Minor Revision

MS: GEB-2017-0341 by Assis et al.

Title: Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling

Dear Editor,

Thank you for your letter concerning our manuscript (GEB-2017-0341) providing marine data layers for bioclimatic modelling. We have made our best effort to modify the manuscript according to all comments and suggestions made by the editors and reviewers. Below we provide a point-by-point response to the comments.

Jorge Assis

-

EDITOR'S COMMENTS TO AUTHORS

Editor: Tittensor, Derek

Comments to the Author:

Bio-ORACLE v 1.0 was a resource for the marine modelling community that synthesized and combined environmental data layers at a consistent spatial resolution. The present manuscript extends this, and furthermore adds an R-package to assist with making the data more accessible. This is an notable improvement and advancement, and a potentially excellent resource. I am also particularly happy to see that the authors attempt to assess the biases and error associated with their interpolations.

Both reviews were positive, and I think that only fairly minor revisions are needed. I urge the authors to address both sets of comments, and note in particular the suggestions of Reviewer 1 on describing the environmental policy context that generates the need for such layers to help inform models for global decision-making, as well as clarifying the key strengths of this data resource versus others.

Dr. Derek Tittensor, Editor

Response to editor:

Thank you for handling our manuscript (GEB-2017-0341). All comments were carefully addressed. Please refer to the section below "Reviewer comments to authors".

-

REVIEWER COMMENTS TO AUTHORS

Referee: 1

Referee's comment 1.1

This manuscript presents an update and extension of the work presented in this same journal in 2012. Having used the version 1 data layers myself, I am pleased to see that more will now be available to users, and not only to species distribution modellers. The fact that bioclimatic analyses can now be supported by these layers is very timely, given the current needs to better predict and communicate the possible effects of climate change on marine biodiversity.

I have not spotted any reference to the current marine biodiversity policy context (e.g. on-going UN negotiations for the Biodiversity Beyond Jurisdiction Agreement BBNJ, the Intergovernmental science-policy Platform on Biodiversity and Ecosystem Services process IPBES, World Ocean Assessment, etc) in this manuscript, and maybe this is something that would provide the big picture? Equally, maybe this is not something that the readership of this specific journal would expect (maybe the Editor could advise)? I pesonally would like a link to be made, e.g. given that the creation of IPBES was greatly inspired by the successes of the IPPC's work. The BIO-Oracle's layer have the potential to help fill the marine biodiversity gaps in knowledge through improved modelling.

Response to reviewer #1.1:

We thank the reviewer by bringing our attention to this point. We now indicate how improved bioclimatic modelling may guide the integration of climate change in conservation strategies, policies and baseline assessments.

Changes in manuscript #1.1:

We added the sentence (Conclusion section):

"The new data layers represent a valuable addition to the spatial information about climate available for the global ocean. Their key features can be used to improve bioclimatic modelling and provide valuable insights into the current and future states of marine biodiversity and indirectly into the services it provides to society. When requested by decision makers, these outcomes may guide important climate change integrated conservation strategies (Hannah et al., 2002; Hobday et al., 2010), and feed baseline

assessments such as those required for the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES)."

Referee's comment 1.2. I have made some comments (plus edits in the Abstract) in the PDF version of the manuscript, which I hope will help the authors to improve it. On a minor point, please define all acronyms...

Response to reviewer #1.2

We thank the reviewer for the comments and edits made in the PDF version of the manuscript. All changes were carefully undertaken with the exception of the title of the manuscript. The reviewer reasons that "the title does not do justice to the manuscript's contents" as it provides "much more than v1 (e.g. iron, phytoplankton...), of relevance to species distribution modellers who are not interested in bioclimatic modelling (future conditions)".The reviewer's argument associates "bioclimatic modelling" to the strict exercise of predicting the distribution of species through time. In our opinion, this is a more generic expression beyond species distribution modelling, as in principle, it is *defined as the association between the environmental aspects and the occurrence of species, allowing to uncover the conditions under which species are likely to maintain viable populations* (Araújo & Peterson, 2012)*. We don't want to enter in the open conceptual debate about this expression, however, it seemed to us that the expression fitted better to address the broad research describing the distribution of species, predict range shifts through time, disentangle the likely anthropogenic pressures leading to population turnover and extinction, address niche-based questions, support biodiversity conservation and management or test other evolutionary hypotheses (e.g., paleoclimatic modeling). Accordingly, we would like to keep the title as it is, however, if the editors or reviewers think we should change it, we are willing to.*

Changes in manuscript #1.2

All points outlined in the PDF version of the manuscript were carefully corrected, with the exception of the title (please refer to the manuscript file with track changes).

Referee's comment 1.3. From a user perspective, packaging these layers into user-friendly formats is key, and offering an R package was an excellent move. Authors might want to consider how they could further promote the use of their dataset by a broader audience on non-GIS specialists. For instance, I have found no mention of a place where these data could be viewed online (see example of the Ocean Data Viewer: http://data.unep-wcmc.org/datasets/36).

Response to reviewer #1.3

We acknowledge the comment of the reviewer and therefore we now provide a simple way for non-GIS specialists to explore (i.e., view) the data layers. This was implemented in the new section "Explore data" of the website where the data is accessible.

Changes in manuscript #1.3

We changed the sentence (Data accessibility section) from:

"The Bio-ORACLE layers are accessible online at http://www.bio-oracle.org as ASCII and TIFF files."

to:

"The Bio-ORACLE layers are accessible online at http://www.bio-oracle.org for download (as ASCII and TIFF files) and preview."

Referee's comment 1.4. I have found that the manuscript is not clear enough regarding the key strengths of the proposed product: for instance, Chlorophyll data are available on numerous web site (e.g. NASA). The BIO-Oracle product goes further by providing a range of cross-compatible (resolution, gridding, etc) layers for seveal variables, that are ready for use.

Response to reviewer #1.4

We acknowledge the concern of the reviewer and therefore we now better synthetize the key strengths of the current update to Bio-ORACLE.

Changes in manuscript #1.4

We changed the sentence (Conclusion section) from:

"Like the first version of Bio-ORACLE, data layers have been produced with bioclimatic modelling in mind and are provided in a user-friendly format. Their global coverage, comparable grid system and comprehensive set of ecologically relevant variables now covering the benthic realm expands the potential of Bio-ORACLE."

to:

"A comprehensive set of new ecologically relevant data layers are provided for bioclimatic modelling in a user-friendly format, with global coverage and comparable grid system. The current update expands the potential of Bio-ORACLE by (1) covering the benthic realm with climate data, (2) adding new variables to present-day conditions (i.e., current velocity, iron, light, phytoplankton, primary productivity and sea ice), (3) adding the new generation of climate change scenarios, (4) providing means of data reliability and uncertainty, as well as (5) a package of functions in the R software environment."

Referee's comment 1.5 Finally, as mentioned above, I had used version 1 in the past, and the main obstacle we had to deal with, that might not necessarily be unique to using BIO-Oracle layers is about a "coastal gap", i.e. a spatial gap of "no data" between the end of the dataset and the start of land. Can you recommend a coastline dataset (e.g. https://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html) that works best with BIO-Oracle data layers, i.e. that minimises this spatial gap? This would be useful to users of the dataset.

Response to reviewer #1.5

We acknowledge the comment of the reviewer and therefore we now recommend the use of GSHHG high-resolution geography dataset to minimize the possible spatial gap between the outputs using the provided data layers and land.

Changes in manuscript #1.5

We added the sentence (Marine data layers section):

"To minimize the possible spatial gap with "no data" between the data layers provided and the available vector shorelines, the global self-consistent, hierarchical, high-resolution geography database (Wessel & Smith, 1996) is recommended"

Referee: 2

Referee's comment 2. Comments to the Author:

The paper titled "Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling" is a well written and clear example of the increasing interest by the international research community to accelerate the collection and synthetization of environmental covariates in the marine realm. This study takes a further step and encourages and supports the collection, dissemination and use of environmental covariates in the third spatial dimension and across time with different climate change scenarios. From the introduction of the manuscript, it becomes apparent how the authors intend to bridge a knowledge gap which is crucial for the creation of environmental niche models and other types of oceanographic geospatial analysis for management/ conservation/ research/ industrial purposes. They do a good job at explaining the strengths and weaknesses of the data presented.

While the availability of biophysical parameters in the ocean is increasing, there are still major deficits which will take years to fill. This effort to aggregate some of the available environmental data at higher spatiotemporal resolutions will help researchers make the most out of the existing data. While there are clear gaps in the spatial coverage of some of the in situ environmental variables, which would help further validate the validation analysis, the authors do a good job at identifying those areas where the reliability of their predictions was weaker; thus, warning potential users of their products of the areas of higher uncertainty in the data. This was one of my main concerns while reading the paper and believe that the authors, with the support of the Supplementary Information (1), have done a good job at explaining the limitations of the products they are presenting and the spatial distribution of the areas where their interpolations do not match the in situ data collected. Furthermore, the R-package tool and the instructions provided for its use are very useful and user-friendly; I am convinced that marine geospatial modelers will use this resource widely when studying the biogeography and ecology of species at regional and global scales.

Below I include some minor comments and suggestions, which I hope will help towards the improvement of the manuscript.

Referee's comment 2.1. - Having access to high spatial and temporal resolution environmental data will allow for cross-sectorial management of ocean resources and the implementation of ecosystem-based management measures. Authors may want to consider further emphasizing the need for high spatiotemporal resolution marine environmental variables by referencing some of the existing literature on Dynamic Ocean Management:

(e.g. Hobday, A.J., Hartog, J.R., Timmiss, T. and Fielding, J., 2010. Dynamic spatial zoning to manage southern bluefin tuna (Thunnus maccoyii) capture in a multi-species longline fishery. Fisheries Oceanography, 19(3), pp.243-253.)

Response to reviewer #2.1:

We are grateful for the reference provided by the reviewer emphasizing the need for high resolution marine environmental variables for implementation of ecosystem-based management.

Changes in manuscript #2.1:

We changed the sentence (Introduction section) from:

"… support biodiversity conservation and management (e.g., Guisan & Thuiller 2005; Guisan et al. 2013; Boavida et al. 2016)…"

to:

"… support biodiversity conservation (e.g., Guisan & Thuiller 2005; Guisan et al. 2013; Boavida et al. 2016) and ecosystem-based management (Hobday et al. 2010), …"

Referee's comment 2.2 - Cross-validating the interpolation methods using other dynamic and static environmental variables (in addition to temperature) may help support this methodology. I believe this would be useful given the differences in spatiotemporal properties and stability of the variables included in the study.

Response to reviewer #2.2

We acknowledge the comment of the reviewer and therefore we have extended the performance test of downscaling methods to different variables.

Changes in manuscript #2.2

We included a new table in Supporting Information (Table S1.1) with the statistical downscaling performance of Kriging and Inverse Distance Weighting (IDW) for different variables.

We changed the sentence (Marine data layers section) from:

"This test cross-validated the interpolation of 1x10^4 random records of ocean temperature with both methods and showed lower root mean square error (RMSE) and mean absolute error (MAE) for Kriging (RMSE: 0.13°C vs. 0.15^{4}C; MAE: 0.06^{4}C vs. 0.07^{4}C), despite the lack of differences in the mean temperatures (nonparametric Kruskal-Wallis p-value: 0.987)."

to:

"This test cross-validated the interpolation of 1x10^4 random records with both methods for different variables, and showed lower root mean square error (RMSE) and mean absolute error (MAE) for Kriging, despite the lack of differences in the mean value of all variables (nonparametric Kruskal-Wallis p-values > 0.05; see Table S1.1 in Supporting Information)"

Referee's comment 2.3 - Authors may want to suggest using their interpolated surfaces to inform which areas of the ocean require better spatial coverage for the collection of in situ samples.

Response to reviewer 2.3:

We acknowledge the comment of the reviewer and therefore we outline the areas requiring better coverage of in situ samples.

