Applicability of a noise-based model to estimate in-traffic exposure to black carbon and particle number concentrations in different cultures

Luc Dekoninck [a,*], Dick Botteldooren [a], Luc Int Panis [b,c], Steve Hankey [d], Grishma Jain [d], Karthik S [d], Julian Marshall [d]

[a] Information Technology, Acoustics Group, Ghent University, Sint-Pietersnieuwsstraat 41, 9000 Ghent, Belgium
[b] Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium
[c] Transportation Research Institute (IMOB), Hasselt University, Wetenschapspark 5 Bus 6, 3590 Diepenbeek, Belgium
[d] Department of Civil Engineering, University of Minnesota, 500 Pillsbury Dr SE, Minneapolis, MN 55455, United States

ABSTRACT

Several studies show that a significant portion of daily air pollution exposure, in particular black carbon (BC), occurs during transport. In a previous work, a model for the in-traffic exposure of bicyclists to BC was proposed based on spectral evaluation of mobile noise measurements and validated with BC measurements in Ghent, Belgium. In this paper, applicability of this model in a different cultural context with a totally different traffic and mobility situation is presented. In addition, a similar modeling approach is tested for particle number (PN) concentration.

Indirectly assessing BC and PN exposure through a model based on noise measurements is advantageous because of the availability of very affordable noise monitoring devices. Our previous work showed that a model including specific spectral components of the noise that relate to engine and rolling emission and basic meteorological data, could be quite accurate. Moreover, including a background concentration adjustment improved the model considerably. To explore whether this model could also be used in a different context, with or without tuning of the model parameters, a study was conducted in Bangalore, India. Noise measurement equipment, data storage, data processing, continent, country, measurement operators, vehicle fleet, driving behavior, biking facilities, background concentration, and meteorology are all very different from the first measurement campaign in Belgium. More than 24 h of combined in-traffic noise, BC, and PN measurements were collected. It was shown that the noise-based BC exposure model gives good predictions in Bangalore and that the same approach is also successful for PN. Cross validation of the model parameters was used to compare factors that impact exposure across study sites. A pooled model (combining the measurements of the two locations) results in a correlation of 0.84 when fitting the total trip exposure in Bangalore. Estimating particulate matter exposure with traffic noise measurements was thus shown to be a valid approach across countries and cultures.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Particulate matter (PM) is currently regulated in Europe, the US, India and other countries based on specific particle size fractions (e.g., PM_{10}, $PM_{2.5}$). Black carbon (BC) and particle number (pN) concentrations are associated with transportation emissions but are typically unregulated. The World Health Organization suggests including BC when evaluating traffic-related health effects (WHO Europe, 2012). Recent epidemiological results for BC suggest that health effects per mass may be up to 10 times higher than PM_{10} (Janssen et al., 2011). Research into the health effects of traffic-related particulates is constrained by the stronger spatial variability for BC and PN concentrations relative to PM_{10} and $PM_{2.5}$. Detailed measurements for near-road settings have shown large spatial gradients for certain aspects of particulate air pollution. For example, ultrafine particles and BC show decreases of over 50% within the first 150 m from the edge of the road (Karner et al., 2010) and significant street-to-street differences in PN and BC have been reported by several authors (Boogaard et al., 2011; Dons et al., 2012, 2013). Building a fixed-site monitoring network for PN and BC to provide robust estimates of exposure patterns, would therefore be a daunting task.

In a previous work a novel way to predict a bicyclist's in-traffic BC exposure was presented based on mobile measurements of traffic-related noise and BC in Ghent (Belgium) (Dekoninck et al., 2013). The noise-based model yields spatially and temporally precise estimates of

* Corresponding author. Tel.: +32 9 264 99 95.
E-mail addresses: luc.dekoninck@intec.ugent.be (L. Dekoninck), dick.botteldooren@intec.ugent.be (D. Botteldooren), luc.intpanis@vito.be, luc.intpanis@uhasselt.be (L.I. Panis), shankey1028@gmail.com (S. Hankey), julian@umn.edu (J. Marshall).

BC based on traffic noise measurements. The additive model successfully split the BC exposure of a bicyclist into a background component and a local, traffic-related component. The local component included the wind speed, a street canyon index and a noise-based characterization of instantaneous traffic volume and dynamics; specifically, an engine throttle noise component and a rolling noise component.

Here we extend our prior work by exploring the same category of modeling (correlating real-time measurement of noise with real-time pollution measurements) in Bangalore, India. Relevant environmental differences between Bangalore and Ghent include vehicle fleet and fuel use, driving patterns, speeds, densities, and behavior; levels of ambient and traffic-related noise and pollution, meteorological conditions, and limited biking facilities. Three important differences between the current investigation (Bangalore) and the prior investigation (Ghent) are: (1) a different location, with very different traffic conditions relative to the prior study, (2) measurement procedures were automated to allow use by bicyclists unfamiliar with the equipment, and (3) we study two pollutants (PN and BC) rather than only BC.

2. Methodology

2.1. Measurement equipment and setup

The original experimental setup for the measurement campaign in Ghent (Dekoninck et al., 2013) was based on a manually operated smartphone GPS, a Type 1 Noise Level Meter (Svantek 959) and a micro-aethalometer. The data was processed and merged manually. These labor intensive procedures needed to be automated to enable large scale, low support measurement campaigns. A low-cost noise measurement setup designed by the acoustics group at the Ghent University was modified to enable automated mobile measurements at a much lower hardware cost (Can et al., 2011a,b; Dauwe et al., 2012; Van Renterghem et al., 2011). The noise measurement module is extended with a GPS (Haicom HI-204 III USB), a micro-aethalometer (AE51, Aethlabs, San Francisco, CA) and a battery for off-grid operation (Fig. 1). Software was developed to automatically capture the 1-second data stream from the micro-aethalometer. The mobile node is designed to be mounted on the handlebar of a bicycle. In addition, a condensation particle counter (CPC 3007, TSI Inc., Shoreview, MN) was carried in a backpack by the bicyclist to measure PN. After each sampling run the mobile node was connected to the internet and the data was uploaded to a database on the server at the University of Ghent. An automated process synchronizes and merges the noise, GPS and BC data based on the timestamp of the mobile node. The CPC data was joined with the other data in a separate post-processing sequence.

