PERFORMANCE VERIFICATION STATEMENT
For the Sunburst SAMI-pH Sensor

TECHNOLOGY TYPE:	pH sensors
APPLICATION:	In situ estimates of pH for coastal moored deployments
PARAMETERS EVALUATED:	Response linearity, accuracy, precision and reliability
TYPE OF EVALUATION:	Laboratory and Field Performance Verification
DATE OF EVALUATION:	Testing conducted from December 2012 through July 2014
EVALUATION PERSONNEL:	T. Johengen, G.J. Smith, D. Schar, M. Atkinson, H. Purcell, D. Loewensteiner, Z. Epperson and M. Tamburri

NOTICE:

ACT verifications are based on an evaluation of technology performance under specific, agreed-upon protocols, criteria, and quality assurance procedures. ACT and its Partner Institutions do not certify that a technology will always operate as verified and make no expressed or implied guarantee as to the performance of the technology or that a technology will always, or under circumstances other than those used in testing, operate at the levels verified. ACT does not seek to determine regulatory compliance; does not rank technologies nor compare their performance; does not label or list technologies as acceptable or unacceptable; and does not seek to determine "best available technology" in any form. The end user is solely responsible for complying with any and all applicable federal, state, and local requirements.

This document has been peer reviewed by ACT Partner Institutions and a technology-specific advisory committee and was recommended for public release. Mention of trade names or commercial products does not constitute endorsement or recommendation by ACT for use.

Questions and comments should be directed to: Dr. Mario Tamburri
c/o Alliance for Coastal Technologies
Chesapeake Biological Laboratory
PO Box 38 / 146 Williams Street
Solomons, Maryland 20688, USA
Email: Tamburri@umces.edu

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Alliance for Coastal Technology (ACT) conducted a sensor verification study of in situ pH sensors during 2013 and 2014 to characterize performance measures of accuracy and reliability in a series of controlled laboratory studies and field mooring tests in diverse coastal environments. A ten week long laboratory study was conducted at the Hawaii Institute of Marine Biology and involved week long exposures at a full range of temperature and salinity conditions. Tests were conducted at three fixed salinity levels (0.03, 22, 35) at each of three fixed temperatures (10, 20, 30 ^{o}C). Ambient pH in the test tank was allowed to vary naturally over the first five days. On the sixth day the pH was rapidly modified using acid/base additions to compare accuracy over an extended range and during rapid changes. On the seventh day the temperature was rapidly shifted to the next test condition. On the tenth week a repeated seawater trial was conducted for two days while the temperature was varied slowly over the $10 - 30$ ^{o}C range.   Four field-mooring tests were conducted to examine the ability of test instruments to consistently track natural changes in pH over extended deployments of 4-8 weeks. Deployments were conducted at: Moss Landing Harbor, CA; Kaneohe Bay, HI; Chesapeake Bay, MD; and Lake Michigan, MI. Instrument performance was evaluated against reference samples collected and analyzed on site by ACT staff using the spectrophotometric dye technique following the methods of Yao and Byrne (2001) and Liu et al. (2011). A total of 263 reference samples were collected during the laboratory tests and between $84 - 107$ reference samples were collected for each mooring test. This document presents the results of the Sunburst SAMI-pH which measures pH_T spectrophotometrically using a dye-based colorimetric reaction.

The SAMI-pH operated continuously throughout the entire ten week lab test and generated 3140 useable measurements out of a total possibility of 3154 timepoints.   The total range of pH measured by the SAMI-pH was 6.273 to 8.441, compared to the range of our discrete reference samples of 6.943 to 8.502.The SAMI-pH showed some differences in accuracy across the various test solutions and pH ranges, as would be expected based on the specific set-up of the instrument in terms of the indicator dye used and its known operational range. For the complete Lab test, the mean of the differences between the SAMI-pH measurement and reference pH was -0.017 ±0.970 (N=263), with a total range of -1.159 to 0.140. The larger offsets only occurred during the rapid pH shifts (see Appendix 1) and comparisons for only the stable testing conditions had a mean difference from reference measurements of -0.010 ±0.058 (N=193), with a total range of -0.048 to 0.063. Instrument measurements conducted at ten weeks with the second seawater trial exhibited a similar level of accuracy with a mean difference from reference of -0.010 ± 0.033 (N=8) compared to 0.014 ± 0.003 (N=27) for measurements in the first week of the lab test.

At Moss Landing Harbor the field deployment test was conducted over 28 days with a mean temperature and salinity of 16.6 ^{o}C and 33. The measured pH range from our 84 discrete reference samples was 7.933 – 8.077 compared to a range of 7.798 – 8.150 reported by the SAMI-pH for 2564 observations conducted continuously at 15 minute intervals. The average and standard deviation of the measurement difference between the SAMI-pH and reference samples over the total deployment was 0.039 ±0.003 with a total range of -0.123 to 0.158. The useable data return for the deployment was 100%.

At Kaneohe Bay the field deployment test was conducted over 88 days with a mean temperature and salinity of 24.5 ^{o}C and 34.4. The measured pH range from our 101 discrete

reference samples was 7.814 – 8.084 compared to a range of 7.678 – 8.220 reported by the SAMI-pH for 4211 observations conducted continuously at 30 minute intervals. The average and standard deviation of the measurement difference between the SAMI-pH and reference samples over the total deployment was -0.014 ±0.016 with a total range in the differences of -0.068 to 0.025. The useable data return for the deployment was 100%.

At Chesapeake Bay the field deployment test was conducted over 30 days with a mean temperature and salinity of 5.9 ^{o}C and 12.8. The measured pH range from our 107 discrete reference samples was 8.024 – 8.403 compared to a range of 8.121 – 8.479 reported by the SAMI-pH for 1842 observations conducted continuously at 15 minute intervals during the first 20 days of operations. Following this date, the instrument reported 175 non-determined values and exhibited a noticeable increased offset so the remaining data was considered compromised. The average and standard deviation of the measurement difference between the SAMI-pH and reference samples for the period up to April 1 was 0.041 ±0.033 (n=69), with the total range of differences between -0.086 to 0.114. The reported data return for the deployment was 94% but again only 69% of the data was considered fully reliable.

At Lake Michigan the field deployment test was conducted over 29 days with a mean temperature and salinity of 21.2 ^{o}C and 0.03. Two SAMI-pH units were deployed at this test site, one equipped with MetaCresol Purple (mCP) as the indicator dye and one with Phenol Red (PR) as the indicator dye. The Phenol Red method was considered to be experimental for the company at the time and was being tested for freshwater applications. The measured range in ambient pH for the 98 reference samples was 8.013 – 8.526. The original range in pH reported by the SAMI-pH (PR) was 6.606 – 9.305 for 2564 observations conducted continuously at 15 minute intervals. After removing fifty-one observations that were identified as outliers based on a difference between the paired SAMI-pH units of more than two times the mean, the measured range was reduced to 7.658 – 8.687.   The average and standard deviation of the measurement difference between the SAMI-pH (PR) and reference samples for the edited dataset was 0.021 ±0.047 with a range of -0.095 to 0.177. The omitted data represents 1.7 % of the returned data but represents a more realistic summary of the expected accuracy and uncertainty for the technology.

The measured range of the SAMI-pH (mCP) samples was 7.813 – 8.405 based on 2665 observations conducted continuously at 15 minute intervals. The average and standard deviation of the measurement difference between the SAMI-pH (mCP) and reference samples over the total deployment was range was -0.162 ±0.045 and the total range of the differences was between -0.333 to -0.068. The useable data return for this unit was 100%.

A summary plot of the SAMI-pH versus dye reference pH for all sites indicated that the SAMI-pH responded consistently among all test sites which covered a salinity range of freshwater to full seawater and temperatures from 3-26 ^{o}C. As expected there was a consistent offset in results reported within the Great Lakes test for the SAMI-pH with MetaCresol Purple indicator dye compared with the unit operating with the Phenol Red dye indicator. The SAMI-pH (PR) results closely match the reference samples since ACT also used Phenol Red as the indicator dye for their analysis.

Lastly, it is worth emphasizing that the continuous 15 – 30 minute time-series provided by the test instrument was able to resolve a significantly greater dynamic range and temporal

resolution than could be obtained from discrete reference samples. There were no obvious changes in the differences between instrument and reference measurements during the duration of the mooring test, indicating that biofouling and instrument drift had not significantly affected measurement performance. Continuous in situ monitoring technologies, such as the Sunburst SAMI-pH, provide critical research and monitoring capabilities for helping to understand and manage important environmental processes such as carbonate chemistry and ocean acidification, as well as numerous other environmental or industrial applications.

BACKGROUND AND OBJECTIVES

Instrument performance verification is necessary so that effective existing technologies can be recognized and so that promising new technologies can be made available to support coastal science, resource management and ocean observing systems. To this end, the NOAA-funded Alliance for Coastal Technologies (ACT) serves as an unbiased, third party testbed for evaluating sensors and sensor platforms for use in coastal environments. ACT also serves as a comprehensive data and information clearinghouse on coastal technologies and a forum for capacity building through workshops on specific technology topics (visit www.act-us.info).

As part of our service to the coastal community, ACT conducted a performance verification of commercially available, in situ pH sensors through the evaluation of objective and quality assured data. The goal of ACT's evaluation program is to provide technology users with an independent and credible assessment of instrument performance in a variety of environments and applications. Therefore, the data and information on performance characteristics was focused on the types of information that users most need.

The fundamental objectives of this Performance Verification were to: (1) highlight the potential capabilities of particular in situ pH sensors by demonstrating their utility in a range of coastal environments; (2) verify the claims of manufacturers on the performance characteristics of commercially available pH sensors when tested in a controlled laboratory setting, and (3) verify performance characteristics of commercially available pH sensors when applied in real world applications in a diverse range of coastal environments.

In 2012 ACT completed a customer needs and use assessment on pH sensors. Scientists, resource managers, and other users of these technologies were asked about their current use or application of these instruments, their perceptions of limitations or problems with the technology, and the most important criteria they use when selecting a sensor or instrument package. The results of these assessments, conclusions from the 2005 Measurement of Dissolved Inorganic Carbon Speciation in Natural Waters workshop, and discussions with the Technical Advisory Committee were used to identify the main applications and key parameters that ACT evaluates in this Technology Verification.

Over 170 coastal resource managers, environmental health agency representatives, manufacturers, and scientific researchers were surveyed regarding their needs, uses and current practices of making pH measurements in the environment. Overall, 42 individuals responded to the survey, all of which are using, or manufacturing, in situ pH sensors. Respondents were allowed to select more than one answer when appropriate so percentages do not add up to 100%. The types of pH sensors being used varied between potentiometric (45%), ion-selective field-effect transistor (32%), and spectrophotometric (using indicator dye) 20%, but with a quarter of

respondents using more than one type. The greatest area of use among respondents was academic research (76%) followed by state and federal resource management (40%). The environmental applications were broad with coastal ocean greatest (74%), followed by estuarine (64%), followed by open-ocean (41%), and followed by freshwater (38%). Correspondingly, these environments included temperature ranges from -5 to 50 ^{o}C, with medians of 5 ^{o}C and 28 ^{o}C when responses were binned into low and high ranges. Similarly, salinities ranged from 0 – 100, with low and high bin medians of 15 and 35, respectively. The range of pH measured by the respondents in these applications was between 4.0 and 11.0, with low and high bin median values of 7.0 and 8.3. Remote deployment was the most common method of use (74%), followed by depth profiling (50%), then hand-held portable use (48%), then flow-through systems (26%). Respondents used a variety of calibration procedures including commercial buffers (68%), CO_2 chemistry (35%), seawater CRMs (23%), pH indicator dyes (18%), and supplied by manufacturer (13%). The four areas where respondents expressed the greatest concern over the use of in situ pH sensors were ruggedness (49%), calibration life (46%), level of measurement uncertainty (43%), and reliability (41%). The complete needs and use assessment reports can be found at:
http://www.act-us.info/Download/Customer_Needs_and_Use/pH/index.html

INSTRUMENT TECHNOLOGY TESTED

The Submersible Autonomous Moored Instrument for pH, SAMI-pH, is Sunburst Sensors' highly accurate colorimetric reagent instrument for in situ pH measurements. The SAMI-pH measures and logs pH in marine environments over long periods with ultra-low power and reagent consumption, providing valuable time-series data to researchers around the world. It can power and log data from up to three external instruments such as PAR, dissolved oxygen, chlorophyll fluorometer, or CTD. One of these instruments can be RS232. The SAMI-pH can be deployed to depths of up to 600 m. The SAMI-pH's log their own data, but can be configured to communicate with the Seabird Inductive Modems or external data loggers. The SAMI-pH measures pH$_T$ (total hydrogen scale) in the marine pH range of 7-9. It does not suffer from the drift that plagues most electrode based pH probes. The client software parses the data at the time of download into a text file that contains the configuration, data records and status records. The user can then open this raw file and the software will calculate the pH based on the salinity value entered by the user of through an attached CTD.

No calibration of the SAMI-pH was performed by ACT staff. Sunburst helped with the post processing of the data after being supplied with only our final salinity time-series. For the freshwater Great Lakes field deployment test, two SAMI units were tested. One contained the normal MetaCresol Purple dye indicator and the other contained Phenol Red as the dye indicator. The use of Phenol Red as the dye indicator for freshwater applications is still considered experimental.

pH SCALES

Four pH scales are commonly used to describe the acidity of an aqueous solution: (1) the free hydrogen ion concentration scale, (2) the total hydrogen ion concentration scale, (3) an H$^+$ concentration scale termed the 'seawater scale' that is numerically quite similar to the total

hydrogen ion concentration scale, and a fourth pH scale that is expressed in terms of an 'apparent activity'. The relationship between the apparent activity scale and the other three scales is not thermodynamically well defined.

The free hydrogen ion concentration scale is expressed as pH = -log [H⁺], where [H⁺] indicates the concentration of a free hydrated hydrogen ion. The total scale (pH_T) is written as pH_T = -log [H+]$_T$ where [H⁺]$_T$ ≈ [H⁺] + [HSO$_4^-$], and the seawater scale pH (pHsws) is written as pHsws = -log[H⁺]$_{sws}$ where [H⁺]$_{sws}$ ≈ [H⁺] + [HSO$_4^-$] + [HF^o].

The various pH scales are inter-related by the following equations:

$$a_H = 10^{-pH(NBS)} = f_H * [H^+]/(1 \text{ mol/kg}), \text{ and} \tag{1}$$

$$[H^+] = [H^+]_{tot} / (1 + {_T}S/K_{SO4}) = [H^+]_{sws} / (1 + {_T}S/K_{HSO4} + {_T}F/K_{HF}) \tag{2}$$

where a_H is an apparent activity, f_H is an apparent activity coefficient for the H⁺ ion, $_T$S and $_T$F are total concentrations of SO$_4^{2-}$ and fluoride ions in seawater, and K$_{HSO4}$ and K$_{HF}$ are HSO$_4^-$ and HF^o dissociation constants in seawater.

Because electrodes transferred between dilute buffers (e.g., NBS buffers) and concentrated electrolytes (such as seawater) are subject to variations in the liquid junction potentials of reference electrodes (an effect whose magnitude is generally specific to a particular electrode), the f_H proportionality factor that relates the apparent activity scale to concentration scales is, in part, influenced by factors that are not strictly thermodynamic. Consequently, use of the apparent activity scale is generally discouraged for measurements in seawater.

Relationships between the different concentration scales depend on temperature, salinity, and pressure. At 25°C, salinity 35, and 1 atm, relationships between seawater scale, total scale and free scale pH (pH$_{sws}$, pH$_T$ and pH) are given approximately as:

$$pH \text{ minus } pH_T = 0.108 \tag{4}$$

$$pH_T \text{ minus } pH_{sws} = 0.008 \tag{5}$$

$$pH_{NBS} \text{ minus } pH_{sws} \approx 0.13 \tag{6}$$

Equation (4) was derived using the K$_{SO4}$ characteristics given in DOE (1994). Alternative descriptions of the HSO$_4^-$ dissociation equilibrium can produce significant variations in the offset between total scale and free scale pH. Determinations of K$_F$ in seawater are sufficiently coherent that deviations from the relationship given in eqn. (5) are quite small. The relationship between pH$_{NBS}$ and pH$_{sws}$ (eqn. 6) is based on the discussion of Culberson (1981). It is important to note that f_H, and therefore eqn. (6), is electrode-dependent. Measurements of a_H on the NBS scale are expressed in terms of mol/kg-H$_2$O, while the other scales express concentrations in terms of mol/kg$_{(SW)}$.

Spectrophotometric measurements of pH in seawater using purified meta-cresol purple are generally expressed on the total H⁺ scale. Using Tris seawater buffers, glass electrodes are also commonly calibrated on the total scale (pH$_T$).

There are significant changes in junction potentials when electrodes are calibrated in dilute media and then utilized for measurements in high ionic strength media such as seawater. In such cases there is no direct way to relate pH_{NBS} values to pH_T or pH_{SWS} values unless electrodes are calibrated in both dilute buffers (NIST buffers) and also artificial seawater buffers such as Tris.

SUMMARY of VERIFICATION PROTOCOLS

The protocols used for this performance verification were developed in conference with ACT personnel, the participating instrument manufacturers and a technical advisory committee. The protocols were refined through direct discussions between all parties during two pH Sensor Performance Verification Protocol Workshops held on 18-20 June, 2012 and 26-28 June, 2013 in Ann Arbor, MI. All ACT personnel involved in this verification were trained on use of instruments by manufacturer representatives and on standardized water sampling, storage, analysis and shipping methods during a training workshop held on 3-6 Dec, 2012 in Kaneohe, HI. The manufacturer representatives and the ACT Chief Scientist verified that all staff were trained in both instrument and sample collection protocols.

This performance verification report presents instrument output in derived pH values reported over time as directly downloaded from the test instruments or captured through independent dataloggers. A summary of the testing protocols is provided below. A complete description of the testing protocols is available in the report, *Protocols for the Performance Verification of In Situ pH Sensors* (ACT PV12-01) and can be downloaded from the ACT website (www.act-us.info/Download/Evaluations/pH/Protocols/index.html).

Analysis of Reference Samples

The reference pH measurements were the pH of discrete water samples as determined by pH indicator dyes, either, meta-cresol purple or phenol red (freshwater) (Liu et al. 2011, Yao and Byrne 2001 respectively). All dyes used throughout the Verification were prepared and characterized by Bob Byrne at the University of South Florida. An initial reading for background correction was taken on each cell and then dye (from Byrne's lab) was introduced into each cell (10 µl of purified 10 mM meta-cresol purple for samples with salinity > 5 and 10 µl of 10 mM phenol red for freshwater samples), mixed thoroughly, and the dye R ratio measured on an Agilent 8453 spectrophotometer (the R ratio is the ratio of absorbance measured at wavelengths corresponding to the peak absorbance of the acid and base forms of the dye).

The R value of each cell was measured a minimum of 5 times in succession in the field test and 10 times in succession for the lab test. pH was calculated from the R ratio, temperature and salinity according to the equations published in Liu et al. 2011 (meta-cresol purple) or Yao and Byrne 2001 (phenol red); both pH total scale and pH free scale are reported (Liu et al. 2011, Eq. 12). The actual sample pH at ambient temperature was calculated using CO_2-Sys with dissociation constants from Millero et al. (2006) and an estimated value of the alkalinity that is based on the average of the samples analyzed by SIO. The anticipated level of accuracy of these measurements is better than 0.01 pH for seawater ($S \geq 20$), but actual laboratory based accuracy and precision levels were quantified and reported for the brackish and freshwater environments.

Laboratory Test

Two thermally insulated, covered, 4.5 m^3 "source-water tanks" were filled with 1 µm filtered seawater, or a mixture of freshwater and filtered seawater. The two tanks were then isolated so that each could be maintained at a specific temperature. The source-water tanks were used to supply water into a third smaller "test tank" (capacity - .8 m^3), where all instrument performance measurements took place. One tank with source-water was used to continuously flush the third test tank (containing the in-situ instruments); and water from the second source tank was used to create a quick transition to a new temperature condition within the test-tank. Test conditions within the source-water tank were set and equilibrated for several days prior to delivery into the test tank. The large volume of equilibrated water in the source-water tank allowed for a rapid transition (10-15 minutes) of temperature and salinity conditions in the test tank. Temperature was maintained within the source water tank to ± 1 ^{o}C using an AquaLogic MT-3 circulating heat exchanger. Water in both of the tanks was mixed continuously with several submerged bilge pumps. Evaporation and heat exchange through the water surface was reduced to a minimum by using a covering on the surface of the water. The test tank was instrumented with the test instruments, as well as three factory calibrated RBR temperature recorders (accurate to 0.02 ^{o}C) placed near the instruments to continuously measure actual temperature conditions experienced by the test instruments. These data were used to help evaluate fine scale variability within the test tank and to correct for temperature offsets that might exist during pH measurement of discrete reference samples.

The test tank pH was also monitored continuously with two glass pH electrodes (Metrohm ECOTRODE PLUS 6.0262.100) measured to 0.1 mv, and spaced across the span of the test instruments. These data were used to create a continuous data record of pH within the tank, and to confirm test conditions during acid/base additions. These pH data will not be used as reference pH data to calculate instrument offsets. The pH probes were calibrated against the dye estimated values obtained on test tank samples during acid-base additions (at the fixed experimental T-S conditions) to get slope responses over a pH range of approximately 7.1 to 8.3. In this way the electrodes did not experience any change in liquid junction potential from either freshwater or saltwater buffers (Easley and Byrne 2012).

Each week testing was conducted at a set combination of temperature and salinity (T-S). Nominal temperature conditions were set for 10, 20 and 30°C, and salinity conditions were set for nominally 0, 20 and 35 S. A week-long test was performed at each T-S combination. After 4 to 6 days of testing at a stable T-S condition and ambient pH, pH was cycled over a reasonable range using acid-base additions to the water of the test tank (7.5 to 8.5 for seawater and 6.5 to 8.8 for freshwater). Two, raised - lowered pH cycles were conducted at each T-S condition over the course of one day. Acid/base additions were done by first mixing known quantities of acid/base into several liters of the current test solution and then adding this solution into the test tank to facilitate mixing and rapid equilibration.

The sequencing of tests was to start with a fixed salinity and the tests were performed for that salinity at the three different temperatures, starting at 10 ^{o}C and increasing sequentially up to 30 ^{o}C. In this way we were able to use the same source water for all three temperature conditions. Each test was scheduled to last 1 week, except when delays were needed to accommodate work schedules. For each new T-S test condition, the test tank was flushed and

filled with new source water while keeping all test instruments submerged and recording continuously. Each new batch of source water was filtered through a 1 μm cartridge prior to use in the tests.

Laboratory Test Reference Sampling - During the stable temperature and salinity period, reference samples were collected and analyzed at 4 timepoints each day. In addition, on one day during the stable cycle, a burst sampling of five independent measurements were collected 5 minutes apart at one of the timepoints to evaluate the repeatability and uncertainty of the reference measurement with respect to the variability observed within the test tank. Reference samples were drawn directly into the spectrophotometer cell from a supply tube within the test tank. After the cell was blanked and prior to the dye addition and spectrophotometric readings, the temperature of the cell was measured and adjusted to within 0.1 degree or less of the tank temperature. 10 readings of the reacted dye were recorded over 1-2 minutes for each sampling timepoint, and a mean and standard deviation of each reference measurement was calculated. The final temperature of the sample in the cell was recorded immediately after reading to define any deviation from the specified reading temperature. The temperature of the sample stream was monitored continuously and the salinity of the sample stream was measured on duplicate sub-samples with a Guideline Portasal. Once during the stable phase of a given T-S test condition a more detailed burst sampling was conducted, where 6-8 samples were collected over the course of an hour to assess the fine-scale variability (temporal and spatial) of test tank conditions. During the pH cycling phase of the test, 2 reference sampling timepoints were completed during each endpoint (approximately an hour after acid/base addition) of the increase or decrease. Each phase (increase or decrease) occurred over roughly 2 hour intervals.

Lastly, duplicate water samples were collected at the beginning and end of the stable test condition period (day 1 and day 5) to characterize all of the CO_2 parameters. Water samples were taken for pCO_2, Dissolved Inorganic Carbon and Total Alkalinity measurements and preserved and stored in accordance with standard oceanic protocols (Dickson et al., 2007). These water samples were shipped to Oregon State University (Burke Hales, OSU) and Scripps Institute of Oceanography (Andrew Dickson, SIO) for analysis. In addition, preserved water samples were shipped to SIO (Andrew Dickson) and University of South Florida (Robert Byrne, USF) for measurements of pH, thereby providing further indication of any possible uncertainty of ACT's reference sample measurements. However, because these samples were not handled identically to the reference samples analyzed real-time and directly from the test tank, they will not be used in any direct assessment against the test instruments without review from the TAC and consent from the manufacturer. Nutrients were also measured on these samples in Atkinson's lab.

Moored Field Tests

Moored field tests were conducted at four ACT Partner Institution sites covering freshwater, estuarine, and open-ocean conditions. The test sites include Moss Landing Marine Laboratories (MLML), CA; Hawaii Institute of Marine Biology (HIMB), HI; Chesapeake Biological Laboratory (CBL), MD; and Lake Michigan, MI. The field tests occurred in sequence, starting with Moss Landing on August 22, 2013. The duration of the moored deployment tests were 4 weeks, except at the Hawaii test site where instruments were deployed

for a period of 12 weeks. Instruments were removed from the water only after the test period was complete.

