Characterization of a novel autonomous analyzer for seawater total alkalinity: Results from laboratory and field tests

Katharina Seelmann [1]*, Steffen Aßmann,[2] Arne Körtzinger[1,3]
[1]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
[2]Kongsberg Maritime Contros GmbH, Kiel, Germany
[3]Christian-Albrechts-Universität zu Kiel, Kiel, Germany

Abstract

High-quality seawater total alkalinity (A_T) measurements are essential for reliable ocean carbon and acidification observations. Well-established manual multipoint potentiometric titration methods already fulfill these requirements. The next step in the improvement of these observations is the increase of the spatial and temporal measuring resolution with minimal personnel and instrumental effort. For this, a rapid, automated underway analyzer meeting the same high requirements as the traditional method is necessary. In this study, we carried out a comprehensive characterization of the flow-through analyzer CONTROS HydroFIA® TA (Kongsberg Maritime Contros GmbH, Kiel, Germany) for automated seawater A_T measurements in the laboratory and in field with overall more than 5000 measurements. Under laboratory conditions, the analyzer featured a precision of ± 1.5 μmol kg^{-1} and an accuracy of ± 1.0 μmol kg^{-1}, combined in an uncertainty of 1.6 – 2.0 μmol kg^{-1}. High precision (± 1.1 μmol kg^{-1}) and accuracy (-0.3 ± 2.8 μmol kg^{-1}), and low uncertainty (2.0 – 2.5 μmol kg^{-1}) were also achieved during field trials of 4 and 6 weeks duration. Although a linear drift appears to be the typical behavior of the system, this can be corrected for by regular reference measurements giving consistent measurement results. Another advantage of regular reference measurements is the early detection of any kind of malfunction due to its direct impact on the measurement performance. Based on the present study, recommendations for automated long-term deployments are provided in order to gain optimal performance characteristics, aiming at the requirements for A_T measurements.

The total alkalinity (A_T) of a seawater sample is defined by Dickson (1981) as "the number of moles of hydrogen ion equivalent to the excess of proton acceptors (bases formed from weak acids with a dissociation constant $K \leq 10^{-4.5}$ at 25°C and zero ionic strength) over proton donors (acids with $K > 10^{-4.5}$) in 1 kg of sample." Therefore, A_T is a measure of the seawater buffering capacity. Together with the other parameters, pH, pCO$_2$, and total dissolved inorganic carbon (C_T), it is one of the four measurable parameters that allow to analytically describe the marine carbonate chemistry using the corresponding thermodynamic relationships (Millero 2007). Therefore, A_T measurements are essential components of ocean carbon observation. However, measuring this

parameter both precisely and accurately is very challenging due to its high background signal (A_T of average seawater ≈ 2300 μmol kg^{-1}) compared to the small natural variability in the open ocean (Millero et al. 1998; Lee et al. 2006) and the required high accuracy for reliable cross calculations. The Guide to Best Practices for Ocean CO$_2$ Measurements (Dickson et al. 2007) describes the most common standard method for measuring A_T based on a manual multipoint potentiometric titration of a seawater sample in an open or closed cell with a strong acid (here, hydrochloric acid). The method described there can achieve a precision (1σ) of better than 1 μmol kg^{-1} and an overall bias of about 2 μmol kg^{-1}. This, however, requires exact weighing of the seawater sample within 0.01 g or a precisely calibrated, thermostatted pipette as a volume-based substitute. Furthermore, the calibration of the pH electrode used for the potentiometric titration must be carried out frequently to ensure proper pH measurements (Millero et al. 1993). Other disadvantages of the traditional method are the relatively long-measurement time per sample (approximately 10–20 min), the need of well-trained technicians in an air-conditioned laboratory, and the fact that the

*Correspondence: kseelmann@geomar.de

measured seawater must be provided as a bottled and typically poisoned discrete sample. This procedure often expands the time period between the seawater sampling during a field campaign and the actual measurement in the laboratory. Furthermore, a potential sampling error can significantly affect the quality of the A_T measurement. Rapid seawater A_T measurements at sea with a simple and robust flow-through analyzer that can also be operated in autonomous mode would overcome most of these challenges. Several authors described different automated flow-through measurement systems for seawater A_T using potentiometric and spectrophotometric pH determination, respectively, with good accordance to the high-quality requirements (e.g., Roche and Millero 1998; Watanabe et al. 2004; Li et al. 2013). But none of these systems have become fully designed, commercially available products. At the time of this study, only the Submersible Autonomous Moored Instrument for alkalinity (SAMI-alk) that was developed and tested by Spaulding et al. (2014) was available as a product for unattended A_T measurements. Its measurement principle is based on a tracer monitored titration approach, introduced by Martz et al. (2006) using a colorimetric tracer for simultaneous pH detection and acid concentration determination.

In this study, we test the CONTROS HydroFIA® TA, a novel commercially available flow-through analyzer for autonomous seawater A_T measurements built by Kongsberg Maritime Contros GmbH. Its general principle is based on open-cell single-point titration with spectrophotometric pH determination. Here, we report the results of a suite of experiments carried out with this novel instrument both in the laboratory and in the field, that is, two major research cruises to the North and South Atlantic Ocean. The goal of this study is to characterize the performance of the analyzer as well as its behavior under laboratory and real field conditions in view of potential long-term deployments. In order to evaluate, whether the measurement quality of the analyzer is suitable for underway A_T measurements in the open ocean, we compare the results with quality targets stated within the oceanographic community's established guides: (1) The "Guide to best practices for ocean CO_2 measurements" by Dickson et al. (2007) provides precision (standard deviation, σ) and accuracy (bias, ΔA_T) requirements for standard open-cell A_T titrators. (2) The "Global Ocean Acidification Observing Network: Requirements and Governance Plan" by Newton et al. (2015) provides uncertainty targets for A_T measurements in order to identify relative spatial patterns and short-term variations ("weather" goal), and to assess long-term trends with a defined level of confidence ("climate" goal), respectively. These targets

are particularly important for the ocean acidification observing community. It must be taken into account that the requirement for the "climate" goal is "only achievable by a very limited number of laboratories and is not typically achievable for all parameters by even the best autonomous sensors" (Newton et al. 2015). The certain targets of both guidelines are outlined in Table 1.

Materials and methods

Measurement principle

The measurement principle of the analyzer is oriented to the open-cell titration as described in the Guide to Best Practices for Ocean CO_2 Measurements (Dickson et al. 2007). Accordingly, a known amount of seawater is titrated with a solution of hydrochloric acid (HCl) to a final pH of about 3.0–3.5. A mixing and degassing procedure allows the escape of all CO_2 deliberated from the sample's dissolved inorganic carbon content. The guide describes a potentiometric pH monitoring of the mixture over the entire titration. Following the definition of total alkalinity (Dickson et al. 2007), A_T at any titration point is given by

$$\frac{-m_{sw} \times A_T + m_A \times C_A}{m_{sw} + m_A} = [H^+]_F + [HF] + [HSO_4^-] + [H_3PO_4] - [HCO_3^-]$$
$$- 2 \times [CO_3^{2-}] - [B(OH)_4^-] - [OH^-] - [HPO_4^{2-}]$$
$$- 2 \times [PO_4^{3-}] - [SiO(OH)_3^-] - [NH_3] - [HS^-]$$
$$(1)$$

where $[H^+]_F$ is the free concentration of hydrogen ions, m_{sw} is the mass of the seawater sample, and m_A is the mass of the added acid with the concentration C_A. Due to the working pH range of 3.0–3.5 and complete CO_2 removal, the majority of the terms in Eq. 1 can be ignored (Dickson et al. 2003). Hence, Eq. 1 can be reduced to

$$\frac{-m_{sw} \times A_T + m_A \times C_A}{m_{sw} + m_A} = [H^+]_F + [HF] + [HSO_4^-] \quad (2)$$

Deviating from the guide, the used analyzer determines the pH spectrophotometrically through a single-point titration of the seawater similar to the A_T measurement principle of Yao and Byrne (1998) and Li et al. (2013). Here, the titrant consists of two separate solutions: An acid (HCl) and an indicator solution of bromocresol green sodium salt (BCG). Based on the definition of Dickson (1981), the added BCG is regarded as a proton donor in the sample-titrant mixture due to its dissociation constant K_I at 25°C being slightly greater than $10^{-4.5}$ (exact definition of K_I later in this section, *see* Eq. 11). Thus, Eq. 2 must be extended by an indicator term

$$\frac{-m_{sw} \times A_T + m_t \times C_t}{m_{sw} + m_t} = [H^+]_F + [HF] + [HSO_4^-] + [HI^-] \quad (3)$$

where $[HI^-]$ is the concentration of the protonated (i.e., acidic) form of BCG, m_t is the sum of the masses of the two titrant

Table 1. Quality targets.

Dickson et al. (2007)	Newton et al. (2015)*
Precision (1σ): ± 1 μmol kg^{-1}	"Weather" goal: u(c) = 10 μmol kg^{-1}
Accuracy (ΔA_T): ± 2 μmol kg^{-1}	"Climate" goal: u(c) = 2 μmol kg^{-1}

*Uncertainties with 68.3% confidence interval.

solutions ($m_t = m_{acid} + m_{indicator}$), and C_t is the acid concentration in the combined titrant solution. Here, it is calculated by:

$$C_t = \frac{C_A \times m_A}{m_t} \tag{4}$$

For a spectrophotometric pH detection using BCG as indicator, the ideal pH range of the sample-titrant mixture is around an absorbance ratio (R) of $R \approx 1$ (further explanation later in this section). This corresponds to a pH range of about 3.5–4.0 and is achieved by adjusting the amount of added acid. Li et al. (2013) show that the reduction of Eq. 1 to Eq. 2 is also valid for this pH range, provided that CO_2 is quantitatively removed. Furthermore, within an autonomous measurement routine, volumes are easier to handle than masses. Hence, Eq. 3 is modified to

$$\frac{-V_{sw} \times \rho_{sw} \times A_T + V_t \times \rho_t \times C_t}{V_{sw} \times \rho_{sw} + V_t \times \rho_t} = [H^+]_F + [HF] + [HSO_4^-] + [HI^-] \tag{5}$$

where V_{sw} and V_t are the volumes of the seawater sample and the added titrant, respectively, and ρ_{sw} and ρ_t are the densities of the seawater sample and the added titrant, respectively. Following Li et al. (2013), a volume mixing ratio γ_v ($\gamma_v = V_{sw}/V_t$), a density ratio γ_ρ ($\gamma_\rho = \rho_{sw}/\rho_t$), and a mass mixing ratio γ ($\gamma = \gamma_v \times \gamma_\rho$) are introduced to simplify the equation:

$$\frac{-\gamma \times A_T + C_t}{1+\gamma} = [H^+]_F + [HF] + [HSO_4^-] + [HI^-] \tag{6}$$

