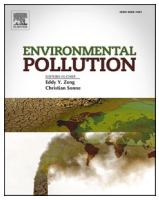
Environmental Pollution 346 (2024) 123664

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol



Harnessing AI to unmask Copenhagen's invisible air pollutants: A study on three ultrafine particle metrics☆

Heresh Amini [a,b,c,*], Marie L. Bergmann [b], Seyed Mahmood Taghavi Shahri [b], Shali Tayebi [b], Thomas Cole-Hunter [b], Jules Kerckhoffs [d], Jibran Khan [e,f], Kees Meliefste [d], Youn-Hee Lim [b], Laust H. Mortensen [g,b], Ole Hertel [h], Rasmus Reeh [i], Christian Gaarde Nielsen [j], Steffen Loft [b], Roel Vermeulen [d], Zorana J. Andersen [b], Joel Schwartz [c]

[a] Department of Environmental Medicine and Public Health, Institute for Climate Change, Environmental Health, and Exposomics, Icahn School of Medicine at Mount Sinai, New York, United States
[b] Department of Public Health, University of Copenhagen, Copenhagen, Denmark
[c] Department of Environmental Health, Harvard TH Chan School of Public Health, Boston, MA, United States
[d] Institute for Risk Assessment Sciences (IRAS), Utrecht University, the Netherlands
[e] Department of Environmental Science, Aarhus University, Roskilde, Denmark
[f] Danish Big Data Centre for Environment and Health (BERTHA), Aarhus University, Roskilde, Denmark
[g] Statistics Denmark, Copenhagen, Denmark
[h] Faculty of Technical Sciences, Aarhus University, Denmark
[i] UrbanDigital, Copenhagen, Denmark
[j] Copenhagen Solutions Lab, Copenhagen Municipality, Copenhagen, Denmark

ARTICLE INFO

Keywords:
Ultrafine particles (UFPs)
Lung deposited surface area (LDSA)
Average particle size (APS)
Machine learning (ML)

ABSTRACT

Ultrafine particles (UFPs) are airborne particles with a diameter of less than 100 nm. They are emitted from various sources, such as traffic, combustion, and industrial processes, and can have adverse effects on human health. Long-term mean ambient average particle size (APS) in the UFP range varies over space within cities, with locations near UFP sources having typically smaller APS. Spatial models for lung deposited surface area (LDSA) within urban areas are limited and currently there is no model for APS in any European city. We collected particle number concentration (PNC), LDSA, and APS data over one-year monitoring campaign from May 2021 to May 2022 across 27 locations and estimated annual mean in Copenhagen, Denmark, and obtained additionally annual mean PNC data from 6 state-owned continuous monitors. We developed 94 predictor variables, and machine learning models (random forest and bagged tree) were developed for PNC, LDSA, and APS. The annual mean PNC, LDSA, and APS were, respectively, 5523 pt/cm^3, 12.0 μm^2/cm^3, and 46.1 nm. The final R^2 values by random forest (RF) model were 0.93 for PNC, 0.88 for LDSA, and 0.85 for APS. The 10-fold, repeated 10-times cross-validation R^2 values were 0.65, 0.67, and 0.60 for PNC, LDSA, and APS, respectively. The root mean square error for final RF models were 296 pt/cm^3, 0.48 μm^2/cm^3, and 1.60 nm for PNC, LDSA, and APS, respectively. Traffic-related variables, such as length of major roads within buffers 100–150 m and distance to streets with various speed limits were amongst the highly-ranked predictors for our models. Overall, our ML models achieved high R^2 values and low errors, providing insights into UFP exposure in a European city where average PNC is quite low. These hyperlocal predictions can be used to study health effects of UFPs in the Danish Capital.

1. Introduction

Air pollution from various sources continues to produce enormous economic and health burden for populations across the world (Cohen et al., 2017; Gakidou et al., 2017; McDuffie et al., 2021). The burden of premature mortality attributable to air pollution has been estimated to

☆ This paper has been recommended for acceptance by Prof. Pavlos Kassomenos.

* Corresponding author. Department of Environmental Medicine and Public Health, Institute for Climate Change, Environmental Health, and Exposomics, Icahn School of Medicine at Mount Sinai, New York, United States.
E-mail address: heresh.amini@mssm.edu (H. Amini).

https://doi.org/10.1016/j.envpol.2024.123664
Received 30 August 2023; Received in revised form 8 February 2024; Accepted 25 February 2024
Available online 29 February 2024

be up to about 9 million deaths per year globally (Burnett et al., 2018). In 2016 alone, ambient air pollution has been estimated to be responsible for about US$4 trillion of economic cost (Egerstrom et al., 2023; Yin et al., 2021).

World Health Organization (WHO) has recommended lower values for most air pollutants in their recent guidelines (World Health Organization, 2021), and recommended more research on smaller particles, specifically particles less than 0.1 μm in aerodynamic diameter or ultrafine particles (UFP), which could be more harmful to health (Ohlwein et al., 2019). Since UFP has small mass compared to larger particles, it has been widely characterized by metrics such as particle number concentrations (PNC) in particles per cubic centimeter of air units (pt/cm^3), or in some studies by lung deposited surface area (LDSA) in squared micrometer of surface per cubic centimeter of air units (μm^2/cm^3) (Brugge and Fuller, 2021). WHO considered PNC levels of 10,000 pt/cm^3 (24-h mean) or 20,000 pt/cm^3 (1-h mean) to be low exposure (Goshua et al., 2022; World Health Organization, 2021). The organization did not recommend a 2021 guideline value for short- or long-term exposure to UFP as the epidemiological evidence for this pollutant was not sufficient, but instead recommended to "utilize emerging science and technology to advance approaches to the assessment of exposure to UFP for their application in epidemiological studies and UFP management" (World Health Organization, 2021).

The research on UFP and health has a challenge of accurate and precise data scarcity, which is mainly due to the fact that it is not a regulated pollutant, and is not part of the air quality monitoring program by regulatory networks in most countries (Bergmann et al., 2023a). Because studies on short-term health effects of UFP typically investigate temporal variations of air pollution and its health outcomes (and other possible temporally variable confounding covariates), such studies are more convenient and cheaper to conduct. Andersen et al. (2010) and Bergmann et al. (2023b) reported novel findings on the association between short-term exposure to UFP, stroke, and mortality in Copenhagen, Denmark (Andersen et al., 2010; Bergmann et al., 2023b). Stafoggia et al. (2017) reported a weak association between short-term exposure to UFP and mortality in eight urban areas (within Finland, Sweden, Denmark, Germany, Italy, Spain, and Greece) between 1999 and 2013 (Stafoggia et al., 2017). Research on short-term exposure to UFP and health continues to evolve with better exposure data, studying more health outcomes, and new study designs. However, it is well known that the health burden from long-term exposure to air pollution is far larger (about 10 times) than that from short-term exposure (Künzli et al., 2001). Research on long-term exposure to UFP is more limited mainly because long-term exposure estimates are less available, and it is well demonstrated that UFP varies over short spatial distances (i.e., in meters) from the producing sources and has short lifetime (Hoek et al., 2011; Saha et al., 2019b). A recent study investigated long-term exposure to UFPs and natural and cause-specific mortality, and reported that PNC was associated with lung cancer and premature natural mortality among adults in the Netherlands independently from other regulated air pollutants (Bouma et al., 2023).

Exposure scientists and environmental epidemiologists have tried to monitor and model long-term exposure to UFP in various ways, with land use regression (LUR) and dispersion models being widely used modeling approaches (Patton et al., 2021). Hoek et el. (2011) by monitoring and modeling UFP in Amsterdam for 2002–2004 using LUR methods (Hoek et al., 2011), and Zwack et al. (2011) by monitoring and modeling UFP in New York for 2007 using dispersion (Zwack et al., 2011a) and LUR methods (Zwack et al., 2011b), were among the first attempts in Europe and North America to characterize spatial and spatiotemporal variations of UFP for applications in environmental epidemiology studies. Since then, many research groups have tried to monitor and model UFP (mainly PNC) not only in North American and European cities (Abernethy et al., 2013; Cattani et al., 2016; Eeftens et al., 2016; Frohn et al., 2021; Kerckhoffs et al., 2021; Kerckhoffs et al., 2016; Ketzel et al., 2021; Lloyd et al., 2023; Patton et al., 2015; Saha et al., 2019a; Weichenthal et al., 2016; Wolf et al., 2017), but also in other parts of the world in Australia (Clifford et al., 2018; Karunasinghe et al., 2015; Rahman et al., 2020), India (Saraswat et al., 2013), China (Ge et al., 2022), or Taiwan (Chang et al., 2021). To our knowledge, only one spatial model exists for UFP metric LDSA in Switzerland (Eeftens et al., 2016), and one Canadian study developed models for mean particle size in the UFP range for Montreal and Toronto (Lloyd et al., 2023). More recently, mobile monitoring using Google Street View cars with intensive repeated day-time on-road measurements during weekdays has been used to develop novel models for UFP (Kerckhoffs et al., 2022a; Kerckhoffs et al., 2022b; Shah et al., 2023).