Instrument Setup – Prior to deployment, all instruments were set up at the field site by a trained ACT staff member following established protocols that had been developed in collaboration with the manufacturers at the training session that occurred in December, 2012 (at the beginning of the ACT laboratory test). Manufacturers that did not conduct an in-person training session sent written protocols and video demonstration of desired handling procedures. As appropriate, and where requested by the manufacturer, submitted test instruments were initially calibrated by ACT staff using a common batch of commercial, non-dyed, NIST traceable buffers at a pH of 7.02 and 10.06. (Each sensor was calibrated with buffer from its own bottle, but they were all from the same LOT and shipment). All calibrations occurred in a constant temperature bath at 20.0^{o}C (except at HI which was done at 25°C). Any deviation from this temperature was noted and used to adjust the stated pH values of the buffers as defined on the bottles.

After calibration and prior to deployment (again where appropriate for each instrument) instruments were exposed to additional certified reference solutions. At Moss Landing, CBL and Hawaii, instruments were exposed to batches of Seawater CRMs produced by the CO2-QC facility at Scripps Institution of Oceanography. (CRMs were batched into a single container before distributing into individual calibration cups for each sensor. Three cuvettes were filled immediately upon mixing the batch and again at the end of pouring out the solution to measure pH and ensure consistency of each aliquot.) For the freshwater test in Lake Michigan instruments were exposed to the NIST buffers or lake water characterized by spectrophotometric analysis by ACT personnel as appropriate for the instrument. All pre-deployment instrument measurements of the reference buffers were in a water bath at a constant temperature (20 ^{o}C, except at HIMB which were at 25 ^{o}C) and all solutions and instruments were pre-equilibrated at this temperature.

For the deployment, instruments were programmed to record data based on a time interval that allowed for a 30 day (or 90 day for HI) deployment. Intervals were selected such that there was a common 30 (or 60) minute interval achieved by all instruments. This schedule allowed us to coordinate our reference sampling for all instruments. Internal clocks were set to local time and synchronized against the time standard provided by www.time.gov. In high flow coastal environments, clock drift could lead to significant bias.

Instrument Deployment – A photograph of each individual instrument and the entire instrument rack was taken just prior to deployment and just after recovery to provide a qualitative estimate of biofouling during the field tests. Instruments were set-up as self-recording on a deployment rack and arranged so that a single representative field sample could be collected within 1 meter of any individual sampling inlet. The deployment frames were arranged so that all of the instruments remain at a fixed depth of 1 m below the water surface (using a float system or fixed dock in environments not affected by tidal changes or strong wave action). Two calibrated SeaBird CTD packages and four RBR Solo thermistors were attached to the mooring at each test site in order to accurately characterize the temperature and salinity heterogeneity surrounding the mooring. In these dynamic coastal regions it was critical to understand the spatial and temporal dynamics of the water column in order to interpret pH measurements appropriately.

Instrument Retrieval – After the instruments were retrieved and cleaned of all removable fouling according to written procedures provided by the manufacturer, a final exposure test was conducted in the CRM buffers as defined above. For instruments that pumped samples through an inlet, the inlet was connected to the manufacturer's supplied storage bag to ensure that no air was introduced prior to exposure in the buffers.

Reference Water Sampling Schedule – The sampling frequency was structured to examine changes in pH over daily and weekly time scales. Specifically, an intensive sampling event was conducted once a week that consisted of 5 sample collections within a day. During four additional days of each week, there was a minimum of twice per day sampling, scheduled in a manner to capture as much diurnal variation as possible. The initial intensive sampling event occurred within the first two days of the deployment after all instruments had been deployed, and the final intensive sampling event occurred during the last two days of the deployment. The schedule provided a higher density of comparative data at the beginning when instruments should have been functioning at optimum performance and again after the challenge of a four or twelve week deployment. The sampling schedule resulted in at least 60 reference samples paired with instrument measurements. For the 12 week deployment test at HIMB, the sampling scheme was modified to spread out a similar number of samples over the extended time period. All sampling times were recorded on logsheets and entered into a database for final data comparisons.

Reference Water Sample Collection – A standard 2.2 or 4.2L Van Dorn water sampler was used at each field test site to collect water samples for reference pH measurements. Water sample collections were timed to correspond directly with the instrument readings or sample intake. The water sampler was lowered to the same depth of the instrument sampling inlets, and as close as physically possible to the inlets (distant by no more than 0.5 m horizontal distance). The water sampler was soaked at sampling depth for 1 minute prior to sampling. If water was not flowing, the sampler was moved to ensure that it was flushed with the ambient water. The water sampler was triggered to match the programmed sampling times of each instrument. Three replicate pH samples were collected in clean 10 cm glass spectrometer cuvettes from each individual field sample. The cuvettes were gravity filled from PTFE tubing connected to the sampling bottle spigot such that the volume of the cuvette was exchanged three times (about 20-30 secs) before a final sample is collected. Care was taken to ensure that no bubbles were left within the cuvette before sealing the ports with their PTFE stoppers. Cuvettes were stored at ambient or slightly cooled conditions during transport to the laboratory (travel times at various field sites ranged from 5 to 20 minutes) until they were placed into the temperature equilibration chamber at the specified temperature for analysis. In addition, at the freshwater site an additional 300 ml BOD bottle was filled to allow for a lab-based electrode pH measurement on each field reference sample. All samples were equilibrated to 25°C and analyzed on a lab electrode calibrated daily before use.

Twice a week (day 1 and day 5) duplicate water samples were collected to characterize all of the CO_2 parameters. (This sample collection occurred in conjunction with our field duplicate sampling and one set of external partner samples came from each Van Dorn. This protocol was mostly to facilitate limitations of volume but also helped to evaluate heterogeneity at the mooring.) Water samples were collected and preserved for pH, pCO_2, TCO_2 and TA measurements following standard oceanic protocols (Dickson et al., 2007). In brief, a 500 mL

glass sample bottle (supplied by Andrew Dickson, SIO) was filled from the Van Dorn sampling bottle using PTFE tubing and preserved with 110 µL of saturated $HgCl_2$ solution. Reference samples shipped to Oregon State University (Burke Hales, OSU) were analyzed for pCO_2 and TCO_2. Reference samples shipped to Scripps Institution of Oceanography (Andrew Dickson) were analyzed for TA and pH.   Reference samples shipped to the University of South Florida (Robert Byrne) were analyzed for pH. These external sample analyses provided further estimation of any uncertainty of ACT's reference sample measurements through both direct pH measurement and calculated pH from carbonate parameters using CO_2-Sys (Pierrot et al. 2006). However, because these samples were not handled in a manner identical to the unpreserved reference samples that were analyzed within a few hours of collection, they will not be used in any published direct assessment against the test instruments without agreement of the TAC.

All sampling procedures were practiced by ACT staff prior to the beginning of the evaluation, and the Chief Scientist verified that all staff were trained in both instrument and sample collection protocols. A preservation test was done at each site to ensure samples that were shipped for external analysis were stable.

Sample Handling and Custody Requirements - All collected reference water samples for all four field tests were handled in the same manner. All reference samples were processed according to methods defined within these protocols and cited Standard Operating Procedure documents (Dickson et.al. 2007). Each reference sample was dated and coded according to site, test condition and sample sequence. The actual sample container was labeled with a number for identification. The reference sample number was used in all laboratory records and Chain-of-Custody (COC) forms to identify the sample. The COC specified time, date, sample location, unique sample number, requested analyses, sampler name, required completion time, date and time of sample transaction, and name of receiving party for the samples. Proper labeling of sample bottles was critical. The COC was a mechanism by which samples could be tracked through the various phases of the process: collection, shipping, receiving, logging, sample prep/extraction, analysis and final data QA/QC review. Transfer of reference samples from field personnel to laboratory personnel was also recorded on the COC and records were maintained in the laboratory with the names and signatures of persons leaving and receiving the custody. All logs were duplicated weekly. The original log was retained at the ACT Partner site and a copy was sent to the ACT Chief Scientist. Accumulated samples to be analyzed by outside laboratories were shipped for analysis at the end of the month long deployment tests, and monthly for the extended HI deployment.   Samples stored on site were routinely inspected by ACT personnel to assure proper preservation and label integrity.   All reference samples not immediately analyzed on site by ACT staff were accompanied by the sample collection sheet and COC forms.

Analytical Methods for Reference Samples - Three spectrometer cells (10 cm cylindrical cell) were filled at the deployment site for each field reference sample, and transported directly to the lab. All analysis was done at a fixed temperature for a given test site. The measurement temperature was 25°C for HIMB, 20°C for MLML, 15°C for CBL and 25°C for Lake Michigan. The temperatures for CBL and Lake Michigan were estimated to be near the mid-point of the expected 30 day ambient range. Filled cells were incubated in a cell warmer (manufactured at USF according to Bob Byrne's specifications) to reach the specified analytical temperature (typically between 30 min to 1 hour). The Agilent had a thermal jacket surrounding the

spectrophotometer cells that was continuously flushed with the same water bath that also supplied the cell warmer. Past experience at Byrne's lab has shown that this system can maintain cells at a constant temperature within $\pm$ 0.1 ^{o}C or better. After the initial dye reading, the sample was re-blanked and then a second 10 µl aliquot of dye was added and the R-ratio re-measured to enable a correction for the effect of the dye addition on the sample pH. By performing the perturbation measurement on each sample we could directly calculate the appropriate adjustment for each sample individually. The final temperature of the solution in the cuvette was measured with a bead thermistor upon completion of the second dye reading and recorded on the datasheets to define any deviation from the specified reading temperature. In addition, at the freshwater site an unpreserved sample collected in a 300 ml BOD bottle was incubated at the same temperature as the equilibration chamber and cell jacket until it reached a constant temperature of 25°C, and the pH was subsequently read on a recent two-point calibrated Metrohm electrode. The electrode was calibrated daily at 25.0 ^{o}C.   After each use the Van Dorn sampler, fill tubing and cuvettes were thoroughly rinsed with deionized water to prevent any buildup of salts or dye.

Ancillary Environmental Data - At each of the mooring test sites, two calibrated CTD packages were attached to the test rack and positioned to best characterize the salinity surrounding the mooring. The CTDs provided an independent record of conductivity and temperature measured at 15 minute intervals. In addition, four RBR Solo temperature sensors were deployed at each site to establish an accurate temperature history for the site. The RBR sensor has a stated accuracy and resolution of 0.002°C and 0.00005°C respectively. In conjunction with each water sample collection, technicians recorded basic site-specific conditions on standardized log sheets including: date and time, weather conditions (e.g., haze, % cloud cover, rain, wind speed/direction), air temperature, recent large weather events or other potential natural or anthropogenic disturbances, tidal state and distance from bottom of sensor rack, and any obvious problems or failures with the instruments. Datasheets were transmitted on a weekly basis to the ACT Chief Scientist for data archiving and ACT personnel performance QA/QC.

Each test site either established or identified the closest meteorological station (and river discharge gauge where appropriate) that could record air temperature, humidity, directional wind speed, precipitation on a continuous basis to help identify the timing and intensity of any event based changes at the field test locations.

Ancillary data was used in a qualitative sense to understand the history of weather patterns and changes in ambient water quality conditions. These data were not used for any direct calibration, correction, or statistical comparison to the reported salinity test data.

RESULTS of LABORATORY TEST

Instruments were tested in an extensive 10 week long laboratory tests under tightly controlled temperature and salinity conditions to examine measurement accuracy, precision, and reliability. Temperature and salinity conditions were maintained within large (4.5m$^{3)}$ storage tanks and the water was constantly circulating into the instrument test tank (0.8 m^3) except during sampling timepoints (Photo 1).
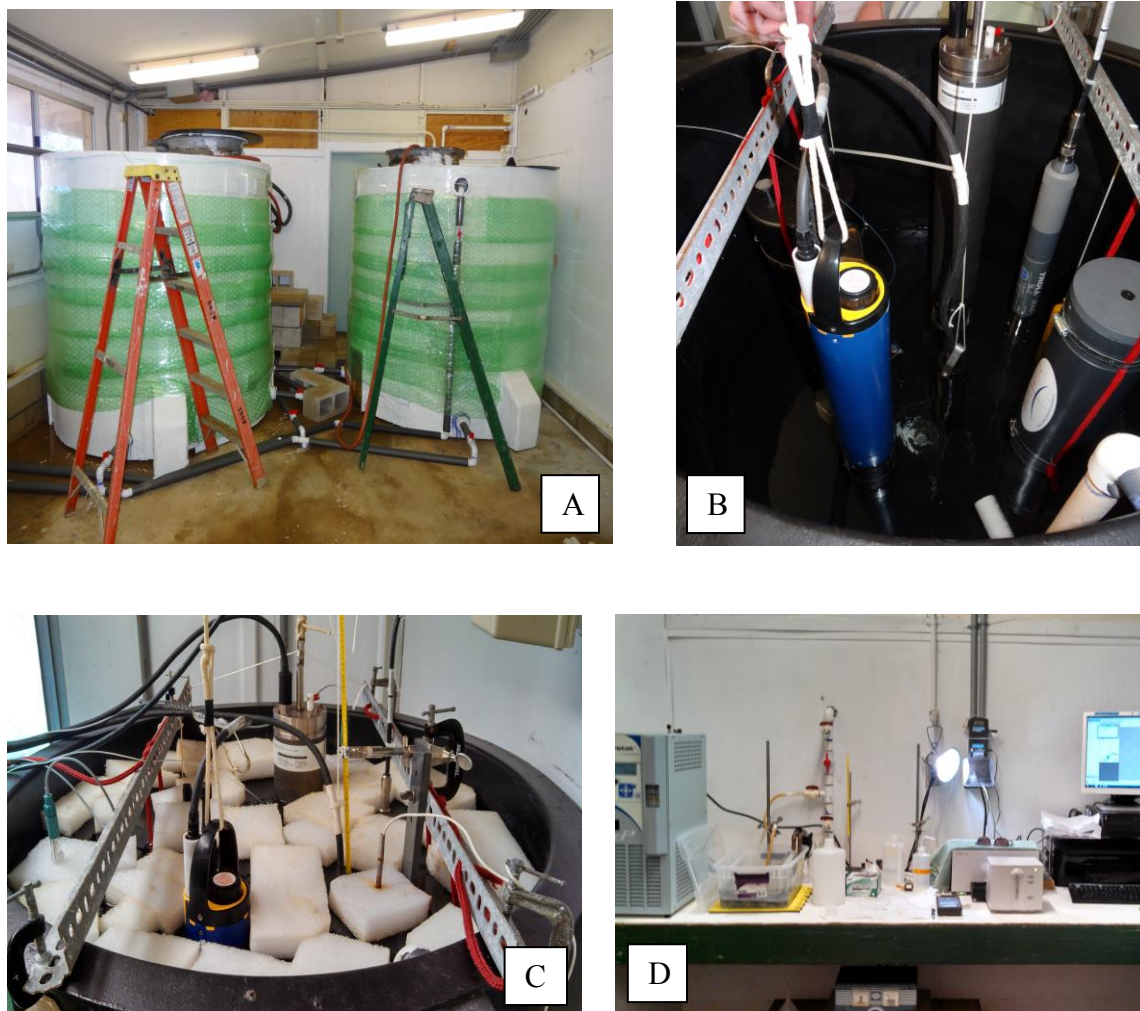
Photo 1. A. The two main source water tanks (4.5 m^3) used to maintain specific salinity and temperature conditions for each trial of the laboratory test. Water was continuously circulated between the source tank and the instrument testing tank in between sampling timepoints. B. Test instruments being deployed in the test tank (750 L). Water was continuously circulated internally within the tank during sampling events. The shown PVC pipe served as the sampling stream inlet and was diverted directly into the laboratory for reference sample withdrawal. Water returns to the test tank at a bottom opening below the sensors. C. Fully instrumented test tank with all company instruments, reference thermistors, and reference pH electrodes. Foam blocks were floated on the surface to minimize evaporation and help maintain temperature set points. D. Laboratory set-up for the dye-based spectrophotometric analysis of pH. The diversion tube coming through the wall from the test tank was used to draw off reference samples. A water bath was used to keep the cuvette cells at the test tank temperature during filling.

Test-tank water was circulating into the laboratory through a bypass tube for reference sample withdrawal. The sequencing of temperature and salinity test conditions is shown in Fig. 1. Each given temperature and salinity condition was maintained at stable conditions for a week. Rapid pH shifts were performed on the sixth day of the week (see Appendix 1) and temperature transitions (usually within 15 minutes unless going into a break) occurred at the end of day seven.

For saltwater trials, salinity averaged 35.19 ±0.03, and temperatures were varied sequentially over the following three ranges 20.3 ±0.2, 10.1 ±0.1, and 29.9 ±0.1 ^{o}C. For brackish water trials, salinity averaged 21.80 ±0.03, and temperatures were varied sequentially over the following three ranges 10.1 ±0.1, 20.2 ±0.1, and 30.4 ±0.1 ^{o}C. For freshwater trials, salinity averaged 0.26 ±0.05 and temperatures were varied sequentially over the following three ranges 10.1 ±0.1, 20.2 ±0.1, and 30.4 ±0.1 ^{o}C. At the end of the nine week-long trials, a new batch of seawater was introduced and temperatures were varied between 30 and 10 ^{o}C over the course of two days.
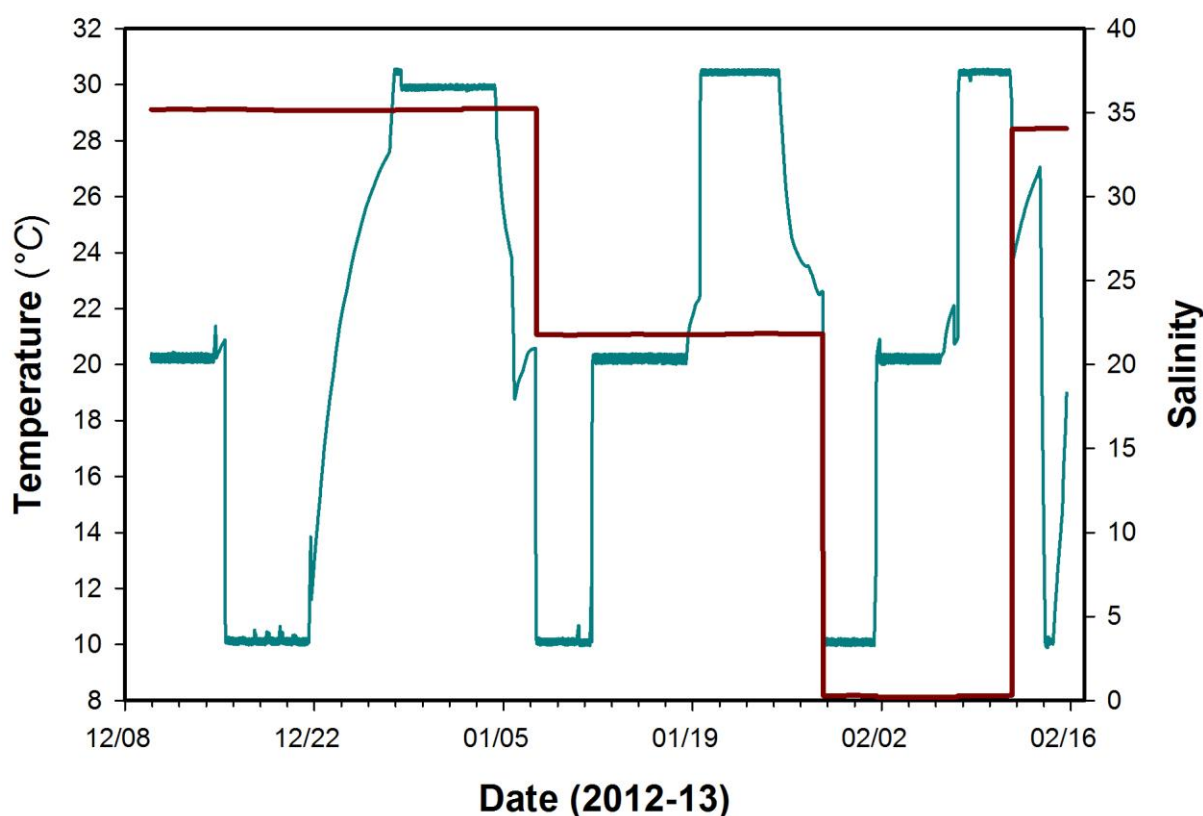
Figure 1. Temperature (Teal) and salinity (Red) conditions maintained throughout the ten week long laboratory test. Sensors were tested at three temperatures (10, 20, 30 ^{o}C) at each of three salinities (0.3, 22, 35). Full seawater salinity conditions were briefly re-established at the end of the test and cycled through the same temperature range to fill in missing instrument coverage.

As defined in the test protocols, two Metrohm electrodes were maintained within the test tank throughout the lab study to provide a continuous, independent measure of pH at 15 minute intervals. These results were not intended to be used as reference values for direct instrument comparison, but because they were calibrated weekly with water from the test tank water against the dye reference pH and they maintained their calibration in between intervals (Fig. 2; slope = 1.006, $r^2 = 0.999$), they provide a high resolution (15 minute) continuous time series of ambient pH within the test tank. Therefore, Metrohm results are included within some of the time series plots but no quantification of instrument accuracy was conducted with these data.
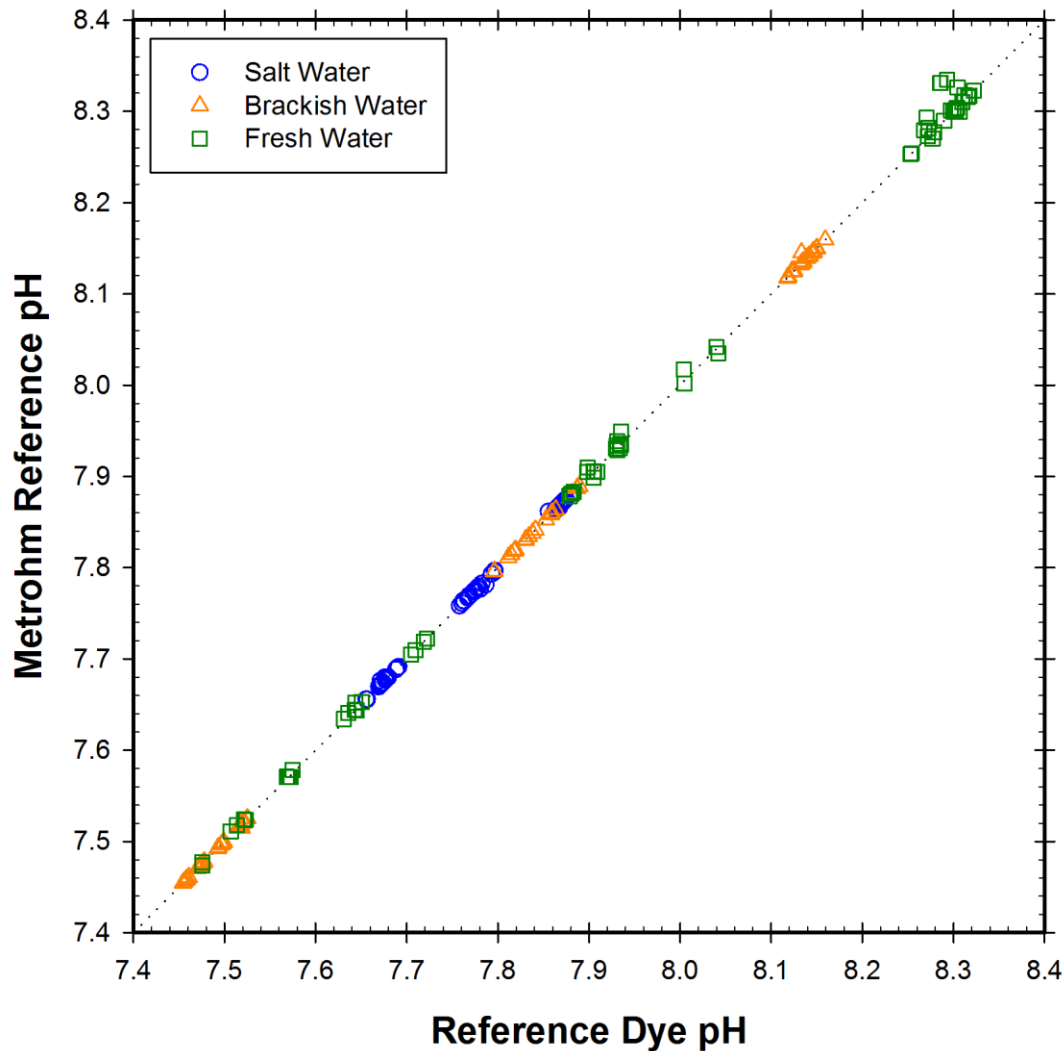
Figure 2. Cross plot of matched Metrohm pH results that measured continuously within the test tank at 15 minute intervals versus dye measured reference pH results. Three-hundred, fifty-one dye based pH measurements were made over the ten week test.