The last three terms in Eq. 6 can be calculated using the dissociation equilibria described in Dickson et al. (2007). An additional rearrangement leads to

$$A_T \times \gamma = C_t - (1+\gamma)$$
$$\times \left([H^+]_F + \frac{S_T \times \frac{\gamma}{1+\gamma} \times [H^+]_F}{[H^+]_F + K_S} + \frac{F_T \times \frac{\gamma}{1+\gamma} \times [H^+]_F}{[H^+]_F + K_F} + \frac{I_T \times \frac{1}{1+\gamma} \times [H^+]_F}{[H^+]_F + K_I} \right) \tag{7}$$

where S_T and F_T are the total sulfate and the total fluoride concentration in the seawater sample, I_T is the total BCG concentration in the titrant, K_S and K_F are the dissociation constants of HF and HSO_4^- and K_I is the second dissociation constant of BCG. The factors $\gamma/(1+\gamma)$ and $1/(1+\gamma)$ represent the dilution factors of the seawater sample and the titrant, respectively. In Eq. 7, all dissociation constants are on the free scale, and concentrations are given in moles per kilogram solution ($mol\,kg^{-1}$). The free concentration of hydrogen ions, $[H^+]_F$, or pH_F, in the sample-titrant mixture is measured spectrophotometrically. Following Breland and Byrne (1993) and Yao and Byrne (1998), pH_F is described by

$$pH_F = -\log K_I + \log\left(\frac{R - e_1}{e_2 - R \times e_3}\right) \tag{8}$$

with

$$e_1 = \frac{\lambda_2 \varepsilon_{HI^-}}{\lambda_1 \varepsilon_{HI^-}} \quad e_2 = \frac{\lambda_2 \varepsilon_{I^{2-}}}{\lambda_1 \varepsilon_{HI^-}} \quad e_3 = \frac{\lambda_1 \varepsilon_{I^{2-}}}{\lambda_1 \varepsilon_{HI^-}} \tag{9}$$

and

$$R = \frac{A_{\lambda_2}}{A_{\lambda_1}} \tag{10}$$

where e_1, e_2, and e_3 represent ratios of absorption coefficients, $\lambda_i \varepsilon_x$ for each indicator form, x, at wavelength 1 (λ_1) and 2 (λ_2), where the acid (HI^-) and the base indicator form (I^{2-}) have their absorbance maxima. R is the absorbance ratio at λ_1 and λ_2. For BCG, the following values are available from the literature: $\lambda_1 = 444$ nm, $\lambda_2 = 616$ nm, $e_1 = 0.0013$, $e_2 = 2.3148$, and $e_3 = 0.1299$; e_1, e_2, and e_3 are considered to be independent of salinity (Breland and Byrne 1993; Yao and Byrne 1998). Breland and Byrne (1993) reported the salinity dependence ($20 \le S \le 35$) of K_I for BCG at 25°C as

$$-\log K_I = 4.4166 + 0.0005946 \times (35 - S) \tag{11}$$

where S is the salinity of the seawater sample. Yao and Byrne (1998) described an advanced salinity range up to 37 for this dependence. Due to the dilution of the seawater sample by the reagents made up in deionized (DI) water, the salinity S of the sample-titrant mixture must be adjusted as follows:

$$i_{sw} = 0.72 \times \left(\frac{S}{35}\right) \tag{12}$$

$$i_{mix} = \frac{i_{sw} \times V_{sw} \times \rho_{sw} + C_A \times V_A \times \rho_A + C_{BCG} \times V_{BCG} \times \rho_{BCG}}{V_{sw} \times \rho_{sw} + V_t \times \rho_t} \tag{13}$$

$$S_{mix} = 35 \times \left(\frac{i_{mix}}{0.72}\right) \tag{14}$$

where i_{sw} and i_{mix} are the ionic strengths of the seawater and the seawater-titrant mixture, respectively, V_{BCG} and ρ_{BCG} are the volume and the density of the added indicator solution with the BCG concentration C_{BCG}, V_A and ρ_A are the volume and the density of the added HCl solution, and S_{mix} is the resulting salinity of the seawater-titrant mixture. This calculation assumes the activity coefficient of BCG to be 1 and that a salinity of 35 corresponds to an ionic strength of 0.72 (Dickson 1990). Hence, the pH of the seawater-titrant mixture at 25°C within the valid salinity range of 20–37 can be calculated as follows:

$$pH_F = 4.4166 + 0.0005946 \times (35 - S_{mix}) + \log\left(\frac{R - 0.0013}{2.3148 - R \times 0.1299}\right) \tag{15}$$

For this calculation, Breland and Byrne (1993) also described the temperature effect on the absorbance measurements between 18°C and 32°C as follows:

$$R_{25} = R_t \times [1 + 0.00907 \times (25 - t)] \qquad (16)$$

where R_{25} and R_t are the absorbance ratios at 25°C and at the exact temperature t (°C) of the sample-titrant mixture, respectively.

Finally, the A_T of a temperature controlled seawater sample with known salinity can be determined by a spectrophotometric pH measurement using Eqs. 7, 15, and 16. Almost all variables in these equations are known or can be calculated. For the present analyzers, the volumes of the added reagents ($V_t = V_{HCl} + V_{BCG}$) are fixed and thus known due to the usage of injections loops with a defined length of tubing (*see* "Instrumental design" section). The densities (ρ_{sw} and ρ_t) at the measured temperatures of the seawater sample with known salinity, and the reagents with known chemical composition were determined using the equations reported in Dickson et al. (2007). The calculation of K_S and K_F is also based on the equations of Dickson et al. (2007) using the salinity of the seawater-titrant mixture; the calculation basis of S_T and F_T is their well-characterized relationship to seawater salinity.

Due to the character of an absolute method, a calibration is principally not necessary. However, the exact volume of the seawater sample V_{sw} is the only unknown variable, which must be practically determined utilizing a one point certified reference material (CRM) measurement. With the known A_T value of the CRM, it is possible to calculate V_{sw} using the same equations as for A_T determination. Consequently, all inevitable uncertainties (e.g., errors in the dissociation constants, the exact concentration of the titrant, impurities of the indicator dye or minor uncertainties in the titrant volume) are combined in V_{sw} and thereby taken into account for subsequent A_T measurements.

Instrumental design

Analyzer setup

For this study, we used two units of the commercially available A_T analyzer CONTROS HydroFIA® TA (Kongsberg Maritime Contros GmbH, Kiel, Germany). For simplification, they are called "red system" and "gray system" in the following sections due to their different housing color. Otherwise, as long as there is no other information provided, the two analyzers can be regarded as being identical. Figure 1 shows the schematic setup of the analyzer showing the involved components in the measurement routine.

The acid and indicator reservoirs are closed and kept in gastight and light-tight bags preventing any alteration of the solutions. Both solutions are separately pumped by piston pumps through injections loops with fixed length and thus fixed volume. These loops are used for injection into the sample circuit using injection valves. The injection valves are connected to the sample circuit in which the solution is pumped by a membrane pump. Depending on the position of the tandem valve, the sample solution is circulated within the sample
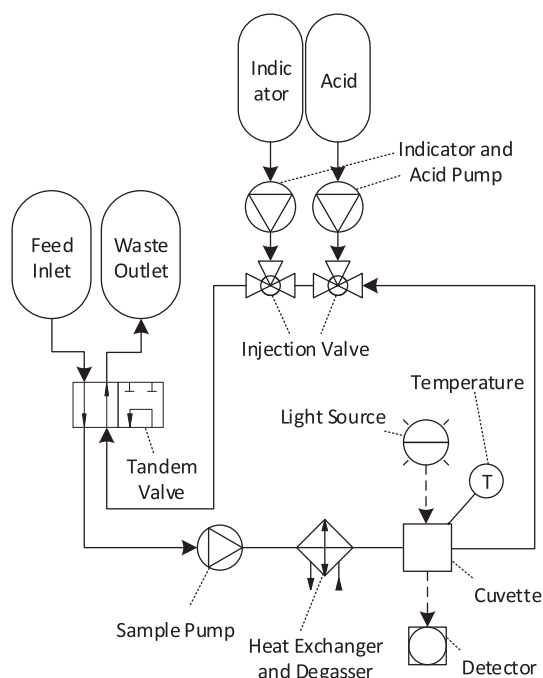
Fig. 1. Schematic diagram of the CONTROS HydroFIA® TA system setup. Solid lines represent fluid paths for sample water, acid, or indicator solution. Dashed lines are light paths for the absorption measurement. Dotted lines are for naming and description. Oval forms represent reservoirs for the respective solutions. The indicator and acid reservoirs are closed while the feed and waste reservoirs are open.

circuit or is pumped through the open circuit to waste. While circulating, the sample is constantly temperature controlled to 25°C by the heat exchanger and at the same time the CO_2 is removed by the degassing unit, which is combined with the heat exchanger (*see* Fig. 2). The temperature control is realized by a Peltier element and temperature measurement directly behind the cuvette. CO_2 removal is done by soda lime behind a membrane. The absorption measurement for the pH detection is realized by means of a flow-through cuvette with 1 cm path length, a broad-band white LED light source and a CCD spectrometer resolving the full absorption spectrum. With this setup, both can be measured simultaneously, the two
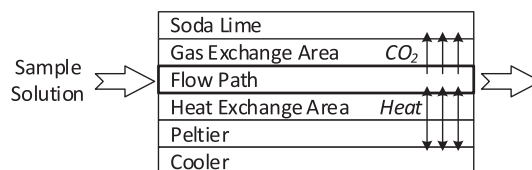
Fig. 2. Schematic diagram of the degassing/heat exchange unit. The sample solution is flowing through the flow path being connected to tubing leading to the other parts of the measurement system (compare Fig. 1). For temperature control of the sample solution, the titanium heat exchange area is used separating the Peltier element from the fluid. For CO_2 stripping, the membrane gas exchange area is used separating the soda lime from the fluid.

absorption maximums of the indicator dye (444 and 616 nm) for pH calculation as well as the nonabsorbing wavelength 730 nm for correction of a potential baseline shift during the measurement routine.

Differences between used systems

For evaluating different working ranges of the analyzers during the second field study (*see* "Experiments" section), each analyzer was equipped with different lengths of acid loop tubing. The red system was equipped with tubing 14.5% longer than the gray system resulting in a lower final pH value after acid addition when measuring at a given A_T. Due to further developments of the analyzer during the course of this study, the red and the gray system were equipped with a modified degasser unit leading into a longer degassing time. The change was done after the first field deployment.

Measurement routine

The A_T measurement routine for each sample is structured as follows: (1) The sample pump flushes all tubing with fresh seawater to remove the residual solution from the previous analysis run. (2) This is followed by a conditioning phase with stopped seawater flow, where the system is conditioned to the pH of seawater to avoid memory effects from the large pH changes during the measurement. (3) Another flush routine collects the seawater sample for the actual measurement (either from a continuous seawater flow or a connected sample bottle), followed by the closing of the sample loop. (4) Now, the sample treatment starts. Dark and blank spectra are measured with the untitrated sample in the cuvette. (5) Both injection loops are filled with HCl solution and BCG solution, respectively, and the reagents are simultaneously injected into the sample loop. (6) This sample-titrant mixture is continuously pumped in the closed loop until completely homogenized. During this, the degasser unit, which is included in the sample loop, removes the CO_2 across a membrane, while constantly controlling the sample temperature to $25.0 \pm 0.1°C$. (7) After equilibration and degassing, which takes about 5 min, the spectra of the CO_2 free and fully temperature controlled sample-titrant mixture are measured in the cuvette, and the A_T of the seawater sample is calculated following the equations in "Measurement principle" section. The whole measurement cycle (maximum measurement frequency) took approximately 7 min during the first field experiment and approximately 10 min with the modified degasser membrane during the second field test.

