Bio-ORACLE v3.0. Pushing marine data layers to the CMIP6 Earth system models of climate change research

Jorge Assis[1,2*], Salvador Jesús Fernández Bejarano[3], Vinícius W. Salazar[4], Lennert Schepers[3], Lidiane Gouvêa[1], Eliza Fragkopoulou[1], Frederic Leclercq[3], Bart Vanhoorne[3], Lennert Tyberghein[3], Ester A. Serrão[1], Heroen Verbruggen[5†], Olivier De Clerck[6†]

[1] Centre of Marine Sciences (CCMAR/CIMAR LA), Universidade do Algarve, Faro, Portugal

[2] Faculty of Bioscience and Aquaculture, Nord Universitet, Bodø, Norway

[3] Flanders Marine Institute, Ostend, Belgium

[4] Melbourne Integrative Genomics, School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia.

[5] School of BioSciences, University of Melbourne, Parkville, Australia

[6] Department of Biology, Phycology Research Group, Ghent University, Ghent, Belgium

[*] Corresponding author (jorgemfa@gmail.com)

[†] Shared senior authorship

Acknowledgements

This work was funded by (1) the Horizon Europe Framework Programme through project MPAEurope (HORIZON-CL6-2021-BIODIV-01-12), (2) the Portuguese National Funds from FCT - Foundation for Science and Technology through projects UIDB/04326/2020 (DOI:10.54499/UIDB/04326/2020), UIDP/04326/2020 (DOI:10.54499/UIDP/04326/2020), LA/P/0101/2020 (DOI:10.54499/LA/P/0101/2020), PTDC/BIA-CBI/6515/2020(DOI:10.54499/PTDC/BIA-CBI/6515/2020), the EU BiodivRestore-253 (FCT: DivRestore/0013/2020), the Individual Call to Scientific Employment Stimulus 2022.00861.CEECIND/CP1729/CT0003 (DOI:10.54499/2022.00861.CEECIND/CP1729/CT0003) and the fellowship SFRH/BD/144878/2019, (3) the European Marine Biological Resource Centre Belgium (GOH3817N, I001621N]) and (4) the Research Foundation - Flanders (FWO) as part of the Belgian contribution to LifeWatch (I002021N) and EMBRC (I001621N). Additional support was received by the University of Melbourne and computational resources provided through the Melbourne Research Cloud. The authors acknowledge the Copernicus Marine Environment Monitoring Service (https://data.marine.copernicus.eu/products), GlobColour (http://globcolour.info; data developed, validated, and distributed by ACRI-ST, France) and the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Biosketch

Jorge Assis is an associate researcher at the Centre of Marine Sciences, University of Algarve (Portugal), where he leads the Biodiversity Data Sciences research group. He is also a professor at Nord University (Norway). His research is focused on developing data science tools to well-inform biodiversity conservation and management.

Data availability statement

The layers are openly available at the ERDDAP instance hosted at the Flanders Marine Institute Data Centre (https://erddap.bio-oracle.org/erddap), through the Bio-ORACLE website (temporary link for peer-review: https://www.bio-oracle.org/downloads-to-email-v3.php) and the R and Python packages available in the permanent GitHub repository (https://github.com/bio-oracle).

Abstract

Motivation: Impacts of climate change on marine biodiversity are often projected with species distribution modelling, using standardized data layers representing physical, chemical and biological conditions. Yet, the available data layers (1) have not been updated to incorporate data of the Sixth Phase of the Coupled Model Intercomparison Project (CMIP6), which comprise the Shared Socioeconomic Pathway (SSP) scenarios; (2) consider a limited number of Earth System Models (ESMs); and (3) miss important variables expected to influence future biodiversity distributions. These limitations might undermine biodiversity impact assessments, by failing to integrate them within the context of the most up-to-date climate change projections, raising the uncertainty in estimates and misinterpreting the exposure of biodiversity to extreme conditions. Here, we provide a significant update of Bio-ORACLE, extending biologically relevant data layers from present-day conditions to the end of the 21st century Shared Socioeconomic Pathway scenarios based on a multi-model ensemble with data from CMIP6. Alongside, we provide R and Python packages for seamless integration in modelling workflows. The data layers aim to enhance the understanding of the potential impacts of climate change on biodiversity and to support well-informed research, conservation and management.

Main types of variable contained: Surface and benthic layers for, chlorophyll-*a*, diffuse attenuation coefficient, dissolved iron, dissolved oxygen, mixed layer depth, nitrate, ocean temperature, pH, phosphate, photosynthetic active radiation, total phytoplankton, total cloud fraction, salinity, silicate, sea-water direction, sea-water velocity, topographic slope, topographic aspect, terrain ruggedness index, topographic position index and bathymetry, and surface layers for air temperature, sea-ice cover and sea-ice thickness.

Spatial location and grain: Global at 0.05° resolution.

Time period and grain: Decadal from present-day to the end of the 21st century (2000-2100).

Major taxa and level of measurement: Marine biodiversity associated with surface and epibenthic habitats.

Software format: A package of functions developed for Python and R software.

