

# International Comparison CCRI(II)-S7 on the analysis of Uncertainty Budgets for $4\pi\beta\gamma$ Coincidence Counting

*Claude J. Bailat (IRA), John Keightley (NPL), Youcef Nedjadi (IRA), Li Mo (ANSTO), Guy Ratel and Carine Michotte (BIPM), Miguel Roteta (CIEMAT), Maria Sahagia and Anamaria C. Wätjen (IFIN-HH), Ming-Chen Yuan (INER), Jong Man Lee, Tae Soon Park, K.B. Lee, Sang Han Lee and Pil Jae Oh (KRISS), Akira Iwahara (LNMRI/IRD), Yasushi Sato, Yasuhiro Unno and Akira Yunoki (NMIJ), Wilfrid M. van Wyngaardt, Joline Lubbe, Martin J. van Staden and Bruce R.S. Simpson (NMISA), Andrey V. Zanevsky (VNIIM), and François O. Bochud (IRA)*

## Abstract

Detailed uncertainty reporting is imperative for proficiency tests and comparison exercises since uncertainties need to be comparable and trusted by all the participants. Even though participants do their best to follow the Guide to the Expression of Uncertainty in Measurement<sup>1</sup>, ambiguities and divergences about uncertainty evaluation remain. Consequently, to analyze the situation, the CCRI (II) Uncertainties Working Group proposed a comparison exercise (CCRI(II)-S7) about the uncertainty evaluation of a relatively simple primary activity measurement: the standardization of a  $^{60}\text{Co}$  source by coincidence counting.

To be able to understand how various NMIs calculate coincidence counting uncertainties, our study focused on two of the dominant uncertainty components commonly quoted for  $4\pi\beta\gamma$  coincidence counting in the International Reference System (IR) submissions and Key Comparison exercises: efficiency-extrapolation and weighing. Participants from twelve different laboratories were sent the same set of measurement data from the analysis of a  $^{60}\text{Co}$  solution standardized at the National Physical Laboratory (NPL).

Our study demonstrated the extent of the different interpretations of the uncertainty components. Some factors causing large discrepancies were isolated and are discussed. Further studies of other techniques using a similar approach would be beneficial for the metrology community.

## 1 Introduction

In an international key comparison (KC), the participants provide their estimate of the measurement value and its associated combined uncertainty. It is critical to know the uncertainty evaluation strategies used by the participants in order to calculate a key comparison reference value (KCRV). Often the relative standard uncertainties reported show large discrepancies, which remain unexplained, because the necessary information is not typically reported. Furthermore, some statistical tools used to calculate the KCRV might necessitate the removal of some outliers. This process is difficult in the absence of detailed knowledge of the uncertainty evaluation.

In order to better understand the discrepancies observed in the uncertainties reported in  $4\pi\beta\gamma$  coincidence activity measurements, the CCRI (II) Uncertainties Working Group, UCWG (II),

---

<sup>1</sup> Evaluation of measurement data — Guide to the expression of uncertainty in measurement (2008); Joint Committee for Guides in Metrology, JCGM 100:2008, GUM 1995 with minor corrections, [http://www.bipm.org/utils/common/documents/jcgm/JCGM\\_100\\_2008\\_E.pdf](http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf)

proposed that a simple comparison exercise be undertaken to examine how various NMIs determine and report uncertainties affecting the coincidence counting technique for radionuclide standardization.

Often, uncertainty budget cannot be compared directly due to different experimental conditions. Consequently, this comparison focuses on two of the dominant uncertainty components commonly quoted for  $4\pi\beta\gamma$  coincidence counting in SIR submissions and Key Comparison exercises: efficiency-extrapolation and weighing. These two evaluations were required and participants had to perform them using a set of provided data. Restricting the scope of the comparison to these two components was necessary because of incompatible experimental methods such as dead time evaluation, live-timing techniques and coincidence count-rate corrections.

However, as an additional exercise, participants were invited to describe the methods they would typically use to evaluate other uncertainty components. They could also estimate the values of these components.

## 2 Method

Participants were sent a set of data to be used to evaluate the associated uncertainties. The data provided for this comparison were from the analysis of a Co-60 solution standardized at the National Physical Laboratory (NPL). The solution had been used to prepare a set of ten VYNS sources, which were then measured for a week using NPL's  $4\pi\beta\gamma$  coincidence system. An ampoule prepared from the same stock solution was submitted to the International Reference System (SIR) in August 2009.

The participants received the raw data of the coincidence counting results and detailed data about the coincidence counting system, as well as the method and the procedure followed during the experiment. Counting conditions of the standardization and background counting values were provided. The data obtained during the standardization, shown in Figure 1, were synthesized in a spreadsheet and sent to the participants. The participants were first asked to fit the data for two selected gates, deduce the activity concentration and report the associated uncertainties. In reporting their results, they were asked to evaluate the different contributions to the uncertainty and more specifically detail the contributions due to weighing and efficiency extrapolation. The exercise focused on the evaluation methods and not on coincidence counting. The Baerg efficiency extrapolation method was chosen and we are aware of different efficiency extrapolation methods, but this choice did not influence the outcome of the comparison.

In order to estimate the uncertainty contribution due to weighing, data about the balances, stability checks as well as detailed logs were provided. The ambient temperature  $T$  (in °C), the atmospheric pressure  $P$  (in hPa) and the relative humidity  $H$  (in percent) were provided to allow the participants to estimate the buoyancy correction and its contribution to the uncertainty.

The weighing data were recorded in such a manner as to allow the use of either the "direct" or the "substitution" weighing method. In the "direct" method of weighing, the data set showed the zero load (or tare) readings before and after the pycnometer was placed onto the balance as well as the pycnometer mass readings before and after dispensing a drop of the solution onto the VYNS support of each source. In the "substitution" method, the provided data reported the pycnometer mass reading, the list of calibration substitution weights, the value of the total nominal mass and the actual recorded mass of the sum of these calibration weights listing these values before and after dispensing a drop of solution onto the VYNS support of each source. The certificate of the calibration weights was provided as well.

The balance used for weighing the pycnometer had been previously calibrated by NPL and the balance's certificate of calibration was made available. Details about the repeatability, off-center loading, linearity and uncertainty of its measurements were given in the certificate.

However, the uncertainty stated in the NPL certificate was not expressed to accommodate every participant. Some NMIs are accustomed to a certificate in which the uncertainty is not constant but rather depends on the weighing range, often proportional to the load. Additionally, the absolute standard uncertainty of a given weighing was provided according to the procedure DKD-R 7-1 of the German calibration service (DKD).

Knowing that some NMIs used the full calibration history of the balance in estimating the uncertainty of its weighing, these data were provided as well. No data about evaporation were made available, and it was left up to NMIs to use their own numbers.

The participants were instructed to report their results using a reporting form, which was sent with all the necessary data. The form was intended to frame the results in order to make them comparable. The results were then extracted from the form and clarifications about problematic results were sought from those specific participants.

At the end, twelve laboratories, listed in Table 1, participated in the interlaboratory comparison.

### 3 Results

#### 3.1 Activity Data

Table 2 and

Figure 2 summarize the values obtained by the various laboratories for the final results reporting the activity concentration. The average standard uncertainty and associated relative standard deviation are shown as well. The activity values are between 354.3 kBq/g and 354.7 kBq/g. The associated relative standard deviations are between 0.10 % and 0.31 %.

Most laboratories evaluated seven contributions: weighing, extrapolation, dead times, resolving times, backgrounds, counting statistics, and half-life. Not many other contributions were quoted by the participants.

Figure 3 shows the square of the relative uncertainty contributions for each laboratory. We chose this representation, because the participants chose to compute the total relative uncertainty by quadratic summation. For most laboratories, the main uncertainty contributions of the total standard uncertainty are weighing and extrapolation. The NMISA chose to take into account only the contributions due to weighing and extrapolation. The ANSTO evaluated the activity concentration with and without additional contributions and decided not to take them into account for the total uncertainty evaluation.

The reported values reflect various choices for truncation. Some reported with two significant digits, as recommended by the GUM under 7.2.6, but others preferred to express all uncertainties to the smallest digit of all uncertainty contributions.

Table 3 displays the type of uncertainty evaluation (A and/or B) reported by the participants. Unanimity can be found only for the contributions due to resolving time and half-life, both reported as pure B-type evaluations.

#### 3.2 Activity Extrapolation Data

Table 4 summarizes the results reported by the participants for the extrapolation data. Except for the ANSTO and the BIPM, all laboratories made a fit for each gate by considering all the sources together. The NMIJ did not fit the data for gate 2 and the BIPM carried out extrapolation for one source at a time and reported the weighted mean of the extrapolated  $y$ -intercepts and slopes. Figure 4 shows the extrapolated  $y$ -intercepts separately for gates 1 and 2 and Figure 5 shows the extrapolated slopes for gates 1 and 2.

Most of the participants used extrapolation methods that do not take into account the  $x$ -uncertainty. The majority of the participants used a least square fitting method to obtain the  $y$ -intercept, the slope and the associated uncertainties. The ANSTO, NMIJ, IRA and NPL took into account  $x$ - and  $y$ -uncertainties. The ANSTO used a least square method and the NMIJ the Deming method. The NPL and IRA used a Monte-Carlo approach to extrapolate the data points and deduce the associated uncertainties.

The average of the intercept and slope distributions were taken for the  $y$ -intercept and slope, and the distribution standard deviations were taken for the associated uncertainties. Some participants used commercial programs, such as Origin, Labfit and XLGenLine, but most of them used homemade calculation codes or a mix of both.

The combination of the values of gates 1 and 2 was generally done using the arithmetic or weighted mean. The KRISSE chose the  $y$ -intercept of gate 1 for the activity value and estimated the uncertainty associated with the extrapolation procedure by the difference between the results of gates 1 and 2. The NMISA as well as the CIEMAT fitted the data of gates 1 and 2 all together. To evaluate the uncertainty of the combined  $y$ -intercepts value, a quadratic sum was calculated by the IFIN-HH, NPL and IRA. The BIPM added quadratically the contribution due to dead times, resolving times, backgrounds, and counting statistics with the extrapolation uncertainty prior to combining the results from the two gates. This approach was chosen by the IRA for the alternative exercise. Two laboratories, the BIPM and VNIIM, evaluated the uncertainty taking into account the correlations of the provided data. The former used the correlation coefficient of the average and the latter chose the lower uncertainty of gates 1 and 2. The NMISA took the difference between the values of gates 1 and 2 divided by two times the square root of three.

Four laboratories (IFIN-HH, LNMRIIRD, NPL and IRA) reported an alternative extrapolation procedure to calculate the activity concentration, see Table 5. The IFIN-HH used homemade software to perform the alternative extrapolation, which is based on a least square fitting algorithm. The LNMRI/IRD chose to discard one source, and then calculated as in the required exercise. At the NPL, an alternative procedure consisted of using the software package XLGenLine v1.0. This software utilizes generalized least squares polynomial fitting. Finally, at the IRA, the associated  $y$ -uncertainties was initially recalculated taking into account the mass uncertainty for each source. Afterward, the evaluations were performed using the MC method as in the required exercise.

### 3.3 Weighing Uncertainty Evaluation

The method that each laboratory used to evaluate the uncertainty due to weighing is summarized in Table 6. Four participants used the balance certificate provided by the NPL and eight preferred the certificate obtained by applying the DKD calibration procedure. The evaluation of the uncertainty is different if one weighs using the direct reading or the substitution method. Therefore, the results are treated separately in the following sections.

#### 3.3.1 Direct reading method

The IFIN-HH and NMISA estimated the contribution to the relative standard uncertainty due to weighing by combining three components due to: drift, balance calibration and buoyancy correction. The IFIN-HH considered the maximum applied correction to be the drift standard uncertainty, but the NMISA preferred taking, for each source, the average drift divided by  $2\sqrt{3}$  and expressing it as a percentage of the source mass to give the relative uncertainty due to drift. Both laboratories estimated the maximum uncertainty of the reading using the DKD-R7-1 formula. For the buoyancy contribution, the NMISA varied each provided components by one standard deviation to increase the buoyancy correction. For each component, the increase in the buoyancy correction was expressed as a percentage of the unvaried buoyancy correction. For each source, the relative values were quadratically summed to

obtain the contribution due to the buoyancy correction. The NMISA relative standard uncertainty due to weighing was calculated using the weighted average (weighted according to the number of data points collected per source) of the total weighing uncertainties for each of the ten sources. The IFIN-HH directly combined their three contributions quadratically.

The NMIJ neglected the contribution due to the buoyancy correction and took the maximum value of the reading uncertainty using the DKD-R7-1 formula for the total weighing uncertainty.

At the CIEMAT, the uncertainty contribution of the balance repeatability was evaluated and, because of its small contribution, they took the balance calibration uncertainty contribution using the NPL certificate as the total standard uncertainty attributable to weighing.

The KRISS estimated the contribution to the relative standard uncertainty due to weighing by quadratically combining five components attributable to: buoyancy correction, repeatability, balance calibration uncertainty, eccentricity and linearity.

At the BIPM, the uncertainty due to repeatability, eccentricity and rounding was evaluated from the results of two consecutive weighings of the same mass. The uncertainty contribution due to drift was taken as the standard deviation of two consecutive tares. The uncertainty on the buoyancy correction was obtained using the drift in temperature, atmospheric pressure and hygrometry during weighing. The uncertainty due to balance non-linearity in the range of measurements was judged negligible from the value of the NPL certificate. Finally, taking into account an evaporation rate obtained from literature and supposing a 5 min interval between two consecutive weighing, a relative uncertainty was calculated.

