Project no. 265432

EveryAware

Enhance Environmental Awareness through Social Information Technologies

http://www.everyaware.eu

Seventh Framework Programme (FP7)

Future and Emerging Technologies of the Information Communication Technologies (ICT FET Open)

D1.2: Final report on: sensor selection, calibration and testing; EveryAware platform; smartphone applications

---

Period covered: from 01/09/2012 to 28/02/2014
Start date of project: March 1st, 2011
Due date of deliverable: Apr 30th, 2014
Distribution: Public

Date of preparation: 28/02/2014
Duration: 36 months
Actual submission date: Apr 30th, 2014
Status: Final

Project coordinator: Vittorio Loreto
Project coordinator organisation name: Fondazione ISI, Turin, Italy (ISI)
Lead contractor for this deliverable: Fondazione ISI, Turin, Italy (ISI)

Contents

List of Figures

Executive Summary

Based on the research previously reported in Deliverable 1.1 the air quality sensor box has been finalized (Fig. 1). AirProbe is an Android application designed to connect to the sensor box via Bluetooth, acquire sensors readings and transit them to the EveryAware servers as soon as a working connection to the Internet becomes available. Airprobe is an essential part in the EveryAware platform because without this application, the sensor box data cannot be accessed nor uploaded to the servers. In addition, the application allows users to visualise the data they collect as tracks on a map, provide estimated black carbon exposure values and timeseries of sensor output. AirProbe also allows the users to make free annotations (tags) to the measurements. AirProbe is freely available for the Android platform and can be installed from Google PlayStore. The sensorbox itself needs to be calibrated against reference measurements to translate raw sensor measurements to pollutant concentrations. We decided the approach the sensor box as a sensor array to estimate black carbon (BC) concentrations. This decision was based on the fact that gas concentrations of CO, NOx, O_3 and VOC correlate with BC, and by the fact that portable BC measurement devices were available for comparison with the BC estimation from the sensorbox. The BC device is approximately 30 times more expensive than the EveryAware sensorbox, and also lacks the AirProbe functionalities such as geographical mapping of the measurements, data tagging, etc. A calibration protocol was worked out, focusing on representativeness of the measurements. Representativeness was increased by including mobile data and indoor data into the calibration dataset, bearing in mind the final application of the sensor box when citizens will carry the sensor box around during daily life. Then, an artificial neural network model was developed to estimate black carbon concentrations from sensorbox data. Results were moderate. Especially mobile test data were hard to reproduce, probably by slower response times of the sensors compared to the reference device. However, results show that the use of a sensor array is preferred over a one-sensor system to map BC in urban areas. Furthermore, the agreement between the estimated BC form sensor box measurements and the reference measurements is substantial. Spatial patterns of BC coincide for many streets when based on sensor box measurements or

Figure 1: Picture of the sensor box.

reference BC measurements in a relative way, meaning that streets or locations with high or low concentrations coincide. Nevertheless, substantial differences between absolute BC concentrations are observed with sensor box measurements often underestimating real black carbon levels.

Outline of the document

Chapter 1 reports the final advances in the AirProbe mobile application. Chapter 2 describes the final advances in the WideNoise application: both Android and iOS versions has been improved with increased stability, the user interaction has been modified to encourage tagging, the protocol has been enhanced to provide more contextual information to the server, and the iOS version has been modified to support the new version of the operating system. Chapter 3 deals with the calibration procedure necessary to provide the user with a reliable and easy to understand parameter related to air quality (BC). An approach involving neural networks was used. In Chapter 4, an in depth evaluation of the sensor box is given, including a comparison of mobile reference and sensor box measurements.

Dissemination of the results

The results from this work have been presented on COST Action TD1105 EuNetAir, 2nd International Workshop EuNetAir, Brindisi (Italy), 25-26 March 2014 as an invited talk about Participatroy air quality sensing. The technical developments of the sensor box, AirProbe platform and calibration excersises is included in a publication that will be submitted soon to an international peer-reviewed journal (PLOS ONE).

Chapter 1

The sensor box for air quality monitoring

1.1 The AirProbe mobile application

AirProbe is an Android application designed to connect to the sensor box via Bluetooth, acquire sensors readings and transit them to the EveryAware servers as soon as a working connection to the Internet becomes available. Without this application, the sensor box data cannot be accessed nor uploaded to the servers. In addition, the application allows users to visualise the data they collect. Specifically, they can see their tracks on a map, calculate an estimated black carbon exposure and follow sensor output in real time plots. While collecting data, users can make free annotations (tags) that will be attached to the recordings and sent to the servers. AirProbe is freely available for the Android platform and can be installed from Google PlayStore. Please refer to delivareble D1.1 for more details about the design principles of AirProbe application.

In the final implementation, AirProbe can operate in different modes: Live Track, Synchronization and Browsing.

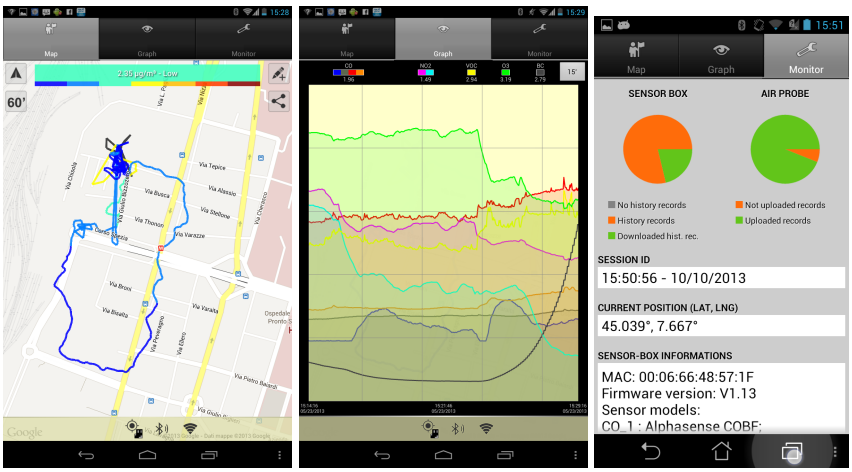

Figure 1.1: AirProbe: Live mode

In order to associate the data uploaded with a specific user, the application must be activated. This process links the application to an existing EveryAware account (which can be created inside the application itself or on the project web site).

1.1.1 Live Track mode

This is the standard way to use AirProbe to collect air quality measurements. In this mode, the application will search for Bluetooth devices nearby and present the user with a list of found devices. EveryAware sensor boxes can be easily identified by their MAC suffix. Once the user has selected the sensor box, AirProbe starts displaying real time data collected by the sensor box, using the Bluetooth connection. In Live Track mode, the interface is composed of three different views accessible from their corresponding tabs (Fig. 1.1):

Map where users can follow their own live track. The track is represented with different colours, depending on real-time black carbon levels. The user can also add annotations and share them on social networks (Facebook/Twitter), using the buttons at the top right corner. The track length to be shown on the map can be of 5, 15, 60 minutes. Camera tracking can be switched on/off, through the top left buttons. The bar at the top represents the black carbon value using a coloured scale (from blue-low value, to brown-high value).

Graph where the user can see black carbon behaviour and the raw data from pollutant sensors, in a variable time interval ranging from 1 to 30 minutes. The user can query the value registered by each sensor by tapping on the series. The graph is updated every two seconds.

Monitor where users can access statistics about collected data, connection information, the status of the sensor box and the installed sensors.

1.1.2 Synchronization mode

In this working mode, AirProbe downloads data from the sensor box and uploads them to the EveryAware server (Fig. 1.2). The sensor box in this case is used as a pure data logger, allowing the user to send data only in suitable conditions (e.g. where battery lifetime and/or connection billing are not a problem).

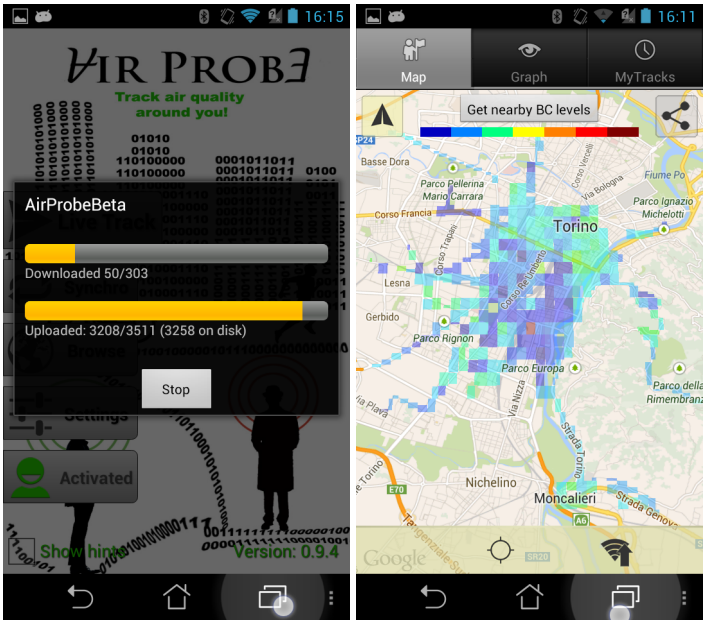

Figure 1.2: AirProbe: Synchronization and Black Carbon map

1.1.3 Browsing mode

This working mode does not requires an active Bluetooth connection to a sensor box. It is composed by three views, accessible from their corresponding tabs:

Map : the user can see the black carbon levels around his current position (Fig. 1.2), by pressing the "Get nearby BC levels" button. If a track from "MyTracks" tab is selected, this is displayed on the map. The black carbon levels and selected track can be shown together.

Graph : it shows the raw pollutant and black carbon behaviour, calculated for a selected track. Only live recorded tracks have black carbon data.

My Track : it shows the list of tracks available on the mobile device. Older tracks are automatically deleted once they have been uploaded to the server and only after a configurable time interval.

Chapter 2

The smartphone for noise pollution monitoring

WideNoise is a mobile application for recording, monitoring and analysing noise pollution. The application is intended to run on mobile devices and more specifically on Android[1] and iOS[2] platforms. It was originally developed by WideTag[3] and then was enhanced by the EveryAware team, who has improved and expanded the data recorded from the mobile device by adding new features. The mobile application sends anonymous data to an application server capable, through RESTful web services, of collecting the acquired data and showing the corresponding information on a map (details in supplementary material). Both sensor data and subjective perceptions are required to create a full sound report, so that the application consists of two main parts: the noise sampling component and the perception tagging. WideNoise application is described in deliverable D1.1. The following changes has been implemented in the final releases:

- Both Android and iOS versions has been improved and make more stable
- The user interaction has been modified to encourage tagging
- The protocol has been enhanced to provide more contextual information to the server (e.g. position accuracy)
- The iOS version has been modified to support the new version of the operating system

[1] http://play.google.com/store/apps/details?id=eu.everyaware.widenoise.android
[2] http://itunes.apple.com/app/id657693514
[3] http://www.widetag.com/

Chapter 3

Calibration of the sensor box

3.1 Introduction

Following the analysis of sensor abilities and conclusions from Deliverable 1.1, we have proceeded with calibration of the sensor boxes. Issues identified by our initial analysis included sensor sensitivity to temperature and humidity, hence the decision to include a temperature an humidity sensor in our sensor array, sensor drift in time and sensitivity to other gasses. Hence one needs to calibrate devices against a reference in order to control for this issues and obtain a measurement meaningful for the user.

Calibration means performing simultaneous measurements with the sensor boxes and a reference device, and then train a model that is able to map the values measured by our sensor array with the values recorded by the reference. We have used artificial neural networks (ANNs) for this regression task. For reference, we used micro-aethalometers (Figure 3.1) , which provide high quality measurements of black carbon (BC), however at a much higher cost (about 30 times more expensive than our sensor box).

Calibration started in Turin, where sensor boxes were built. There were several stages in data collection and model training, due to lessons learned on the way. The analysis and experience from Turin was then applied to the other cities for the calibration of the sensor boxes used during the AirProbe International Challenge. In the following we describe the approach taken. We first give more details about the different stages of data collection and analysis, the ANN calibration model, finalising with the calibration for the large scale international test case.

3.2 EveryAware sensorbox calibration approach

In this section we introduce the basic idea of the EveryAware SensorBox (EA-SB) calibration. Considering the amount of EA-SBs that will be employed in the network (around 100 EA-SB) and the difficulties and cost to calibrate each sensor present in each EA-SB we have opted to employ a statistical approach to calibrate each EA-SB to a target pollutant directly in an outdoor environment (later referred as field-calibration) rather than calibrate each sensor in a laboratory. The

Figure 3.1: Microaethalometer: device used as a reference for calibration.

Table 3.1: Cross-correlation between reference gas measurements at urban environment.

