Science & Justice 59 (2019) 9–19

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

Lessons learned from inter-laboratory studies of carbon isotope analysis of honey

Philip J.H. Dunn[a,*], Sarah Hill[a], Simon Cowen[a], Heidi Goenaga-Infante[a], Mike Sargent[a], Ahmet Ceyhan Gören[b,1], Mine Bilsel[b], Adnan Şimşek[b], Nives Ogrinc[c], Doris Potočnik[c], Paul Armishaw[d], Lu Hai[e], Leonid Konopelko[f], Yan Chubchenko[f], Lesley A. Chesson[g], Gerard van der Peijl[h], Cornelia Blaga[h], Robert Posey[i], Federica Camin[j], Anatoly Chernyshev[k], Sadia A. Chowdhury[l]

[a] LGC Limited, Queens Road, Teddington, Middlesex TW11 0LY, UK
[b] TUBITAK Ulusal Metroloji Enstitüsü (TÜBİTAK UME), Turkey
[c] Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
[d] National Measurement Institute Australia, 105 Delhi Road, North Ryde, NSW 2113, Australia
[e] National Institute of Metrology, 18 Bei San Huan Dong Lu, Chaoyang District, Beijing 100013, PR China
[f] D.I. Mendeleev Institute for Metrology, St. Petersburg, Russia
[g] IsoForensics, Inc., 421 Wakara Way, Suite 100, Salt Lake City, UT 84108, USA
[h] The Netherlands Forensic Institute, PO Box 24044, 2490 AA, Den Haag, the Netherlands
[i] Food Forensics, Innovation Centre, Norwich Research Park, Norwich NR4 7GJ, UK
[j] Department of Food Quality and Nutrition, Research and Innovation Centre, Fondazione Edmund Mach (FEM), Via E. Mach, 1, 38010 S. Michele all'Adige, TN, Italy
[k] Analytica Laboratories, Ruakura Research Centre, 10 Bisley Road, Hamilton 3240, New Zealand
[l] Queensland Health Forensic and Scientific Services, P.O. Box 594, Archerfield, Queensland 4108, Australia

ARTICLE INFO

Keywords:
Isotope ratio
Inter-laboratory comparison
Performance metrics
Metrology

ABSTRACT

Forensic application of carbon isotope ratio measurements of honey and honey protein to investigate the degree of adulteration with high fructose corn syrup or other C_4 plant sugars is well established. These measurements must use methods that exhibit suitable performance criteria, particularly with regard to measurement uncertainty and traceability – low levels of adulteration can only be detected by methods that result in suitably small measurement uncertainties such that differences of 1‰ or less can be reliably detected. Inter-laboratory exercises are invaluable to assess the state-of-the art of measurement capabilities of laboratories necessary to achieve such performance criteria. National and designated metrology institutes from a number of countries recently participated in an inter-laboratory assessment (CCQM-K140) of stable carbon isotope ratio determination of bulk honey. The same sample material was distributed to a number of forensic isotope analysis laboratories that could not participate directly in the metrological comparison. The results from these studies have demonstrated that the majority of participants provided isotope delta values with acceptable performance metrics; that all participants ensured traceability of their results; and that where measurement uncertainties were reported; these were fit-for-purpose. A number of the forensic laboratories only reported precision rather than full estimates of measurement uncertainty and this was the major cause of the few instances of questionable performance metrics. Reporting of standard deviations in place of measurement uncertainties is common practice outside metrology institutes and the implications for interpretations of small differences in isotopic compositions are discussed. The results have also highlighted a number of considerations that are useful for organisers of similar inter-laboratory studies in the future.

* Corresponding author.
E-mail address: philip.dunn@lgcgroup.com (P.J.H. Dunn).
[1] Present address: Bezmialem Vakıf University, Faculty of Pharmacy, Department of Analytical Chemistry, 34093 Istanbul, Turkey.

https://doi.org/10.1016/j.scijus.2018.08.003
Received 19 February 2018; Received in revised form 26 July 2018; Accepted 10 August 2018

1. Introduction

The determination of stable carbon isotope ratios as isotope delta values ($\delta^{13}C$ values) has many applications within forensic sciences including source inference (e.g. [1]), detection of food adulteration and authenticity (e.g. [2]) and distinguishing materials that are otherwise chemically and physically identical (e.g. [3]). Forensic applications of isotope delta values require analyses that meet the needs of stakeholders such as the police, courts and regulators and it is therefore essential that measurement results are compatible regardless of which laboratory provides them. Carbon isotope ratio testing of honey is usually performed to detect the addition of high-fructose corn syrup or other C_4 plant sugars within C_3 plant-based honey. AOAC has a standard method for this test which depends on the determination of the carbon isotope ratio of the bulk honey as well as of the extracted honey protein to estimate the amount of adulteration using the following equation [4]:

$$\%\ C_4\ Sugar = \frac{\delta^{13}C_{Protein} - \delta^{13}C_{Honey}}{\delta^{13}C_{Protein} - (-9.7)} \times 100 \quad (1)$$

Most unadulterated honeys will yield ≤7% C_4 sugars [4] and as the $\delta^{13}C_{VPDB\text{-}LSVEC}$ difference between the honey and protein increases, so too does the calculated degree of adulteration. The application of this method requires adherence to the common method protocol, isotope ratio analyses that afford traceability to the internationally agreed carbon isotope delta scale to ensure compatibility of results and that provide adequate measurement uncertainty to detect low levels of adulteration.

Achieving compatibility of results between laboratories at the highest metrological level is one of the fundamental aims of the inter-laboratory comparisons (ILCs) organised by the Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology (CCQM). The Inorganic Analysis Working Group (IAWG) of the CCQM has previously organised a number of ILCs of isotope ratio measurements for the national and designated metrology institutes (NMIs and DIs) which participate in its activities. The aim was to demonstrate compatibility of their results, including investigating light element isotope delta values of methionine, strontium isotope ratios in wine and lead isotope ratios in bronze [5].

Under most circumstances metrology institutes aim to obtain measurement data traceable to the SI. However, the requirement in forensic and other applications to distinguish very small variations in isotope ratios of elements such as carbon is presently achievable only using isotope delta values. The IAWG participants were, therefore, granted an exception to use delta values with traceability to the reference materials that define the position of the zero point until a scale with SI-traceability can be established at the required level of uncertainty [6]. There were a number of incentives for conducting a new comparison on isotope delta values: firstly, the need for IAWG members to publicly demonstrate their measurement capabilities; secondly, to assess the compatibility of metrology institute results reported using an isotope delta scale; and finally it had been almost a decade since light element isotope ratio measurements were the subject of a CCQM comparison. Two metrology institutes, LGC in the UK and Scientific and the Technological Research Council of Turkey's National Metrology Institute (TÜBİTAK UME), therefore recently coordinated an ILC of the stable carbon isotope analysis of bulk honey (CCQM-K140) [7]. This CCQM comparison highlighted that there was good agreement in obtained $\delta^{13}C$ values between the participants despite the significant differences in methods applied, particularly in regard to the data reduction and measurement uncertainty estimation approaches [7].

CCQM comparisons demonstrating compatibility between participants underpin the measurement services offered by metrology institutes (e.g. production and certification of reference materials and provision of reference values). On the other hand, field laboratories applying isotope ratio analysis to forensic investigations/questions generally only provide a commercial analytical service and simply report analytical results to their stakeholders. As forensic laboratories address different needs and use different approaches, the comparison of data obtained by metrological institutes to the forensic stable isotope community is useful. The results of such comparisons demonstrate the extent to which metrological principles are adopted outside of metrology institutes and more importantly should encourage the forensic laboratories to appreciate the benefits of a metrological approach. There are also implications from measurement uncertainty for forensic application of carbon isotope ratio measurements of honey as well as isotope ratios in other materials.

In this work, we briefly review the results from the CCQM comparison study and compare them to those obtained by a number of forensic stable isotope laboratories contacted via the Forensic Isotope Ratio Mass Spectrometry (FIRMS) Network that analysed the same honey material. This comparison includes discussions of the relative merits of the application of various performance metrics, measurement and data handling methods employed by all participants, comparison of the reported results, discussion of the performance of the forensic laboratories and implications for forensic analysis of honey.