Once a week multiple reference samples were collected from the test tank in a burst, i.e. sampled as rapidly as possible to examine heterogeneity within the test tank and repeatability of the reference pH results. The average standard deviation of the reference pH value for ten sampling events was 0.0025 with a range of 0.0055 to 0.0005 (Table 1).

Table 1. Results of burst sample analysis for the Lab Test to demonstrate consistency of test tank conditions and repeatability of dye pH measurements.

Date	Measurement Temperature (^{o}C)	Salinity	Time Interval (min) (# Obs)	Mean	Std Dev	CV
12/11/12	20.4	35	24 (n=5)	7.783	0.0021	0.026
12/15/12	20.7	35	10 (n=3)	7.762	0.0035	0.045
12/18/12	11.0	35	16 (n=4)	7.867	0.0017	0.022
1/4/13	29.8	22	8 (n=2)	7.973	0.0018	0.023
1/8/13	12.0	22	30 (n=5)	8.141	0.0055	0.068
1/15/13	20.5	22	29 (n=5)	7.860	0.0005	0.001
1/31/13	12.5	0.3	35 (n=5)	8.277	0.0054	0.065
2/6/13	20.4	0.3	65 (n=5)	7.881	0.0024	0.030
2/8/13	30.4	0.3	60 (n=5)	7.644	0.0013	0.018
2/12/13	25.5	34	39 (n=5)	7.934	0.0010	0.013
Overall Average					**0.0025**	**0.032**

The complete ten week time series for the SAMI-pH and corresponding reference pH results are shown in Fig. 3. Although temperature and salinity were tightly controlled and the test tank received a single batch of continuously circulated source water, the pH varied considerably during each of the trials due to biological activity. These time series results also include the rapid pH shifts that were conducted once a week under each of the test T/S conditions using acid/base additions. Detailed plots of each pH transitions are provided in Appendix I. In general, pH was varied by around 1.2 pH units during the acid/base additions and conditions maintained at each new value for approximately an hour.

The SAMI-pH operated continuously throughout the entire lab test and generated 3140 useable measurements out of a total possibility of 3154 timepoints. The total range of pH measured by the SAMI-pH was 6.273 to 8.441, compared to the range of our discrete reference samples of 6.943 to 8.502. The SAMI-pH showed some differences in accuracy across the various test solutions and pH ranges, as would be expected based on the specific set-up of the instrument in terms of the indicator dye used and its operational range. For the complete Lab test, the mean of the differences between the SAMI-pH measurement and reference pH was -0.017 ±0.970 (N=263), with a total range of -1.159 to 0.140. The larger offsets only occurred during the rapid pH shifts (see Appendix 1 and Stable Periods section below). Instrument measurements conducted at ten weeks with the second seawater trial exhibited a similar level of

accuracy with a mean difference from reference of -0.010 ± 0.033 (N=8) compared to 0.014 ± 0.003 (N=27) for measurements in the first week of the lab test (data not shown). The higher variance may have resulted from the final trial including continuously changing temperature conditions.
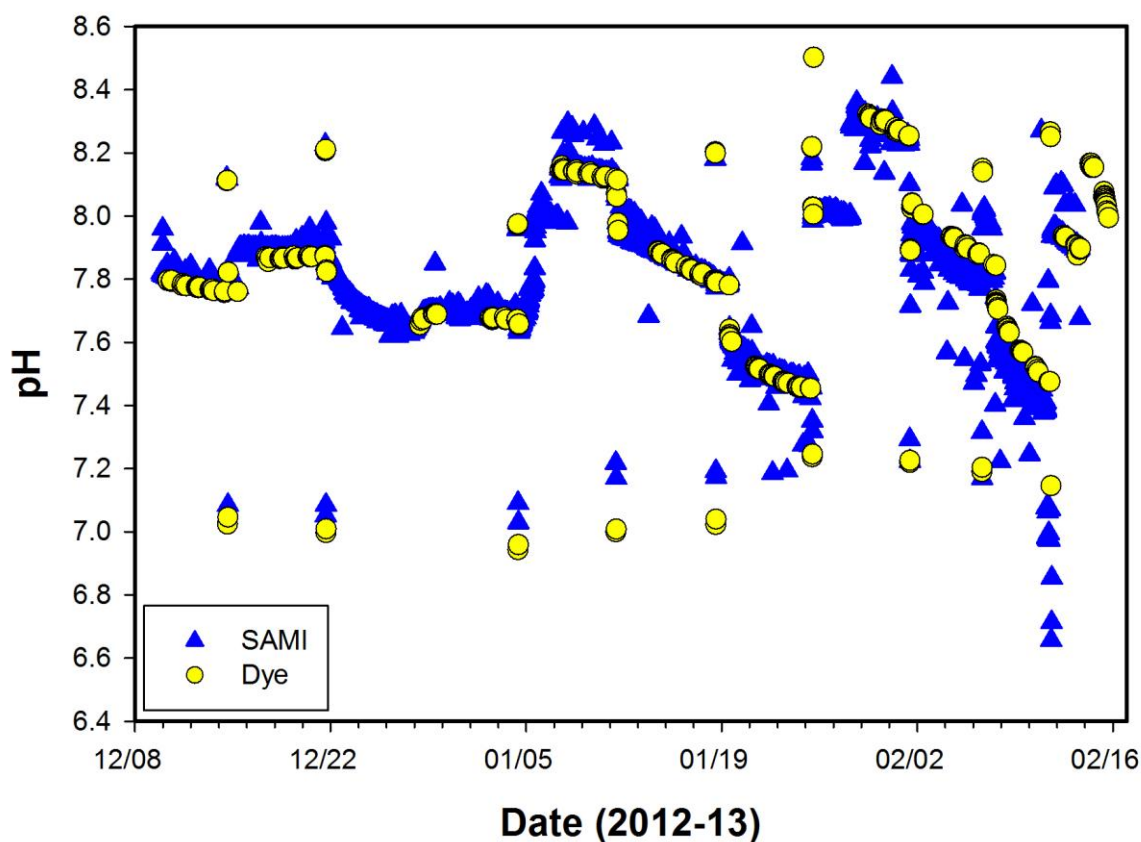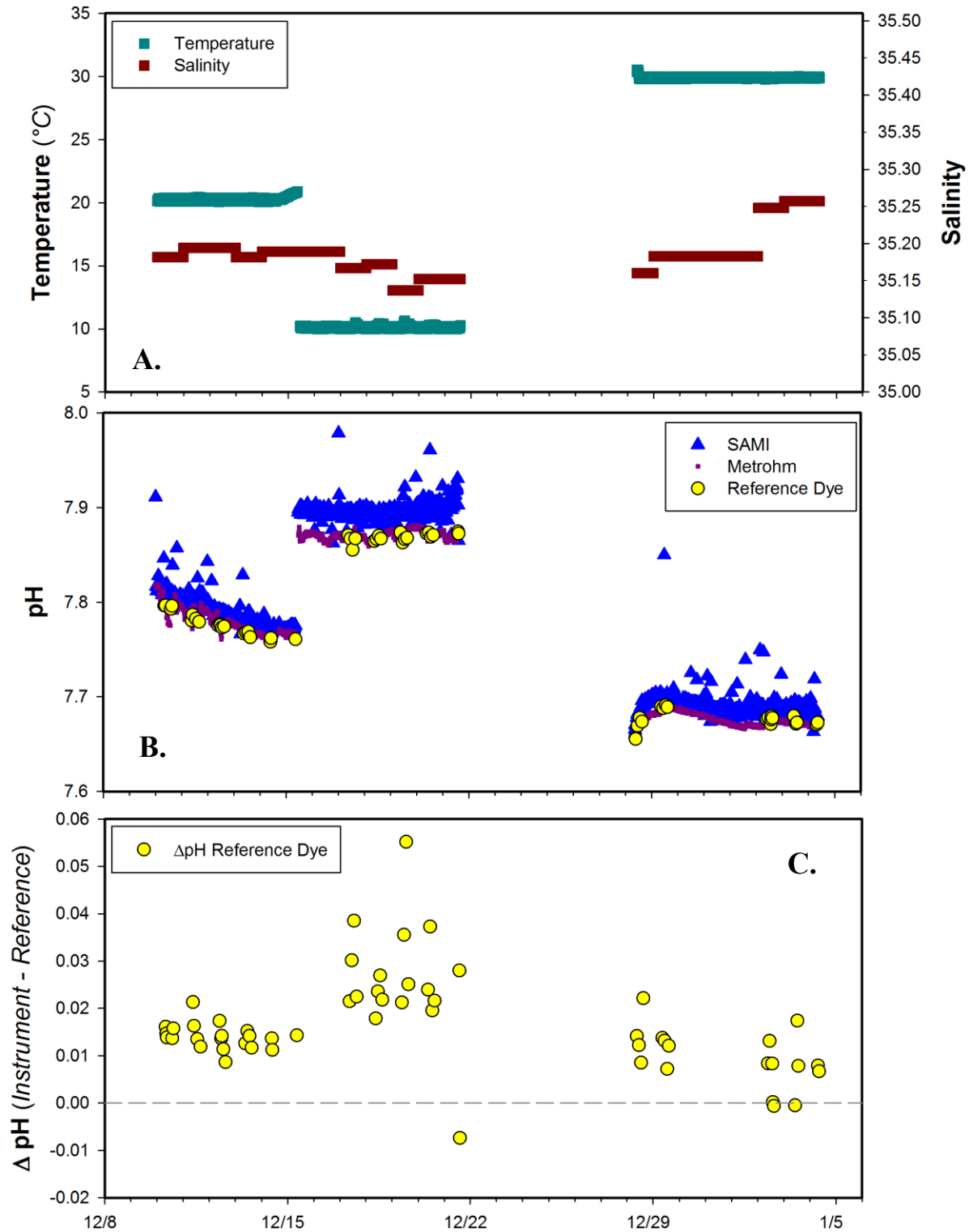

Figure 3. Time series of the Sunburst Sami pH test instrument pH results along with the dye reference pH results for the complete ten week lab test. Results include nine rapid pH transition tests conducted at each temperature and salinity condition. Detailed plots of each transition test are provided in Appendix I.

Lab testing results were further examined for the stable portions of the test for each of the source waters and these data do not include any of the pH shifts or temperature shifts. For the saltwater trials, salinity was maintained within 0.12 units (range = 35.14 to 35.26) with an average of 35.19 and temperature was maintained to within 0.8 ^{o}C for each set point (Fig. 4 top). The first temperature trial ranged from 20.07 to 20.88 ^{o}C, with an average of 20.27 ^{o}C., the second temperature trial ranged from 9.98 to 10.64 ^{o}C, with an average of 10.12 ^{o}C, and the third temperature trial ranged from 29.77 to 30.53 ^{o}C, with an average of 29.90 ^{o}C. Time series results of the SAMI-pH, Metrohm data and reference pH data are shown in middle panel.   The time

series of the calculated difference between Sami-pH and reference pH are shown in the bottom panel. The average difference between SAMI-pH and the reference pH was of 0.016 ± 0.010, with a total range of -0.007 to 0.055 (N=59).

During the brackish water trials, salinity was maintained to within 0.11 units (range = 21.75 to 21.86) with an average of 21.80 and temperature was maintained to within 0.7 ^{o}C for each set point (Fig. 5 top). The first temperature trial ranged from 9.95 to 10.67 ^{o}C, with an average of 10.09 ^{o}C, the second temperature trial ranged from 20.02 to 20.40 ^{o}C, with an average of 20.20 ^{o}C, and the third temperature trial ranged from 30.18 to 30.54 ^{o}C, with an average of 30.43 ^{o}C. Time series results of the SAMI-pH, Metrohm data and reference pH data are shown in middle panel.   The time series of the calculated difference between Sami-pH and reference pH are shown in the bottom panel. The average difference between SAMI-pH and the reference pH for the brackish water trial was 0.008 ±0.015, with a total range of -0.018 to 0.117 (N=76).

During the freshwater trials, salinity was maintained to within 0.11 units (range = 0.19 to 0.30) with an average of 0.26 and temperature to within 0.9 ^{o}C for each set point (Fig. 6 top). The first temperature trial ranged from 9.96 to 10.28 ^{o}C, with an average of 10.08 ^{o}C, the second temperature trial ranged from 20.00 to 20.91 ^{o}C, with an average of 20.22 ^{o}C, and the third temperature trial ranged from 29.73 to 30.55 ^{o}C, with an average of 30.42 ^{o}C. Time series results of the SAMI-pH, Metrohm data and reference pH data are shown in middle panel.   The time series of the calculated difference between Sami-pH and reference pH are shown in the bottom panel. The average difference between SAMI-pH and the reference pH for the freshwater trial was -0.057 ± 0.085, with a total range of -0.0480 to 0.063 (N=62).

.

Figure 4. Time series results of the Sunburst SAMI pH sensor for the stable portion of the seawater trials. Measurements taken during temperature and pH shifts have been removed. Panel (a) temperature and salinity conditions during the seawater trials; panel (b) measured pH from the Sunburst Sami pH plotted against both Metrohm and Dye reference pH measurements; and, panel (c) the difference between Sunburst and reference pH values.
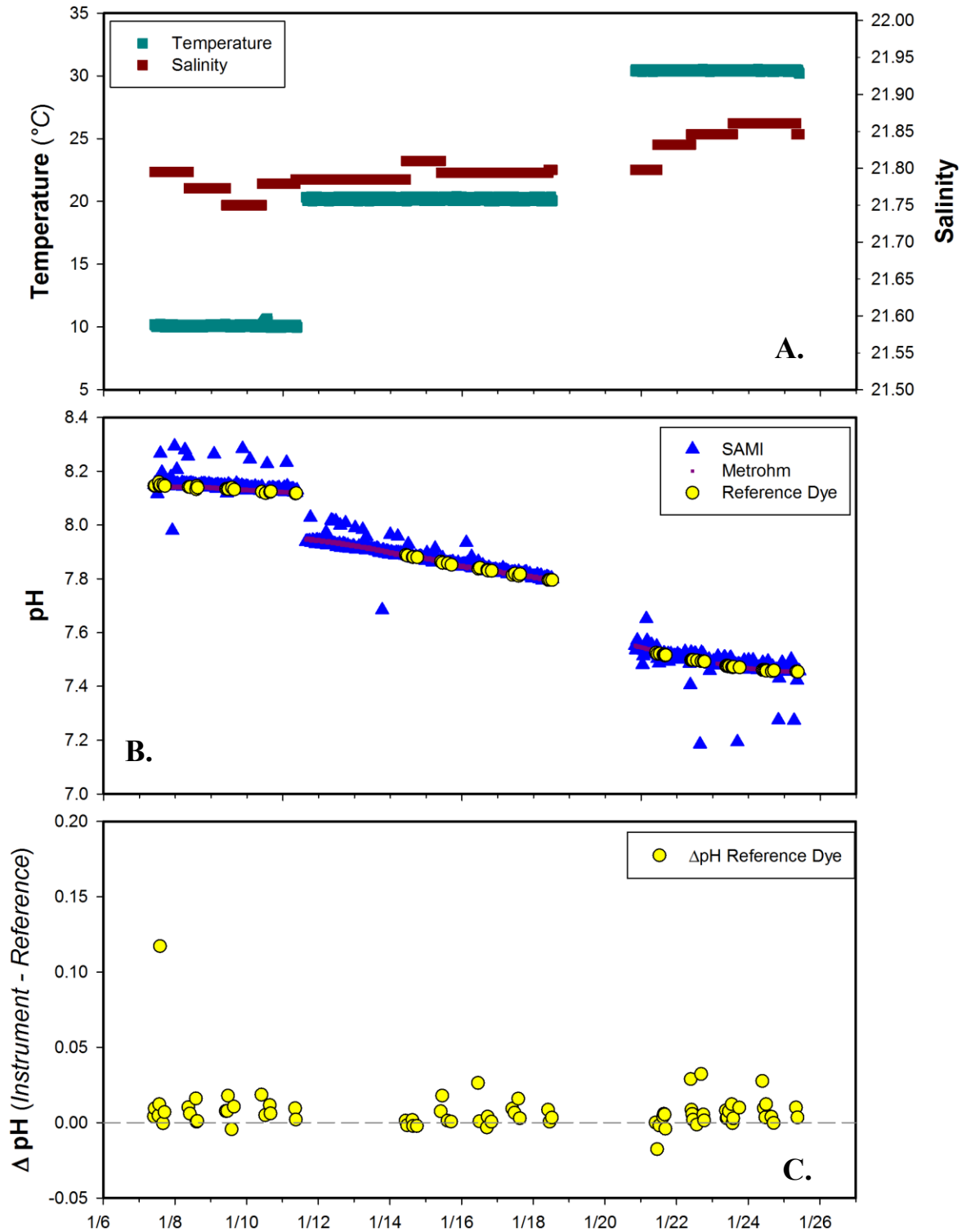
Figure 5. Time series results of the Sunburst Sami pH sensor for the stable portion of the brackish water trials. Measurements taken during temperature and pH shifts have been removed. Panel (a) temperature and salinity conditions during the seawater trials; panel (b) measured pH from the Sunburst Sami pH plotted against both Metrohm and Dye reference pH measurements; and, panel (c) the difference between Sunburst and reference pH values.
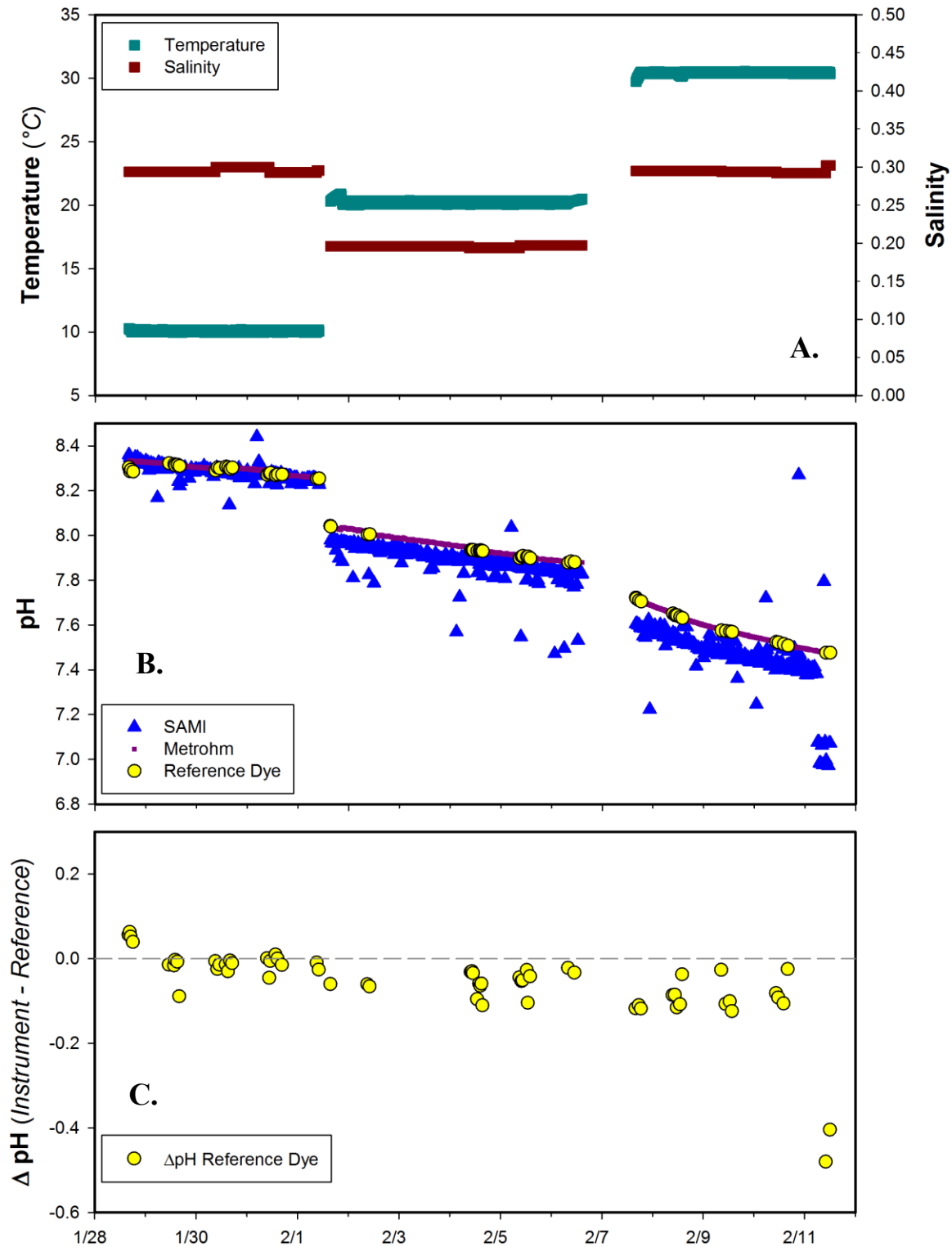
Figure 6. Time series results of the Sunburst Sami pH sensor for the stable portion of the freshwater trials. Measurements taken during temperature and pH shifts have been removed. Panel (a) temperature and salinity conditions during the seawater trials; panel (b) measured pH from the Sunburst Sami pH plotted against both Metrohm and Dye reference pH measurements; and, panel (c) the difference between Sunburst and reference pH values.

A cross plot of all corresponding SAMI-pH versus reference pH results in shown in figure 7. Results from the SAMI-pH agreed closely over the entire test range of 7.4 to 8.3 pH. The slopes and regression coefficients ranged from 1.082 (r^2=.99), 1.183 (r^2 = .99), and 1.183 (r^2=0.97) for seawater, brackish water, and freshwater tests respectively. The instrument response factor was more variable for the freshwater trials, as may be expected with the use of MetaCresol purple dye indicator for SAMI compared to Phenol Red for the reference samples. The overall linear regression of the data had a slope of 1.037 with an r^2 of 0.96. The mean and standard deviation of the calculated difference between the SAMI-pH measurement and the reference pH for all matched pairs within the three stable testing periods was -0.010 ±0.058 (N=193).
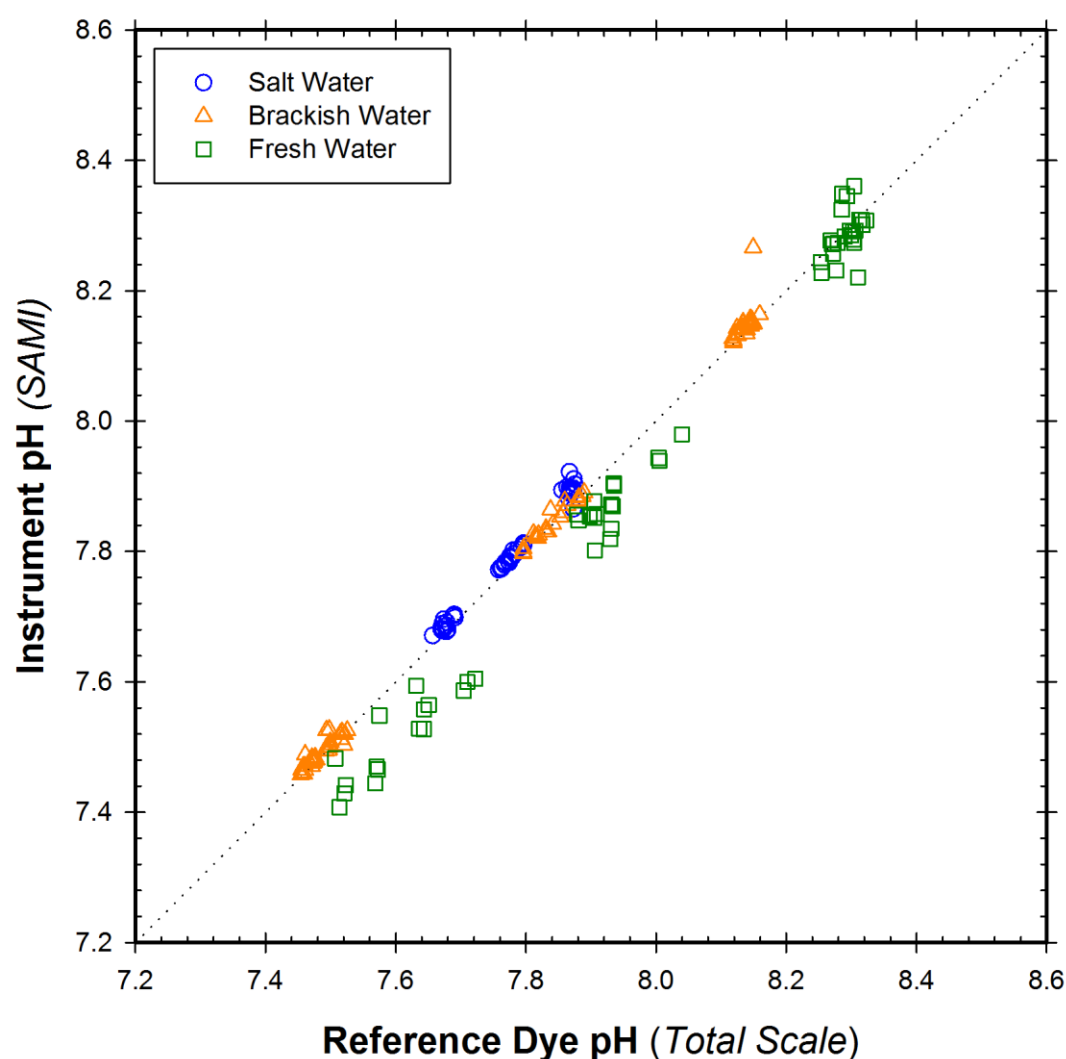
Figure 7. Cross plot of Sunburst Sami pH measurements against the dye reference pH measurements for 201 matched observations during the stable Temperature and Salinity portions of the entire lab test.

