An instantaneous spatiotemporal model to predict a bicyclist's Black Carbon exposure based on mobile noise measurements.

Non Peer-reviewed author version

Dekoninck, Luc; Botteldooren, Dick & INT PANIS, Luc (2013) An instantaneous spatiotemporal model to predict a bicyclist's Black Carbon exposure based on mobile noise measurements.. In: ATMOSPHERIC ENVIRONMENT, 79, p. 623-631.

DOI: 10.1016/j.atmosenv.2013.06.054
Handle: http://hdl.handle.net/1942/15937

An instantaneous spatiotemporal model to predict a bicyclist's Black Carbon exposure based on mobile noise measurements.

Luc Dekoninck[a)]
Dick Botteldooren[b)]
Information Technology, Acoustics Group, Ghent University
St-Pietersnieuwstraat 41, 9000 Ghent, Belgium[a,b]

Luc Int Panis[c,d)]
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]

ABSTRACT

Several studies have shown that a significant amount of daily air pollution exposure, in particular black carbon (BC), is inhaled during trips. Assessing this contribution to exposure remains difficult because on the one hand local air pollution maps lack spatio-temporal resolution, at the other hand direct measurement of particulate matter concentration remains expensive. This paper proposes to use in-traffic noise measurements in combination with geographical and meteorological information for predicting BC exposure during commuting trips. Mobile noise measurements are cheaper and easier to perform than mobile air pollution measurements and can easily be used in participatory sensing campaigns.

The uniqueness of the proposed model lies in the choice of noise indicators that goes beyond the traditional overall A-weighted noise level used in previous work. Noise and BC exposures are both related to the traffic intensity but also to traffic speed and traffic dynamics. Inspired by theoretical knowledge on the emission of noise and BC, the low frequency engine related noise and the difference between high frequency and low frequency noise that indicates the traffic speed, are introduced in the model. In addition, it is shown that splitting BC in a local and a background component significantly improves the model. The coefficients of the proposed model are extracted from 200 commuter bicycle trips. The predicted average exposure over a single trip correlates with measurements with a Pearson coefficient of 0.78 using only four parameters: the low frequency noise level, wind speed, the difference between high and low frequency noise and a street canyon index expressing local air pollution dispersion properties.

Highlights

- We performed combined black carbon and traffic noise measurements by bicycle.
- After correcting for the black carbon background concentrations a successful model is build.
- Personal noise measurements can be used as a proxy for black carbon exposure for bicyclists.

Keywords: Black Carbon, Vehicle Noise, Personal exposure, Bicyclists, Traffic

[a)] email: luc.dekoninck@intec.ugent.be (corresponding author, tel +32 9 264 99 95)
[b)] email: dick.botteldooren@intec.ugent.be
[c)] email: luc.intpanis@vito.be; luc.intpanis@uhasselt.be

1. INTRODUCTION

Exposure to particulate matter is currently regulated in PM standards, that only distinguish between the size of the particles (PM10, PM2.5, …). The soot fraction, Black Carbon (BC) is the part of the PM directly related to combustion processes. Recent evidence, summarized by the world health organization, documents the relevance of BC for evaluating traffic related health effects (WHO Europe, 2012). The first epidemiological results suggest health effects that are up to 10 times higher than a similar evaluation based on PM10 (Janssen et al., 2011). Further research into health effects is hampered by the difficulty to measure or model BC concentrations. An important reason for this is the strong spatial variability of BC compared to PM10 (Karner et al., 2010). Building a monitoring network for BC would be a daunting task because of the large spatio-temporal gradients. In addition, efforts are now made to standardize BC measurements as a first step to including BC in the set of official air pollution standards. For these reasons, this paper takes a closer look at an innovative way to predict a bicyclist's Black Carbon exposure.

Large personal exposure measurement campaigns prove the relevance of the in-traffic exposure contribution (Dons et al., 2011, 2012). Technology for mobile air pollution measurements is however scarce and expensive. On the one hand, cheap implementations do not meet quality requirements, on the other hand high quality equipment is often highly demanding on the operator (e.g. changing filters or liquids, limited portability). In contrast, mobile noise measurements can be done with low intrusive measurement equipment like dosimeters and new mobile technologies. Mobile noise measurements are a popular theme in noise exposure modeling (Eisenman et al., 2009; Kanjo et al., 2010, Maissonneuve et al., 2009). Noise levels are strongly correlated with traffic related air pollution levels and might be a good proxy to model personal air pollution exposure (Dekoninck et al., 2012; Can et al., 2010, 2011; Eisenman et al. 2009). Since in-traffic personal air pollution exposure is a major component of the total personal diurnal exposure and diurnal activity patterns are very diverse within the population, epidemiology would benefit from including exposure differences due to different activity patterns when investigating the health effects of air pollution (von Klot et al., 2011; Dons et al., 2012). The use of a proxy which is easy to monitor could result in improved personal exposure estimates on larger population samples at a reasonable cost.

Numerous efforts have been made to quantify the exposure and health effects of cycling in dense traffic since the current trends in sustainable mobility focus on establishing modal shifts towards biking and

walking (de Nazelle et al., 2011; Int Panis et al., 2010; Berghmans et al., 2009; de Hartog et al., 2010). Exposure of cyclists is directly influenced by the distance to the local traffic, strongly related to local traffic conditions (congestion, traffic lights etc.) and highly influenced by meteorological conditions. Since cyclists often travel along low-traffic-density roads however, there is in general no traffic data available from either counting loops or traffic models. The most important exposure parameter is therefore unknown in most studies. Local traffic on low density roads is also highly variable in both space and time. A suitable traffic description should reflect these short-term effects with an adequate spatial and temporal resolution. Mobile noise exposure therefore has the potential to become this key indicator to predict the local component of traffic related air pollution exposure.

Based on theoretical aspects of traffic dynamics, the relationship between noise and particulate matter emission and the potential to extract one from the other have been discussed earlier (Dekoninck et al. 2012). The focus there was on the selection of noise descriptors correlating best with vehicular particulate matter emissions for typical traffic dynamics. However, noise exposure close to the source is not strongly influenced by the meteorological conditions while air pollution is strongly affected by the meteorological conditions. This is one of the reasons why the average of repeated noise exposure measurements will converge faster compared to repeated air pollution exposure measurements. So, if mobile noise measurements are proven to be a valid proxy for air pollution exposure, fewer measurements will be needed to predict the personal exposure at a higher spatial and temporal resolution.

