National Institute
for Public Health
and the Environment

Report 680708002/2008

R. Beijk | D. Mooibroek | J. van de Kassteele | R. Hoogerbrugge

PM_{10}: Equivalence study 2006

Demonstration of equivalence for the automatic PM_{10} measurements in the Dutch National Air Quality Monitoring Network

A technical background report

rivm

RIVM Report 680708002/2008

PM_{10}: Equivalence study 2006

Demonstration of equivalence for the automatic PM_{10} measurements in the Dutch National Air Quality Monitoring Network
A technical background report

R. Beijk
D. Mooibroek
J. van de Kassteele
R. Hoogerbrugge

Contact:
Ruben Beijk
Laboratory for Environmental Monitoring
Ruben.Beijk@rivm.nl

This investigation has been performed by order and for the account of the Directorate-General for the Environment, within the framework of the Dutch National Air Quality Monitoring Network

rivm

Abstract

PM_{10}: Equivalence study 2006

An equivalence study is carried out with the aim of ensuring the quality of PM_{10} measurements in the Dutch National Air Quality Monitoring Network (NAQMN). The results have led to an improvement in the quality of the measurements and introduce an appropriate calibration for the PM_{10} measurements in the NAQMN. The resulting relative measurement uncertainty for particulate matter (PM_{10}) measurements – performed by configurations currently in use in the NAQMN – is approximately 17%.

In the NAQMN, PM_{10} is measured at various locations across the Netherlands using automatic beta-adsorption monitors. European (EU) legislation allows this measurement method if equivalence with the official reference method (gravimetric measurements) is demonstrated. The NAQMN comprises various measurement configurations, and equivalence has been determined for each configuration. A distinction is made between three different devices and between urban and rural sites.

This study is primarily based on the equivalence guideline as recommended by the Clean Air For Europe (CAFE) steering group. Orthogonal regression is used in all cases to determine equivalency, and in situations with an insignificant intercept, orthogonal regression without intercept is applied. The equations for orthogonal regression without intercept and the corresponding uncertainty are presented. The technical background information of the steps taken to demonstrate equivalence is elaborated on in more detail in this report.

Key words:
PM_{10}, equivalence, calibration, beta, reference, orthogonal, uncertainty

rivm

Rapport in het kort

PM10: Equivalentie studie 2006

In een zogeheten equivalentiestudie door het RIVM is de onzekerheidsfactor bepaald tussen de fijnstofmetingen (PM_{10}) in het Landelijk Meetnet Luchtkwaliteit (LML) en de Europese referentiemethode. De in de equivalentiestudie bepaalde meetonzekerheid van de huidige PM_{10}-meetinstrumenten in het meetnet bedraagt circa 17 procent. De in de Europese richtlijn vermelde maximaal toegestane onzekerheid is 25 procent. De equivalentiestudie is uitgevoerd om de kwaliteit van deze fijnstofmetingen in Nederland te waarborgen. Tevens heeft de studie geleid tot verbeteringen in de kwaliteit van de PM_{10}-metingen in het LML.

In dit rapport wordt de technische achtergrond van de gevolgde stappen en de bijbehorende resultaten nader beschreven. Het LML meet PM_{10} op circa veertig locaties met behulp van automatische bèta-adsorptiemonitoren. Hoewel deze methode afwijkt van de voorgeschreven referentiemethode, staat de Europese regelgeving deze toe mits gelijkwaardigheid met de referentiemethode wordt aangetoond. Dat is voor de verschillende PM_{10}-meetopstellingen in het meetnet onderzocht en aangetoond. Hierbij is onderscheid gemaakt tussen drie apparaattypen en tussen meetlocaties binnen en buiten het stedelijk gebied.

Het equivalentieonderzoek is grotendeels gebaseerd op de *guideline* die de Europese werkgroep Clean Air For Europe (CAFE) aanbeveelt. Op één onderdeel is daarvan afgeweken. Voor dat onderdeel is een alternatieve wiskundige methode ontwikkeld, die ook in deze studie wordt gepresenteerd.

Trefwoorden:
PM10, equivalentie, kalibratie, fijn stof, bètastof, orthogonaal, regressie, onzekerheid

rivm

Contents

rivm

Summary

Monitoring of the national air quality in the Netherlands is carried out within the framework of the National Air Quality Monitoring Network (LML). This monitoring network is operated by the Dutch National Institute for Public Health and the Environment (RIVM) and comprises approximately 49 monitoring stations distributed throughout the Netherlands. Particulate matter (PM_{10}) at 40 of these stations is measured using automated beta-gauge monitors. European (EU) legislation, however, prescribes the gravimetric method as the reference method (RM), although methods other than the RM are allowed when equivalence with the RM is demonstrated. The RIVM therefore performed an equivalence study, expanded with additional experiments, to improve the quality of the PM_{10} measurements.

The equivalence study is carried out following the guideline of the European Commission (EC) Working Group on Guidance for the Demonstration of Equivalence (2005). This report elaborates on the steps that are taken to demonstrate equivalence. Results of these steps are presented and discussed.. Both the final calibration factors and resulting measurement uncertainty are presented. Equivalence is demonstrated on the basis of these results.

Calibration factors for beta-gauge devices are determined for various configurations. Four candidate methods are defined – the new FH62 I-R monitor at rural sites, the same monitor at urban sites, the old FH62 I-N monitor with an optimized heating system at urban sites and the old FH62 I-N model with the original heating system at urban sites. Each configuration is characterized by a distinct calibration function, which varies between 1.17*x* (FH62 I-N old heating system) and 1.30*x* (FH62 I-N optimized heating system). Only the new FH62 I-R at rural sites shows a calibration with an intercept ($1.17x + 2.7$ μg/m^3). For the FH62 I-R at urban sites a calibration of 1.20*x* is determined.

After calibration of the datasets, the resulting combined relative measurement uncertainties are approximately 17% for all candidate methods, with the exception of the old FH62 I-N with the original heating system, for which the uncertainty is approximately 21%. The combined relative measurement uncertainties for each individual site – after the calibration of the corresponding candidate method has been applied – varies between 10% and 24%, with the exception of one rural site in the north of the Netherlands (uncertainty is approximately 30%). The number of outliers in each of the dataset falls within the scope of that statistically expected, and no outliers have been removed.

Although the guideline of the EC Working Group is used as a guidance for this study, the authors suggest that the guidance document is inconsistent where it concerns the determination of the calibration. Orthogonal linear regression is recommended for the determination of a possible calibration factor. However, in the case of an insignificant intercept, no recommendations are provided by the EC Working Group for orthogonal regression through the origin due to the lack of an algebraic expression for the associated uncertainty. Therefore, an algebraically method is derived in this study and subsequently validated by the statistical bootstrap approach. The derived method is also compared with a modified version of the less complicated formulae given in the guideline. Both present fairly similar resulting uncertainties of the slope that do not differ by more than approximately 0.001 μg/m^3.

Gravimetric samples must be conditioned at 50% (±5%) relative humidity (RH) and 20 °C (± 1 °C) during weighing. A limited number of samples used in this study are weighed at an average RH below the prescribed minimum of 45% – at approximately 43%. An experiment is carried out to determine the

effect of this deviation on the measurement results, with the results showing a correction factor of 1.03 for samples weighed at 43% RH. The samples in question are corrected using this correction factor prior to the determination of equivalence. The associated uncertainty of this correction, weighted by the fraction of affected samples per candidate method, is added to the combined uncertainty.

rivm

1 Introduction

Ambient air particulate matter (PM_{10}) is associated with adverse effects on human health and can cause a significant reduction in life expectancy (WHO, 2004; 2005; 2006). The monitoring and analysis of PM_{10} is therefore required and laid down in European legislation. The European ambient air quality framework, directive 1996/62/EC (EC, 1996) and its follow-up directive, daughter directive 1999/30/EC (EC, 1999), provides the legal structure for the assessment and management of air quality. One of the primary requirements of these directives is to monitor ambient PM_{10} levels and provide hourly air quality information to the general public.

Measuring particulate matter is, however, a rather complex process, and the different methods that have been developed for this purpose each have their own intrinsic advantages and disadvantages, possibly leading to different results. In general, a distinction is made between the semi-automatic gravimetric method and various fully automated methods. The gravimetric method is often carried out using a low volume sampler (LVS) in which PM_{10} particles are collected on a filter. The filter is changed at 24-h intervals and manually weighed in an climate-controlled weighing room. The difference in weight before and after sampling determines the daily average concentration. In contrast, the fully automatic methods do not require manual weighing. Three different types of automatic methods are widely used: the tapered element oscillating microbalance system, the beta-ray absorption analyser and the light-scattering system.

To avoid inconsistent results stemming from differing measuring methods, the Comité Européen de Normalisation (CEN) introduced a European PM_{10} standard. The standard increases harmonization, consistency and comparability between measurements carried out using different methods. The gravimetric method is the European reference method (RM) and is described in CEN standard 12341 (1998). However, as this RM is not suitable for acquiring up-to-minute air quality information due to its limited time resolution, fully automated non-gravimetric measurement methods are allowed – when these methods can demonstrate an equivalence with the gravimetric reference method. Procedures for the demonstration of equivalence are described by the EC Working Group on Guidance for the Demonstration of Equivalence (2005). The application of these procedures is recommended by the Clean Air for Europe (CAFE) steering group.

In the Netherlands, the National Institute for Public Health and the Environment (RIVM) operates the National Air Quality Monitoring Network (NAQMN). In this network, PM_{10} is measured at 40 different locations using fully automatic beta-ray absorption analysers. An equivalence study, based on the guideline of the EC Working Group (2005), with some modifications, is conducted to determine the equivalence between the beta-ray analysers in the NAQMN and the European RM. An earlier published report provides a general and informative overview of the first results and consequences of both the 2006 revalidation of PM_{10} data and this on-going QA/QC (Beijk et al., 2007). The aim of this report is to demonstrate equivalence conform the recommendations. The structure of this report is therefore largely defined by the guideline. While the previous report was primarily focused on the consequences of the determined calibration (and preceding revalidation), this report elaborates on the technical background of the steps taken to demonstrate equivalence, especially those steps that deviate from the guideline of the EC Working Group.

2 Methodology

2.1 Equivalence protocol

Equivalence between the automatic PM_{10} measurements, the candidate method (CM) and the European RM is examined. The methodology used to assess equivalence is largely based upon the guideline of the CAFE Working Group (EC Working Group on Guidance for the Demonstration of Equivalence, 2005). The study consists of several steps, all of which were performed for each CM separately.

FIRST, parallel measurements are performed with the RM and CM simultaneously for a period of approximately 3 years. The Working Group recommends a minimum of four comparisons (sites), with 40 or more samples taken from each site. In this study, samples have been collected at sixteen different sites during various meteorological seasons to account for variations in composition and meteorological conditions. The characteristics of the measurement devices, validation procedure and dataset used in this study are presented in the following paragraphs.

