Updated Methods for Global Locally-Interpolated Estimation of Alkalinity, pH, and Nitrate

Carter, B. R.[1,2], Feely, R. A.[2], Williams, N. L.[3], Dickson, A. G.[4], Fong, M. B.[4], Takeshita, Y.[5],

[1]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, 3737 Brooklyn Avenue, Seattle, WA, USA, 98105

[2]Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA, USA, 98115

[3]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA, 97331

[4]Scripps Institution of Oceanography, University of California San Diego, San Diego, CA, USA, 92093

[5]Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA, USA 95039

Abstract

We have taken advantage of the release of version 2 of the Global Data Analysis Project (GLODAPv2) data product (Olsen et al. 2016) to refine the Locally Interpolated Alkalinity Regression (LIAR) code for global estimation of total titration alkalinity of seawater (A_T), and to extend the method to also produce estimates of nitrate (N) and *in situ* pH (total scale). The updated MATLAB software and methods are distributed as supplementary materials for this paper and referred to as LIAR version 2 (LIARv2), Locally Interpolated Nitrate Regression (LINR), and Locally Interpolated pH Regression (LIPHR). Collectively they are referred to as Locally Interpolated Regressions (LIRs). Relative to LIARv1, LIARv2 has an 18% lower average A_T estimate RMSE, improved uncertainty estimates, and fewer regions in which the method has little or no available training data. LIARv2, LINR, and LIPHR produce estimates globally with skill that is comparable to or better than regional alternatives used in their respective regions. LIPHR pH estimates have an optional adjustment to account for ongoing ocean acidification. We have used the improved uncertainty estimates to develop LIR functionality that selects the lowest-uncertainty estimate from among possible estimates. Current and future versions of LIR software will be available on GitHub at https://github.com/BRCScienceProducts/LIRs.

Introduction

The LIAR method and software was developed to estimate A_T globally from other measurable seawater properties (Carter et al. 2016b). The original application for the method was providing A_T estimates as a second carbonate parameter for use with data from the emerging network of biogeochemical floats that measure pH (Wanninkhof et al. 2016; Johnson and Claustre 2016; Johnson et al. 2016). However, LIAR may also prove useful for studies or models interested in

estimating a climatological A_T baseline with limited variability or deviations from such a baseline (e.g. Carter et al. 2016a).

LINR and LIPHR are primarily intended to provide cross-comparisons for nitrate (N) and pH sensor measurements that can be used to assess potential float sensor errors or measurement drifts. Profiling biogeochemical floats cannot typically be retrieved for sensor recalibration, so it is important to have independent means to assess such problems that may arise during or after float deployment. A common approach to this problem is to use known atmospheric, surface, or climatological concentrations (Bushinsky et al. 2016; Plant et al. 2016; Takeshita et al. 2013) to recalibrate sensors, but such known values are not always available for N and pH. LINR and LIPHR are designed to provide estimated values in the stable 1000-2000 m depth range of the ocean as alternatives. All three LIRs have secondary scientific applications when A_T, N, or pH estimates are desirable and some seawater property information is available.

By default, LIRs have the limitation that they are unable to capture changes in the relationships between the estimated properties and the predictor properties. An example of such an unresolved change comes from the influence of ocean acidification (OA), the effect of continually increasing ocean storage of anthropogenic carbon dioxide (CO_2) on seawater pH. LIPHR contains an option to adjust for the effects of OA on pH, but we expect OA induced pH changes to result in LIPHR estimates becoming less skillful over time even when this adjustment is used because the adjustment does not account for regional or temporal variations in the rate of OA. All three LIRs are expected to be most skillful at reproducing measurements below the ocean surface where the effects of OA and other changes are smaller, or for estimates made close in time and space to the measurements used to train the LIRs. Another limitation of these algorithms is that they break down any time relationships between predictors and the estimated properties become significantly nonlinear. An example of a region where estimate skill would be expected to be diminished by this limitation would be on the margins of O_2 deficient zones where the influences of both denitrification and aerobic respiration can be important.

Regressions for estimating pH, N, and A_T have been reported numerous times. A_T regressions are the most common variant (e.g. McNeil and Sasse 2016; Lee et al. 2006; Alin et al. 2012; Velo et al. 2013; Bostock et al. 2013; Millero et al. 1998; Sasse et al. 2013) with regressions for pH being less frequently reported (e.g. Juranek et al. 2011; Alin et al. 2012; Williams et al. 2016) and nitrate regressions being even less frequently reported still (e.g. Williams et al. 2016, supplementary information). The LIRs presented here make improvements over earlier versions with respect to global applicability, ease of use, and the ability to scale uncertainty estimates based on input uncertainties. Critically, they also produce estimates that reproduce pH measurements at least as skillfully as earlier versions. The bulk of the improvement results from the larger quantity and span of data available through the GLODAPv2 data product (Olsen et al. 2016) than was available to train earlier methods. A similar method to the LIRs developed

recently is the Carbonate system and Nutrients concentration from hYdrological properties and Oxygen using a Neural-network (CANYON) (Sauzède et al. 2017). CANYON was also trained using the GLODAPv2 data product and is capable of estimating pH, A_T, silicate (*Si*), *N*, total dissolved inorganic carbon (C_T), and pCO_2 globally from O_2, temperature, salinity (*S*), latitude, longitude, depth, and day of year. We expect the LIRs we propose here will provide complementary estimates to those provided by CANYON for most applications, and note that the LIRs presented here do not require O_2 and temperature as measurement inputs.

In the remainder of this paper we describe version 2 of the LIAR software (LIARv2) in the context of the improvements relative to version 1 (LIARv1: Carter et al. 2016b), and extend the LIR approach to nitrate and *in situ* total scale seawater pH estimates with LINR and LIPHR. Particular attention is paid to new procedures required to address complications with extending the LIR framework to pH measurements.

2 Methods

2.1 Summary of LIR methods

As with LIARv1, the LIR methods developed here use regression coefficients that are determined at each location on a 5° latitude and longitude grid with 33 depth surfaces (44,957 total locations). Each set of regression coefficients is determined using a robust multiple linear regression of the subset of measurements from the global training dataset that are found within a volume defined by latitude, longitude, and depth/density windows of the grid coordinates (the same grid used by Carter et al., 2016b). The windows used are 5° for latitude, ($10^\circ / \cos(\text{latitude})$) for longitude, and either 0.01 kg m^{-3} for potential density or 50 m for depth (whichever is more inclusive). The dimensions of these windows are iteratively scaled by a factor of the iteration number until at least 100 measurements are selected to train each regression. When generating estimates, the LIAR software then interpolates between regression coefficients specific to these grid locations to arbitrary locations where the user desires regression estimates. LIARv2 works with 16 different combinations of the predictor variables: salinity *S*, potential temperature θ, nitrate *N*, apparent oxygen utilization AOU, and silicate *Si*. LINR uses the same combinations as LIAR with phosphate *P* in place of *N* in the 8 regressions that included *N*. LIPHR uses the same predictors as LIAR, but also includes depth (*z*) in meters as a predictor. This additional predictor is intended to allow for the effects of pressure on *in situ* pH. The specific combinations of variables used are indicated in sections 3.1 through 3.3. A full description of the LIARv1 method is provided by Carter et al. (2016b). In this update we focus on how LIARv2, LIPHR, and LINR adapt and improve upon the LIARv1 methods.

In some instances where spectrophotometric pH measurements are unavailable, we use *in situ* total scale pH as calculated from A_T and C_T. These calculations were made with carbonate constants from Lueker et al. (2000), borate dissociation coefficients from Dickson (1990), total

borate from Lee et al. (2010), and KF from Perez and Fraga (1987). Calculations are performed using the CO2SYS for MATLAB routine by van Hueven et al. (2011).

2.2 Data products used to train and test LIRs

The primary improvement in LIARv2 relative to LIARv1 stems from regression coefficients having been re-estimated using the Global Data Analysis Project version 2 data product (GLODAPv2). All measured and calculated values in GLODAPv2 were used except those from 161 cruises (40,303 measurements) that had A_T quality control (QC) adjustments of ±10 μmol kg^{-1} or greater, were flagged as poor data, or were not quality controlled for A_T (Olsen et al. 2016). The new training data set is comprised of 236,852 A_T measurements and A_T estimates from CO_2-calculations based on other CO_2 parameters, 211,704 of which had the property measurements required for training all 16 regressions (Figure 1). The LIAR test data set omits the 2,279 calculated A_T values that are included in the training data set. We use the coefficient re-estimation strategy used by Carter et al. (2016b) to allow overlap between our training and test data sets without compromising the validity of the assessments (described in Section 3).

LINR regression coefficients were estimated using 684,475 *N* measurements, 569,761 of which had associated property measurements required for training all 16 regressions. This training dataset is all GLODAPv2 data product *N* measurements excepting those from 187 cruises that had multiplicative adjustments greater than 10%, that were not QC'd, or that were flagged as having poor quality measurements. GLODAPv2 QC protocols changed reported negative *N* values to 0 μmol kg^{-1}. The LINR code does likewise. The LINR test data set is identical to the training data set.