2.2. Measurement location, strategy and processing

Measurements were carried out north of the city center of Bangalore, India, (elevation: 920 m; metropolitan population: ~10
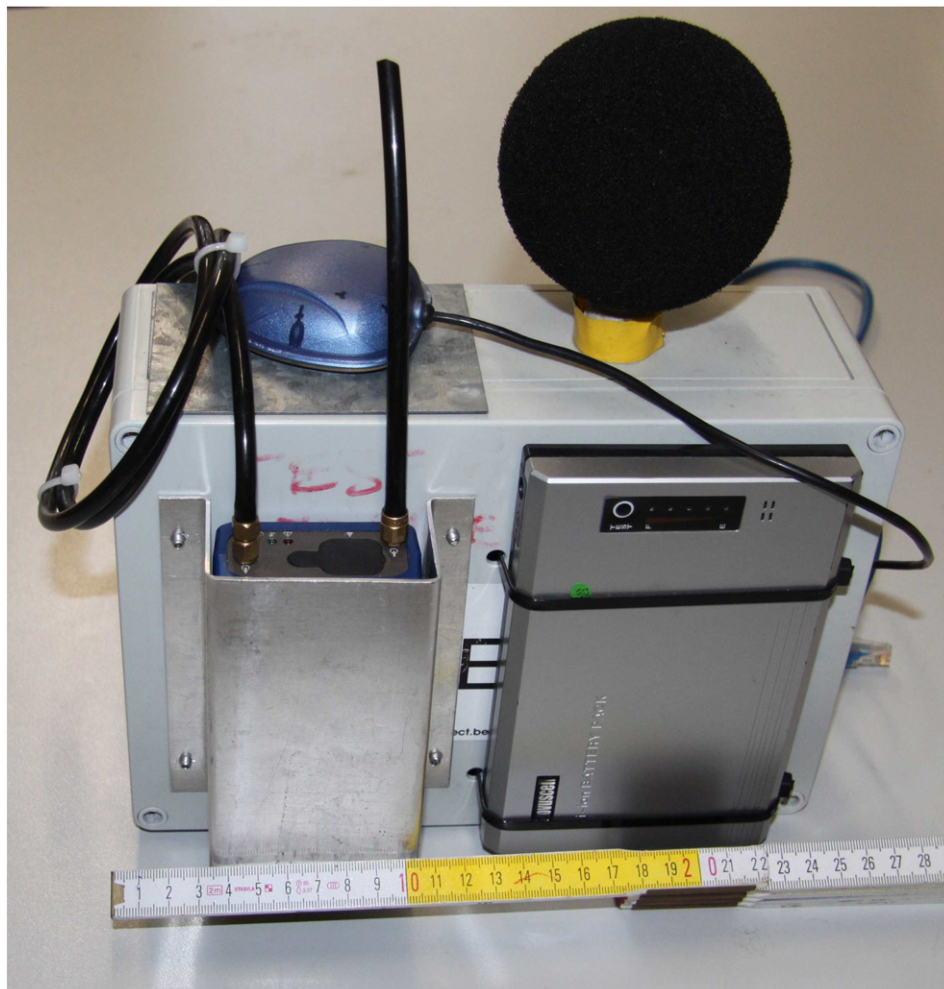

Fig. 1. Inexpensive mobile noise measurement equipment including the micro-aethalometer and GPS. The box containing all instruments is 23 cm by 9 cm wide and 17 cm high.

million; land area: 1276 km^2). The mobile measurements were performed while cycling a predefined route around the Armana Nagar neighborhood including roads with different traffic conditions instead of the random sampling used by Dekoninck et al. (2013). The route was selected to alternate on a regular basis between high and low traffic roads to assess small scale spatial and temporal differences in exposure. The sampling route was completed 15 times during the morning rush-hour and 5 times in the evening rush-hour during the sampling period (11 November 2013–15 January 2014; total duration: 24.5 h). The sampling route is shown in Fig. 2. The sampling runs also included stationary roadside measurement to explore the possibility of using stationary noise measurements. The results of the stationary measurements are not reported in this article. Some additional 'random' runs (not on the sampling route) were included to increase the coverage of small roads in several neighborhoods.

The noise measurements are performed in one-third octave band spectra, at a sampling rate of 8 measurements per second; these sound measurements are later aggregated to 1-second averages, to match the temporal resolution of the GPS, BC and PN measurements. The basic traffic-related noise parameters $L_{OLF}(t)$ and $L_{HFmLF}(t)$ as defined in Dekoninck et al. (2013) are then calculated. The parameter $L_{OLF}(t)$ is the sum of the (A-weighted) 100 to 200 Hz one-third octave bands and describes the engine noise of the nearby traffic at timestamp t. It increases with increasing engine throttle, increasing traffic volume, and increasing fraction of heavy vehicles. The second parameter $L_{HFmLF}(t)$ is related to the ratio of rolling noise and engine noise. The high frequencies in rolling noise are caused by tire–road interaction and are related to the size and pattern of the tire profile and the road surface grain size (see for example Sandberg and Ejsmont, 2002). $L_{HFmLF}(t)$ is defined as the difference between the high (1000 to 2000 Hz) and the low frequencies (100 to 200 Hz) in the noise spectrum

at timestamp t. High levels of $L_{HFmLF}(t)$ indicate that rolling noise dominates over engine noise, which indicates higher driving speed.

The noise, BC and PN measurements are smoothed by applying a running average of 20 s on the 1-second time series. The smoothed 1-second time series are then aggregated to the mean of each 10-second interval. The dataset with 10 second temporal resolution is used as input for the exposure models.

The noise–BC relation is evaluated on the uncorrected standard output of the aethalometer or CPC and reported on this basis in the detailed results, but several BC and PN correction functions to improve the quality of the data processing are available in the literature. Some measurement quality issues were related to bumps along the bicycle route as reported in the past by other authors (Apte et al., 2011; Cai et al., 2013). We checked the validity of the noise-BC relation for a selection of commonly used correction functions to test the sensitivity of our results. Two correction functions aim at cleaning the short term time series of BC by removing spikes and potentially erroneous aethalometer readings due to vibration errors (Apte et al., 2011; Hagler et al., 2011). Two particle loading correction functions for BC and one correction function for PN are applied as well (Kirchstetter and Novakov, 2007; Virkkula et al., 2007; Westerdahl et al., 2005). Each of the correction functions are applied to the raw output of the aethalometer or CPC. Where the correction function requires the use of the ATN value of the BC measurements it is applied as documented in the references.

2.3. Black carbon: background and local contribution

The approach used to model the local BC exposure in Dekoninck et al. (2013) was based on an adjustment for the background concentration at a fixed measurement site in the official air pollution network. Since no fixed BC background measurement stations were available at
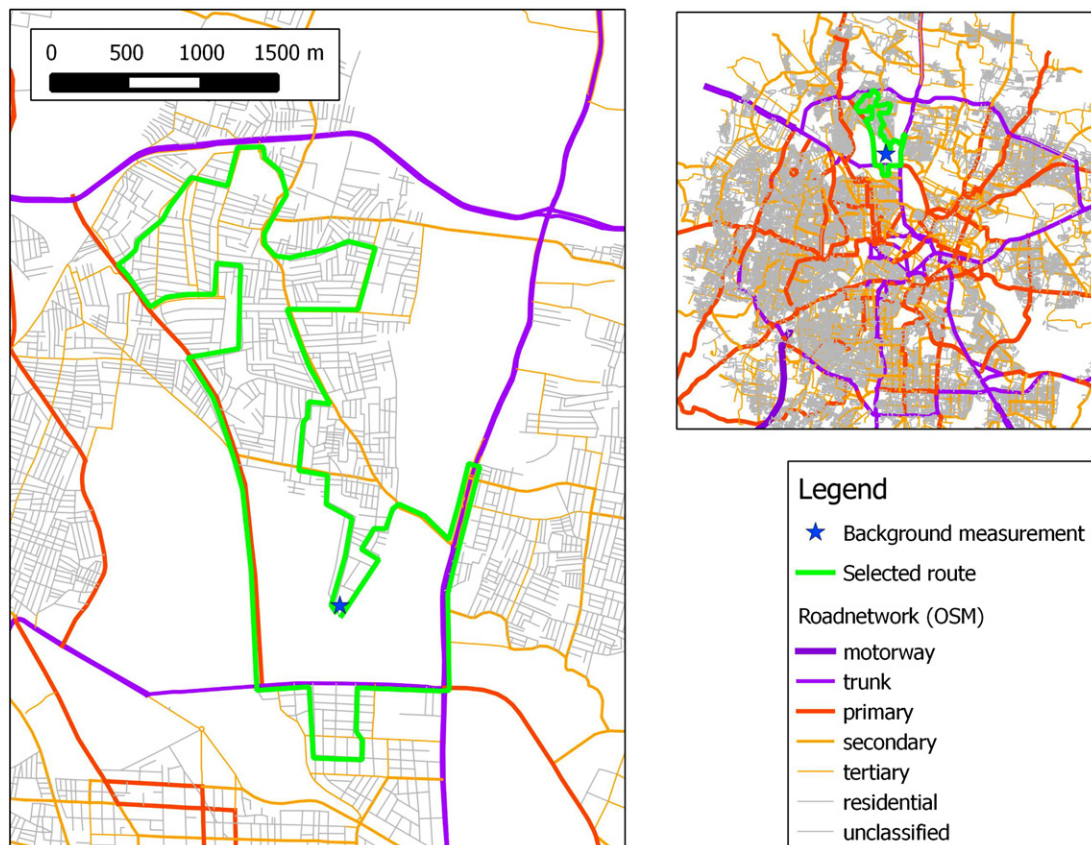