In this study, we aimed to develop very fine spatial resolution machine learning (ML) models to estimate long-term mean UFP metrics, namely PNC, LDSA, and average particle size (APS) in the capital city of Copenhagen, Denmark, using standard designed campaigns and technologies for monitoring and state-of-the-art techniques for modeling.

2. Materials and methods

2.1. Study area

The study area, Danish capital city of Copenhagen, is about 150 km^2 (Statistics Denmark, 2018) with a total population of >800,000 people (Worldpop, 2020). The city has a mild climate with an annual average temperature of ~9 °C. Copenhagen receives significant rainfall (annual average ~700 mm) (Liu and Jensen, 2017).

2.2. Monitored data

We conducted a monitoring campaign over one year in Copenhagen to estimate the annual mean of PNC, LDSA, and APS in selected locations. In brief, outdoor PNC, LDSA, and APS were monitored using miniature diffusion size classifiers ('DiSCmini' [DM]; Testo SE & Co. KGaA, Germany) at 27 residences' facades for approximate 72-h per site in two identical campaigns, each run in a warm or cool season (Fig. 1). We enlisted volunteers for residential measurements based on availability, ensuring a geographically representative spread across the study area. We did not intentionally oversample areas affected by specific sources, like major roads. The DM instruments measure PNC (1000–1,000,000 particles per cubic centimeter of air (pt/cm^3)), LDSA, and APS (range of 10–300 nm average particle size with an impactor for APS cut-off at 700 nm) at 1-s intervals. The exact number of DM devices used was three (one reference site and two rotating samplers). Indeed, when we refer to the APS as measured by the DiSCmini device, what we're actually referring to is the estimated modal diameter of the particles, not their precise size. This estimation is derived from a comparison of the electrical currents measured at two different stages within the device. The currents correspond to the number of particles collected at each stage, which in turn is related to the size of the particles. By comparing these currents, the device can predict the most common, or modal, particle diameter in the sample. This method provides a practical way to estimate particle size distribution in real-time, although it's important to note that it provides an estimate rather than a precise measurement.

The warmer season campaign was from July 08 to November 08, 2021, and the cooler season was from February 10 to May 29, 2022. To approximate the annual mean in each of the 27 sites, we additionally monitored PNC, LDSA, and APS at a reference site within a University of Copenhagen campus away from traffic sources from May 29, 2021, to May 29, 2022. The DM instruments were set up on the ground or first floor level, in house entrances, on windowsills, or balconies facing the street in weather-proof plastic boxes to represent concentration levels immediately close to the residences. Further details of the monitoring campaign are explained elsewhere (Bergmann et al., 2023a).

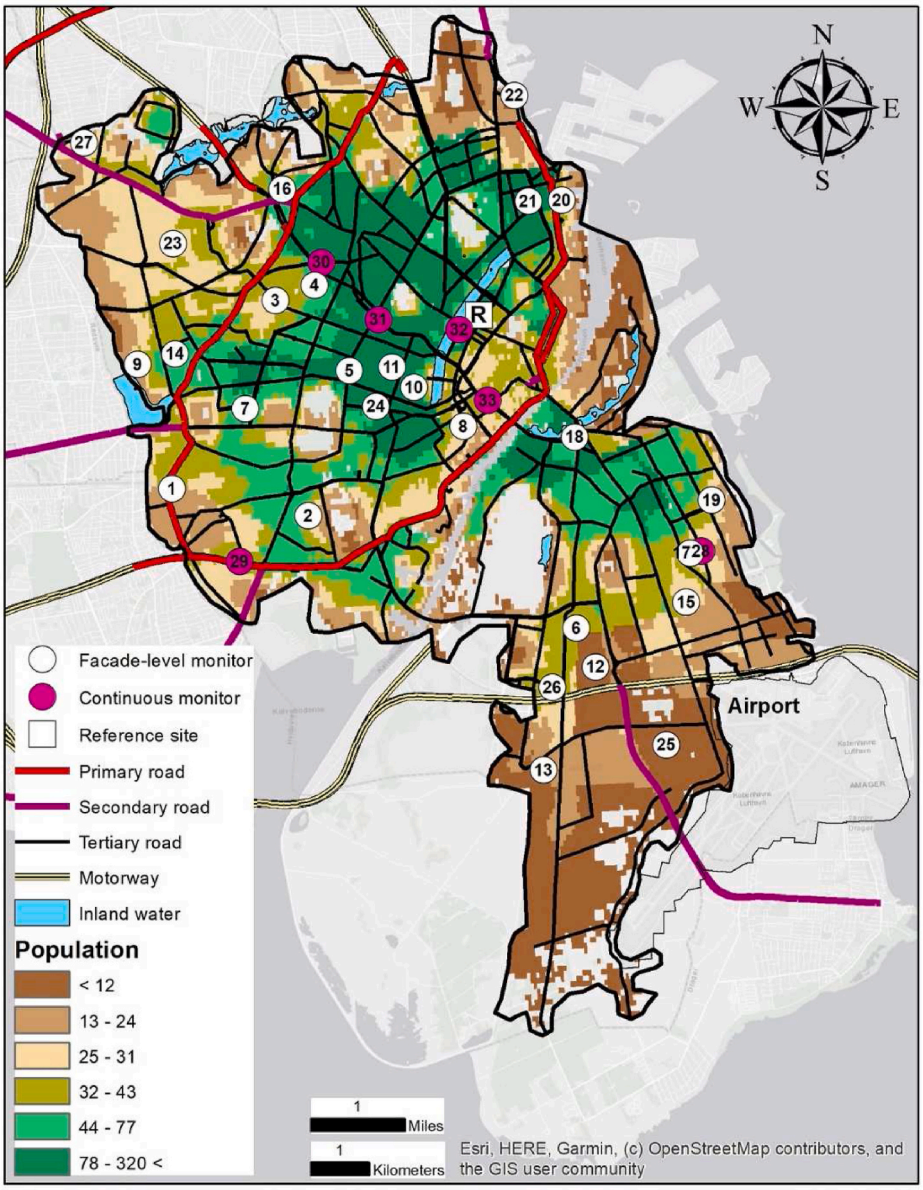


Fig. 1. The location of monitoring sites in Copenhagen, Denmark. Sites 1 to 27 were façade-level monitors that measured particle number concentration, lung deposited surface area and average particle size by DiSCmini instruments, and sites 28 to 33 were six state-owned continuous monitors. The population counts are per 100 m grid cells.

2.3. Quality assurance/control (QAQC)

We conducted 'zero checks' immediately before and after DM measurements using a HEPA filter, and also evaluated the accuracy and precision of the instruments to the best possible extent. To evaluate accuracy, we co-located our three DM instruments thrice ("Co-location 1": November 09–17, 2021, "Co-location 2": May 31-June 07, 2022, and "Co-location 3": August 23-September 06, 2022) with a regulatory network site that measured PNC using a Scanning Mobility Particle Sizer (SMPS) instrument. In these three occasions, we also evaluated precision by co-locating the DM instruments together and compared hourly means of PNC with each other. The monitoring period was not impacted by COVID-19 containment response policies (Bergmann et al., 2022; Bergmann et al., 2021). The QAQC procedures resulted in high accuracy and precision for our measurements.

2.4. Site-specific annual mean estimation

After careful data cleaning and processing, the annual means at each of the 27 sites were estimated using the average of ratio and difference methods based on the two short-term monitoring campaigns, and temporally adjusted using data from the reference site (Amini et al., 2017a; Amini et al., 2017b; Eeftens et al., 2015). The reference site was used to monitor the temporal variation of PNC, LDSA, and APS throughout the year. This temporal information was then used to adjust the measurements from the 27 sites, which were not monitored continuously, to approximate the annual mean. The spatial monitoring at the 27 sites was necessary to capture the spatial variation across the city. The combination of temporal data from the reference site and spatial data from the 27 sites allowed us to model the annual means at each of the 27 sites. More details about the temporal adjustment methods in our monitoring campaign is available elsewhere (Bergmann et al., 2023a). For PNC modeling, we additionally included the annual mean data of six state-owned continuous monitors that measured PNC

mostly near traffic sites from May 29, 2021, to May 29, 2022. The instruments used in five of these monitors were GRIMM 5421 condensation particle counter (CPC) that had a lower detection limit up to 7 nm to a greater limit of 3 μm and in one monitor it was General Scanning Mobility Particle Sizer (SMPS) with size range from 1 nm to 1000 nm.

