

2019-01-21

Estimating Uncertainties in Oceanographic Trace Element Measurements

Worsfold, Paul

<http://hdl.handle.net/10026.1/13360>

10.3389/fmars.2018.00515

Frontiers in Marine Science

Frontiers Media

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

1 **Estimating Uncertainties in Oceanographic Trace Element Measurements**

2

3 Paul J. Worsfold^{a*}, Eric P. Achterberg^b, Antony J. Birchill^a, Robert Clough^a, Ivo Leito^c, Maeve C. Lohan^d,
4 A. Milne^a and Simon J. Ussher^a.

5

6 ^aBiogeochemistry Research Centre, School of Geography, Earth and Environmental Sciences, Plymouth
7 University, Plymouth, PL4 8AA, United Kingdom.

8 ^bGEOMAR Helmholtz Centre for Ocean Research, Wischhofstraße 1-3, D-24148 Kiel, Germany.

9 ^cInstitute of Chemistry, University of Tartu, Ravila 14a, Tartu, 50411, Estonia.

10 ^dOcean and Earth Science, University of Southampton, National Oceanography Centre, Southampton,
11 SO14 3ZH, United Kingdom.

12

13 **Keywords:** Uncertainty estimation, metrology, trace elements, modelling approach, empirical approach.

14

15 **Abstract**

16 A realistic estimation of uncertainty is an essential requirement for all analytical measurements. It is
17 common practice, however, for the uncertainty estimate of a chemical measurement to be based on the
18 instrumental precision associated with the analysis of a single or multiple samples, which can lead to
19 underestimation. Within the context of chemical oceanography such an underestimation of uncertainty
20 could lead to an over interpretation of the result(s) and hence impact on, e.g. studies of biogeochemical
21 cycles, and the outputs from oceanographic models. Getting high quality observational data with a firm
22 uncertainty assessment is therefore essential for proper model validation. This paper describes and
23 compares two recommended approaches that can give a more holistic assessment of the uncertainty
24 associated with such measurements, referred to here as the “bottom up” or modelling approach and the
25 “top down” or empirical approach. “Best practice” recommendations for the implementation of these
26 strategies are provided. The “top down” approach combines the uncertainties associated with day to day
27 reproducibility and possible bias in the complete data set and is easy to use. For analytical methods that
28 are routinely used, laboratories will have access to the information required to calculate the uncertainty
29 from archived quality assurance data. The determination of trace elements in seawater is a significant
30 analytical challenge and iron is used as an example for the implementation of both approaches using real
31 oceanographic data. Relative expanded uncertainties of 10 – 20% were estimated for both approaches
32 compared with a typical short term precision (rsd) of $\leq 5\%$.

33

34 **Keywords**

35 Uncertainty estimation, iron, seawater, bottom up approach, top down approach.

37 **1. Introduction: The oceanographic context**

38 The 20th century was a productive period for development of analytical techniques for oceanographic
39 chemical measurements and the motivation to increase the number of analytes determined in seawater via
40 innovative analytical applications continues to this day. However, with an ever increasing suite of
41 established methods adapted for global oceanographic studies, more importance is now placed on the
42 determination and documentation of accuracy, repeatability (within laboratory) and reproducibility
43 (between laboratories) and more rigorous uncertainty estimates to accompany reported data.

44

45 The need for increased confidence in analytical measurements has gained prominence in recent years, as
46 data produced by the oceanographic community have become more critical for informing the decisions of
47 governments and international organisations on the functioning of the global climate system, such as the
48 Intergovernmental Panel for Climate Change (IPCC, 2013), the Global Ocean Observing System (GOOS;
49 <http://www.goosocean.org/>) and Optimising and Enhancing the Integrated Atlantic Ocean Observing
50 Systems (AtlantOS; <https://www.atlantos-h2020.eu/>).

51

52 One of the most important sectors within the oceanographic community that requires analytical rigour is
53 time-series datasets at strategic global locations. These critical sites provide the baseline data for many
54 comparisons and can also be used to ground truth and test remote sensing techniques for ocean
55 monitoring. To help maintain standards at these study sites, it was recognised that handbooks were
56 required that describe standard operating procedures for the methods and calibration procedures, such as
57 the Bermuda Atlantic Time-series Study's Analytical Methods Handbook (Knap and Gundersen, 1997)
58 and the Hawaii Ocean Time-series methods (<http://hahana.soest.hawaii.edu/hot/methods/results.html>).
59 More recent handbooks for recommended oceanographic measurements are now available; e.g. the
60 GEOTRACES handbook (<http://www.geotraces.org/sic/intercalibrate-data/cookbook>) and the GO-SHIP
61 Repeat Hydrography Manual (Hood, 2010), which replaced the documented methods of the 1994 WOCE
62 Hydrographic Programme.

63

64 The study of trace elements in the ocean is at the forefront of good practice. Data quality is particularly
65 important as many of the methods used operate at close to their limit of detection (i.e. sub-nanomolar
66 concentrations) for open ocean waters. Results for early oceanographic intercomparison exercises
67 (Bewers et al., 1981; Landing et al., 1995) were inconsistent, with inaccuracies in calibration and
68 variability in the quantification of the analytical blank. Two exercises that focussed on iron were
69 IRONAGES, with samples from the Atlantic Ocean (Bowie et al., 2003; Bowie et al., 2006), and SAFe
70 (Sampling and Analysis of Fe), with samples from the Central North Pacific (Johnson et al., 2006;

71 Johnson et al., 2007). Both exercises provided seawater reference materials (RMs) for the oceanographic
72 community with the latter providing “consensus values” for nine elements, including iron.