RESULTS of MOORED FIELD TESTS

Moored field tests were conducted to examine the performance of the SAMI-pH to consistently track natural changes in pH over extended deployment durations of 4-8 weeks. In addition, the field tests examined the reliability of the instrument, i.e., the ability to maintain integrity or stability of the instrument and data collections over time. Reliability was determined by quantifying the percent of expected data that was recovered and useable. The performance of the SAMI-pH was examined in four separate field deployment tests at various ACT Partner test sites to include a range of biogeochemical conditions. The range and mean for temperature and salinity for each test site is presented in Table 2. The final reference temperature data was computed from the mean of two RBR thermistors and the SeaBird SBE that were mounted at the same sampling depth as the test instrument. The reference temperature data and salinity were used to compute ambient pH results from the laboratory results using CO2-Sys program (Pierrot et.al. 2006). Immediately before each deployment the test instrument was exposed to a certified or internal reference sample for 3-4 measurements to confirm good working status and provide an initial offset against its latest calibration (Table 3).

Table 2. Range and average for temperature, and salinity at each of the test sites during the sensor field deployment. Temperature and salinity were measured by two RBR temperature loggers and a SeaBird SBE 26 (or SBE26plus) mounted on the instrument rack and the duration of the deployment.

SITE (deployment period/duration)		Temperature (^{0}C)	Salinity
Moss Landing Harbor	Min.	14.9	30.7
26Aug – 22Sept	Max.	18.6	33.7
(n = 28 days)	Mean	16.6	33.0
Kaneohe Bay	Min.	22.3	32.8
15Nov – 10Feb	Max.	26.5	35.2
(n = 88 days)	Mean	24.5	34.4
Chesapeake Bay	Min.	3.4	9.3
12Mar – 10Apr	Max.	10.1	14.0
(n = 30 days)	Mean	5.9	12.8
Lake Michigan	Min.	11.6	0.024
21Jun – 19Jul	Max.	24.9	0.035
(n = 29 days)	Mean	21.2	0.031

Table 3. Results of the pre-deployment check for the SAMI-pH at each deployment site. The pH of the reference solution is based on ACT's spectrophotometric analysis completed at the time of exposure. Seawater CRM provided by Andrew Dixon's Lab (Scripps Institution of Oceanography). Lakewater CRM based on freshly collected ambient site water and analyzed spectrophotometrically by ACT.

Deployment Site	Reference Solution	Measurement Temperature	Reference Solution pH	SAMI pH MCP	SAMI pH PR
MLML	Seawater CRM	20 ^{o}C	7.97	7.97	
HIMB	Seawater CRM	25 ^{o}C	7.86	7.60	
CBL	Seawater CRM	20 ^{o}C	8.00	7.99	
GLERL	Lakewater CRM	20 ^{o}C	8.23	8.13	8.26

## Moss Landing Marine Laboratories Field Deployment Site

A month-long moored field test was conducted at the Moss Landing Marine Laboratories from August 26 to September 22, 2013. The deployment site was located at 36.80°N, 121.79°W, in the Small Boat Facilities area of Moss Landing Harbor (Photo 2). This secure deployment site was located in the Harbor on the junction of the northern tributary of the Salinas River and Elkhorn Slough National Estuarine Reserve on the central coast of California. Instrumentation was deployed off a floating dock in waters with a tidal range of 2 meters and a maximum depth below the dock of 4 meters.
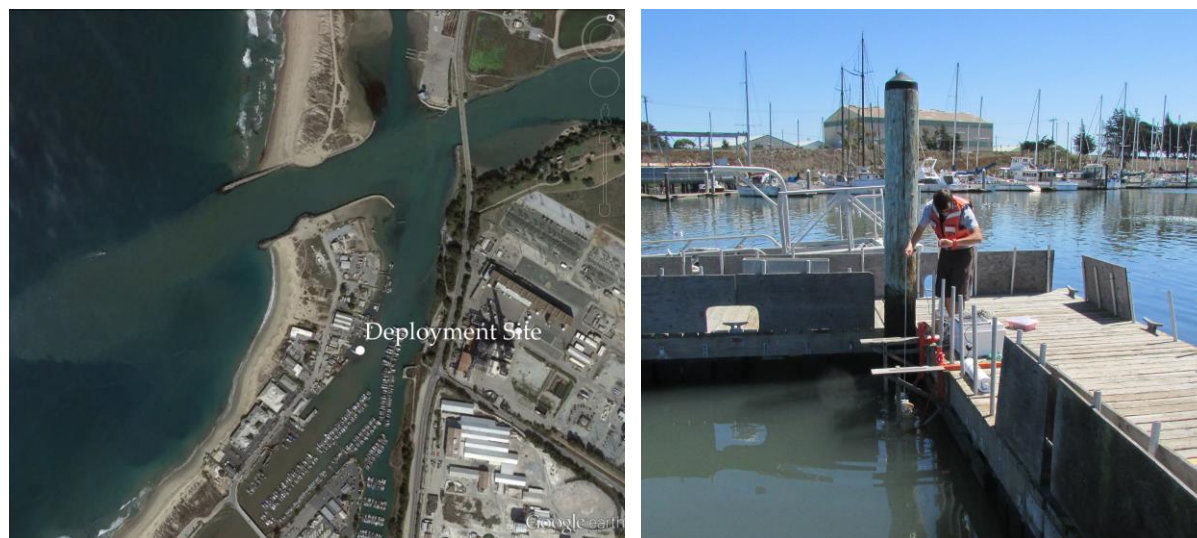
Photo 2. Aerial view of MLML Harbor (left) and dockside mooring deployment (right).

Time series results of ambient conditions for tidal height, temperature, and salinity are given in figure 8. Temperature ranged from 14.9 to 18.6 and salinity from 30.7 to 33.7 over the duration of the field test. The bottom panel displays the maximum difference recorded between all reference thermistors (RBR Solo and SBE26) mounted at the same depth and different locations across the mooring rack. The average temperature difference observed was 0.11 ^{o}C with a maximum of 0.8 ^{o}C. The difference between the test instruments instantaneous temperature reading and the final reference temperature will contribute to the variability and offset to the calculated ambient reference pH. It is not possible to quantify the exact difference between the averaged reference temperature and that measured by the test instrument, but a difference of 1 ^{o}C would amount to an approximate offset of 0.015 pH units at ambient conditions. Additional differences resulting from variation in the actual chemistry of the water across this space may exist but will be minimized because the sampling bottle integrates across the mooring space.

The SAMI-pH operated successfully throughout the entire deployment and generated 2564 observations based on its 15 minute sampling interval. Time series results of the SAMI-pH and corresponding reference pH results are given in figure 9. The range of pH measured by Sami-pH was 7.798 to 8.150 compared to the range captured by reference samples of 7.933 to 8.077. The bottom panel presents the time series of the difference between the two results for each matched pair (N=84 observations). The average and standard deviation of the measurement difference over the total deployment was 0.039 ±0.003 with a total range of -0.123 to 0.158.
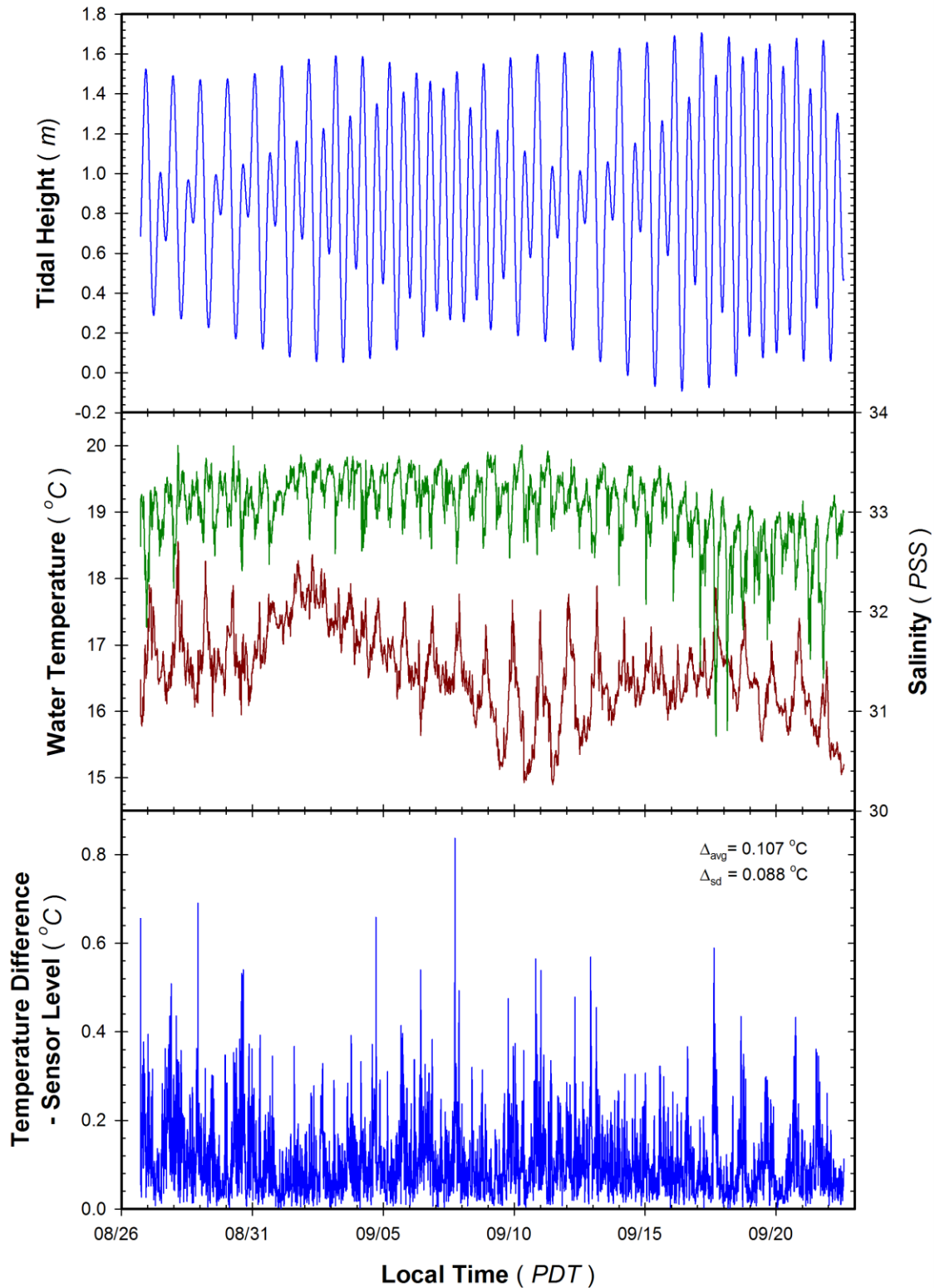
Figure 8. Environmental conditions encountered during deployment at the MLML Small Boats Dock. Test sensor array deployed at 1 m fixed depth, variation in local tidal heights indicate active water flow around instrument (*Top Panel*). Variation in salinity (green) and temperature (red) at depth of instrument sensor detected by an SBE 26 and two RBR Solo thermistors (*Middle Panel*). Temperature range determined from max-min temperatures detected by RBR and SBE thermistors spanning instrument sensor array (*Bottom Panel*).
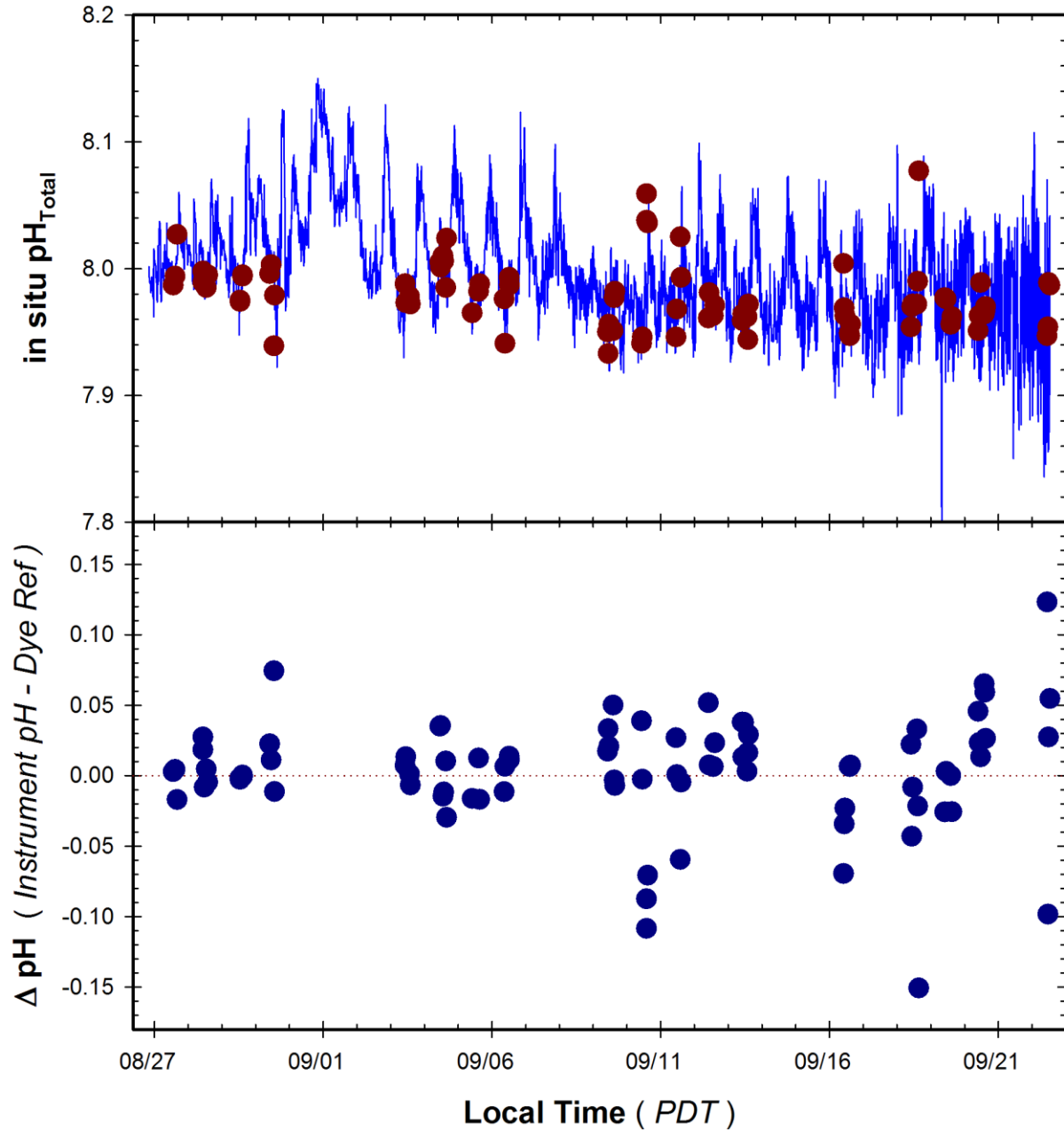
Figure 9. Time series of pH measured detected by a SAMI-pH(mCP) instrument deployed during the 1 month MLML field trial. *Top Panel*: Continuous pH_{Total} recordings from instrument (blue line) and pH_{Total} of adjacent grab samples determined by mCP dye (red circles). *Bottom Panel*: Difference in measured pH relative to reference samples (Instrument – Dye Ref) observed during deployment. Zero offset is represented by dashed line as pH_{Total} scale is reported by both the SAMI-pH sensor and the dye reference method.

A cross-plot of the matched observations is given in figure 10. The regression was not statistically significant (p=0.55, r^2 =0.004) and there was significant scatter around a fairly small pH range within the observations. There were slightly more SAMI-pH observation below the reference value (51) than above (33).
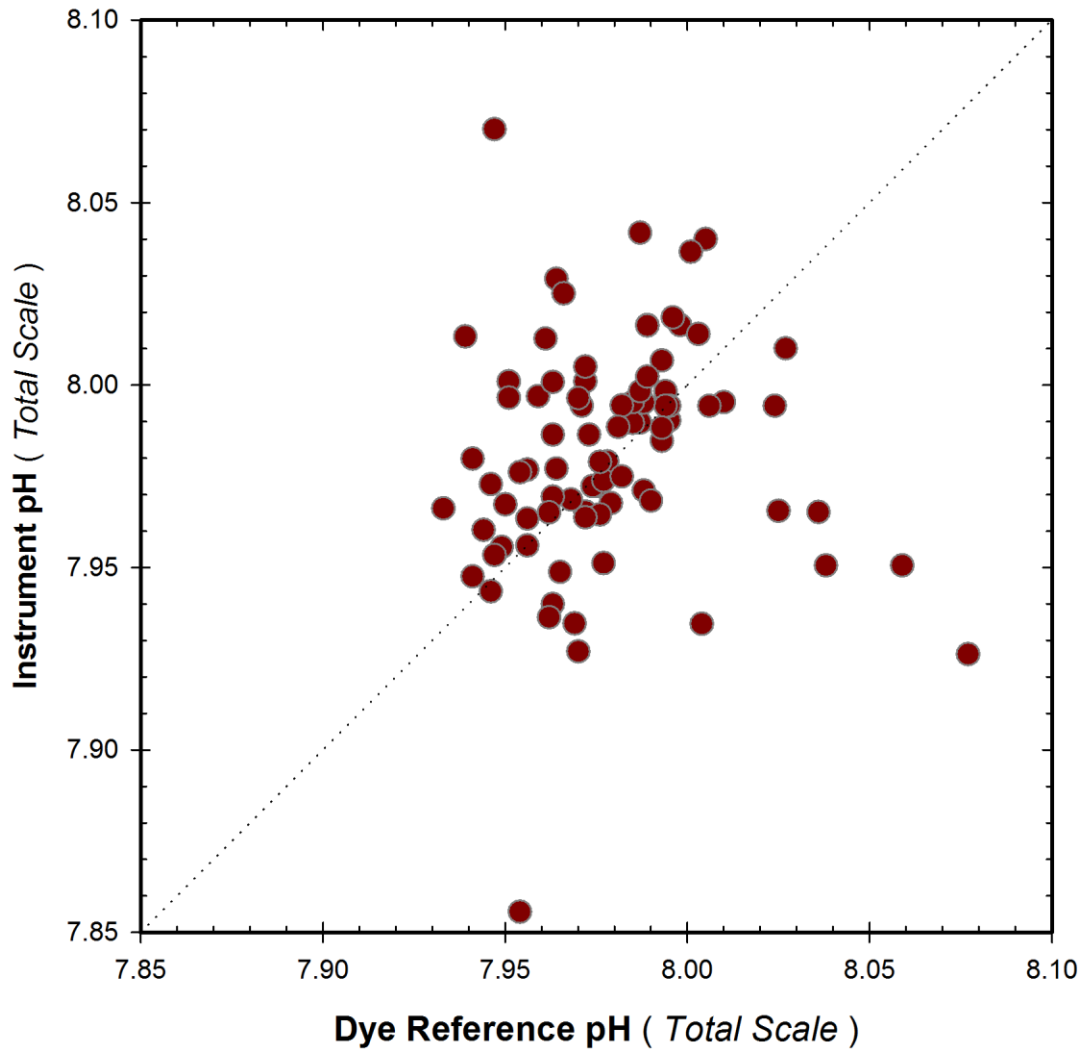
Figure 10. MLML field response plot of a SAMI-pH(mCP) instrument compared to reference pH$_{Total}$ samples. Fine dotted line represents 1:1 correspondence, as pH$_{Total}$ scale is reported by the SAMI-pH instrument and the dye reference method.
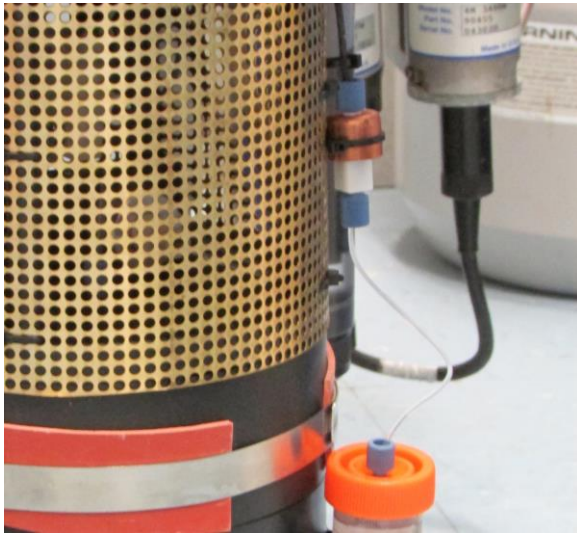
Photo 3. The SAMI pH prior to and following the 1 month deployment for the MLML field test.

Moored Deployment off Coconut Island in Kaneohe Bay, Hawaii

A three month-long moored field test was conducted in Kaneohe Bay from November 15, 2013 to February 10, 2014. The deployment site was located at 21.46° N, 157.80° W in the back-reef region of the barrier reef offshore of Coconut Island (HIMB) in a depth of 3 meters (Photo 3). Kaneohe Bay, located on the eastern side of Oahu, Hawaii, is a complex estuarine system with a large barrier coral reef, numerous patch reefs, fringing reefs, and several riverine inputs. Tides in Kaneohe Bay are semi-diurnal with mean tidal amplitude of approximately 68 cm day. The pH instruments were mounted on the Crimp II Buoy (right), about on meter below the surface.

Photo 4. HIMB deployment site (left) and sampling at deployment buoy (right).

Time series results of ambient conditions for tidal height, temperature, and salinity are given in figure 11.   Temperature ranged from 22.3 to 26.5 and salinity from 32.8 to 35.2 over the duration of the field test. The bottom panel displays the maximum difference recorded between all reference thermistors mounted at the same depth but located across the mooring rack. The average temperature difference observed across the space of the mooring rack was 0.04 ±0.10 $^{\circ}$C, with a maximum of 0.65 $^{\circ}$C.   As noted above, it is not possible to quantify the exact difference between the averaged reference temperature and that measured by the test instrument, but differences resulting from this variation and variation in the chemistry of the water across will be minimized because the sampling bottle integrates across the mooring space.

