

Use of accumulated proficiency testing scheme results to evaluate laboratory performance in forensic ethanol analysis

Marcellé Archer^{id}, Dominique N. Marajh^{id},
Maria S. Visser^{id} and Maria Fernandes-Whaley^{id*}

National Metrology Institute of South Africa, Private Bag X34, Lynnwood Ridge, 0040,
South Africa

(Received April 28, 2017; Revised June 8, 2017; Accepted June 13, 2017)

Abstract: In 2004, the National Metrology Institute of South Africa, NMISA (then the National Metrology Laboratory of the CSIR, South Africa) was requested by forensic testing laboratories to coordinate a proficiency testing scheme for the determination of ethanol in blood. This led to an annual proficiency testing scheme where the test materials used are aqueous ethanol certified reference materials prepared and certified by NMISA. This article describes the development of the scheme and the progress made by the regular participants in the scheme over twelve years. The ethanol concentrations of the test samples were between 0.01 g/ 100 g and 0.6 g/ 100 g. The measurement uncertainty and the expected method performance were conveniently estimated from the data collected over the years.

Keywords: Ethanol; proficiency testing; blood alcohol analysis; metrological traceability. © 2017 ACG Publications. All rights reserved.

1. Introduction

The accurate determination of ethanol content in various matrices is critical for regulatory forensic and trade purposes. For forensic purposes, ethanol is determined for the effective regulation of blood alcohol levels (e.g., driving under the influence of alcohol, autopsies, etc.). As a trading commodity, ethanol content is determined for the classification, control and taxation of alcoholic beverages.

The National Metrology Institute of South Africa (NMISA) has demonstrated its capability to accurately determine forensic ethanol concentrations through participation in the Comité Consultatif pour la Quantité de Matière (CCQM) international comparisons, CCQM-K27.1 [1], CCQM-K27.2 [2] and CCQM-K79 [3], coordinated through the International Committee for Weights and Measures (BIPM) [4]. NMISA currently has calibration and measurement capabilities (CMCs) in the International Bureau of Weights and Measures [4] key comparison database (KCDB) for the determination of aqueous ethanol solutions.

NMISA is ISO Guide 34 accredited to gravimetrically prepare and certify aqueous ethanol solutions in the concentration range 0.01 g/ 100 g to 20 g/ 100 g. The gravimetric concentration is confirmed titrimetrically. Aliquots taken from the aqueous ethanol solution are oxidised with a known quantity of potassium dichromate and then a redox back-titration of the excess potassium dichromate with iron ammonium sulphate is performed [5].

* Corresponding author E-mail: mfwhaley@nmisa.org; Phone: +27 841 2105; Fax: +27 086 509 5996

Proficiency testing (PT) as described in the IUPAC [6] and ISO [7] protocols is a widely-recognised practice for monitoring analytical performance and which, with appropriate additional activities, improves the performance of analytical laboratories [8]. It is not only essential for laboratories and their customers to have ongoing confidence in their competence, but also for stakeholders such as regulators, laboratory accreditation bodies and other organisations.

Starting in 2004, the NMISA has been coordinating proficiency testing schemes for South African laboratories for ethanol analysis. For the first round in 2004, blood and serum certified reference materials were purchased and re-labelled before distribution to the participants. A need for continuous and affordable proficiency testing was identified and thus annual proficiency testing schemes followed, using aqueous ethanol certified reference solutions as the test materials. In-house commutability tests at that time confirmed that headspace analysis of ethanol in a blood matrix and in an aqueous matrix yielded the same results.

Each proficiency testing scheme initially consisted of four rounds per annum with three samples of different concentrations per round, but in 2008 the number of rounds per year was reduced to three (still with three samples per round).

In 2011, NMISA also coordinated an inter-Africa PT and an international (AFRIMETS.QM-K27 [9]) interlaboratory comparison for ethanol analysis. Participants from laboratories in six African countries participated in the former scheme, while there were participants from ten international laboratories in the latter study.

This article will focus on the PT for South African testing laboratories in order to highlight the benefits of regular participation in PT schemes and how data accumulated from proficiency testing results over time may provide useful information on the expected method performance and measurement uncertainty of the applied analytical methods [10, 11].

For over forty years, ethanol in blood and urine forensic samples have been analysed by headspace gas chromatography [12]. The majority of participants in the NMISA ethanol analysis PT schemes have been applying headspace gas chromatography with flame ionisation detection on a routine basis. Few participants apply gas chromatography-mass spectrometry for this analysis.

