

Land use regression modelling of ultrafine particles, ozone, nitrogen oxides and markers of particulate matter pollution in Augsburg, Germany.

Kathrin Wolf^{a,*}, Josef Cyrys^{a,b}, Tatiana Harciníková^c, Jianwei Gu^{a,b}, Thomas Kusch^b, Regina Hampel^a,
Alexandra Schneider^a, Annette Peters^a

^aHelmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Institute of Epidemiology II, Neuherberg, Germany.

^bEnvironmental Science Center, University of Augsburg, Augsburg, Germany.

^cComenius University in Bratislava, Faculty of Natural Sciences, Department of Cartography, Geoinformatics and Remote Sensing, Bratislava, Slovakia.

CORRESPONDING AUTHOR:

Dr. Kathrin Wolf

Institute of Epidemiology II – Research Group “Environmental Risks”

Ingolstädter Landstr. 1

85764 Neuherberg

Phone: +49-89-3187-4563

Fax: +49-89-3187-3380

Email: kathrin.wolf@helmholtz-muenchen.de

Word count abstract: 223

Word count text: 5,634

No. of references: 41

No. of tables: 4

No. of figures: 1

Supplementary material: 1 document including 2 tables and 1 figure

KEY WORDS: Land use regression, particle number concentration, ultrafine particles, ozone, particulate matter

HIGHLIGHTS

- We constructed land use regression models for annual averages of ultrafine particles and ozone
- Models for ultrafine particles and ozone performed very well for Augsburg, Germany
- Models for PM₁₀, PM_{2.5}, soot and nitrogen oxides also performed well
- PNC was moderately correlated with PM_{2.5} and ozone, but highly correlated with NO_x

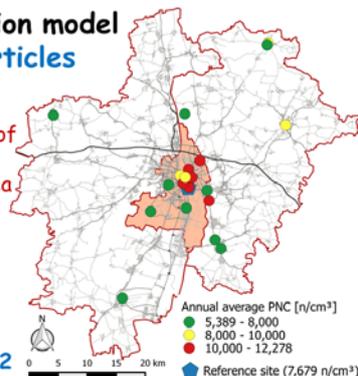
GRAPHICAL ABSTRACT

Landuse regression model for ultrafine particles

GIS predictors:

- Total traffic load of major roads
- % of industrial area
- % of forest and seminatural areas
- % of green area
- Area of buildings

Adj R² = 0.89
LOOCV Adj R² = 0.82

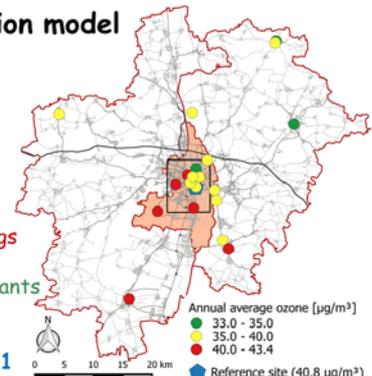


Landuse regression model for ozone

GIS predictors:

- Elevation
- Total traffic load of major roads
- X coordinate
- Number of buildings
- Area of buildings
- Number of inhabitants

Adj R² = 0.88
LOOCV Adj R² = 0.81



ABSTRACT

Important health relevance has been suggested for ultrafine particles (UFP) and ozone, but studies on long-term effects are scarce, mainly due to the lack of appropriate spatial exposure models. We designed a measurement campaign to develop land use regression (LUR) models to predict the spatial variability focusing on particle number concentration (PNC) as indicator for UFP, ozone and several other air pollutants in the Augsburg region, Southern Germany. Three bi-weekly measurements of PNC, ozone, particulate matter (PM_{10} , $PM_{2.5}$), soot ($PM_{2.5abs}$) and nitrogen oxides (NO_x , NO_2) were performed at 20 sites in 2014/15. Annual average concentrations were calculated and temporally adjusted by measurements from a continuous background station. As geographic predictors we offered several traffic and land use variables, altitude, population and building density. Models were validated using leave-one-out cross-validation. Adjusted model explained variance (R^2) was high for PNC and ozone (0.89 and 0.88). Cross-validation adjusted R^2 was slightly lower (0.82 and 0.81) but still indicated a very good fit. LUR models for other pollutants performed well with adjusted R^2 between 0.68 (PM_{coarse}) and 0.94 (NO_2). Contrary to previous studies, ozone showed a moderate correlation with NO_2 (Pearson's $r=-0.26$). PNC was moderately correlated with ozone and $PM_{2.5}$, but highly correlated with NO_x ($r=0.91$). For PNC and NO_x , LUR models comprised similar predictors and future epidemiological analyses evaluating health effects need to consider these similarities.

1. INTRODUCTION

Ultrafine particles (UFP), a subset of particulate matter (PM) with a diameter <100 nm, contribute only slightly to PM mass concentration but may cause health effects independently of larger particles as they can penetrate deeper into the lung and even translocate into the blood stream.¹ However, the literature on spatial models for particle number concentration (PNC) as indicator for UFP is limited.²⁻¹² As longer term measurement of UFP over several days or weeks are extremely cost- and labor-intensive, most of these studies used mobile or short-term monitoring campaigns. With these, a very good spatial coverage can be obtained which is essential for UFP due to their in general quite heterogeneous spatial distribution. However, the ability to predict the true long-term exposure is limited as measurements are usually performed for certain periods of the day and are often restricted to weekdays. So far, only one study from Switzerland⁴ conducted repeated longer term UFP measurements. Regarding epidemiological analyses, only one study from California assessed the health effects of long-term UFP exposure so far.¹³ The authors used a chemical transport model with a resolution of 4 km to estimate the spatial variation of UFP mass concentration and components and observed significant positive associations between ischemic heart disease mortality and UFP species.

Ozone is a secondary pollutant which can be reduced by nitric oxide in fresh motor vehicle exhaust, but which can also be regenerated during transport. Therefore, ozone concentrations are usually higher in suburban and rural areas downwind of the sources than in urban areas.¹⁴ Several modeling attempts have been made using interpolation¹⁵⁻¹⁷, Bayesian maximum entropy methods¹⁸ and chemical transport models.¹⁴ However, the resolution of these methods is generally quite coarse and models were mainly developed for large geographical regions. Thus, they are less suitable to assess small scale urban variability which is fundamental for ozone exposure.¹⁴ Only a few land use regression (LUR) models have been developed so far: one European model using a 1 km grid¹⁹ and two fine scale models from Sweden²⁰ and the Netherlands.²¹ Epidemiological studies investigating health effects of long-term ozone exposure are still limited and evidence is not equally conclusive on

detrimental effects as short-term studies though there is some suggestion of a causal relationship from animal toxicological studies.¹⁴

LUR is a useful tool to model spatial variability of long-term outdoor air pollution and has become more and more popular in the last decade.^{22, 23} Multiple regression models are built on the basis of air pollution measurements at several monitoring sites and potential predictor variables in the vicinity of these sites like traffic, land use, topography and population data. Within the multi-center ESCAPE (European Study of Cohorts for Air Pollution Effects) project, a standardized method has been applied to develop local LUR models for 36 European regions for NO_x, nitrogen dioxide (NO₂)²⁴ and for 20 regions for PM <10µm (PM₁₀), 2.5-10µm (PM_{coarse}), <2.5µm (PM_{2.5}), and the reflectance of PM_{2.5} (PM_{2.5abs})²⁵ as well as one European-wide LUR model.²⁶ These models are also available for the Augsburg-Munich region, however, they are mainly based on measurements conducted in Munich located approximately 80 km east of Augsburg. We therefore set up a new measurement campaign with a much denser network of monitoring sites in Augsburg specifically designed to model long-term exposure to ultrafine particles, ozone and other pollutants.

The specific objectives of this study were 1) to develop LUR models for PNC and ozone, 2) to refine and update the LUR models for PM₁₀, PM_{coarse}, PM_{2.5}, PM_{2.5abs}, NO₂, and NO_x for the Augsburg region based on measurements conducted in 2014/15, and 3) to explore the correlation between the pollutants. The models will later be applied to the residential addresses of the Cooperative Health Research in the Region of Augsburg (KORA) cohort²⁷ to assess the health effects of long-term exposure to these pollutants.

2. METHODS

The study region consisted of the city of Augsburg (280,000 inhabitants) and the two adjacent rather rural counties (372,000 inhabitants) covering an area of 147 km² and 1,854 km², respectively. Air pollution measurements and LUR modeling was based on the standardized ESCAPE approach^{24, 25} but

extended by further air pollutants (PNC and ozone) as well as predictor variables (building number and area). We conducted an intensive measurement campaign to determine annual average concentrations of PNC, ozone, PM₁₀, PM_{2.5}, PM_{2.5abs}, NO_x, and NO₂. Predictor variables were gathered from European-wide and local Geographic Information System (GIS) databases.

