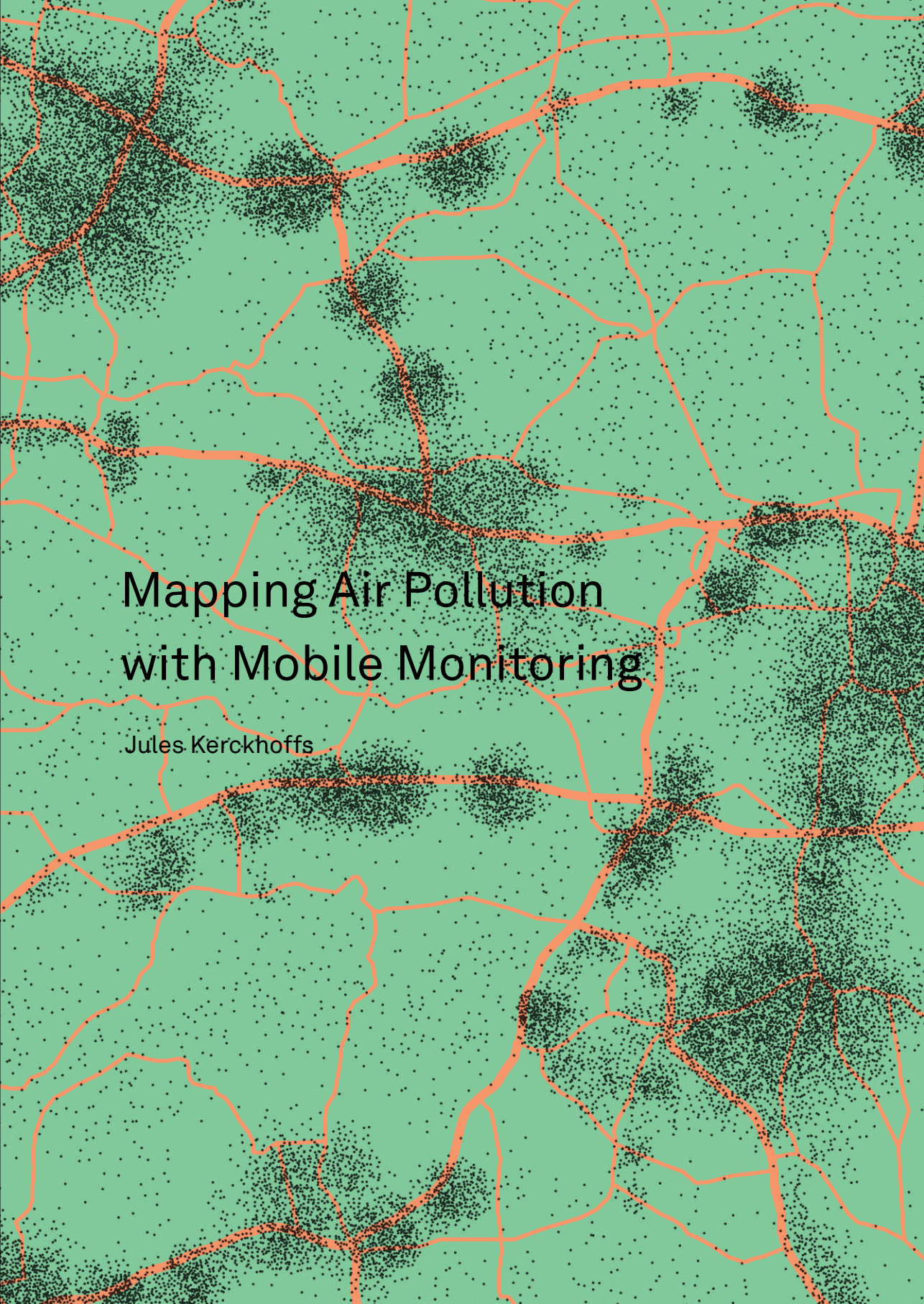


Mapping Air Pollution with Mobile Monitoring

Jules Kerckhoffs

Graphic design
Saus (Frank Slangen)

Printing
Ipskamp Printing

ISBN
978-90-39373835

Mapping Air Pollution with Mobile Monitoring

Het gebruik van mobiele monitoring om luchtvervuiling in kaart te brengen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 8 juni 2021 des middags te 2.15 uur

door

Jules Kerckhoffs

geboren op 28 september 1988
te Maastricht

Promotoren:
Prof. dr. R.C.H. Vermeulen
Prof. dr. B. Brunekreef

Copromotor:
Dr. G. Hoek

The studies in this thesis were supported by an ASPASIA grant from the Dutch Research Council (NWO; project number 015.010.044), the Environmental Defense Fund, EXPOSOME-NL (NWO; project number 024.004.017) and EXPANSE (EU-H2020 Grant number 874627).

Roadmap

Introduction

Ambient air pollution accounts worldwide for an estimated 4.2 million deaths per year due to stroke, heart disease, lung cancer and chronic respiratory diseases[1]. And even though air pollution standards set by the European Union are mostly met, concentration levels in Europe are largely above WHO (World Health Organization) air quality guidelines. On top of that, recent studies found elevated risks to adverse health-effects at levels well below the current WHO standards [2–5].

A considerable part of air pollution is caused by motorized traffic, especially in urban environments[6,7]. The variation in concentration levels is therefore also largest in urban environments, opposed to rural areas. Since epidemiological studies link air quality to health-effects it is important to characterise intra-urban air pollution as best as possible. And in order to assess people's exposure to air pollution (about 75% of the European population lives in urban environments[8]) exposure maps with a high spatial resolution are crucial. At the same time could this benefit (local) policy as well.

Two air pollutants that show the biggest variation in cities are ultrafine particles (UFP) and black carbon (BC). UFP are particles smaller than 100nm and have varying composition. They are present in the air due to natural sources but are increasingly subscribed to anthropogenic activities[6]. BC is the dark carbonaceous component of particulate matter (PM), also known as soot, originates from coal and wood burning and in Europe mainly from (diesel) engines[9].

See box 1.1 for additional information on UFP and BC.

Where larger particles (>PM_{10}) are mainly blocked in the throat and nose, UFP and BC can enter the lung alveoli, with UFP as particular concern due to their even smaller size; allowing them to penetrate deeper into the lungs, cross biological barriers, enter the bloodstream and reach other organs[6,10–12]. Though epidemiological studies to long-term UFP exposure are scarce, multiple toxicological studies suggest that UFP have a higher toxicity per mass unit than larger particles.

Important characteristic of these air pollutants is that they are highly variable is space and time and exhibit large spatial variability within city limits[18], making it difficult to assess their variation. However, ambient air pollution measurement stations are generally scarce and most of them are intentionally located far from emission sources. This enables them to measure air pollution with high precision and excellent temporal coverage but with limited spatial information[19].

UFP and BC measurements are therefore increasingly performed with mobile platforms. In the first place because UFP and BC instruments are not suited to measure a lot of locations for an extended amount of time, for example like NO_2, which can relatively easily be measured with passive samplers.

See box 1.2 for a description of the mobile platform.

Box 1.1

UFP and BC

↓

UFP

Ultrafine particles (UFPs) are defined as particles with diameters smaller than 0.1 µm (100nm). They are expressed as number of particles per volume or air, mostly particles/cm³. Sometimes they are also referred to as particle number concentration (PNC). In urban areas, UFP exposure mostly originates from combustion processes by motorized vehicles (and other forms of transport, like air traffic), with peak concentrations near airports, highways and major roadways[13], even up to tenfold higher than background concentrations [14].

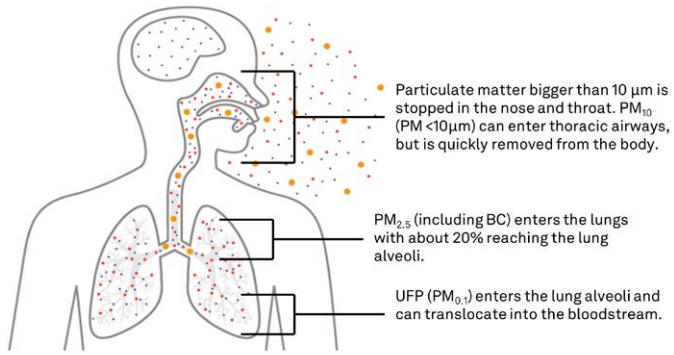


Figure 1.1. Size of particles and their ability to enter the human body.

BC

Black carbon is a constituent of fine particles ($PM_{2.5}$) and emitted from combustion-related sources including biomass burning, residential heating and cooking, industry, and transportation[15]. BC concentrations near streets will therefore fluctuate much more than concentrations at background locations[16]. BC also varies more in space and time than $PM_{2.5}$, creating larger exposure contrasts, and hence better assess local combustion sources opposed to $PM_{2.5}$. Moreover, studies have indicated that BC is related to multiple health outcomes and valuable additional marker for health effects next to $PM_{2.5}$[17].

And secondly as a mobile platform provides the possibility to sample more spatially diverse environments in less time, with a limited number of (costly) monitoring devices. Together with advancements in air monitoring instrumentation, such as higher time resolution and greater portability, these platforms can capture the high variability of UFP and BC in space and time in a complex urban terrain[18,20].

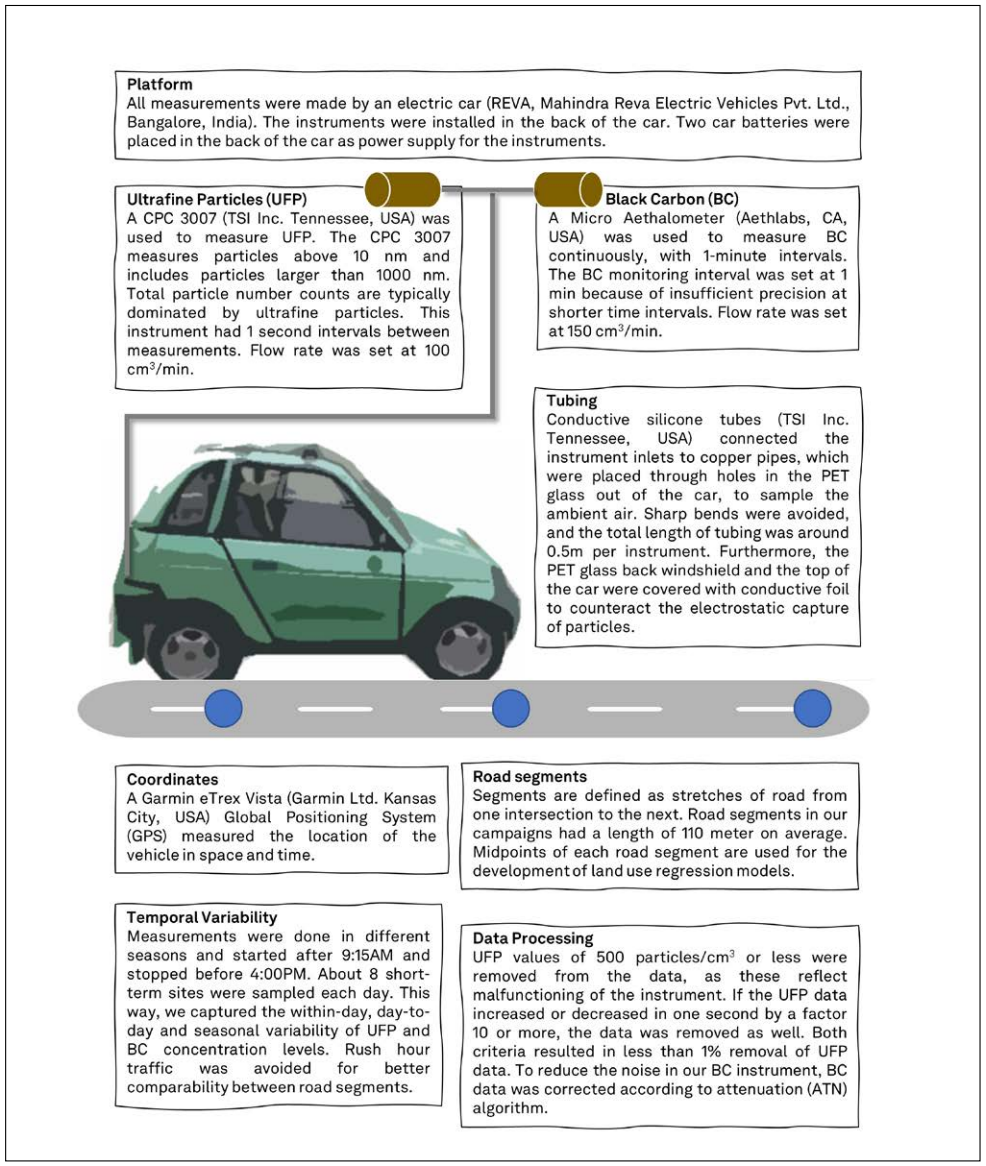
Terminology on the use of a mobile platform to measure air pollution is not used unambiguously in literature, as mobile measurements can be regarded as short-term measurements and vice-versa. Throughout this thesis, I will use short-term (stationary) measurements for measurements that

were sampled when standing still, usually about 15 to 60 minutes at each site[21–24]. Mobile measurements are considered as measurements done without stopping, by using for example cars[9,25–27], bikes[28,29], backpacks[30] and combinations thereof[31].

Box 1.2

Mobile Platform

↓



Disadvantages of mobile monitoring are the temporal resolution (as sites/segments are mostly sampled for seconds to minutes) and the fact that measurements are done on-road and might not reflect residential exposure. Intuitively, mobile measurements are not suited for assessing long-term exposure.

See box 1.3 for an explanation of LUR models.

However, for the assessment of spatial variation it might not be needed to correctly assess all sites/segments. We can alleviate this by building (land-use regression; LUR) models. Montagne et al[22] showed that a model based on short-term measurements was able to predict the spatial variation of independent long-term measurements moderately high, performing better than its model R^2. In this thesis, I will assess if this conclusion can be extended to mobile measurements.

Like short-term measurements, mobile measurements do not only consist of spatial variation, but also have a temporal component. The goal of this thesis is to develop spatial maps for UFP and BC, meaning there is no need to explain this temporal variability. We therefore only offer spatial predictor variables to models in order to explain the spatial variation. So, since models are not offered predictor variables that explain the temporal variation, the chance to develop a good model is very small[32]. However, regression-based techniques might not require a very good model to predict the long-term spatial variation of air pollution. This means that long-term measurements will be essential to truly assess the performance of models based on short-term and mobile measurements.

Additional element to consider for mobile measurements is that they are done on-road, whereas short-term measurements are usually performed on the side of the road, as close to the façade of a building as possible. This could potentially lead to bias when the models are applied to residential addresses, as UFP and BC are known to have steep concentration gradients[13,33].

The first research question in this thesis is therefore if (and how) mobile monitoring can be used to develop land use regression models for estimating long-term exposure to UFP and BC. How does the limited measuring time per segment (decreased temporal coverage) and the increased number of roads that are measured (increased spatial coverage) affect model performance?

See box 1.4 for the different campaigns used in this thesis.

In chapter 2, I start with answering question 1 by comparing mobile monitoring to short-term stationary monitoring. To do this I will use UFP and BC data from a single measurement campaign (MUSiC, see box 1.4), where mobile and short-term measurements were conducted alternately. Both subsets (mobile and short-term) are used to develop land-use regression (LUR) models. These models are then used to predict concentration levels

on a set of 1000 random addresses in the Netherlands and compared to one another. In chapter 3, I will use a second measurement campaign (EXPOsOMICS, see box 1.4), also measuring UFP and BC and with a sampling strategy including mobile and short-term measurements alternately. On top of that, this campaign consisted of measuring longer-term average (3 x 24 hours) concentrations of UFP at 42 locations in Amsterdam and Utrecht. This way, I am also able to compare predictions from mobile and short-term stationary LUR models to actual long-term measurements.

Box 1.3

Land Use Regression Models

↓

Land use regression (LUR) models are used to describe the relationship between a dependent variable (in this case UFP and BC) and several geographical information system (GIS) variables. The midpoint of each road segment was identified and used as a coordinate for obtaining GIS predictors for LUR modelling. The average concentration adjusted for temporal variation of both pollutants per road segment was used as dependent variables.

GIS variables used in the different campaigns are:

- Traffic variables, including traffic intensity and road length variables (in 25m to 1000m buffers).
- Land use (e.g., port, airport, industry, urban green) in buffers from 100 to 5000m
- Population / household density in buffers from 100 to 5000m.
- Distance to highways, railways, restaurants, gas stations and traffic lights were offered in the latter campaigns as well.

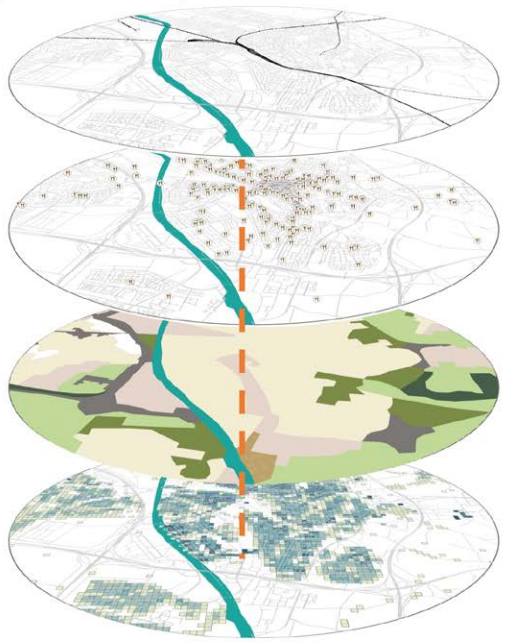


Figure: GIS variables in Utrecht. From top to bottom: road and rail network, restaurants, land-use and population density.

LUR model development
The development of our supervised forward stepwise LUR models starts with an empty (intercept only) model and then adds variables based on goodness of fit, determined by the (adjusted) R^2 value. Variables are selected that lead to the largest increase in adjusted R^2; the coefficient conforms with a predefined direction of effect and the direction of effect of predictors already in the model does not change. Model building stops when new variables were not able to improve the adjusted R^2. Predictor variables in the final model are removed from the model when they have a p-value less than 0.1, or a variance inflation factor over 3.

Since all models in chapter 2 and 3 are based on linear regression methods, I explore the possibilities to create different prediction algorithms based on mobile and short-term stationary data in the second part of this thesis. For a range of prediction algorithms, including regularisation, non-linear and machine learning approaches, I analyse if they can improve predictions compared to the widely used linear regression methods in chapter 4.

In the third part, I will look at the possibility to spatially scale up exposure maps for UFP. So far, UFP models are generally restricted to single city's boundaries, limiting their ability to be used in wide-scale epidemiological studies including peri-urban and rural areas [34]. Nation-wide epidemiological studies including administrative cohort studies have been very informative to assess health effects of air pollutants such as $PM_{2.5}$ and NO_2, because of the large size of the population, avoidance of selection bias, and the increased contrast in exposure. These studies have become possible because of national exposure models, typically based on nation-wide routine monitoring and modelling, incorporating variation at the regional background, urban background, and local scale. More specifically, the inclusion of regional background concentrations in LUR models advanced exposure assessment and health-effect studies substantially by increasing exposure contrast [35]. This approach is not directly possible for UFP because of the lack of nationwide routine monitoring. Therefore, we combined targeted mobile monitoring and long-term regional background monitoring in a new campaign (RUN, see box 1.4) to develop national UFP models. First by exploring the regional differences in UFP concentrations in the Netherlands in chapter 5 and then by combining all efforts learned in the previous chapters to create nationwide LUR models in chapter 6. I will also look at the possibility to separately model the local and the background concentrations.

In chapter 7, I will discuss all findings in the general discussion and come back to the 3 main research questions stated in this introduction.

1. Can we (and how can we) use mobile measurements to develop fine resolution spatial maps for long-term average UFP and BC concentrations?

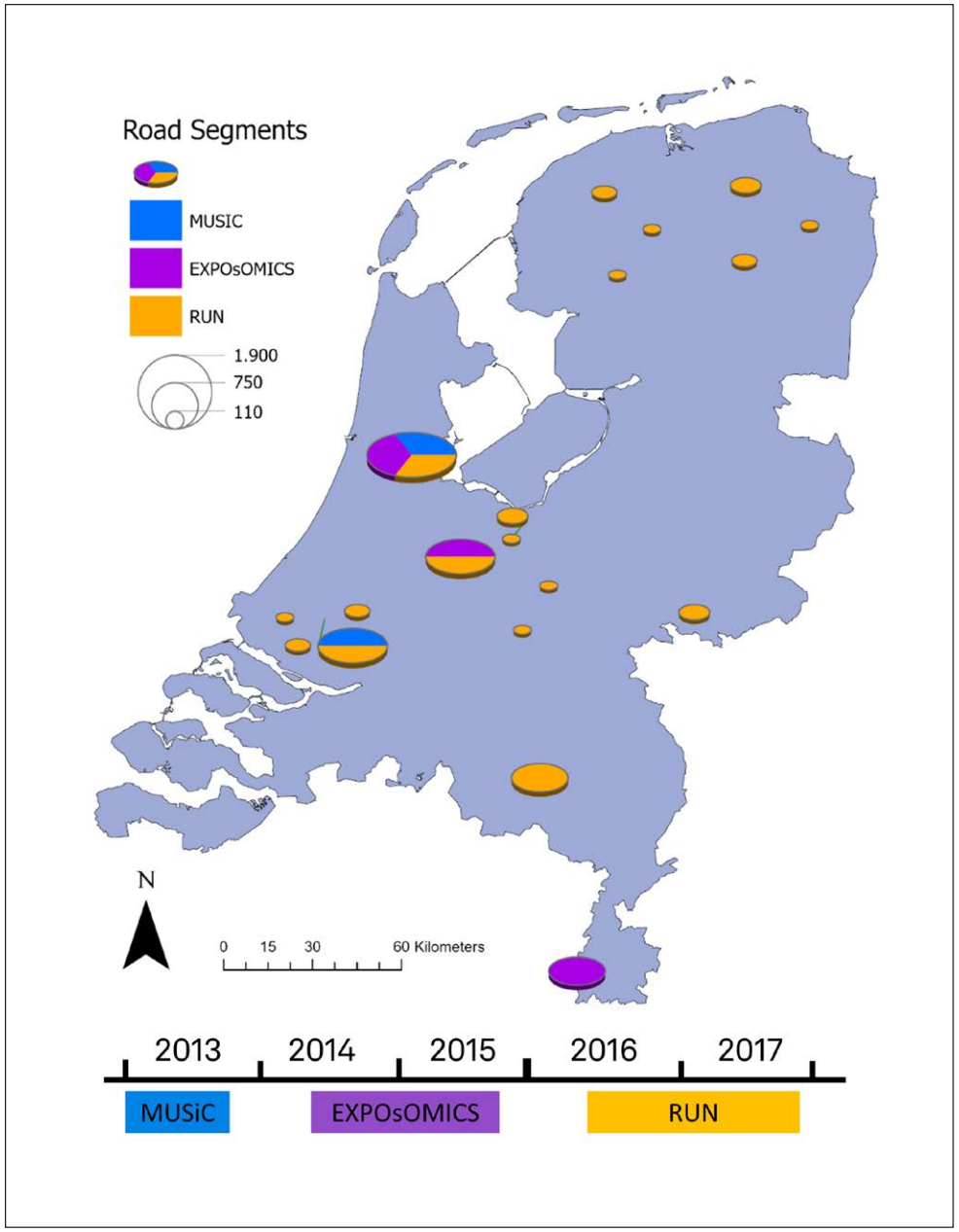
2. Can we improve long-term average concentration maps by applying different prediction algorithms?

3. Can we use a mobile monitoring design for application in nation-wide cohorts, by accounting for regional background concentrations?

Box 1.4

Different measurement campaigns.

↓



References

(1) WHO. Ambient air pollution https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/ambient-air-pollution (accessed Dec 2, 2020).

(2) Hvidtfeldt, U. A.; Severi, G.; Andersen, Z. J.; Atkinson, R.; Bauwelinck, M.; Bellander, T.; Boutron-Ruault, M. C.; Brandt, J.; Brunekreef, B.; Cesaroni, G.; Chen, J.; Concin, H.; Fecht, D.; et al. Long-Term Low-Level Ambient Air Pollution Exposure and Risk of Lung Cancer – A Pooled Analysis of 7 European Cohorts. *Environ. Int.* **2021**, *146*, 106249. https://doi.org/10.1016/j.envint.2020.106249.

(3) Liu, S.; Jørgensen, J. T.; Ljungman, P.; Pershagen, G.; Bellander, T.; Leander, K.; Magnusson, P. K. E.; Rizzuto, D.; Hvidtfeldt, U. A.; Raaschou-Nielsen, O.; Wolf, K.; Hoffmann, B.; Brunekreef, B.; Weinmayr, G.; Hoek, G.; Andersen, Z. J. Early View Long-Term Exposure to Low-Level Air Pollution and Incidence of Asthma: The ELAPSE Project. https://doi.org/10.1183/13993003.030992020.

(4) Di, Q.; Wang, Y.; Zanobetti, A.; Wang, Y.; Koutrakis, P.; Choirat, C.; Dominici, F.; Schwartz, J. D. Air Pollution and Mortality in the Medicare Population. *N. Engl. J. Med.* **2017**, *376* (26), 2513–2522. https://doi.org/10.1056/nejmoa1702747.

(5) Crouse, D. L.; Peters, P. A.; Hystad, P.; Brook, J. R.; van Donkelaar, A.; Martin, R. V.; Villeneuve, P. J.; Jerrett, M.; Goldberg, M. S.; Pope, C. A.; Brauer, M.; Brook, R. D.; Robichaud, A.; Menard, R.; Burnett, R. T. Ambient $PM_{2.5}$, O_3, and NO_2 Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environ. Health Perspect.* **2015**, *123* (11), 1180–1186. https://doi.org/10.1289/ehp.1409276.

(6) Morawska, L.; Ristovski, Z.; Jayaratne, E. R.; Keogh, D. U.; Ling, X. Ambient Nano and Ultrafine Particles from Motor Vehicle Emissions: Characteristics, Ambient Processing and Implications on Human Exposure. *Atmospheric Environment*. November 2008, pp 8113–8138. https://doi.org/10.1016/j.atmosenv.2008.07.050.

(7) Ghassoun, Y.; Ruths, M.; Löwner, M. O.; Weber, S. Intra-Urban Variation of Ultrafine Particles as Evaluated by Process Related Land Use and Pollutant Driven Regression Modelling. *Sci. Total Environ.* **2015**, *536*, 150–160. https://doi.org/10.1016/j.scitotenv.2015.07.051.

(8) Environmental, E.; Agency. Urban environment http://www.eea.europa.eu/themes/urban#tab-publications (accessed Jul 25, 2016).

(9) Larson, T.; Henderson, S. B.; Brauer, M. Mobile Monitoring of Particle Light Absorption Coefficient in an Urban Area as a Basis for Land Use Regression. *Environ. Sci. Technol.* **2009**, *43* (13), 4672–4678. https://doi.org/10.1021/es803068e.

(10) HEI Review Panel. Understanding the Health Effects of Ambient Ultrafine Particles. *Heal. Eff. Inst.* **2013**, No. January, 122.

(11) Ohlwein, S.; Kappeler, R.; Kutlar Joss, M.; Künzli, N.; Hoffmann, B. Health Effects of Ultrafine Particles: A Systematic Literature Review Update of Epidemiological Evidence. *International Journal of Public Health*. Springer International Publishing May 1, 2019, pp 547–559. https://doi.org/10.1007/s00038-019-01202-7.

(12) Oberdörster, G.; Oberdörster, E.; Oberdörster, J. Nanotoxicology: An Emerging Discipline Evolving from Studies of Ultrafine Particles. *Environ. Health Perspect.* **2005**, *113* (7), 823–839. https://doi.org/10.1289/ehp.7339.

(13) Simon, M. C.; Hudda, N.; Naumova, E. N.; Levy, J. I.; Brugge, D.; Durant, J. L. Comparisons of Traffic-Related Ultrafine Particle Number Concentrations Measured in Two Urban Areas by Central, Residential, and Mobile Monitoring. *Atmos. Environ.* **2017**, *169*, 113–127. https://doi.org/10.1016/j.atmosenv.2017.09.003.

(14) Karner, A. A.; Eisinger, D. S.; Niemeier, D. A. Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. *Environ. Sci. Technol.* **2010**, *44* (14), 5334–5344. https://doi.org/10.1021/es100008x.

(15) Luben, T. J.; Nichols, J. L.; Dutton, S. J.; Kirrane, E.; Owens, E. O.; Datko-Williams, L.; Madden, M.; Sacks, J. D. A Systematic Review of Cardiovascular Emergency Department Visits, Hospital Admissions and Mortality Associated with Ambient Black Carbon. *Environment International*. Elsevier Ltd October 1, 2017, pp 154–162. https://doi.org/10.1016/j.envint.2017.07.005.

(16) Dons, E.; Van Poppel, M.; Kochan, B.; Wets, G.; Int Panis, L. Modeling Temporal and Spatial Variability of Traffic-Related Air Pollution: Hourly Land Use Regression Models for Black Carbon. *Atmos. Environ.* **2013**, *74*, 237–246. https://doi.org/10.1016/j.atmosenv.2013.03.050.

(17) Janssen, N. A. H.; Hoek, G.; Simic-Lawson, M.; Fischer, P.; van Bree, L.; ten Brink, H.; Keuken, M.; Atkinson, R. W.; Anderson, H. R.; Brunekreef, B.; Cassee, F. R. Black Carbon as an Additional Indicator of the Adverse Health Effects of Airborne Particles Compared with PM_{10} and $PM_{2.5}$. *Environ. Health Perspect.* **2011**, *119* (12), 1691–1699. https://doi.org/10.1289/ehp.1003369.

(18) Zwack, L. M.; Hanna, S. R.; Spengler, J. D.; Levy, J. I. Using Advanced Dispersion Models and Mobile Monitoring to Characterize Spatial Patterns of Ultrafine Particles in an Urban Area. *Atmos. Environ.* **2011**, *45* (28), 4822–4829. https://doi.org/10.1016/j.atmosenv.2011.06.019.

(19) Healtheffects.org. Health Effects https://www.healtheffects.org/sites/default/files/Apte_0.pdf (accessed Dec 3, 2020).

(20) Zwack, L. M.; Paciorek, C. J.; Spengler, J. D.; Levy, J. I. Modeling Spatial Patterns of Traffic-Related Air Pollutants in Complex Urban Terrain. *Environ. Health Perspect.* **2011**, *119* (6), 852–859. https://doi.org/10.1289/ehp.1002519.

(21) Abernethy, R. C.; Allen, R. W.; McKendry, I. G.; Brauer, M. A Land Use Regression Model for Ultrafine Particles in Vancouver, Canada. *Environ. Sci. Technol.* **2013**, *47* (10), 5217–5225. https://doi.org/10.1021/es304495s.

(22) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712–8720. https://doi.org/10.1021/es505791g.

(23) Rivera, M.; Basagaña, X.; Aguilera, I.; Agis, D.; Bouso, L.; Foraster, M.; Medina-Ramón, M.; Pey, J.; Künzli, N.; Hoek, G. Spatial Distribution of Ultrafine Particles in Urban Settings: A Land Use Regression Model. *Atmos. Environ.* **2012**, *54*, 657–666. https://doi.org/10.1016/j.atmosenv.2012.01.058.

(24) Saraswat, A.; Apte, J. S.; Kandlikar, M.; Brauer, M.; Henderson, S. B.; Marshall, J. D. Spatiotemporal Land Use Regression Models of Fine, Ultrafine, and Black Carbon Particulate Matter in New Delhi, India. *Environ. Sci. Technol.* **2013**, *47* (22), 12903–12911. https://doi.org/10.1021/es401489h.

(25) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An Hourly Regression Model for Ultrafine Particles in a Near-Highway Urban Area. *Environ. Sci. Technol.* **2014**, *48* (6), 3272–3280. https://doi.org/10.1021/es404838k.

(26) Shairsingh, K. K.; Jeong, C. H.; Wang, J. M.; Evans, G. J. Characterizing the Spatial Variability of Local and Background Concentration Signals for Air Pollution at the Neighbourhood Scale. *Atmos. Environ.* **2018**, *183*, 57–68. https://doi.org/10.1016/j.atmosenv.2018.04.010.

(27) Weichenthal, S.; Farrell, W.; Goldberg, M.; Joseph, L.; Hatzopoulou, M. Characterizing the Impact of Traffic and the Built Environment on Near-Road Ultrafine Particle and Black Carbon Concentrations. *Environ. Res.* **2014**, *132*, 305–310. https://doi.org/10.1016/j.envres.2014.04.007.

(28) Farrell, W.; Weichenthal, S.; Goldberg, M.; Valois, M. F.; Shekarrizfard, M.; Hatzopoulou, M. Near Roadway Air Pollution across a Spatially Extensive Road and Cycling Network. *Environ. Pollut.* **2016**, *212*, 498–507. https://doi.org/10.1016/j.envpol.2016.02.041.

(29) Hankey, S.; Marshall, J. D. Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, $PM_{2.5}$, Particle Size) Using Mobile Monitoring. *Environ. Sci. Technol.* **2015**, *49* (15), 9194–9202. https://doi.org/10.1021/acs.est.5b01209.

(30) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Evans, G. J. The Application of Wavelet Decomposition to Quantify the Local and Regional Sources of Ultrafine Particles in Cities. *Atmos. Environ.* **2014**, *95*, 249–257. https://doi.org/10.1016/j.atmosenv.2014.05.035.

(31) Minet, L.; Gehr, R.; Hatzopoulou, M. Capturing the Sensitivity of Land-Use Regression Models to Short-Term Mobile Monitoring Campaigns Using Air Pollution Micro-Sensors. *Environ. Pollut.* **2017**, *230*, 280–290. https://doi.org/10.1016/j.envpol.2017.06.071.

(32) Tan, Y.; Robinson, A. L.; Presto, A. A. Quantifying Uncertainties in Pollutant Mapping Studies Using the Monte Carlo Method. *Atmos. Environ.* **2014**, *99*, 333–340. https://doi.org/10.1016/j.atmosenv.2014.10.003.

(33) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Reali, C.; Sun, T.; Evans, G. J. Development of a Land-Use Regression Model for Ultrafine Particles in Toronto, Canada. *Atmos. Environ.* **2015**, *110*, 84–92. https://doi.org/10.1016/j.atmosenv.2015.02.018.

(34) Hoek, G. Methods for Assessing Long-Term Exposures to Outdoor Air Pollutants. *Current environmental health reports*. Springer December 1, 2017, pp 450–462. https://doi.org/10.1007/s40572-017-0169-5.

(35) Beelen, R.; Hoek, G.; van den Brandt, P. A.; Goldbohm, R. A.; Fischer, P.; Schouten, L. J.; Jerrett, M.; Hughes, E.; Armstrong, B.; Brunekreef, B. Long-Term Effects of Traffic-Related Air Pollution on Mortality in a Dutch Cohort (NLCS-AIR Study). *Environ. Health Perspect.* **2008**, *116* (2), 196–202. https://doi.org/10.1289/ehp.10767.

Comparison of Ultrafine Particles and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model

Jules Kerckhoffs
Gerard Hoek
Kyle P. Messier
Bert Brunekreef
Kees Meliefste
Jochem O. Klompmaker
Roel Vermeulen

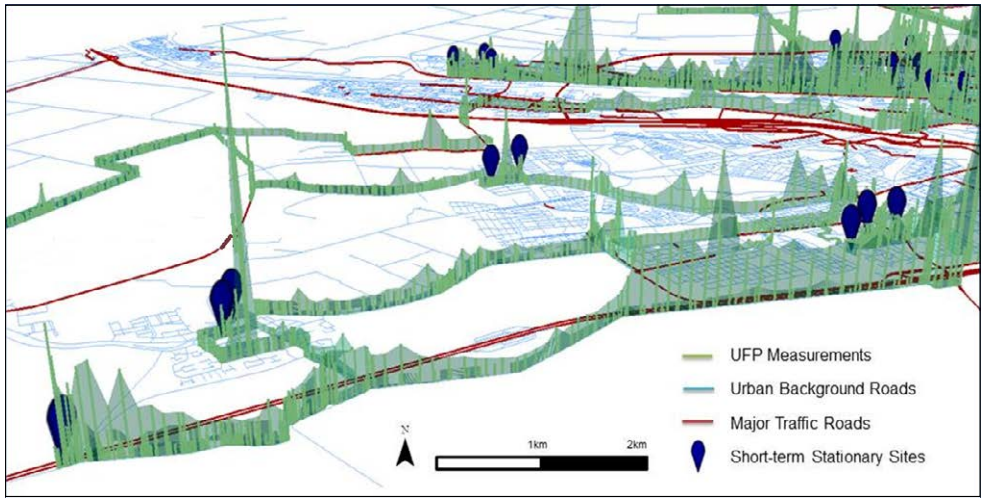
Abstract

Mobile and short-term monitoring campaigns are increasingly used to develop land use regression (LUR) models for ultrafine particles (UFP) and black carbon (BC). It is not yet established whether LUR models based on mobile or short-term stationary measurements result in comparable models and concentration predictions. The goal of this paper is to compare LUR models based on stationary (30 minutes) and mobile UFP and BC measurements from a single campaign. An electric car collected both repeated stationary and mobile measurements in Amsterdam and Rotterdam, The Netherlands. A total of 2,964 road segments and 161 stationary sites were sampled over two seasons. Our main comparison was based on predicted concentrations of the mobile and stationary monitoring LUR models at 12,682 residential addresses in Amsterdam. Predictor variables in the mobile and stationary LUR model were comparable resulting in highly correlated predictions at external residential addresses (R^2 of 0.89 for UFP and 0.88 for BC). Mobile model predictions were on average 1.41 and 1.91 times higher than stationary model predictions for UFP and BC respectively. LUR models based upon mobile and stationary monitoring predicted highly correlated UFP and BC concentration surfaces, but predicted concentrations based on mobile measurements were systematically higher.

TOC/Abstract Art

Typical trajectory of the electric car, shown in 3D.

↓



① Introduction

Multiple studies have shown negative relationships between outdoor particulate matter air pollution and health[1]. Both animal and human studies provide evidence for respiratory and cardiovascular effects associated with exposure to outdoor air pollution, with the ultrafine particle (UFP) fraction[2,3] and Black Carbon (BC)[4] as valuable indicators of pollution mixtures produced by local combustion sources.

Assessing spatial variation of UFP and BC can be challenging as these concentrations are highly variable in space and time, especially in urban environments[5]. Land-use regression (LUR) modelling has proven to be a useful tool to predict (long-term) intra-urban spatial variation of outdoor air pollution[6]. LUR models for components such as nitrogen dioxide (NO_2), particles smaller than 2.5 µm ($PM_{2.5}$) and BC are usually based on 20-100 locations measured repeatedly over one or two week time periods[6]. For UFP, long-term sampling is complicated because instruments need frequent quality control and maintenance and are expensive compared to measurement instruments for NO_2 and $PM_{2.5}$[7]. Although long-term LUR models exist for UFP[8], most monitoring campaigns are based on short-term measurements ranging from 15 minutes to 1 hour per site[9–12] or true mobile monitoring with measurements obtained from a platform moving in traffic measuring often at a one second to one minute interval[13–17].

Short-term stationary and mobile monitoring reduces total measurement time significantly, allowing measurement of a larger number of sites. However, they present challenges in separating spatial and temporal variability compared to study designs with long-term fixed sites measured simultaneously. In mobile monitoring, this is complicated further by the even shorter time per location (typically road segment) compared to short-term monitoring. On-road monitoring may further not be representative for modelling of residential exposures as dwellings can, depending on the urban topography, be several meters away from the roadside.

Previous studies have used either short-term stationary or mobile monitoring, but not in one study area. Short-term stationary monitoring campaigns for UFP and BC have been conducted in Vancouver, Canada[10] Girona, Spain[11], Amsterdam and Rotterdam, the Netherlands[9] and New Delhi, India[12]. Typically, models explained a moderate amount of variation in measured concentrations (R^2 30-50%). The moderate R^2 has been attributed to the large variability of short-term stationary measurements compared to campaigns using repeated 1-2 week measurement periods[9]. True mobile monitoring for UFP has been performed in Toronto and

Montreal, Canada[15–17], Somerville, USA[14] and Minneapolis, USA[7]. Model R^2 was between 40 and 70%. Model R^2 can however not be easily compared between studies as some are spatial and some are spatiotemporal models, and sampling campaigns differ in duration and number of sites.

To date there is no systematic comparison between LUR models of UFP and BC based on mobile and short-term stationary monitoring in the same study. We performed a measurement campaign in which we collected short-term stationary measurements (30 minutes) and measured in a mobile fashion in between the short-term stationary measurements using an electrical car. We previously showed that LUR models can be developed based on short-term stationary measurements in this region[9,18]. In this paper, we extend this work by developing LUR models based on mobile measurements and compare both methods in their ability to predict UFP and BC concentration surfaces of residences in the study area.

② Methods

2.1 Study Design

The monitoring campaign and site selection has been described previously[18]. Briefly, ultrafine particles were measured each second using a CPC 3007 (TSI Inc. Tennessee, USA). Black Carbon was measured each minute using the micro Aethalometer (Aethlabs, CA, USA). Instruments were installed in the back of an electric vehicle (REVA, Mahindra Reva Electric Vehicles Pvt. Ltd., Bangalore, India), connected to inlets on the outside of the car. A Garmin eTrex Vista (Garmin Ltd. Kansas City, USA) Global Positioning System (GPS) measured the location of the electric vehicle in space and time. Sites were selected with a wide range of spatial contrast in traffic characteristics and land use[9,18]. Specifically, we selected traffic sites (>10,000 vehicles per day), urban background sites, sites near urban green, rural sites (outside the city) and sites near waterways.

Complete measurements of both air pollution and GPS were collected for 42 days in two seasons, winter and spring. 2,964 unique road segments were monitored between January 16th and May 22nd of 2013, when driving from one stationary measurement to the next. Measurements were performed between 9:15 and 16:00 to increase comparability between sites (avoiding rush hour). Of those 2,964 road segments 745 (514) road segments had measurements of UFP and BC in both winter and spring. The abstract art shows a typical trajectory of the electric car.

2.2 Data Aggregation

Following previous mobile monitoring studies[15,16], we averaged air pollution per road segment, defined as a part of a road between two consecutive intersections. Because all measurements were performed on the road, GPS points were snapped to the nearest road segment along the predetermined route to correct for small positional errors of the GPS. Road segments in tunnels or on bridges were deleted from the dataset (n=30), as they are not representative for residential streets. Road segments were on average 130 meters long and comprised on average 12 seconds of UFP data. BC concentrations were based on one-minute values assigned to every road segment the car was on in that minute. On average this means three road segments were assigned the same BC measurement. A cut-off point of 10,000 vehicles per day was used to distinguish traffic and urban background road segments, following the definition for stationary sites[9]. Approximately 40% of road segments were considered as traffic and 60% as urban background road segments.

2.3 Air Pollution Data Processing

UFP data was removed if the concentration was below 500 particles per cm^3 and if the UFP concentration decreased or increased within a second by a factor 10, following the procedures of Klompmaker et al.[18]. These criteria resulted in less than 1% of observations for short-term and mobile sites being removed. Concentrations lower than 500 particles per cm^3 were mostly zero, reflecting instrument malfunctioning.

For mobile monitoring, it is considered important to identify sampling events close to high emission vehicles[19], as these events can influence LUR model development. We defined samples influenced by local exhaust plumes if an UFP concentration was three standard deviations above the previous measurement second, based on the concentrations distribution for that day. Observations remained flagged until they dropped beneath the day average plus one standard deviation. This method is based on the method used by Drewnick et al.[20]. For the main analyses we used all measurements, including road segments with local exhaust plumes. For sensitivity analysis, we excluded them.

Our BC instrument generates minute averages, but this still can be a too short time period to produce reliable concentration levels, because of a too small change in attenuation (ATN). Data with a too small ATN or negative ATN were corrected using the algorithm proposed by Hagler et al.[21]. BC concentrations were only assigned when a threshold of 0.05 ATN change was fulfilled. Minute values with less than 0.05 ATN change were averaged over time until the criteria was met. The algorithm leads to significant noise reduction in our instrument while preserving high-resolution temporal variation. 92 percent of the data have a three-minute time resolution or less. Only 20 percent of the data had a one-minute time resolution. Local exhaust plumes for BC were based on road segments for which UFP had a local exhaust plume, as 1-minute BC measurements are too long to detect local concentration peaks.

2.4 Temporal Variation

A reference site with the same equipment as the electric vehicle was set up in Utrecht (about 30km from Amsterdam and 50km from Rotterdam), The Netherlands, to correct for temporal variation. We used the difference method for correcting the spatial data, following previous work in the stationary campaign[18]. First, the overall mean concentration of the entire campaign at the reference site was calculated. Next, for each minute at the reference site an average of 30 minutes around time x was calculated which was subtracted from the overall mean concentration at the reference site. The difference is then used to adjust the original concentration measured at the sampling locations.

2.5 Model Development

The midpoint of each road segment was identified and used as a coordinate for obtaining GIS predictors for LUR modelling. The average concentration adjusted for temporal variation of both pollutants per road segment was used as dependent variables in a linear regression analysis with multiple GIS variables as independent variables. We used the original concentration scale in order to predict arithmetic averages for epidemiological studies. We used data from all road segments, even when only one measurement was available. We observed similar standard errors of regression coefficients with this LUR modelling approach compared to using only the 745 (of 2,964) road segments with at least two observations. GIS predictor variables were described previously[9] and summarized in Supplement Table S1. Briefly, a range of traffic variables was defined, including traffic intensity and road length variables (in 25m to 1000m buffers). Inverse distance to major roads was also used in the stationary model development but not in the mobile monitoring model, as this variable cannot be computed for major roads (distance zero for major road segments). Additionally, land use (e.g. port, industry, urban green) and population / household density in buffers from 100 to 5000m were potential predictor variables. We additionally included airports as potential variable as recent studies found a 3 to 4-fold increase in UFP concentrations near airports[22,23]. These studies found elevated UFP concentrations up to 10km from the airport, so for this variable we included a 5 and 10km buffer as potential variable. The GIS variables were selected using a supervised stepwise selection procedure following our previous study[9]. The direction of the effect for the variables was determined a priori (Table S1) and the variable with the highest adjusted R^2 was entered in the model. Model building stopped when new variables were not able to improve the adjusted R^2. The variables in the resulting models were checked for p-value (removed when p-value >0.10), collinearity (variance inflation factor > 3 were removed), and influential observations (if Cook's D > 1 the model was further examined). We accounted for autocorrelation using a first order autoregressive (AR-1) term in the ARIMA procedure in SAS. Based upon the partial autocorrelation function, an AR-1 term was sufficient to characterize autocorrelation of the residuals. This correlation structure was also found to be most suitable in a mobile monitoring study by Farrell et al.[24]. If after adding an AR-1 term to the identified model, variables were no longer significant (p>0.10), they were removed from the model. We did not use universal kriging to account for the autocorrelation of the data, because in this method the actual measured concentrations at the monitoring sites unduly influence model predictions. Measured concentrations are not precise due to the short duration of the measurements. The strength of mobile measurements is not the precision of individual observations, but the amount of them.

For sensitivity analyses, we also developed models excluding observations flagged as influenced by local exhaust plumes. To test the impact of accounting for autocorrelation, we compared models with and without additional modelling of autocorrelation.

As stationary models in the previous paper were based on three seasons and the true mobile monitoring models on two seasons, stationary LUR models were redeveloped with the same method as above, based on values in winter and spring and only used if both measurements were available. Montagne et al.[9] did not consider airports as potential variable, so we included the area of airports in a 5km and 10km buffer in the new short-term stationary model.

2.6 Comparison between Short-term Stationary and Mobile Monitoring

To compare stationary and mobile monitoring, four different analyses were performed: 1) stationary and mobile measurements were compared by identifying road segments adjacent to stationary sites. Measurement averages on these pseudo co-locations were compared using Pearson and Spearman correlation coefficients. For BC this analysis was not performed because of the one-minute time resolution of the instrument; 2) mobile LUR model predictions at the stationary measurement sites were compared with measured concentrations at these sites; 3) the stationary LUR model predictions were compared with concentrations measured at the mobile road segments. To perform this analysis, road segment coordinates with a value of zero for distance to nearest major road (mobile observations on a major road) were assigned a value of 10 meter, because inverse distance variables (predictor in the stationary models) could not be calculated otherwise. To check whether this number unduly affected the results, other distances (between 4 and 15m) were considered; 4) both LUR models were used to predict concentration levels at residences using an external dataset. GIS variables were used from a cohort study in Amsterdam, consisting of 12,682 residential addresses spread over the city, including urban background and traffic addresses with different land-use characteristics. We only used address information and further details of this cohort can be found elsewhere[25]. The range of predictor variables was truncated to the range observed at the monitoring locations. As application of the LUR models in epidemiological studies was the main goal of model development, this comparison was considered the central comparison between mobile and stationary monitoring models.

③ Results

Figure 1 and table 1 illustrate the variability of average UFP and BC concentrations from mobile and stationary monitoring, stratified by traffic and urban background (UB) sites. Mobile measurements were on average 1.7 times higher than stationary measurements for UFP (21,167 and 12,630 particles per cm³ for mobile and stationary measurements, respectively). Black Carbon concentrations collected on road were on average 2.1 times higher as stationary measurements, 2.83 and 1.35 µg/m³ respectively. Especially the higher percentiles of mobile measurements were about 3 times higher than the stationary measurements (2.8 for UFP and 3.1 for BC for the 95th percentile), likely related to the shorter averaging time of the mobile measurements. Mobile measurements were higher than the stationary measurements both for traffic and for urban background locations.

Table 1

Comparison between average UFP and BC concentrations of all Mobile and short-term stationary Measurements.

↓

Method		Number of Observations	Mean	5th Pct.	Median	95th Pct.
Stationary Average [a]	UFP (in particles/cm³)	128	12,630	5,089	11,474	21,362
	BC (in µg/m³)	141	1.35	0.56	1.13	2.67
Mobile Average [b]	UFP (in particles/cm³)	2,964	21,167	4,391	15,057	59,628
	BC (in µg/m³)	2,336	2.83	0.48	1.79	8.41

[a] Stationary average consists of 2 times 30 minutes. [b] Mobile average consists on average of 18 seconds.

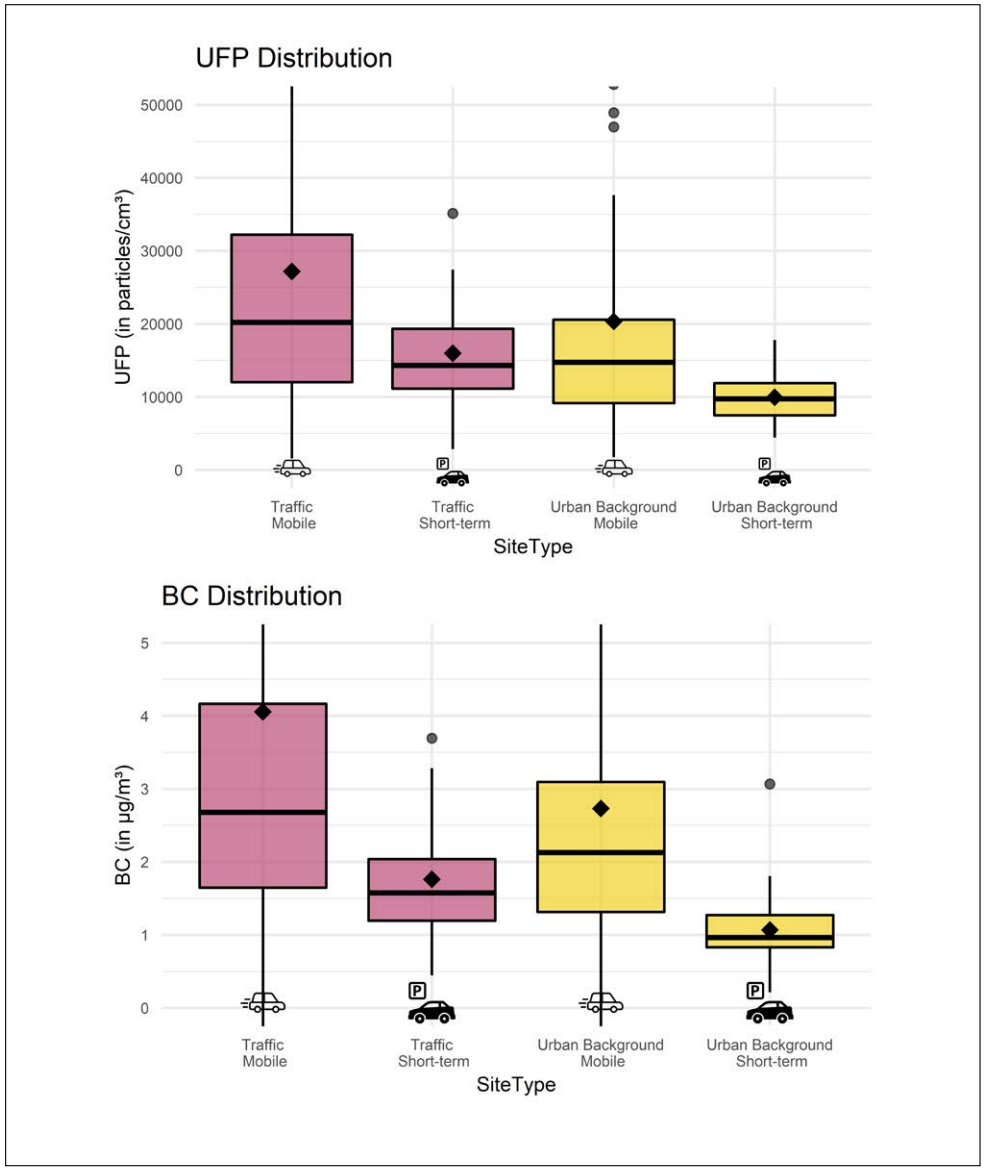
3.1 Mobile and Short-term Stationary Monitoring LUR Models

The LUR models based upon mobile monitoring for UFP and BC are shown in table 2. The model included predictor variables describing traffic (in small and large buffers) and population density for both pollutants. For UFP, airports and ports were also included in the model. As models were developed including an AR-1 term, we cannot report standard R^2 values of our main models. Instead, the reported R^2 value is calculated by regressing the predicted concentration based on the parameter estimates of the mobile AR-1 model without the AR-1 term. R^2 values were low (0.13 and 0.12 for UFP and BC respectively) reflecting the large temporal variability of the short duration measured concentrations. Models developed on road segments with at least two repeats (n=745), increased the R^2 to 0.18 for UFP and 0.30 for BC (Table S2). As the explained variance remains low, all further analyses are based on all road segments. Standard errors of the regression slopes were similar, while losing about 75% of the data. We previously argued that a model with a low model R^2 can provide robust spatial models[9,17].

Figure 1

Distribution of average stationary (30-minute) and road segment UFP and BC concentrations.

↓



Each box shows the median and the 25th and 75th percentiles. The diamond shape represents the group mean. Axes truncated. There were 26 and 18 UFP observations higher than 100,000 p/cm³ at traffic and background locations with a maximum of 221,243 p/cm³. There were 57 and 32 BC observations higher than 10 µg/m³ at traffic and background locations with a maximum of 43 µg/m³.

When models were developed using only road segments without local exhaust plumes (n=2,907), models were very similar to the models including all road segments (Tables S3 and S5) with the exception that some small-scale traffic variables were dropped from the UFP and BC models (traffic intensity on the nearest road). Not accounting for spatial autocorrelation changed the model only modestly (Tables S3 and S5). UFP and BC predictions based on the mobile models with and without AR-1 term and with and without local exhaust plumes were very highly correlated at the 12,682 addresses of the external dataset (Tables S4 and S6), suggesting that these modelling choices do not affect the model predictions substantially.

Table 2

Land-Use Regression Models based upon Mobile Measurements for UFP and BC.

↓

Variables in LUR model	UFP (in particles/cm³)	BC (in µg/m³)
Intercept	5,656 (2,675)	0.48 (0.60)
Population Density in a 5000 meter buffer	8,064 (1,947)	1.15 (0.48)
Airport Area in a 5000 meter buffer	4,669 (1,185)	
Port Area in a 1000 meter buffer	2,499 (1,248)	
Nature Area in 5000 meter buffer	-2,557 (1,357)	
Major Road Length in a 50 meter buffer	6,868 (1,071)	0.61 (0.15)
Traffic Intensity on the Nearest Road		0.30 (0.14)
Traffic Load on Major Roads in a 100 meter buffer	1,928 (1,095)	
Traffic Load in a 500 meter buffer	2,917 (1,514)	
Traffic Load in a 1000 meter buffer		0.88 (0.36)
R^2 of model	0.13[b]	0.12
Number of road segments used for model development	2,964	2,336

[a]Regression slopes and standard error (between parentheses), multiplied by the difference between 10th and 90th percentile for all predictors to allow comparison of the effect of predictors with different units and distribution on measured concentrations. [b] R^2 of AR-1 model without AR-1 term.

The stationary models developed based on two seasons (Table S7) were very similar to the models previously published based on three seasons[9] (Table S8). The stationary models contained fewer but similar predictor variables than the mobile monitoring models.

3.2 Comparison between Short-term Stationary and Mobile Measurements

UFP measurements of both datasets were compared by creating a pseudo co-location of the stationary measurement. Of the possible 322 (161 x 2) stationary measurements, 184 also had a valid mobile measurement on the same road segment. Comparisons for UFP concentrations are given in table 3 and figure S1, showing a moderately high correlation between mobile and stationary measurements (r_p=0.48). Correlations were higher for urban background than for traffic sites (r_p=0.67 versus 0.39). Spearman correlations were higher than Pearson correlation values, indicating a non-linear relation between stationary and mobile measurements. Mobile measurements at the pseudo co-location were on average 1.12 times higher than stationary measurements, consistent with the comparison of all stationary sites and road segments (Table 1). However, the difference for the pseudo co-locations is much lower than the overall difference. Mobile measurements were thus not only higher because of the distance to the road but are also related to the car using relatively busy roads to travel between stationary sites. As seen in figure S1, the measurements were similar at the lower end of the concentration distributions and the difference increased at sites with high mobile measurements, hence the higher Spearman correlation compared to the Pearson correlation.

Table 3

Correlation between Mobile and Stationary UFP Measurements sampled at the same Road Segment.

↓

Sitetype	Number of observations	Pearson (Spearman)	Median Difference In particles/cm³ (Ratio)
Total	184	0.48 (0.70)	1,674 [a] (1.12)
Traffic sites	100	0.39 (0.56)	2,010 (1.12)
Urban Background sites	84	0.67 (0.81)	1,450 (1.11)

[a] Difference is mobile minus stationary measurement.

3.3 Comparison between Mobile LUR model Predictions and Short-term Stationary Measurements

Predictions of the mobile LUR models were compared to the stationary measurements (Figure 2a). The mobile UFP model explained 26% of the spatial variability, two times more than the R^2 of the mobile model itself (Table 2). The mobile model explained less variation of stationary measurements than the stationary model (Table S6) itself: 36%. For BC, the explained variance of stationary measurements was 20% for the mobile model (compared to 12% for the mobile BC model in Table 2) and 28% for stationary predictions respectively. In this comparison, it should be noted that the stationary models were developed on the stationary measurement data. Both mobile models predicted higher concentrations than the measured concentrations at the short-term stationary sites. Mean difference was 4,805 particles/cm³ (95% CI: 3,336; 6,251) for UFP and 1.11 µg/m³ (95% CI: 0.93; 1.28) for BC. Regressing the predicted UFP concentration based on the mobile measurements to the stationary measurements shows a slope of 0.70 (95% CI: 0.49; 0.91). For BC, the slope was 0.49 (95% CI: 0.32; 0.67).

3.4 Comparison between Mobile Measurements and Short-term Stationary LUR model predictions

We also compared the stationary LUR model predictions to the mobile road segment average measurements, which resulted in a similar explained variance as for the mobile LUR models itself (R^2 0.13 versus 0.13 and 0.11 versus 0.12 for UFP and BC, respectively). Results are shown in figure 2b. This comparison is somewhat hampered as an inverse distance is included in the stationary LUR model while all mobile measurements were taken on road. We therefore set a minimum distance of 10 meter to a major road and explored if varying this distance (4 to 15m) resulted in marked differences. The results of these sensitivity analyses were essentially similar to the model assuming a minimum distance of 10m.

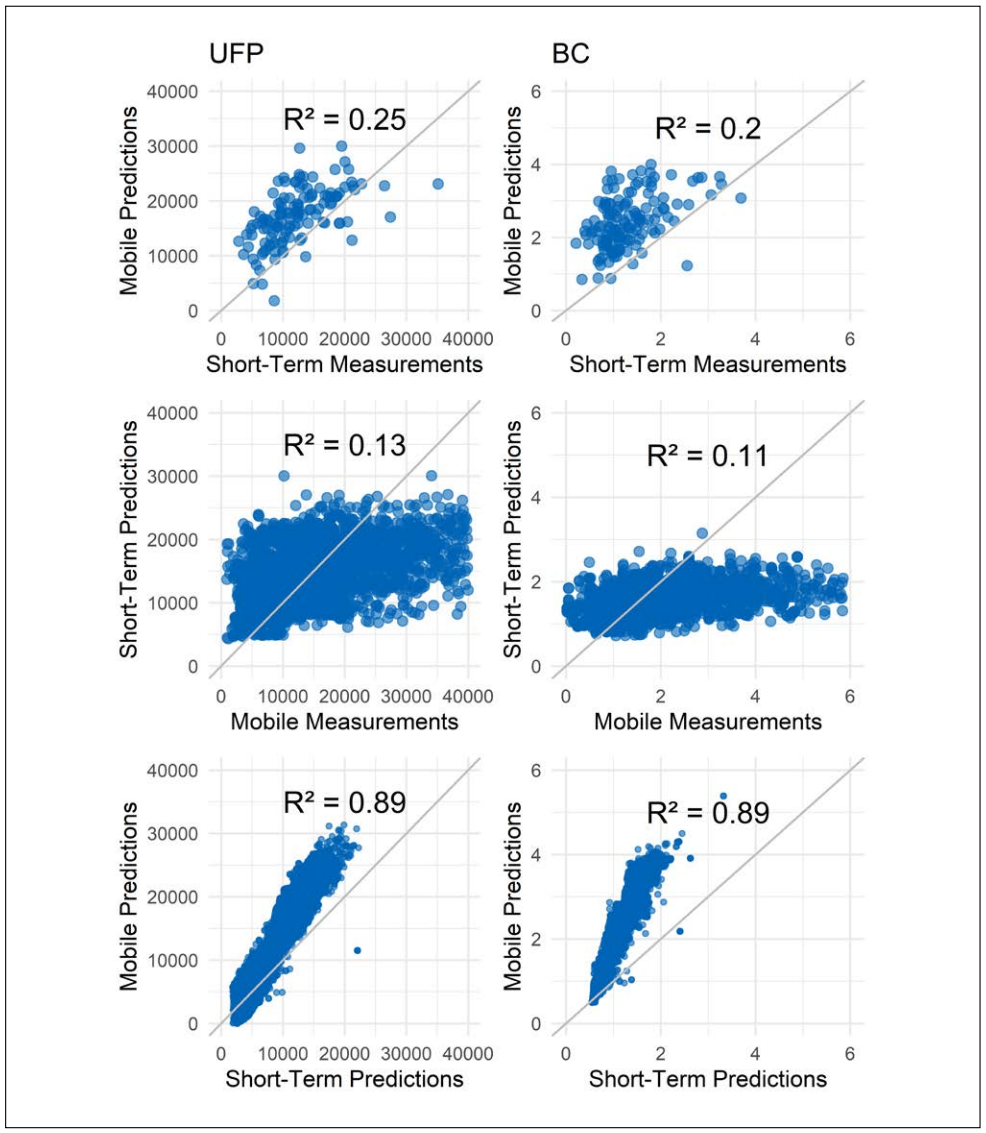
3.5 Comparison between Mobile and Short-term Stationary LUR Model predictions on External Dataset

The developed stationary and mobile LUR models were used to predict concentration levels at residences from an ongoing cohort study in Amsterdam, consisting of 12,682 residential addresses. Predictions at these addresses were very highly correlated for both sets of LUR models (Figure 2c, R^2 of 0.89 for UFP and 0.88 for BC). Mobile models did, however, predict higher concentrations than stationary models, especially at higher concentrations. This difference is larger for BC (factor 1.91) than for UFP (factor 1.41) and reflective of the overall difference observed in measured concentrations in the stationary and mobile datasets (factor 2.1 and 1.7, respectively). Mean absolute difference was 4,163 particles/cm³ (95% CI: 4,027; 4,299) for UFP and 1.08 µg/m³ (95% CI: 1.07; 1.10) for BC.

Figure 2

Comparison between Stationary and Mobile Measurements and Predictions for UFP and BC.

↓



(a) Predicted concentrations at stationary sites based on mobile model compared with stationary measurements. (b) Predicted concentrations at road segments based on the stationary model compared with mobile measurements. (c) Mobile versus stationary predicted UFP and BC concentrations at residences from an ongoing cohort study in Amsterdam, consisting of 12,682 residential addresses.

④ Discussion

We developed LUR models based on mobile and short-term stationary measurements and compared both models in their ability to predict UFP and BC concentration levels at the stationary measurement locations and at the external dataset with over 12,000 addresses in and around Amsterdam. Mobile monitoring LUR models included similar predictor variables compared to the short-term stationary models. Predictions of the mobile and stationary LUR models at residential addresses of the cohort study were very highly correlated (R^2 >0.88), but predictions based on the mobile model were on average 1.41 (UFP) and 1.91 (BC) times higher than predictions of the stationary model.

4.1 Mobile vs short-term stationary LUR Models

The variables in the mobile and stationary models were comparable. Both UFP models include the population density in a 5000 meter buffer and port area while the inverse distance to a major road variable in the stationary model is replaced by major road length in a 50 meter buffer in the mobile model. Traffic intensity in a 100 meter buffer is also very similar to the intensity of traffic on the nearest street, used in the mobile model. The mobile UFP model also included traffic load in a 500 meter buffer and the area of airports in 5000 meter buffer. The presence of an airport in a 5km buffer in the UFP model is consistent with recent studies evaluating the impact of airports on air pollutants[22,23]. These studies found elevated UFP (but not BC) concentrations up to 10km downwind of airports. In agreement with these previous studies, our LUR model for BC did not include airport. Due to the similar structure of both models, the predicted UFP concentrations at over 12,000 addresses were highly correlated with a R^2 value of 0.89.

The mobile BC model predictions were also highly correlated with the stationary model predictions on the external dataset (R^2 of 0.88). Variables in the BC mobile model were again similar to the stationary model variables. Here, population density and traffic intensity on the nearest road were present in both models and the variable describing the inverse distance to major road in the stationary model was replaced by major road length in a 50 meter buffer in the mobile model, as in the UFP model. These results indicate that both measurement approaches result in very similar model structures and highly correlated outdoor UFP and BC concentration estimates on a population level.

Explained variances of the mobile models were low and even lower than for the stationary models in this and our previous paper[9]. In a previous analysis,

it was documented that the short stationary monitoring duration resulted in much higher within (temporal) to between (spatial) sites concentration variance ratios compared to monitoring campaigns with 1 or 2-week average samples[9,18]. Mobile monitoring involves even shorter monitoring per road segment, and hence an even less favorable ratio between the within to between sites concentration variance ratios, resulting in low explained variances with the use of spatial predictor variables only. Consistently, the mobile UFP and BC models explained a larger percentage of the variation of the short-term stationary measurements than of the variability of the mobile measurements on which they were developed. We previously documented that, despite the low R^2, robust spatial models can be developed as the large number of sites in short-term stationary measurement campaigns compensates the low precision of the averages for each site[9]. The explanation offered was that measurement error in a continuous outcome variable in linear regression does not lead to biased regression slopes but does lead to lower R^2 values. In the present analyses we extend this observation to mobile monitoring showing that the mobile LUR models, on a population level, result in a similar rank-order of estimated outdoor concentrations, albeit that the mobile models systematically overestimated concentrations at residential addresses.

Consistent with the above reasoning and despite, the low mobile model R^2 values, variables in the model were able to predict differences of up to 6,997 particles/cm³ for UFP and 1.15 µg/m³ for BC between the 10^{th} and 90^{th} percentile of predictor variables. These are substantial contrasts, given average urban background values of about 10,000 particles/cm³ and 1.00 µg/m³ for UFP and BC, respectively.

Some mobile and short-term monitoring studies reported higher model R^2 values than we report here. In a bicycle-based mobile monitoring study in Minneapolis, models explained about 50% of the particle number concentration (PNC) and 30-40% of the measured BC variability[7]. Averages consisted of 12 to 30 runs on the same road segment, which is far more than our study in which most of the road segments were only sampled once. In studies by Weichenthal et al.[16,17] the average road segment consisted of 10 minutes of UFP data or consisted of at least 200 data points per segment, resulting in explained variances of 67% and 62%. Sabaliauskas et al.[15] only reported their R^2 of the mobile model compared with measurements at seven fixed sites to validate their model (R^2 of 0.68). Here, averaging time of UFP measurements was also between 5 and 10 minutes. Other differences between studies may be related to the overall variability of the measured concentrations, related to proximity to and magnitude of sources and the complexity of the study area.

4.2 Overestimation of mobile models

An average of 1.41 (UFP) and 1.91 (BC) times higher predicted values for mobile models compared to stationary models was found on residential addresses of the external dataset. The overestimation is likely caused by the fact that mobile measurements are on-road while the stationary measurements were taken at the sidewalks as close as possible to the facade of buildings. While monitoring studies have documented large gradients of UFP and BC within meters of major roads[3], our models contained buffer variables for traffic with radii of 50 to 1000m and these were unable to sufficiently explain differences between concentrations from on road to residential addresses located at typically 5 – 20 m from the side of the road. We excluded inverse distance to major roads from modelling, as this variable cannot be calculated for on-road monitoring on major roads (distance zero). Assigning small distances to the road for on-road measurements was not an option, as the distance chosen is arbitrary and if chosen small essentially results in a dummy variable for major roads versus non-major roads. In addition, it is unlikely that distance variables would be able to provide the needed scaling, given the limited precision of GIS calculated distances in compact urban areas[8].

The higher overestimation for BC compared to UFP can possibly be ascribed to the coarser time resolution of the BC instrument (1 minute versus 1 second for UFP) and the used data aggregation method. BC measurements were averaged over multiple minutes when the attenuation change was too low, while road segments with high concentration levels are more likely to remain on one-minute resolution. Road segments with expected low concentrations are therefore assigned relative higher concentration levels. This method produces higher parameter estimates when modelling BC concentrations opposed to UFP. When UFP values were averaged with the same approach as the BC measurements (based on ATN values of the BC instrument), the absolute difference increased to 1.67 (Figure S2). Due to the precision of the instrument and data aggregation method, it seems that mobile monitoring while driving a car limits the application of our BC device. For mobile monitoring campaigns at lower driving speeds based for example upon cycling and especially walking, the 1-minute resolution may be appropriate. In study areas with higher concentrations than observed in our study, the instrument will support shorter duration measurements.

4.3 Mobile versus Stationary Monitoring

The advantage of mobile monitoring is that many locations can be measured in a relatively short amount of time with a limited number of

monitoring devices. These locations include more complex, but realistic locations, such as near intersections[9]. An additional advantage is that little preparation is needed for site selection because the mobile platform does not need to be stationed anywhere for 30 minutes. However, selecting monitoring routes to cover relevant spatial settings and avoiding bias due to temporal variation remains important. We tried to minimize temporal bias by restricting to measurements outside the rush hours and by having a background reference site with identical equipment. Other approaches that could be taken is by smart-driving patterns in which several locations are re-visited during the day or by having more than one platform driving at the same time.

Mobile monitoring may be affected by wind and vehicle speed effects, though the impact may be limited for submicrometer particles[26]. Figure S1 suggests this has not been a main issue in our study, as stationary and nearby mobile measurements did not differ substantially. Furthermore, we did not find a correlation between driving speed and measured concentrations in our dataset.

One of the methodological challenges when using mobile data is to account for the inherent autocorrelation structure in the data. We used a first order autoregressive model for the residuals, which assumes regular space and time intervals and that autocorrelation remains constant over time. However, as our measurements were not taken with the same lag-time in between road segments this method is unlikely to be optimal. We therefore performed sensitivity analyses where we did not account for the autocorrelation. This indicated that the correlations on a population level between the models accounting or not accounting for autocorrelation are high: 0.99 for UFP and 0.97 for BC (Table S4 and S6).

We aggregated the monitoring data to road segments and used the midpoint to obtain GIS predictor variables. This adds some uncertainty to the analysis, but in a study in Minneapolis, Hankey and co-workers found no difference in model performance for aggregation at 50, 100 or 200m spatial resolution[7].

The exclusion of road segments with local exhaust plumes for model development did not affect LUR models much in our study. An argument against removing local exhaust plumes of high emitting vehicles is that busy road segments have a higher frequency of local exhaust plumes. In our modelling approach, we checked for the influence of potential outliers using Cook's D statistic. The large number of road segments in the dataset probably prevented that the highest values influenced the developed model. Figure 1 illustrates that we measured few extreme concentrations. Consistently, only about 2% of our observations were flagged as local exhaust plume by our algorithm. In other study areas, this may be different.

An interesting option for model development is the combination of mobile and short-term stationary monitoring. This hybrid approach needs further work, likely involving definition of weights to take into account, the number of sites, the time at each site and allowing for heterogeneity in variance structure. A further possible development is to build spatiotemporal models incorporating meteorology.

We focused our paper on the use of mobile and short-term stationary monitoring LUR models to characterize residential exposures for cohort studies. Mobile on-road monitoring models may also provide useful information to characterize commuter exposures. When coupled with time activity information, e.g. tracking by smartphone, individual exposure can be calculated incorporating both residential and commuter exposures[27].

4.4 Implications for epidemiology

The spatial models that we developed may be useful for long-term exposure studies. On a population level the predictions of the mobile and short-term model were highly correlated, implying that significant associations with health observed with one model would be detected with the other model as well. The overestimation of mobile models does not necessarily induce biased risk estimates in epidemiological studies, if there is a systematic overestimation of the concentration predictions. Figure 2c however suggests that the relationship between the two model predictions is better described with a ratio than with a constant difference. Contrasts in exposure are higher for the mobile than for stationary monitoring models, possibly leading to lower effect estimates per unit of exposure in epidemiological studies for the mobile models.

In addition, the overestimation of the mobile models was only modest; 30% for UFP and 50% for BC. Such differences can also be produced by measuring in one season only; excluding or including rush hour traffic and measuring with different sampling devices. As such, the overestimation observed here is not likely to be an important factor. Our study thus suggests that mobile models can be used to predict exposures in epidemiological studies, taking into account that predictions on absolute level may not reflect residential exposures fully.

References

(1) Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y.; Luepker, R. V.; Mittleman, M. A.; et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the american heart association. *Circulation* **2010**, *121* (21), 2331–2378.

(2) Delfino, R. J.; Sioutas, C.; Malik, S. Potential role of ultrafine particles in associations between airborne particle mass and cardiovascular health. *Environ. Health Perspect.* **2005**, *113* (8), 934–946.

(3) HEI Review Panel. Understanding the Health Effects of Ambient Ultrafine Particles. *Heal. Eff. Inst.* **2013**, No. January, 122.

(4) Janssen, N. A. H.; Hoek, G.; Simic-Lawson, M.; Fischer, P.; van Bree, L.; ten Brink, H.; Keuken, M.; Atkinson, R. W.; Anderson, H. R.; Brunekreef, B.; et al. Black Carbon as an Additional Indicator of the Adverse Health Effects of Airborne Particles Compared with PM_{10} and $PM_{2.5}$. *Environ. Health Perspect.* **2011**, *119* (12), 1691–1699.

(5) Zwack, L. M.; Hanna, S. R.; Spengler, J. D.; Levy, J. I. Using advanced dispersion models and mobile monitoring to characterize spatial patterns of ultrafine particles in an urban area. *Atmos. Environ.* **2011**, *45* (28), 4822–4829.

(6) Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42* (33), 7561–7578.

(7) Hankey, S.; Marshall, J. D. Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, $PM_{2.5}$, Particle Size) Using Mobile Monitoring. *Environ. Sci. Technol.* **2015**, *49*, 9194–9202.

(8) Hoek, G.; Beelen, R.; Kos, G.; Dijkema, M.; van der Zee, S. C.; Fischer, P. H.; Brunekreef, B. Land use regression model for ultrafine particles in Amsterdam. *Environ. Sci. Technol.* **2011**, *45* (2), 622–628.

(9) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712–8720.

(10) Abernethy, R. C.; Allen, R. W.; Mckendry, I. G.; Brauer, M. A Land Use Regression Model for Ultra fine Particles in Vancouver, Canada. *Environ. Sci. Technol.* **2013**, *47* (10), 5217–5225.

(11) Rivera, M.; Basagana, X.; Aguilera, I.; Agis, D.; Bouso, L.; Foraster, M.; Medina-Ramón, M.; Pey, J.; Künzli, N.; Hoek, G. Spatial distribution of ultrafine particles in urban settings: A land use regression model. *Atmos. Environ.* **2012**, *54*, 657–666.

(12) Saraswat, A.; Apte, J. S.; Kandlikar, M.; Brauer, M.; Henderson, S. B.; Marshall, J. D.; Vt, B. C.; Apte, J. S.; Brauer, M.; Henderson, S. B.; et al. Spatiotemporal land use regression models of fine, ultrafine, and black carbon particulate matter in New Delhi, India. *Environ. Sci. Technol.* **2013**, *47* (22), 12903–12911.

(13) Larson, T.; Henderson, S. B.; Brauer, M. Mobile monitoring of particle light absorption coefficient in an urban area as a basis for land use regression. *Environ. Sci. Technol.* **2009**, *43* (13), 4672–4678.

(14) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An hourly regression model for ultrafine particles in a near-highway urban area. *Environ. Sci. Technol.* **2014**, *48* (6), 3272–3280.

(15) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Reali, C.; Sun, T.; Evans, G. J. Development of a land-use regression model for ultrafine particles in Toronto, Canada. *Atmos. Environ.* **2015**, *110*, 84–92.

(16) Weichenthal, S.; Van Ryswyk, K.; Goldstein, A.; Shekarrizfard, M.; Hatzopoulou, M. Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: A land use regression model. *Environ. Pollut.* **2016**, *208*, 241–248.

(17) Weichenthal, S.; Ryswyk, K. Van; Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A land use regression model for ambient ultrafine particles in Montreal, Canada: A comparison of linear regression and a machine learning approach. *Environ. Res.* **2016**, *146*, 65–72.

(18) Klompmaker, J.; Montagne, D.; Meliefste, K.; Hoek, G.; Brunekreef, B. Spatial variation of ultrafine particles and black carbon in two cities: results from a short-term measurement campaign. *Sci. Total Environ.* **2015**, *508*, 266–275.

(19) Brantley, H. L.; Hagler, G. S. W.; Kimbrough, E. S.; Williams, R. W.; Mukerjee, S.; Neas, L. M. Mobile air monitoring data-processing strategies and effects on spatial air pollution trends. *Atmos. Meas. Tech.* **2014**, *7* (7), 2169–2183.

(20) Drewnick, F.; Böttger, T.; Von Der Weiden-Reinmüller, S. L.; Zorn, S. R.; Klimach, T.; Schneider, J.; Borrmann, S. Design of a mobile aerosol research laboratory and data processing tools for effective stationary and mobile field measurements. *Atmos. Meas. Tech.* **2012**, *5* (6), 1443–1457.

(21) Hagler, G. S. W.; Yelverton, T. L. B.; Vedantham, R.; Hansen, A. D. A.; Turner, J. R. Post-processing method to reduce noise while preserving high time resolution in aethalometer real-time black carbon data. *Aerosol Air Qual. Res.* **2011**, *11* (5), 539–546.

(22) Keuken, M. P.; Moerman, M.; Zandveld, P.; Henzing, J. S.; Hoek, G. Total and size-resolved particle number and black carbon concentrations in urban areas near Schiphol airport (the Netherlands). *Atmos. Environ.* **2015**, *104*, 132–142.

(23) Hudda, N.; Gould, T.; Hartin, K.; Larson, T. V.; Fruin, S. A. Emissions from an international airport increase particle number concentrations 4-fold at 10 km downwind. *Environ. Sci. Technol.* **2014**, *48* (12), 6628–6635.

(24) Farrell, W.; Weichenthal, S.; Goldberg, M.; Valois, M. F.; Shekarrizfard, M.; Hatzopoulou, M. Near roadway air pollution across a spatially extensive road and cycling network. *Environ. Pollut.* **2016**, *212*, 498–507.

(25) Van Eijsden, M.; Vrijkotte, T. G. M.; Gemke, R. J. B. J.; Van der Wal, M. F. Cohort profile: The Amsterdam born children and their development (ABCD) study. *Int. J. Epidemiol.* **2011**, *40* (5), 1176–1186.

(26) Morawska, L.; Salthammer, T. *Indoor environment: Airborne particles and settled dust*; 2006.

(27) De Nazelle, A.; Seto, E.; Donaire-Gonzalez, D.; Mendez, M.; Matamala, J.; Nieuwenhuijsen, M. J.; Jerrett, M. Improving estimates of air pollution exposure through ubiquitous sensing technologies. *Environ. Pollut.* **2013**, *176*, 92–99.

Supplement information for

Comparison of Ultrafine Particles and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model

Table S1

Spatial predictor variables with units, a priori defined directions of effect and buffer sizes.

↓

Predictor variable	Variable name	Units	Direction	Buffer (m)
Industry	INDUSTRY	m^2	+	100, 300, 500, 1000, 5000
Port	PORT	m^2	+	100, 300, 500, 1000, 5000
Airport	AIRPORT	m^2	+	5000, 10000
Urban green	URBGREEN	m^2	-	100, 300, 500, 1000, 5000
Semi-natural and forested areas	NATURAL	m^2	-	100, 300, 500, 1000, 5000
Population data	POPEEA	m^2	+	100, 300, 500, 1000, 5000
Traffic intensity on nearest road	TRAFNEAR	$Veh.day^{-1}$	+	
Distance to nearest road [a]	DISTINVNEAR1, DISTINVNEAR2	m^2	+	
Product of traffic intensity on nearest road and inverse distance to the nearest road and inverse distance squared [a]	INTINVDIST, INTINVDIST2	$Veh.day^{-1}m^{-1}$, $Veh.day^{-1}m^{-2}$	+	
Traffic intensity on nearest major road	TRAFMAJOR	$Veh.day^{-1}$	+	
Distance to nearest major road [a]	DISTINVMAJOR1, DISTINVMAJOR2	m^2	+	
Product of traffic intensity on nearest major road and inverse of distance to the nearest major road and distance squared [a]	INTMAJORINVDIST, INTMAJORINVDIST2	$Veh.day^{-1}m^{-1}$, $Veh.day^{-1}m^{-2}$	+	
Total traffic load of major roads in a buffer (sum of (traffic intensity* length of all segments))	TRAFMAJORLOAD	$Veh.day^{-1}m$	+	25, 50, 100, 300, 500, 1000
Total traffic load of roads in a buffer (sum of (traffic intensity * length of all segments))	TRAFLOAD	$Veh.day^{-1}m$	+	25, 50, 100, 300, 500, 1000
Heavy-duty traffic intensity on nearest road	HEAVYTRAFNEAR	$Veh.day^{-1}$	+	
Product of heavy-duty traffic intensity on nearest road and inverse of distance to the nearest road and distance squared	HEAVYINTINVDIST, HEAVYINTINVDIST2	$Veh.day^{-1}m$, $Veh.day^{-2}m$	+	
Heavy-duty traffic intensity on nearest major road	HEAVYTRAFMAJOR	$Veh.day^{-1}$	+	
Total heavy-duty traffic load of major roads in a buffer (sum of (heavy-duty traffic intensity*length of all segments)	HEAVYTRAFMAJORLOAD	$Veh.day^{-1}m$	+	25, 50, 100, 300, 500, 1000
Total heavy-duty traffic load of all roads in a buffer (sum of (heavy-duty traffic intensity* length of all segments))	HEAVYTRAFLOAD	$Veh.day^{-1}m$	+	25, 50, 100, 300, 500, 1000
Road length of all roads in a buffer	ROADLENGTH	m	+	25, 50, 100, 300, 500, 1000
Road length of all major roads in a buffer	MAJORROADLENGTH	m	+	25, 50, 100, 300, 500, 1000
Distance to nearest road, inverse distance (m^{-1}) and inverse squared distance (m^{-2}) [a]	DISTINVNEARC1, DISTINVNEARC2	m^{-1}/m^{-2}	+	
Distance to nearest major road [a]	DISTINVMAJORC1, DISTINVMAJORC2	m^{-1}/m^{-2}	+	

[a] Variables were not used for mobile model development, due to values being zero.

Table S2

Land-Use Regression Models based upon Mobile Measurements for UFP and BC with at least two repeats.

↓

Variables in LUR model	UFP (in particles/cm³)	BC (in µg/m³)
Intercept	557 (2,104))	0.71 (0.26)
Population Density in a 5000 meter buffer	10,091 (1,607)	
Airport Area in a 5000 meter buffer	5,085 (978)	0.83 (0.12)
Number of households in a 5000 meter buffer		0.67 (0.23)
Nature Area in 5000 meter buffer		
Major Road Length in a 100 meter buffer	6,885 (1,340)	0.83 (0.18)
Heavy Traffic Intensity on the Nearest Road	2,556 (1,288)	0.85 (0.17)
Heavy Traffic Load on Major Roads in a 500 meter buffer	5,047 (1,248)	
Traffic Load on Major Roads in a 500 meter buffer		0.59 (0.20)
Traffic Load in a 1000 meter buffer		
R^2 of model	0.18	0.30
Number of road segments used for model development	745	514

[a] Regression slopes and standard error (between parentheses), multiplied by the difference between 10th and 90th percentile for all predictors to allow comparison of the effect of predictors with different units and distribution on measured concentrations.

Table S3

Land-Use Regression models using mobile UFP measurements (in particles/cm³).

↓

	With Local Exhaust Plumes		Without Local Exhaust Plumes	
Variables in LUR model	**Linear Regression**	**AR-1 Model**	**Linear Regression**	**AR-1 Model**
Intercept	2381 (1746)	5656 (2675)	1414 (1532)	1244 (2007)
Population Density in a 5000 meter buffer	7894 (1212)	8064 (1947)	8833 (1021)	9394 (1587)
Airport Area in a 5000 meter buffer	4703 (728)	4669 (1185)	5329 (663)	5240 (1077)
Port Area in a 1000 meter buffer	2155 (838)	2499 (1248)	2070 (730)	2007 (1101)
Port Area in a 5000 meter buffer	2837 (1094)		3002 (959)	3116 (1500)
Nature Area in 5000 meter buffer	-1830 (851)	-2557 (1357)	-1362 (745)	
Major Road Length in a 50 meter buffer	8540 (943)	6868 (1071)	6994 (827)	6186 (847)
Traffic Load on Major Roads in a 100 meter buffer	3562 (1055)	1928 (1095)	2780 (1018)	
Traffic Intensity on the Nearest Road			2387 (969)	3240 (825)
Traffic Load in a 500 meter buffer	2514 (1159)	2917 (1514)		
R^2 of model	0.14	0.13	0.16	0.16
Number of sites used for model development	2,964	2,964	2,907	2,907

Different methods were explored for developing mobile LUR models. Mobile models were generated with and without local exhaust plumes, shown as linear regression models. For both methods, an AR-1 term was added to the model to correct for autocorrelation. [a] Regression slopes were multiplied by the difference between 10th and 90th percentile for all predictors. Standard error between brackets.

