Atmospheric Environment 343 (2025) 120990

Contents lists available at ScienceDirect

Atmospheric Environment

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

Optimal measurement strategy for air quality combining official and low-cost measurements

Ronald Hoogerbrugge, Sjoerd van Ratingen, Koen Siteur, Joost Wesseling *

National Institute for Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA, Bilthoven, the Netherlands

HIGHLIGHTS

- Optimizing the combined use of model results, reference - and low-cost measurements.
- Combination of citizen science and operational monitoring of air quality.
- Minimum uncertainty of model after bias removal.

ARTICLE INFO

Keywords:
Air quality modelling
Measurement strategy
Bias removal
Calibration uncertainty
Citizen science

ABSTRACT

Air pollution affects the health of people and therefore monitoring of the air quality is important both for the public and policy makers. Efficient monitoring of air quality requires a combination of measurements and modelling. Both current and annual average concentrations as well as future concentrations on all locations where people live are required. This information on exposure to pollutants can only be achieved at high spatial resolution at all locations by using air-quality models. Therefore, model calibration is a major objective in air quality measurement strategies. Measurement results of reference instruments (or equivalent) as defined in the EU air quality directive offer a high-quality basis for model calibration and validation. Over the last years, low-cost sensors/samplers have shown a rapid development and promising results. In this paper, a statistical framework is presented to evaluate measurement strategies that apply a combination of reference measurement instruments and low-cost measurements, like diffusion tubes and sensors. For some practical situations the introduction of sensors at only twenty locations gives a significant improvement of the calibration of an air quality model. The calibration of the low-cost measurements themselves with respect to the reference instruments is critical for any application. This calibration largely determines the model quality improvement due to the addition of low-cost measurements. The results shown in this paper can be used to optimize measurement strategy using low-cost measurements and/or sensors with established performance characteristics. The results can also be used to define the quality of the low-cost measurements that is required for useful applications. Using low-cost measurements can improve the quality of the calibrated model, even with a simultaneous reduction of the number of reference instruments. I.e., improved quality of information, at reduced costs.

1. Introduction

Air quality information is important because of the impact of air pollution on the health of people. In the European Union, legal limit values are set for ambient levels for a number of air pollutants (EC, 2008). All countries in the EU must report on the concentration levels in their territories and ensure that the levels of air pollutants are below the relevant limit values. When exceedances are observed or expected, plans have to be developed to ensure the exceedances are prevented or eliminated as soon as possible. The concentration levels of the present legal limit values are no guarantee that there are no adverse health effects due to exposure to these levels. According to the World Health Organisation, much stricter target values are needed to avoid health effects (WHO, 2021). The European Commission has announced a new concept Ambient Air Quality Directive that aims for much lower legal limit values (EC, 2022). The new Ambient Air Quality Directive has recently been accepted by the European Parliament and is expected to come into force in the EU in 2024. In order to assess air quality

* Corresponding author.
E-mail address: joost.wesseling@rivm.nl (J. Wesseling).

https://doi.org/10.1016/j.atmosenv.2024.120990
Received 5 July 2024; Received in revised form 8 November 2024; Accepted 11 December 2024
Available online 16 December 2024

everywhere in a country, measurements alone do not suffice and air quality models are employed. As with all models, results of air quality models are subject to uncertainties and possible biases. Determining the quality of models is therefore very important. The European FAIRMODE project focusses on providing methodologies to establish a harmonised way to assess model quality, based on a comparison with measurements (Janssen, 2020; Janssen, 2022; Thunis et al., 2022). So, both for direct air quality assessment and for providing data for model comparisons, measurements are essential to determine the air quality objectively. However, ambient air quality measurements are usually only representative for a relatively small area around the measurement location. For application to all other locations, and the people living there, model calculations are essential. For adequate determination of population exposure in a country, calibration and validation of air quality models may be objectives of a measurement network. In this article, the achievable accuracy of model calibration is assumed to be a relevant and objective indicator for the evaluation of the performance of measurement and monitoring strategies. Therefore, the maximum achievable accuracy of calibrated models is essential in the evaluation and design of measurement and monitoring strategies.

The air quality directive (EU, 2008) describes the required minimum number and quality of the measurement system, i.e. the number and uncertainty of measurements and locations for all relevant substances. All countries in the EU provide a yearly description of the national measuring networks, for the Netherlands the last description was provided by (Coolen et al., 2023). For some substances, like nitrogen dioxide (NO_2) and ammonia (NH_3), alternative low-cost measurements using diffusion tubes have been available for a long time (Palmes et al., 1976). The quality and applicability of these low-cost measurements have been extensively investigated and reported (Bush et al., 2001; Hafkenscheid et al., 2009; Heal et al., 2019). Palmes tubes have been used in several large-scale experiments involving citizens to facilitate and sometimes also perform the measurements (De Craemer et al., 2020; Lauriks et al., 2022).

In recent years, new types of measurement equipment, like low-cost digital sensors for different sizes of particulate matter $PM_{2.5}$, PM_{10} and gasses like NO_2, are developed. This raises the opportunity to develop alternative measurement and monitoring strategies that go beyond the use of fixed reference measurements only. In several cases the alternative measurements are setup by professional organisations in the field of air quality using large groups of citizens to facilitate and perform measurements and sometimes help analyse the results (Irwin, 2018; Wesseling et al., 2019; Woutersen et al., 2022). For particulate matter, both stationary and mobile measurements are performed using digital sensors (Gressent, 2019; Wesseling et al., 2021). Although the digital PM sensors have substantial quality issues, several calibration strategies have been developed and tested to deal with issues like the effects of environmental conditions on the measurements (Wesseling et al., 2024). For NO_2, several digital sensors are available, but an accurate calibration is much harder to achieve than with PM sensors (Ratingen et al., 2021).

There have been many studies on the design of measurement networks, a recent review of which is presented by Verghese and Nema (2022). In many of these studies, relatively advanced statistical techniques are used, like mixed integer linear programming, multivariate statistics, geostatistical techniques and information gain and entropy. Most of the studies focus on the optimal spatial distribution of (new) air quality measurement sites, and assume uniform measurement hardware. In some cases, use of high-end precision measurement equipment is found to be optimal, whereas for the assessment of an areal mean concentration lower-cost measurements seem more effective.

In our approach, we consider existing official networks, where (usually many) low-cost measurements are added to a limited set of high-end official reference measurements. The combined information is used to calibrate the results of air quality models. Although the objective of model calibration has been included in studies on optimizing measurement network design, networks combining low-cost measurements and high-end official reference measurements for model calibration have not been considered in these studies. Pickett and Whiting (1981) did study the cost effectiveness of using either only low-cost, low precision monitors, or only high-cost, higher precision monitors in the contexts of model calibration performance. Although this may be relevant for cases where a new measurement network is designed from scratch, it is less relevant in most practical cases where an existing network of high-end reference measurements is extended with low-cost measurements. In our work we seek for an optimal combination of reference and low-cost measurements, given the uncertainties in both measurement methods and model results, and using relatively simple statistics. Summarizing, what are the optimal combinations of these measurements?

In this paper we will show how relatively standard application of well-known statistics can be used to calculate the uncertainties of new combinations of measurement data. We will show how official reference measurements can be combined with different numbers and configurations of low-cost measurements to yield the overall best calibration of model results. We first briefly describe some statistical relations between calibration accuracy and measurement network design (the details are provided in Appendix A). Next, the main relation between measurement strategy and calibration accuracy will be used to evaluate several measurement strategies for two types of low-cost measurements. A low-cost measurement can, in our analysis, be an electronic measuring device producing real-time data as well as a passive sampler producing, for example, weekly or monthly time series of concentrations. Most measurement campaigns using low-cost measurements and input from citizens have a relatively short running time, many last for one or a few years. Some campaigns perform measurements for only one month and try to generalize the results to a yearly average concentration. Good examples of large-scale experiments were the recent Belgian “Curieuze Neuzen” campaigns, with 3000 up to 20000 measurement locations in Belgium (De Craemer et al., 2020).

Our focus is on the use of low-cost measurements in addition to official measurements, mainly to obtain long-term (yearly/annual) average concentrations. The statistical performance of alternative measurement strategies will be calculated and compared to the performance of the official measurement requirement. This statistical approach enables optimization of the measurement system with respect to both the costs and the quality/performance.

Although the main application of the strategies discussed here is in a country with a relatively dense measurement network, the ideas for optimization may also be applied in areas with only limited numbers of official reference measurements and little budget to expand the measurements.

2. Relation between measurement accuracy and model calibration

For any quantitative or statistical comparison between measurement strategies, one needs a minimum set of performance indicators. In this paper we look at three key elements in the measurement strategies for determining air quality: 1) reference instruments, 2) low-cost measurements (sensors) and 3) air quality models. For each of these elements known random uncertainties are assumed:

Standard deviation of the random uncertainty of the low-cost measurement (E) s_E

Standard deviation of the random uncertainty of the model (M) s_M

Standard deviation of the random uncertainty of the reference instrument (R) s_R

In this paper the uncertainty of the model for a specific location is interpreted as the overall standard deviation of the uncertainty of the model results, which is a combination of the mathematical algorithms that are used in the model and the necessary emission data and other relevant input data. The area being studied is assumed to be within the applicable range of the model. We assume that reference instruments are available at N locations in the domain of interest. At K of these N locations both sensor measurements and model results are available. There are in total L low-cost measurement devices/sensors available. The low-cost measurement devices/sensors will be calibrated using the reference instruments and the model will be calibrated using both low-cost data and reference instrument data.