	Reference monitors				
	CO	NO	NO_2	O_3	BC
CO	1.00	0.77	0.62	-0.55	*0.83*
NO	0.77	1.00	0.76	-0.51	*0.89*
NO_2	0.62	0.76	1.00	-0.53	*0.81*
O_3	-0.55	-0.51	-0.53	1.00	*-0.54*
BC	0.83	0.89	0.81	-0.54	*1.00*

Table 3.2: Performance of the ANN, SVM and RF techniques to model BC from sensor box data on independent mobile data.

	R^2	rmse
ANN	0.26	2.10
SVM	0.23	1.70
RF	-0.13	1.72

field-calibration has been inspired by the works of Carotta et al. [Carotta et al., 2000] [Carotta et al., 2001] [Carotta et al., 2007], Tsujita et al. [Tsujita et al., 2005], Kamionka et al. [Kamionka et al., 2006] and De Vito et al. [De Vito et al., 2008] [De Vito et al., 2009]. The basic idea relies in using a supervised learning technique to model an unknown concentration of a target pollutant from sensor array measurements (based on low cost gas sensors, temperature and relative humidity sensors). The supervised learning model is parameterised by using a training dataset consisting of sensorbox measurements and simultaneous target pollutant concentration measurements. The target pollutant selected in this study was black carbon. The selection of black carbon as a target was motivated by following three reasons: (1) black carbon is a relevant pollutant in urban environment by its adverse health effects [EPA, 2010], (2) black carbon is correlated with the gases that are measured by the sensor box, and (3) the availability of portable black carbon measurement devices (micro-aethalometers, AethLabs) which makes it possible to collect mobile black carbon data. The correlation of black carbon with some of the gases that are measured by the sensor box is given in Table 3.1. Moderate (O_3) to high (CO, NO and NO_2) correlations with BC are observed.

The first calibration datasets were obtained in Antwerp in autumn 2012 at an air quality monitoring station at a traffic site and in Turin in spring 2013 from a two weeks long monitoring with sensor boxes and micro-aethalometers positioned at the entrance gate of the ISI company building near a busy road. These datasets were used to compare the performance different supervized learning techniques (a conceptual representation of how these empirical models work is given in D1.1). We have explored four different possible models to use for mapping of sensor output to the reference measurements. These were Random Forests, Support Vector Machines (SVMs), a custom air quality index and Artificial Neural Networks (ANNs). After comparing these (Table 3.2), SVMs and ANNs obtained similar behaviour, better than the other two options, but training of ANNs appeared to be faster so we decided to adopt them for our model.

ANNs[Mitchell, 1997] are regression models that mimic the behaviour or neuronal networks. They consist of interconnected computing units (neurons) that can have several inputs and an output. In each unit, two operations happen: compute a weighted sum of the inputs and apply a sigmoid (activation) function to obtain the output. The network can have several layers feeding into one another: One input and one output layer plus a number of hidden layers (see Figure 3.2).

In order to train a network, one needs to select a topology and find the values of the input weights for each neuron. To select the topology we have performed an empirical analysis that led to the usage of a network with one hidden layer of 10 neurons. The final topology is displayed in Figure

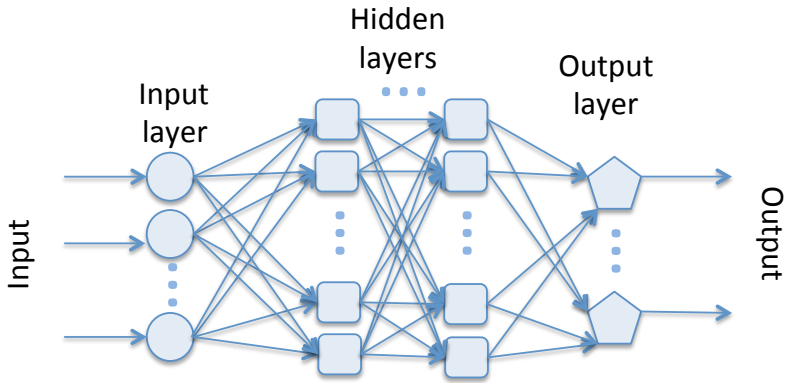

Figure 3.2: General structure for artificial neural networks.

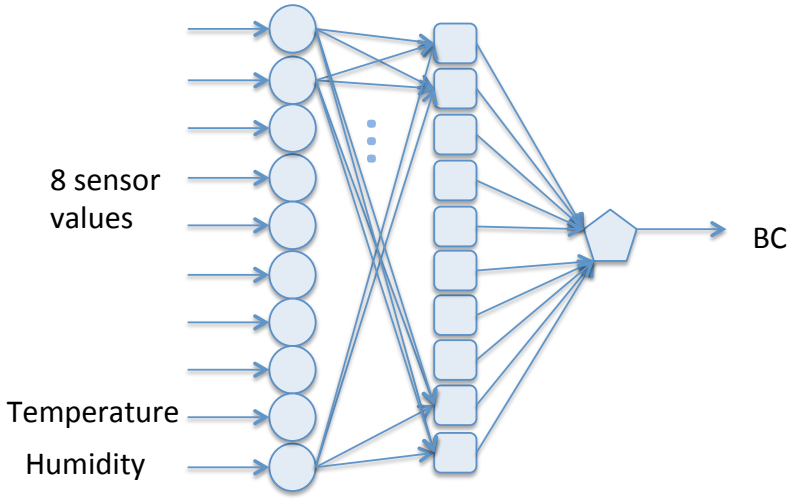

Figure 3.3: ANN topology for our calibration problem.

3.3 After this, we used backpropagation, a standard algorithm for ANN training [Mitchell, 1997], to obtain the network weights.

Emperical tests were performed to investigate the use of secondary features (computed from the original sensor data) or the inclusion of time of the day as an extra feature. None of these additional features improved the model performance substantially. Another important assessment was related to the use of one model for several sensor boxes. Bearing the APIC in mind, it would be computationally beneficial to have one model for each city. From the first calibration data sets we observed that sensor boxes behaved similarly when exposed simultaneously (the absolute sensor values differ between boxes, the fluctuations in time are similar for the same sensors). This means that sensor box rescaling (i. e. normalization of the sensor signals) could be used to scale the different sensor boxes within the same range, end parameterize a model on the standardized data, which is then applicable to all the sensor boxes. Tets indicated no significant differences in model performance using sensor box specific models or a more general model that is applied to sensor box specific rescaled sensor box signals. For APIC we decided to use one calibration model for each city, and to derice the sensor box specific rescaling terms form the calibration data.

The ANN model was implemented both in the AirProbe application, to give the user real time feedback from the sensor box, but also server side. This approach was taken due to the fact that the sensor box has two working modes, one online and one offline. Computing model output for all offline records would have been too computationally expensive for an average smartphone, while

server side this was not an issue.

3.3 Calibration Stages

The sensor box calibration involves different tasks. Firstly, all sensor boxes are installed in a representative environment, i. e. in an environment where they will be used in future applications. The empirical nature of the calibration model requires that future measurements are captured within the training data, these models are not good at extrapolating beyond the training ranges. This is particularly challenging for an outdoor air quality environment, where seasonal, daily, hourly and higher frequency fluctuations are observed. It implies that the validity of a calibration model is likely to be restricted in time because pollutant dynamics change through time (e. g. winter versus summer period). The first task is thus the collection of representative data of simultaneous measurements with the sensor boxes and the reference BC devices. Simultaneous measurements mean that the instrumentation is positioned at the same location and that the time resolution of the instruments is (about) the same. Further, a calibration model has to be parameterized based on training data. Finally, the calibration model has to be tested (cross-validation and fully independent testing) to investigate the model performance. An overview of how these tasks were performed in EveryAware is given below.

3.3.1 Stationary Calibration in Turin

A first attempt at calibration was performed in January 2013, when 20 sensor boxes were used to collect stationary data, along with two aethalometers. Figure 3.4 shows the setup of the boxes, which were installed at a sheltered location close to a main street (one way street with three lanes, an important artery in Turin). Data was collected over 2 weeks, however after initial analysis the measuring range of the boxes had to be readjusted so only 5 days of correct data were collected and used for calibration in the end.

Due to high variability both in BC values and sensor response, the data was smoothed by computing averages over moving time windows of five minutes. This value was deemed suitable by comparing outputs from the two aethalometers, which become highly correlated at this resolution. So the BC value obtained from the model represents an average over the last five minutes of exposure.

3.3.2 Parameterization of the calibration model

Although initially the possibility of building one calibration model for each box was intended, this would not have scaled very well, so we explored the possibility of building one model for all sensor boxes. This has also the advantage that data from multiple boxes can be used, resulting, possibly, in better modelling performance. For this, we first compared sensor output for the different sensor boxes. Gas sensors produce a signal with a value in $[0, 5]$, however only part of this interval is actually used. It is the setting of the potentiometer (done by hand by our engineers) that determines the exact range, so differences between the output of the same sensors on different sensor boxes are impossible to avoid. However, in principle, the fluctuations should correlate. Figure 3.5 displays an example of two collocated sensor boxes and the sensor responses. While some sensors have similar ranges, some others do not (for instance the dashed red sensor). However it appears that a linear scaling could bring the values in the same interval.

Considering this, a scaling procedure was employed in order to enable the use of one model for all boxes. Using the entire stationary data when all boxes were performing measurements together, the active range of each sensor was determined and then all data rescaled so that the active range falls to interval $[0.2, 0.7]$. This allows for measurements outside the range shown during the

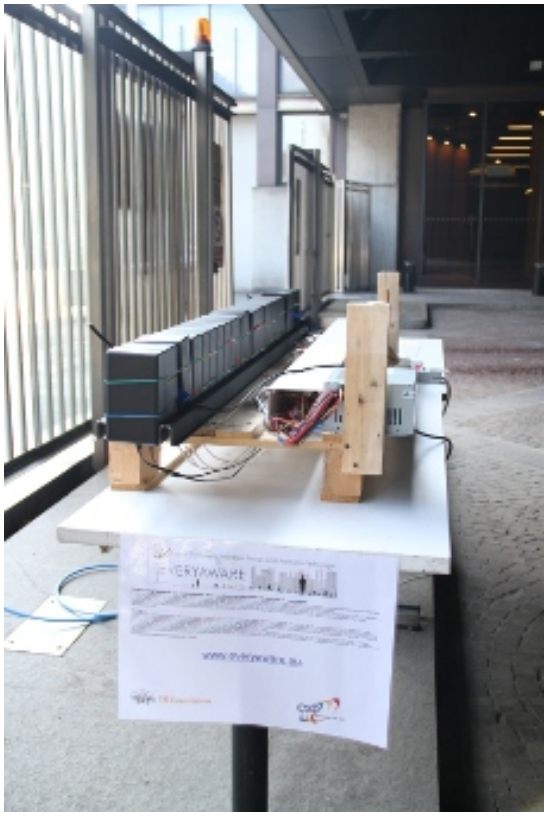

Figure 3.4: Setup of sensor boxes for stationary data collection.

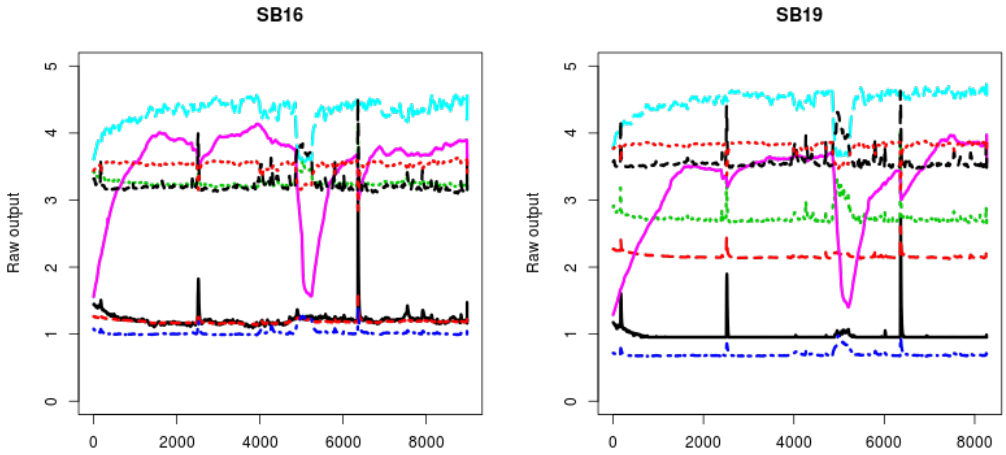

Figure 3.5: Comparison of sensor output for two sensor boxes.

stationary measurements to appear in the future. Any future data will be scaled using the same scaling parameters extracted from the stationary collocated data. Temperature and humidity, that also vary from one box to another (due to small differences in air flow or position) were scaled to interval $[0.1, 0.4]$.

From the calibration data min and max values were determined for each of the sensors and for each of the sensor boxes. The min and max values are used for data scaling, prior to application of the model. For a given sensor box, $\min_i$ and $\max_i$ correspond to the minimal and maximal sensor value during the calibration period for sensor i where i corresponds to the co_alpha, co_mics_1, co_mics_2, co_figaro, no2_mics, no2_figaro, voc_as, o3_mics, temperature or relative humidity sensor.

