

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

Geerke H. Floor,<sup>1,a</sup> Robert Clough,<sup>2</sup> Maeve C. Lohan,<sup>2</sup> Simon J. Ussher,<sup>2</sup> Paul J. Worsfold,<sup>2</sup> Christophe R. Quétel\*<sup>1</sup>

<sup>1</sup>Institute for Reference Materials and Measurements, Joint Research Centre–European Commission, Geel, Belgium

<sup>2</sup>Biogeochemistry Research Centre, School of Geography, Earth and Environmental Sciences, Plymouth University, Plymouth, United Kingdom

### 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  $<0.2 \mu\text{m}$  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 normalised intensities (in  $V \text{ kg}^{-1}$ ) vs. values of the Fe amount content (in  $\text{nmol kg}^{-1}$ ) 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% (coverage factor  $k = 2$ ), respectively. The most important contributory factors were the uncertainty on the sensitivity coefficient (i.e., calibration slope) and the within-sequence-stability (i.e., the signal stability over several hours of operation; here 32 h). For GD, using peak height measurements, these factors contributed respectively 69.7% and 21.6% while the short-term repeatability accounted for only 7.9%. 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.

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 half of the world ocean (Moore et al. 2001) 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: (1) the use of reference materials, (2) the comparison of results achieved with other methods, (3) inter-laboratory comparison, (4) systematic assessments of the factors influencing the result and (5) 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

Additional Supporting Information may be found in the online version of this article.

<sup>†</sup>Present address: GFZ German Research Centre for Geosciences, Helmholtz Centre Potsdam, Potsdam, Germany

\*Correspondence: Christophe.Quétel@ec.europa.eu

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

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 2012). 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 reference materials 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<sub>3</sub>, 20–22%) and glacial acetic acid (CH<sub>3</sub>CO<sub>2</sub>H), all SpA (Super Pure Acid) 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 7 d. 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.

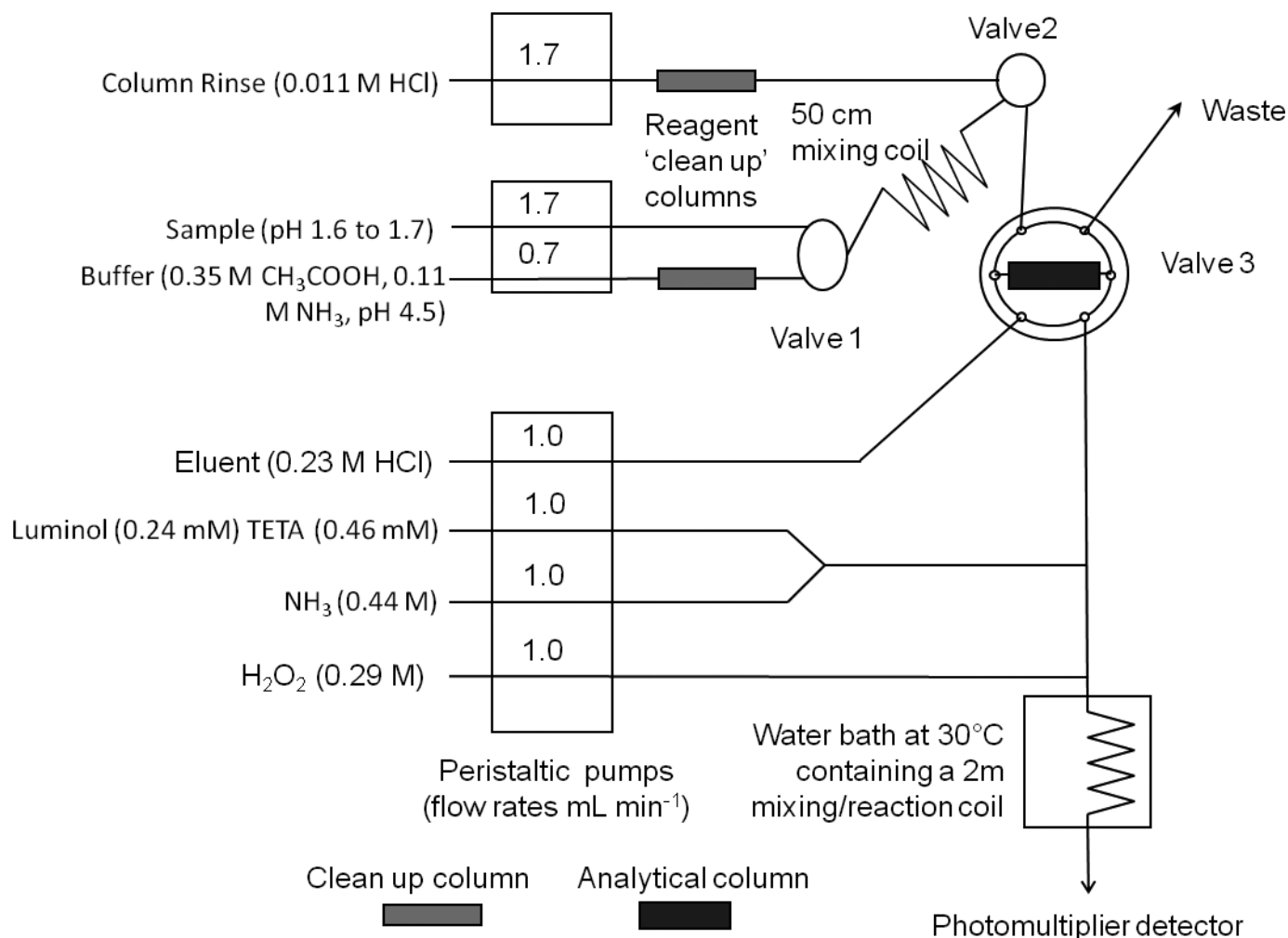
The main characteristics of the seawater samples used for this project are described in Table 1. Briefly, all samples were filtered at sea and then acidified either at sea or at Plymouth University (PU). Seawater samples, referred to as SWA, SWB, and SWC, containing  $\leq 0.5$  nmol kg<sup>-1</sup> 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<sup>-1</sup> 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 (Johnson 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

