Data Management Workflow and Challenges in Developing a California Ocean Acidification and Hypoxia Data Portal

Marine Lebrec, CeNCOOS DMAC

Axiom DATA SCIENCE

MBARI

CENTRAL & NORTHERN CALIFORNIA OCEAN OBSERVING SYSTEM

SOUTHERN CALIFORNIA COASTAL OCEAN OBSERVING SYSTEM

Sea Grant California

OCEAN PROTECTION COUNCIL

Background

Project goal: Build a centralized information access hub, the California OAH Portal, to serve automated, high-quality and interoperable data and synthesis products

Project objectives

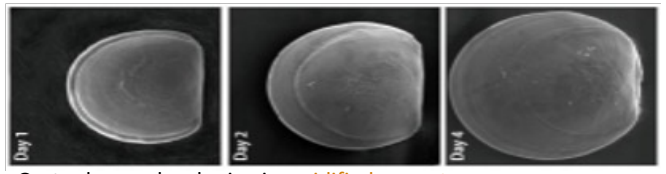
Objective 1. Improve data quality, interoperability, and access.

Objective 2. Streamline data ingestion for enhanced visualization and synthesis.

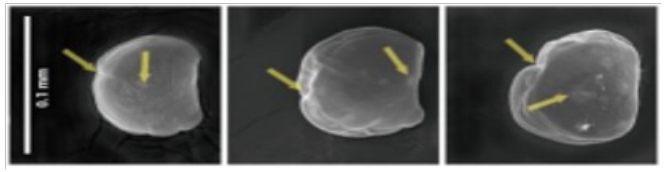
Objective 3. Establish a state-wide information hub for managers and stakeholders.

Objective 4. Develop and automate synthesis products and biological indicators.

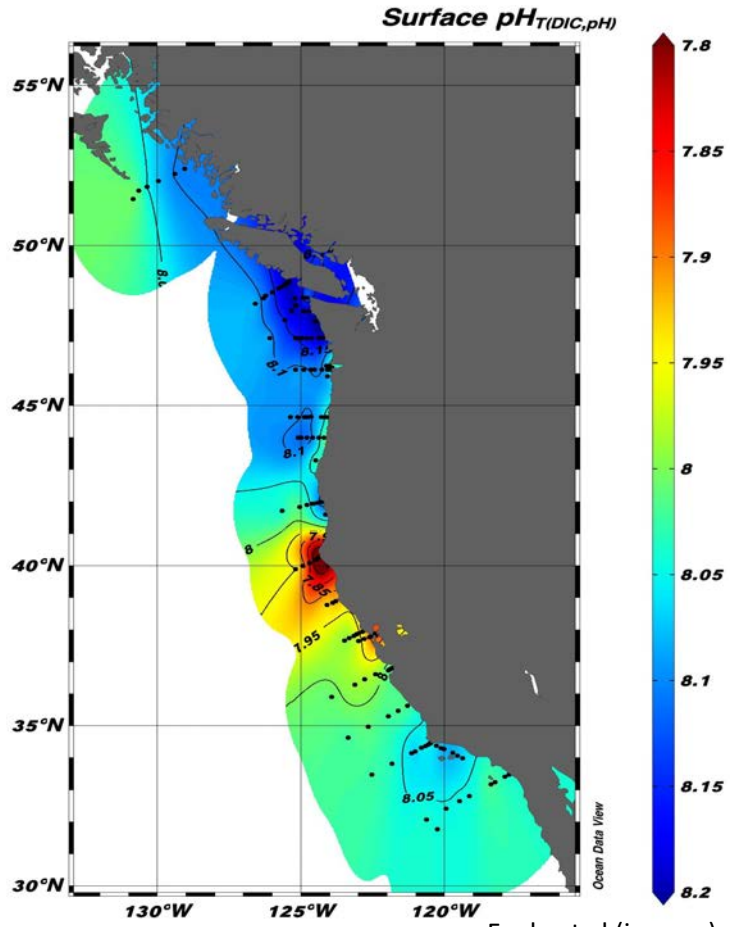
Oyster larvae developing in: normal seawater



Oyster larvae developing in: acidified seawater



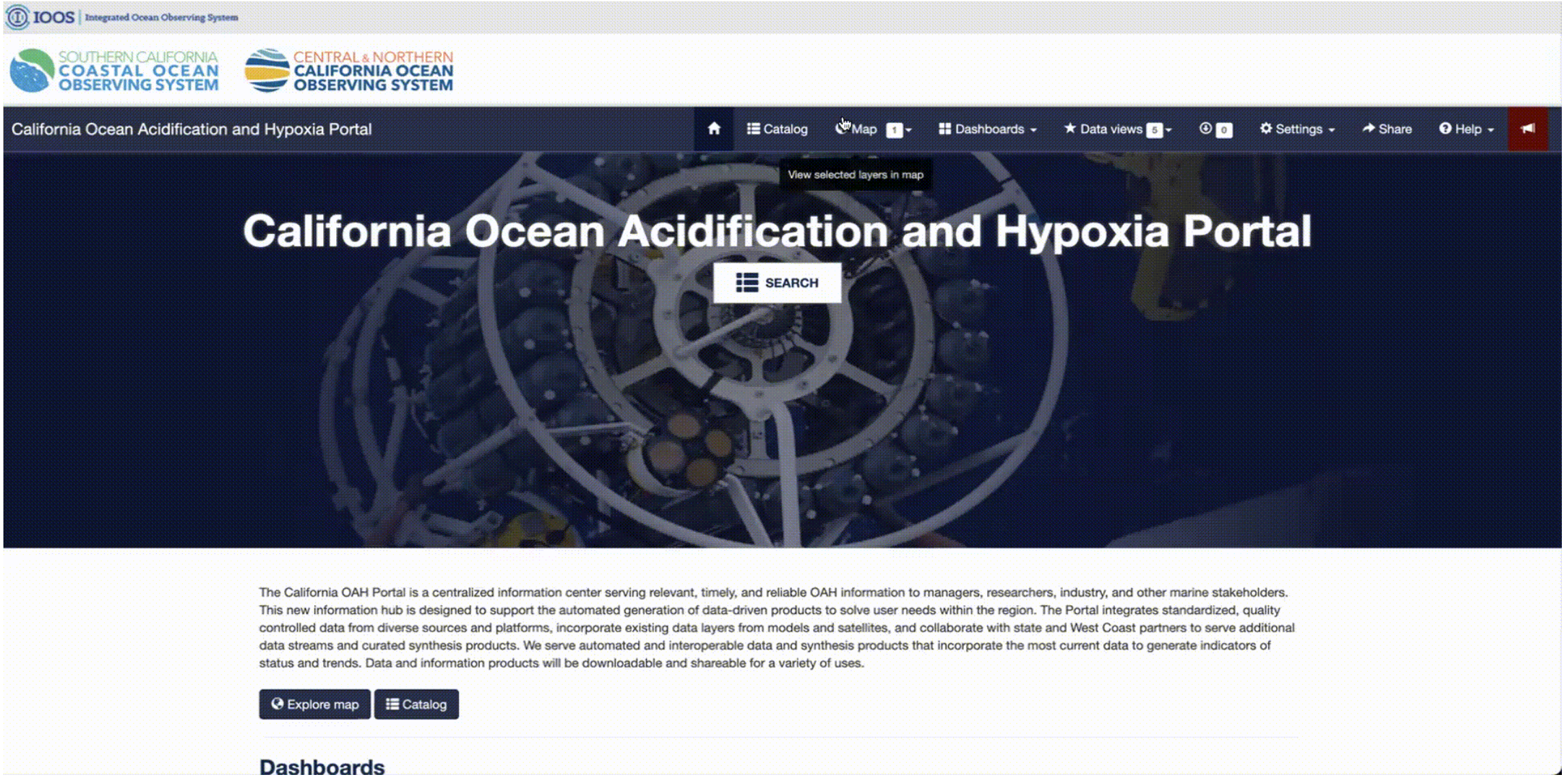
Elizabeth Brunner & George Waldbusser, Oregon State University



