Atmospheric Environment 273 (2022) 118967

Contents lists available at ScienceDirect

Atmospheric Environment

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

Modified Target Diagram to check compliance of low-cost sensors with the Data Quality Objectives of the European air quality directive

Sinan Yatkin [a,*], Michel Gerboles [a], Annette Borowiak [a], Silvije Davila [b], Laurent Spinelle [c,d], Alena Bartonova [e], Frank Dauge [e], Philipp Schneider [e], Martine Van Poppel [f], Jan Peters [f], Christina Matheeussen [g], Marco Signorini [h]

[a] European Commission, Joint Research Centre (JRC), Via Enrico Fermi 2749, 21027, Ispra, VA, Italy
[b] Institute for Medical Research and Occupational Health, Ksaverska cesta 2, Zagreb, Croatia
[c] Institut national de l'Environnement Industriel et des Risques (INERIS), Parc technologique Alata, BP 2, F-60550, Verneuil-en-Halatte, France
[d] Laboratoire Central de Surveillance de la Qualité de l'Air (LCSQA), Parc technologique Alata, BP 2, F-60550, Verneuil-en-Halatte, France
[e] NILU – Norwegian Institute for Air Research, PO Box 100, 2027, Kjeller, Norway
[f] Flemish Institute for Technological Research (VITO), Boeretang 200, 2400, Mol, Belgium
[g] Flanders Environment Agency, Dokter De Moorstraat 24-26, 9300, Aalst, Belgium
[h] Liberaintentio Srl, Malnate, 21046, Italy

HIGHLIGHTS

- A graphical tool to evaluate performances of sensors was developed.
- Uncertainty of sensors and its' contributors can be displayed.
- Checking compliance of sensors with legislation is easy and fast.
- Several other information can be extracted.

ARTICLE INFO

Keywords:
Target diagram
Measurement uncertainty
Low-cost sensors

ABSTRACT

The modified Target Diagram (MTD) was developed to evaluate the performance of low-cost sensors (LCS) for air quality monitoring in comparison with reference methods by reporting relative expanded uncertainty and its contributors. An MTD provides several pieces of information, including compliance with regulation, sources of error and how to diminish them, completeness and validity of LCS calibration etc. It allows the user to examine the effect of selecting different regression types and residual fitting on the LCS measurement uncertainty. The ordinary least squared regression with fitted residuals and dynamic between reference analyser uncertainty rather than constant ones yielded more realistic LCS measurement uncertainty compared to other options. The MTD is a fast visual tool to extract several pieces of information on evaluation of any candidate method against reference method.

1. Introduction

Air Quality Monitoring Stations (AQMS) implement the reference measurement methods for the regulated pollutants defined in the European Air Quality Directive (AQD) to provide the best estimation of true air pollution levels (EC, 2008). In the last decade, the Taylor Diagram (Taylor, 2001) and the Target Diagram (Jolliff et al., 2009) have been widely used to evaluate the agreement between air quality models and reference measurements of AQMS (Cuvelier et al., 2007; Thunis et al., 2012; Vautard et al., 2007). The Taylor Diagram is a composite tool visualising of several statistics that summarise agreement between two datasets, e.g., correlation, root-mean-squared error (RMSE), and ratio of corresponding variances (Taylor, 2001). Jollif et al. (2009) developed the Target Diagram, which displays RMSE with an orthogonal decomposition of bias and centred RMSE to quantify their corresponding contributions to RMSE. The major advantage of the Target Diagram over the Taylor Diagram is that it takes bias between two datasets into

* Corresponding author.
E-mail addresses: sinan.yatkin@ec.europa.eu (S. Yatkin), michel.gerboles@ec.europa.eu (M. Gerboles).

https://doi.org/10.1016/j.atmosenv.2022.118967
Received 18 December 2021; Received in revised form 17 January 2022; Accepted 19 January 2022
Available online 22 January 2022

Abbreviations

AQD	Air Quality Directives
AQMS	Air quality monitoring stations
AT	Alert threshold
CL	Critical level
DQO	Data quality objectives
DR	Deming regression
GAM	General additive models
GDE	Guide to demonstrate the equivalence
MPC	Model Performance Criteria
MTD	Modified Target Diagram
LAT	Lower assessment threshold
LCS	low-cost sensors
LV	Limit value
OLS	Ordinary least squares regression
OR	Orthogonal regression
RB	Relative bias
RMSE	Root-mean-squared-error
RSi	Squared-residual
RSS	Sum-squared of residuals
RR	Relative residuals
u(bs,RM)	Between reference method standard uncertainty
u(bs,s)	Between sensors standard uncertainty
$U_R(Y_i)$	Relative expanded uncertainty of Y_i (LCS results for a testing duration of i)

account.

As a further version, the FairMode (Forum for air quality modelling in Europe) initiative developed the Model Performance Criteria (MPC) to check whether paired differences between model outputs and reference measurements remain within corresponding uncertainties (Pernigotti et al., 2013; Thunis et al., 2012, 2013). FairMode makes the Delta Tool available to display Target Diagrams whether or not the MPC are satisfied (https://aqm.jrc.ec.europa.eu/index.aspx).

Although both diagrams are suited for evaluating agreement between model outputs and reference measurements, none of them are appropriate to evaluate whether the Data Quality Objectives (DQO) set in the AQD are met. The AQD states that reference, equivalent to reference and indicative measurements as well as model outputs shall meet the DQO, which are the maximum relative expanded uncertainties of regulated pollutant measurements not to be exceeded at the corresponding Limit Values (LV). The AQD also states that the methodology described in the Guide for the Demonstration of Equivalence of Measurement Methods (GDE) shall be followed to show the fulfilment of DQO (European Commission, 2010). This method estimates measurement uncertainty using a quadratic summation of a) the residuals sum of squares (RSS) of orthogonal regression and b) the bias between this orthogonal regression line and the perfect line of agreement (slope of 1 and intercept of 0) at each pollutant level.

The aforementioned methodologies, namely the GDE and the Target Diagram, suggest different approaches and thus metrics to check compliance of measurements/modelling with the DQO, either directly or indirectly. These approaches can be merged in a modified Target Diagram, where RMSE is replaced with measurement uncertainty calculated in line with the GDE method, thus adding some slight changes. Therefore, the GDE and the Target Diagram can be combined to display a variety of useful information, e.g., meeting of the DQO, over or under-estimation, lack of sensitivity by showing the measurement uncertainty as the distance of each data to origin of the Target Diagram etc.

The Modified Target Diagram (MTD) proposed in this study has been particularly designed for evaluating of the performance of low-cost sensors (LCS), which have been receiving an increasing interest not only from researchers but also from public communities (Kumar et al., 2015; Lewis et al., 2018; Rai et al., 2017). From a European legislative point of view, the LCS would make air pollution monitoring possible at much lower cost compared to reference methods. The LCS are identified as emerging devices for indicative measurements or objective estimations as defined in the AQD (Aleixandre and Gerboles, 2012; Karagulian et al., 2019; Spinelle et al., 2017). However, their performances are often questionable (Lewis and Edwards, 2016; Spinelle et al., 2015) and protocols for evaluation of LCS for air pollution are being developed to be able to gather comparable harmonized metrics about LCS data quality (ASTM, 2018, 64899; CEN, 2021; Collier-Oxandale et al., 2020; Papapostolou et al., 2017; US EPA, 2021a; US-EPA, 2021b).

The MTD can be used to compare any measurement time series to co-located reference data giving a harmonized indicator. Applying MTD to any past or future studies with provided time series is easy and does not require additional information except for brand of reference analyser. Using the MTD for the performance evaluation of both dispersion model output and the LCS would make implementing the same assessment method in line with the AQD requirements possible.

This manuscript presents the equations required for computing the MTD, a guidance for extracting information from the MTD and several examples of MTDs for LCS studies.

2. Theoretical basis

2.1. Modified Target Diagram

When the LCS are collocated with the AQMS, a linear relationship between two datasets is expected as shown in Eq. (1). In case of a non-linear relationship between the LCS and reference data sets, the proposed methodology in this paper cannot be implemented due to a non-linear bias between two sets.

$$Y_i = b_0 + b_1 X_i \quad (1)$$

where

- Y_i is the LCS final data series. Generally, the raw LCS data are transformed to the final LCS data using a known or unknown LCS calibration model. Hereafter, the LCS data refers to final data in the same unit of reference data, unless otherwise indicated;
- X_i is reference data series, which is the best estimation of true values;
- b_0 and b_1 are respectively the intercept and slope of the regression line of Y_i against X_i.