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

Sample name	SAFE campaign			GEOTRACES campaigns		
	SWA	SWB	SWC	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° 47.2' S, 07° 12.9' 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 1 mL of Romil UpA grade HCl per L seawater	Acidified at PU with 2 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
Salinity	34.1	30.6	34.2		36.7	34.9
Consensus dissolved Fe ± 2 s.d. (nmol kg <sup>-1</sup> )	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)	-	Wyatt et al. (2014)	Lohan et al. (2006)	Johnson et al. (2007)	



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

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  $\mu\text{L}$ ). The peristaltic pump tubing used was two stop Accu-Rated<sup>TM</sup> PVC (Elkay, Basingstoke, UK) and all other manifold tubing was 0.8 mm i.d. PTFE. The pumps were turned on and run for 2 h before any measurements were made. If the pump tubing was changed it was conditioned by running the pump slowly overnight. The system used two poly(methyl methacrylate) columns (1 cm long, 1.5 mm i.d., volume 70  $\mu\text{L}$ ), loaded with Toyopearl AF Chelate 650 resin (Tosoh Bioscience, Stuttgart, Germany) retained with HDPE frits (BioVion F, 0.75 mm thick, 22–57  $\mu\text{m}$  pore size), to clean up the buffer and column rinse solutions (the clean-up column on the rinse solution line is not strictly necessary). The analytical column, also loaded with Toyopearl AF Chelate 650 resin, was made of polyethylene with LDPE frits with an internal volume of 200  $\mu\text{L}$  (Global FIA, Fox Island, U.S.A.). Further details of the

physico-chemical properties of the resin can be found in Shelley et al. (2010).

Peristaltic pump, 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 the chemistry reported in Obata et al. (1993), was as follows. A working solution of approximately 0.35  $\mu\text{mol kg}^{-1}$  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  $\text{nmol kg}^{-1}$  to 0.9  $\text{nmol kg}^{-1}$  Fe in 0.15  $\text{nmol kg}^{-1}$  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  $\mu\text{L}$  aliquot of a 10 mM  $\text{H}_2\text{O}_2$  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 luminol was obtained by dissolving 0.177 g of luminol and 0.25 g of  $\text{Na}_2\text{CO}_3$  in 20 mL of HPW. This stock was then diluted to give a 0.24 mM working solution. The post column reagents for the chemiluminescence reaction was a mixture of 0.23 M HCl, 0.44 M  $\text{NH}_3$ , 0.24 mM luminol/0.46 mM TETA and 0.31 M  $\text{H}_2\text{O}_2$ . The acidified reference samples and standards of seawater were buffered on-line to pH 3.5 with 0.35 M  $\text{CH}_3\text{CO}_2\text{H}$  and 0.11 M  $\text{NH}_3$ . To precondition and wash the column, 0.011 M HCl was used.

