Environmental Science & Technology

Article

pubs.acs.org/est

Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation

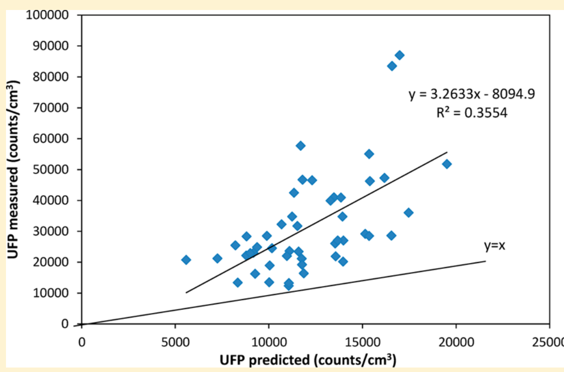
Denise R. Montagne,[†] Gerard Hoek,[*,†] Jochem O. Klompmaker,[†] Meng Wang,[†] Kees Meliefste,[†] and Bert Brunekreef[†,‡]

[†]Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, 3584 CK Utrecht, The Netherlands

[‡]Julius Center for Health Sciences and Primary Care, University Medical Center, University of Utrecht, 3584 CK Utrecht, The Netherlands

Supporting Information

ABSTRACT: Health effects of long-term exposure to ultrafine particles (UFP) have not been investigated in epidemiological studies because of the lack of spatially resolved UFP exposure data. Short-term monitoring campaigns used to develop land use regression (LUR) models for UFP typically had moderate performance. The aim of this study was to develop and evaluate spatial and spatiotemporal LUR models for UFP and Black Carbon (BC), including their ability to predict past spatial contrasts. We measured 30 min at each of 81 sites in Amsterdam and 80 in Rotterdam, The Netherlands in three different seasons. Models were developed using traffic, land use, reference site measurements, routinely measured pollutants and weather data. The percentage explained variation (R^2) was 0.35−0.40 for BC and 0.33−0.42 for UFP spatial models. Traffic variables were present in every model. The coefficients for the spatial predictors were similar in spatial and spatiotemporal models. The BC LUR model explained 61% of the spatial variation in a previous campaign with longer sampling duration, better than the model R^2. The UFP LUR model explained 36% of UFP spatial variation measured 10 years earlier, similar to the model R^2. Short-term monitoring campaigns may be an efficient tool to develop LUR models.



■ INTRODUCTION

Studies of health effects of outdoor (traffic related) air pollution have focused on particulate matter with a diameter of less than 2.5 ($PM_{2.5}$) or 10 μm (PM_{10}), black carbon (BC), and nitrogen dioxide (NO_2). It has been suggested that ultrafine particles (UFP) have a high penetration rate and are biologically more reactive than larger particles.[1,2] UFP are airborne nanoparticles with a diameter less than 100 nm, which account for a large fraction of the total particle number, while contributing little to ambient particle mass.[3] BC is created by incomplete combustion. BC may be a useful indicator of health effects related to particulate matter especially at the local scale.[4]

Frequently, land use regression (LUR) models are used in epidemiological studies to estimate long-term exposure to ambient air pollution for participants in such studies.[5] LUR models for particulate matter combine measurements at typically 20−40 locations and predictor variables (traffic, land use) in an empirical statistical model. Only a few LUR models have been developed for UFP,[6−11] because of monitoring issues such as the cost of equipment and problems related to leaving equipment unattended for periods of 1−2 weeks, the typical duration of purpose-designed sampling campaigns. UFP is usually not monitored in routine monitoring networks. To capture the high spatial variation of UFP, most previous studies have used mobile or short-term monitoring campaigns, typically with short (15 min to hours rather than days to weeks) observation periods at each measurement site.[6−10] The only UFP LUR model based upon fixed monitoring was derived in Amsterdam using data collected for evaluation of spatiotemporal patterns across the city in a large EU funded study.[11] In most short-term monitoring studies a site was measured only once.[6−8] These short-term measurements likely have more temporal variability and might therefore be less precise in determining spatial variation of long-term average concentrations. Most previous LUR studies have developed models that explain variation of measured average air pollution concentrations with time-invariant spatial predictors. Because of the inherently larger temporal variation of short-term measurements, LUR studies have applied spatiotemporal modeling.[6,8] Spatiotemporal models included temporal predictors (hourly measurements of routinely measured pollutants or weather from a fixed site) as well as spatial variables to explain variation of measured air quality. Another approach

Received: November 28, 2014
Revised: June 15, 2015
Accepted: June 16, 2015
Published: June 16, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/es505791g
Environ. Sci. Technol. 2015, 49, 8712−8720

consists of on-road mobile monitoring with typically even shorter sampling at specific locations but more repeats.[10,12] Studies have used mobile and short-term campaigns to develop models for BC as well.[6,12] LUR models for $PM_{2.5}$ absorbance (a marker for BC) have also been developed based on study designs with sampling times of weeks.[13,14] Independent validation of the models has received limited attention in LUR modeling studies based upon mobile campaigns. Recent methodological work has shown the importance of independent validation of LUR models based upon the typically relatively small number of sites using fixed monitoring.[15−17] It is not clear whether the same applies to mobile or short-term monitoring campaigns with larger number of sites but shorter sampling durations.

The aim of the Measurements of UFP and Soot in two Cities (MUSiC) project, was to develop and evaluate LUR models for UFP and BC in Rotterdam and Amsterdam (The Netherlands) based on short-term monitoring. The second aim was to assess the validity of the LUR models in predicting previously and independently measured spatial contrasts of UFP and BC.[11,14] The design of the monitoring campaign and an evaluation of within site temporal and between site spatial concentration variability has been reported separately.[18] The short-term monitoring study had a much higher ratio of within-site to between-site concentration variability, compared to studies with longer sampling times.[18] We develop spatiotemporal models in addition to spatial models, to account for the high within-site temporal variability.

