Atmospheric Environment 210 (2019) 66–75

Contents lists available at ScienceDirect

Atmospheric Environment

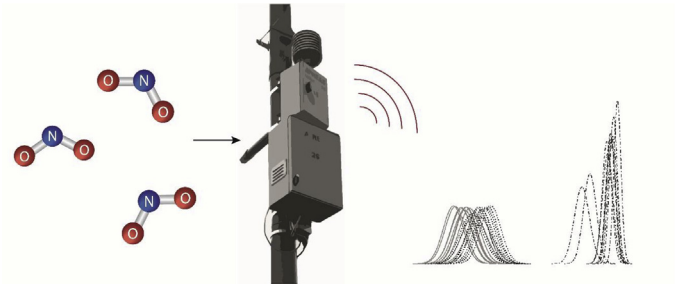
journal homepage: www.elsevier.com/locate/atmosenv

Calibration of low-cost NO_2 sensors in an urban air quality network

Vera van Zoest[a,*], Frank B. Osei[a], Alfred Stein[a], Gerard Hoek[b]

[a] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, Enschede, 7500 AE, Enschede, the Netherlands
[b] Institute for Risk Assessment Sciences (IRAS), Utrecht University, PO Box 80178, 3508 TD, Utrecht, the Netherlands

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Data quality
Calibration
Air quality
Sensor network
Nitrogen dioxide
INLA

ABSTRACT

Low-cost air quality sensors measuring air quality at fine spatio-temporal resolutions, typically suffer from sensor drift and interference. Field calibration is typically performed at one location, while little is known about the spatial transferability of correction factors. We evaluated three calibration methods using a year of hourly nitrogen dioxide (NO_2) observations from low-cost sensors, collocated at two sites with a conventional monitor as reference: (1) an iterative Bayesian approach for daily estimation of the parameters in a multiple linear regression model, (2) a daily updated correction factor and (3) a correction factor updated only when concentrations are uniformly low. We compared the performance of the calibration methods in terms of temporal stability, spatial transferability, and sensor specificity. We documented drift within the 1-year period. The correction factor updated under uniformly low concentrations performed poorly. The iterative Bayesian approach and daily correction factor reduced the root mean squared error (RMSE) by 21–46% at the calibration locations, but did not reduce RMSE at the other location. By examining the posterior distributions of the regression coefficients, we found that the poor spatial transferability is consistent with different responses of individual sensors to environmental factors. We conclude that the spatial and temporal variability in the calibration parameters requires them to be updated regularly, including sensor-specific recalibrations.

1. Introduction

The interest in the use of low-cost air quality sensors at the city level to extend and densify conventional air quality monitoring is increasing (Jerrett et al., 2017; Snyder et al., 2013). A dense air quality monitoring network can be used for modelling and mapping air quality on a finer resolution in space and time than conventional monitoring networks that typically include only one or two monitors per city (Schneider et al., 2017). Such models and maps may be of added value e.g., for health studies and policy making. The data quality of low-cost air quality sensor networks however is often poor or unknown, potentially leading to wrong policy decisions or bias when applied in

* Corresponding author.
E-mail address: v.m.vanzoest@utwente.nl (V. van Zoest).

https://doi.org/10.1016/j.atmosenv.2019.04.048
Received 12 November 2018; Received in revised form 18 April 2019; Accepted 23 April 2019
Available online 25 April 2019

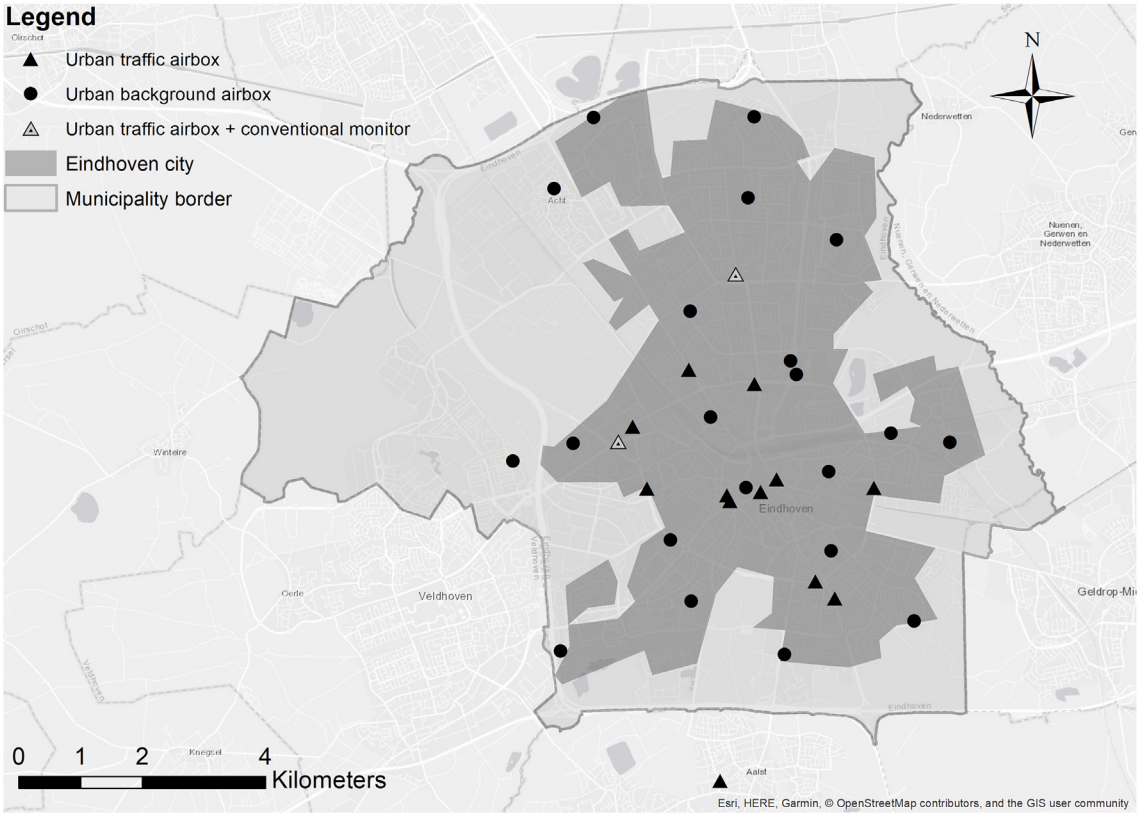


Fig. 1. Locations of the airboxes and conventional monitors in Eindhoven.

epidemiological studies (Snyder et al., 2013). In order to adjust, correct or improve sensor observations, calibration is important in the development and maintenance of such networks (Lewis and Edwards, 2016).

Calibration starts in the lab with finding an optimal function to convert absorbance or conductivity to pollutant concentrations (Neri et al., 2002). Performance indicators include sensitivity, selectivity, stability, response time, saturation, sensitivity to humidity, and the limit of detection (LoD) (Colin et al., 1998; Morales et al., 2002; Penza et al., 1998; Santos, Serrini, O'Beirn and Manes, 1997). Lab calibration, however, is not sufficient for field deployment of air quality sensors. The sensors perform less accurately under changing weather conditions and when exposed to different mixtures of gases as compared to calibration in the lab (De Vito, Piga, Martinotto and Di Francia, 2009; Kamionka et al., 2006). For long-term performance in the field, there are two main challenges: drift and interference effects.

The electrochemical cells typically used in low-cost gas sensors are more prone to lose sensitivity as compared to conventional monitors. This leads to sensor drift: an increasing bias in the sensor response. Xiang et al. (2016) showed that the measurement error due to drift in nitrogen dioxide (NO_2) sensors increased by a factor three within two months. Low-cost air quality sensors are also sensitive to relative concentration distribution changes, for example caused by seasonality or pollution events (De Vito et al., 2009; Moltchanov et al., 2015). The gas sensor can show interaction and interference effects: an unwanted response to other pollutants, gases, temperature and relative humidity (Kamionka et al., 2006; Mead et al., 2013). We refer to these unwanted responses as interference effects. Field calibration is needed to estimate the parameters which correct the sensor response such that it represents the true concentrations as good as possible, accounting for drift and interference.

