

The impact of zooplankton calcifiers on the marine carbon cycle

¹Nielja S. Knecht, ¹Fabio Benedetti, ¹Urs Hofmann Elizondo, ^{2,3}Nina Bednaršek, ^{4,5,6}Sonia Chaabane, ⁷Catharina de Weerd, ^{7,8}Katja T. C. A. Peijnenburg, ⁶Ralf Schiebel, ¹Meike Vogt

¹Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zurich,

Switzerland

²National Institute of Biology, Marine Biological Station, Piran, Slovenia

³Cooperative Institute for Marine Resources Studies, Oregon State University, Oregon, USA

⁴Aix-Marseille Université, CNRS, IRD, INRAE, CEREGE, Aix-en-Provence, France

⁵French Foundation for Research on Biodiversity (FRB-CESAB), Paris, France

⁶Department of Climate Geochemistry, Max-Planck-Institute for Chemistry, Mainz, Germany

⁷Plankton Diversity and Evolution, Naturalis Biodiversity Center, Leiden, The Netherlands

⁸Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

Key Points:

- Calcifying zooplankton
- Species distribution models
- Carbonate fluxes

Abstract

Shelled pteropods and planktic foraminifers are calcifying zooplankton that contribute to the biological carbon pump, but their importance for regional and global plankton biomass and carbon fluxes is not well understood. Here, we modelled global annual patterns of pteropod and foraminifer total carbon (TC) biomass and total inorganic carbon (TIC) export fluxes over the top 200 m using an ensemble of five species distribution models (SDMs). An exhaustive newly assembled dataset of zooplankton abundance observations was used to estimate the biomass of both plankton groups. With the SDM ensemble we modeled global TC biomass depending on multiple environmental parameters. We found hotspots of mean annual pteropod biomass in the high Northern latitudes and the global upwelling systems, and in the high latitudes of both hemispheres and the tropics for foraminifers. This largely agrees with previously observed distributions. For the biomass of both groups, surface temperature is the strongest environmental correlate, followed by chlorophyll-*a*. We found mean annual standing stocks of 52 Tg TC (48 Tg TC to 57 Tg TC) and 0.9 Tg TC (0.6 Tg TC to 1.1 Tg TC) for pteropods and foraminifers, respectively. This translates to mean annual TIC fluxes of 14 Tg TIC yr⁻¹ (9 Tg TIC yr⁻¹ to 22 Tg TIC yr⁻¹) for pteropod shells and 11 Tg TIC yr⁻¹ (3 Tg TIC yr⁻¹ to 27 Tg TIC yr⁻¹) for foraminifer tests. These results are similar to previous estimates for foraminifers standing stocks and fluxes but approximately a factor of ten lower for pteropods. The two zooplankton calcifiers contribute approximately 1.5% each to global surface carbonate fluxes, leaving 40%–60% of the global carbonate fluxes unaccounted for. We make suggestions how to close this gap.

1 Introduction

Marine calcifying plankton play a key role in the ocean's carbon cycle, particularly through the formation, sinking, and dissolution of their CaCO₃ shells (J. L. Sarmiento & Gruber, 2006). These processes impact the carbonate system throughout the water column and thus also affect the oceanic CO₂ uptake (Takahashi & Bé, 1984; J. Sarmiento & Gruber, 2006). Annually, the inorganic carbon export flux from the surface ocean amounts to 0.6 Pg C yr⁻¹ to 1.4 Pg C yr⁻¹ (Iglesias-Rodriguez et al., 2002; Lee, 2001; Berelson et al., 2007; Jin. et al., 2006; Schiebel, 2002). However, there are significant uncertainties regarding the spatial and seasonal carbon flux patterns and the relative contribution of the different plankton groups to global calcification rates.

The major groups of calcifying plankton are coccolithophores, shelled pteropods and planktic foraminifers (Schiebel & Hemleben, 2017; Stepien, 1980; Lalli & Gilmer, 1989; Schiebel, 2002; Bednaršek, Mozina, et al., 2012). Shelled pteropods from the sub-order Thecosomata (in the following referred to as pteropods) build shells of aragonite, a metastable form of calcium carbonate (Lalli & Gilmer, 1989) with adults ranging from 1 mm to 30 mm in size (Bednaršek, Mozina, et al., 2012; Bednaršek, Tarling, et al., 2012). Aragonite is 50% more soluble than calcite (Mucci, 1983), which makes pteropods more sensitive to ocean acidification than calcite-shelled organisms (Fabry et al., 2008; Bednaršek et al., 2016; Manno et al., 2016). Pteropods are flux feeders, i.e., they secrete a floating mucus web to trap sinking organic particles (Gilmer & Harbison, 1986). They are active swimmers and some species perform diel vertical migration (DVM), feeding at night at the surface and spending the day at depths between 100 m and in some cases up to 1000 m (Bé & Gilmer, 1977; Bednaršek, Tarling, et al., 2012) to avoid predation. Foraminifers build calcareous tests that can reach diameters ranging between 100 µm and 1 mm (Frerichs et al., 1972; Schiebel & Hemleben, 2017). They are generally omnivorous and can capture prey actively, but feeding preferences differ between species (Rhumbler, 1911; Caron & Bé, 1984; Spindler et al., 1984; Anderson et al., 1979) with some species also harboring facultative photosymbionts (Hemleben et al., 1989). The global abundances and habitat suitability of pteropods and foraminifers are known to be controlled by a range of environmental parameters, including temperature (Beaugrand et al., 2010; Helaouët

& Beaugrand, 2009; Hofmann Elizondo et al., 2021; Jonkers et al., 2019; Bednaršek et al., 2022), chlorophyll-a as a proxy for food availability (Vereshchaka et al., 2022; Pinkerton et al., 2020; Thibodeau et al., 2019), and parameters related to physical mixing that influence and phytoplankton growth through light availability and particle sinking rates (Longhurst, 2007; Rothschild & Osborn, 1988; Boyce et al., 2010; Seuront et al., 2001; Govoni et al., 2010; Mackas et al., 2005; Bednaršek et al., 2022).

The relative importance of the different calcifying plankton groups for global carbonate fluxes remains uncertain. Coccolithophores were long thought to dominate the inorganic carbon export (Rost & Riebesell, 2004; Rembauville et al., 2016; Anglada-Ortiz et al., 2021; Iglesias-Rodriguez et al., 2002; Schiebel, 2002). However, in global observation-based estimates, they only accounted for 26%–52% of global carbonate fluxes, which leaves a significant fraction of the carbonate fluxes unattributed (Buitenhuis, Vogt, et al., 2013; C. J. O’Brien, 2015). This discrepancy shifted the attention towards the contribution of the two calcifying zooplankton groups, pteropods and foraminifers. Recent observational studies estimated pteropods to contribute more than previously thought to global surface carbonate fluxes with a fraction of 20% to 42% (Bednaršek, Mozina, et al., 2012). Foraminifer carbon flux estimates vary by a factor of 100 (Schiebel & Movellan, 2012; Schiebel, 2002; Buitenhuis et al., 2019; Buitenhuis, Vogt, et al., 2013). However, recent studies based on newly available observations find significantly lower fluxes. Finally, a recent mechanistic modelling study found pteropods to dominate upper subsurface CaCO_3 export, with contributions ranging between 33% - 89% (Buitenhuis et al., 2019). These results further suggest the key role of pteropods and foraminifers for the oceanic inorganic carbon cycle.

To derive the magnitude of carbon export mediated by zooplankton calcifiers, we first need to quantify the global biomass standing stocks and characterize the global distribution patterns of these groups. Earlier descriptions of the global patterns based on global plankton sampling data were made by the MARine Ecosystem DATA (MARE-DAT) project (Buitenhuis, Vogt, et al., 2013). Additionally, large-scale observational datasets have been collected by the Continuous Plankton Recorder (CPR) survey (Richardson et al., 2006). However, the existing observations are usually confined to specific ocean regions and have an overall low data coverage in the central oceanic basins (Bednaršek, Mozina, et al., 2012; Schiebel & Movellan, 2012; de Garidel-Thoron et al., 2022). Furthermore, plankton distributions are generally patchy in space and time (Boltovskoy, 1971; Beckmann et al., 1987; Siccha et al., 2012; Buitenhuis, Vogt, et al., 2013), which causes high variability in the observed abundances. Different sampling techniques and varying sampling depths and mesh sizes introduce additional variation (Wells, 1973). The derivation of continuous global biomass maps and standing stock estimates for zooplankton calcifiers hence requires us to account for these data gaps and biases by employing statistical methods.

As statistical techniques, species distribution models (SDMs) empirically learn the relation between the target variable and a range of environmental predictors through response curves and can then extrapolate said target variable to un-sampled regions by projecting these response curves on predictor values (Guisan & Zimmermann, 2000; Elith & Leathwick, 2009; Merow et al., 2014). They have been successfully used in marine macroecology to model plankton species distributions based on occurrence data (presence/absence) (Righetti et al., 2019; Benedetti et al., 2021; Brun et al., 2016; Barton et al., 2016; Brun et al., 2015; Bednaršek et al., 2022) and are increasingly being used to model continuous abundance values (Waldock et al., 2022; Pinkerton et al., 2010; De Broyer et al., 2014). In the present work, we apply an SDM framework to estimate global biomasses for calcifying zooplankton.

To this end, we use newly compiled global data compilation of pteropod and foraminifer abundances and species-specific biomass conversion methods to calculate biomass concentrations over the top 200 m. We combine the global gridded biomass data with an en-

semble of SDMs to address the following questions: (1) What are the biogeographic patterns and main environmental covariates of global total carbon (TC) biomass for pteropods and foraminifers (on a monthly, $1 \times 1^\circ$ gridded scale of the upper open ocean)? (2) What is the magnitude and range of uncertainty of the associated annual total inorganic carbon (TIC) fluxes from pteropods and foraminifers?

2 Methods

We model the biomass patterns and associated carbon fluxes of pteropods and foraminifers at a global scale using SDMs and updated abundance datasets for the two groups. To this end, we use a multi-step modelling pipeline as shown in figure 1.

