

Evaluation of marine pH sensors under controlled and natural conditions for the Wendy Schmidt Ocean Health XPRIZE

Remy R. Okazaki ^{1,2*} Adrienne J. Sutton,^{1,2} Richard A. Feely,² Andrew G. Dickson,³ Simone R. Alin,² Christopher L. Sabine,² Paul M. E. Bunje,⁴ Jyotika I. Virmani⁴

¹Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington

²Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, Washington

³Scripps Institution of Oceanography, University of San Diego, La Jolla, California

⁴XPRIZE Foundation, Culver City, California

Abstract

The annual anthropogenic ocean carbon uptake of 2.6 ± 0.5 Gt C is changing ocean composition (e.g., pH) at unprecedented rates, but our ability to track this trend effectively across various ocean ecosystems is limited by the availability of low-cost, high-quality autonomous pH sensors. The Wendy Schmidt Ocean Health XPRIZE was a year-long competition to address this scientific need by awarding \$2 million to developers who could improve the performance and reduce the cost of pH sensors. Contestants' sensors were deployed in a series of trials designed to test their accuracy, repeatability, and stability in laboratory, coastal, and open-ocean settings. This report details the validation efforts behind the competition, which included designing the sensor evaluation trials, providing the conventional true pH values against which sensors were judged, and quantifying measurement uncertainty. Expanded uncertainty (coverage factor $k = 2$, corresponding to 95% confidence) of validation measurements throughout the competition was approximately 0.01 pH units. A custom tank was designed for the coastal trials to expose the sensors to natural conditions, including temporal variability and biofouling, in a spatially homogenous environment. The competition prioritized the performance metrics of accuracy, repeatability, and stability over specific applications such as high-frequency measurements. Although the XPRIZE competition focused on pH sensors, it highlights considerations for testing other marine sensors and measuring seawater carbonate chemistry.

The global oceans have absorbed about 30% of anthropogenic CO₂ emitted since the start of the Industrial Revolution (Sabine et al. 2004; Le Quéré et al. 2015), resulting in an approximately 28% increase in [H⁺], corresponding to a 0.11 decrease in pH (Gattuso et al. 2015). The rate of this decline in pH, termed ocean acidification (OA), is faster than at any other time in the geologic record (Zeebe 2012). Ocean acidification is expected to have deleterious effects on many marine taxa and ecosystems (Fabry et al. 2008; Doney et al. 2009; Somero et al. 2016), including reduced growth in marine calcifiers (Hendriks et al. 2010; Kroeker et al. 2010) and impaired sensory function in fishes (Nilsson et al. 2012). The effects of ocean acidification on foundational species like coral, phytoplankton, and shellfish will have cascading effects on community structure, food webs,

biogeochemical cycling, and commercial fisheries (Hofmann and Schellnhuber 2009; Narita et al. 2012; Kroeker et al. 2013; Punt et al. 2014; Ekstrom et al. 2015; Mathis et al. 2015).

Despite the documented effects of altered seawater carbonate chemistry on many ecosystems, ocean acidification data are limited in both coverage and quality (Borges et al. 2009). Global ocean climate datasets like the World Ocean Climate Experiment (WOCE), Hawaii Ocean Time-series (HOT), and Bermuda Atlantic Time-series (BATS) have included carbon variables only since the late 1980s. These large-scale sampling efforts and long-term time series have mostly concentrated on the open ocean. Coastal areas experience more variability in CO₂ partial pressure ($p\text{CO}_2$) and represent 10–20% of the oceanic CO₂ sink (Chen and Borges 2009; Doney et al. 2009; Takahashi et al. 2009), yet most $p\text{CO}_2$ time-series in coastal waters are less than a decade old (Borges et al. 2009; Feely and Dickson 2009). This relative paucity of data can be attributed to both economic and technological limitations.

*Correspondence: Remy.Okazaki@noaa.gov

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Recent developments in floats, profilers, and automated underwater vehicles are increasing scientists' capacity to measure ocean conditions across time and space by increasing the number of observations and decreasing the cost of obtaining them (Byrne 2014). Equally important to the development of these new sensing platforms are high-quality, cost-effective autonomous sensors that perform comparably to established laboratory methods and that have effective in situ quality control. Furthermore, sensors need to be capable of long deployments, which requires in situ calibration, anti-biofouling materials, and low power consumption (Schuster et al. 2009; Byrne 2014). As platforms add more sensing packages, the capacity for sensors to integrate with others and communicate both bi-directionally and remotely with users becomes more important.

While greater distribution of OA measurements throughout the oceans is important, properly defining uncertainty in those measurements is also critical. The Global Ocean Acidification Observing Network (GOA-ON) has defined two tiers for OA measurement quality: (1) a "climate" standard, whereby long-term anthropogenic effects on OA can be distinguished from natural variability, and (2) a broader "weather" standard that allows spatial patterns and short-term natural variation to be adequately resolved (Newton et al. 2015). The "weather" standard defines the target relative standard uncertainty in carbonate ion concentration $[\text{CO}_3^{2-}]$ as 10%, corresponding to absolute uncertainties at 400 μatm seawater $p\text{CO}_2$ of approximately 0.02 in pH, 10 $\mu\text{mol kg}^{-1}$ in total alkalinity (TA) and dissolved inorganic carbon (DIC), and 10 μatm in $p\text{CO}_2$. The "climate" standard defines the target relative standard uncertainty in a change in $[\text{CO}_3^{2-}]$ as 1%, corresponding to uncertainties of approximately 0.003 in pH, 2 $\mu\text{mol kg}^{-1}$ in TA and DIC, and 2 μatm in $p\text{CO}_2$. Currently, some laboratories are capable of achieving "climate" quality measurements, but these measurements are costly and labor-intensive, often requiring the collection of discrete bottle samples. These costs have contributed to the lack of spatiotemporal coverage of OA measurements based on established methods.

Tracking ocean acidification requires values for at least two of the aforementioned carbon parameters to fully characterize the CO_2 system chemistry. Currently, only $p\text{CO}_2$ and pH can be measured autonomously on commercially available sensors (Martz et al. 2015). These two parameters are not the ideal pair for characterizing CO_2 system chemistry because they covary. Even under a best-case scenario with uncertainties of 0.002 pH units and 2 μatm $p\text{CO}_2$, uncertainties of at least 15 $\mu\text{mol kg}^{-1}$ would be propagated into the estimates of DIC and TA (Millero 2007). Autonomous sensors for TA and DIC are in various stages of development (Byrne 2014; Spaulding et al. 2014; Martz et al. 2015). Until they become widely available, one approach to increasing OA measurements is the co-deployment of $p\text{CO}_2$ and pH sensors along with continued efforts to reduce

uncertainty in these sensors. Gas-equilibrating surface sensors for $p\text{CO}_2$ are believed to be able to measure at "climate" level targets of $<2 \mu\text{atm}$ and are present on many moorings and ships of opportunity (Friederich et al. 1995; Pierrot et al. 2009; Sutton et al. 2014). Subsurface $p\text{CO}_2$ sensors have approached this target (Tamburri et al. 2011; Atamanchuk et al. 2014; Fietzek et al. 2014; Jiang et al. 2014). Conversely, pH sensors have experienced relatively slower development. For example, the commonly used potentiometric glass electrode has changed little since its development in 1908 (Haber and Klemensiewicz 1909) and subsequent commercialization in the 1930s. Glass electrodes tend to measure inconsistently as a consequence of a variety of random and systematic effects that are inherent in such measurements, introducing errors of unknown magnitude. As a result, they are limited to uncertainties >0.005 pH units (Spitzer and Pratt 2011) and are not suited for long-term OA monitoring activities.

Newer autonomous pH-sensing technologies, namely potentiometry with ion sensitive field effect transistors (ISFETs) and spectrophotometry with pH-sensitive indicator dyes, have the potential to meet the GOA-ON "climate" standard. ISFETs were initially tested for seawater applications by Le Bris and Birot (1997) and have been commercialized in the last 7 yr (Martz et al. 2010). Spectrophotometric methods for pH, regularly used in oceanography, were developed 15 yr earlier (Robert-Baldo et al. 1985), and sensors based on this approach were commercialized in the 2000s (Martz et al. 2003; Seidel et al. 2008). These divergent approaches each have some distinct advantages. Sensors employing ISFETs are low-power, have rapid response times, and do not require moving parts. Spectrophotometric sensors are very precise and do not require calibration beyond characterizing the apparent absorption profiles of the purified indicator dye species (Liu et al. 2011). Both technologies are capable of long-term deployments with little drift (Gray et al. 2011; Hofmann et al. 2011). However, they are expensive and too complicated for many users, limiting their widespread use. If pH sensors performed as reliably as temperature or salinity sensors and were available at similar price points, they could dramatically improve our understanding of ocean acidification processes.

The XPRIZE Foundation targets technology gaps such as this through competitions that catalyze research and development in areas with potential benefit to humanity (Virmani and Bunje 2015). The goals of the Wendy Schmidt Ocean Health XPRIZE (WSOHXP) were to improve the performance and reduce the cost of robust pH sensing technology. The competition's \$2 million prize was therefore divided into two categories: Performance and Affordability. The WSOHXP was announced in September 2013, and the evaluations were conducted from September 2014 through May 2015. This competition satisfied two important requirements in advancing sensor technology: (1) testing new

Table 1. Scoring criteria for the Wendy Schmidt Ocean Health XPRIZE.

Phase	Measurement criterion	Accuracy prize purse		Affordability prize purse	
		Minimum standards	Points %	Minimum standards	Points %
2A	Accuracy	≤ 0.02 pH from “accepted value”	20	≤ 0.04 pH from “accepted value”	20
2B	Stability	Interdecile range ≤ 0.05	10	Interdecile range ≤ 0.1	10
2B	Repeatability	1 SD ≤ 0.01 pH units	5	1 SD ≤ 0.02 pH units	5
3	Stability	Interdecile range ≤ 0.05	10	Interdecile range ≤ 0.1	7.5
3	Repeatability	1 SD ≤ 0.01 pH units	10	1 SD ≤ 0.02 pH units	7.5
3	Ease-of-use	—	5	—	25
3	Affordability	$\leq \$15,000$	5	$\leq \$1,000$	25
4	Accuracy	≤ 0.02 pH from “accepted value”	15	—	—
4	Repeatability	1 SD ≤ 0.01 pH units	15	—	—
4	Ease-of-use	—	5	—	—
Total			100	100	

SD, standard deviation.

sensors against conventional methods under varied yet appropriate conditions, and (2) focusing on commercialization as a means of ensuring widespread adoption (Byrne et al. 2010). Seventeen sensors competed, of which 12 were prototypes from non-commercial entities. The new technologies and increased participation illustrate how prizes can leverage research and development.

