RESULTS FROM THE 8TH CIRCUIT INTERLABORATORY FOR DIOXINS (CIND)

Raccanelli Stefano [1] and Libralato Simone [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, Borgo Grotta Gigante 42/c, 34010, Sgonico (TS), Italy.

Abstract

The Circuit INter-laboratory for Dioxins (CIND) represents an important opportunity for laboratories for inter-comparing their analytical performances in measuring dioxins. In the 8th CIND edition two samples were sent to laboratories and analyses for PCDD/F, PCBs dioxin-like and PAHs were requested in three replicates. The two samples consisted in environmental matrix from the Lagoon of Venice (sediment) and in fly ash obtained from a municipal crematory. Among the 55 laboratories invited to participate to this CIND edition 48 provided results within the set deadline, for a total of 3618 and 3604 valid data for sediment and fly ash samples, respectively. For detecting extremes and outliers in data provided by laboratories a method based on nonparametric statistics (median and quartiles) was implemented. Such method showed to be more efficient than methods based on parametric statistics, and allowed to robustly and efficiently detect extreme and outlier values in one single run. Data eliminated by these values were used to calculate indicators of precision (z-scores) and accuracy (r-scores) thanks to the availability of analyses in three replicates providing meaningful insights.

Introduction

In order to give to all Italian laboratories the possibility of inter-comparing their analytical performances in measuring dioxins, the Consorzio Interuniversitario "La Chimica per l'Ambiente" (Consorzio INCA), prompted in the 2000 the 1st Circuit INter-laboratory for Dioxins (CIND), i.e. the first Italian Intercalibration study. This study was repeated in the following years inviting also foreign laboratories in order to increase the number of participants (only 3 Italian laboratories participated in the 2000 edition of the international inter-calibration studies [1,2]) and to give a broader perspective to the study. The 1st CIND regarded only PCDD/F whereas PCBs dioxin-like were included in the list of the congeners in 2001 and PAHs were included in 2004. These further POPs were added since, according to recent literature, their toxicity might be of the same order of PCDD/F in some instances[3]. The number of laboratories participating steady increase during past editions, especially due to foreign participation[4].

In the 8th CIND two samples were send to 55 laboratories and analyses were received from 48 of them. Laboratories were asked to provide analyses in 3 replicates for each sample. The main peculiarity of the 8th CIND consisted in the use of a method for detecting outliers and extremes that is based on no-parametric statistical indexes (median and quartiles) that has proven to be mode efficient than those based on parametric statistics as used in previous editions. Moreover, analyses in replicates allowed to obtain useful insights on the laboratories performances in terms of both accuracy and precision.

Material and Methods

Two samples have been sent to the laboratories asking to have measurements of PCDD/F, PCBs and PAH congeners in three replicates. The first sample consists in an environmental matrix (sediment) collected in the Lagoon of Venice. Large debris (>1cm) were separated by hand and homogenised "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 second sample consisted in fly ash obtained from a municipal crematory. The sample treatment was the same of the sediment sample.

For the 8th CIND edition, for each sample an amount sufficient for three replicates was delivered in October 2008 to 55 laboratories for the analysis of PCDD/F (17 congeners), PCB (12 congeners) and PAH (7 congeners). 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. In this CIND edition three replicates for each sample were asked for

all the analyses to all laboratories. To each laboratory was assigned an identificative number (LAB#) used for assuring an objective data treatment and the privacy for all participants. The Consorzio Interuniversitario "La Chimica per L'ambiente" (Consorzio INCA) store opportunely a portion of each sample for its future use as reference material (upon request by participants).
The main statistical indexes were estimated on the checked data matrix. For each congener (j) were calculated across laboratories: average concentration ($\overline{x}_j$), standard deviation (s_j), coefficient of variance (CVj= $s_j / \overline{x}_j$ %), Maximum and Minimum value observed (MAX_j and MIN_j, respectively), the median (M_j) and the 1st and 3rd quartiles ($Q25_j$ and $Q75_j$ respectively). The statistical indexes were calculated, for each sample, pooling together laboratories (i) and replicates (k) values. Average and standard deviation, for example, were calculated as:

$$\overline{x}_j = \frac{\sum_{i=1}^{n}\sum_{k=1}^{r} x_{i,k,j}}{\sum_{i=1}^{n}\sum_{k=1}^{r} 1} \qquad \text{(Eq. 1)}$$

