Evaluation of the inter-laboratory comparison exercise for SO2, CO, O3, NO and NO2 \*(13-16 May 2019, Ispra)

Barbiere M., Lagler F., Borowiak A.

last update 15 July 2019, 11:45

## 1. Acknowledgements

In collaboration with: [BILIC Marijo, HECTOR David, DAVIES Ben, HERCOG Predrag, LEPRI Petra, MARLIERE Fabrice, MIHAJLOVIĆ Domagoj, MRVOS Luka, NATHALIE Bocquet, PANAYOTOV Nikolay, PARVANOVA Milena, SILHAVY Jan, VOKOUN Miroslav, ZOBAJ Eugen].

## 2. Abstract

Within the harmonisation programme of Air Quality monitoring in Europe the European Reference Laboratory of Air Pollution (ERLAP) organises Inter-Laboratory Comparison Exercises (ILC). From the 13th to the 16th of May 2019, nine Laboratories of AQUILA (Network of European Air Quality Reference Laboratories) met for a laboratory comparison exercise in Ispra (IT) to evaluate their proficiency in the analysis of inorganic gaseous air pollutants (NO, NO2, SO2, CO and O3) covered by the European Air Quality Directive 2008/50 EC [1] and its recent amendments 2015/1480/EC [42]. The proficiency evaluation, where each participant’s bias was compared to two criteria, provides information on the current situation and capabilities to the European Commission and can be used by participants in their quality control system. On the basis of adopted criteria, 95.6% of the results reported by AQUILA laboratories were good both in terms of measured values and reported uncertainties. The rest of the results had good measured values, but the reported uncertainties were either too high (1.3%) or too small (3.1%). Based on the z’-score evaluation, no values were found to be questionable or unsatisfactory. Comparability of results among AQUILA participants at the highest generated concentration levels is satisfactory for measurements of all pollutants.

## 3. Introduction

The Directive 2008/50/EC [1] on ambient air quality and cleaner air for Europe sets a framework for a harmonised air quality assessment in Europe. One important objective of the Directive [1] is that the ambient air quality shall be assessed on the basis of common methods and criteria. It deals with the air pollutants sulphur dioxide (SO2), nitrogen dioxide (NO2) and monoxide (NO), particulate matter, lead, benzene, carbon monoxide (CO) and ozone (O3). Among others it specifies the reference methods for measurements and Data Quality Objectives (DQOs) for the accuracy of measurements. The European Commission (EC) has supported the development and publication of reference measurement methods for CO [2], SO2 [3], NO-NO2 [4] and O3 [5] as European standards. Appropriate calibration methods [6], [7] and [8] have been standardised by the International Organization for Standardization (ISO). As foreseen in the Air Quality Directive, the European Reference Laboratory of Air Pollution (ERLAP) of the Directorate for Energy, Transport and Climate at the Joint Research Centre (JRC) organises inter-laboratory comparison exercises (ILC) to assess and improve the status of comparability of measurements of National Reference Laboratories (NRL) of the Member States of the European Union. The World Health Organization Collaborating Centre for Air Quality Management and Air Pollution Control, Berlin (WHO CC) is carrying out similar activities since 1994 [9] [10], [24], [31], [35], [38] and [45] but with a view to obtaining harmonised air quality data for health related studies. Their programme integrates within the WHO EURO region, which includes public health institutes and other national institutes - especially from the Central Eastern Europe, Caucasus and countries from Central Asia. Starting in 2004, it has been decided to bring together the efforts of both the JRC-ERLAP and WHO CC and to coordinate activities as far as possible, with a view to optimise resources and improve international harmonisation. The following report deals with the ILC that took place from 4th to the 7th of June 2018 in Ispra (IT). Since 1990 ERLAP has organised ILC in order to evaluate the comparability of measurements carried out by NRLs and promote information exchange among the expert laboratories. Recently, a more systematic approach has been adopted, in agreement with the Network of National Reference Laboratories for Air Quality (AQUILA) [11], aiming to both provide an alert mechanism for the purposes of the EC legislation and support the implementation of quality schemes by NRLs. The methodology for the organisation of ILC was developed by ERLAP in collaboration with AQUILA and is described in a paper on the organisation of laboratory comparison exercises for gaseous air pollutants [12]. This evaluation scheme was adopted by AQUILA in December 2008 and is applied to all ILC since then. It contains common criteria to alert the EC on possible performance failures which do not rely solely on the uncertainty claimed by participants. The evaluation scheme implements the z’-score method [13] with the uncertainty requirements for calibration gases stated in the European standards [2], [3], [4] and [5], which are consistent with the DQOs of European Directives. According to the above-mentioned document, NRLs with an overall unsatisfactory performance in the z’-score evaluation (one unsatisfactory or two questionable results per parameter) ought to repeat their participation in the following ILC in order to demonstrate remediation measures [12]. In addition, considering that the evaluation scheme should be useful to participants for accreditation according to ISO 17025, they are requested to include their measurement uncertainty. Hence, participants’ results (measurement values and uncertainties) are compared to the assigned values applying the En–score method [13]. Beside the proficiency of participating laboratories, the repeatability and reproducibility of standardised measurement methods [14], [15] and [16] are evaluated as well. These group evaluations are useful indicators of trends in measurement quality over different ILC.

All references here []

## 4. Inter-laboratory organization

The ILC was announced in February 2018 to the members of the AQUILA network and the WHO CC representative. Registration was opened in April 2019 and closed at the end of May 2019. The participants were required to bring their own measurement instruments, data acquisition equipment and travelling standards (to be used for calibrations or checks during the ILC). The participants were invited to arrive on Monday, 13th of May 2019, for the installation of their equipment. The calibration of NOx and O3 analysers was carried out on Tuesday morning and the generation of NOx and O3 gas mixtures started at 11:00. The calibration of SO2 and CO analysers was carried out on Wednesday afternoon and the generation of CO and SO2 gas mixtures started at 20:00. The test gases generation and measurements finished on Thursday at 9:00.

### 4.1 Participants

All participants were organisations dealing with the routine ambient air monitoring or institutions involved in environmental or public health protection. The national representatives came from Croatia, France, United Kingdom, Bulgaria, Czech Republic, Croatia, Italy, Croatia.

Table 1: List of participating organizations.

|  |  |  |  |
| --- | --- | --- | --- |
| acronim | full\_name | country\_name | code |

|  |  |  |  |
| --- | --- | --- | --- |
| EKONERG | Energy and Environmental Protection Institute | Croatia | A |
| INERIS | Institut National de lâ€™Environnement Industriel et des Risques | France | B |
| AEA | RICARDO-AEA Technology | United Kingdom | C |
| EEA | Executive Environmental Agency | Bulgaria | D |
| CHMI | Czech Hydrometereological Institute | Czech Republic | E |
| DHZ-CAL | Meteorological and Hydrological Service - CALIBRATION | Croatia | F |
| ERLAP | European Reference Laboratory for Air Pollution | Italy | G |
| DHZ-TES | Meteorological and Hydrological Service - TEST | Croatia | H |

Table 2 reports the manufacturer and model of the instrumentation used by every participant during the inter-laboratory comparison exercise including those used in the calculation of the assigned values. The instrumentation used to analyse all parameters was manufactured by three different companies. The list contains the information reported by participants and cannot be considered as an implicit or explicit endorsement by the organisers of any specific instrumentation.

Table 2: List of instruments used by participants.

[GAS LAB CODE INSTRUMENT]

### 4.2 Preparation of test mixtures

The ERLAP ILC facility has been described in several reports [17], [18]. During this ILC, gas mixtures were prepared for SO2, CO, O3, NO and NO2 at concentration levels around limit values, critical levels and assessment thresholds set by the European Air Quality Directive [1]. The test mixtures were prepared by the dilution of gases from cylinders containing high concentrations of NO, SO2 or CO using thermal mass flow controllers [8]. O3 was added using an ozone generator and NO2 was produced applying the gas phase titration method [19] in a condition of NO excess. The participants were required to report three half-hour-mean measurements for each concentration level (run) in order to evaluate the repeatability of standardised measurement methods. Zero concentration levels were generated for one hour and one half-hour-mean measurement was reported. The sequence programme of generated test gases is given in Table 3.

Table 3: Sequence program of generated test gases with indicative pollutant concentrations

# A tibble: 29 x 12

day start time duration parameter installation calibration   
1 h   
2 1st 0.375 5 / X   
3 2nd 0.333333333~ 3 / X  
4 2nd 0.458333333~ 1 NO-NO2-O3   
5 2nd 0.5 2 NO-NO2   
6 2nd 0.583333333~ 2 NO-NO2   
7 2nd 0.666666666~ 2 O3   
8 2nd 0.75 2 NO-NO2   
9 2nd 0.833333333~ 2 NO-NO2   
10 2nd 0.916666666~ 2 O3   
# … with 19 more rows, and 6 more variables: Zero Air , NO , # NO2 , O3 , CO , SO2

## 5. The evaluation of laboratory’s measurement proficiency

To evaluate the participant’s measurement proficiency, the methodology described in ISO 13528 [13] was applied. It has been agreed among the AQUILA members to take the measurement results of ERLAP as the assigned/reference values for the whole ILC [12]. The traceability of ERLAP’s measurement results and the method applied to validate them are presented in 0. In the following proficiency evaluations, the uncertainty of test gas homogeneity (0) was added to the uncertainties of ERLAP’s measurement results. All data reported by participating laboratories are presented in Annex A. As described in the position paper [12], the proficiency of the participants was assessed by calculating two performance indicators. The first performance indicator (z’-score) tests whether the difference between the participants measured value and the assigned/reference value remains within the limits of a common criterion. The second performance indicator (En-score) tests if the difference between the participants measured values and assigned/reference value remains within the limits of a criterion, that is calculated individually for each participant, from the uncertainty of the participants measurement result and the uncertainty of the assigned/reference value.

### 5.1 z’–score

The z’- score statistic is calculated according to ISO 13528 [13] as:

Equation 1

where xi is a participant’s average value for each run, X is the assigned/reference value, σp is the standard deviation for proficiency assessment and uX is the standard uncertainty of the assigned value. For a and b see [Table 4]. In the European standards [2], [3], [4] and [5] the uncertainties for calibration gases used in ongoing quality control are prescribed. In fact, it is stated that the maximum permitted expanded uncertainty for calibration gases is 5% and that ‘zero gas’ shall not give instrument reading higher than the detection limit. As one of the tasks of NRLs is to supply calibration gas mixtures, the ‘standard deviation for proficiency assessment’ (p) [13] is calculated in fitness-for-purpose manner from requirements given in European standards. Over the whole measurement range is calculated by linear interpolation between 2.5% at the calibration point (75% of calibration range) and the limit of detection at zero concentration level. The limits of detection of studied measurement methods were evaluated from the data of previous ILC. The linear function parameters of p are given in [Table 4].

Table 4: Standard deviation for proficiency assessment (). { is a linear function of concentration (c) with parameters: slope (a) and intercept (b).