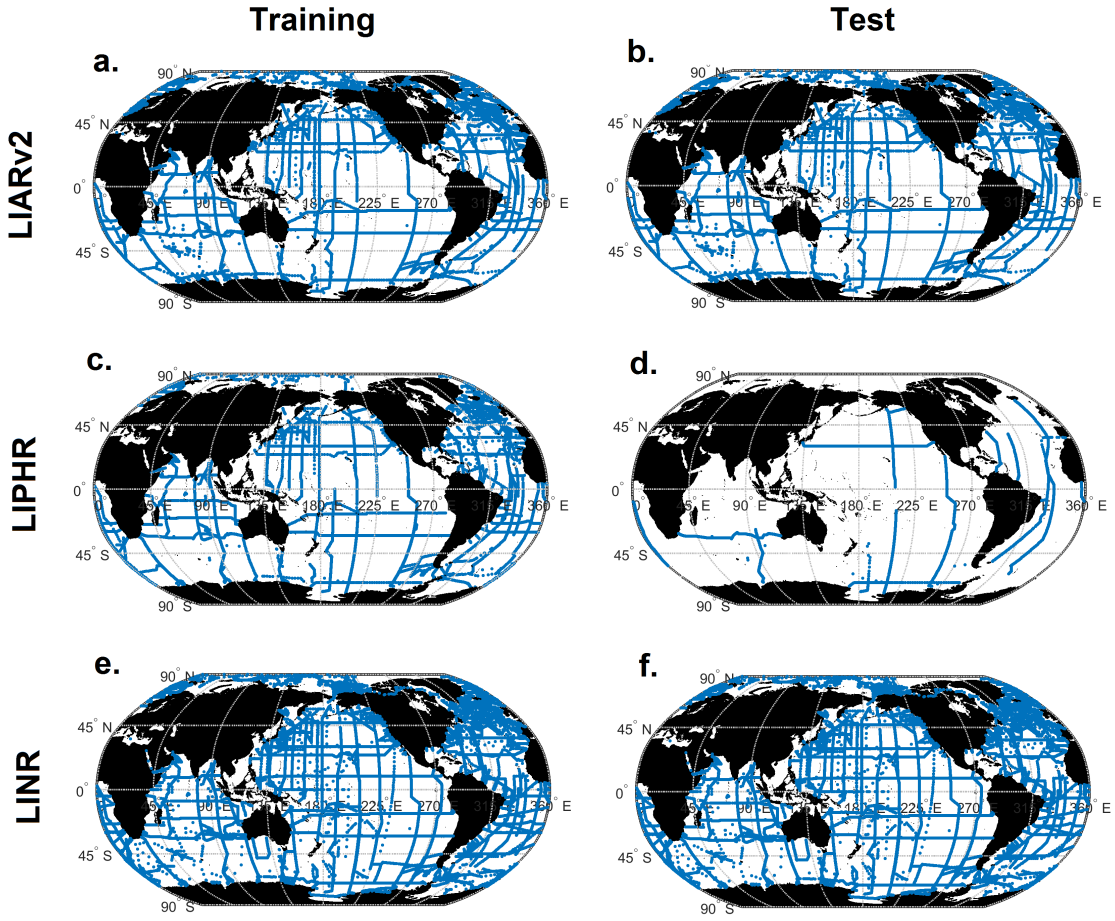

Figure 1. Maps of the data used for the training (left) and test (right) data sets for LIARv2 (top, a. and b.), LIPHR (middle, c. and d.), and LINR (bottom, e. and f.) regression coefficients.

There are several additional difficulties for constructing a consistent data product for training LIPHR that originate from changes in ocean pH and in pH measurement practices over time. Dealing with these inconsistencies requires understanding several adjustments that we and others (Olsen et al. 2016) have made to pH measurements and estimates. We list these adjustments here and explain them in this section and the next.

1. GLODAPv2 adjustments: These are recommended adjustments to cruise pH, A_T, and C_T measurements based on deep crossovers (Olsen et al. 2016). We do not use these adjustments for pH, though we do use them for A_T and C_T.
2. Impure-dye adjustments: These are adjustments to pH measurements that we make for pH values measured using impure dye (*i.e.* commercially available indicator dye that has not been specially purified). These adjustments are intended to bring these values in line with pH calculated from A_T and C_T. They are detailed below.
3. Calculation-to-purified-dye-pH adjustment: This is a single adjustment we apply to impure-dye measurements (after they have been adjusted by the impure-dye adjustment) and to calculated pH values. This adjustment is intended to bring these values in line

with pH measurements made with purified dyes. LIPHR includes optional code to apply the inverse of this adjustment to returned pH estimates if the user desires estimates that match what would be expected for pH calculations from A_T and C_T. This adjustment is also detailed below.

4. Ocean acidification adjustment: This is an optional adjustment applied to LIPHR pH estimates to reflect the impacts of ongoing ocean acidification on seawater pH (detailed in 2.3).

The primary additional difficulty for pH stems from the variety of ways pH is measured or calculated, as well as the evolution of accepted best practices for pH measurement over the decades for which GLODAPv2 contains data. GLODAPv2 contains a mixture of pH calculated from carbonate system measurements, pH measured using electrodes, and pH measured spectrophotometrically. Also, although the spectrophotometric pH method has been used since the early 1990s (e.g. Clayton and Byrne 1993), Yao et al. (2007) revealed that impurities in the indicator dye used can significantly bias spectrophotometric pH measurements, and Liu et al. (2011) subsequently published calibration equations that allow seawater pH measurements to be made using purified m-cresol purple dye. Others (Carter et al. 2013; Patsavas et al. 2015; Williams et al. 2017) have since shown that measurements with purified dyes appear to have an (unexplained) broadly-consistent-but-pH-dependent discrepancy from the pH calculated from combinations of A_T, C_T, and pCO_2 whether calculated at *in situ* or laboratory conditions (Figure 2c). This pH dependent discrepancy is not unique to a single pH sample handling approach, as it exists for both manual and automated pH measurements. It exists also for multiple carbonate constant sets (Carter et al. 2013). It exists for multiple characterizations of the properties of purified dyes: There is a small pH-dependent discrepancy between spectrophotometric pH obtained from various sets of purified dye coefficients (Liu et al., 2011; DeGrandpre et al., 2014), but the discrepancy (ranging from ~0.006 at a pH of 8.2 to ~0.002 at a pH of 7.4) is too small to account for the differences between calculated pH and pH measured with purified dyes. The pH-dependent pH discrepancy is less apparent for electrode pH measurements (Figure 2a) and impure dye measurements (Figure 2b) considered collectively across many cruises. However there are many strongly-differing discrepancy relationships visible when impure dye measurements are considered on a cruise-by-cruise basis (see Supplementary Materials figures), with some discrepancies increasing and some decreasing with pH.

A second complication arises in the GLODAPv2 data product QC process. This data product relies on deep crossovers to obtain measurement adjustments intended to bring measurements from various cruises in line with one another. However, the variety of pH-dependent pH discrepancies found in various cruises casts doubt on the comparability of deep-ocean pH measured on different cruises. Adjustments based on forcing an agreement at depth between pH distributions obtained with different approaches could therefore create, exacerbate, or inadequately capture a discrepancies at the surface.

Our approach to these challenges is to first divide the data into three subsets and then apply linear adjustments to the first two subsets to make them comparable to the third. The first subset is the earlier measurements made with impure dyes. The second subset is pH calculated from A_T and C_T. These two subsets collectively comprise the majority of the GLODAPv2 data product. The third subset is the subset of the GLODAPv2 data product where pH was measured with purified dyes. We augment the purified dye subset with 11 cruises conducted too recently to appear in the GLODAPv2 data product (Expocodes: 096U20160108, 096U20160426, 29HE20130320, 318M20130321, 320620140320, 320620151206, 33AT20120324, 33RO20150410, 33RO20150525, and 33RO20161119). We further add data from two recent cruises measured with impure dye to the impure-dye subset (33RO20130803, 33RO20131223). Data from one additional recent cruise using purified dyes along the I09N transect (33RR20160208) is withheld from the pH training data set entirely to provide a completely independent assessment (Section 3.5). Linear pH-dependent adjustments ($\Delta_{1\to2}$, adjustment 2) are applied separately to each cruise measured with impure dyes to make the pH measurements comparable to the "calculated pH" subset. The coefficients for these adjustments are determined with a robust linear regression of the pH discrepancy (measured minus calculated) against measured pH. Coefficients for these adjustments are supplied as Supplementary Materials. Next, a single pH dependent adjustment ($\Delta_{2\to3}$, adjustment 3, ~ +0.004 to -0.020, Fig. 2b) is applied to the combination of the second subset and the adjusted first subset to make them comparable to the third "purified-dye" subset. The $\Delta_{2\to3}$ adjustment is (Figure 2c):

$$\text{adjustment 3} \equiv \Delta_{2\to3} \equiv -0.3168 + 0.0404\text{pH} \quad (1)$$

After applying $\Delta_{2\to3}$, the combined training pH data set has a pH-dependent pH discrepancy with calculated pH (Figure 3). Adjustments to the impure data are designed to take the place of the recommended GLODAPv2 adjustments (adjustment 1), and—except when noted—pH data presented herein do not include the GLODAPv2 adjustments. Supporting the decision to omit the GLODAPv2 pH adjustments, the algorithms we produce have a ~3% smaller RMSE and 4% smaller average bias when reproducing the unadjusted data than adjusted-data-trained algorithms have when reproducing adjusted data.