2. Materials and methods

2.1. Participation

The participants within the studies described in this work are those listed under the author affiliations. No indication is given as to which participants participated in which of the two aspects (the CCQM comparison and the FIRMS study) to preserve anonymity. Note that Analytica Laboratories reported two separate results due to having used two different measurement approaches and these two results are listed separately in results tables, figures and within the text (hence the total number of results is thirteen while there are only twelve institutions). Also, for administrative reasons, one of the CCQM laboratories (CCQM Laboratory A) participated in a separate study conducted in parallel with CCQM-K140 and its data were not included in the CCQM-K140 report (but are described herein). Studies were coordinated by LGC and TÜBİTAK UME.

2.2. Honey material

Each participant within these studies (CCQM and FIRMS) received at least one amber borosilicate vial containing 2 g of honey. This was the same material that is now commercially available from the TÜBİTAK UME as “UME CRM 1312 – Honey (Unadulterated).” Prior to use in these studies, TÜBİTAK UME had investigated the homogeneity of the units of honey as well as performing experiments to determine the short and long-term stability of the honey material. The results of these preliminary studies can be found in the published final report for the CCQM study [7] as well as on the certificate for UME CRM 1312 [8].

2.3. Participants' methods

For CCQM participants, other than CCQM Laboratory A, full details of participant sample handling, instrumental and data handling methods can be found in the published final report [7]; however a brief summary is presented in Table 1. CCQM Laboratory A and participants within the FIRMS parallel study were also free to use whatever analytical measurement method they chose, with the proviso that full method details including source of traceability to the VPDB-LSVEC scale were reported. This differs from the AOAC method where the protocol is specified. Participants were required to report the $\delta^{13}C_{VPDB\text{-}LSVEC}$ of the honey together with an indication of measurement uncertainty. For the FIRMS laboratories where measurement uncertainty estimation might not be a common practice, the standard deviation of

Table 1
Sample preparation, instrumental and data handling information for the twelve laboratories participating in this study (one participant used two different techniques and therefore reported two results).

Laboratory	Principle	Mass of honey analysed	Corrections						Normalisation		Quality control material(s)
		μg	^{17}O	Blank	Linearity	Drift	Memory	Other	Type	RMs Used	
FIRMS 1	EA-IRMS	120 ± 50	SSH	n	y (Sharp)	n	n	n	Two point	IAEA-CH-7 USGS24	In-house tape
FIRMS 2	EA-IRMS	1100 to 1200	SSH	n (blank below detection limit)	n (mass controlled)	y (using QC material)	n	n	Two point	USGS40 USGS41	Cane sugar
FIRMS 3	CM-CRDS	1100 to 1200	n/a	n (blank below detection limit)	n (mass controlled)	y (using QC material)	n	n	Two point	USGS40 USGS41	Cane sugar
FIRMS 4	EA-IRMS	1200 ± 120	SSH	n (none observed)	n (mass controlled)	n (none observed)	n (none observed)	n	Two point	USGS40 USGS41	In-house glutamic acid
FIRMS 5	EA-IRMS	500 to 1000	Craig	n	y (no detail)	n	n	n	Multiple point	NBS-22 IAEA-CH-6 LSVEC	None reported
FIRMS 6	EA-IRMS	1500 to 2500	SSH	n	y (no detail)	n	n	n	Two point	Two in-house glucoses	IAEA-CH-7
FIRMS 7	EA-IRMS	3000 ± 300	Craig	y (automatic)	n (none observed)	y (spl/std. bracketing)	n	n	Two point	In-house flour and sucrose	NBS-22 IAEA-CH-6
CCQM 1	EA-IRMS	103 ± 66[a]	IUPAC	y (mass balance)	n (mass controlled)	n (none observed)	n (none observed)	n	Multiple point	NBS-22 USGS40 IAEA-CH-6 USGS41	FIRMS Olive Oil FIRMS Chitin In-house Glycine USGS 42
CCQM 2	EA-IRMS	200 ± 10	SSH	y (mass balance)	n (mass controlled)	n (none observed)	n (none observed)	n	Two point	IAEA-CH-6 NBS-22	None reported
CCQM 3	EA-IRMS	2000 to 2500	Craig	n	n	y (using QC material)	n	n	Two point	IAEA-CH-6 IAEA-CH-7	None reported
CCQM 4	EA-IRMS	900 ± 100	SSH	n (none observed)	n (mass controlled)	y (spl/std. bracketing)	n (avoided)[c]	n	Two point	NBS-22 USGS24	USGS40
CCQM 5	EA-IRMS	600 to 1000	IUPAC	y (mass balance)	n (none observed)	y (using QC material)	n (none observed)	n	Multiple point	IAEA-CH-7 IAEA-CH-6 USGS40 IAEA-600	Two in-house honey materials
CCQM A	CM-CRDS	90,000 ± 9000[b]	n/a	n	n	y	n	n	Two point	IAEA-CO-8 IAEA-CH-7	None reported

[a] Standard deviation of masses analysed.
[b] Custom offline combustion used significantly larger amount of honey than the online processes used by other participants.
[c] Carryover from graphite into subsequent analyses was observed; however the affected analyses were discarded rather than a memory correction being applied.

the independent replicates was requested as an indication of method precision if the participants were inexperienced with estimation of measurement uncertainty. A brief description of the instrumental and data handling methods applied by these participants can also be found within Table 1.

2.4. Performance metrics

Within the CCQM study, the reference value was determined according to the method described within the published final report, i.e. the arithmetic mean of participant results (excluding CCQM Laboratory A), which was $\delta^{13}C_{VPDB\text{-}LSVEC} = -24.095 \pm 0.107‰$ (expanded uncertainty, $k = 2.776$) [7]. CCQM inter-comparison studies also report a degree of equivalence (DoE) which is a measure of metrological compatibility and it is expressed quantitatively by two terms: the deviation from the reference value and the uncertainty of this deviation (at a 95% level of confidence). The method used for calculation of DoEs will depend on the statistical method used to determine the reference value [9] and for CCQM-K140 details can be found in the final report [7].

Z-scores for participants were calculated as follows:

$$z = \frac{(x - ref)}{U_{ref}} \quad (2)$$

where x is the value reported by the participant in this study and *ref* and its associated expanded uncertainty, U_{ref}, (with $k = 2.776$) were the reference value taken from the CCQM study [7].

The participant results and reference values were also used to calculate ζ-scores which provide a means to examine the plausibility of participants' measurement uncertainty estimate. The ζ-scores were calculated as follows:

$$\zeta = \frac{(x - ref)}{\sqrt{(u_x)^2 + (u_{ref})^2}} \quad (3)$$

where x is the value reported by the participant, u_x is the standard uncertainty in the value as reported by the participant (or the reported standard deviation where the measurement uncertainty was not reported) and *ref* and its associated standard uncertainty, u_{ref}, (with $k = 1$) were the reference value taken from the CCQM study [7].

E_n numbers for the participant results were also calculated. These are very similar to the ζ-scores but use expanded rather than standard uncertainties for both the participant result and reference value:

$$E_n = \frac{(x - ref)}{\sqrt{(U_x)^2 + (U_{ref})^2}} \quad (4)$$

where x is the value reported by the participant, Ux is the expanded uncertainty in the value as reported by the participant (here a k factor of 2 has been applied to the standard deviations reported in place of measurement uncertainties by some participants) and ref and its associated expanded uncertainty, Uref, (with k = 2.776) were the reference value taken from the CCQM study [7].

Unless specifically stated, all measurement uncertainties within this manuscript are expanded uncertainties (U). These are standard uncertainties (u) multiplied by a k-factor (usually $k = 2$ and so $U = 2u$) so that the resulting uncertainty range has a 95% confidence level of including the true value.

3. Results

3.1. Instrumental and data handling methods

3.1.1. Instrumental approach

The majority of participants across all three studies employed elemental analyser-isotope ratio mass spectrometry (EA-IRMS) to determine the $\delta^{13}C_{VPDB\text{-}LSVEC}$ value of the bulk honey. Only two participants, FIRMS Laboratory 3 and CCQM Laboratory A, did not use EA-IRMS – both used a combustion module either directly coupled to cavity ring down spectroscopy instrumentation (CRDS, a form of isotope ratio infrared spectroscopy) or using an offline sample preparation and transfer.