To establish an instantaneous relationship between noise exposure and air pollution exposure meteorological conditions have to be taken into account. A major concern is the influence of long term meteorology and long distance air pollution transport on the background concentrations influencing the actual personal exposure. In this paper two major research questions are addressed: (1) is a prediction model for BC exposure improved by separating out long term variations in the background concentration; (2) can the instantaneous local Black Carbon exposure be predicted based on instantaneous local noise exposure and meteorological conditions. For the latter, the question how to derive the local traffic dynamics that influence the instantaneous Black Carbon exposure from noise measurements is addressed. Section 2 will address the methodology including the definition of the models and the noise exposure parameters. Section 3 gives the results of the models and the model validations. Results are discussed in section 4.

2. Methodology

2.1 Measurement equipment and setup

The experimental setup contains a basic GPS (in a HTC Desire smart phone), a Type 1 Noise Level Meter (Svantek 959) and a micro-aethalometer (Model AE51 MageeScientific, 2009) to measure Black Carbon. The measurements were performed while commuting by bicycle from the villages to the west of Ghent (Belgium) into the city center, thus covering the sparsely build areas in the villages, the city center, open recreational areas and natural reserves in between. A total of 209 biking trips were performed, covering a distance of 2300 kilometer, a total measurement time of 128 hours at an average speed of 18 km per hour. More than 75 km of distinct roads were sampled at least 3 times. Almost all measurements were made between 7:30h-9:30h and 16:30h-18:30h. Some of the longer trips, for example when sampling in the vicinity of the highway, were partially made outside these time windows.

The details of the measurement setup, temporal resolution, preprocessing, meteorological data and the spatial mapping on aggregation points p_x along the network with a spatial resolution of 50 m are available in the supplementary data. In the instantaneous model only one spatial attribute, the street canyon index $StCan_{p_x}$ at aggregation point p_x is included, identifying 'street canyon likeliness' of the built-up area along the trip trajectories. More detail on the calculation of $StCan_{p_x}$ is available in the supplementary data.

2.2 Black carbon, background and local contribution

The BC exposure during a cycling trip consists of a contribution of local sources and a background contribution. The latter varies only little over a large area and can thus easily be obtained from a well-located fixed measurement station. The background contribution strongly depends on long-term meteorological conditions. The proposed model assumes that the dominant source of BC in the vicinity of the cyclist is the local traffic on the travelled road. An additive approach is used; the BC exposure is viewed as the sum of the background level $BC_{j,bg}$ during trip j and the "local" contribution BC_{loc}. Similar procedures are used in exposure estimations where regional and local scale models are added to estimate personal exposure (Isakov et al., 2009). The available background measurements are averaged concentrations over half an hour (see supplementary data). Subtracting a fixed measurement with a temporal resolution of a half hour from mobile BC measurements sampled at 1 second time interval is not trivial. When sampling air pollution at a shorter

time interval, concentrations below the "background" concentrations can be measured at some of the low traffic locations. Adjusting for the background concentration could then result in negative local concentrations. The proposed models will use a logarithm of BC as an outcome variable because noise is also measured on a logarithmic scale and hence negative values cannot be allowed. For this reason the background concentration is not removed completely but replaced by a typical but constant low background concentration. The choice of this constant is not very critical since it will be added again to the model outcome and constants do not affect the model. This approach enables the dataset to retain spatial variation also for low density roads. Nevertheless a physical reference to the ambient concentration is useful, therefore the long term first quartile concentration over the whole measurement period, $Q1(BC_{bg,lt})$, equal to 775 ng/m^3, is used. This value does not depend on the route choice of the sampled trip which is an additional benefit. The relation between the measured BC_{raw} and the local contribution BC_{loc} for a location *i* and during a trip *j*, can be written as:

$$BC_{raw,i,j} = BC_{loc,i,j} + (BC_{bg,j} - Q1(BC_{bg,lt})) \quad (1)$$

Where $BC_{bg,j}$ is the background concentration obtained from the continuous measuring station averaged over the whole duration of the trip j. At each aggregation points p_x on the network, the arithmetic average of all *n* measurements of trip j in this collection is calculated as:

$$BC_{loc,p_x,j} = \frac{1}{n}\sum_{\in P_{px,j}} BC_{loc,i,j} \quad (2)$$

2.3 Noise parameters and physical interpretation

Details of noise parameter calculations are available in the supplementary data. For an even more detailed description of the theoretical and empirical relations of the noise parameters with the BC exposure the reader is referred to Dekoninck et al, 2012. The main arguments for selecting particular noise indicators to be included in the model, based on their physical and technical interpretations, are briefly reminded to the reader. Three noise parameters were included in the instantaneous modeling. The harmonized calculation method used for noise map calculations for the European Union (END Directive) separates the noise emission into an engine contribution and a rolling noise contribution. Engine noise is dominant in the low frequencies; rolling noise is dominant in the high frequencies at higher speeds. The two first parameters are directly related to these emission features. The first parameter $L_{OLF,p_x,j}$ (100 – 200 Hz) describes the engine noise of the nearby traffic at point p_x for trip j. High throttle increases the engine noise. $L_{HLF,p_x,j}$ (1000 – 2000 Hz) is

related to the rolling noise. The second parameter $L_{HFmLF,p_x,j}$ is the difference between high and low frequencies in the noise spectrum at point p_x for trip j. High levels of $L_{HFmLF,p_x,j}$ indicate a relatively stronger contribution of high frequencies, compared to low frequencies, indicating more rolling noise than engine noise due to the nearby traffic, hence traffic at higher speed. The third parameter $(L_{Aeq} - L_{Amin})_{p_x,j}$ referred to as the short term dynamics of the noise at point p_x for trip j, indicates the presence of noise events. If noise levels $L_{Aeq,100\ ms}$ within a single second change rapidly, the passing vehicles can individually be detected in the noise measurements. If $(L_{Aeq} - L_{Amin})_{p_x,j}$ is low, noise levels are constant, indicating constant flow traffic with many sources in the vicinity.

