

Intercomparison of dissolved trace elements at the Bermuda Atlantic Time Series station

Journal Article**Author(s):**

Middag, Rob; Séférian, Roland; Conway, Tim M.; John, Seth G.; Bruland, Kenneth W.; de Baar, Hein J.W.

Publication date:

2015-12-20

Permanent link:

<https://doi.org/10.3929/ethz-b-000106881>

Rights / license:

[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International](#)

Originally published in:

Marine Chemistry 177(3), <https://doi.org/10.1016/j.marchem.2015.06.014>

1 **Intercomparison of Dissolved Trace Elements at the Bermuda Atlantic Time Series**

2 **Station**

3

4 R. Middag^{1,2,3*}, R. Séférian⁴, T.M. Conway^{5,6}, , S.G. John^{5,7}, K.W. Bruland² and H.J.W de
5 Baar³

6

7 ¹ Department of Chemistry, University of Otago, Dunedin, Otago, New Zealand

8 ² Department of Ocean Sciences & Institute of Marine Sciences, University of California
9 Santa Cruz, CA 95064, USA.

10 ³Royal Netherlands Institute for Sea Research (Royal NIOZ), Den Burg, Texel, The
11 Netherlands.

12 ⁴ CNRM-GAME, Centre National de Recherche Météorologique-Groupe d'Etude de
13 l'Atmosphère Météorologique, Météo-France/CNRS, 42 Avenue Gaspard Coriolis, 31100
14 Toulouse, France

15 ⁵ Department of Earth and Ocean Sciences, University of South Carolina, SC 29208 USA.

16 ⁶ Department of Earth Sciences, ETH Zürich, 8092 Zürich, Switzerland.

17 ⁷ Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089
18 USA

19

20

21 *Corresponding Author: Rob Middag, rob.middag@otago.ac.nz, +64 03 479 7907

22

23 Roland (cnrm) rseferian.cnrm@gmail.com

24 Tim Conway conway.tm@gmail.com

25 Seth John sjohn@geol.sc.edu

26 Kenneth Bruland bruland@ucsc.edu
27 Hein de Baar Hein.de.Baar@nioz.nl
28
29 Running head: BATS intercomparison
30
31 Keywords:
32 Intercomparison
33 Cross-over station
34 Trace metals
35 GEOTRACES
36

37 **Abstract**

38 The international GEOTRACES program is emphasizing the use of the GEOTRACES
39 reference samples program and the importance of cross-over or baseline stations where all
40 aspects of sample collection, filtration, processing and analytical methods can be compared.
41 These intercomparison efforts are of crucial importance to merge or link data from different
42 origins together. The occupation of the Bermuda Atlantic Time Series (BATS) station
43 ($31^{\circ}45.92'N$, $64^{\circ}04.95'W$) by the Netherlands GEOTRACES (GA02; 13 June 2010) as well
44 as US GEOTRACES (GA03; 19-21 November 2011) Atlantic sections expeditions provided
45 an intercomparison opportunity to confirm the compatibility of the 2 different sampling
46 systems as well as different analytical techniques used. In order to compare the data from
47 different sampling systems or analytical techniques, a new statistical approach was developed
48 to include the analytical uncertainty. Furthermore, modifications to an existing multi-element
49 technique (Biller and Bruland, 2012) were made that increased the number of elements
50 analysed and that allows the time consuming extractions to be done shipboard. Overall, we
51 show excellent agreement between data generated by different sampling systems and
52 analytical techniques. However, while both sampling systems are capable of collecting
53 uncontaminated samples for all the GEOTRACES key elements, some apparent outliers are
54 present. The intercomparison between the different analytical techniques also reveals subtle
55 differences that would have gone unnoticed if only reference samples would have been used,
56 underlining the importance and power of intercomparison stations. The results show cross-
57 over station data can be used to assess consistency between datasets if these stations have
58 been carefully planned and analysed in combination with (internal) seawater reference
59 samples to assure intra-dataset consistency.

60

61

62 **Introduction**

63 The GEOTRACES program is an international study which aims to improve the
64 understanding of the global marine biogeochemical cycles and large-scale distribution of
65 trace elements and their isotopes. As GEOTRACES is an international effort with scientists
66 from approximately 30 nations participating, the emerging data sets are produced using
67 differing sampling systems and analytical methods. These datasets are merged in the
68 Intermediate Data Product (The GEOTRACES group, in press). To assure consistency
69 between datasets, the international program is emphasising the concurrent analysis of
70 reference samples from the GEOTRACES reference sample program for concentrations of 9
71 key dissolved trace elements, as well as the planning of cross-over or baseline stations where
72 all aspects of sample collection, filtration, processing and analytical methods can be
73 compared. The intercomparison of data for the Intermediate Data Product has been carried
74 out by the GEOTRACES Standards and Intercalibration committee, which shares the results
75 only with contributors to encourage participation. This paper, independently of the findings
76 of that committee, describes the comparison of 2 sampling systems and four analytical
77 techniques, as well as statistical methods to compare trace metal profiles.

78 The US and Netherlands designed new trace metal clean sampling systems specifically for
79 their respective GEOTRACES programs to obtain accurate values of the distributions of key
80 trace metal elements and isotopes along full depth sections. These are the Titan sampling
81 system of the Netherlands (De Baar et al., 2008) that was recently upgraded with novel
82 PVDF samplers (de Baar et al., 2012) as described in the companion article (Rijkenberg et
83 al., submitted; this issue) and the US GEOTRACES carousel (Cutter and Bruland, 2012).
84 Both systems utilise a conducting cable with a CTD instrument as part of the sampling
85 system and can collect 24 samples, but they differ in the materials used and the procedures of
86 recovery and deployment. For example, the whole Titan system with its 24.4 litre PVDF

87 samplers goes into its own cleanroom van for sub-sampling whereas the 12 litre, Teflon-
88 coated, GO-FLO sampling bottles are removed from the US GEOTRACES carousel and are
89 brought into a dedicated cleanroom van for filtration and sub-sampling (for more details see
90 (Cutter and Bruland, 2012; De Baar et al., 2008; Rijkenberg et al., submitted; this issue)).

91 Besides the development of new sampling equipment, considerable progress has also been
92 made in the development of new multi-element methods (e.g. Biller and Bruland, 2012; Lee
93 et al., 2011; Milne et al., 2010; Sohrin et al., 2008) using chelating resins for off-line
94 extraction, with subsequent detection with a high-resolution, magnetic sector, inductively
95 coupled plasma mass spectrometer (ICP-MS). Most of the data presented in this paper has
96 been analysed using a modified version of the Biller and Bruland (2012) method that will be
97 described here.

98 With these sampling and processing systems, datasets of unprecedented resolution are
99 being produced that are linked together in a global database (The GEOTRACES group, in
100 press). Therefore, we must be confident the sampling systems as well as different analytical
101 techniques obtain comparable and uncontaminated samples. The occupation of the Bermuda
102 Atlantic Time Series (BATS) station by the Netherlands GEOTRACES during a meridional
103 section of the western Atlantic in June 2010 (GA02) and the US GEOTRACES during a
104 zonal section of the North Atlantic in November 2011 (GA03) provided an opportunity to
105 collect dissolved trace metal samples from two occupations of the same ‘cross-over’ station
106 with different sampling systems, which were analysed for a suite of trace metals by the same
107 analyst with the same analytical method. For Al, Fe and Mn, shipboard dissolved
108 concentration data values are available from the June 2010 Netherlands BATS occupation
109 allowing further intercomparison with shipboard data. Additionally, for Cd, Fe and Zn,
110 concentration data is available from the US occupation using a double spike multi-collector
111 ICP-MS method (Conway et al., 2013). More data is available from the US occupation as

112 well as historic occupations of the BATS station, but for the sake of conciseness not all data
113 and methods can be described and compared in this paper and we contrast shipboard methods
114 and 2 recent ICP-MS methods that apply standard addition versus isotope dilution double
115 spiking as the calibration method.

116 Besides comparing the results of the different sampling systems and analytical techniques
117 (Table 1), this paper also aims to describe the modifications made to the method of Biller and
118 Bruland (2012). In order to compare the data from different sampling systems or analytical
119 techniques, a new statistical approach had to be developed to include the analytical
120 uncertainty that is described in this paper. In the companion paper the new PVDF samplers
121 are described (Rijkenberg et al., submitted; this issue).

122

123 **Methods**

124 For the Titan sampling system and the US GEOTRACES carousel the reader is referred to
125 the respective papers (Cutter and Bruland, 2012; De Baar et al., 2008) and the companion
126 paper on the PVDF samplers (Rijkenberg et al., submitted; this issue). The only differences in
127 sample treatment are the filtration and acidification. Seawater samples were acidified to a
128 concentration of 0.024 M hydrochloric acid (HCl) which results in a pH of 1.7 to 1.8 with
129 either Baseline[®] HCl (Seastar Chemicals Inc.) for the samples from the Netherlands cruises,
130 or Q-HCl (produced at UCSC as 6 N Q-HCl with a sub-boiling quartz still) for samples from
131 the US cruise analysed with the multi element standard addition ICP-MS method. Seawater
132 samples from the US cruise analysed with the double spike ICP-MS method were acidified to
133 a concentration of 0.012 M HCl (which results in a pH of 2.0) with Aristar Ultra HCl (VWR
134 International). Any difference between these acids is accounted for in the blank
135 determination. On the Netherlands GEOTRACES cruise, the samples were filtered using 0.2
136 μm pore size Sartobran capsule filters, whereas on the US GEOTRACES cruise 0.2 μm pore

137 size Acropak capsule filters were used. The use of these different brands of filter cartridge
138 does not result in significant differences in trace metal concentrations (Cutter and Bruland,
139 2012). All samples were stored in LDPE bottles that were acid washed according to the
140 ‘Sample and Sample-handling Protocols for GEOTRACES cruises’
141 (<http://www.geotraces.org>).

142

143 *Double spike ICP-MS and shipboard methods*

144 The details of the double spike ICP-MS method for Fe, Zn and Cd by Thermo Neptune
145 multicollector ICP-MS are described by Conway et al. (2013). Accuracy of the double spike
146 method for dissolved concentrations has been previously demonstrated by agreement with the
147 consensus values for SAFe D1, D2 and S reference standards Conway et al. (2013), which
148 were analysed within the same ICP-MS analytical sessions at the Center for Elemental Mass
149 Spectrometry (University of South Carolina) as the samples presented in this
150 intercomparison. Procedural blanks for this technique were determined as 5 pmol kg⁻¹ (Fe), 1
151 pmol kg⁻¹ for (Zn) and 35 fmol kg⁻¹ for Cd (Conway et al., 2013). The details on the
152 shipboard methods of the Netherlands occupation for Fe, Mn and Al were described by
153 Rijkenberg et al. (2014), Middag et al. (2011a) and Middag et al. (in press), respectively.
154 Typical precisions and detection limits for shipboard Al were 3.3% for multiple
155 measurements of a 6.5 nM sample (n=40) and a detection limit of 0.05 nM (defined as three
156 times the standard deviation of the lowest concentration observed) (Middag et al., in press).
157 For shipboard Mn these were 5.1% for a 0.45 nM sample (n=37) and a detection limit of 0.01
158 nM. For shipboard Fe the precision was 4% for replicate measurements of a 0.94 nM
159 reference sample (n=29) and a limit of detection of 0.01 nM (Rijkenberg et al., 2014).

160

161 *Multi-element standard addition ICP-MS Method*

162 *Reagents*

163 This method includes the analysis of yttrium (Y), lanthanum (La), titanium (Ti) and
164 gallium (Ga), in addition to manganese (Mn), iron (Fe), cobalt (Co), nickel (Ni), copper (Cu),
165 zinc (Zn), cadmium (Cd) and lead (Pb) that were determined in the original method by Biller
166 and Bruland (2012). Moreover, we introduce a new ‘element dilution’ approach that can be
167 used for extractions performed at sea that is less labour intensive than the gravimetical
168 method described by Biller and Bruland (2012) as the weighing of the samples has been
169 excluded. The extraction of the samples is the process where the trace metals of interest are
170 separated from the original seawater matrix to remove interfering ions, as well as
171 concentrating the samples via the use of a chelating column (Nobias PA1 chelating resin in
172 this method). This pre-concentration is necessary due to the low concentrations of trace
173 metals in the open ocean and the high background salt matrix of seawater.

174 The reagents used for the trace metal analyses are the same as those described by Biller
175 and Bruland (2012), with the exception of a Lutetium (Lu)-Indium (In) spike and the use of
176 Hydrofluoric Acid (HF) as described below. A 4000 nM stock of Lu and In was made in
177 0.024 M Nitric Acid (HNO₃) (triple quartz distilled) from dilutions of the respective 1000 µg
178 g⁻¹ SPEX standards. A 2.4 M and 1.2 M working solution of HF were made by diluting
179 concentrated HF (50%, Optima grade, Fisher Scientific). The HF addition was found to
180 increase the recovery of notably Ti (see *Modified Multi-element standard addition ICP-MS*
181 *Method*). A short outline of the method is given below that mainly focusses on the differences
182 with the method described by Biller and Bruland (2012). For more details on the method and
183 method verification in general, the reader is referred to Biller and Bruland (2012). The
184 method uses Nobias-chelate PA1 resin to extract the metals from the seawater. The metals are
185 subsequently eluted with 1 M nitric acid and the extracts are run on a sector field ICP-MS.

