European Commission

ERM
European Reference Materials

CERTIFICATION REPORT

The certification of the mass fraction and mass concentration of total Hg in seawater:
ERM®-CA400

EUR 28264 EN – 2016

Abstract

This report describes the production of ERM-CA400, a seawater material certified for the mass concentration and mass fraction (calculated from the (certified) density) of total Hg. The material was produced following ISO Guide 34:2009.

Approximately 500 litres of surface seawater were collected from The Netherlands and transferred into a pre-cleaned polyester tank. Upon arrival at the JRC Geel, the water was acidified and passed first through a 0.8 μm, then through a 0.45 μm filter into another pre-cleaned tank. After filtration, the water was spiked with Hg^{2+} to achieve a final concentration of total Hg of approximately 20 ng/L. Aliquots of 100 mL were dispensed into borosilicate glass ampoules which were flame sealed and sterilised by gamma ray irradiation.

Between ampoule-homogeneity was quantified and stability during dispatch and storage were assessed following ISO Guide 35:2006.

The material was characterised by means of an inter-comparison among laboratories of demonstrated competence and adhering to ISO/IEC 17025 for Hg and by a primary method of measurement, confirmed by independent results for density. Technically invalid results were removed, but no outlier was eliminated on statistical grounds only.

Uncertainties of the certified values were calculated in compliance with the Guide to the Expression of Uncertainty in Measurement (GUM) and include uncertainties related to possible inhomogeneity, instability and characterisation.

The material is intended for the quality control and assessment of method performance. As any reference material, it can also be used for control charts or validation studies. The CRM is available as a set of three borosilicate glass ampoules each containing 100 mL of acidified seawater. The minimum amount of sample to be used is 10 g.

The CRM was accepted as European Reference Material (ERM®) after peer evaluation by the partners of the European Reference Materials consortium..

The certification of the mass fraction and mass concentration of total Hg in seawater:
ERM®-CA400

Raquel Sanchez Romero, Mitja Vahčič, James Snell, Albert Oostra, Håkan Emteborg, Thomas P.J. Linsinger, Andrea Held

Disclaimer

Certain commercial equipment, instruments, and materials are identified in this paper to specify adequately the experimental procedure. In no case does such identification imply recommendation or endorsement by the European Commission, nor does it imply that the material or equipment is necessarily the best available for the purpose.

Summary

This report describes the production of ERM-CA400, a seawater material certified for the mass concentration and mass fraction (calculated from the (certified) density) of total Hg. The material was produced following ISO Guide 34:2009 [1].

Approximately 500 litres of surface seawater were collected from The Netherlands and transferred into a pre-cleaned polyester tank. Upon arrival at the JRC Geel, the water was acidified and passed first through a 0.8 μm, then through a 0.45 μm filter into another pre-cleaned tank. After filtration, the water was spiked with Hg^{2+} to achieve a final concentration of total Hg of approximately 20 ng/L. Aliquots of 100 mL were dispensed into borosilicate glass ampoules which were flame sealed and sterilised by gamma ray irradiation.

Between ampoule-homogeneity was quantified and stability during dispatch and storage were assessed following ISO Guide 35:2006 [2].

The material was characterised by means of an inter-comparison among laboratories of demonstrated competence and adhering to ISO/IEC 17025 for Hg and by a primary method of measurement, confirmed by independent results for density. Technically invalid results were removed, but no outlier was eliminated on statistical grounds only.

Uncertainties of the certified values were calculated in compliance with the Guide to the Expression of Uncertainty in Measurement (GUM) [3] and include uncertainties related to possible inhomogeneity, instability and characterisation.

The material is intended for the quality control and assessment of method performance. As any reference material, it can also be used for control charts or validation studies. The CRM is available as a set of three borosilicate glass ampoules each containing 100 mL of acidified seawater. The minimum amount of sample to be used is 10 g.

The CRM was accepted as European Reference Material (ERM®) after peer evaluation by the partners of the European Reference Materials consortium.

The following values were assigned:

	Certified value [2]	Uncertainty [3]	Unit
Hg mass concentration[1]	16.8	1.1	ng/L
Hg mass fraction[1]	16.4	1.0	ng/kg
Density at 20 °C	1.0226	0.0003	g/cm³

1) As determined by cold vapour (CV) based methods.

2) Certified values are values that fulfil the highest standards of accuracy. The certified value for the Hg mass concentration represents the unweighted mean value of the means of accepted sets of data, each set being obtained in a different laboratory and/or with a different method of determination. The certified value for density represents the result obtained by a primary method of measurement (gravimetry), confirmed by results using oscillating tube density meters. The certified value for the Hg mass fraction was derived from mass concentration and density. The certified values and its uncertainty are traceable to the International System of Units (SI).

3) The uncertainty is the expanded uncertainty of the certified value with a coverage factor $k = 2$ corresponding to a level of confidence of about 95 % estimated in accordance with ISO/IEC Guide 98-3, Guide to the Expression of Uncertainty in Measurement (GUM:1995), ISO, 2008.

Disclaimer

Table of Contents

Glossary

AFS	Atomic Fluorescence Spectrometry
ASTM international	ASTM International (previous American Society for Testing and Materials)
ANOVA	Analysis of variance
b	Slope in the equation of linear regression $y = a + bx$
BCR$^®$	One of the trademarks of CRMs owned by the European Commission; previous Community Bureau of Reference
BP	Bingham Pycnometer
c	Mass concentration $c = m / V$ (mass / volume)
CI	Confidence interval
CRM	Certified reference material
CV-AFS	Cold vapour atomic fluorescence spectrometry
CV-ICP-MS	Cold vapour ICP- mass spectrometry
EC	European Commission
EQS	Environmental quality standards
ERM$^®$	Trademark of European Reference Materials
EU	European Union
ICP	Inductively coupled plasma
ICP-QMS	ICP-quadrupole mass spectrometry
ICP-SFMS	ICP-sector field mass spectrometry
ID	Isotope dilution
IDMS	Isotope dilution mass spectrometry
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
JRC	Joint Research Centre of the European Commission
k	Coverage factor
LOD	Limit of detection
LOQ	Limit of quantification
$MS_{between}$	Mean of squares between-unit from an ANOVA
MS_{within}	Mean of squares within-unit from an ANOVA
n	Number of replicates per unit
p	Number of technically accepted datasets in the characterisation study
N	Number of samples (units) analysed
n.a.	Not applicable
NIST	National Institute of Standards and Technology (USA)
OT	Oscillating tube electronic density meter
QC	Quality control
rel	Index denoting relative figures (uncertainties etc.)
RM	Reference material
RSD	Relative standard deviation
RSE	Relative standard error (=RSD/$\sqrt{n}$)
s	Standard deviation
s_{bb}	Between-ampoule standard deviation; an additional index "rel" is added when appropriate
se	Standard error of the mean
SI	International System of Units
s_{within}	Standard deviation within groups as obtained from ANOVA; an additional index "rel" is added as appropriate
s_{wb}	Within-ampoule standard deviation
T	Temperature

t	Time
t_i	Time point for each replicate
$t_{\alpha, df}$	Critical t-value for a t-test, with a level of confidence of 1-α and df degrees of freedom
t_{sl}	Proposed shelf life
u	standard uncertainty
U	expanded uncertainty
u^*_{bb}	Standard uncertainty related to a maximum between-ampoule inhomogeneity that could be hidden by method repeatability/intermediate precision select as appropriate; an additional index "rel" is added as appropriate
u_{bb}	Standard uncertainty related to a possible between-ampoule inhomogeneity; an additional index "rel" is added as appropriate
u_c	combined standard uncertainty; an additional index "rel" is added as appropriate
u_{cal}	Standard uncertainty of calibration
u_{char}	Standard uncertainty of the material characterisation; an additional index "rel" is added as appropriate
u_{CRM}	Combined standard uncertainty of the certified value; an additional index "rel" is added as appropriate
U_{CRM}	Expanded uncertainty of the certified value; an additional index "rel" is added as appropriate
u_Δ	Combined standard uncertainty of measurement result and certified value
u_{lts}	Standard uncertainty of the long-term stability; an additional index "rel" is added as appropriate
u_{meas}	Standard measurement uncertainty
u_{rec}	Standard uncertainty related to possible between-ampoule inhomogeneity modelled as rectangular distribution; an additional index "rel" is added as appropriate
u_{sts}	Standard uncertainty of the short-term stability; an additional index "rel" is added as appropriate
u_t	Standard uncertainty of trueness
$\bar{x}$	Arithmetic mean
$\bar{x}_{ns}$	Arithmetic mean of all results of normal stock samples
$\bar{x}_{ref}$	Arithmetic mean of results of reference samples
α	significance level
Δ_{meas}	Absolute difference between mean measured value and the certified value
$\nu_{s,meas}$	Degrees of freedom for the determination of the standard deviation s_{meas}
$\nu_{MSwithin}$	Degrees of freedom of MS_{within}

1 Introduction

1.1 Background

The Water Framework Directive (WFD) [4] provides a list of priority substances that present a risk for the good chemical status of the aquatic environment defined in terms of compliance with all the environmental quality standards (EQS) established in the daughter Directive 2013/39/EU [5].The EQS for mercury expressed as a maximum allowable concentration is 0.07 µg/L. Also, Directive 2009/90/EC [6] states that laboratories performing the water monitoringl shall demonstrate their competence by analysing reference materials that are representative of collected samples. While certified reference materials (CRM) for trace elements in natural waters are widely available, there is a limited number of CRMs for mercury at or below the level of the EQSs [7], especially for seawater matrix. Therefore, it was decided to produce a CRM for total Hg in seawater that is close to the EQS level and is in line with the requirements of current EU directives.

1.2 Choice of the material

A variety of approaches using different spectrometric methods are available for the determination of total Hg in water and seawater samples. However, seawater represents a unique challenge for the determination of total Hg, mainly because of high salt and halogen ion contents and the possibility of sample contamination with ambient Hg present in the air. In obtaining representative and valid results, it is necessary to check the method using a matrix matched CRM.

Because the typical concentration of the total Hg found in seawaters is below 10 ng/L [8, 9], it was decided to target the concentration for the new CRM at approximately 1/3 of the EQS (around 20 ng/L). To make the certified values immediately useful for laboratories working in mass fractions, it was decided to certify mass fractions as well. This necessitated the certification of the density of the material.

1.3 Design of the CRM project

ERM-CA400 was characterised by the mass concentration of Hg in seawater. After establishing the seawater density, the Hg mass fraction was calculated.

The total Hg concentration in ERM-CA400 was established through inter-comparison study. Eleven laboratories with experience and recognised expertise in the field were involved in the characterisation of total Hg in ERM-CA400. Expertise was assessed on beforehand by using data from proficiency tests and other CRM production projects. Method bias during the study was assessed using a quality control material (QCM) of known Hg mass concentration. The laboratories used several different sample preparation techniques followed by atomic fluorescence and mass spectrometry, both combined with a cold vapour (CV) generation step.

The density of the material was based on a primary method of measurement (gravimetry), confirmed by measurements with oscillating tube density meters.

2 Participants

The organisations participating in the project are listed below. Their accreditation status is mentioned only if the accreditation also covers the task performed. For example, no accreditation number is given for a laboratory accredited for the measurement of Hg in groundwater, but not seawater.

