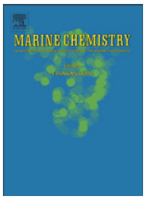
Marine Chemistry 171 (2015) 36–43

Contents lists available at ScienceDirect

Marine Chemistry

journal homepage: www.elsevier.com/locate/marchem



An inter-laboratory comparison assessing the quality of seawater carbon dioxide measurements

Emily E. Bockmon *, Andrew G. Dickson

Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0244, USA

ARTICLE INFO

Article history:
Received 25 April 2014
Received in revised form 7 January 2015
Accepted 4 February 2015
Available online 11 February 2015

Keywords:
Carbonate chemistry
Intercomparison
Carbon dioxide
Alkalinity
Dissolved inorganic carbon
pH
Ocean acidification

ABSTRACT

Seawater CO_2 measurements are being made with increasing frequency as interest grows in the ocean's response to changing atmospheric CO_2 levels and to climate change. The ultimate usefulness of these measurements depends on the data quality and consistency. An inter-laboratory comparison was undertaken to help evaluate and understand the current reliability of seawater CO_2 measurements. Two seawater test samples of different CO_2 content were prepared according to the usual method for the creation of seawater reference materials in the Dickson Laboratory at Scripps Institution of Oceanography. These two test samples were distributed in duplicate to more than 60 laboratories around the world. The laboratories returned their measurement results for one or more of the following parameters: total alkalinity (A_T), total dissolved inorganic carbon (C_T), and pH, together with information about the methods used and the expected uncertainty of the measurements. The majority of laboratories reported A_T and C_T values for all their measurements that were within 10 μmol kg^{-1} of the assigned values (i.e. within ±0.5%), however few achieved results within 2 μmol kg^{-1} (i.e. within ±0.1%), especially for C_T. Results for the analysis of pH were quite scattered, with little suggestion of a consensus value. The high-CO_2 test sample produced results for both C_T and pH that suggested in many cases that CO_2 was lost during analysis of these parameters. This study thus documents the current quality of seawater CO_2 measurements in the various participating laboratories, and helps provide a better understanding of the likely magnitude of uncertainties in these measurements within the marine science community at the present time. Further improvements will necessarily hinge on adoption of an improved level of training in both measurement technique and of suitable quality control procedures for these measurements.

1. Introduction

Carbonate chemistry measurements of seawater have become routine in recent decades. Large-scale, regular observations of CO_2 parameters began in the 1970s with the Geochemical Ocean Sections (GEOSECS) program. However, disagreement in total alkalinity (A_T) and total dissolved inorganic carbon (C_T) was sometimes greater that 1% of the ambient values, requiring large adjustments to create complete data sets for comparison (Feely et al., 2001). Even now, it is common practice to recommend adjustments to particular data-sets so as to achieve basin-scale consistency for measurements from various oceanographic expeditions (e.g. Key et al. (2004) and Key et al. (2010)). In 1988 an intercomparison of CO_2 measurements (A_T, C_T, pH, and $p(CO_2)$) was undertaken. Seawater at four different salinities, prepared by the IAPSO Standard Seawater Service, was distributed to 14 laboratories for analysis. Although precision within each laboratory was quite high, the accuracy of the measurements was low. The results disagreed considerably, with differences in mean A_T and C_T of 20–30 μmol kg^{-1} for seawater with salinities in the range appropriate to the open ocean (Poisson et al., 1990). Another intercomparison of 14 laboratories which were using the extraction/coulometric procedure for the determination of C_T was carried out in 1990–91 and showed similar disagreement (Dickson, 1992). The desired accuracy of these measurements for the Joint Global Ocean Flux Study (JGOFS) and World Ocean Circulation Study (WOCE) programs was ~1 μmol kg^{-1} (UNESCO, 1992), far smaller than the agreement found, prompting a call for suitable reference materials to help increase measurement accuracy (Poisson et al., 1990; UNESCO, 1990).

The Dickson lab has been producing seawater-based reference materials for C_T since 1990 (Dickson, 2001), and began to certify them for A_T in 1996 (Dickson et al., 2003). In 2012, the lab began to measure the pH of these reference materials using a spectrophotometric technique (Carter et al., 2013) using purified *m*-cresol purple (Liu et al., 2011). This reference material project began originally as a response to the need to standardize CO_2 measurements made during the JGOFS program (Dickson, 2001) and has grown to a process that distributes nearly 10,000 bottles of reference material every year, sending them to approximately 250 laboratories around the world. Since the introduction of these reference materials, there has been substantial improvement in the quality of seawater CO_2 measurements. For example,

---

* Corresponding author.
E-mail address: ebockmon@gmail.com (E.E. Bockmon).

http://dx.doi.org/10.1016/j.marchem.2015.02.002

where in the past C_T measurements made on deep seawater of intersecting cruises might have disagreed by 15–20 µmol kg^{-1}, high-quality measurements now often agree within 2 µmol kg^{-1} (Dickson, 2010).

Various efforts were made in the early 1990s to improve the accuracy of carbon measurements across the community, including the creation of CO_2 in seawater reference materials (Dickson, 1992; UNESCO, 1991) and the documentation of CO_2 measurement techniques (DOE, 1994). This continues to be especially relevant now that many more researchers have become interested in the implications of the ocean's response to increasing atmospheric carbon dioxide concentrations, and the use of seawater reference materials to calibrate instruments for carbonate chemistry measurements has increased significantly. There are an increasing number of laboratory manipulations with high CO_2 treatments, as well as an emergence of interest in and monitoring of coastal upwelling zones experiencing low oxygen and low pH seawater (Feely et al., 2008). Each of these research themes is growing and each requires frequent measurement of seawater samples with high C_T and low pH. This has made it desirable to have a seawater reference material with a high CO_2 content so as to enable a reduction in carbonate chemistry measurement uncertainties (Hoppe et al., 2012). If a reference material with high CO_2 content is used in conjunction with an "ambient level" reference material to calibrate or evaluate an instrument, much more information can be learned about the performance of the instrument.

