RESULTS OF THE 10TH CIRCUIT INTERLABORATORY FOR DIOXINS (CIND)

Raccanelli S[1], Libralato S[2]

[1] Consorzio I.N.C.A., VEGA-Edificio Cygnus, Via delle Industrie 21/8, 30175 Marghera (VE), Italy; [2] Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Borgo Grotta Gigante 42/c, 34010, Sgonico (TS), Italy.

Introduction

In the 2000 the Consorzio Interuniversitario "La Chimica per l'Ambiente" (Consorzio INCA), prompted the 1st Circuit INter-laboratory for Dioxins (CIND), i.e. the first Italian Intercalibration study for giving to all Italian laboratories the possibility for comparing their analytical performances in measuring dioxins for free. Since then the Circuit has broaden the perspective of the study by inviting also foreign laboratories and by including in the list of analytes, other than dioxins and furans (PCDD/F), also PCBs dioxin-like and PAHs. The number of participants increased steadily during the past editions and now CIND represents an important international intercalibration study that is going to be accredited for UNI CEI EN ISO/IEC 17043:2010.

The 10th CIND edition was initiated in October 2010 by sending (upon request) two samples to more than 50 laboratories, asking them to provide analyses of PCDD/Fs, PCBs and PAHs in 3 replicates for each sample. In March 2011 the analysis received by 48 laboratories were statistically analysed for data quality control (QC) and evaluation of performance of laboratories. Results were reported in a specifically organized congress open to all participating laboratories (10th CIND; Auditorium Santa Margherita, Venice, Italy; 1st April 2011). Here we report QC method based on no-parametric statistical indexes (median and quartiles) that has proven to efficient for detecting outliers and extremes. Moreover, the main results of the 10th CIND edition are reported in terms of both accuracy and precision performances of participating laboratories.

Materials and methods

Samples and analyses required.

The first sample consists in an environmental matrix (sediment) collected in the Lagoon of Venice. Large debris (>1cm) were separated by hand and homogeneized "in situ". The second sample consisted in fly ash obtained from a steel plant. Three sets of 50 kg each for each of the two samples were dried at low temperature (approximately 40°C), grinded and sieved through a 100µm sieve. The material obtained, 20kg per set, was homogenized again and divided into four parts, which were analyzed twice, in order to ascertain their homogeneity. After this test, the samples were stored in amber glass containers. The two samples have been sent (October 2010) to the laboratories asking to have measurements of PCDD/F (17 congeners), PCB (12 congeners) and PAH (7 congeners) in three replicates. Laboratories were also asked to report total TEQ for PCDD/F, PCB, PAH, and their sums (PCDD/F+PCB; PCDD/F+PCB +PAH) and the sum of PAH (Lower and Upper bound). Each laboratory was assigned an identificative number (LAB#) used for assuring an objective data treatment and the privacy for all participants. A portion of each of the materials sent is opportunely stored by Consorzio Interuniversitario "la Chimica per l'Ambiente" (Consorzio INCA) for its future use as reference material (upon request by participants). The form for transmitting results and the analysis of data received have been standardized thanks to a collaboration between CIND and INTERCAL [1] intercalibration studies.

Statistical treatment of data.

For each congener (j) were calculated across laboratories: average concentration ($\bar{x}_j$), standard deviation (s_j), coefficient of variance (RSD%j= $s_j / \bar{x}_j$ %), Maximum and Minimum value observed (MAXj and MINj, respectively), the median (Mj) and the 1st and 3rd quartiles (Q25j and Q75j respectively). The statistical indexes were calculated, for each sample and congener, pooling together laboratories (i) and replicates (k) values as in the following for mean and standard deviation:

$$\bar{x}_j = \left(\sum_{i=1}^{n} \sum_{k=1}^{r} x_{i,k,j} \right) / \left(\sum_{i=1}^{n} \sum_{k=1}^{r} 1 \right) \qquad \text{(Eq. 1)}$$

$$s_j = \sqrt{\frac{\sum_{i=1}^{n}\sum_{k=1}^{r}\left(x_{i,k,j} - \bar{x}_j\right)^2}{\left(\sum_{i=1}^{n}\sum_{k=1}^{r}1\right)-1}} \quad \text{(Eq. 2)}$$

where $x_{i,k,j}$ are the values reported, with i = laboratory (i=1...48); k= replicates (k=1...3); j = congener. Moreover, defining $P(x_{i,k,j} < X)$ the cumulative probability of having values smaller than X, the median M_j and quartiles $Q25_j$ and $Q75_j$ are defined as:

Median: $P(x_{i,k,j} < M_j) \leq 0.5$ and $P(x_{i,k,j} \leq M_j) \geq 0.5$; (Eq. 3)