SECOND, data from both the CM and RM are evaluated. The measurements are validated according to the standard procedure described in the following paragraphs, and the uncertainty between samplers (devices) of the same type is evaluated. The Working Group recommends a maximum between-sampler uncertainty for the RM and CM of 2 and 3 $\mu g/m^3$, respectively; equivalence can not be declared if the between-sampler uncertainty exceeds this level, The fulfillment of the between-sampler prerequisites is discussed for both the RM and CM in sections 3.3 and 4.3, respectively.

THIRD, the CM measurements are compared with the concurrent RM measurements. The comparison is made for the entire CM dataset, a subset with values above 50% of the European limit value ($0.5 \times 50\ \mu g/m^3$) and for each comparison (monitoring site) separately. Outliers are not removed from the dataset. The lack of comparability is assessed by means of linear orthogonal regression and the resulting combined relative uncertainty. A correction factor may be applied if this uncertainty (multiplied by the appropriate number of degrees of freedom) exceeds the European quality standard objective.

FOURTH, if needed, a correction factor (calibration) is determined based on the regression analyses on the entire CM dataset. The regression analyses used to determine the calibration of the CM are performed using a slightly different approach than the one recommended by the Working Group. This is discussed in more detail in section 2.2.

LAST, after the CM samples have been calibrated, a new combined relative uncertainty is determined, including the uncertainty of the original regression that was used to define a correction factor. The final expanded uncertainty is again compared with the European quality objective. Equivalence is declared if it does not exceed 25% (see the data quality requirement in the EU guideline: Annex VIII of EC, 1999)

2.2 Statistical method

In this section, both a possible calibration and the lack of comparability between the CM and RM are examined using linear regression analyses. The Working Group on equivalence recommends orthogonal regression, in which an error is assumed in both variables. The equations to determine the slope, intercept and associated uncertainty of these parameters are given in Appendix B of the equivalence document (EC Working Group on Guidance for the Demonstration of Equivalence, 2005). A calibration may be applied prior to determination of equivalence. This factor is calculated using the same method of orthogonal regression, provided that the results present a slope or intercept significantly deviating from one or zero, respectively. The slope and intercept are defined as being significant when the absolute value is more than twofold greater than its uncertainty.

For the determination of a possible calibration, the guideline always calculate a linear slope with a free intercept regardless of whether the intercept appears to be insignificant or not. When the results give an insignificant intercept, the correction factor (calibration) is based on a model with intercept while the intercept correction itself is neglected. This approach is rather inconsistent, which is also recognized by its authors. However, at the time of the guideline, there was no alternative available to calculate an estimate for the slope uncertainty when the regression is forced through the origin.

In Appendix A of this report, we provide a maximum-likelihood method that can be used to determine the uncertainty in an orthogonal regression slope forced through the origin. Based on these equations the standard method, as described in the guideline, can be rewritten. In order to do so, the equations are slightly adjusted (Appendix A) by calculating all sums of squares without the subtraction of the origin. The slope b for a fit using orthogonal regression is calculated with the following equation, both for fits with a free intercept and fits with a forced intercept through the origin:

$$b = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}}{2S_{xy}} \tag{1}$$

When performing regression analyses with a forced intercept through the origin, the terms for S_{yy}, S_{xx} and S_{xy} are then no longer expressed as $S_{xx} = \sum (x_i - \overline{x})^2$, etc., but as follows:

$$S_{xx} = \sum x_i^2 \tag{2}$$

$$S_{yy} = \sum y_i^2 \tag{3}$$

$$S_{xy} = \sum x_i y_i \tag{4}$$

where x_i is the individual reference sample, and y_i is the parallel CM sample. The variance of the slope $u^2(b)$ is then estimated as:

$$u^2(b) = \frac{S_{yy} - \frac{(S_{xy})^2}{S_{xx}}}{(n-1)\cdot S_{xx}} \qquad (5)$$

where n is the number of data pairs, and S_{yy}, S_{xx} and S_{xy} are as defined in Eqs. (2) to (4). The validity of Eq (5) is confirmed by a maximum likelihood derivation (see Appendix A) and a bootstrap simulation in which alternative datasets are sampled from the original dataset. The spread in the results of the alternative datasets equals the results of Eq. (5).

When the CM is corrected for an intercept significantly deviating from zero, the contributing uncertainty is calculated as described in the guideline:

$$u^2(a) = u^2(b)\frac{\sum x^2}{n} \qquad (6)$$

2.3 Determination of relative measurement uncertainty

The uncertainty in the results of the comparison between the CM and RM after calibration (if needed) consists of several terms. If corrected for a deviating slope, intercept or both, then the associated uncertainty does contribute to the relative measurement uncertainty. The general equation describing the uncertainty due to a lack of comparability is based on the equation given in the guideline of the EC Working Group. The definition consists of several terms:

TERM A: Residual sum of squares for the dataset to be tested for equivalence
TERM B: Lack of comparability between the CM and RM
TERM C: Uncertainty of the applied calibration, if any; discussed and defined in previous paragraph.
TERM D: Subtraction of RM standard uncertainty; see section 3.4
TERM E: Uncertainty due to RM correction, added for this study specifically; see Eq 9.

$$u^2(y_i) = \boxed{\frac{RSS}{(n-2)}}_{A} + \boxed{\left(c + (d-1)x_i\right)^2}_{B} + \boxed{x_i^2 u^2(b) + u^2(a)}_{C} - \boxed{u^2(x_i)}_{D} + \boxed{u^2_{crm}}_{E} \qquad (7)$$

where n is the number of sample-pairs, x_i is the limit value, c is the regression slope of the (calibrated) dataset, d is the resulting intercept , $u^2(b)$ is the uncertainty of the original slope if used for calibration, $u^2(a)$ is the uncertainty of the original intercept if used for calibration and $u^2(x_i)$ is the between-sampler uncertainty.

If no calibration is applied the variables, c and d represent the regression slope and intercept of the original dataset, respectively. In this case, TERM C can be neglected (whereas $u^2(b) = 0$ and $u^2(a) = 0$). When corrected for a slope significantly deviating from 1, the associated uncertainty $u^2(b)$ is defined and calculated as presented in section 2.2; when corrected for an intercept significantly deviating from 0, the associated uncertainty $u^2(a)$ is calculated using Eq. (6). The root summed square (RSS) is calculated using Eq. (8).

$$RSS = \sum_{i=1}^{n} (y_i - c - d \cdot x_i)^2 \quad (8)$$

Correction RM samples contributing to uncertainty budget
The RM samples are weighed at approximately 43% relative humidity (RH) for a limited period of time, then corrected as described in section 3.4. The uncertainty of this correction contributes to the relative measurement uncertainty of the CM and is therefore added to Eq. (7). If only a part of the samples within a single comparison is corrected, the contributing uncertainty is equally reduced by multiplication with the appropriate ratio:

$$u_{crm}^2(y_i) = u_{rm\_c}^2 \frac{n_{RM,corrected}}{n_{RM}} \quad (9)$$

where $u^2_{rm\_c}$ is as defined in Eq. (11), n is the number of corrected RM samples and n_{RM} is the total number of RM samples.

2.4 Testing equivalence

Following the guideline of the Working Group, the CM comparison uncertainty $u^2(y_i)$ is multiplied by a coverage factor of two (representing approximately a 95% confidence interval) to obtain the expanded uncertainty. Finally, to test for equivalence, the expanded uncertainty is divided by 50 µg/m^3 to calculate the expanded relative (measurement) uncertainty at the relevant European limit value. The expanded relative measurement uncertainty is then compared to the European quality objective (a maximum measurement uncertainty of 25%) to determine if equivalence may be declared.

3 Reference method

3.1 Characteristics

The reference method for PM_{10} measurements is the gravimetric method as described in the EN12341 standard (CEN, 1998). All reference measurements are carried out with a low volume sampler (LVS) conform to this standard. The technical specifications of the LVS used in this study are listed in Table 1. The LVS makes use of filters that are weighed before and after sampling. When the LVS is operating in the field, a constant airflow passes through this filter. Particles ≤ 10 μm are collected on this filter, and the filter is replaced at 24-h intervals. The daily average PM_{10} concentration is obtained by calculating the difference in weight between the filters before and after sampling and then dividing the result by the airflow volume using the following equation
Equation (10).

$$PM_{10,DA} = \frac{F_{DL} - F_B}{V_{AF}} \qquad (10)$$

where $PM_{10,DA}$ is the daily average PM_{10} concentration measured in μg/m^3, F_{DL} is the dust-loaded filter-weight after sampling (μg), F_B is the blank filter-weight before sampling and V_{AF} is the total volume of the airflow used during the entire sampling period (m^3/24 h) .

Two filter types have been used: QF20 and, more recently, QMA filters; both are quartz filters. The results of an experiment (details provided in Appendix B) reveal that there is no significant difference between the two filters in terms of PM_{10} concentrations, The analyses of the filters are carried out in an air-conditioned weighing room maintained at the pre-prescribed conditions of 50% (±5%) RH and 20 °C (±1 °C). New filters are acclimatized at these conditions for at least 2 days prior to field usage, and the 24-h filters used in the field measurements are weighed under the same conditions. The samples are validated following the weighing process.

3.2 Validation

The gravimetric PM_{10} measurements are validated based on five conditions:

- All new filters are weighed twice before being used. The maximum allowed difference between the two measurements is 40 μg.
- The airflow during sampling is not allowed to deviate more than 2.5 m^3 per 24-h period.
- After sampling, all filters are weighed another two times. The difference between these two measurements cannot exceed 60 μg.
- The temperature and RH level in the weighing room is continuously monitored. If the RH or temperature deviates more than the maximum allowed variance (±5% RH and ±1 °C, respectively), then the sample in question will be rejected. However, until the beginning of 2006 the weighing room appears to have systematically operated at a RH below 45%. An experiment has been conducted to determine a correction factor for this deviation instead of all samples being rejected. See section 3.4.

- Remarks of individuals involved in the whole measurement process are considered during the validation process. For example, the result from a filter contaminated by parts of insects can be rejected during validation.