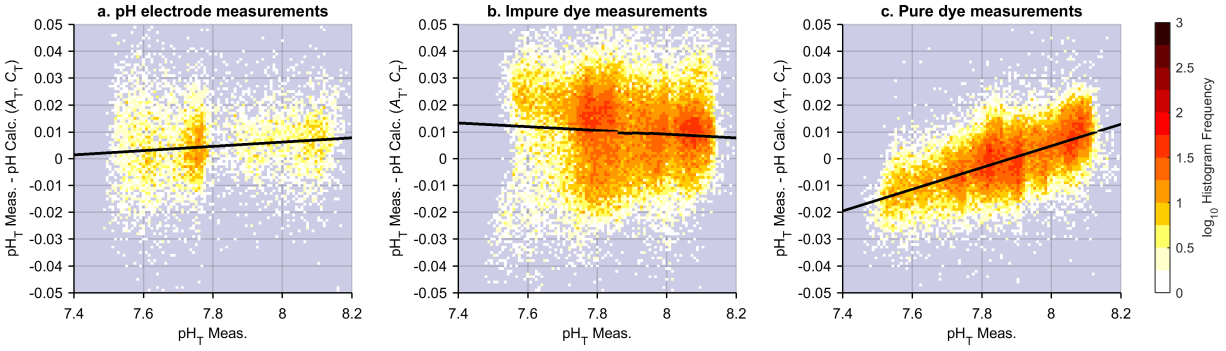

Figure 2. 2-dimensional histograms showing the number of *in situ* total scale pH measurements

falling within bins of discrepancy between measured and calculated pH (y-axis) and measured pH (x-axis) for (**a.**) the electrode-based measurements; for (**b.**) the impure-dye subset primarily measured prior to 2011, and (**c.**) our test data set, which is the purified dye measurement subset predominantly made since 2011.

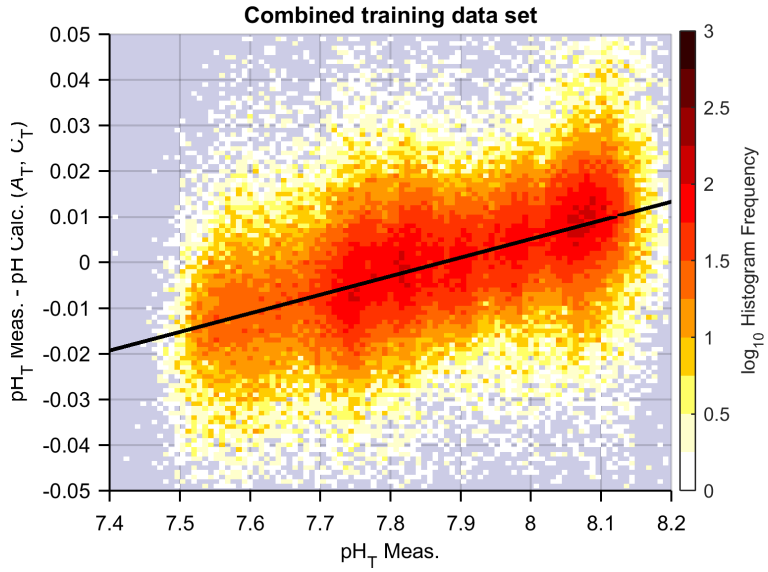

Figure 3. A 2-dimensional histogram showing the number of *in situ* total scale pH measurements falling within bins of discrepancy between measured and calculated pH (y-axis) and measured pH (x-axis) for the LIPHR training data set after adjustments 2 and 3 are applied.

Our use of the purified-dye adjustment (adjustment 3) reflects our need for a consistent training data product and not any confidence that purified dye measurements are necessarily more accurate representations of the "true" seawater pH than pH calculations. The apparent pH-dependent pH discrepancy remains an unresolved challenge to our carbonate system knowledge. Our strategy is to allow LIPHR users to decide whether pH estimates specific to purified dye measurements or pH calculations with Lueker et al. (2000)'s carbonate chemistry coefficients are more appropriate for their own applications. LIPHR therefore includes an optional counter-adjustment for adjustment 3 ($\Delta_{3\to2}$) derived from equation 1 to return pH estimates that are consistent with pH calculated from A_T and C_T. Broadly, we recommend the default "purified dye estimates" without this counter-adjustment when pH is the parameter of interest, and "calculation-pH estimates" with this adjustment when LIPHR estimates are being used as one of two constraints to estimate another carbonate system parameter. Whichever is used, the user should be aware of this mismatch in our understanding of carbonate system chemistry.

In total, the LIPHR training data set consists of 51,325 impure-dye measurements (adjusted with $\Delta_{1\to2}$ and $\Delta_{2\to3}$); 99,061 calculated pH values (adjusted with $\Delta_{2\to3}$); and 35,383 unadjusted purified dye measurements (185,769 total measurements). The test data set contains only the 35,383 purified dye measurements. These data sets exclude 416 electrode pH measurements and 14,983 impure dye measurements for which no calculated pH value was available. These totals also exclude measurements and calculations from cruises that either had GLODAPv2 pH adjustments estimated to be larger than ±0.015 pH units, that were calculated from cruises with (applied) total dissolved inorganic carbon (C_T) or total seawater titration alkalinity (A_T) GLODAPv2 adjustments greater than ±10 μmol kg^{-1}, or that were flagged as having poor quality pH measurements. When viable pH measurements and calculations were both available for a sample, only the pH measurements were included. We also omitted data from 7 cruises (expocodes: 49K619990523, 49HG19950414, 49HG19940413, 49HG19930807, 49HG19930413, 33RR19971202, 318M19940327) either because they came from series of cruises with large and variable GLODAPv2 adjustments or because the calculated and measured pH values did not agree with a ±0.03 or less RMS or ±0.015 average difference. A full list of cruises and how they were classified is provided in Supplementary Materials.

2.3 An ocean acidification adjustment for pH estimates

Johnson et al. (2017) find that recent profiling float sensor pH measurements are significantly lower than most nearby pH stations in the GLODAPv2 record, and that these disagreements are largest in the better-ventilated surface ocean. LIPHR includes an optional adjustment (on by default) to reflect these expected effects of OA on modern and future seawater pH (adjustment 4). For this adjustment, the rate of pH change (γ_{OA}) is approximated from the robust regression:

$$\mathrm{pH}_{\mathrm{TestDat}} - \mathrm{pH}_{\mathrm{LIPHR}} = \gamma_{OA}\left(D_{\mathrm{TestData}} - D_{\mathrm{TrainingData}}\right) \quad (2)$$

This is a regression between the reconstruction error ($\mathrm{pH}_{\mathrm{TestDat}} - \mathrm{pH}_{\mathrm{LIPHR}}$) as the dependent variable and the difference ($D_{\mathrm{TestData}} - D_{\mathrm{TrainingData}}$) between the mean decimal years of the training measurements used to estimate the regression coefficients ($D_{\mathrm{TrainingData}}$) and the decimal years of the test data (D_{TestData}) as the independent variable. The term "decimal years" is used to mean the year (C.E.) with a decimal added to represent the fraction of 365 days elapsed in that year (such that a measurement on the 200th day of 2020 would be represented by ~2020.55). This regression has been performed for the reconstructions of 10 subsets of the GLODAPv2 data product used separated by every 10th percentile of potential density (σ_θ) (Figure 4). If the OA adjustment is enabled in the LIPHR code, γ_{OA} is linearly interpolated to the σ_θ estimated for the query data location and the adjusted LIPHR estimate ($\mathrm{pH}^*_{\mathrm{LIPHR}}$) is supplied as:

$$\mathrm{pH}^*_{\mathrm{LIPHR}} = \mathrm{pH}_{\mathrm{LIPHR}} + \gamma_{OA}\left(D_{\mathrm{QueryData}} - D_{\mathrm{TrainingData}}\right) \quad (3)$$

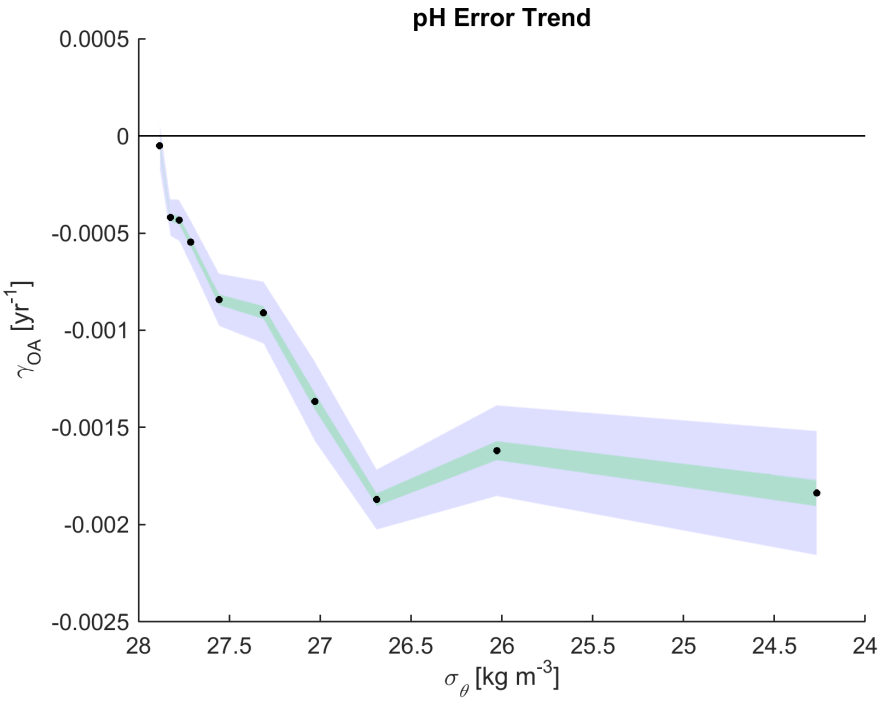

Figure 4. The average annual rate of ocean acidification (OA)-related impacts on LIPHR estimate errors (γ_{OA}) calculated for every 10th percentile of potential density (σ_θ) in the GLODAPv2 data product. If the optional OA adjustment is used (Equation 3), LIPHR uses user-provided dates with this relationship to adjust estimates it returns for the effects of ocean acidification. The green envelope indicates 95% confidence intervals of the fits. The blue envelope shows the larger confidence intervals obtained if one degree of freedom is assumed for each cruise rather than each measurement. Values in this figure are calculated using regression 7 (of the 16 regressions LIPHR can employ, see Table 2). Values for the other 15 regressions would be within ~±0.0005 yr^{-1} of these.

