



## Practical field calibration of electrochemical NO<sub>2</sub> sensors for urban air quality applications

Bas Mijling<sup>1</sup>, Qijun Jiang<sup>2</sup>, Dave de Jonge<sup>3</sup>, Stefano Bocconi<sup>4</sup>

<sup>1</sup>Royal Netherlands Meteorological Institute (KNMI), Postbus 201, 3730 AE, De Bilt, The Netherlands

5 <sup>2</sup>Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands

<sup>3</sup>Public Health Service of Amsterdam (GGD), Nieuwe Achtergracht 100, 1018 WT, Amsterdam, The Netherlands

<sup>4</sup>Waag Society, Nieuwmarkt 4, 1012 CR, Amsterdam, The Netherlands

Correspondence to: Bas Mijling ([mijling@knmi.nl](mailto:mijling@knmi.nl))

10 **Abstract.** In many urban areas the population is exposed to elevated levels of air pollution. However, air quality is usually only measured at a few locations. These measurements provide a general picture of the state of the air, but they are unable to monitor local differences. Since a few years new low-cost sensor technology is available, which has the potential to extend the official monitoring network significantly. These sensors, however, are still in an experimental stage and suffer from various technical issues which limit their applicability.

15 This study explores the added value of alternative air quality measurements, focusing on nitrogen dioxide (NO<sub>2</sub>) in Amsterdam, the Netherlands. 16 low-cost air quality sensor devices were built and distributed among volunteers living close to roads with high traffic volume for a two-month measurement campaign.

Careful calibration of individual sensors is essential to measure ambient concentrations of NO<sub>2</sub> significantly. Field calibration was done next to an air monitoring station during an 8-day period, resulting in  $R^2$  ranging from 0.3 to 0.7. The  
20 NO<sub>2</sub> accuracy can be improved by including temperature and humidity measurements from an additional low-cost sensor,  $R^2$  ranging from 0.6 to 0.9. Recalibration is crucial, as all sensors show significant signal drift after the two-month measurement campaign. The measurement series between the calibration periods can be corrected in hindsight by taking a weighted average of the calibration coefficients.

Validation against an independent air monitoring station shows good agreement. Using our approach, the standard deviation  
25 of a typical sensor device for NO<sub>2</sub> measurements was found to be 7  $\mu\text{g m}^{-3}$ . This shows that, if properly treated, low-cost sensors based on the current generations of electrochemical NO<sub>2</sub> sensors may provide useful complementary data on local air quality in an urban setting.



## 1 Introduction

Because air pollution is difficult to measure, instrumental and operational costs of official measurement stations are usually high. Air quality networks in cities, if present at all, are therefore usually sparse. Emerging low-cost sensor technology has the potential to extend the official monitoring network significantly, and improve our understanding of local urban air pollution. Miniaturized and affordable sensors enable citizens to measure their environment in more detail in space and time (Kumar et al., 2015). However, most sensors are still in an experimental stage and suffer from various technical issues which limit their applicability. The poor data quality is of concern to health authorities, scientists and citizens themselves. Before conclusions can be drawn from the measurements, comprehensive calibration and validation is essential (e.g. Lewis and Edwards, 2016; Lewis et al., 2016).

Several studies have been done to explore the performance of low-cost air quality sensors, e.g. Jiao et al., 2016, Duvall et al., 2016; Mead et al., 2013; Moltchanov et al., 2015. For NO<sub>2</sub> monitoring, mostly metal oxide and electrochemical sensors are used (Borrego et al., 2016; Spinelle et al., 2015b; Thompson, 2016). Typical ambient concentrations of NO<sub>2</sub> are at part-per-billion (ppb) level. The main problems encountered in NO<sub>2</sub> sensor evaluations in these real-world environments are low sensitivity, poor selectivity, low precision and accuracy, and drift. Especially metal oxide sensors are not very stable (Spinelle et al., 2015b; Thompson, 2016) and suffer from lower selectivity. Therefore, in this study, we opted for electrochemical sensors to measure NO<sub>2</sub>.

Mead et al. (2013) already noted the strong interference of ozone and other ambient factors in electrochemical NO<sub>2</sub> sensors. The performance can be increased significantly when adding additional measurements of e.g. temperature and humidity in a regression model or neural network, as shown by e.g. Piedrahita et al. (2014), Spinelle et al. (2015b), Masson et al. (2015). Coping with sensor degradation remains a serious issue. Some studies, such as Jiao et al. (2016), include an additional temporal term in their linear regression which improves the predicted NO<sub>2</sub> slightly.

The following sections will further explore the applicability of electrochemical NO<sub>2</sub> sensors for measurements of urban air quality, using a practical method for in-field calibration and regression modelling for assessment of accuracy and sensor degradation.

## 2 The Urban AirQ project

The Urban AirQ project explores the added value of alternative air quality measurements in the city. It focusses on a 2×1 km<sup>2</sup> area around Valkenburgerstraat, a primary road in the East-central part of Amsterdam, see Figure 1. Its dense traffic causes regular exceedances of the European annual limit value for nitrogen dioxide (40 µg m<sup>-3</sup>).

Two town hall meetings were organized in which residents of this area were invited to raise their concerns about air pollution in their neighborhood and to formulate related research questions. Topics included the relation between traffic density and air pollution, the difference between a main road and a side street, the front side of an apartment compared to its backside, the influence of apartment height, and the influence of cut-through traffic at nighttime. The residents were invited



to participate in finding answers to their questions by measuring their outdoor air quality with 16 experimental low-cost sensor devices, built for this purpose by Waag Society.

Measurements were done from June to August 2016. Beforehand, the sensor devices were calibrated using side-by-side measurements next to an official air quality measurement station. With a second calibration period after the campaign,  
5 individual sensor drift was assessed and compensated in hindsight.

### 3 NO<sub>2</sub> sensor devices

The concept of the Urban AirQ sensor is building a device with low-cost electronic components which is easy to operate, so citizens can do their own air quality measurements. It builds on the basic design described by Jiang et al. (2016), having an improved power supply, weather resistant housing, WiFi connectivity, and additional sensors for temperature, relative  
10 humidity, and particulate matter. The sensor development is part of an open hardware project; detailed technical information can be found at <https://github.com/waagsociety/making-sensor>.

Central is the microcontroller board (Arduino UNO) which handles the reading of the sensors and sends the data to the WiFi module, see Figure 2.

For NO<sub>2</sub> measurements, an amperometric electrochemical cell is used from Alphasense Ltd (Essex, United Kingdom). The  
15 cell contains four electrodes. The target gas, NO<sub>2</sub>, diffuses through a membrane where it is chemically reduced at the Working Electrode, generating a current signal. This electric current is balanced by a opposite current from the Counter Electrode. The Reference Electrode sets the operating potential of the Working electrode. The sensor also includes an Auxiliary Electrode, which is used to compensate for baseline changes in the sensor. To get full sensor performance, low noise interface electronic is necessary. An individual sensor board, also provided by Alphasense, is used to guarantee a low  
20 noise environment and to optimize the sensor resolution at low ppb levels. The sensor signal is read by a 16-bit analog to digital (A/D) converter (ADS1115). 14 sensor devices contain model NO2-B43F for NO<sub>2</sub> measurements, the other two use model NO2-B42F.

All devices are equipped with a DHT22 sensor from Aosong Electronics measuring temperature and relative humidity (RH).  
12 of the 16 sensor boxes are also equipped with a Shinyei PPD42NS sensor in order to measure particulate matter optically.  
25 The present paper, however, will focus only on the assessment of the NO<sub>2</sub> measurements.

#### 3.1 Averaging and filtering

Raw sensor measurements are stored in a central database on a one minute base. However, the calibration analysis is based on hourly averages to enable direct comparison between the ground truth (also provided as hourly values), and to improve the signal to noise ratio.

30 The NO<sub>2</sub> sensor measurements are done at the Working Electrode ( $S_{WE}$ ) and the Auxiliary Electrode ( $S_{AE}$ ). They are provided as counts from the A/D converter. Sensor readings of temperature and RH are converted according to the indication



of the manufacturer to degrees Celsius and percentages respectively. Raw, hourly averaged, sensor data is shown in Figure 3. Careful filtering is needed before the data can be further processed. We have applied the following rules:

- Raw, minute-based,  $S_{WE}$  and  $S_{AE}$  measurements outside a  $\pm 10\%$  range of their mean value during the entire measuring period are considered outliers. This affects 0.33% of all measurements.
- 5 • All readings at sensor temperatures above  $30^{\circ}\text{C}$  are discarded to avoid non-linear temperature dependence of the electrochemical  $\text{NO}_2$  sensor (see Sect. 4.4). This affects 4.53% of the measurements during the entire period.
- At least 20 valid minute-based measurements are required to calculate a representative hourly mean.

During the first calibration period, the sensors were measuring 79% of the time on average. After applying the criteria above, this resulted in 70% valid hourly measurements. During the measurement campaign, the sensors produced 79% valid hourly  
10 measurements on average, with the uptime dropping to 50% in places where sensors experienced connectivity problems due to limited range of the participant's WiFi network.