# A tibble: 7 x 3

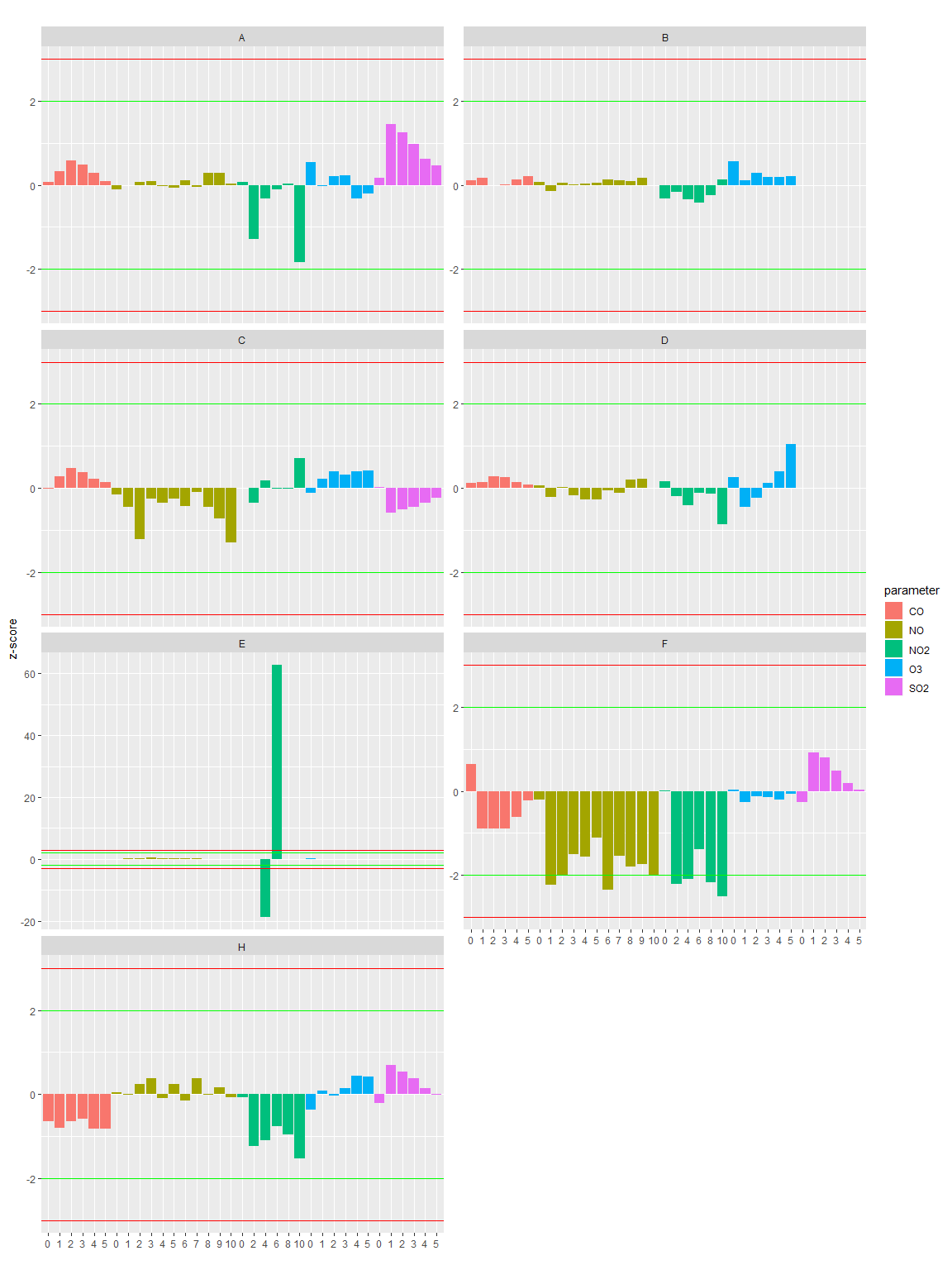
Gas a b (nmol/mol) 1 SO2 0.022 1 2 CO 0.024 100 3 O3 0.02 1 4 NO 0.024 1 5 NO2 0.02 1 6 NA NA 7 sp=a·c+b NA NA

$$ \sigma \_p = (a times c) + b $$

The assessment of results in the z‘-score evaluation is made according to the following criteria: • |z’| ≤2 are considered satisfactory. • 2 < |z’| < 3 are considered questionable. • |z’| ≥ 3 are considered unsatisfactory. Scores falling in this range are very unusual and are taken as evidence that an anomaly has occurred that should be investigated and corrected.

The results of z’-score evaluation are presented in bar plots (Figure 1 to Figure 5) in which the z’-scores of each participant are grouped together, and assessment criteria are presented as z’=±2 and z’=±3 lines.

Figure 1: Z’-score evaluations of SO2 measurements Scores are given for each participant and each tested concentration level (run). Run number order (with nominal concentration) is: 0 (0 nmol/mol), 1 (124 nmol/mol), 2 (67 nmol/mol), 3 (36 nmol/mol), 4 (11 nmol/mol), 5 (21 nmol/mol). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results.



