This metadata document is posted by the Mid-Atlantic Ocean Data Portal for the Chlorophyll-a and pH Seasonal Cycle Data (2018-2023)

Source/originator: Authored by Nichole Ruiz, 2024 Fellow for Mid-Atlantic Coastal Acidification Network (MACAN), and student at Barnard College of Columbia University; data compiled from various station sources (listed below in the metadata for Chlorophyll-a and pH); data visualized by the Mid-Atlantic Ocean Data Portal.

Description

These data represent the seasonal cycle for chlorophyll-a and pH data collected from 2018-2023 for specified sites in the Mid-Atlantic Bight, Barnegat Bay, Chesapeake Bay, Delaware Bay, and offshore New Jersey. Seasons were categorized as active or inactive (relative to phytoplankton biomass growth) using the first order derivative of chlor-a, and are site-specific. For each season, pH and chlor-a values in that period were averaged.

Chlorophyll-a Data:

Chlorophyll-a satellite product (Gap-filled DINEOF: NOAA S-NPP NOAA-20 VIIRS and Copernicus S-3A OLCI). Science Quality, 2km Resolution, 01-01-2018 to 12/31/2023, Daily Files. Downloaded from NOAA CoastWatch program with support from the NOAA NESDIS Ocean Color Science Team between March 11th and 22nd. Optical penetration depth was not provided, but is understood to generally be <= 3m, with pixel depth variations due to differences in mixed layer depth.

pH Data:

pH measurements were taken from multiple in-situ sources, including gliders, cruises, and continuous monitoring buoys. Listed below are the sampling campaigns/missions whose data was used for calculating site-specific pH and whose sample coordinates were then used to index for nearby chlor-a pixels.

- CSIO: Citizen Science Initiative: Oyster (data provided by Emily Rivest, Virginia Institute of Marine Science)
- EOTB: Eyes on the Bay (www.eyesonthebay.net)
- National Estuary Program, Barnegat Bay Partnership (data QA/QC'ed by Janet Reimer, provided by Nicole Petersen and Stephen Pacella)
- Rutgers University RU30 Glider (dataset authors: Grace K. Saba, Lori M. Garzio, Laura J. Nazzaro, John Kerfoot)
- NYOS: New York Bight NYOS Glider (data provided by Janet Nye).
- The third NOAA East Coast Ocean Acidification (ECOA3) Cruise
- Delaware Estuary Water Quality Monitoring Program Boat Run data provided by John Yagecic.

Coordinate values are used to index the chlor-a pixel closest to each pH measurement. For each row containing a window range, the averaged value of coordinate pairs during that period is provided. These values were used to plot the chlor-a pixel value ($2km^2$ in size) as well as the pH value for that window.

Note some point positions were moved for map visualization purposes by/on the MARCO Portal (original point location overlaps with other station data), as indicated in the data attributes.

Purpose: The goal of the project and datasets is to understand how pH and chlor-a dynamics may impact each other under increasing pCO_2 conditions in the ocean surface and coastal zone. Carbonate chemistry, aka, acidification layers, are intended to provide information for natural resource management and planning purposes.

Access Constraints: All data are provided as is and no warranty is made as to the currency, completeness, accuracy, or utility of any specific data. Please read all metadata documentation for information about the data, including descriptions, purpose, contacts, use and citations/sources. For questions about full dataset used to create the acidification data layers, please contact jreimer@midatlanticocean.org or info@midatlanticocean.org The full dataset is not yet publicly available.

Contact Information: Janet Reimer, MARCO Science Program Manager, jreimer@midatlanticocean.org or info@midatlanticocean.org.

