NOT equal. Extremely low p-values are represented here as 0.

	KD SW	**KD SE**	**KD NW**	**KD NE**	**Lemgo 1st floor**
KD SW					
KD SE	0.48				
KD NW	***0***	***0***			
KD NE	***0.0058***	***0.0003***	0.053		
Lemgo 1st floor	***0***	***0***	***0***	***0***	

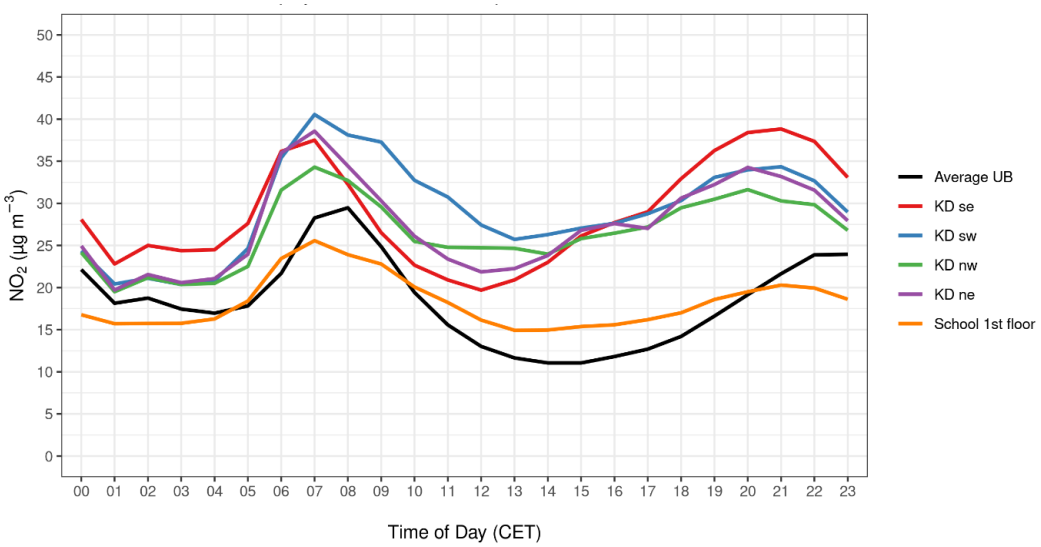

Figure 4-3 NO_2 diurnal pattern of all stationary Zephyrs as well as the urban background (UB) average for the duration of the campaign. The KD Zephyr locations correspond to cardinal directions related to their position on the street, e.g., se = southeast.

4.3.2 Mobile measurements

To assess the impact of the bike lane on cyclists' personal exposure, NO_2 concentrations from the primary and secondary route from before and after its construction were compared. Figure 4-4 shows these data in box-and-whisker plots, grouped according to the presence of the bike lane. As each measurement contained an associated uncertainty, error was propagated when comparing the before and after samples[3]. For the primary route, a decrease in the median normalized NO_2 of 8.4 ± 7.4 µg.m^{-3} was measured whereas for the secondary route the decrease in the median normalized NO_2 was 2.5 ± 7.4 µg.m^{-3}. In both cases, tests with the Mann-Whitney U-test were significant at a p-value of 0.05, indicating the distributions of the data from before and after the implementation of the bike lane are different. At the 95th percentile, the reduction in normalized NO_2 for the primary route was 14 ± 7.4 µg.m^{-3}, whereas for the secondary route an increase of 2.4 ± 7.4 µg.m^{-3} was found.

To more closely inspect these results, the data from each route were broken down into smaller segments. For the primary route, the trend of decreasing concentrations was found in every segment, and all were statistically significant. Two of the segments along the route, at Hermannplatz and Kottbusser Tor, showed the same trend in decreasing concentrations even though no changes to cycling infrastructure were implemented at these sites. The same analysis for the secondary route was inconclusive, as only three of the twelve segments were statistically significant, with some segments showing no decrease or even a slight increase in NO_2 concentrations.

[3] Propagation of error for differences between medians and the 95th percentile were calculated using the formula $Q = \sqrt{(e_1)^2 + (e_2)^2}$, where Q is the propagated error and e_1 and e_2 are the individual uncertainties of the two measurements being compared.

4.3.3 Community space measurements

Figure 4-5 shows box-and-whisker plots for the Zephyr located at the primary school, grouped according to whether the community space was in place or not. On Wednesdays between 14:00 and 18:00, when the community space was in place (April – September), the median normalized NO_2 was 3.7 ± 11.2 µg.m^{-3} less than when there was no community space in place (February, March, and October). A Mann-Whitney U-test indicated that these distributions are statistically different (p = 0.012). To confirm that the difference seen in Figure 4-5a (the comparison of Wednesdays with and without the community space) was indeed owing to the community space, a parallel comparison for all other weekdays was also carried out (Figure 4-5b). This difference in median normalized NO_2 from 14:00 – 18:00 was 0.42 ± 11.2 µg.m^{-3} under the same conditions and was not statistically significant (p = 0.61). With this analysis, it was assured that the changes measured on Wednesdays could be associated with the community space, as all other weekdays during the same time window did not exhibit any significant changes. A comparison to weekend concentrations was not made, as weekend NO_2 concentrations were significantly different than on weekdays and were not related to the street closures. Finally, an analysis to control for holiday days that fell on weekdays was done and showed that the effect was negligible.

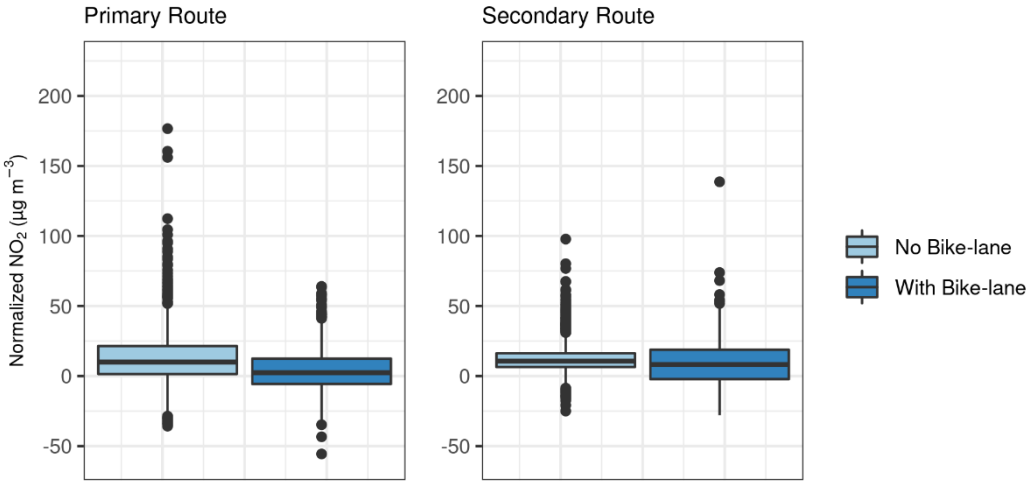

Figure 4-4 (left) Box-and-whisker plots of mobile measurements along KD normalized to the five-minute average NO_2 of the four lamppost Zephyrs, with and without the presence of a bike lane, not including measurements from Kottbusser Tor or Hermannplatz; (right) Box-and-whisker plots of mobile measurements along the side-streets of KD normalized to five-minute average NO_2 of the Nansenstrasse urban background station, with and without the presence of a bike lane.

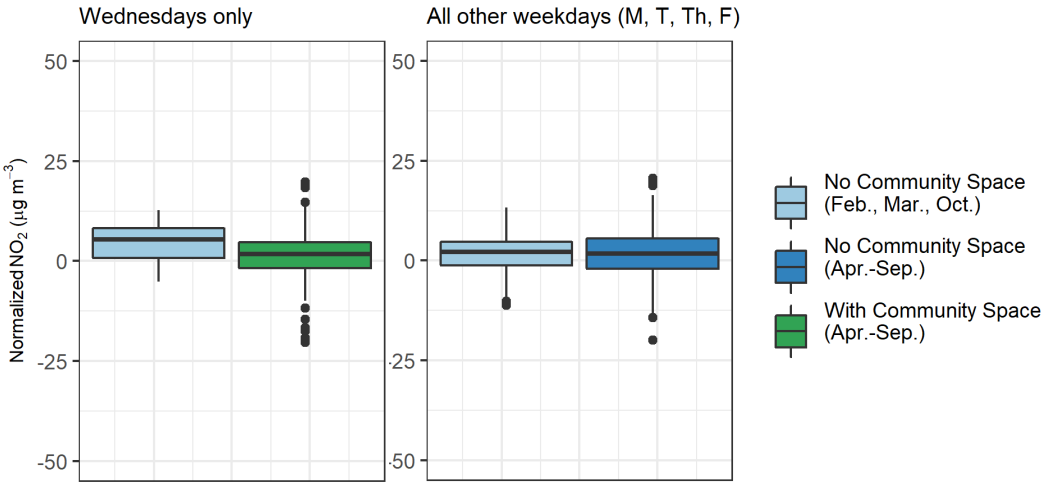

Figure 4-5 Box-and-whisker plots of measurements from the Zephyr located at the primary school, normalized to the hourly average of four urban background stations, grouped by time periods where the community space was or was not in place on a) Wednesdays only and b) All weekdays excluding Wednesdays.

4.4 Discussion

This study has successfully shown that low-cost sensors can be deployed to assess changes in air quality in connection with local transport measures. In general, the Zephyrs accurately captured the expected diurnal trends of NO_2 on KD as well as on the Böckhstrasse, demonstrating their utility for such work. Despite the higher level of uncertainty in the measurements, these low-cost sensors have proven their value in assessing small-scale spatial changes in air quality in cities, which has implications for future research aimed at understanding local changes in air quality, including before-after measurements in connection with transport measures. In the context of the mobility transition in Germany, these sensors can provide valuable information to policymakers with regards to the effect on air quality for the various measures they may implement.

Here, a reduction in exposure to NO_2 for cyclists following the implementation of the bike lane, after normalization to local conditions, was measured. Furthermore, the reduction in exposure along KD was higher at the 95th percentile, which suggests that cyclists were exposed to fewer extreme peaks in concentrations associated with proximity to tailpipe emissions from motor vehicles. These reductions were not seen along the side-streets, indicating that the effect is local in nature and can be associated with the construction of the bike lane. The exposure to peak concentrations can be quite important; Evi Dons et al. (2019) found that only 5.5% of participants' daily exposure was attributed to peak concentrations, but these contributed to 21% of their total exposure (Evi Dons et al., 2019). As such, cyclists along KD appear to have profited from this measure, not only from the increased safety of a protected bike lane, but also from the reduction in exposure to both overall and peak NO_2 concentrations. An effect of the bike lane on overall air quality on KD could not be determined due to a lack of data representative of local conditions following the implementation of the bike lane that are free of effects from restrictive COVID-19 health measures. While no other studies exist in the literature that are directly comparable, related studies have shown that cycling infrastructure influences the particle number concentration cyclists are exposed to (Boogaard, Borgman, Kamminga, & Hoek, 2009; E. von Schneidemesser et al., 2019a), whereas others quantify the role of urban infrastructure on air pollution, such as green walls or low emission zones (LEZ) (Boogaard et al., 2012; Paull, Krix, Torpy, & Irga, 2020), but these do not assess changes to cyclists'

exposure to air pollution. Other studies that explicitly measured cyclists' exposure to air pollution did not explicitly connect their mobile measurements to specific transport measures (Samad & Vogt, 2021). Therefore, this study is one of the first to measure changes in cyclists' exposure to NO_2 in direct connection with changes in cycling infrastructure.

While quantifying the health benefits associated with this reduction is beyond the scope of this study, a reduction in NO_2 exposure is a positive influence on cyclists' health, given the established connections between NO_2 exposure and mortality (R. W. Atkinson et al., 2018; Faustini et al., 2014; S. Huang et al., 2021; Huangfu & Atkinson, 2020; Sorensen et al., 2017). In addition, recent research has shown that pop-up bike lanes increased cycling across European cities (Kraus & Koch, 2021). That study estimated that the associated health benefits in terms of kilometres of new bike lanes per square kilometre were estimated to be worth between \$1 and \$7 billion annually, or \$1.2 – \$3.5 in terms of kilometres per capita, if the uptake in cycling is maintained (Kraus & Koch, 2021). This is in line with other research estimating the health benefits across 167 European cities, in which the expansion of cycling networks could lead to the avoidance of 10,000 premature deaths annually (Mueller et al., 2018). These findings indicate that a measured increase in cycling uptake along KD confers additional health benefits to Berlin cyclists alongside the reduction in NO_2 exposure associated with new bike lane infrastructure.

The implementation of the community space also led to a measured reduction in NO_2 concentrations. On other weekdays where the community space was not in place, there was no discernible difference in NO_2 concentrations. While this reduction was smaller than that on KD, it highlights the relationship between NO_2 and vehicle traffic. Böckhstrasse is already traffic-calmed, but if cars are no longer allowed to traverse the street, there will be further local reductions in NO_2 concentrations. The overall decrease, however, is constrained by urban background levels of NO_2 pollution, which would require larger-scale changes in emissions sources to change.

4.4.1 Limitations

This study and its findings are subject to several key limitations. Primarily, the COVID-19 pandemic disrupted the plans for this measurement campaign. Due to restrictive lockdowns implemented in Berlin in response to the pandemic, traffic patterns and behaviours were substantially altered during the campaign. Furthermore, the bike lane was implemented as temporary cycling infrastructure months earlier than anticipated. These circumstances led to a substantial reduction in the amount of data collected before and after the bike lane was in place. In addition, this study would have benefited from additional measurements that quantitatively assessed traffic patterns, composition, and behaviour along KD and its side-streets. This data would allow for a more detailed analysis of the impact of the bike lane on local traffic, individual transport decisions, and the concomitant influences on air quality. For this study, such data were unavailable. In addition, no tests on sensor performance pertaining to mobile deployment were conducted and therefore the potential influences of mobile use on the sensors (i.e., vibration) in this study are unknown. However, as the analysis focuses on gas-phase species (not particulate matter) isokinetic sampling is a non-issue, and previous mobile deployments of measurement devices have not shown vibration to cause any interference. As there are few studies assessing low-cost sensor performance in mobile conditions, more research is needed to identify potential interferences on measurement quality. Last, it should be noted that the deployment of the sensors in this study did not

follow regulatory guidelines for site selection, nor did the sensor go through any certification process for assessing their performance relative to regulatory standards. As such, the data should not be used to assess exceedances of air quality limit values. However, the results of this study still have a high relevance for human health in urban areas, and data are presented with associated uncertainties.

4.5 Conclusions

This study has demonstrated the utility of small sensors for both stationary and mobile measurements in an urban environment to measure small scale spatial changes in air quality. This is one of the first studies to implement such small sensors to accompany the implementation of a mobility policy to quantify the effect on air pollution, including exposure. Results showed that the implementation of a (pop-up) bike lane, in which cyclists went from cycling in the street with traffic, to a dedicated bike lane largely protected from motor vehicle traffic by a lane of parked cars, resulted in a reduction of 8.4 ± 7.4 $\mu g.m^{-3}$ in NO_2 or 22% ± 19% that they were exposed to. This underlines the importance of infrastructure for the protection of human health in urban areas. Additional studies are needed to understand how representative and transferrable these results are. As the mobility transition in Berlin and across Europe proceeds, these types of measurements will prove invaluable for decisionmakers.

Acknowledgements

The authors would like to thank Katja Grunow, Philipp Tödter, and Marcel Kyrsiak (Berlin Senate Department for the Environment, Transport, and Climate Protection), Felix Weisbrich (Roads and Green Spaces Department, District Friedrichshain-Kreuzberg), Ms. Albert and Mr. Weinandt (Lemgo primary school), Tarik Mustafa and Martin Wittau (Bundesvereinigung Nachhaltigkeit e.V.), and the entire EarthSense team for their support of this work. We would also like to thank Sophia Becker, Dirk von Schneidemesser, and Katharina Götting (RIFS) for their collaboration on this work and as part of the LuftMODE interdisciplinary research group.

Funding

The research of EvS, SS, AC is supported by RIFS, with financial support provided by the Federal Ministry of Education and Research of Germany (BMBF) and the Ministry for Science, Research and Culture of the State of Brandenburg (MWFK).

5 Low-cost system application for policy assessment: A case study from Berlin

Abstract

Local policies are part of the toolbox available to decision makers to improve air quality but their effectiveness is under-evaluated and underreported. We evaluate the impact of the closure of a street in the city centre of Berlin on the local air pollution. Nitrogen dioxide (NO_2) was measured on the street where the policy was implemented and on two parallel streets using low-cost sensor systems supported by periodic calibrations against reference-grade instruments and constrained by passive samplers. Further measurements of NO_2 were conducted with a reference-grade instrument mounted on a mobile platform. The concentrations were evaluated against the urban background to isolate the policy-related signal from natural fluctuations, long-term trends and the COVID-19 lockdown. The intervention reduced NO_2 concentrations to the level of the urban background on week- days. Kerbside NO_2 concentrations exhibited substantial differences to the concentrations measured at lampposts highlighting the difficulty for such measurements to capture personal exposure. The results have implications for policy, showing that an intervention on the local traffic patterns can be very effective in improving local air quality. The versatile and fast deployment capabilities of low-cost systems offer a quick information return regarding policy. Sampling locations must be chosen carefully in order to assess human exposure.

Submitted to Atmospheric Environment: X on 26.08.2022 as:

Caseiro, Alexandre, Seán Schmitz, Andreas Kerschbaumer, and Erika von Schneidemesser. 'Low-cost system application for policy assessment: a case study from Berlin'. Atmospheric Environment: X.

5.1 Introduction

Cities are hot spots of atmospheric emissions and urban populations. Worldwide, cities have struggled with air quality issues for decades (EEA, 2007; Lelieveld, Evans, Fnais, Giannadaki, & Pozzer, 2015; Liu et al., 2019). A vast majority (over 70 %) of the population of the European Union (EU) lives in cities. Despite improvements, 77 % of the EU-28 urban population was still exposed to PM2.5 concentrations above the World Health Organization (WHO) Air Quality Guidelines value in 2019 (EEA, 2019, 2020a, 2020c). Even low concentrations of air pollutants pose a threat to human health (Dominici et al., 2019; Stanaway et al., 2018), and the WHO has recently recognized this research finding by updating its guidelines based on a systematic review of the body of literature (WHO, 2021). The presence of pollutants in the air impacts the population by means of increased incidence of diseases and years of life lost, mainly via cardiovascular diseases (Burnett et al., 2018; HEI, 2020). For instance, in 2018, 54000 premature deaths in the EU-28 (of which 9200 in Germany) were attributed to NO_2 alone (EEA, 2020a). In the context of the ongoing SARS-CoV-2 pandemic, previous exposure to air pollution was identified as a relevant factor increasing mortality risk from the COVID-19 disease (Pozzer et al., 2020; Erika von Schneidemesser et al., 2021).

In 2013, relevant and ambitious goals were set in Europe to tackle the issue: (1) the long-term goal to reach the WHO guidelines, even upon update (European Commission, 2013; WHO, 2021), (2) full compliance with existing legislation by 2020, and (3) the halving (in 2030 relative to the 2005 levels) of the premature deaths in the EU due to air pollution (European Commission, 2013). Although achieving the objective for 2020 was deemed unlikely in all member states already in 2018 (EEA, 2018), the adoption of novel instruments (e.g. the National Emission reduction Commitments Directive of 2016) allows some degree of optimism regarding the attainment of the long-term objectives, with significant disparities between member states, depending on if the member-states comply with the obligations laid out in their National Air Pollution Control Programmes (EC, 2019, 2021; Sicard, Agathokleous, De Marco, Paoletti, & Calatayud, 2021).

Fossil fuel burning has been identified as a dominant source of air pollution-related health impacts (Lelieveld, Klingmüller, et al., 2019). The role of road traffic in the concentration increase of air pollutants in urban areas with respect to background levels has long been acknowledged (Lenschow et al., 2001; Oliveira et al., 2010). For NO_2 in particular, exceedances of the standards in European cities are attributed mainly to the high levels of road traffic and domestic combustion (Dias et al., 2018). This has spurred the emergence of a concept known as the mobility transition. Indeed, over half of the policies implemented by member states to reduce air pollution were traffic-related, targeting urban mobility either by addressing technological issues to lower end-of-pipe emissions or reducing traffic (urban road tolls, low emission zones (LEZ), and access regulation schemes) (EEA, 2018).

Traffic reduction policies, technological measures and behavioural change are factors which have been recognized as having the potential to mitigate exposure to air pollution in an ever more urbanized anthropocentric world (Fuglestvedt, Berntsen, Myhre, Rypdal, & Skeie, 2008; Gallardo et al., 2018; Kelly & Zhu, 2016; Shindell et al., 2011; Yan et al., 2014). The EU has evolved towards more stringent emission standards (Magueta, Madaleno, Ferreira Dias, & Meireles, 2018). Despite the technological advances prompted by the evolution of legislation, which have considerably decreased the fleet's

emission factors, the increased use of personal motorized transportation has offset those gains (Kelly & Zhu, 2016). European cities therefore find themselves in a situation where the reduction in per vehicle emissions has not been sufficient to alleviate the detrimental effects of traffic emissions on human health.

The policy treated in the present work goes along with strategies such as charging schemes, LEZs, vehicle speed management or the prohibition of circulation for highly polluting vehicles, which have been adopted by European cities in recent years (Rodriguez-Rey et al., 2022). Such strategies are translated into policies and put in place in order to reduce the externalities related to traffic, such as congestion, noise, accident risk and air pollution (Bernardo, Fageda, & Flores-Fillol, 2021a, 2021b; De Borger & Proost, 2013; Morton, Mattioli, & Anable, 2021; Sfendonis, Basbas, Mintsis, Taxiltaris, & Politis, 2017; Tretvik, Nordtømme, Bjerkan, & Kummeneje, 2014). While charging schemes aim to reduce the flow of traffic within a city, LEZs are defined areas where access by more polluting vehicles is restricted or banned, in an attempt to control the technology mix of the emitters. There is indeed some evidence that freight transport and fleet composition are affected by the implementation of LEZs, but the change, relative to e.g., the national average, appears to be short-lived (André, Carteret, Pasquier, & Liu, 2017; Ellison, Greaves, & Hensher, 2013; Peters, Burguillo, & Arranz, 2021; Settey, Gnap, & Beňová, 2019; Ye, Qin, & Chen, 2021). LEZs have been implemented in over 250 European cities in the last decade and a half, with varying approaches and rules (Aguayo, Reichmuth, & Weintraub, 2021; Cruz & Montenon, 2016). In Berlin (the capital city of Germany, with 3.7 million people), some of the policies implemented to improve air quality over the past decade include a ban on older, high NO_2-emitting vehicles on eight highly polluted streets, the establishment of a LEZ and the fostering of clean transport modes ((e-)bikes and electric buses), among others, and can be understood as a part of the mobility transition, reinforced by the Berlin mobility Act of 2018. Although the situation for particulate matter has been improving over the last years (EEA, 2018), NO_2 exposure is still a serious problem (EEA, 2021b).

LEZs and similar policies are under-evaluated in the scientific literature. The study of the impact of LEZs may be conducted via modelling exercises, either ex-ante in the planning phase or after their implementation (Börjesson, Bastian, & Eliasson, 2021; Carslaw & Beevers, 2002; B. Degraeuwe et al., 2021; Dias, Tchepel, & Antunes, 2016; Host et al., 2020; Keuken, Jonkers, Zandveld, Voogt, & Elshout van den, 2012; Poulhès & Proulhac, 2021; Sánchez, Ortega, López-Lambas, & Martín, 2021). The number of studies which evaluate the impact of LEZs on air quality by means of measurements is limited and the conclusions are sometimes contradictory, e.g., Boogaard et al. (2012) and Panteliadis et al. (2014) (the Netherlands), H. E. Wood et al. (2015) and Mudway et al. (2019) (London), or Ferreira et al. (2015) and Santos, Gómez-Losada, and Pires (2019) (Lisbon). Indeed, the effect on air pollutant concentrations, and public health indicators, appears to be of diminished importance relative to pre-LEZ levels (Ellison et al., 2013; Ferreira et al., 2015; Jianwei Gu et al., 2022; Jones, Harrison, Barratt, & Fuller, 2012; H. E. Wood et al., 2015). Therefore, results of such studies are largely affected by which sources of bias are taken into consideration and how they are controlled for (Gehrsitz, 2017; Holman, Harrison, & Querol, 2015; Malina & Scheffler, 2015a, 2015b) (Morfeld, Groneberg, & Spallek, 2015a, 2015b). The heterogeneity of the nature of LEZs (e.g., which vehicles are affected or how large the LEZ is) further hinders drawing a general conclusion on their effectiveness (Josef Cyrys, Wichmann, Rückerl, & Peters, 2018; Holman et al., 2015; Lurkin, Hambuckers, & van Woensel, 2021). For example, results from Germany, where light-duty vehicles (LDVs)

have been affected by LEZs along with heavy-duty vehicles (HDVs), show a clearer effect than in places where only HDVs, and not LDVs, are the subject of policies (Holman et al., 2015; Jiang, Boltze, Groer, & Scheuvens, 2017; Pestel & Wozny, 2021). In the particular case of air pollution, most studies have evaluated the pre- and post-policy levels by means of data from monitoring stations (Ferreira et al., 2015; Gehrsitz, 2017). Although suitable for large inner-city areas, monitoring data do not have the necessary spatial resolution which allows catching the fine nature of the impact of a local policy. Furthermore, confounders such as meteorology and pre-policy technological trends, influence the outcome of the before/after analysis and are only seldom accounted for in measurements-based assessments (Mudway et al., 2019; Salas, Perez-Villadoniga, Prieto-Rodriguez, & Russo, 2021; Santos et al., 2019; Tartakovsky, Kordova – Biezuner, Berlin, & Broday, 2020).