Feely et al (in prep)

State-wide OAH data portal: oah.caloos.org

Axiom DATA SCIENCE



Ingesting new datasets & products into the CeNCOOS/Axiom DMAC pipeline

- **NOAA PMEL West Coast Ocean Acidification cruises** via NCEI:
  - CTD data & discrete samples
  - Underway PCO_2
  - Pteropod abundance/dissolution severity
- Co-located environmental data + zooplankton monitoring **(Trinidad Head, N. California**) via ERDDAP
- **MBARI underway PCO_2 time-series** via MySQL db
- Water quality dashboards for shellfish growers (**CeNCOOS shore station network - ERDDAP**)

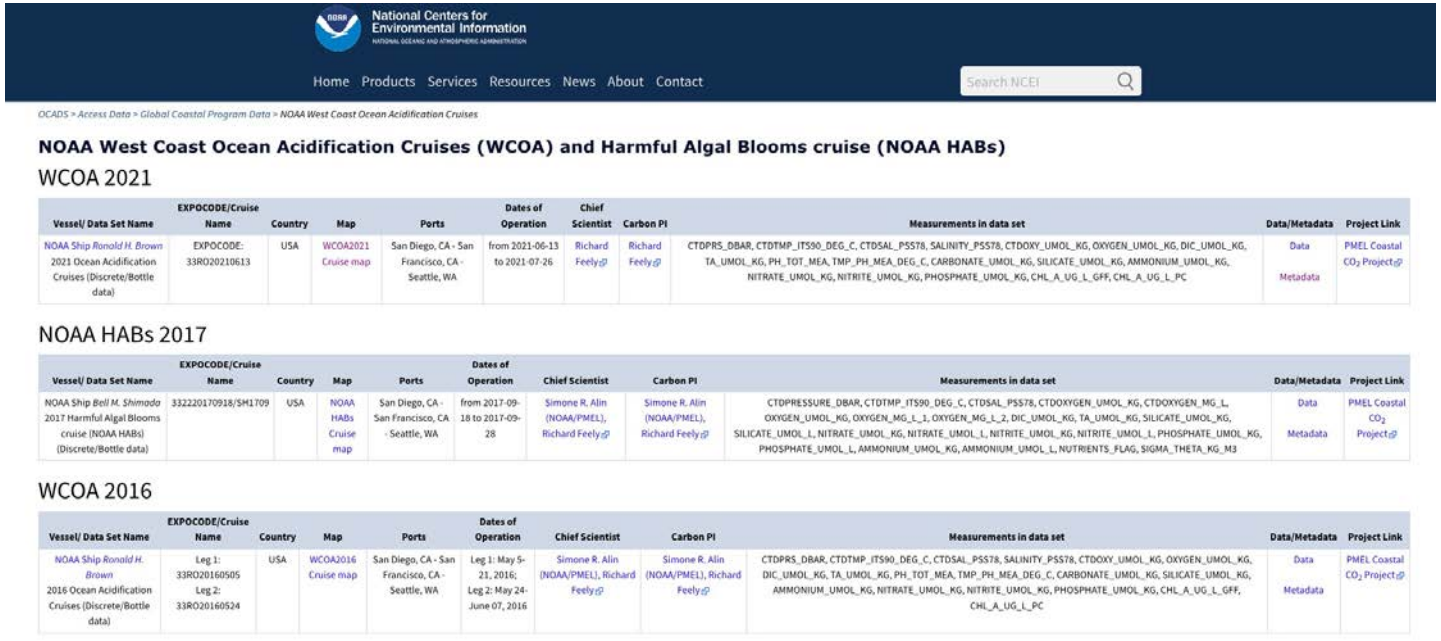
- Data access
- Processing / QC'ing steps
- Challenges, lessons learned

NOAA PMEL West Coast Ocean Acidification cruises: data access

Background

- West-Coast wide cruises to obtain snapshots of key carbon, physical, biogeochemical parameters relating to OA (2007 – present)

- All data (.xlsx) + metadata (HTML or XML) are published on NCEI Ocean Carbon and Acidification Data System (OCADS)



National Centers for Environmental Information

Home Products Services Resources News About Contact

Search NCEI

OCADS > Access Data > Global Coastal Program Data > NOAA West Coast Ocean Acidification Cruises

NOAA West Coast Ocean Acidification Cruises (WCOA) and Harmful Algal Blooms cruise (NOAA HABs)