To operate the FI-CL instrument, the LabVIEW software was opened and the baseline signal from the PMT monitored to check for stability. The pump controlling the eluent and post-column reagents was then activated and the baseline chemiluminescence signal recorded after the signal had stabilised. Each analytical session started with the measurement of a procedural blank (by application of the “closed sample line” method). For this, the sample flow was stopped, by closing one port on valve 1, so that only the rinse solution and ammonium acetate buffer passed over the column. The FI-CL system was then operated by loading and injecting SWA for at least 30 min to monitor stability. Subsequently, calibration seawater standards and samples were analysed. The FI-CL manifold was fully automated and one replicate measurement consisted of the following analytical cycle. The column was conditioned for 10 s with 0.011 M HCl. Then the sample and buffer were loaded simultaneously for 60 s. The loading pH was optimised for maximum retention of Fe (Clough et al. 2015). The column was washed with 0.011 M HCl for 20 s. The Fe was then eluted with 0.23 M HCl for 120 s (total time for one analytical measurement = 210 s). Six replicate measurements were made for each sample or standard solution. The mass of loaded sample or standard solution was gravimetrically determined for each replicate by differential weighing. Between each sample the sample flow path was washed with HPW for 30 s followed by uptake of the fresh sample for 180 s to flush the line up to valve 2 (total time for one analytical cycle of measurement and washing for six replicates = 21 min). After each analytical session all fluid paths were flushed with 0.011 M HCl for 10 min and then with HPW for 15 min and HPW was left in the lines.

### Data treatment

Data integration was also performed with the custom build software run in LabVIEW. The baseline, and the start and end points of the peak were set manually for each transient signal. The main calculations in this study were carried out on the basis of peak height data, as this is common practice for FI-CL measurements in the oceanographic community (and the wider FI community). Peak area measurements were also made and some of the differences observed when using peak areas are discussed below. Further data treatment, including calculations for the estimation of standard uncertainties, was carried

out in Excel®. The combined uncertainties were obtained by propagating together individual uncertainty components according to the GUM (JCGM 100 2008). In practice, a dedicated software program was used (Metrodata GmbH 2003). The reported combined uncertainties are expanded uncertainties and reported as  $U = ku_c$  where  $u_c$  is the combined standard uncertainty and  $k$  is a coverage factor equal to 2. If “the probability distribution characterized by  $\gamma$  and  $u_c(\gamma)$  is approximately normal and the effective degrees of freedom of  $u_c(\gamma)$  is of significant size” (“greater than 10”), “taking  $k = 2$  produces an interval having a level of confidence of approximately 95%” (JCGM 100 2008).

## Assessment

### Description of the measurand

The GUM states that a measurement begins with an appropriate specification of the measurand, the particular quantity intended to be measured (JCGM 100 2008). Iron exists in different physico-chemical forms in seawater. Traditionally, filtration is performed to differentiate between the different physical size fractions (Wu et al. 2001; Ussher et al. 2004, 2010a). Additionally, iron occurs in two oxidation states; Fe(II) and Fe(III). Generally, Fe(III) predominates in oxygenated waters, of which most (80–99%) is strongly complexed by organic ligands (Achterberg et al. 2001; Mawji et al. 2008; Gledhill and Buck 2012). In this study, the measurand is the amount content of Fe present in  $<0.2 \mu\text{m}$  filtered and acidified samples and is regarded as the dissolved fraction of the Fe present in the seawaters. The aim was to obtain the Fe amount content in specific samples and therefore the uncertainties associated with the sampling process and/or the sample conditioning phase have not been considered.

### Experimental design

Three different types of experiment were performed in different analytical sessions. First, the stability of the analytical procedure was checked with five measurements (six replicates of each) performed over a period of 32 h for SWC and a procedural blank using the closed sample line approach to obtain the within-sequence stability. This experiment was termed the “stability experiment.” The FI-CL manifold was run continuously for this period, i.e., the pumps remained on and the same batch of reagents was used throughout. Thirty two hours is the maximum time that the manifold can operate continuously with a single batch of reagents. Second, the effect of small variations in the matrix was investigated by comparing the sensitivity factor using three different seawaters as standards (Table 1). On the first day, SWA was compared with SWB while on the second day SWA was compared with SWC (“matrix experiment”). Third, the FI-CL based procedure was applied to the determination of iron in samples of three filtered and acidified seawater reference materials using SWA for calibration (“reference material experiment”).

**Table 2.** Mathematical equations for quantification of the Fe amount content using 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:  $C_S = \frac{\bar{I}_S - \bar{I}_B}{F}$
2. Normalised signal intensity for the sample  $\bar{J}_S$ 
  - a. Normalised signal intensity for the sample accounting for all sources of uncertainty:  $\bar{J}_S = \bar{J}_{R,S} \cdot \delta_{rep,S} \cdot \delta_{stab,S}$
  - b. Average normalised raw signal intensity for consecutive replicates:  $\bar{J}_{R,S} = \frac{1}{n} \sum_i \frac{I_{S,j}}{m_{S,j}}$
3. Normalised signal intensity for the analytical blank  $\bar{J}_B$ 
  - a. Normalised signal intensity for the analytical blank accounting for all sources of uncertainty:  $\bar{J}_B = \bar{J}_{R,B} \cdot \delta_{stab,B} \cdot \delta_{rep,B} \cdot \delta_{matrix,B}$
  - b. Average normalised raw signal intensity for consecutive replicates under closed sample line conditions:  $\bar{J}_{R,B} = \frac{1}{n} \sum_i \frac{I_{B,j}}{\bar{m}_S}$
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 vs. the amount added Fe:  $F_{reg} = \frac{r \sum C_{std,j} \cdot \bar{J}_{std,j} - \sum C_{std,j} \cdot \sum \bar{J}_{std,j}}{r \sum C_{std,j}^2 - (\sum C_{std,j})^2}$
5. Amount content of the added Fe in the calibration standards
  - a. Added Fe amount in the calibration standard:  $C_{std,j} = \frac{m_{stock,j}}{(m_{stock,j} + m_{calSW,j})} \cdot C_{stock}$
  - b. Amount in the stock solution:  $C_{stock} = \frac{m_{mother,aliquot}}{(m_{stock} + m_{mother,aliquot})} \cdot C_{mother}$