3.1.2. ^{17}O correction

For EA-IRMS measurements of carbon isotope ratios of CO_2 gas, the isotopic composition of the oxygen must be taken into account by application of a so-called ^{17}O correction. These corrections use one of three different sets of algorithms and are typically carried out automatically within instrumental software packages but can also be applied offline. The Craig correction [10] is the simplest to apply but makes assumptions regarding oxygen isotopic fractionation which are not supported by experimental evidence. CCQM Laboratory 3 and FIRMS Laboratories 5 and 7 used the Craig correction approach for ^{17}O. The improved algorithm suggested by Santrock, Studley and Hayes (SSH) [11] is available within some IRMS software packages. This is considered an exact approach but requires an iteration procedure to determine the $^{18}O/^{16}O$ ratio in the sample CO_2. Six participants, CCQM Laboratories 2 and 4 as well as FIRMS Laboratories 1, 2, 4 and 6, applied the SSH algorithm to correct for ^{17}O. The final approach, which is endorsed by the Commission on Isotopic Abundances and Atomic Weights (CIAAW) of the International Union of Pure and Applied Chemistry (IUPAC), is a linear approximation rather than an exact solution and also uses more up-to-date values for the absolute isotope ratios of VPDB than the SSH approach [12]. Only two participants, CCQM Laboratories 1 and 5 used the CIAAW-recommended approach.

Provided that raw $\delta^{13}C$ values of samples and the reference materials (RMs) used for scale normalisation are measured against the same working reference, and that the same ^{17}O correction is applied to all materials within an analytical sequence, then the bias in scale-calibrated $\delta^{13}C_{VPDB\text{-}LSVEC}$ values introduced via the choice of ^{17}O correction approach will be < 0.001‰.

3.1.3. Blank correction

Three of the CCQM participants (CCQM Laboratories 1, 2 and 5) carried out a blank correction using the mass balance approach described in the FIRMS Good Practice Guide for IRMS [13]. The other two did not report the need to carry out a correction for the blank contribution. CCQM Laboratory A did not carry out a blank correction as a result of using an offline combustion process and a significantly larger amount of honey material. Only one FIRMS laboratory, FIRMS Laboratory 7, found a blank level that required correction and therefore carried out a blank correction of their raw δ values using an automated procedure within the instrumental software. The remaining FIRMS laboratories found negligible blank contributions and therefore deemed a blank correction unnecessary.

The effect of blank correction on measured isotope delta values will depend on the difference in peak size between sample and blank signals and also upon the difference in isotopic composition. Where the blank is significantly smaller than the same signal (i.e. by a factor of 100 or more), then the difference between the measured isotope delta value before and after application of a blank correction will be minimal, as too will be the contribution of blank to the measurement uncertainty.

3.1.4. Linearity correction

All participating laboratories controlled the mass of honey analysed to some degree, with some having very low tolerances while others having much wider ranges of acceptable weight (Table 1). Within the CCQM study, no participants reported application of a correction to measured data to account for variation in sample mass analysed. FIRMS Laboratories 2, 3, 4 and 7 deemed a linearity correction was unnecessary within the ranges of mass of honey analysed. FIRMS Laboratory 1 applied a linearity correction as described by Sharp [14] while FIRMS Laboratories 5 and 6 applied a linearity correction based on the change in delta value of repeated injections of working gas with

Table 2
Reported $\delta^{13}C_{VPDB\text{-}LSVEC}$ values, standard uncertainties or standard deviations, expanded uncertainties and performance metrics for each participating laboratory together with the reference value from CCQM-K140 [7]. Questionable performance metric scores are highlighted in bold text; there were no unsatisfactory performance metric scores in these studies.

Laboratory	n	δ13CVPDB-LSVEC	u	U (k = 2)	z-score	ζ-score	En Number
		‰	‰	‰			
FIRMS 1	5	−24.25	0.03*		−1.45	**−3.17**	**−1.26**
FIRMS 2	5	−24.137	0.038*		−0.39	−0.78	−0.32
FIRMS 3	5	−24.117	0.153*		−0.21	−0.14	−0.07
FIRMS 4	5	−24.089	0.021*		0.06	0.14	0.05
FIRMS 5	11	−24.08	0.15	0.30	0.14	0.10	0.05
FIRMS 6	5	−23.93	0.2*		1.54	0.81	0.40
FIRMS 7	7	−23.88	0.08*		**2.01**	**2.42**	**1.12**
CCQM 1		−24.20	0.09	0.18	−0.98	−1.07	−0.50
CCQM 2		−24.146	0.139	0.278	−0.48	−0.35	−0.17
CCQM 3		−24.09	0.05	0.11	−0.05	0.08	0.03
CCQM 4		−24.03	0.05	0.11	0.61	1.03	0.42
CCQM 5		−23.993	0.042	0.084	0.95	1.79	0.75
CCQM A		−24.20	0.45	0.90	−0.98	−0.23	−0.12
Reference value		−24.095 ± 0.107‰ (expanded uncertainty, k = 2)					

* This is a standard deviation (Section 4.3).

varying intensity.

IRMS instrument manufacturers typically specify control limits for change in isotope delta with sample signal. These might be for example 0.05‰ nA^{-1} and therefore if variation in signal amplitude is > 1 nA, there will be > 0.05‰ variability in measured isotope delta value resulting from the linearity effect, which will be reflected in the standard deviation of replicate analyses. The degree to which a correction for linearity contributes to measurement uncertainty will depend on the magnitude of the effect being corrected as well as the mathematical approach employed.

3.1.5. Drift correction

CCQM Laboratories 1 and 2 as well as FIRMS Laboratories 1, 4, 5 and 6 did not find any evidence of drift within their analytical sequences and therefore did not apply a drift correction. CCQM Laboratories 3 and 5 as well as FIRMS Laboratories 2 and 3 applied a drift correction based upon the analysis of QC materials dispersed throughout their analytical sequences while CCQM Laboratories 4 and 6 and FIRMS Laboratory 7 employed a sample-standard bracketing technique within each sequence to account for instrumental drift.

The use of QC materials to correct for drift within a sequence of analyses will result in additional contributions to measurement uncertainty for each sample result from the QC results. If the drift correction also shifts results such that the QC result post-correction matches the expected result (e.g. after [15]), then there will also be an uncertainty contribution from the expected value for the QC material. The magnitude of these contributions will depend on the extent of observed drift and the correction algorithm applied.

3.1.6. Normalisation

All participants applied normalisation using at least two RMs as recommended by the CIAAW [16] thereby providing adequate traceability to the reporting scale. While all participants within the CCQM study used secondary reference materials directly for normalisation, this was only the case with five of the seven FIRMS laboratories. These secondary materials included USGS24, USGS40, USGS41, NBS 22, IAEA-CH-6, LSVEC, IAEA-CH-7, IAEA-600 and IAEA-CO-8 and the assigned values were taken from the relevant IUPAC Technical Report [17]. The remaining participants (FIRMS Laboratories 6 and 7) used in-house RMs that had been previously calibrated against primary and/or secondary reference materials, resulting in a slightly longer traceability chain. Both of these laboratories did include secondary reference materials amongst their quality control materials. Only one participant across both studies used exactly matrix-matched QC materials, i.e. honey (CCQM Laboratory 5).

The contribution to measurement uncertainty resulting from normalisation of results to the reporting scale will include not only the uncertainty of the known values of the RMs (for those listed above the standard uncertainties are 0.05‰ or less), but also the uncertainty related to the analysis of those RMs as well as the sample. The choice of RMs, both in terms of calibration range that they afford but also in their associated measurement uncertainty, will impact the uncertainty in normalized isotope delta values obtained. These issues have been discussed in previous publications [e.g. 18–21]. It is also important to consider the nature of the RMs used for normalisation of results. When analysing organic materials it is preferable to use organic RMs, however provided that quantitative conversion of the RM to the analyte gas can be demonstrated and that the same is also true for samples analysed using the same instrumental method, then any RM can be used. For many materials, proving that conversion to the analyte gas is complete is difficult and therefore the use of matrix-matched RMs is critical.