Much research has already been done to find the best methods for field calibration of air quality sensors. Multiple linear regression (MLR) has been used in the past few years to build calibration functions in which covariates account for environmental and meteorological variability (Kizel et al., 2018; Piedrahita et al., 2014; Spinelle et al., 2015). Feed-forward artificial neural networks (ANNs) have also been used to adjust for interference effects (De Vito et al., 2009). Xiang et al. (2016) also accounted for multiple sensors drifting simultaneously.

In air quality sensor networks, one commonly used approach is that only one sensor is collocated with a reference monitor during regular operation. Before operation and in regular intervals during operation, sensors may be collocated to be calibrated or recalibrated. In addition, some studies report automatic calibration procedures, aiming at reducing calibration costs. When recalibrating the sensors in a sensor network to account for drift and varying meteorological conditions in this way, it is assumed that all sensors behave similarly, and are influenced by the same conditions and similar drift. Although there is evidence that this assumption is not valid (Barakeh et al., 2016), calibration parameters established at a location with a reference instrument are often assumed to be transferable to other locations in the network.

The aim of this study is to evaluate the spatio-temporal variability and sensor specificity of calibration parameters. To do so, we evaluate three different calibration methods that account for interference effects and drift. The methodology is applied to the city of Eindhoven, where a low-cost sensor network was set up with NO_2 measurements at 25 locations (Hamm et al., 2016). At two locations the low-cost sensors were co-located with reference monitors, allowing us to assess spatial and temporal variability of correction factors. The focus of this study is on NO_2, because earlier passive sampler measurements have documented large spatial variation in European cities related especially to road traffic (Cyrys et al., 2012).

2. Methods

We evaluated the performance of (1) an iterative Bayesian approach for daily estimation of the parameters in a multiple linear regression model, (2) a daily updated simple correction factor and (3) a uniform concentration correction similar to the method suggested in Tsujita

et al. (2005). For Bayesian inference we used Integrated Nested Laplace Approximations (INLA). We evaluated the methods in terms of the stability of the correction factors or calibration parameters in time, the usability of the correction factors or calibration parameters at other locations within the same urban area, and the transferability of the calibration parameters to different sensors of the same type.

2.1. Data

The air quality sensor network in Eindhoven (Fig. 1) was established by the AiREAS civil initiative (Close, 2016) in November 2013 and has continuously been operating since. The sensor network consists of 35 fixed airboxes, manufactured by the Energy Research Centre of the Netherlands. The airboxes contain an array of sensors measuring particulate matter (PM), temperature and relative humidity (RH). Some airboxes also measure ozone (O_3), NO_2 or both. The focus of this study is on NO_2, which is measured in 25 airboxes since 2016, after an initial calibration at the end of 2015 (Hamm et al., 2016). NO_2 is measured using the electrochemical cell Citytech Sensoric NO_2 3E50 in a differential measurement setup. A switching valve and reagent cartridges are used in front of the electrochemical cell to dry the air. Observations are discarded when temperature and humidity are outside acceptable ranges. All airboxes are installed at 2.5–3 m height and at the roadside, as lamp posts were used to supply electricity. Sites were purposely selected reflecting background and traffic sites (Fig. 1). Background sites were located in minor residential roads; traffic sites in roads with typically more than 10,000 vehicles/day. The two conventional monitors sites are also designated as traffic sites in the Dutch national air quality measurement network. Further details are described elsewhere (Van Zoest et al., 2018).

Two conventional chemiluminescence monitors of the Dutch national air quality measurement network (Buijsman, 2013), operated by the National Institute for Public Health and the Environment, are both located in similar urban traffic locations in Eindhoven and are used as reference instruments. With each conventional monitor, an airbox is collocated. Airbox NO_2 measurements are averaged to hourly values for analysis of the calibration methods, similar to the temporal resolution of the conventional monitors. Data cleaning and outlier detection were performed as described in Van Zoest et al. (2018). We refer to $S = \{s_1, ..., s_{25}\}$ as the collection of airboxes measuring NO_2 and $Z = \{z_1, z_2\}$ as the collection of conventional monitors, where $Z \subset S$.

In this study we use covariates $c \in C = \{NO2,\ O3,\ RH,\ T,\ WS,\ WD\}$. Nitrogen dioxide ($NO2$), relative humidity ($RH$) and temperature ($T$) are measured within the airbox. Ozone ($O3$) data in the airboxes at z_1 and z_2 were missing for most of the year due to sensor failure. Therefore we used the O_3 data available at the conventional monitor at location z_1 (O_3 not measured at z_2). Wind speed (WS) and wind direction (WD) were obtained from the Royal Netherlands Meteorological Institute weather station in Eindhoven (KNMI, 2016). NO_2 concentrations were square root transformed to approximate a normal distribution. Where needed, the covariates were also transformed to obtain distributions closer to the normal distribution and to obtain a more linear relationship between the covariate and square root transformed NO_2 concentrations. An overview of the potential covariates, their sources and transformations is given in Table S1 in the supplementary materials.

2.2. Sensor drift

Sensor drift is caused by the loss of sensitivity of the electrochemical cell measuring NO_2. We examine the average drift of the sensor network by plotting a time series of the difference $\Delta_{d,t}$ between the mean NO_2 concentration of the two conventional monitors and the mean NO_2 concentration observed by all airboxes deployed in the city, for all hourly observations in 2016:

$$\Delta_{d,t} = \frac{\sum_S (x_{NO2,d,t,s})}{|S|} - \frac{\sum_Z (y_{NO2,d,t,z})}{|Z|} \quad (1)$$

where $x_{NO2,d,t,s}$ is the hourly NO2 concentration measured at day d and hour $t = 1, ..., 24$ at airbox location s for $|S|$ number of airboxes, $y_{NO2,d,t,z}$ is the hourly NO2 concentration measured at day d and hour $t = 1, ..., 24$ at conventional monitor location z for $|Z|$ number of conventional monitors. For calculation of $\Delta_{d,t}$ we use the actual NO_2 observations before square root transformation. We take a smoothed line through the time series of $\Delta_{d,t}$ to largely separate the effect of drift from temporally varying spatial variability.

2.3. Multiple linear regression model

2.3.1. Calibration

The term calibration refers to two processes. First, it establishes a relationship between indicative measurements and standard (reference) measurements, i.e., estimating the parameters of the calibration function; second, it uses an established relationship, i.e., the calibration function, for obtaining a measurement result from an indicative measurement (Rasch et al., 1994). In this paper, our focus is on the first process, and we will use the term correction for the second process.

MLR functions have been widely used to build calibration functions accounting for environmental and meteorological variables (Piedrahita et al., 2014; Spinelle et al., 2015). We adapted the method, (1) to allow for transformations of the response variable and covariates to obtain better linear relationships, (2) to estimate the calibration parameters including their uncertainty using Bayesian inference, and (3) to iteratively update the calibration parameters on a daily basis using observations of the previous 30 days.

The calibration function resembles a generalized additive model (GAM):

$$y_{d,t,z} = \beta_{0,d,z} + \sum_c \beta_{c,d,z} g_c(x_{c,d,t,z}) + \varepsilon_{d,t,z} \quad (2)$$

where $y_{d,t,z}$ is the square root transformed reference NO_2 level at day d, hour $t = 1, ..., 24$ and location z, $g_c(x_{c,d,t,z})$ are covariate-dependent known functions or transformations applied to covariate $x_{c,d,t,z}$, $\beta_{0,d,z}$ is the intercept and $\beta_{c,d,z}$ are the unknown coefficients for covariates c for day d at location z and the error is assumed $\varepsilon_{d,t,z} \sim N(0, \sigma^2)$. The coefficients $\beta_{c,d,z}$ have a posterior distribution with mean $\mu_{\beta_{c,d,z}}$ and precision $\tau_{\beta_{c,d,z}}$. The covariate-dependent transformations $g_c(x_{c,d,t,z})$ are chosen such that $x_{c,d,t,z}$ approximates a normal distribution, and $g_c(x_{c,d,t,z})$ has an approximately linear relation with $y_{d,t,z}$.

