Coastal Ocean Data Analysis Product in North America (CODAP-NA) - An internally consistent data product for discrete inorganic carbon, oxygen, and nutrients on the North American ocean margins

Li-Qing Jiang[1,2], Richard A Feely[3], Rik Wanninkhof[4], Dana Greeley[3], Leticia Barbero[4,5], Simone Alin[3], Brendan R. Carter[3,6], Denis Pierrot[4], Charles Featherstone[4], James Hooper[4,5], Chris Melrose[7], Natalie Monacci[8], Jonathan Sharp[3,6], Shawn Shellito[9], Yuan-Yuan Xu[4,5], Alex Kozyr[1,2], Robert H. Byrne[10], Wei-Jun Cai[11], Jessica Cross[3], Gregory C. Johnson[3], Burke Hales[12], Chris Langdon[13], Jeremy Mathis[14], Joe Salisbury[9], David W. Townsend[15].

[1]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland 20740, USA.

[2]National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Silver Spring, Maryland 20910, USA.

[3]Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, Washington 98115, U.S.A.

[4]Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, 4301 Rickenbacker Causeway, Miami, Florida 33149, USA.

[5]Cooperative Institute for Marine and Atmospheric Studies, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA.

[6]Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington, 3737 Brooklyn Ave NE, Seattle, Washington 98105, USA.

[7]Northeast Fisheries Science Center, National Oceanic and Atmospheric Administration, 28 Tarzwell Drive, Narragansett, Rhode Island 02882, USA.

[8]Ocean Acidification Research Center, College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, 2150 Koyukuk Drive, Fairbanks, Alaska 99775, USA.

[9]Ocean Process Analysis Lab (OPAL), University of New Hampshire, 8 College Rd, Durham, New Hampshire 03824, USA.

[10]College of Marine Science, University of South Florida, 140 7th Avenue South, St. Petersburg, Florida 33701, USA.

[11]School of Marine Science and Policy, University of Delaware, 261 S. College Ave, Newark, Delaware 19716, USA.

[12]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, 2651 SW Orchard Avenue, Corvallis, Oregon 97331, USA.

[13]Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA.

[14]Walsh School of Foreign Service, Georgetown University, Washington, District of Columbia 20057, USA.

[15]School of Marine Sciences, University of Maine, Orono, ME 04469 U.S.A.

Correspondence to: Li-Qing Jiang (Liqing.Jiang@noaa.gov)

Abstract. Internally-consistent, quality-controlled data products play an important role in promoting regional to global research efforts to understand societal vulnerabilities to ocean acidification (OA). However, there are currently no such data products for the coastal ocean where most of the OA-susceptible commercial and recreational fisheries and aquaculture industries are located. In this collaborative effort, we compiled, quality controlled (QC), and synthesized two decades of discrete measurements of inorganic carbon system parameters, oxygen, and nutrient chemistry data from the North American continental shelves, to generate a data product called the Coastal Ocean Data Analysis Product for North America (CODAP-NA). There are few deep-water (>1500m) sampling locations in the current data product. As a result, cross-over analyses, which rely on comparisons between measurements on different cruises in the stable deep ocean, could not form the basis for cruise-to-cruise adjustments. For this reason, care was taken in the selection of data sets to include in this initial release of CODAP-NA, and only data sets from laboratories with known quality assurance practices were included. New consistency checks and outlier detections were used to QC the data. Future releases of this CODAP-NA product will use this core data product as the basis for cruise-to-cruise comparisons. We worked closely with the investigators who collected and measured these data during the QC process. This version of the CODAP-NA is comprised of 3,391 oceanographic profiles from 61 research cruises covering all continental shelves of North America, from Alaska to Mexico in the west and from Canada to the Caribbean in the east. Data for 14 variables (temperature; salinity; dissolved oxygen concentration; dissolved inorganic carbon concentration; total alkalinity; pH on the Total Scale; carbonate ion concentration; fugacity of carbon dioxide; and concentrations of silicate, phosphate, nitrate, nitrite, nitrate plus nitrite, and ammonium) have been subjected to extensive QC. CODAP-NA is available as a merged data product (Excel, CSV, MATLAB, and NetCDF, doi:10.25921/531n-c230, https://www.ncei.noaa.gov/data/oceans/ncei/ocads/metadata/0219960.html) (Jiang et al., 2021). The original cruise data have also been updated with data providers' consent and summarized in a table with links to NOAA's National Centers for Environmental Information (NCEI) archives (https://www.ncei.noaa.gov/access/ocean-acidification-data-stewardship-oads/synthesis/NAcruises.html).

1 Introduction

Anthropogenic ocean acidification (OA) refers to the process by which the ocean's uptake of excess anthropogenic atmospheric carbon dioxide (CO_2) reduces ocean pH and calcium carbonate mineral saturation states (Feely et al., 2004; Orr et al., 2005; Jiang et al., 2019; IPCC, 2011). OA is making it more difficult for marine calcifiers to build shells and skeletal structures and is endangering coral reefs and other marine ecosystems (Doney et al., 2009; Gattuso and Hanson, 2011). Coastal ecosystems account for most of the economic activities related to commercial and recreational fisheries and aquaculture industries, supporting about 90% of the global fisheries yield and 80% of known species of marine fish (Cicin-

Sain et al., 2002). Studies have shown that OA has the potential to significantly impact both the fisheries and aquaculture industries, and change the way humans make their living, run their communities, and live their lives in coastal regions around the world (Cooley and Doney, 2009; Barton et al., 2012, 2015).

The Global Ocean Data Analysis Project (GLODAPv2) offers an internally consistent data product for discrete sampling-based, open-ocean carbonate chemistry, nutrient chemistry, isotopes, and transient tracer data (Olsen et al., 2016; Olsen et al., 2020), allowing for a slew of new research products related to OA and its temporal trends in the global ocean (e.g., Lauvset et al., 2015; Jiang et al., 2015a; Gruber et al., 2019; Lauvset et al., 2020). While there are several coastal surface water partial pressure of CO_2 (pCO_2) data products and climatologies (e.g., Bakker et al., 2016; Laruelle et al., 2017; Roobaert et al., 2019; Takahashi et al., 2020), internally consistent data products for water column carbonate and nutrient chemistry data in the coastal ocean currently do not exist. Such products would contribute significantly to our understanding of the current status of OA and its temporal trends, and help guide OA mitigation and adaptation efforts in coastal oceans.

The impact of OA on North American ocean margins is expected to vary significantly from region to region, with distinct regional drivers amplifying or mitigating overall coastal acidification. Anthropogenic carbon dioxide (CO_2) invasion has been identified as the primary driver of open ocean acidification over decadal time scales, but coastal ocean acidification is influenced by many other physical, biological, and anthropogenic processes that can oppose or amplify the anthropogenic CO_2 uptake. The continental West Coast (WC) and East Coast (EC) are in two vastly different ocean basins (Pacific vs. Atlantic) with different amounts of net organic matter remineralization in deeper waters flowing along the path of the Global Thermohaline Circulation (Broecker, 1991; Feely et al., 2008; Jiang et al., 2010; Wanninkhof et al., 2015). In the surface ocean, latitudinal variation of sea surface temperature (SST) and the ratio of dissolved inorganic carbon (DIC) to total alkalinity (TA) result in significantly different pH and calcium carbonate mineral saturation states between the Alaska Coast and Gulf of Mexico (Jiang et al., 2019; Cai et al., 2020). Upwelling can bring deep waters with corrosive OA chemistry (resulting from large respiratory CO_2 loads) to the surface, while onshore surface flow can bring less-corrosive open ocean waters to the coastline (Hales et al., 2005; Feely et al., 2008, 2016). Riverine input of low-pH water is found to intensify OA shoreward of the shelf break on the EC (Hunt et al., 2011; Xue et al., 2016). However, riverine water composition also varies significantly and the Mississippi River is a source of high-TA water to the Gulf of Mexico (Cai et al., 2008; Stets et al., 2014; Gomez et al., 2020). Eutrophication (enhancement of biological production of organic matter through addition of nutrients) causes high pH and calcium carbonate mineral saturation states in surface waters of the coastal ocean, and can lead to subsurface hypoxia (via subsequent respiration of that production), which is associated with low pH and calcium carbonate mineral saturation (Borges and Gypens, 2010; Cai et al., 2011; Laurent el al., 2017; Feely et al., 2016, 2018). The lack of OA synthesis efforts on North American ocean margins hampers our understanding of the geographic pattern and relative regional progression rates of OA along these coastlines (Cai et al., 2020).

Deleted:

Deleted:

Deleted: where most of the global fisheries and aquaculture industries are focused.

Deleted: U.S.

Carbonate data in the coastal ocean are often collected by multiple laboratories with different methods and instruments.
Many of the data sets may have never been shared with any major data centers, nor have these data sets gone through

120 rigorous quality control (QC) and inter-comparison analyses. The lack of observations in intermediate and deep water (water depth >1500 m) makes it challenging to adjust the data based on constancy of parameters in deep water (i.e., cross-over analyses) as is done for the open ocean (Lauvset and Tanhua, 2015). All these factors contribute to the lack of internally consistent data products for these important coastal environments. In this study, we compiled and QCed discrete sampling-based data for inorganic carbon, oxygen, and nutrient chemistry, and hydrographic parameters collected from the entire

125 North American continental shelves. We serve both the internally-consistent climate quality data product, as well as the QCed original cruise data through the NOAA National Centers for Environmental Information (NCEI). This effort will promote future OA research, modeling, and data synthesis in critically important coastal regions to help advance the OA adaptation, mitigation, and planning efforts of North American coastal communities. While we only partially address limitations associated with the lack of deep and intermediate data in this study, we do produce a data product that can be

130 used as the basis to address these limitations and incorporate additional coastal cruises going forward. We hope this release will be considered analogous to GLODAPv2 (Olsen et al. 2016), in the sense that the new data sets added in the subsequent GLODAPv2.2019 and .2020 updates (Olsen et al., 2019; 2020) were brought to be internally consistent with the fully quality-controlled data in the original GLODAPv2 product.

2 Study area

135 From a geopolitical perspective, the term "continental shelf" is defined as the region between the coastline (excluding estuaries) and a distance of 200 nautical miles (~370 km) offshore. While this definition is not as mechanistic as one based on a change in bathymetric gradient or a hydrographic condition such as chlorophyll or salinity levels, it is regionally and seasonally invariant, and captures the full extent of coastal influences (Hales et al., 2008). This version of the data product is focused on the continental shelves of the North American (NA) coasts (Figure 1), including:

140 - Alaska Coast (AC) – including the large marine ecosystems (LMEs) of Gulf of Alaska, East Bering Sea, Northern Bering-Chukchi Seas, and Beaufort Sea (see Sherman et al., 2009 for more information on the LMEs).
- West Coast (WC) – including the LMEs of California Current and Gulf of California.
- East Coast (EC) – including the LMEs of Northeast U.S. and Southeast U.S. continental shelf regions.
- Gulf of Mexico (GMx)

145 Data beyond continental shelves will be included if they are collected from a cruise that predominately covers parts of the North American ocean margins.