WCOA 2021

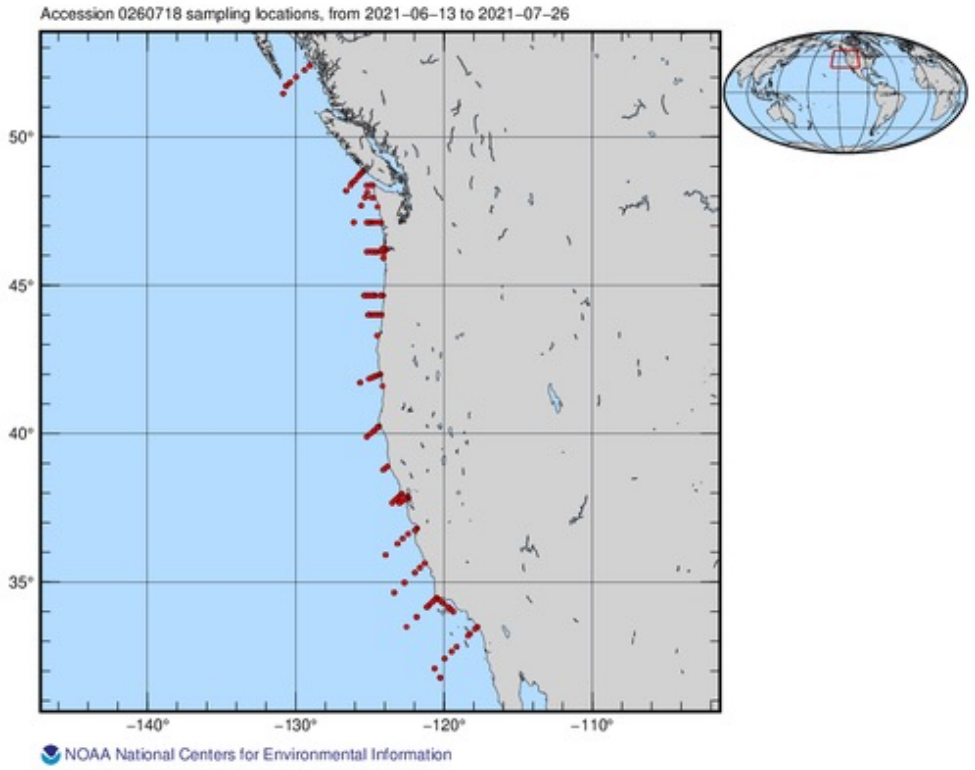
Vessel/ Data Set Name	EXPOCODE/Cruise Name	Country	Map	Ports	Dates of Operation	Chief Scientist	Carbon PI	Measurements in data set	Data/Metadata	Project Link
NOAA Ship *Ronald H. Brown* 2021 Ocean Acidification Cruises (Discrete/Bottle data)	EXPOCODE: 33RO20210613	USA	WCOA2021 Cruise map	San Diego, CA - San Francisco, CA - Seattle, WA	from 2021-06-13 to 2021-07-26	Richard Feely	Richard Feely	CTDPRS_DBAR, CTDTMP_ITS90_DEG_C, CTDSAL_PSS78, SALINITY_PSS78, CTDOXY_UMOL_KG, OXYGEN_UMOL_KG, DIC_UMOL_KG, TA_UMOL_KG, PH_TOT_MEA, TMP_PH_MEA_DEG_C, CARBONATE_UMOL_KG, SILICATE_UMOL_KG, AMMONIUM_UMOL_KG, NITRATE_UMOL_KG, NITRITE_UMOL_KG, PHOSPHATE_UMOL_KG, CHL_A_UG_L_GFF, CHL_A_UG_L_PC	Data Metadata	PMEL Coastal CO_2 Project

NOAA HABs 2017

Vessel/ Data Set Name	EXPOCODE/Cruise Name	Country	Map	Ports	Dates of Operation	Chief Scientist	Carbon PI	Measurements in data set	Data/Metadata	Project Link
NOAA Ship *Bell M. Shimada* 2017 Harmful Algal Blooms cruise (NOAA HABs) (Discrete/Bottle data)	332220170918/SM1709	USA	NOAA HABs Cruise map	San Diego, CA - San Francisco, CA - Seattle, WA	from 2017-09-18 to 2017-09-28	Simone R. Alin (NOAA/PMEL), Richard Feely	Simone R. Alin (NOAA/PMEL), Richard Feely	CTDPRESSURE_DBAR, CTDTMP_ITS90_DEG_C, CTDSAL_PSS78, CTDOXYGEN_UMOL_KG, CTDOXYGEN_MG_L, OXYGEN_UMOL_KG, OXYGEN_MG_L_1, OXYGEN_MG_L_2, DIC_UMOL_KG, TA_UMOL_KG, SILICATE_UMOL_KG, SILICATE_UMOL_L, NITRATE_UMOL_KG, NITRATE_UMOL_L, NITRITE_UMOL_KG, NITRITE_UMOL_L, PHOSPHATE_UMOL_KG, PHOSPHATE_UMOL_L, AMMONIUM_UMOL_KG, AMMONIUM_UMOL_L, NUTRIENTS_FLAG, SIGMA_THETA_KG_M3	Data Metadata	PMEL Coastal CO_2 Project

WCOA 2016

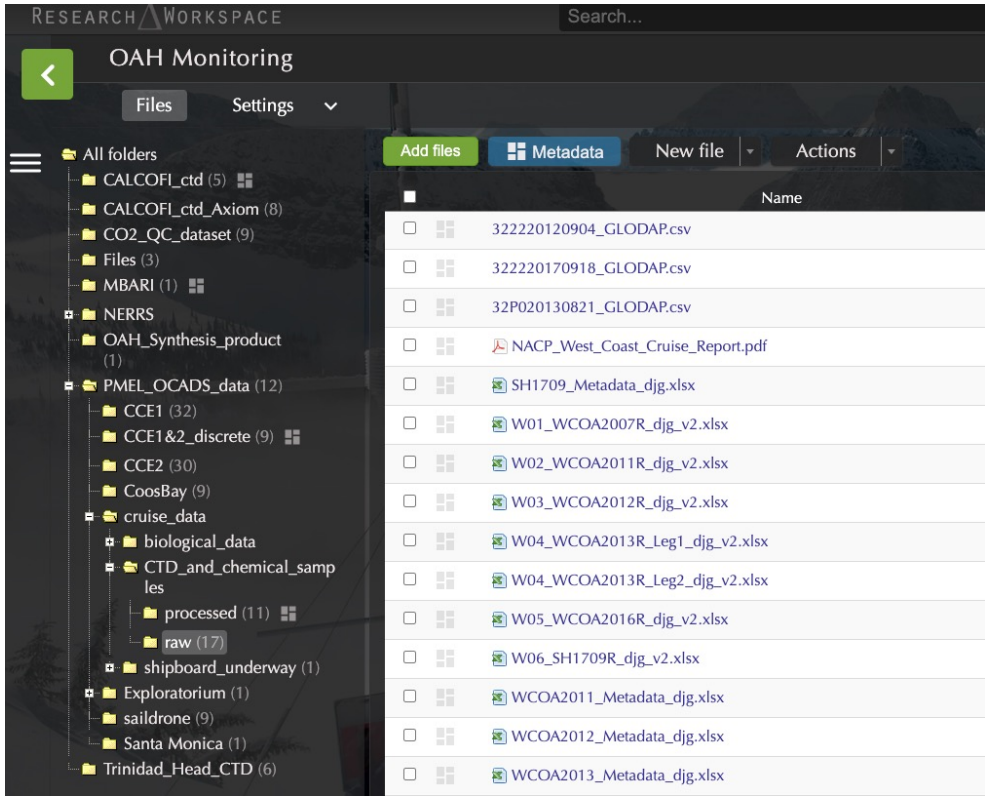
Vessel/ Data Set Name	EXPOCODE/Cruise Name	Country	Map	Ports	Dates of Operation	Chief Scientist	Carbon PI	Measurements in data set	Data/Metadata	Project Link
NOAA Ship *Ronald H. Brown* 2016 Ocean Acidification Cruises (Discrete/Bottle data)	Leg 1: 33RO20160505 Leg 2: 33RO20160524	USA	WCOA2016 Cruise map	San Diego, CA - San Francisco, CA - Seattle, WA	Leg 1: May 5-21, 2016; Leg 2: May 24-June 07, 2016	Simone R. Alin (NOAA/PMEL), Richard Feely	Simone R. Alin (NOAA/PMEL), Richard Feely	CTDPRS_DBAR, CTDTMP_ITS90_DEG_C, CTDSAL_PSS78, SALINITY_PSS78, CTDOXY_UMOL_KG, OXYGEN_UMOL_KG, DIC_UMOL_KG, TA_UMOL_KG, PH_TOT_MEA, TMP_PH_MEA_DEG_C, CARBONATE_UMOL_KG, SILICATE_UMOL_KG, AMMONIUM_UMOL_KG, NITRATE_UMOL_KG, NITRITE_UMOL_KG, PHOSPHATE_UMOL_KG, CHL_A_UG_L_GFF, CHL_A_UG_L_PC	Data Metadata	PMEL Coastal CO_2 Project



