

A community-wide intercomparison exercise for the determination of dissolved iron in seawater

Andrew R. Bowie ^{a,b,c,*}, Eric P. Achterberg ^c, Peter L. Croot ^d, Hein J.W. de Baar ^d,
Patrick Laan ^d, James W. Moffett ^e, Simon Ussher ^c and Paul J. Worsfold ^c

^a *Antarctic Climate & Ecosystems CRC, Hobart, Tasmania, Australia*

^b *Australian Centre for Research on Separation Science, University of Tasmania, Hobart, Tasmania, Australia*

^c *School of Earth, Ocean and Environmental Sciences, University of Plymouth, Plymouth, United Kingdom*

^d *Royal Netherlands Institute for Sea Research, Texel, The Netherlands*

^e *Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole (MA), USA*

*Corresponding author. E-mail: Andrew.Bowie@utas.edu.au, Fax: +61-3-62262973

Last revised 09 August 2005

Submitted to Marine Chemistry

Abstract

The first large-scale international intercomparison of analytical methods for the determination of dissolved iron in seawater was carried out between October 2000 and December 2002. The exercise was conducted as a rigorously “blind” comparison of 7 analytical techniques by 24 international laboratories. The comparison was based on a large volume (700 L), filtered surface seawater sample collected from the South Atlantic Ocean (the “IRONAGES” sample), which was acidified, mixed and bottled at sea. Two 1 L sample bottles were sent to each participant. Integrity and blindness were achieved by having the experiment designed and carried out by a small team, and overseen by an independent data manager. Storage, homogeneity and time-series stability experiments conducted over 2.5 years showed that inter-bottle variability of the IRONAGES sample was good (<7 %), although there was a decrease in iron concentration in the bottles over time (from 0.8-0.5 nM) before a stable value was observed. This raises questions over the suitability of sample acidification and storage.

For the complete dataset of 45 results (after excluding 3 outliers not passing the screening criteria), the mean concentration of dissolved iron in the IRONAGES sample was 0.59 ± 0.21 nM, representing a coefficient of variation (%CV) for analytical comparability (“community precision”) of 36 % (1s), a significant improvement over earlier exercises. Within-run precision (5-10 %), inter-run precision (15 %) and inter-bottle homogeneity (<7 %) were much better than overall analytical comparability, implying the presence of: (1) random variability (inherent to all intercomparison exercises); (2) errors in quantification of the analytical blank; and (3) systematic inter-method variability, perhaps related to secondary sample treatment (e.g. measurement of different physicochemical fractions of iron present in seawater) in the community dataset. By grouping all results for the same method, analyses performed using flow injection – luminol chemiluminescence (with FeII detection after sample reduction) [Bowie et al., 1998. *Anal. Chim. Acta* 361, 189] and flow injection – catalytic

spectrophotometry (using the reagent DPD) [Measures et al., 1995. Mar. Chem. 50, 3] gave significantly ($P=0.05$) higher dissolved iron concentrations than analyses performed using isotope dilution ICPMS [Wu and Boyle, 1998. Anal. Chim. Acta 367, 183]. There was, however, evidence of scatter within each method group (CV up to 59 %), implying that better uniformity in procedures may be required. This paper does not identify individual data and should not be viewed as an evaluation of single laboratories. Rather it summarises the status of dissolved iron analysis in seawater by the international community at the start of the 21st century, and can be used to inform future exercises including the SAFE iron intercomparison study in the North Pacific in October 2004.

Keywords

Iron, seawater, determination, intercomparison, IRONAGES, large volume sample

1. Introduction

The last quarter century has witnessed a revolution in our understanding of trace metal distributions in the World's oceans and one element, iron, has changed our thinking more than any other (Coale et al., 1999). In order to understand the factors controlling the functioning of marine ecosystems and their effect on carbon cycling (Moore et al., 2002), it is imperative that iron is measured routinely and accurately during oceanographic expeditions. During an international symposium of SCOR-IUPAC Working Group 109 on the "Biogeochemistry of Iron in Seawater" held in Amsterdam in November 1998, it became apparent that concentrations of iron in the World's surface oceans varied over several orders of magnitude (de Baar and de Jong, 2001). The lack of rigorous intercomparison exercises, quality control of trace metal data and appropriate certified reference materials (CRMs) for sub-nanomolar concentrations of iron (the National Research Council of Canada NASS-5 solution contains 3.71 ± 0.63 nM Fe – at least tenfold greater than typical open-ocean concentrations) has meant that the scientific community has little ability to correlate these observations, and distinguish between environmental variability, analytical data quality and measurement drift. Moreover, such uncertainties in the global distribution of iron preclude the development of accurate biogeochemical models of iron limitation.

Although attempts to measure iron in seawater extend back to the 1930s (Cooper, 1935), it is recognised that modern methods began with publication of a technique based on chelation-solvent extraction followed by analysis using graphite furnace atomic absorption spectrometry (GFAAS) (Danielsson et al., 1978). Further improvements in clean sampling techniques (including not using iron hydrowires; Betzer and Pilson, 1975) were reported by Bruland et al. (1979), using a similar extraction method developed independently. The use of this analytical method resulted in reported dissolved iron concentrations in seawater decreasing by up to two orders of magnitude (Achterberg et al., 2001), and enabled oceanographers to obtain the first

reliable open-ocean profiles (Gordon et al., 1982). The last two decades of the 20th century saw a great deal of interest in developing new and improved land and shipboard techniques for measuring iron in seawater, many based on portable flow injection (FI) systems (Worsfold et al., 2002); to date, at least 10 different analytical methods have been reported. Whilst errors in reported iron concentrations may result from contamination or analyte loss during sampling (an extremely challenging task on research vessels partly constructed from iron), filtration or post-collection preservation (e.g. acidification) and storage of samples, it is essential to be able to attribute differences in reported concentrations for different analytical methods using a common sampling, treatment and storage protocol.

Only two, relatively small-scale, intercomparison exercises for trace metals have previously been conducted, both somewhat unsuccessful for iron compared to other elements. During a study sponsored by the International Council for the Exploration of the Sea (Bewers et al., 1981), poor inter-laboratory precision was reported over the range 14.5-31.5 nM Fe for the analysis of acidified samples by seven different laboratories. On a 1990 Intergovernmental Oceanographic Commission baseline survey (Landing et al., 1995), three laboratories reported iron data for a single profile in the North Atlantic. Results showed up to an order of magnitude degree of variability over the concentration range 0.1-5.2 nM Fe. More recently, Measures and Vink (2001) and Bowie et al. (2003) have conducted shipboard intercomparison exercises for iron using a variety of popular analytical methods. Their results indicate that concentration offsets are due to the different sensitivities of the methods to organic and colloidal fractions of iron present in seawater, and specifically the differential extraction of these fractions during preconcentration. Bowie et al. (2004) have further argued that extended storage (>6 months) of filtered, acidified seawater samples may be necessary to enable full detection of colloidal and organic iron fractions present within the dissolved phase. It remains clear, however, that

additional larger-scale exercises are required, including a thorough examination of each stage of the overall procedure (sample collection, filtration, storage, extraction and analysis).

Therefore, to promote a more standardised international capability for the measurement of iron in seawater, a second workshop was held in San Antonio in January 2000, where plans for the collection and distribution of a large volume (approximately 700 L), low iron sample taken from the surface Atlantic Ocean were formulated (hereafter known as the “IRONAGES” sample). This material was filtered, acidified, mixed and sub-sampled at sea, and subsequently distributed to interested parties for laboratory analyses. Thirty-one laboratories in 11 different countries took part in this SCOR-IUPAC sponsored exercise. Sample bottles were coded to protect laboratory anonymity. Twenty-four laboratories reported data (Table 1), using 7 different analytical techniques. These methods offered a wide range of complexity, cost, transportability and experience, and collectively represented the best available technology at the time for measuring dissolved iron in seawater.

Results from this “blind” intercomparison of the IRONAGES sample were initially submitted to an independent data manager, who was not actively involved with the measurement of dissolved iron in seawater. At a third workshop held in San Francisco in December 2002, data were disseminated to the iron community, revealing which laboratory generated which data, and resulting in a lively and open exchange of ideas. The conclusions reached at this meeting were integrated into a preliminary report which was distributed to the National Science Foundation and SCOR in April 2003. It is envisaged that the results from this timely exercise will lead to reliable methods for the routine production of commercially-available CRMs suitable for low level, open-ocean iron measurements. Although the IRONAGES intercomparison has now closed, samples are still available on request from the University of Plymouth, in order to aid new analysts or calibrate new instruments.