Table S4

Comparison between predicted UFP concentrations at 12,682 residential addresses from a cohort in Amsterdam, based on different specifications of the LUR model. (in particles/cm³).

↓

Dataset	N	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile	Pearson correlation with main model (With peaks and AR-1)
Stationary Short-Term Model	12,682	3,882	6,153	10,657	13,339	16,217	0.94
Model With Peaks	12,682	1,288	6,287	14,299	18,105	23,254	0.99
Model With Peaks and AR-1	12,682	2,531	8,494	15,629	19,318	24,247	/
Model Without Peaks	12,682	1,472	5,298	13,622	17,008	22,138	0.98
Model Without Peaks and AR-1	12,682	4,082	7,091	14,572	17,692	22,019	0.96

Different methods were explored for developing mobile LUR models. Mobile models were generated with and without local exhaust plumes, indicated by with and without peaks. For both methods, an AR-1 term was added to the model to correct for autocorrelation.

Table S5

Land-Use Regression models using mobile BC measurements (in µg/m³).

↓

	With Local Exhaust Plumes		Without Local Exhaust Plumes	
Variables in LUR model	**Linear Regression**	**AR-1 Model**	**Linear Regression**	**AR-1 Model**
Intercept	-1.22 (0.37)	0.48 (0.60)	-0.52 (0.23)	-0.02 (0.54)
Major Road Length in a 50 meter buffer	0.95 (0.18) [a]	0.61 (0.15)	0.85 (0.15)	0.29 (0.12)
Population Density in a 5000 meter buffer	1.01 (0.25)	1.15 (0.48)	1.34 (0.19)	0.84 (0.43)
Road Length in a 1000 meter buffer	0.65 (0.23)			
Traffic Load in a 100 meter buffer	0.45 (0.23)			
Traffic Load in a 1000 meter buffer	0.86 (0.23)	0.88 (0.36)	0.94 (0.19)	1.74 (0.28)
Traffic Intensity on the Nearest Road	0.73 (0.20)	0.30 (0.14	0.89 (0.15)	
R^2 of model	0.13	0.12	0.12	0.12
Number of sites used for model development	2336	2336	2234	2234

Different methods were explored for developing mobile LUR models. Mobile models were generated with and without local exhaust plumes, shown as linear regression models. For both methods an AR-1 term was added to the model to correct for autocorrelation.
[a] Regression slopes were multiplied by the difference between 10th and 90th percentile for all predictors. Standard error between brackets.

Table S6

Comparison between predicted BC concentrations at 12,682 residential addresses from a cohort in Amsterdam, based on different specifications of the LUR model (in µg/m³).

↓

Dataset	N	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile	Pearson correlation with main model (With peaks and AR-1)
Stationary Short-Term Model	12,682	0.71	0.86	1.09	1.33	1.50	0.93
Model With Peaks	12,682	-0.16	0.79	1.48	2.53	3.45	0.97
Model With Peaks and AR-1	12,682	0.94	1.51	2.00	2.87	3.44	/
Model Without Peaks	12,682	0.01	0.66	1.24	2.24	2.91	0.99
Model Without Peaks and AR-1	12,682	0.41	1.09	1.65	2.72	3.53	0.98

Different methods were explored for developing mobile LUR models. Mobile models were generated with and without local exhaust plumes, indicated by with and without peaks. For both methods an AR-1 term was added to the model to correct for autocorrelation.

Table S7

LUR Models for UFP and BC based upon stationary measurements.

↓

Variables in LUR model	UFP (in particles/cm³)	BC (in µg/m³)
Intercept	1807 (1965)	0.54 (0.21)
Inverse Distance to Major Road	5403 (1047) [a]	0.53 (0.13)
Population Density in 5000 meter buffer	4886 (1364)	0.41 (0.16)
Port Area in an 5000 meter buffer	2238 (1299)	
Airport Area in an 5000 meter buffer	612 (204)	
Traffic Load in 100 meter buffer	1722 (863)	
Traffic Intensity on the nearest street		0.21 (0.13)
R^2 of model	0.36	0.28
Number of sites used for model development	128	141

Stationary models were developed with the average of 30 minute measurements in two seasons. The old stationary model, based on three seasons, which has been published before is shown in table S7.
[a] Regression slopes were multiplied by the difference between 10th and 90th percentile for all predictors. Standard error between brackets.

Table S8

Previously published LUR Models for UFP and BC based upon stationary measurements, based on three seasons.

↓

Variables in LUR model	UFP (in particles/cm³)	BC (in µg/m³)
Intercept	3,221	0.54
Inverse Distance to Major Road	4,552 [a]	0.52
Population Density in 5000 meter buffer	3,959	0.37
Port Area in an 5000 meter buffer	2,255	
Traffic Load in 100 meter buffer	1,740	
Traffic Intensity on the nearest street		0.30
R^2 of model	0.37	0.35
Number of sites used for model development	159	160

The old stationary model based on three seasons which has been published before[9]. Airport data was not included in the paper by Montagne et al[9].
[a] Regression slopes were multiplied by the difference between 10th and 90th percentile for all predictors.
Standard error between brackets.

Figure S1

Stationary versus mobile UFP measurements on the same road segment (n=184). Concentration levels in particles/cm³.

↓

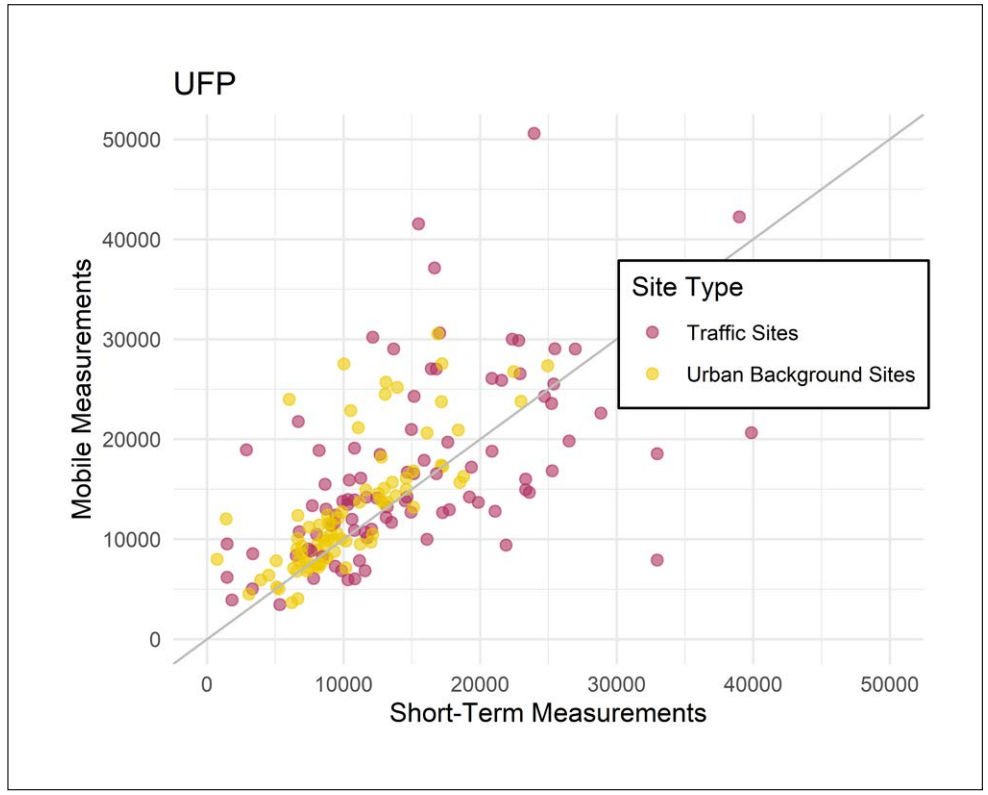
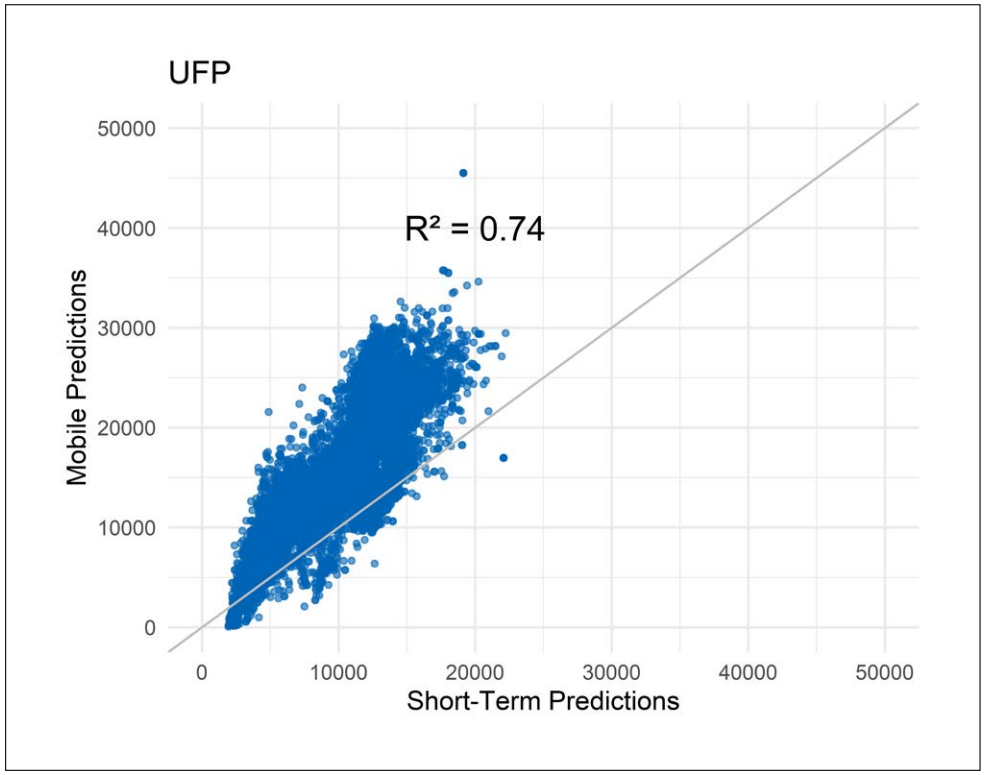


Figure S2

Mobile versus Stationary predicted concentrations on an external dataset for UFP averages based on the same averaging method as the BC instrument.

↓



Mobile predictions are based on UFP road segment averages that were calculated with the same method as the BC instrument, so based on the minimal attenuation change of 0.05 of the BC device.

Robustness of Intra Urban Land-Use Regression Models for Ultrafine Particles and Black Carbon based on Mobile Monitoring

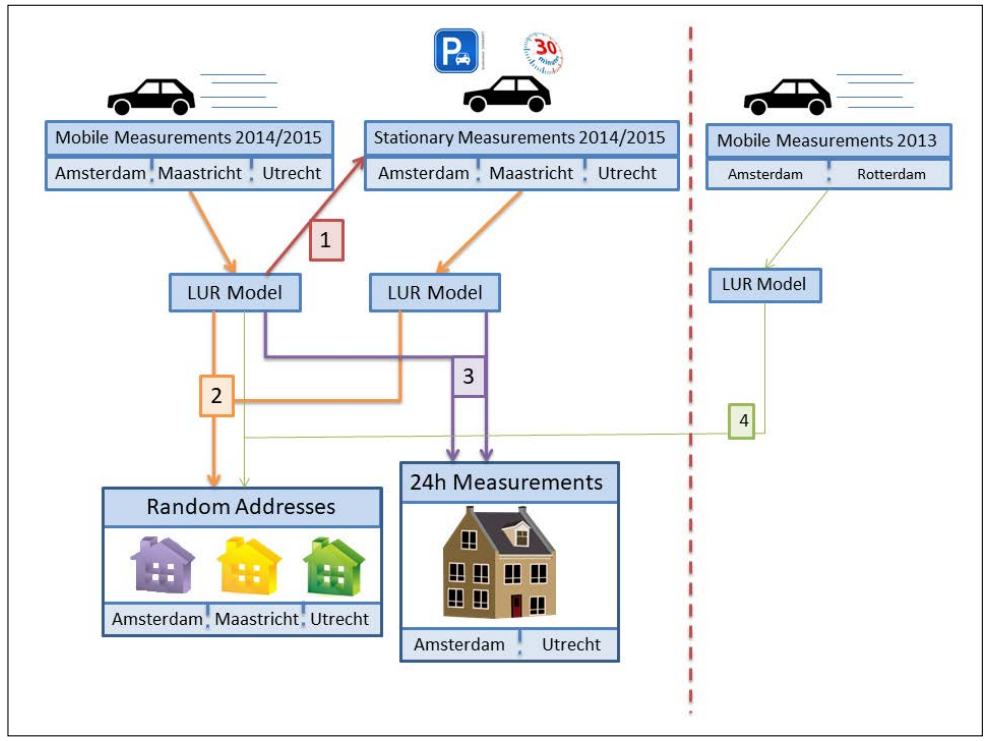
Jules Kerckhoffs
Gerard Hoek
Jelle Vlaanderen
Erik van Nunen
Kyle Messier
Bert Brunekreef
John Gulliver
Roel Vermeulen

Abstract

Land-use regression (LUR) models for ultrafine particles (UFP) and Black Carbon (BC) in urban areas have been developed using short-term stationary monitoring or mobile platforms in order to capture the high variability of these pollutants. However, little is known about the comparability of predictions of mobile and short-term stationary models and especially the validity of these models for assessing residential exposures and the robustness of model predictions developed in different campaigns.

We used an electric car to collect mobile measurements (n=5236 unique road segments) and short-term stationary measurements (3x30min, n=240) of UFP and BC in three Dutch cities (Amsterdam, Utrecht, Maastricht) in 2014-2015. Predictions of LUR models based on mobile measurements were compared to (i) measured concentrations at the short-term stationary sites, (ii) LUR model predictions based on short-term stationary measurements at 1500 random addresses in the three cities, (iii) externally obtained home outdoor measurements (3x24hour samples; n=42) and (iv) predictions of a LUR model developed based upon a 2013 mobile campaign in two cities (Amsterdam, Rotterdam).

Despite the poor model R^2 of 15%, the ability of mobile UFP models to predict measurements with longer averaging time increased substantially from 36% for short-term stationary measurements to 57% for home outdoor measurements. In contrast, the mobile BC model only predicted 14% of the variation in the short-term stationary sites and also 14% of the home outdoor sites. Models based upon mobile and short-term stationary monitoring provided fairly high correlated predictions of UFP concentrations at 1500 randomly selected addresses in the three Dutch cities (R^2 =0.64). We found higher UFP predictions (of about 30%) based on mobile models opposed to short-term model predictions and home outdoor measurements with no clear geospatial patterns. The mobile model for UFP was stable over different settings as the model predicted concentration levels highly correlated to predictions made by a previously developed LUR model with another spatial extent and in a different year at the 1500 random addresses (R^2=0.80). In conclusion, mobile monitoring provided robust LUR models for UFP, valid to use in epidemiological studies.



① Introduction

Traffic is considered a major source of intra-urban air pollution[1,2]. Multiple studies have linked traffic proximity and traffic related air pollution to increased risks of adverse health effects[3,4]. With about 75% of the population living in urban environments in Europe[5], it is important to characterise intra-urban air pollution with high spatial-resolution, especially for primary pollutants that exhibit large spatial variability within city limits such as ultrafine particles (UFP) and black carbon (BC)[1,6,7]. UFP and BC measurements are therefore increasingly performed with densely distributed networks or mobile platforms. Mobile monitoring provides the possibility to sample more spatially diverse environments in less time, with a limited number of monitoring devices. This is cost-effective and especially within city limits, it can capture the high variability of UFP and BC in a complex urban terrain[8,9].

Several land use regression (LUR) models for UFP and BC have been developed using mobile measurements in North America[10–16] and Europe[17,18], with promising results for effective exposure assessment. Mobile monitoring campaigns that developed LUR models used bikes[10,11], cars[12,13,16,18], public transport[17] or walking with backpacks[14,15] to collect their data. In a previous study, we developed UFP and BC models based on mobile measurements and found a high correlation (R^2 ~ 0.88) of model predictions with LUR models based on short-term stationary measurements (30min) from a combined (mobile and stationary) measurement campaign in two cities in The Netherlands[18]. The mobile model for UFP and BC did predict substantially (30-50%) higher concentrations than the short-term stationary model.

Although these results were encouraging for the application of LUR models based on mobile monitoring campaigns in epidemiological research some questions remain. First, we want to confirm our previous observation of high correlation of mobile versus short-term models in a new campaign involving additional cities in a different year. Second, in contrast to our previous study we added home outdoor measurements (3 times 24 hours) allowing an unbiased comparison of the validity of both approaches. Third, we address the systematic difference in predicted concentration levels between mobile and short-term stationary models by exploring several methodologies to try to correct for this systematic difference. Fourth, we were interested if the derived LUR models are stable over space and time by comparing models derived from two independent measurements campaigns performed in 2013 and 2014/2015.

② Methods

2.1 Study design

We used five different sets of data as can be seen in the TOC Art and supporting information table A.1. Four of them (on the left of the red dotted line) were collected and retrieved from the EXPOsOMICS campaign, conducted in 2014/2015. Mobile measurements from the MUSiC campaign in 2013 (right side) were used in additional analyses. The MUSiC measurements and models have been extensively described in previous publications[18–20]. Data from the EXPOsOMICS campaign[21] in the Netherlands consists of mobile, short-term stationary, and home outdoor 24h air pollution measurements. The study design and models, based upon short-term stationary monitoring in six study areas including the Netherlands, have been reported before[22].

We gathered mobile measurements between short-term stationary measurements (30 min) when driving from one site to the next; 240 short-term stationary sites and 5,236 unique road segments were sampled in the winter, spring and summer in 2014/2015. Measurements were about equally divided over 84 days and started after 9:15AM and stopped before 4:00PM. About 8 short-term sites were sampled each day over 8-10 routes per city and per season. This way, we captured the within-day, day-to-day and seasonal variability of UFP and BC concentration levels[23]. Rush hour traffic was avoided for better comparability between road segments. Short-term stationary sites were selected with a wide range of traffic characteristics and land use in and around the cities of Amsterdam, Utrecht and Maastricht, The Netherlands. We selected traffic sites (>10,000 vehicles per day[24]), urban background sites, industrial areas, sites near urban green, regional background sites and sites near rivers or canals[22]. In further comparisons between traffic sites and urban background sites, all sites that are not traffic sites are considered urban background sites.

Short-term stationary and on-road measurements were made using an electric vehicle (REVA, Mahindra Reva Electric Vehicles Pvt. Ltd., Bangalore, India). A condensation particle counter (TSI, CPC 3007) and a micro Aethalometer (Aethlabs, CA, USA) were used to monitor UFP and BC concentrations respectively. The CPC had a measurement every second, whereas the Aethalometer averaged measurements over one minute. The geographical location of the electric car was recorded using a Global Positioning Unit (GPS, Garmin eTrex Vista) and linked to the instruments in the car based on date and time.

To compare the predictions of UFP and BC exposure from mobile and short-term LUR models in the general population we used 1500 randomly selected addresses equally divided between Amsterdam, Utrecht and Maastricht. Furthermore, three temporally adjusted 24-hour measurements of UFP and $PM_{2.5}$ absorbance (as a proxy for BC) were performed at home (outdoor) addresses at 42 locations in Utrecht and Amsterdam, according to protocols described by van Nunen et al[22] and Eeftens et al.[25] UFP measurements were monitored using MiniDiSCs (Testo AG, Lenzkirch, Germany) which sampled every second. Previous studies have shown good agreement between CPCs and MiniDiSCs with limited differences in absolute values[26,27]. $PM_{2.5}$ absorbance samples were measured using Harvard Impactors and were found to be highly correlated with black carbon[25].

These external addresses are referred to as “home outdoor sites” and used to compare LUR estimates at the home location from the mobile and short-term stationary LUR models (external validation).

2.2 Data Aggregation

Following our previous mobile monitoring measurement campaign[18], we corrected for small spatial errors of the GPS by assigning all GPS points to the nearest road they were supposed to be on. Then we calculated average concentration levels of UFP per road segment, defined as a part of a road between two consecutive intersections[11,12,15]. Road segments in tunnels or on bridges were deleted from the dataset, as they are not representative for concentrations at residential addresses. Road segments were on average 110 meters long and accumulated 25 seconds of UFP data over the study period.

BC concentrations were sampled at a one-minute interval, but this is often too short to detect reliable changes in concentration levels[18,28]. To reduce the noise of the instrument Hagler et al.[28] proposed a method to only assign minute averages when the attenuation value of the filter in the instrument increased sufficiently. In our campaign this meant that about one measurement was obtained every two or three minutes. So, minute values with a too small change in attenuation (>75% of the values) were averaged over time until the criteria was met. These values were then assigned to every road segment the car was on in that period (on average 7 road segments, ~ 140 sec). When the BC measurement changed during a road segment, an average was calculated.

2.3 Data Processing

UFP values of 500 particles/cm³ or less were removed from the data set, as these reflect malfunctioning of the instrument. If the UFP data increased or decreased in one second by a factor 10 or more, the data was removed as well. Both criteria were used in previous studies[18–20] and resulted in less than 1% removal of UFP data. We defined observations during mobile monitoring influenced by local exhaust plumes if UFP concentration was three standard deviations above the previous measurement second, based on the concentrations distribution for that day. Observations remained flagged until they dropped beneath the day average plus one standard deviation. This is based on methods used by Drewnick et al.[29] and Ranasinghe et al[30]. For the main analyses we used all measurements, including road segments with local exhaust plumes. For a sensitivity analysis, we excluded them.

2.4 Temporal Variation

A reference site with the same equipment as the electric vehicle and the home outdoor measurement sites was set up near Utrecht (about 2km outside the city border of Utrecht, 40km to Amsterdam and 140km to Maastricht), The Netherlands, to correct for temporal variation. We used the difference method for correcting the spatial data, following previous work in the stationary campaign[22] and the previous mobile monitoring campaign[18]. First, the overall mean concentration of the entire campaign at the reference site was calculated. Next, for each minute at the reference site an average of 30 minutes around time x was calculated which was subtracted from the overall mean concentration at the reference site. The difference is then used to adjust the original concentration measured at the sampling locations. We co-located instruments when the instruments were transferred between cities to check comparability and found a median ratio (averaged over 1 minute) for the CPCs of 1.09 (SD=0.16) and 0.98 (SD=0.63) for the Aethalometers.

2.5 Model Development

In accordance with our previous and most other mobile monitoring studies[11,12,15,18], we identified the middle of each road segment and used this coordinate to acquire GIS predictors for LUR modelling (overview of GIS predictors see Table A.2). In summary, a range of traffic variables was defined, including traffic intensity and road length variables (in 50m to 1000m buffers); ii) land use (e.g. port, industry, urban green, airports) and population / household density in buffers from 100 to 5000m. Inverse distance to roads was used in the stationary model development, but not in the mobile monitoring model as this variable cannot be computed (distance is 0).

Variable selection was done using a supervised forward stepwise selection procedure[18,19]. The direction of the effect for the variables was determined a priori (Table A.2) and the variable with the highest adjusted R^2 was entered first in the model. Model building stopped when new variables were not able to improve the adjusted R^2. The variables in the resulting models were checked for p-value (removed when p-value >0.10), collinearity (variance inflation factor > 3 were removed), and influential observations (if Cook’s D > 1 the model was further examined). We accounted for autocorrelation in the mobile measurements using a first order autoregressive (AR-1) term in the ARIMA procedure[9,11,14,31,32]. If after adding an AR-1 term to the identified model, variables were no longer significant (p>0.10), they were removed from the model.

2.6 Mobile LUR models versus Short-term stationary LUR Models

Mobile models of the 2014/2015 campaign were compared to short-term stationary models using different analyses, schematically shown and according to the numbers in the TOC art. First, we predicted concentration levels at stationary measurement sites using the mobile LUR model and compared them to their respective short-term stationary measurements (1). Second, we compared mobile and short-term stationary models by predicting concentration on 500 random addresses in each city (2). Third, we compared stationary and mobile LUR model predictions to external average home outdoor measurements based upon three times 24hour monitoring periods (3). In all data sets the GIS predictors were truncated to the range observed in the mobile monitoring campaign.

2.7 Overestimation of Mobile LUR models

We compared differences in predicted concentrations from mobile and stationary measurements for both the 2014/2015 and 2013 campaign to help understand the overestimation of mobile models from the 2013 campaign[18]. We explored four methodologies: *i.* using the distance between the road and the site where the prediction is made as an explanatory variable for the over-prediction; *ii.* LUR analyses with the delta (difference between predicted concentrations based on mobile model and observed short-term measurement) as a dependent variable with the available LUR GIS variables, *iii.* using a global correction based on the absolute and *iv.* relative differences between the predicted and measured concentration on the short-term stationary sites. Predictions based on the mobile model could then be subtracted by an absolute or relative value.

2.8 Robustness of Mobile LUR models

Stability of mobile LUR models was tested by comparing predictions of the mobile LUR models presented in this paper based on measurements in 2014/2015 with mobile LUR models based on measurements in 2013[18] (4). To rule out geographical differences between the campaigns analyses were restricted to Amsterdam, which was the only city represented in both campaigns. Other sensitivity analyses included the addition of a fixed city effect to the model, exclusion of the autocorrelation procedure, and exclusion of local emission peaks before model development.

③ Results

3.1 Distribution of UFP and BC

The distribution of road segment averaged UFP and BC measurements is shown in figure 1 and appendix table A.3 and figure A.1. Observed UFP and BC levels were on average higher on the road than at the short-term stationary sites, particularly the frequency of high UFP and BC concentrations is higher for mobile road segment averages than for short-term stationary averages. Stationary measurements are averaged over 30min, while mobile measurements are averaged over a road segment (about 25sec), thus partly explaining the lower variability in stationary measurements.

In figure A.1, the distribution of UFP and BC measurements are stratified by city and site type (urban background (UB) and traffic). Measurements in Amsterdam were on average higher than measurements in the other two cities. Mobile UFP measurements were on average 1.44 times higher than short-term stationary UFP measurements. For BC, mobile measurements were on average 1.92 times higher (Table A.3).

3.2 UFP: Mobile LUR models versus Short-term stationary LUR Models

The developed LUR models based on UFP mobile and short-term stationary measurements are shown in table 1. Both the short-term stationary and mobile models include similar population density and traffic related variables. The short-term stationary model includes industry in a 500m buffer whereas the mobile LUR model includes the area of ports and urban green area in the final model. As models were developed including an AR-1 term, we cannot report standard R^2 values of our main mobile models. Instead, the reported R^2 value is calculated by regressing the predicted concentrations based on the parameter estimates of the mobile AR-1 model without the AR-1 term. Due to the very short duration of measured concentrations and the large temporal variability, the R^2 value of the mobile monitoring model is low (15 %).

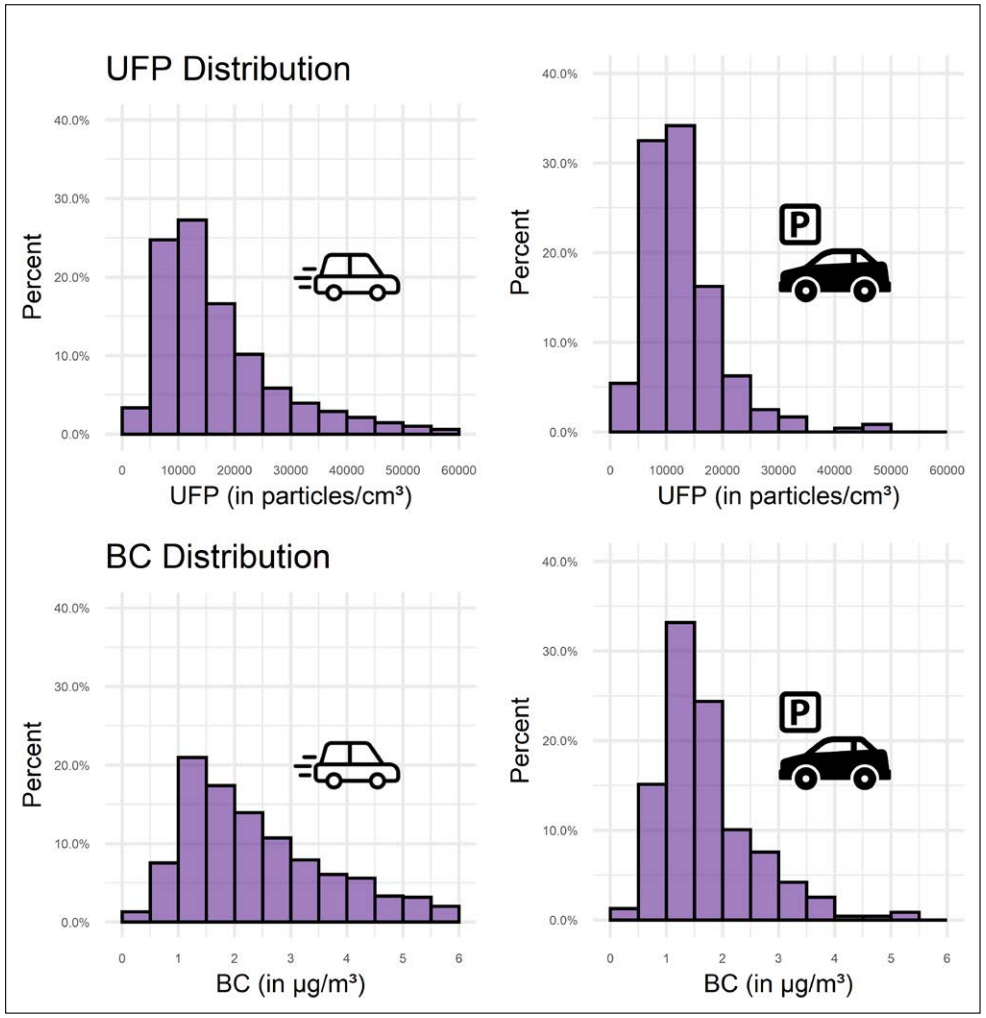
Models were also developed including a fixed effect for city. These models did not differ substantially from the original models (Table B.1). Other sensitivity analyses include models excluding the AR-1 term from model development and first excluding measurements flagged as local exhaust plumes before model development. All models are very similar and predicted concentrations based on these models on 1500 random addresses (500 per measurement city) are highly correlated (R^2 ~ 0.98; table B.1).

Although the LUR model for UFP explained only a small percentage of the variance in mobile measurements the model explained a much larger proportion of the variance of the short-term stationary measurements. The mobile LUR model for UFP explained 36% of the variance in the short-term stationary measurements (Figure 2a), which is more than two times higher than the mobile model is able to explain its own measurements (15%).

Figure 1

Distribution of mobile and stationary UFP/ BC measurements in 2014/2015.

↓



The number of mobile measurements does not match the total of road segments (n=5,236), as the figure for UFP is cropped to a maximum 60,000 particles per cm³ and 10 µg/m³ for BC (Max UFP=209140 particles per cm³, max BC=38 µg/m³).

Table 1

Mobile and Short-Term Stationary UFP Models.

↓

	UFP (particles/cm³)	
Variable	**Short-Term**	**Mobile AR-1**
Intercept	7,784 (582)	8,072 (968)
Population Density:		
Population density in a 5000m buffer	4,720 (977) [a]	
Residential land area in a 5000m buffer		7,763 (1,155)
Traffic:		
Traffic intensity on the nearest road	2,499 (860)	2,244 (756)
Heavy traffic intensity on the nearest road		989 (536)
Traffic intensity in a 50m buffer	3,459 (782)	
Length of major roads in a 50m buffer	2,873 (998)	
Length of major roads in a 100m buffer		4,588 (524)
Land Use:		
Area of industry in a 500m buffer	854 (450)	
Port area in a 5000m buffer		3,457 (995)
Urban green area in a 500m buffer		-1,001 (494)
R^2 of model	0.46	0.15 [b]
Number sites used for model development	240	5,236

[a] Regression slopes and standard error (between brackets), multiplied by the difference between 10th and 90th percentile for all predictors. [b] R^2 of model without AR-1 term.

Comparing predicted concentrations at random addresses (n=1500) revealed a strong correlation (R^2=0.64) between mobile and short-term stationary model predictions (Figure 2b). This correlation was reasonably similar for traffic and urban background sites (R^2 of 0.71 vs. 0.60; results not shown).

Figure 3 shows the correlation between predicted UFP concentrations for 42 home outdoor measurement sites and their respective average of 3 times 24h-measurements, based on the mobile (Figure 3a) and short-term stationary model (Figure 3b). The mobile model for UFP predicts 57% of the variation in the home outdoor measurements, whereas the short-term stationary model predicts 46% of the variation. These results were consistent with new analyses of our previous campaign. The mobile model based on measurements from 2013 predicted 51% of the variance of home outdoor concentration levels in 2014/2015 (Figure B.1).

Figure 2

(a) Predicted concentration levels (particles/cm³) at stationary sites based on mobile LUR model compared to stationary measurements. (b) Comparison of predicted concentration levels based on mobile and stationary LUR models at 1,500 random addresses in Amsterdam, Utrecht and Maastricht.

↓

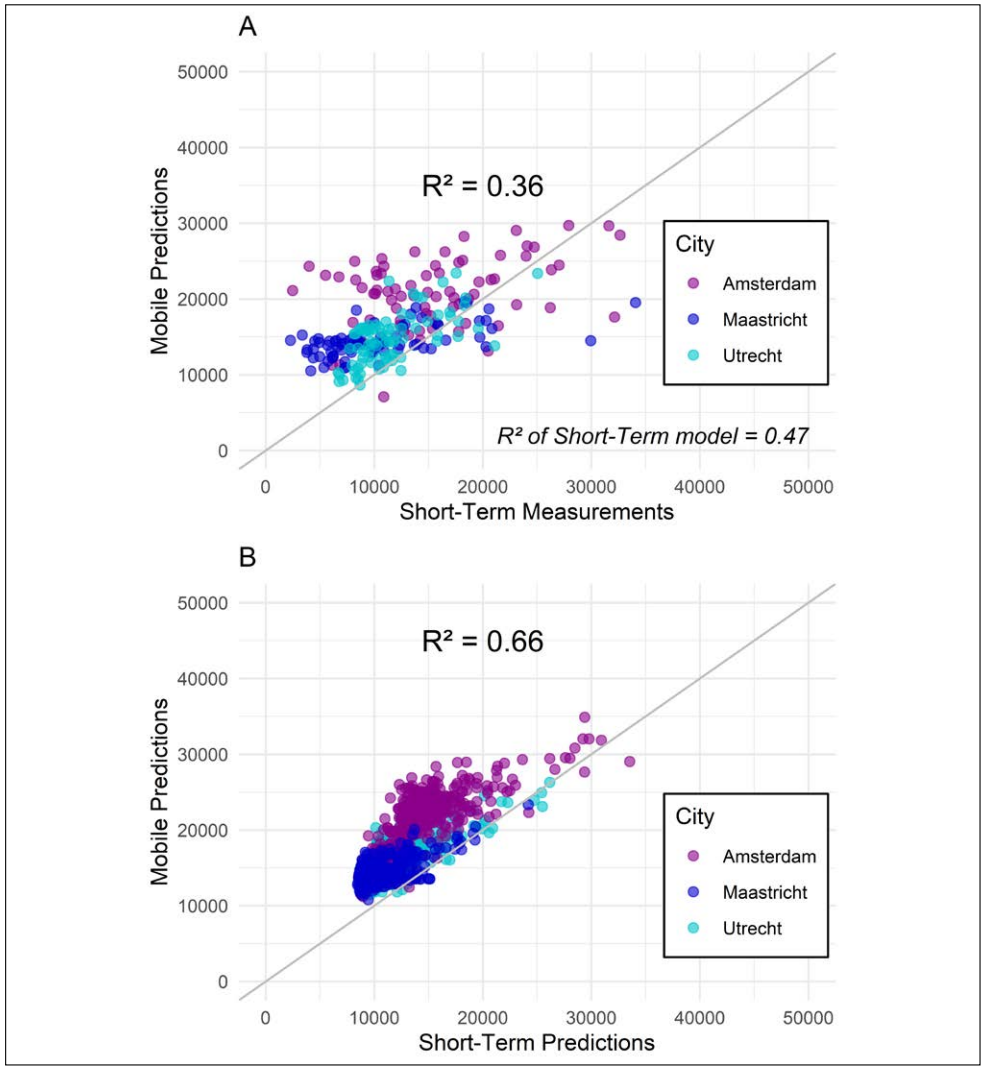
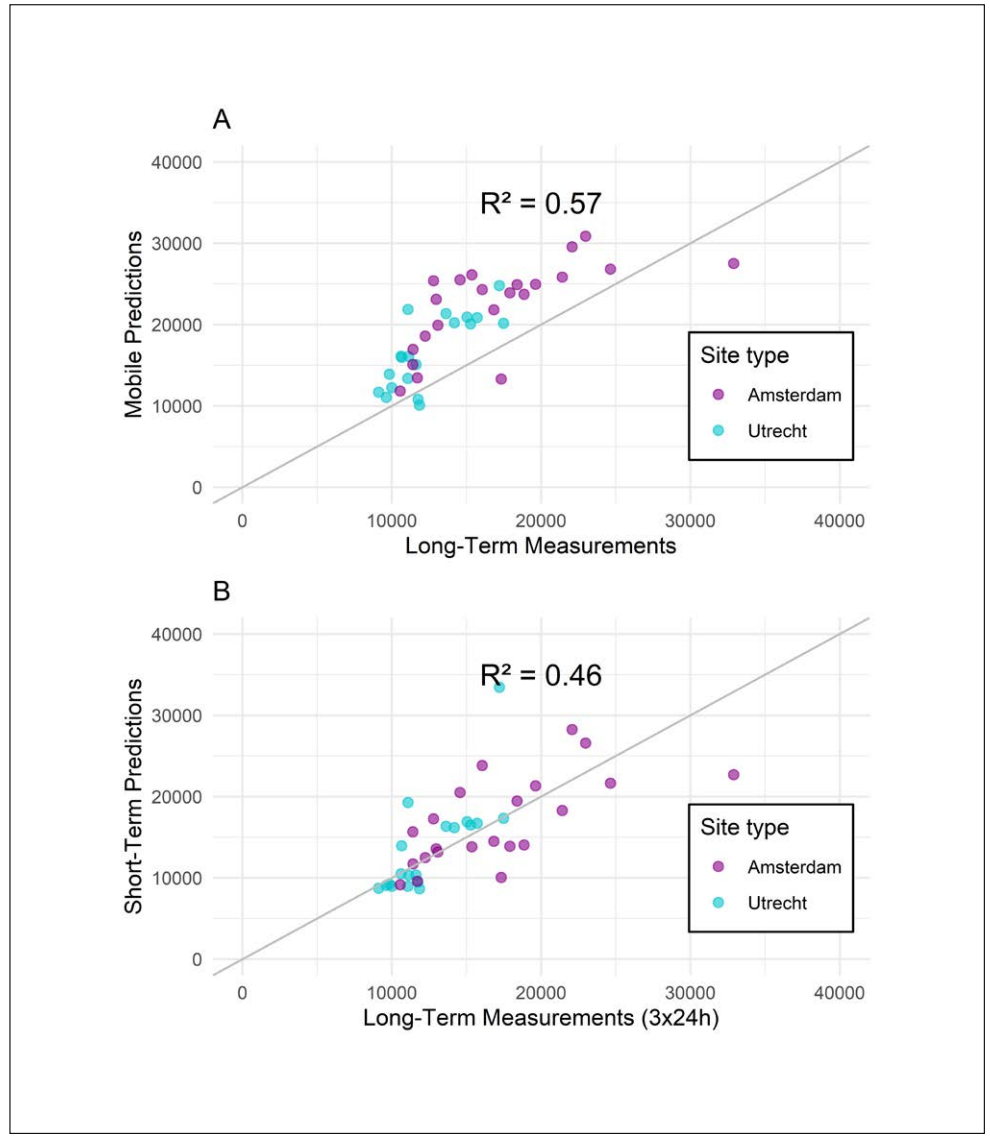


Figure 3

Predicted concentration levels (particles/cm³) at home outdoor sites (n=42) based on mobile models (a) and short-term stationary models (b) compared to the average of 3 x 24h measurements at home addresses.

↓



3.3 BC: Mobile LUR models versus Short-term stationary LUR Models

Like the UFP models, the mobile and short-term stationary LUR models for BC include population density and nearby traffic variables in both models. For BC, urban green area is also included in the mobile model, similar as to the UFP mobile model. The LUR model and figures related to BC can be found in appendix C. The LUR model poorly explains the spatial variation in the mobile measurements (R^2=0.10; Table C.1), comparable to the UFP model. Similarities with UFP stop when we try to use the model to predict concentration levels at the short-term stationary and home outdoor sites. The mobile model explained only 14% of the variance in the short-term stationary measurements and 14% of the variation in the home outdoor measurements (Figure C.1/C.2). The stationary model explained 44% of the spatial variation in the stationary measurements (Table C.1) and 38% of the home outdoor measurements (Figure C.2). Mobile BC model predictions at 1500 random households were only moderately correlated to the short-term stationary model predictions (R^2=0.37; Figure C.1).

Where the UFP mobile LUR was able to predict measurements with longer averaging periods (3x24h) with greater accuracy, the mobile BC model could not. Predictions made by the mobile model based on 2013 BC measurements were also poorly correlated to home outdoor measurements in the current study (R^2=0.17). Results are shown in figure C.3, together with the results from the short-term stationary model predictions. Since mobile LUR models for BC (from 2013 and 2014/2015) did not predict the measurements with longer averaging periods well, we did not precede with further analyses of the BC LUR models in this paper. It appears, due to the long averaging time of the instrument, that our measurement device is unable to capture the fine spatial scale needed in urban settings.

3.4 Exploration of overestimation of mobile UFP LUR models

In all analyses we observed higher predicted concentration levels based on mobile UFP models than predictions made by short-term stationary models, consistent with our previous work[18]. Predictions made on randomly selected addresses were on average about 5000 particles/cm³ and 30% higher than models based on short-term stationary measurements (Table 2). No significant differences in overestimation were found between traffic and urban background sites. Predicted UFP concentrations based on mobile models also overestimated 24h home outdoor measurements. The 2014/2015 mobile model overestimated the home outdoor measurements by 27% (about 4100 particles/cm³ on average), whereas the short-term stationary models did not over-predict concentrations.

We explored four methodologies to correct for the difference between mobile and short-term stationary predictions. Distance to the road was not related to the difference between mobile predictions and measured UFP for the short-term stationary sites and the home outdoor sites (Figure B.2). We also developed several LUR models with the delta as dependent variable, but could not derive a reasonable and interpretable LUR Model. The other two methods considered are to compensate the overestimation of mobile LUR models by reducing the mobile predicted levels overall by 30% or 5000 particles/cm³. These methods were also compared to the short-term stationary predictions on random addresses. In these analyses, the relative reduction of 30% to the mobile model predicted concentration seems to have a better agreement with the short-term stationary model predictions (Figure B.3).

Table 2

Differences between the 2013 and 2014/2015 mobile measurement campaigns.

↓

	2014-2015 Campaign	**2013 Campaign**
Cities	Amsterdam, Utrecht, and Maastricht	Amsterdam and Rotterdam
Seasons	Winter, Spring and Summer	Winter and Spring
UFP over-prediction [a]	33% (5000 particles/cm³)	29% (4200 particles/cm³)
Traffic	25% (3600 particles/cm³)	31% (6000 particles/cm³)
Urban Background	35% (5200 particles/cm³)	29% (4000 particles/cm³)

[a] Difference between predicted concentration levels based on mobile and short-term stationary LUR models, tested on 500 random addresses in Amsterdam.

3.5 Robustness of mobile LUR models

As we conducted measurement campaigns in 2013 and 2014/15 we were interested to see if the model predictions were similar when using measurements from different geographical and temporal settings for model development. Mobile models from the 2013 campaign (Table B.3) are based on measurements in Rotterdam and Amsterdam, both industrialised and busy cities with the presence of a harbour. The mobile models from the 2014/2015 are based on the cities of Amsterdam, Utrecht and Maastricht. The cities of Utrecht and Maastricht do not have a port area and are smaller cities with less traffic than Amsterdam and Rotterdam. UFP models from both time periods were used to predict concentration levels at 1500 random addresses in Amsterdam, Maastricht and Utrecht. These predictions were highly correlated as shown in figure 4 (R^2=0.80). Predictions made by the two short-term stationary models were also highly correlated as shown in figure B.4 (R^2=0.60), but less than the mobile models.

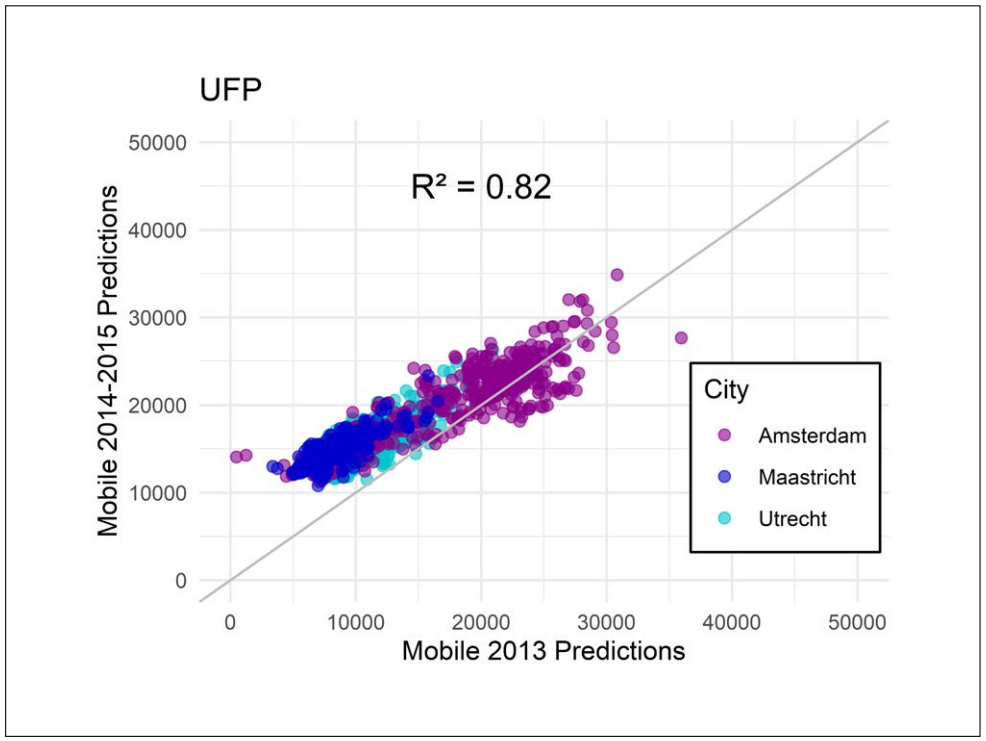
The mobile UFP model from 2013 had a lower intercept and included natural area in a 5000m buffer, resulting in the observed deviance in absolute concentration predictions at the lower end of the concentration range. Most of these sites are located in Maastricht, a less urban area compared to Rotterdam and Amsterdam. To exclude the influence of geographical differences, mobile LUR models were also created for the city of Amsterdam only.

These LUR models are shown in tables B.2 and B.3. Correlation between the 2013 and 2014/2015 mobile models is less than models with all cities included (R^2=0.51; figure B.5). Random variability due to developing models on a smaller number of sites may have contributed to the lower correlation between the two mobile models.

Figure 4

Mobile predictions (particles/cm³) based on 2013 measurement campaign versus mobile predictions based on measurement presented in this paper (2014/2015), on 1500 random addresses in Amsterdam, Utrecht and Maastricht.

↓



④ Discussion

Our novel analyses demonstrate many scenarios in which LUR model predictions for UFP are robust from data collection design and sampling temporal range. Models based upon mobile and short-term stationary monitoring provided highly correlated predictions of UFP concentrations at 1500 randomly selected addresses in three Dutch cities (R^2 =0.64). Mobile and short-term models explained 57% and 46% of the variability in measured average home outdoor UFP concentrations at 42 external sites in Amsterdam and Utrecht. We found a high correlation (R^2=0.80) between predicted UFP levels based on the mobile LUR model and a previously developed mobile LUR model (with another spatial extent and in a different year) at 1500 random addresses in Amsterdam, Maastricht and Utrecht. Predicted UFP concentrations made by the mobile models were on average 30% higher than predicted by the stationary models. Distance to the road and land-use/traffic predictors did not explain the overprediction.

In contrast, mobile model predictions for BC correlated only moderately with those of short-term stationary BC models. Mobile BC models did not explain home outdoor BC concentrations at the 42 external sites well (R^2 = 14%).

4.1 Mobile versus Short-term Stationary Monitoring models for UFP

Our mobile UFP LUR model explains 36% of the spatial variability of the short-term stationary measurements, which is more than two-fold the explained variance of the mobile measurements where the model is based on (15%). Similar results were found in the 2013 campaign[18], where the mobile LUR model was able to explain 26% of the short-term stationary measurements, two times higher than the explained variability of the mobile measurements the mobile model was based on (13%). In this study we were additionally able to compare mobile and short-term models to external measurements with longer averaging periods (3x24 hours) and found that UFP mobile models predicted an even larger fraction of the variability of these longer term measurements (R^2=0.57). This analysis further supports the assertion that despite the low R^2 of mobile UFP LUR models they provide robust exposure estimates at residential addresses.

The low model R^2 has been attributed to the high temporal variability in measured concentrations of very short duration per site[18,19]. Temporal predictors are purposely left out model development as we set out to develop a spatial model. We have now documented in two combined short-term and mobile monitoring studies that the explained variance

of measurements increases when the model is compared with measurements with longer duration[18,19,22]. This is due to the significant decrease in total variance from temporal averaging. LUR models based on longer term UFP monitoring campaigns[34–38] explained spatial variability of their own measured UFP concentrations a lot better than our study, with R^2 values ranging from 0.48 to 0.89. In the current study, an increase in the averaging time of measurements led to an increase of the ability of mobile models to predict these measurements; from 15% to mobile measurements (median 25sec), 36% to short-term stationary measurements (3x30min) and 57% to home outdoor measurements (3x24h). Consistently, studies that have repeated mobile monitoring at the same road segment more often than in our studies have reported fairly high model and validation R^2 values[10,12,15].

For the 2014/2015 campaign, the model predictions of the mobile and short-term model at external addresses (n=1500) were fairly highly correlated (R^2=0.64), replicating, albeit somewhat lower, our previous observation based on the 2013 monitoring campaign (R^2=0.92). The lower correlation in our current work could be due to the larger and more diverse study area. The mobile model was slightly better than the short-term stationary model in predicting concentration levels on the home outdoor sites (57 versus 46%). For the 2013 campaign, mobile and short-term stationary models explained 51 and 55% of the concentration variability at the home outdoor sites (Figure B.1). We conclude that mobile and short-term stationary monitoring lead to very similar predictions of spatial exposure contrasts, with no consistent difference in validity.

4.2 Mobile versus Short-term Stationary Monitoring models for BC

The moderate agreement between mobile and short-term stationary model predictions for black carbon in the current study (R^2=0.37) is inconsistent with our previous evaluation, based on a mobile monitoring campaign in 2013 (R^2=0.88) [18]. When we compared the mobile model predictions with the home outdoor measurements from 2014/2015, we poorly explained the variability in monitored concentrations (14%). The predicted levels on these sites based on the mobile model from 2013 was also poorly correlated with the measurements (R^2=0.17). The short-term stationary models in both campaigns explained more variation of the home outdoor sites (R^2=0.38 and 0.28; Figure C.2 and C.3).

The BC measurement device used in the 2013 and 2014/15 campaign had a temporal resolution of one minute, which was later adjusted to two or three minutes because of noise of the instrument. This is too long to detect the high spatial variation of BC, especially within city limits. The derived mobile LUR model has a relatively large estimate for residential land area in a 5000m buffer, probably representing the difference between cities. Variation within cities could not be sufficiently assessed by our BC instrument using mobile monitoring by car driving. In contrast, short-term stationary monitoring can be performed with a Micro-Aethalometer as each measurement consist of 30 1-minute averages. The Micro-Aethalometer may be useful in mobile monitoring in much higher pollution environments and in mobile monitoring campaigns using slow moving platforms such as bicycles and backpacks (whilst walking). Lonati et al[39] used bicycles to measure BC in a city in Northern Italy and found that the 1-min time resolution of the Micro-Aethalometer always exceeded the suggested attenuation threshold. Hankey and Marshall[10] also needed at least 1min averages to smooth the noise of the instrument and reported moderate model R^2 for cycling-based mobile monitoring for BC (35-49%), though lower than for particle number (58-61%).

4.3 Over-prediction of mobile UFP models

The mobile UFP LUR models generated higher predicted concentrations than short-term stationary models for the same locations. In our previous study, we could not distinguish between overprediction by the mobile model and under prediction of the short-term model or a combination of both. In our current study this is corroborated in the comparison of the mobile and short-term at home outdoor sites for which we had independent measurements available. The mobile but not the short-term stationary model over predicted home outdoor concentrations. This mostly related to mobile measurements being taken on-road where concentration levels are likely to be higher than at roadside residential addresses. Multiple studies have observed sharp UFP and BC gradients in near-road urban environments with gradients similar to what was observed in our previous study[1,6,7,24,40–44]. However, no studies have measured actual difference between measuring on-road and near the side of the road. Ragettli et al[45] compared measurements of UFP on the sidewalk and at the façade of buildings and found a difference in concentration levels of about 20%. Kaur et al[41] found a difference between measuring at the edge of the curb side near the road and measuring at the side of the building. They observed pedestrian exposure whilst walking curb side of about 86,000 particles/cm³, while an average of about 73,000 particles/cm³ was measured walking along the building side of the pavement (difference about 13,000 particles/cm³ which amounts to 15%). These relative differences are in the range of the finding in this paper with concentration differences between on-road and sidewalk of about 30%. This was also found in the 2013 campaign, suggesting that this number is not significantly affected by geographical differences within The Netherlands.

In our dataset we also found no correlation between the mobile model overprediction and the distance of the short-term measurement sites to the road. On top of that, LUR analyses of the delta (difference between predicted and observed at the short-term measurement site) generated no interpretable results. One of the reasons for this is probably the lack of accuracy of GIS and GPS of the measurements when it comes to differences in the range of 5 -20 meters. Short-term stationary sites were mostly located within 2 to 10m from the edge of the road. Within these distances the mobile models are not able to scale down concentration levels to residential addresses[18]. Furthermore, mobile monitoring campaigns usually do not have short-term stationary measurements to make adjustments based on distance or LUR analyses. For the use in epidemiology, we suggest to either perform no corrections at all, as relative ranking are preserved, or use of an empirical determined factor to scale down mobile LUR model predictions, based on study-area specific data.

A rationale for no adjustment is that other factors can influence the over- or under-prediction of mobile LUR models. All our measurements are sampled between 9:15AM and 4:00PM, excluding rush hour. This could lead to some underestimation of our LUR models. The exclusion of nighttime period could in contrast lead to an over-prediction of 24hour average concentrations. Other studies only sampled during rush hour[10,11,14] or only sampled in the summer season[10,11,15], which respectively would cause some overestimation and underestimation[1,46] of concentration levels. As such the observed difference here between mobile and short-term stationary LUR models may well be within the error of other limitations in these campaigns.

4.5 Robustness of mobile LUR Models

We compared LUR models developed from two different monitoring campaigns (including different cities) and found highly correlated predicted concentration levels at 1500 random addresses, providing further support for the robustness of LUR models based upon mobile monitoring data. The comparability of the two models is consistent with previous observations of stable spatial contrast of air pollution over short periods (here 1-2 years), and a previous analysis of the 2013 campaign suggesting no difference between the combined city model and city-specific models[19]. In comparable Dutch cities, similar predictor variables (mainly small-scale traffic), explain a major fraction of UFP spatial variability.

In general, models from the 2013 and this campaign included similar predictors, which was also found by Hatzopoulou et al.[47] reviewing LUR models of several Canadian cities. Both Dutch models include a large scale population density buffer, the length of major roads in a small buffer, the area of natural land, the presence of a nearby port and traffic intensity variables. The area of airports was included in the model from 2013, but not in the 2014/2015 LUR model. It could be that the area of airports was not included our LUR model because of limited measurements near airports (only Amsterdam in the 2014/2015 campaign).

4.6 Advantages and limitations of mobile monitoring

Mobile monitoring is a cost-effective method to generate LUR models, as a wide range of conditions can be captured in a limited amount of time and with a limited amount of instruments[30,46]. A high spatial density of measurements can be obtained, sampling more sites which are more representative for people's exposures such as near-intersections and close proximity to traffic lights. Conversely, mobile monitoring decreases sampling time significantly opposed to stationary measurement campaigns leading to substantial uncertainties in concentration fields[30]. This is reflected in our study in very low R^2 values for mobile models explaining the spatial variability in mobile measurements. For LUR model development, however, the short sampling time per road segment is likely counterbalanced by the increased spatial variability[47,48], which explains the consistent selection of explanatory variables and good external dataset prediction.

Several mobile monitoring studies suggest to use a minimum temporal resolution[12,49,50] or minimum number of visits[11] to adequately assign average concentrations per road segment. Hatzopoulou et al[47] looked into the amount of visits needed per road segment to characterise its average concentration and found an increase in model R^2 with an increasing number of visits. 20% of the road segments in our data set consist of 10 seconds or less. Excluding these road segments from model development increases our model R^2 to 0.20 (results not shown). This model however does not improve predictions to short-term stationary measurements (R^2 remains 0.36) and home outdoor measurements (R^2 of 0.56 compared to 0.57 for all road segments).

Of note, LUR models were developed using linear regression and adjusted by adding an AR-1 term to the model to correct for spatial autocorrelation. The AR-1 term assumes regular time and space intervals and that the autocorrelation remains constant over time. This method is unlikely to be optimal, but is considered the best option in several mobile monitoring campaigns[11,14,51]. Other mobile campaigns include a Local Indicator

of Spatial Analysis (LISA)[10], extend the averaging period[52] or disregard the issue[13]. Performing sensitivity analyses on the autoregressive models did not yield significant different results from the original models (Table B.1) and also in our previous campaign[18] and in a study by Weichenthal et al[14].

⑤ Conclusions

Models based upon mobile and short-term stationary monitoring provided fairly high correlated predictions of UFP concentrations at 1500 randomly selected addresses in three Dutch cities. Mobile and short-term models explained 57% and 46% of the variability in measured average home outdoor UFP concentrations at external sites. In contrast, mobile BC models did not explain home outdoor BC concentrations at the external sites well (R^2 = 14%). We found a high correlation (R^2=0.80) between predicted UFP levels based on the mobile LUR model and a previously developed mobile LUR model (with another spatial extent and in a different year) at 1500 random addresses. Because of on-road measurements predicted UFP concentrations made by the mobile models were on average 30% higher than predicted by the stationary models. Distance to the road and land-use / traffic predictors did not explain the overprediction. Overall, our study supports that robust LUR models for UFP can be developed based on mobile monitoring.

Supplement Information

The Supporting information is divided into three subsections. Appendix A contains general information concerning both UFP and BC. Appendix B contains supporting information about UFP and Appendix C about BC.

Acknowledgements

This study was funded by the Environmental Defence Fund, The Netherlands Royal Academy of Science and the EU seventh Framework Program EXPOsOMICS Project.

References

(1) Morawska, L.; Ristovski, Z.; Jayaratne, E. R.; Keogh, D. U.; Ling, X. Ambient nano and ultrafine particles from motor vehicle emissions: Characteristics, ambient processing and implications on human exposure. *Atmospheric Environment*. 2008, pp 8113–8138.

(2) Ghassoun, Y.; Ruths, M.; Löwner, M.-O.; Weber, S. Intra-urban variation of ultrafine particles as evaluated by process related land use and pollutant driven regression modelling. *Sci. Total Environ.* **2015**, *536*, 150–160.

(3) Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y.; Luepker, R. V.; Mittleman, M. A.; et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the american heart association. *Circulation* **2010**, *121* (21), 2331–2378.

(4) Hoek, G.; Boogaard, H.; Knol, A.; De Hartog, J.; Slottje, P.; Ayres, J. G.; Borm, P.; Brunekreef, B.; Donaldson, K.; Forastiere, F.; et al. Concentration response functions for ultrafine particles and all-cause mortality and hospital admissions: Results of a European expert panel elicitation. *Environ. Sci. Technol.* **2010**, *44* (1), 476–482.

(5) Environmental, E.; Agency. Urban environment http://www.eea.europa.eu/themes/urban#tab-publications (accessed Jul 25, 2016).

(6) Van den Bossche, J.; Peters, J.; Verwaeren, J.; Botteldooren, D.; Theunis, J.; De Baets, B. Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset. *Atmos. Environ.* **2015**, *105*, 148–161.

(7) Peters, J.; Van den Bossche, J.; Reggente, M.; Van Poppel, M.; De Baets, B.; Theunis, J. Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. *Atmos. Environ.* **2014**, *92*, 31–43.

(8) Zwack, L. M.; Hanna, S. R.; Spengler, J. D.; Levy, J. I. Using advanced dispersion models and mobile monitoring to characterize spatial patterns of ultrafine particles in an urban area. *Atmos. Environ.* **2011**, *45* (28), 4822–4829.

(9) Zwack, L. M.; Paciorek, C. J.; Spengler, J. D.; Levy, J. I. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. *Environ. Health Perspect.* **2011**, *119* (6), 852–859.

(10) Hankey, S.; Marshall, J. D. Land Use Regression models of on-road particulate air pollution (Particle Number, Black Carbon, $PM_{2.5}$, Particle Size) using mobile monitoring. *Environ. Sci. Technol.* **2015**, *49* (15), 9194–9202.

(11) Farrell, W.; Weichenthal, S.; Goldberg, M.; Valois, M. F.; Shekarrizfard, M.; Hatzopoulou, M. Near roadway air pollution across a spatially extensive road and cycling network. *Environ. Pollut.* **2016**, *212*, 498–507.

(12) Weichenthal, S.; Van Ryswyk, K.; Goldstein, A.; Shekarrizfard, M.; Hatzopoulou, M. Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: A land use regression model. *Environ. Pollut.* **2016**, *208*, 241–248.

(13) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An hourly regression model for ultrafine particles in a near-highway urban area. *Environ. Sci. Technol.* **2014**, *48* (6), 3272–3280.

(14) Weichenthal, S.; Farrell, W.; Goldberg, M.; Joseph, L.; Hatzopoulou, M. Characterizing the impact of traffic and the built environment on near-road ultrafine particle and black carbon concentrations. *Environ. Res.* **2014**, *132*, 305–310.

(15) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Reali, C.; Sun, T.; Evans, G. J. Development of a land-use regression model for ultrafine particles in Toronto, Canada. *Atmos. Environ.* **2015**, *110*, 84–92.

(16) Larson, T.; Henderson, S. B.; Brauer, M. Mobile monitoring of particle light absorption coefficient in an urban area as a basis for land use regression. *Environ. Sci. Technol.* **2009**, *43* (13), 4672–4678.

(17) Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; Thiele, L. Deriving high-resolution urban air pollution maps using mobile sensor nodes. In *Pervasive and Mobile Computing*; 2015; Vol. 16, pp 268–285.

(18) Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of Ultrafine Particles and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environ. Sci. Technol.* **2016**, 1–31.

(19) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712–8720.

(20) Klompmaker, J.; Montagne, D.; Meliefste, K.; Hoek, G.; Brunekreef, B. Spatial variation of ultrafine particles and black carbon in two cities: results from a short-term measurement campaign. *Sci. Total Environ.* **2015**, *508*, 266–275.

(21) Vineis, P.; Chadeau-Hyam, M.; Gmuender, H.; Gulliver, J.; Herceg, Z.; Kleinjans, J.; Kogevinas, M.; Kyrtopoulos, S.; Nieuwenhuijsen, M.; Phillips, D.; et al. The exposome in practice: Design of the EXPOsOMICS project. *International Journal of Hygiene and Environmental Health*. 2016.

(22) Nunen, E. van; Vermeulen, R.; Tsai, M.-Y.; Probst-Hensch, N.; Ineichen, A.; Davey, M. E.; Imboden, M.; Ducret-Stich, R.; Naccarati, A.; Raffaele, D.; et al. Land use regression models for Ultrafine Particles in six European areas. *Environ. Sci. Technol.* **2017**, acs.est.6b05920.

(23) Padró-Martínez, L. T.; Patton, A. P.; Trull, J. B.; Zamore, W.; Brugge, D.; Durant, J. L. Mobile monitoring of particle number concentration and other traffic-related air pollutants in a near-highway neighborhood over the course of a year. *Atmos. Environ.* **2012**, *61*, 253–264.

(24) Weijers, E. P.; Khlystov, A. Y.; Kos, G. P. A.; Erisman, J. W. Variability of particulate matter concentrations along roads and motorways determined by a moving measurement unit. *Atmos. Environ.* **2004**, *38* (19), 2993–3002.

(25) Eeftens, M.; Tsai, M. Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; et al. Spatial variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2 - Results of the ESCAPE project. *Atmos. Environ.* **2012**, *62*, 303–317.

(26) Asbach, C.; Kaminski, H.; Von Barany, D.; Kuhlbusch, T. A. J.; Monz, C.; Dziurowitz, N.; Pelzer, J.; Vossen, K.; Berlin, K.; Dietrich, S.; et al. Comparability of portable nanoparticle exposure monitors. In *Annals of Occupational Hygiene*; 2012; Vol. 56, pp 606–621.

(27) Meier, R.; Clark, K.; Riediker, M. Comparative Testing of a Miniature Diffusion Size Classifier to Assess Airborne Ultrafine Particles Under Field Conditions. *Aerosol Sci. Technol.* **2013**, *47* (1), 22–28.

(28) Hagler, G. S. W.; Yelverton, T. L. B.; Vedantham, R.; Hansen, A. D. A.; Turner, J. R. Post-processing method to reduce noise while preserving high time resolution in aethalometer real-time black carbon data. *Aerosol Air Qual. Res.* **2011**, *11* (5), 539–546.

(29) Drewnick, F.; Böttger, T.; Von Der Weiden-Reinmüller, S. L.; Zorn, S. R.; Klimach, T.; Schneider, J.; Borrmann, S. Design of a mobile aerosol research laboratory and data processing tools for effective stationary and mobile field measurements. *Atmos. Meas. Tech.* **2012**, *5* (6), 1443–1457.

(30) Ranasinghe, D. R.; Choi, W.; Winer, A. M.; Paulson, S. E. Developing high spatial resolution concentration maps using mobile air quality measurements. *Aerosol Air Qual. Res.* **2016**, *16* (8), 1841–1853.

(31) Buonocore, J. J.; Lee, H. J.; Levy, J. I. The influence of traffic on air quality in an urban neighborhood: a community-university partnership. *Am. J. Public Health* **2009**, *99 Suppl 3*.

(32) Hsu, H. H.; Adamkiewicz, G.; Houseman, E. A.; Spengler, J. D.; Levy, J. I. Using mobile monitoring to characterize roadway and aircraft contributions to ultrafine particle concentrations near a mid-sized airport. *Atmos. Environ.* **2014**, *89*, 688–695.

(33) Hofman, J.; Staelens, J.; Cordell, R.; Stroobants, C.; Zikova, N.; Hama, S. M. L.; Wyche, K. P.; Kos, G. P. A.; Van Der Zee, S.; Smallbone, K. L.; et al. Ultrafine particles in four European urban environments: Results from a new continuous long-term monitoring network. *Atmos. Environ.* **2016**, *136*, 68–81.

(34) Abernethy, R. C.; Allen, R. W.; Mckendry, I. G.; Brauer, M. A Land Use Regression Model for Ultra fine Particles in Vancouver, Canada. *Environ. Sci. Technol.* **2013**, *47* (10), 5217–5225.

(35) Hoek, G.; Beelen, R.; Kos, G.; Dijkema, M.; van der Zee, S. C.; Fischer, P. H.; Brunekreef, B. Land use regression model for ultrafine particles in Amsterdam. *Environ. Sci. Technol.* **2011**, *45* (2), 622–628.

(36) Eeftens, M.; Meier, R.; Schindler, C.; Aguilera, I.; Phuleria, H.; Ineichen, A.; Davey, M.; Ducret-Stich, R.; Keidel, D.; Probst-Hensch, N.; et al. Development of land use regression models for nitrogen dioxide, ultrafine particles, lung deposited surface area, and four other markers of particulate matter pollution in the Swiss SAPALDIA regions. *Environ. Heal.* **2016**, *15* (1), 53.

(37) Wolf, K.; Cyrys, J.; Harciníková, T.; Gu, J.; Kusch, T.; Hampel, R.; Schneider, A.; Peters, A. Land use regression modeling of ultrafine particles, ozone, nitrogen oxides and markers of particulate matter pollution in Augsburg, Germany. *Sci. Total Environ.* **2017**, *579*, 1531–1540.

(38) Cattani, G.; Gaeta, A.; Di Menno di Bucchianico, A.; De Santis, A.; Gaddi, R.; Cusano, M.; Ancona, C.; Badaloni, C.; Forastiere, F.; Gariazzo, C.; et al. Development of land-use regression models for exposure assessment to ultrafine particles in Rome, Italy. *Atmos. Environ.* **2017**, *156*, 52–60.

(39) Lonati, G.; Ozgen, S.; Ripamonti, G.; Signorini, S. Variability of Black Carbon and Ultrafine Particle Concentration on Urban Bike Routes in a Mid-Sized City in the Po Valley (Northern Italy). *Atmosphere (Basel).* **2017**, *8* (2), 40.

(40) Hagler, G. S. W.; Baldauf, R. W.; Thoma, E. D.; Long, T. R.; Snow, R. F.; Kinsey, J. S.; Oudejans, L.; Gullett, B. K. Ultrafine particles near a major roadway in Raleigh, North Carolina: Downwind attenuation and correlation with traffic-related pollutants. *Atmos. Environ.* **2009**, *43* (6), 1229–1234.

(41) Kaur, S.; Nieuwenhuijsen, M. J.; Colvile, R. N. Pedestrian exposure to air pollution along a major road in Central London, UK. *Atmos. Environ.* **2005**, *39* (38), 7307–7320.

(42) Fujita, E. M.; Campbell, D. E.; Arnott, W. P.; Johnson ed, T.; Ollison, W. Concentrations of mobile source air pollutants in urban microenvironments. *J. Air Waste Manag. Assoc.* **2014**, *64*, 743–758.

(43) Baldwin, N.; Gilani, O.; Raja, S.; Batterman, S.; Ganguly, R.; Hopke, P.; Berrocal, V.; Robins, T.; Hoogterp, S. Factors affecting pollutant concentrations in the near-road environment. *Atmos. Environ.* **2015**, *115*, 223–235.

(44) Hitchins, J.; Morawska, L.; Wolff, R.; Gilbert, D. Concentrations of submicrometre particles from vehicle emissions near a major road. *Atmos. Environ.* **2000**, *34* (1), 51–59.

(45) Ragettli, M. S.; Ducret-Stich, R. E.; Foraster, M.; Morelli, X.; Aguilera, I.; Basagaña, X.; Corradi, E.; Ineichen, A.; Tsai, M. Y.; Probst-Hensch, N.; et al. Spatio-temporal variation of urban ultrafine particle number concentrations. *Atmos. Environ.* **2014**, *96*, 275–283.

(46) Rizza, V.; Stabile, L.; Buonanno, G.; Morawska, L. Variability of airborne particle metrics in an urban area. *Environ. Pollut.* **2017**, *220*, 625–635.

(47) Hatzopoulou, M.; Valois, M. F.; Levy, I.; Mihele, C.; Lu, G.; Bagg, S.; Minet, L.; Brook, J. Robustness of Land-Use Regression Models Developed from Mobile Air Pollutant Measurements. *Environ. Sci. Technol.* **2017**, *51* (7), 3938–3947.

(48) Peters, J.; Theunis, J.; Van Poppel, M.; Berghmans, P. Monitoring PM_{10} and ultrafine particles in urban environments using mobile measurements. *Aerosol Air Qual. Res.* **2013**, *13* (2), 509–522.

(49) Weichenthal, S.; Ryswyk, K. Van; Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A land use regression model for ambient ultrafine particles in Montreal, Canada: A comparison of linear regression and a machine learning approach. *Environ. Res.* **2016**, *146*, 65–72.

(50) Apte, J. S.; Messier, K. P.; Gani, S.; Brauer, M.; Kirchstetter, T. W.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Vermeulen, R. C. H.; Hamburg, S. P. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environ. Sci. Technol.* **2017**, *51* (12), 6999–7008.

(51) Zwack, L. M.; Paciorek, C. J.; Spengler, J. D.; Levy, J. I. Characterizing local traffic contributions to particulate air pollution in street canyons using mobile monitoring techniques. *Atmos. Environ.* **2011**, *45* (15), 2507–2514.

(52) Fruin, S.; Westerdahl, D.; Sax, T.; Sioutas, C.; Fine, P. M. Measurements and predictors of on-road ultrafine particle concentrations and associated pollutants in Los Angeles. *Atmos. Environ.* **2008**, *42* (2), 207–219.

Supplement Information for

Robustness of Intra Urban Land-Use Regression Models for Ultrafine Particles and Black Carbon based on Mobile Monitoring

Table A.1 Appendix A: General Information

Overview of different data sets.

↓

Data set	Year	Location	Number	Resolution	Instruments
Mobile Measurements	2014/2015	Amsterdam, Maastricht and Utrecht	5236 Road Segments	~25seconds	CPC (UFP) and Aethalometer (BC)
	2013	Amsterdam and Rotterdam	2964 Road Segments	~20seconds	CPC (UFP) and Aethalometer (BC)
Short-Term Stationary Measurements	2014/2015	Amsterdam, Maastricht and Utrecht	240 Sites	3 times 30 minutes	CPC (UFP) and Aethalometer (BC)
Randomly selected addresses	/	Amsterdam, Maastricht and Utrecht	1500 Addresses	/	/
Home Outdoor Measurements	2014/2015	Amsterdam and Utrecht	42 Homes	3 times 24 hours	MiniDisc (UFP) and Harvard Impactors (BC)

Table A.2

Spatial predictor variables with units, a priori defined directions of effect and buffer sizes in the mobile and short-term stationary data sets.

↓

Predictor variable	Units	Direction of effect	Buffer (m)	Mobile			Short-term Stationary		
				10th perc	Mean	90th Perc	10th perc	Mean	90th Perc
Industry area	m2	+	**100**	0	1400	0	0	440	0
			300	0	13633	37122	0	7237	6463
			500	0	41895	166610	0	29522	110419
			1000	0	197751	711267	0	174742	619476
			5000	2305345	5591752	8349138	1931709	5353372	8062690
Port area	m2	+	**100**	0	247	0	0	208	0
			300	0	2473	0	0	1786	0
			500	0	8142	0	0	7167	0
			1000	0	45285	0	0	52363	0
			5000	0	2428278	9225706	0	2116359	8617504
Airport area	m2	+	**5000**	0	136650	0	0	25216	783
Urban green area	m2	-	**100**	0	490	0	0	963	0
			300	0	7711	22125	0	10843	37904
			500	0	32633	132893	0	37932	152378
			1000	0	190110	579390	0	192426	551366
			5000	1372895	5233285	9714816	1122221	4590130	9281079
Natural and forested areas	m2	-	**100**	0	222	0	0	167	0
			300	0	2103	0	0	1588	0
			500	0	7244	0	0	6077	0
			1000	0	52377	165299	0	51926	144940
			5000	1334268	4944915	8182331	1328510	5172157	8159990
Residential land area	m2	+	**100**	0	26434	31375	0	25954	31375
			300	66208	226402	282618	47422	222740	282618
			500	231301	593328	785191	189582	579130	785191
			1000	842139	2082317	3050349	518003	1967973	3005396
			5000	15002050	28689475	46595685	11680371	27225930	46124341
Population density	n	+	**100**	7	245	524	15	270	561
			300	311	2117	4370	480	2096	4429
			500	1236	5469	11421	956	5208	10991
			1000	4805	19270	42504	2533	17452	39812
			5000	85535	251242	539468	71083	227169	531610
Household density	n	+	**100**	3	132	292	6	142	324
			300	144	1136	2476	200	1111	2553
			500	563	2940	6507	453	2773	6309
			1000	2079	10416	24315	1012	9380	22482
			5000	39485	134927	307588	32832	121927	303001
Traffic intensity on nearest road	Veh. day-1	+		82	8656	25785	30	4090	14943
Traffic intensity on nearest major road	Veh. day-1	+		5736	18232	36470	5649	18579	34240
Heavy-duty traffic intensity on nearest road	Veh. day-1	+		0	324	1005	0	125	420
Heavy-duty traffic intensity on nearest major road	Veh. day-1	+		67	982	1950	48	1206	1769
Road length of all roads	m	+	**50**	102	258	404	98	190	308
			100	488	838	1200	363	716	1059
			300	3997	6359	8494	3180	5925	8209
			500	9994	16603	21857	7734	15605	21532
			1000	33179	60248	80412	27773	55926	78365
Road length of all major roads	m	+	**50**	0	79	203	0	42	174
			100	0	194	507	0	107	390
			300	0	1030	2259	0	664	1760
			500	0	2600	4951	0	2080	4164
			1000	3161	9707	15328	2165	8471	14235
Traffic intensity on all roads (sum of (traffic intensity * length of all segments))	Veh. day-1m	+	**50**	40082	1257727	3400359	7963	757087	2207469
			100	185957	3417972	8691268	77072	2255473	5463763
			300	3136239	22142804	47797106	1315772	15681335	36067592
			500	11762571	58068646	120273812	5579975	46420718	98239045
			1000	63847576	226759175	452445437	31635299	196221848	417062321

Predictor variable	Units	Direction of effect	Buffer (m)	Mobile			Short-term Stationary		
Traffic intensity on all major roads (sum of (traffic intensity* length of all segments))	Veh. day-1m	+	**50**	0	1127175	3336962	0	688087	2185933
			100	0	2998942	8304429	0	1942440	4963848
			300	0	18940815	44304127	0	12729584	31697479
			500	4808393	49735352	108364418	0	38751720	89534363
			1000	44458796	196518241	410822257	18085365	168852113	377180297
Heavy-duty traffic intensity on all roads (sum of (heavy-duty traffic intensity* length of all segments))	Veh. day-1m	+	**50**	440	54021	141389	0	36679	92230
			100	3122	155742	378676	568	131698	213495
			300	72173	1121670	2637212	41345	760615	1541019
			500	328397	3078547	7798998	157767	2233684	5463944
			1000	2075563	13058286	26737180	1077166	10937816	23631745
Heavy-duty traffic intensity on major roads (sum of (heavy-duty traffic intensity*length of all segments)	Veh. day-1m	+	**50**	0	48165	134636	0	33903	92230
			100	0	137648	357616	0	120977	210302
			300	0	988569	2470260	0	651430	1335757
			500	106797	2727455	7255375	0	1931699	5094488
			1000	1133803	11678977	25328256	624627	9707896	21738224
Inverse Distance to nearest road [a]	m-1	+		na	na	na	0.073	0.488	0.756
Inverse Distance to nearest major road [a]	m-1	+		na	na	na	0.002	0.065	0.135

[a] Variables were not used for mobile model development, due to values being zero.

Table A.3

Distribution of UFP and BC in mobile and short-term stationary data set.

↓

Pollutant	Type of Measurements	No. of observations	Mean	10th percentile	Median	90th percentile
UFP (in particles /cm³)	Mobile	5236	18623	6850	14013	35512
	Stationary	240	12910	6453	11062	20663
BC (in µg/m³)	Mobile	5186	3.32	1.07	2.29	6.40
	Stationary	240	1.73	0.85	1.50	2.94

Figure A.1

Distribution of UFP and BC in mobile and short-term stationary data set.

↓

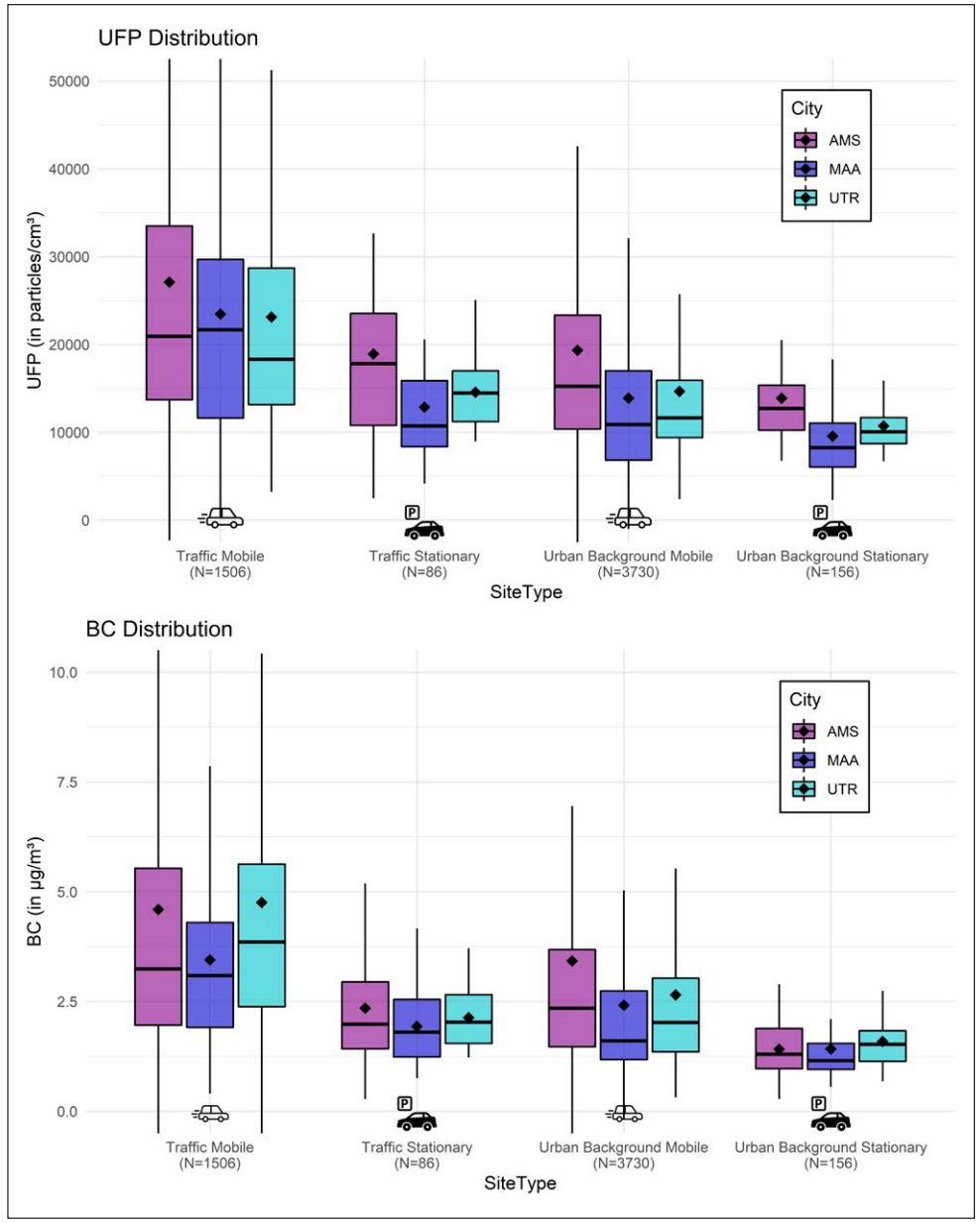


Table B.1

UFP Land-Use Regression Models based upon Mobile Measurements with and without AR-1 term and local exhaust plumes.

↓

Variable	Original (With AR-1 term and with peaks)	With peaks, without AR-1 term	With AR1- term, without Peaks	Without AR-1 term and without peaks	Fixed City effect model (Without AR-1 term)
Intercept	8072 (968)	9002 (578)	8296 (596)	8603 (413)	Amsterdam 7858 (1060) Maastricht 8950 (658) Utrecht 8719 (641)
Population Density:					
Residential Land Area in a 5000m buffer	7763 (1155) [a]	5591 (703)	4493 (710)	4182 (504)	5955 (758)
Traffic:					
Traffic Intensity on the Nearest Road	2244 (756)	3727 (656)	2755 (504)	2876 (462)	3760 (656)
Heavy Traffic Intensity on the Nearest Road	989 (536)	1790 (499)	878 (381)	952 (355)	1754 (501)
Major Road Length in a 100m buffer	4588 (524)	5057 (465)	1727 (476)	2445 (473)	5095 (469)
Major Road Length in a 300m buffer			2069 (587)	1656 (533)	
Land Use:					
Port Area in a 5000m buffer	3457 (995)	3882 (586)	4195 (594)	4525 (415)	4599 (812)
Urban Green Land in a 500m buffer	-1001 (494)	-1018 (354)			-926 (361)
Urban Green Land in a 1000m buffer				-803 (306)	
Number Road Segments used for model development	5236	5236	5164	5164	5236
Model R^2	0.15 [b]	0.15	0.18 [b]	0.19	/
Pearson Correlation with Original Model on 1500 Random Addresses	/	0.99	0.97	0.98	/

[a] Regression slopes and standard error (between brackets), multiplied by the difference between 10th and 90th percentile for all predictors to allow comparison of the effect of predictors with different units and distribution on measured concentrations. Predictions in particles/cm³.b R^2 of model without AR-1 term.

Table B.2

UFP Land-Use Regression Models from the 2014-2015 campaign based upon Mobile Measurements.

↓

Variable	Combined	Amsterdam
Intercept	8072 (968)	4053 (3015)
Population Density:		
Residential Land Area in a 5000m buffer	7763 (1155)[a]	8528 (1873)
Traffic:		
Traffic Intensity on the Nearest Road	2957 (740)	2817 (1141)
Heavy Traffic Intensity on the Nearest Road	989 (536)	
Heavy Traffic Intensity on the Nearest Major Road		1156 (633)
Traffic Intensity in a 100m buffer		
Major Road Length in a 100m buffer	4588 (524)	3100 (1034)
Land Use:		
Port Area in a 5000m buffer	3457 (995)	4911 (1993)
Urban Green Land in a 500m buffer	-1001 (494)	
Number Road Segments used for model development	5,236	1,991
R^2 of model compared to short-term stationary measurements	0.36	0.21

[a] Regression slopes and standard error (between brackets), multiplied by the difference between 10^{th} and 90^{th} percentile for all predictors to allow comparison of the effect of predictors with different units and distribution on measured concentrations. Predictions in particles/cm³.

Table B.3

UFP Land-Use Regression Models from the 2013 campaign based upon Mobile Measurements.

