ELSEVIER

Journal of Nuclear Materials 246 (1997) 30–36

journal of nuclear materials

The Davies–Gray titration for the assay of uranium in nuclear materials: a performance study

M. Bickel

European Commission, Joint Research Centre, Institute for Reference Materials and Measurements, IRMM Retieseweg, B-2440 Geel, Belgium

Received 10 June 1996; accepted 27 March 1997

Abstract

An interlaboratory comparison in two phases was organized to assess the precision and accuracy concerning the assay of uranium in nuclear materials by the potentiometric titration method. This contribution presents the results of this exercise in terms of method performance. Variations to be expected between different laboratories and within a single laboratory are estimated. In general, the method proved again very reliable. © 1997 Elsevier Science B.V.

1. Introduction

The uranium titration method introduced by Davies and Gray [1] (and later improved) [2] is probably the most widely used analytical method for the potentiometric titration of uranium in nuclear materials. It is based upon the reduction of U(VI) to U(IV) followed by a subsequent titration of the U(IV) with potassium dichromate. The uranium sample normally is present as uranium(VI) in nitric acid solution. The largest part of the nitric acid is eliminated by evaporating to dryness and taking up the residue in water. Any excess of nitric acid is destroyed by sulphamic acid, after which the uranium(VI) is reduced to uranium(IV) using an excess of iron(II) in a phosphoric acid medium. The excess iron(II) is then eliminated by nitric acid using ammonium molybdate as a catalyst.

The titration of uranium (IV) to uranium (VI), using standardized potassium dichromate then takes place in a sulfuric acid medium, adding vanadyl sulphate to sharpen the endpoint.

Interference from other elements (in moderate concentrations) is low. There are, however, some disturbing elements such as, for example,

– Al, Hg, which cause sluggish end points;

– Ag, As, Sb, Cl, Br, V, Mo, noble metals, which cause positive bias;

– I, which causes negative bias.

The institute for reference materials and measurements (IRMM) has been applying this method for many years [3], mainly with the goal of certification of reference materials or intercomparison samples, but also for routine analyses [4–6]. Overall uncertainties lie in the order of 0.02 to 0.04% relative in the case of pure uranium materials, somewhat higher for, e.g., uranium/plutonium mixed oxides. The technique also is widely used both by plant operators and for verification analyses by International and Regional Safeguards. Hence, it is important to have knowledge about its general performance in terms of precision and accuracy in order to decide whether or not deviations between results from different laboratories are significant.

Therefore, upon request of the working group on 'low enriched uranium conversion and fuel fabrication plants' of the European safeguards research and development association (ESARDA) IRMM organized an interlaboratory comparison exercise for the determination of uranium by potentiometric titration. The special aim of this intercomparison, apart from giving the laboratories the opportunity to intercompare their results and their individual laboratory deviations from the certified values, was the general assessment of the precision and accuracy of this analytical method.

The exercise was executed in two phases. The first phase dealt with pure uranyl–nitrate solutions [7–9], the second was conducted with similar solutions doped with impurities [10,11].

The whole exercise was an experiment to evaluate precision. Methods of measurement are characterized by their precision and their accuracy. Precision is used as a

PII S0022-3115(97)00040-8

general term for the closeness of agreement between replicate results, by applying an experimental procedure under prescribed conditions. Accuracy is defined as the closeness of agreement between the results of a measurement and the true value. The precision experiment is used as the test method to investigate the type of the variability of the data, and the difference between the results and the certified value(s).

Fourteen and eight laboratories participated in the first and second phase, respectively.

2. Arrangement of the exercise

2.1. General procedure

For the distribution of the samples to be analyzed IRMM prepared two solutions (called A and B) of uranyl–nitrate slightly different in concentration. Both solutions had a nominal uranium concentration of about 7.50 mg U per gram of solution, which had been certified by IRMM. Samples of the solutions A and B were distributed in sealed glass ampoules, each containing about 60 ml. Each of the fourteen participating laboratories received simultaneously three ampoules of solution A and three ampoules of solution B. The certified values were not communicated to the participating laboratories before the report of the results.

In detail the procedure was as follows.

– Two slightly different solutions were prepared (A and B).

– Participants were provided with six ampoules, three of solution A and three of solution B.