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 (Dominici et al., 2002, Pearce et al., 2011). The additive model in the context of spatial exposure modeling can be written in the form:

$$\log(BC_{px,j}) = \sum_{z=1}^{n} s_z(v_{z,px,j}) + \varepsilon_{x,j} \qquad (3)$$

Where v_z is the z^{th} covariate evaluated for trip j at location p_x; $s_z(v_{z,px,j})$ is the smooth function of z^{th} covariate, n is the total number of covariates, and $\varepsilon_{x,j}$ is the corresponding residual with $var(\varepsilon_{x,j}) = \sigma^2$, which is assumed 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).

Two modeling approaches will be discussed. The first option is to model the measurements on the basic aggregation level: one value for each parameter for each trip at each aggregation point p_x. This dataset contains about 37.700 records and will be referred to as the basic dataset (BDS). The second approach starts the modeling after aggregating the BDS to a dataset averaging the BC exposure for the classified parameters included in the aggregation models. Each of the retained parameters is classified according to a set of

predefined percentile classes (see supplementary data). For each of these two approaches the models will be evaluated for both the raw BC result (BC_{raw}) as for the local contribution BC_{loc} to assess the relevance of handling background concentrations separately.

3. Results

In this section, the following terminology will be used to refer to the different models that will be compared:

- BC_{bg} model: this model simply assumes that exposure during cycling trips equals concentrations measured at a regional background measurement station.
- BC_{raw} model: uses GAM to obtain BC directly from noise, weather, and geometrical data.
- BC_{loc} model: uses GAM to obtain BC_{loc} and adds measured BC_{bg} to obtain the overall concentration as shown in Eqs. (1).
- Aggregated BC_{raw} and BC_{loc} models: same as above but with predictive variables categorized in percentile classes.

3.1 BC_{loc} and BC_{raw} model

The BC_{loc} model is based on the BDS dataset and includes the three noise parameters, wind speed WS_{trip}, temperature $Temp_{trip}$ and street canyon index $StCan_{p_x}$. The parameters of the GAM models are shown in Table 1. The quality of a GAM model and the relative strength of its parameters are described by the deviance explained, the intercept, the F 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_{loc} model the intercept and the $L_{OLF,p_x,j}$ have a similar strength; the wind speed is about half the strength of $L_{OLF,p_x,j}$. The plots of the splines created by the GAM modeling show the relation of the parameter to the outcome $\log(BC_{loc})$ (Figure 1). In the BC_{loc} model the $L_{OLF,p_x,j}$ is the strongest component and has a linear relation with $\log(BC_{loc})$. $L_{HFmLF,p_x,j}$ saturates for higher levels. $\log(BC_{loc})$ decreases for high wind speeds. The street canyon effect is visible as a steady increase of the exposure with a higher street canyon index $StCan_{p_x}$. The temperature is not very important, the steep upward trend is based on only a

fewtrips performed at temperatures above 20 °C. Log(BC_{loc}) increases strongly for very small values of $(L_{Aeq} - L_{Amin})_{p_x,j}$ indicating episodes with almost constant noise levels result in higher BC concentrations.

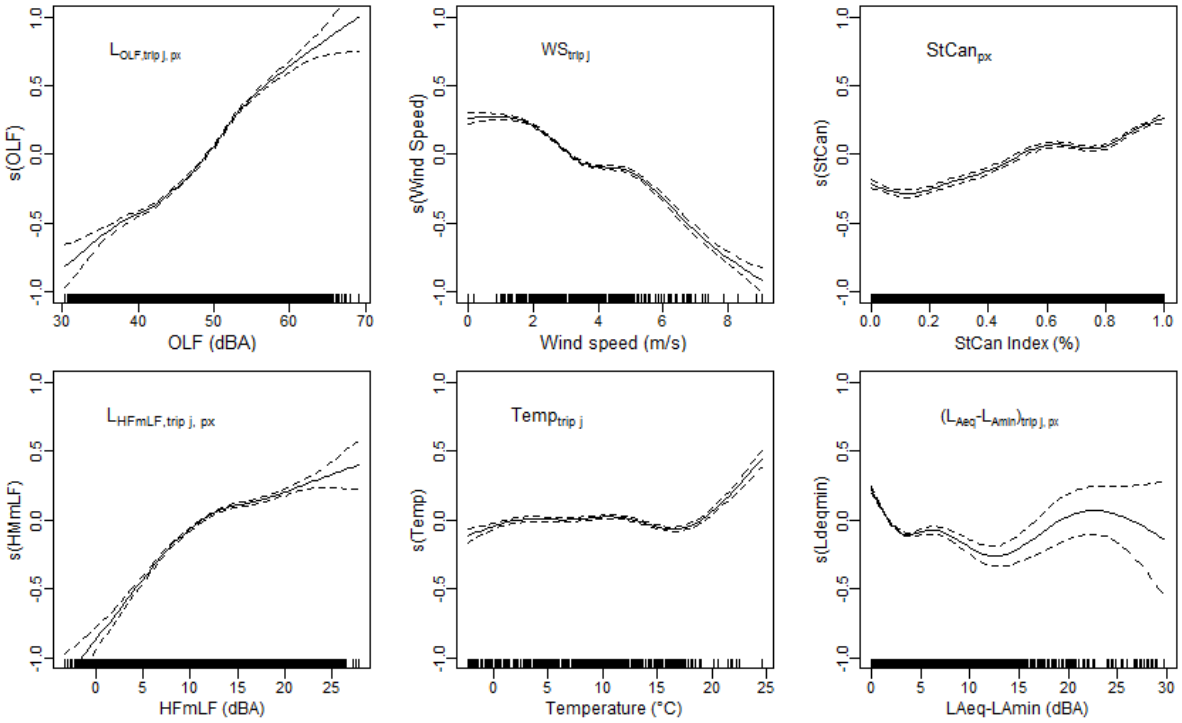

Figure 1: Splines of the six covariates of the BC_{loc} GAM model, ordered by strength within the model (top-left to bottom-right).

The BC_{raw} model shows a slightly higher deviance explained than the BC_{loc} model (Table 1). The BC_{raw} model also has a higher intercept. Wind speed is the strongest component. The low frequency noise level L_{OLF} and temperature have a similar strength.

Table 1: Comparing the results of the BC_{loc} and BC_{raw} models, F-value and p-value.

