



Estimating the capabilities of the TIMS TE analytical technique using random effects models applied to quality control measurements

Alexander Venzin¹ · Monika Sturm¹ · Andreas Koepf¹ · Joe Hiess¹ · Sergei Boulyga¹

Received: 28 October 2022 / Accepted: 10 December 2022 / Published online: 10 January 2023
© Akadémiai Kiadó, Budapest, Hungary 2023

Abstract

The International Atomic Energy Agency uses the Thermal Ionization Mass Spectrometry Total Evaporation method routinely to analyze the $n(\text{U-235})/n(\text{U-238})$ atom amount ratios in inspection and quality control samples. This paper uses a top-down statistical model applied to quality control measurements of certified reference materials (CRM) to study the capabilities of the technique. The technique is shown to produce measurements with a relative standard deviation not more than 0.03% when applied to a wide range of certified reference materials.

Keywords TIMS TE · Mixed effects models · Intermediate precision · Analysis of variance · Nuclear material analysis · Nuclear safeguards

Introduction

Multi-collector thermal ionization mass spectrometry (MC-TIMS) is optimally suited for isotope ratio measurements of U and Pu in nuclear material samples because it possesses superior accuracy and reliability, and it can be relatively easily implemented in a radioanalytical laboratory. Due to the high precision attainable with state-of-the-art MC-TIMS instruments and comparatively few molecular and polyatomic interferences, MC-TIMS is considered a benchmark technique for isotope ratio analysis of these elements in nuclear material samples [1, 2]. The simultaneous multi-collection of isotopes allows the implementation of the so-called “total evaporation” (TE) method [3, 4], for which the

sample is completely evaporated within a relatively short measurement time, and the isotope ratios are calculated from the integrated (i.e., summed) intensities, thus minimizing the isotope mass fractionation effects. This technique has been demonstrated to be sufficiently precise and accurate for major isotope-amount ratio measurements of uranium and plutonium in nuclear material samples [2]. It has been used for routine nuclear safeguards measurements for more than 30 years and is still the most frequently used method today.

The TE method has some disadvantages such as peak tailing contributions that cannot be accounted for in situ due to the nature of static measurements [2]. Also signal spikes, which can be caused by voltage arcing in the TIMS ion source, might bias the measured isotope ratios. Such signal spikes cannot be identified and filtered out by the currently used TIMS data evaluation software. The magnitude and the frequency of such spikes could be reduced by using the ion-source design described by Siegmund et al. [5]. However, depending on the ion source, some residue signal spikes might still occur in TIMS TE analyses. It should be mentioned that peak tailing and signal spikes are particularly crucial for the determination of minor isotopes, e.g. U-234 and U-236, but these effects might also introduce some bias to the measured $n(\text{U-235})/n(\text{U-238})$ atom amount ratio, especially in samples with low U-235 abundance.

Measurements produced using this method include several sources of uncertainty coming from both chemical sample preparation and mass spectrometric analysis. One could

✉ Alexander Venzin
alex.venzin@gmail.com

Monika Sturm
m.sturm@iaea.org

Andreas Koepf
a.koepf@iaea.org

Joe Hiess
j.hiess@iaea.org

Sergei Boulyga
s.bulyha@iaea.org

¹ Department of Safeguards, Office of Safeguards Analytical Services, International Atomic Energy Agency, Friedenstrasse 1, Seibersdorf 2444, Lower Austria, Austria

model these sources of uncertainty using either first principles or bottom-up uncertainty models like those described in the Guide to Uncertainty in Measurement (GUM) [6] or top-down statistical models (i.e. random effects models). Most laboratory analysts have some experience using GUM methods to estimate uncertainties, but it is not trivial to derive uncertainties for every single variable. Top-down methods can be useful for deriving agglomerate uncertainty estimates if repeated measurements are available. These models rely on defining a sampling plan that carefully groups measurements together to study the variables of interest. These models can be mathematically complex, however, there are open-source software packages available that take care of the computational burden for the researcher(s).

Experimental

Three instruments are used for TIMS TE analysis. There is one Triton® and two Triton Plus® instruments (ThermoFisher Scientific, Bremen, Germany). In this study these instruments are referred to as “Triton 2”, “Triton 3” and “Triton 4”, respectively. Triton 2 and Triton 3 were re-validated in 2015 after their transfer to the International Atomic Energy Agency’s Nuclear Material Laboratory (NML) from an older laboratory building. The Triton 4 instrument was purchased new and validated in the NML.

Uranium isotopic standards NBL CRM 112A, NBL CRM U005A, NBS CRM U030A, NBL CRM U100, NBL CRM U200, NBL CRM U500, NBL CRM U930 [7–13] and IRMM 183, IRMM 184, IRMM 185, IRMM 186, IRMM 187 [14] are used as quality control (QC) samples. Samples represented pure U solutions diluted in purified water with 3 M nitric acid.

Theory

Exploratory data analysis

In the period from January 2016 to July 2020, 4154 quality control TIMS TE measurements were produced from 12 different reference materials as can be seen in Fig. 1. The measurements were performed to determine atom amount ratios of uranium.

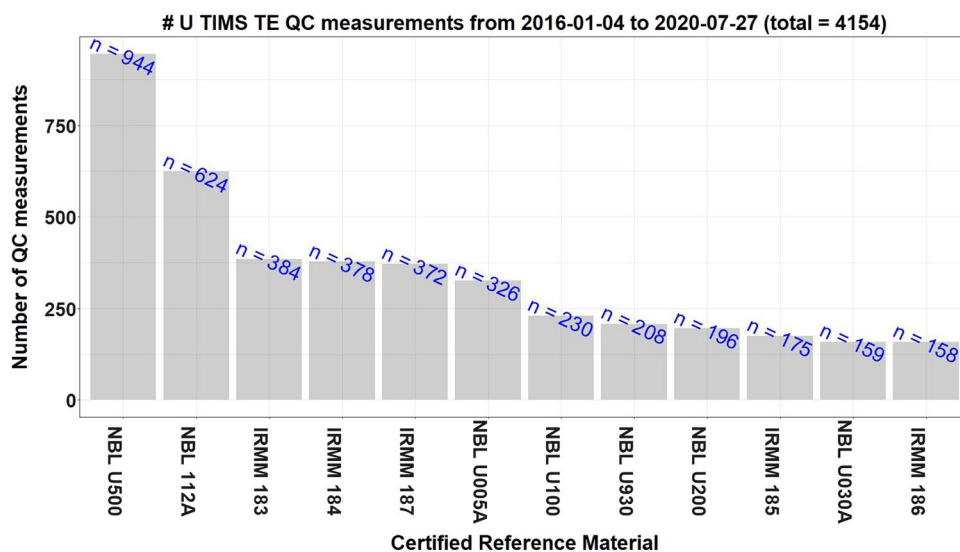
Description of quality control measurements

The TIMS TE analytical method is primarily used to determine the “major” isotope ratios like $n(\text{U-235})/n(\text{U-238})$ and $n(\text{Pu-240})/n(\text{Pu-239})$, as well as $n(\text{U-233})/n(\text{U-238})$ and $n(\text{Pu-242})/n(\text{Pu-239})$ atom amount ratios in samples that have been spiked for isotope dilution (ID) analysis. Measurements of atom amount ratios involving minor isotopes (U-234, U-236 and Pu-238, Pu-241, Pu-242) are known to be biased for the TIMS TE method. This article is focused entirely on studying TIMS TE QC measurements of the $n(\text{U-235})/n(\text{U-238})$ atom amount ratio because it is the measurement of interest for most applications of the technique.