Introduction

Research quantifying the impacts of climate change on biodiversity relies heavily on species distribution modelling (Peterson *et al.*, 2011) using high-resolution data layers representing physical, chemical and biological conditions at the global scale (Tyberghein *et al.*, 2012; Assis *et al.*, 2017b; Fick & Hijmans, 2017). Initiatives providing essential climate layers uniform in extent and resolution, like MERRAclim (Vega et al., 2017) and WorldClim (Fick & Hijmans, 2017) for terrestrial environments and the World Ocean Atlas (Levitus *et al.*, 2013), Hexacoral (Fautin & Buddemeier, 2008), Marspec (Sbrocco & Barber, 2013), Bio-ORACLE (Tyberghein *et al.*, 2012; Assis *et al.*, 2017b) and OCLE (de la Hoz et al., 2018) for marine environments, established solid standards to estimate future climate-induced pressures on the world's ecosystems and associated ecosystem services (Assis *et al.*, 2017a, 2022b; Gouvêa *et al.*, 2020, 2022; Martins *et al.*, 2021). But despite their relevance, current marine datasets have three main limitations that can preclude the development of well-informed conservation and management strategies (Arneth *et al.*, 2020; Arafeh-Dalmau *et al.*, 2021).

First, current standardized climate layers have not been updated to incorporate data of the Sixth Phase of the Coupled Model Intercomparison Project (CMIP6), which serves as the core contribution to the latest Intergovernmental Panel on Climate Change Assessment Report (IPCC). The CMIP6 dataset includes an updated set of variables (Séférian *et al.*, 2020), which better capture the spatial and temporal variability of physical and biogeochemical properties of the global ocean (Séférian *et al.*, 2020), and the Shared Socioeconomic Pathway (SSP) scenarios, a set of narratives on possible global change trajectories throughout the 21st century, with and without climate policy responses (Riahi *et al.*, 2017; van Vuuren *et al.*, 2017). Compared to the previous Representative Concentration Pathway scenarios (van Vuuren *et al.*, 2011), despite producing similar forcing pathways, the SSPs use more up-to-date socioeconomic assumptions (Burgess *et al.*, 2023) and encompass the Paris Agreement expectations, which allows estimating the biodiversity benefits of compliance to international climate policies (Sanderson *et al.*, 2016; Martins *et al.*, 2021).

Second, current standardized climate layers projecting future conditions consider a limited number of Earth System Models (ESMs) (Schoeman et al., 2023). These are intensive computer simulations tuned by specific parameters and conditions (e.g., greenhouse gas concentrations, solar forcing, land-use, as well as atmospheric and ocean dynamics (Mechoso & Arakawa, 2015)) that can resolve global atmospheric and oceanic conditions across different time scales. There are, however, inherent uncertainties of ESMs projections due to the complexity of the Earth's climate system, as well as the limited understanding and the challenges in incorporating particular processes in the models.

Accordingly, the multi-model ensemble offers a straightforward approach in species distribution modelling (Araújo & New, 2007; Assis *et al.*, 2017b; Gouvêa *et al.*, 2022), yet at the cost of incorporating ESM that do not perform well in reproducing historical climates over time (please refer to the 'hot model' problem; Hausfather et al., 2022). The previous version of Bio-ORACLE (Assis *et al.*, 2017b) was based on a multi-model ensemble, yet limited to two ESMs from the dozens available, limiting the breadth of projections of climate change impact.

Third, current standardized climate layers miss projections of biologically meaningful variables beyond temperature, salinity and sea ice conditions (Tyberghein *et al.*, 2012; Assis *et al.*, 2017b). For instance, dissolved oxygen, primary productivity and pH (among other variables) are expected to change in the coming decades and are emerging as limiting factors for marine biodiversity (Krumhardt *et al.*, 2017; Fragkopoulou *et al.*, 2021; Martins *et al.*, 2021; Shi *et al.*, 2021). Failing to consider them in climate change projections might misinterpret biodiversity exposure to novel detrimental conditions.

To address these gaps in the marine climate data space, we provide a substantial extension of Bio-ORACLE, including a set of essential variables for bioclimatic modelling from present-day conditions to the end of the 21st century SSP scenarios, for both surface and benthic (i.e., along the seafloor) conditions. These were built by ensembling numerous ESMs provided by the CMIP6, from the higher (SSP1, in line with the Paris Agreement) to the lower mitigation (SSP5) trajectories of global change. These new layers are provided under the FAIR principle (Findability, Accessibility, Interoperability, and Reusability (Wilkinson *et al.*, 2016)) with a finer spatial and temporal resolution (0.05° spatial resolution; decadal climatologies). Alongside the data layers, R and Python packages were developed for programmatic data access and easy integration in available bioclimatic modelling workflows (Thuiller *et al.*, 2009; Naimi & Araújo, 2016). The extension of Bio-ORACLE aims to improve understanding of the potential impacts of climate change on marine organisms and support well-informed research and management.

Marine data layers

The current data descriptor provides information on the approach used to generate a set of essential variables for bioclimatic modelling. This comprises the acquisition of physical, chemical, biological and topographic data and the techniques used to generate variables for present-day conditions and the SSP scenarios of future climate change.