Figure 1. A map showing all the sampling profiles of the CODAP-NA data product (v2021, a total of 3,391 profiles). Magenta dots show the sampling profiles in the Alaska Coast (AC). Blue ones are for the West Coast (WC), green ones are for the East Coast (EC), and the red ones are for the Gulf of Mexico (GMx). Numbers within the parentheses indicate the total number of profiles in the region.

Parameters / variables

For the current version of the CODAP-NA, inorganic carbon system parameters, oxygen, nutrients, and related hydrographic parameters were included (Table 1). CTDPRES, CTDTEMP, CTDSAL, and CTDOXY were commonly measured with pressure, temperature, conductivity, and oxygen sensors, respectively, mounted on a CTD rosette. In some cruises with surface samples collected from flow-through systems, temperature and salinity were also provided in columns reserved for CTDTEMP and CTDSAL, respectively. Water samples were collected and measured onboard or later in a shore-based laboratory for discrete salinity, discrete dissolved oxygen concentration (DO), dissolved inorganic carbon concentration (DIC), total alkalinity (TALK), pH, carbonate ion concentration ($[CO_3^{2-}]$), fugacity of carbon dioxide ($f\text{CO}_2$), and concentrations of silicic acid, phosphate, nitrate, nitrite, nitrate plus nitrite, and ammonium. For discrete pH on the Total Scale, $[CO_3^{2-}]$, and $f\text{CO}_2$, both measured and calculated values were presented. Saturation states of aragonite (Ω_{arag}) and calcite (Ω_{calc}) could only be calculated. The carbonate system calculations were conducted using the MATLAB version 3.01 (Sharp et al., 2020) of the CO2SYS program (Lewis and Wallace, 1998), with the dissociation constants for carbonic acid of Lueker et al. (2000), bisulfate (HSO_4^-) of Dickson (1990), hydrofluoric acid (HF) of Perez and Fraga (1987), and with the total borate equations of Lee et al., (2010).

225 **Table 1.** Parameters that are included in the CODAP-NA (v2021) data product.

Abbreviation	Variable name	Unit	Measured/ calculated
CTDPRES	Water pressure recorded from sensors on a CTD rosette. For surface samples collected from an onboard flow-through system, its pressure is equal to the depth of the water inlet. When such info is not available, it is assumed to be 5 dbar.	dbar	measured
CTDTEMP_ITS90	Temperature on the International Temperature Scale of 1990 (ITS-90) from sensors on a CTD rosette. For surface samples collected from an onboard flow-through system, temperature has also been merged into the CTDTEMP_ITS90 variable.	°C	measured
CTDSAL_PSS78	Salinity on the Practical Salinity Scale 1978 (PSS-78) from sensors on a CTD rosette. For surface samples collected from an onboard flow-through system, salinity from the thermosalinograph (TSG) has been merged into the CTDSAL_PSS78 variable.	-	measured
Salinity_PSS78	Salinity on the PSS-78 scale measured as discrete samples	-	measured
recommended_Salinity_PSS78	Discrete salinity with some missing values filled in using CTDSAL	-	measured
CTDOXY	Dissolved oxygen concentration from sensors on a CTD rosette	$\mu mol\ kg^{-1}$	measured
Oxygen	Dissolved oxygen concentration measured as discrete samples from Winkler titration	$\mu mol\ kg^{-1}$	measured
recommended_Oxygen	Discrete dissolved oxygen concentration from Winkler titration with some missing values filled in using CTDOXY	$\mu mol\ kg^{-1}$	measured
AOU	Apparent oxygen utilization	$\mu mol\ kg^{-1}$	calculated
DIC	Dissolved inorganic carbon concentration	$\mu mol\ kg^{-1}$	measured
TALK	Total alkalinity	$\mu mol\ kg^{-1}$	measured
pH_TS_measured	pH on total hydrogen scale (TS) measured as discrete samples at measurement temperature and ambient pressure	-	measured
TEMP_pH	Temperature of pH measurement	°C	measured
pH_TS_insitu_measured	pH on total hydrogen scale (TS) measured as discrete samples and adjusted to *in-situ* conditions	-	measured
pH_TS_insitu_calculated	pH on total hydrogen scale (TS) at *in-situ* conditions calculated from DIC, TA and other parameters using CO2SYS	-	calculated
Carbonate_measured	Carbonate ion concentration measured as discrete samples at measurement temperature and ambient pressure	$\mu mol\ kg^{-1}$	measured
TEMP_Carbonate	Temperature of carbonate ion measurement	°C	measured
Carbonate_insitu_measured	Carbonate ion concentration measured as discrete samples and adjusted to *in-situ* conditions	$\mu mol\ kg^{-1}$	measured
Carbonate_insitu_calculated	Carbonate ion concentration at *in-situ* conditions calculated from DIC, TA and other parameters using CO2SYS	$\mu mol\ kg^{-1}$	calculated
$f\mathrm{CO_2}$_measured	Fugacity of carbon dioxide measured as discrete samples at measurement temperature and ambient pressure	μatm	measured
TEMP_$f\mathrm{CO_2}$	Temperature of $f\mathrm{CO_2}$ measurement	°C	measured
$f\mathrm{CO_2}$_insitu_measured	Fugacity of carbon dioxide measured as discrete samples and adjusted to *in-situ* conditions	μatm	measured
$f\mathrm{CO_2}$_insitu_calculated	Discrete fugacity of carbon dioxide at *in-situ* conditions calculated from DIC, TA and other parameters using CO2SYS	μatm	calculated
Aragonite	Aragonite saturation state at *in-situ* conditions calculated from DIC, TA and other parameters using CO2SYS	-	calculated

Calcite	Calcite saturation state at *in-situ* conditions calculated from DIC, TA and other parameters using CO2SYS	-	calculated
Revelle_Factor	Revelle Factor calculated from DIC, TA and other parameters using CO2SYS	-	calculated
Silicate	Silicate concentration	μmol kg^{-1}	measured
Phosphate	Phosphate concentration	μmol kg^{-1}	measured
Nitrate	Nitrate concentration	μmol kg^{-1}	measured
Nitrite	Nitrite concentration	μmol kg^{-1}	measured
Nitrate_and_Nitrite	Nitrate and Nitrite concentrations combined	μmol kg^{-1}	measured
recommended_Nitrate_and_Nitrite	Nitrate_and_Nitrite concentration, along with Nitrate concentration when Nitrate_and_Nitrite data are not available	μmol kg^{-1}	measured
Ammonium	Ammonium concentration	μmol kg^{-1}	measured

4 Data sources

CODAP-NA was focused on chemical oceanographic data (inorganic carbon system parameters, oxygen, and nutrients) collected from discrete sampling-based observations. This also included discrete samples taken from shipboard flow-through systems rather than solely water collected in sampling rosette bottles. Carbon parameters recorded from continuous underway measurements by inline analytical instruments were excluded, as they had been QCed and included within the Surface Ocean CO$_2$ Atlas (SOCAT) (Bakker et al., 2016). The same was true for carbon parameters from time-series moorings. Data from large open estuaries (e.g., Salish Sea, Chesapeake Bay, Bay of Fundy) are excluded during this first round of analysis, but these are among the data that may be able to benefit from secondary QC against CODAP-NA. When a cruise spans ocean margins and also contains a subset of measurements within estuaries, the estuarine data from that cruise is retained for this data product.

We started with the highest quality coastal data sets to define a protocol for consistent QC and inter-comparison, which will subsequently be applied to other compiled coastal data sets. As a first step, only climate-quality discrete measurements (core data sets) with known quality and metadata from the Atlantic Oceanographic and Meteorological Laboratory, Pacific Marine Environmental Laboratory, University of South Florida, University of Miami, University of Alaska Fairbanks, University of New Hampshire, and University of Delaware were included (Table 2). These data sets will serve as a reference for QCing future data sets.

Table 2. List of cruises that are included in this version (v2021) of the CODAP-NA data product. Refer to Table 1 for the full names of the abbreviations and their units, and Table 3 for definitions of the Cruise_flags. CTD is short for conductivity, temperature, and depth, and refers to a package of electronic instruments that measure these properties. Start date and End date refer to the dates when data were first and last collected, respectively. For samples collected from flow-through systems, temperature and salinity were also stored in CTDTEMP and CTDSAL, respectively. GMx is short for Gulf of Mexico. FT is short for Flow-through systems.

NO	Region	EXPOCODE	Observation type	Cruise flag	CRUISE ID	Start date	End date	Variables measured
1	Alaska Coast	33HQ20080329	Niskin	B	HLY0802	2008-04-01	2008-05-06	CTDTEMP, CTDSAL, DIC, TALK, CTDOXY, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
2	Alaska Coast	33HQ20080703	Niskin	B	HLY0803	2008-07-04	2008-07-30	CTDTEMP, CTDSAL, DIC, TALK, CTDOXY, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
3	Alaska Coast	33HQ20090403	Niskin	B	HLY0902	2009-04-05	2009-05-10	CTDTEMP, CTDSAL, DIC, TALK, Silicate, Phosphate, Nitrate, Nitrite
4	Alaska Coast	33HQ20100907	Niskin	B	HLY1003	2010-09-07	2010-09-08	CTDTEMP, CTDSAL, DIC, TALK
5	Alaska Coast	33HQ20111003	Niskin	B	HLY1103	2011-10-06	2011-10-19	CTDTEMP, CTDSAL, TALK, pH
6	Alaska Coast	33HQ20121005	Niskin	B	HLY1203	2012-10-10	2012-10-20	CTDTEMP, CTDSAL, DIC, TALK
7	Alaska Coast	33HQ20131005	UW	D	HLY1303	2013-10-05	2013-10-30	CTDTEMP, CTDSAL, DIC, TALK
8	Alaska Coast	316N20090614	Niskin	B	KN195	2009-06-22	2009-07-13	CTDTEMP, CTDSAL, DIC, TALK, Silicate, Phosphate, Nitrate
9	Alaska Coast	31FN20090924	Niskin	B	MF0904	2009-09-26	2009-10-09	CTDTEMP, CTDSAL, DIC, TALK, CTDOXY, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
10	Alaska Coast	33RO20150713	Niskin	B	RB1504	2015-07-17	2015-07-31	CTDTEMP, CTDSAL, DIC, TALK, CTDOXY, Silicate, Phosphate, Nitrate, Nitrite
11	Alaska Coast	325020100509	Niskin	B	TN249-10	2010-05-13	2010-07-12	CTDTEMP, CTDSAL, DIC, TALK, CTDOXY, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
12	East Coast	316G20120202	Niskin	C	DE1202	2012-02-06	2012-02-19	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
13	East Coast	33GG20150619	Niskin	B	ECOA1	2015-06-20	2015-07-23	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Nitrate_and_Nitrite
14	East Coast	33HH20180625	Niskin	B	ECOA2	2018-06-25	2018-07-29	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate_and_Nitrite
15	East Coast	334A20140510	Niskin	C	EX1403	2014-05-10	2014-05-17	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
16	East Coast	33RO20070710	Niskin	A	GOMECC1	2007-07-11	2007-08-02	CTDTEMP, CTDSAL, Salinity, DIC, TALK, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
17	East Coast	33RO20120721	Niskin	A	GOMECC2	2012-07-22	2012-08-13	CTDTEMP, CTDSAL, Salinity, DIC, TALK, pH, Carbonate, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Nitrate_and_Nitrite
18	East Coast	33GG20130609	Niskin, FT	C	GU1302	2013-06-09	2013-06-23	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
19	East Coast	33GG20131113	Niskin	C	GU1305	2013-11-14	2013-11-24	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
20	East Coast	33GG20140301	Niskin	C	GU1401 Leg2	2014-03-01	2014-03-08	CTDTEMP, CTDSAL, DIC, TALK, pH, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
21	East Coast	33GG20151012	Niskin	C	GU1506 Leg2	2015-10-13	2015-10-24	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
22	East Coast	33GG20160521	Niskin	C	GU1608 Leg1	2016-05-23	2016-06-02	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium

23	East Coast	33GG20160607	Niskin	C	GU1608 Leg2	2016-06-08	2016-06-12	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
24	East Coast	33GG20170516	Niskin, FT	C	GU1701 Leg1	2017-05-17	2017-05-25	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
25	East Coast	33GG20170530	Niskin, FT	C	GU1701 Leg2	2017-05-31	2017-06-05	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
26	East Coast	33GG20170610	Niskin, FT	C	GU1702	2017-06-12	2017-06-21	CTDTEMP, CTDSAL, DIC, TALK, pH
27	East Coast	33GG20171031	Niskin	C	GU1706	2017-11-01	2017-11-07	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
28	East Coast	33GG20180822	Niskin, FT	C	GU1804	2018-08-23	2018-08-29	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite
29	East Coast	33HH20120531	Niskin, FT	C	HB1202	2012-06-02	2012-06-13	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
30	East Coast	33HH20130314	FT	D	HB1301	2013-03-17	2013-05-09	CTDTEMP, CTDSAL, DIC, TALK, pH, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
31	East Coast	33HH20140908	FT	D	HB1405 Leg1	2014-09-10	2014-09-18	CTDTEMP, CTDSAL, DIC, TALK, pH
32	East Coast	33HH20140923	FT	D	HB1405 Leg2	2014-09-25	2014-09-30	CTDTEMP, CTDSAL, DIC, TALK, pH
33	East Coast	33HH20141028	FT	D	HB1405 Leg4	2014-11-04	2014-11-05	CTDTEMP, CTDSAL, DIC, TALK, pH
34	East Coast	33HH20150519	Niskin	C	HB1502	2015-05-20	2015-06-02	CTDTEMP, CTDSAL, DIC, TALK, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
35	East Coast	33HH20170211	Niskin, FT	C	HB1701	2017-02-12	2017-02-22	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
36	East Coast	33HH20180523	Niskin, FT	C	HB1803	2018-05-23	2018-06-04	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
37	East Coast	334B20121026	Niskin	C	PC1207	2012-10-27	2012-11-13	CTDTEMP, CTDSAL, DIC, TALK, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
38	East Coast	334B20141103	Niskin	C	PC1405	2014-11-04	2014-11-18	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
39	East Coast	334B20160807	Niskin	C	PC1604	2016-08-09	2016-08-19	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
40	East Coast	334B20161018	Niskin	C	PC1609	2016-10-19	2016-10-19	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
41	East Coast	33H520181102	Niskin	C	S11802	2018-11-02	2018-11-12	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite, Ammonium
42	East Coast	AGSK20031205	FT	D	SKO0313	2003-12-06	2003-12-14	CTDTEMP, CTDSAL, DIC
43	East Coast	AGSK20040403	FT	D	SKO0406	2004-04-04	2004-04-11	CTDTEMP, CTDSAL, DIC
44	East Coast	AGSK20040625	FT	D	SKO0410	2004-06-26	2004-07-02	CTDTEMP, CTDSAL, DIC, TALK
45	East Coast	AGSK20041015	FT	D	SKO0414	2004-10-16	2004-10-22	CTDTEMP, CTDSAL, DIC, TALK
46	East Coast	AGSK20050916	FT	D	SKO0510	2005-09-17	2005-09-23	CTDTEMP, CTDSAL, DIC, TALK
47	East Coast	AGSK20060403	FT	D	SKO0604	2006-04-18	2006-04-27	CTDTEMP, CTDSAL, DIC, TALK
48	East Coast	AGSK20061014	FT	D	SKO0611	2006-10-15	2006-10-23	CTDTEMP, CTDSAL, DIC, TALK
49	East Coast	AGSK20070525	FT	D	SKO0721	2007-05-26	2007-06-02	CTDTEMP, CTDSAL, DIC
50	East Coast	AGFO20140607	FT	D	SKO1406	2014-06-20	2014-06-28	CTDTEMP, CTDSAL, DIC, TALK, pH
51	East Coast	46SL20181115	FT	D	Selfoss846	2018-11-15	2018-11-22	CTDTEMP, CTDSAL, DIC, TALK, pH, Silicate, Phosphate, Nitrate_and_Nitrite

52	GMx	33RO20170718	Niskin	A	GOMEC C3	2017-07-18	2017-08-20	CTDTEMP, CTDSAL, Salinity, DIC, TALK, pH, Carbonate, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Nitrate_and_Nitrite
53	GMx	33WA20141201	Niskin	C	WS1418	2014-12-03	2014-12-04	CTDTEMP, CTDSAL, DIC, TALK, pH, CTDOXY, Silicate, Phosphate, Nitrate_and_Nitrite
54	GMx	33WA20150921	Niskin, FT	C	WS15264	2015-09-23	2015-09-24	CTDTEMP, CTDSAL, DIC, TALK, pH, Silicate, Phosphate, Nitrate_and_Nitrite
55	West Coast	332220170918	Niskin	A	SH1709	2017-09-18	2017-09-28	CTDTEMP, CTDSAL, DIC, TALK, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
56	West Coast	32WC20070511	Niskin	A	WCOA2007	2007-05-14	2007-06-12	CTDTEMP, CTDSAL, Salinity, DIC, TALK, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite
57	West Coast	32WC20110812	Niskin	A	WCOA2011	2011-08-12	2011-08-30	CTDTEMP, CTDSAL, Salinity, DIC, TALK, pH, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
58	West Coast	332220120904	Niskin	A	WCOA2012	2012-09-05	2012-09-16	CTDTEMP, CTDSAL, Salinity, DIC, TALK, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
59	West Coast	317W20130803	Niskin	A	WCOA2013	2013-08-05	2013-08-10	CTDTEMP, CTDSAL, Salinity, DIC, TALK, pH, Carbonate, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
60	West Coast	32P020130821	Niskin	A	WCOA2013	2013-08-21	2013-08-28	CTDTEMP, CTDSAL, Salinity, DIC, TALK, pH, Carbonate, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Ammonium
61	West Coast	33RO20160505	Niskin	A	WCOA2016	2016-05-08	2016-06-06	CTDTEMP, CTDSAL, Salinity, DIC, TALK, pH, Carbonate, CTDOXY, Oxygen, Silicate, Phosphate, Nitrate, Nitrite, Ammonium

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.

Zero level QC by Submitting PIs	Primary level QC conducted by the data product development community		Second level QC
(focused on outliers within a cruise) - Property to depth plots - Property to property plots	**Data loading & calculation** - Calculate depth from pressure and vice versa. - Use DIC & TALK to calculate pH, $[CO_3^{2-}]$, and fCO_2. - Adjust pH, $[CO_3^{2-}]$ and fCO_2 from report temperature to in-situ temperature. - Calculate AOU and Percent oxygen. - Use algorithms to estimate properties from CTDTEMP, CTDSAL and CTDOXY.	**Property vs. depth plots** (All variables against their sampling depths) - CTDTEMP, CTDSAL - CTDOXY, Oxygen, AOU, and Percent oxygen - DIC, TALK, pH, $[CO_3^{2-}]$, fCO_2 - Silicate, Phosphate, Nitrate - Nitrite, Ammonium	(focused on cruise to cruise offsets) - Crossover analyses - Offset or factor adjustment
	Property vs. property - CTDSAL vs. CTDETMP - TALK vs. Salinity - TALK vs. Silicate - TALK vs. DIC - DIC vs. fCO_2 (20°C) - DIC vs. CTDOXY - Phosphate vs. Nitrate - Nitrate vs. Silicate - Nutrients vs. CTDOXY	**Internal consistency** (Property vs. property, delta value vs. depth, delta value vs. property): - CTDSAL vs. Salinity (discrete) - CTDOXY vs. Oxygen (Winkler) - Measured pH vs. calculated pH - Measured $[CO_3^{2-}]$ vs. calculated $[CO_3^{2-}]$ - Measured fCO_2 vs. calculated fCO_2. **Property vs. algorithm derived values** - Oxygen - DIC - TALK - Silicate - Phosphate - Nitrate	

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). $[CO_3^{2-}]$ is carbonate ion concentration, fCO_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, O₂, 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). All bottle measurement flags with a corresponding Niskin_flag of 3 or 4 were replaced with the corresponding Niskin_flags. For example, if a discrete salinity measurement has a Salinity_flag

of 2, but the corresponding Niskin_flag (QC flag of the Niskin bottle where the sample was drawn) is 3, the original Salinity_flag will be updated from 2 to 3.

325

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

330 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:

335
    (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

340
    (f) apparent oxygen utilization (AOU);

    (g) recommended_Nitrate_and_Nitrite;

    (h) calculated pH, carbonate ion, and fCO$_2$ 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 fCO$_2$ from their respective values at their measurement conditions.

345

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:

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

350

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.

355 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.

365 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

370 "recommended_Oxygen. (Table 1).

Conservative temperature (Θ) is proportional to the potential enthalpy and is recommended as a replacement for potential temperature (θ), 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

375 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.,

380 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

385 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

390 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 fCO$_2$_insitu_measured (Table 1) were recalculated from their respective values at measurement conditions (i.e., pH_TS_measured, Carbonate_measured, and fCO$_2$_insitu_measured) with

395 the CO2SYS program, using the dissociation constants as described above. TALK was preferentially used as the second

Deleted: 2

Deleted: nitrate plus

Deleted: together

Deleted: be approximate to

Deleted: 2

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, fCO₂_insitu_calculated, aragonite saturation state, calcite saturation state, and Revelle_Factor were calculated from DIC and TALK, along with *in-*

405 *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
410 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.,
415 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
420   (c) TALK against Silicate
      (d) TALK against DIC
      (e) DIC vs. fCO₂ (20°C),
      (f) DIC against CTDOXY,
      (g) Phosphate vs Nitrate
425   (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.
430   (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 ([CO₃²⁻]) measured with a spectrophotometer vs. [CO₃²⁻] calculated with CO2SYS from DIC,
          TALK and other parameters

(e) Discrete fCO$_2$ measured with a non-dispersive infrared analyzer vs. fCO$_2$ 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. All rows with a Niskin_flag of "4" (Table 4) were removed. 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". 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". 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

6 Data products

The data product is available in Excel, CSV, MATLAB, and NetCDF formats at NOAA/NCEI with a DOI of [10.25921/531n-c230] and NCEI Accession Number of [0219960] (Jiang et al., 2021). All parameters in Table 1, along with their Cruise_flags (Table 3) and primary level QC flags (Table 4) are presented. The chosen primary level QC flag convention is the same as the GLODAPv2 project (Olsen et al., 2020). Note the difference between the WOCE primary level QC flags (e.g., 2, 3, 4, 9, etc.) and the Secondary QC flags as used by the GLODAPv2 (a choice of either 0 or 1). In the current version (v2021) of the CODAP-NA, there are 3,391 discrete chemical oceanographic profiles, and a total of 28,206 data points. They were collected on 61 cruises in the ocean margins of North America from December 6, 2003 to November 22, 2018. There are on average eight sampling depth levels (a median of seven) for each profile. The total count of data points for each parameter and their minimum, maximum, and mean values are listed in Table 5.