The OA pH change rates we find here are consistent with previous estimates (e.g. Feely et al. 2009). These simplistic OA adjustments may be poor estimates of the impacts of OA on seawater pH generally because they treat all water of a given density identically despite strong regional differences in the degree of water mass ventilation and C_{anth} storage. Nevertheless, we believe the optional adjustment is useful for LIPHR pH estimates made in the coming decades, and note that including the adjustment decreases mean estimate bias by 85% and RMSE by ~51%. Due to the progressive effects of OA, we contend this adjustment will be yet more important for modern estimates than for our test data set. Limited experimentation suggested additional cruises would be needed to adequately constrain regional differences in this adjustment. The LIPHR code therefore contains an option for users to input γ_{OA} estimates that

are specific to the OA rates found in their study regions, if desired. The assessment values we report in section 3 include the OA adjustment.

2.4 Update to uncertainty estimation

The LIRs generate uncertainty estimates for each property estimate returned. As with LIARv1, uncertainty estimates (E_{Est}) are quantified as:

$$E_{\mathrm{Est}} = \sqrt{{E_{\mathrm{Meas}}}^2 + {E_{\mathrm{MLR}}}^2 + \sum_{j=1}^{n} (U_j \alpha_j)^2} \quad (4)$$

E terms refer to the RMS uncertainties as assessed in section 3. E_{Meas} represents A_{T}, N, and pH measurement uncertainties in our data product, and is assumed to be a constant 2.8 µmol kg^{-1} A_{T}, 0.3 µmol kg^{-1} N, and 0.005 pH units, respectively (Olsen et al. 2016). U_j are the n input uncertainties for the predictor properties provided by the user, or default uncertainties if no U values are provided. The default uncertainties are now 0.005 for S, 0.005 °C for θ, 1% O_2, and 2% of N, P, and Si. The α_j terms are the n regression coefficients used in the estimate. E_{MLR} represents the component of the overall uncertainty inherent to regression based estimates. It is estimated for LIR outputs using estimates of E_{MLR} that are specific to each of the 16 equations and to 10 depth ranges (for N and pH) or 50 ranges of depth and S (for A_{T}). These ranges correspond to every 10$^{\mathrm{th}}$ percentile of depth and/or salinity in the training data product (with a single range spanning the 20$^{\mathrm{th}}$ through 80$^{\mathrm{th}}$ percentile of salinity). E_{MLR} is estimated for the various ranges using the assessment data with known E_{Est} and solving equation (4) for E_{MLR}. These range-specific E_{MLR} estimates are then interpolated by these properties to the depth and/or salinity inputs for the E_{Est} calculations. LINR and LIPHR errors also scale slightly with salinity, but not as strongly as LIAR errors do because of the smaller impact of freshwater cycling on N and pH than on A_{T}. All LIR uncertainties increase near the surface due to a larger impact of seasonality, episodic biogeochemical cycling, and gas exchange.

2.5 Minimum uncertainty estimates

One difficulty with LIRs is choosing between up to 16 possible estimates. We have added (optional, on by default) functionality to all LIR routines that automatically picks the estimate with the smallest estimated uncertainty from among all estimates it is possible to generate using the suite of input predictor data provided by the user. This feature is intended in part to address a limitation of the method, being that some LIR equations have too many terms (i.e. are over-fit) for some of the >2 million combinations of predicted variables, predictor variables, and grid locations. Over-fitting leads to larger-magnitude regression coefficients due to "Variance Inflation." Larger magnitude coefficients (α_j) propagate through equation 4 to return larger uncertainty estimates. Once the increase in E_{Est} from having more and larger-magnitude coefficients (i.e. from over-fitting) balances the typically lower E_{MLR} values for the equations

with more terms, this functionality automatically selects the less complex and less over-fit equation. This feature therefore selects an equation that minimizes overall error from over-fitting, input uncertainties, and method errors generally. This option modestly decreases estimate RMSE by 0 to 11% and, more importantly, makes the function easier to use without compromising estimate skill. The estimate improvement becomes more marked with (known) larger input uncertainties such as those that will be common with sensor measurements. For example, the A_T estimate RMSE improvement with this feature increased from 3% to 10% after simulated errors were applied to AOU (these were normally distributed offsets with a mean of 0 and a standard deviation of 5 μmol kg^{-1} O_2).

3 Assessment

Estimate bias and RMS errors are calculated in the same way as the error estimates provided by Carter et al. (2016), except using the subsets of the GLODAPv2 data product and additional cruises specified as "test data" sets in section 2.2. These values are presented as "bias (±RMSE)." The bias is the mean residual for the assessment and can be positive or negative. LIR bias estimates are small compared to RMSE at the global level, suggesting the LIR estimates are appropriately centered on the measured values. However, bias grows (in an absolute sense) as the number of measurements averaged decreases, so the bias estimates are presented alongside RMSE as potentially useful indicators of how correlated LIR errors are for various regions. Bias estimates are also useful when comparing assessments from various algorithms. In particular, lower biases for LIPHR than for other pH algorithms highlight the importance of the OA adjustment and the dye-impurity-related adjustments applied to the training data set. An important feature of the error estimation method used is that a separate set of regression coefficients is estimated for each data point in our test data sets, and is estimated without using any data from the cruise that produced that particular test pH value. Data from the same cruise is omitted to avoid under-estimating error by including numerous measurements in the training dataset found proximally in time and space to the test measurement.

3.1 LIARv2

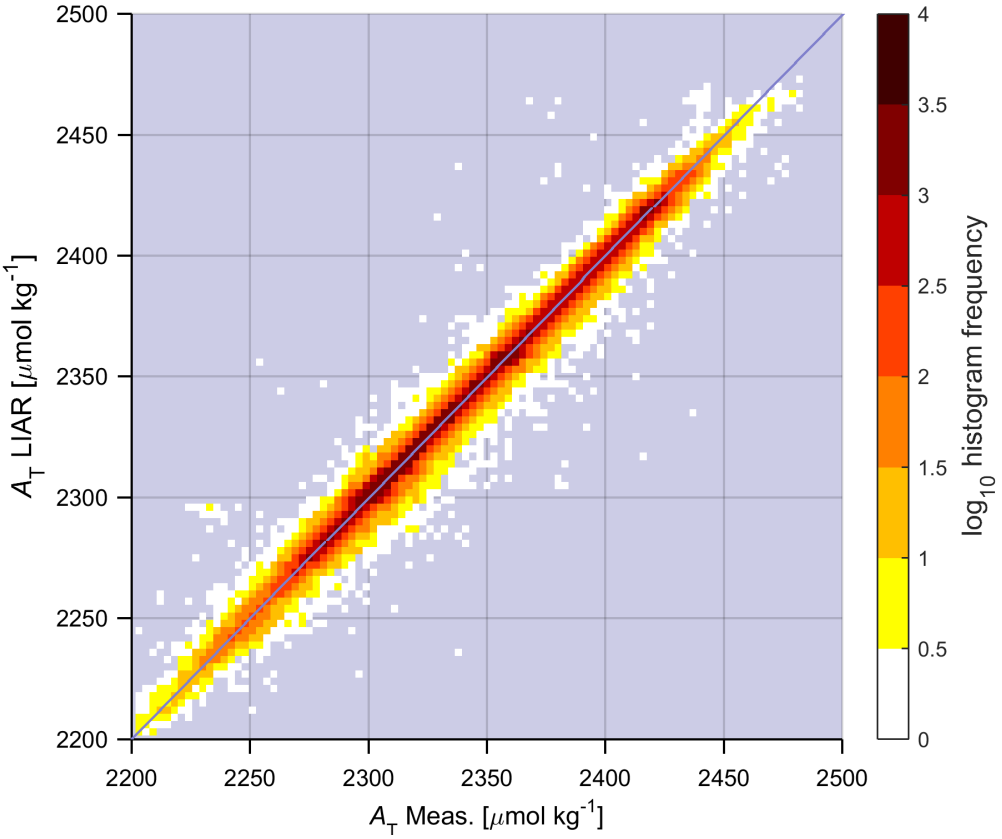

Figure 5. A 2-dimensional histogram of measured A_T (x-axis) against estimated A_T (y-axis). Darker colors along the thin blue 1:1 line indicate orders of magnitude more measurements fall on the line than in the light colored histogram bins off the line.