3.1.7. Estimation of measurement uncertainty

Details of the approaches used by CCQM Laboratories 1–5 to estimate their measurement uncertainties can be found in the final report for CCQM-K140 [7]. CCQM Laboratory A combined the uncertainty components arising from measurements (type A), from sample preparation (type B), from the expected $\delta^{13}C_{VPDB\text{-}LSVEC}$ value of one RM and from their normalisation approach. CCQM Laboratory A was relatively lacking in experience with the instrumentation used and hence why they participated in a parallel study (results from which were excluded during calculation of the reference value) and it is unsurprising that their estimation of measurement uncertainty turned out to be very conservative. FIRMS Laboratory 5 combined an estimate of within-laboratory reproducibility obtained from QC data with the uncertainty component arising from method and laboratory bias, which was estimated from proficiency testing (PT) data. Both of these participants combined the individual uncertainty components using the square root sum of squares approach.

Other FIRMS participants explicitly reported standard deviations of replicate analyses rather than measurement uncertainties.

3.2. Participant results and performance metrics

The $\delta^{13}C_{VPDB\text{-}LSVEC}$ values reported by the participants can be found within Table 2 and Fig. 1, together with the reference value from the CCQM study for comparison. The calculated performance metrics can also be found in Table 2. All participants within the FIRMS parallel

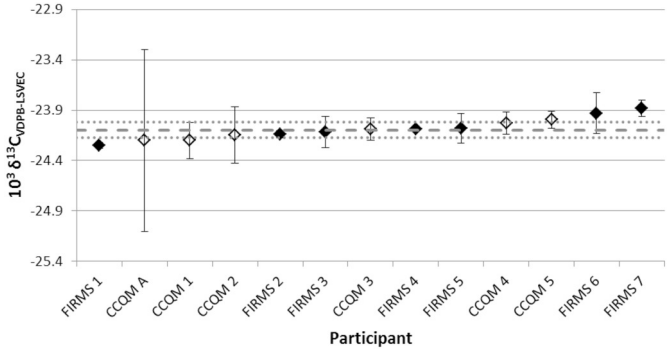

Fig. 1. Reported $\delta^{13}C_{VPDB\text{-}LSVEC}$ values for the honey for all participants within the CCQM study (open diamonds) and FIRMS parallel study (filled diamonds). The error bars represent the reported standard deviations of replicate analyses for the majority of FIRMS laboratories, but expanded uncertainties ($k = 2$) for FIRMS Laboratory 5 and the CCQM laboratories. The dashed and dotted grey lines are the reference value plus or minus its expanded uncertainty ($k = 2$) [7].

study reported the standard deviation of replicate analyses as an indication of measurement uncertainty with the exception of FIRMS Laboratory 5, which estimated a standard uncertainty (Section 3.1.7).

Z-scores within the range $-2 < z < +2$ are satisfactory, within the ranges of $-3 < z < -2$ and $+2 < z < +3$ are questionable while those within the ranges $z < -3$ and $z > +3$ are unsatisfactory. Only one participant, FIRMS Laboratory 7, produced a questionable z-score (+2.01). No participants within the three studies reported results that were unsatisfactory (Table 2). As with z-scores, ζ-scores within the range $-2 < \zeta < +2$ are satisfactory, within the ranges of $-3 < \zeta < -2$ and $+2 < \zeta < +3$ are questionable while those within the ranges $\zeta < -3$ and $\zeta > +3$ are unsatisfactory. If the magnitude of the E_n number, $|E_n|$, is < 1, then the result is satisfactory, where $|E_n|$ is > 1, it is an indication of unsatisfactory performance. Only two participant results produced ζ-scores and E_n numbers that were questionable (FIRMS Laboratories 1 and 7, Table 2).

4. Discussion

4.1. Assessment of performance metrics

It is z-scores that are most commonly applied to provide information of performance during inter-laboratory studies. Indeed, the two commercial PT schemes available for light element isotope ratio analysis both rely on z-scores for the assessment of participant performance. These PT schemes are provided by LGC which organises an isotope ratio PT scheme in collaboration with the FIRMS Network that is accredited to ISO/IEC 17043:2010 [22] and by Eurofins Scientific which organises the Food analysis using Isotopic Techniques – Proficiency Testing Scheme (FIT-PTS) – focussed on official methods of analysis such as those from AOAC [23].

While z-scores are useful performance metrics, they do not take into account the measurement uncertainty reported by participating laboratories. It is therefore possible to report a value that results in a z-score outside of the satisfactory range of $-2 < z < +2$ but for which the measurement uncertainty encompasses part of - if not all of - the reference value range with 95% confidence. This is more likely when the z-score is calculated on the basis of a fixed uncertainty in the reference value than when a parameter such as the robust standard deviation between participants, which may vary between different rounds of the same PT scheme, is used. The use of ζ-scores and E_n numbers, which take into account the uncertainties reported by participants, are therefore beneficial in terms of performance assessment.

Both ζ-scores and E_n numbers increase as either the deviation from the reference value increases or as the uncertainty reported by the participant decreases; therefore larger ζ-scores and E_n numbers can indicate a large bias, an underestimated uncertainty, or a combination of both [24]. These metrics therefore provide avenues for further investigation should poor results be obtained by individual participants. This is not the case for z-scores, where no indication as to the reason for poor performance is given.

While there was one laboratory with questionable z-score (FIRMS Laboratory 7), there were two (FIRMS Laboratories 1 and 7) with questionable ζ-scores and unacceptable E_n number. Naturally, whichever performance metrics are employed to investigate performance within PT schemes or other ILC exercises, there will be subtly different results, particularly if there are laboratories whose results fall close to the thresholds of acceptability for particular metrics.

4.2. CCQM study overview

The published report for the CCQM study [7] contains a detailed discussion of the participant results and therefore only a brief summary is included here. The most significant findings were the lack of agreement in data handling including corrections applied to raw data, specific RMs used for normalisation of results and measurement uncertainty estimation approach. Nevertheless good agreement was observed between all participants with none reporting results leading to questionable performance metrics. Given that the reference value is derived from the CCQM participants (excluding CCQM Laboratory A), this is not altogether surprising.

Measurement uncertainty budgets reported by the metrology institutes were particularly interesting given the large differences in contributions to uncertainty resulting from some parameters. For example, the contribution of uncertainty in the assigned values of the RMs used during scale realisation ranged from approximately 10% to just over 60%. This lack of agreement is perhaps the result of the difficulty in estimating the measurement uncertainty for isotope delta values where correlation between the various input parameters can be difficult to account for coupled to the variety of approaches available for uncertainty estimation. The expanded uncertainties (with $k = 2$) reported by the CCQM-K140 participants were large in comparison to the overall standard deviation of the reported values. The reference value was determined as the arithmetic mean with the value: $\delta^{13}C_{VPDB\text{-}LSVEC} = -24.095 \pm 0.107$‰ (expanded uncertainty, $k = 2.776$, [7]).

4.3. Comparison of approaches between CCQM-K140 and FIRMS participants

There was little difference in instrumental approach between the metrology institutes and the FIRMS laboratories, with EA-IRMS being employed by the majority and two (or more) point scale realisation used by all participants. There were some differences in the corrections applied to raw data, for example only metrology institutes used the CIAAW-recommended ^{17}O correction, while only FIRMS laboratories employed a linearity correction to account for differences in sample mass between replicates. While only CCQM Laboratory 5 used exactly matrix-matched QC materials (in-house honey). The commercial availability of the test material used in these studies, UME CRM 1312 – Honey (Unadulterated) as well as UME CRM 1313 Honey (Adulterated) that was not investigated in these studies, should improve this situation in the future. Exactly matrix matched QC materials facilitate the monitoring of data handling procedures such as corrections applied to raw data and also provide evidence of complete conversion of honey samples to the analyte gas during online or offline combustion. The major difference was in the reporting of measurement uncertainty, with all but one FIRMS laboratory providing standard deviations of replicate analyses (i.e. a precision estimate) rather than an estimate of measurement uncertainty.

