2015-09-18

Combined uncertainty estimation for the determination of the dissolved iron amount content in seawater using flow injection with chemiluminescence detection

Floor, GH

10.1002/lom3.10057
LIMNOLOGY AND OCEANOGRAPHY-METHODS
Wiley

Disclaimer: This is a pre-publication version. Readers are recommended to consult the full published version for accuracy and citation. Published in Limnology and Oceanography: Methods, 13, 673–686 (2015), doi: 10.1002/lom3.10057.

Combined uncertainty estimation for the determination of the dissolved iron amount content in seawater using flow injection with chemiluminescence detection.

G. H. Floor[1a], R. Clough[2], M. C. Lohan[2], S. J. Ussher[2], P. J. Worsfold[2], C. R. Quétel[1]*.

1. Institute for Reference Materials and Measurements, Joint Research Centre–European Commission, 111 Retieseweg, 2440 Geel, Belgium.

2. Biogeochemistry Research Centre, School of Geography, Earth and Environmental Sciences, Plymouth University, Plymouth, PL4 8AA, United Kingdom

a: Currently at GFZ German Research Centre for Geosciences, Helmholtz Centre Potsdam, Telegrafenberg, 14473 Potsdam, Germany

* Corresponding author: Christophe.Quetel@ec.europa.eu

Acknowledgements

This work was financially supported by the EMRP via JRP-ENV05 (Metrology for ocean salinity and acidity, GF) and JRP-ENV05-REG1 (RC). The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union. ML was supported by a Natural Environment Research Council (NERC) grant NE/H004475/1. SU was supported by a Marie Curie Career Integration Grant (PCIG-GA-2012-333143 DISCOSAT) from the European Commission. GF acknowledges J. Snell and T. Linsinger for fruitful discussions.

Abstract

This work assesses the components contributing to the combined uncertainty budget associated with the measurement of the Fe amount content by flow injection chemiluminescence (FI-CL) in <0.2 µm filtered and acidified seawater samples. Amounts of loaded standard solutions and samples were determined gravimetrically by differential weighing. Up to 5% variations in the loaded masses were observed during measurements, in contradiction to the usual assumptions made when operating under constant loading time conditions. Hence signal intensities (V) were normalised to the loaded mass and plots of average normalized intensities (in V kg^{-1}) versus values of the Fe amount content (in nmol kg^{-1}) added to a 'low level' iron seawater matrix were used to produce the calibration graphs. The measurement procedure implemented and the uncertainty estimation process developed were validated from the agreement obtained with consensus values for three SAFe and GEOTRACES reference materials (D2, GS and GD). Relative expanded uncertainties for peak height and peak area based results were estimated to be around 12% and 10% (k=2) respectively. The most important contributory factors were the uncertainty on the sensitivity coefficient (i.e. calibration slope) and within-sequence-stability (i.e. the signal stability measured over several hours of operation; in this case 32 h). Therefore, an uncertainty estimation based on the intensity repeatability alone, as is often done in FI-CL studies, is not a realistic estimation of the overall uncertainty of the procedure.

Introduction

The ocean acts as both a sink and a source for carbon dioxide and plays an important role in regulating the global climate system (Boyd and Elwood, 2010). The dynamics of the ocean and its interaction with the atmosphere are strongly linked to the properties of seawater. Elements such as Fe limit marine primary production in approximately one third of the world's oceans (Ussher et al., 2013) and thus may have a profound effect on plankton communities and the global carbon cycle (Martin and Fitzwater, 1988; Mills et al., 2004). More reliable determinations of micronutrient elements in marine waters are thus essential to enhance our understanding of their impact on ocean productivity and processes (e.g. ocean acidification). Therefore, robust and fully validated measurement procedures are necessary, accompanied by an estimation of the overall uncertainty budget.

The international standard ISO/IEC 17025 (2005) states that the performance of a measurement procedure should be evaluated based on one or a combination of the following approaches: a) the use of reference materials, b) the comparison of results achieved with other methods, c) inter-laboratory comparison, d) systematic assessments of the factors influencing the result and e) the assessment of the uncertainty of the results. The Fe content of commercially available certified reference materials is at least one order of magnitude higher than most open ocean waters and are thus of limited use for method development. Therefore, test materials from inter-laboratory comparison exercises are often used instead, e.g. those collected as part of the IRONAGES, SAFe and GEOTRACES studies. However, Bowie et al. (2006) observed that discrepancies between results obtained in different laboratories during the IRONAGES comparison remained too large (e.g. up to 59% variability when using the same procedure) and differed significantly at the 95 % confidence level. Factors thought to explain these results included: (1) variations in the efficiency of the extraction of iron from the matrix during pre-concentration (resulting in different procedures measuring different fractions of iron), (2) errors in the quantification of the analytical blank, (3) inaccuracies in the system calibration and (4) underestimation of the stated uncertainty (Bowie et al, 2003; Petrov et al. 2007). Hence iron data from these exercises for the same water mass were distinctly inconsistent. Points (1) and (2) have been addressed by the SAFe (Johnson et al., 2007) and GEOTRACES (GEOTRACES, 2013) exercises but not points (3) and (4). It is thus useful to revisit these two factors and determine how realistic uncertainties can be estimated for the most commonly applied measurement procedures (particularly shipboard procedures) (see also Ussher et al., 2010b). In this respect flow injection with chemiluminescence detection (FI-CL) was chosen for this study as it is a technique that allows high temporal and spatial resolution measurements at sea without the need for sample storage and transport.

According to the international nomenclature, the measurement uncertainty is a "*non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used*" (JCGM 200, 2008). The basic purpose of an uncertainty statement is to propose a

range of possible 'true' values. There are various ways of estimating uncertainties. For instance, combined uncertainty estimates can be based on data obtained by inter-laboratory or intra-laboratory studies (see e.g. Analytical Methods Committee, 1995; Nordic Committee on Food Analysis, 1997). The uncertainty estimation proposed in the Guide for Uncertainty in Measurements (GUM) is based on combining the contributions of all known sources of uncertainty (JCGM 100, 2008). In this approach, the measurement procedure is described by a mathematical model and the values and associated standard uncertainties of the different components (the input quantities) in the model must be established. The model and input data are then used to calculate the measurement result including its associated combined uncertainty.

The aim of this work was to study the application of the 'GUM approach' to the FI-CL measurement procedure. The specific objectives were to; (1) propose a set of mathematical equations (a model) describing this measurement process and allowing the estimation of a measurement uncertainty, (2) discuss the best way to assess the uncertainties of the different components in the model, (3) apply this uncertainty model to present the measurement results with their estimated combined uncertainties obtained for seawater samples from the SAFe and GEOTRACES campaigns (Lohan et al., 2006; Johnson et al., 2007) and, from the above, (4) propose a simplified equation to estimate the measurement uncertainty.

Materials and procedures

Reagents, materials and samples

Concentrated hydrochloric acid (HCl), ammonia (NH_3, 20 – 22%) and glacial acetic acid (CH_3CO_2H), all SpA grade, were purchased from Romil (Cambridge, UK). Hydrogen peroxide, Merck Suprapur grade was obtained from VWR (Lutterworth UK). Luminol (5-amino-2,3-dihydro-1,4-phthalazinedione), sodium carbonate and triethylenetetramine (TETA) were purchased from Sigma Aldrich (Gillingham, Dorset, UK). All high purity water (HPW), 18.2 MΩ·cm, was drawn from an ElgaStat Maxima system (Marlow, UK). All weighing was performed using an analytical balance (OH1602/C, Ohaus, Thetford, UK). The accuracy of the balance was checked daily before use using F1 Class certified weights (KERN, Albstadt, Germany). All facilities were managed under ISO 9001:2008 certification.

To ensure low blank Fe amount content all sample and reagent handling was undertaken in an ISO 14644-1 Class 5 laminar flow hood (Bassaire, Southampton, UK) situated within an ISO 14644-1 Class 5 clean room. Reagent and sample containers were made of low density polyethylene (LDPE; Nalgene, Fisher Scientific, UK) and were cleaned using established cleaning protocols for trace metals. Containers were immersed in ~ 1.1 M trace metal grade HCl (Fisher Scientific) for at least seven days. Subsequently,

the containers were rinsed in copious amounts of HPW, filled with 0.01 M HCl and stored in double re-sealable plastic bags until use.