↓

Variable	Combined	Amsterdam
Intercept	5656 (2675)	-1254 (2974)
Population Density:		
Population density in a 5000m buffer	8064 (1947) [a]	8323 (2865)
Traffic:		
Traffic Intensity on Major Roads in a 100m buffer	1928 (1095)	5722 (1641)
Traffic Intensity in a 500m buffer	2917 (1514)	
Traffic Intensity in a 1000m buffer		7694 (2919)
Major Road Length in a 50m buffer	6868 (1071)	3884 (1567)
Land Use:		
Port Area in a 500m buffer		2102 (633)
Port Area in a 1000m buffer	2499 (1248)	
Airport Area in a 5000m buffer	4669 (1185)	
Natural Land in a 5000m buffer	-2557 (1357)	
Number Road Segments used for model development	2964	1427
R^2 of model	0.13	0.18

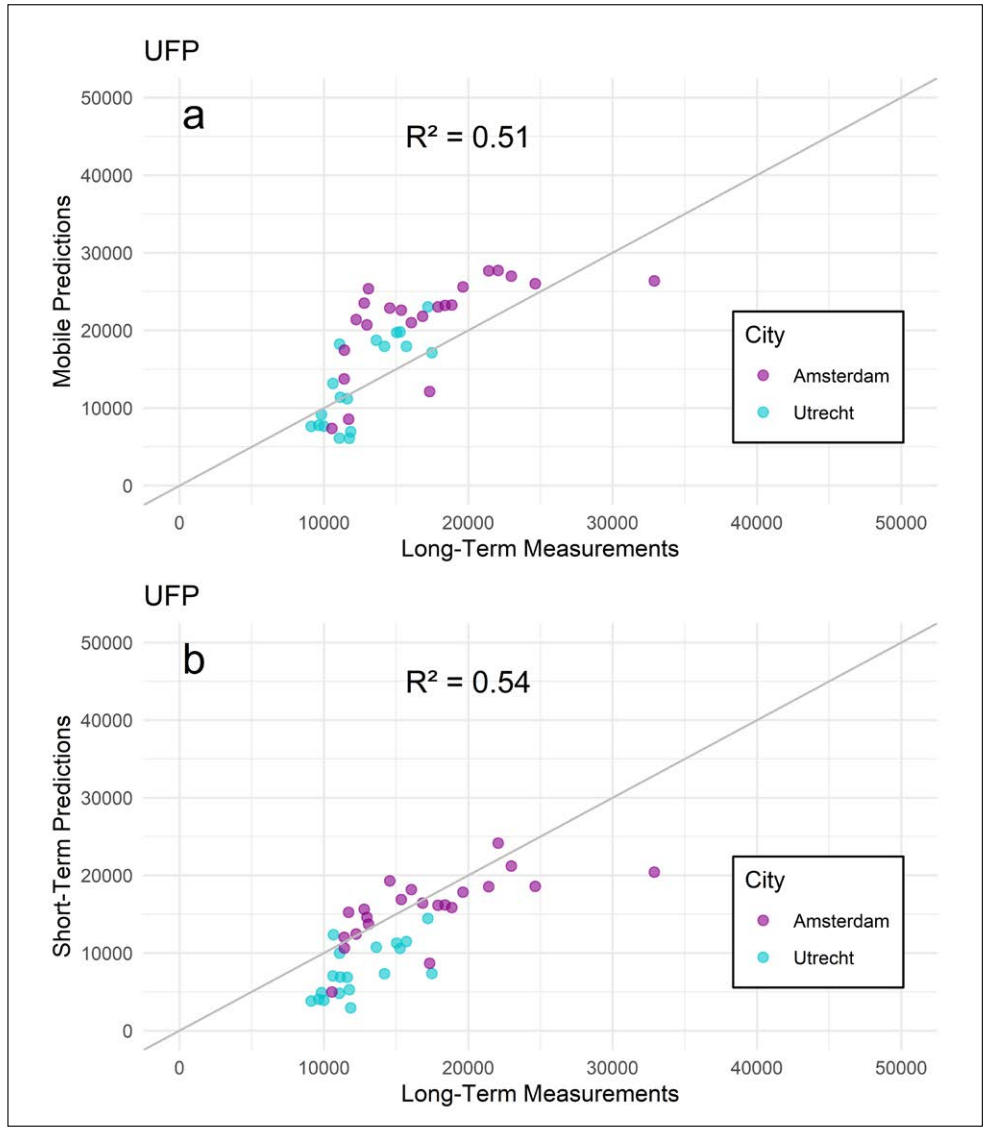
[a] Regression slopes and standard error (between brackets), multiplied by the difference between 10th and 90th percentile for all predictors to allow comparison of the effect of predictors with different units and distribution on measured concentrations. Predictions in particles/cm³.

[b] R^2 of short-term stationary model is between brackets.

Figure B.1

Predicted concentration levels at home outdoor sites (n=42) based on mobile UFP 2013 model (a) and short-term stationary UFP 2013 model (b) compared to 3 x 24h measurements from 2014/2015.

↓

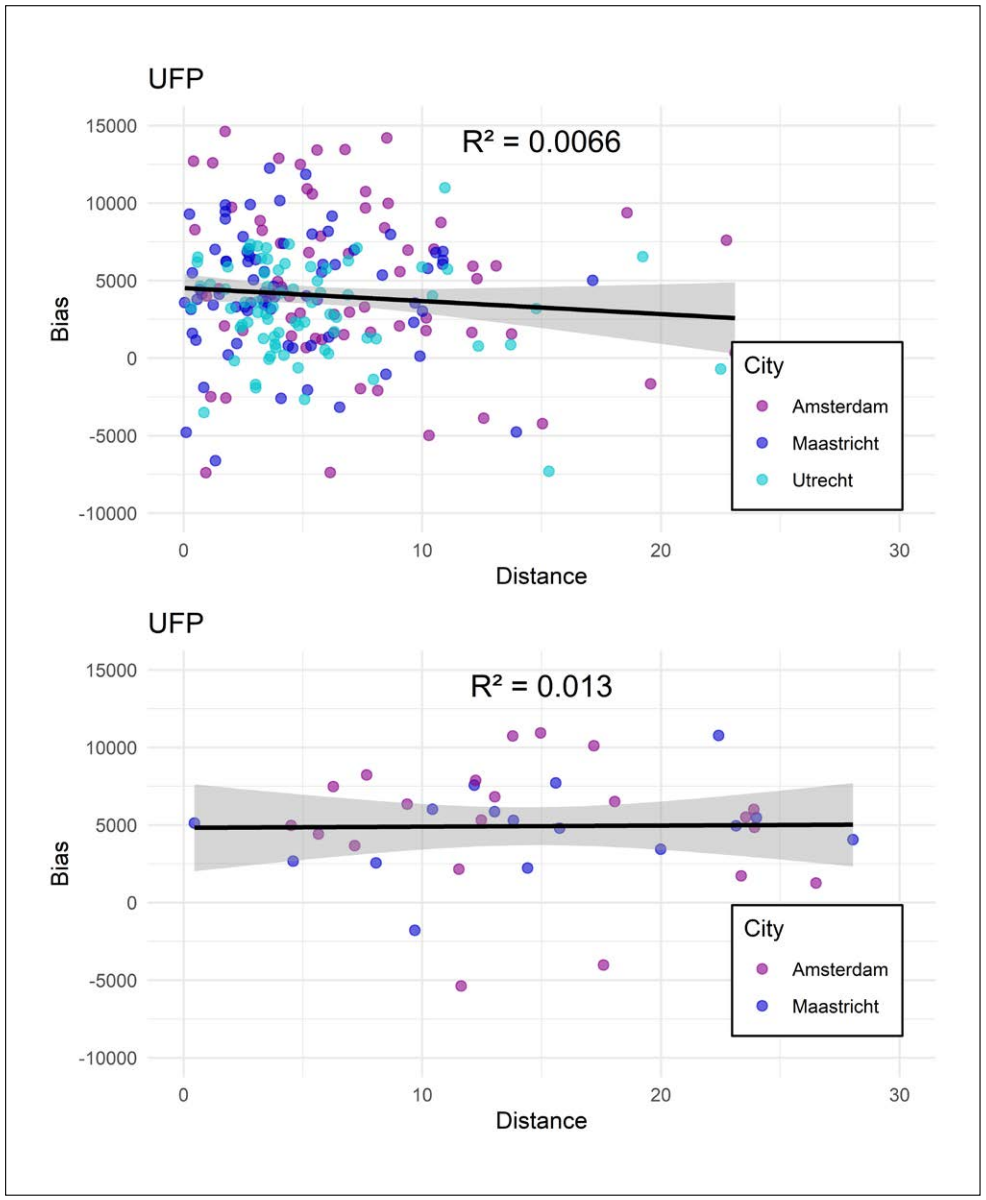


UFP levels in particles/cm³.

Figure B.2

Bias of predicted UFP counts on the short-term stationary sites (a) and home outdoor 24h sites (b) vs. distance of the measurement site to the nearest road.

↓



Bias UFP levels in particles/cm³.

Figure B.3

Mobile UFP predictions compared to short-term stationary UFP predictions on 1500 random addresses, with Bland Altman plots and for both measurement campaigns.

↓

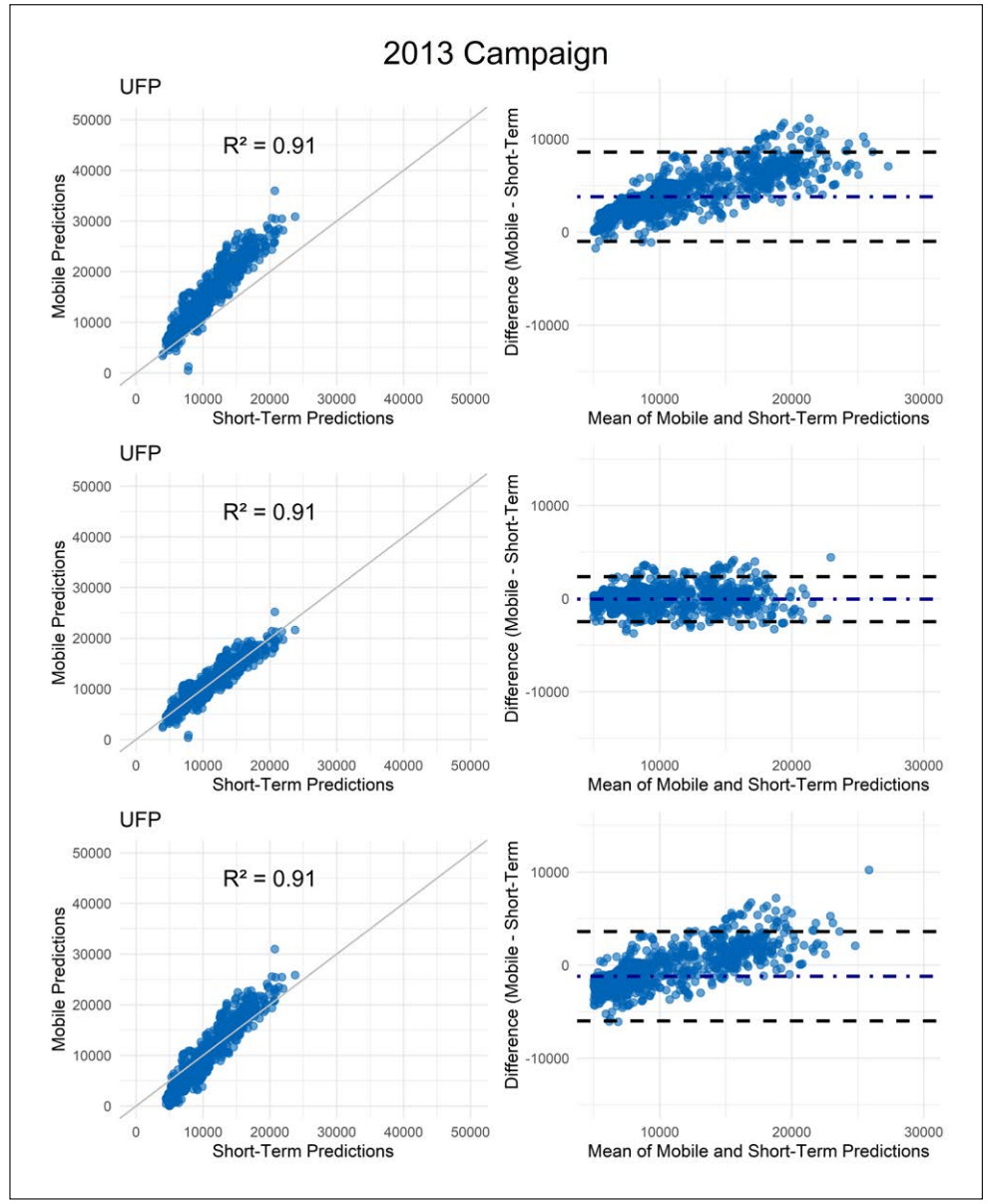


Figure B.3

Continued.

↓

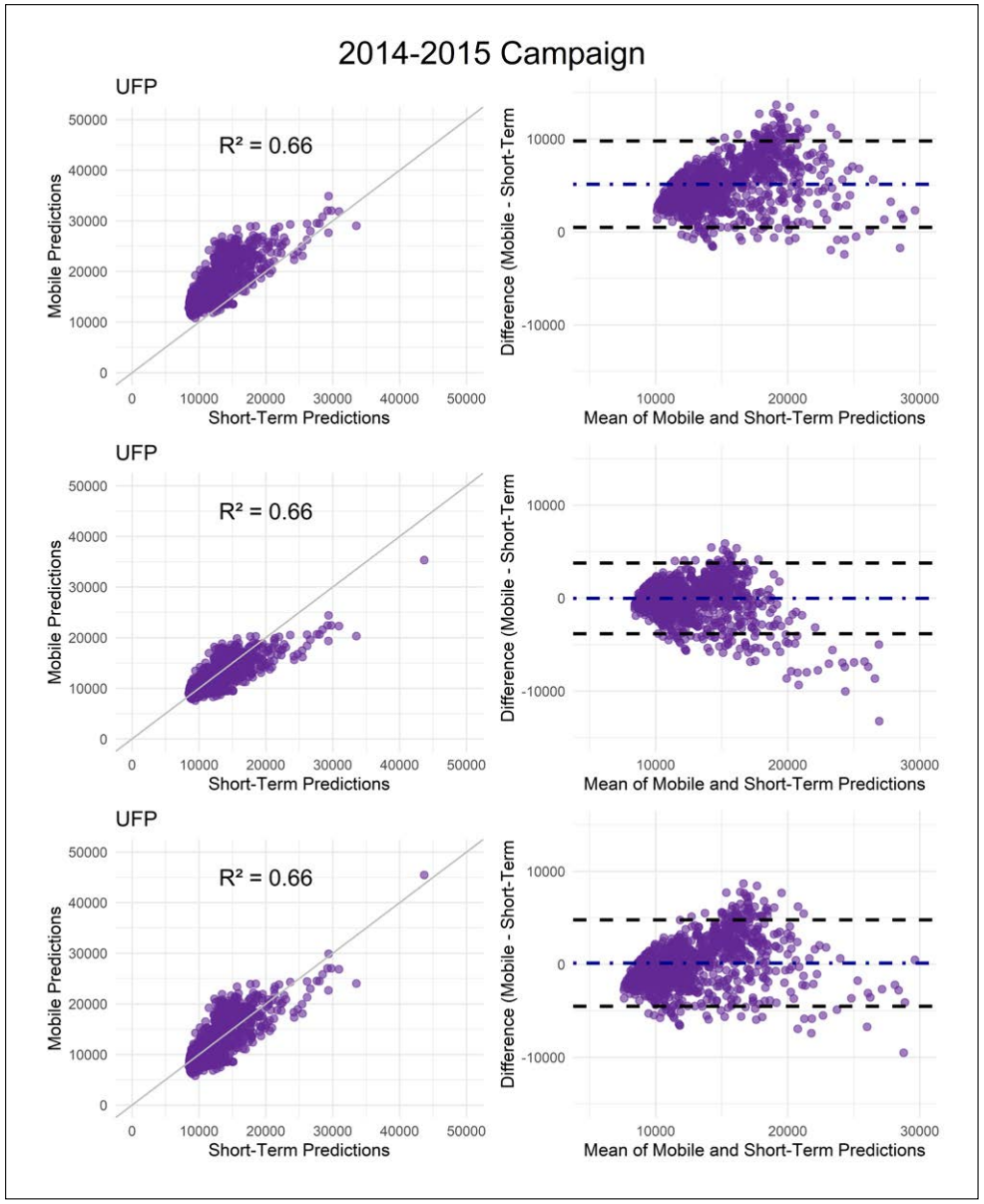
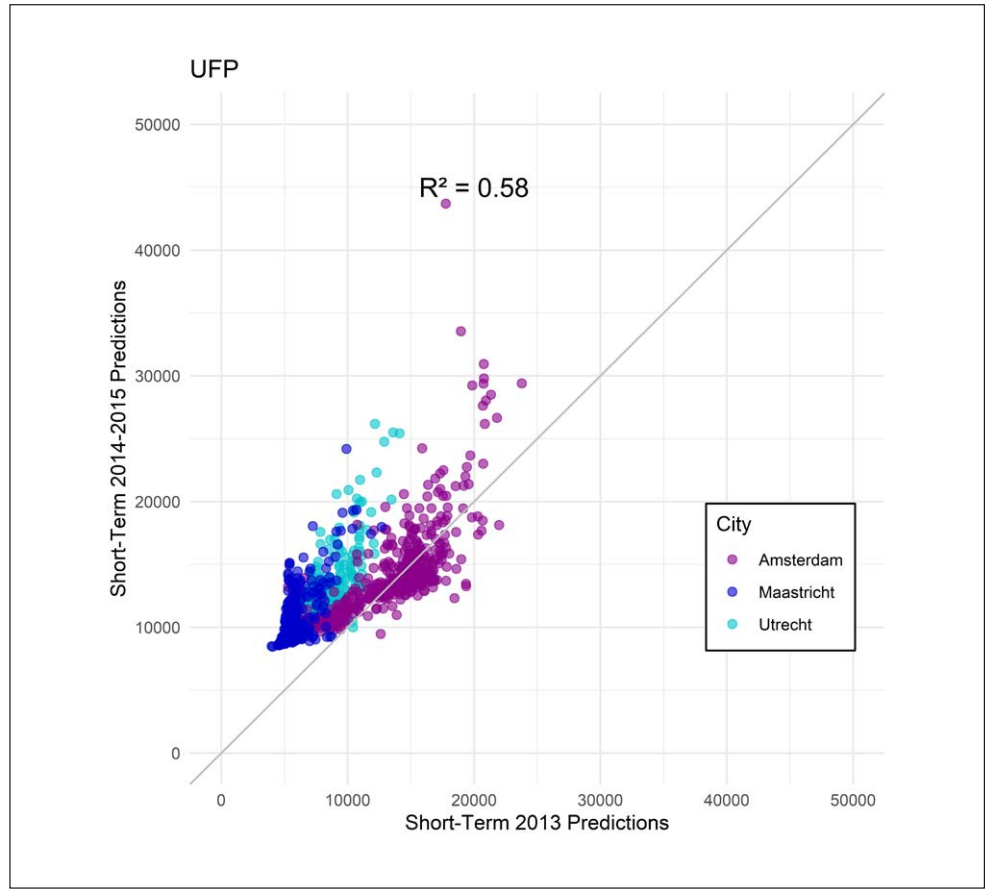


Figure B.4

Short-Term Stationary UFP predictions based on measurements from the 2013 campaign versus Short-Term Stationary predictions based on the measurements in 2014/2015.

↓

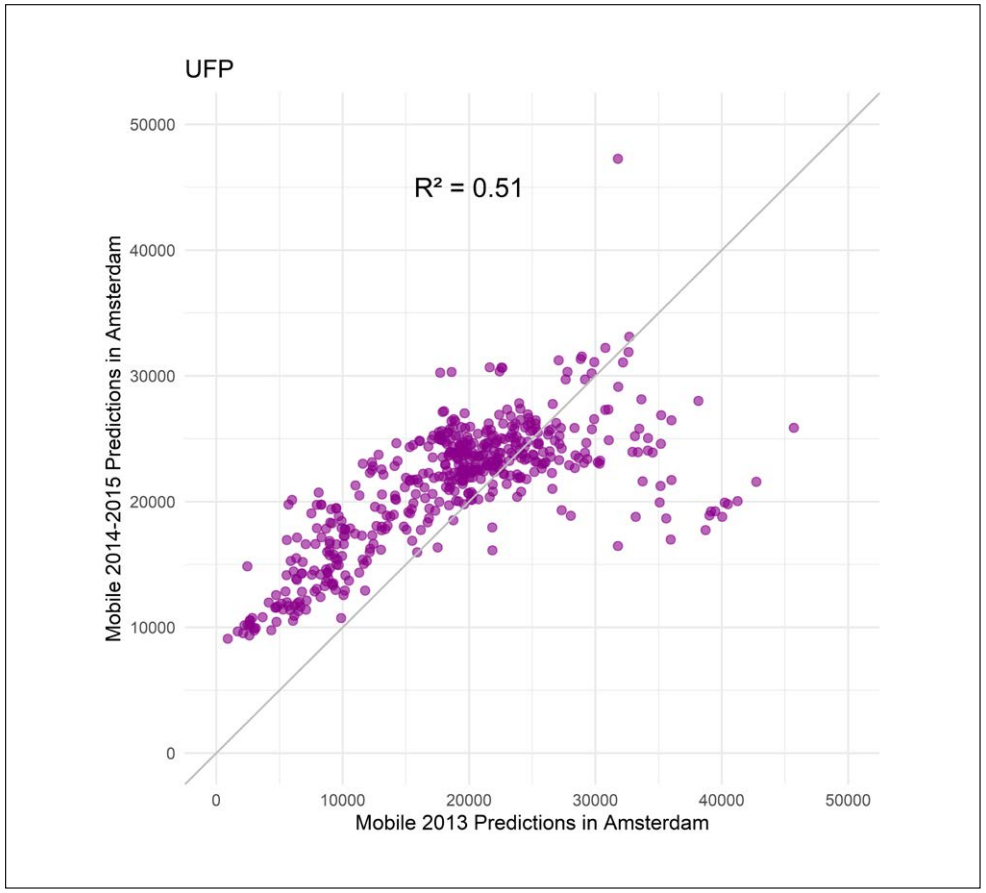


Predictions in particles/cm³.

Figure B.5

Comparison of predicted UFP counts based on mobile LUR models in 2013 and 2014/2015 in Amsterdam.

↓



Predictions in particles/cm³.

Table C.1 Appendix C: Black Carbon

Mobile and Short-Term Stationary BC Models.

↓

	BC (in µg/m³)	
Variable	**Short-Term**	**Mobile AR-1**
Intercept	1.20 (0.07)	1.00 (0.28)
Population Density:		
Household density in a 1000m buffer	0.33 (0.14)	
Residential land area in a 5000m buffer		2.43 (0.29)
Traffic:		
Traffic intensity on the nearest road	0.29 (0.12)	0.36 (0.10)
Traffic intensity in a 50m buffer	0.63 (0.12)	
Length of major roads in a 100m buffer	0.37 (0.13)	0.27 (0.10)
Land Use:		
Urban green in a 1000m buffer		-0.35 (0.16)
R^2 of model	0.44	0.10 [b]
Number sites used for model development	240	5,169

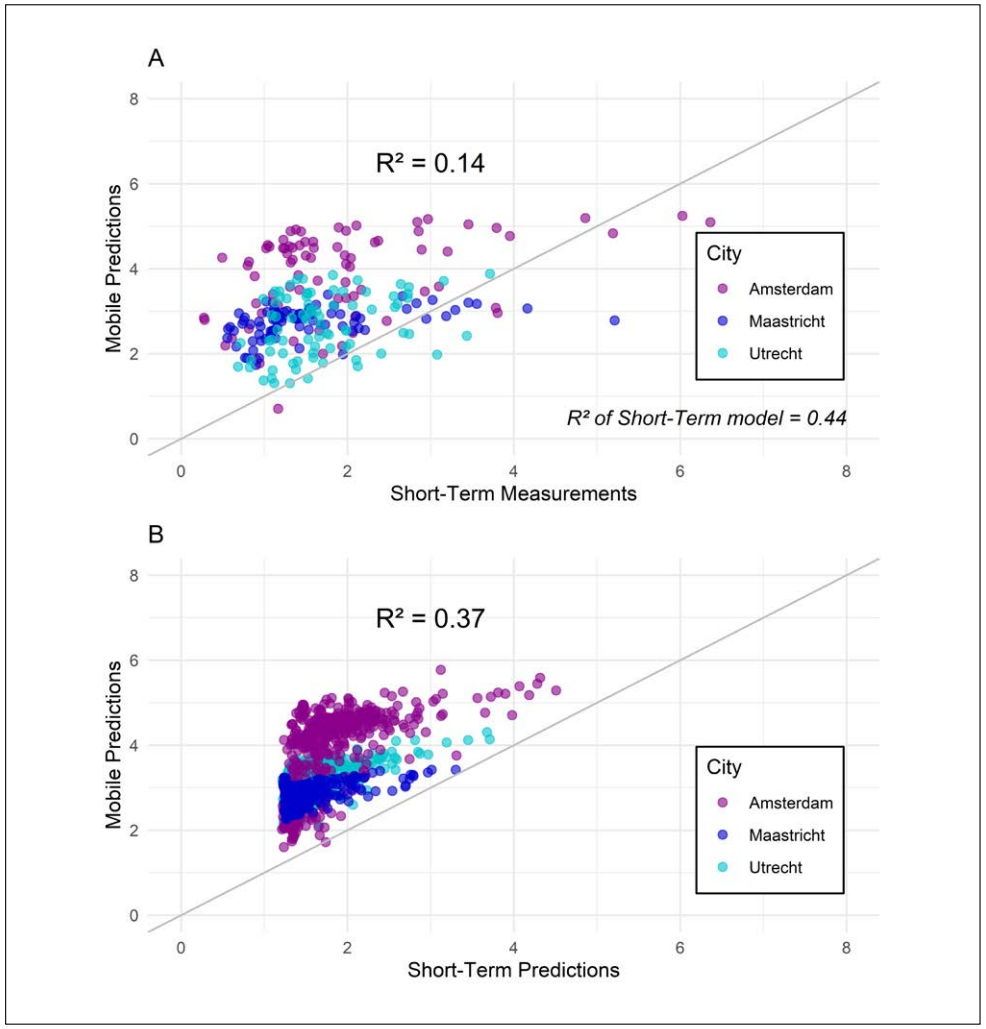
[a] Regression slopes and standard error (between brackets), multiplied by the difference between 10th and 90th percentile for all predictors.

[b] R^2 of model without AR-1 term.

Figure C.1

(a) Predicted concentration levels at stationary sites based on mobile LUR model compared to stationary measurements. (b) Comparison of predicted concentration levels based on mobile and stationary LUR models at 1,500 random addresses in Amsterdam, Utrecht and Maastricht.

↓

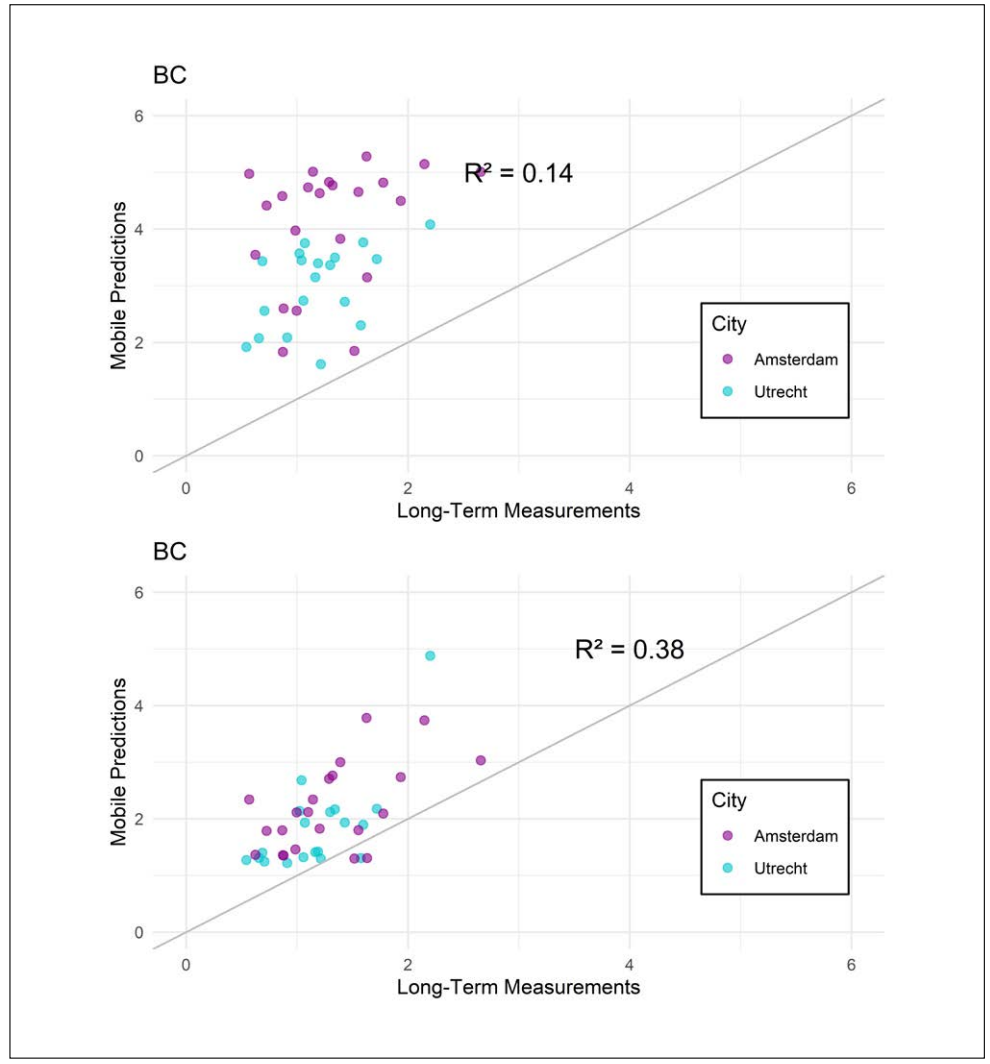


BC predicted levels in µg/m³.

Figure C.2

Predicted concentration levels at home outdoor sites (n=42) based on mobile models (a) and short-term stationary models (b) compared to 3 x 24h measurements.

↓

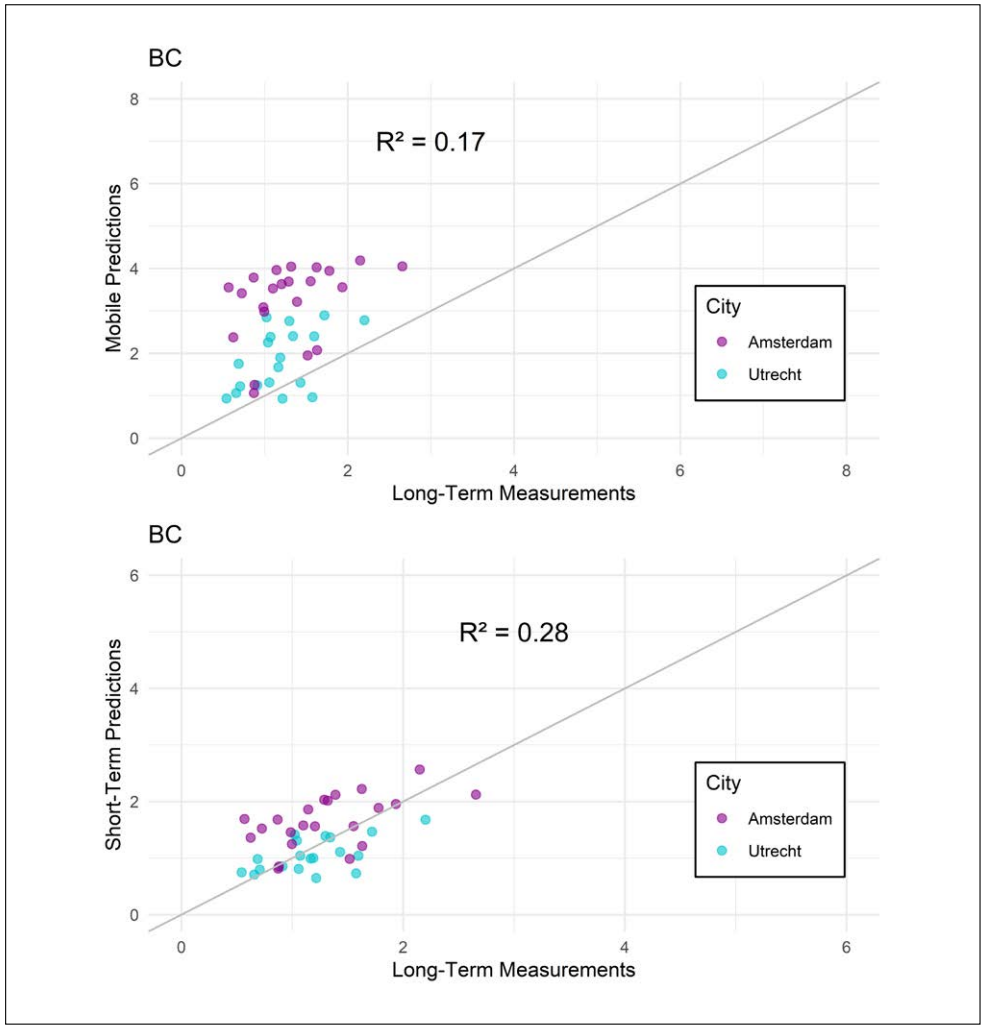


BC predicted levels in µg/m³.

Figure C.3

Predicted concentration levels at home outdoor sites (n=42) based on mobile BC 2013 model (a) and short-term stationary BC 2013 model (b) compared to 3 x 24h measurements from 2014/2015.

↓



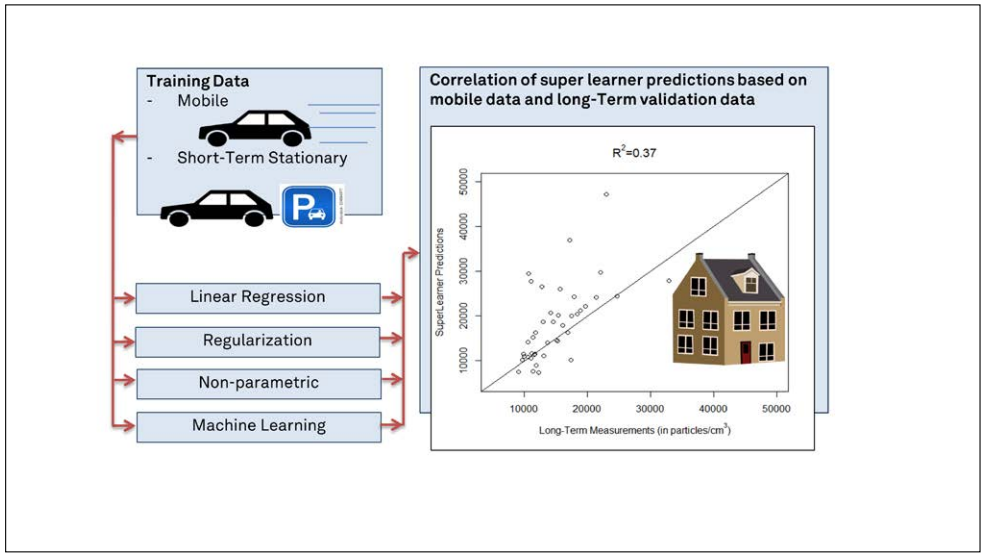
BC predicted levels in µg/m³.

Performance of Prediction Algorithms for Modeling Outdoor Air Pollution Spatial Surfaces

Jules Kerckhoffs
Gerard Hoek
Lützen Portengen
Bert Brunekreef
Roel C.H. Vermeulen

Abstract

Land use Regression (LUR) models for air pollutants are often developed using multiple linear regression techniques. However, in the past decade linear (stepwise) regression methods have been criticized for their lack of flexibility, their ignorance of potential interaction between predictors and their limited ability to incorporate highly correlated predictors. We used two training sets of ultrafine particles (UFP) data (mobile measurements (8200 segments, 25 seconds monitoring per segment), and short-term stationary measurements (368 sites, 3x30 minutes per site)) to evaluate different modeling approaches to estimate long-term UFP concentrations by estimating precision and bias based on an independent external dataset (42 sites, average of three 24-hour measurements). Higher training data R^2 did not equate to higher test R^2 for the external long-term average exposure estimates, arguing that external validation data are critical to compare model performance. Machine learning algorithms trained on mobile measurements explained only 38-47% of external UFP concentrations whereas multivariable methods like stepwise regression and elastic net explained 56-62%. Some machine learning algorithms (bagging, random forest) trained on short term measurements explained modestly more variability of external UFP concentrations compared to multiple linear regression and regularized regression techniques. In conclusion, differences in predictive ability of algorithms depend on the type of training data and are generally modest.



① Introduction

A common approach to predict spatial concentration levels of air pollutants is to use land-use regression (LUR) models. LUR modeling is an empirical technique with the measured concentration of a pollutant as dependent variable and potential predictors such as road type, traffic count, elevation and land cover as independent variables in a multiple regression model[1]. Model structures and the criteria for variables to be included in the model (variable selection) used in LUR modeling, though, differ between studies.

In a review of Hoek[2] it was found that most LUR studies use linear regression techniques that rely on least-squares or maximum likelihood fitting to develop prediction models. These models are simple, fast and often provide interpretable coefficients of predictors[3]. Forward, backward and best-subsets automatic selection methods are used in most of these settings. Often linear regression is applied with restrictions, typically allowing only slopes that conform with physical reality such as positive traffic intensity slopes. However, in the past decade linear (stepwise) regression methods have been criticized for the assumed linearity of predictor pollution relationships[3], limited inclusion of potential interactions and incorporating highly correlated predictors[4]. Next, standard linear regression methods are prone to overfitting, especially when few training sites are used for model development along with a large number of predictor variables[5,6]. Linear regression therefore may not identify the optimal model.

To overcome these issues new modeling techniques have been introduced into air pollution epidemiology[7]. For example, non-linear relationships can been obtained with general additive[8,9] or kernel based models[10]. Furthermore, machine learning techniques (such as neural networks[11] and random forests[12]) offer possibilities to create spatial models of air pollutants by learning the underlying relationships in a training data set, without any predefined constrictions. In a review by Bellinger et al.[7] on applications in epidemiology and references therein it was concluded that machine learning can be an effective tool for building accurate predictive models[13,14], especially with increasing size and complexity of data sets[7].

Comparisons between the performance of modeling techniques are scarce. Reid et al.[15] compared 11 modeling techniques for predicting spatiotemporal variability of $PM_{2.5}$ concentrations during wildfire events in California. General boosting, random forest and support vector machines performed better than standard linear regression modeling, LASSO and elastic net. Van den Bossche et al.[16] reported no significant differences

between linear regression, LASSO and support vector regression (SVR) to create LUR models for black carbon in a mobile monitoring campaign. Random forest performed better than stepwise selection to develop prediction models for elements in PM measured at 24 sites in Cincinnati, OH (Brokamp, 2017). Weichenthal et al.[10] found minor differences between a machine learning method (KRLS) and linear stepwise regression for mobile monitoring data in Montreal. KRLS had a higher training model R^2 compared to linear regression, but differences decreased when external data were used to compare predictions.

In previous studies we reported LUR-models for ultrafine particles (UFP) based on mobile monitoring[17,18] and short-term stationary measurements[19,20] developed with a supervised forward linear regression approach. Now, we use the UFP measurements and GIS predictors from both data sets to evaluate different modeling approaches by estimating precision and bias based on an external dataset. Since the main goal of the models is to predict concentrations for the use in epidemiological studies of chronic diseases, we used long-term measurements as the test set. These measurements were completely independent of the mobile and short-term measurements. As each modeling algorithm is likely not optimal, we also explored if a combination of models (stacking) could increase predictive performance. Stacking methods are commonly used to gain predictive power and to average out biases as compared to individual models[4]. In additional analysis we therefore used an algorithm (Super Learner[21]) that works by stacking results from a wide range of different modeling techniques.

② Materials and Methods

2.1 Data description

We combined data from two measurement campaigns; MUSiC (Measurements of UFP and Soot in Cities) and EXPOsOMICS (Combining Exposure and Omics data). The MUSiC campaign was conducted in 2013 and entailed both mobile and short-term air pollution monitoring. The EXPOsOMICS campaign (2014-2015) had a similar monitoring setup, but also included three repeated 24 hour home outdoor measurements in addition to the mobile and short-term stationary measurements. An overview of all data, also showing the spatial and temporal resolution is given in table 1. Measurements and models from the mobile MUSiC campaign[18], the short-term stationary MUSiC campaign[19], the mobile EXPOSOMICS campaign[17] and the short-term stationary EXPOSOMICS campaign[20] have been described in detail elsewhere.

In summary, in both campaigns short-term stationary and on-road UFP measurements were made using a condensation particle counter (TSI, CPC 3007) attached to an electric vehicle (REVA, Mahindra Reva Electric Vehicles Pvt. Ltd., Bangalore, India). UFP concentrations were recorded every second alongside a Global Positioning Unit (GPS, Garmin eTrex Vista). For better comparability between sites we avoided rush hour traffic and started sampling after 9:15 and stopped before 16:00 with about 8 short-term sites sampled each day. These sites were monitored for 30 minutes at 3 different days, spread over 3 seasons. Mobile measurements were collected when driving from one stationary site to the next and then averaged over the road segment the measurement was conducted on and all days the road segment was sampled on. A reference site with the same equipment as the electric vehicle was used to temporally correct all measurements. In both campaigns we used the difference method, which first calculates the overall mean concentration at the reference site over the entire campaign. Next, using the data at the reference site, the average reference concentration of 30 minutes around the time of sampling was subtracted from the overall mean. This difference was used to adjust the measured concentration at the sampling locations. Sites were selected with varying traffic intensities and land use characteristics. About 40% were traffic sites (>10,000 vehicles a day), 40% urban background sites and some were industrial sites and sites near urban green or rivers.

In this paper we combined the data sets from both campaigns, creating one mobile data set and one short-term stationary data set. Independent home outdoor measurements were collected during the EXPOsOMICS campaign. 42 outdoor home locations where repeatedly sampled for 24 hours during three different seasons. Measurements were averaged over all available seasons (at least two) after adjustment for temporal variation at a reference site. Measurements were performed with DiscMinis, which were found to have a ratio of almost 1 compared with co-located CPC 3007 measurements on three occasions in the measurement campaign[20].

Table 2

Overview of UFP monitoring data used for model development and validation.

↓

Data	Reference	Cities a	Duration	Sites b	Instrument	Year
MUSiC Short-Term	Montagne et al (2015)	Amsterdam and Rotterdam	3x30min	128	CPC 3007	Winter, Spring and Summer 2013
MUSiC Mobile	Kerckhoffs et al (2016)	Amsterdam and Rotterdam	¬20sec	2964	CPC 3007	Winter and Spring 2013
EXPOsOMICS Short-term	Van Nunen et al (2017)	Amsterdam, Maastricht and Utrecht	3x30min	240	CPC 3007	Winter, Spring and Summer 2014/2015
EXPOsOMICS Mobile	Kerckhoffs et al (2017)	Amsterdam, Maastricht and Utrecht	¬25sec	5236	CPC 3007	Winter, Spring and Summer 2014/2015
EXPOsOMICS Home Outdoor	Van Nunen et al (2017)	Amsterdam and Utrecht	3x24h	42	DiscMini	Winter, Spring and Summer 2014/2015

[a] Indication of size (inhabitants) of cities: Amsterdam: 820.000, Maastricht: 120.000, Rotterdam:620.000, Utrecht:330.000.
[b] For mobile measurements this refers to the number of road segments.

2.2 Statistical analysis

Land use regression models were developed for mean UFP concentrations per street segment in the mobile and per site for the short-term stationary data set separately. All model techniques were offered the same set of 72 GIS-derived predictors, shown in appendix S1. Because of the inoperable predictors related to inverse distance to road in the mobile data set the number of predictors is 70. These include a range of traffic variables, including traffic intensity and road length variables (in 50-1000m buffers), land use variables (e.g. port, industry, urban green, airports) and population density variables.

2.3 Comparing prediction algorithms

We compared several different model algorithms for creating LUR models, schematically shown in the TOC art. Stepwise regression is by far the most widely used prediction algorithm for developing land use regression models for air pollution[22]. We selected different forms of stepwise regression (Forward, Backward, Both and a Supervised method)[6,21]. Note that "supervised" in this case means a stepwise procedure that is customized and not the fact that the input and output variables are known, a definition that is more common in machine learning terminology. We keep the term supervised as an option for stepwise regression for its extensive use in literature. Supervised regression approaches usually constrain slopes of predictors to a predefined direction. Related to stepwise regression is the Deletion Substitution Addition (DSA) algorithm[6,24]. The difference with stepwise regression is that model complexity is determined by cross-validation and the final model is selected by minimizing the residual mean squared error (RMSE) using cross-validation[24]. To deal with possible correlation of predictors we selected regularization techniques (LASSO, RIDGE and Elastic Net). They can deal with correlated predictors by imposing a penalty on the absolute size of regression coefficients. Non-linear methods are represented in our analyses by a log-transformed generalized linear model (GLM)[25], multivariate adaptive regression splines (MARS) and generalized additive modeling (GAM)[8,9,26]. More flexibility in the association between predictors and pollution can be obtained by data mining and machine learning methods. We used kernels (Kernel-Based Regularized Least Squares (KRLS)[10]), neural networks[11], support vectors (support vector machines (SVM)[27]), or combining weak classifiers/trees in an ensemble (random forests[12], boosting and bagging). Algorithms are described more extensively in the supporting information (SI.2.1-SI.2.11).

In an additional analysis we combined all models in a 'stacking' method called Super Learner, which selects and combines different prediction algorithms. Each algorithm is assigned a weight, which is based on the difference between the training model R^2 and cross validated R^2. The super learner then uses a weighted combination of the algorithms which were assigned the highest weight. The supervised stepwise approach, DSA and KRLS are not supported in the super learner algorithm and are therefore not included in this analysis.

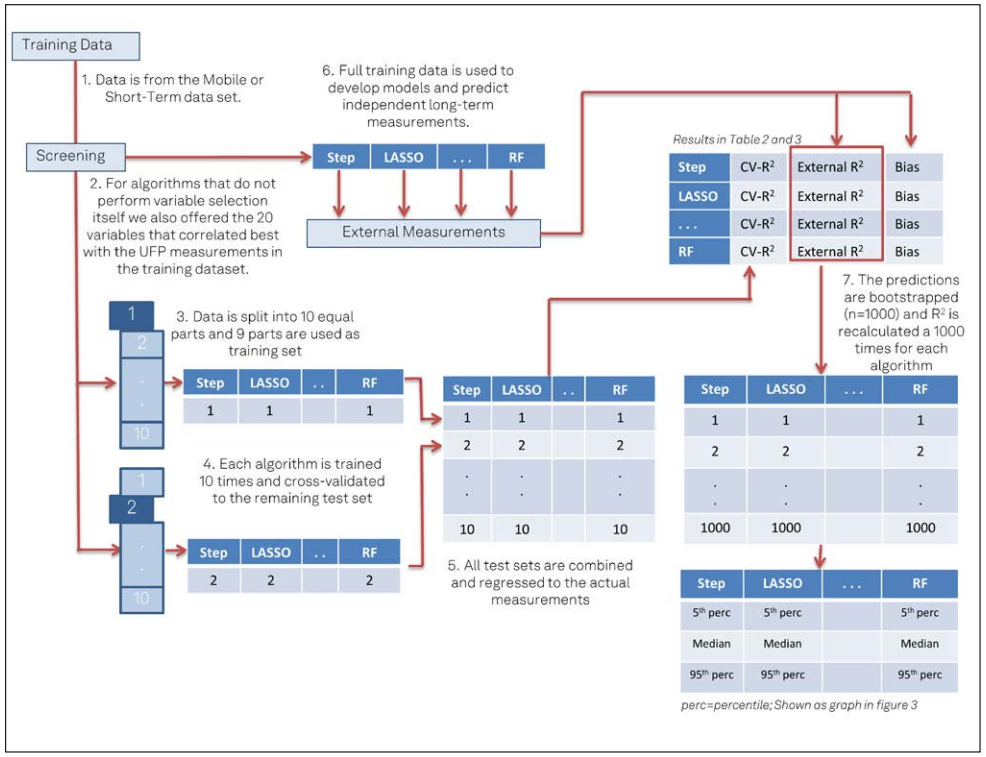
We present the cross-validated training model R^2, and the R^2 value and bias of model predictions compared to the external measurements (Figure 1). We used 10-fold cross-validation with repeated random selection of training and test sets. In each cross-validation set models were redeveloped, creating 10 models for each algorithm. Predictions were combined and then regressed to the actual measurements. Training model R^2 (without cross-validation) can be found in the supporting information. To investigate the precision of our external test set regression coefficients, we bootstrapped (n=1000) all external test set model predictions and show the median, IQR and 90% probability interval of the estimated R^2 values. Specifically, we randomly selected 42 sites from the original 42 sites with replacement, thus allowing sites to be included multiple times in the

bootstrap sample. This procedure has been shown to give correct estimates of the precision of the R^2 [4]. We also ranked all models according to their median R^2 values based on the 1000 bootstrap samples in sensitivity analyses. Bias was calculated as the mean difference between the external measurement and the predicted concentration at the sampling location.

Figure 1

Flow Diagram of analysis, with 'STEP', 'LASSO' and 'RF' as example algorithms (stepwise regression, LASSO and random forest).

↓



2.4 Variable Selection

When a certain algorithm does not perform variable subset selection, exposure assessment studies often use screening of variables. This could be a maximum number of variables to be able to enter the model[6,24], stepwise regression screening[9,10] or a preselection of variables[8,28]. For algorithms that do not perform variable selection we also offered the 20 variables that correlated best with the UFP measurements in the training dataset. We selected 20 as the cut-off because it restricts the number of variables considerably (by a third) and algorithms that do variable selection never included more than 20 variables in their model (Table S3). For sensitivity analysis (Tables S4 and S5) we also offered other amounts of variables to the models, one based on the absolute value of the Pearson correlation coefficient (variable included when it correlates at least 0.1 to the UFP measurement) and by first using a stepwise regression procedure (variables from the customized stepwise regression (SI.2.1)). The overview of GIS-derived predictors (Table S1) shows which variables pass a specific screening method.

③ Results

3.1 Distribution

Figure 2 and table S2 show the distribution of UFP measurements collected while driving (n=8200) and during the short-term stationary (n=368) and long-term stationary measurements (n=42). Mean concentration collected while driving is about 50% higher than measurements collected on short-term measurement sites. Especially the frequency at which high concentrations occur is higher while driving (90^{th} percentile of mobile measurements is almost double the 90^{th} percentile of stationary measurements, whereas the 10^{th} percentiles are similar in the mobile and short-term data sets). The external validation set had a similar variation of UFP data as the short-term data, but the average of UFP was slightly higher (about 2000 particles/cm³).

3.2 Comparing prediction algorithms

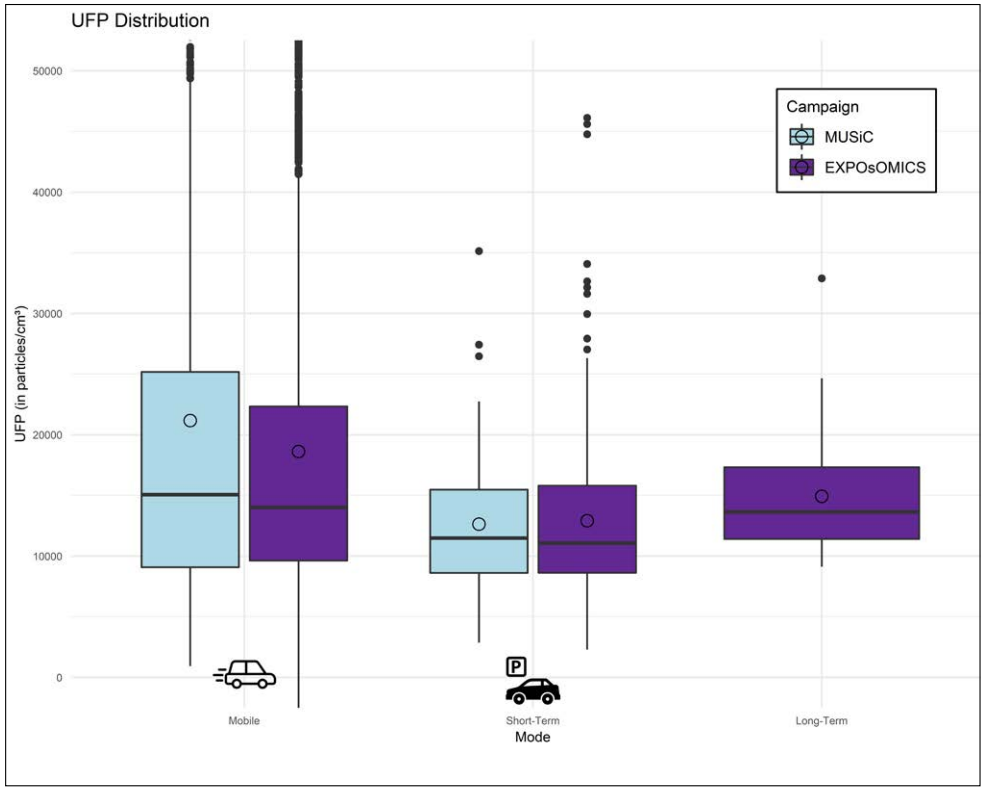
Table 2 (mobile data) and table 3 (short-term stationary data) shows the performance of all individual algorithms (with default settings used in the super learner algorithm). Pages SI.2.1 to SI.2.11 provide scatterplots of measured versus predicted concentrations for selected algorithms, as well the settings used in the various algorithms. Variables selected in models that perform variable selection can be found in table S3.

In general, the highest training model CV-R^2 in the mobile data set were based on some of the data mining approaches, such as support vector machines, boosting and random forest (CV-R^2 range of 0.22-0.24; table 2). The widely used stepwise regression algorithm predicted (only) 13% of the spatial variance of the mobile measurements.

Figure 2

Distribution of site averaged concentrations in the mobile and short-term training datasets and the external long-term validation dataset (in particles/cm³).

↓



Distributions extent from 5th to 95th percentile with the IQR in the solid box and the circle representing the mean.

Table 2

Comparison of prediction algorithms with Mobile UFP data.

↓

			Training	External prediction	
	Algorithm	**Variables [a]**	**Model CV- R^2**	**R^2**	**Mean Bias [b] (in particles/cm³)**
Linear regression	**Supervised Forward [c]**	All	0.13	**0.62**	6902
	Stepwise Both (R^2)	All	0.14	**0.54**	3525
	Stepwise Both (AIC)	All	0.13	**0.53**	3307
	Stepwise Backward (R^2)	All	0.14	**0.54**	3525
	Stepwise Forward (R^2)	All	0.15	**0.60**	3293
	Deletion/Substitution/ Addition (DSA) [c]	All	0.14	**0.56**	6866
Regularization	**LASSO**	All	0.13	**0.62**	3374
	Elastic net (alpha=0.25)	All	0.13	**0.63**	3182
	Elastic net (alpha=0.50)	All	0.13	**0.63**	3163
	Elastic net (alpha=0.75)	All	0.14	**0.62**	3366
	Ridge	All	0.14	**0.63**	3144
		Top20	0.12	**0.61**	4367
Non-Linear	**Generalized Linear Model (GLM)**	All	0.12	**0.58**	-518
		Top20	0.11	**0.57**	348
	Multivariate Adaptive Regression Splines (MARS)	All	0.12	**0.41**	3261
		Top20	0.11	**0.43**	3750
	General Additive Model (GAM)	All	0.17	**0.48**	2978
		Top20	0.14	**0.55**	4782
Data mining	**Kernel Based Regularized Least Squares (KRLS) [c]**	Regression Screening	0.18	**0.46**	5993
	Neural network	All	0.06	**0.26**	4431
		Top20	0.05	**0.28**	4519
	Support Vector Machine	All	0.23	**0.26**	1132
		Top20	0.17	**0.45**	1915
	Random forest	All	0.24	**0.40**	5688
		Top20	0.22	**0.45**	5501
	Gradient Boosting Machine (GBM)	All	0.15	**0.49**	4268
		Top20	0.14	**0.50**	4561
	Extreme Boosting	All	0.23	**0.31**	5533
		Top20	0.18	**0.42**	5947
	Bagging	All	0.16	**0.46**	4088
		Top20	0.15	**0.46**	4795
	Super Learner	All	/	**0.37**	3383

[a] For algorithms that do not perform variable selection itself we also offered the 20 variables that correlated best with the UFP measurements in the training dataset. [b] Bias = Predicted concentration – Measured concentration.
[c] Algorithm is the only algorithm that considers direction of effect and is not included in the Super Learner method.

Data mining algorithms, however, generated lower R^2 values (max R^2=0.50) when external long-term measurements were predicted. Regularized regression techniques (LASSO, elastic net and Ridge), supervised stepwise regression and GLM were able to explain external measurements equally well (R^2 0.61-0.62).

In the short-term stationary data set most data mining methods, linear regression and regularization algorithms predicted the short-term measurements equally well (R^2 range of 0.26-0.39, Table 3). In the short-term stationary data set, some of the data mining algorithms explained variation in the external validation data by 58-70% (Table 3) compared to 53% for the customized stepwise regression model. Prediction of external measurements improved for most algorithms when only 20 predictors were offered. Short-term stationary predictions were much less affected by bias as the mobile predictions. Mobile predictions are on average 3900 particles/cm³ higher than actual measurements, whereas the bias in the short-term prediction is close to zero (-260 particles/cm³).

Figure 3 and figure S1 illustrate the precision of predictive performance at the 42 external sites in 1000 bootstrap samples. Most algorithms showed an IQR of about 10% to 15% difference in R^2 (solid boxes in figure 3) with increasing variability with more flexible model algorithms, meaning that such algorithms are more dependent on the sites used for validation. The precision illustrates that the performance of algorithms based on regularization, (supervised) linear regression, DSA and some machine learning methods does not differ consistently. These algorithms ranked in the top ten (out of 31 algorithms) in almost all bootstrapped samples (Figure S1). Figures SI.2.1 to SI.2.10 illustrate that some of the modest difference in R^2 between models is partially due to how well the highest measured concentration is explained. Deletion of this point from the external data set diminished the difference in predictive performance between the algorithms (Figure S2).

The super learner 'stacking' algorithm did not improve upon the individual algorithms in our study. In the mobile monitoring data, model techniques that exploited training data with great detail (such as SVM, random forest and extreme boosting) were chosen in the Super Learner (Table S4). These algorithms were individually able to generate a high training model R^2 but are unable to predict external long-term average concentrations. For example, the R^2 was 0.37 for the super learner algorithm, while regularization methods were able to predict external measurements by 62-63%. For the short-term stationary data different prediction algorithms were chosen in the super learner algorithm: stepwise backward regression, regression splines (with screening) and extreme boosting (Table S5).

Table 3

Comparison of algorithms with Short-Term Stationary UFP data.

↓

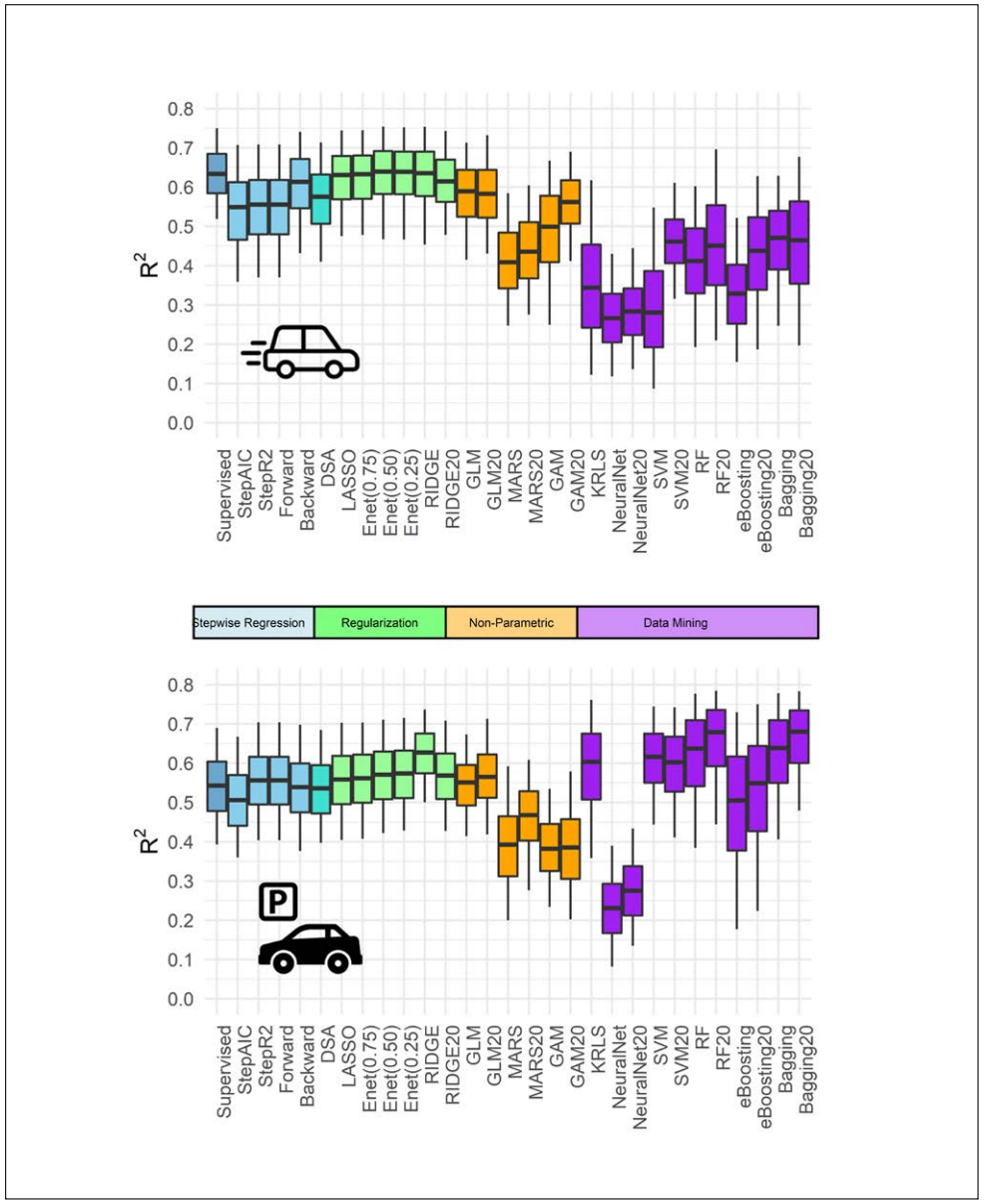
			Training	External prediction	
	Algorithm	**Variables [a]**	**Model CV- R^2**	**R^2**	**Mean Bias [b] (in particles/cm³)**
Linear regression	**Supervised Forward [c]**	All	0.35	**0.53**	151
	Stepwise Both (R^2)	All	0.36	**0.54**	428
	Stepwise Both (AIC)	All	0.29	**0.49**	-88
	Stepwise Backward (R^2)	All	0.36	**0.54**	428
	Stepwise Forward (R^2)	All	0.36	**0.52**	322
	Deletion/Substitution/ Addition (DSA) [c]	All	0.35	**0.51**	508
Regularization	**LASSO**	All	0.30	**0.55**	-167
	Elastic net (alpha=0.25)	All	0.30	**0.56**	-389
	Elastic net (alpha=0.50)	All	0.30	**0.56**	-250
	Elastic net (alpha=0.75)	All	0.30	**0.55**	-207
	Ridge	All	0.28	**0.61**	-142
		Top20	0.31	**0.55**	-30
Non-Linear	**Generalized Linear Model (GLM)**	All	0.23	**0.53**	-249
		Top20	0.28	**0.55**	-892
	Multivariate Adaptive Regression Splines (MARS)	All	0.04	**0.38**	740
		Top20	0.33	**0.45**	101
	General Additive Model (GAM)	All	0.03	**0.37**	475
		Top20	0.02	**0.37**	973
Data mining	**Kernel Based Regularized Least Squares (KRLS) [c]**	Regression Screening	0.39	**0.60**	-625
	Neural network	All	0.06	**0.22**	-1283
		Top20	0.10	**0.27**	-1166
	Support Vector Machine	All	0.32	**0.61**	-796
		Top20	0.33	**0.60**	-731
	Random forest	All	0.30	**0.66**	-229
		Top20	0.33	**0.70**	-54
	Gradient Boosting Machine	All	0.31	**0.66**	-1225
		Top20	0.32	**0.64**	-1142
	Extreme Boosting	All	0.26	**0.51**	-978
		Top20	0.29	**0.56**	-775
	Bagging	All	0.31	**0.65**	-566
		Top20	0.34	**0.70**	-347
	Super Learner	All	/	**0.60**	63

[a] For algorithms that do not perform variable selection itself we also offered the 20 variables that correlated best with the UFP measurements in the training dataset. [b] Bias = Predicted concentration – Measured concentration.
[c] Algorithm is the only algorithm that considers direction of effect and is not included in the Super Learner method.

Figure 3

Distribution of R-squares when predictions of Mobile (top) and Short-Term Stationary (bottom) models are compared to external measurements, based on 1000 bootstrapped samples.

↓



④ Discussion

We compared different algorithms for developing land use regression models of spatial concentration variations of UFP based on mobile and short-term stationary monitoring campaigns. In general, we found modest differences in performance when external measurements were used for validation. In our mobile data set algorithms based on regularization (LASSO, elastic net and Ridge), (supervised) linear regression and DSA explained variance in the long-term measurements almost equally and slightly better than data mining approaches. In the short-term data set however, data mining approaches, such as random forest, boosting and bagging, explained variance in external measurements slightly better than stepwise regression and regularization techniques. Algorithms based on regularization (LASSO, elastic net and Ridge), (customized) linear regression and DSA explained variance in the long-term measurements almost equally. The super learner algorithm did not improve external prediction in both data sets but performed reasonably well when short-term stationary data was used and had marginally lower bias estimates.

4.1 Comparing prediction algorithms

All algorithms in both the mobile and short-term stationary data set showed higher external prediction performance than the (cross-validated) performance in the training data (Tables 2 and 3). Training data was based on mobile or short-term data, with a resolution of ~25 seconds and 3x30minutes of sampling per site respectively. Hence, training data contains more noise (with mobile measurements more so) than the long-term measurements of 3 times 24 hours used for validation. We previously discussed and explained the higher R^2 in validation samples[17,19,20].

For algorithms with a higher cross-validated training model R^2, we did not always find a higher R^2 for the external long-term average concentrations, stressing the importance of an external validation data set reflecting the temporal resolution of what one wants to predict. For example, the random forest algorithm, which predicted mobile measurements best (24%), predicted 40% of the variance in the external measurements (Table 2), whereas the customized stepwise regression procedure, regularization methods and DSA predicted 13-15% of the mobile measurements and 61-62% of the external measurements. Data mining methods train models in great detail, possibly assigning too much value to individual mobile measurements and create models that reflect patterns that are not present at the external sites using longer averaged measurements. We note that our study

developed models based on training data with a different time scale than the validation data. The data further differ in their spatial features: on-road mobile monitoring versus residential outdoor monitoring, typically near the façade of homes. The validation data are important because they represent the locations where the models are mostly applied for epidemiological studies. Our study setting is therefore relevant for other studies based on mobile monitoring as well. We note that the data mining methods were developed using internal cross-validation using randomly selected sets. Auto-correlation present in the mobile monitoring set may have contributed to some overfitting of the models. We therefore performed an additional cross-validation analysis based on clusters of driving days and found that the CV-R^2 dropped for all methods (shown as clustered CV-R^2 in table S4). Models that are used for external prediction are based on the full data set, it did not alter external prediction performance of individual algorithms.

In the short-term data, several machine learning algorithms (bagging, random forest) trained on short term measurements explained modestly more variability of external UFP concentrations compared to linear regression and regularization techniques. We did not observe higher training model R^2 for the data mining methods. The short-term stationary measurements are an average of 3 times 30 minutes and hence averages are more stable (less noise) than those based upon the mobile measurements. This may explain that certain features picked up by data mining algorithms were also helpful to predict long-term measurements.

This is also reflected in the algorithms that are selected in the super learner algorithms. In the mobile data three machine learning algorithms are chosen, whereas stepwise backward regression is given the most weight in the short-term training data. These algorithms are chosen because the CV-R^2 shows that they do not overfit the training data. In other words, the predictive ability of that particular algorithm in the training data is high, and the risk of choosing that algorithm is low.

In both datasets, (supervised) linear regression model and regularization approaches provided similar and relatively high external prediction R^2. These algorithms were also found to describe data equally well in a study by Van de Bossche et al.[16], evaluating LUR models for black carbon based on mobile measurements. LASSO, Ridge and elastic net can (better) deal with correlated predictors as they do not use a custom variable selection method but a regularization penalty[16]. Compared to data mining algorithms, an advantage of the regularization algorithms is that they keep their interpretability, similar as linear stepwise regression procedures, and in addition can perform variable selection not necessitating any a-priori selection of variables. We note that a screening step based on simple correlation between predictors and pollution, takes away some of the attractiveness of applying sophisticated data mining methods.

We have focused extensively on explained variance (R^2) of the models, as generalization of exposure contrast in an epidemiological study is a key goal of application of these models. We additionally asses bias. Models based upon the mobile monitoring but not the short-term monitoring overestimated long-term outdoor home façade measurements for most model algorithms. In previous studies we observed an average overestimation of 5000 particles/cm^3 (30%) when models were developed with mobile monitoring data[17] and concluded that this is due to the fact that mobile measurements were taken on the road itself while short-term and long-term stationary measurements were taken on the sidewalk near the façade of homes. This is similar to a study by Simon et al.[29], who compared UFP concentrations in Chelsea (Massachusetts, USA) based on residence and mobile monitoring. They found average differences of 5300 particles/cm^3, which is about the same as found in our studies.

Although no strong conclusions can be drawn regarding which algorithm is best, some recommendations can be given. The fact that most algorithms generate similar predictions, above all, is encouraging. Predictors that are selected or given the most weight in our exposure models are also similar. One could argue that with more robust training data (i.e. longer-term average), machine learning algorithms tend to operate better. Correlation structures in the data are probably more important. When the relationships between UFP and predictor variables are not complex (no major interactions and no major deviations from linear associations), there is not much to gain for machine learning algorithms opposed to 'simple' linear regression. More detailed predictor data could benefit certain algorithms as well. Computational runtime is not a limiting factor. Individual algorithms never took more than 15 minutes to run on a standard desktop computer. Total runtime of the super learner algorithm was about 12 hours for the mobile data and 2 hours for the short-term data. We therefore recommend using multiple approaches (that differ in model structure) in future studies. An external dataset to test predictions may be valuable when training data differ in temporal resolution or when spatial features differ from the locations to which the models are applied (short term on road mobile monitoring versus long-term average residential address exposure). Such long-term UFP measurements as a test set of sufficient size is however difficult to obtain.

4.2 Limitations

Although the availability of long-term external residential level validation data at 42 sites is a major strength not typically available in other studies

on UFP, we acknowledge that the limited number of sites requires that differences in external R^2 values between algorithms must be interpreted with caution. We therefore used a bootstrapping approach to quantify the precision of the R^2 and interpret the differences as modest. Most algorithms showed an IQR of about 10% to 15% difference in R^2 (solid boxes in figure 3) with increasing variability with more flexible model algorithms, meaning that such algorithms are more dependent on the sites used for validation.

On top of that, it is possible that different settings (instead of default parameters) for certain algorithms can change the predictive performance. If a random forest or boosting model would be trained with parameters that are constrained in such a way that the model does not try to aggressively explain every single observation in the training data, it could improve external prediction performance. In the extreme case of extreme boosting, training model R^2 (without cross-validation) can go all the way to 1 in the short-term stationary data (Table S5). Limiting the number of steps in a boosting model or restrict the number of trees or nodes in a random forest could alleviate this issue. Farrell et al.[25] and Patton et al.[9] restricted the number variables in their models (GLM and GAM respectively) by regression screening, which in our study increased R^2 values for most algorithms when external measurements were predicted. For GLM, the external prediction R^2 increased from 0.57-0.58 to 0.62 in the mobile data and remained the same in the short-term data set. Predictions from the GAM models were 0.48 and 0.55 in the mobile data and 0.37 in the short-term data. Using regression screening increased performance to 0.60 and 0.53 in the mobile and short-term data set respectively. As there usually is no external long-term data available it is difficult to choose parameters, especially when models are trained on data sets with different timescales.

4.3 Previous model algorithm comparisons

A few studies compared linear regression techniques to typically one or a few other algorithms. Weichenthal et al.[10] and Zou et al.[30] respectively compared KRLS and GAM to linear regression and found minor differences between methods, especially when tested on an external data set. Basagaña et al.[6] compared DSA to linear regression and found that predictive power was more related to the number of measurements sites than the applied algorithm. Brokamp et al.[31] compared linear regression to random forest and found that regression models showed higher prediction error variance with cross validation in 24 fixed sites. A difference with the settings in our study is the much smaller number of training sites, the more stable pollution data and that the authors used a variable selection method based on the variable importance in a random forest approach. This makes it more difficult to compare these results to algorithms that use other variable selection methods[31].

In our study, we compared a much larger number of algorithms. The different findings in our study for the mobile and short-term training data and the somewhat inconsistent findings in previous studies, preclude drawing strong conclusions based on empirical performance of models. This was also found by van den Bossche et al[16], who compared LASSO and Support Vector Regression (SVR) to linear regression in a mobile monitoring campaign of black carbon. To minimize overfitting of their models they performed cross-validation with a fully rebuilt model (including variable selection) on every training set and found only small differences between the techniques. Worse predictive performance was found for algorithms that did not limit the number of predictor variables[16]. Performance of our SVM model was also better in predicting external measurements when only the variables were used that correlated best with the outcome in the training data set (0.45 opposed 0.36 in the mobile data set and 0.60 opposed to 0.50 in the short-term data set). In general, data mining algorithms tended to benefit from variable reduction in the mobile data set (Table S4).

Cross validation and a subset of variables was also used by Reid et al.[15] predicting spatial variation of PM concentrations during wildfire events. They compared eleven algorithms, each with their optimal number of variables. Data mining methods (boosting, random forest and SVM) were better in explaining $PM_{2.5}$ related to wildfire events (based on cross validated R^2 values) than regularization methods and additive models[15], but difference were small and no external data set was used to assess their performance. The relative performance of these algorithms may depend on number of sites, noise in air pollution measurement, spatial contrast, extent of study area, number of and correlation among predictors.

Supporting Information

Supporting information consists of two parts: Part 1 are general tables and figures listing the GIS predictors and showing distributions of UFP measurements and sensitivity analyses. Part 2 elaborates on the different algorithms separately (in text, scatterplots, and parameter settings).

References

(1) Ryan, P. H.; LeMasters, G. K. A Review of Land-Use Regression Models for Characterizing Intraurban Air Pollution Exposure. *Inhal. Toxicol.* **2007**, *19* (sup1), 127–133. https://doi.org/10.1080/08958370701495998.

(2) Hoek, G. Methods for Assessing Long-Term Exposures to Outdoor Air Pollutants. *Curr. Environ. Heal. Reports* **2017**, *4* (4), 450–462. https://doi.org/10.1007/s40572-017-0169-5.

(3) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; 2000. https://doi.org/10.1007/978-1-4614-7138-7.

(4) Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning. *Elements* **2009**, *1*, 337–387. https://doi.org/10.1007/b94608.

(5) Wang, M.; Beelen, R.; Basagana, X.; Becker, T.; Cesaroni, G.; de Hoogh, K.; Dedele, A.; Declercq, C.; Dimakopoulou, K.; Eeftens, M.; Forastiere, F.; Galassi, C.; Gražulevičienė, R.; Hoffmann, B.; Heinrich, J.; Iakovides, M.; Künzli, N.; Korek, M.; Lindley, S.; Mölter, A.; Mosler, G.; Madsen, C.; Nieuwenhuijsen, M.; Phuleria, H.; Pedeli, X.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Stempfelet, M.; Tsai, M.-Y.; Lanki, T.; Udvardy, O.; Varró, M. J.; Wolf, K.; Weinmayr, G.; Yli-Tuomi, T.; Hoek, G.; Brunekreef, B. Evaluation of Land Use Regression Models for NO 2 and Particulate Matter in 20 European Study Areas: The ESCAPE Project. *Environ. Sci. Technol.* **2013**, *47* (9), 4357–4364. https://doi.org/10.1021/es305129t.

(6) Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Vila, J.; Künzli, N. Effect of the Number of Measurement Sites on Land Use Regression Models in Estimating Local Air Pollution. *Atmos. Environ.* **2012**, *54*, 634–642. https://doi.org/10.1016/j.atmosenv.2012.01.064.

(7) Bellinger, C.; Mohomed Jabbar, M. S.; Zaïane, O.; Osornio-Vargas, A. A Systematic Review of Data Mining and Machine Learning for Air Pollution Epidemiology. *BMC Public Health* **2017**, *17* (1), 907. https://doi.org/10.1186/s12889-017-4914-3.

(8) Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; Thiele, L. Deriving High-Resolution Urban Air Pollution Maps Using Mobile Sensor Nodes. In *Pervasive and Mobile Computing*; 2015; Vol. 16, pp 268–285. https://doi.org/10.1016/j.pmcj.2014.11.008.

(9) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An Hourly Regression Model for Ultrafine Particles in a Near-Highway Urban Area. *Environ. Sci. Technol.* **2014**, *48* (6), 3272–3280. https://doi.org/10.1021/es404838k.

(10) Weichenthal, S.; Ryswyk, K. Van; Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A Land Use Regression Model for Ambient Ultrafine Particles in Montreal, Canada: A Comparison of Linear Regression and a Machine Learning Approach. *Environ. Res.* **2016**, *146*, 65–72. https://doi.org/10.1016/j.envres.2015.12.016.

(11) Liu, W.; Li, X.; Chen, Z.; Zeng, G.; León, T.; Liang, J.; Huang, G.; Gao, Z.; Jiao, S.; He, X.; Lai, M. Land Use Regression Models Coupled with Meteorology to Model Spatial and Temporal Variability of NO_2 and PM_{10} in Changsha, China. *Atmos. Environ.* **2015**, *116*, 272–280. https://doi.org/10.1016/j.atmosenv.2015.06.056.

(12) Song, B.; Wu, J.; Zhou, Y.; Hu, K. Fine-Scale Prediction of Roadside CO and NOx Concentration Based on a Random Forest Model. *J. Residuals Sci. Technol.* **2014**, *11*, 83–89.

(13) Lima, A. R.; Cannon, A. J.; Hsieh, W. W. Nonlinear Regression in Environmental Sciences Using Extreme Learning Machines: A Comparative Evaluation. *Environ. Model. Softw.* **2015**, *73*, 175–188. https://doi.org/10.1016/j.envsoft.2015.08.002.

(14) Lary, D. J.; Alavi, A. H.; Gandomi, A. H.; Walker, A. L. Machine Learning in Geosciences and Remote Sensing. *Geosci. Front.* **2016**, *7* (1), 3–10. https://doi.org/10.1016/j.gsf.2015.07.003.

(15) Reid, C. E.; Jerrett, M.; Petersen, M. L.; Pfister, G. G.; Morefield, P. E.; Tager, I. B.; Raffuse, S. M.; Balmes, J. R. Spatiotemporal Prediction of Fine Particulate Matter during the 2008 Northern California Wildfires Using Machine Learning. *Environ. Sci. Technol.* **2015**, *49* (6), 3887–3896. https://doi.org/10.1021/es505846r.

(16) Van den Bossche, J.; De Baets, B.; Verwaeren, J.; Botteldooren, D.; Theunis, J. Development and Evaluation of Land Use Regression Models for Black Carbon Based on Bicycle and Pedestrian Measurements in the Urban Environment. *Environ. Model. Softw.* **2018**, *99*, 58–69. https://doi.org/10.1016/J.ENVSOFT.2017.09.019.

(17) Kerckhoffs, J.; Hoek, G.; Vlaanderen, J.; van Nunen, E.; Messier, K.; Brunekreef, B.; Gulliver, J.; Vermeulen, R. Robustness of Intra Urban Land-Use Regression Models for Ultrafine Particles and Black Carbon Based on Mobile Monitoring. *Environ. Res.* **2017**, *159*, 500–508. https://doi.org/10.1016/j.envres.2017.08.040.

(18) Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of Ultrafine Particles and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environ. Sci. Technol.* **2016**, *50* (23), 12894–12902. https://doi.org/10.1021/acs.est.6b03476.

(19) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712–8720. https://doi.org/10.1021/es505791g.

(20) Nunen, E. van; Vermeulen, R.; Tsai, M.-Y.; Probst-Hensch, N.; Ineichen, A.; Davey, M. E.; Imboden, M.; Ducret-Stich, R.; Naccarati, A.; Raffaele, D.; Ranzi, A.; Ivaldi, C.; Galassi, C.; Nieuwenhuijsen, M. J.; Curto, A.; Donaire-Gonzalez, D.; Cirach, M.; Chatzi, L.; Kampouri, M.; Vlaanderen, J.; Meliefste, K.; Buijtenhuijs, D.; Brunekreef, B.; Morley, D.; Vineis, P.; Gulliver, J.; Hoek, G. Land Use Regression Models for Ultrafine Particles in Six European Areas. *Environ. Sci. Technol.* **2017**, *51* (6), 3336–3345. https://doi.org/10.1021/acs.est.6b05920.

(21) van der Laan, M. J.; Polley, E. C.; Hubbard, A. E. Super Learner. *Stat. Appl. Genet. Mol. Biol.* **2007**, *6* (1), Article25. https://doi.org/10.2202/1544-6115.1309.

(22) Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A Review of Land-Use Regression Models to Assess Spatial Variation of Outdoor Air Pollution. *Atmos. Environ.* **2008**, *42* (33), 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057.

(23) Eeftens, M.; Tsai, M. Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; De Nazelle, A.; de Vocht, F.; Declercq, C.; Dedele, A.; Eriksen, K.; Galassi, C.; Gražulevičiene, R.; Grivas, G.; Heinrich, J.; Hoffmann, B.; Iakovides, M.; Ineichen, A.; Katsouyanni, K.; Korek, M.; Krämer, U.; Kuhlbusch, T.; Lanki, T.; Madsen, C.; Meliefste, K.; Mölter, A.; Mosler, G.; Nieuwenhuijsen, M.; Oldenwening, M.; Pennanen, A.; Probst-Hensch, N.; Quass, U.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Udvardy, O.; Vaskövi, É.; Weinmayr, G.; Brunekreef, B.; Hoek, G. Spatial Variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ Absorbance and PMcoarse Concentrations between and within 20 European Study Areas and the Relationship with NO2 - Results of the ESCAPE Project. *Atmos. Environ.* **2012**, *62*, 303–317. https://doi.org/10.1016/j.atmosenv.2012.08.038.

(24) Beckerman, B. S.; Jerrett, M.; Martin, R. V.; Van Donkelaar, A.; Ross, Z.; Burnett, R. T. Application of the Deletion/Substitution/Addition Algorithm to Selecting Land Use Regression Models for Interpolating Air Pollution Measurements in California. *Atmos. Environ.* **2013**, *77*, 172–177. https://doi.org/10.1016/j.atmosenv.2013.04.024 Short communication.

(25) Farrell, W.; Weichenthal, S.; Goldberg, M.; Valois, M. F.; Shekarrizfard, M.; Hatzopoulou, M. Near Roadway Air Pollution across a Spatially Extensive Road and Cycling Network. *Environ. Pollut.* **2016**, *212*, 498–507. https://doi.org/10.1016/j.envpol.2016.02.041.

(26) Zwack, L. M.; Paciorek, C. J.; Spengler, J. D.; Levy, J. I. Modeling Spatial Patterns of Traffic-Related Air Pollutants in Complex Urban Terrain. *Environ. Health Perspect.* **2011**, *119* (6), 852–859. https://doi.org/10.1289/ehp.1002519.

(27) Lu, W.-Z.; Wang, W.-J. Potential Assessment of the "Support Vector Machine" Method in Forecasting Ambient Air Pollutant Trends. *Chemosphere* **2005**, *59* (5), 693–701. https://doi.org/10.1016/j.chemosphere.2004.10.032.

(28) Luna, A. S.; Paredes, M. L. L.; de Oliveira, G. C. G.; Corrêa, S. M. Prediction of Ozone Concentration in Tropospheric Levels Using Artificial Neural Networks and Support Vector Machine at Rio de Janeiro, Brazil. *Atmos. Environ.* **2014**, *98*, 98–104. https://doi.org/10.1016/j.atmosenv.2014.08.060.

(29) Simon, M. C.; Hudda, N.; Naumova, E. N.; Levy, J. I.; Brugge, D.; Durant, J. L. Comparisons of Traffic-Related Ultrafine Particle Number Concentrations Measured in Two Urban Areas by Central, Residential, and Mobile Monitoring. *Atmos. Environ.* **2017**, *169*, 113–127. https://doi.org/10.1016/j.atmosenv.2017.09.003.

(30) Zou, B.; Chen, J.; Zhai, L.; Fang, X.; Zheng, Z. Satellite Based Mapping of Ground $PM_{2.5}$ Concentration Using Generalized Additive Modeling. *Remote Sens.* **2016**, *9* (1), 1. https://doi.org/10.3390/rs9010001.

(31) Brokamp, C.; Jandarov, R.; Rao, M. B.; LeMasters, G.; Ryan, P. Exposure Assessment Models for Elemental Components of Particulate Matter in an Urban Environment: A Comparison of Regression and Random Forest Approaches. *Atmos. Environ.* **2017**, *151*, 1–11. https://doi.org/10.1016/j.atmosenv.2016.11.066.

Supporting Information for

Performance of Prediction Algorithms for Modeling Outdoor Air Pollution Spatial Surfaces

Part 1: General

Table S1
Spatial predictor variables with units, a priori defined directions of effect and buffer sizes in the mobile and short-term stationary data sets.

Table S2
Distribution of site averaged concentrations in the mobile and short-term training datasets and the external long-term validation dataset (in particles/cm³).

Table S3
Overview selected variables in mobile monitoring data models that perform variable selection.

Figure S1
Distribution of Ranks in the Mobile (top) and Short-Term Stationary (bottom) data.

Figure S2
Distribution of R-squares when predictions of Mobile (top) and Short-Term Stationary (bottom) models are compared to external measurements without highest observation.

Table S4/S5
Comparison of prediction algorithms with training model R^2 and weights in Super Learner with Mobile/Short-Term UFP data.

Table S6/S7
Combining all algorithms and screening procedures in SuperLearner with Mobile/Short-Term UFP data.

2.1 Stepwise Regression (Customized, Forward and Backward)

2.2 Deletion Substitution Addition (DSA) Algorithm

2.3 Regularization algorithms (LASSO, Elastic Net and Ridge)

2.4 Generalized Linear Models (GLM)

2.5 Additive Models (GAM and MARS)

2.6 Kernel Regularized Least Squares (KRLS)

2.7 Neural Networks

2.8 Support Vector Machines (SVM)

2.9 RandomForest

2.10 Boosting and Bagging

2.11 SuperLearner

Table S1

Spatial predictor variables with units, a priori defined directions of effect and buffer sizes in the mobile and short-term stationary data sets.