Parameter		Index	
C	Fe amount content (nmol kg <sup>-1</sup> )	S	Sample
I	Signal intensity (V)	B	Blank
$\bar{J}$	Average mass normalised intensity (V kg <sup>-1</sup> )	R	Raw
		Std	Calibration Standard
F	Sensitivity coefficient (slope, V nmol <sup>-1</sup> )	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 <sup>th</sup> sample replicate
m & $\bar{m}$	Mass and average mass (kg)	j	Index referring to the x <sup>th</sup> standard
		Reg	Sensitivity coefficient (calibration slope) obtained by linear regression
		calSW	a "low iron" seawater substrate used to produce the calibration curves
$\delta$	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 or volumetrically

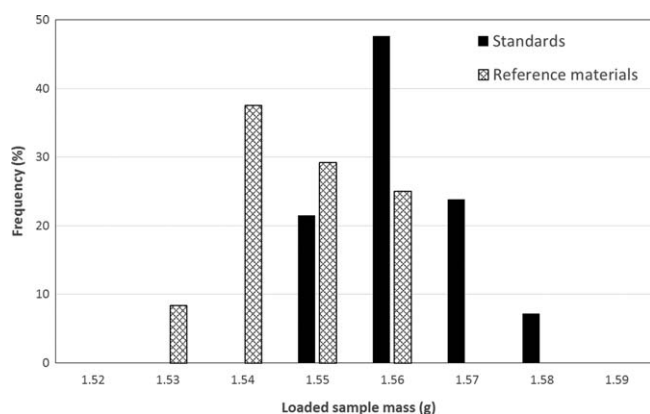
**Calculating the dissolved Fe amount content in the samples and mathematical description of the measurement procedure**

Implicit in the GUM "is the assumption that a measurement can be modelled mathematically to the degree imposed by the required accuracy of the measurement" (JCGM 100 2008). A measurand Y is determined from various input quantities X<sub>i</sub> through a functional relationship. These input quantities "may themselves be viewed as measurands and may themselves depend on other quantities, including corrections and correction factors" "that can contribute a significant component of uncertainty to the result of the measurement" (JCGM 100 2008). A mathe-

tical description of the FI-CL measurement procedure is given through Eqs. 1–5 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 (Eq. 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



**Fig. 2.** Frequency of variation (in %) of loaded masses for reference materials and calibration standards during the “reference material experiment.”

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 32 h long “stability experiment,” with about 5% decrease in the sample mass loaded from the first to the last measurement (data not shown). During the “reference material experiment” the average loaded mass for samples (which were all run at the end of the sequence) was lower than for the standards (Fig. 2). The observed changes in mass loaded over time are likely to be due to changes in manifold parameters such as flow rates and column hydrodynamics. 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.

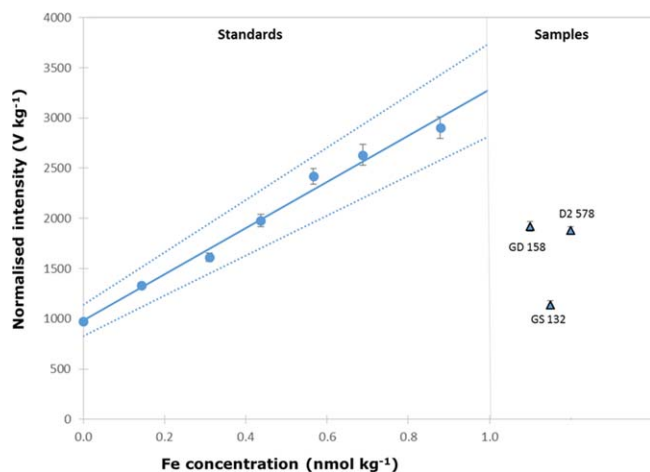
As a result of this finding, mass normalised signals (symbol J, in  $V \text{ kg}^{-1}$ ) were used throughout this study for the calculations (Eq. 2b; Table 2). Following the example given in Quétel et al. (2001), in Eqs. 2a, 3a, 4a, and 6 unity multiplicative factors were introduced to carry standard uncertainties associated with signal stability, matrix effects and differences during mass loadings. Applying the rules of uncertainty propagation to calculations involving several intensity results (such as average calculations, for instance), and thus combining together repeatability values associated to every single signal intensities, could lead to an overestimation of the resulting combined uncertainty. To avoid this risk, instead of applying the rules of uncertainty propagation to these calculations, these calculation results were multiplied by factors equated to 1 and carrying relative uncertainty estimations considered realistic and representative for these cal-