– Three days of analyses were foreseen: within two weeks after receipt, four weeks after first analysis and eight weeks after first analysis.

– Ampoules should be opened on day of analysis.

– On each day four analyses were to be performed, two aliquots of each solution.

– Routine methods should be applied.

2.2. Samples

Solutions were prepared from EC-NRM 110 (UO_2 pellets) [6] for both phases. To stay within the precondition of a realistic solution, in the second phase the following impurities were chosen in the respective concentrations:

– Mo(VI): 150 ppm on a U-basis;

– Al(III): 500 ppm on a U-basis;

– V(V): 10 ppm on a U-basis.

Since Mo interferes only at relatively high HNO_3-levels, acidity was decided to be 4 M HNO_3 in the samples. The characterization was done by IRMM after sealing the ampoules to assure that

– samples exactly equal to the participants' samples were analyzed;

– possible changes in concentration during the sealing procedure (evaporation) would not affect the results of the intercomparison.

The following information was given to the participants:

– The samples are realistic for a LEU fuel fabrication plant.

– Uranium concentrations lie in the region of 7.5 mg/g solution.

– The acidity is approximately 4 M HNO_3 (second phase only).

– There are impurities present, chosen from the following list: Fe, Ca, Al, Si, Mo, Cr, Ni, V, Cl, F (second phase only).

2.3. Analysis and result submission

Apart from the submission of the twelve analytical results, the laboratories were asked to give a detailed description of their procedure, as well as their calibrations results including information about the reference materials used. The results of the uranium determinations (expressed as mg g^{-1} solution) and the description of the method were sent to IRMM, where they were coded before the data treatment.

2.4. Data treatment

Many reasons may lead to the variability of data. For this experiment involving many laboratories, factors such as differences in operators, equipment, calibration, environment and sample preparation on different days may give a large contribution to the variability. In practical situations two numerical measures of variability are recommended to take into account differences in the circumstances of new measurements: repeatability and reproducibility. Repeatability refers to the variability of results obtained at short intervals of time in one laboratory by one operator using the same apparatus. Reproducibility refers to the variability of results obtained in different laboratories, which implies different operators, equipment and/or at different times.

2.4.1. Tools applied

2.4.1.1. Preliminary graphical survey. A first impression is obtained by presenting the results in ascending order, together with their (doubled) standard deviations (Fig. 1). The results are compared to the certified values with their overall uncertainties.

Special attention to the search for outliers can be given by means of Youden diagrams [12–14] (Fig. 2), where the results of the measurements for two series (A and B) constitute the coordinates of points in a plot. One pair of results yields a single point in the diagram; furthermore, horizontal and vertical lines denoting the median values

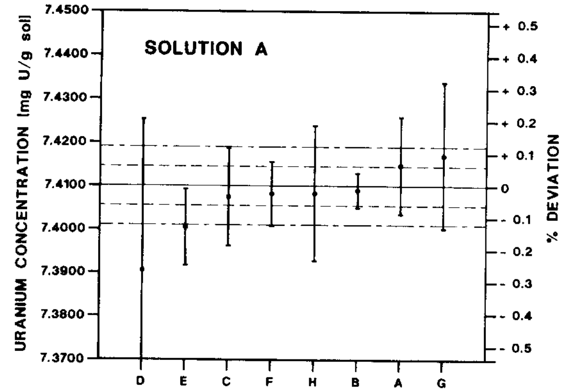

Fig. 1. Presentation of results in ascending order, together with their (doubled) standard deviations for solution A of phase 2 (laboratories coded A to H). The solid line corresponds to the certified value, the different broken lines to its 1s and 2s uncertainties.

are presented on all results in a series. In an ideal situation the points are expected to constitute an elliptic cluster around the point with median value coordinates. Because the two solutions are nearly equal in uranium concentration, it is possible to have an indication of the precision of the data in the figures by means of a circle (centered on the intersection of the median lines). A radius of about 2.5 to 3 times the standard deviation gives the smallest region that can be expected when a normal distribution approximately applies.

Another way to get a rough insight into the homogeneity of a data set is to use a box-and-whisker plot (Fig. 3). There, a box is a rectangle extending in direction of the adscissa between the upper and lower quartile (25 and 75 percentiles) with a central line at the median and whiskers out to the largest and smallest values. If data have a distance from the box larger than 1.5 times the length of the box, they are displayed as single points, i.e., recognized as extreme values. Extensions in direction of the ordinate are arbitrary and have no particular meaning, i.e., the symmetry of the resulting figure in the direction of the abscissa is important.