In the present study, we attempt to overcome the limitations listed above by using localized, calibrated, NO_2 measurements (from reference instrumentation on a mobile platform and static, calibrated, low-cost systems (LCS) and passive samplers), normalized to the urban background, to assess the impact of a street closure to motor vehicle traffic on air quality. Additionally, we investigate the relevance of sampling location at the micro-environment scale for exposure assessment. LCS are a popular new technology in the field of atmospheric chemistry, especially within the last 10-15 years. Much research has been done investigating their accuracy (Cross et al., 2017; Karagulian et al., 2019; Malings et al., 2019; Rai et al., 2017; Zimmerman et al., 2018), potential applications (Bigi et al., 2018; Castell et al., 2017; McKercher et al., 2017; Morawska et al., 2018), and technological advancement (Fishbain et al., 2017; Peterson et al., 2017; Spandonidis et al., 2020), but few studies deploy them in a policy evaluation context (Schmitz, Caseiro, et al., 2021). Despite current accuracy and precision limitations, LCS are highly versatile are highly versatile and offer the possibility a quick information return upon a flexible deployment.

The policy was implemented by the city of Berlin as a trial in the context of the mobility transition. The idea behind the policy was primarily to increase the attractiveness of the street, the Friedrichstrasse, thereby increasing the volume of visitors to the heavily commercial street. In addition, a new bike-lane would be implemented for cyclists traversing the street. This trial street closure was put in place in August 2020 and was planned through the end of January 2021. It was later extended until the end of October 2021. This study presents the results of that measurement campaign and discusses their implications. We show that LCS have the ability to quickly inform policymakers about the change in ambient air quality produced by a policy. We further show that the sampling location must be chosen carefully if the measurements are to be used to assess human exposure to urban air pollution.

5.2 Methods

5.2.1 Instrumentation

In this study, a variety of instruments were used to gather data on air pollution at the study site and generally in Berlin. In the present study, we focus on the measurement of NO_2, owing to the substantial contribution of vehicle emissions to NO_2 concentrations in Berlin and urban areas generally.

5.2.1.1 EarthSense Zephyrs

To gather high-resolution data at multiple sites in the study area, three Zephyrs manufactured by EarthSense Systems Ltd. were installed on lampposts at about 3 m height. These sensor systems contain a variety of components including: 1) electrochemical (EC) sensors that measure NO_2 and O_3; 2) micro-optical sensors for the measurement of PM; 3) a global positioning system (GPS) unit; 4) internal temperature and relative humidity monitors; 5) an internal fan for air intake and expulsion; 6) a lithium-ion battery; and 7) a Global System for Mobile Communications (GSM) unit for sending logged data to an external database (EarthSense, 2021). These sensor systems are part of a new generation of air quality measurement devices that are lower-cost, smaller, and easier to use in comparison to standard reference instrumentation. These were installed on Glinkastrasse, Friedrichstrasse, and Charlottenstrasse on June 13th, 2020, and provided continuous measurements throughout the entire campaign (until January 31st, 2021).

As the low-cost sensors within the Zephyrs do not directly measure concentrations of pollutants such as NO_2, they must be calibrated with reference-grade instrumentation. By doing so, statistical models can be trained using reference data and the raw data from the Zephyrs as response and predictor variables, respectively. To achieve this, the Zephyrs were co-located at a BLUME roadside monitoring station (MC117) in Steglitz, Berlin. Over the course of the campaign, four co-locations were performed at the same site. These were spaced accordingly with seasonal variation, so as to cover as wide a range of meteorological conditions as possible. Statistical models, in this case using multiple linear regression (MLR), were then built using the seven-step method (Schmitz, Towers, et al., 2021b) and were used to predict NO_2 concentrations throughout the measurement campaign.

5.2.1.2 Sensitivity analysis

In the process of obtaining NO_2 concentrations, an uncertainty range in the form of upper and lower bounds are also produced. In the present study, the ratio of the average upper (lower) bounds to the individual 5-minutes data are 1.31, 1.34 and 1.27 (0.69, 0.65 and 0.73) for Friedrichstrasse, Glinkastrasse and Charlottenstrasse. These bounds were calculated in the final step of the seven-step method, in which reference instrument error was combined with the prediction uncertainty of the statistical MLR models calculated during validation with co-location data. As low- cost sensors' field performance are affected by many factors, this calculation produces representatively large ranges of uncertainty. However, this uncertainty is associated with the magnitude of changes in concentrations, not with the direction of the changes, as the Zephyrs capture the diurnal patterns of NO_2 concentrations. A first sensitivity analysis was conducted using the upper and lower bounds of the individual 5-minutes data as input for the analysis and comparing its outcome to the that of the main analysis.

The seven-step methodology produces NO_2 concentrations based on co-location data. Co-locations within the sampling campaign (co-locations 2, 3 and 4, roughly between June 13th, 2020 and January 31st, 2021, see Table 4-1) were used for the main analysis. Data from three other co-locations (co- locations 1, 5 and 6, see Table 4-1) were added to the calibration process and the corresponding NO_2 concentrations used as input for the analysis for sensitivity analysis purposes.

To compare results using different methods, the seven-step methodology was used to produce NO_2 concentrations with MLR and the random forest machine-learning

technique (RF). These produced consistent results, but MLR was selected as it more accurately captured extreme high and low concentrations, which the RF model did not. The analysis was repeated with the RF-based dataset as a third sensitivity analysis to check the robustness of the calibration.

5.2.1.3 *Passive samplers*

As a part of the test phase of the car-free Friedrichstrasse policy, the Berlin Senate Department for the Environment, Urban Mobility, Consumer Protection and Climate Action (SenUMVK) expanded their network of passive samplers to include more locations in the study area. Of these, three were located on lampposts at about 3 m height one block north of the EarthSense Zephyrs on each street and were therefore selected for comparison with the Zephyrs (Figure 4-1b).

Passive samplers capture ambient NO_2 as the gas molecules adsorb to a reactive chemical on the surface of the tube. This mixture is then extracted and chemically analysed to measure the total amount of NO_2 collected. All samplers were deployed for periods of two weeks, whereafter they were collected for analysis and replaced with new ones. In this case, an external lab provided the passive samplers and conducted the laboratory analysis. For samplers MP 701 and MP 702, these measurements began on the 16th of June 2020 and continued through the end of the campaign. Sampler MP 562 was already in place as part of the measurement network run by the SenUMVK and therefore has continuous data from well before the start and through the end of the campaign.

5.2.1.4 *Cargo bike*

In addition to the measurements with passive samplers and EarthSense Zephyrs, measurements with reference instruments were conducted on Charlottenstrasse. These took place on the same city block as the EarthSense Zephyr installed there. To achieve this, three reference-grade instruments were installed in a specially designed e-cargo bike (LuftRad) which was parked on the street, allowing for the instrument inlets to face the street at about 1–1.5 m height. A GRIMM 11-R mobile particulate matter (PM) monitor, a 2B-Technologies Ozone Monitor, and a Teledyne T-200 NOx monitor were used to this end. The devices were powered by a rechargeable Lithium-ion GreenPack® battery from Ansmann AG. Each set of measurements lasted as long as the battery-life, which totalled roughly seven hours. In total, five days of measurements using this cargo bike were conducted on Charlottenstrasse.

5.2.1.5 *BLUME*

The five BLUME stations used in this analysis employ HORIBA APNA-370 NO_X Monitors to measure NO_2 concentrations in ambient air in accordance with DIN EN 14211 (DIN, 2012). For quality assurance, these NO_X Monitors are subject to an automatic daily function check with NO_2 permeation test gas generators, a two-monthly calibration with NO and NO_2 test gas cylinders and an annual maintenance and adjustment in the BLUME test and calibration laboratory (TÜV, 2006). All test gases used are traceable to national standards of the German Environment Agency (Umweltbundesamt, UBA).

5.2.2 Meteorology

5.2.2.1 *Meteorological stations*

In the present study we have used the pressure and wind speed data measured at three stations within, or in the direct vicinity of, the city of Berlin (Berlin Brandenburg, Berlin-Tegel and Berlin-Tempelhof, see Figure 4-1) and operated by the German Weather Service

(Deutscher Wetter Dienst, DWD). The data were downloaded from the DWD Open Data Hub (https://opendata.dwd.de/), accessed on August 10, 2021. The vertical temperature profiles were retrieved from the DWD Open Data Hub for the Lindenberg station (52.2◦ N, 14.1◦ E).

5.2.2.2 *Reanalysis*

The boundary layer height (BLH) data was retrieved from the Copernicus Atmosphere Monitoring Service (CAMS) global greenhouse gas reanalysis (EGG4). The data were downloaded from the Copernicus Atmosphere Monitoring Service Atmosphere Data Store (ADS) (https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-ghg-reanalysis-egg4), accessed on August 10, 2021.

5.2.3 Traffic counts

Traffic counters were installed at the intersections between Friedrichstrasse, Glinkastrasse and Charlottenstrasse and the perpendicular streets Französische Strasse, Jägerstrasse, Taubenstrasse, Mohrenstrasse, Kronenstrasse and Leipziger Strasse (see Figure 4-1). This resulted in counts for five segments (a subsection of the street between two intersections) for Friedrichstrasse and the parallel streets Glinkastrasse and Charlottenstrasse, from North to South: segment 1 (between Französische Strasse and Jägerstrasse), segment 2 (between Jägerstrasse and Taubenstrasse), segment 3 (between Taubenstrasse and Mohrenstrasse), segment 4 (between Mohrenstrasse and Kronenstrasse), segment 5 (between Kronenstrasse and Leipziger Strasse). The traffic counts were conducted on seven days throughout the experiment: 2 days before the closure of Friedrichstrasse (2020-07-14 and 2020-08-13, Tuesday and Thursday), 4 days after the closure and before the lockdown (2020-09-10, 2020-10-06, 2020-11-05 and 2020-12-15, Thursday, Tuesday, Wednesday, and Tuesday) and one day after the closure and during the lockdown (2021-01-21, Thursday).

5.2.4 Measurement site

For this study, several sources of in-situ data from the measurement campaign, as well as from the Berlin air quality measurement network (BLUME) were used. The measurement site was located in the Mitte city district of Berlin and covers a three by five city-block area (Figure 4-1). The car free Friedrichstrasse policy led to the closure of the street to car-traffic between Französische Strasse and Leipziger Strasse (5 blocks). Within this study area, passive samplers and EarthSense Zephyrs were installed on Glinkastrasse, Friedrichstrasse, and Charlottenstrasse. These locations were selected to assess the effect of the policy on the air quality on Friedrichstrasse itself, as well as on parallel streets associated with potential changes in traffic patterns. Also incorporated in this study were data collected from measurement stations that are a part of BLUME. Figure 4-1 (upper panel) shows the location of four urban background stations (MC 010, MC 018, MC 042, MC 171) and one roadside station (MC 117) that were selected for use in comparison to the study area. These urban background stations surround the measurement site and together were considered representative of urban background conditions for Berlin. Relevant dates for the experiment are given in Table 4-1.

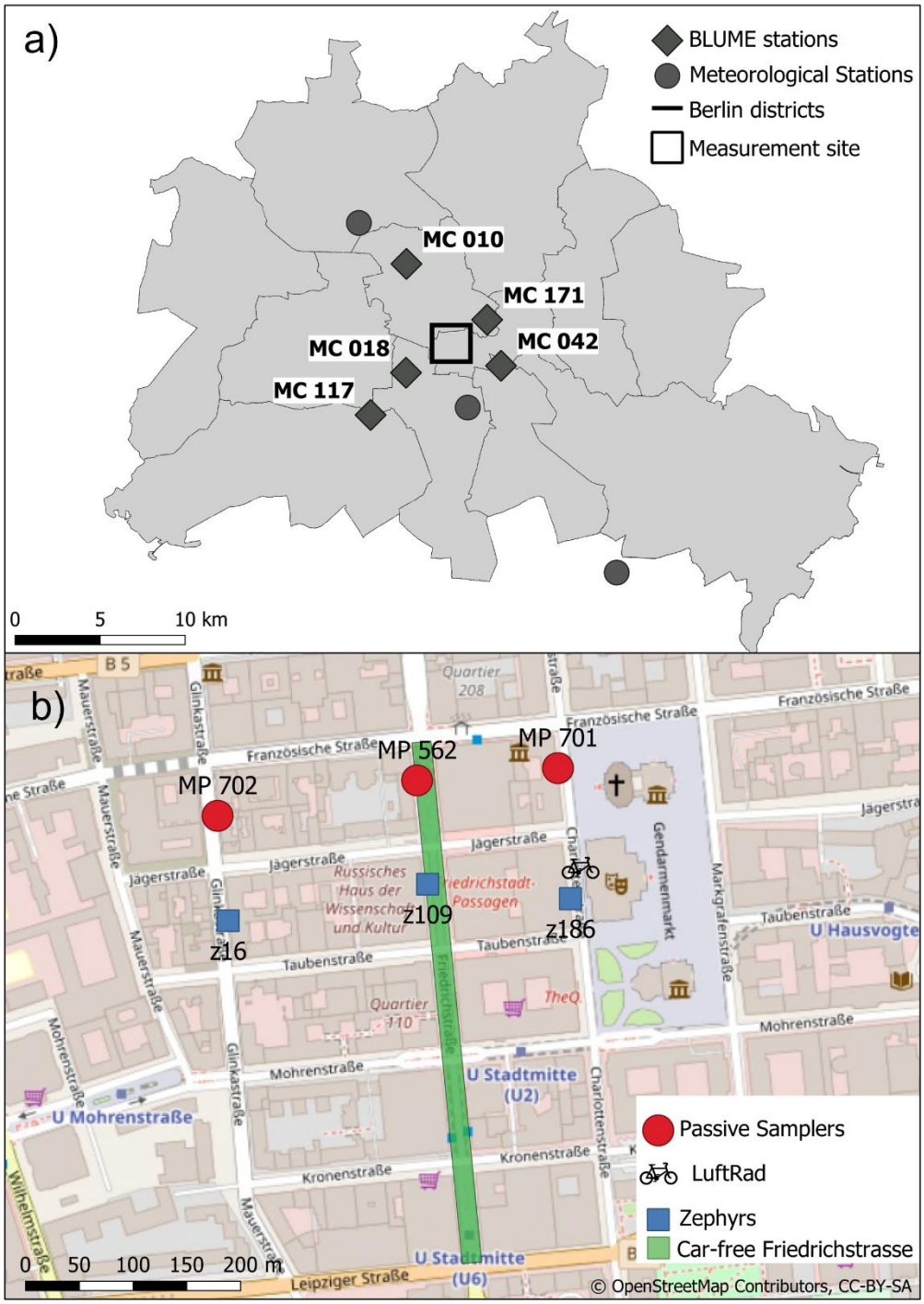

Figure 5-1 a) Map of the city districts of Berlin, including the locations of the four urban background BLUME stations, three meteorological stations, one roadside BLUME station, and the location of the main measurement site; b) map of the measurement site, including the location of the passive samplers, the EarthSense Zephyrs, and the LuftRad. Highlighted in green is the portion of Friedrichstrasse closed to car traffic for the policy.

Table 5-1 Relevant dates regarding the sampling campaign.

	Start	End
Co-location 1	05.02.2020	18.02.2020
Co-location 2	14.05.2020	02.06.2020
Experiment start	13.06.2020	
Motorized traffic allowed		21.08.2020
Co-location 3	24.07.2020	30.07.2020
Public space adaptation	21.08.2020	28.08.2020
Reopening without motorized traffic	28.08.2020	
Co-location 4	06.11.2020	18.11.2020
Stringent lockdown	16.12.2020	
Experiment end		01.02.2021
Co-location 5	04.03.2021	16.03.2021
Co-location 6	24.06.2021	07.07.2021

5.3 Results

5.3.1 City centre nitrogen dioxide

Figures 5-2, 5-3 and 5-4 show the time series of the Zephyr data at the city centre sites. The dashed vertical lines represent the closure and the reopening of the Friedrichstrasse, between which arrangements for the new, car-free, street space were implemented. In the measurement period before the closure of the Friedrichstrasse, the NO_2 hourly concentrations measured by the Zephyrs were in the ranges (first and ninety-ninth percentiles) 0–56, 0–47 and 0–50 µg.m^{-3} at Friedrichstrasse, Glinkastrasse and Charlottenstrasse, respectively. The concentration ranges when the Friedrichstrasse was car-free were 2.7–57, 1.3–59 and 4.2–58 µg.m^{-3} at Friedrichstrasse, Glinkastrasse and Charlottenstrasse, respectively.

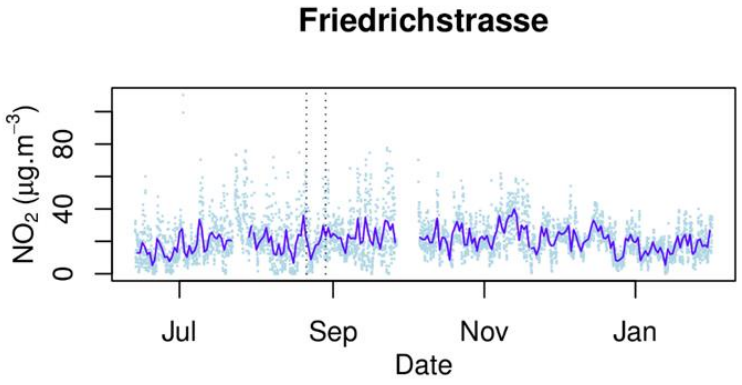

Figure 5-2 NO_2 concentration time series at the Friedrichstrasse site. The dots are the hourly values, the line shows the daily averages. The dashed vertical lines represent the closure and the reopening of the Friedrichstrasse, after which the street was car-free. The periods without data are the times when the Zephyr instruments were taken for co-location.

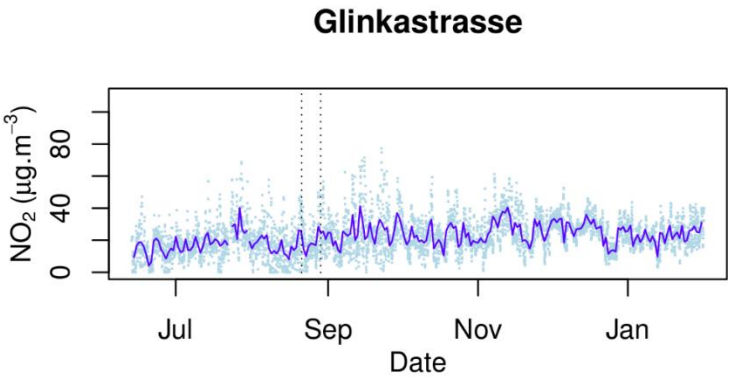

Figure 5-3 NO_2 concentration time series at the Glinkastrasse site. The dots are the hourly values, the line shows the daily averages. The dashed vertical lines represent the closure and the reopening of the Friedrichstrasse, after which the street was car-free. The periods without data are the times when the Zephyr instruments were taken for co-location.

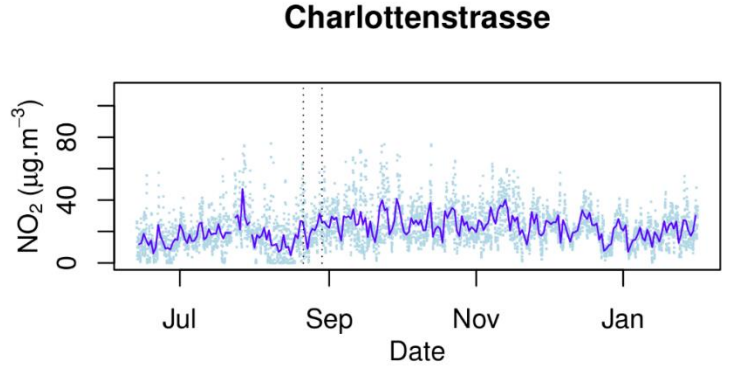

Figure 5-4 NO_2 concentration time series at the Charlottenstrasse site. The dots are the hourly values, the line shows the daily averages. The dashed vertical lines represent the closure and the reopening of the Friedrichstrasse, after which the street was car-free. The periods without data are the times when the Zephyr instruments were taken for co-location.

5.3.2 Comparison of the LCS data with the passive samplers

In order to confirm, beyond the calibration procedure, the reliability of the Zephyr data, we performed a comparison with the data from the passive samplers deployed one block away (see Figure 5-1). The passive samplers have a temporal resolution of two weeks. Figure 5-5 presents the comparison of the NO_2 concentrations from the passive samplers with the average NO_2 concentrations and the 10^{th}, 25^{th}, 75^{th} and 90^{th} from the Zephyrs during the same two weeks. The datasets compare well given the difference in measurement method and measurement location. The passive sampler measurements lie within the inter-quartile range of the Zephyr measurements except in three cases, where it is slightly outside this range.

5.3.3 Traffic data

Figure 5-6 shows the traffic counts at Friedrichstrasse and the two parallel streets Glinkastrasse and Charlottenstrasse. The first two days for which traffic data is available, both before the Friedrichstrasse closure, took place during the summer, a time of year where traffic intensity throughout the city is lower. The Friedrichstrasse closure is clear in the data. The effect of the lockdown is also clear on the counts from 2021-01-21, clearly lower (by approximately one third) than the peaks on 2020-09-10, 2020-10-06, and 2020-

12-15. The sampling strategy for the traffic counts does not allow to evaluate if the closure of Friedrichstrasse had an impact on the traffic at the two parallel streets. The phenomena identified as the sources of the fluctuations at the three streets are expected to uniformly impact the city centre as a whole.

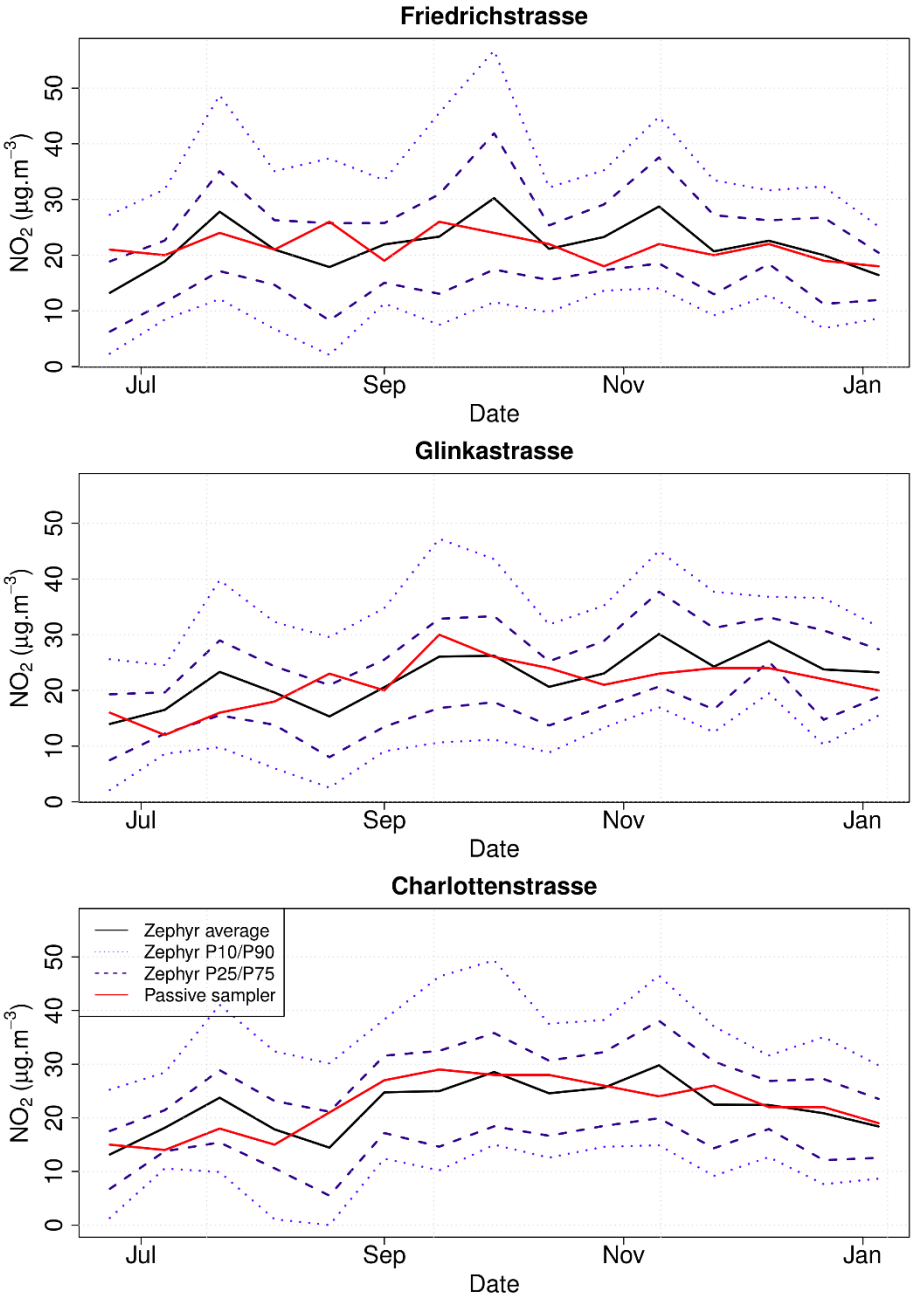

Figure 5-5 NO_2 concentration, comparison between the passive samplers and the Zephyrs average, 10th, 25th, 75th and 90th percentiles over the corresponding passive sampling resolution (two weeks).