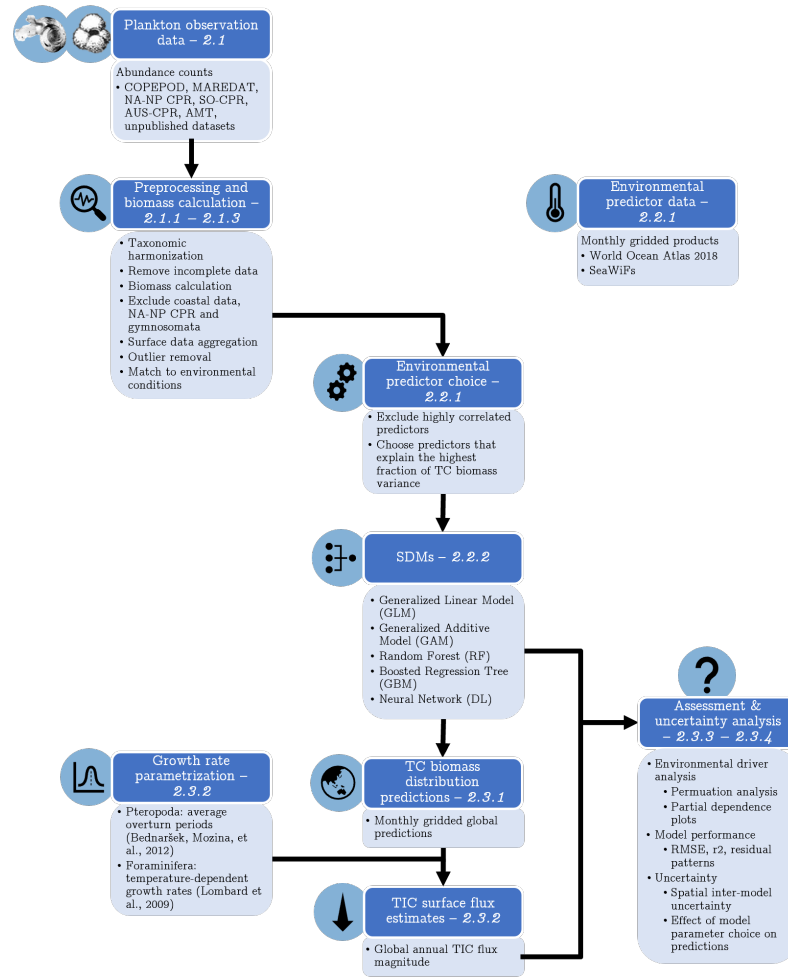


Figure 1. Flow diagram illustrating the pipeline of numerical analyses implemented for the present study. The various steps taken from the raw data to the final total carbon (TC) biomass distributions and total inorganic carbon (TIC) flux estimates using species distribution models (SDMs) are shown. The numbers in italics indicate the subsection of the Methods where the corresponding step is described.

2.1 Plankton data

2.1.1 Data collection and pre-processing

We updated the original MAREDAT pteropod and foraminifer abundance and biomass datasets of Schiebel and Movellan (2012) and Bednaršek, Mozina, et al. (2012) by aggregating abundance concentration data from large scale sampling campaigns, existing data compilation efforts, and unpublished sampling data (figure 1). The main data sources (figure S1) for both plankton groups included the Southern Ocean Continuous Plankton Recorder (SO-CPR) (Hosie, 2021), the Australian CPR (Aus-CPR) (IMOS, 2022), the North Atlantic and North Pacific CPR (NA-NP CPR) (Johns, 2021), and the Coastal and Oceanic Plankton Ecology, Production and Observation Database (COPEPOD) (T. D. O'Brien, 2010). For pteropods, we added data from the Tara Oceans expeditions (Brandão et al., 2021), the Atlantic Meridional Transect (AMT24) (Burridge et al., 2017) and AMT27 (Peijnenburg, 2021), as well as unpublished sampling data from the North Atlantic (Schiebel, 2021). For foraminifers, we also gathered data from various individual sampling campaigns (Schiebel et al., 1995; Schiebel & Hemleben, 2000; Schiebel et al., 2001; Schiebel, 2002; Schiebel et al., 2002, 2004; Jentzen et al., 2018).

We took several pre-processing steps to ensure the quality of the biological observations. To harmonize all classifications across datasets and correct for potential deprecated scientific species names, we matched all taxonomic information against the list of accepted taxon names of the World Register of Marine Species (WoRMS) (Horton et al., 2017). Observations lacking complete sampling metadata (date, depth, longitude, latitude, and abundance value) and observations of body parts were removed (21303 points for pteropods, mainly due to observations of body parts and larvae, and 522 for foraminifers). Additionally, pteropod abundance values from the Ecosystem Monitoring - Ships Of Opportunity surveys (EcoMon-SOOP) in the Gulf of Maine from the COPEPOD dataset were corrected by dividing them by a factor of 100 as the units in the original dataset had been erroneously reported. We did not standardize the abundance estimates between the various mesh sizes used in the different sampling cruises as there were not yet any published correction factors that we were aware of for these two specific plankton groups.

The final, quality-controlled pteropod abundance dataset (figure S2) contains 841239 data points at 309921 individual locations, collected at a mean sampling depth (\pm sd) of 38.15 ± 190.89 m over the 1938–2021 period (2001.25 ± 15.23). Abundances range between 0 ind/m³ and 1066.67 ind/m³, with a mean of 4.38 ± 79.86 ind/m³. The median abundance (0.00 ind/m³) is low due to the CPR datasets which make up 91.15% of the data, and contain 92.06% absence observations. 50.19% of the data is resolved only to the order-level, whereas 24.03% of the observations are species-resolved and 22.41% resolved to the genus level (see table S2). The dataset contains observations on 33 species out of 165 currently recognized pteropod species (Peijnenburg et al., 2020) (see table S2). The largest contributions to total abundance summed over all observations stem from *Limacina helicina* sensu lato (47.7% of the total species-resolved abundance), *Heliconoides inflatus* (26.7%), and *L. retroversa* s.l. (10.0%).

The final, quality-controlled foraminifer abundance dataset (figure S2) consists of 1021283 points at 308641 unique locations, with a mean sampling depth of 108.06 ± 340.49 m and collected during the 1939–2021 period (mean 2000.36 ± 13.30). Foraminifer abundances range between 0 ind/m³ and 152170.00 ind/m³, with a mean abundance of 3.63 ± 163.08 ind/m³. There is a high prevalence of CPR data (74.35% of the total data) with 89.72% zero abundance observations, which causes a low median abundance value of 0.00 ind/m³. 59.79% of the data are species resolved, followed by 33.07% of the observations on a phylum level (see table S4). This dataset contains observations on 42 of the 47 extant foraminifer species (Schiebel & Hemleben, 2017). Most of the total abundance is composed of *Glo-*

bigerina bulloides (25.6% of the total species-resolved abundance), *Neogloboquadrina incompta* (23.7%), *Turborotalita quinqueloba* (13.3%), and *Globigerinita glutinata* (11.3%).

For model training, we performed additional data quality controls to ensure sensible relations between environmental predictors and biomass values could be derived. The NA-NP CPR dataset was flagged and discarded for modelling as it contained discrete medians of abundance bins instead of continuous values (removal of 340250 points for pteropods and 250620 points for foraminifers). Additionally, we excluded data from neritic sampling locations associated with a climatological salinity < 30 PSU from the analysis to avoid observations influenced by terrestrial freshwater and nutrient inputs (Brun et al., 2015) (removal of 18725 data points for pteropods and 17207 points for foraminifers). Lastly, observations for pteropods from the clades Gymnosomata and Pseudothecosomata were removed for modelling, as only some of the latter are calcifiers (Lalli & Gilmer, 1989), and there is very little literature on their role in the carbon cycle (removal of 106929 points). The final datasets used for modelling contain 375336 points for pteropoda and 770663 points for foraminifers as shown in figure 2.

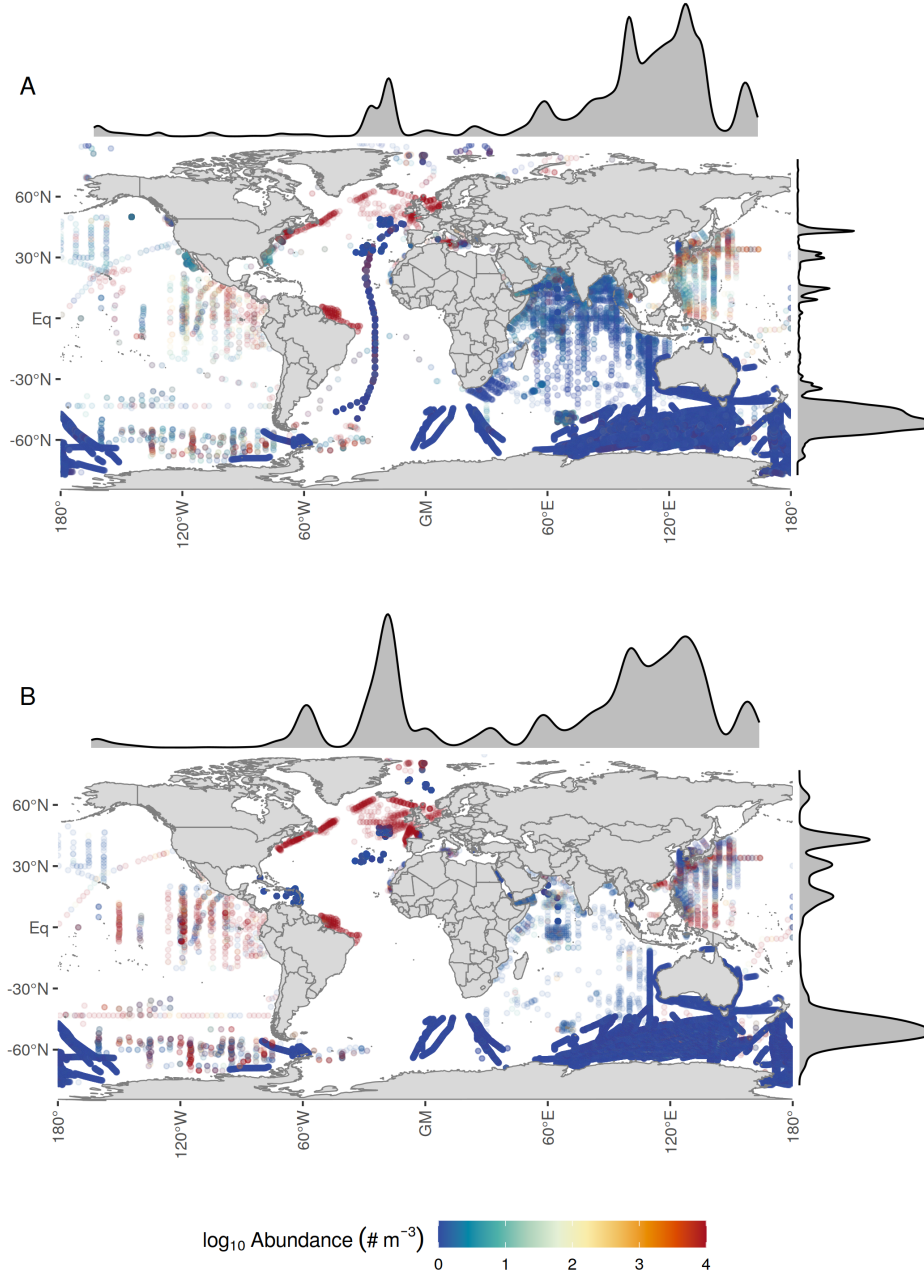


Figure 2. Global distribution of the final quality-controlled observations of pteropod (A) and foraminifer (B) abundance used for modelling. The marginal plots show the density of observations and highlight the dominant role of the Southern Ocean Continuous Plankton Recorder (SO-CPR) survey as well as a spatially confined, highly resolved dataset in the North Atlantic. This plot shows the dataset used for modelling, i.e., the dataset after removing the North Atlantic and North Pacific CPR data, coastal observations with surface salinity ≤ 30 , and observations of naked pteropods (Gymnosomata) as described above. For the full collected dataset, see figure S2.

2.1.2 Biomass calculations

To estimate calcifying zooplankton biomass and subsequent carbon fluxes, we converted the abundances to biomass data based on morphology-based conversion factors (cf. figure 1). To this end, we grouped species of similar morphology into shape groups and derive biomass as a function of average body size (maximum elongation) based on shape-specific conversion equations. Generally, we applied all conversions on the lowest taxonomic level available and only used shape-group or phylum-wise averages where the species identification was not available.

