PROFICIENCY TESTING USING NATURAL MATRIXES AND NON-PARAMETRIC STATISTICS: EXPERIENCES OF TWO YEARS OF THE INTERCIND STUDY.

Libralato S[1], van Bavel B [2]* and Raccanelli S[3]

[1]Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Borgo Grotta Gigante 42/c, 34010, Sgonico (TS), Italy; [2]MTM Research Center, School of Science and Technology, Örebro University - SE 701 82 Örebro, Sweden ; [3] Chimico etico ambientale, via S. Cecilia 40, 30174 Venezia, Italy

Introduction

The multiannual experiences of Circuit Inter-laboratory for Dioxins (CIND)[1] and INTERCAL[2], highlighted the importance on objective methods for quality control and quality assurance (QC/QA) of laboratories data in laboratory intercalibration and intercomparison studies. In 2012 the two intercalibration studies joined together to set up a unified international proficiency test under the name of InterCinD.

In its first two editions (2012-13; 2013-14), InterCinD has broadened the perspective of the intercalibration studies and proficiency test by including several POPs and analytes, other than dioxins and furans (PCDD/F), such as PCBs dioxin-like, PCB-ICES 6, PAH, PBDE and further included heavy metals (HM) in the second edition, special for InterCinD is i) the use of "natural matrixes" (environmental, industrial and food matrixes) as sample other than standards and solutions; ii) the request of results, for each lab, in three replicates permitting evaluation of the lab's precision; iii) the use of an input form opportunely built in a spreadsheet helping data introduction; iv) treatment of data using non parametric statistical analysis[3] to identify the "consensus value" as the potential true concentration of the analyte in the sample. Moreover, InterCinD which has a significant number of participating laboratories worldwide and represents an important international intercalibration study that is going to be accredited for UNI CEI EN ISO/IEC 17043:2010.

The scope of this work is to present the results of the 1st and 2nd InterCinD proficiency tests, giving particular emphasis at 1) the statistical analysis used that has proven to efficient for detecting outliers and extremes, 2) the main results of the two editions reported in terms of both accuracy and precision performances of participating laboratories.

Materials and methods

Samples

The reference materials used in the InterCIND are all NATURAL MATRIXES prepared and characterized in agreement with the international guidelines on the production of reference materials (ISO Guide 34:2009). These materials have the needed characteristics of homogeneity and stability that allow their use in the interlaboratory circuits. For the InterCIND editions were given 3 different samples of "natural" matrix, plus standards and solutions.

Table 1: Environmental, industrial, food matrixes and solutions used as samples in the first two InteerCinD.

1st InterCinD	2nd InterCinD
Natural lagoon Sediment (SEDIMENT);	Sediment of the Venice Lagoon (SEDIMENT)
Dust from a steel plant (FLY ASH);	Ash from an incinerator (FLY ASH);
Fish (Silurus glanis) (FISH);	Food additive (FEED)
Solution CC (dioxins and furans (PCDD/F), PCBs dioxin-like)	Solutions Mix PCDD/F
Solution DD (PBDE)	Solution mix PBDE
Solution EE (PCBs ICES-6)	Solution mix PCB
Solution FF (PAHs)	Solution ICUS3748
Solution EDF 5008 (PCDD/F)	
Solution EC 4987 (PCBs dioxin-like)	

The samples indicated as "sediment" are two different environmental matrix (sediment) collected in the Lagoon of Venice. Large debris (>1cm) were separated by hand and homogenized "in situ". Subsequently, three sets of 50 kg each of sediment 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 analysed twice, in order to ascertain their homogeneity. After this test, the sample was stored in amber glass containers. The samples indicated as "fly ash" consisted in dust obtained from a steel plant (year 2013) and incinerator ash (2014). Each sample was grinded and sieved through a 100µm sieve. The material obtained 20kg per set, was homogenized again and divided into four parts, which were analysed twice, in order to ascertain their homogeneity. After this test, the sample was stored in amber glass containers. The sample, indicated as "fish" in the following, consisted in a large individual of Silurus glanis known to be contaminated by POPs. The fish was grinded, homogenized and freeze-dried. The material obtained, 20kg, was homogenized again and divided into four parts, which were analysed twice, in order to ascertain their homogeneity. After this test, the sample was stored in amber glass containers. The sample indicated as Feed (2014) is an extract of "Calendula Officinalis" used as an additive in feed for laying hens
A portion of each of the natural matrixes sent to the laboratories is opportunely stored for its future use as reference material (upon request by participants).