Seawater sample treatment

Bottled samples (i.e., CRM) were connected to the analyzer with PVC tubing using the same inlet as for underway measurements without any pretreatment. For autonomous underway measurements, the system was installed in bypass to a continuously pumped seawater flow using PVC tubing. Due to the very small tubing diameters of 0.8 mm inner diameters inside the system, the seawater was filtered using a flow-through filter with 50 μm pore size on the first cruise and a cross-flow filter with 0.2 μm pore size on the second cruise. These filters only remove particular matter (e.g., sediment particles) which is important for not having particles dissolved during the sample treatment routine of the analyzer thus altering the A_T measurement result. An adsorption of dissolved organic matter (DOM) onto the filter material interfering with A_T measurement is not given to our knowledge. Furthermore, within the scope of this study, DOM contributions to A_T are not significant in the open ocean (e.g., McElligott et al. 1998; Lee et al. 2000; Millero et al. 2002, Ko et al. 2016).

Solutions and standards

The concentration of the used HCl and BCG solution was 0.1 mol kg^{-1} and 0.002 mol kg^{-1}, respectively. Both reagents were made up in DI water and provided custom-made and in ready-to-use cartridges by Kongsberg Maritime Contros GmbH. The BCG solution was made up by dissolving the sodium salt of BCG. The used BCG was not purified, but the development of a high perfromance liquid chromatography (HPLC) purification method for BCG and testing its impact on the A_T measurement is in progress and will be described elsewhere. CRM (batches 142, 143, 150, and 160) was obtained from A. G. Dickson at the Scripps Institution of Oceanography of the University of California, San Diego. For laboratory experiments, a seawater substandard was prepared out of leftover seawater samples poisoned with mercury chloride. For this, the samples were freshly mixed and the resulting A_T was measured using the reference method (see below).

Reference measurements

For accuracy monitoring, the analyzer measured CRM daily throughout all field campaigns. Every morning and evening, five repetitive CRM measurements were carried out. Furthermore, the measurements of the analyzer were compared to the results from discrete seawater samples measured on a standard open-cell alkalinity system using potentiometric titration (VINDTA 3S, Marianda, Germany). For these measurements, discrete samples were on average taken twice per day throughout all field campaigns and measured in the home lab (GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany) following the recommendations in Dickson et al. (2007).

Statistical calculations

For evaluating the precision of the system both in the laboratory and in field, the standard deviation σ of consecutive measurements of a reference sample (e.g., CRM) is calculated as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \qquad (17)$$

with

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^{n} x_i \qquad (18)$$

where n is the number of measurements, x_i is the ith measurement of n measurements, and $\bar{x}$ is the mean of the measurements.

For evaluating the accuracy in field, the bias ΔA_T between the A_T value of the system and the A_T value of the reference sample (CRM and discrete samples) is calculated as follows:

$$\Delta A_T = A_{T,\text{Analyzer}} - A_{T,\text{Reference}} \qquad (19)$$

The accuracy in the laboratory is determined in a different way and will be explained in the "Results and discussion" section.

In order to compare the measurement quality of the analyzer with the targets of the ocean acidification observing community (Newton et al. 2015), an approximation of the standard uncertainty both in the laboratory and in field is necessary. Due to the usage of CRM for "calibration" and validation of the analyzer, we utilize the within-laboratory validation approach of measurement uncertainty estimation known as "top down" for the first approximation of the measurement uncertainty. The best-known formalization of this approach is the so-called Nordtest™ described by Magnusson et al. (2017), which is based on the guide by Ellison and Williams (2012). Therefore, the combined standard uncertainty u(c) (approximates to a 68.3% confidence interval) is calculated by:

$$u(\text{c}) = \sqrt{u(\text{Rw})^2 + u(\text{bias})^2} \qquad (20)$$

where u(Rw) is the uncertainty estimate of the precision (random effects) and u(bias) is the uncertainty estimate of possible laboratory and procedural bias (systematic effects).

Experiments

Laboratory experiments

Scope

The first part of this study consists of experiments carried out under laboratory conditions. This means that the analyzer did not run for longer than 200 consecutive measurements (equaling approximately up to 24 h) and was set up in an air-conditioned room. Furthermore, the system was shut off overnight between the measurement days. At the start of each measurement day, the analyzer carried out several conditioning measurements to ensure good system stability.

Performance characteristics

Tests on the performance characteristics in the laboratory were carried out as standard addition experiment. Therefore, a stable seawater sample (relatively high A_T) was titrated with a HCl solution (0.1 mol kg^{-1}) to lower its A_T in five steps

(general range of resulting A_T: 2000–2450 μmol kg^{-1}). The titration was carried out by adding different precisely known volumes of HCl to a known volume of seawater resulting in five seawater aliquots with stepwise decreasing A_T. The theoretical A_T ($A_{T,\text{theo}}$) was calculated from the volumes of added acid and seawater, the concentration of the acid, and the original A_T of the seawater. To determine the practical A_T ($A_{T,\text{prac}}$), each of these aliquots was repeatedly measured with the analyzer for five times. This experiment was carried out for each analyzer before and after the cruises.

Overlapping Allan experiment

Regular reference measurements are obligatory for quality assurance and performance monitoring of the analyzer during long-term deployments. To achieve best results, the optimal number of repeated reference measurements with the smallest averaging error had to be determined. For this, we performed a stability estimation by determining the overlapping Allan deviation at different averaging times. To improve the confidence of the stability estimate, we used the overlapping Allan deviation instead of the normal Allan deviation. The overlapping Allan deviation $\sigma_y(\tau)$ makes maximum use of the data set by utilizing all possible combinations of samples at each averaging time τ (Riley 2008). It is estimated by the expression

$$\sigma_y(\tau) = \sqrt{\frac{1}{2 \times \text{AF}^2 \times (n - 2 \times \text{AF} + 1)} \times \sum_{j=1}^{n-2\times\text{AF}+1} \left\{ \sum_{i=j}^{j+\text{AF}-1} [y_{i+\text{AF}} - y_i] \right\}^2}$$
$$(21)$$

where n is the total number of measured samples, τ is the averaging time that is calculated as $\tau = \text{AF} \times \tau_0$, where AF is the averaging factor and τ_0 is the basic measurement interval, and y_i is the ith of n fractional frequency values averaged over τ. In this experiment, n was the total number of A_T measurements ($n = 30$), τ_0 was the measurement interval of the analyzer ($\tau_0 = 10$ min), AF was the number of averaged replicates of the reference measurement, and y represented the A_T values. In an optimal system with only statistical noise, a higher number of averaged replicates, that means longer averaging time, would lead to a higher precision of the measurement. However, due to long-term drift effects on the analyzer, the Allan deviation starts to increase again at some point. The minima in the overlapping Allan plot ($\sigma_y(\tau)$ vs. AF) indicate the optimal number of averaged replicates. For this experiment, a stable seawater substandard was repeatedly measured 30 times in row with a measurement interval of 10 min on four different measurement days.

Field experiments

Scope

In order to test the performance of the analyzer under field conditions, we participated in two major research cruises: RV

Meteor cruise 133 (M 133), from Cape Town, South Africa to Port Stanley, Falkland Islands; 15 December 2016–13 January 2017, and RV *Maria S. Merian cruise 68/2* (MSM 68/2), from Emden, Germany to Mindelo, Cape Verde; 03 November 2017–14 November 2017. In both cases, the analyzer measured continuously pumped surface seawater (underway measurement mode) during the entire cruise with the fastest measurement interval of 7 min during M 133 and 10 min during MSM 68/2, except when separate experiments were carried out. The different intervals were due to the degasser membrane change as mentioned in "Instrumental design" section.

While we only used the red system on cruise M 133, we had the possibility to run both the red and gray system in parallel on cruise MSM 68/2.

At the beginning of both cruises, each analyzer performed several conditioning measurements to ensure system stability. After stabilization, the internal seawater sample volume was determined with a freshly opened CRM.

Working range

Due to the measurement principle of the system, the working range of the analyzer is limited by the pKa value of BCG, its absorption coefficients of the acid and base form and the absorbance ratio R of the sample-titrant mixture, and therefore by the resulting pH of the sample-titrant mixture. The final pH value of a sample with given A_T after acidification can be freely adjusted by the amount of added acid or its concentration to meet the range of seawater A_T in the measured area. The A_T range is limited to seawater with salinities between 20 and 37 as specified by the characterization of BCG (Breland and Byrne 1993; Yao and Byrne 1998). Due to the constant temperature control of the sample water to 25°C, there is only the limitation of the analyzer's temperature controlling capability ranging from 5°C to 30°C for in situ temperatures. To take advantage of two analyzers running in parallel during the MSM 68/2 cruise, the influence of two different acid volumes was tested. For this, each analyzer was equipped with different lengths of acid loop tubing (*see* "Instrumental design" section). The goal of this experiment was to investigate the influence of different pH ranges on the performance of the measurements. This experiment was only carried out on cruise MSM 68/2.

Performance characteristics

The precision under field conditions was evaluated by measuring CRM on both cruises. Additionally, during the cruise M 133, a long-term precision experiment was conducted. For this, a stable seawater substandard was prepared and measured 178 times consecutively with a measurement interval of 7 min.

The accuracy evaluation was carried out by comparing the measurements of the analyzer with both the certified values of the CRM (twice per day throughout both cruises), and the A_T values of taken discrete samples (on average twice per day throughout both cruises) measured with the reference system VINDTA 3S.

Initial drift after idle time

For longer idle times (≥ 1 d), it is recommended to flush the analyzer with DI water to avoid any deposits inside the tubing, for example, from the last colored and acidified sample. These idle times could be necessary, for example, during harbor time between field campaigns. Harbor seawater often is very dirty and should not be run through the system. Since the analyzer is flushed with DI water, the system is conditioned to low ionic strength, causing an extended stabilization phase (initial drift) when measuring again after these idle times. To examine the extent of such a drift, the system was flushed with DI water and did not measure any sample for 48 h, except for the very first measurements at the beginning of the cruise. Therefore, the initial idle time matched the storage and transportation time of the analyzer before the cruise (≈ 3 months). Afterward, a seawater substandard taken during the cruise was measured until the measurements were stable (standard deviation of the last three measurements ≤ 2 μmol kg^{-1}). This 48-h idling experiment was carried out three times during the whole cruise M 133: at the beginning, after 1304 measurements, and after 2183 measurements.