Biomass calculation for pteropods To convert pteropod abundance into carbon biomass, we used corrected species-specific biomass conversion equations from Bednaršek, Mozina, et al. (2012) to calculate wet weight (WW) as shown in table S1. These equations are based on six different morphological shape groups and relate an individual species' body length in millimeters to its biomass. For observations without morphometric data (99.8%), we used the species-average lengths from (Bednaršek, Mozina, et al., 2012). We used pteropod shell length whenever given in Bednaršek, Mozina, et al. (2012), otherwise we used the body length values from the same source. Table S2 shows the average length value used for each species, their respective shape group, and the number of observations for each species. WW was then transformed to dry weight (DW) as per Davis and Wiebe (1985) (equation 1)

$$DW = WW \cdot 0.28 \quad (1)$$

and subsequently transformed to total carbon (TC) following Larson (1986) (equation 2).

$$TC = DW \cdot 0.25 \quad (2)$$

Finally, total inorganic carbon (TIC) was computed (equation 3) following Bednaršek, Mozina, et al. (2012).

$$TIC = 0.27 \cdot TC \quad (3)$$

This TC-TIC conversion factor is based on data for *L. helicina antarctica* and hence probably not representative for all pteropod species and life stages (Hofmann Elizondo & Vogt, 2022). To account for the lack of species-specific TC-TIC conversion factors in literature, we added an uncertainty range of $\pm 20\%$ to the conversion factor, based on the range of TIC values reported in Bednaršek, Tarling, et al. (2012). The effect of this parameter choice is assessed according to the methodology in section 2.3.4.

Biomass conversion for foraminifers A morphological approach was also carried out for converting foraminifer abundances to TC concentrations. We were not aware of any published shape class definitions for foraminifers. Thus, we defined eight morphological shape groups based on similar adult test shape and structure as shown in table S3.

To derive biovolume-to-biomass conversion equations, we constructed species and group-specific maximum test length to biomass functions from the literature. We collected species-specific test weight measurements per plankton size class from Schiebel and Hemleben (2000) and Takahashi and Bé (1984). We fitted linear functions to calculate biomass as a function of length per species and per shape group, where the biomass of a shape group is calculated as the average of all species within the group (figure S7). To compare the ranges of the conversion factors to published equations for the entire foraminifera phylum, we used the equation provided by Michaels et al. (1995) (figure S7). This function computes foraminifer cytoplasm carbon (i.e., total organic carbon, TOC) as a function of test length. The TC biomass is calculated based on the following conversion fac-

tors (Schiebel & Movellan, 2012):

$$TIC = 0.36 \cdot TOC \quad (4)$$

$$TIC = 0.265 \cdot TC \quad (5)$$

To compute TC from the test weight measurements of Schiebel and Hemleben (2000) and Takahashi and Bé (1984), we used the following molar relationship:

$$m(C) = \frac{m(\text{CaCO}_3)}{M(\text{CaCO}_3)} \cdot M(C) = \frac{m(\text{CaCO}_3)}{100.09 \text{ g mol}^{-1}} \cdot 12.01 \text{ g mol}^{-1}, \quad (6)$$

where m denotes the mass and M the molar weight.

The biomass conversion factors (BCF) shown in table S5 are the coefficients of the linear relation between foraminifer TC biomass and their biovolume. To apply the conversion factors, the BCF values were substituted into the following equation:

$$TC = a * L^3 * BCF, \quad (7)$$

where TC represents the TC biomass of foraminifers in μg , a denotes foraminifer abundance and L the species' length in μm .

We collected average length values for all species from the images of (Schiebel & Hemleben, 2017). These average length values as well as the number of observations per species can be found in table S4.

2.1.3 Surface ocean aggregation

To reduce spatio-temporal patchiness and noise in the data, we conducted a surface ocean aggregation (C. J. O'Brien, 2015). To this end, we re-gridded all data onto the $1 \times 1^\circ$ grid of the World Ocean Atlas 2018 (WOA18; Boyer et al. (2018)). For each grid cell, we summed all biomass concentrations from the same sampling event, as different species were sometimes counted as separate measurements. Next, we averaged all biomass and abundance values per grid cell and month of the year over the top 200 m. This depth cutoff was deemed reasonable as 99.1% and 99.4% of the summed abundance of pteropods and foraminifers, respectively, stem from the top 200 m (figure S3).

To better approximate a normal distribution, TC mass values were log-transformed with a $\log_{10}(\text{TC}+1)$ transformation for further analyses. Lastly, to dampen the effect of plankton patchiness and bloom dynamics, we flagged outliers in the surface aggregated values based on the z-score criterion (Burba & Anderson, 2005). Hence, for modelling, we excluded high biomass observations with a score of $z > 3$, i.e., more than three standard deviations away from the sample mean.

2.2 Modelling

2.2.1 Environmental predictor selection

To identify the set of predictors used for training the biomass-based SDMs, we collected gridded monthly climatologies of meaningful environmental predictors as shown in table 1 and figure 1. Whenever necessary, the fields were averaged and re-gridded to monthly climatologies at a $1 \times 1^\circ$ resolution. Depth-resolved predictors from the WOA18 were averaged over the climatological mixed layer depth (MLD). As many pteropods actively migrate vertically (on a daily or seasonal basis) and both groups are passively vertically mixed within the water column (Mackas et al., 2005; Lalli & Gilmer, 1989; Schiebel & Hemleben, 2017; Wormuth, 1981; Myers, 1968), the depth-averaged environmental predictors are more representative of the conditions they experience rather than the surface values. However, as sampling devices are often towed vertically or obliquely, the reported water depth interval of each observation is not directly representative of the depth

Table 1. Environmental predictor variables used in the univariate predictor evaluation.

WOA18 refers to the 2018 edition of the World Ocean Atlas (Boyer et al., 2018), SeaWiFS denotes the Sea-viewing Wide Field-of-view Sensor satellite data (OB.DAAC, 2018) and SODA describes the Simple Ocean Data Assimilation project (Carton et al., 2018).

Predictor	Source	Reference
Temperature	WOA18	Locarnini et al. (2018)
Chlorophyll-a	SeaWiFS	NASA OB.DAAC (2018a)
Mixed layer depth (MLD)	SODA3.4.2	Carton et al. (2018)
Eddy kinetic energy (EKE)	Copernicus	Copernicus (2021)
Salinity	WOA18	Zweng et al. (2019)
Dissolved oxygen	WOA18	Garcia et al. (2019b)
Nitrate	WOA18	Garcia et al. (2019a)
Phosphate	WOA18	Garcia et al. (2019a)
Depth of the euphotic layer (z_{eu})	SeaWiFS	NASA OB.DAAC (2018c)
Photosynthetically active radiation (PAR)	SeaWiFS	NASA OB.DAAC (2018e)
Particulate backscattering coefficient at 443 nm (BBP_{443})	SeaWiFS	NASA OB.DAAC (2018d)
Diffuse attenuation coefficient for downwelling irradiance at 490 nm (Kd_{490})	SeaWiFS	NASA OB.DAAC (2018b)
Total alkalinity (TA)	OceanSODA-ETHZ	Gregor and Gruber (2021)
Dissolved inorganic carbon (DIC)	OceanSODA-ETHZ	Gregor and Gruber (2021)
Partial pressure of CO ₂ (pCO_2)	OceanSODA-ETHZ	Gregor and Gruber (2021)
Calcite saturation state (Ω_{Ca})	OceanSODA-ETHZ	Gregor and Gruber (2021)
Aragonite saturation state (Ω_{Ar})	OceanSODA-ETHZ	Gregor and Gruber (2021)

an organism dwells at over the entire day or even through its life span. Hence, we assume that pteropods and foraminifers move within the mixed layer, where the majority of the organic matter is present (Sallée et al., 2021; Soviadan et al., 2022). For all depth-resolved environmental predictors considered, the average over the top 200m, the values at the surface and the MLD-averaged predictor values are each correlated with a Pearson correlation coefficient of $r > 0.99$, so this simplification is deemed reasonable. For dissolved oxygen concentration, we used the value at 200 m depth to avoid the strong collinearity with the sea surface temperature (SST) values. The distribution of chlorophyll-a concentrations, nutrient variables, MLD, and eddy kinetic energy (EKE) were right-skewed (figure S9), therefore we log-transformed those variables so their distribution is closer to a normal one. Then, we collocated the environmental parameters with the gridded monthly pteropod and foraminifer biomass fields.

To select the most meaningful environmental predictors for the final biomass-based SDMs we used a multi-step approach for each zooplankton group. First, we identified clusters of collinear predictors (Pearson correlation coefficient $|r| > 0.7$ calculated from the values matched up with the monthly biomass climatologies, Brun et al. (2020)). Second, we excluded all but one predictor in each cluster, which improves model performance (Dormann et al., 2013; Brun et al., 2020) (figures S10 and S11). Thus, for each cluster we first chose the most normally distributed predictor as assessed by the Shapiro-Wilk test (Shapiro & Wilk, 1965), and second, we chose predictors whose effect are easier to interpret from an ecological point of view (e.g. chlorophyll-a over Kd_{490} , the remotely sensed light attenuation at a wave length of 490 nm, which is an indirect measure of surface productivity and turbidity). This selection procedure resulted in the following seven candidate predictors for both foraminifers and shelled pteropods: surface chlorophyll-a, MLD, temperature averaged over the MLD, surface EKE, oxygen at 200 m depth, salin-

ity averaged over the MLD, partial pressure of CO₂ (pCO₂), photosynthetically active radiation (PAR), and particulate backscattering coefficient at 443 nm (BBP₄₄₃).

The exclusion of a predictor variable does not mean that it is not ecologically relevant for the organisms modelled. The calcite and aragonite saturation states are known to influence habitat suitability for foraminifers and pteropods, respectively (Lischka et al., 2011; Lischka & Riebesell, 2012; Manno et al., 2016; Bednaršek et al., 2016, 2022). However, the matched saturation states were determined to be highly correlated with water temperature averaged over the MLD (Pearson $r > 0.99$). As previous studies have shown temperature to be more biologically relevant in influencing large-scale biogeographic distribution patterns (Bednaršek et al., 2018; Howes et al., 2015; Beaugrand et al., 2013; MacKas & Galbraith, 2012), we excluded the saturation states as predictors. Exchanging temperature for the aragonite saturation state in the pteropod models does not have a significant effect on the biomass distribution pattern or the annual TC fluxes (figure S22).

To choose the final predictor set, we assessed the variance of the TC biomass explained by each of the seven candidate predictors using univariate regression models (figure S12). For this, we calculated both 1° pixel-wise and latitudinal 10°, 5°, and 1° monthly means of the TC biomass and the environmental predictors to identify the large-scale effects of the environmental predictors. To model variations in the TC biomass as a function of each environmental predictor, we trained two Generalized Linear Models (GLMs) with a Gaussian response function (one with only a linear term and the second with both a linear and a quadratic term) and a Generalized Additive Model (GAM) with a cubic regression spline. Then, we assessed the percentage of deviance explained by each predictor (Hosmer Jr et al., 2013; Nelder & Wedderburn, 1972). We retained all predictors that explained $\geq 5\%$ of variability at any of the spatial aggregation levels. For pteropods, the resulting set of predictors included: MLD-averaged temperature, surface chlorophyll-a, and MLD. For foraminifers, we retained the MLD-averaged temperature, surface chlorophyll-a, and EKE (figure S12 and figure S13 for mean annual maps of the predictors).