This paper presents technical information on the preparation of the intercomparison exercise, shipboard sampling procedures and subsequent distribution of the IRONAGES sample. We report the community-wide results of this exercise and attempt to correlate reported differences with procedural and methodological variations. In addition, data from a series of long-term stability, homogeneity and storage studies on a subset of 10 IRONAGES samples are presented, and recommendations for follow-up exercises outlined.

2. Materials, methods and data handling

2.1 Design of the intercomparison

The primary goal of the first intercomparison was to obtain a set of at least 200 homogeneous and stable samples containing filtered seawater with a low iron concentration (<1.0 nM). Integrity and blindness were achieved by having the experiment designed and managed by a small team from the University of Plymouth (UK) and the Royal Netherlands Institute for Sea Research (NIOZ). In June 2001, two 1 L bottles containing IRONAGES samples were sent to each of the 31 laboratories who expressed an interest in joining the exercise. No blank or pure water reference samples were included in this preliminary study. Participants were requested to report results by 1 July 2002 to an independent data manager (James Moffett, Woods Hole Oceanographic Institution). When the samples were sent, information on the shipboard collection, filtration and bottling procedures were outlined, although no concentration range was provided. Data were reported on a standard spreadsheet which contained the following information: bottle number, iron concentration (in nM), standard deviation (1s), number of replicates, method blank, detection limit (3s), date of analysis, analytical method and literature reference for the method. These raw data were further processed by the first author and sent to participants in February 2003 to verify that no errors had been made in the data transcription.

Submitted results from each bottle were treated independently, rather than calculating a mean for each laboratory, in order to aid in the identification of outlier bottles. Table 2 shows the timetable of events during the IRONAGES intercomparison exercise.

2.2 Cleaning

Prior to departure, a 1040 L polyethylene (HDPE) cubic collection vessel supported within a polyurethane coated cage (Eco Fut, Van Leer, The Netherlands) was washed at NIOZ using 1-2 % DECON solution (1 week) and subsequently filled to the brim with ~20 % HCl (AnalaR) and left for 6 weeks. This tank was drained and thoroughly rinsed five times with ultra-high purity (UHP) water prior to shipping and all taps and inlets securely sealed. Unfortunately, on arrival at the research vessel's departure port, it was discovered that the tank had been opened (probably at customs), most likely resulting in its inadvertent contamination. Therefore, the tank was also rinsed three times at sea using unfiltered, surface (1-2 m) open-ocean seawater delivered from a towed torpedo fish and a trace metal clean pumping system (Bowie et al., 2003). For each shipboard rinse, the seawater was acidified to pH <2 with HCl (AnalaR, Merck BDH), mixed and left for at least 2 days prior to draining. The tank was completely covered with a double layer of polyethylene sheeting at all times.

Two hundred 1 L LDPE (Nalgene) sample bottles were cleaned at the University of Plymouth following procedures documented in Table 3 in Achterberg et al. (2001). This washing protocol consisted of room temperature immersion in a mild DECON (1-2 %) solution followed by 2 weeks in each of 6 M HCl and 3 M HNO₃ (both Aristar grade, BDH Merck), with thorough (5x) UHP water rinses between each stage. Bottles were stored filled with ~0.01 M quartz-distilled HCl (Q-HCl) until sampling. Bottle blanks (i.e. the concentration of iron in ~0.01 M Q-HCl in UHP water) were checked at regular intervals during the cleaning and were satisfactory (<30 pM Fe).

2.3 *Sampling location*

The IRONAGES bulk sample was collected from the Atlantic Ocean during cruise ANT XVIII/1 of R/V Polarstern in October 2000. During a north-south transect of the eastern Atlantic from 27°N to 19°S, underway surface dissolved iron levels were monitored by 4 separate research groups using different analytical methods (Bowie et al., 2003). On passage through the equatorial South Atlantic, surface dissolved iron concentrations reported by all 4 teams were uniformly low (0.15 ± 0.08 nM; $n=26$). The bulk sample was therefore collected in this region between 05°20.5' S, 06°11.9' W and 06°44.8' S, 05°04.8' W. The mean surface salinity and temperature during collection were 35.74 ± 0.06 and 23.6 ± 0.1 °C, respectively.

2.4 *Sample collection*

Surface (1-2 m) seawater was fed directly into the HDPE cubic tank using an underway sampling system consisting of a towed polyurethane-coated torpedo-shaped fish (1 m long, 50 kg weight), fitted with a Teflon FEP nose tube and deployed off the crane arm of a hydrographic winch at a distance of approximately 5 m from the ship's starboard side (Bowie et al., 2003). The fish was towed at 13-14 knots during filling of the tank. Seawater was pumped on-board through acid-washed braided PVC tubing using a variable speed high volume peristaltic pump (model 7591-00, Cole Palmer Instrument Co.), fitted with silicone pump tubing and filtered through a Sartobran-P polypropylene cartridge unit with a cellulose acetate filter membrane (0.45 µm pre-filter and 0.2 µm final filter, Sartorius Ltd.). The tube from the fish was connected directly to a dispensing tap at the base of the tank and a 0.2 µm PTFE membrane sterile air venting filter unit (Millex-FG) attached to a hole in the lid of the tank to release pressure build-up during the filling. The tank was filled at a rate of 1.5-1.8 L min⁻¹ to approximately 700 L, between 02:30 and 10:30 h on 16 October 2000. Every hour the flow to the tank was reduced as the seawater line was partially diverted (using a flow splitter) for ~10

min in order to collect 1 L of seawater for the on-going shipboard intercomparison exercise reported in Bowie et al. (2003).

2.5 *Preservation and bottling of the bulk sample*

The bulk sample was acidified to approximately pH 2.0 using 700 mL of ~10 M Q-HCl (quadruple sub-boiled quartz-distilled HCl, prepared in-house at NIOZ) and the seawater gently mixed by shaking and rocking the tank on the ship's deck. After a period of 1 day (during which sub-aliquots were taken from the tank for shipboard analysis), the sample was pumped through acid-washed Teflon FEP tubing leading directly from the tank into a laminar flow hood housed in a class-100 clean container. The first 100 L of sample was used for rinsing the tubing and discarded. Two hundred 1 L sample bottles were then filled on 19-20 October 2000, after 3 thorough rinses with the water from the tank. The exterior of the bottles were allowed to dry before sealing within two zip-lock plastic bags. Bottles were labelled sequentially, but were coded in order to prevent bias during analysis. Bottles #200-280 (batch 1) were first filled, followed by bottles #001-120 (batch 2). These samples were stored in the dark at room temperature at the University of Plymouth until distribution. Every 5th sample was collected in separate 1 L bottles prepared at NIOZ which were not distributed as part of the intercomparison exercise. After every 10th bottle, the effluent was sub-sampled in pre-cleaned 100 mL HDPE bottles (Emergo) for immediate shipboard measurement.

2.6 *Shipboard and time-series analyses*

At sea, the concentration of iron in the acidified seawater in the tank was monitored during the cleaning steps and the final fill. Analyses were performed using two versions of flow injection – luminol chemiluminescence: FI-CL [FeII] (with FeII detection after reduction of Fe(III) to Fe(II) using sodium sulfite; Bowie et al., 1998; 2002) and FI-CL [FeIII] (with FeIII detection after natural oxidation (>1 h) of sample, using H₂O₂ as an added oxidant in the manifold; de

Jong et al., 1998). Laboratory-based homogeneity, stability and storage studies were initiated at the University of Plymouth in January 2001. Ten bottles were randomly selected, five from batch 1 (#204, 216, 236, 243, 276) and five from batch 2 (#027, 053, 072, 084, 117). These studies were continued at the University of Tasmania in Hobart (Australia) at approximately 1-2 month intervals from March 2001 until October 2002, and again in March and July 2003. All time-series analyses were performed by a single scientist (A. Bowie) using the same FI-CL [FeII] instrumentation, which required only a small volume (~20 mL) of sample water for analysis (Bowie et al., 1998).