In addition to the measurement results, PT participants are also requested to submit raw data. The raw data is used to compare the measurement uncertainties submitted by participants against the proposed performance model experimentally determined from the accumulated proficiency test result data.

Participant result data is evaluated according to ISO 13528 statistical principles [13]. The PT reports provided are very detailed and give specific feedback and recommendations to improve participant performance, where required.

2. Experimental

2.1. Samples, chemicals and instrumentation

Aqueous ethanol solutions were prepared by mixing a known mass of pure ethanol (Sigma-Aldrich 99.8% purity) and ultra-pure water (Milli-Q 18.2 M Ω .cm). The solutions were bottled in 25 ml or 50 glass containers.

Depending on batch size, the ethanol in eight to twelve aliquots of the solution from selected containers was oxidised with known quantities of potassium dichromate (4.9 g/ kg, CRMs from National Institute for Standards and Technology, NIST, or National Metrology Institute of Japan, NMIJ) in the presence of sulphuric acid. The excess dichromate was titrated with freshly prepared iron ammonium sulphate (0.05 mol/ l, Sigma-Aldrich), using Mettler-Toledo DL53, DL55 or T70 autotitrators with a combined platinum ring electrode, and the concentration of the ethanol determined [5]. This procedure also evaluates the purity of the ethanol and the homogeneity of the solutions.

Extensive stability studies have been done on these unpreserved solutions since 2000; solutions with ethanol concentrations lower than 0.02 g/ 100 g are certified stable for twelve months from date of preparation and solutions with higher concentrations are certified stable for twenty-four months from date of preparation.

2.2. Protocol

The scheme is designed so that the concentrations for each round test a specific aspect of the analysis, such as whether the method can distinguish between two concentrations close to each other and examines the magnitude of the measurement uncertainties around the legal blood ethanol limit. The ethanol concentrations ranged between 0.01 g/ 100 g to 0.6 g/ 100 g.

Since 2008, each proficiency testing scheme consisted of three rounds per year and three samples of different ethanol concentrations were sent to participants in each round. Participants were asked to analyse two aliquots from each sample, preferably on two different analytical systems under routine analytical conditions for their laboratories. More than one analyst per laboratory could participate.

Participants were informed that the results from previous proficiency testing schemes would be used to estimate the expected performance of the participants. These limits would provide a quality control guideline for participants, but laboratories were encouraged to estimate and set their own acceptable quality control parameters for methods in the laboratory.

3. Results and discussion

The aqueous ethanol samples distributed to the participants were solutions with certified concentrations. The evaluation of the results was therefore much easier with samples of exactly known concentration, than with samples for which consensus values had to be agreed on. The number of PT participants and participant competence levels are also no longer a constraint when the metrologically traceable certified value of the CRM is used as the assigned value for the PT. The main challenge was to propose a realistic fit-for-purpose standard deviation (s) for proficiency assessment based on sound statistical principles.

Initially, the Horwitz equation [14] was used to estimate the worst acceptable precision. It is generally accepted that values falling within two standard deviations ($\pm 2s$) of the target concentration are satisfactory, while values falling between $\pm 2s$ and $\pm 3s$ need investigation. Values outside the $\pm 3s$ limits are unacceptable.

The Horwitz model provided rather wide limits: $\pm 2s$ of 12.6% at the 0.05 g/ 100 g ethanol level (the legal blood alcohol limit in South Africa) lowering to around 8.5% for $\pm 2s$ at an ethanol concentration of 0.5 g/ 100 g. After two schemes of four rounds each (twenty four samples) it was apparent that the headspace gas chromatography methods used by the participants were capable of producing more accurate ethanol results. Additionally, since the blood alcohol results were to be used in law enforcement, a measurement uncertainty as small as possible is preferred.

In 2008, the general model of Thompson, *et al* [15] was utilised to evaluate the precision of the data. The model was adapted for the simplest approach available, with Microsoft Excel to process the data, so that sophisticated statistical software need not be used.