The main characteristics of the seawater samples used for this project are described in Table 1. Briefly, all samples were filtered at sea and then acidified either at sea or at Plymouth University (PU). Seawater samples, referred to as SWA, SWB, and SWC, containing ≤0.5 nmol kg^{-1} Fe were selected to prepare three different sets of calibration standards, by addition of controlled amounts of iron from a CPI International (Amsterdam, Netherlands) ICP-MS standard containing 0.17 mol kg^{-1} Fe. Experiments in this work were carried out with 0.5 L reference samples from large volumes of homogenised, bulk seawater samples (SAFe D2 and GEOTRACES GS and GD consensus mean reference materials). More details regarding the sampling, pre-treatment and bottling procedures for these materials can be found elsewhere (Jonhson et al., 2007; GEOTRACES 2013).

The FI-CL based measurement procedure

Figure 1 describes the FI-CL manifold used for these experiments. It consists of three peristaltic pumps (Minipuls 3, Gilson, Luton, UK), one PTFE manually operated three port valve (Valve 1; Omnifit), one three port solenoid valve (Valve 2), one two-way six port electronically actuated valve (Valve 3; VICI, Valco Instruments, Schenkon, Switzerland), a thermostatic water bath (Gran, Cambridge, UK) and a photomultiplier tube (PMT; Hamamatsu H 6240-01, Hamamatsu Photonics, Welwyn Garden City, UK) containing a coiled, transparent PVC flow cell (volume 40 µL). The peristaltic pump tubing used was two stop accu-rated™ PVC (Elkay, Basingstoke, UK) and all other manifold tubing was 0.8 mm i.d. PTFE. The system used two poly(methyl methacrylate) columns (1 cm long, 1.5 mm i.d., volume 70 µL), loaded with Toyopearl AF Chelate 650 resin (Tosoh Bioscience, Stuttgart, Germany) retained with HDPE frits (BioVion F, 0.75 mm thick, 22-57 µm pore size), to clean up the buffer and column rinse solutions. The analytical column, also loaded with Toyopearl AF Chelate 650 resin, was made of polyethylene with LDPE frits with an internal volume of 200 µL (Global FIA, Fox Island, USA).

Peristaltic pump and valve control and data acquisition were performed using custom built hardware and software (Ruthern Instruments, Bodmin, UK) run under Labview v 7.1 (National Instruments, Newbury, UK). The measurement procedure, based on Obata et al. (1993), was as follows. A working solution of approximately 0.35 µmol kg^{-1} Fe was prepared gravimetrically by serial dilution of the CPI International stock solution. This working solution was then used to gravimetrically prepare calibration standards and achieve added levels ranging from 0.15 to 0.9 nmol kg^{-1} Fe in 0.15 nmol kg^{-1} increments. All calibration standards were prepared at least 12 h before use to allow for complete equilibration of the added Fe with that present in the calibration seawater. A 20 µL aliquot of a 10 mM H_2O_2 solution was added to all calibration standards at least 2 h before use, to ensure that all Fe present was as Fe(III) (Lohan et al., 2006). The following solutions were also prepared at least 12 h before use. A 48 mM stock solution of

147 luminol was obtained by dissolving 0.177 g of luminol and 0.25 g of Na_2CO_3 in 20 mL of HPW. This stock
148 was then diluted to give a 0.24 mM working solution. The post column reagents for the
149 chemiluminescence reaction was a mixture of 0.23 M HCl, 0.44 M NH_3, 0.24 mM luminol / 0.46 mM TETA
150 and 0.31 M H_2O_2. The acidified reference samples and standards of seawater were buffered to pH 3.5
151 with 0.35 M CH_3CO_2H and 0.11 M NH_3. To precondition and wash the column, 0.011 M HCl was used.
152 To operate the FI-CL instrument, the LabVIEW software was opened and the baseline signal from the
153 PMT monitored to check for stability. The pump controlling the eluent and post-column reagents was
154 then activated and the baseline chemiluminescence signal recorded after the signal had stabilised. Each
155 analytical session started with the measurement of a procedural blank (by application of the "closed
156 sample valve" method). For this, the sample flow was stopped, by closing one port on valve 1, so that
157 only the wash solution and ammonium acetate buffer passed over the column. The FI-CL system was
158 then operated by loading and injecting SWA for at least 30 min to monitor stability. Subsequently,
159 calibration seawater standards and samples were analysed. The FI-Cl manifold was fully automated and
160 one replicate measurement consisted of the following analytical cycle. The column was conditioned for 10
161 s with 0.011 M HCl. Then the sample and buffer were loaded simultaneously for 60 s. The column was
162 washed with 0.011 M HCl for 20 s. The Fe was then eluted with 0.23 M HCl for 120 s. The mass of loaded
163 sample or standard solution was gravimetrically determined for each replicate by differential weighing.
164 Between each sample the sample flow path was washed with HPW for 30 s followed by uptake of the
165 fresh sample for 180 s. After each analytical session all fluid paths were flushed with 0.01 M HCl for 10
166 min and then with HPW for 15 min and HPW was left in the lines.
167
168 ***Data treatment***
169 Data integration was also performed with the custom build software run in LabVIEW. The baseline, and
170 the start and end points of the peak were set manually for each transient signal. The main calculations in
171 this study were carried out on the basis of peak height data, as this is the commonly used practice for FI-
172 CL measurements in the oceanographic community (and the wider FI community). Peak area
173 measurements were also made and some of the differences observed when using peak areas are
174 discussed below. Further data treatment, including calculations for the estimation of standard
175 uncertainties, was carried out in Excel®. The combined uncertainties were obtained by propagating
176 together individual uncertainty components according to the GUM (JCGM 100, 2008). In practice, a
177 dedicated software program was used (Metrodata GmbH, 2003). The reported combined uncertainties
178 are expanded uncertainties and reported as $U = k u_c$ where u_c is the combined standard uncertainty and k
179 is a coverage factor equal to 2. If "*the probability distribution characterized by y and $u_c(y)$ is*
180 *approximately normal and the effective degrees of freedom of $u_c(y)$ is of significant size*" ("*greater than*

181 *10*"), "*taking k=2 produces an interval having a level of confidence of approximately 95 %*" (JCGM 100,
182 2008).
183
184 **Assessment**
185 ***Definition of the measurand***
186 The GUM states that a measurement begins with an appropriate specification of the measurand, the
187 particular quantity intended to be measured (JCGM 100, 2008). Iron exists in different physico-chemical
188 forms in seawater. Traditionally, filtration is performed to differentiate between the different physical size
189 fractions (Ussher et al., 2004, 2010a, Wu et al., 2001). Additionally, iron occurs in two oxidation states;
190 Fe(II) and Fe(III). Generally, Fe(III) predominates in oxygenated waters, of which most (80–99%) is
191 strongly complexed by organic ligands (Achterberg et al., 2001; Mawji et all, 2008; Gledhill and Buck,
192 2012). In this study the measurand is the amount content of Fe present in <0.2 µm filtered and acidified
193 samples and is regarded as the dissolved fraction of the Fe present in the seawaters. The aim was to
194 obtain the Fe amount content in specific samples and therefore the uncertainties associated with the
195 sampling process and/or the sample conditioning phase have not been considered.
196
197 ***Experimental design***
198 Three different types of experiment were performed. Firstly, the stability of the analytical procedure was
199 checked with 5 measurements (6 replicates of each) performed over a period of 32 h for SWC (with and
200 without H_2O_2 addition) and a procedural blank (this was termed the "stability experiment"). Secondly, the
201 effect of small variations in the matrix was investigated by comparing the sensitivity achieved for the
202 three different seawaters (Table 1). On the first day, SWA was compared with SWB while on the second
203 day SWA was compared with SWC ("matrix experiment"). Thirdly, the FI-CL based procedure was applied
204 to the determination of iron in samples of three filtered and acidified seawater reference materials using
205 SWA for calibration ("sample experiment").
206
207 ***Calculating the dissolved Fe amount content in the samples and mathematical description of***
208 ***the measurement procedure***
209 Implicit in the GUM "*is the assumption that a measurement can be modelled mathematically to the*
210 *degree imposed by the required accuracy of the measurement*" (JCGM 100, 2008). A measurand Y is
211 determined from various input quantities X_i through a functional relationship. These input quantities "*may*
212 *themselves be viewed as measurands and may themselves depend on other quantities, including*
213 *corrections and correction factors*" "*that can contribute a significant component of uncertainty to the*
214 *result of the measurement*" (JCGM 100, 2008). A mathematical description of the FI-CL measurement
215 procedure is given through equations 1 to 5 described in Table 2. The main equation in this procedure is

the calculation of the dissolved Fe amount content in a sample by dividing the blank corrected sample intensity by the sensitivity of the system (Equation 1 in Table 2). The way the equations controlling these three input parameters were established is discussed below.