Data for present-day conditions (period 2000 to 2020) were acquired at a monthly basis from the Global Ocean Physics Reanalysis and Forecast and the Global Ocean Biogeochemistry Analysis and Forecast, two pre-processed re-analyses of Copernicus Marine Environment Monitoring Service (https://data.marine.copernicus.eu/products), provided at a 0.08º and 0.25º degree resolution with 50 and 75 vertical levels, respectively. These integrate and assimilate a comprehensive range of satellite and in-situ data (Jean-Michel *et al.*, 2021). Specifically for Photosynthetic Active Radiation and Diffuse Attenuation Coefficient, data were acquired as monthly averages from GlobColour (https://www.globcolour.info), a dataset at a ~0.04º resolution that merges MERIS, MODIS and SeaWiFS sensors (Maritorena *et al.*, 2010). Data for the SSP1-1.9, SSP1-2.6, SSP2-4.6, SSP3-7.0 and SSP5.8.5 scenarios (period 2000 to 2100) were acquired at the monthly basis from numerous ESMs provided by the CMIP phase 6 (Earth System Grid Federation), namely, ACCESS-ESM1-5 (Australian Community Climate and Earth System Simulator; 1.875×1.25º resolution and 38 vertical levels), CanESM5 (Canadian Earth System; 1º resolution and 45 vertical levels), CESM2-WACCM (Community Earth System Model 2 – Whole Atmosphere Community Climate Model; ~0.88x0.56º resolution and 60 vertical levels), CNRM-ESM2-1 (Centre National de Recherches Météorologiques; 0.25º resolution and 75 vertical levels), GFDL-ESM4 (Geophysical Fluid Dynamics Laboratory; 0.25º resolution and 75 vertical levels), GISS-E2-1-G (Goddard Institute for Space Studies; 1×1.25º resolution and 40 vertical levels), IPSL-CM6A-LR (Institut Pierre Simon Laplace; 1º resolution and 75 vertical levels), MIROC-ES2L (Model for Interdisciplinary Research on Climate; 1º resolution and 63 vertical levels), MPI-ESM1-2-LR (Max Plank Institute; 1.5º resolution and 40 vertical levels), MRI-ESM2-0 (Meteorological Research Institute; 1º resolution and 63 vertical levels) and UKESM1-0-LL (United Kingdom Earth System Model; 1º resolution and 75 vertical levels).

The monthly data of present-day and future conditions were used to generate six statistics per decade and variable: the average, maximum and minimum records of a given decade, long-term average of the yearly maxima and minima of a given decade (e.g., the average temperature of the warmest month in the period 2000-2010), and range, which represents the average absolute difference between the maximum and minimum records per year. These statistics were produced for air temperature, chlorophyll-a, diffuse attenuation coefficient, dissolved iron, dissolved oxygen, mixed layer depth, nitrate, ocean temperature, pH, phosphate, photosynthetically active radiation, primary productivity (total phytoplankton), total cloud fraction, salinity, silicate, sea-ice cover, sea-ice thickness and sea-water direction and velocity (Table 1).

To produce gridded layers with uniform spatial extent and resolution, the statistics generated for present-day conditions were interpolated to 0.05 degrees (approx. 5.5 km at the equator) with inverse distance weighting (IDW). This is a well-established algorithm in climate change research (Ozelkan *et al.*, 2016; Assis *et al.*, 2017b) and was chosen based on a performance and tuning test against the kriging algorithm (Supplement 1). The inverse distance weighting algorithm fitted the 8 nearest values of each focal cell with an inverse distance power of 2. For surface layers, the algorithm performed bilinear interpolation, considering the position of each cell (i.e., longitude and latitude) and the information comprised in the top vertical levels (i.e., defining the surface of the ocean) of each dataset (e.g., Copernicus) and variable. For benthic layers, the algorithm performed trilinear interpolation, by further considering the depth of each cell and the multiple vertical levels of each dataset and variable (Boavida *et al.*, 2016; Assis *et al.*, 2017b). To this end, the depth of each cell was extracted from the general bathymetric chart of the oceans (GEBCO_2023 Grid), a global terrain model providing elevation data at a 0.004 degrees resolution (GEBCO Bathymetric Compilation Group 2023, 2023). Because focal cells at 0.05º comprise a wide range of depth values, the benthic layers were developed for the minimum, average and maximum depth within focal cells, as in Bio-ORACLE version 2.0 (Assis *et al.*, 2017b). The process of interpolation does not add local detail in climate data, but rather smooths it across the generated layers.

The layers projecting future conditions were produced with the change-factor method (also known as the delta change method). This involved adding the changes projected in climate by the ESMs (as interpolated anomalies, i.e., differences) to high-resolution climatologies representing present-day conditions (Assis *et al.*, 2017b; Maraun *et al.*, 2017; Schoeman *et al.*, 2023). In detail, the method was based on (1) computing the climate change anomaly (i.e., difference) between the future and the historical conditions with data of each ESM, at their native spatial resolution, and for the period of 2000-2014 (ESM projections start in the year 2015); (2) interpolating the climate change anomalies to the common 0.05º resolution with inverse distance weighting to be uniform in extent and resolution, as described for the present-day statistics; (3) averaging the interpolated climate change anomalies of the multiple ESM; and (4) applying the averaged downscaled anomalies to the downscaled layers defining the conditions of the present-day period. This approach, considering the magnitude of climate change (i.e., the change-factor), allowed removing mean state biases of the ESM (Schoeman *et al.*, 2023), while providing a high confidence level in climate change projections (Hall & Hall, 2014), linking greenhouse gas emissions with global climate change with and without climate policy responses (Riahi *et al.*, 2017). The change-factor approach was not used in diffuse attenuation and photosynthetically active radiation, as these variables are not provided by the CMIP. The standard deviation of the ESM projected data was also computed (Supplement 2).