The expanded measurement uncertainty of the LCS, $U(Y_i)$, is estimated using Eq. (2) according to the GDE and elsewhere (CEN, 2021):

$$U(Y_i) = k\left(\frac{RSS}{n-2} - u^2(bs,\ RM) + u^2(bs,s) + [b_0 + (b_1 - 1)X_i]^2\right)^{½} \quad (2)$$

where

- k is the coverage factor accounting for the probability distribution of uncertainty of a measurement with a selected probability level. The corresponding standard (CEN, 2021) imposes k to be 2 for a large number of experimental results available that gives approximately the 95% confidence interval for a normal distribution (1.96 at 95% confidence level).
- RSS is the sum of the squared residuals, computed using Eq. (3), when the RSS is constant over the X_i range:

Table 1
Level of upper or lower assessment thresholds or the alert or critical levels and data quality objectives for indicative measurements, objective estimation and class 3 set in the Air Quality Directive and in TS 17660.

	Averaging time	LV	UAT	LAT	AT or CL	DQO of indicative measurements	DQO of objective estimations	DQO of Class 3 of TS 17660
		μg/m^3	μg/m^3	μg/m^3	μg/m^3	%	%	%
Sulfur dioxide, SO_2	1 h	350			500	25	75	200
SO_2	24 h	125	75	50	20	25	75	200
Nitrogen dioxide, NO_2	1 h	200	140	100	400	25	75	200
NO_2	1 year	40	32	26		25	75	200
Ozone, O_3	8 h	120	84	60		30	75	200
O_3	1 h				240			
Benzene	1 year	5	3.5	2		30	100	200
Carbon monoxide, CO	8 h	10[a]	7[a]	5[a]		25	75	200
PM_{10}	24 h/1 h	50	35	25		50	100	200
$PM_{2.5}$	24 h/1 h	25	17.5	12.5		50	100	200

Note 1: LV is the limit or target value, UAT is the upper assessment threshold, LAT is the lower assessment threshold, AT is the alert threshold, CL is the critical levels and DQO is the data quality objective. Please note that some LVs might change after finalizing the ongoing review of the Air Quality Directive and the update of LVs in the new World Health Organization guidelines (WHO, 2021).
Note 2: For O_3, an averaging period of 1 h instead of 8 h shall be used for the evaluation of LCS at the LV.
Note 3: For benzene, an averaging period of 1 h instead of 1 year shall be used for the evaluation of LCS at the LV.
Note 4: For CO, an averaging period of 1 h instead of 8 h shall be used for the evaluation of LCS at the LV.
Note 5: For PM_{10} and $PM_{2.5}$, technical standard regarding the testing of LCS is yet to be approved, and it is likely that an averaging period of 1 h shall be set in addition to 24 h for the evaluation of LCS at the LV.
[a] In mg/m^3.

$$\mathrm{RSS} = \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2 \quad (3)$$

- n is the number of data pairs;
- u(bs,s) is the between LCS standard uncertainty;
- u(bs,RM) is the between reference method standard uncertainty.

u(bs,s) is optionally added to Eq. (2) when performance is evaluated for multiple sensors of the same brand. Usually, a subset of LCS of the same brand is tested at the AQMS for calibration and/or performance evaluation purposes. Thus, u(bs,s) can be computed using Eq. (4):

$$u(bs,s) = \sqrt{\left(\frac{\sum_{l=1}^{N}\sum_{j=1}^{p}\left(Y_{l,j} - \overline{Y}_l\right)^2}{N(p-1)}\right)} \quad (4)$$

where:

- $Y_{l,j}$ is the measurement of LCS j for period of l;
- $\overline{Y}_l$ is the mean result for period l from p collocated LCS;
- N is the number of measurements over time when u(bs,s) is estimated;
- p is the number of collocated LCS.

The RSS includes the random uncertainty of reference measurements. u(bs,RM) shall be subtracted from the RSS, thanks to (RSS/(n-2))$^{1/2}$ and u(bs,RM) being independent and therefore their variances are additive. When multiple reference instruments are deployed in field tests, the between reference method random uncertainty, u(bs,RM), should be calculated using Eq. (4). In this case, $Y_{l,j}$ corresponds to the reference data for period l of reference instrument j. When only a single reference instrument is deployed, previously calculated u(bs,RM) shall be used. These values can be found for many reference analysers elsewhere (CEN, 2021; "QAL1 - Certified measuring- and evaluating-systems," available at: https://qal1.de/en/hersteller/thermo.htm).

Eq. (5) is derived from Eq. (2) to compute the relative expanded measurement uncertainty of LCS, $U_R(Y_i)$, as the ratio of $U(Y_i)$ to X_i. $U_R(Y_i)$ is the square root of the quadratic summation of:

- RR, see Eq. (6), the relative residuals of the regression line corrected for u(bs, RM) and optionally for u(bs,s);
- RB, see Eq. (7), the relative bias between the regression line and the perfect line of agreement with slope of 1 and intercept of 0.

$$U_R(Y_i) = \frac{U(Y_i)}{X_1} = k\left\{\left[\frac{\left(\frac{\mathrm{RSS}}{n-2} - u^2(bs,RM) + u^2(bs,s)\right)^{\frac{1}{2}}}{X_i}\right]^2 + \left[\frac{b_0 + (b_1 - 1)X_i}{X_i}\right]^2\right\}^{\frac{1}{2}} \quad (5)$$

$$RR = k\left[\frac{\frac{\mathrm{RSS}}{(n-2)}}{X_i} + \frac{-u^2(bs,\ RM) + u^2(bs,s)}{X_i}\right]^{1/2} \quad (6)$$

$$RB = k\left[\frac{b_0}{X_i} + (b_1 - 1)\right] \quad (7)$$

In Eq. (5), since X_i is the best estimate of the true value, it is more rational to normalize $U_R(Y_i)$ to X_i rather than Y_i since bias in Y_i could distort $U_R(Y_i)$. In addition, by normalising $U_R(Y_i)$ to X_i, it is possible to check whether the DQO are met at the true air pollutant levels including the LV, upper/lower assessment thresholds (UAT and LAT), and alert/critical levels (AT/CL) (see Table 1).

In Eq. (5), $(b_1\text{-}1)X_i + b_0$ gives the distance between the regression line of Eq. (1) and the perfect agreement line with slope of 1 and intercept of 0 while random errors are represented by the remaining terms, namely the RSS, u(bs,s) and u(bs,RM). In Eq. (6), RSS/(n-2) - u^2(bs,RM) shall not be negative, otherwise, u(bs,RM) is likely to be overestimated. Moreover, a contribution of u(bs,s) is optionally added to Eq. (6) to account for the variability in LCS units. If variability of LCS is not tested and evaluated, u(bs,s) shall be discarded. In that case, $U_R(Y_i)$ and RR will only apply to individual LCS and cannot be generalised to other LCS of the same brand.

