

SUPPORTING INFORMATION

Appendix S2: Extended materials and methods used to build Bio-ORACLE dataset

General Approach

A set of 23 macroecological variables representing environmental dimensions assumed to influence the distribution of marine shallow water organisms was assembled. Our data collecting approach consisted of the compilation of preprocessed remotely sensed and in situ measured oceanographic data. We limited our analyses to variables relevant at the macroecological level and with global coverage. This supplementary material aims to give an overview and more detailed information of the Bio-ORACLE variables.

Our data processing pipeline consisted of the following steps:

- Data acquisition (WOD09 and Ocean Color Web)
- In case of WOD09 data, raster generation by interpolation with DIVA
- Conversion of raster to a convenient format: ESRI ascii
- Calculation of metrics (e.g. rasters with mean, minimum and maximum values)
- Application of uniform landmask
- Clipping rasters to 70°N – 70°S

Recommendations & Availability

The authors advise against the use of the Bio-ORACLE variables at latitudes higher than 70° N/S to reduce possible biases and potential errors in the data at high latitudes (see text for explanation).

Bio-ORACLE is released under the GNU General Public License and is available for download on <http://www.bio-oracle.ugent.be>.

Data layers

Remotely sensed data

Calcite concentration (mol/m ³)	
Original Spatial Resolution	5 arcmin (9.2 km)
Sensor	Aqua-MODIS
Data	Seasonal climatologies
Temporal Range	2002 - 2009
Brief description	Calcite concentration indicates the concentration of calcite (CaCO ₃) in oceans.
Manipulation	Derivation of metric: Mean
Source	Reference: (Feldman & McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/

Chlorophyll A concentration (mg/m ³)	
Original Spatial Resolution	5 arcmin (9.2 km)
Sensor	Aqua-MODIS
Data	Monthly climatologies
Temporal Range	2002 - 2009
Brief description	Chlorophyll A concentration indicates the concentration of photosynthetic pigment chlorophyll A (the most common “green” chlorophyll) in oceans. Please note that in shallow water these values may reflect any kind of autotrophic biomass.
Manipulation	Derivation of metrics: Mean, Minimum, Maximum, Range
Source	Reference: (Feldman & McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/

Cloud fraction (%)	
Original Spatial Resolution	6 arcmin (11 km)
Sensor	Terra-MODIS
Data	Monthly images
Temporal Range	2005 - 2010
Brief description	Cloud fraction indicates how much of the earth is covered by clouds.
Manipulation	Derivation of metrics: Mean, Minimum, Maximum Bilinear interpolation (10 km → 9.2 km)
Source	Reference: (NASA 2010) URL: http://neo.sci.gsfc.nasa.gov/Search.html

Diffuse attenuation coefficient at 490 nm (m^{-1})	
Original Spatial Resolution	5 arcmin (9.2 km)
Sensor	Aqua-MODIS
Data	Monthly climatologies
Temporal Range	2002 - 2009
Brief description	The diffuse attenuation coefficient is an indicator of water clarity. It expresses how deeply visible light in the blue to the green region of the spectrum penetrates in to the water column.
Manipulation	Derivation of metrics: Mean, Minimum, Maximum
Source	Reference: (Feldman & McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/

Photosynthetically Available Radiation ($Einstein/m^2/day$)	
Original Spatial Resolution	5 arcmin (9.2 km)
Sensor	SeaWiFS
Data	Monthly climatologies
Temporal Range	1997 - 2009
Brief description	Photosynthetically Available Radiation (PAR) indicates the quantum energy flux from the Sun (in the spectral range 400-700 nm) reaching the ocean surface.
Manipulation	Derivation of metrics: Mean, Maximum Minimum PAR was considered, but excluded due to the high level of artifacts in original data.
Source	Reference: (Feldman & McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/

Sea Surface Temperature ($^{\circ}C$)	
Original Spatial Resolution	5 arcmin (9.2 km)
Sensor	Aqua-MODIS
Data	Monthly climatologies
Temporal Range	2002 - 2009
Brief description	Sea surface temperature is the temperature of the water at the ocean surface. This parameter indicates the temperature of the topmost meter of the ocean water column.
Manipulation	Derivation of metrics: Mean, Minimum, Maximum, Range
Source	Reference: (Feldman & McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/

In situ measured oceanographic data

Dissolved oxygen (ml/l)	
Database	World Ocean Database 2009
Data	Standard Level Data: Ocean Station Data (OSD); High-resolution Conductivity-Temperature-Depth (CTD) (Surface)
Temporal Range	1898 - 2009
Number of data points	540582
Brief description	Dissolved oxygen concentration [O_2]
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: http://www.nodc.noaa.gov/