2.1 Project management and evaluation

European Commission, Joint Research Centre, Directorate F- Health, Consumers and Reference Materials, Geel, BE
(accredited to ISO Guide 34 for production of certified reference materials, BELAC No. 268-RM)

2.2 Processing

European Commission, Joint Research Centre, Directorate F- Health, Consumers and Reference Materials, Geel, BE

2.3 Homogeneity and stability studies

European Commission, Joint Research Centre, Directorate F- Health, Consumers and Reference Materials, Geel, BE
(measurements for total Hg under the scope of ISO/IEC 17025 accreditation BELAC No. 268-TEST)

2.4 Characterisation studies

2.4.1 Characterisation of the total Hg mass concentration

Ultra Traces Analyses Aquitaine (UT2A), Pau, FR

ALS Scandinavia AB, Luleå, SE
(measurements performed under the scope of ISO/IEC 17025 accreditation, SWEDAC, accreditation number 2030)

Brooks Rand Labs (BR), Seattle, USA
(measurements performed under the scope of ISO/IEC 17025 accreditation, ANSI-ASQ, accreditation number ADE-1447)

Energieonderzoek Centrum Nederland (ECN), Petten, NL
(measurements performed under the scope of ISO/IEC 17025 accreditation, NEN-EN-ISO/IEC 17025 accreditation certificate, number L 135)

Flett Research Ltd., Winnipeg, CA
(measurements performed under the scope of ISO/IEC 17025 accreditation, CALA, number A3306)

Institut "Jozef Stefan" (IJS), Department of Environmental Sciences, Ljubljana, SI
(measurements performed under the scope of ISO/IEC 17025 accreditation, Slovenska Akreditacija-LP090)

IVL Svenska Miljöinstitutet AB, Göteborg, SE
(measurements performed under the scope of ISO/IEC 17025 accreditation, SWEDAC, accreditation number 1213)

Laboratoire National d'Essais (LNE), Paris, FR
(measurements performed under the scope of ISO/IEC 17025 accreditation, COFRAC, accreditation number 2-54 rev. 3)

Rijkswaterstaat - Institute for Inland Water Management and Waste Water Treatment (RIZA)-Waterdienst, Lelystad, NL
(measurements performed under the scope of ISO/IEC 17025 accreditation, NEN-EN-ISO/IEC 17025 accreditation certificate, number: L 521)

SGS – Belgium NV, Antwerp, BE

Vlaamse Instelling voor Technologisch Onderzoek (VITO), Mol, BE
(measurements performed under the scope of ISO/IEC 17025 accreditation, BELAC number 045-TEST)

2.4.2   Characterisation of the density

Nederlands Meetinstituut, VSL, Delft, NL
(measurements performed under the scope of ISO/IEC 17025 accreditation as calibration laboratory, RVA, accreditation number K999)

Paragon Scientific, Prenton, UK
(measurements performed under the scope of ISO/IEC 17025 accreditation as calibration laboratory, UKAS, accreditation number 0649)

3   Material processing and process control

3.1   Origin of the starting material

The base material for ERM-CA400 coastal surface seawater, collected from Neeltje Jans (Oosterschelde), Zeeland, the Netherlands into polyester tank. Of this material, approximately 500 litres were transferred into perfluoroalkoxy - lined polyester tank and transferred to the JRC Geel.

A mercury ICP standard solution (mercury nitrate solution ($Hg(NO_3)_2$) in 10 % m/m nitric acid (HNO_3) with a concentration of 1000 mg/L and density of 1.054 g/mL) was used for spiking the seawater.

3.2   Processing

The seawater was acidified to a pH ≤ 2 with suprapure hydrochloric acid (HCl) (Merck KgA, Darmstadt, DE) upon arrival, filtered in a clean cell over a capsule membrane filter (Pall 12131 Versaflow filter with a 0.8 μm pre-filter and a 0.45 μm main filter; membrane consisting of an acrylic copolymer) and transferred into another pre-cleaned polyester tank using a Watson Marlow Peristaltic pump. After filtration, the seawater was spiked with a Hg solution to achieve a target concentration of approximately 20 ng/L of total Hg in seawater.

Mixing was effected by a pneumatically driven inert bellow-pump. The flow per pump was about 30 L per minute resulting in vigorous mixing of the contents. The seawater was mixed for >8 h per day during four days. Thereafter the water was left to stand and equilibrate for 2 weeks before one last filtration took place (Pall 12131 Versaflow filter with a 0.8 mm pre-filter and a 0.45 mm main filter; membrane consisting of an acrylic copolymer). Afterwards, the material was transferred into 3300 glass borosilicate ampoules, which were flame sealed after filling. Each ampoule was filled with approximately 100 mL of the solution. The flame-sealed ampoules were labelled and packed in Styrofoam boxes in sets of three ampoules (Figure 1).

Ampoules were allocated to the boxes in a way that each set contains one ampoule with a number, between 1-1099, one with a number between 1100 and 2199 and one with a number between 2200 and 3300. The finalised CRM units were sterilized by gamma ray irradiation.

Figure 1: ERM-CA400

3.3 Process control

The polyester tanks used for sampling and spiking of the seawater were tested for Hg contamination before use. After acid cleaning, the Hg concentration in the tanks was below the limit of detection.

After acidification and filtration of the seawater, the incipient total Hg concentration was measured in the seawater in three samples. The measurements showed a mean concentration of 1.24 ± 0.01 ng/L of total Hg present in the seawater before spiking.

The analyses of three samples taken immediately after the spiking and through the mixing of the seawater solution showed that the mean concentration agreed with the target concentration.

The total Hg was determined in all samples by cold vapour generation coupled to inductively coupled plasma mass spectrometry (CV-ICP-MS). The sample size used for analysis was 10 g.

During the CRM characterisation, some ampoules with numbers below 500 randomly showed very low Hg mass concentrations. This may have been caused by insufficient pre-conditioning of the filling equipment, resulting in loss of Hg through adsorption. Therefore, all ampoules with numbers lower than 500 were removed from the batch.

4 Homogeneity

A key requirement for any reference material is the equivalence between the various CRM units. In this respect, the important issue is whether the variation between units is significant compared to the uncertainty of the certified value. In contrast to that, it is not relevant if this variation between units is significant compared to the method repeatability in the homogeneity study. Consequently, ISO Guide 34 requires RM producers to quantify the between-unit variation. This aspect is covered in between-unit (in this case: between-ampoule) homogeneity studies.

The within-unit inhomogeneity does not influence the uncertainty of the certified value when the minimum sample intake is respected, but determines the minimum size of an aliquot that

is representative for the whole unit. Quantification of within-unit (in this case: within-ampoule) inhomogeneity is, therefore, necessary to determine the minimum sample intake.

4.1 Between-ampoule homogeneity

The between-ampoule homogeneity was evaluated to ensure that the certified values of the CRM are valid for all units of the material within the stated uncertainty.

4.1.1 Hg mass fraction

Five sets (15 ampoules) were selected using a random stratified sampling scheme covering the whole batch for the between-ampoule homogeneity test. This number of selected ampoules corresponds to approximately the cubic root of the total number of the produced ampoules. For the selection, the batch was divided into five groups (with a similar number of sets), and one set was selected randomly from each group. Six independent subsamples of 10 g were taken from each selected ampoule and analysed by cold vapour generation coupled to inductively coupled plasma mass spectrometry (CV-ICP-MS). The measurements were performed in a randomised block design because the necessary number of measurements cannot be included in a single run due to instrumental constraints (drift towards the end of a long run). Under such circumstances better precision (measured as the within-ampoule standard deviation) can be obtained using several short runs in a randomised block design than what can be obtained in a single run.

The randomised block design chosen for the measurement of six replicates on each of the 15 ampoules of ERM-CA400 consisted of three measurement runs (each of them on a different day), on each of which six replicates on five ampoules were analysed. Runs were randomised individually in a manner that could separate a potential analytical drift from a trend in the production sequence. The results are shown in Annex A.

Measurement results from ampoules below 500 were removed, because the results obtained for Hg in the characterisation study showed that some ampoules in that range had very low Hg mass concentrations.

The evaluation of the remaining data was performed in the following order:

1) Regression analyses to evaluate potential trends in each analytical run. A trend that was statistically significant at a 99 % confidence level was detected for the day 2. The trends of day 1 and 3 were not statistically significant at a 95 % confidence level.

2) Correction of the dataset for the significant analytical trend of the measurements of day 2. The correction of biases, even if they are statistically not significant, was found to combine the smallest uncertainty with the highest probability to cover the true value [10]. Correction of trends is therefore expected to improve the sensitivity of the subsequent statistical analysis through a reduction in analytical variation without masking potential between-ampoule heterogeneities. As the analytical sequence and the ampoule numbers were not correlated, trends significant at, at least a 95 % confidence level were corrected as shown below:

$$x_T(r,i) = x(r,i) - b(r) \cdot i \qquad\qquad \textit{Equation 1}$$

i position of the result in the analytical run

r number of the analytical run from 1 to 3

$b(r)$ slope of the linear regression for the analytical run r

$x(r,i)$ measurement results on the position i in the analytical run r

$x_T(r,i)$ corrected results for analytical trend on the position i in the analytical run r

3) The datasets for each day were checked for outliers and normality. One replicate of sample 3027 showed a result that was 30 % higher than the others. This result was flagged as an outlier at a 99 % confidence level by the Grubbs test. As the five other results of this sample agreed with the other results of this day, this high result is most likely caused by contamination and was, therefore, removed. No outlying average value was found for any of the other ampoules, and the results followed a normal distribution on each day.

4) The analytical trend-corrected dataset was evaluated for the statistically significant difference between analytical runs (95 % confidence level) using one-way ANOVA. A statistically significant difference between analytical runs was observed for mercury at 95 % confidence level.

5) The results for each day were normalised to the average of each day by dividing each result by the average of this particular day. The data were combined into one dataset, the data of which did not contain any outliers (Grubbs test on a 95 % confidence level). The ampoule means and individual results followed a normal distribution, but a regression analysis of ampoule mean against ampoule number showed a significant increase at a 95 % confidence level. This increase might have been caused by insufficient pre-conditioning of the filling equipment, leading to loss of Hg by adsorption in the first part of the filling sequence.

6) Due to the significant trend, the between-ampoule uncertainty of the Hg mass fraction ($u_{bb,rel}$) was estimated from a rectangular distribution between the highest and lowest ampoule arithmetic mean according to

$$u_{bb,rel} = \frac{|max - \min|}{2 * \sqrt{3} * average}$$

Equation 2

with max and min being the highest and lowest ampoule mean, respectively and average being the average of the all normalised ampoule means.

4.1.2 Density

The data from the long-term stability study were used for the homogeneity assessment, as the long-term stability showed no change over time (see section 5.2). 32 ampoules taken from twelve sets selected using a randomly stratified sampling scheme were analysed. Measurement results from ampoules below 500 were removed, because the results obtained for Hg in the characterisation study showed that some ampoules in that range had very low Hg mass concentrations. After discarding of these ampoules, 32 ampoules covering the whole batch remained for the homogeneity assessment.

Each ampoule was measured in quadruplicate, using a Densito 30PX density meter (Mettler-Toledo, Schwerzenbach, CH). The measurements were performed in one analytical run but split into four measurement blocks, where each ampoule was measured once. Measurements in each block were performed in a randomised order. The results are shown in Annex A.

The average result was higher than the eventually certified value (see 7.1), but the difference was covered by the uncertainty of the assigned value and the maximum bias specified for the instrument.

Regression analyses were performed to evaluate potential trends in the measurement sequence, as well as trends in the filling sequence. No trends in the filling sequence or the analytical sequence were observed at a 95 % confidence level. The dataset was assessed for consistency using Grubbs outlier tests at a confidence level of 99 % on the individual results and on the ampoule means. Two outlying individual results but no outlying ampoule means were detected. These outlying individual results are most likely caused by the limited resolution of the instrument (0.0001 g/cm^3), which means that, a result deviating one or two minimum steps is flagged as an outlier. These two values were retained, because flagging them as outliers are statistical artefacts.