Firstly, the percentage deviation of the mean of each pair of results from the certified concentration was calculated from Equation (1).

$$\%deviation\ from\ certified\ value = \left(\frac{\bar{x} - \mu}{\mu}\right) * 100 \quad (1)$$

Where:

\bar{x} is the mean of two measurements for the same sample in g/ 100 g submitted by the analyst and
 μ is the certified ethanol concentration in g/ 100 g of the sample

These deviations were then plotted as a function of concentration. Figure 1 shows a plot of the data from PT schemes from 2005 to 2008, where there are 155 pairs of results. From this graph it is

evident that the observed deviations are not constant relative, as the percentage deviations would then be constant over the whole range. The deviations also do not become progressively and smoothly smaller from high to low concentration, thus the observed deviations are not constant absolute.

There is no evidence of bias since the deviations seem to fall approximately equally on either side of the zero deviation line.

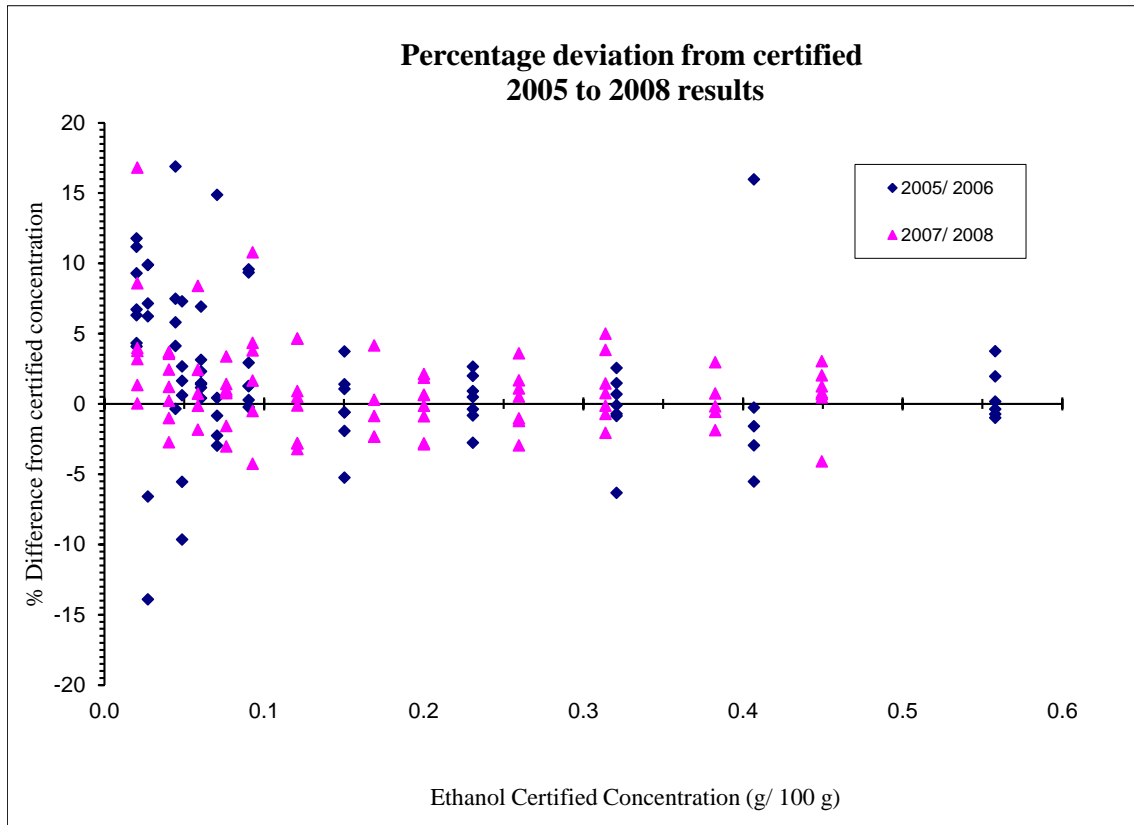


Figure 1. Plot of percentage deviation from the certified concentration of results between 2005 and 2008, as a function of ethanol concentration.

Equation (2) proposes a reasonable model [15] which is easy to understand and is defensible in the literature. In equation (2) u is the expected standard deviation for the proficiency testing scheme.

$$f(x) = u = \sqrt{a^2 + (bx)^2} \quad (2)$$

Where:

a is the intercept,

b is the slope and

x is the ethanol concentration, in g/ 100 g

The relative percentage difference between the expected uncertainty for each concentration (u in Equation (2)) and the analyst's deviation from the certified concentration is calculated. The sum square of this range of values divided by the number of values is then used in the Microsoft Excel Solver equation to minimise for a and b .