There are two experimental variables of interest that have been documented with all measurements, and these are instrumentation (Triton 2, Triton 3 and Triton 4) and the names of aliquots from which measurements were produced. These two variables have been used to define a grouping structure for the data for statistical analysis.

Aliquots are chemical preparations of a subsample of a given CRM. After preparation and delivery to the TIMS laboratory, the new aliquot is analysed to assess if it is suitable for quality control. The results produced from new aliquots are plotted on control charts whose parameters are

Fig. 1 The number of quality control (QC) measurements produced from each uranium certified reference material (CRM) over the time period 2016/01/04 to 2020/07/27



defined by the certified reference value and its uncertainty. During this characterization period, if there is an aliquot that is already being used for quality control, then the new and old aliquot may be measured in parallel. If the results produced from the new aliquot are satisfactory, then it is accepted for use and the old one is discarded. Otherwise, another new aliquot is requested.

There may also be differences between mass spectrometers that introduce some amount of measurement heterogeneity. The TIMS laboratory operates two models of TIMS: The Triton® (Triton 2) and the Triton Plus® (Triton 3 and Triton 4). These instruments are produced by the same manufacturer, ThermoFisher Scientific. The Triton Plus® offers an option for extension of an ion detector collector, however, in the current configuration, all three TIMS are considered equivalent with respect to their performance for U isotope analysis.

The measurement results were produced under varying conditions during the time period. Most QC measurements were produced by several analysts in parallel with routine inspection sample analysis, whereas others were training measurements produced by staff learning how to apply the TIMS TE method. In accordance with the TIMS TE procedure, the name of the analysts performing analyses are recorded in loading sheets and in the Laboratory Information Management System (LIMS), but not in the QC charts and thus, these data were not used in the current evaluation. Instead, all other sources of uncertainty, including those associated with the analyst were combined into a single residual effect.

One can visualize how measurement varies with changes in aliquots and instruments using panel graphics. For

example, each panel in Fig. 2 corresponds to measurements of natural uranium reference material IRMM 184 produced by each Triton instrument. The measurements are ordered by time, but the dates on the horizontal axis are free to vary for each instrument. The measurements are colour coded by aliquot ID and the plot symbols correspond to the different instruments. The certified reference value is plotted as a red dashed line in each panel. An interval for the uncertainty of the certified value was not plotted because the primary interest is visualizing the intermediate precision of the QC measurements. These measurements are reasonably homogeneous from instrument to instrument and aliquot to aliquot except for some extreme results produced from instruments Triton 2 and Triton 3.

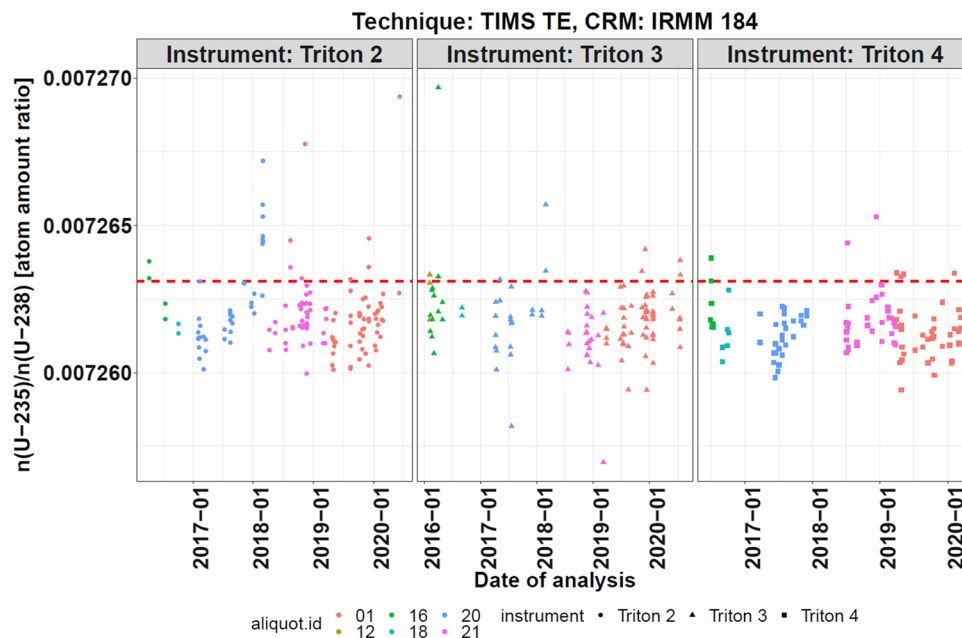
Atom amount ratio measurements produced from different aliquots will also vary to some extent (Fig. 3). The variation in measurement could be attributed to differences in chemical preparation of subsamples of the CRM. There is also the possibility that some of the data structure can be explained by other unknown variables.

Using a random effects model to analyze the data

Variation in the $n(\text{U-235})/n(\text{U-238})$ atom amount ratio measurements may be associated with differences between instruments and aliquots. The following questions can be studied using a random effects model:

- (1) By what amount can measurements differ from instrument to instrument?
- (2) By what amount can measurements differ from aliquot to aliquot?