73
74 The GEOTRACES programme (<http://www.geotraces.org/>) facilitates the study of marine
75 biogeochemical cycles of elements, with one objective being to undertake intercalibration exercises
76 (Cutter, 2013) in order to achieve the best possible accuracy (i.e. the lowest random and systematic
77 errors). With regard to iron and other trace element measurements, GEOTRACES conducted two
78 intercalibration cruises, one in the North Atlantic Ocean at the BATS (Bermuda Atlantic Time Series) site
79 in 2008 and one at the SAFe site in the oligotrophic North Pacific in 2009, and collected seawater at both
80 sites in order to prepare RMs (GEOTRACES GS surface sample and GEOTRACES GD deep (2000 m)
81 sample). These intercalibration exercises have contributed to improving the accuracy of dissolved iron
82 and other trace element measurements in seawater, thereby enhancing the ability of the oceanographic
83 community to compare datasets that vary temporally and/or spatially or are obtained by different
84 researchers using different analytical methods. To maintain consistent data quality, the GEOTRACES
85 programme continues to recommend that laboratories undertake intercalibration exercises and has
86 produced specific protocols for the sampling and analysis of trace elements
87 (<http://www.geotraces.org/sic/intercalibrate-data/cookbook>). Moreover, a key aspect of intercalibration
88 for GEOTRACES sections is the use of cross-over stations to take into account different sampling
89 systems alongside analytical procedures.

90
91 It is also common practice for chemical oceanographers to estimate bias (i.e. a quantitative estimate of
92 trueness) by analysing a CRM with a certified value or a RM with a consensus value. In addition to using
93 CRM/RM data to estimate bias, the internal instrumental precision is commonly used to estimate the
94 uncertainty of a measurement result. This approach may however underestimate measurement uncertainty
95 because it neglects any contributions from systematic effects. A more robust approach to uncertainty
96 estimation is to use a mathematical model that combines all of the individual uncertainties, including
97 those causing systematic effects. This approach allows the major contributing factors to the overall
98 uncertainty to be identified, as discussed for the determination of dissolved cobalt in seawater (Worsfold
99 et al., 2013) and the determination of ^{210}Po and ^{210}Pb activities in seawater (Rigaud et al., 2013), and also
100 indicates where to focus efforts to meet the target uncertainty. It should be noted that for the collection of
101 oceanographic samples there will also be uncertainties associated with the sampling process and this is
102 discussed elsewhere (Clough et al., 2016).

103
104 The International Organization for Standardization/International Electrotechnical Commission (ISO/IEC,
105 2005) states that “*the range and accuracy of the values obtainable from validated methods (e.g. the*

106 *uncertainty of the results, detection limit, selectivity of the method, linearity, limit of repeatability and/or*
107 *reproducibility, robustness against external influences and/or cross-sensitivity against interference from*
108 *the matrix of the sample/test object), as assessed for the intended use, shall be relevant to the customer's*
109 *needs.” This article focuses on approaches for assessing the uncertainty associated with chemical*
110 *oceanographic measurements and provides examples of good practice for chemical oceanographers to*
111 *enhance the usefulness of their analytical data. It is recommended that these practices be routinely used*
112 *when reporting chemical oceanographic data.*

114 **2. Uncertainty in chemical oceanography**

115 Metrology is a fundamental aspect of chemical measurements (see e.g.
116 <https://sisu.ut.ee/measurement/uncertainty>) and although metrological concepts have increasingly been
117 applied in analytical chemistry in recent years, challenges still remain. The analyte is often determined in
118 the presence of other substances in the sample matrix, with some potentially at higher concentrations than
119 the analyte, and these may contribute to the analytical signal. To achieve sufficient selectivity many
120 analytical methods therefore include at least one separation step to remove interferences but this can also
121 remove a fraction of the analyte, leading to biased results. Similarly, a preconcentration step is often
122 included in the overall method and this can also lead to biased results. Therefore, the main uncertainty
123 contributions in chemical oceanography measurements, particularly for the determination of trace
124 elements, usually come from the seawater sample under investigation rather than the measurement
125 technique itself.

126
127 Comparing experimental results with an independent reference value for the same sample is useful for
128 confirming that the results have acceptable trueness and that the measurement uncertainty estimate is fit
129 for purpose. In addition, good agreement between the experimental result and the reference value
130 suggests that selectivity is probably adequate and robustness is good. The result of such a comparison can
131 be expressed in different ways, e.g. as a zeta or E_n score (ISO/IEC, 2005) or as a bias (Magnusson et al.,
132 2012).

133
134 A CRM is commonly used to validate a method as the certified value(s) are *de facto* reference values.
135 These values should be the same as those required in the target sample(s) and the matrix and
136 concentration range(s) of the analyte(s) in the CRM should be similar to those likely to be encountered in
137 the samples. Often, however, there is no CRM available for the required analyte-matrix-concentration
138 parameters which has led to the “in-house” production of the IRONAGES, SAFe and GEOTRACES
139 seawater RMs. Satisfactory in-house RMs can be obtained from appropriate solutions, e.g. open ocean

140 seawater collected from a cruise, providing that the matrix and any spiked analyte are stable and fully
141 homogenized.

142
143 To evaluate precision or trueness (e.g. using a CRM or RM as described above) it is essential that
144 replicate measurements are made. When performed in a single day replicate measurements enable
145 repeatability (s_r) to be obtained whereas when performed over a longer time period they can be used to
146 determine intermediate precision (s_{Rw}), also known as within-laboratory reproducibility (Floor et al.,
147 2015). For uncertainty estimation, intermediate precision is more useful than repeatability because it takes
148 a larger number of effects into account. This is because effects that are systematic within a single day can
149 become random over a longer time period, e.g. laboratory temperature, personnel, detector response. The
150 longer the measurement time period, the more effects are included and hence the more useful this
151 characteristic becomes. Fewer measurement values collected over a longer time period are therefore
152 better than more measurement values collected over a shorter period, providing that the sample remains
153 stable and homogeneous over the longer period. The same is true when using a CRM to evaluate
154 trueness/bias, i.e. it is better to make, say, four replicate measurements over several weeks rather than on
155 one day. The average value obtained can then be compared with the reference value and/or used for bias
156 calculation.