The scaling of the gas sensors is done by following rescaling formula:

$$xs_i = \frac{x_i - \min_i}{\max_i - \min_i} \times 0.5 + 0.2. \tag{3.1}$$

The scaling of the temperature and relative humidity data is done by using this formula:

$$xs_i = \frac{x_i - \min_i}{\max_i - \min_i} \times 0.3 + 0.1. \tag{3.2}$$

The rescaled ranges are quite narrow so that future measurements lower than $\min_i$ or higher than $\max_i$ probably stay within the $[0, 1]$ interval. The rescaled data are used to parameterize a ANN calibration model. A file with model parameters for was saved, including the vector b1 (10 elements), the 10 by 10 matrix of weights of the hidden layer, the vector w2 (10 elements) and on the last one the value of b2 (one element), which are the output layer weights. These parameters are be used to model the black carbon concentration from xs_i as:

$$ann = tanh(b2 + \sum_{j=1}^{10} [w2_j \times tanh(b1_j + \sum_{i=1}^{10} [w1_{ij} \times xs_i])]). \tag{3.3}$$

The value *ann* is scaled back to real black carbon value:

$$bc = 21441.63 \times \frac{ann + 1}{2} \tag{3.4}$$

Using scaled sensor values, a training procedure was initiated for the ANN. To speed up training, data was sampled every five seconds. Cross validation with test data has been performed. All training has been repeated several times and the model with best behaviour on the test data was selected. Model performance was evaluated using 3 criteria: R^2, root mean squared error (RMSE) and Pearson correlation coefficient between the modelled and the measured BC time series.

3.3.3 Testing of the calibration model

An initial test concentrated on obtaining a model for one sensor box only. Figure 3.6 shows performance on both training an test data: the first $\sim$3 days of data (first 8000 points) were used for training, the last $\sim$2 days for testing. Two time series are shown, one in red, representing the model output, one in blue representing the data measured by the aethalometer (this applies to all plots from here on). While performance on training data is quite good, on test data there is a shift in the modelled values, which could be due to drift in time.

The same model, obtained from data for these days from one sensor box, was applied to data from a different box, after scaling. This was meant to evaluate how much a model can be translated between boxes. Figure 3.7 shows model performance in this case. RMSE and R^2 values are comparable to those in Figure 3.6, indicating that indeed, after scaling, one model can be used for all boxes.

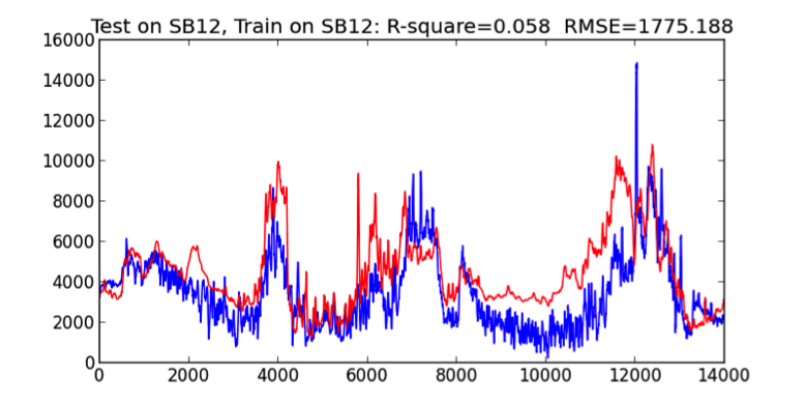

Figure 3.6: Train on SB12, test on SB12.

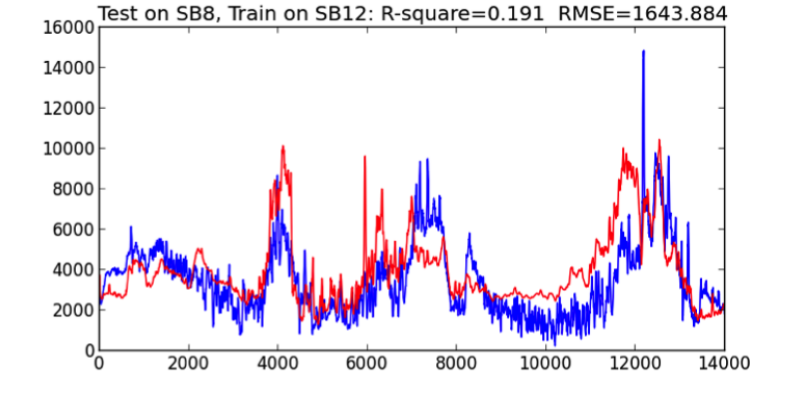

Figure 3.7: Train on SB12, test on SB8.

Since we were aiming for only one model, data integration from multiple boxes was performed to boost model performance. Five boxes were used for training (five days of data for each box) to obtain a final model from stationary data. This was tested on additional sensor boxes (Figure 3.8 shows performance on a sixth box), with better performance compared to using one box only (R^2 larger that 0.6).

Since the sensor boxes are intended for mobile measurements, we also tested the model on mobile data, not used for training (a dataset collected in March 2013, consisting of three hours of measurements, including two indoor periods, with one sensor box). Figure 3.9 shows performance of the model trained with data from five boxes on this dataset. Outdoors, the modelled BC levels match well enough the data (R^2 would be 0.26 if indoor periods were removed), but not as well as the static measurements. Indoors the model performs very badly, with a negative correlation to the real data. This because of sudden changes in humidity and temperature that are not included in the training data.

The results obtained until this point of our analysis indicates that while stationary data is useful, performance of the model on mobile data is reduced. At the same time, indoor data is necessary otherwise the sensor box response correlates negatively with reality. Although the sensor box is designed for outdoor use, this behaviour can be very confusing for the user, so should be avoided.

General trends in the data are captured by the models, however local fluctuations and large peaks are not visible. This can be due to the lower sensitivity of the sensors compared to the aethalometer, and also due to a delay in sensing. This problem might be exacerbated also by the usage of a large time window for averaging. Hence, in the following experiments we change the averaging window to 1 minute for the sensors (the aethalometer data is still averaged over 5 minutes). So,

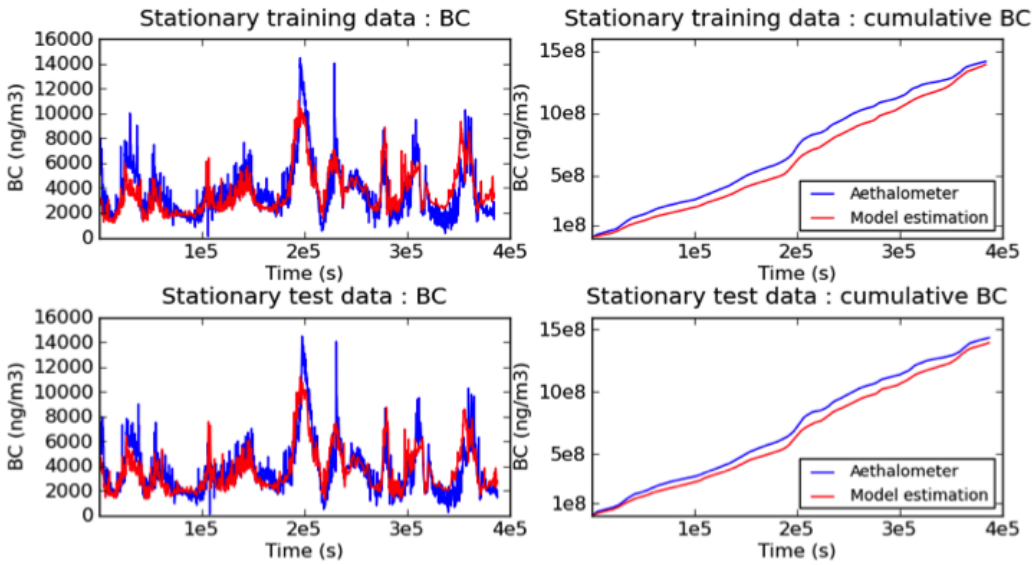

Figure 3.8: Train on five boxes, test on a sixth. The R^2 values are 0.62 for training and 0.67 for testing, while RMSE values are 1299 and 1225, respectively.

the modelled BC will still represent an average over the last 5 minutes, but we process only the last 60 seconds since some of the delay is incorporated in the data itself.

3.4 Mobile and Indoor Calibration for Turin

The first results presented in the previous section showed the need for mobile and indoor data for calibration. Considering this, we collected one dataset in March 2013, of about three hours, with one SB, then four additional datasets in June 2013 (2 SBs). The setup used for mobile measurements is displayed in Figure 3.10.

With these new datasets, a new model was trained. Raw data was averaged, scaled, as described perviously, and then used to parametrise an ANN using the following procedure:

- Cross validatation: For training: (1) outdoor stationary data from January-February (1 SB), (2) indoor stationary measurements from June (2 SBs), (3) mixed mobile measurements from June (2 SBs). For testing: (4) mixed mobile measurements from March (the same test dataset used for the previous calibration).
- Moving averages were computed over windows of 60 seconds for sensors and 300 seconds for aethalomenters.
- Data scaling - based on dataset (1), after averaging (using the same min max values previously derived). The range seen is these data has been scaled to [0.2,0.7] for gas sensors, [0.1,0.4] for temperature and humidity, [0,0.7] for BC.
- Dimensionality was reduced by considering one point every 5 seconds instead of all.
- Ten training sessions have been performed and the best one (in terms of R^2) was chosen.

Figure 3.9 displays the performance of the previous model (obtained from outdoor measurements) on mixed data, while Figure 3.11 shows the best model obtained from both indoor and outdoor data. Both qualitatively and quantitatively, the performance is increased. The new model displays higher variability, due to the difference in averaging windows. The cumulative is underestimated more than by the old model, but only because the old model strongly overestimated indoor pollution. This behaviour demonstrates that mobile and indoor data can help boost model performance.

Individual sensor effects The same training procedure has been repeated for subsets of sensors, in order to test wether some sensors can be removed from the sensor box. A special interest was in testing wether the Alphasense sensor could be avoided, since this would reduce SB costs significantly. Each subset of sensors was obtained by eliminating one of the gas sensors. Figure 3.12 shows the distribution of R^2 obtained in 10 training sessions for each subset. It appears that removal of the NO2_mics sensor is beneficial, significantly improving both best and average performance. The Alphasense sensor appears to be important since its removal decreases the model fit significantly, so no cost reduction could be obtain by eliminating this sensor.

Considering this, the CO Alphasense and NO2_mics sensors have been also removed together. However, performance is still reduced compared to using all sensors or removing the NO2_mics.

The results we obtained show that one model for all boxes is possible. However each location requires a new calibration session due to different response of the sensors in different concentration of gasses (which changes from location to location due to changes in fuel usage). Our experience shows that three types of data are required for calibration. To enable scaling, stationary collocated data should be collected for at least 2-3 days. Mobile data is required to account for changes in environment and allow the model to learn from more diverse data. Indoor data is also necessary to teach the model the transition between a more ventilated and a closed environment. While mobile data requirements have been envisioned prior to our tests, indoor behaviour was unexpected so required an adjustment of the data collection strategies planned.

3.5 Calibration for the international test case

3.5.1 Collection of representative data

Before the start of the Airprobe International Challenge (see D3.2) a parallel calibration data collection was set up in the four participating cities (30th of September to 11th of October 2013). In Turin, due to large availability of previous stationary data, only mobile measurements were performed. In the other cities, stationary data was also collected, using set-ups similar to that in Turin: all the sensor boxes were positioned at a near-traffic site together with two micro-aethalometers. Guidelines for the SB calibration were communicated among the project partners including a detailed description about site selection, installation of the sensor boxes and the micro-aethalometers, a maintenance procedure for the instruments, data upload from the instruments and data sharing, and the time schedule for specific tasks.

The calibration sites were within the area that was selected for final deployement at each city. The calibration measurements were carried out at a station from the Flemish air quality monitoring network in Antwerp. The station (Borgerhout, 42R801, see www.ircel.be) is situated at a traffic location along a double lane main street with an average daily traffic volume of 43,381 vehicles (42,961 cars and 420 heavy duty vehicles, data from the Traffic Centre Flanders). The calibration in Kassel was done near the front-roof of the City Hotel Kassel[1] in Wilhelmshöher Allee 38–42. The Wilhelmshöher Allee is one of the largest streets in Kassel with four lanes.

Continuous measurements at 1-second resolution were made. The clocks of the micro-aethalometers were synchronized with the sensor box time. After three consequtive days of mea-

[1] http://www.city-hotel-kassel.de/de/index.php

surements, data from each individual sensor box were screened for sensor failures or out-of-range measurement series (i. e. measurements outside the sensor range). None of the sensors was failing, no replacements needed to be done. Some sensors went out-of-range, and internal electrical circuits had to be adapted (by increasing/decreasing the resistivity) inside the sensor box. After this intervention, the sensorboxes and micro-aethalometers were continuously measuring (day and night, weekdays and weekend) during 10 days. A regular inspection of the set-up was carried out every two days, micro-aethalometer filters were replaced during these visits as well.