Changes in manuscript #2.3:

We added the sentence (in the section on Reliability of marine data layers):

"The spatial distribution of errors further showed that dissolved molecular oxygen, phosphate, salinity, silicate and temperature have good coverage of in situ samples (GLODAP dataset), while nitrate and chlorophyll are mostly uncovered throughout the globe."

Literature cited

Araújo, M.B. & Peterson, A.T. (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**, 1527–1539.

Title

Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling

Running title

Marine data layers for bioclimatic modelling

Jorge Assis[1]*, Lennert Tyberghein[2], Samuel Bosch[2,3], Heroen Verbruggen[4], Ester A. Serrão[1], Olivier De Clerck[3]

[1] Centre for Marine Sciences, CCMAR-CIMAR, University of Algarve, Campus Gambelas, 8005-139 Faro, Portugal.

[2] Flanders Marine Institute (VLIZ), InnovOcean site, Wandelaarskaai 7, 8400 Ostend, Belgium

[3] Phycology Research Group, Biology Department, Ghent University, 9000 Ghent, Belgium.

[4] School of BioSciences, University of Melbourne, Victoria 3010, Australia.

* Corresponding author (jorgemfa@gmail.com | +351 912 361 127)

Acknowledgements

This study was supported by the Pew Foundation (EAS) and the Foundation for Science and Technology (FCT) of Portugal through a fellowship to Jorge Assis (SFRH/BPD/111003/2015) and projects PTDC/MAR-EST/6053/2014, BIODIVERSA/004/2015, CCMAR/Multi/04326/2013. Samuel Bosch and Olivier De Clerck are indebted to EU FP7 ERANET (Project SEAS-ERA/INVASIVES SD/ER/010). Heroen Verbruggen was supported by the Australian Research Council (FT110100585). The datasets ARMOR, ORAP and PISCES were made available by the E.U. Copernicus Marine Service Information (http:// marine.copernicus.eu). The GlobColour dataset (http://globcolour.info) used in this study has been developed, validated, and distributed by ACRI-ST, France. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (listed in Table 2 of this paper) for producing and making available their model output. 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.

Abstract

Motivation: The availability of user-friendly, high-resolution global environmental datasets is crucial for bioclimatic modelling. For terrestrial environments, WorldClim has served this purpose since 2005 but equivalent marine data only became available in

2012, with pioneer initiatives like Bio-ORACLE providing data layers for several ecologically relevant variables. Currently, the available marine data packages have not yet been updated to the most recent Intergovernmental Panel on Climate Change (IPCC) predictions nor to present times, and are mostly restricted to the top surface layer of the oceans, precluding modelling of a large fraction of the benthic diversity that inhabits deeper habitats. To address this gap, we present a significant update of Bio-ORACLE for new future climate scenarios, present-day conditions, and benthic layers (near sea bottom). The reliability of data layers was assessed using a cross-validation framework against *in situ* quality controlled data. This test showed a generally good agreement between our data layers and the global climatic patterns. We further provide a package of functions in the R software environment (*sdmpredictors*) to facilitate listing, extraction and management of data layers and allow easy integration with the available pipelines for bioclimatic modelling.

Main types of variable contained: Surface and benthic layers for water temperature, salinity, nutrients, chlorophyll, sea ice, current velocity, phytoplankton, primary productivity, iron and light at bottom.

Spatial location and grain: Global at 5 arcmin (approximately 0.08 degrees or 9.2 km at the equator).

Time period and grain: Present (2000-2014) and future (2040-2050 and 2090-2100) environmental conditions based on monthly averages.

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

Software format: ASCII and TIFF grid formats for geographic information systems and a package of functions developed for R software.

Keywords

Bio-ORACLE, bioclimatic modelling, environmental data, global, kriging, macroecology, marine, species distribution modelling.

Introduction

Early attempts to model the relationship between the occurrence or abundance of species and their natural environment relied heavily on environmental variables measured *in situ* (Sutherst & Maywald, 1985) and often involved complex software pipelines specifically developed to extract, organize and visualize data (Kemp *et al.*, 2012; Lima-Ribeiro *et al.*, 2015). The spatial and temporal resolution of environmental data also showed high variability, precluding smooth integration and comparison of bioclimatic analyses (Lima-Ribeiro *et al.*, 2015). Alongside the development of geographic information systems (GIS), the advent of cutting-edge spatial interpolation resulted in data layers representing global environmental conditions conform in extent and resolution. Pioneer initiatives like the Climatic Research Unit Terrestrial Climatology (New *et al.*, 1999) and WorldClim (Hijmans *et al.*, 2005a) significantly pushed the application of bioclimatic modeling in ecology, biogeography, conservation biology and evolution. Yet, these gridded datasets were tailored for terrestrial climates only, and the availability of marine data layers lagged significantly behind (Robinson *et al.*, 2011). National Oceanic and Atmospheric Administration's World Ocean Atlas (Levitus, 2001), AquaMaps (Kaschner *et al.*, 2008) and Hexacoral (Fautin & Buddemeier, 2008) were only recently enhanced by the more comprehensive and higher-resolution datasets Bio-ORACLE (Tyberghein *et al.*, 2012) and Marspec (Sbrocco & Barber, 2013).

The increased accessibility of marine data layers allowed an emerging body of research to describe the global distribution of species (e.g., Parravicini *et al.* 2013; Hill & Terblanche 2014; Chefaoui *et al.* 2015; Stuart-Smith *et al.* 2015; Chaudhary *et al.* 2017), address niche-based questions (e.g., Verbruggen *et al.* 2009; Assis *et al.* 2015; Lee-Yaw *et al.* 2016), support biodiversity conservation (e.g., Guisan & Thuiller 2005; Guisan *et al.* 2013; Boavida *et al.* 2016) and ecosystem-based management (Hobday *et al.*, 2010), as well as infer the likely anthropogenic pressures leading to population turnover and extinction (e.g., Scherner *et al.* 2013). The establishment of standard protocols (e.g., Coupled Model Intercomparison Project; CMIP) delivering the outputs of atmosphere-ocean general circulation models (AOGCM) for past and future climate scenarios (Otto-Bliesner *et al.*, 2009; Moss *et al.*, 2010) further expanded the applications for marine data layers, for instance, to predict range shifts through time (e.g., Thomas *et al.* 2004; Burrows *et al.* 2014; Neiva *et al.* 2015; Assis *et al.* 2017) or test relevant evolutionary hypotheses such as the location of marine biodiversity hotspots free from past bottlenecks and extinctions (*i.e.*, climatic refugia; Waltari *et al.* 2007; Assis *et al.* 2014; Chefaoui et al. 2017).

The marine datasets currently available, however, are nearly exclusively restricted to the top surface layer of the oceans (e.g., Bio-ORACLE), and those including benthic layers adjacent to the seabed are particularly coarse in resolution (Hexacoral to ~56 km^2 and World Ocean Atlas to ~112 km^2, at the equator) or limited to biophysical features extracted from bathymetric profiles (MARSPEC, Sbrocco & Barber, 2013). These constraints significantly limit the potential for modelling benthic species (Davies & Guinotte, 2011; Reiss *et al.*, 2014; Boavida *et al.*, 2016), which include a large proportion of marine biodiversity. For instance, the exploration of deep cryptic refugia for marine species is suboptimal when using surface data only (Perry *et al.*, 2005; Graham *et al.*, 2007; Assis *et al.*, 2016).

To address this gap, we present a significant extension of the marine data layers available in Bio-ORACLE. New ecologically relevant surface and benthic layers tailored for mechanistic and correlative modelling (Kearney & Porter, 2009; Peterson *et al.*, 2011) are provided for present conditions and the new generation of climate change scenarios (Moss *et al.* 2010). Besides the extension of Bio-ORACLE to include benthic layers for temperature, salinity, nutrients and chlorophyll (Table 1), we also provide new data on sea ice, current velocity, phytoplankton, primary productivity, iron and light at bottom for a better understanding of marine macroecological processes. We further determine the reliability of data layers (as stressed by Hall & Hall 2014) using a cross-validation framework against *in situ* quality controlled data. We provide a package of functions in the R software environment (R Development Core Team, 2016) for easy integration with the available pipelines for bioclimatic modeling (e.g., Thuiller *et al.* 2009; Naimi & Araújo 2016).

Marine data layers

Marine data layers for present conditions were produced with climate data describing monthly averages for the period 2000-2014, obtained from pre-processed global ocean reanalyses combining satellite and *in situ* observations at regular two- and three-dimensional spatial grids (Table 1). Future layers were produced for 2040-2050 and 2090-2100 by averaging data from distinct AOGCM provided by the CMIP 5 (Table 2). The available data (temperature, salinity, current velocity and sea ice thickness) were obtained for the new representative concentration pathway scenarios (RCP): the RCP26, a peak-and-decline scenario ending in very low greenhouse gas concentration levels by the end of the 21th century, the RCP45 and RCP60 in which levels stabilize, and the RCP85, a scenario of increasing emissions over time leading to high greenhouse gas concentration levels (reviewed by Moss *et al.*, 2010).

The monthly averages for the present and future were used to produce 6 distinct predictors per variable for bioclimatic modelling: the long-term average, the minimum and maximum records, the long-term average of the minimum and maximum records per year (e.g., temperature of the warmest month, on average), and range, given by the average of the absolute difference between the minimum and maximum records per year. These predictors were statistically downscaled (i.e., from coarse to fine scale resolution) to a common spatial resolution of 5 arcmin (approximately 0.08 degrees or 9.2 km at the equator) by fitting a Kriging model based on the 12 nearest values of each focal cell (e.g., Hofstra *et al.* 2008; Lima-Ribeiro *et al.* 2015). The choice of Kriging over other interpolation methods was based on studies showing higher performance for this method (e.g., Hofstra et al., 2008; Lima-Ribeiro et al., 2015) and also on *a priori* test performed against Inverse Distance Weighting (IDW; e.g., Kemp *et al.*, 2012; Assis *et al.*, 2014). This test cross-validated the interpolation of 1×10^4 random records with both methods for different variables, and showed lower root mean square error (RMSE) and mean absolute error (MAE) for Kriging, despite the lack of differences in the mean value of all variables (nonparametric Kruskal-Wallis p-values > 0.05; see Table S1.1 in Supporting Information).

The downscaling process for benthic layers considered the geographic position and depth of cells (e.g., Assis *et al.*, 2016; Boavida *et al.*, 2016) as inferred from the general bathymetric chart of the oceans (GEBCO, 2015). Because focal cells included a range of depth values, the benthic layers were produced for the minimum, average and maximum depths. The future layers were downscaled using the change-factor approach (Wilby *et al.*, 2004; Lima-Ribeiro *et al.*, 2015; Varela *et al.*, 2015). This technique is based on applying the predicted magnitude of climate change to the data layers produced for the present. For this purpose, data for the period 2000-2014 was also obtained from the AOGCMs to determine the difference (change-factor) between the present conditions and the future scenarios of change, at the native resolution of each AOGCM (Table 2). Then, the change-factor was downscaled to 0.08º resolution with Kriging (as previously described) and applied to the corresponding baseline layer for the present conditions.

The layers for current velocity and light at the bottom were further post-processed. While current velocity was determined with the Pythagoras theorem on the meridional (along the longitude circle) and zonal (along the latitude circle) components of ocean currents, light at the bottom used a standard depth-dependent exponential function (Graham *et al.*, 2007; Assis *et al.*, 2016) based on photosynthetically active radiation (PAR) and diffuse attenuation coefficient (Kd490):

Light at bottom = PAR • exp(-Kd490 • z)

where z is depth inferred from the general bathymetric chart of the oceans (GEBCO, 2015).

All layers were exported to ASCII (ESRI) and TIFF grid formats for easy download and integration in modern GIS technologies (e.g., ESRI and QGIS). To minimize the possible spatial gap with "no data" between the data layers provided and the available vector shorelines, the global self-consistent, hierarchical, high-resolution geography database (Wessel & Smith, 1996) is recommended.