Eq. (5) allows to plot a MTD with RB on y-axis and RR on x-axis. $U_R(Y_i)$ is represented by the hypotenuse of the triangle RB, RR, and $U_R(Y_i)$. $U_R(Y_i)$ can be directly compared to the DQO that are shown with bold contour lines centred on the origin (0,0), here after called target

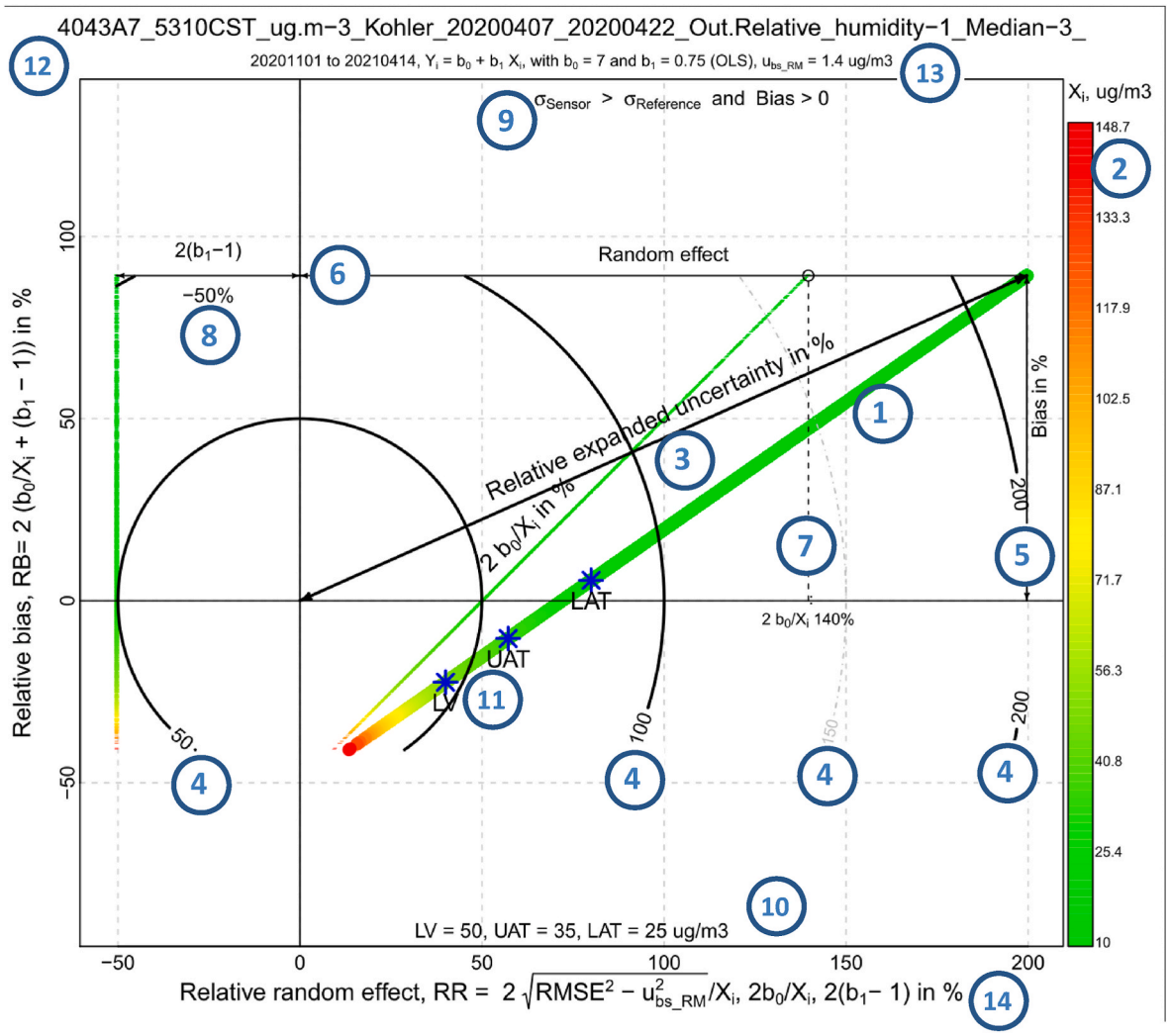

Fig. 1. Example of MTD for a PM_{10} sensor.

circles. Eq. (6) and Eq. (7) show that the RR and RB depend on X_i, while (b_1-1) is the only constant value. The RR and b_0/X_i decrease with increasing X_i, indicating that the relative expanded uncertainty given by the DQO might be exceeded at lower X_i whereas it is possible to be met at the LV, UAT and LAT, and AT/CL. Applying a coloured X_i scale at the MTD is an effective visualisation solution to display this complexity, clearly identifying $U_R(Y_i)$ at corresponding X_i. This illustration helps to evaluate whether the DQO are met at X_i approaching to the regulated pollutant levels.

2.2. Regression types for comparison of data sets

The choice of regression type to compare two data sets, e.g., orthogonal regression (OR) or Ordinary Least Squares Regression (OLS), may have a significant effect on slope and intercept and thus on $U_R(Y_i)$ computed using Eq. (5). The major difference of OR compared to the OLS is that it accounts for perpendicular distances from data pairs of candidate and reference to the regression line. A generalization of OR, so-called the Deming regression (DR, Deming, 1943), aims at minimizing a weighted RSS (wRSS), both along x-axis and y-axis (Eq. (8)). The DR has only an analytical solution if δ, the ratio of the variances of y over x, is known (Eq. (9)). Eq. (10) and Eq. (11) give respectively slope, b_1, and intercept, b_0, with intermediary calculations given in Eq. (12) to Eq. (14) of a regression line assuming both x and y are measured with non-negligible errors. With equal error variances of y and x thus $\delta = 1$, The DR becomes an OR. Since u(bs,RM) can be found for almost all reference analysers available in the market, the DR seems to be applicable by computing u(bs,s) employing multiple LCS in field experiment. However, the direct inspection of approval test reports reveals that some of the reported u(bs,RM) are evaluated questionable along X_i range, since these tests were performed usually at a single pollutant level, which is usually considerably higher than typical ambient air levels, thus overestimated for low X_i. u(bs,RM) should ideally decrease with decreasing X_i. Consequently, augmented u(bs,RM) leads to an underestimated δ and thus erroneous slope and intercept of the DR line.

The GDE imposes to fit an OR to demonstrate equivalence of a candidate method to a reference method, both being expected to show similar between instrument uncertainty. However, in case of LCS being a candidate method, u(bs,s) is expected to be higher than u(bs,RM), as reference methods have the highest metrological quality for pollution monitoring. Therefore, the OR cannot be used and instead the OLS aiming at minimizing errors only parallel to *y*-axis (LCS) is better suited and advised for evaluating the LCS meeting the DQO, since the error on the x-axis (Reference) can be assumed to be much smaller than the ones on the y-axis (Eq. (3)). For the OLS, the intercept and slope of line are computed using Eq. (11) and Eq. (15), respectively.

$$\mathrm{wRSS} = \sum_{i=1}^{n} \frac{(Y_i - (b_1 X_i + b_0))^2}{u(bs,s)^2} + \frac{\left(X_i - Y_i - b_0/b_1\right)^2}{u(bs,RM)^2} \quad 8$$

$$\delta = \left(\frac{u(bs,s)}{u(bs,RM)}\right)^2 \quad 9$$

$$b_1 = \frac{S_{yy} - \delta\, S_{xx} + \sqrt{\left(S_{yy} - \delta\, S_{xx}\right)^2 + 4\,\delta\, S_{xy}^2}}{2S_{xy}} \quad 10$$

$$b_0 = \overline{Y} - b\overline{X} \quad 11$$

$$S_{xx} = \frac{1}{n-1}\sum (\mathrm{X} - \overline{X})^2 \quad 12$$

$$S_{yy} = \frac{1}{n-1}\sum (Y_i - \overline{Y})^2 \quad 13$$

$$S_{xy} = \frac{1}{n-1}\sum (X_i - \overline{X})\,(Y_i - \overline{Y}) \quad 14$$

$$b_1 = \frac{\sum[(X_i - \overline{X})\,(\mathrm{Y} - \overline{Y})]}{\sum (X_i - \overline{X})^2} \quad 15$$

where $\overline{Y}$ and $\overline{X}$ are the mean LCS and reference data, respectively.

The MTD can be plotted selecting several fitting options including the OLS (advised for the LCS, the method prescribed in CEN, 2021), the OR as stated in the GDE and the DR.

2.3. Fitting sum squared of residuals

In Eq. (3), the residuals, $(Y_i - (b_1 X_i + b_0))$, are assumed to be approximatively constant over X_i range to attribute an X_i-regardless value to the RSS. Eq. (3) is valid only if the variance of residuals remains constant over X_i range, so-called homogeneity of variance or homoscedasticity. Conversely, the situation of the variance of residuals varying over X_i is called heterogeneity of variance or heteroscedasticity (Zuur et al., 2010). Heteroscedasticity can be detected using the Breusch Pagan test (Breusch and Pagan, 1979).

