APPLICATION OF THE POWER-MODERATE WEIGHTED MEAN (PMM) CONCEPT IN THE CALCULATION OF REFERENCE VALUES OF INTERLABORATORY COMPARISONS

M. SAHAGIA**, 1, A. LUCA1, A. ANTOHE1, M.-R. IOAN1, E. GARCIA-TORANO2

1 “Horia Hulubei” National Institute for R&D in Physics and Nuclear Engineering, IFIN-HH, P.O Box MG-6, RO-077125, Bucharest, Romania
2 Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
** Corresponding author: msahagia@nipne.ro
Received June 25, 2015

Recently, the Consultative Committee for Ionizing Radiations Section II, Measurement of Radionuclides, of the International Committee of Weights and Measures [CIPM-CCRI(II)], decided to apply the new concept of power-moderate weighted mean (PMM) in calculation of the Key Comparison Reference Values (KCRV), standard uncertainties and degrees of equivalence, Document CCRI(II)/13–37. The paper will exemplify the calculations of $^{226}$Ra activity concentration of a slag sample, using the PMM mode, for which we had recommended before the use of the weighted mean. This value will also be compared with a recently published one.

Key words: power-moderate weighted mean, interlaboratory comparison (ILC), comparison reference value (CRV).

1. INTRODUCTION

The problem of the most adequate method for treatment of experimental data in the calculation of some representative parameters describing the whole population is of prime importance in many situations. It includes two essential steps in judgement:

(i) The critical evaluation of individual results, in order to decide which of them are part of the distribution and elimination of the outliers;
(ii) The choice of the most convenient method to combine the individual results, in order to obtain the best evaluation of the mean value of the final result and its uncertainty.

The most frequent situations found in our field of work are the following:

– Evaluation of Nuclear Decay Data for radionuclides within the international programs, as the Decay Data Evaluation Program (DDEP), the Evaluated Nuclear Structure Data File (ENSDF), etc. from published experimental data, such as: half-life, emitted radiation energies and their intensities, which are then published in Data Tables [1–3] and recommended to be used by the entire scientific community in the field.

– Calculation of the Key Comparison Reference Value (KCRV) within the key and supplementary comparisons, organized within the CIPM-MRA (Mutual Recognition Arrangement) in order to establish: (a) The reference value for the response of the International System of Reference (SIR), existing at BIPM and (b) The degree of equivalence of the National and Designated Metrology Institutes, signatories of the CIPM-MRA, in order to support their claims for Calibration and Measurement Capabilities (CMCs) [4]. In these particular cases, many efforts were done in order to calculate the best KCRV, recognizing its importance on international scale. Up to now, the responsible persons for evaluation of data usually processed them in several steps [5]: analysis of the data from the national participants, application of criteria for identifying outliers, calculation of the indicators: arithmetic mean, weighted mean, median, weighted median and their uncertainties. The external and internal uncertainties and their ratio, Birge ratio [6], were calculated, in order to detect some underestimation or overestimation of uncertainties. This evaluation, although very comprehensible, involves much effort and consequently an equilibrated, simplified, method was looked for. The final calculation of the KCRV was based on the elimination of outliers and calculation of the arithmetic mean [7]. Recently, the Consultative Committee for Ionizing Radiations Section II, Measurement of Radionuclides of the International Committee of Weights and Measures (CIPM-CCRI(II)), decided to apply the new concept of power-moderate weighted mean (PMM) in calculation of the Key Comparison Reference Values (KCRV), standard uncertainties and degrees of equivalence, Document CCRI(II)/13–37 [8], to be found at the address http://www.bipm.org. The concept, introduced in the publication EUR 25355 of the JRC-IRMM, 2012, by S. Pommé [9], takes the advantage of balancing between the arithmetic and the weighted mean calculations, as an improvement of the Mandel-Paule mean calculation [10], which in some cases tends to be similar to the arithmetic mean. The new method was already applied in calculation of the KCRV in the case of some new key or supplementary comparisons [11, 12].

– Calculation of the consensus value of a reference material as a result of an intercomparison, on international as well as national scale, such as presented in [13].

The paper focuses on a similar situation, a comparison between two Romanian and three Spanish laboratories, aimed to determine the $^{226}\text{Ra}$ activity concentration of a slag sample, proposed to be used for spiking with a standard
226Ra solution, for which in the published paper [14] we had recommended the use of the weighted mean for calculation of the reference values. This result will also be compared with a recently published one regarding the subject [15].