Throughout the article, we will regularly use “uncertainty” as a shortcut to “standard deviation of the random uncertainty”. In most of the formula’s the square of the standard deviation is used, i.e. the variance. Note that all relations are valid for different averaging times. If the random uncertainties of the low-cost measurement, the reference measurements and the model are all for hourly values or annual values, all results presented here are for the same averaging period. In the examples we will use yearly/annual uncertainties for the measurements and model results.

2.1. Measurements and uncertainties

The statistical framework presented in this article, to combine different types of measurements and model results, can be applied to real-life situations. We use the framework to obtain the combined uncertainties for a number of practical examples. In this section we discuss

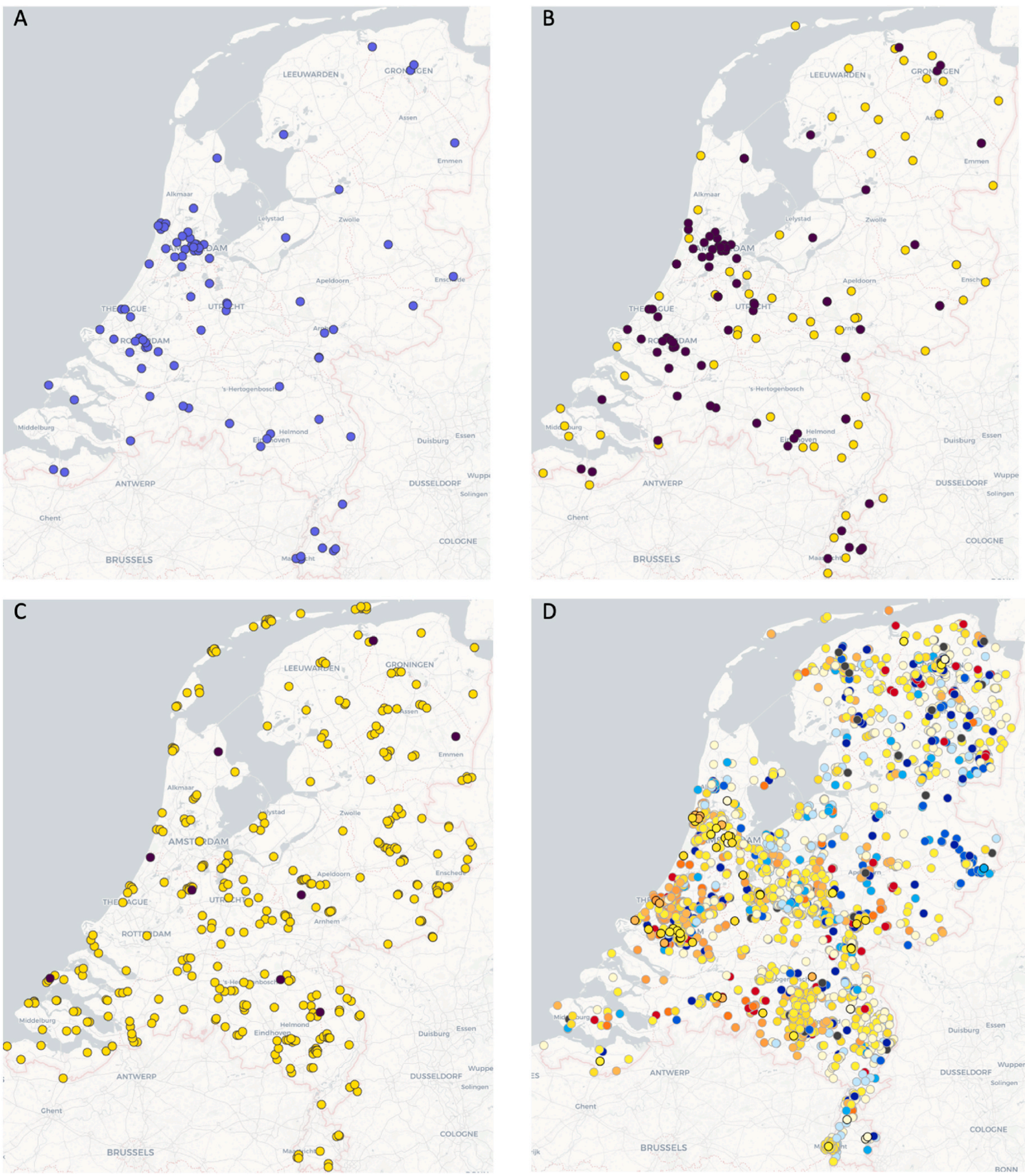

Fig. 1. Air quality networks in the Netherlands. A) Locations of the NO_2, NH_3, PM10 and PM2.5 hourly reference measurements in the Dutch National Air Quality Monitoring Network (LML). B) NO_2 Palmes tubes in support of the national modelling (yellow), reference locations are also shown (brown). C) NH_3 diffusion tubes in support of modelling of nitrogen deposition (yellow). Locations of reference measurements of NH_3 are also shown (brown). D) PM2.5 sensor measurements (hourly) provided by both institutional projects and citizen projects, with colors between blue (0 μg/m^3) and red (60 μg/m^3), status November 05, 2024 at 19:00. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the uncertainties that are used in the examples.

An overview of the locations where concentrations of NO_2, NH_3, PM10 and PM2.5 are measured in the Netherlands is provided in Fig. 1A. The passive measurements of NO_2 (Fig. 1B) and NH_3 (Fig. 1C), that are described in this article, also cover the whole country. The website https://sensors.rivm.nl/, run by RIVM, continuously shows the results of the low-cost sensors (Fig. 1D) as well as separate pages with the results of passive samples in the Netherlands.

2.1.1. Official measurements

Measurements performed by official authorities, like the Dutch National Institute for Public Health and the Environment (RIVM), are performed according to strict criteria, specified in official EU regulations (EU, 2008). The measurements can also be performed using other methods, that are subsequently shown (using prescribed methods) to be *equivalent* to the reference methods. Practical calibration methods used in official networks guarantee unbiased results. There remain random uncertainties, that are required not to exceed prescribed values. The official authorities are obligated to test and report the quality of the official measurements. Usually, the practical uncertainties are less than the maximum allowed values. The 95% confidence intervals of the air quality measurements performed by RIVM in the National Air Quality Monitoring Network (www.luchtmeetnet.nl), have been determined using the methods described in the European Directive and additional regulations. Fig. 1A also shows the measurement sites of regional networks (DCMR and GGD-Amsterdam) with comparable performance.

Component	Period	Concentration level (μg/m^3)	Uncertainty (95%CI)	Remarks
NO_2	Hour	200	7.2%	Mooibroek (2014)
NO_2	Year	40	9.1%	Mooibroek (2014)
PM10	Day	50	17%	Automatic equivalent (Mooibroek, 2014)
PM10	Day	50	7.7%	Mooibroek (2014)
PM10	Year	40	6.2%	Mooibroek (2014)
PM2.5	Day	30	16%	(Hafkenscheid, 2014)
PM2.5	Year	25	9.3%	Mooibroek (2014)

Because the instruments and methods applied in the Netherlands comply with the regulations from the EU directive, the statistical data are assumed to be representative for many other European countries.

2.1.2. Low-cost diffusion tubes for NO_2

Palmes tubes are well known small plastic tubes containing a chemical reagent which can be used to measure nitrogen dioxide concentration. These tubes can be used to measure nitrogen dioxides (NO_2) in a simple way at locations where there is no official, continuous measurement (Palmes et al., 1976). As was mentioned before, the quality and applicability of these low-cost measurements have been extensively investigated and reported (Bush et al., 2001; Hafkenscheid et al., 2009; Heal et al., 2019). Palmes tubes have been used in several large-scale experiments involving citizens to facilitate and sometimes also perform the measurements (De Craemer et al., 2020; Lauriks et al., 2022). In a set of tests performed at RIVM (Nguyen, 2016) results of Palmes tubes were compared to those of official measurements. The uncertainty of individual measurements with Palmes tubes was estimated at 24.2% (95%CI). This was for the practical case that volunteers were used to perform the actual measurements.

Component	Period	Concentration level (μg/m^3)	Uncertainty (95%CI)	Remarks
NO_2	Year	20–40	24.2	Diffusion tube (Nguyen, 2016)

The measurement technique using Palmes tubes are relatively simple and commercially available. Therefore, the uncertainty data found in The Netherland are expected to be feasible in many other countries. The quality can be assessed by using sufficient representative locations with co-located reference instruments and Palmes tubes. Presently, there are some 60 measurement locations run by RIVM, using Palmes tubes.

2.1.3. Low-cost diffusion tubes for NH_3

In the Dutch MAN (Ammonia in Nature Areas) Measurement Network, run by RIVM, air concentrations of ammonia are measured using passive samplers. The measurement network was established in 2005 and now consists of 110 areas and 330 measurement locations. Noordijk (2020) reported the relevant measurement uncertainties of the ammonia measurements and monitoring. Comparing multiple passive samplers at the same location yields a standard deviation of random uncertainty on a monthly basis of $\sqrt{0.8^2 + 0.21^2 x^2}$, with x the measured concentration. Combining 12 of these random uncertainties results in an annual uncertainty of approximately $s_E = \sqrt{0.3^2 + 0.06^2 x^2}$. The average concentration of the whole measurement network is 4.7 μg/m^3, resulting in a typical standard deviation of the uncertainty of $s_E = 10\%$. The standard deviation of the uncertainty of the reference instrument (s_R) on a yearly basis is estimated as 3.4 %, which is quite comparable with the uncertainty of a NO2 reference instrument (5 %).