Analyses required.
The samples were sent in October of the year before (2012 and 2013 respectively for 1st and 2nd InterCind) to the laboratories asking to have measurements of PCDD/F (17 congeners), PCB (12 congeners), PCB-ICES 6 (6 congeners), PAH (7 congeners), PBDE (24 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). In the second Intercind was also asked to privide analyses in triplicate for heavy metals too (HM). Each laboratory was assigned an identification number (LAB#) used for assuring an objective data treatment and the privacy for all participants. The form for transmitting results from Lab to organizers has been standardized thanks to past collaboration between CIND and INTERCAL [1] intercalibration studies.

Statistical treatment of data.
The statistical indexes were calculated, for each matrix and congener, pooling together laboratories (i) and replicates (k) values. For each congener or compound(j) was calculated 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). as in the following for mean and standard deviation:

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

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

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:

$$IQ_j = Q75_j - Q25_j \qquad \text{(Eq. 5)}$$

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

extremes: $Q1_j - 3IQR_j > x_{i,k,j}$ and $Q3_j + 3IQR_j < x_{i,k,j}$ (Eq. 6)

outliers: $Q1_j - 1.5IQR_j > x_{i,k,j}$ and $Q3_j + 1.5IQR_j < x_{i,k,j}$ (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[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:

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

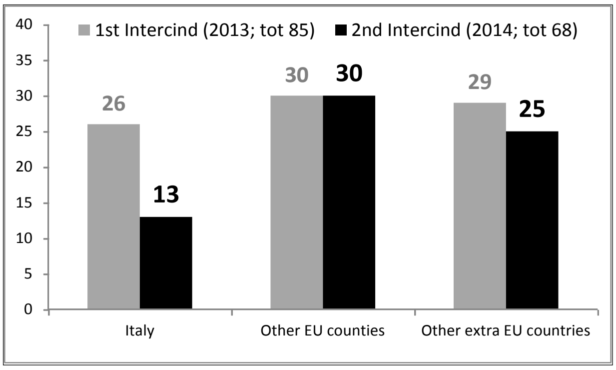
Z-scores assume that the "true value" is the average (and standard deviation) calculated for all laboratories and all replicates after elimination of extremes and outliers. The performance of each participant laboratory in term of precision (closeness between laboratory replicates) for each congener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

$$r_{i,j} = \frac{\sqrt{\frac{\sum_k (x_{i,k,j} - \bar{x}_{i,j})^2}{n-1}}}{\bar{x}_{i,j}} \qquad \text{(Eq. 9)}$$

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

The laboratories that reported results by the set deadline were 85 in the first InterCind and 68 in the second InterCind and laboratories were from 26 and 23 countries worldwide, respectively. Italy was the country most represented, whose laboratories participation reduction in 2014 was reflected to the total number of participants (Fig. 1).



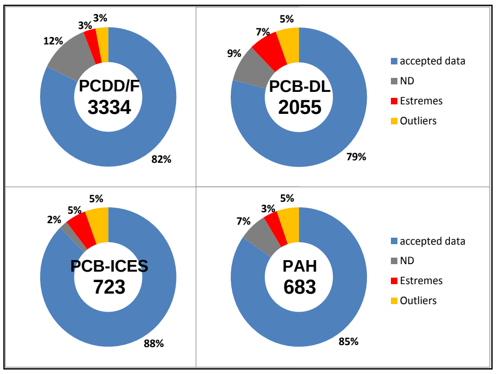


Figure 1. (Left). *Countries of the laboratories participating to InterCinD in 2013 and 2014*

Figure 2 (Right). *Representation of the proportions of laboratories' data after identification of the ouliers and extremes for each class of compounds (PCDD/F, PCB-DL, PCB-ICES and PAH). The example report the results for matrix "Sediment 2013" (1st Intercind).*

RSD% are estimated both for the original dataset and the dataset excluded from identified outliers and extremes values. The example report the results for matrix "Sediment 2013" (1st Intercind). The method applied allowed identify the outliers and extremes (see Figure 2, related to Sediment sample 2013, as an example). Statistical treatment allowed to reduce substantially the coefficient of variance of data (RSD%) as it is shown for the same exemple in Figure 3. The valid data were used to calculate also z-scores and r-scores for each congener and TEQ values and for each laboratory.