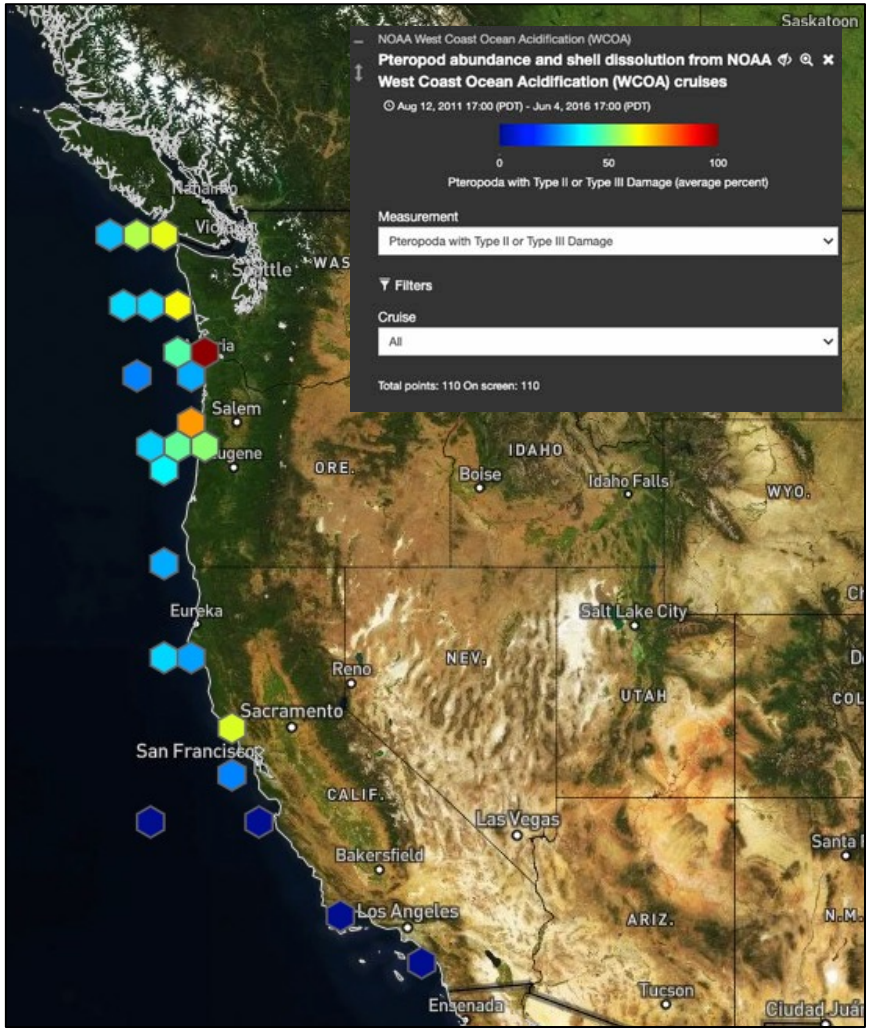
NOAA PMEL West Coast Ocean Acidification cruises: data processing

- Using Axiom's Research Workspace to share notebooks, files, metadata, documentation
- Data cleaning
    - Multiple files for multiple legs per cruise
    - Variable units included as rows in certain cruises but not others
    - Multiple tabs of data
    - Inconsistencies in variable names between cruises; converting to CF format
    - Reformatting time to ISO
- Converting cruise data to netCDF, use IOOS compliance checker