The VNIIM estimated the contribution to the relative standard uncertainty due to weighing as the relative standard uncertainty of the mean mass. The value was obtained by dividing the balance calibration uncertainty value divided by the average mass of the sources and the square root of the number of sources.

The IRA estimated the relative standard uncertainty due to weighing by using a conservative approach. The uncertainty adopted was the one of the lightest source included in the efficiency extrapolation, which is the source with the largest relative uncertainty. They estimated the contribution to the relative standard uncertainty by quadratically combining two main components: the buoyancy correction uncertainty and the balance calibration uncertainty using the DKD-R1-7 formula.

### **3.3.2 Substitution method**

At the ANSTO, each sample mass uncertainty was obtained by quadratically combining uncertainties on the calibration substitution weights before and after deposition, using the DKD 7-1 data as well as the uncertainty of the standard weights. The square of the relative uncertainty was then obtained as the weighted average of the squared individual sample mass uncertainties. The individual weights were chosen as the inverse of the variance of the activity values determined source by source.

At the INER, the weighing uncertainty was a combination of the uncertainties which were derived from the DKD 7-1 formula for each mass and the uncertainty of the standard weights.

The LNMRI/IRD obtained the standard uncertainty attributable to weighing by taking into account the largest mass and calculating the balance calibration uncertainty.

The NPL neglected the buoyancy uncertainty contribution, but estimated the relative uncertainty by quadratically combining the uncertainty on each substitution weight. The relative uncertainty on any single mass measurement was found between 0.2 % and 0.4 %. However, because they estimated that the source masses enter into the efficiency extrapolation, and thus create scatter in the data sets, they decided to assign an estimate of the overall relative uncertainty on the weighing procedure of 0.15 %.

## 4 Discussion

### 4.1 Activity Data

The values of the activity concentration are all compatible and the relative standard deviation of the activity concentration is 0.03 %. In light of the several different methods used for evaluating the activity concentration, it shows how robust the coincidence counting method is.

The average standard uncertainty is 0.62 kBq/g with a relative standard deviation of 38 %, which is significantly larger than the relative standard deviation of the activity concentration. Such an observation motivates these type of interlaboratory comparisons. The spread of the total uncertainty and its separate contributions are graphically represented in Figure 6. The spread shows more discrepancies for the separate contributions than for the activity. The spread is especially large for the contributions attributable to weighing, resolving times, counting statistics and half-life. Not unexpectedly, the spread of the uncertainty on the activity is smaller, which indicates that the deviating contributions are compensated by other ones. To further synthesized the data, we report the average, minimum and maximum uncertainty reported by the participants for each main contribution (

Table 7). The ratio of the maximum to the minimum value was calculated and found to be quiet varied. For contributions due to extrapolation, dead times and background, this ratio is lower than 10, while for the half-life contribution, it is 250. The ratio of the maximum and minimum uncertainty on the activity concentration is around 3. This clearly confirms that the uncertainties reported are strongly dependent on the laboratory. The large discrepancy of the half-life contribution is not important in this particular case due to the relatively long half life of the nuclide. However, when measuring short-lived radionuclides, which are common in nuclear medicine, this contribution will be more important and will carry more weight in the uncertainty budget.

The determination of the type of evaluation (A, B or A+B) remains a subjective matter and shows the limitations of the GUM at the laboratory level. An agreement was reached on certain contributions, but the laboratories perceive most contributions differently and, therefore, reported different types of evaluation.

### 4.2 Extrapolation Data

The associated relative standard uncertainties of the extrapolated y-intercepts are small and show that the y-intercept value is well determined, see Figure 4. The associated relative standard uncertainties of the slopes for gates 1 and 2 are large and reach some dozen of percents (Figure 5).

Figure 7 shows the extrapolation results for gates 1 and 2 superimposed on the provided experimental data, as well as the average slope and the minimum and maximum curves. Although all participants used the exact same data, the linear fits are very different. However, the combined y-intercepts are not very sensitive on the choice of the slope, and its average standard uncertainty is 0.29 kBq/g. The y-intercept standard uncertainties is 0.16 kBq/g and 0.14 kBq/g for gates 1 and 2 respectively. The ratio of the maximum to the minimum relative standard uncertainty is about 10 for the y-intercepts, see Table 4. The slope ratio is around 7 for gate 1 and 18 for gate 2. Even if the slope values showed more discrepancy than the y-intercepts, the slope ratio of the maximum to the minimum relative standard uncertainty is not systematically larger.

Most laboratories did not evaluate the combined y-intercepts uncertainty by direct uncertainty propagation using the equation of the arithmetic or weighted average. This choice is practical

and is motivated by the fact that the two extrapolated y-intercepts are not compatible. The combined value ratio of the maximum to the minimum relative standard uncertainty is 144, see Table 4. In comparison, this ratio is under 10 for gate 1 and under 20 for gate 2. This large discrepancy is due to the different uncertainty evaluation strategies used.

The average standard uncertainty on the final reported activity concentration is nearly the double of the average standard uncertainty on the combined y-intercepts, which is due to the additional contributions on the final reported uncertainty (also visible when comparing Figure 2 and Figure 8).

The results show that the robustness of the method lies in the evaluation of the y-intercept. In order to preserve the method's strength, the extrapolation has to be performed not too far from the experimental data points, which makes the slope determination less influential.

The relative standard uncertainties of  $\rho_\beta/\varepsilon_\beta$  as a function of  $(1-\varepsilon_\beta)/\varepsilon_\beta$  are shown in Figure 10. The relative uncertainties are independent of  $(1-\varepsilon_\beta)/\varepsilon_\beta$ , which means that the weighting of the data is homogeneous across the x-axis. In order to refine the analysis of the extrapolation, the residues were calculated.

Figure 11 shows the residues calculated for the average extrapolation. Next, the distribution of the residues was calculated and is also shown in

Figure 11 for the average. Figure shows the distribution obtained for each participant and gate. Because the y-uncertainties don't depend on the x-values, see Figure 10, the distribution of residues is expected to center on zero and be symmetrical. However, the participant residue distributions were not always centered on zero. The fitting procedure of some laboratories (IFIN-HH, NMISA, NMIJ, and CIEMAT) induced a shift of the residue distribution towards 0.4 for gate 2 and 0.2 for gate 1. No particularities regarding the fitting procedure of these laboratories could be identified, but some weighting choices must have been made differently. Some laboratories may have removed some data points they considered outliers and did not indicate which ones. It is likely that some homemade calculation codes include statistical tests which remove points from the fitting procedure, and not always to the knowledge of the users.

Four laboratories chose to use an alternative extrapolation method and the maximum difference between the two methods was 0.2 % (for the LNMRI/IRD). The alternative methods did not significantly influence the resulting activity concentrations values, which is partly due to the robustness of the coincidence counting method. Indeed, the participants found very different values for the slope, but not for the y-intercept.

### 4.3 Weighing Uncertainty Evaluation

Figure 9 shows the value of the evaluated uncertainty labeled with the weighing methods and the chosen certificate type. No strong correlation between the weighing method or the certificate and the results was found. However, the uncertainty evaluated using the DKD certificate seems slightly larger. The participants reported comparable values with the exception of the LNMRI/IRD, which estimated the uncertainty to be much smaller. The ratio between the maximum and minimum uncertainty value is 26, which is not surprising in light of the different evaluation methods. Nevertheless, this contribution carries strongly influences the uncertainty budget and needs to be evaluated carefully.

## 5 Conclusions

This interlaboratory comparison helped define the state of the art of a specific measurement method, and could probably be extended to other techniques. With 12 participants, its success reflects the openness of the radionuclide metrology community.

Thanks to the robustness of the coincidence counting method and in spite of contrasting evaluating methods, the values of the estimated activity concentrations were compatible.

The reported uncertainty values diverged by as much as 250 fold for certain contributions and were often not compatible. At an even more basic level, we noticed that the participants labeled the type of evaluation (A, B or A+B) very differently. All this indicates that the GUM may not be sufficiently clear on this matter or may not be uniformly understood at the laboratory level.

The extrapolation results, obtained with all participants sharing the same raw data, are surprisingly different. This was explained partially by the use of different algorithms and by procedures that disregard specific data points. It shows the importance of knowing what the data processing tools are doing in general, and, in this case, how they consider outliers.

The participants proposed various solutions to combine the extrapolated values for gates 1 and 2. Counter-intuitively, combining two values with known uncertainties was not a trivial and standardized operation, which indicates incomplete harmonization of procedures.

In conclusion, we showed that the uncertainty evaluation can be very diverse, even when all participants share the same raw data from the same instrument. Clearly, participants have different interpretations of the uncertainties. A standardized approach to uncertainty evaluation might be beneficial, but would be complex. However, if the uncertainties were to be considered in computing reference values, such as key comparison reference values, the present study would need to be extended. The community might design recommendations leading to better uniformity and higher confidence. Focusing on the most significant contributions to the uncertainty would help to reach consensus efficiently. Clearly, the way uncertainties are currently reported is not sufficient to understand how they are evaluated, nor to what they are precisely referring. Some progress could readily be made by encouraging laboratories to provide detailed descriptions of the contributions to the reported uncertainty.

## 6 Tables

Table 1: List of participants to the interlaboratory comparison

Acronym	Full name	Acronym	Full name
<b>IFIN-HH</b>	"Horia Hulubei" National Institute of Research and Development for Physics and Nuclear Engineering, Romania	<b>LNMRI/IRD</b>	Laboratório Nacional de Metrologia das Radiações Ionizantes, Instituto de Radioproteção e Dosimetria, Brazil
<b>ANSTO</b>	Australian Nuclear Science and Technology Organisation, Australia	<b>KRISS</b>	Korea Research Institute of Standards and Science, Republic of Korea
<b>NMISA</b>	National Metrology Institute of South Africa, South Africa	<b>BIPM</b>	International Bureau of Weights and Measures/Bureau International des Poids et Mesures, France
<b>INER</b>	Institute of Nuclear Energy Research, Atomic Energy Council, Executive Yuan, Taiwan	<b>VNIIM</b>	D.I. Mendeleyev Institute for Metrology, Rostekhnregulirovaniye of Russia, Russian Federation
<b>NMIJ</b>	National Metrology Institute of Japan, National Institute of Advanced Industrial Science and Technology, Japan	<b>IRA</b>	Institute of Radiation Physics, Switzerland
<b>CIEMAT</b>	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain	<b>NPL</b>	National Physical Laboratory, United Kingdom

Table 2: Values reported by the participants as final results and the associated uncertainty contributions.

	Activity Concentration [kBq/g]	Total Standard Uncertainty [kBq/g]	Total Relative Standard Uncertainty [%]	Contribution due to [ $u_i \times 10^4$ ]							
				Weighing	Extrapolation	Dead times	Resolving times	Backgrounds	Counting Statistics	Half-life	Others
IFIN-HH	354.62	0.6	0.17	13.6	9.1	2.5	0.9	1.2	1.4	0.13	
ANSTO	354.57	0.38	0.11	10	3						
NMISA	354.59	0.67	0.19	16.69	8.682	5	5	6	5	0.008	Afterpulses: 5
INER	354.62	0.50	0.14	9.4	2.3	8.1	0.6	0.8	5.6	0.02	
NMIJ	354.30	0.50	0.14	12.8	3.3						
CIEMAT	354.37	0.61	0.17	13	10	3.6	0.05	3.8	0.2	0.007	
LNMRI/IRD	354.6	1	0.28	0.99	3.55	11.1	2	0.6	3	0.004	Delay mismatch: 24.9
KRISS	354.38	0.75	0.21	6.282	15.802	4.5	0.7	10	6.013	0.01	Timing: 1
BIPM	354.7	0.4	0.12	6	9					0.011	Out of channel events for gate 1: 5
VNIIM	354.547	0.362	0.102	5.42	3.2	7.43	0.43	3.05	0.38		
IRA	354.614	1.110	0.313	25.7	11.1	12.5	3.4	1.0	5.0	0.01	Timing: 0.02
NPL	354.5	0.6	0.18	15	7.3	4	1	2	3	1	Pile-up: 2
<b>Average</b>	354.53	0.62	0.18	11.77	7.19	6.53	1.56	2.31	3.29	0.13	
<b>Std dev.</b>	0.12	0.23	0.07	6.50	4.18	3.49	1.63	1.88	2.25	0.33	
<b>Rel. std dev. (%)</b>	0.03	37.65	36.74	55.26	58.11	53.50	104.50	81.34	68.31	245.58	

Table 3: Type of uncertainty evaluation reported by the participants.

	Contribution due to									Out of channel events for gate 1
	Weighing	Extrapolation	Dead times	Resolving times	Backgrounds	Counting Statistics	Half-life	Delay mismatch	Timing	
IFIN-HH	B	A+B	B	B	A+B	A	B			
ANSTO	B	A+B								
NMISA	B	B	B	B	B	A	B		B	
INER	B	A	B	B	A	A	B			
NMIJ	A+B	A								
CIEMAT	B	B	B	B	A	A	B			
LNMRI/IRD	B	A	B	B	A	A	B	B		
KRISS	A+B	B	A+B		A+B	A	B		B	
BIPM	A+B	A+B	B	B	A	A	B			B
VNIIM	B	A+B	B	B	B	B				
IRA	B	A	B	B	B	A	B		B	
NPL	B	A+B	B	B	B	B	B			

Table 4: Extrapolation results reported by the various participants for the required exercise.