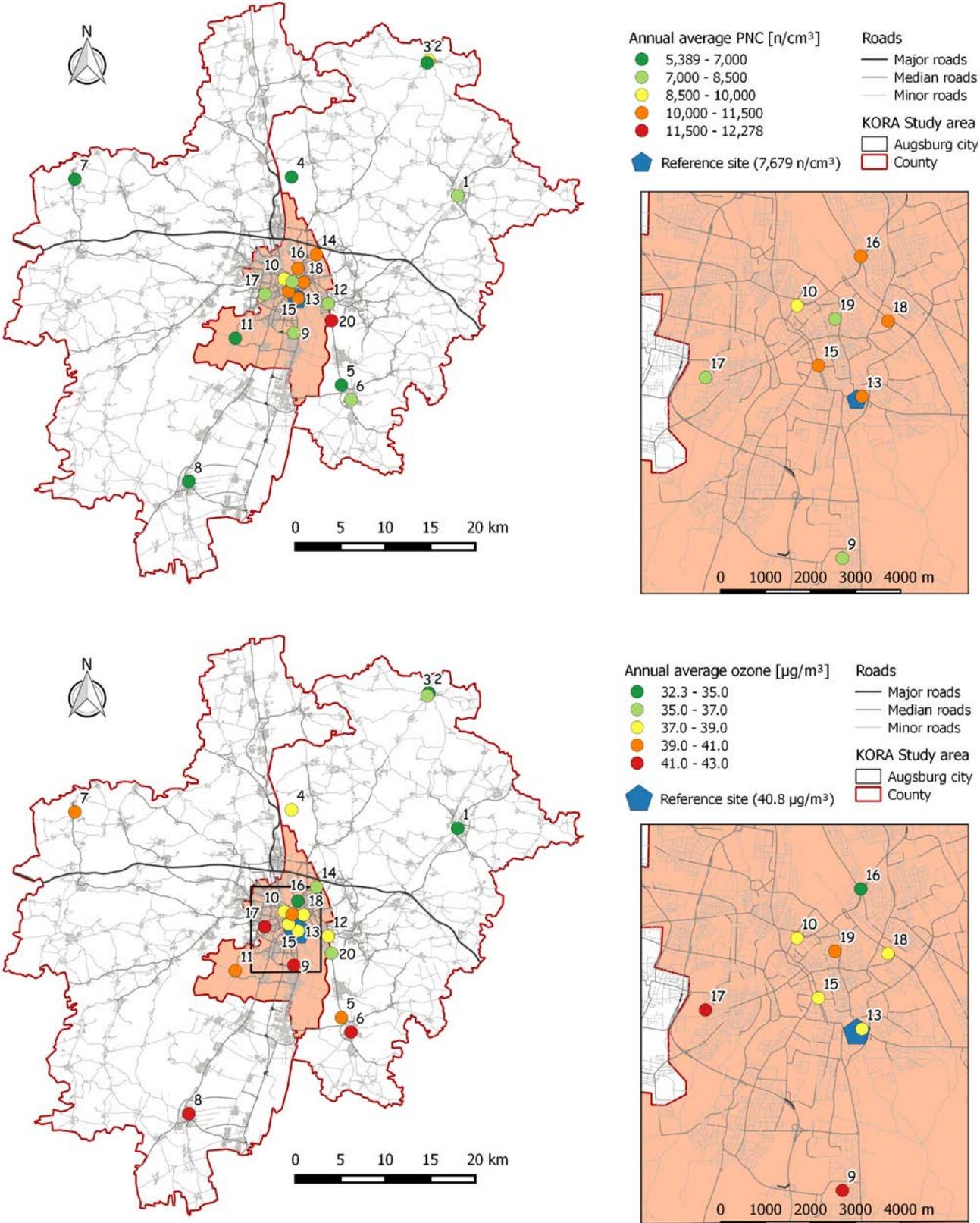
2.1. Sampling campaign

Air pollutants were measured at 20 locations within the study region (Figure 1). Of these, 12 sites were located in the city of Augsburg and 8 sites in the two adjacent counties. The site selection was based on the spatial variation of air pollution at residential addresses of the KORA participants and included a mixture of urban traffic (n=6), urban background (n=5), regional traffic (n=4) and regional background (n=4) monitoring sites and one industrial site. As several of the KORA participants were residing in predominantly rural areas of the counties, we placed one of the regional background sites in the countryside to enhance the performance of the model prediction also in the lower concentration range.

The measurements were carried out at the 20 monitoring sites between March 2014 and April 2015. Four sets of instruments were measuring at four sites for two weeks until moving to the next four sites. Thus, it took ten weeks to complete one measurement round. Overall, we had three complete measurement rounds intended to cover the warm, cold and intermediate seasons. In addition, measurements were carried out continuously at one urban background site (reference site) over the whole study period to adjust the discontinuous site measurements to the long-term average.

PNC was measured by three GRIMM ultrafine particle counters (model EDM 465 UFPC, GRIMM aerosol, Ainring, Germany) measuring total PNC with a cut-off at 7 nm and one NanoScan SMPS Nanoparticle Sizer (model 3910, TSI, Shoreview, MN, USA) measuring PNC in 13 size channels in the size range from 10 to 420 nm until July 18th, 2014. A diffusion dryer was used in the sheath air loop of the Nanoscan-SPMS in order to minimize the influence of particle growth under conditions of high

Figure 1 (2-column fitting image). Location of the 20 monitoring sites (circles) and the reference station (pentagon). The colors of the circles indicate the measured annual average concentrations of particle number concentrations (PNC, top) and ozone (bottom) at the monitoring sites.



relative humidity. From August 21st, 2014 on, a fourth GRIMM UFP counter replaced the NanoScan which broke down due to water damage. At the fixed urban background reference site PNC was measured continuously throughout the whole measurement period in the size range from 3 nm to 10 μm by use of a combination of custom-made Twin Differential Mobility Particle Spectrometry (TDMPS, size range 3 to 800 nm) based on Birmili et al²⁸ and an aerodynamic particle sizer (APS, Model 3321, TSI Inc., U.S., size range 0.8 to 10 μm). For more details please see Pitz et al.²⁹

Comparisons of all instruments conducted every two weeks showed in general a very good correlation between the collocated instruments with R^2 ranging from 0.97 to 0.98 (mean of Pearson r : 0.99). Also, PNC levels measured by all devices were very similar with differences below 5%. Only one Grimm device had to be corrected for the last seven bi-weekly measurements by factors ranging from 1.1 to 1.3. Ozone and NO_x were measured with Ogawa passive samplers (Ogawa&Co. USA Inc.) following the Ogawa protocol.³⁰ A detailed description of the ozone analysis has been published recently.²¹ The sampling rate for the Ogawa passive sampler given in the Ogawa protocol was used as a constant of 21.8 mL min⁻¹.³⁰ PM₁₀ and PM_{2.5} were sampled using Harvard Impactors and PM_{coarse} was later calculated as their difference. Reflectance was measured on both PM₁₀ and PM_{2.5} filters and transformed into absorbance. The limit of detection was determined as three times the standard deviation of the blanks (ozone: 0.85 $\mu\text{g}/\text{m}^3$; PM₁₀: 0.60 $\mu\text{g}/\text{m}^3$; PM_{2.5abs}: 0.068 10^{-5} m^{-1} ; NO_x: 6.52 $\mu\text{g}/\text{m}^3$; NO₂: 1.33 $\mu\text{g}/\text{m}^3$). Coefficients of variance based on duplicate measurements were determined following Eeftens et al³¹ and ranged from 1.9% for ozone to 8.4% for NO_x. As the correlation between PM_{10abs} and PM_{2.5abs} was extremely high with Pearson $r = 0.98$ and the concentration levels were similar, we decided to restrict the LUR model building to PM_{2.5abs}. The temporal adjustment of the discontinuous site measurements was conducted by correcting each measurement period at each site by the difference between the average concentration for each period at the reference site and the annual mean at the reference site. To account for the higher variability of PNC, we performed the correction procedure for each day of the measurements

accordingly. All data shown in this paper are temporally adjusted and are representative for the long-term average concentrations.

As Global Positioning System (GPS) coordinates showed some variations between the site visits, we manually determined the geographic coordinates of the sampling sites in GIS to ensure the accurate position.