3.2 Aggregated BC_{loc} and BC_{raw} models

The aggregated models are built including respectively 3, 4 and 5 parameters. The temperature is not included since it proved to be of little relevance in the BC_{loc} model. In Table 2 the results of the aggregated models are assembled, including the number of unique combinations of the classified parameters. Again similar changes between the $BC_{raw,Xp}$ and $BC_{loc,Xp}$ models can be detected (where X is the number of aggregation parameters). In Figure 2 the splines of the five parameter model $BC_{loc,5p}$ are shown. The deviance

explained is higher and the splines are smoother compared to the BC_{loc} model due to the aggregation process. In the 3 parameter model L_{OLF} the only acoustic parameter describing the source has the strongest importance in the model, but as other traffic descriptors, L_{HFmLF} and (L_{Aeq}-L_{Amin}) are added, wind speed becomes the most important covariate.

Table 2: Comparing the results of the aggregated models.

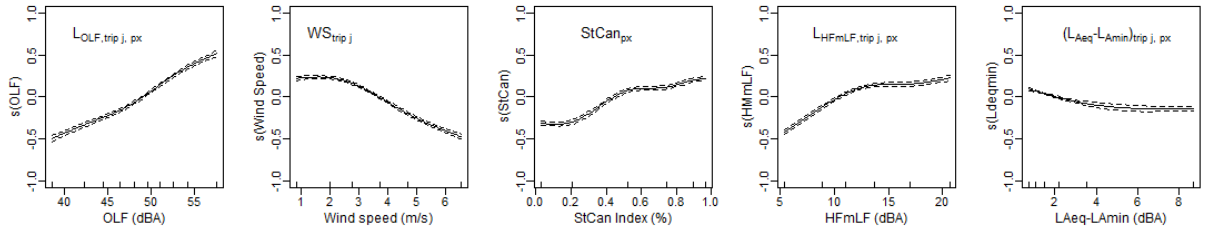

Figure 2: Splines of the five covariates of the GAM $BC_{loc,5p}$ model.

3.3 Comparing the fit of the models

The BC exposure during individual trips is reconstructed based on the models presented above. The results for the BC_{bg}, BC_{loc} and BC_{raw} models are shown in Figure 3. The second row shows the results of the BC_{loc} model, and the third row the results of the BC_{raw} model. Each point in de charts represents a single trip. The x-axis shows the measured BC exposure, the y-axis the model outcome. The first column shows the total trip exposure (as ng/m^3 * seconds), the second column the average exposure over the trip. Both are relevant, but the correlation in the first column is strongly influenced by the duration of the trip. The predictive quality of the models is best determined by the ability to predict the trip averaged exposure. In each of these charts, the diagonal (red) and the linear fit (green) on the results are shown and the correlation and the spearman's correlation between model and measurement are given. The properties (intercept and slope) of the linear fit are shown to evaluate the model fits. The third column shows the distribution of the relative prediction: total model BC divided by total measured BC, presenting the under- and overshoots of the trip total exposure prediction which is identical for the total and averaged evaluation.

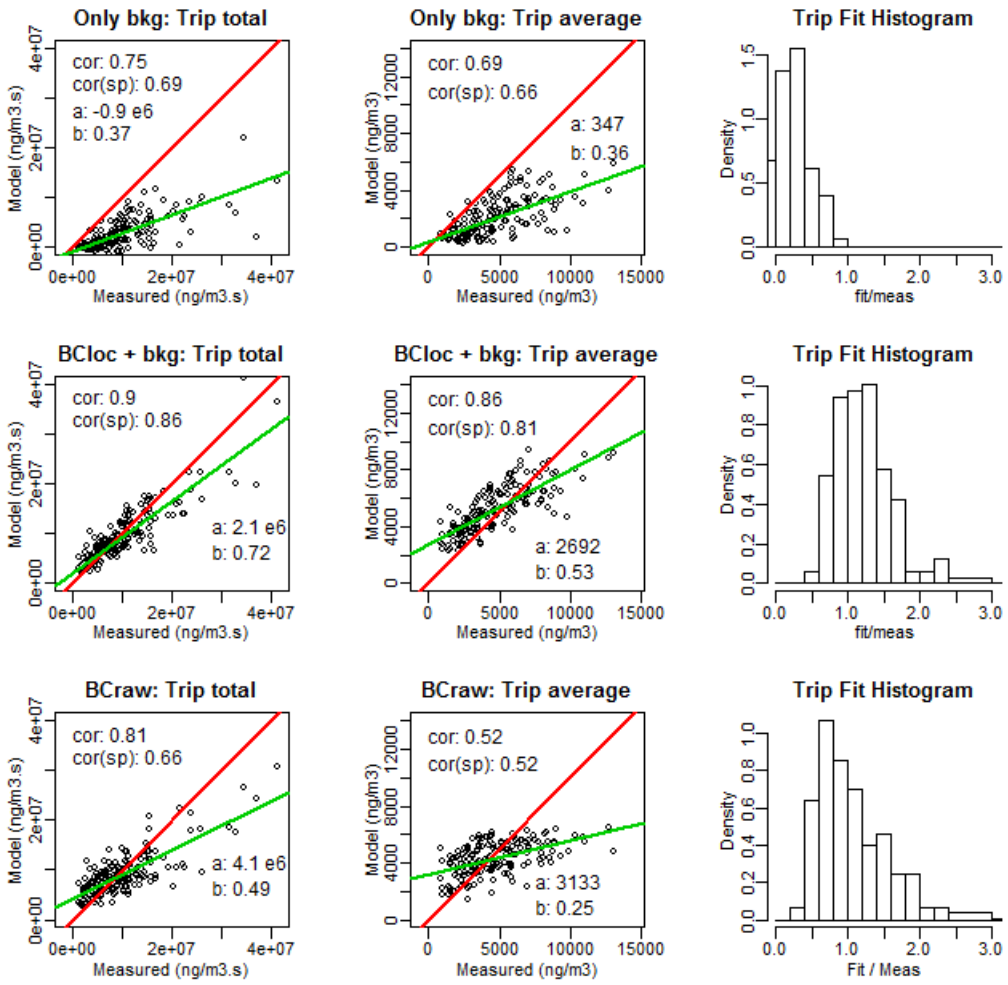

Figure 3: Evaluation of the model fit of the BC_{bg}, BC_{loc} and BC_{raw} models. Each point in de charts represents a single trip. The first row shows the results of the BC_{bg} model, the second row the results of the BC_{loc} model, the third row the results of the BC_{raw} model.