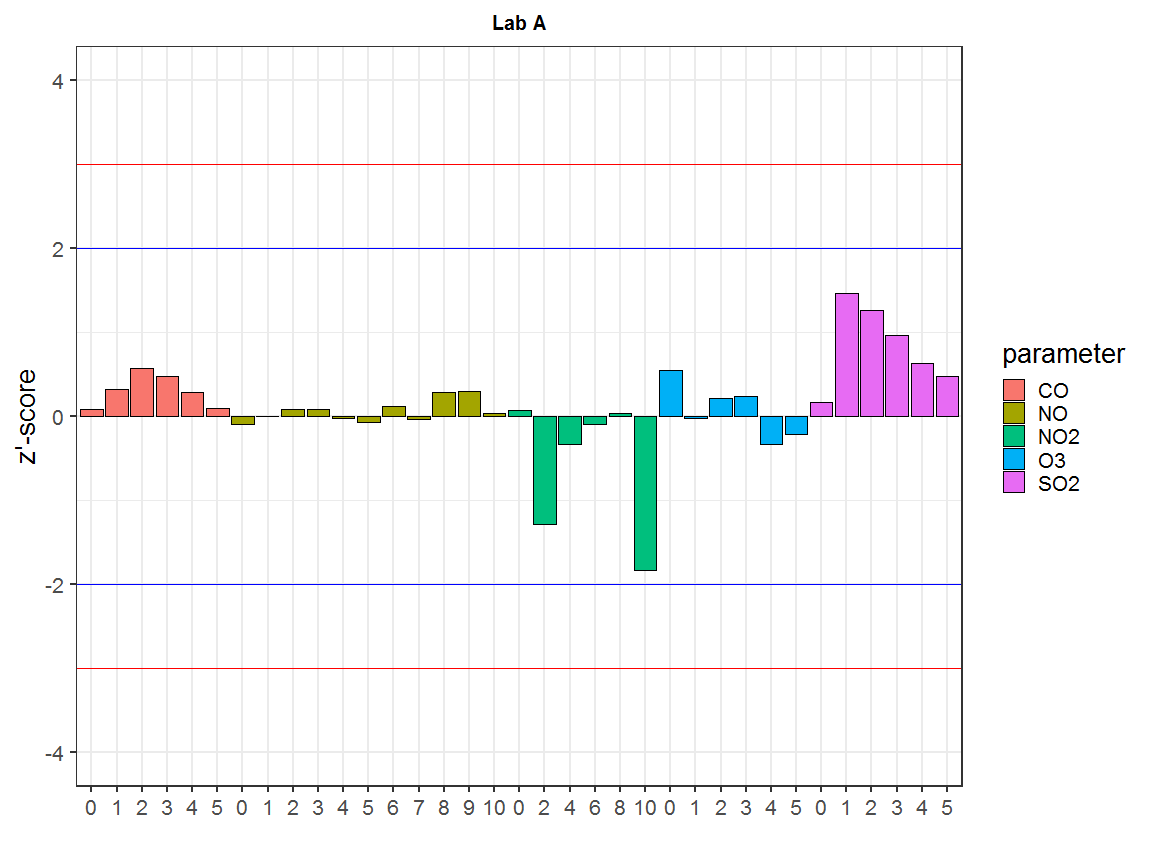
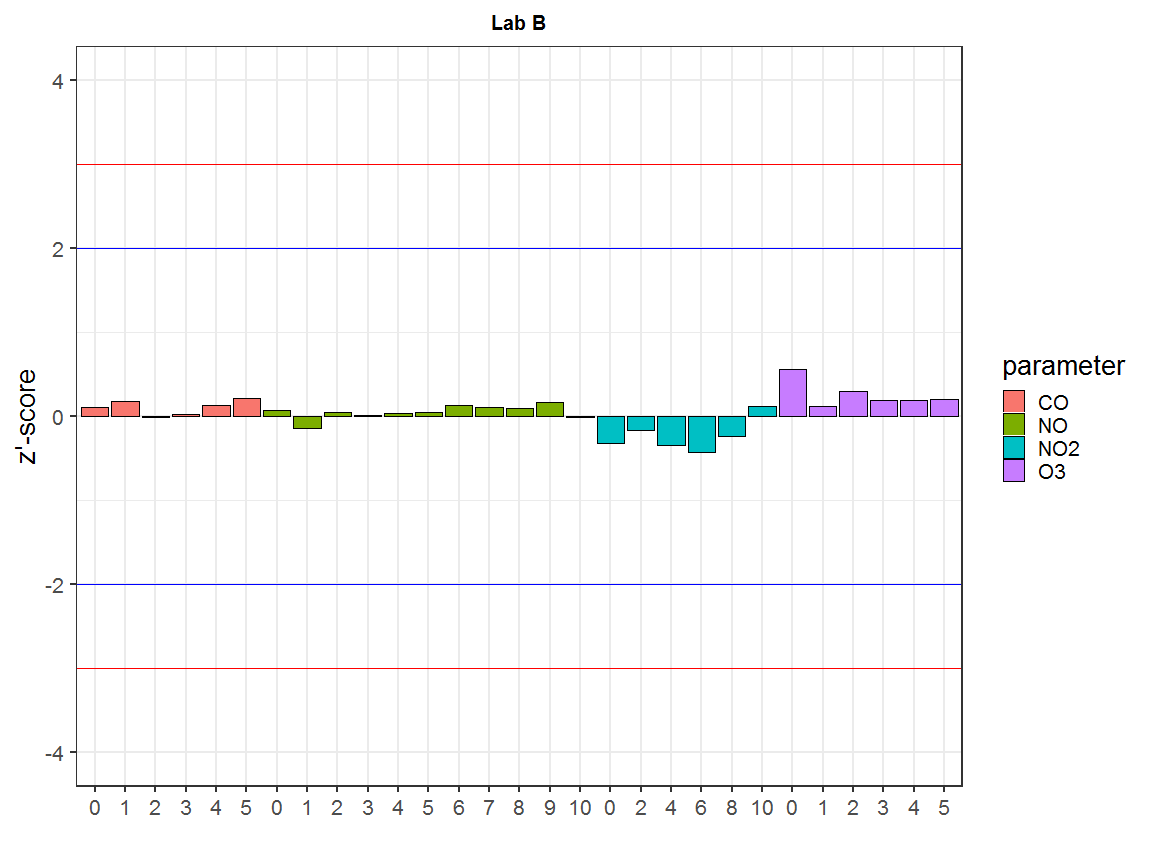
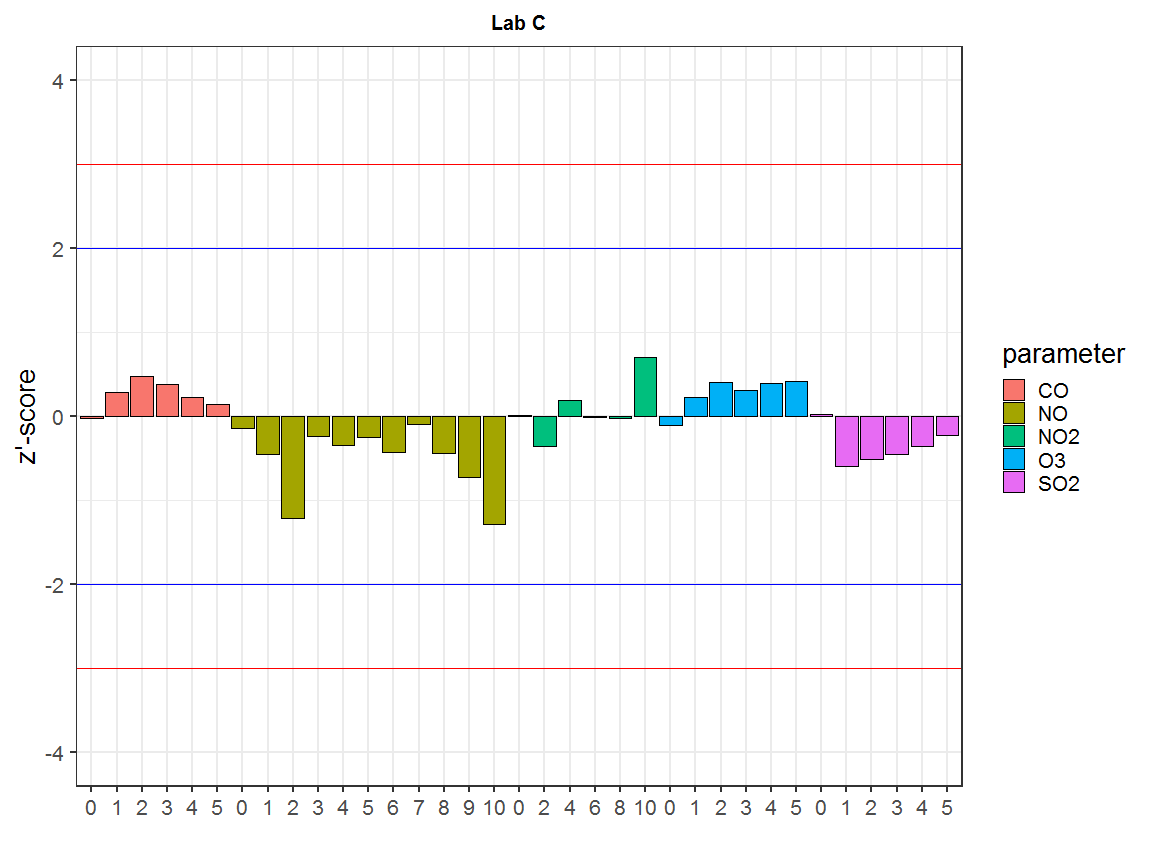
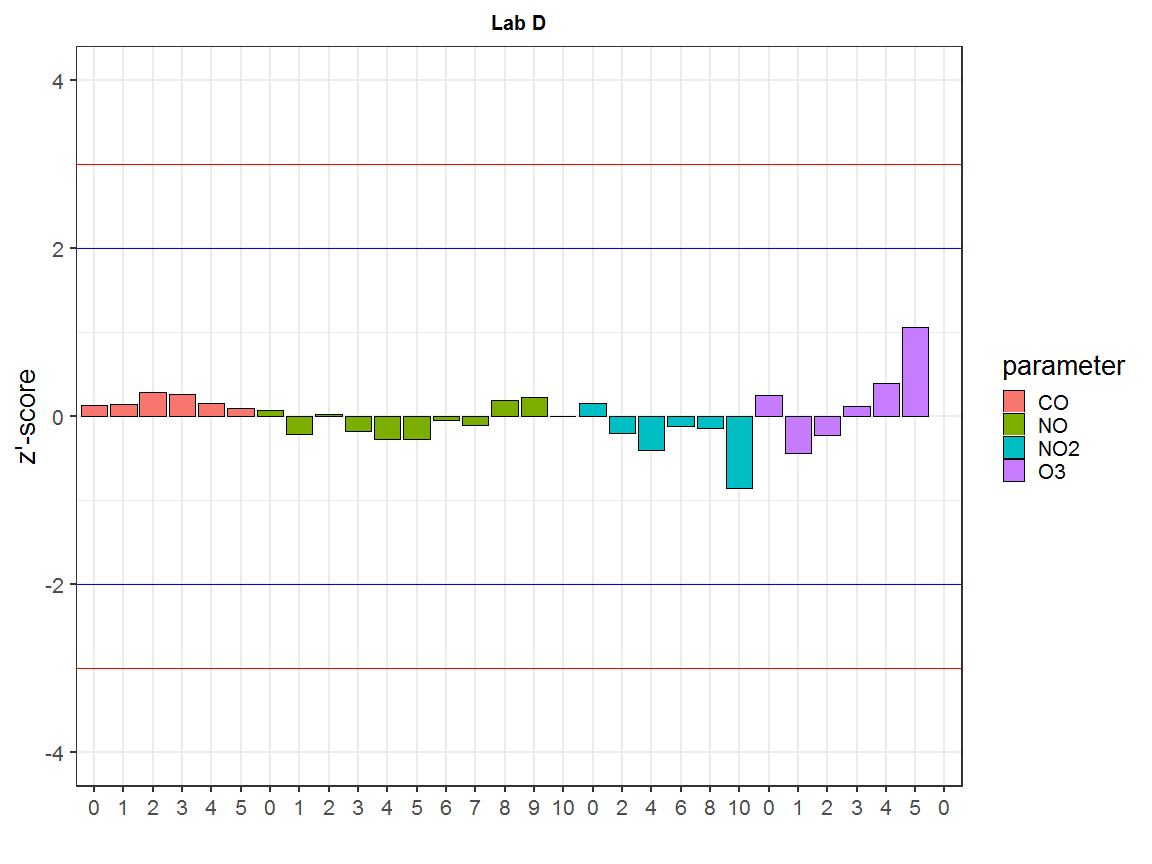
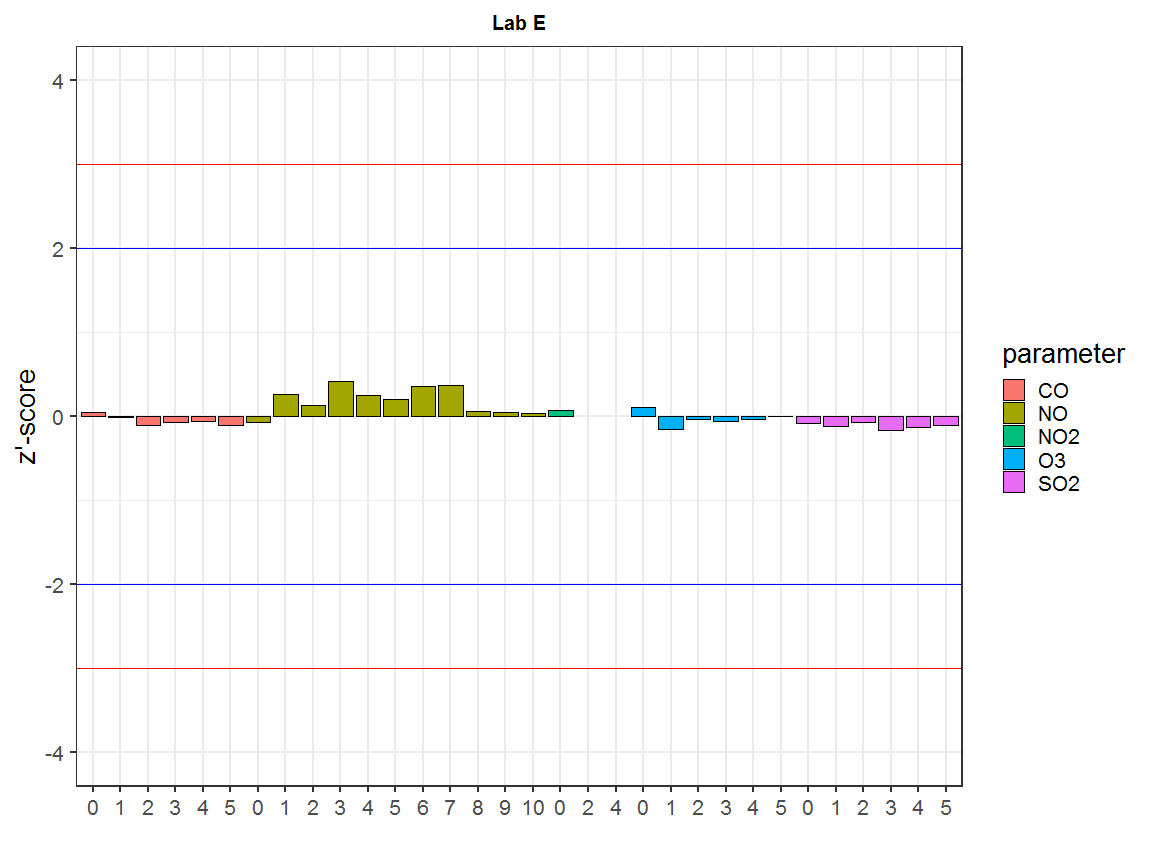
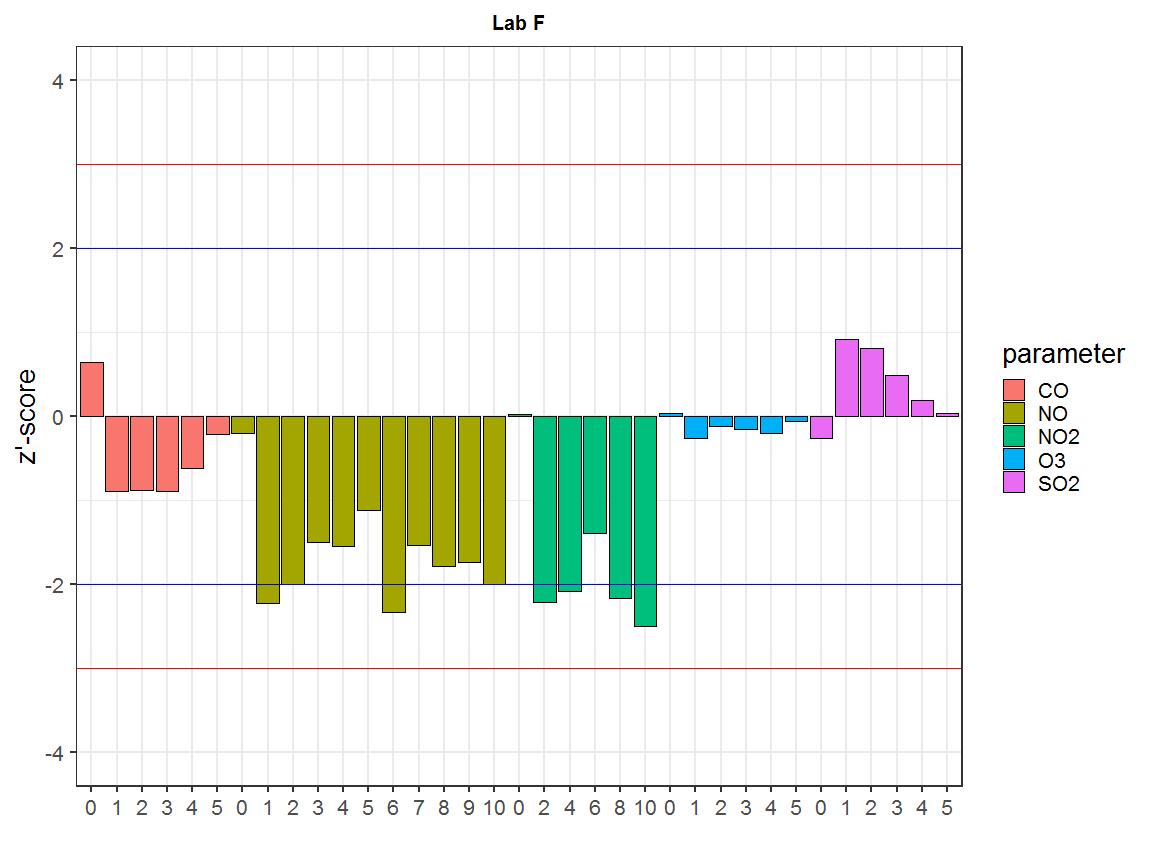
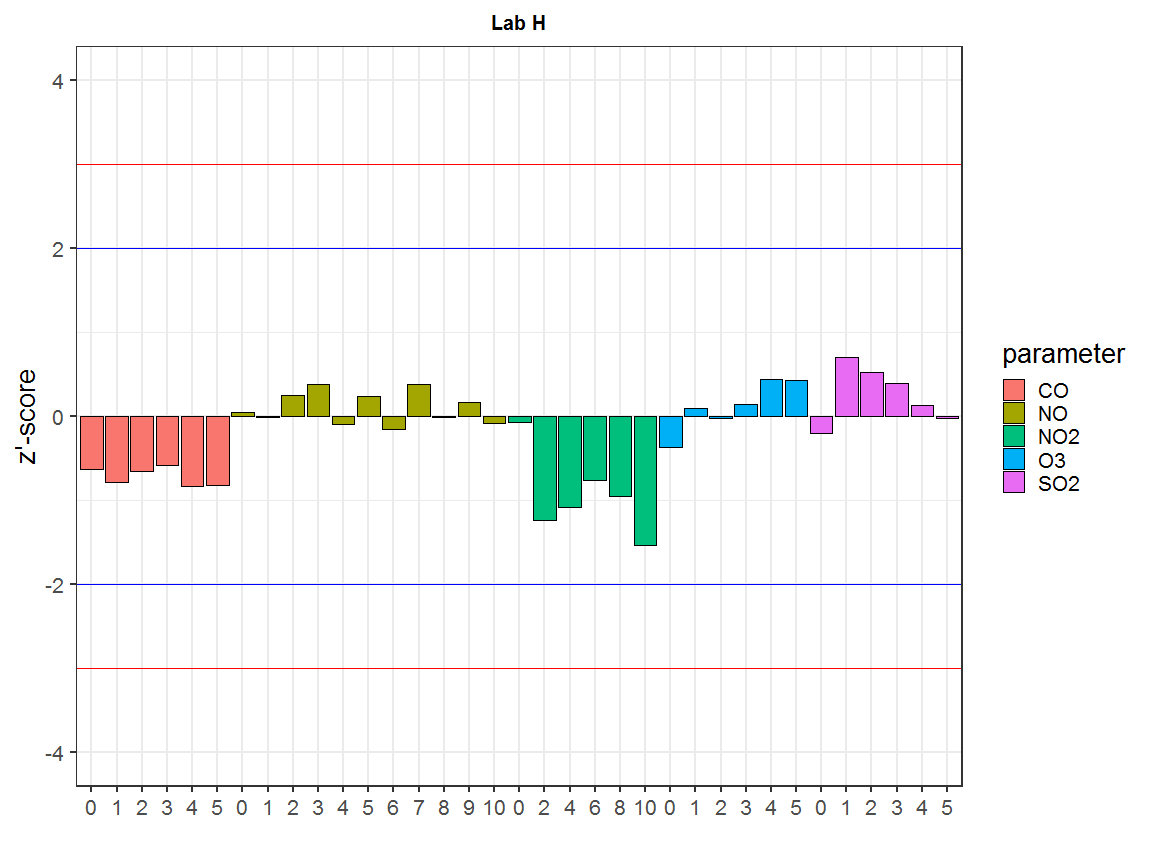
[1] Figure 1: Z’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 2: Z’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 3: Z’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 4: Z’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 5: Z’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 6: Z’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 7: Z’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results 