157
158 Reporting the mean and standard deviation (s.d.) of any measured value is clearly essential. It is also
159 important however to estimate and report the overall uncertainty of observational trace element data to
160 support biogeochemical cycling studies and ensure proper validation of input data for global scale
161 models. The following section therefore describes two approaches for estimating uncertainty known as
162 “bottom up” and “top down”. In both approaches the objective is a realistic assessment of the combined
163 standard uncertainty (u_c), which is often reported as a relative term, i.e. a percentage of the mean (u_{c_rel}).
164 The combined uncertainty takes into account contributions from all of the important uncertainty sources.
165 The combined standard uncertainty represents a probability of approximately 68% (i.e. one standard
166 deviation) and hence a combined expanded uncertainty (U_c) is often calculated by multiplying u_c by a
167 coverage factor (k). A coverage factor of $k = 2$, which represents a probability of approximately 95% (i.e.
168 two standard deviations), is most commonly used, and in relative terms is designated U_{c_rel} .

170 **3. Measurement of iron in seawater – an example application**

171 3.1. The bottom-up approach

172 In the “bottom up” or modelling approach, the standard uncertainties that are associated with each
173 component of the overall measurement procedure are estimated and subsequently combined using
174 uncertainty propagation laws. In this way, the effect of each component of the measurement procedure,
175 e.g. sample manipulation and blank correction, on the combined uncertainty estimate can be calculated.

176 This information can be used as a diagnostic tool to refine the analytical method in order to minimise
177 these effects and hence lower the combined standard uncertainty. Two example studies are used to
178 demonstrate this approach in a chemical oceanography context; (i) the determination of dissolved Co, Fe,
179 Pb and V in seawater using flow injection with solid phase preconcentration (on Toyopearl AF-Chelate-
180 650 resin) and detection by collision/reaction cell-quadrupole ICP-MS (Clough et al., 2015) and (ii) the
181 determination of dissolved iron in seawater using flow injection with chemiluminescence detection (Floor
182 et al., 2015). The flow injection manifolds used for these two examples are shown in Fig. 1 (ICP-MS
183 detection) and Fig. 2 (chemiluminescence detection).

184
185 In the former example, the relative expanded uncertainty (U_{c_rel}) of each analytical result was estimated
186 using the numerical differentiation method of Kragten (1994). This approach is easy to adopt and
187 estimates the effect of each parameter in the measurement equation on the analytical result using a simple
188 spreadsheet. Worked examples can be found in the Eurachem guide (Ellison and Williams, 2012), as well
189 as on the website http://www.ut.ee/katsekoda/GUM_examples/. It is important to provide a clear
190 statement of what is being measured (the measurand) and the measurement equation used to calculate the
191 result. It should be stressed that the conversion of a signal to a concentration must be done carefully,
192 taking into consideration (on a case by case basis) how to deal with non-zero intercepts and non-linear
193 calibration curves. In the first example the measurands were the concentrations of dissolved Co, Fe, Pb
194 and V in seawater and the following equation was used as the model for calculating C_S (the analyte
195 concentration in the sample);

$$C_S = \frac{\left(\frac{(I_S - I_{WB}) \times V_1}{F}\right) - (B_C \times B_V)}{V_2} \quad \text{Equation 1}$$

196
197
198
199 I_S = analyte signal (area measurement). The standard uncertainty for I_S was calculated from the peak area
200 precision ($n = 3$).

201 I_{WB} = wash blank signal (area measurement). The standard uncertainty for I_{WB} was calculated from the
202 peak area precision ($n = 10$).

203 V_I = volume of the sample + added pH buffer solution. The standard uncertainty of V_I was taken from the
204 manufacturers certificates for the pipettes and plastic laboratory ware used.

205 F = slope of the calibration line (i.e. the sensitivity coefficient). The uncertainty of the slope of the
206 calibration line was calculated using regression statistics (Miller and Miller, 2010) and assumed that
207 systematic effects affecting all calibration line points in the same direction (e.g. standard substance purity)
208 were negligible.

209 B_C = analyte concentration in the buffer solution. The standard uncertainty of B_C was taken as the
210 standard deviation of five replicate measurements of the pH adjustment buffer.

211 B_V = volume of the buffer solution. The standard uncertainty of B_V was taken from the manufacturers
212 certificates for the pipettes and plastic laboratory ware used.

213 V_2 = initial sample volume. The standard uncertainty of V_2 was taken from the manufacturers certificates
214 for the pipettes and plastic laboratory ware used.

215 This approach does not take into account other potential systematic effects such as incomplete selectivity
216 and analyte losses. Details of the analytical method can be found elsewhere (Clough et al., 2015).

217

218 The analytical figures of merit for the measurement of iron in three seawater RMs are shown in Table 1
219 and the relative uncertainty contributions of each parameter are given in Table 2. In summary, the relative
220 expanded uncertainty (U_{c_rel}) for the concentration of iron in three RMs obtained by Clough et al. ranged
221 from 8% - 18%, with the largest uncertainty associated with the lowest concentration (GEOTRACES GS;
222 0.50 nmol L⁻¹). For the GS RM, measurement of the sample peak area (I_S) was the major contributor to
223 the overall uncertainty (73%), with lesser contributions from the slope of the calibration line (F ; 8%) and
224 the wash blank (I_{WB} ; 17%). In this instance, a longer sample loading time could potentially lead to a lower
225 uncertainty for I_S but the trade-off would be the time required for each analytical cycle. The certified
226 concentration for the NASS-6 CRM was \approx twenty times higher (9.64 nmol L⁻¹) and in this case the main
227 contributor was the slope of the calibration line (81%), with lesser contributions from the sample peak
228 area and the wash blank.