Additional topographic layers, namely, slope (expressing depth changes over distance), aspect (expressing the direction that slope faces), terrain ruggedness index (expressing depth changes between adjacent focal cells), topographic position index (comparing the depth of focal cells to the mean elevation of adjacent focal cells), as well as the minimum, average and maximum depth of each focal cell, were generated at 0.05 degrees resolution based on the general bathymetric chart of the oceans (GEBCO Bathymetric Compilation Group 2023, 2023).

All layers were archived as NetCDF (network Common Data Form) and deposited into an ERDDAP (Environmental Research Division's Data Access Program) server to facilitate filtering and downloading of the layers in common data formats (Wilson *et al.*, 2020). Additionally, Python (pyo_oracle) and R (biooracler) packages were developed for facilitated data retrieval and improved integration in available frameworks of bioclimatic modelling (Thuiller *et al.*, 2009; Naimi & Araújo, 2016). Such packages act as clients for ERDDAP's REST API, an interoperable web protocol for data transfer, and thus can be used for integration into most generic web-based applications.

Data Records

The dataset (Assis, 2023) comprises: (1) decadal time series of six statistics for each of 19 essential physical, chemical and biological variables at the global scale, for surface and benthic conditions, at a spatial resolution of 0.05 degrees, and a temporal resolution of 10 decadal steps, from 2000 to 2100, under the scenarios SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP4-6.0 and SSP5-8.5; (2) seven topographic layers at a spatial resolution of 0.05 degrees (Table 1).

Technical Validation

The reliability of data layers was estimated by means of cross-validation against quality-controlled data provided by the Global Ocean Data Analysis Project (GLODAP (Olsen *et al.*, 2016; Lauvset *et al.*, 2021)), for the variables available in this dataset (dissolved oxygen, nitrate, phosphate, salinity, silicate and temperature). Following additional applications (Davies & Guinotte, 2011; Assis *et al.*, 2015, 2017b, 2022a; Boavida *et al.*, 2016), GLODAP records, which are in situ observations, were compared at every given location with the data used to develop the layers. To this end, the data was interpolated with inverse distance weighting, as previously described, to the locations reported on GLODAP (i.e., longitude, latitude and depth). Then, the paired records were statistically analyzed with Pearson's correlation coefficient, root mean square error, and mean absolute error. The difference between paired records was also mapped globally onto a 2.5 degrees grid (average of differences within grid cells (Davies & Guinotte, 2011; Assis *et al.*, 2017b).

Overall, the layers mirrored the climatic patterns of quality-controlled data, as verified elsewhere (Assis *et al.*, 2017b, 2022a). All variables retrieved high correlation coefficients, ranging from 0.93 to 0.99, and low error rates (Supplement 3). Nitrate, phosphate, salinity and temperature displayed discrete anomalies restricted to specific regions (e.g., the transition between the warm and cold temperate Northwest Atlantic and Pacific oceans), with no relationship with depth. The errors for dissolved molecular oxygen and silicate were mostly distributed in the Southern Ocean, also with no relationship with depth (Supplement 3).

The marine data layers produced with information from the Global Ocean Physics Reanalysis and Forecast and the Global Ocean Biogeochemistry Analysis and Forecast from the Copernicus Marine Environment Monitoring Service, and the numerous ESMs from CMIP phase 6, have climate data in all gridded cells. However, 7.81 % of the cells of Photosynthetic Active Radiation and 4.22 % of the cells of Diffuse Attenuation Coefficient have missing information at latitudes above 70° N / S. This is due to missing satellite information from MERIS, MODIS and SeaWiFS sensors (Maritorena *et al.*, 2010) at such higher latitudes, as reported elsewhere (Tyberghein *et al.*, 2012).

Usage notes

This major update of Bio-ORACLE presents twenty-six physical, chemical, biological and topographic marine data layers, with global coverage and uniform grid system. These layers, provided with improved 0.05° spatial resolution (to better capture geomorphological features; Supplement 4) and per decade, have numerous potential applications in biogeography, ecology, evolutionary biology and climate change research, which can enhance our understanding of anthropogenic impacts on biodiversity and support well-informed conservation and management strategies. In particular, the layers can be used in Species Distribution Modeling (SDM) (Peterson *et al.*, 2011) to predict the distribution of species at the global scale (Chefaoui *et al.*, 2015; Fragkopoulou *et al.*, 2022), including non-native species (Assis *et al.*, 2015), address niche-based questions (Lee-Yaw *et al.*, 2016; Hu *et al.*, 2021; Song *et al.*, 2021) and phylogeographic hypotheses (Neiva *et al.*, 2014), identify biodiversity hotspots (Fragkopoulou *et al.*, 2022) and support the conservation and management of marine biodiversity (Hobday *et al.*, 2010; Boavida *et al.*, 2016). In the scope of

SDM, by letting users filter historical data into two time periods (decades 2000-2010 and 2010-2020), the current version allows generating independent data for temporal cross-validation, which can assist in evaluating model performance and prediction error (Ko *et al.*, 2013). Moreover, the development of biologically meaningful variables for future climate change scenarios (e.g., dissolved oxygen, primary productivity and pH) allows more realistic estimates of the anthropogenic pressures that may lead to extinction and turnover of populations (Martins *et al.*, 2021; Assis *et al.*, 2022b; Gouvêa *et al.*, 2022).