2.7 Distribution, community analysis and methods

Two individual bottles of the IRONAGES bulk sample were shipped to 31 worldwide laboratories participating in the intercomparison, with 24 reporting data (Table 1). Some of the investigators are no longer at the institution indicated. Seven laboratories received samples but either did not analyse them or did not report the results. Most laboratories analysed samples in the 2 bottles originally distributed, although 5 laboratories requested extra bottles. Some samples were analysed by more than one laboratory (and sometimes by more than one method), although each laboratory reported data using only one analytical method. In general, to standardise the exercise, where a laboratory provided results from the analysis of additionally requested samples, data from only 2 randomly-selected bottles were used for processing. In total, results from 48 sets of analyses are reported here. The number of analyses on sub-samples taken from each bottle and the number of replicates per sub-sample varied between laboratories.

Twenty-one laboratories analysed their samples and presented results within the 13-month reporting window (1 June 2001 – 1 July 2002), although at different times within this period. Laboratory 15 analysed samples in April 2001, laboratory 22 analysed samples in July 2002 and laboratory 25 analysed samples in October 2002. Laboratories 22 and 25 therefore reported

their data after the official deadline for results and hence after the community-wide results had been disseminated to the group. Their results are therefore not blind but have been included in the overall dataset because they add to the overall discussion. Ten samples were used for the time-series studies. The stability of the samples is discussed below.

Seven different analytical methods were used in the intercomparison exercise (Table 3). Three methods were based on flow injection (FI) systems (with chemiluminescence or spectrophotometric detection), one on cathodic stripping voltammetry (CSV), whilst the other three were based on laboratory-based GFAAS or inductively coupled plasma mass spectrometry (ICP-MS). All methods required some form of analyte preconcentration and/or matrix elimination, either using a resin micro-column, a competing ligand, batch solvent extraction or a co-precipitation step. Every method (except CSV) was used by at least two laboratories. Flow injection with luminol chemiluminescence detection (with FeIII detection after natural oxidation of sample), following variants of the method originally reported by Obata et al. (1993), was the most popular method and was used by 8 laboratories. Many of the methods listed have been slightly modified from the exact design reported in the original paper, and some laboratories that used essentially identical methods often operated their instruments in subtly different ways. In particular, differences in sample buffering and pre-treatment were highlighted as key issues at the San Francisco workshop. Furthermore, many laboratories have now found new and improved ways of purifying their reagents to lower blanks and detection limits. Further details can be provided by the principal investigators listed in Table 1.

2.8 *Screening criteria for results*

Two iterations of Grubb's test (recommended by the International Organisation for Standardisation) were used to check for suspected outlier data (i.e. values at the ends of the range), based on the assumption that the population had a Gaussian distribution. In total, 83

bottles were mailed out to participants. Of the 48 sets of reported results, 3 did not pass the screening criterion ($P=0.05$) and were excluded from the overall dataset (6% of the total). These values are reported in the table of results (Table 4), but are not included in the data processing, figures or discussion below. These outliers, all at the high extreme of the range, may have arisen from inadvertent contamination of the bottle on sample filling or handling in the investigator's laboratory. There does not appear to be a serious problem of sporadic gross contamination of samples. Every method used in the intercomparison had at least one set of analyses with "passing" data.

3. Results and discussion

3.1 Integrity of tank cleaning procedures

The concentration of iron in the IRONAGES bulk sample was monitored during shipboard cleaning and rinsing. A sub-aliquot of acidified unfiltered seawater was taken from the dispensing tap of the tank after the first shipboard fill and clean process. The concentration of iron in the tank at this time was ~50 nM (analysis using FI-CL [FeIII]). After a second drain, rinse, fill and acidification (AnalaR HCl, Merck BDH) cycle, 10.4 nM of iron was present in the bulk seawater sample (analysis using FI-CL [FeII]). After a third clean, the tank sample contained 3.11 nM of iron (analysis using FI-CL [FeII]). From the assay provided by Merck BDH for this batch of AnalaR grade HCl, the maximum contribution from the iron impurities in the acid to the concentration of iron in the tank sample would be ~4.2 nM, and therefore the elevated iron concentration in the bulk sample was most likely due to impurities added during sample acidification. The tank was subsequently filled with filtered seawater and acidified to ~0.01M Q-HCl (Section 2.5) prior to sub-sampling into the 1 L bottles for the intercomparison exercise.

3.2 Shipboard verification

The concentration of iron in sub-samples of effluent taken from the tank was measured after every 10th bottle collected during the filling process (“time zero” analyses). Values ranged from 0.21 to 0.35 nM with a mean of 0.26 ± 0.03 nM ($n=21$; RSD=12 %) (analysis using FI-CL [FeIII]; Section 2.6) (Figure 1), confirming the homogeneity of the bulk sample inside the tank and that the mixing process was satisfactory. Sample bottle #011 contained 0.23 nM. The iron concentration in surface seawater of this region during filling of the tank ranged from 0.02 to 0.31 nM, with a mean of 0.15 ± 0.08 nM (Bowie et al., 2003), indicating that there was no evidence of severe contamination during sampling.

In addition, on 19 October 2000, three days after the final filling of the tank with filtered seawater, a sub-aliquot of the acidified bulk sample was collected from the dispensing tap. This sample was reduced using sodium sulfite and analysed on five separate occasions over the next 10 h using FI-CL[FeII] (Section 2.6). Iron levels were initially 0.2-0.3 nM (consistent with the iron concentrations in the tank effluent), but were observed to increase to 0.53 nM with increasing time of sulfite reduction (and increasing acidification time) (Figure 2(a)). It is likely that the observed increase was due to the quantitative reduction of Fe(III) to Fe(II) in the sample by the added sulfite. Stringent clean procedures were followed at every stage of the sample collection, handling and analysis.

3.3 Homogeneity

The results of the homogeneity studies on a subset of 10 randomly-selected bottles (five from batch 1 and five from batch 2) of the IRONAGES bulk sample are shown in Figure 3. Analyses were carried out using FI-CL[FeII] (Section 2.6) on 6 March 2002, a date within the 13-month reporting window for the community analyses. One sample (bottle #117) was contaminated (*Q*-test, $P=0.05$) and has thus been rejected as an outlier. Results from the remaining 9 bottles

(Fe = 0.49 ± 0.03 nM, RSD=6.5 %) showed that there was no significant difference (*t*'-test, P=0.05) in iron concentration between samples, nor was there any correlation between values obtained and bottle sequence. Thus it is assumed that the iron concentrations of all bottles, as measured by one analytical technique, are not significantly different and the IRONAGES bulk sample was homogeneous. It is important to note that it is not possible to state categorically that any one individual sample bottle does not have a higher iron concentration due to inadvertent contamination at some stage of the overall sampling process.

3.4 Stability

The results of the time-series stability studies conducted between January 2001 and July 2003 are shown in Figure 2(b). All 9 bottles were analysed on each occasion (except July 2003 when only 2 bottles were analysed; Section 2.6), yielding the means and standard deviations in Figure 2(b). The concentration of iron (0.81 ± 0.03 nM) measured in the sample bottles on 24 January 2001 (100 days after sampling) was noticeably higher than that observed at the end of the shipboard analysis (0.53 ± 0.03 nM), representing a 53 % increase. Since the seawater entering the tank was filtered (0.2 μ m), the concentration increase was presumably due to the release of iron into solution (after acidification) from particles transferred from the tank into the 1 L bottles during sub-sampling. Higher iron concentrations were observed in all 9 sample bottles and thus the low-level contamination was associated with the tank itself, and not individual 1 L bottles. It appears likely that the tank cleaning procedures adopted were not rigorous enough for sampling open-ocean seawater for trace iron, or that the tank was inadvertently contaminated during transit or on-board ship. This highlights the extreme difficulties faced when attempting to cleanly sample and preserve a large volume (hundreds of litres) seawater sample for trace (sub-nM) iron determination.

Interestingly, over the following 10 months (January–November 2001, up to 400 days after sampling), the concentration of iron in the IRONAGES sample bottles decreased from 0.81 ± 0.03 to 0.53 ± 0.06 nM, representing a 35 % reduction. We are uncertain whether this decrease was due to the loss of iron to the walls of the sample bottles (unlikely since the bottles were made from LDPE and the sample was acidified to pH ~ 2.0) or the conversion of the iron into an analytically undetectable form. Independent experiments (Bowie et al., 2004) however, suggest the contrary in that storage of acidified seawater samples over periods of months to years results in the solubilisation/dissociation of colloidal or organic iron species. Towards the latter months of the stability trials (March 2002 – July 2003), the concentration of iron in the IRONAGES bottles converged to a stable concentration (0.48 ± 0.03 nM). There does not appear to be any relationship between the ratio of air-to-seawater in the bottles and the iron concentration, although clearly the number of times the bottles are opened and re-sealed increases the chance of inadvertent sample contamination. We investigate any relationship between the community-wide results and sample stability below (Section 3.6). Stability trials will continue beyond 2004 and until the supply is exhausted. Bottle blanks (i.e. the concentration of iron in bottles that were not filled with the IRONAGES sample) remained low (<30 pM Fe) and stable throughout.