Table 1 Technical specifications of the PM10 reference method

	Gravimetric Reference method
Device characteristics	
Device	Sven Leckel SEQ 47/50
Filter type	Quartz (QF20, QMA)
Filter pre-conditioning	48 h at 50% RH and 20 °C
Airflow	55.2 m^3/24 h
Sample frequency	24 h
Maximum filter storage time after sampling	1 month
Temperature regulation	20 °C (at most rural sites) or fan cooling (at most urban sites)
Validation criteria	
Technical deviations	57.7 > [Airflow m^3/24 h] > 52.7
Blank filter pre-sampling test	≤ 40 µg difference between two consecutive laboratory tests
Filter post-sampling test	≤ 60 µg difference between two consecutive laboratory tests
Measurement of room conditions	Maximum deviation of 5% RH and 1 °C air temperature

3.3 Comparability

The comparability of RM devices is described by the random uncertainty term, $u^2(x_i)$, which in the RM used in this study is determined in a filter dependency experiment. In this experiment, a set of eleven RM devices are placed together, and equitability between different choices of filter material are examined together with the between-sampler uncertainty (see Appendix B). The results reveal a between-sampler uncertainty of 1.6 µg/m^3, although lower between-sampler uncertainties have been obtained elsewhere; see, for example, the UK Equivalence Programme (Harrison et al., 2006). Another RIVM experiment that examined a possible difference between devices placed in open air and those inside an air-conditioned measurement room also obtained a smaller uncertainty (0.8 µg/m^3). The current filter dependency experiment contains the most extensive and recent dataset. Therefore, a between-sampler uncertainty of 1.6 µg/m^3 is used for the RM in this study. This uncertainty is subtracted from the uncertainty budget when the relative measurement uncertainty of the CM is being calculated [see Eq. (7)].

3.4 Correction of samples weighed at 43% relative humidity

Some of the RM samples have been analysed at an average RH of 43% due to shortcomings of the concerned device in the weighing room for a limited period of time. The European standard prescribes the RH to be 50% (±5%). Consequently, an experiment has been conducted to determine the effect of this deviation on the measurement results.

The change in filter mass can be examined either by increasing the RH from 43% to 50% or by decreasing it from 50% to 43%. However, hysteresis may occur when the RH is increased. In addition, the atmospheric RH in the Netherlands rarely reaches levels below 50%. Therefore, to mimic the field conditions as closely as possible, the experiment is performed by lowering the RH in the weighing room. A 40% RH was chosen for practical reasons. Assuming a linear relation, the result can then be interpolated to 43%.

The weight of the blank filter itself changes when it is acclimatized at different humidity levels. As this change has also to be accounted for when determining the effect on the final concentration, the experiment is carried out in two steps. A sample set of 64 blank filters are first weighed, each filter individually, at 50% RH. The weighing room is then brought to 40% RH and the sample filters weighed a second time. The difference between the first and second measurements provides an estimate of the effect of weighing the blank filters at 40% RH instead of 50%. Second, a sample set of 63 filters are initially weighed at 50% RH and the filters placed into the field at five different locations (sites 131, 636, 240, 448 and 544). Table 2 lists the characteristics of each set of samples. After the appropriate sampling interval, the field sample filters are weighed at 50% RH and again at 40% RH. The difference between the first and second measurements provides an estimate of the effect of the deviating RH levels during weighing. Prior to determining the final correction factor, the effect on the blank filters is subtracted from the field samples. The final correcting factor is then computed using ordinary least square analysis, with the corrected field samples weighed at 40% RH as the dependant variable and the (uncorrected) field samples weighed at 50% RH as the independent variable (see Figure 1).

The result is a relative difference of –4.4% ($R^2 = 0.99$) when the field samples are corrected with the offset found for the blank filters (–0.94 µg/m^3) (see Figure 1). Assuming a linear relation, historic measurements weighed at 43% RH need to be corrected with a factor of 1.03 (100% + 4.4% × 0.7).

A contribution to the relative uncertainty has to be accounted for in corrected RM measurements included in the equivalence study. This contribution ($u^2_{rm\_c}$) to the relative measurement uncertainty is defined in Eq. (11). Using this equation, the contribution is calculated to be 0.6 µg/m^3 (0.83 µg/m^3× 0.7).

$$u^2_{rm\_c} = \frac{\sum (x_i - y_i)^2}{2n} \qquad (11)$$

where x_i and y_i are the uncorrected and corrected RM measurements, respectively, used to determine the correction factor, and n is the number of samples used for determining the correction factor.

Table 2 Characteristics of the dataset used for the gravimetric measurement correction

	131	240	448	544	636	All samples
Location	Vredepeel	Breda	Rotterdam	Amsterdam	Utrecht	
Type	Rural	Urban	Urban	Urban	Urban	
Samples (*n*)	14	14	9	14	12	63
Average µg/m^3 (RH = 50%)	25.0	25.6	19.8	13.2	23.8	21.5
Standard deviation (RH = 50%)	12.9	6.3	4.5	3.3	4.8	8.7
Average µg/m^3 after blank filter correction (RH = 40%)	23.5	24.4	19.0	13.5	22.8	20.7
Standard deviation (RH = 40%)	12.2	5.9	4.1	3.3	4.5	8.0

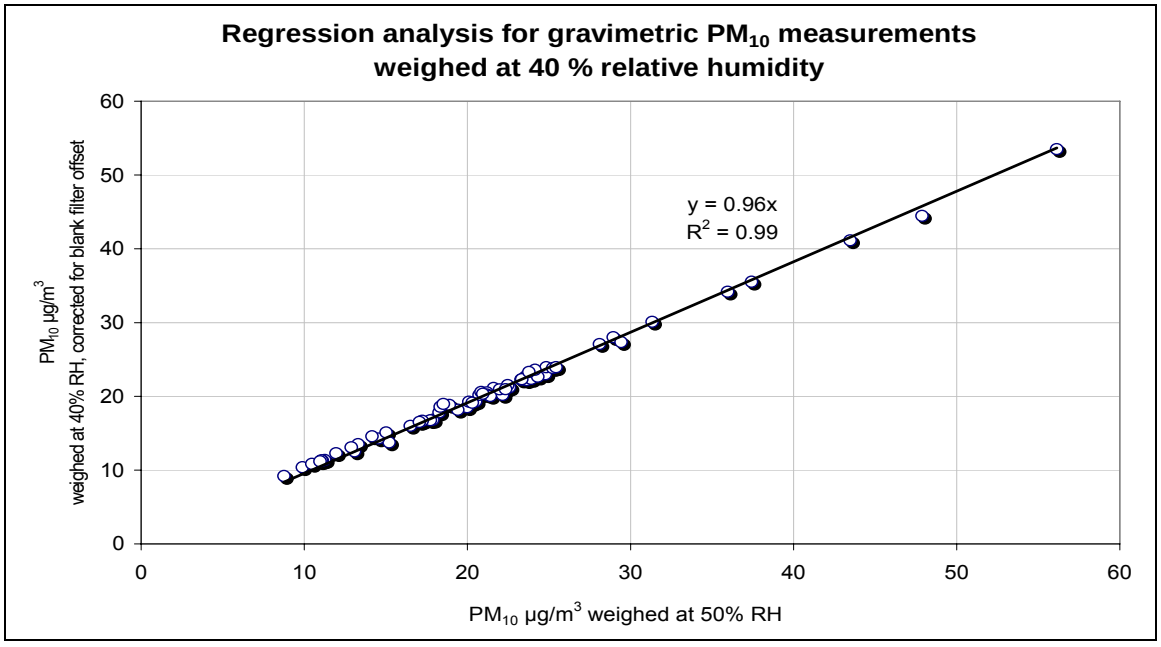


Figure 1: Regression analysis (ordinary least square) for PM_{10} reference samples weighed at 40% RH versus samples measured at 50% RH. The samples weighed at 40% are corrected for the offset that appeared while acclimatizing the blank filters at 40% RH.

4 Candidate method

4.1 Characteristics

In the Dutch National Air Quality Monitoring Network (NAQMN), the number of automated PM_{10} monitoring sites expanded from approximately 20 to 40 between 2003 and 2006. In 2006, roughly half of the monitoring sites were equipped with an old monitor model (FH62 I-N), while the other half were equipped with the newer model (FH62 I-R). All old models will be replaced with the new type in 2007 and 2008. The equivalence for both models is examined for consistency between historical measurements. For technical details of each candidate method, see Table 3.

The heating system of the old model has in recent years been replaced by a new type of heating system due to discontinuance in the distribution of the old heating instrument. The method of heating has an important influence on the possible loss of volatile material. The new heating device is more than fivefold longer than the old one and, therefore, the two devices differ significantly (Table 3). Consequently, a distinction is made between the FH 62 I-N model with an old heating system and that with the new one, thereby introducing a third candidate method.

Table 3 Overview of candidate methods (CM) for PM_{10} measurements in the Netherlands.

	Candidate method I	Candidate method II	Candidate method III
Brand	Thermo ESM Andersen	Anderson	Anderson
Type/version	FH 62 I-R	FH 62 I-N	FH 62 I-N
Heating method	On pipe, Δ 180 cm	On pipe, Δ 170 cm	Trough mantle, Δ 30 cm
Heating parameter	Ambient + 10 °C	50 °C	50 °C
Temperature regulation	At approximately 20 °C	Cooling fan	Cooling fan
I/O method	Analogue 4–20 mA	Analogue 4–20 mA	Analogue 4–20 mA
Sample frequency	Hourly	Hourly	Hourly
Filter change	25th hour	25th hour	25th hour
Absolute minimum	–5 μg/m^3	–5 μg/m^3	–5 μg/m^3
Absolute maximum	1000 μg/m^3	1000 μg/m^3	1000 μg/m^3
Hour after filter change	Value deleted	Value deleted	Value deleted
Correction for ambient temperature and pressure	Actual	Manually	Manually
Allow negative values	≥ –5 μg/m^3	≥ –5 μg/m^3	≥ –5 μg/m^3

4.2 Validation

The automatic PM_{10} measurements (CM) are validated based on four criteria:

- Values below –5 $\mu g/m^3$ are automatically rejected.
- In the case of mechanical alerts and technical malfunctions, a signal of 2 mA is given by the monitor; these values are automatically rejected.
- A comparison with comparable monitoring stations and other measurements at the station are considered in order to determine a possible malfunctioning of the device.
- The first sample taken after each filter change is automatically rejected as this sample is typically different. The daily average concentration is based on the European Union's criteria of a minimum of 13 validated samples per 24-h period (Guideline 2001/752/EC).

4.3 Comparability

Two experiments are carried out in which a between-sampler uncertainty is determined for the FH62-IR. In the first experiment, PM_{10} concentrations are measured with four parallel samplers for a period of 19 days. Based on the results of this experiment, a between-sampler uncertainty of 0.76 $\mu g/m^3$ is calculated (see Appendix D). In a second experiment, two samplers are placed together for approximately 1 year. Based on the results of this experiment, a between-sampler uncertainty of 2.56 $\mu g/m^3$ is calculated. Both determined uncertainties comply with the between-sampler limit as recommended by the EC Working Group (see also Appendix D).

4.4 Recalculation

The procedures, device configurations and settings in the NAQMN were checked thoroughly during the course of 2006 with the aim of improving the quality of the measurements and to prevent technical deviations from influencing the equivalence study. The observations have led to a revalidation of measurements and an improvement of procedures. The equivalence study is carried out using the revalidated data only. More information on the entire revalidation process can be found in the RIVM publication (in Dutch) of Beijk et al. (2007).