### 3.2 Calibration periods

Calibration of the sensors devices have been done by placing the 16 sensors side by side on the rooftop of the air quality station at Vondelpark, operated by the Public Health Service of Amsterdam (GGD). This station is classified as a city  
15 background station. It measures nitrogen dioxide, nitrogen monoxide (NO), ozone ( $\text{O}_3$ ), particulate matter ( $\text{PM}_{10}$ ,  $\text{PM}_{2.5}$ , particle number and size distribution), black carbon, and carbon monoxide (CO). For NO and  $\text{NO}_2$  measurements, GGD alternates Teledyne API 200E and Thermo Electron 42I NO/ $\text{NO}_x$  analysers, both based on chemiluminescence. The validated measurements used in this study are considered to be the ground truth. The calibration period spanned several days  
20 to be able to test the sensors under a wide range of ambient conditions. To assess the stability of the calibration, the sensors were brought back after the two-month measurement campaign to the calibration facility for a second calibration period. The Urban AirQ campaign consisted therefore of three phases.

The first field calibration period at GGD Vondelpark station started at 2 June 2016, 00h LT (local time), and ended at 10 June 2016, 10h (8.5 days; 204 hours). Due to connectivity problems sensor data was missing between 4 June 19h and 6 June 9h.

25 During the following citizen campaign, 15 sensors were distributed among the participants. One sensor (55303) was kept at the Vondelpark station as a reference. The first sensor was installed and connected at 13 June 2016, 18h, the last sensor connected at 17 June 2016, 17h. At 15 August 2016, 9h, the first sensor was disconnected, at 16 August 2016, 18h, the last sensor was disconnected.

The second field calibration period at GGD Vondelpark station started at 18 August 2016, 15h, and ended at 29 August  
30 2016, 00h (10.4 days; 249 hours). Due to connectivity problems sensor data was missing between 26 August 12h and 27 August 11h.

Figure 4 shows the distribution of temperature, relative humidity,  $\text{NO}_2$ , and  $\text{O}_3$  during the different periods. The calibration periods are characterized by higher temperatures and ozone levels than the campaign period. The range of hourly  $\text{NO}_2$



concentrations at the Vondelpark station in the calibration periods is larger than in the campaign, reaching more frequently higher  $\text{NO}_2$  values. During the campaign the sensor are more closely located to the GGD station at Oude Schans.  $\text{NO}_2$  values measured here are generally a few  $\mu\text{g m}^{-3}$  higher than at Vondelpark. The Oude Schans site does not measure ozone.

#### 4 $\text{NO}_2$ calibration

- 5 Electrochemical sensors such as the Alphasense  $\text{NO}_2$ -B series, are known to be sensitive to interfering species and ambient factors. Especially ozone, temperature, and relative humidity influence the sensor reading (see e.g. Spinelle et al., 2015a).

##### 4.1 Explaining the $\text{NO}_2$ sensor signal

To understand better the behavior of the  $\text{NO}_2$  sensor, we study its sensitivity to different ambient factors. We use the first calibration period to test the correlation of the measured  $S_{\text{WE}}$  and  $S_{\text{AE}}$  signal with  $\text{NO}_2$ , ozone, temperature and humidity by  
10 making a best fit through the hourly time series, e.g.

$$S_{\text{WE}}(t) = c_0 + c_1 \text{NO}_2(t) \quad (1)$$

Temperature and RH were not available from the obtained GGD station data. Instead of taking from ambient air measurements, we take temperature and RH from the average readings from the DHT22 sensors, as these reflect better the internal sensor conditions.

Figure 5 shows scatter plots for an average performing sensor and the  $R^2$ , the coefficient of determination. The measured  $S_{\text{WE}}$  signal can be explained by ambient  $\text{NO}_2$  ( $R^2=0.20$ ), but better by its anti-correlation with ozone ( $R^2=0.49$ ). Temperature alone is an even better predictor for the sensor signal ( $R^2=0.73$ ), probably because of direct temperature dependence of the sensor, and indirect dependence (temperature being a reasonable proxy for both  $\text{NO}_2$  and  $\text{O}_3$  concentrations). Also the correlation with humidity is very strong ( $R^2=0.73$ ). The measured  $S_{\text{WE}}$  signal can best be explained as a linear combination of  $\text{NO}_2$ ,  $\text{O}_3$ , T, and RH together, resulting in a correlation of 0.98 ( $R^2=0.96$ ).

- 20 The  $S_{\text{AE}}$  signal is practically insensitive to  $\text{NO}_2$ . This suggests that a combination of  $S_{\text{WE}}$  and  $S_{\text{AE}}$  is more sensitive to  $\text{NO}_2$  and less to the other interfering factors, as intended by the manufacturer.

##### 4.2 $\text{NO}_2$ calibration models

The sensor manufacturer suggests to correct both Working Electrode and Auxiliary Electrode for a zero-offset with  $S_{\text{WE},0}$  and  $S_{\text{AE},0}$  respectively. Then a sensitivity constant  $s$  is applied to convert from mV to ppb  $\text{NO}_2$ :

$$\text{NO}_2[\text{ppb}] = \frac{(S_{\text{WE}} - S_{\text{WE},0}) - (S_{\text{AE}} - S_{\text{AE},0})}{s} \quad (2)$$

- 25 In practice, the factory-supplied constants  $S_{\text{WE},0}$ ,  $S_{\text{AE},0}$ , and  $s$  do not result in realistic values of  $\text{NO}_2$ . As an alternative, we propose a linear combination of signal  $S_{\text{WE}}$  and  $S_{\text{AE}}$



$$\text{NO}_2[\mu\text{g m}^{-3}] = c_0 + c_1 S_{\text{WE}} + c_2 S_{\text{AE}} \quad (3)$$

with the coefficients to be determined with data from the calibration period using ordinary least squares (OLS). Table 1 shows the fit results and the corresponding correlation with true  $\text{NO}_2$  signal. As can be seen, within the batch of sensors there is a large variability of direct sensitivity to ambient  $\text{NO}_2$ .

During the calibration period, hourly ozone values (also taken from the Vondelpark station) happened to be a good proxy for the ambient  $\text{NO}_2$  concentration:  $\text{NO}_2(t) = 44.6 - 0.40 \cdot \text{O}_3(t)$  in  $[\mu\text{g m}^{-3}]$ , with  $R^2$  of 0.49 (see Figure 6).

When compared with Table 1, one can see that direct sensor readings from a fair part of the sensors cannot outperform this result. To improve the results we use additional measurements and their statistical relation to  $\text{NO}_2$ . We fit different calibration models with multiple linear regression (using OLS). The calibration models which were tested are listed in Table 2.

10 Temperature and RH are taken from the DHT22 sensor. There is no need to calibrate the individual T and RH sensor values beforehand; the calibration coefficients for  $\text{NO}_2$  are determined for the specific set of all sensors in the box. However, this means that if an individual sensor is replaced, new calibration parameters for the sensor box have to be derived.

### 4.3 Calibration results

A complete overview of fit results for all models can be found in the supplement. The sign of the calibration parameters can be easily understood. As the electrochemical sensor loses sensitivity at higher temperatures, coefficients  $c_3$  are positive to compensate for this effect. The additional sensor response due to cross-sensitivity with ozone is compensated by negative values for  $c_5$ .

From the fit results we see that Model C (including RH) performs better than Model A, but model B (including T) outperforms model C. Model D (including both RH and T) only marginally improves the results of Model B. This can be understood from the strong sensor dependence on temperature directly, and indirectly on temperature as a proxy for ozone. The better performance of model C with respect to model A can be explained by RH being a reasonably proxy for temperature. Note that measuring RH is essential for guarding the data quality of electrochemical sensors, as these sensors are very sensitive to *sudden changes* in RH, see e.g. AAN (2013) and Pang et al. (2016).

The best calibration results (i.e.  $R^2$  values closer to 1) are obtained by including ozone (Model E). The ozone values were obtained from the GGD station, as the sensor devices do not measure ozone.

As local ozone measurements were only available during the calibration periods, we used Model D for the Urban AirQ campaign, i.e. generating an  $\text{NO}_2$  value based on a linear combination of  $S_{\text{WE}}$ ,  $S_{\text{AE}}$ , T, and RH. The regression analysis of Model D and correlation with the  $\text{NO}_2$  ground truth can be found in Table 3.

The two worst performing sensor boxes (14560051 and 1184206) contain the older  $\text{NO}_2$ -B42F sensor. It is not clear if their poor performance can be attributed to the different sensor model, or to their longer operating time (both sensors have been used in previous experiments for more than a year). Again, one can see that even within the same batch of sensors, there is a



significant spread in performance, around a median value for  $R^2$  of 0.83. Figure 7 shows the results for the different calibration models for an average performing sensor. The time series in Figure 7(b) show clearly how the performance of a typical sensor device improves when temperature and humidity are included in the calibration analysis.

#### 4.4 Dependency on temperature

5 Calibrated, but uncorrected, data show occasionally strong negative values, see Figure 8 below. These negative peaks coincide with internal sensor temperatures exceeding 30 °C. This behavior can be explained from the dependency of the electrochemical sensor on temperature becoming non-linear, see Figure 8(b). In this regime, the response of the sensor cannot be described well with our multilinear regression approach. As temperatures during the measurement period only rose occasionally above 30 °C, we decided to filter these measurements out.

#### 10 4.5 Startup time

When the sensors are switched on after an unused period they need time to stabilize. Figure 9 give some examples of 4 sensors which are switched on at their campaign sites after being offline for a couple of days. The startup-effect is translated by the calibration model as a strong positive NO<sub>2</sub> peak. After 4 hours most sensors are sufficiently stabilized. Note that this startup effect should not be confused with the response time, which is determined to be less than 2 minutes in Mead et al.  
15 (2013) and Spinelle et al. (2015a).