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

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 [5]:

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

$P(xi,k,j < Q25i) \leq 0.25$ and $P(xi,k,j \leq Q25i) \geq 0.25$ (Eq. 4)

$P(xi,k,j < Q75i) \leq 0.75$ and $P(xi,k,j \leq Q75i) \geq 0.75$ (Eq. 5)

These non-parametric indexes (median and quartiles) were used to identify extremes and outliers according to statistical packages[6]. In particular, extremes were identified as the values that are outside the range defined by three times the quartiles out of the median, thus extremes are the values $x_{i,k,j}$ for which:

$$M_j - 3 \cdot Q25_j < x_{i,k,j} \quad \text{and} \quad x_{i,k,j} > M_j + 3 \cdot Q75_j \qquad \text{(Eq. 6)}$$

Similarly outliers were defined as those values that are outside the range defined by 1.5 times the quartiles out of the median, that is:

$$M_j - 1.5 \cdot Q25_j < x_{i,k,j} \quad \text{and} \quad x_{i,k,j} > M_j + 1.5 \cdot Q75_j \qquad \text{(Eq. 7)}$$

This method for defining extremes and outliers is reported in widely used statistical packages for producing box plots[6] and is much more efficient than the method based on the definition of outlier when data is falling outside the range of two standard deviations from the average value (as in previous CIND editions)[4]. The method reported above allowed identifying extremes and outliers that were excluded from the data set for calculations of

statistical indexes calculated across laboratories. The statistical indexes computed on the remaining data set, named "treated data", are reported in the following for each class of compounds.
The performance of each participant laboratory in term of accuracy 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 mean of all laboratories/all replicates calculated after elimination of extremes and outliers. The performance of each participant laboratory in term of precision (dispersion of values) for each congener was estimated by the range between maximum and minimum value reported in the three replicates relative to the average of the replicates, thus:

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

This 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 [3].

Results and Discussion

Among all laboratories that received the samples, 48 laboratories reported results by the set deadline (February 2009). Considering the 36 congeners to analyse and the 3 replicates per laboratory a total of 36x3x48=5184 data were expected for each sample. Raw data for sediment sample accounted for 6.6 % (342) of ND, 14.5 % (751) for NA of measured values (proportions calculated with respect to the total expected data). Similarly, raw data for fly ash sample accounted for 3.3 % (173) of ND, 17.3 % (896) for NA of measured values. Some laboratories did not reported results in three replicates and/or did not measured some congeners thus 4091 and 4115 data were received for sediment and fly ash samples, respectively ("raw data" in the following). In some cases laboratories indicated the specific detection limit (e.g. " < 0.0001"), that was used in the data treatment and statistical analyses (thus set as, e.g. 0.0001). The first screening evidenced that 473 values for sediment sample and 511 data for fly ash sample were evident unchanged background values that prompted for their definition as NA values that therefore became the 23.61 % and 27.14% of expected entries. The resulting matrix of data is therefore the "checked data" matrix for statistical analyses, consisting in 3618 and 3604 data values for sediment and fly ash samples, respectively (see Figure 1) .

