Optimally Interpolated O2 anomalies based on World Ocean Database 2018

Website: https://www.bco-dmo.org/dataset/886218
Data Type: model results, Cruise Results
Version: 1
Version Date: 2023-01-11

Project
» Mapping Dissolved Oxygen using Observations and Machine Learning (DO Machine Learning)

Contributors	Affiliation	Role
Ito, Takamitsu	Georgia Institute of Technology (GA Tech)	Principal Investigator
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
OIO2 is a gridded data product of dissolved oxygen interpolated from shipboard observations archived in the World Ocean Database 2018 (WOD18). The quality-controlled WOD18 data are averaged for each bin at 1°x1° and monthly resolution where mean, variance, and sample size are recorded from 1965 to 2014 for the bottle data, and from 1987 to 2014 for the CTD-O2 data.

Table of Contents

Coverage

Spatial Extent: **N**:90 **E**:180 **S**:-90 **W**:-180
Temporal Extent: 1965 - 2012

Dataset Description

This is a gridded data product generated from many observational profiles from the World Ocean Database 2018. There are two types of measurement methods included in the data source: bottle O2 data measured by the Winkler titration method and data from the CTD-O2 sensor. The data product is available as a NetCDF file (OIO2_g1x1v47_1967_2012.nc).

Methods & Sampling

The monthly mean climatology is determined by calculating the climatological monthly mean combining the bottle and CTD-O2 data and then filling data gaps. The gridded map of climatological O2 is assembled to estimate oxygen anomalies for all 1°x1° grid cells following the optimal interpolation method which provides the least-square estimate of O2 anomaly field on regularly spaced grid cells. Stationary and isotropic Gaussian covariance is assumed throughout this study, with the e-folding length scale of 1,000 kilometers (km). This particular choice of length scale controls how far an observation can influence the far field.

The optimal interpolation is applied to each basin separately except for the Southern Ocean, for which data points from all basins southward of 30°S are used. As a consistency check, the calculated climatology is compared to the World Ocean Database 2018 and it is confirmed that the optimally interpolated O2 climatology closely matches that of WOA18. Departures from the monthly climatology are recorded as O2 anomalies. The binned data is very sparse at a monthly timescale. For each year, the monthly anomaly data is averaged into yearly anomalies neglecting the months with missing data. This step increases data coverage significantly while averaging out high-frequency variability in the data including changes shorter than the yearly timescale. In addition, a 5-year moving window (pentadal) averaging is applied to the yearly anomaly neglecting the years with missing data. This further increases the data coverage, while averaging out variability on a timescale shorter than 5 years. The resulting pentadal O2 anomaly data covers the 46-year period from 1967 to 2012.

Finally, the optimal interpolation is applied for each year to yield the optimally interpolated O2 anomalies. The resulting data product is saved in the netCDF format, and it includes O2 climatology, O2 anomalies, and standard error.

Data Processing Description

Data Processing:
Optimal interpolation is performed using the Gaussian covariance function with the e-folding scale of 1,000 kilometers (km). Briefly, it produces the best-fit O2 distribution in the least square sense, given the covariance structure in the dataset. The binned data is very sparse at a monthly timescale. For each year, the monthly anomaly data is averaged into yearly anomalies neglecting the months with missing data. This step increases data coverage significantly while averaging out high-frequency variability in the data including changes shorter than the yearly timescale. In addition, a 5-year moving window (pentadal) averaging is applied to the yearly anomaly neglecting the years with missing data. This further increases the data coverage, while averaging out variability on a timescale shorter than 5 years. The resulting pentadal O2 anomaly data covers the 46-year period from 1967 to 2012. Finally, the optimal interpolation is applied for each year to yield the optimally interpolated O2 anomalies.

[table of contents | back to top]

Data Files

File
OIO2_g1x1v47_1967_2012.nc (NetCDF, 2.11 GB) MD5:e5e1248be2f0006559d3b67c83398762 File header: netcdf OIO2_g1x1v47_1967_2012 { dimensions: lon = 360 ; lat = 180 ; depth = 47 ; time = 46 ; variables: double lon(lon) ; lon:_FillValue = -99999. ; lon:standard_name = "lon" ;