Fig. 2. Map of the sampling route in Bangalore. The blue star shows the location of the background measurement location (BG).

the time of the measurements in Bangalore, an alternative procedure was designed for the background adjustment. A second micro-aethalometer was set up in Bangalore near a dwelling in a low exposure area next to a park (background measurement location [BG] shown in Fig. 2). The BC concentration at the background location was measured simultaneously with the mobile concentrations at a 1 second temporal resolution. At the same site a semi-professional meteorological station was set up to acquire wind speed and temperature conditions. To mimic the resolution of the background concentrations available in the original (Belgium-based) model, the basic statistics of the BC at the background location BG are evaluated at a temporal resolution of 15 min. Instantaneous values of the in-traffic measurements can be lower than the 15 minute background evaluations. Due to the frequent occurrence of episodes with high background exposure concentrations in the Bangalore experiment, the background adjustment had to be restricted. A maximum adjustment is introduced in the procedure. The background adjustment for Bangalore is thus defined as:

$$BC_{loc,\,j}(t) = \max\left(BC_{raw,j}(t) - \min\left(BC_{bc,limit},\, BC_{bg,15min,Q1,j}\right),\, 500\right) \tag{1}$$

where $BC_{bg,15\,min,Q1,j}$ is the first quartile of background concentration of the 15 minute interval at the start of the bicycle trip j. The minimum level for the $BC_{loc}(t)$ component was set to 500 ng/m^3, a slightly higher value than in Belgium. Setting a minimum value is necessary to reduce the range of $\log(BC_{loc}(t))$ for episodes with very low exposure. The value $BC_{bg,limit}$ is chosen while evaluating the GAM model on $\log(BC_{loc}(t))$, ensuring a maximum positive effect on the prediction of the trip segments with low amounts of local traffic.

A similar approach was developed for PN. No simultaneous measurements at the background location were performed since only one CPC 3007 was available. Instead, to adjust for day-to-day variability in background concentrations, all instruments were co-located at the reference site for 30 min before and 30 min after each sampling run. This data was used to assess the background concentration of PN during the bicycle trip. To be consistent with the background adjustment for BC the average of the first quartile of the 15 minute periods before and after the measurement period is used to estimate the background concentration of PN. The PN_{loc} adjustment is defined as:

$$PN_{loc,\,j}(t) = \max(PN_{raw,j}(t) - \min\left(PN_{bc,limit},\, mean_j\left(PN_{bg,15min,Q1}\right),\, 500\right) \tag{2}$$

where $mean_j(PN_{bg,15min,Q1})$ is the mean of the first quartile of the 15 minute intervals at the background location before and after the trip j. Variants of the local components are calculated for the different correction functions listed in the previous section.

2.4. GAM modeling

Generalized additive models (GAMs) are regression models where smoothing splines are used instead of linear coefficients for the covariates. This approach has been found to be particularly effective for handling the complex non-linearity associated with air pollution research (Dekoninck et al. 2013; Dominici et al., 2002; Pearce et al., 2011). The mathematical form of GAM models is reduced to a time stamp of the measurement and a reference to the measurement node j and can be written in the form:

$$\log\left(BC(t)\right) = \sum_{z=1}^{n} s_z(v_z(t)) + \varepsilon(t) \tag{3}$$

where $v_z(t)$ is the zth covariate evaluated at timestamp t; $s_z(v_z(t))$ is the smooth function of zth covariate, n is the total number of covariates, and $\varepsilon(t)$ is the corresponding residual with $\mathrm{var}(\varepsilon) = \sigma^2$, which is assumed to be normally distributed. Smooth functions are developed through a combination of model selection and automatic smoothing parameter selection using penalized regression splines, which optimize the fit and try to minimize the number of dimensions in the model. The main advantage of GAM modeling is the possibility to adjust for non-linear relationships between the covariate and the outcome. The analysis was constructed using the GAM modeling function in the R environment for statistical computing (R development Core Team, 2009) with the package 'mgcv' (Wood, 2006). The covariates are restricted to L_{OLF}, L_{HFmLF}, wind speed and BC background.

3. Results

3.1. General statistics and background correction

Basic statistics of the BC and PN measurements are assembled in Table 1. The distributions are presented in Fig. 3A. The background correction limit $BC_{bg,limit}$ is set to 6000 ng/m^3 in this measurement campaign, close to the average background. Higher correction limits did not improve the BC_{loc} models. In contrast to BC, the PN concentrations at the background locations are small compared to the in-traffic PN concentration. The $PN_{bg,limit}$ could be set to 16,000 pt/cc, close to the maximum of the PN background concentration without disturbing the PN_{loc} model. This could express a physical difference between the PN and BC particle behavior at the background location, but also a difference in the measurement technique. Particles coagulate while the distance to the source increases. This strongly influences the particle count, but has less influence on mass related measurements.

Box plots of air pollution measurements and the model covariates are shown in Fig. 3A and B. The wind speed in Bangalore is very low and never exceeded 2 m/s (wind speeds during the sampling in Belgium averaged 3 m/s). The noise measurements in Bangalore are high with L_{Aeq} reaching values over 100 dBA and L_{OLF} reaching 90 dB compared to Ghent where these levels were 70 dBA and 65 dB respectively. The difference between L_{Aeq} and L_{OLF} is shown in Fig. 3C. Differences larger than 25 dB are not unusual in Bangalore and are related to the regular use of (high frequency) vehicle horns. Several reports are available on the frequent honking while driving in India (Banerjee et al., 2009; Sen et al., 2010). No data was removed to accommodate the potential disturbances of the models due to the honking. In Fig. 3D the $BC_{10s}(t)$ concentration is shown as a function of $L_{OLF,10s}(t)$ as an illustration of the physical relation between noise and particulate matter exposure.

3.2. Black carbon and particle number models

Models of BC and PN were developed using two types of dependent variables: measurements (1) unadjusted and (2) adjusted for background concentrations. The BC_{raw} model is based on measured BC concentrations $BC_{raw}(t)$ and investigates the relationship between $BC_{raw}(t)$ and the covariates L_{OLF}, L_{HFmLF}, wind speed and the logarithm of the background concentration $\log(BC_{bkg}(t))$. The BC_{loc} model is based on background adjusted BC exposure $BC_{loc}(t)$ as described in Eq. (1) and the same set of covariates. The BC_{loc} model investigates the potential to describe the actual BC exposure in an additive model similar to the approach in Dekoninck et al., 2013. The parameters of

Table 1
Descriptive statistics of the BC and PN (background, raw and local) datasets in ng/m^3 for BC, pt/cc for PN.