Quantification of between-ampoule inhomogeneity was undertaken by analysis of variance (ANOVA), which separates the between- ampoule variation (s_{bb}) from the within- ampoule variation (s_{wb}). The latter is equivalent to the method repeatability if the individual samples were representative for the whole ampoule.

Evaluation by ANOVA requires that the mean values per ampoule follow at least a unimodal distribution and that results for each ampoule follow unimodal distributions with approximately the same standard deviations. The distribution of the mean values per ampoule was visually tested using histograms and normal probability plots. Too few data are available for each ampoule to make a clear statement of the distribution of individual results. Therefore, it was checked visually whether all individual data follow a unimodal distribution using histograms and normal probability plots and no indication of a deviation from a normal distribution was found.

It should be noted that $s_{bb,rel}$ and $s_{wb,rel}$ are estimates of the true standard deviations and are therefore subject to random fluctuations. Therefore, the mean square between groups ($MS_{between}$) can be smaller than the mean squares within groups (MS_{within}), resulting in negative arguments under the square root used for the estimation of the between-ampoule variation, whereas the true variation cannot be lower than zero. In this case, u^*_{bb}, the maximum inhomogeneity that could be hidden by method repeatability, was calculated as described by Linsinger *et al.* [11]. u^*_{bb} is comparable to the LOD of an analytical method, yielding the maximum inhomogeneity that might be undetected by the given study setup.

Method repeatability ($s_{wb,rel}$), between–ampoule standard deviation ($s_{bb,rel}$) and $u^*_{bb,rel}$ were calculated as:

$$s_{wb,rel} = \frac{\sqrt{MS_{within}}}{\overline{y}} \qquad \qquad Equation\ 3$$

$$s_{bb,rel} = \frac{\sqrt{\dfrac{MS_{between} - MS_{within}}{n}}}{\overline{y}} \qquad \qquad Equation\ 4$$

$$u^*_{bb,rel} = \frac{\sqrt{\dfrac{MS_{within}}{n}} \sqrt[4]{\dfrac{2}{v_{MSwithin}}}}{\overline{y}} \qquad \qquad Equation\ 5$$

MS_{within} mean square within ampoule from an ANOVA

$MS_{between}$ mean squares between ampoule from an ANOVA

$\overline{y}$ mean of all results of the homogeneity study

n mean number of replicates per ampoule

$v_{MSwithin}$ degrees of freedom of MS_{within}

13

4.1.3 Results

The results of the evaluation of the between- ampoule variation are summarised in Table 1. The resulting values from the above equations were converted into relative uncertainties.

Table 1: Results of the homogeneity study for total Hg and density

	Mean	$s_{bb,\,rel}$ [%]	$s_{wr,\,rel}$ [%]	$u^{*}_{bbr,\,rel}$ [%]	$u_{bb,rel}$ [%]
Hg	n.a.[1)]	n.a.[1)]	2.66	n.a.[1)]	1.63
Density	1.02322 g/cm^3	0.00006	0.0046	0.00087	0.00087

[1)] n.a.: not applicable, as the between-day variance and the trend in the filling sequence makes any grand mean or between- ampoule standard deviation meaningless.

4.2 Within- ampoule homogeneity and minimum sample intake

The material is a true solution and is not expected to have any relevant inhomogeneity. This assumption was confirmed by the stability-study, where sample intakes of 10 g were found to give acceptable repeatability.

Measurements during the characterisation study confirmed the homogeneity of density at a 1-2 mL level.

As it is unlikely that the material is used as a density standard, the minimum sample intake was set as for Hg, at 10 g or 9.8 mL.

5 Stability

Time, temperature and radiation were regarded as the most important influences on the stability of the materials. The influence of light was minimised by choice of the container and packaging which eliminates most of the incoming light. In addition, materials are stored and dispatched in the dark, thus practically eliminating the possibility of degradation by radiation. Additionally, the material was acidified, and sterilised by γ-irradiation/heat treatment to eliminate microbial growth. Therefore, only the influences of time and temperature needed to be investigated.

Stability testing is necessary to establish conditions for storage (long-term stability) as well as conditions for dispatch to the customers (short-term stability). During transport, especially in the summer, temperatures of up to 60 °C could be reached, and stability under these conditions must be demonstrated if transport at ambient temperature will be applied.

The stability studies were carried out using an isochronous design [12]. In this approach, samples are stored for a given time at different temperature conditions. Afterwards, the samples are moved to conditions where further degradation can be assumed to be negligible (reference conditions). At the end of the isochronous storage, the samples are analysed simultaneously under repeatability conditions. Analysis of the material (after various exposure times and temperatures) under repeatability conditions significantly improves the sensitivity of the stability tests.

5.1 Short-term stability study

5.1.1 Hg mass fraction

Samples were stored at 18 and 60 °C for 0, 1, 2 and 4 weeks. The reference temperature was set to 4 °C. Two sets (each containing three ampoules) per storage time and temperature were selected using a random stratified sampling scheme. From each ampoule,

three subsamples were measured by CV-ICP-MS. The measurements were performed in a randomised block design because the number of replicates of all units cannot be included in a single run due to instrumental constraints (drift towards the end of a long run). Under such circumstances, better precision (measured as the within- ampoule standard deviation) can be obtained using several short runs in a randomised block design than what can be obtained in a single run. In a randomised block design with three replicates on each ampoule of ERM-CA400, the simplest randomised block design involves four measurement runs, and eleven units are measured each day. The selection of the samples was made in a random order. Runs were randomised individually in a manner that can separate a potential analytical drift from a trend over storage time.

Measurement results from ampoules below 500 were removed, because the results obtained for Hg in the characterisation study showed that some ampoules in that range had very low Hg mass concentrations.

The data evaluation was performed in the following order:

1) Regression analyses to evaluate potential trends in each analytical run were performed. The trends in the analytical sequence were significant (95 % confidence level) on the second and third measurement day, pointing to a signal drift in the analytical system.

2) Correction of the dataset for significant analytical trend (95% confidence level). The correction of biases, even if they are statistically not significant, was found to combine the smallest uncertainty with the highest probability to cover the true value [10]. Correction of trends is therefore expected to improve the sensitivity of the subsequent statistical analysis through a reduction in analytical variation without masking the potential between- ampoule heterogeneities. As the analytical sequence and the ampoule numbers were not correlated, trends significant at, at least a 95 % confidence level were corrected as shown in Equation 1.

3) The analytical trend-corrected dataset was evaluated for the statistically significant difference between analytical runs (95 % confidence level) using one-way ANOVA. A statistically significant difference between analytical runs was observed for mercury on 95 % confidence level.

4) Normalisation of data, showing statistically significant difference between analytical run (95 % confidence level). As it is assumed that run-effects and unit-effects are independent, differences between analytical runs on at least a 95 % confidence level were corrected as shown below:

$$x_R(r,i) = \frac{x_T(r,i)}{\bar{x}_T(r)}$$

Equation 6

    i   position of the result in the analytical run

    r   number of the analytical run from 1 to 4

    $\bar{x}_T(r)$ mean results of the analytical run *r* after correction for the trend in analytical sequence

    $x_T(r,i)$ corrected results for analytical trend on the position *i* in the analytical run *r*

    $x_R(r,i)$ normalised results on the position *i* in the analytical run *r*

5) The obtained data were evaluated individually for each temperature. The results were screened for outliers using the single and double Grubbs test. No outliers were found at the 99 % confidence level.

Furthermore, the normalised data were evaluated against storage time and regression lines of normalised mass fraction versus time were calculated. The slopes of the regression lines were tested for statistical significance (loss/increase due to shipping conditions). At 18 °C, the slope of the regression line was not significantly different from zero (at 95 % confidence

level). However, at 60 °C the slope of the regression line (0.5 % /week) was significantly different from zero (on 95 % confidence level). The fact that the slope is positive (0.5 % per week) and that it is not significant at a 99 % level casts doubts on the reality of this change. Also, the extent of the slope is small compared to the other uncertainties.

All measurement results are shown in Annex B.

5.1.2 Density

Samples were stored at 60 °C for 0, 1, 2 and 4 weeks. The reference temperature was set to 4 °C. Two sets of three ampoules each per storage time were selected using a random stratified sampling scheme. From each ampoule, two samples were measured using the Mettler Toledo Densito 30PX density meter.

The dataset was tested for a trend in the analytical run or ampoule numbers and was screened for outliers using the single and double Grubbs test. No significant trend was detected, and no outliers were found at the 99 % confidence level.

All results are shown in Annex B.

5.2 Long-term stability study

5.2.1 Hg mass fraction

Data from two isochronous stability studies have been combined to assess the stability of the CRM.

For the first isochronous study, two sets of three ampoules per storage time and temperature were selected using a random stratified sampling scheme samples and were stored at 18°C for 0, 4, 8 and 12 months. The reference temperature was set to 4 °C. For the second isochronous study, two sets of three ampoules per storage time and temperature were selected using a random stratified sampling scheme samples and were stored at 18°C for 0, 8, 16 and 23 months. The reference temperature was set to 4 °C. From each ampoule, six samples were measured by CV-ICP-MS. This design allows separation of a potential analytical drift from a trend over storage time. The measurements were performed in a randomised block design because the number of replicates of all units cannot be included in a single run due to instrumental constraints (drift towards the end of a long run). Under such circumstances, better precision (measured as the within-ampoule standard deviation) can be obtained using several short runs in a randomised block design than what can be obtained in a single run. In a randomised block design with six replicates on each ampoule of ERM-CA400, the simplest randomised block design involves four measurement runs, and six ampoules are measured each day. The selection of the samples was made in a random order. Runs were randomised individually in a manner that can separate a potential analytical drift from a trend over storage time. A normalisation was applied to take into account differences between the two studies.

Measurement results from ampoules below 500 were removed, because the results obtained for Hg in the characterisation study showed that some ampoules in that range had very low Hg mass concentrations.

The data evaluation was performed in the following order:

1) Regression analyses to evaluate potential trends in each analytical run. Significant (95 % confidence level) trends in the analytical sequences were visible in some of the measurements for 12 and 24 months, pointing at a signal drift in the analytical system.

2) Correction of the dataset for significant analytical trend (95% confidence level). The correction of biases, even if they are statistically not significant, was found to combine the smallest uncertainty with the highest probability to cover the true value [10]. Correction of trends is therefore expected to improve the sensitivity of the subsequent statistical analysis

through a reduction in analytical variation without masking the potential between- ampoule heterogeneities. As the analytical sequence and the ampoule numbers were not correlated, trends significant at, at least a 95 % confidence level were corrected as shown in Equation 1.

3) The analytical trend-corrected dataset was evaluated for the statistically significant difference between analytical runs (95 % confidence level) using one-way ANOVA. A statistically significant difference between analytical runs was observed for mercury at 95 % confidence level.

4) Normalisation of data, showing statistically significant difference between analytical run (95 % confidence level). As it is assumed that run-effects and unit-effects are independent, differences between analytical runs on at least a 95 % confidence level were corrected as shown in Equation 6.

5) The obtained data were evaluated individually for each study. The results were screened for outliers using the single and double Grubbs test. No outliers were found at the 99 % confidence level for both studies.

Furthermore, the data were plotted against storage time and regression lines of mass fraction versus time were calculated. The slope of the regression lines was tested for statistical significance (loss/increase due to storage conditions). For both studies, the slopes of the regression lines were not significantly different from zero (99 % confidence level).

6) Afterwards, the normalised results of the two isochronous studies were combined into one single dataset. The combined data were plotted against storage time and regression line of mass fraction versus time was calculated. The slope of the regression line was tested for statistical significance. The slope of the regression lines was not significantly different from zero (95 % confidence level).