File

```
lon:long_name = "longitude" ;
lon:units = "degrees_east" ;

double lat(lat) ;
lat:_FillValue = -99999. ;
lat:standard_name = "lat" ;
lat:long_name = "latitude" ;
lat:units = "degrees_north" ;

double depth(depth) ;
depth:_FillValue = -99999. ;
depth:standard_name = "depth" ;
depth:long_name = "depth from the surface ocean" ;
depth:units = "m" ;
depth:bounds = "depth_bnds" ;

double time(time) ;
time:_FillValue = -99999. ;
time:standard_name = "time" ;
time:long_name = "time" ;
time:units = "days since 1950-01-01" ;
time:calendar = "proleptic_gregorian" ;

double o2(time, depth, lat, lon) ;
o2:_FillValue = -99999. ;
o2:long_name = "objective map of dissolved oxygen anomaly" ;
o2:units = "micro-molO2/L" ;

double o2_standard_error(time, depth, lat, lon) ;
o2_standard_error:_FillValue = NaN ;

double o2_climatology(depth, lat, lon) ;
o2_climatology:_FillValue = NaN ;

// global attributes:
```

File
:_NCProperties = "version=2,netcdf=4.8.1,hdf5=1.12.1" ;
}

[table of contents | back to top]

Related Datasets

References

Boyer, T.P., O.K. Baranova, C. Coleman, H.E. Garcia, A. Grodsky, R.A. Locarnini, A.V. Mishonov, C.R. Paver, J.R. Reagan, D. Seidov, I.V. Smolyar, K. Weathers, M.M. Zweng,(2018): World Ocean Database 2018. A.V. Mishonov, Technical Ed., NOAA Atlas NESDIS 87. https://www.ncei.noaa.gov/sites/default/files/2020-04/wod_intro_0.pdf

[table of contents | back to top]

Parameters

Parameters for this dataset have not yet been identified

[table of contents | back to top]

Project Information

Mapping Dissolved Oxygen using Observations and Machine Learning (DO Machine Learning)

Coverage: Global

NSF Award Abstract:
Oxygen is produced by algae in the sunlit surface waters and is released into the atmosphere. This process contributes to about the half of atmospheric oxygen. However, there is a growing consensus in the scientific community that the global ocean oxygen inventory has declined in recent decades. Ocean heat uptake causes the reduction of solubility, and changes in circulation and biogeochemical processes associated with the ocean warming can further change ocean oxygen content. The reduction of dissolved oxygen can have far-reaching impacts on the marine habitats. Recent estimates of the global oxygen decline are in the range of 0.5-3.3% over the period of 1970- 2010. Distribution of the historical O_2 measurements is irregular and sparse, causing significant uncertainty in these estimates. The objective of this project is to determine changes in the dissolved oxygen content of the oceans based on observational data and machine learning techniques. The overarching hypothesis of this project is that there are significant, regional relationships between O_2 and other observed quantities. Dissolved oxygen is ultimately controlled by the combination of ocean circulation, air-sea gas transfer and biological processes. These processes can be linked with other observed quantities such as temperature (T) and salinity (S), but such relationships can be complex and non-linear. Therefore, it is difficult to determine a universal relationship that governs the distribution of O_2 based on the first principle. However, machine learning algorithms can extract empirical relationships between O_2 and other variables from existing observations, allowing us to estimate O_2 where direct observation is not available. The work will also support one graduate and one undergraduate student research and outreach activities at local events.

In this project, machine learning will be used to fill data gaps in the historical O_2 dataset and to generate an improved, gridded estimates of O_2 from 1960 to present. This approach takes advantage of the large amount of accumulated in-situ observations over multiple decades including not only O_2 itself but also other related variables such as T and S. The proposed work revolves around three hypotheses. First, the current estimates of global O_2 trend and variability are strongly influenced by relatively data-rich regions such as North Atlantic and North Pacific. Machine-learning based O_2 dataset with an improved gap-fill approaches is hypothesized to better represent relatively data-poor regions such as tropics and southern hemisphere oceans. Secondly, the current estimates indicate that less than half of O_2 decline is explained by the solubility effect. The global O_2-heat relationship measures the reinforcing effects of ocean ventilation and biogeochemistry. Machine learning

can estimate empirical relationships between O2, T and other physical variables, which can be manipulated to perform sensitivity experiments. The empirical model of O2 can constrain the regional and global O2-heat relationship. Thirdly, it is hypothesized that observed O2 decline in the tropical thermocline are driven by the combination of natural climate variability and long-term trends. In the proposed work, sensitivity experiments are performed with the empirical model of O2 to evaluate the influences of long-term trends and decadal-scale changes associated with the modes of natural climate variability.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-2123546

[table of contents | back to top]