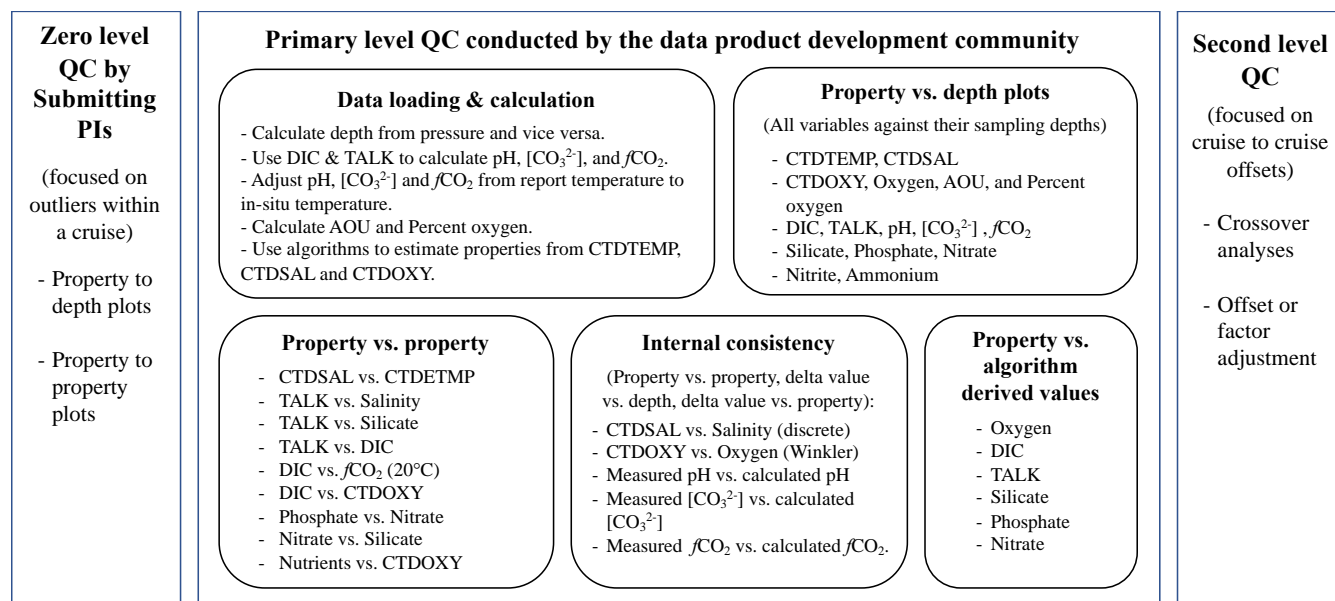


## 5 Technical Approach and Methodology

Cruise data set quality control often involves two steps: primary QC and secondary QC (Tanhua et al., 2010). These steps should follow initial, sometimes called “0-level” QC which is performed for individual measurements based on instrument readings and observations collected during the analyses. Primary QC is the process of identifying outliers and obvious errors within an individual cruise data set using measurement metadata or approaches like property-to-property plots (Figure 2). It should largely be done by the investigators responsible for the measurements. In addition, it is critical to provide additional uniform primary QC to all cruises within a data product using common tools and common thresholds to help identify any issues that have been missed by the data producers. These issues are communicated back to the investigators so that the issues could be reviewed and, if necessary, addressed. This additional layer of primary QC is often performed by the data product synthesis community. Secondary QC is a process in which data from one cruise are objectively compared against data from another cruise or a previously synthesized dataset in order to quantify systematic differences in the reported values. The secondary QC process often entails cross-over analysis (Lauvset and Tanhua, 2015), and increasingly regional Multiple Linear Regression (MLR) and inversions (Olsen et al., 2019; 2020).

Due to the scarcity of cross-over stations at depths where parameters were not likely to be influenced by temporal variations (sampling depth >1500 m, Olsen et al., 2020) on coastal cruises, secondary QC was not conducted for this version of the CODAP-NA and no cruise-wide offsets or multiplicative adjustments were applied. Instead, the QC relied on (a) stringent criteria for the selection of data sources, and (b) an enhanced primary QC procedure with rigorous consistency checks. This version of the CODAP-NA only accepted data from laboratories with direct involvement in the CODAP effort and with a track record of producing high-quality data and following best practices, making secondary quality control less essential. It is likely that there are other very high-quality coastal cruise data sets that are not yet included in this version of CODAP-NA.