The marine layers can be used in additional analyses beyond SDM, such as those based on univariate algorithms of climate velocity (Burrows *et al.*, 2014) or multivariate algorithms of climate analogs (Mahony *et al.*, 2017), to estimate biodiversity exposure to climate change (Mackintosh *et al.*, 2023), or identify climate connectivity corridors promoting future distribution range shifts (Garciá Molinos *et al.*, 2016). In this scope, the availability of the Shared Socioeconomic Pathway scenarios, comprising the mitigation strategies of the Paris Agreement, coupled with the ensemble of multiple ESMs, allows supporting international climate policies and agendas (Martins *et al.*, 2021). The increased interoperability via the ERDDAP data server (Wilson *et al.*, 2020) opens the door for Bio-Oracle to be used in digital twinning, such as the European Digital Twin of the Ocean, or private initiatives like the Microsoft Planetary Computer. This aligns with the United Nations Decade of Ocean Science for Sustainable Development, which aims to boost the development of Digital Twins of the Ocean for supporting blue growth and global governance.

Data availability statement

The layers are openly available at the ERDDAP instance hosted at the Flanders Marine Institute Data Centre (https://erddap.bio-oracle.org/erddap), through the Bio-ORACLE website (temporary link for peer-review: https://www.bio-oracle.org/downloads-to-email-v3.php) and the R and Python packages available in the permanent GitHub repository (https://github.com/bio-oracle).

Code availability

The R language code used to generate the layers is available in the permanent GitHub repository at https://github.com/bio-oracle/data-layers.

Author contributions

J.A., L.T., H.V. and O.D.C. conceived the study. J.A. designed the data pipelines. J.A. wrote the manuscript. V.S., F. L. and B. V. set up the ERDDAP server and F.L. and B.V. will maintain it. S.F. and V. S. wrote the documentation, published the dataset on ERDDAP and developed the R and Python clients. All authors reviewed the manuscript.

Competing interests

The authors declared no conflict of interest.

Tables

Table 1. Marine data layers and units, realm (surface layers, benthic layers, or both, where appropriate), accuracy assessed with in situ quality-control data (for the variables comprised in GLODAP (Olsen *et al.*, 2016; Lauvset *et al.*, 2021); MAE: mean absolute error; RMSE: root mean square error; Cor: Pearson's correlation; additional estimates per depth in Supplement 3), number of quality-control records (n), range of values for present conditions, and change (anomaly) between the present (decade 2010-2020) and the future (decade 2090-2100) under contrasting scenarios SSP1-1.9 and SSP5-8.5.

Layer	Surface	Benthic	MAE	RMSE	Cor	n	Range	Future change (SSP1-1.9; SSP5-8.5)
Air temperature (ºC)	X		-	-	-	-	-57.96; 40.97	0.72±1.08; 4.67±3.53
Chlorophyll-a concentration (mmol · m^{-3})	X	X	-	-	-	-	0; 21.95	0±0.03; 0.01±0.03
Diffuse attenuation coefficient (m^{-1})	X		-	-	-	-	0; 1.19	-
Dissolved iron concentration (mmol · m^{-3})	X	X	-	-	-	-	0; 0.101	-0.01±0.01; -0.01±0.01
Dissolved oxygen concentration (mmol · m^{-3})	X	X	5.222	25.648	0.930	558,720	41.70; 456.98	-1.26±4.93; -14.33±11.33
Mixed layer depth (m)	X		-	-	-	-	0; 3728.89	0.27±11.36; -7.19±19.35
Nitrate concentration (mmol · m^{-3})	X	X	0.467	3.082	0.946	642,133	0; 303.75	-0.38±0,91; -1.11±1.29
Ocean temperature (ºC)	X	X	0.141	0.823	0.991	823,531	-1.94; 36.49	0.38±0.40; 2.97±0.98
pH	X	X	-	-	-	-	7; 8.53	-0.03±0.02; -0.39±0.04
Phosphate concentration (mmol · m^{-3})	X	X	0.028	0.218	0.948	612,203	0; 4.38	-0.03±0.09; -0.10±0.11
Photo. active radiation (E · m^{-2} · yr^{-3})	X		-	-	-	-	0; 69.13	-
Total phytoplankton concentration (mmol · m^{-3})	X	X	-	-	-	-	0.02; 57.20	-0.01±0.16; -0.10±0.29
Total cloud fraction (fraction)	X		-	-	-	-	0; 1	0.01±0.04; -0.01±0.05
Salinity	X	X	0.012	0.217	0.974	824,173	0; 47.54	-0.01±0.32; -0.22±0.51
Silicate concentration (mmol · m^{-3})	X	X	1.762	9.484	0.929	645,495	0.23; 680.78	-0.83±3.33; -1.60±3.53
Sea-Ice Cover (fraction)	X		-	-	-	-	0; 1	-0.02±0.05; -0.09±0.20
Sea-Ice Thickness (m)	X		-	-	-	-	0; 7.87	-0.13±0.37; -0.28±0.68
Sea-water direction (º)	X	X	-	-	-	-	0; 360	0.45±40.66; 2.59±45.39
Sea-water velocity (m / s)	X	X	-	-	-	-	0; 3.05	0.02±0.05; 0.02±0.05
Topographic slope		X	-	-	-	-	0; 30.71	-
Topographic aspect		X	-	-	-	-	0; 360	
Terrain ruggedness index		X	-	-	-	-	0; 2512.92	-

Topographic position index		X	-	-	-	-	-1963; 2512	-
Minimum depth (m)		X	-	-	-	-	-10363; 0	-
Average depth (m)		X	-	-	-	-	-10699; 0	-
Maximum depth (m)		X	-	-	-	-	-10977; 0	-

Table 2. List of R and Python functions to facilitate listing and extraction of data layers (refer to the permanent GitHub repository for additional information; https://github.com/bio-oracle).