To assess the impact of this predictor selection procedure on SDM outputs, we also trained the models for both plankton groups on a Principle Component Analysis (PCA) transformation of the full initial predictor set (table 1). There was no significant difference between the PCA-based global annual TIC fluxes and those calculated based on our final choice of predictors ($p > 0.05$ for both plankton types as assessed with a Kruskal-Wallis test (Kruskal & Wallis, 1952), see figure S23). This shows that the predictor selection procedure did not substantially affect the SDMs estimates.

2.2.2 Multivariate modelling

We used the identified predictors to train an ensemble of five SDMs of increasing complexity: a GLM, a GAM, a Random Forest (RF), a Gradient Boosting Machine (GBM) and a Neural Network/Deep Learning Model (DL) (see figure 1). GLMs, GAMs and RFs have been widely and successfully used in the modelling of global marine plankton distributions (Righetti et al., 2019; Benedetti et al., 2021; Brun et al., 2016). The more complex models have also been used for modelling plankton distributions, though less frequently (GBMs in Pinkerton et al. (2020, 2010), DL models in C. J. O'Brien et al. (2016); Benedetti et al. (2021)). For an extensive description of the more complex model types, we refer to Boehmke and Greenwell (2019e, 2019b, 2019a) and sources within. All modelling was conducted with the `h2o 3.36.0.3` R package (H2O.ai, 2021).

For the GLM, we included both first and second-order dependencies on the predictors and assumed a normal distribution of the target variable with an identity link function (Nelder & Wedderburn, 1972). In the GAM, we fitted smoothing terms for all predictor variables using cubic regression splines, the most common smoothing algorithm

(Hastie & Tibshirani, 1990), and a normal distribution with the identity function as link for the target variable. For the RF, GBM, and DL, the hyperparameters were tuned using a grid search (Boehmke & Greenwell, 2019d). Tables S6, S7, and S8 show the grid of parameters evaluated for each model. The final setup of the RF as determined from the tuning process (table S6) included 830 trees for pteropods and 330 for foraminifers. At each tree node, one and two environmental predictors were evaluated (m_{try}) for pteropods and foraminifers, respectively, and the minimum number of rows at each final node (min_{rows}) was set to three and two. The maximum tree size was constrained to 30 for pteropods and 10 for foraminifers. For each bootstrap replicate of the tree, we chose a fraction (r_{sample}) of 0.8 and 0.632 of the total dataset. For the GBM, we determined a maximum depth ($max_{depth} = 5$) and minimum number of observations per terminal node ($min_{rows} = 1$) for each individual tree for both plankton groups (see also table S7). The learning rate (r_{learn}) was determined to be 0.01 and each individual tree is trained on a fraction of 0.75 and 0.5 of the total dataset for pteropods and foraminifers, respectively, using all of the predictor columns ($r_{samplecolumns}$). The DL (see also table S8) was determined to have a Tanh activation function for both plankton groups. The pteropod model has two hidden layers with 20 neurons each and the foraminifer model has two three hidden layers of 15 neurons each. To avoid overfitting, L_1 and L_2 regularizations were included (Boehmke & Greenwell, 2019a) with weight factors $\lambda_{L_1} = 0$ and $\lambda_{L_2} = 1 * 10^{-3}$ for pteropods, and $\lambda_{L_1} = 1 * 10^{-3}$ and $\lambda_{L_2} = 1 * 10^{-5}$ for foraminifers.

We assessed the effect of the hyperparameter tuning on the global annual TIC fluxes by comparing the fluxes calculated using the tuned models (for the RF, GBM, and DL) to those based on the untuned models with standard hyperparameter set-up (see tables S6–S8). As expected, the tuned models showed a better model performance, but the global annual TIC fluxes did not differ significantly ($p > 0.05$ as assessed with a Kruskal-Wallis test (Kruskal & Wallis, 1952) for each plankton group). Tuning the models hence does not introduce unfounded model complexity or biases.

To train the SDMs and assess their performance, we split the dataset into a training and a testing set (Boehmke & Greenwell, 2019d). For a conservative estimate of model performance, we randomly assigned 75% of the values to the training dataset. On the training dataset, we performed a 5-fold cross validation, where we (i) split the training dataset into five equally-sized, randomly chosen, non-overlapping subsets, (ii) train the SDMs on four of the subsets, and (iii) evaluate the model performance of the trained SDM on the remaining subset based on the average root mean squared error (RMSE). This procedure was repeated until each of the five subsets of the data were used four times for training and once for validation. Finally, we evaluated the trained SDM on the testing set.

2.2.3 Model performance

We assessed model performance using three metrics (figure 1). The root mean squared error (RMSE) is an error metric estimating the deviation between predicted and true values. Pearson’s coefficient of correlation, R^2 indicates the magnitude of correspondence between trends in the predicted and observed values. Finally, the Nash-Sutcliffe-efficiency (NSE; Nash and Sutcliffe (1970)) compares the model performance to a null model, i.e., the mean of all observations. Positive NSE values indicate that the assessed model performs better than the null model. Each performance metric was calculated on both the training and the testing set of the data (cf. section 2.2.2).

2.3 Model inference

2.3.1 Global total carbon (TC) biomass patterns

We used the SDMs to project global monthly TC biomass values as a function of the monthly climatological environmental predictors (see figure 1). Projections were made for each grid cell and month where environmental data were available. We flagged and excluded all predictions of negative biomass values, because they correspond to unrealistic predictions (0.33% of all predicted values for pteropods and 0.06% for foraminifers). Many complex SDMs suffer from low transferability into novel environmental conditions due to non-linear response curves (Elith et al., 2010; Qiao et al., 2019; Bell & Schlaepfer, 2016). Thus, for each grid cell we evaluated whether the environmental conditions lie within the range of the training dataset or are considered non-analog using a Multivariate Environmental Similarity Surfaces (MESS) analysis (Elith et al., 2010). The MESS analysis assesses the similarity between the environmental conditions at any given point and the training dataset of each SDM. To avoid including unrealistically high values in the flux calculations and global summaries of calcifying zooplankton biomass, we excluded the biomass values from regions where non-analog environmental conditions were detected by the MESS analysis (3.25% of the values for pteropods and 4.03% for foraminifers). To analyze the spatial biomass patterns, we defined hotspots as unusually high biomass concentrations that lie above the 90th percentile for each plankton group.

2.3.2 Annual total inorganic carbon (TIC) export fluxes

We computed TIC fluxes from the projected global TC biomass values and environmental conditions (see figure 1). To compare our results to those of Buitenhuis et al. (2019), biomass values were calculated for TC, while export fluxes were based only on the inorganic shells, i.e., on TIC. Hence, we assumed that the carbon export flux is dominated by the sinking and empty shells.

TIC export flux calculation for pteropods To compute the annual pteropod TIC flux, we applied a simplified approach based on an average overturn time of one year, following the methodology of Bednaršek, Mozina, et al. (2012). Based on grid cell-wise mean annual biomass concentrations, we computed the global annual mean biomass as the spatially weighted mean of the average concentrations of each grid cell, multiplied by the TIC-TC factor (Bednaršek, Mozina, et al., 2012), the depth of 200 m, and the global open ocean area excluding shelf seas ($362e6 \text{ km}^2$) (Bednaršek, Mozina, et al., 2012; Dietrich et al., 1975). To represent the variability and uncertainty in turnover times between various pteropod species and regions, we added an uncertainty factor of $\pm 20\%$ to the flux conversion equation based on the range of values given in the review study by Wang et al. (2017). The effect of this parameter choice is evaluated according to the uncertainty analysis described in section 2.3.4.

TIC export flux calculation for foraminifers To calculate foraminifer TIC fluxes, we used the phylum-resolved temperature-dependent growth rates from Lombard et al. (2009). To calculate annual TIC fluxes, we multiplied the daily growth rate at each grid cell and month by the current biomass concentration, the TIC-TC factor (equation 4 in section 2.1.2), and the depth of 200 m, and weighted the result by grid-cell area. To represent uncertainty in the growth rate, we calculated the minimum and maximum growth rates by computing all combinations within the parameter uncertainty range. Then, we chose those parameter combinations that would minimize or maximize the integral of the growth rate as a function of temperature from 0 to 30°C , while maintaining ecologically-sensible response shapes (see figure S8 for an illustration of the growth rate options). The effect of this choice was evaluated according to the methodology described in section 2.3.4.

2.3.3 Environmental predictor analysis

To examine how underlying ecological processes were captured by the SDMs (figure 1), we assessed the models' dependence on the predictor variables in two ways. First, we assessed the overall effect of each environmental predictor based on a permutation analysis using the Fisher-Yates algorithm (Fisher & Yates, 1953). Second, we characterized the biological relevance of the response curve learned by each SDM using partial dependence plot (PDP) curves. The PDP curves were calculated by computing biomass prediction at 25 evenly spaced points across each predictor's range while keeping all other predictors constant at their mean value (Boehmke & Greenwell, 2019c).

2.3.4 Uncertainty quantification

We assessed the three main sources of uncertainty underlying our SDMs predictions: SDM choice (Thuiller et al., 2019), TIC-TC factor, and growth rate parametrization (figure 1). First, we identified potential non-normal relationships based on the patterns of the model residuals. Second, we quantified the effect of different model and parameter choices (see sections 2.1.2, 2.2.2 and 2.3.2 for details on the uncertainty setup) on the carbon flux predictions using a multivariate Analysis of Variance (mANOVA; Weinfurt (1995)) whose target variable was the monthly TIC flux values at each grid cell. We used the model type, the growth rate definition, the TIC-TC conversion factor, and the interactions between these three factors as input for the mANOVA.

3 Results

3.1 Global biogeographic total carbon (TC) biomass patterns

The global mean annual TC biomass (\pm sd) is $0.701 \pm 0.648 \text{ mg TC m}^{-3}$ for pteropods, and $13.5 \pm 28.7 \text{ } \mu\text{g TC m}^{-3}$ for foraminifers, implying that pteropod biomass is a factor of 50 larger than foraminifer biomass. The projected global mean biomass patterns are shown in figure 3A and 3B for pteropods and foraminifers, respectively. For both plankton groups, high biomass concentrations are found in the tropics and at latitudes $\geq 50^\circ\text{N}$. Lower biomass concentrations (mean values of $0.31 \text{ mg TC m}^{-3}$ and $5 \text{ } \mu\text{g TC m}^{-3}$ are found between 40°S and 50°S for pteropods and between 30° and 40° in both hemispheres for foraminifers. Contrary to pteropods, we find high biomass concentrations of up to $880 \text{ } \mu\text{g TC m}^{-3}$ for foraminifers in the Southern Ocean south of 50°S .

On a regional scale, the North Atlantic Ocean is associated with biomass hotspots (values above the 90th percentile) for both plankton groups, but particularly for foraminifers. A trail of high foraminifer biomasses with a mean value of $150 \text{ } \mu\text{g TC m}^{-3}$ is found across the North Atlantic that is likely associated with the Gulf Stream. Other regions of high biomass are associated with tropical and coastal upwelling systems. Pteropod biomass concentrations are particularly high in the coastal Eastern Boundary Upwelling Systems (EBUS) with an average concentration of 3 mg TC m^{-3} . For foraminifers, regions of high biomass are associated with the equatorial upwelling region.