Previous studies offered valuable insights for the experimental design of the WSOHXP sensor assessments. A year earlier, the industry consortium Alliance for Coastal Technologies (ACT) tested seven commercially available pH sensors in a 10-week laboratory deployment and several 4–12 week coastal deployments (Johengen et al. 2015). Validation samples were collected in close proximity to sensors (< 1 m) at synchronized sampling times to reduce variability between the validation and instrument measurements arising from changes in water mass or diffusion of heat/ CO_2 within the system. In the field trials, observed differences between instruments' and validation measurements increased when environmental conditions fluctuated. The largest spatial variation was observed during a storm event where the maximum recorded temperature difference across the test space (< 2 m) was 2.5°C , resulting in potential pH errors of at least 0.04 pH units. In a separate evaluation for $p\text{CO}_2$ sensors, the ACT estimated $p\text{CO}_2$ uncertainty of 2–15 μatm across the test spaces (< 1 m) during their evaluations at two coastal sites (Tamburri et al. 2011). They observed large, rapid fluctuations of temperature (1°C h^{-1}) and $p\text{CO}_2$ (400 μatm within hours) during these tests. Bubbles near sensor interfaces were another concern during the pH sensor evaluations, especially for deployments near the surface. Outside of the ACT trials, developers have tested their instruments in laboratory, underway, and moored field deployments where pH was compared directly against spectrophotometric pH measurements or indirectly against other CO_2 parameters (Bellerby et al. 2002; Liu et al. 2006; Seidel et al.

2008; Martz et al. 2010). In these tests, variability among sensors and validation measurements was attributed to spatial differences in water masses, temporal mismatching, biofouling, and interference from bubbles and detritus. For example, Seidel et al. (2008) observed water mass variability of 2°C between sensors located 2 m apart at the same depth. Though they did not report the resulting pH variability, a 2°C difference could result in pH measurement differences upwards of 0.03 pH units. These studies illustrate how environmental factors can complicate sensor evaluations, and how sampling protocols can mitigate potential sources of error.

The WSOHXP competition was designed to test the performance of sensors in a series of environments representing real-world applications. The four competition phases in chronological order were Phase 1 registration, Phase 2 laboratory trials, Phase 3 coastal trials, and Phase 4 open-ocean trials. Phase 2 required sensors to measure both Tris-artificial seawater buffer and seawater in stable, regulated systems. This phase was designed to directly assess sensor performance with minimal uncertainty from extraneous, uncontrolled sources of variability. The Phase 3 coastal trials tested sensors for a month in more naturally varying estuarine conditions with the added possibility of biofouling. The Phase 4 open-ocean trials focused on the ability of finalist sensors to record pH over a range of depths up to 3000 m. The sensors themselves are not evaluated in this report because they were judged by a separate, independent panel. However, the judging criteria (Table 1) are described because they influenced the design of each phase, and anonymized sensor results are discussed in terms of competition design. The performance criteria consisted of accuracy, repeatability, and stability. Accuracy was defined as the pH residual, i.e., the difference between the value of pH reported by a sensor and the conventional true value of pH for that seawater. The *conventional true value* of pH for a seawater sample was

defined as the pH measured with a spectrophotometer using purified *m*-cresol purple indicator dye and calculated from the calibration values of Liu et al. (2011). By definition, there is no uncertainty in this conventional true value, only in the realization of it. Repeatability, a short-term estimate of precision, was defined as the agreement among successive measurements of pH carried out under the same conditions over the course of each phase, expressed as a standard deviation. This metric assumes the pH in the test space does not change significantly in the time interval between two consecutive measurements. Stability, a long-term estimate of precision, was defined as the interdecile range of pH residuals observed over the trial deployment. Stable sensors should have similar pH offsets from validation measurements for the entire phase, and therefore a small interdecile range. This range, encompassing 80% of the pH residuals, corresponds to 2.56 standard deviations of a normal distribution. Two additional judging criteria were manufacturing cost and ease-of-use. Manufacturing cost was directly assessed from the materials cost estimate to manufacture a sensor and from supporting documentation provided by competition participants. Ease-of-use was defined as the ease with which devices can be calibrated (or self-calibrated), deployed, maintained, and the data downloaded, taking into consideration physical size, weight, durability, accessibility, and related characteristics.

Critically important to the competition were (1) ensuring the validation estimates of conventional true pH reflected the environmental conditions experienced by all the sensors and (2) characterizing the uncertainty of those estimates. These validation goals were accomplished in each competition phase through engineering, monitoring, modeling, and QA/QC protocols. Test spaces were designed to be spatiotemporally homogenous with respect to pH during validation measurements, and variation in pH of the test spaces with time was characterized through regular monitoring. Magnitudes of sampling and analysis uncertainty were evaluated from validation instruments, measurements, and comparisons with an independent laboratory. Tradeoffs in the experimental design and sensor evaluations are considered in the context of sensor performance. The topics covered here are likely relevant to the broader topics of seawater chemistry measurements and sensor evaluations.

Methods

Phase 2: Laboratory trials

The laboratory trials consisted of an accuracy test in Tris-artificial seawater buffer (Phase 2A) and repeatability and stability tests in seawater over 7 weeks (Phase 2B). The laboratory trials were conducted at the Monterey Bay Aquarium Research Institute (MBARI) in Moss Landing, California, U.S.A. from September to November 2014. The pH of the Tris buffer test solution was unknown to contestant teams, and this test solution was composed of approximately

0.03 : 0.05 mol kg⁻¹ Tris : Tris-HCl in synthetic seawater of nominal salinity 35 (DelValls and Dickson 1998). Each sensor was submerged in a container of Tris buffer and after equilibrating with the buffer, recorded 10 measurements at a rate of one measurement per minute or as quickly as possible for slower instruments. After the first set of pH measurements, additional HCl was added to the buffer to lower pH by approximately 0.4 pH units. Validation samples were collected in duplicate at the midpoint of the sensor's regular and lowered pH measurements. The Tris buffer and HCl solutions were composed of the same synthetic seawater background to maintain the same activity coefficients of their constituent species. Sodium bromide was included to ensure stable behavior in sensors with chloride ion-sensitive electrodes. The Phase 2A trials were conducted in a temperature-controlled laboratory, and temperature of the Tris test buffer solution was monitored throughout the tests (Fluke 1521 and 1504 readouts with Amphenol AS115-4-wire thermistors, manufacturer-specified accuracies ± 0.005 , ± 0.002 , ± 0.001 K, respectively). Teams were allowed to submit up to three sensors. Following the Tris tests, teams were allowed to review preliminary results, recalibrate sensor(s), and choose a single sensor to compete for the remaining phases of the competition. They were not allowed to make substantial modifications that affected sensor performance. Sensors were securely stored by XPRIZE personnel between competition phases.

Phase 2B long-term repeatability and stability tests were conducted in MBARI's seawater test tank ($L \times W \times D = 14 \times 9 \times 10$ m) following a 1-d test deployment. The test tank is stable, with low diurnal variability in temperature and only gradual increases in salinity and pH from evaporation and out-gassing (Martz et al. 2010). To provide a larger range of pH for sensors to measure, weekly step changes in pH of up to 0.2 pH units were planned. Sensors were programmed to record pH every 2 h. Validation samples were collected in duplicate from three sampling locations in the tank, twice daily, four times a week (i.e., 48 samples per week). Sensors were deployed at 3 m depth within 1.2 m of a sampling location at the same 3 m depth (Fig. 1a). Seawater was recirculated within the tank with minimal turbulence at a turnover rate of 12 h. Tank conditions were monitored with a CTD (Seabird 16plusV2, manufacturer-specified temperature accuracy ± 0.005 K, conductivity accuracy ± 0.00005 S m⁻¹) and pH sensor (Satlantic SeaFET, manufacturer-specified precision ± 0.004 pH units). Temporal variation was estimated as the standard deviations of these measurements, pooled by hour. Spatial variation across the tank was assessed weekly with additional discrete pH samples or CTD surveys. The tank was ozonated once a week to limit biofouling, and the resulting microbubbles were allowed to out-gas for at least another day to limit potential interference with the sensors. No sensor data were evaluated nor validation samples collected during these periods.

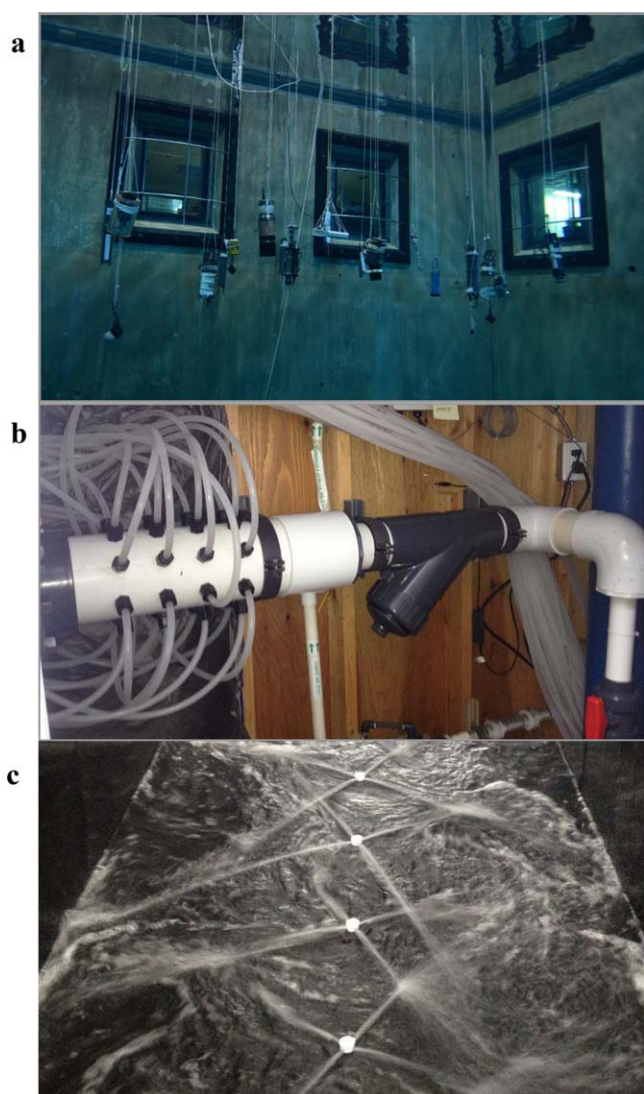


Fig. 1. Contestant sensors were (a) suspended at 3 m in the MBARI test tank during Phase 2B. Seawater in Phase 3 was (b) pumped into a manifold then divided over 32 equal lengths of tubing that terminated in (c) four-way diffusers located along the bottom of the tank. The tank is empty in (c) to show the water flow through the diffusers. (Photos: S. Haddock and R. Kiel).