The updates to LIAR decreased the overall reconstruction errors (E_{LIARv2}) for all 16 regressions relative to E_{LIARv1} by 7% to 26% (average 18%) when both sets of errors are calculated using the newer test dataset. The largest improvements are for regressions with the fewest predictors. We attribute the majority of the improvements to the increased size, quality, and consistency of the subset of the GLODAPv2 data product we used relative to the merged data product we used for LIARv1 (Figure 5). LIARv1 compared favorably to regional A_T regressions in literature (many are compared in Carter et al., 2016) and Table 1 shows LIARv2 does somewhat better still. CANYON A_T estimates reproduce our entire test dataset with errors of -0.2 (±5.4) µmol kg^{-1} while LIARv2 (Regression 7) has errors of -0.1 (±5.1) µmol kg^{-1}. These errors are slightly smaller at -0.5 (±5.2) µmol kg^{-1} for CANYON and 0.2 (±4.4) µmol kg^{-1} for LIARv2 when limited to the open ocean test regions used by Sauzède et al. (2017).

Interestingly, regression 3 (S, θ, AOU, and Si) slightly outperforms regression 1 (S, θ, N, AOU, and Si) on average, and there is little difference between the error estimates for the various equations for A_T. This suggests that regression 1 and possibly others are over-fitting A_T in places

374 (this observation does not hold true if we include the test data in the training data). See section
375 2.5 for how the LIR minimum-uncertainty functionality automatically avoids using over-fit
376 relationships despite this.
377

Table 1. Error estimates expressed as "bias (±RMSE)" for the subset of our data product found within the open-ocean salinity range of 33 to 38. E_{MLR} is uncertainty inherent to the use of a MLR approach, E_{Input} is error arising from uncertainties in the input data (i.e. the summed term in equation 4), E_{LIARv2} is the overall estimate uncertainty for LIARv2. GLODAPv2 data product is used as test data for all estimates. Errors are expressed as standard errors in μmol A_T kg^{-1}.

Reg. #	Parameters used	E_{MLR}	E_{Input}	E_{LIARv2}
1	*S*, *θ*, *N,* AOU, *Si*	(±3.6)	(±0.8)	0.1 (±5.0)
2	*S*, *θ*, *N, Si*	(±3.7)	(±0.7)	0.1 (±5.0)
3	*S*, *θ*, AOU, *Si*	(±3.6)	(±0.7)	0.1 (±4.9)
4	*S*, *θ*, *Si*	(±3.7)	(±0.6)	0.1 (±5.0)
5	*S*, *θ*, *N,* AOU	(±3.8)	(±0.9)	0.0 (±5.1)
6	*S*, *θ*, *N*	(±4.0)	(±0.9)	0.1 (±5.3)
7	*S*, *θ*, AOU	(±3.8)	(±0.7)	-0.1 (±5.1)
8	*S*, *θ*	(±4.4)	(±0.5)	0.2 (±5.5)
9	*S*, *N,* AOU, *Si*	(±3.6)	(±0.8)	0.1 (±5.0)
10	*S*, *N, Si*	(±3.7)	(±0.7)	0.1 (±5.0)
11	*S*, AOU, *Si*	(±3.6)	(±0.6)	0.1 (±5.0)
12	*S*, *Si*	(±3.7)	(±0.6)	0.1 (±5.0)
13	*S*, *N,* AOU	(±4.6)	(±1.2)	-0.1 (±5.8)
14	*S*, *N*	(±4.4)	(±1.0)	-0.1 (±5.6)
15	*S*, AOU	(±4.6)	(±0.8)	-0.2 (±5.7)
16	*S*	(±5.1)	(±0.4)	0.1 (±6.1)

378
379
380 *3.2 LIPHR*
381 LIPHR pH estimates reconstruct the test pH data set well (Table 2, Figure 6). We separately
382 estimate error between 1000 and 2000 m as these estimates are more likely to be used to
383 compare with float data (Table 2).
384

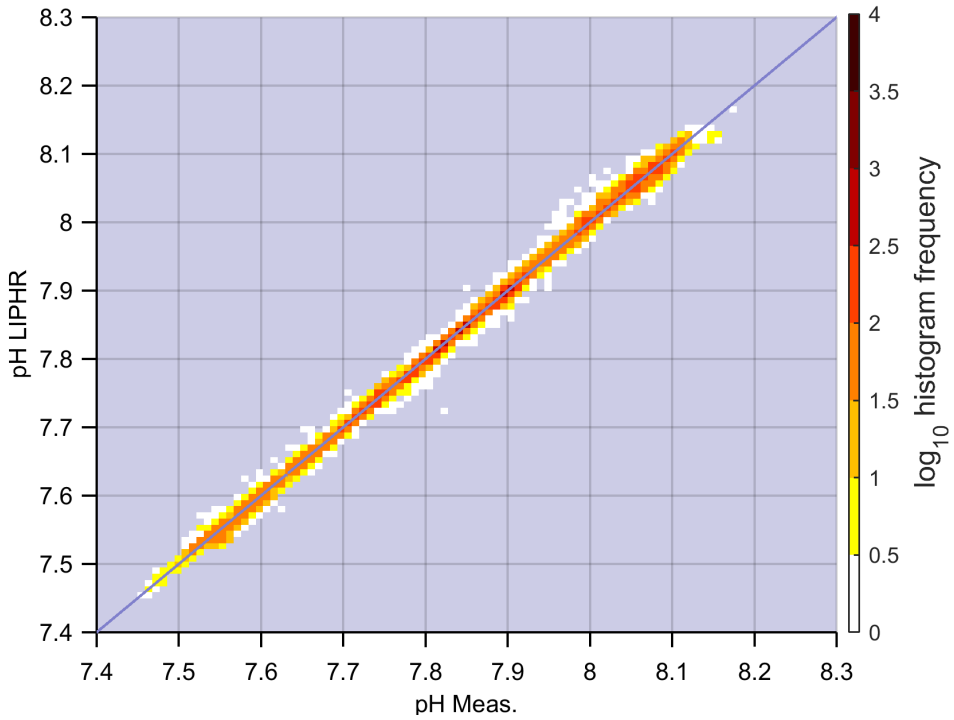

Figure 6. A 2-dimensional histogram of measured or calculated pH (x-axis) against OA-adjusted estimated pH (y-axis). Darker colors along the thin blue 1:1 line indicate orders of magnitude more measurements fall close to the line than in light colored histogram bins off the line.

LIPHR estimates compare well to the few published pH regression estimates. Williams et al. (2016) designed regression estimates for south of 45°S between 2006 and 2017 and between 0 and 2100 m depth. For the subset of our data product within these bounds and omitting their S04P and P16S training cruises, their published regressions have errors of -0.006 (±0.017) and -0.006 (±0.016), while similar LIPHR regressions (6 and 7 respectively) have errors of -0.001 (±0.010) and -0.001 (±0.011). Williams et al. (2016) also report a regression for estimates in the same region but trained specifically for estimates between 1000 and 2100 m depth, the depth range most useful for assessment of biogeochemical profiling float sensor performance. For the relevant subset of our test data product, their algorithm has errors of -0.001 (±0.005), while the LIPHR regression 7 has errors of 0.002 (±0.005). LIPHR (also regression 7) estimates have errors of 0.004 (±0.014) in the California Current Ecosystem specific window of 114°N to 124°W, 27°N to 36°N and 15 to 500 m depth after 1994 where the algorithm from Alin et al. (2012) uses temperature and O_2 measurements to generate estimates with errors of -0.008 (±0.015). CANYON pH estimates reproduce our entire test dataset with errors of 0.009 (±0.017) while LIPHR (Regression 7) has errors of 0.000 (±0.010). At mid depths (1000 m to 2000 m), these estimates are 0.013 (±0.017) for CANYON and 0.000 (±0.006) for LIPHR. The CANYON error estimates are the same at this precision when the GLODAPv2 adjustments are retained.

Table 2. LIPHR error estimates expressed as "bias (±RMSE)" for the subset of our data product found within the open-ocean salinity range of 33 to 38. E_{MLR} is the uncertainty inherent to the use of a MLR approach, E_{Input} is error arising from uncertainties in the input data (i.e. the summed term in Equation 4), and E_{LIPHR} is the overall estimate uncertainty. $E_{\mathrm{LIPHR2000m}}$ is the uncertainty estimate for pH measurements between 1000 and 2000 m, or the approximate depth range at which biogeochemical floats will require pH estimates for cross-comparison.