229

230 Whilst on-line preconcentration with ICP-MS detection is well suited to the determination of iron in
231 seawater in the laboratory, and has the added benefit of simultaneous multi-element detection, it is not
232 suitable for use on-board ship. For such applications, portable techniques such as flow injection with
233 chemiluminescence detection (FI-CL) are preferred (Worsfold et al., 2014). The possible major sources of
234 uncertainty for calculating C_S (the dissolved iron concentration in the sample in ng kg⁻¹) using this
235 approach are shown in an Ishikawa diagram (Fig. 3). The relative expanded uncertainty (U_{c_rel}) for the
236 determination of iron in seawater using FI-CL has been rigorously assessed (Floor et al., 2015). The
237 following equation incorporates all of the uncertainty components used in the model;

238

$$C_S = \frac{\bar{I}_{R_S} \cdot \delta_{rep_S} \cdot \delta_{stab_S} \cdot \delta_{WtoV_S} - \bar{I}_{R_B} \cdot \delta_{stab_B} \cdot \delta_{rep_B} \cdot \delta_{matrix_B}}{F_{reg} \cdot \delta_{matrix_std}} \quad \text{Equation 2}$$

240

241 Where R refers to the raw data, S refers to the sample, B refers to the blank and std refers to the
242 calibration standard. δ terms are unity multiplicative correction factors carrying the relative uncertainty
243 associated with the parameter considered. I is the analytical signal intensity (V), F is the sensitivity
244 coefficient (slope, V L nmol⁻¹) and reg is the sensitivity coefficient (calibration slope) obtained by linear

245 regression. *rep* is the uncertainty arising from the intensity repeatability, *stab* is the uncertainty arising
246 from the intensity stability over an analytical sequence, *WtoV* is the uncertainty related to the difference in
247 loaded mass of the analyte (whether it is done by weighing or volumetrically), and *matrix* is the
248 uncertainty arising from matrix effects on the sensitivity. Whilst the nomenclature is different to that used
249 in Equation 1, it has followed that used in the original publication. Further details of the analytical
250 method can be found in Floor et al. (2015).

251
252 The analytical figures of merit for the measurement of iron with FI-CL in the GS and GD reference
253 materials are shown in Table 3. The combined expanded relative uncertainty for the concentration of iron
254 (U_{c_rel}) was $\approx 12\%$ for peak height measurements and $\approx 10\%$ for peak area measurements. Gravimetric
255 loading gave a slightly lower combined uncertainty compared with volumetric loading (e.g. 12% cf. 13%
256 for GD). RM results for GS and GD were in agreement with consensus values within uncertainty
257 statements based on the methodology reported by Linsinger (2005) in ERM-1, free to download from
258 <https://ec.europa.eu/jrc/en/reference-materials/application-notes>. For comparative purposes, the major
259 relative uncertainty contributions for GD were the within-sequence-stability (intermediate precision;
260 assessed by making 5 measurements, each of 6 replicates, over 32 h) at 22% and the sensitivity
261 coefficient (slope of the calibration line) at 70%. The normalised signal intensity repeatability (i.e. the
262 short term instrumental precision) accounted for only 7.9% of the total uncertainty and highlights the fact
263 that reporting only the instrumental precision can seriously underestimate the overall uncertainty. Floor et
264 al. therefore suggested that it is most beneficial to have a low uncertainty on the calibration slope and
265 hence recommended the use of a sufficient number of replicates (6) and standards (at least the non-spiked
266 standard and 5 spiked levels). They also highlighted the importance of correctly estimating the within-
267 sequence-stability, which should be done under the same measurement conditions as for the samples.
268 They concluded that “*Results obtained indicate that an uncertainty estimation based on the signal*
269 *repeatability alone, as is often done in FI-CL studies, is not a realistic estimation of the overall*
270 *uncertainty of the procedure.*” (Floor et al., 2015).

271 272 3.2. The top-down approach

273 Many laboratories will have historical data on the intermediate precision and analytical bias of a method.
274 These data can easily be applied to the “safe”, (i.e. overestimated) within-laboratory validation approach
275 of measurement uncertainty estimation known as “top down”. Its best-known formalization is the
276 Nordtest™ approach (Magnusson et al., 2012) which uses the following equation:

$$277 \quad u_c = \sqrt{(u(R_w))^2 + u(bias)^2} \quad \text{Equation 3}$$

280 Where u_c is the combined standard uncertainty (approximates to a 68.3% confidence interval), $u(R_w)$ is
281 the uncertainty estimate of within laboratory reproducibility (random effects) and $u(bias)$ is the
282 uncertainty estimate of possible laboratory and procedural bias (systematic effects). This approach
283 therefore combines the uncertainties associated with day to day reproducibility and possible bias in the
284 complete data set and is easy to use. For analytical methods that are routinely used, laboratories will have
285 access to the information required to calculate u_c from archived quality assurance data.

286
287 If a laboratory is implementing a new method it is best to start with a limited objective but new data
288 should be added regularly. So, an intermediate precision value obtained from data collected over four
289 weeks is not sufficient; data should be collected over several months and preferably over at least one year
290 (Magnusson et al., 2012) and data from the analysis of just one CRM is generally not enough for bias
291 evaluation. However, limited data (e.g. intermediate precision data from four weeks and bias estimate
292 from one CRM) can be used as a first approximation. Intermediate precision can be recalculated using
293 longer time intervals, bias can be re-estimated using several reference values and the measurement
294 uncertainty estimate can then be recalculated. Constant improvement in the quality and amount of data is
295 therefore the key to producing reliable analytical results.

296
297 Several examples of the application of the Nordtest™ approach to analytical datasets can be found at
298 http://www.ut.ee/katsekoda/GUM_examples/, including annotated Excel files. The example of most
299 relevance to the oceanographic community estimates the uncertainty associated with the determination of
300 Al, V, Fe and Cd in marine suspended particles using ICP-MS detection. Replicate measurements were
301 performed over 10 months using a plankton matrix CRM. All of the method steps, including sample
302 preparation and ICP-MS determination, were carried out on each day of analysis using a method detailed
303 in Milne et al. (2017) which was a modification of a previously published method (Ohnemus et al., 2014).
304 The random uncertainty component was evaluated via intermediate precision and the systematic
305 component was evaluated using the found and certified values of the CRM. Relative expanded
306 uncertainties (U_{c_rel}) ranged from 16% - 30% (see the Excel file provided in supplementary material S1
307 for further details). A similar approach was used by Rapp et al. to calculate the relative expanded
308 uncertainties (U_{c_rel}), which ranged from 13% - 25%, for the determination of Cd, Co, Cu, Fe, Mn, Ni, Pb
309 and Zn in seawater by on-line preconcentration and high-resolution sector field ICP-MS detection (Rapp
310 et al., 2017).