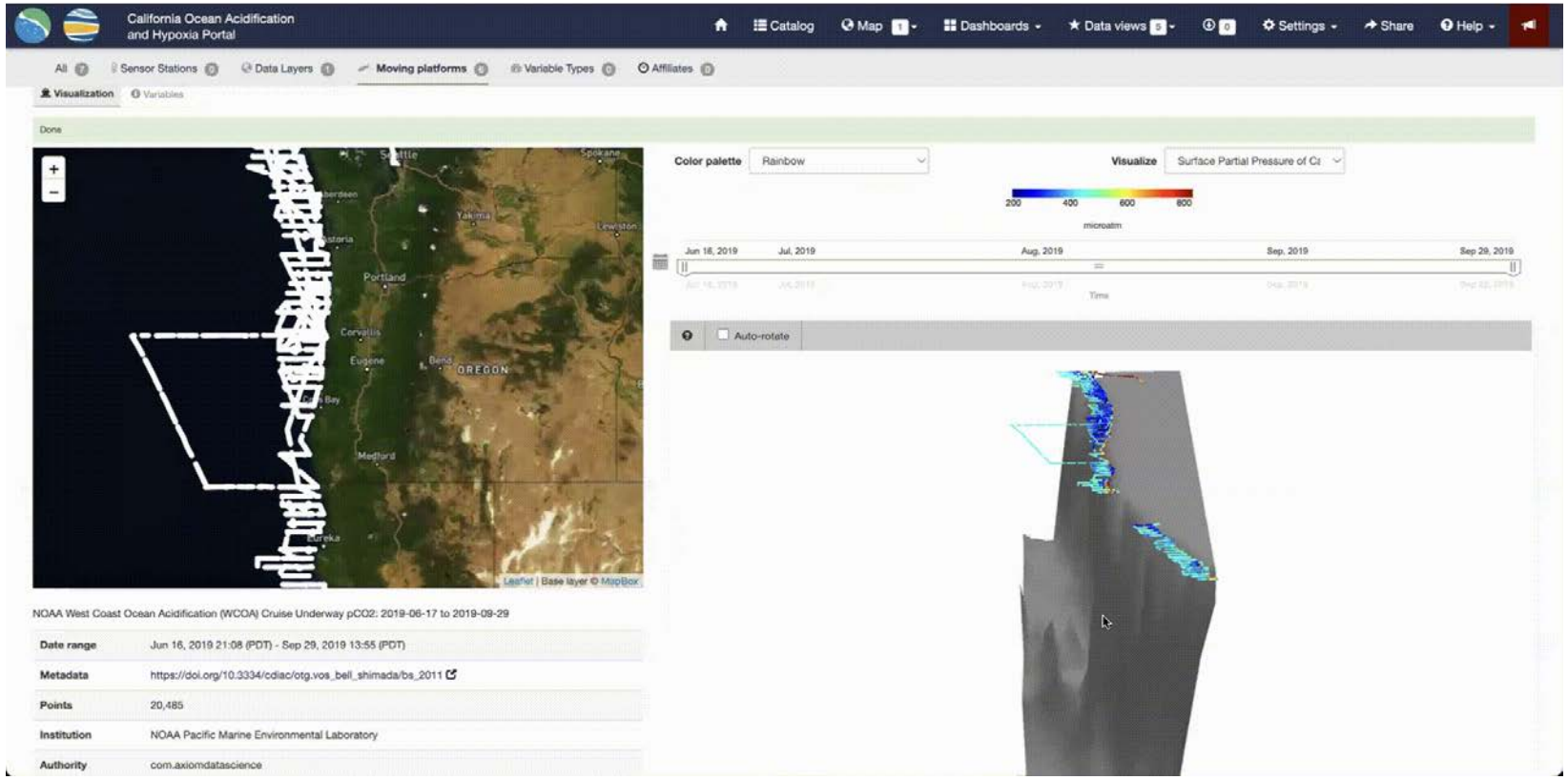
```
new_columns = {

    "LATITUDE_DECIMAL":"latitude",
    "LONGITUDE_DECIMAL":"longitude",
    "DATE_UTC": "time",
    "CTDPRESSURE_DBAR": "depth",
    "Depth": "depth",
    "CTDTEMP_ITS90": "sea_water_temperature",
    "CTDTEMP_ITS90_DEG_C": "sea_water_temperature",
    "CTDTMP_ITS90_DEG_C": "sea_water_temperature",
    "CTDSAL_PSS78":"sea_water_salinity",
    "Oxygen":"mole_concentration_of_dissolved_molecular_oxygen_in_sea_water",
    "OXYGEN_UMOL_KG":"mole_concentration_of_dissolved_molecular_oxygen_in_sea_water",
    "DIC":"moles_of_dissolved_inorganic_carbon_per_unit_mass_in_sea_water",
    "DIC_UMOL_KG":"moles_of_dissolved_inorganic_carbon_per_unit_mass_in_sea_water",
    "TA":"sea_water_alkalinity_expressed_as_mole_equivalent",
    "TA_UMOL_KG":"sea_water_alkalinity_expressed_as_mole_equivalent",
    "Silicate":"mole_concentration_of_silicate_in_sea_water",
    "SILICATE_UMOL_KG":"mole_concentration_of_silicate_in_sea_water",
    "Nitrate":"moles_of_nitrate_per_unit_mass_in_sea_water",
    "NITRATE_UMOL_KG":"moles_of_nitrate_per_unit_mass_in_sea_water",
    "Nitrite":"mole_concentration_of_nitrite_in_sea_water",
    "NITRITE_UMOL_KG":"mole_concentration_of_nitrite_in_sea_water",
    "Phosphate":"mole_concentration_of_phosphate_in_sea_water",
    "PHOSPHATE_UMOL_KG":"mole_concentration_of_phosphate_in_sea_water",
    "PH_TS":"sea_water_ph_reported_on_total_scale",
    "PH_TOT_MEA":"sea_water_ph_reported_on_total_scale",
    "pH_T_measured":"sea_water_ph_reported_on_total_scale",
    "Ammonium":"mole_concentration_of_ammonium_in_sea_water",
    "AMMONIUM_UMOL_KG":"mole_concentration_of_ammonium_in_sea_water",
    "Chl_a":"mass_concentration_of_chlorophyll_a_in_sea_water",
    "CHL_A _UG_L":"mass_concentration_of_chlorophyll_a_in_sea_water",
    "PON_UMOL_KG":"mole_concentration_of_particulate_organic_nitrogen_in_sea_water",
    "POC_UMOL_KG":"mole_concentration_of_particulate_matter_expressed_as_carbon_in_sea_water",
    "POC_PON_RATIO_MOL_MOL":"mole_ratio_of_particulate_organic_carbon_to_particulate_organic_nitrogen_in_sea_water",
    "CARBONATE_UMOL_KG":"mole_concentration_of_carbonate_expressed_as_carbon_in_sea_water",
    "Carbonate_measured":"mole_concentration_of_carbonate_expressed_as_carbon_in_sea_water",
    "Methane_CH4":"mole_concentration_of_methane_in_sea_water"
}
```



NOAA PMEL West Coast Ocean Acidification cruises: data discovery & visualization



Pteropod with type II or type III shell damage

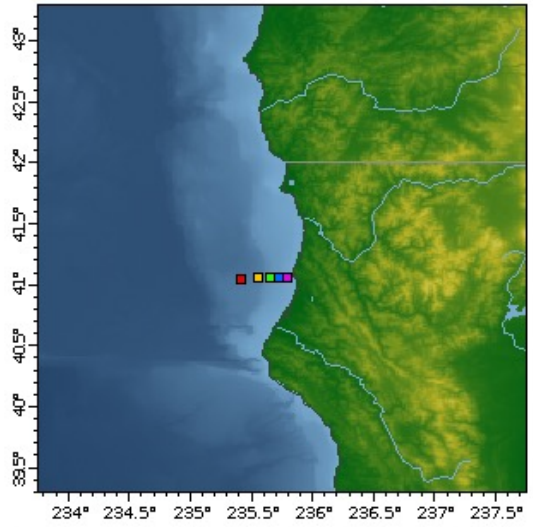


Underway seawater pCO_2 from 2019 WCOA cruise

Co-located biological and chemical datasets: Trinidad Head

Background

Ocean observing cruises since 2007 led by Southwest Fisheries Science Center (SWFSC) and Cal Poly Humboldt (CPH) to characterize seasonal and interannual variability in the plankton ecosystem in coastal waters off northern California.