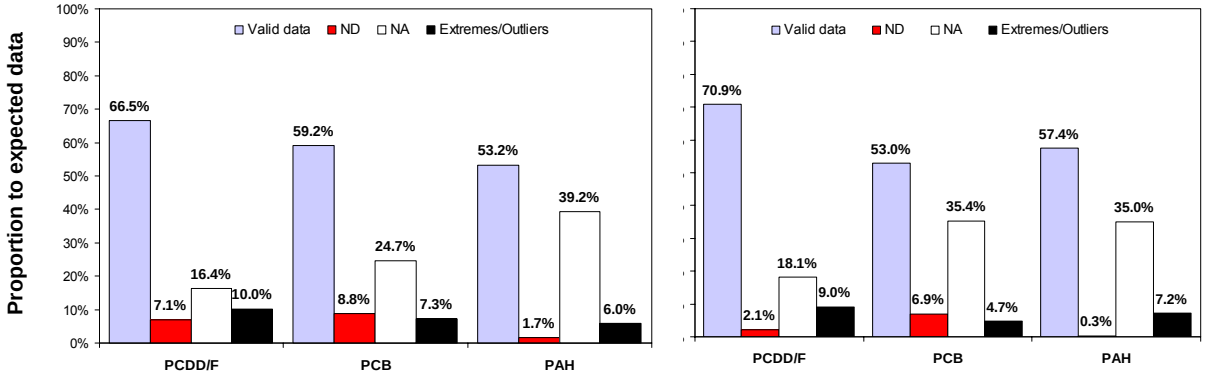

Figure 1

Proportions of data to be treated for statistical analysis after identification of the outliers and extremes. Proportions reported for each class of compounds, for sediment and fly ash left and right panel, respectively. Global figures (proportions of valid data against expected) are reported in the text.

The statistical treatment led to the identification of 280 extremes in the sediment data set (185, 72 and 23 for PCDD/F, PCB and PAH, respectively) and 274 extremes for the fly ash sample (188, 48, 38 extreme values for PCDD/F, PCB and PAH, respectively). Outliers were 152 in the sediment sample (61, 54 and 37 for PCDD/F, PCB and PAH, respectively) and 100 in the fly ash sample (32, 33 and 35 for PCDD/F, PCB and PAH, respectively). The method based on non parametric statistical indexes showed to be very powerful. In fact, CIND data are typically not distributed symmetrically and parametric methods based on standard deviation and mean appear highly inefficient. In this 8th CIND edition, for example, the identification of outliers on the basis of two standard deviations from the average value was not successful even after repeating reiteratively the identification process twice. Conversely, the above reported method based on non parametric statistics (median and quartiles) was proven to be much more efficient and identified opportunely extremes and outliers on the first run (verified by the normal distribution of accepted data).

All statistical indexes were calculated on the final database in which extremes and outliers were removed. The results in terms of coefficient of variance (CV), for PCDD/F, PCB and PAH in the two samples are represented in Figure 2. As shown, the coefficient of variation among PCDD/F ranged from 84% (1,2,3,7,8,9-HxCDF) to as 25% (for 2,3,7,8-TeCDF) in the sediment sample and from 90% (1,2,3,7,8,9-HxCDF) and 25% (1,2,3,4,6,7,8-HpCDF) in the fly ash. Among PCBs the coefficient of variation ranged from 67% (PCB 81) to 19% (PCB 118) for sediment and from 76% (PCB 105 and PCB 118) and 38% (PCB 77) for fly ash. Regarding PAH the CV ranged from 52% (calculated for Indeno[1,2,3cd]pyrene) and 28% (Benzo[b+j+k]fluoranthene) for sediment, whereas the range for fly ash was from 61% (Dibenzo[a,h]anthracene) and 42% (Benzo[ghi]perylene).

Valid data were used to calculate also z-scores and r-scores for each congener and TEQ value and for each laboratory. As an example, Figure 3 reports z-scores (average of three replicates) and r-score for PCDD/F toxicity. As a general result one can see that precision and accuracy is not necessarily related to the method used to make the analysis (high resolution vs low resolution). However evaluation of accuracy and precision should be evaluated directly by congener and results for the total toxicity (sum of WHO-TEQ for PCDD/F, PCB and PAH) should be considered with caution.