Function	Language	Description
list_layers()	R and Python	Lists the data layers available in the Bio-ORACLE ERDDAP server, either as a list or a dataframe containing metadata. Users may subset layers based on attributes, such as variable or SSP scenario.
download_layers()	R and Python	Downloads one or more data layers from the Bio-ORACLE ERDDAP server. Users may filter data using attributes, such as sets of coordinates or time periods, and select the data format to be downloaded*.
list_local_data()	R and Python	Lists local data that has been downloaded by the Bio-ORACLE client.
config	R and Python	Shows configuration values, *i.e.* the path to the local data directory and the address of the Bio-ORACLE server.

* Refer to the ERDDAP server for the complete list of data formats (https://erddap.bio-oracle.org/erddap/griddap/documentation.html#fileType).

Additional information

Supplementary information 1: Performance and tuning test between inverse distance weighting and kriging algorithms.

Supplementary information 2: Standard deviation of the ensemble of Earth System Models.

Supplementary information 3: Reliability of climate layers estimated with cross-validation.

Supplementary information 4: Comparison between the current and the previous spatial resolution of Bio-ORACLE datasets (version 2.0 vs version 3.0).

Supplementary information 5: Availability of Earth System Models per variable and shared socioeconomic pathway scenario.

References

Arafeh-Dalmau, N., Brito-Morales, I., Schoeman, D.S., Possingham, H.P., Klein, C.J. & Richardson, A.J. (2021) Incorporating climate velocity into the design of climate-smart networks of marine protected areas. *Methods in Ecology and Evolution*, **12**, 1969–1983

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Arneth, A., Shin, Y.J., Leadley, P., Rondinini, C., Bukvareva, E., Kolb, M., Midgley, G.F., Oberdorff, T., Palomo, I. & Saito, O. (2020) Post-2020 biodiversity targets need to embrace climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **117 (49)**, 30882-30891.

Assis, J. (2023) Bio-ORACLE, Marine data layers for ecological modelling. *https://www.bio-oracle.org*.

Assis, J., Berecibar, E., Claro, B., Alberto, F., Reed, D., Raimondi, P. & Serrão, E.A. (2017a) Major shifts at the range edge of marine forests: the combined effects of climate changes and limited dispersal. *Scientific Reports*, **7 (44348)**, 1–10.

Assis, J., Neiva, J., Bolton, J.J., Rothman, M.D., Gouveia, L., Paulino, C., Mohdnasir, H. & Serrão, E.A. (2022a) Ocean currents shape the genetic structure of a kelp in southwestern Africa. *Journal of Biogeography*, **01**, 1–14.

Assis, J., Serrão, E.A., Duarte, C.M., Fragkopoulou, E. & Krause-Jensen, D. (2022b) Major expansion of marine forests in a warmer Arctic. *Frontiers in Marine Science*, **9**.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E.A. & De Clerck, O. (2017b) Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, **27**, 277–284.

Assis, J., Zupan, M., Nicastro, K.R., Zardi, G.I., McQuaid, C.D. & Serrão, E.A. (2015) Oceanographic Conditions Limit the Spread of a Marine Invader along Southern African Shores. *Plos One*, **10**, e0128124.

Boavida, J., Assis, J., Silva, I. & Serrão, E.A. (2016) Overlooked habitat of a vulnerable gorgonian revealed in the Mediterranean and Eastern Atlantic by ecological niche modelling. *Scientific Reports*, **6**, 36460.

Burgess, M.G., Becker, S.L., Langendorf, R.E., Fredston, A. & Brooks, C.M. (2023) Climate change scenarios in fisheries and aquatic conservation research. *ICES Journal of Marine Science*, **80**, 1163–1178.

Burrows, M.T., Schoeman, D.S., Richardson, A.J., Molinos, J.G., Hoffmann, A., Buckley, L.B., Moore, P.J., Brown, C.J., Bruno, J.F., Duarte, C.M., Halpern, B.S., Hoegh-Guldberg, O., Kappel, C. V, Kiessling, W., O'Connor, M.I., Pandolfi, J.M., Parmesan, C., Sydeman, W.J., Ferrier, S., Williams, K.J. & Poloczanska, E.S. (2014) Geographical limits to species-range shifts are suggested by climate velocity. *Nature*, **507**, 492–5.

Chefaoui, R.M., Assis, J., Duarte, C.M. & Serrão, E.A. (2015) Large-Scale Prediction of Seagrass Distribution Integrating Landscape Metrics and Environmental Factors: The Case of Cymodocea nodosa (Mediterranean–Atlantic). *Estuaries and Coasts*, **39**, 123–137.

Davies, A.J. & Guinotte, J.M. (2011) Global Habitat Suitability for Framework-Forming Cold-Water Corals. *Plos One*, **6**, 1–15.

de la Hoz, C.F., Ramos, E., Acevedo, A., Puente, A., Losada, J.I. & Juanes, J.A. (2018) OCLE: A European open access database on climate change effects on littoral and oceanic ecosystems, *Progress in Oceanography*, **168**, 222-231

Fautin, D.G. & Buddemeier, R.W. (2008) Biogeoinformatics of the hexacorals.

Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302-4315.