↓

Predictor variable	Units	Direction of effect [b]	Mobile				Short-term Stationary			
			Buffer	10th percentile	Mean	90th Percentile	Buffer	10th percentile	Mean	90th Percentile
INDUS_ Industry area	**m²**	**+**	**100**	0	1400	0	**100**	0	440	0
			300	0	13633	37122	**300**	0	7237	6463
			500	0	41895	166610	**500**	0	29522	110419
			1000	0	197751	711267	**1000**	0	174742	619476
			5000	2305345	5591752	8349138	**5000**	1931709	5353372	8062690
PORT_ Port area	**m²**	**+**	**100**	0	247	0	**100**	0	208	0
			300	0	2473	0	**300**	0	1786	0
			500	0	8142	0	**500**	0	7167	0
			1000	0	45285	0	**1000**	0	52363	0
			5000*	0	2428278	9225706	**5000***	0	2116359	8617504
URBG_ Urban green area	**m²**	**-**	**100**	0	490	0	**100**	0	963	0
			300	0	7711	22125	**300**	0	10843	37904
			500	0	32633	132893	**500**	0	37932	152378
			1000	0	190110	579390	**1000**	0	192426	551366
			5000*	1372895	5233285	9714816	**5000***	1122221	4590130	9281079
NATUR_ Natural and forested areas	**m²**	**-**	**100**	0	222	0	**100**	0	167	0
			300	0	2103	0	**300**	0	1588	0
			500	0	7244	0	**500**	0	6077	0
			1000	0	52377	165299	**1000***	0	51926	144940
			5000	1334268	4944915	8182331	**5000***	1328510	5172157	8159990
RES_ Residential land area	**m²**	**+**	**100**	0	26434	31375	**100**	0	25954	31375
			300	66208	226402	282618	**300**	47422	222740	282618
			500	231301	593328	785191	**500***	189582	579130	785191
			1000*	842139	2082317	3050349	**1000***	518003	1967973	3005396
			5000**	15002050	28689475	46595685	**5000****	11680371	27225930	46124341

Predictor variable	Units	Direction of effect [b]	Mobile				Short-term Stationary			
			Buffer	10th percentile	Mean	90th Percentile	Buffer	10th percentile	Mean	90th Percentile
POP_ Population density	**n**	**+**	**100**	7	245	524	**100**	15	270	561
			300	311	2117	4370	**300**	480	2096	4429
			500*	1236	5469	11421	**500****	956	5208	10991
			1000*	4805	19270	42504	**1000****	2533	17452	39812
			5000**	85535	251242	539468	**5000****	71083	227169	531610
HHOLD_ Household density	**n**	**+**	**100**	3	132	292	**100**	na	na	na
			300	144	1136	2476	**300**	na	na	na
			500*	563	2940	6507	**500**	na	na	na
			1000**	2079	10416	24315	**1000**	na	na	na
			5000**	39485	134927	307588	**5000**	na	na	na
TRAFNEAR Traffic intensity on nearest road	**Veh. day^{-1}**	**+**	****	82	8656	25785	****	30	4090	14943
TRAFMAJOR Traffic intensity on nearest major road	**Veh. day^{-1}**	**+**		5736	18232	36470	*	5649	18579	34240
HTRAFNEAR Heavy-duty traffic intensity on nearest road	**Veh. day^{-1}**	**+**	****	0	324	1005	****	0	125	420
HTRAFMAJOR Heavy-duty traffic intensity on nearest major road	**Veh. day^{-1}**	**+**		67	982	1950		48	1206	1769
RDL_ Road length of all roads	**m**	**+**	**50***	102	258	404	**50***	98	190	308
			100*	488	838	1200	**100***	363	716	1059
			300*	3997	6359	8494	**300***	3180	5925	8209
			500*	9994	16603	21857	**500***	7734	15605	21532
			1000*	33179	60248	80412	**1000***	27773	55926	78365
MRDL_ Road length of all major roads	**m**	**+**	**50****	0	79	203	**50****	0	42	174
			100**	0	194	507	**100****	0	107	390
			300**	0	1030	2259	**300***	0	664	1760
			500**	0	2600	4951	**500***	0	2080	4164
			1000*	3161	9707	15328	**1000***	2165	8471	14235
TLOA_ Traffic intensity on all roads (sum of (traffic intensity * length of all segments))	**Veh. day^{-1}m**	**+**	**50****	40082	1257727	3400359	**50****	7963	757087	2207469
			100**	185957	3417972	8691268	**100****	77072	2255473	5463763
			300**	3136239	22142804	47797106	**300****	1315772	15681335	36067592
			500**	11762571	58068646	120273812	**500****	5579975	46420718	98239045
			1000**	63847576	226759175	452445437	**1000****	31635299	196221848	417062321
TMLOA_ Traffic intensity on all major roads (sum of (traffic intensity* length of all segments))	**Veh. day^{-1}m**	**+**	**50****	0	1127175	3336962	**50****	0	688087	2185933
			100**	0	2998942	8304429	**100****	0	1942440	4963848
			300**	0	18940815	44304127	**300****	0	12729584	31697479
			500**	4808393	49735352	108364418	**500****	0	38751720	89534363
			1000**	44458796	196518241	410822257	**1000****	18085365	168852113	377180297
HMLOA_ Heavy-duty traffic intensity on all roads (sum of (heavy-duty traffic intensity* length of all segments))	**Veh. day^{-1}m**	**+**	**50***	440	54021	141389	**50****	0	36679	92230
			100*	3122	155742	378676	**100***	568	131698	213495
			300*	72173	1121670	2637212	**300***	41345	760615	1541019
			500*	328397	3078547	7798998	**500***	157767	2233684	5463944
			1000	2075563	13058286	26737180	**1000***	1077166	10937816	23631745
HMLOA_ Heavy-duty traffic intensity on major roads (sum of (heavy-duty traffic intensity*length of all segments)	**Veh. day^{-1}m**	**+**	**50***	0	48165	134636	**50****	0	33903	92230
			100*	0	137648	357616	**100***	0	120977	210302
			300*	0	988569	2470260	**300***	0	651430	1335757
			500*	106797	2727455	7255375	**500***	0	1931699	5094488
			1000	1133803	11678977	25328256	**1000***	624627	9707896	21738224
DINVNEAR Inverse Distance to nearest road [a]	**m^{-1}**	**+**		na	na	na		0.073	0.488	0.756
DINVMAJOR Inverse Distance to nearest major road [a]	**m^{-1}**	**+**		na	na	na	*	0.002	0.065	0.135
Campaign	**N**			0	0.64	1		0	0.60	1

[a] Variables were not used for mobile model development, due to distance values being zero.
[b] Used in the supervised forward regression only
* Absolute Pearson Correlation with UFP > 0.1
** Among best 20 correlated predictors

Table S2

Distribution of site averaged concentrations in the mobile and short-term training datasets and the external long-term validation dataset (in particles/cm³).

↓

Campaign	Observations	Mean	Std Dev	P10	P90
Total Mobile	**8200**	**19542**	**17842**	**6466**	**37947**
MUSiC	2964	21167	21424	5647	42313
EXPOsOMICS	5236	18672	15359	6897	35555
Total Stationary	**368**	**12812**	**6809**	**6343**	**20473**
MUSiC	128	12630	6735	6269	20050
EXPOsOMICS	240	12910	6860	6453	20663
External	**42**	**14932**	**4888**	**10561**	**21419**

Table S3

Overview selected variables in mobile monitoring data models that perform variable selection.

↓

	Supervised	Stepwise (AIC)	Stepwise (R^2)	Forward	Backward	LASSO	Elastic Net α=0.75	Elastic Net α=0.50	Elastic Net α=0.25
Population Density variables		LDRES_5000			LDRES_5000	LDRES_5000	LDRES_5000	LDRES_5000	LDRES_5000
					POP_500				
	POP_5000		POP_5000	POP_5000	POP_5000	POP_5000	POP_5000	POP_5000	POP_5000
					HHOLD_1000				
					HHOLD_5000			HHOLD_5000	HHOLD_5000
Road length variables	MRDL_50	MRDL_50	MRDL_50	MRDL_50	MRDL_50	MRDL_50	MRDL_50	MRDL_50	MRDL_50
						MRDL_100	MRDL_100	MRDL_100	MRDL_100
					RDL_100				
					RDL_500				
		RDL_1000	RDL_1000	RDL_1000		RDL_1000			
									TMLOA_50
					TMLOA_100		TMLOA_100	TMLOA_100	TMLOA_100
					TMLOA_300				
					TMLOA_1000				
									TLOA_50
Traffic Intensity Variables	TLOA_100	TLOA_100	TLOA_100	TLOA_100		TLOA_100	TLOA_100	TLOA_100	TLOA_100
					TLOA_500				
	TRAFNEAR	TRAFNEAR	TRAFNEAR	TRAFNEAR	TRAFNEAR	TRAFNEAR	TRAFNEAR	TRAFNEAR	TRAFNEAR
					HLOA_1000				
					HMLOA_50				
					HMLOA_300				
					HMLOA_1000				
Other	PORT_5000	PORT_5000	PORT_5000	PORT_5000	PORT_5000				
	Campaign	Campaign	Campaign	Campaign					
					NATUR_500				
		NATUR_1000	NATUR_1000	NATUR_1000	NATUR_1000				
Training Model R^2	R^2=0.13	R^2=0.14	R^2=0.13	R^2=0.13	R^2=0.15	R^2=0.14	R^2=0.14	R^2=0.14	R^2=0.14
Number of variables	n=6	n=8	n=8	n=8	n=20	n=7	n=7	n=8	n=10

Figure S1

Distribution of Ranks when predictions of Mobile (top) and Short-Term Stationary (bottom) models are compared to external measurements, based on 1000 bootstrapped samples. Algorithms are ranked for every bootstrap sample and the average and distribution is shown.

↓

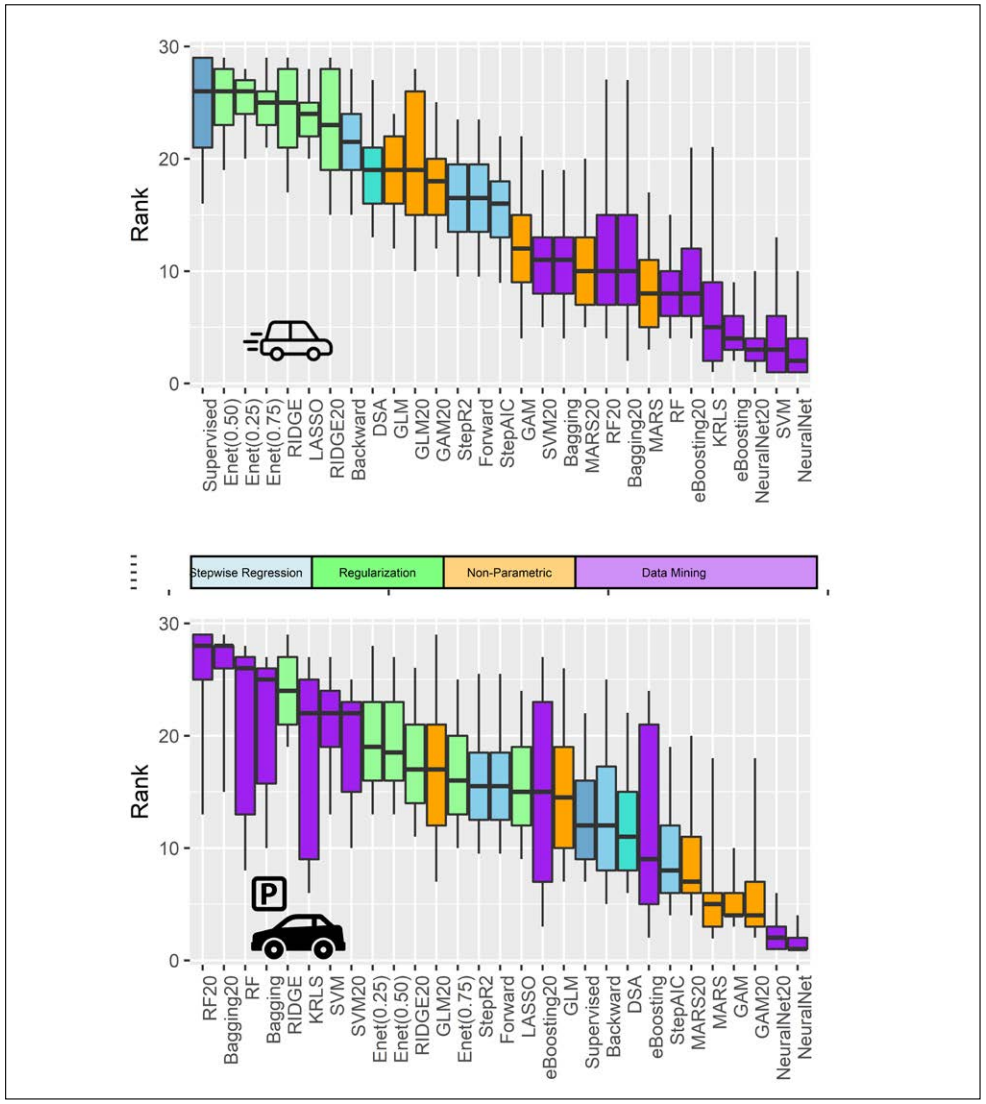


Figure S2

Distribution of R-squares when predictions of Mobile (top) and Short-Term Stationary (bottom) models are compared to external measurements when the highest observation in the external dataset is excluded (n=41).

↓

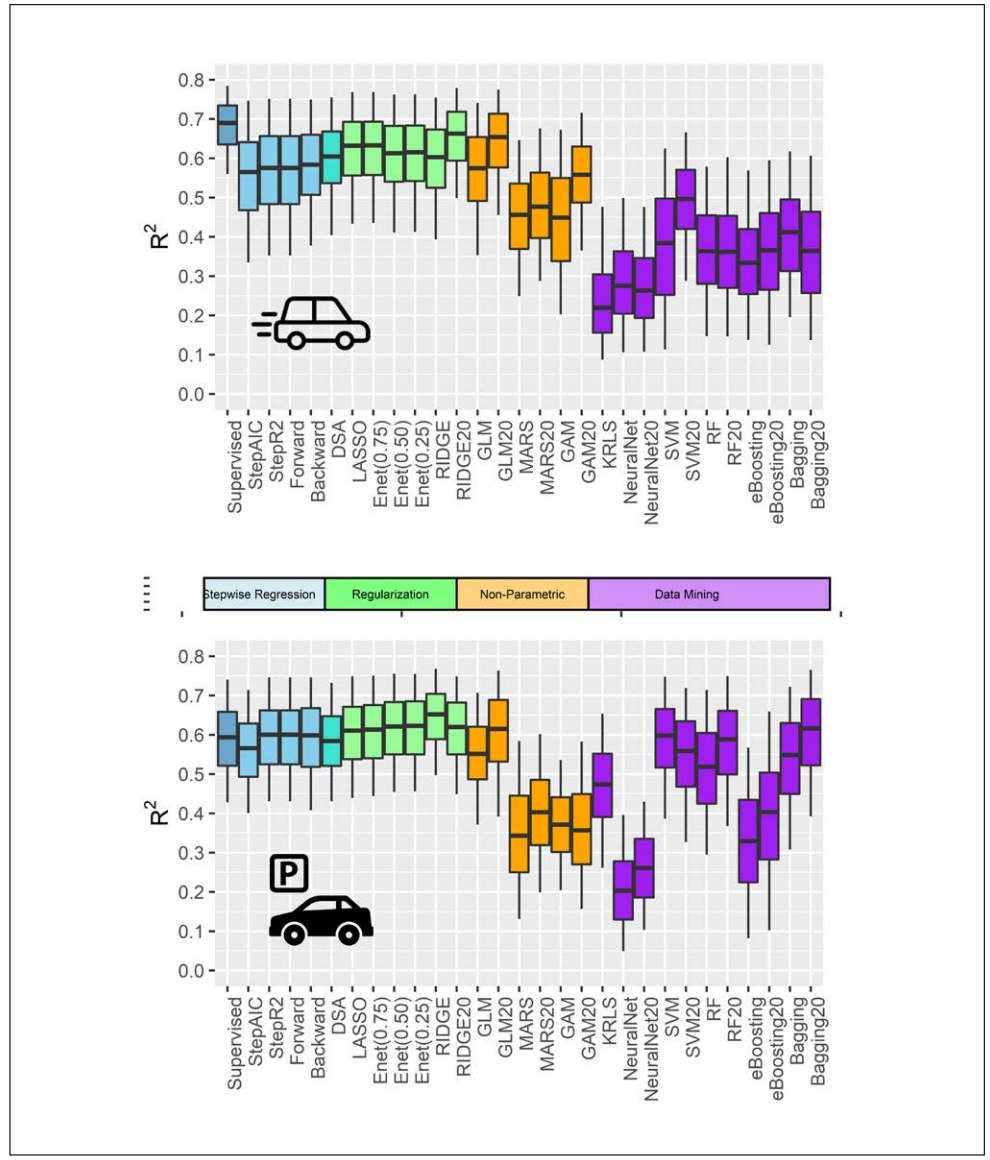


Table S4

Comparison of prediction algorithms in Super Learner with Mobile UFP data.

↓

			Training			External	
	Algorithm	**Variables**	**R^2**	**Random CV- R^2**	**Clustered CV- R^2**	**R^2**	**Weight [a]**
Linear regression	**Stepwise Both (R^2)**	All	0.14	0.14	0.1	0.54	0
	Stepwise Both (AIC)	All	0.14	0.13	0.09	0.53	0
	Stepwise Backward (R^2)	All	0.14	0.14	0.08	0.54	0
	Stepwise Forward (R^2)	All	0.15	0.15	0.1	0.60	0
Regularization	**LASSO**	All	0.15	0.13	0.08	0.62	0
	Elastic net (alpha=0.25)	All	0.15	0.13	0.08	0.63	0
	Elastic net (alpha=0.50)	All	0.15	0.13	0.08	0.63	0
	Elastic net (alpha=0.75)	All	0.15	0.14	0.08	0.62	0
	Ridge	All	0.15	0.14	0.08	0.63	0
		Top20	0.13	0.12	0.1	0.61	0
Non Linear	**Generalized Linear Model (GLM)**	All	0.09	0.12	0.07	0.58	0
		Top20	0.08	0.11	0.08	0.57	0
	Multivariate Adaptive Regression Splines (MARS)	All	0.10	0.12	0.08	0.41	0
		Top20	0.12	0.11	0.06	0.43	0
	General Additive Model (GAM)	All	0.20	0.17	0.08	0.48	0
		Top20	0.16	0.14	0.09	0.55	0
Data mining	**Neural network**	All	0.07	0.06	0.04	0.26	0
		Top20	0.06	0.05	0.04	0.28	0
	Support Vector Machine	All	0.34	0.23	0.08	0.26	0.40
		Top20	0.23	0.17	0.09	0.45	0
	RandomForest	All	0.25	0.24	0.08	0.40	0.19
		Top20	0.23	0.22	0.07	0.45	0
	Gradient Boosting Machine (GBM)	All	0.16	0.15	0.1	0.49	0
		Top20	0.16	0.14	0.1	0.50	0
	Extreme Boosting	All	0.86	0.23	0.06	0.31	0.41
		Top20	0.82	0.18	0.05	0.42	0
	Bagging	All	0.26	0.16	0.09	0.46	0
		Top20	0.23	0.15	0.08	0.46	0
	Super Learner	All	/	/	/	0.37	1

[a] Weights are based on the difference between the training R^2 and the random cross validated linear regression R^2.

Table S5

Comparison of prediction algorithms in Super Learner with Short-term UFP data.

↓

			Training		External	
	Algorithm	**Variables**	**R^2**	**CV- R^2**	**R^2**	**Weight [a]**
Linear regression	**Stepwise Both (R^2)**	All	0.36	0.36	0.54	0
	Stepwise Both (AIC)	All	0.36	0.29	0.49	0
	Stepwise Backward (R^2)	All	0.36	0.36	0.54	0.63
	Stepwise Forward (R^2)	All	0.36	0.36	0.52	0
Regularization	**LASSO**	All	0.36	0.30	0.55	0
	Elastic net (alpha=0.25)	All	0.36	0.30	0.56	0
	Elastic net (alpha=0.50)	All	0.36	0.30	0.56	0
	Elastic net (alpha=0.75)	All	0.36	0.30	0.55	0
	Ridge	All	0.42	0.28	0.61	0
		Top20	0.36	0.31	0.55	0
Non Linear	**Generalized Linear Model (GLM)**	All	0.45	0.23	0.53	0
		Top20	0.36	0.28	0.55	0
	Multivariate Adaptive Regression Splines (MARS)	All	0.55	0.04	0.38	0
		Top20	0.55	0.33	0.45	0.19
	General Additive Model (GAM)	All	0.63	0.03	0.37	0
		Top20	0.52	0.02	0.37	0
Data mining	**Neural network**	All	0.21	0.06	0.22	0
		Top20	0.20	0.10	0.27	0
	Support Vector Machine	All	0.60	0.32	0.61	0
		Top20	0.55	0.33	0.60	0
	RandomForest	All	0.32	0.30	0.66	0
		Top20	0.35	0.33	0.70	0
	Gradient Boosting Machine (GBM)	All	0.48	0.31	0.66	0
		Top20	0.47	0.32	0.64	0
	Extreme Boosting	All	1.00	0.26	0.51	0.18
		Top20	1.00	0.29	0.56	0
	Bagging	All	0.66	0.31	0.65	0
		Top20	0.63	0.34	0.70	0
	Super Learner	All	/	/	0.60	1

[a] Weights are based on the difference between the training R^2 and the CV-R^2.

Table S6

Combining all algorithms and screening procedures with Mobile UFP data.

↓

	Model	Variables	CV - R^2	External measurements R^2	External measurements Bias (in particles/cm³)	Weight
Linear Regression	**Stepwise Both (R^2)**	All	0.13	0.56	6768	0
	Stepwise Both (AIC)	All	0.13	0.56	6828	0
	Stepwise Backward (R^2)	All	0.15	0.59	7326	0
	Stepwise Forward (R^2)	All	0.13	0.56	6768	0
Regularization	**LASSO**	All	0.14	0.56	7981	0
	Elastic net (alpha=0.25)	All	0.14	0.56	8280	0
	Elastic net (alpha=0.50)	All	0.13	0.59	6598	0
	Elastic net (alpha=0.75)	All	0.13	0.56	7953	0
	RIDGE	All	0.14	0.62	6575	0
		Cor 0.1	0.13	0.67	3371	0
		Top20	0.12	0.61	4366	0
		SupReg	0.13	0.62	6882	0
Non-Linear	**Generalized Linear Model (GLM) (link=log)**	All	0.12	0.58	8906	0
		Cor 0.1	0.12	0.60	-358	0
		Top20	0.11	0.57	348	0
		SupReg	0.11	0.66	5086	0
	Multivariate Adaptive Regression Splines (MARS)	All	0.11	0.33	2360	0
		Cor 0.1	0.11	0.41	5357	0
		Top20	0.11	0.43	5965	0
		SupReg	0.14	0.25	4596	0
	General Additive Model (GAM)	All	0.17	0.49	9244	0
		Cor 0.1	0.16	0.60	8092	0
		Top20	0.14	0.55	7320	0
		SupReg	0.13	0.60	6459	0
Data Mining	**Neural network (size=10)**	All	0.05	0.06	7176	0
		Cor 0.1	0.02	0.14	4347	0
		Top20	0.02	0.06	4490	0
		SupReg	0.04	0.11	5097	0
	Support Vector Machine	All	0.26	0.36	13056	0.27
		Cor 0.1	0.21	0.32	284	0
		Top20	0.17	0.45	1915	0
		SupReg	0.18	0.39	6659	0
	RandomForest	All	0.28	0.48	12661	0.03
		Cor 0.1	0.24	0.41	5430	0
		Top20	0.22	0.44	5598	0
		SupReg	0.28	0.53	5160	0.31
	Gradient Boosting Machine (GBM)	All	0.15	0.50	4312	0
		Cor 0.1	0.14	0.49	4451	0
		Top20	0.14	0.49	4680	0
		SupReg	0.12	0.45	4211	0
	Extreme Boosting	All	0.29	0.35	18857	0.39
		Cor 0.1	0.22	0.27	6182	0
		Top20	0.18	0.43	5982	0
		SupReg	0.20	0.61	4937	0
	Bagging	All	0.17	0.48	4586	0
		Cor 0.1	0.16	0.47	4255	0
		Top20	0.15	0.44	4864	0
		SupReg	0.15	0.52	4076	0
	SuperLearner	All		0.57	11621	1

Table S7

Combining all algorithms and screening procedures with Short-Term UFP data.

↓

	Model	Variables	CV - R^2	External measurements R^2	External measurements Bias (in particles/cm³)	Weight
Linear Regression	**Stepwise Both (R^2)**	All	0.35	**0.53**	152	0
	Stepwise Both (AIC)	All	0.31	**0.47**	6954	0
	Stepwise Backward (R^2)	All	0.40	**0.57**	1042	0.37
	Stepwise Forward (R^2)	All	0.35	**0.53**	152	0
Regularization	**LASSO**	All	0.31	**0.55**	4424	0
	Elastic net (alpha=0.25)	All	0.31	**0.56**	4588	0
	Elastic net (alpha=0.50)	All	0.31	**0.55**	5467	0
	Elastic net (alpha=0.75)	All	0.3	**0.54**	5517	0
	RIDGE	All	0.30	**0.59**	4058	0
		Cor 0.1	0.28	**0.59**	84	0
		Top20	0.29	**0.55**	-30	0
		SupReg	0.32	**0.53**	97	0
Non-Linear	**Generalized Linear Model (GLM) (link=log)**	All	0.23	**0.52**	-4638	0
		Cor 0.1	0.21	**0.47**	-254	0
		Top20	0.27	**0.55**	-892	0
		SupReg	0.32	**0.49**	-721	0
	Multivariate Adaptive Regression Splines (MARS)	All	0.22	**0.39**	575	0
		Cor 0.1	0.31	**0.45**	165	0
		Top20	0.31	**0.45**	100	0
		SupReg	0.37	**0.54**	-411	0.27
	General Additive Model (GAM)	All	0.18	**0.32**	5758	0
		Cor 0.1	0.09	**0.35**	826	0
		Top20	0.05	**0.37**	973	0
		SupReg	0.34	**0.60**	137	0
Data Mining	**Neural network (size=10)**	All	0.05	**0.02**	-1869	0
		Cor 0.1	0.05	**0.03**	-1725	0
		Top20	0.04	**0.16**	-1755	0
		SupReg	0.16	**0.14**	-1999	0
	Support Vector Machine	All	0.32	**0.50**	239	0
		Cor 0.1	0.29	**0.60**	-969	0
		Top20	0.32	**0.60**	-730	0
		SupReg	0.4	**0.62**	-540	0.20
	RandomForest	All	0.33	**0.59**	2356	0
		Cor 0.1	0.30	**0.68**	-170	0
		Top20	0.33	**0.70**	-47	0
		SupReg	0.35	**0.64**	-157	0
	Gradient Boosting Machine (GBM)	All	0.31	**0.62**	-416	0
		Cor 0.1	0.30	**0.66**	-1250	0
		Top20	0.32	**0.64**	-1153	0
		SupReg	0.33	**0.59**	-1024	0
	Extreme Boosting	All	0.26	**0.58**	1587	0.16
		Cor 0.1	0.21	**0.63**	0	0
		Top20	0.22	**0.65**	-502	0
		SupReg	0.25	**0.49**	580	0
	Bagging	All	0.30	**0.65**	736	0
		Cor 0.1	0.30	**0.69**	-519	0
		Top20	0.33	**0.66**	-384	0
		SupReg	0.34	**0.63**	-274	0
	SuperLearner	All		**0.66**	450	1

2.1 Stepwise Regression

Stepwise regression is by far the most widely used prediction algorithm for developing land use regression models for air pollution[1]. It starts with an empty (intercept only) model and then adds or replaces variables based on goodness of fit, which can either be determined by AIC criterion or the (adjusted) R^2 value. Two variations of this process are the forward and backward stepwise regression. Forward regression is similar to stepwise regression but only adds variables to the model, whereas backward stepwise starts with all variables in the model and removes predictors with the least impact on the fit[2]. Applied linear regression methods differ with respect to constraining predictors (only allow in model if direction of slopes conforms with physical reality versus no restriction) and dealing with correlated predictors (pre-selection from groups of predictors or offering all and judging model diagnostics).

In the current study we used a supervised forward method constraining to plausible associations as applied in the ESCAPE study and fully unconstrained forward, backward and stepwise methods. The supervised stepwise method is described in Kerckhoffs et al.[3]. Variables are selected that lead to the largest increase in adjusted R^2; the coefficient conforms with a predefined direction of effect and the direction of effect of predictors already in the model does not change. Predictor variables in the final model are removed from the model when they have a p-value less than 0.1, or a variance inflation factor over 3. The customized stepwise LUR models are shown below and are used in the comparisons between prediction algorithms and for the screening of variables in the sensitivity analysis (Table S4/S5).

Table S8

Supervised stepwise LUR models for UFP (in particles/cm³)

↓

	Short-term	Mobile
Variable	**Estimate (Std error)**	**Estimate (Std error)**
Intercept	7063 (581)	8491 (977)
Population Density in 1000m buffer	2034 (948)*	
Population Density in 5000m buffer	3232 (1148)	7932 (1180)
Major Road Length in 50m buffer		3701 (511)
Major Road Length in 100m buffer	3991 (800)	
Traffic Intensity in a 50m buffer	2736 (558)	
Traffic Intensity in a 100m buffer		2624 (521)
Traffic Intensity on the nearest road		2071 (517)
Port area in 5000m buffer		3047 (1150)
EXPOsOMICS Campaign		2077 (767)
Model R^2	0.35	0.13
Number of Sites/Segments used for model development	368	8200

*All estimates and standard error are multiplied by the difference between 10th and 90th percentile for all predictors

R-Script and parameters:

Stepwise R^2: SignifReg(Y~., X, alpha=0.10, direction = “both”, criterion=”r-adj”, correction=”FDR”)

Stepwise AIC: MASS::stepAIC(g0, scope, direction = “both”, steps = 30)

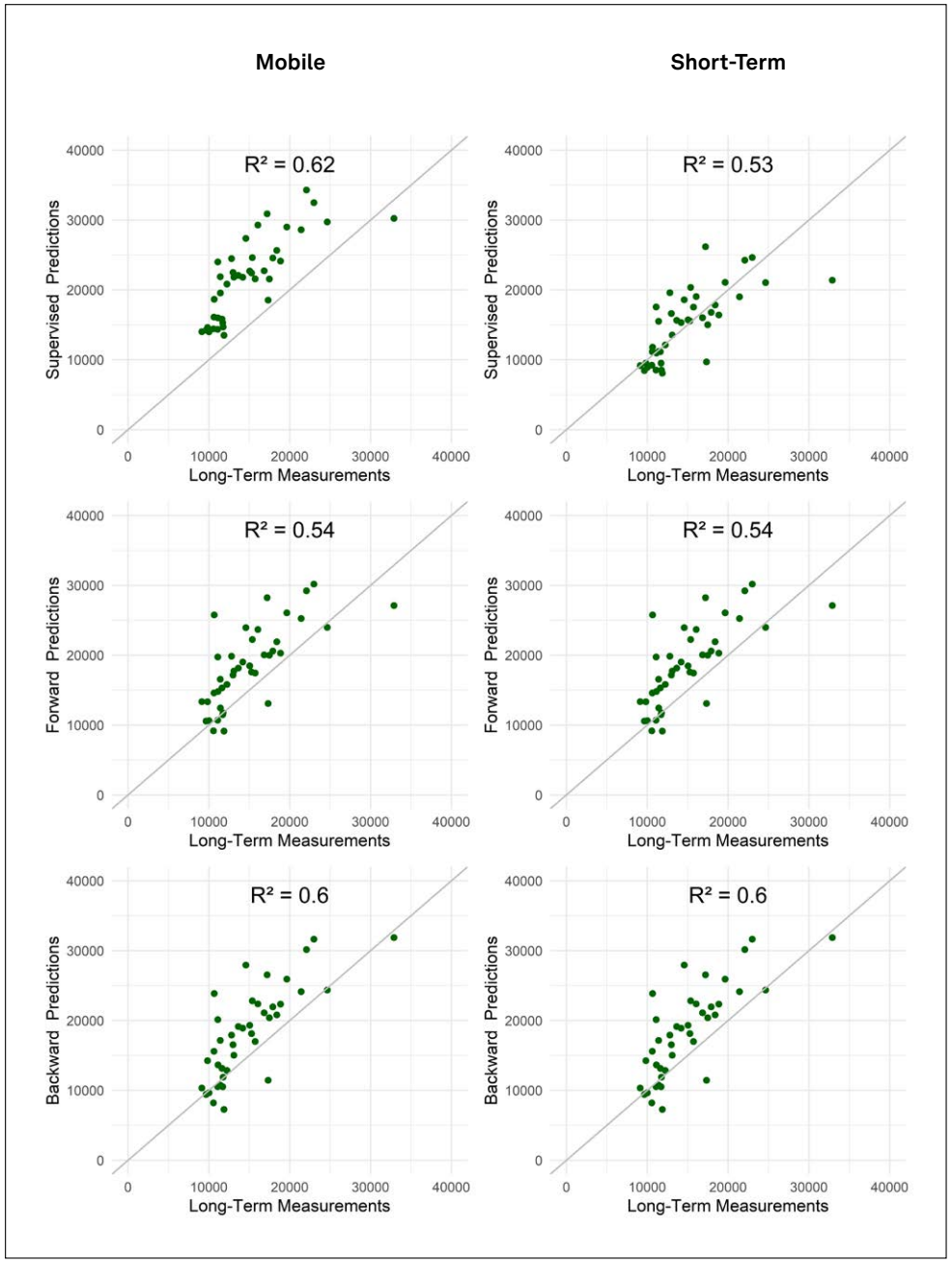
Stepwise Backward: SignifReg(Y~., X, alpha=0.10, direction = “backward”, criterion=”r-adj”, correction=”FDR”)

Stepwise Forward: SignifReg(Y~., X, alpha=0.10, direction = “forward”, criterion=”r-adj”, correction=”FDR”)

Figure S3

Predicted UFP versus Long-Term Measurements.

↓



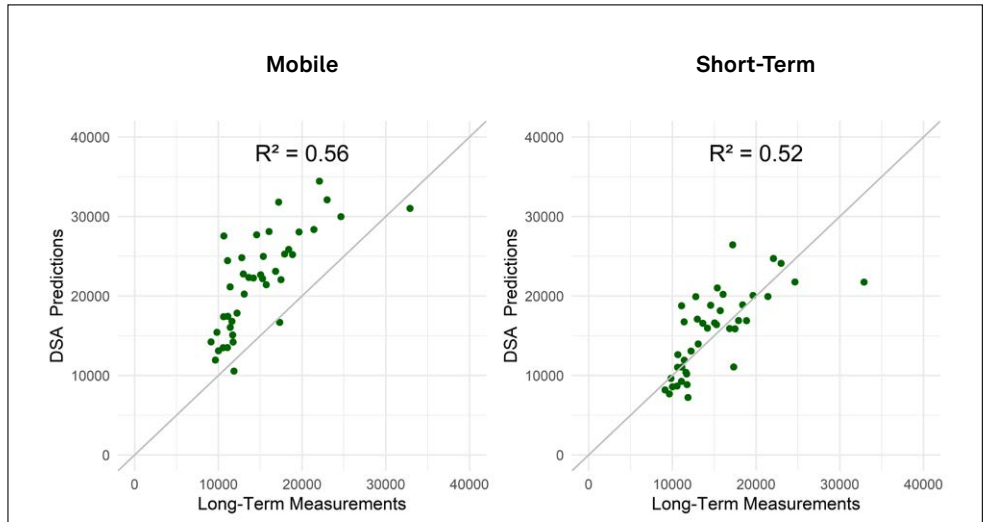
2.2 Deletion Substitution Addition (DSA) Algorithm

The DSA algorithm is related to stepwise regression in the sense that it is a covariate search algorithm that starts with an empty model and uses deletion (removing a variable from the model), substitution (replacing a variable in the model by another not in the model) or addition (adding a variable in the model) moves to find the final model. The final model is selected by minimizing the residual mean squared error (RMSE) using cross-validation[4]. It is possible to include interactions and polynomial transformations, but for this study, variables were kept in the original scale and interactions were not allowed. The DSA-algorithm was not able improve on the stepwise regression methods, even without supervision. Our dense data set reduced the advantages of the DSA approach, which has advantages in case of small (temporally unbalanced) data sets. External predictions based on the DSA method were like stepwise regression methods (Table 2). This was also found by Basagaña et al[5] who compared the DSA approach to a forward linear regression method.

Figure S4

Predicted UFP versus Long-Term Measurements.

↓



R-Script and parameters:

DSA: DSA(Y, X, maxsize = 12, maxorderint = 1, maxsumofpow = 1)

2.3 Regularization algorithms (LASSO, Elastic Net and Ridge)

Regularized regression techniques, like Ridge and LASSO regression, are regression techniques optimized for prediction. They can deal with correlated predictors by imposing a penalty on the absolute size of regression coefficients. Effects of two highly correlated variables are shrunken towards each other in Ridge regression whereas the LASSO tends to attribute the full effect to only one of the variables, shrinking the other towards zero. LASSO regression tends to result in effects for some variables to be shrunken all the way to zero, and therefore can be used easily for variable selection. Elastic net combines both a LASSO and a Ridge penalty. This combination encourages highly correlated variables to be averaged (ridge element), while the LASSO element encourages a sparse solution in the coefficients. The tuning parameter alpha that determines the relative weight of the ridge and LASSO penalties was set to 0.25, 0.50 and 0.75 [2].

R-Script and parameters:

LASSO: cv.glmnet(X, Y, alpha = 1, nfolds = 10, nlambda = 100, useMin=TRUE, loss = “deviance”)

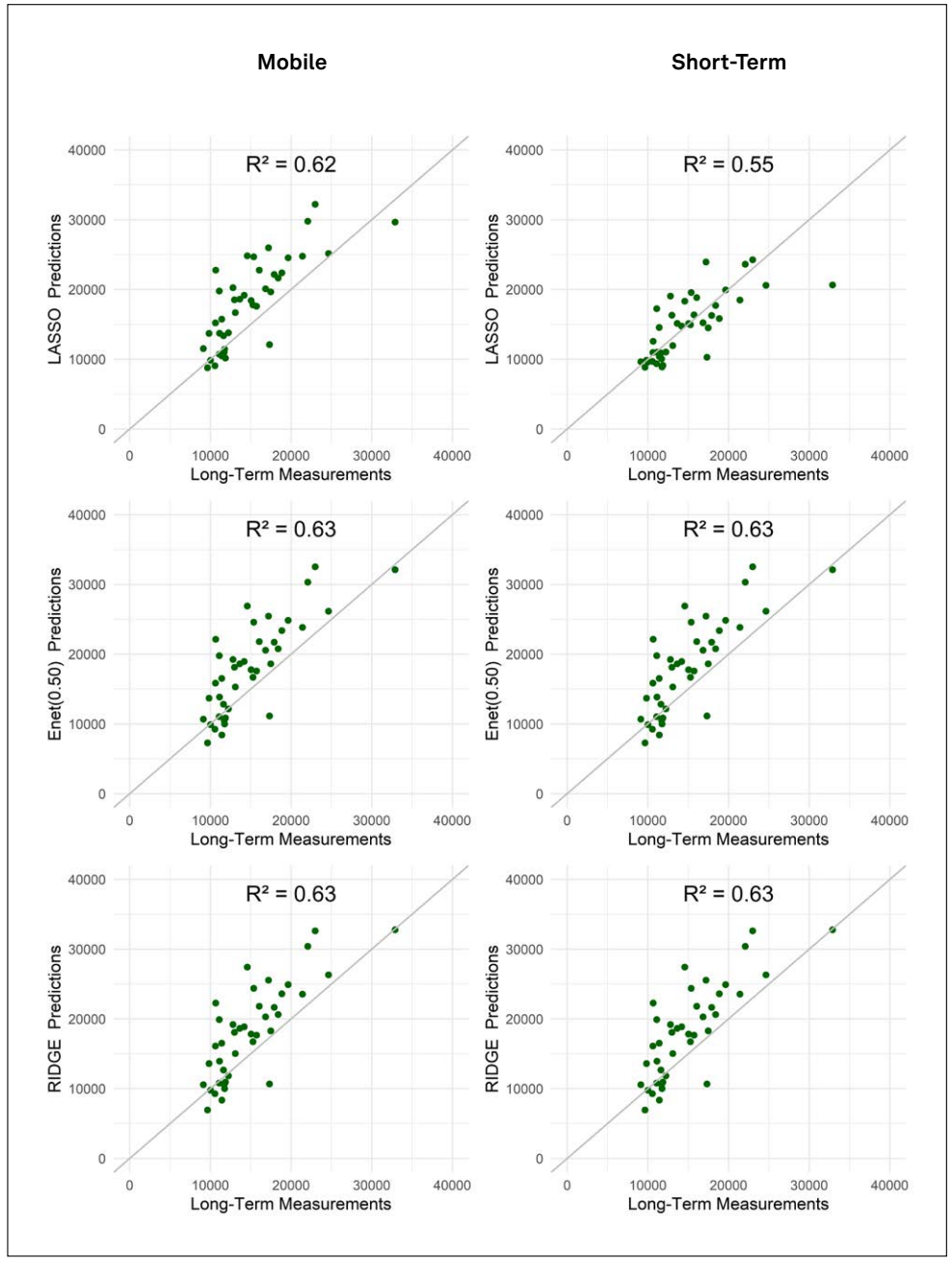
Elastic net: cv.glmnet(X, Y, alpha = 0.25/0.50/0.75, nfolds=10, nlambda=100, useMin=TRUE, loss=“deviance”)

RIDGE: MASS::lm.ridge(Y~., X)

Figure S5

Predicted UFP versus Long-Term Measurements.

↓



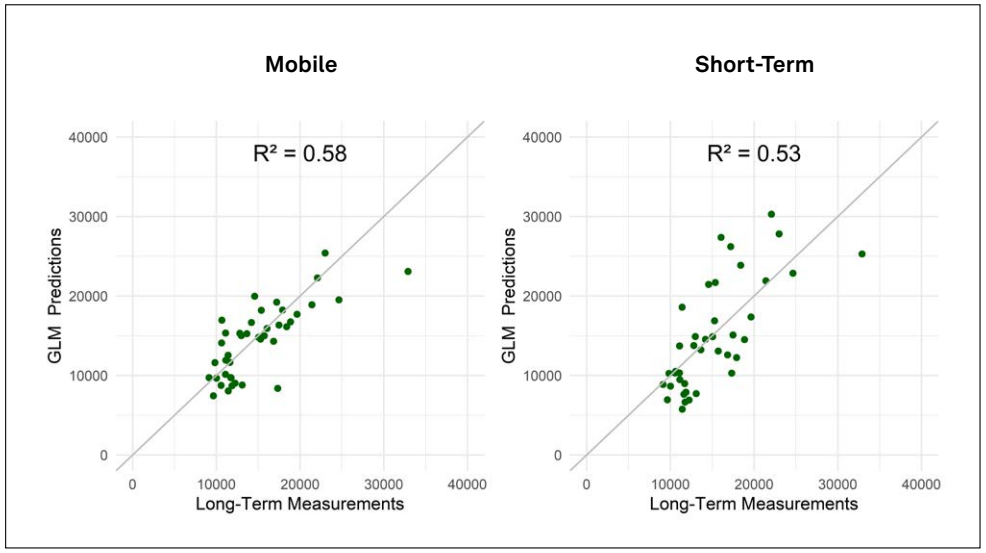
2.4 Generalized Linear Models

Generalized linear models (GLMs) offer an alternative to stepwise regression when the dependent variable is not normally distributed. In most studies that use a GLM to predict air pollutant concentrations, the dependent variable is transformed with the natural log[6]. In this analysis we applied a natural log transformation to the UFP data. Both training model R^2 and the ability to predict external measurements were similar to stepwise regression procedures (without log transformation) in both the mobile and short-term data set. Our data was slightly lognormally distributed, like in the study by Farrel et al[6].

Figure S6

Predicted UFP versus Long-Term Measurements.

↓



R-Script and parameters:

GLM: glm(log(Y) ~ ., X)

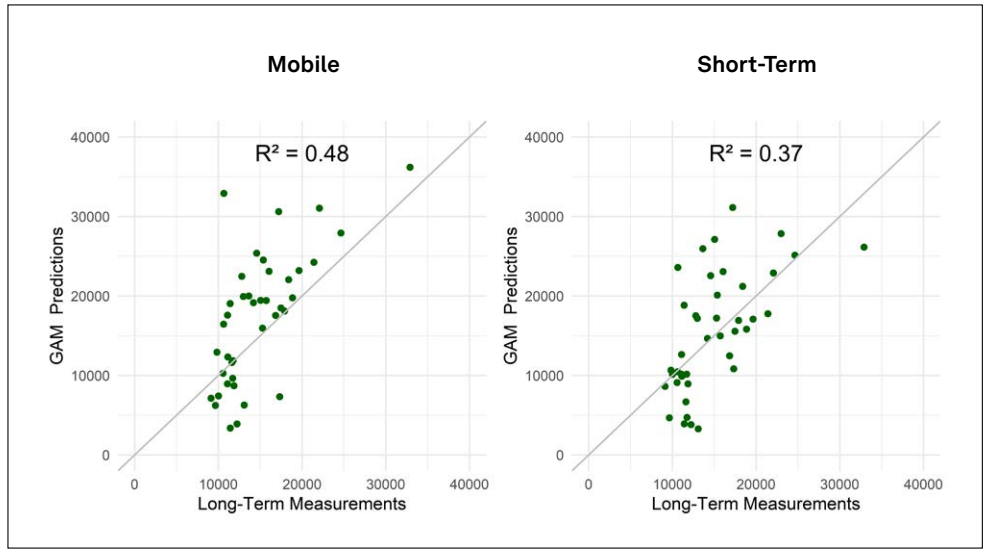
2.5 Additive Models (GAM)

When the relationship between predictor and outcome is non-linear, general additive models (GAM) are a solution, replacing the linear term in a regression model by a more general functional form[7]. It fits the nonlinear effects in flexible way, while retaining much of their interpretability[8]. Hasenfratz et al.[9] used GAMs to create 989 spatiotemporal surfaces explaining different time frames based on measurements collected with a mobile platform (average R^2 = 0.38). Patton et al.[7] also collected mobile measurement and used a GAM with local regression to create spatiotemporal models (R^2 0.43 to 0.53). Limitation of most (generalized) additive model algorithms is that it uses all offered variables which is often not feasible or desirable, especially when predictors are correlated[2]. Non-linearity can also be tackled by adaptive regression splines (e.g. Multivariate Adaptive Regression Splines (MARS)). MARS is an adaptive procedure for regression which uses piecewise linear base functions with knots at all unique observed values of each predictor. At every iteration all pairs are considered and the product that decreases the residual error the most is added to the model, up until all offered variables are considered[2].

Figure S7

Predicted UFP versus Long-Term Measurements.

↓



R-Script and parameters:

GAM: gam(gam.model, data = X, control=gam.control (maxit = 50, bf.maxit = 50)) with deg.gam = 2 and cts.num = 4

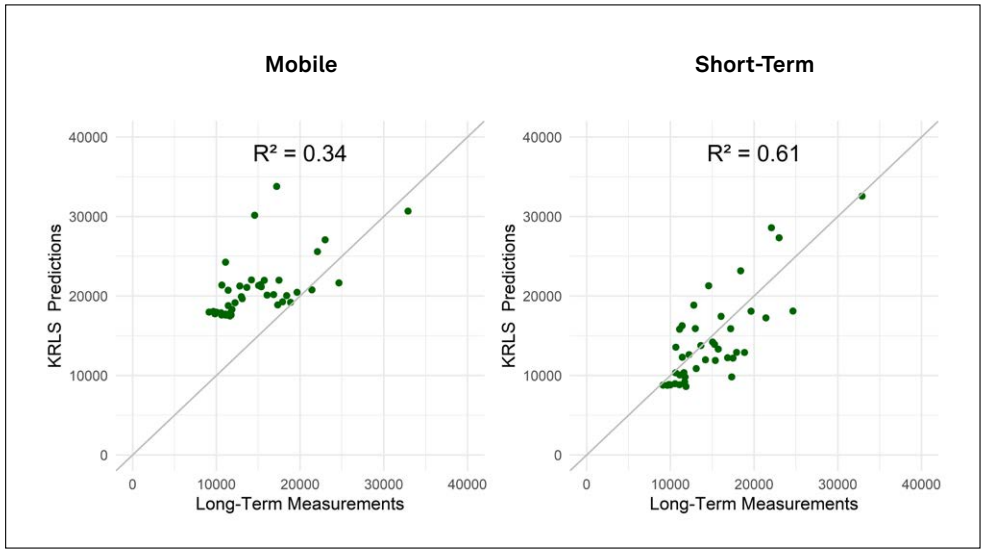
2.6 Kernel Regularized Least Squares (KRLS)

The KRLS estimator operates in an even larger space of possible functions based on the idea that observations with similar covariate values are expected to have similar outcomes on average[10]. The flexible KRLS estimator learns the functional form from the data, thereby protecting inferences against misspecification bias. Yet it nevertheless allows for interpretability and inference in ways similar to outputs of ordinary regression models. Furthermore, KRLS employs regularization which amounts to a prior preference for smoother functions over erratic ones. This allows KRLS to minimize over-fitting and reducing the variance and fragility of estimates[11]. KRLS was used for the development of a LUR model in Montreal, Canada[11] and included the same predictor variables as the supervised stepwise regression model but was able to create a better model evaluated with 10-fold cross validation. However, predictions on an external data set were similar for KRLS and multivariable linear regression. Same results were found in this study. Training model R^2 in the mobile data was 0.21 (compared to 0.13 for stepwise regression), but external prediction the stepwise linear procedures outperformed KRLS (0.62 vs. 0.46). Also, in the short-term data the KRLS predicted the training data better than stepwise regression (0.55 vs. 0.35).

Figure S8

Predicted UFP versus Long-Term Measurements.

↓



R-Script and parameters:

KRLS: krls(X, Y)

2.7 Neural Networks

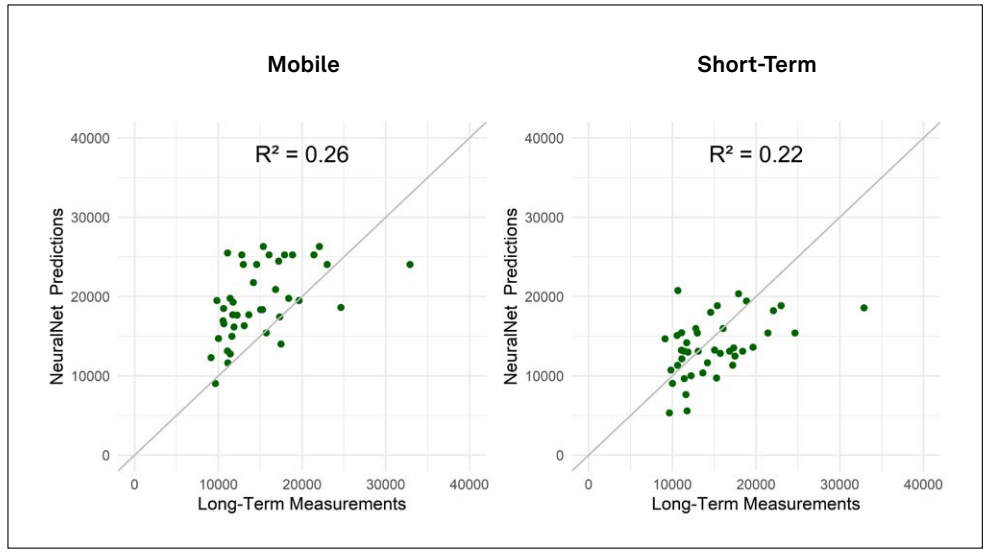
Flexibility can be also being obtained by using neural networks. A neural network can be thought of as the human brain with each unit representing a neuron. All units in a layer are connected to all units in the previous layer and so on. Most neural networks have one hidden layer which can be seen as a basis expansion of the original inputs[2]. The neural network is then a standard linear model using these transformations as inputs. Disadvantage of this technique is that it is quite difficult to train the model[12], and computational requirements can rise significantly when a lot of predictors in combination with a lot of units in the hidden layer are used.

None of our neural network approaches was able to create good spatial models for UFP. Neural networks do have computational efficiency and nonlinear characteristics and therefore often used in forecasting and spatiotemporal modeling[13]. Liu et al.[14], for example developed spatial LUR models and tried to update these models using meteorological data and a neural network to create yearly models. The authors found a significant increase in the predictive power when the neural network was added to the model opposed to the addition of the meteorological variables as independent variables.

Figure S9

Predicted UFP versus Long-Term Measurements.

↓



R-Script and parameters:

Neural Network: nnet(X, Y, size = 100, MaxNWts=7500, linout = TRUE, trace = FALSE, maxit = 500)

2.8 Support Vector Machines

Support Vector Machines (SVM) deals with flexibility issues by enlarging the feature space. In regression this is often done with smoothing kernels, where each point is replaced by a predefined distribution (for example gaussian). SVMs try to create a hyperplane/decision boundary were points on the correct side of the boundary and far away from it are ignored in the optimization. These low error points are the ones with the small residuals. All weight is given to points that are close to the decision boundary (called support vectors). These elements would change the position of the decision boundary if removed. SVM is therefore effective in high dimensional data, but causes problems when the noise in the data becomes larger[2]. Support vector machines are often used to predict temporal concentration levels[15,16]. Although these studies conclude that SVMs are a promising alternative to other methods, we do not see this in our results. The difference in the time resolution in the training data opposed to the external measurements could be an explanation for this, as SVMs are not suitable when there is lot of noise in the data[2]. For the short-term data set, where noise is much lower, models were similar or slightly better than linear stepwise regression. In the analyses of Van den Bossche et al.[17] they also included support vector regression (SVR) and concluded that SVR did not improve predictions compared with linear regression.

Figure S10

Predicted UFP versus Long-Term Measurements.

↓

R-Script and parameters:

SVM: e1071::svm(Y, X, , type.reg = “nu-regression”, type.class = “nu-classification”, kernel = “radial”, nu = 0.5, degree = 3, cost = 1, coef0 = 0)

2.9 RandomForest

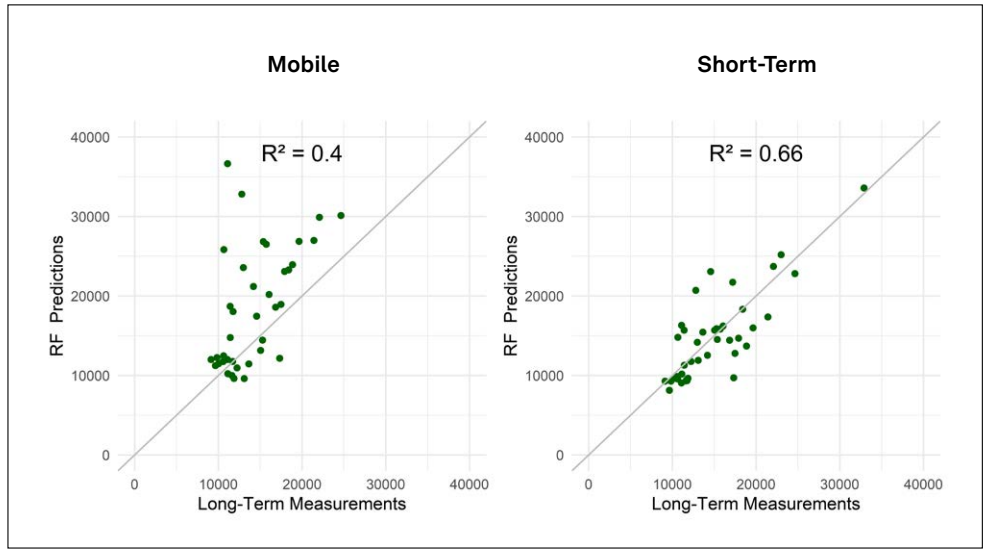
Another option is to use the average of a lot of predictions, by combining many weak classifiers/trees to a strong one, represented in our paper by bagging, boosting and random forest. A weak classifier is one whose error rate is only slightly better than random guessing.

Random forest restricts all regression trees to a limited number of variables, creating a large collection of ‘weak’ trees[2]. Each tree selects a specified number of variables at random and chooses the best variable to split the data. The same procedure applies to two splits and continues till all observations are partitioned[18]. Results of such algorithms lack some interpretability as no overview of how variables interact and contribute to the final prediction is given. Known disadvantages of such models is that they are harder to tune than other models used in this study and easily overfit data[2].

Figure S11

Predicted UFP versus Long-Term Measurements.

↓



R-Script and parameters:

randomForest(Y, X, mtry = ncol(X)/3, ntree = 1000, nodesize = 5, maxnodes = NULL)

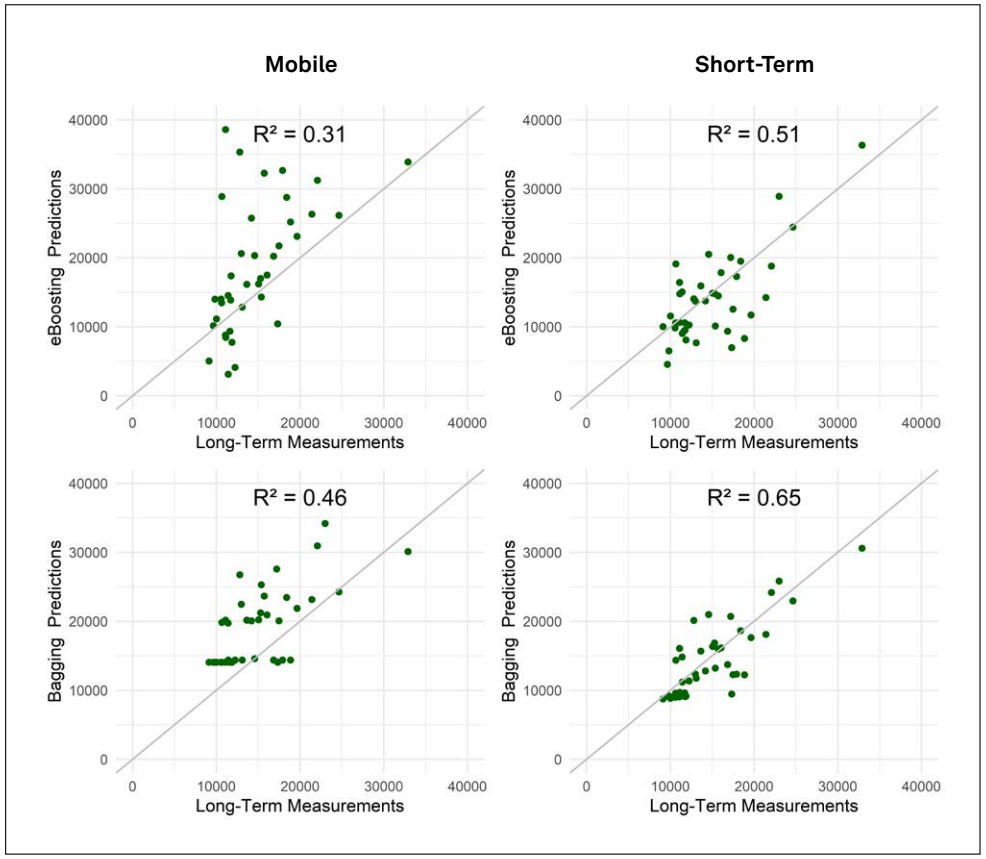
2.10 Boosting and Bagging

Bootstrap aggregating (bagging) averages the prediction over a collection of bootstrap samples, thereby reducing its variance. Each bootstrap tree will typically involve different features than the original and might have a different number of terminal nodes. The bagged estimate is then the average prediction from all trees[2]. Boosting builds on this principle, but is more powerful as it can modify the original input data by giving more weight to previously poorly predicted observations than well-predicted observations[19].

Figure S12

Predicted UFP versus Long-Term Measurements.

↓



R-Script and parameters:

GBM: gbm(gbm.model, X, n.trees = 1000, interaction.depth = 2, shrinkage = 0.001, cv.folds = 5)

Extreme Boosting: xgboost(X, ntrees=1000, max_depth = 4, shrinkage = 0.1, minobspernode=10, params = list(), nthread = 1)

Bagging: ipredbagg(Y,nbagg = 100, control = rpart.control(xval = 0, maxsurrogate = 0, minsplit = 20, cp = 0.01, maxdepth = 30)

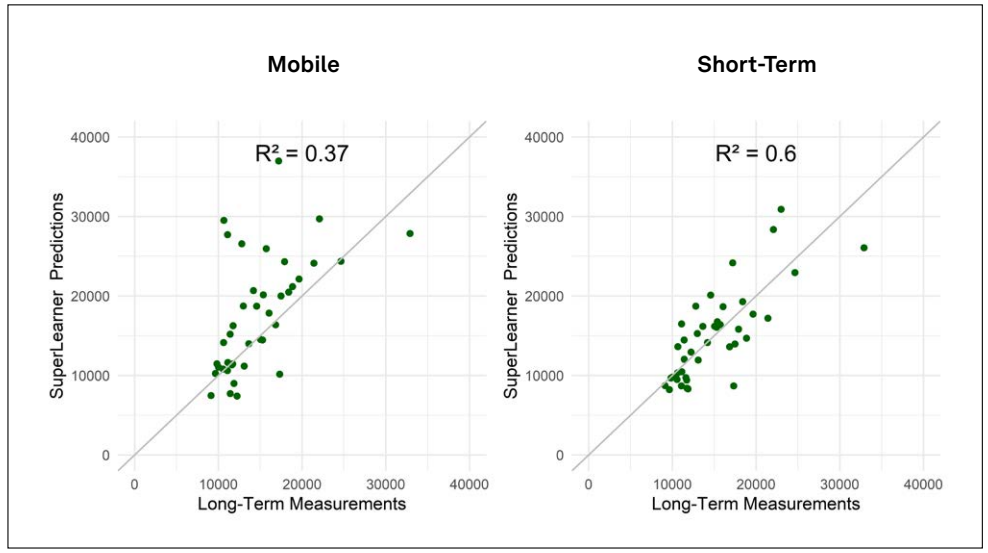
2.11 SuperLearner

Choosing one statistical model a priori limits the search for the best predictive model for the data. A stacking algorithm like Super Learner could therefore improve predictive performance. The super learner combines a specified collection of models and assigns weights with respect to results from least squares linear regression in the training data. Is uses cross validated predictions to avoid giving unfairly high weights to models with higher complexity[2]. This method differs in the subset of variables, the base functions, the loss functions, the searching algorithm, and the range of tuning parameters[20]. The super learner algorithm has been demonstrated previously in other settings to perform at least as well as the best choice among the library of candidate algorithms[21]. We tried to use default settings for a range of model algorithms, but for neural networks we changed a parameter to adjust the size as these were too computationally intensive.

Figure S13

Predicted UFP versus Long-Term Measurements.

↓



R-Script and parameters:

SuperLearner(Y, X, newX, SL.library = SL.library, method="method.NNLS", verbose=TRUE)

References

(1) Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A Review of Land-Use Regression Models to Assess Spatial Variation of Outdoor Air Pollution. *Atmos. Environ.* **2008**, *42* (33), 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057.

(2) Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning. *Elements* **2009**, *1*, 337–387. https://doi.org/10.1007/b94608.

(3) Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of Ultrafine Particles and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environ. Sci. Technol.* **2016**, *50* (23), 12894–12902. https://doi.org/10.1021/acs.est.6b03476.

(4) Beckerman, B. S.; Jerrett, M.; Martin, R. V.; Van Donkelaar, A.; Ross, Z.; Burnett, R. T. Application of the Deletion/Substitution/Addition Algorithm to Selecting Land Use Regression Models for Interpolating Air Pollution Measurements in California. *Atmos. Environ.* **2013**, *77*, 172–177. https://doi.org/10.1016/j.atmosenv.2013.04.024 Short communication.

(5) Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Vila, J.; et al. Effect of the Number of Measurement Sites on Land Use Regression Models in Estimating Local Air Pollution. *Atmos. Environ.* **2012**, *54*, 634–642. https://doi.org/10.1016/j.atmosenv.2012.01.064.

(6) Farrell, W.; Weichenthal, S.; Goldberg, M.; Valois, M. F.; Shekarrizfard, M.; Hatzopoulou, M. Near Roadway Air Pollution across a Spatially Extensive Road and Cycling Network. *Environ. Pollut.* **2016**, *212*, 498–507. https://doi.org/10.1016/j.envpol.2016.02.041.

(7) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An Hourly Regression Model for Ultrafine Particles in a Near-Highway Urban Area. *Environ. Sci. Technol.* **2014**, *48* (6), 3272–3280. https://doi.org/10.1021/es404838k.

(8) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; 2000. https://doi.org/10.1007/978-1-4614-7138-7.

(9) Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; Thiele, L. Deriving High-Resolution Urban Air Pollution Maps Using Mobile Sensor Nodes. In *Pervasive and Mobile Computing*; 2015; Vol. 16, pp 268–285. https://doi.org/10.1016/j.pmcj.2014.11.008.

(10) Ferwerda, J.; Hainmueller, J.; Hazlett, C. *KRLS: A Stata Package for Kernel-Based Regularized Least Squares*; 2013. https://doi.org/10.2139/ssrn.2325523.

(11) Weichenthal, S.; Ryswyk, K. Van; Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A Land Use Regression Model for Ambient Ultrafine Particles in Montreal, Canada: A Comparison of Linear Regression and a Machine Learning Approach. *Environ. Res.* **2016**, *146*, 65–72. https://doi.org/10.1016/j.envres.2015.12.016.

(12) Lima, A. R.; Cannon, A. J.; Hsieh, W. W. Nonlinear Regression in Environmental Sciences Using Extreme Learning Machines: A Comparative Evaluation. *Environ. Model. Softw.* **2015**, *73*, 175–188. https://doi.org/10.1016/j.envsoft.2015.08.002.

(13) Caselli, M.; Trizio, L.; De Gennaro, G.; Ielpo, P. A Simple Feedforward Neural Network for the PM_{10} Forecasting: Comparison with a Radial Basis Function Network and a Multivariate Linear Regression Model. *Water. Air. Soil Pollut.* **2009**, *201* (1–4), 365–377. https://doi.org/10.1007/s11270-008-9950-2.

(14) Liu, W.; Li, X.; Chen, Z.; Zeng, G.; León, T.; Liang, J.; Huang, G.; Gao, Z.; Jiao, S.; He, X.; et al. Land Use Regression Models Coupled with Meteorology to Model Spatial and Temporal Variability of NO_2 and PM_{10} in Changsha, China. *Atmos. Environ.* **2015**, *116*, 272–280. https://doi.org/10.1016/j.atmosenv.2015.06.056.

(15) Lu, W.-Z.; Wang, W.-J. Potential Assessment of the "Support Vector Machine" Method in Forecasting Ambient Air Pollutant Trends. *Chemosphere* **2005**, *59* (5), 693–701. https://doi.org/10.1016/j.chemosphere.2004.10.032.

(16) Luna, A. S.; Paredes, M. L. L.; de Oliveira, G. C. G.; Corrêa, S. M. Prediction of Ozone Concentration in Tropospheric Levels Using Artificial Neural Networks and Support Vector Machine at Rio de Janeiro, Brazil. *Atmos. Environ.* **2014**, *98*, 98–104. https://doi.org/10.1016/j.atmosenv.2014.08.060.

(17) Van den Bossche, J.; De Baets, B.; Verwaeren, J.; Botteldooren, D.; Theunis, J. Development and Evaluation of Land Use Regression Models for Black Carbon Based on Bicycle and Pedestrian Measurements in the Urban Environment. *Environ. Model. Softw.* **2018**, *99*, 58–69. https://doi.org/10.1016/J.ENVSOFT.2017.09.019.

(18) Song, B.; Wu, J.; Zhou, Y.; Hu, K. Fine-Scale Prediction of Roadside CO and NOx Concentration Based on a Random Forest Model. *J. Residuals Sci. Technol.* **2014**, *11*, 83–89.

(19) Reid, C. E.; Jerrett, M.; Petersen, M. L.; Pfister, G. G.; Morefield, P. E.; Tager, I. B.; Raffuse, S. M.; Balmes, J. R. Spatiotemporal Prediction of Fine Particulate Matter during the 2008 Northern California Wildfires Using Machine Learning. *Environ. Sci. Technol.* **2015**, *49* (6), 3887–3896. https://doi.org/10.1021/es505846r.

(20) Polley, E. C.; van der Laan, M. J. Super Learner in Prediction. *U.C. Berkeley Div. Biostat. Work. Pap. Ser.* **2010**, Paper 266.

(21) Pirracchio, R.; Petersen, M. L.; Van Der Laan, M. Improving Propensity Score Estimators' Robustness to Model Misspecification Using Super Learner. *Am. J. Epidemiol.* **2015**, *181* (2), 108–119. https://doi.org/10.1093/aje/kwu253.

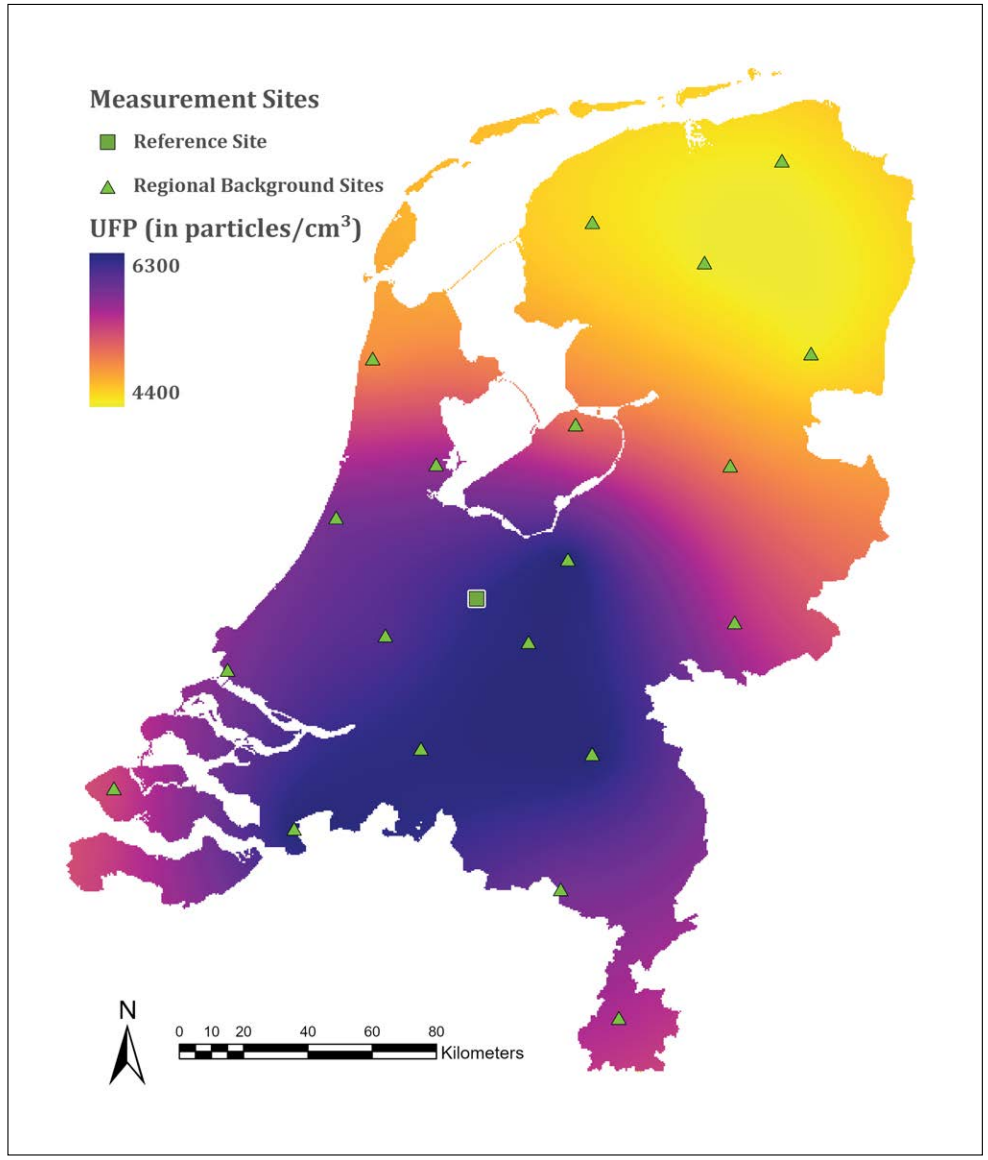
Spatial and Spatiotemporal Variability of Regional Background Ultrafine Particle Concentrations in The Netherlands

Esther van de Beek*
Jules Kerckhoffs*
Gerard Hoek
Geert Sterk
Kees Meliefste
Ulrike Gehring
Roel Vermeulen

*Contributed equally

Abstract

Studies of health effects of ultrafine particles (UFP) in large nationwide cohorts are currently hampered by a lack of knowledge about spatial and spatiotemporal variation of regional background UFP. We measured UFP (10-300 nm) at 20 regional background locations (3 x 2-weeks) across the Netherlands and a reference site continuously over a total period of 14 months in 2016-2017. We compared overall averages for each site and used kriging to create a regional background spatial map of the Netherlands. Spatiotemporal variability was analysed by correlating time-series of 2- and 24-hour average concentrations. Overall average measured UFP concentrations at the 20 locations ranged from 3814 to 7070 particles/cm³. We found spatial correlation in UFP concentrations up to 180 km and clear differences between the north and the more populated southern parts of the country. Average temporal correlation between 2-hour and 24-hour average UFP concentrations was 0,50 (IQR: 0,36-0,61) and 0,58 (IQR: 0,44-0,75), respectively. Temporal correlation declined weakly with distance between sites, from 0,58 for sites within 80km of each other to 0,47 for sites farther away. The substantial spatial variation in regional background UFP concentrations, suggests that regional variation may contribute importantly to exposure contrast in nationwide health studies of UFP.



① Introduction

Exposure to ambient air pollutants, such as particulate matter, is associated with adverse health effects [1]. Ultrafine particles (UFP; nanoparticles (<100nm)) behave differently than the larger fractions of particulate matter as they can penetrate deeper into the lungs and can enter the bloodstream [2,3]. However, evidence on the associations between long-term UFP exposure and cardiovascular and inflammatory health effects is still incomplete [4]. While models have been applied across longer timeframes using short-term UFP monitoring [5,6], there is a need for long-term monitoring of UFP and improved deterministic and empirical modelling to gain understanding of health effects of long-term UFP exposure[7–9]. So far, little is known about the spatial and spatiotemporal variability of UFP concentrations over larger distances.

Most UFP monitoring campaigns for epidemiological studies have been performed in single cities, often with mobile or short-term measurements [10–13], or focussing on spatiotemporal variability of UFP in urban environments [7,14–16]. The focus on urban areas is related to the concept that UFPs are mainly emitted by traffic and due to coagulation have a much shorter atmospheric lifetime than fine particles [17]. However, little empirical information is available about UFP concentrations at regional background locations. Consequently, health effects of long-term exposure to UFP cannot be readily studied in nation-wide cohorts, which have been very powerful in studies of health effects of $PM_{2.5}$ and NO_2 [7,18–21]. A study in California used a chemical transport model at 4 by 4 km resolution to calculate the background UFP concentration over large areas in California and suggested large variability (IQR: 778-1.747 ng/m^3) in background concentrations [22].

The aim of this study is to explore the spatial and spatiotemporal variation of UFP regional background concentrations across the Netherlands. Long-term measurements were conducted three times for two weeks on 20 different sites across the country. The first aim of this paper is to assess spatial variation of measured and interpolated average UFP concentrations based on a kriging approach. The second aim is to assess the temporal correlation between simultaneously measured sites in relation to distance between sites.

② Methods

2.1 Study design

We selected 20 regional background sites evenly spread across the Netherlands (Figure 1). For the purpose of this study, we defined regional background sites as locations in towns with a maximum of 10.000 inhabitants, outside of major urban areas. We specifically selected small towns instead of stations outside any settlement (as is common in routine monitoring networks), as our goal was to contribute to human exposure assessment of UFP. Sampling locations were selected to avoid local sources, such as gas stations or parking lots, restaurants, and local industry. Locations also did not have major roads (more than 10.000 vehicles/day) within a buffer of 3 km or were close to ports or airports. Sampling equipment was placed at ground-level in gardens of houses, away from local sources on the microlevel (e.g. extractor hood kitchen) similar as earlier studies [23].

UFP concentrations were measured (in 1-second resolution) three times for two weeks at each measurement location. Five instruments were available, so four sampling sites and the reference site were measured simultaneously. In each period we selected one site from each region to limit temporal variation influences on spatial variation. The set of four sites was equal for all three rounds. Every sampling site was measured during different seasons, totalling to 75 measurements (3 times 20 sampling sites and 15 reference site measurements). Because not all 20 sites were measured at the same time we used the reference site to temporally correct all site measurements [23]. All measurements were conducted between June 2016 and November 2017.

UFP number concentrations were monitored each second with Miniature Diffusion Size Classifiers (MiniDisc) [24]. This device measures particles between 10 – 300nm. UFP are generally defined as <100nm, but the major part of the total particle count is in 10-100nm range [8,14,17]. Therefore, we use UFP to refer to our total particle count number measurements. We did not include a comparison with other measurement devices in this study, but the monitors were compared with Condensation Particles Counters (CPCs) in a study by van Nunen et al. 2017, using the same instruments as our study. The ratio between devices was found to be 1.0.

2.2. Data cleaning

Measurements were removed when 1) there were error messages from the instrument (low flow, voltage); 2) the ratio between the UFP concentration and the UFP concentration of the subsequent second was above 10 or below 0,1 (with subsequent unrealistic data removed as well) [16]; or 3) the count was lower than 500 particles/cm³ (indicating unrealistic values). Individual 1-second measurements were removed in less than 0,1% of all cases. 2-hour average, 24-hour average and an overall mean concentration were then calculated for all sites and used for further analysis. When less than 75% of the data of a site were available (because of defective power supply) we did not calculate an overall mean concentration for that site. Because UFP concentrations were slightly skewed we log-transformed all averages for sensitivity analysis. From a health point of view, all results in the main paper are based on non-transformed data.

2.3 Data correction

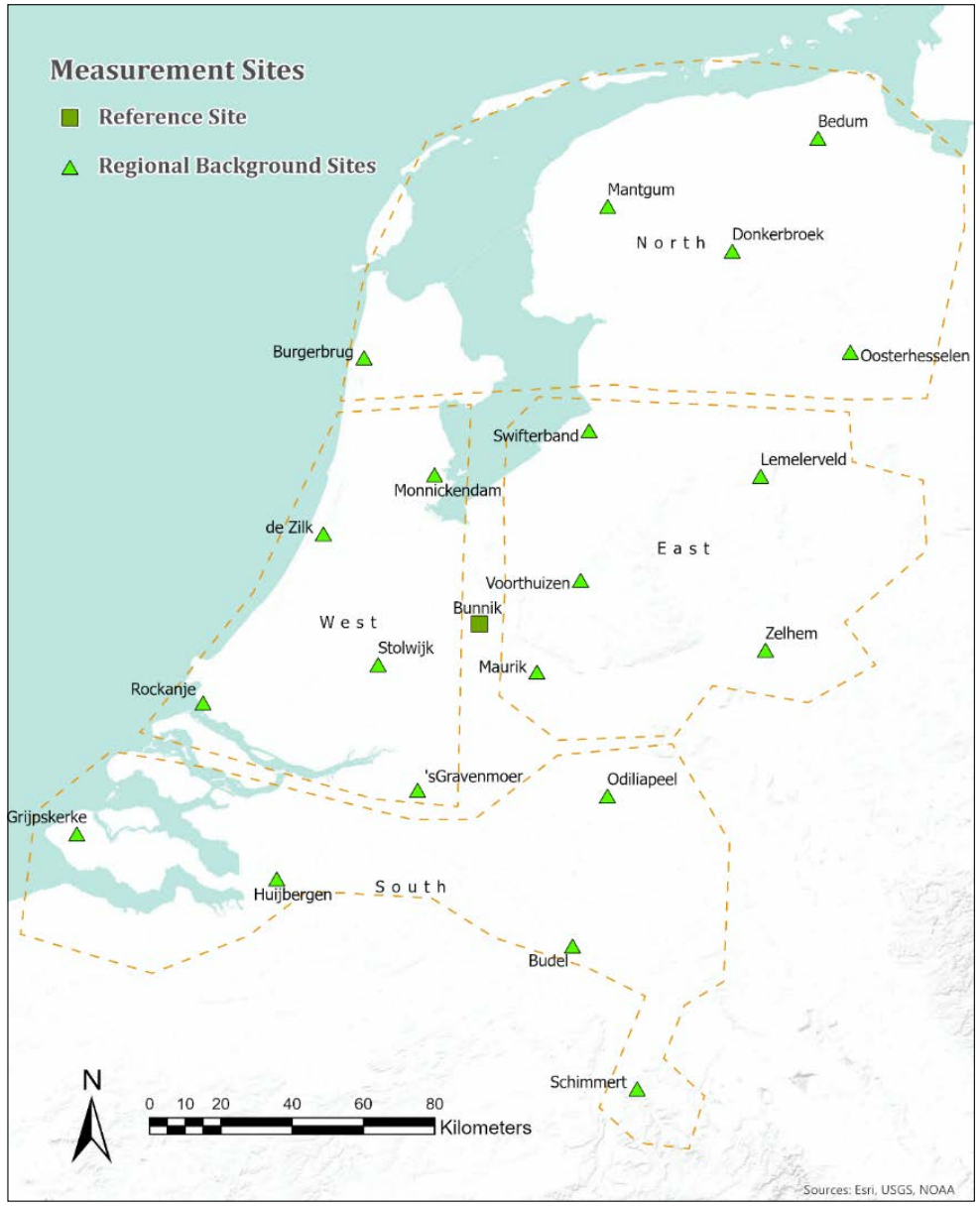
We corrected for differences between devices by co-locating them in a laboratory setting and corrected temporal differences for the spatial analyses using a reference site. These procedures are more elaborately described in the supplement information (Supplement A). In brief, differences between devices were assessed by co-locating all 5 devices for 1-2 days before and after every two-week measurement. We corrected each measurement campaign based on calculated median ratios between the reference device and other devices of each co-located comparison. All individual ratios are shown in supplement table A.1 with average medians ranging from 0,99 to 1,19.

To correct for temporal variation in the spatial analyses, we used a single reference site in the middle of the country with the same equipment as all other sites (Figure 1). The reference site had the same data correction procedure and restrictions as the sampling locations regarding potential local sources. Correction was performed using the difference method [25]. The overall average of the reference site is subtracted from the average of each of the 15 measurement periods at the reference site. This difference is then added to the measurements of the corresponding period at the sampling locations.

Figure 1

Distribution of the sampling locations in the Netherlands.

↓



Bunnik is the reference location (dark green square). Orange dashed line indicates the sites considered as North, East, West and South.

2.4 Spatial variability

To assess the spatial variability of UFP concentrations we performed a geostatistical space-time procedure and kriging interpolation [26]. In brief, this method assumes that the spatial correlation for all measurements combined equals the spatial correlation structure for all three measurement periods separately. The three measurements at the same sampling locations are treated as independent temporal replicates. This procedure is an adaption of a methodology that has previously been used to spatially model wind-blown mass transport based on a limited number of point measurements [26]. A full description of the space-time procedure and kriging approach (including semi-variogram) can be found in the supporting information (Supplement B).

2.5 Spatiotemporal correlation

We used 2-hour (within-day variation) and 24-hour (day-to-day variation) moving average for the spatiotemporal analysis. Correlations between all sampling locations (including the reference site) were examined by calculating Pearson, Spearman, and cross-correlations. Cross-correlation refers to correlation between concentrations with different lagged time periods. For example, the concentration at a certain site is correlated with the concentration at another site a number of hours earlier or later. In our study we assessed the maximum cross-correlation with lags up to 3 hours in both directions. We also assessed the influence of distance on the temporal correlation between two sampling sites.

We further checked the efficacy of the difference method to adjust for temporal variation using one reference site for the entire country. First, we applied the difference method to each 2-hour and 24-hour averaged data point in the time series. Next, we compared the variance of the temporally corrected time series and the variance of the unadjusted time series using the formula: *1 – variance(adjusted)/variance(unadjusted)* to calculate the proportion of variance explained (PVE). This means that a PVE between 0 and 1 relates to a decrease in variance. A PVE value below 0 means that the variance increased after using the difference method.

③ Results

Figure 2 shows the distribution of concentrations per period of measurements. Missing data due to equipment failure occurred for 10 out of 75 measurements (13%). For one reference measurement only 20% of the data was available. We decided to exclude this reference measurement and the measurements at the sampling locations of the same period. This resulted in 10 sampling locations with all three measurements available, 9 sampling sites with 2 measurements and one location with one measurement (Supplement A). All individual 2-week average concentrations per site are shown in figure 2 and supplement A. Temporally corrected biweekly UFP concentrations ranged from 3028 to 8202 particles/cm³. Concentrations were generally higher in the first round of measurements and were consistently lower in the North of the country compared to other regions of the Netherlands. Round 1 was mostly in the fall of 2016, round 2 in winter 2016-2017 and round 3 in summer and fall of 2017. The measured average concentrations over the 3 periods, ranged from 3814 to 7080 particles/cm³. UFP mean and median (less affected by high values) were highly correlated, documenting that outliers did not affect the pattern of 14-day average concentrations.

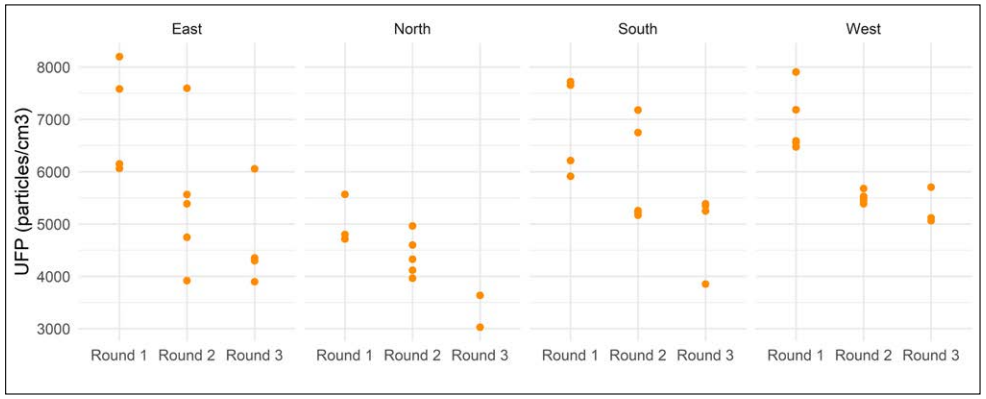
3.1 Spatial variability

The regional differences are shown in Figure 3. A spherical model was fitted through the computed standardized semi-variogram values (Figure B.1). The fitted spherical model has a range of 180,9 km, which indicates the presence of spatial correlation up until 181 km. The nugget (the value that the semi-variogram model attains at lag=0: intercept) and sill (the value that the semi-variogram model attains at the range) values are respectively 0,193 and 0,582. All interpolated maps (Figure B.2, i.e., maps per measurement period and the average map) showed that kriging resulted in a moderate loss of the variation compared to the measured variation. Predicted concentrations in the average kriging map range from 4.414 to 6.221 particles/cm³, whereas the overall measured average concentrations in the 3 periods, range from 3.814 to 7.080 particles/cm³. The error maps of the individual kriging interpolations (Figure B.2) and leave-one-out-cross-validation (LOOCV) (Figure B.3, B.4 and B.5) show that the uncertainty of the prediction increases as distance from observation locations increases and uncertainty is greater towards the edges of the map.