The SAMI-pH operated successfully throughout the entire deployment and generated 4211 observations based on its 30 minute sampling interval, with 17 timepoints used for internal calibration cycles. Time series results of the SAMI-pH and corresponding reference pH results are given in Figure 12. Ambient pH measured by the SAMI ranged from 7.678 to 8.220 compared to the range captured by reference samples of 7.814 to 8.084. The bottom panel presents the time series of the difference between the two results for each matched pair (N=101 observations). The average and standard deviation of the difference for the total deployment was -0.014 ±0.016 with a total range in the differences of -0.068 to 0.025.
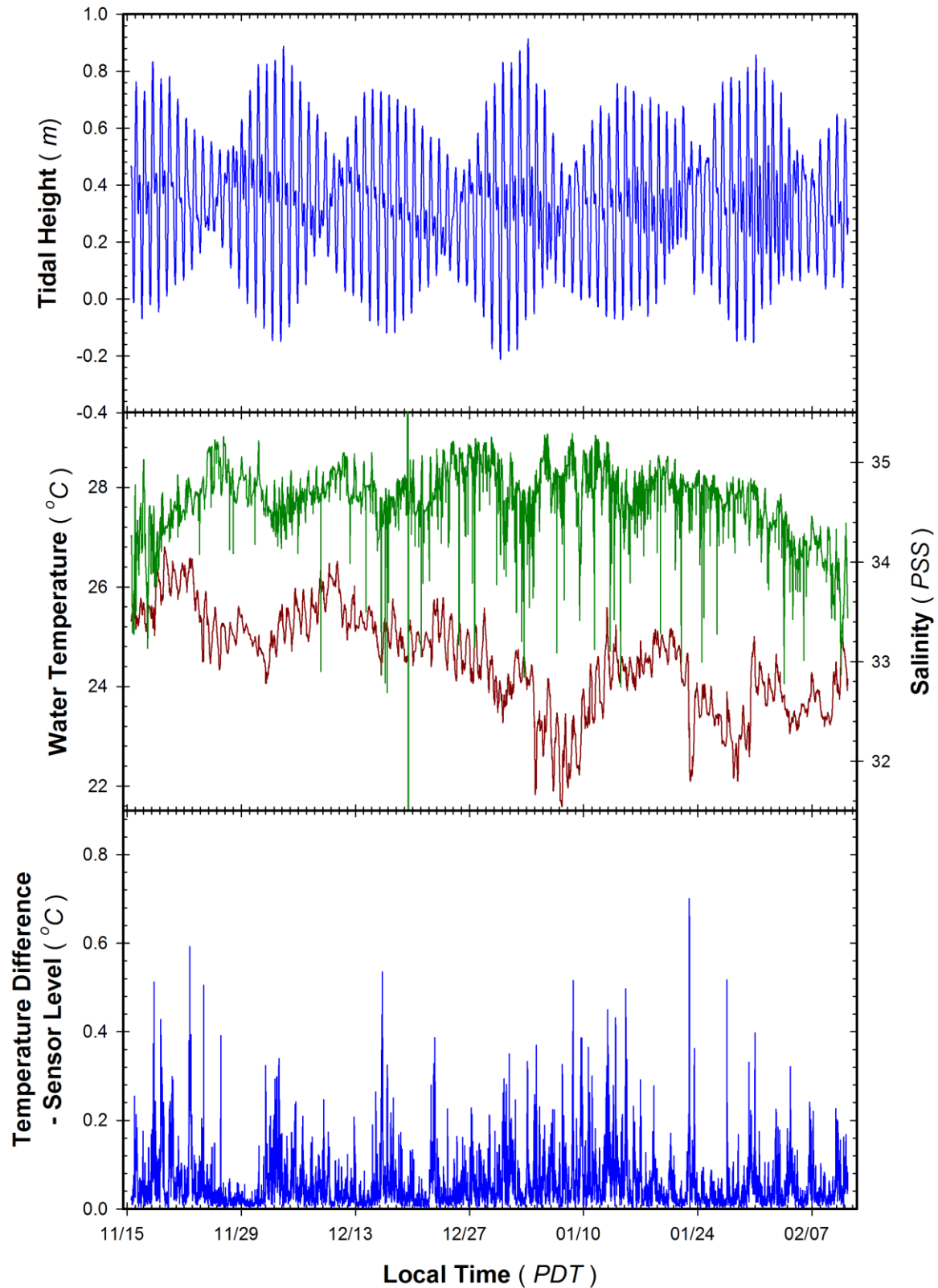
Figure 11. Environmental conditions encountered during the 3 month HIMB offshore deployment on the NOAA CRIMP Buoy. Test sensor array deployed at 1 m fixed depth, variation in local tidal heights indicate active water flow around instrument (*Top Panel*). Variation in salinity (green) and temperature (red) at depth of instrument sensor detected by an SBE 26 and two RBR Solo thermistors (*Middle Panel*). Temperature range determined from max-min temperatures detected by RBR and SBE thermistors spanning instrument sensor array (*Bottom Panel*).
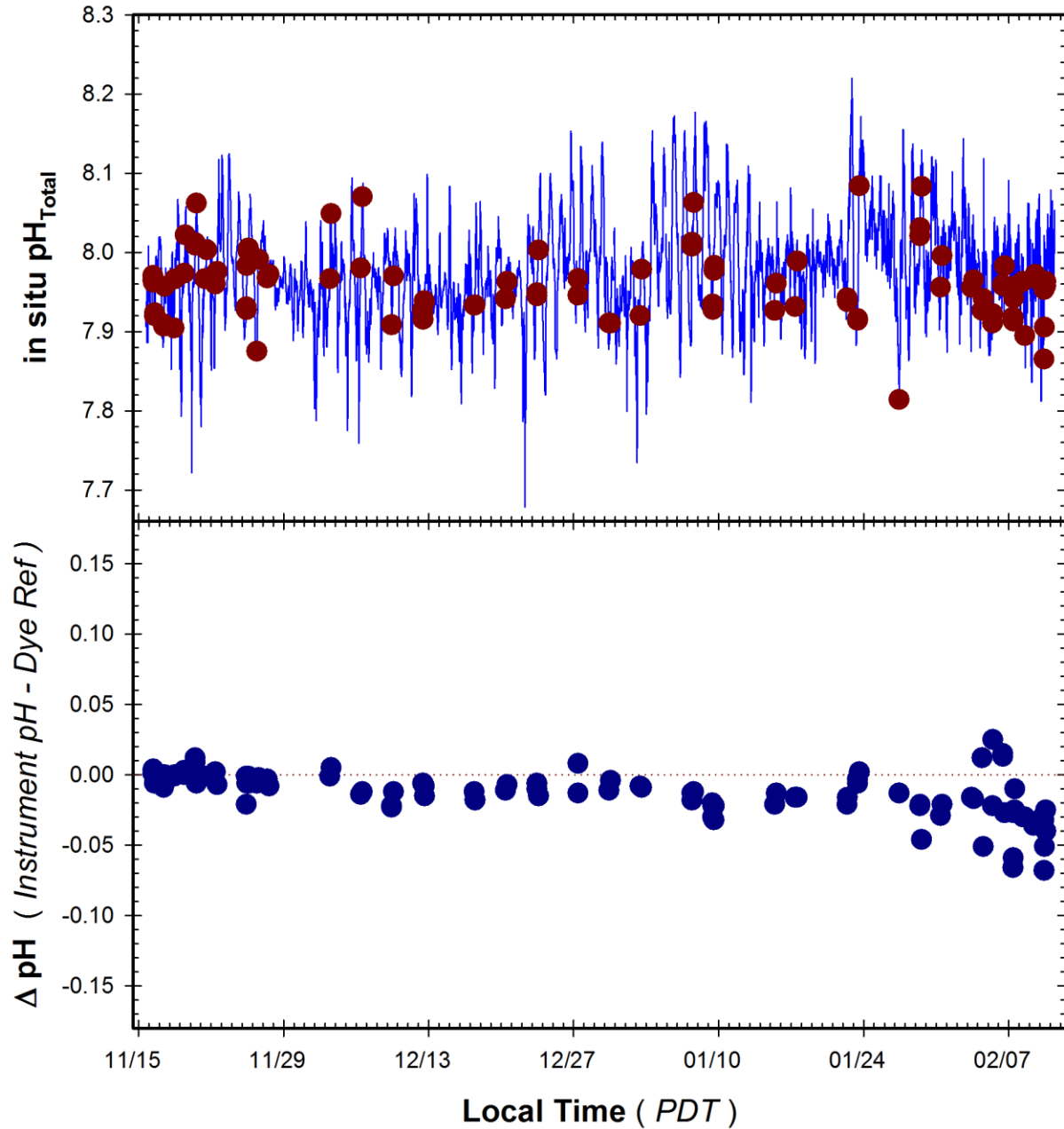
Figure 12. Time series of pH measured by a SAMI-pH(mCP) instrument deployed during the 3 month HIMB field trial. *Top Panel*: Continuous pH_{Total} recordings from instrument (blue line) and pH_{Total} of adjacent grab samples determined by mCP dye (red circles). *Bottom Panel*: Difference in measured pH relative to reference samples (Instrument – Dye Ref) observed during deployment. Zero offset is represented by dashed line as pH_{Total} scale is reported by the SAMI-pH sensor and the dye reference method.

A cross-plot of the matched observations is given in figure 13. A linear regression was statistically significant (p<0.001, r^2 =0.87) with a slope of 0.954. The SAI-pH under-predicted pH for 84 of the observations and over-predicted for 17 of the observations.
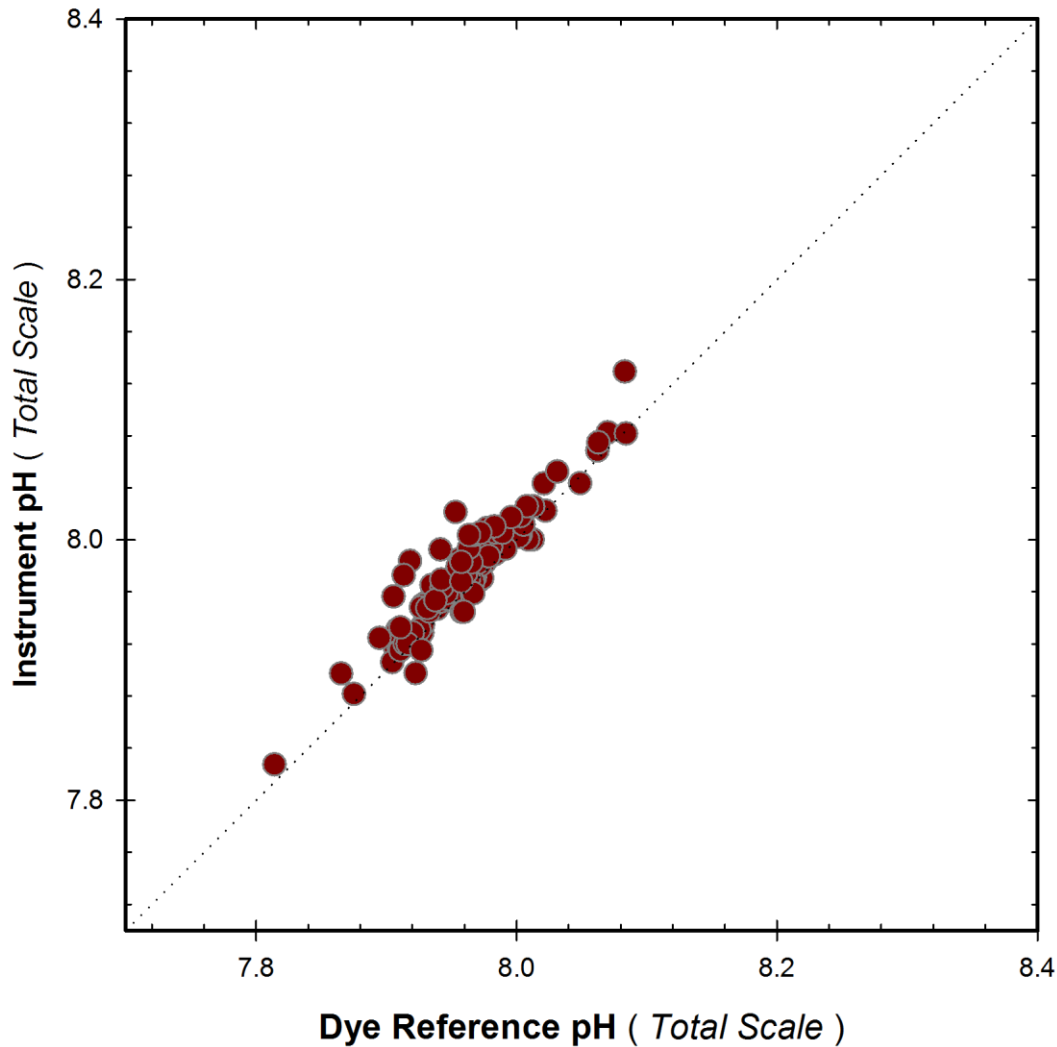
Figure 13. HIMB field response plot of a SAMI-pH(mCP) instrument compared to reference pH_{Total} samples. Fine dotted line represents 1:1 correspondence, as pH_{Total} scale is reported by the SAMI-pH instrument and the dye reference method.
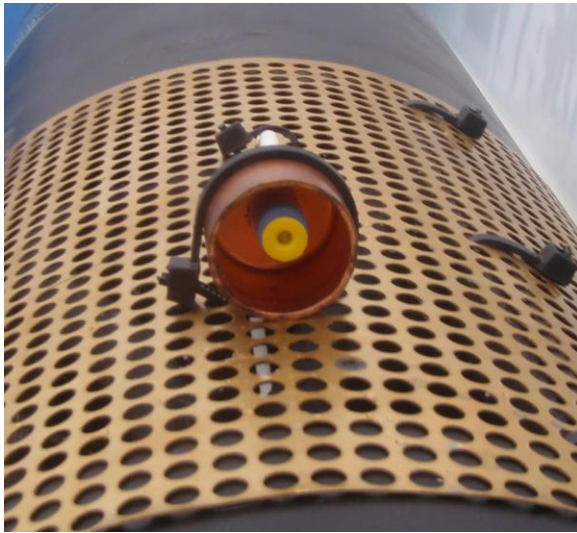
Photo 5. The SAMI pH prior to and following the 3 month deployment for the HIMB field test.

Moored Deployment at Chesapeake Biological Laboratory (CBL)

A month-long moored field test was conducted in Chesapeake Bay from March 3 to April 10, 2014. The deployment was located at 38.32°N, 76.45°W attached to the end of a pier at the mouth of the Patuxent River (Photo 6.) The average water depth of the test site was 2.2 m. The site was brackish with salinity ranging from 9.1 PSU to 14.3 PSU during the deployment and temperature ranging from 3.3 °C to 10.2 °C.
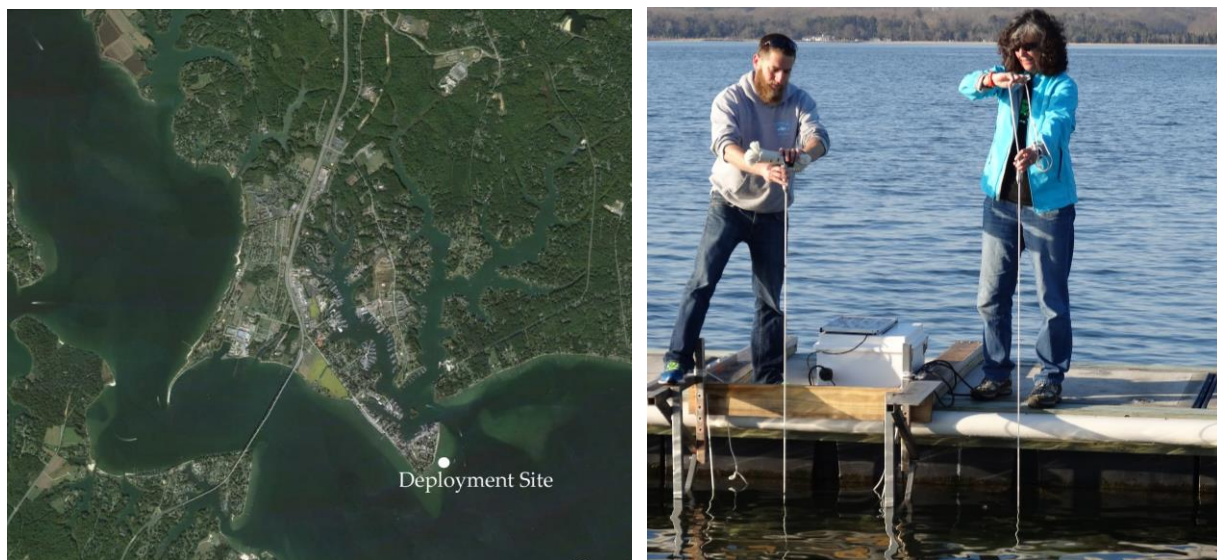
Photo 6. Aerial view of CBL deployment site (left) and duplicate sampling at mooring rack.

Time series results of ambient conditions for tidal height, temperature, and salinity are given in figure 14. Temperature ranged from 3.4 to 10.1 and salinity from 9.3 to 14.0 over the duration of the field test. The bottom panel displays the maximum difference recorded between all reference thermistors (RBR Solo and SBE26) mounted at the same depth and different locations across the mooring rack. The average temperature difference observed across the space of the mooring rack was 0.04 ±0.06 °C, with a maximum of 1.38 °C. As noted above, it is not possible to quantify the exact difference between the averaged reference temperature and that measured by the test instrument, but differences resulting from this variation and variation in the chemistry of the water across will be minimized because the sampling bottle integrates across the mooring space.

The SAMI-pH operated successfully from March 12 through April 1 and generated 1842 observations based on its 15 minute sampling interval. Time series results of the SAMI-pH and corresponding reference pH results are given in figure 15. From April 1-10, there were 175 no data returns and the remaining 714 reported observations showed increasing deviations from the reference results. Ambient pH measured by the SAMI from March 12 through April 1 ranged from 8.121 to 8.479, compared to the range captured by the reference measurements of 8.024 to 8.403. The bottom panel presents the time series of the difference between the two results for each matched pair (N=89 observations). The average and standard deviation of the measurement difference for the period up to April 1 was 0.041 ±0.033 (n=69), with the total range of differences between -0.086 to 0.114.
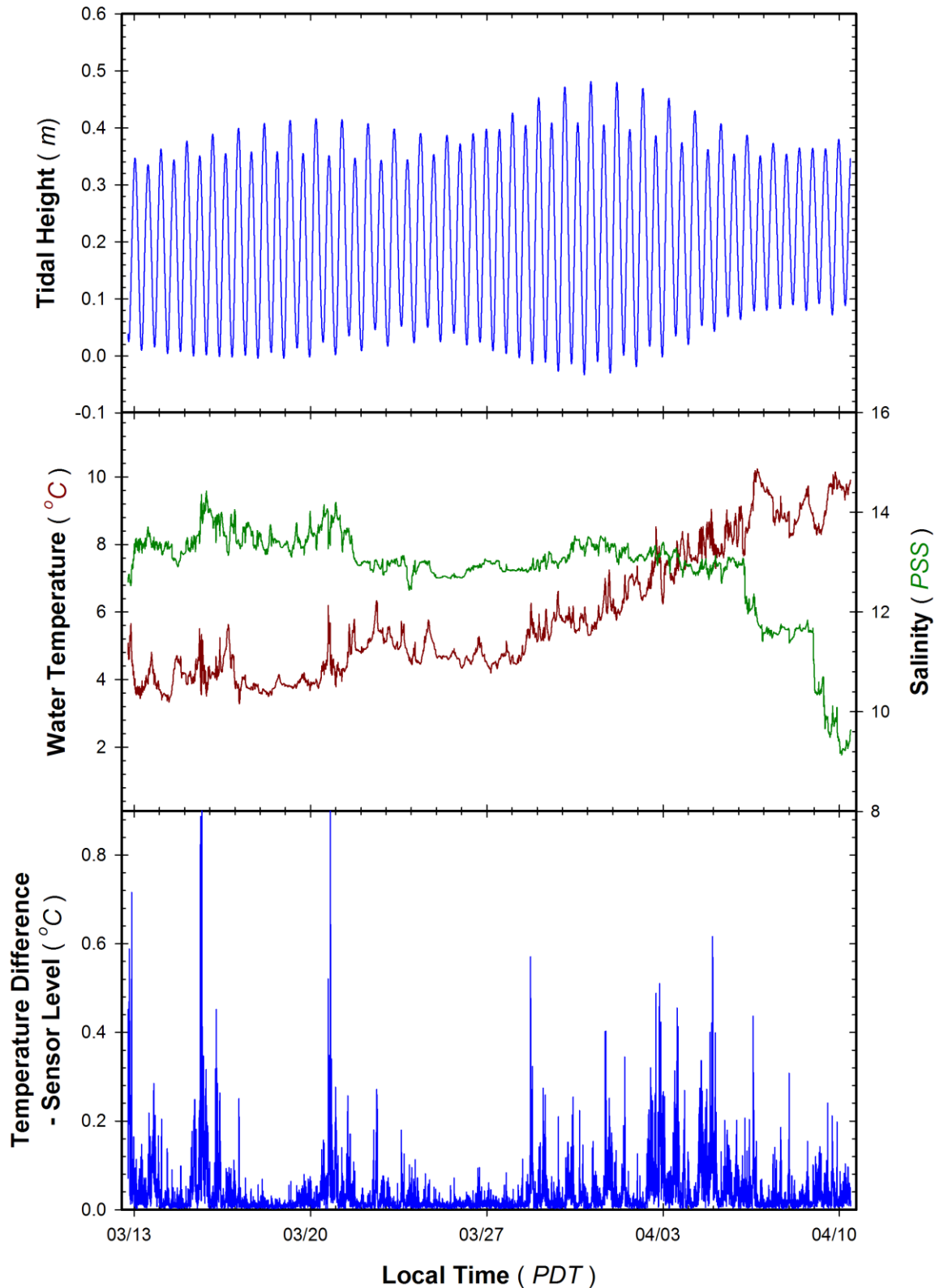
Figure 14. Environmental conditions encountered during the 1 month CBL floating dock deployment. Test sensor array deployed at 1 m fixed depth, variation in local tidal heights indicate active water flow around instrument (*Top Panel*). Variation in salinity (green) and temperature (red) at depth of instrument sensor detected by an SBE 26 and two RBR Solo thermistors (*Middle Panel*). Temperature range determined from max-min temperatures detected by RBR and SBE thermistors spanning instrument sensor array (*Bottom Panel*).
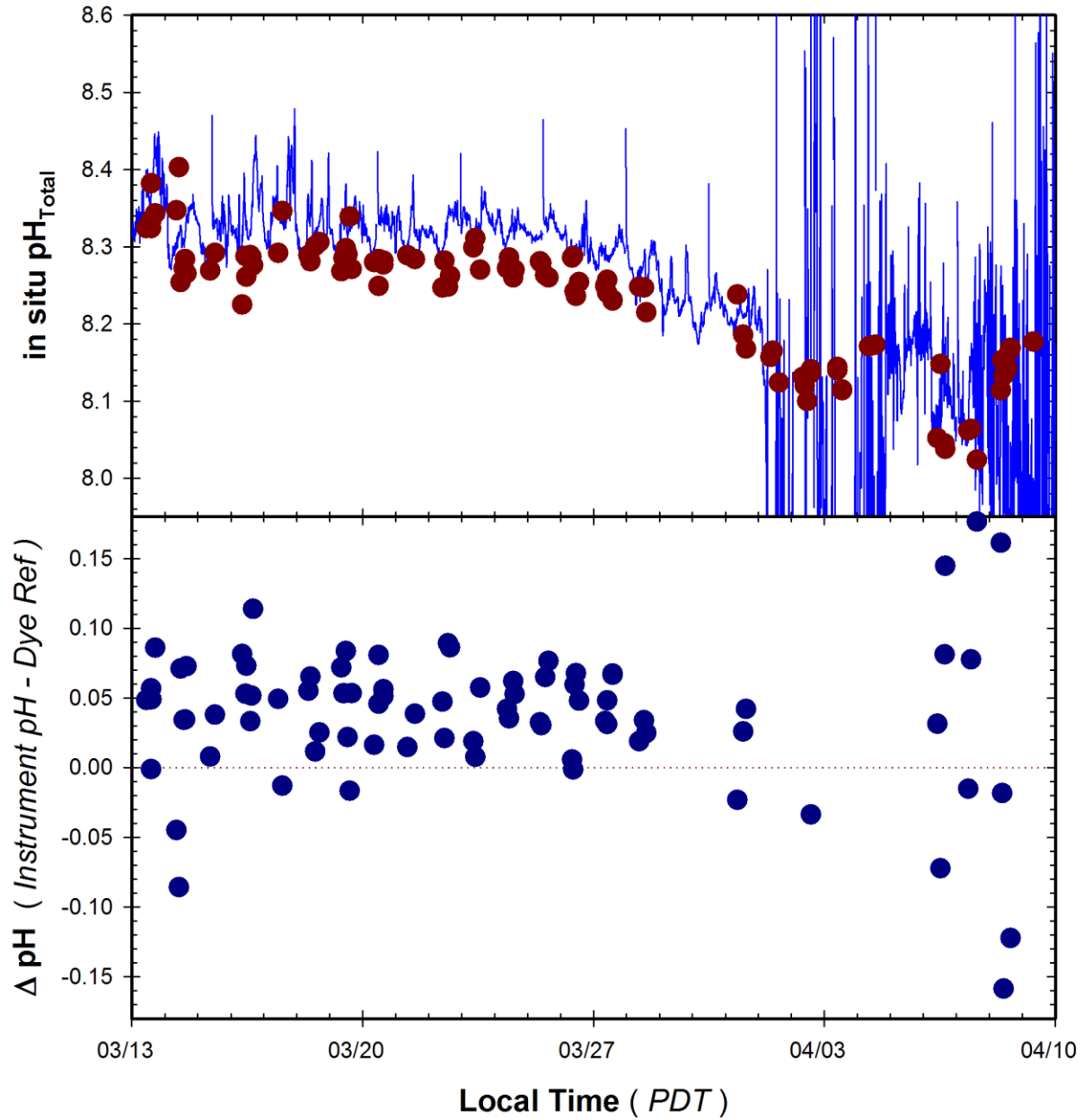
Figure 15. Time series of pH measured by a SAMI-pH(mCP) instrument during the 1 month CBL field trial. *Top Panel*: Continuous pH_{Total} recordings from instrument (blue line) and pH_{Total} of adjacent grab samples determined by mCP dye (red circles). *Bottom Panel*: Difference in measured pH relative to reference samples (Instrument – Dye Ref) observed during deployment. Zero offset is represented by dashed line as pH_{Total} scale is reported by the SAMI-pH sensor and the dye reference method.