culations. The ways these relative combined uncertainty components were estimated are discussed in the “Assessing the standard uncertainties” section 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  $\text{H}_2\text{O}_2$ , HCl, the buffer and rinse 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 used to generate the chemiluminescence reaction is included in the baseline. Baseline subtraction for the determination of net peak height or peak area signals, as commonly applied in FI methods, should therefore remove this possible bias. The influence of additions of chemical reagents for the purpose of preserving and/or conditioning the samples prior to the measurements (e.g., acid,  $\text{H}_2\text{O}_2$ ) 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, 2004; Klunder et al. 2011). The major contribution to any blank signal arising from the reagents is therefore most likely to come from the buffer solution.

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). In this work using FI-CL the procedural blank was considered to be the signal obtained with the “closed sample line” method as described above, i.e., loading only buffer (Bowie et al. 2004; Ussher et al. 2010a). Alternative measurement procedures for blank determination are the field blank approach (which requires a matrix containing no analyte) or varying sample loading times and extrapolating back to time zero (not accounting for the buffer solution) (Bowie et al. 2004). There is a risk that matrix effects and pH changes could influence final results due to fluctuations in the blank values determined using the “closed sample line” method and this is discussed further in the “Uncertainty on blank corrections” section below.

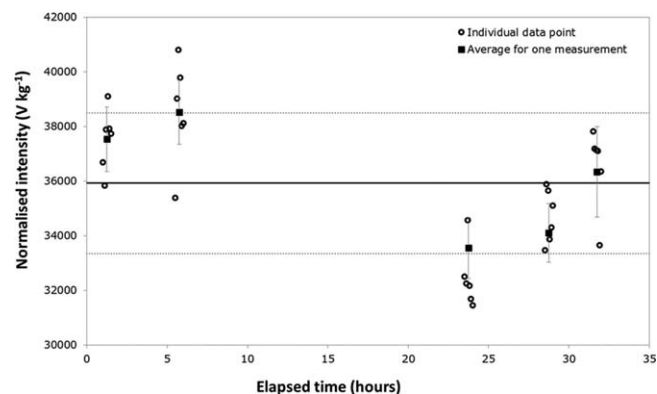


**Fig. 3.** Unweighted calibration using average data for the regression. Blue dotted lines delimit a 95% confidence interval around the regression graph. Signal intensities observed for samples GD158, GS132 and D2578 are also reported.

Normalised signal intensities were calculated by division by the average loaded sample mass (Eq. 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 (Eq. 3a) and are discussed in more detail in the “Assessing the standard uncertainties” section 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 nmol kg<sup>-1</sup> to 0.9 nmol kg<sup>-1</sup>. Since measurements were repeated six times for each calibration point, a total of  $7 \times 6 = 42$  results were obtained. A linear regression was plotted (not shown), 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 six 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,



**Fig. 4.** Stability over the 32 h “stability experiment” with seawater C using mass normalised peak height based results. Vertical bars indicate the standard deviation of the average of the six replicates. Horizontal lines indicate the average and standard deviations for the groups of five repeat measurements.

in nmol kg<sup>-1</sup>) on the  $x$  axis and the corresponding average normalised intensities ( $J$ , in V kg<sup>-1</sup>) on the  $y$  axis (Fig. 3). The sensitivity coefficient ( $F$ , in V nmol<sup>-1</sup>), i.e., the slope, is obtained using Eq. 4b from Table 2. 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

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,  $n = 6$ ) during the “reference material experiment” and between 2.4% and 4.9% RSD ( $n = 6$ ) during the “stability experiment.” These variations in RSD cannot be explained by variations in the sample since variable RSD was also observed in the “stability experiment” where the same solution was measured. Together with a short term source of variability a longer term component, with the within-sequence stability, was also involved and influencing the intensity values (Fig. 4). Over the 32 h long analytical



**Table 3.** Slopes and their associated standard uncertainties depending on the regression calculations 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 ( <i>k</i> = 1)
Weighted regression	7 ( <i>r</i> )	2301	83
Unweighted regression	Average values	2297	118
	All individual data	2297	56

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 ( $\delta_{rep\_S}$  and  $\delta_{stab\_S}$ ). Applying ANOVA to data from the “stability experiment” gave 4.1% and 6.3% as, respectively, the intensity repeatability and the relative within-sequence-stability component. Assuming independence between the intensity values used to calculate both types of RSDs, the relative standard uncertainties associated to  $\delta_{rep\_S}$  and  $\delta_{stab\_S}$  were estimated using these RSDs divided by square root 6 (the number of intensity replicates per measurement) and square root 5 (the number of repeat measurements in 32 h) respectively, to give values of 1.7% and 2.8%. Previously published work (Ussher et al. 2005) suggests that the major source of this uncertainty is associated with the column (loading and elution).