In case of heteroscedasticity, using the RSS in Eq. (5) is not possible since the squares of residuals, RS_i, are not constant over X_i range. However, a simple approach to account for possible heteroscedasticity consist of fitting a General Additive Model (GAM) (Hastie and Tibshirani, 1987, 1990) describing the relationship between RS_i and X_i, and thus the modifications of Eq. (5) and Eq. (6) respectively to Eq. (16) and Eq. (17) for computing $U_R(Y_i)$ and the RR.

$$U_R(Y_i) = \frac{U(Y_i)}{X_1} = \mathrm{k}\left\{\left[\frac{(RS_i - u^2(bs,RM) + u^2(bs,s))^{\frac{1}{2}}}{X_i}\right]^2 + \left[\frac{b_0 + (b_1 - 1)X_i}{X_i}\right]^2\right\}^{½} \quad 16$$

$$RR = k\left[\frac{RS_i}{X_i} + \frac{-u^2(bs,\ RM) + u^2(bs,s)}{X_i}\right]^{½} \quad 17$$

The Breusch Pagan test should be initially applied to determine whether the variance of residuals is not constant over X_i range. In such case, a GAM should be fitted to account for the variation of RS_i versus X_i, and $U_R(Y_i)$ and RR should be computed using Eq. (16) and Eq. (17) instead of Eq. (5) and Eq. (6). In case of fitting RS_i with a GAM, the effect of random errors included in the RR on $U_R(Y_i)$ would be better estimated and represented.

3. Results

3.1. Guidance of reading MTD

A wide range of different information can be extracted from the MTD. Using the example given in Fig. 1, the following guidance for reading the MTD is given where the numbered list refers to blue numbers in Fig. 1:

1. Firstly, identify the bold line in the plot with a coloured scale indicating the ascendingly sorted X_i from green to red.
2. The colour scale indicates the X_i range from light green to red, corresponding to 10.0 and 148.7 μg/m^3 in the example of Fig. 1.
3. The distance between any point of the bold coloured line and the origin indicates $U_R(Y_i)$ computed using Eq. (5). In Fig. 1, the smallest X_i of 10.0 μg/m^3 has a corresponding $U_R(Y_i)$ of 215%.
4. Each $U_R(Y_i)$ can be visually estimated by interpolation of the target circles shown in bold solid lines (50, 100 and 200% in Fig. 1) indicating the DQO or dotted lines (150% in Fig. 1).
5. $U_R(Y_i)$ corresponds to the hypotenuse of triangle made up of two contributors: the first one is from the RR (Relative random effect) describing random error of LCS data computed using Eq. (6). In Fig. 1, the smallest X_i has a corresponding RR of 200%.
6. And the second is one from the RB (Relative bias) between the LCS and reference data computed using Eq. (7). In Fig. 1, the smallest X_i has a corresponding RB of 90%.
7. The RB is divided into two contributors to $U_R(Y_i)$: the first one is from the intercept of the regression line, which equals to $2(b_0/X_i)$. This contribution is indicated by an oblique solid coloured line respecting the coloured scale of X_i (see point 2). $2(b_0/X_i)$ values are read on the x-axis.
8. And the second one is a constant contribution to $U_R(Y_i)$ from the slope of the regression line, which equals to $2(b_1\text{-}1)$. This contribution is indicated by a vertical solid coloured line using the coloured scale of X_i. $2(b_1\text{-}1)$ values are read on the x-axis. In Fig. 1, at the smallest X_i, the RB is 90%, as the sum of b_0 effect (140%) and b_1 effect (−50%).
9. Comparison of standard deviations of the LCS and reference data and bias being negative or positive, which are also identified by the location of the bold coloured line. 9.1) If the standard deviation of LCS data is lower than the one of reference data, the bold coloured line locates at the left of the y-axis with values on the x-axis increasing from right to left, otherwise, the bold coloured line stands at the right of the y-axis. 9.2) When the RB is positive or negative, $U_R(Y_i)$ are plotted in the above or below the x-axis, respectively. This topic is further discussed in Section 4.1. The line of Fig. 1 being at the right of the y-axis shows that the LCS has higher sensitivity than the reference data. The bold coloured line crossed the x-axis at about 30 μg/m^3, thus the line for $X_i < 30$ μg/m^3 being above the x-axis shows that the LCS overestimated the reference data. Conversely, for $X_i > 30$ μg/m^3, the bold coloured line being below the x-axis shows that the LCS underestimated the reference data.
10. The current AT, LV, UAT and LAT set in the AQD are displayed when covered by the X_i range (see Table 1).
11. All the legislative levels covered by the X_i range are marked on the bold coloured line using a blue asterisk with an indicative label. In Fig. 1, the LV of 50 μg/m^3, the UAT of 35 μg/m^3 and LAT of 25 μg/m^3 are covered by the range of X_i between 10 and 149 μg/m^3.
12. In the example of Fig. 1, the title is "4043A7_5310CST_ug.m-3_Kohler_20200407_20200422_Out.Relative_humidity-1_Median-3_". Although there are no requirements, the title should identify the brand of sensor, e. g. 4043A7_5310CST. In Fig. 1, the title also indicates the unit of the raw sensor (μg/m^3), the type of calibration model (Kohler model), the time interval of the calibration datasets (between 2020-04-07 and 2020-04-22), the covariate used in the calibration model (relative humidity), and the method to determine the coefficients of the model (Median-3 refers to the median of all models set for 3 days interval within the calibration period).
13. The 2nd line of the title gives the interval time of prediction (between 2020-11-01 and 2021-04-14), b_0 and b_1 the intercept and slope of the regression line of LCS data versus reference data in Eq. (1) with the type of regression being fitted (the OLS, the OR

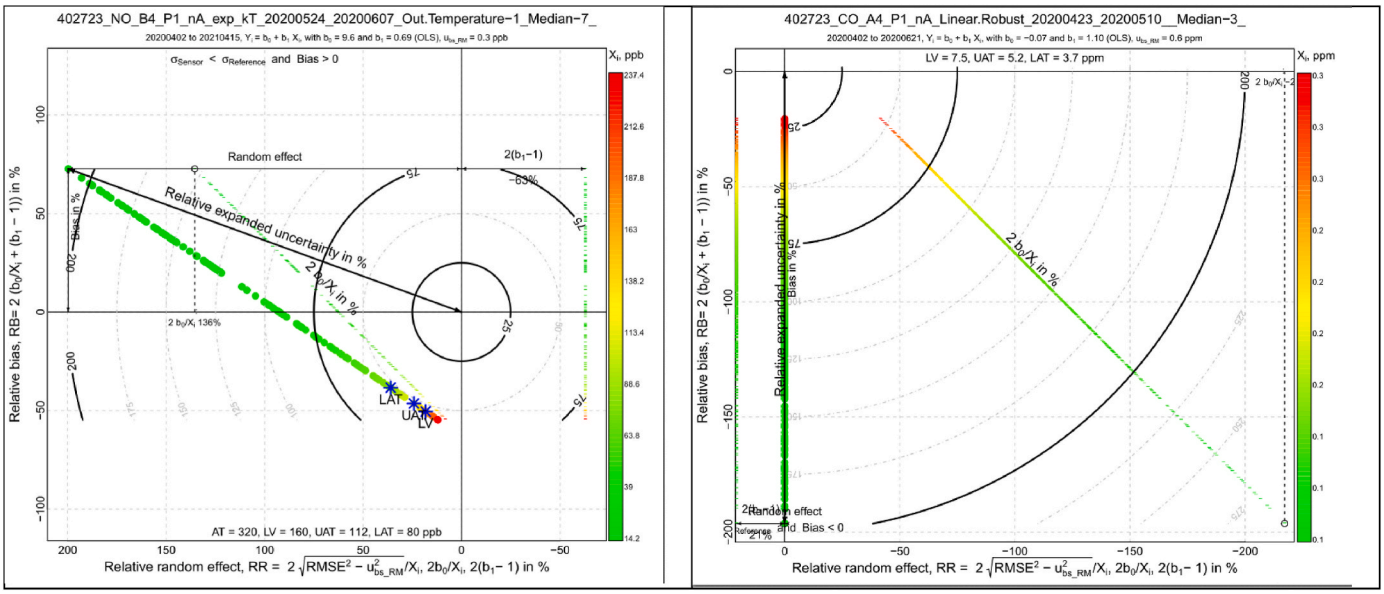

Fig. 2. Examples of MTD with $U_R(Y_i)$ dominated by the contribution of the slope (NO sensor, left) and intercept (CO sensor, right) of regression line of the LCS data versus reference data.

or the DR), and u_{bs_RM} the between reference analyser uncertainty. In Fig. 1, for an PM_{10} reference analyser Fidas, u(bs,RM) is 1.4 μg/m^3 given elsewhere (The approval test, https://qal1.de/report/0000040212_21227195C_palas_Fidas200S_en.pdf, extrapolated from Table 26 for the LV of 50 μg/m^3).