In order to investigate this apparent loss of iron over time, two additional experiments were performed. Firstly, on 18 October 2002, sub-aliquots taken from 4 bottles (#027, 216, 236, 276) of the IRONAGES sample were transferred to acid-cleaned quartz tubes and subjected to 6 h UV oxidation using a 1.1 kW Hg vapour lamp. The concentration of iron in these samples was compared to their control counterparts (sub-aliquots from the same 4 bottles but without UV oxidation). No significant differences (t -test, $P=0.05$) were observed in the UV oxidised samples (Fe = 0.54 ± 0.08 nM) compared with the controls (Fe = 0.49 ± 0.05 nM). Secondly, two samples (#027, 276) were subjected to further acidification with the addition of an extra 1 mL

of ~10 M Q-HCl per 1 L of seawater, resulting in an approximate final pH of 1.7. The extra acid was added to both 60 mL sub-aliquots of these samples as well as to the original bottles themselves. The concentration of iron in these samples was compared to two control bottles (#072, 243) on 29 October 2002, after a one month storage period. Results show that there was no significant difference (*t*-test, $P=0.05$) between the samples with extra acid ($\text{Fe} = 0.46 \pm 0.03$ nM) and the controls ($\text{Fe} = 0.40 \pm 0.04$ nM). There was also no significant difference between extra acid treatments in the 60 mL sub-aliquots and original 1 L LDPE bottles.

3.5 Effect of storage

During the stability studies, three different storage conditions were adopted. Sample bottle #216 was stored in the fridge (~4 °C) in the dark. Sample bottle #236 was stored under ambient light at room temperature (15-25 °C). All other samples (controls) were stored in the dark at room temperature. There was no significant difference between samples stored in the dark compared with those stored in the light, both at room temperature ($P=0.05$; Figure 4). However, the changes in iron concentration found in the sample stored in the fridge were significantly different than the controls ($P=0.05$), showing a more rapid decrease over time between January and November 2001 compared with the samples stored at room temperature. The long-term gradual increase in iron concentration (from 1.1 to 1.8 nM) in sample bottle #117 confirms that this sample was contaminated.

3.6 Community results

Data for all 48 community analyses of dissolved iron in the IRONAGES sample are listed in Table 4, showing the mean value \pm one standard deviation (1 SD) and ancillary figures of merit for each analysis. The reported values were derived from separate calibration and blank quantification for each laboratory. Since the IRONAGES samples are considered here for the purpose of intercomparison only (rather than oceanographic study), the acid blank associated

with sample acidification has not been included in the determination by any laboratory; i.e. the acid blank has *not* been subtracted from the measured sample concentration. Previous work (Bowie et al., 2004) has shown this acid blank to be small (<30 pM).

When plotted in order of increasing iron concentration (Figure 5), the distribution shows a broadly “S”-type shape, with no clear evidence of clustering. This is similar to the results of earlier DOC (Hedges et al., 1993) intercomparison experiments, where much of the difference were attributed to uncertainty in determining the instrument blank. The variability of all analyses was 36 % (coefficient of variation, expressed as a percentage of the mean based on one standard deviation, %CV), with a mean iron concentration of 0.59 ± 0.21 nM (n=45) and a range of 0.29-1.19 nM (3 outliers excluded). A histogram of the community results (Figure 6) binned over the range 0.2-1.2 nM show the data do not follow a normal (Gaussian) distribution, and have a slight positive skew. This may be the result of the superposition of several normal distributions (for each method) with similar means but different variances. There was no correlation between iron concentration and bottle number or order of collection (data not shown), confirming the earlier measurements of dissolved iron in the tank effluent during bottling (Section 3.2), and demonstrating that the bottle filling process was not biased. A negative (albeit weak) correlation existed between iron concentration and date of analysis, consistent with the results from the stability trials, which showed a decrease in iron concentration in all sample bottles over the period January-November 2001 (Section 3.4). During, the 13-month community analysis and reporting window (1 June 2001 - 1 July 2002), a maximum decrease of 0.2 nM Fe was observed in the stability trials (from 0.68-0.48 nM; Figure 2(b)). Since the magnitude of this linear decrease in iron concentration was considerably less than the range of reported data, stability changes alone cannot account for the observed community-wide differences, although they may have contributed to the variability.

Method blanks and detection limits

Detection limits (3s) ranged from 0.009 to 0.12 nM (community mean of 0.043 nM) and method blanks ranged from 0.013 to 0.5 nM (community mean of 0.115 nM) (Table 4). Most laboratories reported their detection limit as “3s of a low iron sample or blank”, although 2 laboratories defined their detection limit as “3x baseline noise”. Eight of the 24 laboratories reported high blank-to-iron concentration ratios (>20 %), which were independent of method type. Since a substantial amount of the random error may reside in the blank, the precision of low-level data for these methods will be particularly affected and these laboratories should focus on ways of minimising their method blank. We acknowledge that differences in the definition of the blank existed during this exercise due to differences in methodology.

Interestingly, those laboratories reporting a high blank precision did not also report high sample precision or high blank-to-iron concentration ratio. There was a general increase in reported iron concentration with increases in method blank and standard deviation, which is not wholly unexpected, although there was also no correlation between iron concentration and detection limit. Future exercises should include a thorough assessment of instrument and procedural blanks, and attempt to standardise these important parameters.

Within- and inter-run variability: comparability with community precision

The precision of analysis for one individual sample (%RSD, “within-run variability”) varied between 1.1 and 21.1 % (n = 2-14) (community mean of 6.3 %) (Table 4). The mean precision for each method (mean of within-run variabilities for laboratories using the same method) ranged between 4.9 and 10.3 % (Table 5). “Inter-bottle variability” was 6.5 % (Section 3.3), although two laboratories (#15 and 18) did report large inter-bottle differences that were not evident in the homogeneity trials. The variability between sample runs (“inter-run variability”) during the community analysis and reporting window was determined in the stability trials using the same 10 sample bottles analysed under identical laboratory conditions using FI-CL

[Fe II] (Section 3.4). Excluding one outlier (bottle #117), inter-run variability was 15.2 %.

Thus whilst the within- and inter-run contributions to systematic variability were broadly comparable, they were generally much lower than both the %CV for each method group and the overall community precision of 36.0 % (Table 5). This suggests the following: (1) the presence of random variability (inherent to all intercomparisons) in the community dataset; (2) that better control of standardisation (calibration) and blank quantification is required; and (3) the presence of systematic inter-method differences, perhaps related to secondary sample treatment (e.g. measurement of different physico-chemical fractions of iron present in seawater; see “Examination of differences in method types” below).

Examination of differences in method types

To evaluate differences between methods, data were pooled into groups of similar analytical techniques and compared using a two-tailed ‘*t*’-test (based on homoscedastic distributions) applied to the pairs of means of each group. Interestingly, the 7 analytical methods can be broadly grouped into 2 categories, according to mean iron concentration (Table 5 and Figure 7): (1) FI-CL [FeIII], ID-ICPMS, CSV-DHN and SE-GFAAS (Fe range = 0.53-0.58 nM); and (2) FI-DPD, SPE-ICPMS and FI-CL [FeII] (Fe range = 0.71-0.74 nM). FI-CL [FeII] and FI-DPD gave significantly ($P=0.05$) higher values than ID-ICPMS. No other pooled means were significantly different from each other at the 95 % confidence interval. The %CV for SE-GFAAS and SPE-ICPMS were large (49-59%), in contrast to the two FI-CL methods (34-37 %), and CSV-DHN, FI-DPD and ID-ICPMS which gave less divergent values (<13 %). Since a relatively small number of analyses were done within each group ($n=2-16$) and the variability was large for some groups, it is not surprising that more discrimination between methods is not evident. The %CV within each method type was often larger than overall analytical comparability, making it difficult to attribute systematic inter-method differences in reported iron concentrations to the measurement of different physico-chemical fractions of iron present

in seawater. This is best achieved through more focussed method intercomparison experiments (e.g. de Jong et al., 2000; Measures and Vink, 2001; Bowie et al., 2003; 2004).