2.2. GIS predictor data for LUR models

All potential predictor variables are summarized in Table 1, together with the a priori choices concerning buffer sizes, variable transformations and anticipated direction of effects. As secondary pollutant, ozone is involved in many chemical reactions and the expected directions of predictor effects were not as clear as for the other pollutants. Ozone concentrations are generally higher in rural and sub-urban regions as nitric oxides reduce the concentrations in urban areas and especially near major roads. Therefore, we specified the opposite direction for the land use predictors, altitude

Table 1: Description of potential predictor variables.

GIS dataset	Predictor variable	Variable name	Unit	Buffer size (radius of buffer in metre)	Direction of effect	
Background variables					All ^a	O ₃ ^b
Local land use	Residential land	hldres	% area in buffer	25, 50, 100, 300, 500, 1000, 5000	+	-
	Industrial, commercial and transport units	industry	% area in buffer	25, 50, 100, 300, 500, 1000, 5000	+	-
	Artificial surfaces (without urbgreen)	tot_built	% area in buffer	25, 50, 100, 300, 500, 1000, 5000	+	-
	Urban green	urbgreen	% area in buffer	25, 50, 100, 300, 500, 1000, 5000	-	+
	Forest and seminatural areas	seminat	% area in buffer	25, 50, 100, 300, 500, 1000, 5000	-	+
	Sum of urbgreen and seminat	green	% area in buffer	25, 50, 100, 300, 500, 1000, 5000	-	+
	Water bodies	water	% area in buffer	25, 50, 100, 300, 500, 1000, 5000	-	+
Building density	Area/number of buildings	abld, nbld	m ² /N	25, 50, 100, 300, 500, 1000, 5000	+	+/-
Population density	Number of inhabitants	pop	N	100, 300, 500, 1000, 5000	+	+/-
Household density	Number of households	hhold	N	100, 300, 500, 1000, 5000	+	+/-
Topography	Square root of altitude	elev_sqrt	m	NA	-	+
Coordinates	Coordinate variables	xcoord, ycoord	m	NA	+/-	+/-
Traffic variables						
Road network	Road length of all (major) roads in a buffer	roadl, roadlm	m	25, 50, 100, 300, 500, 1000	+	-
	Total traffic load of all (major) roads in a buffer (sum of (traffic intensity*length of all segments))	trafload, trafloadm	Veh. day ⁻¹ m	25, 50, 100, 300, 500, 1000	+	-
	Total heavy-duty traffic load of all (major) roads	heavytrafload, heavytrafloadm	Veh. day ⁻¹ m	25, 50, 100, 300, 500, 1000	+	-

Traffic intensity on nearest (major) road	intnear, intnearm	Veh. day ⁻¹	NA	+	-
Inverse distance and inverse distance squared to the nearest (major) road	distnearinv, distnearminv distnearinv2, distnearminv2	m ⁻¹ , m ⁻²	NA	+	-
Product of traffic intensity on nearest (major) road and inverse distance to the nearest (major) road and distance squared	intdistnearinv, intdistnearminv intdistnearinv2, intdistnearminv2	Veh. day ⁻¹ m ⁻¹ Veh. day ⁻¹ m ⁻²	NA	+	-
Heavy-duty traffic intensity on nearest (major) road	Heavyintnear, heavyintnearm	Veh. day ⁻¹	NA	+	-
Product of heavy-duty traffic intensity on nearest road and inverse of distance to the nearest road and distance squared	heavyintdistnearinv heavyintdistnearinv2	Veh. day ⁻¹ m ⁻¹ Veh. day ⁻¹ m ⁻²	NA	+	-

^aAll air pollutants except ozone.

^bO₃: ozone

and traffic compared to the other pollutants. However, for building, population and household density the expected direction was not as clear and we thus let it open. We used GIS to extract the predictor information at the monitoring sites out of the following GIS source data:

Digital road network

We used a local road network which is based on the basic digital landscape model for road traffic for 2009 with an accuracy <5m from the Bavarian Survey Office. We linked the roads with traffic counts routinely collected by the South Bavarian Street Directorate for regional major roads, and by the Augsburg Municipal works service and the Augsburg Environmental Agency for urban streets. Missing traffic counts were estimated based on the 5th percentile of the observed values for the corresponding street type. Major roads were defined as roads with more than 5,000 motor vehicles per 24 hours. Total traffic load in vehicles*metres was calculated as the length of a road segment*the traffic intensity on that road segment. We computed road length, total traffic load and total heavy traffic load of all roads and all major roads for buffers of 25, 50, 100, 300, 500, and 1,000 m. In addition, we calculated the (heavy-duty) traffic intensity on the nearest road and on the nearest major road and the (inverse/inverse squared) distance to these roads as well as products of the (heavy-duty) traffic intensity on the nearest (major) road and the inverse distances to that roads.

Land use data

Local Bavarian land use data with a resolution<5m were obtained from the Bavarian Survey Office for the year 2009. Land use categories were compiled in accordance to the CORINE nomenclature³² and reclassified on the basis of the APMOSPHERE³³ and ESCAPE^{24, 25} projects and adjusted to the local conditions. We computed the percentage of surface area of several land use categories in buffers of 25, 50, 100, 300, 500, 1,000, and 5,000 m including residential land, industry, built-up area, urban green, forest and seminatural areas, and water bodies.

Building density data

Footprints of buildings were available from the Bavarian Survey Office with an accuracy <5m for the year 2015. We calculated the area and the number of buildings in buffers of 25, 50, 100, 300, 500, 1,000, and 5,000 m.

Population density data

Population and household density information were gathered at a 125m*125m grid from a private company (WiGeoGIS) for the year 2008 as no official data were available in a high spatial resolution due to data protection issues. The number of inhabitants and number of households was calculated for buffers of 100, 300, 500, 1,000, and 5,000 m.

Altitude

Digital elevation data (SRTM 30 m) were obtained through the Shuttle Radar Topographic Mission, and available globally from CGIAR-CSI GeoPortal³⁴. The data have a resolution of 1 arc second (approximately 30 m at the equator and 23 m in the study area).

2.3. Pre-selection of predictor variables

Following the strategy of Eeftens and colleagues⁴ we tried to improve the stability of the LUR models by including only predictor variables where at least five sites exhibited differing values and the minimum or maximum value lay within the threefold of the 10th to 90th percentile range below or above the 10th and 90th percentile. In this way, the selection of specific predictors which included mainly zeros (e.g. major roads within 25 m) or extreme outliers was prevented aiming to avoid unstable coefficients reflected by large differences between the model R² and cross-validated R².

2.4. LUR model building and evaluation

The development of the LUR models mainly followed the standardized ESCAPE approach using a supervised stepwise selection procedure^{24,25}. For each pollutant separately, we conducted linear

regression models with the average concentration at the monitoring sites as outcome and the potential predictor variables as explaining covariates. First, we ran a univariate regression model for all potential predictors and chose the predictor with the highest R^2 . Then, we added step by step further predictor variables which maximized the adjusted R^2 if i) the increase in adjusted R^2 was more than 1%, ii) the direction of the effect was as expected, iii) there was no change in the direction of other predictor estimates and iv) the variance inflation factor was below 3 to avoid multicollinearity. In the final step, we sequentially removed variables with a p-value above 0.1 starting with the least significant one. In addition, we examined influential observations identified by Cook's D above 1. If changes in the model coefficients, p-values or model R^2 were large after rerunning the model without the corresponding site, we excluded the variable from the set of eligible predictors for the specific pollutant and repeated the whole model development. We also examined heteroscedasticity (visual inspection of residuals plotted against fitted values), normality (visual inspection of normal Q-Q plots) and spatial autocorrelation (Moran's I) of residuals to assess the independence assumption. We evaluated the model performance by leave-one-out cross-validation (LOOCV). Sequentially, each site was left out and the model was refitted with the same predictors potentially leading to changes in the parameter estimates. The model was then used to predict the concentrations at the left-out sites and the R^2 between these predicted concentrations and the actually measured ones was calculated.

GIS analyses were conducted with QGIS version 2.6.1. Statistical analyses were conducted using R version 3.2.2.

3. RESULTS

Descriptive statistics of the temporally adjusted annual average air pollution concentrations at the 20 monitoring sites are shown in Table 2. Spatial contrasts were quite substantial for our study region for PNC, NO_x and NO_2 whereas $\text{PM}_{2.5}$ showed only little variation.

Table 2: Distribution of PNC, ozone, PM_{2.5}, PM_{2.5}absorbance, PM₁₀, PM_{coarse}, NO₂, and NO_x at 20 measurement sites.