On a seasonal scale, biomass hotspots shift towards high latitudes during global summer (figures S16 and S17). The seasonal variation in biomass is stronger in the Northern Hemisphere (NH) than in the Southern Hemisphere (SH) with a difference in variability V ($V_{NH} - V_{SH}$) of $+0.73 \text{ mg TC m}^{-3}$ for pteropods and of $+45.59 \text{ } \mu\text{g TC m}^{-3}$ for foraminifers ($p < 2 \cdot 10^{-16}$ for both groups; t-test (Student, 1908) where seasonal variability is computed from the maximum difference between the monthly mean surface ocean biomass concentrations at each grid point per model type). Foraminifers display a higher seasonal variation than pteropods ($+0.28$, $p < 2 \cdot 10^{-16}$ when comparing the maximum seasonal variation at each grid point normalized by the mean global biomass between the plankton groups with a t-test).

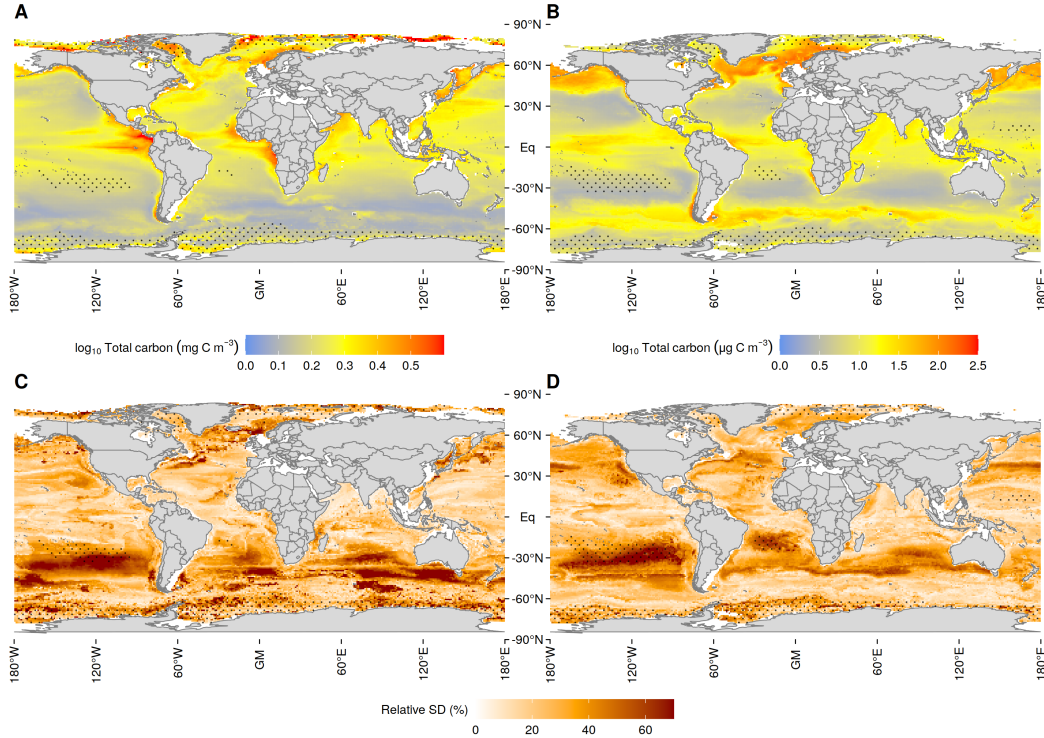


Figure 3. Global mean annual total carbon (TC) biomass concentration for pteropods (left, **A**) and foraminifers (right, **B**), averaged over all months and models. Values are shown as $\log_{10}(TC + 1)$, note also the different color scales for pteropods and foraminifers. Stippled regions in plots **A** - **D** indicate grid points where the environmental conditions were outside the training dataset for more than six months of the year as calculated with the Multivariate Environmental Similarity Surfaces (MESS) analysis. The lower panel plots **C** and **D** show the mean annual relative standard deviation of the model predictions for pteropods (left) and foraminifers (right), normalized with the mean prediction value at each grid point to facilitate comparability.

3.2 Model performance

To assess model performance of the five SDMs, we evaluated each model using the root mean squared error (RMSE), the R^2 and the Nash-Sutcliffe-Efficiency (NSE) as shown in table 2 for both plankton groups. Compared to the GLM and GAM, the more complex model types (RF, GBM, and DL) have a lower RMSE, a higher R^2 , and a higher NSE, i.e., they generally perform better across all three metrics (table 2, see also section 2.2.3 for a description of the metrics). For both pteropods and foraminifers, the RF performs best, followed by the GBM. However, the GBM's R^2 is significantly higher on the training set than on the testing set (2), which indicates model overfitting. The same pattern is visible for RMSE (table 2). In contrast, the RF achieves similar performances on the training and testing set, which indicates a robustly high performance. All model types perform better than using the mean observation value as prediction, which is indicated by the positive NSE values (table 2). Comparing the R^2 values between the plankton groups shows that the pteropod models generally perform better and can explain a higher fraction of the biomass variability (table 2). For the complex non-parametric models (RF, GBM, DL), R^2 is not an optimal metric (Spiess & Neumeier, 2010). However,

Table 2. Model performance for the pteropod and foraminifer models. Each model metric was calculated on both the training set (X_{train}) and the testing set (X_{test}). R^2 ranges from $-\infty$ to $+1$, with a perfect fit of the model and full variance explained indicated by a value of $+1$. The root mean squared error (RMSE) is an error measure, hence smaller values show higher accuracy. The Nash-Sutcliffe-efficiency (NSE) indicates improvement of the model predictions over using the observation mean, with perfect model performance indicated by a value of $+1$ and a value of 0 indicating that the models perform no better than the observation mean. The models are ranked by their performance over the five metrics.

	Model	R^2_{train}	R^2_{test}	$RMSE_{train}$	$RMSE_{test}$	NSE_{train}	NSE_{test}	Ranking
Pteropoda	GLM	0.1113	-0.6427	0.2612	0.3633	0.1113	0.1442	5
	GAM	0.1299	0.1678	0.2585	0.2586	0.1299	0.1678	4
	RF	0.2332	0.2805	0.2408	0.2404	0.5581	0.2805	1
	GBM	0.409	0.2674	0.2114	0.2426	0.3652	0.2674	2
	DL	0.1597	0.1822	0.2521	0.2563	0.1625	0.1822	3
Foraminifers	GLM	0.0503	-0.0279	0.8554	0.8789	0.0503	0.0491	5
	GAM	0.1116	0.0823	0.8274	0.8304	0.1116	0.0823	4
	RF	0.2424	0.2003	0.7586	0.7752	0.4252	0.2003	1
	GBM	0.3999	0.1926	0.6751	0.7789	0.3594	0.1926	2
	DL	0.1718	0.1367	0.7931	0.8054	0.1780	0.1367	3

as it is frequently reported in plankton studies as a measure of the fraction of variance explained (Zurell et al., 2020; Pinkerton et al., 2010, 2020), we chose to still include it.

All models tend to underestimate the total biomass on a global scale (-35% for pteropods and -5% for foraminifers of log-transformed biomass), with a stronger underestimation of the top 10th percentile biomass hotspots (on average -78% for pteropods and -53% for foraminifers). However, this underestimation is less pronounced in the more complex models (figures S19 and S20). On a basin-scale, highly productive regions are generally underestimated and low productivity areas overestimated with an average overestimation of the lowest 50% of log-transformed biomass by a factor of 8.7 for pteropods and a factor of 2.5 for foraminifers. Hence, biomass concentrations of both plankton groups are underestimated in the North Atlantic Ocean and the tropical Pacific and Atlantic, whereas predictions in the Indian Ocean and the region around Australia are on average too high (figures S19 and S20).

3.3 Environmental covariates

In general, the modeled responses of biomass to the fitted predictors converges across the ensemble members, except near the outer ranges of the predictor values, and for EKE (figure 4). Temperature shows an overall positive relation to pteropod biomass and a bimodal relation for foraminifer biomass with peaks around 5° to 7°C and above 25°C . Chlorophyll-a is positively related to both pteropod and foraminifer biomass. At high chlorophyll-a concentrations ($\text{Chl} - a > 1 \text{ mg m}^{-3}$), biomass concentrations stagnate for pteropods and decrease slightly for foraminifers. MLD has a negative parabolic relation to pteropod biomass. Deepening MLD up to 30 m causes a decrease in biomass while a further deepening of the MLD leads to an increase in biomass concentrations. The effect of EKE on foraminifer biomass varies across the models, with a strong positive effect in the simpler GLM and GAM, a near neutral effect in the RF and GBM, and a negative influence in the DL (see figure 4).

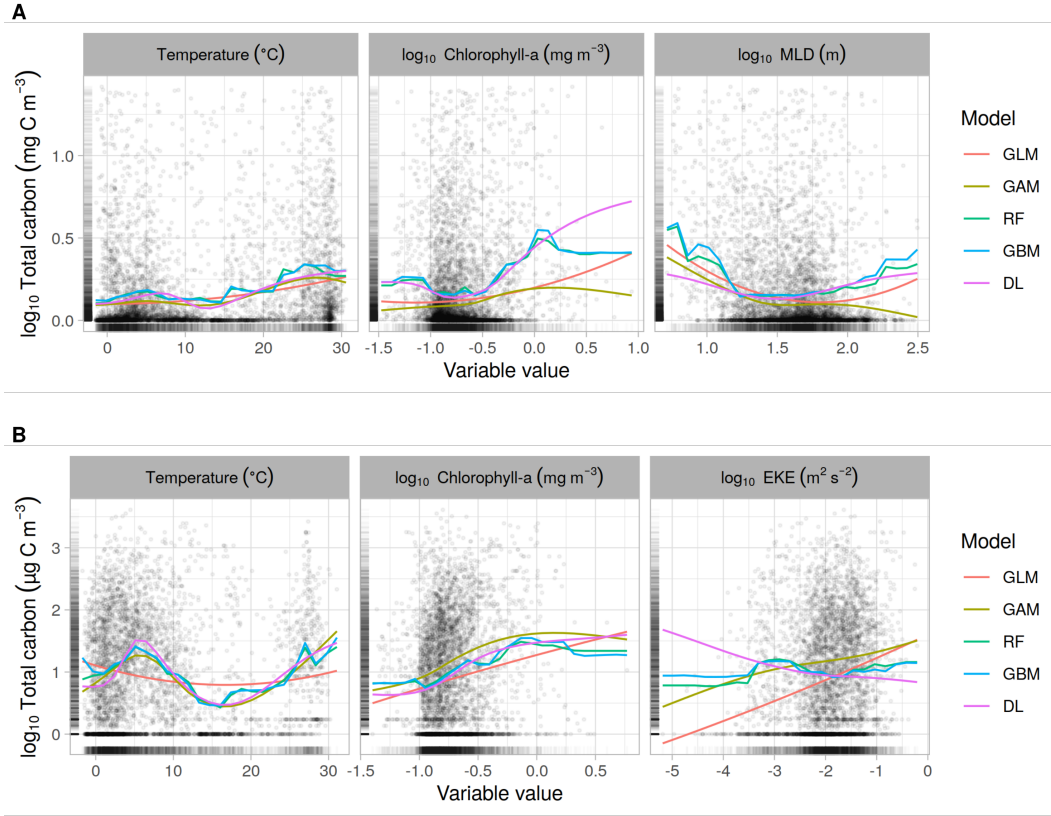


Figure 4. Partial dependence plots (PDP) for the environmental predictors in the pteropod (A) and foraminifer (B) models. The curves indicate the relations learned by the different SDMs and the rug on the x- and y-axis represents the distribution of the training data. MLD refers to the mixed layer depth, EKE to the eddy kinetic energy. The different model types are the Generalized Linear Model (GLM), Generalized Additive Model (GAM), Random Forest (RF), Boosted Regression Tree (GBM) and Neural Network (DL).