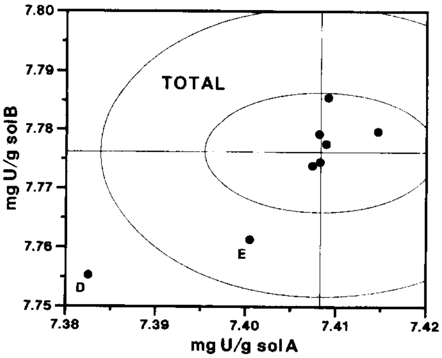

Fig. 2. Youden diagram for solution A (phase 2).

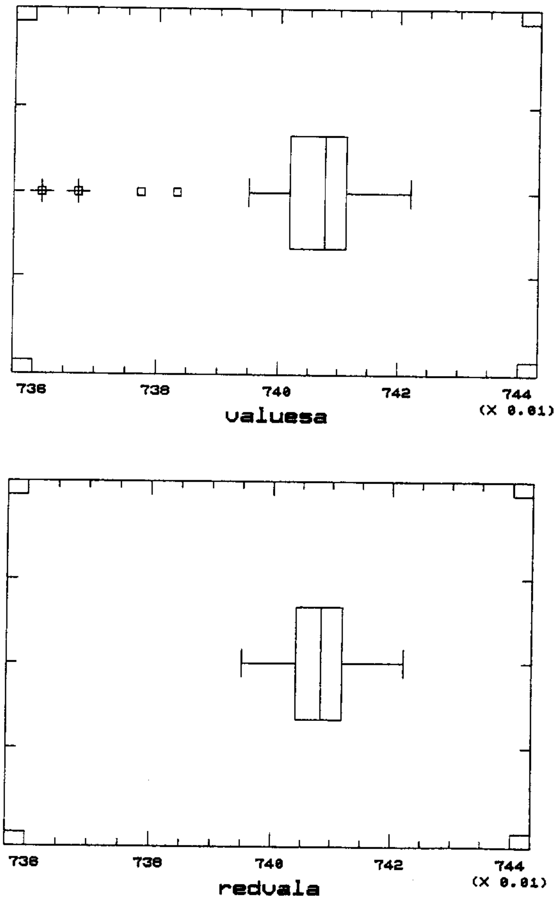

Fig. 3. Box-and-Whisker plots for solution A (Phase 2). Top: Complete data set; Bottom: Reduced data set.

2.4.1.2. Outlier testing. An outlying observation may be an extreme manifestation of the random variability inherent in the data. If this is true, the value should be retained and processed in the same manner as the other observations in the sample.

On the other hand, an outlying observation may be the result of gross deviation from prescribed experimental procedure or an error in calculating or recording the numerical value. In such cases, it may be desirable to institute an investigation to ascertain the reason for the aberrant value. The observation may even actually be rejected as a result of the investigation, though not necessarily so. At any rate, in subsequent data analysis the outlier or outliers will be recognized as probably being from a different population than that of the other sample values.

Three tests were applied, the Dixon Q test [15–17], the T-test recommended for single outliers in [16] and

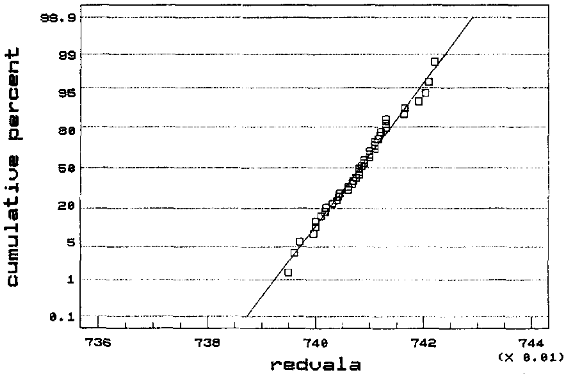

Fig. 4. Normal probability plot for solution A (phase 2).

Cochran's maximum variance test [18], depending on the respective situation.

Attention must be paid that for nearly all outlier tests the assumption is made that the population examined follows a normal distribution.