Table 5. The minimum, maximum, mean, and data point counts of the parameters that are included in the final product. Refer to Table 2 for their full parameter names and units.

Abbreviation	Units	Min	Max	Mean	Count
CTDTMP_ITS90	°C	-1.79	31.74	10.84	26,679
CTDSAL_PSS78	-	1.38	37.61	33.4	26,641
Salinity_PSS78	-	0.12	36.97	34.8	3,485
recommended_Salinity_PSS78	-	0.12	37.61	33.4	26,648
CTDOXY	µmol kg⁻¹	3.1	616	215.6	24,198
Oxygen	µmol kg⁻¹	2.7	472	184	11,869
recommended_Oxygen	µmol kg⁻¹	2.7	616	216.7	24,259
AOU	µmol kg⁻¹	-299.3	313.5	58.4	24,203
DIC	µmol kg⁻¹	886.4	2621	2104	18,341
TALK	µmol kg⁻¹	878	2853	2251	18,351
pH_TS_insitu_measured	-	7.43	8.58	7.93	10,575
pH_TS_insitu_calculated	-	7.37	8.54	7.93	17,031
Carbonate_insitu_measured	µmol kg⁻¹	21.7	325	121.7	5,021
Carbonate_insitu_calculated	µmol kg⁻¹	21.4	306.9	114.2	17,031
fCO2_insitu_measured	µatm	198.3	1171	449.9	3,173
fCO2_insitu_calculated	µatm	105.9	2074	574.7	17,031
Aragonite	-	0.34	5.1	1.72	17,031
Calcite	-	0.58	7.67	2.68	17,031
Revelle_Factor	-	8.26	19.7	14.31	17,031
Silicate	µmol kg⁻¹	0	234.4	24.38	18,772
Phosphate	µmol kg⁻¹	0	9.36	1.22	18,709
Nitrate	µmol kg⁻¹	0	97.1	15.63	15,808
Nitrite	µmol kg⁻¹	0	3.76	0.09	14,846
Nitrate_and_Nitrite	µmol kg⁻¹	0	97.1	14.15	17,564
recommended_Nitrate_and_Nitrite	µmol kg⁻¹	0	97.1	14.25	18,640
Ammonium	µmol kg⁻¹	0	13.98	0.54	10,452

Of the 3,391 profiles, 2,869 have both DIC and TALK measurements, thus the full list of carbonate system parameters (pH,

fCO$_2$, [CO$_3^{2-}$], aragonite saturation state, calcite saturation state, and Revelle Factor) can be calculated (Figure 3). In

addition, there are 1,501 profiles with discrete pH measurements from a spectrophotometer-based method (Byrne and

675   Breland, 1989; Clayton and Byrne, 1993; Dickson, 1993), 412 profiles with discrete carbonate ion measurements (Byrne and

Yao, 2008; Sharp and Byrne, 2019), and 278 profiles with discrete fCO$_2$ measurements (Wanninkhof and Thoning, 1993).

There is also good coverage of oxygen and nutrients measurements (Figure 3).

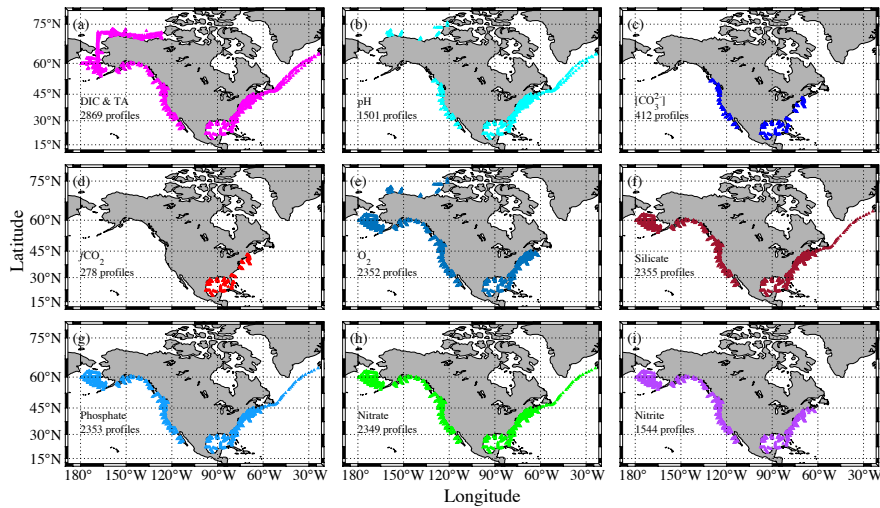



Figure 3. Sampling profiles for certain parameters. A profile is plotted if it has at least one measured value. Panel (a)

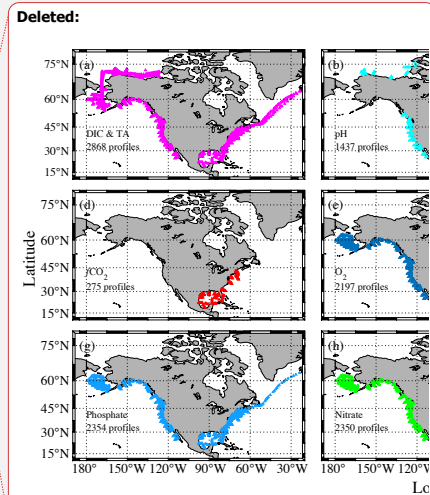
680   **only includes profiles that have both dissolved inorganic carbon (DIC) and total alkalinity (TA) measured. Panel (b)**

is for profiles with discrete pH measurements from a spectrophotometer. Panel (c) is for profiles with discrete

carbonate ion concentration ([CO$_3^{2-}$]) measurements from a spectrophotometer. Panel (d) is for profiles with discrete

fugacity of carbon dioxide (fCO$_2$) measurements from flasks. Panel (e) is for profiles with either

recommended_Oxygen values (see Table 1 for more details). Panels (f-i) are for profiles with nutrient measurements.

685

One major difference between the CODAP-NA and the GLODAPv2 is the shallower sampling depths of the former (Figure 4). About 80% of the 3,391 profiles have a maximum sampling depth of < 250 m, and 30% of them have maximum sampling depth of < 25 m, with a lot of them being surface-only measurements. Only 195 profiles (< 6% of the total 3,391 profiles) have at least one sampling depth level below 1500 m, which has commonly been used as a threshold for subsurface cross-over analyses (Figure 4). Most of these deep-water profiles are found off the West Coast, Gulf of Mexico, and a few offshore stations in the Mid-Atlantic Bight. On average, the sampling depth is 298 m, with a median sampling depth of only 65 m.

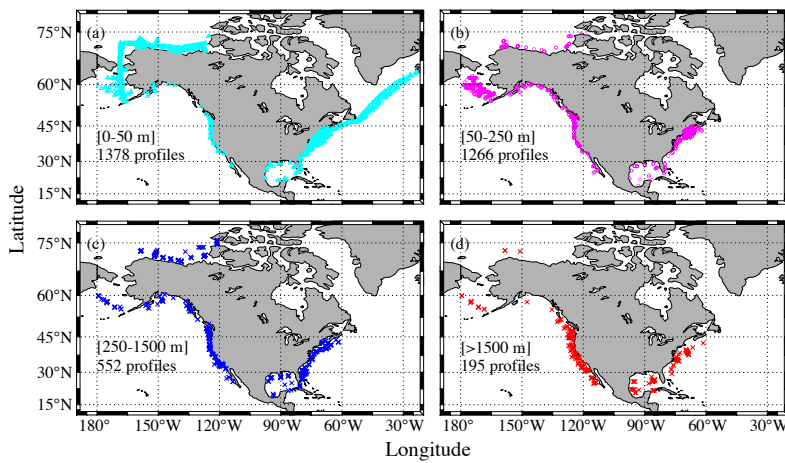
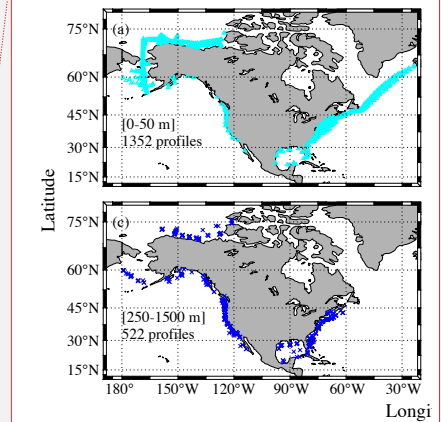



Figure 4. Sampling depths of the profiles: (a) profiles with maximum depths ranging from 0 to 50 m, (b) profiles with maximum depths ranging from 50 to 250 m, (c) profiles with maximum depths ranging from 250 to 1500 m, (d) profiles with maximum depths greater than 1500 m.

Another distinctive feature of coastal oceans is their large magnitude of seasonal variation. For a lot of parameters, their seasonal variation, along with the diel and intertidal variations often eclipse their long-term variation. Understanding the seasonal variation and de-seasonalizing the observation data are often critical steps in the process of deciphering the long-term change. Like most data products, this version of the CODAP-NA is summer- and fall-biased, with spring, summer, fall and winter having 676, 1554, 1059, and 102 profiles, respectively (Figure 5). All coasts have good summer data coverage, but the only area with meaningful winter data coverage is the northeastern U.S. coast (Figure 5, Table 6).

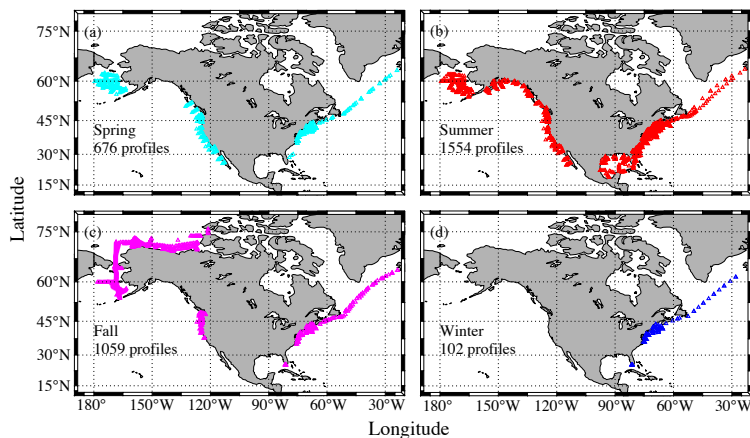
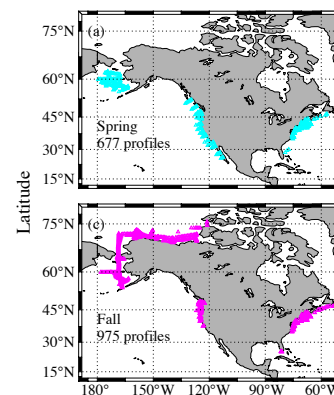



Figure 5. Sampling profiles in each of the four seasons: (a) Spring (March – May), (b) Summer (June – August), (c) Fall (September – November), (d) Winter (December – February).