311
312 The example discussed in more detail in this article uses the same FI-CL manifold (with minor
313 modifications) for the measurement of iron in seawater as was used for the second “bottom up” example
314 (Floor et al., 2015) and therefore provides a direct comparison of the results from the two approaches to

315 uncertainty estimation. The minor manifold differences were a loading pH of 3.5-3.7 (rather than 3.1), a
316 column rinse time of 20 s (rather than 40 s), flow rates for the sample and rinse lines of 1.5 mL min⁻¹
317 (rather than 1.7 mL min⁻¹) and flow rate for the buffer line of 0.6 mL min⁻¹ (rather than 0.7 mL min⁻¹).

318
319 The blank signal associated with the eluent and post-preconcentration column reagents was included in
320 the baseline; therefore if any blank signal was detected it would likely have been from the ammonium
321 acetate buffer, HCl wash and/or the manifold. The blank contribution from these potential sources,
322 determined by shutting off the sample line and loading only buffer, were typically below the limit of
323 detection. This showed that the clean-up columns effectively removed any contribution from the buffer
324 and wash solutions and that the cleaning procedures used helped to maintain a trace metal clean manifold.
325 Further blank contributions could have arisen due to the manipulation of samples e.g. by the addition of
326 H₂O₂ or HCl, but such contributions are negligible if highly pure reagents are used (Bowie et al., 2004;
327 Klunder et al., 2011).

328
329 The accuracy of this method was evaluated using SAFe D1, D2 (deep water) and S (surface water) RMs
330 and the NASS-5 CRM. The results were in agreement with consensus/certified values within uncertainty
331 statements (Linsinger, 2005) as shown in Table 4. Due to the limited quantity of these materials available,
332 internal quality control standards were developed and run daily to assess reproducibility; the results
333 obtained for both the consensus material and the internal quality control standards were then used to
334 calculate a combined uncertainty estimate.

335
336 In this example the dataset was comprised of \approx 2 years of analyses, enabling a more robust estimate of the
337 analytical uncertainty than the short term internal instrumental precision associated with replicate
338 measurements of a single sample (typically <5%). Both RMs with consensus values and internal quality
339 control materials (open ocean seawater with dissolved iron concentrations in the range 0.69 - 1.49 nmol
340 L⁻¹) were used to calculate a “top down” estimate of the combined measurement uncertainty using the
341 Nordtest™ approach. The calculated relative combined standard uncertainty (u_{c_rel}) was 9.5% and hence
342 the relative expanded uncertainty (U_{c_rel}) was 18.9%. For further details see the Excel file provided in
343 supplementary material S2. This uncertainty estimate compares well with a ‘bottom up’ assessment
344 (Floor et al., 2015). Further examination of the combined uncertainty estimate shows that within
345 laboratory reproducibility ($u(R_w)_{rel}$) contributed 65% of the analytical uncertainty, with 35% coming
346 from possible laboratory and procedural bias ($u(bias)_{rel}$). The work of Floor et al. (2015) suggests the
347 main reason for the uncertainty contribution from the within laboratory reproducibility is likely to be
348 associated with the calibration slope for the FI-CL method.

350 The contribution from possible laboratory and procedural bias includes the uncertainty of the published
351 consensus concentrations (e.g. SAFe D2 0.956 ± 0.024 nmol L⁻¹ (1 s.d.)) and the possible bias estimated
352 from the differences between the published mean concentrations and those determined in the course of the
353 analysis reported here (see supplementary material S2). The results of the estimate indicated that the
354 uncertainty associated with the published consensus concentrations ($u(C_{ref_rel})$) contributed 86% of the
355 uncertainty associated with possible laboratory and procedural bias. This is an important point as it shows,
356 for this particular example, that any reduction in the bias of the analysis would have minimal impact on
357 the combined expanded uncertainty (U_c).

358
359 The iron concentration of seawater samples, particularly from transects running from coastal, through
360 shelf to open ocean waters, can span several orders of magnitude (< 0.1 to > 100 nmol L⁻¹ for filtered and
361 unfiltered samples) (Birchill et al., 2017) and the available RMs cover a narrow range of iron
362 concentrations (see e.g. the concentrations of the GEOTRACES and SAFe reference materials). Hence
363 the combined uncertainty estimates obtained using these reference materials would not be applicable over
364 the entire concentration range, even if uncertainty calculations were carried out with relative quantities.
365 Therefore, data for the SAFe S reference material (0.12 ± 0.02 nmol L⁻¹) shown in supplementary
366 material S2 used the short term uncertainty associated with replicate measurements, as is common
367 practice. An assessment of the within laboratory reproducibility can be estimated using in house quality
368 control material. At lower concentrations, e.g. close to the limit of detection, the relative uncertainty
369 increases and hence it is generally recommended that absolute values are used for sample concentrations
370 near the limit of detection when using the NordtestTM. The decision on the threshold for using absolute
371 values should be made based on experience rather than a mathematical algorithm. Using this approach the
372 relative expanded uncertainty (U_{c_rel}) for iron concentrations in seawater was 14% (for 4 - 16 nmol L⁻¹ Fe)
373 and 4% for 49 - 70 nmol L⁻¹ Fe and the expanded uncertainty (U_c) was 0.04 nmol L⁻¹ for 0.14 - 0.24 nmol
374 L⁻¹ Fe.