2.4.1.3. Exploratory statistics. This includes calculation and assessment of basic statistical data, e.g., mean, median, variance, standard deviation, coefficient of variation, skewness and kurtosis.

2.4.1.4. Distribution testing. A simple visual way of seeing whether a set of data is consistent with the assumption of normality is to plot a cumulative frequency curve on normal probability paper (Figs. 4 and 5). If the data come from a normal distribution, one would expect them to lie approximately on a straight line in this presentation.

Another way of looking at distributions is by using a frequency histogram (Fig. 6). A frequency histogram is a display of the distribution of a set of quantitative observations. This procedure first divides the data values into a set of non-overlapping intervals of equal width. Then it plots bars for each interval. The height of each bar is proportional to the number of data values that fall within the interval.

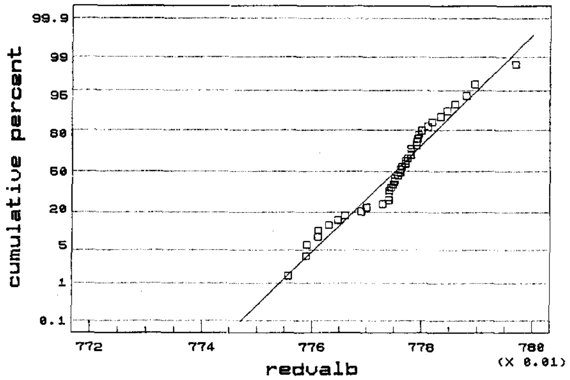

Fig. 5. Normal probability plot for solution B (phase 2).

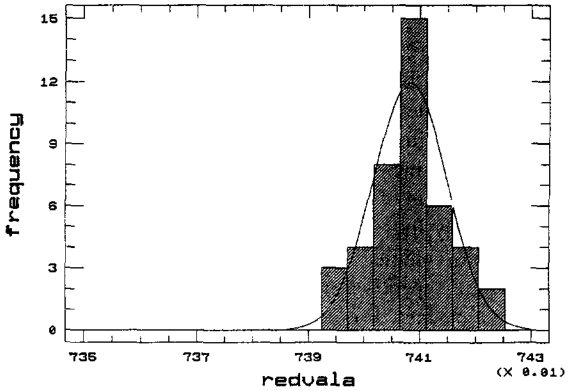

Fig. 6. Frequency histogram with fitted normal distribution for solution A (phase 2).

To enable more quantitative statements the Kolmogorov–Smirnov (K–S) test was performed on the data sets. Its principle involves comparing the cumulative frequency curve of the data to be tested with the cumulative frequency curve of the hypothesized distribution. When the hypothetical and experimental curves have been drawn, the test statistic is obtained by finding the maximum vertical difference between them, and comparing this value in the usual way with a set of tabulated values. If the experimental data depart substantially from the expected distribution, the two curves will be expected to be widely separated over part of the cumulative frequency diagram: if, however, the data are closely in accord with the expected distribution, the two curves will never be very far apart.

As a result, the K–S test leads to the rejection of the hypothesis of normality if its critical value exceeds the respective (tabulated) value. In other words, it answers the question: What is the probability a Gaussian distribution would produce this particular critical value?

2.4.1.5. Analysis of variance. For the interpretation of the results obtained with this exercise, the analysis of variance (ANOVA) technique is a very powerful tool. It gives a good indication of the interlaboratory as well as of the intralaboratory variance and it investigates whether the average outcomes of the laboratories are equal taking into account the experimental uncertainties.

The ANOVA method tests the null-hypothesis that the statistical samples originate from populations with the same mean value. In other words, it investigates the significance of differences between the laboratory outcomes.

From its results, repeatabilities and reproducibilities can be calculated.

2.4.1.6. Comparison with DoD method. A possible drawback of the straightforward application of conventional definitions for averages as well as for standard deviations used is that extreme data can influence the results of these calculations in a serious way. For this reason, and in particular concerning the interlaboratory standard deviation, a comparison with the distribution of differences (DoD) [19–21] between the results of measurements is of interest. The method has proved to be suitable to derive meaningful estimates of the standard deviations for data groups comprising outliers without necessity for identification and deletion of these outliers.