Pollutant	Mean	SD	Min	10%	25%	Median	75%	90%	Max
PNC (particles/cm ³)	8,311	2,026	5,489	6,382	6,791	7,818	10,051	10,647	13,232
Ozone (µg/m ³)	38.2	3.1	32.3	33.7	36.5	38.4	40.39	42.4	43.0
PM ₁₀ (µg/m ³)	17.4	1.7	14.0	15.3	16.3	17.8	18.5	19.1	21.6
PM _{coarse} (µg/m ³)	5.7	1.2	2.7	4.6	5.3	5.73	6.1	6.9	8.4
PM _{2.5} (µg/m ³)	11.8	0.9	10.1	10.3	11.2	11.9	12.6	12.6	13.2
PM _{2.5} abs (10 ⁻⁵ m ⁻¹)	1.2	0.2	0.8	1.0	1.1	1.2	1.38	1.4	1.7
NO ₂ (µg/m ³)	16.9	5.5	9.1	11.1	12.8	15.9	22.28	24.6	25.9
NO _x (µg/m ³)	27.7	9.5	15.5	17.4	19.6	26.1	35.91	37.7	47.8

The pre-selection of predictor variables led to the exclusion of eleven predictors: eight variables had less than five observations with differing values (industry_25, urbgreen_25, forest_25, water_25, industry_50, forest_50, water_50, industry_100), two displayed a maximum above the threefold of the 10th to 90th percentile range (heavytrafload_100, heavytrafloadm_100) and one had a minimum below this criterion (water_300). The distribution of the finally selected predictor variables can be found in Supplementary material Table 1.

The final LUR models and corresponding performance measures are summarized in Table 3. Model adjusted R² indicated a good fit for all pollutants ranging from 68% for PM_{coarse} to 94% for NO₂. Especially the traffic-related air pollutants (PNC, PM_{2.5abs}, NO₂, and NO_x) and ozone performed well with an adjusted LOOCV R² being less than 10% lower than the adjusted R². For the mass fractions, the models were not as robust with the adjusted LOOCV R² being 12% (PM₁₀, PM_{2.5}) to 19% (PM_{coarse}) lower than the adjusted R².

The LUR models consisted of four (PM_{coarse}, NO₂, and NO_x) to seven predictors (PM₁₀) and contained at least one traffic predictor in a rather small buffer (25m-100m), industry in the medium to distant vicinity (300m to 5km) and one predictor for green areas. The model for PNC included building footprints and traffic in the close vicinity (25m and 50m), seminatural and industrial areas in a 100m and 300m buffer, respectively, and green area within 500m. Models for the other pollutants tended to select rather larger buffers. The ozone model included altitude, traffic load of major roads within 100m, the X coordinate, number and area of buildings within 500m and 50m, respectively, and the population density within 300m. Except for altitude and the building number, the effect direction was negative. Altitude was also predictive for PM_{2.5}, however in the negative direction. The X coordinate, too, showed a negative association with PM_{coarse} but a positive association with PM_{2.5abs}.

The site in the countryside (Figure 1, site 4) was identified as an outlier (Cook's D>1) for the PM_{coarse} model and led to changes in the model coefficient of the traffic predictor. Thus, we excluded this site from the model selection for PM_{coarse}. Similarly, we had to exclude the traffic site 20 east of the city

Table 3. Final land use regression models for PNC, ozone, PM₁₀, PM_{2.5}, PM_{2.5}absorbance, NO₂, and NO_x.

Pollutant	N	LUR model	R ²	Adj R ²	LOOCV R ²	LOOCV Adj R ²	Moran's I (p-value)
PNC	20	6845 + 0.0023 * trafloadm_50 + 75.88 * industry_300 + 52.99 * seminat_100_neg + 44.86 * green_500_neg + 2.49 * abld_25	0.92	0.89	0.83	0.82	-0.05 (0.99)
Ozone	20	637.4 + 1.30 * elev_sqrt + 0.00000036 * trafloadm_100_neg + 0.00014 * xcoord_neg + 0.0043 * nbld_500 + 0.0013 * abld_50_neg + 0.00095 * pop_300_neg	0.92	0.88	0.82	0.81	-0.10 (0.61)
PM ₁₀	20	13.67 + 0.0078 * roadlm_100 + 0.000079 * roadl_1000 + 0.044 * industry_300 + 0.098 * urbgreen_500_neg + 0.0022 * abld_25 + 0.0000016 * trafloadm_25	0.91	0.87	0.78	0.76	-0.18 (0.15)
PM _{coarse} ^a	19	168.7 + 0.031 * tot_build_1000 + 0.030 * intdistnearminv2 + 0.025 * industry_300 + 0.000037 * xcoord_neg	0.75	0.68	0.57	0.55	-0.11 (0.60)
PM _{2.5}	20	19.47 + 0.0099 * roadl_50 + 0.041 * seminat_1000_neg + 0.41 * elev_sqrt_neg + 0.0094 * tot_build_25 + 0.012 * industry_300	0.84	0.79	0.70	0.69	-0.12 (0.49)
PM _{2.5} abs ^b	19	-14.07 + 0.000000037 * trafload_100 + 0.00047 * abld_25 + 0.0087 * industry_1000 + 0.000058 * nbld_1000 + 0.040 * water_5000_neg + 0.0000034 * xcoord	0.93	0.89	0.84	0.83	-0.14 (0.37)
NO ₂ ^b	19	12.57 + 0.22 * industry_5000 + 0.015 * roadlm_100 + 0.10 * seminat_1000_neg + 0.15 * industry_1000	0.95	0.94	0.90	0.89	-0.10 (0.64)
NO _x	20	28.06 + 0.0000084 * trafload_50 + 0.27 * green_1000_neg + 0.20 * industry_300 + 0.25 * seminat_100_neg	0.91	0.89	0.83	0.82	-0.05 (0.94)

^awithout rural site 4 (see Figure 1).

^bwithout traffic site 20 (see Figure 1).

See Table 1 for a detailed explanation of the variable names. Variables with _X (e.g. trafload_50) are buffers with _X indicating the radius of the buffer in meters. The suffix "_neg" indicates predictors with a negative sign

(where maximum PM_{2.5abs} concentration was measured) for the PM_{2.5abs} and the NO₂ model development as it led to large changes in R² and the p-values, respectively. Moran's I value indicated spatial autocorrelation in the residuals for PM₁₀. As the number of predictors was quite large we decided to leave out the last included predictor (traffic intensity on nearest road) which reduced the spatial autocorrelation but obviously also the model performance with an adjusted LOOCV R² of 76% compared to 83%. The scatter plots of measured and predicted values together with linear regression lines generally showed a good agreement (Supplementary material Figure 1).

A comparison of the updated LUR models for PM₁₀, PM_{coarse}, PM_{2.5}, PM_{2.5abs}, NO_x and NO₂ based on measurements from 2014/15 with the ESCAPE LUR models from 2008/09 can be found in Supplementary material Table 2. Except for PM_{coarse}, the adjusted R²s were similar or higher for the new models compared to the ESCAPE models with increases between 0% for PM_{2.5abs} and 11% for NO₂.

Pearson correlation coefficients are presented in Table 4 for the annual averages and the LUR predicted concentrations. In general, the discrepancy between the measured and predicted correlation coefficients was comparably low (difference in $r < 28\%$). PNC and NO_x were extremely highly correlated ($r > 0.9$). While measured PNC concentration was highly correlated ($r > 0.8$) with PM_{2.5abs}, NO₂ and NO_x, the predicted PNC concentration was highly correlated only with NO_x. Measured and predicted ozone showed a low to moderate negative correlation with the other pollutants except for PM_{coarse} which was not correlated at all.

Table 4: Pearson correlations between annual air pollution concentrations using measured (lower half) and predicted (upper half) concentrations at 20 measurement sites. Coefficients above 0.80 are marked in italic, coefficients above 0.90 are marked in bold.