■ METHODS

Study Design. The development of the LUR models involved monitoring of UFP and BC at 161 sites in Amsterdam and Rotterdam, the collection of predictor variables using a geographic information system (GIS) and regression modeling to link monitoring results and predictor variables.[5]

The monitoring campaign included 81 sites in Amsterdam and 80 sites in Rotterdam, representing a large spatial contrast in traffic characteristics and land use. Approximately 30 sites per city were traffic sites. Other site types were urban background, regional background, urban green, highway sites, and sites adjacent to water bodies. Sites near water bodies were chosen to determine the influence of shipping emissions. Traffic sites were defined as sites at roads with more than 10 000 vehicles per day.

Measurements were performed during 30 min per site. Measurements were conducted at each site in three seasons; winter, spring and summer. In total, 483 measurements at 161 sites were conducted between January and July 2013. Measurements were taken between 9:30 am and 4:00 pm, to increase comparability of measurements across sites. The equipment was placed in an electric vehicle (REVA, Mahindra Reva Electric Vehicles Pvt. Ltd., Bangalore, India). The sampling heights were approximately 1.5 m. During the entire measurement campaign, measurements were performed at a single reference site in Utrecht using the same sampling equipment. One reference site was selected in the center of the country to be able to combine measurements from Amsterdam and Rotterdam using a single source for temporal adjustment.

UFP was monitored each second with CPC 3007 instruments (TSI, Tennessee, USA). The CPC 3007 measures particles from 10 nm to above 1 μm and thus does not specifically measure UFP. However, UFP typically dominates total particle number counts.[3] BC was measured averaged over each minute, using the micro Aethalometer (Aethlabs, CA, USA). All further calculations use the 30 min average concentration. QA/QC included zero checks and colocation of instruments.[18]

Adjustment for Temporal Variation. The concentrations at the reference site were used to adjust the concentrations at the short-term monitoring sites for temporal variability (day to day and within day), so that the spatial contrast between sites can be determined. First, the corresponding 30 min mean concentration at the reference site is subtracted from the overall mean reference site concentration. This difference is added to the 30 min mean concentration at the sites. Finally the average concentration from the three adjusted 30 min mean concentrations per site was calculated. The difference method performed better than the ratio adjustment method in our study.[18]

LUR Model Development. The LUR models were developed using a methodology that has previously been successfully applied in The Netherlands to model the spatial variation of the concentration of $PM_{2.5}$, NO_2, and the soot content of fine particles.[14] This methodology was developed within the framework of the European Study of Cohorts for Air Pollution Effects (ESCAPE). In brief, the average concentration per site was used as the dependent variable entered in a linear regression analysis with several GIS variables as independent variables. The offered predictor variables can be found in the Supporting Information (Table S1). The GIS variables were selected using a supervised stepwise selection procedure. The direction of the effect for the variables was determined a priori and the variable with the highest adjusted R^2 (coefficient of determination) was entered in the model. The next variable was added when the adjusted R^2 increased more than 1%. The variables in the resulting models were checked for p-value (removed when p-value >0.10), colinearity (variance inflation factor >3 were removed), and influential observations (if Cook's D > 1 the model was further examined). The final models were evaluated by Hold-out validation (HV) and Moran's I to detect possible spatial autocorrelation in the residuals. We used universal Kriging models to further evaluate spatial autocorrelation, using the GSTAT package in R3.02 to calculate variograms. Hold-out validation (against data from sites not included in model building) reflects the true prediction ability of LUR models to the independent locations not used for modeling. For our HV, half of the sites were randomly selected for model building and the other half were used to validate the predicted values.[17] This procedure was repeated 10 times. The data sets were stratified to ensure that half of the traffic sites were included in every training and test set.

For the spatial model site-specific adjusted averages were used following the procedure outlined above. Models were developed for Amsterdam and Rotterdam separately and for the two cities pooled. Sites with <2 samples were excluded from all analyses, resulting in the exclusion of one site for BC and two different sites for UFP.

Spatiotemporal models were developed using the individual unadjusted 30 min mean measurements, with different temporal predictors in the model to account for temporal variation. We evaluated the corresponding reference site 30 min mean concentrations, weather data from the nearest weather station and NO_2 measurements in the cities at routine urban background stations and combinations of these predictors. To further interpret differences between spatial and spatiotemporal

DOI: 10.1021/es505791g
Environ. Sci. Technol. 2015, 49, 8712−8720

models, a fixed spatiotemporal model was developed, which included the spatial variables from the pooled spatial models.

The coefficients for each predictor in the final models were multiplied with the difference between the 90th and the 10th percentile of the predictor variable to compare the impact of predictors with different variability and units on the concentrations.

Comparison with Previously Measured Concentrations. To further assess the ability of the developed LUR models to predict spatial concentration contrasts at sites not used in model development, we collected data from two previous Dutch LUR studies, RUPIOH and ESCAPE.[11,14] In the ESCAPE study, annual average $PM_{2.5}$ absorbance concentrations were measured at 40 monitoring sites spread over The Netherlands/Belgium. The averages were based upon three 14-day average concentrations measured in three seasons in 2009 and adjusted for temporal variation using the same reference site as in the current study.[14] In the RUPIOH study, averages were based upon at most seven 24-h average UFP concentrations at 48 locations across Amsterdam between October 2002 and April 2004, adjusted for temporal variation using UFP data from an urban background reference site in the city.[11] LUR model predictions were compared with past measured adjusted average concentrations.