As already indicated, the absolute differences between all possible combinations of the individual observations representing repetitive analytical determinations are used as basic elements. These differences are sorted as a function of their absolute value, after which their frequency is determined. By plotting the cumulative frequency Q of their occurrence versus their absolute values, an impression is obtained of the probability of observing a certain discrepancy.

2.4.2. Procedure used

The data treatment in the first phase included:

- tabulation of data;
- test of homogeneity of the laboratory variances;
- search of the raw data to trace and eliminate outliers;
- analysis of variance;
- calculation of the intra-laboratory variance;
- calculation of the inter-laboratory variance;
- comparison of results with the DoD method;
- comparison of laboratory means;
- comparison of the laboratory means with the certified value.

For the second phase, it was decided to perform the statistical analysis in a less sophisticated manner than for the first phase for the following reasons:

– Due to the reduced number of participants and, therefore, decreased number of measurement results in comparison with the first phase any statistical statements become weaker.

– A simpler statistical analysis would consume less time; therefore, participants could be provided with results faster.

The data treatment in the second phase included:

- graphical survey;
- outlier testing;
- exploratory statistics;
- distribution testing;
- analysis of variance;
- calculation of the intra-laboratory variance;
- calculation of the inter-laboratory variance;
- comparison of results with the DoD method;
- comparison of laboratory means;
- comparison of the laboratory means with the certified value.

3. Results

The number of single results was 168 for the first phase, 96 for the second phase.

3.1. Methods and reference materials

The laboratories reported information in a non-uniform manner going from no information or only a few words to a full description.

There were important variations of the working procedure from lab to lab according to the use of the original method [1] or the use of the method calling upon addition of vanadium as described by Eberle et al. [2]. Other variations (concentration and volume of reagents, intervals of time between the successive steps of the chemical reactions, approach towards the end-point) are certainly present. Some laboratories used automatic titrators, some did not mention the type of equipment. It was, thus, difficult to establish a link between procedures applied. It also seemed that some labs had good experience in applying the method, others being less acquainted with it. This information was confirmed after separate discussions with the labs.

All the reference materials used, except in the case of one laboratory in the first phase (use of metal chips), seem to have been suitable reference materials. Nevertheless, some of them have better confidence limits attached to their certified value due to a more elaborated characterization. The handling of the reference material by the laboratory might influence the quality of the reference solution because some of the materials used need a preliminary pretreatment like a calcination or the cutting of a large piece followed by a chemical cleaning of the surface. The failure to follow rigorous procedures when dissolving the material (avoiding loss of material when heating solutions, contaminations,...) might influence in a particular laboratory the accuracy of the calibration.

3.2. Generally occurring errors

The supplementary information as supplied by the laboratories gave an indication that in the first phase only one laboratory had incorporated buoyancy correction in its results. The correction is necessary to take into account the difference in the densities of uranium materials, solutions and (brass) weights, and its magnitude depends on the mass values for the analysis. For the determination of the certified uranium concentration of solution A and B this correction was also applied. For this reason the results coming from the other laboratories were also corrected for this effect by IRMM using a common factor (0.105%).

In phase 2 the situation had somewhat improved: in only two cases buoyancy correction was not applied. These values again have been adjusted by IRMM using a correction of 0.105% as for the first phase.

Only one laboratory determined the atomic weight of the uranium via the isotopic composition using mass spectrometry. EC-NRM 110, of which the solutions were made, contains depleted uranium. Its atomic weight is given in the certificate as 238.041 g/mol. Natural uranium has an atomic mass of, for example, 238.03 as given in the Certificate of CBNM-NRM 106 [5]. This leads to an error of 0.005% relative, if participants used the data of natural uranium instead of doing a mass spectrometric analysis. That error still lies one order of magnitude below the variances in the intercomparison and, therefore, does not affect the results significantly. Hence, no corrections were applied in this case.

3.3. Outlying data

Examination of the results of phase 1 made it clear that two (very extreme in comparison with the laboratory variances) outliers were present for two laboratories at nearly all levels (Fig. 1). Therefore, it was concluded that it was justified to discard all the data coming from these laboratories. After this deletion some statistical outliers still remained at single levels for three other laboratories. Due to the repeated application of the tests and the fact that these outliers occur only at single levels it was not thought to be justified to discard these data at this stage of the analysis.

