Intercomparison of carbonate chemistry measurements on a cruise in northwestern European shelf seas

M. Ribas-Ribas[1], V. M. C. Rérolle[1,2], D. C. E. Bakker[3], V. Kitidis[4], G. A. Lee[3], I. Brown[4], E. P. Achterberg[1,5] and T. Tyrrell[1]

[1] {Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, United Kingdom}

[2] {National Oceanography Centre Southampton, Southampton, United Kingdom}

[3] {Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, United Kingdom}

[4] {Plymouth Marine Laboratory, Prospect Place, Plymouth, United Kingdom}

[5] {GEOMAR, Helmholtz Centre for Ocean Research, Kiel, Germany}

Correspondence to: M. Ribas-Ribas (M.Ribas-Ribas@soton.ac.uk)

Abstract

Four carbonate system variables were measured in surface waters during an ocean acidification cruise traversing northwestern European shelf seas in the summer of 2011. High resolution surface water data were collected for partial pressure of carbon dioxide ($p\mathrm{CO}_2$; using two independent instruments) and pH using the total pH scale (pH_T), in addition to discrete measurements of total alkalinity and dissolved inorganic carbon. We thus overdetermined the carbonate system (four measured variables, two degrees of freedom) which allowed us to evaluate the level of agreement between the variables on a cruise whose main aim was not intercomparison and thus where conditions were more representative of normal working conditions. Calculations of carbonate system variables from other measurements generally compared well with direct observations of the same variables (Pearson's correlation coefficient always ≥ 0.94; mean residuals were similar to the respective accuracies of the measurements). We therefore conclude that four of the independent datasets of carbonate chemistry variables were of high quality. A diurnal cycle with maximum amplitude of 41 µatm was observed in the difference between the $p\mathrm{CO}_2$ values obtained by

the two independent analytical pCO_2 systems, and this was partly attributed to irregular seawater flows to the equilibrator and partly to biological activity inside the seawater supply and one of the equilibrators. We discuss how these issues can be addressed to improve carbonate chemistry data quality on future research cruises.

1 Introduction

Accurate determination of the inorganic carbon system is a key requirement for ocean acidification studies, as it forms the basis for assessments of biological and biogeochemical responses to changes in ocean carbonate chemistry as a result of rising atmospheric CO_2 concentrations. It is also essential for the determination of the air-sea fluxes of CO_2, calculation of carbon budgets and estimation of anthropogenic CO_2 concentrations in different water masses. When the carbonate system is overdetermined, it is possible to test if the different variables are consistent with one another. This requires that more than two of the measurable variables (total dissolved inorganic carbon (C_T), total alkalinity (A_T), pH_T, and partial pressure or fugacity of CO_2 (pCO_2, fCO_2)) are determined.

Several at-sea intercomparison studies have taken place in recent years that compared different pCO_2 instruments. Körtzinger et al. (1996) carried out what may have been the first intercomparison study in coastal waters between two similarly designed underway pCO_2 systems. They found a remarkable agreement between the two simultaneously measured pCO_2 datasets even though the spatial variability in surface pCO_2 in the North Sea was high. The average difference was 0.2 µatm (standard deviation = 1.2 µatm), indicating no systematic difference. The difference tended to be highest during the most pronounced pCO_2 gradients. Körtzinger et al. (2000) reported on a comprehensive shipboard, international intercomparison exercise which used one discrete and seven underway systems for the measurement of fCO_2. This exercise showed that underway fCO_2 can be determined to a high level of precision (± 2 µatm) with a variety of equilibrator and system designs.

Other workers have undertaken at-sea intercomparisons of different variables. For instance, Johnson et al. (1999) compared C_T, fCO_2 and A_T measurements during the same intercomparison exercise as reported for fCO_2 by Körtzinger et al. (2000). These scientists found a systematic fCO_2 overestimation of 9 µatm when calculated from C_T and A_T measurements relative to observed fCO_2. Lamb et al. (2001) investigated 25 cruises in the Pacific Ocean where at least two of the four inorganic carbon variables were determined.

They examined the consistency of the dataset using Certified Reference Material (CRM) analyses, precision of at-sea replicate analyses, agreement between shipboard analyses and replicate shore based analyses, comparison of deep water values at locations where two or more cruises overlapped or crossed, consistency with other hydrographic parameters and internal consistency between multiple carbon variables measurements. Using all this evidence the carbonate data was adjusted for inconsistencies and a combined dataset was constructed, which showed that C_T and A_T had an estimated overall accuracy of 3 µmol kg^{-1} and 5 µmol kg^{-1}, respectively.

Other studies pointed out some inconsistences: Millero et al. (2002) noted that the use of pH_T and C_T from field measurements from the Atlantic, Indian, Southern and Pacific oceans yielded standard errors (1σ) of ± 22.3 µatm in calculated pCO_2 and ± 4.3 µmol kg^{-1} in calculated A_T. Lueker et al. (2000) noted that observed values of pCO_2 above 500 µatm were by, on average, 3.35 % (if fCO_2 was 500 µatm that will be 17 µatm) higher than pCO_2 calculated from C_T and A_T. This tendency towards a larger differences between measured pCO_2 and calculated pCO_2 at higher pCO_2 levels was also observed by McElligott et al. (1998), suggesting that it might result from inaccuracy in the formulation of the solubility coefficient of CO_2 in seawater (K_0). However, this apparent discrepancy has not yet been explained satisfactorily (Dickson, 2010). It is possible that an unidentified acid-base system affects the calculation of pCO_2 or that one or more dissociation constants for acid-base equilibria are not well parameterised at high pCO_2 (Dickson, 2010).

The aims of our study were to evaluate the quality of our observations of inorganic carbon variables and to investigate differences between observed and calculated variables in order to identify means of improving data quality. Our study differs from some previous work in two respects: firstly, our study was undertaken in surface waters of shelf seas where spatial variability is high; and, secondly, the study was not designed at the outset as an intercomparison exercise, which normally involves placing all the instruments in one laboratory, sampling from a single seawater supply and an intense focus on every aspect of the carbonate chemistry measurements. Instead, the instruments were in three separate laboratories, with samples taken from four different seawater outlets and the operators conducting multiple tasks as part of the multidisciplinary research activities undertaken on the cruise. Therefore our findings are more representative of a typical multidisciplinary research cruise.

2 Material and methods

The data used in this study were collected in the period 06 June to 07 July 2011 during the RRS *Discovery* research cruise D366 in northwestern European shelf seas. The cruise formed part of the UK Ocean Acidification Research Programme. Two variables of the carbonate system (pCO$_2$ and pH$_T$), plus salinity and sea surface temperature, were measured at a high temporal resolution (every 5 min for one pCO$_2$ system (pCO$_2$-1), every 6 min for pH$_T$, and every 1 min for a second pCO$_2$ system (pCO$_2$-2)). These instruments received a continuous flow of water from the ship's underway continuous seawater supply (intake positioned at ca. 5 m depth). In addition, nutrients (nitrate plus nitrite, phosphate and silicate), C_T and A_T were sampled every 2 h from the underway supply, and also collected in surface waters sampled by CTD casts (samples obtained from the sampling bottle closest to 5 m depth; typically between 2.0 and 8.2 m). Continuous temperature and conductivity data were obtained from a Sea-Bird Electronics SBE45 thermosalinograph (TSG) installed on the ship's underway supply. Discrete surface water samples for salinity (S) were collected every 4 h in order to calibrate the conductivity measurements. Discrete salinity samples were analysed using a salinometer (Guildline Autosal 8400B). Photosynthetically Active Irradiance (PAR), radiation between 400 and 700 nm was measured as part of the ship's meteorological parameters with a 2-pi sensor (Skye Instruments, model SKE 510) positioned at 10 m height.

2.1 Carbonate chemistry analysis

2.1.1 Partial pressure of CO$_2$

Quasi-continuous measurements of pCO$_2$ in surface water and marine air were undertaken using two different instruments (hereafter pCO$_2$-1 and pCO$_2$-2). The pCO$_2$-1 and pCO$_2$-2 systems undertook 6,187 and 26,671 measurements of surface water pCO$_2$ during the cruise, respectively.

System 1: System pCO$_2$-1 was an underway pCO$_2$ instrument (PML-Dartcom *Live* pCO$_2$) as described in detail by Hardman-Mountford et al. (2008), with the modified 'vented' equilibrator introduced by Kitidis et al. (2012). The instrument was located in a mid-ship chemistry laboratory. The system used a vented-showerhead equilibrator, with ambient light blocked out, to equilibrate seawater CO_2 with a headspace. In order to maintain atmospheric pressure in the equilibrator headspace, the unit was vented to a second equilibrator, which in turn was vented to the atmosphere via a 2 m coil of stainless steel tubing (1.5 mm internal diameter). The equilibrator was fitted with 2 platinum resistance thermometers (Pico Technology, model PT100) and a water-jacket supplied with seawater from the ship's underway seawater system. A seawater flow of 1.6 L min^{-1} was maintained across the equilibrator. The average warming between the ship's underway seawater intake and the equilibrator was 0.5 °C (standard deviation = 0.1 °C). Atmospheric measurements of CO_2 were taken from an intake located forward on the deck above the ship's bridge. Both gas streams from the equilibrator headspace and the air inlet were dried in a Peltier cooler (-20 °C). Mixing ratios of CO_2 and water in the marine air and equilibrator headspace were determined by infrared detection (LI-840, LI-COR). Measurements were referenced against secondary calibration gases from BOC Gases (UK) with known CO_2 mixing ratios (0, 251.3 and 446.9 µmol CO_2 mol^{-1}) in synthetic air mixtures (21 % oxygen and 79 % nitrogen). All calibration gases underwent pre- and post-cruise calibration against certified primary standards from the National Oceanic and Atmospheric Administration (NOAA), which had values of 244.9 and 444.4 µmol CO_2 mol^{-1}.

System 2: System pCO_2-2 was an underway pCO_2 system located in a container laboratory positioned on the aft-deck of the ship. The instrument setup and calibration procedures are as described by Bakker et al. (2007), with the exception of the vented equilibrator. The percolating packed bed-type equilibrator was identical to the one described by Schuster and Watson (2007). The equilibrator of transparent perspex was positioned next to the window of the container without blinds. Atmospheric samples were taken from an air inlet located forward on the deck above the ship's bridge. Samples from the equilibrator headspace and marine air were partially dried by being passed through an electric cool box at about 2 °C, prior to analysis. Mixing ratios of CO_2 and water in the marine air and equilibrator headspace were determined by infrared detection with a LI-COR LI7000. The LI-COR was calibrated using secondary gas standards BOC Gases (UK) with CO_2 mixing ratios of 2.4, 260.9, 364.2 and 473.1 µmol CO_2 mol^{-1} in an artificial air mixture (21 % oxygen, 79 % nitrogen). All calibration gases underwent pre- and post-cruise calibration against certified primary standards from the NOAA, which had values of 251.6, 347.2 and 448.8 µmol CO_2 mol^{-1}. The seawater flow to the container laboratory was highly variable throughout the cruise. This was due to the location of the container downstream of an intermittently large water demand for an experiment. The water flow was regulated to a maximum of 1.8 L min^{-1}, to avoid flooding of the equilibrator and CO_2 analyser during sudden spikes in supply. The water flow tended to gradually decrease to very low flow over 6 to 12 h. Two platinum resistance thermometers positioned in the upper and lower part of the seawater stream determined the temperature of the seawater in the equilibrator (a PT probe (Omega) with modified electronics). Average warming of the seawater between the intake and the equilibrator was estimated as 0.5 °C (standard deviation = 0.4 °C). The large temperature deviations reflected the irregular seawater flow to the equilibrator. In addition, the post-cruise temperature calibration of the PT100 sensors showed excessive drift of 4-5 °C relative to the pre-cruise calibration. The absolute calibration of the temperature sensors was therefore deemed unreliable. Equilibrator temperatures from 13 June (09:57) to 17 June (17:12) have been reduced by 0.7 °C, while equilibrator temperatures after 19 June (21:57) have been increased by 0.7 °C to remove the negative temperature changes.