376 4. Conclusions and recommendations

377 Quantifying the concentration of iron in seawater is undoubtedly challenging, particularly when one
378 considers the different physical and chemical forms of the element and the complexity of the sample
379 matrix. It is therefore imperative that results for the measurement of iron (and other trace elements) are
380 accompanied by a realistic assessment of uncertainty. It is common practice to state an uncertainty based
381 on the internal instrumental precision of replicate measurements of a sample using the method of choice.
382 However a more holistic and robust approach that considers all of the factors contributing to the overall
383 uncertainty provides a more realistic evaluation. This in turn aids interpretation of oceanographic

384 measurements that are used to elucidate biogeochemical cycles and provide input data for oceanographic
385 models.

386
387 The two approaches described in this paper are the “bottom up” or modelling approach and the “top down”
388 or empirical approach. The former (“bottom up”) is a more rigorous but time consuming approach and is
389 therefore not practical for many laboratories. However it has the advantage of providing information on
390 the relative contributions of the different factors to the overall uncertainty. This information is very useful
391 for method development as it pinpoints which factors make the largest contribution and should therefore
392 be targeted for improvement. The latter (“top down”) is easier to apply but does require long term data
393 from the analysis of reference materials. It is important to emphasise however that many laboratories will
394 already have the necessary information from archived quality assurance data to use the Nordtest™
395 approach to calculate the combined uncertainty. Furthermore, this spreadsheet based approach is very
396 easy to use.

397
398 The uncertainties reported for the determination of iron in seawater using FI-CL show good agreement
399 between the two approaches and suggest that a relative expanded uncertainty (U_{c_rel}) of around 10% - 20%
400 is the best that can be achieved, depending on the sample concentration. These values provide a more
401 realistic estimation of uncertainty than values of $\leq 5\%$ that are typically reported for the short term
402 instrumental precision. For further guidance on the estimation of measurement uncertainty in chemical
403 analysis the reader is referred to an on-line course at <https://sisu.ut.ee/measurement/uncertainty>.

405 **Acknowledgements**

406 This paper was supported by funding from the European Union’s Horizon 2020 research and innovation
407 programme under the AtlantOS programme, grant agreement No 633211. Angela Milne and Antony
408 Birchill were supported by funding from the UK Natural Environment Research Council under grants
409 NE/K001779/1 and NE/L501840/1, respectively.

411 **Author contribution statement**

412 PW, EA, ML and SU were co-investigators for the project and IL was a contributor to the project. AB,
413 RC and AM provided data for the case studies. All authors were involved in the preparation and editing of
414 the manuscript.

416 **Conflict of interest statement**

417 There are no conflicts of interest to declare.

- 420 Bowers, J.M., Dalziel, J., Yeats, P.A., and Barron, J.L. (1981). An Intercalibration for Trace-
421 Metals in Seawater. *Marine Chemistry* 10(3), 173-193. doi: 10.1016/0304-
422 4203(81)90040-2.
- 423 Birchill, A.J., Milne, A., Woodward, E.M.S., Harris, C., Annett, A., Rusiecka, D., et al. (2017).
424 Seasonal iron depletion in temperate shelf seas. *Geophysical Research Letters* 44(17),
425 8987-8996. doi: 10.1002/2017gl073881.
- 426 Bowie, A.R., Achterberg, E.P., Blain, S., Boye, M., Croot, P.L., de Baar, H.J.W., et al. (2003).
427 Shipboard analytical intercomparison of dissolved iron in surface waters along a north-
428 south transect of the Atlantic Ocean. *Marine Chemistry* 84(1-2), 19-34. doi:
429 10.1016/S0304-4203(03)00091-4.
- 430 Bowie, A.R., Achterberg, E.P., Croot, P.L., de Baar, H.J.W., Laan, P., Moffett, J.W., et al. (2006).
431 A community-wide intercomparison exercise for the determination of dissolved iron in
432 seawater. *Marine Chemistry* 98(1), 81-99. doi: 10.1016/j.marchem.2005.07.002.
- 433 Bowie, A.R., Sedwick, P.N., and Worsfold, P.J. (2004). Analytical intercomparison between flow
434 injection-chemiluminescence and flow injection-spectrophotometry for the determination
435 of picomolar concentrations of iron in seawater. *Limnology and Oceanography-Methods*
436 2(2), 42-54. doi: 10.4319/lom.2004.2.42.
- 437 Clough, R., Floor, G.H., Quérel, C.R., Milne, A., Lohan, M.C., and Worsfold, P.J. (2016).
438 Measurement uncertainty associated with shipboard sample collection and filtration for
439 the determination of the concentration of iron in seawater. *Analytical Methods* 8(37),
440 6711-6719. doi: 10.1039/c6ay01551d.
- 441 Clough, R., Sela, H., Milne, A., Lohan, M.C., Tokalioglu, S., and Worsfold, P.J. (2015).
442 Uncertainty contributions to the measurement of dissolved Co, Fe, Pb and v in seawater
443 using flow injection with solid phase preconcentration and detection by collision/reaction
444 cell - Quadrupole ICP-MS. *Talanta* 133, 162-169. doi: 10.1016/j.talanta.2014.08.045.
- 445 Cutter, G.A. (2013). Intercalibration in chemical oceanography-Getting the right number.
446 *Limnology and Oceanography-Methods* 11(JULY), 418-424. doi:
447 10.4319/lom.2013.11.418.
- 448 Ellison, S.L.R., and Williams, A. (eds.). (2012). *Eurachem/CITAC guide: Quantifying Uncertainty*
449 *in Analytical Measurement*. Eurachem.
- 450 Floor, G.H., Clough, R., Lohan, M.C., Ussher, S.J., Worsfold, P.J., and Quérel, C.R. (2015).
451 Combined uncertainty estimation for the determination of the dissolved iron amount
452 content in seawater using flow injection with chemiluminescence detection. *Limnology*
453 *and Oceanography: Methods* 13(12), 673-686. doi: 10.1002/lom3.10057.
- 454 Hood, E.M., C.L. Sabine, B.M. Sloyan (2010). "The GO-SHIP Repeat Hydrography Manual: A
455 Collection of Expert Reports and Guidelines, IOCCP Report Number 14, ICPO
456 Publication Series Number 134".).
- 457 IPCC (2013). "Climate Change 2013: The Physical Science Basis. Contribution of Working
458 Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate
459 Change", (ed.) T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A.
460 Nauels, Y. Xia, V. Bex And P.M. Midgley. (Cambridge, UK: Cambridge University Press).
- 461 ISO/IEC (2005). *General requirements for the competence of testing and calibration*
462 *laboratories*. Geneva: International Organization for Standardization/International
463 Electrotechnical Commission.
- 464 Johnson, K., Boyle, E., Bruland, K., Coale, K., Measures, C., and Moffett, J. (2006). Sampling
465 and Analysis of Fe: The SAFE Iron Intercomparison Cruise. *Ocean Sciences Meeting,*
466 *Hawaii, USA*.
- 467 Johnson, K.S., Boyle, E., Bruland, K., Coale, K., Measures, C., Moffett, J., et al. (2007).
468 Developing standards for dissolved iron in seawater. *Eos* 88(11), 131-132.