The aim of this study is to document the current quality of discrete seawater CO_2 measurements (A_T, C_T, and pH) in the various participating laboratories. It will provide a better understanding of the current likely uncertainties in these measurements, which are central not only to research into the changing marine carbon cycle, but also which underpin our understanding of ocean acidification and its implications for marine organisms. When judging the "quality" of analytical measurements it is important to clarify the scientific application that they are required for and the maximum uncertainty that is considered appropriate for that application. A recent report (Newton et al., 2014) describing plans for a Global Ocean Acidification Observing Network (GOA-ON) articulates two such applications: a "weather" goal and a "climate" goal, and notes that in all cases it is desirable to fully characterize the seawater CO_2 system by measurement.

The "weather" goal is defined as measurements of quality sufficient to identify relative spatial patterns and short-term variations, supporting mechanistic response to and impact on local, immediate ocean acidification dynamics. The GOA-ON report proposes that this objective requires the carbonate ion concentration (used to calculate $CaCO_3$ saturation states) to have a relative standard uncertainty of ≤10%. This, in turn, implies an uncertainty of ~0.02 in pH; of ~10 µmol kg^{-1} in measurements of A_T and C_T; and a relative uncertainty of ~2.5% in $p(CO_2)$.

The "climate" goal is defined as measurements of quality sufficient to assess long-term trends with a defined level of confidence, supporting detection of the long-term anthropogenically driven changes in hydrographic conditions and carbon chemistry over multi-decadal time scales. This objective is far more demanding, and requires that a change in the carbonate ion concentration be estimated at a particular site with a relative standard uncertainty of 1%. This, in turn, implies an uncertainty of ~0.003 in pH; of ~2 µmol kg^{-1} in measurements of A_T and C_T; and a relative uncertainty of about 0.5% in $p(CO_2)$.

These goals serve as a point of comparison for the quality of measurements. Similar objectives have been articulated in a draft document from the Global Ocean Observing System (GOOS, 2014) that proposes the carbonate system as an Essential Ocean Variable for biogeochemistry. Our discussion is framed in terms of these quality objectives, and tries to evaluate whether current laboratory quality control approaches are adequate to meet either of these objectives reliably.

2. Materials and methods

2.1. Preparation of the sample materials

Two batches of seawater (Batches A and B) were prepared according to the standard technique for the preparation of seawater reference materials in the Dickson Laboratory at the Scripps Institution of Oceanography (UC San Diego). For each batch a large volume of seawater was filtered, poisoned with mercury(II) chloride and bottled in 500 mL Corning Pyrex® reagent bottles sealed with greased ground-glass stoppers and leaving ~1% headspace. These bottles were cleaned by first baking in an annealing oven (at 590 °C) and then rinsing thoroughly with 18 MΩ water (UNESCO, 1991).

Batch A had a lower $p(CO_2)$ typical of a normal reference material batch; Batch B was modified to have a higher CO_2 by bubbling with CO_2 gas. Bubbling with CO_2 to increase C_T modifies the seawater in a way that mimics the expected future conditions of the ocean, since it does not change the A_T, while it does increase $p(CO_2)$ and decrease pH. The bubbling took place three days before the bottling was to occur, during normal recirculation of the reference material seawater, which allowed time for complete mixing. The $p(CO_2)$ of the seawater was monitored during this recirculation using a CONTROS HydroC^{TM} CO_2 FT sensor and temperature was monitored using a DirecTemp^{TM} surface thermistor (Model # DTU6022).

During the several hours that it took to transfer the reference material seawater from the large container into individual bottles, the CO_2 level of the headspace in the large container was controlled dynamically to match the $p(CO_2)$ in the seawater. This control ensured that the seawater did not change its CO_2 concentration during the bottling process, even with the elevated CO_2 of Batch B. The appropriate gas mixture for the headspace was calculated from the initial $p(CO_2)$ measured by the CONTROS, A_T measured on a discrete sample taken on the first day of recirculation, and the instantaneous temperature of the seawater in the jug, which can change over the course of the day. This headspace gas mixture was created dynamically by combining N_2, O_2, and CO_2 gas streams, using three mass flow controllers managed by software written in LabVIEW^{TM} (National Instruments), and was introduced into the large container at the top, at a rate slightly above the water removal rate during bottling.

2.2. Assignment of values to the sample materials

It is essential to ensure both homogeneity and stability of a batch of reference materials. The homogeneity was assured by initial thorough mixing of the seawater, together with the dynamic control of the headspace outlined above and was confirmed by subsequent analysis. The stability was confirmed by a series of analyses made over a three month period, as is usual for the certification of the Scripps CO_2 in seawater reference materials. Such reference materials have been shown to be stable for at least three years with respect to changes in A_T and C_T. The measurements performed on these sample materials were identical to those that are performed on each new batch of reference material.

C_T was assayed by the vacuum extraction/manometric procedure originally developed in Dr. C. D. Keeling's laboratory and based on the work of Wong (1970). A weighed sample is acidified with phosphoric acid; the CO_2 evolved is then extracted under vacuum and condensed in a trap cooled by liquid nitrogen. The water and CO_2 are separated from one another by sublimation and the CO_2 is transferred into an electronic constant-volume manometer. There its pressure, volume, and temperature are measured and the amount of CO_2 separated is computed from a virial equation of state.

A_T was determined by a two-stage, potentiometric, open-cell titration using coulometrically analyzed hydrochloric acid. A weighed sample of reference material is acidified to a pH between 3.5 and 4.0 with an aliquot of titrant. The solution is stirred while bubbling with air to allow the evolved CO_2 to escape. The titration is then continued to a pH of about 3.0 and the equivalence point evaluated from titration points in the pH region 3.0–3.5 using a non-linear least squares procedure that corrects for the reactions with sulfate and fluoride ions (Dickson et al., 2003).