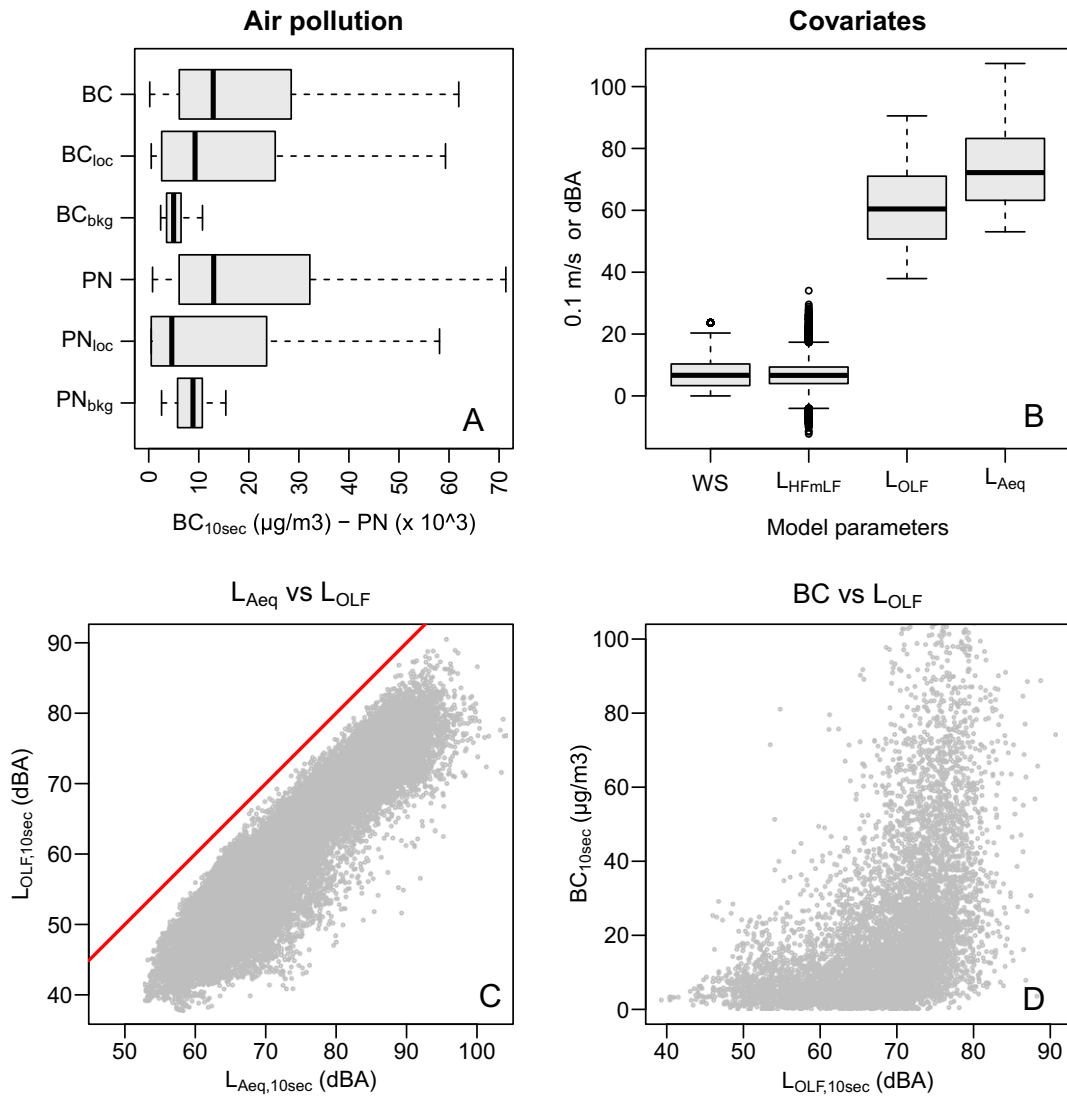|	P10	P25	Median	Mean	P75	P90
$BC_{bkg,15\,min,Q1}$	2993	3567	4966	6687	6457	10,983
$BC_{raw}(t)$	3347	6096	12,890	26,240	28,470	59,412
$BC_{loc}(t)$	500	2597	9284	22,840	25,260	54,956
$PN_{bkg,15\,min,Q1}$	5070	5765	8842	8854	10,710	13,302
$PN_{raw}(t)$	3786	5946	12,940	23,210	32,190	59,210
$PN_{loc}(t)$	500	500	4259	15,980	23,570	50,075

Fig. 3. Distributions of the air pollution measurements: BC and BC_{loc} while biking, BC at the background location, PN and PN_{loc} while biking and PN at the background (A), distributions of the model covariates, wind speed, L_{HFmLF}, L_{OLF} and L_{Aeq} (B), relation between $L_{Aeq,10s}$ and $L_{OLF,10s}$ (C), relation between BC_{10s} while biking and $L_{OLF,10s}$ (D). Boxes represent interquartile range; whiskers represent the 5th and 95th percentiles. Plot B shows outliers as dots.

the GAM models are shown in Table 2. The quality of a GAM model and the relative strength of its parameters are described by the deviance explained, the intercept, as well as the F-parameter and p-value for each of its covariates. Since the number of data points in the models is large compared to the number of covariates, the degrees of freedom is large and the p-values are in general too small to be used to compare the covariates. The F-parameters present the relative strength of the

covariates instead. In the BC_{raw} model the intercept and the L_{OLF} have similar strengths; the other covariates are almost ten times less important compared to L_{OLF}. The plots of the splines for both models show the relation of the parameter to the outcome $\log(BC_{loc})$ (Fig. 4, top and second row). In the BC_{loc} model, L_{OLF} is even stronger compared to the BC_{raw} model. The influence of the background adjustment is visible in the stronger linearity of the spline of the L_{OLF} covariate for smaller

Table 2
Results of the BC_{raw}, BC_{loc}, PN_{raw} and PN_{loc} exposure models.