Historically, the old heating device (model FH62 I-N) was not equipped with an ambient temperature and pressure sensor. Prior to 2003 and, in some cases, up to 2005, particulate matter concentrations were therefore reported using standard conditions (20 °C, 1013 hPa). European legislation states that PM_{10} concentrations are to be based on prevailing atmospheric conditions. Consequently, historical data are recalculated to meet this demand and to ensure consistency in trends. Neither atmospheric temperature nor pressure is simultaneously available with all PM_{10} measurements, and meteorological data from the Dutch Royal Meteorological Institute (KNMI) are used to recalculate the PM_{10} measurements. Although the distance between KNMI stations and that between PM_{10} monitors may differ, the effect of this distance on the recalculated concentrations is negligible. The revalidation is based on the ideal gas law:

$$PM10_{nieuw}(t) = PM10_{oud}(t) \times \frac{p(t)}{1013} \times \frac{293}{273 + T(t)} \qquad (12)$$

where for each measured hour-average *t*, PM_{10} ($\mu g/m^3$) concentrations are corrected with the ambient and standard pressure (hPa) ratio and with the standard and environment temperature ratio (°C). The recalculation is further discussed and elucidated in a report by Berkhout et al. (2008).

5 Equivalence results

The results are given for the entire dataset of each CM, for a split dataset (samples higher than 50% of the European upper assessment threshold) and for each individual comparison (site). A possible calibration of a CM is based on the outcome for the entire dataset. Results of the split dataset and those of the individual sites illustrate the extent to which the results of the full CM dataset reflect prevailing field characteristics at the various locations.

5.1 Data characteristics

In total, 763 daily average samples at eight different sites are used for the comparison between the RM and the new (automatic) sampler at rural locations (CM-I rural). A total of 463 samples measured at four different sites comprises the dataset for urban sites (CM-I urban). Reference measurements for comparison with the old monitor model are only available at urban sites: 181 samples at two different sites for those with the original (static) heating device (CM-II) and 239 samples at two different sites for the old model with the new heating device (CM-III). The geographical position of each sampling sites is given in Figure 2. Outliers are not removed from any of the datasets during the analyses carried out in this study. The number of statistical outliers (samples outside the 99% confidence interval) are listed along with the results.

Candidate methods II and III are one and the same device. It was initially considered to treat these as one CM. However, the alteration in the inlet heating device leads to an important difference, and the dataset is therefore separated into two datasets. The consequence of separating the dataset is that the now following two candidate methods no longer comply with the recommended minimum number of comparisons. Both candidate methods are no longer technically supported and are replaced with a new type in the NAQMN. Despite the limited datasets the differences between CM-II and CM-III are considered to be of such magnitude that treating them separately would meet the field conditions better opposed to treating them as one.

An overview of the characteristics and meteorological conditions of each CM is given in Table 4. The latter is given in percentiles to provide an impression of the distribution of prevailing meteorological conditions within each dataset. Where available, aerosol measurements are also presented (in percentiles).

5.2 Calibration and equivalence of the four candidate methods

The regression results for the CM (full datasets) assessed here are shown in Figure 3 to Figure 6. The results for the new device (CM-I: FH62 I-R) at rural sites show a slope of 1.17, with a 2.7 $\mu g/m^3$ offset; this is the only category with a statistical significant intercept. The relative (measurement) uncertainty for this category is 17%. Results for the same device but at urban sites present a slope of 1.20 based on orthogonal regression forced through the origin, with a relative uncertainty of 17%. Both sets of results comply with the European quality objective of 25% (see also Table 5 for the statistical results for each CM). The regression results for the old device (CM-II and CM-III; FH62 I-N new and old heating, respectively) at urban sites shows a slope of 1.30 and 1.20 for the configuration with the new and old

heating, respectively. The relative uncertainty of the former is 18%, and that of the latter is 21%. Consequently, both uncertainties are within the boundaries of the European quality objective of 25%.

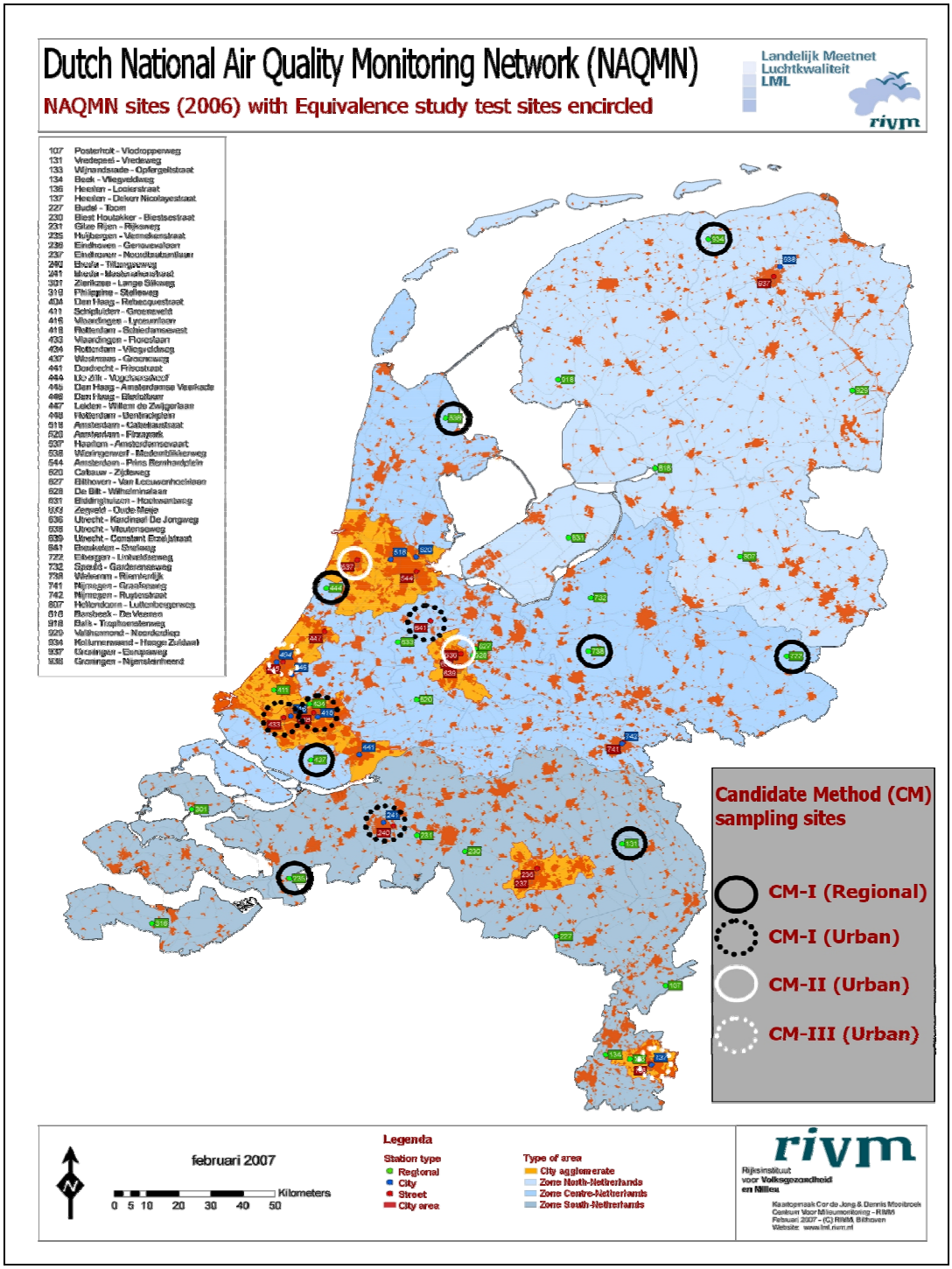


Figure 2 Overview of sampling locations in the NAQMN for the purpose of the 2006 equivalence study. Locations are indicated for each candidate (measurement) method separately.

Table 4 Dataset characteristics of each CM

	CM-I **FH62 I-R**		**CM-II** **FH62 I-N** New heating device	**CM-III** **FH62 I-N** Old heating device
Location type	Rural	Urban	Urban	Urban
Representatives				
PM_{10} total samples	763	463	239	181
PM_{10} summer samples	282	220	97	84
PM_{10} winter samples	481	243	142	97
Temperature °C (percentile 0.25)	4.4	5.2	9.9	2.5
Temperature °C (percentile 0.75)	15.2	16.2	16.6	11.9
Relative humidity (percentile 0.25)	77	76	81	73
Relative humidity (percentile 0.75)	90	88	89	88
Pressure (percentile 0.25)	1010	1009	1007	1012
Pressure (percentile 0.75)	1022	1022	1021	1024
Aerosol samples	417	0	0	0
NH_4 µg/m^3 (percentile 0.25)	0.7	n/a	n/a	n/a
NH_4 µg/m^3 (percentile 0.75)	2.1	n/a	n/a	n/a
NO_3 µg/m^3 (percentile 0.25)	1.5	n/a	n/a	n/a
NO_3 µg/m^3 (percentile 0.75)	4.5	n/a	n/a	n/a
SO_4 µg/m^3 (percentile 0.25)	1.2	n/a	n/a	n/a
SO_4 µg/m^3 (percentile 0.75)	2.9	n/a	n/a	n/a

(n/a, Data not available)

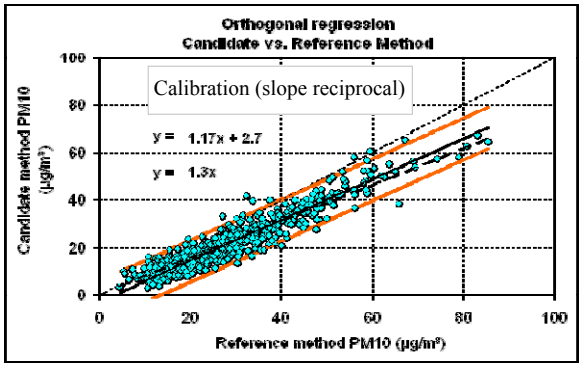


Figure 3: Calibration curve for CM-I (FAG62-IR; y-axis) versus reference method (x-axis) at rural **sites. The confidence interval (99%) is drawn above and below the calibration curve.**

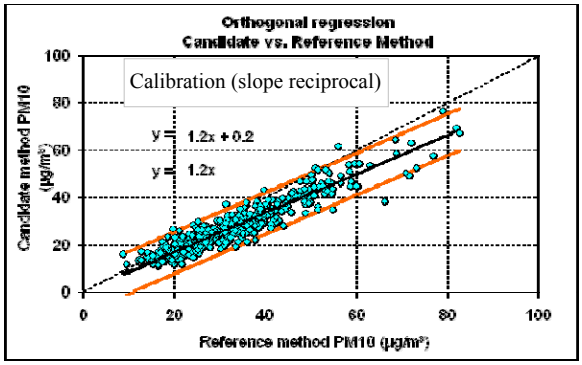


Figure 4: Calibration curve for CM-I (FAG62-IR, y-axis) versus reference method (x-axis) at urban **sites. The confidence interval (99%) is drawn above and below the calibration curve.**