In phase 2 there was evidence for the results of one laboratory being (slightly) aberrant values; however, it was considered doubtful to reject these results on statistical grounds, with a physical reason not recognizable (see above, Section 2.4.1).

Several important observations could be made by comparing results from the complete and the reduced (possible outlyers deleted) data sets. In all cases

- the means were closer to the certified values,
- the means and medians were closer together,
- the variances and hence standard deviations were significantly smaller,
- the skewness and kurtosis were significantly smaller for the reduced data sets in comparison with the complete ones. This fact is in agreement with above results, i.e., results from some laboratories seemed to have some disturbing effect on the ensemble of observations.

From Fig. 3 it is clearly seen that, e.g., in the case of solution A, phase 1, four results in this presentation would have to be considered as extreme. They all came from one laboratory. With the reduced data set, the picture looks much more symmetric. In the case of solution B of the same phase this was not as clear. The box-and-whisker plot became more symmetric for the reduced data set, however some more extreme values seemed to be present both on the lower and upper end of the distribution.

In such cases, one problem of successive outlier rejection becomes evident: with one (or some) aberrant values excluded the distribution becomes so tight that suddenly other data would have to be tested, leading in the extreme to the situation with a very tight distribution of several data with the rest all being outliers. However, it does not seem advisable to do so, since for neither of these data physical reasons are known which would justify a rejection.

After fitting of the respective distributions (phase 2) it was obvious that the reduced sets at least were much closer to normal distributions than the complete ones. From the results of the K–S test one would have had to reject the hypothesis of normality only for one of the complete data sets. In agreement with previous conclusions, the large difference between the complete and reduced data sets pointed towards existing problems for the complete ones. In spite of this fact, the distributions had to be accepted as normal since there was not enough evidence to reject this hypothesis.

Hence, the conclusion in both phases was that the following statistical treatment had to be done for both the complete and the reduced data sets.

3.4. Repeatabilities and reproducibilities

ANOVA was performed on all eight data sets (complete and reduced, two solutions in two phases) according to the procedure described in [9]. What is, again, common to both phases is the fact that the exclusion of the laboratories under question leads to a drastic change in the statistical properties of the data sets. Furthermore, at least one laboratory mean value of each of the reduced data sets deviated significantly from the others. It was, however, not decided to exclude these laboratories due to the reasons mentioned above.

For the calculation of repeatabilities and reproducibilities, only the reduced data sets were considered. The procedure is described in [9]. Results are shown in Table 1. Therein, also, the standard deviations as derived using the DoD method are listed.

3.5. Accuracies

The deviations of a lot of laboratory mean values with respect to the certified values are larger than the standard deviation stated for certification, in spite of the common buoyancy correction applied to the uranium contents reported by the laboratories. This was particularly the case in

Table 1
Precision parameters (repeatability and reproducibility from ANOVA, standard deviation from DoD) (relative values)

	Phase 1	Phase 2
Repeatability standard deviation		
Solution A	0.08%	0.08%
Solution B	0.20%	0.11%
Inter-laboratory standard deviation		
Solution A	0.11%	0.05%
Solution B	0.16%	0.21%
Reproducibility standard deviation		
Solution A	0.13%	0.09%
Solution B	0.26%	0.24%
Standard deviation derived by DOD using all data		
Solution A	0.15%	0.11%
Solution B	0.18%	0.12%

phase 1. The means of the reduced data sets were closer to the certified values than those of the complete sets.

For calibration purposes of the titration method used, verification by comparison with an appropriate internationally recognized uranium reference material is necessary. Either pure uranium metal (for example EC-NRM 101 or NBS 960) or pure U_3O_8 (for example NBS 950 b), or sintered uranium dioxide (EC-NRM 110) of known purity may be used. Part of the calibration procedure is the dissolution of reference material. Considering a possible dissolution effect, in phase 1 the laboratories were clustered according to the type of reference material used. Based on these data the general idea arose that the use of uranium oxide as reference material causes roughly a two or three times higher standard deviation.

Also, in the second phase, the means of the reduced data sets were closer to the certified values than those of the complete sets. This qualitative finding was investigated by statistical testing of the hypothesis: the observed mean equals the certified value, versus the alternative 'not equal'. Again earlier observations were confirmed: the reduced data set for Solution A was rather homogeneous and in disagreement with the certified value, while for solution B there still seemed to be some 'trouble-causing' data even in the reduced set.