All measurement results are shown in Annex C.

No technically unexplained outliers were observed, and none of the trends was statistically significant at a 95 % confidence level. The material can, therefore, be stored at 18 °C.

5.2.2 Density

Three sets of three ampoules per storage time were selected using a random stratified sampling scheme and were stored for 8, 16 and 23 months at 18 °C. Each ampoule was measured in quadruplicate using the Mettler Toledo Densito 30PX density meter. Measurement results from ampoules below 500 were removed, because the results obtained for Hg in the characterisation study showed that some ampoules in that range had very low Hg mass concentrations.

All results are shown in Annex C.

The dataset was tested for a trend in the analytical run, a trend over the ampoules and was screened for outliers using the single and double Grubbs test. No significant trend was detected. Two outlying individual results, but no outlying ampoule means were detected. These outlying individual results are most likely caused by the limited resolution of the instrument (0.0001 g/cm^3), which means that a result that deviates one or two minimum steps is flagged as an outlier. These two values were retained, as they are statistical artefacts.

5.3 Estimation of uncertainties

Due to the intrinsic variation of measurement results, no study can rule out degradation of materials completely, even in the absence of statistically significant trends. It is, therefore, necessary to quantify the potential degradation that could be hidden by the method repeatability, i.e. to estimate the uncertainty of stability. This means, even under ideal conditions, the outcome of a stability study can only be "degradation is 0 ± x % per time".

Uncertainties of stability during storage and dispatch was estimated as described in [13] for all studies without a significant slope (density for dispatch and storage; Hg mass fraction for storage). For this approach, the uncertainty of the linear regression line with a slope of zero is calculated. The uncertainty contributions u_{sts} and u_{lts} are calculated as the product of the chosen transport time/shelf life and the uncertainty of the regression lines as:

$$u_{lts,rel} = \frac{RSD}{\sqrt{\sum (x_i - \bar{x})^2}} \cdot t_{sl}$$

Equation 7

RSD   relative standard deviation of all results of the stability study

x_i result at time point *i*

$\bar{x}$ mean results for all time points

t_{sl} chosen shelf life (36 months at 18 ⁰C)

The same approach was used to estimate the uncertainty during dispatch under cooling conditions for one week, i.e. with temperatures below 18 °C ($u_{sts,rel,\ 18\ °C}$). Also, the uncertainty for dispatch under worst-case conditions (up to 60 °C) for one week ($u_{sts,rel,\ 60°C}$) was calculated. As the mass fraction of Hg shows a significant trend at 60 °C, the uncertainty due to the change over this time (u_{deg}) was combined with an uncertainty contribution reflecting the uncertainty of this change (u_b)

$$u_{sts,rel,60°C} = \sqrt{u_{deg}^2 + u_b^2}$$

Equation 8

$$u_{deg} = \frac{b \cdot t_{tt}}{\sqrt{3}}$$

Equation 9

$$u_b = \frac{s_{y.x}}{\sqrt{\sum (x_i - \bar{x})^2}} \cdot t_{tt}$$

Equation 10

`

$u_{deg,}$ ....uncertainty contribution due to degradation

b.........................slope of the regression line

u_b ....uncertainty due to lack of fit of the regression line at the time t_{tt}

s_{yx}.......................standard error of the estimation

t_{tt} ....chosen transport time (1 week at 60 ⁰C)

The following uncertainties were estimated:

-     $u_{sts,rel}$, the uncertainty of degradation during dispatch. This was estimated from the 60 °C study. The uncertainty describes the possible change during a dispatch at 60 °C lasting for one week.

- $u_{lts,rel}$, the stability during storage. This uncertainty contribution was estimated from the combined 18 °C studies. The uncertainty contribution describes the possible degradation during 36 months storage at 18 °C.

The results of these evaluations are summarised in Table 2.

Table 2: Uncertainties related to stability during dispatch and storage. $u_{sts,rel}$ was calculated for a temperature of 60 °C for 1 week; $u_{lts,rel}$ was calculated for a storage temperature of 18 °C and 3 years

Measurand	$u_{sts,rel, 18 °C}$ [%]	$u_{sts,rel, 60 °C}$ [%]	$u_{lts,rel}$ [%]
Hg mass fraction	0.24	0.36	0.63
Density	Not tested	0.00086	0.0017

The material shows no change of density but a significant change of the Hg mass fraction at 60 °C. The extent of change is small compared to the uncertainties of homogeneity and characterisation and the material will be shipped with cooling elements.

No change was observed for storage at 18 °C and the material will therefore be stored at 18 °C.

After the certification campaign, the material will be subjected a regular stability monitoring programme to control its further stability.

6   Characterisation

The material characterisation is the process of determining the property values of a reference material.

The material characterisation of the total Hg mass concentration was based on an intercomparison among expert laboratories, i.e. the total Hg mass concentration of the material was determined in different laboratories that applied different sample preparation procedures and different quantification methods to demonstrate the absence of a measurement bias. This approach aims at randomisation of laboratory bias, which reduces the combined uncertainty. While different sample preparation and detection methods were applied, all methods were based on cold vapour (CV) generation. It was therefore impossible to assess whether this CV step influences the analytical result and therefore needs to be reflected in the definition of the measurand.

The material characterisation for density was based on gravimetry, which is a primary method of measurement, confirmed by an independent method. A primary method of measurement (also called "primary reference method" in the International Vocabulary of Metrology (VIM) [14]) is a method that does not require calibration with a standard of the same measurand and does not depend on a chemical reaction. Such methods are of highest metrological order and often yield results with low uncertainties. However, it is prudent to demonstrate the absence of bias or gross errors using an independent method of lower metrological order.

Hg mass fraction was then calculated from the Hg mass concentration and the density.

6.1 Selection of participants

For the characterisation of the total Hg mass fraction, 11 laboratories were selected based on criteria that comprised both technical competence and quality management aspects. Each participant was required to operate a quality system and to deliver documented evidence of its laboratory proficiency in the field of element measurements in relevant matrices by submitting results for inter-comparison exercises or method validation reports.

For density, three different laboratories were selected that had shown their proficiency in the determination of the density of water in the characterisation study of ERM-CA403 [15].

Having a formal accreditation was not mandatory, but meeting the requirements of ISO/IEC 17025 was obligatory. Where measurements are covered by the scope of accreditation, the accreditation number is stated in the list of participants (Section 2.4).

6.2 Study setup A

6.2.1 Hg mass concentration

Each laboratory received 2 units (6 ampoules) of ERM-CA400 and was requested to provide six independent results, one per ampoule. The units for material characterisation were selected using a random stratified sampling scheme and covered the whole batch. The sample preparations (if necessary) and measurements had to be spread over at least two days to ensure intermediate precision conditions.

Each participant received a sample of spiked BCR-579 (coastal seawater) as a blinded quality control (QC) sample. BCR-579 had a certified value of 1.85 ± 0.20 ng/kg and each unit was individually fortified with Hg to a final mass fraction of 22.64 ± 2.26 ng/kg. The results for this sample were used to support the evaluation of the characterisation results.

Laboratories were also requested to give estimations of the expanded uncertainties of the mean value of the six results. No approach for the estimation was prescribed, i.e. top-down and bottom-up were regarded as equally valid procedures.

6.2.2 Density

Each laboratory received 3 sets (9 ampoules) of ERM-CA400 and was requested to provide two independent measurements of density at 20 °C on each of the sets (i.e. not all ampoules were measured). Also, each laboratory received one ampoule of ERM-CA403 [15] as a quality control (QC) sample. The results for this sample were used to support the evaluation of the characterisation results.

Laboratories were also requested to give estimations of the expanded uncertainties of the mean value of the six results. No approach for the estimation was prescribed, i.e. top-down and bottom-up were regarded as equally valid procedures.

6.3 Methods used

A variety of sample pretreatment methods (dilution, digestion with HNO_3, digestion with BrCl and digestion with aqua regia) with different quantification steps (isotope-dilution CV-ICP-MS, CV-AFS) were used to characterise the material for the total Hg mass fraction. The combination of results from methods based on different sample preparation principles mitigates undetected bias in this step.

One laboratory determined the density using a primary method of measurement (Bingham Pycnometer, ASTM D480). The results of this primary method were confirmed by two sets of measurements obtained from electronic density meters based on the oscillating tube principle obtained independently from two different laboratories.

All methods used during the characterisation study are summarised in Annex D. The laboratory code (e.g. L01) is a random code and does not correspond to the order of laboratories in Section 2.4. The lab-method code consists of a number assigned to each laboratory (e.g. L01) and abbreviation of the measurement method used, (e.g. CV-AFS).

6.4 Evaluation of results

The characterisation study resulted in 11 datasets for total Hg mass concentration and three datasets for density. All individual results of the participants, grouped per property are displayed in a tabular and graphical form in Annex E. As was done for homogeneity and stability studies, data from ampoules with sample numbers below 500 were excluded from the evaluation regardless of agreement/disagreement with other results.

6.4.1 Technical evaluation

The data obtained were first checked for compliance with the analysis protocol and for their validity based on technical reasons. The following criteria were considered during the evaluation:

- compliance with the analysis protocol: sample preparations and measurements performed in two days and the analytical sequence.

- absence of values given as below limit of detection or below limit of quantification.

- method performance, i.e. whether the results agreed with the assigned value of the QC materials on a 95 % confidence level, considering both the expanded uncertainty of the assigned value and the expanded uncertainty of the reported measurement results. Datasets that did not meet this criterion were rejected, regardless whether the results on the candidate CRM agreed with the other datasets or not.

All laboratories complied with the analysis protocol and no laboratory submitted data given as below the limit of detection. Because of deviations from the assigned value of the QCM, data for Hg from L01, L05, L07, L08 and L09 were therefore excluded from further data treatment on technical grounds.

All measurement results for density fulfilled all criteria and all datasets were therefore accepted on technical grounds.

6.4.2 Statistical evaluation

The technically accepted datasets for Hg mass concentration were tested for normality of dataset means using kurtosis/skewness tests and normal probability plots and were tested for outlying means using the Grubbs test and using the Cochran test for outlying standard deviations, (both at a 99 % confidence level). Standard deviations within (s_{within}) and between ($s_{between}$) laboratories were calculated using one-way ANOVA. The results of these evaluations are shown in Table 3.

Table 3: Statistical evaluation of the technically accepted datasets for Hg. p*: number of technically valid datasets*

	p	Outliers		Normally distributed	Statistical parameters			
		Means	Variances		Mean [ng/L]	*s* [ng/L]	$s_{between}$ [ng/L]	s_{within} [ng/L]
Hg mass concentration	6	no	no	yes	16.80	1.00	0.95	0.70

The laboratory means for Hg follow a normal distribution. None of the data contains outlying means and variances. The datasets are therefore consistent, and the mean of laboratory means is a good estimate of the true value. The standard deviation among laboratories is

considerably larger than the standard deviation within laboratories, showing that confidence intervals of replicate measurements are unsuitable as an estimate of measurement uncertainty. The relative uncertainty of characterisation ($u_{char,rel)}$ was estimated as the standard error of the mean of laboratory means as shown in Equation 11.

$$u_{char,rel} = \frac{s}{\overline{x} \cdot \sqrt{p}_{tt}}$$ *Equation 11*

$\overline{x}$	arithmetic mean of laboratory means
s	standard deviation of laboratory means
p	number of technically accepted datasets

For density, the results from the measurements made by oscillating tube agreed with the gravimetric results from the Bingham Pycnometer. The results from the oscillating tube, therefore, confirm the absence of undetected errors in the gravimetric method, the results of which is, therefore, a good estimate of the true value. The uncertainty provided by the laboratory was used as u_{char}.