A cross-plot of the matched observations is given in figure 16. The linear regression of the data from the initial, well performing period, was statistically significant ($p<0.001$, $r^2 = 0.40$) with a slope of 0.647 and intercept of 2.895.
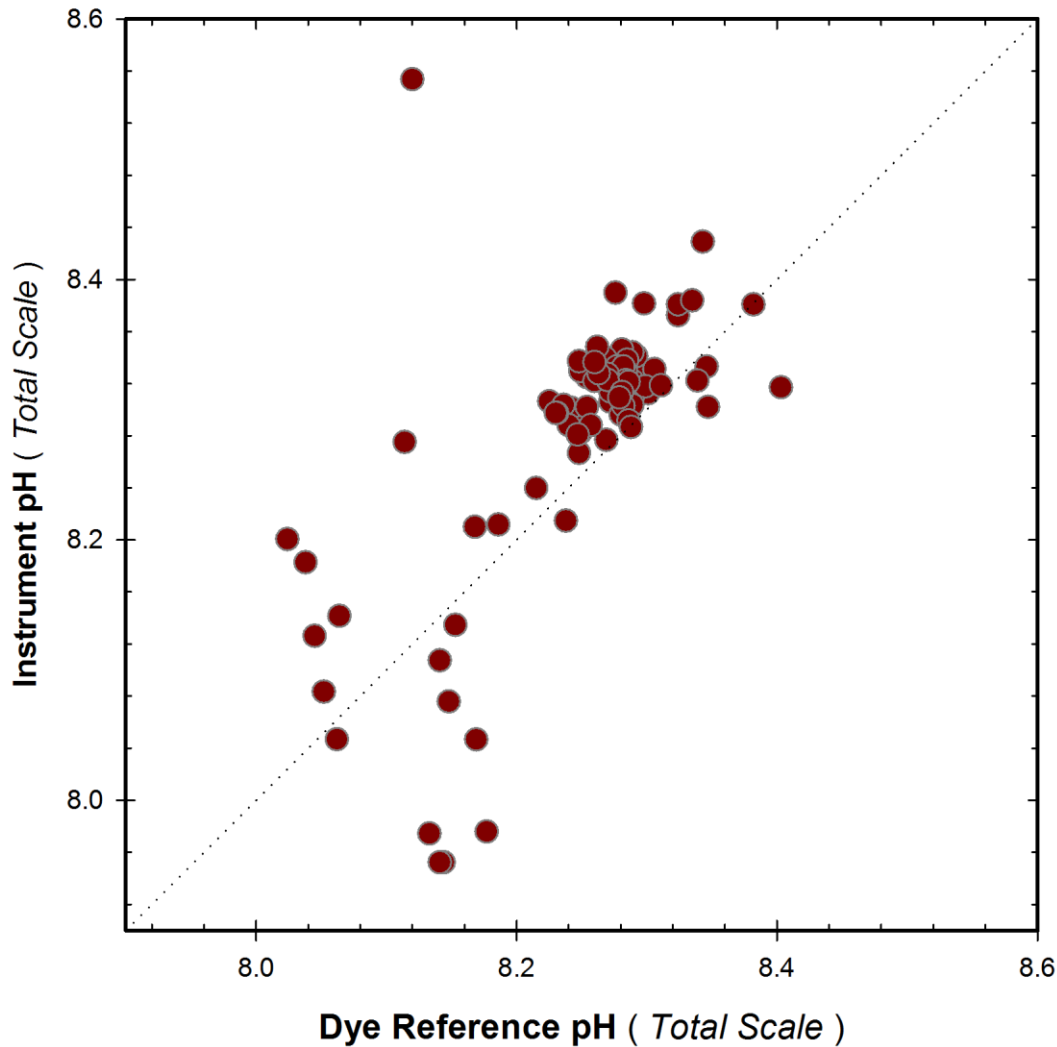
Figure 16. CBL field response plot a SAMI-pH(mCP) instrument compared to reference pH_{Total} samples. Fine dotted line represents 1:1 correspondence, as pH_{Total} scale is reported by both the SAMI-pH instrument and the dye reference method.
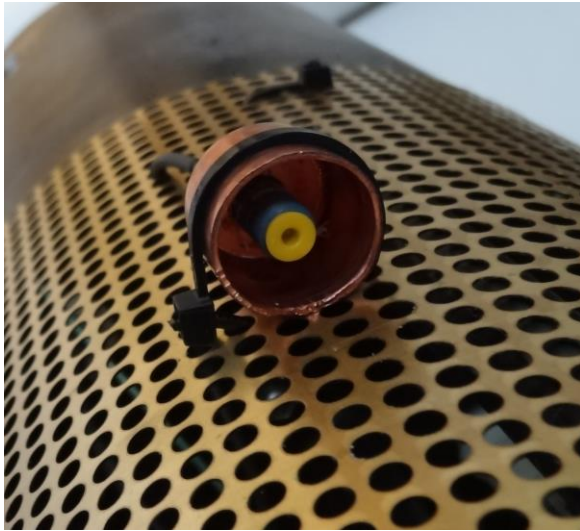
Photo 7. The SAMI pH prior to and following the 1 month deployment for the CBL field test.

Great Lakes Field Test Site

A month-long moored field test was conducted in Lake Michigan from June 21 to July 19, 2014. The Great Lakes deployment site was located at 43.23°N, 86.34°W on a fixed pier at the Lake Michigan Field Station of the NOAA Great Lakes Environmental Research Laboratory, in Muskegon, Michigan. The site is located at the outfall of Lake Muskegon into Lake Michigan with a depth of approximately 3 meters. The temperature range during deployment was 11.6°C through 24.9°C and salinity averaged 0.031.
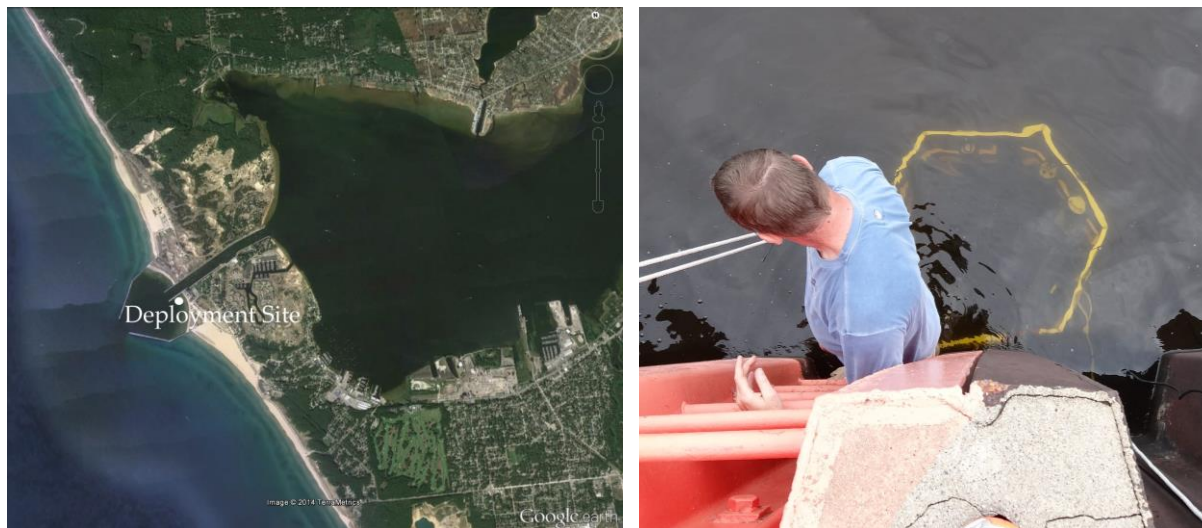
Photo 8. Aerial view of Lake Michigan and Muskegon Lake (left) and rack deployment at GL Site

Two SAMI-pH units were deployed at this test site, one equipped with MetaCresol Purple (mCP) as the indicator dye and one with Phenol Red (PR) as the indicator dye. The Phenol Red dye was considered to be experimental for the company at the time and being tested for appropriateness in freshwater. The ACT reference pH samples were determined using a Phenol Red dye prepared in Dr. Bob Byrne' lab at the University of S. Florida. Time series results of ambient conditions for tidal height, temperature, and salinity are given in figure 17. Temperature ranged from 11.6 to 24.9 and salinity from 0.024 to .035 over the duration of the field test. While there are no tidal cycles at this site there was significant exchange between the open lake and the protected basin where the mooring was located as evidenced by significant changes in salinity and water depth. The bottom panel displays the maximum difference recorded between all reference thermistors (RBR Solo and SBE26) mounted at the same depth and different locations across the mooring rack. The average temperature difference observed across the space of the mooring rack was 0.08 ±0.18 °C, with a maximum of 2.53 °C. As noted above, it is not possible to quantify the exact difference between the averaged reference temperature and that measured by the test instrument, but differences resulting from this variation and variation in the chemistry of the water across will be minimized because the sampling bottle integrates across the mooring space.

The SAMI-pH (PR) unit operated throughout the entire deployment and generated 2654 observations based on its 15 minute sampling interval, plus 11 timepoints used for internal calibration (Fig. 18). However, 51 of the observations had significantly higher offsets from the SAMI-pH (mCP) unit, as well as the dye reference, and those data are of uncertain quality and

were omitted from statistical summaries.   The original range in ambient pH measured by the SAMI-pH (PR) was 6.606 to 9.305 compared to 8.013 to 8.526 for the reference pH measurements. If we exclude the 51 observations that are suspect, the range was reduced to 7.658 – 8.687, which agrees much closer to that of the reference data. The bottom panel presents the complete time series of the difference between the SAMI-pH (PR) and reference sample for each matched pair (n=98 observations).   If we exclude seven outliers in the matched dataset during the malfunction the average difference of the SAMI-pH (PR) and reference was 0.021 ± 0.047 (N=92), with a total range of -0.095 to 0.177. A cross-plot of the matched observations is given in figure 19. If we exclude the suspect data, the linear regression was highly significant with a slope of 0.91, an intercept of 0.76 and $r^2 = 0.80$.

The SAMI-pH (mCP) unit operated continuously during the entire deployment and generated 2665 observations based on its 15 minute sampling interval. No internal calibration cycles were included in the program on this unit. Time series results of the SAMI-pH (mCP) and corresponding reference pH results are given in figure 20. The range in ambient pH measured by the SAMI-pH (mCP) was 7.813 to 8.405 compared to 8.013 to 8.526 for the reference pH measurements The bottom panel presents the time series of the difference between the two results for each matched pair (n=98 observations). The average and standard deviation of the difference for the entire deployment was -0.162 ± 0.045 with a total range of -0.333 to -0.068. A cross-plot of the matched observations is given in figure 21. The linear regression was highly significant (<0.001, $r^2 = 0.81$) with a slope of 0.816 and an intercept of 1.352.

A summary plot of the SAMI-pH versus dye reference pH for all sites is shown in figure 22. The SAMI-pH responded consistently among all test sites which covered a salinity range of freshwater to full seawater and temperatures from 3-26 ^{o}C. As expected based on its characterized operational properties, there was a consistent offset in results reported within the Great Lakes test for the SAMI-pH with MetaCresol Purple indicator dye compared with the unit operating with the Phenol Red dye indicator. After removing the 51 outliers, the average difference between the two SAMI-pH units used for the freshwater test was 0.177 ± 0.079, with a total range of -0.375 to 0.393.
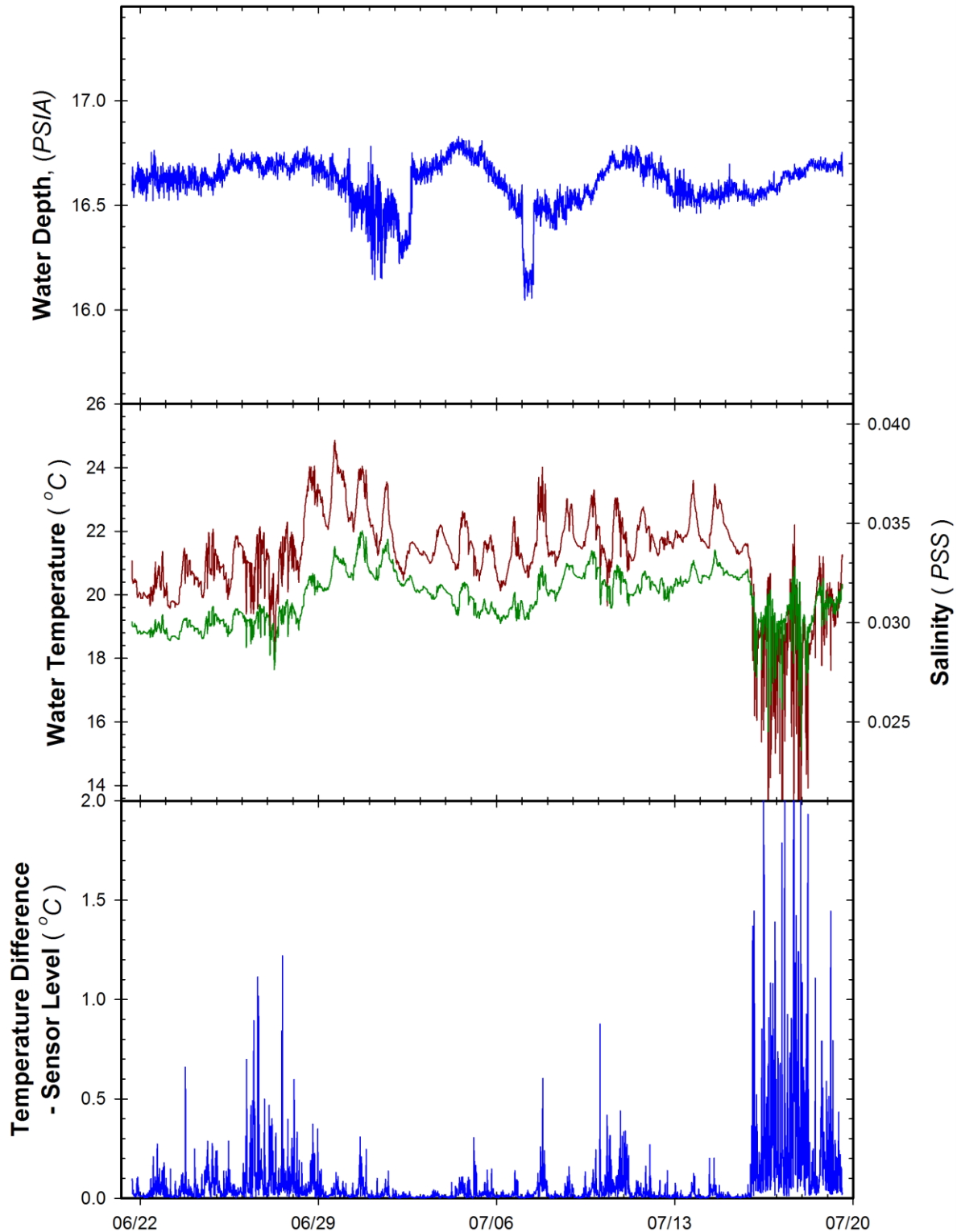
Figure 17. Environmental conditions encountered during the 1 month freshwater deployment adjacent to the NOAA GLERL Pier. Test sensor array deployed on bottom, variation in local water depth measured as PSIA by an SBE 26PLUS Wave gauge indicating active water flow around instrument (*Top Panel*). Variation in salinity (green) and temperature (red) at depth of instrument sensor detected by an SBE26 and two RBR Solo thermistors. (*Middle Panel*). Temperature range determined from max-min temperatures detected by RBR and SBE thermistors spanning instrument sensor array (*Bottom Panel*). A significant storm with winds over 25 mph event contributed to higher variation in environmental conditions near end of deployment period.
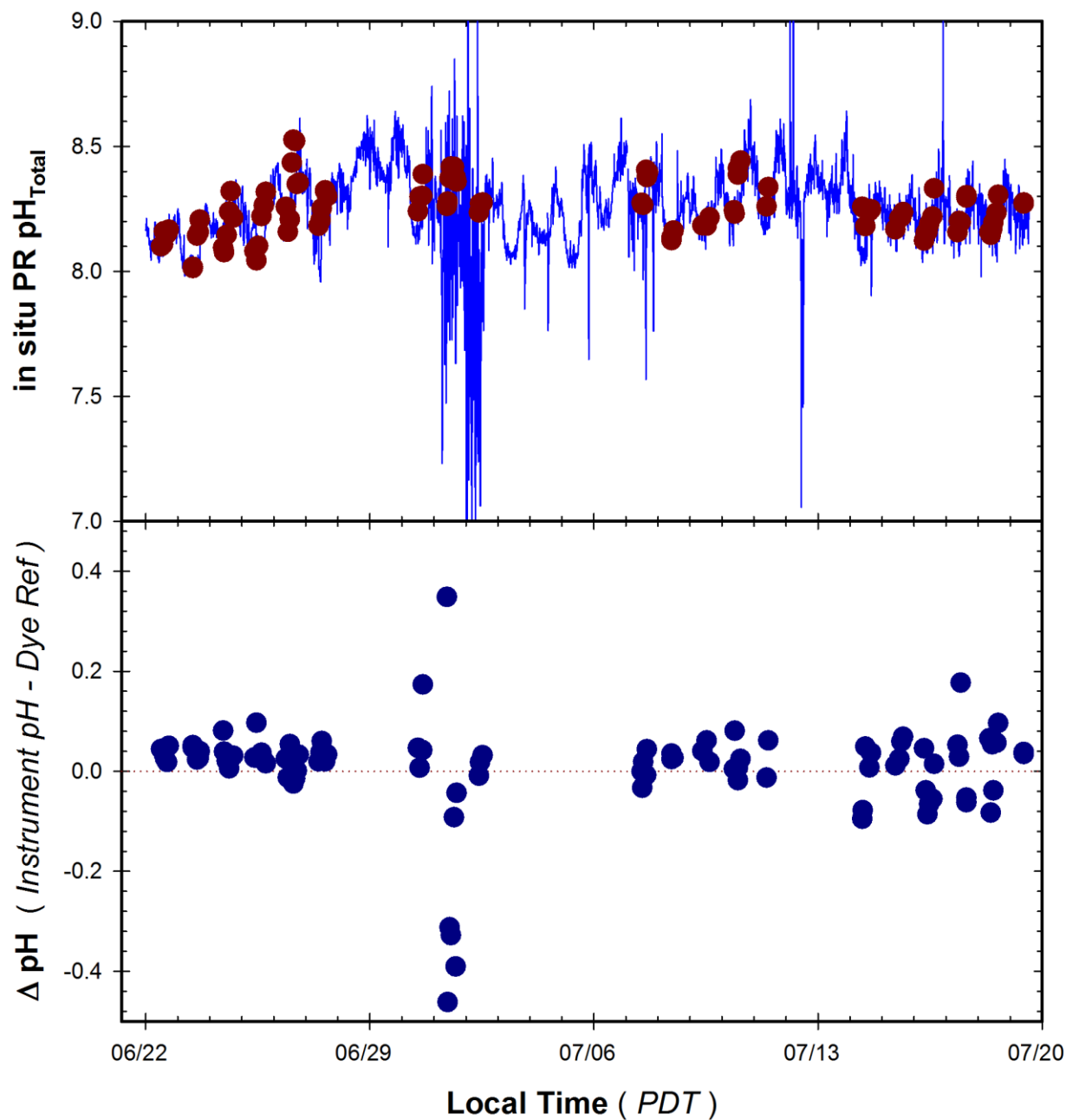
Figure 18. Time series of pH measured by a SAMI-pH(PR) instrument deployed during the 1 month GLERL field trial. *Top Panel*: Continuous pH$_{Total}$ recordings from instrument (blue line) and pH$_{Total}$ of adjacent grab samples determined by phenol red dye (red circles). *Bottom Panel*: Difference in measured pH relative to reference samples (Instrument − Dye Ref) observed during deployment. Zero offset is represented by dashed line as pH$_{Total}$ scale is reported by both the SAMI-pH sensor and the dye reference method.
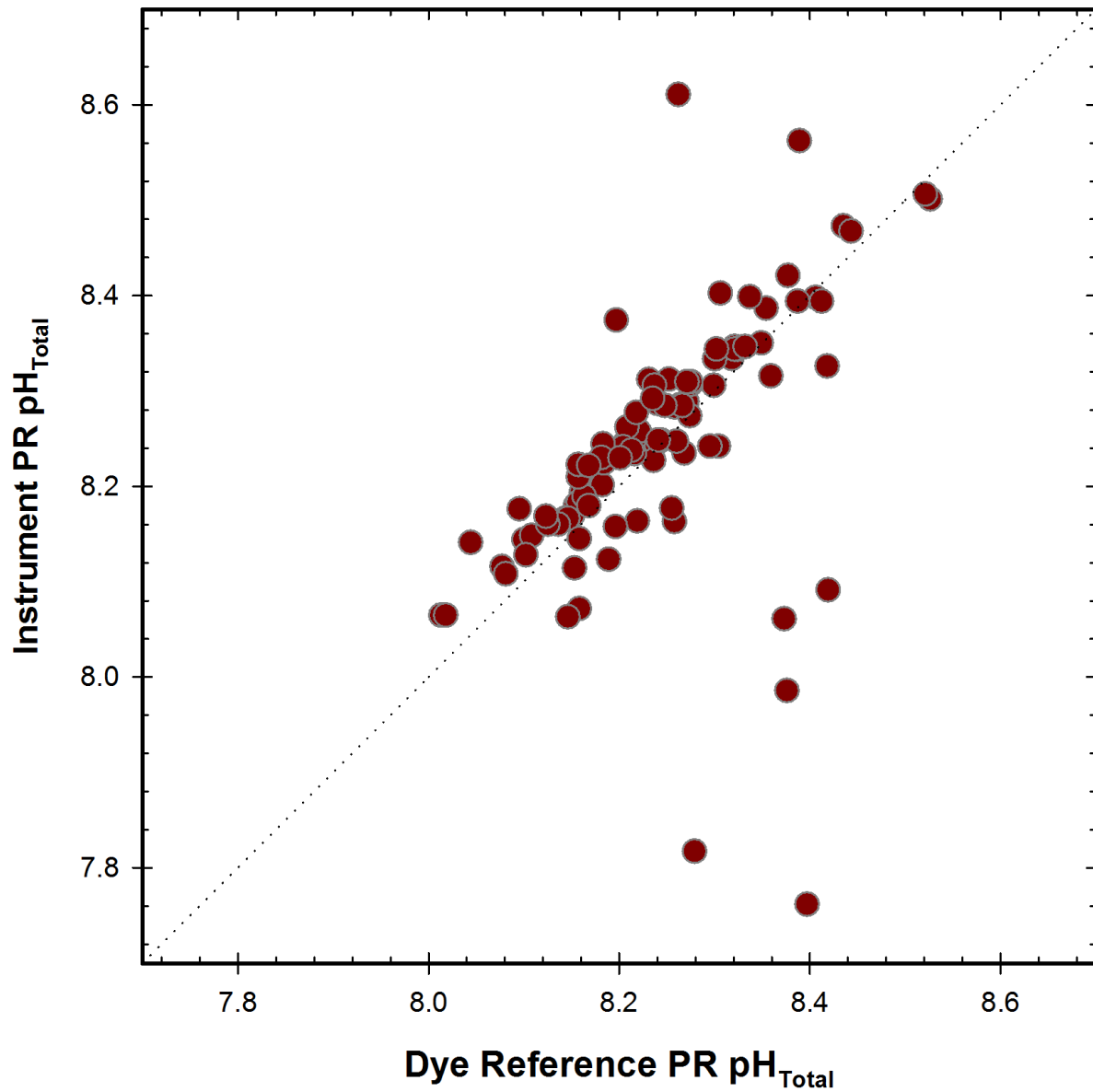
Figure 19. GLERL field response plot for a SAMI-pH(PR) instrument compared to reference pH_{Total} samples. Fine dotted line represents 1:1 correspondence, as pH_{Total} scale is reported by both the SAMI-pH instrument and the dye reference method.
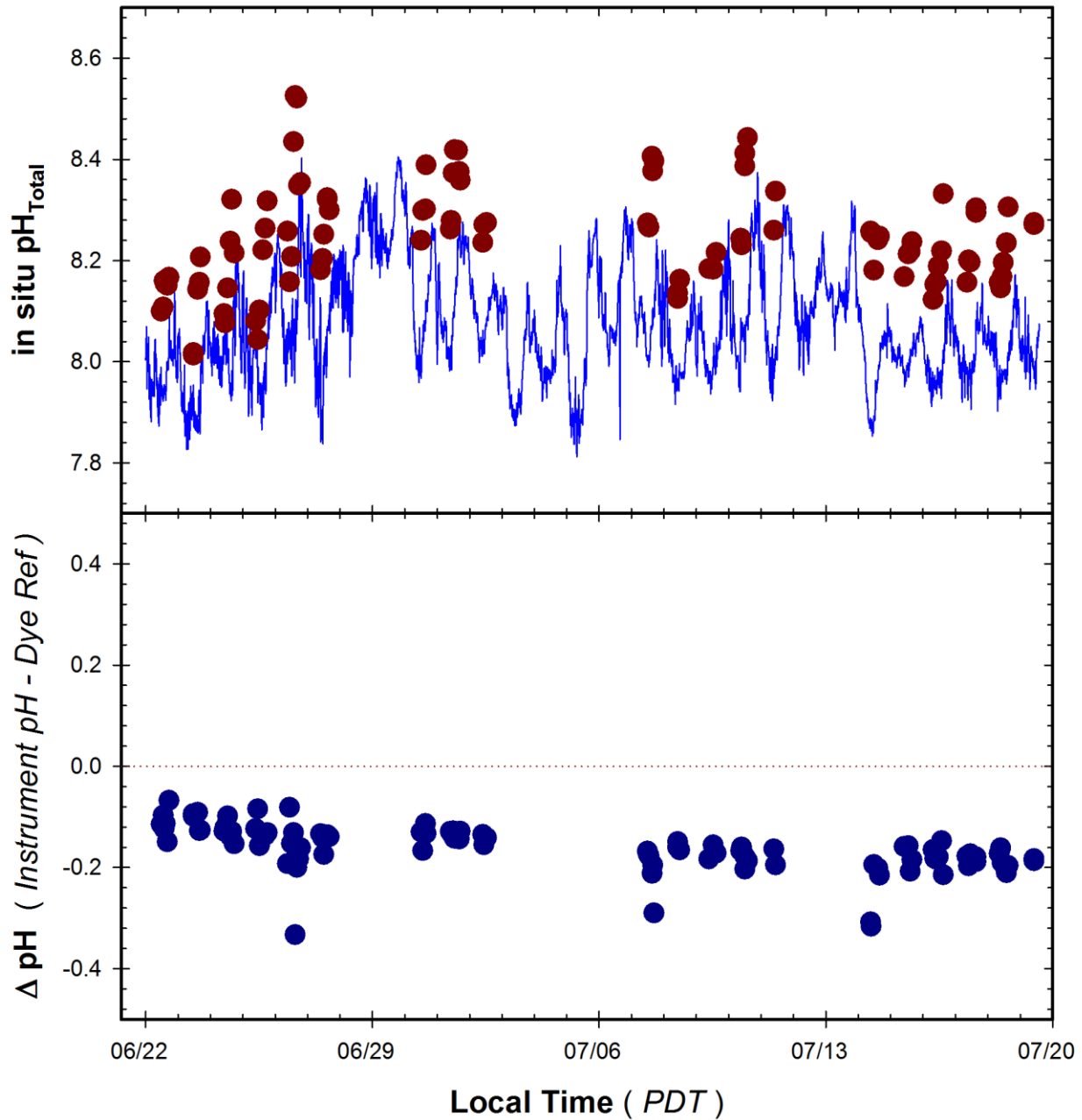
Figure 20. Time series of pH measured by a SAMI-pH(mCP) instrument deployed during the 1 month GLERL field trial. *Top Panel*: Continuous pH$_{Total}$ recordings from instrument (blue line) and pH$_{Total}$ of adjacent grab samples determined by phenol red dye (red circles). *Bottom Panel*: Difference in measured pH relative to reference samples (Instrument – Dye Ref) observed during deployment. Zero offset is represented by dashed line as pH$_{Total}$ scale is reported by both the SAMI-pH sensor and the dye reference method.
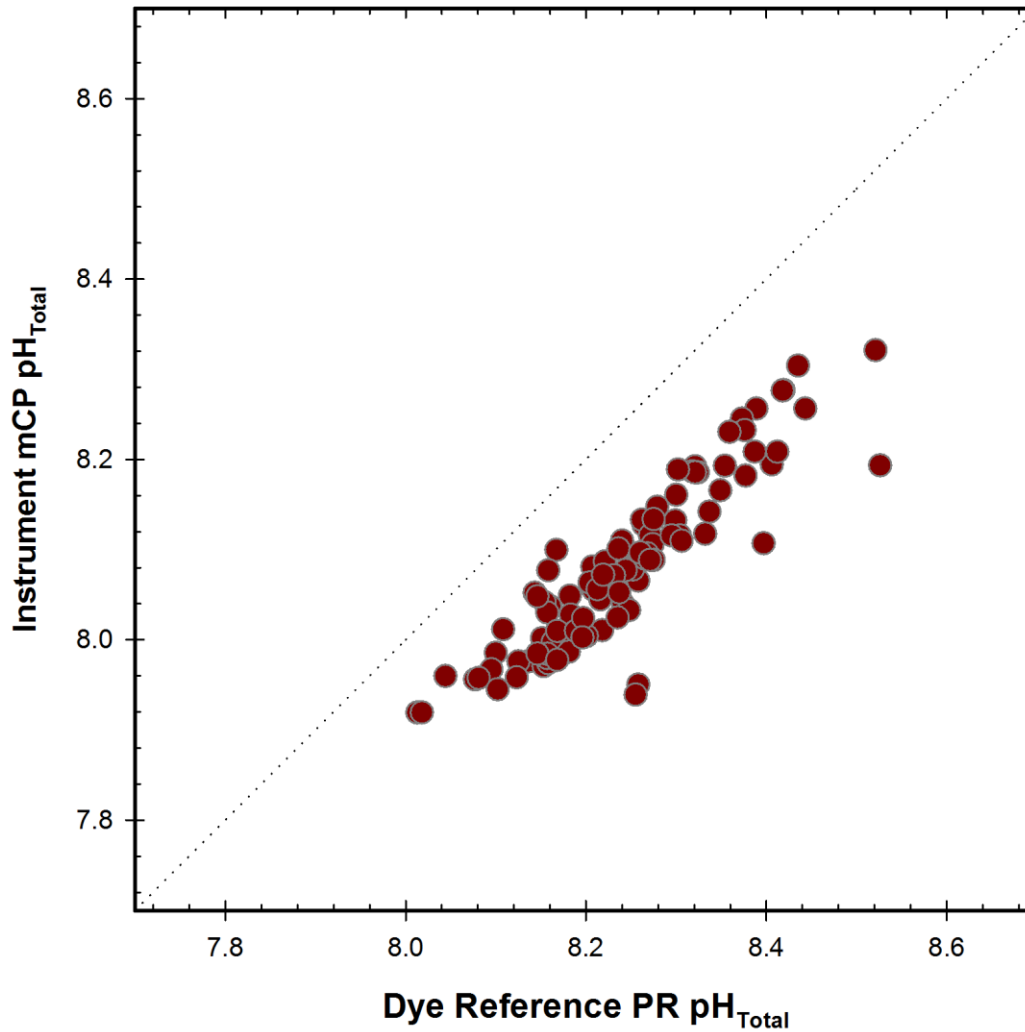
Figure 21. GLERL field response plot for a SAMI-pH(mCP) instrument compared to reference pH_{Total} samples. Fine dotted line represents 1:1 correspondence, as pH_{Total} scale is reported by both the SAMI-pH instrument and the dye reference method. Offset likely represents calibration differences in mCP and PR dye ratio algorithms at low salinity, high pH waters (*see Fig. 19*).
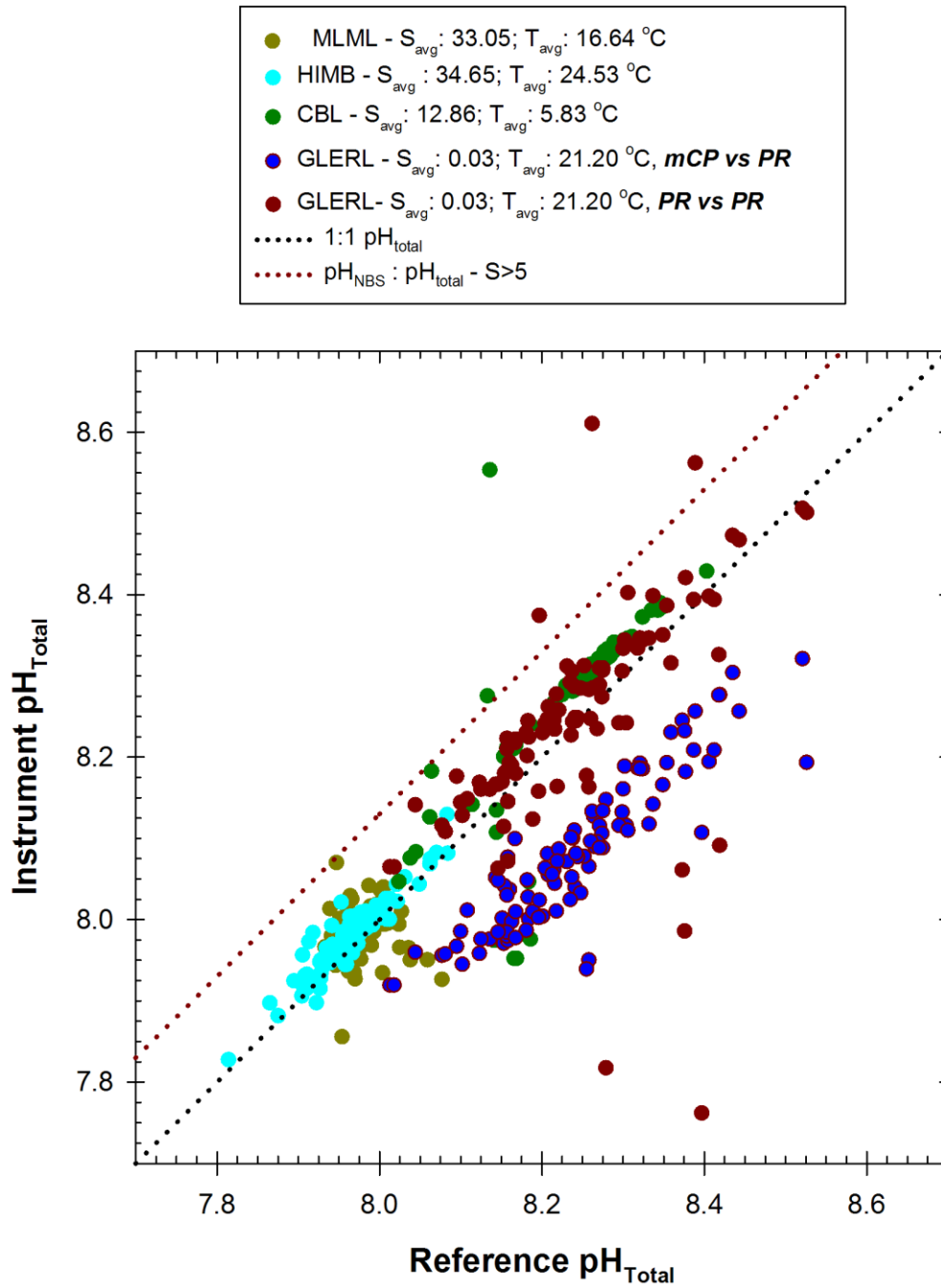
Figure 22. Global response plot for SAMI-pH(mCP) and SAMI-pH(PR) observed during ACT field trials. Fine dotted line represents 1:1 correspondence, coarse dotted line represents correspondence predicted based on difference between pH_{NBS} and pH_{Total} in seawater ($pH_{NBS} = pH_{Total} + 0.13$).
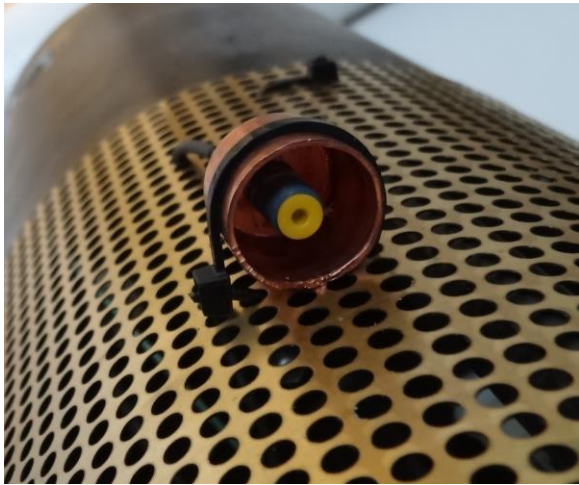
Photo 9. The SAMI pH prior to and following the 1 month deployment for the Great Lakes field test.