	IFIN-HH	ANSTO	NMISA	INER	NMIJ	CIEMAT	LMNRI/IRD	KRISS	BIPM	VNIM	IRA	NPL	Ratio of maximum and minimum
<b>Gate 1</b>													
Y [kBq/g]	354.34	354.01	354.01	354.34	354.30	354.02	354.014	354.38	354.39	354.014	354.014	354.34	1
Y Standard Uncertainty [kBq/g]	0.23	0.12	0.05	0.12	0.12	0.55	0.0486	0.21	0.19	0.118	0.049	0.073	11
Y Relative Standard Uncertainty [ $u_i \times 10^4$ ]	6.5	3.4	1.373	3.3	3.3	15	1.37	6.013	5	3.3	1.4	2.6	11
Slope [kBq/g]	1.24	2.47	2.47	1.25	1.472	1.36	2.470	0.9198	1.0	2.47	2.474	1.25	3
Slope Standard Uncertainty [kBq/g]	0.71	0.83	0.34	0.72	0.724	0.45	0.339	0.8075	0.9	0.83	0.338	0.47	3
Slope Relative Standard Uncertainty [ $u_i \times 10^4$ ]	5726	3360	1372	5760	4915	3350	1372	8779	9000	3370	1364.5	3760	7
<b>Gate 2</b>													
Y [kBq/g]	354.90	355.08	355.08	354.90	-	354.60	355.081	354.94	354.88	355.081	355.080	354.90	1
Y Standard Uncertainty [kBq/g]	0.23	0.11	0.04	0.12	-	0.43	0.0418	0.17	0.15	0.114	0.042	0.108	11
Y Relative Standard Uncertainty [ $u_i \times 10^4$ ]	6.4	3.1	1.178	3.2	-	12	1.18	4.928	4	3.2	1.2	8.6	10
Slope [kBq/g]	-0.53	-2.39	-2.38	-0.55	-	-0.47	-2.383	-0.7983	-0.4	-2.38	-2.382	-0.54	6
Slope Standard Uncertainty [kBq/g]	0.71	0.74	0.27	0.69	-	0.46	0.269	0.9050	0.8	0.74	0.269	0.37	3
Slope Relative Standard Uncertainty [ $u_i \times 10^4$ ]	1340	3096	1130	1255	-	4588	1130	11336	20000	3092	1129.2	6850	18
<b>Combining gate 1 and 2</b>													
Activity concentration [kBq/g]	354.62	354.57	354.59	354.62	354.30	354.37	354.0	354.38	354.70	354.547	354.547	354.52	1
Standard Uncertainty [kBq/g]	0.32	0.38	0.31	0.082	0.12	0.35	0.07	0.56	0.33	0.114	0.533	0.26	8
Relative Standard Uncertainty [ $u_i \times 10^4$ ]	9.1	0.11	8.682	2.3	3.3	10	1.9	15.802	9	3.2	15.0	7.3	144

Table 5: Extrapolation results reported by the four participants for the alternative extrapolation procedure.

	IFIN-HH	LNMRI/IRD	IRA	NPL
<b>Gate 1</b>				
Y [kBq/g]	354.02	354.4	354.219	354.01
Y Standard Uncertainty [kBq/g]	0.25	0.18	0.219	0.05
Y Relative Standard Uncertainty [ $u_i \times 10^4$ ]	7.1	505	6.2	1.4
Slope [kBq/g]	2.48	-	1.783	2.473
Slope Standard Uncertainty [kBq/g]	0.71	-	1.444	0.338
Slope Relative Standard Uncertainty [ $u_i \times 10^4$ ]	2863	-	8102.1	1370
<b>Gate 2</b>				
Y [kBq/g]	355.08	354.9	355.008	355.08
Y Standard Uncertainty [kBq/g]	0.25	0.16	0.223	0.04
Y Relative Standard Uncertainty [ $u_i \times 10^4$ ]	7.0	440	6.3	1.18
Slope [kBq/g]	-2.38	-	-1.092	-2.386
Slope Standard Uncertainty [kBq/g]	0.71	-	1.447	0.269
Slope Relative Standard Uncertainty [ $u_i \times 10^4$ ]	2983	-	13256.5	1130
<b>Combining gate 1 and 2</b>				
Activity concentration [kBq/g]	354.55	354.6	354.614	354.62
Standard Uncertainty [kBq/g]	0.53	0.13	0.395	0.53
Relative Standard Uncertainty [ $u_i \times 10^4$ ]	15	355	11.1	14.3

Table 6: Weighing method and uncertainty evaluation type reported by the various participants.

	<b>Weighing method</b>	<b>Balance certificate</b>	<b>Weighting uncertainty evaluation type</b>
<b>IFIN-HH</b>	Direct reading	DKD	B
<b>ANSTO</b>	Substitution	DKD	A
<b>NMISA</b>	Direct reading	DKD	B
<b>INER</b>	Substitution	DKD	B
<b>NMIJ</b>	Direct reading	DKD	A+B
<b>CIEMAT</b>	Direct reading	NPL	B
<b>LNMRI/IRD</b>	Substitution	DKD	B
<b>KRISS</b>	Direct reading	NPL	A+B
<b>BIPM</b>	Direct reading	NPL	A+B
<b>VNIIM</b>	Direct reading	NPL	B
<b>IRA</b>	Direct reading	DKD	B
<b>NPL</b>	Substitution	DKD	B

Table 7: Average, minimum and maximum values reported for the main uncertainty contributions to the activity concentration.

	Contribution due to [ $u_i \times 10^4$ ]						
	Weighing	Extrapolation	Dead times	Resolving times	Backgrounds	Counting Statistics	Half-life
<b>Average</b>	11.8	7.2	6.5	1.6	2.3	3.3	0.13
<b>Std dev.</b>	6.5	4.2	3.5	1.6	1.9	2.3	0.3
<b>Rel. std dev. (%)</b>	55.3	58.1	53.5	104.5	81.3	68.3	245.6
<b>Minimum value</b>	0.99	2.3	2.5	0.05	0.6	0.2	0.004
<b>Maximum value</b>	25.7	15.8	12.5	5.0	6.0	6.0	1.0
<b>Ratio of the maximum and minimum value</b>	26.0	6.9	5.0	100.0	10.0	30.1	250.0

# 7 Figures

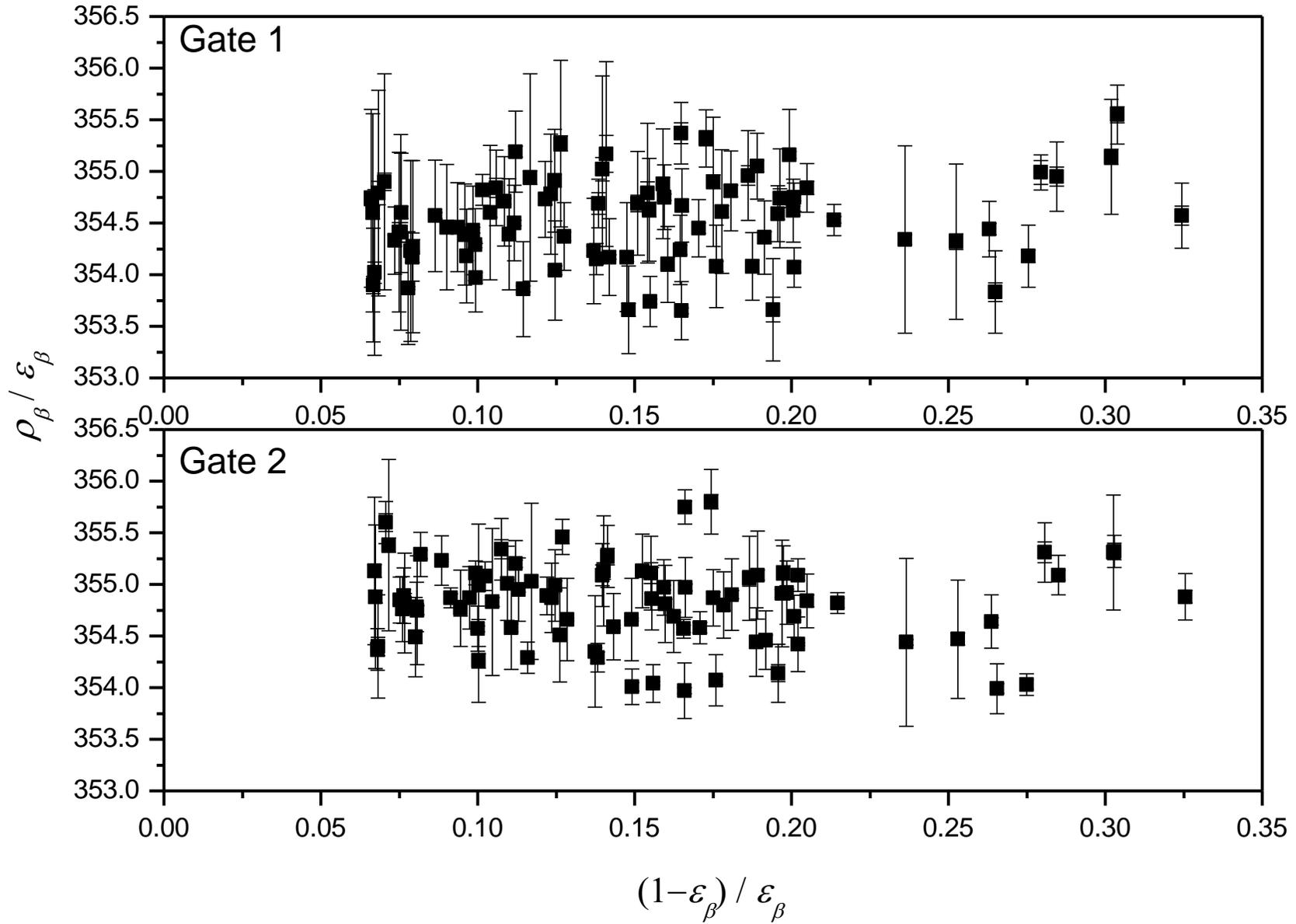


Figure 1: NPL Co-60 coincidence counting data

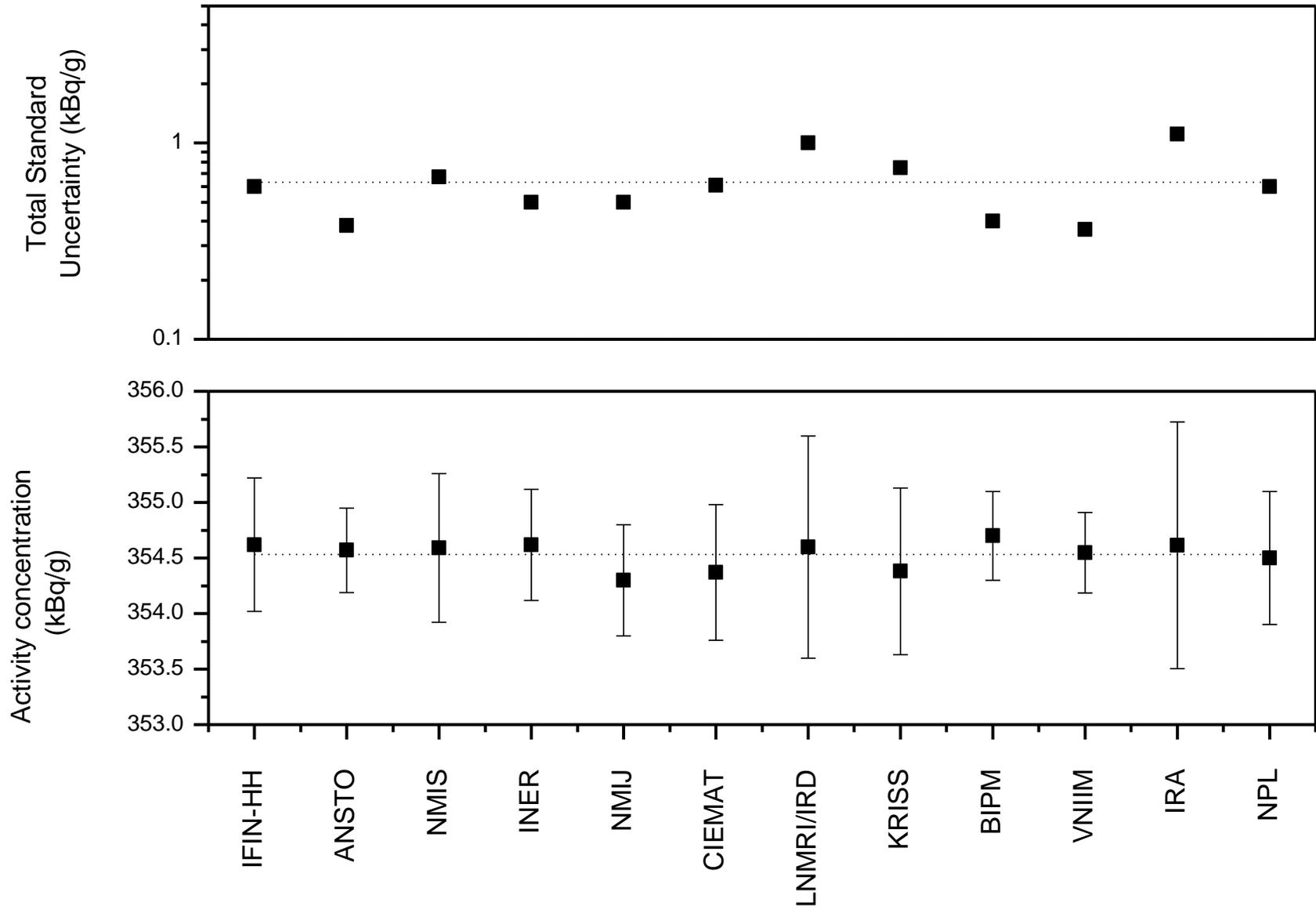


Figure 2: Plot of the final result values obtained by each laboratory.