Fig. 2 Measurements of the $n(\text{U-235})/n(\text{U-238})$ atom amount ratio for standard IRMM 184. The red dashed line in each panel is the certified reference value



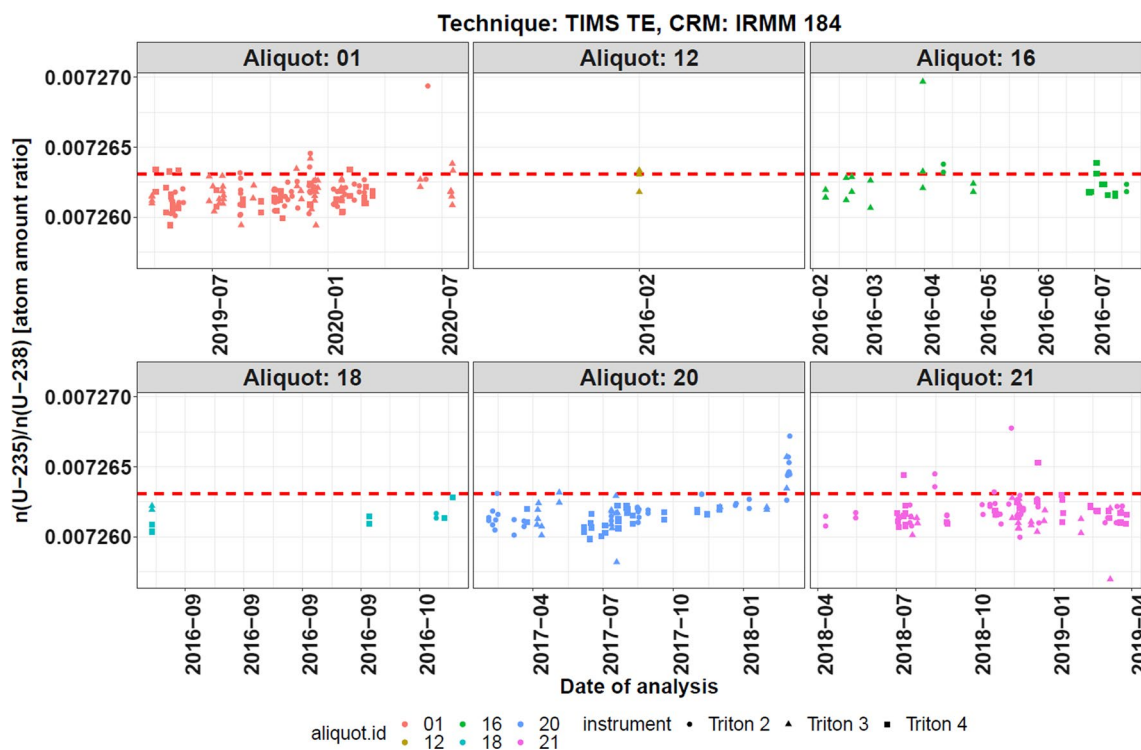


Fig. 3 Measurements of IRMM 184 grouped by aliquot ID. The red dashed line in each panel is the $n(\text{U-235})/n(\text{U-238})$ atom amount ratio certified reference value

- (3) By what amount can measurements differ when subsamples of the same aliquot are analysed using different instruments and vice versa?
- (4) How much of the measurement variation cannot be explained by either instrument or aliquot heterogeneity?

The (statistical) experimental design that most closely resembles how the data was collected is called a “factorial” or “crossed” sampling design. The aliquots are crossed with the instruments in this sampling model because subsamples of the same aliquots are analyzed on all instruments. An example of a factorial sampling design is shown in Table 1 where each row corresponds to a different aliquot and each column corresponds to a different instrument.

A measurement produced from a sampling design with two factors and an interaction component subject to additive errors can be written algebraically as follows:

$$y_{ijk} = \mu + \varepsilon_{A_i} + \varepsilon_{I_j} + \varepsilon_{(A_i \times I_j)} + \varepsilon_{residual_{ijk}} \tag{1}$$

y_{ijk} is the k th TIMS TE measurement of the $n(\text{U-235})/n(\text{U-238})$ atom amount ratio produced by the j th instrument subsampled from the i th aliquot.

μ is the estimate of the mean atom amount ratio derived from measurements of a chosen CRM.

$\varepsilon_{A_i} \stackrel{IID}{\sim} N(0, \sigma_{aliquot}^2)$ is assumed to be an independent and identically distributed (IID) mean zero Gaussian random variable that accounts for atom amount ratio heterogeneity from aliquot to aliquot.

$\varepsilon_{I_j} \stackrel{IID}{\sim} N(0, \sigma_{instrument}^2)$ is assumed to be an IID mean zero Gaussian random variable that accounts for atom amount ratio heterogeneity from instrument to instrument.

$\varepsilon_{(A_i \times I_j)} \stackrel{IID}{\sim} N(0, \sigma_{interaction}^2)$ is assumed to be an IID mean zero Gaussian random variable that accounts for atom amount ratio heterogeneity from the combined effect due to

Table 1 Example of a balanced factorial sampling design with factors aliquot and instrument. Each aliquot is measured three times on each instrument

	Instrument 1			Instrument 2			Instrument 3		
Aliquot 1	y_{111}	y_{112}	y_{113}	y_{121}	y_{122}	y_{123}	y_{131}	y_{132}	y_{133}
Aliquot 2	y_{211}	y_{212}	y_{213}	y_{221}	y_{222}	y_{223}	y_{231}	y_{232}	y_{233}
Aliquot 3	y_{311}	y_{312}	y_{313}	y_{321}	y_{322}	y_{323}	y_{331}	y_{332}	y_{333}

aliquot and instrument heterogeneity after accounting for their individual uncertainty contributions.

$\varepsilon_{residual_{ijk}} \stackrel{IID}{\sim} N(0, \sigma_{residual}^2)$ is an IID mean zero Gaussian random variable that accounts for the uncertainty that cannot be explained by the aliquot and instrument effects.

The curious reader can find more information about the theory and computation of linear mixed effects models in the book *Variance Components* by Searle et al. [15]. A more hands-on approach to model fitting with the R statistical programming language can be found in Galecki and Burzykowski's textbook [16].

Interpretation of the model for a mass spectrometry analyst

Instead of trying to model the sources of uncertainty for individual measurements like peak tailing effects, detector deadtime, etc., suppose that we assign uncertainty to high-level sampling factors like instrumentation and aliquot preparation.

The aliquot effect, ε_{A_i} , can be interpreted as a random variable modelling the differences in chemical preparations of samples. The estimate of uncertainty associated with this effect is $\sigma_{aliquot}$. All sources of uncertainty that can be attributed to aliquot preparation are contained within the parameter estimate $\sigma_{aliquot}$.

By similar argument, the instrument effect, ε_{I_j} , can be interpreted as a random variable modelling the differences in measurement outcomes associated with the instruments. The variation due to sources like peak tailing and detector deadtime are therefore incorporated into the combined estimate, $\sigma_{instrument}$.

There is also a model parameter for the combined effect due to aliquot and instrument, $\varepsilon_{(A_i \times I_j)}$. Most analysts assume that this experimental effect does not exist, because otherwise it means that measurement outcomes depend on specific combinations of instrument and aliquot. One could omit this term from the model, but it can be useful to actively model it and confirm that $\sigma_{interaction} \approx 0$ for all data sets.