1st quartile: $P(x_{i,k,j} < Q25i) \leq 0.25$ and $P(x_{i,k,j} \leq Q25i) \geq 0.25$ (Eq. 4)

3rd quartile: $P(x_{i,k,j} < Q75i) \leq 0.75$ and $P(x_{i,k,j} \leq Q75i) \geq 0.75$ (Eq. 5)

These non-parametric indexes (median and quartiles) were used to indentify extremes and outliers. In particular, the range between the two quartiles for each congener was defined as:

$$U_j = Q75_j - Q25_j \quad \text{(Eq. 5)}$$

and extremes and outliers were identified as the values that are outside the range defined as[2]:

extremes: $Q25_j - 3 \cdot U_j > x_{i,k,j}$ and $Q75_j + 3 \cdot U_j < x_{i,k,j}$ (Eq. 6)

outliers: $Q25_j - 1.5 \cdot U_j > x_{i,k,j}$ and $Q75_j + 1.5 \cdot U_j < x_{i,k,j}$ (Eq. 7)

This method for defining extremes and outliers based on non parametric statistical indexes is much more efficient than parametric methods in cases of QC of largely scattered data that can be the result of data misreported in terms of unit of measure (as in previous CIND editions)[3]. Extremes and outliers identified were excluded from the data set for calculations of the "true value" and statistical indexes calculated across laboratories.

Accuracy and precision performances of laboratories.

The performance of each participant laboratory in term of accuracy (closeness of laboratory data to true value) was estimated by means of the z-scores coefficients, z_i, calculated as:

$$z_{i,k,j} = \frac{x_{i,k,j} - \bar{x}_j}{s_j} \quad \text{(Eq. 8)}$$

Z-scores assume that the "true value" is the average (and standard deviation) calculated for all laboratories and all replicates after elimination of extremes and outliers. The performance of each participant laboratory in term of precision (closeness between laboratory replicates) for each congener was estimated by what we called r-score:

$$r_{i,j} = \frac{\max(x_{i,1,j}, x_{i,2,j}, x_{i,3,j}) - \min(x_{i,1,j}, x_{i,2,j}, x_{i,3,j})}{\bar{x}_{i,j}} \quad \text{(Eq. 9)}$$

The r-score ($r_{i,j}$) represent a measure of dispersion of values for congener (j) provided by the laboratory (i) with respect to the average of the values provided. The z-scores and r-scores were calculated for all concentrations provided by each laboratory and for TEQ values for PCDD/F, PCB, PAH and their sum. The TEQ of PAH were computed according to literature[4].

Results and discussion:

Among all laboratories that received the sample, 47 and 43 laboratories reported results for SEDIMENT and FLY ASH, respectively, by the set deadline. Considering the 36 congeners to analyse and the 3 replicates per laboratory a total of 5076 (36x3x47) and 4644 (36x3x43) data were expected for SEDIMENT and FLY ASH, respectively. However, some laboratories did not report results in three replicates and/or did not measure some congeners. A total of 884 and 747 data were not analysed (NA), thus data received were 4192 and 3897 for SEDIMENT and FLY ASH samples, respectively. Raw data included not detected (ND) values: 730 and 33 for SEDIMENT and FLY ASH, respectively. The form opportunely prepared for reporting results forced the laboratories to use uniform specification of detection limit (e.g. " < 0.0001"), which was used in the analyses.

The statistical treatment led to the identification of 161 extremes in the sediment dataset (45, 114 and 2 for PCDD/F, PCB and PAH, respectively) and 220 extremes for the fly ash sample (124, 91, 5 extreme values for PCDD/F, PCB and PAH, respectively). Outliers were 104 in the SEDIMENT sample (50, 38 and 16 for PCDD/F, PCB and PAH, respectively) and 181 in the FLY ASH sample (111, 55 and 15 for PCDD/F, PCB and PAH, respectively). Figure 1 reports the percentages of valid, NA, ND, outliers and extremes of data after QC.