The precision of both LI-COR's $p\mathrm{CO}_2$ measurements was 1 µatm, established using standard gases. We estimated different accuracies for the two systems: 4 µatm for the system $p\mathrm{CO}_2$-1 and 10 µatm for system $p\mathrm{CO}_2$-2. The $p\mathrm{CO}_2$ was computed from the CO_2 mixing ratios and the ship's barometric pressure corrected from 18 m height to sea level, and corrected for seawater vapour pressure (Weiss and Price, 1980). Sea surface $p\mathrm{CO}_2$ data were corrected to sea surface temperature to account for the warming between the seawater intake and the equilibrators (Takahashi et al., 1993). The accuracies of the temperature measurements inside the equilibrators were estimated to be 0.02 °C and 5 °C for $p\mathrm{CO}_2$-1 and $p\mathrm{CO}_2$-2 respectively.
$p\mathrm{CO}_2$-1 measurements were backdated by 1 min and $p\mathrm{CO}_2$-2 measurements by 3 min to account for the travel time of the seawater between the seawater intake and the respective equilibrators. The time offsets (1 min and 3 min) between seawater intake and equilibrators were chosen objectively as those producing the minimum standard deviation between paired equilibrator and intake temperatures. The intercomparison exercise was carried out on the datasets after they had been adjusted according to the procedures just described, including corrections to in-situ seawater temperature described above.

2.1.2 pH_T

Surface water pH_T was measured continuously with an automated instrument located in the mid-ship chemistry laboratory and connected to the ship's underway seawater supply. The pH_T system undertook 29,950 measurements. The surface distribution and the processes that controls pH_T in this cruise have been described by Rérolle et al. (2014). The measurement technique, described by Rérolle et al. (2013), was based on a colorimetric method using Thymol Blue as pH indicator (Clayton and Byrne, 1993; Rérolle et al., 2012). pH_T was determined on the total pH scale. Measurements were made every 6 min with a precision of 1 mpH (Rérolle et al., 2013). Three bottles of Tris pH buffer provided by Dr Andrew Dickson (Scripps Institution of Oceanography, Marine Physical Laboratory, University of California San Diego, USA) were analysed at the beginning, middle and end of the cruise to check the accuracy of the pH_T measurements, which was 4 mpH. The Thymol Blue extinction coefficients were determined in the laboratory following the cruise, applying the salinity and temperature ranges observed during the cruise, with the indicator´s dissociation constant taken from Zhang and Byrne (1996). Measurements at sea were made at the seawater temperature plus 0.2 °C due to warming between the seawater intake and the pH instrument. In order to minimise absorbance interference by particulates an in-line filter (0.45 µm pore size, Millex HP syringe filter MilliporeExpress® (PES) membrane 33 mm diameter, Millipore) was placed at the entry of the sample tube. Chromophoric Dissolved Organic Matter (CDOM) only absorbs weakly in the visible where Thymol Blue absorbance is measured (<3 % at 435 nm and <1% at 596 nm) and is accounted for in the seawater blank. CDOM absorbance interference is thereby cancelled out. Additionally, measurements at the wavelength 750 nm (not affected by Thymol Blue indicator) were used to monitor for sample turbidity and instrument drift.

2.1.3 Dissolved inorganic carbon and total alkalinity

Samples for C_T and A_T analysis were collected from the underway seawater supply (322 samples in total), and shallow depths sampled (64 samples) using Ocean Test Equipment bottles on the CTD frame following procedures detailed in Bakker et al. (2007). All samples were fixed with 50 µL of saturated mercuric chloride ($HgCl_2$) solution per 250 mL seawater. The samples were analysed in duplicate on replicate 250 mL samples bottles. Two VINDTAs 3C (Versatile Instrument for the Determination of Titration Alkalinity; Marianda) were used to determine C_T and A_T, with CRMs (batch 107) analysed in duplicate for C_T and A_T at the beginning, middle and end of each use of a coulometric cell. One VINDTA was used for surface water samples and the other one for CTD casts. For consistency, we checked offsets between those underway and CTD samples which were less than 30 min apart. Because underway sampling was often stopped when we were on station, only 19 stations could be checked. The average offsets between CTD and surface water samples from the continuous supply were 0.3 and 0.5 $\mu mol\,kg^{-1}$ for C_T and A_T, respectively. This is below the accuracy of the method and so we merged the data from the two types of sampling and the two VINDTA 3C instruments. The concentration of C_T was determined using coulometric analysis (Johnson et al., 1987). Analysis for A_T was carried out by potentiometric titration with hydrochloric acid to the carbonic acid end point (Dickson, 1981). The accuracies of the C_T and A_T measurements were 2.0 and 1.5 $\mu mol\,kg^{-1}$ and the precisions 1.7 and 1.2 $\mu mol\,kg^{-1}$, respectively (159 CRMs analysed in duplicate). The combined carbonate chemistry dataset is available via the British Oceanographic Data Centre at https://www.bodc.ac.uk/data/published_data_library/catalogue/10.5285/f56e35bc-635e-0ab5-e044-000b5de50f38/ with doi: 10.5285/f56e35bc-635e-0ab5-e044-000b5de50f38.

2.2 Nutrients

Analyses of nitrate and nitrite, phosphate and silicate were undertaken using a segmented flow auto-analyser (Skalar San+) following methods described by Kirkwood (1989). Samples were stored in 25 mL polycarbonate vials and kept refrigerated at approximately 4 °C until analysis (conducted within 12 h after sampling). Nutrient concentrations were used for the calculations of the carbonate chemistry system.

2.3 Carbonate chemistry calculations

We applied the CO2SYS programme (MATLAB version) (Lewis and Wallace, 1998; Van Heuven et al., 2011) to all possible pairs of pH_T, pCO_2, C_T and A_T measurements to calculate the other variables, using the carbonate equilibria constants described by Mehrbach et al. (1973) and refitted by Dickson and Millero (1987) (here after Mehrbach constants). We used the Mehrbach constants because they led to the smallest inconsistencies between different high-accuracy measurements in previous observational studies (Clayton et al., 1995; Lee et al., 1997; McElligott et al., 1998; Wanninkhof et al., 1999). However, we also compared against results calculated using constants from Roy et al. (1993), Lueker et al. (2000) and Millero et al. (2006) (Sect. 3.3).

For the dissociation constant of boric acid we used Dickson (1990b), for bisulphate ions Dickson (1990a) and for the ratio of total boron to salinity, we used Uppström (1974), but also compared against results calculated using Lee et al. (2010) (Sect. 3.3).

2.4 Intercomparison

The sampling frequencies of $p\mathrm{CO}_2$-1 and pH_T were both around 5 min, but the measurements were not synchronised and were undertaken simultaneously (within 1 min) on only 208 occasions. However, it was possible to interpolate pH_T determinations (with a maximum interval of 5 min) and thereby obtain values at comparable times.

Data from the underway temperature, salinity, PAR, $p\mathrm{CO}_2$ and pH_T measurements were retrieved at the times of nutrient, C_T and A_T measurements.

Statistical analyses were used to determine the level of agreement between observed and calculated carbonate system variables:

- Pearson’s correlation coefficient (r): A measure of the degree of linear dependence between two variables.
- Mean Residual (MR): Average difference between two variables, e.g. between the observed values and the values calculated from measurements of a pair of other carbonate variables; MR will be negative if the observed values are on average lower than the calculated values.
- Root Mean Square Error (RMSE): Square root of the mean of the squared differences between the observed and calculated values.

Some properties of individual variables (as opposed to comparisons) are also used in Sect. 3:

- Accuracy is an expression of the lack of bias and relates to the degree of agreement of a measured value with the true value (as determined using a CRM).
- Uncertainty characterizes the range of values within which the true value is asserted to lie with some level of confidence. Uncertainty is derived from inaccuracy and imprecision of measurements, and also from propagation of errors for calculated variables.

In this study, uncertainties in calculated values were determined by a Monte Carlo approach as follows: 1) The original carbonate chemistry variable values in the dataset were input into the CO2SYS program (MATLAB version) (Van Heuven et al., 2011); 2) Artificial random errors (normally distributed according to the central limit theorem, with a mean of zero and standard deviation equal to the accuracy of measurement) were calculated using a random number generator; 3) New carbonate chemistry variable values (the original ones plus the randomly generated errors) were input into CO2SYS. Calculated $p\mathrm{CO_2}$-1 and calculated $p\mathrm{CO_2}$-2 have the same uncertainty because they depend only on the accuracies of the variables from which they are calculated. They are therefore identical for both $p\mathrm{CO_2}$ systems. The calculated uncertainty of parameters calculated from $p\mathrm{CO_2}$-2 as one of the input variables is higher than those from $p\mathrm{CO_2}$-1 because the measurement accuracy was higher. This Monte Carlo approach has previously been used by Juranek et al. (2009) to calculate uncertainties in calcium carbonate saturation states.

In this paper we use accuracies and calculated uncertainties as benchmarks. We use them to provide an assessment of what it is reasonable to expect for a match between observed and calculated values. We compare values of accuracy with MRs, and calculated uncertainties with RMSEs, to evaluate if the calculated variables are in “good” agreement with the measured variables. All comparison resulted in MR less than or equal to accuracy (with the exception of predictions based on $p\mathrm{CO_2}$ and pH and A_T from C_T and $p\mathrm{CO_2}$-1). In terms of RMSE, all comparisons resulted in RMSE less than twice the uncertainty, except those involving measured $p\mathrm{CO_2}$-2. On this basis we conclude that there is generally good agreement between measured and calculated variables, except those comparison involving measured $p\mathrm{CO_2}$-2 or calculations from $\mathrm{pH_T}$ and $p\mathrm{CO_2}$. For example, if the MR and RMSE between measured and calculated (from A_T and $p\mathrm{CO_2}$-1) C_T are -1.7 $\mu\mathrm{mol\,kg^{-1}}$ and 3.9 $\mu\mathrm{mol\,kg^{-1}}$, whereas the accuracy is 2.0 $\mu\mathrm{mol\,kg^{-1}}$ and the calculated uncertainty is 2.5 $\mu\mathrm{mol\,kg^{-1}}$, then we conclude that there is good agreement between the calculated and measured C_T values. As another example, a RMSE of 0.008 and a MR of 0.001 demonstrates a good agreement between measured and calculated (from C_T and A_T) $\mathrm{pH_T}$ if the calculated uncertainty according to the Monte Carlo approach is 0.005 and the measurement accuracy is 0.004. A smaller MR on its own does not demonstrate a better agreement; it should always be put in context with RMSE, accuracy and uncertainty.