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 Eq. 6 as described below:

$$C_S = \frac{\bar{I}_{R\_S} \cdot \delta_{rep\_S} \cdot \delta_{stab\_S} \cdot \delta_{WtoV\_S} \cdot \bar{I}_{R\_B} \cdot \delta_{stab\_B} \cdot \delta_{rep\_B} \cdot \delta_{matrix\_B}}{F_{reg} \cdot \delta_{matrix\_std}} \tag{6}$$

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<sup>-1</sup>) and has the units V/nmol L<sup>-1</sup>. Secondly, an extra unity multiplicative correction factor ( $\delta_{WtoV\_S}$ ) was introduced to take account of the difference in the mass loading between samples and standards (Fig. 2). Using this dataset 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 corresponding to the intensity repeatability and within-sequence-stability component, the uncertainties 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* = 6) and the within-sequence-stability component (*n* = 5). A unity multiplicative factor  $\delta_{matrix\_B}$  with a value of 1.0 ± 0.2 was conservatively applied in Eq. 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 procedural 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 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 six rather than three replicates.

Small matrix differences between the three seawaters tested in the “matrix experiment” (see Table 1 for salinities) did not lead to significant differences ( $t_{calc} < 1.96$  for all comparisons; *p* = 0.05) between the slopes obtained for SWA, SWB, and SWC. Iron binding ligands may also affect the linearity of calibration depending on the time allowed for equilibration after spiking seawater with Fe. The

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

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

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

Sample	Obtained Fe amount content					
	Peak height		Peak area		Consensus Fe amount content	
	Value (nmol kg <sup>-1</sup> )	Relative uncertainty (%)	Value (nmol kg <sup>-1</sup> )	Relative uncertainty (%)	Value (nmol kg <sup>-1</sup> )	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

concentrations of iron binding ligands were not measured in these matrices and literature data give a range of  $\approx 0.4$ – $5.0$  nM for Atlantic waters (Gledhill and Buck 2012). However, since SWC, D2, GS and GD were acidified to pH 1.7/1.8 the organic ligands should not impact on the calibration for these matrices (Lohan et al. 2006). 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 six replicates each time, the “closed sample line” approach for blank assessment and a least square regression calibration line with seven levels (no Fe added + 6 levels of added Fe) in SWA. This was the “reference material 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 Supporting Information Table S1. The relative

**Table 6.** Relative contributions (%) to the combined uncertainty budget estimated for the dissolved Fe level measured by FI-CL in the GD sample from the GEOTRACES campaign (symbols as in Table 2). The intermediate result refers to the parameters used in Eq. 1 of Table 2, in which all associated uncertainties are included. The GUM Workbench dedicated software package (Metrodata GmbH 2003) was used for the uncertainty propagation calculations.

Quantity		Gravimetric loading	
		Peak height	Peak area
Average normalised signal intensity for sample $\bar{J}_S$ ( $V \text{ kg}^{-1}$ )	Intermediate result	29.5	44.4
	$\bar{J}_{R,S}$ (treated as constant)	-	-
	$\delta_{\text{rep}_S}$	7.9	9.4
	$\delta_{\text{stab}_S}$	21.6	35.0
Average normalised signal intensity for blank $\bar{J}_B$ ( $V \text{ kg}^{-1}$ )	Intermediate result	0.6	1.4
	$I_B$ (treated as constant)	-	-
	$\bar{m}_S$	0.0	0.0
	$\delta_{\text{rep}_B}$	0.0	0.6
	$\delta_{\text{stab}_B}$	0.1	0.0
	$\delta_{\text{matrix}_B}$	0.5	0.8
Sensitivity coefficient (or slope) $F$ ( $V \text{ nmol}^{-1}$ )	Intermediate result	69.7	54.3
	$F_{\text{reg}}$	69.7	54.3
	$\delta_{\text{matrix}_\text{std}}$	0.0	0.0

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 component (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 five spiked levels, Table 4). 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 Eq. 7:

$$U(C_s) \approx 2 \cdot C_s \sqrt{\frac{\bar{J}_S^2 \cdot \left[ \left( \frac{u(\delta_{\text{rep}_S})}{\delta_{\text{rep}_S}} \right)^2 + \left( \frac{u(\delta_{\text{stab}_S})}{\delta_{\text{stab}_S}} \right)^2 \right]}{(\bar{J}_S - \bar{J}_B)^2} + \left( \frac{u(F)}{F} \right)^2} \quad (7)$$