Mass normalisation of the measurement signal

In most flow analysis methods incorporating a pre-concentration column, the amount of sample loaded is assumed to remain the same for constant loading times and the resulting peak height signals (expressed in V) are used for the calculations. Variations in the loaded mass are thus not corrected for. However, this was found to be an issue as variations in sample mass were observed to be significant during the "stability experiment", with about 5% decrease in the sample loaded from the first to the last measurement (data not shown). During the "sample experiment" the average loaded mass for samples was lower than for the standards, probably due to the fact that the samples were all run at the end of the sequence and were therefore more likely to be affected by wear on the pump tubing, increased back pressure on the analytical column and / or changes in the relative flow rates of the sample and buffer lines (Figure 2). These results show the importance of weighing the amount of seawater loaded each time and of normalising the peak signal (symbol I, in V) to the loaded mass (in kg). In addition, gravimetric measurement, coupled with calibration of the analytical balance, provides tighter traceability to SI (the kg) of the amounts of loaded samples than loading by volumetric means.

As a result of this finding, mass normalised signals (symbol J, in V kg^{-1}) were used throughout this study for the calculations (Equation 2b, Table 2). Following the example given in Quétel et al. (2001), in equations 2a, 3a and 4a unity multiplicative factors were introduced to carry standard uncertainties associated with signal stability, data integration and matrix effects. Since these unity factors do not influence the final results, but enable the propagation of sources of uncertainty, they are discussed in more detail below.

Blank corrections

Assessment of overall blank levels that reflect the reality of sample contamination during the measurement procedure is necessary. In the international inter-laboratory comparison exercise IRONAGES, blanks were reported to range between 6 and 290% of the Fe content in the seawater sample (Petrov et al., 2007). Moreover, participants had diverse ways of defining and assessing their blanks (Bowie et al., 2006) and were, therefore, possibly overlooking different aspects of the contamination process. Sources of contamination during FI-CL measurements include the Fe present in reagents (i.e. the added H_2O_2, the buffer and wash solutions and the chemiluminescence reagents) and Fe leaching from laboratory ware and parts of the experimental set-up. Sample manipulations could also be a major contributor to the analytical blank as was shown to be the case by Petrov et al. (2007) during

isotope dilution inductively coupled plasma mass spectrometry measurements using co-precipitation with magnesium hydroxide for sample preparation. The Fe from the reagents of the chemiluminescence reaction is normally included in the baseline. Baseline subtraction for the determination of net peak height or peak area signals should therefore remove this possible bias. The influence of additions of chemical reagents for the purpose of preserving and/or conditioning the samples (e.g. acid, H_2O_2) can be assessed using double spiking of the reagents. Previous studies using FI-CL have shown their contribution to be low / negligible if care is taken to select high purity reagents (Bowie et al., 2003: Bowie et al., 2004; Klunder et al. 2010).

Descriptions of what a blank may represent are available from the International Union of Pure and Applied Chemistry (IUPAC). A "procedural blank" is "*where the analytical procedure is executed in all respects apart from the addition of the test portion"* (McNaught and Wilkinson, 1997; Inczedy et al., 1998). Alternative measurement procedures for blank determination, such as the field blank approach or varying sample loading times and extrapolating back to time zero (Bowie et al. 2004), were not suitable as the former requires a matrix containing no analyte and the latter only accounts for reagents that are loaded for a constant time e.g. the wash solution, but not those for which the amount loaded is variable e.g. the pH adjustment solution.

In FI-CL, the signal obtained with the "closed sample valve" method as described above, i.e. loading only buffer (Bowie et al., 2004; Ussher et al., 2010a), can be considered as a procedural blank. This method was applied as no better alternatives could be found for estimating the level of contamination. The risk that matrix effects and pH changes could influence final results due to fluctuations in the blank values determined in this way is discussed below. Normalised signal intensities were calculated by division by the average loaded sample mass (equation 3b). These blank values were 50-100 times lower than the signals for the seawater samples. Unity multiplicative correction factors were used to propagate uncertainties on stability and matrix effects (equation 3a) and are discussed in more detail below.

Calculation of the calibration slope

The FI-CL method has a different sensitivity for seawater than for ultra-pure water because of matrix related effects (Bucciarelli et al., 2001). Thus, a common approach for the calibration under matrix-matching conditions is to use a low level Fe seawater and fortify it with increasing amounts of Fe (Bucciarelli et al., 2001; Bowie et al., 2004; Ussher et al., 2010a; Klunder et al., 2011). In this work, in addition to the low level seawater alone (termed the 'zero' standard), six calibration standards were prepared with Fe amount content ranging from 0.15 to 0.9 nmol kg^{-1}. Since measurements were repeated 6 times for each calibration point, a total of 7 x 6 = 42 results were obtained. A linear regression was plotted, with the masses of Fe loaded (in kg, obtained by multiplication of the standard Fe mass fraction by the loaded mass of the replicate) on the x axis and the corresponding measured signal

intensities (in V) on the y axis. The 'behaviour' of the data was nearly the same irrespective of the scale of observation, with replicate results spread randomly around the regression graph in more or less the same way for all 6 standards prepared and tested. Common practice is to produce 3-4 replicates per Fe level and work with average values. Thus, a more practical way of establishing the calibration curve consists of plotting a linear regression between the group of 6+1 Fe amount content (C, in nmol kg^{-1}) on the x axis and the corresponding average normalized intensities (J, in V kg^{-1}) on the y axis. The sensitivity coefficient (F, in V $nmol^{-1}$), i.e. the slope, is obtained using equation 4b from Table 2 (see Figure 3). Weighted regression can also be performed but the calculations are more complex. In a weighted regression the higher the uncertainty on a y value the smaller the contribution of the y value to the regression slope. This is especially important if the increase of values on the x axis can be related to an increase of the standard uncertainty on corresponding values on the y axis. There was no difference with this dataset at the 95% confidence level between weighted and unweighted regressions. This is probably because the increase in the standard uncertainty with increased normalised intensity is limited. The comparison between these two approaches is further discussed in the next section.

Assessing the standard uncertainties

Individual uncertainty components and the factors influencing their standard uncertainties were evaluated. This is necessary to enable a combined uncertainty estimation of the Fe amount content results.

Uncertainty on mass normalised measurement signals

During the "sample experiment", the repeatability (short term signal stability) of mass-normalised intensities (peak height based signals) for one measurement varied between 1.9 and 4.0% RSD (relative standard deviation, n=6) while for the "stability experiment" repeatabilities varied between 2.4 and 4.9 %. These variations in RSD cannot be explained by variations in the specific characteristics of the sample replicates since the same solution was measured throughout the "stability experiment". Moreover, as illustrated in Figure 4, there was also a longer term variability component involved (within-sequence-stability), and thus two sources of instability influencing the intensity values. Over the 32 h long analytical sequence there was no clear trend, and as a result correction for drift was not possible. Therefore, the approach proposed is to estimate typical values for both components from the outcome of an ANOVA analysis and multiply the sample average mass normalised intensities by unity correction factors carrying the uncertainty for these two components (δ_{rep_S} and δ_{stab_S}). Repeated intensity values per sample and average intensity values from replicate samples were approximately normally distributed. The intensity repeatability and the within-sequence-stability, determined using data from the "stability experiment", were 4.1% and 6.3% respectively. Assuming independence between the intensity values used to

calculate both types of standard deviation, uncertainty estimations were carried out using these standard deviations divided by square root 6, i.e. the number of replicates and square root 5, i.e. the number of repeat measurements analysed in each case, to give values of 1.7% and 2.8% respectively.