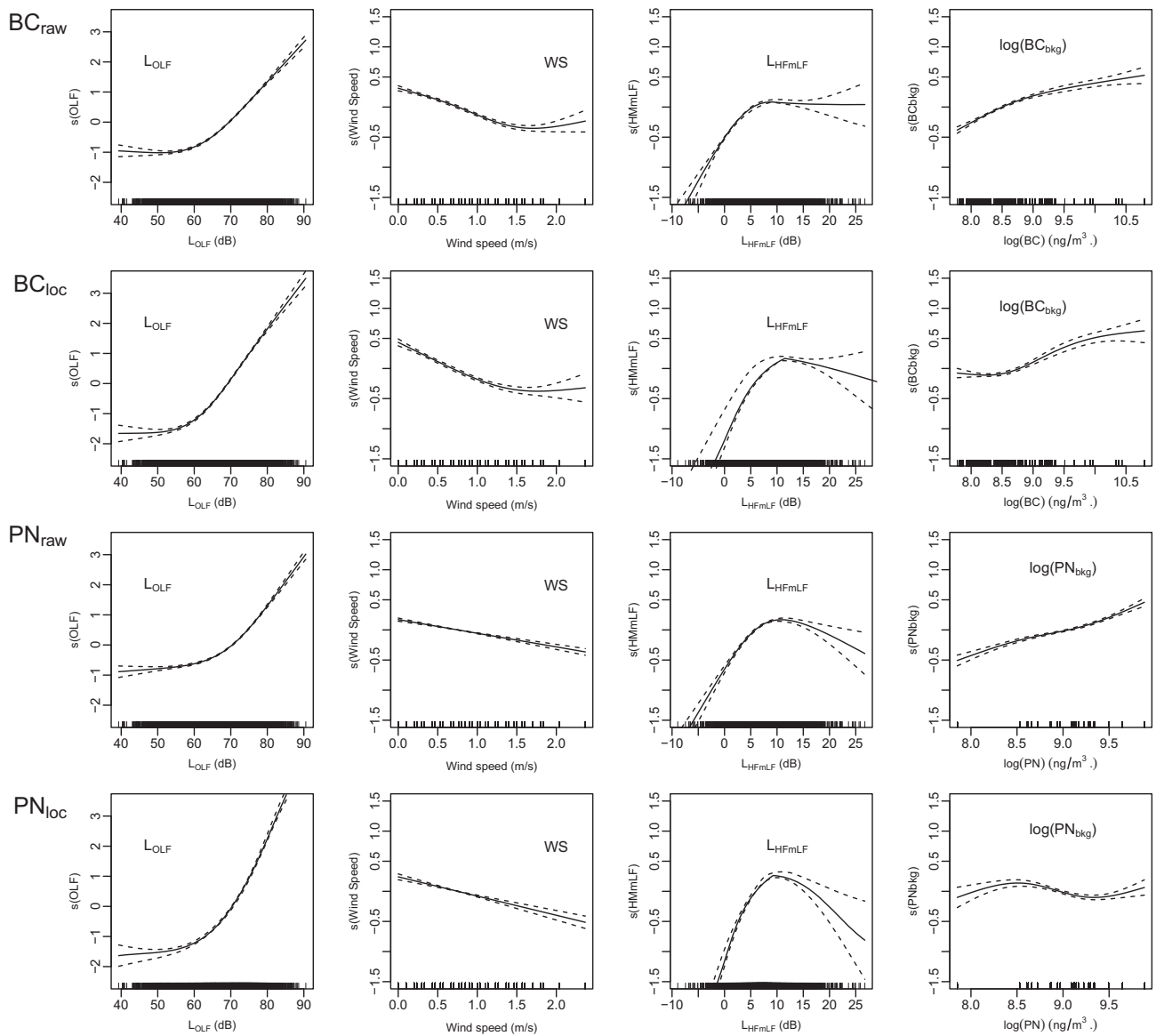|	BC_{raw}		BC_{loc}		PN_{raw}		PN_{loc}	
	F	p-Value	F	p-Value	F	p-Value	F	p-Value
Intercept	9.50		8.96		9.54		8.29	
	(13,359 ng/m³)		(8103 ng/m³)		(13,800 cm⁻³)		(3,984 cm⁻³)	
Intercept (t value)	1050		706		1079		516	
L_{OLF}	1692	<2e-16	1792	<2e-16	1421	<2e-16	1376	<2e-16
WS	160	<2e-16	121	<2e-16	176	<2e-16	100	<2e-16
L_{HFmLF}	116	<2e-16	118	<2e-16	187	<2e-16	165	<2e-16
$\log(BC_{bkg})$	136	<2e-16	51	<2e-16				
$\log(PN_{bkg})$					133	<2e-16	11	6.0e-7
Deviance explained	45.7%		43.6%		35.8%		33.6%	
Number of datapoints	8819		8 819		8 819		8 819	

Fig. 4. Splines of the four covariates for the BC_{raw}, BC_{loc}, PN_{raw} and PN_{loc} GAM models. Note the different scales used for L_{OLF} as compared to the other covariates since this covariate strength is an order of magnitude larger compared to the other covariates.

L_{OLF}. The relative strength of the wind speed and L_{HFmLF} covariate compared to $log(BC_{bkg})$ in the BC_{loc} model is higher than in the BC_{raw} model illustrating the effect of the background adjustment. Compared to Dekoninck et al. (2013), the strength of the wind speed covariate is low and this is likely due to the low variability of the wind speed in Bangalore during the measurement campaign. Note that the intercept of the BC_{raw} model in Bangalore is more than three times the value observed in Ghent and twice the value of the BC_{loc} model. The splines in the Bangalore model are similar to the model for Ghent.

We also developed a model for particle number concentrations to explore if PN behaves similarly to BC. The results are available in Table 2 and the splines are shown in the two bottom rows of Fig. 4. The PN_{raw} and PN_{loc} models behave similarly to the BC_{raw} and BC_{loc} models for all covariates. The intercept of the PN_{raw} model is 13.8×10^3 cm^{-3} and is reduced to 3900 for the PN_{loc} model due to the background adjustment. L_{OLF} is the strongest component in both models; L_{HFmLF} is the second strongest component. The importance of the background PN concentration and the wind speed is relatively low in both models and drops in importance in the PN_{loc} model compared to the PN_{raw} model and is more effective compared to the BC models.

The wind speed covariate is weaker in the PN models than the BC models. The strength of the L_{OLF} covariate increases in most of the local models, expressing an improved noise-BC/PN relation after background adjustment.

In Table 3 the results of the variants of the GAM models for the different correction functions are presented. The basic model using unadjusted BC measurements shows the highest F-value for the L_{OLF} covariate and the highest deviance explained. The noise-BC relation is consistent in all other models. In the background adjusted models the strength of L_{OLF} increases and the strength of $log(BC_{bkg})$ decreases; the deviance explained does not increase in these models due to loss of resolution in the dataset for low BC exposure. The alternative approach, removing the measurements below the instantaneous background does not improve the local models.

3.3. International comparison of the noise-based models

Upon completion of data collection in Bangalore it was possible to compare model results from Belgium and India. To make this comparison the total BC exposure is reconstructed by applying the prediction

Table 3
GAM model evaluation for the BC_{raw} and BC_{loc} models for different correction and loading functions, sorting based on $F(L_{OLF})$. The models based on the unadjusted data from Table 2 are added as a "base" case to allow for comparison. The abbreviated names relate to the first authors of the applied correction functions as listed in Section 2.2.

Model	Intercept	$F(L_{OLF})$	F(WS)	$F(L_{HFmLF})$	$F(\log(BC_{bkg}))$	Deviance explained	Count	AIC
BC raw models								
BC (base)	13,324	1692	159	115	135	45.7%	8819	22,204
BC (Virk)	19,456	1652	140	119	136	44.9%	8819	21,772
BC (Kirch)	20,936	1595	131	117	126	43.9%	8819	21,840
BC (Hagler)	13,039	1482	126	115	137	43.0%	8819	23,587
BC (Apte)	13,165	1456	150	78	143	46.0%	7834	19,499
BC (Hagler + Virk)	18,962	1385	105	114	133	41.1%	8819	23,615
BC (Apte + Kirch)	20,267	1371	129	79	145	44.3%	7834	19,269
BC (Hagler + Kirch)	20,388	1331	98	112	125	40.1%	8819	23,741
BC local models (background adjusted)								
BC_{loc} (base)	7862	1792	121	118	50	43.6%	8819	28,109
BC_{loc} (Hagler)	7907	1750	118	128	57	43.3%	8819	28,171
$BC_{loc\_}$(Virk)	13,684	1673	106	120	77	43.0%	8819	26,506
BC_{loc} (Kirch)	15,079	1605	102	119	75	42.0%	8819	26,417
$BC_{loc\_}$(Hagler + Virk)	13,655	1599	100	130	81	42.2%	8819	26,810
BC_{loc} (Apte)	7646	1589	137	81	53	44.5%	7834	24,624
BC_{loc} (Hagler + Kirch)	15,023	1529	95	128	78	41.2%	8819	26,773
BC_{loc} (Apte + Kirch)	14,459	1416	113	87	86	42.9%	7834	23,149
PN models (raw + background adjusted)								
PN (base)	13,800	1421	176	187	133	35.8%	8819	21,724
PNloc (base)	3984	1376	100	165	11	33.6%	8819	32,600
PN (Westerdahl)	14,045	1422	175	185	132	35.8%	8819	21,828
PNloc (Westerdahl)	4004	1401	100	164	11	33.6%	8819	32,714