186 The 1 M nitric acid used in this method contains 0.6 mM HF in addition to the 10 ng g⁻¹ Rh in
187 the original method.

188

189 *Calibration*

190 A multi-element stock standard with natural isotopic abundances of Mn, Fe, Co, Ni, Cu,
191 Zn, Cd, Y, La, Ti, Ga and Pb, in 0.024 M HNO₃ was used to make standard additions to
192 natural seawater with low concentrations of metals for calibration. Five standard additions
193 were used for calibration in this method, and every sample and standard were spiked with Lu
194 and In to obtain a concentration of 5 nM for both elements. In addition to the seawater
195 standard additions, 5 standards were made up in the acid used to elute the metals from the
196 pre-concentration columns (elution acid standards). These standards (also spiked with Lu-In)
197 and the extracted seawater standards were analysed on each run.

198 The resin recovery for the modified multi-element ICP-MS method was monitored as
199 described by Biller and Bruland (2012). Briefly, the slope of the relationship between the
200 ICP-MS signal and metal standard additions to seawater is compared with the slope of the
201 relationship between the ICP-MS signal and metal additions to elution acid that was not
202 passed through the extraction system. This was done for every set of samples that were
203 extracted and analysed by ICP-MS. The recovery for all elements with the exception of Ga
204 and Ti is quantitative (recovery >98%) and agrees with previous results for this resin (Biller
205 and Bruland, 2012; Sohrin et al., 2008).

206 Lutetium and In are two trace metals, naturally existing at sub pM concentrations (Alibo et
207 al., 1999; Alibo and Nozaki, 1999; Amakawa et al., 1996; Greaves et al., 1991), that are
208 quantitatively concentrated by the extraction procedure (recovery >98%). The 5 nM spike
209 overwhelms the natural concentration in the samples. This Lu-In spiking, when done to an
210 exact volume (see *preconcentration and analysis*) of seawater, serves as an internal standard

211 or, analogous to isotope dilution, as an ‘element dilution’ calibration technique. This is in
212 addition to the standard additions to a seawater sample: the standard additions serve to
213 determine the calibration sensitivity for each metal (i.e. the signal per unit of metal) while the
214 Lu-In spike represents the concentration factor of each sample due to the extraction method.
215 The Lu : ‘metal of interest’ or In : ‘metal of interest’ ratio does not change after addition of
216 the spike to the sample and will be the same in the sample and the eluent when recoveries are
217 quantitative. Additionally, after addition of the Lu-In spike to the seawater samples, any
218 evaporation of the seawater samples or the eluants after extraction is irrelevant as that does
219 not change the ratios. If the recovery of a metal is not quantitative, the method will only work
220 if the recovery is stable, i.e. does not vary between samples and standards. In this case the
221 calibration sensitivity will be relatively low, but the non-quantitative recovery will be
222 accounted for in this calibration sensitivity.

223 By dividing the ICP-MS signal for every metal by the signal for Lu, the signal is
224 normalized with respect to the concentration factor as the Lu signal is proportional to the
225 amount of sample loaded and the amount of elution acid used. Similarly, the signal for every
226 metal can be (separately) divided by the In signal to have 2 independent determinations of the
227 concentration factor with the added benefit that a ‘failed’ extraction shows up in large
228 deviations of the resulting concentrations or Lu/In ratios. The Lu and In signals were
229 determined in both medium and low resolution, and were applied appropriately to metals
230 measured in medium resolution (^{47}Ti , ^{55}Mn , ^{56}Fe , ^{59}Co , ^{60}Ni , ^{63}Cu , ^{66}Zn , ^{69}Ga) or low
231 resolution (^{89}Y , ^{111}Cd , ^{139}La and ^{208}Pb). The Rh spike to the elution acid served to quickly
232 highlight any aspiration problems or drifts in sensitivity on the ICP-MS for the original
233 gravimetric method. The data presented here was determined by the gravimetric method
234 based upon the sensitivity of the standard additions to the seawater. The Lu-In element

235 dilution method was tested versus the gravimetric method (see *Modified Multi-element*
236 *standard addition ICP-MS Method*).

237

238 *Pre-concentration and analysis*

239 The off-line pre-concentration set-up and procedures are as described by Biller and
240 Bruland (2012) with some small but essential differences. Most importantly, to 40 ml of
241 seawater, 20 μL of the 1.2 M HF was added (resulting in a 0.6 mM fluoride ion
242 concentration), followed by a 50 μL addition of the Lu-In spike. This 40 ml was taken using a
243 10 ml pipette from a slightly larger volume of seawater. This seawater was UV oxidised to
244 destroy organic metal binding ligands for dissolved Co and Cu concentration determinations
245 (this was found to be particularly important for obtaining accurate dissolved Co
246 concentrations and significant for obtaining accurate dissolved Cu concentrations (Biller and
247 Bruland, 2012)). The overall imprecision of the pipetting the 40 ml of water was determined
248 gravimetrically and found to be $\sim 0.9\%$. The accuracy of the pipette was monitored and no
249 drift was observed over the course of 2 months.

250 Samples were capped and vigorously shaken (5 seconds) and left overnight (14 hours) to
251 homogenise after the spiking. Immediately prior to extraction, the pH of the samples was
252 adjusted to 5.8 \pm 0.2 with $\sim 500 \mu\text{L}$ of a 3.7 M NH_4Ac solution. About 35 ml of the sample
253 was loaded onto the columns in ~ 45 minutes (the exact amount of the sample loaded was
254 gravimetrically determined (Biller and Bruland, 2012)). The leftover 5 ml was used to retro-
255 actively verify that the correct pH was achieved. The columns were rinsed with a 0.05 M
256 NH_4Ac rinse/conditioning solution for 1 and 2 minutes, respectively, before and after sample
257 loading to condition the columns at the correct pH, and remove interfering sea salt before
258 elution. Samples were eluted with ~ 2 ml of the elution solution in ~ 5 minutes (the exact
259 amount of the elution acid was gravimetrically determined), resulting in a concentration

260 factor close to 17.5. Using 8 columns in this off-line manifold allowed the simultaneous
261 extraction of 8 samples.

262 Samples were analysed on a Thermo Element XR Magnetic Sector ICP-MS at the Institute
263 of Marine Science at UCSC with nickel sample and skimmer cones (Spectron) and an ESI-
264 PC3 Peltier cooled spray chamber (4 °C). The sample gas rate was 0.75 mL min⁻¹, and the
265 additional gas rate was optimised every run (usually around 0.2 mL min⁻¹). Sample was
266 pumped into the instrument at a rate of 120 µL min⁻¹. Molybdenum (⁹⁵Mo) was measured to
267 correct for Molybdenum oxide interferences on the Cd signal (Biller and Bruland, 2012) that
268 became significant at low surface water concentrations of dissolved Cd.

269 Unused (i.e. not passed through the extraction set-up) elution acid was used to track the
270 baseline associated with the acid and ICP-MS instrument and was subtracted from the raw
271 signals. An elution acid standard (elution acid with added metals) served as a drift standard to
272 track the sensitivity of the ICP-MS during a run even though the Rh (gravimetric method)
273 and the Lu-In spikes ('element dilution' method) are used to correct for any change in
274 sensitivity. System blanks were determined by performing the normal extraction procedure,
275 but loading a low metal concentration seawater sample (the same as used for the standard
276 additions) onto the column for only 30 seconds versus ~45 minutes for an ordinary sample.
277 This way the column is loaded with sea salts, but only minimally with metals. A system blank
278 was determined for every column, but consistency between the blanks indicated no noticeable
279 difference between columns. A reagent blank was determined by adding double the normal
280 amount of HCl, buffer, HF and Lu-In spike to a sample of the same seawater used for the
281 standard additions. The resulting concentration was compared to the 'zero addition' point of
282 the calibration that only had normal amounts of reagents added. No noticeable reagent blank
283 could be determined for the metals except a small Pb blank of 1.2 pmol kg⁻¹. Blanks and limit
284 of detection (LOD, defined as three times the standard deviation of the system blank) are

285 shown in Table 2. SAFE and GEOTRACES reference samples were analysed in every run
286 (Table 3).

287

288 **Comparing profiles with analytical uncertainty**

289 To compare datasets, traditionally paired two-tailed t-tests or 95% confidence intervals
290 (CI's; expressed as 2 x standard deviation) have been used as well as the root-mean-square
291 (RMS) deviations of the reported values from the mean values (Bowie et al., 2003; Landing
292 et al., 1995). The t-tests have the disadvantage that analytical uncertainty is ignored, for
293 example Bowie et al. (2003) reported significant differences (at the 0.05 significance level)
294 between different analytical techniques based on paired two-tailed t-tests, whereas 95% CI's
295 where often overlapping. In addition, a t-test assumes a normal distribution of the data which
296 is rarely the case for an oceanographic profile. Indeed, the distribution of the data is generally
297 skewed towards higher or lower concentrations in the shape of the vertical profile of the
298 chemical species. Solely looking at 95% CI's or RMS deviations, on the other hand, does not
299 give a quantitative measure of comparability for an entire profile.

300 In order to objectively and reproducibly assess the comparability of analytical methods or
301 sampling systems while acknowledging there is analytical uncertainty, another approach is
302 needed. Here we describe a relatively straight forward test that can be done in a spreadsheet
303 or statistical program. This approach, however, is limited to comparing analytical techniques
304 based on results from the same samples or subsamples thereof. It is less suited to compare
305 oceanographic samples collected at different occasions since samples will have been
306 collected at (slightly) different depths for which another test was developed (see *Different*
307 *samples approach*).

308

309 *Same sample approach*

310 To compare analytical results from the same samples (or sub-samples thereof) first a
311 standard paired two-tailed t-test ($\alpha=0.05$) was done as described by Bowie et al. (2003). As
312 an example, in case of an oceanographic profile of 24 samples analysed by two different
313 analytical techniques in triplicate, the test compares the 24 analytical result pairs. If this does
314 not result in a statistical difference, the results of the two methods are not significantly
315 different regardless of analytical uncertainty. As mentioned above, this paired two-tailed t-
316 test assumes a normal distribution of the data. This constitutes a potential caveat for this
317 approach. A statistical program such as Sigmaplot 13.0 as used in this paper, will test for
318 normality of the data when performing a t-test and, if assumptions are violated, an alternative
319 non parametric test such as a Wilcoxon signed-rank test can be performed instead. The test
320 outlined here is for the comparison of 2 datasets, for multiple datasets the t-test and signed
321 rank test can be replaced by an ANOVA or a Friedman Repeated Measures Analysis of
322 Variance on Ranks, with a Tukey post hoc test to identify differences. If the t-test or signed-
323 rank test does give a significant difference, the results of the two methods are subsequently
324 assessed with consideration of the analytical uncertainty. For this, the absolute difference
325 between results of method 1 and method 2 was calculated for each sample. Subsequently the
326 standard deviation of the difference was calculated using equation 1:

$$327 \quad \sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2} \quad (\text{eq. 1})$$

328 Where σ_d is the standard deviation of the difference between the analytical method 1 and
329 analytical method 2, and σ_1 and σ_2 are the standard deviations of the results for method 1 and
330 method 2 respectively. The latter standard deviation is calculated based on repeat
331 measurements of subsamples of the same sample and can also be referred to as the standard
332 error of the mean. If no standard deviation can be calculated for one of the methods (e.g.
333 because no multiple subsamples were analysed) the analytical uncertainty needs to be
334 assessed by other means. Notably for ICP-MS methods, often a single sample is analysed by

335 the instrument a number of times and the average and standard deviation of those
336 measurements is reported. This standard deviation divided by $n^{0.5}$, where n is the number of
337 measurements (instrument scans), is the 'internal' standard error for the measurement
338 (instrument variability). It is known that this 'internal' standard error of one sample analysis
339 approximates the standard deviation of replicate analysis of subsamples (standard error of the
340 mean). However, this does not hold in case of techniques, such as the standard addition ICP-
341 MS method, where a considerable amount of variation is not due to instrument variability, but
342 due to variation in extractions on resin, difference in sample bottles or other non-instrument
343 related variability. This shows in an 'internal' standard error for the measurement of one
344 subsample that is smaller than the standard deviation of replicate analysis of multiple
345 subsamples (standard error of the mean). The best determination of the standard deviation
346 (standard error of the mean) is to analyse subsamples of the same samples an infinite number
347 of times. The best approximation of this in a practical sense is probably the analyses of
348 references samples/standards that are analysed with every run. In this paper standard
349 deviation will be denoted with the symbol σ and is based on repeat measurements of
350 separately extracted subsamples of the same sample (standard error of the mean).