#### 4.6 Sensor drift, aging, and uncertainty estimation

Almost all electrochemical sensors have some degree of drift because of aging and poisoning (Di Carlo et al., 2011; Hierlemann and Gutierrez-Osuna, 2008). This becomes a serious complication when the drift is in the order of the strength of the signal of interest. The idea of keeping sensor 55303 next to the reference station during the whole campaign was to  
20 study sensor degradation in more detail. Unfortunately, the sensor was removed temporarily from 10 to 14 July for service, which introduced a sudden and unexplained offset in measurements. By introducing a second calibration period after the measurement campaign, we have another possibility to assess the stability of the sensors, and calibrate the measurements in hindsight. All sensors were brought back to the GGD station at the Vondelpark. In Figure 10, the sensor signals (calibrated with coefficients from the first calibration period) are compared to the official station measurements. As can be seen in  
25 Figure 10(b), most sensors have been drifting in the intermediate two-month period. Note that part of the drift could also be partly related to the aging of the DHT22 temperature and RH sensor.

We describe this degradation effect as a bias  $b$  between the mean of the hourly estimated NO<sub>2</sub> values  $\hat{x}_i$  and the mean of the hourly true NO<sub>2</sub>  $x_i$  during the calibration period:



$$b = \frac{1}{N} \sum_{i=1}^N \hat{x}_i - \frac{1}{N} \sum_{i=1}^N x_i \quad (4)$$

and the root-mean-square error (RMSE) of the difference between the bias corrected calibrated measurement and the ground truth. The latter is the same as the standard deviation of the residuals (SDR)  $\hat{x}_i - x_i$ :

$$\text{SDR} = \sqrt{\frac{1}{N} \sum_i ((\hat{x}_i - b) - x_i)^2} \quad (5)$$

As can be seen in Table 4 below, the bias is mostly positive. Note that sensor 54911 and 1184206 had a limited uptime in the second period, which makes their bias and RMS calculation not very representative.

5 The strongest bias after two months is found for 14560051 and 1184206, both of model NO2-B42F and having been used in others experiments for more than one year. These sensors have also the largest RMSE in the first calibration period (see also Table 3), another indication of their poor performance. The range in RMSE of the remaining sensors is  $4.5 - 7.2 \mu\text{g m}^{-3}$  for the first period. The bias corrected RMSE increases to  $5.3 - 9.3 \mu\text{g m}^{-3}$  for the second period. The latter is a more conservative yet more realistic estimation of the precision of the NO<sub>2</sub> estimates, as they are based on measurements which  
 10 were not used for calibration. Based on our results listed in the last column of Table 4, we take  $7 \mu\text{g m}^{-3}$  as a typical uncertainty for the estimated NO<sub>2</sub> values.

The increase of SDR is also due to a loss of sensitivity over time. The aging of the sensors can be further investigated by recalibrating the devices, i.e. determining the coefficients of regression model D, using the data of the second calibration period (see the Supplemental Material).

15 The panels in Figure 11 show how the calibration coefficients change after two months of deployment. Having in mind that the  $S_{\text{WE}}$  signal is the only component which has direct sensitivity to NO<sub>2</sub>, one can see in Figure 11(b) (all dots below the  $y=x$  line) that all sensors suffer from sensitivity loss to NO<sub>2</sub>. This results in lower  $R^2$  values in Figure 11(f), although the performance loss is partly compensated by the other components in the multivariate linear regression.

The older Alphasense models NO2-B42F (red dots in Figure 11(b)) are the most insensitive to NO<sub>2</sub>, and have the largest  
 20 sensitivity loss, which the regression tries to compensate with an increased temperature dependence (Figure 11(d)), although this can not avoid that they have the worst performance and the worst performance loss in terms of  $R^2$ .

#### 4.7 Weighted calibration

Taking  $18 \mu\text{g m}^{-3}$  as a typical NO<sub>2</sub> concentration in an urban environment (Figure 4), the sensor drift as listed in Table 4 is a significant error component, even after a two month period. It is impossible to predict the progressing bias for an individual  
 25 sensor. However, using the second calibration period we can compensate for signal drift in hindsight. If  $\hat{x}_1(t)$  represents the estimated NO<sub>2</sub> value at time  $t$  based on the first calibration period (starting at  $t_1$ ), and  $\hat{x}_2(t)$  the estimated NO<sub>2</sub> value based on



the second calibration period (ending at  $t_2$ ), then we take for intermediate times  $t_1 \leq t \leq t_2$  a weighted average of both calibrations:

$$\hat{x}(t) = (1 - f(t))\hat{x}_1(t) + f(t)\hat{x}_2(t) \quad (6)$$

Assuming that the sensor degradation is linear in time we select

$$f(t) = (t - t_1)/(t_2 - t_1) \quad (7)$$

such that  $f(t_1)=0$  and  $f(t_2)=1$ .

## 5 4.8 Validation against Oude Schans station

From 14 June to 16 August, sensor 54200 was placed at Korte Koningsstraat (ground floor/street side), which happens to be 120m from another GGD station at Oude Schans, also classified as an urban background station. The Korte Koningsstraat characterizes as a side street, away from traffic arteries. The proximity to a reference station enables us to perform an independent validation of the sensor measurements, as the calibration of the sensor is based on side-by-side measurements with Vondelpark station, at 3km distance. As can be seen from Table 5, the sensor readings agree very well with the official measurements. Figure 12(a) and 12(b) show the time series and the scatter plot.

Using the weighted calibration explained in the previous section, the measurement bias largely disappears. The RMSE is comparable to the RMSE found during the calibration period (see Table 4). The results give confidence that our calibration method is independent of location, and that our assumption of sensor degradation being linear in time is acceptable.

## 15 5 Discussion

As all electrochemical  $\text{NO}_2$  sensors, the Alphasense  $\text{NO}_2\text{-B4}$  sensor is not very selective to the target gas. The sensor response can best be explained as a linear combination of  $\text{NO}_2$ ,  $\text{O}_3$ , temperature and humidity signals ( $R^2 \approx 0.9$ ).

As a consequence, a linear combination of the Working Electrode and the Auxiliary Electrode alone give poor indication of ambient  $\text{NO}_2$  concentrations. The accuracy varies greatly between different sensors ( $R^2$  between 0.3 and 0.7). For the Urban AirQ campaign, temperature and relative humidity were included in a multilinear regression approach. The results improve significantly with  $R^2$  values typically around 0.8. This corresponds well with the findings of Jiao et al. (2016), who find an adjusted  $R^2=0.82$  for the best performing electrochemical  $\text{NO}_2$  sensor in their evaluation, when including T and RH.

Best results are obtained by also including ozone measurements in the calibration model:  $R^2$  increases to 0.9. Spinelle et al. (2015b) used a similar regression and found  $R^2$  ranging from 0.35 to 0.77 for 4 electrochemical  $\text{NO}_2$  sensors during a two-week calibration period, but dropping to 0.03—0.08 when applied to a successive 5-month validation period. Low  $\text{NO}_2$  values at their semi-rural site partly explains this poor performance, but most likely also unaccounted effects such as changing sensor sensitivity and signal drift.



The sensor devices were tested in an Amsterdam urban background in summertime, with  $\text{NO}_2$  values ranging from  $3 \mu\text{g m}^{-3}$  to  $78 \mu\text{g m}^{-3}$ , and median values around  $15 \mu\text{g m}^{-3}$ . During the 3-month period most sensors show loss of sensitivity and significant drift, ranging from  $-9$  to  $21 \mu\text{g m}^{-3}$ . After bias correction we found a typical value for the accuracy of the  $\text{NO}_2$  measurements of  $7 \mu\text{g m}^{-3}$ .

- 5 This error consists of several components. The reference measurements by the  $\text{NO}/\text{NO}_x$  analysers have an estimated hourly error of 3.65% (certified validation at a  $200 \mu\text{g m}^{-3}$   $\text{NO}_2$  concentration), which would contribute to  $0.5 \mu\text{g m}^{-3}$  under typical conditions. The low-cost DHT22 sensor has a reported error of  $0.5 \text{ }^\circ\text{C}$  for temperature and 2–5% for RH. For a single measurement, this would contribute to an error of approximately  $1 \mu\text{g m}^{-3}$  and  $0.5 \mu\text{g m}^{-3}$ , respectively (Figure 11(d) and 11€). It should be noted, however, that binning minute-based measurements to hourly averages removes large part of the
- 10 variability, while determining the best fitting regression model for each sensor device removes large part of the remaining systematical biases. The largest part of the error term is therefore introduced by the linear regression model itself, which does not include all interfering species or meteorological quantities, and is not able to describe non-linear dependencies of its variables. One should therefore be careful extrapolating the calibration model for conditions different than the calibration period.
- 15 The found sensor accuracy after two calibrations and corrections is good enough to complement official measurements by providing additional information on local air quality between reference stations, and detect unexpected hot spots (or low values) of urban  $\text{NO}_2$ . However, it must be further investigated if the regression method used here would provide realistic estimates for peak values (such as the EU hourly limit value,  $200 \mu\text{g m}^{-3}$ ).

The necessity for recalibration troubles practical applications in operational urban networks. Sensors must be brought back

20 to a calibration facility on a regular basis, or must be recalibrated on the spot by a travelling reference instrument. New data driven techniques, such as Bayesian networks (e.g. Xiang et al., 2016), might offer a solution for this problem.