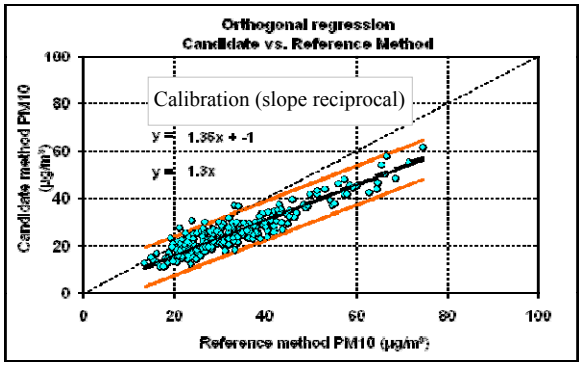


Figure 5: Calibration curve for CM-II (FAG62-IN new heating; y-axis) versus the reference method at urban **sites. The confidence interval (99%) is drawn above and below the calibration curve.**

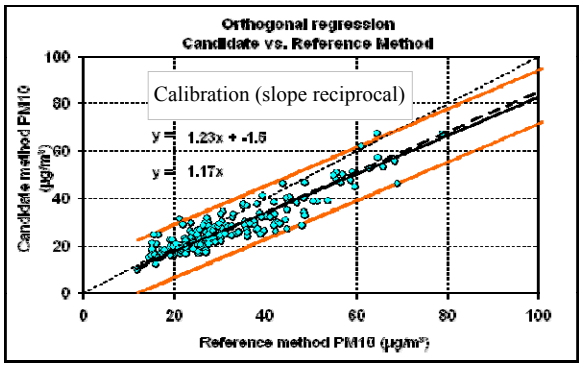


Figure 6: Calibration curve for CM-III (FAG62-IN old heating, y-axis) versus the reference method at urban **sites. The confidence interval (99%) is drawn above and below the calibration curve.**

Table 5 Calibration and equivalence test results for each CM using the full datasets

	CM-1: FH62 I-R		CM-2: FH62 I-N New (optimized) heating device	CM-3: FH62 I-N Old (original) heating device
Location type	**Rural**	**Urban**	**Urban**	**Urban**
Calibration results				
Samples	763	463	239	181
Slope correction	1.17 x	1.20 x	1.30 x	1.17 x
Intercept correction	2.7 µg/m^3	*Not significant*	*Not significant*	*Not significant*
Outliers (99% level) [1]	16 (2 %)	12 (3 %)	5 (2 %)	4 (2 %)
Mean RM	26.9 µg/m^3	31.9 µg/m^3	32.3 µg/m^3	32.3 µg/m^3
RM samples > LV	46	46	22	20
RM samples > 0.5 LV	363	298	159	122
CM samples > LV (calibrated)	50	47	22	19
CM samples > 0.5 LV (calibrated)	340	295	166	125
Equivalence test results				
CM calibration	1.17 x + 2.7	1.20 x	1.30 x	1.17 x
Relative uncertainty	16.6%	17.1%	17.5%	20.8%

[1]) Detected, not deleted

RM, Reference method LV, limit value

5.3 Level of representativeness

The results of the split datasets with only samples greater than 50% of the limit values are given in Table 6. Although the regression results differ slightly from those obtained with the full datasets, the resulting relative measurement uncertainty, which is calculated after the application of the appropriate CM calibration, remains beneath 25% for all split datasets

The results for each individual comparison (monitoring site) are given in Table 7 to Table 10. To determine the relative uncertainty of each comparison separately, each sub-dataset is calibrated by applying the earlier determined CM calibration – not by using the individual regression result. After the sub-datasets are calibrated, the results for all individual comparisons except one present a relative uncertainty that lies beneath the quality objective. Site 934 (Kollumerwaard, in the north of the Netherlands) presents a relative uncertainty of 29%, which is somewhat above the quality objective.

Table 6 Regression and equivalence test results using datasets with only samples *greater than or equal to 50% of the limit value* (≥ 25 µg/m³). Calibration is based on the results of the full dataset.

	CM-1: FH62 I-R		CM-2: FH62 I-N New heating device	CM-3: FH62 I-N Old heating device
Location type	**Rural**	**Urban**	**Urban**	**Urban**
Calibration results				
Samples	363	298	159	122
Slope correction	1.09 x	1.13 x	1.32 x	1.19 x
Intercept correction	5.6 µg/m^3	2.4 µg/m^3	*Not significant*	*Not significant*
Outliers (99% level) [1]	7 (2%)	7 (2%)	3 (2%)	1 (1%)
Equivalence test results				
Applied CM calibration	1.17 x + 2.7	1.20 x	1.30 x	1.17 x
Relative uncertainty	19.9%	19.3%	18.3%	22.4%
Mean RM	37.0 µg/m^3	38.3 µg/m^3	37.9 µg/m^3	38.4 µg/m^3
Mean CM (calibrated)	36.5 µg/m^3	37.7 µg/m^3	37.4 µg/m^3	37.4 µg/m^3

[1]) Detected, not deleted

Table 7 Regression and equivalence test results for *individual sites* (*rural background stations, CM-I*). The calibration is based on the results of the full dataset of the appropriate candidate method.

Site ID	**131**	**235**	**437**	**444**
Candidate Method I	**FH62 I-R**	**FH62 I-R**	**FH62 I-R**	**FH62 I-R**
Location type	**Rural**	**Rural**	**Rural**	**Rural**
Calibration results				
Samples	247	39	52	64
Slope correction	1.15 x	1.25 x	1.35 x	1.23 x
Intercept correction	5.8 µg/m^3	1.6 µg/m^3	*Not significant*	*Not significant*
Outliers (99% level) [1]	13 (5%)	0 (0%)	1 (2%)	1 (2%)
Equivalence test results				
CM calibration	1.17 x + 2.7	1.17 x + 2.7	1.17 x + 2.7	1.17 x + 2.7
Relative uncertainty	15.7%	11.5%	10.8%	18.3%

[1]) Detected, not deleted

Table 8 Regression and equivalence test results for *individual sites* (*rural background stations, CM-I*). The calibration is based on the results of the full dataset of the appropriate candidate method.

Site ID	538	722	738	934
Candidate Method I	**FH62 I-R**	**FH62 I-R**	**FH62 I-R**	**FH62 I-R**
Location type	**Rural**	**Rural**	**Rural**	**Rural**
Calibration results				
Samples	58	105	134	64
Slope correction	1.38 x	1.16 x	1.12 x	1.10 x
Intercept correction	–2.7 μg/m^3	3.7 μg/m^3	1.7 μg/m^3	*Not significant*
Outliers (99% level) [1]	1 (2 %)	4 (4 %)	5 (4 %)	3 (5 %)
Equivalence test results				
CM calibration	1.17 x + 2.7	1.17 x + 2.7	1.17 x + 2.7	1.17 x + 2.7
Relative uncertainty	8.4 %	15.2%	21.5%	30.3%

[1]) Detected, not deleted

Table 9 Regression and equivalence test results for *individual sites* (*urban stations, CM-I*). The calibration is based on the results of the full dataset of the appropriate candidate method.

Site ID	240	433	448	641
Candidate Method I	**FH62 I-R**	**FH62 I-R**	**FH62 I-R**	**FH62 I-R**
Location type	**Urban**	**Urban**	**Urban**	**Urban**
Calibration results				
Samples	69	67	248	79
Slope correction	1.36 x	1.25 x	1.20 x	1.18 x
Intercept correction	*Not significant*	*Not significant*	*Not significant*	*Not significant*
Outliers (99% level) [1]	6 (9%)	2 (3%)	7 (3%)	2 (2%)
Equivalence test results				
CM calibration	1.20 x	1.20 x	1.20 x	1.20 x
Relative uncertainty	23.7 %	18.8 %	15.6 %	18.5 %

[1]) Detected, not deleted

Table 10 Regression and equivalence test results for *individual sites* (*urban stations, CM-II/III*). The calibration is based on the results of the full dataset of the appropriate candidate method.

Site ID	**537**	**636**	**137**	**404**
Candidate Method II & III	**FH62 I-N New heating**	**FH62 I-N New heating**	**FH62 I-N Old heating**	**FH62 I-N Old heating**
Location type	**Urban**	**Urban**	**Urban**	**Urban**
Calibration results				
Samples	120	119	92	89
Slope correction	1.36 x	1.40 x	1.11 x	1.24 x
Intercept correction	–3.0 µg/m^3	*Not significant*	*Not significant*	*Not significant*
Outliers (99% level) [1]	4 (3%)	8 (7%)	2 (2%)	0 (0%)
Equivalence test results				
Applied calibration	1.30 x	1.30 x	117 x	1.17 x
Relative uncertainty	19.1%	18.2%	20.5%	23.9%

[1]) Detected, not deleted

5.4 Aerosols (ammonium nitrate)

Ammonium nitrate is assumed to be important as a volatile component of PM_{10}. Due to the preheating of the airflow in the sampler, this component may become lost before the actual measurement inside this unit actually takes place. This component may therefore influence the difference between the CM and the RM method with respect to the calibration term(s).

In a first attempt to study the effect of ammonium nitrate on the calibration, the dataset is split in two subsets: one in which the PM_{10} samples contain a high (≥ 15) percentage of ammonium nitrate and one with a low (<15) percentage. Both of these data subsets comprise slightly more than 200 daily average samples, of which somewhat more than 25% are measured during the 'summer' half year (April–September) at rural sites.

The resulting calibration curve for both datasets are illustrated in Figure 6. In both cases, the results are nearly identical: a slope of 1.19, with an intercept of approximately 2.6 µg/m^3. The slope as well as the intercept are statistically significant for both datasets (twofold greater than its uncertainty). No clear relation between the level of ammonium nitrate and the calibration can currently be identified.