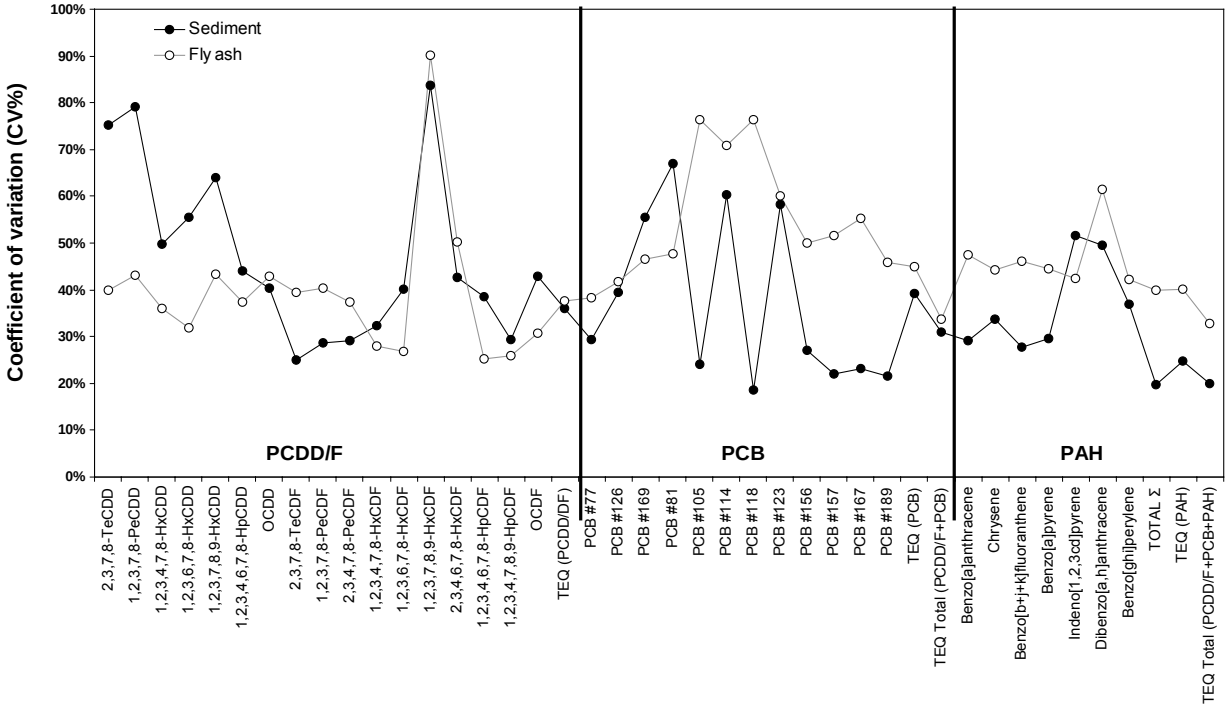

Figure 2

Coefficient of variation (CV%) of treated data represented for all congeners and for TEQ over all laboratories. CV% are estimated excluding outliers and extremes. Values provide basis for evidencing the average accuracy (over all laboratories) related to each congener.