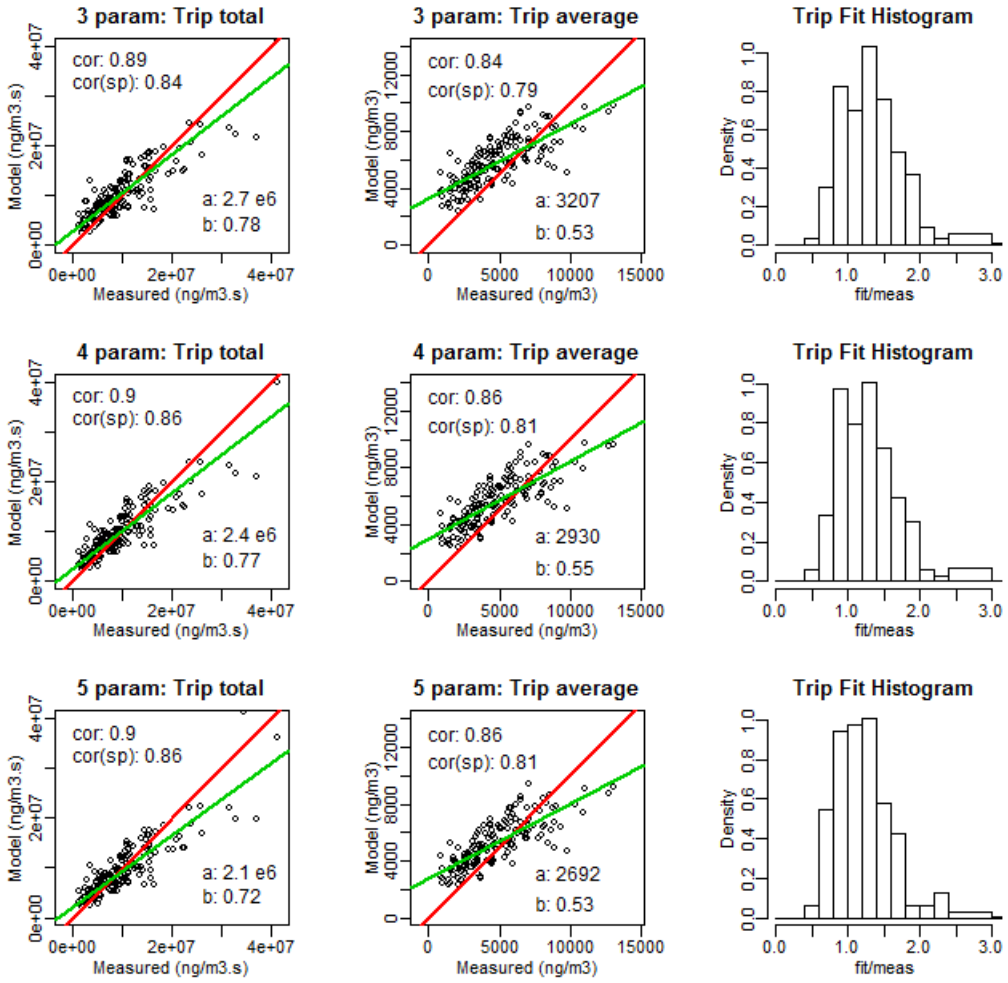

Figure 4: Evaluation of the model fit of the aggregated BC_{loc} models for 3, 4 and 5 parameters.

A similar procedure is performed for the aggregated models. For each passage at an aggregation point the corresponding classification is used to estimate BC exposure. Summing all results over the full trajectory reconstructs the total trip exposure. In Figure 4 the fitting properties of the reconstructed trips of the $BC_{loc,Xp}$ models are shown for the three aggregated models on BC_{loc}. The aggregated BC_{raw} model evaluations are not shown.

3.4 Comparing the predictive strength of the models

The predictive strength of the models is checked by building variants of the models based on a random subset of 75% of the available trips and then predicting the remaining trips. This procedure is repeated 50 times. For each trip, a set of approximately 13 predictions for different model variants are thus available. The

average BC trip exposure is calculated for the trips used to build the model (referred to as 'fitted trips') and predicted for the trips left out of the model (referred to as 'predicted trips'). The Pearson correlation of the fitted trips versus the measurements and the predicted tips versus the measurements are calculated for each model variant. The distributions of the correlations of the fitted trips of the model variants and the distributions of correlations of predicted trips out of the model variants are shown by model type in Figure 5. The mean correlations of the total exposure of the fitted trips for the models (as ordered in Figure 5) are 0.90, 0.86, 0.88, 0.89 and 0.89, the mean correlations of the total exposure of the predicted trips are 0.88, 0.84, 0.89, 0.90 and 0.90. The mean correlations of the averaged exposure of the fitted trips for the models are 0.78, 0.71, 0.75, 0.77 and 0.77; the mean correlations of the averaged exposure of the predicted trips are 0.75, 0.67, 0.75, 0.78 and 0.74. Models BC_{loc}, $BC_{loc,4p}$ and $BC_{loc,5p}$ have similar distributions for both the fitted trips and the predicted trips. The predicted trips show wider distributions compared to the fitted trips. BC_{raw} and $BC_{loc,3p}$ perform significantly worse than BC_{loc}, $BC_{loc,4p}$ and $BC_{loc,5p}$. The correlation of the predicted trips of BC_{raw} model is extremely sensitive to the trip selection. Figure 5C shows the relation between the correlations of the fitted and predicted trips for the individual model variants. Correlations of the predicted trips are not necessarily lower than the fitted trips. The correlation of the predictions can be higher than the correlation of the fitted trips, especially for the BC_{loc} and $BC_{loc,xp}$ models (Figure 5C).