Phase 3: Coastal trial

The 4-week coastal trial tested the repeatability and stability of sensors under approximately naturally varying seawater conditions at the Seattle Aquarium (Pier 59) utilizing seawater pumped from Elliott Bay, Washington, U.S.A. in February 2015. Tests were conducted in a custom $10 \times 1 \times 1$ m (L $\times$ W $\times$ D) tank designed to reduce spatial variation while still tracking external temporal changes in near real time. Seawater was pumped from 10 m depth in Elliott Bay at 400 L min^{-1} into the tank with a distribution manifold through 32 equal lengths of tubing terminating at diffusers positioned along the bottom of the tank (Fig. 1b,c). Seawater flowed into the tank

evenly from the bottom and exited through two skimmer pipes running the length of the tank with minimal turbulence. An awning provided protection from the elements. Sensors were positioned at mid-depth (0.5 m) and within 1 m of one of three evenly spaced validation sampling locations at the same 0.5 m depth. During sampling, inflow was stopped and power-heads provided water circulation. Validation samples were collected in duplicate at each sampling location twice daily, 5 d per week (i.e., 60 measurements per week). Tank monitoring and surveys of spatial pH variability were conducted as in Phase 2B. The pH of Elliott Bay water was monitored in situ with two glass pH electrodes (YSI 6600 V2) at 1 m and 10 m maintained by King County (<http://green2.kingcounty.gov/marine-buoy/SeattleAquariumData.aspx>; King County 2015). The sensor at 10 m is located at the seawater intake for the aquarium. A third estimate of pH was derived from a salinity-TA relationship based on discrete samples collected during the experiment and a flow-through $p\text{CO}_2$ equilibrator (General Oceanics GO8050) on the Seattle Aquarium's seawater intake. Although these additional data provided further information on the intake water, only the validation measurements were used to evaluate sensor performances.

Phase 4: Open-ocean trials

The oceanographic deployment tested the accuracy and repeatability of five finalist sensors on nine vertical hydrocasts with max depths increasing from 250 m to 3000 m. These trials were conducted at the HOT Station ALOHA (A Long-Term Oligotrophic Habitat Assessment; $22^\circ 45' \text{ N}$, $158^\circ 00' \text{ W}$) during 15 May 2015–19 May 2015. The water column was profiled by a CTD lowered at a rate of 30 m min^{-1} and discrete samples were collected in 12 L Niskin bottles at 15 depths on each upcast. Before each cast, sensors were attached to a rosette containing the CTD and Niskin bottles just before each cast using hose clamps, stainless steel bolts, fiberglass strut, and cable ties. Sensors were programmed to begin sampling at a specified time and at a sampling rate of 1 Hz. Sensors that sampled at slower rates were penalized under the Ease-of-use category. The longest sampling period of the finalist sensors was 6 min. Consequently, the rosette was stopped for a minimum of 6 min at each sampling depth, and the Niskin was tripped at the 3-min mark. The rosette CTD (Seabird SBE 9/11plus with 6800-m rated pressure sensor, SBE 3P, SBE 4C, and SBE 43 sensors; manufacturer-specified temperature accuracy $\pm 0.001 \text{ K}$, conductivity accuracy $\pm 0.0003 \text{ S m}^{-1}$, depth accuracy $\pm 1 \text{ m}$) recorded ambient conditions at 1 Hz. Once the rosette returned to the deck and sensors were removed, water samples were collected in the following order: validation pH, audit pH, dissolved oxygen (DO), and salinity. This procedure prioritized pH and was a departure from common oceanographic practice where DO is sampled before pH (Dickson et al. 2007).

Validation procedures for all phases of the competition

For the purposes of the competition, “pH” is defined as the negative decadic logarithm of total hydrogen ion concentration of seawater expressed in moles per kilogram of solution. The conventional true value of pH used to validate the sensors was based on spectrophotometric pH measurements using a stock solution of 10 mM purified *m*-cresol purple indicator dye (Byrne Laboratory, University of South Florida) and calculated using the parameters of Liu et al. (2011) after correcting for any pH change resulting from the addition of dye to the sample (see below).

Sampling and measurements were conducted according to best practices (Dickson et al. 2007). Briefly, samples were collected with minimal turbulence in borosilicate glass bottles after pre-rinsing and at least one volume overflow. Seawater samples were left with ~ 1% headspace and poisoned with a 0.04% (by volume) addition of saturated HgCl₂ solution. Samples were poisoned as a precautionary measure for any laboratory problems that might delay analyses and for methodological consistency with the audit samples that were analyzed in a different laboratory days-to-weeks later. Validation measurements were performed with a 10 cm path length cell on single-beam spectrophotometers (Agilent Cary 8453, 8454, manufacturer-specified absorbance accuracy ± 0.005 absorbance units, AU) manually in Phase 2 and with an automated system (Carter et al. 2013) in Phases 3 and 4. Dye perturbation measurements were made on >15% of seawater samples by measuring those samples with a double addition of dye. With the manual system in Phase 2B, the dye correction was based on an added 10 μ L dye addition to the same sample (Dickson et al. 2007). Dye corrections with the automated system were based on a second, double-dye measurement, and the amount of dye added was calculated from isosbestic absorbances at 488 nm (Carter et al. 2013). Analysis temperature was 15°C in Phase 2A and 20°C for the remainder of the competition. Analysis temperatures were regulated by a custom cell incubator in the manual method and by a custom water-jacketed spectrophotometer cell in the automated method. Temperature was recorded immediately following measurements (Fluke 1521, 1504 readouts with Amphenol AS115-4-WIRE thermistors, manufacturer-specified accuracies of ± 0.005 , ± 0.002 , and ± 0.001 K, respectively). Validation pH was reported, as the mean of duplicate measurements, at in situ temperature and pressure conditions on the total scale using equilibrium constants K_1 and K_2 from Lueker et al. (2000); $K_{\text{H}_2\text{SO}_4}$ from Dickson (1990a); $K_{\text{B(OH)}_3}$ from Dickson (1990b); $K_{\text{H}_3\text{PO}_4}$, $K_{\text{H}_2\text{PO}_4}$, K_{HPO_4} , $K_{\text{H}_2\text{O}}$, and K_{Si} from Millero (1995); and K_{HF} from Perez and Fraga (1987) (or Dickson and Riley 1979) for temperatures < 9°C). Pressure corrections for equilibrium constants follow Millero (1995) and were calculated on the seawater scale before conversion to the total scale. Total alkalinity was derived from audit measurements, and carbonate alkalinity was calculated from total concentrations of boron (Uppström 1974), sulfur (Morris and Riley 1966), and

fluoride (Riley 1965) in proportion to salinity. Phosphate and silicate concentrations were assumed to be 0 $\mu\text{mol kg}^{-1}$. In Phase 2A, the DelValls and Dickson (1998) temperature dependence for the pH of equimolar (0.04 : 0.04 mol kg⁻¹ Tris : Tris-HCl) buffer was used to adjust pH to in situ temperature for both test solutions, assuming it represents the change in pK with temperature. To adjust seawater pH to in situ temperature and pressure, DIC was first calculated from the pH measured under laboratory conditions and TA, then TA and DIC were used to estimate in situ pH (at the in situ temperature and pressure). The in situ adjustments of seawater samples were calculated using the *seacarb* package v3.0 (Gattuso et al. 2014) in R v3.1 (R Core Team 2014) and CO2Sys package v25 (Pelletier et al. 2015) in Microsoft Excel.

Quality assurance and quality control (QA/QC)

The quality assurance plan required pH measurements to be made using appropriate equipment operated by qualified individuals, targeting a standard uncertainty in pH of 0.003–0.005 based on state-of-the-art methods and “climate standards” (Carter et al. 2013; Newton et al. 2015). Appropriate equipment was defined as new (< 1 yr old) equipment with traceable calibrations and/or older equipment passing periodic testing relative to newer equipment and with reference materials. Throughout the competition, spectrophotometers passed diagnostic tests using holmium oxide and absorbance filter reference materials. Primary and backup spectrophotometers were compared in side-by-side tests during Phases 2B and 3. Thermistors were batch-tested in baths, measuring within 0.01 K of one another. The purified dye composition was verified by the National Institute of Standards and Technology. All validation measurements were performed by a single analyst, while additional measurements of the analyst were compared with colleagues at the start of the competition, with an audit laboratory, and against reference materials throughout the competition. An advisory board provided feedback on protocols and results. Operating procedures, diagnostics, notes, and data were recorded in a notebook and digitally with regular back-ups.

Quality control measures included replicate measurements, regular measurement of suitable reference materials, comparisons with an independent laboratory, and comparisons of secondary measurements. Duplicate validation bottle samples were collected and analyzed at every validation point. During each analytical run, validation bottle samples were only analyzed after confirming acceptable pH measurements of Tris or seawater reference materials (all reference materials were obtained from the Dickson Laboratory at the Scripps Institution of Oceanography, University of California, San Diego). The reference materials used had pH values differing by up to ~ 0.3 pH units. Additional discrete samples were collected in duplicate at places and times corresponding to approximately 15% of validation samples, and these audit samples were analyzed by the Dickson Laboratory for pH, total alkalinity, and salinity. Secondary measurements of pH, other carbonate

Table 2. Sources of uncertainty for each phase of the competition, summed in quadrature to yield the combined standard uncertainty.