Sample loading and standard preparation cannot be performed gravimetrically on board ship and therefore this is done volumetrically, which may cause additional sources of uncertainty. In this case, the set of equations described in Table 2 will change slightly and result in equation 6 as described below:

$$C_S = \frac{\overline{I}_{R\_S} \cdot \delta_{rep\_S} \cdot \delta_{stab\_S} \cdot \delta_{\mathrm{WtoV\_S}} - \overline{I}_{R\_B} \cdot \delta_{stab\_B} \cdot \delta_{rep\_B} \cdot \delta_{matrix\_B}}{F_{reg} \cdot \delta_{matrix\_std}} \qquad \text{equation 6}$$

As a consequence of not using mass normalization, the sensitivity factor is determined by regression of the intensity (expressed in V) with the concentration (nmol L^{-1}) and has the units V/nmol L^{-1}. Secondly, an extra unity multiplicative correction factor (δ_{WtoV_S}) was introduced to take account of the difference in the mass loading between samples and standards (Figure 2). Using this data set and assuming constant loading (i.e. without mass normalisation) its contribution to the final uncertainty budget was a few percent. Lastly, although the same approach can be used to quantify the uncertainty on the unity multiplicative factors of the intensity repeatability and within-sequence-stability, the values will be higher than in the case of mass normalization. It must be noted that the within-sequence-stability during on-board measurements might be different than in controlled laboratory conditions, but a specific assessment was not possible within the time frame of this study.

Uncertainty on blank corrections

The evaluation of the uncertainty on blank measurement signals was approached in a similar way as for the sample measurement signals. ANOVA analysis of the "stability experiment" results indicated 6.9% and 10% respectively for the intensity repeatability (n=5) and the within-sequence-stability. A unity multiplicative factor δ_{matrix_B} with a value of 1±0.2 was conservatively applied in equation 3a to account for the matrix differences between the blank samples and the standards used for calibration purposes. However, since the signal intensity for the analytical blank was about 50-100 lower than the intensity for the seawater samples in this project, this source of uncertainty on the blank correction had no influence on the combined uncertainties estimated for the Fe amount content in the samples investigated.

Uncertainty on the calibration slope

As discussed above, there are different statistical approaches that can be used to calculate the slope of the regression line (Miller, 1991: Press et al., 2012). Values obtained using different regression approaches are not significantly different at the 95% confident interval, but associated standard

uncertainties do vary (Table 3). The standard uncertainty on the slope when using average normalised intensity values is the same whether the regression is weighted or unweighted. It is lower when using all individual data in the unweighted regression because there are more data points that follow a normal distribution. The importance of the number of standards and replicates on the size of the estimated standard uncertainty of the slope was studied. In Table 4 it can be seen that the number of standards used is a more important criterion than the number of replicates, but nevertheless the uncertainty on the sensitivity factor also improves using 6 rather than 3 replicates. Small matrix differences between the three seawaters tested in the "matrix experiment" did not lead to significant differences between the slopes obtained for SWA, SWB and SWC. Therefore, no uncertainty factor for differences in the calibrant matrix was applied.

Discussion

Application to seawater samples from the SAFe and GEOTRACES campaigns

Since consensus values are available for the Fe amount content in samples from the SAFe and GEOTRACES campaigns (GEOTRACES, 2013), these data were compared with results obtained by application of the model for combined uncertainty estimation and the calculations described above. Samples D2, GS and GD were analyzed using 6 replicates each time, the "closed sample valve" approach for blank assessment and a least square regression calibration line with 7 levels (no Fe added + 6 levels of added Fe) in SWA. This was the "sample experiment", and results obtained are reported in Table 5. Estimated expanded (coverage factor k=2) relative combined uncertainties were around 12% on a peak height basis, and around 10% on a peak area basis. Using this dataset, the combined uncertainty was slightly higher using volumetric loading compared with gravimetric loading. For example, for sample GD the combined expanded uncertainty increased from 12 to 13% for peak height integration. It can be seen that both peak height and peak area based results are systematically lower than the consensus values. Results obtained for GS and GD (peak height and peak area basis) and peak area results for D2 were in agreement with consensus values within uncertainty statements. These conclusions were reached from the observation that the expanded combined uncertainty (k=2) on the difference between a measured and the corresponding consensus value was greater than the difference itself in all cases (calculations according to a methodology reported in Linsinger, 2010). For the peak height results for the D2 sample, the expanded uncertainty on the difference was smaller than the difference itself but only by less than 3%. These results validate the measurement procedure implemented and the uncertainty estimation process developed. They nevertheless point to the presence of a systematic effect not yet (sufficiently) corrected for.

An overview of the values of the input parameters and their associated standard uncertainties for these experiments is given in supplementary Table S1. The relative contributions of the different input parameters to the uncertainty budget are given for sample GD in Table 6 as an example. The normalised signal intensity repeatability accounts for only 7.9% of the total uncertainty. The within-sequence-stability (assessed over 32 h) and the uncertainty on the sensitivity coefficient (calibration slope) are the most important contributors to the combined uncertainty with relative contributions of 21.6 and 69.7%. Therefore, it is beneficial to have a low uncertainty on the calibration slope. For this reason, it is favourable to use sufficient replicates (6) and number of standards (at least the non-spiked standards and 5 spiked levels). Moreover, correctly estimating the within-sequence-stability is key and should be done under the same measurement conditions as for the samples.

Results obtained indicate that an uncertainty estimation based on the signal repeatability alone, as is often done in FI-CL studies, is not a realistic estimation of the overall uncertainty of the procedure. However, taking into account only the major contributions, the combined expanded uncertainty could be approximated using equation 7:

$$U_{C_s} \approx 2 \cdot C_S \sqrt{\frac{\overline{J}_S{}^2 \cdot \left[\left(\frac{u_{\delta_{rep\_S}}}{\delta_{rep\_S}} \right)^2 + \left(\frac{u_{\delta_{\text{stab}\_S}}}{\delta_{\text{stab}\_S}} \right)^2 \right]}{\left(\overline{J}_S - \overline{J}_B \right)^2} + \left(\frac{u_F}{F} \right)^2} \qquad \text{equation 7}$$

In this, the standard uncertainty on the intensity repeatability and within-sequence-stability can be assessed using ANOVA analyses of repeat measurements of the same solution. The uncertainty on the calibration slope can be obtained using statistical tools. This simplified approach assumes that the blank does not significantly contribute to the uncertainty and should therefore have a much lower intensity compared with the sample (as was the case in this study). When using data from this project the uncertainty obtained with equation 7 was nearly identical to the uncertainty calculated above (for example the difference was less than 0.2% for GD using peak height data). Therefore, if the assumptions are valid this simplified approach provides a realistic uncertainty estimate.

Peak area versus peak height

The bias between results and consensus values was around -12% for D2 and GS and -20% for GD, on a peak height basis, and around -8% for D2 and GS and -16% for GD, on a peak area basis. This also means that peak height results were systematically lower than the peak area results by approximately 4-

5%. The cause for this trend is not well understood. It is unlikely to be related to an error in the placement of the baseline for integration, as this affects height less than area (Dyson et al., 1998). In contrast, the asymmetry of the FI-CL peaks could be a possible source of error during peak height measurement, since peak area is less sensitive to peak asymmetry than peak height (Dyson et al., 1998). It can also be observed in Table 5 that estimated combined uncertainties are larger for peak height than for peak area based results. This is mainly related to a larger uncertainty associated with the sensitivity coefficient for peak height compared with peak area (Table S1). The intensity repeatability and the within-sequence-stability are also slightly better for peak area than for peak height data, which can be related to count statistics. Area integration is considered the 'true' measure of the amount of solute (Dyson et al., 1998) and possible problems specific to peak area data such as peak overlap and / or low S/N ratios (Dyson et al., 1998) are not an issue with FI-CL measurements. These observations lead to the conclusion that peak area data may be preferable to peak height data with FI-CL measurement results, contrary to common practice. Additionally, users should routinely and systematically describe the way peak data are processed.

Comments and recommendations

The amount content of dissolved Fe in marine waters is measured to elucidate the biogeochemical cycling of this element and its role in the oceanic sequestration of atmospheric CO_2. However, quantifying the amount of Fe present in <0.2 µm filtered and acidified seawater samples remains a difficult analytical task, and achieving reliable results is a challenging objective. Moreover, the uncertainty as part of the measurement results is easily underestimated.