Component	Period	Concentration level (μg/m^3)	Uncertainty (95%CI)	Remarks
NH_3	Year	5	6.8	Reference NH_3 (Blank, 2001)
NH_3	Year	5	20	Diffusion tube (Noordijk, 2020)

2.1.4. Low-cost digital samplers/sensors for PM2.5 and PM10

In the Netherlands, RIVM has been running an infrastructure making it possible for citizens and other parties to provide and display data from their own low-cost sensors measuring PM2.5 and PM10 (Wesseling et al., 2019), see https://sensors.rivm.nl/. Most of the sensors that are presently used in the Netherlands are of the type SDS011 from Nova Fitness Co., Ltd. There are some 1800 sensors of this type, providing concentration data every 5–15 min (status summer 2024). These sensors are popular in Europe and are readily available in webstores. For a recent benchmark on sensor calibration (Wesseling et al., 2024), only the SDS011 sensors were used. This has resulted in extensive experience with the uncertainties involving the SDS011 sensors. The benchmark also showed that is possible to use a subset of sensors to estimate an hourly calibration of a large set of sensors.

Recently, other sensors, like the Sensirion SPS30, have become popular. Presently, some additional 500 Sensirion sensors are used in several cities. There were also mobile tests. Some 400–500 Sensirion SPS30 sensors were used in an experiment where the sensors were mounted on bikes of volunteers. In an article about this experiment (Wesseling et al., 2021), the uncertainty of these sensors was assessed. The annual average $PM_{2.5}$ concentration in 2020, measured using a set of sensors, was shown to have a bias of −1.2 μg/m^3, and a standard deviation of 0.8 μg/m^3. The yearly average $PM_{2.5}$ concentration in the area being studied was 9.35 μg/m^3. After applying an average calibration scheme, the bias was almost completely eliminated while the standard deviation increased slightly to 1.1 μg/m^3, resulting in a relative uncertainty of 23.5%.

In several countries in Europe, extensive sensor networks are active, consisting to a large extent of only a hand full of types of sensors. Several networks of sensors run in parallel. For the present paper we therefore assume that all the sensors that are included in the type of analysis discussed here are of the same type. For the present analyses we assume the use of low-cost sensors with uncertainties like that of the Sensirion

SPS30.

Component	Period	Concentration level (μg/m^3)	Uncertainty (95%CI)	Remarks
PM2.5	Year	10	23.5	Sensirion SPS30 (Wesseling et al., 2021)

2.2. Model calibration uncertainty using reference instruments only

The model is calibrated by comparing the results to those of reference measurements. We assume that all N reference locations can be used for the model calibration. It is furthermore assumed that the model bias can be calculated as the average of the difference between the model results (M) and the reference measurements (R) at all N locations. The reference measurements are assumed to have only a negligible bias. This is guaranteed by the calibration procedures for official measurements. So, the uncertainty of the reference measurements consists of random uncertainties only. After calibration of the model to the reference measurements, by adding the observed average bias to the model results, there is no remaining bias in the model result. The uncertainty in the calibration, neglecting correlation between measurement and model uncertainties, can then be estimated from a detailed analysis, presented in Appendix A. The resulting variance of the uncertainty is given by:

$$s^2_{\overline{\Delta_{RM^c}}} = \frac{s_M^2 + s_R^2}{N} \tag{1}$$

The subscript "$\overline{\Delta_{RM^c}}$" in (1) identifies the type of calibration of the model: uncertainty of model calibrated using reference measurements. Because of the assumed lack of bias of the reference measurements, only random uncertainties play a role in (1). Usually, the random uncertainty of model calculations is (much) larger than that of reference measurements ($s_M > s_R$). As a result, the resulting uncertainty (i.e. the standard deviation or variance) of model calibration is dominated by the random uncertainty of the model.

2.3. Model calibration using calibrated low-cost measurements

Instead of using the reference measurements, the model can be calibrated using the in total L available low-cost measurements. The analysis is similar to above. It is assumed that all N reference instruments are used to calibrate the low-cost measurements with m low-cost measurements co-located with each reference instrument, the L-mN additional calibrated low-cost measurements are then used to calibrate the model. In this combined approach the uncertainty in the model calibration depends not only on the basic (instrumental) uncertainty of the low-cost measurements but also on the uncertainty of the calibration of the low-cost measurements. For the analysis of the calibration uncertainty of the model using low-cost measurements we must keep in mind that the uncertainty of the low-cost measurements consists of a random part and a systematic part. This latter part is due to the calibration of the low-cost measurements to the reference measurements. Whereas the uncertainty due to the random part reduces with increasing total number of low-cost measurements (L), the uncertainty due to the systematic part only depends on the available number of reference measurements (N) to calibrate the low-cost measurements. The final variance of the uncertainty in the calibration can be written (see Appendix A) as:

$$s^2_{\overline{\Delta_{EM^c}}} = \frac{s_E^2 + s_M^2}{L - mN} + \frac{s_E^2/m + s_R^2}{N} \tag{2}$$

Usually, the uncertainty of the low-cost measurements is larger than that of the reference measurements; the uncertainty of the low-cost measurement's calibration is then dominated by the uncertainty of the low-cost measurements. In practice, there will be (much) more low-cost measurements than there are reference measurements. As a result, the dominating term in the uncertainty will often be the uncertainty of low-cost measurements calibration, which may be reduced by applying several of these (m) at the reference locations, reducing the variance of the random uncertainty of the average sensor measurement at the locations by a factor $1/m$.

2.4. Model calibration using both reference and low-cost measurements

In a situation where both reference measurements and low-cost measurement data are available, the model calibration will be performed using both data sets. The N locations with reference measurements are used to calibrate the model directly, and the L-mN locations with low-cost measurements only will also be used for calibration of the model. The model calibration will then be a weighted sum of the model/reference and model/low-cost measurements calibrations. The weights are determined to minimize the overall uncertainty. The variance of the uncertainty in the final combined calibration of the model results can be calculated by:

$$s^2_{\overline{\Delta}} = \frac{s^2_{\overline{\Delta_{EM^*}}}\, s^2_{\overline{\Delta_{RM^*}}}}{s^2_{\overline{\Delta_{EM^*}}} + s^2_{\overline{\Delta_{RM^*}}}} + \frac{S_R^2}{N} \tag{3}$$

Here $s^2_{\overline{\Delta_{RM^*}}}, s^2_{\overline{\Delta_{EM^*}}}$ are the variances of the calibration of the model to the N reference measurements and L low-cost measurements, respectively, see equations (1) and (2) and the Appendix. To prevent an underestimation of the variance of the uncertainty due to the reference measurements it is eliminated from the intermediate terms $s^2_{\overline{\Delta_{RM^*}}}, s^2_{\overline{\Delta_{EM^*}}}$ and later added to the overall result (see Appendix A, relation (A.26)).

An interesting question is how many reference measurements can be replaced by introducing low-cost measurements while maintaining the same overall uncertainty in the model calibration. In Appendix A it is shown that the standard deviation of the uncertainty obtained using K reference measurements combined with L low-cost measurements (diffusion tubes/sensors) can also be obtained using reference measurements only, but with more of these measurements. I.e. combining low-cost measurements (like sensors and diffusion tubes) and reference measurements means that less reference measurements are needed to obtain the same uncertainty. Several examples will be discussed in chapter 3.

The above results were derived assuming that all available reference measurements are used to calibrate both the model and the additional co-located low-cost measurements. The calibration obtained for the co-located low-cost measurements is used to calibrate the remaining low-cost measurements, that are in turn used to calibrate the model at the locations with only low-cost measurements and model results. In practice, not all locations with reference measurements may be used to calibrate the low-cost measurements. In that case K out of N locations are used for the calibration of the low-cost measurements whereas N locations are used for the calibration of the model using the reference measurements. The statistics of the uncertainties becomes slightly more complicated, as is shown below and derived and explained in Appendix A, relation (A.29).

$$s^2_{\overline{\Delta}} = \frac{s^2_{\overline{\Delta_{EM^*}}}\, s^2_{\overline{\Delta_{RM^*}}}}{s^2_{\overline{\Delta_{EM^*}}} + s^2_{\overline{\Delta_{RM^*}}}} + \frac{s_R^2}{N}\left[1 + \left(\frac{s^2_{\overline{\Delta_{RM^*}}}}{s^2_{\overline{\Delta_{EM^*}}} + s^2_{\overline{\Delta_{RM^*}}}}\right)^2 \frac{N-K}{K}\right] \tag{4}$$

2.5. Validation and generalization of analytic expressions

The above equations are derived with a number of approximations, these are discussed in Annex A. Several applications of the derived relations are used in the next chapters, both in the form of figures and as tables. The validity of the approach and derived relations is tested by numerical simulation of both the calibration of the low-cost measurements with respect to the reference instruments and of the calibration of the model with respect to the reference instruments and the low-cost measurements. An R script was developed for the numerical tests, see Appendix D. All the uncertainties presented in the examples shown in the figures and tables in this article were checked using the numerical scripts. In all cases presented in the figures and tables, the results were practically identical. In situations with low numbers of measurement locations, either total numbers or numbers of low-cost measurements, there will be differences between the analytic results and the numerical results. In our simulations, we performed at least 10000 runs for every situation to ensure that the numerical results are representative of the situations described by the analytic relations. The numerical simulation set up, presented in Annex D, can be used to evaluate more complicated (real-life) variations in experimental set up. Examples are the use of data sets with actual measured concentration distributions or using (empirically derived) complex distributions for the standard deviation of the uncertainty of the low-cost measurements.