Results and discussion

Laboratory experiments

Performance characteristics

The comparison and discussion of the laboratory performance before and after a field deployment is most useful for an analyzer without any hardware problems during this deployment. Consequently, only the results for the red system before and after the MSM 68/2 cruise are shown and interpreted in the following part as the gray system suffered from a leakage in the degasser unit (*see* "General information" section later in this study). Figure 3 shows the results of the standard addition experiment observed with the red system before and after the MSM 68/2 cruise. For accuracy evaluation in the laboratory, the root mean square error (RMSE) of the measured A_T values was determined as follows:

$$\text{RMSE} = \pm \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} \left(A_{T,\text{fitted},i} - A_{T,\text{prac},i}\right)^2} \qquad (22)$$

where n is the number of titration steps, $A_{T,\text{fitted},i}$ is the ith A_T value calculated with the linear regression equation with $A_{T,\text{theo},i}$ as x variable. Before the campaign, the RMSE was determined with ± 5.5 μmol kg^{-1}; afterward, it is improved to ± 1.0 μmol kg^{-1}. This big difference is due to a change of the experiment procedure. Before the cruise, the titrated seawater samples were manually changed and each measurement was started by hand. Afterward, a more optimized experiment procedure was applied using an autosampler for these purposes. This custom-made autosampler is part of the system calibration setup at the Kongsberg Maritime Contros GmbH laboratory and is used for routine calibrations automatically changing sample solutions of defined A_T
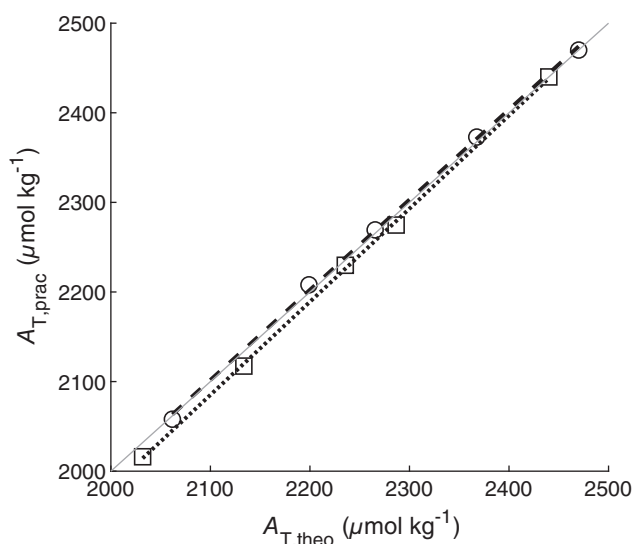
Fig. 3. $A_{T,prac}$ as a function of $A_{T,theo}$ of each titration step measured with the red system before (open squares) and after (open circles) the MSM 68/2 cruise in the laboratory. $A_{T,prac}$ is the average of five repeated measurements for each aliquot. The black dashed and dotted line indicates the linear fit of the data points with the following resulting linear equations: Before: $y = (1.03 \pm 0.01) \times x - (97.6 \pm 28.4)$, $R^2 = 0.999$, and after: $y = (1.01 \pm 0.02) \times x - (11.4 \pm 40.6)$, $R^2 = 0.999$. The gray solid line indicates the 1-to-1 line of this plot.

levels. By using this autosampler, uncertainties, caused by the operator, are partly removed resulting in better performance of the experiment itself. Furthermore, due to the worse slope and the large intercept of the linear regression of the data set before the cruise, it is possible that some unknown additional errors occurred during the experiment. However, the after-cruise evaluation shows a very satisfactory correlation between $A_{T,prac}$ and $A_{T,theo}$ with a slope of 1.01 ± 0.02 and an intercept of -11.4 ± 40.6 which are as expected (slope = 1, intercept = 0) within their found uncertainty. Furthermore, its laboratory accuracy of $\pm 1.0 \, \mu\text{mol kg}^{-1}$ is in full agreement with the requirements of Dickson et al. (2007) for the standard A_T titration methods for which an accuracy of $\pm 2 \, \mu\text{mol kg}^{-1}$ is required.

The precision in the laboratory is determined by using the standard deviation of the five single measurements at each titration step. General precision for each titration step was found to be approximately $\pm 2 \, \mu\text{mol kg}^{-1}$ for this analyzer (data not shown). This general precision also agrees with precision values determined by Kongsberg Maritime Contros GmbH for any HydroFIA analyzer. The explained laboratory performance characteristic is a standard procedure at Kongsberg Maritime Contros GmbH for each CONTROS HydroFIA® TA system and is carried out regularly. A performance characteristic test carried out with the gray system after the MSM 68/2 cruise and maintenance of the manufacturer (no leakage) shows an overall precision of $\pm 1.5 \, \mu\text{mol kg}^{-1}$. Because both analyzers are treated similarly in the laboratory, and the modified method with autosampler is

stable, robust and part of the quality management system of the company, it is possible to generalize this precision for all laboratory performance characteristics.

For estimating the combined standard uncertainty of this experiment in the laboratory (only shown for after-cruise experiment), the laboratory precision was utilized as random uncertainty component and the RMSE of the measured A_T values as systematic uncertainty component, respectively. Both components were estimated with a freshly "calibrated" analyzer using CRM. The relative combined laboratory standard uncertainty was estimated at 0.08%, which results in a combined laboratory standard uncertainty of 1.6–2.0 $\mu\text{mol kg}^{-1}$ at A_T values from 2000 to 2450 $\mu\text{mol kg}^{-1}$ (*see* Supporting Information for more details of the calculations). This laboratory uncertainty approximation is in full agreement with the "weather" goal requirements of Newton et al. (2015). Even the very high requirement of the "climate" goal is achieved. Thus, the laboratory standard uncertainty of the analyzer is sufficient for ocean acidification measurements.

Overlapping Allan experiment

Figure 4 shows the results of the overlapping Allan analysis with each curve representing one specific measurement day. As expected, each Allan plot shows a minimum representing the optimal number of averaging replicate measurements. The minima range from $A = 3$–6 with resulting overlapping Allan deviations $\sigma_y(\tau)$ of 0.5–1.0 $\mu\text{mol kg}^{-1}$, each determined at $A_T \approx 2270 \, \mu\text{mol kg}^{-1}$ (pH$_{\text{sample-titrant mixture}} \approx 3.6$). This means that a reference sample used for performance monitoring should be repeatedly measured at least three times to minimize the impact of statistical noise. On the other hand, more than six
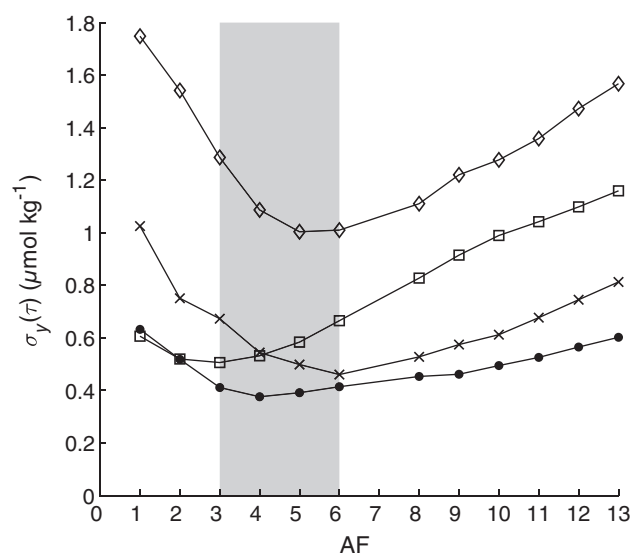
Fig. 4. Overlapping Allan deviations $\sigma_y(\tau)$ as a function of the number of averaged replicates of the reference measurement, AF. Each curve represents one measurement day. The gray area indicates the range of the observed minima in the experiments.

repetitions lead into a regime affected by instrument drift or changes of environment/sample solution and causes the precision to deteriorate. The ideal number of repeated reference measurements depends on different factors: The available volume of stable reference seawater during the deployment, the length of the deployment, and the required number of quality assurance measurements per day. In addition, possible outliers within these reference measurements should be taken into account. For the performance monitoring during the two research cruises, we decided to repeatedly measure the reference samples (here, CRM) five times every morning and evening.

Field experiments

General information

During the MSM 68/2 cruise, the red analyzer ran without any hardware problems. Consequently, its performance characterization is discussed in detail in the following parts. However, because of the early development level of the system, both the M 133 analyzer as well as the gray analyzer during the MSM 68/2 cruise suffered from a leakage in the degasser unit. The effects of such a malfunction on the performance are briefly discussed afterward.

Underway measurements

To give an overview of the measured underway variables (A_T, sea surface salinity [SSS] and sea surface temperature [SST]) in the monitored regions, Fig. 5a,b shows their time series over the course of each cruise (note: shown A_T values are corrected). The red filled circles in the A_T time series (Fig. 5a) represent the discrete samples taken during each cruise. Figure 5c illustrates the track of the M 133 and MSM 68/2 cruise, respectively. The scientific interpretation of these underway data is not part of this report as the focus here lies on the assessment of instrument performance under typical field deployment conditions. However, to get an rough idea of the consistency between the underway A_T values measured by the analyzer and the A_T range and variability in the measured region, we compared the corrected A_T data sets to calculated A_T values based on the parameterization described by Lee et al. (2006). This calculation utilizes SSS and SST data. The consistency is estimated by calculating the RMSE of the $A_{T,Analyzer}$ and $A_{T,Calculated}$ following Eq. 20. A plot of the comparison is shown in Supporting Information Fig. S1. An RMSE of ± 12.7 μmol kg^{-1} and ± 4.9 μmol kg^{-1} was calculated for the M 133 cruise (South Atlantic Ocean) and the MSM 68/2
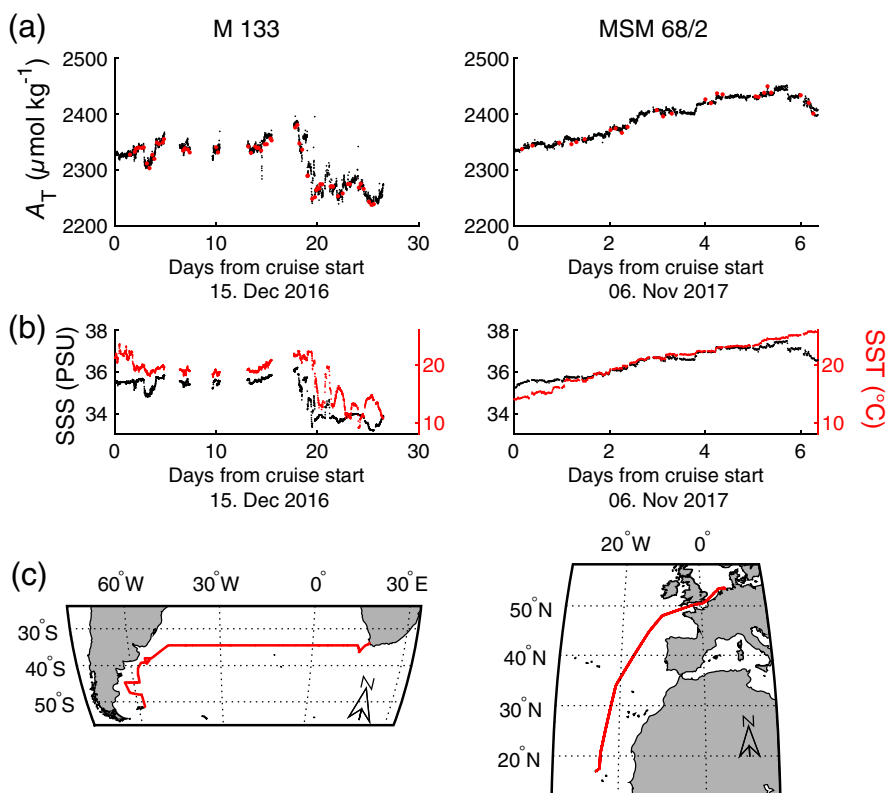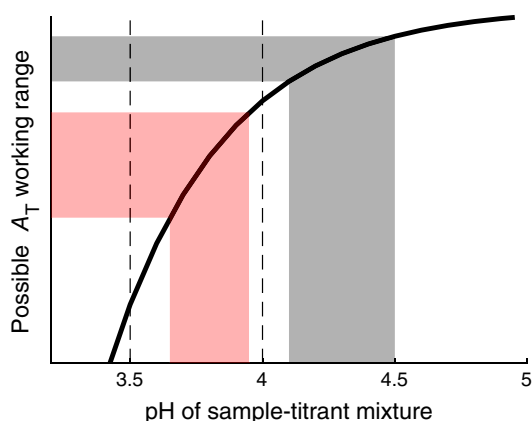

Fig. 5. (**a**) Time-series of the measured A_T values during the M 133 and MSM 68/2 cruise, respectively. Each filled red circle represents the A_T value of a discrete sample taken at a certain time during the cruise. (**b**) Time-series of the measured SSS with the left *y*-axis and SST with the right *y*-axis during the M 133 and MSM 68/2 cruise, respectively, where the black filled circles represent SSS and the red filled circles represent SST. (**c**) Track of M 133 cruise (start: Cape Town, South Africa; end: Stanley, Falkland Islands) and MSM 68/2 cruise (start: Emden, Germany; end: Mindelo, Cape Verde).

cruise (North Atlantic Ocean), respectively. The error of the parameterization is $\pm\,8.4$ μmol kg^{-1} and $\pm\,6.4$ μmol kg^{-1} for the North and South Atlantic Ocean, respectively. By taking these errors into account, both field data sets seem to be consistent with the A_T range and variability in the measured region, which is also proved by the comparison to discrete samples (*see* "Performance characteristics" section).