■ RESULTS

Figure 1 illustrates the substantial variability of the adjusted mean UFP and BC concentrations at the 161 sites.

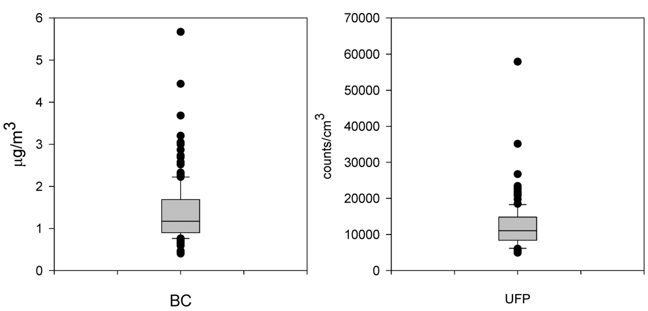


Figure 1. Boxplot of BC ($\mu g/m^3$) and UFP ($counts/cm^3$) concentrations (adjusted average per site) $N = 161$.

Spatial Models Per City. The LUR models for UFP and BC in Amsterdam and Rotterdam are described in Table 1. The Cook's D test showed no influential observations. All Moran's I values were small and generally nonsignificant indicating no evidence of spatial autocorrelation. For the Rotterdam UFP model there was statistically significant spatial autocorrelation but with a near-zero Moran's I of 0.013. The percentage of explained variation (R^2) for BC was 0.40 and 0.41 in Amsterdam and Rotterdam. For UFP, R^2 was 0.33 and 0.42 in Amsterdam and Rotterdam, respectively.

Population within a 5000 m buffer was a predictor in the Rotterdam and Amsterdam models, both for UFP and BC. The models included two to four variables and every model included traffic variables. As indicated by the coefficients multiplied by the difference between the 90th and the 10th percentile of predictors, the models predict sizable contrasts.

In Amsterdam, the training set models from the HV predicted 16% less variation for BC and 13% for UFP in the test sets compared to the full model R^2. In Rotterdam the gap between the full model R^2 and HV was larger, 26% for BC and 22% for UFP. Absolute values for HV R^2 were low for all models.

Spatial Pooled Models. Table 2 shows the LUR models for the pooled adjusted BC and UFP concentrations. The explained variation for the pooled model was slightly lower than the R^2 for the models per city for BC and in between for UFP. The pooled BC model included three and the UFP model four variables. For BC, population density and inverse distance to the nearest major road were present in both city models and in the pooled model, with a slope that was comparable to the slope in the Amsterdam model. For UFP, only the population variable was present in all three models with similar slopes. Furthermore, the pooled model contained two traffic variables and a variable for port area.

For both BC and UFP, the gap between the full model R^2 and the R^2 of HV test set predictions was about 10%, substantially lower than for the city-specific models. Absolute HV values were higher than for the city-specific models but remained low.

Both for BC and UFP, the variogram showed no patterns of semivariance with distance, further supporting the lack of spatial autocorrelation. In urban study areas, with sites affected differently by local sources located at relatively small distances from each other, typically no spatial autocorrelation of LUR model residuals has been found.[5]

To interpret differences in model structures, correlations between GIS predictor variables are included in Supporting Information Table S2. Several variables representing small-scale traffic impacts were highly correlated (for example traffic on nearest street and heavy traffic on nearest street).

The pooled model structure applied to the individual cities had very similar model R^2 values compared to the city-specific models: R^2 = 0.40 and 0.34 for BC in Amsterdam and Rotterdam and R^2 = 0.36 and 0.40 for UFP in Amsterdam and Rotterdam, respectively. While the Rotterdam UFP model contains different variables than the pooled model, the variables from the pooled model were all significant when applied to Rotterdam only with an R^2 of 0.40, virtually identical to the R^2 of 0.42 for the Rotterdam model in Table 1.

Spatiotemporal Models. In Table 3 and Supporting Information Table S3, the spatiotemporal LUR models are described. The model R^2 values were very similar to the pooled spatial models. The BC model included two spatial and three temporal variables. The spatial variables inverse distance to the nearest major road and traffic intensity on the nearest road were present in the pooled spatial and in the spatiotemporal model. The slopes for both variables had similar magnitude as in the pooled models. The population variable of the spatial model was not included in the spatiotemporal model, because it did not add more than 1% explained variability. When forced into the model, the slope of the population variable was only 22% lower than the slope of the pooled model (Supporting Information Table S3). The gap between the full model R^2 and the median HV R^2 was 29%, substantially larger than for the pooled spatial model.

For UFP, three spatial and three temporal variables were included in the model. The three spatial variables were identical (inverse distance to the nearest major road and traffic load within a buffer of 100 m) or similar (population in a 1000 m buffer instead of a 5000 m buffer) to the pooled spatial model. The port variable of the spatial model was not included in the spatiotemporal model and had a smaller slope when forced into the model (Supporting Information Table S3). For UFP the

DOI: 10.1021/es505791g
Environ. Sci. Technol. 2015, 49, 8712−8720

Table 1. Spatial LUR Models for the Adjusted Average BC and UFP Concentrations Per City

variables in LUR model	BC ($\mu g/m^3$) model		UFP ($counts/cm^3$) model	
	Amsterdam	Rotterdam	Amsterdam	Rotterdam
intercept	0.46	0.46	4172	4648
inverse distance major road	0.38[a]		6308	
inverse distance major road squared		0.16		
traffic nearest street	0.53			
heavy traffic nearest street				1580
traffic load 100 m buffer				
major road length 50 m buffer		0.67		3887
port 5000 m buffer				
population 1000 m buffer				
population 5000 m buffer	0.42	0.56	4373	4443
industry 500 m buffer		0.01		
R^2 of model	0.40	0.41	0.33	0.42
HV median[b] (min−max) R^2	0.24 (0.11−0.45)	0.15 (0.08−0.27)	0.20 (0.13−0.45)	0.20 (0.12−0.28)
number of sites[c]	81	79	80	79
mean (min−max) measured concentrations	1.39 (0.45−5.67)	1.36 (0.40−3.68)	12780 (5282−57897)	11480 (4902−23414)

[a]Regression slopes were multiplied by the difference between the 10th and 90th percentile for all predictors. See Supporting Information (Table S1) for more detailed explanation of the variables. [b]Hold-out validation median (minimum−maximum) squared correlation (R^2) of the 10 training set models predictions for the test sets. [c]Number of sites that have been used for model development.