2.5. Predictors

In total, 94 variables were developed as predictors in the model. These were in five main classes, namely population, satellite observation, land use, traffic, and distance to variety of features (Table S1). The population was the total number of people living within five buffers of the location (100, 200, 300, 400, and 500 m radii) sourced from Worldpop (2020). The satellite observation class included 6 predictors for normalized difference vegetation index (NDVI) from Landsat satellite for the year 2019 at ~30 m spatial resolution, also at different buffers (30, 100, 200, 300, 400, and 500 m radii). The land use class comprised of 10 predictors for the area of parks within 100, 200, 300, 400, and 500 m buffers, and additionally area of residential land use again within 100, 200, 300, 400, and 500 m buffer radii at a spatial resolution of 1 m. The traffic class included 31 predictors at a 1 m spatial resolution, such as 2017 annual average daily traffic, and length of various major roads in different categories and speed limits within buffers 100, 125, 150, 200, 250, 300, 350, 400, 450, and 500 m. Finally, the distance class included 42 predictors, also at a 1 m spatial resolution, such as distance to Open Street Map (OSM) major roads, major roads with different speed limits (40, 50, 60, 70, and 100 km/h), one way roads, two-way roads, primary, secondary, and tertiary roads, tunnels, bridges, airport, gas stations, parking lots, traffic signals, parking areas, schools, kindergartens, railways, taxi stations, ferry terminals, bus stops, bus stations, restaurants, café, or pubs, supermarkets, malls, hotels, parks, industries, inland water bodies, amongst others. Overall, the spatial resolution of 84 of 94 predictors (~90%) was 1 m.

2.6. Model development

The ML models were developed using the Caret package in R (Kuhn, 2008). Pre-processing of all predictors was conducted, and they were centered and scaled. Next, Random Forest (RF) and Bagged Tree (BT) ML algorithms were trained. These models were chosen based on prior experience on their very good performance for prediction of environmental data (Weichenthal et al., 2016). The details of these algorithms are explained elsewhere (Biau and Scornet, 2016; De'ath, 2007; Foley, 2020). Briefly, the RF model is created from multiple decision trees (DTs). The DTs determine the best split to subset the data. To create a low correlation forest of decision trees, the RF algorithm uses bagging and feature randomness (Biau and Scornet, 2016). Bagging, or Bootstrap Aggregating, is a method used to improve the stability and accuracy of machine learning algorithms. It works by creating multiple subsets of the original data, with replacement (meaning some samples may be repeated), and then training a separate model on each subset. The final output is typically the average of the predictions from each model. This method helps to reduce variance and prevent overfitting, making the model more robust and accurate. Bagging is commonly used in decision tree algorithms. In our work for RF model, the tuning parameter 'min.node.size' was held constant at a value of 5, and root mean square error (RMSE) was used to select the optimal model using the smallest value. The amount of granularity in the tuning parameter grid was refined by setting the tuneLength as 93 for RF. The final values used for the PNC RF model were mtry = 23 and splitrule = variance. The mtry is the number of variables randomly sampled as candidates at each split, and the splitrule determines the splitting rule (the options in the caret package are "gini", "extratrees", or "variance"). These for LDSA and APS were mtry = 2 and splitrule = extratrees.

The BT method in the caret package is a bagged CART model. It is a special case of RF models where B regression trees are built using B bootstrapped training sets and the resulting predictions are averaged. These trees grow deep and are not pruned. Therefore, each individual tree has high variance but low bias, and averaging the B trees reduces the variance (De'ath, 2007; Foley, 2020). No tuning was done for BT as Caret has no hyperparameters to tune with this model. For our BT model, B was the default value of 25.

Our ML models were cross-validated (10 fold, repeated 10 times). In other words, the data sample was shuffled in each repetition resulting in developing different splits of the sample data, and finally, the mean performance and accuracy across all 10 repetitions was calculated. The coefficient of determination (R^2) was calculated for the performance evaluation, and RMSE was calculated to measure accuracy. The RMSE was used to select the optimal model using the smallest value.

2.7. Performance evaluation

The final model predictions were further evaluation by visualizing the difference between monitored and predicted concentrations using 1:1 plots. Additionally, the dependence of differences between observed and predicted values on the mean of their associated values was assessed using Bland-Altman (BA) plots.

2.8. Predictions

Although most predictors (90%) were available at a 1 m spatial resolution, for computational efficiency and applicability of exposure prediction for each residence, prediction grid cells were created for centroid of 5 × 5 m grid cells. The total number of prediction grids in the study area was 4,116,529 grid cells and 94 predictors were sampled for each grid cell.

2.9. Software and high-performance computing (HPC)

Google Earth Engine was used to retrieve satellite observations (Gorelick et al., 2017), and R and RStudio environment was used for statistical analyses (Allaire, 2012; Ihaka and Gentleman, 1996), visualizations, and model development. In particular, the following packages were used: raster, rgdal, sf, sp, zoo, caret, caretEnsemble, mgcv, doParallel, future, ggExtra, ggpubr, BlandAltmanLeh, iml, and knitr. Features of ArcMap were used for geo-visualizations (Mitchel, 2005). For HPC, a supercomputer at Harvard University running a Linux-based operating system was used.

3. Results

3.1. Summary statistics and correlations for monitored data

The monitored annual mean PNC across 33 locations was 5523 (range: 3730–8975) pt/cm^3. The monitored annual mean LDSA across 27 locations was 12.00 (range: 9.6–14.5) μm^2/cm^3, while these for mean APS were 46.1 (39.2–55.5) nm (see more detailed statistics in Table 1). The Spearman correlation between monitored PNC and LDSA was 0.84, and both of these measures were negatively correlated with APS (−0.65 for PNC and APS, and −0.36 for LDSA and APS) (Fig. S1).

Table 1
Descriptive statistics for monitored annual mean particle number concentration (PNC) (pt/cm^3), lung deposited surface area (LDSA) (μm^2/cm^3), and average particle size (APS) (nm) in Copenhagen, Denmark.

UFP Metric	Observations	Min	25% ile	50% ile	Mean	75% ile	Max
PNC	33	3730	4741	5309	5523	6312	8975
LDSA	27	9.6	11.2	11.5	12.0	12.8	14.5
Particle size	27	39.2	44.3	45.3	46.1	47.0	55.5

3.2. Machine learning models

3.2.1. PNC model

The cross validated (10-fold, repeated 10 times) R^2 using the best RF model was 0.65 while for the BT model this was 0.71. However, the RMSE was smaller for RF model when predictions were regressed against the monitored PNC across 33 monitors (Table 2). The R^2 of the final selected RF model was 0.93. The predicted annual mean PNC by RF model across 4.1 million grid cells was 5573 (4272–7709) pt/cm^3 (Table 3 and Table S2). The top five main predictor variables for PNC were length of major roads within buffer of 125 m and 150 m, distance to bus stops, length of major roads within buffer of 100 m, and distance to traffic signals (Table 4, and see Table S5 for top 20 predictors).

3.2.2. LDSA model

The cross validated R^2 using the best RF model was 0.67 with an RMSE of 0.48 μm^2/cm^3, and for the BT model these were, respectively, 0.69 and 0.99 μm^2/cm^3 (Table 2). The R^2 of the final selected RF model was 0.88. The predicted annual mean LDSA by RF across 4.1 million grid cells was 12.00 (11.0–13.2) μm^2/cm^3 (Table 3 and Table S3). The top rank 5 main predictor variables for LDSA were length of major roads within buffer of 125 m, length of major roads with speed limit less than 60 km/h within buffer of 100 m, length of major roads with speed limit less than 60 km/h within buffer of 125 m, length of major tertiary roads with within buffer of 125 m, and distance to parks or green spaces (Table 4, and see Table S6 for top 20 predictors).

3.2.3. Particle size model

The cross validated R^2 using the best RF model was 0.60 with an RMSE of 1.60 nm and for the BT model these were, 0.61 and 2.46 nm, respectively (Table 2). The R^2 of the final selected RF model was 0.85. The predicted annual mean APS by RF across 4.1 million grid cells was 45.9 (42.3–50.0) nm (Table 3 and Table S4). The top five main predictor variables for APS were distance to bus stops, distance to major roads with a speed limit more than 50 km/h, length of major roads within a buffer of 200 m, length of major tertiary roads within a buffer of 350 m, and distance to motorway junctions (Table 4, and see Table S7 for top 20 predictors).

3.3. Correlation of predictions

The Spearman correlation between predicted PNC and LDSA was 0.82 and like in the monitored data both of these measures were negatively correlated with the APS (−0.74 for PNC and APS, and −0.79 for LDSA and APS) (Fig. S2).

Table 2
– The model performance metrics for particle number concentration, lung deposited surface area and average particle size using random forest and bagged tree machine-learning algorithms in Copenhagen, Denmark.

UFP Metric	Machine Learning Algorithm	R^2	CV R^2[a]	RMSE[b]	MAE	Bias	Slope
PNC	Random Forest	0.93	0.65	296	740	−1593	1.29
	Bagged Tree	0.72	0.71	584	672	−1802	1.33
LDSA	Random Forest	0.88	0.67	0.48	1.19	−12.70	2.06
	Bagged Tree	0.50	0.69	0.99	1.31	−11.49	1.95
Particle size	Random Forest	0.85	0.60	1.60	3.23	−52.72	2.33
	Bagged Tree	0.65	0.61	2.46	3.41	−61.61	2.14

[a] The cross validation for random forest and bagged tree ML algorithms was 10-fold repeated 10 times.