QUALITY ASSURANCE AND QUALITY CONTROL

A variety of procedures were established for this verification to ensure the highest quality of data possible. Foremost was the development of detailed, scientifically reviewed testing protocols and state of the art analytical methodologies. Each reference sample generated was analyzed in triplicate and any result with a standard deviation of greater than 0.01 pH units was initially flagged. If there was an obvious reason to exclude any of the triplicates because of handling errors that was done and a new value computed from the remaining two replicates. Reference values with a standard deviation of greater than 0.02 were discarded. External standards were run weekly to verify consistency in performance of the spectrophotometric dye analysis.

Quality Control Samples

Each site also conducted weekly field duplicates and one field trip blank, with the exception that no field trip blank was conducted in Hawaii and only one field duplicate was conducted during the Moss Landing moored deployment. Results of these QA/QC samples are presented below in Tables 4 – 7. The global average of the standard deviation among field duplicates for all field test sites was 0.005 ±0.008 (N=21), with two values exceeding our expected quality threshold of better than 0.01 pH units. These results attest to the representativeness of our sampling to water mass being analyzed by the test instruments and to consistent sample handling. The three field trip blanks that were performed had a standard deviation among the paired samples of 0.003, 0.002, and 0.002. The absolute difference between the paired samples was 0.004, 0.003, and 0.003. The field trip blanks attest to proper cleaning and handling of sampling equipment and sample processing.

Table 4. Results of Field Duplicates (FD) and Field Trip Blank (FTB) for the Moss Landing Harbor, CA mooring test. The field trip blank was determined with a Seawater CRM from the Dickson Lab at Scripps Institute of Oceanography.

Date/Time	Rep	Temp	Salinity	pH	Std Dev	Mean	Std Dev	ABS Diff
9-10-13 14:00	FD1	15.0	33.0	8.059	.003	8.049	.015	.021
	FD2			8.038	.001			
9-5-13 11:20	FTB1	22.0	35.0	7.974	.001	7.972	.003	.004
	FTB2			7.970	.001			

Table 5. Results of Field Duplicates (FD) for the Kaneohe Bay, HI mooring test. No field trip blank was determined at this test site.

Date/Time	Rep	Temp	Salinity	pH	Std Dev	Mean	Std Dev	ABS Diff
11-17-13 10:00	FD1	25.2	33.64	7.907	.001	7.908	.002	.003
	FD2			7.910	.001			
11-20-14 11:00	FD1	26.1	34.36	8.012	.001	8.011	.002	.003
	FD2			8.009	.002			
11-27-14 10:00	FD1	24.7	35.03	7.968	.001	7.968	0	0
	FD2			7.968	.001			
12-9-13 10:00	FD1	25.7	34.72	7.909	.001	7.909	0	0
	FD2			7.909	.001			
12-23-14 10:30	FD1	25.0	24.7	7.949	.007	7.947	.003	.004
	FD2			7.945	.001			
1-7-14 10:30	FD1	22.6	34.89	8.013	.001	8.010	.003	.005
	FD2			8.008	.001			
1-23-14 10:00	FD1	22.5	34.76	7.914	.001	7.915	.002	.002
	FD2			7.916	0			
2-6-14 10:00	FD1	23.4	33.99	7.958	.002	7.959	.001	.002
	FD2			7.959	0			
2-9-14 9:30	FD1	23.4	33.43	7.968	.001	7.968	0	0
	FD2			7.968	.001			

Table 6. Results of Field Duplicates (FD) and Field Trip Blank (FTB) for the Chesapeake Bay, MD mooring test. The field trip blank was determined with a mixture of Seawater CRMs from the Dickson Lab at Scripps Institution of Oceanography.

Date/Time	Rep	Temp	Salinity	pH	Std Dev	Mean	Std Dev	ABS Diff
3-20-14 11:30	FD1	4.4	13.87	8.111	.003	8.113	.034	.0048
	FD2			8.115	.004			
3-20-14 15:00	FD1	4.6	13.98	8.124	.004	8.121	.004	.005
	FD2			8.119	.003			
3-27-14 10:00	FD1	4.3	13.02	8.065	.004	8.065	.000	.000
	FD2			8.065	.001			
3-27-14 14:00	FD1	4.7	12.87	8.090	.002	8.090	.001	.0013
	FD2			8.089	.002			
4-3-14 9:30	FD1	6.9	13.19	7.992	.001	7.991	.001	.0015
	FD2			7.991	.009			
4-3-14 13:00	FD1	7.5	13.27	8.054	.001	8.055	.001	.001
	FD2			8.056	.007			
4-10-14 8:00	FTB1	9.5	9.28	8.122	.003	8.120	.002	.003
	FTB2			8.118	.003			

Table 7. Results of Field Duplicates (FD) and Field Trip Blank (FTB) for the Great Lakes, MI mooring test. The field trip blank was determined with a mixture of commercial NBS buffers.

Date/Time	Rep	Temp	Salinity	pH	Std Dev	Mean	Std Dev	ABS Diff
6-23-14 11:15	FD1	19.7	0.029	8.013	.003	8.016	.004	.005
	FD2			8.018	.003			
6-27-14 14:30	FD1	21.0	0.030	8.324	.004	8.322	.002	.003
	FD2			8.321	.008			
7-8-14 10:15	FD1	21.1	0.031	8.136	.001	8.131	.008	.011
	FD2			8.125	.004			
7-10-14 12:00	FD1	21.0	0.031	8.387	.010	8.400	.018	.025
	FD2			8.412	.006			
7-17-14 15:15	FD1	20.4	0.032	8.304	.008	8.300	.006	.009
	FD2			8.295	.007			
7-19-14 10:15	FD1	20.4	0.032	8.275	.002	8.273	.003	.004
	FD2			8.271	.008			
7-14-14 9:15	FTB1	22.0	0.032	8.258	.005	8.257	.002	.003
	FTB2			8.255	.003			

External Quality Assurance and Quality Control Assessments

An effective assessment program is an integral part of ACT's quality system for performance evaluations of marine sensor technologies. The ACT Quality Assurance (QA) Manager independently conducted three technical systems audits (TSA), an audit of data quality (ADQ), and a Data Quality Assessment (DQA) for the pH verification. The audits were conducted in accordance with the procedures described in EPA's *Guidance on Technical Audits and Related Assessments for Environmental Data Operations (EPA QA/G-7)*.

Technical System Audit - A Technical Systems Audit (TSA) is a thorough, systematic, on-site qualitative audit of ACT's sampling and measurement processes and procedures associated with a specific technology verification. The objective of a TSA is to assess and document the conformance of on-site testing procedures with the requirements of the Test Protocols, and associated SOPs. The ACT QA Manager conducted the following three TSAs over the course of the verification:

- The laboratory tests at the Hawai'i Institute of Marine Biology (HIMB), Coconut Island, Kaneohe Bay, Oahu, HI during December 11-15, 2012.
- The field tests at HIMB during November 19-21, 2013.

- The field tests at the Chesapeake Biological Laboratory, Solomons, MD during March 11-14, 2014

All of the audits included a review of staff, test procedures (sample collection, sample analysis, data processing, etc.), facilities, and documentation to assure compliance with the ACT Test Protocols and Standard Operating Procedures (SOPs) for the pH sensor verification.

The TSA of the laboratory tests at HIMB determined that there were several deviations in the conduct of the tests from the Test Protocols. Minor deviations, such as schedule changes and the collection of triplicate versus duplicate samples, were documented in laboratory records. Other deviations from the Test Protocols included deviations in the sequence of test conditions and dropping one type of burst sampling. These deviations also were documented and explanations of the rationale for the changes were provided. It was determined that none of the deviations in the Test Protocols had any effect on data quality for the verification and no corrective action was required.

The TSA's findings for the two field tests were positive in nature and indicated that these components of the pH sensor verification were being implemented in a manner consistent with the Test Protocols and SOPs. All phases of the implementation of the test reviewed during the TSA were acceptable and performed in a manner consistent with ACT data quality goals. The overall quality assurance objectives of the test were met.

ACT personnel are well-qualified to implement the verification, and demonstrated expertise in pertinent procedures. Communication and coordination among all personnel was frequent and effective. ACT's internal record keeping and document control is well organized. The ACT staff understands the need for QC, as shown in the conscientious development and implementation of a variety of QC procedures.

All samples and instrument measurements were collected, analyzed and cataloged as described in the Test Protocols and SOPs. For samples transported and analyzed by other affiliated laboratories, chain-of-custody (COC) protocols were practiced, specifying time, date, sample location, unique sample number, requested analyses, sampler name, time and date of transaction between field and laboratory staff, and name of receiving party at the laboratory. Examination of maintenance and calibration logs provided evidence of recent and suitable calibration of sampling and analytical equipment.

Audit of Data Quality - The ACT QA Manager conducted an Audit of Data Quality (ADQ) on verified data to document the capability of ACT's data management system (hardcopy and electronic) to collect, analyze, interpret, and report data as specified in the Test Protocols and SOPs. The ADQ implemented an examination of data after they had been collected and verified by ACT personnel. A representative set of no less than 10% of the data on core parameters was traced in detail from raw data from field and laboratory logs and instrument readouts, data transcription or transference, data manipulation, data reduction to summary data, data calculations, and final reported data.

The ADQ determined that ACT's sample measurement system performed in accordance with the performance goals specified in the Test Protocols and that the data were accumulated, transferred, reduced, calculated, summarized, and reported correctly. There is sufficient documentation of all procedures used in the data collection and analysis to verify that the data

have been collected in accordance with ACT quality objectives.

Data Quality Assessment - A Data Quality Assessment (DQA) is a scientific and statistical evaluation of validated data to determine if the data are of the right type, quality, and quantity to support their intended use. Data review is conducted to ensure that only sound data that are of known and documented quality and meet ACT technology verification quality objectives are used in making decisions about technology performance. Two steps were followed to determine the validity of the analytical data – data verification and validation.

Data verification evaluates the completeness, correctness, and consistency of the test data sets against the requirements specified in the Test Protocols. For the pH sensor verification, this step verified that:

- The raw data records were complete, understandable, well-labeled, and traceable;
- All data identified in the Test Protocols were collected;
- Instrument calibration and QC criteria were achieved;
- Data calculations were accurate.

Data validation established:

- Required sampling methods were used;
- Sampling procedures and field measurements met performance criteria;
- Required analytical methods were used;
- QC measures were obtained and criteria were achieved.

The DQA determined that ACT's data quality objectives were achieved. The DQA supports conclusions that:

- The sampling design performed very well and was very robust with respect to changing conditions.
- Sufficient samples were taken to enable the reviewer to see an effect if it were present.
- Data on the performance of the pH sensors are unambiguous, and the vendors and buyers can make informed choices about the performance of a sensor with a high level of certainty.

REFERENCES

American National Standards Institute and American Society for Quality Control (ANSI/ASQC). 2004. Quality Systems for Environmental Data Collection and Environmental Technology Programs: Requirements with Guidance for Use. E4-2004. Milwaukee, WI.

Culberson, C.H. 1981. Direct potentiometry, in: Whitfield, M., and Jagner, D. (Eds.), Marine Electrochemistry: a Practical Introduction. Wiley, pp. 187–261.

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

Easley R.A. and Byrne, R.H. 2012. Spectrophotometric calibration of pH electrodes in seawater using purified m-cresol purple. Environmental Science and Technology 46:5018-5024.

Liu, X., Patsavas, M., and Byrne, R.H. 2011. Purification and characterization of meta-cresol purple for spectrophotometric seawater pH measurements. Environmental Science and Technology 45:4862-4868.

Millero F.T., Graham T.B., Huang, F., Bustos-Serrano, H., and Pierrot, D. 2006. Dissociation constants of carbonic acid in seawater as a function of salinity and temperature. Marine Chemistry 100:80-94.

National Environmental Laboratory Accreditation Conference (NELAC). 2003. NELAC Standard. EPA/600/R-04/003. Washington, DC.

Pierrot, D., Lewis, E., and Wallace, D.W.R. 2006. MS Excel Program Developed for CO2 System Calculations. ORLN/CDIAC-105, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy. Oak Ridge, TN.

United States Department of Energy (U.S. DOE). 1994. Handbook of Methods for the Analysis of the Various Parameters of the Carbon Dioxide System in Sea Water; version 2, A.G. Dickson and C. Goyet, (eds.) ORNL/CDIAC-74.

United States Environmental Protection Agency (U.S. EPA). 2003. Guidance on Assessing Quality Systems. EPA/240/R-03/002: EPA QA/G-3. Washington, DC.

United States Environmental Protection Agency (U.S. EPA). 2006a. Guidance on Technical Audits and Related Assessments for Environmental Data Operations. EPA QA/G-7: EPA/600/R-99/080. Washington, D.C.

United States Environmental Protection Agency (U.S. EPA). 2006b. Data Quality Assessment: A Reviewer's Guide. EPA/240/B-06/002: EPAQA/G-9R. Washington, DC.

United States Environmental Protection Agency (U.S. EPA). 2006c. Data Quality Assessment: Statistical Tools for Practitioners EPA/240/B-06/003: EPA QA/G-9S. Washington, DC

United States Environmental Protection Agency (U.S. EPA). 2008. Guidance on Environmental Data Verification and Data Validation EPA/240/R-02/004: EPA QA/G-8. Washington, DC

Yao, W., and Byrne, R. H. 2001. Spectrophotometric determination of freshwater pH using bromocresol purple and phenol red. Environmental Science and Technology 35:1197-1201.

ACKNOWLEDGEMENTS:

February 20, 2015
———————————
  Date

Approved By: Dr. Mario Tamburri
ACT Executive Director
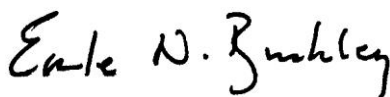
February 20, 2015
———————————
  Date

Approved By: Dr. Tom Johengen
ACT Chief Scientist

February 20, 2015
———————————
  Date

**Approved By: Dr. Earle Buckley**
Quality Assurance Supervisor

APPENDIX I

*Results of Rapid pH Shifts during Laboratory Testing using Acid/Base Additions*

Laboratory tests involved week long testing at a set combination of three temperatures (10, 20 and 30°C) and three salinities (0.3, 22 and 35 psu). After 4 to 6 days of testing at a stable T-S condition and ambient pH, pH was cycled over a reasonable range using acid-base additions to the water of the test tank (7.5 to 8.5 for seawater and 6.5 to 8.8 for freshwater). One or two, raised - lowered pH cycles were conducted at each T-S condition over the course of three to six hours. Acid/base additions were done by first mixing known quantities of acid/base into several liters of the current test solution and then adding this solution into the test tank to facilitate mixing and rapid equilibration. Shifts in pH were typically complete in 10 – 15 minutes, so usually within 1 measurement cycle of the instrument. There was no attempt to calculate a specific response time because the shifts were not implemented on a specific schedule relative to the sampling rate.