3.4 Global annual total inorganic carbon (TIC) export fluxes

Global mean annual biomass standing stocks are 52.2 Tg TC (ranging from 49.2 Tg TC to 57.3 Tg TC across SDM types) for pteropods and 0.9 Tg TC (0.6 Tg TC to 1.1 Tg TC) for foraminifers (table 3).

The corresponding global annual TIC fluxes were calculated based on growth rate parametrizations (section 2.3.2) and are on average 14.1 Tg TIC yr⁻¹ (13.3 Tg TIC yr⁻¹ to 15.5 Tg TIC yr⁻¹; table 3) for pteropods. Foraminifer TIC fluxes amount to on average 10.9 Tg TIC yr⁻¹ (8.5 Tg TIC yr⁻¹ to 14.3 Tg TIC yr⁻¹; table 3). The inter-SDMs range of the TIC fluxes increases by a factor of approximately 4–5 if the modelling uncertainty associated with the TIC-TC factor and the growth rate parametrization are included (table 3, see also section 3.5).

3.5 Uncertainty quantification

To assess the effects of SDM choice, growth rate parametrization and TIC-TC factor parametrization on the TIC flux predictions, we conducted a mANOVA and eval-

Table 3. Global mean annual surface pteropod and foraminifer total carbon (TC) biomass standing stocks and annual total inorganic carbon (TIC) flux estimates as calculated by the five species distribution model (SDMs).

Model	Pteropoda		Foraminifers	
	Standing stock (Tg TC)	Carbon flux (Tg TIC yr ⁻¹)	Standing stock (Tg TC)	Carbon flux (Tg TIC yr ⁻¹)
GLM	49.2	13.3 (8.5 - 19.1)	0.6	8.9 (4.0 - 16.4)
GAM	49.2	13.3 (8.5 - 19.1)	0.7	8.5 (3.0 - 19.0)
RF	57.3	15.5 (9.9 - 22.3)	1.0	14.2 (5.0 - 26.8)
GBM	56.9	15.4 (9.8 - 22.1)	1.1	13.2 (4.8 - 24.7)
DL	48.3	13.1 (8.4 - 18.9)	1.1	9.8 (3.7 - 20.3)
Average	52.2	14.1	0.9	10.9

uated spatial patterns of standard deviation between model predictions. The main sources of variability in global mean annual TIC fluxes differ between the plankton groups (figure S21). For pteropods, the growth rate and TIC-TC conversion factor choice are the major sources of uncertainty, as each explains 27% of the variability. SDM choice explains 10% of the variability in fluxes for pteropods. In contrast, the TIC flux variability for foraminifers is dominated by the parametrization of the foraminifer growth rate (71%), followed by the model choice (11%), and the TIC-TC factor (< 10%).

From a spatial point of view, relative inter-SDMs variability is highest in regions of low productivity and where environmental conditions are outside the range of the training dataset (figure 3, panels **C** and **D**). This encompasses the Southern Hemispheric (SH) oceanic gyres and the low-productivity latitudinal band around 45°S for pteropods and around 30°S for foraminifers. Absolute biomass predictions differ the most in regions of high biomass, i.e., mainly the North Atlantic for both groups (figure 3, panels **A** and **B**).

4 Discussion

4.1 Biogeographic biomass patterns

The biogeographic distribution patterns found for pteropods and foraminifers largely agree with previous findings (Lalli & Gilmer, 1989; Bednaršek, Mozina, et al., 2012; Buitenhuis et al., 2019; Lombard et al., 2011; Schiebel, 2002). We found high biomass concentrations for both plankton groups in the warm tropical waters, at the high northern latitudes and in the upwelling systems.

The global warm-water belt around the equator has previously been identified as a region of high biomass for pteropods (Lalli & Gilmer, 1989; Bednaršek, Mozina, et al., 2012; Burridge et al., 2017) and foraminifers (Schiebel & Movellan, 2012). High biomass concentrations in the equatorial region for the two plankton groups are representative of total global mesozooplankton distributions (Moriarty et al., 2013; Strömberg et al., 2009), which also show peaks in the tropical ocean.

Earlier studies also found the high latitudes to be regions of high biomass for both plankton groups (Lalli & Gilmer, 1989; Bednaršek, Mozina, et al., 2012; Schiebel & Movellan, 2012; Hunt et al., 2008). Contrary to previous studies (Bednaršek, Mozina, et al., 2012; Lalli & Gilmer, 1989; Hunt et al., 2008), the Southern Ocean was not identified as a region of major pteropod productivity in our study. This is likely due to the influ-

ence of the SO-CPR dataset, which included a high fraction (95.8%) of absences. Removing all CPR data from our training dataset (i.e., SO-CPR and Aus-CPR) significantly increases biomass concentrations for pteropods by a factor of 4–8 in the Southern Ocean but not in other basins (figures S24 and S25). However, removing the CPR data also leads to significantly less well constrained PDP curves for low temperatures, which increases the uncertainty of these CPR-depleted SDMs projections. As previous studies were based on much fewer and spatially confined observations in the Southern Ocean (e.g., 141 data points south of 60° S in Bednaršek, Mozina, et al. (2012)), it is possible that they constitute local upper bound estimates of pteropod abundance and biomass in the Southern Ocean.

Similar to our findings, upwelling regions have previously been found to be associated with high abundances of pteropods (Dadon & Masello, 1999; Koppelman et al., 2013; McGowan, 1967; Burridge et al., 2017) and foraminifers (Schiebel et al., 2004; Naidu & Malmgren, 1996; Ivanova et al., 1999). Upwelling systems are characterized by recurrent nutrient inputs that trigger high local primary productivity (Kämpf & Chapman, 2016), which produces optimal conditions for opportunistic foraminifer and pteropod species (Kucera, 2007; Schiebel & Hemleben, 2017). However, the upwelling systems are also associated with the upwelling of low pH waters (Joint et al., 2011; Hauri et al., 2013), and the shoaling of the calcite and aragonite saturation horizon (Leinweber & Gruber, 2013; Frenger et al., 2018). The effects of these changes in water chemistry in upwelling systems on pteropod and foraminifer abundances are discussed in section 4.2.

Overall, the modelled biogeographic patterns of pteropods and foraminifers are generally in line with earlier work (Bednaršek, Mozina, et al., 2012; Lalli & Gilmer, 1989; Schiebel & Movellan, 2012). Deviations, as the lower pteropod biomass in the Southern Ocean are probably caused by previously low sampling density and current biases in sampling methodology.

4.2 Environmental drivers

In agreement with other studies (Pinkerton et al., 2020; Beaugrand et al., 2013; Meiland et al., 2016; Jentzen et al., 2018), temperature was the strongest statistical covariate for the biomass distributions of pteropods and foraminifers in our study. This is not surprising since temperature influences all scales of biological processes, from intra-cellular reaction rates to species interactions (Chappon & Seuront, 2011; Kirby & Beaugrand, 2009; Schmidt-Nielsen, 1997; Brown et al., 2004). Temperature is also related to the water column stratification, which in turn can affect plankton biomass by influencing nutrient availability (see section 2.2.1) and primary productivity (Chiswell et al., 2014). The present global dependencies of biomass on temperature can differ from the results of local studies (e.g., a negative dependency of pteropod biomass on temperature as found in Bednaršek et al. (2022)). As the geographic scale of the analysis is different, distinct effects are captured by the models, such as for example large-scale latitudinal effects in contrast to local upwelling influences.

The modelled bimodal structure of the biomass dependency of our SDMs on temperature for foraminifers, and—to a lesser extent—pteropods (peaks around 5–7°C and above 25°C, figure 4), likely reflects the existence of distinct assemblages of warm-water species and cold-water species within these groups (Bradshaw, 1959). Some foraminifer species are associated with one end of the temperature spectrum - for instance, *Neoglobobulimina pachyderma* is associated with temperatures below 10°C and *Globigerinoides ruber (white)* with temperatures above 18°C (Kucera, 2007; G. A. Schmidt & Mulitza, 2002), which broadly matches our identified peaks (Morard et al., 2015; Antell et al., 2021; Rillo et al., 2022). Furthermore, the temperature interval around 17°C constitutes a minimum in the foraminifer biomass dependency curve and it can be associated with the subtropical front (D. N. Schmidt et al., 2004). As very dynamic dispersal barriers, fronts

are regions of significant environmental variability where foraminifer body sizes were found to be significantly smaller (D. N. Schmidt et al., 2004). This could help explain why the subtropical front was associated with lower foraminifer biomass. However, most foraminifer species display wide thermal tolerances of around 10°C (Schiebel & Hemleben, 2017). Additionally, the sampling density was high in the cold regions of the Southern Ocean and at high temperatures in the tropics, but few data points (18.0% for pteropods and 23.1% for foraminifers) stem from the intermediate temperature range between 10°C and 20°C for both plankton groups. Hence, the bimodal structure might also be skewed due to an uneven sampling distribution (as seen in the uneven density of the x-axis rug plot in figure 4A).

In our SDMs, surface chlorophyll-a concentration emerged as the second-most important environmental covariate for the biomass of both plankton groups, which is also supported by the literature (Pinkerton et al., 2020; Schiebel et al., 2001; Meilland et al., 2016; Schiebel et al., 1995). Generally, a positive near-linear relationship between chlorophyll-a concentrations and pteropod and foraminifer biomasses is observed, particularly in the well-constrained range of the PDP curve (figure 4). As a measure of food availability, chlorophyll-a can be directly positively linked to zooplankton abundances and biomass (Schiebel et al., 2001; Pinkerton et al., 2020; Strömberg et al., 2009). However, both pteropods and foraminifers also feed on non-phytoplankton prey and organic particles to varying degrees (Lalli & Gilmer, 1989; Rhumbler, 1911; Caron & Bé, 1984; Spindler et al., 1984). Pteropods typically feed on particles that are one 100–1000th of their own size (Conley et al., 2018), while some foraminifers can digest prey larger than themselves (Schiebel & Hemleben, 2017). This can explain the smaller-scale deviations of the PDP curves from the near-linear trend and a certain decoupling at low chlorophyll-a concentrations (figure 4) as the zooplankton can feed on alternative organic particles.

The overall importance of the environmental variables driving biomass in models may vary with the spatio-temporal scale at which the analysis is conducted (Corney et al., 2006). Both MLD and EKE were found to be of minor importance as driving variables in our SDMs, which might be due to their predominantly mesoscale effect on mixing and food availability. MLD negatively influences pteropod biomass concentrations over most of the assessed range (figure 4). As flux-feeders, pteropods rely on a steady downward flux of particles, which can be hindered by a deep and turbulent water column mixing (Tsurumi et al., 2005). From a viewpoint of ecological successions over seasons, the shoaling of the deep winter mixed layer in spring is one of the main factors triggering spring phytoplankton blooms (Chiswell et al., 2014). Following these blooms, zooplankton productivity increases to feed on the remaining phytoplankton (Romagnan et al., 2015). This might explain the increase in pteropod biomass for shallow MLD values as an indirect consequence. EKE shows a slight positive impact on foraminifer biomass in the simpler models (figure 4). At the mesoscale, eddies can sustain increases in foraminifer biomass, as they can drive the mixing of the deep chlorophyll-a maximum into shallower surface layers, i.e., into the habitat of foraminifers (Turner, 2015; Kupferman et al., 1986; Beckmann et al., 1987; Fallet et al., 2011; Steinhardt et al., 2014; Schiebel et al., 1995). However, the effect of eddies varies as their direction of rotation determines the dominant vertical direction of water movement (Dufois et al., 2016). The direct large-scale effects of MLD and EKE on biomass patterns are not frequently assessed in the literature (exceptions for MLD are Pinkerton et al. (2020) and Schiebel et al. (2001)). On a local, short-term scale, however, they might have a strong influence on zooplankton biomass that cannot be captured by our global-scale monthly model.