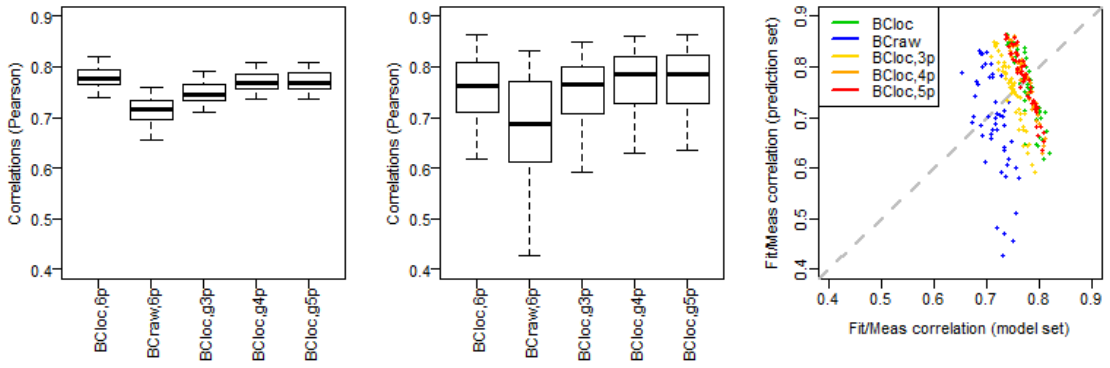

Figure 5: Distributions of the model variants correlation of the fitted trips (A), distributions of the model variants correlation of the predicted trips (B) and the relations between the correlation of the fitted trips and the correlation of the predicted trips for the individual model variants (C).

For each set of predictions of a single trip obtained from a model variant, the average prediction and the standard deviation of the trip predictions is calculated. The relative prediction (average of the total trip predictions divided by the total measured for trip j) and the standard deviation of the trip j predictions divided by the trip j total exposure are used to compare the predictive strength and sensitivity of the model

variants to the random sampling strategy. The results are shown in two charts in Figure 6. The BC_{loc} model is slightly underestimating the exposure, BC_{raw} and $BC_{loc,3p}$ are overestimating, the $BC_{loc,4p}$ and $BC_{loc,5p}$ models are centered around 1.0. The interquartile range of the distribution of the relative trip predictions for the BC_{loc} model is 0.38, the ranges are slightly larger for the $BC_{loc,4p}$ and $BC_{loc,5p}$ models, respectively 0.47 and 0.43. The interquartile range of the BC_{loc} model is 3.1, the BC_{raw} model performs much worse (IQR 5.8). The best IQR is found for the $BC_{loc,5p}$ model with 1.8. All aggregated models perform better than the BC_{loc} model for the distribution of the standard deviation of the trip predictions.

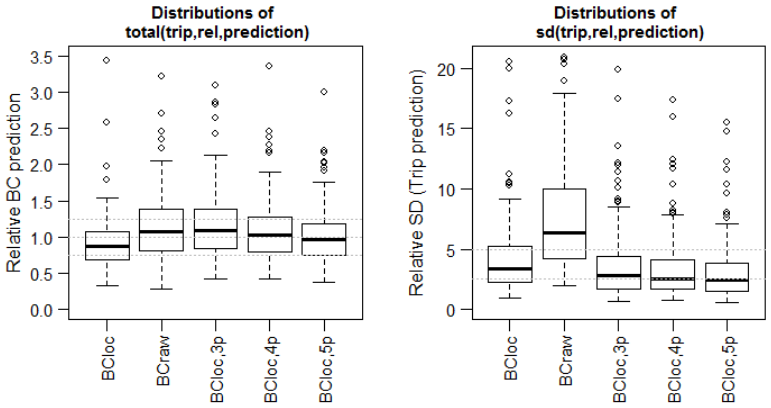

Figure 6: Distribution of the average of the relative trip predictions (A) and the distribution of the standard deviations of the trip prediction relative to the trip total measured BC (B), presented by model.

4. DISCUSSION

The first research question investigates whether a prediction model for BC exposure could be improved by separating out long term variations in the background concentration; the second research question whether the instantaneous local Black Carbon exposure could be predicted on the basis of local noise exposure and meteorological conditions. The first research question is addressed by comparing BC_{bg}, BC_{loc} and BC_{raw} models. The second research question is addressed by evaluating the features of BC_{loc}, $BC_{loc,3p}$, $BC_{loc,4p}$, $BC_{loc,5p}$ models.

The BC_{bg} model accounts on average for 30-40% of the total trip exposure only, but the slope in the relation between modeled and measured exposure is as large as 0.36 (Figure 3). High average trip exposure is partially related to higher background concentrations. Although the BC_{raw} model on average results in a good estimate of the average exposure during cycling trips, the trend line connecting prediction to measurement (Figure 3) slopes at 0.25. The strong influence of wind speed and temperature indicates that this model

306 mainly tries to resolve the temporal variations in the background contributions and fails to include indicators
307 for local exposure. Integrating the GAM model for BC_{loc} with measured BC_{bg} combines best of both worlds:
308 temporal variability of the BC_{bg} with spatial variability of the BC_{loc} model and results in a slope between
309 modeled and measured average trip exposure of 0.53.

310 A similar conclusion can be drawn when evaluating the predictive power of the models in Section 3.4. The
311 wide distribution of the correlations of the predicted trips in the model variants of the BC_{raw} model (Figure
312 5B) and the high relative standard deviation of the individual trip prediction (Figure 6B) illustrate the
313 sensitivity of the BC_{raw} model to the trip selection. All variants of the BC_{loc} model perform significantly better.
314 The BC_{loc} model is slightly sensitive to over fitting for the low exposure values, which is reflected in the
315 underestimation of the relative trip prediction (Figure 6A).