Sources of uncertainty	Estimated standard uncertainty	Contribution to pH uncertainty			
		Phase 2A	Phase 2B	Phase 3	Phase 4
Sample analysis	—	0.0042	0.0030	0.0017	0.0010
Sample handling	—	—	0.0022	0.0054	0.0034
Adjustments to in situ conditions	$T_{\text{Fluke1521}} = 0.012 \text{ K}$ $T_{\text{Fluke1504}} = 0.0013 \text{ K}$ $T_{\text{SBE16}} = 0.003 \text{ K}$ $T_{\text{SBE3}} = 0.0006 \text{ K}$ $P_{\text{SBE911}} = 0.6 \text{ dbar}$	0.0006	0.0002	0.0001	0.0001
Constants in pressure correction	8%, i.e., 0.0033 pH units/1000 dbar	—	—	—	0– 0.0101
Spectrophotometer absorbance	0.002 AU	0.0037	0.0033	0.0019	0.0024
Dye perturbation	—	—	0.0004	0.0014	0.0013
Standard uncertainty	0 dbar 3050 dbar	0.0056	0.0050	0.0061	0.0045– 0.0110
Expanded uncertainty ($k = 2$)	0 dbar 3050 dbar	0.011	0.010	0.012	0.009– 0.022

parameters, and dissolved oxygen provided additional information on the likely quality of the validation pH measurements and testing environments.

Sensor scoring

Contestant sensors were judged by a separate, independent panel from the validation team. After the conclusion of the competition, contestants' measurements at every validation sampling point were provided to the validation team. These data were used to estimate accuracy and stability, which were both estimated from differences between the measured sensor pH and the validation pH as described earlier. By definition, a sensor's repeatability was calculated as the standard deviation of differences between consecutive measurements, regardless of their accuracy relative to validation pH values. Here, repeatability can only be approximated based on the available data, with consecutive measurements spanning up to 6 h apart for Phases 2B and 3. Repeatability was not estimated for Phase 4 because pH naturally changes by as much as 0.5 pH units over the depth profile. The competition guidelines indicated sensors scored full points for accuracy and stability values within the reported validation expanded uncertainty corresponding to 95% confidence. For repeatability, full points could be scored for 0 pH unit standard deviations. Sensors scored no points for values greater than the minimum standards (Table 1). Values falling between these thresholds were awarded linearly scaled points, and total scores in each phase were weighted according to the maximum possible score for each criteria in each phase (Table 1). Similarly, if a sensor stopped collecting data

during any point of the competition, its score was penalized in proportion to the amount of data missing.

Assessment

Validation pH uncertainty

The main goal of validation efforts in the WSOHXP competition was to ensure that the reported conventional true pH values were representative of the seawater around the test sensors. Efforts to characterize uncertainty in validation measurements were as important as the measurements themselves because the judges used such uncertainty estimates in the scoring of sensor performance. The total expanded validation pH uncertainty (coverage factor $k = 2$, corresponding to 95% confidence) was approximately 0.01 pH units throughout the competition (Table 2) and is comparable to estimates made previously (Carter et al. 2013; Hammer et al. 2014; Bockmon and Dickson 2015). This uncertainty consisted of contributions from analysis, sample handling, corrections to in situ conditions, spectrophotometer absorbance, and dye perturbation corrections. These sources of uncertainty were summed in quadrature to obtain the combined standard uncertainty for the validation pH measurements in each phase (Ellison and Williams 2012).

All analyses were conducted by the same operator in the same mobile laboratory, housed in a shipping container that traveled with the competition. Outside environmental conditions, electrical power supply, internal temperature fluctuations, and so forth varied within and among the different phases of the competition. Their contributions to analytical uncertainty were estimated from the standard deviation (SD)

of measurements on various different reference materials over time (pooled by j groups). For the short-term Phase 2A tests, analysis repeatability was calculated as the pooled SD of all measurements made on the Tris buffer samples. The calculated Phase 2A measurement pooled SD was 0.0042 pH units ($n = 238$, $j = 119$ sample pairs). Standard deviations of measurements on reference materials in Phase 2B, 3 and 4 were 0.0030 ($n = 52$, $j = 3$ batches of reference materials, $t = 55$ d), 0.0017 ($n = 34$, $j = 3$ batches, $t = 25$ d), and 0.0010 ($n = 26$, $j = 2$ batches, $t = 7$ d) pH units, respectively. The improvement in analytical uncertainty over time between Phase 2B and Phases 3 and 4 was likely a result of the automated system's reduced and more standardized sample processing time.

Sample handling uncertainty was defined as uncertainties in measurement observed among different labs analyzing the same sample materials, including other unknown contributions. Sample handling uncertainty was estimated from the standard deviation of differences between the paired audit and validation pH measurements. Sample handling uncertainties for Phases 2B, 3, and 4 were 0.0020 ($n = 26$), 0.0054 (19), 0.0034 (25) pH units, comprising 10–50% of the combined standard uncertainty in each phase. Phase 2A did not include a sample handling comparison because audit samples were collected from the initial mixing carboy while validation samples were collected from testing containers with instruments.

Uncertainty from corrections to in situ conditions includes uncertainties in salinity, temperature, and pressure measurements of laboratory and CTD sensors. These uncertainties were assumed to have a uniform distribution within the manufacturer-specified accuracy ranges. Salinity uncertainties of $\leq 0.01\%$ have negligible effects on pH. The uncertainty of the laboratory thermistors used in Phases 2A and 2B were conservatively assumed to be 0.02 K based on tests in a controlled water bath and intercomparisons with other thermistors. Pressure sensor accuracy was irrelevant in Phases 2B and 3 and up to 1 dbar in Phase 4. The resulting pH uncertainties from temperature and pressure uncertainties were 0.0001–0.0006 pH units. Laboratory-measured pH values were corrected to in situ conditions, and the magnitude of the adjustment increases with increasing pressure. Similarly, uncertainty in the estimated in situ pH consequent on the uncertainties in the constants used in pressure adjustments will increase proportionately with depth. This uncertainty was assumed to be 8% of the difference between pH calculated, at in situ temperature, at the surface and at the actual sample depth. The assumed 8% pressure correction uncertainty corresponds to 0 pH units at the surface and 0.0101 pH units at 3000 m.

The contribution of the spectrophotometers to pH uncertainty was calculated assuming absorbance uncertainty of 0.005 AU. Treating this range as a triangular distribution implies a standard uncertainty in absorbance measurement

of $0.005/(6^{1/2}) = 0.002$. This value was then used to estimate the relative standard uncertainty in the absorbance ratio R (dR/R), using the relationship

$$\left(\frac{dR}{R}\right)^2 = \left(\frac{dA_1}{A_1}\right)^2 + \left(\frac{dA_2}{A_2}\right)^2 \quad (1)$$

together with the measured absorbance values at 434 nm and 578 nm (A_1 and A_2) for all samples in each competition phase. The maximum “smoothed” value over the range of absorbances was used as the resulting spectrophotometer contribution to uncertainty in pH for each phase. Reported uncertainties from spectrophotometer absorbance were 0.0019–0.0037 pH units across all three phases. Lower spectrophotometer uncertainty in the later phases likely resulted from an increase in the amount of dye added to samples with the automated system that subsequently reduced the ratio of absorbance error to absorbance measured.

The addition of indicator dye can affect the sample seawater pH, and there is an uncertainty associated with the empirically derived correction to this dye perturbation. This dye-correction intercept uncertainty was estimated as one-fourth of the largest adjustment. In Phase 2A, dye perturbation was assumed to be negligible as the samples were strongly buffered Tris/Tris- H^+ mixtures.

Phase 2A

Tris seawater buffer pH is twice as sensitive to temperature as seawater pH. Therefore, additional steps were taken to reduce the potential contributions of temperature variability to pH uncertainty, including regulating the testing lab temperature, allowing sensors to acclimate to the test room, and submerging test solutions in water baths to slow temperature fluctuations. Thermistors were cross-checked in water baths to verify their performance. Despite these efforts, temperature variability was introduced by larger sensors' thermal mass in the water baths and movements of personnel to and from both the laboratory and the testing room. Ultimately, when temperature could not be held constant, monitoring and recording temperatures during sample collection and analysis was critical. The advantage of testing sensors in Tris buffer as opposed to seawater is that the pH of a Tris buffer solution would not be expected to drift during the test period, even for the new solution obtained after the addition of HCl. The tradeoffs include the aforementioned temperature sensitivity and the added cost. Phase 2A required mixing approximately 800 L of Tris buffer to test approximately 50 sensors.

Sensor accuracy varied considerably, with nine of 17 teams measuring pH within the 0.02 pH unit performance threshold and another two measuring within the 0.04 pH unit affordability threshold (Fig. 2).

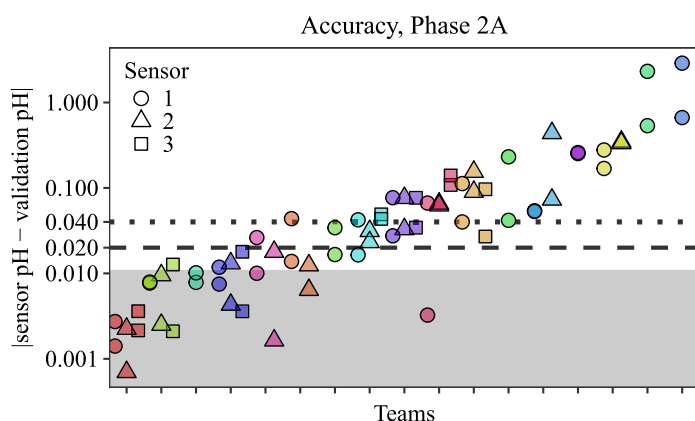


Fig. 2. Phase 2A accuracy in Tris seawater buffer as residuals between sensor and validation pH (y-axis, log-scale) for each team's sensor(s) (colored symbols along x-axis, ranked from left to right in order of increasing residuals). Teams were allowed to test up to three replicate sensors (shapes). Accuracy trials consisted of measurements in Tris seawater buffer at two different pH levels for each sensor (i.e., two points per shape). Residuals within the expanded uncertainty of validation pH measurements (gray shaded region) scored full points. Residuals between validation uncertainty and the Performance (dashed line) and Affordability (dotted line) category thresholds scored linearly scaled points from 0 to the maximum indicated in Table 1.