Predicted \ Measured	PNC	Ozone	PM ₁₀	PM _{coarse}	PM _{2.5}	PM _{2.5abs}	NO ₂	NO _x
PNC _{mean} (particles/cm ³)		-0.47	0.75	0.71	0.58	0.61	0.71	0.91
Ozone (µg/m ³)	-0.46		-0.18	0.08	-0.51	-0.42	-0.26	-0.34
PM ₁₀	0.71	-0.17		<i>0.84</i>	0.69	0.76	0.68	0.74
PM _{coarse}	0.52	0.10	<i>0.86</i>		0.33	0.52	0.75	0.77
PM _{2.5} (µg/m ³)	0.64	-0.45	0.74	0.30		0.79	0.49	0.55
PM _{2.5abs} (10 ⁻⁵ m ⁻¹)	<i>0.81</i>	-0.39	0.75	0.49	<i>0.77</i>		0.57	0.55
NO ₂ (µg/m ³)	<i>0.88</i>	-0.32	0.76	0.64	0.58	0.65		<i>0.88</i>
NO _x (µg/m ³)	0.91	-0.28	0.78	0.65	0.63	0.76	0.95	

4. DISCUSSION

We developed LUR models for PNC, ozone, PM₁₀, PM_{coarse}, PM_{2.5}, PM_{2.5abs}, NO_x, and NO₂ based on measurements conducted in the Augsburg region in 2014/15. In general, all LUR models performed well with the adjusted R² ranging from 68% (PM_{coarse}) to 94% (NO₂). However, models for traffic-related air pollutants (PNC, PM_{2.5abs}, NO₂, and NO_x) and ozone performed better with an adjusted LOOCV R² between 82% and 89% compared to models for the different size fractions of PM mass concentration (PM_{2.5}, PM₁₀, PM_{coarse}) with an adjusted LOOCV R² between 55% and 76%.

4.1. Comparison with previous LUR models for PNC

Only one study from Switzerland⁴ measured PNC repeatedly over a longer period comparable to our campaign and used a similar modelling approach. Their PNC model is based on measurements from 67 sites located in four different regions and performed similarly well with an adjusted R² of 85% and a LOOCV R² of 82% compared to 89% and 83% for PNC in our study. A study from Amsterdam, the Netherlands measuring seven days at 50 sites yielded adjusted R² of 65% and LOOCV R² of 57%.³ In our study, measured annual average PNC ranged from 5,489 to 13,232 particles/cm³ similar to the spatial contrasts observed by the German Ultrafine Aerosol Network which involves rural, urban background and traffic sites across Germany and therefore should reflect the typical range between less polluted rural sites and heavy polluted traffic sites in Germany.³⁵ The multi-annual mean of hourly PNC (20-800 nm) measured at 17 monitoring sites from 2009 to 2014 (data coverage two to six years) ranged between 3,000 to 5,500 particles/cm³ at rural sites and 9,000 to 10,500 particles/cm³ at urban traffic sites. Our concentrations lay also within the range of the Swiss study (3,361 to 22,896 particles/cm³) which covered with four areas a much larger and very heterogeneous study region. The concentrations of the Dutch study were substantially higher ranging from 12,248 to 86,902 particles/cm³ with an annual mean concentration of 19,272 particles/cm³ at the urban background reference site. However, the study region was much smaller covering only an urban area (Amsterdam) and the single weekly measurements, although temporally adjusted, might have inflated the annual average concentrations.

Main sources for UFP are in general traffic-related combustion processes, but also industrial sources, home heating and biomass burning play a role.¹ Of these, UFP are either emitted directly or formed secondarily through chemical processes in the atmosphere. For the city of Augsburg, we identified seven sources of particles by positive matrix factorization based on particle size distribution in a previous study.³⁶ However, only three sources contributed significantly to PNC: fresh traffic emissions (24.9%), aged traffic emissions (40.3%), and stationary combustion (26.1%). As shown in a further study from the area,³⁷ traffic-related particles were rather heterogeneously distributed over the city area compared to combustion-related particles (mean coefficient of divergence for the traffic source was 0.62 compared to 0.31 for the combustion source). In fact, traffic load of major roads within 50m was selected as the first predictor to enter our LUR model for PNC reflecting the major impact of traffic in the close vicinity. Likewise, the two previous models comprised nearby traffic predictors (Switzerland: traffic load within 250m, road length within 100m and major road length within 50m; Amsterdam: traffic intensity*inverse distance to the nearest road squared). Neither ours nor the previous LUR models included heavy duty traffic predictors which might have been expected to contribute through diesel UFP emissions. But also industrial processes (e.g. from stationary combustion sources) within the medium surrounding (300m) seemed to play an important role for increased PNC in our study region entering as the second predictor which however was not selected in the previous models. Forest, seminatural and green areas in the medium surrounding (100-500m) decreased the concentrations in Augsburg as was the case for Switzerland (urban green within 1km) but not for Amsterdam. Our model additionally included building density in the close vicinity (25m), whereas the Amsterdam model comprised address density within 300m as well as ports within 3 km. The Swiss multi-site model was mainly dominated by area indicators.

Other previous pure spatial LUR models for PNC were all based on mobile or short-term measurements and the performance was comparably lower with R²s between 22% and 53%.^{7, 9, 11, 12} Several studies also using mobile or short-term measurements constructed spatiotemporal LUR models by incorporating meteorological predictors, and thus cannot be compared with our models directly.^{2, 5-8, 10, 12} The explained variance of these studies ranged from 14% to 72%.

4.2. Comparison with previous LUR models for ozone

For ozone, two studies with comparable predictor variables and development procedures have been published so far.^{20,21} Kerckhoffs et al²¹ developed a national model for the Netherlands and reported similar annual average concentrations of 25.0 to 47.8 $\mu\text{g}/\text{m}^3$ compared to our study region with 32.3 to 43.0 $\mu\text{g}/\text{m}^3$. The range was likely higher due to the larger and more heterogeneous area covering coastal and very rural areas but also the capital Amsterdam with a population size three times as big as Augsburg. A study from Sweden²⁰ observed substantially higher concentrations with 60 to 83 $\mu\text{g}/\text{m}^3$ in Malmö and 36 to 63 $\mu\text{g}/\text{m}^3$ in Umeå. However, the levels were calculated as mean concentrations of three weekly measurements conducted in April, May/June and August, thus not covering the months with minimum concentrations.

The LUR model of the Dutch study explained 77% of the variability of the annual average concentration though 29% was explained by an indicator variable for North. Other included predictors were traffic intensity and major road length within 50m, urban green space within 500m and low density residential land within 5 km. The authors also predefined the directions of predictor coefficients being negative for traffic, industry and residential land and positive for green areas. In addition, they specified a positive direction for population density which was, however, not selected. Contrary to the Dutch study and ours, the direction of predictor effects was allowed to vary freely for separate models for the two Swedish cities resulting in an adjusted R^2 of 40% (LOOCV $R^2=17\%$) for Malmö and of 67% (48%) for Umeå. The latter included similar predictors as the Dutch model except for natural areas within 5km with a negative sign.

In our study, altitude being generally higher in the rural surrounding of Augsburg and a western location (X coordinate with a negative sign) were selected as predictors increasing ozone concentrations. We speculate that NOx concentrations are higher in the wind vane of the city leading to reduced ozone concentrations in the east. Similarly to the Dutch study, our model contained traffic load of major roads within 100m with a negative sign reflecting lower ozone concentrations close to traffic sources. In addition, area of buildings within 50m and the population density within 300m were selected decreasing ozone concentrations whereas number of buildings within 500m increased the levels. The differing directions are actually not

implausible for our study region as population density and area of buildings in the close vicinity are generally higher in the urban area with city blocks and apartment houses while the rural area is characterized by small detached houses reflected in a higher building density.

4.3. Comparison with ESCAPE LUR models

Our study region was also part of the ESCAPE project and LUR models were already developed for PM₁₀, PM_{coarse}, PM_{2.5}, PM_{2.5abs}, NO_x and NO₂ based on measurements from 2008/09.^{24,25} However, the Augsburg area was combined with the Munich region covering a study area of about 27,000 km². Compared to the Augsburg region, the Munich area exhibited generally higher air pollutant levels mainly due to a higher population and traffic density. Thus with the new measurement campaign, we aimed to refine and update the existing LUR models but also to compare the performances of the rather spread and heterogeneous ESCAPE models with the denser and homogeneous new models. Except for PM_{coarse}, the explained variance was generally similar or higher for the new models compared to the ESCAPE models. Especially for NO₂, the adjusted LOOCV R² increased from 66% to 89%. Whereas mainly traffic predictors were included in the ESCAPE model (traffic load and road length of all and major roads within 50m and product of traffic intensity on nearest major road and inverse distance to the nearest major road), only road length of major roads within 100m was selected for the Augsburg model (Supplementary material Table 2). The latter was dominated by industry and seminatural areas (negative sign) in the rather distant neighborhood (1-5km) which were completely missing in the ESCAPE model. As 75% of the ESCAPE measurement sites were located in Munich or its surroundings and only 25% in the Augsburg area, the ESCAPE model was mainly dominated by Munich specific predictors. We assume that important predictors for Augsburg were not selected in the joint modeling due to the heterogeneous nature of the Munich-Augsburg area with a much higher building, population and traffic density in the Munich region leading to this difference in model performance. As measurement devices and predictor variables were similar for PM_{coarse} in the current and the ESCAPE project, we suppose that the lower spatial variation ranging from 2.7 to 8.4 µg/m³ in Augsburg

compared to 4.9 to 15.9 $\mu\text{g}/\text{m}^3$ in the Munich/Augsburg area might be responsible for the poorer fit of the LUR model. In general, ESCAPE models included more traffic predictors. For ESCAPE $\text{PM}_{\text{coarse}}$ and NO_2 , population density in a 5 km buffer was selected which was, however, not included in the new Augsburg models. Building density was not yet available in ESCAPE but was quite prominent in the new models being selected for four pollutants (PNC, ozone, PM_{10} , and $\text{PM}_{2.5\text{abs}}$).