3 Results and discussion

3.1 Comparison between two pCO_2 systems

The pCO_2 datasets obtained using the pCO_2-1 and pCO_2-2 systems were significantly correlated (r = 0.956, $p < 0.001$, df = 2679) (Fig. 1). We used a major axis model II regression because both pCO_2 datasets included uncertainty (pCO_2-1 and pCO_2-2 was not one dependent controlled and one independent variable). We used the R code for Model II Regression (Legendre, 2014). The resulting equation of the regression is $pCO_2\text{-}1 = 0.9\ (\pm\ 2.1) + 0.99\ (\pm\ 0.01) \times pCO_2\text{-}2$. Confidence intervals are used for testing the null hypothesis of a slope of 1 and a y-intercept of 0 (Quinn and Keough, 2002). The 95 % confidence interval of the slope does include the value 1 and the 95 % confidence interval of the intercept does include the value 0. The comparison between both sets of pCO_2 data revealed a mean residual of -2 µatm (pCO_2-1 minus pCO_2-2). The RMSE was 10 µatm.

Körtzinger et al. (2000) reported that even after correction of all differences between equilibrator temperature readings and following a time synchronization procedure, the remaining mean residual in their study was ca. 2 µatm for most of their cruise. The study by Körtzinger et al. (2000) can be described as an ideal open ocean exercise with all instruments sharing a common seawater supply, positioned in the same laboratory and sharing common calibration gases. Considering that our cruise took place in coastal waters with strong gradients in temperature, salinity and chemical variables (Rérolle et al., 2014), with the instruments situated in different laboratories and using different calibration gases, and known water flow problems (Sect. 2.1.1), we conclude that an average difference of 2 µatm and a RMSE of 10 µatm were good outcomes. The non-ideal conditions of our intercomparison did not negatively affect the overall consistency (we obtained comparable results to the study by Körtzinger et al. (2000)). This result is also comparable with a previous (the only other) coastal water intercomparison, described by Körtzinger et al. (1996). In this study, where there was highly variable spatial pCO_2 distribution in the southern North Sea, the average difference between observed values was 0.2 µatm (standard deviation = 1.2 µatm).

There have been a number of intercalibration exercises of pCO_2 systems in an indoor seawater pool at the National Institute for Environment Studies, Japan (in 1993, 1998, 2003 and 2009) (Katayama et al., 1999; IOCCP, 2004; Pierrot et al., 2009). Most of the instruments showed good agreement (within 2 µatm).

The MR of the observed pCO_2-1 compared with pCO_2 calculated from C_T and A_T was 3 µatm (n = 43), and of pCO_2-2 was 1 µatm (n = 156), both MR within the measurement accuracy of the instrument (Table 1). The accuracies of C_T and A_T were ± 2.0 µmol kg^{-1} and ± 1.5 µmol kg^{-1} respectively, and this translates into a propagated pCO_2 uncertainty of 4 µatm. This is a clear example of where the MR does not provide whole story, because pCO_2-2 compared to calculated from C_T and A_T had a smaller MR but higher RMSE and lower r than pCO_2-1 (Table 1). Lower MR does not necessarily mean that there is a better agreement, because positive values may compensate negative values, as is the case here. The ranges of the residuals were (-7 to 14) µatm and (-29 to 39) µatm, for pCO_2-1 and pCO_2-2 respectively. Both of the mean residuals were well within the expected accuracy of pCO_2 calculated from C_T and A_T measurements (Millero, 2007). The residuals of the two pCO_2 datasets are presented in Fig. 2. A diurnal cycle was observed in the residuals after 8 days of the cruise, with the amplitude increasing over time and reaching a maximum difference between the two pCO_2 systems on julian day 177 (Figs. 2 and 3). The average difference (pCO_2-1 minus pCO_2-2) was -2 µatm and the maximum difference was 41 µatm.

Temperature forms a critical parameter for pCO_2 calculations, with for example temperature differences of 2 ºC translate into pCO_2 differences of 32 µatm. The two analytical pCO_2 systems were not in the same laboratory and therefore water spent different lengths of time and took different routes between the ship's seawater intake and the two equilibrators, and therefore warmed differently.This temperature effect was more important when the ship sailed through strong surface water temperature gradients. Rapid changes in seawater temperature of up to 2-3 °C min^{-1} were observed on julian day 180. However, such rapid temperature changes at the seawater intake may not fully translate into pCO_2 changes in the equilibrator, as equilibrator pCO_2 is effectively integrated over the equilibration time (around 8 min for CO_2 in both equilibrators). Small differences in the equilibration time between the two systems, i.e. how rapidly they respond to a change in seawater pCO_2, may account for some of the observed differences between observations by pCO_2-1 and pCO_2-2. Moreover, no trend was observed between the difference in measured pCO_2-1 and pCO_2-2 versus the difference between the temperature in equilibrator 1 and the sea surface temperature, whereas a positive relationship was observed for equilibrator 2 (Fig. 4). The discrepancy between the two pCO_2 systems was negative at low light levels (at night pCO_2-1 was smaller than pCO_2-2) and positive during daylight hours (pCO_2-1 larger than pCO_2-2) (Fig. 3). This pattern is consistent with respiration at night and photosynthesis during the day in the seawater supply to the

pCO$_2$-2 equilibrator or in the equilibrator itself. The length of the seawater pipes to equilibrator 2 was about twice as long as to equilibrator 1. Furthermore, the equilibrator of pCO$_2$-2 was subject to direct daylight, compared to the pCO$_2$-1 equilibrator which was shielded from light. A multiple regression analysis was performed to estimate the relative importance of two factors (the temperature difference between the two equilibrators (ΔT_{eq}) and PAR) in determining the size of the pCO$_2$ differences (ΔpCO$_2$ = pCO$_2$-1 - pCO$_2$-2). A correlation coefficient of 0.47 between the predictor variables (ΔT_{eq} and PAR) indicated that they are not strongly correlated. Tolerance (or the inverse of the variance inflation factor) is 0.78 indicating no colinearity problems as this value is well above the tolerance threshold of 0.1 (Quinn and Keough, 2002). The analysis indicated that the correlation of ΔpCO$_2$ with ΔT_{eq} and PAR was statistically significant ($p < 0.0001$; $F = 587.6$) and that the two parameters together explained 38 % of the ΔpCO$_2$ variance. ΔT_{eq} and PAR were found to individually explain 17 % and 18 % respectively of the variance in ΔpCO$_2$. Thus we conclude that the diurnal pattern in the pCO$_2$ differences between the two instruments were caused by a combination of: (a) biological activity in the seawater system or in equilibrator 2, and (b) variations in water flow and temperature perturbations in pCO$_2$-2 (Sect. 2.1.1). The first issue should be addressed in future studies by protection of seawater tubing and equilibrators from light and by regular cleaning of the seawater intake and equilibrators. In addition, the flow rate of the ship's seawater supply should be kept constant, while the water flow to equilibrators should be sufficiently large and should be kept constant by using a water flow controller. Furthermore, the temperature measurements inside the equilibrator need to be accurate, in agreement with Körtzinger et al. (2000) and as implemented in the Surface Ocean CO$_2$ Atlas (SOCAT) by making an accuracy of 0.05 °C for the equilibrator temperature a pre-requisite for data set quality flags A and B (Pfeil et al., 2013; Bakker et al., 2014).

3.2 Intercomparison of measured and calculated variables

The results of the intercomparison between observed carbonate chemistry variables and those calculated from different pairs of measured variables are presented in Table 1. Statistical techniques were used to evaluate the agreement between the observed and the calculated values. The comparison between observed pH$_T$ and pH$_T$ calculated from observed C_T and A_T showed a mean residual of 0.001 pH$_T$ units and a RMSE of 0.008 pH$_T$ units (Table 1). This compared to the 0.004 pH$_T$ units accuracy of the measurements. The linear correlation coefficient (r) between observed and calculated values was 0.952. pH_T calculated from C_T and A_T had an uncertainty of 0.005 pH_T units. The RMSE corresponded to twice the accuracy of the pH_T measurements, and slightly less than twice the uncertainty expected from the calculation, so we conclude that there was good agreement between calculated and measured pH_T.

The value of pH_T calculated from pCO$_2$-1 and A_T had a RMSE of 0.006 when compared to measured pH_T, and the same calculation with pCO$_2$-2 led to a RMSE of 0.013 (Table 1). The calculated values of pH_T using pCO$_2$-1 and A_T were therefore better (lower RMSE) than pH_T calculated from C_T and A_T, whereas the calculations using pCO$_2$-2 and A_T had a higher RMSE. Calculations of pH_T from the combination of pCO$_2$ with either C_T or A_T may be expected to yield more accurate estimates than calculations of pH_T from C_T and A_T (Table 1), because they do not require reliable estimates of the second dissociation constant of carbonic acid and are relatively insensitive to uncertainties in C_T and A_T (Millero, 2007). However, this is not always the case, as shown here.

pCO$_2$ calculated from C_T and A_T compared to the pCO$_2$-1 and pCO$_2$-2 observational datasets showed RMSE values of 6 µatm and 12 µatm, and mean residuals of 3 and 1 µatm (Table 1). This compares to an accuracy associated with direct measurement of pCO$_2$ of 4 µatm for pCO$_2$-1 and 10 µatm for pCO$_2$-2. pCO$_2$ calculated from C_T and A_T is predicted to have an uncertainty of 4 µatm. We therefore conclude that our calculated dataset is in good agreement with the measured dataset.

Lueker et al. (2000) report that the mean relative difference between measured fCO$_2$ and fCO$_2$ calculated from C_T and A_T (for fCO$_2$ less than 500 µatm) was 0.07 % (standard deviation = 0.50 %). For example, if fCO$_2$ was 400 µatm that will be 0.3 µatm (standard deviation = 2.0 µatm). For fCO$_2$ above 500 µatm, there was a mean relative difference of 3.3 % (standard deviation = 1.2 %). For example, if fCO$_2$ was 500 µatm that will be 16.5 µatm (standard deviation = 6.0 µatm).

pCO$_2$ and pH_T do not make a good pair for predicting other variables because CO$_2$ and hydrogen ion concentration are smaller than carbonate and bicarbonate concentration. Therefore, relatively small errors in CO$_2$ and/or hydrogen ion propagate into relatively large errors in carbonate and bicarbonate concentration, when the system is computed from pCO$_2$ and pH_T. This is also indicated by our data, where the mean residuals for A_T and C_T (observed minus calculated) are comparable to the calculated uncertainties, but are both one order of magnitude greater from pCO_2 and pH_T than the accuracy and precision of observations (Table 1).

C_T calculated from A_T and pCO_2 had RMSE values of 3.9 µmol kg^{-1} and 7.2 µmol kg^{-1} compared to the measurement datasets of systems pCO_2-1 and 2, and mean residuals of -1.7 and -0.3 µmol kg^{-1} (Table 1). This compares to an accuracy associated with direct measurement of C_T of 2 µmol kg^{-1}. A_T calculated from C_T and pCO_2-1 had a MR of to 2.1 µmol kg^{-1} and a RMSE value of 4.6 µmol kg^{-1} and the calculated uncertainty was 3 µmol kg^{-1} (Table 1). This compares to an accuracy associated with direct measurement of A_T of 1.5 µmol kg^{-1}.

Several papers have raised the issue of the impact of organic acids on computations of the CO_2 system in coastal waters (Kim et al., 2006; Hernández-Ayón et al., 2007; Kim and Lee, 2009). Dissolved organic matter produced by phytoplankton during photosynthesis can potentially make a significant contribution to seawater total alkalinity although we saw no evidence for this in our study (analysis not shown).