**Figure 2. A diagram showing major steps of the quality control (QC) process. Note uncertainty is separated into outliers (scatter) and systematic offset (all data from the cruise has a bias).  $[\text{CO}_3^{2-}]$  is carbonate ion concentration,  $f\text{CO}_2$  is fugacity of carbon dioxide. Refer to Table 1 for the rest of the abbreviations.**

We worked directly with the data providers who knew their data best to conduct these primary QC procedures in order to leverage all of the resources related to a measurement: details related to the methods, instrumentation, reference standards, access to the raw data, and the analysts' recollection of the measurements. As part of the QC process, comparisons were made between many combinations of measured values. For a subset of properties, inter-consistency calculations and algorithm estimates based on other measurements allowed additional checks. Below are the 5 major steps of the QC procedures used for CODAP-NA (Figure 2). A new suite of QC tools is under development to allow these many comparisons and calculations to be performed quickly and efficiently, and these tools will be made available to the public soon with a separate paper dedicated to their rationales, development details, and instructions (Jiang et al., in prep.). A prototype version was used for CODAP-NA, though many software packages would, in principle, allow the comparisons and plots we use.

**Step One** was to ensure all of the cruise data files were ingested into NCEI's archives and documented with a rich metadata record (Jiang et al., 2015b). Maintaining a cruise data table allowing future users of the data product to access the original data files is an important component of any synthesis effort. For this study, a table with key metadata is available through this link: <https://www.ncei.noaa.gov/access/ocean-acidification-data-stewardship-oads/synthesis/NAcruises.html>. The following fields are listed in the table: A sequential number of the individual cruise data set (NO), expedition code (EXPOCODE), flags indicating the quality of the cruise (Cruise\_flag, see Table 3), cruise identifier (Cruise\_ID), Start\_date, End\_date, measured parameters, and links to NCEI's archive).

**Table 3.** Cruise flags used for this product.

Flag value	Meaning
A	These were dedicated OA cruises that were executed following Best Practices for global ocean work as outlined in Hood et al. (2011) and other documents as can be found on GO-SHIP site*. Colloquially these are referred to as GO-SHIP quality. Traceable standards and certified reference materials were used, and deep stations (> 2500 m) were sampled to allow using near-constant deep-water concentrations as anchor points. A third inorganic carbon system parameter, such as pH or carbonate ion concentration were often measured, allowing consistency checks.
B	These are dedicated OA cruises that had onboard inorganic carbon measurements performed according to Best Practices (Dickson et al. 2007), and many other parameters to highest accuracy through use of standards and certified reference materials. However, the cruises did not necessarily have all other parameters analyzed to highest standards, such as freezing nutrients for shoreside analyses; not taking oxygen and nutrients samples on most Niskins; not normalizing CTD oxygen trace to Winkler oxygen values, insufficient metadata etc. There often are insufficient deep stations to compare data with open ocean data.
C	These were opportunistic cruises where OA parameters were measured in the water column. They include standard hydrographic, carbon, and OA parameters; T, S, O <sub>2</sub> , nutrients, TALK, DIC, pH. Many parameters, including carbon and OA parameters were measured shoreside; CTD oxygen data were not adjusted to Winkler oxygen values. Generally, no dedicated OA personnel were onboard.
D	Underway samples only. These cruises had no CTD casts, and only had samples taken from the seawater supply line, with often a limited amount of other hydrographic parameters. T and S were obtained from thermosalinographs with limited or no salinity check samples.

(\*<https://www.go-ship.org/HydroMan.html>)

**Step Two** was to load the measurement values from the original cruise data files into MATLAB and conduct necessary calculations (Figure 2). All missing values were replaced with “-999” during this process. Variables without a QC flag from the original cruise data file were assigned an initial flag of 2 (good values, Table 4). Variables that were clearly out of range (e.g., a DIC value of < 0) were automatically assigned with a QC flag of “4” (bad values). The QC flags for all “-999” values or missing values were replaced with “9” (missing values).

Some surface samples from a few coastal cruises were collected from flow-through systems onboard research vessels, instead of Niskin bottles on sampling rosettes. In such cases, the temperature and salinity values were stored under the CTDTEMP and CTDSAL columns, respectively, although they were not measured from sensors mounted on a CTD rosette. Similarly, their sampling depth values were extracted from the metadata as the depth of the water inlet and stored under CTDPRES (Table 1). When water inlet depth information was not available, its sampling pressure was set to be 5 dbar. There is a column named “Observation\_type” in the CODAP data product file to indicate whether a sample is from a “Flow-through” system or a “Niskin” bottle.