function of the GAM local models on an identical sequence of L_{OLF} while keeping other covariates constant and adding a background adjustment for the matching location. Wind speed was set to 1 m/s, L_{HFmLF} to 5 dB (relatively low traffic speed) and the street canyon index (Belgium-based model only) is set to a value that represents an urban environment (0.7). As background exposure the median of the difference between BC_{raw} and BC_{loc} is used for each location. The results are plotted in Fig. 5A for Ghent and Bangalore; predictions outside the range of measured values are shown as dashed lines. The model in Bangalore lacks resolution for the small L_{OLF} values, resulting in an almost constant exposure below 65 dB. The BC concentrations in Bangalore are almost exponential for the higher values of L_{OLF}. A second observation is that for the range of 50 to 70 dB, BC concentrations in Ghent are higher compared to Bangalore. In addition, a third model was developed for pooled data from the two sampling campaigns (giving each dataset equal weight in the GAM model). The pooled data model could only include covariates that were available in both cities and is therefore restricted to L_{OLF}, wind speed and L_{HFmLF}. The authors are aware that the pooled data model implies similar noise-BC emission relations for both locations, despite the extremely different fleet composition. Higher air

pollution emission could relate to the noise emission of the vehicles and is assumed valid within the scope of this comparison. In Fig. 5B the resulting simulations are presented for the pooled model with the local matching BC background (3600 ng/m^3 for Bangalore and 1600 ng/m^3 for Ghent). The simulation is evidently higher for Bangalore compared to Ghent over the full range.

The second step to compare the Ghent and Bangalore models is to perform a cross validation by predicting the measurements in Ghent with the model based on the measurements in Bangalore and vice versa and predicting both measurement campaigns with the pooled data model. The predicted exposure is aggregated for each sampling trip and compared to the actual measurements for that trip (see Fig. 6). The measured vs. modeled prediction correlations are weaker when transferring models between cities; this result parallels the L_{OLF}-BC relationship for the two models shown in Fig. 5A. The slope and absolute prediction of the Ghent trips is lower due to the underestimates of the Bangalore model for L_{OLF} of 50 to 70 dB; the opposite effect is visible for the trips in Bangalore predicted with the Ghent model for L_{OLF} larger than 70 dB. The pooled data model is almost identical to the Ghent model for L_{OLF} below 70 and performs well on the Ghent trips.
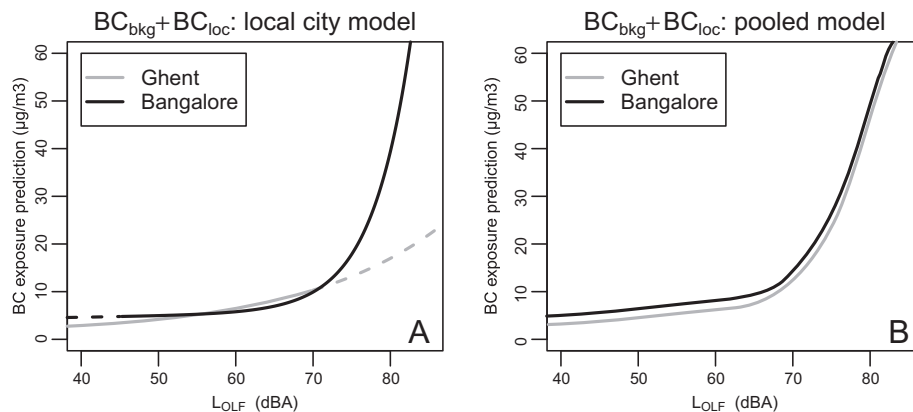

Fig. 5. Comparison of Ghent and Bangalore BC prediction for fixed wind speed and L_{HFmLF} values, dashed lines present extrapolated results outside the measured range of L_{OLF} (A) and the BC prediction by location based on the pooled data model with matching background adjustment by location (B).
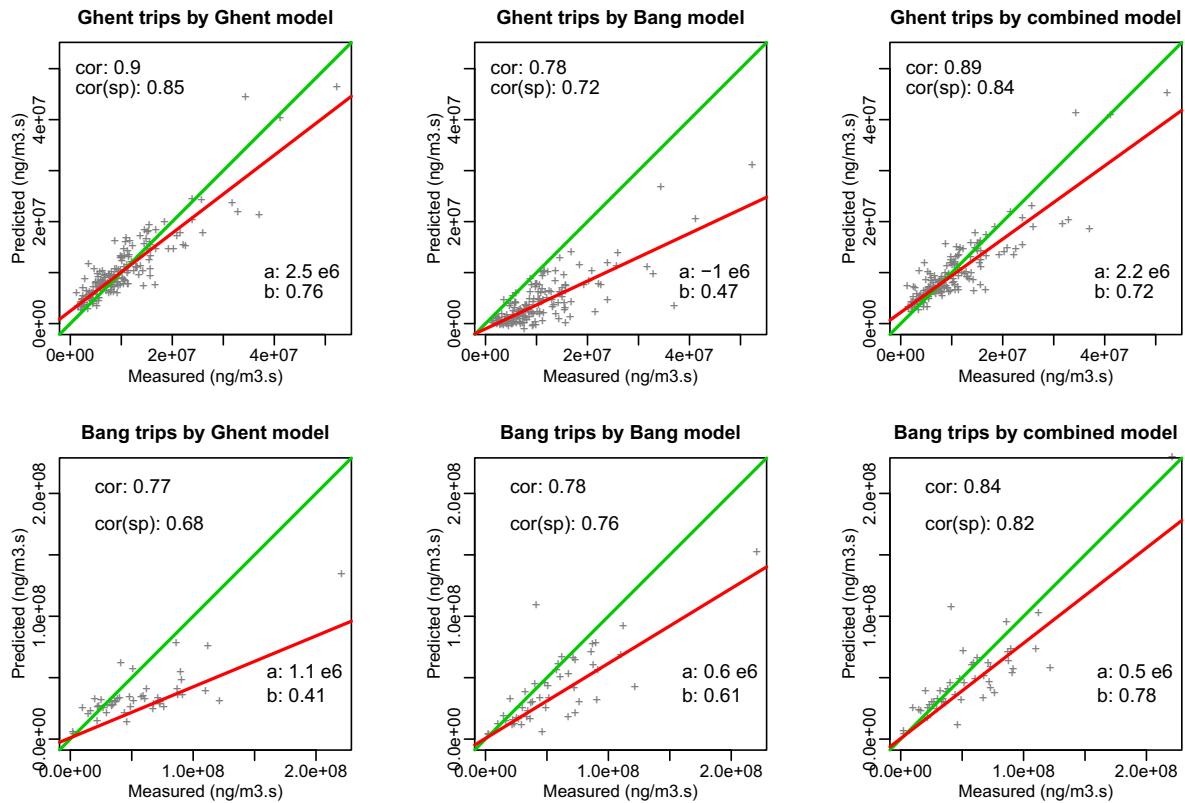
Fig. 6. Trip-based comparison and cross validation of the Ghent and Bangalore models. Each dot represents the total predicted exposure by trip compared to the measurement. The green line indicates the perfect fit; the red line presents the linear fit on the total trip exposure predictions. Correlations (Pearson and Spearman) and linear fit parameters of the trip prediction are shown for each plot.