We used hierarchical Bayesian estimation and inference. Bayesian inference provides a posterior distribution for each $\beta_{c,d,z}$ rather than a single estimate, and therefore allows for the comparison of estimates of different airboxes, including their uncertainty. The parameters of the posterior distributions were estimated using Integrated Nested Laplace Approximation (INLA). INLA provides fast and accurate Bayesian parameter estimates through Laplace approximations. The advantage of INLA over Markov Chain Monte Carlo (MCMC) simulations is that the computation time is significantly shorter, while INLA gives an approximation which is as good or better (Rue et al., 2009).

We built a set of multiple linear regression models including combinations of covariates which are often used in calibration of NO_2 sensors, as they are known for causing interference effects or sensor bias. All covariates are scaled and centered to zero before running INLA. The calibration equation is updated every day d to account for gradual drift and meteorological variability using all non-missing observations in the previous 30 days, adding up to maximum 720 hourly observations. Moltchanov et al. (2015) found variability in calibration parameters over shorter periods of time, e.g. 4 days, using 30-min averages. We chose a period of 30 days based on prior analyses (Table S2), aiming to include enough variation in air pollutant concentrations while minimizing the longer term effects of drift and seasonality. We

further note that our calibration parameters change daily in a smooth manner. Our method does not distinguish between day and night periods in calibration parameters. Using R-INLA (Martins et al., 2013), we built a model on the calibration set to estimate the parameters of the posterior distribution $\theta_{c,d,z} = \left(\mu_{\beta_{c,d,z}}, \tau_{\beta_{c,d,z}}\right)$ of the coefficients $\beta_{c,d,z}$. The models are built at the two locations z_1 and z_2 where a conventional monitor is collocated with an airbox. The model is rebuilt for every day in 2016, such that there is an overlap of 29 days between the data used for calibration on day d and for calibration on the next day d_{+1}.

2.3.2. Calibration performance measures

The fit of a Bayesian model is commonly evaluated using posterior predictive checks or leave-one-out cross-validation. For the first check the posterior predictive p-values,

$$p(y^*_{d,t} \leq y_{d,t}|\boldsymbol{y}) \tag{3}$$

for replicate observations $y^*_{d,t}$, are evaluated to be uniformly distributed. For none of the models explored in this study the distribution of the posterior predictive p-values was uniformly distributed. Wang et al. (2018) however argue that in some cases the posterior predictive p-values can be affected by the nature of the data in such a way that they would never be uniformly distributed even in the case of a perfect model. Therefore they suggest using the probability integral transform (PIT) instead:

$$PIT_{d,t} = p(y^*_{d,t} \leq y_{d,t}|\boldsymbol{y}_{-d,t}) \tag{4}$$

where $\boldsymbol{y}_{-d,t}$ are all observations except for the observation at time stamp t on day d. The performance of the different calibration models is evaluated based on the Deviance Information Criterion (DIC), a generalization of the Akaike Information Criterion (AIC), accounting for both model complexity and fit in a Bayesian model (Spiegelhalter et al., 2002):

$$DIC = \bar{D} + p_D \tag{5}$$

where $\bar{D}$ is the posterior mean of the deviance and p_D is the effective number of parameters. A smaller DIC denotes a better fit.

2.3.3. Validation

For temporal validation, we use the calibration function from Eq. (2) and replace the unknown β by $\hat{\beta}$ to predict $y_{d,t,z}$:

$$\hat{y}_{d,t,z} = \hat{\beta}_{0,d,z} + \sum_c \hat{\beta}_{c,d,z} g_c(x_{c,d,t,z}) + \varepsilon_{d,t,z} \tag{6}$$

for which we now know the posterior distributions of $\hat{\beta}_{0,d,z}$ and $\hat{\beta}_{c,d,z}$. On every day d, Eq. (6) is applied on $t = 1...24$ using the parameters of $\hat{\beta}_{0,d,z}$ and $\hat{\beta}_{c,d,z}$ estimated during the calibration phase. The calibration and validation are repeated daily, so each 24 h period is validated with a new set of calibrated parameters based on the hourly data available in the previous 30 days.

The number of locations with a collocated reference monitor is always sparse in low-cost air quality sensor networks. If the drift and the influence of external variables are similar for each airbox and location, a calibration model built at one location can be transferred to the other locations. To test this, we apply spatiotemporal validation by adjusting Eq. (6) to predict $\hat{y}_{d,t,z}$ at a different location (z_i) from where the model is built (z_j):

$$\hat{y}_{d,t,z_i} = \hat{\beta}_{0,d,z_j} + \sum_c \hat{\beta}_{c,d,z_j} g_c\left(x_{c,d,t,z_j}\right) + \varepsilon_{d,t,z_i} \tag{7}$$

for the two locations where a conventional monitor is located. Note the different subscripts of z to denote the different locations used in spatiotemporal validation. Similar to the temporal validation, Eq. (7) is applied on $t = 1...24$ on the current day d using the parameters of $\hat{\beta}_{0,d,z}$ and $\hat{\beta}_{c,d,z}$ estimated during the calibration phase, but now at a different location z_j.

2.3.4. Validation performance measures

Prediction performance is based on the Root Mean Squared Error (RMSE). We consider two RMSE values: the RMSE before calibration ($RMSE_{pre}$) and the RMSE after calibration ($RMSE_{post}$). $RMSE_{pre}$ is obtained as:

$$RMSE_{pre} = \sqrt{\frac{\sum_{t=1}^{T_d} (y_{t,d} - x_{NO2,t,d})^2}{T_d}} \tag{8}$$

where $y_{t,d}$ is the observed reference NO2 concentration ($NO2_{ref}$) and $x_{NO2,t,d}$ is the observed airbox NO2 concentration ($NO2_{ab}$) for time stamp t in $1, ..., T_d$ where T_d is the total number of non-missing hours in validation day d. $RMSE_{post}$ is obtained as:

$$RMSE_{post} = \sqrt{\frac{\sum_{t=1}^{T_d} (y_{t,d} - \hat{y}_{t,d})^2}{T_d}} \tag{9}$$

where $\hat{y}_{t,d}$ is the predicted NO_2 concentration. A smaller RMSE denotes better prediction.

The calibration approach using INLA is compared to two other techniques that are often used for correction of low-cost air quality sensor networks: (1) a simple daily updated correction factor and (2) a correction factor which is updated when the concentrations are low and uniform across the sensor network.

2.4. Daily correction factor

Miskell et al. (2018) calibrated low-cost O_3 sensors using a conventional monitor in the vicinity with similar land use type. This method assumes that the drift between the airboxes is similar and that high peaks missed by the airboxes are due to meteorological factors which are the same across the sensor network. We apply a similar method to NO2, distinguishing between an absolute correction factor and a relative correction factor. On each day d we find the relative difference correction factor $\gamma_{rel,d,z}$:

$$\gamma_{rel,d,z} = \sum_{t=1}^{T_d} \left(\frac{y_{t,d,z}}{x_{NO2,t,d,z}}\right) \times \frac{1}{T_d} \tag{10}$$

and the absolute difference correction factor $\gamma_{abs,d,z}$:

$$\gamma_{abs,d,z} = \frac{\sum_{t=1}^{T_d} (y_{t,d,z} - x_{NO2,t,d,z})}{T_d} \tag{11}$$

The correction factors are computed on a location z where an airbox is collocated with a conventional monitor. On a daily basis, $\gamma_{rel,d,z}$ or $\gamma_{abs,d,z}$ corrects all hourly airbox measurements of that day, for all airboxes located at a similar site type (urban traffic or urban background). The spatial transferability of the correction factor is evaluated by applying it at the other airbox location in Z and comparing the corrected airbox NO_2 concentrations with the observations of the conventional monitor at that location. Since both conventional monitors are located at urban traffic locations, we could only evaluate the method for this site type. The RMSE is calculated before and after correction.