Table 2. Spatial Pooled LUR Models for the Adjusted Average BC ($\mu g/m^3$) and UFP ($counts/cm^3$) Concentrations

variables in LUR model	BC ($\mu g/m^3$)	UFP ($counts/cm^3$)
intercept	0.62	3221
inverse distance major road	0.52[a]	4552
inverse distance major road squared		
traffic nearest street	0.30	
heavy traffic nearest street		
traffic load 100 m buffer		1740
major road length 50 m buffer		
port 5000 m buffer		2255
population 1000 m buffer		
population 5000 m buffer	0.37	3959
industry 500 m buffer		
R^2 of model	0.35	0.37
HV median[b] (min−max) R^2	0.24 (0.12−0.39)	0.29 (0.19−0.48)
number of sites[c]	160	159
mean (min−max) measured concentrations	1.38 (0.40−5.67)	12134 (4902−57897)

[a]Regression slopes were multiplied by the difference between the 10th and 90th percentile for all predictors. [b]Hold-out validation median (minimum−maximum) squared correlation (R^2) of the 10 training set models predictions for the test sets. [c]Number of sites that have been used for model development.

13% difference between full model and HV R^2 was larger than for the pooled spatial model difference (8%).

The best spatiotemporal models for UFP (BC) were developed with the UFP (BC) measured at the reference site in Utrecht, the NO_2 concentration from routine monitoring within Amsterdam or Rotterdam and weather data as temporal predictors. Models with only the reference site measurement or routinely measured NO_2 had R^2 values of 0.32−0.33 for both components (Supporting Information Table S3), substantially lower than the full model (Table 3). The spatial component of the models was similar between models.

Prediction of Previously Measured Spatial Contrasts. The pooled spatial BC model predicted spatial variation of $PM_{2.5}$ absorbance at 40 sites measured in 2009 across The Netherlands very well (Figure 2a). Remarkably, the R^2 of 0.61 was larger than the model R^2 and the hold-out validation R^2. Absolute concentration levels cannot be directly compared because of the different metrics. The pooled spatial UFP model predicted spatial variation at 48 sites measured in 2002−2004 in Amsterdam fairly well (Figure 2b). The R^2 of 0.36 was similar to the model R^2 and higher than the hold-out validation R^2. Absolute concentration levels were much higher in the 2002−2004 campaign than predicted by the model. The spatial model developed for Amsterdam specifically (Table 1) predicted slightly less variation of UFP in 2002−2004(R^2 = 0.33). The pooled spatial BC model predicted annual average PM2.5 absorbance in 2002−2004 across the Amsterdam sites better (R^2 = 0.41) than the model R^2 of 0.35 suggested (Figure 2c).

■ DISCUSSION

Land use regression models have been widely applied to model the spatial variation of outdoor air pollution, particularly to assign long-term average air pollution exposures to participants of epidemiological studies.[5] A recent development is the use of mobile and/or short-term monitoring to provide the

DOI: 10.1021/es505791g
Environ. Sci. Technol. 2015, 49, 8712−8720

Table 3. Spatiotemporal LUR Models for BC and UFP, Including Temporal and Spatial Variables

variables in LUR model	BC ($\mu g/m^3$)	UFP ($counts/cm^3$)
intercept	−0.71	2304
spatial predictors		
inverse distance major road	0.47[a]	4362
inverse distance major road squared		
traffic nearest street	0.44	
heavy traffic nearest street		
traffic load 100 m buffer		1603
major road length 50 m buffer		
port 5000 m buffer		
population 1000 m buffer		3821
population 5000 m buffer		
industry 500 m buffer		
temporal predictors		
reference site BC/UFP	0.92	5355
NO_2 routine monitoring	1.13	5781
temperature	0.59	
relative humidity		−2812
R^2 of model	0.50	0.39
HV median[b] (min−max) R^2	0.21 (0.08−0.30)	0.26 (0.16−0.38)
number of samples[c]	446	417
mean (min−max) measured concentrations	1.42 (−0.23−6.26)	12261 (1746−74009)

[a]Regression slopes were multiplied by the difference between the 10th and 90th percentile for all predictors. [b]Hold-out validation median (minimum−maximum) squared correlation (R^2) of the 10 training set models predictions for the test sets [c]Number of samples that have been used for model development.

monitoring base for development of LUR models for particulate matter.[12] Mobile monitoring involves very short monitoring (<1 h) at a larger number of locations than in most previous LUR studies based on repeated 1−2 week sampling. We will discuss three main issues of mobile monitoring. First, we provide an interpretation of the typically lower explained variance (R^2) of LUR models based upon short-term monitoring compared to monitoring with longer sampling times. Second, we will discuss the merits of different modeling approaches, specifically spatial versus spatiotemporal models and local city-specific versus pooled models, including a comparison of model structures. Third, advantages and disadvantages of short-term monitoring versus fixed monitoring will be discussed.