FIRMS Laboratory 5 was the only FIRMS participant to estimate a measurement uncertainty (0.30‰ with $k = 2$), which was larger than the uncertainties reported by CCQM Laboratories 1–5 although it was

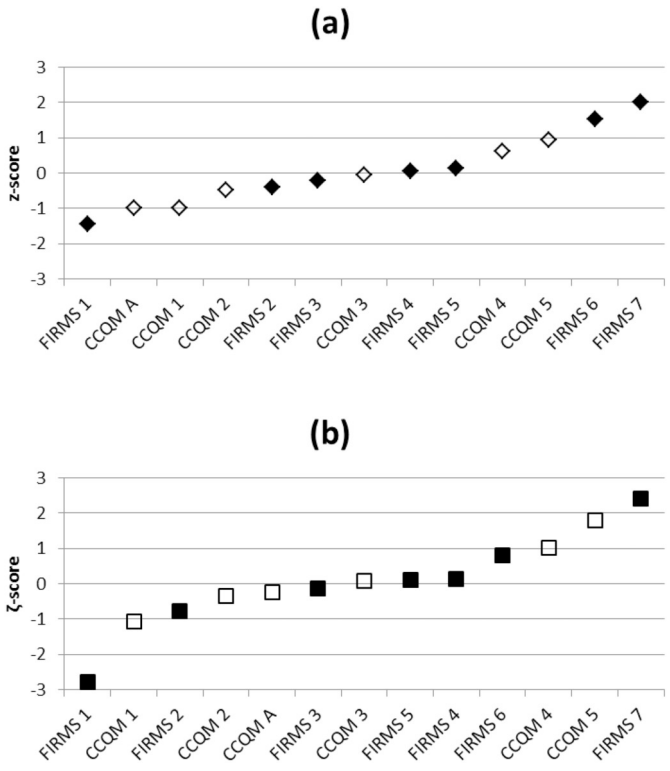

Fig. 2. Performance metrics for all laboratories: (a) z-scores for participants within the CCQM comparison (open diamonds) and the FIRMS parallel study (filled diamonds); (b) ζ-scores for participants within CCQM (open squares) and the FIRMS parallel study (filled squares).

smaller than the uncertainty reported by CCQM Laboratory A. This may be a reflection of the approach used by the FIRMS laboratory (assessing bias from PT results) which is different from the metrological approach. On the other hand, there was no single approach to estimating measurement uncertainty for carbon isotope delta values applied by the CCQM participants, for example, which parameters to include in the uncertainty budget and the relative contributions of these parameters to the uncertainty budget.

In contrast, of the standard deviations (from $n = 5$ to 7 independent replicates, Table 2) reported by FIRMS participants, those of FIRMS Laboratories 1, 2, 4 and 7 were all smaller than the lowest uncertainty reported in the CCQM study (0.084‰), while the standard deviations of FIRMS Laboratories 3 and 6 were within the range of uncertainties reported during the CCQM comparison (note that FIRMS Laboratory 3 was reporting the outcomes of CRDS measurements, which are known to produce approximately three times higher standard deviation than the IRMS-based methods [25]).

Despite these differences, comparison of the results submitted by the participants in the FIRMS study with those from CCQM laboratories shows that there is mostly good agreement between all reported values (Fig. 2), very few instances of questionable performance metrics and no unsatisfactory results (Table 2). The fitness-for-purpose of the reported results in terms of forensic honey analysis is discussed below in section 4.5.

4.4. Discussions of poor performance

Although the normalisation procedures and quality control materials employed by the laboratories that produced one or more questionable performance metrics (FIRMS Laboratories 1 and 7) were very similar to those that did not, there are several factors that might explain why poor performance was observed. Firstly, FIRMS Laboratory 1 used amongst the smallest amount material for analysis (120 ± 50 µg). Secondly, of the two participants that did not directly use secondary reference materials for normalisation (FIRMS Laboratories 6 and 7, Table 1) one produced questionable performance metrics. In addition, the reporting of precision estimates rather than measurement uncertainties by FIRMS participants is likely to be another explanation of the poor performance, particularly for the performance metrics where reported uncertainties are considered. Finally, there is the possibility that the reported results are biased. These factors are discussed in more detail within the following sections.

4.4.1. Sample mass and the "linearity" effect

Sample mass used for analysis certainly has the potential to contribute to variance between reported $\delta^{13}C_{VPDB\text{-}LSVEC}$ values. The so-called "linearity effect" for EA-IRMS analyses is well known and occurs when different isotope ratios for the same material are obtained depending on the mass of sample analysed. This can occur even when the material in question is isotopically homogeneous and could result in poor repeatability of results should there be significant differences in mass of the same analyte. Manifestation of the linearity effect can be avoided by constraining the sample mass range for replicate analyses of the same material or be corrected for via determination of magnitude of the effect using a working reference or quality control material analysed at different amount levels. Both of these approaches were applied by participants within these studies (section 4.1.4).

Fig. 3(a) shows that the CCQM participants' results do not appear to vary with mass of honey analysed. The regression line of best fit accounting for error in both variables by using Williamson's method [26,27] has a gradient of 0.040‰ mg^{-1} with a standard error of 0.052‰ mg^{-1} and is therefore indistinguishable from zero. The FIRMS results presented in Fig. 3(b) however show a larger residual bias with a slope of 0.129‰ mg^{-1} (standard error 0.020‰ mg^{-1}). It is difficult to untangle the effect of smaller ranges in the participant reported delta

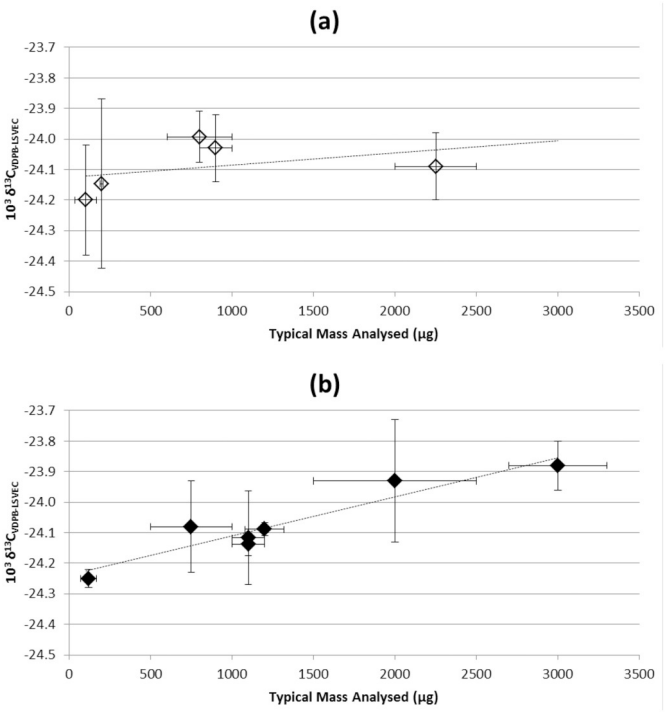

Fig. 3. Relationship between typical mass of honey analysed and reported $\delta^{13}C_{VPDB\text{-}LSVEC}$ value for the participants using EA-IRMS: (a) results for CCQM Laboratories 1–5 (open diamonds); (b) results for the FIRMS laboratories (filled diamonds). Error bars in both plots represent the reported expanded uncertainty (or standard deviation) for the $\delta^{13}C_{VPDB\text{-}LSVEC}$ values and the reported ranges of sample mass. The dashed lines are regression lines of best fit accounting for error in both variables by using Williamson's method [26,27].

values (due to reporting precision rather than measurement uncertainty) from any underlying residual bias from the range of sample mass using the data obtained in this study. Whether the linearity effect has contributed to the poor performance metrics for FIRMS Laboratory 1 is therefore difficult to establish.

Separate from the linearity effect, the degree of homogeneity of the sample material also has the potential to impact the accuracy of isotope ratio measurement results at small amount levels. Therefore, prior to the commencement of the CCQM and FIRMS studies, ten of the units of honey were tested for homogeneity by TÜBİTAK UME. The honey material isotopic composition was found not to be significantly different between vials following ANOVA analysis of the analytical results [7,8]. The honey material was therefore considered to be homogenous in terms of carbon isotopic composition – but only down to the amount level used during the homogeneity study (200–230 μg). Two participants within the CCQM and FIRMS studies, FIRMS Laboratory 1, who returned results with questionable performance metrics, and CCQM Laboratory 1, who did not, used amounts of honey smaller than this. The results from these participants do not indicate that the honey material was the source of the poor performance of FIRMS Laboratory 1. For future inter-laboratory exercises of this type a minimum sample amount to use for each analysis should be recommended based upon the amount used during the homogeneity investigation. This is an aspect that is covered during the production of a reference material under ISO 17034:2016 (previously ISO Guide 34:2009) accreditation [28,29] and a minimum sample amount of 0.2 mg is indeed specified in the TÜBİTAK UME certificate for the honey material [8].