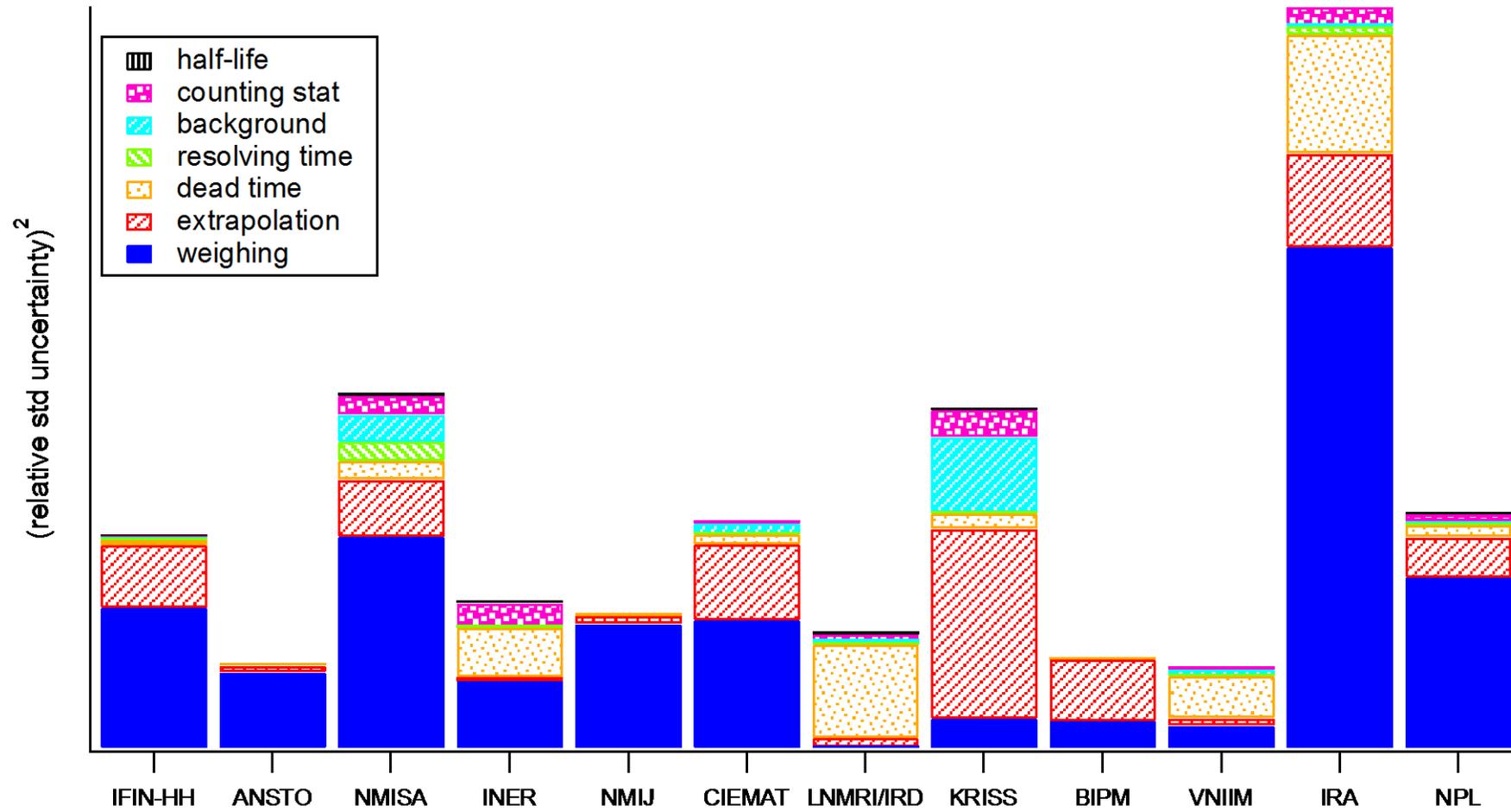


Figure 3: Square of the relative-uncertainty contributions for each laboratory.

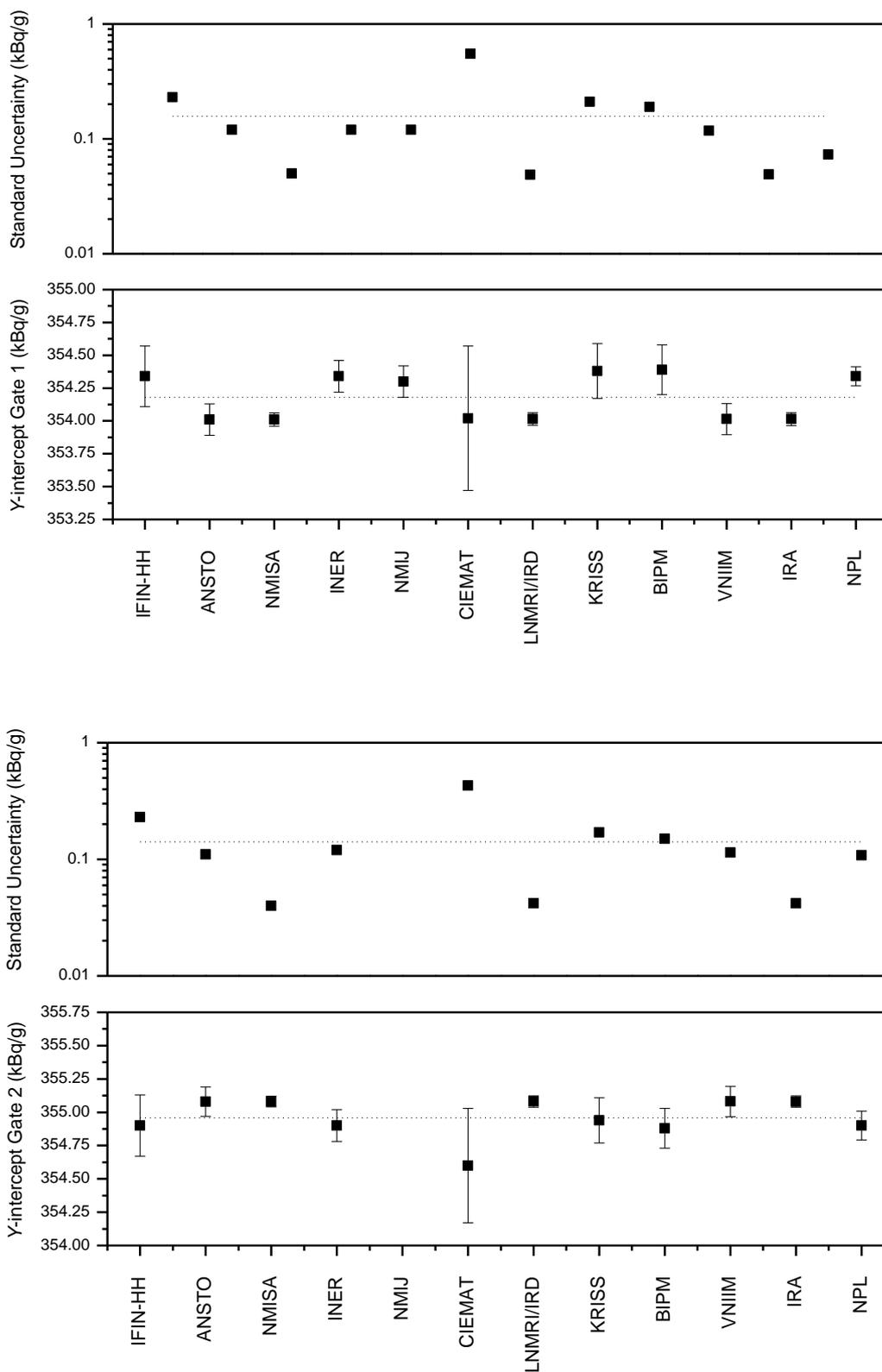


Figure 4: Plots of the extrapolated y-intercepts for gates 1 and 2 reported by each participant.