The 7 analytical methods used during this intercomparison are based on widely different chemistries and instrument types. The grouping of methods into 2 distinct categories (based on mean of iron concentration) does not appear to be directly related to the nature or complexity of analysis, with both FI and ICPMS methods spanning the range of reported iron concentrations. Moreover, the more automatic methods with less sample handling do not produce lower iron data or possess smaller within-method %CVs, indicating sample contamination not to be a serious problem. Potential interferences to each analytical technique have been thoroughly examined during method development (see references in Table 3), and are thus unlikely to cause the observed differences. Rather it is probable that subtle variations in: (1) secondary sample preservation (acidification time, strength and type of acid – HCl, HNO₃, formic acid); (2) sample pre-treatment (reduction/oxidation, microwave or UV digestion); (3) sample buffering (type of buffer, pH modifications); and (4) analyte extraction (type of chelating resin or competing ligand competition) contribute to differences in reported data through the recovery and measurement of different physico-chemical fractions of iron in seawater. These procedural variations exist both between and within groups of similar methods.

Several cases highlight the difficulty in attributing differences in reported iron concentrations to differences in method types. First, results obtained using the most popular method (FI-CL [FeIII]) span almost the entire range for the whole exercise (0.23-0.83 nM). Second, laboratory 20 (SE-GFAAS) reported significantly higher data than the other laboratories using this instrumentation, despite the use of an identical solvent extraction step. Third, laboratory 9 (FI-CL [FeII]) adopted a preconcentration step based on modified Mg(OH)₂ co-precipitation (rather than microcolumn extraction using an 8-hydroxyquinoline (8HQ) resin) and reported data

approximately double that of the other laboratories using this method. Fourth, laboratory 4 (using FI-CL [FeIII]) re-analysed bottle #227 after the addition of 10 μM H_2O_2 to the sample and observed a significant increase in iron concentration (0.56 nM compared to a control value of 0.23 nM). This increase was attributed to the re-oxidation of Fe(II) in the sample, which was not detected at low pH in the sample without added peroxide. The last point highlights the importance of complete oxidation (peroxide) or reduction (sulfite) of the sample, depending on the choice of FI-CL method.

Despite these somewhat conflicting observations, it is encouraging that the community data reported here appear broadly consistent with the few intercomparison exercises for dissolved iron in seawater that have taken place to date (i.e. FI-CL [FeIII] \approx CSV-DHN, de Jong et al., 2000; FI-DPD \geq SE-GFAAS, Measures and Vink, 2001; FI-CL [FeII] $>$ FI-CL [FeIII], Bowie et al., 2003; FI-CL Fe[II] \geq FI-DPD, Bowie et al., 2004). In addition, an independent objective inspection of Figure 7 suggests that there may be 5 extreme high data points, #31 & 32 (SE-GFAAS), #39 (SPE-ICPMS), and #44 & 45 (FI-CL [FeII]). Although these results are not “officially” outliers (and thus have been included here in the data processing and figures), their exclusion from the community data would considerably improve the coefficients of variation for these method types and the overall community precision (Table 5). The ranges of all the other methods would also then lie within the range for the most popular method (FI-CL [FeIII]).

4. Summary of exercise

This exercise was essentially a first round intercomparison, designed to establish the range of variability in the measurement of dissolved iron across the present community of marine scientists and to determine if patterns in the measurement values might be related to specific instrument characteristics or data treatments. We are now able to make several important

conclusions about the determination of dissolved iron in seawater. First, at an international level, although only moderate agreement (CV=36 %) was found amongst a diverse range of independent analysts and methods, there has certainly been a vast improvement in community analytical capability over the past decade. The overall range of the IRONAGES sample (0.23-1.19 nM) is unacceptably large though and, if plotted as a single depth profile, these data would appear oceanographically inconsistent. Second, none of the methods appears to be grossly inaccurate, although additional data from certain methods (e.g. CSV-DHN and SPE-ICPMS) would be useful in this regard. Third, the more rapid shipboard techniques (FI and CSV) do not appear to underestimate dissolved iron concentrations compared to the lab-based techniques based on larger instrumentation (GFAAS and ICPMS). Fourth, the inter-method differences in reported dissolved iron concentrations may (in part) be explained by the availability of different physico-chemical fractions of iron present in seawater to the different analytical methods, although such observations are obscured by a high %CV within each method group.

It is important to state that this study was conducted at an iron concentration higher than that observed in high-nutrient low-chlorophyll regions of the ocean, and significantly greater than the half-saturation constant for phytoplankton growth (~0.05-0.3 nM; Fitzwater et al., 1996; Coale et al., 2003). To obtain a trace metal seawater standard at bio-limiting iron concentrations will be a challenge for the next intercomparison exercise. We are still some way off the routine production of a commercially-available CRM for iron in seawater (the current “best” option is NASS-5, which contains 3.71 nM Fe). Furthermore, it is not reassuring that the concentration of dissolved iron in the IRONAGES sample changed over time. Although the initial increase (days-weeks) appears to be due to low-level contamination in the tank, the subsequent gradual decrease (months) in sub-sample bottles is harder to explain and raises important questions regarding wall-loss artefacts and the suitability of the acidification process.

Looking further at the results of this exercise, it should be emphasized that groups of analytical methods were probably not singular methods run in a consistent manner from lab-to-lab, and multiple causes for the disparities remain. This exercise used uniform samples to overcome differences that may be due to sample collection, filtration and preservation. However, different secondary treatments (acidification, extraction, reduction/oxidation, microwave or UV digestion) were used in different laboratories and were an important additional source of variability. Whilst it is beyond the scope of this paper to thoroughly compare the different methods within each group, we believe it is important for analysts to examine the finer details of their analytical techniques. In particular, incomplete reduction (by sulfite) or oxidation (by peroxide) of the sample could result in “missing” iron for the two FI-CL methods. The sample buffering pH, type of column preconcentration material and iron redox chemistry under acidification also appear to be key parameters of the overall analytical process warranting further investigation, as well as the actual chemistry of individual methods.

Moreover, the definition of analytical blank by each investigator is diverse and future exercises should seek to standardise this important parameter. Several of the authors of this paper are currently working on smaller interlaboratory comparison (“round-robin”) exercises to further resolve these differences. While the array of methods used in this intercomparison offer clear differences in detection limit, precision and rapidity of analysis, there are also differences in cost, ease of use, weight and ruggedness that are important to appreciate when an instrument is used in the field. In addition to these factors, the most important consideration when choosing a dissolved iron analytical method may be “what quality of data is required to answer your specific research question?”.

Finally, it should be noted that this experiment was guided by the desire to evaluate why data were disparate, and not just which data were “best” or “most accurate”. This exercise should

not be viewed as a qualification exercise for the analysts, but rather a broad preliminary test of the comparison of each group of methods in routine use worldwide, as well as an important assessment of the problems faced in cleanly collecting, preserving and sub-sampling several hundred litres of seawater. In large part, this was accomplished. Open evaluation exercises such as this one will hopefully stimulate experimentation in individual laboratories that will help to resolve the tenth of nanomolar differences in reported iron concentrations that remain.

5. Set of Recommendations

A follow-up iron intercomparison exercise is planned for October 2004 in the Pacific Ocean. During project “SAFE” (*Sampling and Analysis of Iron*), 32 scientists will openly examine each stage of the complete procedure for the determination of iron in seawater, including sample collection, filtration, storage and analysis. This exercise will focus on a wide variety of sampling systems currently in use, including Go-Flo and Niskin bottles suspended on Kelvar wire, towed torpedo fish (surface water only), trace metal clean rosette systems and automated, in situ closing devices (e.g. MITESS; Bell et al., 2002), as well as scrutinising filtration procedures (membranes, cartridges, syringe units). Here, we present a set of recommendations for this and other future exercises from lessons learnt during the IRONAGES intercomparison.

Firstly, the clean collection and preservation of a large volume sample is not a trivial task, whilst being key to the overall exercise. During the IRONAGES exercise, changes in iron concentration in both the tank and the sample bottles themselves hindered interpretation of the data. As well as the obvious step of stringent cleaning of the large volume collection unit, in-line re-filtering of the acidified sample during bottling may be considered, using a filter cartridge resistant to mild acid conditions. Protocols for stabilisation of the seawater sample

must also be reviewed, including the use of stronger acidification. Individual acidification of each sample bottle is not recommended.