3. Examples of calibration using passive samplers

3.1. Yearly NO_2 with Palmes tubes

3.1.1. Optimal deployment of NO_2 palmes tubes

The statistical framework described above enables interesting evaluations using existing data. For the measurements and modelling of the annual average NO_2 concentrations, accurate performance data are available. Monitoring of the annual average NO_2 concentration is important, both because of the EU Air Quality Directive and the estimated health effects. In this example we will describe how a combination of reference measurements and low-cost Palmes tubes can be used to reduce the uncertainty in the calibration of a model. The uncertainty in the calibration of the model that can be obtained using different combinations is discussed below and the results are presented in Fig. 2.

For the measurements of yearly average concentrations with the reference instruments, the standard deviation of the random uncertainty is estimated as $s_R = 5\%$ and for individual measurements with passive samplers (Palmes tubes), as $s_E = 10\%$. The red curve in Fig. 2 shows the uncertainty of the model calibration using between 1 and 40 reference measurements only. These values are calculated using relation (1) of section 2.1, which is the same as relation (4) in Appendix A. As discussed in Appendix A, the calibration assumes that the same numerical value of the difference can be used in the whole domain being considered. For different calibration scenarios, we have assumed a situation where 100 Palmes tubes are available for measurements of yearly average concentrations. I.E. there are 100 Palmes tubes available, for each of 13 periods of 4 weeks in a year, resulting in 100 annual/yearly average measurements with Palmes tubes. The model is calibrated using the Palmes tubes only. There are 40 reference measurements available during the whole year. For different strategies we can distribute the Palmes tubes partly over the reference locations, for calibration of the tubes, and use the remainder at other locations. The results of the calibration obtained at these other locations are subsequently used to calibrate the model. In practice, the calibration of Palmes tubes may be different for several types of locations, like traffic, urban, rural locations. In that case the analysis is valid for only one of these groups of types of locations. When different types of locations are combined, the variance of the calibration uncertainty will usually increase.

A straightforward strategy is to locate only 1 Palmes tube at only one reference location and use that comparison to calibrate the other 99 Palmes tubes. These 99 results are subsequently used to calibrate the model. We then have an example of the situation described in section C of Appendix A, where relation (14) provides the resulting total uncertainty in an individual model result that was calibrated using L = 99 free Palmes's tube measurements that were calibrated using K = 1 reference

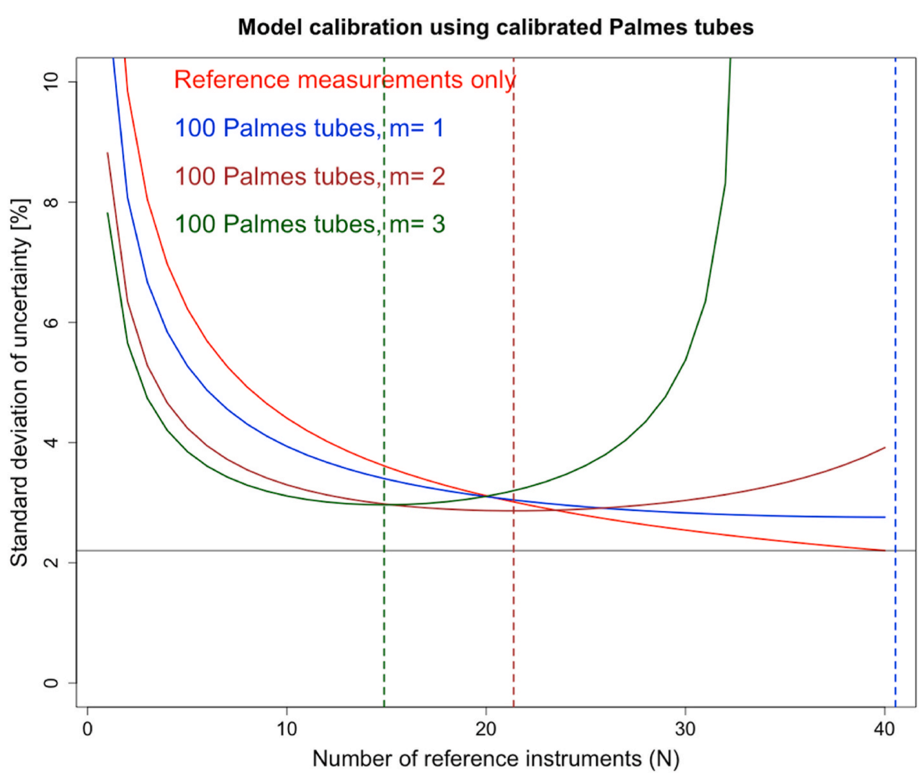

Fig. 2. Example of yearly average model calibration using calibrated Palmes tubes only. Shown are the resulting standard deviation of the uncertainty in the calibration when m = 1, 2 and 3 Palmes tubes are co-located with K = 1, 2, 3, …, 40 reference measurements and the remainder of 100 tubes are used to calibrate the model. For comparison, the red curve shows the uncertainty if only reference measurements are used. The dashed vertical lines show the optimum number of collocation locations for 100 Palmes tubes and m = 1,2,3. The standard deviation of the uncertainty of the reference measurement is estimated as $s_R = 5\%$, for the model as $s_M = 13\%$, and for individual measurements with Palmes tubes as $s_E = 10\%$. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

measurements, so the total number of Palmes tubes equals 100. The blue curve (labelled "… m = 1") in Fig. 2 represents the resulting uncertainty in the model calibration when 1 Palmes tube is co-located with K = 1, 2, 3, …, 40 reference measurements each, respectively. The results of the remaining 99, 98, 97, …, 60 Palmes tubes are then used to calibrate the model, in this scenario the reference stations are not used for calibration of the model. Similarly, the brown curve (labelled "… m = 2") represents the resulting uncertainty in the model calibration when 2 Palmes tubes are co-located with K = 1, 2, 3, …, 40 reference measurements, respectively, and the remaining 98, 96, 94, …, 20 tubes are used to calibrate the model. The other green curve shows the model uncertainties when 3 Palmes tubes are co-located at reference stations. However, the total number of Palmes tubes at any moment is limited to 100. This means that, when 3 Palmes tubes are co-located at reference stations, only 33 reference stations can be equipped with Palmes tubes, leaving only 1 Palmes tube to calibrate the model. The blue, brown and green curves were calculated using relation (2) of this article, that is the same as relation (13) in Appendix A.

Fig. 2 shows that calibration of the Palmes tubes on a single reference location (the points at K = 1) results in a poor result for the calibration. For a single sampler the standard deviation of the uncertainty is roughly 10%, which improves by mounting multiple Palmes tubes at this single reference location. Calibration accuracy improves substantially by calibrating the Palmes tubes at more reference locations. However, in case not enough Palmes tubes are left to calibrate the model, the standard deviation of the uncertainty in the model calibration will increase, this explains the behaviour of the green m = 3 curve in Fig. 2. So, clearly there is an optimum number of reference stations to be equipped with Palmes tubes, depending of the number of tubes that is co-located at each reference station. Too few and the calibration of the Palmes tubes is limited, leading to relatively large uncertainty in the model calibration that is performed using the Palmes tubes. Too many Palmes tubes used for co-location means that there are too few for a sufficient calibration of the model. In Appendix A, the optimum value of the number of collocated Palmes tubes is derived, see equation (17). The dashed vertical lines in Fig. 2 show these optimum number of collocation locations for 100 Palmes tubes and 1, 2 and 3 Palmes tubes at every location. The results are 15, 21 and 41 locations, respectively. Appendix A gives the equation to estimate the optimum values for m for other situations, in relation (18). Fig. 2 furthermore shows that, given a sufficient calibration, low-cost measurements can clearly reduce the uncertainty of model calibration in case of sparse networks.

For a range of calibration strategies, results can be quite comparable in the situation described above. In general, 20–50 Palmes tubes at reference sites and correspondingly 80-50 free Palmes tubes give a standard deviation of the uncertainty in a model calibration parameter of roughly 2.5 %. In our hypothetical situation, the best result is obtained with 41 sites with a single Palmes tube. Increasing the number of reference measurements clearly also reduces the uncertainty. However, in this analysis we are mainly interested in the effect of adding low-cost measurements.