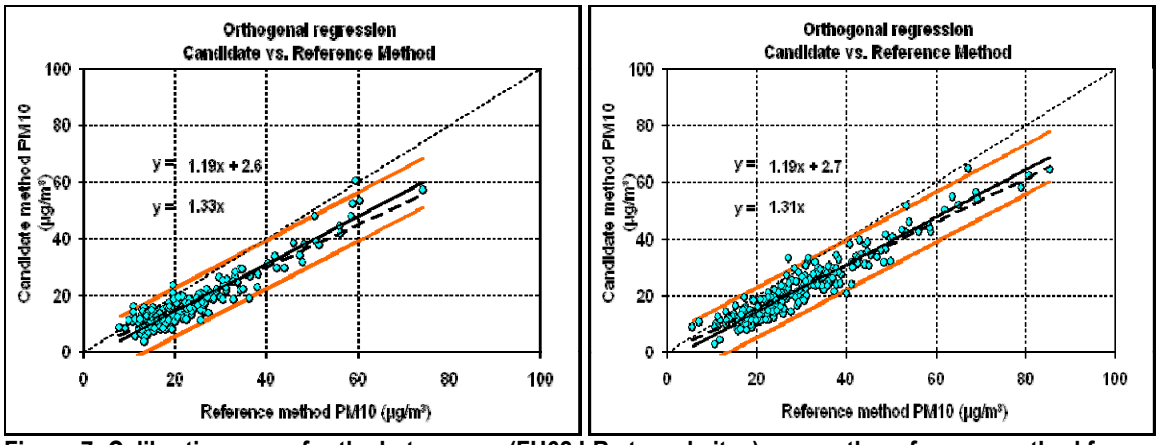


Figure 7: Calibration curve for the beta-gauge (FH62 I-R at rural sites) versus the reference method for samples with less than 15% ammonium nitrate (left) and those with ≥ 15% (right).

6 Conclusion

Outcome
The automatic PM_{10} measuring method (beta-attenuation) in the NAQMN is compared with the European RM. Based on the results of this comparison, a calibration is introduced for four different configurations: the new model (FH62 I-R) at rural locations, the new model at urban locations, the old model (FH62 I-N) with a new heating system and the old model with an old heating system. Equivalence is demonstrated for all four configurations after applying the assessed calibration. The resulting relative uncertainty for each CM varies between 16% and 21%. Following calibration of the CM measurements, fifteen of the sixteen locations included in this study comply with the European quality objective.

Where possible, a distinction is made between rural and urban locations. A sufficient number of parallel measurements are available for the new model to make this distinction. The old model, however, is in the process of being replaced with the new model and is steadily being phased out of the NAQMN. Parallel reference measurements with the old model are limited to only urban locations; consequently, it is not possible to distinguish between urban and rural locations using this model. Although a comparison at rural locations is necessary to reconstruct a calibration for historical data, such a comparison is of little consequence for future PM_{10} measurements since the old model is phased out of the monitoring network.

Definition of CM
The distinction between rural and urban locations raises a question relating to the definition of the CM. Two different beta-attenuation devices (old and new) are used in the NAQMN. The old model is physically modified through the replacement of the heating inlet system. While the monitoring device is basically the same, this modification might cause a difference in calibration. In addition, the same device at different locations is likely to lead to differences in calibration due to typical variations in particle composition – which is the reason underlying the EC Working Group's recommendation for the test sites to be representative of 'typical conditions'. However, there is currently no definition for determining whether or not a certain configuration is within the scope of these 'typical conditions'.

In this study, a CM is defined as a distinctive measuring device. Modification of the heating system leads to a device that is to be distinguishable from the original. Because of expected differences in composition between urban (both city background and street locations) and rural sites (background locations with possibly a larger fraction of secondary aerosols), a distinction is also made between the these two location types, leading to two subcategories.

Calibration method and regression forced trough the origin
Orthogonal regression is used for the determination of a possible calibration for each CM. Orthogonal regression without intercept is applied in the case of an insignificant intercept. The equations for orthogonal regression without intercept and corresponding uncertainty are not available in the equivalence guideline. Such an approach was considered necessary; therefore, the statistical equations are presented and applied in this study.

Uncertainties

Conditions in the weighing room in which the RM samples are weighed appear to have deviated from the required 50% RH (±5%) up to 2006. This deviation has been corrected for, leading to a small contribution to the estimated relative uncertainty of the CM. This contribution is accounted for while testing equivalence. Another uncertainty that may have an effect when determining equivalence is the preconditioning of the RM filters. Although all filters are preconditioned for a minimum of 2 days – as prescribed in the CEN 12341 standard – dry filters may require a prolonged conditioning to reach an equilibrium. Currently applied procedures in the NAQMN have been modified so that all new filters reach their equilibrium before being used for sampling. Also, the weighing room has been completely modified and is ISO 17025 accredited. Both activities reduce uncertainties in current and future RM measurement results.

rivm

References

Beijk, R., Hoogerbrugge, R., Hafkenscheid, T.L., Arkel, F.Th. van, Stefess, G.C., Meulen, A. van der, Wesseling, J.P., Sauter, F.J., Albers, R.A.W. (2007) *PM10: Validatie en equivalentie 2006* (in Dutch). Bilthoven: National Institute for Public Health and the Environment (RIVM). Report 680708001.

Berkhout, J.P.J., Mooibroek, D., Pul, W.A.J., Hoogerbrugge, R. (2008) *Herberekening historische meetgegevens fijnstof* (in Dutch). Bilthoven: National Institute for Public Health and the Environment (RIVM). Report 680708005.

CEN (1998) *Air quality – Determination of the PM10 fraction of suspended particulate matter – reference method and field test procedure to demonstrate reference equivalence of measurement methods.* Brussels: CEN 12341.

EC Working Group on Guidance for the Demonstration of Equivalence (2005) *Demonstration of equivalence of ambient air monitoring methods*. Published on the internet, February 28, 2007: Available at: http://ec.europa.eu/environment/air/pdf/equivalence_report2.pdf.

EC (1996) *Council directive 96/62/EC of 27 September 1996 on ambient quality assessment and management*. Official journal L 296, 21/11/1996 p. 0055-0063.

EC (1999) *Council Directive 1999/30/EC of 22 April 1999 relating to limit values for sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air*. Official Journal L 163: 29/06/1999 p. 0041 – 0060.

Harrison, D., Maggs, R., Booker, J. (2006) *UK Equivalence Programme for Monitoring of Particulate Matter*. Bureau Veritas, Report Reference BV/AQ/AD202209/DH/2396.

Isobe, T., Feigelson, E.D., Akriteas, M.G., Babu, G.J. (1990) *Linear regression in Astronomy.* Astrophysical Journal, vol. 364, pp. 104-113

WHO (2004) *Health aspects of air pollution. Results from the WHO project "Systematic review of health aspects of air pollution in Europe"*. European centre for Environment and Health – Bonn Office. Denmark: World Health Organization Europe. Publication E83080.

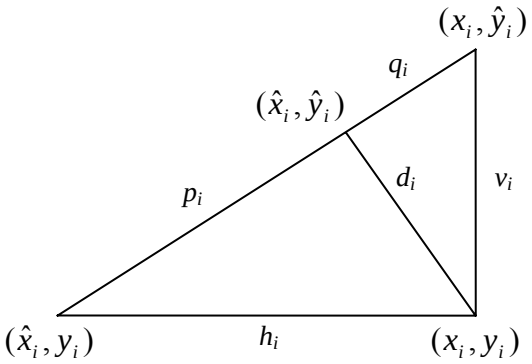
WHO (2005) *Effects of air pollution on children's health and development. A review of evidence.* European centre for Environment and Health – Bonn Office. Denmark: World Health Organization Europe, Publication E86575.

WHO (2006) *Health risk of particulate matter from long-range transboundary air pollution.* Joint WHO/Convention Task Force on the health aspects of air pollution. Denmark: World Health Organization Europe, Publication E88189.

Appendix A: Orthogonal regression

Derivation of the slope parameter in orthogonal regression with zero intercept

In this section a derivation for the slope coefficient and associated standard error is given. We will use the least squares principle and minimize the sum of the squared perpendicular distances d_i^2 from the data points (x_i, y_i) to the line $\hat{y}_i = a + b\hat{x}_i$. This is schematically shown in the figure below.



The horizontal distance from a point to the line is $h_i = x_i - \hat{x}_i$. The vertical distance is $v_i = y_i - \hat{y}_i$. According to Pythagoras' law $h_i^2 + v_i^2 = (p_i + q_i)^2$ and the sum of areas of the small triangles equals the area of the big triangle, so $d_i(p_i + q_i) = h_i v_i$. Combining these two equations and solving for d_i, we find:

$$d_i = \frac{h_i v_i}{\sqrt{h_i^2 + v_i^2}} \tag{13}$$

Using the above equations, this can be written in terms of x_i, y_i, a and b:

$$d_i = \frac{y_i - a - bx_i}{\sqrt{1 + b^2}} \tag{14}$$

As in regular linear regression, where the vertical distances, or residuals, v_i are assumed to be normally distributed with zero expectation and a common variance, we may also assume that the perpendicular distances d_i are normally distributed with zero expectation and a common variance σ^2. Using maximum likelihood theory, we can now derive orthogonal regression expressions for both a and b and their standard errors. Because we wish to derive an expression for the slope b only, we set $a = 0$. The likelihood function for one observation then becomes:

$$f(x_i, y_i \mid b, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - bx_i)^2}{2(1 + b^2)\sigma^2} \right) \tag{15}$$

The likelihood for *n* independent observations is the product *n* times the above expression. This function is maximized. The log is usually taken first, so the log-likelihood for *n* observations becomes:

$$\log[f(x,y \mid b,\sigma^2)] = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n}\frac{(y_i - bx_i)^2}{2(1+b^2)\sigma^2}, \qquad (16)$$

which can be written as:

$$\log[f(x,y \mid b,\sigma^2)] = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{Syy - 2bSxy + b^2Sxx}{2(1+b^2)\sigma^2} \qquad (17)$$

In the above equation, $Sxx = \sum_{i=1}^{n} x_i^2$, $Syy = \sum_{i=1}^{n} y_i^2$ and $Sxy = \sum_{i=1}^{n} x_i y_i$. In order to find the maximum, we calculate the partial derivatives of the above equation with respect to *b* and σ^2. These should be equal to zero. After a little simplification, we find:

$$\frac{\partial \log[f(x,y \mid b,\sigma^2)]}{\partial b} = \frac{Sxy - bSxx}{(1+b^2)\sigma^2} + b\frac{Syy - 2bSxy + b^2Sxx}{(1+b^2)^2\sigma^2} = 0, \qquad (18)$$

$$\frac{\partial \log[f(x,y \mid b,\sigma^2)]}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{Syy - 2bSxy + b^2Sxx}{2(1+b^2)\sigma^4} = 0. \qquad (19)$$

The first equation is solved for *b*. This appears to be a quadratic equation with two solutions. After a rearrangement of terms, the expression for the slope parameter we are interested in is:

$$b = \frac{Syy - Sxx + \sqrt{(Syy - Sxx)^2 + 4Sxy^2}}{2Sxy} \qquad (20)$$

The second equation is solved for σ^2. After a rearrangement of terms, the expression for the residual variance is:

$$\sigma^2 = \frac{Syy - 2bSxy + b^2Sxx}{n(1+b^2)}. \qquad (21)$$

The next step is the derivation of the standard error of *b*. In maximum likelihood theory, this is achieved by calculating the inverse Hessian minus the log-likelihood function. The variances of the parameters are then along the diagonal. However, since we are only interested in the standard error of *b*, we can substitute the expression for the residual variance minus the log-likelihood function and simply take the reciprocal of the second order derivative with respect to *b*. This is then equal to var(*b*). After substituting σ^2 in the log-likelihood function and some simplification, the minus log-likelihood function becomes:

$$-\log[f(x,y \mid b)] = \frac{n}{2}\left[1 + \log\left(\frac{2\pi}{n}\right) + \log(Syy - 2bSxy + b^2Sxx) - \log(1+b^2)\right] \qquad (22)$$

After some simplification the second order derivative becomes:

$$\frac{\partial^2 - \log[f(x,y \mid b)]}{\partial b^2} = n\left[\frac{(Syy + 2bSxy - b^2 Sxx)Sxx - 2Sxy^2}{(Syy - 2bSxy + b^2 Sxx)^2} + \frac{b^2 - 1}{(1+b^2)^2}\right] \quad (22)$$

Finally, var(b) is the reciprocal of the above expression:

$$\mathrm{var}(b) = \frac{1}{n}\left[\frac{(Syy + 2bSxy - b^2 Sxx)Sxx - 2Sxy^2}{(Syy - 2bSxy + b^2 Sxx)^2} + \frac{b^2 - 1}{(1+b^2)^2}\right]^{-1} \quad (23)$$

The standard error of b is the square root of the variance of b.