2.5. Uniform concentration correction

Tsujita et al. (2005) proposed a method for automatic calibration of low-cost air quality sensor networks. The method differs from the previously mentioned correction factor, in the sense that the correction factor is only updated under conditions of uniform low NO2 concentrations. We tested a similar method. When NO_2 concentrations are uniform and low for any hourly timestamp, the baseline of the low-cost sensors is adjusted to the mean of the conventional monitors. All hourly NO_2 observations are corrected using a fixed correction factor γ_{uni} which is the same for all airboxes. This correction factor is based on the ratio

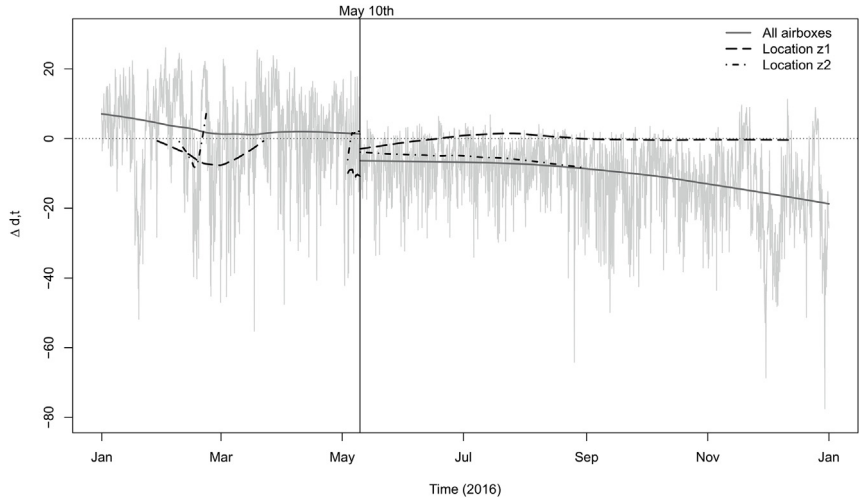


Fig. 2. Difference between mean airboxes and mean conventional monitors over time (light gray line) and fitted smooth curves before and after May 10th (dark gray solid lines). Vertical solid black line: May 10th, after which we observe a sudden decrease in $\Delta_{d,t}$. Black dashed and dot-dashed lines: fitted smooth curves of difference between conventional monitor and airbox at location z_1 and location z_2, respectively. The line at z_2 has been drawn until September. Missing values in December would otherwise influence the line up towards the high levels in November, making comparison with z_1 and the mean of all airboxes impossible.

between the mean of the two conventional monitors and the mean of all low-cost sensors. The correction factor is applied to all future observations of the low-cost sensors, until it is updated at time stamp d, t when standard deviations of the NO2 concentrations drop below a threshold $\delta_{d,t}$ and the mean NO2 concentrations drops below a threshold $\eta_{d,t}$:

$$\sqrt{\frac{\sum_S (x_{NO2,s,d,t} - \bar{x}_{NO2,d,t})}{|S-1|}} < \delta_{d,t} \ \wedge \ \frac{\sum_S x_{NO2,s,d,t}}{|S|} < \eta_{d,t} \tag{12}$$

We evaluated different values of $\delta_{d,t} = \{4,\ 5,\ 6\}$ µg m^{-3} and values $\eta_{d,t} = \{10,\ 12,\ 15\}$ µg m^{-3}.

Following Moltchanov et al. (2015) we also applied night-time calibrations. Every night the concentrations are assumed to be uniform, and a new airbox-specific correction factor $\gamma_{night,d,s}$ is retrieved from the ratio between each individual airbox and the average between the two conventional monitors between 1:00–4:00 a.m. The correction factor is used to correct the airbox values during the next day.

2.6. Sensitivity of individual airboxes to environmental factors

We evaluated whether individual airboxes reacted similarly to interfering gases, temperature and humidity. First, using INLA we examined the relationship between $NO2_{ab}$ and the other variables RH and T measured in the airbox. For the 25 airboxes the posterior mean estimates were compared based on slope direction and strength. The differences in posterior mean estimates between the 25 non-collocated airboxes reflect both spatial variability in the calibration parameters, as well as inter-sensor variability (Broday (2017)). Second, for a comparison independent of the airbox location, we compared the posterior mean estimates of ten airboxes which were simultaneously collocated with a conventional monitor for ten days ($n = 240$ hourly observations per airbox). A separate model, with $NO2_{ref}$ as the response variable, was built using INLA for each covariate measured in the airbox: $NO2_{ab}$, RH, and T. The posterior mean estimates in this case solely reflect inter-sensor variability.

2.7. Temporal autocorrelation

A first-order random walk model is added to Eq. (2) to account for possible temporal autocorrelation, as suggested in Blangiardo and Cameletti (2015):

$$y_{d,t,z} = \beta_{0,d,z} + \sum_c \beta_{c,d,z} g_c(x_{c,d,t,z}) + \ \mathrm{rw}_1(d,\ t) + \varepsilon_{d,t,z} \tag{13}$$

where $\mathrm{rw}_1(d,\ t)$ is a first-order random walk function on the time series of d, t. For both locations z_1 and z_2, and for both the model with and without random walk component, the $RMSE_{pre}$ and $RMSE_{post}$ are obtained. A lower $RMSE_{post}$ for the model with random walk component compared to the model without random walk component suggests the presence of temporal autocorrelation. The models in Eqs. (2) and (13) are also compared using a full year of data, for which 80% of all hourly observations in 2016 are randomly selected for calibration and the remaining 20% of the observations is used for validation.

3. Results

3.1. Descriptive statistics

Based on the initial calibration, the airbox showed good agreement with the conventional monitors. The correlation coefficient for the full year 2016 for hourly NO2 concentrations is 0.75 at z_1 and 0.83 at z_2. Scatterplots for both locations nevertheless showed substantial differences of individual observations (supplementary materials, Fig. S1).

The percentage of missing values in the dataset of 2016 was 23.1% for the airboxes, of which 4.4% was removed during data cleaning and outlier detection (Van Zoest et al., 2018). The remainder is caused by sensor malfunctioning (6.6%) or displacement during maintenance periods (12.1%). To reduce the service costs of the sensor network, maintenance periods were long. The airbox located at z_1 was removed for maintenance from the 22nd of March until the 3rd of May 2016. The airbox at z_2 was removed from the 22nd of February until the 3rd of May 2016. In future development of the network, maintenance time clearly needs to be reduced. For the two conventional monitors, 2.8% of the hourly NO_2 observations was missing in 2016.

3.2. Sensor drift

Fig. 2 shows the time series of $\Delta_{d,t}$, the difference between the mean NO2 concentration of the two conventional monitors and the mean NO2 concentration observed by all airboxes. Since all airboxes are at different locations and their mean concentrations are not necessarily equal to the mean of the conventional monitors which are both located at a traffic location, we do not require $\Delta_{d,t}$ to be zero. Neither do we require $\Delta_{d,t}$ to be stable throughout the year, as it could possibly vary

with the seasonality of NO2, meteorological conditions or interfering gases. However, the downward trend in $\Delta_{d,t}$ in Fig. 2 suggests sensor drift. After four months of deployment, there is a sudden decrease in $\Delta_{d,t}$, which leads to a systematic decrease of $\sim 10\,\mu g\,m^{-3}$. This is related to a change in the initial calibration factor around the 10th of May. The bias further increases with time. Compared to the average drift of all airboxes in the sensor network, the two collocated airboxes do not show the same decreasing trend. At z_1 the fitted smooth curve of the difference is around zero after the 10th of May, showing no signs of drift. At z_2, the fitted smooth curve follows that of the average drift after the 10th of May until the end of August. Due to malfunctioning of the sensor there are no data available for the last month of the year. The high values at the end of November would therefore strongly influence the fitted curve to increase from September onwards. For a better comparison with z_1 and the mean of all airboxes, the line of z_2 is drawn until September.

3.3. Multiple linear regression model

3.3.1. Calibration

The histogram and Uniform Q-Q plot of the PIT values (Eq. (4)) are created for every daily iteration of the models. A sample is visually inspected to check for uniformity; an example is shown in Fig. S2. The PIT values show uniformity for all models, which means that the models suitably fit the data. The model performance, based on the DIC, can thus be evaluated for the models. In Table 1 we present the DIC values for different models. Since the INLA model is iteratively rebuilt, giving a new DIC value every day, we report the mean DIC and median DIC for 2016. The lowest DIC, indicating the best model fit, is found for model 9 including all covariates. Model 8 (excluding wind direction) has the next lowest DIC, only slightly higher than model 9. All models show a better fit at z_2 compared to z_1.