Figure 2: Z’-score evaluations of CO measurements Scores are given for each participant and each tested concentration level (run). Run number order (with nominal concentration) is: 0 (0 μmol/mol), 1 (8.3 μmol/mol), 2 (3 μmol/mol), 3 (2.1 μmol/mol), 4 (1 μmol/mol), 5 (4.8 μmol/mol). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results

Figure 3: Z’-score evaluations of O3 measurements Scores are given for each participant and each concentration level (run). Run number order (with nominal concentration) is: 0 (0 nmol/mol), 1 (110 nmol/mol), 2 (90 nmol/mol), 3 (40 nmol/mol), 4 (60 nmol/mol), 5 (25 nmol/mol). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results.

Figure 4: Z’-score evaluations of NO measurements Scores are given for each participant and each tested concentration level (run). Run number order (with nominal concentration) is: 0 (0 nmol/mol), 1 (490 nmol/mol), 2 (380 nmol/mol), 3 (300 mol/mol), 4 (200 nmol/mol), 5 (65 nmol/mol), 6 (25 nmol/mol), 7 (135 nmol/mol), 8 (70 nmol/mol), 9 (40 nmol/mol), 10 (15 nmol/mol). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (redline). They represent the limits for the questionable and unsatisfactory results.

Figure 5: Z’-score evaluations of NO2 measurements Scores are given for each participant and each concentration level (run). Run number order (with nominal concentration) is: 0 (0 nmol/mol), 1 (110 nmol/mol), 2 (100 nmol/mol), 3 (40 nmol/mol), 4 (65 nmol/mol), 5 (25 nmol/mol). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results

### 5.2 En-score

The normalised deviations [13] (En) were calculated according to:

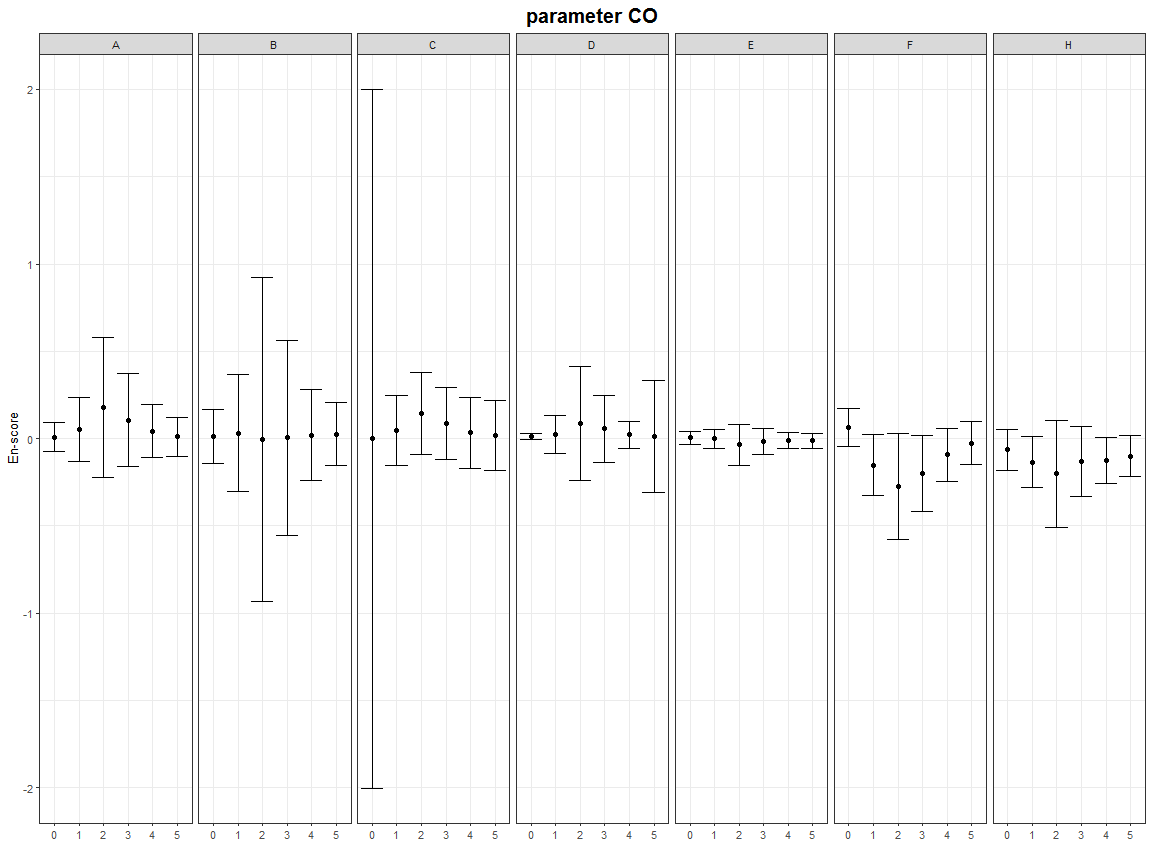
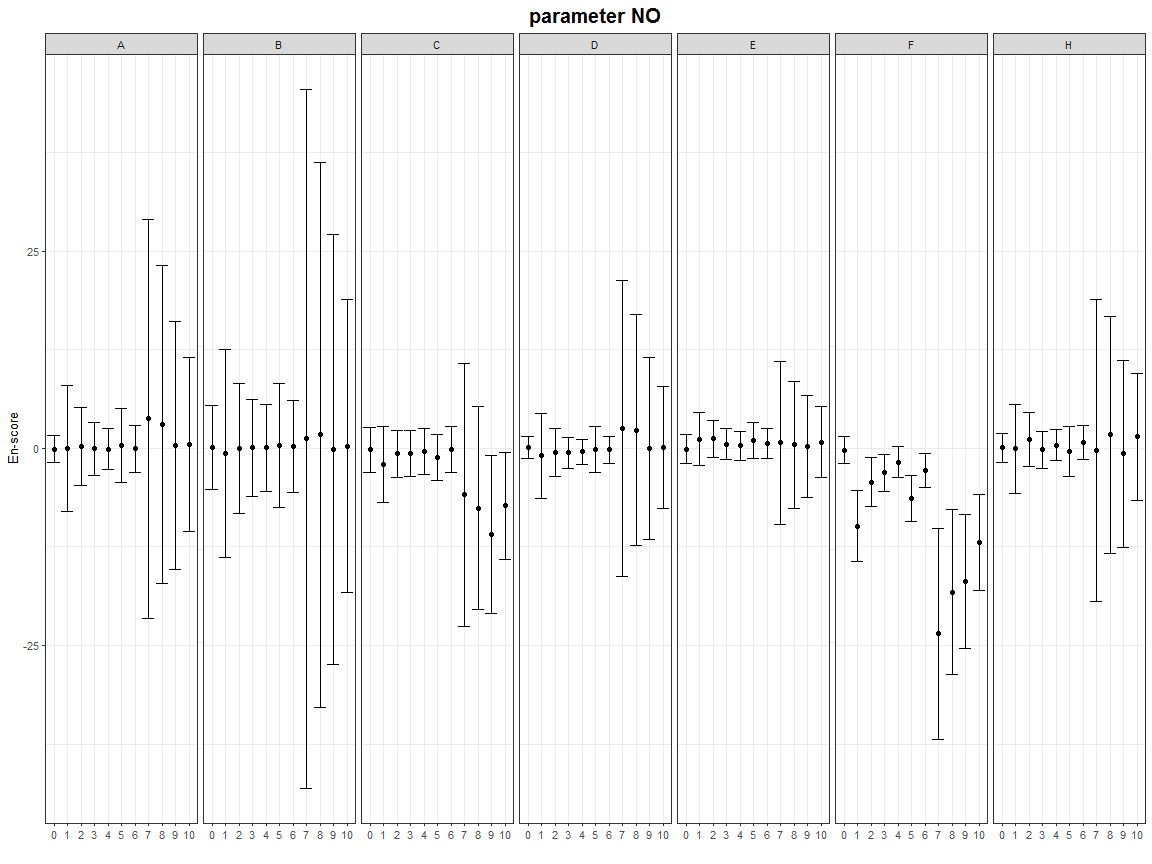
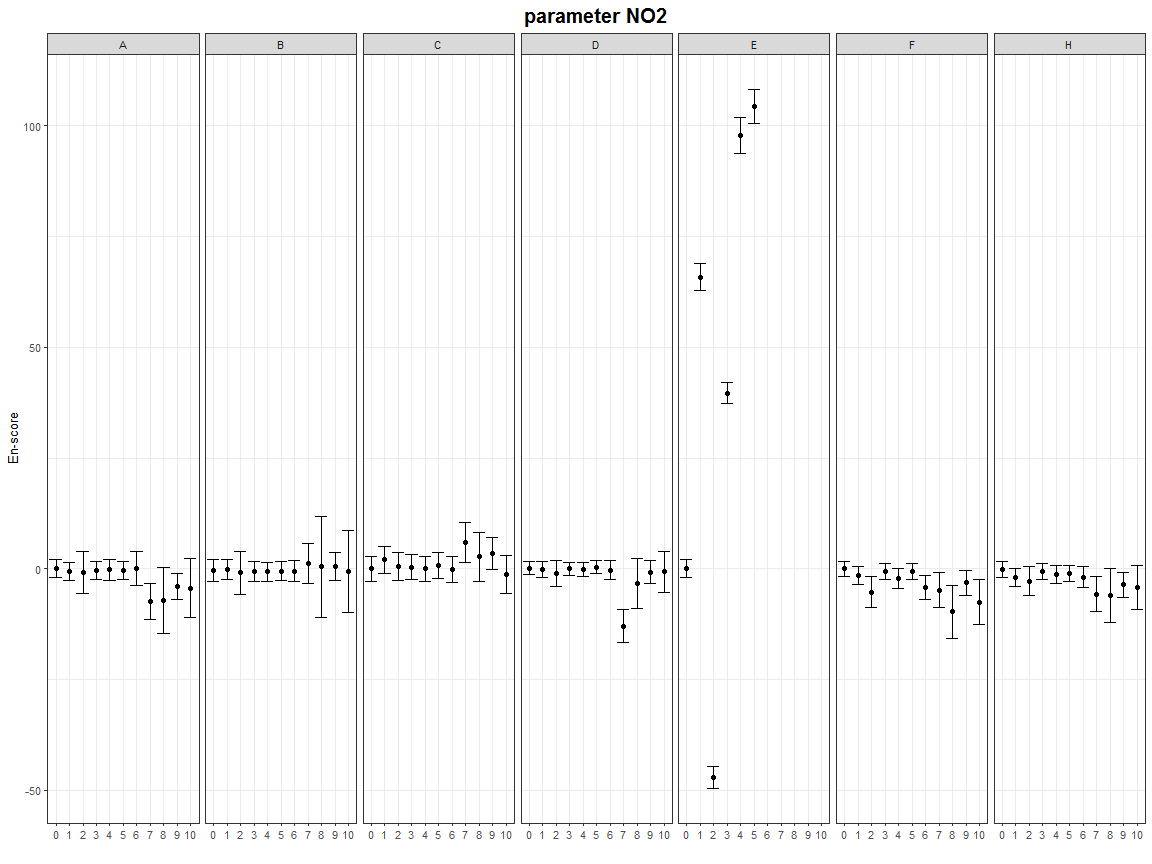
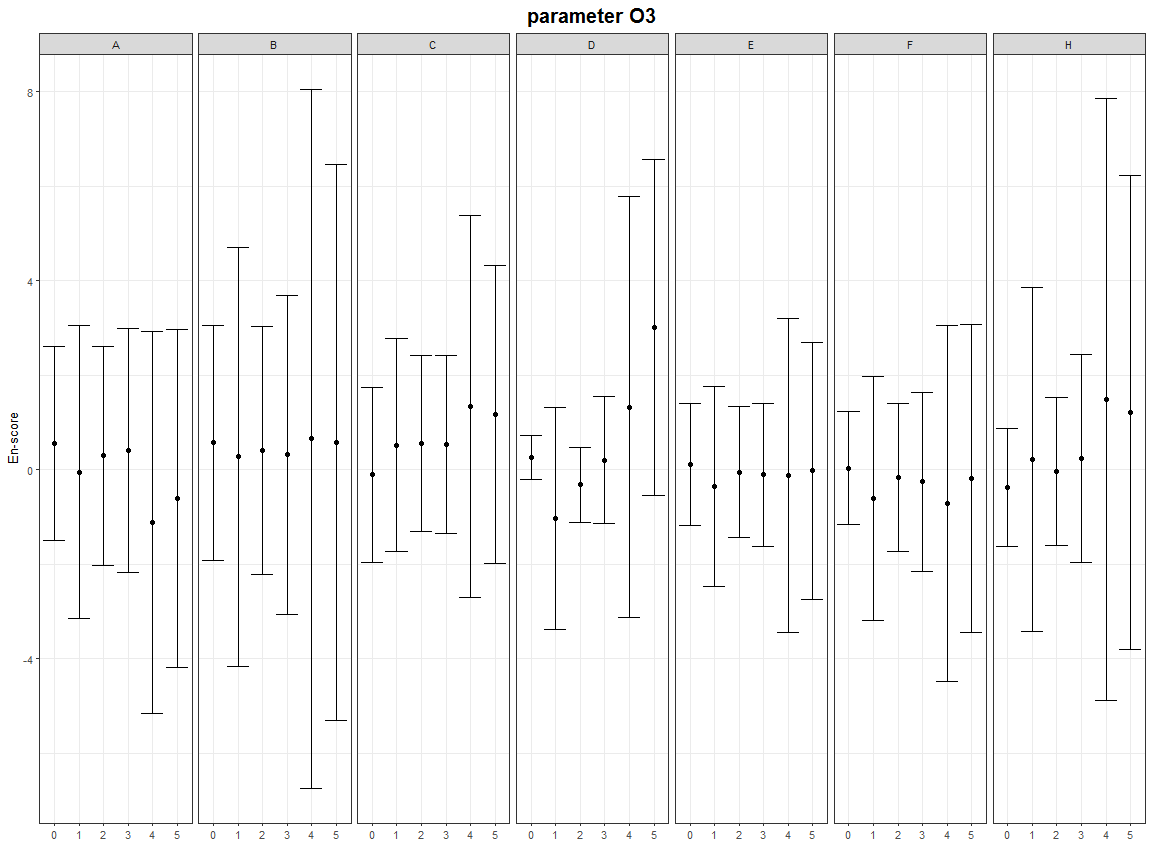
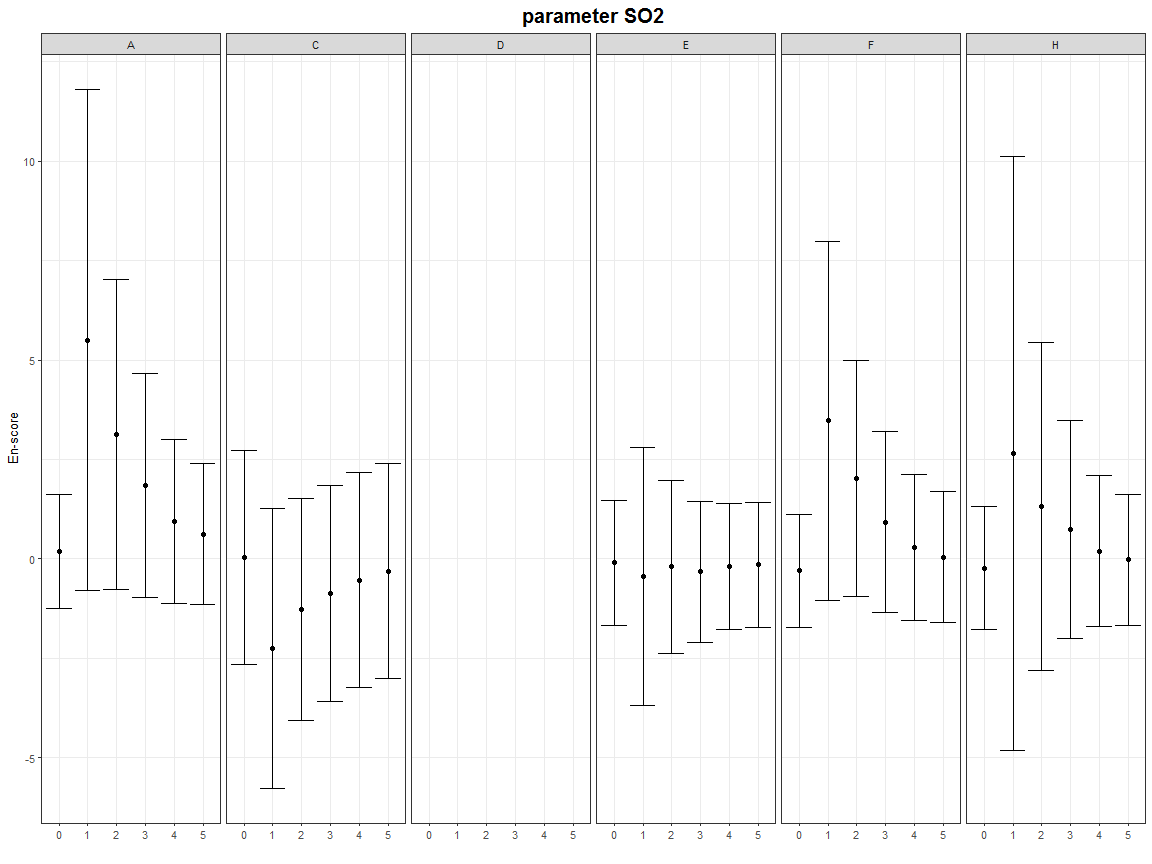
Equation 2