Honey is a liquid and therefore homogeneity in terms of carbon isotope ratio for a bulk sample might be expected. Nonetheless, it is also possible that the viscous nature of honey could result in within-vial variation in carbon isotope ratio, or that some precipitation or crystallisation may go unnoticed during sample storage, preparation and/or analyses. Storage temperature for the honey material was not specified during the study, although it was recommended that vials be kept at "room temperature." No participants reported using a particular storage method for the vials for example at elevated temperature to avoid crystallisation; however no participants reported crystallisation occurring in their honey. The subsequent certificate for the honey reference material provides (+20 ± 5) °C as the storage temperature [8]. Only one of the participants reported using a homogenisation procedure prior to the transfer of honey into tin capsules for analysis (CCQM Laboratory 5). This involved heating, vortex mixing and ultrasonication – a procedure which that laboratory applies for all honey carbon isotope analyses. It is unlikely that the heating resulted in loss of volatile species from the honey as this might be expected to result in lower $\delta^{13}C$ values via preferential loss of ^{13}C-depleted species, while CCQM Laboratory 5 reported a $\delta^{13}C_{VPDB\text{-}LSVEC}$ value that was one of the most ^{13}C-depleted.

4.4.2. Reported measurement uncertainties

Of the corrections performed on the measured isotope ratios (Table 1) it is normalisation and the ^{17}O correction that are essential for every EA-IRMS analysis (and only the former for optical isotope ratio methods); the others - such as blank, drift, linearity and memory corrections - depend on whether there is an analytical need to perform them as indicated by QC materials. While each of these corrections will contribute to measurement uncertainty to some degree (e.g. [21,30]), they will not be reflected in a precision estimate exemplified by the standard deviation of replicate analyses reported by the majority of the FIRMS participants. An estimate of measurement uncertainty that takes into account the effect of normalisation and other corrections performed on raw instrumental data will therefore be larger than simply the precision alone. Furthermore, precision in isotope delta values is sometimes misleadingly reported giving the impression that it is indeed a measurement uncertainty - although in these studies the FIRMS laboratories concerned were all explicitly reporting a standard deviation. Reporting measurement uncertainties that are simply standard deviations can lead to unwarranted confidence in very small variations of isotope delta values and, as stated earlier, an aim of collaboration between the forensic and metrology communities is to change this attitude.

For FIRMS Laboratory 1, if the reported standard deviation (or indeed a reported measurement uncertainty) had been 0.07‰ or larger, the resulting ζ-score and E_n number would have been −1.94 and − 0.88, respectively and fall within the acceptable performance ranges for each metric. Likewise, for FIRMS Laboratory 7, if their reported standard deviation had been larger than 0.1‰, their ζ-score and E_n number would have been +1.99 and + 0.94, respectively and also fall within the acceptable range. These are relatively small increases from the reported standard deviations of approximately 0.03‰ in both cases and therefore the question arises whether simply including the essential data handling processes (^{17}O correction and normalisation) within an uncertainty budget would result in sufficient increase over the reported precision to yield acceptable performance metrics and to provide a more realistic measurement uncertainty that can give a better indication of the significance of small differences in isotopic composition between two materials.

It is unlikely that the ^{17}O correction has a large contribution to measurement uncertainty (Section 3.1.2); however the effect of normalisation is significant and has been previously studied [13,18–21]. As noted in Section 3.1.6, the selection of RMs used for normalisation will impact the achievable uncertainty not only through the Type B contribution from the assigned values (although in some instances such as NBS 19 this is defined as zero), but also a Type A contribution resulting from the uncertainty in their measurement [13]. The Type B contribution to measurement uncertainty reported by CCQM-K140 participants comprised between 10 and 60% of the uncertainty budget depending on the method used to estimate measurement uncertainty. This highlights the potential danger of relying on a precision estimate which does not consider these Type B contributions as these can be significant within the uncertainty budget.

The additional uncertainties arising from (i) normalisation via extrapolation rather than interpolation; and (ii) use of inorganic carbon RMs to normalise organic carbon samples where complete conversion of carbon to the analyte CO_2 has not been demonstrated are also potentially important but not applicable for any of the results reported herein and are therefore not considered further. An extended traceability chain (such as the use of in-house RMs for normalisation that have been themselves characterised against commercially available RMs) will result in a larger measurement uncertainty associated with calibration of measured delta values to the reporting scale. The RMs used by the participants (listed in Section 3.1.6 and Table 1) have correlated assigned values due to being characterised during the same study against the same two reference materials (NBS 19 and LSVEC) [31]. Accounting for this correlation within an uncertainty budget is not straightforward and falls outside of the scope of this manuscript.

It is possible to model the effect of normalisation on measurement uncertainty of the reported isotope delta values from FIRMS Laboratories 1 and 7 provided that some assumptions are made. Firstly it is assumed that normalisation does not alter the raw isotope delta values for sample measurement (i.e. the measured isotope delta values for the RMs are exactly equal to their expected values – although the uncertainty associated with the measured value is different to the uncertainty in the assigned value). As a consequence, the standard deviation of their raw delta values for sample measurements are assumed to be equal to the standard deviations reported; indeed normalisation does not typically alter the reported standard deviation of results by a significant amount. It is also assumed that the replicate analyses of the RMs used for normalisation have the same standard deviation as replicate analyses of the sample material.

The supplementary spreadsheet S1 contains estimations of the measurement uncertainties for FIRMS Laboratories 1 and 7 as

determined by use of the Kratgten spreadsheet approach detailed in the Good Practice Guide for IRMS [13,21,32]. Estimations for FIRMS Laboratory 1 result in a combined standard uncertainty associated with normalisation of 0.05‰ in comparison to the reported standard deviation of 0.03‰. This is still slightly lower than the 0.07‰ required to result in acceptable performance metrics; however FIRMS Laboratory 1 also implemented a linearity correction (Table 1) in addition to normalisation, which is not accounted for. For FIRMS Laboratory 7, the use of in-house RMs with larger assigned uncertainties that take into account the calibrations throughout the traceability chain (expanded uncertainties at the 95% confidence level for their flour and sucrose reference materials were 0.13 and 0.10‰, respectively) resulted in a combined standard uncertainty associated with normalisation of 0.12‰ which would have led to satisfactory performance metrics.

That the differences between the reported precisions and required measurement uncertainty to achieve acceptable ζ-scores and E_n numbers for FIRMS Laboratories 1 and 7 can be accounted for simply by consideration of the uncertainty introduced by normalisation and other corrections to instrumental data highlights the need to report measurement uncertainties rather than simple precision estimates if data compatibility is to be ensured. Moreover, in both of the examples above, the normalisation Type A and B contributions account for over half of the estimated uncertainty. This was also the case for CCQM-K140 participants where some even reported normalisation accounting for over 90% of the measurement uncertainty budget [7]. It is clear that a precision estimate is therefore an overly optimistic measure of the uncertainty.

The best estimate of any result is that its true value lies somewhere within the range defined by its measurement uncertainty; limiting this range to one defined solely by precision data may imply that there is better knowledge of the true value than is justified by the data. This may in turn lead to differences in isotope ratio between two samples being interpreted as significant when this would not be the case if measurement uncertainty were considered, particularly for expanded uncertainties at the 95% confidence level.

4.4.3. Bias in reported values

For FIRMS Laboratories 1 and 7 it is also possible to determine by how much their reported values would need to change (assuming no change to the reported standard deviations) to result in satisfactory ζ-scores and E_n numbers. For FIRMS Laboratory 1 the reported $\delta^{13}C_{VPDB\text{-}LSVEC}$ value would need to be −24.19‰ or higher for an acceptable ζ-score and higher than −24.21‰ for an acceptable E_n number. For FIRMS Laboratory 7 the reported $\delta^{13}C_{VPDB\text{-}LSVEC}$ value would need to be lower than −23.92‰ for an acceptable ζ-score and lower than −23.91‰ for an acceptable E_n number. These are shifts of < 0.05‰ which are small and there is also no indication that the reported results from FIRMS Laboratories 1 and 7 are outliers or not part of a normal distribution of $\delta^{13}C_{VPDB\text{-}LSVEC}$ values. It is therefore unlikely that the poor performance metrics from these participants is the result of reporting biased $\delta^{13}C_{VPDB\text{-}LSVEC}$ values.