14. The caption of x-axis gives the equations and units to compute the RR and its contributions from b_0 and b_1 that are all read on the x-axis. In Fig. 1, u(bs,s) was not included in Eqs. (5) and (6) hence not in the label of x-axis either.

The MTD presented in this paper was plotted using R scripts developed by the authors (R Core Team, 2021), in particular the "Target.Diagram" function in the "Functions4ASE.R″ file (Gerboles et al., 2021).

3.2. Decomposition of bias between regression line and perfect agreement line

The RB is decomposed into two elements, $2b_0/X_i$ and $2(b_1\text{-}1)$. These elements are plotted on the x-axis of MTDs to give visual information whether the RB is dominated by either element in relation to X_i. In fact, $(b_1\text{-}1)$ is constant over X_i range while b_0/X_i decreases with X_i. An intersection point where these two elements cancel each other may occur, which is evidenced by the bold X_i coloured line crossing the x-axis. Fig. 1 shows an example of the effects of $2b_0/X_i$ (50%) and $2(b_1\text{-}1)$ (−50%) cancel each other at X_i about 30 μg/m^3. Fig. 2-right shows an example where the effect of b_0 and b_1 never cancel each other since the bold coloured line does not cross the x-axis.

Fig. 2-left gives an example of RB being dominated by the contribution of b_1 that is evidenced by the larger distance between the y-axis and the $2(b_1\text{-}1)$ line than the distance between the y-axis and the $2b_0/X_i$ line for the highest X_i. Conversely, Fig. 2-right shows an example where the RB is dominated by the contribution from b_0 that is demonstrated by the larger horizontal distance between the y-axis and the $2b_0/X_i$ line than the horizontal distance between the y-axis and the $2(b_1\text{-}1)$ line. Fig. 2-left suggests that a slope adjustment would be beneficial while Fig. 2-right suggests an offset adjustment.

In Fig. 2-right, the RR always equals to zero, resulting from subtracting a high u(bs,RM) found in the approval test (0.6 ppm), which is even higher than the X_i range (0.1–0.3 ppm CO) plotted in the MTD. The effect of constant versus variable u(bs,RM) on the RR is further discussed in section 4.3.

4. Discussion

4.1. Fast extraction of information from the MTD

The primary use of MTD is to check whether the LCS meet the DQOs of the AQD at the legislative levels presented in Table 1, for indicative measurements, modelling and objective estimation. In Fig. 1, since the LV at 50 μg/m^3 marked with a blue asterisk on the bold coloured line is within the first target circle of 50%, it can be visually confirmed that the LCS is meeting the DQO for indicative measurements with $U_R(Y_i)$ of 46% being lower than the requirement (50%, Table 1). According to the scheme developed in (CEN, 2021), this LCS could be granted to classification 1.

A simple scatterplot of $U_R(Y_i)$ versus X_i could have drawn the same conclusion. However, the MTD gives additional information such as:

- The bold coloured line being above the x-axis (1st or 4th quadrant) indicates an overestimation of the LCS compared to reference data. Conversely, the bold coloured line being below the x-axis (2nd or 3rd quadrant) indicates an underestimation of the LCS compared to reference data.
- Additionally, the bold coloured line being at the right of y-axis (1st or 2nd quadrant) indicates a higher sensitivity of the LCS compared to reference data. Conversely, the bold coloured line being at left of the y-axis (3rd or 4th quadrant) indicates a lower sensitivity of the LCS compared to reference data.
- Variation of contributors to the RB: High contribution from b_0 would indicate an offset between the LCS and reference data, possibly correctable by offset subtraction, while high contribution from b_1 may indicate an erroneous slope of calibration, possibly correctable by re-calibration or readjustment. Significant contribution from b_0 and/or b_1 are evidenced when corresponding thin coloured line(s) are far from y-axis.
- The comparison of RR and RB: In Fig. 1, the RR at the LV (40%) being two times higher than the absolute value of RB (20%) shows that $U_R(Y_i)$ is dominated by the random errors, as indicated by the RR, likely resulting from several parameters including the electronic noise of LCS.
- Improvement by adjustment of b_0 and/or b_1 in order to set the RB to zero: in Fig. 1, the adjustment of b_0 and b_1 could allow to set the RB to zero for the entire X_i range with a rotation of the bold coloured line

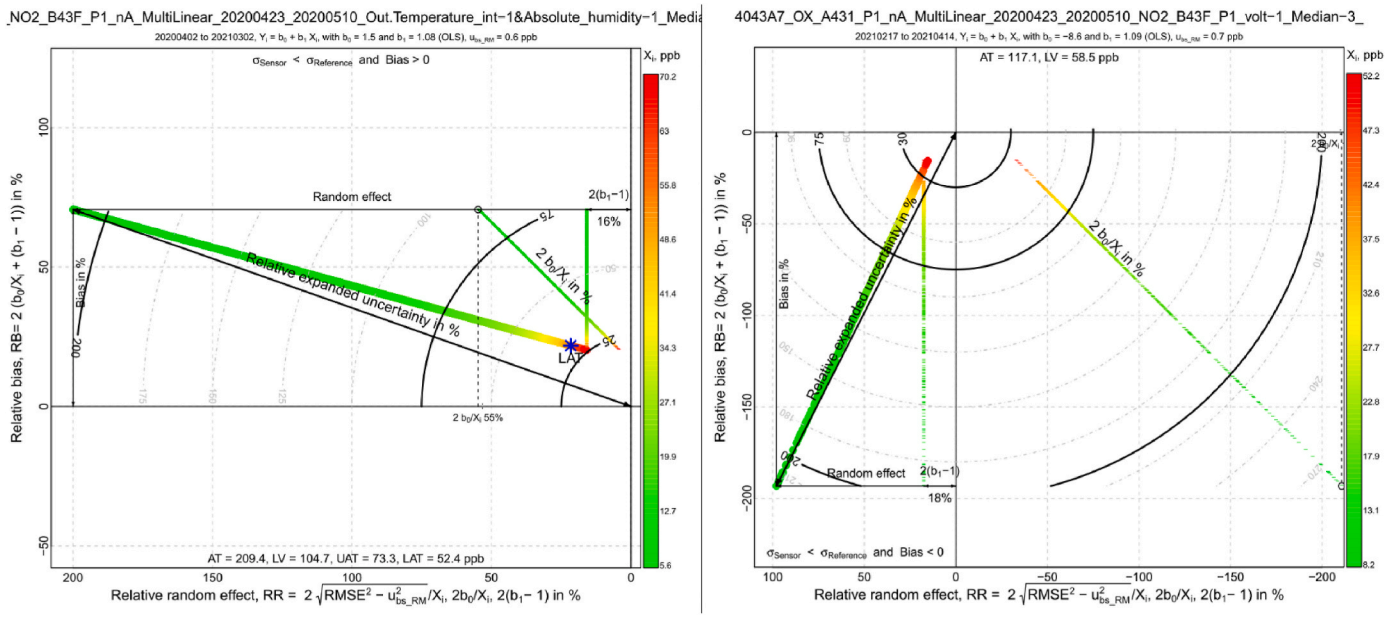

Fig. 3. Examples of MTD for NO_2 (brand name NO2_B43) and O_3 (brand name OX_A431) sensors with the effects of b_0 and b_1 not cancelling each other. At left, the contributions of b_0 and b_1 are located in the same side of the y-axis, and at right, the contributions of b_0 and b_1 are located in different side of the y-axis with overwhelming contribution from b_0.