[b] The RMSE units are pt/cm^3 for PNC, μm^2/cm^3 for LDSA, and nm for particle size.

Table 3
– Descriptive statistics for predicted annual mean particle number concentration (PNC) (pt/cm^3), lung deposited surface area (LDSA) (μm^2/cm^3), and average particle size (APS) (nm) by the random forest models in Copenhagen, Denmark.

UFP Metric	# of Predictions	Min	25% ile	50% ile	Mean	75% ile	Max
PNC	4,116,529	4272	5041	5309	5573	6153	7709
LDSA	4,116,529	11.0	11.7	12.0	12.0	12.4	13.2
Particle size	4,116,529	42.3	45.2	45.9	45.9	46.7	50.0

Table 4
– Top five important predictors for prediction of annual mean particle number concentration, lung deposited surface area, and average particle size using random forest (RF) and bagged tree (BT) machine-learning algorithms in Copenhagen, Denmark.

UFP Metric	Rank	RF Feature Importance	BT Feature Importance
PNC	1	Length of major roads within buffer of 125 m	Length of major roads within buffer of 100 m
	2	Length of major roads within buffer of 150 m	Length of major roads within buffer of 150 m
	3	Distance to bus stops	Distance to major roads with speed limit more than 40 km/h
	4	Length of major roads within buffer of 100 m	Distance to traffic signals
	5	Distance to traffic signals	Distance to parking areas
LDSA	1	Length of major roads within buffer of 125 m	Length of major roads within buffer of 150 m
	2	Length of major roads with speed limit less than 60 km/h within buffer of 100 m	Distance to industrial land use
	3	Length of major roads with speed limit less than 60 km/h within buffer of 125 m	Distance to major roads with speed limit more than 40 km/h
	4	Length of major tertiary roads with within buffer of 125 m	Area of residential land use within 100 m buffer
	5	Distance to parks or green spaces	Length of major tertiary roads with within buffer of 125 m
Particle size	1	Distance to bus stops	Distance to major roads with speed limit more than 50 km/h
	2	Distance to major roads with speed limit more than 50 km/h	Length of major tertiary roads with within buffer of 350 m
	3	Length of major roads within buffer of 200 m	Distance to supermarkets
	4	Length of major tertiary roads with within buffer of 350 m	Distance to gas stations
	5	Distance to motorway junctions	Distance to bus stops

3.4. Performance evaluation

The 1:1 plots for PNC, LDSA, and APS are shown in Figs. S3–S5. The BA plots suggested that albeit larger differences were typically seen in higher mean values, the cloud of the points were mostly within the limits of agreement (Figs. S6–S8).

3.5. Predictions

The within-city spatial predictions for PNC and LDSA showed substantial reductions away from major roads with high traffic levels and away from crosses, traffic signals, and bus stops (Fig. 2, Fig. 3, and Figs. S9–S10). The spatial patterns of PNC and LDSA seemed somewhat different with smaller predicted LDSA values around airport areas while having high PNC values in their roads, and having larger predicted LDSA values in wider buffers around the major streets. Importantly, the

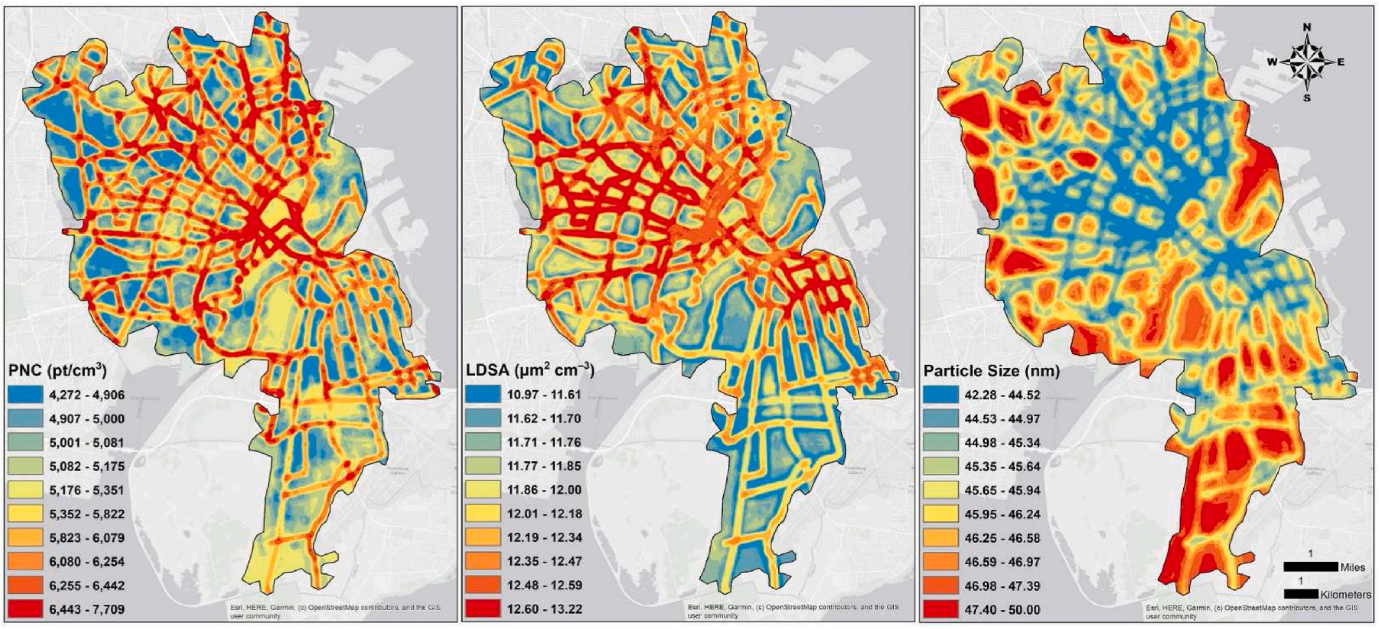


Fig. 2. Predicted annual mean concentrations of particle number concentration, lung deposited surface area and average particle size by random forest model in Copenhagen, Denmark. The spatial resolution of prediction grids is 5 m (4,116,529 predicted grid cells for each pollutant). The legend breaks are based on deciles, i.e., the 5th category includes the median values.

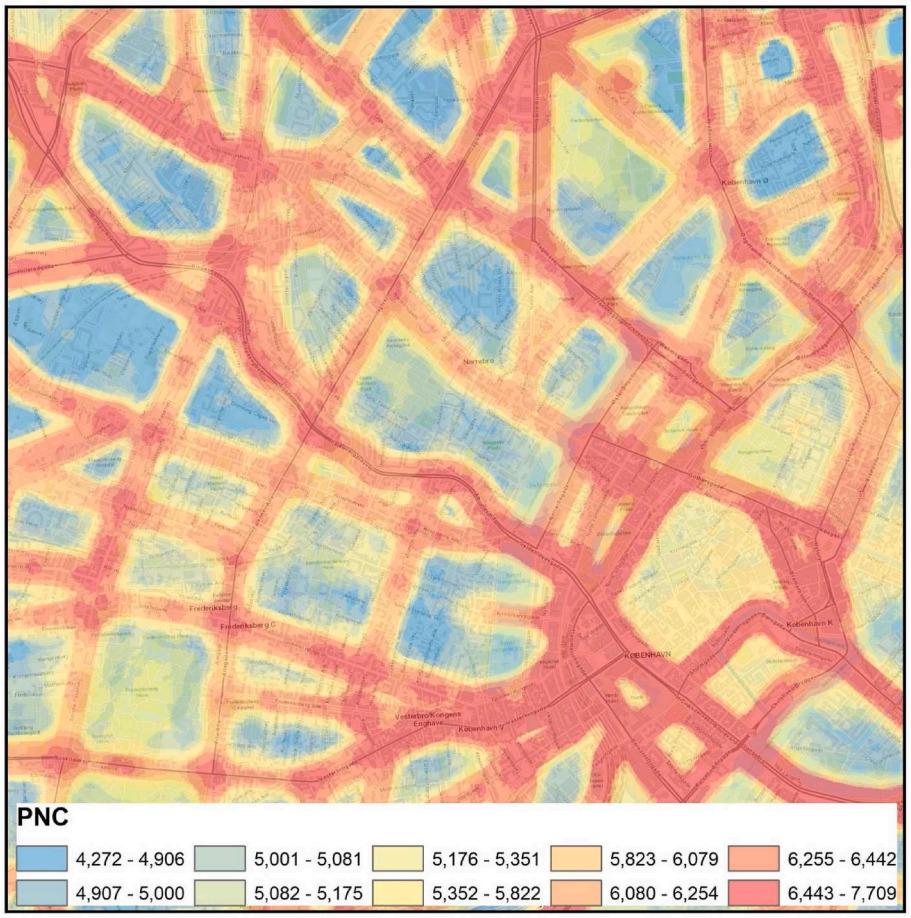


Fig. 3. – Predicted within-city annual mean particle number concentrations (PNC) by random forest model in Copenhagen, Denmark. The legend breaks are based on deciles.

selected buffers for the impact of roads on PNC were about 100–150 m while for LDSA this was until 100–350 m radii.