## 6 Conclusions and outlook

The current generation of low-cost  $\text{NO}_2$  sensors has some serious issues which trouble straightforward application. To make electrochemical  $\text{NO}_2$  sensor measurements accurate, careful filtering of the raw data is necessary. There is a strong spread in

25 sensor performance, even if the sensors come from the same batch, which make individual calibration essential. The accuracy of the measurements can be improved by including temperature and humidity measurements from other low-cost sensors in a multilinear regression approach. A practical calibration method is measuring side-by-side to an air monitoring station. This measurement period should be as long as possible (but at least a few days), to capture the sensors behavior under a wide range of pollution levels and meteorological conditions.

30 Startup time of sensors is estimated 4 hours. To avoid nonlinear response of the electrochemical sensor at elevated temperatures, we filter out measurements above  $30 \text{ }^\circ\text{C}$ . This is not a serious restriction for applicability in moderate climates such as in the Netherlands, provided that the sensor is protected from direct sunlight. However, for warmer regions or during



heat waves this may reduce the data stream considerably, unless the temperature dependencies are better captured by more advanced regression models.

The calibration coefficients seem to be location independent, as independent validation in the proximity of a second monitoring station suggests. However, calibration coefficients are not constant in time. During the 3-month period most sensors suffer from significant sensitivity loss and drift. The standard deviation of the random error is estimated 7  $\mu\text{g m}^{-3}$  for a typical sensor. The strongest drift and largest uncertainty are found for the older NO<sub>2</sub>-B42F sensors. It remains unclear if the poorer performance is related to the sensor model or the longer usage in field experiments.

Individual sensor drift can be compensated in hindsight by taking a weighted average of the calibration coefficients determined before and after the campaign, assuming that the sensor degradation is linear in time. The sensor degradation makes it necessary to think about smart re-calibration programs when one wants to use electrochemical sensors operationally in a low-cost urban networks. More research is needed to gain better insight of how sensors age in field applications. This will provide better calibration strategies which improve data quality.

To further improve accuracy of electrochemical NO<sub>2</sub> measurements in low-cost sensor devices we recommend to include an additional ozone sensor to better resolve cross-sensitivity issues. Even imperfect ozone measurements will improve the NO<sub>2</sub> estimation, as large part of the sensor's cross-dependency issues are solved by the linear regression approach. The RH sensor signal should be used more cleverly to detect and filter for sudden changes in humidity. Adding a local data logger is also recommended, to be able to recover data for periods when the WiFi connection to the central database is lost.

### Data availability

A complete overview of fit results for all models can be found in the supplement. The hourly Urban AirQ sensor data, calibrated in hindsight by interpolating the calibration in time between two calibration periods, can be downloaded at <https://github.com/waagsociety/making-sensor>.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The Urban AirQ project was partly funded by a 2016 Stimulus Grant from AMS (Advanced Metropolitan Solutions). The project is also part of Making Sense, funded by European Union's Horizon 2020 research and innovation programme. Qijun Jiang is supported by the China Scholarship Council for his Ph.D. research.

### References

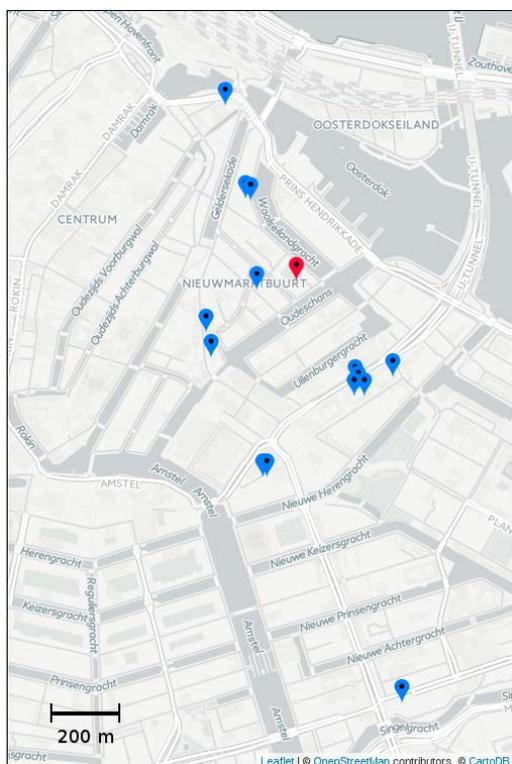
AAN 110, Alphasense Application Note on Environmental Changes: Temperature, Pressure, Humidity, 2013, [http://www.alphasense.com/WEB1213/wp-content/uploads/2013/07/AAN\\_110.pdf](http://www.alphasense.com/WEB1213/wp-content/uploads/2013/07/AAN_110.pdf)



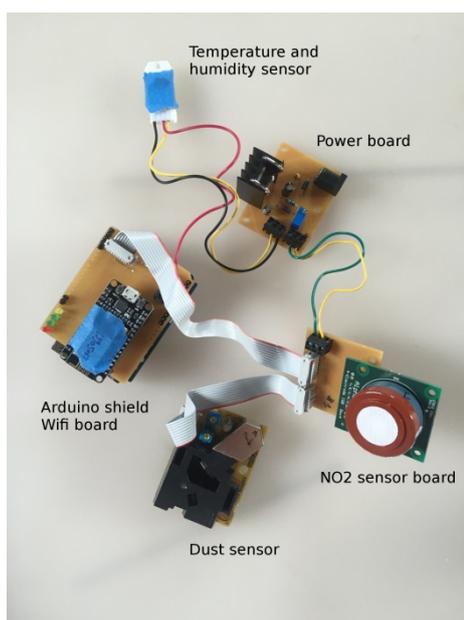
- ADS, Alphasense Data Sheet for NO<sub>2</sub>-B43F, April 2016, <http://www.alphasense.com/WEB1213/wp-content/uploads/2016/07/NO2-B43F.pdf>
- Borrego, C., Costa, A. M., Ginja, J., Amorim, M., Coutinho, M., Karatzas, K., . . . Penza, M. (2016). Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise. *Atmospheric Environment*, 147, 246-263. doi: <http://dx.doi.org/10.1016/j.atmosenv.2016.09.050>
- 5 Di Carlo, S., Falasconi, M., Sanchez, E., Scionti, A., Squillero, G., & Tonda, A. (2011). Increasing pattern recognition accuracy for chemical sensing by evolutionary based drift compensation. *Pattern Recognition Letters*, 32(13), 1594-1603. doi: <http://dx.doi.org/10.1016/j.patrec.2011.05.019>
- Duvall, R., Long, R., Beaver, M., Kronmiller, K., Wheeler, M., & Szykman, J. (2016). Performance Evaluation and  
10 Community Application of Low-Cost Sensors for Ozone and Nitrogen Dioxide. *Sensors*, 16(10), 1698.
- Hierlemann, A., & Gutierrez-Osuna, R. (2008). Higher-Order Chemical Sensing. *Chemical Reviews*, 108(2), 563-613. doi: 10.1021/cr068116m
- Jiang Q., Frank Kresin, Arnold K. Bregt, et al. (2016), "Citizen Sensing for Improved Urban Environmental Monitoring", *Journal of Sensors*, vol. 2016, Article ID 5656245, 9 pages, 2016. doi:10.1155/2016/5656245
- 15 Jiao, W., G. Hagler, R. Williams, R. Sharpe, R. Brown, D. Garver, R. Judge, M. Caudill, J. Rickard, M. Davis, L. Weinstock, S. Zimmer-Dauphinee, K. Buckley (2016), Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States, *Atmospheric Measurement Techniques*, 9, 11, pp 5281-5292, doi 10.5194/amt-9-5281-2016
- Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., . . . Britter, R. (2015). The rise of low-  
20 cost sensing for managing air pollution in cities. *Environment International*, 75, 199-205. doi: <http://dx.doi.org/10.1016/j.envint.2014.11.019>
- Lewis, A. and P. Edwards (2016), "Validate personal air-pollution sensors", *Nature* 535, 29–31, doi:10.1038/535029a
- Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., . . . White, A. (2016). Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday Discussions*, 189(0), 85-103. doi:  
25 10.1039/c5fd00201j
- Masson, N.; Piedrahita, R.; Hannigan, M. Quantification Method for Electrolytic Sensors in Long-Term Monitoring of Ambient Air Quality (2015). *Sensors*, 15, 27283-27302.
- Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., . . . Jones, R. L. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70,  
30 186-203. doi: 10.1016/j.atmosenv.2012.11.060
- Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M., & Fishbain, B. (2015). On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Science of The Total Environment*, 502, 537-547. doi: <http://dx.doi.org/10.1016/j.scitotenv.2014.09.059>