We calculated or assigned the below parameters:

- (a) Sample\_ID if not already included (Equation 1)
- (b) depth from pressure and vice versa;
- (c) recommended\_Salinity\_PSS78 (Table 1);
- (d) conservative temperature, absolute salinity, sigma-theta;
- (e) recommended\_Oxygen
- (f) apparent oxygen utilization (AOU);
- (g) recommended\_Nitrate\_and\_Nitrite;

- (h) calculated pH, carbonate ion, and  $f\text{CO}_2$  at *in-situ* conditions using CO2SYS from DIC and TALK, along with temperature, salinity, pressure, and nutrients; and
- (i) *in-situ* pH, carbonate ion, and  $f\text{CO}_2$  from their respective values at their measurement conditions.

Sample\_IDs were calculated from STATION\_ID (station identification number), CAST\_NO (cast number) and NISKIN\_ID (Niskin identification) based on equation (1), if they were not already available:

$$\text{Sample\_ID} = \text{Station\_ID} \times 10000 + \text{Cast\_number} \times 100 + \text{Niskin\_ID} \quad (1)$$

For example, at station 15, the 2nd cast, a Niskin\_ID of 3 will have a Sample\_ID of 150203. In cases when they could not be calculated (e.g., Station\_ID is non-numerical), Sample\_ID was assigned as 1, 2, 3, ... from the first row to the last row of the original cruise data file.

Sampling depth (Depth) and pressure (CTDPRES) were calculated from one another where applicable using the equations of “gsw\_z\_from\_p”, and “gsw\_p\_from\_z”, respectively, from the International Thermodynamic Equation of Seawater 2010 (TEOS-10; IOC et al., 2010). When both values were available, CTDPRES values were preferentially used, and the calculated Depth values were used to replace the original Depth values.

The “recommended\_salinity\_PSS78” column was created by merging the discrete salinity and CTDSAL columns. Data were preferentially chosen from the discrete measurements provided their QC flags were equal to 2 or 6. If these values were not available, CTDSAL values with QC flags of 2 or 6 were chosen. In the absence of these two, discrete salinity measures with QC flags other than 2 or 6 were chosen. Lastly, the CTDSAL values with other QC flags were chosen. The same principles were applied to merge the oxygen data. The merged discrete oxygen and CTDOXY data were stored in the column named “recommended\_Oxygen. (Table 1).

Conservative temperature ( $\Theta$ ) is proportional to the potential enthalpy and is recommended as a replacement for potential temperature ( $\theta$ ), as it more accurately represents the heat content (IOC et al., 2010). Absolute Salinity ( $S_A$ ) is the mass fraction of salt in seawater (unit: g/kg) based on conductivity ratio plus a regional correction term as opposed to the practical salinity scale (SP, Practical Salinity Scale 1978, or PSS-78, unitless, based solely on the conductivity ratio) (Le Menn et al., 2018). Conservative temperature, absolute salinity, and sigma-theta were calculated using the equations of “gsw\_CT\_from\_t”, “gsw\_SA\_from\_SP”, and “gsw\_sigma0”, respectively, from the TEOS-10 (IOC et al., 2010). Apparent oxygen utilization (AOU) was calculated based on absolute salinity, conservative temperature, latitude, longitude, CTDPRES, and recommended\_Oxygen variable using the function “gsw\_O2sol” as described in the TEOS-10 (IOC et al., 2010). Oxygen solubility is estimated with the combined equation from Garcia and Gordon (1992).

In order to measure nitrate, it is first reduced to nitrite and then this new nitrite is measured alongside the nitrite originally in seawater (Hydes and Hill, 1985). The concentration of nitrite in ocean water is usually much lower than nitrate. When nitrite is not reported, it is often because its concentration is too low to be detectable. For the CODAP-NA data product, when Nitrate values were not available, but both Nitrate\_and\_Nitrite and Nitrite values with QC flags of 2 or 6 were available, Nitrate values were calculated by subtracting Nitrite from Nitrate\_and\_Nitrite. Similarly, when Nitrate\_and\_Nitrite values were not available, but both Nitrate and Nitrite values with QC flags of 2 or 6 were available, Nitrate\_and\_Nitrite values were calculated by adding Nitrate and Nitrite concentrations together. The “recommended\_Nitrate\_and\_Nitrite” column was created by preferentially using Nitrate\_and\_Nitrite values. In cases when Nitrate\_and\_Nitrite values were not available but Nitrate values with a QC flag of 2 or 6 were available (Nitrite values not available), the Nitrate\_and\_Nitrite values were assumed to equal the Nitrate values.