Phase 2B

Protocols were designed to limit natural variability in physical conditions throughout the test tank, ensuring sensors were compared fairly to the assigned conventional true pH value of the test solutions. These protocols included regular monitoring of spatiotemporal variability in pH and temperature during Phases 2B and 3. Periodic surveys with discrete pH sampling and CTD/pH sensors further assessed spatial variability. Discrete pH samples collected across the MBARI tank had a pooled standard deviation of 0.0027 pH units ($n=96$, $j=4$ surveys), comparable to the analysis uncertainty in this phase of 0.0030 pH units, and exhibited no consistent spatial pattern. The standard deviation of all validation samples pooled across sampling events, i.e., reflecting spatial heterogeneity between sampling locations, was 0.0017 pH units ($n=293$, $j=50$ events). Temperatures across the tank measured within 0.005 K of the tank center with an increasing temperature gradient from the west to east wall. This slight temperature gradient would result in negligible pH uncertainty of 4×10^{-5} pH units, assuming no differences in salinity and TA. Given the lack of measured pH differences throughout the tank and the negligible temperature effects, validation pH was reported as the average of the three sampling locations for each sampling event. Hourly standard deviations ($n=6$ h⁻¹) of the SeaFET pH sensor used for tank monitoring were generally less than the manufacturer-specified precision of 0.004 pH units, indicating stable pH over hour periods. During Phase 2B, accelerated stray current corrosion was observed in the MBARI test tank. Consequently, the tank pH manipulations were cancelled and testing was halted after 7 weeks to protect

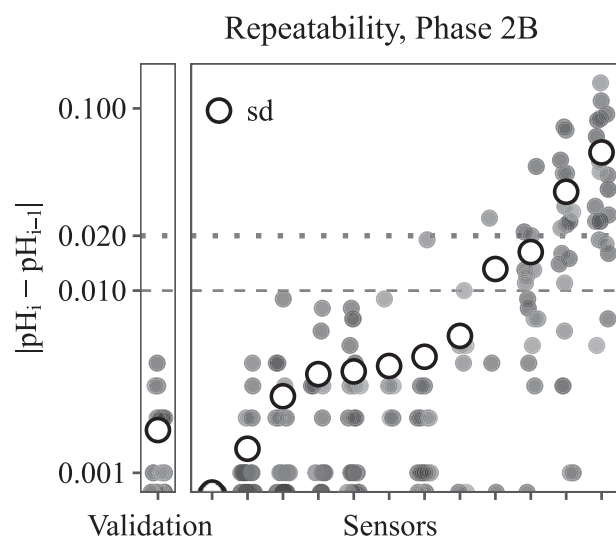


Fig. 3. Approximate Phase 2B sensor repeatability in the MBARI seawater test tank as the standard deviation (white circles) of differences in consecutive validation measurements (gray circles) for validation bottle samples (left panel) and each team's sensor (right panel; x-axis ticks, ranked from left to right in order of increasing deviations). Only measurements associated with validation samples are shown instead of sensors' full bihourly data log, and validation samples were collected 1–6 h apart. Therefore, this figure is an approximation of the actual sensor repeatability. Points for deviations were linearly scaled from 0 (perfect repeatability) to the Performance (dashed line) and Affordability (dotted line) category thresholds.

the sensors, cutting this phase short from the originally planned duration of 3 months. Subsequent electrical tests were unable to determine the source of the corrosion. Damage to sensors was limited to metal support bolts and sacrificial anodes.

The repeatability and stability of about half the sensors were within the Performance and Affordability thresholds (Figs. 3, 4). One sensor malfunctioned during this phase and reported only a single pH value for every timestamp, effectively earning it a perfect score in repeatability. Because the tank pH only increased by 0.002 pH units over 6 weeks, this sensor also performed relatively well in the stability category, whereas in a more dynamic environment this malfunction would have more severely penalized the sensor's score. Five sensors failed to record measurements.

Phase 3

In Phase 3, pre-experiment pH surveys with the pH probe indicated a pH range across the Seattle Aquarium test tank of ≤ 0.005 units with no discernible spatial pattern. Similarly, temperature readings ranged ≤ 0.005 K with no consistent gradients. However, discrete pH samples detected an intermittent pH gradient up to 0.008 pH units across the tank, though the overall tank pooled SD was 0.001 pH units ($N=240$, $j=40$ events). Because of this potential gradient, the validation pH was reported for each sampling location. Reporting mean tank

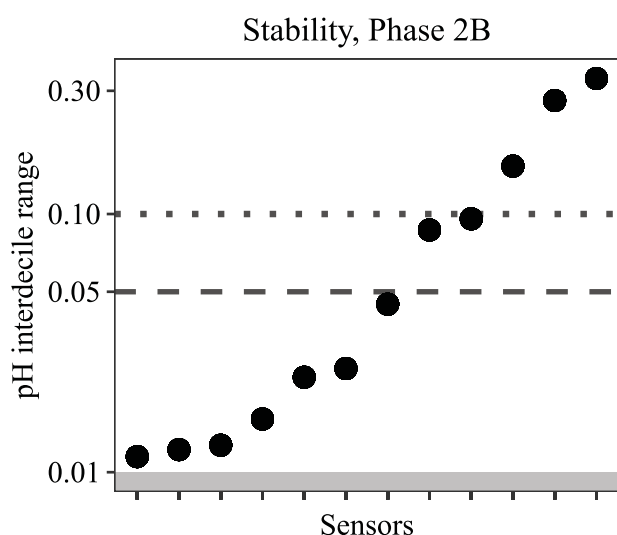


Fig. 4. Phase 2B stability over 6 weeks in the MBARI seawater tank. The interdecile range (IDR) of residuals (sensor pH - validation pH) (y-axis, log-scale) represents the stability of sensors over time. Sensors are displayed as x-axis ticks in order of increasing IDR. The gray shaded region represents expanded uncertainty of validation pH measurements. Performance (dashed line) and Affordability (dotted line) thresholds were 0.05 and 0.10 pH units.

values would have introduced an additional source of variability disproportionately affecting sensors at the far ends of the tank. This observation of ephemeral pH gradients illustrates the importance of continually surveying the test area. As in the previous phase, the pH variability on hourly timescales in Phase 3 was less than the auxiliary pH sensor repeatability. Discrete samples were generally collected within 20 min. Within these timeframes, no detectable pH variability could be attributed to time in both Phase 2B and 3, and therefore no additional steps were taken to account for temporal variability when reporting results.

The pH of Elliott Bay increased from 7.66 to a peak of 7.76 pH units over the first week, dropped to the original value over the next week, and increased again to 7.71 pH units 2 weeks later (Fig. 5). These pH movements reflected changes in primary productivity on weekly timescales. The King County buoy pH sensors tracked the relative changes of validation pH measurements well, but their offsets from the validation measurements drifted over time and even reversed upon redeployment on 18 February 2015. The performances of these glass electrode sensors suggest their utility was limited to tracking relative changes in pH on weekly timescales. The agreement in relative fluctuations of all buoy parameters (pH, dissolved oxygen, chlorophyll fluorescence) and validation pH over time indicated that the tank represented environment conditions. This tank design appeared to be well-suited for sensor testing in a variable coastal environment.

Phase 3 was conducted in February when biological productivity and pH variability are lower compared to the

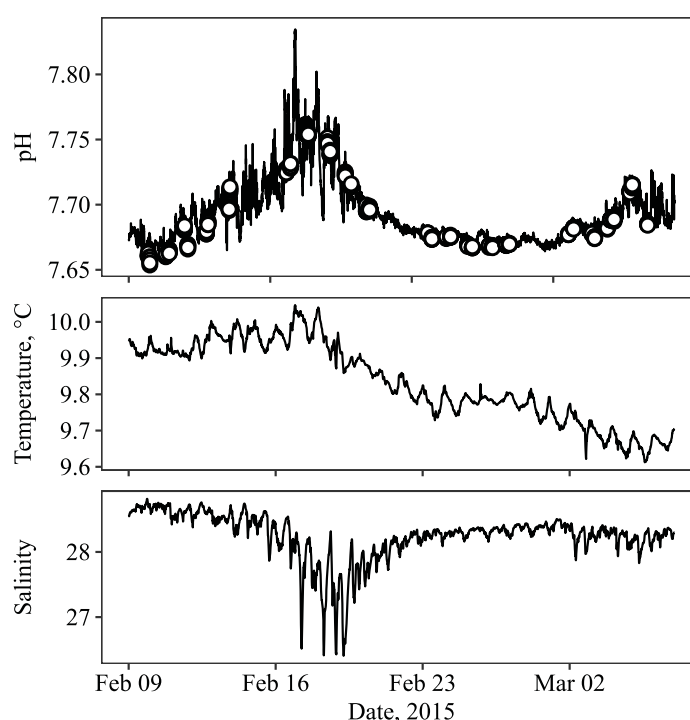


Fig. 5. Test tank pH, temperature, and salinity during Phase 3. Top panel includes spectrophotometric pH from discrete validation bottle samples (white circles) and continuous pH (black line) estimated as a function of TA (from salinity) and $p\text{CO}_2$. Temperature and salinity were recorded from a CTD in the test tank.

summer. Nonetheless, the background pH variability was higher than in Phase 2B. As in Phase 2B, half of the competing sensors recorded a significant proportion of their measurements within the Performance and Affordability thresholds for both repeatability and stability (Figs. 6, 7). The largest deviations in repeatability and stability were an order of magnitude higher than in Phase 2B, likely due to the increased environmental variability and biofouling compared to Phase 2B. These results illustrate the differences between variable field and controlled laboratory environments.