4.4. Correlations between the pollutants

As a general note we would like to highlight that correlations among annual averages of air pollutants are expected to be higher than among daily or hourly concentrations since any differing temporal patterns are completely averaged. The measured and predicted PNC concentrations were highly correlated with NO_x both with $r = 0.91$ indicating common local sources for our study area. Consistently, the predictor variables were similar with traffic and industry being the major factors. A recent European multi-center study reported that the temporal correlation between PNC and NO/NO_2 was high in Augsburg and in general higher than in the other cities Dresden, Germany, Prague, Czech Republic, and Ljubljana, Slovenia.³⁸

For ozone, a highly negative correlation with NO_2 with $r = -0.87$ was reported for measured yearly average concentrations in the Netherlands²¹ whereas for our study region, the correlation was rather low with $r = -0.32$. The correlation between measured ozone and $\text{PM}_{2.5}$ was moderately negative with $r = -0.45$ but slightly smaller than within the Dutch study ($r = -0.67$). A study from California estimated ozone via inverse distance weighting interpolation and observed no correlation with NO_2 ($r = -0.0071$) but a moderately positive correlation with $\text{PM}_{2.5}$ ($r = 0.56$) both derived from LUR models.¹⁵ However, these results are not directly comparable to our study mainly due to the intense photochemical effects on the ozone - NO_x associations and common processes affecting the formation of background ozone and secondary fine particles in California. Moreover, PM_{10} with $\text{PM}_{\text{coarse}}$ and NO_2 with NO_x were highly spatially correlated in our study region. Eeftens and colleagues constructed LUR models for PM_{10} , $\text{PM}_{\text{coarse}}$, $\text{PM}_{2.5}$, $\text{PM}_{2.5\text{abs}}$, NO_2 , PNC and lung deposited surface area for Swiss regions and reported generally higher correlations.⁴ For the predicted concentrations, Pearson's r for PNC ranged between 0.82 for $\text{PM}_{2.5}$ and 0.94 for both $\text{PM}_{2.5\text{abs}}$

and PM_{coarse} (0.58, 0.61 and 0.71 in our study, respectively). Even the correlation between PM_{coarse} and $PM_{2.5}$ was comparably high with $r = 0.74$ compared to 0.33 in our study. The main reason for the higher correlations might be the prominent role of traffic predictors in the Swiss models, but measured concentrations were also slightly higher correlated.

For the subsequent epidemiological analyses it is important to be aware of the high correlation between PNC and NO_x and similar predictors for the LUR models. It might be necessary to use rather sophisticated models to be able to disentangle the health effects of PNC and NO_x if possible at all.

4.5. Limitations

A major limitation is the comparatively few number of 20 monitoring sites which might have inflated the predictive ability of our models.^{39, 40} However, we tried to minimize the risk of over-fitting by strict a priori defined inclusion criteria for the predictor variables. We strengthened the ESCAPE approach by additional requirements regarding the distribution of the predictor variables to increase the stability of the coefficients. In addition, the restriction to the Augsburg region resulted in a smaller and rather homogenous study region and a much denser monitoring network compared to ESCAPE. Moreover, the performance of our LUR models might have been overestimated by the use of LOOCV as recently demonstrated.⁴¹ But especially the devices for the PNC measurements were extremely costly which limited the number of feasible sites and thus the possibility to conduct the more reliable hold-out or v-fold cross validation. We tried to reduce measurement error by conducting parallel measurements before and after each bi-weekly measurement round and applied correction factors where necessary. Further limitations were the discontinuous and non-concurrent measurements in different periods and the temporal adjustment process which usually apply to LUR models in general due to lack of alternatives.

We could not include wind data as appropriate data were either not available (wind direction) or the spatial variation was too small in the available resolution of 1 km (wind speed). However, the thorough characterization and inclusion of local and regional wind patterns did not lead to any improvements in LUR models for PNC in Vancouver, Canada.¹¹

GIS data sources for traffic and land use were based on the year 2009, population density on the year 2008 and thus somewhat older than our measurement campaign. However, significant spatial transformations in the road network, land cover and the population spread were not the case for our study region and usually take several decades.

5. CONCLUSIONS

There is an increasing societal interest concerning health effects of UFP and ozone, especially in urban environments. However, reliable data on long-term effects of ambient air UFP and ozone are limited mainly due to the lack of exposure models assessing the necessary small scale spatial variability. We were able to build LUR models for PNC and ozone which performed very well for our study region. These models will be applied to the residential addresses of the KORA cohort to deliver epidemiological evidence of the role of long-term exposure to these pollutants on the incidence of chronic diseases and thus, to impact on future air quality standards by supporting policy makers to consider UFP for regulation. The refined models for PM_{10} , PM_{coarse} , $PM_{2.5}$, $PM_{2.5abs}$, NO_2 , and NO_x performed generally better than the ESCAPE models except for PM_{coarse} . Ozone was only moderately correlated with all pollutants. PNC was moderately correlated with $PM_{2.5}$ and ozone, but highly correlated with NO_x ($r=0.91$). For PNC and NO_x , LUR models comprised similar predictors and future epidemiological analyses evaluating health effects need to consider these similarities and specifically the role of traffic load and road length at the residence of study participants.

ACKNOWLEDGMENTS

We would like to thank Uwe Hartz for collecting the exposure data and Kees Melifste from the Institute for Risk Assessment Sciences (IRAS), Utrecht, for the help in preparation, transportation, and analysis of the passive samplers and PM filters. We also thank the Bavarian Environment Agency for air pollutant data used for validation purposes. We gratefully acknowledge the Bavarian Environment Agency, several

institutions, companies, kindergarden, schools, and private persons for the possibility to conduct the measurements at their premises.

FUNDING SOURCES

This work was supported by intramural funding for Environmental Health projects of Helmholtz Zentrum München - German research Center for Environmental Health.

COMPETING INTERESTS

The authors declare no competing financial interest.

LIST OF REFERENCES

1. HEI *Understanding the Health Effects of Ambient Ultrafine Particles* Health Effects Institute: Boston, MA, 2013.
2. 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. *Atmospheric Environment* **2015**, *110*, 84-92.
3. 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), 622-8.
4. Eeftens, M.; Meier, R.; Schindler, C.; Aguilera, I.; Phuleria, H.; Ineichen, A.; Davey, M.; Ducret-Stich, R.; Keidel, D.; Probst-Hensch, N.; Kunzli, N.; Tsai, M. Y. 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* **2016**, *15*, 53.
5. Weichenthal, S.; Ryswyk, K. V.; Goldstein, A.; Bagg, S.; Shekarrizfard, 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.
6. 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. *Environmental pollution (Barking, Essex : 1987)* **2016**, *208*, (Pt A), 241-8.
7. Montagne, D. R.; Hoek, G.; Klompaker, 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), 8712-20.
8. 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), 12903-11.
9. Hankey, S.; Marshall, J. D. Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, PM2.5, Particle Size) Using Mobile Monitoring. *Environ Sci Technol* **2015**, *49*, (15), 9194-202.
10. Patton, A. P.; Zamore, W.; Naumova, E. N.; Levy, J. I.; Brugge, D.; Durant, J. L. Transferability and generalizability of regression models of ultrafine particles in urban neighborhoods in the Boston area. *Environ Sci Technol* **2015**, *49*, (10), 6051-60.
11. 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), 5217-25.
12. 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. *Atmospheric Environment* **2012**, *54*, 657-666.
13. 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), 549-56.
14. EPA *Final Report: Integrated Science Assessment of Ozone and Related Photochemical Oxidants*; U.S. Environmental Protection Agency: 2013.
15. Jerrett, M.; Burnett, R. T.; Beckerman, B. S.; Turner, M. C.; Krewski, D.; Thurston, G.; Martin, R. V.; van Donkelaar, A.; Hughes, E.; Shi, Y.; Gapstur, S. M.; Thun, M. J.; Pope, C. A., 3rd. Spatial analysis of air pollution and mortality in California. *Am J Respir Crit Care Med* **2013**, *188*, (5), 593-9.
16. Liu, L. J. S.; Rossini, A. J. Use of kriging models to predict 12-hour mean ozone concentrations in Metropolitan Toronto—A pilot study. *Environment international* **1996**, *22*, (6), 677-692.
17. Jerrett, M.; Finkelstein, M. M.; Brook, J. R.; Arain, M. A.; Kanaroglou, P.; Stieb, D. M.; Gilbert, N. L.; Verma, D.; Finkelstein, N.; Chapman, K. R.; Sears, M. R. A cohort study of traffic-related air pollution and mortality in Toronto, Ontario, Canada. *Environ Health Perspect* **2009**, *117*, (5), 772-7.
18. Nazelle, A. d.; Arunachalam, S.; Serre, M. L. Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: An Application for Attainment Demonstration in North Carolina. *Environ Sci Technol* **2010**, *44*, (15), 5707-5713.
19. 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), 1852-67.