Figure 2

Distribution of temporally adjusted UFP concentrations.

↓



Distribution of temporally adjusted UFP concentrations for the three measurement periods by region. Dots represent individual mean of a measurement site in a single round.

3.2 Spatiotemporal correlations

Figure 4 shows the correlation between all simultaneously measured sampling sites, with all 1-second measurements averaged over 2 hours (Figure 4a) and with all measurements averaged over 24 hours (Figure 4b). Median correlations are 0,50 between 2-hour average concentrations at simultaneously measured sites and 0,58 between the 24-hour average measurements. A large variability in pairwise correlations is observed. Interquartile range for 2-hour averages was 0,36 to 0,61 and 0,44 to 0,75 for 24-hour averages. We found one site (Round 1 in “De Zilk”) with negative correlation to the reference site in our data (both for 2-hour and especially 24-hour average). Especially in the first week of the measurement there seem to be local influences that we cannot account for (Figure C.1). The average diurnal variation followed the same pattern at all sites including the reference site (Figure C.3). This indicates that measurements were not systematically impacted by local sources. Inspection of individual hourly average plots per study period per location generally gave no indication of incidental local sources, except for the first measurement round in one site (Figure C.1). The correlations for log-transformed PNC were slightly higher than for non-transformed (Table C.2).

Figure 3

Map of average UFP regional background concentrations in the Netherlands.

↓

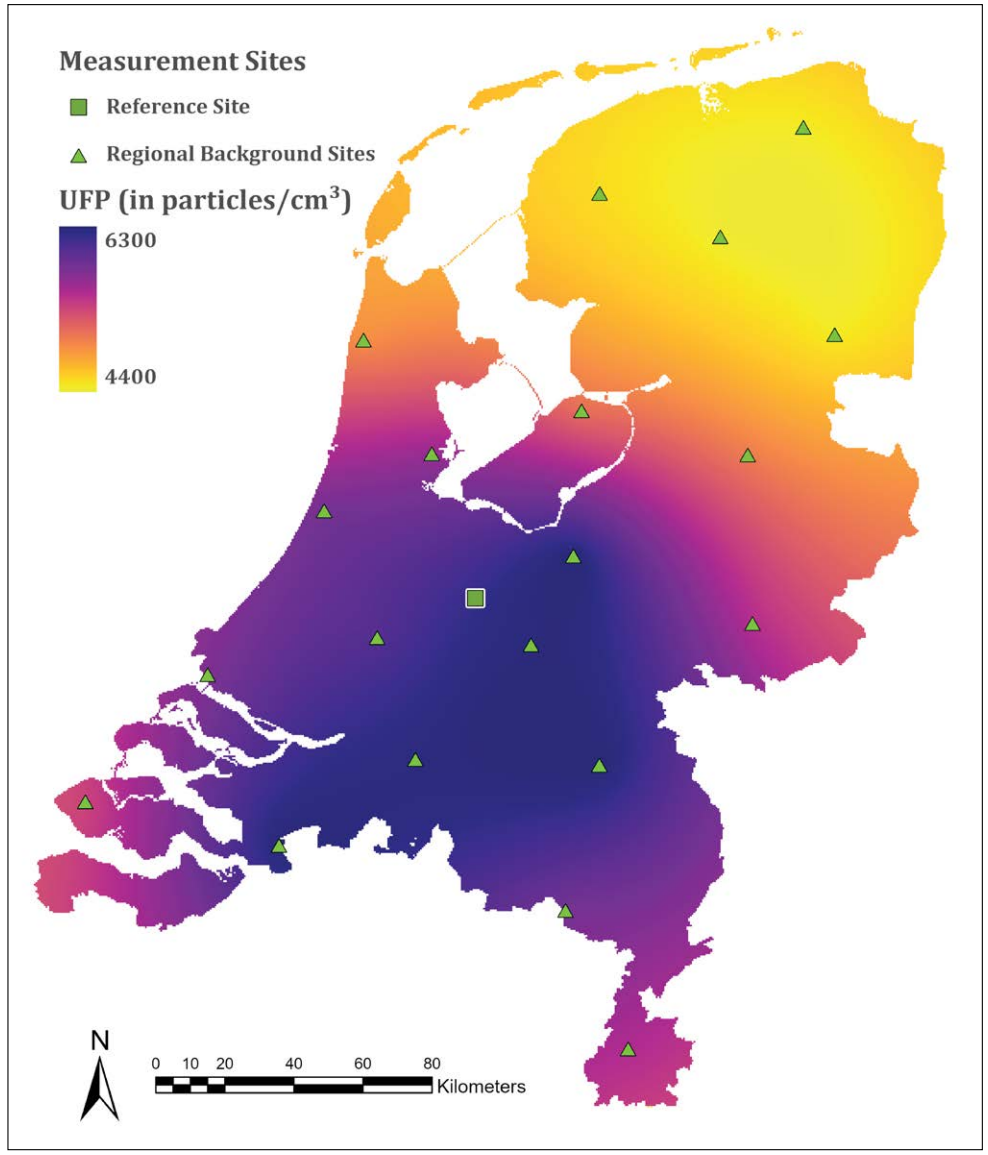
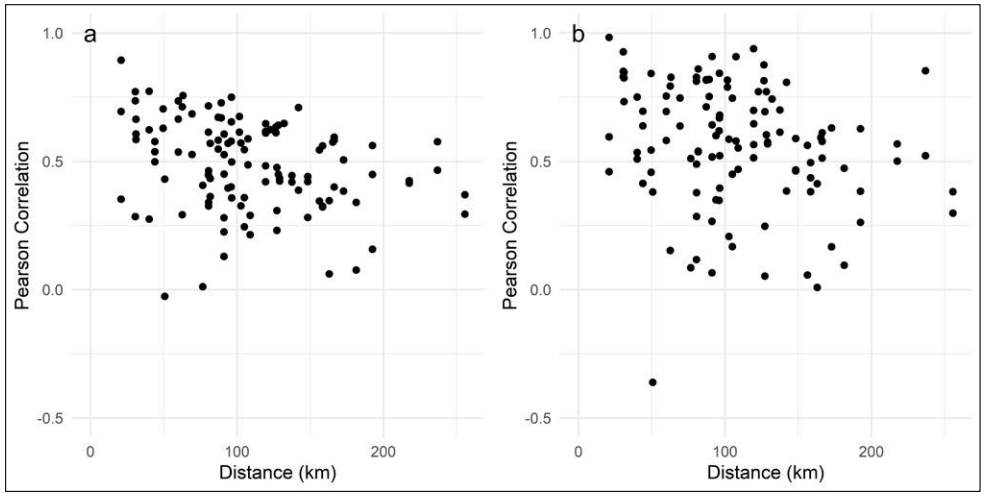


Figure 4

Scatterplot of pearson correlation versus distance.

↓



Pearson correlation calculated based on 2-hour (a) and 24-hour (b) average UFP concentrations at two simultaneously measured sampling sites, per sampling round of two weeks.

For all 2-hour average concentrations, we also show time-series plots in the supporting information (Figure C.1), with two examples shown in Figure 5. These plots show that UFP concentrations behave similar in time and magnitude for both the sampling location and reference site, independent of the actual concentrations of both sites. This can also be observed by the average diurnal pattern of each measurement site (Figure C.2). Figure 5 shows time series graphs of a sampling location close to the reference site (Maurik, 25 km, in the west of the country) and one further away (Donkerbroek, 125 km in the North of the country). The variation in concentrations of Maurik and the reference site is very similar, both in behaviour at base level (coarse variation across days), magnitude, and short-term variation (within day variation). The measurement at Donkerbroek still shows similar behaviour at base level as the reference site, but not regarding the short-term variation. Furthermore, the time series plot also demonstrates a considerable, though consistent difference in magnitude of approximately 4000 particles/cm³ when comparing Donkerbroek to the reference location. This is also reflected in the difference between the overall means of this period of Donkerbroek and the reference site.

The correlations and cross-correlations calculated in addition to the time series graphs corroborate these findings. Table C.1 shows the (cross-)correlations and corresponding time lags. The median maximum cross-correlation was slightly higher than the Pearson correlation: 0,54 (95% confidence interval 0,47 – 0,57). Time lags for the maximum cross-correlation exceeded 2 hours only in 2 occasions. Cross-correlations within sampling locations are quite consistent (Table C.1).

After correcting for temporal variation with the difference method, the variation decreased for 31 (63%) of the measurements (Table C.1), with an overall average of 8%. Variation decreased in 41 sites (84%) when only day-to-day (24hr) variation was assessed, with an overall average of 18%.

Table 1 shows the relationship between distance and correlation between simultaneously measured sites. Measurement sites were grouped based on the distance between sites in such way that each group contained about equal number of pairs. For sites that are within 50km of each other, the median Pearson correlation is 0,58, whereas the median correlation between sites more than 80km apart decreased to 0,44. Also, the proportion of explained variance is higher for sites that are close to the reference site, especially when day-to-day variance (PVE 24h) is used (Table 1). Median Pearson correlation values increased when data was log-transformed. Pearson correlation scores were 0,69 (>50km), 0,61 (50-80km) and 0,55 (>80km), see also table C.2.

Table 1

Influence of distance between sites on spatiotemporal correlation.

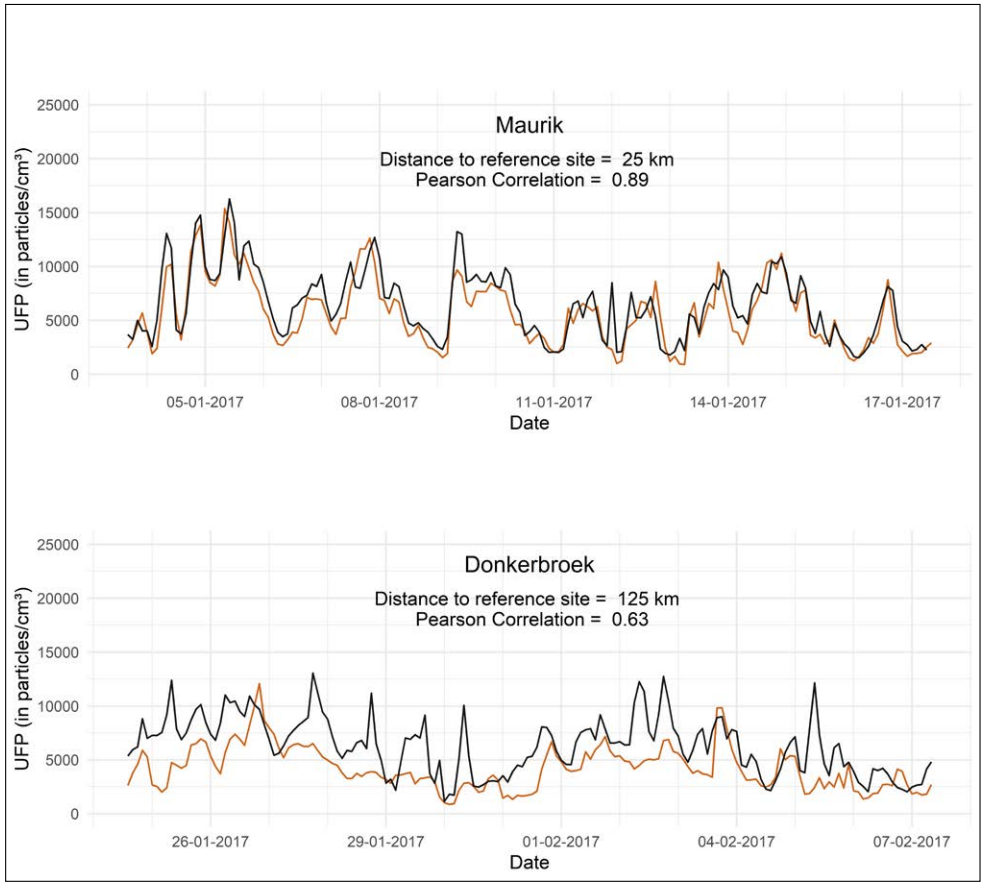
↓

Distance between sites	Median Pearson correlation 2h	Median cross correlation 2h	Median cross correlation 24h	Median PVE 2h	Median PVE 24h
< 50 km	0,58	0,58	0,50	0,30	0,24
50 - 80km	0,57	0,58	0,55	0,15	0,24
> 80km	0,44	0,47	0,43	0,23	0,05

Figure 5

Temporal variation of 2-hourly averaged UFP concentrations (particles/cm³) at the reference site (black) and the sampling locations (brown) Maurik and Donkerbroek for one round.

↓



③ Discussion

This paper describes the spatial and spatiotemporal variability of regional background UFP concentrations in the Netherlands. We found substantial spatial differences between regional background UFP concentrations, especially between the north and the more populated southern parts of the Netherlands. We found a moderate temporal correlation between the UFP concentrations at simultaneously measured sampling locations which varied by distance.

4.1 Spatial variability

We designed a long-term UFP measurements campaign because routine monitoring, nationwide chemical transport modelling (CTM) and satellite observations of UFP are not available in the Netherlands and in most countries. We then used a pooled kriging approach to assess spatial variation in this study, because of the limited number of sites. While relatively old, kriging can still be considered state-of the art modelling in settings with smoothly changing concentration patterns and limited local sources. Alternative methods, such as LUR methods are not well suited to describe large scale concentration patterns related to e.g. coagulation of ultrafine particles or new particle formation, especially when the number of sites drops below 20 [27,28]. Moreover, LUR models are limited to represent differences in background concentrations well for scales larger than about 10 km [29,30].

Regional background UFP concentrations differed across the Netherlands, with higher concentrations in the more urbanized southern parts of the Netherlands and lower concentrations in the North. This north-south difference is consistent with measured and modelled concentrations of other combustion source related air pollutants (such as soot, $PM_{2.5}$ and NO_2) in the Netherlands [7,31]. The lower concentration in the North of the country has been attributed to a combination of lower regional emissions from traffic, households and industry and larger distance to major foreign sources areas in Germany and Belgium particularly [23]. Relative differences (maximum/minimum) in that study were 1,5 for $PM_{2.5}$ and 2,2 for NO_2, respectively. In our study we found a ratio of 1,9.

Interpretation of spatial variation of UFP has mostly focussed on the urban and local (roadside) scale [17]. Relatively little is known about sources and factors at the regional scale. Compared to fine particles, UFP is expected to travel over shorter distances due to its short atmospheric lifetime [32,33]. While our monitoring sites were located away from major roads and small

local sources, there is increasing evidence that large source areas do affect UFP concentrations at distances of several 10-s of kilometres [34–36]. Several studies have now documented contributions of major airports to UFP concentrations at least 10 km away from the airport [34–37]. In the Netherlands, measurable UFP contributions of a major airport and a large industrial area including refineries were found at a regional background site up to 40 km away from the source area [34,35]. Ultrafine particles may travel relatively large distances because of the relatively slow coagulation processes in the diluted plumes or new particles may be formed from gaseous precursors. Our measurements may further be affected by residential sources including domestic heating, which in the Netherlands is primarily based upon natural gas. Wood burning is an increasing residential source as well. Regional nucleation events can have a large impact on UFP concentrations [17]. Due to the Netherlands being a flat and small country, the Netherlands can be considered as one airshed. Regional nucleation events therefore would contribute to smooth spatial patterns. Background concentrations of UFP are mainly driven by wind speed, precipitation, temperature, and relative humidity [38]. These weather factors do not differ much across the country.

Understanding regional differences in UFP concentrations is of great importance. While most people live in urban areas, it is not feasible to measure in every city or town in a country. Our measurements suggest that cities equal in local and urban sources but are located in different regions of the country experience different concentrations due to variation in regional background concentration. On top of that, regional differences we found in this study are often in the same order of magnitude as the predicted differences between streets within a city. For example, measured annual average UFP concentrations in the city of Augsburg, Germany ranged from 7.000 to 13.000 particles/cm^3 [12], and predictor variables in a previous developed UFP model in the Netherlands were able to distinguish differences up to 5.000 particles/cm^3 between the 90th and 10th percentile for each predictor [11]. Klompmaker et al.[25] measured short-term concentrations in two Dutch cities and found an average difference of 5.601 particles/cm^3 between traffic (15.464 particles/cm^3) and urban background sites (9.863 particles/cm^3). Regional background differences up to 3.000 particles/cm^3 are therefore important to consider in health studies covering populations in a large geographic area. Regional background variation may contribute significantly to the overall variability in exposure in a nationwide health study. We specifically used the kriging map of this study to represent the regional background component in a national land use regression model. The national LUR is now further derived from mobile monitoring data and local and urban source predictors[39]. The importance of the regional background data is therefore mostly in distinguishing UFP exposure in towns and cities in different parts of the country.

4.2 Spatiotemporal correlation

The median cross-correlation of 0,54 indicates a moderate temporal relationship between the sampling sites and the reference location. Hofman et al.[14] compared long-term UFP measurements in four different European cities and found similar patterns. The highest correlation they found was between London and Leicester (143 km, Spearman correlation = 0,50). Such moderate correlations support the use of a reference site to correct for temporal variation over distances up 200km (like the Netherlands), but suggest that this adjustment is of limited value, as supported by our variance reduction analysis. Therefore, sampling designs to minimize bias due to temporal variation in assessing spatial variation are needed. Studies analysing acute health effects of UFP exposure, like time series studies, are still restricted to measurement stations that are close to all participants.

Most studies that sample air pollution concentrations at different locations at various time points use a central reference site to adjust for temporal variation. The range of the variogram of 181km indicates the presence of spatial autocorrelation up to about 200 km. The maximum distance between the sampling sites and the reference location in this study was 165 km. We did find higher correlations for sampling locations close to the reference site but found only a small decrease in correlation with increasing distance. Furthermore, the difference method for correction of temporal variation decreased the variation in the time series of the sampling location in only 63% of the cases.

The time lags of the maximum cross correlations between concentrations at a sampling location and the reference site were 2 hours or smaller in almost all cases. This supports our averaging period of 2 hours for all temporal analyses. Correlations with (cross-correlation) and without (Pearson) lags were therefore similar. Not many studies explore time lags between time series of air pollution concentrations between sampling sites. A study in China, examining $PM_{2.5}$ concentration time series lags between Beijing and Zhangjiakou, reported time lags of 6 to 50 hours between the two cities, approximately 200 km apart [40]. A lag of one or two hours between sites that are more than 100 km apart based on transport between the two locations seems unlikely, because this would require a constant wind speed of 50-100 km/h, whereas the average wind speed in the Netherlands is 8km/h.

4.3 Strengths and Limitations

In this study, we were able to measure regional background UFP across a wide geographic area over a larger period of time. So far, most UFP

measurement campaigns have been conducted within cities as local UFP sources are concentrated in major cities and monitoring is costly and labour intensive. The sample size of this study (20 sampling locations) was still rather small to produce nationwide maps. Yet, with our kriging approach, we were able to make interpolations with limited prediction error despite the small number of locations and measurements. The distribution of the predicted concentrations throughout the country was consistent for all three measurement periods (Figure B.2). This indicates that the interpolated map gives us a reasonably accurate general description of the spatial variability of regional background UFP concentration in the Netherlands, though we cannot exclude that we have missed specific regional patterns due to the limited number of sites. Higher density of the monitoring network would increase the accuracy of the prediction.

The method of producing a pooled variogram assumes independence between the three measurement periods. However, the repeated measurements of concentrations at a specific sampling site are likely (temporally) correlated between periods because we expect the emission sources that influence the difference between various sampling locations to be largely consistent over time. We consider the temporal independence between measurements at the same sites to be large enough to justify the use of a pooled variogram, because the measurements were far apart enough in time. On average, measurements at the same site were 165 days apart. Our study was primarily empirical and not designed to elucidate the physicochemical processes explaining the factors contributing to spatial and temporal variation.

Supporting Information Available. Supporting information contains extra information on the data correction procedures (Supplement A), elaborate explanation of pooled kriging approach with spatial variation of UFP over the different rounds (Supplement B), spatiotemporal correlation statistics and all time-series figures (Supplement C). This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) HEI. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. *HEI* **2010**, *Special Re* (January), 1–386.

(2) Oberdörster, G.; Oberdörster, E.; Oberdörster, J. Nanotoxicology: An Emerging Discipline Evolving from Studies of Ultrafine Particles. *Environ. Health Perspect.* **2005**, *113* (7), 823–839. https://doi.org/10.1289/ehp.7339.

(3) Peters, A.; Veronesi, B.; Calderón-Garcidueñas, L.; Gehr, P.; Chen, L.; Geiser, M.; Reed, W.; Rothen-Rutishauser, B.; Schürch, S.; Schulz, H. Translocation and Potential Neurological Effects of Fine and Ultrafine Particles a Critical Update. *Part. Fibre Toxicol.* **2006**, *3* (1), 13. https://doi.org/10.1186/1743-8977-3-13.

(4) Ohlwein, S.; Kappeler, R.; Kutlar Joss, M.; Künzli, N.; Hoffmann, B. Health Effects of Ultrafine Particles: A Systematic Literature Review Update of Epidemiological Evidence. *International Journal of Public Health*. Springer International Publishing May 1, 2019, pp 547–559. https://doi.org/10.1007/s00038-019-01202-7.

(5) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712–8720. https://doi.org/10.1021/es505791g.

(6) Jones, R. R.; Hoek, G.; Fisher, J. A.; Hasheminassab, S.; Wang, D.; Ward, M. H.; Sioutas, C.; Vermeulen, R.; Silverman, D. T. Land Use Regression Models for Ultrafine Particles, Fine Particles, and Black Carbon in Southern California. *Sci. Total Environ.* **2020**, *699*, 134234. https://doi.org/10.1016/j.scitotenv.2019.134234.

(7) Eeftens, M.; Phuleria, H. C.; Meier, R.; Aguilera, I.; Corradi, E.; Davey, M.; Ducret-Stich, R.; Fierz, M.; Gehrig, R.; Ineichen, A.; Keidel, D.; Probst-Hensch, N.; Ragettli, M. S.; Schindler, C.; Künzli, N.; Tsai, M. Y. Spatial and Temporal Variability of Ultrafine Particles, NO_2, $PM_{2.5}$, $PM_{2.5}$ Absorbance, PM_{10} and PMcoarse in Swiss Study Areas. *Atmos. Environ.* **2015**, *111*, 60–70.

(8) Hennig, F.; Quass, U.; Hellack, B.; Küpper, M.; Kuhlbusch, T. A. J.; Stafoggia, M.; Hoffmann, B. Ultrafine and Fine Particle Number and Surface Area Concentrations and Daily Cause-Specific Mortality in the Ruhr Area, Germany, 2009–2014. *Environ. Health Perspect.* **2018**, *126* (2). https://doi.org/10.1289/EHP2054.

(9) Downward, G. S.; van Nunen, E. J. H. M.; Kerckhoffs, J.; Vineis, P.; Brunekreef, B.; Boer, J. M. A.; Messier, K. P.; Roy, A.; Verschuren, W. M. M.; van Der Schouw, Y. T.; Sluijs, I.; Gulliver, J.; Hoek, G.; Vermeulen, R. Long-Term Exposure to Ultrafine Particles and Incidence of Cardiovascular and Cerebrovascular Disease in a Prospective Study of a Dutch Cohort. *Environ. Health Perspect.* **2018**, *126* (12). https://doi.org/10.1289/EHP3047.

(10) Hatzopoulou, M.; Valois, M. F.; Levy, I.; Mihele, C.; Lu, G.; Bagg, S.; Minet, L.; Brook, J. Robustness of Land-Use Regression Models Developed from Mobile Air Pollutant Measurements. *Environ. Sci. Technol.* **2017**, *51* (7), 3938–3947. https://doi.org/10.1021/acs.est.7b00366.

(11) Kerckhoffs, J.; Hoek, G.; Vlaanderen, J.; van Nunen, E.; Messier, K.; Brunekreef, B.; Gulliver, J.; Vermeulen, R. Robustness of Intra Urban Land-Use Regression Models for Ultrafine Particles and Black Carbon Based on Mobile Monitoring. *Environ. Res.* **2017**, *159*, 500–508. https://doi.org/10.1016/j.envres.2017.08.040.

(12) Wolf, K.; Cyrys, J.; Harciníková, T.; Gu, J.; Kusch, T.; Hampel, R.; Schneider, A.; Peters, A. Land Use Regression Modeling of Ultrafine Particles, Ozone, Nitrogen Oxides and Markers of Particulate Matter Pollution in Augsburg, Germany. *Sci. Total Environ.* **2017**. https://doi.org/10.1016/j.scitotenv.2016.11.160.

(13) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An Hourly Regression Model for Ultrafine Particles in a Near-Highway Urban Area. *Environ. Sci. Technol.* **2014**, *48* (6), 3272–3280. https://doi.org/10.1021/es404838k.

(14) Hofman, J.; Staelens, J.; Cordell, R.; Stroobants, C.; Zikova, N.; Hama, S. M. L.; Wyche, K. P.; Kos, G. P. A.; Van Der Zee, S.; Smallbone, K. L.; Weijers, E. P.; Monks, P. S.; Roekens, E. Ultrafine Particles in Four European Urban Environments: Results from a New Continuous Long-Term Monitoring Network. *Atmos. Environ.* **2016**, *136*, 68–81. https://doi.org/10.1016/j.atmosenv.2016.04.010.

(15) Von Bismarck-Osten, C.; Birmili, W.; Ketzel, M.; Massling, A.; Petäjä, T.; Weber, S. Characterization of Parameters Influencing the Spatio-Temporal Variability of Urban Particle Number Size Distributions in Four European Cities. *Atmos. Environ.* **2013**, *77*, 415–429. https://doi.org/10.1016/j.atmosenv.2013.05.029.

(16) Puustinen, A.; Hämeri, K.; Pekkanen, J.; Kulmala, M.; de Hartog, J.; Meliefste, K.; ten Brink, H.; Kos, G.; Katsouyanni, K.; Karakatsani, A.; Kotronarou, A.; Kavouras, I.; Meddings, C.; Thomas, S.; Harrison, R.; Ayres, J. G.; van der Zee, S.; Hoek, G. Spatial Variation of Particle Number and Mass over Four European Cities. *Atmos. Environ.* **2007**, *41* (31), 6622–6636. https://doi.org/10.1016/j.atmosenv.2007.04.020.

(17) HEI Review Panel. Understanding the Health Effects of Ambient Ultrafine Particles. *Heal. Eff. Inst.* **2013**, No. January, 122.

(18) Milanzi, E. B.; Koppelman, G. H.; Smit, H. A.; Wijga, A. H.; Oldenwening, M.; Vonk, J. M.; Brunekreef, B.; Gehring, U. Air Pollution Exposure and Lung Function until Age 16 Years: The PIAMA Birth Cohort Study. *Eur. Respir. J.* **2018**, *52* (3). https://doi.org/10.1183/13993003.00218-2018.

(19) Pope, C. A.; Burnett, R. T.; Thun, M. J.; Calle, E. E.; Krewski, D.; Ito, K.; Thurston, G. D. Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to Fine Particulate Air Pollution. *J. Am. Med. Assoc.* **2002**, *287* (9), 1132–1141. https://doi.org/10.1001/jama.287.9.1132.

(20) Fischer, P. H.; Marra, M.; Ameling, C. B.; Hoek, G.; Beelen, R.; de Hoogh, K.; Breugelmans, O.; Kruize, H.; Janssen, N. A. H.; Houthuijs, D. Air Pollution and Mortality in Seven Million Adults: The Dutch Environmental Longitudinal Study (DUELS). *Environ. Health Perspect.* **2015**, *123* (7), 697–704. https://doi.org/10.1289/ehp.1408254.

(21) Beelen, R.; Raaschou-Nielsen, O.; Stafoggia, M.; Andersen, Z. J.; Weinmayr, G.; Hoffmann, B.; Wolf, K.; Samoli, E.; Fischer, P.; Nieuwenhuijsen, M.; Vineis, P.; Xun, W. W.; Hoek, G. Effects of Long-Term Exposure to Air Pollution on Natural-Cause Mortality: An Analysis of 22 European Cohorts within the Multicentre ESCAPE Project. *Lancet* **2014**, *383* (9919), 785–795. https://doi.org/10.1016/S0140-6736(13)62158-3.

(22) Ostro, B.; Hu, J.; Goldberg, D.; Reynolds, P.; Hertz, A.; Bernstein, L.; Kleeman, M. J. Associations of Mortality with Long-Term Exposures to Fine and Ultrafine Particles, Species and Sources: Results from the California Teachers Study Cohort. *Environ. Health Perspect.* **2015**, *123* (6), 549–556. https://doi.org/10.1289/ehp.1408565.

(23) Eeftens, M.; Tsai, M. Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; De Nazelle, A.; de Vocht, F.; Declercq, C.; Dedele, A.; Eriksen, K.; Galassi, C.; Gražulevičiene, R.; Grivas, G.; Heinrich, J.; Hoffmann, B.; Iakovides, M.; Ineichen, A.; Katsouyanni, K.; Korek, M.; Krämer, U.; Kuhlbusch, T.; Lanki, T.; Madsen, C.; Meliefste, K.; Mölter, A.; Mosler, G.; Nieuwenhuijsen, M.; Oldenwening, M.; Pennanen, A.; Probst-Hensch, N.; Quass, U.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Udvardy, O.; Vaskövi, É.; Weinmayr, G.; Brunekreef, B.; Hoek, G. Spatial Variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ Absorbance and PMcoarse Concentrations between and within 20 European Study Areas and the Relationship with NO_2 - Results of the ESCAPE Project. *Atmos. Environ.* **2012**, *62*, 303–317. https://doi.org/10.1016/j.atmosenv.2012.08.038.

(24) Fierz, M.; Houle, C.; Steigmeier, P.; Burtscher, H. Design, Calibration, and Field Performance of a Miniature Diffusion Size Classifier. *Aerosol Sci. Technol.* **2011**, *45* (1), 1–10. https://doi.org/10.1080/02786826.2010.516283.

(25) Klompmaker, J. O.; Montagne, D. R.; Meliefste, K.; Hoek, G.; Brunekreef, B. Spatial Variation of Ultrafine Particles and Black Carbon in Two Cities: Results from a Short-Term Measurement Campaign. *Sci. Total Environ.* **2015**, *508*, 266–275. https://doi.org/10.1016/j.scitotenv.2014.11.088.

(26) Sterk, G.; Stein, A. Mapping Wind-Blown Mass Transport by Modeling Variability in Space and Time. *Soil Sci. Soc. Am. J.* **1997**, *61* (1), 232. https://doi.org/10.2136/sssaj1997.03615995006100010032x.

(27) Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Vila, J.; Künzli, N. Effect of the Number of Measurement Sites on Land Use Regression Models in Estimating Local Air Pollution. *Atmos. Environ.* **2012**, *54*, 634–642. https://doi.org/10.1016/j.atmosenv.2012.01.064.

(28) Wang, M.; Beelen, R.; Basagana, X.; Becker, T.; Cesaroni, G.; De Hoogh, K.; Dedele, A.; Declercq, C.; Dimakopoulou, K.; Eeftens, M.; Forastiere, F.; Galassi, C.; Gražulevičiene, R.; Hoffmann, B.; Heinrich, J.; Iakovides, M.; Künzli, N.; Korek, M.; Lindley, S.; Mölter, A.; Mosler, G.; Madsen, C.; Nieuwenhuijsen, M.; Phuleria, H.; Pedeli, X.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Stempfelet, M.; Tsai, M. Y.; Lanki, T.; Udvardy, O.; Varró, M. J.; Wolf, K.; Weinmayr, G.; Yli-Tuomi, T.; Hoek, G.; Brunekreef, B. Evaluation of Land Use Regression Models for NO2 and Particulate Matter in 20 European Study Areas: The ESCAPE Project. *Environ. Sci. Technol.* **2013**, *47* (9), 4357–4364. https://doi.org/10.1021/es305129t.

(29) Kerckhoffs, J.; Wang, M.; Meliefste, K.; Malmqvist, E.; Fischer, P.; Janssen, N. A. H.; Beelen, R.; Hoek, G. A National Fine Spatial Scale Land-Use Regression Model for Ozone. *Environ. Res.* **2015**, *140*, 440–448. https://doi.org/10.1016/j.envres.2015.04.014.

(30) Hoek, G.; Eeftens, M.; Beelen, R.; Fischer, P.; Brunekreef, B.; Boersma, K. F.; Veefkind, P. Satellite NO2 Data Improve National Land Use Regression Models for Ambient NO2 in a Small Densely Populated Country. *Atmos. Environ.* **2015**, *105*, 173–180. https://doi.org/10.1016/j.atmosenv.2015.01.053.

(31) Schmitz, O.; Beelen, R.; Strak, M.; Hoek, G.; Soenario, I.; Brunekreef, B.; Vaartjes, I.; Dijst, M. J.; Grobbee, D. E.; Karssenberg, D. Data Descriptor: High Resolution Annual Average Air Pollution Concentration Maps for the Netherlands. *Sci. Data* **2019**, *6*. https://doi.org/10.1038/sdata.2019.35.

(32) Junkermann, W.; Hacker, J. M. Ultrafine Particles in the Lower Troposphere Major Sources, Invisible Plumes, and Meteorological Transport Processes. *Bull. Am. Meteorol. Soc.* **2018**, *99* (12), 2587–2602. https://doi.org/10.1175/BAMS-D-18-0075.1.

(33) Moran, M. D.; Dastoor, A.; Morneau, G. Long-Range Transport of Air Pollutants and Regional and Global Air Quality Modelling. In *Air Quality Management*; Springer Netherlands: Dordrecht, 2014; pp 69–98. https://doi.org/10.1007/978-94-007-7557-2_4.

(34) Keuken, M. P.; Moerman, M.; Zandveld, P.; Henzing, J. S.; Hoek, G. Total and Size-Resolved Particle Number and Black Carbon Concentrations in Urban Areas near Schiphol Airport (the Netherlands). *Atmos. Environ.* **2015**, *104*, 132–142. https://doi.org/10.1016/j.atmosenv.2015.01.015.

(35) Keuken, M. P.; Moerman, M.; Zandveld, P.; Henzing, J. S. Total and Size-Resolved Particle Number and Black Carbon Concentrations near an Industrial Area. *Atmos. Environ.* **2015**, *122*, 196–205. https://doi.org/10.1016/j.atmosenv.2015.09.047.

(36) Hudda, N.; Gould, T.; Hartin, K.; Larson, T. V.; Fruin, S. A. Emissions from an International Airport Increase Particle Number Concentrations 4-Fold at 10 Km Downwind. *Environ. Sci. Technol.* **2014**, *48* (12), 6628–6635. https://doi.org/10.1021/es5001566.

(37) Shairsingh, K. K.; Jeong, C. H.; Wang, J. M.; Brook, J. R.; Evans, G. J. Urban Land Use Regression Models: Can Temporal Deconvolution of Traffic Pollution Measurements Extend the Urban LUR to Suburban Areas? *Atmos. Environ.* **2019**, *196*, 143–151. https://doi.org/10.1016/j.atmosenv.2018.10.013.

(38) Morawska, L.; Ristovski, Z.; Jayaratne, E. R.; Keogh, D. U.; Ling, X. Ambient Nano and Ultrafine Particles from Motor Vehicle Emissions: Characteristics, Ambient Processing and Implications on Human Exposure. *Atmospheric Environment*. November 2008, pp 8113–8138. https://doi.org/10.1016/j.atmosenv.2008.07.050.

(39) J, K.; G, H.; R, V. A Nationwide Land Use Regression Model for Ultrafine Particles. *Environ. Epidemiol.* **2019**, *3*, 195. https://doi.org/10.1097/01.ee9.0000607960.60423.0e.

(40) Liu, J.; Li, W.; Wu, J.; Liu, Y. Visualizing the Intercity Correlation of $PM_{2.5}$ time Series in the Beijing-Tianjin-Hebei Region Using Ground-Based Air Quality Monitoring Data. *PLoS One* **2018**, *13* (2), 1–14. https://doi.org/10.1371/journal.pone.0192614.

Supporting Information for

Spatial and Spatiotemporal Variability of Regional Background Ultrafine Particle Concentrations in The Netherlands

Supplement A: Data correction:

Supplement B: Spatial Analysis:

Supplement C: Spatiotemporal Analysis:

Supplement A: Data Correction

Differences between devices were assessed by co-locating all 5 devices for 1-2 days before and after every two-week measurement. All devices were placed in a laboratory setting with all inlets connected to the same air-stream. Concentrations were fluctuated between 0 and 100.000 particles/cm³ throughout the colocation measurement. We corrected each measurement campaign based on calculated median ratios between the reference device and other devices of each co-located comparison. All individual ratios are shown in table A.1 with average medians ranging from 0,99 to 1,19.

Temporal Correction

We used a reference site in the middle of the country (Bunnik) to correct for temporal variation. This reference site had the same restrictions as the sampling locations regarding potential local sources. Reference equipment was mounted into a bike-trailer and was placed at ground level in the back garden of a relatively secluded house.

The difference method we used is as follows:

1. Calculate the overall average UFP concentration on the reference site for the full study period.
2. Calculate for the reference site the difference between the overall average UFP concentration and each of the 15 averages of the measurement periods.
3. Add this difference to the time-corresponding individual site average concentrations on all sampling locations.

Table A.1

Ratios for all individual co-location measure-ments.

↓

Date	DM1:DM5	DM2:DM5	DM3:DM5	DM4:DM5
06/09/2016	1,20	1,23	1,20	0,80
23/09/2016	1,16	1,16	1,07	0,85
26/09/2016	/	1,19	1,15	0,85
13/10/2016	1,18	1,25	1,15	0,93
14/10/2016	1,10	1,23	1,12	0,96
07/11/2016	1,03	1,20	0,99	0,98
25/11/2016	0,94	1,13	1,01	0,88
29/11/2016	1,10	1,22	1,00	1,02
16/12/2016	1,22	1,20	0,99	1,01
21/12/2016	1,24	1,16	1,02	1,02
20/01/2017	1,14	1,22	0,94	1,01
23/01/2017	1,11	1,22	0,96	1,06
09/02/2017	1,12	1,23	0,98	0,98
10/02/2017	0,98	1,23	0,92	0,98
02/03/2017	1,13	1,17	1,04	1,05
23/03/2017	1,16	1,30	1,03	1,11
27/03/2017	1,18	1,27	1,08	1,12
28/03/2017	1,17	1,22	1,07	1,02
13/04/2017	1,19	1,26	1,13	1,05
10/07/2017	0,88	0,88	0,98	0,87
01/08/2017	/	1,11	1,16	1,01
04/09/2017	1,17	1,20	1,47	1,20
03/10/2017	1,08	0,93	1,07	0,98
23/10/2017	1,25	1,18	1,09	1,01
14/11/2017	1,38	1,32	1,09	1,07
Total	1,14	1,19	1,07	0,99

Table A.2

Individual 2-week average concentrations per site

↓

Site	Round 1	Round 2	Round 3	Total	Region
Bedum	5569	3964	/	4767	North
Budel	5914	5257	/	5586	South
Burgerbrug	4800	4966	3637	4468	North
De Zilk	7908	5449	/	6678	West
Donkerbroek	/	4600	3028	3814	North
Grijpskerke	6215	5171	3854	5080	South
Huybergen	7658	7180	5355	6731	South
Lemelerveld	6065	4749	/	5407	East
Mantgum	/	4330	/	4330	North
Maurik	8202	5565	4352	6040	East
Monnickendam	6593	5680	5064	5779	West
Odiliapeel	7723	6751	5252	6575	South
Oosterhesselen	4717	4117	/	4417	North
Rockanje	6477	5480	5121	5693	West
s Gravenmoer	7185	5530	5707	6141	West
Schimmert	/	5228	5390	5309	South
Stolwijk	6561	5388	/	5975	West
Swifterband	/	3920	3899	3909	East
Voorthuizen	7584	7599	6056	7080	East
Zelhem	6153	5388	4300	5280	East

Concentrations in particles/cm³

Supplement B: Spatial Analysis

Geostatistics

Geostatistics can be applied to create maps of a spatially sampled property, a so-called regionalized variable (Cressie, 1991). The most widely used geostatistical technique is kriging, which provides estimates with minimum error variance of the regionalized variable at unsampled locations (Davis, 2002). In order to apply kriging, first the spatial correlation between observations of the regionalized variable needs to be quantified. The concept used in geostatistics is that observations of a regionalized variable at locations that are close to each other are more similar than observations that are further apart. This spatial correlation can be quantified with the semi-variogram, which is a model of the semi-variance between spatial observations at a certain distance apart (the lag distance). In a 2-dimensional plane the semi-variance between pairs of observations (z_i, z_{i+h}) which are separated by a distance falling in a distance class h is calculated as (Davis, 2002):

$$\gamma_h = \frac{1}{2}\frac{\sum_{i=1}^{Nh}\left(z_i - z_{i+h}\right)^2}{Nh} \qquad \text{S1}$$

where y_h is the semi-variance at lag distance class h, and Nh is the total number of observation pairs in lag distance class h. By fitting an equation through the calculated semi-variogram a model of the semi-variance is obtained. This model is characterized by three parameters: the nugget, the range and the sill. The nugget is the semi-variance at distance zero. In theory this semi-variance is zero, but due to measurement errors or spatial variation at small lag distances that are not included in the observations there can be a non-zero semi-variance. The range of the semi-variogram is the spatial distance over which spatial correlation exists. Beyond the range any two point are no longer related to each other. The sill is the semi-variance that is reached at the range. It is approximately equal to the variance of all observations.

Webster and Oliver (1992) recommended that semi-variograms should be computed using at least 50 observations, but preferably 100 or more observations of the regionalized variable should be used. In practice, often less observations of a regionalized variable are available due to varying problems, such as lack of instruments, time and financial resources. To overcome this problem of an insufficient number of observations, Sterk and Stein (1997) developed a space-time procedure to compute stable semi-variograms. In their procedure multiple events with observations of the regionalized variable taken at different moments in time are used to calculate a standardized semi-variogram. This standardized semi-variogram is subsequently converted to an event-based semi-variogram, by re-calculating the nugget and sill, which are proportional to the variance of the observations for the event. The range of the semi-variograms is assumed equal for all events, assuming that the spatial correlation structure is unchanged. The Sterk and Stein (1997) space-time procedure was used in this study to calculate separate semi-variograms for each of the three UFP measurement times.

In the first step the average concentrations per location and measurement period were standardized by withdrawing the mean of all 49 observations from each individual observation, and by dividing with the standard deviation of all 49 observations. The semi-variogram was calculated using equation S1, but to avoid that calculations between the observations at the three measurement periods were used, the coordinates were shifted 1000 km per time interval. The maximum distance of pairs of observations was set at 500 km, which guaranteed that only pairs of observations within a certain time period were used. A spherical model was fitted through the computed standardized semi-variogram values (Figure B1). The spherical model is defined as (Davis, 2002):

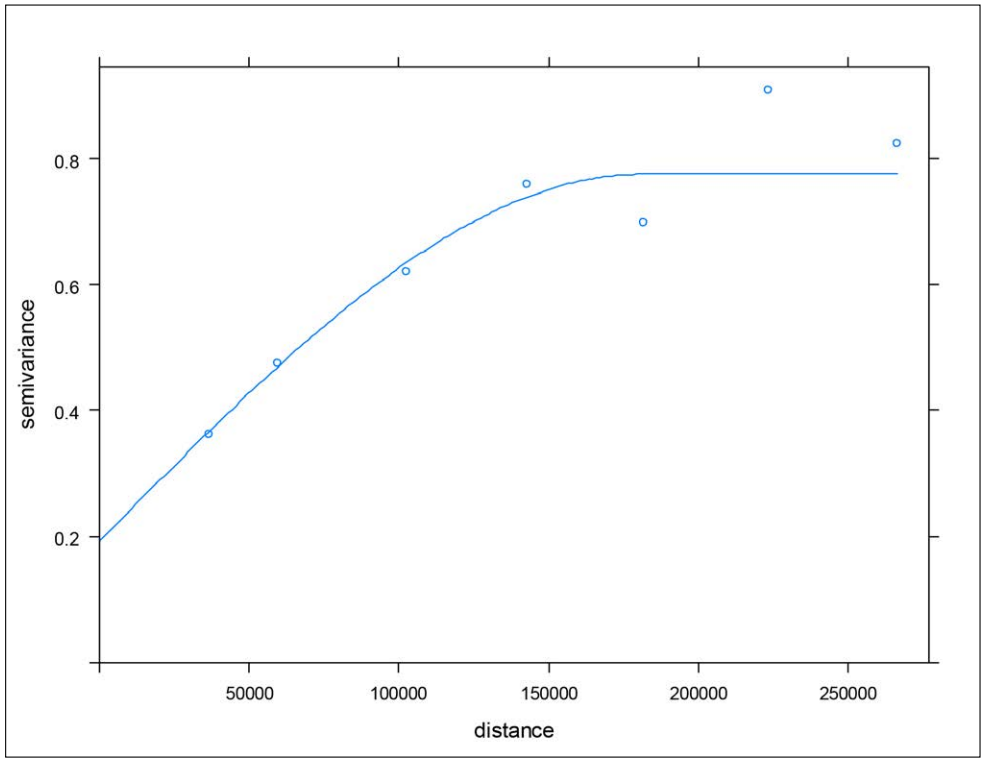
$$\begin{aligned} \gamma_h &= \gamma_0 + (\sigma_0^2 - \gamma_0)\left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right) && \text{for } h < a \\ \gamma_h &= \sigma_0^2 && \text{for } h \geq a \end{aligned} \qquad \text{S2}$$

where γ_0 is the nugget, σ_0^2 is the sill and a is the range of the semi-variogram.

Figure B.1

Fitted semi-variogram model through calculated semi-variance values of standardized UFP measurements in the Netherlands.

↓



The fitted spherical model has a range of 180.9 km, which indicates the presence of spatial correlation up until 181 km. The nugget and sill values are respectively 0.193 and 0.582. The standardized semi-variogram was converted into three semi-variograms, one for each measurement time. This was done by multiplying the nugget and sill of the standardized semi-variogram with the variance of the UFP values per measurement time, while the range was kept constant. The parameters of the three spherical semi-variogram models are provided in Table B.1.

Table B.1

Fitted spherical semi-variogram models for three UFP measurement periods in the Netherlands.

↓

Observation time	**Spherical semi-variogram**		
	Nugget (particles cm^{-3})²	Sill (particles cm^{-3})²	Range (m)
Round 1	217606	657807	180893
Round 2	184765	558532	180893
Round 3	151692	458553	180893

The three semi-variogram models and the original UFP concentrations of every measurement period were used to make three separate kriging maps, applying ordinary kriging (Davis, 2002):

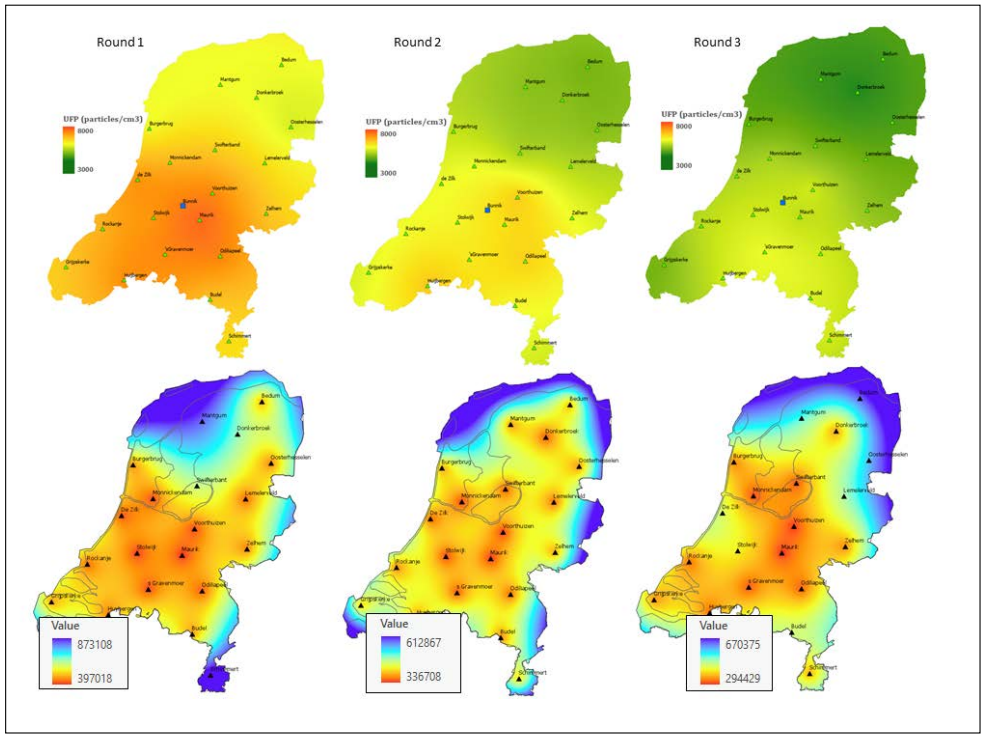
$$\hat{Z}(x_o) = m + \sum_{i=1}^{k} \lambda_i [Z(x_i) - m] \qquad \text{S3}$$

where $\hat{Z}(x_o)$ is the kriging estimate at position x_o, m is the mean of the observations, λ_i are the kriging weights, and $Z(x_i)$ the UFP observations. The kriging weights are calculated using the semi-variogram model, and determine the weight of each observation for the kriging estimate. A raster was used with a cell size of 500 x 500m. The three produced kriging maps (Figure B.2) were then averaged in ArcGIS pro, to create one final mean kriging map of the Netherlands. Kriging error maps (i.e. the variance of the kriging estimates) for each measurement period were made to estimate the uncertainty of the kriging prediction over the area (Figure B.2). In sensitivity analysis we performed a leave-one-out-cross-validation to further analyse potential error in the map. While deletion of a site would affect the standardisation procedure and development of the variogram, we assumed these to be unchanged. So, we created 20 kriging maps, each based on 19 sites and the same variogram. Predictions made for the remaining site was then compared to the same site in the kriging map with all 20 sites included. Results are shown in figure B.3. All geostatistical analyses were done using the gstat package in R.

Figure B.2

Maps of UFP regional background concentrations in the Netherlands produced by kriging.

↓



Dates of measurement periods are June 8 – Nov 23 2016 (Round 1), Nov 30 2016 – Mar 22 2017 (Round 2) and Mar 29 - Nov 8 2017 (Round 3). Bottom: Maps of the kriging variance for the corresponding periods, providing a measure of the uncertainty of the kriging estimates in the top row figures. Values are the root of the standard error.

Figure B.3

Scatterplot of UFP predictions generated by leave-one-out-cross-validation and kriged values with the full data set.

↓

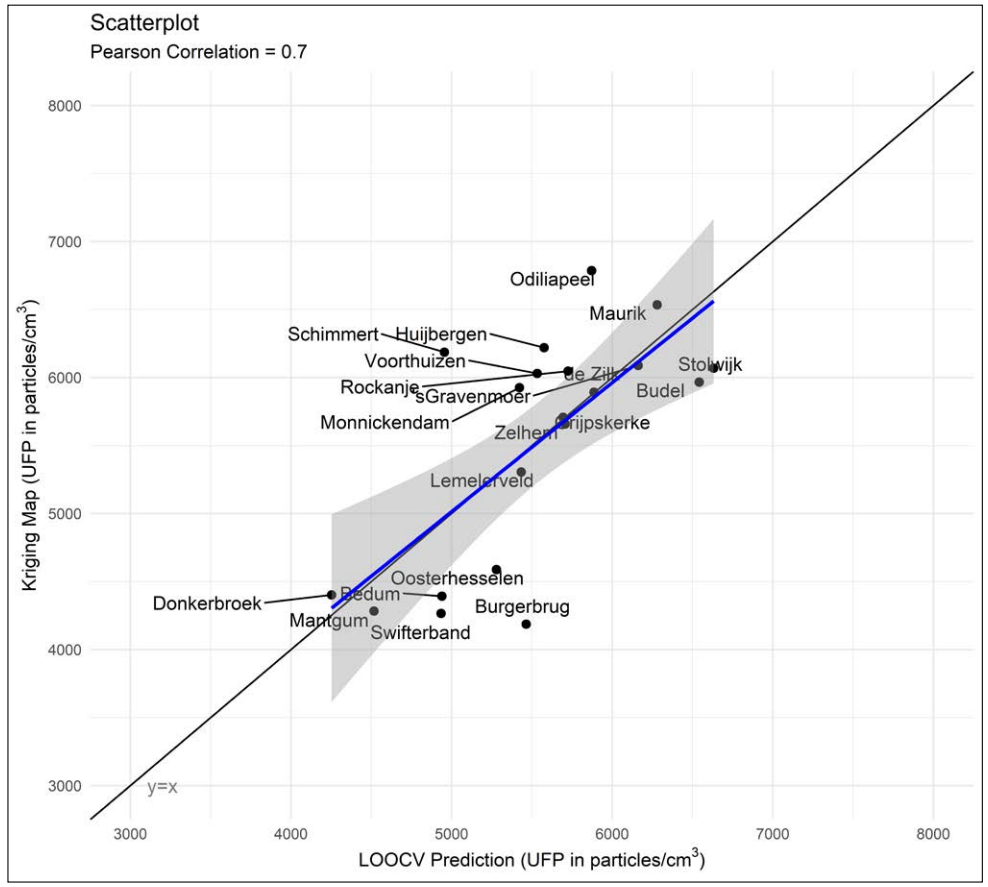


Figure B.4

Scatterplot of UFP predictions generated by leave-one-out-cross-validation and measured concentrations.

↓

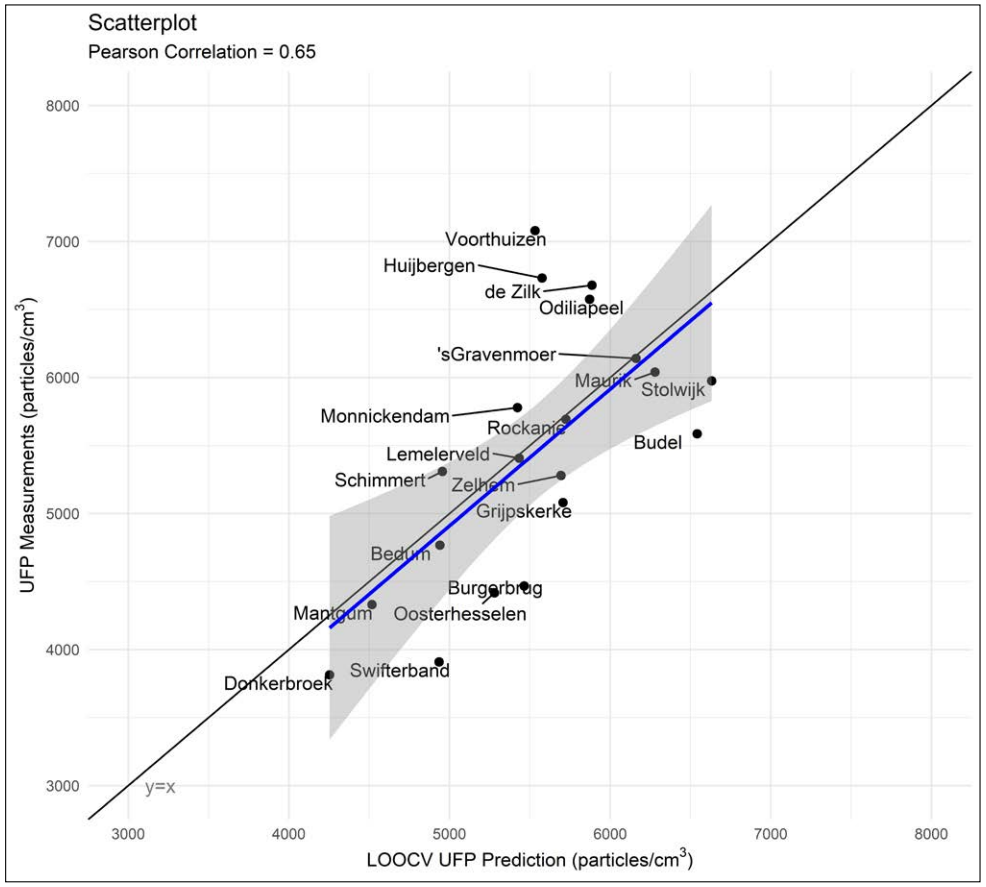


Figure B.5

Scatterplot of UFP concentrations generated by kriging with the measured concentrations.

↓

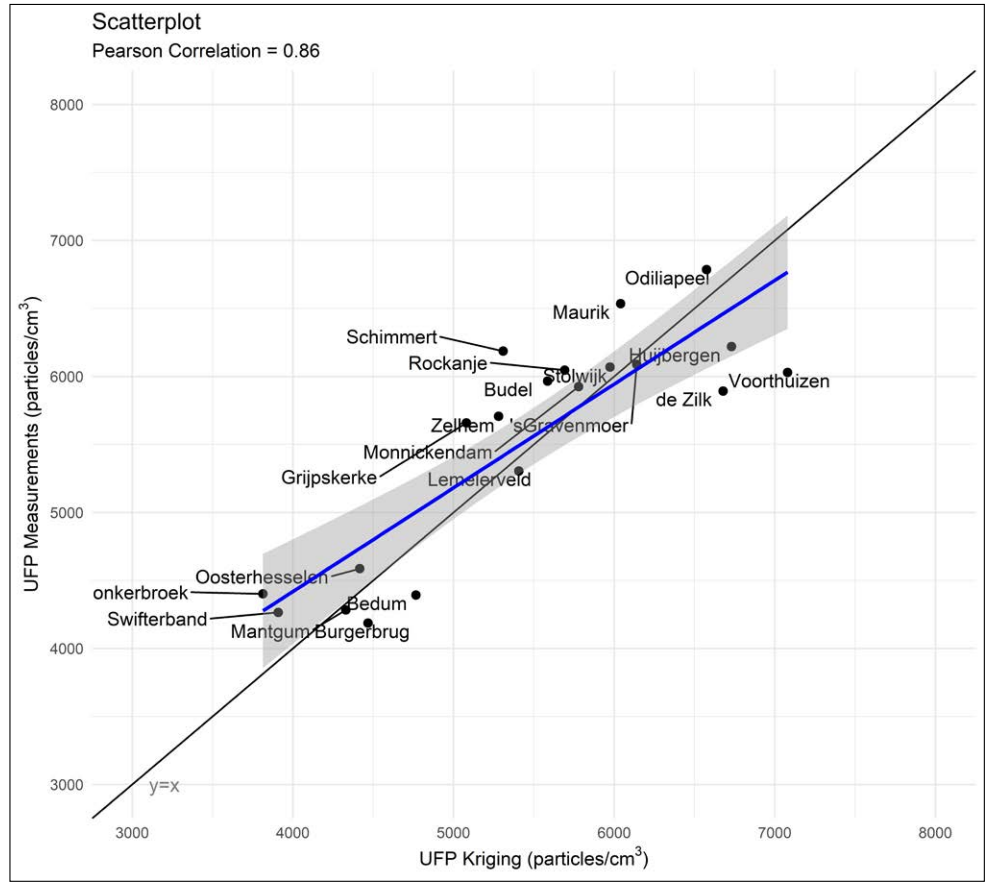


Table C.1

Distance, Pearson and Spearman correlations, maximum cross-correlations and proportion of variance explained (PVE), between the sampling locations and the reference site (Bunnik).

↓

Site	Distance	Pearson	Spearman	2hr Cross-Correlation	Time lag (Hours)	24hr Cross-Correlation	PVE 2hr	PVE 24hr
Bedum	165	0,59	0,50	0,59	0	0,39	0,33	0,15
	165	0,58	0,57	0,58	0	0,52	-0,13	0,02
Budel	100	0,42	0,38	0,44	1	0,26	0,11	0,05
	100	0,54	0,64	0,55	1	0,56	0,08	0,27
Burgerbrug	77	0,43	0,53	0,45	1	0,40	-0,25	0,11
	77	0,57	0,62	0,58	0,5	0,66	0,27	0,41
	77	0,36	0,38	0,46	2[v]	0,58	-0,05	0,15
De Zilk	46	-0,05	-0,05	-0,24	3[v]	-0,24	-0,41	-0,28
	46	0,44	0,44	0,45	-0,5	0,33	-0,06	0,00
Donkerbroek	125	0,63	0,68	0,67	1,5	0,67	-0,06	0,37
	125	0,61	0,62	0,61	-0,5	0,65	-0,03	0,39
Grijpskerke	125	0,28	0,27	0,33	-1,5	0,38	-0,16	-0,04
	125	0,48	0,51	0,49	0,5	0,57	-0,31	0,22
	125	0,23	0,34	0,25	-1,5	0,23	-0,47	-0,02
Huijbergen	92	0,60	0,66	0,60	0	0,56	0,36	0,27
	92	0,53	0,55	0,54	0,5	0,48	0,20	0,17
	92	0,45	0,50	0,45	0,5	0,53	-0,22	0,12
Lemelerveld	90	0,73	0,70	0,73	0	0,54	0,47	0,29
	90	0,67	0,63	0,67	0	0,63	0,31	0,34
Mantgum	120	0,63	0,65	0,63	0	0,65	0,22	0,36
Maurik	25	0,35	0,87	0,38	-1,5	0,17	0,12	0,03
	25	0,90	0,90	0,90	0	0,75	0,78	0,55
	25	0,70	0,71	0,70	0	0,50	0,36	0,24
Monnickendam	40	0,57	0,35	0,58	-1	0,38	0,32	0,13
	40	0,53	0,66	0,53	0	0,51	0,20	0,23
	40	0,49	0,51	0,49	0	0,49	0,07	0,19
Odiliapeel	65	0,73	0,74	0,74	-0,5	0,61	0,49	0,37
	65	0,66	0,65	0,66	0	0,60	0,38	0,36
	65	0,53	0,53	0,55	1	0,45	0,12	0,19
Oosterhesselen	130	0,44	0,47	0,48	-1	0,43	-0,53	-0,21
	130	0,44	0,42	0,47	1,5	0,33	-0,30	-0,03
Rockanje	78	0,32	0,42	0,36	-2,5	0,36	-0,23	0,12
	78	0,45	0,59	0,46	0,5	0,55	0,15	0,24
	78	0,34	0,35	0,34	0	0,25	-0,57	-0,20
s'Gravenmoer	52	0,70	0,71	0,70	0	0,52	0,47	0,23
	52	0,63	0,65	0,63	0	0,59	0,09	0,33
	52	0,63	0,65	0,63	0	0,30	0,23	0,03
Schimmert	142	0,44	0,48	0,45	-1	0,43	-0,14	0,05
	142	0,42	0,41	0,42	0,5	0,41	-0,32	0,04
Stolwijk	28	0,78	0,83	0,78	0	0,56	0,59	0,31
	28	0,74	0,74	0,74	-0,5	0,59	0,29	0,24
Swifterband	62	0,71	0,67	0,72	0,5	0,69	0,05	0,40
	62	0,29	0,39	0,30	1	0,41	-0,60	-0,10
Voorthuizen	32	0,66	0,76	0,67	0,5	0,62	0,42	0,29
	32	0,60	0,66	0,60	0	0,58	0,36	0,28
	32	0,59	0,57	0,59	0	0,59	0,31	0,30
Zelhem	84	0,44	0,50	0,50	2	0,26	0,12	0,04
	84	0,72	0,71	0,72	0	0,67	0,21	0,41
	84	0,62	0,63	0,65	1	0,61	0,15	0,36

Table C.2

Influence of distance between sites on spatiotemporal correlation, with log-transformed UFP concentrations.

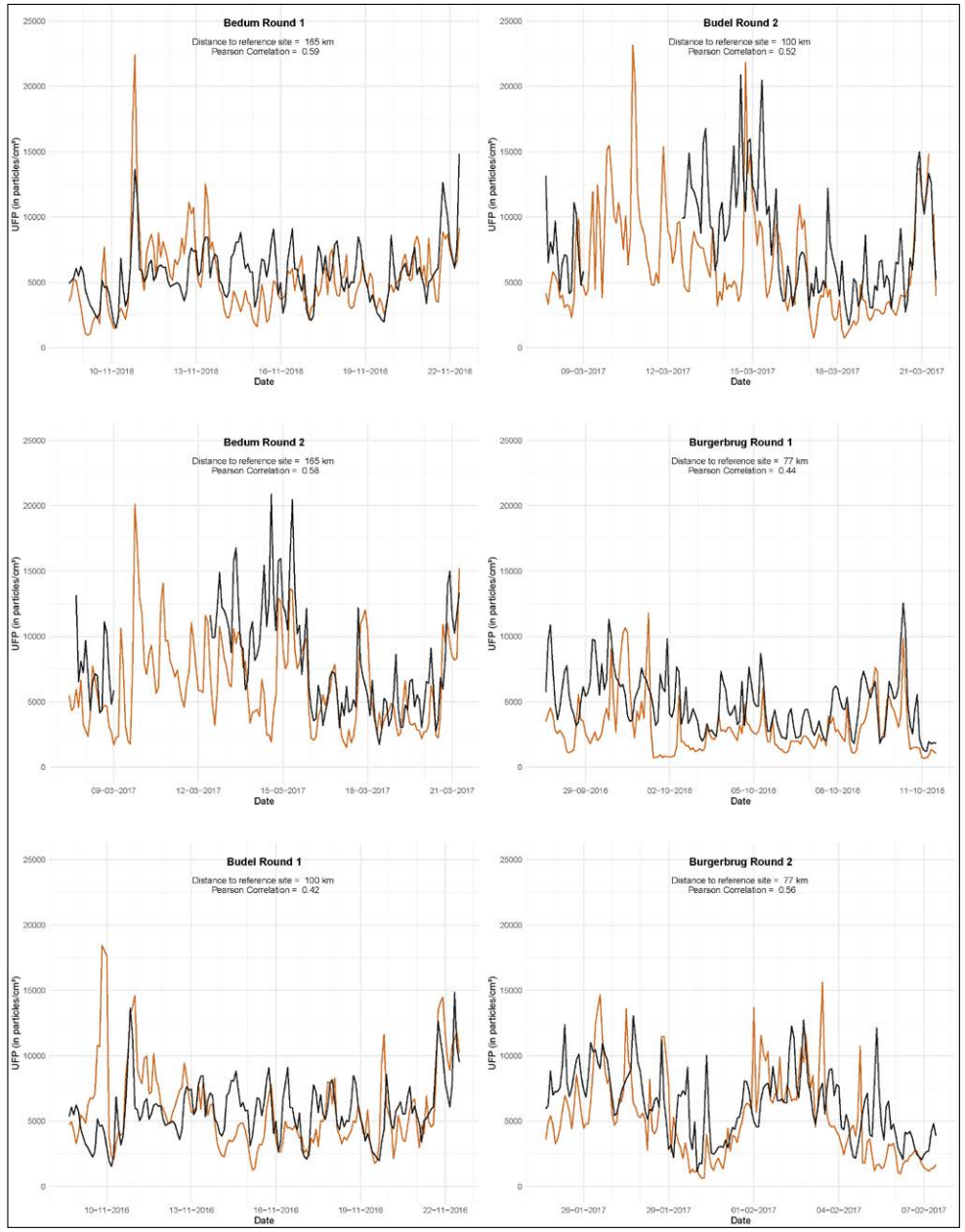
↓

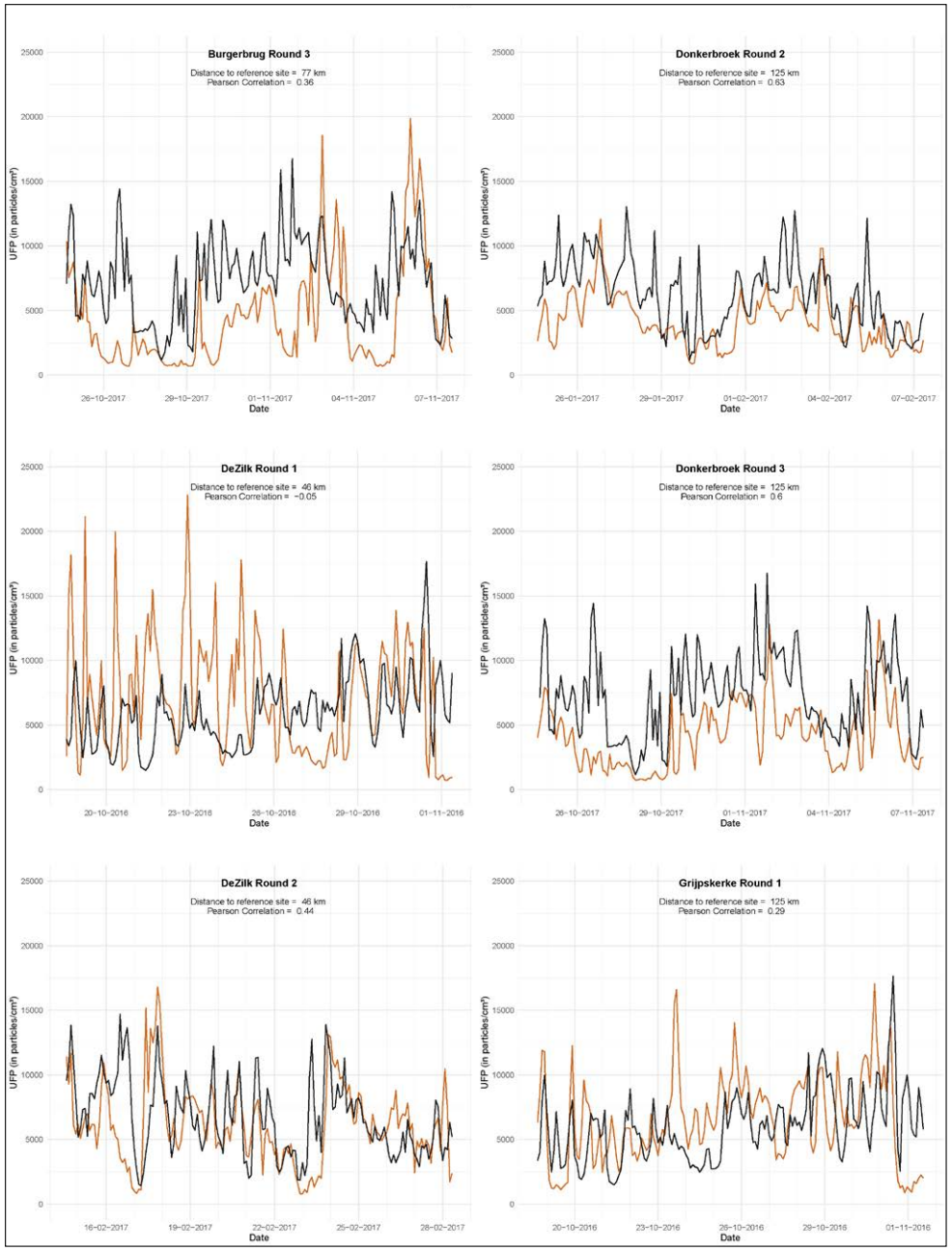
Distance between sites	Median Pearson correlation 2h	Median cross correlation 2h	Median cross correlation 24h	Median PVE 2h	Median PVE 24h
< 50 km	0,69	0,69	0,60	0,47	0,35
50 - 80km	0,61	0,62	0,53	0,24	0,24
> 80km	0,55	0,56	0,52	0,18	0,23

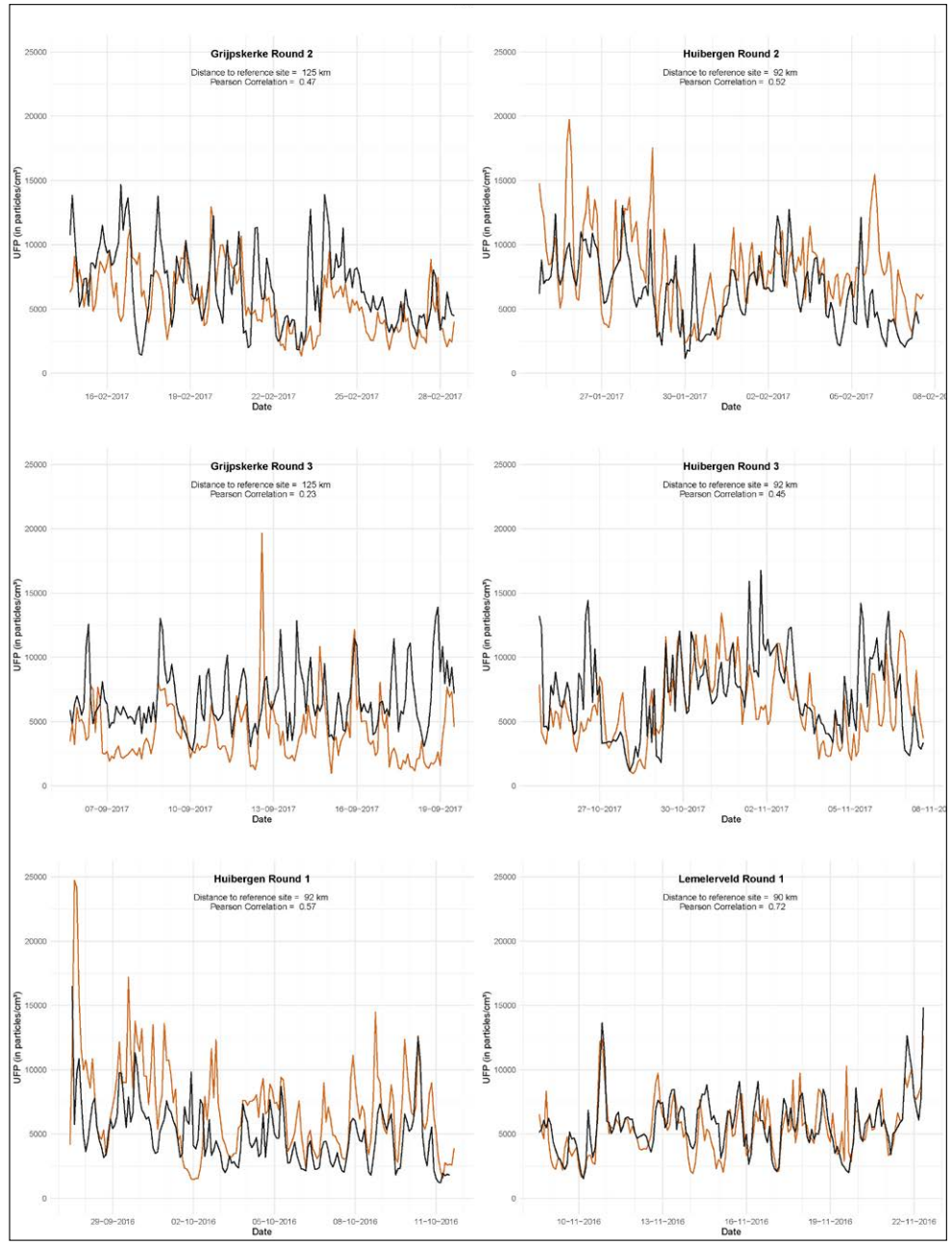
Figure C.1

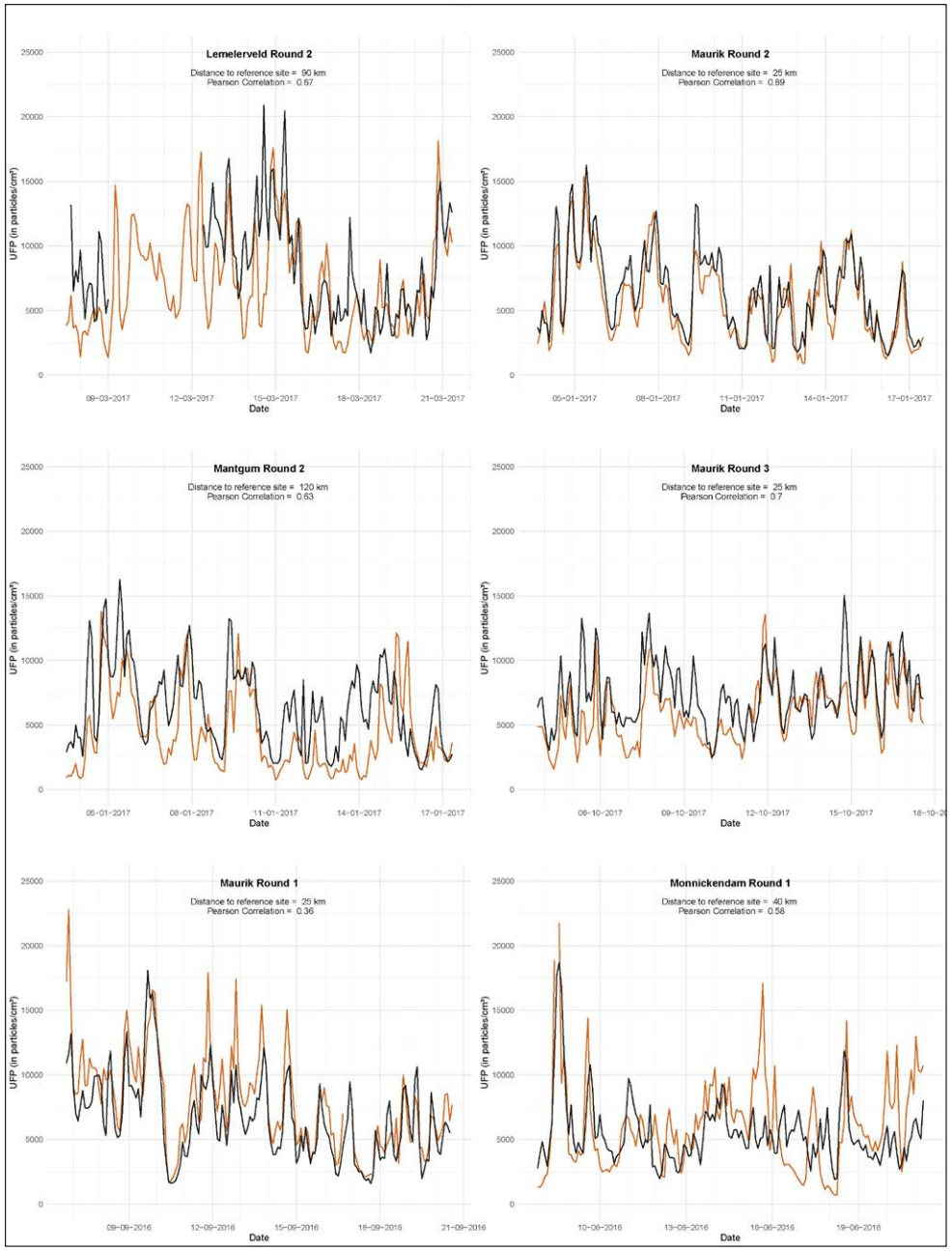
Temporal variation of 2-hourly averaged UFP concentrations (particles/cm³) at the reference site (black) and the sampling locations (brown).