where X is the assigned/reference value with an expanded uncertainty UX and xi is the participant’s average value with an expanded uncertainty UXi. Satisfactory results are the ones for which “absolute En<1”. In Figure 6 to Figure 10 the bias of each participant (xi-X) is plotted and error bars are used to show the value of denominator of equation 2

These plots represent also the En-score evaluations where, considering the En criterion ( ), all results with error bars touching or crossing the x-axis are satisfactory. Reported standard uncertainties (Annex A) that are larger than the “standard deviation for proficiency assessments” (p, Table 4) are considered not fit-for-purpose and are denoted with “\*” in the x-axis of each figure. The En evaluation showed few unsatisfactory results for different parameters and concentrations, as reported in table 5.

Table 5: Unsatisfactory results according to En-score

Figure 6: Bias of participant’s SO2 measurement results. Expanded uncertainty of bias for each run is presented as error bar. The results with error bars touching or crossing the x-axis are satisfactory. For each evaluation the run number (numbers 0 to 5) together with the participants rounded run average (nmol/mol) is given. The ‘\*’ mark indicates reported standard uncertainties bigger than p.

[1] Figure 1: En’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 2: En’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [2] Figure 3: En’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 3: En’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 4: En’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results [1] Figure 5: En’-score evaluations of CO measurements are given for each participant and each tested concentration level (run). The assessment criteria are presented as z’=±2 (blue line) and z’=±3 (red line). They represent the limits for the questionable and unsatisfactory results 

```

Figure 7: Bias of participant’s CO measurement results Expanded uncertainty of bias for each run is presented as error bar. Results with error bars touching or crossing the x-axis are satisfactory. For each evaluation the run number (numbers 0 to 5) together with the participants rounded run average (μmol/mol) is given. The ‘\*’ mark indicates reported standard uncertainties bigger than p.

Figure 8: Bias of participant’s O3 measurement results Expanded uncertainty of bias for each run is presented as error bar. Results with error bars touching or crossing the x-axis are satisfactory. For each evaluation the run number (numbers 0 to 5) together with the participants rounded run average (nmol/mol) is given. The ‘\*’ mark indicates reported standard uncertainties bigger than p.

Figure 9: Bias of participant’s NO measurement results Expanded uncertainty of bias for each run is presented as error bar. Results with error bars touching or crossing the x-axis are satisfactory. For each evaluation the run number (numbers 0 to 10) together with the participants rounded run average (nmol/mol) is given. The ‘\*’ mark indicates reported standard uncertainties bigger than p.

Figure 10: Bias of participant’s NO2 measurement results Expanded uncertainty of bias is presented as error bar for NO2 run numbers 0, 2, 4, 6, 8 and 10 (see Table 3). Results with error bars touching or crossing the x-axis are satisfactory. For each evaluation the run number together with the participants rounded run average (nmol/mol) is given. The ‘\*’ mark indicates reported standard uncertainties bigger than p.

## 6. Performance characteristics of individual laboratories

Individual participants’ biases were evaluated and are presented in chapter 3.2 (Figure 6 - 10). Since the results of NO2 runs 1, 3, 5, 7 and 9 were not treated in proficiency evaluation the bias of these runs are presented in Figure 11.

Figure 11: Bias of participant’s NO2 measurements with error bars representing expanded uncertainty for run numbers 1, 3, 5, 7 and 9. Within these test gas mixtures there is no gas phase titration to produce NO2 (see table 3). For each evaluation the run number together with the participants rounded run average (nmol/mol) is given.

### 6.1 Converter efficiencies of NO2-to-NO for NOX analysers

Since NO and NO2 test gases were produced by gas phase titration it is possible to evaluate the efficiency of the NO2-to-NO converter of each participant’s NOX analyser. The evaluation takes each participant’s NO and NO2 measurements before and after oxidation by O3. However, possible minor instabilities in the preparation of the test gas mixtures were not taken into account. The converter efficiency () is calculated using Equation 3 [4]:

Equation 3

Ideal value for  is 100%. The evaluation of equation 3 for each participant at different concentration levels are given in Table 6.

Table 6: Efficiency of NO2-to-NO converters.

## 7. Discussion

For a general assessment of the quality of each result a decision diagram was developed (Figure 12) that results in seven categories (1 to 7). The general comments for each category are: -1: measurement result is completely satisfactory -2: measurement result is satisfactory (z’-score satisfactory and En-score ok) but the reported uncertainty is too high -3: measured value is satisfactory (z’-score satisfactory) but the reported uncertainty is underestimated (En-score not ok) -4: measurement result is questionable (z’-score questionable) but due to a high reported uncertainty can be considered valid (En-score ok) -5: measurement result is questionable (z’-score questionable and En-score not ok) -6: measurement result is unsatisfactory (z’-score unsatisfactory) but due to a high reported uncertainty can be considered valid (En-score ok) -7: measurement result is unsatisfactory (z’-score unsatisfactory and En-score not ok)

Figure 12: Decision diagram for general assessment of proficiency results.

The results of the ILC were assigned to categories according to the diagram given in Figure 12 and are presented in the following Table 7. Table 7: General assessment of proficiency results. “n.r.” is referring to values not reported. “U n.r.” is referring to Uncertainty values not reported.

## 8. Assigned values

The assigned values of tested concentration levels (run) were derived from ERLAP’s measurements which are calibrated against the certified reference values of CRMs and are traceable to international standards. In this perspective the assigned values are reference values as defined in the ISO 13528 [13]. To foster its reference function ERLAP is participating regularly to key comparisons of the Gas Analysis Working Group within the framework of BIPM’s CCQM. During this ILC ERLAP’s SO2, CO and NO analysers were calibrated according to the methodology described in the ISO 6143 [6]. Reference gas mixtures were produced from the primary reference materials (produced and certified by NMi Van Swinden Laboratorium) by dynamic dilution method using mass flow controllers [8]. All flows were measured with a certified molbloc/molbox1 system. For O3 measurements, the analysers were calibrated using the JRC SRP42 primary standard (constructed by NIST) which has been compared to BIPM primary standard [26]. The photometer absorption cross section uncertainty (1.06%) was included in the uncertainty budget [27], [28]. The reference gas mixture and the calibration experiment evaluation were carried out using two computer applications, the “GUM WORKBENCH” [29] and “B-least” [30] respectively. For extending calibration from the NO to NO2 channel of NOX analyser the GPT test was performed to establish the efficiency of NO2-converter. ERLAP’s measurement results were verified by comparison to the group statistics (x\* and s\*) for every parameter and concentration level of the ILC. These statistics are calculated from participants, applying the robust method described in the Annex C of the ISO 13528 [13]. The verification is taking into account ERLAP’s measurement result (X) and its standard uncertainty (uX) as given in Equation 4 [13]:

Equation 4

Where x\* and s\* represent robust average and robust standard deviation respectively and p is the number of participants. In table 8 all inputs for Equation 4 are given and all ERLAP’s measurement results are confirmed to be valid. As a group evaluation robust average (x*) and robust standard deviation (s*) were calculated (applying the procedure described in Annex C of ISO 13528) for each run, and are presented in the following tables.

Table 8: Verification of assigned values (X).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| run | MinOfunit | yref,j | \_x\* | p |
| \*\*NO \_0\*\* | nmol/mol | 0.12 | 0.066 | 8 |
| \*\*NO \_10\*\* | nmol/mol | 201.2 | 201.4 | 8 |
| \*\*NO \_9\*\* | nmol/mol | 302.4 | 302.1 | 8 |
| \*\*NO \_5\*\* | nmol/mol | 66.06 | 65.94 | 8 |
| \*\*NO \_6\*\* | nmol/mol | 26.77 | 26.78 | 8 |
| \*\*NO \_1\*\* | nmol/mol | 137.7 | 137.2 | 8 |
| \*\*NO \_4\*\* | nmol/mol | 17.01 | 16.89 | 8 |
| \*\*NO \_2\*\* | nmol/mol | 71.87 | 71.87 | 8 |
| \*\*NO \_3\*\* | nmol/mol | 35.82 | 35.6 | 8 |
| \*\*NO \_7\*\* | nmol/mol | 493 | 493.2 | 8 |
| \*\*NO \_8\*\* | nmol/mol | 383.6 | 384.4 | 8 |
| \*\*NO2 \_9\*\* | nmol/mol | 3.71 | 2.655 | 7 |
| \*\*NO2 \_7\*\* | nmol/mol | 7.317 | 3.92 | 7 |
| \*\*NO2 \_8\*\* | nmol/mol | 117.1 | 113.8 | 7 |
| \*\*NO2 \_0\*\* | nmol/mol | 0.01 | 0.028 | 8 |
| \*\*NO2 \_2\*\* | nmol/mol | 66.2 | 64.69 | 8 |
| \*\*NO2 \_3\*\* | nmol/mol | 0.17 | 0.07 | 8 |
| \*\*NO2 \_4\*\* | nmol/mol | 18.89 | 18.6 | 8 |
| \*\*NO2 \_1\*\* | nmol/mol | 0.357 | 0.288 | 8 |
| \*\*NO2 \_5\*\* | nmol/mol | 0.183 | 0.143 | 8 |
| \*\*NO2 \_6\*\* | nmol/mol | 39.44 | 39.08 | 7 |
| \*\*NO2 \_10\*\* | nmol/mol | 104.9 | 102.8 | 7 |
| \*\*O3 \_5\*\* | nmol/mol | 89.29 | 89.78 | 8 |
| \*\*O3 \_2\*\* | nmol/mol | 18.25 | 18.32 | 8 |
| \*\*O3 \_4\*\* | nmol/mol | 114.2 | 114.6 | 8 |
| \*\*O3 \_1\*\* | nmol/mol | 63.6 | 63.51 | 8 |
| \*\*O3 \_0\*\* | nmol/mol | -0.06 | 0.054 | 8 |
| \*\*O3 \_3\*\* | nmol/mol | 33.8 | 33.97 | 8 |
| \*\*SO2 \_4\*\* | nmol/mol | 18.13 | 18.21 | 6 |
| \*\*SO2 \_0\*\* | nmol/mol | 0.12 | 0.046 | 6 |
| \*\*SO2 \_5\*\* | nmol/mol | 10.08 | 10.05 | 6 |
| \*\*SO2 \_1\*\* | nmol/mol | 122.1 | 123.6 | 6 |
| \*\*SO2 \_3\*\* | nmol/mol | 37.92 | 38.3 | 6 |
| \*\*SO2 \_2\*\* | nmol/mol | 64.21 | 65.03 | 6 |
| \*\*CO \_4\*\* | Âµmol/mol | 2.028 | 2.032 | 8 |
| \*\*CO \_1\*\* | Âµmol/mol | 2.839 | 2.847 | 8 |
| \*\*CO \_3\*\* | Âµmol/mol | 5.083 | 5.078 | 8 |
| \*\*CO \_2\*\* | Âµmol/mol | 8.649 | 8.637 | 8 |
| \*\*CO \_0\*\* | Âµmol/mol | 0.002 | 0.008 | 8 |
| \*\*CO \_5\*\* | Âµmol/mol | 0.909 | 0.909 | 8 |

By comparison to the robust averages (x*) with taking into account the standard uncertainties of assigned values (uX), and robust standard deviations (s*) as denoted by Equation 4. The homogeneity of test gas was evaluated from measurements at the beginning and end of the distribution line. The relative differences between beginning and end measurements are calculated.

Equation 5

The upper and lower limits of bias due to homogeneity were evaluated to be smaller than 0.5% which constitutes the relative standard uncertainty of 0.3% of each concentration level assuming a rectangular distribution of the biasis. The standard uncertainties of assigned/reference values (uX) were calculated with Equation 5 and used in the proficiency evaluations of chapter 3. All calculations about the homogeneity testing data are retained by ERLAP, they are not published in this ILC report but are available on request.

## 9. Conclusions

The proficiency evaluation scheme has provided an assessment of the participants measured values and their evaluated uncertainties. In terms of the criteria imposed by the European Directive (p) 95.6% of the results reported during this ILC (see Table 7) by AQUILA laboratories fall into category ‘1’ and are satisfactory both in terms of measured values and evaluated uncertainties. Among the remaining all results presented satisfactory measured values, but the evaluated uncertainties were either too high, category ‘2’ (1.3%), or too small, category ‘3’ (3.1%). No values were found questionable (category 5) or not satisfactory for both value and uncertainty (category 6 and 7).

Table 9: Flags summary

As in previous ILC, the adopted criteria for high concentrations were the standard deviations for proficiency assessment, deriving from the European Standards’ uncertainty requirements.  