351 In case of the double spike ICP-MS method where no multiple subsamples of the same
352 seawater samples were analysed for the whole profile, the analytical uncertainty was
353 estimated to be 2% (Conway et al., 2013) and these estimates were used as the σ . The σ_d and
354 absolute difference was calculated for every pair of results; in the case of an oceanographic
355 profile of 24 samples analysed by two different analytical techniques, this results in 24
356 absolute differences and 24 associated standard deviations of the difference. Subsequently a
357 one-tailed paired t-test or signed-rank test can be performed to assess whether σ_d is larger
358 than the absolute difference for the entire profile. If this is the case, it is concluded the
359 analytical results of the two methods are not statistically different within uncertainty as the

360 uncertainty is larger than the difference. Otherwise the conclusion is that the two methods do
361 not give comparable results.

362 This analytical uncertainty assessment can be repeated using a 2σ -based analytical
363 uncertainty or 95% CI's, but for this paper the more conservative approach of 1σ was chosen.
364 Assuming a normal distribution, the 95% CI denotes the interval that contains the actual
365 mean with 95% probability. This can be calculated by multiplying the standard error
366 (standard error= $\sigma \times n^{-0.5}$; note this is *not* equivalent to the standard error of the mean defined
367 above) with the critical t-value for the applicable number of degrees of freedom and chosen
368 significance level. For large numbers of replicates ($n > 20$) the critical t-value at $\alpha = 0.05$ is ~ 2 ,
369 making the 95% CI about 2 standard errors whereas for $n = 3$ the critical t-value equals 4.3. In
370 other words, the 95% CI is a measure of certainty about the mean reported value and this
371 certainty increases with a larger number of replicates (smaller 95% CI). Additionally, for a
372 normal distribution and a large number of replicates, 95% of the results lie within $\sim 2\sigma$ of the
373 mean, which confusingly is sometimes also referred to as the 95% CI. This number of 2σ is a
374 measure of the spread around the mean and can therefore be considered more appropriate as a
375 measure of analytical uncertainty. However, commonly samples are analysed in triplicate for
376 shipboard flow injection methods, or single sample extracts are analysed for ICP-MS
377 methods, making the estimation of the 95% CI less reliable or impossible due to the lack of
378 degrees of freedom. Similarly, the assumption of 2σ encompassing 95% of the variation
379 around the mean is not valid with a small numbers of replicates and we opted to use 1σ as
380 the most reliable measure of analytical uncertainty.

381 Additional information can be gained by dividing the difference between result pairs
382 (either positive or negative) by the σ_d after John and Adkins (2010; their equation 18),
383 resulting in values that are basically individual t-values for each pair of results. Subsequently
384 the additive inverse ($-t$) of each t-value is added to the list of t-values, giving a collection of t-

385 values with an average that equals zero. The standard deviation of this collection of t-values
386 can be used to differentiate between intra-lab and inter-lab variability. If the standard
387 deviation is ≤ 1 , all differences can be related to intra-lab variability (i.e. the analytical
388 uncertainties). If the standard deviation is ≥ 1 , the number indicates how much greater the
389 actual observed differences are than expected based on intra-lab variability, e.g. if the
390 standard deviation is 1.3 or 4, it means the inter-lab variability is 30% or 300% greater than
391 expected based on intra-lab variability, respectively. A standard deviation greater or smaller
392 than 1, however, does not indicate statistical significance at the 95% significance level as this
393 test just predicts differences based on the reported analytical uncertainty and does not test for
394 differences. For the latter, the above or below described tests are to be used.

395

396 *Different samples approach*

397 As stated above, the above approach is limited to comparing results from the same
398 samples or subsamples thereof and is not suited to compare oceanographic samples collected
399 by different sampling systems or at different times, as samples will have been collected at
400 (slightly) different depths and/or locations. Therefore a script was developed in the freeware
401 program R (R Core Team, 2013) that interpolates oceanographic profiles and the analytical
402 uncertainty and subsequently tests the shape of two profiles within a prescribed depth
403 interval. The interpolation allows comparing a sampled depth on one profile with an
404 interpolated value in the other profile. Besides, errors in interpolation on mean concentrations
405 and standard deviation of each profile are used to determine the standard deviation of the
406 interpolated values assuming a 95%-significance confidence interval. This interpolation is
407 based on depth, but since water masses might occupy different depths between sampling
408 events, conservative tracers such as potential or neutral density, salinity or potential
409 temperature might be better suited. However, for the surface ocean, salinities, temperatures

410 and densities observed in the US BATS occupation were not observed in the NL BATS
 411 occupation, making interpolation based on these parameters unfeasible. For the sake of clarity
 412 we used depth, as in the deep ocean where comparison is feasible (see section *Intercomparison*
 413 *between the different sampling systems*), depth and density matched very well between the
 414 two occupations. Given that thickness of water masses increases with depth (contracting at
 415 the bottom of the ocean), using depth makes comparison between two profiles easier, notably
 416 in graphs. After interpolation, the shape of the two profiles is compared simultaneously using
 417 two statistical approaches. The first approach consists of comparing samples between each
 418 other across depth using a simple T-based distribution relying on a pooled variance
 419 assumption as follows:

$$420 \quad t_z = \frac{\bar{m}_1 - \bar{m}_2}{\sqrt{\left(n_1\sigma_1^2 + n_2\sigma_2^2\right) \frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{n_1+n_2-1}$$

421 Where \bar{m}_i and σ_i represent respectively the mean concentration at a given z depth of the
 422 profile i and the standard deviation relative to the number of samples, n_i , analysed at this
 423 depth. This gives a p -value for the comparison at each individual depth that is compared and
 424 thus gives a p -value profile, a profile of p -values which provides insight into where the
 425 largest discrepancies and the best agreements are along the depth profile.

426

427 The second approach aims at comparing the shape of the two profiles within a depth interval
 428 using a Fisher-based test. This test consists in computing both the cumulative sum of the
 429 Welch's t_z^2 (Welch, 1947) and the cumulative sum of the randomly generated Fisher
 430 distribution F_{1,n_1+n_2-2} . A chosen number of random simulations (usually 100,000) of the
 431 Fisher F_{1,n_1+n_2-2} distribution are used to estimate the distribution of the cumulative Fisher

432 law. This distribution is then used to compute an ‘integrated’ p -value which represents an
433 objective metric to determine how far the two profiles are consistent between each other
434 within the depth interval. The script is provided in the supplementary material with this
435 paper.
436 The use of both these approaches ensures a good understanding of the profile’s consistency.
437 The ‘integrated’ p -value allows quantifying the match between profiles with one metric, i.e.
438 whether or not there is a significant difference, while the T-based p -value profiles ensures to
439 decipher the reason for a good or poor match between profiles.
440 That said, the computation of the ‘integrated’ p -value is sensitive to the number of samples
441 and to the resolution of sampling over depth, i.e. if the sampling resolution is low in one
442 profile, the interpolation will inherently miss fine scale gradients if present. The ‘integrated’
443 p -value estimated from the F or a χ^2 test can differ by about 10 to 20% due to the
444 sensitivity on the number of samples on the computed confidence intervals. Therefore this
445 approach cannot be considered alone without a proper analysis of profiles miss-match based
446 on the p -value profiles. In the following, the ‘integrated’ p -value will be noted as ‘ i - p -value’
447 and reports the analysis that is based on computation over the profiles (deeper than 1000 m
448 depth, see *Intercomparison between the different sampling systems*).

449

450 **Results and Discussion**

451 *Modified Multi-element standard addition ICP-MS Method*

452 The recovery for Mn, Fe, Co, Ni, Cu, Zn, Cd, Y, La, and Pb was quantitative (>98%) at pH
453 5.8. Two elements demonstrated non-quantitative recovery; Ga (75 %; $\sigma=4\%$) and Ti (50%
454 $\sigma=5\%$). The recovery was assessed over a range of pH and this showed most metals were
455 quantitatively recovered over a pH range of 4.5 to 9, but the recovery of Mn is not
456 quantitative below pH 5.5 and Ga has its optimum recovery around a pH of 5.5. The recovery

457 of Ti, zirconium and scandium benefited from the addition of fluoride (see supplementary
458 material for details). The highest recovery for Ti (75% with 0.6 mM HF) was at pH 4.5 and
459 the recovery of Ti was relatively constant around 50% at a pH of 5.8, but multiple analyses of
460 reference samples indicated modest precision ($\sigma \sim 10\%$).

461 The recovery of zirconium and scandium was unstable, resulting in concentrations with
462 low accuracy. Further development needs to be carried out for these elements and they are
463 currently not included in the method. The fluoride addition had a perceived positive effect on
464 the analysis of Fe and Ga. This was mainly visible in the increased 'smoothness' of oceanic
465 profiles (i.e. rather than meandering the profiles showed uniform trends with increasing
466 depth) and the reproducibility of reference samples. Before addition of HF, occasionally
467 values were measured for the reference samples that were a lot lower than expected (outside
468 the consensus range), whereas after addition of HF this was no longer observed. The
469 improvement of the visual appearance of the profiles and more consistent reference samples
470 results might possibly be due to hindering of precipitation after pH adjustment in the presence
471 of HF, but increased operator experience cannot be excluded either. The effect of fluoride
472 addition to the seawater on the recovery was assessed at concentrations up to 10 mM fluoride
473 and it appeared there is an optimum of about 0.6 mM fluoride for the recovery of Ti. It is
474 hypothesised that the addition of fluoride has an advantageous effect for the elements that
475 form hydrolysis complexes in seawater. Potentially, the formation of metal-fluoride
476 complexes diminishes precipitation and wall adsorption after pH adjustment to pH 5.8,
477 leading to an improved recovery. Alternatively, the formation of metal-fluoride complexes
478 results in an increased lability of the metal-water complexes, leading to a greater chance for
479 the resin to form a complex with the metal. For example in the case of Fe, at the analytical
480 loading pH (5.8) the $\text{Fe}(\text{OH})_2^+$ complex would dominate (Glasby and Schulz, 1999) and this
481 is relatively slow with respect to water exchange (Hudson et al., 1992). The increased lability

482 of the metal-water complexes could increase the recovery on the resin as the chelating resin is
483 competing for the metals with the complex forming agents in the seawater. The only
484 exception was Al; recovery of Al at pH 5.8 was about 25% without fluoride and after the
485 fluoride addition this decreased to near 0%, indicating the Al-fluoride complex is stronger
486 than the Al-resin complex or kinetically inert on the time scale of its passage through the
487 resin. Overall, the pH of 5.8 +/- 0.2 and an addition of 0.6 mM fluoride to the samples
488 resulted in the best recovery for most elements.

489 SAFe and GEOTRACES reference samples were analysed in every run and the results are
490 in line (1σ intervals overlap) with consensus values for the metals that have those (Table 3).
491 Currently there are no consensus values for Ti on reference samples, but Croot (2011)
492 published results for his analysis of Ti on the reference samples (Table 3). The results for the
493 higher concentrations agree reasonably well (1σ intervals overlap), but there appears to be
494 more discrepancy for the lower concentrations, most notably for SAFe S. However, it should
495 be noted that for one of the SAFe S samples (SAFE S 474) analysed by Croot (2011) the
496 result were in range with our results for SAFe S. This indicates the analysis of Ti is
497 challenging and more method development needs to be done in order to attain reliable data on
498 the distribution of Ti in the open ocean. In this case the method should benefit from using
499 isotope dilution to account for the non-quantitative and potentially variable recovery on the
500 resin. For Y, La and Ga there are currently no consensus values either, and the obtained
501 vertical profiles at BATS are compared to published data below (*Intercomparison between*
502 *the different sampling systems*) to assess the method's accuracy. There is data on reference
503 samples for La (Hatje et al., 2014; van de Flieddt et al., 2012) that agrees (1σ intervals
504 overlap) with the here reported values (Table 3).

505 To accurately measure the sub-pM concentrations of Lu and In the concentration factor
506 would need to be increased, i.e. much more seawater would need to be pre-concentrated on

507 the resin. Therefore we opted instead to use these two elements as an internal standard (see
508 calibration). This ‘element dilution’ calibration technique was assessed by comparing the
509 resulting values versus the gravimetric method as described by Biller and Bruland (2012).
510 No discernible difference was observed between the ‘element dilution’ and gravimetric
511 method (Fig. 1), either with a two-tailed paired t-test ($p=0.20$) or the script for profile
512 comparison that includes analytical uncertainty ($p=0.97$). The benefit of the element dilution
513 calibration technique is that the gravimetric determination of the pre-concentration factor is
514 no longer necessary, making it possible to do the time consuming extractions at sea. The
515 results presented in this paper are based on extractions done in the shore based lab, but
516 meanwhile the extractions for the subsequent Netherlands Mediterranean GEOTRACES
517 expedition were carried out aboard RV Pelagia by the Middag group and analysed at the
518 University of California Santa Cruz during a visit to the Bruland lab. These results are written
519 up separately, but the results for the GEOTRACES reference samples that were extracted
520 shipboard are included in Table 4 and are not distinguishable from the results for the here
521 reported samples ($P>0.05$ in all cases; two-tailed homoscedastic t-test). The good agreement
522 between the reference samples and the comparison of the gravimetric and the element
523 dilution method shows the off-line extractions are no longer required to be carried out in the
524 home laboratory and can also be done at sea, where gravimetric measurements are not
525 possible. After returning with the extracted samples to the home laboratory all samples can be
526 run in 1 long ICP-MS run (150 hours for the Mediterranean data where a total of roughly
527 2100 profile samples, reference samples, standards and blanks were analysed) creating the
528 possibility to produce the data shortly after return. We believe this is a significant
529 improvement over the shore-based extractions that often take many months to complete.
530 Taking back-up seawater samples to return to the shore laboratory is always recommended in
531 case (i) some extractions fail, or (ii) there is inadvertent contamination of some samples, or

532 (iii) if a future interest arises in other trace elements that require different extraction
533 techniques.