3.1.2. *Optimal combination of reference instruments and NO_2 palmes tubes*

The improvements in model calibration, as shown in Fig. 2, also enable the possibility to optimize the overall monitoring efficiency. Since sensors/samplers are much cheaper than the reference instruments, the replacement of some reference instruments by a large(r) number of samplers might lead to both an improvement in quality and a reduction of costs. In the previous examples, the model was calibrated using the Palmes tubes only. In this next step we combine calibration of

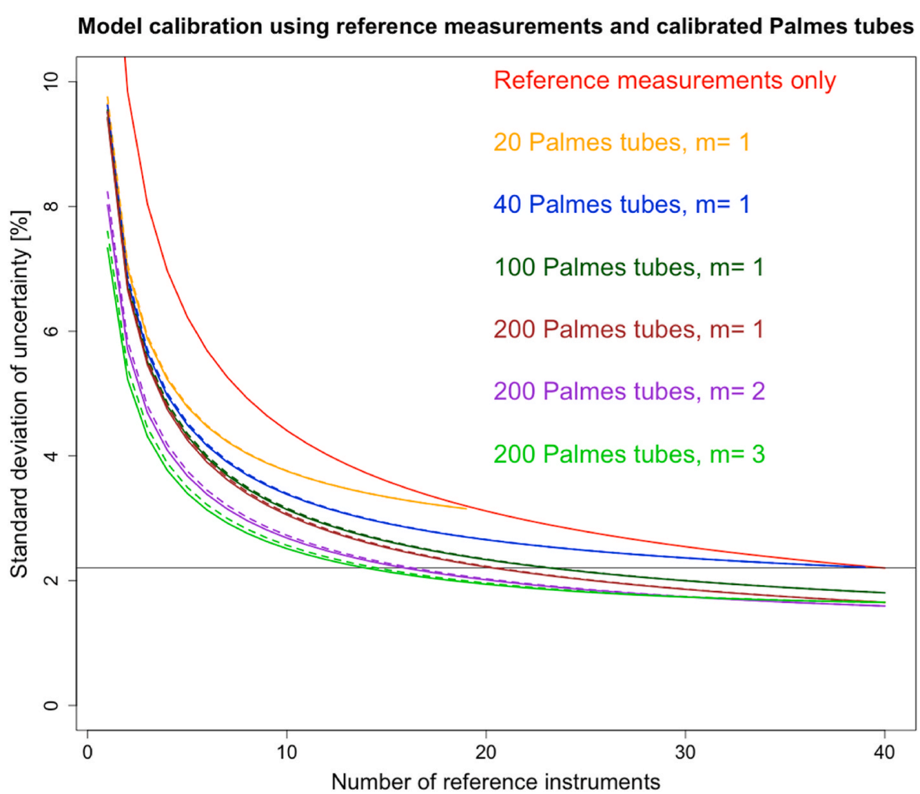

Fig. 3. Example of yearly average model calibration using both reference measurements and calibrated Palmes tubes. The standard deviation of the uncertainty in the model calibration is shown, using up to 40 reference measurements and a varying number of Palmes tubes calibrated using 1, 2, 3 Palmes tube co-located at every available reference location for calibration of the Palmes tubes that are used to additionally calibrate the model. The solid lines are calculated using relation (3) and the dashed using (4). The horizontal black line corresponds to the standard deviation of the model calibration obtained using 40 reference instruments (2.2 %). The standard deviation of the uncertainty of the reference measurement is estimated as $s_R = 5\%$, for the model as $s_M = 13\%$, and for individual measurements with Palmes tubes as $s_E = 10\%$.

both the low-cost measurements and the model using reference measurements. The results obtained for several combinations of reference measurements and Palmes tubes are presented in Fig. 3.

Using the same set-up of all 40 reference instruments with $s_R = 5\%$ and for the model $s_M = 13\%$ as baseline situation the standard deviation of the yearly average model calibration amounts to 2.2%. In case of less reference instruments the standard deviation of the model calibration increases. The red line in Fig. 3 (below) again shows the uncertainty in case of only up to 40 reference measurements. The uncertainty of the model calibration is also shown in Fig. 3 for a varying number of Palmes tubes co-located with the reference measurements. The plots show the estimated accuracies for an assumed base line situation with 40 reference instruments, in combination with in total 20, 40, 100 and 200 Palmes tubes measurements on top of the K reference measurements. The Palmes tubes are calibrated on all available K = 10.40 locations using 1, 2 or 3 Palmes tube(s) at each available location (… m = 1,2,3), depending on the scenario. The uncertainties are calculated using relations (3) (dashed curves) and relation (4) (other curves) with input from relations (1) and (2). Note that the curves can only be calculated for situations with L-mK > 0 in relation (2).

As expected, the calibration strategies in Fig. 3 seem quite sensible. Using additional Palmes tubes to calibrate the model means that less than 40 reference measurements are needed to obtain the standard deviation of the uncertainty otherwise obtained using the 40 reference measurements only. Employing 100 Palmes tubes, 20 for calibration and 80 at other locations in combination with only 20 reference instruments (50 % of the available number) yields the same overall model calibration accuracy, 2.19% using the combination versus 2.20% with 40 reference measurements only. The differences between the dashed curves (calculated using relation 3) and the other curves (calculated using relation 4) is limited. In practical situations relation (3) will be sufficient to estimate the uncertainties.

The above schemes to calculate uncertainties can be used to optimize the application of, for example, 100 samplers using both samplers and reference measurements to calibrate the model. Table 1, also calculated using the relations presented in Annex A, shows the results for measurement strategies with 30 and 40 reference instruments. For the four scenarios in the table, 100 Palmes tubes are added to the reference measurements, resulting in a reduction of the calibration uncertainty from 2.5 % to 2.0 % for a scenario with a single calibration of Palmes tubes at 30 locations and 70 additional Palmes tubes. In case calibration of the Palmes tubes is performed using groups of 3 tubes at a maximum of 10 reference locations, the overall uncertainty in the model calibration drops from 2.5% to 2.1%. When 40 reference measurements are available, the two scenarios discussed above result in a reduction of the overall uncertainty in the model calibration, from 2.2% to 1.8% or 1.9%, depending on the scenario.

3.1.3. *Real life example using volunteers (citizen science)*

Since 2017, almost 20 volunteers in the Netherlands have been helping the National Institute for Public Health and the Environment (RIVM) to measure nitrogen dioxide (NO_2) throughout the country (Siteur, 2024). Every month, the volunteers replace the Palmes tubes used for this purpose by new tubes and send the used tubes to RIVM, where they are analyzed. The volunteers mostly live in places where few official measurements are available. This way more information is collected about air quality in these areas. These measurements are an important tool for calibrating model calculations, increasing the accuracy in areas with few reference measurements.

The Palmes tubes are calibrated using 11 sets of 2 tubes co-located with official measurements. In practice a spatial model calibration is applied. Therefore, the main improvement in accuracy is observed in areas with a low density of reference instruments. These are in general the areas outside the highly populated agglomerations.

3.1.3.1. Small network and local sources. The examples discussed above concern situations with many reference measurements in a large area. In the vicinity of local sources in a smaller area, like traffic, the yearly average model quality may be very different from the quality in background situations. This situation occurs frequently in large cities. An effective local model needs a much higher spatial resolution than models for background situations. The local model result is more sensitive to the accuracy of emission strength and the exact location of individual sources. To calibrate such a model is therefore even more challenging than the calibration of the background model, even more so as the available resources for local studies may be limited. To study the influence of a measurement strategy, we again use equation (4) with input data that is appropriate for local situations. For many common situations, like a large city or similar area, we assume the availability of 5 or 2 reference instruments and the additional use of a set of Palmes tubes. We also compare the resulting uncertainties with a hypothetical situation with 10 reference measurements without Palmes tubes. In Table 2 the advantages of a mixed measurement strategy, using both reference measurements and Palmes tubes, are shown. The uncertainty of the model calibration using 10 reference measurements is 4.4%. The scenario of replacing 3 reference instruments out of the assumed 5 by 20 Palmes tubes results in a lower combined standard deviation of 5.5%, versus 6.2% when using only 5 reference measurements.

The results of Table 2 are valid for the calibration of a specific model for a specific type of situation, in this case a model describing the background concentrations on the scale of a large city. Model calibration is also relevant in other situations. There are many other types of situations, for example, measurements close to urban roads are used to calibrate the Dutch dispersion models for the calculation of traffic contributions to the NO2 concentrations [Wesseling et al.]. In the vicinity of a large local source, the model uncertainty itself will usually be larger than was assumed in the above (background) examples. Table 3shows the calibration accuracies for a local model with a larger model uncertainty, with a standard deviation $s_m = 25\%$. In case of model calibration with only 10 reference instruments the overall uncertainty increases with almost a factor of 2 with respect to the results in Table 2, going from 4.4% to 8.1%. When 5 reference measurements and Palmes tubes are combined, the overall uncertainty can be below that of using 10 reference measurements only. Combining 2 reference measurements and Palmes tubes yields an overall uncertainty below that of using 5 reference measurements only. This example illustrates how, in situations with relatively uncertain models, the application of Palmes tubes more easily improves the total calibration result than in situations with more accurate models.

Table 1
Calculated uncertainties for different configurations of measurements using 30 or 40 reference instruments, combined with 2 scenarios to distribute 100 Palmes tubes in the domain, assuming $s_R = 5\%$, $s_E = 10\%$ and $s_M = 13\%$.

Nr. reference Locations	Nr. calibration Locations	Nr. tubes per ref. location	Total number of Palmes tubes	standard deviation of the uncertainty using reference only	Uncertainty using tubes only	Combined standard deviation of the uncertainty
30	30	1	100	2.5%	2.8%	2.0%
30	10	3	100	2.5%	3.1%	2.1 %
40	40	1	100	2.2%	2.8%	1.8%
40	10	3	100	2.2%	3.1%	1.9%

Table 2
Calculated yearly average uncertainties for different configurations of measurements on a local scale, using 5 and 2 reference instruments, respectively, combined with 2 scenarios to distribute 20 Palmes tubes, assuming $s_R = 5\%$, $s_E = 10\%$ and $s_m = 13\%$.