An absolute model was also used in Solver to minimise for a and b , using Equation (3) as the solver formula.

$$\sqrt{\frac{(\text{analyst deviation} - u)^2}{n}} \quad (3)$$

Where:

u is calculated from Equation (2) and
 n is the number of data points

The values for a and b using the relative model and the absolute model were very similar. From these values, the $\pm 1s$ and $\pm 2s$ limits could be calculated using $a = 0.00148$ and $b = 0.0272$. These limits were then plotted against the actual percentage deviations for each pair of results submitted. A graph showing the results from December 2005 to March 2008 is given in Figure 2. The continuous red lines closest to zero bias represent the expected $\pm 1s$ limits and the continuous red lines furthest from zero bias represent the $\pm 2s$ (expected U , the expanded uncertainty) limits. For this model to be valid, five percent or less of the results should fall outside the $\pm 2s$ limits (in a normal distribution approximately 95% of the results will lie within $2s$ of the mean and 99.7% of the results will lie within $3s$ of the mean). Fourteen results out of one hundred and fifty-five values fell outside these limits (9%).

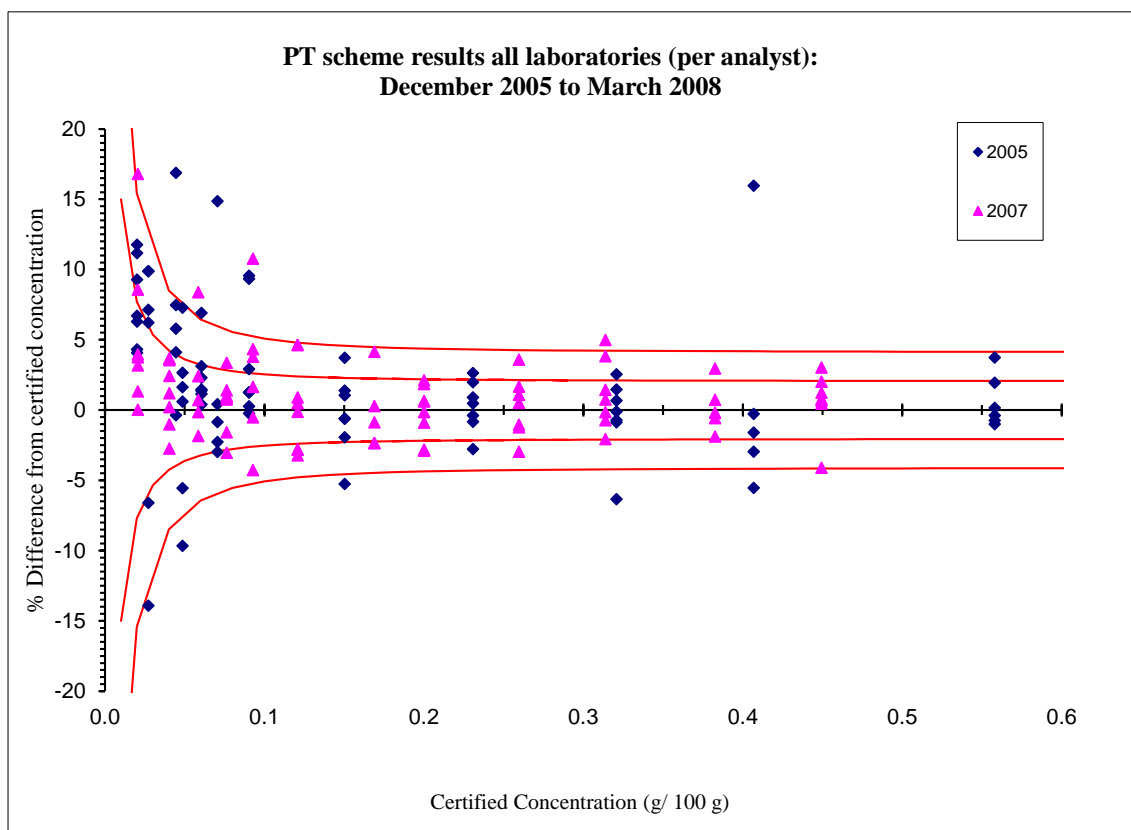


Figure 2. Graph of the pairs of the results ($n = 155$) for the proficiency testing schemes between December 2005 and March 2008, using the standard deviations calculated from the proposed models. The continuous red lines closest to zero bias represent the expected $\pm 1s$ limits. The continuous red lines furthest from zero bias represent the $\pm 2s$ (expected U) limits.