Table 6. Number of profiles and data points (the sum of all depth levels at each profile) in all seasons of each region.

	Spring (Mar – May)		Summer (Jun – Aug)		Fall (Sep – Nov)		Winter (Dec – Feb)	
	Profiles	Data points	Profiles	Data points	Profiles	Data points	Profiles	Data points
Alaska Coast	300	2053	425	2807	681	2580	0	0
West Coast	167	3024	250	3874	128	1524	0	0
East Coast	209	484	728	8776	235	554	91	235
Gulf of Mexico	0	0	151	2269	15	15	11	11

To demonstrate the large seasonal amplitude (defined here as the difference between the maximum and minimum values of a variable on an annual cycle) in the study area, an analysis was conducted to group surface stations (with at least one sampling depth < 25 m) that are within 1 km distance and have at least one measurement between December and March and one measurement between June and October. The results, which are based on 135 groups of stations (most of them in the northeastern U.S. coast), show large seasonal variations for nearly all the variables (Figure 6). The average seasonal amplitudes, and their percentage changes are: CTDTEMP (13.9 °C), CTDSAL (2.3), TALK (112 μmol kg^{-1}, 5%), DIC (126 μmol kg^{-1}, 6%), fCO$_2$ (170 μatm, 39%), [CO$_3^{2-}$] (61 μmol kg^{-1}, 45%), pH (0.16), aragonite saturation state (0.99, 47%), calcite saturation state (1.47, 45%). Note the "seasonal amplitudes" here represent the sum of effects of all changes including changes from freshwater input, mixing, upwelling, warming and cooling, biological cycling, and diurnal cycling within a season.

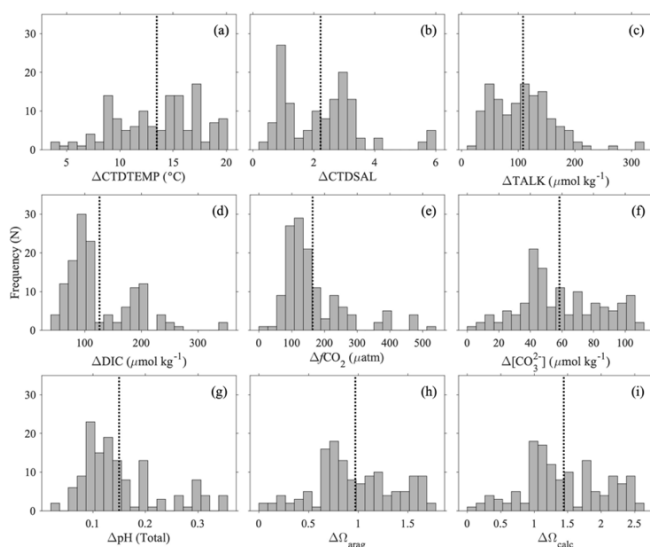
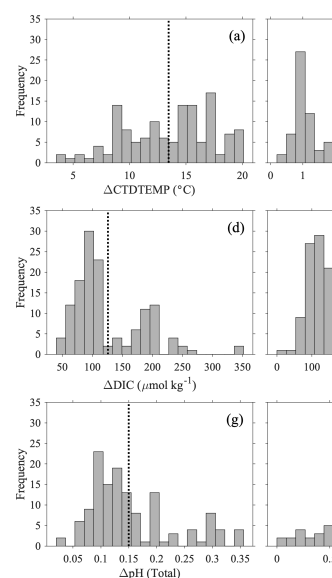



Figure 6. Seasonal amplitudes (maximum minus minimum values within a group of close by stations) of (a) temperature (CTDTEMP), (b) salinity (CTDSAL), (c) total alkalinity (TA), (d) dissolved inorganic carbon (DIC), (e) fugacity of carbon dioxide (fCO$_2$), (f) carbonate ([CO$_3^{2-}$], (g) pH on the Total Scale, (h) aragonite saturation state (Ω_{arag}), and (i) calcite saturation state (Ω_{calc}) in the surface water. The Y-axis "Frequency (N)" refers to the number of group of stations. The dotted lines show the average value of the variabilities. This analysis is based on groups of profiles that are within 1 km apart from each other.

To present a rough estimate of the measurement uncertainties of these variables, a similar approach was used to group deep water stations with a maximum sampling depth of >1500 m. Due to the scarcity of deep-water stations, a radius of 10 km and 200 m depth difference were used to find the comparison pairs. This analysis is limited to certain cruises with deep water sampling (~5% of the data) only, thus the uncertainty estimates only hold true for these "reference" cruises, mostly with a cruise flag of A (Table 3). They do not apply to the rest of the cruises. Results show that the DIC and TA uncertainties (0.1% and 0.2%, respectively) are about the same as previously reported by the GLODAPv2 group (Figure 7, Table 7) by this metric. Some variables like Nitrite and Ammonium show uncertainties as large as ~70% with this metric due, primarily, to the low average values of these measurements at depth. The average CTDTEMP precision of 0.06 °C is significantly higher than that of 0.01 °C as previously reported for the GLODAPv2 (Olsen et al., 2020). The measurement uncertainties could be overestimated, because this analysis includes natural gradients due to the large radius and depth differences, as well as any temporal changes within the 1 to 10 years (average 6 years) period.

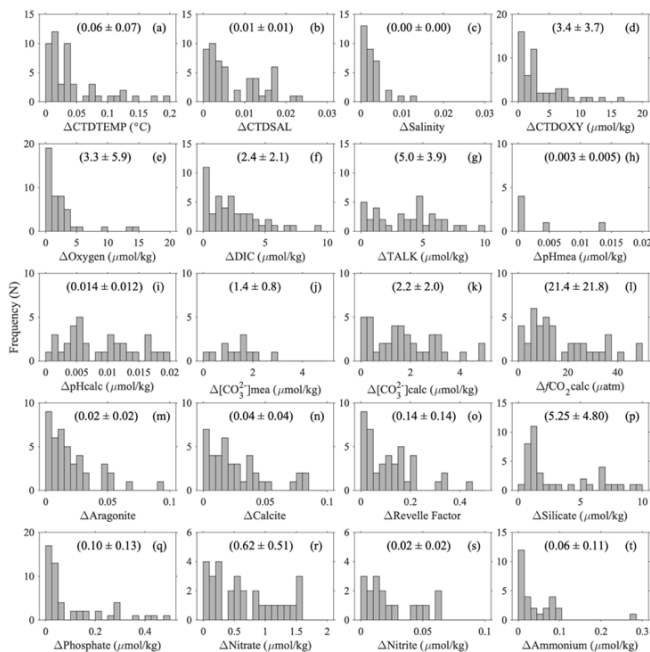
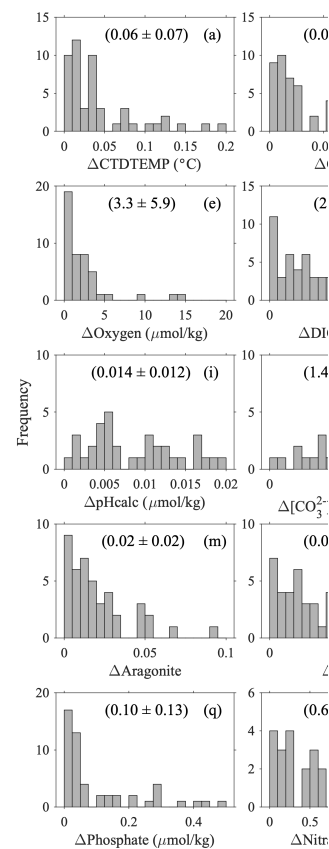



Figure 7. Uncertainties of some parameters based on deep water comparison analyses: (a) Temperature (ITS-90) measured with CTD sensors, (b) Salinity (PSS78) measured with CTD sensors, (c) discrete salinity (PSS78), (d) Dissolved oxygen measured with CTD sensors, (e) Dissolved oxygen concentration measured with Winkler titration, (f) Dissolved inorganic carbon concentration, (g) Total alkalinity, (h) pH on Total Scale measured with spectrophotometers, (i) pH on Total Scale calculated from DIC, TA, and other parameters, (j) Carbonate ion concentration measured with spectrophotometers , (k-o) Carbonate ion concentration, fugacity of carbon dioxide, aragonite saturation state, calcite saturation state, and Revelle Factor calculated from DIC, TA and other parameters, (p-t) Silicate, Phosphate, Nitrate, Nitrite, and Ammonium concentrations. **The Y-axis "Frequency (N)" refers to the number of group of stations.** The values inside the parentheses are (mean values ± standard deviations).

Table 7. Uncertainties of some variables based on an analysis that groups deep-water stations (>1000 m sampling depth) within 10 km radius, and 200 m depth difference. Stdev is the short for standard deviation. Numbers in parenthesis are expected errors based on propagating uncertainties in carbonate system calculations using the CO2SYS companion errors.m program. Refer to Table 1 for their full names and respective units.

Abbreviation	Units	Mean ± stdev	Percentage	Number of pairs
CTDTEMP_ITS90	°C	0.06 ± 0.07	-	54
CTDSAL_PSS78	-	0.007 ± 0.007	0.02%	53
Salinity_PSS78	-	0.003 ± 0.003	0.01%	33
CTDOXY	µmol kg⁻¹	3.4 ± 3.7	4%	51
Oxygen	µmol kg⁻¹	3.3 ± 5.9	3%	47
DIC	µmol kg⁻¹	2.4 ± 2.1	0.1%	48
TALK	µmol kg⁻¹	5.0 ± 3 .9	0.2%	45
pH_TS_insitu_measured	-	0.003 ± 0.005		6
pH_TS_insitu_calculated	-	0.01 ± 0.01 (0.02 ± 0.01)		44
Carbonate_insitu_measured	µmol kg⁻¹	1.4 ± 0.8	2%	12
Carbonate_insitu_calculated	µmol kg⁻¹	2.2 ± 2.0 (4.1 ± 0.7)	3%	44
fCO₂_insitu_measured	µatm	-	-	
fCO₂_insitu_calculated	µatm	21 ± 22 (30 ± 16)	3%	44
Aragonite	-	0.02 ± 0.02	3%	44
Calcite	-	0.04 ± 0.04	3%	44
Revelle_Factor	-	0.14 ± 0.14	1%	44
Silicate	µmol kg⁻¹	5.3 ± 4.4	5%	50
Phosphate	µmol kg⁻¹	0.10 ± 0.13	5%	51
Nitrate	µmol kg⁻¹	0.6 ± 0.5	2%	29
Nitrite	µmol kg⁻¹	0.02 ± 0.02	69%	17
Ammonium	µmol kg⁻¹	0.06 ± 0.11	72%	29

For aragonite and calcite saturation states, uncertainty comes primarily from the use of an empirical equation to approximate the real-world apparent solubility product (K_{sp}'). Despite the 3% number shown in Table 7, the real uncertainty of aragonite and calcite saturation states is likely >5% (Mucci, 1983; Jiang et al., 2015a; Orr et al., 2018). Best practices for oceanic carbonate system calculations have been recommending the dissociation constants of Lueker et al., (2000) (Dickson et al.,

Deleted: 6

Deleted: Refer to Table 2 for their

Inserted Cells

Deleted: 0.04%

Deleted: 0.2%

Deleted: -

Inserted Cells

Inserted Cells

Deleted: their

815   2007). However, a recent study finds that in colder regions, where water temperature is < 8 °C, the constants of Lueker et al. (2000) may underestimate fCO$_2$ and overestimate pH and [CO$_3^{2-}$], meaning that cold ocean regions are more undersaturated than expected with respect to calcium carbonate mineral (CaCO$_3$) saturation states (Sulpis et al., 2020). This applies to many Alaska coast stations. In brackish water (salinity < 20), the relative uncertainty in carbonate ion concentration is worse than that in open ocean water (Dickson et al., 2007; Orr et al., 2018). In addition, due to the way calcium concentration is derived

820   in the CO2SYS (Riley and Tongudai, 1967; Millero, 1995), the calculated saturation states could suffer from uncertainties up to 12% for not directly measuring the calcium concentration in certain very low-salinity regions (Beckwith et al., 2019; Dillon et al., 2020).