There are other sources of uncertainty that are confounded with the chosen factors, but they cannot be accounted for without changing the sampling model. The other sources of uncertainty are combined into a single residual effect, $\varepsilon_{residual_{ijk}}$, and the combined uncertainty of these other effects is $\sigma_{residual}$. Note that the word “residual” is descriptive because $\sigma_{residual}$ is the uncertainty remaining after accounting for effects due to aliquot heterogeneity, instrument heterogeneity and their combined effects.

There are limits to what can be statistically quantified due to the conditions under which the data was collected. This model is a first approximation for developing more

refined sampling models for monitoring quality control measurements.

Results and discussion: application of the model to TIMS TE measurements of NBL 112A

All statistical analysis has been done using the R statistical programming language. Usage of italicized font indicates that the object is either a function or library in the R language. The *rlmer* function from the *robustlmm* library [17] was used to analyze all data sets. This software was chosen because it is designed for robustness against outliers under the assumed model.

Certified reference material NBL 112A is a natural uranium standard. A total of $N = 624$ TIMS TE QC measurements of NBL 112A were produced during the time period 2016-01-04 to 2020-07-24. The measurements can be partitioned into groups defined by aliquot and instrument as indicated in Table 2.

Boxplots are useful visual tools for examining measurements split into groups. If groups of measurements are visually distinguishable across aliquots and instruments, then it is evidence of potential heterogeneity.

The boxplots indicate that the distribution of measurements within each group is approximately symmetric, but there are more extreme events than expected for Gaussian distributions. It is assumed that these extreme events are outliers, but it is also possible to augment the model so that the Gaussian assumptions are exchanged for other probability distributions. It can be loosely inferred from Fig. 4 that there is likely aliquot heterogeneity and some small amount of instrumental heterogeneity. It is unclear if a significant interaction effect exists after accounting for the individual effects due to aliquots and instruments.

Estimating standard errors and intervals for standard deviation estimates is impractical for this data set because only three instruments and five aliquots have been observed. The accuracy of statistical intervals computed from small

Table 2 Breakdown of sample size by aliquot and instrument

	Triton 2	Triton 3	Triton 4
Aliquot 01	59	50	69
Aliquot 13	2	2	0
Aliquot 15	42	33	43
Aliquot 18	78	56	72
Aliquot 19	35	33	50

For example, 59 measurements of aliquot 01 were produced using instrument Triton 2, 50 measurements of aliquot 01 were produced using Triton 3, etc.

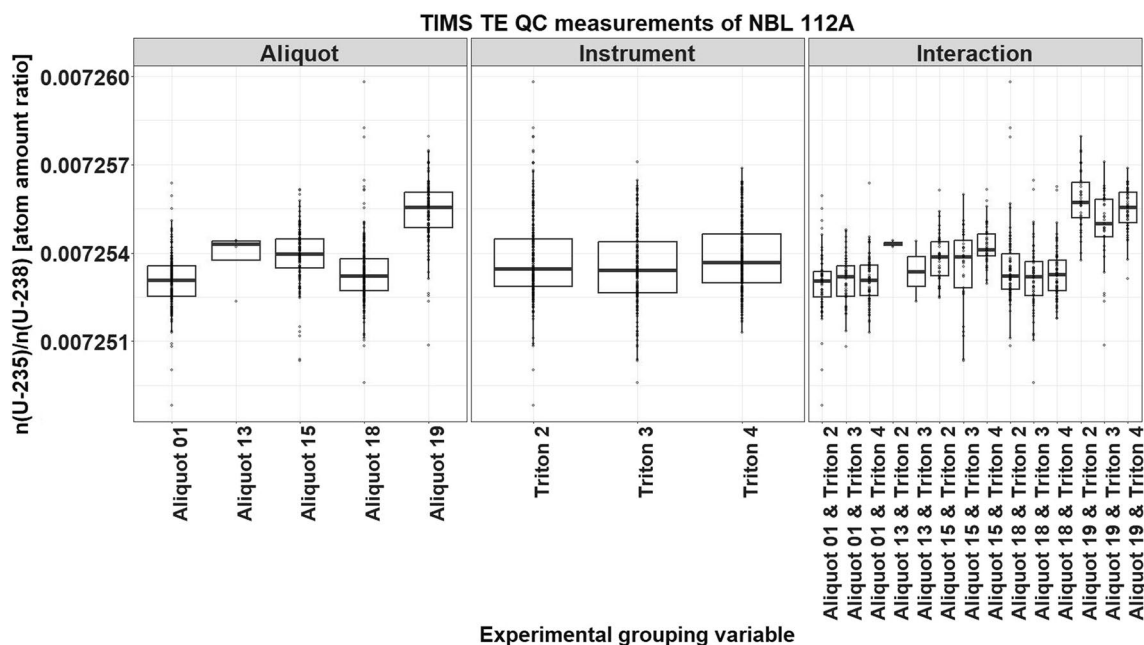


Fig. 4 Boxplots of measurements grouped by aliquot ID, instrument, and the combination of the factors

sample sizes is heavily model dependent and they are often so wide as to provide limited information. Hence, we are primarily interested in point estimates of uncertainty components that will provide a rough indication of the magnitude of the heterogeneity associated with each experimental factor. A reasonably precise standard error estimate for the mean, however, can be computed. The estimates are shown in Table 3.

The σ_{aliquot} parameter is the uncertainty due to differences in the observed $n(\text{U-235})/n(\text{U-238})$ atom amount ratio between aliquots of NBL 112A. The model estimates suggest that aliquot heterogeneity is a more significant

uncertainty contributor than instrument heterogeneity and about the same magnitude as residual variation.