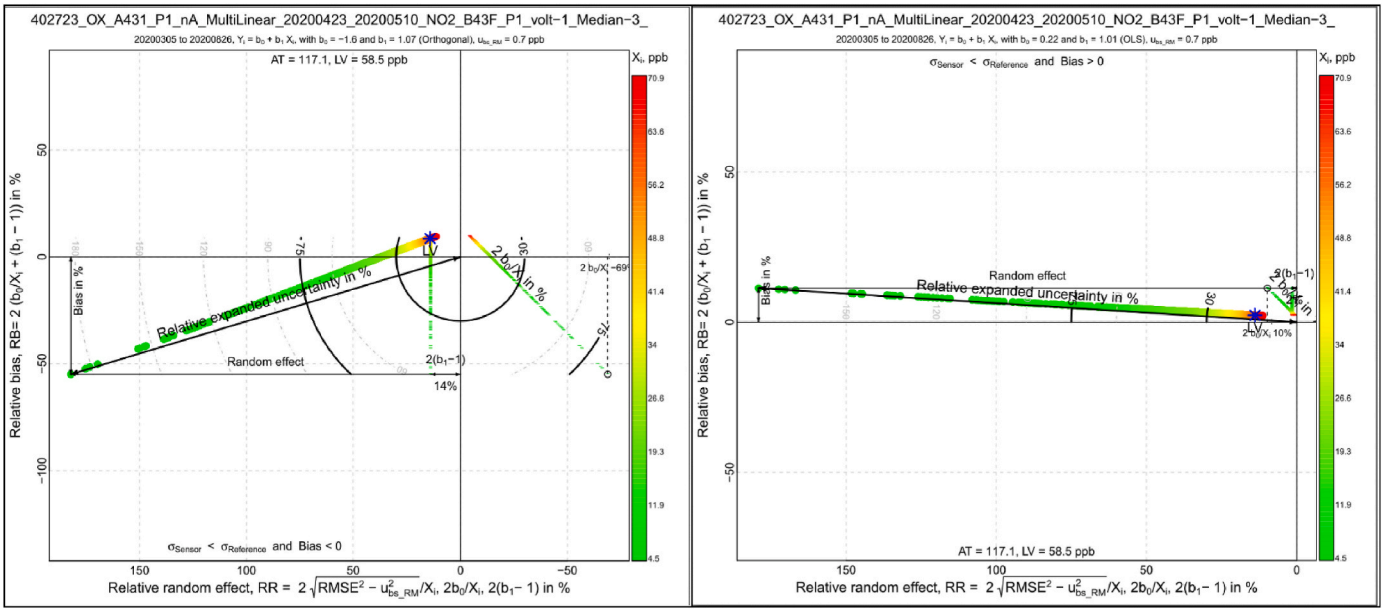

Fig. 4. Examples of MTD for an O_3 sensor (brand name OX_A431) being compared to reference data using OLS and OR at right and left, respectively.

upon the x-axis. However, since $U_R(Y_i)$ is dominated by the RR, setting the RB to zero by adjustment of b_0 and b_1 would not allow any significant decrease of $U_R(Y_i)$.

There are cases when the effects of b_0 and b_1 never cancel each other and thus the RB never crosses the x-axis. Such cases might happen due to either:

- the contributions of b_0 and b_1 are located on the same side of the y-axis with the combinations of b_0 negative and $b_1 < 1$, or, b_0 positive and $b_1 > 1$, see Fig. 3, left;
- the contributions of b_0 and b_1 are located on different sides of the y-axis but one of these contributions is overwhelming, see Fig. 3, right with the b_0 effect being higher than the b_1 effect.

4.2. *Effect of the regression line type*

Comparing the slopes and intercepts of OLS and OR reveals significant differences as shown in Fig. 4. The OLS and OR result in respectively positive and negative biases which the latter has an increased b_1 contribution of 13% to $U_R(Y_i)$ compared to 0%. This is mainly caused by the assumption of OR that uncertainties of sensor and reference data are very similar. This assumption is generally wrong as well as the robust u (bs,RM) value necessary to apply the DR is absent, therefore, the OLS is the best choice.

Even if a more realistic estimation of u(bs,RM) is available to use the DR, it is likely that u(bs,RM) and u(bs,s) are not constant over the whole X_i and Y_i ranges, respectively. In particular, u(bs,RM) and u(bs,s) approach the corresponding limits of detection of reference analyser and

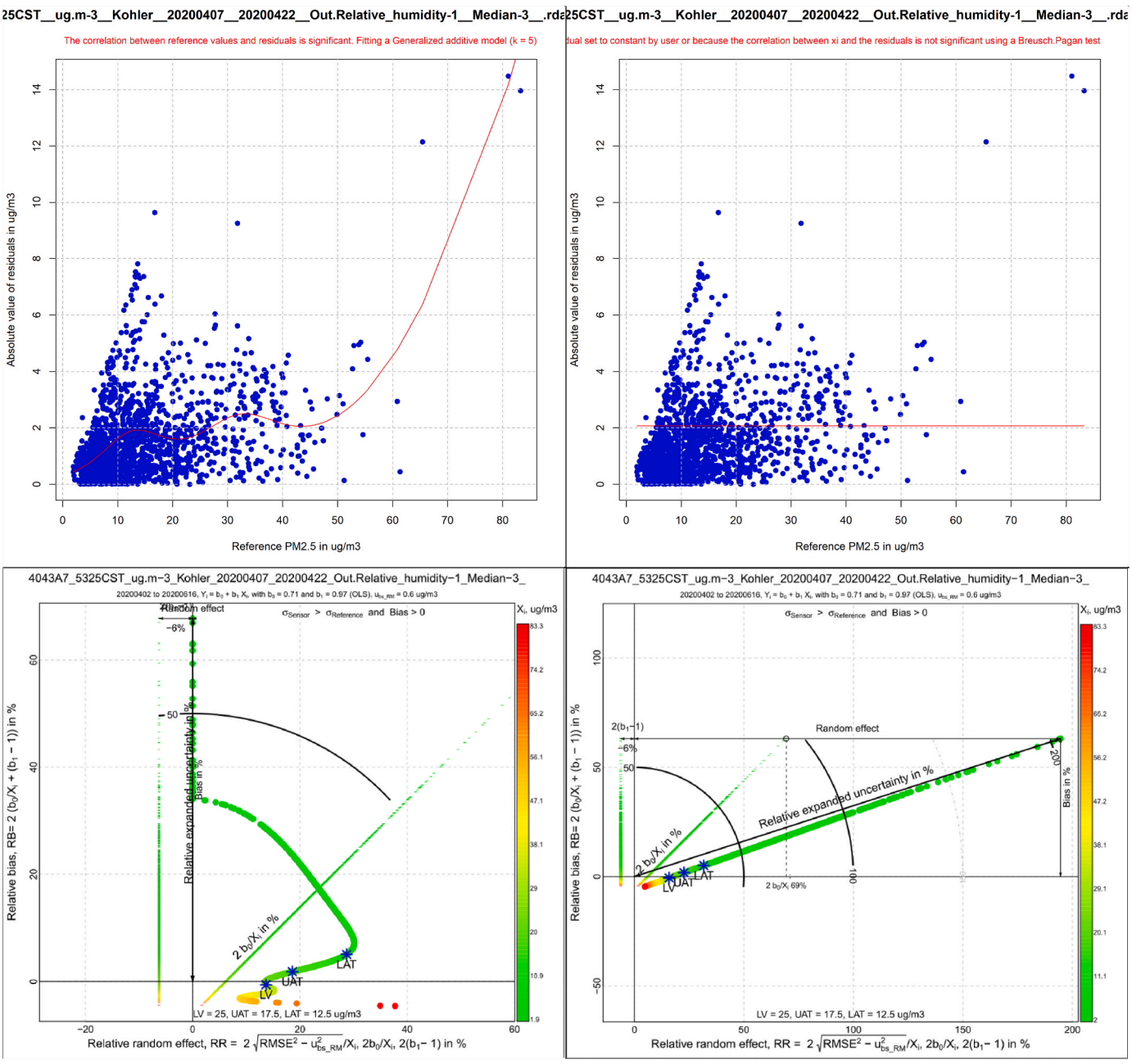

Fig. 5. Example of MTD with GAM-fitted residuals (at left) compared to constant residuals (with RSS averaged as $\frac{RSS}{(n-2)}$, at right).

LCS when X_i approaches zero. When u(bs,s) and u(bs,RM) are known along X_i and Y_i ranges, the weighted linear regression, one the most efficient methods to estimate the intercept and slope of regression line (Gerboles et al., 2005), can be implemented, although its computation is more complicated.