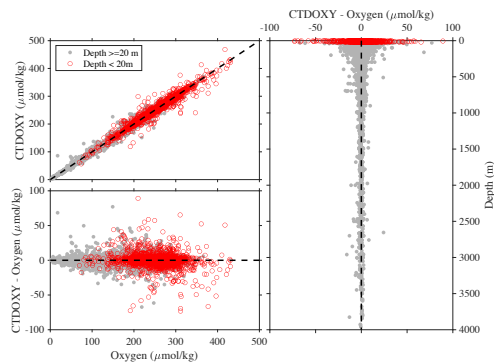



Figure 8. Comparison plots of dissolved oxygen measured from sensors mounted on CTD (CTDOXY) and dissolved
825   **oxygen that is measured from Winkler titration.**

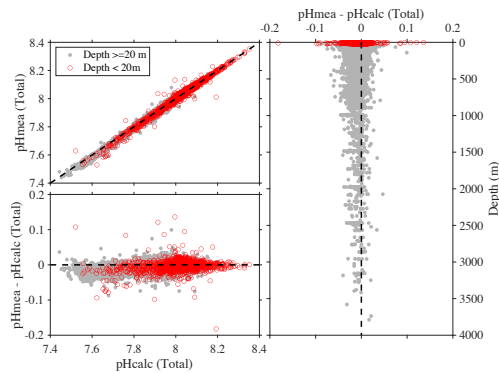



Figure 9. Comparison plots of in-situ pH (Total scale) that is measured using spectrophotometers (pHmea) and in-situ pH (Total scale) that is calculated from DIC and TALK and other parameters (pHcalc).

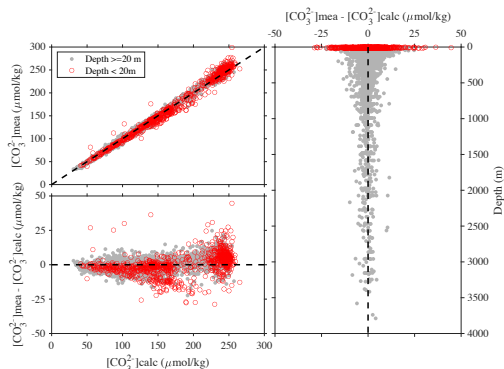



Figure 10. Comparison plots of in-situ carbonate ion concentration that is measured using spectrophotometers ([CO$_3^{2-}$]mea) and that is calculated from DIC, TALK and other parameters ([CO$_3^{2-}$]calc).

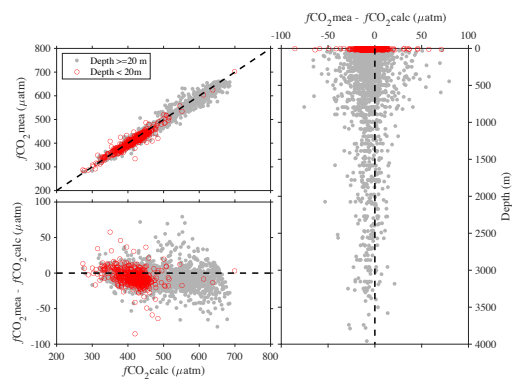



Figure 11. Comparison plots of fugacity of carbon dioxide (fCO$_2$mea) that is measured from discrete bottle samples and fCO$_2$ that is calculated from DIC, TALK and other parameters (fCO$_2$calc).

Note the above uncertainty analyses are based on deep water stations only, and these data are usually collected from cruises with a Cruise_flag of A or B (Table 3). The uncertainties of data points from cruises with Cruise_flags of C, and D are

 expected to be much larger. Internal consistency checks of measured versus calculated values and validation checks of

values measured using different methods show that differences increase quickly towards the surface (Figure 8-11). Some apparent "outliers" end up being in surface samples where the Niskin vs. CTD values are offset due to highly stratified surface conditions in the coastal ocean. We contend that these Winkler and CTD values are likely "good" data from the measurement point of view, so, for such instances, the QC flags are kept as "2", despite their poor internal consistency.

7 Data availability

The Coastal Ocean Data Analysis Product for North America (CODAP-NA) is available as a merged data product in the formats of Excel, CSV, MATLAB, and NetCDF [doi:10.25921/531n-c230, NCEI Accession: 0219960], and can be accessed with the link: https://www.ncei.noaa.gov/data/oceans/ncei/ocads/metadata/0219960.html (Jiang et al., 2021). An Excel spreadsheet listing all of the QC related changes is also included as part of the data package. The original cruise data files have also been updated with data providers' consent and summarized in a table with the link: https://www.ncei.noaa.gov/access/ocean-acidification-data-stewardship-oads/synthesis/NAcruises.html.

8 Summary and conclusions

In this study, we relied on consistency checks performed in direct collaboration with the data providers who originally collected and measured the samples to QC and synthesize two decades of discrete measurements of inorganic carbon system parameters, oxygen, and nutrient chemistry data from North America's coastal oceans. The generated data product is called Coastal Ocean Data Analysis Product for North America (CODAP-NA). It is composed of 3,391 oceanographic profiles from 61 research cruises covering all continental shelves in North America (West Coast, East Coast, Gulf of Mexico, and Alaska coast) from December 6, 2003 to November 22, 2018.

It is strongly recommended to measure a third carbon-related variable for consistency check purposes. The large majority of coastal OA cruises have already measured DIC and TALK, with a lot of them also measuring pH using high-precision spectrophotometric methods. Recently, laboratories have increasingly begun to include carbonate ion concentration ($[CO_3^{2-}]$) as an additional measurable parameter of the seawater CO_2 system (Byrne and Yao, 2008; Sharp and Byrne, 2019). Uncertainty analyses suggest that cross-over adjustments could be applied to future coastal data QC. All major coastal cruises in the future are recommended to take deep water samples (>1500 m) when feasible, ideally at agreed-upon reference stations for QC purposes.

Quality control of coastal data is an ongoing challenge that is not fully resolved by this effort, but CODAP-NA provides a foothold for future efforts toward continuously updating CODAP-NA as an internally-consistent data product for the coastal environment. Perhaps more significantly, the CODAP-NA product greatly improves the findability, accessibility, interoperability, and reusability of these data sets. Findability is improved with this manuscript highlighting the data sets, accessibility is improved through data ingestion of the cruises in Table 2 into a coherent data product, interoperability is

improved by providing the product in multiple machine-readable formats, and reusability is improved by assigning a static DOI for this initial version of the product.

Author contribution:

All authors contributed to the writing of the paper. L-QJ coordinated with the overall effort, designed and built the tools for the QC process, participated in the QC effort, and prepared the first draft of the paper. RAF, RW, DG, LB, SA, BRC, DP, and CF contributed data to the data product, helped develop the QC strategy, provided comments on the QC tools, participated in all steps of the QC effort, and provided guidance to L-QJ on this overall effort. RW minted the new cruise flags that were used by this data product. JH, CM, NM, JS, SS, and Y-YX (ranked alphabetically based on their last names) participated in the QC process. AK provided data management support to this product development. RHB, W-JC, JC, GCJ, BH, CL, JM, and JS (ranked alphabetically based on their last names) contributed data to this data product. DWT and his group measured nutrients data for the ECOA2 cruise and all of the Northeast Fisheries Science Center (NOAA/NFSC)'s Ecosystem Monitoring Program (EcoMon) cruises.

Competing interest

The authors declare that they have no conflict of interest.

Acknowledgements

Funding for this work comes from the National Oceanic and Atmospheric Administration (NOAA) Ocean Acidification Program (OAP, Project #: OAP 1903-1903). L-QJ is grateful to Are Olsen (University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway), Robert Key (Princeton University, Princeton, New Jersey, USA), Nico Lange and Toste Tanhua (GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany), and Siv Lauvset (NORCE Norwegian Research Centre, Bergen, Norway) for generously sharing their QC experiences. We thank Dr. Jon Hare (NOAA Northeastern Fisheries Science Center) for contributing nutrient data collected from U.S. Northeast coast from 2009 to 2016 (doi:10.7289/v5hq3wv3). This research was carried out in part under the auspices of the Cooperative Institute for Marine and Atmospheric Studies (CIMAS), a Cooperative Institute of the University of Miami and the National Oceanic and Atmospheric Administration, cooperative agreement #NA20OAR4320472. RW, RAF, and BRC thank the National Oceanic and Atmospheric Administration's Global Ocean Monitoring and Observation division for funding the Carbon Data Management and Synthesis Grant (Fund ref. no. 100007298). We thank Nico Langue for answering multiple QC related questions along the way. L-QJ also wants to thank Christopher Sabine (University of Hawaii at Manoa) for the guidance and encouragement from his experiences of working on the 1st version of the Global Ocean Data Analysis Project (GLODAP)

Deleted: LJ

Deleted: LJ

long before this project was funded. We thank Maura Thomas (University of Marine) for analyzing nutrient data from some cruises, and Chris Taylor and Tamara Holzwarth-Davis (Northeast Fisheries Science Center) for help with sampling and data analysis for DIC, TALK, and DO in some cruises. We thank Xuewu (Sherwood) Liu (University of South Florida) for assistance with the original pH data from Cruise HLY1103. We also thank Dr. Jia-Zhong Zhang (NOAA Atlantic Oceanographic and Meteorological Laboratory) and Dr. Calvin W. Mordy (NOAA Pacific Marine Environmental Laboratory) for help answering questions related to the conversion between Nitrate and Nitrate and Nitrite. This is contribution number 5161 from the NOAA Pacific Marine Environmental Laboratory.

References

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., et al.,: A multi-decade record of high-quality fCO$_2$ data in version 3 of the Surface Ocean CO2 Atlas (SOCAT), Earth Syst. Sci. Data, 8, 383–413, https://doi.org/10.5194/essd-8-383-2016, 2016.