Data

T, S, Chl-a, DO, pH + QARTOD flagging

Zooplankton samples (abundance, size) @ TH04, TH05

ERDDAP: ERD Projects
Easier access to scientific data

ERDDAP > info > cciea_EI_KRILLEN

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	WMS	Source Data Files	Acces-sible	Title	Sum-mary	FGDC, ISO, Metadata	Back-ground Info	RSS	Institution	Dataset ID
	set	data	graph		files	public	Euphausia pacifica (krill) lengths and biomass Trinidad Line Stations TH04, TH05		M	background	RSS	NOAA SWFSC; Cal P...	cciea_EI_KRILLEN

The Dataset's Variables and Attributes

Row Type	Variable Name	Attribute Name	Data Type	Value
attribute	NC_GLOBAL	cdm_data_type	String	Other
attribute	NC_GLOBAL	contributor_email	String	eric.bjorkstedt at noaa.gov
attribute	NC_GLOBAL	contributor_name	String	Bjorkstedt
attribute	NC_GLOBAL	Conventions	String	COARDS, CF-1.6, ACDD-1.3, NCCSV-1.1
attribute	NC_GLOBAL	defaultGraphQuery	String	time,krill_length_biomass&measurement="Euphausia pacifica (krill) mean length"&.draw=linesAndMarkers
attribute	NC_GLOBAL	history	String	last updated: 2023-01-17 06:00:01
attribute	NC_GLOBAL	infoUrl	String	https://www.fisheries.noaa.gov/west-coast/science-data/ocean-and-ecosystem-observations-trinidad-head-line
attribute	NC_GLOBAL	institution	String	NOAA SWFSC; Cal Poly Humboldt
attribute	NC_GLOBAL	license	String	The data may be used and redistributed for free but is not intended for legal use, since it may contain inaccuracies. Neither the data Contributor, ERD, NOAA, nor the United States Government, nor any of their employees or contractors, makes any warranty, express or implied, including warranties of merchantability and fitness for a particular purpose, or assumes any legal liability for the accuracy, completeness, or usefulness, of this information.
attribute	NC_GLOBAL	sourceUrl	String	(local files)
attribute	NC_GLOBAL	standard_name_vocabulary	String	CF Standard Name Table v70
attribute	NC_GLOBAL	subsetVariables	String	measurement
attribute	NC_GLOBAL	summary	String	Krill (Euphausia pacifica) data were provided by Dr. Eric Bjorkstedt (eric.bjorkstedt@noaa.gov), NMFS/SWFSC and Cal Poly Humboldt, and R. Robertson (roxanne.robertson@noaa.gov), Cooperative Institute for Marine, Earth, and Atmospheric Systems (CIMEAS) at Cal Poly Humboldt. Additional Calculations: Krill body length was measured from the back of the eye to the base of the telson. ;Krill (Euphausia pacifica) data were provided by Dr. Eric Bjorkstedt (eric.bjorkstedt@noaa.gov), NMFS/SWFSC and Cal Poly Humboldt, and R. Robertson (roxanne.robertson@noaa.gov), Cooperative Institute for Marine, Earth, and Atmospheric Systems (CIMEAS) at Cal Poly Humboldt. Additional Calculations: Carbon biomass of krill is calculated from body length measurements following length to weight conversions in Fisher et al., 2020.

ERDDAP
Easier access to scientific data

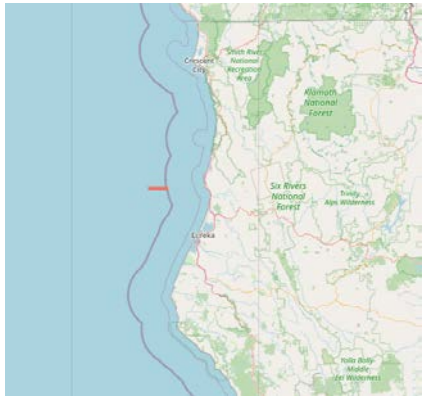
ERDDAP > info > th04-trinidad-head-line

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	WMS	Source Data Files	Title	Sum-mary	FGDC, ISO, Metadata	Back-ground Info	RSS	Institution	Dataset ID
		data	graph		files	TH04, Trinidad Head Line		F I M	background	RSS	Cal Poly Humboldt	th04-trinidad-head-line

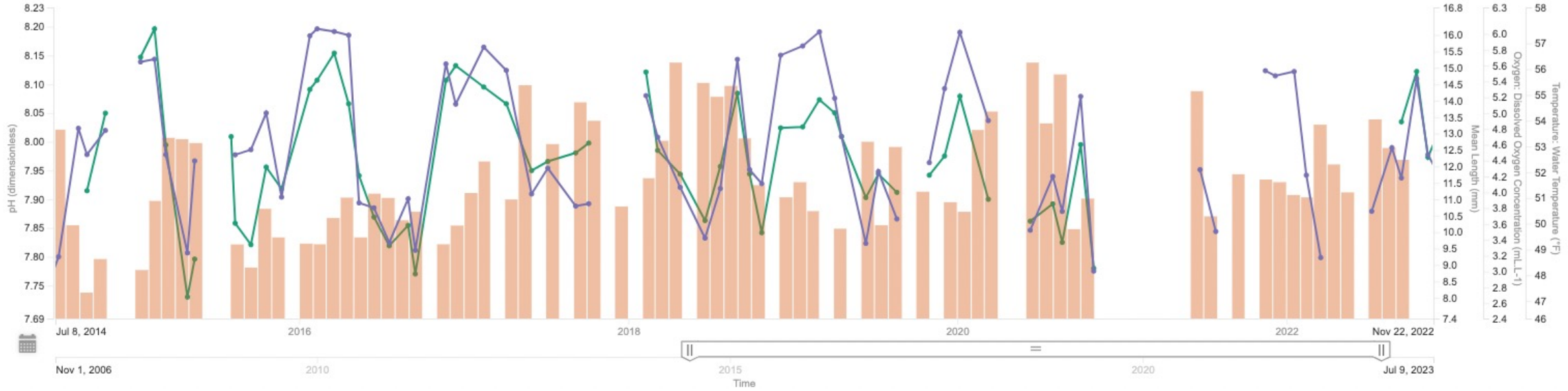
The Dataset's Variables and Attributes

Row Type	Variable Name	Attribute Name	Data Type	
attribute	NC_GLOBAL	_NCProperties	String	version=2,netcdf=4.7.4,hdf5=1.10.6
attribute	NC_GLOBAL	cdm_altitude_proxy	String	z
attribute	NC_GLOBAL	cdm_data_type	String	TimeSeriesProfile
attribute	NC_GLOBAL	cdm_profile_variables	String	time
attribute	NC_GLOBAL	cdm_timeseries_variables	String	station,longitude,latitude
attribute	NC_GLOBAL	contributor_email	String	feedback at axiomdatascience.com
attribute	NC_GLOBAL	contributor_name	String	Axiom Data Science
attribute	NC_GLOBAL	contributor_role	String	processor
attribute	NC_GLOBAL	contributor_role_vocabulary	String	NERC
attribute	NC_GLOBAL	contributor_url	String	https://www.axiomdatascience.com
attribute	NC_GLOBAL	Conventions	String	IOOS-1.2, CF-1.6, ACDD-1.3
attribute	NC_GLOBAL	creator_country	String	USA
attribute	NC_GLOBAL	creator_email	String	None
attribute	NC_GLOBAL	creator_institution	String	Cal Poly Humboldt
attribute	NC_GLOBAL	creator_name	String	Cal Poly Humboldt
attribute	NC_GLOBAL	creator_sector	String	academic
attribute	NC_GLOBAL	creator_type	String	institution
attribute	NC_GLOBAL	creator_url	String	http://cencoos.humboldt.edu/
attribute	NC_GLOBAL	defaultDataQuery	String	sea_water_ph_reported_on_total_scale_qc_agg,sea_water_temperat
attribute	NC_GLOBAL	Easternmost_Easting	double	-124.436691
attribute	NC_GLOBAL	featureType	String	TimeSeriesProfile
attribute	NC_GLOBAL	geospatial_lat_max	double	41.058019