2. PRESENTATION OF THE POWER-MODERATE WEIGHTED MEAN (PMM) CALCULATION

According to the literature data, the most significant quantities used in describing a set of \( N \) experimental data, with individual values \( x_i \), having individual uncertainties \( u_i \), is described by the parameters presented in Table 1 [8].

<table>
<thead>
<tr>
<th>No.</th>
<th>Quantity</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Arithmetic mean</td>
<td>( \bar{x} = \frac{\sum x_i}{N} )</td>
</tr>
<tr>
<td>2</td>
<td>Weight of a result</td>
<td>( W_{a_i} = \frac{1}{N} )</td>
</tr>
<tr>
<td>3</td>
<td>Standard deviation of the mean (External)</td>
<td>( u(x) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N(N-1)}} )</td>
</tr>
<tr>
<td>4</td>
<td>Internal uncertainty of the mean value</td>
<td>( u(x) = \frac{1}{N} \sqrt{\sum u_i^2} )</td>
</tr>
<tr>
<td>5</td>
<td>Weighted mean</td>
<td>( x_{\text{weigh}} = \frac{\sum x_i / u_i^2}{\sum 1/u_i^2} ) or ( x_{\text{weigh}} = u(x_{\text{weigh}}) \sum x_i / u_i^2 )</td>
</tr>
<tr>
<td>6</td>
<td>Weight</td>
<td>( W_{\text{weigh} \ i} = \frac{1/u_i^2}{\sum 1/u_i^2} )</td>
</tr>
<tr>
<td>7</td>
<td>Standard deviation of the weighted mean (Internal)</td>
<td>( u(x_{\text{weigh}}) = \frac{1}{\sqrt{\sum 1/u_i^2}} )</td>
</tr>
<tr>
<td>8</td>
<td>Reduced observed chi-squared value</td>
<td>( \chi^2 = \frac{1}{N-1} \sum (x_i - x_{\text{weigh}})^2 / u_i^2 )</td>
</tr>
</tbody>
</table>

Comments on Table 1

When reporting a result as arithmetic (unweighted) mean, it is advisable to compare the two quantities, rows (3) and (4) in Table 1 and take into account the maximum of them;
Normally, the reduced observed chi-squared value, row (8) in Table 1, should be equal to unity; in the contrary case, mainly when the uncertainties are underestimated, the set is non consistent.

**Application of the exclusion criteria**

In the case of discrepant data, it is necessary to reject outlier values, *i.e.* those values in the set which do not belong to the distribution, supposed to be normal, and which could distort the calculation of the mean value. The most used exclusion criterion is known as the Chauvenet’s criterion [16], [17]. It is based on the calculation of the ratio:

$$ p = \frac{(x_i - \bar{x})}{u(x)} $$

for which an upper limit of acceptability is defined. This parameter is calculated from the arithmetic mean and its standard deviation. The limit value depends on the number of data, \( N \). In [16] \( p \) is approximated by the relation:

$$ p = 0.91772 + 1.0168 \log N. $$(2)

In the paper [17] it is given as an equivalent table.

Both the arithmetic and weighted means were considered as extreme cases of calculation and consequently other intermediate calculations were taken into account. In paper [16] several alternatives are proposed for application, when discrepant data are to be evaluated, one of them being to use weights according to relation:

$$ w_{\text{weight}} \left( \sum w_{\text{weight}} \right) \leq 0.5. $$

A well-known, recommended calculation method is that of Mandel-Paule (M-P) [10], situated between the two extreme calculations, mainly applied when the set is non consistent, that is when the reduced observed chi-squared value \( \chi^2 \) is higher than unity. In this method, an additional uncertainty “\( s \)” is calculated such as that the new chi squared value should become:

$$ \chi^2 = \frac{1}{N-1} \sum (x_i - x_{\text{ref}})^2 / (u_i^2 + s^2) \leq 1. $$

A new development of the M-P method was done by S. Pommé and Y. Spasova [18] and published in the concise document EUR 25355 [9]. It consists in the following: after determining the value of \( s \), the new M-P weight, weighted mean and its uncertainty are calculated as:

$$ w_{(M-P)i} = [1/(u_i^2 + s^2)] / \sqrt{\sum 1/(u_i^2 + s^2)}; $$
Application of the power-moderated weighted mean (PMM) concept

\[ X_{M-P} = \sum \frac{x_i}{(u_i^2 + s^2)} = u^2(X_{M-P}) \sum \frac{x_i}{u_i^2 + s^2} \]

and

\[ u(X_{M-P}) = 1/\sqrt{\sum 1/(u_i^2 + s^2)} \]

(5)

The concept of power-moderated weighted mean, introduced in [9], is based on the use of a new parameter, “\( \alpha \)”, describing the degree of confidence in the reported uncertainties “\( u_i \)”, and reflected in different types of mean evaluation, as indicated in Table 2.