Values for pH at 25 °C were estimated spectrophotometrically (on the total hydrogen ion concentration scale) using purified *m*-cresol purple indicator dye (supplied by Dr. DeGrandpre, U. Montana), the equations of Liu et al. (2011), and a procedure similar to that described by Carter et al. (2013).

The results for these various analyses are reported in Table 1 as means and standard deviations, and these means are used as the assigned values for this inter-laboratory comparison. The methods for C_T and A_T have been evaluated for overall uncertainty (unpublished work), and we believe that the assigned values are within 3 μmol kg^{-1} of their "true" values (95% confidence). The assigned pH is, as yet, an *operational* value (based on a particular dye and a particular calibration of that dye) and its likely uncertainty is not well known, though we believe it to be within 0.01 in pH of the "true" value (unpublished work).

2.3. Distribution and collection of results

We are very grateful to the CO_2 chemistry community for their participation in this inter-laboratory comparison. Every effort was made to include as many laboratories as wished to participate. Test samples were shipped to 66 participants: 30 in the United States and 36 in 18 other countries. We received results back from 59 participants (Table 2). All laboratories were assured anonymity and accordingly their results are presented without any designation. Several interested laboratories were not able to participate due to constraints of time, instrumentation, and/or budget.

Laboratories were responsible for their own sample handling and data reporting. Several laboratories did not perform measurements of all parameters. Thus, the total number of results reported for each parameter differs, and furthermore separate results for a parameter may in fact reflect work done in the same lab, but on different instruments. Two laboratories measured $p(CO_2)$ directly; their results are not included in this report.

Participants were asked to provide final, calibrated, measurement results for each of the parameters they determined, noting whenever more than one analysis for a particular parameter was done out of a single bottle. Additionally, laboratories were asked to indicate how the measurements were made (including equipment); how the measurement data were calibrated and an assessment of the uncertainty; and finally, if any adjustment was made to the data prior to reporting.

It is generally not a good idea to try to measure both C_T and pH from the same sample as some CO_2 transfer may occur before the second analysis is done, and the results will be sensitive to this. Of the laboratories that made multiple measurements of C_T and pH, some took multiple sub-samples out of the same bottle, while others purchased separate sets of bottles for each instrument. Any consequent errors, while possibly small, will be reflected in the results reported.

Table 1
Assigned values for total alkalinity, total dissolved inorganic carbon, and pH (25 °C; total scale) for the test samples. Values are expressed as *mean ± standard deviation* (*number of analyses*).

	Batch A	Batch B
Salinity	33.190	33.186
Total alkalinity	2215.08 ± 0.49 (24) μmol kg^{-1}	2216.26 ± 0.52 (18) μmol kg^{-1}
Total dissolved inorganic carbon	2015.72 ± 0.74 (9) μmol kg^{-1}	2141.94 ± 0.37 (6) μmol kg^{-1}
pH (25 °C; total scale)	7.8796 ± 0.0019 (18)	7.5541 ± 0.0020 (18)

Table 2
Participating organizations. Several labs from a single organization may have participated.

Country	Organization
Australia	Australian Institute of Marine Science
	CSIRO — Marine and Atmospheric Research (Hobart)
	Geoscience Australia
	Southern Cross University
	University of Tasmania
Belgium	Université de Mons
	Université Libre de Bruxelles
Canada	Institute of Ocean Sciences
China	Xiamen University
France	Station Biologique de Roscoff
Germany	Alfred Wegener Institute
	GEOMAR Helmholtz Centre for Ocean Research Kiel
	Leibniz Institute for Baltic Sea Research
Italy	Istituto di Scienze Marine — Consiglio Nazionale delle Ricerche
	Istituto Nazionale di Oceanografia e di Geofisica Sperimentale
Japan	Japan Agency for Marine-Earth Science and Technology
	Meteorological Research Institute, Japan Meteorological Agency
	Tokyo University of Marine Science and Technology
Netherlands	Royal Netherlands Institute for Sea Research, NIOZ
New Zealand	National Institute of Water and Atmospheric Research
	University of Otago
Norway	Institute of Marine Research
	University of Bergen
South Africa	Council for Scientific and Industrial Research
	Department of Environmental Affairs, Oceans and Coasts
Spain	Instituto de Investigaciones Marinas de Vigo — CSIC
UK	National Oceanography Centre
	Plymouth Marine Laboratory
USA	California State University, Northridge
	NOAA — Alaska Fisheries Science Center, Auke Bay Laboratories
	NOAA — Alaska Fisheries Science Center, Kodiak Laboratory
	NOAA — Atlantic Oceanographic & Meteorological Laboratory
	NOAA — Northeast Fisheries Science Center, Milford Laboratory
	NOAA — Northeast Fisheries Science Center, Sandy Hook Laboratory
	NOAA — Northwest Fisheries Science Center
	NOAA — Pacific Marine Environmental Laboratory
	Oregon State University
	Sunburst Sensors, LLC
	Texas A&M University
	Texas A&M University — Corpus Christi
	U.S. Geological Survey — St. Petersburg Coastal and Marine Science Center
	University of California, San Diego — Scripps Institution of Oceanography
	University of California, Santa Barbara
	University of Georgia
	University of Hawaii
	University of Miami
	University of Montana
	University of New Hampshire
	University of Washington — Friday Harbor Laboratories
	Woods Hole Oceanographic Institution

3. Results

Altogether 59 separate groups participated; 61 sets of results were returned for $A_{T,}$ 58 for C_T, and 33 for pH. For each of these parameters, the difference between the average measured value reported and the assigned value for the test samples (Table 1) has been plotted in Figs. 1–3. Whenever more than a single value was reported for a bottle, the average was taken and used to represent the value for the bottle, unless instructed otherwise by the lab returning results. Nearly all laboratories reported using reference materials provided by Prof. Dickson's laboratory to calibrate the A_T and C_T instrumentation or results, however a few used no correction.