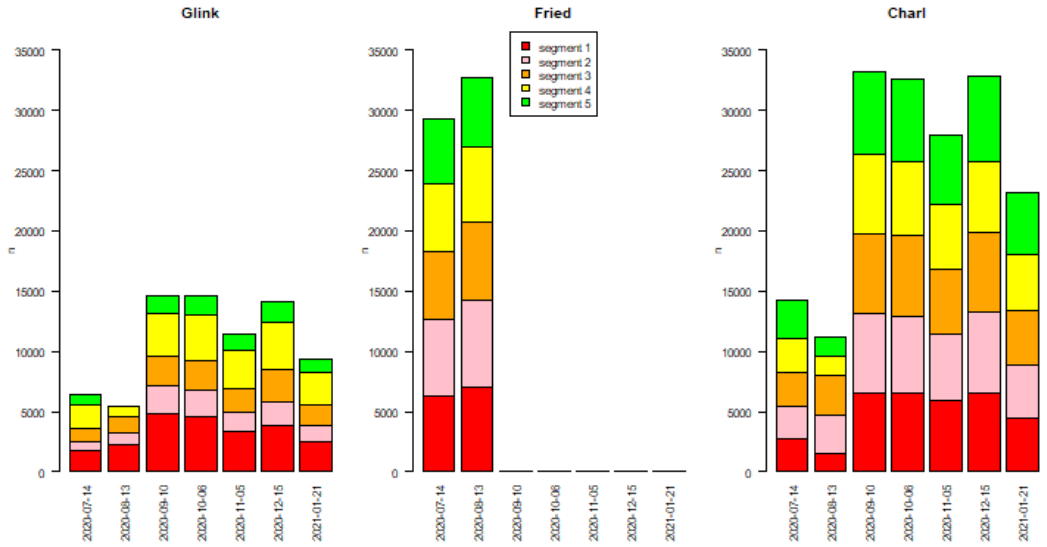

Figure 5-6 Traffic counts at Friedrichstrasse and the two parallel streets Glinkastrasse and Charlottenstrasse. The counts represent the light traffic between 7:00 and 19:00, local time, on 7 days: 2 days before the closure of the Friedrichstrasse (2020-07-14 and 2020-08-13), 4 days after the closure but before the lockdown measures (2020-09-10, 2020-10-06, 2020-11-05 and 2020-12-15) and 1 day after the closure and during the lockdown measures (2021-01-21). The counts are detailed for five segments (see Section 2.3), each representing a subsection of the street between two intersections.

5.3.4 Comparison before and after the intervention

Vehicles were allowed on the Friedrichstrasse until 2020-08-21 04:00:00, when construction started. The street reopened one week later on 2020-08-28 22:00:00, without vehicle traffic. In the following analysis, we subtract (on an hourly basis) the average urban background (UB) concentration from the concentration measured at the experiment sites. UB sites are representative for several square kilometres, influenced by the integrated contribution from all sources upwind and not dominated by a single source (European Parliament, 2008). The resulting variable is called the NO_2 roadside impact and can be understood as normalized NO_2 concentration. With this approach we not only take into consideration the weather variability but also other phenomena which impacted the air quality within the city, including the effects of the various COVID-19-related lockdown phases, in an effort to isolate the signal produced by the intervention. The UB stations considered (Figure 5-1) are: Wedding (MC 010) to the Northwest, Schöneberg (MC 018) to the Southwest, Mitte (MC 171) to the East, and Neukölln (MC 042) to the Southeast.

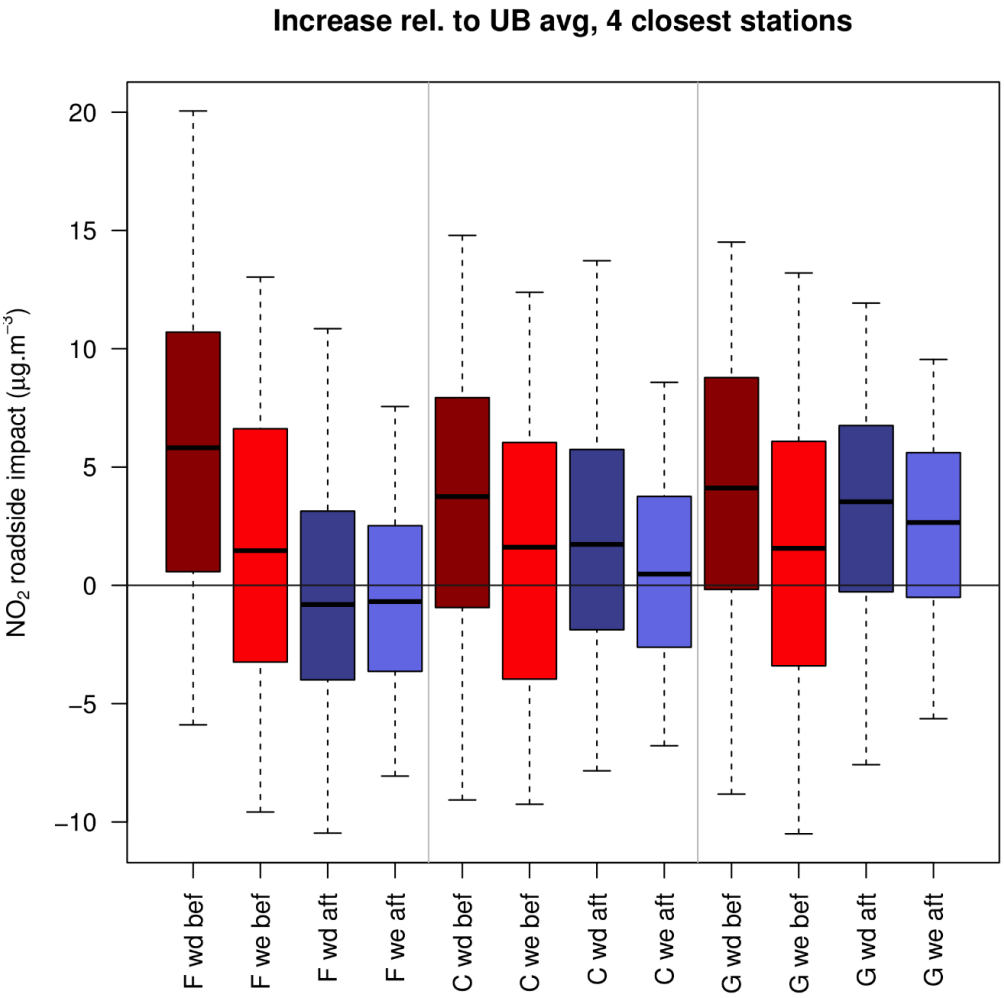

Figure 5-7 shows that before the intervention, NO_2 concentrations at the three streets were clearly above the urban background on weekdays (Monday to Friday), with a lesser difference on weekends (Saturday – Sunday). After the intervention, the concentrations at Friedrichstrasse were clearly lower than before, relative to the urban background, with little to no differentiation between weekdays and weekends. For Charlottenstrasse, there is a drop on weekends but not so much on weekdays, whereas for Glinkastrasse the increment relative to the urban background was maintained after the intervention (weekdays) or even increased (weekends). Besides the closure to traffic, the introduction of stringent COVID-19 lockdown policies introduced in mid-December 2020 is another factor that could have strongly influenced local concentrations.

To test the hypothesis that local concentrations were reduced and became similar to urban back- ground levels and in order to disentangle the effects of the intervention and that of the stringent lockdown related to the COVID-19 pandemic, we compare the concentrations at the three experiment sites with the average of the 4 closest UB stations (see Figure 5-1) by means of the Student's t-test, the Wilcoxon-Mann-Whitney U-test and the Kolmogorov-Smirnov test (K-S-test). The null hypothesis of the Student's t-test states that the true difference in means is equal to 0. The null hypothesis of the Wilcoxon-Mann-Whitney U-test and of the K-S-test (both two-sided) state that both the distributions are equal. If the p-value of the test is very low, the null hypothesis is rejected. Table 5-2 shows the p-values of the statistical tests conducted to compare the concentrations at Friedrichstrasse, Glinkastrasse and Charlottenstrasse with those at the 4 closest UB stations. The tests were conducted for the time intervals before (June 16th – August 21st, 2020), after the intervention without stringent COVID-19 lockdown policies (August 28th – December 16th, 2020), and after the intervention with stringent COVID-19 lockdown

policies (December 16th, 2020 – February 1st, 2021), discriminated by weekdays and weekends.

Table 5-2 Results of the statistical tests (p-value) comparing the measurements in the city centre area and the urban background measurements from the Berlin monitoring network. For a simplified reading, p-values which indicate a difference are written in red, whereas those indicating a similitude are written in blue.

	t-test	*U-test*	*K-S-test*
Friedrichstrasse			
Before			
weekdays (Mon-Fri)	5.8×10^{-39}	9.8×10^{-45}	0
weekends (Sat-Sun)	5.4×10^{-2}	9.3×10^{-3}	9.1×10^{-87}
After, no lockdown			
weekdays (Mon-Fri)	2.3×10^{-1}	5.9×10^{-3}	1.9×10^{-3}
weekends (Sat-Sun)	1.1×10^{-2}	1.1×10^{-3}	2.4×10^{-4}
After, lockdown			
weekdays (Mon-Fri)	2.1×10^{-11}	2.5×10^{-11}	5.1×10^{-11}
weekends (Sat-Sun)	4.4×10^{-7}	6.2×10^{-6}	2.4×10^{-4}
Glinkastrasse			
Before			
weekdays (Mon-Fri)	3.4×10^{-20}	2.3×10^{-35}	0
weekends (Sat-Sun)	5.5×10^{-2}	1.1×10^{-2}	6.3×10^{-12}
After, no lockdown			
weekdays (Mon-Fri)	3.4×10^{-10}	1.7×10^{-21}	0
weekends (Sat-Sun)	1.0×10^{-5}	8.2×10^{-9}	1.9×10^{-8}
After, lockdown			
weekdays (Mon-Fri)	1.7×10^{-13}	4.5×10^{-15}	9.0×10^{-14}
weekends (Sat-Sun)	8.4×10^{-5}	5.5×10^{-7}	2.0×10^{-6}
Charlottenstrasse			
Before			
weekdays (Mon-Fri)	2.7×10^{-13}	3.7×10^{-20}	0
weekends (Sat-Sun)	1.5×10^{-1}	7.2×10^{-3}	2.1×10^{-9}
After, no lockdown			
weekdays (Mon-Fri)	2.7×10^{-13}	7.2×10^{-22}	0
weekends (Sat-Sun)	3.1×10^{-4}	2.6×10^{-6}	2.2×10^{-6}
After, lockdown			
weekdays (Mon-Fri)	6.5×10^{-1}	5×10^{-1}	6.9×10^{-3}
weekends (Sat-Sun)	1.3×10^{-2}	6.2×10^{-2}	1.4×10^{-1}

The density plots (Figure 5-8) of the NO_2 concentrations show a difference at Friedrichstrasse: before the intervention (blue lines), the concentrations are clearly higher than at the UB sites on weekdays, but more similar on weekends. This is confirmed by the statistical tests (Table 5-2): the means and the distribution are different before the

intervention on weekdays, but similar on weekends. Following the intervention but before the hard lockdown was put in place (green lines on Figure 5-8), Friedrichstrasse became, in terms of NO_2 concentrations, similar to an urban background site, as confirmed by the statistical tests (no differences in means and similarity of the distributions).

Before the intervention, the weekdays concentrations at the side streets Glinkastrasse and Charlottenstrasse (Figures 5-9 and 5-10, blue lines) were also higher than at the UB sites, which is confirmed by the very low p-values. On weekends, concentrations were more similar. Unlike for Friedrichstrasse, the intervention (until the stringent lockdown policies were put in place) may have had some impact on the concentrations, in that the distribution shape is closer to the shape of the distribution at UB. However, the low p-values indicate that these are not significant.

After the stringent lockdown policies were put in place (16 December 2020), the very low p-values of the statistical tests show that concentrations were again significantly different between Friedrichstrasse and the UB, both in terms of averages and in terms of distribution. However, unlike before the intervention, the density plots show that concentrations at Friedrichstrasse were lower than at the UB (Figure 5-8, red lines). The side street Charlottenstrasse became, in terms of NO_2 concentrations, similar to an UB site (Figure 5-10, red lines), whereas Glinkastrasse still exhibited higher NO_2 concentrations. A possible explanation for that difference could be that Charlottenstrasse is more commercial whereas Glinkastrasse is starting to become more residential. The three streets under investigation are dominated by services and residents are very few. During the stringent lockdown, with the closure of commerce, traffic probably dropped there more than in other parts of the city, leading to localized lower NO_2 concentrations.

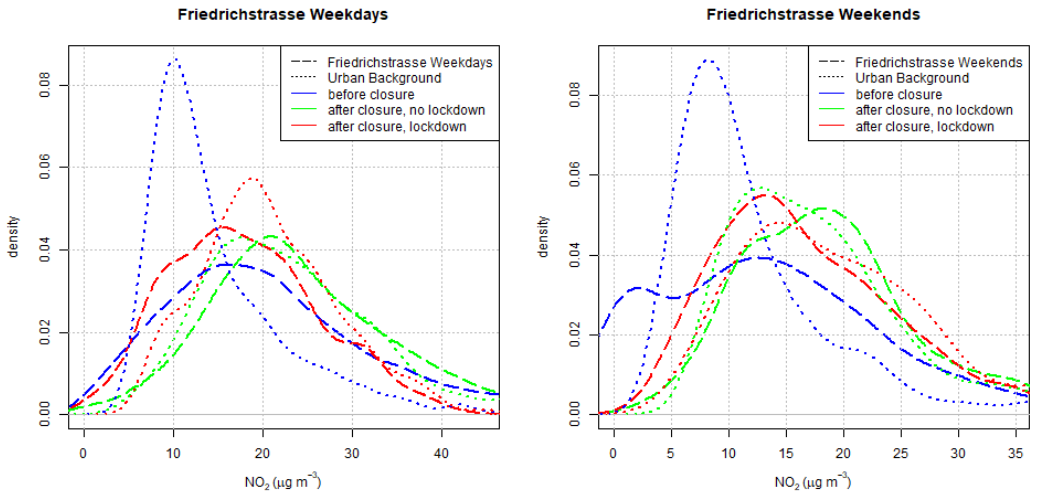

Figure 5-8 Density plots for the NO_2 concentration at Friedrichstrasse on weekdays and weekends: before the intervention, after the intervention and before the stringent lockdown was put in place, after the stringent lockdown was put in place.

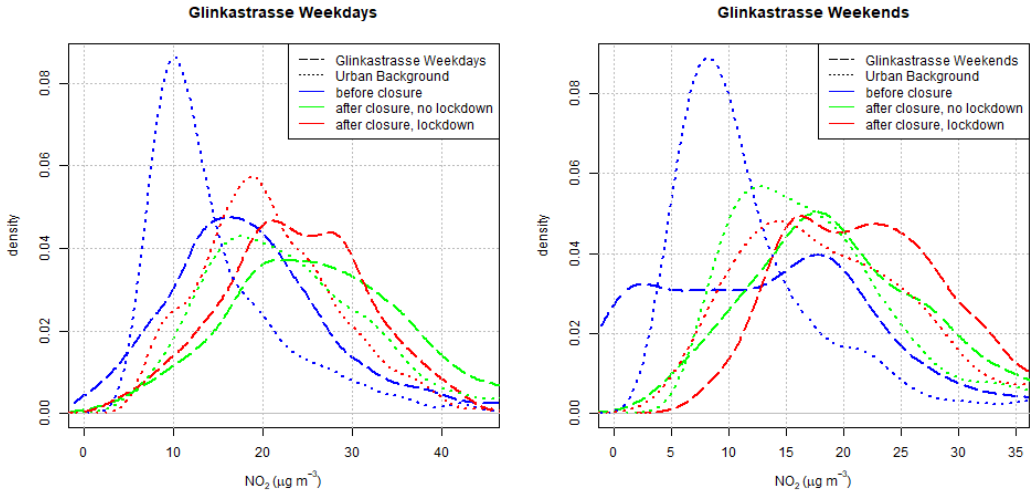

Figure 5-9 Density plots for the NO_2 concentration at Glinkastrasse on weekdays and weekends: before the intervention, after the intervention and before the stringent lockdown was put in place, after the stringent lockdown was put in place.

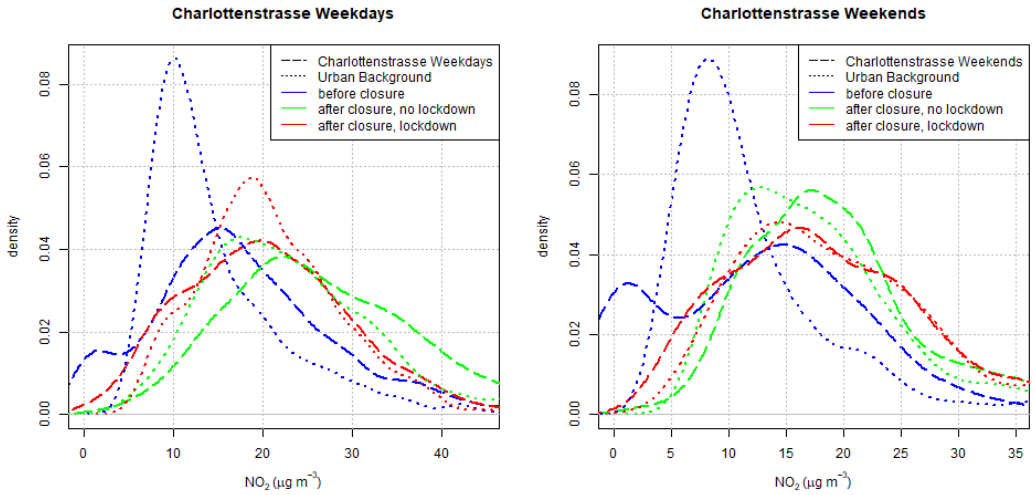

Figure 5-10 Density plots for the NO_2 concentration at Charlottenstrasse on weekdays and weekends: before the intervention, after the intervention and before the stringent lockdown was put in place, after the stringent lockdown was put in place.

5.3.4.1 *Sensitivity analysis*

The seven-step methodology proposed by Schmitz, Towers, et al. (2021b) and used in the present work also outputs upper and lower bounds for the individual, 5-minute, measurements of the Zephyrs. That range represents the combination of the uncertainty from the model and the uncertainty from the reference instrument used to calibrate the LCS units. In order to check our results against that particular source of uncertainty, the same statistical tests were run with the upper and with the lower bounds of the Zephyr measurements. The results are presented in Table 5-3 (for weekdays only).

The derived p-values show that when using the lower bounds, the NO_2 levels at Friedrichstrasse are of the same magnitude as those at the urban background before the intervention. At Glinkastrasse and Charlottenstrasse the concentrations before the intervention are above the urban background level. After the intervention the

concentrations at the three locations evolve, with the different lockdown situations, towards being clearly lower than at the urban background. When considering the upper bounds, concentrations in the city centre are always higher than at the urban background. This sensitivity analysis shows that the interval given by the seven-step methodology for the city centre concentrations measured by the Zephyrs encompasses the urban background levels. This highlights the fact that LCS technologies, together with the associated evaluation algorithm, are not a replacement for reference hardware and although the change produced by the intervention can be evaluated, its exact magnitude is uncertain.

Table 5-3 Results of the statistical tests (p-value) comparing the upper and lower bounds of the measurements in the city centre area and the urban background measurements from the Berlin monitoring network (weekdays only).

Lower bounds	*t-test*	*U -test*	*K-S-test*
Friedrichstrasse			
Before	7.7×10^{-1}	3.9×10^{-5}	0
After, no lockdown	1.7×10^{-57}	2.1×10^{-69}	0
After, lockdown	1.8×10^{-114}	1.5×10^{-100}	0
Glinkastrasse			
Before	7.9×10^{-31}	1.6×10^{-47}	0
After, no lockdown	1.4×10^{-56}	4.1×10^{-57}	0
After, lockdown	1.6×10^{-47}	1.9×10^{-40}	0
Charlottenstrasse			
Before	1.4×10^{-6}	3.4×10^{-15}	0
After, no lockdown	3.3×10^{-13}	5.7×10^{-13}	8.2×10^{-14}
After, lockdown	5.6×10^{-44}	2.0×10^{-42}	0
Upper bounds			
Friedrichstrasse			
Before	8.2×10^{-142}	3.3×10^{-174}	0
After, no lockdown	8.1×10^{-75}	2.0×10^{-106}	0
After, lockdown	3.2×10^{-28}	6.7×10^{-26}	0
Glinkastrasse			
Before	5.8×10^{-176}	8.2×10^{-200}	0
After, no lockdown	1.1×10^{-164}	1.1×10^{-202}	0
After, lockdown	5.2×10^{-156}	1.1×10^{-130}	0
Charlottenstrasse			
Before	1.6×10^{-78}	2.6×10^{-106}	0
After, no lockdown	3.9×10^{-101}	4.8×10^{-133}	0
After, lockdown	1.6×10^{-39}	4.9×10^{-36}	0

A second sensitivity analysis was conducted using different co-location data. The analysis was repeated using data from co-locations 1–6, whereas the main analysis was conducted using a calibration derived from data collected during co-locations 2–4 (which occurred within the experiment dates). The repeated analysis (Table 5-4) shows a similar trend as the main analysis (NO_2 concentrations at Friedrichstrasse were higher before the intervention, lower after the lockdown measures were put in place and intermediate in-between). The statistical significance of the relation to the UB NO_2 concentrations is different: the concentrations at Friedrichstrasse were still significantly higher after the

intervention, a situation that only changed when the lockdown measures were put in place. We trace back such behaviour to the introduction of higher concentrations in the calibration process (the additional co-location periods took place within the city and outside the COVID-19-related lockdown) and show the importance of using appropriate co-location data, as argued by Schmitz, Towers, et al. (2021b).

A third sensitivity analysis was done comparing the output from two models: MLR (used for the main analysis) and RF (also output in the calibration process). Results from the RF model (p-values in Table 5-5) show the same trend as the MLR-based analysis: NO_2 concentrations at Friedrichstrasse were clearly larger than at the UB before the intervention, similar after the intervention without lockdown and clearly lower after the lockdown measures were put in place. This shows that our findings are robust with respect to the model used to derive calibrated NO_2 concentrations.

Table 5-4 Results of the statistical tests (p-value) comparing the NO_2 concentrations obtained from 6 co-location experiments for the city centre area and the urban background measurements from the Berlin monitoring network (weekdays only).

	t-test	U -test	K-S-test
Friedrichstrasse			
Before	1.3×10^{-122}	1.1×10^{-169}	0
After, no lockdown	1.1×10^{-7}	5.8×10^{-16}	0
After, lockdown	4.3×10^{-3}	7.3×10^{-3}	6.9×10^{-3}
Glinkastrasse			
Before	2.7×10^{-33}	8.2×10^{-53}	0
After, no lockdown	3.4×10^{-32}	2.9×10^{-53}	0
After, lockdown	9×10^{-40}	9.0×10^{-40}	0
Charlottenstrasse			
Before	5.5×10^{-43}	1.6×10^{-66}	0
After, no lockdown	3.5×10^{-47}	2.0×10^{-71}	0
After, lockdown	1.5×10^{-13}	1.3×10^{-12}	5.9×10^{-12}

Table 5-5 Results of the statistical tests (p-value) comparing the NO_2 concentrations obtained from random forest (RF) models for the city centre area and the urban background measurements from the Berlin monitoring network (weekdays only).

	t-test	U -test	K-S-test
Friedrichstrasse			
Before	9.5×10^{-67}	7.8×10^{-101}	0
After, no lockdown	4.3×10^{-1}	7.9×10^{-2}	6.5×10^{-2}
After, lockdown	1.8×10^{-10}	2.1×10^{-10}	2.5×10^{-11}
Glinkastrasse			
Before	2.5×10^{-93}	3.3×10^{-145}	0
After, no lockdown	6.4×10^{-16}	2.6×10^{-28}	0
After, lockdown	2.9×10^{-7}	5.5×10^{-7}	3.9×10^{-6}
Charlottenstrasse			
Before	2.8×10^{-43}	5.4×10^{-79}	0
After, no lockdown	5.4×10^{-14}	1.2×10^{-22}	0
After, lockdown	9.6×10^{-1}	9.4×10^{-1}	5.0×10^{-2}

5.3.5 Kerbside measurements

At Charlottenstrasse we also conducted measurements with reference-grade air quality monitoring equipment installed on a cargo bike (see Section 5.2.1.4). The cargo bike measures NO_2 near exhaust height (1–1.5 m), closer to the breathing zone than the 3 m height at which the Zephyrs are deployed (on lampposts). Figures 5-11 and 5-12 show the data from the cargo bike together with the data from the closest Zephyr on the following days: Thursday 17 September, Wednesday 30 September, Tuesday 13 October, Monday 26 October and Saturday 21 November.