316 As it is now established that the BC_{loc} models outperform models that aim at predicting the whole
317 exposure at once, let us now look in detail at the choice of predictive variables for these models. In the BC_{loc}
318 model $L_{OLF,p_x,j}$ is the strongest component and wind speed the second strongest. The negligible importance of
319 temperature in the model can be explained as follows. Temperature is expected to affect emission and is an
320 indicator for meteorological conditions that influence dispersion. The background concentration indeed
321 shows a distinctive seasonal pattern and temperature is related to season (see supplementary data 1.9 figure
322 S3). BC_{bg} seems to include the temperature dependence sufficiently thereby reducing the relevance of
323 temperature in the BC_{loc} model. Wind speed has a similar relationship with background concentration as
324 temperature (supplementary data 1.9 figure S4). The strength of the wind speed covariate is therefore also
325 reduced in the BC_{loc} compared to its strength in the BC_{raw} model but in contrast to the temperature, it is still
326 highly relevant. This is explained by the observation that wind speed not only influences background
327 concentrations but also instantaneous dispersion of the local emissions. Higher wind speeds reduce the local
328 exposure of cyclists. StCAN and L_{HFmLF} are also stronger in the BC_{loc} model compared to the BC_{raw} model
329 probably because major confounders are eliminated by treating BC_{bg} separately. The BC_{loc} model becomes
330 more sensitive to spatially varying variables such as vehicle speed and local air pollution accumulation. The
331 latter effect and its dependence on street canyon geometry (measured here as StCAN) is confirmed by
332 physical calculations of street canyon accumulation, validated with measurements, such as the one presented
333 by Berkowicz et al (2008).

High-degree-of-freedom models such as the GAM model risk to over fit the data that they are based on while losing the ability to generalize to other situations. This can be prevented by reducing the number of input variables (covariates) and –in the particular case of GAM models – also by reducing the number of values that the variable can take. Classification of the covariates indeed reduces the degrees of freedom (Table 2) in the model. Because of the change in number of data points, the goodness-of-fit parameters and the strength of the covariates in the models cannot be compared directly, the quality of the models is determined by evaluating the properties of the predictions. In addition, a lower number of variables reduces the complexity of the model and simplifies implementation.

Turning to the model prediction evaluation shown in Figure 5 and Figure 6, it is clear that the $BC_{loc,g3p}$ model performs worse than the BC_{loc} model in predicting the average exposure. This is not unexpected as the information used in the prediction is significantly reduced. The reduction in number of covariates and the clustering of values is expected to make the model less sensitive to trip sampling. For $BC_{loc,g3p}$ the aggregation is too strong to observe this positive effect. The $BC_{loc,g4p}$ and $BC_{loc,g5p}$ models slightly outperform the BC_{loc} model when evaluated for correlation between model and the measurement data not used for fitting (Figure 5B). The distribution of relative trip predictions (Figure 6A) is wider for the $BC_{loc,g4p}$ and $BC_{loc,g5p}$ models compared to the BC_{loc} model, but the distributions are centered around 1. The relative standard deviation of the individual trip prediction (Figure 6B) is slightly better. The $BC_{loc,g5p}$ outperforms the BC_{loc} model despite the fact that it only reduces the number of data points from 37722 to 8832. The aggregation process removes the lowest values in the basic dataset, resulting in a better prediction of the total trip BC exposure.

It is more difficult to distinguish between the $BC_{loc,g4p}$ and $BC_{loc,g5p}$ models. The only relevant difference is the reduction of the intercept in the $BC_{loc,g5p}$ model. The short term dynamics (L_{Aeq}-L_{Amin}) covariate – the variable differentiating $BC_{loc,g5p}$ from $BC_{loc,g4p}$ – distinguishes between two different traffic conditions: low values can be linked to situations of congested traffic, many cars with constant noise emission (idling) result in constant noise levels; high values indicate short distinct events, typically a single car passing by at a relatively high speed. This last traffic situation is however rare during the rush hour and is under represented in our (rush hour only) database. When similar measurements would be performed outside the rush hour, this covariate might prove more relevant. With the available measurements it cannot be concluded that the $BC_{loc,5p}$ model is better than $BC_{loc,4p}$ model.

In general, the simplification of the model improves its quality as long as the 4 or 5 most significant parameters are kept. The models suggested in this paper perform well mainly due to the unique choice of noise indicators that are directly related to relevant traffic dynamics. The L_{OLF} covariate detects traffic volume including acceleration both linked to higher particulate emissions. L_{HFmLF} indicates the traffic speed; emissions increase with speed at low speeds, but saturate at higher speeds because constant high speeds result in more efficient combustion processes and lower particulate emissions. Similar relationships were reported for nanoparticles in Uhrner et al., 2011 and for gaseous emissions in De Vlieger et al., 2000. This also explains why models based on L_{Aeq} are less successful: L_{Aeq} is related to human loudness perception and does allow distinguishing situations with different traffic dynamics (Boogaard et al., 2008). Ross et al., 2011 relates a spectral noise evaluation to air pollution on a fixed monitoring station. Although these measurements are not directly comparable with the measurements in this paper, the same frequency bands are found to be relevant.

5. CONCLUSIONS

This paper proves that it *is* possible to predict instantaneous BC concentrations based on mobile noise measurements. The in-traffic exposure to Black Carbon of bicyclists is determined by background concentration, distance to source, local traffic density and speed, local traffic dynamics, local street geometry, and meteorological conditions. Spectral evaluation of (mobile) noise measurements can be used to implicitly detect local traffic conditions directly related to the local BC emission. Predicting personal BC exposure of cyclists proved only successful after splitting the model into a background contribution and a local component. In particular, it was shown that the spatial variability due to the local traffic contribution can be modeled using four parameters: the low frequency noise L_{OLF} related to the traffic volume and engine throttle, the difference between high and low frequencies L_{HFmLF} related to the traffic speed, the instantaneous wind speed and the street canyon index both related to the local accumulation of BC. The overall trip exposure is predicted by the four categorized parameters GAM model, $BC_{loc,g4p}$, with a correlation of 0.90. The average trip exposure is predicted with a correlation of 0.78.

The structure of the model presented in this paper can be expected to be valid for different areas in the world since parameters are chosen with a physical background in mind and because the model has been well

validated. Since fleet composition and driving behavior might differ between different parts of the world, it is suggested that each mobile noise measurement campaign is accompanied with a partial BC sampling to re-establish the model coefficients. The measurement underlying the specific model presented here, are performed in a typical (European) environment, a medium size city including suburbs and green areas, representative for the living environment of the majority of the population in Europe. The exact same model including model coefficients is expected to be applicable in any city sufficiently similar to this situation.

Mobile noise measurements on bicycles have the potential to provide the spatial detail and high temporal resolution that is necessary to predict the urban exposure to black carbon, including the local effects of route choice. Mobile noise measurements therefore have the power to replace large scale in-traffic personal air pollution exposure measurements and can be performed on larger populations at a significantly lower cost than traditional participatory sampling techniques. In this way the results obtained in this study could be useful for raising public awareness, changing personal behavior by selecting low exposure routes and to perform large scale epidemiological research on the impact of personal BC exposure on health.