Phase 4

The high cost of ship time limits the testing of sensors in oceanographic settings. Most of the finalists' prototype sensors had never been exposed to pressures >1000 dbar. Casts were gradually increased in depth to allow sensors to collect some data in the event their housings were compromised at higher pressures. Indeed, one sensor's battery compartment was crushed at approximately 2000 m, precluding it from participating in the last cast. Another advantage of progressively deeper casts was increased sampling resolution where pH changes more rapidly at depths <1000 m. The in situ pH profiles decreased from 8.1 pH units in surface waters to a minimum of 7.6 between 700 dbar and 800 dbar, corresponding to the oxygen minimum zone (Fig. 8). The

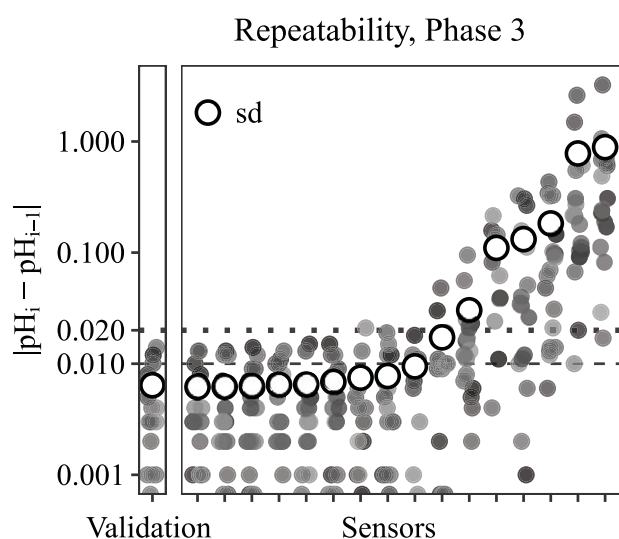


Fig. 6. Approximate Phase 3 sensor repeatability in the Seattle Aquarium test tank as the standard deviation (white circles) of differences in consecutive validation measurements (gray circles) for validation bottle samples (left panel) and each team's sensor (right panel; x-axis ticks, ranked from left to right in order of increasing deviations). Only measurements associated with validation samples are shown instead of sensors' full hourly data log, and validation samples were collected 1–6 h apart. Therefore, this figure is an approximation sensor repeatability. Points for deviations were linearly scaled from 0 (perfect repeatability) to the Performance (dashed line) and Affordability (dotted line) category thresholds.

reported expanded pH uncertainty increased from 0.009 pH units at the surface to 0.022 pH units at 3050 dbar due to the assumed uncertainty in pressure correction. This pressure correction uncertainty is currently unknown, but a lower bound of 5% can be inferred from Millero (1979). The 8% pressure correction uncertainty used here resulted in a pH uncertainty of 0.0101 pH units at 3050 dbar, which provided judges the scope to assess sensor accuracy within the competition's minimum standard of 0.02 pH units.

Hawaii Ocean Time-series Station ALOHA has measured pH since 1992 and is the longest sustained monthly seawater pH time-series in the world. This extensive time-series provided a record against which both sensor and validation measurements could be compared. Validation pH values were generally within 0.01 pH unit of historical April–June HOT Station ALOHA pH values (Fujieki et al. 2015), after recalculating the HOT ALOHA data using the same constants from Liu et al. (2011) (Fig. 8). The HOT-ALOHA pH measurements are made with uncharacterized, impure dye, resulting in an additional contribution to HOT-ALOHA pH uncertainty of 0.02 pH units (Liu et al. 2011). Despite these uncertainties, the validation pH-depth profiles were consistent with the historical record.

In addition to stepping casts to progressively deeper maximum depths, the 6-min rosette stop time at each depth was designed to ensure the collection of sensor measurements that could be compared to validation measurements obtained

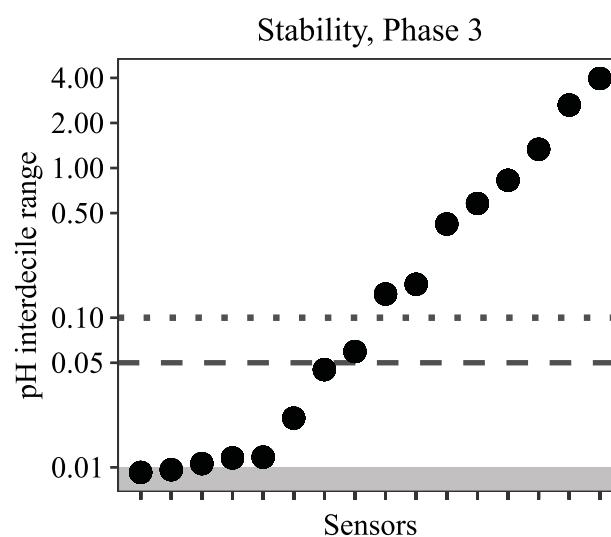


Fig. 7. Phase 3 stability over 4 weeks in the Seattle Aquarium test tank for each sensor. The interdecile range (IDR) of residuals (sensor pH – validation pH) (y-axis, log-scale) represents the stability of sensors over time. Sensors are displayed as x-axis ticks in order of increasing IDR. The gray shaded region represents expanded uncertainty of validation pH measurements. Performance (dashed line) and Affordability (dotted line) thresholds were 0.05 and 0.10 pH units.

from Niskin bottles. Normally, stop times are less than a minute. Consequently, the competition protocol eliminated the handicap that slower-sampling technologies, namely spectrophotometric methods, would have encountered in collecting high-resolution time-series. Slower instruments were instead penalized under the qualitative Ease-of-use category, which was worth fewer points (Table 1). Accuracy appeared to decrease with increasing pressure for all sensors (Fig. 9), potentially reflecting uncertainties due to the pH adjustment from laboratory to in situ pressure in addition to instrument measurement uncertainties resulting from the changing conditions. Temperature changes from approximately 26°C to 2°C along the cast profiles also could have contributed errors to sensors' pH readings if sensor thermistors were unable to equilibrate before the pH measurements were made. A third potential source of sensor pH errors could arise if in situ temperature, salinity, or pressure conditions extended beyond the ranges that teams used to calibrate their sensors. Repeatability cannot be assessed here because only the sensor measurements associated with Niskin samples were provided to the validation team, and Niskin bottles were collected at different depths. The overall results from Phase 4 reflect the tradeoffs of evaluating sensor performance vs. application, the unquantified uncertainties due to pressure, and the time/cost constraints of oceanographic deployments.

Discussion

The WSOHXP competition demonstrates how sensor accuracy, repeatability, and stability can be evaluated in

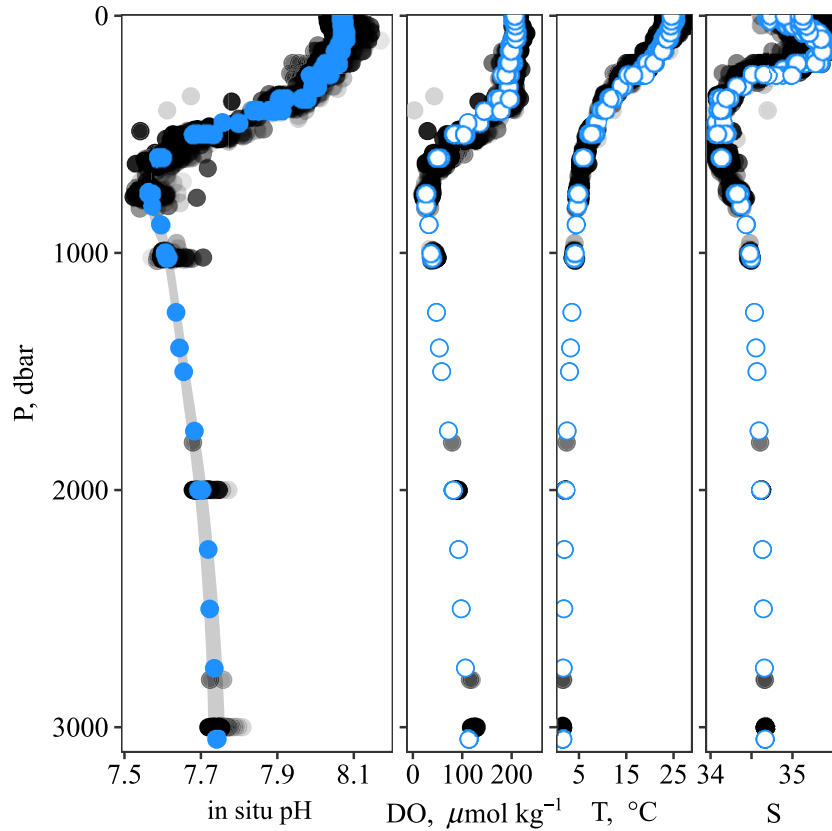


Fig. 8. Depth profiles of discrete pH and CTD dissolved oxygen (DO), temperature (T), and salinity (S) at HOT-ALOHA from the 2015 XPRIZE cruise (blue = validation measurements, white = CTD measurements) and 2000–2015 University of Hawaii cruises (Fujiieki et al. 2015) (grayscale from light gray in 2000 to black in 2015). The HOT-ALOHA pH values are recalculated according to Liu et al. (2011). The gray shaded region in the pH panel represents the expanded uncertainty of the validation pH measurements.