20. Malmqvist, E.; Olsson, D.; Hagenbjörk-Gustafsson, A.; Forsberg, B.; Mattisson, K.; Stroh, E.; Strömgren, M.; Swietlicki, E.; Rylander, L.; Hoek, G.; Tinnerberg, H.; Modig, L. Assessing ozone exposure for epidemiological studies in Malmö and Umeå, Sweden. *Atmospheric Environment* **2014**, *48*, 241-248.
21. Kerckhoffs, J.; Wang, M.; Meliefste, K.; Malmqvist, E.; Fischer, P.; Janssen, N. A.; Beelen, R.; Hoek, G. A national fine spatial scale land-use regression model for ozone. *Environ Res* **2015**, *140*, 440-8.
22. Ryan, P. H.; LeMasters, G. K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal Toxicol* **2007**, *19 Suppl 1*, 127-33.
23. 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* **2008**, *42*, (33), 7561-7578.
24. Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.-Y.; Künzli, N.; Schikowski, T.; Marcon, A.; Eriksen, K. T.; Raaschou-Nielsen, O.; Stephanou, E.; Patelarou, E.; Lanki, T.; Yli-Tuomi, T.; Declercq, C.; Falq, G.; Stempfelet, M.; Birk, M.; Cyrus, J.; von Klot, S.; Nádor, G.; Varró, M. J.; Dédélé, A.; Gražulevičienė, R.; Mölter, A.; Lindley, S.; Madsen, C.; Cesaroni, G.; Ranzi, A.; Badaloni, C.; Hoffmann, B.; Nonnemacher, M.; Krämer, U.; Kuhlbusch, T.; Cirach, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Bellander, T.; Korek, M.; Olsson, D.; Strömgren, M.; Dons, E.; Jerrett, M.; Fischer, P.; Wang, M.; Brunekreef, B.; de Hoogh, K. Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmospheric Environment* **2013**, *47*, (10), 10-23.
25. Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.; Galassi, C.; Gražulevičienė, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek, M.; Lanki, T.; Lindley, S.; Madsen, C.; Molter, A.; Nador, G.; Nieuwenhuijsen, M.; Nonnemacher, M.; Pedeli, X.; Raaschou-Nielsen, O.; Patelarou, E.; Quass, U.; Ranzi, A.; Schindler, C.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Tsai, M. Y.; Yli-Tuomi, T.; Varro, M. J.; Vienneau, D.; Klot, S.; Wolf, K.; Brunekreef, B.; Hoek, G. Development of Land Use Regression models for PM_{2.5}, PM_{2.5} absorbance, PM₁₀ and PM_{coarse} in 20 European study areas; results of the ESCAPE project. *Environ Sci Technol* **2012**, *46*, (20), 11195-205.
26. Wang, M.; Beelen, R.; Bellander, T.; Birk, M.; Cesaroni, G.; Cirach, M.; Cyrus, J.; de Hoogh, K.; Declercq, C.; Dimakopoulou, K.; Eeftens, M.; Eriksen, K. T.; Forastiere, F.; Galassi, C.; Grivas, G.; Heinrich, J.; Hoffmann, B.; Ineichen, A.; Korek, M.; Lanki, T.; Lindley, S.; Modig, L.; Molter, A.; Nafstad, P.; Nieuwenhuijsen, M. J.; Nystad, W.; Olsson, D.; Raaschou-Nielsen, O.; Ragettli, M.; Ranzi, A.; Stempfelet, M.; Sugiri, D.; Tsai, M. Y.; Udvardy, O.; Varro, M. J.; Vienneau, D.; Weinmayr, G.; Wolf, K.; Yli-Tuomi, T.; Hoek, G.; Brunekreef, B. Performance of multi-city land use regression models for nitrogen dioxide and fine particles. *Environ Health Perspect* **2014**, *122*, (8), 843-9.
27. Holle, R.; Happich, M.; Lowel, H.; Wichmann, H. E. KORA--a research platform for population based health research. *Gesundheitswesen* **2005**, *67 Suppl 1*, S19-S25.
28. Birmili, W.; Stratmann, F.; Wiedensohler, A. Design of a DMA-based size spectrometer for a large particle size range and stable operation. *Journal of Aerosol Science* **1999**, *30*, (4), 549-553.
29. Pitz, M.; Birmili, W.; Schmid, O.; Peters, A.; Wichmann, H. E.; Cyrus, J. Quality control and quality assurance for particle size distribution measurements at an urban monitoring station in Augsburg, Germany. *J. Environ. Monit.* **2008**, *10*, (9), 1017-1024.
30. Ogawa Protocol for Ozone Measurement Using the Ozone Passive Sampler Badge, Revision 3; 2001.
31. Eeftens, M.; Tsai, M. Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach, M.; Cyrus, J.; Hoogh, K. d.; Nazelle, A. D.; Vocht, F. d.; Declercq, C.; Dedele, A.; Eriksen, K.; Galassi, C.; Gražulevičienė, 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. M., A.; Mosler, G.; Nieuwenhuijsen, M.; Oldenwening, M.; Pennanen, A.; Probst-Hensch, N.; Quass, U.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E. S., D.; Udvardy, O.; Vaskövi, E.; Weinmayr, G.; Brunekreef, B.; Hoek, G. Spatial variation of PM_{2.5}, PM₁₀, PM_{2.5} absorbance and PM_{coarse} concentrations between and within 20 European study areas and the relationship with NO₂ - results of the ESCAPE project. *Atmospheric Environment* **2012**, *46*, 303-317.
32. CORINE Land Cover Classes. <http://uls.eionet.europa.eu/CLC2000/classes/index.html>. (Accessed 19.11.2016).

33. 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 PM₁₀. *Environ Sci Technol* **2013**, *47*, (23), 13555-64.
34. GeoPortal <http://srtm.cgiar.org/>. . (Accessed 19.11.2016).
35. Birmili, W.; Sun, J.; Weinhold, K.; Merkel, M.; Rasch, F.; Spindler, G.; Wiedensohler, A.; Bastian, S.; Löschau, G.; Schladitz, A.; Quass, U. T. A.; Kuhlbusch, J.; Kaminski, H.; Cyrus, J.; Pitz, M.; Gu, J.; Peters, A.; Flentje, H.; Meinhardt, F.; Schwerin, A.; Bath, O.; Ries, L.; Gerwig, H.; Wirtz, K.; Weber, S. Atmospheric aerosol measurements in the German Ultrafine Aerosol Network (GUAN) - Part 3: Black Carbon mass and particle number concentrations 2009 to 2014. . *Gefahrstoffe – Reinhaltung der Luft* **2015**, *11/12*, 479-488.
36. Gu, J.; Pitz, M.; Schnelle-Kreis, J.; Diemer, J.; Reller, A.; Zimmermann, R.; Soentgen, J.; Stoelzel, M.; Wichmann, H. E.; Peters, A.; Cyrus, J. Source apportionment of ambient particles: Comparison of positive matrix factorization analysis applied to particle size distribution and chemical composition data. *Atmospheric Environment* **2011**, *45*, (10), 1849-1857.
37. Gu, J.; Schnelle-Kreis, J.; Pitz, M.; Diemer, J.; Reller, A.; Zimmermann, R.; Soentgen, J.; Peters, A.; Cyrus, J. Spatial and temporal variability of PM₁₀ sources in Augsburg, Germany. *Atmospheric Environment* **2013**, *71*, (0), 131-139.
38. UFIREG Environmental Health Report. [http://www.ufireg-central.eu/files/Downloads/UFIREG Environmental Health Report.pdf](http://www.ufireg-central.eu/files/Downloads/UFIREG_Environmental_Health_Report.pdf). (Accessed 19.11.2016).
39. 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. *Atmospheric Environment* **2012**, *54*, 634-642.
40. Wang, M.; Beelen, R.; Eeftens, M.; Meliefste, K.; Hoek, G.; Brunekreef, B. Systematic evaluation of land use regression models for NO₂. *Environ Sci Technol* **2012**, *46*, (8), 4481-9.
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.; Grazuleviciene, R.; Hoffmann, B.; Heinrich, J.; Iakovides, M.; Kunzli, N.; Korek, M.; Lindley, S.; Molter, 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.; Varro, 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), 4357-64.