FI-CL is a technique commonly applied because of its portability and hence suitability for shipboard deployment. This paper proposes that the relative expanded (k=2) combined uncertainty of the measurement results using FI-CL in the described configuration cannot be better than about 10 to 15% for seawater samples containing 0.5 to 1 nmol kg^{-1} of dissolved Fe. When applied on-board ship the minimum achievable uncertainty is likely to be even larger owing to the more challenging working conditions compared with shore-based laboratories. Moreover, this paper emphasises the fact that it will be beneficial to researchers to refine measurement practices in order to improve the likelihood of achieving lower uncertainty targets. For FI-CL, the uncertainty associated with the calibration slope and the within-sequence-stability are shown to be much greater sources of uncertainty than the intensity repeatability alone. Experimental planning must therefore systematically address the identification of strategies aimed at quantifying and minimising the role of these uncertainty contributors. These strategies include the use of as many calibration standards as possible (ideally 5 plus the 'zero' standard measured with 6 replicates) and measurements repeated regularly for the same sample over the entire analytical sequence. It is also shown that more attention needs to be paid to the way FI-CL peak data are

452 collected and processed, as this could lead to significant errors with respect to the size of the combined
453 uncertainties. To enhance the transparency of these aspects it is recommended that more comprehensive
454 descriptions of the methods used to validate the measurement procedures (including the way peak data
455 collection/processing is performed) are included in publications and reports. Moreover, a simple equation
456 to approximately estimate the uncertainty has been proposed, which is valid if the blank levels are
457 significantly lower than the levels of interest.

References

- Achterberg, E.P., T.W. Holland, A.R. Bowie, R.F.C. Mantoura and P.J. Worsfold, 2001. Determination of iron in seawater. Analytica Chimica Acta, 442: 1-14.
- Analytical Methods Committee, 1995. Uncertainty of measurement: implications of its use in analytical science. Analyst 120: 2303-2308.
- Bowie, A.R., E.P. Achterberg, S. Blain, M. Boye, P.L. Croot, P.L., H.J.W. de Baar, P. Laan, G. Sarthou and P.J. Worsfold, 2003. Shipboard analytical intercomparison of dissolved iron in surface waters along a north–south transect of the Atlantic Ocean. Marine Chemistry 84: 19-34.
- Bowie, A.R., E.P. Achterberg, P.L. Croot, H.J.W. de Baar, P. Laan, P., J.W. Moffett, S. Ussher and P.J. Worsfold, 2006. A community-wide intercomparison exercise for the determination of dissolved iron in seawater. Marine Chemistry 98: 81-99.
- Bowie, A.R., P.N. Sedwick and P.J. Worsfold, 2004. Analytical intercomparison between flow injection-chemiluminescence and flow injection-spectrophotometry for the determination of picomolar concentrations of iron in seawater. Limnology and Oceanography: Methods 2: 42-54.
- Boyd, P.W. and M.J. Elwood, 2010. The biogeochemical cycle of iron in the ocean. Nature Geoscience 3: 675-682.
- Bucciarelli, E., S. Blain and P. Tréguer, 2001. Iron and manganese in the wake of the Kerguelen Islands (Southern Ocean). Marine Chemistry 73: 21-36.
- Dyson, N., 1998. Chapter 2: Errors in peak area measurement. In Chromatographic Integration Methods. The Royal Society of Chemistry. pp. 35-88.
- GEOTRACES, 2013. Dissolved Iron – values in nmol/kg Consensus values (± 1 std. dev.) for SAFe Reference Samples as of May 2013. http://www.geotraces.org/science/intercalibration/322-standards-and-reference-materials (last accessed July 15, 2014)
- Gledhill, M.B, and K.N. Buck, 2012. The organic complexation of iron in the marine environment: A review. Frontiers in microbiology 3: 1-17.
- Inczedy, J., T. Lengyel, A. Ure, A. Gelencsér and A. Hulanicki, 1998. Compendium of analytical nomenclature. Blackwell Science Ltd., Oxford, UK.
- ISO/IEC, General requirements for the competence of testing and calibration laboratories, ISO/IEC 17025:2005, International Organization for Standardization/International Electrotechnical Commission, Geneva, 2005.
- JCGM 100, Evaluation of measurement data — Guide to the expression of uncertainty in measurement; Joint Committee for Guides in Metrology (BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, ILAC), BIPM, www.bipm.org, Paris, 2008.

- JCGM 200, International vocabulary of metrology — Basic and general concepts and associated terms (VIM); Joint Committee for Guides in Metrology (BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, ILAC), BIPM, www.bipm.org, Paris, 2008
- Johnson, K. S., E.Boyle, K. Bruland, K. Coale, C. Measures, J. Moffett, A. Aguilar-Islas, K. Barbeau, B. Bergquist, A. Bowie, K. Buck, Y. Cai, Z. Chase, J. Cullen, T. Doi, V. Elrod, S. Fitzwater, M. Gordon, A. King, P. Laan, L. Laglera-Baquer, W. Landing, Lohan, J. Mendez, A. Milne, H. Obata, L. Ossiander, J. Plant, G. Sarthou, P. Sedwick, G. J. Smith, B. Sohst, S. Tanner, S. Van den Berg and J. Wu, 2007. Developing standards for dissolved iron in seawater. Eos, Transactions of the American Geophysical Union 88: 131-132.
- Klunder, M., P. Laan, R. Middag, H. de Baar and J. Van Ooijen, 2011. Dissolved iron in the Southern Ocean (Atlantic sector). Deep Sea Research Part II: Topical Studies in Oceanography 58: 2678-2694.
- Linsinger, T., 2010. Comparison of a measurement result with the certified value. European Reference Materials: Application Note 1.
- Lohan, M.C., A.M. Aguilar-Islas and K.W. Bruland, 2006. Direct determination of iron in acidified (pH 1.7) seawater samples by flow injection analysis with catalytic spectrophotometric detection: Application and intercomparison. Limnology and Oceanography: Methods 4: 164-171.
- Martin, J. and S. Fitzwater, 1988. Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. Nature 331: 341-343.
- Mawji, E., M. Gledhill, J.A. Milton, G.A. Tarran, S. Ussher, A. Thompson, G.A. Wolff, P.J. Worsfold, and E.P. Achterberg, 2008. Hydroxamate siderophores: Occurrence and importance in the Atlantic Ocean. Environmental Science & Technology 42: 8675-8680.
- McNaught, A.D. and A. Wilkinson, 1997. Compendium of chemical terminology, Blackwell Science Oxford.
- Metrodata GmbH, 2003. Gum Workbench. The software tool for the expression of uncertainty in measurement. Metrodata GmbH, D-79639 Grenzach-Wyhlen, Germany.
- Miller, J.N., 1991. Basic statistical methods for analytical chemistry. Part 2. Calibration and regression methods. A review. Analyst 116: 3-14.
- Mills, M.M., C. Ridame, M. Davey, J. La Roche and R.J. Geider, (2004) Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. Nature 429: 292-294.
- Nordic Committee on Food Analysis (NMKL), 1997. Pathogenic Vibrio species. Detection and enumeration in foods. Method n°156 (UDC 576.851)
- Obata, H., H. Karatani and E. Nakayama, (1993) Automated determination of iron in seawater by chelating resin concentration and chemiluminescence detection. Analytical Chemistry 65: 1524-1528.

- Petrov, I., C.R. Quétel and P.D.P. Taylor, (2007) Investigation on sources of contamination, for an accurate analytical blank estimation, during measurements of the Fe content in seawater by isotope dilution inductively coupled plasma mass spectrometry. Journal of Analytical Atomic Spectrometry 22, 608-615.
- Press, W.H., 2012. Numerical Recipes in C: The Art of Scientific Computing.
- Quétel, C.R., T. Prohaska, S.M. Nelms, J. Diemer and P.D.P. Taylor, 2001. ICP-MS applied to isotope abundance ratio measurements: performance study and development of a method for combining uncertainty contributions from measurement correction factors. In: G. Holland and S. Tanner (Editors), Plasma Source Mass Spectrometry: The New Millennium. 7th Durham Conference Royal Society of Chemistry, Cambridge, pp. 257
- Ussher, S.J., E.P. Achterberg, C. Powell, A.R. Baker, T.D. Jickells, R. Torres and P.J. Worsfold, 2013. Impact of atmospheric deposition on the contrasting iron biogeochemistry of the North and South Atlantic Ocean. Global Biogeochemical Cycles 27: 1-12.
- Ussher, S.J., E.P. Achterberg, G. Sarthou, P. Laan, H.J.W. de Baar and P.J. Worsfold, 2010a. Distribution of size fractionated dissolved iron in the Canary Basin. Marine Environmental Research 70: 46-55.
- Ussher, S.J., E.P. Achterberg and P.J. Worsfold, 2004. Marine Biogeochemistry of Iron. Environmental Chemistry 1: 67-80.
- Ussher, S.J., I. Petrov, C.R. Quétel and P.J. Worsfold, 2010b. Validation of a portable flow injection–chemiluminescence (FI-CL) method for the determination of dissolved iron in Atlantic open ocean and shelf waters by comparison with isotope dilution–inductively coupled plasma mass spectrometry (ID-ICPMS). Environmental Chemistry 7: 139-145.
- Wu, J.F., E. Boyle, W. Sunda and L.S. Wen, 2001. Soluble and colloidal iron in the oligotrophic North Atlantic and North Pacific. Science 293: 847-849.
- Wyatt, N.J., A. Milne, E.M.S. Woodward, A.P. Rees, T.J. Browning, H.A. Bouman, P.J. Worsfold and M.C. Lohan, 2014. Biogeochemical cycling of dissolved zinc along the GEOTRACES South Atlantic transect GA10 at 40°S. Global Biogeochemical Cycles 10.1002/2013GB004637.