Table 1: Uncertainty of characterisation for ERM-CA400. The data for density show the results from the Bingham pycnometer only, for which all 6 replicate measurements gave the same result.

	p	Mean	s	u_{char}	$u_{char, rel}$
Hg	6	16.80 ng/L	1.00 ng/L	0.41 ng/L	2.43 %
density	1	1.0226 g/cm^3	0.0000 g/cm^3	0.0001 g/cm^3	0.00978 %

7 Value Assignment

Certified values were assigned to this material.

Certified values are values that fulfil the highest standards of accuracy. Procedures applied for JRC's certified reference materials require pooling not less than 6 datasets to assign certified values or using data from a primary method of measurement confirmed by results from another method. Full uncertainty budgets in accordance with the 'Guide to the Expression of Uncertainty in Measurement' [3] were established.

7.1 Certified values and their uncertainties

The unweighted mean of the means of the accepted datasets as shown in Table 3 was assigned as certified value for the Hg mass concentration. The value obtained by the Bingham Pycnometer was assigned as the certified value for density.

The assigned uncertainty consists of uncertainties relating to characterisation, u_{char} (Section 6.4.2), potential between-ampoule inhomogeneity, u_{bb} (Section 4.1), and potential degradation during transport, u_{sts}, and long-term storage, u_{lts} (Section 0). These different contributions were combined to estimate the relative expanded uncertainty of the certified value ($U_{CRM, rel}$) with a coverage factor k as follows:

$$U_{CRM, rel} = k \cdot \sqrt{u_{bb, rel}^2 + u_{sts, rel}^2 + u_{lts, rel}^2 + u_{char, rel}^2} \qquad \text{Equation 12}$$

- u_{char} was estimated as described in Section 6.4.2.
- u_{bb} was estimated as described in Section 4.1.
- u_{sts} and u_{lts} were estimated as described in Section 0.

The numbers of the degrees of freedom of the different uncertainty contributions add up to at least 11 effective degrees of freedom calculated according to the Welch-Satterthwaite equation [3]. Therefore, a coverage factor k of 2 was applied to obtain the expanded uncertainties. The uncertainty for density was rounded to the nearest 0.0001 g/cm³ to stay in line with the reported number of digits for the density measurements themselves.

The certified values and their uncertainties are summarised in Table 5.

Table 2: Certified values and their uncertainties for ERM-CA400. The value for the Hg mass fraction was calculated from the certified values for Hg mass concentration and density.

	Certified value	$u_{char, rel}$ [%]	$u_{bb, rel}$ [%]	$u_{sts, rel}$ [%]	$u_{lts, rel,}$ [%]	U_{CRM}
Hg mass concentration	16.8 ng/L	2.43	1.63	0.36	0.63	1.1 ng/L
Hg mass fraction	16.4 ng/kg	2.43	1.63	0.36	0.63	1.0 ng/kg
density	1.0226 g/cm³	0.0098	0.0009	0.0009	0.0017	0.0003 g/cm³

8 Metrological traceability and commutability

8.1 Metrological traceability

Identity

The participants used different methods for the sample preparation, as well as for the final determination, demonstrating the absence of measurement bias in these two steps. As all quantification methods used in this study are based on CV generation, it is impossible to assess whether this step has an influence of the result. The measurand is therefore defined as "as determined by CV based methods".

Density is a clearly defined physical property which is independent of any measurement method.

Quantity values

Only validated methods were used for the determination of the assigned values for the Hg mass concentration and Hg mass fraction. Different calibrants of known purity and specified traceability of their assigned values were used, and all relevant input parameters were calibrated. The individual results are therefore traceable to the SI, as it is also confirmed by the agreement among the technically accepted datasets. As the assigned values are combinations of agreeing results individually traceable to the International System of Units (SI), the assigned quantity values themselves are also traceable to the SI.

Density was determined by a primary method of measurement, the values of which were confirmed by independent measurements in other laboratories. All relevant input parameters were calibrated, as shown by the agreement of the result with the assigned value of the QC material. The individual results are, therefore, traceable to the SI, as it is also confirmed by the agreement among the technically accepted datasets. As the assigned values are combinations of agreeing results individually traceable to the International System of Units (SI), the assigned quantity values themselves are also traceable to the SI.

8.2 Commutability

Many measurement procedures include one or more steps which select specific (or specific groups of) analytes from the sample for the subsequent whole measurement process. Often, the complete identity of these 'intermediate analytes' is not fully known or taken into account. Therefore, it is difficult to mimic all analytically relevant properties of real samples within a CRM. The degree of equivalence in the analytical behaviour of real samples and a CRM on various measurement procedures (methods) is summarised in a concept called 'commutability of a reference material'. Various definitions define this concept. For instance, the CLSI Guideline C53-A [16] recommends the use of the following definition for the term *commutability*:

> "The equivalence of the mathematical relationships among the results of different measurement procedures for an RM and for representative samples of the type intended to be measured."

The commutability of a CRM defines its fitness for use and is, therefore, a crucial characteristic when applying different measurement methods. When the commutability of a CRM is not established, the results from routinely used methods cannot be legitimately compared with the certified value to determine whether a bias does not exist in calibration, nor can the CRM be used as a calibrant.

ERM-CA400 was produced from an acidified and spiked natural coastal seawater. The spike constitutes that largest part of the Hg concentration and consists entirely of Hg^{2+}. This is not the case for natural seawaters, in which Hg^{2+} often is the main fraction, but which also

contain other species, amongst others dissolved gaseous Hg, monomethylHg and dimethylHg [17]. Therefore, the analytical behaviour of this CRM may differ from a routine sample of coastal seawater and users consequently should assess the commutability of this CRM for their method.

9   Instructions for use

9.1   Safety information

The general laboratory safety measures apply.

9.2   Storage conditions

The materials should be stored at (18 ± 5) °C in the dark.

Please note that the European Commission cannot be held responsible for changes that happen during storage of the material at the customer's premises, especially for opened ampoules.

9.3   Opening of ampoules

Use a sharp glass cutter to make a deep scratch 1.5 – 2 cm below the top of the ampoule (several scratches are usually needed to obtain a sufficiently deep scratch). Then break off the top part, either by hitting it with the glass cutter, by hand or using a specific ampoule opening tool.

Note: The ampoules are rather thick and can be difficult to break. When not using an ampoule opener, wrapping the ampoules into a tissue to avoid injuries is advisable.

During all operations, take precautions to avoid contamination of the material.

9.4   Minimum sample intake

The minimum sample intake representative for the material is 10 g or 9.8 mL

9.5   Use of the certified value

The main purpose of these materials is to assess method performance, i.e. for checking the accuracy of analytical results. Like any reference material, it can be used for establishing control charts or validation studies or calibration.

Use as a calibrant

It is not recommended to use this matrix material as a calibrant. Nevertheless, if used, the uncertainty of the certified value shall be taken into account in the estimation of the measurement uncertainty.

Comparing an analytical result with the certified value

A result is unbiased if the combined standard uncertainty of measurement and certified value covers the difference between the certified value and the measurement result (see also ERM Application Note 1, www.erm-crm.org [18].

When assessing the method performance, the measured results on the CRMs are compared with the certified values. The procedure is summarised here:

- Calculate the absolute difference between mean measured value and the certified value (Δ_{meas}).

- Combine the measurement uncertainty (u_{meas}) with the uncertainty of the certified value (u_{CRM}): $u_\Delta = \sqrt{u_{meas}^2 + u_{CRM}^2}$

- Calculate the expanded uncertainty (U_Δ) from the combined uncertainty (u_Δ) using an appropriate coverage factor, corresponding to a level of confidence of approximately 95 %

- If $\Delta_{meas} \leq U_\Delta$ then no significant difference exists between the measurement result and the certified value, at a confidence level of approximately 95 %.

Use in quality control charts

The materials can be used for quality control charts. Using CRMs for quality control charts has the added value that a trueness assessment is built into the chart.

10 Acknowledgments

11 References

1 ISO Guide 34, *General requirements for the competence of reference materials producers*, International Organization for Standardization, Geneva, Switzerland, 2009

2 ISO Guide 35, *Reference materials – General and statistical principles for certification*, International Organization for Standardization, Geneva, Switzerland, 2006

3 ISO/IEC Guide 98-3, *Guide to the Expression of Uncertainty in Measurement*, (GUM 1995), International Organization for Standardization, Geneva, Switzerland, 2008

4 European Commission, Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy, Off. J. Eur. Comm., L 327, 2000

5 European Commission, Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 on Amendment Directives 2000/60/EC and 2008/105/EC as Regards Priority Substances in the Field of Water Policy, Off. J. Eur. Union, L 226, 2013, p. 1

6 European Commission, Directive 2009/90/EC of the European Parliament and of the Council of 31 July 2009 on Technical Specifications for Chemical Analysis and Monitoring of Water Status, Off. J. Eur. Union, L 201, 2009

7 M. Ricci, I. Kourtchev, H. Emons H *Chemical water monitoring under the Water Framework Directive with Certified Reference Materials,* TRAC-Trend Anal Chem 36(2012):47-57

8 D. Cossa, M-h. Cotte-Krief, R.P. Mason, J. Bretaudeau-Sanjuan, *Total mercury in the water column near the shelf edge of the European continental margin*, Marine Chemistry 90(2004) 21-29

9 M. Coquery, D. Cossa, *Mercury speciation in surface waters of the north sea*, Netherlands Journal of Sea Research 34(1995) 245-257

10 G. O'Donnell, D. Brynn, Hibbert, *Treatment of bias in estimating measurement uncertainty*, Analyst 130 (2005) 721-729

11 T.P.J. Linsinger, J. Pauwels, A.M.H. van der Veen, H. Schimmel, A. Lamberty, *Homogeneity and stability of reference materials*, Accred. Qual. Assur. 6 (2001) 20-25

12 A. Lamberty, H. Schimmel, J. Pauwels, *The study of the stability of reference materials by isochronous measurement*s, Fres. J. Anal. Chem. 360 (1998) 359-361

13 T.P.J. Linsinger, J. Pauwels, A. Lamberty, H. Schimmel, A.M.H. van der Veen, L. Siekmann, *Estimating the Uncertainty of Stability for Matrix CRMs*, Fres. J. Anal. Chem. 370 (2001) 183-188

14 ISO Guide 99, *International vocabulary of metrology -- Basic and general concepts and associated terms (VIM)*, International Organization for Standardization, Geneva, Switzerland, 2007

15 R. Sánchez, E. de Vos, J. Snell, H. Emteborg, J. Snell, A. Held, *The certification of the mass concentration of As, Cd, Co, Cu, Mn, Mo, Ni and Pb, and density in seawater: ERM®-CA403*, , Publications Office of the European Union, Luxembourg (in press)

16 H. Vesper, H. Emons, M. Gnezda, C. P. Jain, W. G. Miller, R. Rej, G. Schumann, J. Tate, L. Thienpont, J. E. Vaks, *Characterization and qualification of commutable reference materials for laboratory medicine; approved guideline*, CLSI document C53-A, Clinical and Laboratory Standards Institute, Wayne, PA, USA, 2010

17 M. Horvat, J. Kotnik, M. Logar, V. Fajon, T. Zvonarić, N. Pirrone*, Speciation of mercury in surface and deep-sea waters in the Mediterranean Sea*, Atmospheric Environment 37(2003) S93-S108

18 T.P.J. Linsinger, *ERM Application Note 1: Comparison of a measurement result with the certified value*, www.erm-crm.org (last accessed on 12.10.2016)