Reliability of marine data layers

The reliability of downscaled data layers was inferred with *in situ* quality controlled data provided by the Global Ocean Data Analysis Project (GLODAP; Olsen *et al.*, 2016) for temperature, salinity, phosphate, nitrate, silicate, dissolved oxygen and chlorophyll. This was performed by cross-validating the outputs of downscaling the raw data used to produce data layers at the locations (geographic position and depth) of each sample available in GLODAP, with the actual data provided by this dataset (e.g., Davies & Guinotte, 2011; Boavida *et al.*, 2016; Assis *et al.*, 2017). The paired relationships between the interpolated and *in situ* data were statistically analysed with mean absolute error, root mean square error and Pearson's correlation. These tests showed the layers mirroring most climatic patterns present in quality-controlled data. All variables retrieved low error rates (Table 1) and high correlation coefficients (Pearson's cor.: > 0.93; Table 1) with a unique exception for chlorophyll, which showed a much lower correlation (Pearson's cor.: 0.56; Table 1; see Fig. S1.1 in Supporting Information). This exception likely results from coupling the high temporal variability known for chlorophyll on a daily basis (e.g., Iriarte *et al.*, 2007; Wang *et al.*, 2009; Wang Hladik *et al.*, 2010), with the monthly averages used to produce the layers (as discussed by Davies & Guinotte, 2011 and Boavida *et al.*, 2016). Moreover, chlorophyll was the variable with the least number of *in situ* observations (5401 for the global ocean; Table 1), a fact that may have precluded a proper assessment of reliability in cross-validation.

The spatial distribution error of the layers was illustrated by mapping the difference between the interpolated and *in situ* data onto a 2.5º grid (e.g., Davies & Guinotte, 2011) and by plotting this difference against depth. In general, the distribution of errors also showed high accuracy for all layers. Temperature, phosphate, nitrate and dissolved molecular oxygen only displayed specific anomalies, highly restricted to discrete regions of the global ocean (e.g., east Siberian Sea and southern North Sea; Figs. S1.4, S1.8, S1.10 and S1.14), and with no relationship with depth (Figs. S1.3, S1.7, S1.9 and S1.13).

The errors for silicate were mostly in the Southern Ocean (Fig. S1.12) and those for salinity were in the top layers (surface waters with higher positive anomaly, Fig. S1.5) of the Canadian Arctic and east Siberian Sea (Fig. S1.6), particularly for values bellow 30 PSS (Fig. S1.5). The spatial distribution of errors further showed that dissolved molecular oxygen, phosphate, salinity, silicate and temperature have good coverage of *in situ* samples (GLODAP dataset), while nitrate and chlorophyll are mostly uncovered throughout the globe.

R-package tool

In addition to providing the layers for download in ASCII and TIFF formats, we also developed the *sdmpredictors* package of functions in R (R Development Core Team, 2016) to facilitate listing, extraction and management of data layers. This package, whose functions are detailed in Table 3, also integrates the layers from the first version of Bio-ORACLE, as well as those of MARSPEC (Sbrocco & Barber, 2013) and BioClim (Hijmans *et al.*, 2005b). The source code and related help files are available via CRAN repository and can be easily installed by entering the following lines of code into the R command prompt:

```
1. install.packages("sdmpredictors")
2. library(sdmpredictors)
```

Conclusion

A comprehensive set of new ecologically relevant data layers are provided for bioclimatic modelling in a user-friendly format, with global coverage and comparable grid system. The current update expands the potential of Bio-ORACLE by (1) covering the benthic realm with climate and environmental data, (2) adding new variables to present-day conditions (i.e., current velocity, iron, light, phytoplankton, primary productivity and sea ice), (3) adding the new generation of climate change scenarios, (4) providing means of data reliability and uncertainty, as well as (5) a package of functions in the R software environment.

The relevance of new data aiming for species associated with sea benthic features (e.g., Assis et al., 2016; Boavida *et al.* 2016) is clearly underlined by the disparity in ocean temperatures between surface and benthic layers, which can amount up to 28.8ºC in the deeper regions of lower latitudes (Fig. 1). The data provided for the new generation of climate change scenarios further diversifies the range of scientific questions that can be addressed using Bio-ORACLE. One such case is the possibility to explore climate-induced depth range shifts (e.g., Assis *et al.* 2016; Assis, Araújo & Serrão 2017), the marine equivalent to elevation range shifts for terrestrial species (Galbreath *et al.*, 2009; Chen *et al.*, 2011). The reliability of new layers has been assessed with *in situ* quality controlled data. This information is quite relevant for marine modelers and showed a good agreement between data layers and the global climatic patterns. The future data using the magnitude of climate changes (change-factor) of different AOGCM, provides the highest confidence level currently attainable (Hall & Hall, 2014). Despite their inherent uncertainties, the AOGCM used represent the present scientific understanding linking greenhouse gas emissions with global climate changes.

The familiar data structure of Bio-ORACLE (rasters for GIS) and its integration with R computing language should allow easy acquisition, exploration and manipulation of data, as well as smooth integration with the available statistical tools (e.g., Thuiller *et al.* 2009; Naimi & Araújo 2016). The new data layers represent a valuable addition to the spatial information about climate available for the global ocean. Their key features can be used to improve bioclimatic modelling and provide valuable insights into the current and future states of marine biodiversity and indirectly into the services it provides to society. When requested by decision makers, these outcomes may guide important climate change integrated conservation strategies (Hannah *et al.*, 2002; Hobday *et al.*, 2010), and feed baseline assessments such as those required for the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES).

References

Assis, J., Araújo, M.B. & Serrão, E.A. (2017a) Projected climate changes threaten ancient refugia of kelp forests in the North Atlantic. *Global Change Biology*.

Assis, J., Berecibar, E., Claro, B., Alberto, F., Reed, D., Raimondi, P. & Serrão, E.A. (2017b) 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., Coelho, N.C., Lamy, T., Valero, M., Alberto, F. & Serrão, E.A. (2016) Deep reefs are climatic refugia for genetic diversity of marine forests. *Journal of Biogeography*, **43**, 833–844.

Assis, J., Serrão, E.A., Claro, B., Perrin, C. & Pearson, G.A. (2014) Climate-driven range shifts explain the distribution of extant gene pools and predict future loss of unique lineages in a marine brown alga. *Molecular Ecology*, **23**, 2797–2810.

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.

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.

Chaudhary, C., Saeedi, H. & Costello, M.J. (2017) Bimodality of Latitudinal Gradients in Marine Species Richness. *Trends in Ecology & Evolution*, **31**, 670–676.

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*, 123–137.

Chen, I.-C., Hill, J.K., Ohlemüller, R., Roy, D.B. & Thomas, C.D. (2011) Rapid range shifts of species associated with high levels of climate warming. *Science (New York, N.Y.)*,

333, 1024–1026.

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

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

Galbreath, K.E., Hafner, D.J. & Zamudio, K.R. (2009) When cold is better: Climate-driven elevation shifts yield complex patterns of diversification and demography in an alpine specialist (american pika, ochotona princeps). *Evolution*, **63**, 2848–2863.

GEBCO (2015) *General bathymetric chart of the oceans*, (ed. by B.O.D. Centre) British Oceanographic Data Centre, Liverpool, UK.

Graham, M.H., Kinlan, B.P., Druehl, L.D., Garske, L.E. & Banks, S. (2007) Deep-water kelp refugia as potential hotspots of tropical marine diversity and productivity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 16576–16580.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.

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

Hannah, L., Midgley, G.F. & Millar, D. (2002) Climate change-integrated conservation strategies. *Global Ecology and Biogeography*, **11**, 485–495.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005a) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005b) WORLDCLIM - a set of global climate layers (climate grids). *International Journal of Climatology*, **25**, 1965–1978.

Hill, M.P. & Terblanche, J.S. (2014) Niche Overlap of Congeneric Invaders Supports a Single-Species Hypothesis and Provides Insight into Future Invasion Risk: Implications for Global Management of the Bactrocera dorsalis Complex. *Plos One*,

9, e90121.

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.

Hofstra, N., Haylock, M., New, M., Jones, P. & Frei, C. (2008) Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research Atmospheres*, **113**, 1–19.

Iriarte, J.L., González, H.E., Liu, K.K., Rivas, C. & Valenzuela, C. (2007) Spatial and temporal variability of chlorophyll and primary productivity in surface waters of southern Chile (41.5–43° S). *Estuarine, Coastal and Shelf Science*, **74**, 471–480.

Kaschner, K., Kesner-Reyes, K., Garilao, C., Rius-Barile, J., Rees, T. & Froese, R. (2008) AquaMaps: Predicted range maps for aquatic species.

Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: Combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.

Kemp, M.U., Loon, E.E. Van, Shamoun-baranes, J. & Bouten, W. (2012) RNCEP : global weather and climate data at your fingertips. **3**, 65–70.

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. (2001) *World Ocean Atlas*, (ed. by NOAA) Silver Spring, Md.

Lima-Ribeiro, M., Varela, S., González-Hernández, J., Oliveira, G., Diniz-Filho, J. & Terribile, L. (2015) ecoClimate: a database of climate data from multiple models for past, present, and future for Macroecologists and Biogeographers. *Biodiversity Informatics*, 1–21.

Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., van Vuuren, D.P., Carter, T.R., Emori, S., Kainuma, M., Kram, T., Meehl, G.A., Mitchell, J.F.B., Nakicenovic, N., Riahi, K., Smith, S.J., Stouffer, R.J., Thomson, A.M., Weyant, J.P. & Wilbanks, T.J. (2010) The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756.

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., Coelho, N.C., Fernandes, F., Pearson, G.A. & Serrão, E.A. (2015) Genes

Left Behind: Climate Change Threatens Cryptic Genetic Diversity in the Canopy-Forming Seaweed Bifurcaria bifurcata. *Plos One*, **10**, e0131530.

New, M., Hulme, M. & Jones, P. (1999) Representing Twentieth-Century Space – Time Climate Variability . Part I: Development of a 1961 – 90 Mean Monthly Terrestrial Climatology. *Journal of Climate*, **12**, 829–856.

Olsen, A., Key, R.M., 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. 297–323.

Otto-Bliesner, B.L., Schneider, R., Brady, E.C., Kucera, M., Abe-Ouchi, A., Bard, E., Braconnot, P., Crucifix, M., Hewitt, C.D., Kageyama, M., Marti, O., Paul, A., Rosell-Melé, A., Waelbroeck, C., Weber, S.L., Weinelt, M. & Yu, Y. (2009) A comparison of PMIP2 model simulations and the MARGO proxy reconstruction for tropical sea surface temperatures at last glacial maximum. *Climate Dynamics*, **32**, 799–815.

Parravicini, V., Kulbicki, M., Bellwood, D.R., Friedlander, A.M., Chabanet, P., Floeter, S.R., Myers, R., Vigliola, L., Agata, S.D. & Mouillot, D. (2013) Global patterns and predictors of tropical reef fi sh species richness. 1254–1262.

Perry, A.L., Low, P.J., Ellis, J.R. & Reynolds, J.D. (2005) Climate change and distribution shifts in marine fishes. *Science*, **308**, 1912–1915.

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.

R Development Core Team (2016) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Reiss, H., Birchenough, S., Borja, A., Buhl-Mortensen, L., Craeymeersch, J., Dannheim, J., Darr, A., Galparsoro, I., Gogina, M., Neumann, H., Populus, J., Rengstorf, A.M., Valle, M., Hoey, G. Van, Zettler, M.L., Degraer, S., van Hoey, G., Zettler, M.L. & Degraer, S. (2014) Benthos distribution modelling and its relevance for marine ecosystem management. *ICES Journal of Marine Science*, **72**, 297.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P. & Richardson, A.J. (2011) Pushing the limits in marine species distribution modelling: Lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, **20**, 789–802.