While most laboratories reported their pH results on the total scale and at 25 °C, there were some exceptions. Seven sets of results were reported at a different temperature, often near 20 °C. For those results, CO2calc (Robbins et al., 2010) was used to adjust the results to 25 °C using the assigned A_T value and the reported pH. Also, an adjustment of −0.14 pH units was applied to the results from the two laboratories

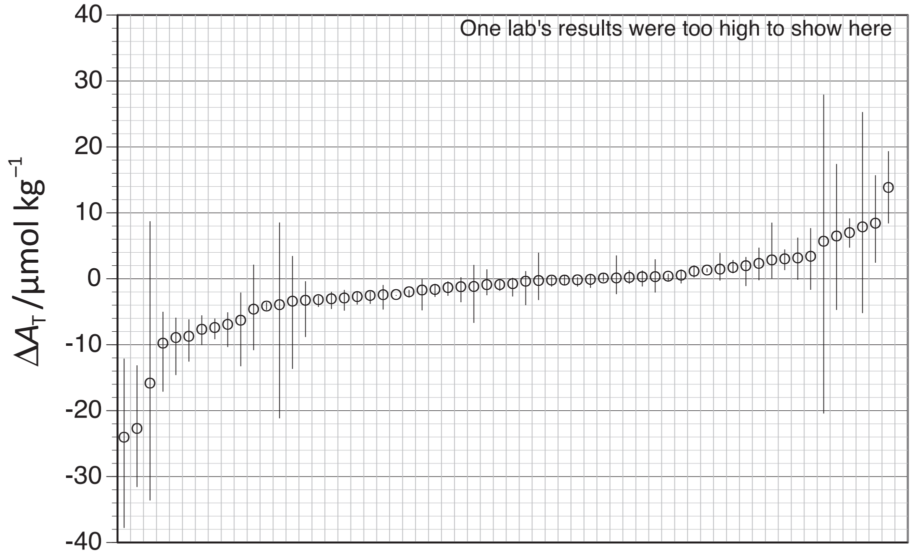


Fig. 1. Differences between the total alkalinity values reported by the participating laboratories and the assigned values for the test samples (Table 1). The open circles indicate the average difference, the lines the range of such differences. Differences for both Batch A and Batch B have been combined, as the two batches were so similar to each other.

that reported their results on the conventional (NBS) pH scale (Dickson, 1984). Although each of these adjustments to the reported pH values is subject to uncertainty, the magnitude of such uncertainty is likely small compared to the range of deviations displayed in Fig. 3.

Results from this inter-laboratory study show that many of the participating laboratories are able to make reasonable measurements for CO_2 in seawater. The majority of laboratories reported A_T and C_T values within ~10 µmol kg^{-1} of the assigned values (i.e. within ±0.5%). However, few laboratories were capable of achieving results within 2 µmol kg^{-1} of the assigned values (i.e. within ±0.1%), especially for C_T. Results for the analysis of pH were quite scattered, with little suggestion of a consensus value (especially for the lower pH sample), probably reflective of the fact that there is currently no widely available certified reference material for this parameter.

3.1. Brief overview of the methods of determination

There was some variety in the instrumentation used for the analyses. In recent years, commercially available instruments have emerged capable of measuring A_T and C_T. However, many measurements are still made using instrumentation custom developed by or for an individual laboratory, especially for A_T. Fig. 4 illustrates the distribution of reported instrumentation for A_T and C_T measurements.

Alkalinity measurements were almost exclusively done by open cell titration, including four participants who performed micro-titrations on sample sizes less than 20 mL. The most common instrument for A_T analysis was the VINDTA (Models 3S and 3C) by Marianda, followed by a significant proportion of custom or in-house designs. Other commercially available instruments used for open cell titration include systems from

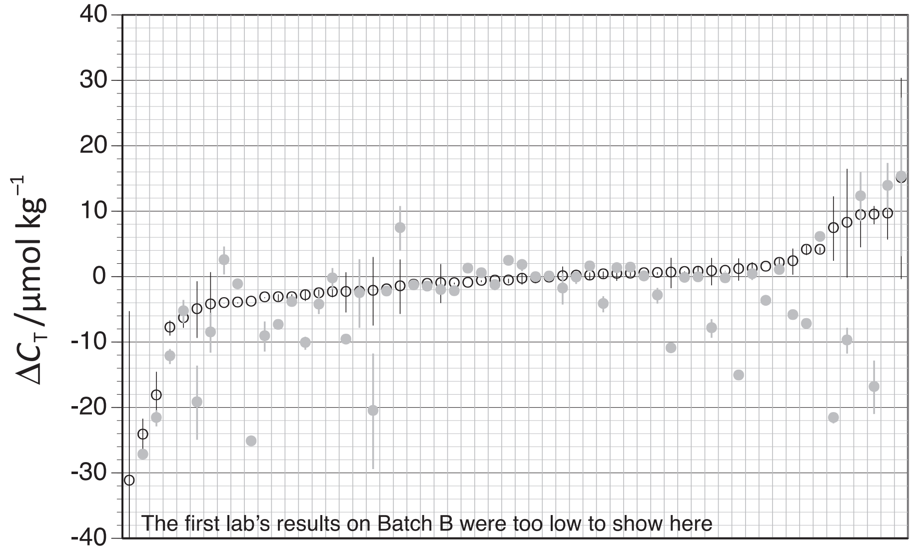


Fig. 2. Differences between the total dissolved inorganic carbon values reported by the participating laboratories and the assigned values for the test samples (Table 1). The open circles indicate the average difference for Batch A, the corresponding black lines the range of such differences; the filled gray circles indicate the average difference for Batch B, the corresponding gray lines the range.