The model can also be used to derive a simple, but useful estimate of combined uncertainty for individual measurements of NBL 112A.

$$u(y_{ijk}) = \sigma_{\text{combined}} = \sqrt{\sigma_{\text{aliquot}}^2 + \sigma_{\text{instrument}}^2 + \sigma_{\text{interaction}}^2 + \sigma_{\text{residual}}^2} \quad (2)$$

The combined standard uncertainty for individual measurements of NBL 112A is about 1.3 ppm [$n(\text{U-235})/n(\text{U-238})$ atom amount ratio] or about 0.02% relative to the

Table 3 Point estimates of parameters derived using the *rlmer* algorithm

Parameter	Point estimate	Interpretation
μ	0.00725387 (46)	The mean estimate of the $n(\text{U-235})/n(\text{U-238})$ atom amount ratio is 0.00725387(46). The standard error / uncertainty of the estimate is 4.6E-7.
σ_{aliquot}	9.8E-7	The measured $n(\text{U-235})/n(\text{U-238})$ atom amount ratio varies by 0.98 ppm from aliquot to aliquot. Relative to the mean estimate, measurements from different aliquots vary by 0.014%.
$\sigma_{\text{instrument}}$	8.2E-8	The measured $n(\text{U-235})/n(\text{U-238})$ atom amount ratio varies by 82 ppb from instrument to instrument. Relative to the mean estimate, measurements from different instruments vary by 0.0011%.
$\sigma_{\text{interaction}}$	0	The measured $n(\text{U-235})/n(\text{U-238})$ atom amount ratio varies by a negligible amount when studying the interaction between aliquots and instruments. ^a
σ_{residual}	9.1E-7	All other sources of variation account for 0.91 ppm $n(\text{U-235})/n(\text{U-238})$ atom amount ratio. Relative to the mean estimate, all other sources of variation are 0.013%.

The *psi2propII* squared robustness weights were used for the *psi* terms with variance components and adjusted for 95% efficiency. The default tuning parameters were used for the fixed effect (mean)

^aIn theory, $\sigma_{\text{interaction}} > 0$, but from a practical point of view, the standard deviation of the interaction term is so small relative to residual variation that it can be removed from the model without significantly changing results

mean estimate ($k = 1$). This estimate of uncertainty can be used to provide IAEA evaluators with an indication of TIMS TE precision for measurements of the atom amount ratio derived from natural U samples.

Application of the model to all TIMS TE QC data sets of interest

The statistical model can be applied to the other sets of certified reference material measurements because the sampling structure is the same. The uncertainty component estimates are useful for investigating the intermediate precision capabilities of the TIMS TE method. The reference materials that have been analyzed cover a wide range of U-235 enrichment and so uncertainty is best described in relative terms. Relative standard deviations are ratios of an uncertainty estimate to the mean derived from the model expressed as a percentage.

$$\widehat{RSD}_{component} = 100\% \frac{\widehat{\sigma}_{component}}{\widehat{\mu}} \tag{3}$$

The circumflex is used to indicate that $\widehat{\sigma}_{component}$ and $\widehat{\mu}$ are uncertain estimates. It is difficult to derive meaningful estimates of $\widehat{u}(\widehat{\sigma}_{component})$ due to the small number of instruments and aliquots studied for each CRM data set. This means that the estimates of relative standard deviation for each uncertainty component and $u(y_{ijk})$ (labelled as “combined”) in Fig. 5 are subject to an unknown amount of uncertainty.

The %RSD estimates are ordered by increasing U-235 abundance such that IRMM 183 contains the least U-235 and NBL U930 is the most highly enriched reference material. Ideally, the major uncertainty component is $\sigma_{residual}$ for all data sets because this would mean that heterogeneity due to differences in instruments and aliquots is negligible relative to other sources of variation. Ratios of standard deviation estimates to the residual standard deviation are shown in Fig. 6. These ratio estimates can be used as rough approximations to the more formal ANOVA F-test procedure that compares ratios of mean squares.

Factors aliquot and instrument are minor uncertainty contributors for most data sets relative to the residual uncertainty component. Important exceptions are data sets NBL U005A, NBL 112A and IRMM 185 that demonstrate potentially significant aliquot to aliquot measurement heterogeneity as demonstrated using boxplots in Fig. 7. There may be some component of instrumental heterogeneity that is associated with measurements of NBL U200.

The estimate of $\sigma_{instrument}$ derived from the measurements of NBL U200 indicates possible instrumental heterogeneity meaning that the observed $n(U-235)/n(U-238)$ atom amount ratio depends on which instrument is used to produce QC measurements (Fig. 8).

The estimates of combined measurement uncertainty, $\widehat{u}(y_{ijk})$, can also be compared to the standard uncertainty of the certified reference values. The uncertainty of the reference value, denoted as $u(\mu_{ref})$ where μ_{ref} is the certified reference value, should preferably be smaller than the individual measurement precision of the analytical technique for

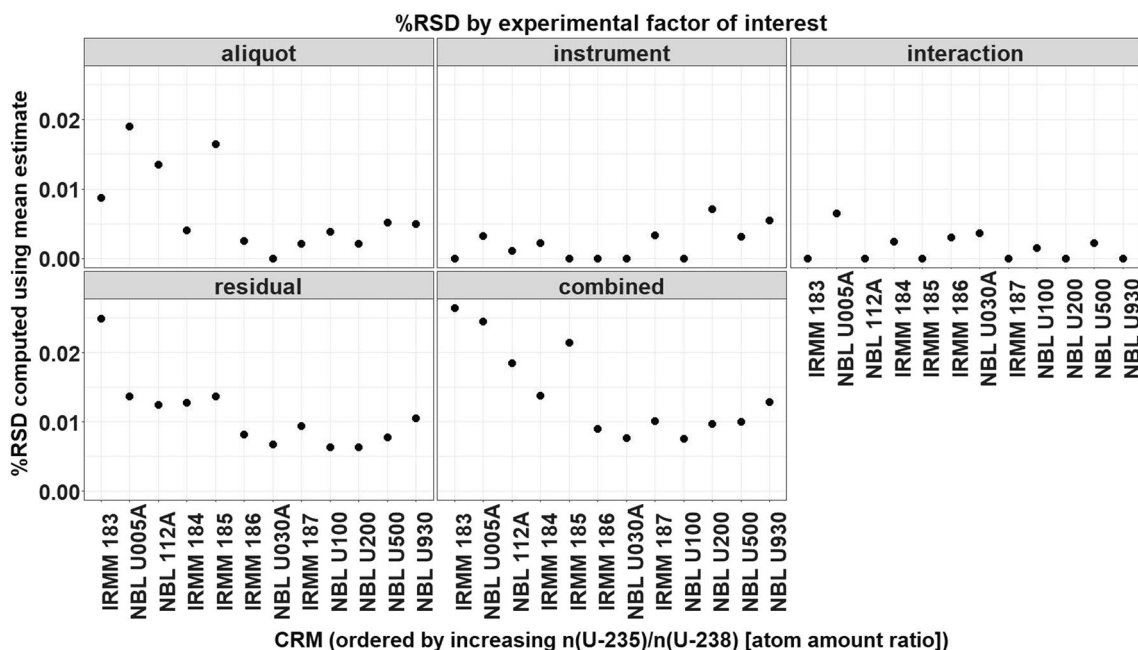


Fig. 5 Plot of estimated %RSD for each uncertainty component (aliquot, instrument, interaction, residual) and the combined %RSD