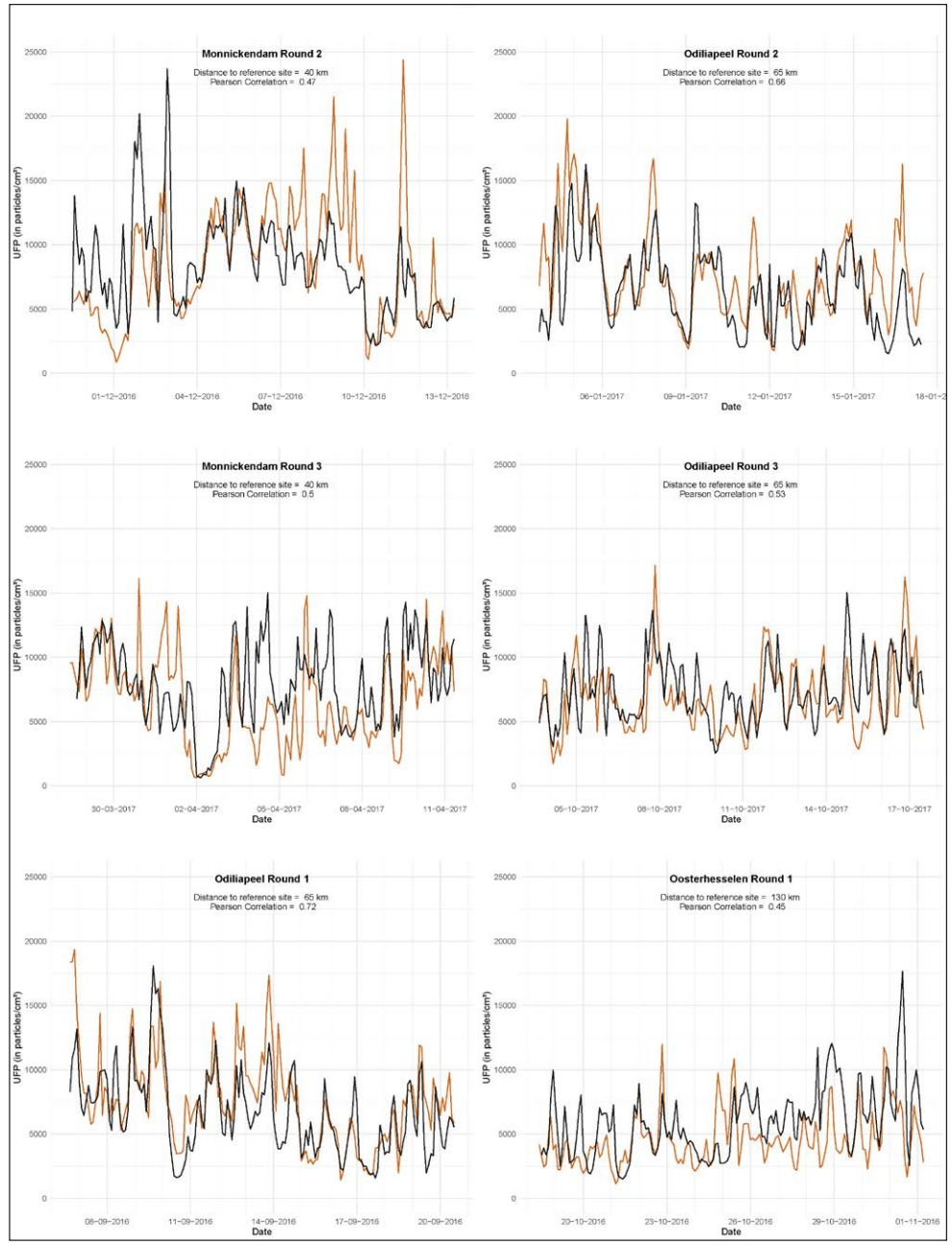
↓

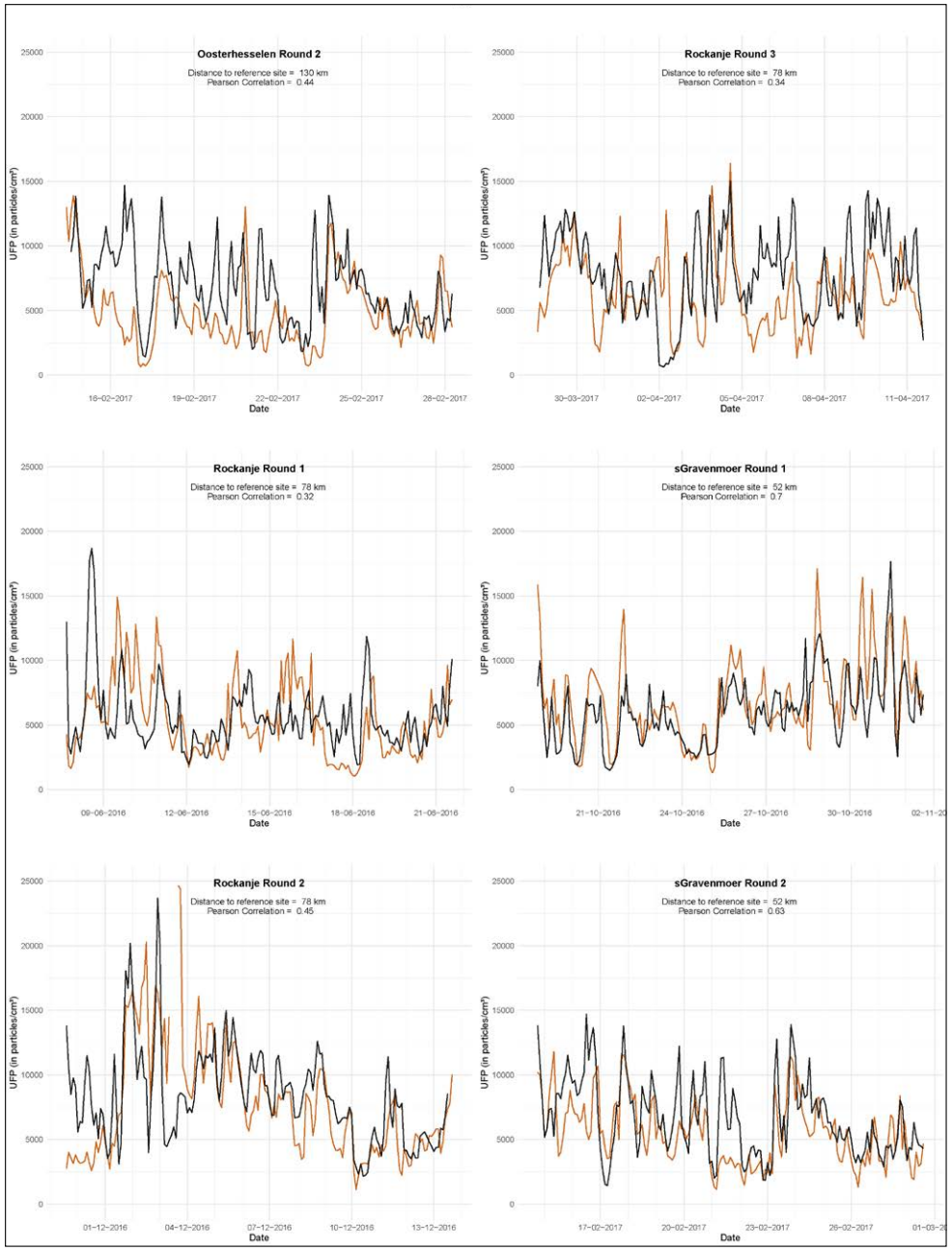


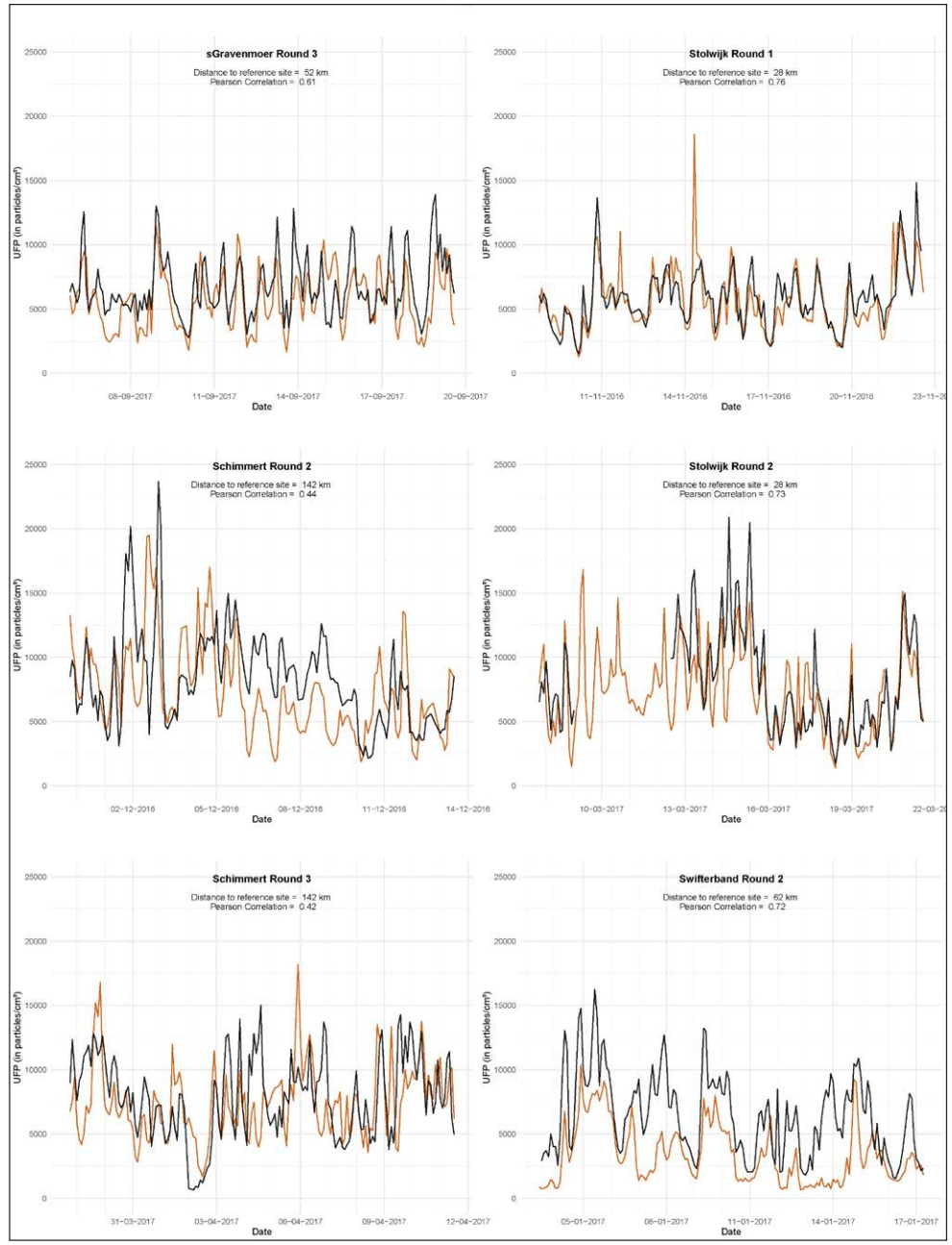

Burgerbrug Round 3
Distance to reference site = 77 km
Pearson Correlation = 0.36
Donkerbroek Round 2
Distance to reference site = 125 km
Pearson Correlation = 0.63
DeZilk Round 1
Distance to reference site = 46 km
Pearson Correlation = −0.05
Donkerbroek Round 3
Distance to reference site = 125 km
Pearson Correlation = 0.6
DeZilk Round 2
Distance to reference site = 46 km
Pearson Correlation = 0.44
Grijpskerke Round 1
Distance to reference site = 125 km
Pearson Correlation = 0.29
UFP (in particles/cm³)
Date

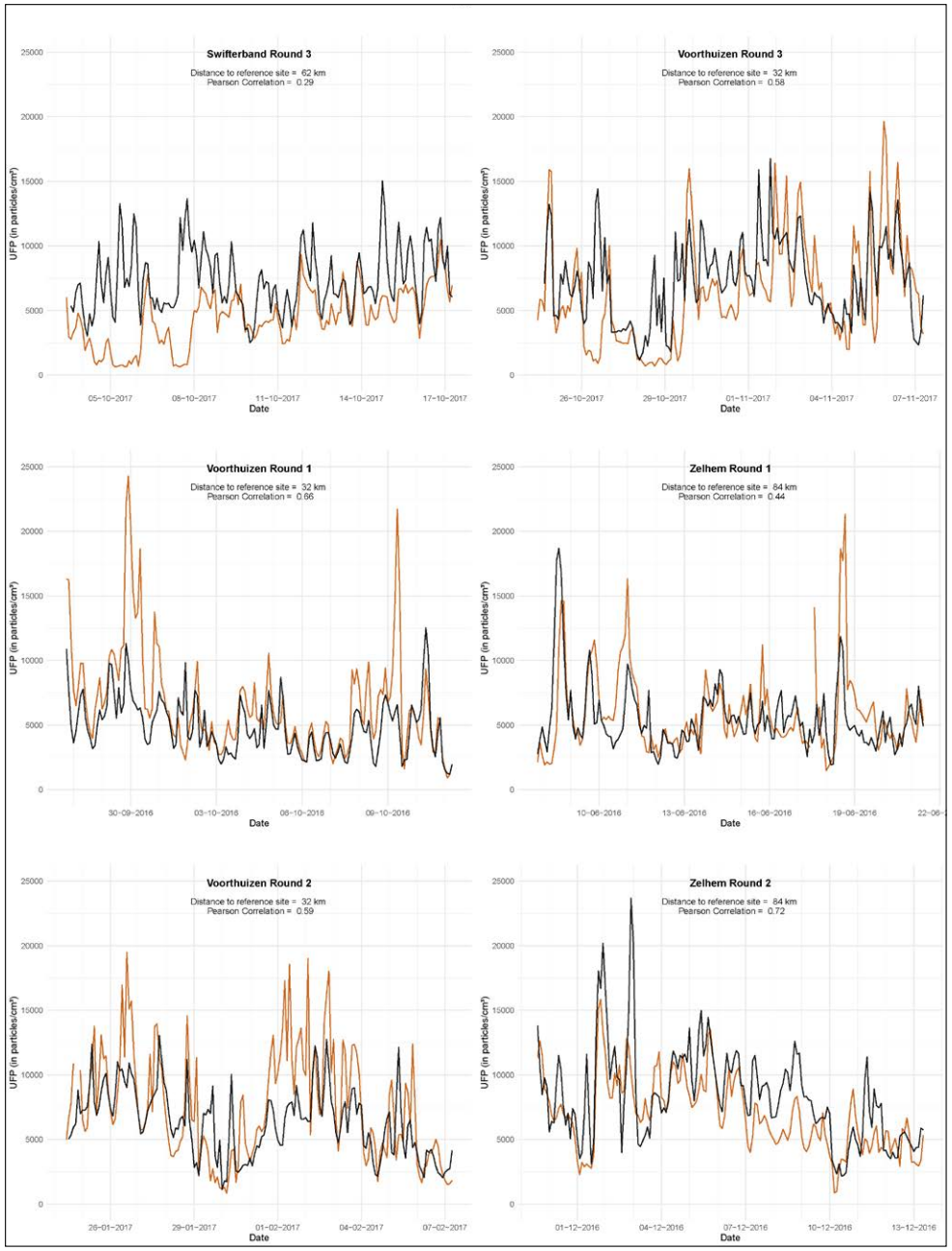

Grijpskerke Round 2
Distance to reference site = 125 km
Pearson Correlation = 0.47
Huibergen Round 2
Distance to reference site = 92 km
Pearson Correlation = 0.52
Grijpskerke Round 3
Distance to reference site = 125 km
Pearson Correlation = 0.23
Huibergen Round 3
Distance to reference site = 92 km
Pearson Correlation = 0.45
Huibergen Round 1
Distance to reference site = 92 km
Pearson Correlation = 0.57
Lemelerveld Round 1
Distance to reference site = 90 km
Pearson Correlation = 0.72
UFP (in particles/cm³)
Date


Lemelerveld Round 2
Distance to reference site = 90 km
Pearson Correlation = 0.67
Maurik Round 2
Distance to reference site = 25 km
Pearson Correlation = 0.89
Mantgum Round 2
Distance to reference site = 120 km
Pearson Correlation = 0.63
Maurik Round 3
Distance to reference site = 25 km
Pearson Correlation = 0.7
Maurik Round 1
Distance to reference site = 25 km
Pearson Correlation = 0.36
Monnickendam Round 1
Distance to reference site = 40 km
Pearson Correlation = 0.58
UFP (in particles/cm³)
Date


Monnickendam Round 2
Distance to reference site = 40 km
Pearson Correlation = 0.47
Odiliapeel Round 2
Distance to reference site = 65 km
Pearson Correlation = 0.66
Monnickendam Round 3
Distance to reference site = 40 km
Pearson Correlation = 0.5
Odiliapeel Round 3
Distance to reference site = 65 km
Pearson Correlation = 0.53
Odiliapeel Round 1
Distance to reference site = 65 km
Pearson Correlation = 0.72
Oosterhesselen Round 1
Distance to reference site = 130 km
Pearson Correlation = 0.45
UFP (in particles/cm³)
Date


Oosterhesselen Round 2
Distance to reference site = 130 km
Pearson Correlation = 0.44
Rockanje Round 3
Distance to reference site = 78 km
Pearson Correlation = 0.34
Rockanje Round 1
Distance to reference site = 78 km
Pearson Correlation = 0.32
sGravenmoer Round 1
Distance to reference site = 52 km
Pearson Correlation = 0.7
Rockanje Round 2
Distance to reference site = 78 km
Pearson Correlation = 0.45
sGravenmoer Round 2
Distance to reference site = 52 km
Pearson Correlation = 0.63
UFP (in particles/cm³)
Date


sGravenmoer Round 3
Distance to reference site = 52 km
Pearson Correlation = 0.61
Stolwijk Round 1
Distance to reference site = 28 km
Pearson Correlation = 0.76
Schimmert Round 2
Distance to reference site = 142 km
Pearson Correlation = 0.44
Stolwijk Round 2
Distance to reference site = 28 km
Pearson Correlation = 0.73
Schimmert Round 3
Distance to reference site = 142 km
Pearson Correlation = 0.42
Swifterband Round 2
Distance to reference site = 62 km
Pearson Correlation = 0.72
UFP (in particles/cm³)
Date



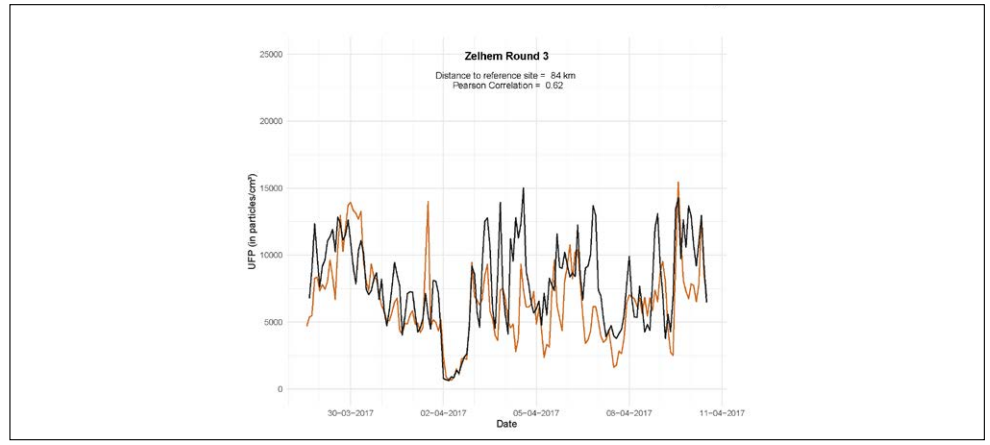


Figure C.2

Average diurnal pattern of UFP for all sites. Reference site in black.

↓

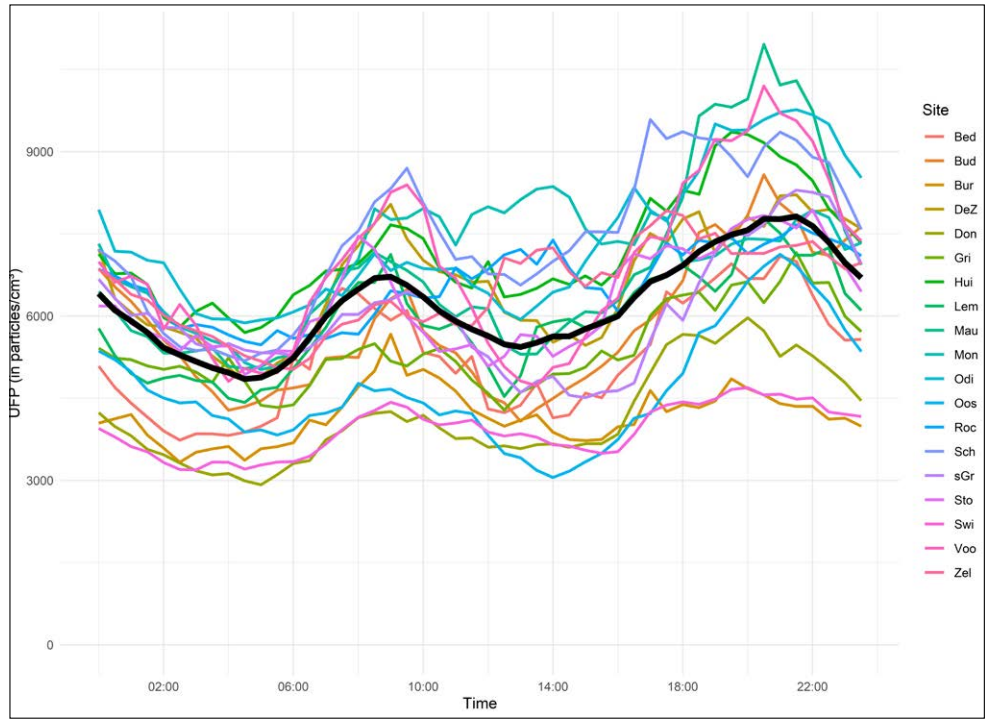


Table C.3

Difference in UFP (in particles/cm³) per site between cold and warm season.

↓

Site	Cold Season	Warm Season
Bed	5.524	2.920
Bud	5.840	NA
Bur	4.262	2.852
DeZ	6.619	NA
Don	4.168	NA
Gri	6.044	4.122
Hui	6.984	10.164
Lem	5.925	NA
Mau	5.807	8.924
Mon	8.562	6.060
Odi	7.216	7.650
Oos	4.669	NA
Roc	8.035	5.664
Sch	7.441	7.351
sGr	6.222	5.650
Sto	6.186	7.091
Swi	3.916	NA
Voo	6.792	9.357
Zel	7.328	5.930

References:

Cressie, N.A.C. 1991. Statistics for Spatial Data. John Wiley & Sons, New York

Davis, J.C. 2002. Statistics and Data Analysis in Geology, 3rd Edition. John Wiley & Sons, New York.

Sterk, G. and A. Stein. 1997. Mapping wind-blown mass transport by modeling variability in space and time. Soil Science Society of America Journal 61: 232-239.

Webster R. and M.A. Oliver 1992. Sample adequately to estimate variograms of soil properties. Journal of Soil Science 43: 177-192.

Modelling Nationwide Spatial Variation of Ultrafine Particles based on Mobile Monitoring

Jules Kerckhoffs
Gerard Hoek
Ulrike Gehring
Roel Vermeulen

Abstract

Background

Large nation- and region-wide epidemiological studies have provided important insights into the health effects of long-term exposure to outdoor air pollution. Evidence from these studies for the long-term effects of ultrafine particles (UFP), however is lacking. Reason for this is the shortage of empirical UFP land use regression models spanning large geographical areas including cities with varying topographies, peri-urban and rural areas. The aim of this paper is to combine targeted mobile monitoring and long-term regional background monitoring to develop national UFP models.

Method

We used an electric car to monitor UFP concentrations in selected cities and towns across the Netherlands over a 14-month period in 2016-2017. Routes were monitored 3 times and concentrations were averaged per road segment. In addition, we used kriging maps based on regional background monitoring (20 sites; 3x2 weeks) over the same period to assess annual average regional background concentrations. All road segments were used to model spatial variation of UFP with three different land-use (regression) approaches: supervised stepwise regression, LASSO and random forest. For each approach, we also tested a deconvolution method, which segregates the average concentration at each road segment into a local and background signal. Model performance was evaluated with short-term (400 sites across the Netherlands; 3x30 minutes) and external longer-term measurements (42 sites in two major cities; 3x24 hours). We also compared predictions of all six models at 1000 random addresses spread over the country.

Results

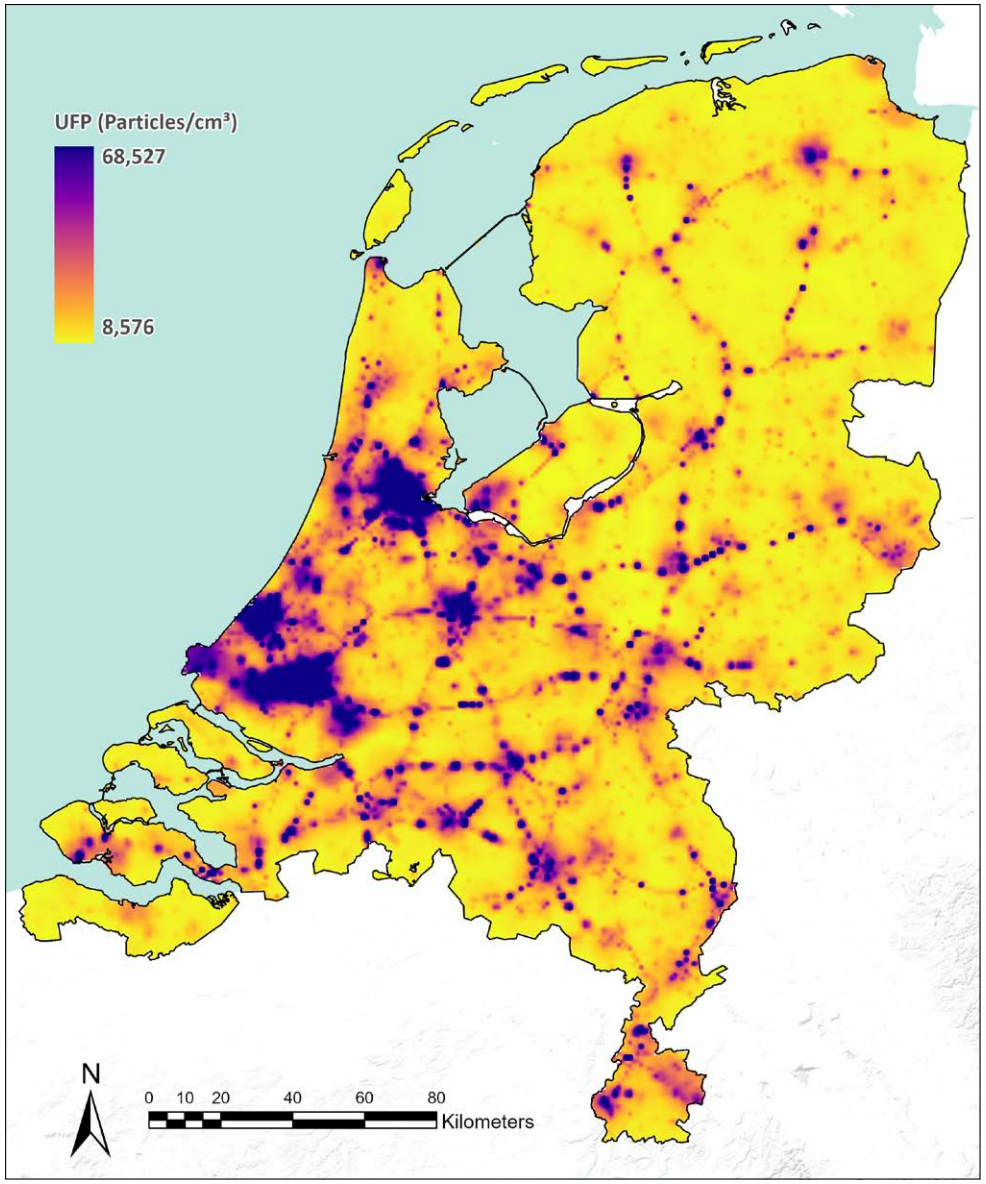
We found similar predictive performance for the six models, with validation R^2 values from 0.25 to 0.35 for short-term measurements and 0.52 to 0.60 for longer-term external measurements. Models with and without deconvolution had similar predictive performance. All models based on the deconvolution method included a regional background kriging map as important predictor. Correlations between predictions at random addresses were high with Pearson correlations from 0.84 to 0.99. Models overestimated exposure at the short-term and long-term sites by about 20-30% in all cases, with small differences between regions and road types.

Conclusion

We developed robust nation-wide models for long-term UFP exposure combining mobile monitoring with long-term regional background monitoring. Minor differences in predictive performance between different algorithms were found, but the deconvolution approach is considered more physically realistic. The models will be applied in Dutch nation-wide health studies.

Keywords

Ultrafine Particles, National LUR model



① Introduction

Most studies of the health effects of ultrafine particles (UFP) so far focused on short-term exposures in a relatively small area. Only a few studies looked at the long-term effects of UFP exposure [1,2] and only one study looked at long-term exposure to UFP over a wide area [3]. The Ostro study was conducted among participants across California, USA, while the former studies were restricted to a few cities. Nation-wide epidemiological studies including administrative cohort studies have been very informative to assess health effects of air pollutants such as $PM_{2.5}$ and NO_2, because of the large size of the population, avoidance of selection bias, and the increased contrast in exposure. These studies have become possible because of national exposure models, typically based on nation-wide routine monitoring and modelling, incorporating variation at the regional background, urban background, and local scale. More specifically, the inclusion of regional background concentrations in LUR models advanced exposure assessment and health-effect studies substantially by increasing exposure contrast [4]. This approach is not directly possible for UFP because of the lack of nationwide routine monitoring. Ongoing standardization work by CEN of particle counters may stimulate increased particle number monitoring in networks [5].

In the past decade, exposure assessment of UFP has been revolutionized using mobile platforms. Together with advancements in air monitoring instrumentation, such as higher time resolution and greater portability, these platforms can capture the high variability of UFP in space and time. To assess the spatial variation of UFP, studies used cars [6–10], bikes [11,12], backpacks [13] and combinations thereof [14] to create land use regression (LUR) models. These campaigns were generally restricted to a single city's boundaries, limiting their ability to be used in wide-scale epidemiological studies including peri-urban and rural areas [15]. We recently showed similarities in UFP models between cities in several European countries [16,17]. However, these campaigns focused only on cities and do not allow the inclusion of peri-urban and rural populations in health evaluations of UFP.

Researchers have used different algorithms to develop air pollution prediction algorithms, including stepwise linear regression, regularization methods and more recently machine learning methods [18–20]. However, little is known about the relative performance of these algorithms. Furthermore, these models did not attempt to separate determinants at different spatial scales. Especially for larger study areas, models have been developed separately for different spatial scales such as the regional and urban background and local scale [4], in order to better characterize spatial variation for the different spatial scales, which can be important in

the development of predictive exposure maps. Recently, a deconvolution approach was proposed to separate local contributions from urban background contributions in mobile monitoring data [21].

Here, we aimed to develop and assess the performance of a national UFP model by combining targeted mobile monitoring in multiple cities and regional background monitoring. A second aim was to assess differences in performance of three different prediction algorithms with and without deconvolution to develop spatial models.

② Methods

2.1 Study Design

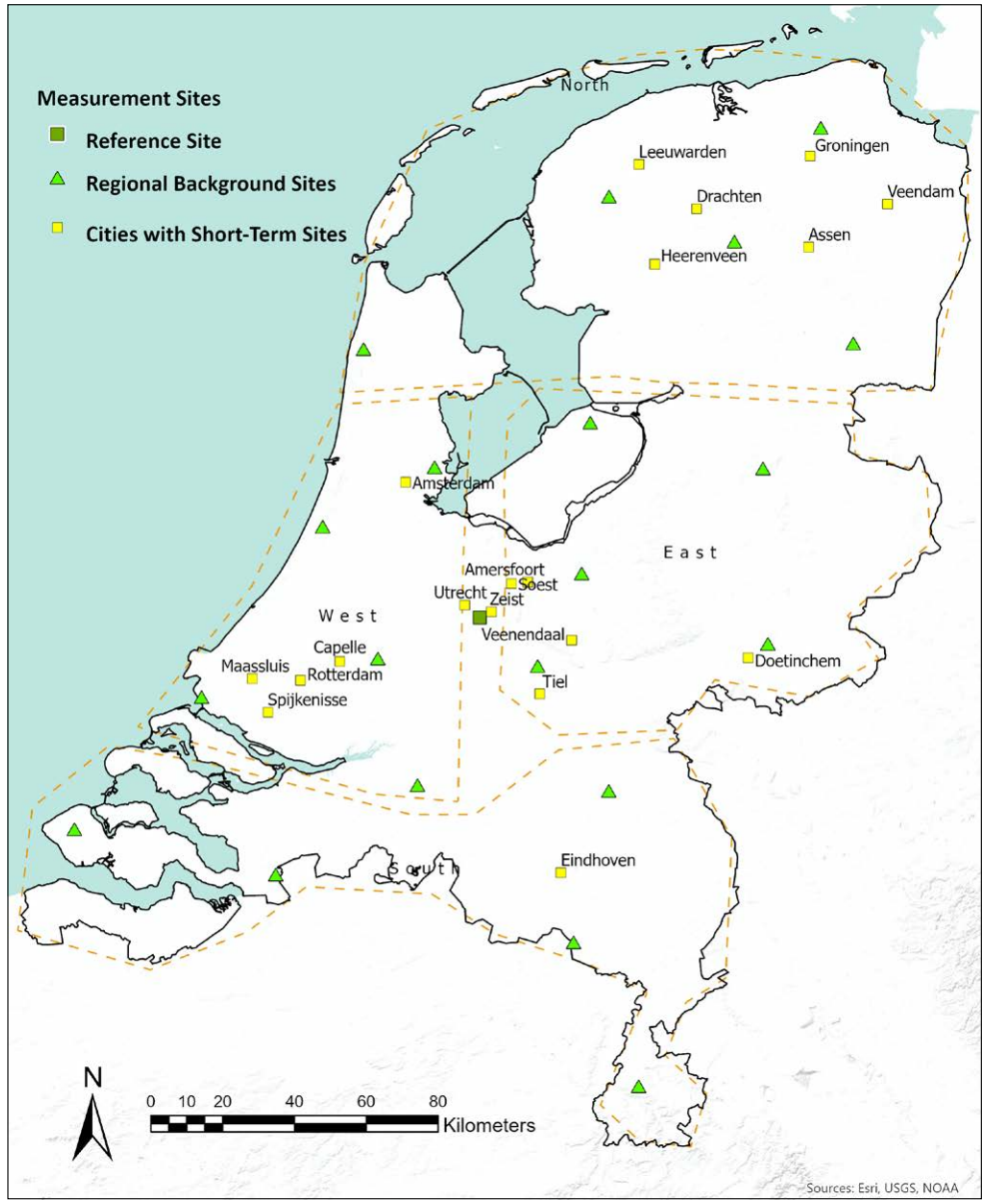
This study combines nation-wide regional background measurements with short-term and true mobile measurements from a mobile platform. The regional background component has been described previously [22]. Briefly, we selected 20 regional background sites across the country (green triangles in figure 1) and measured each site three times for periods of 14 days with a stationary setup. Individual regional background measurements ranged from 3,028 to 8,202 particles/cm³. We used kriging to estimate annual average regional background concentrations across the Netherlands. The sampling periods, concentration statistics, and interpolated map used in this study can be found in the supporting information (Figures A2 and A3).

Short-term and mobile monitoring was done alternately with the same mobile platform and measurements were done in towns that were selected to cover the study areas of selected cohort studies in the Netherlands, supplemented with towns to increase the national coverage. Most major cities (Amsterdam, Rotterdam, Utrecht, Eindhoven, Groningen) were included as well as several medium sized towns. Mobile measurements were performed when driving from one short-term site to the next and then aggregated over a road segment. We elaborate on the different data sets in the paragraphs describing the training (2.2) and test data (2.4). In brief, data for all road segments (n= 14,392) were used for model development (training data). The short-term and long-term measurements were used as test data. The short-term measurements consisted of 400 sites (traffic and urban background) spread across the Netherlands for repeated stationary short-term measurements (30 minutes). The long-term measurements were used from a previous study [16], consisting of 42 sites in the cities of Amsterdam and Utrecht with 3 times 24 hour measurements. We furthermore selected 1000 random addresses in the Netherlands to compare models developed with different algorithms. Figure 2 shows a schematic overview of the measurements (used for training and testing) and models used in this study.

Figure 1

Map of towns with mobile and short-term UFP measurements.

↓

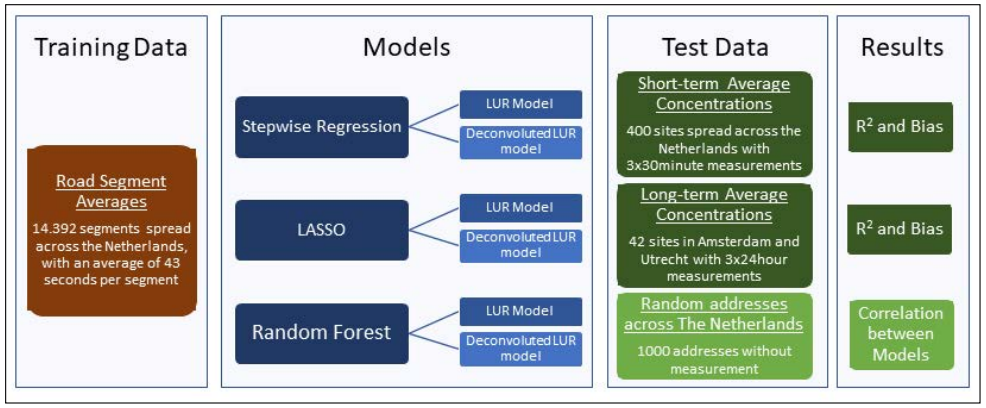


Map of towns with mobile and short-term UFP measurements in yellow (multiple sites per town) and regional background sites in green. Green square is the reference site (Bunnik).

Figure 2

Schematic overview of measurements and models.

↓



Measurements started after 9:15 AM and stopped before 4:00 PM, to avoid rush hours and increase comparability between road segments, following previous campaigns [10,16,23]. Avoiding rush hour resulted in lower spatial contrast but we assessed the lack of comparability to be more important than the loss of contrast. All measurements took place between June 2016 and November 2017. With this design, we captured the day-to-day, seasonal and a large part of the within-day variability of UFP concentration levels.

2.2 Training Data

All measurements were performed using a condensation particle counter (TSI, CPC 3007), installed in the back of an electric car (REVA, Mahindra Reva Electric Vehicles Pvt. Ltd., Bangalore, India). The setup of the car and equipment has been extensively described by Klompmaker et al. (2015). In brief, the CPC was connected via conductive silicon tubes to the outside of the car and performed measurements every second. In previous analysis we also found that speed does not affect the measurements [10]. The geographical location of the electric car was recorded using a Global Positioning Unit (GPS, Garmin eTrex Vista) and linked to the instruments in the car based on date and time. Following our previous mobile monitoring measurement campaigns [10], we corrected the GPS signal for small spatial errors by assigning all GPS points to the nearest road. We made sure that the GPS points were assigned to the correct road by creating a 30 m buffer around the route that was driven. Only GPS points within this buffer were assigned to roads that were completely in this buffer. Then we calculated average concentration levels of UFP per road segment. Road segments were on average 110 m long (SD: 68 m) and accumulated on average 43 seconds of UFP data (IQR: 9-44 seconds) over the study period.

UFP values were removed from the data set when the concentration was 500 particles/cm³ or less or if the UFP number counts increased or decreased in one second by a factor of 10 or more, as these reflect malfunctioning of the instrument. [10]. In total, less than 0.01% of the UFP data were removed due to these criteria.

As not all measurements were performed at the same time, we corrected all mobile measurements for temporal variation. This was done using a reference site with the same equipment (and tubing) as the electric vehicle, set up in the middle of the country (Bunnik; see figure 1). In our previous paper [22] we showed moderately high correlations between background concentrations of sites up to 200 km apart indicating that for UFP the Netherlands can be regarded as one airshed. We used the difference method to correct for temporal variation, following procedures in previous mobile monitoring campaigns [10,17]. First, the overall mean concentration for the entire campaign at the reference site was calculated. Next, we created a moving average of reference site measurements of 30 minutes (15 minutes before to 15 minutes after) for every timepoint of the mobile monitoring campaign. This value was then subtracted from the overall mean concentration at the reference site. Finally, this difference was added to the time-corresponding road segment average concentrations.

Every three weeks, the CPC used in the car and the CPC at the reference site were co-located in the laboratory to check comparability. This procedure is more elaborately described in the supplement information (Table A1). We found a median ratio (averaged over 1 min) for the two CPCs of 0.88 (CI: 0.78-1.01). Because 14% of the reference site UFP measurements were missing, we imputed UFP using a DiscMini that was placed as backup at the same reference site. Pearson correlation of hourly concentrations between the two instruments was 0.81. The formula used for imputation was CPC = 1345 + 0.957*DiscMini. Both instruments have a lower cutoff value of 10 nm.

2.2.1 Deconvolution

The idea of deconvoluting is to separate local UFP contributions from background concentrations. To do this, we use an approach developed by

Brantley et al. (2014) and Shairsingh et al. (2018) using a spline of minimums. First, measurements were averaged over different time scales (ranging from 60s to 2000s) and then correlated with a local traffic variable (Figure A1). The time scale for which the local traffic variable does not correlate with the average concentration anymore is considered the averaging period for the background concentration.

In figure A1, correlations are given for UFP measurements and traffic intensities in 50- and 1000-meter buffers, respectively. The local influence drops after an averaging period of 150 seconds. Taking a moving minimum value of ~200 seconds would represent a background signal in this case. With an average driving speed of about 30 km/h this would mean that background concentrations shift over a distance of about 1.6 km. Then, a regression spline is applied to interpolate between the minimum values from the time series. The minimum value in a time frame of 200 seconds around every measurement is thus the background concentration. The local signal is derived by subtracting the background concentration from the total concentration. We validated the urban background signal derived from deconvoluting by comparing it to short-term concentrations measured at urban background sites.

2.3 Models

Because there are inconclusive results as to which algorithms are best to be used in developing empirical air pollution models [18,25], we used three different land-use (regression) approaches (stepwise regression, LASSO and random forest) to assess the spatial variation of UFP across the Netherlands. For each approach, we also tested the deconvolution method.

2.3.1 Predictor variables

In accordance with our previous and most other mobile monitoring studies [10,17], we identified the middle of each road segment and used this coordinate to acquire spatial GIS predictors for LUR modelling (for an overview of GIS predictors see table A2). Besides the regional background predictions from the kriging map, a range of traffic variables was defined, including traffic intensity and road length variables (in 50–1000 m buffers); land use (e.g. port, industry, urban green, airports) and population / household density in buffers from 100 to 5000 m.

2.3.2 Stepwise Regression

We used a supervised forward stepwise regression approach that starts with an empty (intercept only) model and then adds variables based on goodness of fit determined by the adjusted R^2 value. This method has been used before for road segment averages [10,17]. Variables that *i)* lead to the largest increase in adjusted R^2, *ii)* for which the coefficient conforms with a predefined direction of effect and *iii)* for which the direction of effect of predictors already in the model does not change, are retained in the model. Predictor variables in the final model are removed from the model when they have a p-value larger than 0.1 or a variance inflation factor over 3.

2.3.3 LASSO

LASSO regression is optimized for prediction, because it deals with correlated predictors by imposing a penalty on the absolute size of the regression coefficients (attributing the full effect to only one of the variables, shrinking the other(s) towards zero). With LASSO regression the effects of some variables may be shrunken all the way to zero, and therefore it can be used for variable selection. We used the *glmnet* package in R (R Core Team (2013)) and default parameters for all LASSO models.

2.3.4 Random Forest

Random forest is a machine learning method that combines many weak classifiers/trees into one strong classifier. A weak classifier is a classifier whose error rate is only slightly better than random guessing. Random forest restricts all regression trees to a limited number of variables, creating a large collection of 'weak' trees. Each tree selects a specified number of variables at random and chooses the best variable to split the data. [26]. We set parameters to a 1000 trees and restricted the number of terminal nodes to 12 using the *randomForest* package.

2.4 Test Data

The first test set is based on short-term measurements (400 sites measured for 3x30minutes). About 50% of these measurements were taken at urban background sites (no major roads within 1000 m) and the rest at sites with high intensity traffic (> 10,000 vehicles/day). An advantage of this dataset is that sites are located across the Netherlands and sampled in the same time frame as the training data. A disadvantage is the short

sampling time (average 90 min) that is expected to result in low explained variances because of relatively instable average concentrations. The second set of test data consists of long-term average outdoor UFP concentrations at 42 different sites (traffic and urban background) in Amsterdam and Utrecht that have been collected according to protocols described by van Nunen et al. (2017) and Eeftens et al. (2012). Measurements were made close to the home, at first floor balconies for major roads and either balconies or gardens in homes located in minor roads. These sites were sampled with a DiscMini (Testo AG, Lenzkirch, Germany) 3 times for 24 hours. Previous studies have shown good agreement between CPCs and DiscMinis with limited differences in absolute values [28,29]. Measurements at these sites are better representatives of average concentrations than short-term measurements at a specific site, but the sites are only in two cities. To compare the two test results directly, we also tested the models on short-term measurements restricted to Amsterdam and Utrecht (n=80). Thirdly, we used a set of 1000 randomly selected home addresses to compare the predictions of the different models across the Netherlands.

③ Results

Table 1 shows the distribution of road segment average UFP concentrations, stratified by region and road type. Concentrations from mobile measurements were on average 15,861 particles/cm³ over an average duration of 43 seconds per street segment. Concentrations measured in Amsterdam (population: 860,000) and Utrecht (population: 350,000) were on average higher than those measured in other regions of the country, mainly because these measurements were done in two of the biggest cities in the Netherlands, while measurements in other regions were also done in smaller cities and municipalities. Lowest concentrations can be found in the North, where population density is small opposed to other regions in the country. Measurements on major roads were on average 55% higher than on minor roads.

3.1 Deconvoluting

Table 2 shows the average background UFP concentrations obtained from the three different measurement strategies (short-term, mobile, and regional long-term), per region. Average concentrations at short-term urban background sites were very similar to the deconvoluted urban background signal calculated from the measurements with the mobile platform. The regional background kriging map was based on measurements outside of cities resulting in concentrations that were considerably lower than the urban background at short-term sites and deconvoluted background signals. However, the same regional pattern can be observed. In figure A4 we show the same comparison but divided by city.

Table 1

Measured road segment average UFP concentrations in particles/cm³

↓

Region	Road Type	Number of Road Segments	Mean	10th Percentile	90th Percentile
AMS-UTR	Minor	1,221	15,557	7,177	26,289
	Major	2,094	21,831	8,887	40,799
East	Minor	1,118	11,495	6,847	16,902
	Major	1,150	17,061	8,110	29,176
North	Minor	2,144	11,044	3,103	18,288
	Major	614	16,290	5,082	28,851
South	Minor	1,490	11,621	5,378	18,609
	Major	1,105	17,697	6,919	32,581
West	Minor	363	12,179	6,253	20,021
	Major	3,016	18,694	7,069	33,132
All	Minor	6,561	12,080	5,061	20,250
	Major	8,316	18,618	7,453	33,907

3.2 External Prediction

Table 3 shows the performance of all individual algorithms based on short-term and longer-term test datasets. In general, all algorithms perform equally well with the proportion of variance explained (R^2) in the external test sets ranging from 0.25 to 0.35 for the short-term test data and from 0.52 to 0.60 for the longer-term test data. Both random forest algorithms seem to perform slightly better on the short-term measurements, but deconvoluted stepwise regression and LASSO explain the longer-term measurements somewhat better. All models tend to overestimate our external test data, usually by between ~2000 and ~4000 particles/cm³. Appendix B provides scatterplots of measured (short-term and longer-term) versus predicted concentrations and Bland Altman plots for all six algorithms. We did not find systematic differences in the extent of bias between different regions or road types (Table A3).

Variables selected in the stepwise regression and LASSO models can be found in table 4. As random forest does not perform variable selection, we show the top 7 variables that were most influential in random forest. Selected variables with regression coefficients and p-values can be found in the supplement (Tables B1-B4). For random forest we show the distribution of minimal depth for the best variables (Figures B5 and B7). Figures B1-B4, B6, and B8 show scatterplots and Bland-Altman plots for the predictions of the mobile models on all 3 test data sets (i.e., short-term, short-term restructured to the Utrecht-Amsterdam region and long-term Utrecht Amsterdam Region) for all 6 algorithms. All models select (or give high importance to) very local traffic intensity variables, population density in a large buffer (5000m) and the presence of a nearby port. The regional background map was selected as predictor in all three (deconvoluted) background models, but not in the standard models.

Table 2

Comparison of average background concentrations per region obtained from short-term monitoring, mobile monitoring, and regional background monitoring.

↓

Region	Average concentration of short-term urban background sites	Average background concentration of deconvoluted mobile measurements	Average concentration of regional background long-term sites
AMS - UTR	11,348	11,420	/
East	7,583	8,493	5,543
North	6,100	6,130	4,359
South	8,732	8,787	5,856
West	10,318	10,356	6,053

Concentrations in particles/cm³.

Table 3

Comparison of predicted concentrations from different algorithms on external test data.

↓

	Short-Term		Short-Term AMS-UTR		Long-Term AMS-UTR	
	R²	**Mean Bias**	**R²**	**Mean Bias**	**R²**	**Mean Bias**
Stepwise Regression	0.28	2,883	0.31	3,422	0.60	2,156
Stepwise Regression Deconvoluted	0.30	3,470	0.29	3,904	0.60	2,690
LASSO	0.29	3,225	0.31	3,092	0.60	1,806
LASSO Deconvoluted	0.25	3,921	0.35	3,655	0.55	2,392
Random Forest	0.34	3,132	0.32	3,025	0.52	2,362
Random Forest Deconvoluted	0.34	3,390	0.31	3,146	0.52	1,818

Bias is expressed in particles/cm³.

Table 4

Predictor variables selected by the different models.

↓

	Stepwise	Stepwise deconvoluted		LASSO	LASSO deconvoluted		RF	RF deconvoluted	
		Background	Local		Background	Local		Background	Local
Background predictors									
Regional background		+			+			+	
Household 5000m[1]	+			+	+		+	+	
Population 5000m		+					+	+	
Port 5000m	+	+		+	+			+	
Industry 5000m									
Highways 1000m					+				
Major roads 1000m									
Heavy traffic intensity 1000m					+				
Industry 5000m					+				
Urban greenery 5000m								+	
Agriculture 5000m								+	
Y-Coordinate								+	
Local predictors									
Traffic intensity near				+					+
Heavy traffic intensity near[2]	+								
Traffic intensity 50m	+		+	+		+	+		+
Traffic intensity major roads 50m	+		+	+			+		+
Traffic intensity 100m				+			+		+
Traffic intensity major roads 100m							+		+
Traffic intensity 300m			+	+			+		+
Household 300m			+						
Restaurants near									+

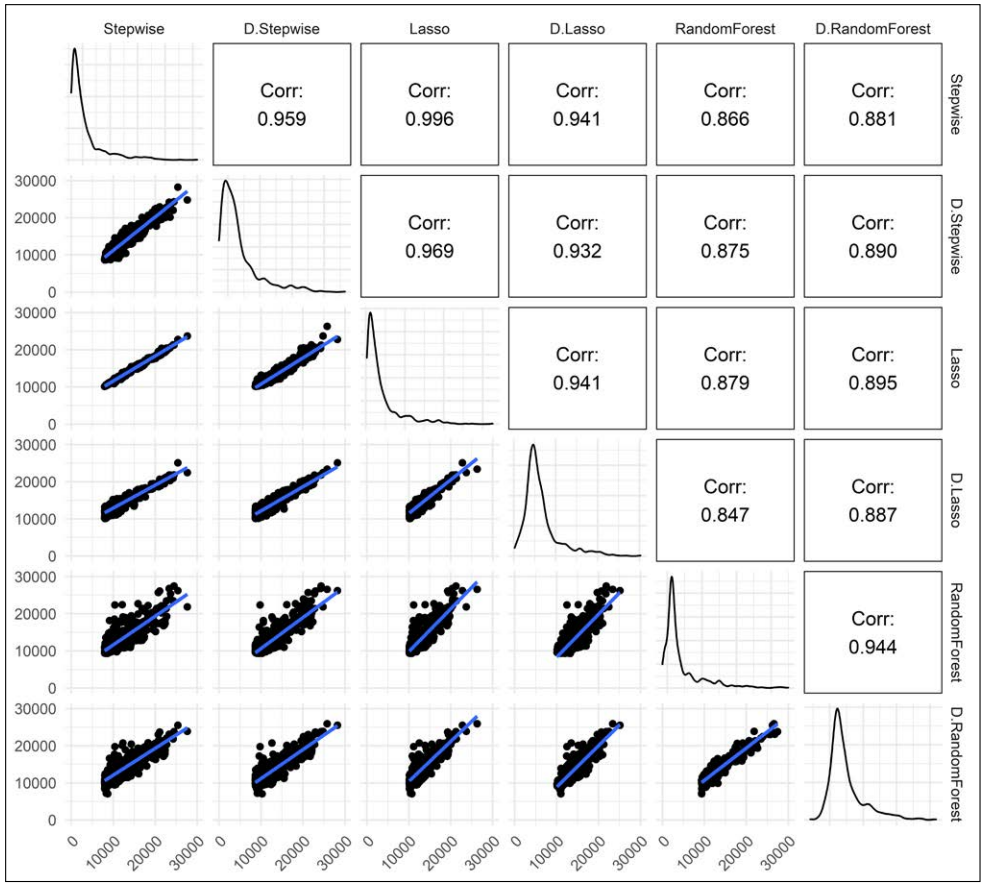
Variables in the random forest column are the top 7 variables across several performance methods.[1]refers to buffer size; [2]near refers to nearest road.

Applying the different algorithms to a set of random addresses (without measurements) across the Netherlands resulted in predictions that were highly correlated, with Pearson correlations ranging from 0.84 to 0.99 (Figure 3). Correlations were higher between stepwise regression and LASSO models than between stepwise/LASSO models and the random forest models. Table A4 shows the distributions of the predicted concentrations for all algorithms.

Figure 3

Comparison of prediction algorithms at randomly selected addresses across the Netherlands (n=1000).

↓



Concentrations in particles/cm³. Abbreviation "D." refers to the deconvoluted method.

④ Discussion

This is the first study showing that it is possible to develop predictive models for UFP concentrations on a national scale (41,543 km²), while accounting for local variability of UFP concentrations in urban areas. We achieved this by combining mobile monitoring with targeted regional background monitoring. We found similar performance for all six algorithms. Models explained 25% to 35% of the variability in measured short-term average concentrations across the Netherlands and 52% to 60% of the variability in longer-term measurements. Correlations between predictions with different algorithms at random addresses were high with Pearson correlations between 0.84 and 0.99.

4.1 Development of a national model

While LUR models have been proven to be useful in assessing small scale variation of UFP [15] and are able to predict retrospective and prospective exposures [30,31], few studies used LUR models to investigate the effect of long-term UFP exposure on adverse health effects [1,2]. Concentrations of UFP vary significantly over space and time and it can therefore be difficult to assess exposure in a large area and over a long time. LUR models of UFP have therefore been restricted to a few cities. There is, however, a limited ability of LUR models measured in one city to be transferred to another city and even more so to rural areas [9]. We therefore had to take into account variation on larger spatial scales, including urban and regional background variation in addition to the more commonly included local scale variation within urban areas. Developing a national model for UFP is more challenging than for more commonly studied pollutants such as $PM_{2.5}$, NO_2 and O_3 for which more advanced modelling frameworks based on deterministic models and emission inventories are available. Such studies can often rely on routine monitoring, chemical transport modelling (CTM) and/or satellite data [32–35]. The only wide-scale model for UFP so far was used in the United States in the California Teachers Study, covering about 424,000km² (about 10 times the size of The Netherlands). That study estimated UFP exposure at a 4 km grid based upon chemical transport modelling[3].

Routine monitoring, CTM, and satellite observations were not available for UFP, so we combined mobile monitoring in many towns across the country with specific long-term regional background monitoring. As mobile monitoring was not feasible in all towns of the Netherlands, our assumption was that the regional background component would account for differences in exposure related to region of the country.

4.2 Comparing prediction algorithms

We only found minor differences in explained variance and bias between different algorithms for developing land use regression models of spatial concentration variations of UFP. The low R^2 values of the short-term test set is primarily due to the influence of temporal variability in the short-term test data (3 x 30 minutes), which do not represent a stable average concentration. For the long-term test data compared to short-term test data we found substantially higher R^2, which can be explained by the longer measurement time of long-term measurements [17,18,30]. We did not have long-term test data at the national scale, but we did observe that the R^2 values based upon short-monitoring tests data in Utrecht-Amsterdam (the cities with longer-term monitoring) and the entire Netherlands were similar. We suspect that on a national scale, the R^2 values for longer-term averages is similar to what we found in Amsterdam-Utrecht.

The minor differences between models (on the same test data) were consistent with findings of a previous study evaluating models based upon mobile and short-monitoring where we tested a much wider range of models and also found minor differences between approaches [18]. While no conclusion can be drawn regarding which algorithm performs best, it is encouraging that different approaches in this study generated robust predictions at randomly selected sites. Chen et al. (2019) compared modelling algorithms in a European setting with long-term measurements of NO_2 and $PM_{2.5}$ and also found minor differences between a wide range of algorithms, including the three algorithms included in the current paper.

Predictors that were selected or given the most weight in the different exposure models that we explored were similar. All models select very local traffic variables, population density in a large buffer and the presence of a port. These variables represent known sources of ultrafine particles and have been included in previous LUR models in the Netherlands [10,16,30,36] and elsewhere [12,13,37–39]. This persists in the deconvolution method, with local traffic variables selected in the local model and population density and port area in the background model.

All three background models in the deconvolution approach selected the regional UFP background map as a predictor into the model, while none of the standard algorithms (without deconvolution) selects the regional background concentration as important predictor. Conceptually, regional variation should be represented in a national model. However, all standard models select mainly local traffic source related variables into the model. Regional differences are apparently also represented by the difference in local traffic intensities or large-scale population density.

4.3 Bias

Predicted UFP concentrations from the different algorithms were on average 20% (~2,000 to ~4,000 particles/cm³) higher than stationary measurements. This is less than the overestimation that we found in previous studies using mobile monitoring [10,17,18]. Previous mobile models were based on measurements performed in urban areas and overestimated measured concentrations on average by ~5,000 particles/cm³, tested on the exact same test dataset with longer-term average concentrations that we used in the present study. In one of those previous studies, mobile stepwise, LASSO and random forest models overestimated exposure by 6,902, 3,374 and 5,688 particles/cm³ respectively [18]. Other studies that used mobile models found similar differences. For example, a difference of 5,300 particles/cm³ has been found between mobile monitoring and home outdoor sites in Chelsea (Massachusetts, USA) [2]. Two other studies compared measurements of UFP on the sidewalk and at the façade of buildings and found a difference in concentration levels of about 15-28% [13,40,41].

Overestimation comes from the fact that these models are based on data collected in the middle of the road, while we assign them to home outdoor residential locations. Of note, some over- and underprediction can also be expected by measuring only during daytime and not measuring during rush hours and night-time, respectively. In a previous study, we documented a high correlation (R^2 >0.95) between UFP exposures at 50 fixed sites in Amsterdam for different times of the day including rush hours, daytime non-rush hours and 24-hour average [1]. We also documented only small differences in the 24-hour average concentration and the average of the period used for mobile monitoring in that study (contrast between traffic and urban background sites was 16,000 and 21,000 particles/cm³ for 9-16 hour and 24 hours, respectively). While the absolute difference in concentrations between on-road driving and at the façade of buildings is higher on major roads, the relative difference is similar for minor and major roads. Therefore, we argue that the overestimation is modest and non-differential across road types. As the modest over-estimation did not differ much across road types, the exposure contrast in epidemiological studies will only be mildly affected. We suspect that the impact on effect estimates in epidemiological studies is likely limited. Errors from land use regression models are generally regarded to be Berkson-like. Berkson type measurement error affects the power of an epidemiological study, but does not bias regression coefficients [42]. A correction factor in the order of 15 to 30% seems to be the most appropriate for correcting on-road measurements if risk is to be expressed per unit of exposure.

4.4 Deconvolution

Adding a deconvolution step to stepwise, LASSO and random forest did not increase the predictive performance and did not reduce bias in our study. Shairsingh et al. (2018), who developed and applied this method, also found only minor differences in predictive performance between deconvoluted and standard models but found less bias when resolved models were used for external validation. A potential explanation for this difference could be the fact that Shairsingh and coauthors used spatio-temporal LUR models for their predictions, while we used purely spatial models. The temporal aspect also allowed them to split the background signal in an urban background and regional background signal, making it a 3-step deconvolution method.

All three deconvolution models included the regional background concentration in their final model, whereas the standard models did not. A national model including the background and local scale fits better with physical reality of spatial distribution of air pollution. Though bias is slightly higher on average in the deconvoluted models, bias is more evenly distributed along the concentration range (major vs minor roads). Separate models for different scales have been developed previously at the national and European scale in empirical LUR models [4,43]. In dispersion modelling the regional background, urban background and local scale are explicitly modelled with separate models as well (e.g Hvidtfeldt et al. 2019).

4.5 Implications for epidemiological studies

Applying models that are developed based on measurements from one city is hampered by an unknown background concentration when they are transferred to another city. This study shows that variation in UFP concentrations is not only influenced by local variability within cities. Residences similar in terms of local (e.g. nearby traffic) and urban sources (e.g. population size) in different regions of the country still differ in exposure for two reasons. First, as previously described in van de Beek et al. (2020), substantial regional differences in background UFP concentrations were found within the Netherlands (Figure A2). Second, background concentrations in Dutch cities ranged from about 3,300 to 13,100 particles/cm³ (Figure A4). Admittingly, these numbers reflect total background concentrations (urban background and regional background). Still, the differences in background concentrations between cities are generally bigger than the difference between traffic and urban background stationary sites within a city in this study (average of 3,500 particles/cm³).

As standard LUR models are often limited to include large-scale background differences [45,46], adding a regional background component creates a better representation of the spatial variation of UFP in the Netherlands. While the regional background map was not selected in all standard models, it was selected in all deconvoluted models. Deconvolution of the mobile monitoring signal is therefore an alternative for specific regional background monitoring, especially when considering larger countries than the Netherlands, that can have even larger differences in regional background concentration. The available resources determine the feasibility of this approach.

Measuring with a mobile platform enabled us to measure spatially diverse environments in a limited amount of time, and with a limited number of monitoring devices. This is cost-effective and especially within city limits, it can capture the high variability of UFP concentrations in complex urban terrains. The goodness of fit of the UFP models was moderate to high when compared to long-term measurements. This performance is comparable to the performance of models successfully applied in epidemiological studies for $PM_{2.5}$, NO_2 and BC [47]. UFP models have also been developed based on longer term monitoring data in a few studies [36,48,49]. The number of sites in these studies was small however (between 20 and 50 sites). The cost of the UFP equipment and the need for regular technician supervision precludes long-term monitoring at a large number of locations at this time. Though we have not formally compared the performance of UFP models based upon mobile and long-term monitoring, we suspect that mobile monitoring is more attractive for UFP.

For developing a national model, separate modelling of background and local scale exposures should be considered. For larger countries, characterization of the regional background is likely best by a combination of fixed site monitoring and deconvolution of the mobile monitoring signal. To assess the potential overestimation by mobile monitoring models, longer-term measurements at a limited number of residential sites spread over the study domain should be part of the study design. Models used in this study include regional, urban background and local information on UFP concentrations, making them applicable to large nation-wide cohorts in the Netherlands.

Supporting Information

The Supporting information is divided into two subsections. Appendix A contains general supporting information. Appendix B contains detailed information for all six prediction algorithms.

Funding

This work was supported by an ASPASIA grant from the Dutch Research Council (NWO) to Dr. Ulrike Gehring (project number 015.010.044), the Environmental Defense Fund, EXPOSOME-NL (NWO grant number 024.004.017) and EXPANSE (EU-H2020 Grant number 874627).

References

(1) Downward, G. S.; van Nunen, E. J. H. M.; Kerckhoffs, J.; Vineis, P.; Brunekreef, B.; Boer, J. M. A.; Messier, K. P.; Roy, A.; Verschuren, W. M. M.; van Der Schouw, Y. T.; Sluijs, I.; Gulliver, J.; Hoek, G.; Vermeulen, R. Long-Term Exposure to Ultrafine Particles and Incidence of Cardiovascular and Cerebrovascular Disease in a Prospective Study of a Dutch Cohort. *Environ. Health Perspect.* **2018**, *126* (12). https://doi.org/10.1289/EHP3047.

(2) Simon, M. C.; Hudda, N.; Naumova, E. N.; Levy, J. I.; Brugge, D.; Durant, J. L. Comparisons of Traffic-Related Ultrafine Particle Number Concentrations Measured in Two Urban Areas by Central, Residential, and Mobile Monitoring. *Atmos. Environ.* **2017**, *169*, 113–127. https://doi.org/10.1016/j.atmosenv.2017.09.003.

(3) Ostro, B.; Hu, J.; Goldberg, D.; Reynolds, P.; Hertz, A.; Bernstein, L.; Kleeman, M. J. Associations of Mortality with Long-Term Exposures to Fine and Ultrafine Particles, Species and Sources: Results from the California Teachers Study Cohort. *Environ. Health Perspect.* **2015**, *123* (6), 549–556. https://doi.org/10.1289/ehp.1408565.

(4) Beelen, R.; Hoek, G.; van den Brandt, P. A.; Goldbohm, R. A.; Fischer, P.; Schouten, L. J.; Jerrett, M.; Hughes, E.; Armstrong, B.; Brunekreef, B. Long-Term Effects of Traffic-Related Air Pollution on Mortality in a Dutch Cohort (NLCS-AIR Study). *Environ. Health Perspect.* **2008**, *116* (2), 196–202. https://doi.org/10.1289/ehp.10767.

(5) Bustin, L.; Tritscher, T.; Spielvogel, J.; Bischof, O. F.; Scheckman, J.; Krinke, T. CEN Standard "Harmonized Counting of Atmospheric Ultrafine Particles" and UFP Monitoring Initiatives in Europe. In *Journal of Physics: Conference Series*; Institute of Physics Publishing, 2019; Vol. 1420, p 12010. https://doi.org/10.1088/1742-6596/1420/1/012010.

(6) Larson, T.; Henderson, S. B.; Brauer, M. Mobile Monitoring of Particle Light Absorption Coefficient in an Urban Area as a Basis for Land Use Regression. *Environ. Sci. Technol.* **2009**, *43* (13), 4672–4678. https://doi.org/10.1021/es803068e.

(7) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An Hourly Regression Model for Ultrafine Particles in a Near-Highway Urban Area. *Environ. Sci. Technol.* **2014**, *48* (6), 3272–3280. https://doi.org/10.1021/es404838k.

(8) Weichenthal, S.; Farrell, W.; Goldberg, M.; Joseph, L.; Hatzopoulou, M. Characterizing the Impact of Traffic and the Built Environment on Near-Road Ultrafine Particle and Black Carbon Concentrations. *Environ. Res.* **2014**, *132*, 305–310. https://doi.org/10.1016/j.envres.2014.04.007.

(9) Shairsingh, K. K.; Jeong, C. H.; Wang, J. M.; Brook, J. R.; Evans, G. J. Urban Land Use Regression Models: Can Temporal Deconvolution of Traffic Pollution Measurements Extend the Urban LUR to Suburban Areas? *Atmos. Environ.* **2019**, *196*, 143–151. https://doi.org/10.1016/j.atmosenv.2018.10.013.

(10) Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of Ultrafine Particles and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environ. Sci. Technol.* **2016**, *50* (23), 12894–12902. https://doi.org/10.1021/acs.est.6b03476.

(11) Hankey, S.; Marshall, J. D. Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, $PM_{2.5}$, Particle Size) Using Mobile Monitoring. *Environ. Sci. Technol.* **2015**, *49* (15), 9194–9202. https://doi.org/10.1021/acs.est.5b01209.

(12) Farrell, W.; Weichenthal, S.; Goldberg, M.; Valois, M. F.; Shekarrizfard, M.; Hatzopoulou, M. Near Roadway Air Pollution across a Spatially Extensive Road and Cycling Network. *Environ. Pollut.* **2016**, *212*, 498–507. https://doi.org/10.1016/j.envpol.2016.02.041.

(13) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Reali, C.; Sun, T.; Evans, G. J. Development of a Land-Use Regression Model for Ultrafine Particles in Toronto, Canada. *Atmos. Environ.* **2015**, *110*, 84–92. https://doi.org/10.1016/j.atmosenv.2015.02.018.

(14) Minet, L.; Gehr, R.; Hatzopoulou, M. Capturing the Sensitivity of Land-Use Regression Models to Short-Term Mobile Monitoring Campaigns Using Air Pollution Micro-Sensors. *Environ. Pollut.* **2017**, *230*, 280–290. https://doi.org/10.1016/j.envpol.2017.06.071.

(15) Hoek, G. Methods for Assessing Long-Term Exposures to Outdoor Air Pollutants. *Current environmental health reports*. Springer December 1, 2017, pp 450–462. https://doi.org/10.1007/s40572-017-0169-5.

(16) van Nunen, E.; Vermeulen, R.; Tsai, M.-Y.; Probst-Hensch, N.; Ineichen, A.; Davey, M.; Imboden, M.; Ducret-Stich, R.; Naccarati, A.; Raffaele, D.; Ranzi, A.; Ivaldi, C.; Galassi, C.; Nieuwenhuijsen, M.; Curto, A.; Donaire-Gonzalez, D.; Cirach, M.; Chatzi, L.; Kampouri, M.; Vlaanderen, J.; Meliefste, K.; Buijtenhuijs, D.; Brunekreef, B.; Morley, D.; Vineis, P.; Gulliver, J.; Hoek, G. Land Use Regression Models for Ultrafine Particles in Six European Areas. *Environ. Sci. Technol.* **2017**, *51* (6), 3336–3345. https://doi.org/10.1021/acs.est.6b05920.

(17) Kerckhoffs, J.; Hoek, G.; Vlaanderen, J.; van Nunen, E.; Messier, K.; Brunekreef, B.; Gulliver, J.; Vermeulen, R. Robustness of Intra Urban Land-Use Regression Models for Ultrafine Particles and Black Carbon Based on Mobile Monitoring. *Environ. Res.* **2017**, *159*, 500–508. https://doi.org/10.1016/j.envres.2017.08.040.

(18) Kerckhoffs, J.; Hoek, G.; Portengen, L.; Brunekreef, B.; Vermeulen, R. C. H. Performance of Prediction Algorithms for Modeling Outdoor Air Pollution Spatial Surfaces. *Environ. Sci. Technol.* **2019**, *53* (3), 1413–1421. https://doi.org/10.1021/acs.est.8b06038.

(19) Weichenthal, S.; Ryswyk, K. Van; Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A Land Use Regression Model for Ambient Ultrafine Particles in Montreal, Canada: A Comparison of Linear Regression and a Machine Learning Approach. *Environ. Res.* **2016**, *146*, 65–72. https://doi.org/10.1016/j.envres.2015.12.016.

(20) Hong, K. Y.; Pinheiro, P. O.; Minet, L.; Hatzopoulou, M.; Weichenthal, S. Extending the Spatial Scale of Land Use Regression Models for Ambient Ultrafine Particles Using Satellite Images and Deep Convolutional Neural Networks. *Environ. Res.* **2019**, *176*, 108513. https://doi.org/10.1016/j.envres.2019.05.044.

(21) Shairsingh, K. K.; Jeong, C. H.; Wang, J. M.; Evans, G. J. Characterizing the Spatial Variability of Local and Background Concentration Signals for Air Pollution at the Neighbourhood Scale. *Atmos. Environ.* **2018**, *183*, 57–68. https://doi.org/10.1016/j.atmosenv.2018.04.010.

(22) van de Beek, E.; Kerckhoffs, J.; Hoek, G.; Sterk, G.; Meliefste, K.; Gehring, U.; Vermeulen, R. Spatial and Spatiotemporal Variability of Regional Background Ultrafine Particle Concentrations in the Netherlands. *Environ. Sci. Technol.* **2020**, *11*, 17. https://doi.org/10.1021/acs.est.0c06806.

(23) Klompmaker, J. O.; Montagne, D. R.; Meliefste, K.; Hoek, G.; Brunekreef, B. Spatial Variation of Ultrafine Particles and Black Carbon in Two Cities: Results from a Short-Term Measurement Campaign. *Sci. Total Environ.* **2015**, *508*, 266–275. https://doi.org/10.1016/j.scitotenv.2014.11.088.

(24) Brantley, H. L.; Hagler, G. S. W.; Kimbrough, E. S.; Williams, R. W.; Mukerjee, S.; Neas, L. M. Mobile Air Monitoring Data-Processing Strategies and Effects on Spatial Air Pollution Trends. *Atmos. Meas. Tech.* **2014**, *7* (7), 2169–2183. https://doi.org/10.5194/amt-7-2169-2014.

(25) Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Katsouyanni, K.; Janssen, N. A. H.; Martin, R. V.; Samoli, E.; Schwartz, P. E.; Stafoggia, M.; Bellander, T.; Strak, M.; Wolf, K.; Vienneau, D.; Vermeulen, R.; Brunekreef, B.; Hoek, G. A Comparison of Linear Regression, Regularization, and Machine Learning Algorithms to Develop Europe-Wide Spatial Models of Fine Particles and Nitrogen Dioxide. *Environ. Int.* **2019**, *130*. https://doi.org/10.1016/j.envint.2019.104934.

(26) Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning. *Elements* **2009**, *1*, 337–387. https://doi.org/10.1007/b94608.

(27) Eeftens, M.; Tsai, M. Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; De Nazelle, A.; de Vocht, F.; Declercq, C.; Dedele, A.; Eriksen, K.; Galassi, C.; Gražulevičiene, R.; Grivas, G.; Heinrich, J.; Hoffmann, B.; Iakovides, M.; Ineichen, A.; Katsouyanni, K.; Korek, M.; Krämer, U.; Kuhlbusch, T.; Lanki, T.; Madsen, C.; Meliefste, K.; Mölter, A.; Mosler, G.; Nieuwenhuijsen, M.; Oldenwening, M.; Pennanen, A.; Probst-Hensch, N.; Quass, U.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Udvardy, O.; Vaskövi, É.; Weinmayr, G.; Brunekreef, B.; Hoek, G. Spatial Variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ Absorbance and PMcoarse Concentrations between and within 20 European Study Areas and the Relationship with NO2 - Results of the ESCAPE Project. *Atmos. Environ.* **2012**, *62*, 303–317. https://doi.org/10.1016/j.atmosenv.2012.08.038.

(28) Asbach, C.; Kaminski, H.; Von Barany, D.; Kuhlbusch, T. A. J.; Monz, C.; Dziurowitz, N.; Pelzer, J.; Vossen, K.; Berlin, K.; Dietrich, S.; Götz, U.; Kiesling, H. J.; Schierl, R.; Dahmann, D. Comparability of Portable Nanoparticle Exposure Monitors. In *Annals of Occupational Hygiene*; 2012; Vol. 56, pp 606–621. https://doi.org/10.1093/anhyg/mes033.

(29) Meier, R.; Clark, K.; Riediker, M. Comparative Testing of a Miniature Diffusion Size Classifier to Assess Airborne Ultrafine Particles Under Field Conditions. *Aerosol Sci. Technol.* **2013**, *47* (1), 22–28. https://doi.org/10.1080/02786826.2012.720397.

(30) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712–8720. https://doi.org/10.1021/es505791g.

(31) Simon, M. C.; Naumova, E. N.; Levy, J. I.; Brugge, D.; Durant, J. L. Ultrafine Particle Number Concentration Model for Estimating Retrospective and Prospective Long-Term Ambient Exposures in Urban Neighborhoods. *Environ. Sci. Technol.* **2020**. https://doi.org/10.1021/acs.est.9b03369.

(32) Vienneau, D.; De Hoogh, K.; Bechle, M. J.; Beelen, R.; Van Donkelaar, A.; Martin, R. V.; Millet, D. B.; Hoek, G.; Marshall, J. D. Western European Land Use Regression Incorporating Satellite- and Ground-Based Measurements of NO2 and PM_{10}. *Environ. Sci. Technol.* **2013**, *47* (23), 13555–13564. https://doi.org/10.1021/es403089q.

(33) Hystad, P.; Setton, E.; Cervantes, A.; Poplawski, K.; Deschenes, S.; Brauer, M.; van Donkelaar, A.; Lamsa, L.; Martin, R.; Jerrett, M.; Demers, P. Creating National Air Pollution Models for Population Exposure Assessment in Canada. *Environ. Health Perspect.* **2011**, *119* (8), 1123–1129. https://doi.org/10.1289/ehp.1002976.

(34) Novotny, E. V.; Bechle, M. J.; Millet, D. B.; Marshall, J. D. National Satellite-Based Land-Use Regression: NO2 in the United States. *Environ. Sci. Technol.* **2011**, *45* (10), 4407–4414. https://doi.org/10.1021/es103578x.

(35) de Hoogh, K.; Gulliver, J.; Donkelaar, A. van; Martin, R. V.; Marshall, J. D.; Bechle, M. J.; Cesaroni, G.; Pradas, M. C.; Dedele, A.; Eeftens, M.; Forsberg, B.; Galassi, C.; Heinrich, J.; Hoffmann, B.; Jacquemin, B.; Katsouyanni, K.; Korek, M.; Künzli, N.; Lindley, S. J.; Lepeule, J.; Meleux, F.; de Nazelle, A.; Nieuwenhuijsen, M.; Nystad, W.; Raaschou-Nielsen, O.; Peters, A.; Peuch, V. H.; Rouil, L.; Udvardy, O.; Slama, R.; Stempfelet, M.; Stephanou, E. G.; Tsai, M. Y.; Yli-Tuomi, T.; Weinmayr, G.; Brunekreef, B.; Vienneau, D.; Hoek, G. Development of West-European $PM_{2.5}$ and NO_2 Land Use Regression Models Incorporating Satellite-Derived and Chemical Transport Modelling Data. *Environ. Res.* **2016**, *151*, 1–10. https://doi.org/10.1016/j.envres.2016.07.005.

(36) Hoek, G.; Beelen, R.; Kos, G.; Dijkema, M.; Zee, S. C. V. Der; Fischer, P. H.; Brunekreef, B. Land Use Regression Model for Ultrafine Particles in Amsterdam. *Environ. Sci. Technol.* **2011**, *45* (2), 622–628. https://doi.org/10.1021/es1023042.

(37) Rivera, M.; Basagaña, X.; Aguilera, I.; Agis, D.; Bouso, L.; Foraster, M.; Medina-Ramón, M.; Pey, J.; Künzli, N.; Hoek, G. Spatial Distribution of Ultrafine Particles in Urban Settings: A Land Use Regression Model. *Atmos. Environ.* **2012**, *54*, 657–666. https://doi.org/10.1016/j.atmosenv.2012.01.058.

(38) Weichenthal, S.; Van Ryswyk, K.; Goldstein, A.; Shekarrizfard, M.; Hatzopoulou, M. Characterizing the Spatial Distribution of Ambient Ultrafine Particles in Toronto, Canada: A Land Use Regression Model. *Environ. Pollut.* **2015**, *208*, 241–248. https://doi.org/10.1016/j.envpol.2015.04.011.

(39) Jones, R. R.; Hoek, G.; Fisher, J. A.; Hasheminassab, S.; Wang, D.; Ward, M. H.; Sioutas, C.; Vermeulen, R.; Silverman, D. T. Land Use Regression Models for Ultrafine Particles, Fine Particles, and Black Carbon in Southern California. *Sci. Total Environ.* **2020**, *699*, 134234. https://doi.org/10.1016/j.scitotenv.2019.134234.

(40) Kaur, S.; Nieuwenhuijsen, M. J.; Colvile, R. N. Pedestrian Exposure to Air Pollution along a Major Road in Central London, UK. *Atmos. Environ.* **2005**, *39* (38), 7307–7320. https://doi.org/10.1016/j.atmosenv.2005.09.008.

(41) Ragettli, M. S.; Ducret-Stich, R. E.; Foraster, M.; Morelli, X.; Aguilera, I.; Basagaña, X.; Corradi, E.; Ineichen, A.; Tsai, M. Y.; Probst-Hensch, N.; Rivera, M.; Slama, R.; Künzli, N.; Phuleria, H. C. Spatio-Temporal Variation of Urban Ultrafine Particle Number Concentrations. *Atmos. Environ.* **2014**, *96*, 275–283. https://doi.org/10.1016/j.atmosenv.2014.07.049.

(42) Armstrong, B. G. Effect of Measurement Error on Epidemiological Studies of Environmental and Occupational Exposures. *Occupational and Environmental Medicine*. 1998, pp 651–656. https://doi.org/10.1136/oem.55.10.651.

(43) Beelen, R.; Hoek, G.; Pebesma, E.; Vienneau, D.; de Hoogh, K.; Briggs, D. J. Mapping of Background Air Pollution at a Fine Spatial Scale across the European Union. *Sci. Total Environ.* **2009**, *407* (6), 1852–1867. https://doi.org/10.1016/j.scitotenv.2008.11.048.

(44) Hvidtfeldt, U. A.; Sørensen, M.; Geels, C.; Ketzel, M.; Khan, J.; Tjønneland, A.; Overvad, K.; Brandt, J.; Raaschou-Nielsen, O. Long-Term Residential Exposure to $PM_{2.5}$, PM_{10}, Black Carbon, NO_2, and Ozone and Mortality in a Danish Cohort. *Environ. Int.* **2019**, *123*, 265–272. https://doi.org/10.1016/j.envint.2018.12.010.

(45) Kerckhoffs, J.; Wang, M.; Meliefste, K.; Malmqvist, E.; Fischer, P.; Janssen, N. A. H.; Beelen, R.; Hoek, G. A National Fine Spatial Scale Land-Use Regression Model for Ozone. *Environ. Res.* **2015**, *140*, 440–448. https://doi.org/10.1016/j.envres.2015.04.014.

(46) Hoek, G.; Eeftens, M.; Beelen, R.; Fischer, P.; Brunekreef, B.; Boersma, K. F.; Veefkind, P. Satellite NO2 Data Improve National Land Use Regression Models for Ambient NO2 in a Small Densely Populated Country. *Atmos. Environ.* **2015**, *105*, 173–180. https://doi.org/10.1016/j.atmosenv.2015.01.053.

(47) de Hoogh, K.; Chen, J.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Katsouyanni, K.; Klompmaker, J.; Martin, R. V.; Samoli, E.; Schwartz, P. E.; Stafoggia, M.; Bellander, T.; Strak, M.; Wolf, K.; Vienneau, D.; Brunekreef, B.; Hoek, G. Spatial $PM_{2.5}$, NO_2, O_3 and BC Models for Western Europe – Evaluation of Spatiotemporal Stability. *Environ. Int.* **2018**, *120*, 81–92. https://doi.org/10.1016/j.envint.2018.07.036.

(48) Cattani, G.; Gaeta, A.; Di Menno di Bucchianico, A.; De Santis, A.; Gaddi, R.; Cusano, M.; Ancona, C.; Badaloni, C.; Forastiere, F.; Gariazzo, C.; Sozzi, R.; Inglessis, M.; Silibello, C.; Salvatori, E.; Manes, F.; Cesaroni, G. Development of Land-Use Regression Models for Exposure Assessment to Ultrafine Particles in Rome, Italy. *Atmos. Environ.* **2017**, *156*, 52–60. https://doi.org/10.1016/j.atmosenv.2017.02.028.

(49) Wolf, K.; Cyrys, J.; Harciníková, T.; Gu, J.; Kusch, T.; Hampel, R.; Schneider, A.; Peters, A. Land Use Regression Modeling of Ultrafine Particles, Ozone, Nitrogen Oxides and Markers of Particulate Matter Pollution in Augsburg, Germany. *Sci. Total Environ.* **2017**. https://doi.org/10.1016/j.scitotenv.2016.11.160.

Supplement Information (Appendix A) for

Modelling Nationwide Spatial Variation of Ultrafine Particles based on Mobile Monitoring

Data Correction

Differences between the devices in the mobile platform and reference station were assessed by co-locating both CPCs 1-2 days before and after every two-week measurement. Devices were placed in a laboratory setting with all inlets connected to the same airstream. Concentrations were fluctuated between 0 and 100,000 particles/cm³ throughout the colocation measurement. We corrected each measurement period based on calculated median ratios between the two devices. All individual ratios are shown in table A1.

Table A1

Ratios for all individual co-location CPC measurements.

↓

Period	Median
5/6 Sep 2016	0.92
23/26 Sep 2016	0.88
13/14 Oct 2016	0.86
3/7 Nov 2016	0.91
25/28 Nov 2016	0.86
19/21 Dec 2016	0.92
20/23 Jan 2017	0.87
10/13 Feb 2017	0.84
3 Mar 2017	0.83
23/27 Mar 2017	0.79
4 Sep 2017	0.96
2 Oct 2017	0.80
23 Oct 2017	1.02
14 Nov 2017	0.93

Figure A1

Pearson correlation between UFP concentrations using spline of minimums with different temporal scales and traffic intensity in 50m and 1000m buffer.

↓

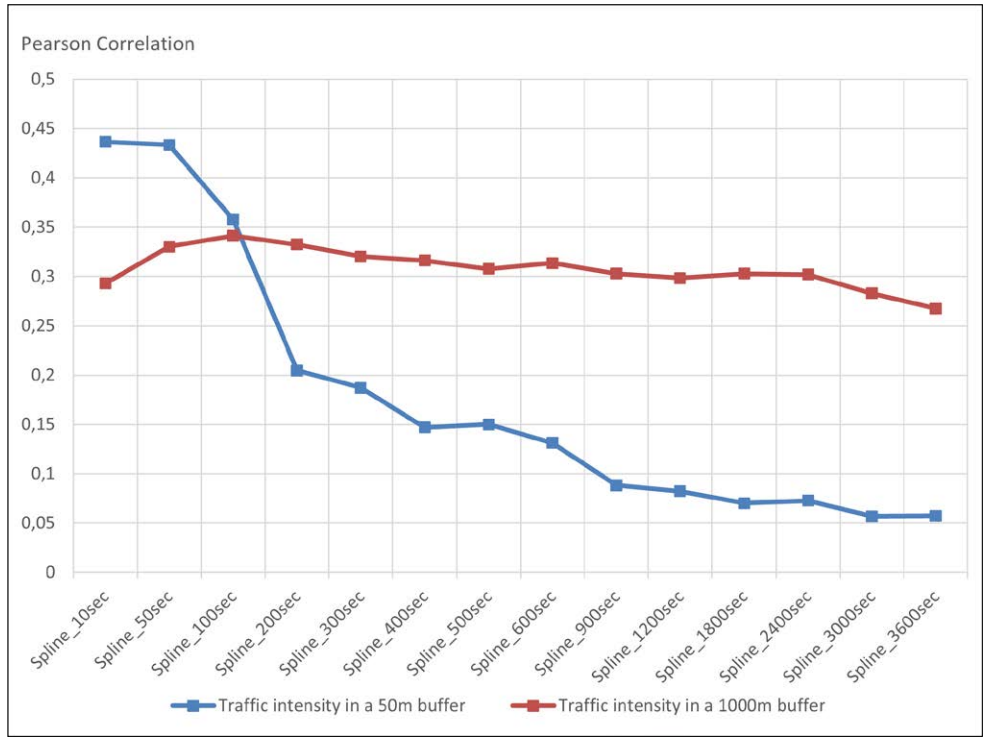


Table A2

Spatial predictor variables with units, a priori defined directions of effect and buffer sizes in the mobile data set.

↓

Predictor variable	Abbreviation	Units	Direction of effect	Buffer	10th percentile	Mean	90th Percentile
Industry area[1]	INDUS_	m²	+	100	0	2059	2566
				300	0	19602	79771
				500	0	57972	240714
				1000	0	257735	825276
				5000	2895308	5584176	8817632
Port area[1]	PORT_	m²	+	100	0	354	0
				300	0	3218	0
				500	0	9490	0
				1000	0	47621	0
				5000	0	2015034	8109209
Airport area[1]	AIR	m²	+	5000	0	449356	2243999
Natural and forested areas[1]	NATUR_	m²	-	100	0	431	0
				300	0	3777	0
				500	0	11555	0
				1000	0	63279	162659
				5000	0	3890813	8651169
Urban Green area[1]	URBG_	m²	-	100	0	800	0
				300	0	9773	27711
				500	0	34155	132889
				1000	0	185077	587273
				5000	1301143	5372652	10916938
Agricultural land area[1]	AGRI_	m²	-	100	0	1718	0
				300	0	17854	58772
				500	0	57144	214116
				1000	0	311041	1044973
				5000	3808456	25612763	49731108
Population density[2]	POP_	n	+	100	0	190	428
				300	134	1724	3627
				500	776	4570	9347
				1000	4481	16319	34311
				5000	50822	206647	430263
Household density[2]	HHOLD_	n	+	100	0	104	253
				300	52	930	2139
				500	321	2453	5549
				1000	1995	8726	19790
				5000	21599	106266	220792
Traffic intensity on nearest road[3]	TRAFNEAR	Veh/day	+			137	9986	22487
Traffic intensity on nearest major road[3]	TRAFMAJOR	Veh/day	+			3115	14212	28042
Heavy-duty traffic intensity on nearest road[3]	HTRAFNEAR	Veh/day	+			0	619	1333
Heavy-duty traffic intensity on nearest major road[3]	HTRAFMAJOR	Veh/day	+			54	980	2138
Road length of all roads[3]	RDL_	m	+	50	101	219	323
				100	399	687	983
				300	2895	5224	7312
				500	7893	13761	19090
				1000	29092	50344	68095
Road length of all major roads[3]	MRDL_	m	+	50	0	116	233
				100	0	285	633
				300	0	1532	3554
				500	0	3787	8800
				1000	3015	13627	31882
Traffic intensity on all roads (sum of (traffic intensity * length of all segments))[3]	TLOA_	(Veh/day)*m	+	50	52447	1274257	2703943
				100	233678	3216247	6627566
				300	4103995	17689365	35419436
				500	13741938	44180744	87625101
				1000	62971674	166892417	324383228

Predictor variable	Abbreviation	Units	Direction of effect	Buffer	10th percentile	Mean	90th Percentile
Traffic intensity on all major roads (sum of (traffic intensity* length of all segments))[3]	TMLOA_	(Veh/day)*m	+	50	0	1072539	2500053
				100	0	2581349	5817327
				300	0	12783795	26453420
				500	0	30877644	64537936
				1000	26948221	116664854	238567679
Heavy-duty traffic intensity on all roads (sum of (heavy-duty traffic intensity* length of all segments))[3]	HLOA_	(Veh/day)*m	+	50	626	79474	160304
				100	3636	199111	388710
				300	89031	1065469	2007866
				500	387896	2664654	6027990
				1000	2550331	10515869	25699398
Heavy-duty traffic intensity on major roads (sum of (heavy-duty traffic intensity*length of all segments)[3]	HMLOA_	(Veh/day)*m	+	50	0	71671	150032
				100	0	172420	339567
				300	0	846066	1614863
				500	0	2033997	4769469
				1000	966747	8018542	22978460
Highw[3]y4	HIGH_	m	+	100	0	15	0
				300	0	82	0
				500	0	207	888
				1000	0	919	3833
Restaurants[4]	REST_	n	+	100	0	1	1
				500	0	11	28
				1000	1	40	122
				5000	59	431	935
Traffic Lights[4]	TL_	n	+	100	0	1	3
				500	0	6	18
				1000	0	22	52
Bus Stops[4]	BS_	n	+	100	0	1	2
				500	2	7	14
				1000	10	26	43
X Coordinate	POINT_X		-		92798	150056	218109
Y Coordinate	POINT_Y		-		384891	463178	570277
Map produced by Kriging	Kriging Map	particles/cm[3]			4455	5678	6115

[1]source: CORINE (Copernicus Land Monitoring Service) 2018, [2]source: CBS (Central Bureau of Statistics Netherlands) 2017, [3]source: NWB (National Road Network Netherlands) 2017, [4]source: OSM (Open Street Map) 2017.

Figure A2

Map of average UFP regional background concentrations in the Netherlands.