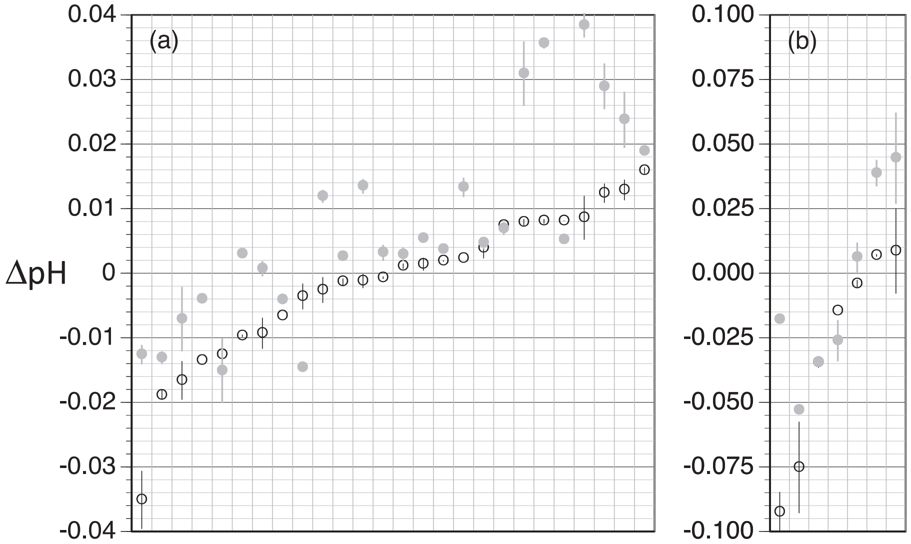


Fig. 3. Differences between the pH values (25 °C; total scale) reported by the participating laboratories and the assigned values for the test samples (Table 1). The open circles indicate the average difference for Batch A, the corresponding black lines the range of such differences; the filled gray circles indicate the average difference for Batch B, the corresponding gray lines the range. The laboratories grouped in panel (a) used spectrophotometric techniques to measure pH; those grouped in panel (b) used electrometric techniques.

Apollo SciTech, Metrohm, Mettler Toledo, and Kimoto. A couple of labs used spectrophotometric methods for measuring A_T including in one case a system built by Nippon ANS Inc., and one lab did a closed cell titration.

The coulometric method for the determination of seawater C_T was adapted by Johnson et al. (1985) from a standard method for CO_2 determination and then automated and calibrated (Johnson et al., 1987). The resulting instrumentation developed at the University of Rhode Island has become known as the SOMMA system and has been used extensively in combination with a UIC Coulometer over the past few decades to monitor C_T in the oceans. Various other instruments are now commercially available based on this method of acidification and coulometric determination including the VINDTA (Models 3D and 3C) by Marianda and an instrument by Nippon ANS Co. In addition, several labs have devised their own custom systems based on a similar coulometric approach. Other C_T systems use a similar acidification technique, but quantify the resulting CO_2 gas using a nondispersive infrared sensor (NDIR — commonly the LiCOR 7000) rather than coulometry. The Apollo SciTech DIC analyzer (AS-C3) was the most common instrument of this type used, followed by custom systems and the AIRICA by Marianda. These NDIR based instruments use a much smaller sample size than the original SOMMA and VINDTA, which can be an advantage if water availability is limited, with some potential trade off in terms of precision and accuracy. Two laboratories used a SEAL Analytical QuAAtro Segmented Flow Analyzer to colorimetrically determine C_T. Several other unique methods of analysis were reported and are classified with those laboratories that failed to report their method in "Other" (Fig. 4).

Most pH measurements were performed spectrophotometrically, although six laboratories used electrodes and one measured pH with an optode (Fig. 3b). The spectrophotometric measurements are grouped in Fig. 3a, and are generally more accurate than those done with a glass electrode (note the different y-axis scales). Although commercially available instruments that perform the entire pH measurement do not exist, there was some consistency in the spectrophotometer and dye used. Spectrophotometers made by Agilent were the most common, followed by spectrophotometers purchased from Cary, Ocean Optics, and Shimadzu. Most laboratories used *meta*-cresol purple indicator dye, but three reported using thymol blue. In the last couple of years, attention has been brought to the impurities present in commercially available *m*-cresol purple (Yao et al., 2007). Methods for purification have been published (Liu et al., 2011; Patsavas et al., 2013) and equations for calculating pH from absorbance values obtained using such purified dye are available (Liu et al., 2011). While a number of labs had access to purified dye for this study, most used commercially available *m*-cresol purple indicator dye. A cell path length of 10 cm was by far the most commonly used, although 2 and 3 cm cells were also reported.

4. Discussion

In our discussion below, we make some necessarily subjective assessments of the overall data quality, and speculate as to the reasons that underlie it. We will use the "weather" and "climate" goals as described for GOA-ON (Newton et al., 2014) as a touchstone of data quality. In most cases, there is no clear relationship between measurement quality and the choice of analytical instrumentation. Some laboratories using a given, commercially available system, obtained excellent agreement with the assigned values, while others using apparently identical equipment did not. This suggests that a significant contribution to the overall uncertainty of measurement can be ascribed to the operator and to the exact laboratory procedure, including quality control practices, used for the analysis.