Nr. reference Locations	Nr. calibration Locations	Nr. tubes per ref. location	Total number of Palmes tubes	standard deviation of the uncertainty using reference only	Uncertainty using tubes only	Combined standard deviation of the uncertainty
10	–	–	–	4.4%	–	–
5	–	–	–	6.2%	–	–
5	5	1	20	6.2%	6.6%	4.8%
5	2	3	20	6.2%	7.0%	4.9%
2	2	1	20	9.8%	8.8%	7.0%
2	2	3	20	9.8%	7.0%	6.1%

Table 3
Example showing yearly average uncertainties for a local-scale model combined with 5 or 2 reference instruments, respectively, with 2 scenarios to distribute 20 Palmes tubes, assuming $s_R = 5\%$, $s_E = 10\%$ and a larger model uncertainty $s_m = 25\%$.

Nr. reference Locations	Nr. calibration Locations	Nr. tubes per ref. location	Total number of Palmes tubes	standard deviation of the uncertainty using reference only	Uncertainty using tubes only	Combined standard deviation of the uncertainty
10	–	–	–	8.1%	–	–
5	–	–	–	11.4%	–	–
5	5	1	20	11.4%	8.6%	7.0%
5	5	2	20	11.4%	9.4%	7.4%
2	2	1	20	18.0%	10.1%	9.1%
2	2	3	20	18.0%	9.0%	8.3%

3.1.3.2. Cost effectiveness. Efficient monitoring usually requires the combination of sufficiently accurate results with a minimum of costs. If we assume that a reference location plus apparatus for NO2 costs about 10000 €/year and a passive sampler (Palmes tube) location 500 €/year (material and transportation, changing the tubes 13 times in a year) we can evaluate the cost effectiveness of several monitoring scenarios. In Table 1 the combination of 30 reference instruments with 100 samplers can yield (using a sensible calibration strategy) a more accurate model calibration (1.79%) than using 40 reference measurements only (2.20%). However, in this case the costs are reduced by 10 * 10000€ (saving 10 reference measurements) - 100 * 500€ (cost new Palmes measurements) = 50000€.

In the local example, Table 3, the replacement of 3 reference instruments by 20 Palmes tubes improves the model calibration performance (from 11.40% down to 8.96%) while the yearly costs are reduced by 3 * 10000€-20 * 500€ = 20000€.

Note that in order for the passive samplers to be really 'low-cost', the calibration and use requires also a very low-cost maintenance and replacement procedure, for example by the assistance of local volunteers.

3.2. Yearly average ammonia with diffusion tubes

High quality measurements of ammonia concentrations is required in the Netherlands as part of the monitoring of nitrogen deposition in the country. Reducing the amount of deposition of nitrogen (of which about 2/3 consisting of ammonia) is an important objective in many "Natura 2000″ areas. From 1993 until 2014 the monitoring of ammonia in the Netherlands was based on 8 automatic AMOR instruments (Wichink Kruit 2020). In 2014 two monitoring locations were closed, leaving 6 automatic monitors. From January 1st, 2016 mini-DOAS instruments gradually replaced the AMOR monitors (Wichink Kruit 2021).

Ammonia concentrations exhibit large gradients on the spatial scale and are also quite variable in time. The effect of ammonia deposition is visible over longer periods (i.e. several years). As a result, the intra-annual variability in time is not very important from the routine monitoring perspective. On the spatial scale however, the gradients are crucial. Therefore, the monitoring of ammonia was extended by RIVM already in 2005 by using passive samplers (diffusion tubes), which are operated on a monthly basis. The network of passive samplers is called "Monitoring Ammonia in Nature (MAN)". Yearly modelling using the OPS model [Sauter et al., 2023] completes the monitoring of ammonia concentrations. In this paragraph we will investigate the monitoring performance by comparing yearly measurement and modelling data. Data are available for the period 2005 until now, on an increasing number of locations (currently more than 300). In general, the number of locations with passive samplers is increasing with the years. Measurement locations and results are presented on the MAN website of the RIVM (see https://man.rivm.).

As mentioned in section 2.1, Noordijk (2020) reported the relevant measurement uncertainties of the ammonia measurements and monitoring. The model calculations are performed using the Dutch OPS model, which produces annual average concentrations. For the whole data set, the differences between the model results and the measurements show a standard deviation of 2.2 $\mu g/m^3$. This difference largely exceeds the measurement uncertainty and is therefore assumed equal to the random model uncertainty, so we take $s_M = 33\%$.

Using the above uncertainties, the uncertainty of the average model calibration (equation (3)) is calculated and shown in Fig. 4. In the original setup, 8 reference instrument were operational. Since 2014 the number of reference instruments is reduced to 6. Recently, installation of two new instrument sites has been accomplished.

Since the model uncertainty is much larger than both types of measurement uncertainties, a limited number of samplers already suffice to compensate for the reduction of the number of reference instruments. Note that this statement applies only to the calculation of annual averages. Uncertainties for several cases are provided in Table 4.

The added value of the additional passive samplers is evident from the reduction of uncertainties.

4. Examples with low-cost digital samplers

As shown above, simple measurements using diffusion tube can be employed to reduce the number of reference measurements needed to calibrate air quality model. Although low-cost NO_2 digital sensors are available, the calibration of these is complicated (Ratingen et al., 2021). For PM_{10} and $PM_{2.5}$ several types of low-cost sensors have become popular over the last few years. These sensors seem capable of providing sensible hourly concentration variations (Wesseling et al., 2019; 2021). Locations where low-cost PM10/PM2.5 are employed are, for instance in cities or on roads (Wesseling et al., 2019; 2021), around large agricultural sources (Woutersen et al., 2022), and over the last few years also

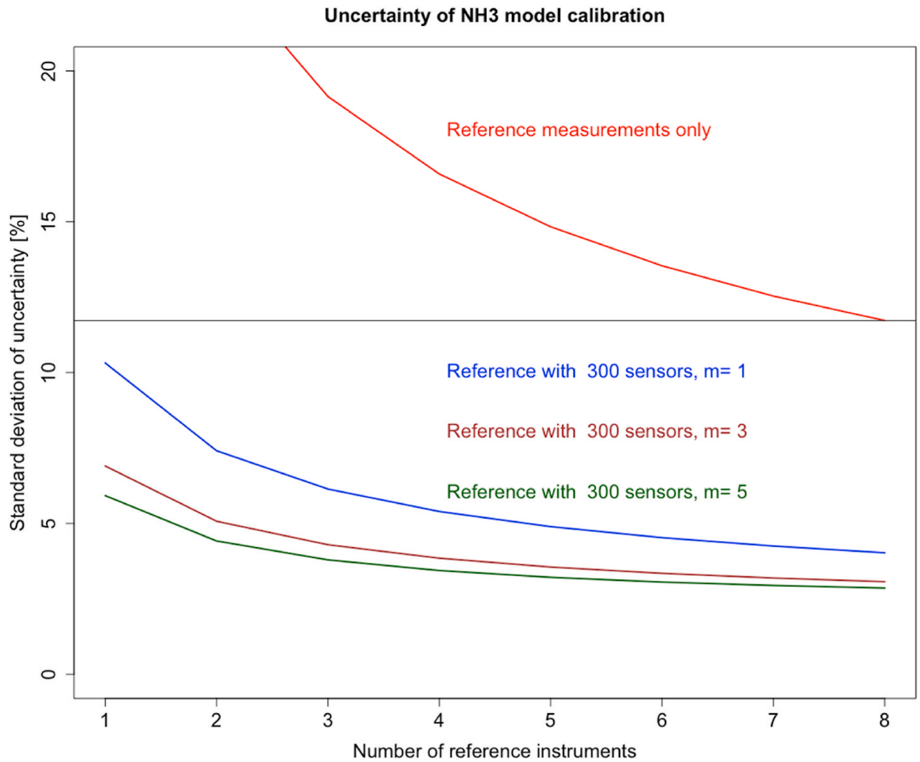

Fig. 4. Uncertainty of NH3 model calibration using up to 8 reference measurements and a varying number of passive samplers calibrated using 1, 3 and 5 passive samplers co-located at every available reference location for calibration of the passive samplers that are used to additionally calibrate the model. The black line is the uncertainty of using only 8 reference measurements, 11.7%.

Table 4
Uncertainty of yearly average NH3 model calibration using 8 resp. 6 reference instruments and 3 measurement strategies to distribute 30 or 300 samplers over the domain. We assume standard deviations of the measurements $s_R = 3.4\%$, $s_E = 10\%$ and for the model uncertainty $s_m = 33\%$.

Nr. reference Locations	Nr. calibration Locations	Nr. tubes per ref. location	Total number of Palmes tubes	standard deviation of the uncertainty using reference only	Uncertainty using tubes only	Combined standard deviation of the uncertainty
8	5	1	30	11.7%	8.4%	6.9%
8	8	1	300	11.7%	4.2%	4.0%
8	8	3	300	11.7%	3.1%	3.1%
6	6	1	300	13.5%	4.8%	4.5%
6	6	3	300	13.5%	3.4%	3.3%
6	6	5	300	13.5%	3.1%	3.1%

around a large steel plant in the Netherlands (https://hollandse-luchten.org). The website https://sensors.rivm.nl/, run by RIVM, continuously shows the results of the low-cost sensors as well as results of passive samples in the Netherlands.

For the present analysis, we assume that the systematic differences can be removed. Then the remaining uncertainty can be treated as random uncertainties. We furthermore assume a normal distribution of the random uncertainties. It is well known that low-cost sensors may show “heavy tails” in the distribution of the random uncertainties. We furthermore assume that the most prominent outliers will be captured by the calibration system for the sensors. So, for the present analysis we use a relative standard deviation for the sensor results on a yearly basis $S_E = 12\%$ (see section 2.1). The standard deviation of yearly average model results for $PM_{2.5}$ is 15% (Hoogerbrugge et al., 2023). Official $PM_{2.5}$ yearly average measurements have a reported uncertainty of 9.3% (95%CI) at 25 μg/m3 (Mooibroek et al., 2012). Some results are provided in Table 5.