The combined model improves the prediction of the Bangalore trips compared to the Bangalore model significantly due to the improvement of resolution in the combined model for L_{OLF} below 70 dB. The pooled data better specifies the L_{OLF} parameter and improves the prediction for the Bangalore data despite the fact that little data was collected in Bangalore below 70 dB. This results in correlation of 0.84 for the total trip exposure fit with the pooled model.

4. Discussion

A monitoring campaign was conducted for independently cross validating a previously designed model for predicting cyclist exposure to black carbon on the basis of noise measurements (Dekoninck et al., 2013) in a completely different socio-economic context. The results show that indeed the exposure to noise and black carbon is significantly higher in Bangalore, India, than in Ghent, Belgium. The bicyclist's exposure to BC in Bangalore exceeds the exposure in Belgium by a factor of four. The range of noise levels observed is at least 10 dBA higher. The observed between-city difference in the range of measured L_{OLF} could be due to a variety of factors. The first candidate is the fleet composition; the number of motorcycles and motorcycle-like vehicles is large compared to Belgium and the individual vehicle noise is high. There are few noise regulations and reducing the individual-vehicle noise is not considered a priority in India. The second candidate is the lack of bicycle facilities. During sampling the cyclists traveled among other vehicles (especially close to motorcycles). Since the concentrations of BC increase exponentially with smaller distance to the tail-pipe, cyclists in India are exposed to less-diluted exhaust emissions and comparatively high noise levels.

In addition to the context, the monitoring campaign conducted in Bangalore also differed slightly in design from previous measurements. An important technical difference is the use of consumer grade

microphones and data acquisition rather than high-quality IEC class 1 graded measurement equipment. As we stated before, using noise as a proxy for particulate air pollution is beneficial because noise measuring equipment is significantly cheaper. However this difference in cost is not high unless noise measurements can be done with consumer grade equipment. It has been shown previously that consumer electrets can be used for traffic noise monitoring over long time intervals (Van Renterghem et al., 2011) but it is also known that these devices may exhibit noise floors of 35 to 40 dBA. Comparing these values to the range of observations in Fig. 3 it becomes clear that they will cause no deviations in measurements of noise in streets. The new back-end database and internet based communication with the mobile measurement equipment showed that it is possible to seamlessly collect, manage and process measurements of noise and particulate matter simultaneously. The automated post-processing of the measurements resulted in quick reporting allowing evaluation of the progress of the measurement campaign on a daily basis. The monitoring campaign in Bangalore has shown the feasibility of internet-based measurement campaigns for noise and air pollution with low latency data storage and processing distributed across the globe (India, Belgium, USA).

Based on a one-year dataset of commuting noise and black carbon measurements, two parameters derived from the noise spectrum were found to be most suitable for explaining the 10 second aggregated BC exposure of cyclists: L_{OLF} and L_{HFmLF}. In addition, it was found that including wind speed measured at a nearby weather station and an indicator for the street geometry — expressing how strong it resembles a canyon — improved the performance of the proposed GAM model. In the current study, street geometry could not be included. Nevertheless, the goodness of fit of the GAM model extracted in this study confirms that the choice of (noise) parameters is valid also in a completely different socio-cultural context. With this fixed choice of noise parameters, the proposed prediction of in-traffic exposure to black carbon can thus

be used in a variety of international settings using a small dataset of combined noise and BC measurements to validate the model coefficients. Aethalometer-based black carbon measurements are sensitive to measurement error and therefore several corrections and loading functions have been proposed in literature (Apte et al., 2011; Hagler et al., 2011; Kirchstetter and Novakov, 2007; Virkkula et al., 2007; Westerdahl et al., 2005). It was found (Table 3) that the choice of correction and loading function has a limited influence on the model coefficients and hardly affects model performance (deviance explained and AIC). The model is robust against these corrections. A possible explanation for this is that the model already smoothed out most of the measurement errors that are also eliminated by the aforementioned corrections.

Based on the theoretical consideration that measured levels of BC have both a local origin, mainly caused by traffic, and a more regional origin including industrial, heating, and distant traffic, the measured BC concentration is split in a local and a background concentration. The rationale for treating a background contribution separately in the proposed models is that noise levels could be an indicator for the local traffic contribution to the overall level but are not expected to be indicative for the origin of background contributions. However, it has also been pointed out that a measured background concentration could also be a good indicator for dispersion of air pollutants in general since it strongly depends on weather conditions (Dominici et al, 2002). In the proposed models the former is included by subtracting the background from the raw measurement data, the latter is partially captured by a multiplicative factor included as a $\log(BC_{bkg})$ factor in the GAM model for $\log(BC_{loc})$. The factor $\log(BC_{bkg})$ in the GAM model for $\log(BC_{raw})$ is added to illustrate the effect of the additive approach in the models. Because of the strong non-linearity of the GAM, there is no strict separation between additive and multiplicative. Yet, the fact that the strength of the $\log(BC_{bkg})$ factor decreases once an additive correction for background has been applied indicates that the additive approach is valid. It results in a model explaining 45.6% of variance and could therefore be a valid alternative for the explicit background correction introduced in Section 2.3. An important issue however is how well the background measurements capture the background exposure variation due to long term meteorological conditions. Here, the background measurement location is inside the city of Bangalore and might be influenced by the diurnal pattern of in city traffic related air pollution exposure partially masking the long term background conditions. The quality and the properties of the BC background measurement location are important when employing an additive modeling approach. Still, this method for taking into account background concentrations could serve as a source of inspiration for mobile air pollution measurement processing in general.

A model using the same noise parameters was also established for particle number. The variance explained by this model is somewhat lower than for the BC model. PN concentration is also mainly related to traffic exhaust, but even more variable than BC, the lower variance explained is therefore expected. The background adjustment did not require strong restrictions through $PN_{bg,limit}$ to achieve a performant BC_{loc} model. This suggests a different spatial behavior of the relation between local and background exposures for PN compared to BC. BC background is probably more sensitive to local non-traffic related exposure compared to PN. Also physical differences between the two quantities could explain this difference. BC is mass based and does not distinguish between the sizes of the particles. PN is known to increase near tail-pipes and decreases strongly due to the coagulation processes at larger distances from the source, reducing the number of particles faster with distance from the source compared to mass based quantities. The meteorological situation is therefore likely to have less influence on the background PN compared to background BC. The noise–PN relationship can be stronger compared to noise–BC relationship.

When comparing the models for Ghent and Bangalore, it is clear that the Bangalore model lacks observations for L_{OLF} below 65 dB and overestimates BC below that level as compared to the model for Ghent (Fig. 5). Similarly, the dataset collected in Ghent lacks observations with L_{OLF} above 70 dB which results in underestimation of BC concentration at high noise levels in an extrapolated model. Extracting model parameters from the pooled data (Fig. 5B) results in a single smooth relationship which is indicative for a common ground truth in the relationship between L_{OLF} and $\log(BC)$. When cross-validating the average exposure to BC during bicycle trips by comparing model predictions to measurements some interesting conclusions can be drawn. As expected, the correlation between model and measurement decreases when model coefficients from the other areas are used. However, using the model based on pooled data, the correlation improves for the Bangalore trips (Fig. 6 bottom right) while it does not deteriorate for the Ghent data (Fig. 6 top right) both compared to the correlation with the model based on local data. This suggests a strong underlying relationship between noise and BC concentrations capable of distinguishing spatial variability beyond the disturbance of background and meteorological effects on the air pollution exposure measurements.