During three days in the same coccolithophore bloom we collected additional samples of filtered C_T and A_T from the underway seawater supply. The filtering was carried out using an in-line filter (Sartorius Sartobran 300 Sterile capsule, 0.45 µm pore size). In this way we studied the effects of the presence of calcite mineral particles in the surface waters on C_T and A_T measurements. The average differences between unfiltered and filtered samples were 2.4 µmol kg^{-1} and 3.7 µmol kg^{-1} for C_T and A_T, respectively, with values in the unfiltered samples being higher. The differences between replicates of filtered samples were on average higher than those of unfiltered samples. The differences between measured pH_T and pH_T calculated from filtered and unfiltered C_T and A_T were -0.003 and 0.005, respectively. The difference between measured pCO_2-1 and pCO_2-2 and pCO_2 calculated from filtered and unfiltered C_T and A_T were -4 µatm in both cases. An influence of dissolution of calcite particle on C_T and A_T measurements would be expected to lead to filtered C_T and A_T being lower than unfiltered with discrepancies twice as large for A_T as for C_T. This ratio was however not exactly observed in our study, and hence it is unclear from our observations whether dissolution of $CaCO_3$ particles affected C_T and A_T measurements. Filtration of samples for C_T potentially introduces a further error through CO_2 loss by turbulence and ebullition, which can affect the C_T measurement, although we took precautions to avoid bubbles in the filter.

So far in this paper we have checked the consistency of the carbonate system and assessed the quality of the carbonate system measurements. Overall, the sizes of the offsets between measured and calculated values agreed well with expectations based on theoretical calculations and expected accuracies of measurements. Four out of the five independent datasets of carbonate chemistry variables are deemed to be of high-quality and therefore suitable to be used as a basis for evaluations of the impacts of ocean acidification by other scientists on the some cruise, see for instance the papers by Poulton et al. (2014) and Young et al. (2014).

Next, we examine possible reasons for discrepancies observed during the $p\mathrm{CO}_2$ measurements using $p\mathrm{CO}_2$-2 (points 1 to 3 below) and make 'good practice' recommendations from investigation of the remaining discrepancies (points 4 and 5 below). Some of these recommendations are new and some are not but we think it is good to compile them all:

1. The duration of seawater transit in the ship's underway system from the seawater intake to the point of sample collection or measurement (1 to 3 min) varied between the carbonate chemistry measurements. It is important to recognise the period of time it takes for the seawater to arrive in the equilibrator of a $p\mathrm{CO}_2$ instrument. This will allow correction for the difference between the intake temperature and the equilibrator temperature. This has already been emphasized by Dickson et al. (2007). The omission of a time correction results in unrealistic spikes in the difference between seawater temperature and the temperature inside the equilibrator. Appropriate temperature correction will reduce artificial variability in $p\mathrm{CO}_2$, even though the average $p\mathrm{CO}_2$ value is still likely to be correct. In addition, both the underway seawater system and the equilibrators tend to smooth out short-lived signals in temperature and $p\mathrm{CO}_2$, because of the time the water spent travelling along the underway water supply and inside the equilibrator. These delay-times can affect the measurement reliability by smoothing out strong gradients. These effects are particularly important in regions with rapid changes in carbonate chemistry and sea water temperature, for example in shelf sea regions with freshwater inputs, in continental shelf break regions with enhanced vertical mixing (internal tides or upwelling) and in regions with sea ice melt. The effects of strong gradients on $p\mathrm{CO}_2$ have been noted by Körtzinger et al. (1996) and Körtzinger et al. (2000).
2. To obtain high quality $p\mathrm{CO}_2$ datasets we recommend special care be taken with the operation of the equilibrator systems, including a) careful control of the seawater supply

and the water flow through the equilibrator (Körtzinger et al., 1996); b) accurate temperature readings (Körtzinger et al., 2000; Pierrot et al., 2009; Bakker et al., 2014) and c) prevention of phototrophic growth in the equilibrator by complete shielding from light. The last recommendation, as far as we know, is new and we recommend that all equilibrators are fully covered.

3. When comparing a pair of variables, it is important to examine the residuals as a function of time as well as constructing a scatterplot of one variable against the other. For example, from examination of Fig. 1 alone we would not have identified the diurnal variation between the two $p\mathrm{CO}_2$ systems (Fig. 2).

4. Differences between the recorded sampling time and the actual sampling time also need to be taken into account. C_T and A_T are discrete measurements, while $p\mathrm{CO}_2$ and pH_T are near-continuous measurements. In addition to the transit times between the intake and the instrument, $p\mathrm{CO}_2$ is an integrated measurement over the timescale of equilibration (around 8 min) and pH_T is an integrated measurement over the timescale of filling the sample chamber (ca. 60 seconds). When comparing carbonate chemistry datasets, corrections should be made for the asynchronous times of sample collection for the different variables.
5. It is recommended to characterise the extinction coefficients of each batch of pH-indicator-dye on the instrument used for ship-board pH analysis, rather than use published values (Clayton and Byrne, 1993; Zhang and Byrne, 1996; Hopkins et al., 2000; Mosley et al., 2004; Gabriel et al., 2005; Liu et al., 2011). This is particularly important where the indicator has not been purified (Yao et al., 2007) and where the detection system has a wider optical bandwidth than that used in the literature to characterize the indicator (here the optical bandwidth used for our pH system is 15-20 nm (Rérolle et al., 2013)). The discrepancy between pH_T values calculated with our coefficients and values calculated with coefficients from Zhang and Byrne (1996) was about 0.02 pH_T units (Rérolle et al., 2013). Rérolle et al. (2013) estimated that about 0.005 pH_T units of the observed discrepancy was due to impurities in the indicator and about 0.015 was due to the wider bandpass detection window in our ship-board pH system. This recommendation has been made before by Liu et al. (2011) and Yao et al. (2007).

3.3 Ratio of total boron to salinity and carbonate constants

The calculations of the carbonate chemistry variables were undertaken for a second time using another ratio of total boron to salinity (Lee et al., 2010) for the combinations of pairs involving A_T (because the ratio of total boron to salinity only influences calculations involving A_T). We then compared the results obtained against those using the original ratio of total boron to salinity from Uppström (1974). Statistically significant differences (ANOVA, p-values < 0.001) were seen for both pCO$_2$ and pH$_T$ calculated from C_T and A_T using the different ratios of total boron to salinity. There were also significant differences between C_T values calculated from A_T and pH$_T$ and from A_T and pCO$_2$-2 using the different constants, and A_T calculated from both C_T and pH$_T$ and C_T and pCO$_2$-2. There were, however, no statistically significant differences when calculating pH$_T$ from pCO$_2$-1 or pCO$_2$-2 and A_T, or pCO$_2$-1 and pCO$_2$-2 calculated from pH$_T$ and A_T (Fig. 5 and Table 2). This was because the calculation of pH from pCO$_2$ and A_T is mainly driven by pCO$_2$ and not A_T (and the same when calculated pCO$_2$ from pH$_T$ and A_T). The residuals were on the whole smaller when using the ratio of total boron to salinity from Uppström (1974) compared to using Lee et al. (2010) (Fig. 5). The differences between mean residuals were 5 µatm, 0.007 pH$_T$ units, 2.9 µmol kg^{-1} and 3.2 µmol kg^{-1} for pCO$_2$, pH$_T$, C_T and A_T, respectively when substracting results obtained using the constants from Uppström (1974) from those obtained using the constants from Lee et al. (2010). These discrepancies might not be so significant in an ocean acidification context, but they are substantial in terms of air-sea flux calculations in coastal waters. Overall, for our ranges of temperature and salinity, Uppström (1974) gave the best results when compared to observed values for carbonate chemistry variables.

The calculations of the carbonate chemistry variables were undertaken for a third time using other sets of carbonate constants (Roy et al., 1993; Lueker et al., 2000; Millero et al., 2006). We then compared the results obtained against those using the original carbonate constants, from Mehrbach and the total boron to salinity ratio of Uppström (1974). We omitted the pair pH$_T$-pCO$_2$ from consideration because they are not a good pair for the calculations.

Statistically significant differences (ANOVA, p-values < 0.001) were noted when using the Roy et al. (1993) constants (hereinafter Roy) to calculate pCO$_2$-1 and pCO$_2$-2 from C_T and A_T, C_T or pH$_T$ or A_T and pH$_T$. The use of Lueker et al.'s (2000) constants (hereinafter Lueker) produces results that do not yield statistically significant differences from Mehrbach when calculating pCO$_2$-1 and pCO$_2$-2 from C_T and A_T, C_T or pH$_T$ or A_T and pH$_T$. The use of Millero et al.'s (2006) constant (hereinafter Millero) did not yield statistically significant differences from Mehrbach when calculating pCO_2-2 from C_T and A_T, C_T or pH_T or A_T and pH_T. The maximum mean residuals were always observed when using Roy's constants (up to 23 µatm when calculating pCO_2-2 from C_T and A_T) (Table 3). The calculated pCO_2 values using the constants of Roy were significantly higher than those by Mehrbach in other studies (Wanninkhof et al., 1999). McElligott et al. (1998) showed a good agreement between measured pCO_2 and pCO_2 calculated from C_T and A_T using Mehrbach at the sea surface.

There were no significant differences when calculating pH_T from pCO_2-1 and either C_T or A_T. However, there were statistically significant differences when using Roy for calculating pH_T from pCO_2-2 and C_T or A_T. Of particular importance, the mean residual is significantly higher when using Roy's constants to calculate pH_T from C_T and A_T (Table 3). McElligott et al. (1998) noted that all four CO_2 parameters measured during the NOAA Equatorial Pacific CO_2 cruises were internally consistent when using the constants of Mehrbach et al. (1973) refitted by Dickson and Millero (1987) at 20-25 °C, if spectrophotometric pH_T values were increased by 0.0038 (value proposed by DelValls and Dickson (1998)).

For calculations of C_T and A_T, the constants by Roy led to a mean residual further from 0 (statistically significant) than using Mehrbach (large differences between paired values).The reasons for the occasional significant differences between constants are not entirely clear and investigating the discrepancies is beyond the scope of this work. Objective comparison of the constants is difficult because of differences between the ranges of temperature and salinity0 over which the constants were measured, and the curve fitting procedures differed (Wanninkhof et al., 1999).

The constants by Mehrbach were determined in artificial seawater using the seawater pH scale with a range in temperature between 2 and 35 °C and in salinity between 20 and 40. The constants by Lueker were measured in natural seawater using the total pH scale with a range in temperature of 2-35 °C and in salinity of 19-43. Lueker et al. (2000) converted the constants by Mehrbach to the total hydrogen ion pH scale. The constants by Roy were measured in artificial seawater using the total pH scale with a range in temperature of 0-45 °C and in salinity of 5-45. The constants by Roy have been used for studies in the polar waters (Jutterström and Anderson, 2005; Chierici et al., 2011). The constants by Millero were measured in natural seawater using the seawater pH scale with a range in temperature of 0-50 °C and in salinity of 1-50. Dickson et al. (2007) reported that the constants by Lueker were in reasonable agreement with those reported by Roy et al. (1993). In our study this did not appear to be the case.

Overall, the use of constants by Lueker gave similar results to those of Mehrbach (perhaps unsurprisingly because they are refitted from constants by Mehrbach) but led to larger mean residuals. Although using constants by Millero did not usually lead to statistically differences compared to Mehrbach, the residuals were overall higher than when using either the constants by Mehrbach or Lueker. Our dataset confirms that the constants by Mehrbach provided the best quality results as previously demonstrated by other studies examining the internal consistency in the laboratory (Lee et al., 1996; Lueker et al., 2000) and in the field (Wanninkhof et al., 1999; Lee et al., 2000).