In practical cases sensor will be employed in specific regions where the model results are relatively uncertain due to many complex sources in that area. Examples are areas with large industrial factories. In that case the number of reference measurements will be limited, probably in the order of 3–5 measurements. In those cases, the addition of low-cost

Table 5
Calculated yearly average uncertainties for different configurations of measurements using digital $PM_{2.5}$ sensors, obtained using 30 and 40 reference measurements and 100 additional low-cost sensors. The assumed uncertainties are $s_R = 0.05$, $s_E = 0.12$ and $s_M = 0.15$.

Nr. reference Locations	Nr. calibration Locations	Nr. sensors per ref. location	Total number of sensors	standard deviation of the uncertainty using reference only	Uncertainty using sensors only	Combined standard deviation of the uncertainty
30	30	1	100	2.9%	3.3%	2.3%
30	10	3	100	2.9%	3.5%	2.3%
40	40	1	100	2.5%	3.2%	2.0%
40	10	3	100	2.5%	3.5%	2.1%

measurements can be quite beneficial. For the analysis, we assume a local model uncertainty of 20%. Some results are provided in Table 6.

In order to have a substantial effect of the low-cost sensors on the overall uncertainty, a relatively large number of sensors is needed. However, given roughly 10–25 calibrated low-cost sensors, the uncertainty in the model calibration can be reduced by 30–40%. More important, using low-cost sensors introduces more robustness in the calibration. With five reference measurements in the area, the overall uncertainty drops from 9% to 12% when two stations drop out. With 25 low-cost sensors in the calibration, the overall uncertainty only drops from just over 5% to just over 6%.

In our experience, electronic sensors can contain significant outliers. The standard deviations strongly depend on the efficiency of the outlier detection. This can influence the monitoring strategy but this is considered outside the scope of this paper. More sophisticated calibration strategies using data fusion/assimilation or AI techniques are being investigated as part of the European FAIRMODE project (Thunis, 2020; Wesseling et al., 2024) as well as by different organisations (Gressent, 2019; 2020; Schneider, 2014; Laakso, 2018).

5. Discussion

In this article we present a set of simple relations to estimate the uncertainty in various combinations of model calibration using official reference measurements and low-cost measurements. These relations are applied in practical measurement strategies, showing also strategies currently applied in the Netherlands. The advantages of the mixed used of reference instruments and low-cost instruments is very clear. The results presented here are also intended to inspire other measurement network operators or users to optimize the network configuration. In these applications the calibration of the low-cost samplers is crucial.

5.1. Model-measurement correlations

In our analysis, we implicitly assume that the differences between model results and measurements are not correlated. However, model errors may well be geometrically correlated, for instance if the processing of meteorological conditions in an area is not optimal in a substantial part of the area. For the present analysis we choose to ignore this possible effect. Strongly correlated comparisons between model results and measurements imply that there is less objective information in the calibration of the model. In practice, the effective number of model comparisons, i.e. the numbers N and K, will be less than the actual number of locations being used in the present analysis (Hastie et al., 2009).

5.2. Representativity

For the analysis of uncertainties presented here, the relations presented in this paper assume a uniform calibration of the model over the domain. In practise this may not be applicable for relevant situations. A simple solution is to divided the area of application into smaller parts with sufficient homogeneity to treat each separately. The split can be necessary due to areas with different weather conditions, altitude, degree of urbanisation etc. After dividing the area of interest into smaller subareas the equations presented can be used but may need to be adjusted to the smaller number of measurements available in the subareas.

Typically, background concentrations are composed of many small contributions from many distant sources. In this situation uncertainties in the magnitude and spatial distribution of emission data are statistically reduced due to the large numbers of contributions. With the present performance data of NO_2 and NH_3 samplers used in this paper their results can also be very useful in hotspot application. In a hotspot situation the uncertainty in the model input parameters of the dominating source heavily influences the uncertainty of the result. Therefore, the model result may show a much larger relative uncertainty than for background situations, which makes the application of low-cost measurements very efficient for the local model calibration, i.e. the low-cost measurements perform (much) better than the model.

The statistical data used in this article are based on the performance of reference instruments and passive tubes/low-cost sensors in The Netherlands. We assume that the performance data for reference instruments are quite representative to other countries since these instruments are the bases for the measurements in the air quality directive (EU, 2008). Data on Palmes tubes for NO_2 and NH_3 are also expected to be quite representative since these tubes are obtained from commercial suppliers.

5.3. Reference instruments

The application of low-cost measurements can reduce the need for large amounts of reference instruments. On the other hand, the importance of reference instruments cannot be overstressed. Without the reference instruments the calibration and the validation of the low-cost measurements is hardly possible and the reliability of the system fades away. The dependency on reference instruments can also appear in the interpretation of the results or when model improvements are required. For example, the limited time resolution of passive samplers is hardly disturbing the assessment of annual limit values or long-term exposure. It may be problematic in the understanding of the underlying processes. The combination of reference and low-cost instruments however still can be very powerful. In such a combination, the locations for the expensive reference instruments should be carefully selected with respect to their representativity. Other locations of interest, like hotspots in emissions, or in concentrations or even in public attention, can be very well dealt with, using properly calibrated and validated low-cost measurements.

5.4. Outliers and representativity of the calibration of low-cost measurements

In our experience, a set of electronic low-cost sensors can contain significant outliers. The uncertainties strongly depend on the efficiency of the outlier detection. This can influence the monitoring strategy but this is considered outside the scope of this paper, see Wesseling et al. (2024) for the results of a benchmark on calibration of low-cost $PM_{2.5}$ sensors. Limited homogeneity in sensor calibration can in first approximation be solved by dividing the area of interest in smaller areas. After some years of experience with both the model and the sensors more

Table 6
Calculated yearly uncertainties for different configurations of measurements using digital $PM_{2.5}$ sensors, obtained using 3 and 5 reference measurements and 8-40 additional low-cost sensors. The assumed uncertainties are $s_R = 0.05$, $s_E = 0.12$ and $s_M = 0.20$.

Nr. reference Locations	Nr. calibration Locations	Nr. Sensors per ref. location	Total number of sensors	standard deviation of the uncertainty using reference only	Uncertainty using sensors only	Combined standard deviation of the uncertainty
3	3	1	8	11.9%	12.9%	9.0%
3	3	3	34	11.9%	6.8%	6.1%
5	5	1	10	9.2%	11.9%	7.5%
5	5	3	25	9.2%	8.3%	6.4%
5	5	3	40	9.2%	6.0%	5.2%

sophisticated calibration strategies using data fusion/assimilation or AI techniques are being investigated (Thunis, 2020).

Calibration of air quality models with a combination of few reference instruments combined with many low-cost and less accurate measurements can outperform a calibration using a small number of very precise measurements only. In all applications, the calibration of the low-cost samplers is a key issue in the accuracy of the results. We observe that, for example in the Netherlands, many stakeholders have larger or smaller networks of low-cost samplers. This implies that the strategic combination of calibration efforts will be much more efficient than a set of individual calibration efforts. Obviously, this will need harmonised procedures on the handling of the samplers and the raw data. Also, some overall design and an overall statistical evaluation of the comparison measurements will improve the accuracy.

5.5. *Citizen science*

The application of low-cost measurement instruments enables the involvement of citizens in the assessment of air quality. This has at least two benefits: the measurements can help to improve the assessment of the air quality and the involvement of citizens may feed a constructive dialog on how to improve the air quality. For both goals a proper calibration of the low-cost measurements will improve the quality of the contribution of the citizens. For the professional owner of a measurement and modelling infrastructure, the collaboration with citizen science initiatives can be very promising provided the professionals organise the calibration strategy of the low-cost measurements and provide sufficient guidance for proper application of the low-cost measurements. Application of low-cost measurements has also been recommended to the European Commission in their upcoming revision of the EU Ambient Air Quality Directives (Thunis et al., 2022).

6. Conclusions

In this article we present a simple statistical framework describing the calibration of a model using different combinations of reference measurements and additional low-cost measurements, like Palmes tubes for NO_2 and NH_3 and PM2.5 digital sensors.

Adding low-cost measurements, like Palmes tubes, to an (existing) network of reference measurements can substantially reduce the uncertainty of model calibration. Alternatively, a similar uncertainty can be achieved using less reference measurements, which are usually expensive.

In networks with limited numbers of reference measurements, adding even a relatively modest number of low-cost measurements is especially beneficial to increase the uncertainty of the model calibration.

Given specific numbers of additional low-cost measurements, there is an optimal distribution of these measurements, with a part used for calibration of the low-cost measurements themselves and the remainder used for additional calibration of the model.

The uncertainty and bias of low-cost digital PM-samplers are larger than those of Palmes tubes. As a result, the effect of using these samplers on the overall model calibration is modest. However, using low-cost sensors introduces more robustness in the calibration.

The presented examples and plots on network optimization will hopefully inspire and assist network operators to balance the use of expensive reference instruments and low-cost sensors.

CRediT authorship contribution statement

Ronald Hoogerbrugge: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Sjoerd van Ratingen:** Writing – review & editing, Formal analysis. **Koen Siteur:** Writing – review & editing, Formal analysis. **Joost Wesseling:** Writing – original draft, Project administration, Methodology, Formal analysis.