Reg. #	Parameters used	E_{MLR}	E_{Input}	E_{LIPHR}	$E_{\mathrm{LIPHR2000m}}$
1	*z, S*, *θ*, *N,* AOU, *Si*	(±0.0080)	(±0.004)	0.002 (±0.010)	0.001 (±0.008)
2	*z, S*, *θ*, *N, Si*	(±0.0110)	(±0.005)	0.002 (±0.013)	0.000 (±0.009)
3	*z, S*, *θ*, AOU, *Si*	(±0.0090)	(±0.003)	0.001 (±0.011)	0.001 (±0.007)
4	*z, S*, *θ*, *Si*	(±0.0190)	(±0.002)	0.001 (±0.020)	-0.003 (±0.014)
5	*z, S*, *θ*, *N,* AOU	(±0.0070)	(±0.004)	0.001 (±0.010)	0.001 (±0.006)
6	*z, S*, *θ*, *N*	(±0.0110)	(±0.004)	0.002 (±0.013)	0.000 (±0.007)
7	*z, S*, *θ*, AOU	(±0.0090)	(±0.003)	0.001 (±0.011)	0.001 (±0.006)
8	*z, S*, *θ*	(±0.0230)	(±0.001)	0.001 (±0.024)	-0.003 (±0.013)
9	*z, S, N,* AOU, *Si*	(±0.0090)	(±0.004)	0.001 (±0.011)	0.001 (±0.007)
10	*z, S, N, Si*	(±0.0120)	(±0.005)	0.002 (±0.014)	0.001 (±0.008)
11	*z, S*, AOU, *Si*	(±0.0100)	(±0.003)	0.001 (±0.011)	0.001 (±0.006)
12	*z, S*, *Si*	(±0.0200)	(±0.002)	0.001 (±0.021)	-0.003 (±0.014)
13	*z, S, N,* AOU	(±0.0090)	(±0.004)	0.001 (±0.011)	0.001 (±0.007)
14	*z, S, N*	(±0.0130)	(±0.004)	0.002 (±0.015)	0.000 (±0.008)
15	*z, S*, AOU	(±0.0100)	(±0.003)	0.001 (±0.011)	0.001 (±0.006)
16	*z, S*	(±0.0300)	(±0.001)	0.001 (±0.031)	-0.003 (±0.015)

3.3 *LINR*

LINR estimates also reproduce the test data product well (Table 3, Figure 7). Williams et al. (2016) provide an *N* estimation algorithm specific to the Pacific sector of the Southern Ocean south of 45°S between 1000 and 2100 m. This algorithm has errors of 0.42 (±0.65) μmol kg^{-1} for the portion of our data product in the target region for this regression. LINR (Regression 7) has an error of -0.11 (±0.45) μmol kg^{-1} for this same subset. CANYON nitrate estimates reproduce our entire test dataset with errors of -0.01 (±0.89) μmol kg^{-1} while LINR (Regression 7) has errors of -0.02 (±0.86) μmol kg^{-1}. These errors are slightly smaller at 0.03 (±0.66) μmol kg^{-1} for CANYON and -0.02 (±0.65) μmol kg^{-1} for LINR when limited to the open ocean test regions used by Sauzède et al. (2017).

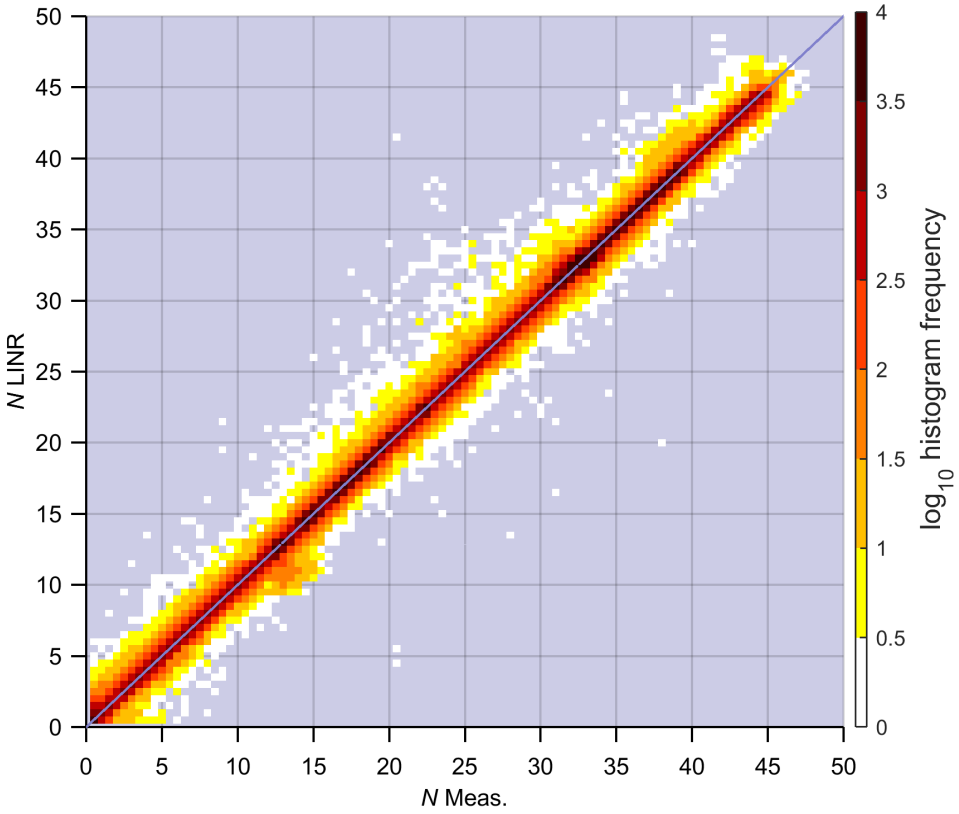

Figure 7. A 2 dimensional histogram of measured *N* (x-axis) against estimated *N* (y-axis). Darker colors along the thin blue 1:1 line indicate orders of magnitude more measurements fall on the line than in the light colored histogram bins off the line.

416

417

Table 3. LINR error estimates expressed as "bias (±RMSE)" for the subset of our data product found within the open-ocean salinity range of 33 to 38. E_{MLR} is the uncertainty inherent to the use of a MLR approach, E_{Input} is error arising from uncertainties in the input data (i.e. the summed term in Equation 4), and E_{LINR} is the overall estimate uncertainty. $E_{\mathrm{LINR2000m}}$ is the uncertainty estimate for *N* measurements between 1000 and 2000 m, or the approximate depth range at which biogeochemical floats will require *N* estimates for cross-comparison.

Reg. #	Parameters used	E_{MLR}	E_{Input}	E_{LINR}	$E_{\mathrm{LINR2000m}}$
1	*S*, *θ*, *P*, AOU, *Si*	(±0.56)	(±0.12)	-0.01 (±0.64)	0.00 (±0.45)
2	*S*, *θ*, *P*, *Si*	(±0.58)	(±0.14)	0.00 (±0.67)	0.02 (±0.47)
3	*S*, *θ*, AOU, *Si*	(±0.81)	(±0.10)	-0.01 (±0.87)	0.00 (±0.84)
4	*S*, *θ*, *Si*	(±1.00)	(±0.09)	0.03 (±1.05)	0.03 (±0.89)

5	*S*, *θ*, *P,* AOU	(±0.56)	(±0.13)	-0.02 (±0.65)	-0.00 (±0.44)
6	*S*, *θ*, *P*	(±0.60)	(±0.16)	0.00 (±0.69)	0.01 (±0.48)
7	*S*, *θ*, AOU	(±0.80)	(±0.11)	-0.02 (±0.86)	0.00 (±0.47)
8	*S*, *θ*	(±1.23)	(±0.07)	0.05 (±1.27)	0.04 (±0.58)
9	*S*, *P,* AOU, *Si*	(±0.58)	(±0.13)	-0.01 (±0.67)	0.00 (±0.44)
10	*S*, *P, Si*	(±0.61)	(±0.15)	0.00 (±0.69)	0.02 (±0.47)
11	*S*, AOU, *Si*	(±0.87)	(±0.10)	-0.01 (±0.92)	-0.00 (±0.83)
12	*S*, *Si*	(±1.06)	(±0.10)	0.05 (±1.11)	0.06 (±0.81)
13	*S*, *P,* AOU	(±0.62)	(±0.14)	-0.03 (±0.70)	-0.00 (±0.44)
14	*S*, *P*	(±0.65)	(±0.17)	-0.01 (±0.74)	0.01 (±0.49)
15	*S*, AOU	(±0.96)	(±0.11)	-0.03 (±1.01)	-0.00 (±0.46)
16	*S*	(±1.68)	(±0.07)	0.06 (±1.71)	0.07 (±0.62)

3.4 Uncertainty estimation skill

With the changes to the error estimation strategy noted in Section 2.4, the overall standard error estimates provided by the software are now greater than or equal to the test data set reconstruction error for 76% of the data product for LIARv2, for 75% for LIPHR, and for 80% for LINR. For perfectly-estimated normally-distributed RMS uncertainties, this number would be 68%. This was true for 87% of the data product with LIARv1.