Fig. 3 shows the change of the coefficients of covariates over time when the model parameters are recalibrated on a daily basis for model 8. We show model 8 (without wind direction) because DIC values are similar to model 9 and wind direction is represented by 9 slopes, increasing the complexity of the figure. The intercept β_0 for the daily INLA models is positive between 3 and 7 $\mu g\,m^{-3}$. At z_1, $\beta_{NO2} < 1$ and at z_2, $\beta_{NO2} > 1$. Transferring the coefficients to another location where the bias is in a different direction will lead to an increase in bias rather than a decrease. Coefficient β_{O3} is negative throughout the year, and β_{RH} is close to zero. Coefficient β_T is mostly positive for both locations. Coefficient β_{WS} shows a pattern close to zero but mostly negative. The month of hourly data used in each calibration iteration should contain enough temporal variability in the covariates to avoid overfitting. However, both locations show a large temporal variability in the coefficients. This is probably due to seasonal variation, as the temporal variability at the two locations is very similar over time. The patterns are smoothened by the overlap of the calibration datasets. When the direction of the slopes would be the same for each location, this would be beneficial for the transferability of the model from one location to the other. However, when coefficients tend to have a different direction at different locations at any point in time, correction may lead to a deterioration.

3.3.2. Validation

A model with only NO_2 improved the RMSE modestly at both locations (Table 2). Adding additional covariates substantially further reduced RMSE. Model 8, with all covariates except wind direction, performed best at location z_1. At z_2, model 9 with all covariates performed best. At both locations, no improvement in RMSE was achieved by correcting the observations using the calibration models built at the other location.

3.4. Daily correction factor

The relative correction factor $\gamma_{rel,d,z}$ shows a higher $RMSE_{post}$ than the absolute correction factor $\gamma_{abs,d,z}$, even exceeding $RMSE_{pre}$ at z_1 (Table 3). Since this method could be applied to all non-missing NO_2 observations at each location, while the Bayesian models could only be applied to observations non-missing for all covariates at each location, the $RMSE_{pre}$ and $RMSE_{post}$ are not directly comparable to those retrieved using INLA (Table 2).

Table 3 also shows the $RMSE_{pre}$ and $RMSE_{post}$ when we tested the correction factor determined at the other airbox location collocated with a conventional monitor. At z_1, RMSE is higher after correction; at z_2 a modest decrease was found using γ_{abs,d,z_1}. The RMSE values can be influenced by a few extremes in the corrected values, especially when an extreme correction factor is established at one location and is transferred to another location. This led to a high $RMSE_{post}$ of 120.86 $\mu g\,m^{-3}$ at z_2. Removing extreme correction factors led to a decrease of this value, however not decreasing below $RMSE_{pre}$. A time series plot of the correction factors illustrates its variability and the extremes (Fig. S3).

3.5. Uniform concentration correction

The results of correction factor γ_{uni} updated under conditions of uniform and low concentrations are shown in Table S3. Depending on the threshold values of standard deviation $\delta_{d,t}$ and mean $\eta_{d,t}$, the number of updates of γ_{uni} in the year ranged between 1 and 39 for the chosen thresholds. For none of the threshold combinations, the

Table 1
DIC performance statistics for different models. A lower DIC denotes better model fit. In each model, the dependent variable is the square root of hourly average reference monitor concentrations of NO_2, $\sqrt{NO2_{ref}}$. $NO2_{ab}$ is NO_2 measured by the low-cost airbox sensor; $O3$ is ozone measured by one reference monitor (z_1). RH is relative humidity and T is temperature, both measured by the airbox. WS is wind speed and WD is wind direction, both measured by the Royal Netherlands Meteorological Institute (KNMI).

#	Covariates	Location z_1		Location z_2	
		DIC mean	DIC median	DIC mean	DIC median
1	$\beta_0 + \beta_{NO2}\sqrt{NO2_{ab}}$	1459	1630	1352	1464
2	$\beta_0 + \beta_{NO2}\sqrt{NO2_{ab}} + \beta_{O3}\log(O3)$	1291	1319	1328	1446
3	$\beta_0 + \beta_{NO2}\sqrt{NO2_{ab}} + \beta_{RH}RH^2$	1444	1619	1312	1439
4	$\beta_0 + \beta_{NO2}\sqrt{NO2_{ab}} + \beta_T T$	1406	1506	1136	1242
5	$\beta_0 + \beta_{NO2}\sqrt{NO2_{ab}} + \beta_{WS}\sqrt{WS}$	1306	1348	1320	1421
6	$\beta_0 + \beta_{NO2}\sqrt{NO2_{ab}} + \beta_{WD}\text{factor(WD)}$	1403	1589	1292	1410
7	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2\sqrt{WS} + \beta_3\log(O3)$	1221	1243	1297	1409
8	$\beta_0 + \beta_{NO2}\sqrt{NO2_{ab}} + \beta_{O3}\log(O3) + \beta_{RH}RH^2 + \beta_T T + \beta_{WS}\sqrt{WS}$	1134	1208	828	877
9	$\beta_0 + \beta_{NO2}\sqrt{NO2_{ab}} + \beta_{O3}\log(O3) + \beta_{RH}RH^2 + \beta_T T + \beta_{WS}\sqrt{WS} + \beta_{WD}\text{factor(WD)}$	1104	1161	778	815

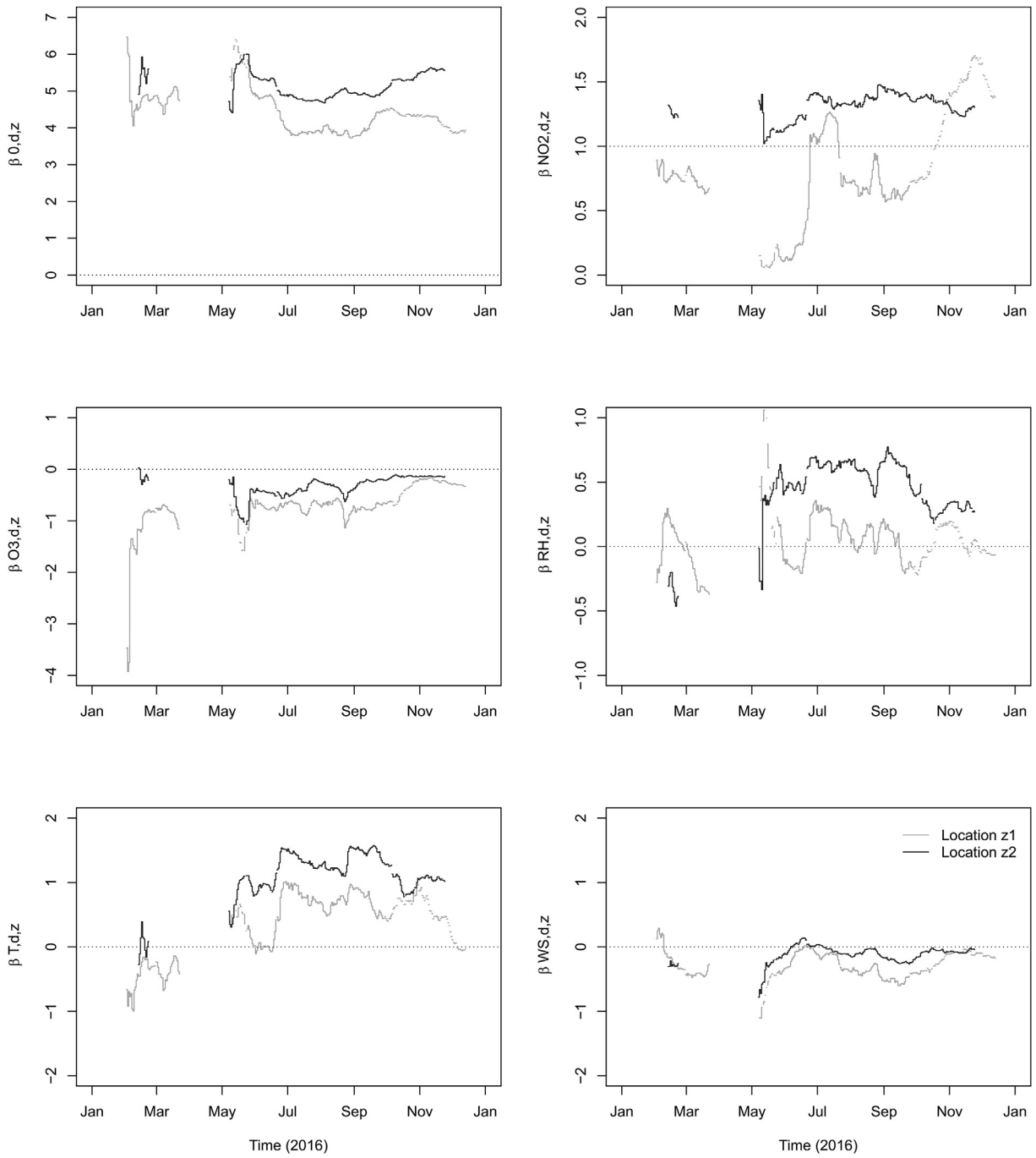


Fig. 3. Time series of the coefficients of the daily INLA models, using model 8. Light gray: location z_1, dark gray: location z_2.