4.1. Total alkalinity

The A_T results seem perhaps the most encouraging (Fig. 1). However, the two batches of seawater used for this inter-laboratory comparison had nearly identical A_T values (2215.08 and 2216.26 μmol kg^{-1}). This is also very similar to the A_T of the seawater reference materials typically

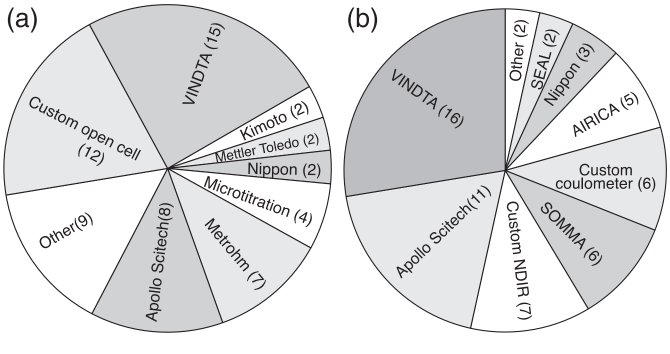


Fig. 4. Types of analytical systems used for (a) total alkalinity determinations; and (b) total dissolved inorganic carbon determinations.

provided by our laboratory and which many of the participants used to calibrate their measurements. Consequently, this study is a less than ideal investigation into a laboratory's ability to make A_T measurements and leaves open the question of whether laboratories that showed themselves capable of getting the correct value here would necessarily be capable of getting the correct answer at significantly lower or higher A_T. Despite this, A_T results were not uniformly perfect. A number of laboratories failed on one or more of their analyses to reproduce the assigned value to within 10 μmol kg^{-1}, the weather objective for data quality. Only 20% met the stricter climate goal of 2 μmol kg^{-1} for all of their samples. In addition, about a quarter of the laboratories had a range in their reported values larger than 10 μmol kg^{-1}, and more than 10% had a range greater than 20 μmol kg^{-1}, suggesting that their measurement technique was not operating reproducibly.

A careful examination of Fig. 1 shows that the majority of participants reported alkalinities that were lower than the assigned values for the samples. It is difficult to be sure of the reason for this, but it seems in part to be associated with the common use of simpler end-point determinations (as opposed to equivalence point determinations based on the use of full equilibrium expressions for seawater acid–base chemistry). In addition, systems that used very small sample sizes (less than 20 mL) had considerable uncertainties associated with their results.

4.2. Total dissolved inorganic carbon

The results reported for C_T (Fig. 2) for Batch A are generally encouraging, though again the test sample is similar in composition to that of the seawater reference materials we distribute. Of the 58 results returned for the C_T of Batch A, only a small number of them were further than the “weather” objective of 10 μmol kg^{-1} from the assigned value. However, for Batch B (with the higher CO_2 level) a significant number of additional laboratories did not meet this standard. Additionally, only 25% of the laboratories achieved the climate objective of 2 μmol kg^{-1} for both Batches A and B. Furthermore, in most cases, the reported values for Batch B tended to be lower than the assigned value whereas for Batch A they were more evenly distributed, both lower and higher. The mean difference of reported values from the assigned values is −1.17 μmol kg^{-1} for Batch A and −4.95 μmol kg^{-1} for Batch B. The exact reason for this discrepancy is not known and may reflect an unidentified calibration problem for some instruments. However, the consistent sign of the discrepancy for Batch B, even with different instruments, is suggestive of a loss of CO_2 at some unidentified point in the analytical process. Such a loss might be expected to be concentration dependent to some extent, and it may well be that some loss also occurred for measurements on Batch A, but that the use of a similar reference material for instrument calibration adjusted for it. Also, the loss could be worse for yet higher CO_2 levels, such as might be found in water samples from certain environments and from intentionally modified seawater intended for ocean acidification studies.

It is not straightforward to use this exercise to distinguish between problems inherent to instrument design, and problems that are due to the use of suboptimal analytical procedures or to operator inexperience. For example, measurements made using SOMMA systems did not exhibit such discrepancies. However, SOMMA instruments have been used extensively, starting with the WOCE/JGOFS Hydrographic Survey in the 1990s when standard operating procedures for use (and quality control) of these instruments were developed as a group effort (DOE, 1994). Furthermore, the laboratories with SOMMA systems have typically owned their instruments for many years, and their operators are very experienced in making high-quality C_T measurements. Other instruments (when viewed as a group) usually lack one or more of these desirable characteristics for optimal operation.

4.3. pH

Results for pH were quite variable (Fig. 3), and without a clear consensus value. Of the 34 sets of results returned, 27 were measured using a pH indicator dye with a spectrophotometer (Fig. 3a); the others were measured with pH probes (Fig. 3b): 6 using glass electrode pH cells, and 1 using an optical pH sensor from PreSens®. While all measurements performed on a spectrophotometer using an indicator dye were within 0.04 units of the assigned value (Fig. 3a), results using a pH probe were as far away as 0.10 pH units. The reasons for the larger range of discrepancies found with pH probes probably reflect a combination of calibration difficulties (including inadequate temperature control) together with sample handling problems. Our remaining discussion will focus on the spectrophotometric pH measurements.

Most laboratories used *m*-cresol purple as the indicator dye for the spectrophotometric measurements. Seven laboratories reported having access to purified *m*-cresol purple indicator dye, and the subsequent results were typically (though not invariably) within 0.004 pH units of the assigned value. Even using impure dyes, the work of Yao et al. (2007) and of Liu et al. (2011) suggest that discrepancies due to the impurities alone are unlikely to be larger than about 0.015 pH units and often smaller. One approach to correct for this (for a number of commercially available dyes) uses measured discrepancies (reported in Fig. 2a of Liu et al., 2011) as an adjustment to the pH provided by the Liu et al. calibration equation (see Carter et al., 2013). However, the reliability of this approach has not been verified.

Fig. 3 also clearly shows that most of the participants had a more positive deviation from the assigned value for Batch B than they did for Batch A, consistent with the suggestion previously made for the C_T results that CO_2 is being lost from the high-CO_2 sample prior to measurement. Again, this may well be due to handling difficulties that could also affect the uncertainty of the results for the lower-CO_2 sample. However, a recent publication from our laboratory (Bockmon and Dickson, 2014) shows that it is possible to filter this high-CO_2 sample without apparently increasing the pH significantly, suggesting that careful sample handling can avoid significant impact. Also, some laboratories, despite showing a substantial discrepancy, had very repeatable results, so we wonder if there is an additional as yet unidentified source of bias in some of the measurements here.