534

535

536 *Intercomparison between the different sampling systems*

537 The BATS intercomparison or cross-over station (31°45.92'N, 64°04.95'W) was used to
538 compare the sampling systems used by the Netherlands and US GEOTRACES programs for
539 12 dissolved trace elements. The data of all trace elements are attached in the supplementary
540 material. Since a different number of samples were collected at different depths between the
541 occupations, the R script is used to interpolate the profiles and assess the differences. Only
542 data below the main thermocline was used (deeper than 1000 m) to avoid temporal variation
543 between the two occupations that were over a year apart and had a negative and positive sea
544 surface height anomaly during the Netherlands and US occupation, respectively, indicative of
545 an up- and down-welling eddy (this was determined using the Colorado center for
546 astrodynamics research SSH data viewer at
547 http://eddy.colorado.edu/ccar/ssh/nrt_global_grid_viewer).

548 Of the first row transition metal elements, Mn, Fe, Co, Ni, Cu, Zn (in order of increasing
549 atomic number (25-30)) are all essential for every living organism (de Baar and La Roche,
550 2003). Within the group 12 elements, the second row element Cd is akin to the first row
551 element Zn, and thus Cd may substitute for Zn in phytoplankton, while an intrinsic Cd-
552 carbonic anhydrase also is found in certain diatoms (Xu et al., 2008). Next, transition metal
553 group 3 is represented by second row Y and third row La, and transition metal group 4 by its
554 first row Ti that serves as a geochemical source tracer of mineral dust (Dammshaeuser et al.,
555 2011) as well as the construction material of the Netherlands sampling frame. The non-
556 transition group 13 is represented by Al and Ga with similar ocean chemistry (Orians and

557 Bruland, 1988), and Al is the construction material of the US sampling frame. Aluminum is
558 discussed last as it is not measured with the ICP-MS method. Finally, group 14 is represented
559 by Pb, a decadal time scale tracer at BATS (Lee et al., 2011) due to the earlier use of, and
560 subsequent phasing out, of Pb in gasoline for automobiles.

561 Dissolved Mn is an element that in the open ocean usually shows a strong surface
562 enrichment due to atmospheric input (Baker et al., 2006; Landing and Bruland, 1980) and
563 photo reduction of Mn oxides (Sunda et al., 1983). Below the surface layer, Mn
564 concentrations are generally low and quite uniform due to (microbially mediated) removal by
565 scavenging (Landing and Bruland, 1980; Statham et al., 1998). The Mn profile at BATS
566 largely follows the classical profile (Fig. 2a). There are some discrepancies in the deep water
567 at the low concentrations that are visible in the *p*-value profile, but overall the results are not
568 statistically different within analytical uncertainty (*i-p*-value=0.32).

569 Dissolved Fe is arguably one of the most analytically challenging elements to determine in
570 the open ocean and considerable effort has been put in intercalibration in the past (e.g.
571 (Johnson et al., 2007)). Dissolved Fe displays characteristics of both a nutrient-type element
572 and a scavenged-type element in the open ocean which is also visible at the BATS station
573 (Fig. 2b). Surface concentrations are elevated, most likely due to recent dust input, but
574 decrease to $\sim 0.1 \text{ nmol kg}^{-1}$ around 100 m depth due to biological uptake. Dissolved Fe
575 concentrations are relatively constant close to 0.5 nmol kg^{-1} in the deep ocean, indicating
576 equilibrium between input and scavenging removal. The profiles for dissolved Fe (Fig. 2b)
577 are not statistically different within analytical uncertainty (*i-p*-value=0.50).

578 Dissolved Co, like Fe, displays characteristics of both a scavenged-type as well as a
579 nutrient-type element. The latter is due to the nutritional requirement of phytoplankton for Co
580 (e.g. Saito and Goepfert, 2008). For Co, the UV oxidation step is of utmost importance, as
581 can be seen by the difference between the data from samples with and without UV oxidation

582 (Fig. 2c). Only one replicate was analysed for the UV oxidised samples and thus the
583 analytical uncertainty was conservatively estimated at 2%. The Fisher law in the R script
584 cannot be run when only one replicate measurement is available for both profiles. To allow
585 comparison, two replicates for each data point were assumed (note that increasing the number
586 of replicates makes the test harsher). With this assumed error, the profiles are not statistically
587 different within analytical uncertainty (i - p -value=0.93).

588 Dissolved Ni is another nutrient-type element (Fig. 2d) as it is used in the breakdown of
589 urea (Price and Morel, 1991) and by nitrogen fixers (e.g.(Ho, 2013)). The general agreement
590 between the data of the Netherlands and US occupation is good, but the data from the US
591 occupation showed four outliers (3000, 2500, 1800 and 874 m depth). Outliers were defined
592 as described by Middag et al. (2011a), where a linear regression was determined between the
593 data above and below the suspect data point and a deviation of more than 10% from the
594 predicted value was used as the rejection criterion. The samples contaminated for Ni were
595 also contaminated for Cu and Zn at a roughly constant ratio of Ni:Cu:Zn but were not
596 contaminated for these metals in profiles analysed by a different PI in the US GEOTRACES
597 program (Jingfeng Wu, pers. com.), so it does not appear to be a problem with the GO-FLO
598 samplers. Additionally, the Zn samples analysed with the double spike ICP-MS method were
599 not contaminated either. When excluding those four outliers, the profiles are not statistically
600 different within analytical uncertainty (i - p -value=0.77) for Ni, further confirming that the US
601 sampling system is not the source of contamination. Either these four subsampling bottles
602 were contaminated for Ni, Cu and Zn or the contamination was introduced during subsequent
603 sample handling. The samples of these four coinciding outliers have been excluded from the
604 profiles for Ni, Cu and Zn.

605 Dissolved Cu has a hybrid type profile (Fig. 2e) with slightly depleted surface
606 concentrations as there is a biological requirement for Cu (e.g. Annett et al., 2008; Peers and

607 Price, 2006), and relatively uniform deep concentrations. For Cu, the UV oxidation step is
608 also of importance as can be seen by the difference between the data from samples with and
609 without UV oxidation (Fig. 2e). As for Co, only one replicate was analysed for the UV
610 oxidised samples and thus the error was conservatively estimated at 2% and the number of
611 replicate analyses was set at 2. With this assumed error, the profiles are not statistically
612 different within analytical uncertainty (i - p -value=0.46). There are two additional outliers (6
613 in total) for Cu at 666 and 500 m depth in the upper ocean of the US profile (not considered
614 in the comparison), implying a contamination issue for Cu of unknown origin. This is further
615 corroborated by the relatively modest agreement between the 2 profiles and further research
616 is required to resolve this, but currently any discrepancies do not seem related to the sampling
617 systems.

618 Dissolved Zn is a nutrient type element due to its biological requirement by
619 phytoplankton, most notably in the carbonic anhydrase enzyme to acquire CO₂ for
620 photosynthesis (Morel et al., 1994). Dissolved Zn is one of the most contamination prone
621 elements and each profile has one outlier (Fig. 2f; US at 2100 m and NL at 4525 m), in
622 addition to the four excluded samples from the US occupation. This demonstrates that the
623 utmost care needs to be taken during sample handling. The profiles for dissolved Zn (Fig. 2f)
624 are not statistically different within analytical uncertainty (i - p -value=0.08). Towards the
625 seafloor (deeper than 4200 m), the US samples trend to lower values, whereas the
626 Netherlands remain similar, resulting in lower p -values in the p -value profile (see
627 supplementary material). These decreasing p -values near the seafloor are also seen for Mn,
628 Fe and Co, indicating natural variation. This variation might be related to the thick benthic
629 nepheloid layer that has been observed at the BATS station during the US occupation (Lam et
630 al., 2014).

631 Dissolved Cd displays nutrient type behaviour in the ocean. This was a long standing
632 subject of debate as Cd was not previously considered a bio-essential metal (Lane and Morel,
633 2000). However, biological roles for Cd have now been identified, e.g. Cd can substitute Zn
634 in carbonic anhydrase in some phytoplankton species (Xu and Morel, 2013). The profiles for
635 dissolved Cd (Fig. 3a) are not statistically different within analytical uncertainty (*i-p-*
636 *value=0.57*).

637 Dissolved yttrium has a narrow range of concentrations at BATS, ranging between ~125
638 and 165 pmol kg⁻¹, with the highest concentrations in the deep waters (Fig. 3b). The sparse
639 amount of data that exists for Y suggests that it has somewhat of a nutrient-type distribution
640 in the oceans with slightly higher concentrations in the older Pacific waters (Alibo et al.,
641 1999). Yttrium has also been reported to have higher concentrations in the Mediterranean due
642 to the high dust input in the latter region (Alibo et al., 1999). The profiles for dissolved Y
643 (Fig. 3b) are not statistically different within analytical uncertainty (*i-p-value=0.34*).

644 The concentrations of dissolved Y have been reported to be tightly correlated with
645 dissolved La (Zhang et al., 1994). At BATS, Y correlates with La with a correlation
646 coefficient, R^2 , 0.84 ($n=60$; not shown). The relatively modest correlation between Y and La
647 is caused by a La maximum around 1500 m depth (Fig. 3c) that is absent in the Y profile,
648 indicating a source to the upper North Atlantic Deep Water for La without accompanying Y.
649 Lanthanum correlates strongly with Si with a R^2 of 0.97 ($n=60$; not shown) over the entire
650 water column. The profiles for dissolved La (Fig. 3c) are not statistically different within
651 analytical uncertainty ($p=0.42$). As part of a GEOTRACES intercalibration effort on rare
652 earth elements (REEs), Pahnke et al. (2012) reported La at several depths at the BATS station
653 from the US GEOTRACES Intercalibration cruise in 2008 and this data agrees with the here
654 reported data within analytical uncertainty (*i-p-value=0.43*). We included La as a typical
655 example of a light REE and Lu works well as a heavy REE, but for the latter the

656 concentration factor needs to be increased. A recent paper has shown that this method does
657 work well for the suite of REE's (Hatje et al., 2014).

658 The recovery for Ti was not quantitative at ~50% recovery, resulting in relatively poor
659 precision and accuracy. The concentrations were calculated based upon the standard addition
660 sensitivities assuming that all samples and standard additions had the same 50% recovery.
661 Nevertheless, the vertical profiles of both occupations agree quite well (Fig. 3e), and are not
662 statistically different within analytical uncertainty (i - p -value=0.80). Due to the low accuracy
663 it might not be suitable to look at the actual concentrations, but the observed trends should
664 represent actual variation in the ocean. However, one thing that is worth noting is that the
665 data of the Netherlands occupation is not significantly higher than the data from the US
666 occupation, indicating there is no noticeable contamination of Ti from the Netherlands
667 titanium frame.

668 Dissolved Ga is a scavenged type element with an atmospheric source to the surface ocean
669 (Orians and Bruland, 1988; Shiller, 1998). This is reflected at BATS in the elevated surface
670 concentrations, followed by a decrease with depth and relatively constant concentrations in
671 the deep waters, indicating equilibrium between input and scavenging removal (Fig. 3e). The
672 agreement between the two occupations is remarkably good considering the low overall
673 precision of the Ga measurements related to the non-quantitative recovery of Ga on the resin
674 and the profiles are not statistically different within analytical uncertainty (i - p -value=0.29).
675 The profile shape is also consistent with previous data from the region (Shiller, 1998).
676 Whereas these results are encouraging, non-quantitative recovery is not ideal and is likely to
677 result in significant variability between different measurement runs and hence, relatively
678 modest accuracy. Using isotope dilution should solve this issue. The results for Ga produced
679 with the current method should mainly be used to assess general trends rather than putting
680 emphasis on the actual concentrations or small differences.

681 Dissolved Pb is an element with a strong anthropogenic atmospheric input related to the
682 use of leaded gasoline that peaked in the mid-1970's. Since leaded gasoline is no longer used,
683 the anthropogenic input to the surface oceans has been reduced and Pb concentrations in the
684 surface and subsurface Atlantic Ocean have been declining (Lee et al., 2011), due to
685 scavenging removal of Pb. Subsurface concentrations of Pb were about 2 pmol kg⁻¹ higher
686 during the NL 2010 occupation than almost a year and a half later during the US occupation
687 of the same station. Below 2000 m, the profiles (Fig. 3f) are not statistically different within
688 analytical uncertainty (*i-p-value*=0.24).