Comparison and validation

The standard error of the orthogonal fit forced through the origin is calculated for each of the candidate methods (CM) defined in this study with no statistical significant intercept as well as for three random individual sites and one small random dataset. The results for three different methods are presented in the table below. The first, u(b), is based on the formulae given in Appendix B of the equivalence guideline with the modification described in section 2.2, the second, $u(b)_{maxlike}$, is based on the maximum likelihood method as described in Appendix A, and the third is based on the bootstrap method (non-analytical). The results for all three are nearly identical. Only in the case of a very small number of samples may a minor difference occur (e.g. a few thousandths µg/m^3). The largest observed difference between the two *analytical* methods is approximately 0.0012 µg/m^3 (5%).

Based on these results it is concluded that all three methods provide fairly similar results, although minor differences of less than 0.001 may occur. These differences are negligible in the determination of the relative combined uncertainty within the scope of this equivalence study. Therefore, the least complicated method (modified version of the equations given by the EC Working Group) is used for this study, while still taking into consideration the usability of possible future adaptations in a revision of the Working Group's guideline.

Table A1 Regression coefficient standard uncertainties (forced trough origin) using different methods

Category	*n*	Sxx	Syy	Sxy	b	u(b)*	$u(b)_{maxlike}$**	$u(b)_{bootstrap}$***
FH62 IR Urban	463	546838	378801	451219	0.83	0.00507	0.00509	0.00626
FH62 IN New heating Urban	239	284798	169323	217588	0.77	0.00675	0.00676	0.00707
FH62 IN Old heating Urban	181	225082	165021	190411	0.85	0.00986	0.00991	0.011150
Random[1]	12	2907	1496	2069	0.72	0.02739	0.02632	0.027130
Site 131	247	239351	119139	166758	0.70	0.00709	0.00711	0.008927
Site 235	39	42500	24475	32122	0.76	0.01106	0.01093	0.011914
Site 934	64	39561	32637	35741	0.91	0.01180	0.01175	0.015355

[1] Small random dataset (n=12) with an average of 15 µg/m^3.
* Standard error calculated with the modified equations as demonstrated in section 2.2.
** Standard error calculated with the maximum likelihood method.
*** Standard error calculated with the bootstrap method.

Appendix B: LVS Filter comparability

Stastical evaluation of the Sven Leckel data.
This appendix describes the statistical evaluation of the Sven Leckel data collected between May 31st and June 27th of 2006. The statistical evaluation focuses on both the differences between multiple monitors as well as the type of filters used.

A predefined scheme was used to load different types of filters onto a total of eight samplers during the analysis period. Due to the initial set-up, the scheme for day 1 and 2 were repeated on days 13, 14, 27 and 28. To prevent the disproportionate existence of certain types of filters, the data collected on these days were omitted. Table B1 shows the PM_{10} concentrations from each sampler on a given day; the filter types are color-coded.

Table B1: PM_{10} concentrations from each sampler on different days. The four filter types (QF20, QMA, Pall and M&N are color coded to illustrate the used scheme.

Concentrations in µg/m³			QF20	QMA	Pall	M&N			
Day	L1	L5	L2	L4	L3	L7	L6	L8	Y_mean
D1	17.2	16.9	16.8	17.7	16.6	15.9	15.4	16.0	16.5
D2	17.0	16.3	14.2	15.1	14.7	14.8	16.3	18.1	15.8
D3	17.7	16.1	14.2	16.0	20.4	19.5	18.6	21.8	18.0
D4	22.8	21.7	25.4	26.9	28.6	27.3	24.9	26.6	25.5
D5	14.7	13.5	13.1	14.8	13.5	11.8	14.6	12.7	13.6
D6	13.8	12.3	10.6	11.8	11.2	10.8	12.5	14.3	12.1
D7	16.2	15.4	13.9	15.3	19.5	18.5	17.9	19.4	17.0
D8	27.1	25.8	30.0	31.8	32.8	32.5	29.9	32.3	30.3
D9	26.4	22.1	21.8	26.1	23.8	22.0	22.0	26.3	23.8
D10	29.3	23.9	21.1	27.5	27.0	24.8	23.5	29.2	25.8
D11	29.1	23.3	25.4	31.1	29.8	26.2	25.3	30.7	27.6
D12	27.2	24.7	22.6	29.1	28.5	25.3	22.1	26.6	25.8
D15	32.2	30.9	32.2	32.3	31.0	29.3	23.9	25.7	29.7
D16	23.3	22.1	21.2	21.1	18.2	17.1	22.7	23.8	21.2
D17	13.9	12.8	11.0	12.2	14.8	13.4	14.1	15.1	13.4
D18	14.7	13.5	13.8	15.8	16.4	15.2	14.0	17.7	15.1
D19	27.8	25.1	25.1	27.8	25.8	25.5	26.4	29.2	26.6
D20	32.5	30.1	27.8	27.9	28.2	27.7	30.5	32.5	29.6
D21	16.5	16.2	15.4	16.1	17.5	16.9	17.9	19.5	17.0
D22	15.0	15.0	18.8	19.4	20.1	19.8	19.1	20.2	18.4
D23	23.8	22.0	22.1	22.5	21.3	20.6	20.7	22.6	21.9
D24	20.4	19.7	17.6	17.7	17.6	16.8	19.5	20.6	18.7
D25	26.7	25.7	24.4	24.3	27.8	26.8	28.0	27.9	26.4
D26	27.2	27.3	34.8	35.0	37.2	36.1	34.6	36.5	33.6
Mean	22.2	20.5	20.5	22.3	22.6	21.4	21.4	23.5	21.8

The difference between samplers (range: 0–15%) was found to be based on the averaged results per sampler. The results for pairs L5 and L2, L7 and L6 and L3 and L4 are similar. However, the average for L8 is 4% higher then the second highest average (found on L3).

Calculation of the between-sampler variation.
The dataset from Table B1 is also used to estimate the between-sampler variation. Based on the results for each sampler pair with the same filter setup (L1 and L5; L2 and L4; L3 and L7; L6 and L8), the between-sampler uncertainty is estimated by:

$$u_{bs}^2 = \frac{\sum_{i=1}^{n}\left(y_{i,1} - y_{i,2}\right)^2}{2n} \qquad (24)$$

in which $y_{i,j}$ is the result of sampler *j* on day *i*, 1 and 2 denote the sampler pairs and *n* is the number of pairs (4 × 24)

The between-sampler uncertainty is ±1.63 µg/m^3 – or ±7.5% – with the averaged concentration of 21.8 µg/m^3. Another study shows results of <1.5 µg/m^3 for the between-sampler variation (Harrison et al., 2006)

Further exploration of the various subsets was considered not to be useful based upon the differences between the averages per sampler for each pair containing the same filter types.

Calculation of the filter-type effect
In order to calculate the effect of the different filter types, the absolute deviation of the daily averaged values $y_{i,k} - \overline{y_i}$ is calculated for each filter type; $y_{i,k}$ represents the result of filter-type *k* on day *i*.

The average of all absolute deviations, split per filter, is an indication of the effect of the different filter types on the measured PM_{10} concentration. Based on this calculation the average for the QF20 and QMA filters are +0.9 and +1.3 µg/m^3, respectively. The averages for the Pall and M&N filter types are –0.5 and –1.7 µg/m^3, respectively, compared to the averaged values over all filter types. The maximum difference between the four filters (QMA vs. M&N) is approximately 14%.

Analysis of variance (ANOVA)
For a more advanced data exploration, we fit a linear model to the PM_{10} concentration using the samplers and filter types as factors in the equation. The original dataset contains a large day-to-day effect, and these differences between days have been removed by normalizing the measurements with the daily average. This approach is only possible if the used filter set does not change over the days. The assumption is also made that there are small negligible differences between the eight samplers.

A two-way ANOVA is performed on the normalized dataset using MATLAB (R2006a) in combination with the Statistics toolbox. An ANOVA is used to test if there are differences between the population averages between two or more groups. Interactions between the factors samplers and filter types can also be observed.

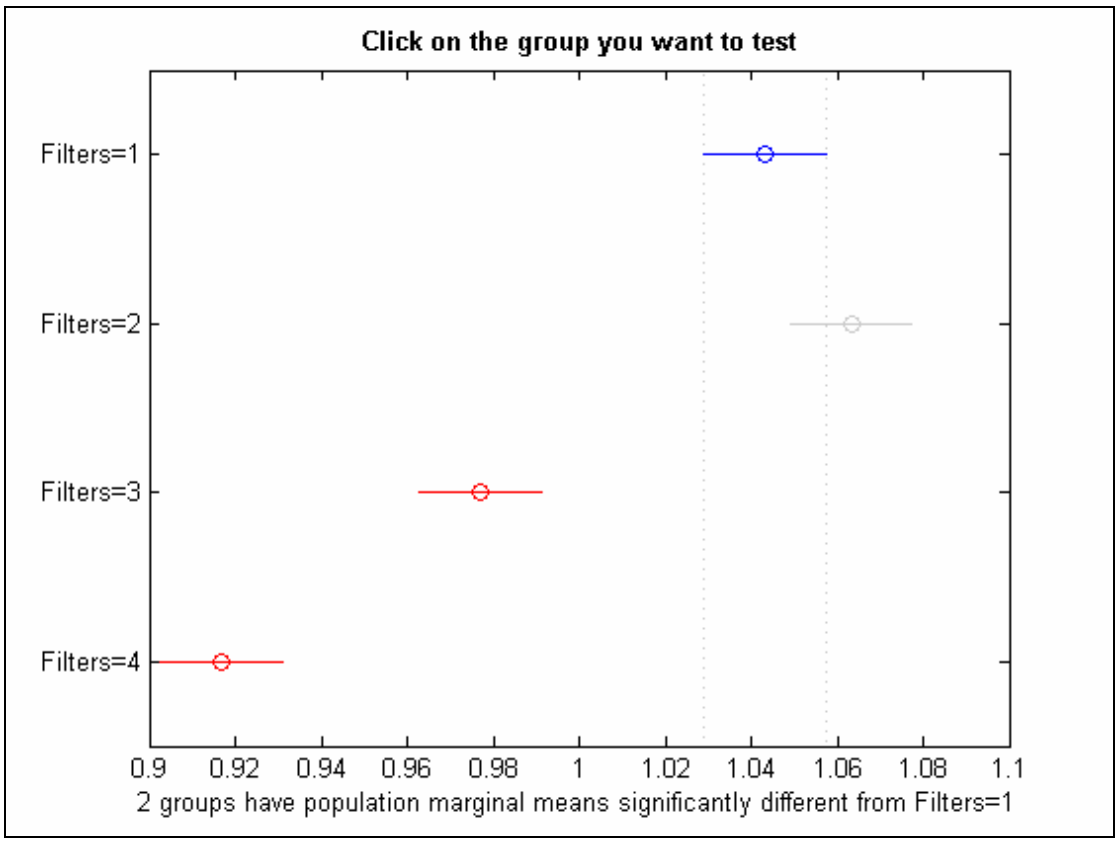


Figure B2: Population averages for the four filter types: filter 1: QF20, filter 2: QMA, filter 3: Pall, filter 4: M&N.