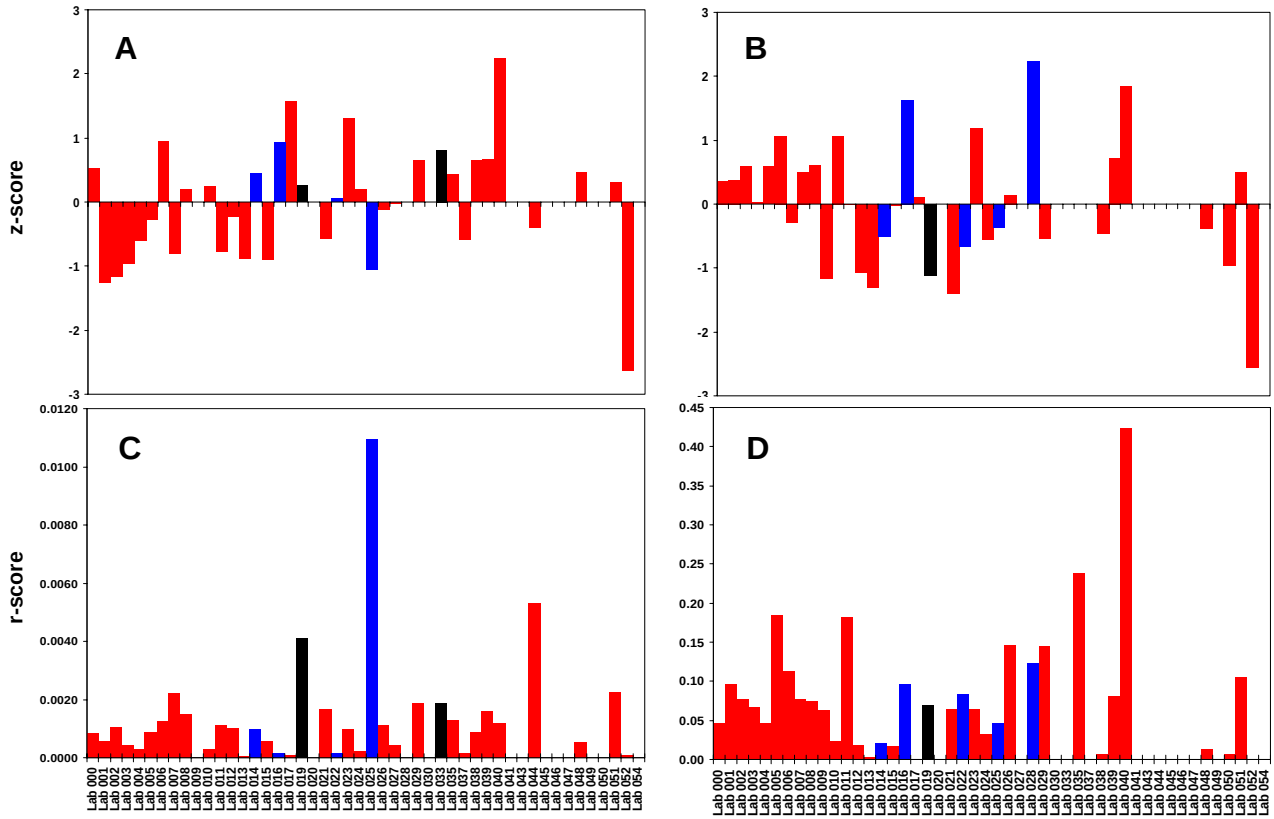

Figure 3
Results obtained for precision (z scores panels A,B) and accuracy (r-scores, panels C,D), for the PCDD/F TEQ taken as an example. Sediment sample in left panels (A,C) and ash fly in right panes (B,D). Red, blue and black bars indicate high, low and unspecified resolution analytical method, respectively. Values consist in the average of the three replicates; outliers and extremes were excluded.

Thus for each laboratory the average z-scores and r-scores for all congeners were calculated, excluding TEQ values. These averages are though to give indication of average accuracy and precision of each laboratory. Combined representation of average accuracy (r-score) and precision (z-score) estimates for each laboratory are reported in Figures 4 and 5 for sediment and fly ash, respectively. These graphs allow to evaluate in a complete form the performance of laboratories and to detect ones with very low precision and low accuracy (Lab 052 for sediment), with very high precision but low accuracy (e.g. Lab 012 for fly ash), with high accuracy and low precision (e.g. Lab 052 for fly ash) and laboratories with high performances in terms of both accuracy and precision (e.g. Lab 004 for sediment and Lab 017 for fly ash).

Conclusions
The innovation of the 8th CIND edition regarded the statistical method for identifying outliers and extremes that was based on non parametric statistical indexes. This method has proven to be robust and more efficient than previous method based on parametric statistics. The present method should be thus recommended in intercalibration analyses. The request of three replicates for the two samples, demonstrated to be extremely useful. Although 3 replicates are still a low number for statistical inference, they allowed for inferring both the accuracy and the precision of the laboratories thus providing meaningful insights. Evaluations of laboratories should be considered on the basis of z-scores and r-scores for each congener or class of compounds. The evaluation based on scores estimated for TEQ, in fact, are deeply influenced by Toxicity Factors (TF). Moreover, since TEQ for PCDD/F, PCB and PAH were simultaneously available in very few cases these overall TEQ (PCDD/F+PCB; PCDD/F+PCB+PAH) need to be considered with caution. The 8th CIND was also

successful for the increased participation. The samples used in all CIND editions are stored and therefore they provide an archive of test samples, which are available upon request and could be used by laboratories for further testing their performances.