The time series plots (Figure 5-11) show that while the peaks tend to be concurrent in time, the concentration closer to the source is much higher than at 3 m height. The distribution of the points in Figure 5-12 further confirms higher concentrations at the kerbside, but with strong contrast between different days. A larger departure from the unity line towards higher kerbside concentrations hap- pens on 30 September and 21 November than on the other 3 days. Meteorological conditions which favour rapid vertical mixing can be one reason for such an observation. A lesser source strength (less concentrated traffic) may also influence such an outcome.

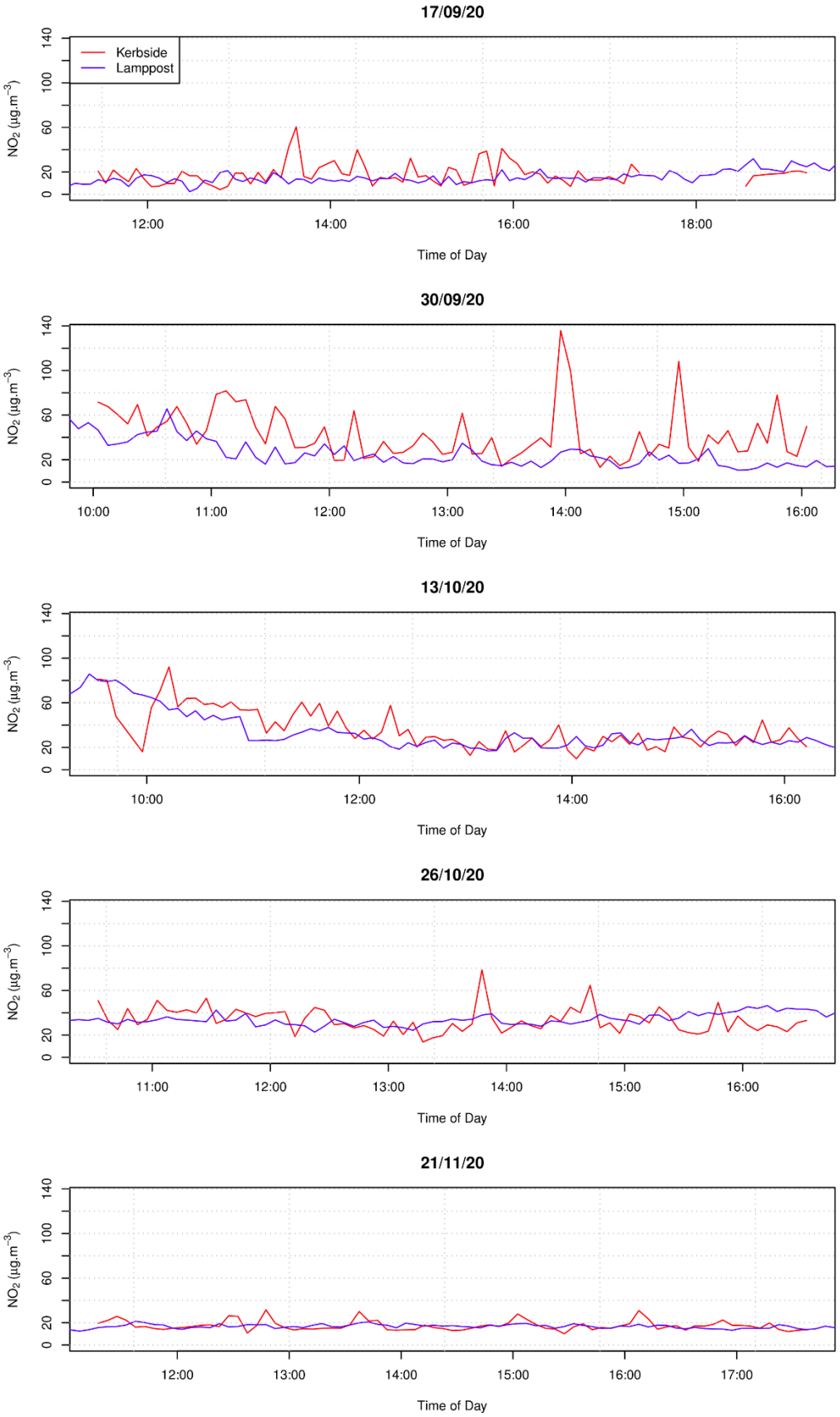

Figure 5-11 Time series of the NO_2 concentrations (5 minutes averages) measured at the lamppost height and at the kerbside at Charlottenstrasse on five different days. Time of day is UTC.

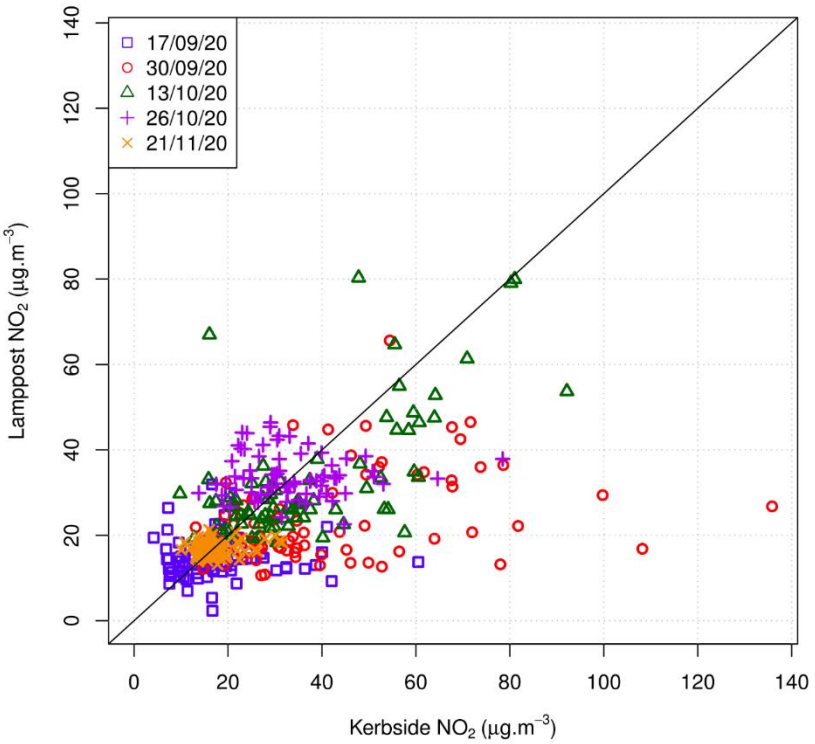

Figure 5-12 Scatter plot of the NO_2 concentrations (5 minutes averages) measured at the lamppost height and at the kerbside at Charlottenstrasse on five different days.

No direct relationship could be found between the extent of the difference in the peaks at the kerbside and on the lampposts and pressure and wind speed (Figure 5-13 and Tables 5-7 and 5-6): poor vertical mixing conditions were observed on September 30th and October 13th: low boundary layer height (Table 5-7) and a stable atmospheric vertical profile. (Figure 5-13). Charlottenstrasse has a roughly N-S orientation, and the sensor was located on the western side of the street. Wind direction influences the horizontal distribution of NO_2 across the street. The most common wind directions, reported in Table 5-7, do not show any matching pattern with the difference in the peaks at the kerbside and on the lampposts. A more complex relationship could possibly be derived, taking into consideration the local traffic, besides meteorological variables. However, detailed local traffic data is not available.

Table 5-6 Distribution statistics of the NO_2 concentrations (5 minutes averages) measured at Charlottenstrasse: lamppost (by Zephyr) and kerbside (by T200).

Date	Time	Location	min	P10	P25	P50	P75	P90	max
17.09.2020	11:25	lamppost	1.6	9.2	12.3	15.8	21	24.3	34.8
Thursday	19:15	kerbside	4.2	7.6	11.1	16.6	20.8	30.5	60.5
30.09.2020	10:00	lamppost	10.6	17	23.4	32.9	45.9	57.4	73.6
Wednesday	16:05	kerbside	13.1	20.8	26	34.7	52.8	71.9	135.8
13.10.2020	09:30	lamppost	15.7	20.7	26	36.4	45.6	65.8	90.4
Tuesday	16:15	kerbside	9.8	17.6	23.6	30.4	48.8	59.8	92.1
26.10.2020	10:30	lamppost	11.7	19.1	22.9	31.3	35.6	39.6	47
Monday	16:35	kerbside	13.8	21.5	25.7	31.4	39.9	44.9	78.5
21.11.2020	11:15	lamppost	8.6	11.3	13.6	16.1	19.6	23	30.8
Saturday	17:40	kerbside	10.1	13.4	14.7	16.5	19	23.2	31.5

Even taking into consideration local meteorology, no clear pattern arises and such results highlight the necessity to take into consideration, when assessing the exposure to air pollutants in cities, the high temporal and spatial variability induced by the meteorology and traffic patterns. This also shows that it is critical to determine measurement location, including height, based on the intended application. In this case, the measurement data shows that to adequately capture population exposure, a difference of a few meters or less in measurement height can have a substantial impact on concentrations, limiting the conclusions one can draw with the data gathered. This is reinforced by the complexity of the relationship between concentrations relevant for exposure and concentrations measured a few metres away vertically.

Table 5-7 Meteorological conditions measured in Berlin on the 5 days where kerbside measurements were made. Pressure, wind speed and direction were measured by the DWD at three stations within the city. The range of the averages for pressure and wind speed and the dominant wind directions are given here. Boundary layer height (BLH) was obtained from the CAMS reanalysis.

Date	Pressure (hPa)	Wind Speed ($m.s^{-1}$)	Wind Direction	BLH (m)
17.09.2020	1021 – 1023	3.2 – 4.8	N and E	726
30.09.2020	1010 – 1011	1.1 – 1.5	N, S and W	34.1
13.10.2020	1009 – 1011	1.8 – 2.9	N, S and W	44.6
26.10.2020	998 – 1000	2.0 – 2.8	S and W	230
21.11.2020	1024 – 1025	4.5 – 6.1	S and SW	154

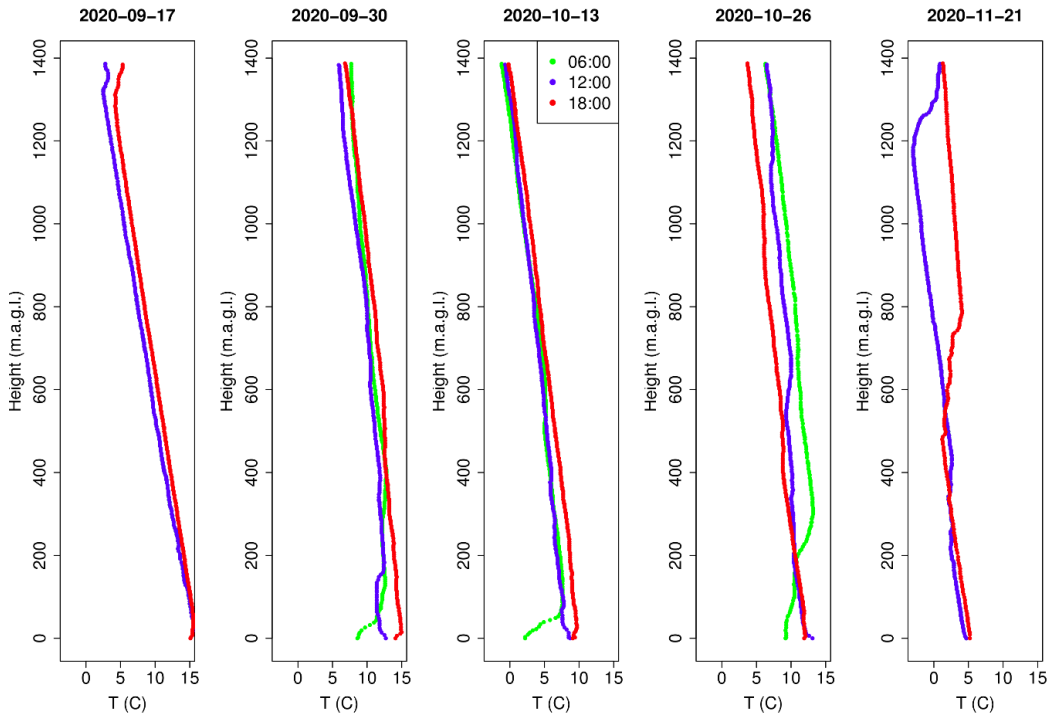

Figure 5-13 Vertical temperature (T in degrees centigrade) profiles (metres above ground level, m.a.g.l.) at Lindenberg during the days where kerbside measurements were taken. The situation evolved from the green line to the blue line and towards the red line over the course of the day (6:00 to 12:00 to 18:00 UTC). Green lines are not shown for the days when the co-location experiment started after 11:00 UTC.

5.4 Discussion

This study has identified several key takeaways regarding the connections between mobility, health, and air pollution. First and foremost, the results of these measurements indicate that mobility policies, such as the street redesign on Friedrichstrasse, which removes cars from the street, reduces local emissions of pollutants, such as nitrogen dioxide and improves local air quality. That the removal or relocation of large volumes of traffic leads to reductions in NO_2 emissions has been known for some time. Such changes were measured during the Beijing Olympics in 2008 (Kelly & Zhu, 2016; Wang & Xie, 2009), have been modelled in various scenario-testing studies (Buchholz, Krein, Junk, Heinemann, & Hoffmann, 2013; Bart Degraeuwe et al., 2017; Holman et al., 2015; Sousa Santos et al., 2020; Steinberga, Sustere, Bikse, Jr, & Kleperis, 2019) and most recently were seen in 2020 as a result of the strict lockdowns in response to the COVID-19 pandemic (Berman & Ebisu, 2020; Brancher, 2021; Gautam, 2020; Skirienė & Stasiškienė, 2021; Erika von Schneidemesser et al., 2021). The results presented here are further proof that a reduction in vehicle traffic leads to concomitant improvements in local air quality but show that this improvement is restricted to levels of urban background pollution in heavily trafficked areas. Other studies have shown that reductions in NO_2 concentrations in connection with traffic- reducing measures may be offset by an associated increase in O_3 pollution, resulting from lower rates of NOx titration under NOx-saturated regimes (Brancher, 2021; Erika von Schneidemesser et al., 2021; Wang & Xie, 2009). To achieve further reductions in air pollution in cities such as Berlin would require holistic measures on a city-wide scale that address emissions of NO_2 from traffic and other key sources, as well as regional scale policies that address O_3 precursor emissions.

These results further highlight the necessity to consider the high temporal and spatial variability induced by changing meteorology and traffic patterns when assessing exposure to air pollutants in cities. Other research has identified poor agreement between kerbside and ambient air pollution measurements (Audrey de Nazelle et al., 2012; Xu et al., 2017). However, we find evidence that even on the same block of the same street, there are differences between 3 m height and kerbside measurements, with clear implications for the calculation of personal exposure to air pollutants. This indicates it is important to carefully consider the impact of measurement location on achieving research goals, and that measurements located a few meters away vertically may already truncate the knowledge one can obtain about a population's air pollution exposure. In this case, even after taking into consideration local meteorology, only a weak relationship between these measurements on Charlottenstrasse arises, reflecting the difficulty to assess exposure from concentrations which were not measured at an ideal height but were rather focused on understanding broader changes to the area.

Last, the results captured in this study were enabled by the flexibility of LCS. Since such small sensors can be readily deployed in higher numbers at lower cost than reference-grade instruments, measurement campaigns can now be conducted at higher spatial and temporal resolution. While other studies have taken advantage of these characteristics to measure in mobile or stationary microenvironments (Lim, Kim, et al., 2019; Lin et al., 2017b; E. von Schneidemesser et al., 2019b) or to establish higher resolution measurement networks (J. M. Barcelo-Ordinas et al., 2019; Mead et al., 2013; Morawska et al., 2018), few studies such as this one use LCS to directly measure changes in air quality in association with mobility policies (Schmitz, Caseiro, et al., 2021). These types of measurements highlight the capacity of LCS to quickly generate valuable insights for decision makers regarding changes in air pollution.

5.4.1 Policy implications

Over the course of the past decade, there has been an increasing policy focus on mitigating air pollution in German cities. With diesel vehicles accounting for a large proportion of air pollutant emissions (NO_2 and PM) in cities, policymakers across Germany have responded by implementing LEZs and other large-scale policies, with only marginal success in improving air quality and human health (J. Cyrys, Peters, Soentgen, & Wichmann, 2014; Margaryan, 2021; Pestel & Wozny, 2021). In light of the recent update to the WHO recommended air pollutant limit values (WHO, 2021), cities will require new policies that are more effective than LEZ to achieve their air quality goals. This study shows that, in the short-term, more stringent policies such as traffic restrictions and street closures can also be successful in reducing local emissions and improving air quality. To achieve larger scale reductions in air pollutant emissions, such as those seen during the lockdowns of the COVID-19 pandemic, cities are likely to require the widespread removal of (diesel) vehicles from their streets. To be effective in the long-term, these policies should be combined with others that promote the use of public transport and alternative mobility options, so as to prevent displacement of emissions and support citizens in switching to alternative transport modes. The Friedrichstrasse case-study exemplifies how such policies can have measurable impacts on local air pollutant concentrations when implemented on a small-scale. With respect to the Berlin Mobility Act and the city's planned mobility transition, improvements in air quality on Friedrichstrasse have already contributed to the narrative of this project as a success story that increased the quality of life for visitors and residents of the street, also serving as evidence for the need of further such policies city-wide (SenUMVK, 2021a, 2021b). To ensure that the impacts of these

policies can be properly quantified going forward, policymakers should further seek to measure changes in air pollution at a higher spatial resolution and in more appropriate locations to better understand the real-world total exposure of their citizens.

5.4.2 Limitations

There are several limitations to this study. First, while it was planned to also measure O_3 during this campaign, there was a technical failure with the reference instrument used during the co-locations that prevented calibration of the raw Zephyr O_3 data. Given that other studies have found increases in O_3 concentrations in connection with traffic-reducing measures, as described previously, the lack of O_3 data in this study prevented us from capturing a more complete picture of changes in air pollution in the study site. In addition, in the present work we did not study the impact of the Friedrichstrasse car-free measure on local concentrations of particulate matter (PM) due to limitations in instrumentation, but a comprehensive assessment of the impact of the measure on the local air quality would have to include it as well. Last, the practicality and lower cost of sensor systems allowed us to conduct this study with a higher spatial resolution of measurements than would have been possible with reference instruments. However, LCS housed in sensor systems are currently less accurate than reference devices. Although we have maximized the accuracy of the data obtained through the LCS and provided representative measures of uncertainty, these are still not, in terms of accuracy, a technological equivalent to reference methods. Therefore, while the general changes in air quality measured in this study are clear, the precise magnitude of the reduction in NO_2 concentrations cannot be determined with precision, as it is shown by the sensitivity analysis. Finally, while real-time traffic count measurements were planned, these ended up being much more limited, also affecting the analysis.

5.5 Conclusions

In the present study, we analyse the effect of an urban intervention – the closure to motorized traffic of a street in central Berlin – on the air quality in the vicinity (the street proper, Friedrichstrasse, and two parallel, adjacent, side streets not closed to traffic, Glinkastrasse and Charlottenstrasse). Nitrogen dioxide concentrations were measured with low-cost systems (EarthSense Zephyrs, one on each street) following a thorough calibration methodology confirmed by a comparison with passive samplers deployed on the same streets.

From our analysis, we can conclude that the intervention had an impact at Friedrichstrasse: the concentrations were brought down to the level of the urban background by the intervention, both on weekdays and weekends. At the side streets Glinkastrasse and Charlottenstrasse, the concentrations of NO_2 after the intervention remained higher than the urban background but did not increase relative to the concentrations prior to the intervention. After the stringent lockdown policies were put in place (16th December 2020), measurements at Friedrichstrasse consistently show lower concentrations than at the urban background. At Friedrichstrasse, the stringent lockdown signal comes on top of the reduction from the intervention. On the side streets, the lockdown brings the concentrations down to the level of the urban background for Charlottenstrasse only.

Our results show that even a relatively small street closure can have a relevant effect on air quality. While the hypothesis was that the emissions on side streets would increase,

due to car drivers using those streets as alternative routes, this was shown to not be the case. We also highlight the difference between measuring at street level versus on lampposts at about 3 m height. While suitable for quantifying the change induced by altered traffic patterns, for exposure purposes we demonstrate that it is important to consider the measurement location carefully, and that it is not straightforward to derive exposure-relevant concentrations from measurements taken even a few meters higher.

The present study was conducted using LCS as the main instruments. The study not only demonstrates the utility of such instruments for urban deployments, but also for evaluation of policy impacts. Despite the calibration methodology used, which focuses on traceability and the maximization of the output concentration accuracy, the concentration output comes associated with an uncertainty range too large to precisely quantify the change in NO_2 concentration associated with the intervention. Instead of a precise quantification, our study based on LCS technology was able to deliver a qualitative, yet policy-relevant, outcome: closing one street in the city centre to traffic brought the NO_2 concentrations in that street down to the level of the urban background, without adversely affecting the concentrations on surrounding streets.

Acknowledgements

We wish to acknowledge the Research Institute for Sustainability, Helmholtz Centre Potsdam for funding the research of Erika von Schneidemesser, Seán Schmitz, and Alexandre Caseiro with financial support provided by the Federal Ministry of Education and Research of Germany (BMBF) and the Ministry for Science, Research and Culture of the State of Brandenburg (MWFK). We also acknowledge the support of the Berlin Senate, Department for the Environment, Urban Mobility, Consumer Protection and Climate Action (SenUMVK), especially Sebastian Clemen and Katja Grunow for their support with the passive sampler deployment and analysis.

Author contributions

Conceptualization: AC, SS, AK, EvS; Data curation: AC, SS; Formal analysis: AC, SS, AK, EvS; Funding acquisition: EvS; Investigation: AC, SS, EvS; Methodology: AC, SS, EvS; Project administration: AK, EvS; Resources: EvS; Software: AC, SS; Supervision: EvS; Validation: AK, EvS; Visualization: AC; Writing - original draft: AC; Writing - review & editing: SS, AK, EvS.

Competing interests

The authors declare no competing interests.

6 Discussion and conclusions

6.1 Summary of findings

This thesis sought to gain several new insights into gas-phase low-cost sensors, including to improve the transparency of their calibration, to understand their suitability for measuring urban NO_2 and O_3 pollution, and to explore new applications for their use, including the production of policy-relevant results. Over the course of 2017-2020 four extensive measurement campaigns were conducted to answer the core research questions posed in this thesis. With these as a framework, this section outlines the key findings and outcomes from Chapters 2-5, as well as recommendations for future work.

(RQ 1) How can the calibration of LCS be standardized, while also made more transparent and accessible for end-users?

With a substantial increase in the number of studies deploying and calibrating LCS in the last several years, standardization of calibration protocols has become a central topic to this field of research. As outlined in Chapter 2, with further contributions from Chapters 3-5, this work has achieved several advances with regards to RQ 1. Primarily, the seven-step methodology was developed and published as open access, demarcating necessary steps for successful, transparent calibration of gas-phase LCS. In a step beyond most published work on calibration, this methodology exemplified best practices for transparency through the reporting of statistical methods and relevant performance metrics at each step, especially during data cleaning and model selection, tuning, and validation. It also served as an example for future research on how to report calibration methods, calling for the calculation of multiple metrics (e.g., R^2, RMSE, MAE) to critically assess model performance. Furthermore, the use of the methodology is not limited to a specific statistical model or technique, though examples with multiple linear regression (MLR) and random forests (RF) are provided. As such, the seven steps are generalized and can be implemented with any model, given that an appropriate level of detail on decision-making regarding model selection, tuning, and validation is provided. The final step established a method for calculating uncertainty associated with the predicted concentrations. By utilizing the median MAE calculated during model validation, an appropriate range of uncertainty was established using field co-located data, ensuring further transparency regarding the accuracy of the LCS measurements. Taken as a whole, this methodology exemplified the potential for calibrating LCS with good performance without the need for black-box proprietary algorithms or complex machine-learning techniques. By publishing all code and data used for the example calibrations in the manuscript in an open-access repository, end-users were able to freely access and reproduce the methodology presented.

With the establishment of this methodology in Chapter 2, a foundation was laid for the work conducted in Chapters 3-5. Once properly calibrated with reference instrumentation using the seven-step methodology, these LCS were able to generate new results regarding the suitability of MOS and EC sensors for the measurement of urban air quality (Chapter 3), and on the impact of changes to transportation infrastructure on local air quality (Chapters 4 and 5). Indeed, the results of these subsequent measurement campaigns strengthen the conclusion that the seven-step methodology produces good calibrations of gas-phase LCS. Of key importance was the repeated field co-location of Zephyrs with reference instruments to account for sensor drift, as well as influences of cross-sensitive species and seasonality. As a result of the good quality in training data and the rigorous pre-processing in the seven-step methodology, the MLR models used throughout Chapters 3-5 performed well in transforming raw sensor data into calibrated NO_2 and O_3 concentrations, as discussed with regards to RQs 2 and 3.