One factor that probably contributes to the lack of reproducibility is that the conventional spectrophotometric method described in Dickson et al. (2007) and based on the original work of Clayton & Byrne (1993) is not automated in any way, and requires some customization to achieve adequate temperature control. It is easy to imagine that each laboratory may be implementing a slightly different procedure without any awareness of possible biases. This is further exacerbated by the lack of a reliable supply of suitable reference material for pH measurements. Given the typically good repeatability of many labs on these test samples, it may be that our widely distributed CO_2 in seawater reference materials could be used by laboratories to confirm that their pH repeatability can be extended to longer time-frames.

Nevertheless, although it is clear that more work needs to be done before we can have high confidence in pH values, many of the laboratories using a spectrophotometric procedure were able to achieve results within 0.02 pH units of the assigned value, i.e. achieving the “weather” objective specified by GOA-ON (see above).

5. Summary and conclusions

The ability of laboratories to perform carbonate chemistry measurements has clearly advanced significantly since the last such study (UNESCO, 1990). This is likely due to three factors: the widespread availability and use of seawater reference materials for A_T and C_T calibrated by our laboratory; the existence of a published Guide describing suitable analytical methods (Dickson et al., 2007); and growing commercial availability of instrumentation for making such measurements.

Table 3
Six principles of Valid Analytical Measurement developed by the Laboratory of the Government Chemist and the National Physical Laboratory in the UK as part of the Valid Analytical Measurement (VAM) programme set up by the UK Department of Trade and Industry.

Valid Analytical Measurement (VAM) principles
1. Analytical measurements should be made to satisfy an agreed requirement.
2. Analytical measurements should be made using methods and equipment which have been tested to ensure they are fit for purpose.
3. Staff making analytical measurements should be both qualified and competent to undertake the task.
4. There should be a regular independent assessment of the technical performance of a laboratory.
5. Analytical measurements made in one location should be consistent with those elsewhere.
6. Organisations making analytical measurements should have well defined quality control and quality assurance procedures.

In this exercise, more than half of the participating laboratories achieved A_T values that agreed with the assigned values to within ± 5 μmol kg^{-1} for every one of the analyses they reported, with more than two-thirds agreeing within ± 10 μmol kg^{-1}. For C_T, nearly three-quarters of the participating laboratories reported all their values within ± 5 μmol kg^{-1} of the assigned value for analyses on Batch A, but only half were within ± 5 μmol kg^{-1} for all their analyses on the high CO_2 Batch B. For pH, more than half the labs were within 0.01 pH units of the assigned value for all their analyses on Batch A, but only about a third were within 0.01 pH units for their analyses on Batch B.

Almost all participating laboratories used reference materials traceable to the Dickson laboratory for the calibration of their A_T and C_T values. Thus, participants tended to do better on samples similar to the typical reference materials (Batch A & Batch B for A_T; Batch A for C_T). However, the repeatability on A_T measurements was noticeably worse than for C_T measurements (median difference for measurements of $A_T \sim 4$ μmol kg^{-1}; for measurements of $C_T \sim 1.5$ μmol kg^{-1}), thus a smaller proportion of participating laboratories consistently achieved ± 5 μmol kg^{-1} for A_T (Batch A and Batch B) than did for C_T (Batch A) despite such calibration. Although there are no widely distributed reference materials for pH measurement, the repeatability was typically very good (median ~0.002 in pH) (The Dickson lab has started to provide an information value for the pH of their seawater reference materials; obtained using the procedure described here.).

As noted above, for many laboratories there is a clear suggestion that CO_2 has been lost in the analysis of Batch B (when compared to the results for the same laboratory for Batch A). This is apparent in the results for C_T, where twice as many labs underestimate the value for B (relative to A) as overestimate it; for pH the proportion that overestimates B (relative to A) is about 5:1. Of course, in a number of these cases, the difference is within the likely analytical reproducibility, and in others, it may reflect a calibration problem. Nevertheless, these results — taken at face value — lead us to believe that C_T or pH measurements on samples with high CO_2 content are typically more uncertain than those with levels closer to current atmospheric values.

Clearly there is a need for more than one reference material for seawater CO_2 properties so that laboratories can better confirm the quality of their measurement procedures, and thus minimize apparent sample handling problems. In the future, our laboratory will try to supply these materials. However, we do note that there is presently too high a reliance on CO_2 in seawater reference materials not only for quality control, but also for calibration purposes. Whenever a reference material is used for calibration, it inherently limits its usefulness as a quality control check. Unless the laboratory is independently confident that the instrument is operating correctly, and with a stable reproducibility (verified, perhaps, over time using a laboratory's own stable test solution), the use of a certified reference material for regular calibration can be misleading.

We also suggest that there has been, as yet, insufficient appreciation of the likely overall uncertainty of these measurements. Although many of the laboratories participating in this study showed themselves capable of achieving the "weather" objective of GOA-ON, few were capable of achieving the "climate" objective. Almost certainly, the seawater CO_2 measuring community would benefit from a more nuanced discussion of what constitutes an acceptable measurement uncertainty for achieving particular scientific goals. We also suspect that for most laboratories, their reported overall uncertainty is likely underestimated, assuming (erroneously) that precision is an adequate estimate of uncertainty — see e.g. the discussion in De Bièvre (2008).

So, how should the CO_2 measurement community work to improve matters? In 1989, the UK Department of Trade and Industry set up its "Valid Analytical Measurement" program. It articulated six principles (Sargent, 1995) that should be emphasized by the marine CO_2 measuring community (Table 3). We plan to do our part.

Acknowledgments

The authors acknowledge financial support from the National Science Foundation through grant OCE-1233648, and from the NOAA Ocean Acidification Program through grant NA10OAR4320156. Emily E. Bockmon was supported by the University of California through a Multi-Campus Research Program supporting research and training in ocean acidification.

Assistance in preparation and analysis of the test materials was provided by Guy Emanuele, Laura Fantozzi, Kristin Jackson, and other staff in the Dickson laboratory. A special thanks is due to each of the principal investigators, students, technicians, and laboratories that participated in this inter-laboratory comparison and to their various sponsors that made it possible.