Working range

Figure 6 illustrates the possible A_T working range as a function of the pH of the sample-titrant mixture observed with analyzers using spectrophotometric pH measurements with BCG. In this figure, the A_T working ranges of the red and the gray system are based on the resulting pH working range observed with the MSM 68/2 underway measurements (pH$_{min}$ to pH$_{max}$, *see* Fig. 7). On the one hand, the configuration of the red system yields in a wider measurement range, where small pH steps correspond to larger A_T steps, which makes it more attractive for regions with high A_T variability. On the other hand, the gray system is more precise due to larger pH steps corresponding to smaller A_T steps. To take advantage of this better precision, the choice of a higher pH range (4.0–4.5) is more useful for regions, where small A_T changes are expected.

Another potential problem with pH ranges above 4.0 is the unknown validity of the A_T calculation following Eq. 7 (*see* "Materials and methods" section). Li et al. (2013) only validate this calculation within pH values of 3.5–4.0. In this range, the concentrations of the terms $[H_3PO_4]$, $[B(OH)_4^-]$, $[OH^-]$, $[HPO_4^{2-}]$, $[PO_4^{3-}]$, $[SiO(OH)_3^-]$, $[NH_3]$, and $[HS^-]$ can be neglected. Following the calculation procedure of Li et al. (2013), recalculation of these concentrations at higher pH values results in an overall increase of the corresponding alkalinity contributions from approximately 0.4 μmol kg^{-1} at pH 4.0 up to approximately 0.9 μmol kg^{-1} at pH 4.5, and, consequently,
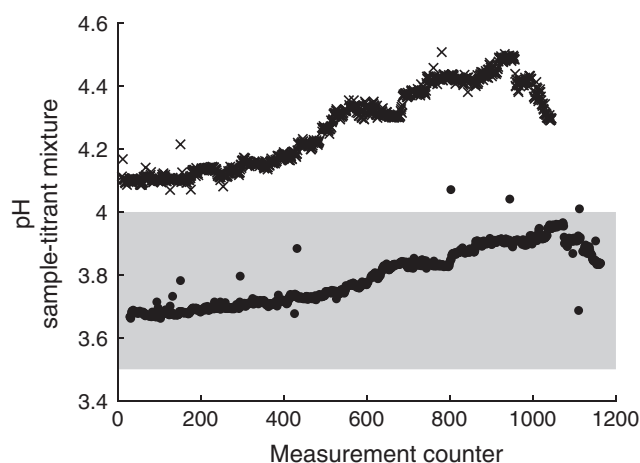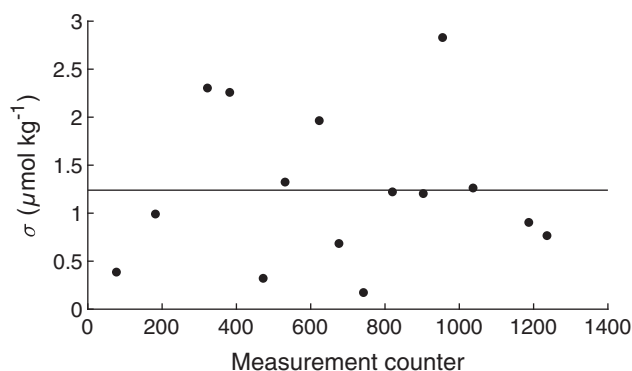

Fig. 7. pH of the sample-titrant mixture (after degassing) of the MSM 68/2 underway measurements as a function of the measurement counter, where filled circles and crosses represent the red and the gray system, respectively. The gray area indicates the optimal pH range of 3.5–4.0 given by Breland and Byrne (1993) and Li et al. (2013).

in an increase in the systematic error of the method. However, it has to be noted that the total concentrations of these species, on which the calculation of Li et al. (2013) is based, are much higher than those of typical open ocean seawater (worst-case scenario). Based on this fact, it can be assumed that the concentrations of the above listed terms can still be neglected within the given instrument uncertainties, thus Eq. 7 is also valid up to a pH of 4.5. Additionally, laboratory performance tests similar to that in "Laboratory experiments" section with another CONTROS HydroFIA® TA system (similar configurations to the gray system, but no leakage) support this finding. There, a linear slope of (1.01 ± 0.01), and (1.000 ± 0.003) measured with a maximum pH of 4.6, and 4.3, respectively, was observable (data not shown), which indicates unbiased performance of the system. Just an increase of the RMSE with increasing pH is detectable (ΔRMSE = 1.8 μmol kg^{-1}). The found decrease in accuracy at those high pH values can be explained by the limit of the spectrophotometer's ability determining the very low concentrations of the remaining indicator acid. Because of that, it is not recommended to measure A_T at pH values greater than 4.6.

Performance characteristics

In this part, only the performance characteristic of the red analyzer during the MSM 68/2 cruise is shown and discussed. Because it had no malfunctions during its deployment, these results are representative for the behavior of the system as such.

For evaluating precision, the standard deviation σ (n = 5) of each repeated CRM measurement was determined. Figure 8 shows these standard deviations as a function of the measurement counter. The averaged field precision is determined at $\pm\,1.2$ μmol kg^{-1}. For an autonomous system, such a level of

Fig. 6. Possible A_T working range as a function of the sample-titrant mixture pH. The red and the gray area indicate the measured A_T range of the red and the gray system, respectively, during the cruise MSM 68/2. The vertical dashed lines indicate the optimal pH range of 3.5–4.0 given by Breland and Byrne (1993) and Li et al. (2013).

Fig. 8. Standard deviation σ of repeated CRM measurements as a function of the measurement counter of the red analyzer during the cruise MSM 68/2. The horizontal black line indicates the averaged standard deviation.

precision is in good agreement with the requirements of Dickson et al. (2007) as stated for standard A_T titration methods for which a standard deviation of better than $\pm\,1\;\mu$mol kg^{-1} is required.

For evaluating accuracy, the bias ($\Delta A_T = A_{T,\text{Analyzer}} - A_{T,\text{Reference}}$) between the analyzer A_T measurements and the A_T values of the reference samples is calculated. Figure 9 shows the biases of the red system as function of the measurement counter both with raw data (Fig. 9a) and with linearly drift corrected data (Fig. 9b). This correction is necessary, because the analyzer shows a linearly increasing ΔA_T from $-2\;\mu$mol kg^{-1} at the beginning up to $+15\;\mu$mol kg^{-1} at the end of the cruise. The conducted CRM measurements are used for this drift correction resulting in a mean bias of $-0.3 \pm 2.8\;\mu$mol kg^{-1} (n = 28) between the red system and the reference data from discrete samples measured by standard open-cell titrator (including the sampling error). Such a level of accuracy is comparable to standard A_T titration methods. Furthermore, by plotting the A_T values of the analyzer (already corrected against the drift using the CRM measurements) against the reference A_T values, the linear function $y = (0.98 \pm 0.01) \times x + (40 \pm 23)$ (R^2 = 0.997) results. Taking all uncertainties in account, this result proofs the good sensitivity of the analyzer over the whole working range.

We discovered that the observed linear drift with increasing measurement number occurs because of material deposits in the optical pathway. As a result, the light intensity decreases and therefore the absorbance A at 444 and 616 nm changes. Usually, such an intensity loss is corrected by the dark and blank spectrum within the measurement routine, but the optical measurements showed systematic deviations over time at the different wavelengths. Figure 10 shows the absorbance changes at 444 and 616 nm of the red system's CRM measurements in dependence of the measurement counter during the MSM 68/2 cruise, and also the resulting change in the absorbance ratio R. Theoretically, the absorbance ratio of a CRM measurement should not change, provided the same batch is
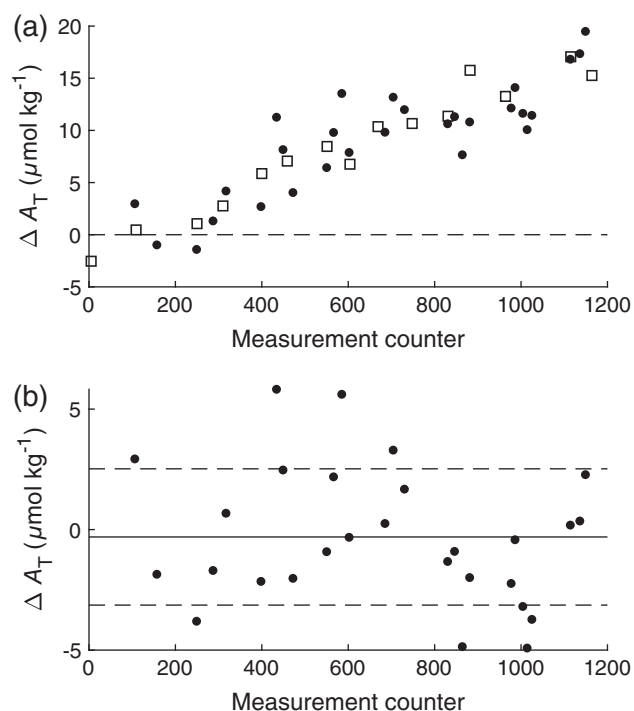
Fig. 9. (**a**) Bias plot for the intercomparison of A_T measurements between the red analyzer and the reference samples during the cruise MSM 68/2, where open squares represent the CRM and filled circles represent the discrete samples measured with the standard open-cell titrator in the laboratory. The horizontal dashed line indicates $\Delta A_T = 0\;\mu$mol kg^{-1}. (**b**) Bias plot (discrete samples) after a linear drift correction using the CRM measurements. The horizontal black line indicates the mean bias, $\Delta \bar{A}_T$, while the dashed lines indicate $\Delta \bar{A}_T \pm \sigma$ (for this data set: $-0.3 \pm 2.8\;\mu$mol kg^{-1}).

measured. In reality, Fig. 10c proves that this is not the case. R increases over time due to the different behavior of the absorbances at 444 and 616 nm, and an increasing R leads to increasing A_T values. We hypothesize that the deposits are caused by colored substances, possibly by a decomposition of the BCG indicator or impurities, which cannot be completely prevented at the moment. Consequently, the observed linear drift toward higher A_T values has to be accepted as typical behavior. Fortunately, the drift can be easily corrected for by regular reference measurements. Even by applying one reference measurement in the beginning and one in the end of a deployment in our case would have been sufficient because of the linear character of the drift.