(RQ 2) How suitable are LCS for the measurement of NO_2 and O_3 in urban environments? What new insights into urban NO_2 and O_3 pollution can be identified with high spatial resolution deployments?

Through the compilation of results generated from several measurement campaigns deploying both MOS and EC LCS in street canyons in Berlin, Chapter 3 provided answers to the issues raised in RQ 2. Generally, both MOS and EC captured diurnal patterns of NO_2 and O_3 well, when compared to reference measurements both on-site and in the urban background. Both technologies performed better in measuring O_3 than NO_2, which in part reflected the greater consistency in diurnal patterns of O_3 pollution, as they are driven more by the confluence of chemistry, transport, and meteorology than for NO_2, the concentrations of which are strongly influenced by local NO_x emissions. The construction of generalized additive models (GAMs) using calibrated LCS concentrations and local meteorological parameters further confirmed their suitability for urban measurement of NO_2 and O_3, as variance in MOS and EC measurements were explained by meteorology to a similar extent as reference measurements across all three measurement campaigns. In addition, relationships between calibrated LCS NO_2 and O_3 concentrations and temperature, relative humidity, mixing layer height, wind speed, and wind direction agreed with expected relationships from the literature, as well as with reference measurements locally and in the urban background. This agreement provided additional evidence in support of the hypothesis that EC and MOS sensors properly capture trends in urban air pollution and are suitable for use in such environments.

Of the two technologies, EC sensors perform better than MOS, largely due to reduced inter-sensor variability and longer-term stability, with reduced levels of uncertainty following calibration. The measurement campaign at Strasse des 17. Juni with MOS further showed that their poorer performance worsened in winter, and they exhibited increased inter-sensor variability. This was reflected in the uncertainty ranges calculated in the final step of calibration, which were larger for MOS than for EC, preventing the establishment of clear differences in measurements across sensors at different sites. Conversely, EC sensors showed good performance across seasons and measured NO_2 and O_3 more accurately than MOS, indicating their greater suitability for the measurement of urban air pollution. In addition, the uncertainty ranges calculated during calibration were smaller than for MOS, reflecting longer-term stability and sensitivity of EC in measuring NO_2 and O_3. As a result, this Chapter concluded that the use of MOS for higher spatial

resolution measurements is inadvisable and that EC are preferred, but that both can provide indicative measurements of NO_2 and O_3 pollution.

In Chapter 3, results showed the potential for EC sensors to generate insights into urban pollution dynamics. Measurement campaigns on Frankfurter Allee and Kottbusser Damm using EC sensors demonstrated that concentrations of NO_2 and O_3 varied with prevailing wind direction, wind speed, and photochemical activity in agreement with published results from by computational fluid dynamic (CFD) modelling studies and wind tunnel experiments for idealized street canyons. Additionally, LCS measurements produced insights into the horizontal and vertical gradients at each site. For example, on Kottbusser Damm concentrations of NO_2 at street level and on the 4^{th} floor were not statistically different, whereas on Frankfurter Allee statistical differences in street level and rooftop NO_2 concentrations depended on prevailing wind direction. At both sites, evidence was found for the important role of photochemistry in local concentrations of O_3 and NO_2. On Frankfurter Allee, O_3 concentrations on the southern side of the street were found to be substantially lower than on the northern side, regardless of wind direction, and tracked well with measurements of internal temperature, a proxy for direct insolation. A similar result was found with regards to NO_2 concentrations on the eastern and western sides of Kottbusser Damm, as daytime NO_2 concentrations on the eastern side of the street were lower than on the western side and then equalized at night. Taken together, these results show that high spatial deployments of LCS not only measure expected patterns of street canyon pollution well, but suggest that with further, targeted deployments they could also be used to improve or validate urban-scale pollution models. As such, the measurement campaigns detailed in Chapter 3 represented both a novel application of LCS not yet described in the literature and a pathway to the development of new applications.

Last, in Chapter 5, LCS measurements of NO_2 agreed well with on-site reference measurements as well as passive samplers. On Glinkastrasse, Charlottenstrasse, and Friedrichstrasse, 2-week average NO_2 concentrations measured by the Zephyrs showed good agreement with the 2-week resolution passive sampler measurements taken nearby. As passive samplers have been established as providing reliable measurements of NO_2 concentrations with uncertainty low enough to measure compliance with air quality criteria (Cape, 2009) and show good agreement with chemiluminescence reference methods (Atkins, 1986), this lends greater reliability to the LCS measurements presented throughout this work. Furthermore, passive samplers have been used in a variety of different studies assessing NO_2 concentrations in urban environments (Laxen & Noordally, 1987; Voordeckers, Meysman, Billen, Tytgat, & M, 2021), indicating that, with the agreement shown here between LCS and passive samplers, LCS are suitable for measuring urban air pollution. The good agreement between reference NO_2 measurements taken with the mobile reference measurement station (LuftRad) in Chapter 5 and the stationary LCS lamppost measurements on Charlottenstrasse also underscores this suitability for the measurement of urban air pollution. However, it was also shown that while the 3 m height stationary Zephyr measurements capture local NO_2 concentrations well, they did not register the large peaks measured by the roadside measurements of the LuftRad. This suggests that 3 m height measurements are neither suitable for capturing peak concentrations of NO_2 associated with traffic emissions, nor for quantifying personal levels of NO_2 exposure. These preliminary results open the door for future research to investigate even higher spatial resolution patterns of air pollution, to better understand true levels of exposure in urban environments.

(RQ 3) Can LCS be deployed alongside mobility policy to generate policy-relevant results for decision makers?

In order to effectively address RQ 3, transdisciplinary methods were utilized in the studies described in Chapters 4 and 5. Through collaboration with partners in the Berlin city administration and the municipal district administration, the pop-up bike-lane on Kottbusser Damm, the community space on Böckhstrasse, and the transformation of the Friedrichstrasse to an open space for pedestrians and bicyclists were identified as policies of interest for which quantifying impacts on air quality would be of added value to decision makers. To this end LCS' flexibility in deployment due to their small size and lower maintenance costs was exploited to design measurement campaigns in connection with these measures. On Kottbusser Damm, this involved the deployment of Zephyrs on lampposts along the street, as well as in mobile measurements mounted on bicycles; on Böckhstrasse, one Zephyr was deployed on the 1st floor balcony of a primary school located in the section of the community space; on Friedrichstrasse, Zephyrs were deployed at the site of the street closure, as well as parallel to it on two side-streets. Using sets of before-and-after measurements, these campaigns produced valuable results regarding each policy's impacts on local NO_2 pollution.

Through the mobile measurements conducted on the bike lane on Kottbusser Damm, a reduction of 8.4 ± 7.4 µg.m^{-3} in the median normalized NO_2 concentrations was measured following its implementation, with no substantial reduction measured on the side-streets. At the 95th percentile, this reduction was 14 ± 7.4 µg.m^{-3}, with a slight increase on side-streets. While the uncertainty in the exact values was notable, their trend was clear: cyclists' exposure to NO_2 pollution was reduced through the introduction of a protected bike lane, especially for peak concentrations. On days where the community space on Böckhstrasse was implemented, a reduction in median normalized NO_2 concentrations of 3.7 ± 11.2 µg.m^{-3} was measured. Again, the trend was towards an improvement, but the propagated uncertainty range was too large to reach a definitive conclusion on this policy's impact on local air quality. On the Friedrichstrasse, following the transformation of the street space and removal of vehicle traffic, local NO_2 concentrations were reduced to the level of the urban background. On the side-streets, there was no increase in NO_2 concentrations following the implementation of the open space, providing evidence that displaced vehicle traffic did not increase pollution on adjacent streets.

These results proved to be robust assessments of the impacts of mobility policies on local air pollution, proving the value of LCS to research for science-policy. Furthermore, similar deployments using measurement techniques such as reference instruments or passive samplers would have been more expensive and likely logistically infeasible in the former and at very low temporal resolutions in the latter. As such, the unique flexibility, high temporal resolution, and lower cost of LCS made them the best choice for deployment alongside the implementation of these mobility policies in Berlin. Last, through continuous collaboration with the local city administrators responsible for carrying out and assessing these policies, the results were quickly discussed with decision makers and other interested societal actors, highlighting the important role played by transdisciplinary methods in successfully addressing RQ 3.

6.2 Discussion and recommendations for future research

6.2.1 New potential applications of LCS in the atmospheric sciences

Since their introduction to the field of atmospheric science, low-cost sensors have expanded rapidly into a host of new applications, including the establishment of high spatial resolution measurement networks (Popoola et al., 2018), personal exposure measurements (Jerrett et al., 2017), citizen science (Ripoll et al., 2019), and even environmental justice (Tanzer, Malings, Hauryliuk, Subramanian, & Presto, 2019). Much of the research so far has concentrated on the lowest hanging fruit of these applications, such as establishing higher spatial resolution LCS networks or getting them in the hands of citizen scientists for identification of pollution hot spots. As such, progress in expanding the literature on potential applications has slowed, with much of the focus still given to ideas presented a decade ago (Snyder et al., 2013). In a step towards filling this gap, the work presented here harnessed the unique features of LCS to investigate two heretofore unexplored lines of research.

First, as described predominantly in Chapter 3 but underscored in Chapters 4 and 5, the concept of deploying LCS at high spatial resolution was expanded upon with targeted deployments in street canyons to investigate local patterns of pollution. Previous research conducting measurements in street canyons relied either upon high spatial and low temporal resolution measurements at low-cost with passive samplers (Dėdelė, Miškinytė, & Česnakaitė, 2019; Voordeckers, Meysman, Billen, Tytgat, & Van Acker, 2021) or low spatial but high temporal resolution measurements with expensive reference instruments (S.-B. Park et al., 2015). This work has highlighted the potential for LCS to produce both high spatial and high temporal resolution measurements in street canyons with relative ease and at comparably low cost. Furthermore, results of these measurement campaigns indicated not only that LCS can reliably measure expected street canyon concentrations of NO_2 and O_3, but that they can identify new insights into their dispersal based on morphology and meteorology. With these results as a foundation, future research could expand the knowledge base with measurement campaigns in new urban environments with diverse topologies to improve our understanding of urban-scale air pollution. In addition, these measurements could be coupled with models to validate their results and inform their continued development.

The second novel application of LCS explored in this work, the coupling of measurements to changes in mobility infrastructure to provide policy advice, was enabled by the success of the first. In demonstrating that LCS could measure high spatial resolution distributions of NO_2 pollution reliably and with good accuracy, Chapter 3 paved the way for the production of policy-relevant results in Chapters 4 and 5. As cities change their transportation infrastructure to accommodate more sustainable forms of travel in their pursuit of carbon neutrality, these transformations will have added benefits for local air quality and human health (Pisoni et al., 2019). Therefore, measuring the extent to which such changes improve air quality and human health will only become more important in the near future for decision makers seeking to evaluate these policies. In this regard, this work has highlighted the potential for LCS to directly measure impacts of mobility policies on local air quality, filling an important research gap and providing a foundation for future research in science-policy for years to come.

6.2.2 Standardization of calibration protocols

As detailed throughout this thesis, a substantial body of research has emerged in the last several years investigating a host of different methods for the calibration of LCS. While many have found success, they often come at the cost of transparency, such as when using machine-learning techniques with black-box algorithms (D. E. Williams, 2019), or with substantial transformation of the original raw data signal, such as with double-calibration involving multiple statistical models (Cordero et al., 2018) or with a data fusion approach (Okafor, Alghorani, & Delaney, 2020). In addition, these methods risk overfitting the models to the training data, as recent research suggests (Russell et al., 2022). Overall, given the discrepancies between studies in 1) the level of transparency in reporting methodological choices, 2) the use of performance metrics, 3) the treatment of data and fusion of external parameters, and 4) experimental conditions, it is currently challenging to discern clear best practices for calibration.

However, thanks to recent efforts, some advances have been made towards standardization of calibration protocols and determining best practices for use of LCS. In particular, two recently published reports from the World Meteorological Organization (WMO) on LCS provide a helpful overview on cost, applications, calibration methods, and technologies, including a small set of generalized best practices in deployment (Lewis et al., 2018; Peltier, 2020). Another helpful resource is Schneider et al. (2019), which makes recommendations for a hierarchical classification of processing levels for LCS measurements based on the treatment of data and parameters used in calibration. With regards to calibration methodologies, only a handful of studies provide an overview of existing methodologies used in the literature (e.g. (Jose M. Barcelo-Ordinas et al., 2019; Delaine, Lebental, & Rivano, 2019)), but do not go so far as to establish a recommended calibration protocol.

To fill this gap, the seven-step methodology developed in this thesis has taken several strides towards establishing best practices and standardizing calibration protocols. Primarily, it transparently codifies the steps needed to effectively process and calibrate LCS raw data in a pairwise fashion with reference instrumentation. In addition, the code and data necessary to reproduce the examples provided in Chapter 2 are open source, with over 70 downloads to date (https://zenodo.org/record/4317521, last accessed 13.10.2022). The work has already been taken up in the research community as in one recently submitted paper where the seven-step methodology was implemented for the re-calibration of LCS to measure particulate matter in Grenoble, France (Aix, Schmitz, & Bicout, 2022, Submitted). Second, the examples provided in Chapter 2, as well as the results from Chapters 3-5, show that the seven-step methodology can produce meaningful, reliable concentrations of NO_2 and O_3 with representative ranges of uncertainty without the need for complex machine-learning techniques. Instead, this work emphasizes the value in using well-understood statistical methods such as MLR combined with rigorous scrutiny of raw data to calibrate LCS. This suggests further that future research seeking to improve LCS accuracy should focus less on developing new, complex statistical methods and more on the technology itself, including sensor system design (Russell et al., 2022; David E. Williams, 2020). Last, it joins in the call for the consistent reporting of performance metrics to ensure agreement and comparison across studies (Jose M. Barcelo-Ordinas et al., 2019). If the best practices and seven-step methodology put forth in this work continue to be taken up in the research community, they can serve as a foundation for future efforts to standardize protocols for the pairwise calibration of LCS with reference instrumentation.

6.2.3 Transdisciplinary research for policy-relevant results

Of the several key contributions of this research to the body of literature, the deployment of LCS alongside mobility policies is perhaps the most important, as it contributes useful results in both scientific and societal discourse. As explored previously, air pollution can be considered a 'wicked' problem that pervades most aspects of human society. Improving air quality, particularly in cities, will require action across all levels of government. At the international, regional, national, and local scales, new policies must be implemented to reduce emissions of primary air pollutants (e.g. NO_x), and reduce the production of secondary pollutants (e.g. O_3, $PM_{2.5}$). While atmospheric models and regional monitoring networks can inform macro-scale policies, up until now there have been few effective methods for directly assessing the influence of local-scale measures. As exemplified with the case studies on Kottbusser Damm and Friedrichstrasse, this work has provided local decision makers with a new tool for assessing the impact of small-scale policies on air pollution and informing their further development. While still inappropriate for measuring compliance with air quality limit values, it is now clear that LCS can be deployed alongside mobility policies to measure changes in air quality resultant from their implementation. This is especially relevant in the context of German cities such as Berlin, that are currently attempting a mobility transition away from fossil-fuel powered vehicles. In these cities, many policies take the form of changes to physical mobility infrastructure and are implemented in large part by local policymakers and administrators, with substantial input from citizen actors (D. von Schneidemesser et al., 2020). It is precisely at this scale where the information LCS can provide is most needed, including in supporting the further uptake of such measures and contributing to efforts to reduce air pollution and its harmful impacts on society.

To achieve the goal of providing policy-relevant results to decision makers, this work profited directly from the implementation of transdisciplinary methods. This work made particular use of the three general phases of transdisciplinary research: Phase 1, problem framing and team-building, Phase 2, co-creation of solution oriented transferable knowledge, and Phase 3, re-integration and application of created knowledge (Lawrence et al., 2022). Prior to the conceptualization of the measurement campaigns on Kottbusser Damm and Friedrichstrasse, a network was established with relevant actors from the NGO Changing Cities e.V., the Berlin Senate Department for the Environment, Urban Mobility, Consumer Protection and Climate Action (SenUMVK), and the Friedrichshain-Kreuzberg city district Department for Roads and Parks. Together with these actors during Phase 1, attention was drawn to the policies planned for Kottbusser Damm and Friedrichstrasse and the need for accompanying measurements of air quality was established. In addition, in connection with the pop-up bike lane on Kottbusser Damm, an interdisciplinary team of researchers from RIFS (the LuftMODE project; Luftqualität, Mobilität und Demokratie; English: Air quality, mobility, and democracy) was assembled to measure the various environmental, societal, and political effects of the policy (Becker et al., 2022). With respect to the removal of vehicles from the Friedrichstrasse, the measurements of air quality described in Chapter 5 were part of an interdisciplinary assessment of the impacts of the measure on businesses, traffic, and the environment, conducted in large part by the SenUMVK (SenUMVK, 2022a). These interdisciplinary efforts in both case studies comprised the core of Phase 2 of the transdisciplinary research process.

In the final Phase 3, both official reports and scientific publications were used to communicate and integrate the knowledge gained from each measurement campaign into public, academic, and political discourse. For the Kottbusser Damm campaign, an

initial press release with an accompanying report was published by RIFS and the Berlin Senate detailing the pop-up bike lane's impacts on air pollution exposure for cyclists (E. von Schneidemesser, Schmitz, & Caseiro, 2021), and two peer-reviewed articles were published. Schmitz et al. (2021) described the specific impacts of the measure on local air quality and cyclists' air pollution exposure, as detailed in Chapter 4 of this thesis. Becker et al. (2022) compiled the results of the interdisciplinary LuftMODE research group, ultimately providing commentary on the nature of the policy-making process itself, recommending that decision makers adopt a learning-by-doing approach to new policies to encourage their more rapid uptake following major disruptions such as the COVID-19 pandemic. With regards to the Friedrichstrasse measurement campaign, the results were directly taken up in the SenUMVK's final report on the policy's impact on businesses, traffic, and the environment (SenUMVK, 2022a). This report was used as a primary basis for decision makers to gauge the success of the policy, using its key messages, including the improvements in air quality, to justify the permanent implementation of the car-free zone on Friedrichstrasse (Thomsen & Lühmann, 2021, 2022). The accompanying peer-reviewed article on the impacts of the policy on local air quality, as detailed in Chapter 5, prepared the results for uptake in academic discussions. Last, in both case studies, the results of this research were taken up into public discourse on the future and success of these policies, particularly by the German media (Goldstein, 2021; Tagesspiegel, 2022). As the mobility transition progresses and discussions on the associated importance of air quality provoked by this thesis continue, new knowledge gaps and thereby more avenues for future research will open, providing a basis for the transdisciplinary process to being anew. This thesis thereby represents an example in the application of transdisciplinary methods to produce policy-relevant results.

6.2.4 Recommendations for future research

There are several directions in which future research should expand upon the work presented here. First, the open-source calibration of LCS should be further encouraged to contribute best practices towards their standardization. While this work highlighted an avenue for achieving this with gas-phase sensors, new research can focus on applying the seven-step methodology to LCS for the measurement of PM, as well as to alternative gas-phase MOS and EC sensors from manufacturers not assessed in this work. In addition, following suit with this research, more studies should publish the data and code associated with their calibration methodologies in open-source repositories, to ensure greater transparency and transferability of results to end-users, ultimately assisting efforts at standardizing calibration methodologies. Most importantly, further research seeking to improve LCS performance following calibration should transition away from developing ever more complex black-box methodologies and focus on improving hardware, especially with regards to sensor sensitivity and selectiveness to specific species.

Second, this research opened up a new application for LCS in the high-resolution measurement of air pollution in urban environments. While this study presented promising preliminary results as to the new insights such LCS measurements could provide, future research can confirm or expand upon them by deploying LCS in a host of new environments with unique morphologies. Ideally, this could take a similar form as the large-scale measurement campaign with passive samplers described in Voordeckers et al., (2021), in which micro-scale pollution dynamics were described in a variety of different street canyons with differing aspect ratios. In addition, while this research was able to measure NO_2 and O_3 with LCS effectively, the results could be expanded by including other

important urban air pollutants such as PM and CO, both of which can be measured with existing LCS technologies. New research would ideally also assess the distribution of VOCs in these environments, given their importance in urban chemistry, but most currently available technologies show a lack of appropriate sensitivity and are not selective to particular species to be truly of value (L. Spinelle, Gerboles, Kok, Persijn, & Sauerwald, 2017). Future studies should also investigate coupling LCS measurements with urban-scale air pollution models, to assess LCS' potential for validating model results and informing their further development.

Last, as shown in this work, LCS have great potential for use in both transdisciplinary research and in science-policy. Future research should continue to take advantage of this by working with partners in policy, practice, and industry (e.g., decision makers, NGOs, manufacturers) to identify further gaps in knowledge with regards to the impacts of societal transformations on air quality and co-create the knowledge necessary to fill them. While this work has provided two case studies utilizing such a research approach to provide policy-relevant results, it by no means serves as the only example. Future research could include, for example: 1) citizen science applications in which LCS are put in the hands of various end-users to measure local air pollution and its sources; 2) campaigns assessing personal levels of exposure to air pollution during daily routines in conjunction with health data, or 3) continued assessment of new mobility policies to assess their impacts on emissions and concentrations of multiple air pollutants, as opposed to the focus placed on NO_2 in this work, thereby expanding understanding of how mobility policies influence air quality.

6.3 Conclusions

This thesis has filled several research gaps in the field of low-cost sensors, including their calibration, reliability in measuring urban air pollution, and potential applications, including through transdisciplinary research and the production of policy-relevant results. With data captured during four unique measurement campaigns in different urban environments in Berlin from 2017-2020: 1) an open-source methodology for the calibration of MOS and EC gas-phase LCS was developed, 2) LCS suitability for measuring high spatial resolution concentrations of NO_2 and O_3 in street canyons was quantified, and 3) three different mobility policies were assessed for their impact on local air quality as case studies in LCS' potential for use in transdisciplinary research in science-policy. Key results include:

- An open-source, seven-step methodology for the calibration of gas-phase LCS was developed, with code and data provided in an online repository. Examples of calibration were provided using multiple linear regression (MLR) and random forests (RF), finding that both performed well against new experimental data, but recommended MLR for use, as it is better understood than RF and can predict outside of the range of the training data, which RF cannot (Chapter 2).

- Several best practices for the calibration of gas-phase LCS were established. These include 1) the transparent reporting of methods used for model selection, validation, and tuning; 2) the use of R^2, mean average error (MAE), and root mean squared error (RMSE) as model performance metrics to enhance comparability across calibration studies; 3) rigorous pre-processing of raw sensor data, including outlier removal and data flagging; and 4) the use of median MAE from model

validation with subsets of co-location data to produce a representative measure of predictive uncertainty (Chapter 2).

- As the first study of its kind, LCS were deployed in two novel applications in several experiments in Berlin, including to measure high spatial resolution patterns of air pollution in street canyons and in connection with mobility measures to produce policy-relevant results. Results from these campaigns indicate not only that LCS are suitable for these applications, but that they can provide valuable insights into urban NO_2 and O_3 pollution and its intersection with mobility policy (Chapters 3 – 5).

- Gas-phase LCS such as MOS and EC can be effectively calibrated in pairwise fashion with reference instrumentation to reproduce urban concentrations of NO_2 and O_3. Of the two technologies, EC perform better in measuring both NO_2 and O_3, with MOS subject to greater inter-sensor variability and instability. Of these two species, O_3 can be calibrated more reliably, especially in the summer (Chapters 2 + 3).

- EC sensors can capture high spatial resolution variability in street canyon NO_2 and O_3 pollution in agreement with expected patterns resultant from the confluence of urban morphology, meteorology, and chemistry. Results highlight the important role played by wind direction, wind speed, and photochemistry in diurnal concentrations of NO_2 and O_3 in street canyons and suggest that LCS can be used to develop and validate urban-scale pollution models (Chapter 3).

- Measurements of NO_2 during the Friedrichstrasse campaign with EC sensors agreed well with nearby passive sampler and reference measurements, further underscoring LCS suitability in measuring air pollution. Stationary measurements on Charlottenstrasse at 3 m height did not capture peaks in NO_2 concentrations measured at roadside by the LuftRad, suggesting that 3 m height measurements do not reflect true street-level exposure to NO_2 (Chapter 5).

- In two case studies in Berlin, changes in local air quality were assessed in connection with three unique mobility policies using stationary and mobile measurements. Results show: 1) the implementation of a pop-up bike lane on Kottbusser Damm reduced cyclists' median NO_2 exposure by 8.4 ± 7.4 $\mu g.m^{-3}$, 2) the creation of a temporary community space on Böckhstrasse reduced median concentrations of NO_2 by 3.7 ± 11.2 $\mu g.m^{-3}$ during times when it was in place; and 3) the opening of the Friedrichstrasse to pedestrians and restriction of vehicle traffic resulted in local NO_2 concentrations being reduced to the level of the urban background, with no significant increases on parallel side streets (Chapters 4 + 5).