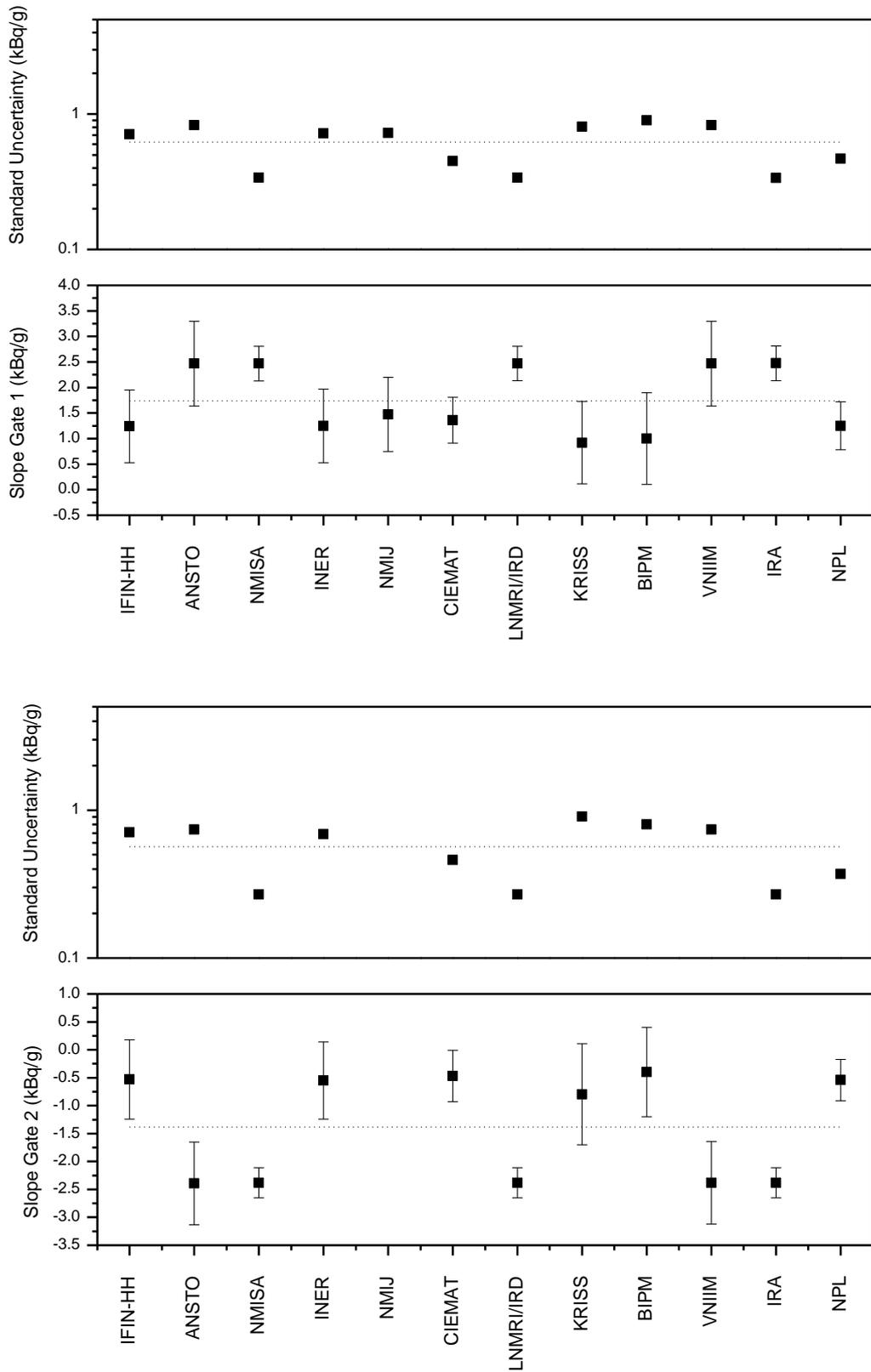


Figure 5: Plots of the reported extrapolated slopes for gates 1 and 2

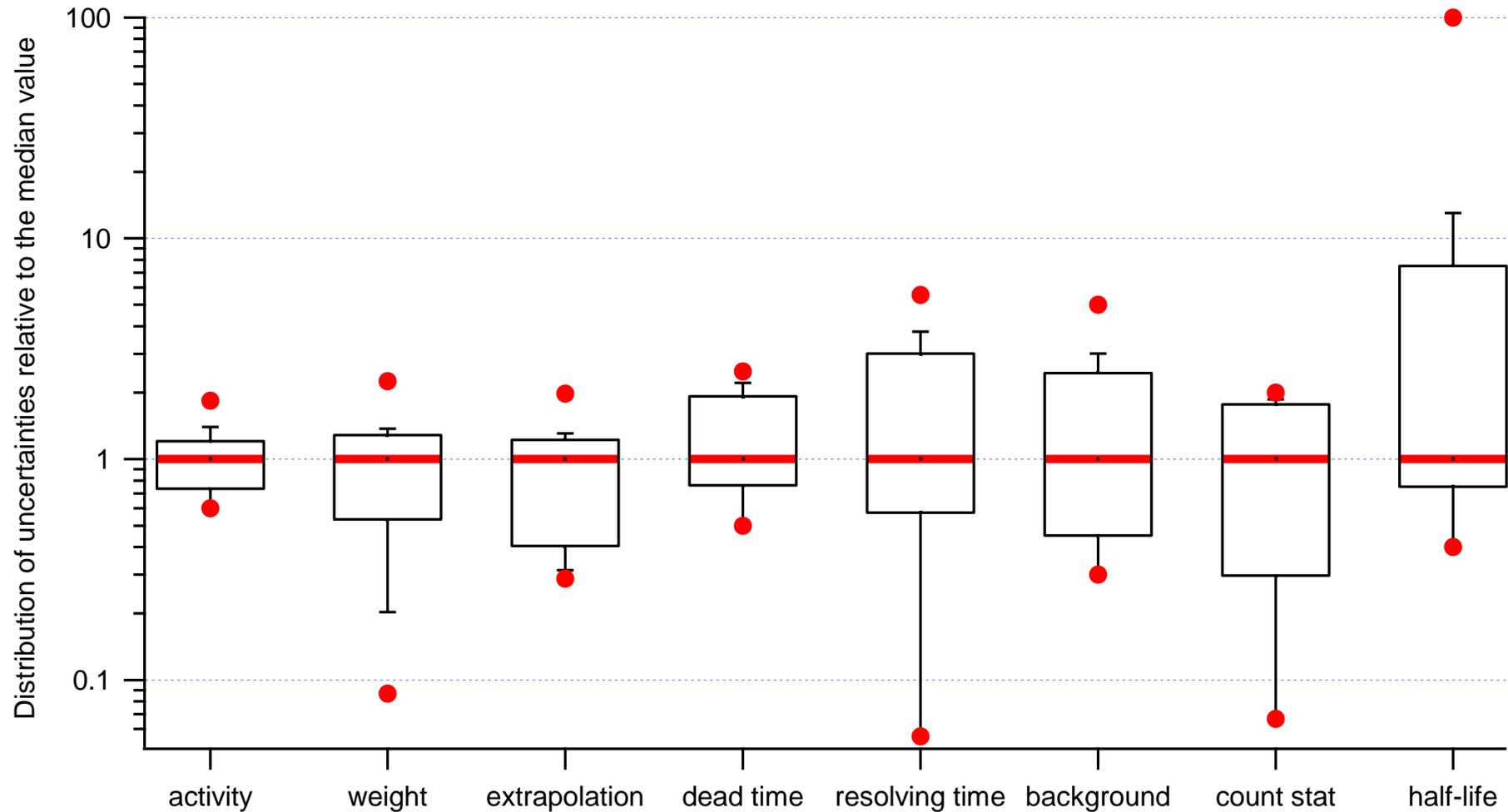


Figure 6: Box-plot showing the spread of the uncertainties estimations. The most left box-plot shows the total uncertainty estimation. The next ones present the different contributions of the uncertainty budget. Each uncertainty value has been normalized by its median value. The central horizontal bold line is the median (set equal to 1). The boxes represent the 25th and the 75th percentiles. The whiskers represent the 10th and the 90th percentiles. The full circles are the points out of the 10th – 90th percentile-interval.

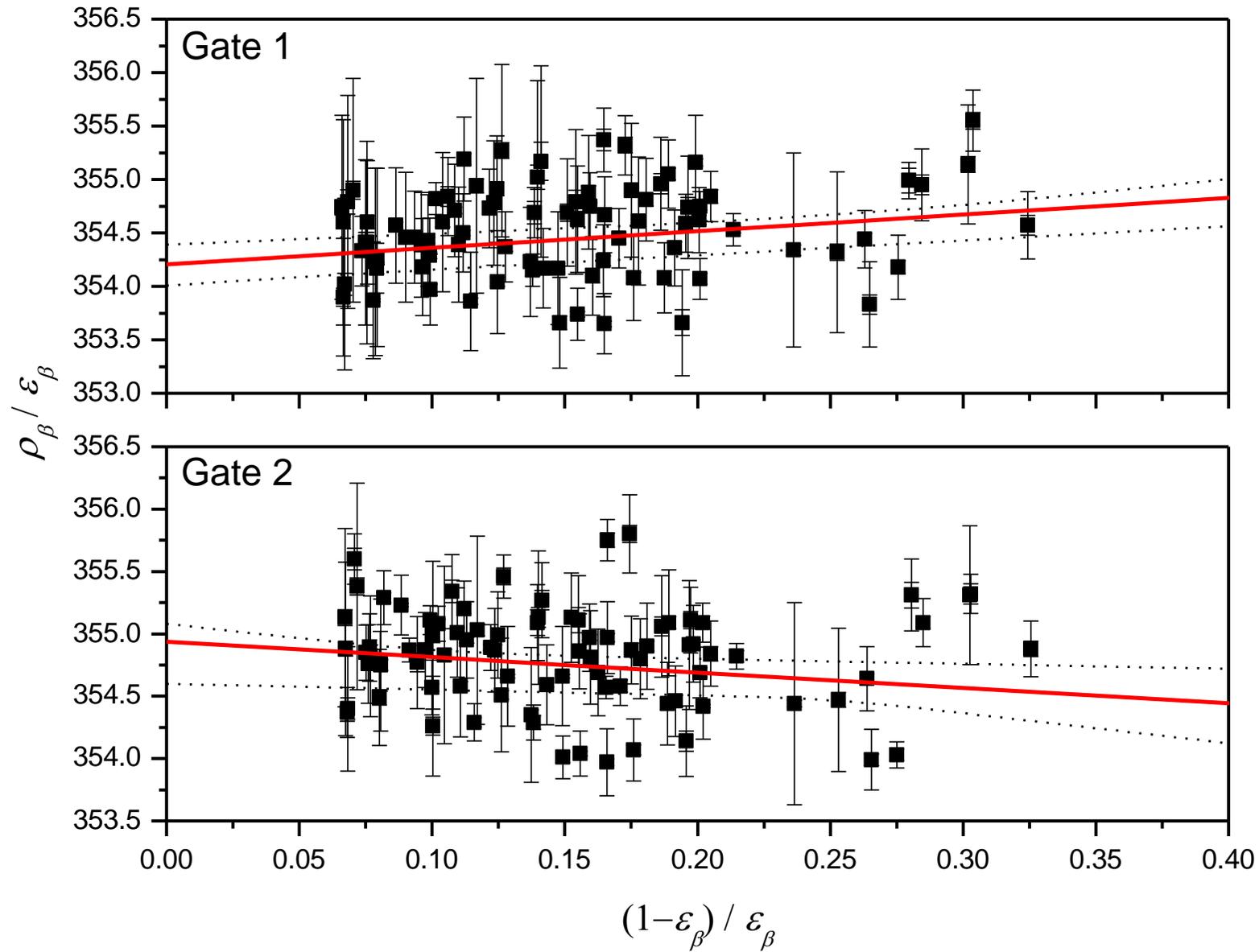


Figure 7: Plot of the extrapolation results for gates 1 and 2 superimposed on the provided experimental data. It shows the average extrapolation results with the minimum and maximum obtained by the participants.

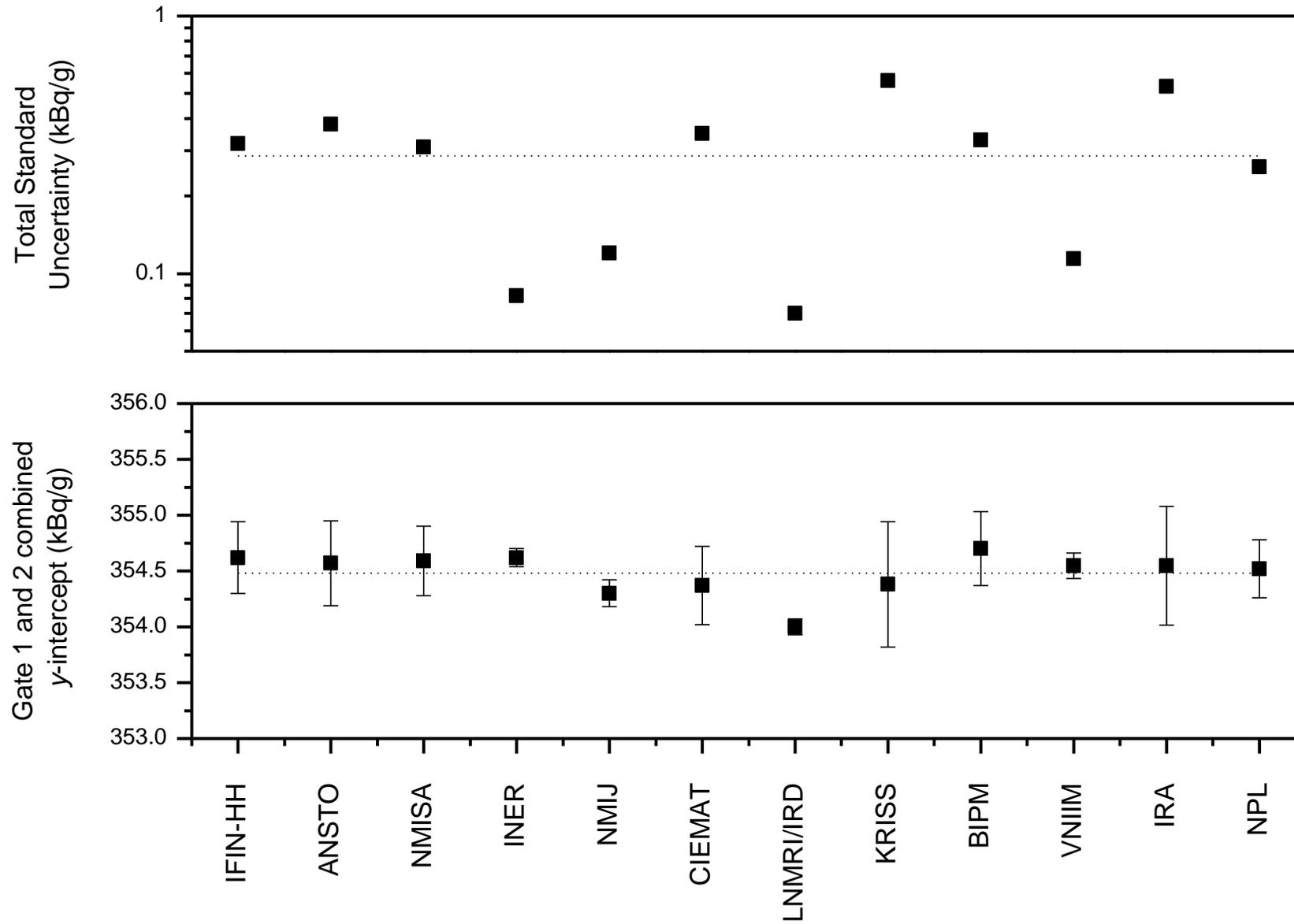


Figure 8: Plot of the reported y-intercept combining the results of gates 1 and 2.

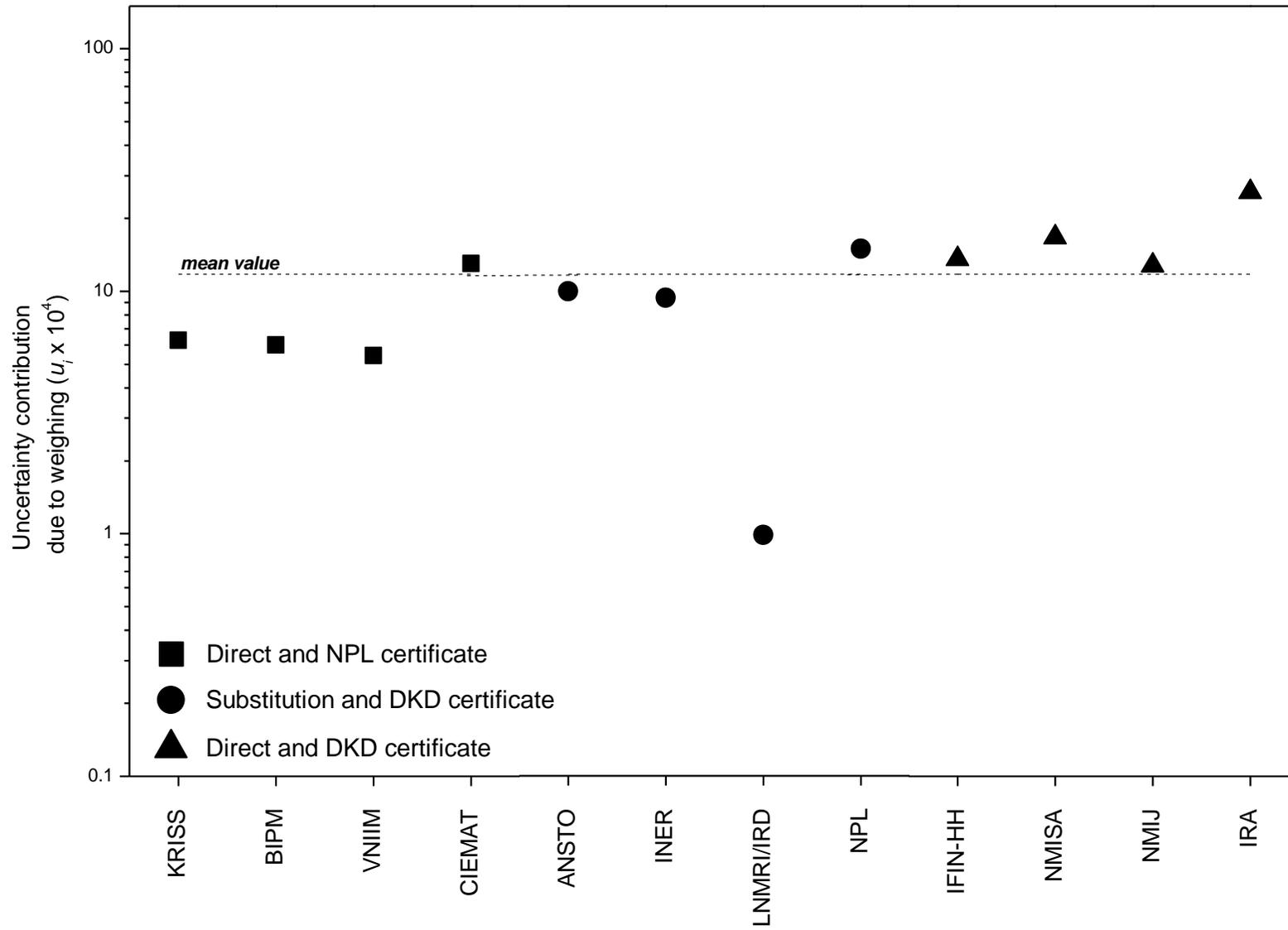


Figure 9: Plot of the uncertainty contribution due to weighing evaluated by each participant.