↓

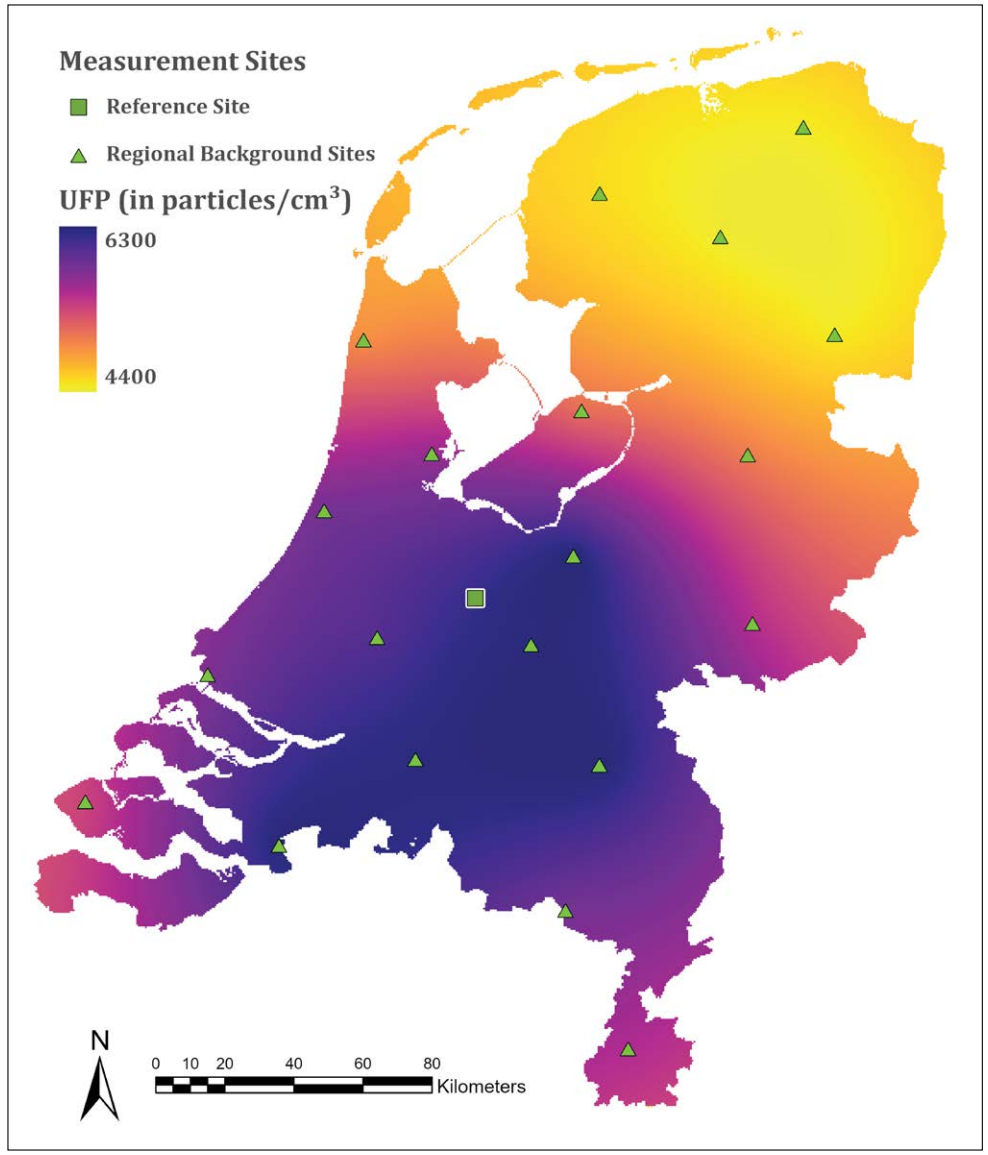
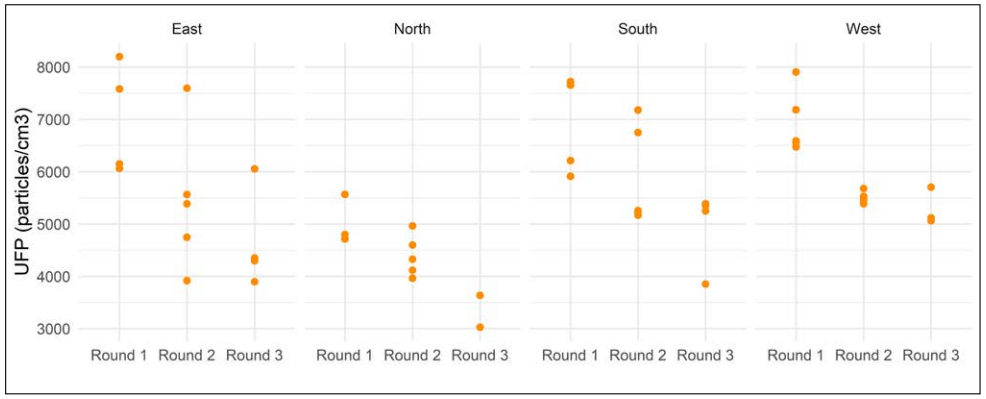


Figure A3

Distribution of temporally adjusted regional background UFP concentrations for the three measurement periods by region. Dots represent individual mean of a measurement site in a single round.

↓



Concentrations in particles/cm³

Table A3

Bias (predicted minus measured) of all models when compared to short-term measurements.

↓

	Stepwise	Deconvoluted Stepwise	LASSO	Deconvoluted LASSO	Random forest	Deconvoluted Randomforest
AMS-UTR	3,422	3,904	3,092	3,616	3,030	3,170
East	1,782	3,046	2,706	4,390	3,395	4,071
North	3,120	4,773	4,147	6,144	3,956	3,876
South	1,943	2,868	2,427	3,603	2,044	3,070
West	4,019	2,677	3,646	1,497	3,046	2,719
Traffic	3,786	3,439	3,529	3,122	3,386	3,607
Background	2,239	3,492	3,009	4,491	2,937	3,236

Table A4

Variation in predictions on 1000 random addresses across the Netherlands.

↓

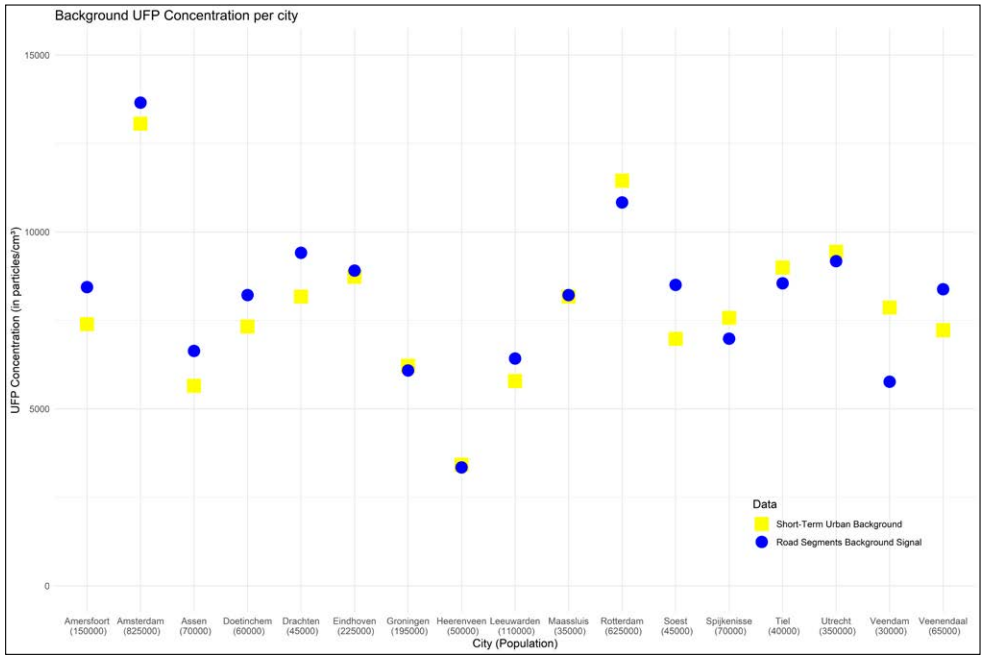
	Stepwise	Deconvoluted Stepwise	LASSO	Deconvoluted LASSO	Randomforest	Deconvoluted Randomforest
Minimum	7,951	8,631	9,932	10,753	9,278	7,076
25th percentile	8,532	9,643	10,376	12,621	10,451	10,783
Median	9,295	10,726	10,985	13,237	10,872	11,570
Mean	10,523	11,660	11,807	13,795	12,010	12,421
75th percentile	10,951	12,342	12,182	14,229	12,327	13,204
Maximum	30,799	28,226	26,746	24,955	26,501	26,119

Concentrations in particles/cm³

Figure A4

Background component of the UFP concentration measured with mobile monitoring compared with short-term urban background sites per city.

↓



Supplement Information (Appendix B) for

Modelling Nationwide Spatial Variation of Ultrafine Particles based on Mobile Monitoring

Table B1

Stepwise Regression Model

↓

Stepwise Regression

Variable	Estimate*	Std. Error	Pvalue
(Intercept)	7843	218	< 0.001
HHOLD_5000	4688	362	< 0.001
HTRAFNEAR	1499	364	4e−05
PORT_5000	4541	324.	< 0.001
TLOA_50	3606	742	< 0.001
TMLOA_50	2800	675.	3e−05
TRAFNEAR	2881	519	< 0.001

* Regression slopes and standard error multiplied by the difference between 10th and 90th percentile for all predictors.

Table B2

Deconvoluted Stepwise Regression Models

↓

Local

Variable	Estimate*	Std. Error	Pvalue
(Intercept)	2179	206	< 0.001
HHOLD_300	1854	285	< 0.001
TLOA_300	1164	291	< 0.001
TLOA_50	4334	626	< 0.001
TMLOA_50	2556	560	1e−05

Background

Variable	Estimate*	Std. Error	Pvalue
(Intercept)	238	269	0.376
POP_5000	3387	91	< 0.001
PORT_5000	2493	83	< 0.001
Kriging Map	1884	80	< 0.001

* Regression slopes and standard error multiplied by the difference between 10th and 90th percentile for all predictors.

Figure B1

↓

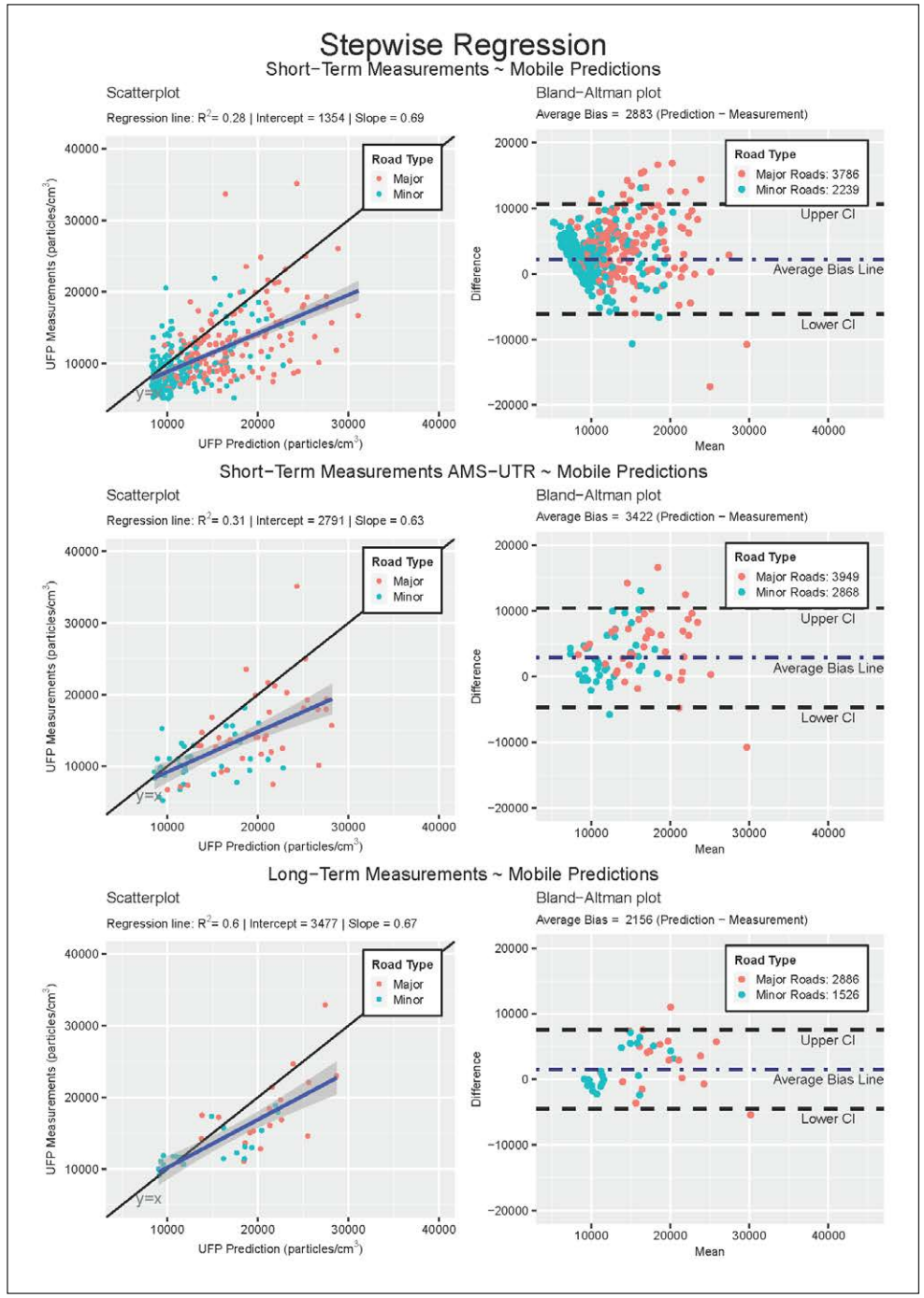


Figure B2

↓

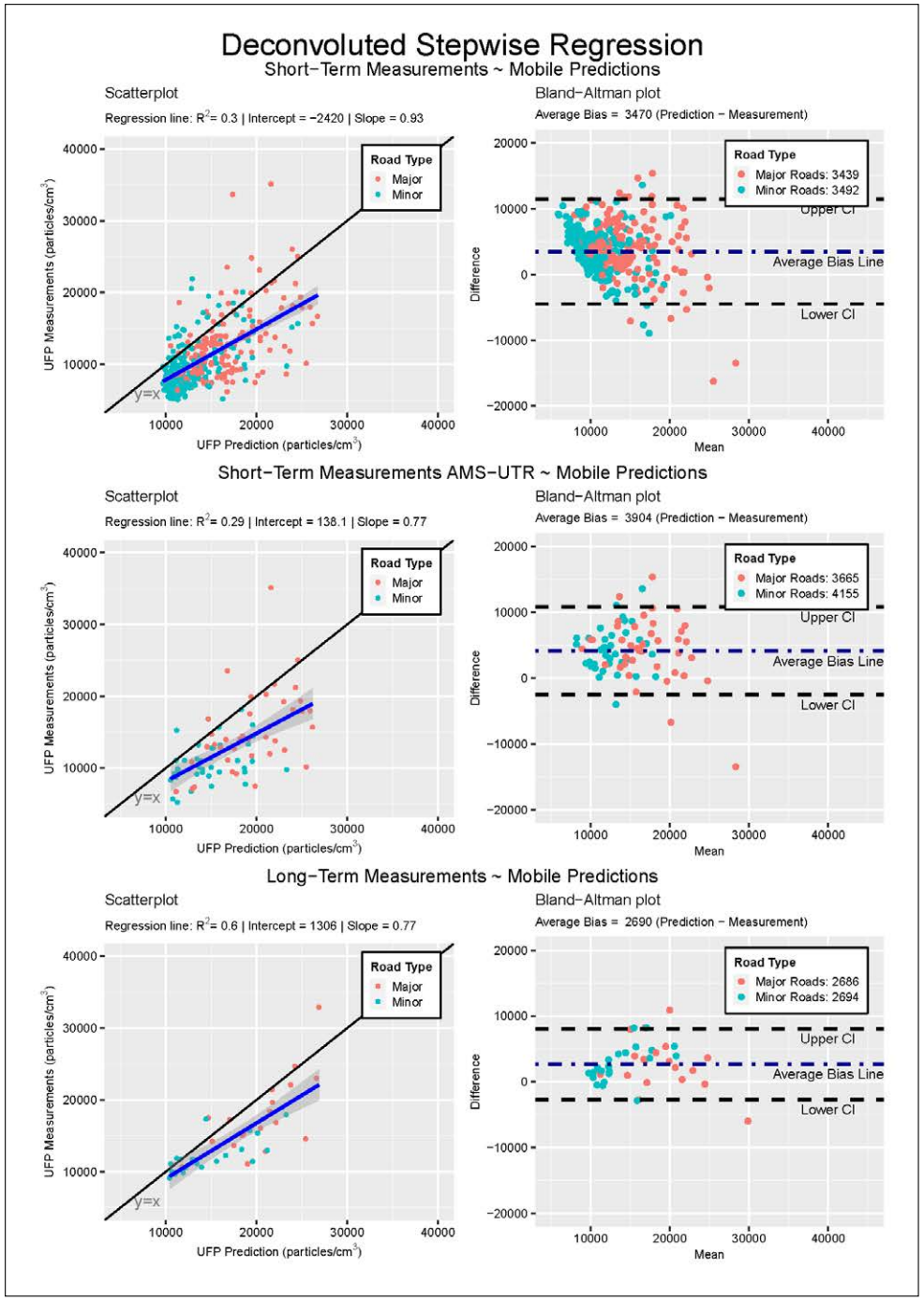


Table B3

LASSO Model

↓

LASSO

Variable	Estimate*
(Intercept)	10108
HHOLD_5000	3189
PORT_5000	2513
TLOA_100	191
TLOA_300	313
TLOA_50	2518
TMLOA_50	2400
TRAFNEAR	2642

* Regression slopes are multiplied by the difference between 10th and 90th percentile for all predictors. LASSO do not state standard errors or p-values.

Table B4

Deconvoluted LASSO Models

↓

Local

Variable	Estimate*
(Intercept)	4943
TLOA_50	4500

Background

Variable	Estimate*
(Intercept)	-797
HLOA_1000	880
INDUS_5000	−580
MRDL_1000	−723
HHOLD_5000	2623
HIGH_1000	-1036
PORT_5000	2988
Kriging Map	2235

* Regression slopes are multiplied by the difference between 10th and 90th percentile for all predictors.

Figure B3

↓

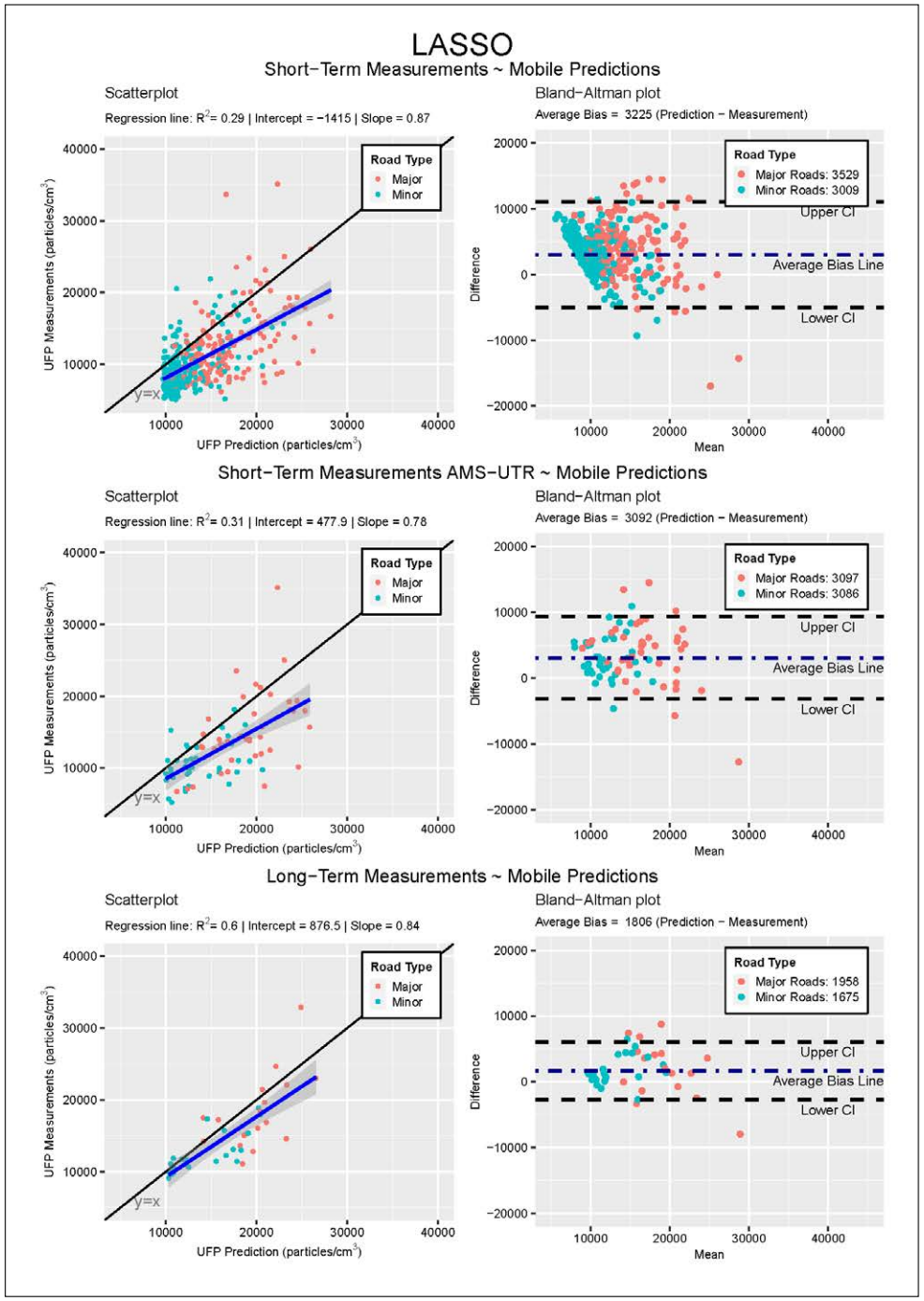


Figure B4

↓

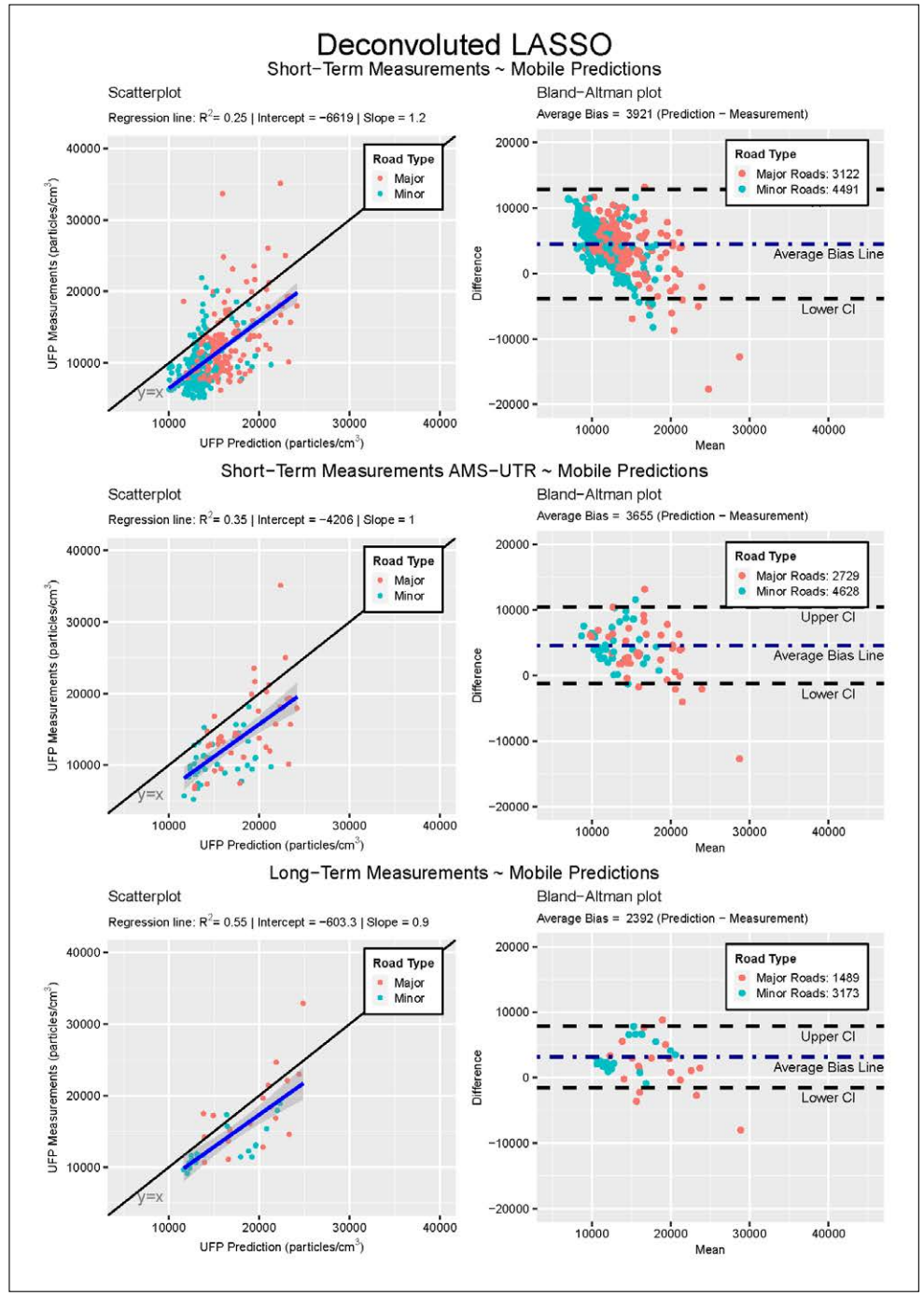
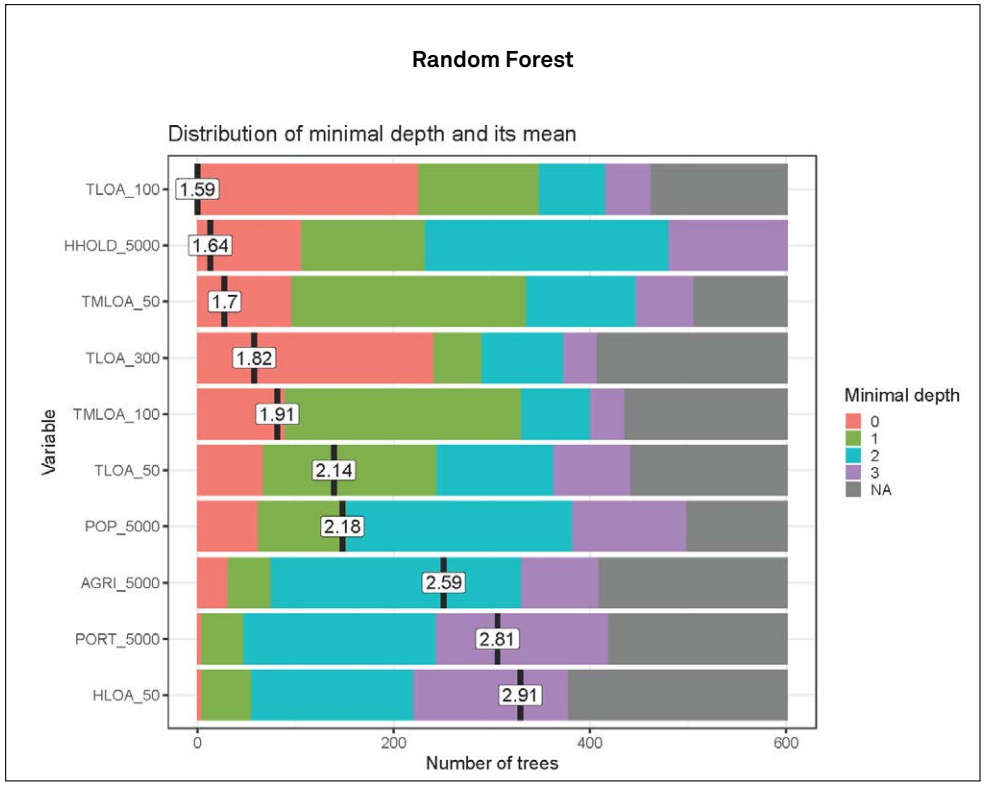


Figure B5

Random Forest Model

↓



Distribution of average tree depth of the top 10 predictors. In other words, the average position in a tree (first branch, second branch etc.) the variable is found. For example, when TLOA_100 is offered to a tree it is mostly selected first (red color) or second (green color), on average 1.59. Since every tree selects about 1/3 of the possible predictors every time, some predictors are offered more than others. The grey area (NA) means that the predictor was not offered.

Figure B6

Random Forest Model
Predictions

↓

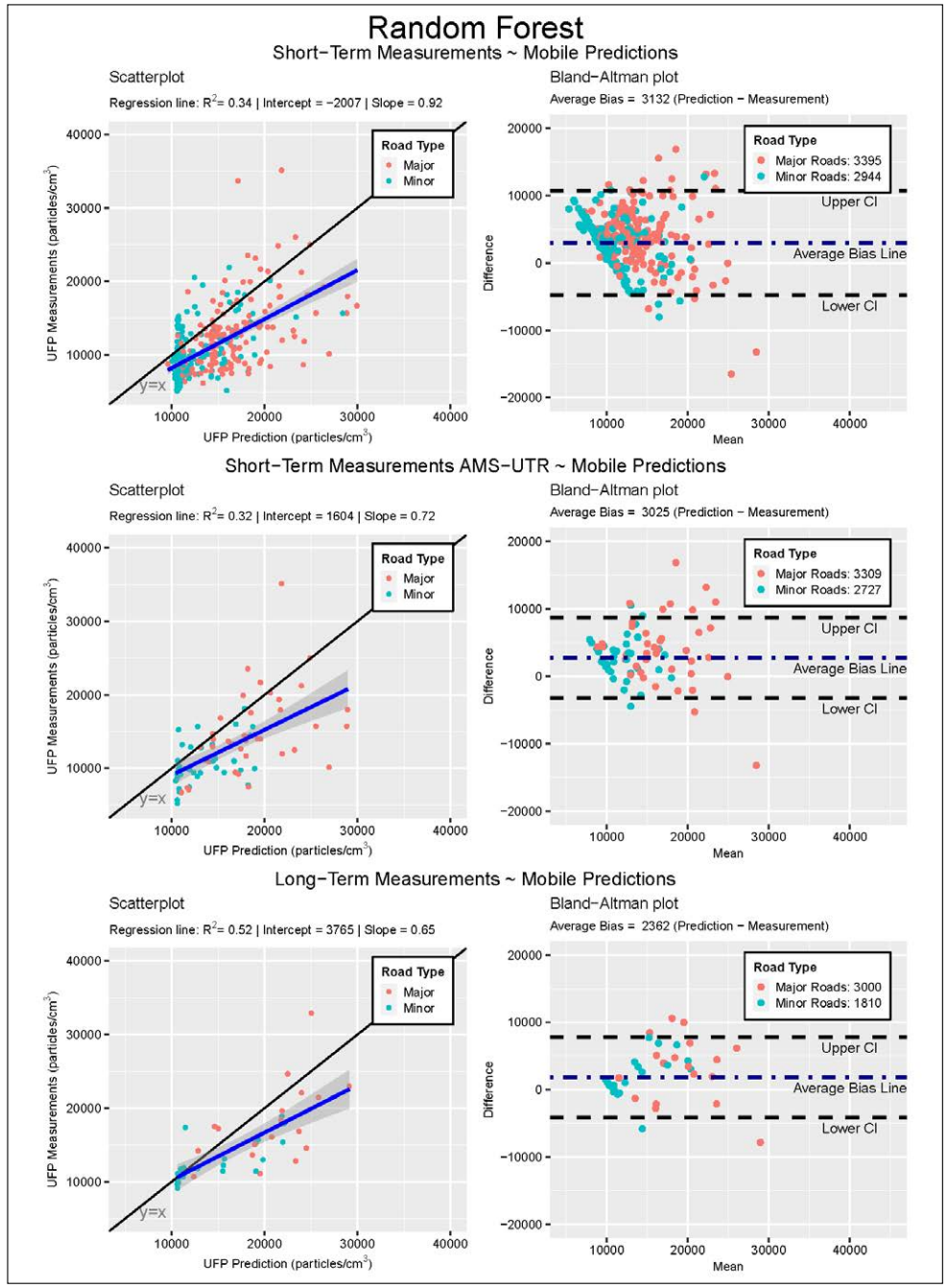
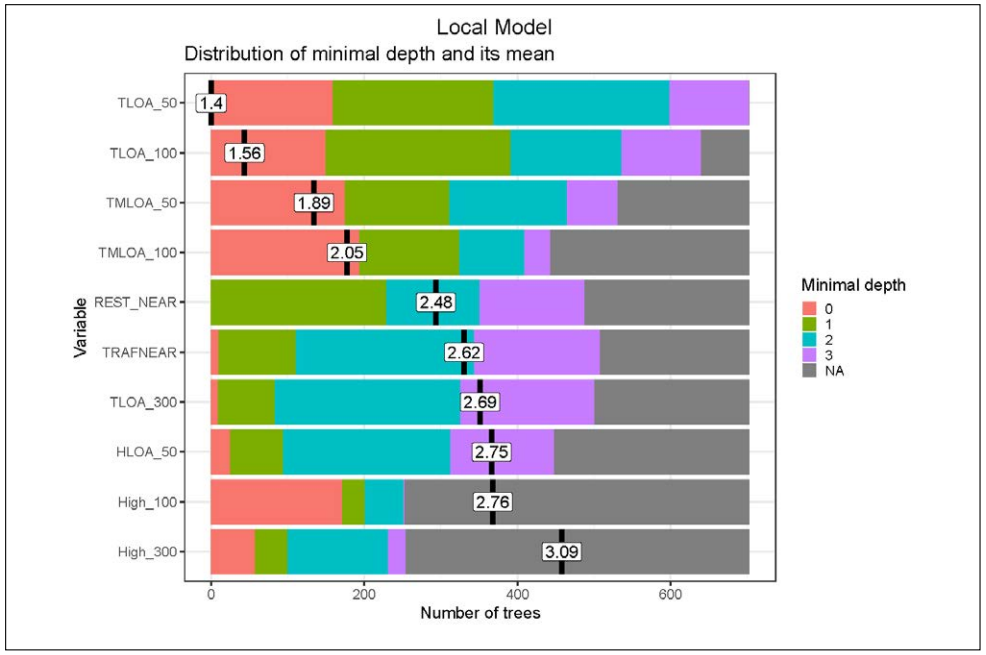


Figure B7

Deconvoluted Random Forest Model

↓



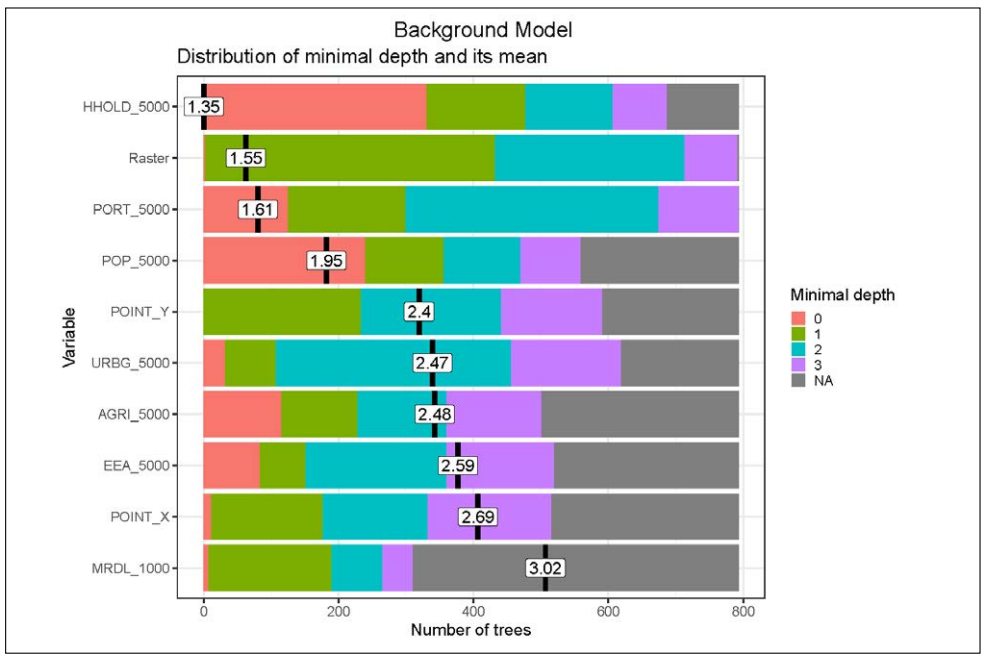
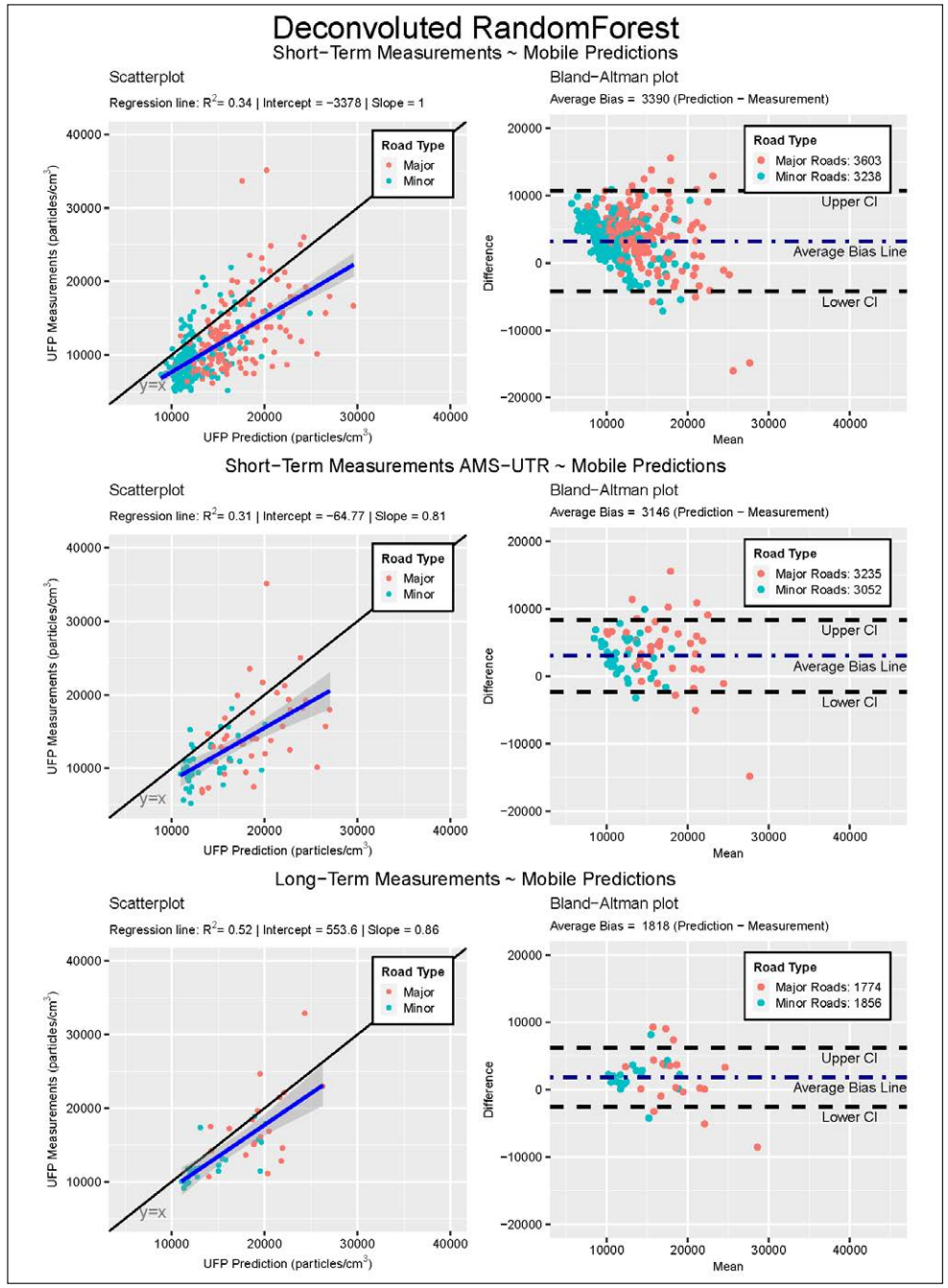


Figure B8

Deconvoluted Random Forest Model Predictions

↓



General Discussion

Throughout this dissertation I showed that mobile monitoring offers valuable insights into traffic-related air pollutants like UFP and BC. Despite the low performance of mobile models tested on mobile measurements, I showed that predictive performance increases substantially when these mobile models are tested on longer term average concentrations (chapter 2 and 3). In chapter 4, I showed that the choice for a prediction algorithm slightly influences the predictive performance. The last part of this dissertation showed that models based on mobile monitoring are scalable to a national extent.

Since all analyses refer to the ability of models to predict concentration levels, I start this general discussion with a brief explanation of performance statistics.

Then I answer the three main questions stated in the introduction (repeated below) and address several important findings separately in different sections. I refer to these specific aspects in the margin of the text. I also want to highlight that this discussion is purposely written in an accessible way, while preserving scientific depth.

1. Can we (and how can we) use mobile measurements to develop fine resolution spatial maps for long-term average UFP and BC concentrations?

2. Can we improve long-term average concentration maps by applying different prediction algorithms?

3. Can we use a mobile monitoring design for application in nation-wide cohorts, by accounting for regional background concentrations?

R^2

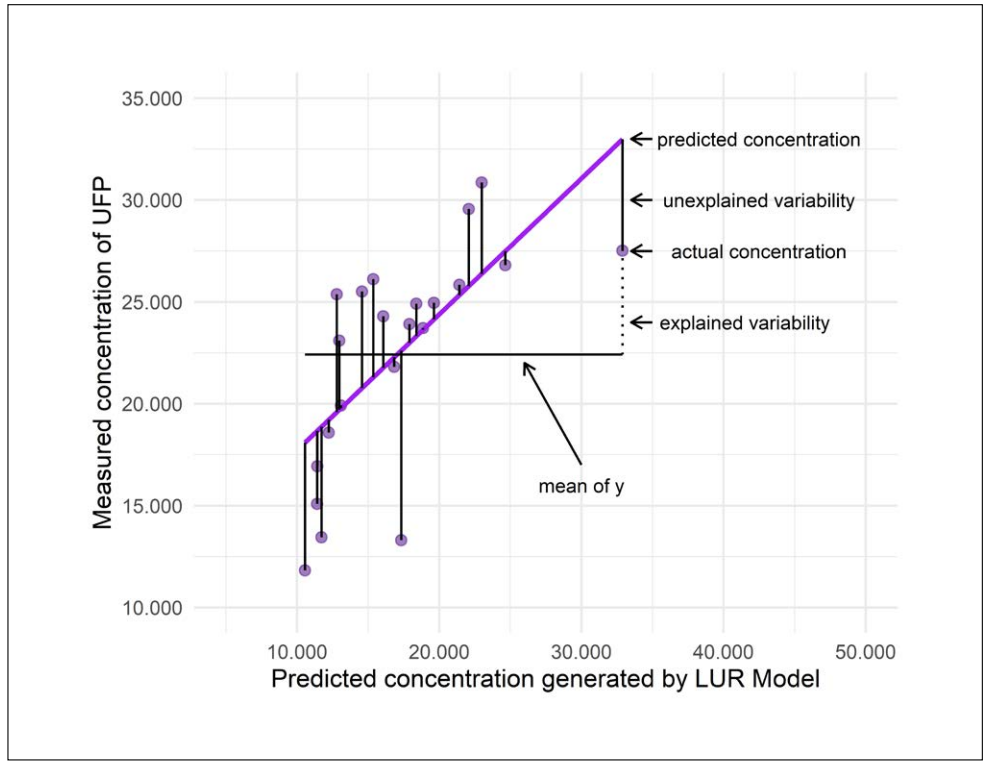
To analyse the spatial variation of ultrafine particles (UFP) and black carbon (BC), we developed land-use regression (LUR) models. LUR models are (linear) regression models with the pollutant (in this case UFP or BC) as dependent variable and different land use characteristics (land-use for e.g., industry, port, population density, traffic intensity, etc) as possible predictor variables (independent variables). LUR models often include a combination of such predictor variables and are typically assessed by their goodness-of-fit (R^2 values; Figure 1) and amount of bias (Figure 2). Throughout this dissertation we mainly focus on these criteria regarding performance of our prediction algorithms.

Figure 1

Explanation of R^2 values in LUR models.

↓

R^2 is the percentage of the total variation of the dependent variable a (linear) model can explain. Without a model, the best prediction would be to take the mean concentration. The R^2 value represents all extra variation the model can predict.



Bias

The bias refers to average difference between the predicted and measured concentration. In the figure below, most points are above the y=x line, indicating that predicted concentration are on average higher than the measured concentrations.

Figure 2

Explanation of bias in LUR models (predictions systematically different from measurements).

↓

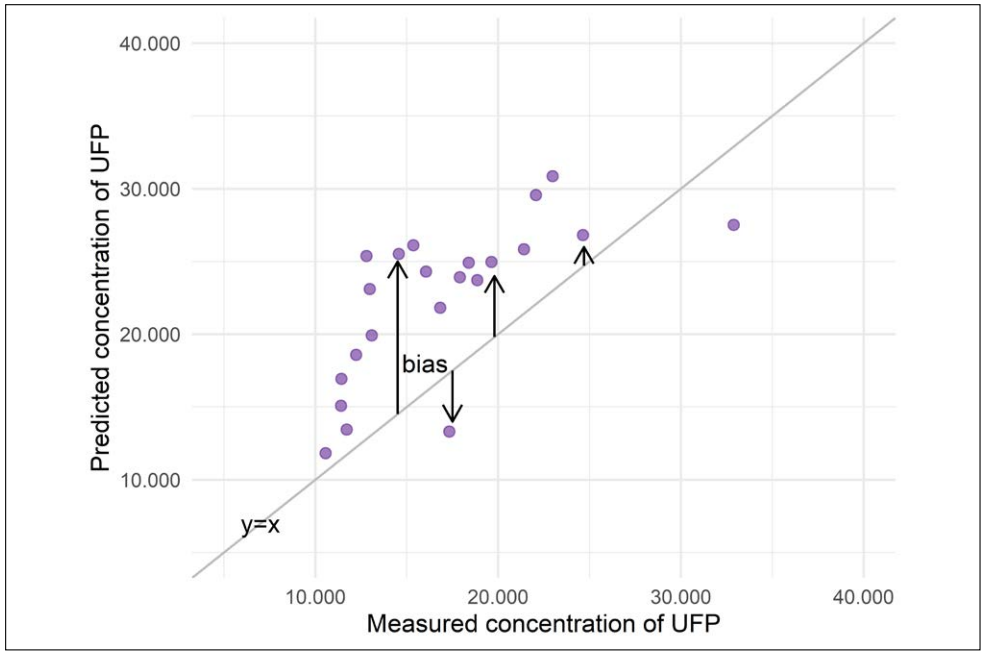


Table 1

Overview of measurement campaigns used in this dissertation.

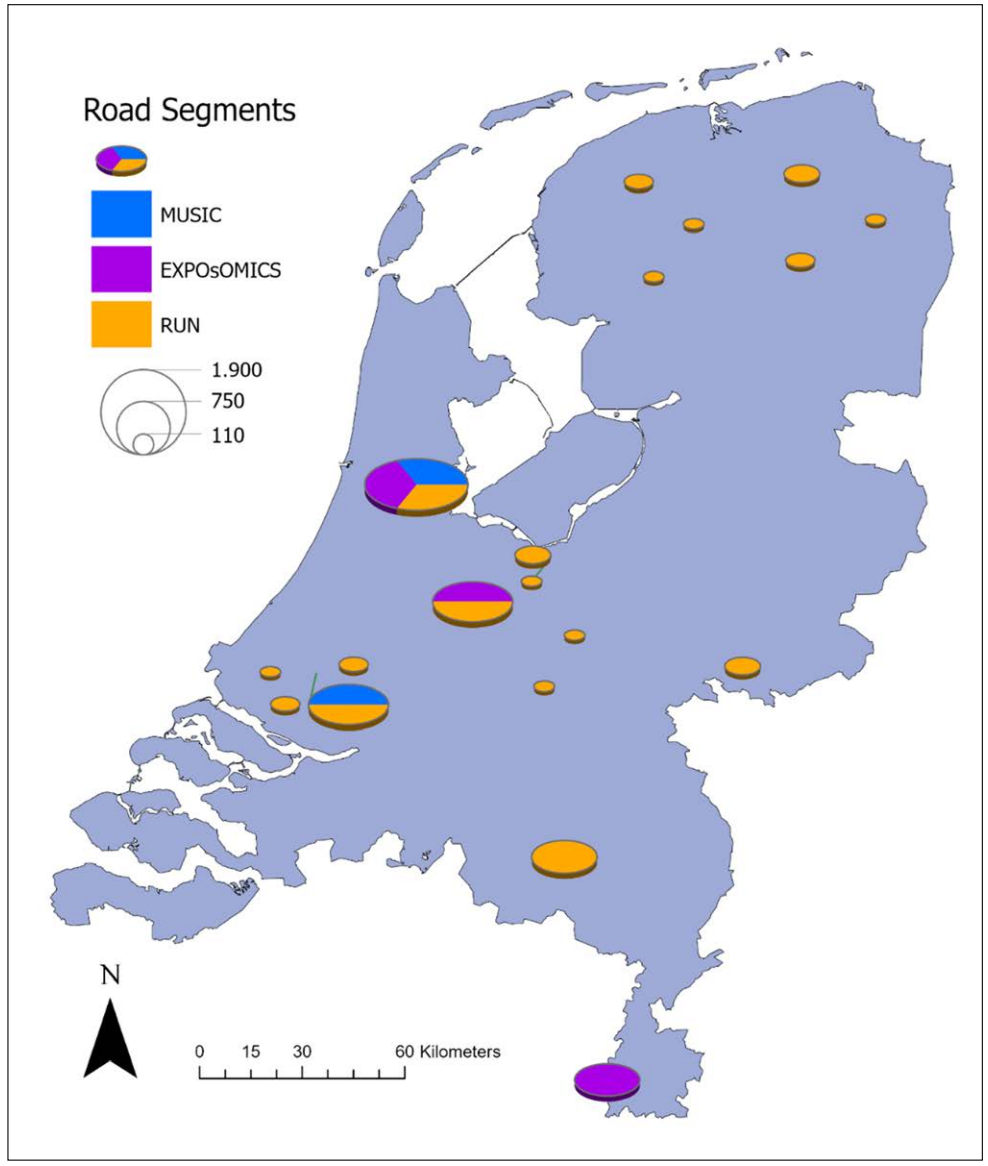
↓

Campaign	Year	Location	Design	Sites	Resolution	Reference
MUSIC	2013	Amsterdam, Rotterdam	Short-term	160	3 x 30 min	Montagne et al.[1]
			Mobile	2,964	~20 seconds	Kerckhoffs et al.[2]
EXPOsOMICS	2014-2015	Amsterdam, Utrecht & Maastricht	Short-term	240	3 x 30 min	van Nunen et al.[3]
			Mobile	5,236	~25 seconds	Kerckhoffs et al.[4]
			Residential	42	3 x 24 hour	van Nunen et al.[3]
RUN	2016-2017	20 cities and towns in the Netherlands	Short-term	400	3 x 30 min	Kerckhoffs et al.
			Mobile	14,936	~40 seconds	Kerckhoffs et al.
			Regional	20	3 x 14 days	Kerckhoffs et al.

Figure 3

Overview of mobile monitoring campaigns in the Netherlands

↓



① Can we (and how can we) use mobile measurements to develop fine resolution spatial maps for long-term average UFP and BC concentrations?

In chapter 2 we compared land-use regression models based on short-term and mobile monitoring, in two major cities in the Netherlands (Amsterdam and Rotterdam, the MUSiC campaign). We showed that, even though the mobile LUR model only explains 13% (UFP) and 12% (BC) of the variation of the mobile measurements and the short-term stationary model explains 36% (UFP) and 28% (BC) of the variation of the short-term stationary measurements, LUR models based upon mobile and short-term stationary monitoring resulted in highly correlated concentrations of both UFP and BC on address level (Figure 4). We also noticed that mobile model predictions are on average 1.41 (UFP) and 1.91 (BC) times higher than short-term stationary model predictions. This can be explained by the fact that mobile measurements are collected on-road, while short-term measurements were collected on the side of the road.

Poor R^2 explained in section B.

Bias in mobile models described in section G.

In chapter 3 we wanted to understand whether these conclusions hold on measured long-term average UFP and BC concentrations. Therefore, 42 sites in Amsterdam and Utrecht were measured for 3 times 24 hours to represent long-term home outdoor concentration and used as test sites. Like chapter 2, LUR models were based on mobile and short-term stationary measurements in major cities in the Netherlands, but now in Amsterdam, Maastricht, and Utrecht (EXPOsOMICS campaign). The resulting mobile model, like the mobile model in chapter 2, explained variation in mobile measurements relatively poorly (R^2 = 0.15). Despite the poor model R^2, the ability of mobile UFP models to predict measurements with longer averaging time increased substantially to 36% for short-term stationary measurements and 57% for long-term home outdoor measurements (Figure 5). Furthermore, the mobile model for UFP was stable over different settings as the model predicted concentration levels highly correlated (r = 0.76) to predictions made by the LUR model from chapter 2. In contrast, the mobile BC model only predicted 14% of the variation of long-term home outdoor measurements.

Poor R^2 explained in section B.

Robustness over time described in section C.

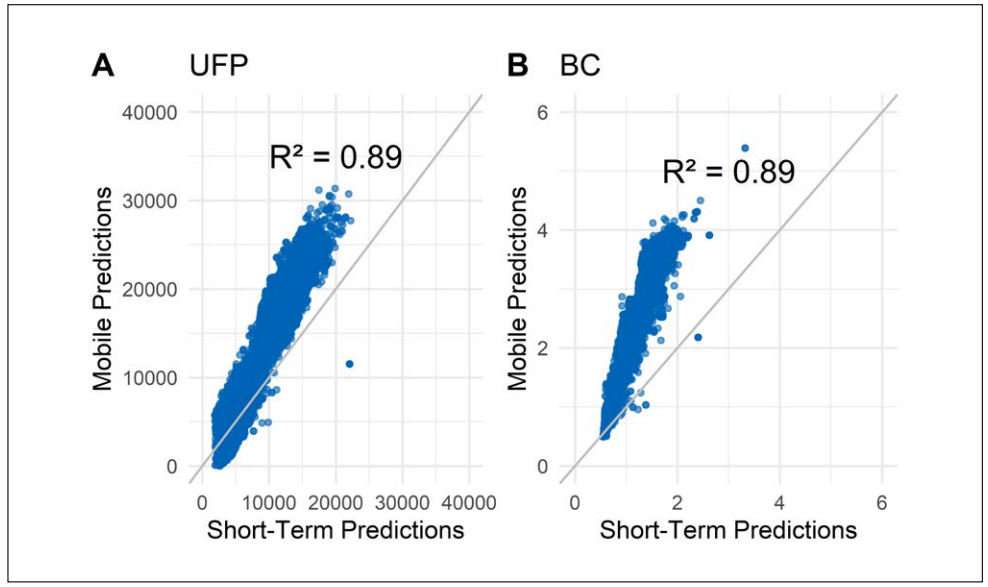
In summary, with mobile monitoring and adequate measurement equipment robust models of residential long-term average exposure can be developed.

BC models described in section A.

Figure 4

Comparison of UFP (top) and BC (bottom) predictions from short-term stationary and mobile LUR models on 12,682 addresses in Amsterdam (MUSiC campaign).

↓

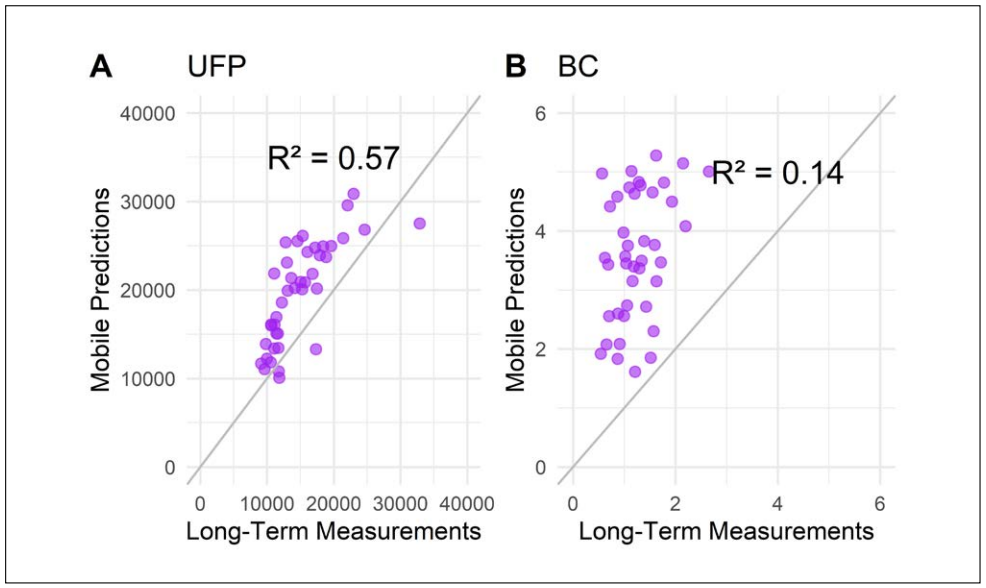


UFP in particles/cm³ and BC in µg/m³

Figure 5

UFP (top) and BC (bottom) predictions from mobile LUR models (EXPOsOMICS campaign) compared with long-term stationary average concentrations in Amsterdam and Utrecht (n=42).

↓



UFP in particles/cm³ and BC in µg/m³.

② Can we improve long-term average concentration maps by applying different prediction algorithms?

Based on chapter 2 and 3, we concluded that it is possible to make valid and robust spatial maps for UFP based on mobile measurements. These models were all based on linear regression. In chapter 4, we explored if other prediction algorithms could improve on the spatial maps for UFP, both for short-term stationary and mobile monitoring. We found that the choice for model algorithm only slightly influences model predictions, both for short-term stationary and mobile sampling (Figure 6). We also observed a small overestimation of the long-term measurements for most model algorithms, similar to what was observed in chapter 2 and 3 for the linear regression method.

Overestimation of mobile models described in section G.

Machine learning algorithms trained on mobile measurements explained 38–47% of external UFP concentrations, whereas multivariable methods like stepwise regression and elastic net explained 56–62%. Some machine learning algorithms (bagging, random forest) trained on short-term stationary measurements explained modestly more variability of external UFP concentrations compared to multiple linear regression and regularized regression techniques. The short-term stationary measurements are an average of 3 times 30 minutes and have less of the inherent noise found in mobile measurements. Machine learning algorithms try to explain this ‘noise’ while this does not reflect long-term stationary concentrations. This could be a reason that some of the data mining techniques outperformed ordinary linear and regularized regression techniques in the short-term stationary data set. Adding other (non-linear) predictor variables to models could change the performance of machine learning techniques. Linear approaches will have more and more difficulties to take into account the inter-variables relationships when the complexity of the data increases.

Machine learning further explained in section D.

Further explained in section B

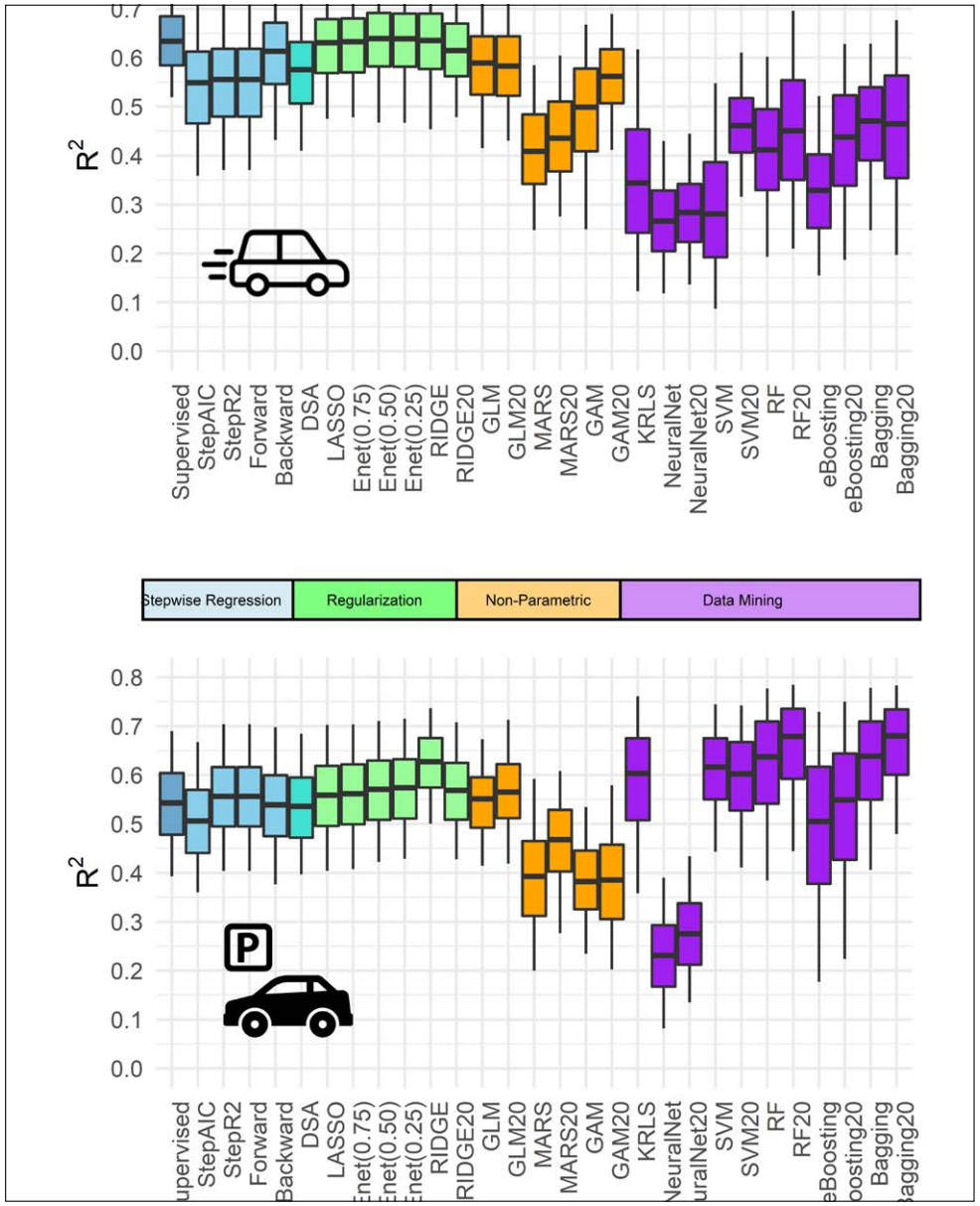
However, we found that higher training model R^2 did not automatically equate to higher R^2 in external measurements, stressing the importance of an external validation data set. This is especially true when transferring models fitted on different timescales.

In summary, only modest improvement in prediction was achieved by alternative algorithms.

Figure 6

Distribution of boot-strapped R^2 values when predictions of mobile models (top) and short-term models (bottom) are compared to long-term measurements.

↓



③ Can we use a mobile monitoring design for application in national cohorts, by accounting for regional background concentrations?

For chapter 5 and 6, we combined short-term stationary and mobile monitoring with long-term regional background measurements across the Netherlands. Up until now, measurements to UFP had only been done in a restricted number of large cities, while a large proportion of people in the Netherlands does not live in the major cities monitored so far.

First, in chapter 5 we examined the regional differences in UFP concentrations in the Netherlands and found a clear difference between concentrations in the north and the more populated southern parts of the country (Figure 7), suggesting that cities equal in local and urban sources but located in different regions of the country experience different concentrations due to variation in regional background concentration. We found spatial correlation with a kriging approach in regional background UFP concentrations up to about 180 km. Average temporal correlation between 2-hour and 24-hour average UFP at sampling sites spread across the Netherlands and the reference location was 0.50 and 0.58, respectively. Cross-correlation yielded similar results with a median average cross-correlation of 0.54 with time lags often below 2 hours.

Spatiotemporal variation of UFP in the Netherlands described in section E.

Second, in chapter 6, we combined all efforts for the creation of nation-wide land-use regression models. In one of the largest mobile monitoring campaigns to date, we developed robust nation-wide models for long-term UFP exposure, with minor differences in predictive performance between three different algorithms (stepwise, LASSO, random forest). All three algorithms included the regional background kriging map, developed in chapter 5 as important predictor. Correlations between predictions at random addresses across the Netherlands were high with Pearson correlations between 0.84 and 0.99. Mobile models overestimated exposure by about 20-30% in all cases, with small differences between regions and road types.

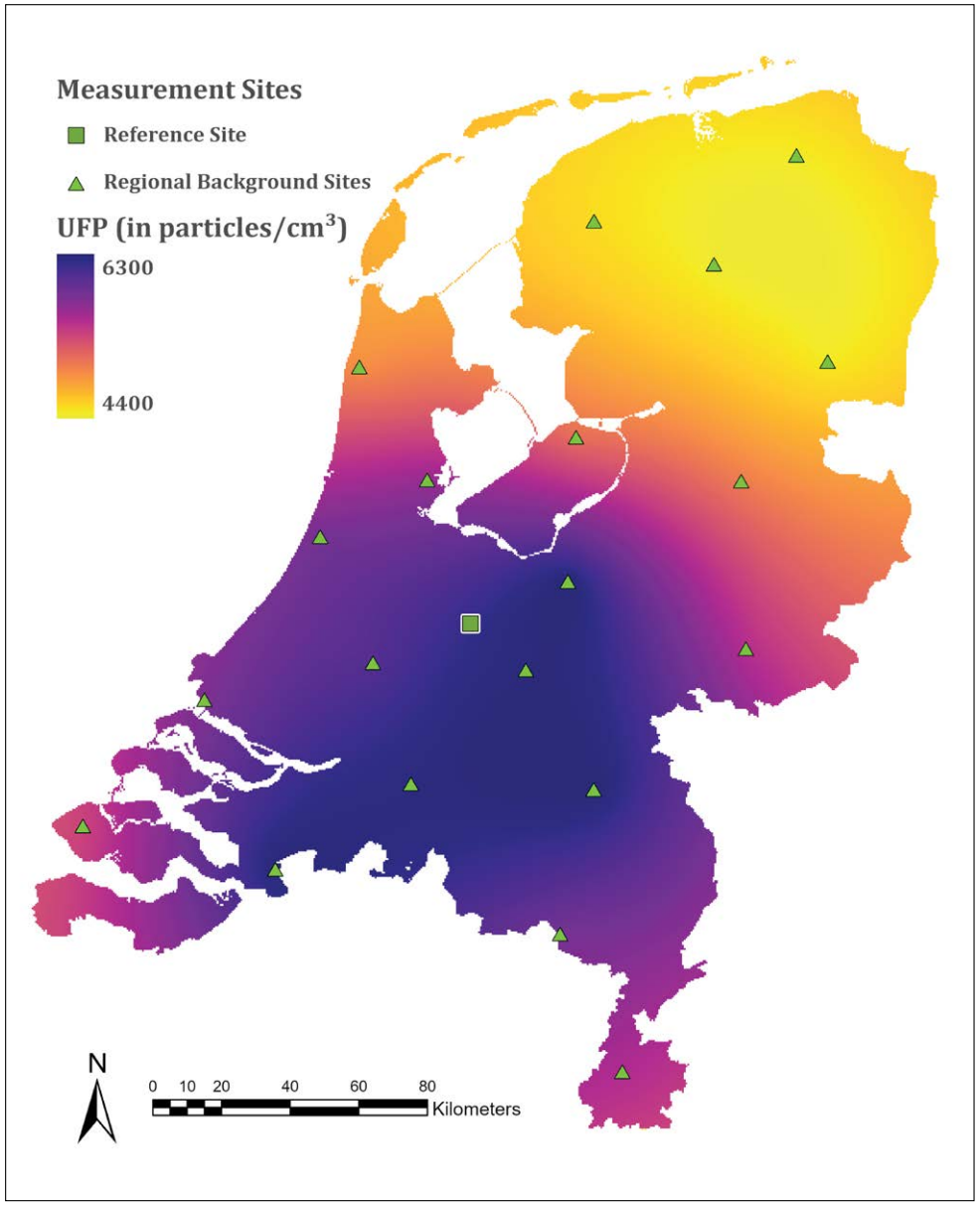
Overestimation of mobile models described in section G.

In summary, regional information on UFP concentrations adds to the current knowledge of spatial variability of UFP, making LUR models scalable to a national extent.

Figure 7

Map of average UFP concentrations across the Netherlands, created with universal kriging.

↓



Section A: Black Carbon LUR models

In this dissertation we developed LUR models based on mobile and short-term stationary measurements of UFP in all three campaigns. For BC, we only developed LUR models based on mobile and short-term stationary observations during the first two campaigns. Although BC predictions made from mobile models correlated highly with short-term stationary predictions in the MUSiC campaign (chapter 2), mobile BC models were not able to predict long-term exposure in the EXPOsOMICS campaign (chapter 3). We therefore concluded that mobile BC models should not be used with the instrumentation used in our studies (micro-Aethalometer AE51). Here, I want to explain why.

Why did the micro-Aethalometer not work well in our setup?

The BC instrument that was used in the mobile monitoring campaigns was set to a 1-minute time resolution, because decreasing the time resolution generates too much noise in Dutch outdoor exposure conditions [5,6]. Hagler et al.[5] performed several experiments with micro-Aethalometers and found noise up to 3.2 µg/m³ and 13% negative values in a mobile set-up with 1-second time resolution and average BC concentration of 3.9 µg/m³. Average concentrations in our studies were about 1-2 µg/m³.

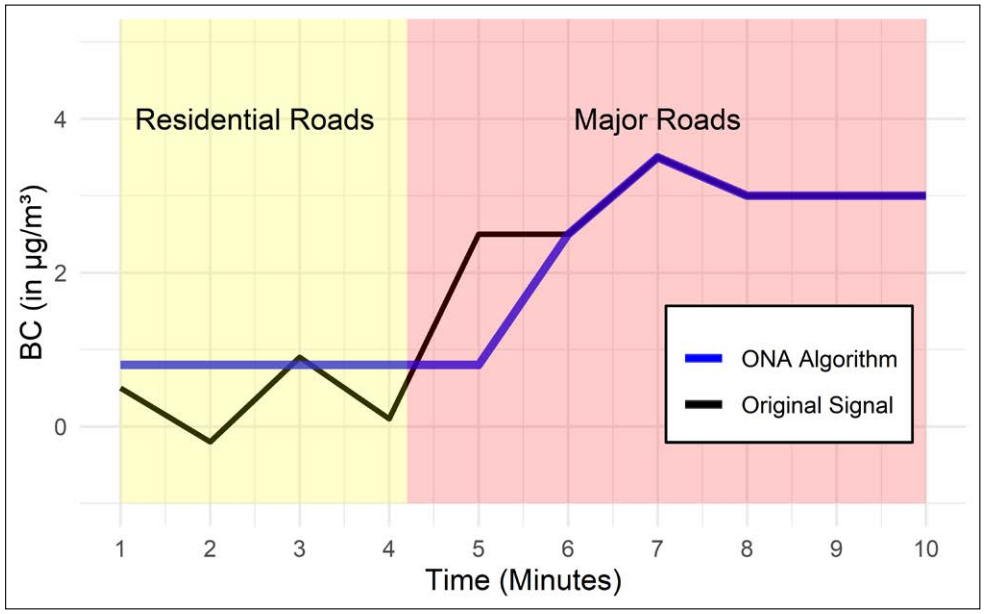
To reduce the noise, Hagler et al.[5] developed an algorithm (Optimized Noise-Reduction Algorithm; ONA) to increase the time resolution where necessary. In short, this method depends on the increasing level of blackness of the filter in the device and if the increase in blackness in a certain time frame is not enough to distinguish measurements from noise, the time frame is extended until that criterium is met.

In our studies, the 1-minute resolution we set for the BC instrument often had to be extended to 2 or 3 minutes by the noise reduction algorithm. While in busy streets it is often possible to keep the 1-minute time resolution, concentrations were averaged over multiple minutes when driving on residential streets. This also means that several times, the criterium was only met when driving on busy streets, averaging the concentration measured on that busy street with the previous residential streets. Some residential streets were therefore assigned higher concentrations, whereas some busy streets were assigned lower concentrations (example shown in figure A.1). It also meant that no distinction could be made between different residential streets. In conclusion, the time resolution of the AE51 was insufficient to capture the high variability of BC concentrations in urban areas.

Figure A.1

Example of Original BC signal versus ONA algorithm.

↓



How to set up BC measurements with mobile monitoring?

Increasing the sampling time on sites/segments reduces the need to average concentrations over multiple segments. Table A.1 shows that BC LUR models based on mobile monitoring is possible. In fact, the first short-term/mobile modelling study of black carbon was a mobile campaign. Larson et al.[7] were able to develop a black carbon model in British Columbia, based on 39 locations. These 39 locations were all intersections driven in a cloverleaf pattern. Traverse times ranged from 5 to 13 minutes. This is similar to averaging times in some short-term stationary measurement campaigns, like Weichenthal et al.[8], who measured BC for 10-20 minutes at each site. Hankey and Marshall[6] used bicycles and a micro-aethalometer (like our studies) and aggregated data over multiple spatial distances. They stated that the 1-minute time resolution was sufficient to develop their LUR models.

Another option is to increase the number of repeats on every segment. Van den Bossche et al.[9] used the micro-aethalometer as well and was able to develop robust LUR models by aggregating 1-second time resolution measurements over 50 meters while walking. Only segments with at least 5 repeats were used to develop their models.

Study by Messier et al further explained in section F.

Last option is to use devices that have less noise and can therefore have smaller time resolutions. Messier et al.[10] developed BC models based on mobile monitoring with a lab-grade dual-spot Aethalometer (Magee Scientific model AE33), measuring at a 1-second resolution, though averaging concentrations over a 30-meter road segment (typically 3-10 seconds). While they drove the same segments numerous times, only a small increase in performance of the LUR models compared to a single drive was found. The same instrument was used by Shairsingh et al.[11] who were also able to develop robust LUR models. In a true mobile setup, it appears to be necessary to use instruments that provide stable readings at a 1 to 5-second time resolution.

Table A.1

Overview of BC LUR models.

↓

Reference	Design	City	Design	Instrument	Number of Sites/Segments	Temporal resolution	Model Structure	Model R^2
Larson et al. 2009	Spatial	Vancouver	Mobile (Car)	PSAP	39 cloverleaf patterns	10-sec	Stepwise Linear Regression	0.68 (LOO: 0.56)
Dons et al. 2013	Spatio-temporal	Flanders	Short-Term	AE51	63 sites	24hours	Stepwise Linear Regression	0.77
Saraswat et al. 2013	Spatio-temporal	New Delhi	Short-Term	AE51	48 sites	1-3 hours	Backward Grouped Regression	Morning: 0.86, Afternoon: 0.69
Weichenthal et al. 2014	Spatial	Montreal	Short-Term	AE51	73 sites	10-20 min	Random Effect Models and Bayes factors	Percent change per IQR for every variable
Montagne et al. 2015	Spatial	Amsterdam & Rotterdam	Short-Term	AE51	160 sites	3 visits	Supervised Stepwise Regression	0.35
Kerckhoffs et al. 2016	Spatial	Amsterdam & Rotterdam	Mobile (Car)	AE51	2,964 segments (100m)	2 or 3 visits	Supervised Stepwise Regression	0.11
Kerckhoffs et al. 2017	Spatial	Amsterdam, Maastricht & Utrecht	Mobile (Car)	AE51	5,236 segments (100m)	2 or 3 visits	Supervised Stepwise Regression	0.12
van den Bossche et al. 2018	Spatial	Antwerp	Mobile (Bicycle and walking)	AE51	1,457 segments (50m) (with at least 5 days)	IQR: 9 to 27 times	Supervised Stepwise Regression	0.43
Messier et al. 2018	Spatial	Oakland	Mobile (Car)	AE33	19,149 segments (30m)	1 to 50 drive days	Stepwise Regression with Kriging	around 0.6
Liu et al. 2019	Spatio-temporal	Shanghai	Mobile (Car)	AE51	116 segments (1000m)	3 visits	Stepwise Linear Regression	0.68 (LOOCV: 0.66)
Shairsingh et al. 2019	Spatial and Spatio-temporal	Toronto	Mobile (Car)	AE33	5,224 segments (50m)	2 or 3 visits	Deconvoluted regression	0.32

*PSAP: particle soot absorption photometer, AE51: micro-Aethalometer, AE33: Aethalometer (reference-grade).

Section B: Poor model R² of models based on mobile measurements.

One important finding in all our campaigns is that low R² values for mobile monitoring LUR models can result in high R² values when applied to long-term stationary measurements.

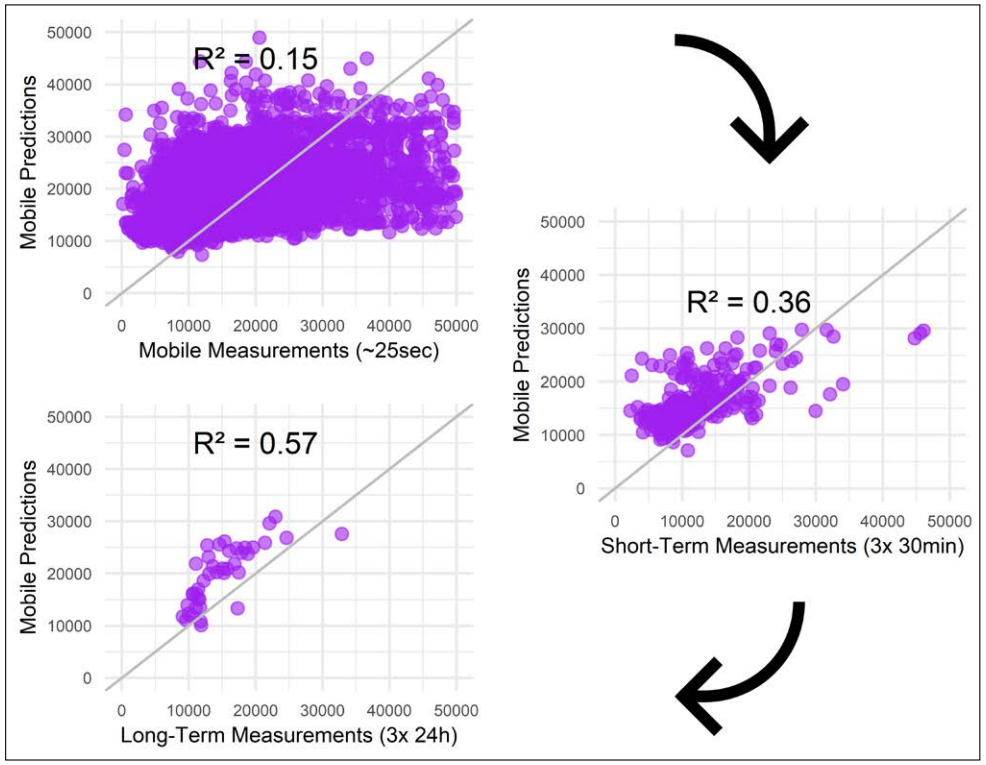
How is that possible?

In chapter 2 and 3 we showed that an increase in the averaging time of measurements led to an increase in the ability of mobile models to predict these measurements; from 13% and 15% for mobile measurements (median 25 s), 28% and 36% for short- term stationary measurements (3 × 30 min) in the MUSiC and EXPOsOMICS study respectively and 57% for home outdoor measurements (3 × 24 h) in the EXPOsOMICS study (Figure B.1).

Figure B.1

LUR model predictions based on mobile monitoring compared to mobile, short-term, and long-term stationary measurements.

↓



Mobile measurements consist of a few seconds per road segment and are therefore very variable. It would be difficult for any model to predict these concentrations accurately, mainly because we do not offer predictors on this temporal resolution. In other words, the variation in measurements can be divided into a within-day variation, between-days variation and a between-segments spatial variation. For the first two, we did not offer possible predictors able to explain this variation in our studies. And because the goal is to assess long-term exposure, we only assess the spatial variation (between-segments variation) and do not need to explain the within-day and between-days variation. The total variation our models are able to explain is therefore restricted to the spatial variation, which explains the low R² values in mobile monitoring campaigns. For example, our short-term models only explained 13% of the variation in the mobile measurements. Other algorithms than the supervised stepwise models used in chapter 2 and 3 also predicted mobile measurements in the range of 12 to 24% (chapter 4). R² values depend more on the time resolution of the test data set, rather than the time resolution of the training data. This also means that it is possible to predict long-term exposure with LUR models based on short-term stationary or mobile measurements [1].

Further explained in section F.

Increasing the accuracy of individual measurements increases R² values of LUR models. This can be done by increasing sampling time or increasing the number of repeats at each road segment. This does not necessarily mean that more robust models will be created.

Overview of UFP LUR models in table B.1

Long-term stationary measurements were used to develop LUR models in Amsterdam[12], Augsburg[13] and Rome[14] and resulted in model R² values of 0.67, 0.92 and 0.71 respectively. Since these models are developed and tested on a limited number of sites, performance statistics can be overestimated.

Studies that have repeated mobile monitoring at the same road segment more often (up to 15 times) than in our studies have reported fairly high model and validation R² values [6,15–18]. These studies have a total of at least 5 minutes of data for every road segment, measured at least 5 times (but often much more). Hatzopoulou et al.[15] compared LUR models developed on road segments with at least 3 visits to segments with at least 16 visits and found that road segments with at least 16 observations achieved a higher adjusted model R² with fewer explanatory variables compared to the model developed with segments having 3+ visits. Ghassoun et al.[19] created LUR models based on short-term stationary sampling (3 minutes), and repeated that 7 times in the summer and 8 times in winter and found R² values of 0.74 and 0.85, respectively. Exception is the LUR model based on short-term stationary measurements (3 x 30 minutes on 215 sites)

developed by Jones et al.[20], with a model R^2 of 0.66.

The reason why we were able to develop robust mobile models is mainly due to similarity between different road segments with comparable characteristics. These similar road segments with comparable characteristics can be viewed as pseudo repeats when it comes to developing LUR models. As stated above, LUR models with high R^2 values can be achieved by repeating measurements multiple times for every segment. For the development of a LUR model, it is therefore not necessary to measure all road segments in a city for example. The total number of segments can be reduced significantly, if the coverage and variation in road network and other predictors is preserved.

Several mobile monitoring studies created LUR models based on subsets of the data and found that the number of road segments for model development can often be reduced by a great amount. Hatzopoulou et al.[15] decreased the number of road segments from 611 to only 100 road segments in steps of 50 and R^2 values remained stable up until 200 road segments. Even LUR models based on 100 segments predicted on average 73% of the variation (opposed to 74% for the full dataset), albeit with a wider confidence interval (55-85% opposed to 70-78% for the full dataset). In unpublished work in our EXPOsOMICS campaign (chapter 3), we analysed the influence of the number of sampling days and road segments on the ability of the LUR model to predict long-term stationary measurements (Figure B.2). We found that sampling days and number of road segments can be reduced significantly when developing a LUR model for predicting home outdoor exposure for UFP. In other words, when the goal of a study is to predict home outdoor exposure by a LUR model, the sampling campaign can be relatively short (less than 40 days and/or 2,000 road segments). Of note, decreasing the number of road segments by taking a random subset is not very realistic. Because it is impossible to randomly 'hop' from one road segment to another road segment, it makes more sense to decrease the number of drive days for subsample analyses[10]. The fact that there is only limited influence of the number of road segments and drive days (above a certain threshold) on LUR model predictions is a major advantage of mobile monitoring.

Figure B.2

Predictions of mobile models from the EXPOsOMICS campaign compared to long-term stationary measurements with decreasing sampling days (top) and decreasing number of road segments (bottom).

↓

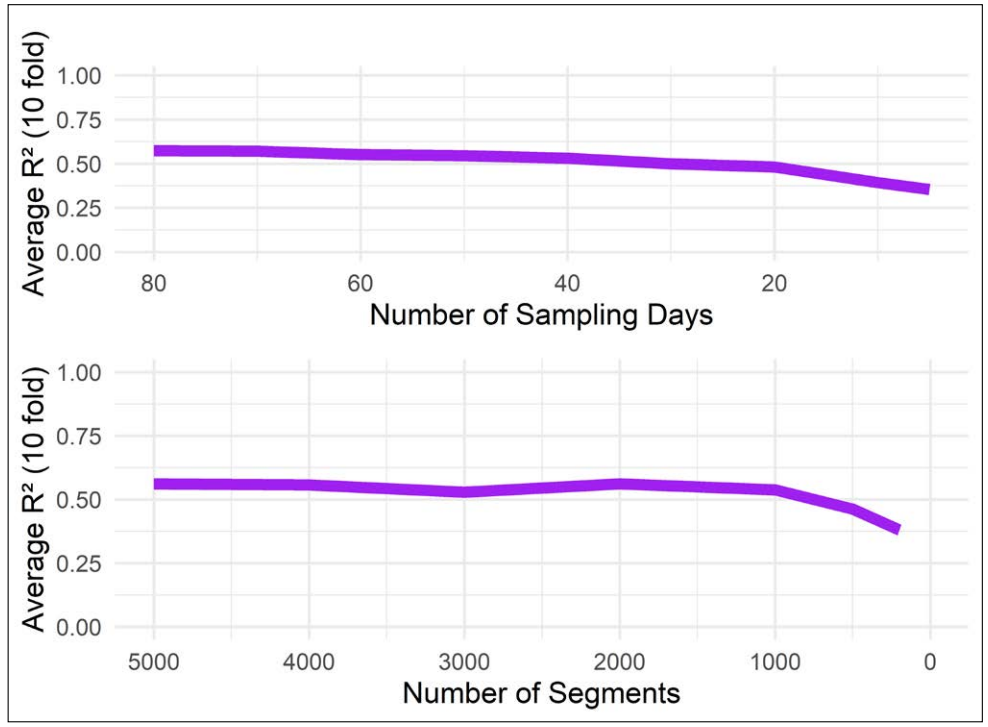


Table B.1

Overview of UFP LUR studies.