Fragkopoulou, E., Serrão, E.A., Clerck, O. De, Costello, M.J., Araújo, M.B., Duarte, C.M., Krause-Jensen, D. & Assis, J. (2022) Global biodiversity patterns of marine forests of brown macroalgae. *Global Ecology & Biogeography*, **31(4)**, 636-648.

Fragkopoulou, E., Serrão, E.A., Horta, P.A., Koerich, G. & Assis, J. (2021) Bottom Trawling Threatens Future Climate Refugia of Rhodoliths Globally. *Frontiers in Marine Science*, **7**, 1–11.

Garciá Molinos, J., Halpern, B.S., Schoeman, D.S., Brown, C.J., Kiessling, W., Moore, P.J., Pandolfi, J.M., Poloczanska, E.S., Richardson, A.J. & Burrows, M.T. (2016) Climate velocity and the future global redistribution of marine biodiversity. *Nature Climate Change*, **6(1)**, 83-88.

GEBCO Bathymetric Compilation Group 2023 (2023) *The GEBCO_2023 Grid - a continuous terrain model of the global oceans and land*, (ed. by N.O.C.N. British Oceanographic Data Centre) British Oceanographic Data Centre, UK.

Gouvêa, L.P., Assis, J., Gurgel, C.F.D., Serrão, E.A., Silveira, T.C.L., Santos, R., Duarte, C.M., Peres, L.M.C., Carvalho, V.F., Batista, M., Bastos, E., Sissini, M.N. & Horta, P.A. (2020) Golden carbon of Sargassum forests revealed as an opportunity for climate change mitigation. *Science of The Total Environment*, **729**, 138745.

Gouvêa, L.P., Serrão, E.A., Cavanaugh, K., Gurgel, C.F.D., Horta, P.A. & Assis, J. (2022) Global impacts of projected climate changes on the extent and aboveground biomass of mangrove forests. *Diversity and Distributions*, **28**, 2349–2360.

Hagedorn, R., Doblas-Reyes, F.J. & Palmer, T.N. (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus, Series A: Dynamic Meteorology and Oceanography*, **57**, 219-233.

Hall, B.A. & Hall, A. (2014) Projecting regional change. *Science*, **346**, 1461–1462.

Hausfather, Z., Marvel, K., Schmidt, G.A., Nielsen-Gammon, J.W. & Zelinka, M. (2022) Climate simulations: recognize the 'hot model' problem. *Nature*, **605**.

Hobday, A.J., Hartog, J.R., Timmiss, T. & Fielding, J. (2010) Dynamic spatial zoning to manage southern bluefin tuna (Thunnus maccoyii) capture in a multi-species longline fishery. *Fisheries Oceanography*, **19**, 243–253.

Hu, Z.-M., Zhang, Q.-S., Zhang, J., Kass, J.M., Mammola, S., Fresia, P., Draisma, S.G.A., Assis, J., Jueterbock, A., Yokota, M. & Zhang, Z. (2021) Intraspecific genetic variation matters when predicting seagrass distribution under climate change. *Molecular Ecology*, **30(15)**, 3840-3855.

Jean-Michel, L., Eric, G., Romain, B.B., Gilles, G., Angélique, M., Marie, D., Clément, B., Mathieu, H., Olivier, L.G., Charly, R., Tony, C., Charles-Emmanuel, T., Florent, G., Giovanni, R., Mounir, B., Yann, D. & Pierre-Yves, L.T. (2021) The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. *Frontiers in Earth Science*, **9**, 698876.

Ko, C.Y., Ko, C.J., Lin, R.S. & Lee, P.F. (2013) Influences of temporal independence of data on modelling species distributions. *Basic and Applied Ecology*, **14**, 309-319.

Krumhardt, K.M., Lovenduski, N.S., Long, M.C. & Lindsay, K. (2017) Avoidable impacts of ocean warming on marine primary production: Insights from the CESM ensembles. *Global Biogeochemical Cycles*, **31**, 114–133.

Lauvset, S.K., Lange, N., Tanhua, T., Bittig, H.C., Olsen, A., Kozyr, A., Álvarez, M., Becker, S., Brown, P.J., Carter, B.R., Cotrim Da Cunha, L., Feely, R.A., Van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jutterström, S., Jones, S.D., Karlsen, M.K., Lo Monaco, C., Michaelis, P., Murata, A., Pérez, F.F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Tilbrook, B., Velo, A., Wanninkhof, R., Woosley, R.J. & Key, R.M. (2021) An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2021. *Earth System Science Data*, **13**.

Lee-Yaw, J.A., Kharouba, H.M., Bontrager, M., Mahony, C., Csergo, A.M., Noreen, A.M.E., Li, Q., Schuster, R. & Angert, A.L. (2016) A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. *Ecology Letters*, **19**, 710–722.

Levitus, S., Antonov, J.I., Baranova, O.K., Boyer, T.P., Coleman, C.L., Garcia, H.E., Grodsky, a I., Johnson, D.R., Locarnini, R. a, Mishonov, a V, Reagan, J.R., Sazama, C.L., Seidov, D., Smolyar, I., Yarosh, E.S. & Zweng, M.M. (2013) The World Ocean Database. *Data Science Journal*, **12**, WDS229–WDS234.

Mackintosh, A., Goldstein Hill, G., Costello, M., Jueterbock, A. & Assis, J. (2023) Modeling Aquaculture Suitability in a Climate Change Future. *Oceanography*, **4**, 211–216.