- 469 Klunder, M.B., Laan, P., Middag, R., De Baar, H.J.W., and van Ooijen, J.C. (2011). Dissolved
470 iron in the Southern Ocean (Atlantic sector). *Deep-Sea Research Part II: Topical Studies*
471 *in Oceanography* 58(25-26), 2678-2694.
- 472 Knap, A.H., Michaels, A. F., Steinberg, D., Bahr, F., Bates, N., Bell, S., Countway, P., Close, A.,
473 Doyle, A., Howse, F., and Gundersen, K., Johnson, R., Little, R., Orcutt, K., Parsons, R.,
474 Rathbun, C., Sanderson, M., Stone, S. (1997). *BATS Methods Manual* [Online]. Woods
475 Hole: U.S. JGOFS Planning Office. Available: http://bats.bios.edu/bats_methods.html
476 [Accessed].
- 477 Kragten, J. (1994). Tutorial review. Calculating standard deviations and confidence intervals
478 with a universally applicable spreadsheet technique. *The Analyst* 119(10), 2161-2165.
- 479 Landing, W.M., Cutter, G.A., Dalziel, J.A., Flegal, A.R., Powell, R.T., Schmidt, D., et al. (1995).
480 Analytical Intercomparison Results from the 1990 Intergovernmental-Oceanographic-
481 Commission Open-Ocean Base-Line Survey for Trace-Metals - Atlantic-Ocean. *Marine*
482 *Chemistry* 49(4), 253-265. doi: 10.1016/0304-4203(95)00016-K.
- 483 Linsinger, T. (2005). "Comparison of a measurement result with the certified value, European
484 Commission - Joint Research Centre (IRMM), Geel, Belgium". (Geel, Belgium: European
485 Commission - Joint Research Centre (IRMM)).
- 486 Magnusson, B., Näykki, T., Hovind, H., and Krysell, M. (2012). "Handbook for calculation of
487 measurement uncertainty in environmental laboratories, 3rd edition". 3rd ed.).
- 488 Miller, J.N., and Miller, J.C. (2010). *Statistics and Chemometrics for Analytical Chemistry*.
489 London, England: Pearson Education.
- 490 Milne, A., Schlosser, C., Wake, B.D., Achterberg, E.P., Chance, R., Baker, A.R., et al. (2017).
491 Particulate phases are key in controlling dissolved iron concentrations in the (sub)tropical
492 North Atlantic. *Geophysical Research Letters* 44(5), 2377-2387. doi:
493 10.1002/2016GL072314.
- 494 Ohnemus, D.C., Auro, M.E., Sherrell, R.M., Lagerström, M., Morton, P.L., Twining, B.S., et al.
495 (2014). Laboratory intercomparison of marine particulate digestions including Piranha: A
496 novel chemical method for dissolution of polyethersulfone filters. *Limnology and*
497 *Oceanography: Methods* 12(AUG), 530-547. doi: 10.4319/lom.2014.12.530.
- 498 Rapp, I., Schlosser, C., Rusiecka, D., Gledhill, M., and Achterberg, E.P. (2017). Automated
499 preconcentration of Fe, Zn, Cu, Ni, Cd, Pb, Co, and Mn in seawater with analysis using
500 high-resolution sector field inductively-coupled plasma mass spectrometry. *Analytica*
501 *Chimica Acta* 976, 1-13. doi: 10.1016/j.aca.2017.05.008.
- 502 Rigaud, S., Puigcorbé, V., Cámara-Mor, P., Casacuberta, N., Roca-Martí, M., Garcia-Orellana,
503 J., et al. (2013). A methods assessment and recommendations for improving calculations
504 and reducing uncertainties in the determination of ²¹⁰Po and ²¹⁰Pb activities in
505 seawater. *Limnology and Oceanography: Methods* 11(OCT), 561-571. doi:
506 10.4319/lom.2013.11.561.
- 507 Worsfold, P.J., Clough, R., Lohan, M.C., Monbet, P., Ellis, P.S., Quetel, C.R., et al. (2013). Flow
508 injection analysis as a tool for enhancing oceanographic nutrient measurements--a
509 review. *Anal Chim Acta* 803, 15-40. doi: 10.1016/j.aca.2013.06.015.
- 510 Worsfold, P.J., Lohan, M.C., Ussher, S.J., and Bowie, A.R. (2014). Determination of dissolved
511 iron in seawater: A historical review. *Marine Chemistry* 166, 25-35. doi:
512 10.1016/j.marchem.2014.08.009.
- 513

514 **Figure captions**

515

516 **Fig. 1.** The FI manifold with on-line preconcentration and ICP-MS detection for the determination of iron
517 in seawater. Reprinted from *Talanta* 133 (2015) 162-169, Clough et al., Uncertainty contributions to the
518 measurement of dissolved Co, Fe, Pb and V in seawater using flow injection with solid phase

519 preconcentration and detection by collision cell - quadrupole ICP-MS, Copyright 2015, with permission
520 from Elsevier (Clough et al., 2015).