4. Discussion

In this study, we have developed the first ML models at a spatial resolution of 5 m for PNC, LDSA, and APS for the Danish capital based on real-world monitored data, predictor variables from a variety of classes, and state-of-the-art modeling techniques in an area with relatively low UFP concentrations. The cross-validation R^2 values (10-fold repeated 10-times) for final models were 0.65, 0.67, and 0.60 for PNC, LDSA, and APS, respectively.

UFPs have various sources for emission into the ambient air that varies over space and time (Daher et al., 2013). They are mainly emitted from transportation sector (traffic exhaust and aircrafts) and nucleation events (Brines et al., 2015), but also from ship terminals and ports, restaurants, agriculture, crop burning and wildfires, industries, natural gas combustion, and from residential areas by heating, wood burning, and cooking (Venecek et al., 2019; Yu et al., 2019).

The PNC values in Copenhagen were relatively low compared to other parts of the world; however, they do vary across space substantially, such as near or far from road traffic intersections, construction sites, bus stops and so forth (Bergmann et al., 2022). The monitored annual mean PNC across six state-owned continuous sites that were located near traffic sources was 6970 (range: 6117–8975) pt/cm^3. The estimated annual mean across the monitored 27 residences in our study was 5201 (range: 3730–6583) pt/cm^3. These values are lower than in any other city where LUR models have been developed (Patton et al., 2021). In response to concerns that our monitoring locations, which were primarily residential, might not accurately represent the land use and traffic patterns across the entire region, we conducted an analysis. We compared the distributions of the top five predictor variables at our monitoring sites with those across the entire region. The results showed that our monitoring sites are relatively closer to bus stops, traffic signals, and major roads compared to the average for the region (Table 5). This suggests that if we were to include more monitoring locations further away from these areas, we might observe even lower annual mean values. It should be considered that our values were including day and night time measurements, and in many other monitoring campaigns only daytime data have been collected. The mean PNC in Pittsburgh, Pennsylvania, United States, over 3–6 weeks of continuous measurement (day and night) has been reported to be as low as 7500 pt/cm^3 (Saha et al., 2019b). In Augsburg, Germany, over three two-week measurements PNC was 8311 pt/cm^3 (Wolf et al., 2017), and in Basel, Switzerland, over three 20 min measurements up to six sites at a time it was 10,100 pt/cm^3 (Ragettli et al., 2014). In some other cities, PNC had

Table 5
– The distributions of the top five predictor variables at our monitoring sites compared with those across the entire region. The units for all variables are meters.

UFP Metric	Predictor	Location	Min	25%ile	50%ile	Mean	75%ile	Max
PNC	Length of major roads within buffer of 125 m	33 monitoring sites	0	0	0	249.3	321.0	1214.0
		Entire region	0	0	0	132.4	236.0	2378.0
	Length of major roads within buffer of 150 m	33 monitoring sites	0	0	198.0	365.2	476.0	1577.0
		Entire region	0	0	0	190.5	331.0	2803.0
	Distance to bus stops	33 monitoring sites	9.5	152.0	252.8	258.0	349.4	585.3
		Entire region	0	222.0	487.0	1065.2	1488.7	5820.1
	Length of major roads within buffer of 100 m	33 monitoring sites	0	0	0	153.0	220.0	740.0
		Entire region	0	0	0	84.8	112.0	1888.0
	Distance to traffic signals	33 monitoring sites	48.3	152.0	248.0	261.3	402.6	539.6
		Entire region	0	236.4	571.5	1217.9	1847.4	6674.3
LDSA	Length of major roads within buffer of 125 m	27 monitoring sites	0	0	0	113.9	221.0	587.0
		Entire region	0	0	0	132.4	236.0	2378.0
	Length of major roads with speed limit less than 60 km/h within buffer of 100 m	27 monitoring sites	0	0	0	60.0	105.0	377.0
		Entire region	0	0	0	55.7	0	1773.0
	Length of major roads with speed limit less than 60 km/h within buffer of 125 m	27 monitoring sites	0	0	0	112.6	204.0	587.0
		Entire region	0	0	0	87.0	56.0	2246.0
	Length of major tertiary roads with within buffer of 125 m	27 monitoring sites	0	0	0	113.9	221.0	587.0
		Entire region	0	0	0	98.6	142.0	2246.0
	Distance to parks or green spaces	27 monitoring sites	20.1	55.5	113.0	156.0	210.3	366.0
		Entire region	0	136.2	413.0	1010.7	1420.1	7379.1
Particle size	Distance to bus stops	27 monitoring sites	58.2	183.0	280.4	291.3	372.3	585.3
		Entire region	0	222.0	487.0	1065.2	1488.7	5820.1
	Distance to major roads with speed limit more than 50 km/h	27 monitoring sites	123.7	335.5	614.2	719.1	958.0	2042.0
		Entire region	0	311.5	732.8	1287.6	1669.0	7354.0
	Length of major roads within buffer of 200 m	27 monitoring sites	0	0	381.0	462.3	550.5	2319.0
		Entire region	0	0	0	338.1	542.0	3767.0
	Length of major tertiary roads with within buffer of 350 m	27 monitoring sites	0	819.5	1394.0	1495.0	2320.0	4351.0
		Entire region	0	0	368.0	769.6	1325.0	7141.0
	Distance to motorway junctions	27 monitoring sites	350.2	1397.9	2671.5	2576.5	3662.5	4993.4
		Entire region	0	1300.0	2523	2860.0	4147.0	9102.0

higher mean values of 20,578 pt/cm^3 in Montreal, Canada, over 23 days of mobile monitoring from June to July 2012 mornings and afternoon rush hours (Zalzal et al., 2019), also near 16,000 pt/cm^3 in a 2020–2021 campaign in same city (mobile monitoring from 7 a.m. to 11 p.m. at random times for four seasons) (Lloyd et al., 2023), 44,000 pt/cm^3 again in Somerville, Massachusetts, United States, over 43 days of mobile monitoring by 3–6 h per day (Patton et al., 2014), or up to a median of 100,000 pt/cm^3 in several Spanish cities by 15 min monitoring in hundreds of rotating sites (Rivera et al., 2012). It is of important note that not all studies have used the same instruments for UFP monitoring (thus, they had different cut-off levels for measurement range), and not of the same monitoring duration and number of locations, thus making the comparisons very challenging given the fact that UFP has high spatiotemporal variation.

LDSA has been suggested to be a better marker for the health relevance of UFP in Switzerland (Aguilera et al., 2016). The measurement of LDSA needs the entire APS distribution where APS is divided into size bins, the sum of particle surface in each size bin is weighted by its lung-deposition probability (based on models), and finally the square micrometer surface per cubic centimeter of air is calculated (Geiss et al., 2016). The estimated annual mean LDSA in Copenhagen across 27 residences was 12.0 μm^2/cm^3, which is again low compared to many other world cities. In Helsinki, Finland, the average measured LDSA varied from 12 to 94 μm^2/cm^3 in a park area and at a traffic site next to a major road, respectively (Kuuluvainen et al., 2016). The mean LDSA near the LAX airport in Los Angeles (LA), California, United States, has been reported to be 47.2 μm^2/cm^3 (Habre et al., 2018), while near LA freeways it has been reported to be about 53 μm^2/cm^3 (Ntziachristos et al., 2007). In London, UK, the average LDSA is reported to be near 18 μm^2/cm^3 (Shah et al., 2023). In Switzerland, across four study areas (Basel, Geneva, Lugano, Wald), the mean LDSA was 32.1 μm^2/cm^3 (Eeftens et al., 2016).

The average APS of UFP strongly relates to its source and various conditions in the production process (Brugge and Fuller, 2021). Most of the total number of particles in the overall size distribution near highways or airports fall in the UFP or its smaller size range (Brugge and Fuller, 2021). The annual average APS in Copenhagen was 46.1 nm in our monitored data across 27 residences. This metric is lower in more polluted cities with higher traffic and other producing sources. In LA, the mean APS was about 30.1 nm near LAX airport (Habre et al., 2018). In Canada, 33.7 nm for Toronto, and 29.7 nm for Montreal has been reported during 2020–2021 mobile monitoring campaign (Lloyd et al., 2023).