Co-located biological and chemical datasets: data-view comparison chart



- Adult mean biomass of *Euphausia pacifica* (krill)
- Dissolved oxygen concentration @50m
- pH @50m



MBARI ship-based underway timeseries (1993-present)

Background

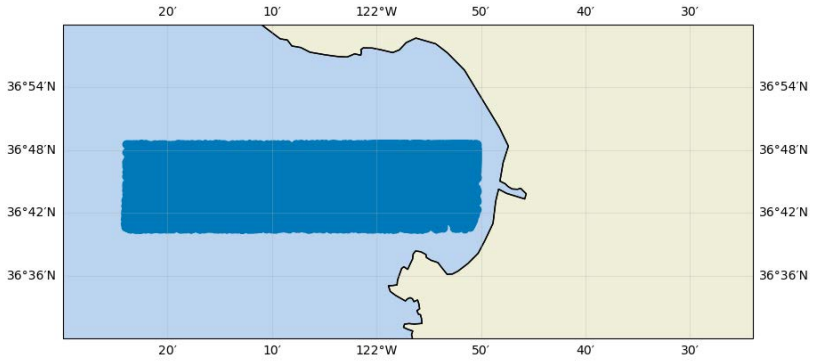
Surface pCO2, temperature, salinity, chlorophyll-a, and nitrate from 1993 to present from MBARI research vessels in Monterey Bay, California, USA.