correction method improved the RMSE value. Instead, the $RMSE_{post}$ was 27–145% higher than the $RMSE_{pre}$, making the method not suitable for NO_2 in this sensor network. Fig. S3 shows the time series of the correction factor for $\delta_{d,t} = 5$ µg m^{-3} and $\eta_{d,t} = 12$ µg m^{-3}. As Fig. 2 suggested, a change occurred in May. This is reflected in an update of the correction factor in Fig. S4. Before May the correction factor remains at 1.

Table S4 shows the results of the night-time calibration. At z_1 the RMSE is almost doubled after night-time calibration, while at z_2 there is a slight improvement in RMSE. The increase in RMSE at z_1 is mainly due to some extreme values for $\gamma_{night,d,s}$ in May (Fig. S5).

3.6. Sensitivity of individual airboxes to environmental factors

Fig. 4 shows a boxplot of the posterior mean estimates for relative humidity and temperature for a model with airbox NO_2 as response variable. When using a full year of data at once, the posterior mean estimate for temperature is negative for each airbox, varying between -0.1 and -0.7 for different airboxes. The posterior mean of relative humidity differed between the different airboxes from -0.5 to $+0.3$. A difference in slope direction can have large influence on the transferability of calibration models to other airboxes.

Fig. S6 shows the posterior distributions of β_{NO2}, β_{RH} and β_T for ten collocated airboxes for a shorter period of time with NO_2 from the conventional monitor as the response (240 h). The posterior distributions of β_{NO2} are around 1. For some of the NO_2 sensors, however, the

Table 2
RMSE before and after temporal and spatiotemporal calibration using different models. In each model, the dependent variable is $\sqrt{NO2_{ref}}$.

#	Covariates	Temporal calibration				Spatiotemporal calibration			
		Location z_1 (N = 5684)		Location z_2 (N = 4816)		Model z_2 validated at z_1 (N = 4751)		Model z_1 validated at z_2 (N = 4885)	
		$RMSE_{pre}$	$RMSE_{post}$	$RMSE_{pre}$	$RMSE_{post}$	$RMSE_{pre}$	$RMSE_{post}$	$RMSE_{pre}$	$RMSE_{post}$
1	$\beta_0 + \beta_1\sqrt{NO2_{ab}}$	9.91	9.44	10.67	8.54	10.00	11.35	10.64	11.39
2	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2\log(O3)$	9.91	8.43	10.67	8.43	10.00	10.81	10.64	11.68
3	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2 RH^2$	9.91	9.52	10.67	8.22	10.00	11.86	10.64	12.70
4	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2 T$	9.91	8.83	10.67	7.26	10.00	12.70	10.64	11.62
5	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2\sqrt{WS}$	9.91	8.05	10.67	8.28	10.00	10.29	10.64	11.10
6	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2\text{factor(WD)}$	9.91	9.47	10.67	8.62	10.00	11.55	10.64	11.56
7	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2\sqrt{WS} + \beta_3\log(O3)$	9.91	7.97	10.67	8.23	10.00	10.27	10.64	11.42
8	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2\log(O3) + \beta_3 RH^2 + \beta_4 T + \beta_5\sqrt{WS}$	9.91	7.62	10.67	5.80	10.00	11.28	10.64	11.55
9	$\beta_0 + \beta_1\sqrt{NO2_{ab}} + \beta_2\log(O3) + \beta_3 RH^2 + \beta_4 T + \beta_5\sqrt{WS} + \beta_6\text{factor(WD)}$	9.91	7.80	10.67	5.74	10.00	11.66	10.64	11.58

beta coefficient was below 1 while for other airboxes the beta coefficient was above 1. The NO_2 concentrations would thus be corrected in the wrong direction when using the estimated coefficients of another airbox. β_{RH} and β_T have posterior distributions around zero, indicating different slope directions for different airboxes. We note that the short duration may have contributed to some uncertainty in the estimates.

3.7. Temporal autocorrelation

In the iterative calibration procedure, the dataset for calibration was not large enough to model the temporal autocorrelation in the NO_2 data. Inclusion of the random walk component (Eq. (13)) did not lead to improvements in $RMSE_{post}$. When applying the model in Eq. (2) to the full year dataset, however, there were clear signs of temporal autocorrelation in the residuals. Including random effects in the model using Eq. (13) led to a significant decrease in $RMSE_{post}$ from 8.30 to 3.12 at z_1, and from 6.76 to 3.71 at z_2. A complete overview is given in Table S5. Inclusion of random effects narrowed the scatterplot closer to the 1:1 line (Fig. S7). A substantial decrease in residuals and removal of the temporal pattern is visible in the residual plot (Fig. S8).

4. Discussion and conclusions

After approximately two to six months after the initial calibration, the airbox NO_2 sensors showed signs of drift. We have evaluated three different methods for regular calibration: daily updated correction factors, corrections based on uniform low concentrations, and a Bayesian regression model. The Bayesian regression model and the daily correction factors both worked very well on the airbox for which they were created, accounting for both systematic bias due to drift and non-systematic errors due to interference effects. However, we found that the transferability of the correction parameters and coefficients to another airbox was limited, though the other airbox was within a short distance and in a similar traffic situation. The poor spatial transferability is consistent with the different sensitivity of individual airboxes to environmental factors including temperature and relative humidity, in agreement with Broday (2017).

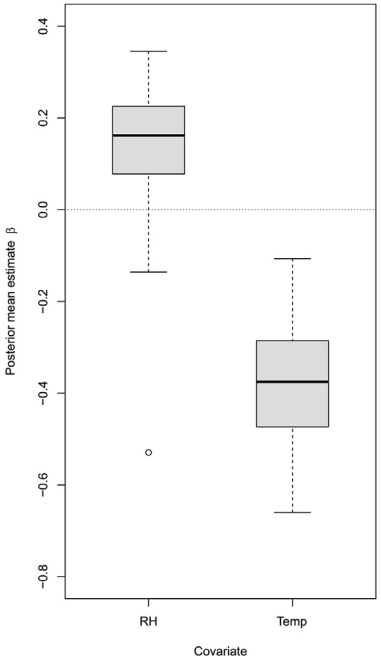


Fig. 4. Posterior mean estimates of different airboxes, for airbox NO_2 vs. covariates RH and temperature measured in the same airbox, full year 2016.

The sensitivity of electrochemical cell NO_2 sensors to temperature and relative humidity has already been evaluated for different types of sensors (Mead et al., 2013; Neri et al., 2002; Phala et al., 2016). With a set of collocated sensors, we found that the interference effects might be

Table 3
RMSE values before and after applying a daily correction factor on hourly values at the same location (temporal calibration) and at the other location (spatiotemporal calibration).