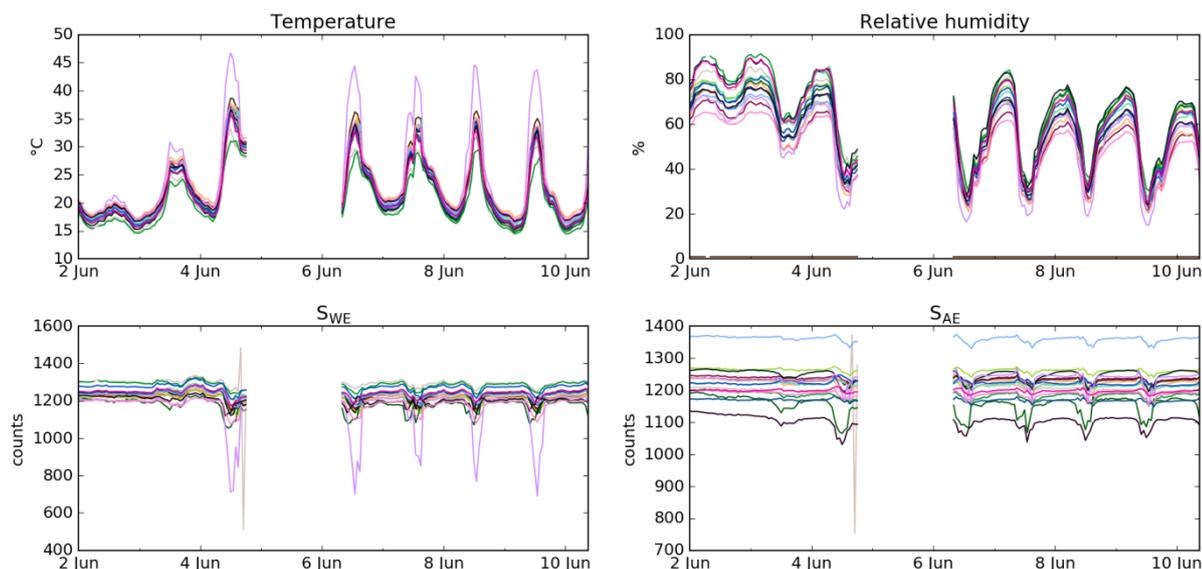
- Pang, X., Shaw, M. D., Lewis, A. C., Carpenter, L. J., & Batchellier, T. (2016). Electrochemical ozone sensors: A miniaturised alternative for ozone measurements in laboratory experiments and air-quality monitoring. *Sensors and Actuators B: Chemical*, 240, 829-837. doi: <http://dx.doi.org/10.1016/j.snb.2016.09.020>
- 5 Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., Li, K., Dick, R. P., Lv, Q., Hannigan, M., and Shang, L. (2014): The next generation of low-cost personal air quality sensors for quantitative exposure monitoring, *Atmos. Meas. Tech.*, 7, 3325-3336, doi:10.5194/amt-7-3325-2014.
- Spinelle, L., M. Gerboles, M. Aleixandre (2015a), EUROSENSORS 2015: Performance evaluation of amperometric sensors for the monitoring of O<sub>3</sub> and NO<sub>2</sub> in ambient air at ppb level, *Procedia Engineering* 00 (2015) 000-000
- Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2015b). Field calibration of a cluster of low-  
10 cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical*, 215, 249-257. doi: <http://dx.doi.org/10.1016/j.snb.2015.03.031>
- Thompson, J. E. (2016). Crowd-sourced air quality studies: A review of the literature & portable sensors. *Trends in Environmental Analytical Chemistry*, 11, 23-34. doi: <http://dx.doi.org/10.1016/j.teac.2016.06.001>
- Xiang, Y., Tang, Y., and Zhu, W. (2016). Mobile sensor network noise reduction and recalibration using a Bayesian  
15 network, *Atmos. Meas. Tech.*, 9, 347-357, doi:10.5194/amt-9-347-2016.



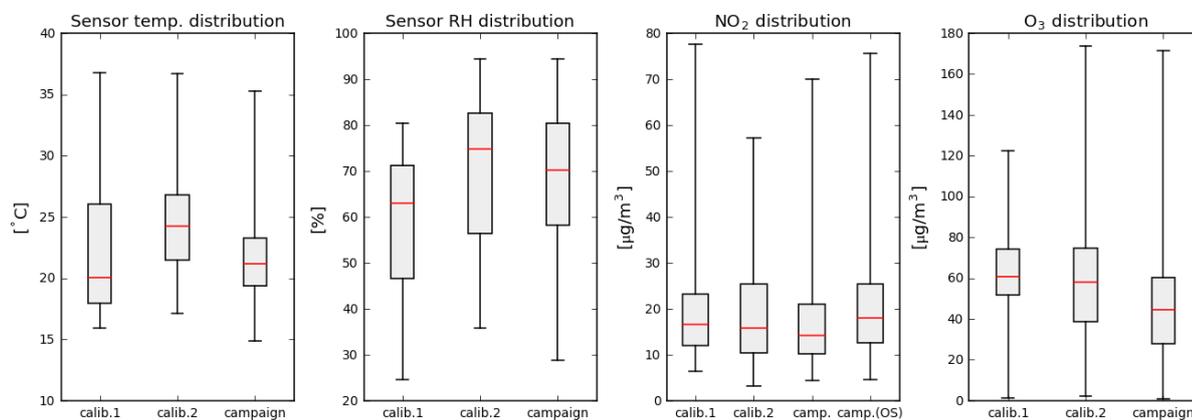
**Figure 1** Locations of the sensor devices during the citizen measurement campaign. The red marker indicates the GGD station at Oude Schans. Not shown is the GGD Vondelpark station, 2.5 km in south-west direction.



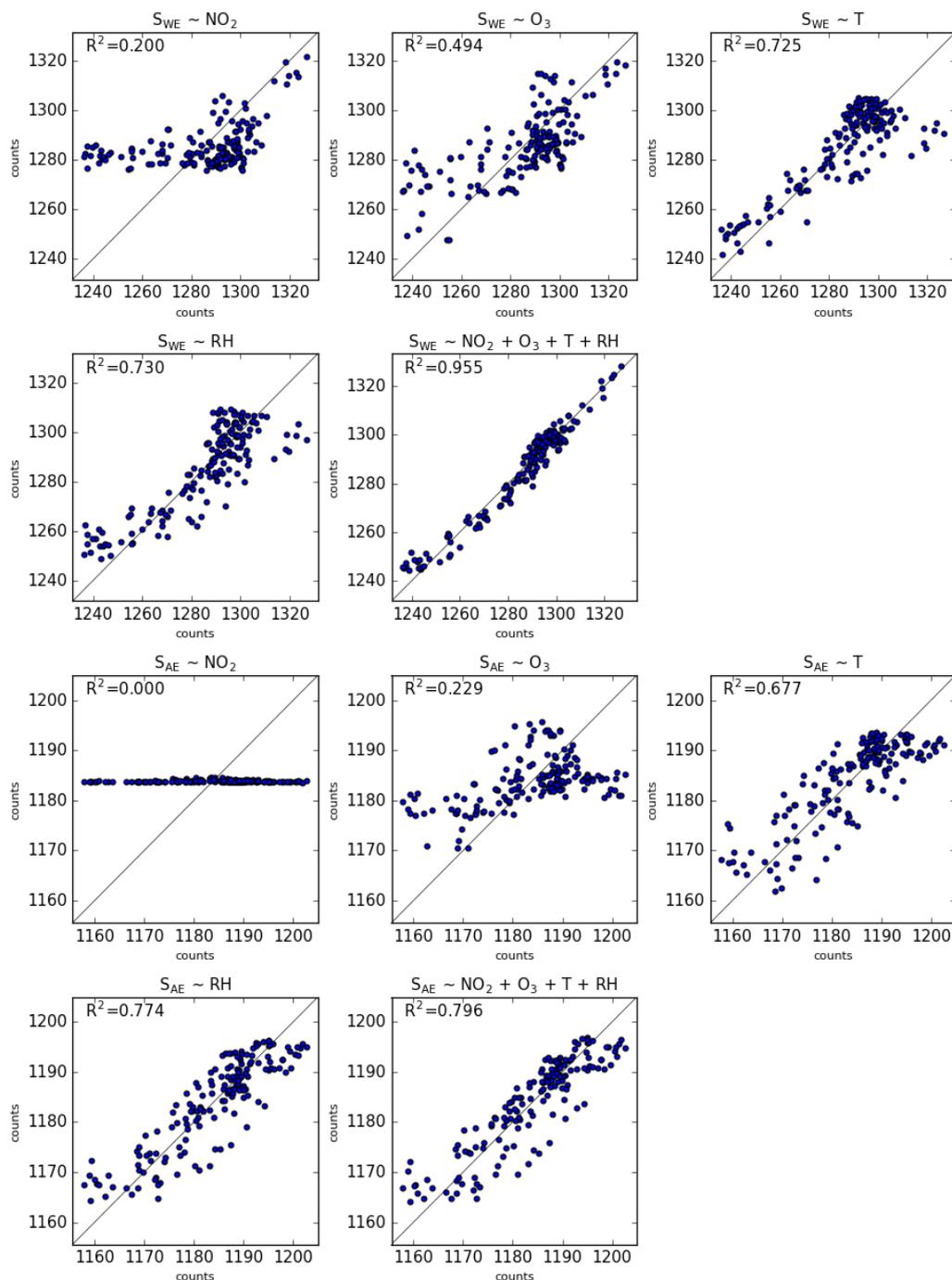
**5 Figure 2** Hardware components of a sensor device (left), and sensors in their housing (right)



**Figure 3** Raw sensor data, unfiltered but hourly averaged, from the 16 sensors during the first calibration period, 2-10 June 2016. The data gap is due to a connectivity problem to the central database.



**5** **Figure 4** Box whisker diagrams of hourly ambient parameters during the calibration period and the measurement campaign. The box edges indicate the 25-75 percentile; the whiskers the minimum and maximum values. The median is indicated in red. Temperature and RH are based on the average values of all sensors devices, NO<sub>2</sub> and ozone are taken from the reference station at Vondelpark. For comparison, NO<sub>2</sub> from the reference station at Oude Schans (OS) is also shown.