The objective is to use the sensor box in a mobile way. Therefore calibration data from mobile use of the sensor box was also required. During the stationary data collection a selection of sensor boxes and micro-aethalometers were taken on a regular basis to perform mobile measurements. Efforts (manpower) is much higher for conducting mobile measurements. In Antwerp for example, we collected 8 hours of mobile data with four different sensor boxes. The final numbers of mobile data was much lower than the stationary data, however, we feel that mobile calibration data are a valuable addition to the training set. Furthermore, mobile data are suitable for model evaluation because the final use of the sensor box will be (mostly) mobile. In Turin, over one hour of additional mobile measurements were collected for each sensor box, resulting in over ten hours of measurements in total. This, together with the previously collected data, formed a more extensive mobile dataset compared to the other locations.

Finally some indoor measurements were also included. A selection of sensor boxes and micro-aethalometers was placed in an indoor environment for some hours. The differences in indoor environments are huge depending on local sources, people density, ventilation characteristics etc. and the indoor measurements included in the calibration data are by no means exhaustive.

3.5.2 Model training and testing

After the data collection period, training datasets were compiled from the sensor box measurements (e. g. date, time, latitude, longitude, sensor values) merged to micro-aethalometer BC measurements based on date and time. One dataset was constructed for each sensor box available. The BC values were post-processed by a noise reduction algorithm [Hagler et al., 2011] to lower the high-frequency instrument noise that is observed when measuring at high frequency. BC levels were further smoothed by averaging over a five minute moving window, while for the sensor values the smoothing was done over 60 second windows (see section 3.3.3). The datasets are now ready to be used for model calibration. The representativity of the datasets is crucial. We increased the representativity of the data by placing the measurement equipment within the final area of deployment (city-specific), by collecting data continuously during day and nighttime, by collection stationary, mobile and some indoor data, and by collecting data over quite a long period (10 days) right before the final deployment (sensor boxes were used in the APIC two weeks later).

Separate ANN training sessions were performed for each location, with four resulting models to be used for APIC. In all locations, stationary data was integrated with all mobile and indoor datasets available, except for one used for testing. Figures 3.13, 3.14, 3.15 and 3.16 display the best model obtained in each city. The Turin model (Figure 3.16) performs best, which demonstrates that collection of large amounts for mobile data is crucial for boosting modelling abilities. The Antwerp dataset also contained larger amounts of mobile data, compared to the other two, so that good performance was obtained as well (Figure 3.13). For Kassel (Figure 3.14) and London (Figure 3.15), although datasets were more restricted, indications were that models obtained were displaying acceptable performance.

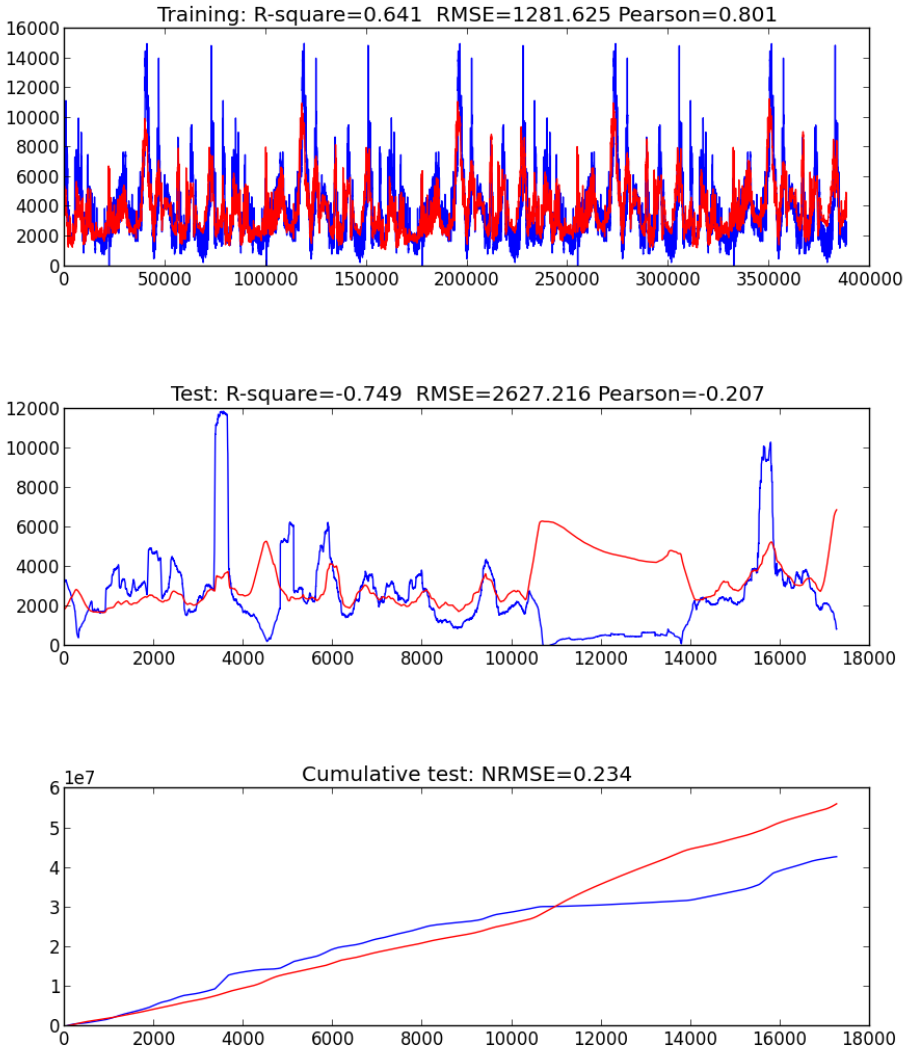

Figure 3.9: Train on five boxes, test on mobile data.

Figure 3.10: Setup of sensor boxes for mobile data collection.

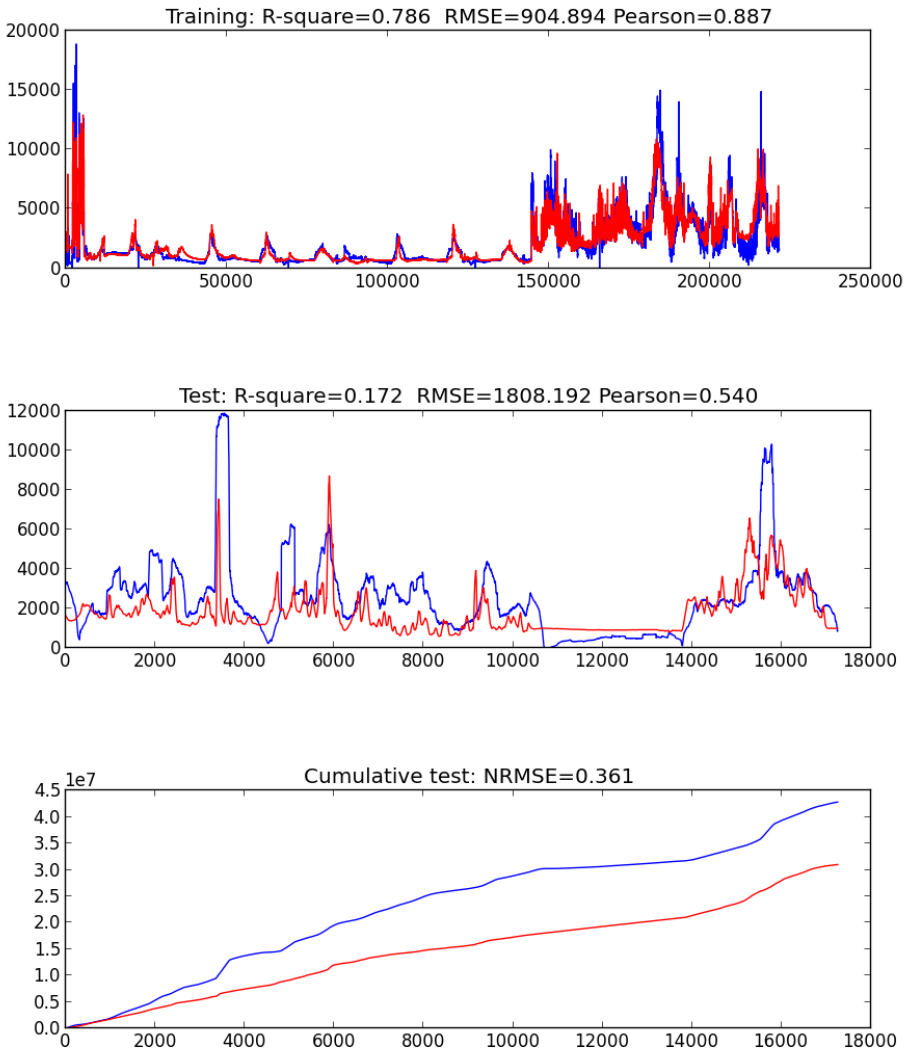

Figure 3.11: New model performance, after addition of mobile and indoor data

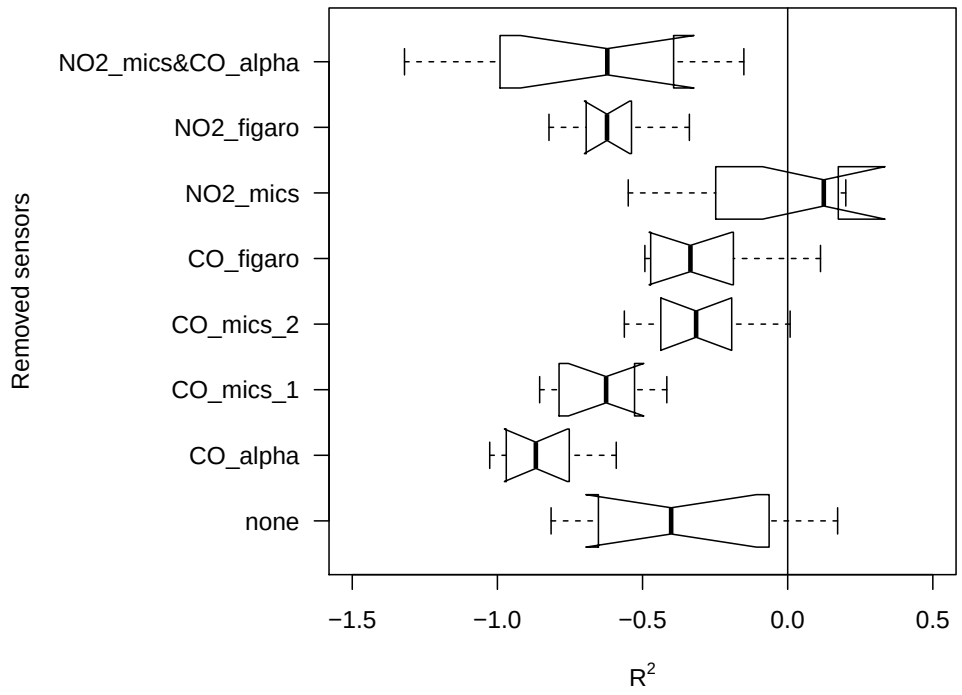

Figure 3.12: R^2 over ten training runs after removal of one sensor at a time or NO2_mics and CO Alphasense together.