By 2009 it became evident that the participant's proficiency had improved steadily and that the data fit into the proposed model much better. Comparing the initial Solver model and the Horwitz model showed that the Horwitz model tended to underestimate the accuracy in the lower concentration range and overestimate the accuracy in the higher ethanol concentration range. Experience also showed that in routine analyses, it was not always possible to reduce the regression uncertainty to

within the Solver model limits. A combination of the realistically achievable regression uncertainty and the Solver model uncertainty was then proposed. Table 1 shows the respective $\pm 1s$ (u) limits for the original Solver model, the Horwitz model and the proposed combined model.

The proposed combined model was applied to the data generated from the results of two ethanol proficiency testing schemes run from May 2009 to March 2010. The model can be seen in Figure 3, where the continuous red lines closest to zero bias represent the expected $\pm 1s$ limits and the continuous red lines furthest from zero bias represent the $\pm 2s$ (expected U) limits. None of the results exceeded $\pm 2s$. Even when the original solver model standard deviations listed in Table 1 were used none of the results exceeded $\pm 2s$. This was excellent because the limits for this model are much lower at higher concentrations than for the Horwitz model, especially around the legal blood alcohol limit for South Africa.

When the results from three proficiency testing schemes run between July 2010 and August 2013 were plotted according to the proposed model, ten out of two hundred and twenty-two results were outside of the $\pm 2s$ limits (4.5%). Adopting this model as the expected ethanol proficiency testing scheme performance is therefore justified. The data are plotted in Figure 4. Similarly, for a scheme run from February 2015 to March 2016, only about 5% of the results fell outside the $\pm 2s$ limits when using the combined model standard deviations in Table 1. Whereas all the results previous to 2015 were generated by headspace gas chromatography with tertiary butanol as an internal standard, the 2015/2016 scheme also included results from a gas chromatography-mass spectrometry (GC-MS) system with ethanol- d_6 as the internal standard. The GC-MS results do not yet quite fit into the model but as a new participant it is not yet clear if this is due to the system or to a need for improved proficiency.

Table 1. Respective $\pm 1s$ (u) limits for the original Solver model, the Horwitz model and the proposed combined model

Ethanol concentration (g/ 100 g)	Solver model $\pm 1s$ (u)	Horwitz model $\pm 1s$ (u)	Combined model $\pm 1s$ (u)
0.01	15.0	8.0	15.0
0.02	7.7	7.2	7.9
0.03	5.4	6.8	5.6
0.04	4.2	6.5	4.6
0.05	3.6	6.3	4.0
0.06	3.2	6.1	3.7
0.07	3.0	6.0	3.5
0.08	2.8	5.9	3.3
0.09	2.6	5.7	3.2
0.1	2.5	5.7	3.1
0.2	2.2	5.1	2.8
0.3	2.1	4.8	2.8
0.4	2.1	4.6	2.8
0.5	2.1	4.4	2.8
0.6	2.1	4.3	2.8