Recent results from other deployments show a similar drift magnitude but over a much longer time period with lower measurement interval of 90 min. Consequently, a small residual deposit is left on the optical window after each measurement leading to accumulation over time as a function of number of measurements. This indicates that this pattern is related to the number of conducted measurements (number of indicator injections) rather than the pure deployment time.
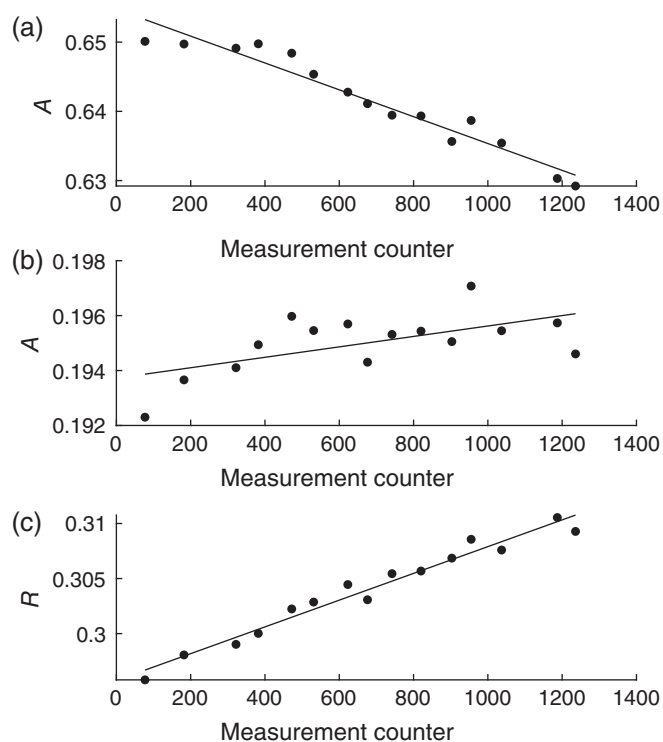
Fig. 10. Absorbance A at the wavelengths of the indicator absorption maxima (**a**) 444 nm, and (**b**) 616 nm of the CRM measurements with the red system as a function of the measurement counter during the MSM 68/2 cruise. (**c**) Absorbance ratio R ($R = A_{616nm}/A_{444nm}$) of these CRM measurements as a function of the measurement counter.

For approximating the combined standard uncertainty in field, the daily CRM measurements were used. The precision of each repeated CRM measurement was utilized as random uncertainty component, and the root mean square of the biases to the certified value of the measured CRM ($\Delta A_{T} = A_{T,CRM} - A_{T,Analyzer}$) as systematic uncertainty component, respectively. Additionally, the uncertainty contribution of the drift correction to the systematic uncertainty was estimated and implemented by using the RMSE of the linear regression. The relative combined field standard uncertainty of the analyzer was estimated with 0.10% at 2212.44 μmol kg^{-1} (certified value of CRM Batch No. 160). Due to the proven linearity of the analyzer over the working range of 2000–2450 μmol kg^{-1}, the combined field standard uncertainty is estimated with 2.0–2.5 μmol kg^{-1} (*see* Supporting Information for more information). This field uncertainty approximation is in full agreement with the "weather" goal requirements of Newton et al. (2015). The very high requirement of the "climate" goal is almost be achieved in the field by being only 0.5 μmol kg^{-1} higher than the target of 2 μmol kg^{-1}. Thus, the field standard uncertainty of the analyzer is sufficient for ocean acidification measurements.

Long-term precision

The result of the long-term precision experiment during the M 133 cruise is shown in Fig. 11. The standard deviation of the long-term measurements is determined with $\pm$ 2.4 μmol kg^{-1}
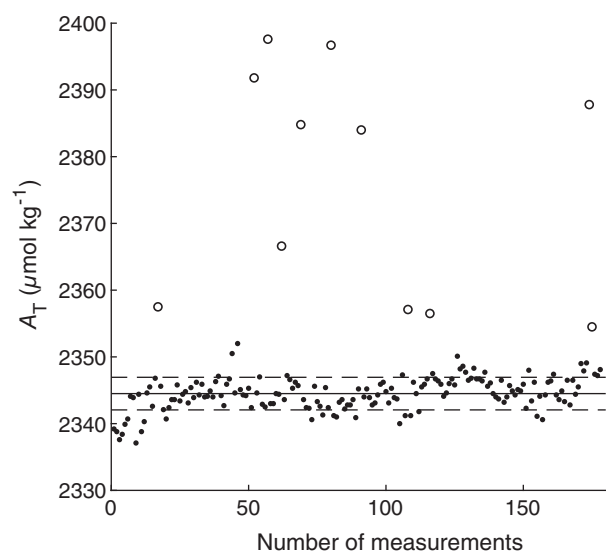

Fig. 11. A_{T} long-term measurements of a stable seawater substandard as a function of the measurement counter of the red system during the cruise M 133, where filled circles indicate the regular measurements, and open circles the outliers in this data set (Grubbs outlier test, $\alpha = 0.05$, two-sided, $n = 178$). The horizontal black line represents the resulting mean A_{T} value, $\bar{A}_{T}$, without outliers. The dashed black lines indicate $\bar{A}_{T} \pm \sigma$.

($n = 178$), which is higher than the averaged short-term precision of $\pm$ 1.2 μmol kg^{-1} observed with the red analyzer during the MSM 68/2 campaign. This reduced precision is due to the instrument long-term drift, and changes of the environmental or sample conditions, that was already reported in the overlapping Allan experiment results. Unfortunately, it has to be mentioned that this experiment was carried out after the leakage in the degasser unit appeared. A functional analyzer would probably show better results. However, while long-term standard deviation of $\pm$ 2.4 μmol kg^{-1} does not reach the requirements for standard open-cell titrators, it still is in an acceptable range. Another outcome of this experiment is the appearance of outliers. Overall, 11 outliers are recognizable in the data set, which is 6.2% of the measurements. This outlier rate seems to be relatively high, but, especially during long-term deployments, the measurement resolution of the CONTROS Hydro-FIA® TA system (measurement time per sample: < 10 min) is high enough to compensate these outliers. Additionally, removing these outliers using an algorithm (e.g., Grubbs outlier test by Grubbs 1974) is very well possible during the postprocessing, since appear as spike outliers in the regular data set. The reasons for the occurrence of these outliers are still unclear. We hypothesize that the sample pump supplies a minimally higher volume of seawater to the sample loop than usual. In addition, bubbles in the sample loop could be possible.

Initial drift after idle time

Figure 12 shows the results of the initial drift experiment during the M 133 cruise. After the very first start of the system following an idle time of approximately 3 months, the required
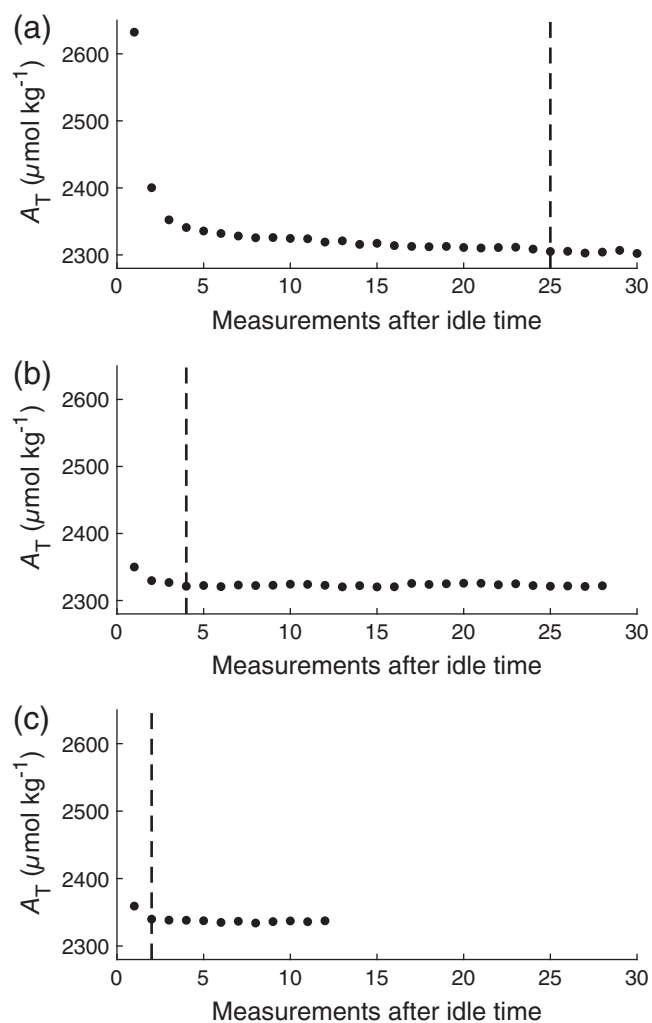
Fig. 12. Measured A_T of a stable substandard as a function of the number of measurements after a DI water flush followed by an idle time of (**a**) $\approx$ 3 months, and (**b**, **c**) 48 h, respectively. The vertical black dashed lines indicate the number of measurements, after which the stability criterion of standard deviation of ≤ 2 μmol kg^{-1} is reached. The experiment was carried out (**a**) at the beginning of the M 133 cruise, (**b**) after 1304 consecutive measurements, and (**c**) after 2183 consecutive measurements, respectively.

standard deviation of ≤ 2 μmol kg^{-1} is reached after 25 measurements. But the longer the analyzer has been operated continuously, the faster stable measurements are reached. After 2183 measurements followed by an idling time of 48 h, the analyzer only needs two measurements to reach stable A_T values. Therefore, relatively short idle times (around 48 h) with a DI water flush during a long-term deployment have a negligible effect on the measurements afterward, provided the analyzer runs constantly between these short idle times.

Observed failure modes

At the time of this study, a leakage in the degasser unit was the only malfunction of the CONTROS HydroFIA® TA system

during its early development phase. The following discussion includes all precision and accuracy results both of the M 133 cruise, and of the gray system during the MSM 68/2 cruise.

Figure 13 shows the precision evaluation of both the red analyzer during the cruise M 133, and the gray analyzer during the cruise MSM 68/2. Table 2 summarizes the results of the precision evaluation. It is obvious that the averaged precision does not differ significantly much to the functional analyzer. However, a systematically increasing spread around the averaged σ over time is observable. This phenomenon is caused by an evolving dead volume within the degasser unit, which has an increasing effect on the random error of the measurement.
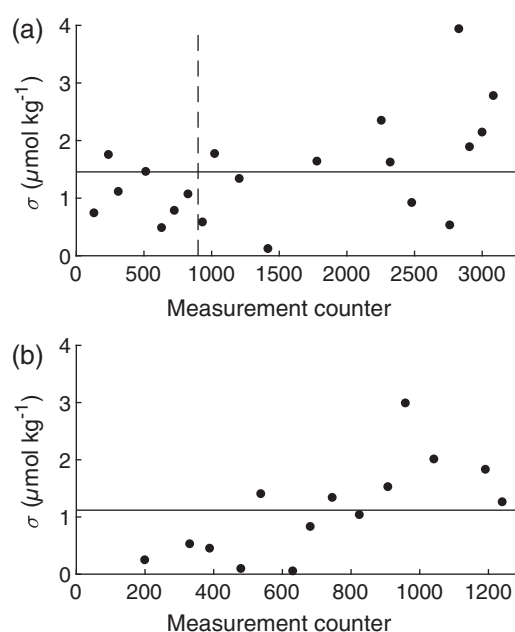
Fig. 13. Standard deviation σ of repeated CRM measurements as a function of the measurement counter of (**a**) the red analyzer during the cruise M 133 and (**b**) the gray analyzer during the cruise MSM 68/2. In both plots, the horizontal black line indicates the averaged standard deviation: (**a**) ± 1.5 μmol kg^{-1} and (**b**) ± 1.1 μmol kg^{-1}. In plot (**a**), the vertical dashed line indicates the appearance of the leakage.