Moderate Explained Variability of Short-Term Monitoring Campaign LUR Models. Our LUR models based on short-term monitoring had only moderate model R^2 and low hold-out validation R^2, consistent with previous studies using mobile or short-term monitoring campaigns (Table 4). The higher model R^2 in Basel than in other short-term studies was especially due to the better prediction by the suburban background UFP concentrations (a temporal predictor), which alone explained 38% of the variation of the 20 min mean UFP concentrations. The model R^2 of our and previously reported UFP and BC models based on mobile campaigns, were lower than reported for LUR models based upon monitoring with longer averaging times for $PM_{2.5}$, NO_2, and BC/EC/$PM_{2.5}$ absorbance.[5,14,20] The model R^2s for BC and UFP in our study were substantially lower compared to the ESCAPE $PM_{2.5}$ absorbance models (Netherlands model R^2 0.92). The *first* and probably main explanation for the lower R^2 in our spatial models is the lower precision of the average concentrations compared to campaigns with longer averaging times. In our study, measurements were performed for 30 min during three seasons, whereas in the ESCAPE study three two week measurement campaigns were performed. The shorter sampling duration resulted in a much larger within site (temporal) variability than for the ESCAPE study.[18] In the current study, the variance ratio (within temporal/between site spatial variation) was 2.44 for BC and 2.17 for UFP after adjusting for temporal variation versus 0.09 for $PM_{2.5}$ absorbance in ESCAPE and 0.31 for UFP in RUPIOH.[18] This indicates that temporal variation remained in the adjusted average concentrations, which cannot be modeled with fixed spatial predictors. *Second*, UFP may be harder to model, because of its reactivity and important local sources, resulting in high spatial and temporal variation. Some support for this explanation is provided by the previous LUR model for Amsterdam, in which the model R^2 was slightly lower for UFP than for $PM_{2.5}$ absorbance (0.67 versus 0.76).[11] *Third*, model R^2 values based on small data sets overestimate the predictive ability in independent test sets.[15−17] Fixed campaigns are typically based on fewer sites than the mobile and short-term campaigns. Validating the models with Hold-out validation (HV) instead of leave one out cross-validation (LOOCV) is preferable if the number of sites allows it. HV has not been applied much in mobile campaign LUR studies. The results from the Hold-out validation (HV) showed that the model R^2 values overestimate the predictive ability by 8−13% for the pooled models based on 161 sites. This is a moderate gap, though larger than reported previously for long sampling durations with a similar number of model training sites.[15−17] Since we used 50% of the sites for validation rather than the full number of sites, it is possible that our HV R^2 underestimated the prediction ability of the full model. The HV R^2 values were substantially lower than the ESCAPE Hold-out validation R^2 values for NO_2 and $PM_{2.5}$ absorbance, suggesting that less overfitting related to the larger number of sampling sites in the short term campaigns does not fully explain the lower R^2.[15]