Correction factor	Temporal calibration				Spatiotemporal calibration			
	Location z_1		Location z_2		Model z_2 validated at z_1		Model z_1 validated at z_2	
	$RMSE_{pre}$	$RMSE_{post}$	$RMSE_{pre}$	$RMSE_{post}$	$RMSE_{pre}$	$RMSE_{post}$	$RMSE_{pre}$	$RMSE_{post}$
$\gamma_{rel,d,z}$	9.94	19.57	10.55	7.34	9.46	13.53	10.68	120.86
$\gamma_{abs,d,z}$	9.94	6.54	10.55	5.78	9.46	10.17	10.68	9.75

different for individual sensors of the same type when deployed in an outdoor environment. We also showed, in line with Kizel et al. (2018), that for each airbox the coefficients strongly vary over time, emphasizing the need for regular recalibration. These results may be different for other pollutants or sensors. Zalel et al. (2015), for example, found a good temporal stability and spatial transferability for benzo(a)pyrene.

Wind speed data was only available at a single location in Eindhoven. In the model we could therefore only include the temporal variation in wind speed but not its spatial variation. Lerner et al. (2015) found that local wind speed affects the NO_2 concentrations. The NO_2 sensors in our study are covered by the airbox to minimize the influence of direct wind. The regression model may be improved by adding measurements of wind speed and wind direction on each airbox location. Ozone data are measured by the airboxes, but were only available for two months in 2016 for the two airboxes collocated with a conventional monitor, due to sensor failure. Therefore we used O_3 data available at one of the conventional monitor sites. This limited the O_3 data to only one location, but the relation between NO_2 and O_3 is strong and the temporal variability in O_3 is typically higher than the spatial variability. This solution was therefore preferred over using no O_3 data or using only two months of data for analysis. Due to a change in initial calibration factor, there is a sudden change in NO_2 concentrations around the 10th of May. This has affected the models based on 30-day periods including May 10th, but not other periods. The change thus does not affect our general conclusions.

We evaluated the performance of the calibration models, their temporal stability and spatial transferability by comparing the RMSE values before and after calibration. This measure is widely used to evaluate sensor performance, but can be influenced by extreme values (Fishbain et al., 2017). We cleaned the data from outliers before the analysis to minimize the influence of extreme values on calibration parameters and their performance (Van Zoest et al., 2018). Besides the RMSE, other performance measures can be used to assess the quality and usability of low-cost air quality sensors. For example, Fishbain et al. (2017) developed a tool kit to evaluate the performance of air quality micro-sensing units. Here, our focus is on calibration performance only.

Miskell et al. (2018) suggested to calibrate low-cost sensors using a conventional monitor in the vicinity with similar land use type. They successfully applied the method on O_3 with an averaging time of 72 h. We applied a similar method on NO_2 using daily absolute and relative correction factors and an averaging time of 24 h. This method accounts for drift and daily variability in interference effects. Despite similar traffic conditions at both locations, and traffic being the major contributor to NO_2 levels in Eindhoven, the correction factors could not successfully be transferred from one location to the other. This is probably due to variability between individual airboxes in the strength of drift or interference.

The uniform concentration correction method was proposed by Tsujita et al. (2005) and only adjusts the correction factor when concentrations of pollutants are uniform over the city. This method accounts mostly for drift but also for long-term variability in interference effects. Moltchanov et al. (2015) applied this method using nighttime calibrations (1:00–4:00) when O_3 concentrations are uniformly negligible. For NO_2 they could not apply the method because they did not find periods of sufficiently long duration with negligible spatial variation in NO_2 concentrations. We faced the same issue for NO_2.

We evaluated the presence of temporal autocorrelation in the residuals by adding a first-order random walk component to the model, as described in Blangiardo and Cameletti (2015). In our case the addition of this random effect only led to an improvement in RMSE when applied on 80% of the full dataset (2279 hourly non-missing observations) rather than iteratively using the hourly observations of the previous 30 days (maximum 720 hourly non-missing observations).

The need for regular recalibration of parameters is clear. The time series plots of the daily correction factors, the correction factors based on uniform low concentrations, and β coefficients of the INLA model show that independent of the method used, there is significant variability in the correction factors and parameters over time. Also, the parameters are dependent on the individual sensor. Hasenfratz et al. (2012) proposed an on-the-fly calibration procedure for gas sensors mounted on public transport vehicles, calibrating the sensors when in each other's vicinity or when in the vicinity of a conventional monitor. In static sensor networks, a moving reference sensor could be used for regular calibration and data quality evaluation of the sensors in the network as suggested by Kizel et al. (2018). A moving reference sensor takes the different response of individual sensors into account, and would be a suitable solution to account for the spatio-temporal variability in the calibration parameters. A disadvantage is the added workload. Besides a moving reference sensor, it would still be of added value to collocate the sensors once a year to compare performance differences.

Low-cost air quality networks provide data of a fine spatial and temporal resolution. They provide valuable opportunities for spatio-temporal modelling and health risk mapping. It can be debated whether one should use modelled values, as derived from the calibration model, as an input for spatiotemporal modelling purposes and health studies, in which the same covariates are likely to be used again as potential confounders in for example time series studies. The purpose of this model is however different, the calibration parameters are optimized for the calibration of NO_2, and the covariates have been transformed and scaled. Therefore we do not expect major issues regarding the use of calibrated values in future modelling.

Calibration procedures are important to correct air pollution data before online publishing, to avoid misinterpretation of the results. This study has highlighted the need for such calibration procedure to not just account for drift and interference effects, but also for the variability in drift and interference effects in space, time and between sensors. The strength of this variability may differ between locations, pollutants and sensors used. Transferability of calibration parameters from one sensor to the other and similarities in drift are often assumed, but this assumption is not always justified. Regular calibration should therefore be performed at the location of the low-cost sensor, for example using a moving reference sensor.

Low-cost air quality sensors are valuable instruments to increase the spatial and temporal resolution of air quality sensor networks. When aware of their limitations, sensor-specific differences and when communicating the uncertainties related to their measurements, they could prove useful in various settings.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Acknowledgements

This work was supported by the Netherlands Organisation for Scientific Research. The authors would like to thank Nicholas Hamm for his initial contributions. The authors would also like to thank the other members of the AiREAS calibration/validation team for the inspiring discussions and good collaboration.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2019.04.048.

References

Barakeh, Z.A., Breuil, P., Redon, N., Pijolat, C., Locoge, N., Viricelle, J.P., 2016. Development of a normalized multi-sensors system for low cost on-line atmospheric pollution detection. Sensor. Actuator. B Chem. 241, 1235–1243. https://doi.org/10.1016/j.snb.2016.10.006.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Bayesian Models with R-INLA. John Wiley & Sons, Chichester.

Broday, D.M., 2017. & the citi-sense project collaborators. Wireless Distrib. Environ. Sens. Netw. Air Poll. Meas. Prom.Curr. Real. Sens. 17 (10), 2263.

Buijsman, E., 2013. De luchtkwaliteitsmeetnetten van het RIV en het RIVM. Houten: Tinsentiep.

Close, J.P. (Ed.), 2016. AiREAS: Sustainocracy for a Healthy City. The Invisible Made Visible Phase 1. Springer.

Colin, F., Shepherd, P.D., Carter, T.J.N., Wright, J.D., 1998. Development of a piezo-optical chemical monitoring system for nitrogen dioxide. Sensor. Actuator. B Chem. 51 (1–3), 244–248. https://doi.org/10.1016/s0925-4005(98)00201-9.

Cyrys, J., Eeftens, M., Heinrich, J., Ampe, C., Armengaud, A., Beelen, R., Bellander, T., Beregszaszi, T., Birk, M., Cesaroni, G., Cirach, M., de Hoogh, K., De Nazelle, A., de Vocht, F., Declercq, C., Dėdelė, A., Dimakopoulou, K., Eriksen, K., Galassi, C., Grąulevičienė, R., Grivas, G., Gruzieva, O., Gustafsson, A.H., Hoffmann, B., Iakovides, M., Ineichen, A., Krämer, U., Lanki, T., Lozano, P., Madsen, C., Meliefste, K., Modig, L., Mölter, A., Mosler, G., Nieuwenhuijsen, M., Nonnemacher, M., Oldenwening, M., Peters, A., Pontet, S., Probst-Hensch, N., Quass, U., Raaschou-Nielsen, O., Ranzi, A., Sugiri, D., Stephanou, E.G., Taimisto, P., Tsai, M.-Y., Vaskövi, É., Villani, S., Wang, M., Brunekreef, B., Hoek, G., 2012. Variation of NO2 and NOx concentrations between and within 36 European study areas: results from the ESCAPE study. Atmos. Environ. 62, 374–390. https://doi.org/10.1016/j.atmosenv.2012.07.080.