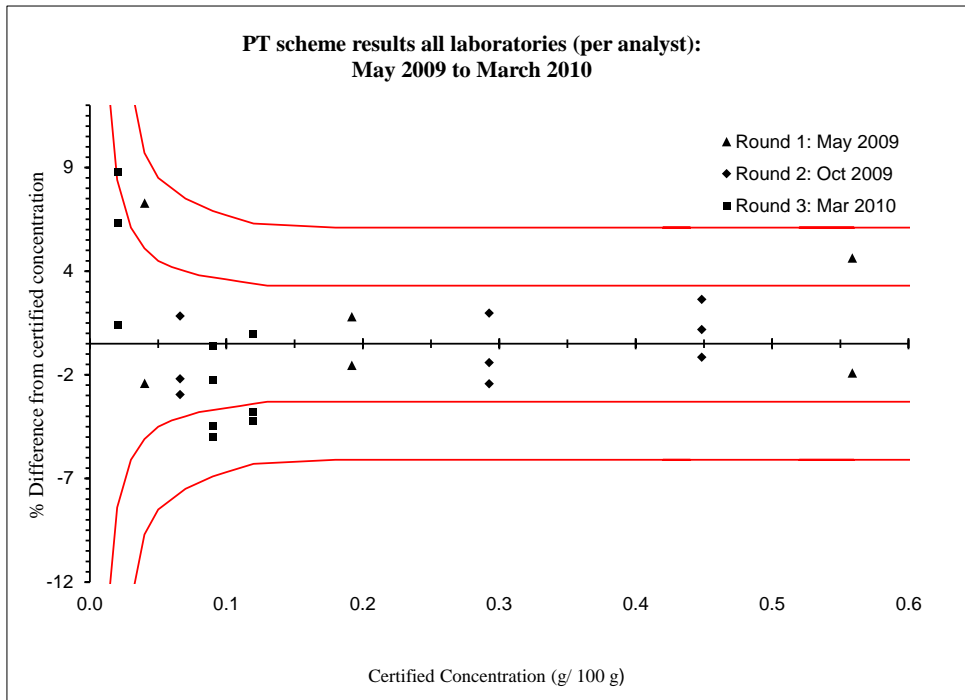


Figure 3. Graph of the pairs of the results ($n = 27$) for two proficiency testing schemes between May 2009 and March 2010, using the standard deviations calculated from the proposed models. The continuous red lines closest to zero bias represent the expected $\pm 1s$ limits. The continuous red lines furthest from zero bias represent the $\pm 2s$ (expected U) limits.

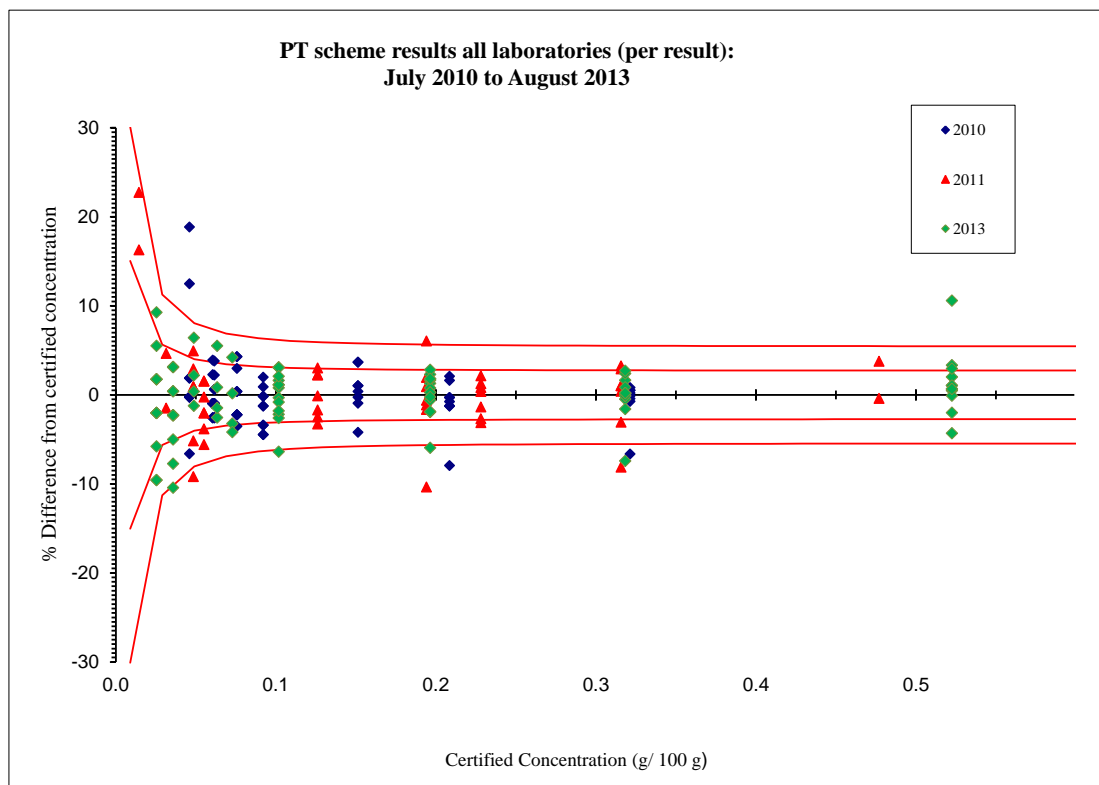


Figure 4. Graph of all the results ($n = 222$) for three proficiency testing schemes between July 2010 and August 2013, using the standard deviations calculated from the proposed models. The continuous red lines closest to zero bias represent the expected $\pm 1s$ limits. The continuous red lines furthest from zero bias represent the $\pm 2s$ (expected U) limits.