The authors suggest extending these measurement campaigns and using similar approaches for other traffic-related air pollutants in different countries. The quality of the background measurement site is an issue worth further investigation for future international deployment of this approach. Detecting and quantifying the contribution of the local traffic to the instantaneous exposure in environments where emissions from local traffic are low as compared to the ambient concentrations is crucial to improve the models. This can and should be achieved by several adjustments in the measurement setup. Extending the measurements to more variable meteorological conditions and extending the scope of the measurements to cover more spatial variability are two potential improvements. This technique attributes mobile measurements based on physical parameters, adding value to the current practice of manual qualification of traffic based on video footage. The simultaneous noise measurements add knowledge to the exposure measurements in an unprecedented temporal resolution enabling the disentanglement of the variation of the in-traffic exposure into a local traffic related component, the meteorological influences and background exposure. It also enables the use of random sampling in participatory measurement campaigns since the traffic is instantaneously documented in space and time. As illustrated in Table 3, the technique can be used in the future to compare the BC correction functions, potentially distinguishing between short term variation related to actual physical changes in the instantaneous exposure and variation related to instrumentation errors.

5. Conclusions

Exposure to traffic-related air pollution is strongly related to local traffic dynamics which can be characterized by spectral noise measurements. The physical relationship between noise and particulate air pollution was confirmed using an inexpensive noise measurement node in an entirely different socio-cultural setting (Bangalore, India) than previous work in Belgium with high-quality noise measurement equipment. The consumer grade noise measurement setup results in mobile noise measurements that are accurate enough to predict air pollution exposure. The full technical setup consisting of inexpensive measurement nodes combined with a back-end database and data post-processer enabled fast and accurate in-traffic sampling and reporting. Despite the significant differences between the local traffic conditions in Ghent, Belgium and Bangalore, India the mobile BC models behave similar. Moreover, trip exposure estimates based on a model fitted on pooled data from both monitoring campaigns correlate better with measurements. Sampling strategy has a strong impact on the properties of the models, low exposure route sampling is as important in the measurement campaigns as achieving proper quantification along the high exposure roads. Splitting the measurements in an instantaneous background concentration and a local contribution gave similar results as

in previous work; however, high background concentrations can reduce the spatial resolution of the method in low exposure streets. The characteristics of the background measurement location have a strong impact on the additive modeling approach and should be carefully considered in future projects. As an alternative, introducing background concentration as a log-additive term in the GAM model also increased the explained variance in a way that is partly complementary to separating background concentrations in an additive way.

Implementing this technique in places where only a few parameters change has the potential to disaggregate the effects of factors influencing in-traffic exposure (e.g., vehicle fleet, biking facilities, driving behavior and meteorology). For example, in the case of Belgium with its high number of diesel vehicles, a comparison with a similar European city but with a different fleet mix could allow for the quantification of the impact of diesel fuel related policy.

An identical model with the same noise covariates is valid as well for the mobile PN measurements, extending the potential of the methodology. The first results on PN for the additive modeling approach suggest that PN is less sensitive to properties of the background measurement location compared to BC. The international validation of this method for two important traffic-related pollutants (BC and PN) demonstrates the potential to evaluate within-city spatial patterns of particulate air pollution based on noise measurements in a variety of urban settings.

References

Apte JS, Kirchstetter TW, Reich AH, Deshpande SJ, Kaushik G, Chel A, et al. Concentrations of fine, ultrafine, and black carbon particles in auto-rickshaws in New Delhi, India. Atmos Environ 2011;45:4470–80.
Banerjee D, Chakraborty SK, Bhattacharyya S, Gangopadhyay A, et al. Appraisal and mapping the spatial–temporal distribution of urban road traffic noise. Int J Environ Sci Technol 2009;6(2):325–35.
Boogaard H, Kos GPA, Weijers EP, Janssen NAH, Fischer PH, van der Zee S, et al. Contrast in air pollution components between major streets and background locations: particulate matter mass, black carbon, elemental composition, nitrogen oxide and ultrafine particle number. Atmos Environ 2011;45:650–8.
Cai J, Yan BZ, Kinney PL, Perzanowski MS, Jung KH, Li TT, et al. Optimization approaches to ameliorate humidity and vibration related issues using the MicroAeth black carbon monitor for personal exposure measurement. Aerosol Sci Technol 2013;47:1196–204.
Can A, Dekoninck L, Rademaker M, Van Renterghem T, De Baets B, Botteldooren D. Noise measurements as proxies for traffic parameters in monitoring networks. Sci Total Environ 2011a;410:198–204.
Can A, Van Renterghem T, Rademaker M, Dauwe S, Thomas P, De Baets B, et al. Sampling approaches to predict urban street noise levels using fixed and temporary microphones. J Environ Monit 2011b;13:2710–9.
Dauwe S, Van Renterghem T, Botteldooren D, Dhoedt B. Multiagent-based data fusion in environmental monitoring networks. Int J Distrib Sens Netw 2012;2012(2012). (http://dx.doi.org/10.1155/2012/324935).
Dekoninck L, Botteldooren D, Int Panis L. An instantaneous spatiotemporal model to predict a bicyclist's black carbon exposure based on mobile noise measurements. Atmos Environ 2013;79:623–31.
Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol 2002;156:193–203.
Dons E, Int Panis L, Van Poppel M, Theunis J, Wets G. Personal exposure to black carbon in transport microenvironments. Atmos Environ 2012;55:392–8.
Dons E, Temmerman P, Van Poppel M, Bellemans T, Wets G, Int Panis L. Street characteristics and traffic factors determining road users' exposure to black carbon. Sci Total Environ 2013;447C:72–9. http://dx.doi.org/10.1016/j.scitotenv.2012.12.076.
Hagler GSW, Yelverton TLB, Vedantham R, Hansen ADA, Turner JR. Post-processing method to reduce noise while preserving high time resolution in aethalometer real-time black carbon data. Aerosol Air Qual Res 2011;11:539–46.
Janssen NAH, Hoek G, Simic-Lawson M, Fischer P, van Bree L, ten Brink H, 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;119:1691–9.
Karner AA, Eisinger DS, Niemeier DA. Near-roadway air quality: synthesizing the findings from real-world data. Environ Sci Technol 2010;44:5334–44.
Kirchstetter TW, Novakov T. Controlled generation of black carbon particles from a diffusion flame and applications in evaluating black carbon measurement methods. Atmos Environ 2007;41:1874–88.
Pearce JL, Beringer J, Nicholls N, Hyndman RJ, Tapper NJ. Quantifying the influence of local meteorology on air quality using generalized additive models. Atmos Environ 2011;45:1328–36.
R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2009.
Sandberg, Ejsmont. Tyre/road noise reference book. INFORMEX; 2002.
Sen R, et al. Horn-ok-please. MobiSys '10 proceedings of the 8th International Conference on Mobile Systems, Applications, and Services. NY, USA: ACM New York 978-1-60558-985-5; 2010. p. 137–50.
Van Renterghem T, Thomas P, Dominguez F, Dauwe S, Touhafi A, Dhoedt B, et al. On the ability of consumer electronics microphones for environmental noise monitoring. J Environ Monit 2011;13:544–52.
Virkkula A, Makela T, Hillamo R, Yli-Tuomi T, Hirsikko A, Hameri K, et al. A simple procedure for correcting loading effects of aethalometer data. J Air Waste Manage Assoc 2007;57:1214–22.
Westerdahl D, Fruin S, Sax T, Fine PM, Sioutas C. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. Atmos Environ 2005;39:3597–610.
WHO Europe. Health effects of black carbon 978 92 890 0265 3; 2012.
Wood SN. On confidence intervals for generalized additive models based on penalized regression splines. Aust N Z J Stat 2006;48.