689 Dissolved aluminium is not part of the ICP-MS multi-element method due to the poor
690 recovery of Al on the resin with the current method, especially in the presence of HF.
691 However, it is included in the discussion here as it is the material of construction for the US
692 sampling system and shipboard results are available. The result for Al by flow injection (FI)
693 method from the NL 2010 occupation and the previous US 2008 occupation with the same
694 US sampling system (Fig. 4) are not statistically different within analytical uncertainty (*i-p-*
695 *value*=0.16) and the accompanying reference samples are within 1 standard deviation from
696 each other as well as the consensus values (Table 3), verifying the accuracy of the Al data. It
697 is worth noting that the data of the Netherlands occupation is not significantly lower than the
698 data from the US occupation, indicating there is no noticeable contamination of Al from the
699 US aluminium frame. The shipboard data for the US 2011 occupation (Measures et al., in
700 press) was significantly higher (*i-p-value*<0.01) than the current data and the 2008
701 occupation. This was also noted by Middag et al. (in press) and the reason for the
702 significantly higher values reported by Measures et al. (in press) in comparison to the 2008
703 and current data is currently unknown (see Middag et al. (in press) for further details).

704

705 *Intercomparison between methods*

706 *Multi-element standard addition ICP-MS versus double spike ICP-MS*

707 The data for Fe, Zn and Cd from the US occupation of the BATS profile as well as 2
708 additional stations (station KN204-1-12 at 29.7°N, 56.8°W and station KN204-1-20 at
709 22.3°N, 35.7°W) from the two different ICP-MS methods were compared with both the t-test
710 and the R-script. The samples were drawn from the same samplers at the same station and
711 cast. These data are attached in the supplementary material and are discussed in
712 oceanographic context elsewhere ((Conway and John, 2014a; Conway and John, 2014b;
713 Conway and John, 2015; Rijkenberg et al., 2014). The results of the standard addition ICP-
714 MS method were triplicate extractions and measurements of the same sample. The standard
715 deviation of the triplicate results was used as the analytical uncertainty. The double spike
716 method is designed to obtain isotope ratios and uses more volume and samples were typically
717 extracted once and analysed twice. A 2% error was used as the analytical uncertainty
718 (Conway et al., 2013).

719 For Fe at the BATS station, the results (Fig. 5a) are not statistically different within
720 analytical uncertainty when using the R-script (*i-p-value*=0.09), but are significantly different
721 when considering the analytical uncertainty using the t-test approach (Table 5). When
722 comparing the 2 additional stations, no significant differences were observed using either test
723 for station 12, but the t-test approach reported a significant difference for station 20 (Table 5).
724 Thus the two ICP-MS methods produce comparable results, but the outcomes of the different
725 tests differ. The different test results between the t-test and the R script are related to the
726 manner of testing, where the R script is more reliable due to the simultaneous testing of both
727 differences and analytical uncertainty as detailed for Zn below.

728 For Zn at the BATS station, the results (Fig. 5b) are not statistically different using the t-
729 test approach (*p*=0.21), but the results are different within analytical uncertainty when using
730 the R-script (*i-p-value*=0.03) (Table 5). The *p-value* profile shows the main discrepancy is in

731 the upper 200m where the Zn concentrations are low (Fig. 5b and supplementary material).
732 When excluding these data points (first 8 of 37), the results are not statistically different
733 within analytical uncertainty (i - p -value=0.19). This implies there is an issue with accuracy for
734 the low Zn numbers, but small differences due to low level contamination cannot be excluded
735 either. The differing results between the t-test and R script, like for Fe, are related to the
736 manner of testing. The initial t-test does not consider analytical error and will give an
737 insignificant difference if the 2 profiles are ‘snaking’ around each other even if individual
738 differences are relatively large, hence the passing of the test for Zn. The secondary test is
739 only performed if the initial test fails (e.g. in case of Fe) and only considers the analytical
740 error and with respect to the absolute difference without regard for the directionality of the
741 difference. In contrast, the R script considers both factors simultaneously. When comparing
742 the 2 additional stations for Zn, no significant differences were observed using either test
743 (Table 5).

744 The low p -values in the p -value profile in the surface (Fig. 5b) were also observed for Fe
745 (Fig. 5a and supplementary material) and might be attributed to lower analytical
746 reproducibility for the lower concentrations, for example because of a greater contribution
747 from analytical blanks or lower signal/noise ratio during analysis, or to low level
748 contamination of the original samples, or to a combination of both factors. For the standard
749 addition ICP-MS method, the standard deviations were higher for the lowest concentrations
750 of Fe and Zn, confirming the lower reproducibility for the lower concentrations. Low level
751 contamination might arise from the used sample bottles or during filtrations, e.g. due to
752 metals leaching from the filter or particulates caught on the filter. More research is needed to
753 verify or exclude the latter hypothesis.

754 For Cd at the BATS station the results (Fig. 5c) are not statistically different within
755 analytical uncertainty when using the R-script (i - p -value=0.30), but are significantly different

756 within analytical uncertainty when using the t-test approach (Table 5). However, when
757 excluding the most deviation data points (more than 30% or 10 pmol/kg; 6 out of 37 data
758 points) the σ_d is significantly larger than the absolute difference for the entire profile and thus
759 are not different within analytical uncertainty (Table 5). When comparing the 2 additional
760 stations, no significant differences were observed using either test (Table 5). Thus the two
761 ICP-MS methods produce very comparable results, but some individual data points do differ.
762 The R script is less sensitive to these individual ‘outliers’ than the t-test approach.

763 Overall, the two ICP-MS based methods give results that are not statistically different for
764 Fe, Zn and Cd. The p-value profiles (Fig. 5 and supplementary material) show that for the
765 low surface ocean concentrations (closer to the LOD), there is more deviation between the
766 two methods. Additionally, the intercomparison shows the manner of testing is very
767 important and the R-script gives results that are more reliable due to the simultaneous testing
768 of the difference in concentration as well as the analytical uncertainty. The agreement
769 between profiles does differ between the different stations (Table 5) indicating that a single
770 cross-over station might not be representative. Multiple cross-over stations between transects
771 will often be unpractical, but repeat sampling at a cross-over station and independent analyses
772 (e.g. in different runs some weeks apart) could provide more representative intercomparisons.

773

774 *Multi-element standard addition ICP-MS versus shipboard flow injection methods*

775 The metals Mn and Fe were also analysed shipboard using flow injection methods. Both
776 methods are based on chemiluminescence produced by the reaction between hydrogen
777 peroxide and luminol which is catalyzed by Fe or Mn (see Klunder et al. (2011) and Middag
778 et al. (2011a) for details). The vertical profiles, at first sight, demonstrate agreement between
779 the shipboard methods and the standard addition ICP-MS method (Fig. 2a, 2b), but for both
780 metals there is a small discrepancy. For Mn, the profiles are not significantly different within

781 analytical uncertainty using the R script (i - p -value=0.05) but the t-test approach does give a
782 significant difference within analytical uncertainty. The shipboard data is about ~ 0.03 nmol
783 kg^{-1} lower at all depths. A similar offset was observed for the analysed reference samples
784 (Table 3), but results for both methods were within the consensus range. For the shipboard
785 data the blank was undeterminable, so no blank was subtracted. For the ICP-MS data a small
786 blank of 0.04 nmol kg^{-1} was determined and subtracted from the data. Perhaps this blank is an
787 underestimation, but that does not seem very likely since Mn is not a contamination sensitive
788 element. Adjusting either dataset for the observed offset makes them undistinguishable when
789 considering the analytical error using either test (i - p -value=0.83 and p =0.91 for the R script
790 and t-test approach, respectively), implying that this indeed is an absolute offset. However,
791 given that the offset is just insignificant based on the R-script, the deviation between the
792 profiles roughly equals the analytical uncertainty.

793 For Fe, the profiles are also not significantly different within analytical uncertainty using
794 the R-script (i - p -value=0.06), but the t-test approach does give a significant difference within
795 analytical uncertainty. The results for reference samples for both methods were within the
796 consensus range, but showed a similar discrepancy as observed for the profiles with the ICP-
797 MS data being $\sim 5\%$ higher. Since the same relative offset is observed at high and low
798 concentrations, this discrepancy appears related to a difference in calibration. For the sake of
799 argument we adjusted the shipboard data towards the ICP-MS data based on the $\sim 5\%$
800 difference observed in the reference samples. This results in the profiles being no longer
801 distinguishable using the t-test approach either (i - p -value=0.18 and p =0.82 for the R script
802 and t-test approach, respectively). As for Mn the deviation between the profiles is just smaller
803 than the overall analytical uncertainty.

804 It should be realised that the consensus values for the reference samples have a range and
805 therefore have limited suitability to be able distinguish small differences. Similarly, when

806 comparing to previously published data, there is inherent variation due to discrepancies in the
807 sampling depths as well as truly natural spatial and temporal variations that can mask small
808 differences between methods. Both the ICP-MS and shipboard data included reference
809 sample analysis and these agreed within the consensus range. When looking in detail, it is
810 noticed that there appears to be a small discrepancy between the vertical profiles (ICP-MS vs
811 shipboard) of dissolved Fe and Mn that is also represented between the reference samples,
812 but usually this would have been ignored as both were within the range of reproducibility and
813 accuracy. The small difference observed between the shipboard and ICP-MS methods is just
814 insignificant, but does warrant monitoring.

815

816 **Conclusion**

817 The intercomparison of the different Netherlands and US sampling system by analysing
818 samples from the BATS cross-over station with the same method and analyst demonstrated
819 that both sampling systems are capable of collecting uncontaminated samples for all the
820 GEOTRACES key elements. Nevertheless, at all times throughout the procedure, utmost care
821 must be taken to avoid inadvertent contamination. All results for reference samples are in
822 agreement with the current consensus values for those elements that have a consensus value
823 available. Moreover, the data reported here for trace metal profiles obtained with different
824 analytical methods are consistent, demonstrating comparability between techniques. This
825 implies the analytical ICP-MS methods and the shipboard FI methods are functioning as
826 expected and produce accurate data that can be reliably compared between cruises, with the
827 caveat of Ga and Ti profiles, for which we did not yet achieve quantitative recovery.

828 Despite the strong consistency between collection and analytical techniques presented
829 here, ultimately it is realised that all measurements have an uncertainty associated with them.
830 Comparison of results from the exact same station i.e. measuring sub-samples from the same

831 sample with different techniques allowed us to discern small differences between the
832 shipboard measurements and the ICP-MS measurements. We were fortunate enough to run
833 sub-samples from the same sample on the different methods, but these comparisons should
834 also be possible from cross-over station as long as that station is sampled at common depths
835 and in a relatively short timeframe to minimize natural variation, highlighting the importance
836 of cross-over stations to the GEOTRACES program.

837

838 *Comments and recommendations*

839 For the design of future crossover stations it is advisable to optimise the coordination of
840 sampling depths, or preferably common neutral density horizons. This would greatly benefit
841 the intercomparison, but might not aid intercomparison in the surface ocean if large variations
842 occur between occupations. For comparisons between analytical techniques or sampling
843 systems the deep ocean is simply best suited to make these comparisons. Comparison in the
844 surface ocean is inherently less accurate due to the larger interpolation necessary and the fact
845 that some processes (such as sinking and remineralising particles) are a function of depth as
846 well as density.

847 In order to merge or link data from different cruises, investigators and nations into a
848 coherent global data set, cross-over stations will be an invaluable tool. This has been done for
849 example by the CARINA program for ocean CO₂ and ancillary measurements (Olsen et al.,
850 2009). It should be realised though, that one cross-over station in a long section is not a
851 definitive answer to resolve potential differences. It will only work if every data set is also
852 validated by regular measurements of the international reference samples to ensure
853 consistency within the international community and merged database.

854 A remark here with respect to the reference samples is that these should not only be
855 analysed, but the results also need to be submitted to the reference sample coordinator (K.W.

856 Bruland; bruland@ucsc.edu) in order to have the most representative updated consensus
857 values available to the community. Furthermore, since the consensus values are not certified
858 or constant, no corrections should be made to datasets based on analyses of reference
859 standards; the values should just be part of the metadata. Internal seawater reference samples
860 are most useful to identify any intra-dataset variability (i.e. variation between analyses
861 performed on different days). The routine use of an internal seawater reference samples to be
862 analysed far more frequently than the international reference samples is recommended,
863 simply due to the scarcity of the international reference samples.