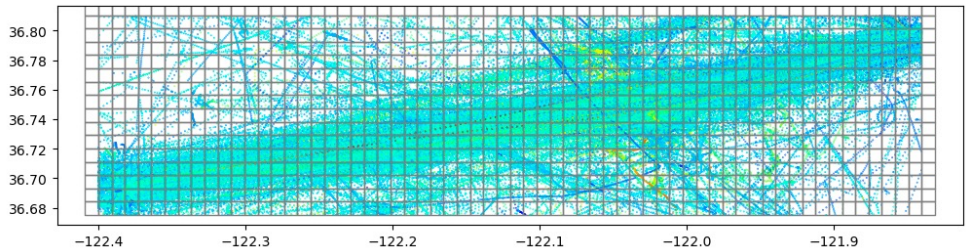
Data access/processing

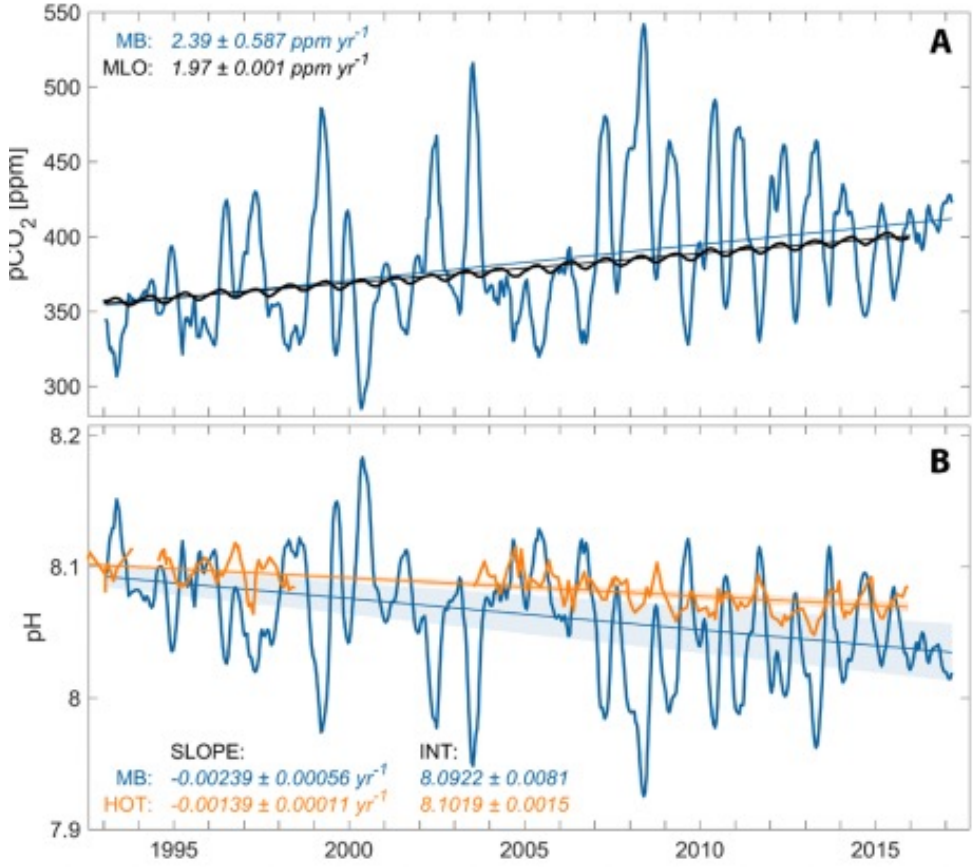
Pymysql to access SQL database

Data queried to region where ship commonly passes

1km grids generated, followed by aggregation







Chavez et al., 2017

Developing dynamic water quality dashboards

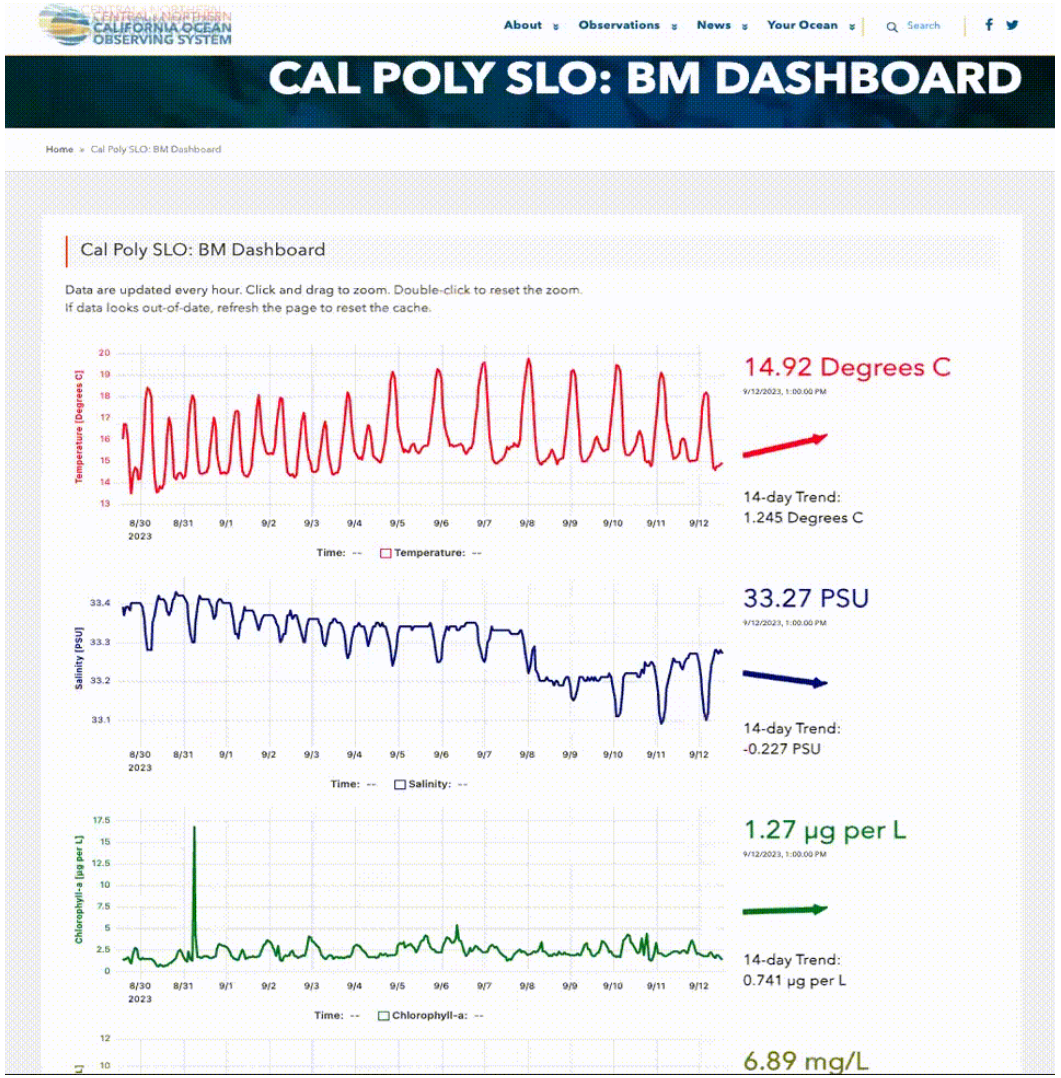
Led by Patrick Daniel, UCSC/CeNCOOS

Background

- Feedback from shellfish growers for user-friendly water quality dashboard interface
- Collaboration with Morro Bay shellfisheries community / researchers

Features

- Updated hourly
- 14-day trend
- QARTOD filtering



Prototype: www.cencoos.org/cal-poly-slo-bs1-dashboard/

Developing dynamic water quality dashboards

Build JSON parameter file for each shore station

```
"erddap-id": "morro-bay-bs1",
        "station-name": "BS Station - Morro Bay",
        "data-provider": "Cal Poly SLO",
        "past_days": 14,
        "data_variables": [
                {
                        "var-id": "sea_water_temperature",
                        "short_name":"Temperature",
                        "units":"°C",
                        "running_mean": "True",
                        "ylims":[12,22]

                },
```

Build URL for ERDDAP query via RESTful web service

Fetch data via erddapy & write to pandas dataframe

Calcualte hourly mean, remove "failed" QARTOD data

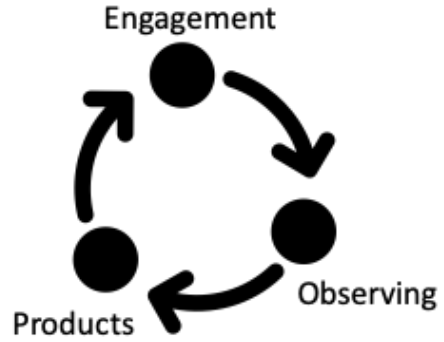
Write clean data to JSON file, push to CeNCOOS webserver

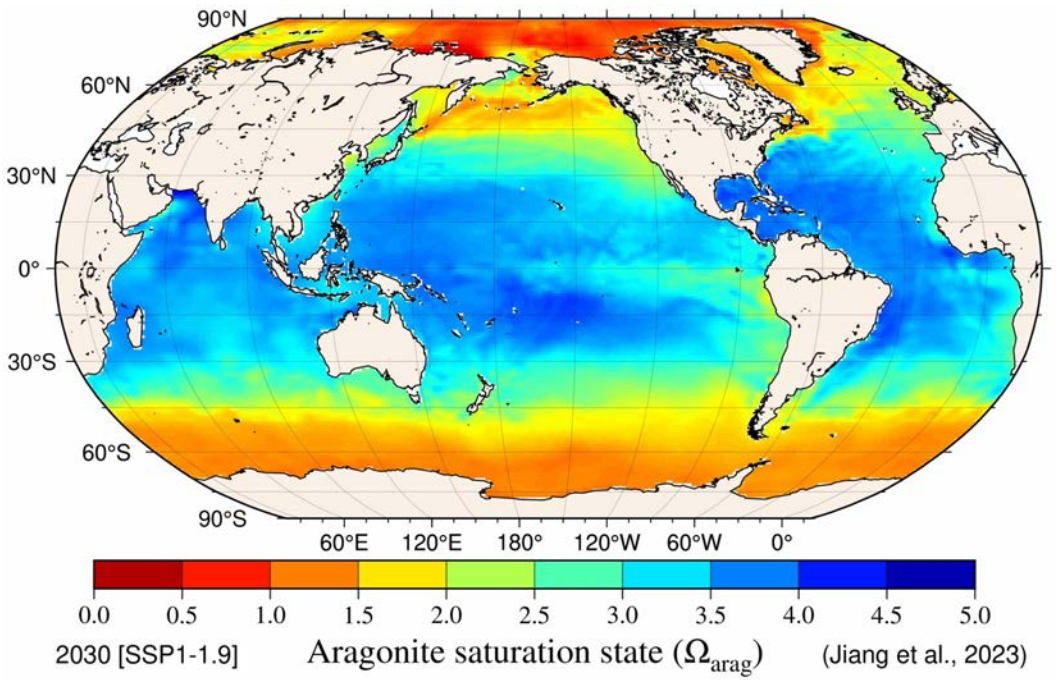
Generate plots using *μ*Plot JS library

Embed into Wordpress page

Future directions for CA OAH data product delivery

- Integration of other high-priority datasets into CA OAH Portal
  - CalCOFI: carbonate chemistry + zooplankton
  - Synthesis products, indicators
  - ROMS BEC model
- Development of next gen dashboards w/ Axiom
  - Shellfish growers
  - Water mgmt
  - Fisheries
  - MPAs
- Notification/alert system for users
- Stakeholder/end user engagement & outreach





Global Surface Ocean Acidification Indicators from 1750 to 2100 (Jiang et al., 2023)
(Available on OCADS in netCDF format)


Thank you!
mlebrec@mbari.org
CENTRAL & NORTHERN
CALIFORNIA OCEAN
OBSERVING SYSTEM
MBARI