**Figure 5** The reading of typical performing NO<sub>2</sub>-B43F sensor (ID 1185325) explained as a linear regression of respectively NO<sub>2</sub>, O<sub>3</sub>, T, RH, and all variables. Top two rows show results for Working Electrode ; bottom two rows for Auxiliary Electrode. On the axis the A/D converter counts, which can be considered as arbitrary units.

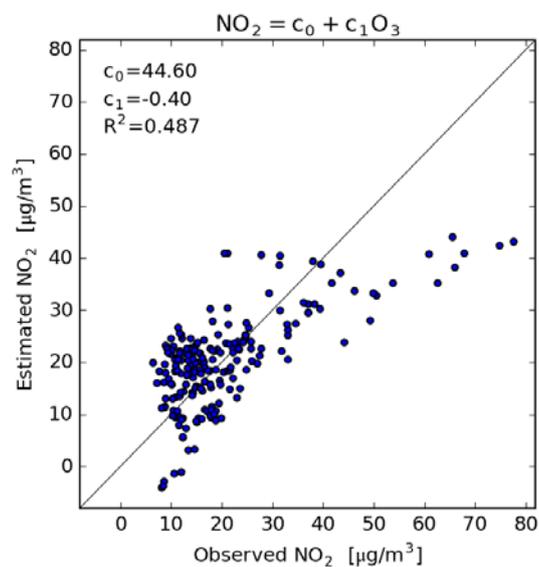


Figure 6 Ozone as a proxy of ambient  $NO_2$ .

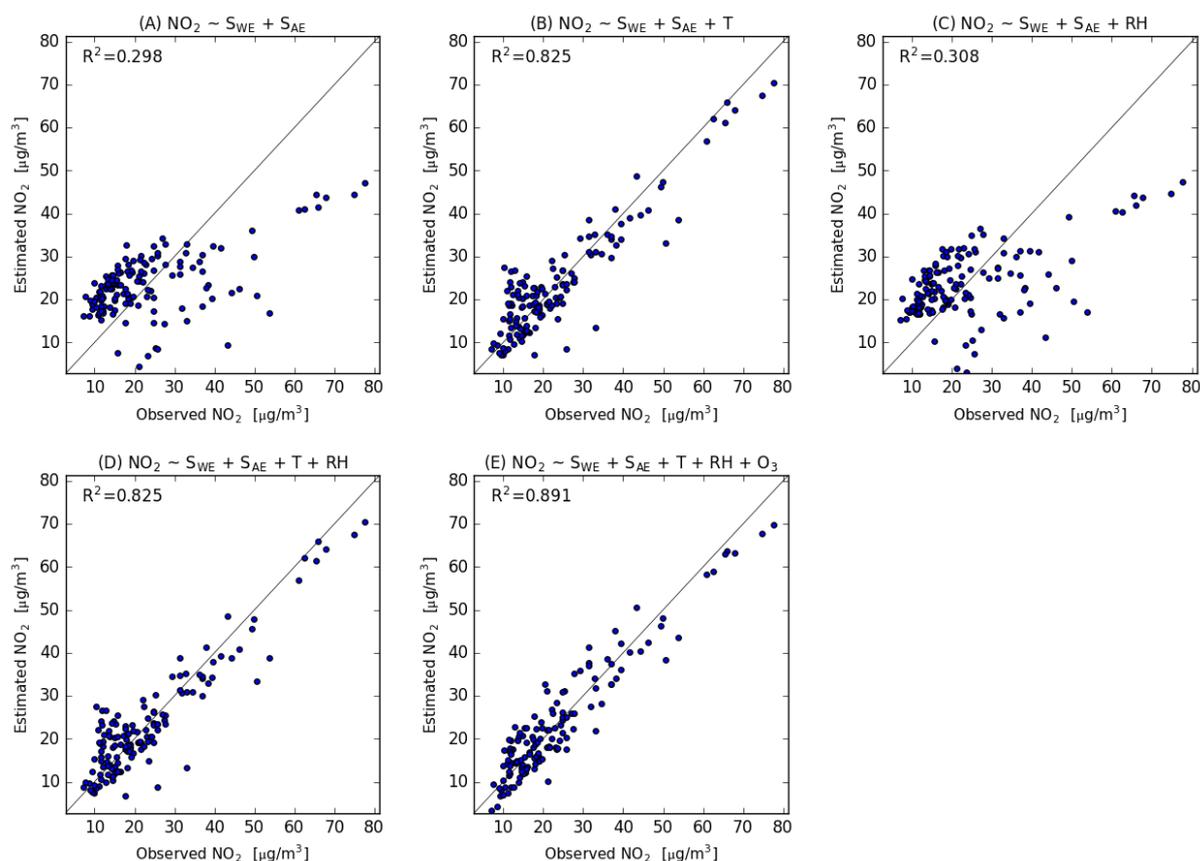
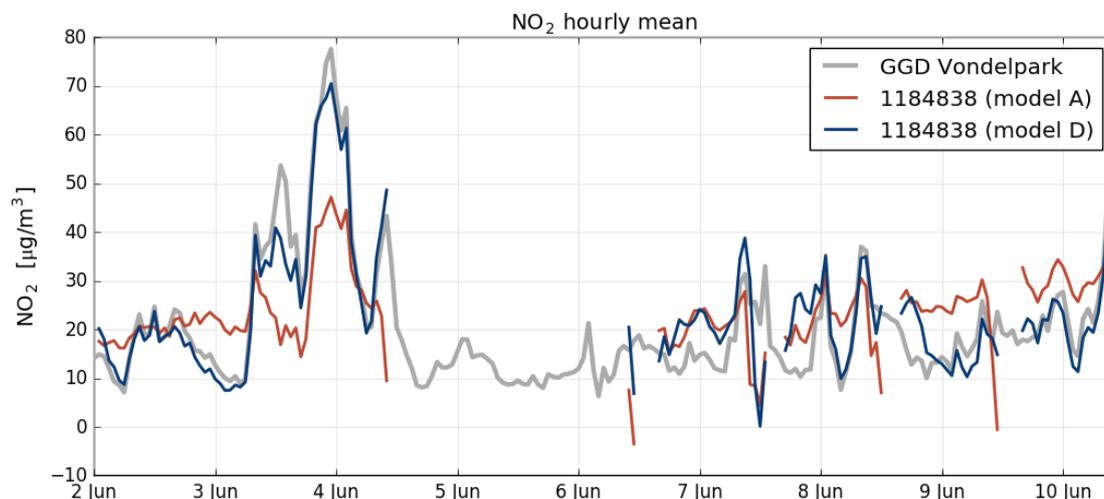


Figure 7(a) Calibration model results for an average performing sensor (ID 1184838). Bottom row shows the recommended calibration by Model D (left), and the results when ozone would be included (right).



5 Figure 7(b) Time series compared to ground truth with calibration parameters of Model A and D.

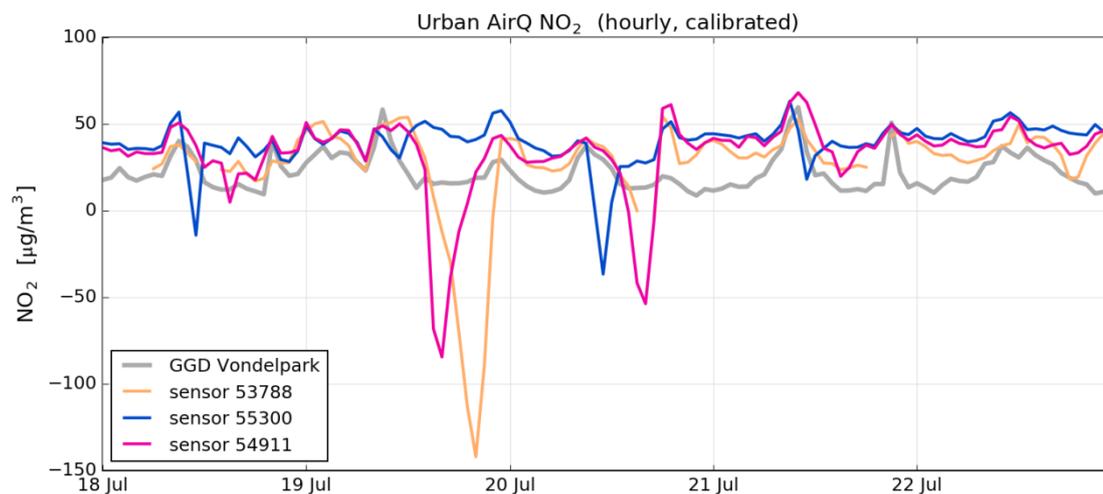


Figure 8(a) Examples of negative spikes in the calibrated NO<sub>2</sub> measurements due to sensor temperatures exceeding 30 °C.

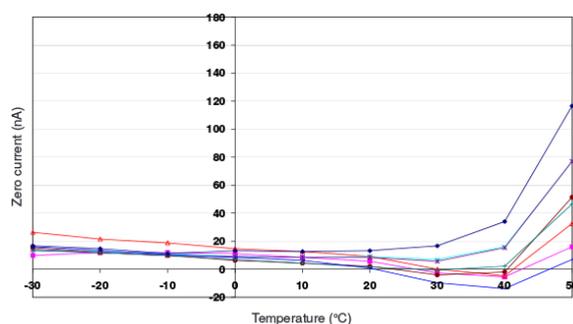


Figure 8(b) Variation of zero output of the working electrode caused by changes in temperature for a typical batch of electrochemical sensors. Image taken from Alphasense Data Sheet for NO<sub>2</sub>-B43F (ADS, 2016).