4 Conclusions

Our results show that it is possible to obtain good consistency between measurements of different variables of the carbonate system, even outside the somewhat artificial conditions of an intercomparison exercise. However, our retrospective intercomparison revealed several sources of discrepancies, leading to the following recommendation for best practice: a) undertake characterization of the pH-indicator-dye in order to obtain correct extinction coefficients for the dye and analytical system in the temperature and salinity range used; b) take into account the transit time of seawater from the intake to the equilibrator, when comparing $p\mathrm{CO}_2$ with other simultaneous measurements; c) examine residuals as a function of time in order to detect temporal biases in measurements; d) prevent phototrophic growth in $p\mathrm{CO}_2$ equilibrators by completely shielding them from exposure to light.

We obtained smaller average residuals when using the ratio of total boron to salinity from Uppström (1974) rather than Lee et al. (2010) and when using the carbonate constants by Mehrbach et al. (1973) (refitted by Dickson and Millero (1987)) rather than Roy et al.'s (1993), Lueker et al.'s (2000) or Millero et al.'s (2006). As found in other studies (Millero, 1995; Cullison Gray et al., 2011), the variables pH_T and $p\mathrm{CO}_2$ are far from an ideal pair for calculation of C_T or A_T, emphasizing the desirability of developing a C_T or A_T sensor capable of autonomous high resolution measurements.

Acknowledgements

This work is a contribution to the UK Ocean Acidification Research Programme (UKOA) which is jointly funded by the Department for Environment, Food and Rural Affairs (DEFRA), the Natural Environment Research Council (NERC) and the Department for Energy and Climate Change (DECC) under grant agreement no. NE/H017348/1. We thank the captain and crew of RRS *Discovery* for assistance during the cruise and N. Wager for her help with e-folding times. N. Greenwood, J.-P. Gattuso and two anonymous reviewers provided valuable comments on earlier versions of this work.

References

Bakker, D. C. E., Nielsdóttir, M. C., Morris, P. J., Venables, H. J., and Watson, A. J.: The island mass effect and biological carbon uptake for the subantarctic Crozet Archipelago, Deep-Sea Res Pt. II, 54, 2174-2190, doi: 10.1016/j.dsr2.2007.06.009, 2007.

Bakker, D. C. E., Pfeil, B., Smith, K., Hankin, S., Olsen, A., Alin, S. R., Cosca, C., Harasawa, S., Kozyr, A., Nojiri, Y., O'Brien, K. M., Schuster, U., Telszewski, M., Tilbrook, B., Wada, C., Akl, J., Barbero, L., Bates, N. R., Boutin, J., Bozec, Y., Cai, W. J., Castle, R. D., Chavez, F. P., Chen, L., Chierici, M., Currie, K., de Baar, H. J. W., Evans, W., Feely, R. A., Fransson, A., Gao, Z., Hales, B., Hardman-Mountford, N. J., Hoppema, M., Huang, W. J., Hunt, C. W., Huss, B., Ichikawa, T., Johannessen, T., Jones, E. M., Jones, S. D., Jutterström, S., Kitidis, V., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Manke, A. B., Mathis, J. T., Merlivat, L., Metzl, N., Murata, A., Newberger, T., Omar, A. M., Ono, T., Park, G. H., Paterson, K., Pierrot, D., Ríos, A. F., Sabine, C. L., Saito, S., Salisbury, J., Sarma, V. V. S. S., Schlitzer, R., Sieger, R., Skjelvan, I., Steinhoff, T., Sullivan, K. F., Sun, H., Sutton, A. J., Suzuki, T., Sweeney, C., Takahashi, T., Tjiputra, J., Tsurushima, N., van Heuven, S. M. A. C., Vandemark, D., Vlahos, P., Wallace, D. W. R., Wanninkhof, R., and Watson, A. J.: An update to the Surface Ocean CO2 Atlas (SOCAT version 2), Earth Syst. Sci. Data, 6, 69-90, 10.5194/essd-6-69-2014, 2014.

Chierici, M., Fransson, A., Lansard, B., Miller, L. A., Mucci, A., Shadwick, E., Thomas, H., Tremblay, J. E., and Papakyriakou, T. N.: Impact of biogeochemical processes and environmental factors on the calcium carbonate saturation state in the Circumpolar Flaw Lead in the Amundsen Gulf, Arctic Ocean, Journal of Geophysical Research: Oceans (1978–2012), 116, 2011.

Clayton, T. D., and Byrne, R. H.: Spectrophotometric sewater pH measurements: total hydrogen ion concentration scale calibration of *m*-cresol purple and at sea results, Deep-Sea Res. Pt. I, 40, 2115-2129, doi: 10.1016/0967-0637(93)90048-8, 1993.

Clayton, T. D., Byrne, R. H., Breland, J. A., Feely, R. A., Millero, F. J., Campbell, D. M., Murphy, P. P., and Lamb, M. F.: The role of pH measurements in modern oceanic CO_2-system characterizations: Precision and thermodynamic consistency, Deep-Sea Res Pt. II, 42, 411-429, 1995.

Cullison Gray, S. E., DeGrandpre, M. D., Moore, T. S., Martz, T. R., Friederich, G. E., and Johnson, K. S.: Applications of *in situ* pH measurements for inorganic carbon calculations, Mar. Chem., 125, 82-90, 2011.

Dickson, A. G.: An exact definition of total alkalinity and a procedure for the estimation of alkalinity and total inorganic carbon from titration data, Deep-Sea Res., 28, 609-623, 1981.

Dickson, A. G., and Millero, F. J.: A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, Deep-Sea Res., 34, 1733-1743, doi: 10.1016/0198-0149(87)90021-5, 1987.

Dickson, A. G.: Standard potential of the reaction $AgCl(s)+1/2H_2(g) = Ag(s)+HCl(aq)$ and the standard acidity constant of the bisulfate ion in synthetic seawater from 273.15 K to 318.15 K, J. Chem. Thermodyn., 22, 113-127, doi: 10.1016/0021-9614(90)90074-z, 1990a.

Dickson, A. G.: Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K, Deep Sea Research Part A. Oceanographic Research Papers, 37, 755-766, 1990b.

Dickson, A. G., Sabine, C. L., and Christian, J. R.: Guide to best practices for ocean CO_2 measurements, PICES Special Publication, 3, 2007.

Dickson, A. G.: Interactive comment on "On CO_2 pertubation experiments: over-determination of carbonate chemistry reveals inconsistencies" by Hoppe, C. J. M. Langer, G., Rokitta, S.D., Wolf-Gladrow, D.A. and Rost, B., Biogeosciences Disc., 7, C1286-C1293, 2010.

Gabriel, M., Forja, J., Rubio, J., and Gómez-Parra, A.: Temperature and salinity dependence of molar absorptivities of thymol blue: Application to the spectrophotometric determination of pH in estuarine waters, Cienc. Mar., 31, 309-318, 2005.

Hardman-Mountford, N. J., Moore, G., Bakker, D. C., Watson, A. J., Schuster, U., Barciela, R., Hines, A., Moncoiffé, G., Brown, J., Dye, S., Blackford, J., Somerfield, P. J., Holt, J., Hydes, D. J., and Aiken, J.: An operational monitoring system to provide indicators of CO_2-related variables in the ocean, ICES J. Mar. Sci., 65, 1498-1503, 2008.

Hopkins, A. E., Sell, K. S., Soli, A. L., and Byrne, R. H.: In-situ spectrophotometric pH measurements: The effect of pressure on thymol blue protonation and absorbance characteristics, Mar. Chem., 71, 103-109, 2000.

IOCCP: Ocean surface pCO_2, data integration and database development, 14-17 January 2004, IOCCP report 2, IOC Worskshop Report 211, available at: http://www.ioccp.org/images/01Underway/WR211_eo.pdf (last access 28 April 2014), UNESCO Tsukuba, Japan, 2004.

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

Johnson, K. M., Körtzinger, A., Mintrop, L., Duinker, J. C., and Wallace, D. W. R.: Coulometric total carbon dioxide analysis for marine studies: measurement and internal consistency of underway TCO_2 concentrations, Mar. Chem., 67, 123-144, doi: 10.1016/s0304-4203(99)00055-9, 1999.

Juranek, L., Feely, R., Peterson, W., Alin, S., Hales, B., Lee, K., Sabine, C., and Peterson, J.: A novel method for determination of aragonite saturation state on the continental shelf of central Oregon using multi-parameter relationships with hydrographic data, Geophys. Res. Lett., 36, L24601, doi:10.1029/2009GL040778, 2009.

Jutterström, S., and Anderson, L. G.: The saturation of calcite and aragonite in the Arctic Ocean, Mar. Chem., 94, 101-110, http://dx.doi.org/10.1016/j.marchem.2004.08.010, 2005.

Katayama, H., Karasudani, T., Ishii, K., Marubayashi, K., and Ueda, H.: Development of a gas-liquid equilibrator for estimating CO_2 flux at the ocean surface, Journal of Atmospheric and Oceanic Technology, 16, 1450-1455, 1999.

Kirkwood, D.: Simultaneous determination of selected nutrients in sea water, International Council for the Exploration of the Sea (ICES) CM, 100, 29, 1989.

Kitidis, V., Hardman-Mountford, N. J., Litt, E., Brown, I., Cummings, D., Hartman, S., Hydes, D., Fishwick, J. R., Harris, C., Martinez-Vicente, V., Woodward, E. M. S., and Smyth,

T. J.: Seasonal dynamics of the carbonate system in the Western English Channel, Cont. Shelf Res., 42, 30-40, doi: 10.1016/j.csr.2012.04.012, 2012.

Körtzinger, A., Thomas, H., Schneider, B., Gronau, N., Mintrop, L., and Duinker, J. C.: At-sea intercomparison of two newly designed underway $p\mathrm{CO}_2$ systems—encouraging results, Mar. Chem., 52, 133-145, 1996.

Körtzinger, A., Mintrop, L., Wallace, D. W. R., Johnson, K. M., Neill, C., Tilbrook, B., Towler, P., Inoue, H. Y., Ishii, M., Shaffer, G., Saavedra, R. F. T., Ohtaki, E., Yamashita, E., Poisson, A., Brunet, C., Schauer, B., Goyet, C., and Eischeid, G.: The international at-sea intercomparison of $f\mathrm{CO}_2$ systems during the R/V *Meteor* Cruise 36/1 in the North Atlantic Ocean, Mar. Chem., 72, 171-192, 2000.

Lamb, M. F., Sabine, C. L., Feely, R. A., Wanninkhof, R., Key, R. M., Johnson, G. C., Millero, F. J., Lee, K., Peng, T. H., Kozyr, A., Bullister, J. L., Greeley, D., Byrne, R. H., Chipman, D. W., Dickson, A. G., Goyet, C., Guenther, P. R., Ishii, M., Johnson, K. M., Keeling, C. D., Ono, T., Shitashima, K., Tilbrook, B., Takahashi, T., Wallace, D. W. R., Watanabe, Y. W., Winn, C., and Wong, C. S.: Consistency and synthesis of Pacific Ocean CO_2 survey data, Deep-Sea Res Pt. II, 49, 21-58, doi: 10.1016/S0967-0645(01)00093-5, 2001.

Lee, K., Millero, F. J., and Campbell, D. M.: The reliability of the thermodynamic constants for the dissociation of carbonic acid in seawater, Mar. Chem., 55, 233-245, 10.1016/s0304-4203(96)00064-3, 1996.

Lee, K., Millero, F. J., and Wanninkhof, R.: The carbon dioxide system in the Atlantic Ocean, J Geophys Res-Oceans, 102, 15693-15707, 10.1029/97jc00067, 1997.