In general, deviations from the certified values were smaller in the second than in the first phase.

4. Conclusions

The original data sets consisting in total of 168 (1st phase) and 96 (2nd phase) observations from fourteen and eight laboratories, respectively, contained some disturbing data, i.e., statistical outliers. Since, on the one hand, no direct physical reason was obvious, and on the other hand only two laboratories had outliers on all levels, the statistical evaluations were performed on the complete, as well as on the reduced data sets. In the course of the evaluations it became more evident that other laboratories might have delivered aberrant values.

The overall variances are two to threefold in the complete data sets in comparison with the reduced ones. For the cleaned data sets no significant deviation from a normal distribution could be proven statistically.

The overall results in terms of laboratory performances are satisfying. The following set of conclusions could be drawn from the statistical evaluation:

- The best estimates for the repeatability standard deviations were in the range 0.08% to 0.20% relative.
- The best estimates for the inter-laboratory standard deviation were in the range 0.11% to 0.21% relative.
- The best estimates for the reproducibility standard deviations were in the range 0.09% to 0.26% relative.
- A relative standard deviation of about 0.25% described very well the total reproducibility variability in the inter-laboratory comparison.
- An experienced laboratory will have a within-laboratory standard deviation of better than 0.08%.
- There are indications that the use of the uranium-oxide reference material causes roughly a two times higher standard deviation than the metal reference material.

In both phases of this intercomparison the Davies–Gray potentiometric titration method proved very reliable for the uranium assay in nuclear materials.

If one compares the results for precision parameters between the two phases, they are mostly slightly lower in the case of solution A and slightly higher in the case of solution B. Hence, it may be concluded that the presence of the particular impurities added to the analyte solutions has no significant influence on the performance of the Davies–Gray method.

The importance of such kind of intercomparisons is reflected in the fact that not only are the participating *laboratories* enabled to check their own performances in terms of precision (comparison with other labs) as well as accuracy (comparison with certified value), but that also a general assessment of the performance of an analytical *method* itself becomes possible.

References

[1] W. Davies, W. Gray, Talanta 11 (1964) 1203.
[2] A. Eberle, R. Lerner, M. Goldbeck, C. Rodden, NBL-252, 1970.
[3] V. Verdingh, A. Michiels, J. Nucl. Mater. 153 (1988) 96.
[4] Y. Le Duigou, EUR-8349, 1983.
[5] Y. Le Duigou, V. Verdingh, Ant. Comm. CEC 4146, 1984.
[6] Y. Le Duigou, EUR-8753, 1984.
[7] V. Verdingh, Y. Le Duigou, EUR-13379, 1991.
[8] V. Verdingh, Y. Le Duigou, M. Bickel, EUR-13686, 1991, p. 195.
[9] D.J. Ketema, R.J.S. Harry, W.L. Zijp, ECN-C-90-043, 1990.
[10] M. Bickel, A. Michiels, F. Hendrickx, EUR-15078, 1995.
[11] M. Bickel, A. Michiels, F. Hendrickx, EUR-15214, 1993, p. 643.
[12] W.J. Youden, Ind. Qual. Contrib. 15 (1959) 24.
[13] W.J. Youden, J. Qual. Techn. 4 (1972).
[14] W.J. Youden, E.H. Steiner, Statistical Manual of the AOAC (AOAC, Arlington, VA, 1975).
[15] J.C. Miller, J.N. Miller, Statistics for Analytical Chemistry (Ellis Horwood, Chichester, 1984).
[16] ASTM publication E-178-80 Standard Practice for Dealing with Outlying Observations, 1989.
[17] W.J. Dixon, Biometrics 9 (1973) 74.
[18] E.S. Pearson, H.O. Hartley, Biometrika Tables for Statisticians (Cambridge University, 1970).
[19] W. Beyrich, W. Golly, G. Spannagel, Proc. 3rd Annual Symp. on Safeguards and Nuclear Material Management, Karlsruhe, May 1981, Report ESARDA-13, CEC/JRC, Ispra, 1981.
[20] W. Beyrich, W. Goily, N. Peter, R. Seifert, KfK-4721 ,1990.
[21] W. Beyrich, W. Goily, N. Peter, R. Seifert, EUR-11398, 1990.