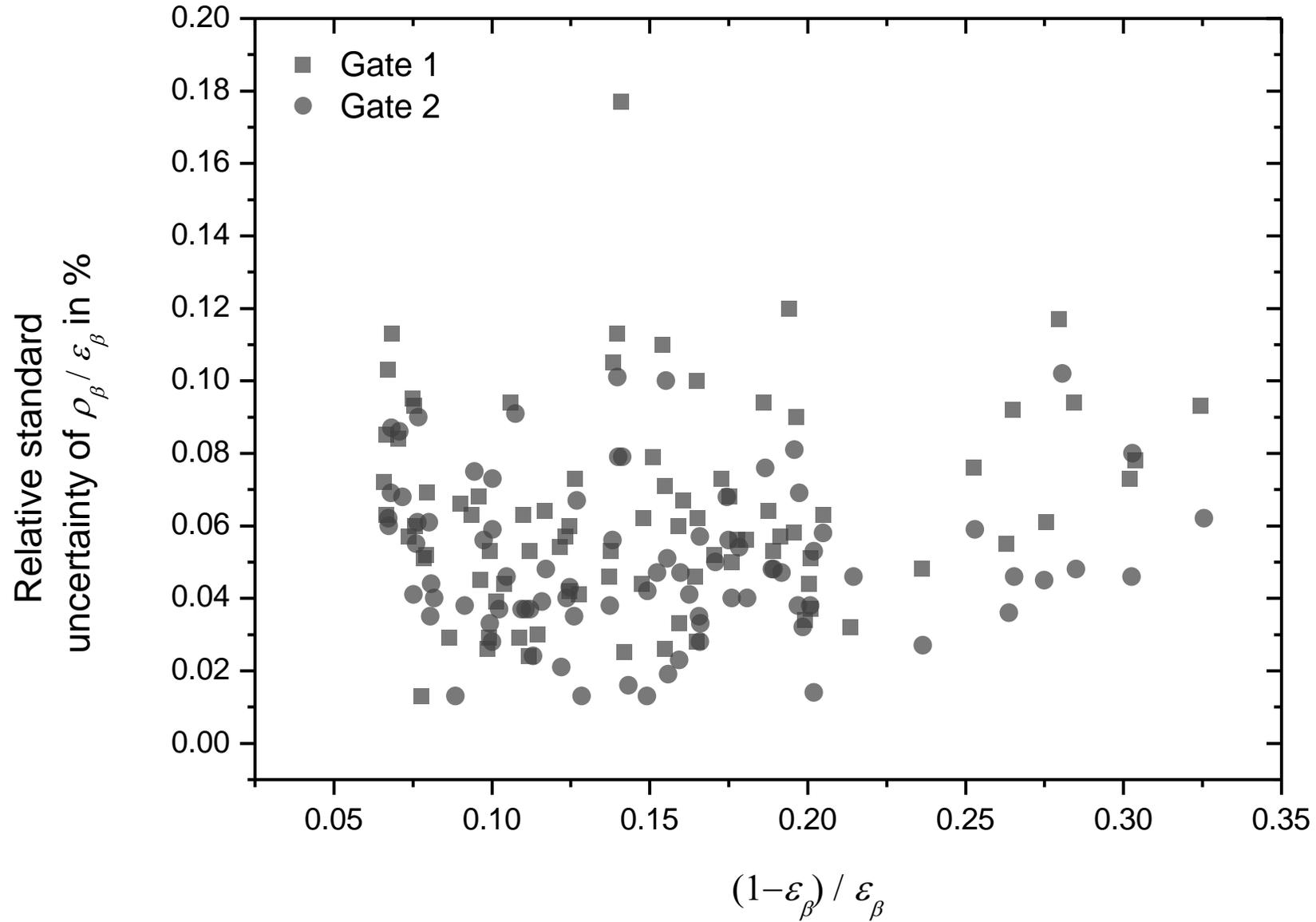


Figure 10: Relative standard uncertainty of  $\rho_\beta/\epsilon_\beta$  as a function of  $(1-\epsilon_\beta)/\epsilon_\beta$ .

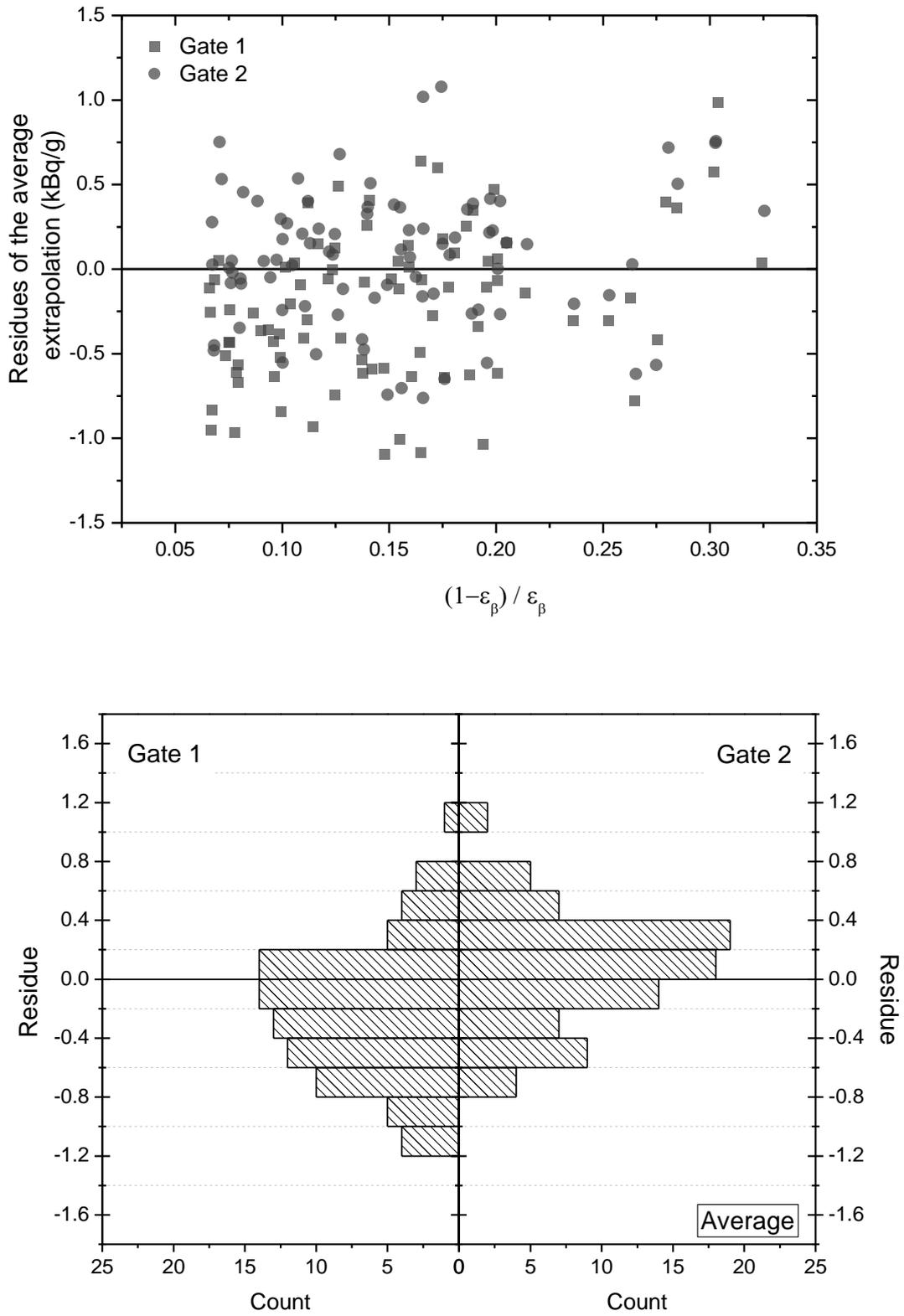


Figure 11: Residues of the average extrapolation as a function of  $(1-\epsilon_\beta)/\epsilon_\beta$  and its distribution.

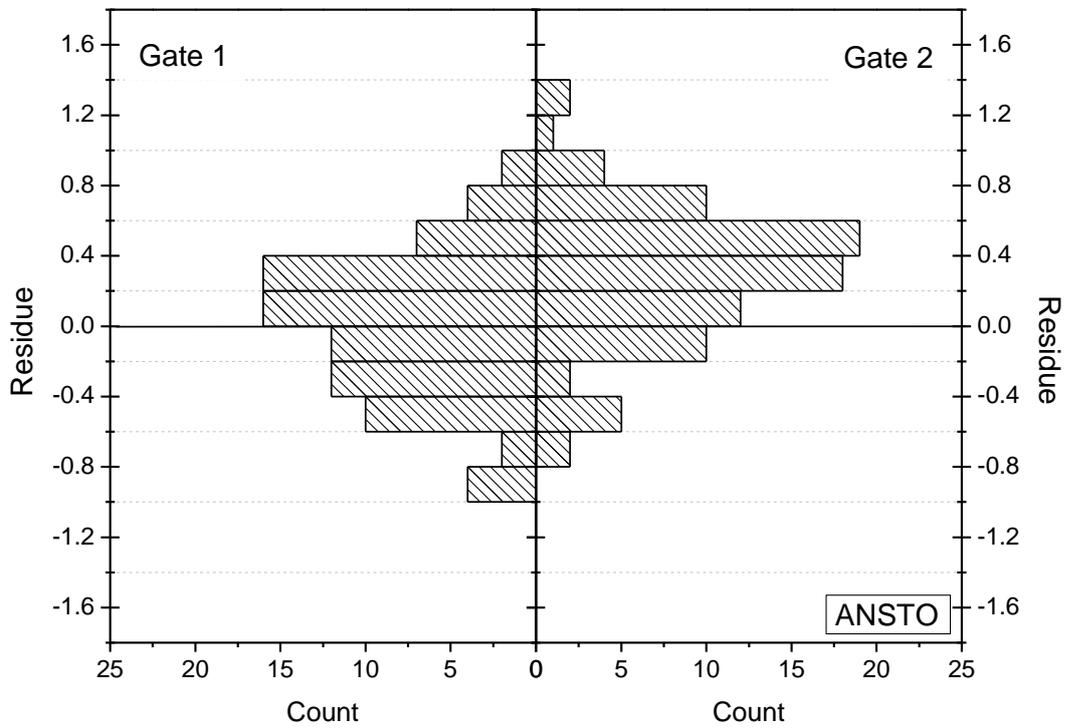
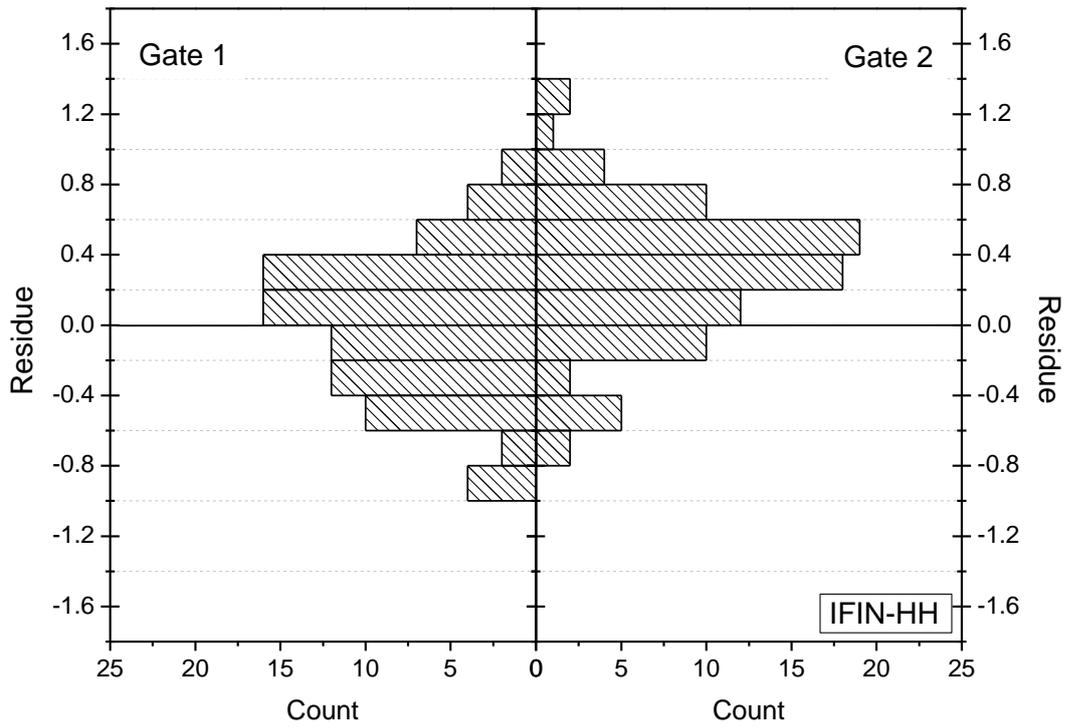


Figure 12: Distribution of the residues for each participant (1/6).