864

865 **Acknowledgements**

866

867 We express our gratitude to the masters and crew of R/V Pelagia and R/V Knorr for their
868 support during the GEOTRACES expeditions GA02-PE321 and GA03-KN204, respectively.
869 The significant contributions of Patrick Laan and Geoffrey Smith in expedition preparation
870 are crucial and invaluable. Major contributions have been made by both the Netherlands 2010
871 and the US 2011 GEOTRACES trace metal teams for shipboard sample collection and
872 filtration. In addition Patrick Laan performed the shipboard dissolved Fe measurements
873 during Pelagia cruise GA02-PE321 (Rijkenberg et al., 2014). The authors thank Matthew
874 Brown for kindly making available his unpublished dissolved aluminium data from BATS
875 collected and analysed on shipboard during the preceding US GEOTRACES IC1-KN193-5
876 cruise aboard RV Knorr in June 2008. Special thanks to Rob Franks for his support and ideas
877 during method development and analysis, this work would not have been possible without
878 him. We also thank A. Rosenberg and B. Bair for technical assistance at the University of
879 South Carolina, and J. Wu for data which informed initial double-spike additions. Last but
880 not least, the authors want to acknowledge the chief scientists of the expeditions Micha

881 Rijkenberg (GA02-PE321), Ed Boyle and Greg Cutter (GA03-KN204) for their support and
882 overall coordination of the BATS intercomparison station. This research is funded by the
883 Netherlands Organization for Scientific Research (NWO) project grants 820.01.014
884 (GEOTRACES Netherlands-USA Joint Effort on Trace Metals in the Atlantic Ocean) and
885 839.08.410 (GEOTRACES, Global Change and Microbial Oceanography in the West
886 Atlantic Ocean), the USA National Science Foundation (NSF) grants: OCE-0961579 and
887 OCE-1131387 and the department of Chemistry of the University of Otago.
888

889 **References**

- 890 Alibo, D.S. and Nozaki, Y., 1999. Rare earth elements in seawater: particle association,
891 shale-normalization, and Ce oxidation. *Geochimica et Cosmochimica Acta*, 63(3–4):
892 363-372.
- 893 Alibo, D.S., Nozaki, Y. and Jeandel, C., 1999. Indium and yttrium in North Atlantic and
894 Mediterranean waters: comparison to the Pacific data. *Geochimica et Cosmochimica*
895 *Acta*, 63(13–14): 1991-1999.
- 896 Amakawa, H., Alibo, D.S. and Nozaki, Y., 1996. Indium concentration in Pacific seawater.
897 *Geophysical Research Letters*, 23(18): 2473-2476.
- 898 Annett, A.L., Lapi, S., Ruth, T.J. and Maldonado, M.T., 2008. The effects of Cu and Fe
899 availability on the growth and Cu : C ratios of marine diatoms. *Limnology and*
900 *Oceanography*, 53(6): 2451-2461.
- 901 Baker, A.R., Jickells, T.D., Witt, M. and Linge, K.L., 2006. Trends in the solubility of iron,
902 aluminium, manganese and phosphorus in aerosol collected over the Atlantic Ocean.
903 *Marine Chemistry*, 98(1): 43-58.
- 904 Biller, D.V. and Bruland, K.W., 2012. Analysis of Mn, Fe, Co, Ni, Cu, Zn, Cd, and Pb in
905 seawater using the Nobias-chelate PA1 resin and magnetic sector inductively coupled
906 plasma mass spectrometry (ICP-MS). *Marine Chemistry*, 130–131(0): 12-20.
- 907 Bowie, A.R. et al., 2003. Shipboard analytical intercomparison of dissolved iron in surface
908 waters along a north–south transect of the Atlantic Ocean. *Marine Chemistry*, 84(1–
909 2): 19-34.
- 910 Conway, T.M. and John, S.G., 2014a. The biogeochemical cycling of zinc and zinc isotopes
911 in the North Atlantic Ocean. *Global Biogeochemical Cycles*, 28(10): 2014GB004862.
- 912 Conway, T.M. and John, S.G., 2014b. Quantification of dissolved iron sources to the North
913 Atlantic Ocean. *Nature*, 511(7508): 212-215.

914 Conway, T.M. and John, S.G., 2015. Biogeochemical cycling of cadmium isotopes along a
915 high-resolution section through the North Atlantic Ocean. *Geochimica et*
916 *Cosmochimica Acta*, 148(0): 269-283.

917 Conway, T.M., Rosenberg, A.D., Adkins, J.F. and John, S.G., 2013. A new method for
918 precise determination of iron, zinc and cadmium stable isotope ratios in seawater by
919 double-spike mass spectrometry. *Analytica Chimica Acta*, 793: 44-52.

920 Croot, P.L., 2011. Rapid Determination of Picomolar Titanium in Seawater with Catalytic
921 Cathodic Stripping Voltammetry. *Analytical Chemistry*, 83(16): 6395-6400.

922 Cutter, G.A. and Bruland, K.W., 2012. Rapid and noncontaminating sampling system for
923 trace elements in global ocean surveys. *Limnology and Oceanography-Methods*, 10:
924 425-436.

925 Dammschaeuser, A., Wagener, T. and Croot, P.L., 2011. Surface water dissolved aluminum
926 and titanium: Tracers for specific time scales of dust deposition to the Atlantic?
927 *Geophysical Research Letters*, 38.

928 de Baar, H.J.W. et al., 2012. Performance of novel ultraclean pristine samplers for trace
929 metals in the GEOTRACES Program: Intercomparison with GOFLO Samplers at the
930 Bermuda Atlantic Site. Oral presentation at the 2012 Ocean Sciences Meeting 20-02-
931 2012, Salt Lake City, USA.

932 de Baar, H.J.W. and La Roche, J., 2003. Trace metals in the oceans: Evolution, biology and
933 global change. *Marine Science Frontiers for Europe*, 79-105 pp.

934 De Baar, H.J.W. et al., 2008. Titan: A new facility for ultraclean sampling of trace elements
935 and isotopes in the deep oceans in the international Geotraces program. *Marine*
936 *Chemistry*, 111(1-2): 4-21.

937 Glasby, G.P. and Schulz, H.D., 1999. E-H, pH diagrams for Mn, Fe, Co, Ni, Cu and As under
938 seawater conditions: Application of two new types of E-H, pH diagrams to the study
939 of specific problems in marine geochemistry. *Aquatic Geochemistry*, 5(3): 227-248.

940 Greaves, M.J., Rudnicki, M. and Elderfield, H., 1991. Rare earth elements in the
941 Mediterranean Sea and mixing in the Mediterranean outflow. *Earth and Planetary
942 Science Letters*, 103(1-4): 169-181.

943 Hatje, V., Bruland, K.W. and Flegal, A.R., 2014. Determination of rare earth elements after
944 pre-concentration using NOBIAS-chelate PA-1 (R) resin: Method development and
945 application in the San Francisco Bay plume. *Marine Chemistry*, 160: 34-41.

946 Ho, T.-Y., 2013. Nickel limitation of nitrogen fixation in *Trichodesmium*. *Limnology and
947 Oceanography*, 58(1): 112-120.

948 Hudson, R.J.M., Covault, D.T. and Morel, F.M.M., 1992. Investigations of iron coordination
949 and redox reactions in seawater using Fe-59 radiometry and ion-pair solvent
950 extraction of amphiphilic iron complexes. *Marine Chemistry*, 38(3-4): 209-235.

951 John, S.G. and Adkins, J.F., 2010. Analysis of dissolved iron isotopes in seawater. *Marine
952 Chemistry*, 119(1-4): 65-76.

953 Johnson, K.S. et al., 2007. Developing standards for dissolved iron in seawater. *Eos,
954 Transactions American Geophysical Union*, 88(11): 131-132.

955 Klunder, M.B., Laan, P., Middag, R., De Baar, H.J.W. and van Ooijen, J.C., 2011. Dissolved
956 iron in the Southern Ocean (Atlantic sector). *Deep-Sea Research Part II-Topical
957 Studies in Oceanography*, 58(25-26): 2678-2694.

958 Lam, P.J., Ohnemus, D.C. and Auro, M.E., 2014. Size-fractionated major particle
959 composition and concentrations from the US GEOTRACES north Atlantic zonal
960 transect. *Deep Sea Research Part II: Topical Studies in Oceanography*(0).

961 Landing, W.M. and Bruland, K.W., 1980. Manganese in the North Pacific. Earth and
962 Planetary Science Letters, 49(1): 45-56.

963 Landing, W.M. et al., 1995. Analytical intercomparison results from the 1990
964 Intergovernmental Oceanographic Commission open-ocean baseline survey for trace
965 metals: Atlantic Ocean. Marine Chemistry, 49(4): 253-265.

966 Lane, T.W. and Morel, F.M.M., 2000. A biological function for cadmium in marine diatoms.
967 Proceedings of the National Academy of Sciences of the United States of America,
968 97(9): 4627-4631.

969 Lee, J.-M. et al., 2011. Analysis of trace metals (Cu, Cd, Pb, and Fe) in seawater using single
970 batch nitrilotriacetate resin extraction and isotope dilution inductively coupled plasma
971 mass spectrometry. Analytica Chimica Acta, 686(1-2): 93-101.

972 Measures, C., Hatta, M., Fitzsimmons, J. and Morton, P., in press. Dissolved Al in the zonal
973 N Atlantic section of the US GEOTRACES 2010/2011 cruises and the importance of
974 hydrothermal inputs. Deep Sea Research Part II: Topical Studies in Oceanography(0).

975 Middag, R., de Baar, H.J.W., Laan, P., Cai, P.H. and van Ooijen, J.C., 2011a. Dissolved
976 manganese in the Atlantic sector of the Southern Ocean. Deep-Sea Research Part Ii-
977 Topical Studies in Oceanography, 58(25-26): 2661-2677.

978 Middag, R., van Hulst, M.M.P., van Aken, H.M., Rijkenberg, M.J.A., Gerringa, L.J.A.,
979 Laan, P., de Baar, H.J.W., in press. Dissolved aluminium in the Ocean Conveyor of
980 the West Atlantic Ocean: effects of the biological cycle, scavenging, sediment
981 resuspension and hydrography.. In press with Marine Chemistry.

982 Middag, R., van Slooten, C., de Baar, H.J.W. and Laan, P., 2011b. Dissolved aluminium in
983 the Southern Ocean. Deep-Sea Research Part Ii-Topical Studies in Oceanography,
984 58(25-26): 2647-2660.

985 Milne, A., Landing, W., Bizimis, M. and Morton, P., 2010. Determination of Mn, Fe, Co, Ni,
986 Cu, Zn, Cd and Pb in seawater using high resolution magnetic sector inductively
987 coupled mass spectrometry (HR-ICP-MS). *Analytica Chimica Acta*, 665(2): 200-207.

988 Morel, F.M.M. et al., 1994. Zinc and carbon co-limitation of marine phytoplankton. *Nature*,
989 369(6483): 740-742.

990 Olsen, A. et al., 2009. Overview of the Nordic Seas CARINA data and salinity
991 measurements. *Earth Syst. Sci. Data Discuss.*, 2(1): 1-25.

992 Orians, K.J. and Bruland, K.W., 1988. The Marine Geochemistry of Dissolved Gallium - a
993 Comparison with Dissolved Aluminum. *Geochimica Et Cosmochimica Acta*, 52(12):
994 2955-2962.

995 Pahnke, K. et al., 2012. GEOTRACES intercalibration of neodymium isotopes and rare earth
996 element concentrations in seawater and suspended particles. Part 2: Systematic tests
997 and baseline profiles. *Limnology and Oceanography-Methods*, 10: 252-269.

998 Peers, G. and Price, N.M., 2006. Copper-containing plastocyanin used for electron transport
999 by an oceanic diatom. *Nature*, 441(7091): 341-344.

1000 Price, N.M. and Morel, F.M.M., 1991. Colimitation of phytoplankton growth by nickel and
1001 nitrogen. *Limnology and Oceanography*, 36(6): 1071-1077.

1002 R Core Team, 2013. R: A language and environment for statistical computing. R Foundation
1003 for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

1004 Rijkenberg, M.J.A. et al., submitted. "PRISTINE", a new high volume sampler for ultraclean
1005 sampling of trace metals and isotopes. *Marine Chemistry*.

1006 Rijkenberg, M.J.A. et al., 2014. The Distribution of Dissolved Iron in the West Atlantic
1007 Ocean. *Plos One*, 9(6).

1008 Saito, M.A. and Goepfert, T.J., 2008. Zinc-cobalt colimitation of *Phaeocystis antarctica*.
1009 *Limnology and Oceanography*, 53(1): 266-275.

1010 Shiller, A.M., 1998. Dissolved gallium in the Atlantic Ocean. *Marine Chemistry*, 61(1-2): 87-
1011 99.

1012 Sohrin, Y. et al., 2008. Multielemental Determination of GEOTRACES Key Trace Metals in
1013 Seawater by ICPMS after Preconcentration Using an Ethylenediaminetriacetic Acid
1014 Chelating Resin. *Analytical Chemistry*, 80(16): 6267-6273.

1015 Statham, P.J., Yeats, P.A. and Landing, W.M., 1998. Manganese in the eastern Atlantic
1016 Ocean: processes influencing deep and surface water distributions. *Marine Chemistry*,
1017 61(1-2): 55-68.