The reproducibility standard deviation obtained at this (Annex B) and previous ILC *[20], [21], [22], [23], [24], [25], [33], [34], [35], [36], [37], [38], [39], [40], [41], [43], [44], [45], [46], [47] and [48]* is comparable to the mentioned criteria. On the other hand, the uncertainty criteria for zero levels were those set in AQUILA’s position paper [12]. In this exercise 100% of the results in the z’-score evaluations were satisfactory.

Table 10: Z’-score summary

Comparability of results among AQUILA participants at the highest concentration level is acceptable for all pollutant measurements. The relative reproducibility limits, at the highest studied concentration levels, are 4.6% for SO2, 4.4% for CO, 2.6% for O3, for NO 3.2% and for NO2 7.5% all within the objective derived from criteria imposed by the European Commission (p see Table 4). During this ILC the performance of all NRL was generally satisfactory. No values were identified as outliers.

## 10. References

## List of abbreviations

AQUILA Network of National Reference Laboratories for Air Quality CEN European Committee for Standardization CO Carbon monoxide CRM Certified Reference Material DQO Data Quality Objective ERLAP European Reference Laboratory for Air Pollution EC European Commission GPT Gas Phase Titration ILC Inter-Laboratory Comparison Exercise ISO International Organization for Standardization JRC Joint Research Centre NO Nitrogen monoxide NO2 Nitrogen dioxide NOX The oxides of nitrogen, the sum of NO and NO2 NRL National Reference Laboratory O3 Ozone SO2 Sulphur dioxide VDI Verein Deutscher Ingenieure WHO-CC World Health Organization Collaborating Centre for Air Quality Management and Air Pollution Control, Berlin

## Mathematical Symbols

 converter efficiency (EN 14211) En En–score statistic (ISO 13528) r repeatability limit (ISO 5725) R reproducibility limit (ISO 5725) σp standard deviation for proficiency assessment (ISO 13528) x\* robust average (Annex C ISO 13528) s\* robust standard deviation (Annex C ISO 13528) sr repeatability standard deviation (ISO 5725) sR reproducibility standard deviation (ISO 5725) UX’ expanded uncertainty of the assigned/reference value (ISO 13528) Uxi expanded uncertainty of the participant’s value uX’ standard uncertainty of the assigned/reference value (ISO 13528) X assigned/reference value (ISO 13528) xi average of three values reported by the participant i (for particular parameter and concentration level) (ISO 5725) xi,j j-the reported value of participant i (for particular parameter and concentration level) (ISO 5725) z’ z’-score statistic (ISO 13528)

## List of figures

List of tables

## Annex A: The results of the ILC

In this annex are reported participant’s results, presented both in tables and graphs. For all mixture concentration generated (run), participants were asked to report 3 results representing 30 minutes measurement each (xi). In this annex are presented the reported data and their uncertainty u(xi) and U(xi) expressed in mol/mol units. For all the runs except concentration levels 0, also average (xi) and standard deviation (si) of each participant are presented. The assigned value is indicated on the graphs with the red line and the individual laboratories expanded uncertainties (Uxi) are indicated with error bars.

## Reported values for SO2

Table 11: Reported values for SO2 run 0.

Figure 13: Reported values for SO2 run 0.

Table 12: Reported values for SO2 run 1

Figure 14: Reported values for SO2 run 1.

Table 13: Reported values for SO2 run 2.

Figure 15: Reported values for SO2 run 2.

Table 14: Reported values for SO2 run 3.

Figure 16: Reported values for SO2 run 3.

Table 15: Reported values for SO2 run 4.

Figure 17: Reported values for SO2 run 4.

Table 16: Reported values for SO2 run 5.

Figure 18: Reported values for SO2 run 5.

## Reported values for CO

Table 17: Reported values for CO run 0.

Figure 19: Reported values for CO run 0.

Table 18: Reported values for CO run 1.

Figure 20: Reported values for CO run 1.

Table 19: Reported values for CO run 2

Figure 21: Reported values for CO run 2.

Table 20: Reported values for CO run 3.

Figure 22: Reported values for CO run 3.

Table 21: Reported values for CO run 4.

Figure 23: Reported values for CO run 4.

Table 22: Reported values for CO run 5.

Figure 24: Reported values for CO run 5.

## Reported values for O3

Table 23: Reported values for O3 run 0.

Figure 25: Reported values for O3 run 0.

Table 24: Reported values for O3 run 1

Figure 26: Reported values for O3 run 1.

Table 25: Reported values for O3 run 2.

Figure 27: Reported values for O3 run 2.

Table 26: Reported values for O3 run 3.

Figure 28: Reported values for O3 run 3.

Table 27: Reported values for O3 run 4.

Figure 29: Reported values for O3 run 4.

Table 28: Reported values for O3 run 5.

Figure 30: Reported values for O3 run 5.

## Reported values for NO

Table 29: Reported values for NO run 0.

Figure 31: Reported values for NO run 0.

Table 30: Reported values for NO run 1.

Figure 32: Reported values for NO run 1.

Table 31: Reported values for NO run 2.

Figure 33: Reported values for NO run 2.

Table 32: Reported values for NO run 3.

Figure 34: Reported values for NO run 3.

Table 33: Reported values for NO run 4.

Figure 35: Reported values for NO run 4.

Table 34: Reported values for NO run 5.

Figure 36: Reported values for NO run 5.

Table 35: Reported values for NO run 6.

Figure 37: Reported values for NO run 6.

Table 36: Reported values for NO run 7.

Figure 38: Reported values for NO run 7.

Table 37: Reported values for NO run 8.

Figure 39: Reported values for NO run 8.

Table 38: Reported values for NO run 9.

Figure 40: Reported values for NO run 9.

Table 39: Reported values for NO run 10.

Figure 41: Reported values for NO run 10.

## Reported values for NO2

Table 40: Reported values for NO2 run 0.

Figure 42: Reported values for NO2 run 0.

Table 41: Reported values for NO2 run 2.

Figure 43: Reported values for NO2 run 2.

Table 42: Reported values for NO2 run 4.

Figure 44: Reported values for NO2 run 4.

Table 43: Reported values for NO2 run 6.

Figure 45: Reported values for NO2 run 6.

Table 44: Reported values for NO2 run 8.

Figure 46: Reported values for NO2 run 8.

Table 45: Reported values for NO2 run 10.

Figure 47: Reported values for NO2 run 10.

## Annex B: precision of standardised measurement methods

For the main purpose of monitoring trends between different ILC undertaken by ERLAP, the precision of standardized SO2, CO, O3 and NOX measurement methods [2], [3], [4] and [5] as implemented by NRLs, was evaluated. Applied methodology is described in ISO 5725-1, 5725-2 and 5725-6 [14], [15] and [16]. The precision experiment has involved a total of nine laboratories, the actual number of labs (pj) is reported in Table 46. Six concentration levels (for run 0 only one value is requested so repeatability cannot be evaluated) were tested for O3, CO, SO2 and NO2, and eleven for NO. Outlier tests were performed and results are reported in Annex D. The repeatability standard deviation (sr) was calculated in accordance with ISO 5725-6 as the square root of average within-laboratory variance. The repeatability limit (r) is calculated using Equation 6 [16]. It represents the biggest difference between two test results found on an identical test gas by one laboratory using the same apparatus within the shortest feasible time interval that should not be exceeded on average more than once in 20 cases in the normal and correct operation of method.

Equation 6

The reproducibility standard deviation (sR) was calculated in accordance with ISO 5725-6 as the square root of sum of repeatability and between-laboratory variance. The reproducibility limit (R) is calculated using Equation 7 [16]. It represents the biggest difference between two measurements on an identical test gas reported by two laboratories, which should not occur on average more than once in 20 cases in the normal and correct operation of method.

Equation 7

The repeatability standard deviation was evaluated with (pj \*(3-1)) degrees of freedom () and reproducibility standard deviation with (pj-1) degrees of freedom. The corresponding critical range student factors (t,) are reported in Table 46.

Table 46: Critical values of t used in the repeatability (r) and reproducibility (R) evaluation.

The repeatability (r) and reproducibility (R) limits of measurement methods are presented from Table 47 to Table 51 and from Figure 48 to Figure 52. Also reported is the ‘reproducibility from common criteria (R (from p))’ calculated by substituting sR in Equation 7 with a ‘standard deviation for proficiency assessment’ (see Table 4). Comparison between R and R (from p) serves to indicate that p is realistic ([13] under 6.3.1) or from the other point of view, that the general methodology implemented by NRLs is appropriate for p.

Table 47: The R and r of SO2 standard measurement method.

Figure 48: The R and r of SO2 standard measurement method as a function of concentration.

Table 48: The R and r of CO standard measurement method.

Figure 49: The R and r of CO standard measurement method as a function of concentration.

Table 49: The R and r of O3 standard measurement method.

Figure 50: The R and r of O3 standard measurement method as a function of concentration.

Table 50: The R and r of NO standard measurement method.

Figure 51: The R and r of NO standard measurement method as a function of concentration.

Table 51: The R and r of NO2 standard measurement method.

Figure 52: The R and r of NO2 standard measurement method as a function of concentration.

## Annex C: scrutiny of results for consistency and outlier test

The precision evaluation (Annex C) focuses on data that are as much as possible the reflection of every day work of NRLs and thus represents the comparability of participant’s standard operating procedures. For that reason, a procedure for the detection of exceptional errors (error during typing, slip in performing the measurement or the calculation, wrong averaging interval, malfunction of instrumentation, etc.) was applied. In this procedure were carried out tests for data consistency and statistical outliers as described in ISO 5725-2. Laboratories showing some form of statistical inconsistency were requested to investigate the cause of discrepancies. Laboratories were allowed to correct their results in case of identification of exceptional errors. Subsequently, data were considered definitive and z’-scores calculation was performed to estimate outliers. Statistical outliers obtained at this stage are not considered as extraordinary errors but due to significant difference in participant’s standard operating procedure. During this ILC, no statistical outlier was identified. The precision of standardised measurement methods reported in Annex B are calculated using the database without outliers. According to z’-score calculation, results between |2| and |3| are considered stragglers and they deserve a specific check.  
During this ILC, neither statistical straggler nor outliers were identified.

## Annex D: Confidentiality

Results of the ILC are published according to the agreements included in the document AQUILA-N37 approved by all NRL. In order to ensure confidentiality of the laboratories information, ERLAP guarantees the submitted data as follows:

* Any administrative information provided by the laboratory is confidential and cannot be communicated to a third party.
* Access to ERLAP facilities is allowed only to members of the Unit C5 and authorized persons (cleaning staff, maintenance staff, safety and security staff etc.)
* Confidential passwords to access the web application for data submission are sent once the registration to ILC is completed. Confidential passwords allow access to the WEB interface and to on-line questionnaire. Passwords are valid until the ILC is closed. Laboratories can change their password online.
* The form LAB-REC-2000 (Confidentiality involvement form) is asked to be signed by the participants during their first participation to an ILC organized by ERLAP.

## Annex E: Accreditation certificates

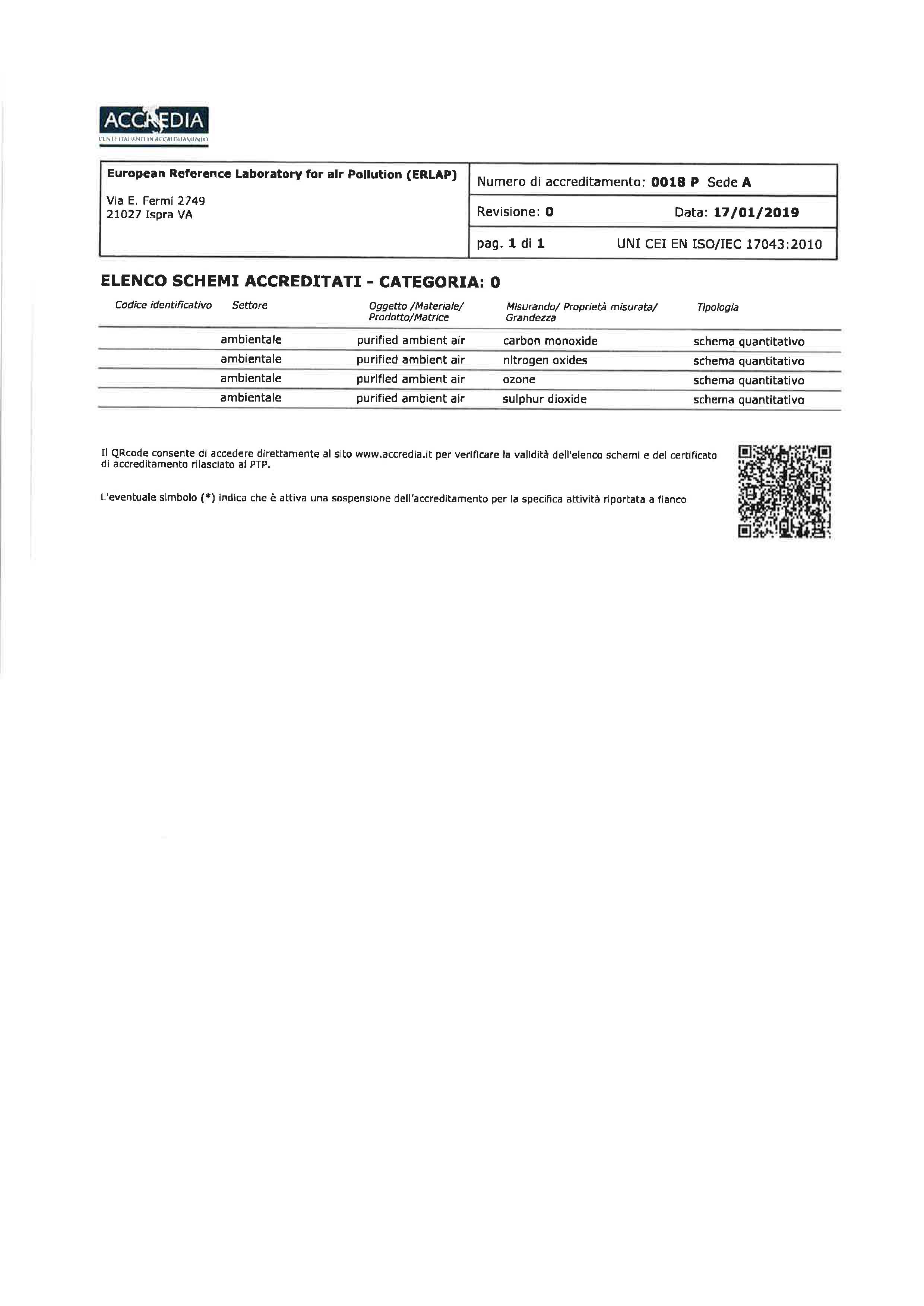
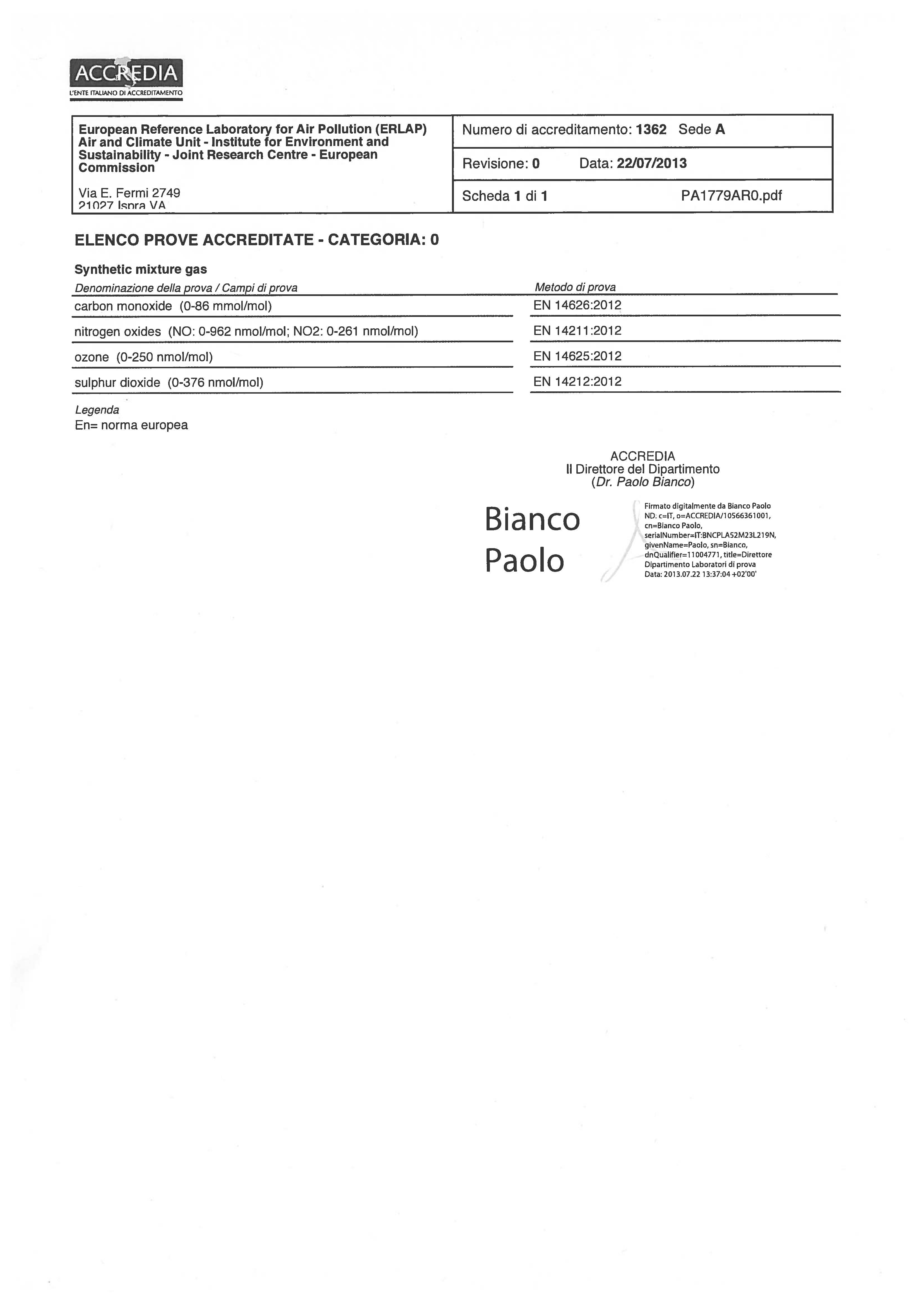
Figure 53: ERLAP ISO 17043 accreditation certificate  

Figure 54: ERLAP ISO 17025 accreditation certificate. 



certificate

END OF REPORT

## Bibliography