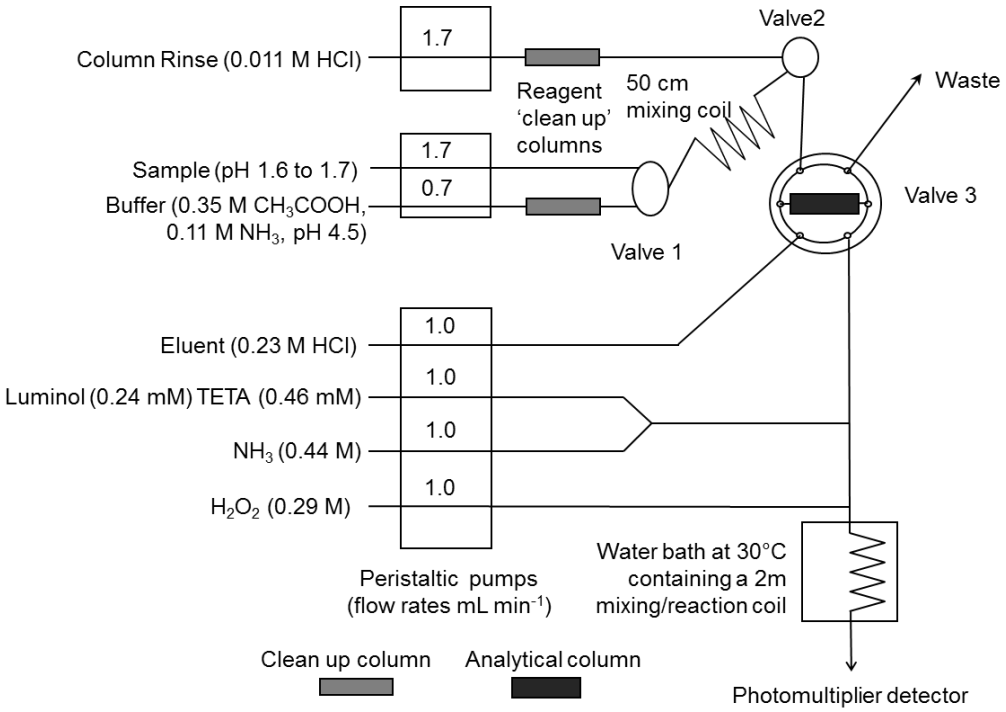


Figure 1: The FI-CL system used for the determination of dissolved Fe levels in seawater.