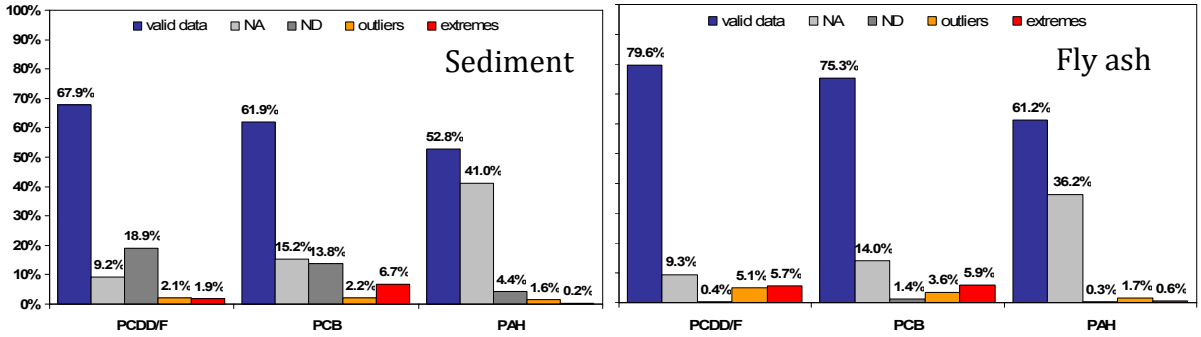

Figure 1. Representation of the proportions of laboratories' data after identification of the ouliers and extremes for each class of compounds. Sediment and fly ash left and right panel, respectively. Data are reported in proportion to expected data, whereas absolute values are reported in the text.

The figure 2 synthetically allow to verify the efficacy of the identification and exclusion of the extremes and outliers for all congeners and TEQ values (reported in x-axis) by reporting the relative standard deviation (RSD%, y-axis) before and after the data treatment. Original data provided by the laboratories showed extremely high RSD% (greater than 100%) especially for PCBs for both samples, 1,2,3,7,8,9 HxCDF for ash sample, and 2,3,7,8 TeCDD and 1,2,3,7,8 PeCDD for sediment sample. The effect of the treatment for identification of outliers and extreme is evident. In particular RSD% of final treated data is far below 100% and close to 20% for all congeners in the ASH sample, with maximum RSD% of 72% and 55%, for 1,2,3,7,8,9 HxCDF and PCB-123, respectively. Although a very large reduction of dispersion of data is recognizable in sediment treated sample (upper panel Figure 2), RSD% for PCB123 and 2,3,7,8 TeCDD remained high (135% and 101% respectively) because of the very low concentrations, close to detection limit.

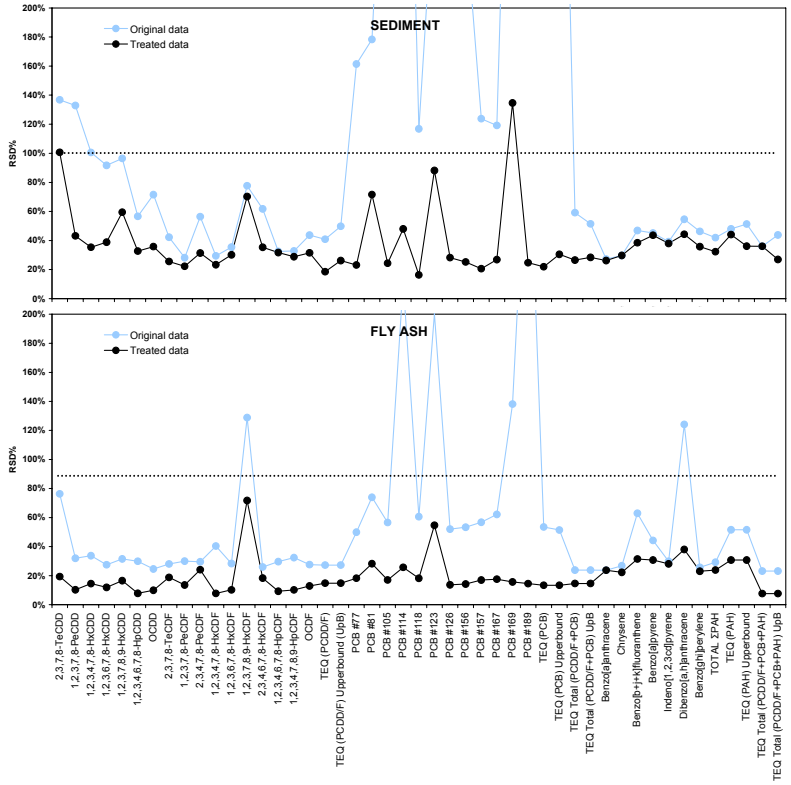

Figure 2. Schematic representation of coefficient of variation for all congeners and derived concentrations (such as TEQ) over all laboratories for sediment and fly ash (upper and lower panel respectively). RSD% are estimated both for the original dataset and that excluded from identified outliers and extremes values. The figures provide basis for evidencing the accuracy related to the evaluation of the true value for each congener.