Nitrate (μmol/l)	
Database	World Ocean Database 2009
Data	Standard Level Data: OSD (Surface)
Temporal Range	1928 - 2008
Number of data points	189530
Brief description	This layer contains both [NO ₃] and [NO ₃ +NO ₂] data. By this we mean chemically reactive dissolved inorganic nitrate and nitrate or nitrite. (It is important to note that data reported as [NO ₃] in the WOD09 should be used with caution because it is difficult to verify that the [NO ₃] (nitrate) data are [NO ₃ +NO ₂] or [NO ₃]. (Boyer <i>et al.</i> 2009))
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: http://www.nodc.noaa.gov/

pH (unitless)	
Database	World Ocean Database 2009
Data	Standard Level Data: OSD (Surface)
Temporal Range	1910 - 2007
Number of data points	117833
Brief description	Measure of acidity in the ocean.
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: http://www.nodc.noaa.gov/

Phosphate (μmol/l)	
Database	World Ocean Database 2009
Data	Standard Level Data: OSD (Surface)
Temporal Range	1922 - 1986
Number of data points	226816
Brief description	Reactive ortho-phosphate concentration [HPO ₄ ⁻²] in the ocean.
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: http://www.nodc.noaa.gov/

Salinity (PSS)	
Database	World Ocean Database 2009
Data	Standard Level Data: CTD (Surface)
Temporal Range	1961 - 2009
Number of data points	532377
Brief description	Salinity indicates the dissolved salt content in the ocean.
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: http://www.nodc.noaa.gov/

Silicate (μmol/l)	
Database	World Ocean Database 2009
Data	Standard Level Data: OSD & CTD (Surface)
Temporal Range	1930 - 2008
Number of data points	234417
Brief description	This variable indicates the concentration of silicate or ortho-silicic acid [Si(OH) ₄] in the ocean.
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: http://www.nodc.noaa.gov/

References

Boyer T.P., Antonov J.I., Baranova O.K., Garcia H.E., Johnson D.R., Locarnini R.A., Mishonov A.V., O'Brien T.D., Seidov D., Smolyar I.V. & Zweng M.M. (2009). *World Ocean Database 2009*. U.S. Gov. Printing Office, Washington D.C.

Feldman G.C. & McClain C.R. (2010). Ocean Color Web. URL <http://oceancolor.gsfc.nasa.gov/>

NASA (2010). NASA Earth observations (NEO). URL <http://neo.sci.gsfc.nasa.gov/Search.html>

Correlation Matrix

[illegible]

Simulation study: Cubic inter/extrapolation

In order to produce rasters with a uniform landmask across the complete dataset two operations were necessary:

- Values that fell on land needed to be masked;
- Missing values needed to be estimated.

The latter operation was done using a cubic inter/extrapolation algorithm. We chose this algorithm because of its stable and smooth characteristics. To assess the quality of the interpolation we performed a simulation study on the remotely sensed data*.

The evaluation consisted of:

- Random selection of a monthly climatology of each remotely sensed variable;
- Random selection of 100 ‘open sea’ pixels (i.e. pixels that are completely surrounded by other sea pixels) from the respective climatology;
- Random selection of 100 ‘coastal’ pixels (i.e. pixels directly adjacent to one or more no-data value) from the respective climatology;
- Removal of these 200 data points from the respective climatology;
- Application of the cubic inter/extrapolation algorithm;
- Extraction of data from both the original and the inter/extrapolated layer;
- Calculation of the average difference between both layers;
- Evaluation of that difference in relation to the total variable range.

The results are shown in the table below:

Variable	<u>Coast</u>		<u>Open Sea</u>	
	$\Delta(\text{avg}(\text{Interpol} - \text{Real}))$	% of total variable range	$\Delta(\text{avg}(\text{Interpol} - \text{Real}))$	% of total variable range
Sea Surface Temperature	0.22675 °C	0.57	0.04161 °C	0.11
Chlorophyll A Concentration	0.38901 mg/m ³	0.60	0.00831 mg/m ³	0.01
Diffuse Attenuation	0.00922 m ⁻¹	0.80	0.00088 m ⁻¹	0.08
Photosynthetically Available Radiation	0.49984 Einstein/m ² /day	0.73	0.18093 Einstein/m ² /day	0.27
Calcite Concentration	0.00086 mol/m ³	1.54	0.000134 mol/m ³	0.24

* Cloud cover data was not taken into account as no pixels needed to be inter/extrapolated.

Supporting figure S1