De Vito, S., Piga, M., Martinotto, L., Di Francia, G., 2009. CO, NO2 and NOx urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. Sensor. Actuator. B Chem. 143 (1), 182–191. https://doi.org/10.1016/j.snb.2009.08.041.

Fishbain, B., Lerner, U., Castell, N., Cole-Hunter, T., Popoola, O., Broday, D.M., Iñiguez, T.M., Nieuwenhuijsen, M., Jovasevic-Stojanovic, M., Topalovic, D., Jones, R.L., Galea, K.S., Etzion, Y., Kizel, F., Golumbic, Y.N., Baram-Tsabari, A., Yacobi, T., Drahler, D., Robinson, J.A., Kocman, D., Horvat, M., Svecova, V., Arpaci, A., Bartonova, A., 2017. An evaluation tool kit of air quality micro-sensing units. Sci. Total Environ. 575, 639–648. https://doi.org/10.1016/j.scitotenv.2016.09.061.

Hamm, N.A.S., Van Lochem, M., Hoek, G., Otjes, R., Van der Sterren, S., Verhoeven, H., 2016. The invisible made visible": science and technology. In: In: Close, J.P. (Ed.), AiREAS: Sustainocracy for a Healthy City. The Invisible Made Visible Phase, vol 1. Springer, pp. 51–78.

Hasenfratz, D., Saukh, O., Thiele, L., 2012. On-the-fly calibration of low-cost gas sensors. In: Paper presented at the European Conference on Wireless Sensor Networks, Trento, Italy.

Jerrett, M., Donaire-Gonzalez, D., Popoola, O., Jones, R., Cohen, R.C., Almanza, E., de Nazelle, A., Mead, I., Carrasco-Turigas, G., Cole-Hunter, T., Triguero-Mas, M., Seto, E., Nieuwenhuijsen, M., 2017. Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science. Environ. Res. 158, 286–294. https://doi.org/10.1016/j.envres.2017.04.023.

Kamionka, M., Breuil, P., Pijolat, C., 2006. Calibration of a multivariate gas sensing device for atmospheric pollution measurement. Sensor. Actuator. B Chem. 118 (1), 323–327. https://doi.org/10.1016/j.snb.2006.04.058.

Kizel, F., Etzion, Y., Shafran-Nathan, R., Levy, I., Fishbain, B., Bartonova, A., Broday, D.M., 2018. Node-to-node field calibration of wireless distributed air pollution sensor network. Environ. Pollut. 233, 900–909. https://doi.org/10.1016/j.envpol.2017.09.042.

KNMI, 2016, 31/08/2016. Uurgegevens Van Het Weer in Nederland - Download. Retrieved from. http://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi.

Lerner, U., Yacobi, T., Levy, I., Moltchanov, S.A., Cole-Hunter, T., Fishbain, B., 2015. The effect of ego-motion on environmental monitoring. Sci. Total Environ. 533, 8–16. https://doi.org/10.1016/j.scitotenv.2015.06.066.

Lewis, A., Edwards, P., 2016. Validate personal air-pollution sensors. Nat. Comm. 535 (7610), 29–31. https://doi.org/10.1038/535029a.

Martins, T.G., Simpson, D., Lindgren, F., Rue, H., 2013. Bayesian computing with INLA: new features. Comput. Stat. Data Anal. 67, 68–83. https://doi.org/10.1016/j.csda.2013.04.014.

Mead, M.I., Popoola, O.A.M., Stewart, G.B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J.J., McLeod, M.W., Hodgson, T.F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J.R., Jones, R.L., 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos. Environ. 70, 186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060.

Miskell, G., Salmond, J.A., Williams, D.E., 2018. Solution to the problem of calibration of low-cost air quality measurement sensors in networks. ACS Sens. 3 (4), 832–843. https://doi.org/10.1021/acssensors.8b00074.

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. Sci. Total Environ. 502, 537–547. https://doi.org/10.1016/j.scitotenv.2014.09.059.

Morales, J.A., Walsh, J.E., Treacy, J., Garland, W.E., 2002. Miniaturized differential optical absorption spectroscopy (DOAS) system for the analysis of NO2. In: Paper presented at the Opto-Ireland 2002: Optics and Photonics Technologies and Applications, vol. 4876. pp. 1229–1235. https://doi.org/10.1117/12.463920 Galway, Ireland.

Neri, G., Bonavita, A., Galvagno, S., Siciliano, P., Capone, S., 2002. CO and NO2 sensing properties of doped-Fe2O3 thin films prepared by LPD. Sensor. Actuator. B Chem. 82 (1), 40–47. https://doi.org/10.1016/s0925-4005(01)00987-x.

Penza, M., Martucci, C., Cassano, G., 1998. NOx gas sensing characteristics of WO3 thin films activated by noble metals (Pd, Pt, Au) layers. Sensor. Actuator. B Chem. 50 (1), 52–59. https://doi.org/10.1016/s0925-4005(98)00156-7.

Phala, K.S.E., Kumar, A., Hancke, G.P., 2016. Air quality monitoring system based on ISO/IEC/IEEE 21451 standards. IEEE Sens. J. 16 (12), 5037–5045. https://doi.org/10.1109/JSEN.2016.2555935.

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., Li, K., Dick, R.P., Lv, Q., Hannigan, M., Shang, L., 2014. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. Atmosph. Meas. Tech. 7 (10), 3325–3336. https://doi.org/10.5194/amt-7-3325-2014.

Rasch, D., Tiku, M.L., Sumpf, D., 1994. Elsevier's Dictionary of Biometry. Elsevier, Amsterdam.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B 71 (2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Santos, J., Serrini, P., O'Beirn, B., Manes, L., 1997. A thin film SnO2 gas sensor selective to ultra-low NO2 concentrations in air. Sensor. Actuator. B Chem. 43 (1–3), 154–160. https://doi.org/10.1016/s0925-4005(97)00115-9.

Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A., Bartonova, A., 2017. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environ. Int. 106, 234–247. https://doi.org/10.1016/j.envint.2017.05.005.

Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S., Shelow, D., Hindin, D.A., Kilaru, V.J., Preuss, P.W., 2013. The changing paradigm of air pollution monitoring. Environ. Sci. Technol. 47 (20), 11369–11377. https://doi.org/10.1021/es4022602.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A.V.D., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B 64 (4), 583–639. https://doi.org/10.1111/1467-9868.00353.

Spinelle, L., Gerboles, M., Villani, M.G., Aleixandre, M., Bonavitacola, F., 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: ozone and nitrogen dioxide. Sensor. Actuator. B Chem. 215, 249–257. https://doi.org/10.1016/j.snb.2015.03.031.

Tsujita, W., Yoshino, A., Ishida, H., Moriizumi, T., 2005. Gas sensor network for air-pollution monitoring. Sensor. Actuator. B Chem. 110 (2), 304–311. https://doi.org/10.1016/j.snb.2005.02.008.

Van Zoest, V.M., Stein, A., Hoek, G., 2018. Outlier detection in urban air quality sensor networks. Water Air Soil Poll. 229 (4), 111. https://doi.org/10.1007/s11270-018-3756-7.

Wang, X., Yue, Y.R., Faraway, J., 2018. Bayesian Regression Modeling with INLA. Chapman and Hall/CRC, New York.

Xiang, Y., Tang, Y., Zhu, W., 2016. Mobile sensor network noise reduction and recalibration using a Bayesian network. Atmosph. Meas. Tech. 9 (2), 347–357. https://doi.org/10.5194/amt-9-347-2016.

Zalel, A., Yuval, Svecova, V., Sram, R.J., Bartonova, A., Broday, D.M., 2015. Modeling airborne benzo(a)pyrene concentrations in the Czech Republic. Atmos. Environ. 101, 166–176. https://doi.org/10.1016/j.atmosenv.2014.11.031.