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

Scherner, F., Antunes, P., Cabral, E., Oliveira, D., Carlos, J., Hall-spencer, J.M., Chow, F., Marcos, J., Nunes, C., Maria, S. & Pereira, B. (2013) Coastal urbanization leads to remarkable seaweed species loss and community shifts along the SW Atlantic. *Marine Pollution Bulletin*, **76**, 106–115.

Stuart-Smith, R.D., Edgar, G.J., Barrett, N.S., Kininmonth, S.J. & Bates, A.E. (2015) Thermal biases and vulnerability to warming in the world's marine fauna. *Nature*, **528**, 88–92.

Sutherst, R.W. & Maywald, G.F. (1985) A computerised system for matching climates in ecology. *Agriculture, Ecosystems & Environment*, **13**, 281–299.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., De Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M. a, Peterson, a T., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.

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.

Varela, S., Lima-Ribeiro, M.S. & Terribile, L.C. (2015) A short guide to the climatic variables of the last glacial maximum for biogeographers. *Plos One*, **10**, e0129037.

Verbruggen, H., Tyberghein, L., Pauly, K., Vlaeminck, C., Nieuwenhuyze, K. Van, Kooistra, W.H.C.F., Leliaert, F. & de Clerck, O. (2009) Macroecology meets macroevolution: Evolutionary niche dynamics in the seaweed Halimeda. *Global Ecology and Biogeography*, **18**, 393–405.

Waltari, E., Hijmans, R.J., Peterson, A.T., Nyári, Á.S., Perkins, S.L. & Guralnick, R.P. (2007) Locating pleistocene refugia: Comparing phylogeographic and ecological niche model predictions. *Plos One*, **2**, e563.

Wang, X., Le Borgne, R., Murtugudde, R., Busalacchi, A.J. & Behrenfeld, M. (2009) Spatial and temporal variability of the phytoplankton carbon to chlorophyll ratio in the equatorial Pacific: A basin-scale modeling study. *Journal of Geophysical Research: Oceans*, **114**.

Wang Hladik, C.., Huang, W., Milla, K., Edmiston, L., Harwell, M.. & Schalles, J.. (2010)

Detecting the Spatial and Temporal Variability of Chlorophyll-a Concentration and Total Suspended Solids in Apalachicola Bay, Florida using MODIS Imagery. *International Journal of Remote Sensing*, **31**, 439–453.

Wessel, P. & Smith, W.H.F. (1996) A global, self-consistent, hierarchical, high-resolution shoreline database. *Journal of Geophysical Research: Solid Earth*, **101**, 8741–8743.

Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P. & Mearns, L.O. (2004) *Guidelines for use of climate scenarios developed from statistical downscaling methods. IPCC Task Group on data and scenario support for impact and climate analysis (TGICA)*, IPCC Data Distribution Centre.

Data accessibility

The Bio-ORACLE layers are accessible online at http://www.bio-oracle.org for download (as ASCII and TIFF files) and preview.

Biosketch

Jorge Assis is a post-doctoral researcher at CCMAR-CIMAR, University of Algarve. His research is focused on ecological niche modelling, past and future climate-driven range shifts and landscape genetics at multiple temporal and spatial scales.

Table 1. Marine data layers, unit, correspondence with the first version of Bio-ORACLE (BO1), range of values (determined for benthic layers, at their average depth), accuracy assessed with quality control data (MAE: mean average error; RMSE: root mean square error; Cor: Pearson's correlation coefficient), number of quality control records (n) and source of climate data.

Layer	**Unit**	**BO1**	**Range**	**MAE**	**RMSE**	**Cor**	**n**	**Source**
Temperature	ºC	Yes	[-1.94;39.22]	0.39	0.75	0.99	445248	ARMOR
Salinity	PSS	Yes	[4.75;41.96]	0.13	0.52	0.93	444925	ARMOR
Sea ice concentration	Fraction	No	[0;1]	-	-	-	-	ORAP
Sea ice thickness	m	No	[0;10.94]	-	-	-	-	ORAP
Current velocity	$m \bullet s^{-1}$	No	[0;2.42]	-	-	-	-	ORAP
Nitrate	$\mu mol \bullet m^{-3}$	Yes	[0;164.51]	1.62	2.47	0.98	93201	PISCES
Phosphate	$\mu mol \bullet m^{-3}$	Yes	[0;3.55]	0.15	0.23	0.97	349074	PISCES
Silicate	$\mu mol \bullet m^{-3}$	Yes	[0.46;316.67]	5.98	9.05	0.99	367629	PISCES
Dissolved molecular oxygen	$\mu mol \bullet m^{-3}$	Yes	[0;789.94]	15.27	23.50	0.96	417790	PISCES
Dissolved iron	$\mu mol \bullet m^{-3}$	No	[0;0.03]	-	-	-	-	PISCES
Chlorophyll	$mg \bullet m^{-3}$	Yes	[0;17.46]	0.1	0.15	0.56	5401	PISCES
Phytoplankton	$\mu mol \bullet m^{-3}$	No	[0;44.94]	-	-	-	-	PISCES
Primary productivity	$g \bullet m^{-3} \bullet day^{-}$	No	[0;0.95]	-	-	-	-	PISCES
Light at the bottom	$E \bullet m^{-2} \bullet yr^{-1}$	No	[0;69.21]	-	-	-	-	GlobColour

ARMOR: Global Observed Ocean Physics Reprocessing (resolution: 0.25º / 33 vertical levels)

ORAP: Global Ocean Physics Reanalysis ECMWF (resolution: 0.25º / 75 vertical levels)

PISCES: Global Ocean Biogeochemistry Non-assimilative Hindcast (resolution: 0.25º / 75 vertical levels)

GlobColour merging MERIS / MODIS / SeaWiFS (resolution: 0.05º / surface)

Table 2. Average and range of climatic anomalies for the future (2090-2100) under different scenarios of change, and source of data (coupled atmosphere-ocean general circulation models). Climatic anomalies inferred by determining the difference between the benthic layers (average depth) produced for the different representative concentration pathway (RCP) scenarios and those for the present.

Layer	**RCP26**	**RCP45**	**RCP60**	**RCP85**	**Source (AOGCM)**
Temperature	0.21 [-2.06;4.45]	0.29 [-2.01;6.72]	0.33 [-1.24;6.89]	0.51 [-2.62;10.98]	CCSM4,HadGEM2-ES,MIROC5
Salinity	-0.03 [-6.03;1.37]	-0.04 [-4.97;1.39]	-0.05 [-4.48;1.34]	-0.07 [-4.33;1.83]	CCSM4,HadGEM2-ES,MIROC5
Current velocity	0.01 [-0.17;0.18]	0.01 [-0.11;0.28]	0.01 [-0.13;0.16]	0.01 [-0.18;0.24]	CCSM4,HadGEM2-ES,MIROC5
Sea ice thickness	-0.12 [-8.33;1.43]	-0.18 [-8.78;1.98]	-0.22 [-8.95;1.54]	-0.27 [-9.01;1.88]	CCSM4,HadGEM2-ES,MIROC5

CCSM4: The Community Climate System Model 4 (resolution: 1.13º x 0.47º / 60 vertical levels)

HadGEM2-ES: Hadley Centre Global Environmental Model 2 (resolution: 1.00º x 0.84º / 40 vertical levels)

MIROC5: Model for Interdisciplinary Research on Climate 5 (resolution: 1.41º x 0.81º / 50 vertical levels)

Table 3. List of functions available in sdmpredictors package, description and arguments that users may specify.

Function	Description	Arguments
list_datasets()	Explore datasets available in the package	Terrestrial (logical) only terrestrial data are returned; marine (logical) only marine data are returned.
list_layers()	Explore layers in a dataset	datasets (character); terrestrial (logical); marine (logical); monthly (logical) only annual and monthly layers are returned (default).
list_layers_future()	Explore future layers in the package	datasets (character); scenario (character) for climate change scenario, e.g., RCP85; year (integer) for the climate change prediction, e.g. 2100; terrestrial (logical); marine (logical); monthly (logical).

get_future_layers()	Get the name of a future climate layer(s) based on the current climate layer(s)	current_layer_codes (character) with the code(s) of the layers either as a vector or dataframe provided by list_layers); scenario (character); year (integer).
load_layers()	Download specific layers to the current directory	layercodes (character) with the codes of the layers to be loaded; equalarea (logical) for Behrmann cylindrical equal-area projection; rasterstack (logical) to stack the layers in a unique object; datadir (character) for the directory to store data.
layer_stats()	Layer statistics	layercodes (character).
layers_correlation()	Pearson's correlation coefficient between layers	layercodes (character).

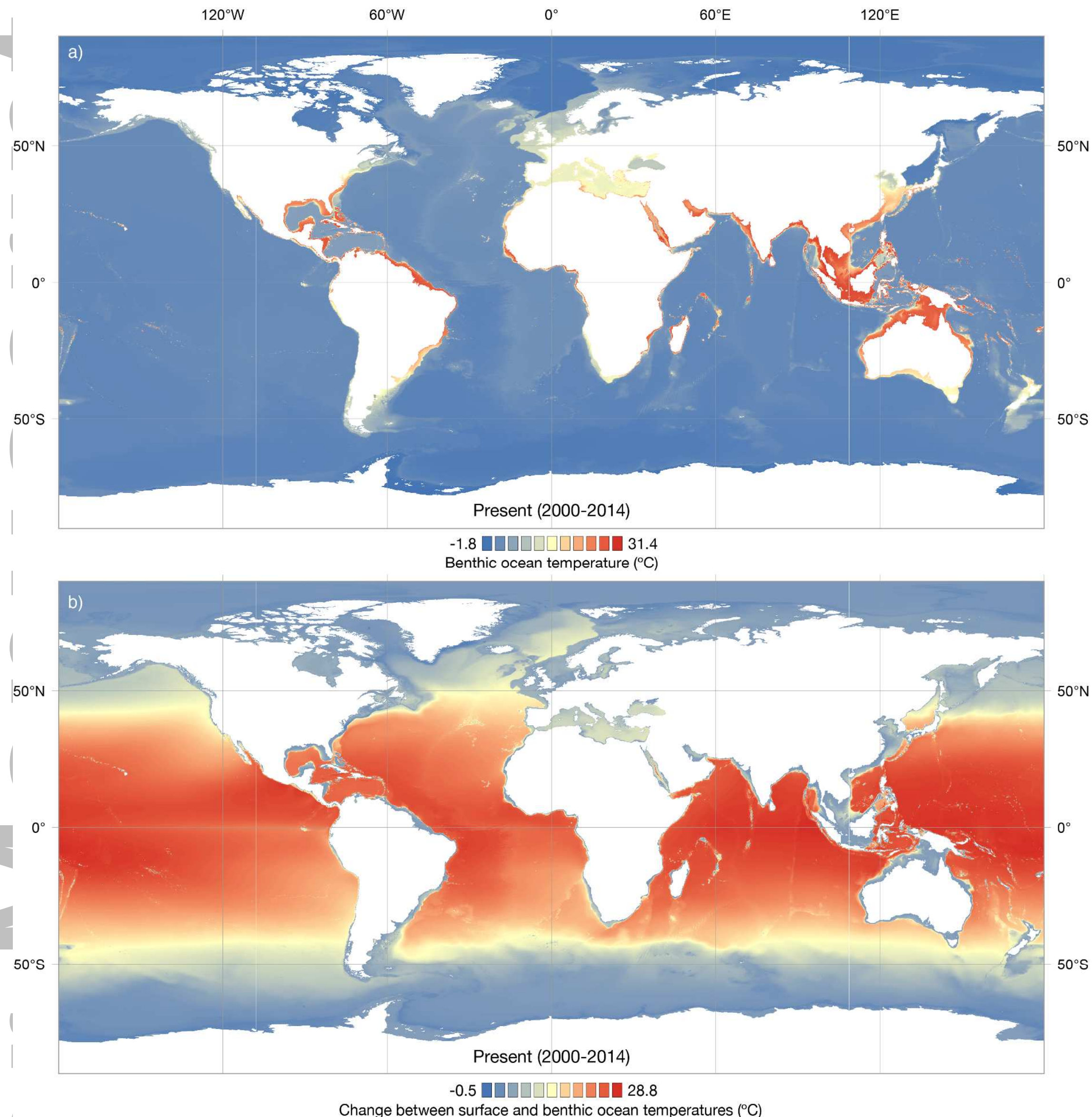


Fig 1. (panel a) Mean benthic ocean temperature for the present (period 2000-2014) and (panel b) change (difference) between surface and benthic ocean temperatures.