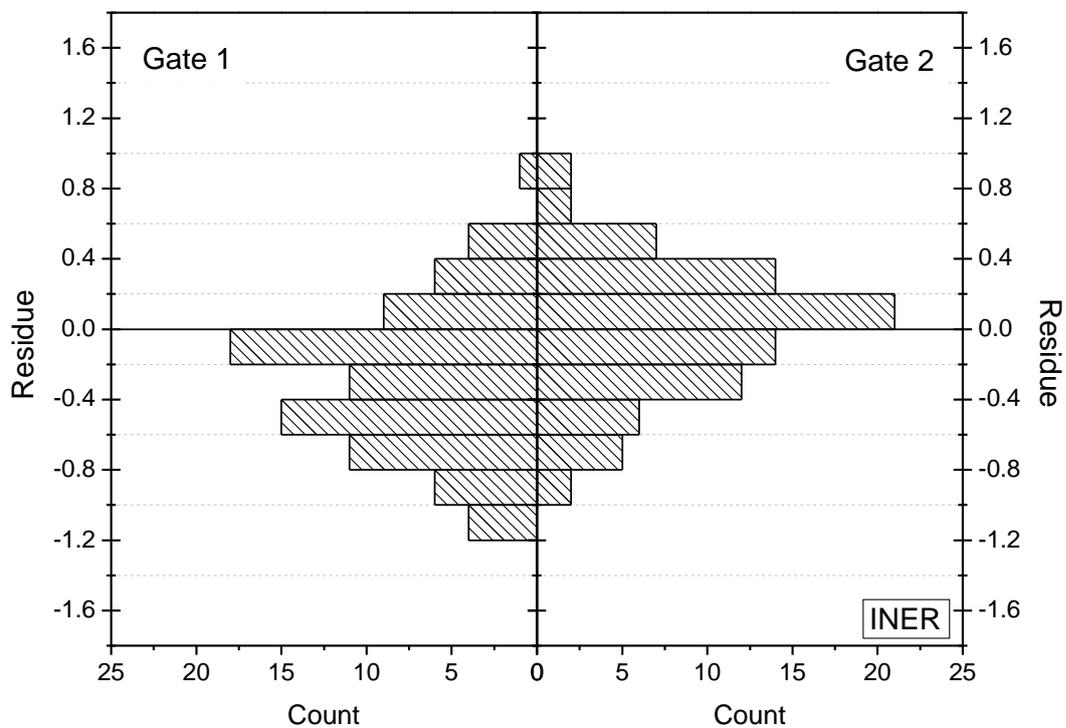
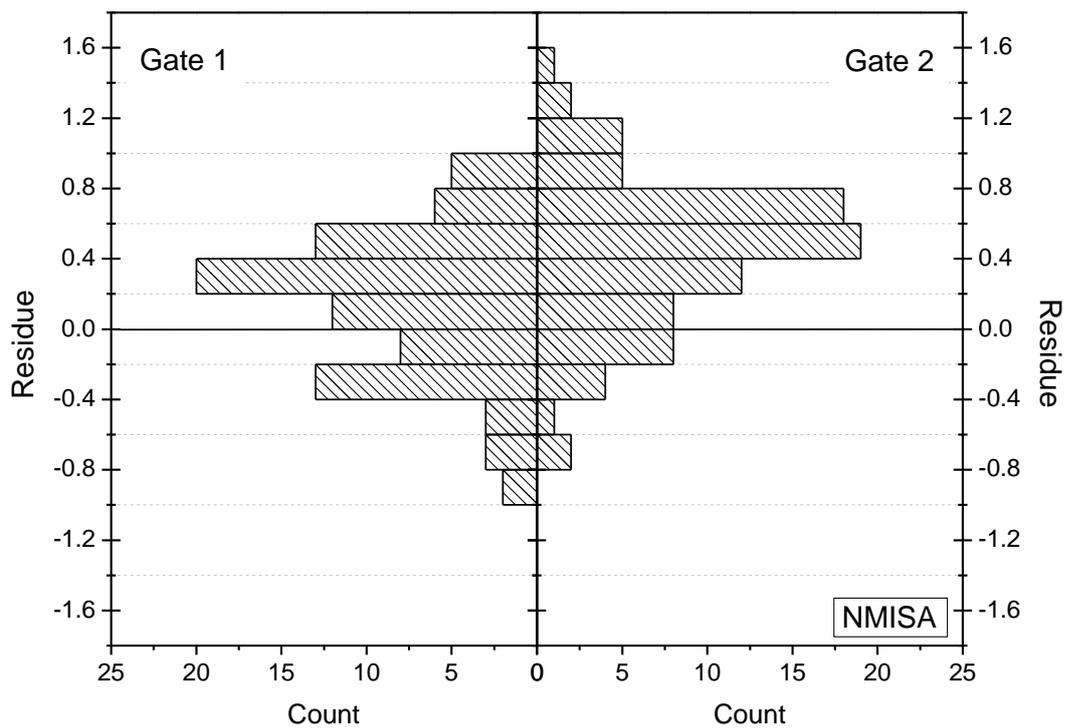


Figure 12: Distribution of the residues for each participant (2/6).

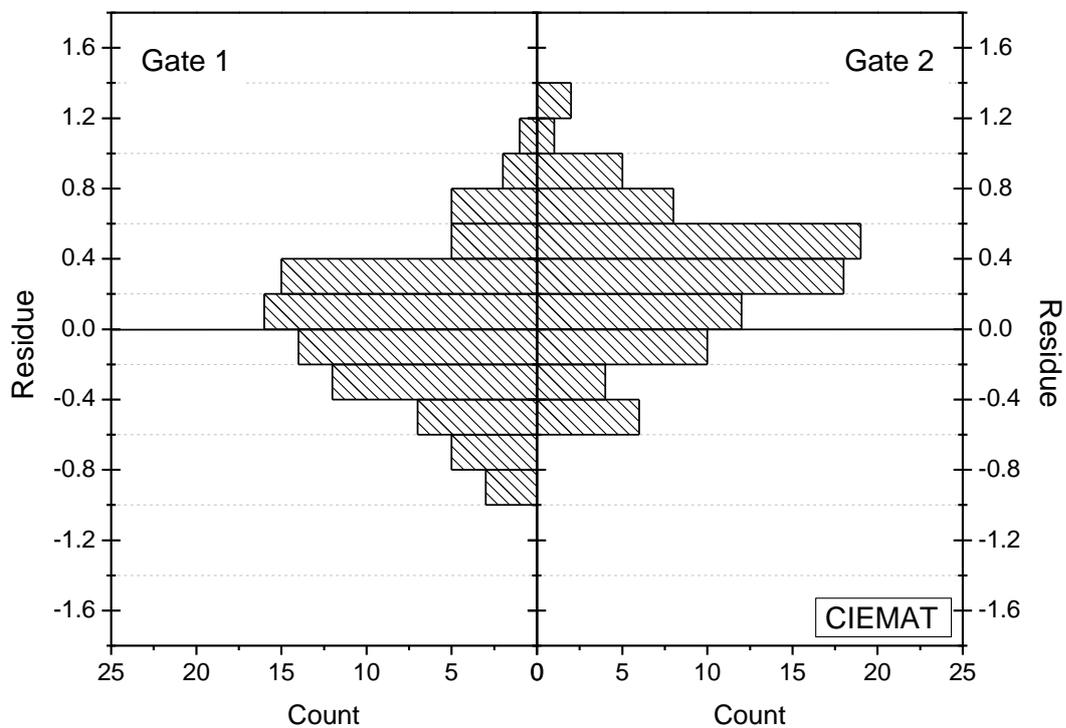
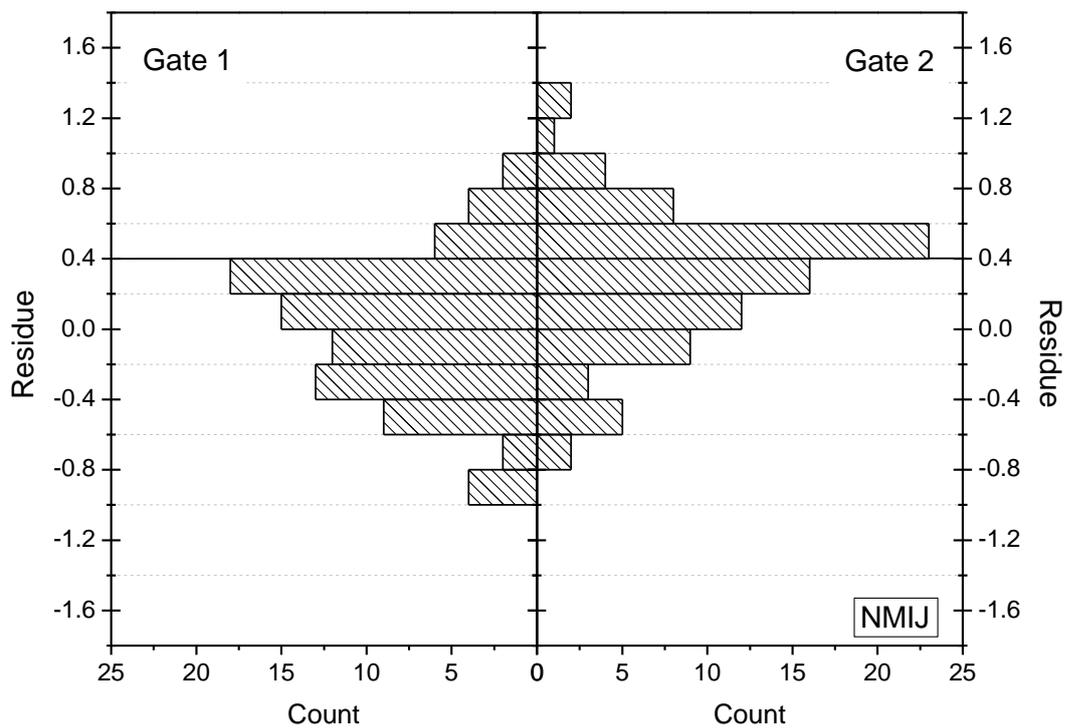


Figure 12: Distribution of the residues for each participant (3/6).