4.3. Effects of fitted squared residuals and u(bs,RM)

Fig. 5, upper, shows scatterplots of GAM-fitted and constant absolute value of residuals (whose squares equals to $\frac{RSS}{(n-2)}$) over the X_i range on plotting of MTD for a $PM_{2.5}$ sensor. The scatterplot of residuals shows how using a constant RSS in Eq. (5) is unrealistic. The scatterplot of GAM-fitted residuals is characterised by two bumps at about 10 and 30 $\mu g/m^3$ and an extremely high increase in RS_i towards the max X_i. For constant RSS, the MTD shows a linear increase in $U_R(Y_i)$ while the MTD with GAM-fitted RS_i follows the same pattern of RS_i: at low X_i the RR is zero until the RS_i being higher than u(bs,RM) followed by two inflexion points of $U_R(Y_i)$ representing two bumps of RS_i around 10 and 30 $\mu g/m^3$. u(bs,RM) of 0.6 $\mu g/m^3$ is drawn from the approval test of the Fidas analyser for $PM_{2.5}$ at about 20 $\mu g/m^3$ (https://qal1.de/report/0000040212_21227195C_palas_Fidas200S_en.pdf, Table 27), which leads the RR to being zero for low X_i. However, the approval test report shows that u(bs,RM) decreases with decreasing X_i, down to 0.1 $\mu g/m^3$ for $PM_{2.5}$ at 10 $\mu g/m^3$. Therefore, $U_R(Y_i)$ for Y_i below 20 $\mu g/m^3$ is likely underestimated.

The aforementioned approval test suggests a linear relationship between u(bs,RM) and X_i yielding u(bs,RM) ~ 3% of X_i. Fig. 6 shows the effect of variable u(bs,RM) set to 3% of X_i on the MTD of the same $PM_{2.5}$ sensor of Fig. 5. The unrealistic zero $U_R(Y_i)$ at low X_i is replaced by much higher values thanks to the better estimation of RR. Conversely, little to no change in $U_R(Y_i)$ can be observed for X_i higher than 10 $\mu g/m^3$.

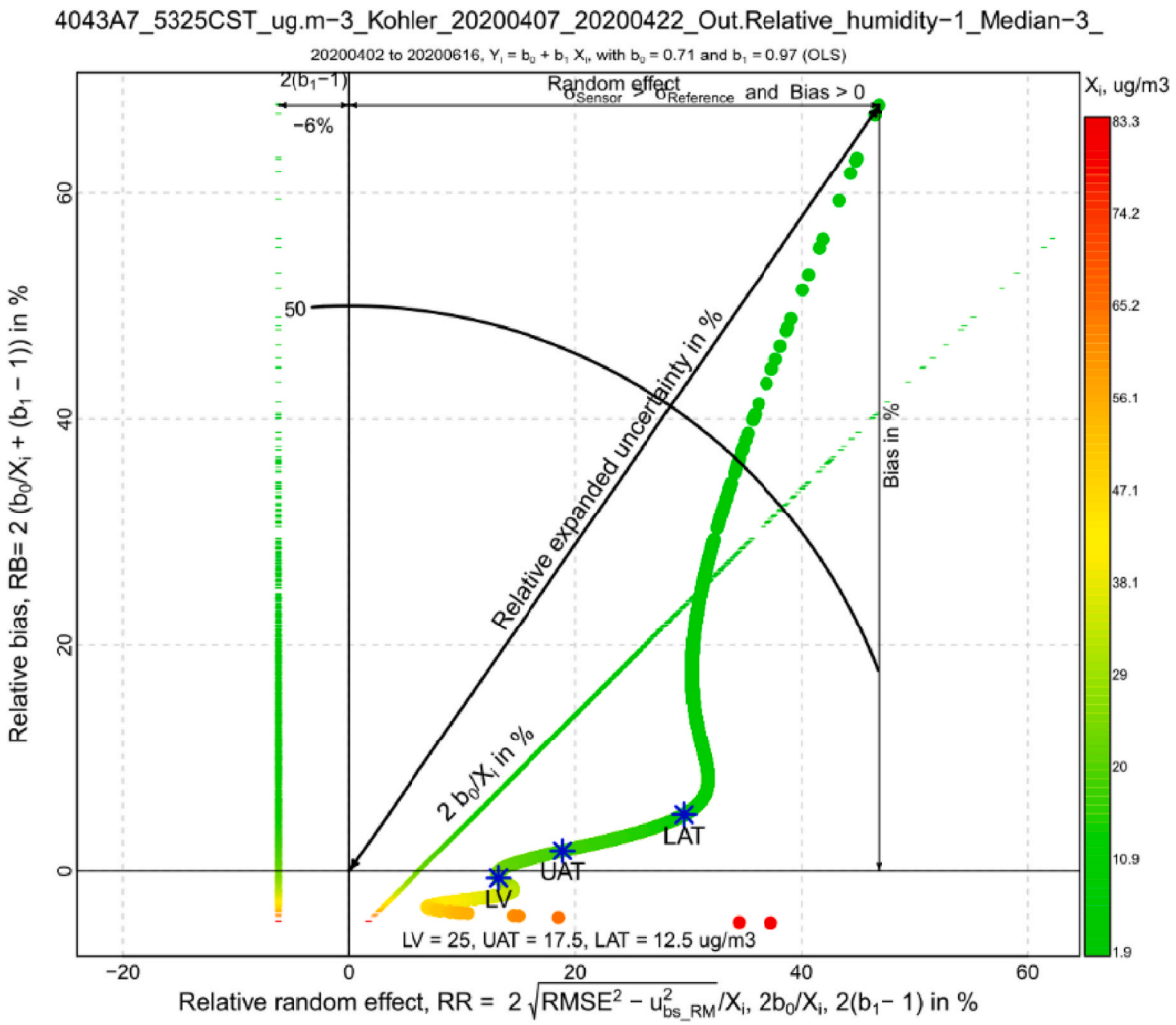

Fig. 6. Example of MTD with GAM –fitted RS_i and variable u(bs,RM) set to 3% of X_i.

5. Conclusion

The proposed MTD requires training regarding its interpretation for unexperienced users in order to extract a range of valuable information about the performance of a given LCS. Firstly, the compliance with the regulations and the DQO of AQD can be easily checked. Secondly, as the major added information, the MTD decomposes sources of errors caused either by bias of LCS data or random errors, and thus provides guidance on how to improve the LCS performance. It allows weighting the importance of error sources. When bias appears to be the major source of error, adjusting slope and intercept of regression line of LCS vs. reference data quantifies improvement of LCS directly. In case of relative expanded uncertainty of the LCS being dominated by random errors, the sensitivity of LCS related to pollutant range must be checked, i.e., pollutant level should be high enough above LCS detection limit to avoid electronic noise interfering signals of measurand. For such cases, the MTD allows the user to diagnose if the LCS calibration model is misdefined, e.g., due to lack of LCS selectivity, possible other factors (so-called covariates) effecting LCS response, aging-related-drifting of coefficients of covariates already included in calibration model, or fluctuations in the LCS baseline.

The choice of u(bs,RM) being constant or dynamic over X_i range, type of regression between LCS vs. reference data, and computation of squared residuals of selected regression line being constant or GAM-fitted over X_i range, all affect the MTD and information that can be extracted from it. In the absence of robust information indicating otherwise, the OLS is the best choice. Although it is tempting to follow the methodology given in the GDE and CEN/TS 17660, the MTD would better characterize the real performance of LCS by using dynamic u(bs, RM) and GAM-fitted RS_i rather than constant corresponding values.

The drawback of MTD is the colour scale making accurate interpolation of X_i and $U_R(Y_i)$ difficult, in particular when X_i range is large. Nevertheless, the MTD provides a wide range of useful information for evaluating not only performance of LCS but also any other candidate method compared to reference method.

Credit authorship contribution statement

Sinan Yatkin: Conceptualization, Methodology, Formal analysis, Writing - original draft. Michel Gerboles: Conceptualization, Methodology, Supervision, Formal analysis, Writing - original draft. Annette Borowiak: Funding acquisition, writing - review & editing. Silvije Davila, Laurent Spinelle, Alena Bartonova, Frank Dauge, Philipp Schneider, Martine Van Poppel, Jan Peters, Christina Matheeussen, and Marco Signorini: Data collection, writing - review & editing.