Annexes

Annex A: Results of the homogeneity measurements
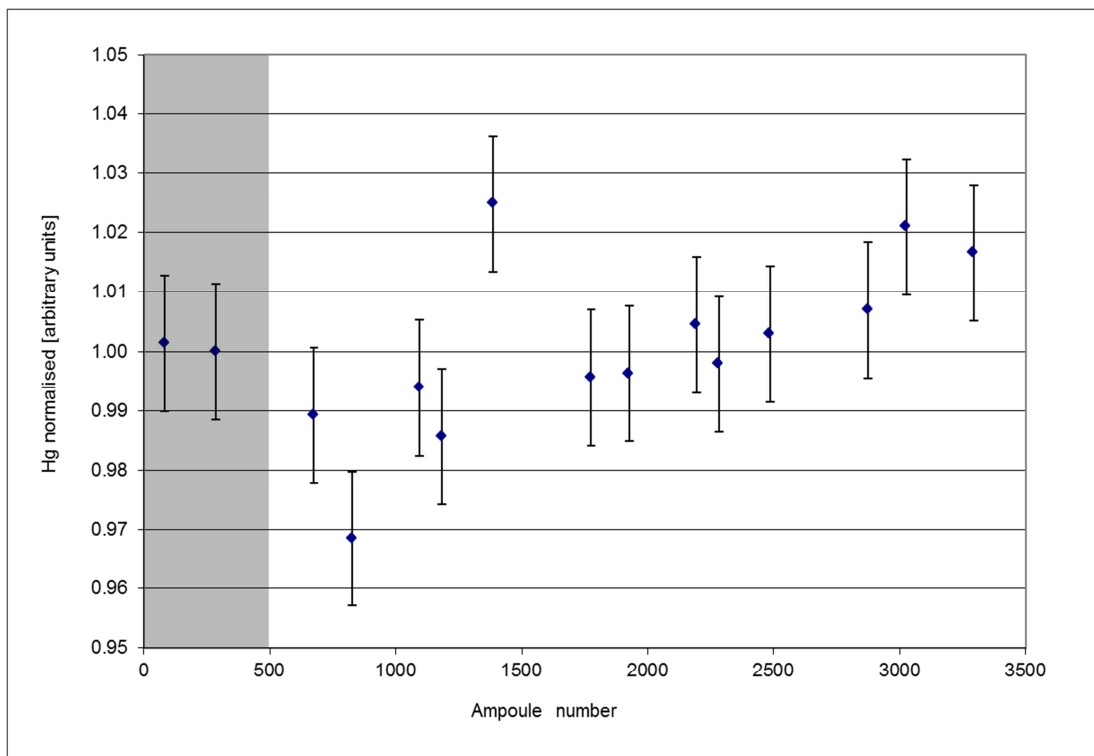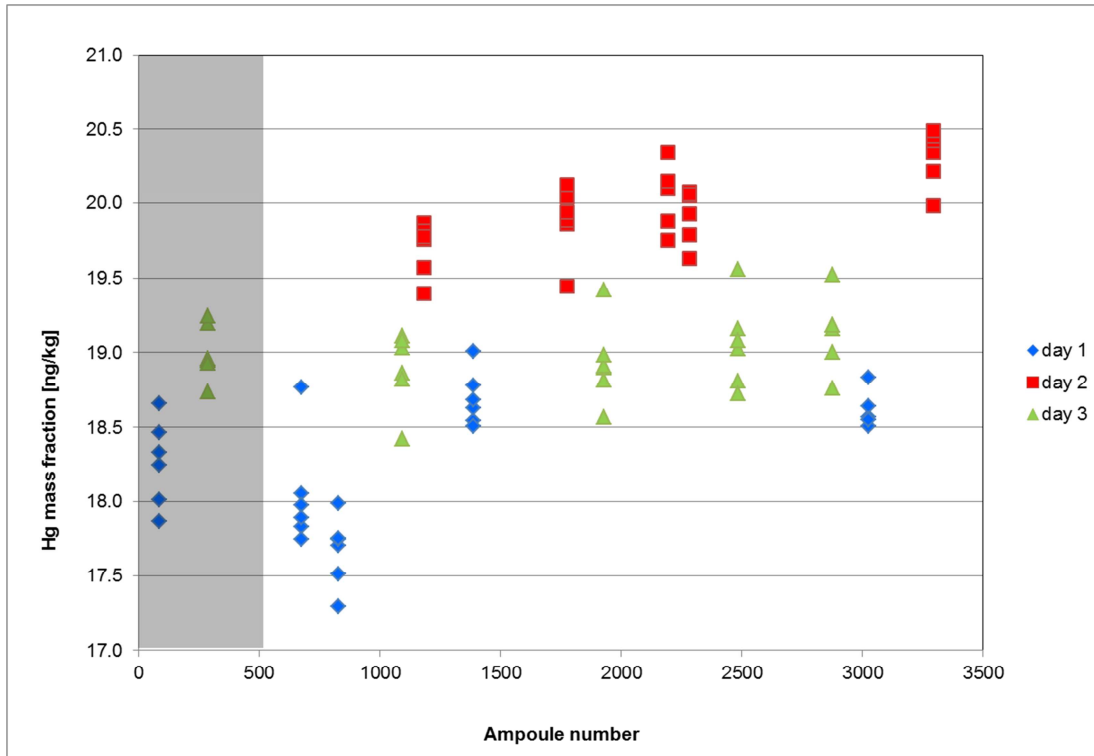
Annex B: Results of the short-term stability measurements

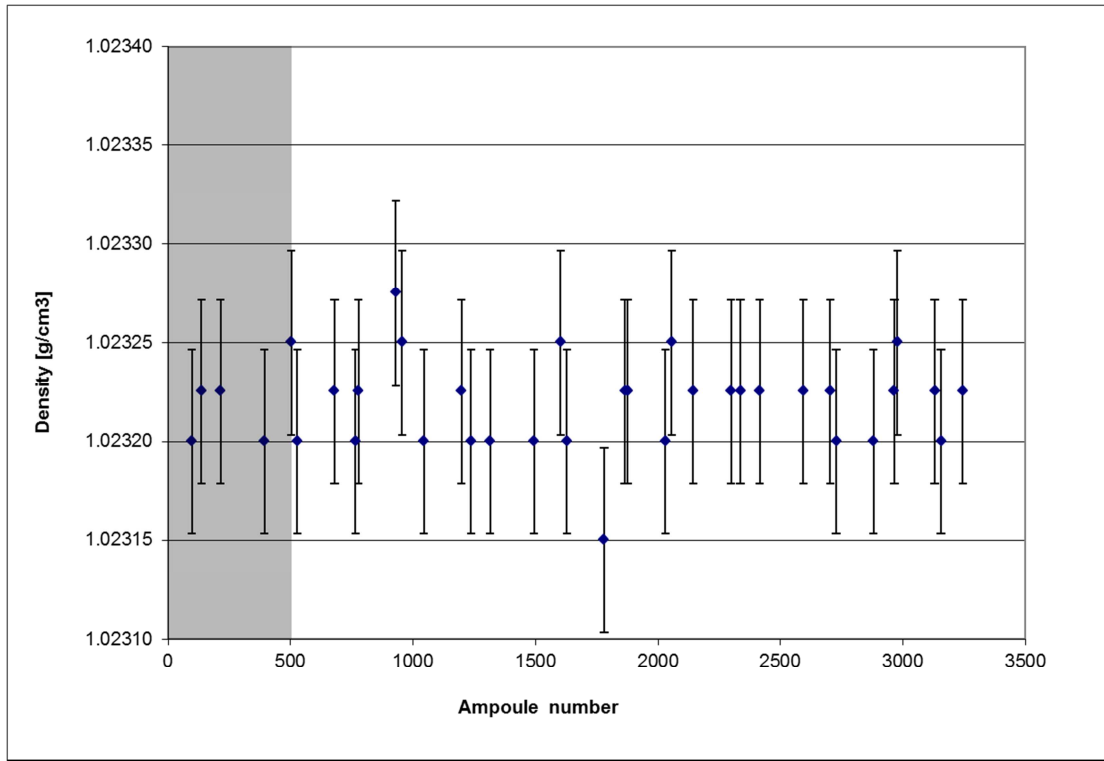**Annex C**: Results of the long-term stability measurements

Annex D: Description of the methods used for characterisation studies

**Annex E**: Results obtained in the characterisation studies

Annex A: Results of the homogeneity measurements: Data shown are the individual results for Hg as well as the mean values and their 95 % confidence interval for Hg and density. The mean values for Hg are based on the normalised and trend-corrected data. The 95 % confidence intervals are based on the within-ampoule standard deviation as obtained from one-way ANOVA.

Results shaded in grey are from ampoules with numbers below 500, which were later removed from the batch.