521

522 **Fig. 2.** The FI manifold with on-line preconcentration and chemiluminescence detection for the
523 determination of iron in seawater. Reprinted from *Limnology & Oceanography Methods* 13 (2015) 673-
524 686, Floor et al., Combined uncertainty estimation for the determination of the dissolved iron
525 concentration in seawater using flow injection with chemiluminescence detection, Copyright 2015, with
526 permission from Wiley Periodicals, Inc. on behalf of the Association for the Sciences of Limnology and
527 Oceanography (Floor et al., 2015).

528

529 **Fig. 3.** The possible major sources of uncertainty for calculating C_s (the dissolved iron concentration in
530 the sample in ng kg^{-1}) using FI-CL. Reprinted from *Analytica Chimica Acta* 803 (2013) 15-40, Worsfold
531 et al., Flow injection analysis as a tool for enhancing oceanographic nutrient measurements – a review,
532 Copyright 2013, with permission from Elsevier (Worsfold et al., 2013).

533

534 **Table 1.** Analytical data for the measurement of dissolved iron (nmol L⁻¹) in seawater using flow injection with ICP-MS detection. Uncertainties for the
 535 experimental values are the calculated expanded (k = 2) uncertainties (U). Uncertainty for the certified value is ± 2 s.d. Uncertainties for the consensus
 536 values are ± 2 s.d. for the results from 22 participating laboratories. Limits of detection (nmol L⁻¹) were 0.33 (NASS-6 CRM), 0.071 (GEOTRACES GD RM)
 537 and 0.23 (GEOTRACES GS RM) and were calculated from three times the combined standard uncertainty of the elemental determination in the wash
 538 solution and pH adjustment buffer. Data sourced from Clough et al. (2015). Original data for the certified value are from the National Research Council
 539 Canada NASS-6 data sheet. Original data for the consensus values are from the GEOTRACES website. All original data have been converted to nmol L⁻¹
 540 dissolved iron and all uncertainties converted to ± 2 s.d.

541

Element	Reference material	Experimental value [Analyte]	Expanded uncertainty (U _c)	Relative expanded uncertainty (U _{c_rel})	Certified value	Consensus value	Statistical agreement
		(nmol L ⁻¹)	(nmol L ⁻¹)	(%)	(nmol L ⁻¹)	(nmol L ⁻¹)	
Fe	NASS-6	9.64	0.84	9	8.86 ± 0.82		Yes
	GEOTRACES GS	0.505	0.089	18		0.560 ± 0.094	Yes
	GEOTRACES GD	1.035	0.079	8		1.03 ± 0.21	Yes

542

543

544

545 **Table 2.** Uncertainty contributions for the measurement of dissolved iron in seawater using flow injection with ICP-MS detection. The symbols used for the
 546 method parameters investigated are defined in the text. Sample volume (V_2), Sample + buffer volume (V_1) and buffer volume (B_v) contributed minimally
 547 (<1%) to the overall uncertainty. Data sourced from Clough et al. (2015).

548

Relative Uncertainty Contribution (%)				
Reference material	Peak area (I_s)	Calibration curve slope (F)	Wash blank (I_{WB})	Buffer blank concentration (B_C)
NASS-6	10	81	8	0
GEOTRACES GS	73	8	17	2
GEOTRACES GD	51	37	11	0

549

550

551
552
553
554
555
556

Table 3. Analytical data for the measurement of dissolved iron (nmol kg^{-1}) in seawater using FI-CL with gravimetric loading. Uncertainties for the experimental values are the calculated expanded ($k = 2$) uncertainties (U). Uncertainties for the consensus values are ± 2 s.d. for the results from 22 participating laboratories. Original data for the consensus values are from the GEOTRACES website. Data sourced from Floor et al. (2015). All original data have been converted to nmol kg^{-1} dissolved iron and all uncertainties converted to ± 2 s.d.

Reference material	Experimental dissolved iron concentration (nmol kg^{-1})				Consensus dissolved iron concentration (nmol kg^{-1})	
	Peak height		Peak area		Value	Relative expanded uncertainty (U_{c_rel})
	Value	Relative expanded uncertainty (U_{c_rel})	Value	Relative expanded uncertainty (U_{c_rel})		
GEOTRACES GS	0.478 ± 0.060	12	0.500 ± 0.051	10	0.546 ± 0.092	16.8
GEOTRACES GD	0.800 ± 0.099	12	0.836 ± 0.084	10	1.0 ± 0.2	20.0

557
558
559

560
561
562
563
564
565
566
567

Table 4. Validation of the FI-CL method, with the dissolved iron concentrations determined for the consensus/reference materials in nmol L⁻¹. Uncertainties for the experimental values are the calculated expanded ($k = 2$) uncertainties (U). Uncertainty for the certified value is ± 2 s.d. Uncertainties for the consensus values are ± 2 s.d. for the results from 29 participating laboratories. Original data for the certified value are from the National Research Council Canada NASS-5 data sheet. Original data for the consensus values are from the GEOTRACES website. All original data have been converted to nmol L⁻¹ dissolved iron and all uncertainties converted to ± 2 s.d. Details of the SAFe sample collection procedures can be found at <https://websites.pmc.ucsc.edu/~kbruland/GeotracesSaFe/kwbGeotracesSaFe.html> .

Reference material	Concentration Determined (nmol L ⁻¹ ± 2 s.d. (<i>n</i>))	Certified value (nmol L ⁻¹)	Consensus value (nmol L ⁻¹)
SAFe D2	0.96 \pm 0.20 (14)		0.956 \pm 0.048*
SAFe D1	0.69 \pm 0.08 (4)		0.69 \pm 0.08*
SAFe S	0.12 \pm 0.02 (4)		0.095 \pm 0.016*
NASS-5	3.77 \pm 0.06 (2)	3.71 \pm 0.63	

*Converted to nmol L⁻¹ from nmol kg⁻¹ using density of 1.025 kg L⁻¹.

568
569