1018 Sunda, W.G., Huntsman, S.A. and Harvey, G.R., 1983. Photo-reduction of manganese oxides
1019 in seawater and its geochemical and biological implications. *Nature*, 301(5897): 234-
1020 236.

1021 The GEOTRACES group, in press. The GEOTRACES Intermediate Data Product 2014.
1022 *Marine Chemistry*.

1023 van de Flieddt, T. et al., 2012. GEOTRACES intercalibration of neodymium isotopes and rare
1024 earth element concentrations in seawater and suspended particles. Part 1:
1025 reproducibility of results for the international intercomparison. *Limnology and*
1026 *Oceanography-Methods*, 10: 234-251.

1027 Welch, B.L., 1947. The Generalization of 'Student's' Problem when Several Different
1028 Population Variances are Involved. *Biometrika*, 34(1/2).

1029 Xu, Y., Feng, L., Jeffrey, P.D., Shi, Y. and Morel, F.M.M., 2008. Structure and metal
1030 exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature*, 452(7183):
1031 56-U3.

1032 Xu, Y. and Morel, F.M.M., 2013. Cadmium in Marine Phytoplankton. In: A. Sigel, H. Sigel
1033 and R.K.O. Sigel (Editors), *Cadmium: From Toxicity to Essentiality. Metal Ions in*
1034 *Life Sciences*, pp. 509-528.

1035 Zhang, J., Amakawa, H. and Nozaki, Y., 1994. The comparative behaviors of Yttrium and
1036 Lanthanides in the seawater of the North Pacific. *Geophysical Research Letters*,
1037 21(24): 2677-2680.
1038
1039

1040 **Tables**

1041 Table 1 overview of the comparisons made in this paper

Comparison	Parameters
Netherlands sampling system vs US sampling system	Mn, Fe, Co, Ni, Cu, Zn, Cd, Y, La, Ti, Ga, Pb via ICP-MS (same analyst)
Netherlands sampling system vs US sampling system	Al via flow injection (different analysts)
Standard addition ICP-MS vs double spike isotope dilution ICP-MS	Fe, Zn and Cd
Standard addition ICP-MS vs shipboard flow injection	Fe and Mn

1042

1043 Table 2 Blanks and Limit of Detection (LOD) defined as three times the standard deviation of
 1044 the blank (average for the 8 different columns).

element	Unit	Blank	LOD
Y	pmol kg ⁻¹	0.36	0.21
Cd	pmol kg ⁻¹	0.11	0.10
La	pmol kg ⁻¹	0.06	0.08
Pb	pmol kg ⁻¹	0.94	0.16
Ti	pmol kg ⁻¹	4.22	3.97
Mn	nmol kg ⁻¹	0.03	0.02
Fe	nmol kg ⁻¹	0.01	0.02
Co	pmol kg ⁻¹	0.52	0.22
Ni	nmol kg ⁻¹	0.07	0.01
Cu	nmol kg ⁻¹	0.01	0.01
Zn	nmol kg ⁻¹	0.03	0.02
Ga	pmol kg ⁻¹	1.57	1.19

1045

1046 Table 3 SAFe and GEOTRACES reference sample results (n=4) and consensus values. Errors

1047 denoted one σ standard deviation.

Metal Unit	Safe S1 91		Safe D2 492		Geotraces S 138		Geotraces D 146	
	Meas.	Cons.	Meas.	Cons.	Meas.	Cons.	Meas.	Cons.
Y pmol kg ⁻¹	62±1		198±2		130±2		138±2	
Cd pmol kg ⁻¹	1.3±0.3	1.1±0.3	977±15	986±23	2.2±0.3	2.1±0.6	262±5	272±6
La pmol kg ⁻¹	4.5±0.1		30.6±0.5		13.3±0.3	14.7±1.1 [^]	22.0±0.5	23.6±1.4 [^]
Pb pmol kg ⁻¹	48.3±1.1	48.0±2.2	27.7±0.5	27.7±1.5	28.7±0.5	28.7±0.8	42.1±1.3	42.3±1.3
Ti pmol kg ⁻¹	23±8	48±14*	140±32	127±33*	66±19	75±12*	221±61	202±18*
Mn nmol kg ⁻¹	0.82±0.02	0.79±0.06	0.36±0.02	0.35±0.06	1.53±0.04	1.50±0.11	0.20±0.01	0.21±0.03
Mn SB nmol kg ⁻¹			0.33±0.01		1.47±0.03		0.18±0.01	
Fe nmol kg ⁻¹	0.10±0.01	0.093±0.008	0.97±0.04	0.933±0.023	0.53±0.04	0.546±0.046	0.94±0.04	1.00±0.10
Fe SB nmol kg ⁻¹			0.89±0.03					
Co pmol kg ⁻¹	3.9±0.3	4.9±1.5	46.9±1	45.8±3.0	32.4±1	31.8±1.3	64.0±0.9	64.5±1.8
Ni nmol kg ⁻¹	2.21±0.09	2.29±0.08	8.50±0.17	8.59±0.28	2.05±0.06	2.09±0.05	4.04±0.07	3.97±0.07
Cu nmol kg ⁻¹	0.5±0.02	0.53±0.05	2.2±0.1	2.32±0.16	0.8±0.1	0.83±0.08	1.5±0.1	1.55±0.13
Zn nmol kg ⁻¹	0.07±0.03	0.068±0.010	7.50±0.20	7.47±0.27	0.04±0.02	0.042±0.008	1.76±0.06	1.73±0.09
Ga pmol kg ⁻¹	11.1±0.8		7.0±0.4		41.4±2		29.0±1.5	
Al SB nmol kg ⁻¹					27.4±0.3	27.5±0.2	17.5±0.3	17.7±0.2
Al SB US nmol kg ⁻¹					27.4±0.5		17.8±0.3	

1048 [^] Not actual consensus data but results reported by van de Flierdt et al. (2014)

1049 *Not actual consensus data but results reported by Croot (2011)

1050 SB denotes shipboard data

- 1051 Table 4 SAFe reference samples comparison of the results in Table 3 with the Mediterranean
 1052 GEOTRACES expedition where shipboard extractions were performed (n=4)

Metal Unit	Safe S1 564	2-tailed paired t-test vs results accompanying BATS	Safe D2 54	2-tailed paired t-test vs results accompanying BATS
Y pmol kg ⁻¹	63±1	0.16	197±2	0.71
Cd pmol kg ⁻¹	1.5±0.1	0.21	976±12	0.91
La pmol kg ⁻¹	4.7±0.1	0.07	31.4±0.4	0.07
Pb pmol kg ⁻¹	49.1±0.2	0.2	27.4±0.2	0.31
Ti pmol kg ⁻¹	16±1	0.14	116±7	0.20
Mn nmol kg ⁻¹	0.81±0.01	0.60	0.38±0.01	0.09
Fe nmol kg ⁻¹	0.10±0.02	0.73	0.94±0.01	0.06
Ni nmol kg ⁻¹	2.28±0.02	0.17	8.62±0.18	0.40
Zn nmol kg ⁻¹	0.08±0.02	0.80	7.44±0.07	0.53
Ga pmol kg ⁻¹	11.2±0.9	0.91	8.5±1.2	0.06

1053

1054 Table 5 Intercomparison results between subsamples of the same profiles for 3 stations in the
 1055 North Atlantic Ocean using Double spike isotope dilution ICP-MS versus standard addition
 1056 ICP-MS. Both statistical testing approaches were used (R script and t-test approach). Please
 1057 not that for the initial t-test a $p > 0.05$ implies no significant differences between the profiles,
 1058 whereas for the secondary t-test a $p < 0.05$ implies the two methods are not statistically
 1059 different within uncertainty.

	Double Spike					
	BATS		KN204-1-12		KN204-1-20	
	t-test <i>p-value</i>	R script <i>i-p-value</i>	t-test <i>p-value</i>	R script <i>i-p-value</i>	t-test <i>p-value</i>	R script <i>i-p-value</i>
Standard addition Fe	0.99*	0.09	0.28#	0.16	0.99*	0.07
Standard addition Zn	0.22#	0.03	0.10#	0.11	0.47#	0.08
Standard addition Cd	0.11*	0.30	0.02\$	0.75	0.41#	0.10

1060 # initial t-test gives no significant difference ($p > 0.05$)

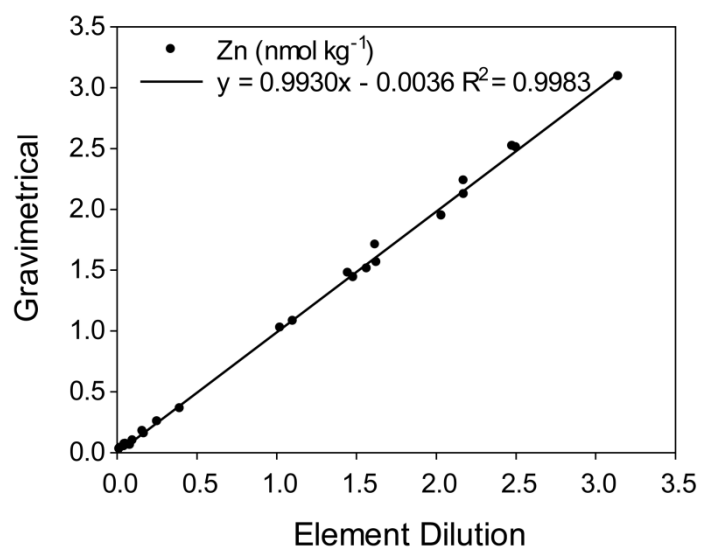
1061 * T-test result when considering the analytical uncertainty > 0.05 (σ_d not greater than
 1062 absolute difference between data points and the analytical results of the two methods are
 1063 statistically different when considering uncertainty)

1064 \$ T-test result when considering the analytical uncertainty < 0.05 (σ_d is greater than absolute
 1065 difference than data points and the analytical results of the two methods are not statistically
 1066 different within uncertainty)

1067 For the R script, a *i-p-value* > 0.05 means the analytical results of the two methods are not
 1068 statistically different within uncertainty

1069

1070 **Figures**

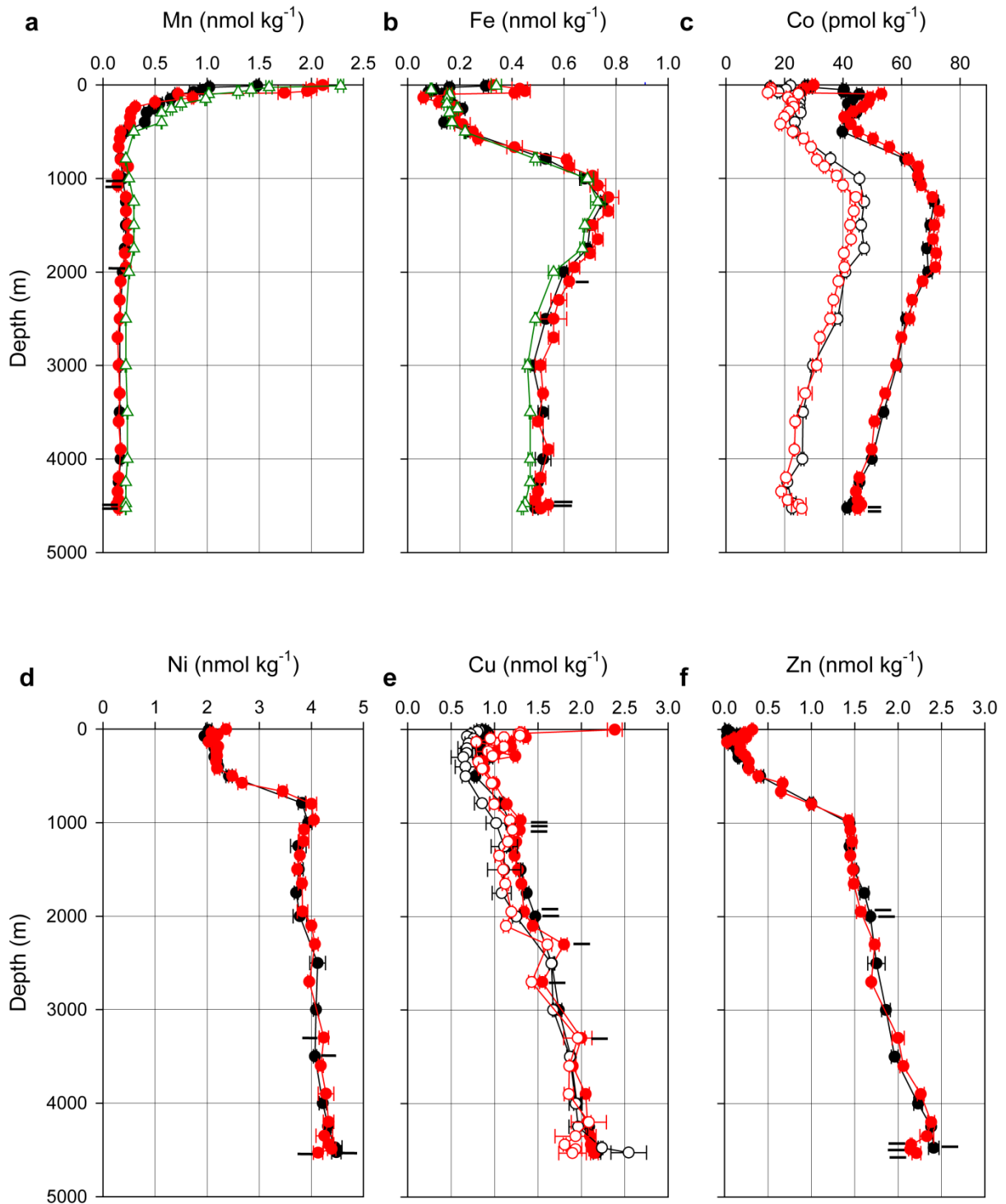


1071

1072 Figure 1. Results of the element dilution calibration vs. the gravimetical calibration for the 24

1073 samples of the Netherlands BATS occupation for dissolved Zn concentration, by standard

1074 addition ICP-MS.