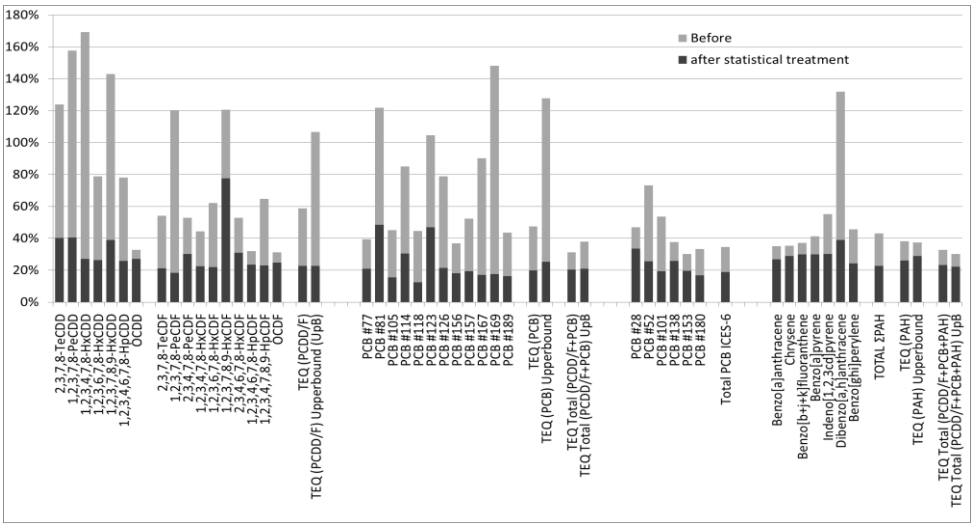


Figure 3*. Schematic representation of coefficient of variation (RSD%=SD/Mean) for all congeners and derived concentrations (such as TEQ) over all laboratories*.

Figure 4, reports z-scores (average of three replicates) and r-score for PCDD/F toxicity of the Sediment matrix analysed in the 1st Intercind as an example. These values provide indication of average accuracy and precision of each laboratory. These graphs allow to evaluate in a complete form the performance of laboratories and to detect ones with very low precision and low accuracy, laboratories with good precision but low accuracy, and Laboratories with problems both in accuracy and precision. These kind of evaluations in Intercind 1 and 2 were reported for all congeners and for the several matrices analysed. Results of these experiences revealed a) the importance of pre-check thanks to input form; b) the importance of the three replicates both for increasing the number of valid data and for evaluating precision of the Lab; c) the natural matrixes revealed to be more challenging for labs than the standards and solutions generally used in other intercalibration studies. Matrixes are kept by InterCind organizers as reference material for requesting labs.

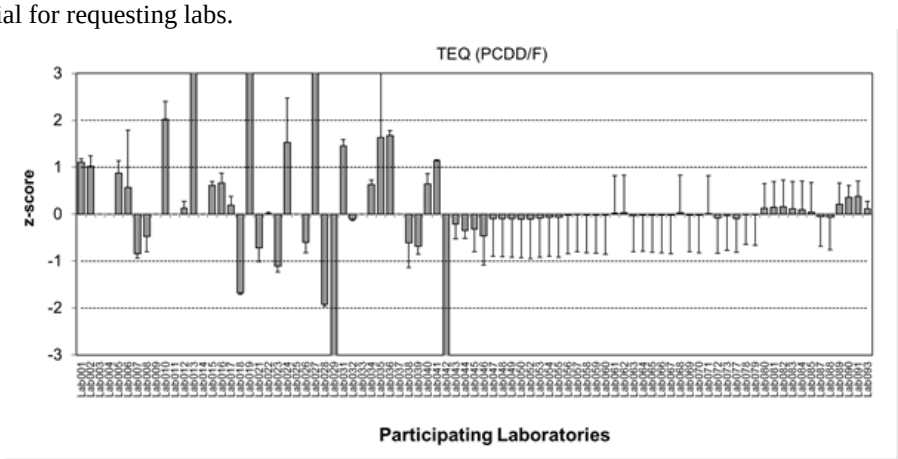


Figure 4. *Graphs of the z-scores and r-range for evaluating accuracy and precision respectively of all laboratories. The example report the results for matrix "Sediment 2013" (1st Intercind).*

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 could not have been held.

References:
1. Raccanelli S., Petrizzo A., Favotto M. and Pastres R. (2007). *Organoh. Comp*. 69: 982-985.
2. van Bavel B., Abad E., 2008. *Anal. Chem*. 80 (2008) 3956-3964.
3. Libralato S., Raccanelli S., van Bavel B., (2011). *Organoh. Comp*. 73: pp 2174-2177.