Our RF PNC model in Copenhagen had a good performance with cross-validated R^2 of 0.65. PNC hybrid models in Canada (a combination of LUR and convolutional neural network models), were able to explain 60% of variation in Montreal, and 73% in Toronto (Lloyd et al., 2023). Overall, based on literature, longer-term measurements for PNC have resulted in better performance (Kerckhoffs et al., 2016). The majority of predictor variables in our PNC model were related to traffic but also included the area of parks within 500 m buffer (Table 4 and Table S5). One novelty of our work is differentiating the impacts of road types and roads with different speed limits on PNC, where distance to major roads with speed limit more than 40 km/h, or length of major roads with speed limit less than 60 km/h within buffer of 100 m were among the important predictors. Traffic intensity and other proxy variables for traffic in addition to other predictors, such as population density, number of restaurants within different buffers, airport and industry related variables have been all predictors well explaining the variation of PNC in urban areas across the world (Abernethy et al., 2013; Kerckhoffs et al., 2021; Ragettli et al., 2014; Shairsingh et al., 2019; Weichenthal et al., 2016). An important finding from our PNC predictions for Copenhagen is that while they have elevated numbers across the main roads as a significant source, they significantly decline away from major roads, which reflects the short lifetime of such particles, and agglomeration to larger particles as shown in our particle size predictions (Fig. 2). Such large spatial variation in PNC are important for epidemiological studies that evaluate the effects of UFP on population health.

Our LDSA model had cross validated R^2 of 0.67. The LDSA models have been rather rare, so far. We found only one study that spatially modeled and predicted LDSA for epidemiological studies, which was developed in the cities of Basel, Geneva, Lugano, and Wald in Switzerland for the SAPALDIA (Study on Air Pollution And Lung Disease In Adults) project (Eeftens et al., 2016). The cross-validated R^2 of the model developed by Eeftens et al. (2016) was 0.87, and LDSA was mainly explained by the total lengths of all major roads within 250 m buffer, total lengths of all roads within 100 m buffer, total traffic load of roads (sum of (traffic intensity × length of each segment)), and altitude (Eeftens et al., 2016). The subsequent epidemiological evidence from Switzerland indicated stronger effects of LDSA than PNC (Aguilera et al., 2016; Endes et al., 2017), which highlights the potential importance of this metric for future epidemiological studies. A recent study in London, UK, monitored LDSA by one year of mobile monitoring using Google Street View cars, and reported sharp spatial patterns for this metric within the city in areas with high density of restaurants. The LDSA in such areas were as high as 25 μm^2/cm^3 in London (Shah et al., 2023). In our modeling practice in Copenhagen, the traffic related variables were the most influential predictors of LDSA, but distance to restaurants, café, or pubs was also an important predictor, supporting the role of such sources in contribution to LDSA (Table S6). Further, as the distance from parks increased, the LDSA levels actually rose, indicating that LDSA values are generally lower within park and green space areas.

Our model for APS in Copenhagen had cross validated R^2 of 0.60. The model like PNC and LDSA was mostly explained by traffic-related predictors, such as distance to bus stops, major roads, and motorway junctions. As UFP ages in the atmosphere, they accumulate oxygen atoms over time and oxidize, and agglomerate with each other, meaning that they transform and become larger farther away from the source. This is evident in our ML model predictions where smaller mean particles are estimated for inner city areas and where the load of traffic as a main source is higher, and larger particles are estimated for parks, and areas farther away from the main sources (Fig. 2). We have found only one study that modeled spatially APS in the UFP range, which was in Canada for Toronto and Montreal using mobile monitored data. Lloyd et al. (2023) developed hybrid models for APS in these two cities and reported R^2 values of 0.55 in Toronto and 0.49 in Montreal in the test sets. They predicted at 100 m spatial resolution and found that busy roads with high traffic to have smaller mean APS, which is a similar finding to our study (Lloyd et al., 2023). APS could be an interesting metric together with PNC and LDSA for epidemiological analyses to see if the effect of PNC modifies based on size of the particles, or for urban planning identifying areas of interest for intervention, e.g., introduction of green buffer zones, creation of green biking routes, and so forth.

Our models have several limitations and strengths. First, the number of sites used for the modeling was rather limited, and the model predictions were generalizations based on information from 27 (for LDSA and APS) to 33 (for PNC) monitoring sites within the city. While more sites are preferable, our ML models even based on these limited number of monitoring sites explain acceptably the variations in long-term monitored data by cross-validated (10-fold, repeated 10 times) R^2 values above 0.60. Second, the true annual mean PNC was available only at six monitors, and for the other 27 locations, they were estimated based on a reference site, field campaigns, and standard methods (i.e., temporal adjustment method). These methods have been demonstrated to well approximate the true annual mean at each site across many studies (Eeftens et al., 2016; Saha et al., 2019a). Third, there is a risk of overfitting for our ML models. We followed a 10-fold and 10-times-repeated cross-validation method to ensure our models were not overfitted. The reasonably generated maps for UFP in our models (higher values near major roads as the main source, and lower values away from the location of main UFP sources) may indicate having

generalizable models, supporting that the models were not overfitted. We believe that the generalized air pollution estimations in our work are valid and agree with our expectations on spatial patterns of these markers in the area. As with any model, external validation with high-quality data would be very useful and warranted, which is recommended in future studies. Fourth, while our model provides predictions at a 5-m grid spatial resolution, we acknowledge that this is an estimate, and the actual resolution may vary. Finally, the challenge of predicting over 4 million data points from just 27–33 observations is indeed significant. However, it's important to note that each of these observations is not a single data point, but rather a rich set of data capturing various aspects of the environment. This, combined with the power of ML algorithms, allows us to extrapolate from these observations to a larger spatial scale.

5. Conclusions

In this paper, we report the first machine-learning models that provide long-term average estimates of PNC, LDSA, and APS in Copenhagen, Denmark. These modeled values have various applications, such as studying the health impacts of different UFP markers or urban planning among others. Our study adds new insights on the spatial variation of UFP exposure metrics in a European city with relatively low PNC. Our high-resolution predictions at 5 m enable investigating the health effects of PNC or LDSA and the role of APS in modifying these effects in the Danish capital where high-quality registry data is available for epidemiological studies.

Author contributions

Steffen Loft: Writing – review & editing. Roel Vermeulen: Writing – review & editing. Shali Tayebi: Writing – review & editing. Youn-Hee Lim: Writing – review & editing. Laust H. Mortensen: Writing – review & editing. Ole Hertel: Writing – review & editing. Rasmus Reeh: Writing – review & editing. Heresh Amini: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing – original draft. Zorana Andersen: Writing – review & editing. Thomas Cole-Hunter: Writing – review & editing. Joel Schwartz: Conceptualization, Methodology, Supervision, Writing – review & editing. Jules Kerckhoffs: Writing – review & editing. Jibran Khan: Writing – review & editing. Kees Meliefste: Writing – review & editing. Christian Gaarde Nielsen: Writing – review & editing. Marie Bergmann: Writing – review & editing. Seyed Mahmood Taghavi Shahri: Writing – review & editing

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by Health Effects Institute (HEI) (#4982-RFA19-2/21-5) and Novo Nordisk Foundation Challenge Programme (NNF17OC0027812). Research described in this article was conducted under contract to the HEI, an organization jointly funded by the United States Environmental Protection Agency (EPA) (Assistance Award CR 83998101) and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI, or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers. Heresh Amini was supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419, and by P30ES023515, National Institutes of Health, United States.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2024.123664.

References

Abernethy, R.C., Allen, R.W., McKendry, I.G., Brauer, M., 2013. A land use regression model for ultrafine particles in Vancouver, Canada. Environ. Sci. Technol. 47, 5217–5225.

Aguilera, I., Dratva, J., Caviezel, S., Burdet, L., de Groot, E., Ducret-Stich, R.E., Eeftens, M., Keidel, D., Meier, R., Perez, L., 2016. Particulate matter and subclinical atherosclerosis: associations between different particle sizes and sources with carotid intima-media thickness in the SAPALDIA study. Environ. Health Perspect. 124, 1700–1706.

Allaire, J., 2012. RStudio, vol. 770. integrated development environment for R, Boston, MA, pp. 165–171.

Amini, H., Hosseini, V., Schindler, C., Hassankhany, H., Yunesian, M., Henderson, S.B., Künzli, N., 2017a. Spatiotemporal description of BTEX volatile organic compounds in a Middle Eastern megacity: tehran study of exposure prediction for environmental health research (Tehran SEPEHR). Environ. Pollut. 226, 219–229.

Amini, H., Schindler, C., Hosseini, V., Yunesian, M., Künzli, N., 2017b. Land use regression models for Alkylbenzenes in a middle eastern megacity: tehran study of exposure prediction for environmental Health Research (Tehran SEPEHR). Environ. Sci. Technol. 51, 8481–8490.

Andersen, Z.J., Olsen, T.S., Andersen, K.K., Loft, S., Ketzel, M., Raaschou-Nielsen, O., 2010. Association between short-term exposure to ultrafine particles and hospital admissions for stroke in Copenhagen, Denmark. Eur. Heart J. 31, 2034–2040.