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 opinions expressed in this article do not necessarily represent those of the European Union. The work was carried out in the framework of a Pilot Project "Integrating smart sensors and modelling for air quality monitoring in cities" proposed by the European Parliament and implemented by the European Commission (Service contract no. 07027747/2019/812686/SER/ENV.C.3).

References

Aleixandre, M., Gerboles, M., 2012. Review of small commercial sensors for indicative monitoring of ambient gas. Chem. Eng. Trans. 30, 169–174. https://doi.org/10.3303/CET1230029.

ASTM, 2018. WK64899 new practice for performance evaluation of ambient air quality sensors and other sensor-based instruments [WWW document]. URL. https://www.astm.org/DATABASE.CART/WORKITEMS/WK64899.htm, 10.26.21.

Breusch, T., Pagan, A., 1979. A simple test for heteroscedasticity and random coefficient variation. Econometrica 47, 1287–1294.

CEN, 2021. CEN/TS 17660-1:2021–Air quality — performance evaluation of air quality sensor systems — Part 1: gaseous pollutants in ambient air [WWW Document]. ITeh Stand. Store. URL. https://standards.iteh.ai/catalog/standards/cen/5bdb236e-95a3-4b5b-ba7f-62ab08cd21f8/cen-ts-17660-1-2021, 10.26.21.

Collier-Oxandale, A., Feenstra, B., Papapostolou, V., Zhang, H., Kuang, M., Der Boghossian, B., Polidori, A., 2020. Field and laboratory performance evaluations of 28 gas-phase air quality sensors by the AQ-SPEC program. Atmos. Environ. 220 (117092) https://doi.org/10.1016/j.atmosenv.2019.117092.

Cuvelier, C., Thunis, P., Vautard, R., Amann, M., Bessagnet, B., Bedogni, M., Berkowicz, R., Brandt, J., Brocheton, F., Builtjes, P., Carnavale, C., Coppalle, A., Denby, B., Douros, J., Graf, A., Hellmuth, O., Hodzic, A., HonorÃ©, C., Jonson, J., Kerschbaumer, A., de Leeuw, F., Minguzzi, E., Moussiopoulos, N., Pertot, C., Peuch, V.H., Pirovano, G., Rouil, L., Sauter, F., Schaap, M., Stern, R., Tarrason, L., Vignati, E., Volta, M., White, L., Wind, P., Zuber, A., 2007. CityDelta: a model intercomparison study to explore the impact of emission reductions in European cities in 2010. Atmos. Environ. Times 41, 189–207. https://doi.org/10.1016/j.atmosenv.2006.07.036.

Deming, W.E., 1943. Statistical Adjustment of Data. John Wiley & Sons ; Chapman & Hall, New York; London.

EC, 2008. Directive 2008/50/EC of the European Parliament and the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.

European Commission, 2010. Guide to the demonstration of equivalence of ambient air monitoring methods, Report by an EC Working, Group on Guidance. European Commission.

Gerboles, M., Buzica, D., Amantini, L., 2005. Modification of the Palmes diffusion tube and semi-empirical modelling of the uptake rate for monitoring nitrogen dioxide. Atmos. Environ. 39, 2579–2592. https://doi.org/10.1016/j.atmosenv.2005.01.012.

Gerboles, M., Spinelle, L., Kotsev, A., Karagulian, F., Signorini, M., 2021. airsenseur-calibration [WWW Document]. GitHub Repos. URL. https://github.com/ec-jrc/airsenseur-calibration.

Hastie, T., Tibshirani, R., 1987. Generalized additive models: some applications. J. Am. Stat. Assoc. 82, 371–386. https://doi.org/10.1080/01621459.1987.10478440.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. CRC Press.

Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R., Arnone, R.A., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. J. Mar. Syst. Skill Assess. Coupled Biol./ Phys. Models Mar. Syst. 76, 64–82. https://doi.org/10.1016/j.jmarsys.2008.05.014.

Karagulian, F., Barbiere, M., Kotsev, A., Gerboles, M., Lagler, F., Borowiak, A., 2019. Review of sensors for air quality monitoring (EUR - scientific and technical research reports). Publ. Off. Eur. Union. https://doi.org/10.2760/568261 (online).

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., Britter, R., 2015. The rise of low-cost sensing for managing air pollution in cities. Environ. Bar Int. 75, 199–205. https://doi.org/10.1016/j.envint.2014.11.019.

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

Lewis, A., Von Schneidemesser, E., Peltier, R., Lung, C., Jones, R., Zellweger, C., Karppinen, A., Penza, M., Amagai, Hüglin, C., Ning, Z., Leigh, R., Hagan, D., Laurent, O., Carmichael, G., Akland, G.G., Cohen, R., Cross, E., Gentner, D., Gerboles, M., Khan, S., Kroll, J., Musu, P., Querol, X., Ruggeri, G.A.K., Tarasova, O., 2018. Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications. World Meteorological Organization, Geneva (Switzerland).

Papapostolou, V., Zhang, H., Feenstra, B.J., Polidori, A., 2017. Development of an environmental chamber for evaluating the performance of low-cost air quality sensors under controlled conditions. Atmos. Environ. 171, 82–90. https://doi.org/10.1016/j.atmosenv.2017.10.003.

Pernigotti, D., Gerboles, M., Belis, C.A., Thunis, P., 2013. Model quality objectives based on measurement uncertainty. PartII: NO2 and PM10. Atmos. Environ. 79, 869–878. https://doi.org/10.1016/j.atmosenv.2013.07.045.

QAL. QAL1-Certified measuring- and evaluating-systems [WWW Document]. n.d. URL. https://qal1.de/en/hersteller/thermo.htm, 10.11.21.

R Core Team, 2021. R: a language and environment for statistical computing. In: R Foundation for Statistical Computing. Austria, Vienna.

Rai, A.C., Kumar, P., Pilla, F., Skouloudis, A.N., Di Sabatino, S., Ratti, C., Yasar, A., Rickerby, D., 2017. End-user perspective of low-cost sensors for outdoor air pollution monitoring. Sci. Total Environ. 607, 691–705. https://doi.org/10.1016/j.scitotenv.2017.06.266.

Spinelle, L., Gerboles, M., Kok, G., Persijn, S., Sauerwald, T., 2017. Review of portable and low-cost sensors for the ambient air monitoring of benzene and other volatile organic compounds. Sensors 17, 1520. https://doi.org/10.3390/s17071520.

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

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106, 7183–7192. https://doi.org/10.1029/2000JD900719.

Thunis, P., Pederzoli, A., Pernigotti, D., 2012. Performance criteria to evaluate air quality modeling applications. Atmos. Environ. 59, 476–482. https://doi.org/10.1016/j.atmosenv.2012.05.043.

Thunis, P., Pernigotti, D., Gerboles, M., 2013. Model quality objectives based on measurement uncertainty. Part I: Ozone. Atmos. Environ 79, 861–868. https://doi.org/10.1016/j.atmosenv.2013.05.018.

US EPA, O. of R. and D, 2021. Performance testing protocols, metrics, and target values for ozone air sensors: use in ambient, outdoor, fixed site, non-regulatory and informational monitoring applications [WWW document]. URL. https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=350784&Lab=CEMM, 10.26.21.

US-EPA, O., 2021. Of R. And Dperformance testing protocols, metrics, and target values for fine particulate matter air sensors: use in ambient, outdoor, fixed site, non-regulatory supplemental and informational monitoring applications [WWW document]. URL. https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=350785&Lab=CEMM, 10.26.21.

Vautard, R., Builtjes, P.H.J., Thunis, P., Cuvelier, C., Bedogni, M., Bessagnet, B., Honoré, C., Moussiopoulos, N., Pirovano, G., Schaap, M., Stern, R., Tarrason, L., Wind, P., 2007. Evaluation and intercomparison of Ozone and PM10 simulations by several chemistry transport models over four European cities within the CityDelta project. Atmos. Environ. 41, 173–188. https://doi.org/10.1016/j.atmosenv.2006.07.039.

World Health Organization, 2021. WHO global air quality guidelines. In: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization.

Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. 1, 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x.