Declaration of competing interest

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.

Acknowledgements

The study was funded by the Dutch Ministry of Infrastructure and Water Management as part of grants M240106 and M240108 and commissioned to the National Institute for Public Health and the Environment.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2024.120990.

Data availability

No data was used for the research described in the article.

References

Blank, F.T., 2001. Measurement Uncertainty National Air Quality Measurement Network (LML), KEMA Report. RIVM (in Dutch) 50050870-KPS/TCM 01-3063.

Bush, T., Smith, S., Stevenson, K., Moorcroft, S., 2001. Validation of nitrogen dioxide diffusion tube methodology in the UK. Atmos. Environ. 35, 289–296.

Coolen, M., Spoor, R., Wietses, M., Rijkse, P., 2023. Report on the assessment of air quality in The Netherlands in 2022. https://www.luchtmeetnet.nl/nieuws (in Dutch).

De Craemer, S., Vercauteren, J., Fierens, F., Lefebvre, W., Meysman, F.J., 2020. Using large- scale NO2 data from citizen science for air-quality compliance and policy support. Environ. Sci. Technol. 54 (18), 11070–11078.

EC, 2022. Proposal for a Directive of the European Parliament and of the Council on Ambient Air Quality and Cleaner Air for Europe. European Commission. Technical Report.

EU, 2008. Directive 2008/50/ec of the European parliament and of the council of 21 May 2008, on ambient air quality and cleaner air for Europe. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:en:PDF.

Gressent, A., Malherbe, L., Colette, A., Rollin, H., Scimia, R., 2020. Data fusion for air quality mapping using low-cost sensor observations: feasibility and added-value. Environ. Int. 143, 105965. https://doi.org/10.1016/j.envint.2020.105965. Epub 2020 Jul 17. PMID: 32688160.

Gressent, Alicia, Malherbe, Laure, Colette, Augustin, 2019. Data fusion for air quality mapping using sensor data: feasibility and added-value through an application in Nantes. In: European Regional Conference of the International Environmetrics Society. (GRASPA 2019), Pescara, Italy, p. 51. ⟨ineris-03319926⟩.

Hafkenscheid, T.L., Vonk, J., 2014. Evaluation of equivalence of the MetOne BAM-1020 for the measurement of PM2.5 in ambient air. RIVM Letter report 2014-0078.

Hafkenscheid, T., Fromage-Mariette, A., Goelen, E., Hangartner, M., Pfeffer, U., Plaisance, H., de Santis, F., Saunders, K., Swaans, W., Tang, Y.S., Targa, J., van Hoek, C., Gerboles, M., 2009. Review of the Application of Diffusive Samplers in the European Union for the Monitoring of Nitrogen Dioxide in Ambient Air. JRC51106. ISBN 978-92-79-12052-7.

Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome, 2009. The Elements of Statistical Learning; Data Mining, Inference, and Prediction, second ed. Springer, p. 588.

Heal, M.R., Laxen, D.P.H., Marner, B.B., 2019. Biases in the measurement of ambient nitrogen dioxide (NO2) by Palmes passive diffusion tube: a review of current understanding. Atmosphere 10, 357. https://doi.org/10.3390/atmos10070357.

Hoogerbrugge, R., Hazelhorst, S., Huitema, M., Siteur, K., Smeets, W., Soenario, I., Visser, S., de Vries, W.J., Wichink Kruit, R.J., 2023. Large-scale Concentration Maps for the Netherlands, RIVM-Report 2023-0113 (in Dutch).

Irwin, A., 2018. No PhDs needed: how citizen science is transforming research. Nature 562, 480–482.

Janssen, S., Thunis, P., 2020. FAIRMODE guidance document on modelling quality objectives and benchmarking. Technical Report EUR 30264 EN. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/15956.

Janssen, S., Thunis, P., 2022. FAIRMODE Guidance Document on Modelling Quality Objectives and Benchmarking (Version 3.3), EUR 31068 EN. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/41988,JRC129254. ISBN 978-92-76-52425-0.

Laakso, Mikko, Johansson, Lasse, 2018. New tools for air quality modelling and forecasting: compact sensors combined with a data fusion model with forecasting capabilities. In: NAQ Conference. Austin, USA.

Lauriks, F., Jacobs, D., Meysman, F.J.R., 2022. CurieuzenAir: Data Collection, Data Analysis and Results. University of Antwerp, p. 50.

Mooibroek, D., 2014. Annual Report on Air Quality, 2013 (in Dutch), RIVM Rapport 2014-0111.

Mooibroek, D., Berkhout, J.P.J., Hoogerbrugge, R., 2012. Yearly Report Air Quality. RIVM report 680704023/2013 (in Dutch).

Nguyen, P.L., Wesseling, J., 2016. Measurement of Nitrogen dioxides (NO2) using so-called Palmes tubes. RIVM report 2016-0089 (in Dutch).

Palmes, E.D., Gunnison, A.F., Dimattio, J., Tomezyk, C., 1976. Personal sampler for nitrogen dioxide. American Industrial Hygiene Association 37, 570–577.

Pickett, E.E., Whiting, R.G., 1981. The design of cost-effective air quality monitoring networks. Environ. Monit. Assess. 1, 59–74. https://doi.org/10.1007/BF00836876.

Ratingen, S., Vonk, J., Blokhuis, C., Wesseling, J., Tielemans, E., Weijers, E., 2021. Seasonal influence on the performance of low-cost NO2 sensor calibrations. Sensors 21. https://doi.org/10.3390/s21237919.

Sauter, Ferd, Sterk, Marina, van der Swaluw, Eric, Kruit, Roy Wichink, de Vries, Wilco, van Pul, Addo, 2023. The OPS-model; description of OPS 5.1.1.0. https://www.rivm.nl/documenten/uitgebreide-modelbeschrijving-van-ops-versie-5110.

Schneider, P., Castell, N., Lahoz, W.A., 2014. Data fusion of observations from low-cost air quality sensors and models for urban-scale air quality mapping. In: Presented at COST Action TD1105 - New Sensing Technologies for Air-Pollution Control and Environmental Sustainability - Third Scientific Meeting, Istanbul, 3-5.

Siteur, K., Vos, A., de Vries, W.J., Sauter, F., Hoogerbrugge, R., 2024. Use of measurements in high-resolution air quality maps. Magazine "Air" 1 (in Dutch).

Thunis, P., Janssen, S., Wesseling, J., Piersanti, A., Pirovano, G., Tarrason, L., Martin, F., Lopez-Aparicio, S., Bessagnet, B., Guevara, M., Monteiro, A., Clappier, A., Pisoni, E., Guerreiro, C., González, Ortiz A., 2022. Recommendations for the revision of the ambient air quality directives (AAQDs) regarding modelling applications. In: EUR 31102 EN. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/761078,JRC129600. ISBN 978-92-76-53279-8.

Thunis, P., Franco, V., González Ortiz, A., Janssen, S., Wesseling, J., Guerreiro, C., Pisoni, E., Tarrasón, L., Guevara, M., Pirovano, G., Clappier, A., Monteiro, A., Durka, P., Martin, F., López-Aparicio, S., 2020. FAIRMODE 2020-2022 Roadmap.

Verghese, S., Nema, A.K., 2022. Optimal design of air quality monitoring networks: a systematic review. Stoch. Environ. Res. Risk Assess. 36, 2963–2978. https://doi.org/10.1007/s00477-022-02187-1.

Wesseling, J., de Ruiter, H., Blokhuis, C., Drukker, D., Weijers, E., Volten, H., Vonk, J., Gast, L., Voogt, M., Zandveld, P., van Ratingen, S., Tielemans, E., 2019. Development and implementation of a platform for public information on air quality, sensor measurements, and citizen science. Atmosphere 10, 445. https://doi.org/10.3390/atmos10080445.

Wesseling, J., Hendricx, W., de Ruiter, H., van Ratingen, S., Drukker, D., Huitema, M., Schouwenaar, C., Janssen, G., van Aken, S., Smeenk, J.W., Hof, A., Tielemans, E., 2021. Assessment of PM2.5 exposure during cycle trips in The Netherlands using low-cost sensors. Int. J. Environ. Res. Publ. Health 18, 6007. https://doi.org/10.3390/ijerph18116007.

Wesseling, J., Drukker, D., Gressent, A., et al., 2024. Using synthetic data to benchmark correction methods for low-cost air quality sensor networks. Air Qual Atmos Health. https://doi.org/10.1007/s11869-023-01493-z.

WHO, 2021. WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization.

Wichink, Kruit R., Bleeker, A., Braam, M., van Goethem, T., Hoogerbrugge, R., Rutledge-Jonker, S., Stefess, G., Stolk, A., van der Swaluw, E., Voogt, M., van Pul, A., 2021. Op Weg Naar Een Optimale Meetstrategie Voor Stikstof. RIVM-briefrapport.

Wichink Kruit, R.J, Ronald Hoogerbrugge, Addo van Pul, "Implementation of a Data Fusion Approach to Assess the Concentration and Dry Deposition of Ammonia in the Netherlands" , Report number: RIVM-2020-0076, DOI: 10.21945/RIVM-2020-0076.

Woutersen, A., de Ruiter, H., Wesseling, J., Hendricx, W., Blokhuis, C., van Ratingen, S., Vegt, K., Voogt, M., 2022. Farmers and local residents collaborate: application of a participatory citizen science approach to characterising air quality in a rural area in The Netherlands. Sensors 22. https://doi.org/10.3390/s22208053.