1075

1076

Figure 2. Concentrations of dissolved trace metals (by standard addition ICP-MS) for the

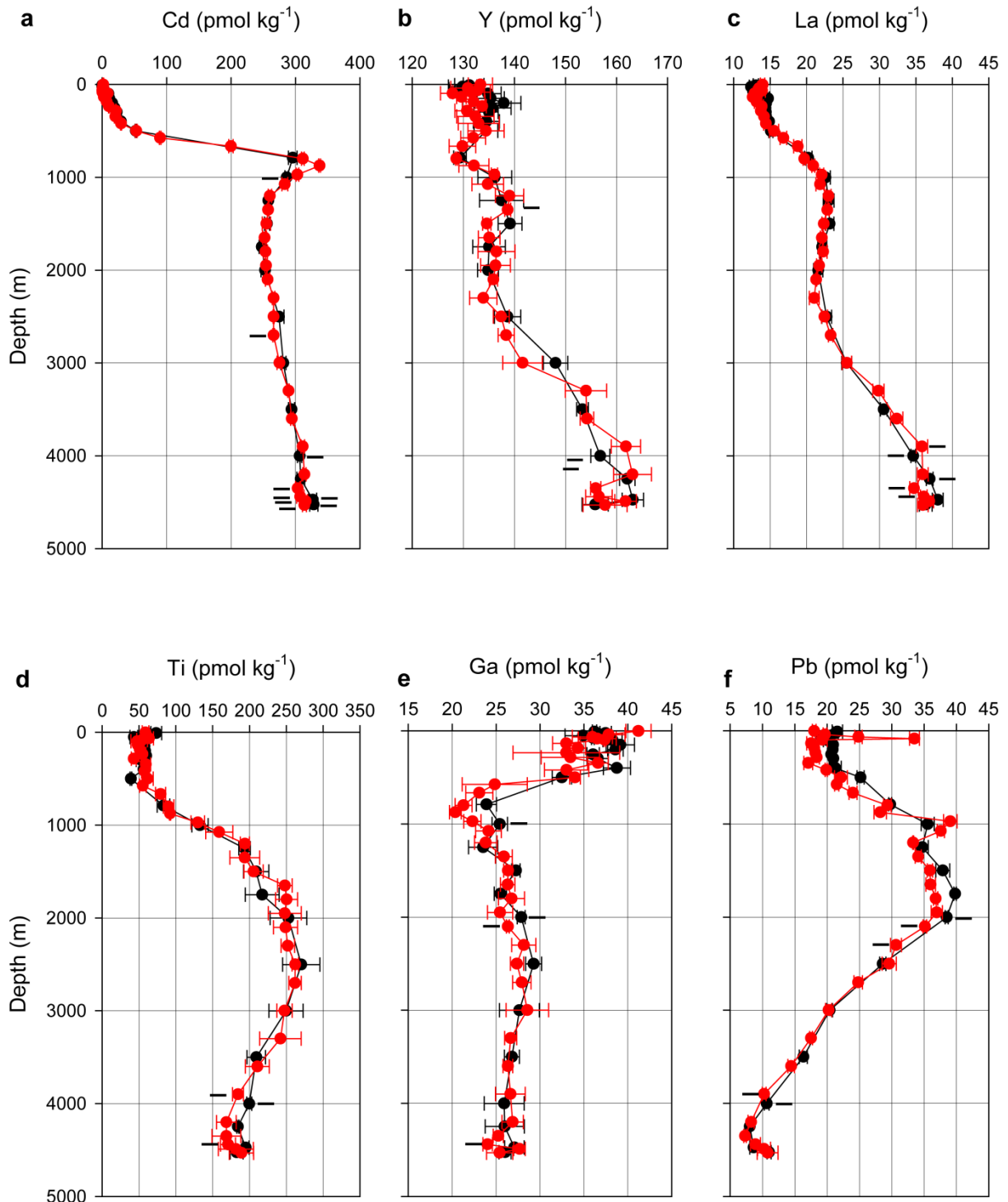
1077 Netherlands (black circles) and US (red circles) occupations of the BATS station for A) Mn

1078 (nmol kg⁻¹); B) Fe (nmol kg⁻¹); C) Co (pmol kg⁻¹) with and without UV oxidation; D) Ni

1079 (nmol kg⁻¹); E) Cu (nmol kg⁻¹) with and without UV oxidation; F) Zn (nmol kg⁻¹). Shipboard

1080 data are denoted by green triangles for Fe and Mn. Open symbols denote no UV oxidation for

1081 Co and Cu. On each graph, a black horizontal bar indicates a significant difference ($p < 0.05$)
1082 between samples at individual depths (deeper than 1000 m) between the US and Netherlands'
1083 profiles (UV oxidised in the case of Co and Cu) observed in the p -value profile (see *Different*
1084 *samples approach* for calculation and the supplementary material for the p -value profiles). .
1085 Note that the integrated p -value (i - p -value) reported in the text is based on the profiles deeper
1086 than 1000 m depth and does not equal the average of the p -values at individual depths (see
1087 *Different samples approach*).



1088

1089

Figure 3. Concentrations of dissolved trace metals (by standard addition ICP-MS) for the US

1090

(red circles) and Netherlands (black circles) occupations of the BATS station for A) Cd (pmol

1091

kg⁻¹); B) Y (pmol kg⁻¹); C) La (pmol kg⁻¹); D) Ti (pmol kg⁻¹) E) Ga (pmol kg⁻¹); F) Pb (pmol

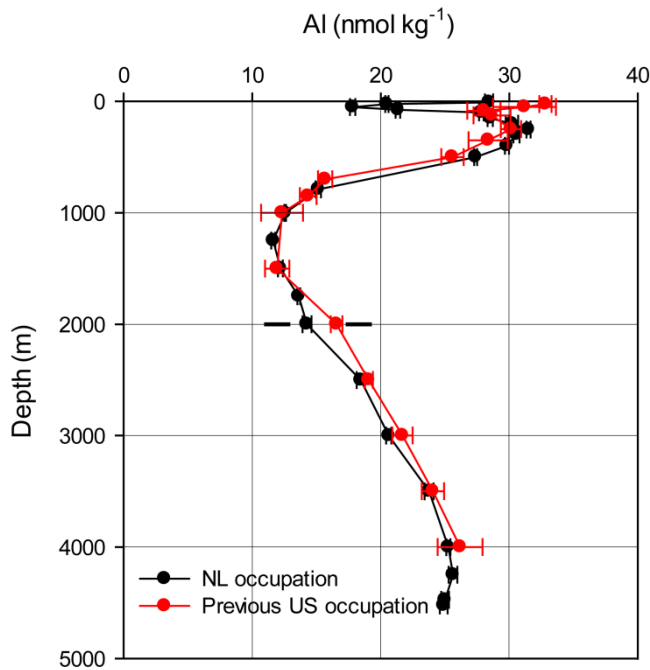
1092

kg⁻¹). On each graph, a black horizontal bar indicates a significant difference ($p < 0.05$)

1093

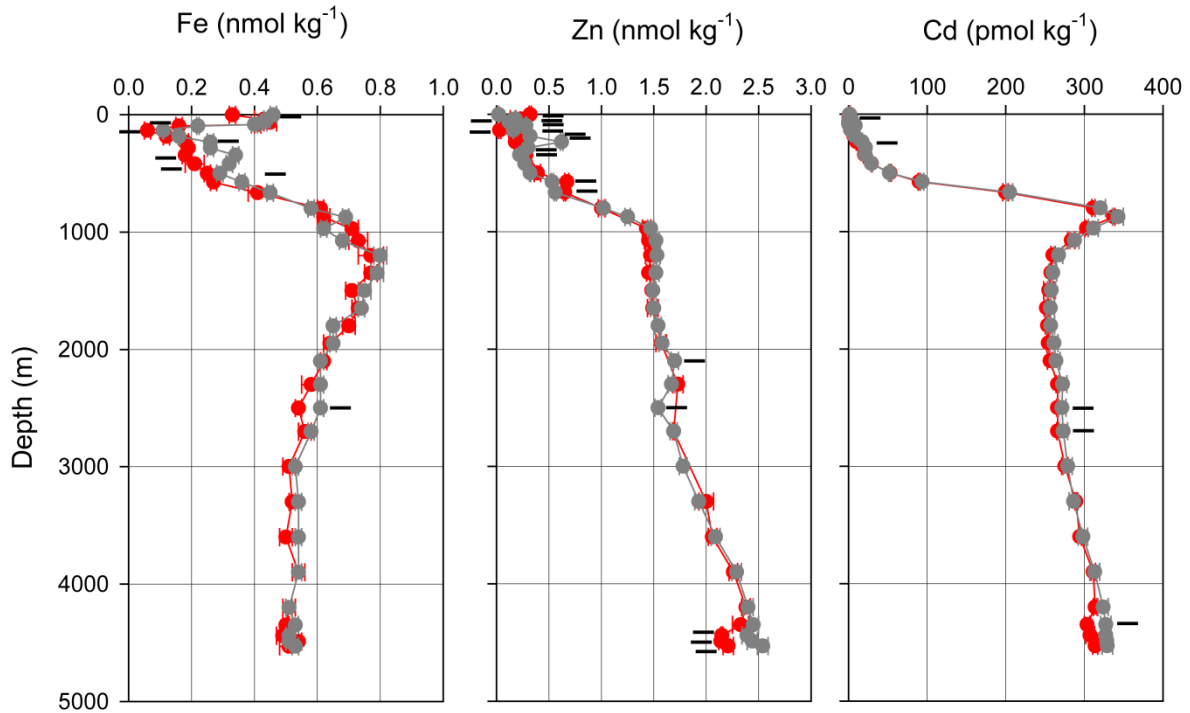
between samples at individual depths (deeper than 1000 m) between the US and Netherlands'

1094 profiles observed in the p -value profile (see *Different samples approach* for calculation and
1095 the supplementary material for the p -value profiles). Note that the integrated p -value (i - p -
1096 value) reported in the text is based on the profiles deeper than 1000 m depth, except for Pb
1097 for which the test was based on depths deeper than 2000 m, and this i - p -value does not equal
1098 the average of the p -values at individual depths (see *Different samples approach*).



1099
1100 Figure 4. Concentrations of Al (nmol kg⁻¹) from shipboard analysis during the Netherlands
1101 (black circles) and a previous (2008) US (red circles) occupation of the BATS station. The
1102 US data was analysed by Matt Brown from the University of California Santa Cruz (UCSC)
1103 and is the mean and standard deviation (error bars) of different samples filtered with different
1104 filtration systems as part of the initial Intercalibration efforts of the GEOTRACES program.
1105 The NL data is the mean and standard deviation (error bars) of triplicate analysis of the same
1106 sample. A black horizontal line denotes a significant difference between samples at
1107 individual depths (in the 1000 to 4000 m depth range) between the two profiles observed in
1108 the p -value profile (see *Different samples approach* for calculation and the supplementary
1109 material for the p -value profile). Note that the integrated p -value (i - p -value) reported in the

1110 text is based on the profiles between 1000 m and 4000 m depth. The excellent agreement is
 1111 firm evidence by ocean crossover station approach of the accuracy of the measurements of
 1112 dissolved Al in seawater by both laboratories (UCSC and NIOZ). This data was previously
 1113 presented in a different form by Cutter and Bruland (2012).



1114
 1115 Figure 5. Concentrations of trace metals for the standard addition ICP-MS method (red
 1116 circles) and double spike isotope dilution ICP-MS method (grey circles) analyses at the US
 1117 BATS station occupation for A) Cd (pmol kg⁻¹); B) Zn (nmol kg⁻¹); C) Fe (nmol kg⁻¹). A
 1118 black horizontal bar indicates a significant difference ($p < 0.05$) between samples at individual
 1119 depths (whole water column) between the US and Netherlands' profiles observed in the p -
 1120 value profile (see *Different samples approach* for calculation and the supplementary material
 1121 for the p -value profiles). Note that the integrated p -value (i - p -value) reported in the text is
 1122 based on the entire profile as these samples were collected simultaneously.

1123
 1124