<table>
<thead>
<tr>
<th>( \alpha ) – values</th>
<th>Type of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0 )</td>
<td>( u_i ) are “noninformative”, nonconfident and not considered – arithmetic mean</td>
</tr>
<tr>
<td>( \alpha = 2 )</td>
<td>( u_i ) are “informative”, fully confident and taken into account – weighted and Mandel-Paule mean</td>
</tr>
<tr>
<td>( \alpha = 2-3/N )</td>
<td>( u_i ) are only “partially informative”, partially confident and are considered as being underestimated – calculation of an intermediary weighted mean</td>
</tr>
</tbody>
</table>

Further one, Pommé and Spasova propose to calculate the parameter “characteristic uncertainty per datum” \( S \), allowing to choose which is the most convenient uncertainty evaluation mode, according to the relation:

\[ S = \sqrt{N \cdot \text{Max}[u^2(\bar{x}), u^2(X_{M-P})]} . \]

(6)

The relations for calculation of the new proposed calculation mean and uncertainty, denominated as power-moderate weighted mean (PMM) were deduced as:

\[ w_{PMM} = 1/\sqrt{(u_i^2 + s^2)^\alpha S^{2-\alpha}} \]

\[ X_{PMM} = u^2(X_{PMM}) \sum \frac{x_i}{(u_i^2 + s^2)^\alpha S^{2-\alpha}} \]

and

\[ u(X_{PMM}) = 1/\sqrt{\sum (\sqrt{u_i^2 + s^2})^{-\alpha} S^{\alpha-2}} . \]

(7)
Relations (7) reduce to the arithmetic, respectively weighted means, for the particular cases: \( s = 0, \alpha = 0 \) and respectively \( s = 0, \alpha = 2 \).

3. CALCULATION OF THE POWER-MODERATE WEIGHTED MEAN FOR THE DETERMINATION OF \( {226}\text{Ra} \) CONCENTRATION IN FURNACE SLAG

The data set, which is presented in Table 3, refers to the content of \( {226}\text{Ra} \) in a slag sample from a Spanish foundry, which was measured by gamma-ray spectrometry by two Romanian and three Spanish laboratories, with a total of six results. The individual values were presented in paper [14] and seem to be rather discrepant and having very different uncertainties, which are due to the difficulty to measure such low activities.

### Table 3

Radionuclide activity concentration of the sample, in Bq kg\(^{-1}\), reported by participants

<table>
<thead>
<tr>
<th>LABORATORY</th>
<th>IFIN-HH, ( \mu \text{Bq} )</th>
<th>IFIN-HH, RML</th>
<th>CIEMAT, LMRI</th>
<th>CIEMAT, LMPR</th>
<th>CIEMAT, LRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radionuclide</td>
<td>( {226}\text{Ra} )</td>
<td>*</td>
<td>*</td>
<td>Prompt ( {222}\text{Rn} ) emission</td>
<td>Equilibrium ( {226}\text{Ra} )-( {222}\text{Rn} )</td>
</tr>
<tr>
<td>( x_i, \text{Bq/kg} )</td>
<td>17.1</td>
<td>14.0</td>
<td>23.5</td>
<td>18.9</td>
<td>–</td>
</tr>
<tr>
<td>( u_i, \text{Bq/kg} )</td>
<td>±2.7</td>
<td>±7.0</td>
<td>±14.3</td>
<td>±11.0</td>
<td>±9.0</td>
</tr>
</tbody>
</table>

* Both laboratories measured the activity from the 186.21 keV Full Absorption Peak (FAP), corrected for the contribution of the 185.72 keV FAP from \( {235}\text{U} \).

Applying the Chauvenet criterion for the arithmetic mean

The quantity \( x_i = 23.5 \text{ Bq/kg} \) seems to be suspect. Indeed, the value \( p = 4.67 \) is superior to \( p = 1.78 \) from relation (2) for \( N = 6 \) and the quantity is an outlier.