4.5. Implications for forensic application of isotope analysis

For an unadulterated honey, $\delta^{13}C_{Honey}$ is generally in the range of −28 to −21‰ while adulteration results in $\delta^{13}C_{Honey}$ values in the range of −21 to −15‰ [33]. Although it is technically possible to observe $\delta^{13}C_{Honey}$ values in the range of −15 to −10‰ these are almost certainly fake “honey” and will consist almost exclusively of sweeteners such as high fructose corn syrup. The $\delta^{13}C_{Protein}$ of an unadulterated honey might be up to 2‰ different to the bulk honey value as a result of bees collecting pollen from different plants to the nectar source of the honey [34]. The proportion of protein within honey is typically 0.1 to 0.2%, although some honeys can contain nearly 1% protein [35].

These indicative values for $\delta^{13}C_{Honey}$ and $\delta^{13}C_{Protein}$ allow a sensitivity analysis of how these terms influence the resultant adulteration percentage, which can be found in supplementary spreadsheet S2. Typically there will be between 5 and 10% increase in reported adulteration level for each permil increase in the numerator of Eq. (1) – this supports the AOAC method whereby 7% adulteration, equivalent to a difference of 1‰, is the lower threshold. Note that this sensitivity analysis simulates how the percentage adulteration might be expected to vary between different honey samples, rather than showing the effect of the measurement uncertainty in the input terms of Eq. (1). The difference in reported $\delta^{13}C_{VPDB\text{-}LSVEC}$ value between the highest (FIRMS Laboratory 7) and lowest (FIRMS Laboratory 1) result was 0.37‰. This equates to a change in reported adulteration of between 2.0 and 3.6%.

The uncertainty in each of the input terms of Eq. (1) will influence the uncertainty in the adulteration percentage obtained. Although the uncertainty in the reference value from CCQM-K140 was 0.107‰ (k = 2.776), we have selected the mean uncertainty of CCQM-K140 participants (0.15‰ with k = 2) as representative of the uncertainty that can be achieved for the $\delta^{13}C_{Honey}$ term. Extraction of the relatively small amount of protein from honey prior to isotope ratio analysis adds additional sample preparation stages that are not required for bulk honey analyses. It is therefore likely that measurement uncertainties associated with the protein carbon isotope ratio measurement are larger than for the bulk honey and consequently the inter-laboratory variability will also likely be larger. We have therefore made the conservative assumption that the spread of results for determination of $\delta^{13}C_{Protein}$ values is double the spread for determination of $\delta^{13}C_{Honey}$ values (i.e. 0.30‰, with k = 2) and used this as an indication of the measurement uncertainty for $\delta^{13}C_{Protein}$.

The two measured terms in Eq. (1) are correlated. For the same honey sample, the $\delta^{13}C_{Honey}$ term includes a contribution from the $\delta^{13}C_{Protein}$ term, although the latter will need to change significantly before the former is affected due to the relatively small amount of protein within bulk honey. For an unadulterated honey the $\delta^{13}C_{Honey}$ term will be reflected in the $\delta^{13}C_{Protein,}$ and this will be a positive correlation, but on the other hand, for an adulterated honey sample, the $\delta^{13}C_{Honey}$ term will not influence the $\delta^{13}C_{Protein}$ as strongly. The correlation coefficient of these two terms is difficult to establish particularly if both adulterated and pure honeys are considered together. Perhaps the easiest approach is to consider the numerator of Eq. (1) as a single quantity (i.e. a difference) which is independent of the $\delta^{13}C_{Protein}$ term in the denominator and therefore correlation need not be considered. The uncertainty in the numerator difference does still need to be known, however this can easily be estimated as a combination of the uncertainties in the $\delta^{13}C_{Protein}$ and $\delta^{13}C_{Honey}$ terms (neglecting to account for correlation within this combination of uncertainties will only result in an overestimation of uncertainty – which is again a conservative approach).

One of the simplest means to estimate the measurement uncertainty in this case is to use a Monte Carlo simulation [36–38]. The supplementary spreadsheet S2 contains the results of Monte Carlo simulations to estimate the measurement uncertainty in the % C_4 adulteration. Assuming that (i) the standard uncertainties of $\delta^{13}C_{Protein}$ and $\delta^{13}C_{Honey}$ are 0.15 and 0.075‰, respectively and therefore the combined uncertainty in their difference is 0.17‰; (ii) that the numerator of Eq. (1) is allowed to vary from 0 to −10‰; (iii) that $\delta^{13}C_{Protein}$ varies from −28 to −20‰; and (iv) that these values are normally distributed, the Monte Carlo simulations produce an expanded uncertainty in the % C_4 adulteration of between 1.8 and 4.3%. Note that this is not a relative uncertainty – the percentage C_4 adulteration varies between 0 and 97% over the same range of calculated results. This degree of uncertainty is certainly fit-for-purpose for application of the AOAC method as the lower limit of adulteration percentage (i.e. the value minus the expanded uncertainty) for differences of 1‰ or less are all < 7%.

In other forensic applications of carbon isotope analysis, differences in reported delta values between laboratories of up to 0.37‰ or a typical measurement uncertainty of 0.15‰ might be more or less

important – it will depend on the scenario in question and upon stakeholder needs. Where differences in isotope delta of < 1‰ between two materials are to be interpreted (which would not be critical for honey adulteration detection), it is vital that the measurement uncertainty associated to the two results at the 95% confidence level is estimated and considered. While there is still some inconsistency in how measurement uncertainty associated to isotope delta values is estimated as illustrated by the uncertainty budgets reported for CCQM-K140 [7], what is certain is that simply using precision data may be misleading and result in differences between two results being interpreted as significant when this would not be the case considering measurement uncertainty at the 95% confidence level.

4.6. Recommendations for future inter-laboratory studies of isotope delta values

The following points should be considered by organisers and participants of future ILCs of isotope ratio measurements, particularly those involving reporting of isotope delta values:

- Protocols for ILCs should include a minimum amount for sample analysis as is the case with certified reference material certificates.
- Participants should be required to report measurement uncertainties either instead of or in addition to standard deviations – and laboratories should be encouraged to do so whenever reporting results. These should be expanded uncertainties with 95% confidence.
- Measurement uncertainties should include contributions from all correction stages to raw data, particularly normalisation.
- Measurement uncertainties submitted by participants should be included in any interpretation of laboratory performance (i.e. ζ-scores and/or E_n numbers in place of z-scores).
- Reporting of all aspects of the measurement methods employed by participants should be encouraged as this can help identify the source(s) of poor performance and thereby make an ILC into a learning exercise rather than simply a measure of performance.

5. Conclusions

The main role of metrology institutes in the field of chemistry is to provide, directly or by provision of traceability to commercial producers, the chemical standards and matrix reference materials used by field laboratories to calibrate and validate their measurements. In order to do this the metrology institutes which participate in the CCQM often use more elaborate and time-consuming methods than is feasible for field laboratories. The organisation of a FIRMS laboratory ILC for delta values of carbon in the same honey sample in parallel with a CCQM comparison has provided an opportunity to extend the impact of the CCQM's activities by directly assessing the validity of routine forensic measurements. Overall there was good agreement between metrology institutes and forensic laboratories in terms of carbon isotope ratio analysis of bulk honey. The maximum difference in $\delta^{13}C_{VPDB\text{-}LSVEC}$ value for the bulk honey reported by a pair of participants was 0.37‰, while the typical expanded measurement uncertainty (with $k = 2$) reported was 0.28‰ and was > 2.5 times the typical standard deviation reported by the FIRMS participants. The exercise has also enabled the investigation of stable isotope laboratory performance using a range of performance metrics. While different performance metrics highlighted different laboratories as having questionable results, those metrics including the participant reported measurement uncertainty were the most useful. Reporting of standard deviations rather than full measurement uncertainties was highlighted as being a major contributing factor to poor performance metrics and also has the potential to impact interpretation of the results of forensic analyses.

Acknowledgements

We would like to thank the two anonymous reviewers for their constructive comments on an earlier version of this manuscript. LAC would like to thank Thuan Chau for assistance with sample preparation and John Howa for assistance with isotope data review and reporting (both IsoForensics, Inc). AC would like to acknowledge Colleen Podmore and Naghmeh Montazer Hojjat of Analytica Laboratories Ltd., for sample and sequence preparation. The work described in this paper was funded in part by the UK government Department for Business, Energy & Industrial Strategy (BEIS) and by EU projects: MASSTWIN - Spreading excellence and widening participation in support of mass spectrometry and related techniques in health, the environment and food analysis (H2020, GA no. 692241) and ERA Chair ISO-FOOD - for isotope techniques in food quality, safety and traceability (FP7, GA no. 621329).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scijus.2018.08.003.