References

Bockmon, E.E., Dickson, A.G., 2014. A seawater filtration method suitable for total dissolved inorganic carbon and pH analyses. Limnol. Oceanogr. Methods 12, 191–195.

Carter, B.R., Radich, J.A., Doyle, H.L., Dickson, A.G., 2013. An automated system for spectrophotometric seawater pH measurements. Limnol. Oceanogr. Methods 11, 16–27.

Clayton, T.D., Byrne, R.H., 1993. Spectrophotometric seawater pH measurements: total hydrogen-ion concentration scale calibration of *m*-cresol purple and at-sea results. Deep-Sea Res. I 40, 2115–2129.

De Bièvre, P., 2008. Measurement uncertainty is not synonym of measurement repeatability or measurement reproducibility. Accred. Qual. Assur. 13 (2), 61–62.

Dickson, A.G., 1984. pH scales and proton-transfer reactions in saline media such as sea-water. Geochim. Cosmochim. Acta 48 (11), 2299–2308.

Dickson, A.G., 1992. The Determination of Total Dissolved Inorganic Carbon in Sea Water Using Extraction/Coulometry: The First Stage of a Collaborative Study. US Department of Energy, p. 43.

Dickson, A.G., 2001. Reference materials for oceanic CO_2 measurements. Oceanography 14 (4), 21–22.

Dickson, A.G., 2010. Standards for ocean measurements. Oceanography 23 (3), 34–47.

Dickson, A.G., Afghan, J.D., Anderson, G.C., 2003. Reference materials for oceanic CO_2 analysis: a method for the certification of total alkalinity. Mar. Chem. 80 (2–3), 185–197.

Dickson, A.G., Sabine, C.L., Christian, J.R., 2007. Guide to Best Practices for Ocean CO2 Measurements. PICES Special, Publication 3, p. 191.

DOE, 1994. Handbook of Methods for the Determination of the Parameters of the Oceanic Carbon Dioxide System. In: Dickson, A.G., Goyet, C. (Eds.), ORNL/CDIAC-74.

Feely, R.A., Sabine, C.L., Hernandez-Ayon, J.M., Ianson, D., Hales, B., 2008. Evidence for upwelling of corrosive "acidified" water onto the continental shelf. Science 320 (5882), 1490–1492.

Feely, R.A., Sabine, C.L., Takahashi, T., Wanninkhof, R., 2001. Uptake and storage of carbon dioxide in the ocean: the global CO_2 survey. Oceanography 14 (4), 18–32.

Global Ocean Observing System, 2014. Nomination of an essential ocean variable (EOV): carbonate system. International Ocean Carbon Coordination Project. http://www.ioccp.org/images/10FOO/EOV_BGC_03_Carbonate%20system_draft.pdf (Accessed October 21, 2014).

Hoppe, C.J.M., Langer, G., Rokitta, S.D., Wolf-Gladrow, D.A., Rost, B., 2012. Implications of observed inconsistencies in carbonate chemistry measurements for ocean acidification studies. Biogeosciences 9 (7), 2401–2405.

Johnson, K.M., King, A.E., Sieburth, J.M., 1985. Coulometric TCO_2 analyses for marine studies; an introduction. Mar. Chem. 16 (1), 61–82.

Johnson, K.M., Sieburth, J.M., Williams, P.J.L., Brandstrom, L., 1987. Coulometric total carbon dioxide analysis for marine studies: automation and calibration. Mar. Chem. 21 (2), 117–133.

Key, R.M., et al., 2004. A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). Glob. Biogeochem. Cycles 18, GB4031.

Key, R.M., et al., 2010. The CARINA data synthesis project: introduction and overview. Earth Syst. Sci. Data 2, 105–121.

Liu, X., Patsavas, M.C., Byrne, R.H., 2011. Purification and characterization of meta-cresol purple for spectrophotometric seawater pH measurements. Environ. Sci. Technol. 45, 4862–4868.

Newton, J.A., Feely, R.A., Jewett, E.B., Williamson, P., Mathis, J., 2014. Global ocean acidification observing network: requirements and governance plan. http://www.goa-on.org/docs/GOA-ON_Plan_final_Sept_2014.pdf.

Patsavas, M.C., Byrne, R.H., Liu, X.W., 2013. Purification of meta-cresol purple and cresol red by flash chromatography: procedures for ensuring accurate spectrophotometric seawater pH measurements. Mar. Chem. 150, 19–24.

Poisson, A., Culkin, F., Ridout, P., 1990. Intercomparison of CO_2 measurements. Deep Sea Res. Part A 37 (10), 1647–1650.

Robbins, L.L., Hansen, M.E., Kleypas, J.A., Meylan, S.C., 2010. CO2calc: a user-friendly seawater carbon calculator for Windows, Mac OS X, and iOS(iPhone). U.S. Geological Survey Open-File Report 2010–1280, p. 17.

Sargent, M., 1995. The VAM Principles, Analytical Proceedings Including Analytical Communications. pp. 201–202.

UNESCO, 1990. Intercomparison of total alkalinity and total inorganic carbon determinations in seawater. UNESCO Tech. Pap. Mar. Sci. 59.

UNESCO, 1991. Reference materials for oceanic carbon dioxide measurements. UNESCO Tech. Pap. Mar. Sci. 60.

UNESCO, 1992. Methodology for oceanic CO_2 measurements. UNESCO Tech. Pap. Mar. Sci. 65.

Wong, C.S., 1970. Quantitative analysis of total carbon dioxide in sea water: a new extraction method. Deep-Sea Res. 17, 9–17.

Yao, W.S., Liu, X.W., Byrne, R.H., 2007. Impurities in indicators used for spectrophotometric seawater pH measurements: assessment and remedies. Mar. Chem. 107 (2), 167–172.