Table 2. Results of the precision evaluation.

|	Min σ–Max σ (μmol kg^{-1})	Mean σ (μmol kg^{-1})
M 133 cruise		
Red system	± 0.5–1.8	± 1.0
Red system*,†	± 0.1–3.9	± 1.8
MSM 68/2 cruise		
Red system	± 0.2–2.8	± 1.2
Gray system†	± 0.1–3.0	± 1.1

*Leakage occurred during the course of the cruise.
†With leakage.

The effects of a leaking degasser on the accuracy of the system are shown in Fig. 14a,b. Different to the typical behavior, as shown in Fig. 9a, there is a downward drift observable. Furthermore, as seen in Fig. 14b, a discrepancy between $\Delta A_{T,CRM}$ and $\Delta A_{T,\text{discrete samples}}$ could occur. By plotting the A_T values of the analyzer (already corrected against the drift using the CRM measurements) against the reference A_T values, the linear function $y = (0.913 \pm 0.005) \times x + (193.4 \pm 13.4)$ ($R^2 = 0.999$) results. Because the slope is significantly smaller than 1.000, there must be a sensitivity loss induced by the leaking degasser. Both phenomena, the downward drift and the sensitivity problem, are caused by the loss of degasser functionality. Additionally, an analyzer, that runs with a leaking degasser for a longer time (approximately > 2000 measurements), could show unpredictable effects like the increasing drift in Fig. 14a after 2500 measurements. The reasons for these effects are still unclear. However, the measured A_T values are still correctable in a way similar to the correction explained in the

"Performance characteristics" section. Deviating from this correction, both the CRM and the discrete sample measurements must be utilized. Figure 14c,d shows the resulting biases of the analyzer A_T values after such a correction to the A_T values measured with a standard open-cell titrator. The mean bias of the corrected M 133 data is 0.2 ± 7.8 μmol kg^{-1}, and 0.2 ± 1.6 μmol kg^{-1} for the corrected MSM 68/2 data observed with the gray system. Obviously, the correction of the data observed with a system, which runs with a malfunction for a longer time (red analyzer during M 133 cruise), results in less accurate values. For the MSM 68/2 campaign, both analyzers show comparable results after correction. Although it has be taken into account, that the correction of the leaking gray analyzer had to rely on discrete samples measured with a reference technique in addition to the regular CRM measurements performed by the analyzer itself.

Summing up, we can say that the effects of a leakage can occur in different unpredictable ways that do not necessarily
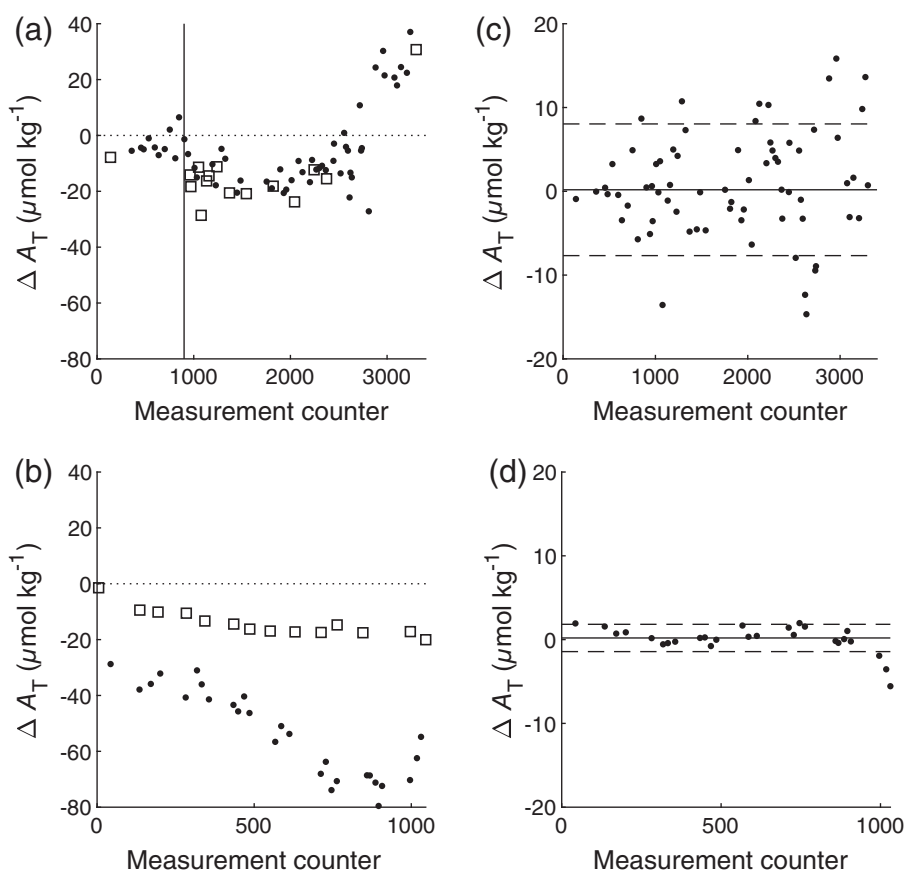
Fig. 14. Results obtained with leaking analyzers: Bias plot for the intercomparison of A_T measurements between the analyzer and the reference samples (**a**) during the cruise M 133 and (**b**) during the cruise MSM 68/2 (gray system), where open squares represent the CRM and filled circles represent the discrete samples measured with the standard open-cell titrator. The horizontal dotted line indicates $\Delta A_T = 0$ μmol kg^{-1}, and the vertical solid line in (**a**) indicates the approximate appearance of the leakage. Bias plots (only discrete samples shown) after correction using both the CRM and discrete sample measurements for (**c**) during the cruise M 133 and (**d**) during the cruise MSM 68/2 (gray system). The horizontal black line indicates the mean bias, $\Delta \bar{A}_T$, while the dashed lines indicate $\Delta \bar{A}_T \pm \sigma$. Here: (**c**) 0.2 ± 7.8 μmol kg^{-1} and (**d**) 0.2 ± 1.6 μmol kg^{-1}.

appear at the same time. Thus, if the system shows a different performance behavior within the quality assurance routine other than the typical bias drift to higher A_T values, a leaking degasser unit might be the reason. However, after the experiments and campaigns carried out in this study, the degassing unit of the analyzer was revised and improved by the manufacturer solving the leakage problem (for further information, *see* "Outlook" section).

Conclusion

The performance tests of the commercially available autonomous analyzer for total alkalinity CONTROS HydroFIA® TA reveal several important features relevant for the future field application of this system.

Table 3 summarizes the laboratory and field performance results obtained with the analyzer. While the system reaches the accuracy requirement for standard open-cell titrators provided by Dickson et al. (2007) both in the laboratory and in field, the precision requirement cannot be completely met. However, for an autonomous analyzer with spectrophotometric pH determination, such a level of precision is still in favorable comparison to the standard A_T titration methods. Furthermore, uncertainty approximations both in the laboratory and in field are in full agreement with the "weather" goal requirements by Newton et al. (2015) for ocean acidification observations. The very high requirement of the "climate" goal is achieved in the laboratory and almost be achieved in the field by being only 0.5 μmol kg^{-1} higher than the target of 2 μmol kg^{-1}. Another important outcome is that the analyzer appears to show a linear drift (offset drift) caused by so far unavoidable colored deposits in the optical pathway. Currently, this drift has to be accepted as typical behavior and therefore must be corrected for by measurements of Seawater CRM with known A_T values or by regular reference samples to reach the required accuracy. Consequently, a regular quality assurance routine has to be implemented for long-term deployments. This routine should contain regular CRM measurements

each consisting of 3–6 repetitive measurements, but if deemed sufficient, it could be lowered to one pre- and one postdeployment CRM measurement due the linear character of the drift. A stability estimation in the laboratory utilizing the overlapping Allan deviation plot showed that fewer than three or more than six replicates are not recommendable due to the effects of statistical noise and long-term drift effects on the analyzer, respectively. Another major advantage of regular quality assurance measurements is the detection (with backward tracking and correction possibility) of malfunctions without the need to perform manual functionality checkups of single components. Fortunately, a leakage always shows a more or less abrupt change within the precision and/or accuracy evaluation of the analyzer. In addition, one of the following observations can point at a malfunction within the system: (1) Higher spread of the standard deviation around the averaged value, (2) abrupt change in the accuracy or offset drift, (3) direction change of the offset drift, and/or (4) discrepancy between CRM and discrete samples biases (loss of sensitivity). Furthermore, it has to be taken into account that not all leakage effects appear necessarily at the same time. Figure 15 summarizes the behavior of the instrument being within the quality assurance routine during long-term deployments.

Experiments dealing with different pH ranges of the system show that its good performance is still maintained at pH values > 4.0, additionally with higher precision. However, the accuracy worsens at pH values between 4.3 and 4.5 (Δ = 1.8 μmol kg^{-1}).

Table 3. Results of the laboratory and field performance tests.

	Precision σ (μmol kg^{-1})	Accuracy ΔA_T (μmol kg^{-1})	Uncertainty u (c)[¶] (μmol kg^{-1})
Laboratory	±1.5	±1.0[*]	1.6–2.0[‖]
Field	±1.1	−0.3 ± 2.8[†]	2.0–2.5[†,‖]
With leakage	±1.3	0.2 ± 7.9[†]	—
		0.2 ± 1.6[†]	—
Required	±1[‡]	±2[‡]	10/2[§]

[*]RMSE of the linear regression curve.
[†]After drift correction against reference measurements.
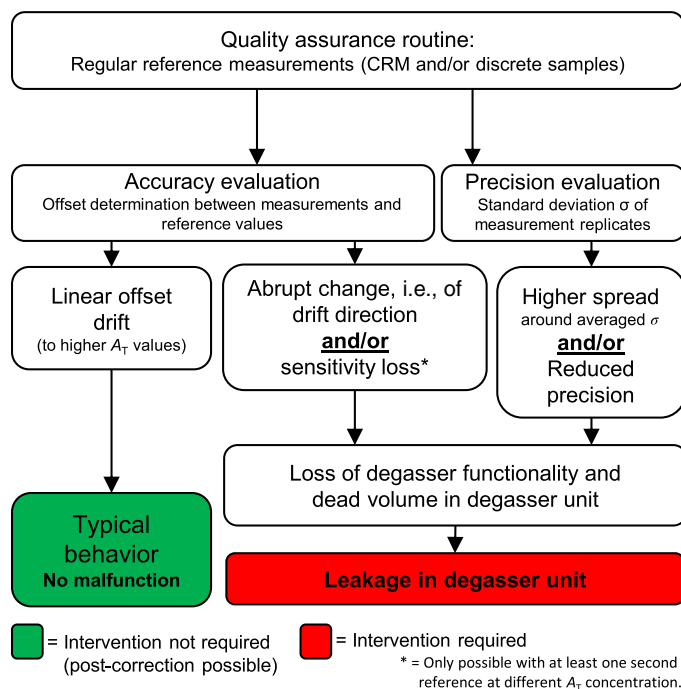[‡]For standard open-cell titrator (Dickson et al. 2007).
[§]"Weather"/"Climate" goal (Newton et al. 2015).
[‖]Working range of 2000–2450 μmol kg^{-1}.
[¶]With 68.3% confidence interval.