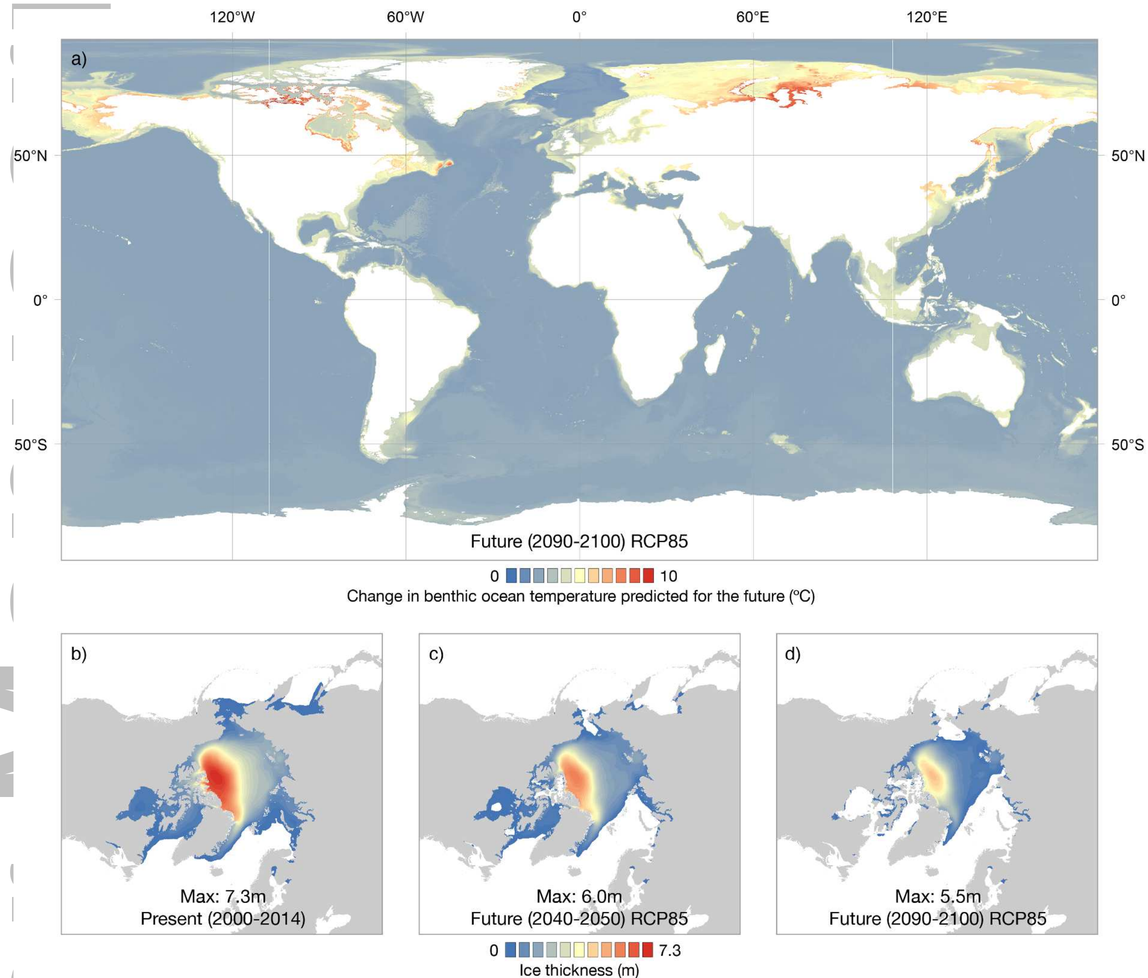
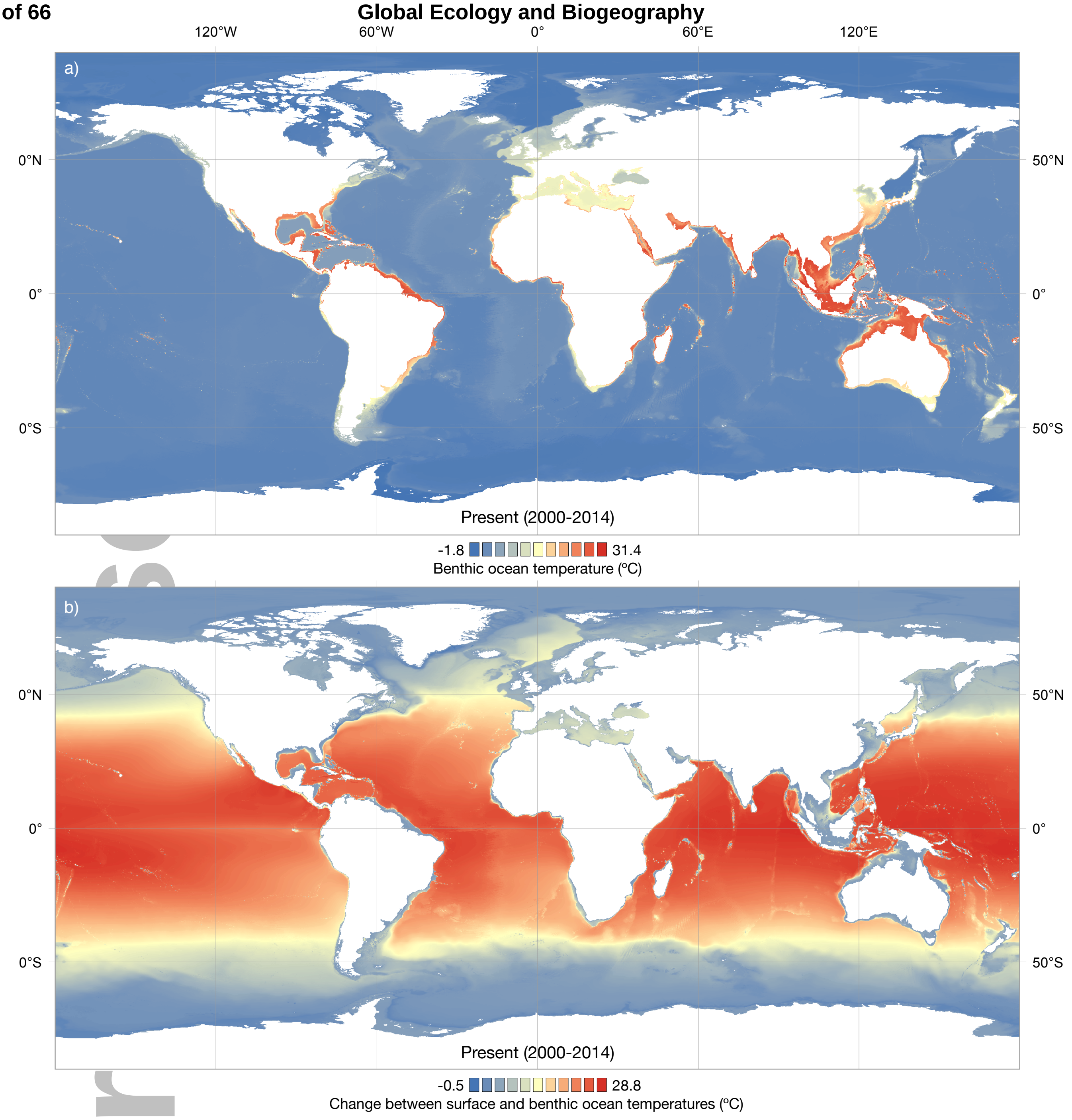
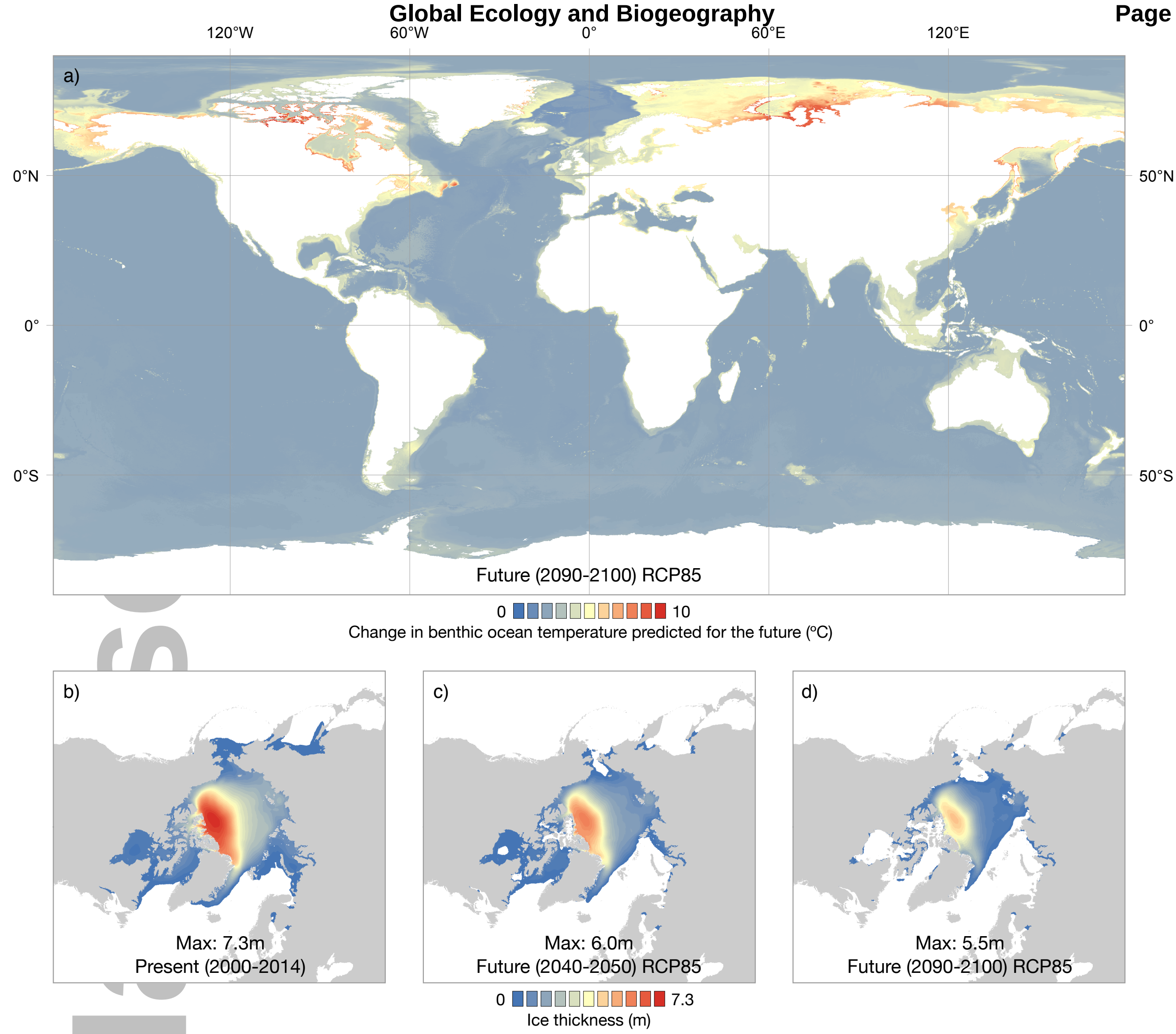


Fig 2. (panel a) Change in the maximum benthic ocean temperature predicted for the period 2090-2100 with RCP85 and arctic mean ice thickness for the periods (panel b) 2000-2014, (panel c) 2040-2050 and (panel d) 2090-2100, predicted with RCP85.