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 Eq. 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 vs. 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 is unlikely to be related to an error in the placement of the baseline for integration, as this affects height less than area (Dyson 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 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 (Supporting Information Table S1). Area integration is considered the “true” measure of the amount of solute (Dyson 1998) and possible problems specific to peak area data such as peak overlap and/or low signal-to-noise

ratios (Dyson 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 (at least for the FI-CL manifold and chemistry described here), 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<sub>2</sub>. 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. From a technological perspective the use of piston pumps (such as micro-sequential injection) has the potential to alleviate some of the issues highlighted in the present manuscript that are associated with the use of peristaltic pumps (Oliveira et al. 2015).

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 10–15% for seawater samples containing 0.5–1 nmol kg<sup>-1</sup> 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 uncertainties 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 five plus the “zero” standard measured with six replicates) and measurements repeated regularly for the same sample over the entire analytical sequence. In view of the long term instability observed during the “stability experiment” a practical recommendation is to analyse a check standard every 2 h and recalibrate if the value is outside of a specified range, e.g., ± 5%. It is also shown that more attention needs to be paid to the way FI-CL peak data are collected and processed, as this could lead to significant errors with respect to the size of the combined uncertainties. To enhance the transparency of these aspects it is recom-

mended that more comprehensive descriptions of the methods used to validate the measurement procedures (including the way peak data collection/processing is performed) are included in publications and reports. Moreover, a simple equation to approximately estimate the uncertainty has been proposed, which is valid if the blank levels are 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. *Anal. Chim. Acta* **442**: 1–14. doi:10.1016/S0003-2670(01)01091-1
- Analytical Methods Committee. 1995. Uncertainty of measurement: Implications of its use in analytical science. *Analyst* **120**: 2303–2308. doi:10.1039/AN9952002303
- Bowie, A. R., and others. 2003. Shipboard analytical inter-comparison of dissolved iron in surface waters along a north–south transect of the Atlantic Ocean. *Mar. Chem.* **84**: 19–34. doi:10.1016/S0304-4203(03)00091-4
- 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. *Limnol. Oceanogr.: Methods* **2**: 42–54. doi:10.4319/lom.2004.2.42
- Bowie, A. R., and others. 2006. A community-wide inter-comparison exercise for the determination of dissolved iron in seawater. *Mar. Chem.* **98**: 81–99. doi:10.1016/j.marchem.2005.07.002
- Boyd, P. W., and M. J. Elwood. 2010. The biogeochemical cycle of iron in the ocean. *Nat. Geosci.* **3**: 675–682. doi:10.1038/ngeo964
- Bucciarelli, E., S. Blain, and P. Tréguer. 2001. Iron and manganese in the wake of the Kerguelen Islands (Southern Ocean). *Mar. Chem.* **73**: 21–36. doi:10.1016/S0304-4203(00)00070-0
- Clough, R., H. Sela, A. Milne, M. C. Lohan, S. Tokalioglu, and P. J. Worsfold. 2015. Uncertainty contributions to the measurement of dissolved Co, Fe, Pb and V in seawater using flow injection with solid phase preconcentration and detection by collision/reaction cell-quadrupole ICP–MS. *Talanta* **133**: 162–169. doi:10.1016/j.talanta.2014.08.045
- Dyson, N. 1998. Errors in peak area measurement, Chapter 2, p. 35–88. *In* Chromatographic integration methods. The Royal Society of Chemistry. doi:10.1039/9781847550514
- GEOTRACES. 2013. Dissolved iron—values in nmol/kg Consensus values (± 1 std. dev.) for SAFe Reference Samples as of May 2013 [accessed 2014 July 15]. Available from <http://www.geotraces.org/science/intercalibration/322-standards-and-reference-materials>
- Gledhill, M. B., and K. N. Buck. 2012. The organic complexation of iron in the marine environment: A review. *Front. Microbiol.* **3**: 1–17. doi:10.3389/fmicb.2012.00069