4. Conclusion

From the results of proficiency testing from 2004 to 2016 for aqueous ethanol samples it was possible to estimate the expected performance of such a scheme, using a relatively simple model. The model can be generated using software that is commonly available and there is no need for special statistical software. The proposed model was shown to be statistically valid over several proficiency testing schemes.

5. Acknowledgements

The authors wish to thank Dr. David L. Duewer (National Institute for Standards and Technology) for his statistical inputs and guidance.

ORCID

Marcellé Archer: [0000-0001-6999-8604](https://orcid.org/0000-0001-6999-8604)

Dominique Marajh: [0000-0001-8330-1527](https://orcid.org/0000-0001-8330-1527)

Ria Visser: [0000-0001-6587-2203](https://orcid.org/0000-0001-6587-2203)

Maria Fernandes-Whaley: [0000-0003-0476-2937](https://orcid.org/0000-0003-0476-2937)

References

- [1] M.M. Schantz, D.L. Duewer, R.M. Parris, W.E. May, M. Archer, C. Mussell, D. Carter, L.A. Konopelko, Y.A. Kustikov, A.I. Krylov and O.V. Fatina (2005). CCQM-K27-Subsequent: Key Comparison (subsequent) for the determination of ethanol in aqueous matrix, *Metrologia*, **42**, Technical Supplement.
- [2] M.M. Schantz, R.M. Parris, W.E. May, A. Rosso, C. Puglisi, J. Marques Rodrigues Caixeiro, G. Massiff, E. Camacho Frías, M. Pérez Urquiza, M. Archer, M.S. Visser and B. de Vos (2013). Final report on CCQM-K27.2: Second Subsequent study: determination of ethanol in aqueous media, *Metrologia*, **50**, Technical Supplement.
- [3] S. Hein, R. Philipp, D.L. Duewer, H. Gasca Aragon, K.A. Lippa and B. Toman (2013) Final report on CCQM-K79: Comparison of value-assigned CRMs and PT materials: Ethanol in aqueous matrix, *Metrologia*, **50**, Technical Supplement.
- [4] <http://www.bipm.org>, accessed 29 March 2017.
- [5] M. Archer, B. de Vos and M.S. Visser (2007). The preparation, assay and certification of aqueous ethanol reference solutions, *Accred Qual Assur*, **12**(3), 188-193.
- [6] M. Thompson, S.L.R. Ellison and R. Wood (2006). The international harmonized protocol for the proficiency testing of analytical chemistry laboratories (IUPAC technical report), *Pure Appl. Chem.*, **78**, 145-196.
- [7] ISO/IEC 17043 (2010): Conformity assessment – general requirements for proficiency testing
- [8] The Royal Society of Chemistry (2010), The role of proficiency testing in method validation, *Accred Qual Assur*, **15**(2), 73-79.
- [9] M. Archer, M. Fernandes-Whaley, R. Visser, J. de Vos, S. Prins, A. Rosso, M. Ruiz de Arechavaleta, I. Tahoun, E. Kakoulides, C. Luvonga, G. Muriira, E. Naujalis, O. Bin Zakaria, M. Buzoianu, J. Bebic, B. Achour Mounir and N. Huy Thanh (2013). Final report on AFRIMETS.QM-K27: Determination of ethanol in aqueous matrix, *Metrologia*, **50**, Technical Supplement.
- [10] D.L. Duewer, M.C. Kline, K.E. Sharpless, J.B. Thomas, K.T. Gary and A.L. Sowell (1999). Micronutrients measurement quality assurance program: helping participants use interlaboratory comparison exercise results to improve their long-term measurement performance, *Anal Chem*, **71**, 1870-1878.
- [11] B. Magnussen, T. Näykki, H. Hovind and M. Krysell (2003). Handbook for calculation of measurement uncertainty in environmental laboratories.
- [12] T. Thompson (1985). Determining blood/ alcohol concentration: Two methods of analysis, *Montana Law Review*, **46**(2), article 9.

- [13] ISO 13528 (2005). Statistical methods for use in proficiency testing by interlaboratory comparisons.
- [14] T. Linsinger and R Josephs (2006). Limitations of the application of the Horwitz equation, *Trends in Analytical Chemistry*, **25(11)**, 1125-1130.
- [15] M. Thompson, K. Mathieson, A.P. Damant and R. Wood (2008). A general model for interlaboratory precision accounts for statistics from proficiency testing in food analysis, *Accred Qual Assur*, **13**, 223-230.

A C G
publications

© 2017 ACG Publications