Supporting Information

S1. Reliability of marine data layers determined with *in situ* quality-controlled data.


120°W
60°W
0°
60°E
120°E
a)
0°N
0°
0°S
50°N
0°
50°S
Present (2000-2014)
-1.8
31.4
Benthic ocean temperature (ºC)
b)
0°N
0°
0°S
50°N
0°
50°S
Present (2000-2014)
-0.5
28.8
Change between surface and benthic ocean temperatures (ºC)


120°W
60°W
0°
60°E
120°E
a)
0°N
50°N
0°
0°
0°S
50°S
Future (2090-2100) RCP85
0 10
Change in benthic ocean temperature predicted for the future (ºC)
b)
Max: 7.3m
Present (2000-2014)
c)
Max: 6.0m
Future (2040-2050) RCP85
d)
Max: 5.5m
Future (2090-2100) RCP85
0 7.3
Ice thickness (m)

Supplementary Information 1

Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling

Reliability of marine data layers determined with *in situ* quality-controlled data

Jorge Assis[1]*, Lennert Tyberghein[2], Samuel Bosch[2,3], Heroen Verbruggen[4], Ester A. Serrão[1], Olivier De Clerck[3]

[1] Centre for Marine Sciences, CCMAR-CIMAR, University of Algarve, Campus Gambelas, 8005-139 Faro, Portugal.

[2] Flanders Marine Institute (VLIZ), InnovOcean site, Wandelaarskaai 7, 8400 Ostend, Belgium

[3] Phycology Research Group, Biology Department, Ghent University, 9000 Ghent, Belgium.

[4] School of BioSciences, University of Melbourne, Victoria 3010, Australia.

* Corresponding author (jorgemfa@gmail.com | +351 912 361 127)

Table 1. Statistical downscaling performance of Kriging and Inverse Distance Weighting (IDW) for different variables. Analyses performed with mean absolute error (MAE), root mean square error (RMSE) and nonparametric Kruskal-Wallis testing the difference between mean values (α = 0.05).

Layer	**Unit**	**MAE**		**RMSE**		**Kruskal-Wallis**
		Kriging	IDW	Kriging	IDW	
Temperature	ºC	0.061	0.073	0.132	0.154	0.987
Salinity	PSS	0.027	0.033	0.059	0.071	0.960
Current velocity	$m \cdot s^{-1}$	0.003	0.004	0.007	0.009	0.986
Nitrate	$\mu mol \cdot m^{-3}$	0.003	0.005	0.008	0.010	0.983
Phosphate	$\mu mol \cdot m^{-3}$	0.003	0.004	0.007	0.009	0.989
Silicate	$\mu mol \cdot m^{-3}$	0.004	0.004	0.008	0.010	0.990
Dissolved molecular	$\mu mol \cdot m^{-3}$	0.004	0.005	0.007	0.009	0.986
Dissolved iron	$\mu mol \cdot m^{-3}$	0.004	0.005	0.007	0.009	0.988
Chlorophyll	$mg \cdot m^{-3}$	0.003	0.004	0.007	0.009	0.985
Phytoplankton	$\mu mol \cdot m^{-3}$	0.003	0.004	0.007	0.011	0.986
Primary productivity	$g \cdot m^{-3} \cdot day^{-1}$	0.003	0.004	0.007	0.010	0.991

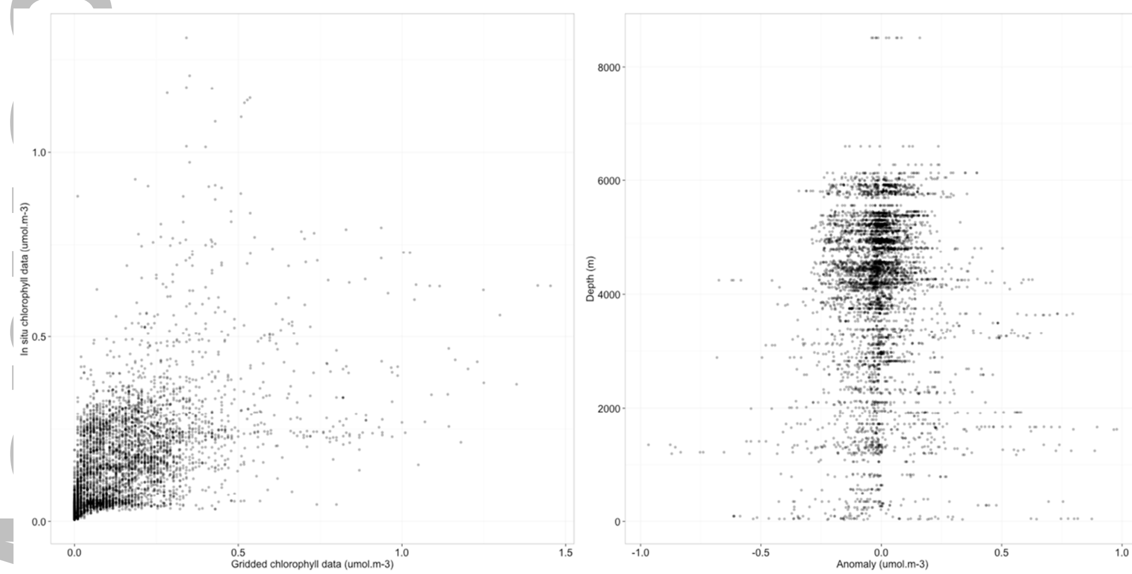


Fig 1. Accuracy of downscaled layer of chlorophyll data. (left panel) Correlation between the interpolated and *in situ* data for chlorophyll. (right panel) Difference (anomaly) between the interpolated and *in situ* data against depth.

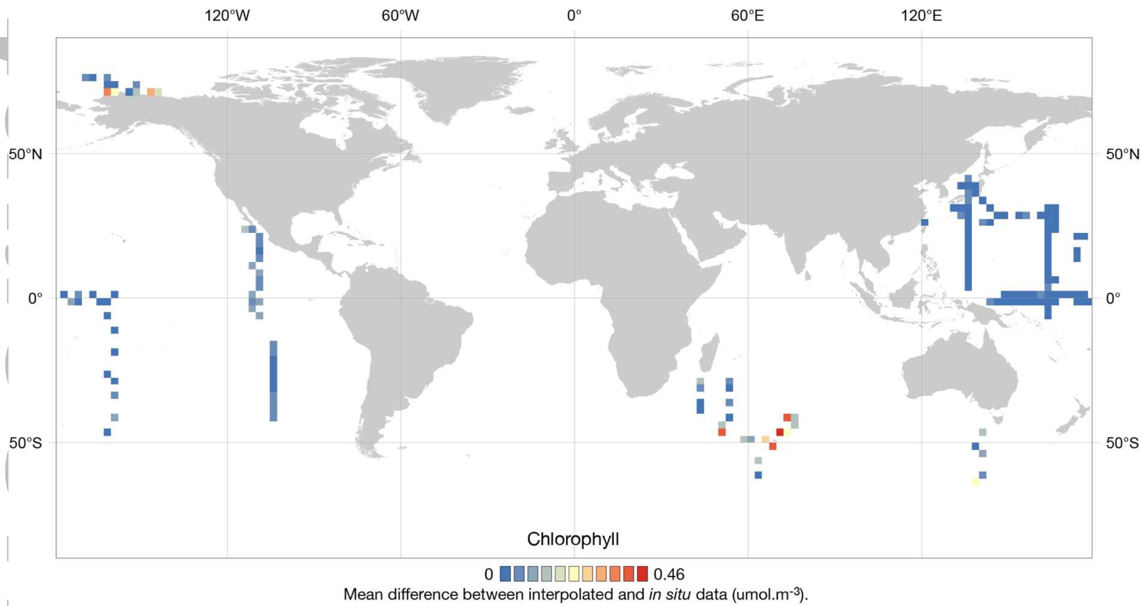


Fig 2. Spatial distribution of the error of chlorophyll data shown as the average difference between the interpolated and *in situ* data.

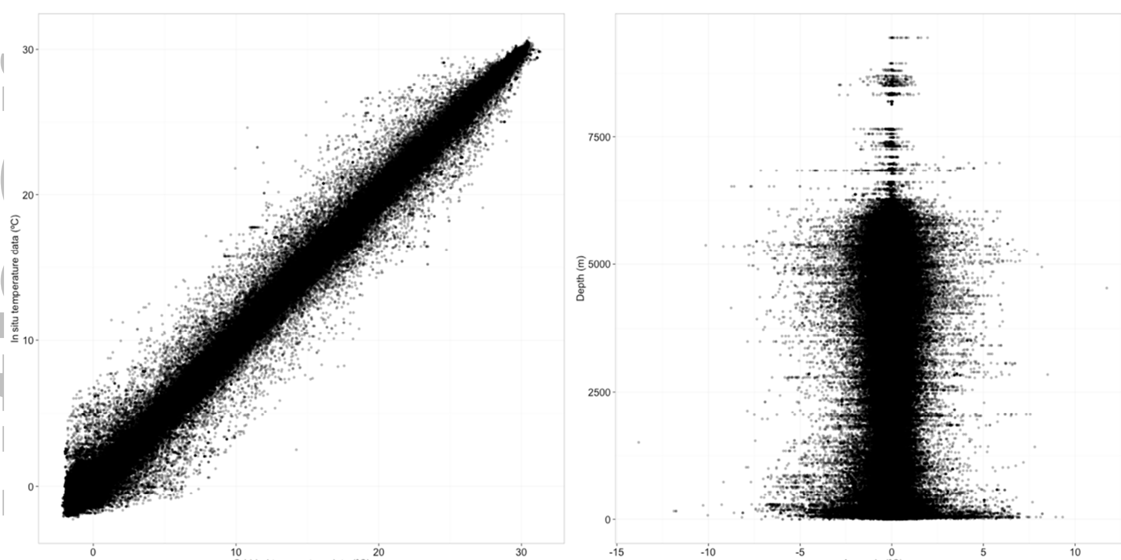


Fig 3. Accuracy of downscaled ocean temperature data. (left panel) Correlation between the interpolated and *in situ* data for temperature. (right panel) Difference (anomaly) between the interpolated and *in situ* data against depth.

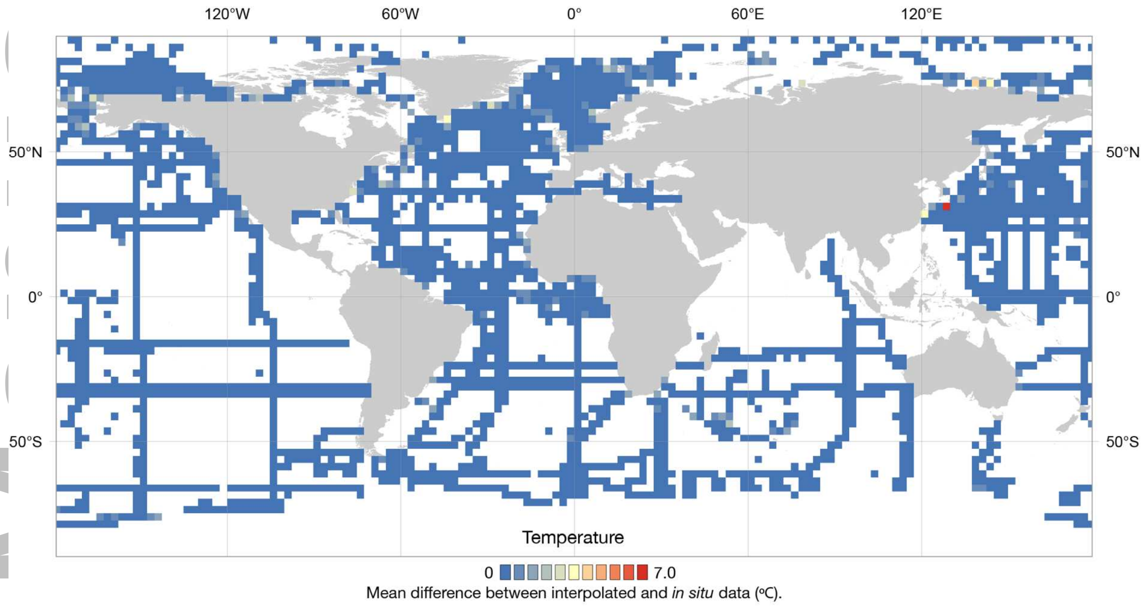


Fig 4. Spatial distribution of the error of ocean temperature data shown as the average difference between the interpolated and *in situ* data.