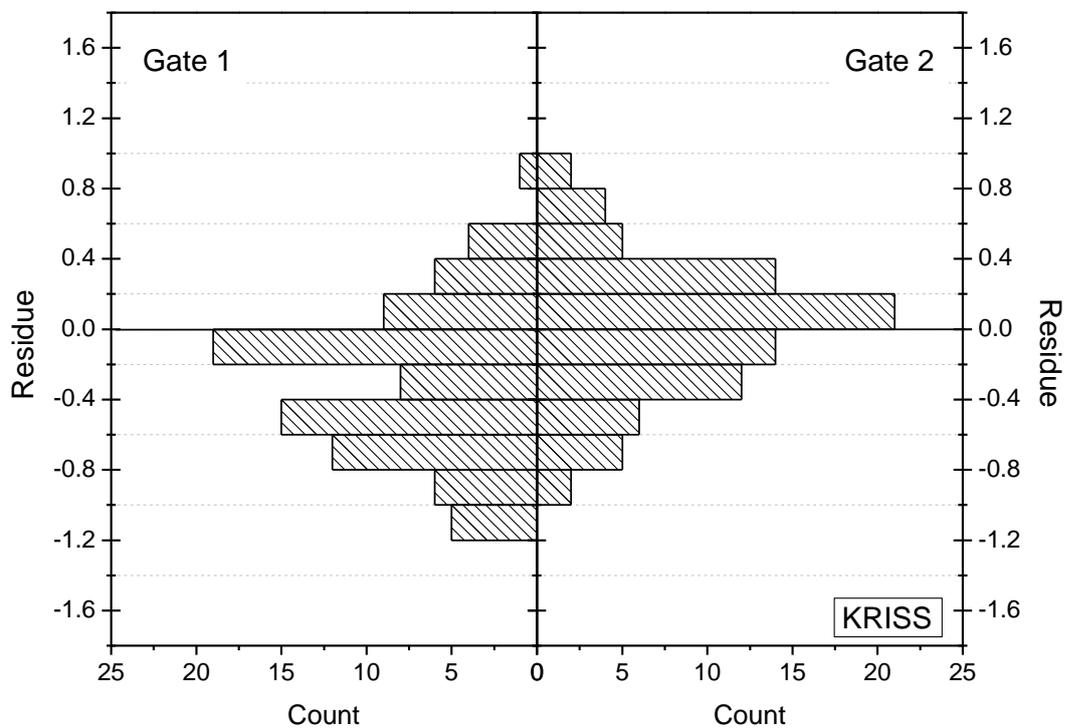
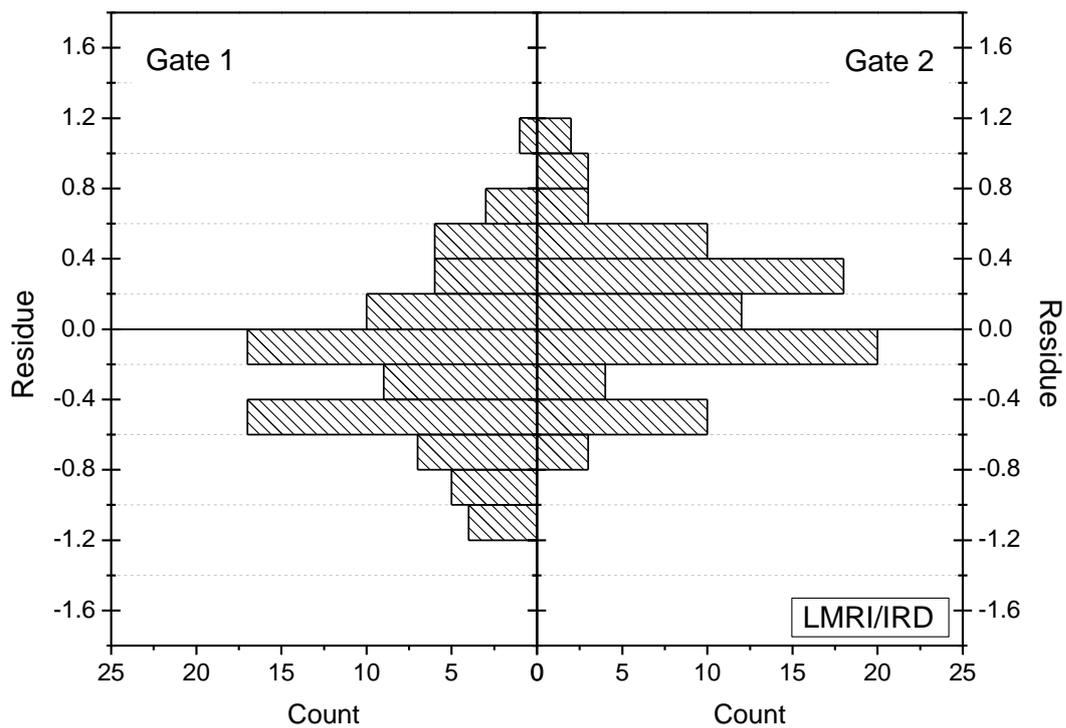


Figure 12: Distribution of the residues for each participant (4/6).

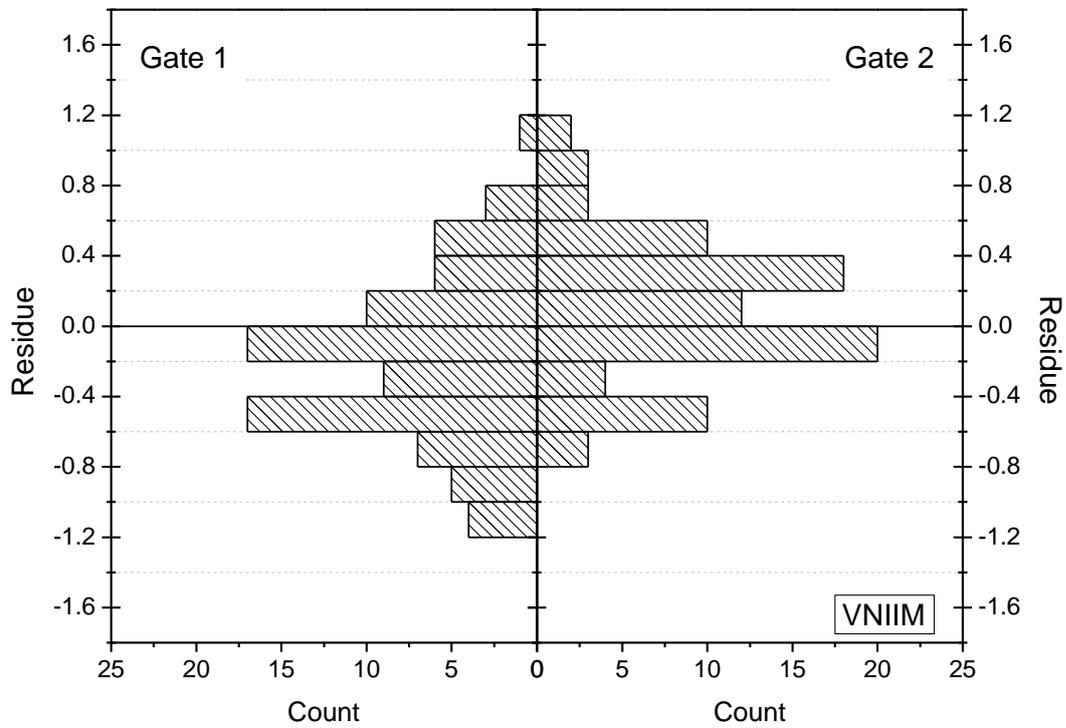
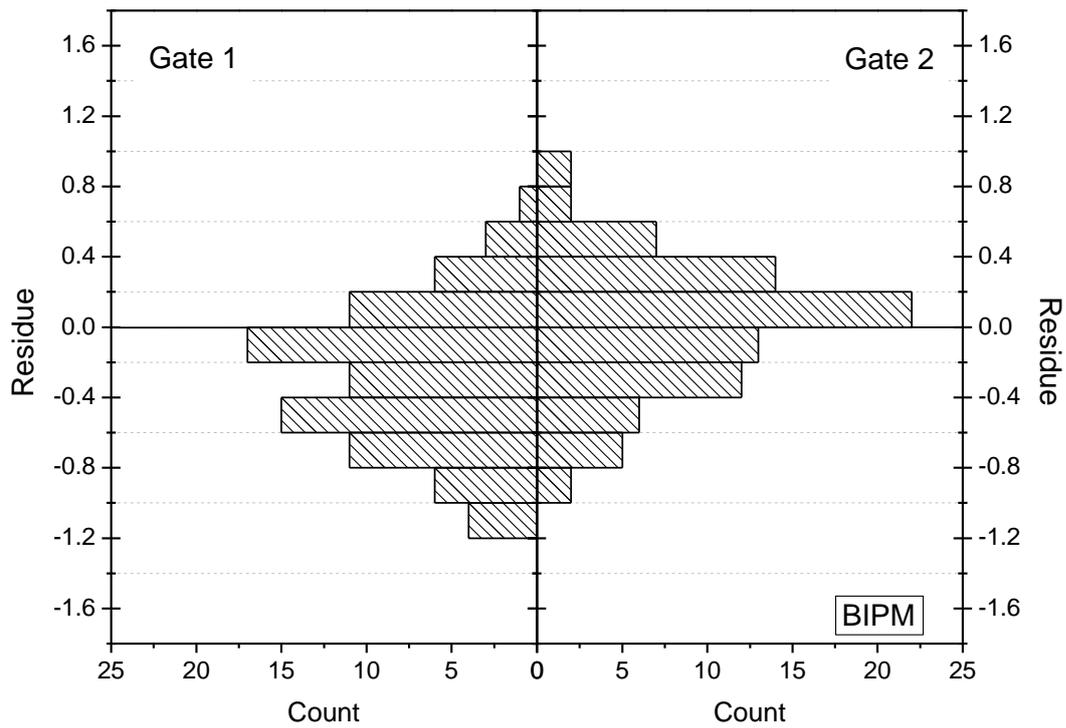


Figure 12: Distribution of the residues for each participant (5/6).

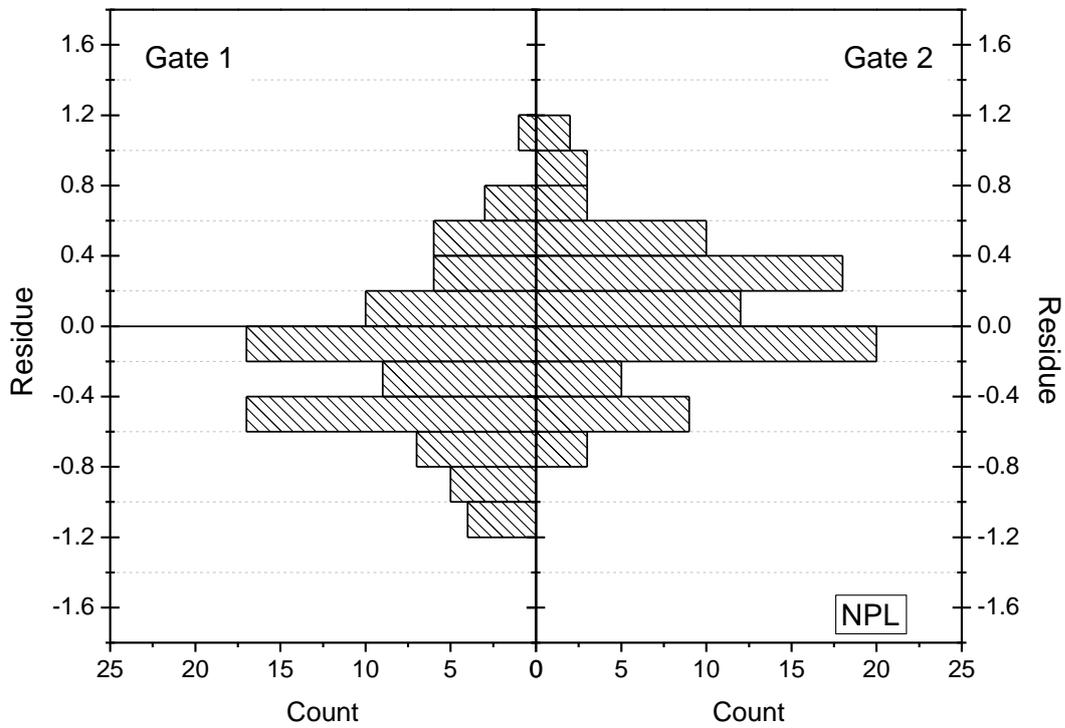
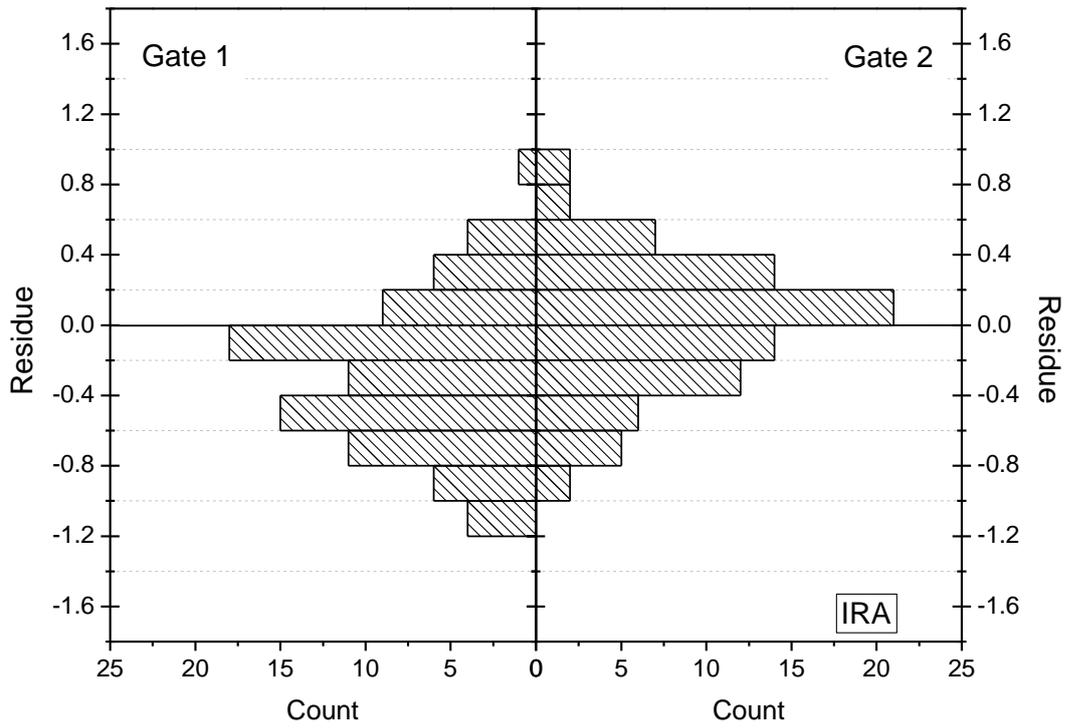


Figure 12: Distribution of the residues for each participant (6/6).