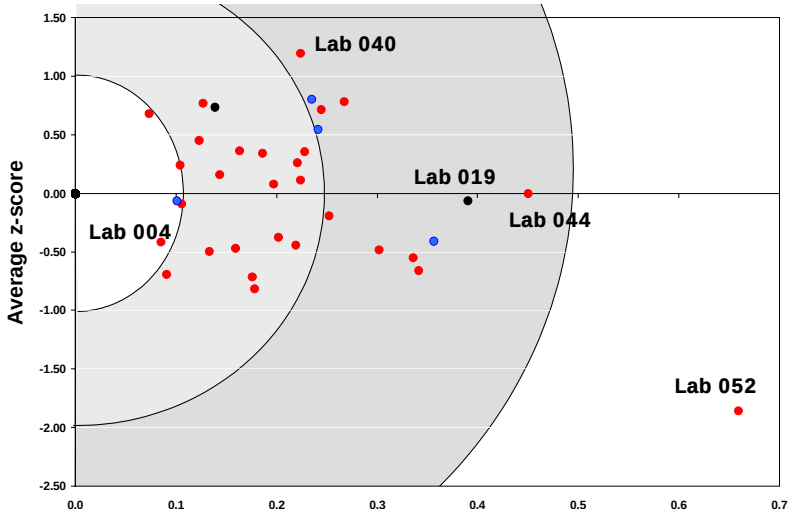

Figure 4
Comparison of laboratory results obtained in terms of average precision (z scores) and average accuracy (r-scores), for all congeners measured in sediment sample. Red, blue and black circles indicate high, low and unspecified resolution, respectively. Grey areas schematically represent decreasing combined values of accuracy and precision.

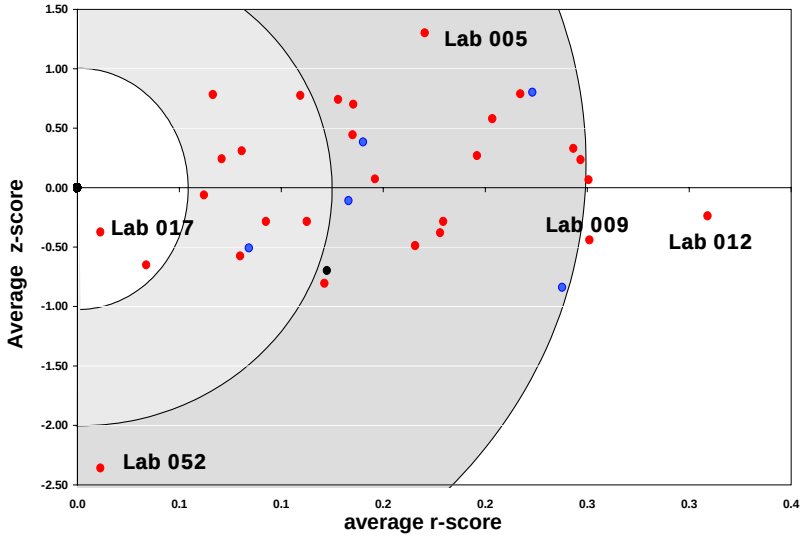

Figure 5
Comparison of laboratory results obtained in terms of average precision (z scores) and average accuracy (r-scores), for all congeners measured in fly ash sample. Symbols used are the same that precedent figure.

Acknowledgments
The authors wish to thank Dr. Ivano Battaglia and Dr. Simona Manganelli (LabService Analytica, Italy), without whom these interlaboratory studies (CIND) could not have been held with yearly frequency.

References
1. van Bavel B., Fifth Round of the International Intercalibration Study, Umea University, 2000.
2. Lindström G., Småsuen Haug L., Nicolaysen T., Intercalibration on Dioxin in Food, Folkehelsa, 2000.
3. Klimm C., Hofmaier A.M., Schramm K.W., Kettrup A. (1999), Using TEF concept for assessing toxic potency of polycyclic aromatic hydrocarbons in industrial samples. *Organoh. Comp.*, 40, 39-42.
4. Raccanelli S., Petrizzo A., Favotto M. and Pastres R. (2007). Results from the 6th round of the italian intercalibration study for PCDD/F, PCB and PAH in fly ash and sediment. *Organoh. Comp.* 69: 982-985
5. Cicchitelli, G., 1995. Probabilità e Statistica. Maggioli Ed., 325 pp.
6. StatSoft, Inc. (2005). STATISTICA (data analysis software system), version 7.1. www.statsoft.com.