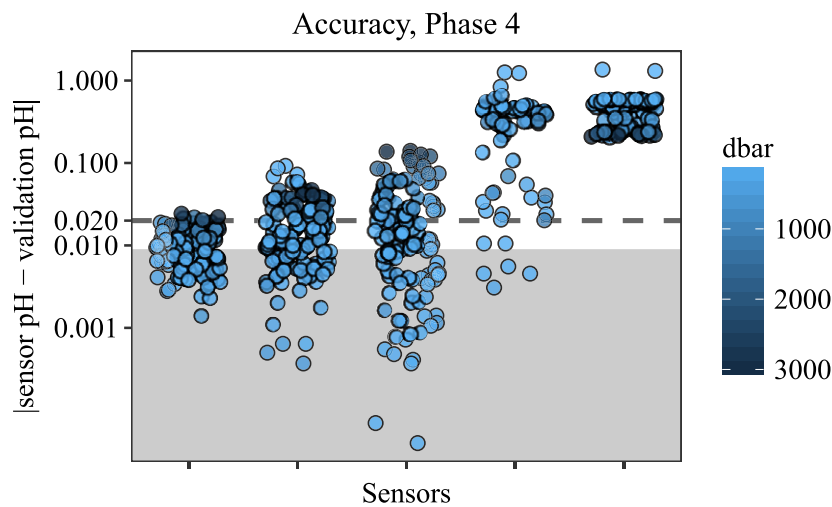


Fig. 9. Phase 4 accuracy at HOT Station ALOHA as residuals between sensor and validation pH (y-axis, log-scale) for each team's sensor (x-axis ticks, ranked from left to right in order of increasing residuals) colored according to pressure. Residuals are plotted with slight horizontal scatter and transparency to reduce overplotting. Expanded uncertainty ranged from 0.009 pH units (gray shaded region) to 0.022 pH units at 3050 dbar. The Performance category threshold (dashed line) was 0.02 pH units.

laboratory, coastal, and open-ocean settings. Temperature, gas exchange, pressure, and biological processes simultaneously affect seawater chemistry over differing spatial and temporal scales. For sensor comparisons, controlling and documenting these sources of variability is a necessary prerequisite to evaluating sensor performance. Collectively, the sensors tested in this competition demonstrated the ability to collect high-quality pH measurements under a range of conditions, including variable coastal and deep ocean environments. Their performances are noteworthy considering a recent inter-laboratory comparison resulted in differences from the assigned true value of up to 0.04 and 0.1 pH units for spectrophotometric and electrometric approaches, respectively, under controlled laboratory conditions (Bockmon and Dickson 2015). Beyond the technical aspects of the competition, the XPRIZE is a means of catalyzing technological development to address scientific and humanitarian issues. In the coming years, the technologies tested in this competition should simultaneously reduce the cost and increase the quality and coverage of OA data. These advances should improve climate models, OA-biological studies, and our understanding of ocean processes.

References

- Atamanchuk, D., A. Tengberg, P. J. Thomas, J. Hovdenes, A. Apostolodis, C. Huber, and P. Hall. 2014. Performance of a lifetime-based optode for measuring partial pressure of carbon dioxide in natural waters. *Limnol. Oceanogr.: Methods* **12**: 63–73. doi:10.4319/lom.2014.12.63
- Bellerby, R. G. J., A. Olsen, T. Johannessen, and P. Croot. 2002. A high precision spectrophotometric method for on-line shipboard seawater pH measurements: The automated marine pH sensor (AMpS). *Talanta* **56**: 61–69. doi:10.1016/S0039-9140(01)00541-0
- Bockmon, E. E., and A. G. Dickson. 2015. An inter-laboratory comparison assessing the quality of seawater carbon dioxide measurements. *Mar. Chem.* **171**: 36–43. doi:10.1016/j.marchem.2015.02.002
- Borges, A. V., and others. 2009. A global sea surface carbon observing system: Inorganic and organic carbon dynamics in coastal oceans. In J. Hall, D. E. Harrison, and D. Stammer [eds.], *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*. ESA.
- Byrne, R. H. 2014. Measuring ocean acidification: New technology for a new era of ocean chemistry. *Environ. Sci. Technol.* **48**: 5352–5360. doi:10.1021/es405819p
- Byrne, R., M. DeGrandpre, R. Short, T. Martz, L. Merlivat, C. McNeil, F. Sayles, R. Bell, and P. Fietzek. 2010. Sensors and Systems for In Situ Observations of Marine Carbon Dioxide System Variables, in *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society* (Vol. 2), Venice, Italy, 21–25 September 2009, J. Hall, D. E. Harrison, and D. Stammer [eds.], ESA Publication WPP-306, doi:10.5270/OceanObs09.cwp.13
- Carter, B. R., J. A. Radich, H. L. Doyle, and A. G. Dickson. 2013. An automated system for spectrophotometric seawater pH measurements. *Limnol. Oceanogr.: Methods* **11**: 16–27. doi:10.4319/lom.2013.11.16
- Chen, C.-T. A., and A. V. Borges. 2009. Reconciling opposing views on carbon cycling in the coastal ocean: Continental shelves as sinks and near-shore ecosystems as sources of atmospheric CO₂. *Deep-Sea Res. Part II* **56**: 578–590. doi:10.1016/j.dsr2.2009.01.001
- DelValls, T. A., and A. G. Dickson. 1998. The pH of buffers based on 2-amino-2-hydroxymethyl-1,3-propanediol ('tris') in synthetic sea water. *Deep-Sea Res. Part I* **45**: 1541–1554. doi:10.1016/S0967-0637(98)00019-3
- Dickson, A. G. 1990a. Standard potential of the reaction: AgCl(s) + 12H₂(g) = Ag(s) + HCl(aq), and the standard acidity constant of the ion HSO₄[−] 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. 1990b. Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 K to 318.15 K. *Deep-Sea Res. Part A* **37**: 755–766. doi:10.1016/0198-0149(90)90004-F
- Dickson, A. G., and J. P. Riley. 1979. The estimation of acid dissociation constants in seawater media from potentiometric titrations with strong base. I. The ionic product of water — Kw. *Mar. Chem.* **7**: 89–99. doi:10.1016/0304-4203(79)90001-X
- Dickson, A. G., C. L. Sabine, and J. R. Christian. 2007. Guide to best practices for ocean CO₂ measurements, p. 176. PICES Special Publication. North Pacific Marine Science Organization.
- Doney, S. C., V. J. Fabry, R. A. Feely, and J. A. Kleypas. 2009. Ocean acidification: The other CO₂ problem. *Ann. Rev. Mar. Sci.* **1**: 169–192. doi:10.1146/annurev.marine.010908.163834
- Ekstrom, J. A., and others. 2015. Vulnerability and adaptation of US shellfisheries to ocean acidification. *Nat. Clim. Chang.* **5**: 207–214. doi:10.1038/nclimate2508
- Ellison, S. L. R., and A. Williams [eds.]. 2012. Eurachem/CITAC guide: Quantifying uncertainty in analytical measurement, 3rd ed. ISBN 978-0-948926-30-3.
- Fabry, V. J., B. A. Seibel, R. A. Feely, and J. C. Orr. 2008. Impacts of ocean acidification on marine fauna and ecosystem processes. *ICES J. Mar. Sci.* **65**: 414–432. doi:10.1093/icesjms/fsn048
- Feely, R. A., and A. G. Dickson. 2009. New evidence for ocean acidification in coastal waters of North America. *Bull. Am. Meteorol. Soc.* **90**: S69. doi:10.1126/science.1155676
- Fietzek, P., B. Fiedler, T. Steinhoff, and A. Körtzinger. 2014. In situ quality assessment of a novel underwater pCO₂ sensor based on membrane equilibration and NDIR spectrometry. *J. Atmos. Ocean. Technol.* **31**: 181–196. doi:10.1175/JTECH-D-13-00083.1