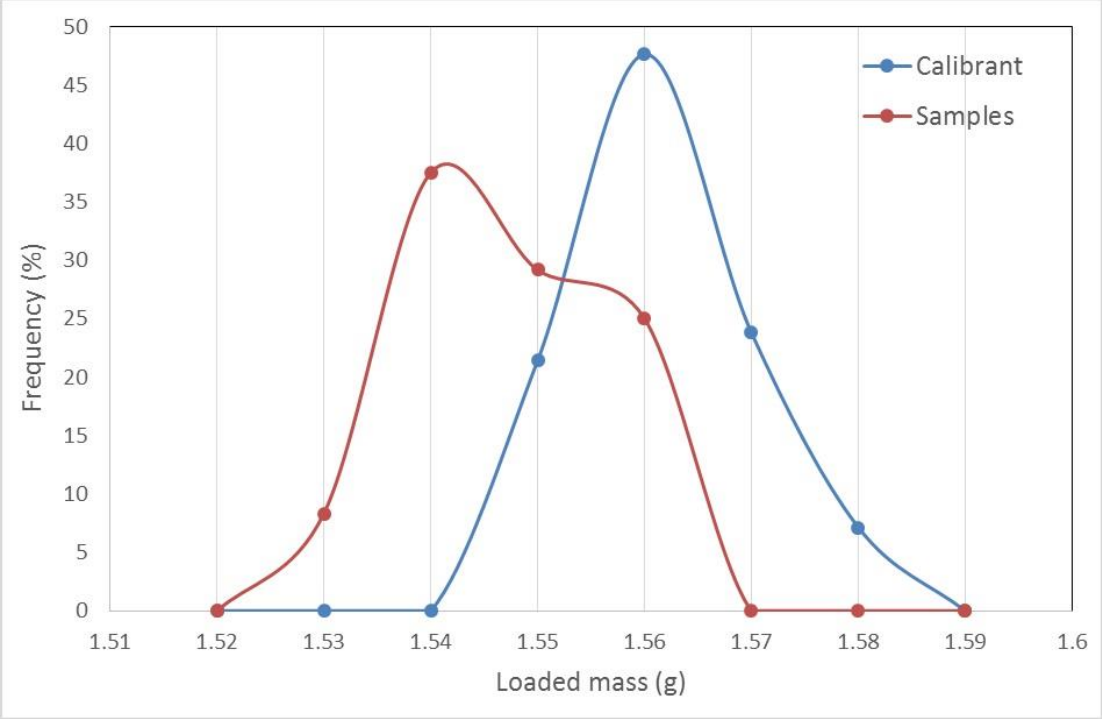
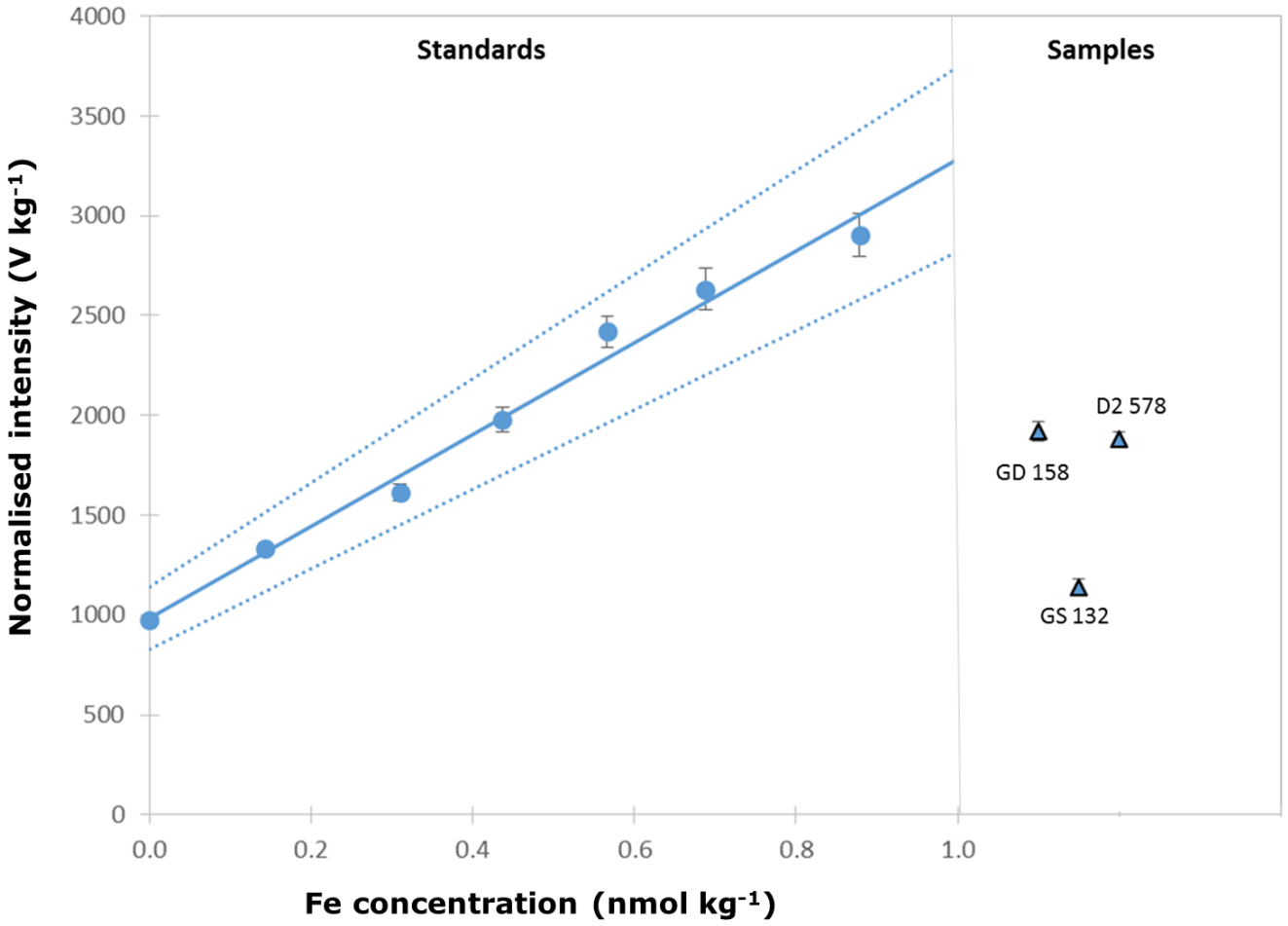
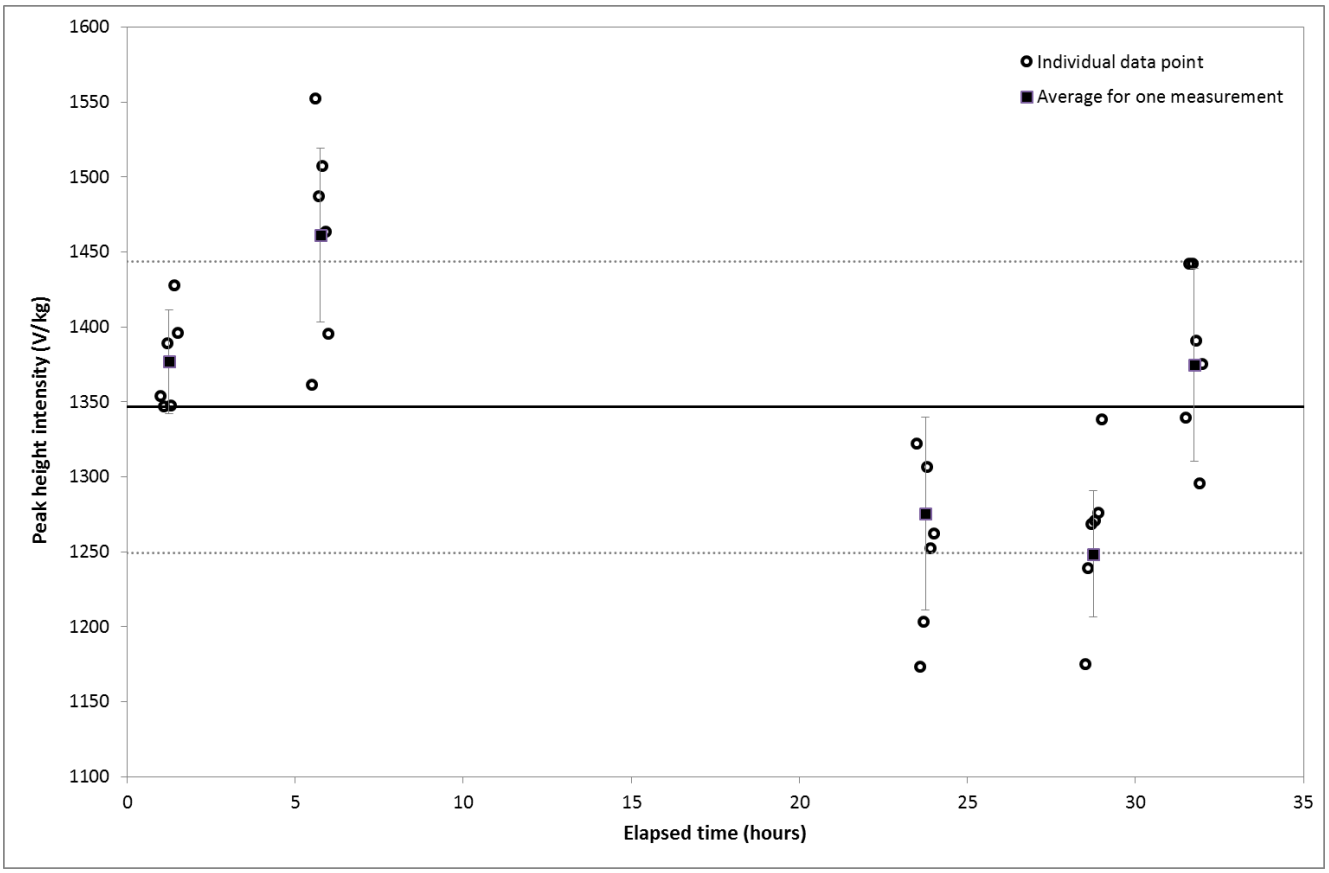


Figure 2: Frequency of variation (in %) of loaded masses for samples and calibration standards during the "sample experiment"



561
562 **Figure 3: Unweighted calibration using average data for the regression. Blue dotted lines delimit a 95%**
563 **confidence interval around the regression graph. Signal intensities observed for samples GD158, GS132 and**
564 **D2578 are also reported.**
565
566
567



568

Figure 4: Stability over the 32 h '"stability experiment" with seawater C using peak height based results. Vertical bars indicate the standard deviation of the average of the 6 replicates. Horizontal lines indicate the average and standard deviations for the groups of 5 repeat measurements.

569

570 **Table 1: Description of the samples used.**

Sample name	SWA	SWB	SWC	SAFe campaign	GEOTRACES campaigns	
				D2-578	GS-132	GD-158
Collection Location	05°20.5′ S, 06°11.9′ W to 06°44.8′ S, 05°04.8 ′W	27° 47.195′ S, 007°1 2.949‘ W	40° S 48.46° W	30° N, 140° W	31°40′ N 64°10′ W	31°40′ N 64°10′ W
Depth	Surface	500 m	Surface	1000 m	Surface	2000 m
Filtration	Sartorius Sartobran-P cartridge. Cellulose acetate 0.45 µm pre-filter then 0.2 µm filter	Whatman GD/X PTFE 0.2 µm filter	Pall Acropak Supor capsule. PES 0.8 pre-filter then 0.2 µm filter.	Polycarbonate track etched 0.45 µm pre-filter, then 0.2-µm filter. Homogenized in 1000 L fluorinated LDPE tanks	Pall Acropak Supor capsule. PES0.8 pre-filter then 0.2 µm filter.	
Acidification	Bulk sample acidified at sea with 700 mL of ~10 M Q-HCl. Homogenized in 1000 L fluorinated LDPE tanks	Acidified at Plymouth University (PU) with 2 mL of Romil UpA grade HCl per L seawater	Acidified at PU with 1 mL of Romil UpA grade HCl per L seawater	Acidified at sea with 2 mL of conc HCl per L seawater.	Homogenized in 500 L fluorinated LDPE tanks.Acidified at sea with 2 mL of conc HCl.	
Final pH	2.0	2.0	1.7	1.8	1.8	1.8
Consensus dissolved Fe ± 2 s.d. (nmol kg^{-1})	0.53 ± 0.20	N/A	N/A	0.933 ± 0.046	0.546 ± 0.092	1.0 ± 0.2
Reference	Bowie et al., 2006	Ussher et al., 2013	Wyatt et al., 2014	Lohan et al., 2006	Johnson et al, 2007	

571

572 **Table 2: Mathematical equations for quantification of the Fe amount content using**
573 **gravimetric loading and FI-CL based procedure.**