↓

Reference	City	Design	Instrument	Number of sites/ road segments	Temporal resolution	Model Structure	Model R^2	GIS Variables*
Hoek et al. 2011	Amsterdam	Long-Term	CPC 3022	50 sites	1 x 7 days	Supervised Stepwise Regression	0.67, CV: 0.57	NT, PD, P
Rivera et al. 2012	Girona	Short-Term	P-Trak	644 sites	1 x 15 min	Supervised Stepwise Regression	0.36, HV: 0.35	NT, PD, MR
Abernethy et al. 2013	Vancouver	Short-Term	CPC 3007	80 sites	1 x 1 hour	Backward Grouped Regression	0.48, LOOCV: 0.32	P, R, MR
Saraswat et al. 2013	New Delhi	Short-Term	CPC 3007	48 sites	1-3 hours	Backward Grouped Regression	Morning: 0.28, Afternoon: 0.23	PD
Patton et al. 2014	Somerville	Mobile (Car)	CPC 3775	Total of 15km (5m)	43 days	Stepwise Regression for screening, General Additive Models	0.43	NT, MR, M
Ragettli et al. 2014	Basel	Short-Term	miniDiSC	60 sites	3 x 20 minutes	Supervised Stepwise Regression	0.68	NT, PD, M
Eeftens et al. 2015	4 areas in Switzerland	Long-Term	miniDiSC	67 sites	3 x 7 days	Supervised Stepwise Regression	0.24, 0.66, 0.66 and 0.47	MR
Hankey and Marshall 2015	Minneapolis	Mobile (Bicycle)	CPC 3007	1000-3000 segments (100m-300m)	42 runs	Stepwise Linear Regression	0.50	PD, MR, I
Hasenfratz et al. 2015	Zurich	Mobile (Trams)	miniDiSC	300-1300 grid of 100m2	50 visits	General Additive Models	0.38	NT, PD
Ghassoun et al. 2015	Braunschweig	Short-Term	SMPS	27 sites	15 x 3 minutes	Backward Grouped Regression	0.74 in summer, 0.85 in winter	PD, MR, I
Montagne et al. 2015	Amsterdam & Rotterdam	Short-Term	CPC 3007	161 sites	3 x 30 minutes	Supervised Stepwise Regression	0.37, HV: 0.29	NT, PD, P
Weichenthal et al. 2015	Toronto	Mobile (Car)	CPC 3007	405 segments (160m)	15 days, 5 min total	Regression Models Single and Multivariable	0.67 (External 0.50)	MR, G, A
Sabaliauskas et al. 2015	Toronto	Mobile (Walking)	FMPS	112 segments (430m)	3 visits, 5-10 min	Stepwise Linear Regression	0.68 on 7 fixed sites	PD, MR, I
Farrel et al. 2016	Montreal	Mobile (Bicycle)	CPC 3007	4,058 segments (100-300m)	1-8 visits, 10min	Generalised Linear Model	0.38-0.43	PD, MR
Kerckhoffs et al. 2016	Amsterdam & Rotterdam	Mobile (Car)	CPC 3007	2,964 segments (120m)	2 or 3 visits	Supervised Stepwise Regression	0.11 (External: 0.51)	PD, A, P, MR, NT, G
Weichenthal et al. 2016	Montreal	Mobile (Bicycle and Car)	CPC 3007	414 segments (370m)	At least 200 data points	KRLS	Linear: 0.62 (HV: 0.60, ext.: 0.55) KRLS: 0.79 (HV:0.67, ext.: 0.58)	NT, PD, G, M
Cattani et al. 2017	Rome	Long-Term	CPC 3007	28 sites	3 x 7 days	Supervised Stepwise Regression	0.71, LOOCV: 0.60	PD, MR, G
Hatzopoulou et al 2017	Montreal	Mobile (Car)	GRIMM and CPC 3775	1,821 segments (100m)	3-16 visits	Supervised Stepwise Regression	0.60 for 3+ visits, 0.74 for 16+ visits	PD, I, A
Kerckhoffs et al. 2017	Amsterdam, Maastricht & Utrecht	Mobile (Car)	CPC 3007	5,236 segments (100m)	2 or 3 visits	Supervised Stepwise Regression	0.12 (External: 0.57)	PD, P, MR, NT, G
van Nunen et al. 2017	Basel, Heraklion, Amsterdam, Norwich, Sabadell, and Turin	Short-Term	CPC 3007	160 sites per city	3 x 30 minutes	Supervised Stepwise Regression	Basel: 0.30, Heraklion: 0.37, Amsterdam: 0.48, Norwich: 0.39, Sabadell: 0.28 and Turin: 0.40	NT, PD
Wolf et al. 2017	Augsburg	Long-Term	GRIMM and SMPS	20 sites	3 x 14 days	Supervised Stepwise Regression	0.92, LOOCV: 0.83	NT, PD, G, I
Shairsingh et al 2019	Toronto	Mobile (Car)	DiSCmini	5,224 segments (140m)	2 or 3 visits	Deconvoluted regression	0.46, HV: 0.30-0.40	PD, R, MR
Jones et al. 2020	Los Angeles	Short-Term	DiSCmini	215 sites	3 x 30 minutes	Supervised Stepwise Regression	0.66 HV: 0.59	PD, A

*NT= Nearby traffic, P=Ports, PD=Population Density, R=Restaurants, MR=Major Roads, G=Greenery/Nature, I=Industry, A=Airports, M=Meteorology

Section C: Robustness of mobile UFP LUR models over time and scale

This dissertation describes 3 campaigns over a span of 5 years. Models created with measurements from different years and cities were very similar. In chapter 3 we compared predicted concentrations from the LUR models of the MUSiC campaign with the EXPOsOMICS campaign and found a high correlation. Also bias and variables selected in these models were very similar. In chapter 4 we combined the data in the MUSiC and EXPOsOMICS campaigns for the analysis of different prediction algorithms. As expected, this model was also very similar to the previous two models. To complete this analysis, we add the national model created with supervised stepwise regression in chapter 6 to this comparison and find that all model predictions to 1000 random addresses across the Netherlands are very similar (Figure C.1), especially the EXPOsOMICS and the national RUN models. The MUSiC campaign was measured in the two largest cities in the Netherlands (Amsterdam and Rotterdam), both located in the West of the country, resulting in high correlations in the West of the country between MUSiC and other campaigns. EXPOsOMICS was measured in Amsterdam, Maastricht and Utrecht and surrounding towns. This created enough variation in LUR predictors to characterise UFP exposure on a national scale, like the RUN campaign which covered the entire Netherlands. Bias (overestimation) was a bit less in the national model compared to the models in the MUSiC and EXPOsOMICS campaigns and will be discussed in section G.

Variables selected in all three supervised stepwise regression models were fairly similar in the sense that all models included a population density predictor in 5000-meter buffer, the presence of a port in the neighbourhood and local traffic variables. See last column of table B.1 for GIS predictors in other UFP LUR models.

Table C.1

Predictor variables selected in the supervised stepwise LUR model in all three campaigns.

↓

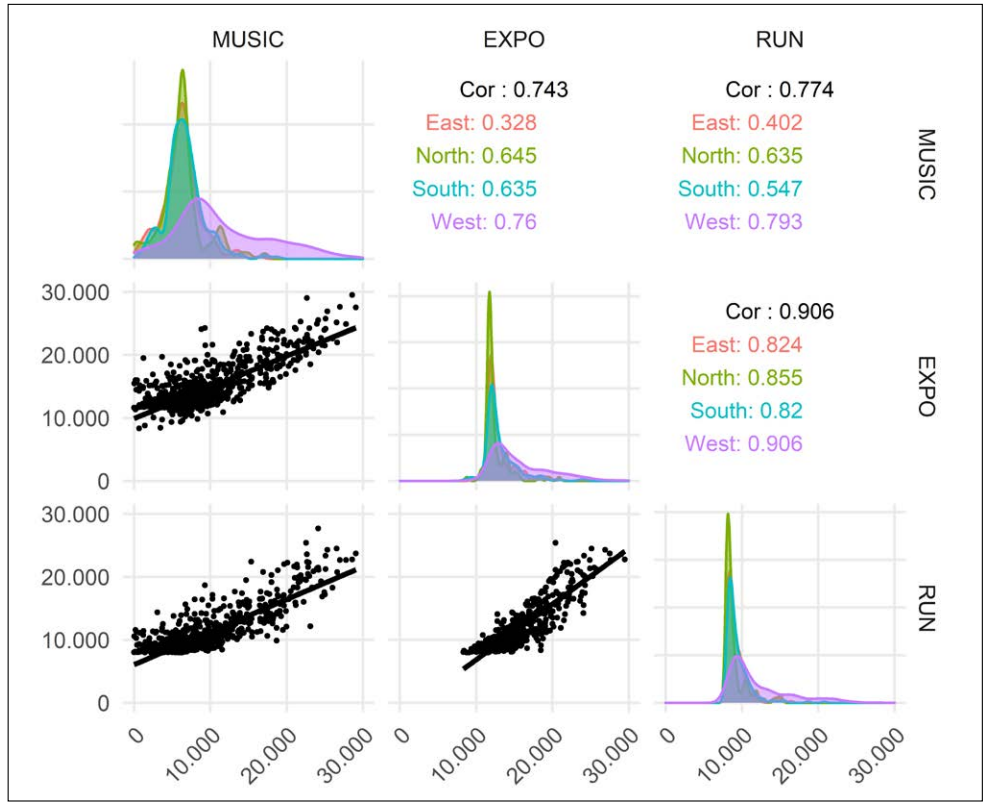
Campaign	Traffic	Population	Industry	Green Space
MUSiC	MRDL_50* TMLOA_100* TLOA_50	POPULATION_5000	PORT_1000 AIRPORT_5000	NATURE_5000
EXPOsOMICS	TRAFNEAR HTRAFNEAR MRDL_100	RESIDENTIAL_5000	PORT_5000	URBANGREEN_500
RUN	TRAFNEAR HTRAFNEAR TLOA_50 TMLOA_50	HOUSEHOLD_5000	PORT_5000	

*MRDL: Major road length, T(M)LOA: traffic intensity on (major) roads, number indicates the buffer size.

Figure C.1

Comparison of exposure predictions on 1000 random addresses in the Netherlands between our mobile monitoring campaigns.

↓



All LUR models are based on supervised forward stepwise approach.

Local Traffic Intensity

It is no surprise all three stepwise models have two or more local traffic intensity variables in the final model (**NT** in last column table B.1). Many studies have conclusively shown that motor vehicle emissions constitute the major source of ultrafine particle pollution in urban environments [21]. Most UFP LUR models include these variables in their final models (references in table B.1). Also, our non-linear algorithms described in chapter 3 gave much importance to traffic intensity.

Road Length

Two out of three supervised stepwise models select the length of major roads in their final model. We interpret that as a surrogate variable for traffic intensity because major roads usually have much more (heavy) traffic then residential roads. Other LUR models often include a major road variable as well (references in table B.1), often with the distance to major road.

Ports

Ports can contribute significantly to UFP concentrations in surrounding areas [22,23]. Concentrations were found to be two times or more higher near shipping routes compared to background locations further away, even surpassing the contribution of traffic[22]. Research in the Netherlands showed that cargo ships contributed 3,000 particles/cm^3 to the UFP concentration near houses along waterways[24]. Increased UFP concentrations also originate from the increased heavy traffic around port activities. Moore et al.[25] and Olvera et al.[26] found that heavy-duty diesel vehicles that load and unload cargo ships have the strongest impact on UFP concentrations. All three supervised stepwise models selected the area of ports as important predictor.

Airports

Emissions from ascending and descending aircrafts can significantly impact UFP concentrations up to 10km downwind[27] and several studies include the vicinity to an airport in their LUR models [15,17,20]. However, mobile monitoring might not be the best design to measure the pollution from airports as this is strongly dependent on wind speed and direction. Dispersion models can include flight schedules and runway direction, making them more suited to assess air pollution from airports then our land-use regression models with limited sites (drive days) near the airport. Airports were only selected in the model based on the MUSiC campaign, with relatively the most measurements near the big airport close to Amsterdam.

Industry

Except for ports and airports, industrial land use was not selected into the models. This likely has to do with the relatively poor representation of industrial sources by the representation of the category industry in the European land-use data (CORINE). Both heavy industry and commercial buildings are labelled as industry in this dataset.

Population Density

Most LUR models for air pollution include a variable that describes urbanisation (PD in last column table B.1). In our models the highly correlated residential land area, population numbers and household numbers in large buffers (mostly 5000m) were selected. These numbers can reflect the difference between small and big towns and/or the difference between the inner and outer city. These differences can be related to a wide variety of sources including the amount of traffic and residential heating for example.

Agriculture

Some air pollutants such as PM_{10}, ammonia and endotoxin have been associated with the vicinity to agricultural activities [28], but we could not find this association for UFP (or BC) in our studies. However, we cannot exclude this influence on UFP concentrations as most models are based on urban measurements. In the national model (chapter 6), agriculture land area was also not selected in the model. Must be said that measurements were not performed in very rural areas, but mostly in towns and cities. Future studies could look at specific agricultural variables, like the number of farms within a certain buffer, to gain more insight on the influence of agriculture on UFP concentrations.

Restaurants

While there is some evidence on the influence of (fast-food) restaurants on UFP concentrations [29], our models did not pick that up. Most restaurants in the Netherlands are in pedestrian friendly environments, making it difficult to measure with a car. Outside Europe restaurants are more often located near roads, so they were picked up as important predictors in two Canadian studies [11,30].

Green Space

Two out of three supervised regression LUR models selected a type of green space in their final model, being the area of nature in a 5000 and urban green in a 500-meter buffer. This likely due to absence of the sources listed above [31,32]. The negative correlation between surrounding green and air pollution was also reported on by Klompmaker et al[33] in the Netherlands.

Section D: Do we need more machine learning?

In previous analysis we assumed that relationships between the pollutant and explanatory variables are (close to) linear. While it is plausible that this is the case for some variables, like traffic intensity and population density, it might be different for others. In chapter 4 we explored the additional value of allowing for more flexible relationships including non-linearity and interactions between variables for model predictions.

However, apart from neural net, the type of model did not influence our predicted concentrations greatly, both for mobile and short-term stationary LUR models (chapter 4). While Weichenthal et al.[18] did find a difference of 17% between model R^2 for linear regression (R^2=0.62) and a machine learning approach (KRLS; R^2=0.79), this diminished when models were applied to an external dataset. Higher training data R^2 does not necessarily equate to higher test R^2 for the external long-term average exposure estimates (chapter 4). This was later verified by Chen et al.[34], who compared several model algorithms based on long-term stationary measurements and found minor difference between approaches. When the relationships between pollutant and predictor variables are not complex (no major interactions and no major deviations from linear associations), there is not much to gain for machine learning algorithms as opposed to “simple” linear regression.

Explained in section B.

This could change with more detailed / other predictor data. Increasing the complexity of the data and the number of dimensions, it becomes more likely that linear approaches will have more and more difficulties to take into account the inter-variables relationships [35]. The downside of machine learning is that such algorithms are generally more difficult to understand the relationship between predictor and outcome. Must be said that there are many advancements made in visualizing machine learning algorithms in the past year(s). Linear regression models are developed in a stepwise fashion, while some machine learning algorithms can consider several (or all) predictor variables at once making it difficult for a human mind to comprehend. This is not a big problem if the only goal of the model is to give the best possible prediction, but it will not tell you what factors cause the spatial variation of air pollution.

Section E: Do we need temporal correction?

The prediction models in all campaigns were solely spatial and were therefore temporally corrected. To do this, we used the same reference site in the three campaigns in the centre of the Netherlands (Bunnik), with the exact same equipment as the mobile monitoring platform. We corrected all on-road measurements with the difference method to exclude as much temporal variation as possible.

Big assumption of this method is that the cities measured in these studies have the same temporal variation over the course of a year/week/day. Especially for cities like Maastricht and Groningen, located 200km away from the reference site, this seems not optimal. However, in chapter 5 we showed that temporal correlation between different regional background locations was moderately high, even over larger distances (up to 180km).

Klompmaker et al[36] analysed the effect of temporally correcting data for UFP and found minor difference in the variance ratio between corrected and uncorrected (30-minute averaged) UFP data. They argued that this is due to the short sampling times. Difference with our mobile studies is that reference data we used was averaged over 2 hours instead of the exact time-corresponding 30-minutes used by Klompmaker et al. 2-hour periods are more stable and less affected by local variations, so we suspect that correction may have a larger impact with 2-hour periods.

This raises the question whether we need temporal correction at all. Only one other mobile monitoring LUR model in table B.1 used a central site to adjust concentrations like our campaigns [6]. Others used other routinely measured traffic related air pollutants [8,17], meteorology in the model [11,37,38] or did not temporally adjust concentrations [15,16,39]. Hankey and Marshall et al.[6] tested how temporal adjustment influenced model performance and found that mean adjusted R^2 decreased by only 0.02.

We already argued that a high number of similar road segments in mobile monitoring campaigns serve as pseudo repeats. This could also be extended to temporal variability. While individual measurements could be adjusted by up to 3,000 particles/cm^3, these adjustments are levelled out over multiple similar segments measured on different days or different times of the day. Unpublished analysis showed that predictions made by the national linear stepwise LUR model (from chapter 6) on 1000 random addresses in the Netherlands correlated 0.87 (without bias) with predictions from the uncorrected UFP model. On top of that, both models performed equally well (R^2 within 0.01) explaining variation of long-term

concentrations in Amsterdam and Utrecht. Important element in this is that there should be no correlation between space and time. All road types should be measured on different parts of the day and on different days over a week and season.

In summary, for the creation of LUR models it seems not vital to temporally correct UFP concentrations. For showing or analysing (individual) measurements it is very important.

Section F: Measurements vs. Models

Taking it one step further, we can ask the question if it is necessary to create models all together. Several projects have been started in the last few years, where Google Street View cars are equipped with fast-response air pollution monitors. Amsterdam, Copenhagen, London, Sydney, and multiple American cities are experimenting with these platforms. First results have been published by Apte et al.[40], monitoring a 30-km² area in Oakland, CA. Resulting maps of annual daytime NO, NO_2, and black carbon at a 30 m-scale reveal stable, persistent pollution patterns with sharp small-scale variability attributable to local sources.

To compare data-only maps to models, Messier et al.[10] measured all streets in a certain neighbourhood at least 50 times and assumed that driving 50 times on different days would generate a stable long-term average concentration. Then they decreased the number of different days a certain street segment was measured and compared data-only maps with LUR models. They found that data-only mapping performed poorly with few (1–2) repeated drives but obtained R^2 values better than the LUR approach within 4 to 8 repeated drive days per road segment. The balance between data-only and LUR-model maps depends on how extensive and detailed predictor variables are available. More and better predictors likely increase the performance of LUR models.

Translation strategies further explained in section G.

Disadvantage of the data-only approach is the fact that the resulting map consists of only road segments and not the locations where people live. So, like all mobile sampling campaigns, there would still be a need to translate on-road measurements to residential addresses. Another disadvantage is the amount of time it takes to measure a certain area, especially when you consider large (national) cohorts. The biggest advantage is the small-scale variation (<50m) and hotspots of air pollution it can detect, whereas LUR models generally have more difficulty characterising small-scale variation.

In summary, data-only mapping seems viable from a policy standpoint, where small-scale variation can be detected and acted upon with interventions. From an epidemiological standpoint, especially regarding large multi area cohorts, it makes more sense to use LUR models. Whether the goal of a study is to create data-only maps or LUR models, there are four main factors to consider in setting up a monitoring campaign for traffic-related air pollutants (Figure F.1):

- Number of measurement sites
- Duration of individual measurement

- Number of replications
- Spatial aggregation (for mobile monitoring)

For mobile monitoring campaigns that aim at developing data-only maps, the number of measurement sites is set by the area you want to measure (slider 1). Also, for car-based monitoring, the duration of individual measurements will always be in the order of seconds (slider 2). This creates some temporal issues when long-term averages are predicted. So, one needs to consider the time of day (daytime, night-time, rush hour), day of the week and season the road segments are measured.

Temporal issues further explained in section C and G.

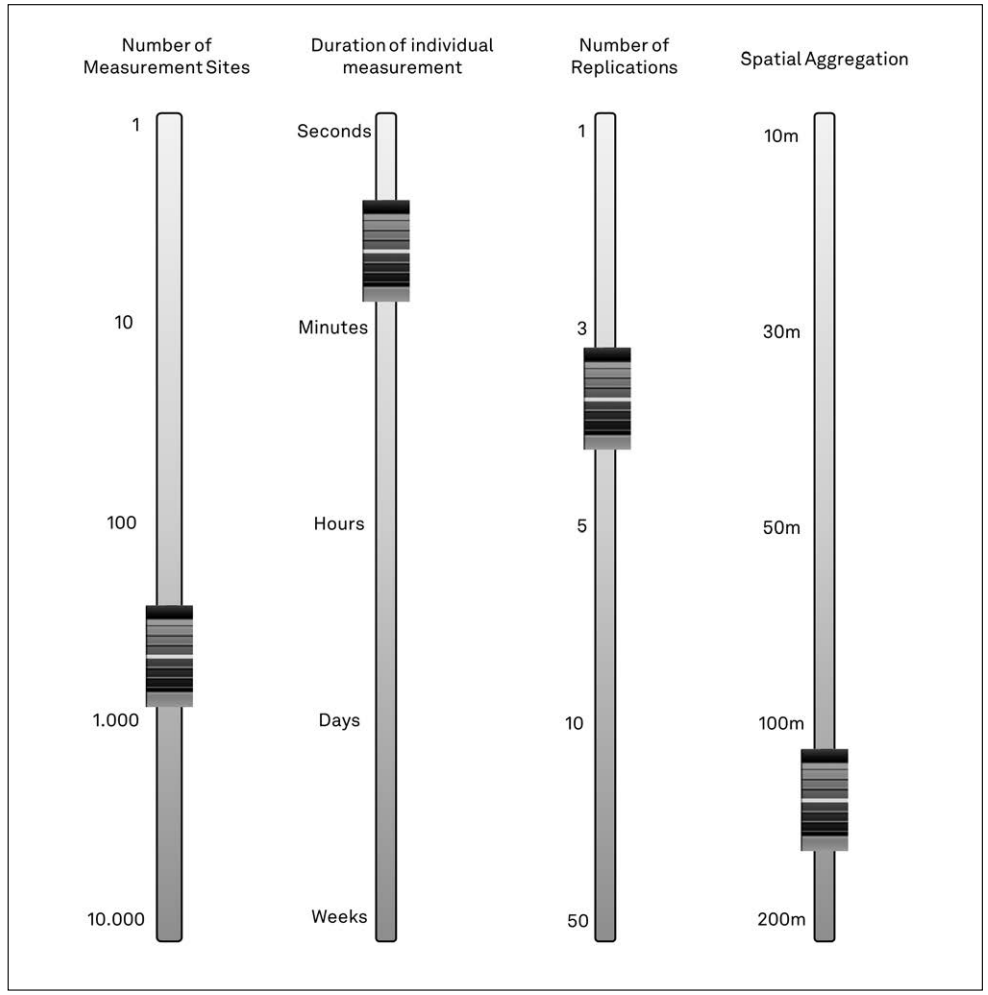
This leaves the number of replications and spatial aggregation as adjustable settings. As stated above, Messier et al.[10] found that 4 to 8 repeats were sufficient to create an exposure map for black carbon and NO better or at par with a LUR model based on the same data. If we assume that 50 drives on a road segment generates a true long-term average concentration for that road segment, then more than 8 drives would be needed to come close to true long-term average concentrations. Figure F.2 shows the correlations between data-only/LUR models and long-term average concentrations related to the number of drives needed to achieve this correlation.

It is evident that with just a few drives on a road segment (1-4 drives) we are not able to characterise long-term average concentrations, but a LUR model can easily achieve a good correlation with only 1 or 2 drives. However, in the study by Messier et al[10], the ability of data-only mapping surpasses the LUR model at about 5 drives when it comes to predicting long-term average (on-road) concentrations. Future research should verify if this holds in other geographical areas and for other pollutants. For example, UFP is more variable than NO in urban environments and mobile measurements might need more repeats to achieve a stable average.

Figure F.1

Four design factors in monitoring campaigns for air pollution.

↓

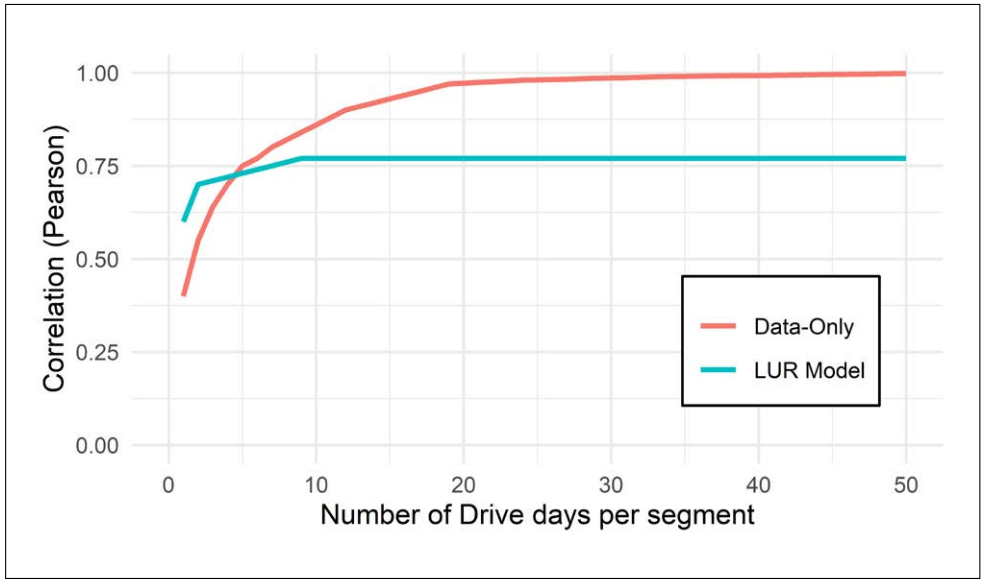


The spatial aggregation used in the study by Messier et al.[10] was 30 meters, meaning every individual segment consisted of only a few seconds. As the goal of data-only mapping is often to find fine-scale variation of air pollution it is best to keep the spatial resolution as low (< 200m) as possible. Since the study by Messier et al.[10] is the first to use a mobile platform to cover a wider geographic area, more research is needed to verify if the amount of 4-8 repeats is sufficient to create stable long-term concentration maps.

Figure F.2

Performance of data-only mapping and LUR models for NO related to the number of drive days per segment (reproduced from Messier et al[19]).

↓



For monitoring campaigns that intend to create LUR models, the listed sliders are more related to each other. Regarding the number of measurement sites, it is necessary to have a minimum of 20 sites [41,42], and often more than double. However, this number is based on long-term (repeated 14-day averages) measurements. For mobile monitoring this is much less an issue. In section B, I showed that the number of road segments for model development can often be reduced by a great amount (when road type variation is preserved). Duration of individual measurements in mobile monitoring is usually in the order of seconds to minutes, depending on the speed and mode of the mobile platform. This will also impact the spatial aggregation than can be used. Most LUR models based on mobile data use the road segments as spatial aggregation for their models, defined as stretch of road from one intersection to the next. Length is mostly between 100 and 300m in urban environments (140 and 110m on average in our studies). Hankey and Marshall [6] analysed the impact of spatial resolution in a mobile monitoring campaign and found very little difference between the performance of LUR models that were based on segments where concentrations were averaged over 50, 100 and 200m.

This leaves the question how many repeats are needed to create LUR models. In section B, I showed that road segments with at least 16 observations achieved a higher adjusted model R^2 with fewer explanatory variables compared to the model developed with road segments having 3+ visits [15]. However, this concerns the training model R^2 (and not external long-term averages). In our studies we showed that with a limited number of repeats (often 2-3 drive days per segment) robust external predictions can be made. With only 30% of the total number of drive days, R^2 values only slightly decreased (Figure B.2). And as stated in several sections, the number of road segments with similar characteristics serve as pseudo repeats, meaning LUR models can be developed based on road segments with mobile measurements only measured once, as long as coverage and distribution of all predictor variables is similar to the prediction sites.

Future research is needed how we can combine data-only mapping with (LUR) models. Creating LUR models on a big scale, while retaining the small-scale variation recorded in data-only mapping could further increase performance of air pollution exposure estimates. This could be done by ranking the precision of measurements versus the performance of LUR models, for example by assigning weights. Since models will never perfectly predict concentration levels there will always be a moment where measurements are more precise than models when measuring long enough. By creating a mixed model with a LUR model as the fixed effects and all road segments as random effects, both LUR model and all measurements influence the predicted concentration per road segment. The more precise the on-road measurements are (i.e., less variation in measurements), the more it influences the output. This could for example be achieved by putting measurements devices on municipal utility vehicles. Firstly, this will create enough data to develop a LUR model in a short amount of time and secondly when more and more data is collected a data-only map could be created.

Section G: Influence on health-effect studies

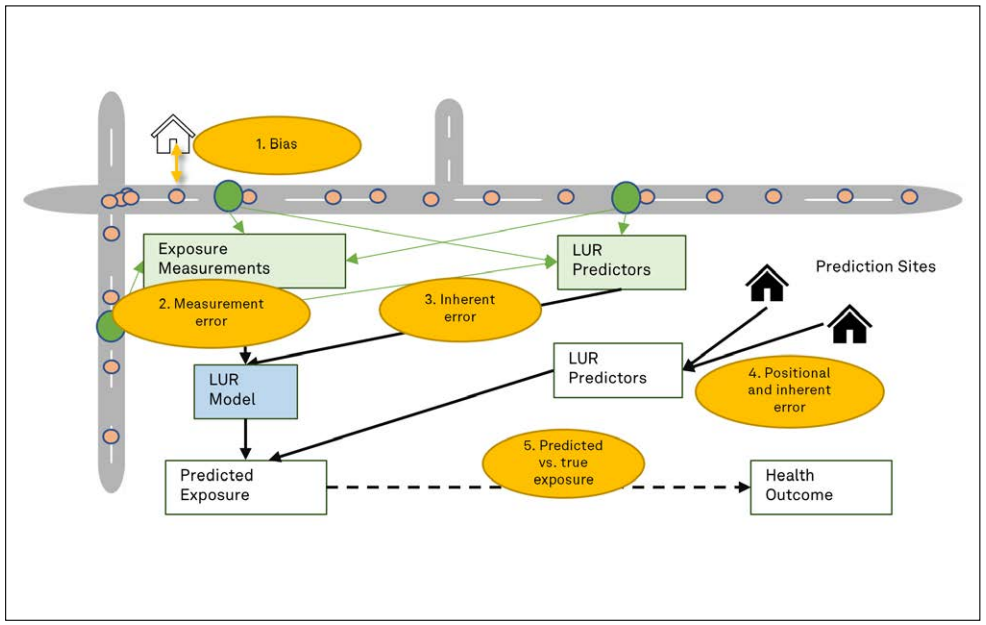
Now that we have robust models for long-term UFP exposure they can be used in health-effect studies. However, epidemiological studies often treat the predicted exposures as known, and plug them into a health model to estimate the association of interest without accounting for measurement error [43]. In this section, I describe the expected error in exposure estimates in mobile monitoring studies, schematically shown in figure G.1 and split into three categories:

- On-road measurements vs. residential exposure predictions (point 1 in figure G.1).
- Measurement error and inherent error (point 2-4)
- Predicted vs. true (personal) exposure (point 5)

Figure G.1

Illustration of the aspects that contribute to error in land-use regression model predictions based on mobile monitoring.

↓



In this section I mainly discuss the first two points as they are unique to mobile sampling campaigns. Inherent error (error in source information), positional error of prediction sites and the difference between predicted and true exposure applies to all LUR models[44–46] and is not discussed here.

On-road measurements vs. residential exposure predictions

As study participants do not live in the middle of the road, prediction sites are per definition not similar to monitoring sites. This is called spatial misalignment and can induce error in exposure estimates [44,47]. In chapter 3 we showed the difference between predicted concentration levels based on mobile and short-term stationary LUR models, tested on 500 random addresses in Amsterdam in the MUSiC and EXPOsOMICS campaign. We found a difference of 29% and 33% between predicted and long-term average concentrations for MUSiC and EXPOsOMICS, respectively. In table G.1 we compare the bias of all three campaigns when models are tested on long-term stationary measurements from the EXPOsOMICS campaign (n=42).

Table G.1

Overestimation of mobile models compared to long-term average concentrations.

↓

	MUSiC	EXPOsOMICS	RUN
Cities	Amsterdam and Rotterdam	Amsterdam, Utrecht, and Maastricht	20 cities and towns in the Netherlands
UFP over-prediction	**31% (4,701 particles/cm³)**	**30% (4,511 particles/cm³)**	**14% (2,156 particles/cm³)**
Traffic	37% (6,612 particles/cm³)	31% (5,566 particles/cm³)	16% (2,886 particles/cm³)
Urban Background	24% (3,051 particles/cm³)	29% (3,599 particles/cm³)	12% (1,526 particles/cm³)

Overprediction caused by the RUN model is less than the overestimation we found in the other two studies. We attributed this to the fact that the RUN study was measured in less urbanized areas compared to previous studies. This means that there was on average less traffic intensity on all road segments, especially on roads characterized as major roads in less urbanized areas. Local traffic was therefore solely assessed by traffic counts in the RUN models instead of also the length of major roads in a small buffer (as in the other two campaigns). The RUN model was therefore able to better capture the distance between local (traffic) sources and prediction sites.

In chapter 6, we found a high similarity between concentrations measured by the urban background short-term stationary sites and the derived background concentration from the on-road concentrations using the deconvolution approach. This suggests that the overestimation originates from on-road driving in general (which occurs on traffic and urban background road segments). Splitting the overprediction from all three campaigns into traffic and urban background parts backs up this statement (Table G.1). Difference in overprediction between road types was very minor for

all three campaigns on a relative scale. In addition, short-term stationary models in the MUSiC and EXPOsOMICS campaigns did not over-predict long-term concentrations.

In chapter 3, we explored four methodologies to correct for the overestimation of mobile models:

- **Distance:** Correct predictions based on the distance to the road.
- **LUR Model:** Using the delta as dependent variable in a LUR Model.
- **Percentage:** Reducing the mobile predicted levels by 30%.
- **Absolute:** Reducing the mobile predicted levels by 4,500 particles/cm³.

As distance and road type do not seem to affect overprediction, relative and absolute correction are considered the best options. To our knowledge, one study measured the difference between mobile and residential levels on the same street. Simon et al.[48] found a difference of 19,000 and 5,300 particles/cm³ between mobile monitoring and residences, averaged over 1 minute, in Boston (n=45) and Chelsea (n=247; Massachusetts, USA), respectively. Two other studies compared measurements of UFP on the sidewalk and at the façade of buildings and found a difference in concentration levels of about 15-20% [49,50]. In addition, multiple studies have observed sharp UFP gradients in near-road urban environments with gradients similar to what was observed in our campaigns [21,51–57].

Sabaliauskas et al.[16] found differences of 28% between predicted concentrations based on a mobile model and measured UFP concentrations at 7 fixed sites. They attributed this to the fact that mobile monitoring was only done during the day, because predicted concentrations did not overestimate measurements between 12:00 and 15:00. All our measurements were done between 9:15 and 16:00, excluding rush hour. This could lead to some underestimation of our LUR models regarding day-time exposure. The exclusion of night-time period could in contrast lead to an overprediction of 24h average concentrations. Measuring during one season only [6,16,38] could cause some overestimation or underestimation as well. It is difficult to separate and rank which of these factors effects the bias in concentration the most, so the best way to correct is to compare predicted concentrations with long-term average measurements at the façade of buildings.

In conclusion, the fact that 1) short-term stationary models (based on measurements during the day) do not overpredict the long-term concentrations levels in our test data, 2) short-term stationary background concentrations are similar to the urban background signal of the mobile measurements and 3) we could not find a relationship between the distance to the road and the overprediction suggests that overprediction can only be attributed to on-road driving. While the absolute difference in concentrations between on-road driving and at the façade of buildings is higher on major roads, the relative difference is similar for minor and major roads. Therefore, a correction factor in the order of 15 to 30 % seems to be the most appropriate for correcting on-road measurements. However, this systematic bias will likely not influence associations between air pollution and health but is important to consider when absolute levels are of interest.

Measurement error in mobile sampling campaigns

Average explained variance (R^2 values) for home outdoor addresses for all UFP models in the different campaigns was between 0.5 and 0.6. Multiple other studies that made UFP models with validations to long-term exposure measurements also found performance in that range or higher (see table B.1), with an average R^2 of around 0.6. Studies that assess spatial variation of other traffic-related pollutant like NO_2 usually create LUR models that describe spatial variation slightly better than UFP models [58] although differences in validation method (external vs internal, duration of long-term average) complicate comparisons across studies and especially pollutants. This makes it difficult to assess how much gain can be achieved in the development of LUR models for UFP. Nevertheless, it is important to know where the error in a model comes from in order to better assess the potential for UFP (and BC) models and its influence on health-effect estimates.

In the previous sections, I described the advantages of having multiple road segments with similar characteristics that serve as pseudo repeats. That means that on-road concentrations are sort of averaged with other roads that have similar road characteristics. Regarding error structure this can be considered as Berkson-like error and the main part of the error in the model. Berkson error affects the power of the study, but does not bias regression coefficients [59]. Vlaanderen et al.[44] purposely introduced error in the predictor variables and found that health-effect estimates where only biased when there was an imbalance in the degree of error between the predictors at the monitoring sites and predictors at the prediction sites. On top of that, Szpiro et al.[60] showed that more accurate exposure prediction does not necessarily lead to improved health-effect estimation. We therefore conclude that bias in health-effect estimates is likely minimally affected by measurement error induced by mobile monitoring specifically.

Conclusion

Mobile monitoring offers additional valuable insights into traffic-related air pollutants like UFP and BC, which are highly variable in space and time. Using a mobile platform enabled us to measure a large number of different exposure settings in a study area with a minimal amount of costly measurement devices. Despite the limited time spent on a road segment we found that it is possible to predict long-term exposure with LUR models based on mobile measurements. R^2 values are equally good as LUR models based on short-term stationary measurements. On top of that, only a limited number of road segments are needed to develop LUR models, as long as there is a large variation in land use and urban topography. It is therefore possible to scale a mobile monitoring design to different cities and/or entire countries. This can greatly help in exposure assessment in health effects studies of long-term exposure to UFP.

References

(1) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712–8720. https://doi.org/10.1021/es505791g.

(2) Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of Ultrafine Particles and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environ. Sci. Technol.* **2016**, *50* (23), 12894–12902. https://doi.org/10.1021/acs.est.6b03476.

(3) van Nunen, E.; Vermeulen, R.; Tsai, M.-Y.; Probst-Hensch, N.; Ineichen, A.; Davey, M.; Imboden, M.; Ducret-Stich, R.; Naccarati, A.; Raffaele, D.; Ranzi, A.; Ivaldi, C.; Galassi, C.; Nieuwenhuijsen, M.; Curto, A.; Donaire-Gonzalez, D.; Cirach, M.; Chatzi, L.; Kampouri, M.; Vlaanderen, J.; Meliefste, K.; Buijtenhuijs, D.; Brunekreef, B.; Morley, D.; Vineis, P.; Gulliver, J.; Hoek, G. Land Use Regression Models for Ultrafine Particles in Six European Areas. *Environ. Sci. Technol.* **2017**, *51* (6), 3336–3345. https://doi.org/10.1021/acs.est.6b05920.

(4) Kerckhoffs, J.; Hoek, G.; Vlaanderen, J.; van Nunen, E.; Messier, K.; Brunekreef, B.; Gulliver, J.; Vermeulen, R. Robustness of Intra Urban Land-Use Regression Models for Ultrafine Particles and Black Carbon Based on Mobile Monitoring. *Environ. Res.* **2017**, *159*, 500–508. https://doi.org/10.1016/j.envres.2017.08.040.

(5) Hagler, G. S. W.; Yelverton, T. L. B.; Vedantham, R.; Hansen, A. D. A.; Turner, J. R. Post-Processing Method to Reduce Noise While Preserving High Time Resolution in Aethalometer Real-Time Black Carbon Data. *Aerosol Air Qual. Res.* **2011**, *11* (5), 539–546. https://doi.org/10.4209/aaqr.2011.05.0055.

(6) Hankey, S.; Marshall, J. D. Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, $PM_{2.5}$, Particle Size) Using Mobile Monitoring. *Environ. Sci. Technol.* **2015**, *49* (15), 9194–9202. https://doi.org/10.1021/acs.est.5b01209.

(7) Larson, T.; Henderson, S. B.; Brauer, M. Mobile Monitoring of Particle Light Absorption Coefficient in an Urban Area as a Basis for Land Use Regression. *Environ. Sci. Technol.* **2009**, *43* (13), 4672–4678. https://doi.org/10.1021/es803068e.

(8) Weichenthal, S.; Farrell, W.; Goldberg, M.; Joseph, L.; Hatzopoulou, M. Characterizing the Impact of Traffic and the Built Environment on Near-Road Ultrafine Particle and Black Carbon Concentrations. *Environ. Res.* **2014**, *132*, 305–310. https://doi.org/10.1016/j.envres.2014.04.007.

(9) Van den Bossche, J.; De Baets, B.; Verwaeren, J.; Botteldooren, D.; Theunis, J. Development and Evaluation of Land Use Regression Models for Black Carbon Based on Bicycle and Pedestrian Measurements in the Urban Environment. *Environ. Model. Softw.* **2018**, *99*, 58–69. https://doi.org/10.1016/j.envsoft.2017.09.019.

(10) Messier, K. P.; Chambliss, S. E.; Gani, S.; Alvarez, R.; Brauer, M.; Choi, J. J.; Hamburg, S. P.; Kerckhoffs, J.; Lafranchi, B.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Roy, A.; Szpiro, A. A.; Vermeulen, R. C. H.; Apte, J. S. Mapping Air Pollution with Google Street View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression. *Environ. Sci. Technol.* **2018**, *52* (21), 12563–12572. https://doi.org/10.1021/acs.est.8b03395.

(11) Shairsingh, K. K.; Jeong, C. H.; Wang, J. M.; Brook, J. R.; Evans, G. J. Urban Land Use Regression Models: Can Temporal Deconvolution of Traffic Pollution Measurements Extend the Urban LUR to Suburban Areas? *Atmos. Environ.* **2019**, *196*, 143–151. https://doi.org/10.1016/j.atmosenv.2018.10.013.

(12) Hoek, G.; Beelen, R.; Kos, G.; Dijkema, M.; Zee, S. C. V. Der; Fischer, P. H.; Brunekreef, B. Land Use Regression Model for Ultrafine Particles in Amsterdam. *Environ. Sci. Technol.* **2011**, *45* (2), 622–628. https://doi.org/10.1021/es1023042.

(13) Wolf, K.; Cyrys, J.; Harciníková, T.; Gu, J.; Kusch, T.; Hampel, R.; Schneider, A.; Peters, A. Land Use Regression Modeling of Ultrafine Particles, Ozone, Nitrogen Oxides and Markers of Particulate Matter Pollution in Augsburg, Germany. *Sci. Total Environ.* **2017**. https://doi.org/10.1016/j.scitotenv.2016.11.160.

(14) Cattani, G.; Gaeta, A.; Di Menno di Bucchianico, A.; De Santis, A.; Gaddi, R.; Cusano, M.; Ancona, C.; Badaloni, C.; Forastiere, F.; Gariazzo, C.; Sozzi, R.; Inglessis, M.; Silibello, C.; Salvatori, E.; Manes, F.; Cesaroni, G. Development of Land-Use Regression Models for Exposure Assessment to Ultrafine Particles in Rome, Italy. *Atmos. Environ.* **2017**, *156*, 52–60. https://doi.org/10.1016/j.atmosenv.2017.02.028.

(15) Hatzopoulou, M.; Valois, M. F.; Levy, I.; Mihele, C.; Lu, G.; Bagg, S.; Minet, L.; Brook, J. Robustness of Land-Use Regression Models Developed from Mobile Air Pollutant Measurements. *Environ. Sci. Technol.* **2017**, *51* (7), 3938–3947. https://doi.org/10.1021/acs.est.7b00366.

(16) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Reali, C.; Sun, T.; Evans, G. J. Development of a Land-Use Regression Model for Ultrafine Particles in Toronto, Canada. *Atmos. Environ.* **2015**, *110*, 84–92. https://doi.org/10.1016/j.atmosenv.2015.02.018.

(17) Weichenthal, S.; Van Ryswyk, K.; Goldstein, A.; Shekarrizfard, M.; Hatzopoulou, M. Characterizing the Spatial Distribution of Ambient Ultrafine Particles in Toronto, Canada: A Land Use Regression Model. *Environ. Pollut.* **2015**, *208*, 241–248. https://doi.org/10.1016/j.envpol.2015.04.011.

(18) Weichenthal, S.; Ryswyk, K. Van; Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A Land Use Regression Model for Ambient Ultrafine Particles in Montreal, Canada: A Comparison of Linear Regression and a Machine Learning Approach. *Environ. Res.* **2016**, *146*, 65–72. https://doi.org/10.1016/j.envres.2015.12.016.

(19) Ghassoun, Y.; Ruths, M.; Löwner, M. O.; Weber, S. Intra-Urban Variation of Ultrafine Particles as Evaluated by Process Related Land Use and Pollutant Driven Regression Modelling. *Sci. Total Environ.* **2015**, *536*, 150–160. https://doi.org/10.1016/j.scitotenv.2015.07.051.

(20) Jones, R. R.; Hoek, G.; Fisher, J. A.; Hasheminassab, S.; Wang, D.; Ward, M. H.; Sioutas, C.; Vermeulen, R.; Silverman, D. T. Land Use Regression Models for Ultrafine Particles, Fine Particles, and Black Carbon in Southern California. *Sci. Total Environ.* **2020**, *699*, 134234. https://doi.org/10.1016/j.scitotenv.2019.134234.

(21) Morawska, L.; Ristovski, Z.; Jayaratne, E. R.; Keogh, D. U.; Ling, X. Ambient Nano and Ultrafine Particles from Motor Vehicle Emissions: Characteristics, Ambient Processing and Implications on Human Exposure. *Atmospheric Environment*. November 2008, pp 8113–8138. https://doi.org/10.1016/j.atmosenv.2008.07.050.

(22) González, Y.; Rodríguez, S.; Guerra García, J. C.; Trujillo, J. L.; García, R. Ultrafine Particles Pollution in Urban Coastal Air Due to Ship Emissions. *Atmos. Environ.* **2011**, *45* (28), 4907–4914. https://doi.org/10.1016/j.atmosenv.2011.06.002.

(23) Karl, M.; Pirjola, L.; Karppinen, A.; Jalkanen, J.-P.; Ramacher, M. O. P.; Kukkonen, J. Modeling of the Concentrations of Ultrafine Particles in the Plumes of Ships in the Vicinity of Major Harbors. *Int. J. Environ. Res. Public Health* **2020**, *17* (3), 777. https://doi.org/10.3390/ijerph17030777.

(24) Van der Zee, S. C.; Dijkema, M. B. A.; Van der Laan, J.; Hoek, G. The Impact of Inland Ships and Recreational Boats on Measured NO x and Ultrafine Particle Concentrations along the Waterways. *Atmos. Environ.* **2012**, *55*, 368–376. https://doi.org/10.1016/j.atmosenv.2012.03.055.

(25) Moore, K.; Krudysz, M.; Pakbin, P.; Hudda, N.; Sioutas, C. Intra-Community Variability in Total Particle Number Concentrations in the San Pedro Harbor Area (Los Angeles, California). *Aerosol Sci. Technol.* **2009**, *43* (6), 587–603. https://doi.org/10.1080/02786820902800900.

(26) Olvera, H. A.; Lopez, M.; Guerrero, V.; Garcia, H.; Li, W. W. Ultrafine Particle Levels at an International Port of Entry between the US and Mexico: Exposure Implications for Users, Workers, and Neighbors. *J. Expo. Sci. Environ. Epidemiol.* **2013**, *23* (3), 289–298. https://doi.org/10.1038/jes.2012.119.

(27) Hudda, N.; Gould, T.; Hartin, K.; Larson, T. V.; Fruin, S. A. Emissions from an International Airport Increase Particle Number Concentrations 4-Fold at 10 Km Downwind. *Environ. Sci. Technol.* **2014**, *48* (12), 6628–6635. https://doi.org/10.1021/es5001566.

(28) Giannadaki, D.; Giannakis, E.; Pozzer, A.; Lelieveld, J. Estimating Health and Economic Benefits of Reductions in Air Pollution from Agriculture. **2017**. https://doi.org/10.1016/j.scitotenv.2017.12.064.

(29) Vert, C.; Meliefste, K.; Hoek, G. Outdoor Ultrafine Particle Concentrations in Front of Fast Food Restaurants. *J. Expo. Sci. Environ. Epidemiol.* **2016**, *26* (1), 35–41. https://doi.org/10.1038/jes.2015.64.

(30) Abernethy, R. C.; Allen, R. W.; McKendry, I. G.; Brauer, M. A Land Use Regression Model for Ultrafine Particles in Vancouver, Canada. *Environ. Sci. Technol.* **2013**, *47* (10), 5217–5225. https://doi.org/10.1021/es304495s.

(31) Hystad, P.; Davies, H. W.; Frank, L.; Van Loon, J.; Gehring, U.; Tamburic, L.; Brauer, M. Residential Greenness and Birth Outcomes: Evaluating the Influence of Spatially Correlated Built-Environment Factors. *Environ. Health Perspect.* **2014**, *122* (10), 1095–1102. https://doi.org/10.1289/ehp.1308049.

(32) Markevych, I.; Schoierer, J.; Hartig, T.; Chudnovsky, A.; Hystad, P.; Dzhambov, A. M.; de Vries, S.; Triguero-Mas, M.; Brauer, M.; Nieuwenhuijsen, M. J.; Lupp, G.; Richardson, E. A.; Astell-Burt, T.; Dimitrova, D.; Feng, X.; Sadeh, M.; Standl, M.; Heinrich, J.; Fuertes, E. Exploring Pathways Linking Greenspace to Health: Theoretical and Methodological Guidance. *Environmental Research*. Academic Press Inc. 2017, pp 301–317. https://doi.org/10.1016/j.envres.2017.06.028.

(33) Klompmaker, J. O.; Hoek, G.; Bloemsma, L. D.; Wijga, A. H.; van den Brink, C.; Brunekreef, B.; Lebret, E.; Gehring, U.; Janssen, N. A. H. Associations of Combined Exposures to Surrounding Green, Air Pollution and Traffic Noise on Mental Health. *Environ. Int.* **2019**, *129*, 525–537. https://doi.org/10.1016/j.envint.2019.05.040.

(34) Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Katsouyanni, K.; Janssen, N. A. H.; Martin, R. V.; Samoli, E.; Schwartz, P. E.; Stafoggia, M.; Bellander, T.; Strak, M.; Wolf, K.; Vienneau, D.; Vermeulen, R.; Brunekreef, B.; Hoek, G. A Comparison of Linear Regression, Regularization, and Machine Learning Algorithms to Develop Europe-Wide Spatial Models of Fine Particles and Nitrogen Dioxide. *Environ. Int.* **2019**, *130*. https://doi.org/10.1016/j.envint.2019.104934.

(35) Champendal, A.; Kanevski, M.; Huguenot, P. E. Air Pollution Mapping Using Nonlinear Land Use Regression Models. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Verlag, 2014; Vol. 8581 LNCS, pp 682–690. https://doi.org/10.1007/978-3-319-09150-1_50.

(36) Klompmaker, J. O.; Montagne, D. R.; Meliefste, K.; Hoek, G.; Brunekreef, B. Spatial Variation of Ultrafine Particles and Black Carbon in Two Cities: Results from a Short-Term Measurement Campaign. *Sci. Total Environ.* **2015**, *508*, 266–275. https://doi.org/10.1016/j.scitotenv.2014.11.088.

(37) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An Hourly Regression Model for Ultrafine Particles in a Near-Highway Urban Area. *Environ. Sci. Technol.* **2014**, *48* (6), 3272–3280. https://doi.org/10.1021/es404838k.

(38) Farrell, W.; Weichenthal, S.; Goldberg, M.; Valois, M. F.; Shekarrizfard, M.; Hatzopoulou, M. Near Roadway Air Pollution across a Spatially Extensive Road and Cycling Network. *Environ. Pollut.* **2016**, *212*, 498–507. https://doi.org/10.1016/j.envpol.2016.02.041.

(39) Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; Thiele, L. Deriving High-Resolution Urban Air Pollution Maps Using Mobile Sensor Nodes. *Pervasive Mob. Comput.* **2015**, *16* (PB), 268–285. https://doi.org/10.1016/j.pmcj.2014.11.008.

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

(41) Wang, M.; Beelen, R.; Basagana, X.; Becker, T.; Cesaroni, G.; De Hoogh, K.; Dedele, A.; Declercq, C.; Dimakopoulou, K.; Eeftens, M.; Forastiere, F.; Galassi, C.; Gražulevičiene, R.; Hoffmann, B.; Heinrich, J.; Iakovides, M.; Künzli, N.; Korek, M.; Lindley, S.; Mölter, A.; Mosler, G.; Madsen, C.; Nieuwenhuijsen, M.; Phuleria,

H.; Pedeli, X.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Stempfelet, M.; Tsai, M. Y.; Lanki, T.; Udvardy, O.; Varró, M. J.; Wolf, K.; Weinmayr, G.; Yli-Tuomi, T.; Hoek, G.; Brunekreef, B. Evaluation of Land Use Regression Models for NO2 and Particulate Matter in 20 European Study Areas: The ESCAPE Project. *Environ. Sci. Technol.* **2013**, *47* (9), 4357–4364. https://doi.org/10.1021/es305129t.

(42) Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Vila, J.; Künzli, N. Effect of the Number of Measurement Sites on Land Use Regression Models in Estimating Local Air Pollution. *Atmos. Environ.* **2012**, *54*, 634–642. https://doi.org/10.1016/j.atmosenv.2012.01.064.

(43) Szpiro, A. A.; Paciorek, C. J. Measurement Error in Two-Stage Analyses, with Application to Air Pollution Epidemiology. *Environmetrics* **2013**, *24* (8), 501–517. https://doi.org/10.1002/env.2233.

(44) Vlaanderen, J.; Portengen, L.; Chadeau-Hyam, M.; Szpiro, A.; Gehring, U.; Brunekreef, B.; Hoek, G.; Vermeulen, R. Error in Air Pollution Exposure Model Determinants and Bias in Health Estimates. *J. Expo. Sci. Environ. Epidemiol.* **2019**, *29*, 258–266. https://doi.org/10.1038/s41370-018-0045-x.

(45) Basagaña, X.; Aguilera, I.; Rivera, M.; Agis, D.; Foraster, M.; Marrugat, J.; Elosua, R.; Künzli, N. Measurement Error in Epidemiologic Studies of Air Pollution Based on Land-Use Regression Models. *Am. J. Epidemiol.* **2013**, *178* (8), 1342–1346. https://doi.org/10.1093/aje/kwt127.

(46) Szpiro, A. A.; Sheppard, L.; Lumley, T. Efficient Measurement Error Correction with Spatially Misaligned Data. *Biostatistics* **2011**, *12* (4), 610–623. https://doi.org/10.1093/biostatistics/kxq083.

(47) Gryparis, A.; Paciorek, C. J.; Zeka, A.; Schwartz, J.; Coull, B. A. Measurement Error Caused by Spatial Misalignment in Environmental Epidemiology. *Biostatistics* **2009**, *10* (2), 258–274. https://doi.org/10.1093/biostatistics/kxn033.

(48) Simon, M. C.; Hudda, N.; Naumova, E. N.; Levy, J. I.; Brugge, D.; Durant, J. L. Comparisons of Traffic-Related Ultrafine Particle Number Concentrations Measured in Two Urban Areas by Central, Residential, and Mobile Monitoring. *Atmos. Environ.* **2017**, *169*, 113–127. https://doi.org/10.1016/j.atmosenv.2017.09.003.

(49) Kaur, S.; Nieuwenhuijsen, M. J.; Colvile, R. N. Pedestrian Exposure to Air Pollution along a Major Road in Central London, UK. *Atmos. Environ.* **2005**, *39* (38), 7307–7320. https://doi.org/10.1016/j.atmosenv.2005.09.008.

(50) Ragettli, M. S.; Ducret-Stich, R. E.; Foraster, M.; Morelli, X.; Aguilera, I.; Basagaña, X.; Corradi, E.; Ineichen, A.; Tsai, M. Y.; Probst-Hensch, N.; Rivera, M.; Slama, R.; Künzli, N.; Phuleria, H. C. Spatio-Temporal Variation of Urban Ultrafine Particle Number Concentrations. *Atmos. Environ.* **2014**, *96*, 275–283. https://doi.org/10.1016/j.atmosenv.2014.07.049.

(51) Baldwin, N.; Gilani, O.; Raja, S.; Batterman, S.; Ganguly, R.; Hopke, P.; Berrocal, V.; Robins, T.; Hoogterp, S. Factors Affecting Pollutant Concentrations in the Near-Road Environment. *Atmos. Environ.* **2015**, *115*, 223–235. https://doi.org/10.1016/j.atmosenv.2015.05.024.

(52) Fujita, E. M.; Campbell, D. E.; Arnott, W. P.; Johnson, T.; Ollison, W. Concentrations of Mobile Source Air Pollutants in Urban Microenvironments. *J. Air Waste Manag. Assoc.* **2014**, *64* (7), 743–758. https://doi.org/10.1080/10962247.2013.872708.

(53) Hagler, G. S. W.; Baldauf, R. W.; Thoma, E. D.; Long, T. R.; Snow, R. F.; Kinsey, J. S.; Oudejans, L.; Gullett, B. K. Ultrafine Particles near a Major Roadway in Raleigh, North Carolina: Downwind Attenuation and Correlation with Traffic-Related Pollutants. *Atmos. Environ.* **2009**, *43* (6), 1229–1234. https://doi.org/10.1016/j.atmosenv.2008.11.024.

(54) Hitchins, J.; Morawska, L.; Wolff, R.; Gilbert, D. Concentrations of Submicrometre Particles from Vehicle Emissions near a Major Road. *Atmos. Environ.* **2000**, *34* (1), 51–59. https://doi.org/10.1016/S1352-2310(99)00304-0.

(55) Peters, J.; Van den Bossche, J.; Reggente, M.; Van Poppel, M.; De Baets, B.; Theunis, J. Cyclist Exposure to UFP and BC on Urban Routes in Antwerp, Belgium. *Atmos. Environ.* **2014**, *92*, 31–43. https://doi.org/10.1016/j.atmosenv.2014.03.039.

(56) Van den Bossche, J.; Peters, J.; Verwaeren, J.; Botteldooren, D.; Theunis, J.; De Baets, B. Mobile Monitoring for Mapping Spatial Variation in Urban Air Quality: Development and Validation of a Methodology Based on an Extensive Dataset. *Atmos. Environ.* **2015**, *105*, 148–161. https://doi.org/10.1016/j.atmosenv.2015.01.017.

(57) Weijers, E. P.; Khlystov, A. Y.; Kos, G. P. A.; Erisman, J. W. Variability of Particulate Matter Concentrations along Roads and Motorways Determined by a Moving Measurement Unit. *Atmos. Environ.* **2004**, *38* (19), 2993–3002. https://doi.org/10.1016/j.atmosenv.2004.02.045.

(58) Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A Review of Land-Use Regression Models to Assess Spatial Variation of Outdoor Air Pollution. *Atmospheric Environment*. October 2008, pp 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057.

(59) Armstrong, B. G. Effect of Measurement Error on Epidemiological Studies of Environmental and Occupational Exposures. *Occupational and Environmental Medicine*. 1998, pp 651–656. https://doi.org/10.1136/oem.55.10.651.

(60) Szpiro, A. A.; Paciorek, C. J.; Sheppard, L. Does More Accurate Exposure Prediction Necessarily Improve Health Effect Estimates? NIH Public Access. *Epidemiology* **2011**, *22* (5), 680–685. https://doi.org/10.1097/EDE.0b013e3182254cc6.

Appendices

EN Summary

Ultrafine particles (UFP) and black carbon (BC) are increasingly associated with adverse health effects. While epidemiological research is not conclusive, toxicological studies to UFP have shown that they are more toxic per mass unit than the larger fractions of particulate matter. UFP and BC can enter the lung alveoli, with UFP able to penetrate deeper into the lungs, cross biological barriers, enter the bloodstream and reach other organs. In this thesis, I assessed the spatial and spatiotemporal variation of UFP and BC, mainly in urban environments.

As UFP and BC are highly variable in space and time, they are difficult to monitor in distributed networks. It would be too expensive to set up a network that fully captures the spatial variability of UFP and BC. These pollutants are therefore increasingly measured with mobile platforms. Mobile monitoring provides the possibility to sample more spatially diverse environments in less time, with a limited number of (costly) monitoring devices. Together with advancements in air monitoring instrumentation, such as higher time resolution and greater portability, these platforms can capture the high variability of UFP and BC in space and time in a complex urban terrain.

To facilitate epidemiological research on UFP and BC and adverse health effects, there is a need for exposure surfaces with long-term exposure estimates. In this dissertation I used mobile monitoring covering about 20.000 street segments and 800 short-term stationary sites (3x30 minutes) to develop fine resolution UFP and BC maps. In chapter 2 I show that models used to develop these maps (mobile and short-term stationary) predicted highly correlated (R^2 of 0.89 for UFP and BC) concentrations on address level. Must be said that models based on mobile measurements have a low model performance. However, in chapter 3 I show that the low model performance (R^2 of 15%) for UFP, can predict measurements with longer averaging time substantially better. Explained variance was 36% for short-term stationary measurements and 57% for the longer-term home outdoor measurements (3 times 24 hours, 42 sites). In contrast, the mobile BC model only predicted 14% of the variation in the short-term stationary sites and also 14% of the home outdoor sites. I also found higher UFP predictions (of about 30%) based on mobile models opposed to short-term model predictions and home outdoor measurements with no clear geospatial patterns.

Next, in chapter 4, I used two training sets of UFP data (mobile and short-term stationary measurements) to evaluate different modelling approaches to estimate long-term UFP concentrations by estimating precision and bias based on an independent external data set, the same as the long-term home outdoor measurements used in chapter 3. I showed that machine learning algorithms trained on mobile measurements explained only 38–47% of external UFP concentrations, whereas multivariable methods like stepwise regression and elastic net explained 56–62%. Some machine learning algorithms (bagging, random forest) trained on short-term measurements explained modestly more variability of external UFP concentrations compared to multiple linear regression and regularized regression techniques. Differences in predictive ability of algorithms seem to be small and may depend on the type of training data used.

Finally, I combined the lessons learned from the previous analyses and designed a measurement campaign to characterise the spatial variability of UFP in the entire Netherlands in chapter 5 and 6. To do this, long-term stationary measurements at regional background locations were combined with mobile monitoring, covering the Netherlands. I showed that different regions of the country experience substantially different concentrations due to variation in regional background concentration while the temporal variation across the Netherlands was fairly similar. Average temporal correlation between 2-hour and 24-hour average UFP concentrations across the 20 regional background sites was 0.50 (IQR: 0.36-0.61) and 0.58 (IQR: 0.44-0.75), respectively, and declining weakly by distance. This showed us that LUR models for UFP are scalable to a national extent using mobile monitoring. We were also able to create a map for the regional background UFP concentration in the Netherlands based on kriging.

In chapter 6, I developed robust nation-wide models for long-term UFP exposure, with minor differences in predictive performance between three different algorithms (stepwise, LASSO, random forest). For each approach, I also tested a deconvolution method, which segregates the average concentration at each road segment into a local and background signal. All models based on the deconvolution method included the regional background kriging map (from chapter 5) as important predictor. Correlations between predictions at random addresses were high with Pearson correlations from 0.84 to 0.99.

In chapter 7 (Section A to G) I highlight several important aspects in this thesis:

A

The BC instrument (AE51) that was used in our mobile monitoring campaigns was not able to maintain the temporal resolution needed in a true mobile setup. This is because the instrument was not able to distinguish

measurements from noise on a small time period. Measurements had to be averaged over multiple minutes when driving, especially on residential streets. So, the instrument was not able to capture the high variability of BC concentrations in an urban environment. Solutions would be to increase the sampling time per road segment or use devices that have less noise and can therefore have smaller time resolutions.

B

An important finding in all our campaigns was that models with low R^2 values for mobile monitoring can result in high R^2 values when applied to long-term stationary measurements. Mobile measurements consist of a few seconds per segment on a certain time and day. Because the goal is to assess long-term exposure, I only assessed the spatial variation (between-segments variation) and do not need to explain the within-day and between-days variation. The total variation our models were able to explain is therefore restricted to the spatial variation, which explains the low R^2 values in mobile monitoring campaigns. The reason why it is possible to develop robust spatial models based on mobile monitoring is the fact that multiple road segments in a city can serve as repeats because they have similar characteristics, regarding traffic intensity and surrounding land use. This means that segments might not have actual repeats, but when the street topography is very similar, they have pseudo repeats. Individually they are not stable (hence the low R^2), but combined they hold enough spatial information to create robust LUR models.

C

LUR models based on mobile monitoring were robust over time and scale. Model predictions for 1000 random addresses in the Netherlands from all three different campaigns correlated very highly to each other. Variables selected in all three supervised stepwise regression models were fairly similar in the sense that all models included a population density predictor in 5000-meter buffer, the presence of a port in the neighbourhood and local traffic variables.

D

Predictions made with models that were developed with machine learning were slightly different to predictions that were made with linear regression models, both for mobile and short-term measurements. The downside of machine learning is that such algorithms are generally difficult to understand. This is not a big problem if the only goal of the model is to give the best possible prediction, but it will not tell you as easily as simple regression techniques what the contribution is from all factors is on the spatial variation of air pollution.

E

Measurements are not taken on the exact same moment in time and have to be temporally corrected in order to be compared. This is because concentrations levels change throughout the course of a day/week. For presenting individual measurements temporal correction is very important as to avoid biases due to temporal variance. However, when mobile measurements are used to develop LUR models, I argue that the high number of similar segments measured over multiple days serve as pseudo repeats. When these similar segments are measured on different times of the day, week and year 'unbiased' spatial maps can be produced.

F

There are several main design factors to consider in a mobile monitoring campaign for measuring traffic-related air pollutants. First, the number of measurement sites; when developing a LUR model this does not need full coverage due to pseudo repeats of similar segments. Second, the duration of individual measurements; in a mobile setup this will always be in the order of seconds to a minute, depending on the mode of transport. Therefore, attention must be paid to the fact that similar segments are measured on different times of the day (daytime, night-time, rush hour), different days of the week and different seasons. Third, the number of replications; more repeats will result in higher training R^2 values, but I showed that with a limited number of repeats (often 2-3 drive days per segment) models can developed that create robust external predictions.

G

Mobile monitoring campaigns face spatial misalignment, as car-based mobile measurements are done on the middle of the road and prediction sites are at people's home address. I found that models based on mobile monitoring were correlated moderately high with home outdoor long-term average concentrations, but overpredicted measurements by about 20-30%. This was irrespective of the distance from the road to a home

address, or whether the home was located on a major or minor road. In addition, short-term stationary models did not over-predict long-term concentrations. Overprediction can therefore be attributed to on-road driving itself. This systematic bias will likely not influence correlations between air pollution and adverse health-effects but is important to consider when absolute levels are of interest.

In conclusion, mobile monitoring offers additional valuable insights into traffic-related air pollutants like UFP and BC, which are highly variable in space and time. Using a mobile platform enabled us to measure a large number of different exposure settings in a study area with a minimal amount of costly measurement devices. Despite the limited time spent on a road segment I found that it is possible to predict long-term exposure with LUR models based on mobile measurements. Validation R^2 values are equally good as LUR models based on short-term stationary measurements. On top of that, only a limited number of road segments are needed to develop LUR models, as long as there is a large variation in land use and urban topography. It is therefore possible to scale a mobile monitoring design to different cities and/or entire countries. This can greatly help in exposure assessment in health effects studies to UFP.

(NL) Samenvatting

Verschillende toxicologische studies hebben aangetoond dat ultrafijne stofdeeltjes (Ultrafine Particles; UFP) en roet (Black Carbon; BC) slecht zijn voor je gezondheid. Deze stoffen zijn giftiger dan grotere deeltjes omdat ze je longblaasjes kunnen bereiken en UFP kan zelfs in je bloedbaan en andere organen terechtkomen. Er kunnen echter nog weinig conclusies uit epidemiologische studies worden getrokken. Hiervoor dient eerst de langdurige blootstelling van de bevolking aan UFP en BC in kaart gebracht te worden. In dit proefschrift analyseer ik de ruimtelijke (geografische) en temporele (verschillende momenten van de dag, week en jaar) variatie van UFP en BC, met een focus op de stedelijke omgeving.

UFP en BC-concentraties zijn heel variabel in tijd en ruimte. Dit maakt ze moeilijk in kaart te brengen met een netwerk van meetstations. Om alle variatie in UFP en BC op straatniveau in kaart te brengen wordt er steeds meer gebruik gemaakt een mobiel platform en al rijdende de luchtkwaliteit te meten. Deze aanpak is efficiënt en financieel aantrekkelijk. Immers wordt in korte tijd en met een beperkt aantal (dure) meetapparatuur meer data verzameld en daardoor ontstaat er een completer beeld van de luchtkwaliteit. Met de gebruikte mobiele meetmethode, bestaande uit 20.000 verschillende wegsegmenten en 800 korte stationaire metingen (3x30minuten), zijn UFP en BC-kaarten met hoge resolutie ontwikkeld.

Uit analyses in hoofdstuk 2 blijkt dat er op adresniveau een sterke correlatie is tussen voorspelde concentraties van modellen gemaakt met mobiele metingen en de modellen op basis van korte stationaire metingen (R^2 = 0.89 voor UFP en BC). De kanttekening hierbij is dat modellen gebaseerd op mobiele metingen, de mobiele metingen niet goed kunnen verklaren (R^2=0.15). In hoofdstuk 3 laat ik echter zien dat het UFP model metingen met langere middelingstijd substantieel beter kan verklaren. Zo kan het model gebaseerd op mobiele metingen, de korte stationaire metingen voor 36% en de lange-termijn metingen (3x24uur op 42 plekken) voor 57% verklaren. In tegenstelling tot het model dat gebaseerd is op de BC-metingen. Dat kon maar 14% van korte stationaire en 14% van lange-termijn metingen verklaren. In deze analyse kwam ook duidelijk naar voren dat modellen op basis van mobiele metingen hogere UFP-concentraties (ongeveer 30%) voorspellen dan modellen op basis van korte stationaire metingen, zonder duidelijke geografische patronen.

Vervolgens, in hoofdstuk 4, vergelijk ik verschillende methodes om een model te ontwikkelen om langdurige UFP-bootstelling te voorspellen. Hiervoor gebruik ik twee verschillende datasets, een met de mobiele

metingen en de andere met korte stationaire metingen. 'Machine learning' algoritmes ontwikkeld met mobiele metingen voorspellen de externe langdurige metingen met 38-47%, terwijl de lineaire algoritmes, zoals 'stepwise regression' en 'elastic net', de langdurige metingen met 26-62% verklaren. Sommige 'machine learning' algoritmes ('bagging', 'random forest') ontwikkeld met korte stationaire metingen waren iets beter in staat de langdurige UFP-metingen te verklaren, vergeleken met lineaire regressie technieken. Verschillen tussen de verschillende methodes zijn klein. Ook hangt het ervan af welke data gebruikt wordt om het model te ontwikkelen.

Tenslotte is er meetcampagne opgezet om de ruimtelijke variatie van UFP in heel Nederland in kaart te brengen. Hiervoor zijn er naast mobiele metingen ook langdurige stationaire metingen op regionale achtergrond locaties verspreid over heel Nederland verricht. Op basis van kriging is hiermee in hoofdstuk 5 een kaart gemaakt met de regionale achtergrondconcentratie van UFP in Nederland. In dat hoofdstuk laat ik ook zien dat verschillende regio's substantieel verschillende UFP-concentraties hebben, terwijl de temporele variatie (weersomstandigheden) door heel Nederland ongeveer gelijk blijft. De gemiddelde temporele correlatie tussen 2-uurs en 24-uurs concentraties op 20 regionale achtergrondlocaties was respectievelijk 0.50 (IQR: 0.36-0.61) en 0.58 (IQR: 0.44-0.75). Dit laat zien dat UFP-modellen opgeschaald kunnen worden naar een nationaal niveau door gebruik te maken van mobiele monitoring.

In hoofdstuk 6 ontwikkelde ik robuuste nationale modellen voor lange-termijn blootstelling van UFP, met weinig verschil tussen de toegepaste algoritmes ('stepwise', 'LASSO' en 'random forest'). In dit hoofdstuk evalueerde ik ook voor elk algoritme een model om de lokale invloed van UFP van de achtergrondconcentratie te scheiden. De drie verschillende modellen voor de achtergrondconcentratie beoordelen de regionale achtergrondkaart uit hoofdstuk 5 als belangrijke input. De voorspelde concentratie op willekeurige adressen in Nederland van alle zes modellen zijn sterk gecorreleerd (r = 0.84 tot 0.99).

In hoofdstuk 7 (Sectie A tot en met G) behandel ik uitgebreid een aantal belangrijke aspecten in dit proefschrift:

A

Het BC-meetinstrument (AE51) dat we hebben gebruikt in de mobiele meetcampagnes was niet stabiel genoeg om de temporele resolutie te verkrijgen die nodig is voor mobiele metingen. Het meetapparaat bleek niet in staat om de metingen van ruis te onderscheiden in een kort tijdsbestek. Om de ruis weg te nemen moesten metingen gemiddeld worden over meerdere minuten, vooral op plekken met lage concentraties, zoals straten in woonwijken. Hierdoor kon het apparaat de hoge variatie van BC-concentraties in stedelijk gebieden niet vastleggen. Oplossingen zouden kunnen zijn om de meettijd per wegsegment te vergroten of meetapparatuur te gebruiken met minder ruis en dus kortere metingen aankan.

B

Een belangrijke bevinding in al onze meetcampagnes was dat modellen gebaseerd op mobiele metingen lage voorspellende scores (R^2) hebben, maar hoge R^2 scores hebben als ze vergeleken worden met lange termijn stationaire metingen. Mobiele metingen duren maar een aantal seconden per wegsegment op een bepaalde dag en tijd. De doelstelling is echter het in kaart brengen van lange-termijn blootstelling en niet de variatie die binnen een dag of tussen dagen plaatsvindt. De totale variatie die de modellen kunnen verklaren is dus gelimiteerd tot de ruimtelijk variatie. Dit verklaart ook de lage R^2 waardes in mobiele meetcampagnes. De reden dat er toch robuuste modellen ontwikkeld kunnen worden met mobiele metingen ligt aan het feit dat meerdere wegsegmenten met vergelijkbare karakteristieken (bijvoorbeeld qua verkeersintensiteit) als herhaalde meting kunnen dienen. Dit betekent dat wegsegmenten geen echte herhaalde metingen hebben, maar wanneer de straatconfiguratie vergelijkbaar is, pseudo-herhaalde metingen hebben. Individueel zijn ze niet stabiel genoeg (dus lage R^2), maar samen bezitten ze genoeg ruimtelijke informatie om robuust modellen te maken.

C

Modellen op basis van mobiele metingen zijn stabiel over tijd en schaal. Voorspelde concentraties voor 1000 willekeurige adressen in Nederland van de 3 verschillende meetcampagnes correleerden sterk aan elkaar. Ook factoren die de variatie van UFP het meeste verklaren lijken in alle drie modellen op elkaar. Alle modellen beschouwen de populatiedichtheid binnen een straal van 5000 meter, de aanwezigheid van een haven en lokale verkeersintensiteit als belangrijke factoren.

D

Voorspellingen van 'machine learning' algoritmes verschillen een beetje met voorspellingen van lineaire regressie methoden, zowel voor mobiele als korte stationaire metingen. Nadeel van 'machine learning' algoritmes is dat

deze over het algemeen moeilijker te begrijpen zijn. Dit hoeft geen probleem te zijn als het geven van de best mogelijke voorspelling de enige doelstelling is, maar het is moeilijker om de specifieke contributie op de ruimtelijke variatie van luchtverontreiniging is van de verschillende factoren te achterhalen.

E

Metingen zijn niet gedaan op precies hetzelfde moment en moeten daarom gecorrigeerd worden om met elkaar vergeleken te worden. Dat komt omdat concentraties variëren gedurende de dag en week. Als individuele metingen gepresenteerd worden is het belangrijk om de meting temporeel te corrigeren om de temporele bias te vermijden. Echter, als de mobiele metingen gebruikt worden om modellen te ontwikkelen, stel ik dat het hoge aantal gereden wegsegment met vergelijkbare straatconfiguratie ook hier kunnen dienen als pseudo-herhaalde metingen. Als deze metingen op verschillende momenten van de dag, week en jaar gedaan worden kun je alsnog kaarten produceren zonder structurele afwijking.

F

Er zijn verschillende aspecten om rekening mee te houden bij het mobiel meten van verkeers-gerelateerde luchtverontreiniging. Ten eerste, het aantal meetlocaties; als er een model ontwikkeld wordt is het door de pseudo-herhaalde metingen niet nodig om het gehele domein te bemeten. Ten tweede, de duur van individuele metingen; in een mobiele campagne zal dit altijd ergens tussen seconden en minuten liggen, afhankelijk van manier waarop de mobiele meting wordt uitgevoerd. Daarom moet er ook gelet worden op het feit dat de metingen plaatsvinden op verschillende momenten van de dag (overdag, nachts en spitstijd), verschillende dagen van de week en verschillende seizoenen. Ten derde, het aantal herhalingen; hoe vaker een wegsegment is gemeten, hoe hoger de R^2 waarde zal zijn. Maar zelfs met een beperkt aantal herhaalde metingen (vaak 2 à 3 meetdagen per wegsegment) is het mogelijk om een model te ontwikkelen dat robuuste voorspellingen kan doen.

G

Omdat mobiele metingen midden op weg uitgevoerd worden verschillen ze van de plek waar de voorspelling voor gedaan dient te worden, namelijk iemand zijn woon of werk adres. Ik heb laten zien dat modellen gebaseerd om mobiele metingen redelijk hoog correleerden met lange termijn concentraties op woonadressen, maar wel 20 tot 30% hogere concentraties voorspelden. Dit was onafhankelijk van de afstand tot de weg en of het adres zich op een drukke doorgaande weg of woonwijk bevond. Daarbij komt dat het model op basis van korte stationaire metingen de lange-termijn metingen niet overschatte. De overschatting van mobiele modellen kunnen we daarom toeschrijven aan het meten midden op weg. Deze afwijking zal waarschijnlijk de correlatie tussen luchtvervuiling en gezondheidseffecten niet beïnvloeden, maar is belangrijk als er gekeken wordt naar de absolute concentratie.

Mijn conclusie is dat mobiele monitoring waardevolle inzichten in verkeers-gerelateerde luchtverontreiniging biedt. Vooral bij stoffen zoals UFP en BC, die heel variabel zijn in tijd en ruimte. Een mobiel platform kan een grote blootstelling variatie in een bepaald studiegebied bemeten met een beperkt aantal kostbare meetapparaten. Ondanks de beperkte meettijd per wegsegment, stel ik dat het toch mogelijk is om mobiele metingen te gebruiken om langdurige blootstelling aan UFP te voorspellen. R^2 waardes zijn even goed als modellen die gebaseerd zijn om korte stationaire metingen. Bovendien is er slechts een beperkt aantal wegsegmenten nodig om deze modellen te ontwikkelen, zolang er maar een grote variatie in land gebruik en stedelijke topografie is. Het is daarom mogelijk om campagnes met mobiele metingen op te schalen naar verschillende steden en of hele landen. Dit kan enorm helpen bij de onderzoeken naar de relatie tussen blootstelling aan UFP en gezondheidseffecten.

Affiliation of contributors

Gerard Hoek. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands.

Roel Vermeulen. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands. & Julius Center for Health Sciences and Primary Care, University Medical Center, University of Utrecht, 3584 CK Utrecht, The Netherlands.

Bert Brunekreef. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands.

Kees Meliefste. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands.

Ulrike Gehring. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands.

Lützen Portengen. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands.

Esther van de Beek. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands.

Jochem Klompmaker. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands. & National Institute for Public health and the Environment, RIVM, 3720 BA Bilthoven, The Netherlands.

Jelle Vlaanderen. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands.

John Gulliver. MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, St Mary's Campus, London, United Kingdom.

Erik van Nunen. Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands.

Kyle Messier. Dept. of Civil, Architectural and Environmental Engineering, University of Texas at Austin, USA & Environmental Defense Fund, Austin, TX, USA

Geert Sterk. Department of Physical Geography, Utrecht University, 3508 TC Utrecht, The Netherlands.

List of publications

Kerckhoffs, J.; Wang, M.; Meliefste, K.; Malmqvist, E.; Fischer, P.; Janssen, N. A. H.; Beelen, R.; Hoek, G. A National Fine Spatial Scale Land-Use Regression Model for Ozone. *Environ. Res.* **2015**, *140*, 440–448. https://doi.org/10.1016/j.envres.2015.04.014.

Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of Ultrafine Particles and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environ. Sci. Technol.* **2016**, *50* (23), 12894–12902. https://doi.org/10.1021/acs.est.6b03476.

Kerckhoffs, J.; Hoek, G.; Vlaanderen, J.; van Nunen, E.; Messier, K.; Brunekreef, B.; Gulliver, J.; Vermeulen, R. Robustness of Intra Urban Land-Use Regression Models for Ultrafine Particles and Black Carbon Based on Mobile Monitoring. *Environ. Res.* **2017**, *159*, 500–508. https://doi.org/10.1016/j.envres.2017.08.040.

Downward, G. S.; van Nunen, E. J. H. M.; Kerckhoffs, J.; Vineis, P.; Brunekreef, B.; Boer, J. M. A.; Messier, K. P.; Roy, A.; Verschuren, W. M. M.; van Der Schouw, Y. T.; Sluijs, I.; Gulliver, J.; Hoek, G.; Vermeulen, R. Long-Term Exposure to Ultrafine Particles and Incidence of Cardiovascular and Cerebrovascular Disease in a Prospective Study of a Dutch Cohort. *Environ. Health Perspect.* **2018**, *126* (12). https://doi.org/10.1289/EHP3047.

Messier, K. P.; Chambliss, S. E.; Gani, S.; Alvarez, R.; Brauer, M.; Choi, J. J.; Hamburg, S. P.; Kerckhoffs, J.; Lafranchi, B.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Roy, A.; Szpiro, A. A.; Vermeulen, R. C. H.; Apte, J. S. Mapping Air Pollution with Google Street View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression. *Environ. Sci. Technol.* **2018**, *52* (21), 12563–12572. https://doi.org/10.1021/acs.est.8b03395.

Kerckhoffs, J.; Hoek, G.; Portengen, L.; Brunekreef, B.; Vermeulen, R. C. H. Performance of Prediction Algorithms for Modeling Outdoor Air Pollution Spatial Surfaces. *Environ. Sci. Technol.* **2019**, *53* (3), 1413–1421. https://doi.org/10.1021/acs.est.8b06038

van de Beek, E; Kerckhoffs, J; Hoek, G; Sterk, G; Meliefste, K; Gehring, U; Vermeulen, R.C.H. Spatial and Spatiotemporal Variability of Regional Background Ultrafine Particle Concentrations in The Netherlands. *Environ. Sci. Technol.* **2021**, *55* (2), 1067–1075. https://doi.org/10.1021/acs.est.0c06806

Kerckhoffs, J; Hoek, G; Gehring, U; Vermeulen, R.C.H. Modelling Nationwide Spatial Variation of Ultrafine Particles based on Mobile Monitoring. *Environment International.* **2021**, *154* (9). https://doi.org/10.1016/j.envint.2021.106569

Curriculum Vitae

Jules Kerckhoffs was born on 28 September 1988 in Maastricht, The Netherlands. In 2007, he started his bachelor studies “Environmental Sciences” in Utrecht, The Netherlands, followed by the master “Toxicology and Environmental Health” at the Institute of Risk Assessment Sciences (IRAS). Part of the master was a 6-month internship on chemical usage in The Bahamas at the Pan American Health Organization (PAHO) in Nassau, The Bahamas. His major 9-month research project was on the development of a national Land-Use Regression (LUR) model for Ozone in the Netherlands. After working as research assistant at IRAS, he started his PhD in 2016 analyzing the possibilities to create (Dutch) air pollution LUR models based on measurements collected with a mobile platform. The results are of this are described in this thesis. Later, his PhD also involved the Google Airview project, where Google Streetview cars were equipped with air quality instruments and drove extensively in Amsterdam and Copenhagen. From June 1st, 2021, he will be employed as PostDoc at the institute of Risk Assessment Sciences, continuing his work on the Google Airview project, among others.

Dankwoord

Wat een avontuur! Het proefschrift is af!

En dat ging niet zonder slag of stoot. Afgebroken meetapparatuur, onvoldoende werkgeheugen voor een hyperparameter berekening in een randomForest model en voetballende Italianen: je komt onderweg allerlei obstakels tegen. Om nog maar te zwijgen over het uitvoeren van een mobiele meetcampagne zonder rijbewijs.

Met veel plezier kijk ik terug op de afgelopen vijf jaar, die erg snel voorbij zijn gegaan. Het was een leerzame ervaring. Ik heb verschillende congressen (Rome, Ottawa, Utrecht) mogen bijwonen en ik mocht deelnemen aan een uitwisselingsprogramma met de Universiteit van Toronto. Hier, en gedurende de hele periode op het IRAS, heb ik veel leuke en interessante mensen mogen ontmoeten. In grotere en kleinere mate hebben allen een bijdrage geleverd aan dit proefschrift. Daarvoor wil ik jullie allemaal graag bedanken.

Als eerste wil ik mijn vrouw Celine bedanken. Met jou is het altijd een feest. Door jou heb ik altijd een zeer plezierige balans tussen werk en privé kunnen houden. Met je oog voor detail en scherpe vragen heb je zeker ook een bijdrage kunnen leveren aan dit proefschrift, tot aan dit dankwoord aan toe. Ik hou van je.

Tevens bedank ik mijn ouders voor hun onvoorwaardelijke steun. Ook al kijken jullie me met grote ogen aan als ik weer eens over mijn onderzoek spreek, jullie steun om verder te leren heb ik altijd gewaardeerd. Hopelijk weten Celine en mijn ouders inmiddels waar de chemische symbolen 0_3 en NO_2 voor staan. Toch schrijf ik het voor de zekerheid nog een keer volledig uit: Ozon is 0_3 en stikstofdioxide is NO_2.

Wellicht dat deze toevoeging ook handig is voor andere familieleden en vrienden. Al weten zij natuurlijk precies waar ik allemaal mee bezig ben: als boswachter ben je bezig met in kaart brengen van de luchtkwaliteit.

Graag bedank ik Frank voor de lay-out en design van mijn proefschrift. Je hebt mijn ideeën op smaak gebracht met een pittig grafisch ontwerp. Ik ben dan ook zeer tevreden over het resultaat.

Dan wil ik natuurlijk de mensen binnen het IRAS bedanken, te beginnen met alle veldwerkers die alle data hebben verzameld. Met name bedank ik Jochem, Daan en Ruud voor het besturen van de TANK. Mijn dank heb ik

overigens al een aantal keren eerder aan jullie getoond door jullie te laten winnen met squash.

En natuurlijk Kees. Ik heb altijd met volle teugen genoten van jouw aanstekelijk enthousiasme. Of dat nu was tijdens de opzet van alle meetcampagnes, tijdens de gezellige lange autoritten voor mijn masteronderzoek over 03 of tijdens onze tripjes in Kopenhagen. Het was altijd leuk. Ik heb ook veel van jou geleerd: de muziek van vroeger is absoluut beter; ik dien mijn data om de paar seconden op te slaan en als ik een computerprobleem heb, dien ik deze gewoon opnieuw op te starten. Hierna zijn vanzelf alle problemen opgelost.

Ook kijk ik met veel plezier terug op alle momenten met IRAS-collega's tijdens spelletjes, zaalvoetbal en PhD- en IRAS-outings. In het bijzonder dank ik mijn kamergenoten. In het begin met Daan, die me toch heeft uit mijn schulp getrokken heeft en me heeft leren praten. De spreek-mok staat nog steeds op mijn bureau. Daarna samen met Marije, Joseph, Edith, Calvin, Daniel, Anran en Ilse op kamer 319b. Naast het harde werken, was er ook altijd tijd voor een leuk gesprek. Of dat nu op kantoor zelf was of tijdens een wandeling, BBQ of bierproeverij. Gelukkig blijft ons koffieclubje nog bestaan.

Als laatste wil ik mijn promotoren Gerard en Roel bedanken voor de kans die zij mij geboden hebben om een promotieonderzoek te beginnen. Dankzij jullie wijsheid hebben jullie mij altijd op de goede route weten te houden. Gerard; vanaf mijn masterscriptie tot aan mijn PostDoc-werkzaamheden ben je al mijn supervisor. Ik waardeer het dat jouw deur altijd open staat voor vragen en discussies. Roel; je had altijd het grote plaatje helder in beeld en hebt me zodoende met deskundige input vaak op weg geholpen. Ook heb ik het altijd fijn gevonden dat jullie mij de ruimte en het vertrouwen hebben gegeven om mijn eigen plan te trekken, zelf alle resultaten mocht presenteren en de vrijheid om dit proefschrift iets populairder wetenschappelijk aan te pakken.

En eigenlijk moet ik nog iemand bedanken die me het vaakst geholpen heeft tijdens mijn PhD: Stackoverflow bedankt!