5

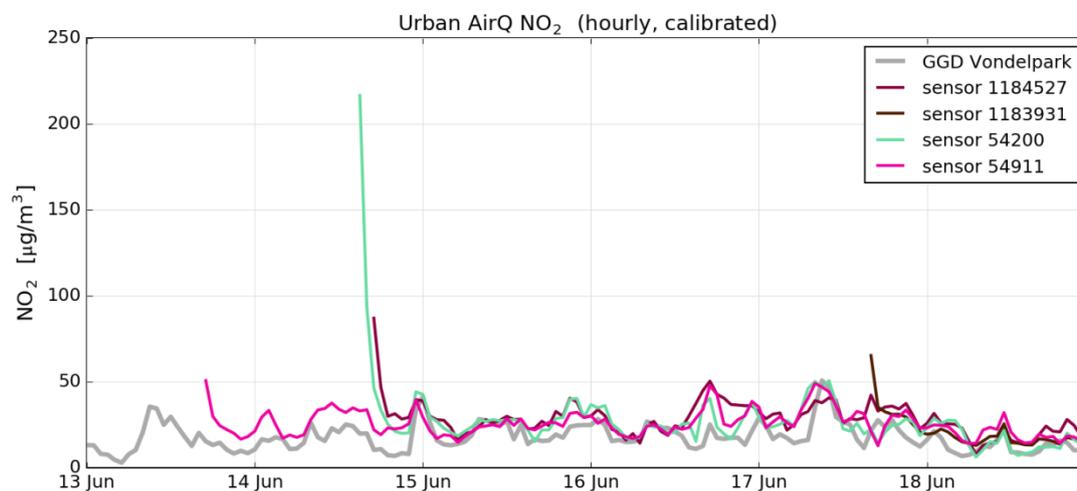


Figure 9 Examples of sensor startup effects when switched on.

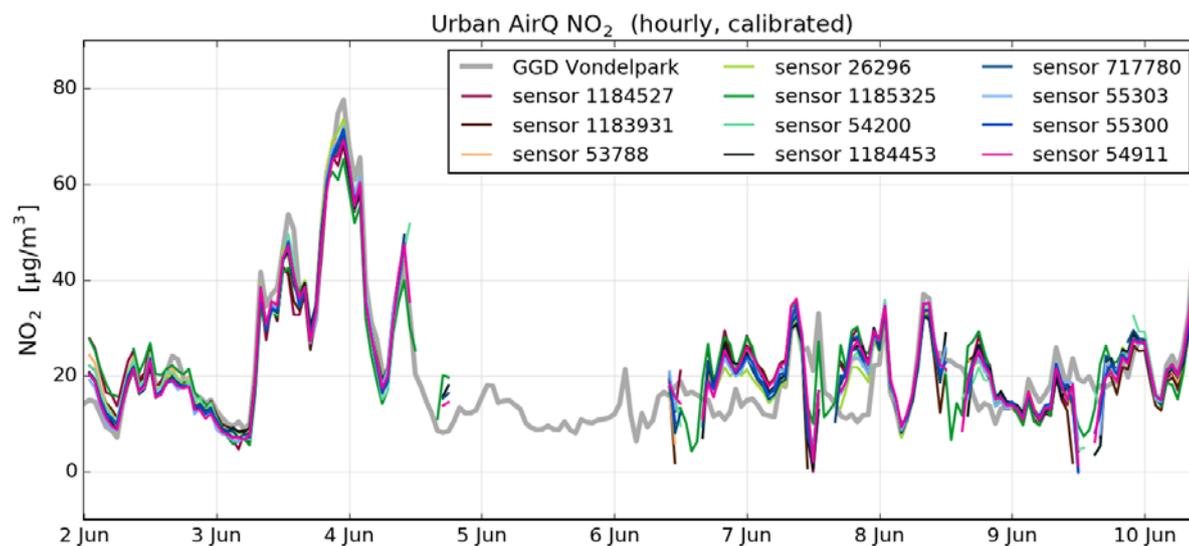
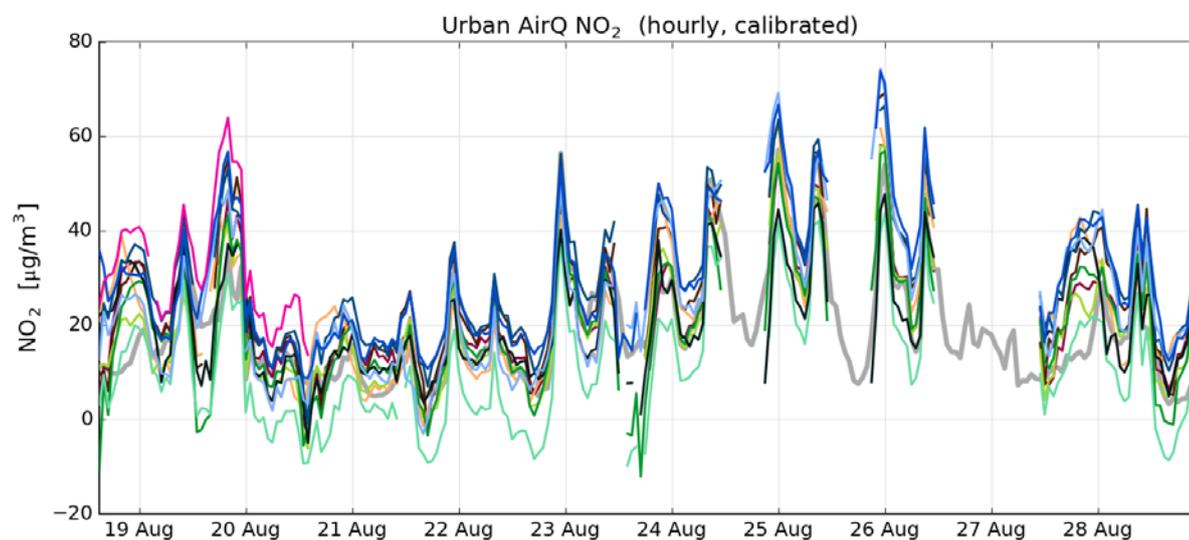
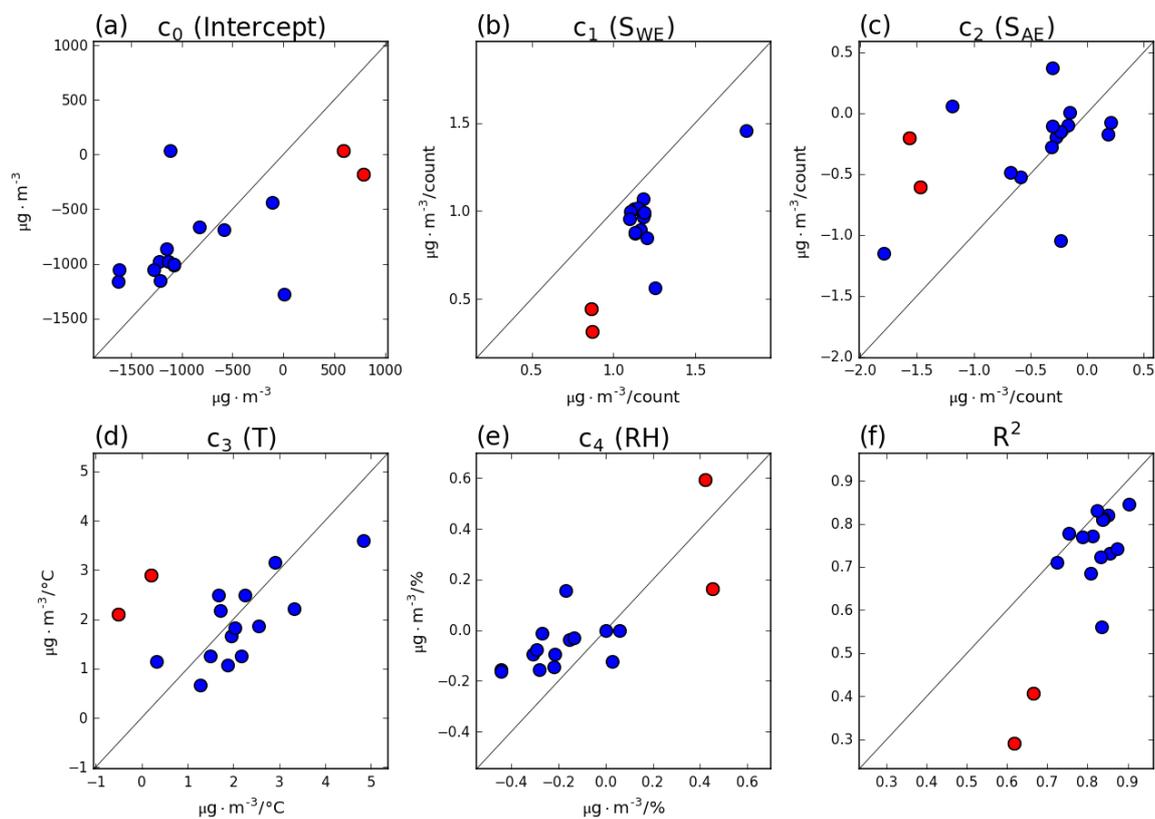