Lee, K., Millero, F. J., Byrne, R. H., Feely, R. A., and Wanninkhof, R.: The recommended dissociation constants for carbonic acid in seawater, Geophys. Res. Lett., 27, 229-232, 2000.

Lee, K., Kim, T. W., Byrne, R. H., Millero, F. J., Feely, R. A., and Liu, Y. M.: The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans, Geochim. Cosmochim. Acta, 74, 1801-1811, doi: 10.1016/j.gca.2009.12.027, 2010.

Legendre, P.: lmodel2: Model II Regression, R package version 1.7-2, http://CRAN.R-project.org/package=lmodel2, 2014.

Lewis, E., and Wallace, D. W. R.: Program developed for CO_2 system calculations. ORNL/CDIAC-105, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S., Department of Energy, Oak Ridge, Tennessee, 1998.

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

Lueker, T. J., Dickson, A. G., and Keeling, C. D.: Ocean $p\mathrm{CO_2}$ calculated from dissolved inorganic carbon, alkalinity, and equations for K_1 and K_2: Validation based on laboratory measurements of CO_2 in gas and seawater at equilibrium, Mar. Chem., 70, 105-119, 2000.

McElligott, S., Byrne, R., Lee, K., Wanninkhof, R., Millero, F., and Feely, R.: Discrete water column measurements of CO_2 fugacity and pH_T in seawater: A comparison of direct measurements and thermodynamic calculations, Mar. Chem., 60, 63-73, 1998.

Mehrbach, C., Culberso, C. H., Hawley, J. E., and Pytkowic, R. M.: Measurement of apparent dissociation constants of carbonic acid in seawater at atmospheric pressure, Limnol. Oceanogr., 18, 897-907, 1973.

Millero, F. J.: Thermodynamics of the carbon dioxide system in the oceans, Geochim. Cosmochim. Acta, 59, 661-677, http://dx.doi.org/10.1016/0016-7037(94)00354-O, 1995.

Millero, F. J., Pierrot, D., Lee, K., Wanninkhof, R., Feely, R., Sabine, C. L., Key, R. M., and Takahashi, T.: Dissociation constants for carbonic acid determined from field measurements, Deep-Sea Res. Pt. I, 49, 1705-1723, 2002.

Millero, F. J., Graham, T. B., Huang, F., Bustos-Serrano, H., and Pierrot, D.: Dissociation constants of carbonic acid in seawater as a function of salinity and temperature, Mar. Chem., 100, 80-94, doi: 10.1016/j.marchem.2005.12.001, 2006.

Millero, F. J.: The marine inorganic carbon cycle, Chem. Rev., 107, 308-341, doi: 10.1021/cr0503557, 2007.

Mosley, L. M., Husheer, S. L., and Hunter, K. A.: Spectrophotometric pH measurement in estuaries using thymol blue and *m*-cresol purple, Mar. Chem., 91, 175-186, 2004.

Pfeil, B., Olsen, A., Bakker, D. C. E., Hankin, S., Koyuk, H., Kozyr, A., Malczyk, J., Manke, A., Metzl, N., Sabine, C. L., Akl, J., Alin, S. R., Bates, N., Bellerby, R. G. J., Borges, A., Boutin, J., Brown, P. J., Cai, W. J., Chavez, F. P., Chen, A., Cosca, C., Fassbender, A. J., Feely, R. A., González-Dávila, M., Goyet, C., Hales, B., Hardman-Mountford, N., Heinze, C., Hood, M., Hoppema, M., Hunt, C. W., Hydes, D., Ishii, M., Johannessen, T., Jones, S. D., Key, R. M., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lourantou, A., Merlivat, L., Midorikawa, T., Mintrop, L., Miyazaki, C., Murata, A., Nakadate, A., Nakano, Y., Nakaoka, S., Nojiri, Y., Omar, A. M., Padin, X. A., Park, G. H., Paterson, K., Perez, F. F., Pierrot, D., Poisson, A., Ríos, A. F., Santana-Casiano, J. M., Salisbury, J., Sarma, V. V. S. S., Schlitzer, R., Schneider, B., Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Suzuki, T., Takahashi, T., Tedesco, K., Telszewski, M., Thomas, H., Tilbrook, B., Tjiputra, J., Vandemark, D., Veness, T., Wanninkhof, R., Watson, A. J., Weiss, R., Wong, C. S., and Yoshikawa-Inoue, H.: A uniform, quality controlled Surface Ocean CO2 Atlas (SOCAT), Earth Syst. Sci. Data, 5, 125-143, 10.5194/essd-5-125-2013, 2013.

Pierrot, D., Neill, C., Sullivan, K., Castle, R., Wanninkhof, R., Lüger, H., Johannessen, T., Olsen, A., Feely, R. A., and Cosca, C. E.: Recommendations for autonomous underway $p\mathrm{CO}_2$ measuring systems and data-reduction routines, Deep-Sea Res Pt. II, 56, 512-522, 2009.

Quinn, G. P., and Keough, M. J.: Experimental design and data analysis for biologists, Cambridge University Press, 2002.

Rérolle, V., Floquet, C. F., Harris, A. J., Mowlem, M. C., Bellerby, R. R., and Achterberg, E. P.: Development of a colorimetric microfluidic pH sensor for autonomous seawater measurements, Anal. Chim. Acta, 786, 124-131, 2013.

Rérolle, V. C. M., Ribas-Ribas, M., Kitidis, V., Brown, I., Bakker, D. C. E., Lee, G. A., Shi, T., Mowlem, M. C., and Achterberg, E. P.: Controls on pH in surface waters of northwestern European shelf seas, Biogeosciences Disc., 11, 943-974, doi: 10.5194/bgd-11-943-2014, 2014.

Rérolle, V. M. C., Floquet, C. F. A., Mowlem, M. C., Bellerby, R., Connelly, D. P., and Achterberg, E. P.: Seawater-pH measurements for ocean-acidification observations, Trends Anal. Chem., 40, 146-157, doi: 10.1016/j.trac.2012.07.016, 2012.

Roy, R. N., Roy, L. N., Vogel, K. M., Porter-Moore, C., Pearson, T., Good, C. E., Millero, F. J., and Campbell, D. M.: The dissociation constants of carbonic acid in seawater at salinities 5 to 45 and temperatures 0 to 45 C, Mar. Chem., 44, 249-267, 1993.

Schuster, U., and Watson, A. J.: A variable and decreasing sink for atmospheric CO_2 in the North Atlantic, J Geophys Res-Oceans, 112, C11006, doi: 10.1029/2006jc003941, 2007.

Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W., and Sutherland, S. C.: Seasonal variation of CO_2 and nutrients in the high-latitutde surface oceans- A comparative study., Global Biogeochem. Cy., 7, 843-878, doi: 10.1029/93GB02263, 1993.

Uppström, L. R.: The boron/chlorinity ratio of deep-sea water from the Pacific Ocean, Deep Sea Research and Oceanographic Abstracts, 21, 161-162, 1974.

Van Heuven, S., Pierrot, D., Rae, J. W. B., Lewis, E., and Wallace, D. W. R.: CO2SYS v 1.1, MATLAB program developed for CO_2 system calculations, ORNL/CDIAC-105b. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. DoE, Oak Ridge, TN., 2011.

Wanninkhof, R., Lewis, E., Feely, R. A., and Millero, F. J.: The optimal carbonate dissociation constants for determining surface water pCO_2 from alkalinity and total inorganic carbon, Mar. Chem., 65, 291-301, 1999.

Weiss, R. F., and Price, B. A.: Nitrous-oxide solubility in water and seawater, Mar. Chem., 8, 347-359, doi: 10.1016/0304-4203(80)90024-9, 1980.

Yao, W., Liu, X., and Byrne, R. H.: Impurities in indicators used for spectrophotometric seawater pH measurements: Assessment and remedies, Mar. Chem., 107, 167-172, 2007.

Zhang, H. N., and Byrne, R. H.: Spectrophotometric pH measurements of surface seawater at in-situ conditions: Absorbance and protonation behavior of thymol blue, Mar. Chem., 52, 17-25, 1996.

Table 1. Results of comparisons between direct measurements and values calculated (using the software CO2SYS) from measurements of other variables. r is Pearson's correlation coefficient, df is degrees of freedom, RMSE is root mean square error and MR is mean of the Residuals. The measurement accuracy is from Sect. 2.1. Note that calculated $p\mathrm{CO_2}$-1 and calculated $p\mathrm{CO_2}$-2 have the same uncertainty because they depend only on the accuracies of the variables from which they are calculated. They are therefore identical for both $p\mathrm{CO_2}$ systems Calculated uncertainty of parameters calculated using $p\mathrm{CO_2}$-2 as an input variable is higher than those using $p\mathrm{CO_2}$-1, because the measurement accuracy is higher.

		Comparison statistics				Benchmarks	
Measured variable	Input variables for calculation	r	df	RMSE[*]	MR[*]	Calculated Uncertainty[*]	Measurement accuracy[*]
$p\mathrm{CO_2}$-1 (µatm)	C_T and $\mathrm{pH_T}$	0.984	31	5	3	4	4
	A_T and $\mathrm{pH_T}$	0.990	27	6	2	4	
	C_T and A_T	0.987	43	6	3	4	
$p\mathrm{CO_2}$-2 (µatm)	C_T and $\mathrm{pH_T}$	0.949	97	11	4	4	10
	A_T and $\mathrm{pH_T}$	0.947	93	11	4	4	
	C_T and A_T	0.937	156	12	1	4	
$\mathrm{pH_T}$	C_T and A_T	0.952	218	0.008	0.001	0.005	0.004
	A_T and $p\mathrm{CO_2}$-1	0.991	27	0.006	0.002	0.003	
	A_T and $p\mathrm{CO_2}$-2	0.951	93	0.013	0.004	0.011	
	C_T and $p\mathrm{CO_2}$-1	0.984	31	0.006	0.004	0.004	
	C_T and $p\mathrm{CO_2}$-2	0.943	97	0.014	0.004	0.013	
A_T (µmol kg^{-1})	C_T and $\mathrm{pH_T}$	0.996	218	4.4	-0.2	3.1	1.5
	C_T and $p\mathrm{CO_2}$-1	0.997	43	4.6	2.1	3.0	
	C_T and $p\mathrm{CO_2}$-2	0.991	156	8.6	0.4	7.0	
	$\mathrm{pH_T}$ and $p\mathrm{CO_2}$-1	0.802	27	36.3	-12.8	35.5	
	$\mathrm{pH_T}$ and $p\mathrm{CO_2}$-2	0.596	93	75.6	-23.1	72.8	
C_T (µmol kg^{-1})	A_T and $\mathrm{pH_T}$	0.994	218	4.0	0.2	2.4	2.0
	A_T and $p\mathrm{CO_2}$-1	0.997	43	3.9	-1.7	2.5	
	A_T and $p\mathrm{CO_2}$-2	0.989	156	7.2	-0.3	5.8	
	$\mathrm{pH_T}$ and $p\mathrm{CO_2}$-1	0.680	31	32.2	-18.8	31.9	
	$\mathrm{pH_T}$ and $p\mathrm{CO_2}$-2	0.528	97	69.8	-21.9	63.4	

*These columns have the same units as the first column.