6. REFERENCES

Berghmans, P.; Bleux, N.; Int Panis, L.; Mishra, V. K.; Torfs, R.; Van Poppel, M., Exposure assessment of a cyclist to PM(10) and ultrafine particles. Science of the Total Environment 2009, 407, (4), 1286-1298.

Boogaard, H.;Borgman F.;Kamminga J.; Hoek G.; Exposure to ultrafine and fine particles and noise during cycling and driving in 11 Dutch cities. Atmosferic environment 43 (2009) 4234-4242.

Berkowicz R., Ketzel M., Jensena S.S, Hvidberg M., Raaschou-Nielsen O., Evaluation and application of OSPM for traffic pollution assessment for a large number of street locations, Environmental Modelling & Software 2008, 296-303.

Can, A.; Botteldooren, D., Towards Traffic Situation Noise Emission Models. Acta Acustica United with Acustica 2011, 97, (5), 900-903.

Can, A.; Leclercq, L.; Lelong, J.; Botteldooren, D., Traffic noise spectrum analysis: Dynamic modeling vs. experimental observations. Applied Acoustics 2010, 71, (8), 764-770.

414 de Hartog, J. J.; Boogaard, H.; Nijland, H.; Hoek, G., Do the health benefits of cycling outweigh the risks?
415 Environ Health Perspect. 2010 August; 118(8): 1109–1116.
416 de Nazelle, A.; Fruin, S.; Westerdahl, D.; Martinez, D.; Matamala, J.; Kubesch, N.; Nieuwenhuijsen, M., Traffic
417 Exposures and Inhalations of Barcelona Commuters. Epidemiology 2011, 22, (1), S77-S78.
418 Dekoninck, L.; Botteldooren, D.; Int Panis, L., Guidelines for participatory noise sensing based on analysis of
419 high quality mobile noise measurements. Internoise 2012 (conference), 394-402
420 De Vlieger I., D De Keukeleere D., Kretzschmar J.G., Environmental effects of driving behaviour and congestion
421 related to passenger cars, Atmospheric Environment, Volume 34, Issue 27, 2000, Pages 4649-4655
422 Dominici, F.; McDermott, A.; Zeger, S. L.; Samet, J. M., On the use of generalized additive models in time-series
423 studies of air pollution and health. American Journal of Epidemiology 2002, 156, (3), 193-203.
424 Dons, E.; Int Panis, L.; Van Poppel, M.; Theunis, J.; Willems, H.; Torfs, R.; Wets, G., Impact of time-activity
425 patterns on personal exposure to black carbon. Atmospheric Environment 2011, 45, (21), 3594-3602.
426 Dons, E.; Int Panis, L.; Van Poppel, M.; Theunis, J.; Wets, G., Personal exposure to Black Carbon in transport
427 microenvironments. Atmospheric Environment 2012, 55.
428 Eisenman, S. B.; Miluzzo, E.; Lane, N. D.; Peterson, R. A.; Ahn, G.-S.; Campbell, A. T., BikeNet: A Mobile Sensing
429 System for Cyclist Experience Mapping. Acm Transactions on Sensor Networks 2009, 6, (1).
430 Int Panis, L.; de Geus, B.; Vandenbulcke, G.; Willems, H.; Degraeuwe, B.; Bleux, N.; Mishra, V.; Thomas, I.;
431 Meeusen, R., Exposure to particulate matter in traffic: A comparison of cyclists and car passengers.
432 Atmospheric Environment 2010, 44, (19), 2263-2270.
433 Isakov, V., J. S. Touma, et al. (2009). "Combining Regional- and Local-Scale Air Quality Models with Exposure
434 Models for Use in Environmental Health Studies." Journal of the Air & Waste Management Association
435 59(4): 461-472.
436 Janssen N.A.H., Hoek G., Simic-Lawson M., Fischer P., Bree L.v., Brink H.t., Keuken M., Atkinson R., Anderson
437 H.R., Brunekreef B. & Cassee F.R. (2011). Black Carbon as an Additional Indicator of the Adverse Health
438 Effects of Airborne Particles Compared to PM10 and PM2.5, Environ Health Perspect, 119(12):1691-
439 1699.
440 Karner, A.A., Eisinger, D.S., Niemeier, D.A., Near-roadway air quality: Synthesizing the findings from real-
441 world data. Environmental Science and Technology 44, 5334-5344, 2010..

Kanjo, E., NoiseSPY: A Real-Time Mobile Phone Platform for Urban Noise Monitoring and Mapping. Mobile Networks & Applications 2010, 15, (4), 562-574.

Maisonneuve, N.; Stevens, M.; Niessen, M. E.; Steels, L., NoiseTube: Measuring and mapping noise pollution with mobile phones. Information Technologies in Environmental Engineering 2009, 215-228.

Pearce, J. L.; Beringer, J.; Nicholls, N.; Hyndman, R. J.; Tapper, N. J., Quantifying the influence of local meteorology on air quality using generalized additive models. Atmospheric Environment 2011, 45, (6), 1328-1336.

Ross Z., Kheirbek I., Clougherty J., Ito K., Matte T., Markowitz S., Eisl H., Noise, air pollutants and traffic: Continuous measurement and correlation at a high-traffic location in New York City. Environmental Research, 2011, 1054-1063.

Uhrner, U., M. Zallinger, et al. (2011). "Volatile Nanoparticle Formation and Growth within a Diluting Diesel Car Exhaust." Journal of the Air & Waste Management Association 61(4): 399-408

von Klot, S.; Cyrys, J.; Hoek, G.; Kuhnel, B.; Pitz, M.; Kuhn, U.; Kuch, B.; Meisinger, C.; Hormann, A.; Wichmann, H. E.; Peters, A., Estimated Personal Soot Exposure Is Associated With Acute Myocardial Infarction Onset in a Case-Crossover Study. Progress in Cardiovascular Diseases 2011, 53, (5), 361-368.

WHO Europe, 2012: Health effects of black carbon, ISBN: 978 92 890 0265 3.

Wood, S. N., On confidence intervals for generalized additive models based on penalized regression splines. Australian & New Zealand Journal of Statistics 2006, 48, (4).