3.5 Example section

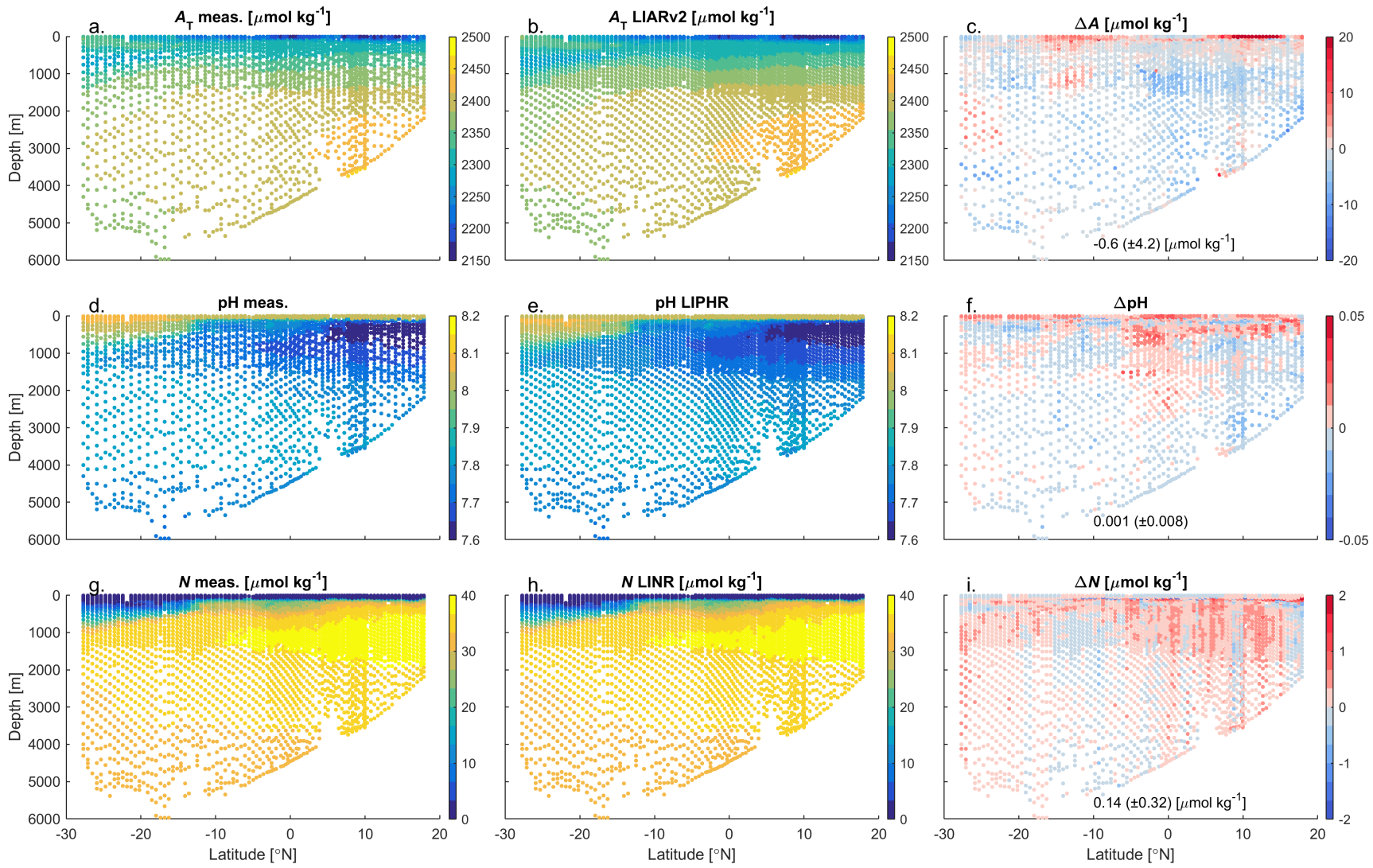
a.
A_T meas. [μmol kg^{-1}]
b.
A_T LIARv2 [μmol kg^{-1}]
c.
ΔA [μmol kg^{-1}]
-0.6 (±4.2) [μmol kg^{-1}]
d.
pH meas.
e.
pH LIPHR
f.
ΔpH
0.001 (±0.008)
g.
N meas. [μmol kg^{-1}]
h.
N LINR [μmol kg^{-1}]
i.
ΔN [μmol kg^{-1}]
0.14 (±0.32) [μmol kg^{-1}]
Depth [m]
Latitude [°N]

Figure 8. Measured (a., d., and g.) and estimated (b., e., and h.) A_T (a., b., and c.), pH (d., e. and f.), and N (g., h., and i.)—and differences between the two (c., f. and i.)—along the I09N section in the Indian Ocean.

Example LIAR, LIPHR, and LINR estimates are derived from hydrographic measurements from the 2016 occupations of the I09 section in the Indian Ocean by the Global Ocean Ship Based Hydrographic Investigations Program (GO-SHIP) program (Figure 8). These estimates provide an independent validation when compared to the measurements made along the cruise because the data from these cruises were not included in either the test or training datasets for the LIRs. The LIRs do an excellent job of reproducing the measurements with errors of -0.6 (±4.2) μmol kg^{-1} for A_T, 0.001 (±0.008) for pH, and 0.14 (±0.32) μmol kg^{-1} N. LIPHR errors increase to -0.014 (±0.017) when the OA adjustment is omitted.

Future Directions

Climatological distributions of carbonate parameters from LIAR A_T and LIPHR pH—or calculated from this pair of properties—may be of interest and would be simply calculated for the measurement-dense World Ocean Atlas climatology (Locarnini et al., 2013; Zweng et al., 2013; Baranova, 2015) or similar products. Such a regression-based climatology—like the A_T climatologies created by Lee et al. (2006) and used by Takahashi et al. (2014)—would be one step further removed from the measurements than gridded climatologies like those provided by Lauvset et al. (2016) and Key et al. (2004). However, it would have the advantage that it could be based on property measurements (such as O_2, S, and temperature) that are more numerous, more broadly spatially and temporally distributed, and less seasonally biased than the carbonate measurements.

With LIAR and LIPHR, it is now possible to estimate two parameters for the carbonate system, thus—in principle—providing a complete carbonate system description. While measurements would be preferable for most applications, this pair of algorithms allows additional context to be added to historical data products.

As Velo et al. (2013) pointed out, regressions can be potentially powerful tools for data quality control. An algorithm that uses many measured properties to estimate many other measured properties and then assesses the various residuals may provide a fast method for identifying apparent outliers and interesting anomalies in property measurement sets. Such automated measures designed to assist human-QC efforts may be of increased importance as growing sensor networks increase the quantity of data being produced relative to the amount of human-effort available for data QC.

The OA rate estimation strategy used (Equation 2) provides a means to incorporate a large number of measurements that are disparate in space and time into unified global trend estimates.

This framework could perhaps be applied to examine the low-signal-to-noise scientific questions of whether long term trends are occurring in A_T (c.f. Carter et al. 2016a), N, or O_2 relative to other measured parameters.

Acknowledgements

This improvement and extension of the LIAR software has been made possible by funding from the Climate Program Office at the National Oceanic and Atmospheric Administration with the US Global Carbon Data Management and Synthesis Project (N8R3CEA-PDM). This is JISAO Publication Contribution Number 2017-02-01 and PMEL contribution number 4661. We are grateful to Joshua Plant, Tanya Maurer, and Kenneth Johnson for beta-testing and helping us improve the code, and to Henry Bittig for sharing the CANYON code with us. We are grateful to Marta Álvarez for help classifying cruises by measurement practices. We are grateful to Kelly Kearney for advice regarding function input parsing and accommodating new code options.

References

Alin, S. R., R. A. Feely, A. G. Dickson, J. M. Hernández-Ayón, L. W. Juranek, M. D. Ohman, and R. Goericke, 2012: Robust empirical relationships for estimating the carbonate system in the southern California Current System and application to CalCOFI hydrographic cruise data (2005-2011). *J. Geophys. Res. Ocean.*, **117**, n/a-n/a, doi:10.1029/2011JC007511. http://doi.wiley.com/10.1029/2011JC007511 (Accessed February 16, 2017).

Baranova, O., 2015: World ocean atlas 2005. http://www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html (Accessed June 14, 2016).

Bostock, H. C., S. E. M. Fletcher, and M. J. M. Williams, 2013: Estimating carbonate parameters from hydrographic data for the intermediate and deep waters of the Southern Hemisphere oceans. *Biogeosciences*, **10**, 6199–6213, doi:10.5194/bg-10-6199-2013. www.biogeosciences.net/10/6199/2013/ (Accessed April 14, 2017).

Bushinsky, S. M., S. R. Emerson, S. C. Riser, and D. D. Swift, 2016: Accurate oxygen measurements on modified argo floats using in situ air calibrations. *Limnol. Oceanogr. Methods*, **14**, doi:10.1002/lom3.10107.

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

Carter, B. R., T. L. Frölicher, J. P. Dunne, K. B. Rodgers, R. D. Slater, and J. L. Sarmiento, 2016a: When can ocean acidification impacts be detected from decadal alkalinity measurements? *Global Biogeochem. Cycles*, **30**, 595–612, doi:10.1002/2015GB005308. http://doi.wiley.com/10.1002/2015GB005308 (Accessed October 10, 2016).

——, N. L. Williams, A. R. Gray, and R. A. Feely, 2016b: Locally interpolated alkalinity regression for global alkalinity estimation. *Limnol. Oceanogr. Methods*, **14**, 268–277, doi:10.1002/lom3.10087. http://doi.wiley.com/10.1002/lom3.10087 (Accessed June 13, 2016).

DeGrandpre, M. D., R. S. Spaulding, J. O. Newton, E. J. Jaqueth, S. E. Hamblock, A. A. Umansky, and K. E. Harris, 2014: Considerations for the measurement of spectrophotometric pH for ocean acidification and other studies. *Limnol. Oceanogr. Methods*, **12**, 830–839, doi:10.4319/lom.2014.12.830. http://doi.wiley.com/10.4319/lom.2014.12.830 (Accessed June 2, 2017).

Dickson, A. G., 1990: Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K. *Deep Sea Res. Part A. Oceanogr. Res. Pap.*, **37**, 755–766, doi:10.1016/0198-0149(90)90004-F. http://linkinghub.elsevier.com/retrieve/pii/019801499090004F (Accessed March 24, 2017).

Feely, R., S. Doney, and S. Cooley, 2009: Ocean Acidification: Present Conditions and Future Changes in a High-CO2 World. *Oceanography*, **22**, 36–47, doi:10.5670/oceanog.2009.95. http://tos.org/oceanography/article/ocean-acidification-present-conditions-and-future-changes-in-a-high-co2-wor (Accessed July 6, 2016).

van Hueven, S., D. Pierrot, J. W. B. Rae, E. Lewis, and D. W. R. Wallace, 2011: *MATLAB program developed for CO_2 system calculations, CO2sys.* Oak Ridge, TN,.