Table 2. Directly measured pH$_{\mathrm{T}}$, partial pressure of CO$_2$ (pCO$_2$-1), dissolved inorganic carbon (C_{T}) and total alkalinity (A_{T}) are compared to values calculated for the combinations of pairs involving A_{T} using different ratios of total boron to salinity. RMSE is root mean square error and MR is mean of the residuals. The two different ratios of total boron to salinity are from Lee et al. (2010) and Uppström (1974). The carbonate constants are from Mehrbach et al. (1973) refitted by Dickson and Millero (1987).

	Ratio of total boron to salinity:	Lee		Uppström	
Measured variable	Calculated variable from	RMSE*	MR*	RMSE*	MR*
pCO$_2$-1 (µatm)	A_{T} and pH$_{\mathrm{T}}$	5	1	6	2
	C_{T} and A_{T}	10	8	6	3
pCO$_2$-2 (µatm)	A_{T} and pH$_{\mathrm{T}}$	11	3	11	4
	C_{T} and A_{T}	14	6	12	1
pH$_{\mathrm{T}}$	C_{T} and A_{T}	0.012	- 0.005	0.008	0.001
	A_{T} and pCO$_2$-1	0.006	0.002	0.006	0.002
	A_{T} and pCO$_2$-2	0.013	0.003	0.013	0.004
A_{T} (µmol kg^{-1})	C_{T} and pH$_{\mathrm{T}}$	6.6	2.9	4.4	-0.2
	C_{T} and pCO$_2$-1	7.3	5.3	4.6	2.1
	C_{T} and pCO$_2$-2	10.7	3.6	8.6	0.4
C_{T} (µmol kg^{-1})	A_{T} and pH$_{\mathrm{T}}$	6.1	-2.7	4.0	0.2
	A_{T} and pCO$_2$-1	6.1	-4.4	3.9	-1.7
	A_{T} and pCO$_2$-2	8.9	-3.0	7.2	-0.3

* These columns have the same units as the first column.

Table 3. Comparison of directly measured pH$_{\mathrm{T}}$, partial pressure of CO$_2$ (pCO$_2$), dissolved inorganic carbon (C_{T}) and total alkalinity (A_{T}) to values calculated using different carbonate constants. RMSE is root mean square error and MR is mean of the residuals. The two

different carbonate constants are from Roy et al. (1993), (Millero et al., 2006), Mehrbach et al. (1973) refitted by Dickson and Millero (1987) and Lueker et al. (2000). The ratio of total boron to salinity is from Uppström (1974).

	Carbonate constants:	Mehrbach		Roy		Millero		Lueker	
Measured variable	Calculated variable from	RMSE[*]	MR[*]	RMSE[*]	MR[*]	RMSE[*]	MR[*]	RMSE[*]	MR[*]
$p\mathrm{CO_2}$-1 (µatm)	C_T and pH$_\mathrm{T}$	5	3	10	11	8	8	6	4
	A_T and pH$_\mathrm{T}$	6	2	9	8	8	7	6	3
	C_T and A_T	6	3	26	25	8	5	6	3
$p\mathrm{CO_2}$-2 (µatm)	C_T and pH$_\mathrm{T}$	11	4	14	11	13	9	12	5
	A_T and pH$_\mathrm{T}$	11	4	14	10	13	9	12	5
	C_T and A_T	12	1	28	23	13	3	12	0
pH$_\mathrm{T}$	C_T and A_T	0.008	0.001	0.022	-0.016	0.007	0.004	0.007	0.002
	A_T and $p\mathrm{CO_2}$-1	0.006	0.002	0.010	0.008	0.009	0.007	0.007	0.003
	C_T and $p\mathrm{CO_2}$-1	0.006	0.004	0.012	0.012	0.009	0.009	0.007	0.005
	A_T and $p\mathrm{CO_2}$-2	0.013	0.004	0.015	0.010	0.015	0.009	0.013	0.005
	C_T and $p\mathrm{CO_2}$-2	0.014	0.004	0.018	0.013	0.016	0.010	0.014	0.006
A_T (µmol kg^{-1})	C_T and pH$_\mathrm{T}$	4.4	-0.2	12.4	9.0	3.9	-1.8	4.0	-1.1
	C_T and $p\mathrm{CO_2}$-1	4.6	2.1	17.8	16.3	5.8	3.5	4.4	1.8
	C_T and $p\mathrm{CO_2}$-2	8.6	0.4	20.0	14.5	9.6	1.8	8.5	0.1
C_T (µmol kg^{-1})	A_T and pH$_\mathrm{T}$	4.0	0.2	11.4	-8.3	3.6	1.6	3.7	1.0
	A_T and $p\mathrm{CO_2}$-1	3.9	-1.7	14.8	-13.6	4.9	-3.0	3.7	-1.5
	A_T and $p\mathrm{CO_2}$-2	7.2	-0.3	16.5	-12.1	8.0	-1.5	7.1	-0.1

[*] These columns have the same units as the first column.

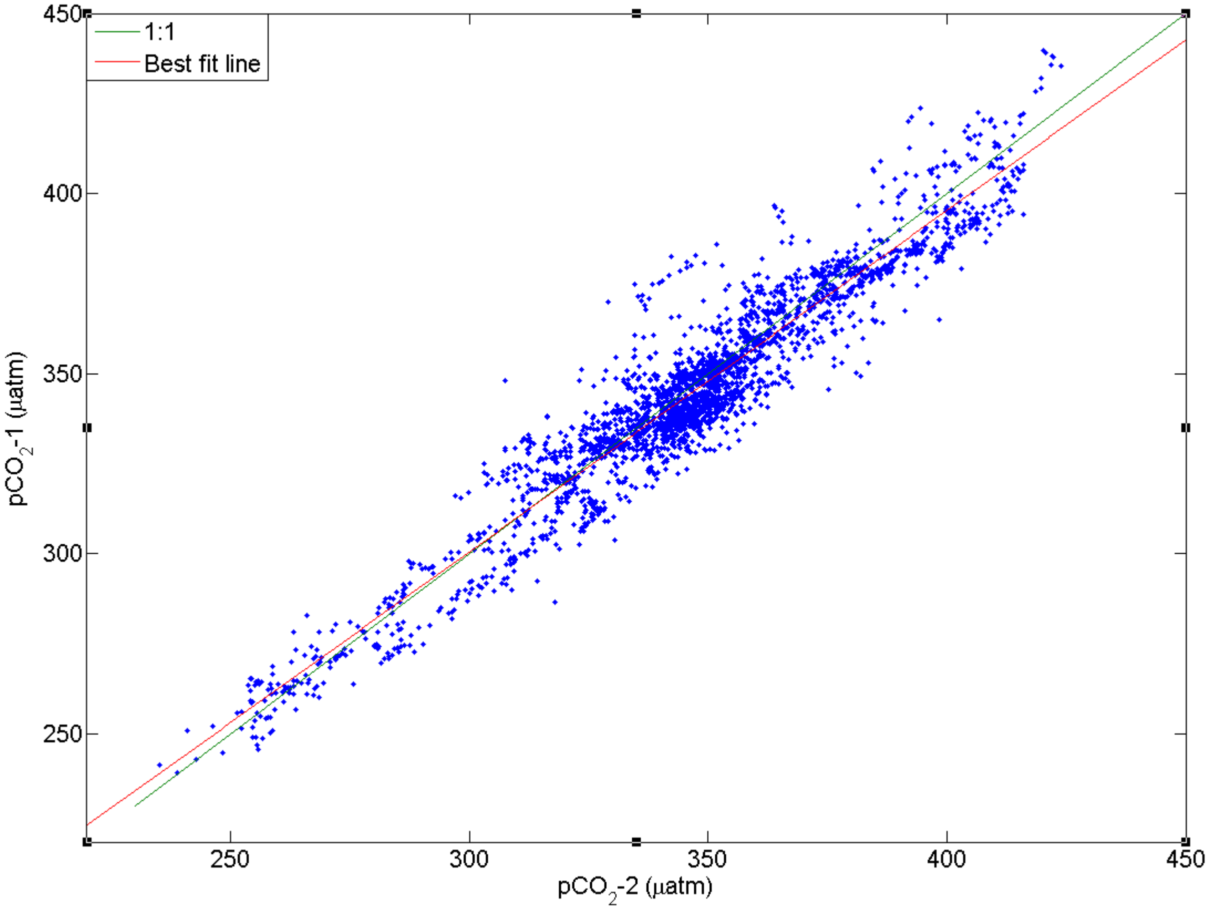

Figure 1. Relationship between two independently measured *p*CO_2 datasets. The line of perfect agreement (1:1 line, in green) and the best fit line (in red) are also shown.

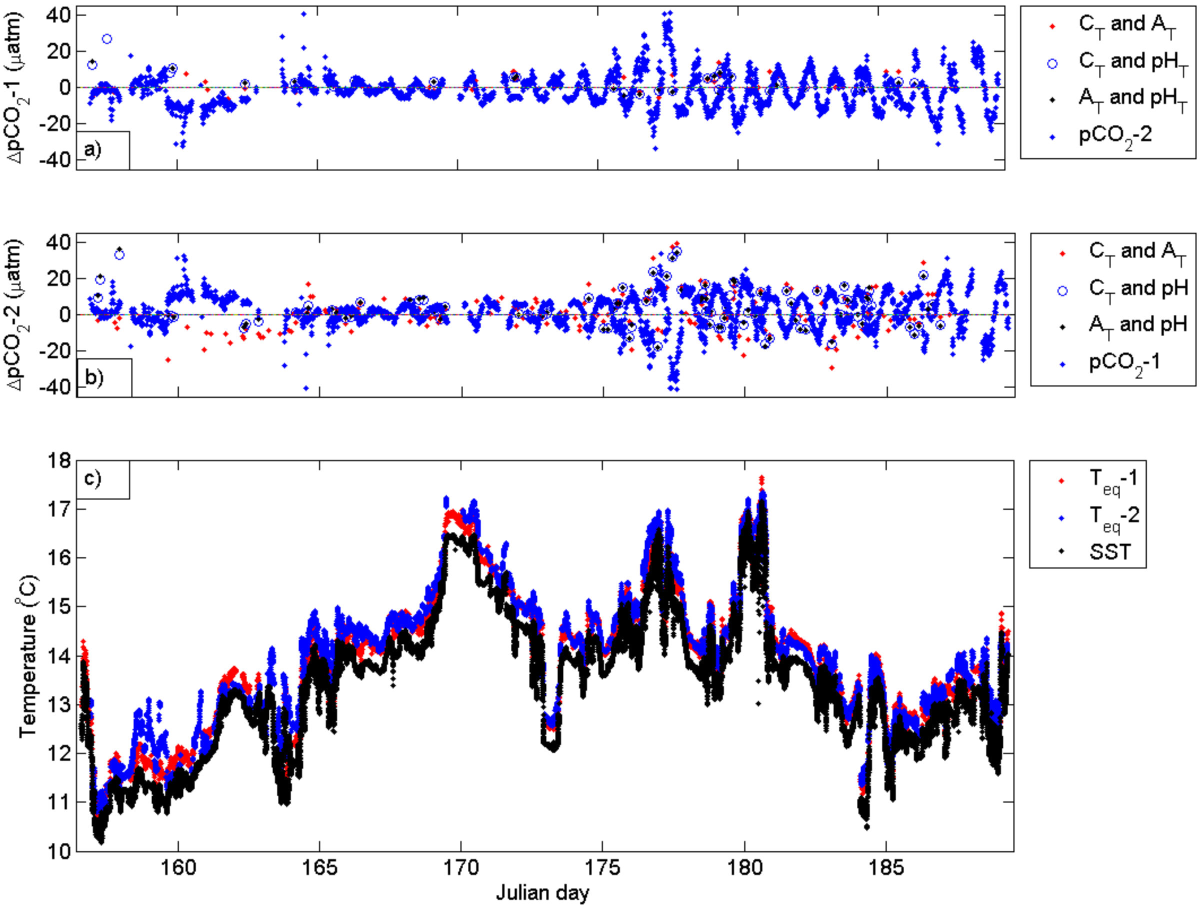

Figure 2. $p\mathrm{CO}_2$ residuals of $p\mathrm{CO}_2$-1 (a) and $p\mathrm{CO}_2$-2 (b) compared to other measured or calculated $p\mathrm{CO}_2$ values. Different symbols show different residuals: against $p\mathrm{CO}_2$ from C_T and A_T (red plus symbols); against $p\mathrm{CO}_2$ from C_T and pH_T (blue circles); against $p\mathrm{CO}_2$ from A_T and pH_T (blue plus symbols); and against $p\mathrm{CO}_2$ from the other measured $p\mathrm{CO}_2$ (2 in a and 1 in b, blue dots). c) In-situ temperature, temperature of equilibrator 1, and temperature of equilibrator 2.