Bergmann, M., Andersen, Z., Amini, H., Khan, J., Lim, Y., Loft, S., Mehta, A., Westendorp, R., Cole-Hunter, T., 2022. Ultrafine particle exposure for bicycle commutes in rush and non-rush hour traffic: a repeated measures study in Copenhagen, Denmark. Environ. Pollut. 294, 118631.

Bergmann, M., Taghavi Shahri, S.M., Tayebi, S., Kerckhoffs, J., Khan, J., Cole-Hunter, T., Hertel, O., Hoek, G., Mortensen, L., Lim, Y.-H., Massling, A., Meliefste, K., Schwartz, J., Vermeulen, R., Loft, S., Andersen, Z., Amini, H., 2023a. Spatial and Temporal Variation of Façade-Level Particle Number Concentrations Using Portable Monitors in Copenhagen, Denmark, SSRN.

Bergmann, M.L., Andersen, Z.J., Amini, H., Ellermann, T., Hertel, O., Lim, Y.H., Loft, S., Mehta, A., Westendorp, R.G., Cole-Hunter, T., 2021. Exposure to ultrafine particles while walking or bicycling during COVID-19 closures: a repeated measures study in Copenhagen, Denmark. Sci. Total Environ. 791, 148301.

Bergmann, M.L., Andersen, Z.J., Massling, A., Kindler, P.A., Loft, S., Amini, H., Cole-Hunter, T., Guo, Y., Maric, M., Nordstrøm, C., 2023b. Short-term Exposure to Ultrafine Particles and Mortality and Hospital Admissions Due to Respiratory and Cardiovascular Diseases in Copenhagen, Denmark. Environmental Pollution, 122396.

Biau, G., Scornet, E., 2016. A random forest guided tour. Test 25, 197–227.

Bouma, F., Janssen, N.A.H., Wesseling, J., van Ratingen, S., Strak, M., Kerckhoffs, J., Gehring, U., Hendricx, W., de Hoogh, K., Vermeulen, R., Hoek, G., 2023. Long-term exposure to ultrafine particles and natural and cause-specific mortality. Environ. Int. 175, 107960.

Brines, M., Dall'Osto, M., Beddows, D.C.S., Harrison, R.M., Gomez-Moreno, F., Nunez, L., Artinano, B., Costabile, F., Gobbi, G.P., Salimi, F., Morawska, L., Sioutas, C., Querol, X., 2015. Traffic and nucleation events as main sources of ultrafine particles in high-insolation developed world cities. Atmos. Chem. Phys. 15, 5929–5945.

Brugge, D., Fuller, C.H., 2021. Ambient Combustion Ultrafine Particles and Health. Nova Science Publishers.

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope III, C.A., Apte, J.S., Brauer, M., Cohen, A., Weichenthal, S., 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proc. Natl. Acad. Sci. USA 115, 9592–9597.

Cattani, G., Gaeta, A., di Bucchianico, A.D.M., De Santis, A., Gaddi, R., Cusano, M., Cesaroni, G., Ancona, C., Forastiere, F., Gariazzo, C., 2016. Development of Land-Use Regression Models for Ultrafine Particles in Rome, Italy. ISEE Conference Abstracts.

Chang, T.-Y., Tsai, C.-C., Wu, C.-F., Chang, L.-T., Chuang, K.-J., Chuang, H.-C., Young, L.-H., 2021. Development of land-use regression models to estimate particle mass and number concentrations in Taichung, Taiwan. Atmos. Environ. 252, 118303.

Clifford, S., Mazaheri, M., Salimi, F., Ezz, W.N., Yeganeh, B., Low-Choy, S., Walker, K., Mengersen, K., Marks, G.B., Morawska, L., 2018. Effects of exposure to ambient ultrafine particles on respiratory health and systemic inflammation in children. Environ. Int. 114, 167–180.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. Lancet 389, 1907–1918.

Daher, N., Hasheminassab, S., Shafer, M.M., Schauer, J.J., Sioutas, C., 2013. Seasonal and spatial variability in chemical composition and mass closure of ambient ultrafine