DOI: 10.1021/es505791g
Environ. Sci. Technol. 2015, 49, 8712−8720

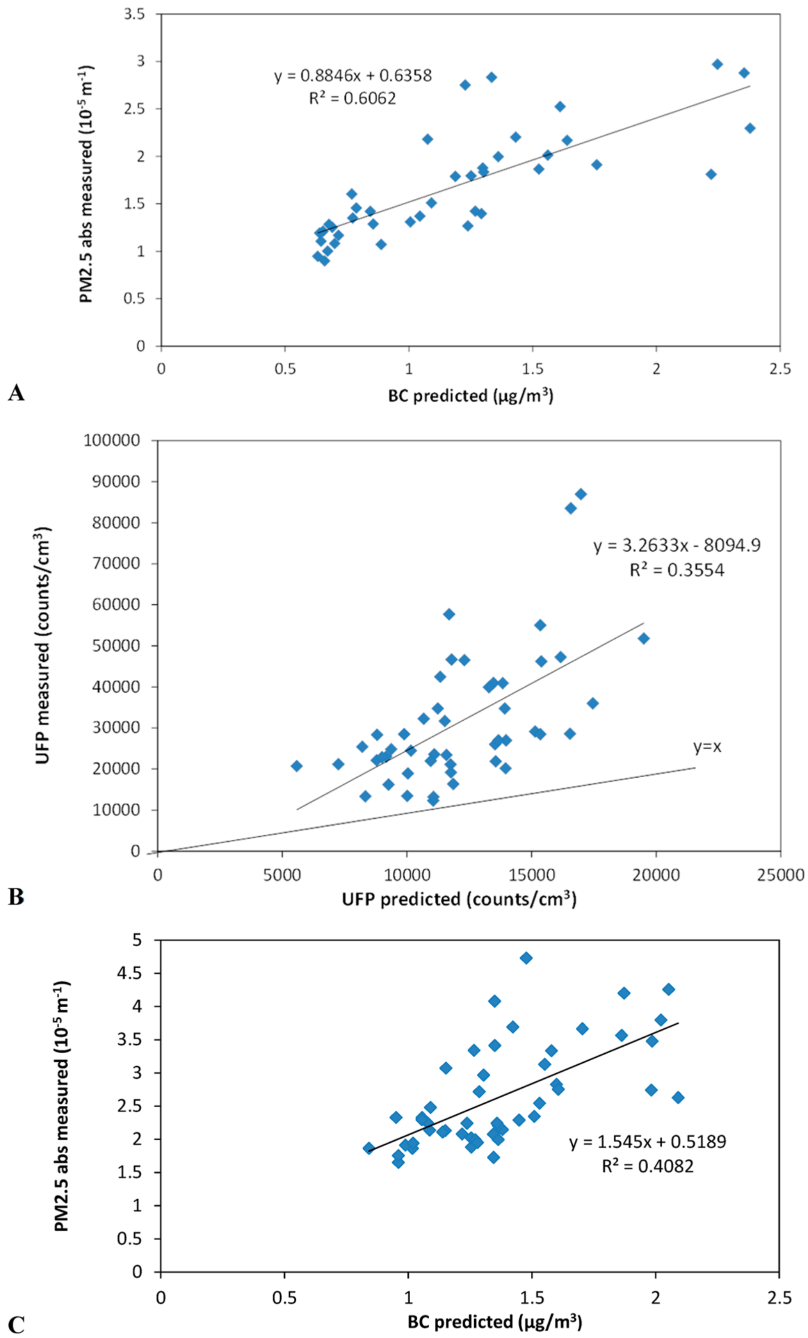


Figure 2. (a) Comparison of LUR predicted and externally measured annual average BC across The Netherlands. (b) Comparison of LUR predicted and externally measured annual average UFP in Amsterdam. (c) Comparison of LUR predicted and externally measured annual average BC in Amsterdam. Panel a was measured from ESCAPE conducted at 40 sites in 2009 across The Netherlands. Panels b and c measured from RUPIOH study conducted in 2002−2004 at 48 sites in Amsterdam.[11] BC and $PM_{2.5}$ absorbances are two highly correlated methods of measuring black carbon.

We conclude that comparing model or hold-out validation R^2 between short-term and longer term monitoring campaigns is likely not appropriate because of the difference in precision of the measurements. Comparison of model predictions based on the same external data set would provide an appropriate comparison, but such evaluations have not been performed. In the last section, we will compare how well our BC model predicted the ESCAPE $PM_{2.5}$ absorbance measurements compared to the hold-out validation of the ESCAPE model, which is based on fixed monitoring.

Prediction of Previous Spatial Contrasts from External Data. Despite the moderate model R^2, the LUR models predicted spatial contrasts of UFP and especially BC determined in fully independent studies in the past well. The percentage explained variation of the spatial contrast observed in these studies was equal (UFP) or even larger (BC) than the model R^2 in the current study and much larger than the holdout validation R^2. The explained variance is remarkably high considering the difference in time period (three and ten years prior to the current sampling campaign), differences in site selection (on the street near the façade in the current campaign versus equipment at homes with traffic locations usually measured at first floor balconies in ESCAPE and RUPIOH) and different monitoring equipment (CPC 3007 in MUSIC

DOI: 10.1021/es505791g
Environ. Sci. Technol. 2015, 49, 8712−8720

Table 4. Comparison with LUR Models from Previous UFP and BC Mobile and Short-Term Sampling Campaigns

reference	location	component	design	model type	model R^2	validation R^{2a}
Larsson, 2009[12]	Vancouver, Canada	BC	mobile monitoring around 39 intersections, 8 3 h afternoon sessions	spatial	0.54−0.72	LOOCV: 0.40−0.62
Zwack, 2011[9]	Brooklyn, NY	UFP	mobile monitoring along fixed walking routes (∼3 h per route)	spatiotemporal	0.22−0.32	NA
Rivera, 2012[8]	Girona, Spain	UFP	single 15 min measurement at 644 sites	spatial	0.36	LOOCV: 0.35
				spatiotemporal	0.51	LOOCV: 0.48
Abernethy, 2013[7]	Vancouver, Canada	UFP	single 60 min measurement at 80 sites	spatial	0.29−0.53	LOOCV: 0.14−0.41
Saraswat, 2013[6]	New Delhi, India	UFP	single 1−3 h measurement at 39 sites.	spatiotemporal	0.23−0.28	LOOE: 0.71−0.59
		BC	single 1−3 h measurement at 39 sites	spatiotemporal	0.69−0.86[b]	LOOE: 0.44−0.34
Patton, 2014[10]	Somerville, MA	UFP	mobile monitoring car driving on 43 days (3−6 h per day)	spatiotemporal	0.41−0.43	NA
Ragettli, 2014[19]	Basel, Switzerland	UFP	three 20 min measurements at 60 sites	spatiotemporal	0.58−0.68[b]	LOOCV: 0.58−0.68
this study	Amsterdam and Rotterdam, NL	UFP	three 30 min measurements at 161 sites	spatial	0.33−0.42	HV: 0.20−0.29
				spatiotemporal	0.39	HV: 0.26
		BC		spatial	0.35−0.41	HV: 0.15−0.24
				spatiotemporal	0.50	HV: 0.21

[a]NA = not available; LOOCV = leave one out cross-validation; LOOE = leave one out evaluation; HV= hold out validation. [b]High model R^2 are mostly due to the temporal variables (reference site concentration).

versus CPC 3022A in RUPIOH). For ESCAPE, the study area included the entire country, whereas the current model was developed in the two major cities only.

The more precise assessment of the site-specific average concentrations due to the longer sampling duration in ESCAPE and RUPIOH likely explains the relatively high validation R^2. The findings further suggest that with a large number of short-term monitoring sites robust models can be developed that predict spatial variation (fairly) well despite the temporal variation. This is consistent with the relatively robust spatial predictor estimates in models developed within our study with different methods (spatial versus spatiotemporal). The observation of a robust spatial model and a moderate R^2 fits with the effect of measurement error in continuous dependent variables in regression analysis.[21] Measurement error in a continuous dependent variable does not result in biased regression coefficients but it does result in a loss in precision and power. Thus, with a sufficiently large sample size, the identified model may be correct, but the explained variance of the model will be lower if more measurement error is present in the dependent variable. For the application in epidemiological studies of long-term exposure to air pollution, the comparison with annual averages at independent sites is more important than the model R^2.