Fig. 6 Ratio of $\frac{\hat{\sigma}_{component}}{\hat{\sigma}_{residual}}$ for each CRM data set. The red dashed line is $\frac{\hat{\sigma}_{component}}{\hat{\sigma}_{residual}} = 1$. Values one or larger indicate potential heterogeneity

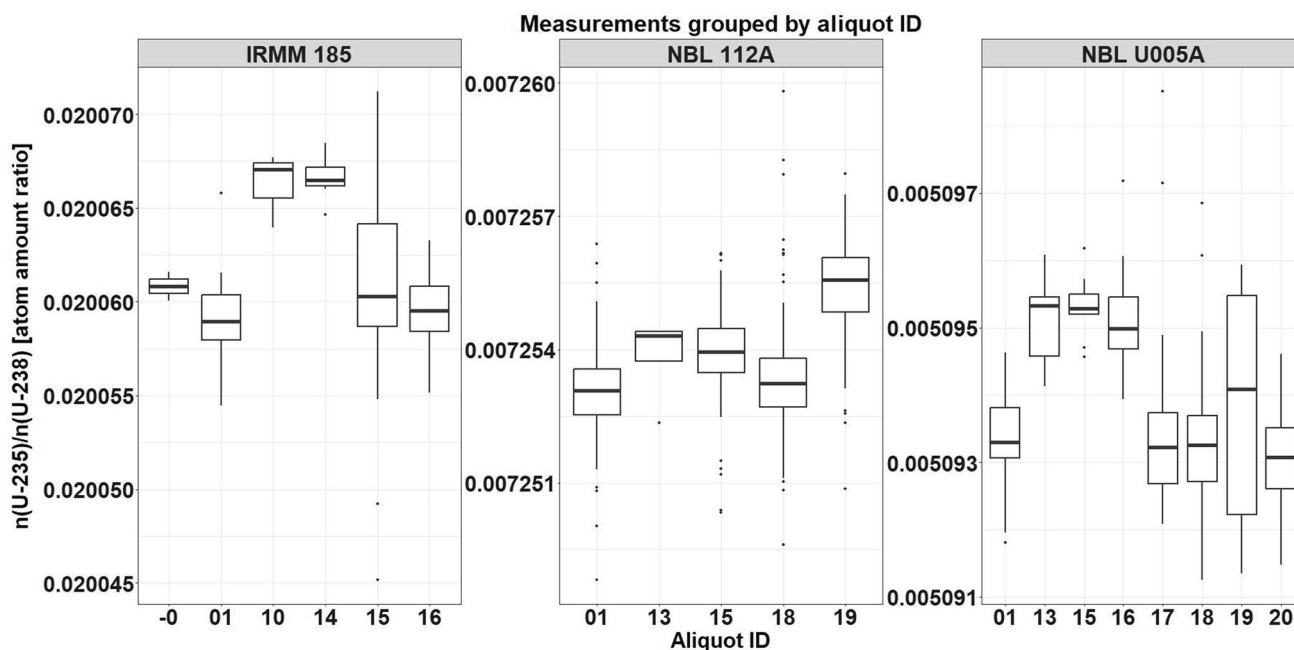
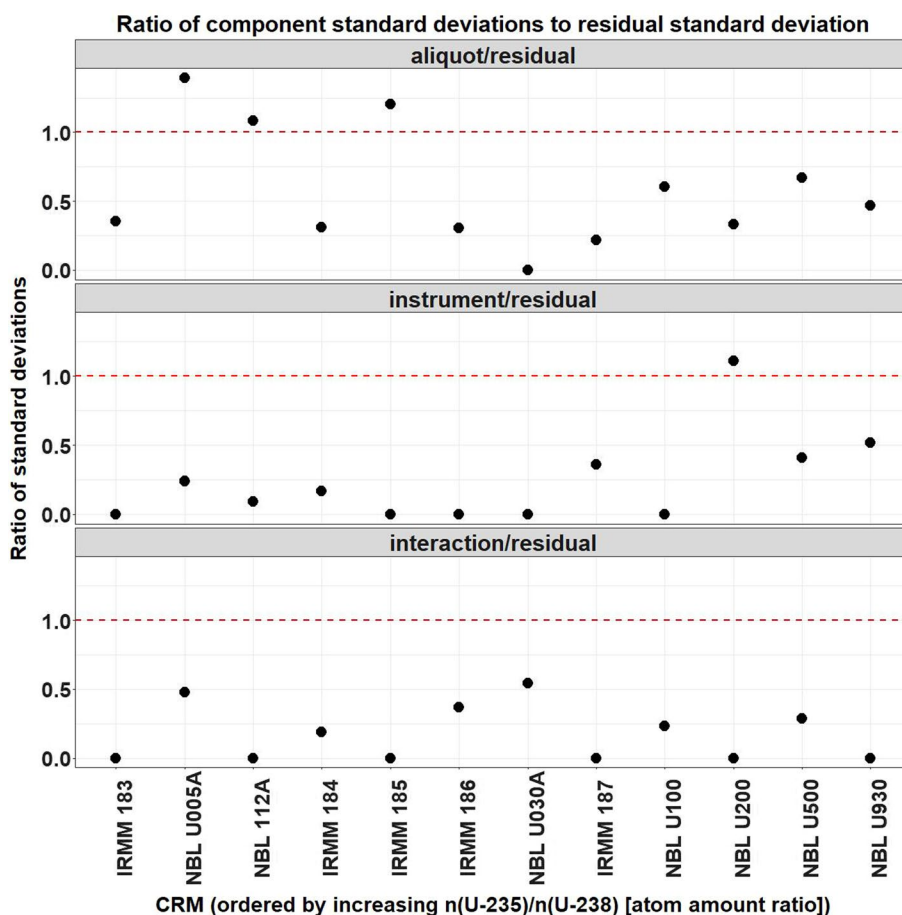
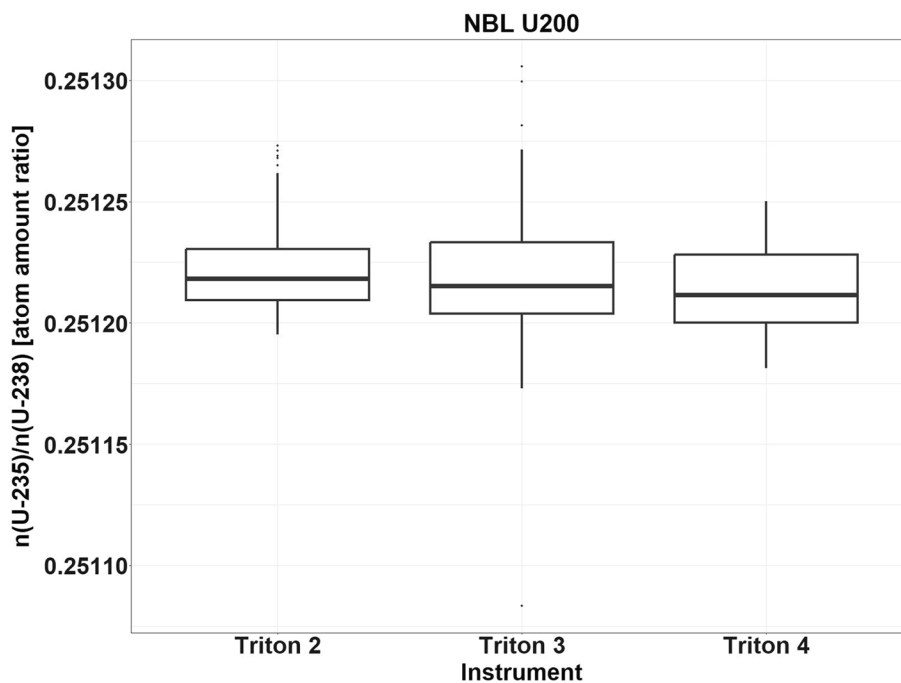