Annex B: Results of the short-term stability measurements: Data shown are the normalised mean values and their 95 % confidence interval for Hg mass fraction and seawater density. The 95 % confidence intervals are based on the within-time standard deviation as obtained from one-way ANOVA. The time points of the different temperatures of the short-term study of Hg were the same, but were separated in the graph to make the points clearly visible
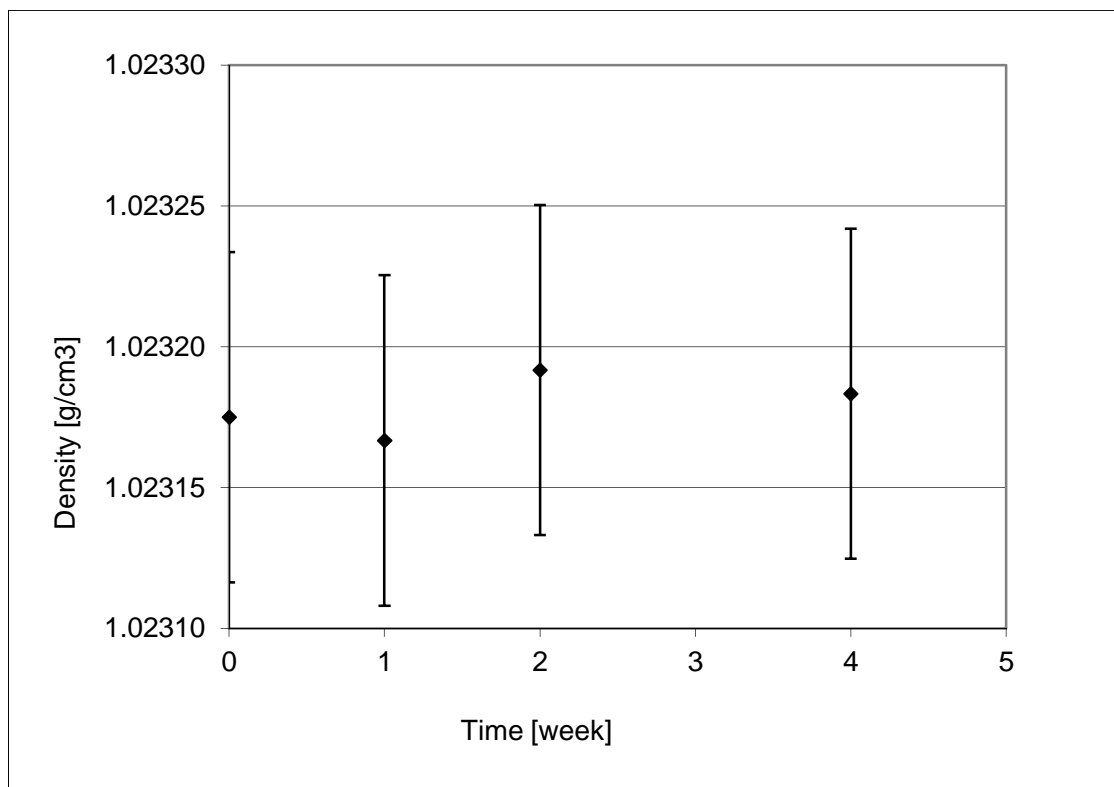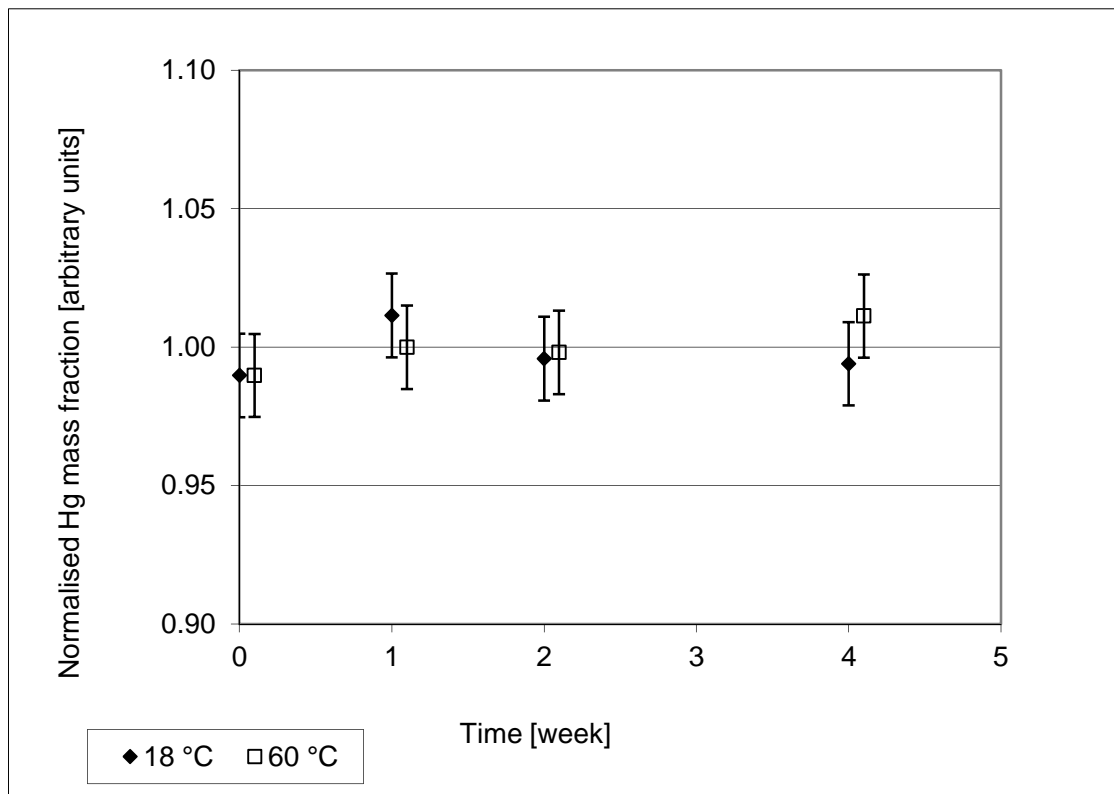
Annex C: Results of the long-term stability measurements: Data shown are the normalised mean values and their 95 % confidence interval for Hg mass fraction and seawater density. The 95 % confidence intervals are based on the standard deviation of time points as obtained from one-way ANOVA.
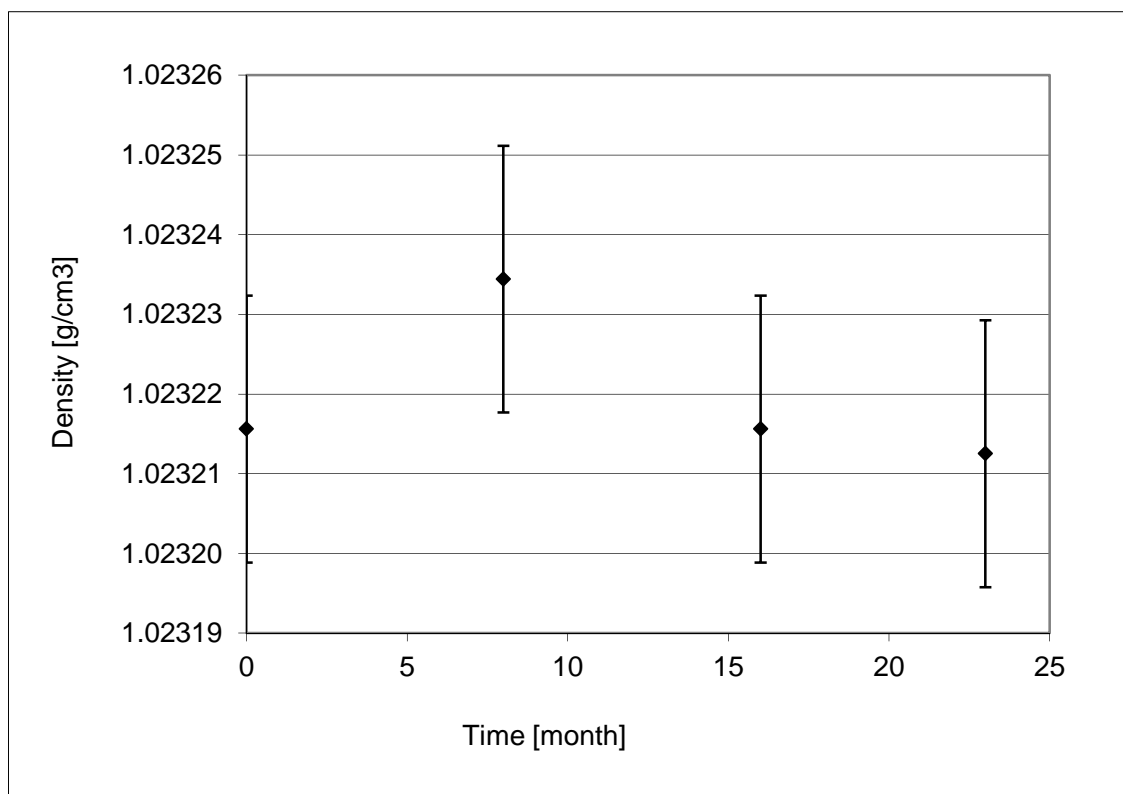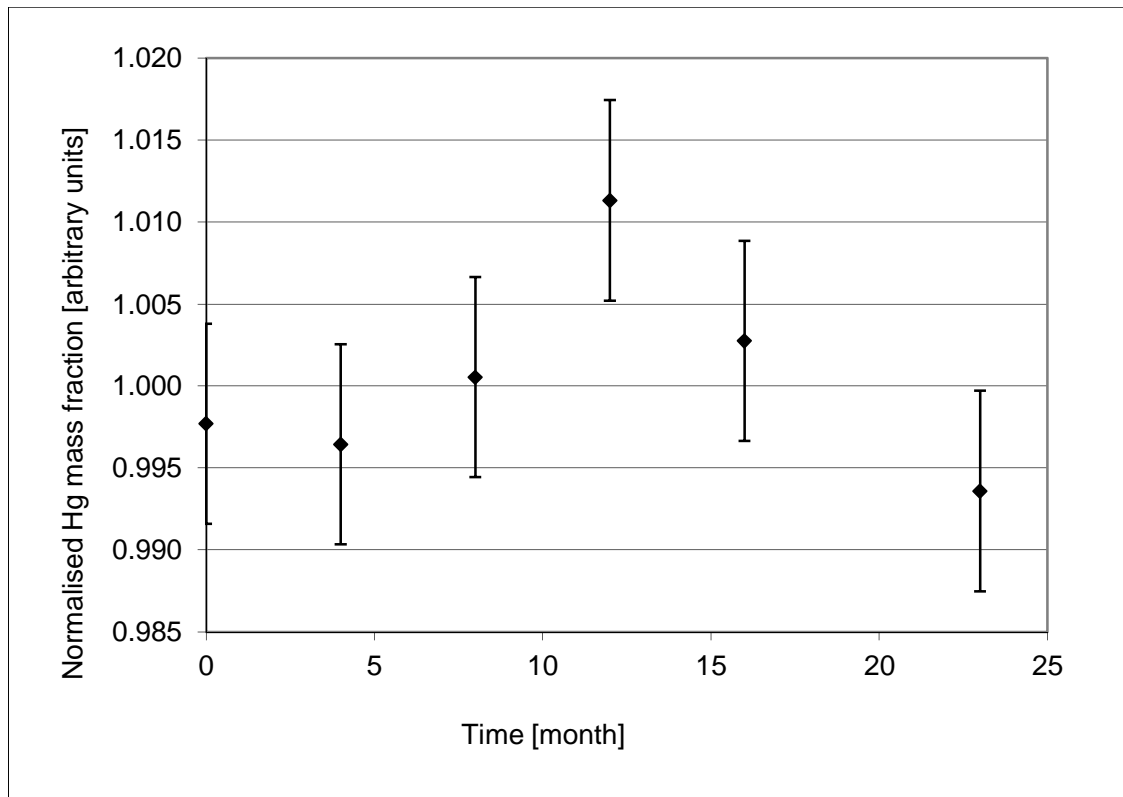
Annex D: Methods used in the characterisation study

D1 Hg mass concentration

Lab-method code	Sample pretreatment	Analytical method	Calibration	LOQ
L01-ID-CV-ICP-SFMS	5 g sample + 0.5 g enriched spike Of each ampoule, 3 sample preparations were made: two were measured as is; one was digested with 2 mL HNO3 in a microwave system	Double ID-MS with a sector field CV-ICP-MS in low- resolution mode with home-made CV module	Double ID-MS using $HgCl_2$ (Sigma), ERM-AE640 (^{202}Hg) Mass bias correction against ERM-AE639 (Hg isotopes)	0.7 ng/L
L02-CV-AFS	9.5 mL sample + 0.5 mL HCl + 0.2 mL $KBr/KBrO_3$ reagent; wait 30 min Immediately before measurements, add 0.025 mL $NH_2OH^.HCl$	CV-AFS	Single element standard from Ultra Scientific; traceability via NIST SRM 3133 Linear calibration	2 ng/L
L03-ID-CV-ICP-QMS	sample + HCl + $KBr/KBrO_3$ reagent; Digestion at 45 °C for at least 4 h	CV-ID-ICP-MS (online reduction of Hg(II) using $SnCl_2$)	IDMS using ^{199}Hg from ISC Science	0.2 ng/L
L04-CV-AFS	Oxidation with BrCl; then reduction with $NH_2OH^.HCl$ and SnCl2	CV-AFS with Au purge and trap	Single element standard from SCP Science; traceability via NIST SRM 3133	0.04 ng/L
L05-CV-AFS	25 g sample + 6 mL HCl s.o. (32 %) + 2 mL HNO_3 s.p. (65 %) digested at 155 °C for 20 min (see ISO 15587-1 Annex C)	CV-AVS with flow – injection system	ICP standard from Inorganic ventures; traceability via NIST SRMs	5.4 ng/L
L06-CV-AFS	100 mL sample + 0.5 mL BrCl; subsequent reduction with SnCl2	CV-AFS with purge and trap system	SRM 1641d	0.1 ng/L