- Through the application of transdisciplinary methods involving the co-creation of results with partners in policy and practice, this work highlighted the potential for LCS to provide timely, policy-relevant advice for decision makers regarding local changes in air quality. It also serves as an example for future work seeking to engage in transdisciplinary research (Chapters 4 + 5).

7 Bibliography

Aguayo, L., Reichmuth, D., & Weintraub, C. (2021). *Low-and Zero-Emissions Zones*. Retrieved from https:/www.ucsusa.org/resources/low-and-zero-emissions-zones

Aix, M.-L., Schmitz, S., & Bicout, D. (2022, Submitted). Open-source calibration of low-cost sensors for high-quality monitoring of fine particulate matter. *Environmental Science & Technology*.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (pp. 267-281). Budapest, Hungary: Akadémiai Kiadó.

Aldred, R., & Goodman, A. (2020). Low Traffic Neighbourhoods, Car Use, and Active Travel: Evidence from the People and Places Survey of Outer London Active Travel Interventions. *Findings*. doi:10.32866/001c.17128

Aldrin, M., & Haff, I. (2005). Generalised additive modelling of air pollution, traffic volume and meteorology. *Atmospheric Environment, 39*(11), 2145-2155. doi:10.1016/j.atmosenv.2004.12.020

André, M., Carteret, M., Pasquier, A., & Liu, Y. (2017). Methodology for characterizing vehicle fleet composition and its territorial variability, needed for assessing Low Emission Zones. *Transportation Research Procedia, 25*, 3286-3298. doi:10.1016/j.trpro.2017.05.174

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., . . . Hamburg, S. P. (2017). High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environ Sci Technol, 51*(12), 6999-7008. doi:10.1021/acs.est.7b00891

Arain, M. A., Blair, R., Finkelstein, N., Brook, J., & Jerrett, M. (2009). Meteorological influences on the spatial and temporal variability of NO2in Toronto and Hamilton. *The Canadian Geographer / Le Géographe canadien, 53*(2), 165-190. doi:10.1111/j.1541-0064.2009.00252.x

Atkins, D. H. (1986). *The measurement of nitrogen dioxide in the outdoor environment using passive diffusion tube samplers*.

Atkinson, R. (2000). Atmospheric chemistry of VOCs and NOx. *Atmospheric Environment, 34*(12-14), 2063-2101. doi:10.1016/s1352-2310(99)00460-4

Atkinson, R. W., Butland, B. K., Anderson, H. R., & Maynard, R. L. (2018). Long-term Concentrations of Nitrogen Dioxide and Mortality: A Meta-analysis of Cohort Studies. *Epidemiology, 29*(4), 460-472. doi:10.1097/EDE.0000000000000847

Barcelo-Ordinas, J. M., Doudoub, M., Garcia-Vidala, J., & Badache, N. (2019). Self-Calibration Methods for Uncontrolled Environments in Sensor Networks: A Reference Survey. *Ad Hoc Networks, 88*, 142-159.

Barcelo-Ordinas, J. M., Ferrer-Cid, P., Garcia-Vidal, J., Ripoll, A., & Viana, M. (2019). Distributed Multi-Scale Calibration of Low-Cost Ozone Sensors in Wireless Sensor Networks. *Sensors (Basel), 19*(11). doi:10.3390/s19112503

Beaver, S., & Palazoglu, A. (2009). Influence of synoptic and mesoscale meteorology on ozone pollution potential for San Joaquin Valley of California. *Atmospheric Environment, 43*(10), 1779-1788. doi:10.1016/j.atmosenv.2008.12.034

Becker, S., & Renn, O. (2019). Akzeptanzbedingungen politischer Maßnahmen für die Verkehrswende: Das Fallbeispiel Berliner Mobilitätsgesetz. In *Akzeptanz und politische Partizipation in der Energietransformation* (pp. 109-130).

Becker, S., von Schneidemesser, D., Caseiro, A., Götting, K., Schmitz, S., & von Schneidemesser, E. (2022). Pop-up cycling infrastructure as a niche innovation for sustainable transportation in European cities: An inter- and transdisciplinary case study of Berlin. *Sustainable Cities and Society, 87*. doi:10.1016/j.scs.2022.104168

Berman, J. D., & Ebisu, K. (2020). Changes in U.S. air pollution during the COVID-19 pandemic. *Sci Total Environ, 739*, 139864. doi:10.1016/j.scitotenv.2020.139864

Bernardo, V., Fageda, X., & Flores-Fillol, R. (2021a). Dealing with Negative Externalities: Low Emission Zones Versus Congestion Tolls. In R. Vickerman (Ed.), *International Encyclopedia of Transportation* (pp. 231-236). Oxford: Elsevier.

Bernardo, V., Fageda, X., & Flores-Fillol, R. (2021b). Pollution and congestion in urban areas: The effects of low emission zones. *Economics of Transportation, 26-27*. doi:10.1016/j.ecotra.2021.100221

Bigi, A., Mueller, M., Grange, S. K., Ghermandi, G., & Hueglin, C. (2018). Performance of NO, NO2 low cost sensors and three calibration approaches within a real world application. *Atmospheric Measurement Techniques, 11*(6), 3717-3735. doi:10.5194/amt-11-3717-2018

Boogaard, H., Borgman, F., Kamminga, J., & Hoek, G. (2009). Exposure to ultrafine and fine particles and noise during cycling and driving in 11 Dutch cities. *Atmospheric Environment, 43*(27), 4234-4242. doi:10.1016/j.atmosenv.2009.05.035

Boogaard, H., Janssen, N. A., Fischer, P. H., Kos, G. P., Weijers, E. P., Cassee, F. R., . . . Hoek, G. (2012). Impact of low emission zones and local traffic policies on ambient air pollution concentrations. *Sci Total Environ, 435-436*, 132-140. doi:10.1016/j.scitotenv.2012.06.089

Borge, R., Narros, A., Artíñano, B., Yagüe, C., Gómez-Moreno, F. J., de la Paz, D., . . . Vardoulakis, S. (2016). Assessment of microscale spatio-temporal variation of air pollution at an urban hotspot in Madrid (Spain) through an extensive field campaign. *Atmospheric Environment, 140*, 432-445. doi:10.1016/j.atmosenv.2016.06.020

Börjesson, M., Bastian, A., & Eliasson, J. (2021). The economics of low emission zones. *Transportation Research Part A: Policy and Practice, 153*, 99-114. doi:10.1016/j.tra.2021.08.016

Borrego, C., Costa, A. M., Ginja, J., Amorim, M., Coutinho, M., Karatzas, K., . . . Penza, M. (2016). Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise. *Atmospheric Environment, 147*, 246-263. doi:10.1016/j.atmosenv.2016.09.050

Borrego, C., Ginja, J., Coutinho, M., Ribeiro, C., Karatzas, K., Sioumis, T., . . . Penza, M. (2018). Assessment of air quality microsensors versus reference methods: The EuNetAir Joint Exercise – Part II. *Atmospheric Environment, 193*, 127-142. doi:10.1016/j.atmosenv.2018.08.028

Bosello, M., Delnevo, G., & Mirri, S. (2020). *On exploiting Gamification for the Crowdsensing of Air Pollution*. Paper presented at the Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good.

Bousiotis, D., Beddows, D. C. S., Singh, A., Haugen, M., Diez, S., Edwards, P. M., . . . Pope, F. D. (2022). A study on the performance of low-cost sensors for source apportionment at an urban background site. *Atmospheric Measurement Techniques, 15*(13), 4047-4061. doi:10.5194/amt-15-4047-2022

Bousiotis, D., Singh, A., Haugen, M., Beddows, D. C. S., Diez, S., Murphy, K. L., . . . Pope, F. D. (2021). Assessing the sources of particles at an urban background site using both regulatory instruments and low-cost sensors – a comparative study. *Atmospheric Measurement Techniques, 14*(6), 4139-4155. doi:10.5194/amt-14-4139-2021

Brancher, M. (2021). Increased ozone pollution alongside reduced nitrogen dioxide concentrations during Vienna's first COVID-19 lockdown: Significance for air quality management. *Environ Pollut, 284*, 117153. doi:10.1016/j.envpol.2021.117153

Brand, C., Götschi, T., Dons, E., Gerike, R., Anaya-Boig, E., Avila-Palencia, I., . . . Nieuwenhuijsen, M. J. (2021). The climate change mitigation impacts of active travel: Evidence from a longitudinal panel study in seven European cities. *Global Environmental Change, 67*. doi:10.1016/j.gloenvcha.2021.102224

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/a:1010933404324

Bright, V., Bloss, W., & Cai, X. (2011). Modelling atmospheric composition in urban street canyons. *Weather, 66*(4), 106-110. doi:10.1002/wea.781

Brunekreef, B., Strak, M., Chen, J., Andersen, Z. J., Atkinson, R., Bauwelinck, M., . . . Hoek, G. (2021). *Mortality and Morbidity Effects of Long- Term Exposure to Low-Level PM2.5, BC, NO2, and O3: An Analysis of European Cohorts in the ELAPSE Project*. Retrieved from Boston, Massachussetts:

Buchholz, S., Krein, A., Junk, J., Heinemann, G., & Hoffmann, L. (2013). Simulation of Urban-Scale Air Pollution Patterns in Luxembourg: Contributing Sources and Emission Scenarios. *Environmental Modeling and Assessment, 18*(3), 271-283. doi:10.1007/s10666

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C. A., 3rd, . . . Spadaro, J. V. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc Natl Acad Sci U S A, 115*(38), 9592-9597. doi:10.1073/pnas.1803222115

Cape, J. N. (2009). The Use of Passive Diffusion Tubes for Measuring Concentrations of Nitrogen Dioxide in Air. *Critical Reviews in Analytical Chemistry, 39*(4), 289-310. doi:10.1080/10408340903001375

Carslaw, D. C., & Beevers, S. D. (2002). The efficacy of low emission zones in central London as a means of reducing nitrogen dioxide concentrations. *Transportation Research Part D: Transport and Environment, 7*(1), 49-64. doi:10.1016/S1361-9209(01)00008-6

Carslaw, D. C., Beevers, S. D., & Tate, J. E. (2007). Modelling and assessing trends in traffic-related emissions using a generalised additive modelling approach. *Atmospheric Environment, 41*(26), 5289-5299. doi:10.1016/j.atmosenv.2007.02.032

Carslaw, D. C., Farren, N. J., Vaughan, A. R., Drysdale, W. S., Young, S., & Lee, J. D. (2019). The diminishing importance of nitrogen dioxide emissions from road vehicle exhaust. *Atmospheric Environment: X, 1*. doi:10.1016/j.aeaoa.2018.100002

Carslaw, D. C., & Taylor, P. J. (2009). Analysis of air pollution data at a mixed source location using boosted regression trees. *Atmospheric Environment, 43*(22-23), 3563-3570. doi:10.1016/j.atmosenv.2009.04.001

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., . . . Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environ Int, 99*, 293-302. doi:10.1016/j.envint.2016.12.007

Cepeda, M., Schoufour, J., Freak-Poli, R., Koolhaas, C. M., Dhana, K., Bramer, W. M., & Franco, O. H. (2017). Levels of ambient air pollution according to mode of transport: a systematic review. *The Lancet Public Health, 2*(1), e23-e34. doi:10.1016/s2468-2667(16)30021-4

Chen, J., & Hoek, G. (2020). Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. *Environ Int, 143*, 105974. doi:10.1016/j.envint.2020.105974

Chew, L. W., Glicksman, L. R., & Norford, L. K. (2018). Buoyant flows in street canyons: Comparison of RANS and LES at reduced and full scales. *Building and Environment, 146*, 77-87. doi:10.1016/j.buildenv.2018.09.026

Cordero, J. M., Borge, R., & Narros, A. (2018). Using statistical methods to carry out in field calibrations of low cost air quality sensors. *Sensors and Actuators B: Chemical, 267*, 245-254. doi:10.1016/j.snb.2018.04.021

Cross, E. S., Williams, L. R., Lewis, D. K., Magoon, G. R., Onasch, T. B., Kaminsky, M. L., . . . 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. doi:10.5194/amt-10-3575-2017

Cruz, C., & Montenon, A. (2016). Implementation and Impacts of Low Emission Zones on Freight Activities in Europe: Local Schemes Versus National Schemes. *Transportation Research Procedia, 12*, 544-556. doi:10.1016/j.trpro.2016.02.010

Cyrys, J., Peters, A., Soentgen, J., & Wichmann, H. E. (2014). Low emission zones reduce PM10 mass concentrations and diesel soot in German cities. *J Air Waste Manag Assoc, 64*(4), 481-487. doi:10.1080/10962247.2013.868380

Cyrys, J., Wichmann, H., Rückerl, R., & Peters, A. (2018). Umweltzonen in Deutschland. *Bundesgesundheitsblatt, 61*(6), 645-655. doi:10.1007/s00103

De Borger, B., & Proost, S. (2013). Traffic externalities in cities: The economics of speed bumps, low emission zones and city bypasses. *Journal of Urban Economics, 76*, 53-70. doi:10.1016/j.jue.2013.02.004

de Nazelle, A., Bode, O., & Orjuela, J. P. (2017). Comparison of air pollution exposures in active vs. passive travel modes in European cities: A quantitative review. *Environ Int, 99*, 151-160. doi:10.1016/j.envint.2016.12.023

de Nazelle, A., Fruin, S., Westerdahl, D., Martinez, D., Ripoll, A., Kubesch, N., & Nieuwenhuijsen, M. (2012). A travel mode comparison of commuters' exposures to air pollutants in Barcelona. *Atmospheric Environment, 59*, 151-159. doi:10.1016/j.atmosenv.2012.05.013

De Vito, S., Esposito, E., Salvato, M., Popoola, O., Formisano, F., Jones, R., & Di Francia, G. (2018). Calibrating chemical multisensory devices for real world applications: An in-depth comparison of quantitative machine learning approaches. *Sensors and Actuators B: Chemical, 255*, 1191-1210. doi:10.1016/j.snb.2017.07.155

Dėdelė, A., Miškinytė, A., & Česnakaitė, I. (2019). Comparison of Measured and Modelled Traffic-Related Air Pollution in Urban Street Canyons. *Polish Journal of Environmental Studies, 28*(5), 3115-3123. doi:10.15244/pjoes/93744

Degraeuwe, B., Pisoni, E., Christidis, P., Christodoulou, A., & Thunis, P. (2021). SHERPA-city: A web application to assess the impact of traffic measures on NO2 pollution in cities. *Environ Model Softw, 135*, 104904. doi:10.1016/j.envsoft.2020.104904

Degraeuwe, B., Thunis, P., Clappier, A., Weiss, M., Lefebvre, W., Janssen, S., & Vranckx, S. (2017). Impact of passenger car NOX emissions on urban NO2 pollution – Scenario analysis for 8 European cities. *Atmospheric Environment, 171*, 330-337. doi:10.1016/j.atmosenv.2017.10.040

Delaine, F., Lebental, B., & Rivano, H. (2019). In Situ Calibration Algorithms for Environmental Sensor Networks: A Review. *IEEE Sensors Journal, 19*(15), 5968-5978. doi:10.1109/jsen.2019.2910317

Derwent, R. (2004). Intercontinental transport and the origins of the ozone observed at surface sites in Europe. *Atmospheric Environment, 38*(13), 1891-1901. doi:10.1016/j.atmosenv.2004.01.008

Dias, D., Amorim, J. H., Sá, E., Borrego, C., Fontes, T., Fernandes, P., . . . Tchepel, O. (2018). Assessing the importance of transportation activity data for urban emission inventories. *Transportation Research Part D: Transport and Environment, 62*, 27-35. doi:10.1016/j.trd.2018.01.027

Dias, D., Tchepel, O., & Antunes, A. P. (2016). Integrated modelling approach for the evaluation of low emission zones. *Journal of Environmental Management, 177*, 253-263. doi:10.1016/j.jenvman.2016.04.031

DIN. (2012). Außenluft - Messverfahren zur Bestimmung der Konzentration von Stickstoffdioxid und Stickstoffmonoxid mit Chemilumineszenz. *Deutsche Fassung EN 14211:2012*. doi:https://dx.doi.org/10.31030/1852617

Dominici, F., Schwartz, J., Di, Q., Braun, D., Choirat, C., & Zanobetti, A. (2019). Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1. *Research Reports: Health Effects Institute, 2019*, 200. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7300216/

Dons, E., Laeremans, M., Anaya-Boig, E., Avila-Palencia, I., Brand, C., de Nazelle, A., . . . Int Panis, L. (2018). Concern over health effects of air pollution is associated to NO2 in seven European cities. *Air Quality, Atmosphere & Health, 11*(5), 591-599. doi:10.1007/s11869-018-0567-3

Dons, E., Laeremans, M., Orjuela, J. P., Avila-Palencia, I., Carrasco-Turigas, G., Cole-Hunter, T., . . . Int Panis, L. (2017). Wearable Sensors for Personal Monitoring and Estimation of Inhaled Traffic-Related Air Pollution: Evaluation of Methods. *Environ Sci Technol, 51*(3), 1859-1867. doi:10.1021/acs.est.6b05782

Dons, E., Laeremans, M., Pablo Orjuela, J., Avila-Palencia, I., de Nazelle, A., Nieuwenhuijsen, M., . . . Int Panis, L. (2019). Transport most likely to cause air pollution peak exposures in everyday life: Evidence from over 2000 days of personal monitoring. *Atmospheric Environment, 213*, 424-432. doi:10.1016/j.atmosenv.2019.06.035

EarthSense. Zephyr® Air Quality Monitor. Retrieved from www.earthsense.co.uk/Zephyr

EC. (2019). Report from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Comittee of the Regions -- The First Clean Air Outlook. (COM(2018) 446 final/2).

EC. (2021). Report from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Comittee of the Regions -- The Second Clean Air Outlook. (COM(2021) 3 final).

EEA. (2007). *Air pollution in Europe 1990--2004*. Retrieved from European Environment Agency, Kongens Nytorv 6, 1050 Copenhagen K, Denmark: http://www.eea.europa.eu/publications/eea_report_2007_2

EEA. (2018). *Europe's urban air quality -- re-assessing implementation challenges in cities*. Retrieved from European Environment Agency, Kongens Nytorv 6, 1050 Copenhagen K, Denmark: https://www.eea.europa.eu/publications/europes-urban-air-quality

EEA. (2019). *Air pollution in Europe -- 2019 report*. Retrieved from European Environment Agency, Kongens Nytorv 6, 1050 Copenhagen K, Denmark: https://www.eea.europa.eu/publications/air-quality-in-europe-2019

EEA. (2020a). *Air pollution in Europe -- 2020 report*. Retrieved from European Environment Agency, Kongens Nytorv 6, 1050 Copenhagen K, Denmark: https://www.eea.europa.eu/publications/air-quality-in-europe-2020-report

EEA. (2020b). *Air quality in Europe - 2020 report*. Retrieved from

EEA. (2020c). Exceedance of air quality standards in Europe. Retrieved from https://www.eea.europa.eu/ds_resolveuid/IND-34-en

EEA. (2021a). *Air quality in Europe 2021*. Retrieved from

EEA. (2021b). Briefing no. 08/2021 -- Europe’s air quality status 2021. doi:10.2800/30241

EEA. (2022). Briefing no. 04/2022 -- Europe’s air quality status 2022. doi:10.2800/125429

Ellison, R. B., Greaves, S. P., & Hensher, D. A. (2013). Five years of London's low emission zone: Effects on vehicle fleet composition and air quality. *Transportation Research Part D: Transport and Environment, 23*, 25-33. doi:10.1016/j.trd.2013.03.010

Elminir, H. K. (2006). Relative influence of air pollutants and weather conditions on solar radiation – Part 1: Relationship of air pollutants with weather conditions. *Meteorology and Atmospheric Physics, 96*(3-4), 245-256. doi:10.1007/s00703-006-0209-4

European Commission. (2013). Environment Action Programme to 2020. Retrieved from https://ec.europa.eu/environment/action-programme/

Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, (2008).

Farrell, W. J., Weichenthal, S., Goldberg, M., & Hatzopoulou, M. (2015). Evaluating air pollution exposures across cycling infrastructure types: Implications for facility design. *Journal of Transport and Land Use*. doi:10.5198/jtlu.2015.702

Faustini, A., Rapp, R., & Forastiere, F. (2014). Nitrogen dioxide and mortality: review and meta-analysis of long-term studies. *Eur Respir J, 44*(3), 744-753. doi:10.1183/09031936.00114713

Ferreira, F., Gomes, P., Tente, H., Carvalho, A. C., Pereira, P., & Monjardino, J. (2015). Air quality improvements following implementation of Lisbon's Low Emission Zone. *Atmospheric Environment, 122*, 373-381. doi:https://doi.org/10.1016/j.atmosenv.2015.09.064

Fishbain, B., Lerner, U., Castell, N., Cole-Hunter, T., Popoola, O., Broday, D. M., . . . Bartonova, A. (2017). An evaluation tool kit of air quality micro-sensing units. *Sci Total Environ, 575*, 639-648. doi:10.1016/j.scitotenv.2016.09.061

Folgerø, I. K., Harding, T., & Westby, B. S. (2020). Going fast or going green? Evidence from environmental speed limits in Norway. *Transportation Research Part D: Transport and Environment, 82*. doi:10.1016/j.trd.2020.102261

Fowler, D., Brimblecombe, P., Burrows, J., Heal, M. R., Grennfelt, P., Stevenson, D. S., . . . Vieno, M. (2020). A chronology of global air quality. *Philos Trans A Math Phys Eng Sci, 378*(2183), 20190314. doi:10.1098/rsta.2019.0314

Fuglestvedt, J., Berntsen, T., Myhre, G., Rypdal, K., & Skeie, R. B. (2008). Climate forcing from the transport sectors. *Proc Natl Acad Sci U S A, 105*(2), 454-458. doi:10.1073/pnas.0702958104

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., . . . Yan, C. (2022). Pollution and health: a progress update. *The Lancet Planetary Health, 6*(6), e535-e547. doi:10.1016/s2542-5196(22)00090-0

Gallardo, L., Barraza, F., Ceballos, A., Galleguillos, M., Huneeus, N., Lambert, F., . . . Véliz, K. D. (2018). Evolution of air quality in Santiago: The role of mobility and lessons from the science-policy interface. *Elementa: Science of the Anthropocene, 6*. doi:10.1525/elementa.293

Gao, Y., Dong, W., Guo, K., Liu, X., Chen, Y., Liu, X., . . . Chen, C. (2016). *Mosaic: A low-cost mobile sensing system for urban air quality monitoring*. Paper presented at the International Conference on Computer Communications IEEE.

Gautam, S. (2020). COVID-19: air pollution remains low as people stay at home. *Air Qual Atmos Health*, 1-5. doi:10.1007/s11869-020-00842-6

Gehrsitz, M. (2017). The effect of low emission zones on air pollution and infant health. *Journal of Environmental Economics and Management, 83*, 121-144. doi:10.1016/j.jeem.2017.02.003

Geiß, A. (2016). *Automated calibration of ceilometer data and its applicability for quantitative aerosol monitoring.* Ludwig–Maximilians–Universität München,

Geiß, A., Wiegner, M., Bonn, B., Schäfer, K., Forkel, R., von Schneidemesser, E., . . . Nothard, R. (2017). Mixing layer height as an indicator for urban air quality?

Atmospheric Measurement Techniques, 10(8), 2969-2988. doi:10.5194/amt-10-2969-2017

Genikomsakis, K. N., Galatoulas, N. F., Dallas, P. I., Candanedo Ibarra, L. M., Margaritis, D., & Ioakimidis, C. S. (2018). Development and On-Field Testing of Low-Cost Portable System for Monitoring PM2.5 Concentrations. *Sensors (Basel), 18*(4). doi:10.3390/s18041056

Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486-489. doi:10.5812/ijem.3505

Giallouros, G., Kouis, P., Papatheodorou, S. I., Woodcock, J., & Tainio, M. (2020). The long-term impact of restricting cycling and walking during high air pollution days on all-cause mortality: Health impact Assessment study. *Environ Int, 140*, 105679. doi:10.1016/j.envint.2020.105679

Goldstein, P. (2021). Pop-up-Radweg sorgt für bessere Luft. Retrieved from https://www.morgenpost.de/bezirke/friedrichshain-kreuzberg/article232031103/Pop-up-Radweg-sorgt-fuer-bessere-Luft.html

Gonzalez Olivardia, F., Zhang, Q., Matsuo, T., Shimadera, H., & Kondo, A. (2019). Analysis of Pollutant Dispersion in a Realistic Urban Street Canyon Using Coupled CFD and Chemical Reaction Modeling. *Atmosphere, 10*(9). doi:10.3390/atmos10090479

Grundström, M., Hak, C., Chen, D., Hallquist, M., & Pleijel, H. (2015). Variation and co-variation of PM10, particle number concentration, NOx and NO2 in the urban air – Relationships with wind speed, vertical temperature gradient and weather type. *Atmospheric Environment, 120*, 317-327. doi:10.1016/j.atmosenv.2015.08.057

Gu, J., Deffner, V., Küchenhoff, H., Pickford, R., Breitner, S., Schneider, A., . . . Cyrys, J. (2022). Low emission zones reduced PM10 but not NO2 concentrations in Berlin and Munich, Germany. *Journal of Environmental Management, 302*, 114048. doi:10.1016/j.jenvman.2021.114048

Gu, J., Deffner, V., Kuchenhoff, H., Pickford, R., Breitner, S., Schneider, A., . . . Cyrys, J. (2022). Low emission zones reduced PM10 but not NO2 concentrations in Berlin and Munich, Germany. *J Environ Manage, 302*(Pt A), 114048. doi:10.1016/j.jenvman.2021.114048

Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, S., Gerland, S., . . . Vose, R. S. (2021). Changing State of the Climate System. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 287-422). Cambridge, United Kingdom and New York, NY, USA.