Previous work identified carbonate chemistry as an important predictor for net calcification on a local scale (Bednaršek & Ohman, 2015; Manno et al., 2017; Lischka et al., 2011; Bednaršek et al., 2022; Mekkes, Renema, et al., 2021). CO₂ - rich waters characterized by low pH, low calcite, and low aragonite saturation states may negatively affect certain calcifying organisms by increasing their dissolution and lowering their cal-

cification rate (Bednaršek, Feely, et al., 2017; Bednaršek et al., 2022; Mekkes, Renema, et al., 2021; Mekkes, Sepúlveda-Rodríguez, et al., 2021). For pteropods, these changes in water chemistry can reduce their metabolic activity, increase shell dissolution, and decrease their growth and survival (Lischka et al., 2011; Lischka & Riebesell, 2012; Maas et al., 2015; Gardner et al., 2017; Manno et al., 2007; Bednaršek et al., 2016; Bednaršek, Klinger, et al., 2017; Bednaršek et al., 2022; Bednaršek, Feely, et al., 2017). Foraminifers are less sensitive to changes in saturation states as their shells are made of calcite (Orr et al., 2005; Weinkauff et al., 2016), but the specific sensitivities are not yet well quantified (Fabry et al., 2008), and calcite saturation has not yet decreased as much as aragonite saturation. As our models did not include any indicator of pteropod physiology or biominerology and were based on climatological environmental conditions, we could not account for these effects in the way that for example individual-based models do (Hofmann Elizondo & Vogt, 2022). So far, field studies have not found large-scale abundance decreases of either pteropods or foraminifers as a result of a changes in the carbonate chemistry (Ohman et al., 2009; Howes et al., 2015; Thibodeau et al., 2019), as the effects of other environmental variables such as temperature tend to prevail (Beare et al., 2013). However, the fitness reduction of individual organisms leads to delayed responses on the population level, such that large-scale changes may happen only in the near future under climate change (Bednaršek et al., 2022).

Overall, the relative importance and response curves shapes (figures 4, S18) of the various environmental predictors are in line with our current state of knowledge. Some response curves are affected by uneven sampling across environmental and geographic space and scale dependencies, but within the most commonly observed ranges of environmental conditions, the response curves of the five SDMs agree well with each other for both plankton groups.

4.3 Current global surface ocean biomass and TIC export fluxes

Estimates of global plankton standing stocks from observations have only become possible during the past decade (Buitenhuis, Vogt, et al., 2013) due to paucity in the available information about marine ecosystems. Hence, there are still large uncertainties, particularly for organisms such as zooplankton with patchy abundance patterns (Buitenhuis, Hashioka, & Quéré, 2013) and strongly uneven sampling distributions and methodologies (see also section 4.4 and figure S5 for an assessment of data patchiness). Estimates of standing stocks are highly uncertain, though less so than in marine systems than terrestrial ones (Bar-On et al., 2018; de Garidel-Thoron et al., 2022). In this context, we deem the partly large deviations of our estimates from previous studies as plausible.

On a global mean annual scale, our estimates of total plankton biomass standing stocks are a factor of 10 lower than previous MAREDAT observation-based estimates for pteropods (Bednaršek, Mozina, et al., 2012) and in the same range for foraminifers (Schiebel & Movellan, 2012) as shown in table 4. For both plankton groups, the previous standing stock estimates were 1) calculated using globally averaged, unweighted biomass concentrations, 2) based on a spatiotemporal subset of our current observational dataset, and 3) based only on non-zero abundance observations.

The discrepancy between our results and those of previous studies decreases when we calculate global standing stocks based on these different configurations (see table 4). Calculating standing stocks based on 1) MAREDAT methodology does not change the standing stock estimates strongly (46 TgTC to 57 TgTC for pteropoda and 0.5 TgTC to 1.1 TgTC for foraminifers). Additionally 2) subsetting our prediction fields at the original MAREDAT sampling points increases pteropod standing stock estimates by approximately 50% to 62 TgTC to 95 TgTC, while foraminifer estimates remain near constant at 0.6 TgTC to 1.9 TgTC. Finally, 3) excluding zero abundance observations before modelling increases standing stock estimates to 91 TgTC to 140 TgTC for pteropods and 2 TgTC

Table 4. Comparison between modeled total carbon (TC) standing stocks and total inorganic carbon (TIC) fluxes for pteropods and foraminifers with previous studies. All values were converted to represent TC and TIC, respectively. The results of the mechanistic studies from Gangstø et al. (2008) and Buitenhuis et al. (2019) denote the reported CaCO_3 production and not the export flux. The export flux calculations include dissolution of the sinking calcium carbonate shells. However, we do not take this into account in the current study. Thus, we compare the production terms before dissolution. The sensitivity analyses are shown in italics. For the comparisons to MAREDAT, the projected biomass maps were sampled at the MAREDAT observation points of the respective plankton group (Bednaršek, Mozina, et al., 2012; Schiebel & Movellan, 2012). To be consistent with the methodology used in Bednaršek, Mozina, et al. (2012) and Schiebel and Movellan (2012), the total standing stocks and fluxes were calculated from global non-weighted mean biomass concentrations and assuming one and nine complete overturn periods for pteropods and foraminifers, respectively.

Source	Pteropoda		Foraminifers	
	Standing stock (Tg TC)	Carbon flux (Tg TIC yr ⁻¹)	Standing stock (Tg TC)	Carbon flux (Tg TIC yr ⁻¹)
<i>Estimates based on mechanistic modelling studies</i>				
Buitenhuis et al. (2019)		152 - 4183 ^a		100 - 141 ^a
Gangstø et al. (2008)		300		
<i>Estimates based on observational data</i>				
Bednaršek, Tarling, et al. (2012)	444 - 505 ^{b,c}	112 - 150 ^{b,c}		
Schiebel and Movellan (2012)			1 - 5 ^b	3 - 12 ^b
Schiebel (2002)				157 - 389 ^{b,d}
Our results	49 - 57	8 - 22	1 - 2	3 - 35
<i>1: MAREDAT methodology</i>	<i>46 - 57</i>	<i>12 - 15</i>	<i>0.5 - 1.1</i>	<i>5 - 10</i>
<i>2: Sampled at MAREDAT points & methodology</i>	<i>62 - 95</i>	<i>17 - 26</i>	<i>1 - 2</i>	<i>1 - 4</i>
<i>3a: W/o zeros</i>	<i>91 - 140</i>	<i>24 - 38</i>	<i>2 - 3</i>	<i>16 - 33</i>
<i>3b: W/o zeros, MAREDAT points & methodology</i>	<i>132 - 220</i>	<i>35 - 60</i>	<i>2 - 9</i>	<i>4 - 20</i>
<i>W/o CPR data</i>	<i>90 - 155</i>	<i>25 - 42</i>	<i>1 - 3</i>	<i>18 - 33</i>

^aBased on calcite production, not flux. ^bBased on subset of observations used in this study; ^cEstimates based on non-zero observations only.

^dFlux at 100 m

to 3 TgTC for foraminifers. Combining all three modifications causes an increase of factor 3–4 for both plankton types. Following the same methodology, pteropod biomass estimates are still a factor of 2–4 lower than the MAREDAT estimates, while foraminifer biomass estimates are in the same range. A potential reason for this difference between the two plankton groups might be the variation in patchiness due to the larger body size of pteropoda. The original MAREDAT pteropod abundance observations are nearly four times as patchy as those of foraminifers (Buitenhuis, Vogt, et al., 2013), which could have led to a higher bias in the pteropod standing stock estimate.

In the context of the marine trophic foodweb, pteropods constitute approximately 6% to 8% of total macrozooplankton biomass, whereas foraminifers make up 0.1% to 0.6% of microzooplankton biomass as shown in figure 5 (Buitenhuis, Vogt, et al., 2013). Each plankton size class encompasses a broad range of taxonomic groups, so that the relatively small contributions of pteropods and foraminifers is logical. In contrast to the other PFTs estimates and the earlier MAREDAT estimates for pteropods and foraminifers, our results are based on global climatological biomass estimates instead of spatially discrete observation data. This causes a lower discrepancy between our mean and median estimates as well as a lower total standard deviation (figure 5), because high biomass extreme events are not as prevalent in our results as in the raw field observations (cf. also section 4.4 and figure S5).

Estimated pteropod TIC fluxes are a factor of 5–100 lower than in previous numerical modelling studies (table 4). The estimates by Buitenhuis et al. (2019) and Gangstø et al. (2008) are based on mechanistic models which used published laboratory evidence for model calibration and observational data from MAREDAT for model evaluation. How-

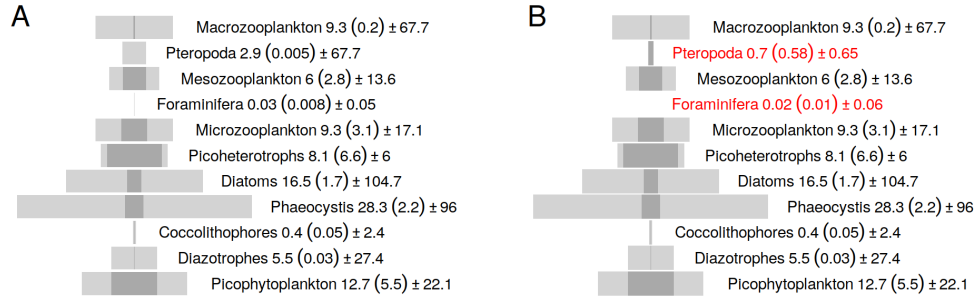


Figure 5. Trophic pyramid of autotrophic and heterotrophic plankton functional types (PFT). The bars show the mean (light grey filling) and median (dark grey filling, value in parentheses) biomass concentrations in $\mu\text{g TCL}^{-1}$ in the surface 200 m. The standard deviation is denoted for each PFT. **A** shows the original results from the MAREDAT project presented in Buitenhuis, Vogt, et al. (2013). **B** shows our updated estimates for pteropods and foraminifers as highlighted in red.

ever, the parametrization of the growth rate is based on copepod observations instead of pteropods in Buitenhuis, Vogt, et al. (2013) and hence could have introduced a bias. The difference in depth at which TIC-fluxes are reported (100 m in Buitenhuis et al. (2019) and 200 m in our study) likely introduce further uncertainties, however as neither of the results include dissolution effects, these are deemed minor. An additional reason for the discrepancy could be an incomplete representation of the true abundances in our observation data due to sampling biases (cf. section 4.4). Examples of such biases include net avoidance, diel vertical migration (DVM), and the use of sub-optimal mesh sizes for the target group (e.g. in the CPR), which can lead to underestimated abundances in our observational data (Zamelczyk et al., 2021; Pinkerton et al., 2020; Doubek et al., 2020). Excluding CPR data from our models approximately doubles the estimated TIC fluxes (table 4), however it also increases uncertainty in the environmental driver dependencies (cf. section 4.1).