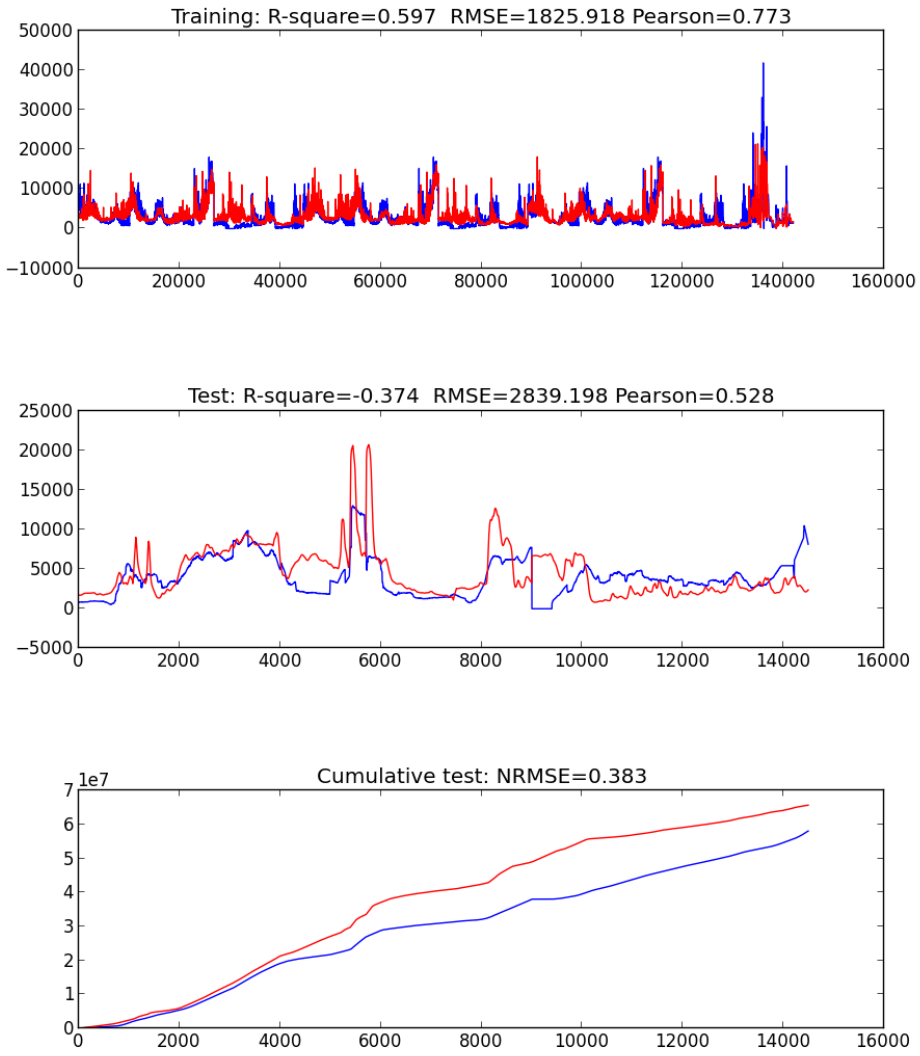

Figure 3.13: Model performance in Antwerp

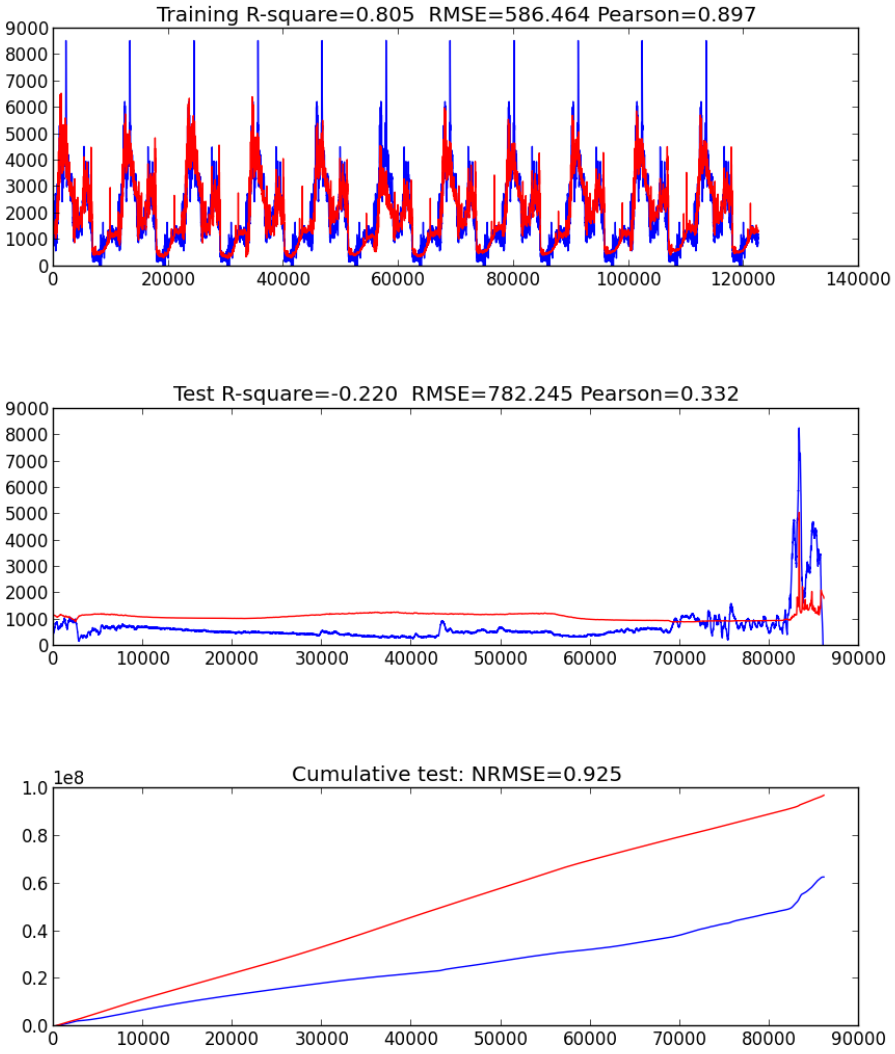

Figure 3.14: Model performance in Kassel

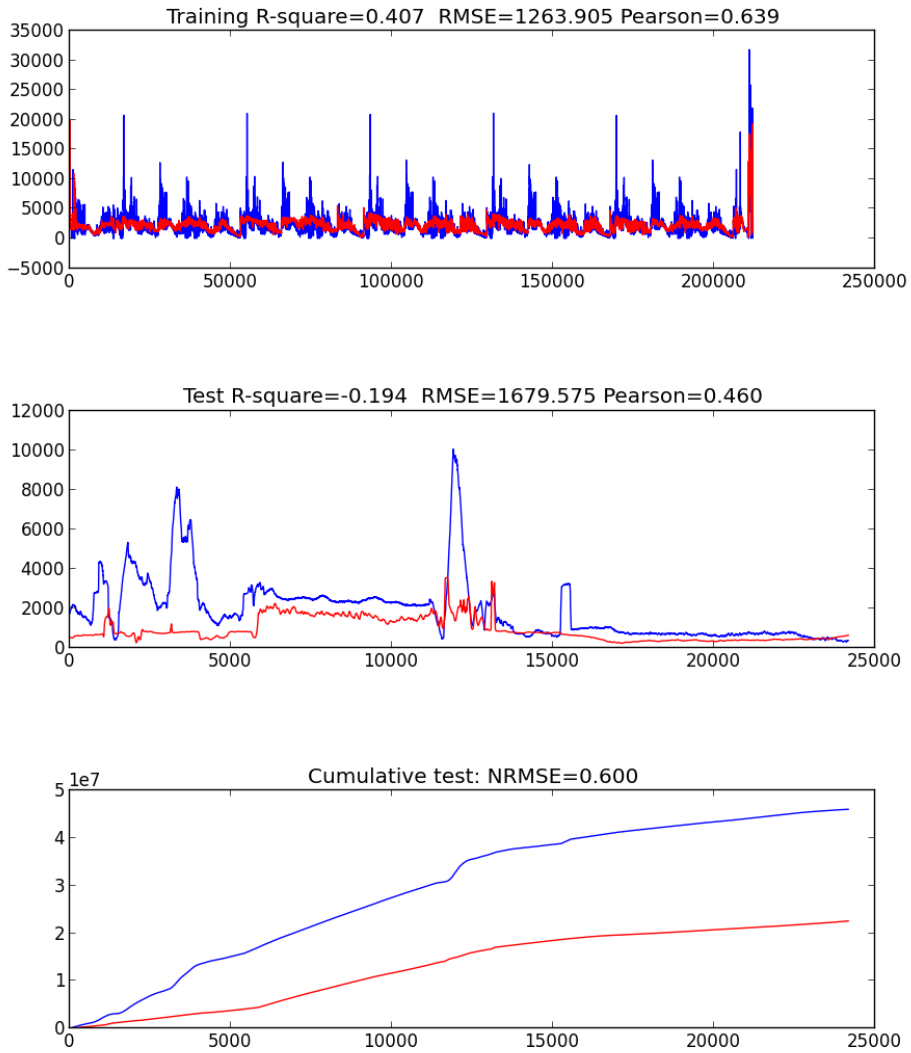

Figure 3.15: Model performance in London

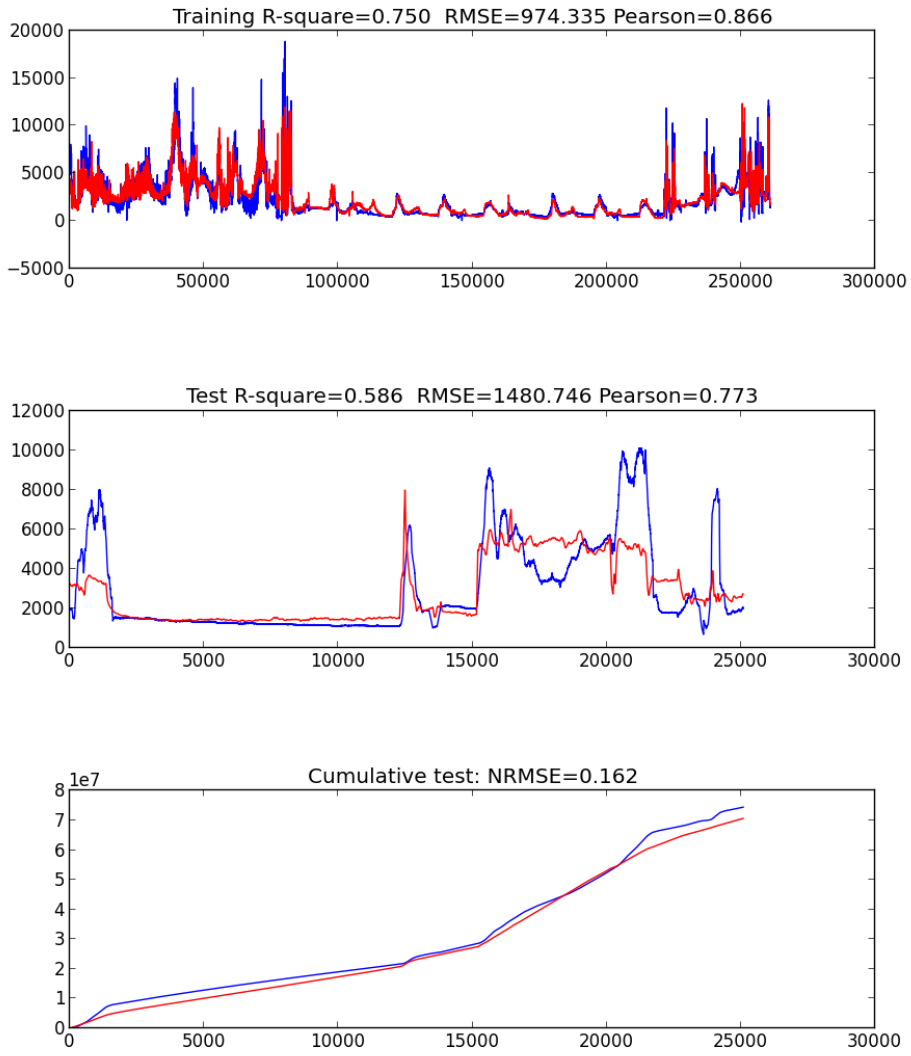

Figure 3.16: Model performance in Turin

Chapter 4

Evaluation of the SensorBox

4.1 Introduction

This chapter reports about additional tests that have been carried out for further evaluation of the sensor box. Specific evaluation tests were carried out in the course of the project, starting with a stationary evaluation of the sensor signals using the final versions of the sensor box compared to reference measurements. Furthermore, a comparison of observed and modelled black carbon concentrations from mobile measurements is made, and finally an assessemnt of spatial black carbon patterns from micro-aethalometer and sensor box measurements is made.

4.2 Comparison between stationary sensor box and reference measurements

Outdoor performance tests were carried out at a station from the Flemish air quality monitoring network. The station (Borgerhout, 42R801, see www.ircel.be) is situated at a traffic location along a double lane main street with an average daily traffic volume of 43,381 vehicles (42,961 cars and 420 heavy duty vehicles, data from the Traffic Centre Flanders). We used 4 sensor boxes for the outdoor performance tests from October 2012 until April 2013. By placing the sensor boxes at an official monitoring station we gained the advantage of having reference data for several pollutants (CO, NO, NO_2, O_3 and Black Carbon (BC)) albeit at a coarser temporal resolution of 30 minutes. The average gas (CO, NO, NO_2 and O_3) concentration and BC concentration during the outdoor tests are given in 4.1. A cross-correlation analysis was performed to compare the 30 minute averaged sensor data with the reference data for several pollutants (Table 4.2). Correlation between the reference data themselves is given in Table 4.3 for comparison.

Low to moderate (0.25–0.60) correlations were found between the CO sensor measurements and the CO reference data. The alphasense CO-BF and e2v MiCS-5525 CO had the highest correlations (0.52 and 0.60, respecively) of the four CO sensors, although the alphasense CO-BF showed

Table 4.1: Average concentration and standard deviation of the CO, NO, NO_2, O_3 and BC by the reference monitors during the outdoor testing period.

	mean	stdev.
CO	0.32 ppm	0.13 ppm
NO	32.27 ppb	35.34 ppb
NO2	26.56 ppb	11.50 ppb
O3	9.87 ppb	10.00 ppb
BC	4.03 μg/m^3	2.73 μg/m^3

Table 4.2: Cross-correlation between sensor measurements and reference gas measurements. Averages of 4 sensor boxes are shown together with the standard deviations between brackets.