The current LUR predicted $PM_{2.5}$ absorbance across The Netherlands in ESCAPE better than UFP and $PM_{2.5}$ absorbance in Amsterdam in RUPIOH, consistent with the longer sampling duration in ESCAPE (three measurements of 2 weeks duration) compared to RUPIOH (seven 24 h measurements) and the resulting lower within/between concentration variance ratios. A second explanation is that the ESCAPE campaign was conducted three years before the MUSIC campaigns and the RUPIOH campaign ten years earlier. In contrast, the BC model was applied to the entire Netherlands, whereas it was developed in the two major cities only; the UFP model was applied in Amsterdam, within the domain of development.

The UFP model predicted much lower UFP concentrations than measured in 2002−2004. Potential reasons include differences in equipment (CPC 3022 used in 2002 measures smaller particles than the CPC 3007 used in 2013), sites and weather circumstances. A more likely explanation for the large difference in measured concentrations is a trend in ultrafine particle concentrations. A recent review documented a reduction in UFP concentrations in European and North American cities.[22] The reduction was attributed to the increased use of diesel particulate filters and a reduced sulfur content of diesel.[22]

Spatial versus Spatiotemporal Models. Spatial and spatiotemporal models have been used to develop models from short-term and mobile monitoring studies (Table 4). We cannot directly compare the model R^2 between spatial and spatiotemporal models, because they explain variability of site average concentrations in spatial models versus individual 30 min concentrations in spatiotemporal models. Furthermore, temporal variables are included in the spatiotemporal models. A merit of spatiotemporal models is that these models can incorporate more temporal predictors than the concentration of UFP at a reference site, whereas the adjustment procedures used in calculating adjusted averages in the spatial models are based upon a single reference site concentration. In our study, a model with reference site UFP (BC), routinely measured NO_2 and weather included as temporal predictors, had substantially higher model R^2 than models with single temporal predictors.

Adjustment for temporal variation was further less effective for our short-term campaign compared to the longer sampling campaigns, as indicated by the much smaller reduction in within/between site variance ratios after adjustment for reference site concentrations,[18] suggesting that adjusted averages may still contain temporal variation.

The spatial variables and their coefficients of the spatiotemporal models were similar to those of the pooled spatial models. Consistently, the explained variance of the identified spatiotemporal model (Table 3) and a spatiotemporal

DOI: 10.1021/es505791g
Environ. Sci. Technol. 2015, 49, 8712−8720

model using the variables of the pooled spatial model was similar (Supporting Information Table S3).

In a study in Girona, adding hour of the day and sampling date to the model, improved the model R^2 from 36% to 51%, but the coefficients of the spatial variables were essentially unchanged.[8] The robustness of spatial predictors is likely due to the design of the sampling campaign which limited the correlation between temporal and spatial variation. We specified that on a specific sampling day all major site types had to be measured and furthermore that traffic sites should not be measured exclusively in the morning but throughout the day.

The coefficients for the spatial predictors did not differ materially between the different spatiotemporal models, suggesting that if prediction of average spatial contrasts is of interest valid models may be obtained even without a reference site for UFP or BC. If spatiotemporal contrasts are of interest, the observation that the model R^2 improved significantly when reference site UFP (BC) was added to a model including routinely measured NO_2, supports the need for a reference site with UFP (BC) measurements. The spatiotemporal models might be useful in time-series or birth cohort studies, which need to predict the exposure in a shorter time frame than the annual average.

Pooled versus City-Specific Models. The pooled model R^2 was slightly smaller than the city specific model R^2 values. However, the gap between the model R^2 and the median HV R^2 was lower for the pooled models (8−11%) than for the city specific models (13−26%). This suggests that pooling the sites increases the predictive ability of the models in independent test sets. Consistently, the pooled model predicted the 2002−2004 UFP spatial contrast in Amsterdam slightly better than the Amsterdam model based upon fewer sites, suggesting there was no benefit in developing local models based on fewer sites. Developing models based upon a larger number of sites has been shown to result in more robust models in studies using longer term sampling as well.[15−17]

Model structures differed somewhat between pooled and city-specific models, largely due to correlation among different small-scale traffic predictors. While we identified a single best model, typically several models predict measured variance in concentrations almost equally well. A good example is the pooled UFP model in Rotterdam that explained virtually the same variance as the city-specific models using different correlated variables. It would be useful to account for the uncertainty in identifying models by applying multiple models in epidemiological analyses.

Our spatial UFP model was similar to the previously developed model from Amsterdam and the Vancouver model.[7,11] All models included a port related variable, probably indicating that ship emissions or (truck) traffic to and from the ports contributes to the spatial variability in UFP. Furthermore, the models from our and the previous two studies include small-scale traffic-related variables.

Advantages and Disadvantages of Short-Term and Mobile Monitoring. The BC model from our short-term monitoring campaign explained 61% of the variance of the absorbance measurements at the 40 ESCAPE measurement sites. Wang et al. reported a hold-out validation R^2 of 76% for the Dutch ESCAPE absorbance model.[15] The hold-out validation R^2 was obtained by developing models based upon 20 sites and evaluating them based upon the remaining 20 sites. This may indicate that the ESCAPE fixed monitoring provided better predictive models than our current short-term monitoring campaign. We cannot generalize this finding, as the monitoring domain of the current short-term and ESCAPE long-term fixed monitoring differed, and no completely independent data were available that could be used to compare predictions from both types of models.

The ***advantage*** of mobile sampling campaigns is that a large number of sites can be measured efficiently. In the current study, we obtain three repeated samples at 161 sites in 7 months' time. In contrast, the ESCAPE study period was performed in 12 months to measure during three campaigns at 40 sites in The Netherlands. Recent studies have documented the increase in model robustness with increasing number of sites.[15−17] Because a field technician is present during sampling in the short term campaigns, there are fewer restrictions with respect to selection of monitoring sites. Leaving valuable equipment unattended in fixed site sampling campaigns limits the locations where measurements can be performed. A larger number of sites allows inclusion of more complex but realistic locations, such as sites near intersections, which are often excluded in LUR monitoring campaigns.