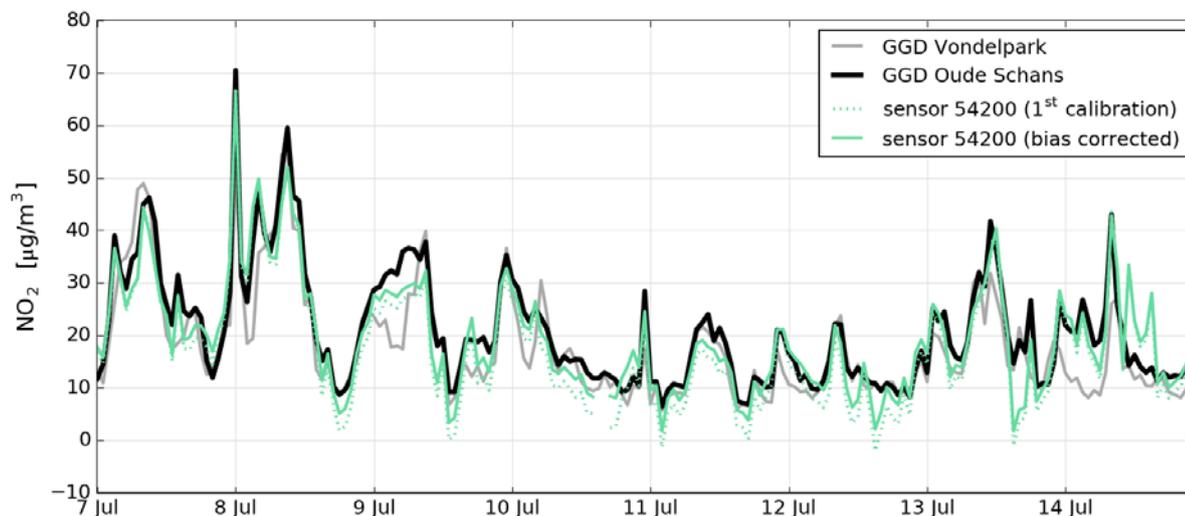
Figure 10(a) Time series of a batch of sensors, calibrated with model D, compared with the reference measurements (grey line).



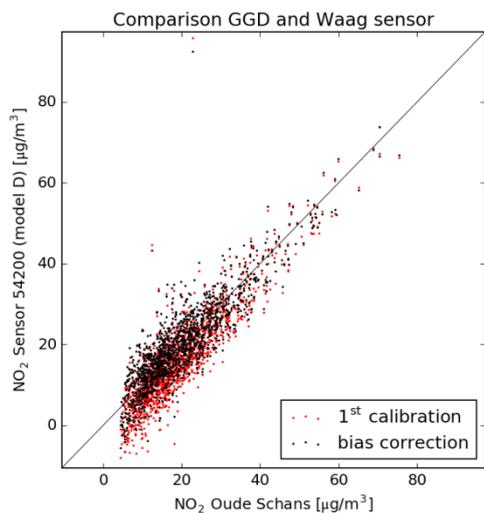
5 Figure 10(b) Comparison of the time series of the same batch of sensors with the reference measurements (grey line), after two months of operation.



5 **Figure 11** Change in calibration coefficients of model D from the first calibration period (horizontal axis) when recalibrating after two months of deployment (vertical axis). The red dots correspond to sensor devices containing the Alphasense NO2-B42F.



**Figure 12(a) Comparison of sensor 54200 NO<sub>2</sub> time series with the nearby Oude Schans station (8-day snap shot), and the effect of bias correction. For comparison, measurements of Vondelpark station are also shown.**



**5 Figure 12(b) Scatterplot of sensor 54200 against Oude Schans station NO<sub>2</sub> measurements during the campaign period.**



**Table 1** Fit results for regression model A, sorted from best to worst sensor

Sensor ID	$c_0$	$c_1 (S_{WE})$	$c_2 (S_{AE})$	$R^2$
55303	-228.6	1.0877	-0.8029	0.72
54200	-968.2	0.9138	-0.1237	0.69
53788	231.4	1.0802	-1.2514	0.68
14560051	355.9	0.8862	-1.2633	0.62
1184453	338.4	0.9823	-1.2246	0.61
55300	-1703.4	0.8218	0.5544	0.60
1183931	-576.4	0.9615	-0.4811	0.57
26296	100.5	0.8669	-0.8952	0.56
717780	-375.2	0.7775	-0.4837	0.54
1185325	342.0	0.8221	-1.1629	0.50
54911	-594.3	0.8007	-0.3192	0.49
1184527	-155.1	0.8368	-0.6841	0.48
1184206	455.4	0.6977	-1.0835	0.47
13905017	162.6	0.8156	-0.9075	0.46
1184739	-141.9	0.6136	-0.5241	0.44
1184838	1211.2	0.9008	-1.8984	0.30

**Table 2** Regression models for  $\text{NO}_2$

Model A	$\text{NO}_2 = c_0 + c_1 \cdot S_{WE} + c_2 \cdot S_{AE}$	Linear combination of Working Electrode and Auxiliary Electrode
Model B	$\text{NO}_2 = c_0 + c_1 \cdot S_{WE} + c_2 \cdot S_{AE} + c_3 \cdot T$	Temperature correction
Model C	$\text{NO}_2 = c_0 + c_1 \cdot S_{WE} + c_2 \cdot S_{AE} + c_4 \cdot \text{RH}$	Relative humidity correction
Model D	$\text{NO}_2 = c_0 + c_1 \cdot S_{WE} + c_2 \cdot S_{AE} + c_3 \cdot T + c_4 \cdot \text{RH}$	Temperature and RH correction
Model E	$\text{NO}_2 = c_0 + c_1 \cdot S_{WE} + c_2 \cdot S_{AE} + c_3 \cdot T + c_4 \cdot \text{RH} + c_5 \cdot \text{O}_3$	Adding also correction for ozone cross-sensitivity



**Table 3** Fit results for regression model D, ordered from best to worst sensor

Sensor ID	$c_0$	$c_1$ ( $S_{WE}$ )	$c_2$ ( $S_{AE}$ )	$c_3$ (T)	$c_4$ (RH)	$R^2$
26296	-586.3	1.1794	-0.6738	2.0415	-0.2192	0.90
55303	-1272.1	1.2045	-0.1492	1.2690	-0.2944	0.87
53788	-1129.7	1.1835	-0.2705	2.2559	-0.2704	0.86
54200	-1613.3	1.1499	0.1818	0.3200	-0.4442	0.85
717780	-1074.9	1.0961	-0.2346	1.4954	-0.2799	0.84
13905017	8.1	1.1860	-1.1889	2.5401	0.0268	0.84
54911	-1215.5	1.2551	-0.3038	2.1742	-0.1333	0.84
55300	-1074.6	1.1294	-0.3058	1.8671	-0.1561	0.83
1184838	-104.5	1.8111	-1.7939	4.8373	0.0596	0.83
1184739	-824.8	1.1850	-0.5839	1.6737	-0.3069	0.81
1184453	-1109.8	1.1055	-0.2339	3.3191	-0.1693	0.81
1183931	-1217.6	1.1305	-0.1642	1.9435	0.0000	0.79
1184527	-1623.1	1.1235	0.2088	1.7161	-0.4430	0.75
1185325	-1152.7	1.1668	-0.3120	2.9112	-0.2147	0.72
14560051	589.2	0.8618	-1.4742	0.2142	0.4204	0.67
1184206	790.9	0.8707	-1.5645	-0.5051	0.4513	0.62



**Table 4 Bias and random error in  $\mu\text{g m}^{-3}$  when calibrated in the first period with model D**

ID	1 <sup>st</sup> calibration period			2 <sup>nd</sup> calibration period		
	Uptime	Bias	SDR	Uptime	Bias	SDR
1185325	77%	0.0	7.2	82%	0.2	7.9
1184453	70%	0.0	6.3	82%	0.5	8.5
26296	68%	0.0	4.5	78%	0.7	5.3
1184739	75%	0.0	6.0	79%	-2.1	6.8
1184527	70%	0.0	7.0	77%	3.0	6.3
53788	68%	0.0	5.4	79%	3.1	9.1
55303	69%	0.0	5.1	82%	5.6	9.3
1183931	67%	0.0	6.6	78%	6.6	6.8
54200	70%	0.0	5.5	81%	-9.2	5.8
55300	70%	0.0	5.8	82%	9.8	7.7
717780	71%	0.0	5.7	77%	10.1	6.0
54911	72%	0.0	5.7	18%	15.6	8.7
13905017	64%	0.0	5.9	84%	16.6	6.9
1184838	69%	0.0	6.0	79%	21.3	6.8
14560051	69%	0.0	8.2	79%	21.4	12.8
1184206	71%	-0.1	8.8	42%	40.1	18.2

**Table 5 Comparison of sensor 54200 with Oude Schans station**

Mean NO <sub>2</sub> , GGD Oude Schans	19.96 $\mu\text{g m}^{-3}$
Mean NO <sub>2</sub> , sensor 54200	19.87 $\mu\text{g m}^{-3}$
Bias	-0.09 $\mu\text{g m}^{-3}$
RMSE residuals	5.2 $\mu\text{g m}^{-3}$
Correlation	0.88