The IRONAGES exercise focused on a single surface sample. Since the presence of colloidal and organic material in the sample is likely to affect the recovery of dissolved iron by different analytical methods, it will be important in the future for the community to compare surface and deep samples from open-ocean and coastal environments, where there will be contrasting iron and ligand concentrations. Moreover, the distribution of a “blank sample” or augmented seawater standard alongside a homogenised sample may aid in unravelling methodological differences.

Finally, there is currently no consensus on the preparation or use of commercially available iron standards. Some laboratories use solid Fe(II) or Fe(III) hydrated salts, others atomic absorption standards (e.g. BDH Merck ‘Spectrosol’). Factors such as standard acidification, matrix conditions for the standard (pure water vs. seawater), and frequency of preparation/dilution vary amongst analysts. An additional source of variance could be inaccurate micropipette calibration, which will result in data yielding excellent precision (and trends in a profile), but inaccurate absolute iron concentrations. Future work should regularly exchange these materials to ensure consistency and stability of iron standards. Since the speciation of the iron standards is unlikely to be representative of the forms of dissolved iron present in seawater, the approach to instrument calibration will also be critical.

Once a consensus on the best form of sampling, filtration, preservation, storage and analysis for the determination of dissolved iron in seawater is achieved, it would undoubtedly be useful for the wider community for such information to be reported in an overall technical manual.

Acknowledgements

The overall project was organised under the Terms of Reference (task description) of SCOR/IUPAC Working Group 109 (Biogeochemistry of Iron in Seawater). SCOR and NSF (Grant No. OCE-0003700 to SCOR) kindly provided financial support for three workshops in Amsterdam (1998), San Antonio (2000) and San Francisco (2002). The European Union provided support for the fieldwork under the IRONAGES project (EVK2-1999-00031). Laboratory studies were funded by the Australian Research Council (X00106765 and DP0342826), ACROSS and the Australian Government's Cooperative Research Centres Programme through the Antarctic Climate and Ecosystems Cooperative Research Centre (ACE CRC). Final preparation of this manuscript was assisted by funding from NERC grant NER/A/S/2003/00489. We thank Charlotte Braungardt and Derek Henon for help with cleaning the IRONAGES bottles, Stéphane Blain, Géraldine Sarthou and Marie Boye for sample bottling at sea, and Bill Landing and two anonymous reviewers for providing constructive comments on the paper. This exercise was only possible through the generous cooperation of many individuals (listed in Table 1) who agreed to run the samples in a blind format, and contributed data and comments on this manuscript.

References

- Achterberg, E.P., Holland, T.W., Bowie, A.R., Mantoura, R.F.C., Worsfold, P.J., 2001. Determination of iron in seawater. *Anal. Chim. Acta* 442, 1-14.
- Bell, J., Betts, J., Boyle, E., 2002. MITESS: a moored *in situ* trace element serial sampler for deep-sea moorings. *Deep-Sea Res. I* 49, 2103-2118.
- Betzer, P.R., Pilson, M.E.Q., 1975. The effect of corroded hydrographic wire on particulate iron concentrations in seawater. *Deep-Sea Res. I* 22, 117-120.
- Bewers, J.M., Dalziel, J.A., Yeats, P.A., Barron, J.L., 1981. An intercalibration for trace metals in seawater. *Mar. Chem.* 10, 173-193.
- Bowie, A.R., Achterberg, E.P., Mantoura, R.F.C., Worsfold, P.J., 1998. Determination of sub-nanomolar levels of iron in seawater using flow injection with chemiluminescence detection. *Anal. Chim. Acta* 361, 189-200.
- Bowie, A.R., Achterberg, E.P., Sedwick, P.N., Ussher, S., Worsfold, P.J., 2002. Real-time monitoring of picomolar concentrations of iron(II) in marine waters using automated flow injection-chemiluminescence instrumentation. *Environ. Sci. Technol.* 36, 4600-4607.
- Bowie, A.R., Achterberg, E.P., Blain, S., Boye, M., Croot, P.L., de Baar, H.J.W., Laan, P., Sarthou, G., Worsfold, P.J., 2003. Shipboard analytical intercomparison of dissolved iron in surface waters along a north-south transect of the Atlantic Ocean. *Mar. Chem.* 84, 19-34.
- Bowie, A.R., Sedwick, P.N., Worsfold, P.J., 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.

- Bruland, K.W., Franks, R.P., Knauer, G.A., Martin, J.H., 1979. Sampling and analytical methods for the determination of copper, cadmium, zinc and nickel at the nanogram per litre level in seawater. *Anal. Chim. Acta* 105, 233-245.
- Coale, K.H., Worsfold, P.J., de Baar, H.J.W., 1999. Iron age in oceanography. *Eos, Trans. Am. Geophys. Union* 80, 377-382.
- Coale, K.H., Wang, X.J., Tanner, S.J., Johnson, K.S., 2003. Phytoplankton growth and biological response to iron and zinc addition in the Ross Sea and Antarctic Circumpolar Current along 170 degrees W. *Deep-Sea Res. II* 50, 635-653.
- Cooper, L.H.N., 1935. Iron in the sea and in marine plankton. *Proc. Roy. Soc. London* B118, 419-438.
- Danielsson, L.G., Magnusson, B., Westerlund, S., 1978. An improved metal extraction procedure for the determination of trace metals in seawater by atomic absorption spectrometry with electrothermal atomization. *Anal. Chim. Acta* 98, 47-57.
- de Baar, H.J.W., de Jong, J.T.M., 2001. Distributions, Sources and Sinks of Iron in Seawater. In: Turner, D.R., Hunter, K.A. (Eds.), *The Biogeochemistry of Iron in Seawater*. Wiley, New York, chapter 5, pp. 123-253.
- de Jong, J.T.M., den Das, J., Bathmann, U., Stoll, M.H.C., Kattner, G., Nolting, R.F., de Baar, H.J.W., 1998. Dissolved iron at sub-nanomolar levels in the Southern Ocean as determined by shipboard analysis. *Anal. Chim. Acta* 377, 113-124.
- de Jong, J.T.M., Boye, M., Schoemann, V.F., Nolting, R.F., de Baar, H.J.W., 2000. Shipboard techniques based on flow injection analysis for measuring dissolved Fe, Mn and Al in seawater. *J. Environ. Monitor.* 2, 496-502.
- Fitzwater, S.E., Coale, K.H., Gordon, R.M., Johnson, K.S., Ondrusek, M.E., 1996. Iron-deficiency and phytoplankton growth in the Equatorial Pacific. *Deep-Sea Res. II* 43, 995-1015.

- Fujishima, Y., Ueda, K., Maruo, M., Nakayama, E., Tokutome, C., Hasegawa, H., Matsui, M., Sohrin, Y., 2001. Distribution of trace bioelements in the sub-Arctic north Pacific Ocean and the Bering Sea (the R/V Hakuho-Marui Cruise KH-97-2). *J. Oceanogr.* 57, 261-273.
- Gordon, R.M., Martin, J.H., Knauer, G.A., 1982. Iron in northeast Pacific waters. *Nature* 299, 611-612.
- Hedges, J.I., Bergamaschi, B.A., Benner, R., 1993. Comparative analyses of DOC and DON in natural waters. *Mar. Chem.* 41, 121-134.
- King, D.W., Lounsbury, H.A., Millero, F.J., 1995. Rates and mechanism of Fe(II) oxidation at nanomolar total iron concentrations. *Environ. Sci. Technol.* 29, 818-824.
- Landing, W.M., Cutter, G.A., Dalziel, J.A., Flegal, A.R., Powell, R.T., Schmidt, D., Shiller, A., Statham, P.J., Westerlund, S., Resing, J., 1995. Analytical intercomparison results from the 1990 Intergovernmental Oceanographic Commission open-ocean baseline survey for trace metals - Atlantic Ocean. *Mar. Chem.* 49, 253-265.
- Measures, C.I., Yuan, J., Resing, J.A., 1995. Determination of iron in seawater by flow injection analysis using in-line preconcentration and spectrophotometric detection. *Mar. Chem.* 50, 3-12.
- Measures, C.I., Vink, S., 2001. Dissolved Fe in the upper waters of the Pacific sector of the Southern Ocean. *Deep-Sea Res. II* 48, 3913-3941.
- Moore, J.K., Doney, S.C., Glover, D.M., Fung, I.Y., 2002. Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep-Sea Res. II* 49, 463-507.
- Obata, H., Karatani, H., Nakayama, E., 1993. Automated determination of iron in seawater by chelating resin concentration and chemiluminescence detection. *Anal. Chem.* 65, 1524-1528.
- Obata, H., van den Berg, C.M.G., 2001. Determination of picomolar levels of iron in seawater using catalytic cathodic stripping voltammetry. *Anal. Chem.* 73, 2522-2528.