particles in the megacity of Los Angeles. Environ. Sci. J. Integr. Environ. Res.: Process. Impacts 15, 283–295.
De'ath, G., 2007. Boosted trees for ecological modeling and prediction. Ecology 88, 243–251.
Eeftens, M., Meier, R., Schindler, C., Aguilera, I., Phuleria, H., Ineichen, A., Davey, M., Ducret-Stich, R., Keidel, D., Probst-Hensch, N., 2016. 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. Health 15, 1–14.
Eeftens, M., Phuleria, H.C., Meier, R., Aguilera, I., Corradi, E., Davey, M., Ducret-Stich, R., Fierz, M., Gehrig, R., Ineichen, A., 2015. Spatial and temporal variability of ultrafine particles, NO2, PM2. 5, PM2. 5 absorbance, PM10 and PMcoarse in Swiss study areas. Atmos. Environ. 111, 60–70.
Egerstrom, N., Rojas-Rueda, D., Martuzzi, M., Jalaludin, B., Nieuwenhuijsen, M., So, R., Lim, Y.H., Loft, S., Andersen, Z.J., Cole-Hunter, T., 2023. Health and economic benefits of meeting WHO air quality guidelines, Western Pacific Region. Bull. World Health Organ. 101, 130–139.
Endes, S., Schaffner, E., Caviezel, S., Dratva, J., Stolz, D., Schindler, C., Kunzli, N., Schmidt-Trucksass, A., Probst-Hensch, N., 2017. Is physical activity a modifier of the association between air pollution and arterial stiffness in older adults: the SAPALDIA cohort study. Int. J. Hyg Environ. Health 220, 1030–1038.
Foley, M., 2020. Bagged Trees.
Frohn, L.M., Ketzel, M., Christensen, J.H., Brandt, J., Im, U., Massling, A., Andersen, C., Plejdrup, M.S., Nielsen, O.-K., Gon, H.D.v.d., Manders-Groot, A., Raaschou-Nielsen, O., 2021. Modelling ultrafine particle number concentrations at address resolution in Denmark from 1979-2018 – Part 1: regional and urban scale modelling and evaluation. Atmos. Environ. 264, 118631.
Gakidou, E., Afshin, A., Abajobir, A.A., Abate, K.H., Abbafati, C., Abbas, K.M., Abd-Allah, F., Abdulle, A.M., Abera, S.F., Aboyans, V., 2017. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 390, 1345–1422.
Ge, Y., Fu, Q., Yi, M., Chao, Y., Lei, X., Xu, X., Yang, Z., Hu, J., Kan, H., Cai, J., 2022. High spatial resolution land-use regression model for urban ultrafine particle exposure assessment in Shanghai, China. Sci. Total Environ. 816, 151633.
Geiss, O., Bianchi, I., Barrero-Moreno, J., 2016. Lung-deposited surface area concentration measurements in selected occupational and non-occupational environments. J. Aerosol Sci. 96, 24–37.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth engine: planetary-scale geospatial analysis for everyone. Rem. Sens. Environ. 202, 18–27.
Goshua, A., Akdis, C.A., Nadeau, K.C., 2022. World Health Organization global air quality guideline recommendations: executive summary. Allergy 77, 1955–1960.
Habre, R., Zhou, H., Eckel, S.P., Enebish, T., Fruin, S., Bastain, T., Rappaport, E., Gilliland, F., 2018. Short-term effects of airport-associated ultrafine particle exposure on lung function and inflammation in adults with asthma. Environ. Int. 118, 48–59.
Hoek, G., Beelen, R., Kos, G., Dijkema, M., Zee, S.C.v.d., Fischer, P.H., Brunekreef, B., 2011. Land use regression model for ultrafine particles in Amsterdam. Environ. Sci. Technol. 45, 622–628.
Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. J. Comput. Graph Stat. 5, 299–314.
Karunasinghe, J., Knibbs, L.D., Clifford, S., Salimi, F., Morawska, L., 2015. Land Use Regression Model (LUR) for Ultrafine Particles in Brisbane. 9th Asian Aerosol Conference (AAC2015).
Kerckhoffs, J., Hoek, G., Gehring, U., Vermeulen, R., 2021. Modelling nationwide spatial variation of ultrafine particles based on mobile monitoring. Environ. Int. 154, 106569.
Kerckhoffs, J., Hoek, G., Messier, K.P., Brunekreef, B., Meliefste, K., Klompmaker, J.O., Vermeulen, R., 2016. Comparison of ultrafine particle and black carbon concentration predictions from a mobile and short-term stationary land-use regression model. Environ. Sci. Technol. 50, 12894–12902.
Kerckhoffs, J., Khan, J., Hoek, G., Yuan, Z., Ellermann, T., Hertel, O., Ketzel, M., Jensen, S.S., Meliefste, K., Vermeulen, R., 2022a. Mixed-effects modeling framework for Amsterdam and Copenhagen for outdoor NO2 concentrations using measurements sampled with Google street view cars. Environ. Sci. Technol. 56, 7174–7184.
Kerckhoffs, J., Khan, J., Hoek, G., Yuan, Z., Hertel, O., Ketzel, M., Jensen, S.S., Al Hasan, F., Meliefste, K., Vermeulen, R., 2022b. Hyperlocal variation of nitrogen dioxide, black carbon, and ultrafine particles measured with Google Street View cars in Amsterdam and Copenhagen. Environ. Int. 170, 107575.
Ketzel, M., Frohn, L.M., Christensen, J.H., Brandt, J., Massling, A., Andersen, C., Im, U., Jensen, S.S., Khan, J., Nielsen, O.K., Plejdrup, M.S., Manders, A., van der Gon, H.D., Kumar, P., Raaschou-Nielsen, O., 2021. Modelling Ultrafine Particle Number Concentrations at Address Resolution in Denmark from 1979 to 2018-Part 2: Local and Street Scale Modelling and Evaluation, vol. 264. Atmospheric Environment.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Software 28, 1–26.
Künzli, N., Medina, S., Kaiser, R., Quenel, P., Horak Jr., F., Studnicka, M., 2001. Assessment of deaths attributable to air pollution: should we use risk estimates based on time series or on cohort studies? Am. J. Epidemiol. 153, 1050–1055.
Kuuluvainen, H., Rönkkö, T., Järvinen, A., Saari, S., Karjalainen, P., Lähde, T., Pirjola, L., Niemi, J.V., Hillamo, R., Keskinen, J., 2016. Lung deposited surface area size distributions of particulate matter in different urban areas. Atmos. Environ. 136, 105–113.
Liu, L., Jensen, M.B., 2017. Climate resilience strategies of Beijing and Copenhagen and their links to sustainability. Water Pol. 19, 997–1013.
Lloyd, M., Ganji, A., Xu, J., Venuta, A., Simon, L., Zhang, M., Saeedi, M., Yamanouchi, S., Apte, J., Hong, K., 2023. Predicting spatial variations in annual average outdoor ultrafine particle concentrations in Montreal and Toronto, Canada: integrating land use regression and deep learning models. Environ. Int. 178, 108106.
McDuffie, E., Martin, R., Yin, H., Brauer, M., 2021. Global Burden of Disease from Major Air Pollution Sources (GBD MAPS): a Global Approach. Health Effects Institute, Research Reports, 2021.
Mitchel, A., 2005. The ESRI Guide to GIS analysis. In: Spartial Measurements and Statistics, ume 2. ESRI press.
Ntziachristos, L., Polidori, A., Phuleria, H., Geller, M.D., Sioutas, C., 2007. Application of a diffusion charger for the measurement of particle surface concentration in different environments. Aerosol. Sci. Technol. 41, 571–580.
Ohlwein, S., Kappeler, R., Kutlar Joss, M., Künzli, N., Hoffmann, B., 2019. Health effects of ultrafine particles: a systematic literature review update of epidemiological evidence. Int. J. Publ. Health 64, 547–559.
Patton, A., Robinson, A., Boogaard, H., 2021. Land use regression models of ultrafine particles for assessment of long-term exposure for health studies. In: Brugge, D., Fuller, C. (Eds.), Ambient Combustion Ultrafine Particles and Health. Nova Science Publishers, New York, pp. 57–92.
Patton, A.P., Collins, C., Naumova, E.N., Zamore, W., Brugge, D., Durant, J.L., 2014. An hourly regression model for ultrafine particles in a near-highway urban area. Environ. Sci. Technol. 48, 3272–3280.
Patton, A.P., Zamore, W., Naumova, E.N., Levy, J.I., Brugge, D., Durant, J.L., 2015. Transferability and generalizability of regression models of ultrafine particles in urban neighborhoods in the Boston area. Environ. Sci. Technol. 49, 6051–6060.
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., 2014. Spatio-temporal variation of urban ultrafine particle number concentrations. Atmos. Environ. 96, 275–283.
Rahman, M.M., Karunasinghe, J., Clifford, S., Knibbs, L.D., Morawska, L., 2020. New insights into the spatial distribution of particle number concentrations by applying non-parametric land use regression modelling. Sci. Total Environ. 702, 134708.
Rivera, M., Basagaña, X., Aguilera, I., Agis, D., Bouso, L., Foraster, M., Medina-Ramón, M., Pey, J., Künzli, N., Hoek, G., 2012. Spatial distribution of ultrafine particles in urban settings: a land use regression model. Atmos. Environ. 54, 657–666.
Saha, P.K., Li, H.Z., Apte, J.S., Robinson, A.L., Presto, A.A., 2019a. Urban ultrafine particle exposure assessment with land-use regression: influence of sampling strategy. Environ. Sci. Technol. 53, 7326–7336.
Saha, P.K., Zimmerman, N., Malings, C., Hauryliuk, A., Li, Z., Snell, L., Subramanian, R., Lipsky, E., Apte, J.S., Robinson, A.L., 2019b. Quantifying high-resolution spatial variations and local source impacts of urban ultrafine particle concentrations. Sci. Total Environ. 655, 473–481.
Saraswat, A., Apte, J.S., Kandlikar, M., Brauer, M., Henderson, S.B., Marshall, J.D., 2013. Spatiotemporal land use regression models of fine, ultrafine, and black carbon particulate matter in New Delhi, India. Environ. Sci. Technol. 47, 12903–12911.
Shah, R.U., Padilla, L.E., Peters, D.R., Dupuy-Todd, M., Fonseca, E.R., Ma, G.Q., Popoola, O.A.M., Jones, R.L., Mills, J., Martin, N.A., Alvarez, R.A., 2023. Identifying patterns and sources of fine and ultrafine particulate matter in London using mobile measurements of lung-deposited surface area. Environ. Sci. Technol. 57, 96–108.
Shairsingh, K.K., Jeong, C.-H., Wang, J.M., Brook, J.R., Evans, G.J., 2019. Urban land use regression models: can temporal deconvolution of traffic pollution measurements extend the urban LUR to suburban areas? Atmos. Environ. 196, 143–151.
Stafoggia, M., Schneider, A., Cyrys, J., Samoli, E., Andersen, Z.J., Bedada, G.B., Bellander, T., Cattani, G., Eleftheriadis, K., Faustini, A., 2017. Association between short-term exposure to ultrafine particles and mortality in eight European urban areas. Epidemiology 28, 172–180.
Statistics Denmark, 2018. Documentation of Statistics for Land Accounts 2018.
Venecek, M.A., Yu, X., Kleeman, M.J., 2019. Predicted ultrafine particulate matter source contribution across the continental United States during summertime air pollution events. Atmos. Chem. Phys. 19, 9399–9412.
Weichenthal, S., Van Ryswyk, K., Goldstein, A., Bagg, S., Shekkarizfard, M., Hatzopoulou, M., 2016. A land use regression model for ambient ultrafine particles in Montreal, Canada: a comparison of linear regression and a machine learning approach. Environ. Res. 146, 65–72.
Wolf, K., Cyrys, J., Harcinfková, T., Gu, J., Kusch, T., Hampel, R., Schneider, A., Peters, A., 2017. Land use regression modeling of ultrafine particles, ozone, nitrogen oxides and markers of particulate matter pollution in Augsburg, Germany. Sci. Total Environ. 579, 1531–1540.
World Health Organization, 2021. WHO Global Air Quality Guidelines: Particulate Matter (PM2. 5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide: Executive Summary.
Worldpop, 2020. Worldpop :: Population Counts.
Yin, H., Brauer, M., Zhang, J.J., Cai, W., Navrud, S., Burnett, R., Howard, C., Deng, Z., Kammen, D.M., Schellnhuber, H.J., 2021. Population ageing and deaths attributable to ambient PM2· 5 pollution: a global analysis of economic cost. Lancet Planet. Health 5, e356–e367.
Yu, X., Venecek, M., Kumar, A., Hu, J., Tanrikulu, S., Soon, S.T., Tran, C., Fairley, D., Kleeman, M.J., 2019. Regional sources of airborne ultrafine particle number and mass concentrations in California. Atmos. Chem. Phys. 19, 14677–14702.

Zalzal, J., Alameddine, I., El Khoury, C., Minet, L., Shekarrizfard, M., Weichenthal, S., Hatzopoulou, M., 2019. Assessing the transferability of landuse regression models for ultrafine particles across two Canadian cities. Sci. Total Environ. 662, 722–734.
Zwack, L.M., Hanna, S.R., Spengler, J.D., Levy, J.I., 2011a. Using advanced dispersion models and mobile monitoring to characterize spatial patterns of ultrafine particles in an urban area. Atmos. Environ. 45, 4822–4829.
Zwack, L.M., Paciorek, C.J., Spengler, J.D., Levy, J.I., 2011b. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. Environ. Health Perspect. 119, 852–859.