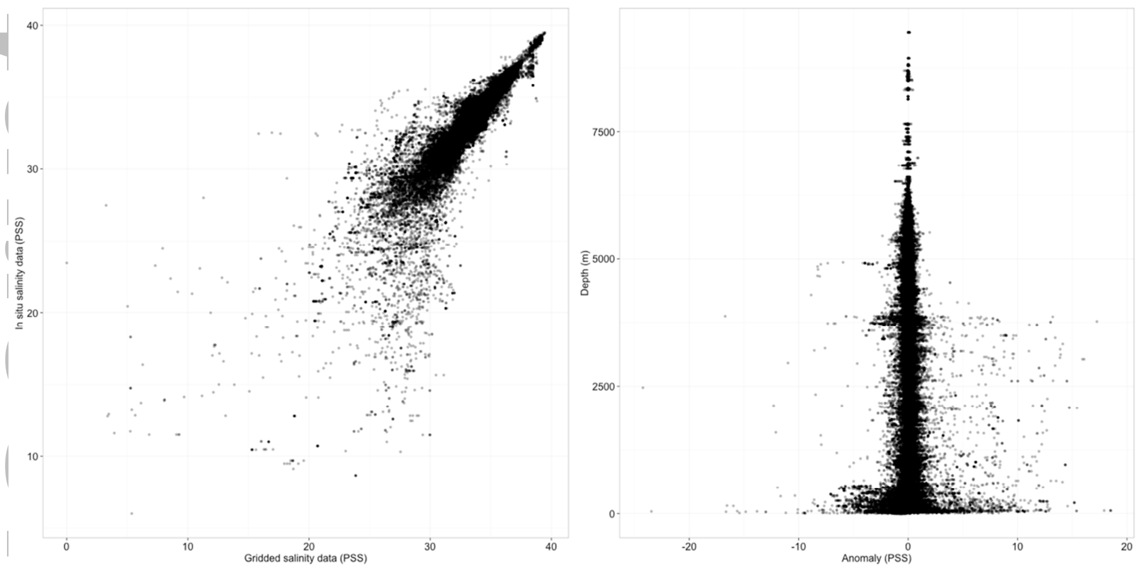


Fig 5. Accuracy of downscaled ocean salinity data. (left panel) Correlation between the interpolated and *in situ* data for salinity. (right panel) Difference (anomaly) between the interpolated and *in situ* data against depth.

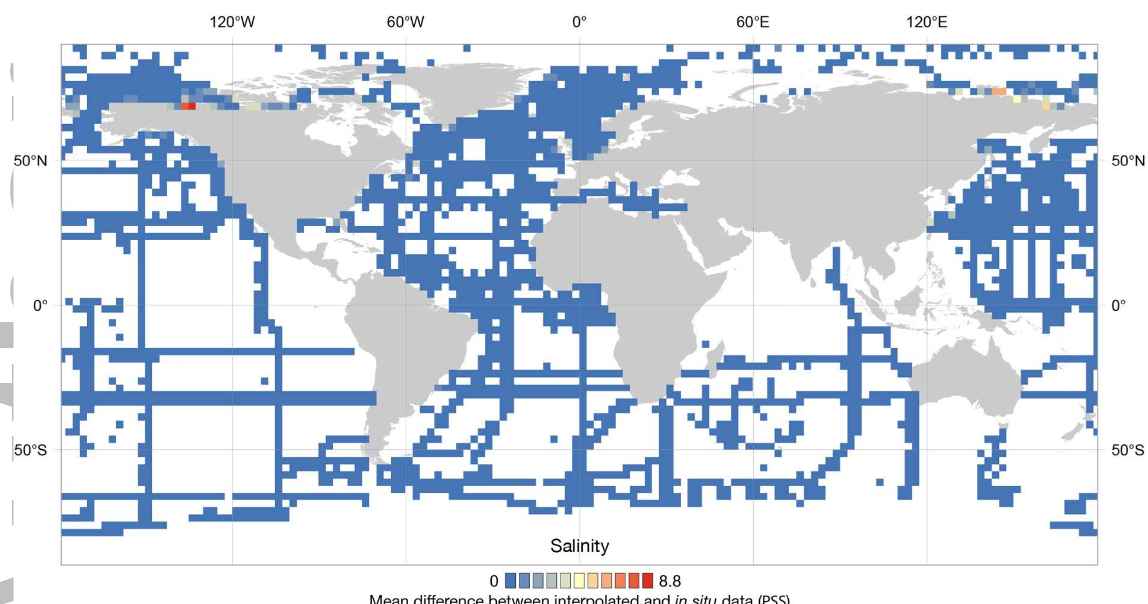


Fig 6. Spatial distribution of the error of ocean salinity data shown as the average difference between the interpolated and *in situ* data.

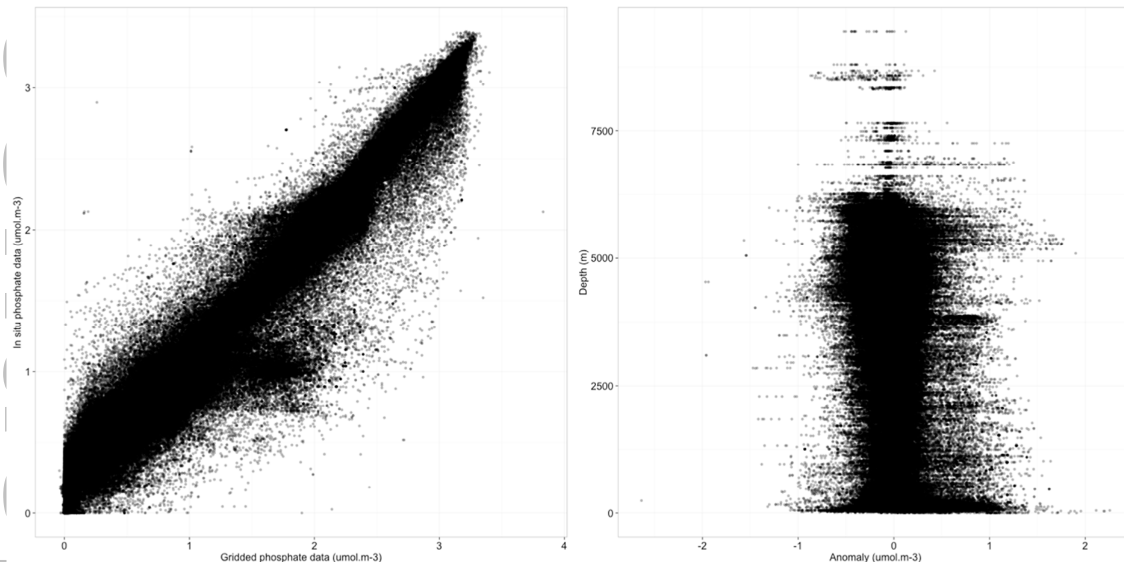


Fig 7. Accuracy of downscaled phosphate data. (left panel) Correlation between the interpolated and *in situ* data for phosphate. (right panel) Difference (anomaly) between the interpolated and *in situ* data against depth.

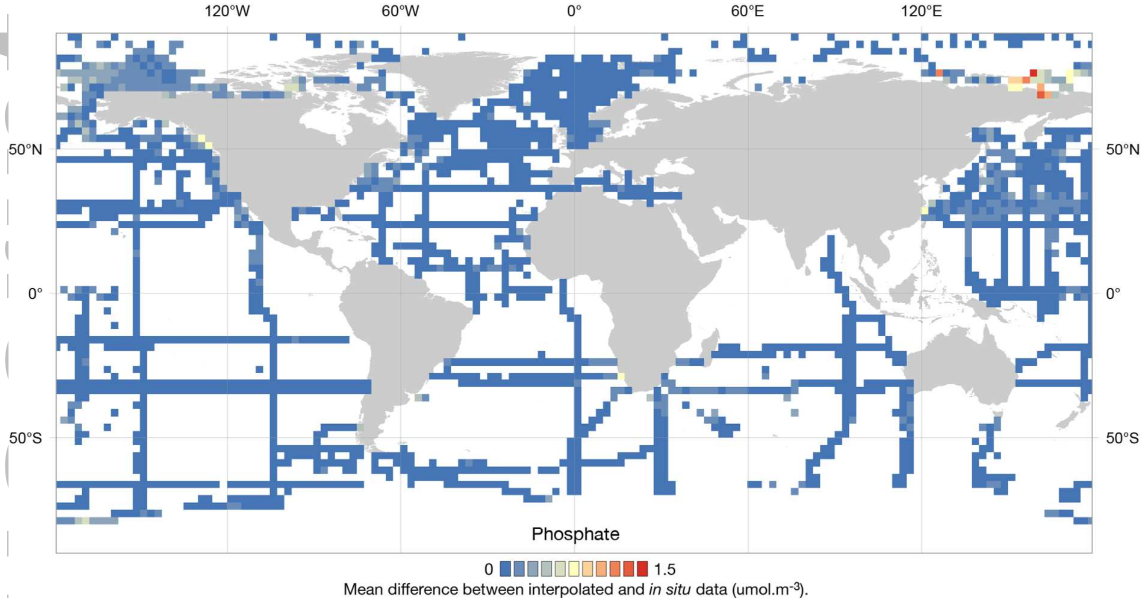


Fig 8. Spatial distribution of the error of phosphate data shown as the average difference between the interpolated and *in situ* data.

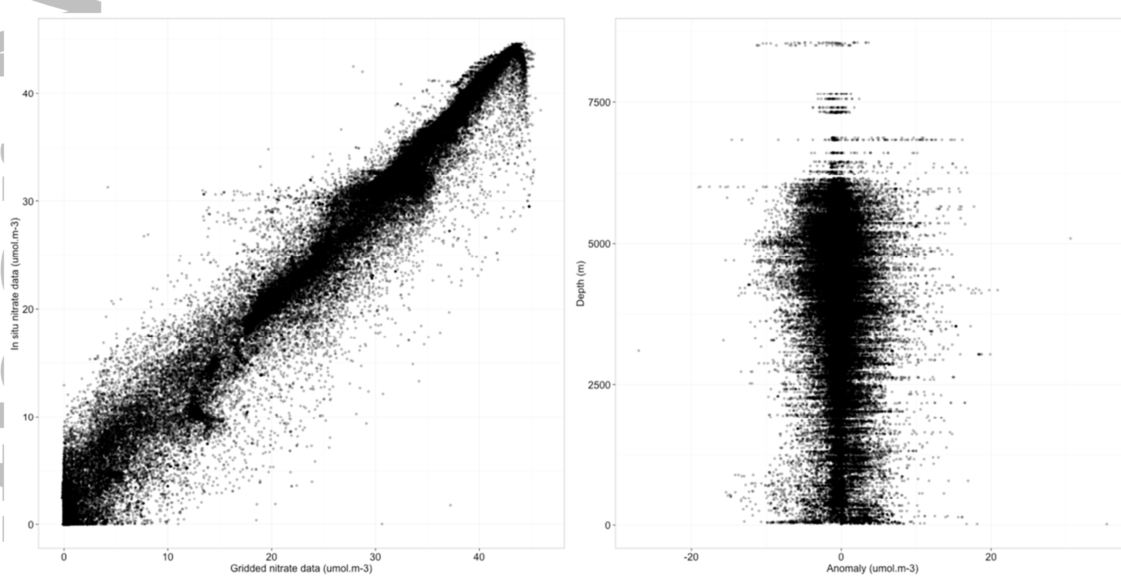


Fig 9. Accuracy of downscaled nitrate data. (left panel) Correlation between the interpolated and *in situ* data for nitrate. (right panel) Difference (anomaly) between the interpolated and *in situ* data against depth.