D1 (continued)

| Lab-method code	Sample pretreatment	Analytical method	Calibration	LOQ
L07-CV-AFS	5 mL sample + 50 mL HCl s.p. (1 %)	CV-AFS with Au purge and trap system	Dissolution of high purity $HgCl_2$ obtained from Strem Chemicals	0.6 ng/L
L08-CV-AFS	39.4 mL sample + 0.1 mL HCl conc + 0.1 mL BrCl; reaction at room temperature under UV overnight Reduction with 40 μL 30 % NH_2OH immediately before analysis	CV-AFS with Au purge and trap system	NIST SRM 3133	0.82 nL
L09-CV-AFS	20 mL sample + BrCl digested for 1 h; excess BrCl is removed by addition of L-ascorbic acid	Dual Amalgamation CV-AFS	Hg standard from CPI; traceability via NIST SRM 3133	1 ng/L
L10-CV-AFS	EPA 1631: sample + BrCl; digestion at room temperature for 12 h Reduction with NH_2OH and $SnCl_2$	CV-AFS	NIST SRM 3133; NIST SRM1641d used as QC	0.51 ng/L
L11-CV-AFS	Sample + BrCl and removal of BrCl using L-ascorbic acid	CV-AFS with Au purge and trap	Standards from Merck; independent control from Spex	

D2 Density

Lab-method code	Sample pretreatment	Analytical method	Calibrant	Instrument
L13a-OT	Sample preparation was done in accordance with ASTM D4052-Standard Test Method for Density, Relative Density, and API Gravity of Liquids by Digital Density Meter	ASTM D4052: A small volume, approximately 1 mL to 2 mL, of liquid sample is introduced into an oscillating sample tube and the change in oscillating frequency caused by the change in the mass of the tube is used in conjunction with calibration data to determine the density of the sample. Each replicate result obtained is the mean of two injected volumes of the sample tested.	Re-boiled and distilled analytical grade H_2O Calibration points: 20°C	Anton Paar DMA 46
L13b-BP	Sample preparation was done in accordance with ASTM D1480-Standard Test Method for Density and Relative Density (Specific Gravity) of Viscous Materials by Bingham Pycnometer	ASTM D1480: The liquid sample is introduced into the pycnometer, equilibrated to the desired temperature, and weighed. The density is then calculated from this weight and the previously determined calibration factor, and a correction is applied for the buoyancy of air. Each replicate result is based on the mean of duplicate pycnometers testing the same fluid.	Re-boiled and distilled analytical grade H_2O Calibration points: 20°C, 25°C, 40°C, 50°C, 60°C, 70°C, 80°C	10 mL Bingham Rte
L14-OT	With a clean syringe (directly from a sealed package) a sample was taken directly from the bottle and immediately injected into the measuring cell. Between each measurement, the cell was cleaned and flushed	Oscillating tube density meter: Before each new measurement is conducted, a check with air and doubly distilled water is performed. Max. permissible deviation is 0.005 mg/ml (for water) and 0.05 mg/ml (air). After that, the cell is flushed two times with the sample before a measurement is done. After the readings, approx. 1 minute apart the cell is cleaned by flushing with doubly distilled water, acetone and drying with filtered air. Due to problems with evaporation and consequent changing of saline concentration (and density), samples are taken directly from the bottle each time, using a clean disposable syringe. The bottle is immediately closed, and the sample is injected.	Doubly distilled water freshly made prior to measurement	Anton Paar DMA 5000

Annex E: Results of the characterisation measurements
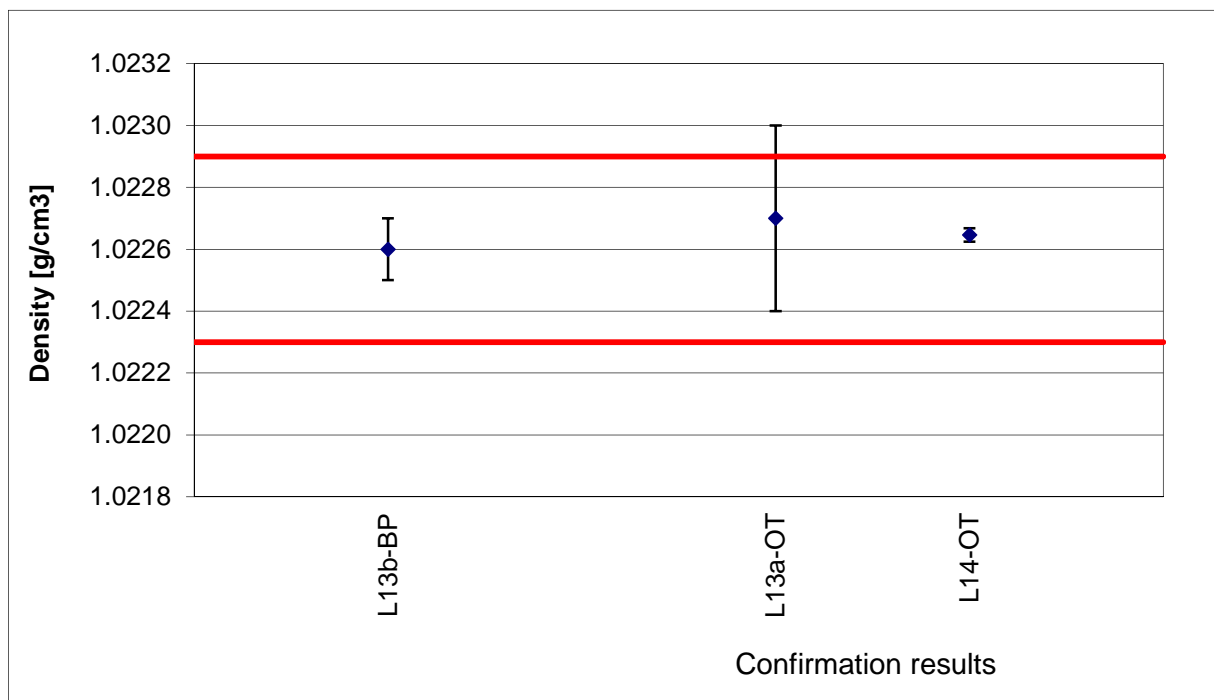
E1: Hg mass concentration

Laboratory	replicate 1 [ng/L]	replicate 2 [ng/L]	replicate 3 [ng/L]	replicate 4 [ng/L]	replicate 5 [ng/L]	replicate 6 [ng/L]	mean [ng/L]	Expanded uncertainty [ng/L]
L02-CV-AFS	16.4	16.3	16.2	16.4	16.2	16.3	16.3	2.6
L03-ID-CV-ICP-QMS	15.892	17.286	16.585	15.31	15.952	16.901	16.3	1.1
L04-CV-AFS	10.5*	17.5	17.3	16.1	17.5	16.5	15.9	2.1
L06-CV-AFS	16.4*	17.6	18.6	17.6	18.5	18.1	18.1	1.2
L10-CV-AFS	18.6	18.0	18.0	15.4	18.3	18.1	17.7	4.4
L11-CV-AFS	14.96	15.12	15.37	15.15	15.32	16.49	15.4	1.8
Results not used for certification								
L01-ID-CV-ICP-SFMS	6.7*	21.1	19.9	11.6*	20.1	20.0	20.3	1.3
L05-CV-AFS	11.9*	9.3	10.9	14.0	15.7	11.6	12.3	1.7
L07-CV-AFS	14.9*	15.1	17.3	Sample lost	16.2	16.4	16.3	0.3
L08-CV-AFS	11.4*	16.0	15.6	14.5*	15.2	15.0	15.5	1.0
L09-CV-AFS	17.5*	17.1	15.6	7.6*	18.9	18.7	17.6	1.4

* Result not used as it came from an ampoule with a number below 500

shaded in grey were not used for certification
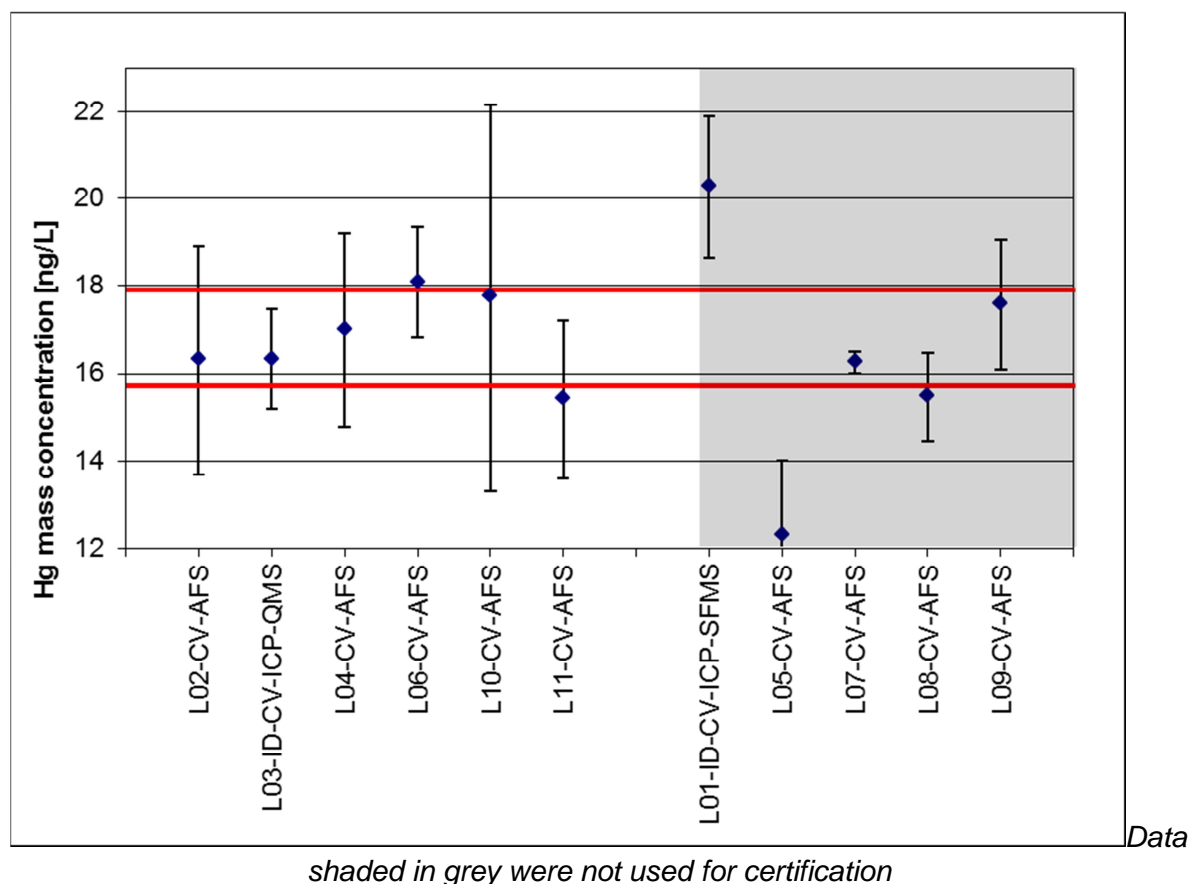
E2: Density

Laboratory	replicate 1 [g/cm^3]	replicate 2 [g/cm^3]	replicate 3 [g/cm^3]	replicate 4 [g/cm^3]	replicate 5 [g/cm^3]	replicate 6 [g/cm^3]	Mean [g/cm^3]	Expanded uncertainty [g/cm^3]
L13b-BP	1.0226	1.0226	1.0226	1.0226	1.0226	1.0226	1.0226	0.0001
Confirmatory results from electronic density meters								
L13a-OT	1.0227	1.0227	1.0227	1.0227	1.0227	1.0227	1.0227	0.0003
L14-OT	1.02265	1.022647	1.02264	1.02263	1.02265	1.02266	1.0226	0.000022

Data shaded in grey were not used for the calculation of the certified value

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security, including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

LA-NA- 28264-EN-N

Publications Office