Gulliver, J., & Briggs, D. J. (2004). Personal exposure to particulate air pollution in transport microenvironments. *Atmospheric Environment, 38*(1), 1-8. doi:10.1016/j.atmosenv.2003.09.036

Hagan, D. H., Isaacman-VanWertz, G., Franklin, J. P., Wallace, L. M. M., Kocar, B. D., Heald, C. L., & Kroll, J. H. (2018). Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. *Atmospheric Measurement Techniques, 11*(1), 315-328. doi:10.5194/amt-11-315-2018

Hanigan, I. C., Rolfe, M. I., Knibbs, L. D., Salimi, F., Cowie, C. T., Heyworth, J., . . . Morgan, G. G. (2019). All-cause mortality and long-term exposure to low level air pollution in the '45 and up study' cohort, Sydney, Australia, 2006-2015. *Environ Int, 126*, 762-770. doi:10.1016/j.envint.2019.02.044

Harkey, M., Holloway, T., Oberman, J., & Scotty, E. (2015). An evaluation of CMAQ NO2using observed chemistry-meteorology correlations. *Journal of Geophysical Research: Atmospheres, 120*(22), 11,775-711,797. doi:10.1002/2015jd023316

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models*. London: Chapman & Hall.

HEI. (2020). *State of Global Air 2020*. Retrieved from Boston, MA, USA: https://www.stateofglobalair.org/sites/default/files/documents/2020-10/soga-2020-report-10-26_0.pdf

Herberg, J., Haas, T., Oppold, D., & von Schneidemesser, D. (2020). A Collaborative Transformation beyond Coal and Cars? Co-Creation and Corporatism in the German Energy and Mobility Transitions. *Sustainability, 12*(8). doi:10.3390/su12083278

Hernández, M. A., Ramírez, O., Benavides, J. A., & Franco, J. F. (2021). Urban cycling and air quality: Characterizing cyclist exposure to particulate-related pollution. *Urban Climate, 36*.

Hoffmann, B., Boogaard, H., de Nazelle, A., Andersen, Z. J., Abramson, M., Brauer, M., . . . Thurston, G. (2021). WHO Air Quality Guidelines 2021-Aiming for Healthier Air for all: A Joint Statement by Medical, Public Health, Scientific Societies and Patient Representative Organisations. *Int J Public Health, 66*, 1604465. doi:10.3389/ijph.2021.1604465

Hofman, J., Do, T. H., Qin, X., Bonet, E. R., Philips, W., Deligiannis, N., & La Manna, V. P. (2022). Spatiotemporal air quality inference of low-cost sensor data: Evidence from multiple sensor testbeds. *Environmental Modelling & Software, 149*. doi:10.1016/j.envsoft.2022.105306

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

Holnicki, P., Nahorski, Z., & Kałuszko, A. (2021). Impact of Vehicle Fleet Modernization on the Traffic-Originated Air Pollution in an Urban Area—A Case Study. *Atmosphere, 12*(12). doi:10.3390/atmos12121581

Host, S., Honoré, C., Joly, F., Saunal, A., Le Tertre, A., & Medina, S. (2020). Implementation of various hypothetical low emission zone scenarios in Greater Paris: Assessment of fine-scale reduction in exposure and expected health benefits. *Environmental Research, 185*, 109405. doi:10.1016/j.envres.2020.109405

Huang, S., Li, H., Wang, M., Qian, Y., Steenland, K., Caudle, W. M., . . . Shi, L. (2021). Long-term exposure to nitrogen dioxide and mortality: A systematic review and meta-analysis. *Sci Total Environ, 776*.

Huang, Y.-D., Hou, R.-W., Liu, Z.-Y., Song, Y., Cui, P.-Y., & Kim, C.-N. (2019). Effects of Wind Direction on the Airflow and Pollutant Dispersion inside a Long Street Canyon. *Aerosol and Air Quality Research, 19*(5), 1152-1171. doi:10.4209/aaqr.2018.09.0344

Huangfu, P., & Atkinson, R. (2020). Long-term exposure to NO2 and O3 and all-cause and respiratory mortality: A systematic review and meta-analysis. *Environ Int, 144*, 105998. doi:10.1016/j.envint.2020.105998

Jacob, D. J., & Winner, D. A. (2009). Effect of climate change on air quality. *Atmospheric Environment, 43*(1), 51-63. doi:10.1016/j.atmosenv.2008.09.051

Jarjour, S., Jerrett, M., Westerdahl, D., de Nazelle, A., Hanning, C., Daly, L., . . . Balmes, J. (2013). Cyclist route choice, traffic-related air pollution, and lung function: a scripted exposure study. *Environ Health, 12*(14).

Jerrett, M., Donaire-Gonzalez, D., Popoola, O., Jones, R., Cohen, R. C., Almanza, E., . . . Nieuwenhuijsen, M. (2017). Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science. *Environ Res, 158*, 286-294. doi:10.1016/j.envres.2017.04.023

Jiang, W., Boltze, M., Groer, S., & Scheuvens, D. (2017). Impacts of low emission zones in Germany on air pollution levels. *Transportation Research Procedia, 25*, 3370-3382. doi:10.1016/j.trpro.2017.05.217

Jones, A. M., Harrison, R. M., Barratt, B., & Fuller, G. (2012). A large reduction in airborne particle number concentrations at the time of the introduction of “sulphur free” diesel and the London Low Emission Zone. *Atmospheric Environment, 50*, 129-138. doi:10.1016/j.atmosenv.2011.12.050

Kallenbach, T. (2020). Narratives of urban mobility in Germany: on the threshold of a departure from the car-centered city? *Sustainability: Science, Practice and Policy, 16*(1), 197-207. doi:10.1080/15487733.2020.1799625

Karagulian, F., Barbiere, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., . . . Borowiak, A. (2019). Review of the Performance of Low-Cost Sensors for Air Quality Monitoring. *Atmosphere, 10*(9). doi:10.3390/atmos10090506

Kazemiparkouhi, F., Eum, K. D., Wang, B., Manjourides, J., & Suh, H. H. (2020). Long-term ozone exposures and cause-specific mortality in a US Medicare cohort. *J Expo Sci Environ Epidemiol, 30*(4), 650-658. doi:10.1038/s41370-019-0135-4

Kelly, F. J., & Zhu, T. (2016). Transport solutions for cleaner air. *Science, 352*(6288), 934-936. doi:10.1126/science.aaf3420

Keuken, M. P., Jonkers, S., Zandveld, P., Voogt, M., & Elshout van den, S. (2012). Elemental carbon as an indicator for evaluating the impact of traffic measures on air quality and health. *Atmospheric Environment, 61*, 1-8. doi:10.1016/j.atmosenv.2012.07.009

Kim, J., Shusterman, A. A., Lieschke, K. J., Newman, C., & Cohen, R. C. (2018). The Berkeley Atmospheric CO2 Observation Network: field calibration and evaluation of low-cost air quality sensors. *Atmospheric Measurement Techniques, 11*(4), 1937-1946. doi:10.5194/amt-11-1937-2018

Kizel, F., Etzion, Y., Shafran-Nathan, R., Levy, I., Fishbain, B., Bartonova, A., & Broday, D. M. (2018). Node-to-node field calibration of wireless distributed air pollution sensor network. *Environ Pollut, 233*, 900-909. doi:10.1016/j.envpol.2017.09.042

Kraus, S., & Koch, N. (2021). Provisional COVID-19 infrastructure induces large, rapid increases in cycling. *Proc Natl Acad Sci U S A, 118*(15). doi:10.1073/pnas.2024399118

Krecl, P., Cipoli, Y. A., Targino, A. C., Castro, L. B., Gidhagen, L., Malucelli, F., & Wolf, A. (2020). Cyclists' exposure to air pollution under different traffic management strategies. *Sci Total Environ, 723*, 138043. doi:10.1016/j.scitotenv.2020.138043

Kukkonen, J., Valkonen, E., Walden, J., Koskentalo, T., Aarnio, P., Karppinen, A., . . . Kartastenpää, R. (2001). A measurement campaign in a street canyon in Helsinki and comparison of results with predictions of the OSPM model. *Atmospheric Environment, 35*(2), 231-243. doi:10.1016/s1352-2310(00)00200-4

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., . . . Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environ Int, 75*, 199-205. doi:10.1016/j.envint.2014.11.019

Kunst, F. (2018). Das Berliner Mobilitätsgesetz: Ein entscheidender Beitrag zur "Verkehrswende" in der Hauptstadt? *Planerin, 4*, 47-49.

Kwak, K. H., Lee, S. H., Seo, J. M., Park, S. B., & Baik, J. J. (2016). Relationship between rooftop and on-road concentrations of traffic-related pollutants in a busy street canyon: Ambient wind effects. *Environ Pollut, 208*(Pt A), 185-197. doi:10.1016/j.envpol.2015.07.030

Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Basu, N., . . . Zhong, M. (2018). The Lancet Commission on pollution and health. *The Lancet, 391*(10119), 462-512. doi:https://doi.org/10.1016/S0140-6736(17)32345-0

Languille, B., Gros, V., Bonnaire, N., Pommier, C., Honore, C., Debert, C., . . . Zeitouni, K. (2020). A methodology for the characterization of portable sensors for air quality measure with the goal of deployment in citizen science. *Sci Total Environ, 708*, 134698. doi:10.1016/j.scitotenv.2019.134698

Lawrence, M. G., Williams, S., Nanz, P., & Renn, O. (2022). Characteristics, potentials, and challenges of transdisciplinary research. *One Earth, 5*(1), 44-61. doi:10.1016/j.oneear.2021.12.010

Laxen, D. P. H., & Noordally, E. (1987). Nitrogen dioxide distribution in street canyons. *Atmospheric Environment (1967), 21*(9), 1899-1903. doi:10.1016/0004-6981(87)90150-8

Lebrusán, I., & Toutouh, J. (2020). Car restriction policies for better urban health: a low emission zone in Madrid, Spain. *Air Quality, Atmosphere & Health, 14*(3), 333-342. doi:10.1007/s11869-020-00938-z

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature, 525*(7569), 367-371. doi:10.1038/nature15371

Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R. T., Haines, A., & Ramanathan, V. (2019). Effects of fossil fuel and total anthropogenic emission removal on public health and climate. *Proceedings of the National Academy of Sciences, 116*(15), 7192-7197. doi:10.1073/pnas.1819989116

Lelieveld, J., Klingmuller, K., Pozzer, A., Poschl, U., Fnais, M., Daiber, A., & Munzel, T. (2019). Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. *Eur Heart J, 40*(20), 1590-1596. doi:10.1093/eurheartj/ehz135

Lenschow, P., Abraham, H. J., Kutzner, K., Lutz, M., Preuß, J. D., & Reichenbächer, W. (2001). Some ideas about the sources of PM10. *Atmospheric Environment, 35*(Supplement 1), 23-33. doi:10.1016/S1352-2310(01)00122-4

Lewis, A., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., . . . White, A. (2016). Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday Discuss, 189*, 85-103. doi:10.1039/c5fd00201j

Lewis, A., von Schneidemesser, E., & Peltier, R. (2018). *Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications*. Retrieved from Geneva, Switzerland:

Lim, C. C., Hayes, R. B., Ahn, J., Shao, Y., Silverman, D. T., Jones, R. R., . . . Thurston, G. D. (2019). Long-Term Exposure to Ozone and Cause-Specific Mortality Risk in the United States. *Am J Respir Crit Care Med, 200*(8), 1022-1031. doi:10.1164/rccm.201806-1161OC

Lim, C. C., Kim, H., Vilcassim, M. J. R., Thurston, G. D., Gordon, T., Chen, L. C., . . . Kim, S. Y. (2019). Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environ Int, 131*, 105022. doi:10.1016/j.envint.2019.105022

Lin, C., Masey, N., Wu, H., Jackson, M., Carruthers, D., Reis, S., . . . Heal, M. (2017a). Practical Field Calibration of Portable Monitors for Mobile Measurements of Multiple Air Pollutants. *Atmosphere, 8*(12). doi:10.3390/atmos8120231

Lin, C., Masey, N., Wu, H., Jackson, M., Carruthers, D. J., Reis, S., . . . Heal, M. R. (2017b). Practical Field Calibration of Portable Monitors for Mobile Measurements of Multiple Air Pollutants. *Atmosphere, 8*(12). doi:10.3390/atmos8120231

Liu, C., Chen, R., Sera, F., Vicedo-Cabrera, A. M., Guo, Y., Tong, S., . . . Kan, H. (2019). Ambient Particulate Air Pollution and Daily Mortality in 652 Cities. *New England Journal of Medicine, 381*(8), 705-715. doi:10.1056/NEJMoa1817364

López, I., Ortega, J., & Pardo, M. (2020). Mobility Infrastructures in Cities and Climate Change: An Analysis Through the Superblocks in Barcelona. *Atmosphere, 11*(4). doi:10.3390/atmos11040410

Lurkin, V., Hambuckers, J., & van Woensel, T. (2021). Urban low emissions zones: A behavioral operations management perspective. *Transportation Research Part A: Policy and Practice, 144*, 222-240. doi:10.1016/j.tra.2020.11.015

Magueta, D., Madaleno, M., Ferreira Dias, M., & Meireles, M. (2018). New cars and emissions: Effects of policies, macroeconomic impacts and cities characteristics in Portugal. *Journal of Cleaner Production, 181*, 178-191. doi:10.1016/j.jclepro.2017.11.243

Mahajan, S., & Kumar, P. (2020). Evaluation of low-cost sensors for quantitative personal exposure monitoring. *Sustainable Cities and Society*(57).

Mahajan, S., Luo, C.-H., Wu, D.-Y., & Chen, L.-J. (2021). From Do-It-Yourself (DIY) to Do-It-Together (DIT): Reflections on designing a citizen-driven air quality monitoring framework in Taiwan. *Sustainable Cities and Society, 66*. doi:10.1016/j.scs.2020.102628

Malina, C., & Scheffler, F. (2015a). The impact of Low Emission Zones on particulate matter concentration and public health. *Transportation Research Part A: Policy and Practice, 77*, 372-385. doi:10.1016/j.tra.2015.04.029

Malina, C., & Scheffler, F. (2015b). The impact of Low Emission Zones on particulate matter concentration and public health: A Rejoinder. *Transportation Research Part A: Policy and Practice, 82*, 257-258. doi:10.1016/j.tra.2015.10.001

Malings, C., Tanzer, R., Hauryliuk, A., Kumar, S. P. N., Zimmerman, N., Kara, L. B., & Presto, A. A. (2019). Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. *Atmospheric Measurement Techniques, 12*(2), 903-920. doi:10.5194/amt-12-903-2019

Margaryan, S. (2021). Low emission zones and population health. *J Health Econ, 76*, 102402. doi:10.1016/j.jhealeco.2020.102402

Masiol, M., Squizzato, S., Chalupa, D., Rich, D. Q., & Hopke, P. K. (2018). Evaluation and Field Calibration of a Low-cost Ozone Monitor at a Regulatory Urban Monitoring Station. *Aerosol and Air Quality Research, 18*(8), 2029-2037. doi:10.4209/aaqr.2018.02.0056

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., . . . Martin, R. V. (2020). A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS). *Earth System Science Data, 12*(4), 3413-3442. doi:10.5194/essd-12-3413-2020

McKercher, G. R., Salmond, J. A., & Vanos, J. K. (2017). Characteristics and applications of small, portable gaseous air pollution monitors. *Environ Pollut, 223*, 102-110. doi:10.1016/j.envpol.2016.12.045

Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., . . . Jones, R. L. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment, 70*, 186-203. doi:10.1016/j.atmosenv.2012.11.060

Melkonyan, A., & Kuttler, W. (2012). Long-term analysis of NO, NO2 and O3 concentrations in North Rhine-Westphalia, Germany. *Atmospheric Environment, 60*, 316-326. doi:10.1016/j.atmosenv.2012.06.048

Mijling, B., Jiang, Q., de Jonge, D., & Bocconi, S. (2018). Field calibration of electrochemical NO2 sensors in a citizen science context. *Atmospheric Measurement Techniques, 11*(3), 1297-1312. doi:10.5194/amt-11-1297-2018

Mills, I. C., Atkinson, R. W., Anderson, H. R., Maynard, R. L., & Strachan, D. P. (2016). Distinguishing the associations between daily mortality and hospital admissions and nitrogen dioxide from those of particulate matter: a systematic review and meta-analysis. *BMJ Open, 6*(7), e010751. doi:10.1136/bmjopen-2015-010751

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

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., . . . Williams, M. L. (2015). Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics, 15*(15), 8889-8973. doi:10.5194/acp-15-8889-2015

Monks, P. S., Granier, C., Fuzzi, S., Stohl, A., Williams, M. L., Akimoto, H., . . . von Glasow, R. (2009). Atmospheric composition change – global and regional air quality. *Atmospheric Environment, 43*(33), 5268-5350. doi:10.1016/j.atmosenv.2009.08.021

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

Morfeld, P., Groneberg, D. A., & Spallek, M. (2015a). The impact of Low Emission Zones on particulate matter concentration and public health: A Comment. *Transportation Research Part A: Policy and Practice, 82*, 255-256. doi:10.1016/j.tra.2015.10.002

Morfeld, P., Groneberg, D. A., & Spallek, M. (2015b). Letter to the Editor: On the effectiveness of low emission zones. *Atmospheric Environment, 122*, 569-570. doi:10.1016/j.atmosenv.2015.10.029

Morton, C., Mattioli, G., & Anable, J. (2021). Public acceptability towards Low Emission Zones: The role of attitudes, norms, emotions, and trust. *Transportation Research Part A: Policy and Practice, 150*, 256-270. doi:10.1016/j.tra.2021.06.007

Mudway, I. S., Dundas, I., Wood, H. E., Marlin, N., Jamaludin, J. B., Bremner, S. A., . . . Griffiths, C. J. (2019). Impact of London's low emission zone on air quality and children's respiratory health: a sequential annual cross-sectional study. *The Lancet Public Health, 4*(1), e28-e40. doi:10.1016/S2468-2667(18)30202-0

Mueller, N., Rojas-Rueda, D., Salmon, M., Martinez, D., Ambros, A., Brand, C., . . . Nieuwenhuijsen, M. (2018). Health impact assessment of cycling network expansions in European cities. *Preventive Medicine, 109*, 62-70. doi:10.1016/j.ypmed.2017.12.011

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., . . . Leigh, R. R. (2015). Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology, 35*(11), 3185-3203. doi:10.1002/joc.4210

Nieuwenhuijsen, M. J. (2020). Urban and transport planning pathways to carbon neutral, liveable and healthy cities; A review of the current evidence. *Environ Int, 140*, 105661. doi:10.1016/j.envint.2020.105661

Ohlwein, S., Kappeler, R., Kutlar Joss, M., Kunzli, N., & Hoffmann, B. (2019). Health effects of ultrafine particles: a systematic literature review update of epidemiological evidence. *Int J Public Health, 64*(4), 547-559. doi:10.1007/s00038-019-01202-7

Okafor, N. U., Alghorani, Y., & Delaney, D. T. (2020). Improving Data Quality of Low-cost IoT Sensors in Environmental Monitoring Networks Using Data Fusion and Machine Learning Approach. *ICT Express, 6*(3), 220-228. doi:10.1016/j.icte.2020.06.004

Oliveira, C., Pio, C., Caseiro, A., Santos, P., Nunes, T., Mao, H., . . . Sokhi, R. (2010). Road traffic impact on urban atmospheric aerosol loading at Oporto, Portugal. *Atmospheric Environment, 44*(26), 3147-3158. doi:10.1016/j.atmosenv.2010.05.027

Ordóñez, C., Mathis, H., Furger, M., Henne, S., Hüglin, C., Staehelin, J., & Prévôt, A. S. H. (2005). Changes of daily surface ozone maxima in Switzerland in all seasons from 1992 to 2002 and discussion of summer 2003. *Atmospheric Chemistry and Physics, 5*(5), 1187-1203. doi:10.5194/acp-5-1187-2005

Orellano, P., Reynoso, J., Quaranta, N., Bardach, A., & Ciapponi, A. (2020). Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis. *Environ Int, 142*, 105876. doi:10.1016/j.envint.2020.105876

Otero, N., Sillmann, J., Mar, K. A., Rust, H. W., Solberg, S., Andersson, C., . . . Butler, T. (2018). A multi-model comparison of meteorological drivers of surface ozone over Europe. *Atmospheric Chemistry and Physics, 18*(16), 12269-12288. doi:10.5194/acp-18-12269-2018

Otero, N., Sillmann, J., Schnell, J. L., Rust, H. W., & Butler, T. (2016). Synoptic and meteorological drivers of extreme ozone concentrations over Europe. *Environmental Research Letters, 11*(2). doi:10.1088/1748-9326/11/2/024005

Panteliadis, P., Strak, M., Hoek, G., Weijers, E., van der Zee, S., & Dijkema, M. (2014). Implementation of a low emission zone and evaluation of effects on air quality by long-term monitoring. *Atmospheric Environment, 86*, 113-119. doi:10.1016/j.atmosenv.2013.12.035

Park, S.-B., Kwak, K.-H., Han, B.-S., Ganbat, G., Lee, H., Seo, J. M., . . . Baik, J.-J. (2015). Measurements of Turbulent Flow and Ozone at Rooftop and Sidewalk Sites in a High-Rise Building Area. *Sola, 11*(0), 1-4. doi:10.2151/sola.2015-001

Park, S.-J., Kim, J.-J., Choi, W., Kim, E.-R., Song, C.-K., & Pardyjak, E. R. (2019). Flow Characteristics Around Step-Up Street Canyons with Various Building Aspect Ratios. *Boundary-Layer Meteorology, 174*(3), 411-431. doi:10.1007/s10546-019-00494-9

Park, S.-J., Kim, J.-J., Kim, M. J., Park, R. J., & Cheong, H.-B. (2015). Characteristics of flow and reactive pollutant dispersion in urban street canyons. *Atmospheric Environment, 108*, 20-31. doi:10.1016/j.atmosenv.2015.02.065

Paull, N. J., Krix, D., Torpy, F. R., & Irga, P. J. (2020). Can Green Walls Reduce Outdoor Ambient Particulate Matter, Noise Pollution and Temperature? *Int J Environ Res Public Health, 17*(14). doi:10.3390/ijerph17145084

Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J., & Tapper, N. J. (2011). Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmospheric Environment, 45*(6), 1328-1336. doi:10.1016/j.atmosenv.2010.11.051

Peltier, R. (2020). *An update on low-cost sensors for the measurement of atmospheric composition*. Retrieved from

Perez, I. A., Garcia, M. A., Sanchez, M. L., Pardo, N., & Fernandez-Duque, B. (2020). Key Points in Air Pollution Meteorology. *Int J Environ Res Public Health, 17*(22). doi:10.3390/ijerph17228349

Pestel, N., & Wozny, F. (2021). Health effects of Low Emission Zones: Evidence from German hospitals. *Journal of Environmental Economics and Management, 109*. doi:10.1016/j.jeem.2021.102512

Peters, J. F., Burguillo, M., & Arranz, J. M. (2021). Low emission zones: Effects on alternative-fuel vehicle uptake and fleet CO2 emissions. *Transportation Research Part D: Transport and Environment, 95*. doi:10.1016/j.trd.2021.102882

Peterson, P. J. D., Aujla, A., Grant, K. H., Brundle, A. G., Thompson, M. R., Vande Hey, J., & Leigh, R. J. (2017). Practical Use of Metal Oxide Semiconductor Gas Sensors for Measuring Nitrogen Dioxide and Ozone in Urban Environments. *Sensors (Basel), 17*(7). doi:10.3390/s17071653

Petetin, H., Bowdalo, D., Soret, A., Guevara, M., Jorba, O., Serradell, K., & Pérez García-Pando, C. (2020). Meteorology-normalized impact of the COVID-19 lockdown upon NO₂ pollution in Spain. *Atmospheric Chemistry and Physics, 20*(18), 11119-11141. doi:10.5194/acp-20-11119-2020

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., . . . Shang, L. (2014). The next generation of low-cost personal air quality sensors for quantitative

exposure monitoring. *Atmospheric Measurement Techniques, 7*(10), 3325-3336. doi:10.5194/amt-7-3325-2014

Pisoni, E., Christidis, P., Thunis, P., & Trombetti, M. (2019). Evaluating the impact of "Sustainable Urban Mobility Plans" on urban background air quality. *J Environ Manage, 231*, 249-255. doi:10.1016/j.jenvman.2018.10.039

Popoola, O. A. M., Carruthers, D., Lad, C., Bright, V. B., Mead, M. I., Stettler, M. E. J., . . . Jones, R. L. (2018). Use of networks of low cost air quality sensors to quantify air quality in urban settings. *Atmospheric Environment, 194*, 58-70. doi:10.1016/j.atmosenv.2018.09.030

Popoola, O. A. M., Stewart, G. B., Mead, M. I., & 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. doi:10.1016/j.atmosenv.2016.10.024

Poulhès, A., & Proulhac, L. (2021). The Paris Region low emission zone, a benefit shared with residents outside the zone. *Transportation Research Part D: Transport and Environment, 98*. doi:10.1016/j.trd.2021.102977

Pozzer, A., Dominici, F., Haines, A., Witt, C., Münzel, T., & Lelieveld, J. (2020). Regional and global contributions of air pollution to risk of death from COVID-19. *Cardiovascular Research, 116*(14), 2247-2253. doi:10.1093/cvr/cvaa288

Pusede, S. E., Gentner, D. R., Wooldridge, P. J., Browne, E. C., Rollins, A. W., Min, K. E., . . . Cohen, R. C. (2014). On the temperature dependence of organic reactivity, nitrogen oxides, ozone production, and the impact of emission controls in San Joaquin Valley, California. *Atmospheric Chemistry and Physics, 14*(7), 3373-3395. doi:10.5194/acp-14-3373-2014

R Core Team. (2019). R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing.