Johnson, K., and H. Claustre, 2016: Bringing biogeochemistry into the Argo age. *Eos. Trans. Am. Geophys. Union*, **97**, doi:10.1029/2016EO062427.

Johnson, K. S., H. W. Jannasch, L. J. Coletti, V. A. Elrod, T. R. Martz, Y. Takeshita, R. J. Carlson, and J. G. Connery, 2016: Deep-Sea DuraFET: A Pressure Tolerant pH Sensor Designed for Global Sensor Networks. *Anal. Chem.*, **88**, doi:10.1021/acs.analchem.5b04653.

——, and Coauthors, 2017: Biogeochemical sensor performance in the SOCCOM profiling float array. *J. Geophys. Res. Ocean.*, **122**, 6416–6436, doi:10.1002/2017JC012838. http://doi.wiley.com/10.1002/2017JC012838 (Accessed November 1, 2017).

Juranek, L. W., R. A. Feely, D. Gilbert, H. Freeland, and L. A. Miller, 2011: Real-time estimation of pH and aragonite saturation state from Argo profiling floats: Prospects for an autonomous carbon observing strategy. *Geophys. Res. Lett.*, **38**, n/a-n/a, doi:10.1029/2011GL048580. http://doi.wiley.com/10.1029/2011GL048580 (Accessed March 31, 2017).

Key, R. M., and Coauthors, 2004: A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochem. Cycles*, **18**, 1–23, doi:10.1029/2004GB002247. http://doi.wiley.com/10.1029/2004GB002247 (Accessed August 23, 2015).

Lauvset, S. K., and Coauthors, 2016: A new global interior ocean mapped climatology: the 1° × 1° GLODAP version 2. *Earth Syst. Sci. Data*, **8**, 325–340, doi:10.5194/ESSD-8-325-2016.

Lee, K., and Coauthors, 2006: Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. *Geophys. Res. Lett.*, **33**, L19605, doi:10.1029/2006GL027207. http://doi.wiley.com/10.1029/2006GL027207 (Accessed February 15, 2017).

——, T.-W. Kim, R. H. Byrne, F. J. Millero, R. A. Feely, and Y.-M. Liu, 2010: The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans. *Geochim. Cosmochim. Acta*, **74**, 1801–1811, doi:10.1016/j.gca.2009.12.027. http://linkinghub.elsevier.com/retrieve/pii/S0016703709007789 (Accessed June 21, 2017).

Liu, X., M. C. Patsavas, and R. H. Byrne, 2011: Purification and Characterization of meta-Cresol Purple for Spectrophotometric Seawater pH Measurements. *Environ. Sci. Technol.*, **45**, 4862–4868, doi:10.1021/es200665d. http://pubs.acs.org/doi/abs/10.1021/es200665d (Accessed June 2, 2017).

Locarnini, R. A., and Coauthors, 2013: World Ocean Atlas 2013. Vol. 1: Temperature. *S. Levitus, Ed.; A. Mishonov, Tech. Ed.; NOAA Atlas NESDIS*, **73**, 40, doi:10.1182/blood-2011-06-357442.

http://www.ncbi.nlm.nih.gov/pubmed/22058116%5Cnhttp://www.bloodjournal.org/cgi/doi/10.1182/blood-2011-06-357442.
Lueker, T. J., A. G. Dickson, and C. D. Keeling, 2000: Ocean pCO2 calculated from dissolved inorganic carbon, alkalinity, and equations for K1 and K2: validation based on laboratory measurements of CO2 in gas and seawater at equilibrium. *Mar. Chem.*, **70**, 105–119, doi:10.1016/S0304-4203(00)00022-0. http://linkinghub.elsevier.com/retrieve/pii/S0304420300000220 (Accessed March 24, 2017).
McNeil, B. I., and T. P. Sasse, 2016: Future ocean hypercapnia driven by anthropogenic amplification of the natural CO2 cycle. *Nature*, **529**, doi:10.1038/nature16156.
Millero, F. J., K. Lee, and M. Roche, 1998: Distribution of alkalinity in the surface waters of the major oceans. *Mar. Chem.*, **60**, 111–130, doi:10.1016/S0304-4203(97)00084-4. http://linkinghub.elsevier.com/retrieve/pii/S0304420397000844 (Accessed April 14, 2017).
Olsen, A., and Coauthors, 2016: The Global Ocean Data Analysis Project version 2 (GLODAPv2) – an internally consistent data product for the world ocean. *Earth Syst. Sci. Data*, **8**, 297–323, doi:10.5194/essd-8-297-2016. http://www.earth-syst-sci-data.net/8/297/2016/ (Accessed February 6, 2017).
Patsavas, M. C., R. H. Byrne, R. Wanninkhof, R. A. Feely, and W.-J. Cai, 2015: Internal consistency of marine carbonate system measurements and assessments of aragonite saturation state: Insights from two U.S. coastal cruises. *Mar. Chem.*, **176**, 9–20, doi:10.1016/j.marchem.2015.06.022. http://www.sciencedirect.com/science/article/pii/S0304420315300074 (Accessed June 2, 2017).
Perez, F. F., and F. Fraga, 1987: Association constant of fluoride and hydrogen ions in seawater. *Mar. Chem.*, **21**, 161–168, doi:10.1016/0304-4203(87)90036-3. http://linkinghub.elsevier.com/retrieve/pii/0304420387900363 (Accessed June 21, 2017).
Plant, J. N., K. S. Johnson, C. M. Sakamoto, H. W. Jannasch, L. J. Coletti, S. C. Riser, and D. D. Swift, 2016: Net community production at Ocean Station Papa observed with nitrate and oxygen sensors on profiling floats. *Global Biogeochem. Cycles*, **30**, 859–879, doi:10.1002/2015GB005349. http://doi.wiley.com/10.1002/2015GB005349 (Accessed March 23, 2017).
Sasse, T. P., B. I. McNeil, and G. Abramowitz, 2013: A new constraint on global air-sea CO_2 fluxes using bottle carbon data. *Geophys. Res. Lett.*, **40**, 1594–1599, doi:10.1002/grl.50342. http://doi.wiley.com/10.1002/grl.50342 (Accessed April 14, 2017).
Sauzède, R., H. C. Bittig, H. Claustre, O. Pasqueron de Fommervault, J.-P. Gattuso, L. Legendre, and K. S. Johnson, 2017: Estimates of Water-Column Nutrient Concentrations and Carbonate System Parameters in the Global Ocean: A Novel Approach Based on Neural Networks. *Front. Mar. Sci.*, **4**, 128, doi:10.3389/fmars.2017.00128. http://journal.frontiersin.org/article/10.3389/fmars.2017.00128/full (Accessed June 16, 2017).
Takahashi, T., S. C. Sutherland, D. W. Chipman, J. G. Goddard, and C. Ho, 2014: Climatological distributions of pH, pCO2, total CO2, alkalinity, and CaCO3 saturation in the global surface ocean, and temporal changes at selected locations. *Mar. Chem.*, **164**, 95–125.
Takeshita, Y., T. R. Martz, K. S. Johnson, J. N. Plant, D. Gilbert, S. C. Riser, C. Neill, and B. Tilbrook, 2013: A climatology-based quality control procedure for profiling float oxygen

601 data. *J. Geophys. Res. Ocean.*, **118**, 5640–5650, doi:10.1002/jgrc.20399.
602 http://doi.wiley.com/10.1002/jgrc.20399 (Accessed March 23, 2017).
603 Velo, A., F. F. Pérez, T. Tanhua, M. Gilcoto, A. F. Ríos, and R. M. Key, 2013: Total alkalinity
604 estimation using MLR and neural network techniques. *J. Mar. Syst.*, **111–112**, 11–18,
605 doi:10.1016/j.jmarsys.2012.09.002.
606 http://www.sciencedirect.com/science/article/pii/S0924796312001728 (Accessed November
607 19, 2015).
608 Wanninkhof, R., and Coauthors, 2016: *An evaluation of pH and NO_3 sensor data from SOCCOM*
609 *floats and their utilization to develop ocean inorganic carbon products*. 1-31 pp.
610 https://soccom.princeton.edu/sites/default/files/files/CWG_white_paper_March_13_2016.p
611 df.
612 Williams, N. L., and Coauthors, 2016: Empirical algorithms to estimate water column pH in the
613 Southern Ocean. *Geophys. Res. Lett.*, **43**, 3415–3422, doi:10.1002/2016GL068539.
614 http://doi.wiley.com/10.1002/2016GL068539 (Accessed February 16, 2017).
615 ——, and Coauthors, 2017: Calculating surface ocean pCO_2 from biogeochemical Argo floats
616 equipped with pH: An uncertainty analysis. *Global Biogeochem. Cycles*, **31**, 591–604,
617 doi:10.1002/2016GB005541. http://doi.wiley.com/10.1002/2016GB005541 (Accessed June
618 2, 2017).
619 Yao, W., X. Liu, and R. H. Byrne, 2007: Impurities in indicators used for spectrophotometric
620 seawater pH measurements: Assessment and remedies. *Mar. Chem.*, **107**, 167–172,
621 doi:10.1016/j.marchem.2007.06.012.
622 http://www.sciencedirect.com/science/article/pii/S0304420307001430 (Accessed June 2,
623 2017).
624 Zweng, M. M., and Coauthors, 2013: *World Ocean Atlas 2013, Volume 2: Salinity*. 227-237 pp.
625