Based upon the calculated population averages, no difference can be found between the QF20 and QMA filter types. The averages for both the Pall and M&N filter types are different from those of the QF20 and QMA filters. There are also differences between the Pall and M&N filter, with the M&N filter giving the lowest results – approximately 15% lower than the QMA filter.

These results confirm the results based on the average of all absolute deviations.

Conclusions

1. No significant difference could be found between the QF20 and QMA filters. Both filters have been used in the National Air Quality Monitoring Network, although the QF20 filters are no longer in use. Based upon the population averages, QMA filters give a slightly higher result than the QF20 filters.
2. Both the M&N and Pall filter give a significantly lower concentration level than the QF20 and QMA filters.
3. The between-sampler variation u_{bs} is ± 1.63 µg/m^3.

Appendix C: LVS inside or outside an outdoor housing

Most of the gravimetric (RM) measurements are carried out with the low volume sampler (LVS) placed inside a temperature-controlled (20 °C) cabin. As the placement of the sampler inside a cabin is not possible for all sites in the NAQMN due to practical limitations, an experiment is carried out to determine possible deviating results for samplers placed outside of the unit. In this experiment, PM_{10} measurements are made at the same location but with samplers placed inside (A and B) and outside (C and D) of a cabin. A possible effect of outside placement of the LVS is to be expected on hot and sunny days. The experiment was carried out during two heat-waves in the Netherlands; as such, the prevailing ambient temperature during sampling leads to a for Dutch standards maximum temperature influence.

Daily average samples are collected for 19 days and, following validation, the samples are examined for possible differences due to placement of the LVS outside of the cabin. Based on the results of the measurements, a between-sampler uncertainty can be calculated using Eq. (26):

$$u_{bs}^2 = \frac{\sum_{i=1}^{n}(y_{i,1} - y_{i,2})^2}{2n} \qquad (25)$$

where $y_{i,1}$ and $y_{i,2}$ are the results of the parallel measurements, and n is the number of samplers.

The results show a between-sampler uncertainty u^2_{bs} of 0.57 and 1.04 µg/m^3 for LVS placed inside and outside a cabin, respectively. Therefore, the resulting average between sampler uncertainty u^2_{bs} is 0.81 µg/m^3 for all samplers together. The average measurement difference between samplers placed concurrently inside and outside of a cabin is 0.06 µg/m^3 with a variance of 0.68 µg/m^3 (see also Table 11. Based on these results it can be concluded that the placement of the LVS inside or outside a cabin does not lead to a significant difference in measurement results nor to an unacceptable (u_{bs} >2 µg/m^3) between-sampler uncertainty.

Conclusion
No statistically significant difference is observed between gravimetric PM_{10} measurements carried out inside and outside an air-conditioned cabin, although the between-sampler uncertainty is slightly larger for a measurement carried out in the open air. The resulting average between-sampler uncertainty u_{bs} – irrespective of air conditioning (inside/outside a cabin) – is approximately 0.90 µg/m^3 (√ 0.81).

Table 11 Results of four parallel gravimetric LVS PM_{10} measurements, with two samplers (C and D) placed outside in the open air and two samplers placed in a temperature-controlled (20 °C) cabin (A and B).

Day	Date	Inside LVS µg/m³			Outside LVS µg/m³			Avg inside – Avg outside	Σ of squared differences
		A	B	Average	C	D	Average	AVG	(A;B;C;D)
D1	31-8-2006	26.99	27.27	27.13	27.45	26.62	27.04	0.1	0.39
D2	1-9-2006	27.37	27.98	27.68	28.24	27.48	27.86	-0.19	0.51
D3	2-9-2006	31.76	32.27	32.02	32.70	31.56	32.13	-0.11	0.79
D4	3-9-2006	16.81	17.22	17.02	17.61	16.27	16.94	0.08	0.99
D5	4-9-2006	15.72	15.47	15.60	16.74	14.15	15.44	0.15	3.41
D6	5-9-2006	27.64	27.27	27.46	28.66	27.12	27.89	-0.43	1.44
D7	6-9-2006	26.95	25.24	26.09	27.66	24.18	25.92	0.17	7.55
D8	7-9-2006	22.84	22.19	22.51	23.77	22.76	23.27	-0.75	1.28
D9	8-9-2006	15.52	15.58	15.55	18.11	17.28	17.69	-2.15	4.95
D10	9-9-2006	16.40	16.53	16.47	17.27	16.76	17.02	-0.55	0.44
D11	10-9-2006	17.91	18.32	18.12	18.16	17.25	17.70	0.41	0.67
D12	11-9-2006	30.93	31.91	31.42	31.80	30.63	31.21	0.21	1.21
D13	12-9-2006	43.17	44.78	43.98	43.70	42.38	43.04	0.93	3.04
D14	13-9-2006	41.30	42.12	41.71	41.28	39.36	40.32	1.39	4.11
D15	14-9-2006	31.10	34.40	32.75	34.14	34.41	34.28	-1.53	7.81
D16	15-9-2006	39.97	40.14	40.06	39.08	40.27	39.67	0.38	0.87
D17	16-9-2006	52.36	53.00	52.68	51.05	52.34	51.69	0.98	2.01
D18	17-9-2006	65.53	66.09	65.81	66.14	65.26	65.70	0.11	0.56
D19	18-9-2006	24.57	23.45	24.01	24.66	23.98	24.32	-0.31	0.95
Average		**30.26**	**30.59**	**30.42**	**30.96**	**30.00**	**30.48**	**-0.06 (var=0.7)**	**2.26 (u_{bs}=0.9)**

Appendix D: FH62 I-R between-sampler uncertainty

The between-sampler uncertainty is examined for the FH62 I-R method at rural locations on two different occasions. A comparison is carried out in which four FH62 I-R samplers are operated in parallel for 19 days during a heat-wave. The measurement results are acquired from the device directly using the digital interface. Based on the results of this experiment, a between-sampler uncertainty is calculated using Eq. (27). A comparison between two FH62 I-R samplers is also carried out at one of the monitoring sites in the NAQMN for a approximately 1 year. The measurement results are acquired using the standard procedures of the NAQMN, including a digital/analogue conversion, for a southern rural location. The results of this comparison are used to calculate the between-sampler uncertainty.

$$u_{bs}^2 = \frac{\sum_{i=1}^{n}(y_{i,s} - \bar{y}_i)^2}{2n} \tag{26}$$

where $y_{i,s}$ is the measurement result of each parallel sampler s on day i, and n is the number of samplers.

Result and conclusion

The results for the four parallel measurements are presented together with the calculated variance in Table 12. The resulting between-sampler u_{bs} based on 19 quadruple sample-pairs, is 0.62 µg/m^3 (u^2_{bs} = 0.38 µg/m^3), and that in the second comparison, based on 318 sample-pairs, is 2.54 µg/m^3. The result of the second comparison is considerably higher than that in the first comparison, and this difference is likely to be associated with the method of acquisition (direct digital output or output after analogue conversion) and the length and period of sampling. Nonetheless, both between-sampler uncertainties lie within the limit prescribed in the guideline of the EC Working Group on Equivalence (u_{bs} <3 µg/m^3). The NAQMN is currently being brought up to date, and part of the modernization includes the transition to purely digital data acquisition. This will lead to a decrease in the between-sampler uncertainty.

Table 12 Results of four parallel FH62-IR PM_{10}-measurements.

Day	LVS µg/m^3	FH62 I-R A (µg/m^3)	FH62 I-R B (µg/m^3)	FH62 I-R C (µg/m^3)	FH62 I-R D (µg/m^3)	Sum of squared differences (µg/m^3)
19-7-2006	31.23	19.39	19.78	17.61	18.94	2.68
18-7-2006	24.67	18.09	16.91	17.02	17.22	0.86
17-7-2006	21.15	14.85	15.04	13.98	14.13	0.82
16-7-2006	17.44	11.37	11.30	10.20	10.46	1.05
15-7-2006	20.82	14.96	14.15	13.54	13.80	1.15
14-7-2006	21.16	16.61	16.22	16.09	15.98	0.23
13-7-2006	22.01	14.22	13.91	13.13	13.37	0.74
12-7-2006	18.46	11.46	11.09	9.57	10.50	2.04
11-7-2006	n/a	12.93	13.68	13.32	13.02	0.34
10-7-2006	n/a	17.85	17.58	19.54	17.71	2.54
9-7-2006	n/a	15.79	15.79	15.69	14.98	0.46
8-7-2006	n/a	12.10	12.27	12.50	12.27	0.08
7-7-2006	n/a	23.85	23.94	24.40	22.96	1.09
6-7-2006	n/a	20.50	20.63	19.73	20.56	0.53
5-7-2006	n/a	23.98	22.67	24.48	24.27	1.98
4-7-2006	n/a	21.00	20.46	22.25	20.88	1.78
3-7-2006	n/a	18.85	19.10	19.40	19.17	0.15
2-7-2006	n/a	17.02	16.54	16.88	17.10	0.18
1-7-2006	n/a	18.04	17.19	15.85	17.85	2.95
Average	**22.12**	**16.99**	**16.75**	**16.59**	**16.59**	**1.14**

RIVM

National Institute
for Public Health
and the Environment

P.O. Box 1
3720 BA Bilthoven
The Netherlands
www.rivm.com