Barton, A., Hales, B., Waldbusser, G. G., Langdon, C., and Feely, R. A.: The Pacific oyster, Crassostrea gigas, shows negative correlation to naturally elevated carbon dioxide levels: Implications for near-term ocean acidification effects, Limnol. Oceanogr., 57, 698-710, doi: 10.4319/lo.2012.57.3.0698, 2012.

Barton, A., Waldbusser, G. G., Feely, R. A., Weisberg, S. B., Newton, J. A., et al.,: Impacts of Coastal Acidification on the Pacific Northwest Shellfish Industry and Adaptation Strategies Implemented in Response, Oceanography, 28, 146-159, doi: 10.5670/oceanog.2015.38, 2015.

Beckwith, S. T., Byrne, R. H., and Hallock, P.: Riverine calcium end-members improve coastal saturation state calculations and reveal regionally variable calcification potential, Front. Mar. Sci., 6, doi: 10.3389/fmars.2019.00169, 2019.

Borges, A. V. and Gypens, N.: Carbonate chemistry in the coastal zone responds more strongly to eutrophication than ocean acidification, Limnol. Oceanogr., 55, 346–353, doi:10.4319/lo.2010.55.1.0346, 2010.

Broecker, W.S.: The great ocean conveyor, Oceanography 4, 79–89, https://doi.org/10.5670/oceanog.1991.07, 1991.

Byrne, R. H. and Breland, J. A.: High precision multiwavelength pH determinations in seawater using cresol red, Deep-Sea Res., 36, 803–810, 1989.

Byrne, R. H. and Yao, W.: Procedures for measurement of carbonate ion concentrations in seawater by direct spectrophotometric observations of Pb(II) complexation, Mar. Chem., 112, 128-135, doi: 10.1016/j.marchem.2008.07.009, 2008.

Cai, W.-J., Guo, X., Chen, C.-T. A., Dai, M., Zhang, L., Zhai, W., Lohrenz, S. E., Yin, K., Harrison, P. J., Wang, Y.: A comparative overview of weathering intensity and HCO$_3^-$ flux in the world's major rivers with emphasis on the Changjiang, Huanghe, Zhujiang (Pearl) and Mississippi Rivers, Cont. Shelf Res., 28, 1538-1549, doi:10.1016/j.csr.2007.10.014, 2008.

Cai, W.-J., Hu, X., Huang, W.-J., Murrell, M. C., Lehrter, J. C., Lohrenz, S. E., Chou, W.-C., Zhai, W., Hollibaugh, J. T., Wang, Y., Zhao, P., Guo, X., Gundersen, K., Dai, M., and Gong, G.-C.: Acidification of subsurface coastal waters enhanced by eutrophication, Nature Geoscience, 4, 766-770, 10.1038/ngeo1297, 2011.

Cai, W.-J., Xu, Y.-Y., Feely, R. A., Wanninkhof, R., Jönsson, B., Alin, S. R., Barbero, L., Cross, J. N., Azetsu-Scott, K., Fassbender, A. J., Carter, B. R., Jiang, L.-Q., Pepin, P., Chen, B., Hussain, N., Reimer, J. J., Xue, L., Salisbury, J. E., Hernández-Ayón, J. M., Langdon, C., Li, Q., Sutton, A. J., Chen, C.-T. A., and Gledhill, D. K.: Controls on surface water carbonate chemistry along North American ocean margins, Nature Communications, 11, 2691, 10.1038/s41467-020-16530-z, 2020.

Carter, B. R., Feely, R. A., Williams, N. L., Dickson, A. G., and Fong, M. B., and Takeshita, Y.: Updated methods for global locally interpolated estimation of alkalinity, pH, and nitrate, Limnol. Oceanogr. – Methods, 16, 119-131,

955   https://doi.org/10.1002/lom3.10232, 2018.

Cicin-Sain, B., Bernal, P., Vandeweerd, V., Belfiore, S., and Goldstein, K.: A Guide to Oceans, Coasts, and Islands at the World Summit on Sustainable Development, Center for the Study of Marine Policy, Newark, Delaware, August 2002.

Clayton, T. D. and Byrne, R. H.: Spectrophotometric seawater pH measurements: total hydrogen ion concentration scale calibration of m-cresol purple and at-sea results, Deep. Sea Res., 40, 2115-2129, 1993.

960 Cooley, S. R. and Doney, S. C.: Anticipating ocean acidification's economic consequences for commercial fisheries Environmental Research Letters. Environ. Res. Lett. 4, doi:10.1088/1748-9326/4/2/02400, 2009.

Dickson, A. G.: Standard potential of the reaction: $AgCl(s) + 1/2\ H_2(g) = Ag(s) + HCl(aq)$, and the standard acidity constant of the ion HSO_4^- in synthetic seawater from 273.15 to 318.15K, J. Chem. Thermodyn., 22, 113–127, doi: 10.1016/0021-9614(90)90074-z, 1990.

965 Dickson, A. G.: The measurement of seawater pH, Mar. Chem., 44, 131–142, 1993.

Dickson, A. G., Sabine, C. L., and Christian, J. R.: Guide to best practices for ocean CO_2 measurement, Sidney, British Columbia, North Pacific Marine Science Organization, PICES Special Publication 3, 191 pp., http://hdl.handle.net/11329/249, 2007.

Dillon, W. D. N., Dillingham, P. W., Currie, K. I., McGraw, C. M.: Inclusion of uncertainty in the calcium-salinity

970   relationship improves estimates of ocean acidification monitoring data quality, Mar. Chem., 226, 103872, doi: 10.1016/j.marchem.2020.103872, 2020.

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean acidification: the other CO_2 problem, Annu. Rev. Mar. Sci., 1, 169–192, 2009.

Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., and Millero, F. J.: Impact of anthropogenic CO_2

975   on the $CaCO_3$ system in the oceans, Science, 305, 362-366, 2004.

Feely, R.A., Sabine, C. L., Hernandez-Ayon, J.M., Ianson, D., Hales, B.: Evidence for upwelling of corrosive "acidified" water onto the Continental Shelf. Science, 320, doi:10.1126/science.1155676, 1490–1492, 2008.

Feely, R. A., Alin, S. R., Carter, B., Bednaršek, N., Hales, B., Chan, F., Hill, T. M., Gaylord, B., Sanford, E., Byrne, R. H., Sabine, C. L., Greeley, D., Juranek, L.: Chemical and biological impacts of ocean acidification along the west coast of

980   North America, Estuarine, Coastal and Shelf Science, 183, Part A, 260-270, doi:10.1016/j.ecss.2016.08.043, 2016.

Feely, R. A., Okazaki, R. R., Cai, W.-J., Bednaršek, N., Alin, S. R., Byrne, R.H., and Fassbender, A.: The combined effects of acidification and hypoxia on pH and aragonite saturation in the coastal waters of the Californian Current Ecosystem and the northern Gulf of Mexico. Cont. Shelf Res., 152, 50–60, doi: 10.1016/j.csr.2017.11.002, 2018.

Garcia, H. E., and Gordon, L. I.: Oxygen solubility in seawater: Better fitting equations, Limnol. Oceanogr., 37, 1307-1312,

985   https://doi.org/10.4319/lo.1992.37.6.1307, 1992.

Gattuso, J.-P. and Hansson, L.: Ocean acidification, Oxford University Press, Oxford, 2011.

990 Gomez, F. A., Wanninkhof, R., Barbero, L., Lee, S. K., & Hernandez Jr, F. J.: Seasonal patterns of surface inorganic carbon system variables in the Gulf of Mexico inferred from a regional high-resolution ocean biogeochemical model, Biogeosciences, 17, 1685-1700, doi:10.5194/bg-17-1685-2020, 2020.

Gruber N., Clement D., Carter, B. R., et al.,: The oceanic sink for anthropogenic CO_2 from 1994 to 2007, Science, 363, 1193-1199, doi: 10.1126/science.aau5153, 2019.

995 Hales, B., Takahashi, T., and Bandstra, L.: Atmospheric CO_2 uptake by a coastal upwelling system, Global Biogeochemical Cycles, 19, GB1009, doi:10.1029/2004GB002295, 2005.

Hales, B., Cai, Wei-Jun, Mitchell, B. G., Sabine, C. L., and Schofield, O.: North American Continental Margins - Report of the North American Continental Margins Working Group for the U.S. Carbon Cycle Scientific Steering Group and Interagency Working Group, https://data.globalchange.gov/assets/48/3f/48c42b8c11bb5e3f5e442a72a7ae/north-american-continental-margins.pdf, 2008.

000

Hood, E. M., Sabine, C. L., and Sloyan, B. M. (Eds.).: The GO-SHIP Repeat Hydrography Manual: A Collection of Expert Reports and Guidelines. IOCCP Report Number 14. Available online at http://www.go-ship.org/HydroMan.html.: ICPO Publication Series number 134.Hugo, G.: Future demographic change and its interactions with migration and climate change, Global Environmental Change, 21, Supplement 1, S21-S33, 2011.

005 Hunt, C. W., Salisbury, J. E., and Vandemark, D.: Contribution of non-carbonate anions to total alkalinity and overestimation of pCO2 in New England and New Brunswick rivers, Biogeosciences, 8, 3069-3076, doi: 10.5194/bg-8-3069-2011, 2011.

Hydes, D. J., and Hill, N.C.: Determination of nitrate in seawater: Nitrate to nitrite reduction with copper-cadmium alloy, Estuarine, Coastal and Shelf Science, 21, 127-130, 1985.

010 IOC, SCOR, and IAPSO: The international thermodynamic equation of seawater - 2010: Calculation and use of thermodynamic properties. Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, UNESCO (English), 196 pp, 2010.

IPCC: Workshop report of the Intergovernmental Panel on Climate Change (IPCC) Workshop on impacts of ocean acidification on marine biology and ecosystems, Field, C. B., Barros, V., Stocker, T. F., Qin, D., Mach, K. J., Plattner, G.-K., Mastrandrea, M. D., Tignor, M., and Ebi, K. L. (Eds), IPCC Working Group II Technical Support Unit, Carnegie Institution, Stanford, California, USA, 164 pp, 2011.

015

Jiang, L.-Q., Cai, W.-J., Feely, R. A., Wang, Y., Guo, X., Gledhill, D. K., Hu, X., Arzayus, F., Chen, F., Hartmann, J. and, Zhang, L.: Carbonate mineral saturation states along the U.S. East Coast, Limnol. Oceanogr., 55, 2424-2432, 2010.

Jiang L.-Q., Feely, R. A., Carter, B. R., Greeley, D. J., Gledhill, D. K., and Arzayus, K. M.: l. Climatological distribution of aragonite saturation state in the global oceans, Global Biogeochem. Cycles, 29, 1656-1673, 2015a.

020

Jiang, L.-Q., O'Connor, S., Arzayus, K., Kozyr, A., and Parsons, A.: A metadata template for ocean acidification data, Earth Syst. Sci. Data, 8, 1-21, doi: 10.5194/essdd-8-1-2015, 2015b.

Jiang, L.-Q., Carter, B. R., Feely, R. A., Lauvset, S., and Olsen, A.: Surface ocean pH and buffer capacity, past, present and future, Sci. Rep., 9, 18624, doi: 10.1038/s41598-019-55039-4, 2019.