A ***disadvantage*** of mobile sampling campaigns is that the shorter sampling times compared to fixed sampling campaigns, leads to a decrease in precision of the site-specific averages because of the larger impact of temporal variation. Another disadvantage of mobile campaigns is that continuous real-time sensors are required, which are not available for certain components (including elemental composition of PM) or may be less reliable than the filter based methods (for example $PM_{2.5}$). Given that technicians need to be present, it is more difficult to include nighttime and weekend periods. A study in Belgium suggested that LUR models for BC differed between weekends and weekdays and between daytime and nighttime.[13]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Definition of GIS predictors, correlation between GIS predictors, and different spatiotemporal models. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/es505791g.

■ AUTHOR INFORMATION

Corresponding Author

*Address: Institute for Risk Assessment Sciences (IRAS), P.O. Box 80178, 3508 TD Utrecht. Telephone: +31 30 253 94 98. Fax: +31 30 253 94 99. E-mail: g.hoek@uu.nl.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was funded by The Netherlands Royal Academy of Sciences. ESCAPE was funded by the European Community's Seventh Framework Program (grant agreement number 211250).

■ REFERENCES

(1) Oberdörster, G.; Oberdörster, E.; Oberdörster, J. Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles. *Environ. Health Perspect.* **2005**, *113*, 823−839.

(2) Utell, M. J.; Frampton, M. W. Acute health effects of ambient air pollution: The ultrafine particle hypothesis. *J. Aerosol Med.* **2000**, *13*, 355−359.

(3) HEI Review Panel. *Understanding the Health Effects of Ambient Ultrafine Particles Report*; Health Effects Institute: Boston, 2013.

DOI: 10.1021/es505791g
Environ. Sci. Technol. 2015, 49, 8712−8720

(4) Janssen, N. A.; Hoek, G.; Simic-Lawson, M.; et al. Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM10 and PM2.5. *Environ. Health Perspect.* **2011**, *119*, 1691−1699.
(5) Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561−7578.
(6) 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*, 12903−12911.
(7) 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*, 5217−5225.
(8) Rivera, M.; Basagaña, X.; Aguilera, I.; Agis, D.; Bouso, L.; Foraster, M.; Medina-Ramón, M.; Pey, J.; Künzli, N.; Hoek, G. Spatial distribution of ultrafine particles in urban settings: A land use regression model. *Atmos. Environ.* **2012**, *54*, 657−666.
(9) Zwack, L. M.; Paciorek, C. J.; Spengler, J. D.; Levy, J. I. Characterizing local traffic contributions to particulate air pollution in street canyons using mobile monitoring techniques. *Atmos. Environ.* **2011**, *45*, 2507−2514.
(10) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An hourly regression model for ultrafine particles in a near-highway urban area. *Environ. Sci. Technol.* **2014**, *48*, 3272−3280.
(11) 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*, 622−628.
(12) Larson, T.; Hernderson, S.; Brauer, M. Mobile monitoring of particle light absorption coefficient in an urban area as a basis for land use regression. *Environ. Sci. Technol.* **2009**, 4672−4678.
(13) Dons, E.; Van Poppel, M.; Kochan, B.; Wets, G.; Int. Panis, L. Modeling temporal and spatial variability of traffic-related air pollution: Hourly land use regression models for black carbon. *Atmos. Environ.* **2013**, *74*, 237−246.
(14) Eeftens, M.; Beelen, R.; de Hoogh, K.; et al. Development of land use regression models for PM(2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas: Results of the ESCAPE project. *Environ. Sci. Technol.* **2012**, *46*, 11195−11205.
(15) Wang, M.; Beelen, R.; Basagana, X.; et al. Evaluation of land use regression models for NO_2 and particulate matter in 20 European study areas: The ESCAPE project. *Environ. Sci. Technol.* **2013**, *47*, 4357−4364.
(16) Basagaña, X.; Rivera, M.; Aguilera, I.; et al. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmos. Environ.* **2012**, *54*, 634−642.
(17) Wang, M.; Beelen, R.; Eeftens, M.; Meliefste, K.; Hoek, G.; Brunekreef, B. Systematic evaluation of land use regression models for NO_2. *Environ. Sci. Technol.* **2012**, *46*, 4481−4489.
(18) Klompmaker, J. O.; Montagne, D. R.; Meliefste, K.; Hoek, G.; Brunekreef, B. Spatial variation of ultrafine particles and black carbon in two cities: Results from a mobile measurement campaign. *Sci. Total Environ.* **2015**, *508*, 266−75.
(19) Ragettli, M. S.; Ducret-Stich, R. E.; Foraster, M.; Morelli, X.; Aguilera, I.; Basagaña, X.; Corradi, E.; Ineichen, A.; Tsai, M.; Probst-Hensch, N.; Rivera, M.; Slama, R.; Künzli, N.; Phuleria, H. C. Spatio-temporal variation of urban ultrafine particle number concentrations. *Atmos. Environ.* **2014**, *96*, 275−283.
(20) Beelen, R.; Hoek, G.; Vienneau, D.; et al. Development of NO_2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe—The ESCAPE project. *Atmos. Environ.* **2013**, *72*, 10−23.
(21) Armstrong, B. G. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ. Med.* **1998**, *55*, 651−656.
(22) Kumar, P.; Morawska, L.; Birmili, W.; Paasonen, P.; Hu, M.; Kulmala, M.; Harrison, R. M.; Norford, L.; Britter, R. Ultrafine particles in cities. *Environ. Int.* **2014**, *66*, 1−10.