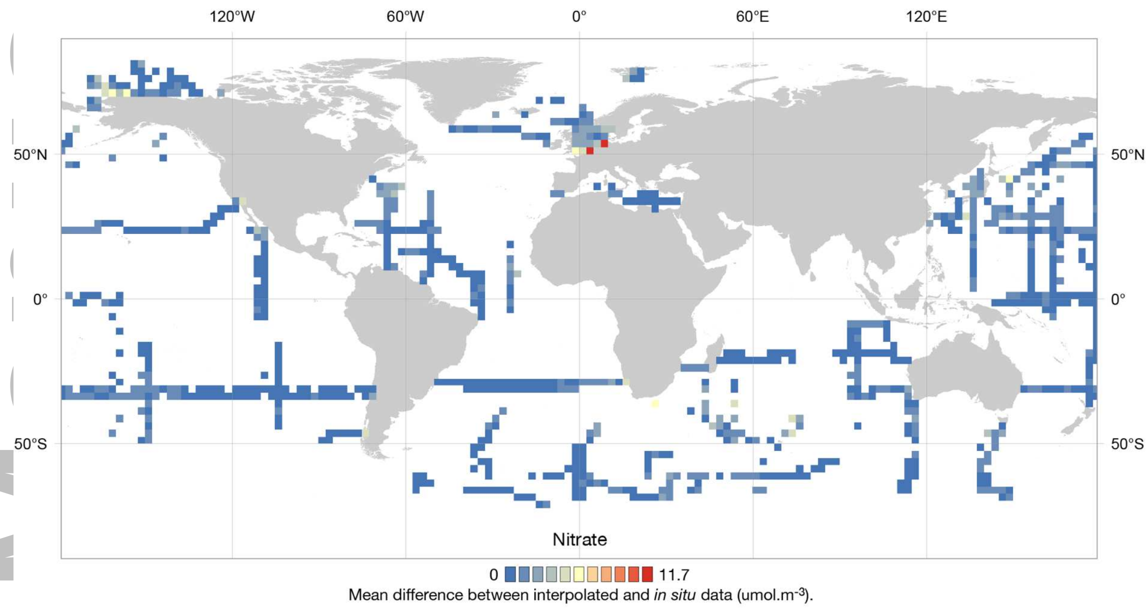


Fig 10. Spatial distribution of the error of nitrate data shown as the average difference between the interpolated and *in situ* data.

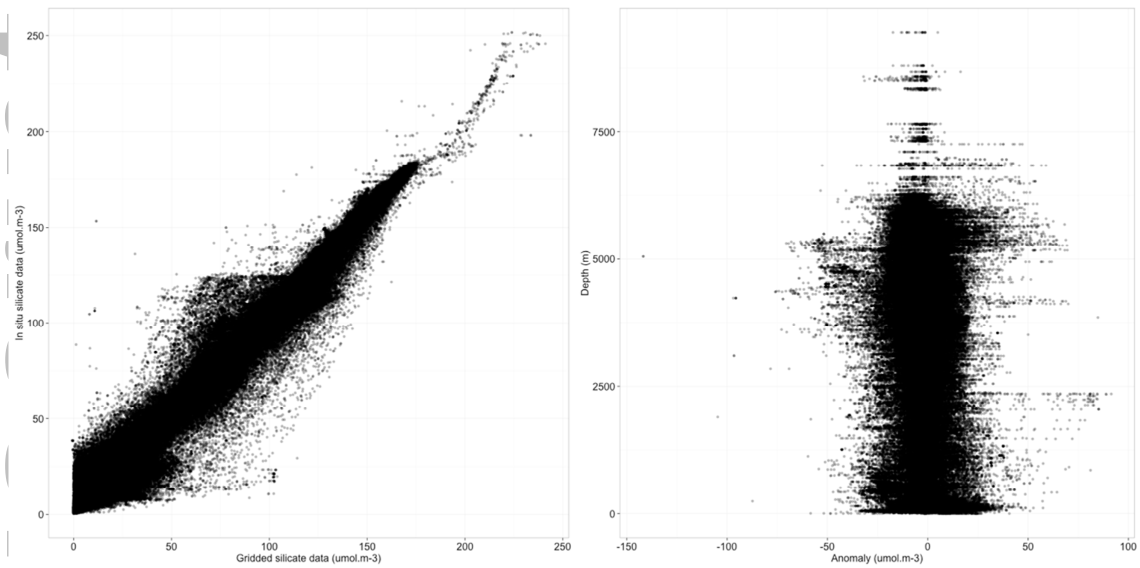


Fig 11. Accuracy of downscaled silicate data. (left panel) Correlation between the interpolated and *in situ* data for silicate. (right panel) Difference (anomaly) between the interpolated and *in situ* data against depth.

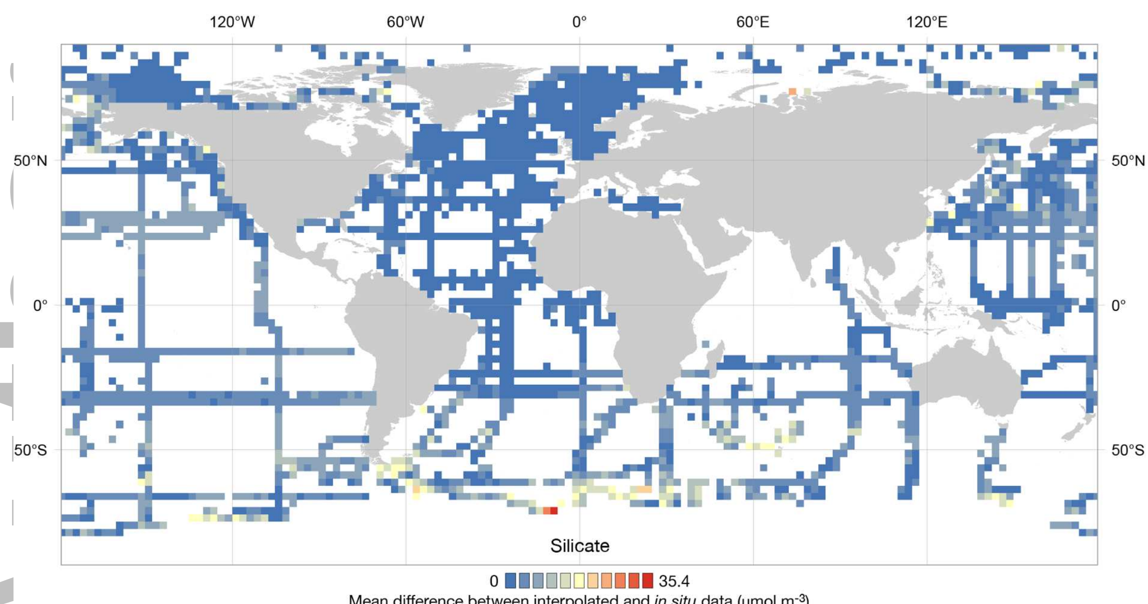


Fig 12. Spatial distribution of the error of silicate data shown as the average difference between the interpolated and *in situ* data.

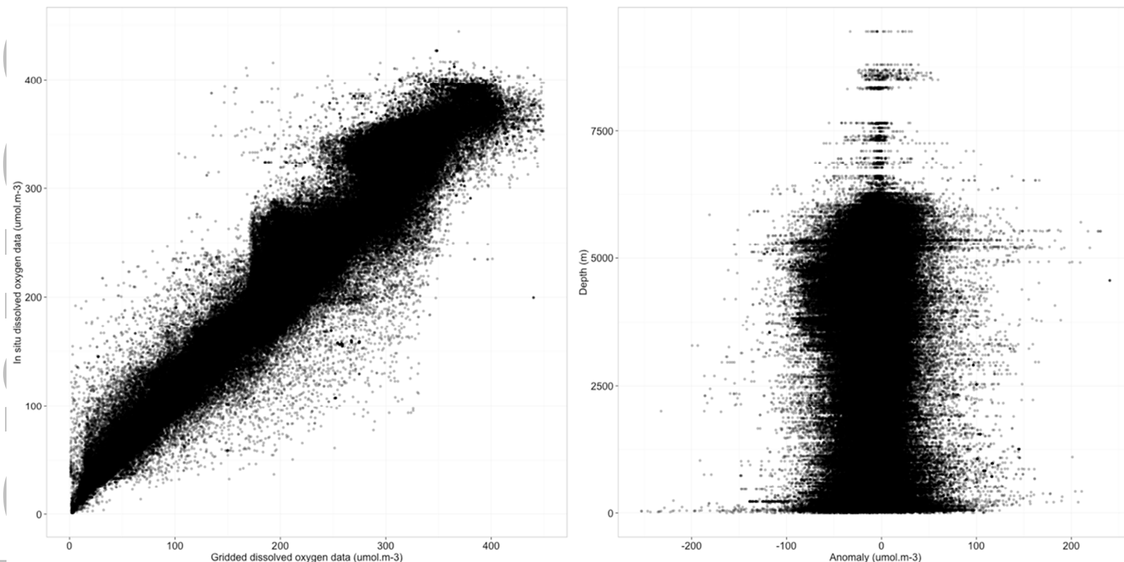


Fig 13. Accuracy of downscaled dissolved molecular oxygen data. (left panel) Correlation between the interpolated and *in situ* data for oxygen. (right panel) Difference (anomaly) between the interpolated and *in situ* data against depth.

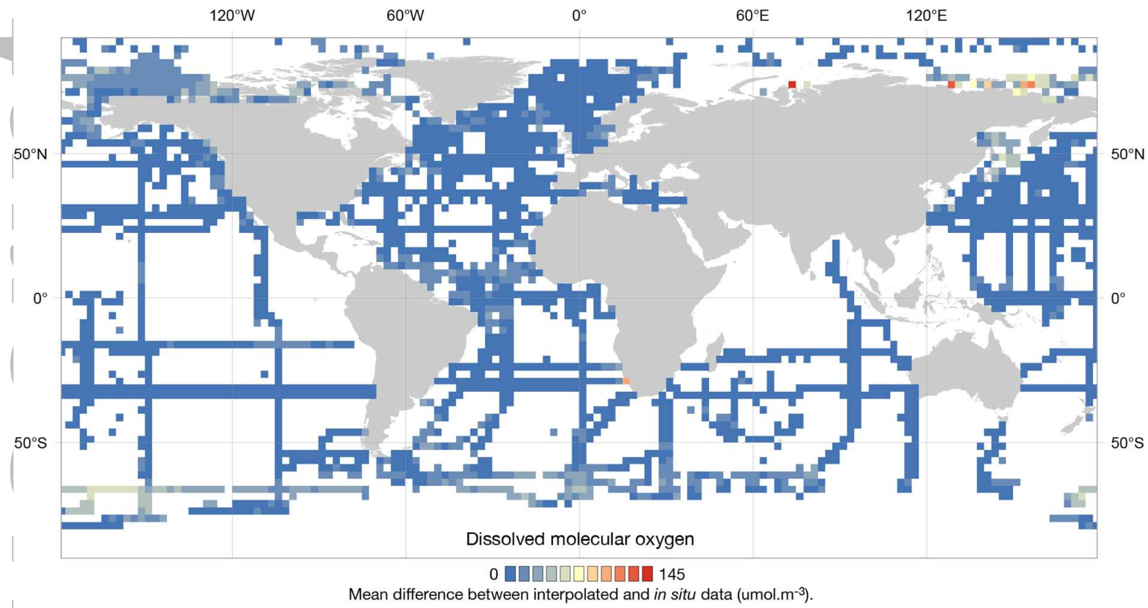


Fig 14. Spatial distribution of the error of dissolved molecular oxygen data shown as the average difference between the interpolated and *in situ* data.