Mahony, C.R., Cannon, A.J., Wang, T. & Aitken, S.N. (2017) A closer look at novel climates: new methods and insights at continental to landscape scales. *Global Change Biology*, **23(9)**, 3934-3955.

Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A. & Mearns, L.O. (2017) *Towards process-informed bias correction of climate change simulations*. *Nature Climate Change*,.

Maritorena, S., d'Andon, O.H.F., Mangin, A. & Siegel, D.A. (2010) Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, **114**, 1791-1804.

Martins, M.R., Assis, J. & Abecasis, D. (2021) Biologically meaningful distribution models highlight the benefits of the Paris Agreement for demersal fishing targets in the North Atlantic Ocean. *Global Ecology & Biogeography*, **30**, 1643–1656.

Mechoso, C.R. & Arakawa, A. (2015) *Numerical Models: General Circulation Models*. *Encyclopedia of Atmospheric Sciences: Second Edition*.

Naimi, B. & Araújo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, **39**, 368–375.

Neiva, J., Assis, J., Fernandes, F., Pearson, G.A. & Serrão, E.A. (2014) Species distribution models and mitochondrial DNA phylogeography suggest an extensive biogeographical shift in the high-intertidal seaweed Pelvetia canaliculata. *Journal of Biogeography*, **41**, 1137–1148.

Olsen, A., Key, R.M., Van Heuven, S., Lauvset, S.K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F.F., Suzuki, T., Heuven, S. Van, Lauvset, S.K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Van Heuven, S., Lauvset, S.K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F.F., Suzuki, T., Heuven, S. Van, Lauvset, S.K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M. & Jutterström, S. (2016) The global ocean data analysis project version 2 (GLODAPv2) - An internally consistent data product for the world ocean. *Earth System Science Data*, **8**, 297–323.

Ozelkan, E., Chen, G. & Ustundag, B.B. (2016) Spatial estimation of wind speed: a new integrative model using inverse distance weighting and power law. *International Journal of Digital Earth*, **9(8)**, 733-747.

Palmer, T.N., Doblas-Reyes, F.J., Hagedorn, R. & Weisheimer, A. (2005) *Probabilistic prediction of climate using multi-model ensembles: From basics to applications*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1463.

Peterson, A.T.T., Soberón, J., Pearson, R.G.R.G., Anderson, R.P.R.P., Martínez-Meyer, E., Nakamura, M. & Bastos Araújo, M. (2011) *Ecological Niches and Geographic Distributions*, Monographs. (ed. by S.E. Simon A. Levin and Henry S. Horn) Princeton University Press, Princeton.

Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L.A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A. & Tavoni, M. (2017) The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, **42**, 153-168.

Sanderson, B.M., O'Neill, B.C. & Tebaldi, C. (2016) What would it take to achieve the Paris temperature targets? *Geophysical Research Letters*, **43**, 7133–7142.

Sbrocco, E.J. & Barber, P.H. (2013) MARSPEC: ocean climate layers for marine spatial ecology. *Ecology*, **94**, 979.

Schoeman, D.S., Gupta, A. Sen, Harrison, C.S., Everett, J.D., Brito-Morales, I., Hannah, L., Bopp, L., Roehrdanz, P.R. & Richardson, A.J. (2023) Demystifying global climate models for use in the life sciences. *Trends in Ecology and Evolution*, **38(9)**, 843-858.

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., Kwiatkowski, L., Aumont, O., Christian, J., Dunne, J., Gehlen, M., Ilyina, T., John, J.G., Li, H., Long, M.C., Luo, J.Y., Nakano, H., Romanou, A., Schwinger, J., Stock, C., Santana-Falcón, Y., Takano, Y., Tjiputra, J., Tsujino, H., Watanabe, M., Wu, T., Wu, F. & Yamamoto, A. (2020) Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. *Current Climate Change Reports*, **6(3)**, 95-119.

Shi, Z., Assis, J. & Costello, M.J.B.T.-R.M. in E.S. and E.S. (2021) *Vulnerability of Marine Species to Low Oxygen Under Climate Change*. *Reference Module in Earth Systems and Environmental Sciences*, Elsevier, 887-894.

Song, X.-H., Assis, J., Zhang, J., Gao, X., Choi, H.-G., Duan, D.-L., Serrão, E.A. & Hu, Z.-M. (2021) Climate-induced range shifts shaped the present and threaten the future genetic variability of a marine brown alga in the Northwest Pacific. *Evolutionary Applications*, **14(7)**, 1867-1879.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - A platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F. & De Clerck, O. (2012) Bio-ORACLE: A global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, **21**, 272–281.

van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S.J. & Rose, S.K. (2011) The representative concentration pathways: An overview. *Climatic Change*, **109**, 5.

van Vuuren, D.P., Riahi, K., Calvin, K., Dellink, R., Emmerling, J., Fujimori, S., KC, S., Kriegler, E. & O'Neill, B. (2017) The Shared Socio-economic Pathways: Trajectories for human development and global environmental change. *Global Environmental Change*, **42**, 148-152.

Vega, C., Pertierra, G., & Olalla-Tárraga, M. (2017) MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. *Scientific Data*, **4**, 170078.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C. ’t, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. a., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, **3**, 160018.

Wilson, C., Robinson, D. & Simons, R.A. (2020) *Erddap: Providing Easy Access to Remote Sensing Data for Scientists and Students*. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3207-3210.