- Wells, M.L., Bruland, K.W., 1998. An improved method for rapid preconcentration and determination of bioactive trace metals in seawater using solid phase extraction and high resolution inductively coupled plasma mass spectrometry. *Mar. Chem.* 63, 145-153.
- Worsfold, P.J., Achterberg, E.P., Bowie, A.R., Cannizzaro, V., Sandford, R.C., Gardolinski, P., 2002. Flow Injection Techniques for the *in situ* Monitoring of Marine Processes. In: Gianguzza, A.; Pelizzetti, E.; Sammartano, S. (Eds.), *Chemistry of Marine Water and Sediments*. Springer-Verlag (Heidelberg), chapter 16, pp. 385-402.
- Wu, J.F., Boyle, E.A., 1998. Determination of iron in seawater by high-resolution isotope dilution inductively coupled plasma mass spectrometry after $\text{Mg}(\text{OH})_2$ co-precipitation. *Anal. Chim. Acta* 367, 183-191.

Figure captions

Figure 1. Concentration of iron (nM) in sub-samples of effluent taken from the tank after every 10th bottle during the filling process. The sample code indicates the point at which the effluent was sampled. Samples were acidified (in the tank) for 1 day before analyses, which was performed using FI-CL [FeIII] (Section 2.6). The horizontal lines represent the mean \pm 1 SD of the dataset.

Figure 2.

(a) Stability of the IRONAGES sample in a sub-aliquot taken from the tank, as measured on-board ship between days 2.8 and 3.2 after collection. Data show a 10 h time-series of analyses performed on a single sample taken from the tank and left sitting after addition of sulfite to the “whole bottle”. They do not represent different sub-aliquots taken from the tank and reduced individually at different times.

(b) Stability of the IRONAGES sample over the period January 2001 to July 2003. Each data point represents the mean \pm 1 SD of 9 randomly-selected bottles (4-14 %), which was generally greater than the analytical uncertainty for replicate analyses of one bottle (n=3, 6 %). Dates show the start of each month. All analyses by FI-CL [FeII].

Figure 3. Results of homogeneity studies for the IRONAGES samples (analyses on 6 March 2002). Error bounds represent \pm 1 SD of replicate analyses of sub-aliquots taken from the same bottle at the same time (n=3, 6 %). Sample #117 was contaminated (Section 3.3).

Figure 4. Effect of different storage conditions on the concentration of iron in the IRONAGES samples. The increase in iron concentration over time in sample bottle #117 (previously shown to be contaminated) has been shown for comparison purposes.

Figure 5. Community results ordered by increasing dissolved iron concentration (nM), 3 outliers excluded. Error bounds represent ± 1 SD of replicate analyses of each bottle (n given in Table 4).

Figure 6. Histogram of results of iron analysis of the IRONAGES samples, binned into 0.1 nM increments.

Figure 7. Dissolved iron concentrations (nM) in the IRONAGES samples by all laboratories, grouped by method. Samples are ordered in increasing mean concentration for each method type and for each analysis within a group. The order on the *x*-axis does not correspond to the bottle or laboratory numbers. Adjacent pairs represent duplicate bottles analysed by the same participant (except laboratory 10 which analysed a combined sample from bottles #231 and 264). Method averages and the overall averages are also shown on the right hand side of the graph. Error bounds represent ± 1 SD on replicate analyses of each bottle (n=1-14, Table 4) or by each method (n=2-16, Table 5).

Table 1. Alphabetical list of the 24 laboratories submitting results to the IRONAGES iron intercomparison.

Principal investigator	Other personnel	Affiliation	Location
Blain, Stéphane	Sarthou, Géraldine	Université de Bretagne Occidentale	Brest, France
Bowie, Andrew		Antarctic CRC; University of Tasmania	Hobart, Australia
Boyle, Ed	Bergquist, Bridget	Massachusetts Institute of Technology	Cambridge, USA
Coale, Kenneth	Gordon, Mike	Moss Landing Marine Laboratories	Moss Landing, USA
Croot, Peter	de Baar, Hein; Laan, Patrick	Royal Netherlands Institute for Sea Research	Texel, The Netherlands
Frew, Russell		University of Otago	Dunedin, New Zealand
Johnson, Ken	Elrod, Virginia	Monterey Bay Aquarium Research Institute	Moss Landing, USA
Kuma, Kenshi		Hokkaido University	Hokkaido, Japan
Landing, William		Florida State University	Tallahassee, USA
Lewis, Brent	Luther III, George	Kettering University; University of Delaware	Flint, USA; Delaware, USA
Measures, Chris	Brown, Matt	University of Hawai'i	Honolulu, USA
Murray, James	Paul, Barbara	University of Washington	Seattle, USA
Nishioka, Jun	Takeda, Shigenobu	Central Research Institute of Electric Power Industry	Chiba, Japan
Obata, Hajime ^a	Nakayama, Eiichiro; Takano, Michiaki; Doi, Takashi	University of Tokyo; University of Shiga Prefecture	Tokyo, Japan; Hikone, Japan
Sañudo-Wilhelmy, Sergio	Tovar-Sanchez, Antonio	Stony Brook University	New York, USA
Sedwick, Peter		Bermuda Biological Station for Research	Bermuda
Sherrell, Rob		Rutgers, State University of New Jersey	Piscataway, USA
Sohrin, Yoshiki		Kyoto University	Uji, Japan
Statham, Peter		Southampton Oceanography Centre	Southampton, UK
Takeda, Shigenobu ^a	Obata, Hajime	University of Tokyo	Tokyo, Japan
Wells, Mark		University of Maine	Orono, USA
Wong, C.S.	Johnson W. Keith; Sutherland, Nes	Institute of Ocean Sciences	Sidney, Canada

Worsfold, Paul
Wu, Jingfeng

Achterberg, Eric; Ussher, Simon

University of Plymouth
University of Alaska

Plymouth, UK
Fairbanks, USA

^a Data submitted by these investigators represent two different laboratories and two different analytical techniques

Table 2. Timetable of events during the IRONAGES iron intercomparison.

Time	Event
January 2000	Plans for IRONAGES intercomparison formulated at a SCOR Working Group 109 workshop
April 2000 - September 2000	Lab-based preparation of bottles, tank and sampling equipment
6-12 October 2000	Shipboard cleaning of tank
8-19 October 2000	Shipboard intercomparison of surface iron concentrations in the eastern Atlantic Ocean (4 laboratories) (see Bowie et al., 2003)
16 October 2000	Shipboard sampling of the bulk sample
18 October 2000	Preservation (acidification) and mixing of bulk sample
19-20 October 2000	Filling of 200 sample bottles
16-21 October 2000	Shipboard analyses of bulk sample
January 2001 – July 2003	Lab-based homogeneity, stability and storage studies
June 2001	World-wide distribution of sample bottles
June 2001 - June 2002	Reporting of results to independent data manager
1 July 2002	Reporting deadline
18-29 October 2002	UV oxidation and additional sample acidification experiments
5 December 2002	Discussion of results at a SCOR Working Group 109 workshop
February 2003	Raw data sent to participants to check for errors

Table 3. Analytical methods used during the IRONAGES iron intercomparison. Laboratory numbers are given in Table 4.