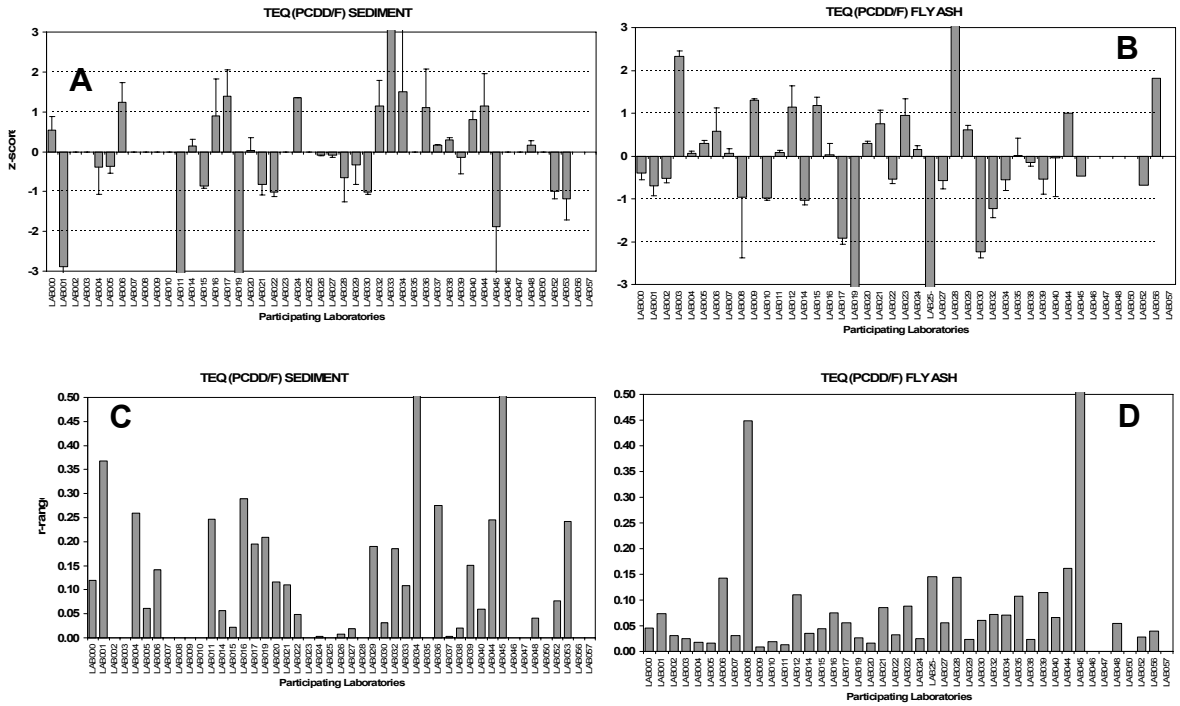

Figure 3.
Results obtained in terms of precision (z scores panels A,B) and accuracy (r-scores, panels C,D), for the PCDD/F TEQ taken as an example. Sediment sample in left panels (A,C) and ash fly in right panes (B,D). Values consist in the average of the three replicates; ouliers and extremes were excluded.

Valid data were used to calculate also z-scores and r-scores for each congener and TEQ value and for each laboratory. Figure 3 reports z-scores (average of three replicates) and r-score for PCDD/F toxicity as an example. These values provide indication of average accuracy and precision of each laboratory: for instance Lab 19 showed intermediate precision for sediment sample (r-range around 0.2) but very low accuracy being z-score minor than -3. Lab030 instead showed very good performances in evaluating sediment sample with z-score around -1 and r-range equal to 0.03. These graphs allow to evaluate in a complete form the performance of laboratories and to detect ones with very low precision and low accuracy in both sediment and ash fly samples (such as Lab 000, Lab 005, Lab 020, Lab048), laboratories with good precision but low accuracy (e.g. Lab 030 for fly ash), and Laboratories with problems both in accuracy and precision (e.g. Lab 045 for sediment). It can be noted that a part from Lab008 and Lab045 the precision of laboratories in evaluating the ash sample was generally very good, whereas this is not true for the sediment sample. This reveals the more difficulties in evaluating the sediment sample due to its low concentration.

Acknowledgements:
LabService Analytica Srl is acknowledged for the support provided. In particular the authors wish to thank Dr. Ivano Battaglia and Dr. Simona Manganelli, without whom these interlaboratory studies (CIND) could not have been held with yearly frequency. Prof. B. Van Bavel is acknowledged for his collaboration in setting common basis for CIND and INTERCAL intercalibration circuits.

References:
1. van Bavel B., Fifth Round of the International Intercalibration Study, Umea University, 2000.
2. StatSoft, Inc. (2005). STATISTICA (data analysis software system), version 7.1. www.statsoft.com.
3. Raccanelli S., Petrizzo A., Favotto M. and Pastres R. (2007). *Organoh. Comp*. 69: 982-985.
4. Klimm C., Hofmaier A.M., Schramm K.W., Kettrup A. (1999). *Organoh. Comp*. 40: 39-42