	Reference monitors						
Sensors	CO	NO	NO_2	O_3	BC	temp	rel. hum.
Alphasense CO-BF	0.52 (0.16)	0.41 (0.11)	0.34 (0.11)	-0.32 (0.14)	0.35 (0.13)	-0.81 (0.11)	0 (0.16)
e2v MiCS-5521 CO	0.31 (0.04)	0.32 (0.04)	0.34 (0.04)	-0.09 (0.11)	0.41 (0.02)	0.89 (0.06)	-0.14 (0.06)
e2v MiCS-5525 CO	0.60 (0.02)	0.51 (0.05)	0.56 (0.05)	-0.71 (0.05)	0.55 (0.06)	0.50 (0.06)	0.25 (0.03)
Figaro TGS 2201 CO	0.25 (0.02)	0.32 (0.01)	0.17 (0.00)	-0.48 (0.01)	0.38 (0.01)	0.45 (0.03)	0.46 (0.07)
Figaro TGS 2201 NOx	-0.78 (0.01)	-0.40 (0.06)	-0.24 (0.05)	0.47 (0.05)	-0.47 (0.06)	-0.40 (0.04)	-0.21 (0.03)
e2v MiCS-2710 NO_2	-0.58 (0.02)	-0.40 (0.06)	-0.31 (0.08)	0.64 (0.07)	-0.49 (0.06)	-0.40 (0.07)	-0.27 (0.02)
e2v MiCS-2610 O_3	-0.67 (0.06)	-0.56 (0.02)	-0.55 (0.05)	0.83 (0.07)	-0.62 (0.03)	-0.18 (0.15)	-0.12 (0.19)
Applied Sensors AS-MLV VOC	0.63 (0.02)	0.43 (0.17)	0.53 (0.15)	-0.44 (0.26)	0.45 (0.19)	0.23(0.22)	0.14 (0.10)

Table 4.3: Cross-correlation between reference gas measurements and meteorological data measured inside the sensor box (average of 4 sensor boxes).

	Reference monitors					sensor box	
	CO	NO	NO_2	O_3	BC	temp ($^{\circ}$C)	% RH
CO	1.00	0.77	0.62	-0.55	0.83	-0.07	-0.09
NO	0.77	1.00	0.76	-0.51	0.89	0.13	-0.06
NO_2	0.62	0.76	1.00	-0.53	0.81	0.11	-0.24
O_3	-0.55	-0.51	-0.53	1.00	-0.54	-0.12	-0.22
BC	0.83	0.89	0.81	-0.54	1.00	0.23	-0.08

* streetside

** backyard (30 m from street)

significant variability between sensors. The non-CO sensors showed higher correlations with the reference CO measurements. By sharing the same pollution sources in the urban environment, i. e. traffic exhaust, it is logical that monitoring signals of different traffic pollutants show high correlations. Correlations between reference gas measurements show that the correlation between CO and other air pollutants are in the same range as observed based on the sensor box measurements (Table 4.3). The high correlations between the non-CO sensors and the reference CO data is therefore not uncommon. The Figaro TGS 2201 NOx and e2v MiCS-2710 NO_2 sensors showed a moderate negative correlation with NO and NO_2, probably by the interference of ozone. Correlations of the NOx sensors are higher with CO and O_3, although these values stayed within the correlation range that was observed for the reference measurements. It is not proven by this experiment that the high correlation with CO and O_3 is due to selectivity problems of the sensor. The e2v MiCS-2610 O_3 showed a high correlation with the reference O_3 measurements. Also the variability between the sensor boxes was limited. Correlations with other gases are negative, which is in line with the physical reality (see reference measurements in Table 4.3). The Applied Sensors AS-MLV VOC sensor shows the highest correlations with CO and NO_2, reference VOC measurements are lacking.

The correlation with BC ranges between 0.35 and -0.62. These correlations are in the same moderate range as the correlations that are found between the sensors and the reference measurements of the respective gases. The reference data show higher correlations between gases and black carbon (>0.8 for CO, NO and NO_2, see Table 4.3).

Important conclusions of these field experiments with respect to further developments and applications of the sensor box are:

(i) correlations between sensor and reference measurements are low to moderate for most of the senors, for e2v MiCS-2610 O_3 sensor the correlation is high with the reference O_3 measurements;

(ii) moderate correlations between the sensor measurements and the BC measurements are observed. For the reference data these correlations are higher;

(iii) from laboratory tests we observed that response times of the sensors are in the minute range rather than in the second range (see D1.1). The slower response of the sensors has

its impact on the mobile use of the sensor box.

4.3 Evaluation of mobile measurements with the sensor box

4.3.1 Spatial patterns in the sensor data

The final case study in Antwerp was different from the case studies in other cities in the sense that participants were equipped with additional monitoring equipment (i. e. the micro-aethalometer). Participants were asked to switch on the micro-aethalometers as soon as they went outdoor to perform air quality monitoring. This was used for data validation: only the sensor box measurements that were collected simultaneously with micro-aethalometer data were used in this analysis.

Sensor rescaling model

Sensor data are not interpretable in an absolute way without calibration. However, sensor values may be used in a relative way to compare measurements from the same sensor at different locations or times. Furthermore, simultaneous measurements can be used for sensor rescaling so that sensor measurements from different sensor boxes become comparable. Here the objective was to create such relative maps of the monitoring area in Antwerp from sensor data, and to investigate the patterns that appear in relation to the local traffic and street characteristics.
The rescaling of the sensor box measurements is based on a period (several days) of simultaneous measurements at a traffic location (see calibration setup, see Chapter 3). The rescaling of the sensor data is done by assigning a 'reference sensor box'. The reference sensor box was picked randomly from the sensor boxes that were used in Antwerp. A linear regression (with intercept) of sensor measurements from the reference sensor box in function of sensor measurements of the other boxes was performed to set the linear rescaling parameters (sensor box and sensor specific). These parameters were then used to for the rescaling of new (mobile) data. The pseuso-code of the rescaling of the sensors is given below:

(1) define a reference sensor box;

(2) for *sb* = 1 : all other sensor boxes do (iteration over all the sensor boxes)

  (3) for *s* = 1 : 10 (iteration over the 8 gas sensors, plus temperature and relative humidity) do

    (i) get sensor *s* values from the reference sensor box;

    (ii) get sensor *s* values from the sensor box *sb*;

    (iii) construct a linear model of the sensor measurements from the reference sensor box in function of the sensor measurements of the other sensor box;

    (iv) save regression coefficients;

    (v) save regression statistics (R^2);

  (4) end;

(5) end;

The r-square statistic of the linear model between a given sensor and the same sensor in the reference box was used for data validation purposes. An overview of the r-square values is given in Table 4.4. SB016 was selected as the reference sensor box. The r-square values range between values as low as 0.02 to 1. For the CO_alpha sensor, a clear difference is observed between the 'older' sensor boxes (SB02 $\rightarrow$ SB11) and the additional sensor boxes that were assembled for the

Table 4.4: R-square statistic of the linear model between sensors. Values below 0.5 are in italics. Sensor box SB16 was selected as the reference sensor box.

	SB02	SB03	SB04	SB05	SB06	SB07	SB08	SB09	SB11	SB16	SB17	SB18	SB19
CO_alpha	*0.18*	*0.05*	0.56	*0.07*	*0.05*	*0.02*	*0.37*	*0.05*	*0.03*	1	0.88	0.89	0.88
CO_mics_1	0.88	0.91	*0.28*	0.89	0.92	0.95	0.92	0.88	0.95	1	0.94	0.96	0.97
CO_mics_2	0.73	0.8	0.57	0.85	0.77	0.5	0.86	*0.25*	0.86	1	0.97	0.99	0.97
CO_figaro	0.97	0.97	0.64	0.98	0.98	0.97	0.89	0.93	0.96	1	0.94	1	0.98
NO2_mics	0.92	0.81	*0.1*	0.84	0.85	0.83	0.93	0.81	0.84	1	0.98	1	0.98
NOx_figaro	0.76	0.78	*0.02*	0.8	0.8	0.79	0.95	0.74	0.83	1	0.86	0.85	0.9
O3_mics	*0*	0.52	0.95	0.88	0.95	0.95	0.99	0.77	0.98	1	0.98	0.99	1
VOC_as	*0.4*	*0.47*	0.88	*0.46*	0.44	0.65	0.93	0.7	0.98	1	0.63	0.99	0.93
temp	0.93	0.97	0.97	0.97	0.97	0.99	0.89	0.93	0.98	1	0.96	1	1
rh	0.96	0.98	0.97	0.97	0.95	0.96	0.8	0.92	0.98	1	0.97	0.99	0.99

Airprobe Challenge (SB16, SB17, SB18 and SB19), with particularly low correlations for the older sensor boxes compared to the new ones, probably due to ageing of the sensors. The diurnal CO pattern with lower concentration during nighttime is not observed very well in the sensor values from the newer sensor boxes. Between other sensors, sensor correlations are mostly moderate to high, with a few exceptions here and there. The r-square statistic is used in the data validation process of this analysis.

Rescaling of mobile data

As indicated in the previous paragraph, the rescaling models were quite satisfactory, except for the CO_alpha sensors and some other specific sensors in some sensor boxes. The r-square of the rescaling model is used as an extra data validation step for this analysis. A 0.5 threshold had to be exceeded to include the data in the analysis. For CO_alpha this leads to a data reduction of 70% of the mobile data. Also for the sensor CO_mics_2 the data reduction was high (-59%). For the other sensors, the data reduction is smaller (between 0 and 13%). The histogram of the rescaled sensor values over the entire monitoring period is given in Figure 4.1. A clear difference is observed between the sensors. For some sensors, the range of measurements is narrow (e. g. CO_mics_1), for others the range of measurements is quite broad (e. g. CO_mics_2).

Mapping of the rescaled measurements

As described in D4.1 mobile measurements carry information about the air quality at a specific location, both in space and time. Aggregation schemes are used to merge this information into interpretable maps. Data aggregation involves the merging of air quality data in a predefined time and space window, usually by a common operation such as averaging. Here we performed weighing averaging to the validated data from the AirProbe Challenge. A point database was constructed with geo-referenced points, 10 m apart, on all the streets of Antwerp city center where the gaming area is located. The mobile measurements were then attributed to these points based on distance. Points further away than 30 m from a point were discarded, other points were weighted based on their distance to the fixed point and averaged. Participants to the AirProbe Challenge were encouraged to concentrate their monitoring efforts to confined areas to have several repeated measurements throughout the duration of the monitoring campaign. As a consequence, the aggregation takes place over time-distributed sensor box measurements, and are thus more representative than the point observations made by the sensor boxes. However, because some participants only collected data at specific moments of the day, e. g. at peak hours, and because participants collected data on different locations on different moments, the results can be biased for some locations and comparison between locations can be distorted.

The first map shown here plots the number of measurements per aggregation point in the gaming area in Antwerp (Figure 4.2). In general, a high number of measurements in Antwerp lacked GPS data, which is very unfortunate (see D4.2). The measurement density was the highest in

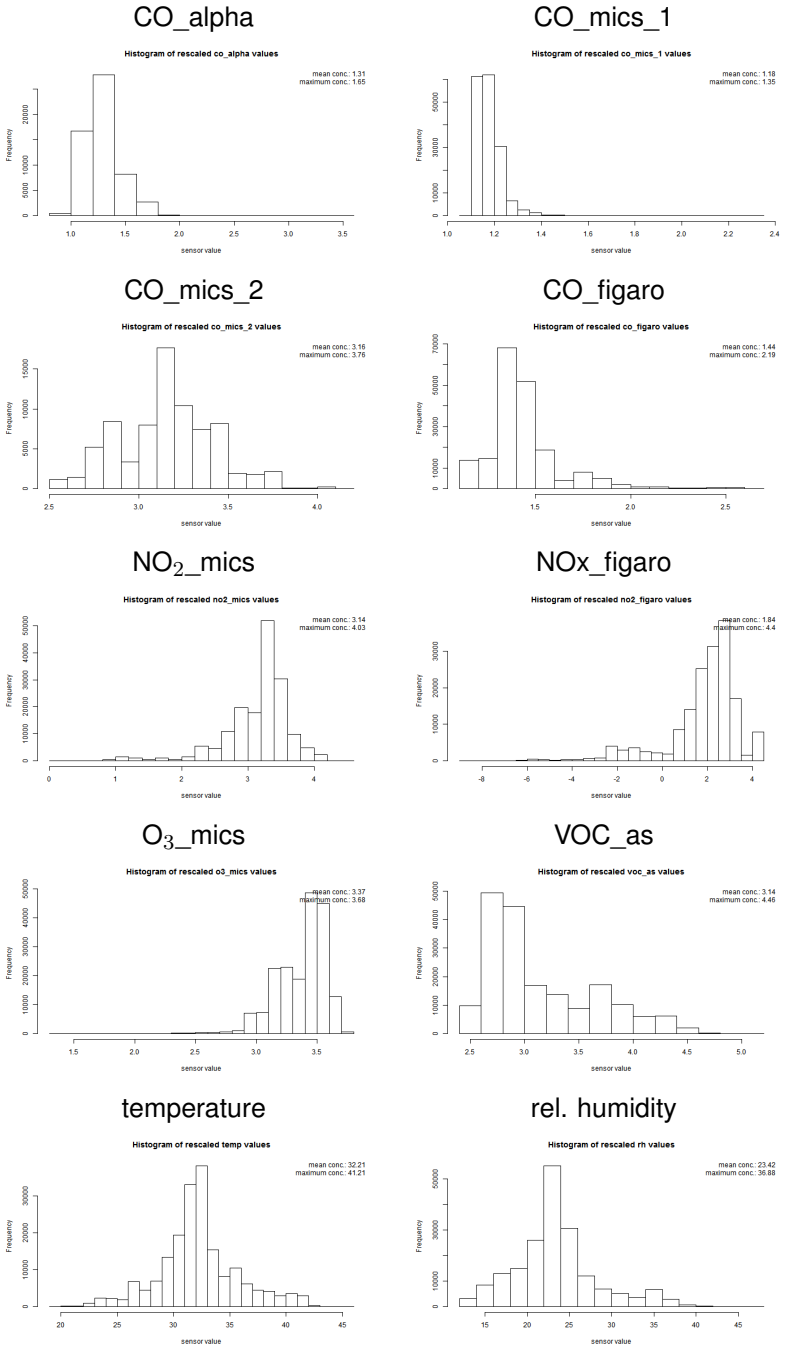

Figure 4.1: Histograms of the rescaled sensor values.