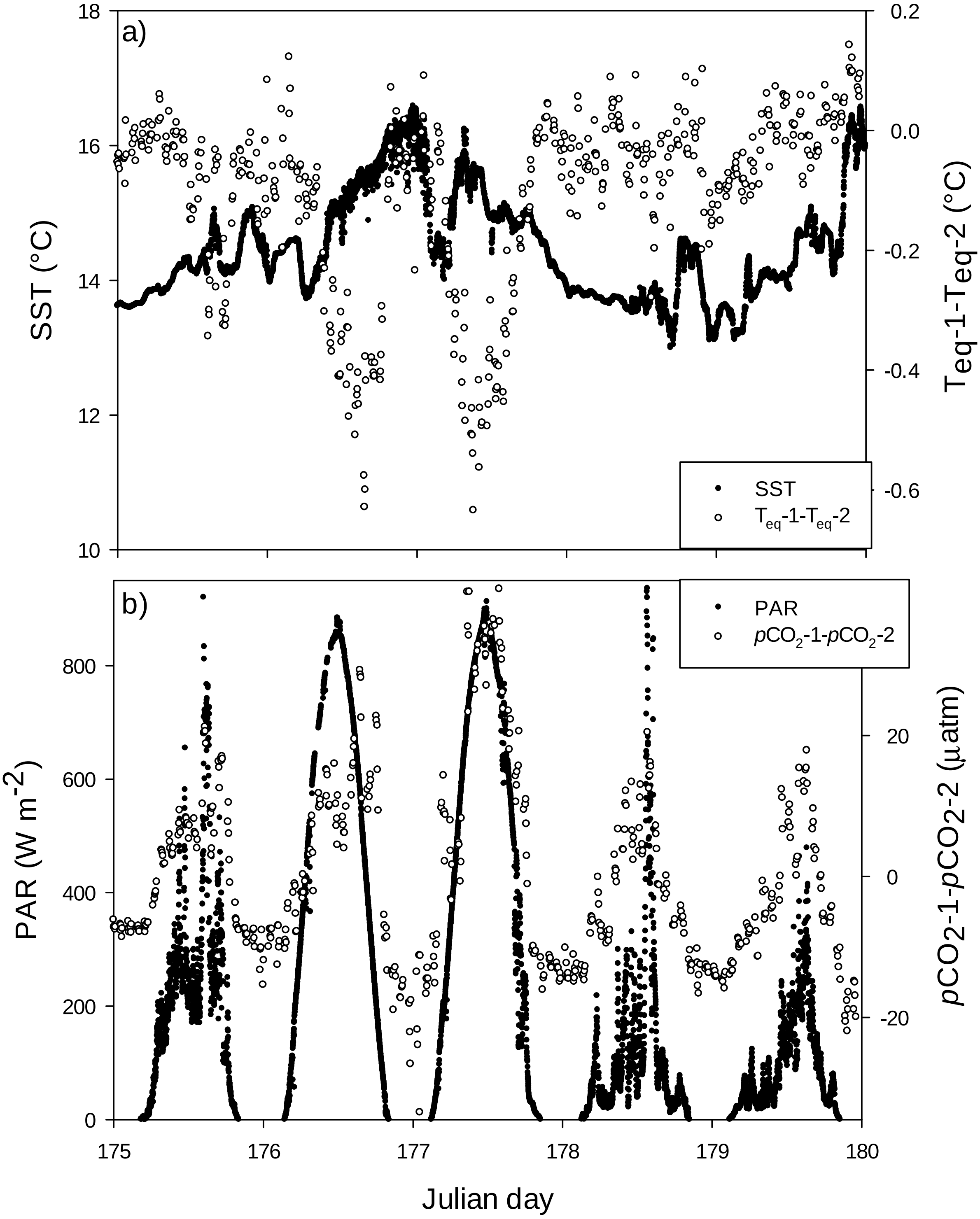

Figure 3. Comparison between a) the in-situ sea surface temperature (SST (°C), in black circles) and the difference in temperatures inside the equilibrators of instruments 1 and 2 (°C; white circles) and b) the phototosynthetically active radiance (PAR (W m^{-2}), in black circles)

and the pCO_2 difference between instruments 1 and 2 (µatm; white circles) measured over five days.

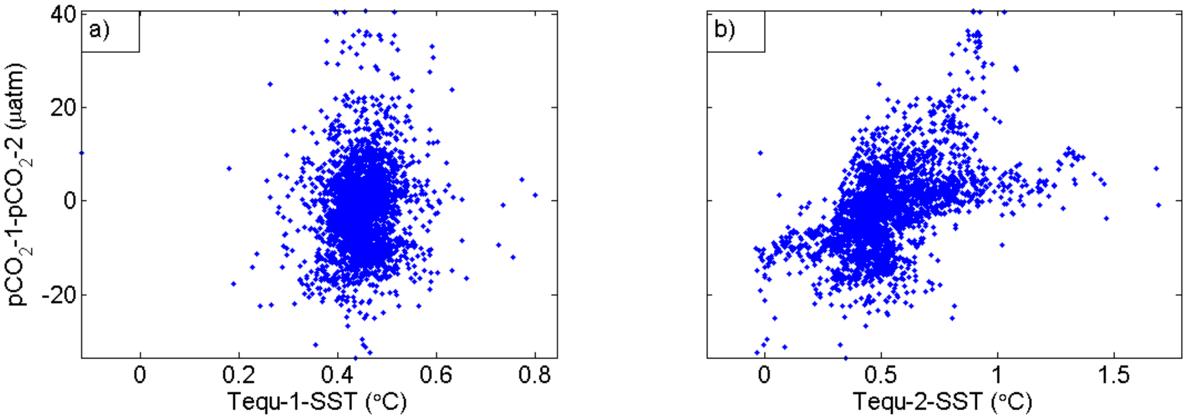

Figure 4. Difference between the two pCO_2 datasets and a) the difference between the temperature inside equilibrator 1 and sea surface temperature, and b) the difference between the temperature inside equilibrator 2 and sea surface temperature.

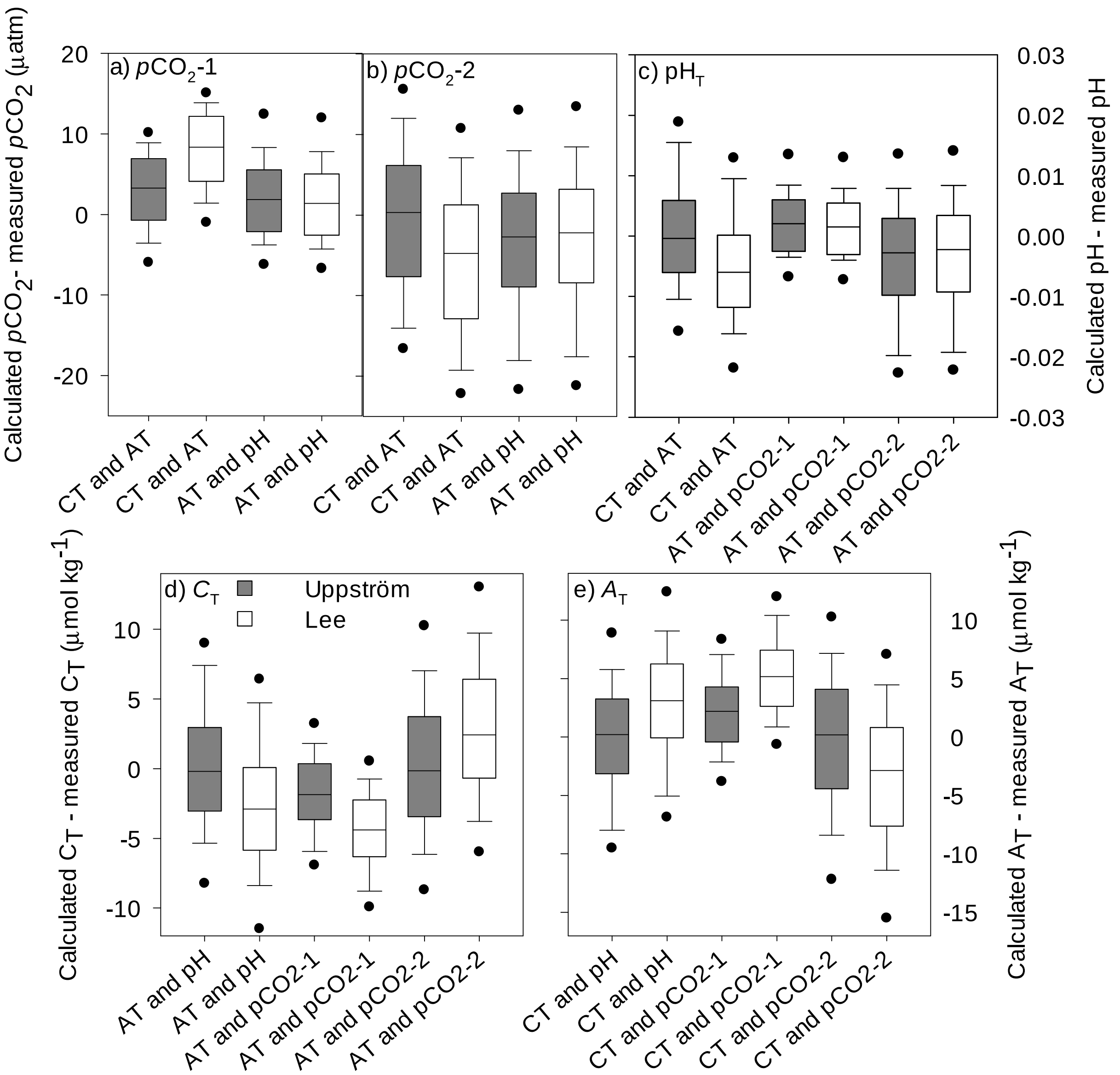

Figure 5. Box-and-whisker plots of the residuals between measured pCO$_2$-1 (a), pCO$_2$-2 (b), pH$_T$ (c), C_T (d), A_T (e) and the respective estimates calculated from different pairs of measured variables (denoted on the x axis) for the two sets of ratios of total boron to salinity. Grey plots depict the distributions of residuals using the constants of Uppström (1974), and white plots denote whose using Lee et al. (2010). The boxes show the median and the 25th and 75th percentiles; dots are the 5th and 95th percentiles. Table 2 presents the RMSE and MR.