Results are presented for each of the three temperatures for a given salinity on the same page. Results for the Metrohm electrode measuring at 15 min intervals are presented in the solid magenta line along with the discrete dye reference pH measurement in yellow circles. The Metrohm electrode was directly calibrated to the dye pH results, hence the direct agreement.
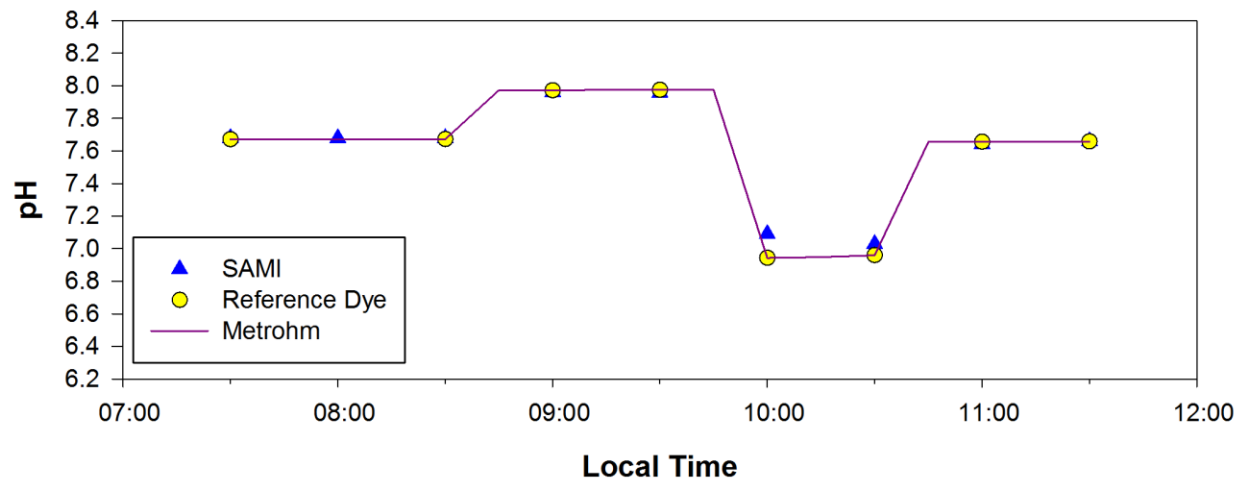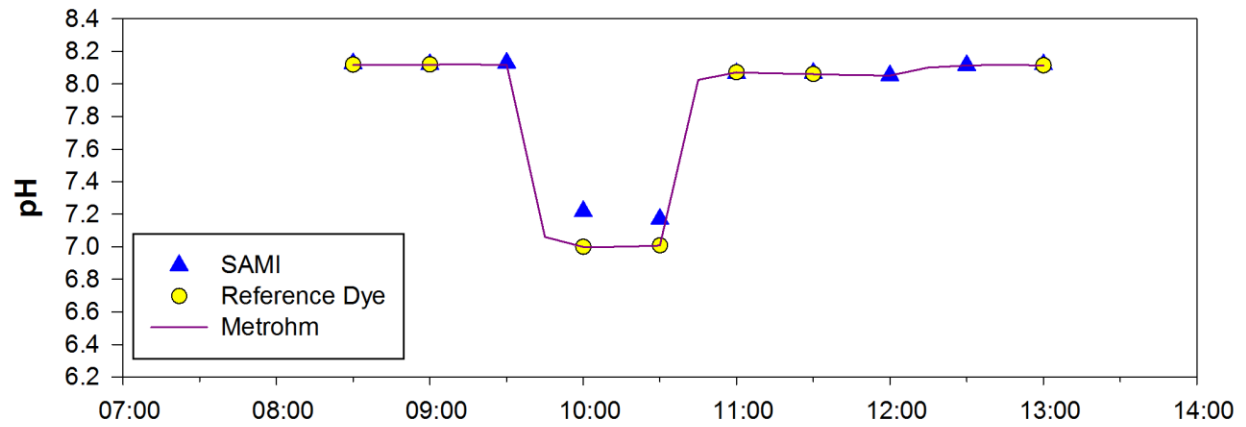
12/14/12 S = 35 T = 20°C
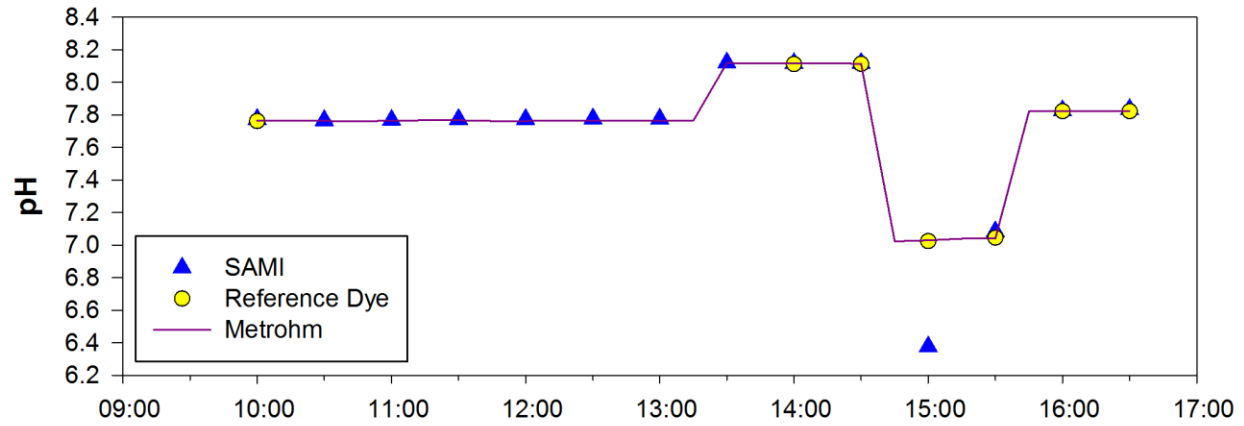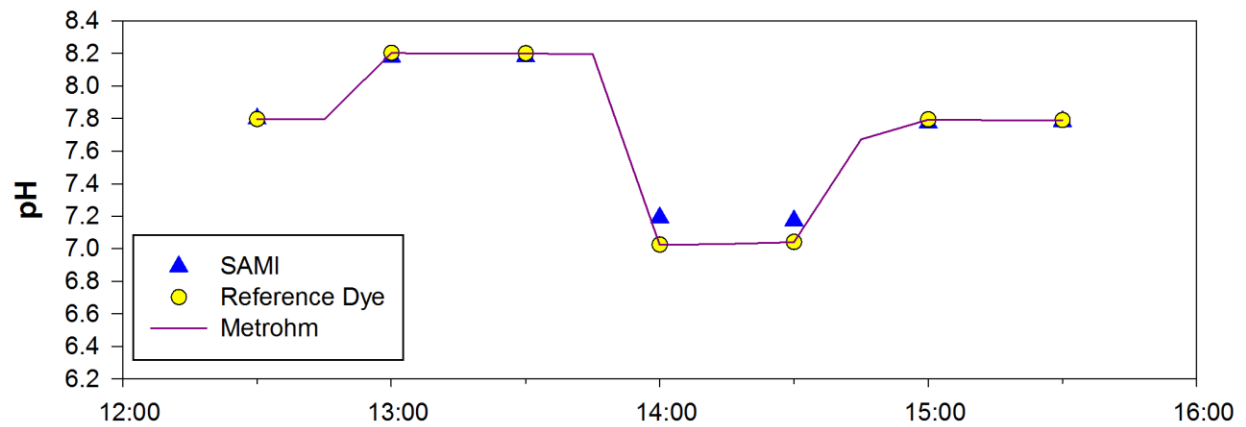

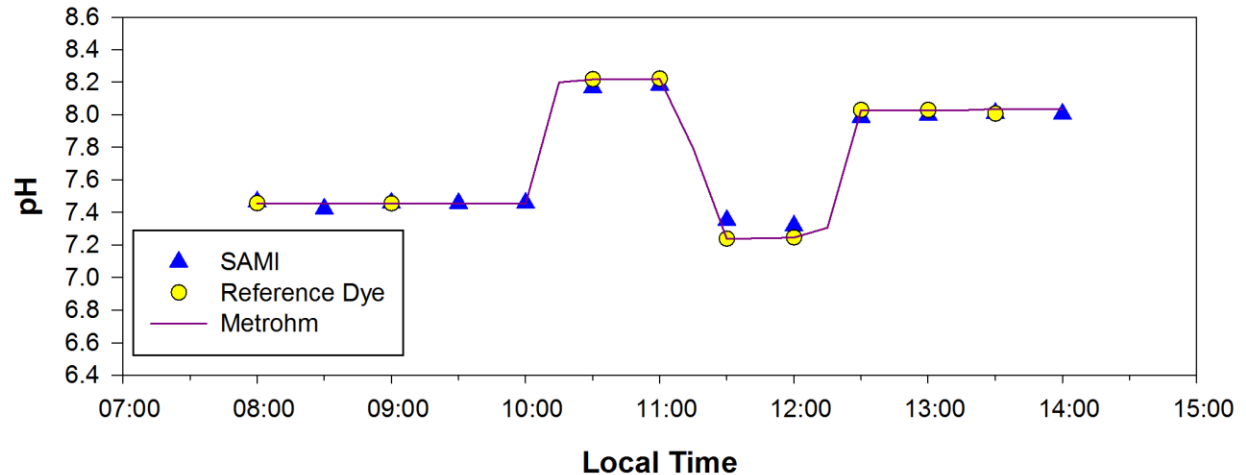**12/21/12 S = 35 T = 10°C**
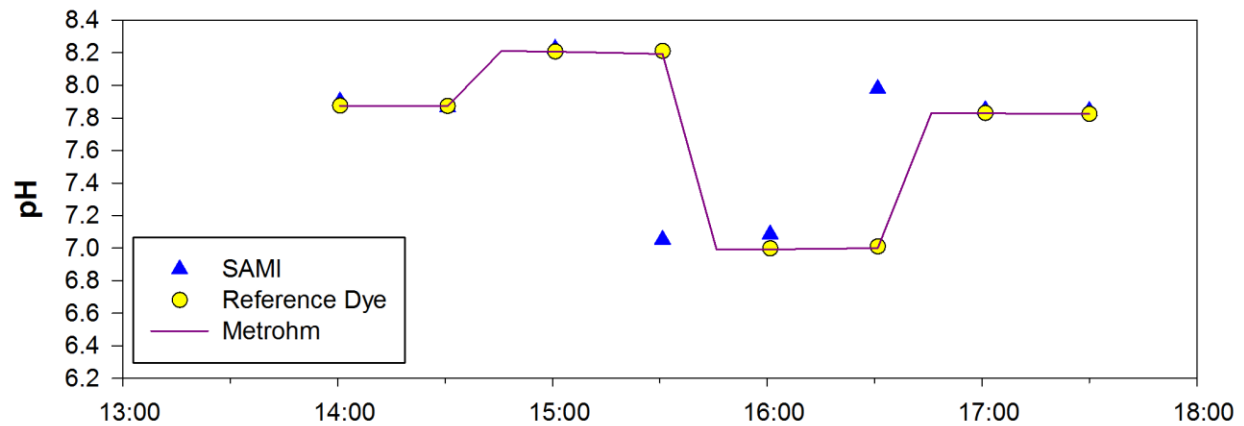
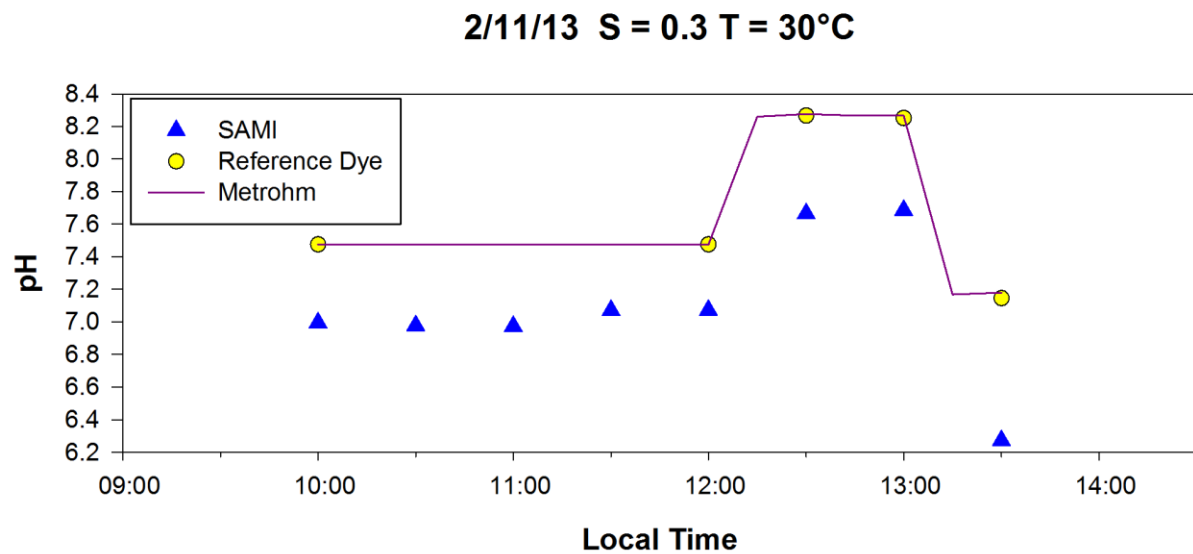**1/4/13 S = 35 T = 30°C**

1/11/13 S = 22 T = 10°C

1/18/13 S = 22 T = 20°C

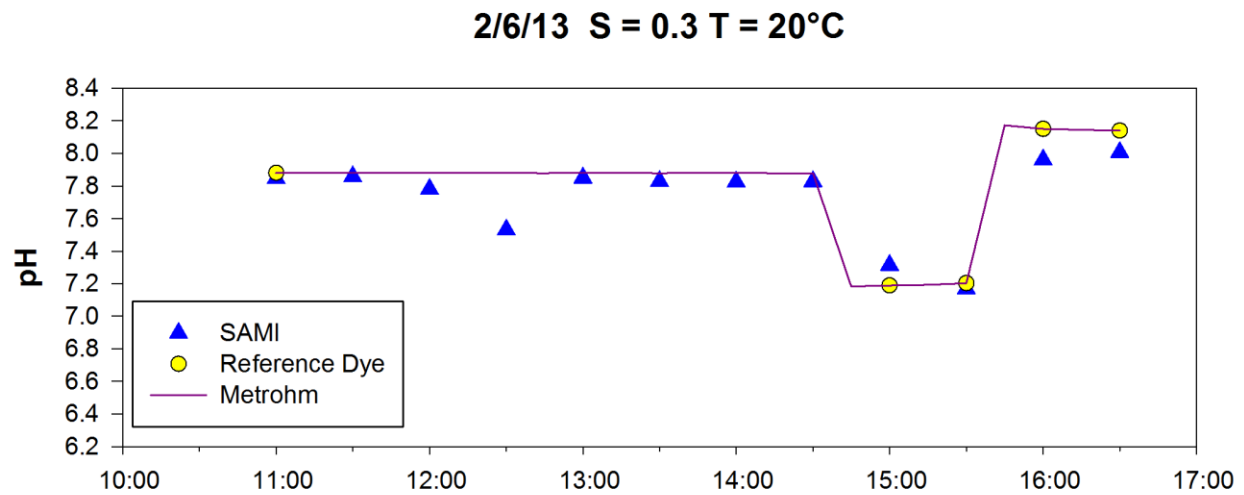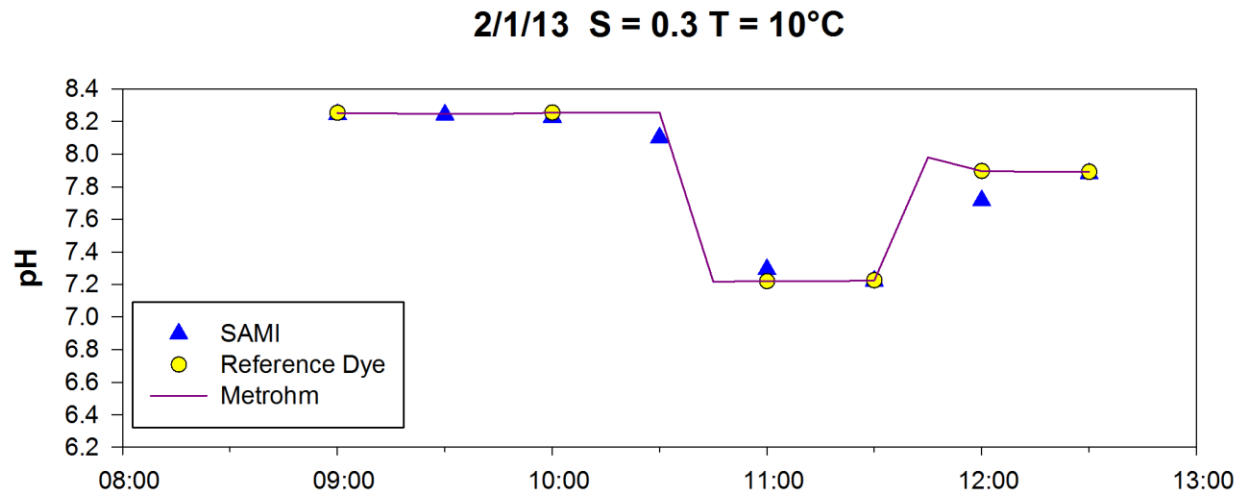### 1/25/13 S = 22 T = 30°C

2/1/13 S = 0.3 T = 10°C

2/6/13 S = 0.3 T = 20°C

2/11/13 S = 0.3 T = 30°C

Sunburst Sensors Response, January 12, 2015

General comments: We thank the ACT personnel and the scientists who participated in the analysis of quality control samples for their dedication to this important project.   Establishing methods to accurately measure ocean pH and to determine precision and accuracy between methods is important in studying the effects of climate change and acidification on marine and freshwater environments.

These tests were conducted in laboratory and field settings with a wide range of temperature, salinity, and pH. These varying conditions test the limits of the instruments, which are generally designed and optimized for more limited conditions. The SAMI-pH was designed to measure pH in the range of 7 - 9 at salinity 25 - 40, and temperatures 0°C - 40°C. These tests were an opportunity for additional rigorous field-testing of the SAMI-pH under more variable conditions. We were also able to test a prototype SAMI-pH unit with Phenol Red (PR) indicator in the freshwater deployment, as this indicator equilibrium range is more appropriate for freshwater pH.

Lab tests allowed collection of QC samples under static pH/salinity/temperature conditions, whereas in the field studies the diurnal pH signal ranged from ~0.1 (MLML) to ~0.5 (GERL). When pH is changing rapidly as in the GERL study, temporal and spatial differences between sensors and collection of QC samples can cause scatter between the methods. Temperature differences between sensors can have a significant effect on pH as well, because QC samples are measured at a constant temperature and corrected to the temperature of the in situ sensors. Considering the challenges, ACT personnel did an excellent job providing both controlled and field testing conditions representing fresh, brackish, and sea water at various temperatures and pH, and collecting and analyzing QC samples for the tests. The comparisons between QC samples and the SAMI-pH are addressed below.

Laboratory Test. Overall, the SAMI-pH had excellent performance, operating through 10 weeks of testing and producing 3140 out of 3154 possible measurements. The overall correlation was excellent and offsets between SAMI-pH and QC pH were good with the largest offsets being in freshwater (Figure 7). The offset originates from the equilibrium constant for the indicator mCP, which is not well known at low salinity, and the precision between pH measurements with mCP and the QC method (Phenol Red) are not well characterized. Bubbles that enter the optical path also likely caused scatter in the SAMI data as shown in Figures 4-6 where significant deviations from the pH baseline are evident. The SAMI-pHs are normally deployed > 3 m depth where bubbles are not significant. This study also provides us valuable data from which to develop better software for filtering outliers from field data.

Field Tests. The SAMI-pH produced excellent field results operating well through 22.5 of 24 weeks and collecting 10,483 useable measurements out of 11,375 possible. The SAMI-pH correlates well with the QC pH in all tests except for Moss Landing, where pH scatter later in the deployment, from a high particulate sample, skewed the data. The SAMI pH difference from the QC pH is excellent at salinity = 35 (Hawaii), since this is the condition that the instrument is designed for, and also very good in the other conditions tested. In freshwater, our experimental Phenol Red SAMI-pH performed better than the mCP SAMI, as expected, because the PR equilibrium constant is more appropriate for freshwater pH and the QC method uses PR. The SAMI-pH did not experience pH drift through deployments as long as 3 months, and pH measurements were not affected by biofouling. Each field test is discussed separately below. The only general concern in these studies was the observed temperature differences between sensors (See Figure A). In each test, two temperature sensors were deployed and differences between the two were recorded (e.g. Figure 8). Temperature differences of > 1°C were seen in some field studies, which would cause errors in the reported pH at in situ conditions of >0.015, after correction from the QC measurement to the in situ conditions. Moreover, these temperature differences also likely correspond to small scale pH (DIC and Alkalinity) spatial gradients making sampling very challenging in these environments. Although these temperature differences were dismissed as insignificant in the report, in fact, they could contribute significantly to the apparent accuracy and precision of the deployed instruments.

Moss Landing Marine Laboratories Field Deployment. During this deployment the SAMI performed well for ~3 out of 4 weeks, when the signal became noisy. This noise was possibly due to particles in the flow cell, as the water at this site (Photograph 2) was very high in particulates. The high particulates probably also contributed to a large difference between duplicate QC samples (0.021 pH units in ACT Report, Table 3), which is similar to the offset between the SAMI-pH and the QC samples (0.039). This was the only data set in which the SAMI-pH did not correlate to the QC samples, and the likely causes are that the data from the last week of the deployment are highly scattered, and that the range of the pH signal is relatively small at this site (≤0.15 Figure 9).
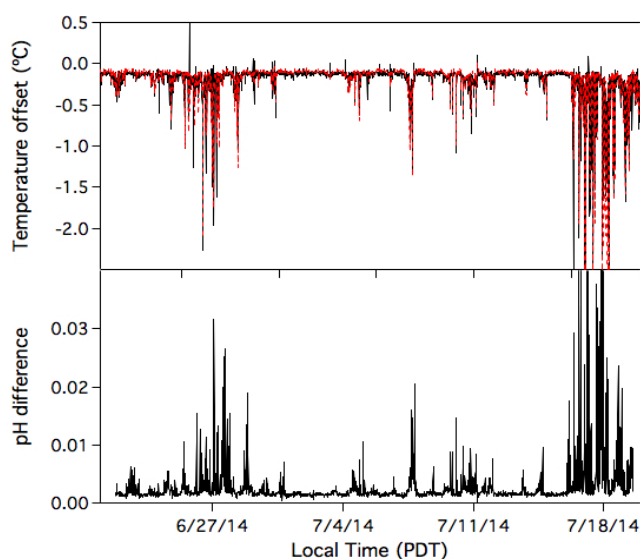
Figure A. Difference in temperature measured by mCP SAMI and SBE (top panel, black), PR SAMI and SBE (top panel, red), and the associated pH error from calculating the QC pH with these temperature errors (bottom panel).

Crimp II Buoy, Coconut Island, Hawaii. The SAMI-pH had excellent performance during this deployment, with excellent accuracy compared to the QC samples (Figures 12 and 13). The HIMB site is a typical SAMI deployment application.

Chesapeake Biological Laboratory. The SAMI-pH performed well for ~3 weeks during this deployment, after which the fluid path appears to have become clogged. The correlation between SAMI-pH and QC pH was good, considering we have done limited testing at this salinity (10-13).

Great Lakes Environmental Research Laboratory. At this site, we deployed 2 SAMIs, one with mCP and one with PR reagent. pH measurements with both instruments were highly correlated to the QC pH, with the PR SAMI accuracy being better, highlighting the importance of the accuracy of the equilibrium constant in the spectrophotometric pH measurement, as discussed previously.

At this site, we had the opportunity to compare temperature from two SAMIs to the QC temperature. Figure A shows that the temperature of the two SAMIs is very similar, with offsets as large as 2°C from the QC temperature (top panel), and associated pH difference as large as 0.03 (bottom panel). The difference between QC and SAMI temperature might be due to stratification, temporal sampling differences during rapidly changing conditions, and undoubtedly contribute to scatter between SAMI and QC pH measurements in all field deployments. These challenges were not discussed in the ACT report.

Summary: These field and lab tests help establish the reliability, accuracy and precision of the SAMI-pH sensor design and have given us opportunity to examine broadening its application. We once again thank ACT personnel for all their hard work performing this evaluation. It is a very valuable service to the scientific research community.

Sincerely,

James Beck
CEO/President
Sunburst Sensors, LLC

Mike DeGrandpre
Director, R & D
Sunburst Sensors, LLC

Reggie Spaulding
Research Chemist
Sunburst Sensors, LLC