Carbonate\_insitu\_measured, pH\_TS\_insitu\_measured, and  $f\text{CO}_2$ \_insitu\_measured (Table 1) were recalculated from their respective values at measurement conditions (i.e., pH\_TS\_measured, Carbonate\_measured, and  $f\text{CO}_2$ \_insitu\_measured) with the CO2SYS program, using the dissociation constants as described above. TALK was preferentially used as the second carbon parameter. When it was not available, DIC was used. If neither of them was available, TALK derived from salinity with the locally interpolated alkalinity regression (LIARv2) method was used for the adjustment from measurement to *in-situ* conditions (Carter et al., 2018). Carbonate\_insitu\_calculated, pH\_TS\_insitu\_calculated,  $f\text{CO}_2$ \_insitu\_calculated, aragonite saturation state, calcite saturation state, and Revelle\_Factor were calculated from DIC and TALK, along with *in-situ* temperature, salinity, pressure, silicate, and phosphate using the same dissociation constants as above (Table 1). When either silicate or phosphate data were unavailable, their mean values during the cruise were used for the calculation. Samples with a salinity of less than 15 were excluded from this calculation, due to the potentially large uncertainties.

**Step Three** was to identify outliers. Outliers were determined by visual inspection. Two types of outlier identification were used for this effort: (a) a broad-scale outlier identification by visually examining the plot of a variable against its sampling depth and other property-to-property plots, and (b) a fine-scale outlier identification based on consistency checks. Here, consistency checks refer to both the “internal consistency checks”, i.e., the comparison of a measurement with its calculated value (e.g., spectrophotometrically-measured pH vs. pH calculated from other carbon parameters using CO2SYS), as well as validation checks, i.e., a measurement with one method against the same measurement made with a different method (e.g., oxygen measured from Winkler vs. a sensor, though in this case the oxygen profile is frequently adjusted to the Winkler titration values, so the measurements are not truly independent). For the broad-scale outlier identification we made plots of all variables against depth (or sigma-theta when only surface values are available), as well as these plots (Figure 2):

- (a) CTDSAL against CTDTEMP
- (b) TALK against Salinity
- (c) TALK against Silicate
- (d) TALK against DIC

- (e) DIC vs.  $f\text{CO}_2$  (20°C),
- (f) DIC against CTDOXY,
- (g) Phosphate vs Nitrate
- (h) Nitrate vs Silicate
- (i) all nutrients (silicate, phosphate, nitrate, nitrite, nitrate plus nitrite, ammonium) against CTDOXY.

Consistency check-based outlier identification was the primary way of finding outliers in this study. Consistency checks were conducted for these below variable pairs. This has been the most effective way of identifying outliers.

- (a) CTDSAL vs. discrete salinity (discrete salinity as the reference value)
- (b) CTDOXY vs. discrete oxygen measured from Winkler titration (Winkler oxygen as the reference value)
- (c) pH measured with a spectrophotometer vs. pH calculated with CO2SYS from DIC, TALK and other parameters
- (d) Carbonate ion ( $[\text{CO}_3^{2-}]$ ) measured with a spectrophotometer vs.  $[\text{CO}_3^{2-}]$  calculated with CO2SYS from DIC, TALK and other parameters
- (e) Discrete  $f\text{CO}_2$  measured with a non-dispersive infrared analyzer vs.  $f\text{CO}_2$  calculated with CO2SYS from DIC, TALK and other parameters.

In addition, the values for dissolved oxygen, DIC, TALK, Silicate, Phosphate and Nitrate were also calculated from existing estimation algorithms (e.g., Carter et al., 2018). These estimates were then compared against the measured CTDOXY, Oxygen, DIC, TALK, Silicate, Phosphate and Nitrate, respectively, to help assess whether cruise-to-cruise biases exist (Figure 2). These algorithms are intended primarily for open-ocean estimation. They are used in the coastal environment only to call attention to measurements that require additional QC, and never to directly assign flags.

For all the aforementioned plots, we enable features to go through each profile individually with all data from a cruise plotted together in the background. Similarly, we are able to go through each cruise individually with all data from all cruises plotted together in the background. These approaches allow us to detect systematic offsets.

**Step four** was to append all of the individual cruise data files one after another into one data product file with all of the variables as listed in Table 1. For surface samples collected from flow-through systems, their Cast\_numbers and Niskin\_IDs were all set to “-999”, and their Niskin\_flags were all set to “9”. The contents of Observation\_type were standardized to be either “Niskin” or “Flow-through”. Data values with QC flags that were not 2 (good), 3 (questionable), or 6 (average of duplicate measurements) were replaced with “-999”, and their corresponding QC flags were changed to “9”. The merged data product file was further QCed by plotting all of the non-missing values for each variable. These plots were examined further, with focus on the outliers falling out of 2.5 times their respective standard deviations.

**Table 4.** World Ocean Circulation Experiment (WOCE) World Hydrographic Program (WHP) (Joyce and Corry, 1994; Swift and Diggs, 2008) QC flags used for this product.

Flag value	Meaning
2	Acceptable
3	Questionable
6	Average of duplicates
9	Missing value