Data lineage: - Data has been complied through various sources (not all data may have been used in the final calculations depending on Chlor-a data availability):

- NOAA Ocean Acidification Program Cruises: ECOA, GOMECC, EcoMon, SOOP. [Data downloads: https://www.ncei.noaa.gov/products/ocean-carbon-acidification-data-system]
    - Cruise List: ECOA1, ECOA2, ECOA3, GOMECC1, GOMECC2, GU1302, GU1305, GU1608 Leg 1 and 2, GU1701 Leg 1 and 2, GU 1702, GU1804, GU1902, GU1905, GU2102, HB1202, HB1701, HB1103 Leg 2, HB1502, HB1803, HB1902, HB1405 Leg 1, HB2204, HB1301 Leg 1-4, PC1207, PC2104, PC2106, PC2205, PC1607, PC1405, DE1202, S11802
    - Structure: discrete, surface and bottom, Mid-Atlantic shelf
    - Years: 2011-2022
- New York Bight Offshore Monitoring Cruises, Stony Brook University and New York State Department of Environmental Conservation. [Data acquisition: NCEI Accession 0294129]
    - Data access: provided by Janet Nye, Baoshan Chen, and Tyler Menz. Future inquiries should go to Janet Nye (jnye@nyelab.org).
    - Cruise List: NYOS1902, NYOS1905, NYOS1907, NYOS1910, NYOS2002, NYOS2006, NYOS2010, NYOS2103, NYOS2105
    - Structure: discrete, surface and bottom, New York Bight shelf and Long Island Sound

    - Years: 2019-2021
- CSI: Oyster water quality sampling, Virginia Institute of Marine Sciences. [Data acquisition: NCEI submission pending; contact Emily Rivest ebrivest@vims.edu]
    - Station list: Merroir, Piankatank River, Urbanna Creek, VIMS Pier, Gray's Inn Marina, Gunston School, Tilghman Farm, Vint Lawrence Dock
    - Structure: discrete, surface, Chesapeake Bay tributaries
    - Years: 2017-2023

Process Steps:

pH Data Processing

1. Acquire data that has been collected using standard equipment, measurement temperatures, QA/QC, use of recommended best practices (Dickson et al., 2007), use of certified reference material (CRM), and has reported uncertainty. In some cases, pH was calculated from other carbonate chemistry variables.
2. Uncertainty constraints
    a. Use uncertainty estimates and methods provided when available.
    b. If there are no uncertainty details provided, use replicate samples (collected at same time, location, and depth) to calculate standard deviation and percent difference between paired replicate samples.
    c. If mean standard deviation or reported uncertainty meet the following criteria, the data is climate quality (Newton et al., 2015):
        i. Total alkalinity: 2 µmol/kg
        ii. Dissolved inorganic carbon: 2 µmol/kg
        iii. pH: 0.003 iv. $p\mathrm{CO_2}$: 0.5%
3. Remove any datasets that do not meet the climate quality criteria as calculated in step 1.
4. Replace replicate samples with their average in the dataset before proceeding, to avoid upweighting those samples.
5. Remove any data points that are flagged as "bad" or "questionable".
6. Use CO2Sys v3.3 to calculate all carbonate chemistry parameters for in-situ temperature and pressure.
    a. Input the data in the DATA tab.
    b. Data requirements: Salinity, input (measurement) temperature, input (measurement) pressure, output (in-situ) temperature, output (in-situ) pressure, two or more carbonate chemistry parameters.
        i. Carbonate chemistry parameters: total alkalinity (TA), dissolved inorganic carbon (DIC or T_C), pH, partial pressure of CO_2 ($p\mathrm{CO_2}$), carbonate ion concentration ($[CO_3^{2-}]$).
        ii. Verify units in dataset match the units in CO2Sys.
        iii. Verify measurement temperatures and whether reported values are for measurement or in situ temperature.
        iv. Remove any samples that do not meet the minimum data requirements.
    c. When available: total phosphorous, total silicon, ammonia, and hydrogen sulfide.
    d. INPUT tab:

i. Constants: for bays use Cai and Wang (1998), for ocean or when salinity is above 15 psu use Lueker (2000). ii. $KHSO_4$: Dickson (1990)

iii. KHF: Perez and Fraga (1987)

iv. $[B]_I$: Uppstrom (1974)

v. pH scale: select the scale in which pH data was provided

vi. Temperature/Salinity: EOS-80 (as long as temperature is in °C and salinity is in ppt or psu)

vii. Error Propagation: Using values in ERROR sheet

viii. In the DATA tab press START to calculate the parameters, then paste the "Output" (in situ) results and errors back into data sheet. Be sure to use "paste as values".