The estimators: mean values and their uncertainties for all modes of calculation – arithmetic, weighted, Mandel-Paule, power-moderate weighted means and medians – are compared in Table 4 and Figure 1, for two situations: use of all reported values and exclusion of the outlier.
Table 4
Mean values and their uncertainties

<table>
<thead>
<tr>
<th>No.</th>
<th>Quantity, $^{226}\text{Ra}$ activity concentration, Bq/kg</th>
<th>Calculated with all reported values</th>
<th>Exclusion of the value 23.5 Bq/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Arithmetic mean</td>
<td>17.2</td>
<td>15.96</td>
</tr>
<tr>
<td>2</td>
<td>Median</td>
<td>17.1</td>
<td>16.0</td>
</tr>
<tr>
<td>3</td>
<td>Standard deviation of the arithmetic mean</td>
<td>$\pm 1.4$</td>
<td>$\pm 0.96$</td>
</tr>
<tr>
<td>4</td>
<td>Weighted mean</td>
<td>14.0</td>
<td>14.0</td>
</tr>
<tr>
<td>5</td>
<td>Standard deviation of the weighted mean &quot;internal error (uncertainty)&quot;</td>
<td>$\pm 0.6$</td>
<td>$\pm 0.6$</td>
</tr>
<tr>
<td>6</td>
<td>Reduced observed chi-squared value</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>7</td>
<td>Paul Mandel mean</td>
<td>$s = 0; 14.0$</td>
<td>$s = 0; 14.0$</td>
</tr>
<tr>
<td>8</td>
<td>Standard deviation of the Mandel-Paule mean</td>
<td>$\pm 0.6$</td>
<td>$\pm 0.6$</td>
</tr>
<tr>
<td>9</td>
<td>S-value; $\alpha$-value</td>
<td>3.43; 1.5</td>
<td>2.15; 1.4</td>
</tr>
<tr>
<td>10</td>
<td>Power-moderated weighted mean</td>
<td>14.3</td>
<td>14.3</td>
</tr>
<tr>
<td>11</td>
<td>Standard deviation of the power-moderated weighted mean</td>
<td>0.8</td>
<td>1.1</td>
</tr>
</tbody>
</table>

![Diagram](image)

Fig. 1
By analyzing Table 4 and Fig. 1, the following conclusions can be obtained:

– In all cases the individual uncertainties of the experimental data were not underestimated.

– When calculating the arithmetic mean, the differences between the mean of all six experimental data and that without the outlier, \( x_i = 23.5 \text{ Bq/kg} \), are relatively high. The median values are very close of the arithmetic means.

– In the case of weighted means the difference vanishes as the reported uncertainty of the value \( x_i = 23.5 \text{ Bq/kg} \), with \( u_i = 14.3 \text{ Bq/kg} \), is high. Both weighted means are smaller than the arithmetic ones, due to the strong influence of the lowest value \( x_i = 13.8 \text{ Bq/kg} \), reported with a very low uncertainty, \( u_i = 0.6 \text{ Bq/kg} \). The two uncertainties are equal, but they seem to be rather small, among others due to the small predominant uncertainty.

– Due to the correct evaluation of the uncertainties, the reduced observed chi-squared value was lower than unity and there was not necessary to adjust the “s” parameter when applying the Mandel-Paule model.

– Application of the power-moderated weighted mean resulted in the mitigation of the influence of the lowest value, \( x_i = 13.8 \text{ Bq/kg} \), and a small rise of both mean values. The two calculations, using all results and rejecting the “outlier”, are equal. The uncertainties are moderately higher as compared with the weighted means.
4. COMPARISON WITH OTHER MEASURED VALUES

The calculation of the power moderate weighed mean, due to the moderation in the use of the weights, provided a value a little higher than the weighted mean, recommended in the paper [14] to be used as reference value of $^{226}$Ra concentration in slag, namely: $x_{(\text{weigh})} = (14.0 \pm 0.6)$ Bq/kg. The new $x_{(\text{PMM})} = (14.3 \pm 0.8)$ Bq/kg proves to be a better estimate, as it accounts for the smaller influences of extreme values; however, it is not statistically different from the former one.

After the publication of our result, a new determination of the $^{226}$Ra concentration, using another sample prepared from the same batch, was performed in the underground laboratory HADES in Mol (Belgium), by a common team from CIEMAT-Spain and JRC-IRMM-Belgium, and was reported in [15]. The new determined and reported value was $(12.8 \pm 2.0)$ Bq/kg, a concentration value in good agreement with our measured and reported values.

5. CONCLUSIONS

– The new method, power-moderate weighted mean, adopted for the calculation of the key comparison reference value (KCRV) in international comparisons, was applied to calculate the $^{226}$Ra concentration in a slag sample within an interlaboratory comparison.

– The method provided satisfactory results in a situation of using discrepant results, due to the moderation in taking into account of the individual reported uncertainties.

Acknowledgements. This paper is a continuation of the work deployed within the EURAMET EMRP Joint Research Project IND04 “MetroMetal” and was supported by the IFIN-HH Research Project 09 37 02 05/ 2015.

REFERENCES