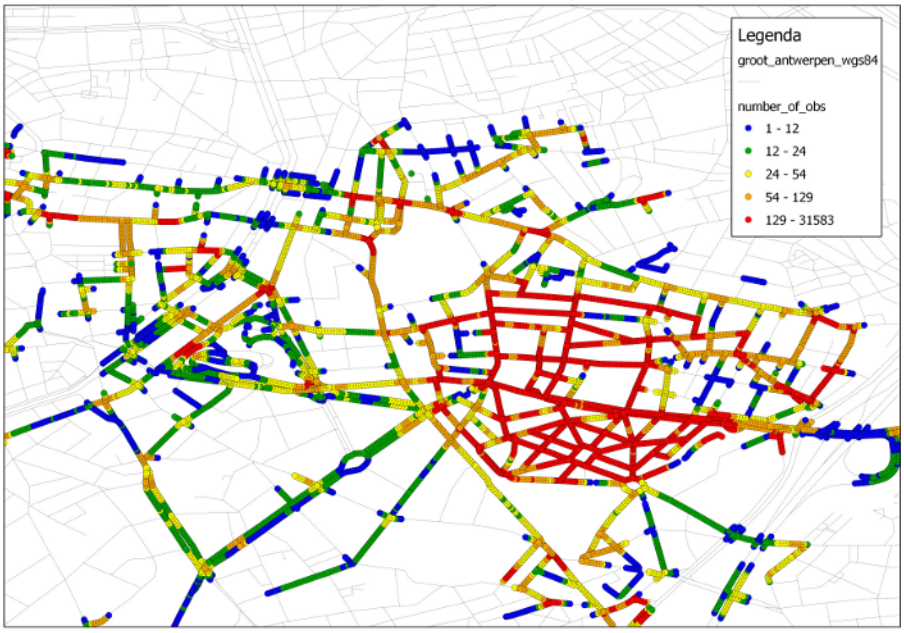

Figure 4.2: Map of the number of validated measurements attributed to fixed points in the monitoring area.

the neighborhoods North and South of Plantin and Moretuslei. Approximately 40% of the points shown on this map were based on more than 54 repeated measurements, 20% of the points were based on more than 130 repeated measurements. 20% of the point with the lowest measurement density were based on less than 14 repeated measurements.

Some contrasting locations (roads, streets and squares) are highlighted on the map of the monitoring area (Figure 4.2). A two-lane city entrance road (Plantin & Moretuslei) is situated in the center of the monitoring area. Traffic intensities are quite high on this road, also for heavy duty vehicles. Other busy streets in the area are Turnhoutsebaan at the North side and Montensstraat with frequent bus traffic. The residential area around Dageraadplaats has low traffic intensity (Dageraadplaats is a traffic free square in the center of this area). An urban green area is also included (Stadspark). The urban green is surrounded by busy two-lane roads. The Franklin Rooseveltplaats is another square at the North side of the monitoring area. Many bus stops are located at both sides of this square. Finally, a pedestrian street in mainly commercial area is also included.

For each sensor separately, a map of aggregated sensor values was constructed (Fig. 4.4). The coloring of the map is based on percentile values. The objective of this analysis was to assess the sensor patterns in the area. Could similar patterns be observed between the different sensors? A visual inspection of the sensor maps at specific locations (A–G, see Fig. 4.3) indicated that:

(A) increased concentrations were observed by the CO_mics_1, CO_figaro and the VOC_as sensors;

(B) low concentrations in the urban green were observed by the VOC_as sensor, other sensors had quite high measurement values, especially at the boundary of the park;

(C) high concentrations were observed by the CO_figaro and O3_mics sensors;

(D) the VOC_as and CO_mics_1 sensors obtained low values in the low-traffic area around Dageraadplaats;

(E) the CO_mics_1 and O3_mics sensors measured quite high values here, other sensors showed low values;

Figure 4.3: Overview of the monitoring area in Antwerp from Google Earth. Some streets and squares are highlighted: (A) Plantin & Moretuslei a two lane road with high traffic density, (B) Stadspark urban green, (C) Turnhoutsebaan street canyon, (D) Dageraadplaats and low-traffic surrounding streets, (E) Montensstraat street with frequent bus traffic, (F) Franklin Rooseveltplaats square with bus stops, and (G) Meir pedestrian area.

(F) the CO_mics_1-2, O3_mics and VOC_as sensors measured quite high concentrations at the Franklin Rooseveltplaats;

(G) only the O3_mics sensor showed low values in the pedestrian commercial area, other sensors showed high values.

In conclusion, a high variability is observed between the maps of the different sensors. At specific locations where low concentrations could be expected based on the traffic situation (B, D and G), high sensor responses were observed for several sensors. At locations with high expected pollution levels (A, C, E, F) several sensors did not have an increased response.

A map comparison was also performed using the Kappa statistic of agreement [Cohen, 1960]. The calculation of the Kappa statistic was as follows: (i) spatially aggregated sensor data from two sensors were merged based on geographical point coordinates, (ii) sensor values were transformed to percentile classes (10% class width), (ii) Cohen's kappa was computed between these classes. A weighing was hereby introduced. Elements on the diagonal of the confusion matrix obtained a 0 weight, for off-diagonal element the weighing equaled the squared distance from the diagonal. The kappa statistic gives an indication on the agreement between the maps of different sensors. Higher kappa values indicate a better agreement (kappa value of 1 shows perfect agreement). The scale of Landis and Koch [1977] is often used for the interpretation of kappa values:

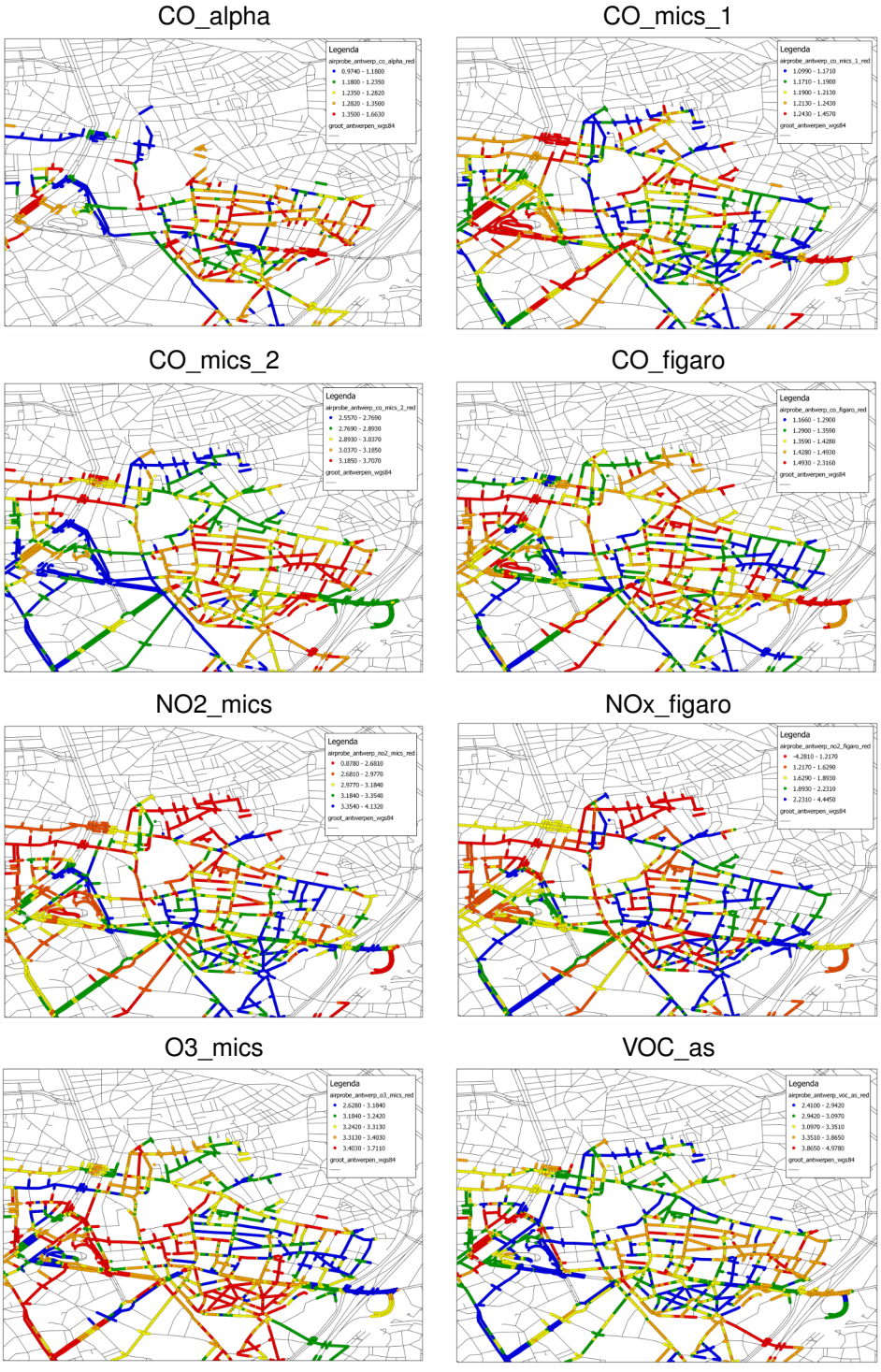

Figure 4.4: Maps of the aggregated and rescaled sensor box data for the different gas sensors. The colour scheme of the NO_2_mics and the NOx_figaro sensors is inverted to be in line with the colour scheme of the other sensors (inverted sensor signal for NO_2_mics and the NOx_figaro)

Table 4.5: Overview of kappa statistics of agreement.

	co_alpha	co_mics1	co_mics2	co_figaro	no2_mics	nox_figaro	o3_mics	voc_as	bc_obs	bc_mod
co_alpha	1	0.11	0.24	0.07	-0.04	0.23	0.07	0	-0.17	-0.08
co_mics1	0.11	1	0.11	-0.09	0.12	0.11	-0.25	0.25	0.17	0.16
co_mics2	0.24	0.11	1	0	-0.19	-0.11	0.04	0.28	-0.07	-0.09
co_figaro	0.07	-0.09	0	1	0.34	0.34	-0.13	-0.02	-0.02	0.06
no2_mics	-0.04	0.12	-0.19	0.34	1	0.45	-0.05	-0.04	0.08	0.2
nox_figaro	0.23	0.11	-0.11	0.34	0.45	1	-0.18	-0.01	-0.11	-0.11
o3_mics	0.07	-0.25	0.04	-0.13	-0.05	-0.18	1	-0.29	-0.11	0.07
voc_as	0	0.25	0.28	-0.02	-0.04	-0.01	-0.29	1	0.15	0.11
bc_obs	-0.17	0.17	-0.07	-0.02	0.08	-0.11	-0.11	0.15	1	0.38
bc_mod	-0.08	0.16	-0.09	0.06	0.2	-0.11	0.07	0.11	0.38	1

kappa	interpretation
< 0	Poor agreement
0.01 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Results are given in Table 4.5. All kappa values indicate poor, slight or fair agreement between the sensor maps. The best agreement between the maps of the CO sensors was found between the co_alphasense and the co_mics_2 sensors. The highest agreement was found between the maps of the no2_mics and nox_figaro sensors (kappa of 0.45).