References

[1] N. Gentile, R.T.W. Siegwolf, P. Esseiva, S. Doyle, K. Zollinger, O. Delemont, Isotope ratio mass spectrometry as a tool for source inference in forensic science: a review, For. Sci. Int. 251 (2015) 139–158.
[2] J.F. Carter, L.A. Chesson (Eds.), Food Forensics: Stable Isotopes as a Guide to Authenticity and Origin, CRC Press, Boca Raton, FL, 2017.
[3] J.F. Carter, S.P. Doyle, B.-L. Phasumane, N. NicDaeid, The role of isotope ratio mass spectrometry as a tool for the comparison of physical evidence, Sci. Justice 54 (2014) 327–334.
[4] AOAC Official Method 998.12. C-4 Plant Sugars in Honey. International Standard Stable Carbon Isotope Ratio Method, (2005).
[5] J. Vogl, Y.-H. Yim, K.-S. Lee, H. Goenaga-Infante, D. Malinowskiy, T. Ren, J. Wang, R.D. Vocke Jr., K. Murphy, N. Nonose, O. Rienitz, J. Noordmann, T. Näykki, T. Sara-Aho, B. Ari, O. Cankur, Final report of the key comparison CCQM-K98: Pb isotope amount ratios in bronze, Metrologia 51 (2014) (Technical Supplement).
[6] Decision CIPM/104-26 from Session I of the 104th meeting of the CIPM, https://www.bipm.org/utils/en/pdf/CIPM/CIPM2015-I-Decisions-EN.pdf, (2015) (accessed 25th October 2017).
[7] P.J.H. Dunn, H. Goenaga-Infante, A.C. Goren, A. Simsek, M. Bilsel, N. Ogrinc, P. Armishaw, L. Hai, CCQM-K140: carbon stable isotope ratio delta values in honey, Metrologia 54 (2017) (Technical Supplement).
[8] TUBITAK UME, Certification report for certified carbon isotope $\delta^{13}C_{VPDB}$) Reference Materials UME CRM 1309 – Sucrose, UME CRM 1310 – Glucose, UME CRM 1311 – Fructose, UME CRM 1312 – Honey (Unadulterated), UME CRM 1313 – Honey (Adulterated), http://www.ume.tubitak.gov.tr/sites/images/ume/ume_crm_1309_1310_1311_1312_1313_certification_report_.pdf, (2016) (accessed 25th October 2017).
[9] BIPM, CCQM guidance note: estimation of a consensus KCRV and associated degrees of equivalence, https://www.bipm.org/cc/CCQM/Allowed/19/CCQM13-22_Consensus_KCRV_v10.pdf, (2013) (accessed 21st June 2018).
[10] H. Craig, Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide, Geochim. Cosmochim. Acta 12 (1957) 133–149.
[11] J. Santrock, S.A. Studley, J.M. Hayes, Isotopic analyses based on the mass spectrum of carbon dioxide, Anal. Chem. 57 (1985) 1444–1448.
[12] W.A. Brand, S.S. Assonov, T.B. Coplen, Correction for the ^{17}O interference in $\delta(^{13}C)$ measurements when analysing CO_2 with stable isotope mass spectrometry (IUPAC Technical Report), Pure Appl. Chem. 82 (2010) 1719–1733.
[13] Good Practice Guide for Isotope Ratio Mass Spectrometry, in: J.F. Carter, V.J. Barwick (Eds.), FIRMS, 2011.
[14] Z. Sharp, Principles of Stable Isotope Geochemistry, Pearson/Prentice Hall, 2007.
[15] K.E. Anders Ohlsson, P. Håkan Wallmark, Novel calibration with correction for drift and non-linear response for continuous flow isotope ratio mass spectrometry applied to the determination of $\delta^{15}N$, total nitrogen, $\delta^{13}C$ and total carbon in biological material, Analyst 124 (1999) 571–577.
[16] T.B. Coplen, Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results, Rapid Commun. Mass Spectrom. 25 (2011) 2538–2560.
[17] W.A. Brand, T.B. Coplen, J. Vogl, M. Rosner, T. Prohaska, Assessment of international reference materials for isotope-ratio analysis (IUPAC Technical Report), Pure Appl. Chem. 86 (2014) 425–467.
[18] J.P. Jasper, Quantitative estimates of precision for molecular isotopic measurements, Rapid Commun. Mass Spectom. 15 (2001) 1554–1557.
[19] G. Skrzypek, R. Sadler, D. Paul, Error propagation in normalization of stable isotope data: a Monte Carlo analysis, Rapid Commun. Mass Spectom. 24 (2010) 2697–2705.

[20] J. Meija, M. M. G. Chartrand, Uncertainty evaluation in normalization of isotope delta measurement results against international reference materials. Anal. Bioanal. Chem. https://doi.org/10.1007/s00216-017-0659-1
[21] P.J.H. Dunn, L. Hai, D. Malinovsky, H. Goenaga-Infante, Simple spreadsheet templates for the determination of the measurement uncertainty of stable isotope ratio delta values, Rapid Commun. Mass Spectom. 29 (2015) 2184–2186.
[22] LGC Ltd, FIRMS PT scheme information, https://www.lgcstandards.com/GB/en/proficiency-testing/forensics/forensic-isotope-ratio-mass-spectrometry-proficiency-testing, (2017) (accessed 25th October 2017).
[23] Eurofins Scientific, FIT-PTS information, https://www.eurofins.com/food-and-feed-testing/food-testing-services/authenticity/fit-pts/, (2017) (accessed 25th October 2017).
[24] M. Thompson, Z-scores and other scores in chemical proficiency testing—their meanings, and some common misconceptions, Anal. Methods 8 (2016) 5553–5555.
[25] M. Mantha, J.R. Urban, W.A. Mark, A. Chernyshev, K.M. Kubachka, Direct comparison of cavity ring down spectrometry and isotope ratio mass spectrometry for detection of sugar adulteration in honey samples, J. AOAC Int. (2018), https://doi.org/10.5740/jaoacint.17-0491 in press.
[26] J.H. Williamson, Least-squares fitting of a straight line, Can. J. Phys. 46 (1968) 1845–1847.
[27] D.L. Möaggart, S.O. Farwell, Analytical use of linear regression. Part II: Statistical error in both variables, J. AOAC Int. 75 (1992) 608–614.
[28] ISO/IEC 17034:2016, General requirements for the competence of reference material producers, (2016).
[29] ISO/IEC Guide 34:2009, General requirements for the competence of reference material producers, (2009).
[30] M. Gröning, Improved water δ^2H and δ^{18}O calibration and calculation of measurement uncertainty using a simple software tool, Rapid Commun. Mass Spectom. 25 (2011) 2711–2720.
[31] T.B. Coplen, W.A. Brand, M. Gehre, M. Gröning, H.A.J. Meijer, B. Toman, R.M. Verkouteren, New guidelines for δ^{18}O measurements, Anal. Chem. 78 (2006) 2439–2441.
[32] J. Kragten, Calculation standard deviations and confidence intervals with a universally applicable spreadsheet technique, Analyst 119 (1994) 2162–2165.
[33] L.R. Croft, Stable isotope mass spectrometry in honey analysis, TrAC 6 (1987) 206–209.
[34] K.M. Rogers, K. Somerton, P. Rogers, J. Cox, Eliminating false positive C_4 sugar tests on New Zealand Manuka honey, Rapid Commun. Mass Spectrom. 24 (2010) 2370–2374.
[35] J.W. White, O.N. Rudyj, The protein content of honey, J. Apic. Res. 17 (1978) 234–238.
[36] ISO/IEC Guide 98-3:2008/Suppl 1:2008, Propagation of distributions using a Monte Carlo method, (2009).
[37] S. L. R. Ellison, A. Williams, (eds) EURACHEM/CITAC Guide: Quantifying uncertainty in analytical measurement, 3rd edition, Eurachem, https://eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf (accessed 21st June 2018).
[38] G. Chew, T. Walczyk, A Monte Carlo approach for estimating measurement uncertainty using standard spreadsheet software, Anal. Bioanal. Chem. 402 (2012) 2463–2469.