1. Amount content in the sample C_S
Blank corrected sample signal intensity divided by the sensitivity (calibration slope) of the measurement procedure : $C_S = \frac{\overline{J}_S - \overline{J}_B}{F}$
2. Normalised signal intensity for the sample $\overline{J}_S$
a. Normalised signal intensity for the sample accounting for all sources of uncertainty: $\overline{J}_S = \overline{J}_{R\_S} \cdot \delta_{rep\_S} \cdot \delta_{stab\_S}$
b. Average normalised raw signal intensity for consecutive replicates: $\overline{J}_{R\_S} = \frac{1}{n}\sum_i \frac{I_{S\_i}}{m_{S\_i}}$
3. Normalised signal intensity for the analytical blank $\overline{J}_B$
a. Normalised signal intensity for the analytical blank accounting for all sources of uncertainty: $\overline{J}_B = \overline{J}_{R\_B} \cdot \delta_{stab\_B} \cdot \delta_{rep\_B} \cdot \delta_{matrix\_B}$
b. Average normalised raw signal intensity for consecutive replicates under closed sample valve conditions: $\overline{J}_{R\_B} = \frac{1}{n}\sum_i \frac{I_{B\_i}}{\overline{m}_S}$
4. Calibration slope F
a. Slope accounting for all sources of uncertainty: $F = F_{reg} \cdot \delta_{matrix\_std}$
b. Slope of least squares regression line of the normalised signal intensity versus the amount added Fe: $F_{reg} = \frac{r\sum C_{\mathrm{std\_j}} \cdot \overline{J}_{\mathrm{std\_j}} - \sum C_{std\_\mathrm{j}} \cdot \sum \overline{J}_{std\_\mathrm{j}}}{r\sum C_{\mathrm{std\_j}}{}^2 - \left(\sum C_{std\_\mathrm{j}}\right)^2}$
5. Amount content of the added Fe in the calibration standards
a. Added Fe amount in the calibration standard: $C_{\mathrm{std\_j}} = \frac{m_{stock\_\mathrm{j}}}{(m_{stock\_\mathrm{j}} + m_{calSW\_\mathrm{j}})} \cdot C_{stock}$
b. Amount in the stock solution: $C_{stock} = \frac{m_{mother\_aliquot}}{m_{stock} + m_{mother\_aliquot}} \cdot C_{mother}$

574

Parameter		**Index**	
C	Fe amount content (nmol kg^{-1})	S	Sample
I	Signal intensity (V)	B	Blank
$\bar{J}$	Average mass normalised intensity (V kg^{-1})	R	Raw
		Std	Calibration Standard
F	Sensitivity coefficient (slope, V $nmol^{-1}$)	stock	Intermediate Fe standard stock solution (prepared dilution of the mother solution)
n	Number of replicates	mother	Mother Fe standard solution (commercial standard)
r	Number of calibration standards	i	Index referring to the x^{th} sample replicate
m & $\bar{m}$	Mass & average mass (kg)	j	Index referring to the x^{th} standard
		Reg	Sensitivity coefficient (calibration slope) obtained by linear regression
		calSW	a 'low iron' seawater substrate used to produce the calibration curves
δ	Unity multiplicative correction factors carrying the relative uncertainty associated to the parameter considered	Stab	Accounts for the uncertainty arising from the intensity stability over an analytical sequence
		matrix	Accounts for the uncertainty arising from matrix effects on the sensitivity
		rep	Accounts for the uncertainty arising from the intensity repeatability
		WtoV	Accounts for the uncertainty related to the difference in loaded mass whether it is done by weighing of volumetrically

575

576

577 **Table 3: Slopes and their associated standard uncertainties depending on the regression calculations**
578 **considered. r is the number of standards and n the number of replicates per standard.**

Regression approach		Data points	Sensitivity coefficient (=slope) (F)	
			Value	Uncertainty (k=1)
Weighted regression		7 (r)	2301	83
Unweighted regression	average values		2297	118
	all individual data	42 (r*n)	2297	56

579

580 **Table 4: Dependence of the relative standard uncertainty (rsu) on the calculated slope/sensitivity coefficient,**
581 **rsu (F), in %, on the number of replicates or calibration standards used.**

n	rsu (F), with n = number of replicates using 7 calibration standards (original + 6 Fe addition levels)	rsu (F), with n = number of calibration standards using 6 replicates for each standard
6	6.6	6.6
5	7.5	6.8
4	7.9	11.5
3	8.6	14.6

582

583 **Table 5: Amount content results with combined expanded uncertainty with a coverage factor (k) of 2 (i.e. 95%**
584 **confidence interval) for the three sea water samples from the SAFe and GEOTRACES campaigns using**
585 **gravimetric loading. Consensus values were downloaded from the GEOTRACES.org website and are from May**
586 **2013.**

Sample	Obtained Fe amount content (nmol kg^{-1})				Consensus Fe amount content (nmol kg^{-1})	
	Peak height		Peak area			
	Value	Relative uncertainty	Value	Relative uncertainty	Value	Relative uncertainty
D2	0.82 ± 0.10	12	0.861 ± 0.086	10	0.933 ± 0.046	4.9
GS	0.478 ± 0.060	12	0.500 ± 0.051	10	0.546 ± 0.092	16.8
GD	0.800 ± 0.099	12	0.836 ± 0.084	10	1.0 ± 0.2	20.0

587
588
589
590
591
592
593
594

595 **Table 6: Relative contributions (%) to the combined uncertainty budget estimated for the dissolved Fe level**
596 **measured by FI-CL in the GD sample from the GEOTRACES campaign (symbols as in Table 2). The intermediate**
597 **result refers to the parameters used in equation 1 of Table 2, in which all associated uncertainties are included.**

Quantity		Gravimetric loading	
		Peak height	Peak area
$\overline{J}_S$ (V/kg)	**Intermediate result**	**29.5**	**44.4**
	$\overline{J}_{R\_S}$ (treated as constant)	-	-
	δ_{rep_S}	7.9	9.4
	δ_{stab_S}	21.6	35.0
$\overline{J}_B$ (V/kg)	**Intermediate result**	**0.6**	**1.4**
	I_B (treated as constant)	-	-
	$\overline{m}_S$	0.0	0.0
	δ_{rep_B}	0.0	0.6
	δ_{stab_B}	0.1	0.0
	δ_{matrix_B}	0.5	0.8
F (sensitivity coefficient or slope) (V/nanomol)	**Intermediate result**	**69.7**	**54.3**
	F_{reg}	69.7	54.3
	δ_{matrix_std}	0.0	0.0

598
599

Supplementary information

Table S1: details of the uncertainty budget associated to the result of the measurement by FI-CL (with gravimetric loading) of the dissolved Fe amount content in the D2 reference material from SAFe. Symbols as in Table 2.

Quantity			Peak height			Peak area		
			Value	Stand Unc (k=1)		Value	Stand Unc (k=1)	
				Absolute	%		Absolute	%
$\overline{J}_S$ (V/kg)	**Intermediate result**	D2	**1918**	**63**	**3.3**	**52179**	**1700**	**3.3**
		GS	**1140**	**37**	**3.3**	**30809**	**1000**	**3.3**
		GD	**1879**	**62**	**3.3**	**50711**	**1700**	**3.3**
	$\overline{J}_{R\_S}$	D2	1918	0	0	52179	0	0
		GS	1140	0	0	30809	0	0
		GD	1879	0	0	50711	0	0
	δ_{rep_S}		1	0.017	1.7	1	0.015	1.5
	δ_{stab_S}		1	0.028	2.8	1	0.029	2.9
$\overline{J}_B$ (V/kg)	**Intermediate result**		**41.8**	**2.2**	**5.3**	**1115**	**190**	**17.0**
	I_B		0.0645	0	0	1.72	0	0.0
	$\overline{m}_S$		0.001542	0.000004	0.3	0.001542	0.000004	0.3
	δ_{rep_B}		1	0.069	6.9	1	0.17	17.0
	δ_{stab_B}		1	0.10	10	1	0	0
	δ_{matrix_B}		1	0.2	20	1	0.2	20
F (sensitivity coefficient or slope) (V/nanomol)	**Intermediate result**		**2297**	**118**	**5.1**	**59330**	**2190**	**3.7**
	F_{reg}		2297	118	5.1	59330	2190	3.7
	δ_{matrix_std}		1	0	0	1	0	0