- Inczedy, J., T. Lengyel, A. Ure, A. Gelencsér, and A. Hulanicki. 1998. *Compendium of analytical nomenclature*. Blackwell Science Ltd.
- ISO/IEC. 2005. *General requirements for the competence of testing and calibration laboratories*, ISO/IEC 17025:2005, International Organization for Standardization/International Electrotechnical Commission, Geneva.
- JCGM 100. 2008. *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, available from [www.bipm.org](http://www.bipm.org), Paris.
- JCGM 200. 2012. *International vocabulary of metrology—basic and general concepts and associated terms (VIM)—3rd ed., 2008 version with minor corrections*; Joint Committee for Guides in Metrology (BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, ILAC), BIPM, available from [www.bipm.org](http://www.bipm.org), Paris.
- Johnson, K. S., and others. 2007. Developing standards for dissolved iron in seawater. *Eos Trans. Am. Geophys. Union* **88**: 131–132. doi:[10.1029/2007EO110003](https://doi.org/10.1029/2007EO110003)
- Klunder, M., P. Laan, R. Middag, H. de Baar, and J. Van Ooijen. 2011. Dissolved iron in the Southern Ocean (Atlantic sector). *Deep-Sea Res. Part II: Top. Stud. Oceanogr.* **58**: 2678–2694. doi:[10.1016/j.dsr2.2010.10.042](https://doi.org/10.1016/j.dsr2.2010.10.042)
- 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. *Limnol. Oceanogr.: Methods* **4**: 164–171. doi:[10.4319/lom.2006.4.164](https://doi.org/10.4319/lom.2006.4.164)
- Martin, J., and S. Fitzwater. 1988. Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. *Nature* **331**: 341–343. doi:[10.1038/331341a0](https://doi.org/10.1038/331341a0)
- Mawji, E., and others. 2008. Hydroxamate siderophores: Occurrence and importance in the Atlantic Ocean. *Environ. Sci. Technol.* **42**: 8675–8680. doi:[10.1021/es801884r](https://doi.org/10.1021/es801884r)
- 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. doi:[10.1039/an9911600003](https://doi.org/10.1039/an9911600003)
- 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. doi:[10.1038/nature02550](https://doi.org/10.1038/nature02550)
- Moore, J. K., S. C. Doney, D. M. Glover, and I. Y. Fung. 2001. Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep-Sea Res.* **49**: 463–507. doi:[10.1016/S0967-0645\(01\)00109-6](https://doi.org/10.1016/S0967-0645(01)00109-6)
- 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. *Anal. Chem.* **65**: 1524–1528. doi:[10.1021/ac00059a007](https://doi.org/10.1021/ac00059a007)
- Oliveira, H. M., M. M. Grand, J. Ruzicka, and C. I. Measures. 2015. Towards chemiluminescence detection in micro-sequential injection lab-on-valve format: A proof of concept based on the reaction between Fe(II) and luminol in seawater. *Talanta* **133**: 107–111. doi:[10.1016/j.talanta.2014.06.076](https://doi.org/10.1016/j.talanta.2014.06.076)
- 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. *J. Anal. At. Spectrom.* **22**: 608–615. doi:[10.1039/b700540g](https://doi.org/10.1039/b700540g)
- Press, W. H. 2012. *Numerical recipes in C: The art of scientific computing*. Cambridge University Press.
- 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, p. 257. *In* G. Holland and S. Tanner [eds.], *Plasma source mass spectrometry: The new millennium*. 7th Durham Conference Royal Society of Chemistry, Cambridge.
- Shelley, R. U., B. Zachhuber, P. N. Sedwick, P. J. Worsfold, and M. C. Lohan. 2010. Determination of total dissolved cobalt in UV-irradiated seawater using flow injection with chemiluminescence detection. *Limnol. Oceanogr.: Methods* **8**: 352–362. doi:[10.4319/lom.2010.8.352](https://doi.org/10.4319/lom.2010.8.352)
- Ussher, S. J., E. P. Achterberg, and P. J. Worsfold. 2004. Marine biogeochemistry of iron. *Environ. Chem.* **1**: 67–80. doi:[10.1071/EN04053](https://doi.org/10.1071/EN04053)
- Ussher, S. J., M. Yaqoob, E. P. Achterberg, A. Nabi, and P. J. Worsfold. 2005. Effect of model ligands on iron redox speciation in natural waters using flow injection with luminol chemiluminescence detection. *Anal. Chem.* **77**: 1971–1978. doi:[10.1021/ac048850a](https://doi.org/10.1021/ac048850a)
- 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. *Mar. Environ. Res.* **70**: 46–55. doi:[10.1016/j.marenvres.2010.03.001](https://doi.org/10.1016/j.marenvres.2010.03.001)
- 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). *Environ. Chem.* **7**: 139–145. doi:[10.1071/EN09092](https://doi.org/10.1071/EN09092)

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. doi:[10.1126/science.1059251](https://doi.org/10.1126/science.1059251)

Wyatt, N. J., and others. 2014. Biogeochemical cycling of dissolved zinc along the GEOTRACES South Atlantic transect GA10 at 40°S. *Glob. Biogeochem. Cycles* **28**: 44–56. doi:[10.1002/2013GB004637](https://doi.org/10.1002/2013GB004637)

#### **Acknowledgments**

GF acknowledges J. Snell and T. Linsinger for fruitful discussions. 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 DISCO-SAT) from the European Commission.

*Submitted 7 December 2014*

*Revised 25 June 2015*

*Accepted 21 July 2015*

*Associate editor: Gregory Cutter*