7. Use R to analyze data.
    a. Give all datasets the same column headers. Use NA for missing data.
    b. Import datasets into R and combine into one dataframe.
    c. Filter out any data points that are outside the Mid-Atlantic
        i. Longitude: 78°W to 69.5°W
        ii. Latitude: 35°N to 41.6°N
    d. Filter data into surface (<= 3m depth) and daytime (8:00 to 16:30 local time) samples.
    e. Use histograms to plot parameter distributions and check for any unrealistic values (e.g. negative carbonate chemistry parameters). Save full datasets as csv files.

Chlorophyll *a* Data Processing

1. Download product (Gap-filled DINEOF: NOAA S-NPP NOAA-20 VIIRS and Copernicus S-3A OLCI) from the NOAA CoastWatch ERDDAP data server using the griddap Data Access tool. Define spatial bounding box and time range, and download dataset as a netCDF file to retain native data layers.
2. For each non-duplicate pH sample with a unique sample location and timestamp, use its coordinate pair to index for the nearest satellite chlor-a pixel. Filter out chlor-a pixels whose value is NaN (and corresponding pH sample values) from the dataset.
3. Growth cycle calculation
    a. Derive seasonal growth signal
        i. Group remaining chlor-a values by day for a generalized year (there should be a maximum of 365 bins, one for each DOY. Fit a linear trend to the daily chlorophyll pixel dataset and use the residual data points to plot the first order derivative of the seasonal chlor-a cycle (dC/dt) with a 30-day rolling average.
    b. Identify Growth Window (GW) and Non-Growth windows (NG)
        i. We identified a growth window to be a DOY range where dC/dt > 0. To account for intraseasonal variability in the growth data, we allowed for adjacent positive windows separated by a gap of <= 20 days of dC/dt < 0 to be merged into a continuous growth window.
        ii. All windows that were not identified as growth windows from this analysis were assigned to be non-growth windows.

Seasonally Averaged pH and Chlorophyll *a* data

1. Use the DOY range corresponding to each GW and NG to group pH and chlorophyll values. Do this for each of the four sites: Barnegat Bay, Chesapeake Bay, Delaware Bay and offshore New Jersey. For each site, average all pH sample coordinates to create a mean coordinate pair for the site data.
2. Compute the average of all pH and chlorophyll values for each respective growth or non-growth window.
3. Save your calculations for pH, chlor-a, latitude, and longitude to a csv in consecutive order of GW and NG windows.

References for the methodology:

Dickson, A. G., Sabine, C. L., and Christian, J. R. (Eds.) 2007. Guide to Best Practices for Ocean CO_2 Measurements. *PICES Special Publication 3*, 191 pp.

Newton, J. A., Feely, R. A., Jewett, E. B., Williamson, P., and Mathis, J. (2015). Global Ocean Acidification Observing Network: Requirements and Governance Plan, Second Edition. http://goa.on.org/documents/general/GOA-ON-2nd-edition_final.pdf

Liu, X., & Wang, M. (2018). Gap filling of missing data for VIIRS global ocean color products using the DINEOF method. IEEE Transactions on Geoscience and Remote Sensing, 56(8), 4464-4476.

Brody, S. R., Lozier, M. S., & Dunne, J. P. (2013). A comparison of methods to determine phytoplankton bloom initiation. Journal of Geophysical Research: Oceans, 118(5), 2345-2357.

Dataset list

- BarnegatBay_Seasonal_Averages.csv
- ChesapeakeBay_Seasonal_Averages.csv
- DelawareBay_Seasonal_Averages.csv
- OffshoreNJ_Seasonal_Averages.csv

Column definitions for the pH data

- DOY Range: Generalized Day of Year
- Window Type: Growth Window (GW) or Non-Growth Window (NG) in chronological order for a given year.
- Avgd pH (Total): pH on the Total scale (mol/kg-SW) averaged from all daytime surface samples taken at a given location for its corresponding DOY range.
- Avg chlor-a (mg m-3): Satellite chlorophyll *a* pixels (in mg/m^3) indexed over a given location and averaged for its corresponding DOY range.
- Avg LAT: Averaged latitude for all pH samples at a given site.
- Avg LON: Averaged longitude for all pH samples at a given site
- Overlapping points: Indicates if the original sampled point position moved for map visualization (original point location overlaps with other station data)