Fig. 7 Boxplots of TIMS TE QC measurements grouped by aliquot ID for data sets NBL 112A and NBL U005A

Fig. 8 Boxplots of measurements of NBL U200 grouped by instrument. The measurements produced on Triton 4 are, on average, smaller in magnitude than those produced on Tritons 2 and 3



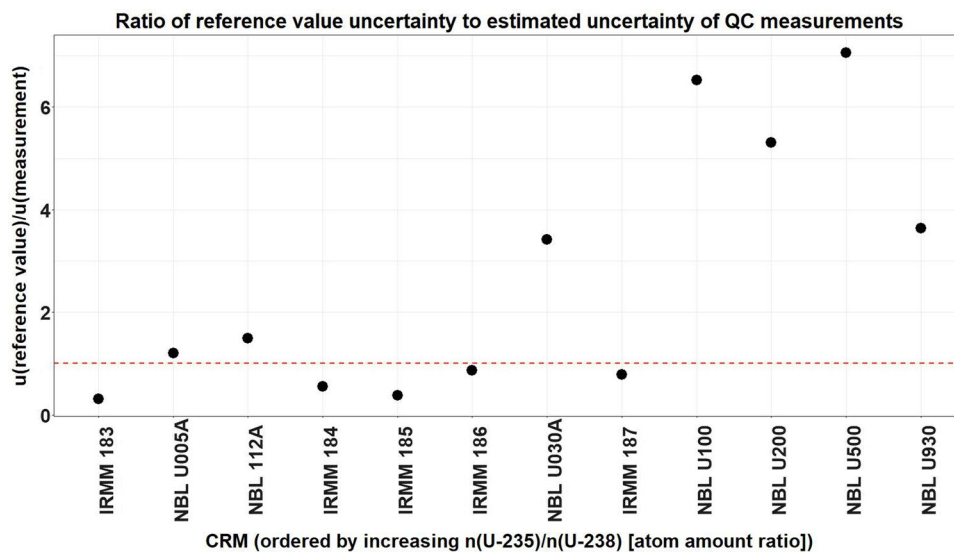
which it is used. One way to visualize this is to plot the ratios of the uncertainty of the reference material to the model estimate of combined uncertainty.

The results displayed in Fig. 9 suggest that the uncertainty of TIMS TE measurements is often as small as the uncertainty of the certified reference values. This could be problematic for improving the TIMS TE technique because reference materials are needed whose certified $n(\text{U-235})/n(\text{U-238})$ atom amount ratio values are known with uncertainties that are not larger than the measurement uncertainty of the analytical technique. This deficiency is most acute for reference materials whose U-235 content is at least 10% [at%]. High enrichment samples are of particular

importance for nuclear safeguards, so finding reference materials that are better characterized than those currently in stock is a priority for quality control. The recent efforts to produce such materials with smaller uncertainties as documented in [14] and [18] play an important role for the development of further technical improvement of measurement procedures and mass spectrometry instruments.

Finally, the model provides estimates of the mean $n(\text{U-235})/n(\text{U-238})$ atom amount ratio from each data set. These mean estimates can be compared to the reference values to check for the existence of possible measurement biases.

Fig. 9 Ratio of standard uncertainty of $n(\text{U-235})/n(\text{U-238})$ atom ratio to the model estimate of uncertainty for individual measurements of the CRMs. The red dashed line is $\frac{u(\mu_{ref})}{\hat{u}(y_{ijk})} = 1$



The statistics plotted in Fig. 10 are called “zeta-scores” and “percent relative differences” in ISO 13528 [19]. The statistics are ordered by increasing U-235 content.

$$\zeta = \frac{\hat{\mu} - \mu_{ref}}{\sqrt{\hat{u}^2(\hat{\mu}) + u^2(\mu_{ref})}} \sim N(0, 1)$$

$$\%RD = 100\% \frac{\hat{\mu} - \mu_{ref}}{\mu_{ref}}; \mu_{ref} > 0$$
(4)

Note that $\hat{\mu}$ and $\hat{u}(\hat{\mu})$ are the mean and standard error estimates, respectively, generated from the random effects models. Under general conditions, one would expect to see the 12 zeta-score estimates randomly distributed within the interval $(-2, 2)$ indicating no significant disagreement between the estimated mean values and the certified reference values. Zeta-scores for CRMs NBL U005A, IRMM 184, IRMM 186 and IRMM 187 are outside of this interval indicating an unexpected result. Relative to the time of writing, the IRMM 183–187 reference values were recently re-certified [14]. While the estimates of uncertainty for the certified values have improved over time, the estimates of the n(U-235)/n(U-238) atom amount ratios have also increased in magnitude. Comparing the 2022 reference values and their uncertainties to reference value estimates from an IRMM publication in 2005 [20] shows a negative bias similar in magnitude to that observed in Fig. 10 for materials IRMM 184–187. The origin of the observed bias is currently unknown and subject to further research.