Sensor maps were constructed based on rescaled sensor data. The rescaling of the data introduces a certain amount of additional uncertainty because the rescaling models were not perfect. However, rescaling models with r-squares less than 0.5 were discarded from this analysis. In conclusion, the gas sensor maps generally show low levels of agreement. The patterns that are observed are also sometimes different from the pattern that was expected from traffic and street characteristics. This can be explained by the low correlations of the different gas sensors with their target gases, and in some cases probably by the high response time.

4.4 Modelled versus observed spatial BC patterns

A benchmark black carbon map for Antwerp is generated using black carbon measurements with the micro-aethalometer (Figure 4.5). The resulting BC map is evaluated visually in relation to street characteristics, traffic conditions and knowledge of the spatial black carbon distribution in the area based on previous monitoring campaigns (e. g. [Peters et al., 2013]). The traffic intensity (e. g. presence of local black carbon sources) has a clear effect on the black carbon distribution in the area. At urban green, pedestrian en low traffic areas, the black carbon concentrations are considerably lower (generally below 4 μg/m^3) compared to the concentration near busy streets and roads. Especially the road section of Plantin and Moretuslei near the ring road has a high black carbon concentration, which has also been observed by previous monitoring campaigns in the area. This is probably caused by high traffic volumes and regular traffic jams at this ring-road access point.

A comparison is made with the map of the modelled black carbon (Fig. 4.6). The agreement between both maps is fair (*sensu* Landis and Koch [1977]) with a kappa coefficient of 0.38 (Table 4.5). These observations are made from a visual comparison between the observed and the modelled BC map:

(+) agreement between both maps

- there is a quite good agreement between both maps at the locations and streets from

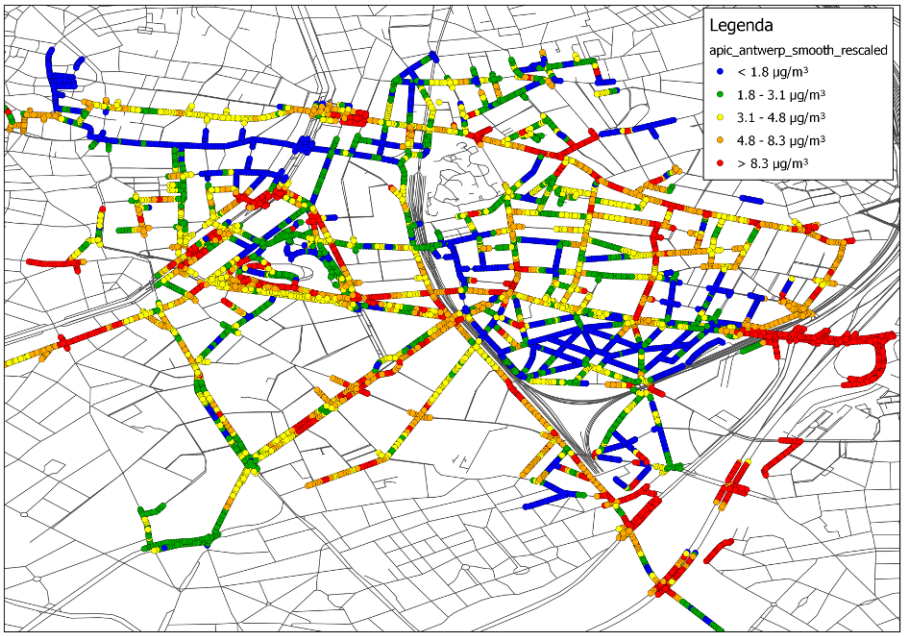

Figure 4.5: Map of the black carbon concentration in Antwerp based on mobile measurements with the micro-aethalometer during the AirProbe Challenge.

Fig 4.3: (i) the outer part of Plantin and Moretuslei (A) near the ring road shows the highest BC concentrations in both maps, (ii) the concentrations for Turnhoutsebaan and Carnotstraat (C) are moderately high on both maps; (iii) the streets with low traffic density near Dageraadplaats (D) show relatively low black carbon concentrations on both maps; moderate concentrations are observed and modelled for the Franklin Rooseveltplaats square (F); at the pedestrian commercial area (G) the measured and observed black carbon concentration is low.

- the agreement between the modelled BC map and the observed BC map is substantially higher than the agreement between the observed BC map and the individual sensor maps.

(-) disagreement between both maps

- at some streets and squares the black carbon concentrations differ substantially: (i) at the interior part of the urban green Stadspark (B) the observed BC concentration is low but the modelled BC concentration is high; (ii) at some of the roads at the Southern monitoring area the difference between observations and modelling results is high.
- although some of the broader patterns coincide between both maps, the within street variability is much larger for the observed BC map than for the modelled one. the within street variability can be substantial e. g. at crossroads etc. This is not always reflected in the modelled maps. The slower sensor response times compared to the fast response of the micro-aethaometers is causing this discrepancy.

So far the comparison between observed and estimated black carbon concentrations is made on a relative basis. Black carbon classes were delineated based on percentile statistics of both data series. In Fig. 4.7 two maps are shown for the gaming area in Antwerp. The first map shows the measured BC the second map the modelled BC. The colouring of the map is based on absolute black carbon concentrations (width of 2 μg/m^3). The variability for the modelled black carbon map is much lower than for the observed black carbon map mainly because the underestimation of the black carbon peak concentrations (see also Chapter 3).

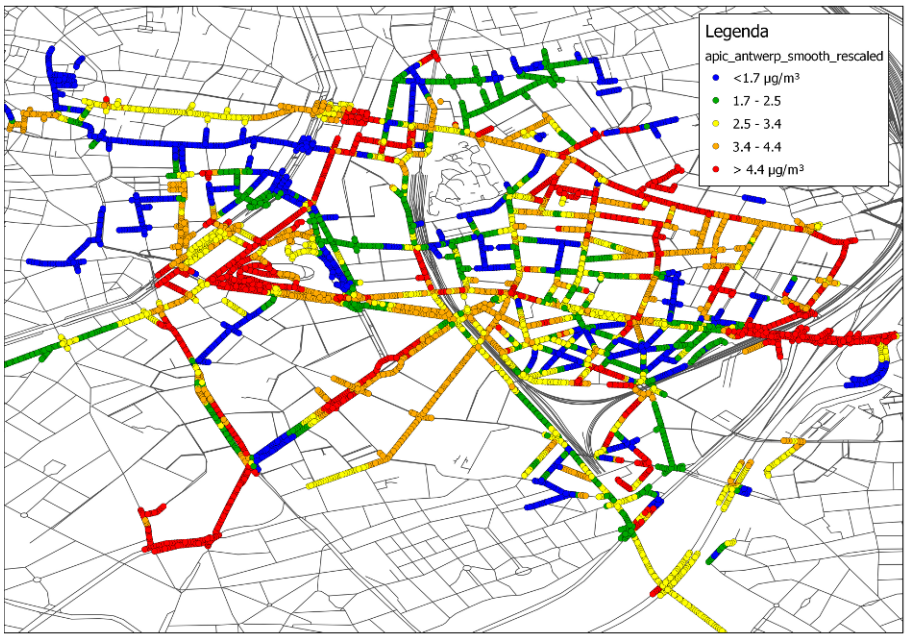

Figure 4.6: Map of the estimated black carbon concentration in Antwerp based on mobile measurements with the sensor box during the AirProbe Challenge.

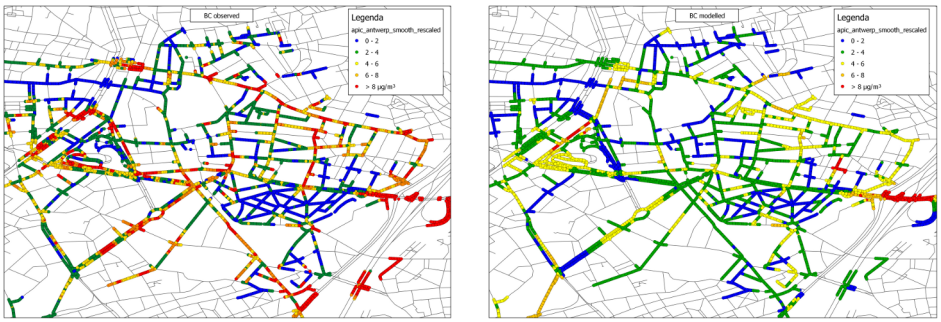

Figure 4.7: Map of the measured and modelled black carbon concentration in Antwerp based on mobile measurements with micro-aethalometer and sensor box during the AirProbe Challenge.

(a) measured BC (b) modelled BC

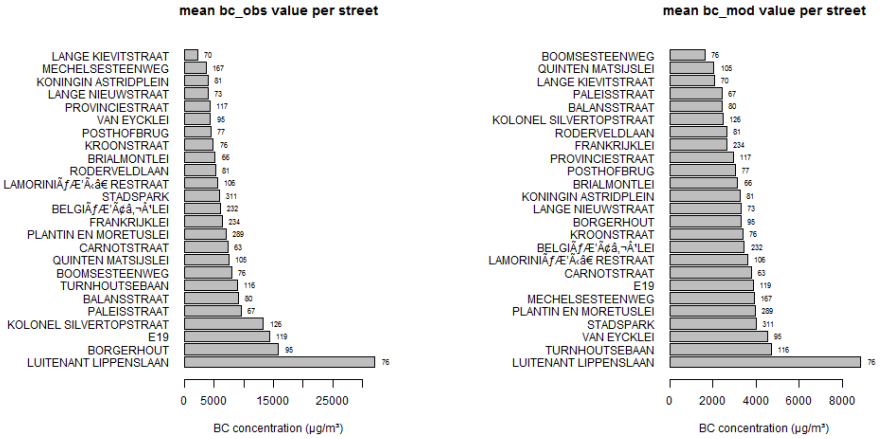

Figure 4.8: Histograms of the average black carbon concentration (measured (a) and modelled (b)) for the 25 streets with the highest number of measurements.

A comparison between the 25 most frequently monitored streets is given in Fig. 4.8. The modelled black carbon concentrations are much lower than the observed BC cocncentraions. The average modelled BC concentration exceeds 5 μg/m^3 at Luitenant Lippenslaan, whereas the average observed BC concentration exceeds 5 μg/m^3 in 18 streets and squares. The absolute BC concentrations differ substantially between the modelled and measured concentrations. The ranking of the streets, however, is similar for some streets (e. g. Luitenant Lippenslaan) and completely different for others (e. g. Van Eycklei).

Bibliography

MC Carotta, G Martinelli, L Crema, M Gallana, M Merli, G Ghiotti, and E Traversa. Array of thick film sensors for atmospheric pollutant monitoring. *Sensors and Actuators B: Chemical*, 68(1 - 3):1 – 8, 2000.

MC Carotta, G Martinelli, L Crema, C Malagu, M Merli, G Ghiotti, and E Traversa. Nanostructured thick-film gas sensors for atmospheric pollutant monitoring: quantitative analysis on field tests. *Sensors and Actuators B: Chemical*, 76(1 - 3):336 – 342, 2001.

MC Carotta, M Benetti, E Ferrari, A Giberti, C Malagu, M Nagliati, B Vendemiati, and G Martinelli. Basic interpretation of thick film gas sensors for atmospheric application. *Sensors and Actuators B: Chemical*, 126(2):672 – 677, 2007.

J Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37 – 46, 1960.

S De Vito, E Massera, M Piga, L Martinotto, and G Di Francia. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2):750 – 757, 2008.

S De Vito, M Piga, L Martinotto, and G Di Francia. Co, no2 and nox urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. *Sensors and Actuators B: Chemical*, 143(1):182 – 191, 2009.

EPA. Report to congress on black carbon - external peer review draft. Technical Report EPA-450/D-11-001, EPA, 2010.

GSW Hagler, TLB. Yelverton, R Vedantham, ADA. Hansen, and JR Turner. Post-processing method to reduce noise while preserving high time resolution in aethalometer real-time black carbon data. *Aerosol and Air Quality Research*, 11:539 – 546, 2011.

M Kamionka, P Breuil, and C Pijolat. Calibration of a multivariate gas sensing device for atmospheric pollution measurement. *Sensors and Actuators B: Chemical*, 118(1 - 2):323 – 327, 2006.

JR Landis and GG Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33:159 – 174, 1977.

Tom M Mitchell. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45, 1997.

J Peters, J Theunis, M Van Poppel, and P Berghmans. Monitoring pm10 and ultrafine particles in urban environments using mobile measurements. *Aerosol and Air Quality Research*, 13:509 – 522, 2013.

Wataru Tsujita, Akihito Yoshino, Hiroshi Ishida, and Toyosaka Moriizumi. Gas sensor network for air-pollution monitoring. *Sensors and Actuators B: Chemical*, 110(2):304 – 311, 2005.