Supplementary material Table 1: Description of predictor variables included in the final LUR models

Predictor variable	Pollutant	Mean	SD	Min	5%	25%	Median	75%	95%	Max
trafload_50	NO _x	586,308	841,295	0	0	29,314	91,503	750,055	2,174,623	2,562,487
trafload_100	PM _{2.5} abs	1,961,176	2,568,870	0	0	249,432	1,218,311	2,477,363	6,714,870	9,777,397
trafloadm_25	PM ₁₀	141,632	289,636	0	0	0	0	62,955	864,620	889,862
trafloadm_50	PNC	534,408	817,094	0	0	0	0	684,722	2,157,786	2,225,753
trafloadm_100	O ₃ ^a	1,776,468	2,590,972	0	0	0	1,001,735	2,271,340	6,193,628	9,777,397
roadl_50	PM _{2.5}	114.8	86.1	0	0	58.6	106.3	179.2	214.4	302.4
roadl_1000	PM ₁₀	28,331	13,511	0	1,0528	1,4586	31,139	39,546	44,662	45,706
roadlm_100	PM ₁₀ , NO ₂	137.9	164.5	0	0	0	142.0	197.6	511.5	544.0
intdistnearminv2	PM _{coarse}	18.03	28.2	0.02	0.04	0.16	1.9	22.6	82.5	83.9
abld_25	PNC, PM ₁₀ , PM _{2.5} abs	198.9	210.5	0.0	0.0	10.0	120.8	336.1	511.5	675.2
abld_50	O ₃ ^a	1,313.6	903.6	0.0	137.9	550.3	1,422.3	1,735.6	2,379.7	3,750.0
nblld_500	O ₃	966.8	462.2	4.0	152.2	749.3	1,117.5	1,262.3	1,543.1	1,677.0
nblld_1000	PM _{2.5} abs	3,081.9	1,747.0	6.0	390.8	1,686.8	3,482.0	4,175.8	5,341.6	6,625.0
pop_300	O ₃ ^a	1,098.1	844.0	1.0	30.5	482.3	980.2	1,528.4	2,327.2	3,029.3
industry_300	PNC, PM ₁₀ , PM _{coarse} , PM _{2.5} , NO _x	8.6	18.9	0.0	0.0	0.0	0.3	5.9	48.7	73.9
industry_1000	PM _{2.5} abs, NO ₂	10.5	9.9	0.0	0.0	2.2	8.3	15.4	21.6	41.8
industry_5000	NO ₂	8.3	7.0	0.5	0.5	1.7	8.1	15.2	18.6	19.1
seminat_100_neg	PNC, NO _x	-3.5	8.7	-33.5	-21.6	-0.7	0.0	0.0	0.0	0.0
seminat_1000_neg	PM _{2.5} , NO ₂	-16.0	16.1	-58.2	-47.7	-22.7	-9.8	-4.6	-2.3	-0.3
urbgreen_500_neg	PM ₁₀	-6.2	6.5	-23.6	-16.0	-10.6	-4.2	-1.6	0.0	0.0
green_500_neg	PNC	-16.1	11.3	-48.4	-34.0	-20.7	-16.3	-6.5	-3.1	-0.4
green_1000_neg	PNC, NO _x	-22.6	14.0	-59.3	-50.6	-26.3	-18.9	-12.3	-8.4	-8.1
water_5000_neg	PM _{2.5} abs	-1.7	1.0	-3.7	-3.0	-2.3	-2.0	-1.0	-0.1	-0.1
tot_build_25	PM _{2.5}	85.3	28.8	0.0	30.4	87.8	100.0	100.0	100.0	100.0
tot_build_1000	PM _{coarse}	57.5	28.3	0.1	21.5	32.9	68.0	81.7	88.8	91.9
elev_sqrt_neg	PM _{2.5} , O ₃ ^a	-21.9	0.8	-23.6	-22.7	-22.3	-22.0	-21.5	-20.3	-20.3

See Table 1 for a detailed explanation of the variable names. Variables with _X (e.g. trafload_50) are buffers with _X indicating the radius of the buffer in meters.

The suffix "_neg" indicates predictors with a negative sign.

^aPredictor variable was multiplied by -1 for the marked pollutant.

Supplementary material Table 2. Comparison between newly developed LUR models for the area of Augsburg for PM₁₀, PM_{coarse}, PM_{2.5}, PM_{2.5}absorbance, NO_x and NO₂ and ESCAPE models.

Pollutant	N	LUR model	R ²	Adj R ²	LOOCV R2	LOOCV Adj R2	Moran's I (p-value)
PM ₁₀	20	13.67 + 0.0078 * roadlm_100 + 0.000079 * roadl_1000 + 0.044 * industry_300 + 0.098 * urbgreen_500_neg + 0.0022 * abld_25 + 0.0000016 * trafloadm_25	0.91	0.87	0.78	0.76	-0.18 (0.15)
PM ₁₀ ESCAPE	20	18.47 + 0.039 * roadlm_50 + 0.57 * natural_100_neg ^c + 0.021 * roadl_50	0.83	0.80	0.75	0.73	-0.08 (0.82)
PM _{coarse} ^a	19	168.7 + 0.031 * tot_build_1000 + 0.030 * intdistnearminv2 + 0.025 * industry_300 + 0.000037 * xcoord_neg	0.75	0.68	0.57	0.55	-0.11 (0.60)
PM _{coarse} ESCAPE	20	4.09 + 0.025 * roadlm_50 + 0.0000042 * pop_5000 + 0.012 * roadl_50	0.81	0.78	0.69	0.67	-0.15 (0.36)
PM _{2.5}	20	19.47 + 0.0099 * roadl_50 + 0.041 * seminat_1000_neg + 0.41 * elev_sqrt_neg + 0.0094 * tot_build_25 + 0.012 * industry_300	0.84	0.79	0.70	0.69	-0.12 (0.49)
PM _{2.5} ESCAPE	20	11.9 + 0.019 * roadlm_50 + 0.000495 * roadl_300 + 0.14 * urbgreen_5000_neg + 0.0000000074 * trafloadm_1000	0.78	0.72	0.62	0.60	-0.13 (0.49)
PM _{2.5} abs ^b	19	-14.07 + 0.000000037 * trafload_100 + 0.00047 * abld_25 + 0.0087 * industry_1000 + 0.000058 * nblld_1000 + 0.040 * water_5000_neg + 0.0000034 * xcoord	0.93	0.89	0.84	0.83	-0.14 (0.37)
PM _{2.5} abs ESCAPE	20	1.34 + 0.000000177 * trafload_50 + 0.0018 * roadl_50 + 0.000216 * trafloadm_1000	0.91	0.89	0.82	0.81	-0.19 (0.21)
NO ₂ ^b	19	12.57 + 0.22 * industry_5000 + 0.015 * roadlm_100 + 0.10 * seminat_1000_neg + 0.15 * industry_1000	0.95	0.94	0.90	0.89	-0.10 (0.64)
NO ₂ ESCAPE	40	7.432 + 0.0000020 * trafload_50 + 0.0014 * intdistnearminv + 0.024 * roadl_50 + 0.000015 * pop_5000 + 0.041 * roadlm_50 + 0.098 * hldres_500	0.86	0.83	0.67	0.66	-0.04 (0.86)
NO _x	20	28.06 + 0.0000084 * trafload_50 + 0.27 * green_1000_neg + 0.20 * industry_300 + 0.25 * seminat_100_neg	0.91	0.89	0.83	0.82	-0.05 (0.94)
NO _x ESCAPE	40	13.34 + 0.0000039 * trafload_50 + 0.0897 * roadlm_50 + 0.0038 * intdistnearminv + 0.000000025 * trafload_1000 + 0.051 * roadl_50 + 0.195 * hldres_1000	0.88	0.85	0.76	0.76	-0.05 (0.65)

^aWithout rural site 4 (see Figure 1).

^bWithout traffic site 20 (see Figure 1).

^cSum of predictors seminat_100 and water_100.

Supplementary material Figure 1. Scatter plots of measured and predicted concentrations together with linear regression lines (one monitor excluded for each PM_{coarse} , $PM_{2.5abs}$ and NO_2).