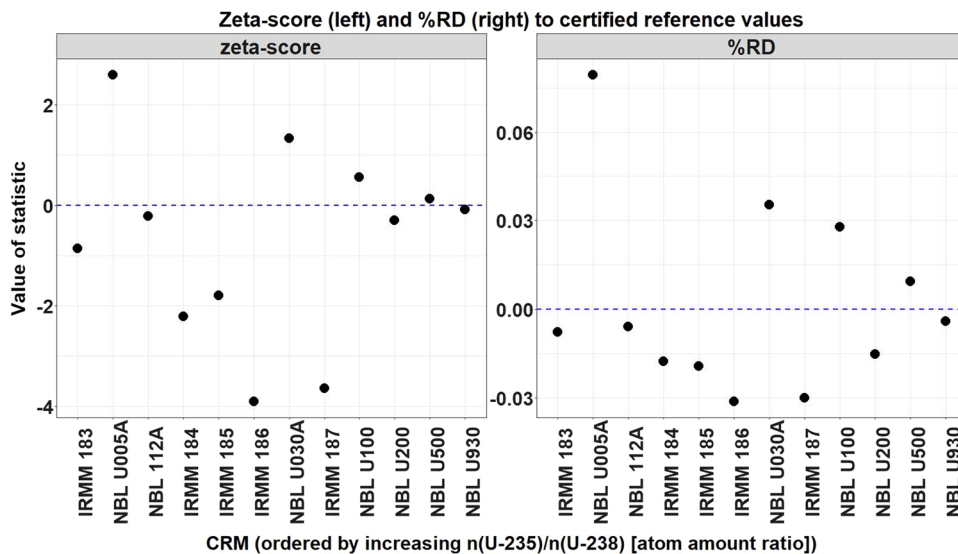
Conclusions

The analysis of the data using a two-factor crossed random effects model indicates that the TIMS TE method is precise enough to detect small, but possibly significant differences in n(U-235)/n(U-238) atom amount ratio measurements attributable to aliquot and/or instrumentation heterogeneity. This approach to uncertainty modelling is an alternative to methods that rely upon defining the uncertainty of each variable of interest and then combining terms using the method of error propagation.

Model estimates and diagnostics were demonstrated to be useful for identifying sources of uncertainty that can be studied in future designed experiments. Boxplots of measurements along with ratios of standard deviation estimates were used to identify variables that are primary contributors to measurement variation. The estimates of the mean measurement results and their uncertainties were also compared to certified reference values using the zeta-score and percent relative deviation statistics. These statistics identified biases between lab results and reference values that can be difficult to observe using only individual measurements.

There is considerable potential for future work applying statistical experimental design methods in mass spectrometry laboratories. Under strictly controlled and expanded sampling conditions, it would be possible to derive statistical intervals for parameters of interest. Interval estimates can be used to quantify the uncertainty in the point estimates generated from this work. Future work should also include more experimental variables to identify potential hidden sources of uncertainty. More advanced sampling models could also be used to decrease the number of QC measurements required to estimate parameters of interest to within a pre-specified precision target.

Fig. 10 Plot of zeta-scores and percent relative differences computed from the model mean estimates and reference values. The dashed blue line indicates perfect agreement between the mean estimate and the certified reference value



Finally, the estimates produced from the model indicate that the uncertainty of the $n(\text{U-235})/n(\text{U-238})$ atom amount ratio values for some reference materials is larger than what is achievable by end-users under routine conditions. This limitation was mainly observed for reference materials whose U-235 content is at least 10% [at%].

Acknowledgements The authors would like to thank the TIMS laboratory analysts Yoshiyasu Kato, Ariana Rugova-Alihajdaraj, Shoichi Saito, Manabu Hagiya and Kenichi Mise who were involved in producing data that have been used in this study.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

- Boulyga S, Konegger-Kappel S, Richter S, Sangely L (2015) Mass spectrometric analysis for nuclear safeguards. *J Anal At Spectrom* 30(7):1469–1489
- Buerger S, Balsley S, Baumann S, Berger J, Boulyga S, Cunningham J, Kappel S, Koepf A, Poths J (2012) Uranium and plutonium analysis of nuclear material samples by multi-collector thermal ionisation mass spectrometry: quality control, measurement uncertainty, and metrological traceability. *Int J Mass Spectrom* 311:40–50
- Mathew K, O’Conner G, Hasozbek A, Kraiem M (2013) Total evaporation method for uranium isotope amount ratio measurements. *J Anal At Spectrom* 28(6):866–876
- ASTM International (2017) Standard test method for determination of uranium or plutonium isotopic composition or concentration by the total evaporation method using a thermal ionization mass spectrometer, ASTM International, West Conshohoken
- Siegmund HHJ, Sturm M, Koepf A, L’Herault C, Boulyga S (2019) Improved TIMS data reliability and precision with new ion source design. *J Anal At Spectrom* 34(5):986–997
- Joint Committee for Guides in Metrology (2008) Evaluation of measurement data — Guide to the expression of uncertainty in measurement, International Bureau of Weights and Measures, Saint-Cloud
- New Brunswick Laboratory (2010) Certificate of Analysis CRM 112-A, New Brunswick Laboratory, Argonne
- New Brunswick Laboratory (2008) Certificate of Analysis CRM U005-A, New Brunswick Laboratory, Argonne
- National Bureau of Standards (1985) Standard reference material U030a. National Bureau of Standards, Gaithersburg
- New Brunswick Laboratory (2008) Certificate of Analysis CRM U100, New Brunswick Laboratory, Argonne
- New Brunswick Laboratory (2008) Certificate of Analysis CRM U200, New Brunswick Laboratory, Argonne
- New Brunswick Laboratory (2008) Certificate of Analysis CRM U500, New Brunswick Laboratory, Argonne
- New Brunswick Laboratory (1987) Certificate of Analysis CRM U930, New Brunswick Laboratory, Argonne
- Richter S, Hennessy C, Jakobsson U, Aregbe Y, Hexel C (2022) Re-Certification of the IRMM-183–187 series of uranium nitrate solution reference materials, Publications Office of the European Union
- Searle SR, Casella G, McCulloch CE (2006) Variance components. John Wiley & Sons Inc., Hoboken
- Galecki A, Burzykowski T (2013) Linear mixed-effects models using R: a step-by-step approach. Springer, New York City
- Koller M (2016) robustlmm: an R package for Robust estimation of linear mixed-effects models. *J Stat Softw* 75(6):1–24
- Richter S, Hennessy C, Venchiarutti C, Aregbe Y, Hexel C (2021) Preparation and certification of highly enriched uranium nitrate solutions IRMM-3000 Series, EUR 30740 EN, Publications Office of the European Union
- International Organization for Standardization (2015) ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparison. International Organization for Standardization, Geneva
- Richter S, Alonso A, De Bolle W, Kühn HVA, Wellum R, Taylor P (2005) Re-certification of a series of uranium isotope reference materials: IRMM-183, IRMM-184, IRMM-185, IRMM-186 and IRMM 187. *Int J Mass Spectrom* 247(1–3):37–39

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.