R Core Team. (2022). R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Ragettli, M. S., Corradi, E., Braun-Fahrländer, C., Schindler, C., de Nazelle, A., Jerrett, M., . . . Phuleria, H. C. (2013). Commuter exposure to ultrafine particles in different urban locations, transportation modes and routes. *Atmospheric Environment, 77*, 376-384. doi:10.1016/j.atmosenv.2013.05.003

Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., . . . Rickerby, D. (2017). End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Sci Total Environ, 607-608*, 691-705. doi:10.1016/j.scitotenv.2017.06.266

Rank, J., Folke, J., & Jespersen, P. H. (2001). Differences in cyclists and car drivers exposure to air pollution from traffic in the city of Copenhagen. *Sci Total Environ, 279*, 131-136.

Raza, W., Forsberg, B., Johansson, C., & Sommar, J. N. (2018). Air pollution as a risk factor in health impact assessments of a travel mode shift towards cycling. *Glob Health Action, 11*(1), 1429081. doi:10.1080/16549716.2018.1429081

Ripoll, A., Viana, M., Padrosa, M., Querol, X., Minutolo, A., Hou, K. M., . . . Garcia-Vidal, J. (2019). Testing the performance of sensors for ozone pollution monitoring in a citizen science approach. *Sci Total Environ, 651*(Pt 1), 1166-1179. doi:10.1016/j.scitotenv.2018.09.257

Rode, P. (2019). Urban planning and transport policy integration: The role of governance hierarchies and networks in London and Berlin. *Journal of Urban Affairs, 41*(1), 39-63. doi:10.1080/07352166.2016.1271663

Rodriguez-Rey, D., Guevara, M., Linares, M. P., Casanovas, J., Armengol, J. M., Benavides, J., . . . Garcia-Pando, C. P. (2022). To what extent the traffic restriction policies applied in Barcelona city can improve its air quality? *Sci Total Environ, 807*(Pt 2), 150743. doi:10.1016/j.scitotenv.2021.150743

Romer, P. S., Duffey, K. C., Wooldridge, P. J., Edgerton, E., Baumann, K., Feiner, P. A., . . . Cohen, R. C. (2018). Effects of temperature-dependent NOx emissions on continental ozone production. *Atmospheric Chemistry and Physics, 18*(4), 2601-2614. doi:10.5194/acp-18-2601-2018

Roustan, Y., Pausader, M., & Seigneur, C. (2011). Estimating the effect of on-road vehicle emission controls on future air quality in Paris, France. *Atmospheric Environment, 45*(37), 6828-6836. doi:10.1016/j.atmosenv.2010.10.010

Russell, H. S., Frederickson, L. B., Kwiatkowski, S., Emygdio, A. P. M., Kumar, P., Schmidt, J. A., . . . Johnson, M. S. (2022). Enhanced Ambient Sensing Environment—A New Method for Calibrating Low-Cost Gas Sensors. *Sensors, 22*(19). doi:10.3390/s22197238

Salas, R., Perez-Villadoniga, M. J., Prieto-Rodriguez, J., & Russo, A. (2021). Were traffic restrictions in Madrid effective at reducing NO2 levels? *Transportation Research Part D: Transport and Environment, 91*. doi:10.1016/j.trd.2020.102689

Samad, A., & Vogt, U. (2021). Mobile air quality measurements using bicycle to obtain spatial distribution and high temporal resolution in and around the city center of Stuttgart. *Atmospheric Environment, 244*. doi:10.1016/j.atmosenv.2020.117915

Sánchez, J. M., Ortega, E., López-Lambas, M. E., & Martín, B. (2021). Evaluation of emissions in traffic reduction and pedestrianization scenarios in Madrid. *Transportation Research Part D: Transport and Environment, 100*, 103064. doi:10.1016/j.trd.2021.103064

Santos, F. M., Gómez-Losada, Á., & Pires, J. C. M. (2019). Impact of the implementation of Lisbon low emission zone on air quality. *Journal of Hazardous Materials, 365*, 632-641. doi:10.1016/j.jhazmat.2018.11.061

Schepers, P., Fishman, E., Beelen, R., Heinen, E., Wijnen, W., & Parkin, J. (2015). The mortality impact of bicycle paths and lanes related to physical activity, air pollution exposure, and road safety. *Journal of Transport & Health, 2*, 460-473.

Scherer, D., Ament, F., Emeis, S., Fehrenbach, U., Leitl, B., Scherber, K., . . . Vogt, U. (2019). Three-Dimensional Observation of Atmospheric Processes in Cities. *Meteorologische Zeitschrift, 28*(2), 121-138. doi:10.1127/metz/2019/0911

Scherer, D., Antretter, F., Bender, S., Cortekar, J., Emeis, S., Fehrenbach, U., . . . Scherber, K. (2019). Urban Climate Under Change [UC]2 – A National Research Programme for Developing a Building-Resolving Atmospheric Model for Entire City Regions. *Meteorologische Zeitschrift, 28*(2), 95-104. doi:10.1127/metz/2019/0913

Schmitz, S., Caseiro, A., Kerschbaumer, A., & von Schneidemesser, E. (2021). Do new bike lanes impact air pollution exposure for cyclists?—a case study from Berlin. *Environmental Research Letters, 16*(8). doi:10.1088/1748-9326/ac1379

Schmitz, S., Towers, S., Villena, G., Caseiro, A., Wegener, R., Klemp, D., . . . von Schneidemesser, E. (2021a). Unraveling a black box: An open-source methodology for the field calibration of small air quality sensors. *Atmos. Meas. Tech. Discuss., 2021*, 1-34. doi:10.5194/amt-2020-489

Schmitz, S., Towers, S., Villena, G., Caseiro, A., Wegener, R., Klemp, D., . . . von Schneidemesser, E. (2021b). Unravelling a black box: an open-source methodology for the field calibration of small air quality sensors. *Atmospheric Measurement Techniques, 14*(11), 7221-7241. doi:10.5194/amt-14-7221-2021

Schneider, P., Bartonova, A., Castell, N., Dauge, F. R., Gerboles, M., Hagler, G. S. W., . . . Williams, R. W. (2019). Toward a Unified Terminology of Processing Levels for Low-Cost Air-Quality Sensors. *Environ Sci Technol, 53*(15), 8485-8487. doi:10.1021/acs.est.9b03950

Seinfeld, J., & Pandis, S. (2006). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (2nd Edition ed.). Hoboken, New Jersey: John Wiley Sons, Inc.

SenUMVK. (2019). *Straßenverkehrszählung Berlin*. Retrieved from

SenUMVK. (2021a). Flaniermeile Friedrichstraße. Retrieved from https://www.berlin.de/friedrichstrasse/

SenUMVK. (2021b). Friedrichstraße soll dauerhaft autofrei bleiben. Retrieved from https://www.berlin.de/sen/uvk/presse/pressemitteilungen/2021/pressemitteilung.1135889.php

SenUMVK. (2022a). *Flaniermeile Friedrichstraße: Abschlussbericht April 2022*. Retrieved from Berlin:

SenUMVK. (2022b). Langjährige Entwicklung der Luftqualität 2021. Retrieved from https://www.berlin.de/umweltatlas/luft/entwicklung-der-luftqualitaet/seit-1989/abbildungen-tabellen/

Settey, T., Gnap, J., & Beňová, D. (2019). Examining the impact of the deployment of low emission zones in Europe on the technological readiness of road freight transport. *Transportation Research Procedia, 40*, 481-488. doi:10.1016/j.trpro.2019.07.070

Sfendonis, N., Basbas, S., Mintsis, G., Taxiltaris, C., & Politis, I. (2017). Investigation of the user's acceptance concerning a Low Emission Zone in the center of Thessaloniki, Greece. *Transportation Research Procedia, 24*, 280-287. doi:https://doi.org/10.1016/j.trpro.2017.05.119

Shindell, D., Faluvegi, G., Walsh, M., Anenberg, S. C., Van Dingenen, R., Muller, N. Z., . . . Milly, G. (2011). Climate, health, agricultural and economic impacts of tighter vehicle-emission standards. *Nature Climate Change, 1*(1), 59-66. doi:10.1038/nclimate1066

Shrestha, A., Mullins, B., Zhao, Y., Selvey, L. A., & Rumchev, K. (2020). Exposure to air pollutants among cyclists: a comparison of different cycling routes in Perth, Western Australia. *Air Quality, Atmosphere & Health, 13*(9), 1023-1034. doi:10.1007/s11869-020-00850-6

Sicard, P., Agathokleous, E., De Marco, A., Paoletti, E., & Calatayud, V. (2021). Urban population exposure to air pollution in Europe over the last decades. *Environmental Sciences Europe, 33*(1), 28. doi:10.1186/s12302

Sillman, S. (1999). The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments. *Atmospheric Environment, 33*(12), 1821-1845. doi:10.1016/s1352-2310(98)00345-8

Skirienė, A. F., & Stasiškienė, Ž. (2021). COVID-19 and Air Pollution: Measuring Pandemic Impact to Air Quality in Five European Countries. *Atmosphere, 12*(3). doi:10.3390/atmos12030290

Smith, K. R., Edwards, P. M., Evans, M. J., Lee, J. D., Shaw, M. D., Squires, F., . . . Lewis, A. C. (2017). Clustering approaches to improve the performance of low cost air pollution sensors. *Faraday Discuss, 200*, 621-637. doi:10.1039/c7fd00020k

Smith, K. R., Edwards, P. M., Ivatt, P. D., Lee, J. D., Squires, F., Dai, C., . . . Lewis, A. C. (2019). An improved low-power measurement of ambient NO2 and O3 combining electrochemical sensor clusters and machine learning. *Atmospheric Measurement Techniques, 12*(2), 1325-1336. doi:10.5194/amt-12-1325-2019

Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S., . . . Preuss, P. W. (2013). The changing paradigm of air pollution monitoring. *Environ Sci Technol, 47*(20), 11369-11377. doi:10.1021/es4022602

Sorensen, M., Wendelboe Nielsen, O., Sajadieh, A., Ketzel, M., Tjonneland, A., Overvad, K., & Raaschou-Nielsen, O. (2017). Long-Term Exposure to Road Traffic Noise and Nitrogen Dioxide and Risk of Heart Failure: A Cohort Study. *Environ Health Perspect, 125*(9), 097021. doi:10.1289/EHP1272

Sousa Santos, G., Sundvor, I., Vogt, M., Grythe, H., Haug, T. W., Høiskar, B. A., & Tarrason, L. (2020). Evaluation of traffic control measures in Oslo region and its effect on current air quality policies in Norway. *Transport Policy, 99*, 251-261. doi:10.1016/j.tranpol.2020.08.025

Spandonidis, C., Tsantilas, S., Sedikos, E., Galiatsatos, N., Giannopoulos, F., Papadopoulos, P., . . . Giordamlis, C. (2020). A compact, modular and low-cost Internet of Things (IoT) platform for air quality monitoring in urban areas. *Journal of Physics: Conference Series, 1710*(1). doi:10.1088/1742-6596/1710/1/012004

Spinelle, L., Gerboles, M., Kok, G., Persijn, S., & Sauerwald, T. (2017). Review of Portable and Low-Cost Sensors for the Ambient Air Monitoring of Benzene and Other Volatile Organic Compounds. *Sensors (Basel), 17*(7). doi:10.3390/s17071520

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2015). Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical, 215*, 249-257. doi:10.1016/j.snb.2015.03.031

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2017). Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2. *Sensors and Actuators B: Chemical, 238*, 706-715. doi:10.1016/j.snb.2016.07.036

Stafoggia, M., Oftedal, B., Chen, J., Rodopoulou, S., Renzi, M., Atkinson, R. W., . . . Janssen, N. A. H. (2022). Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: results from seven large European cohorts within the ELAPSE project. *The Lancet Planetary Health, 6*(1), e9-e18. doi:10.1016/s2542-5196(21)00277-1

Stanaway, J. D., Afshin, A., Gakidou, E., Lim, S. S., Abate, D., Abate, K. H., . . . Murray, C. J. L. (2018). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet, 392*(10159), 1923-1994. doi:10.1016/S0140-6736(18)32225-6

Steinberga, I., Sustere, L., Bikse, J., Jr, J. B., & Kleperis, J. (2019). Traffic induced air pollution modeling: scenario analysis for air quality management in street canyon. *Procedia Computer Science, 149*, 384-389. doi:10.1016/j.procs.2019.01.152

Stevenson, D. S., Young, P. J., Naik, V., Lamarque, J. F., Shindell, D. T., Voulgarakis, A., . . . Archibald, A. (2013). Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP). *Atmospheric Chemistry and Physics, 13*(6), 3063-3085. doi:10.5194/acp-13-3063-2013

Szopa, S., V. Naik, B. Adhikary, P. Artaxo, T. Berntsen, W.D. Collins, . . . Zanis, P. (2021). Short-Lived Climate Forcers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Tagesspiegel. (2022). Kottbusser Damm wird clean: Pop-up-Radwege verbessern Luftqualität. Retrieved from https://checkpoint.tagesspiegel.de/telegramm/4OA2jzs4zoxUETZADukc1w

Tainio, M., Jovanovic Andersen, Z., Nieuwenhuijsen, M. J., Hu, L., de Nazelle, A., An, R., . . . Sa, T. H. (2021). Air pollution, physical activity and health: A mapping review of the evidence. *Environ Int, 147*, 105954. doi:10.1016/j.envint.2020.105954

Tang, J., McNabola, A., & Misstear, B. (2020). The potential impacts of different traffic management strategies on air pollution and public health for a more sustainable city: A modelling case study from Dublin, Ireland. *Sustainable Cities and Society, 60*. doi:10.1016/j.scs.2020.102229

Tanzer, R., Malings, C., Hauryliuk, A., Subramanian, R., & Presto, A. A. (2019). Demonstration of a Low-Cost Multi-Pollutant Network to Quantify Intra-Urban Spatial Variations in Air Pollutant Source Impacts and to Evaluate Environmental Justice. *Int J Environ Res Public Health, 16*(14). doi:10.3390/ijerph16142523

Tartakovsky, D., Kordova – Biezuner, L., Berlin, E., & Broday, D. M. (2020). Air quality impacts of the low emission zone policy in Haifa. *Atmospheric Environment, 232*, 117472. doi:10.1016/j.atmosenv.2020.117472

Thompson, M. L., Reynolds, J., Cox, L. H., Guttorp, P., & Sampson, P. D. (2001). A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Environment, 35*(3), 617-630. doi:10.1016/s1352-2310(00)00261-2

Thomsen, J., & Lühmann, S. (2021). Friedrichstraße soll dauerhaft autofrei bleiben [Press release]. Retrieved from https://www.berlin.de/sen/uvk/presse/pressemitteilungen/2021/pressemitteilung.1135889.php

Thomsen, J., & Lühmann, S. (2022). Autofreie Friedrichstraße: Der Fußverkehr erhält den Vorrang [Press release]. Retrieved from https://www.berlin.de/sen/uvk/presse/pressemitteilungen/2022/pressemitteilung.1202275.php

Tretvik, T., Nordtømme, M. E., Bjerkan, K. Y., & Kummeneje, A.-M. (2014). Can low emission zones be managed more dynamically and effectively? *Research in Transportation Business & Management, 12*, 3-10. doi:10.1016/j.rtbm.2014.08.002

Umweltbundesamt. (2022). Nationale Trendtabellen für die deutsche Berichterstattung atmosphärischer Emissionen seit 1990, Emissionsentwicklung 1990 bis 2020. Retrieved from https://www.umweltbundesamt.de/sites/default/files/medien/384/bilder/dateien/3_tab_emi-ausgew-luftschadst_2022.pdf

Vardoulakis, S., Fisher, B. E. A., Pericleous, K., & Gonzalez-Flesca, N. (2003). Modelling air quality in street canyons: a review. *Atmospheric Environment, 37*(2), 155-182. doi:10.1016/s1352-2310(02)00857-9

Venkatram, A., Snyder, M., Isakov, V., & Kimbrough, S. (2013). Impact of wind direction on near-road pollutant concentrations. *Atmospheric Environment, 80*, 248-258. doi:10.1016/j.atmosenv.2013.07.073

Vohra, K., Vodonos, A., Schwartz, J., Marais, E. A., Sulprizio, M. P., & Mickley, L. J. (2021). Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem. *Environ Res, 195*, 110754. doi:10.1016/j.envres.2021.110754

Voiculescu, M., Constantin, D. E., Condurache-Bota, S., Calmuc, V., Rosu, A., & Dragomir Balanica, C. M. (2020). Role of Meteorological Parameters in the Diurnal and Seasonal Variation of NO2 in a Romanian Urban Environment. *Int J Environ Res Public Health, 17*(17). doi:10.3390/ijerph17176228

von Schneidemesser, D., Herberg, J., & Stasiak, D. (2020). Re-claiming the responsivity gap: The co-creation of cycling policies in Berlin's mobility law. *Transportation Research Interdisciplinary Perspectives, 8*. doi:10.1016/j.trip.2020.100270

von Schneidemesser, E., Monks, P. S., Allan, J. D., Bruhwiler, L., Forster, P., Fowler, D., . . . Sutton, M. A. (2015). Chemistry and the Linkages between Air Quality and Climate Change. *Chem Rev, 115*(10), 3856-3897. doi:10.1021/acs.chemrev.5b00089

von Schneidemesser, E., Schmitz, S., & Caseiro, A. (2021). Der Effekt eines neuen Radweges und einer Spielstrasse auf die Luftverschmutzung: Ein Real-Experiment in Berlin. In I. f. A. S. S. e. V. (IASS) (Ed.), (pp. 1-9). Potsdam.

von Schneidemesser, E., Sibiya, B., Caseiro, A., Butler, T., Lawrence, M. G., Leitao, J., . . . Salvador, P. (2021). Learning from the COVID-19 lockdown in berlin: Observations

and modelling to support understanding policies to reduce NO2. *Atmospheric Environment: X, 12*. doi:10.1016/j.aeaoa.2021.100122

von Schneidemesser, E., Steinmar, K., Weatherhead, E. C., Bonn, B., Gerwig, H., & Quedenau, J. (2019a). Air pollution at human scales in an urban environment: Impact of local environment and vehicles on particle number concentrations. *Science of the Total Environment, 688*, 691-700. doi:10.1016/j.scitotenv.2019.06.309

von Schneidemesser, E., Steinmar, K., Weatherhead, E. C., Bonn, B., Gerwig, H., & Quedenau, J. (2019b). Air pollution at human scales in an urban environment: Impact of local environment and vehicles on particle number concentrations. *Sci Total Environ, 688*, 691-700. doi:10.1016/j.scitotenv.2019.06.309

Voordeckers, D., Lauriks, T., Denys, S., Billen, P., Tytgat, T., & Van Acker, M. (2021). Guidelines for passive control of traffic-related air pollution in street canyons: An overview for urban planning. *Landscape and Urban Planning, 207*. doi:10.1016/j.landurbplan.2020.103980

Voordeckers, D., Meysman, F. J. R., Billen, P., Tytgat, T., & M. (2021). The impact of street canyon morphology and traffic volume on NO2 values in the street canyons of Antwerp. *Building and Environment, 197*, 107825. doi:https://doi.org/10.1016/j.buildenv.2021.107825

Voordeckers, D., Meysman, F. J. R., Billen, P., Tytgat, T., & Van Acker, M. (2021). The impact of street canyon morphology and traffic volume on NO2 values in the street canyons of Antwerp. *Building and Environment, 197*. doi:10.1016/j.buildenv.2021.107825

Wager, S., Hastie, T., & Efron, B. (2014). Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife. *Journal of Machine Learning Research, 15*, 1625-1651.

Wagner, P., & Schäfer, K. (2017). Influence of mixing layer height on air pollutant concentrations in an urban street canyon. *Urban Climate, 22*, 64-79. doi:10.1016/j.uclim.2015.11.001

Wallace, J., & Kanaroglou, P. (2008). Modeling NOx and NO2 emissions from mobile sources: A case study for Hamilton, Ontario, Canada. *Transportation Research Part D: Transport and Environment, 13*(5), 323-333. doi:10.1016/j.trd.2008.04.001

Wang, T., & Xie, S. (2009). Assessment of traffic-related air pollution in the urban streets before and during the 2008 Beijing Olympic Games traffic control period. *Atmospheric Environment, 43*(35), 5682-5690. doi:10.1016/j.atmosenv.2009.07.034

Weichenthal, S., Kulka, R., Dubeau, A., Martin, C., Wang, D., & Dales, R. (2011). Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists. *Environ Health Perspect, 119*(10), 1373-1378. doi:10.1289/ehp.1003321

WHO. (2021). *WHO global air quality guidelines. Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.* Geneva: World Health Organization.

Williams, D. E. (2019). Low Cost Sensor Networks: How Do We Know the Data Are Reliable? *ACS Sens, 4*(10), 2558-2565. doi:10.1021/acssensors.9b01455

Williams, D. E. (2020). Electrochemical sensors for environmental gas analysis. *Current Opinion in Electrochemistry, 22*, 145-153. doi:10.1016/j.coelec.2020.06.006

Williams, R., Duvall, R., Kilaru, V., Hagler, G., Hassinger, L., Benedict, K., . . . Ning, Z. (2019). Deliberating performance targets workshop: Potential paths for emerging PM2.5 and O3 air sensor progress. *Atmos Environ X, 2*, 100031. doi:10.1016/j.aeaoa.2019.100031

Williams, R., Kilaru, V., Snyder, E., Kaufman, A., Dye, T., Rutter, A., . . . Hafner, H. (2014). *Air Sensor Guidebook*. Retrieved from Washington, DC:

Wood, H. E., Marlin, N., Mudway, I. S., Bremner, S. A., Cross, L., Dundas, I., . . . Griffiths, C. J. (2015). Effects of Air Pollution and the Introduction of the London Low Emission Zone on the Prevalence of Respiratory and Allergic Symptoms in Schoolchildren in East London: A Sequential Cross-Sectional Study. *PLoS One, 10*(8), 1-12. doi:10.1371/journal.pone.0109121

Wood, S. N. (2017). *Generalized Additive Models* (2nd Edition ed.).

World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). (2019). Retrieved from New York:

Xie, S., Zhang, Y., Qi, L., & Tang, X. (2003). Spatial distribution of traffic-related pollutant concentrations in street canyons. *Atmospheric Environment, 37*(23), 3213-3224. doi:10.1016/s1352-2310(03)00321-2

Xu, J., Jiang, H., Zhao, H., & Stephens, B. (2017). Mobile Monitoring of Personal NOx Exposures during Scripted Daily Activities in Chicago, IL. *Aerosol and Air Quality Research, 17*(8), 1999-2009. doi:10.4209/aaqr.2017.02.0063

Yan, F., Winijkul, E., Streets, D. G., Lu, Z., Bond, T. C., & Zhang, Y. (2014). Global emission projections for the transportation sector using dynamic technology modeling. *Atmospheric Chemistry and Physics, 14*(11), 5709-5733. doi:10.5194/acp-14-5709-2014

Ye, J., Qin, Z., & Chen, X. (2021). Adapt by adopting cleaner vehicles? — Evidence from a low-emission zone policy in Nanchang, China. *China Economic Review, 66*, 101598. doi:10.1016/j.chieco.2021.101598

Zhang, K., Chen, G., Zhang, Y., Liu, S., Wang, X., Wang, B., & Hang, J. (2020). Integrated impacts of turbulent mixing and NOX-O3 photochemistry on reactive pollutant dispersion and intake fraction in shallow and deep street canyons. *Sci Total Environ, 712*, 135553. doi:10.1016/j.scitotenv.2019.135553

Zheng, X.-y., Orellano, P., Lin, H.-l., Jiang, M., & Guan, W.-j. (2021). Short-term exposure to ozone, nitrogen dioxide, and sulphur dioxide and emergency department visits and hospital admissions due to asthma: A systematic review and meta-analysis. *Environment International, 150*. doi:10.1016/j.envint.2021.106435

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., & Robinson, A. L. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques, 11*(1), 291-313. doi:10.5194/amt-11-291-2018