Jiang, L.-Q., Feely, R. A., Wanninkhof, R., Greeley, D., Barbero, L., Alin, S. R., Carter, B. R., Pierrot, D., Featherstone, C., Hooper, J., Melrose, C., Monacci, N., Sharp, J., Shellito, S., Xu, Y.-Y., Kozyr, A., Byrne, R. H., Cai, W.-J., Cross, J., Johnson, G. C., Hales, B., Langdon, C., Mathis, J., Salisbury, J., Townsend, D. W.: Coastal Ocean Data Analysis Product in North America (CODAP-NA, Version 2021) (NCEI Accession 0219960), NOAA National Centers for Environmental Information, Dataset, https://doi.org/10.25921/531n-c230, 2021.

Jiang, L.-Q., Pierrot, D., Sharp, J., and Carter, B.: A suite of internal consistency based primary-level quality control tools for discrete bottle based chemical oceanographic data, Limnol. Oceanogr.-Methods, in prep.

Joyce, T. and Corry, C.: Chapter 4. Hydrographic Data Formats, in Requirements for WOCE Hydrographic Programme Data Reporting, WOCE Hydrographic Programme Office, Woods Hole Oceanographic Institution, Woods Hole, MA, 1994.

Laruelle, G. G., Landschützer, P., Gruber, N., Tison, J., Delille, B., & Regnier, P.: Global high-resolution monthly pCO$_2$ climatology for the coastal ocean derived from neural network interpolatiom. Biogeosciences, 14, 4545-4561, 2017.

Laurent, A., Fennel, K., Cai, W.-J., Huang, W.-J., Barbero, L., & Wanninkhof, R.: Eutrophication-induced acidification of coastal waters in the northern Gulf of Mexico: Insights into origin and processes from a coupled physical-biogeochemical model, Geophys. Res. Lett., 44(2), 946-956, doi:10.1002/2016gl071881, 2017.

Lauvset S. K., Gruber, N., Landschutzer, P., Olsen, A., Tjiputra, J.: Trends and drivers in global surface ocean pH over the past 3 decades. Biogeosciences 12:1285-1298, 2015.

Lauvset, S. K. and Tanhua, T.: A toolbox for secondary quality control on ocean chemistry and hydrographic data, Limnol. & Oceanogr. - Methods, 13, 601-608, https://doi.org/10.1002/lom3.10050, 2015.

Lauvset, S.K., Carter, B. R., Perez, F. F., Jiang, L.-Q., Feely, R. A., Velo, A., Olsen, A.: Processes Driving Global Interior Ocean pH Distribution, Global Biogeochem. Cycles, 34, e2019GB006229, doi: 10.1029/2019GB006229, 2020.

Le Menn, M., Giuliano, Albo, P. A. G., Lago, S., Romeo, R., and Sparasci, F.: The absolute salinity of seawater and its measurands. Metrologia 56, 015005, https://doi.org/10.1088/1681-7575/aaea92, 2018.

Lee, K., Kim, T.-W., Byrne, R. H., Millero, F. J. Feely, R. A., and Liu, Y.-M.: The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans, Geochimica et Cosmochimica Acta, 74, 1801-1811, https://doi.org/10.1016/j.gca.2009.12.027, 2010.

Lewis, E., and Wallace, D. W. R.: Program Developed for CO$_2$ System Calculations, ORNL/CDIAC-105 (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge, Tennessee, 1998.

Lueker, T. J., Dickson, A. G. and Keeling, C. D.: Ocean pCO$_2$ 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, 2000.

Millero, F. J.: Thermodynamics of the carbon dioxide system in the oceans, Geochim. Cosmochim. Acta, 59, 661–677m 1995.

Mucci, A.: The solubility of calcite and aragonite in seawater at various salinities, temperatures, and one atmosphere total pressure, Am. J. Sci., 283, 781–799, 1983.

070   Olsen, A., Key, R. M., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterstrom, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., and Suzuki, T.: 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, 2016.

Olsen, A., Lange, N., Key, R. M., Tanhua, T., Álvarez, M., Becker, S., Bittig, H. C., Carter, B. R., Cotrim da Cunha, L.,
075   Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jones, S. D., Jutterström, S., Karlsen, M. K., Kozyr, A., Lauvset, S. K., Lo Monaco, C., Murata, A., Pérez, F. F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Telszewski, M., Tilbrook, B., Velo, A., and Wanninkhof, R.: GLODAPv2.2019 – an update of GLODAPv2, Earth Syst. Sci. Data, 11, 1437–1461, https://doi.org/10.5194/essd-11-1437-2019, 2019.

Olsen, A., Lange, N., Key, R. M., Tanhua, T., Bittig, H. C., Kozyr, A., Àlvarez, M., Azetsu-Scott, K., Becker, S., Brown, P.
080   J., Carter, B. R., Cotrim da Cunha, L., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jutterström, S., Landa, C. S., Lauvset, S. K., Michaelis, P., Murata, A., Pérez, F. F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Tilbrook, B., Velo, A., Wanninkhof, R., and Woosley, R. J.: GLODAPv2.2020 – the second update of GLODAPv2, Earth Syst. Sci. Data, doi:10.5194/essd-2020-165, 2020.

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., et al.,: Anthropogenic ocean acidification over the twenty-first
085   century and its impact on calcifying organisms, Nature, 437, 681–686, 2005.

Orr, J. C., Epitalon, J.-M., Dickson, A. G., Gattuso, J.-P.: Routine uncertainty propagation for the marine carbon dioxide system, 207, 84-107, 2018.

Perez, F. F. and Fraga, F.: Association constant of fluoride and hydrogen ions in seawater, Mar. Chem., 21, 161-168, doi: 10.1016/0304-4203(87)90036-3, 1987.

090   Riley, J. P., and Tongudai, M.: The major cation/chlorinity ratios in sea water, Chemical Geology, 2, 263-269, 1967.

Roobaert, A., Laruelle, G. G., Landschützer, P., Gruber, N., Chou, L., and Regnier, P.: The spatiotemporal dynamics of the sources and sinks of CO_2 in the global coastal ocean. Global Biogeochemical Cycles, 33, 1693-1714, doi:10.1029/2019GB006239, 2019.

Sharp, J. D. and Byrne, R. H.: Carbonate ion concentrations in seawater: spectrophotometric determination at ambient
095   temperatures and evaluation of propagated calculation uncertainties, Mar. Chem., 209, 70-80, https://doi.org/10.1016/j.marchem.2018.12.001, 2019.

Sharp, J.D., Pierrot, D., Humphreys, M.P., (2020). CO2-System-Extd, v3.0.1, MATLAB (MathWorks), http://doi.org/10.5281/zenodo.3952803.

Sherman, K., Aquarone, M.C. and Adams, S.: Sustaining the World's Large Marine Ecosystems. Gland, Switzerland: IUCN. viii+142p., 2009.

Stets, E. G., Kelly, V. J., and Crawford, C. G.: Long-term trends in alkalinity in large rivers of the conterminous US in relation to acidification, agriculture, and hydrologic modification, Sci. Total Environ., 488-489, 280-289, doi:10.1016/j.scitotenv.2014.04.054, 2014.

Sulpis, O., Lauvset, S. K., and Hagens, M.: Current estimates of K_1 and K_2 appear inconsistent with measured CO_2 system parameters in cold oceanic regions, Ocean Sci., 16, 847-862, doi: 10.5194/os-16-847-2020, 2020.

Swift, J. H., and Diggs, S. C.: Description of WHP-Exchange Format for CTD/Hydrographic Data. https://cchdo.github.io/hdo-assets/documentation/WHP_Exchange_Description.pdf, 2008.

Takahashi, T., Sutherland, S. C., and Kozyr, A.: Global Ocean Surface Water Partial Pressure of CO_2 Database (LDEO Database Version 2019): Measurements Performed During 1957-2019 (NCEI Accession 0160492), NOAA National Centers for Environmental Information, Dataset, doi:10.3334/cdiac/otg.ndp088(v2015), 2020.

Tanhua, T., van Heuven, S., Key, R. M., Velo, A., Olsen, A., and Schirnick, C.: Quality control procedures and methods of the CARINA database, Earth Syst. Sci. Data, 2, 35-49, doi:10.5194/essd-2-35-2010, 2010.

Wanninkhof, R. and Thoning, K.: Measurement of fugacity of CO_2 in surface water using continuous and discrete methods, Mar. Chem., 44, 189-204, 1993.

Wanninkhof, R., Barbero, L., Byrne, R., Cai, W.-J., Huang, W.-J., Zhang, J.-Z., Baringer, M., and Langdon, C.: Ocean acidification along the Gulf Coast and East Coast of the USA, Cont. Shelf Res., 98:54-71, doi:10.1016/j.csr.2015.02.008, 2015.

Xue, L., Cai, W.-J., Hu, X., Sabine, C., Jones, S., Sutton, A., Jiang, L-Q., and Reimer, J. J.: Sea surface carbon dioxide at the Georgia time series site (2006-2007): Air-sea flux and controlling processes, Prog. Oceanogr, 140, 14–26, doi:10.1016/j.pocean.2015.09.008, 2016.

105

110

115

120

125

Page 17: [1] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [1] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [1] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [1] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [1] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [1] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [1] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [1] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [2] Formatted	Liqing Jiang	4/18/21 8:34:00 PM

Font: Not Bold

Page 17: [3] Deleted	Liqing Jiang	4/18/21 8:34:00 PM

Page 17: [4] Formatted	Liqing Jiang	4/18/21 8:34:00 PM

Font: Times New Roman, Not Bold

Page 17: [5] Formatted	Liqing Jiang	4/18/21 8:34:00 PM

Font: 10 pt, Bold

Page 17: [6] Inserted Cells	Liqing Jiang	4/18/21 8:34:00 PM

Inserted Cells

Page 17: [7] Inserted Cells	Liqing Jiang	4/18/21 8:34:00 PM

Inserted Cells

Page 17: [8] Inserted Cells	Liqing Jiang	4/18/21 8:34:00 PM

Inserted Cells

Page 17: [9] Inserted Cells	Liqing Jiang	4/18/21 8:34:00 PM

Inserted Cells

Page 17: [10] Inserted Cells	Liqing Jiang	4/18/21 8:34:00 PM

Inserted Cells

Page 17: [11] Inserted Cells	Liqing Jiang	4/18/21 8:34:00 PM

Inserted Cells

Inserted Cells

Page 22: [13] Deleted	Liqing Jiang	4/18/21 8:34:00 PM
Page 22: [13] Deleted	Liqing Jiang	4/18/21 8:34:00 PM
Page 22: [13] Deleted	Liqing Jiang	4/18/21 8:34:00 PM
Page 22: [13] Deleted	Liqing Jiang	4/18/21 8:34:00 PM
Page 22: [13] Deleted	Liqing Jiang	4/18/21 8:34:00 PM
Page 22: [13] Deleted	Liqing Jiang	4/18/21 8:34:00 PM
Page 22: [13] Deleted	Liqing Jiang	4/18/21 8:34:00 PM
Page 22: [13] Deleted	Liqing Jiang	4/18/21 8:34:00 PM