Method code	Acronym	Summary	Laboratory(s) using this method	Reference(s)
A	CSV-DHN	Competitive ligand equilibration - cathodic stripping voltammetry (ligand: 2,3-dihydroxynaphthalene)	1	Obata and van den Berg (2001)
B	FI-CL [FeII]	Flow injection – luminol chemiluminescence (using dissolved O ₂ , sulfite reduction to FeII); preconcentration on 8HQ resin	9 ^a , 19, 21	King et al. (1995); Bowie et al. (1998)
C	FI-CL [FeIII]	Flow injection - luminol chemiluminescence (using H ₂ O ₂ in manifold, natural oxidation to FeIII); preconcentration on 8HQ resin	2, 4, 5, 6, 11, 12, 14, 22	Obata et al. (1993); de Jong et al. (1998)
D	FI-DPD	Flow injection - catalytic spectrophotometry (reagent: <i>N,N</i> -dimethyl- <i>p</i> -phenylenediamine dihydrochloride); preconcentration on 8HQ resin	13, 15	Measures et al. (1995)
E	ID-ICPMS	Mg(OH) ₂ co-precipitation, isotope dilution - inductively coupled plasma mass spectrometry	7, 8, 23, 24	Wu and Boyle (1998)
F	SE-GFAAS	Chelation solvent extraction - graphite furnace atomic absorption spectrometry (ligand: APDC/DDDC)	3, 16, 18, 20	Bruland et al. (1979)
G	SPE-ICPMS	Solid phase extraction - inductively coupled plasma mass spectrometry	10, 17 ^b	Wells and Bruland (1998); Fujishima et al. (2001)

^a Laboratory 9 used the Mg(OH)₂ co-precipitation method of Wu and Boyle (1998) for preconcentration. The method was modified to reduce 40 mL of sample to 10 mL, for a ~4x preconcentration

^b Laboratory 10 added the ligand bis(2-hydroxyethyl) dithiocarbamate (HEDC) to seawater samples and recovered the metal-ligand complex on a C₁₈ column; laboratory 17 used a column containing 8HQ immobilised on a silica gel (MAF) (developed by Obata et al., 1993)

Table 4. Community results for the IRONAGES iron intercomparison (data in nM). Method details are given in Table 3. Certificate values for National Research Council of Canada open-ocean CRMs NASS-4 and NASS-5 are included for information purposes only and have not been included in the summary statistics at the foot of the table. Outlier data (Grubb's test, $P=0.05$; shown in square brackets) have not been included in the calculations.

Lab #	Method	Bottle #	Fe concentration (nM)	SD (nM)	RSD (%)	Number of replicates	Blank (nM)	Detection limit (3s) (nM)	Date of analysis (dy/mo/yr)
1a	A	217	0.58	0.023	4.0	5	0.035	0.030	29-Nov-01
1b	A	222	0.55	0.046	8.4	5	0.035	0.030	29-Nov-01
2a	C	223	0.47	0.030	6.4	5	0.045	0.020	15-Dec-01
2b	C	252	0.39	0.020	5.1	3	0.045	0.020	15-Dec-01
3a	F	229	0.38	0.080	21.1	2	0.032	0.069	21-Dec-01
3b	F	271	0.36	0.060	16.7	2	0.032	0.069	21-Dec-01
4a	C	232	0.29	0.022	7.7	13	0.035	0.060	10-Aug-01
4b	C	227	0.23 ^a	0.012	5.2	4	0.035	0.060	24-Sep-02
5a	C	234	0.37	0.015	4.1	8	0.013	0.010	20-Dec-01
5b	C	273	0.40	0.010	2.5	8	0.013	0.010	20-Dec-01
6a	C	203	0.63	0.045	7.1	4	0.180	0.031	30-Nov-01
6b	C	208	0.53	0.032	5.9	5	0.180	0.031	30-Nov-01
7a	E	267	0.51	0.050	9.8	3	0.160	0.060	29-Apr-02
7b	E	211	0.52	0.020	3.8	3	0.160	0.060	29-Apr-02
8a	E	247	0.68	0.040	5.9	14	0.240	0.060	15-Jan-02
8b	E	202	[6.79]	0.210	3.1	3	0.240	0.060	15-Jan-02
8c ^b	E	214	0.58	0.030	5.2	6	0.240	0.060	15-Jan-02
9a	B	263	1.03	0.145	14.0	2	0.212	0.100	05-Oct-01
9b	B	248	1.06	0.071	6.8	2	0.212	0.100	05-Oct-01
10a ^c	G	231/264	1.19	0.070	5.9	4	0.120	0.060	04-Jan-02
11a	C	244	0.77	0.065	8.4	3	0.180	0.014	21-Jan-02

11b	C	246	0.82	0.070	8.5	3	0.180	0.014	21-Jan-02
12a	C	224	0.83	0.009	1.1	3	0.041	0.009	11-Nov-01
12b	C	241	0.74	0.011	1.5	3	0.041	0.009	11-Nov-01
13a	D	273	0.74	0.070	9.5	4	0.100	0.020	20-Mar-02
13b	D	277	0.76	0.110	14.5	4	0.100	0.020	20-Mar-02
14a	C	226	0.63	0.017	2.7	6	0.036	0.023	28-Nov-01
14b	C	242	0.66	0.013	2.0	6	0.036	0.023	28-Nov-01
15a	D	243	0.58	0.044	7.5	2	0.057	0.013	20-Apr-01
15b	D	204	0.74	0.055	7.4	2	0.057	0.013	20-Apr-01
16a	F	269	[6.76]	ND ^d	ND	ND	0.300	ND	ND
16b	F	254	[1.98]	0.100	5.1	ND	0.300	ND	ND
17a	G	251	0.47	ND	ND	ND	0.083	0.084	20-Nov-01
17b	G	261	0.46	ND	ND	ND	0.049	0.084	20-Nov-01
18a	F	201	0.53	0.028	5.3	2	0.029	0.036	28-Jun-02
18b	F	279	0.34	ND	ND	1	0.029	0.036	28-Jun-02
19a	B	236	0.50	0.024	4.9	3	0.075	0.029	06-Mar-02
19b	B	276	0.52	0.048	9.2	3	0.075	0.029	06-Mar-02
20a	F	212	0.92	0.040	4.3	2	0.500	0.120	03-Dec-01
20b	F	238	0.94	0.040	4.3	2	0.500	0.120	03-Dec-01
21a	B	213	0.69	0.046	6.7	4 ^e	0.234	0.017	12-Sep-01
21b	B	233	0.62	0.018	2.9	4 ^e	0.159	0.017	12-Sep-01
22a	C	052	0.35	0.032	9.3	3	0.035	0.020	12-Jul-02
22b	C	078	0.36	0.006	1.6	3	0.035	0.020	12-Jul-02
23a	E	267	0.52	0.020	3.8	6	0.210	0.060	20-Nov-01
23b	E	211	0.52	0.030	5.8	5	0.210	0.060	20-Nov-01
24a	E	209	0.48	0.016	3.2	8	0.052	0.051	17-Oct-02
24b	E	258	0.46	0.007	1.4	7	0.052	0.051	17-Oct-02

NASS-4	1.88	0.145	7.7		
NASS-5	3.71	0.315	8.5		
Mean	0.59 nM		6.3 %	0.115 nM	0.043 nM
SD	0.21 nM				
%CV	36.0 %				
Range	0.23-1.19 nM		1.1-21.1 %	0.013-0.5 nM	0.009-0.12 nM
n	45		44	45	45

^a Laboratory 4 reported Fe=0.56 nM in bottle #227 after addition of 10 μ M H₂O₂ to sample

^b Laboratory 8 reported data from a third bottle (due to one of the initial two bottles being contaminated)

^c Laboratory 10 combined the sample water in their two bottles in order to provide sufficient volume for determination

^d ND = not determined or not reported

^e Replicate FI peaks (not complete repeat analyses)

Table 5. Analysis of the IRONAGES sample grouped by 7 method types (data in order of increasing mean iron concentration, in nM). Method details are given in Table 3.

Method code	Acronym	Mean Fe concentration (nM)	Mean RSD (%) ^a	Number of analyses	SD ^b (nM)	%CV ^c
C	FI-CL [FeIII]	0.53	4.9	16	0.20	37.4
E	ID-ICPMS	0.53	4.9	8	0.07	12.9
A	CSV-DHN	0.57	6.2	2	0.02	3.8 ^d
F	SE-GFAAS	0.58	10.3	6	0.28	48.5
D	FI-DPD	0.71	9.7	4	0.08	11.9
G	SPE-ICPMS	0.71	5.9	3	0.42	59.2
B	FI-CL [FeII]	0.74	7.4	6	0.25	33.8
Overall community data		0.59	6.3	45	0.21	36.0

^a Mean precision for replicate analyses by one method type (from Table 4 column 6) (“within-run variability”)

^b One standard deviation of means for each method type

^c Coefficient of variation (%) calculated as 1 SD/mean Fe for each method type

^d Only 1 laboratory reported data using this method