- Friederich, G. E., P. G. Brewer, R. Herlien, and F. P. Chavez. 1995. Measurement of sea surface partial pressure of CO₂ from a moored buoy. *Deep-Sea Res. Part I* **42**: 1175–1186. doi:[10.1016/0967-0637\(95\)00044-7](https://doi.org/10.1016/0967-0637(95)00044-7)
- Fujieki, L., F. Santiago-Mandujano, C. Fumar, R. Lukas, and M. Church. 2015. Hawaii Ocean Time-series Program Data Report 24: 2012. In U. Hawaii [ed.]. http://hahana.soest.hawaii.edu/hot/reports/rep_y24.pdf
- Gattuso, J.-P., J.-M. Epitalon, and H. Lavigne. 2014. seacarb: Seawater carbonate chemistry. R package version 3.0. <https://CRAN.R-project.org/package=seacarb>
- Gattuso, J.-P., and others. 2015. Contrasting futures for ocean and society from different anthropogenic CO₂ emissions scenarios. *Science* **349**. doi:[10.1126/science.aac4722](https://doi.org/10.1126/science.aac4722)
- Gray, S. E. C., M. D. DeGrandpre, T. S. Moore, T. R. Martz, G. E. Friederich, and K. S. Johnson. 2011. Applications of in situ pH measurements for inorganic carbon calculations. *Mar. Chem.* **125**: 82–90. doi:[10.1016/j.marchem.2011.02.005](https://doi.org/10.1016/j.marchem.2011.02.005)
- Haber, F., and Z. Klemensiewicz. 1909. Concerning electrical phase boundary forces. *Z. Phys. Chem.* **67**: 385–431.
- Hammer, K., B. Schneider, K. Kulinski, and D. E. Schulz-Bull. 2014. Precision and accuracy of spectrophotometric pH measurements at environmental conditions in the Baltic Sea. *Estuar. Coast. Shelf Sci.* **146**: 24–32. doi:[10.1016/j.ecss.2014.05.003](https://doi.org/10.1016/j.ecss.2014.05.003)
- Hendriks, I. E., C. M. Duarte, and M. Alvarez. 2010. Vulnerability of marine biodiversity to ocean acidification: A meta-analysis. *Estuar. Coast. Shelf Sci.* **86**: 157–164. doi:[10.1016/j.ecss.2009.11.022](https://doi.org/10.1016/j.ecss.2009.11.022)
- Hofmann, G. E., and others. 2011. High-frequency dynamics of ocean pH: A multi-ecosystem comparison. *PLoS One* **6**: e28983. doi:[10.1371/journal.pone.0028983](https://doi.org/10.1371/journal.pone.0028983)
- Hofmann, M., and H. J. Schellnhuber. 2009. Oceanic acidification affects marine carbon pump and triggers extended marine oxygen holes. *Proc. Natl. Acad. Sci. USA* **106**: 3017–3022. doi:[10.1073/pnas.0813384106](https://doi.org/10.1073/pnas.0813384106)
- Jiang, Z.-P., and others. 2014. Application and assessment of a membrane-based pCO₂ sensor under field and laboratory conditions. *Limnol. Oceanogr.: Methods* **12**: 264–280. doi:[10.4319/lom.2014.12.264](https://doi.org/10.4319/lom.2014.12.264)
- Johengen, T., G. J. Smith, D. Schar, M. Atkinson, H. Purcell, D. Loewenstainer, Z. Epperson, and M. Tamburri. 2015. Performance demonstration for autonomous pH sensor technologies. UMCES Technical Report Series. Alliance for Coastal Technologies.
- King County. 2015. Marine Moorings: Seattle Aquarium. In King County, Department of Natural Resources and Parks and Marine and Sediment Assessment Group [eds.]. <http://green2.kingcounty.gov/marine-buoy/SeattleAquariumData.aspx>
- Kroeker, K. J., R. L. Kordas, R. N. Crim, and G. G. Singh. 2010. Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms. *Ecol. Lett.* **13**: 1419–1434. doi:[10.1111/j.1461-0248.2010.01518.x](https://doi.org/10.1111/j.1461-0248.2010.01518.x)
- Kroeker, K. J., F. Micheli, and M. C. Gambi. 2013. Ocean acidification causes ecosystem shifts via altered competitive interactions. *Nat. Clim. Chang.* **3**: 156–159. doi:[10.1038/nclimate1680](https://doi.org/10.1038/nclimate1680)
- Le Bris, N., and D. Birot. 1997. Automated pH-ISFET measurements under hydrostatic pressure for marine monitoring application. *Anal. Chim. Acta* **356**: 205–215. doi:[10.1016/S0003-2670\(97\)00533-3](https://doi.org/10.1016/S0003-2670(97)00533-3)
- Le Quéré, C., and others. 2015. Global carbon budget 2015. *Earth Syst. Sci. Data* **7**: 349–396. doi:[10.5194/essd-7-349-2015](https://doi.org/10.5194/essd-7-349-2015)
- Liu, X. W., Z. H. A. Wang, R. H. Byrne, E. A. Kaltenbacher, and R. E. Bernstein. 2006. Spectrophotometric measurements of pH in-situ: Laboratory and field evaluations of instrumental performance. *Environ. Sci. Technol.* **40**: 5036–5044. doi:[10.1021/es0601843](https://doi.org/10.1021/es0601843)
- 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](https://doi.org/10.1021/es200665d)
- Lueker, T. J., A. G. Dickson, and C. D. Keeling. 2000. Ocean pCO₂ calculated from dissolved inorganic carbon, alkalinity, and equations for K₁ and K₂: Validation based on laboratory measurements of CO₂ in gas and seawater at equilibrium. *Mar. Chem.* **70**: 105–119. doi:[10.1016/S0304-4203\(00\)00022-0](https://doi.org/10.1016/S0304-4203(00)00022-0)
- Martz, T. R., J. J. Carr, C. R. French, and M. D. DeGrandpre. 2003. A submersible autonomous sensor for spectrophotometric pH measurements of natural waters. *Anal. Chem.* **75**: 1844–1850. doi:[10.1021/ac020568l](https://doi.org/10.1021/ac020568l)
- Martz, T. R., J. G. Connery, and K. S. Johnson. 2010. Testing the Honeywell Durafet® for seawater pH applications. *Limnol. Oceanogr.: Methods* **8**: 172–184. doi:[10.4319/lom.2010.8.172](https://doi.org/10.4319/lom.2010.8.172)
- Martz, T. R., K. L. Daly, R. H. Byrne, J. H. Stillman, and D. Turk. 2015. Technology for ocean acidification research: Needs and Availability. *Oceanography* **28**: 40–47. doi:[10.5670/oceanog.2015.30](https://doi.org/10.5670/oceanog.2015.30)
- Mathis, J. T., and others. 2015. Ocean acidification risk assessment for Alaska's fishery sector. *Prog. Oceanogr.* **136**: 71–91. doi:[10.1016/j.pocean.2014.07.001](https://doi.org/10.1016/j.pocean.2014.07.001)
- Millero, F. J. 1979. Thermodynamics of the carbonate system in seawater. *Geochim. Cosmochim. Acta* **43**: 1651–1661. doi:[10.1016/0016-7037\(79\)90184-4](https://doi.org/10.1016/0016-7037(79)90184-4)
- Millero, F. J. 1995. Thermodynamics of the carbon dioxide system in the oceans. *Geochim. Cosmochim. Acta* **59**: 661–677. doi:[10.1016/0016-7037\(94\)00354-0](https://doi.org/10.1016/0016-7037(94)00354-0)
- Millero, F. J. 2007. The marine inorganic carbon cycle. *Chem. Rev.* **107**: 308–341. doi:[10.1021/cr0503557](https://doi.org/10.1021/cr0503557)
- Morris, A. W., and J. P. Riley. 1966. The bromide/chlorinity and sulphate/chlorinity ratio in sea water. *Deep-Sea Res. Oceanogr. Abstr.* **13**: 699–705. doi:[10.1016/0011-7471\(66\)90601-2](https://doi.org/10.1016/0011-7471(66)90601-2)
- Narita, D., K. Rehdanz, and R. S. J. Tol. 2012. Economic costs of ocean acidification: A look into the impacts on global shellfish production. *Clim. Change* **113**: 1049–1063. doi:[10.1007/s10584-011-0383-3](https://doi.org/10.1007/s10584-011-0383-3)

- Newton, J. A., R. A. Feely, E. B. Jewett, P. Williamson, and J. Mathis. 2015. Global Ocean Acidification Observing Network: Requirements and Governance Plan. GOA-ON.
- Nilsson, G. E., D. L. Dixon, P. Domenici, M. I. McCormick, C. Sørensen, S.-A. Watson, and P. L. Munday. 2012. Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function. *Nat. Clim. Chang.* **2**: 201–204. doi:10.1038/nclimate1352
- Pelletier, G., E. Lewis, and D. Wallace. 2015. CO2SYS.XLS: A calculator for the CO₂ system in seawater for Microsoft Excel/VBA. Wash. State Dept. of Ecology/Brookhaven Nat. Lab.
- Perez, F. F., and F. Fraga. 1987. Association constant of fluoride and hydrogen ions in seawater. *Mar. Chem.* **21**: 161–168. doi:10.1016/0304-4203(87)90036-3
- Pierrot, D., and others. 2009. Recommendations for autonomous underway *p*CO₂ measuring systems and data-reduction routines. *Deep-Sea Res. Part II* **56**: 512–522. doi:10.1016/j.dsr2.2008.12.005
- Punt, A. E., D. Poljak, M. G. Dalton, and R. J. Foy. 2014. Evaluating the impact of ocean acidification on fishery yields and profits: The example of red king crab in Bristol Bay. *Ecol. Modell.* **285**: 39–53. doi:10.1016/j.ecolmodel.2014.04.017
- R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Riley, J. P. 1965. The occurrence of anomalously high fluoride concentrations in the North Atlantic. *Deep-Sea Res. Oceanogr. Abstr.* **12**: 219–220. doi:10.1016/0011-7471(65)90027-6
- Robert-Baldo, G. L., M. J. Morris, and R. H. Byrne. 1985. Spectrophotometric determination of seawater pH using phenol red. *Anal. Chem.* **57**: 2564–2567. doi:10.1021/ac00290a030
- Sabine, C. L., and others. 2004. The oceanic sink for anthropogenic CO₂. *Science* **305**: 367–371. doi:10.1126/science.1097403
- Schuster, U., A. Hannides, L. Mintrop, and A. Körtzinger. 2009. Sensors and instruments for oceanic dissolved carbon measurements. *Ocean Sci.* **5**: 547–558. doi:10.5194/os-5-547-2009
- Seidel, M. P., M. D. DeGrandpre, and A. G. Dickson. 2008. A sensor for in situ indicator-based measurements of seawater pH. *Mar. Chem.* **109**: 18–28. doi:10.1016/j.marchem.2007.11.013
- Somero, G. N., J. M. Beers, F. Chan, T. M. Hill, T. Klinger, and S. Y. Litvin. 2016. What changes in the carbonate system, oxygen, and temperature portend for the Northeastern Pacific Ocean: A physiological perspective. *BioScience* **66**: 14–26. doi:10.1093/biosci/biv162
- 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
- Spitzer, P., and K. W. Pratt. 2011. The history and development of a rigorous metrological basis for pH measurements. *J. Solid State Electrochem.* **15**: 69–76. doi:10.1007/s10008-010-1106-9
- Sutton, A. J., and others. 2014. A high-frequency atmospheric and seawater *p*CO₂ data set from 14 open-ocean sites using a moored autonomous system. *Earth Syst. Sci. Data* **6**: 353–366. doi:10.5194/essd-6-353-2014
- Takahashi, T., and others. 2009. Climatological mean and decadal change in surface ocean *p*CO₂(*i*), and net sea-air CO₂ flux over the global oceans. *Deep-Sea Res. Part II* **56**: 554–577. doi:10.1016/j.dsr2.2008.12.009
- Tamburri, M. N., and others. 2011. Alliance for coastal technologies: Advancing moored *p*CO₂ instruments in coastal waters. *Mar. Technol. Soc. J.* **45**: 43–51. doi:10.4031/MTSJ.45.1.4
- Uppström, L. R. 1974. The boron/chlorinity ratio of deep-sea water from the Pacific Ocean. *Deep-Sea Res. Oceanogr. Abstr.* **21**: 161–162. doi:10.1016/0011-7471(74)90074-6
- Virmani, J. I., and P. M. E. Bunje. 2015. Incentivizing innovation for the oceans and beyond. *Mar. Technol. Soc. J.* **49**: 27–29. doi:10.4031/MTSJ.49.3.5
- Zeebe, R. E. 2012. History of seawater carbonate chemistry, atmospheric CO₂, and ocean acidification. *Annu. Rev. Earth Planet. Sci.* **40**: 141–165. doi:10.1146/annurev-earth-042711-105521

Acknowledgments

The authors thank the staff and researchers at the Monterey Bay Research Aquarium Institute and the Seattle Aquarium, the crew of the R/V *Kilo Moana*, B. Carter, C. Cosca, A. Fassbender, D. Greeley, R. Kiel, M. Langis, and N. Williams for technical assistance. This publication was primarily funded by the Wendy Schmidt Ocean Health XPRIZE with additional support from the Joint Institute for the Study of the Atmosphere and Ocean (JISAO), the NOAA Ocean Acidification Program, and the Pacific Marine Environmental Laboratory (PMEL). PMEL Contribution 4606 and JISAO Contribution 2016-03-57.

Conflict of Interest

None declared.

Submitted 12 December 2016

Revised 30 March 2017

Accepted 21 April 2017

Associate editor: Mike DeGrandpre