Fig. 15. Overview of the behavior of the CONTROS HydroFIA® TA system within the quality assurance routine during a long-term deployment.

The use of such high pH ranges might be useful for regions, where small A_T changes must be detected. Due to the detection limit of the spectrophotometer, it is not recommended to measure at pH values > 4.5.

Another important outcome of this study is that the stability of a continuously running analyzer (i.e., an analyzer running continuously for more than 2000 measurements without any idle times of $\gg$ 48 h) is not affected by short idle times (up to about 48 h) with previous DI water flush.

A long-term precision experiment reveals, additionally to a long-term precision of $\pm$ 2.4 μmol kg^{-1}, how often outliers could occur during a measurement campaign. Approximately 6.2% of the measurements are outliers showing always clearly higher A_T values than the regular data. Consequently, an automated detection and removal routine is very well possible during the postprocessing. Furthermore, the high measurement resolution of the analyzer compensates the loss of data caused by outliers.

In summary, the commercial autonomous A_T analyzer CONTROS HydroFIA® TA is suitable for autonomous underway measurements of the marine carbonate system and for ocean acidification observations. In comparison to traditional A_T measurement methods, it provides similar qualitative measurement results, provided that regular reference measurements are carried out for drift correction and, at the same time, monitoring the functionality of the system. The knowledge gained from this study forms the basis for defining such best practices for automated long-term operations with this system.

Outlook

Recently, the degasser unit of the analyzer, a frequent cause for malfunctions in early versions of the instrument, was revised and improved by the manufacturer. Its newly developed membrane is more robust, and test measurements in the laboratory confirm the high resistance against leakages. A long-term deployment of the system with this new degasser unit in field is ongoing. Furthermore, this new membrane is more robust against organic solvents, allowing to flush the system with isopropanol to remove the colored deposits in the optical pathway. Future work dealing with possible improvements within the measurement routine will possibly overcome the drifting behavior of the analyzer. So, one option could be a regular cleaning procedure with isopropanol (only possible with the improved degasser membrane) to reduce the influence of the material deposits in the optical pathway and eliminating the observed bias. Another option may be the use of purified BCG indicator to possibly minimize its decomposition in solution. Using purified BCG could also improve the spectrophotometric measurement, similar to meta-cresol purple for spectrophotometric seawater pH measurements (Yao et al. 2007; Liu et al. 2011). For this, a preparative HPLC purification method is under development. Another crucial task

during automated long-term deployments is the provision of enough reference seawater for regular quality assurance measurements. A standard CRM bottle of 500 mL, as provided by A. G. Dickson, is not sufficient for an automated long-term deployment. Due to the autonomous character of such deployments, changing the standard bottles by hand is also no option. Hence, an alternative stable larger volume storage (minimum 5 L) for standard seawater must be found. For this purpose, we are currently testing several types of containers, such as gas sampling bags, infusion bags, different canisters, or bottles.

Recommendations for automated long-term deployments

Due to the fact, that the CONTROS HydroFIA® TA system is a commercially available analyzer, it is already being used by the oceanographic community. But not all of these users have the time or resources to fully characterize and test the system for their purposes. Based on our experiences and the present study, we provide recommendations for automated long-term deployments of this analyzer:

- After very long idle times ($\gg$ 48 h, e.g., after storage and/or transportation of the system), flush the system with 0.1 mol L^{-1} HCl solution to shorten stabilization phase (initial drift) before starting measurements.
- At the beginning of a deployment, carry out stabilization measurements. Stabilization measurements should be carried out with stable seawater substandard. The absolute A_T value is not important as long as it is in the working range of the analyzer. The measurements are considered stable when a standard deviation of $\leq$ 2 μmol kg^{-1} of the last three measurements is reached.
- After stabilization, "calibrate" (sample volume determination) the system always with a freshly opened CRM standard.
- Pumped underway seawater must be filtered, for example, using a cross-flow filter before running it through the analyzer. Particles or particulate matter must be avoided in the measured sample water.
- Carry out regular quality assurance measurements (e.g., with CRM or any other suitable seawater standard). For this, 3–6 replicates are recommended. These quality assurance routine is important for the drift correction of the data during the postprocessing. The frequency depends on the lengths of the deployment. For shorter deployments (< 20 d), daily measurements are recommended (based on our experiences during field campaigns). The more reference data are collected, the better the drift correction of the system is. For longer campaigns, they can be reduced to every 2 or 3 d.
- Evaluate the quality assurance measurements on regular basis. These can be used to identify malfunctions (e.g., leakage in the degasser unit) without the need to

manually inspecting the system. For identifying such malfunctions, use Fig. 15 as a guidance.

- In case quality assurance measurements indicate a problem of the analyzer functionality, increase the frequency of standard measurements to verify this.

- If there is a leakage in the degasser unit, stop the deployment as soon as possible for instrument maintenance by the manufacturer. The longer a leaking system runs, the more difficult the data postcorrection becomes. The measured A_T values lose plausibility because of increasingly unpredictable effects caused by the leakage.

- Higher pH ranges (4.0–4.5) of the acidified sample may be used for regions with small A_T variability to detect small changes more precisely. The pH range can be adjusted by changing the length of the acid loop tubing (only by manufacturer) or by adjusting the acid concentration (consultation of manufacturer should be taken; postprocessing needed due to changed parameters).

References

Breland, J. A., and R. H. Byrne. 1993. Spectrophotometric procedures for determination of sea water alkalinity using bromocresol green. Deep-Sea Res. Part I Oceanogr. Res. Pap. **40**: 629–641. doi:10.1016/0967-0637(93)90149-W

Dickson, A. G. 1981. An exact definition of total alkalinity and a procedure for the estimation of alkalinity and total inorganic carbon from titration data. Deep-Sea Res. A **28**: 609–623. doi:10.1016/0198-0149(81)90121-7

Dickson, A. G. 1990. Standard potential of the reaction: AgCl(s) + $^1/_2$H$_2$(g) = Ag(s) + HCl(aq), and the standard acidity constant of the ion HSO$_4^-$ in synthetic sea water from 273.15 to 318.15 K. J. Chem. Thermodyn. **22**: 113–127. doi:10.1016/0021-9614(90)90074-z

Dickson, A. G., J. D. Afghan, and G. C. Anderson. 2003. Reference materials for oceanic CO_2 analysis: A method for the certification of total alkalinity. Mar. Chem. **80**: 185–197. doi:10.1016/S0304-4203(02)00133-0

Dickson, A. G., C. L. Sabine, and J. R. Christian. 2007. Guide to best practices for ocean CO_2 measurements. PICES Special Publications 3. North Pacific Marine Science Organization.

Ellison, S. L. R., and A. Williams [eds.]. 2012. Eurachem/CITAC guide: Quantifying uncertainty in analytical measurement, 3rd Edition. Eurachem.

Grubbs, F. E. 1974. Procedure for detecting outlying observations in samples. Technometrics **11**: 53. doi:10.1080/00401706.1969.10490657

Ko, Y. H., K. Lee, K. H. Eom, and I. S. Han. 2016. Organic alkalinity produced by phytoplankton and its effect on the computation of ocean carbon parameters. Limnol. Oceanogr. **61**: 1462–1471. doi:10.1002/lno.10309

Lee, K., F. J. Millero, R. H. Byrne, R. A. Feely, and R. Wanninkhof. 2000. The recommended dissociation constants for carbonic acid in seawater. Geophys. Res. Lett. **27**: 229–232. doi:10.1029/1999GL002345

Lee, K., and others. 2006. Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. Geophys. Res. Lett. **33**: L19605. doi:10.1029/2006GL027207

Li, Q., F. Wang, Z. A. Wang, D. Yuan, M. Dai, J. Chen, J. Dai, and K. A. Hoering. 2013. Automated spectrophotometric analyzer for rapid single-point titration of seawater total alkalinity. Environ. Sci. Technol. **47**: 11139–11146. doi:10.1021/es402421a

Liu, X., M. C. Patsavas, and R. H. Byrne. 2011. Purification and characterization of meta-cresol purple for spectrophotometric seawater pH measurements. Environ. Sci. Technol. **45**: 4862–4868. doi:10.1021/es200665d

Magnusson, B., T. Näykki, H. Hovind, M. Krysell, and E. Sahlin. 2017. Handbook for calculation of measurement uncertainty in environmental laboratories. Nordtest Report TR 537 (ed. 4). Available from www.nordtest.info. Last accessed on May 13, 2019.

Martz, T. R., A. G. Dickson, and M. D. DeGrandpre. 2006. Tracer monitored titrations: Measurement of total alkalinity. Anal. Chem. **78**: 1817–1826. doi:10.1021/ac0516133

McElligott, S., R. Byrne, K. Lee, R. Wanninkhof, F. Millero, and R. Feely. 1998. 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. doi:10.1016/S0304-4203(97)00080-7

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

Millero, F. J., J. Zhang, K. Lee, and D. M. Campbell. 1993. Titration alkalinity of seawater. Mar. Chem. **44**: 153–165. doi:10.1016/0304-4203(93)90200-8

Millero, F. J., K. Lee, and M. Roche. 1998. Distribution of alkalinity in the surface waters of the major oceans. Mar. Chem. **60**: 111–130. doi:10.1016/S0304-4203(97)00084-4

Millero, F. J., D. Pierrot, K. Lee, R. Wanninkhof, R. Feely, C. L. Sabine, R. M. Key, and T. Takahashi. 2002. Dissociation constants for carbonic acid determined from field measurements. Deep-Sea Res. Part I Oceanogr. Res. Pap. **49**: 1705–1723. doi:10.1016/S0967-0637(02)00093-6

Newton, J. A., R. A. Feely, E. B. Jewett, P. Williamson, and J. Mathis. 2015. Global ocean acidification observing network: Requirements and governance plan, 2nd Edition. GOA-ON, Available from http://www.goa-on.org/docs/GOA-ON_plan_print.pdf. Last accessed on May 13, 2019.

Riley, W. J. 2008. Handbook of frequency stability analysis. NIST Special Publication 1065. NIST.

Roche, M. P., and F. J. Millero. 1998. Measurement of total alkalinity of surface waters using a continuous flowing spectrophotometric technique. Mar. Chem. **60**: 85–94. doi:10.1016/S0304-4203(97)00087-X

Spaulding, R. S., M. D. DeGrandpre, J. C. Beck, R. D. Hart, B. Peterson, E. H. De Carlo, P. S. Drupp, and T. R. Hammar. 2014. Autonomous in situ measurements of seawater alkalinity. Environ. Sci. Technol. **48**: 9573–9581. doi:10.1021/es501615x

Watanabe, A., and others. 2004. A rapid, precise potentiometric determination of total alkalinity in seawater by a newly developed flow-through analyzer designed for coastal regions. Mar. Chem **85**: 75–87. doi:10.1016/j.marchem.2003.09.004

Yao, W., and R. H. Byrne. 1998. Simplified seawater alkalinity analysis: Use of linear array spectrometers. Deep-Sea Res. Part I Oceanogr. Res. Pap. **45**: 1383–1392. doi:10.1016/S0967-0637(98)00018-1

Yao, W., X. Liu, and R. H. Byrne. 2007. Impurities in indicators used for spectrophotometric seawater pH measurements: Assessment and remedies. Mar. Chem. **107**: 167–172. doi:10.1016/j.marchem.2007.06.012

Conflict of Interest

None declared.

Associate editor: Mike DeGrandpre