Our foraminifer TIC flux estimates are of the same order of magnitude as the most recent observation-based estimates and mechanistic model-based studies, albeit on the lower end for the latter (Table 4). The earlier observational study by Schiebel (2002) is based on much smaller datasets with a spatial bias towards the highly productive North Atlantic Ocean and found substantially higher TIC fluxes. However, our results align well with the flux estimate calculated by more recent studies, such as Schiebel and Movellan (2012). In our work, we account only for large adults due to mesh size limitations, but including juvenile biomass might double foraminifer biomass and flux estimates (Schiebel and Movellan (2012), see also section 4.4). This uncertainty could also explain the deviations of our results from the mechanistic model-based estimate by Buitenhuis et al. (2019), which is a factor of 1.5–50 higher than our global annual TIC flux estimate. Similar as for pteropods, excluding CPR data prior to modelling approximately doubles the estimated global annual TIC fluxes (cf. table 4), which might be indicative of non-optimal representation of foraminifer abundances in this dataset.

Pteropoda contribute 0.5%–2.2% to total annual global carbonate fluxes and foraminifers contribute 0.2%–3.5%, assuming annual global fluxes amount to $1.0 \text{ Pg TIC yr}^{-1}$ to $1.6 \text{ Pg TIC yr}^{-1}$ (Iglesias-Rodriguez et al., 2002; Lee, 2001; Berelson et al., 2007). We can assume that the carbon fluxes calculated in our study represent a lower bound estimate due to biases and incompleteness of the observation dataset (see section 4.4). Coccolithophores are estimated to contribute 26%–52% to global carbonate fluxes (C. J. O’Brien, 2015),

which still leaves 40%–70% of global carbonate fluxes unaccounted for and points to an underestimation of the contribution from the calcifying zooplankton. Additional minor contributors to the marine CaCO_3 budget are fishes, atlantid heteropods, pseudothecosomes (particularly the fully shelled *Peracle* species), calcifying ostracods, dinoflagellates, ciliates and the larvae of both benthic molluscs and gymnosomates (Buitenhuis et al., 2019). However, their contribution to global carbonate fluxes is not well constrained, but may range between 3%–15% per group (Wilson et al., 2009; Schiebel, 2002; Buitenhuis et al., 2019), and hence warrants further investigation.

4.4 Limitations and uncertainties

Here, we use large global datasets and an exhaustive model ensemble approach to estimate pteropod and foraminifer biomass. We quantify and discuss the uncertainty arising from the model choice and key parametrizations and estimates for the growth rate and the TIC-TC factor. However, our biomass and carbon flux estimates are affected by the characteristics and errors underlying the observational data and the simplifying assumptions made for the model setup. These include the interaction of spatio-temporal biases in sampling effort with the inherent patchiness of plankton distribution, variations in sampling net mesh sizes, and limited taxonomic resolution for biomass conversions (de Garidel-Thoron et al., 2022).

Patchy sampling across space and time leads to spatiotemporal biases in the training dataset (figures 2, S3 and S4). Data coverage is low in the low productivity oligotrophic gyres and during the less productive months (figure S4). We find that a large fraction of the inter-model variability is due to environmental conditions outside of or at the outer ranges of the training data (section 3.5 and figures 3 and 4). Nonetheless, a large fraction of the global environmental space of our predictor variables is covered by the abundance datasets, which allows us to predict biomass values with higher certainty (figure S6). Furthermore, plankton distributions are generally characterized by a high level of seasonal and spatial patchiness (figure S5, Boltovskoy (1971); Beckmann et al. (1987); Siccha et al. (2012); Buitenhuis, Vogt, et al. (2013)). This introduces high variance in the observed abundances (figure S5) and a mismatch between the gridded monthly climatologies used as environmental predictors and the mesoscale-affected biomass patterns (Righetti et al., 2019; Benedetti et al., 2021). However, previous studies found no significant benefit of using highly temporally resolved data over climatologies (Pinkerton et al., 2020), as the environmental conditions an organism experiences are based on their Lagrangian movement over time (Hofmann Elizondo & Vogt, 2022). Finally, the use of coarse mesh sizes for sampling relatively small zooplankton can underestimate the true abundances as small and/or mobile individuals are missed (Tseng et al., 2011; Wells, 1973; Miloslavić et al., 2014; Mack et al., 2012; Skjoldal et al., 2013; Fabry, 1989; Zamelczyk et al., 2021). This is particularly relevant for the SO-CPR and Aus-CPR observations which make up 91% and 73% of our training data for pteropods and foraminifers, respectively (section 2.1.1) due to the large mesh size of 270 μm used (Richardson et al., 2006). These sampling data constraints hence cause our biomass and flux estimates to be lower end estimates.

Further uncertainties in the standing stock and flux estimates come from the simplified abundance to biomass conversions and the biomass to carbon flux derivation. We assumed species-level or group-level averages for the size-based biomass conversion functions (section 2.1.2). Yet, in practice these values vary based on ontogenetic stage, subspecies, ambient temperature (Bradshaw, 1959), and food availability (Meilland et al., 2016; Schiebel et al., 2001; Schiebel & Hemleben, 2005). These factors vary with latitude and we could not account for them explicitly in the present carbon conversions (cf. section 2.1.2) due to a lack of available parametrizations. Therefore, we likely underestimated the global latitudinal variability in our biomass predictions. To convert biomasses to TIC fluxes, growth rates and the TIC-TC conversion factor were based on spatially

constrained data and a limited number of species due to data availability (cf. section 2.1.2). The choices made for the growth rate function and the TIC-TC factor had a significant impact on flux estimates for pteropods, and for foraminifers to a lesser extent (section 3.5). Plankton observations identified at a finer taxonomic level and species-specific laboratory-based conversion factors and growth rates would likely increase the accuracy of our calculations. To estimate export fluxes at depth, particle sinking velocities and dissolution rates need to be considered (Takahashi & Bé, 1984; Schiebel et al., 2007). During periods of peak biomass production, high pulses of fast-sinking organisms occur and can drive higher export efficiency (Schiebel, 2002). However, the relative species abundances observed in our upper ocean foraminifera data (figure S26) are in good agreement with those found in sediment trap data in previous studies Kretschmer et al. (2018); Lombard et al. (2011). This shows that the foraminifer surface export fluxes and patterns found in our study are representative of export patterns found in the deeper ocean. For pteropods, to our knowledge, no comprehensive global sediment trap data analysis has yet been conducted. To assess comparability between fluxes at the surface and the deep ocean, such an analysis is hence much needed.

5 Conclusion

The aim of this study was to predict global monthly and annual patterns and drivers of shelled pteropod and planktic foraminifer TC biomass distributions, and their associated TIC fluxes, and to assess the importance of these groups for the global biogeochemical cycling of carbon and CaCO_3 .

Globally, pteropods contribute 6%–8% and foraminifers 0.1%–0.6% to total global macrozooplankton and microzooplankton TC standing stocks, respectively. The sinking of their shells and tests constitutes approximately 1.5% each of the total global annual surface TIC fluxes. We found biomass hotspots for both plankton groups in the high Northern latitudes, around the equator, and in the upwelling systems. Temperature and chlorophyll-a concentrations were the two most important environmental covariates for modelling the biomass patterns.

Based on newly assembled abundance data for different organism groups, we can use our modelling pipeline to project global biomass patterns for various plankton functional groups (Le Quéré et al., 2005). Thus, we can validate newly developed mechanistic marine ecosystem models (Le Quéré et al., 2016; Clerc et al., 2022) of increased complexity and higher diversity in zooplankton functional types. Additionally, the models can be employed to assess future changes in plankton biomass by projecting the present models on future environmental fields under climate change scenarios (Benedetti et al., 2021; Tittensor et al., 2021). This is particularly relevant considering the high sensitivity of, for example, pteropods to ocean acidification and warming (Bednaršek et al., 2016; Manno et al., 2016). Thus, we can identify hotspots of future biomass changes (complementary to future changes in diversity as modeled in Benedetti et al. (2021)) and potentially link these to risk assessments based on other ocean health indices (Halpern et al., 2012).

Furthermore, the pipeline can be used to model other types of quantitative data, such as sediment trap data (Kretschmer et al., 2018) or measurements based on novel approaches like underwater imaging techniques or omics (Pesant et al., 2015). As a large fraction of the carbonate export fluxes still remains unaccounted for based on our results, we could calculate flux contributions of different organism groups such as fish and shelled heteropods (Wilson et al., 2009; Buitenhuis et al., 2019; Wall-Palmer et al., 2016). Comparing estimates based on upper ocean data with those based on sediment traps could help to improve our understanding of export patterns driven by different organism groups as well as the impacts of carbonate dissolution and sinking rates on such patterns.

Open Research Section

The observational datasets used to train the models as well as the model outputs will be made publicly available on AtlantECO's GeoNode portal (<https://atlanteco-geonode.eu/>) upon acceptance of the manuscript. An adapted version of the modelling pipeline, applicable to any species abundance or biomass dataset in the AtlantECO format is available on the GitHub account of N.K. (<https://github.com/nielja>).

Acknowledgments

We thank all contributors involved in the zooplankton field sampling and identification and we acknowledge the efforts made by the community to share such data through publicly available archives. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 862923. This output reflects only the author's view, and the European Union cannot be held responsible for any use that may be made of the information contained therein. We thank David Johns for preparing and sharing the calcifying pteropod counts of the CPR survey. We are grateful to Anthony J. Richardson and Claire Davies for facilitating the access to the Aus-CPR survey. The Aus-CPR data was sourced from Australia's Integrated Marine Observing System (IMOS) – IMOS is enabled by the National Collaborative Research Infrastructure Strategy (NCRIS). We are also very grateful to John Kitchener for facilitating the access to the SO-CPR survey. Nina Bednaršek acknowledges support from the Slovene Research Agency (ARRS "Biomarkers of subcellular stress in the Northern Adriatic under global environmental change", #J12468).

References

- Anderson, O. R., Spindler, M., Bé, A. W. H., & Hemleben, C. (1979). Trophic activity of planktonic foraminifera. *Journal of the Marine Biological Association of the United Kingdom*, 59(3), 791–799. Retrieved from <https://www.cambridge.org/core/article/trophic-activity-of-planktonic-foraminifera/F9AC18BAA017EC7EAE64CB0D6E3A805C> doi: DOI:10.1017/S002531540004577X
- Anglada-Ortiz, G., Zamelczyk, K., Meilland, J., Ziveri, P., Chierici, M., Fransson, A., & Rasmussen, T. L. (2021). Planktic Foraminiferal and Pteropod Contributions to Carbon Dynamics in the Arctic Ocean (North Svalbard Margin). *Frontiers in Marine Science*, 8(June). doi: 10.3389/fmars.2021.661158
- Antell, G. S., Fenton, I. S., Valdes, P. J., & Saupe, E. E. (2021). Thermal niches of planktonic foraminifera are static throughout glacial-interglacial climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 118(18), 1–9. doi: 10.1073/pnas.2017105118
- Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, 115(25), 6506–6511. doi: 10.1073/pnas.1711842115
- Barton, A. D., Irwin, A. J., Finkel, Z. V., & Stock, C. A. (2016). Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proceedings of the National Academy of Sciences of the United States of America*, 113(11), 2964–2969. doi: 10.1073/pnas.1519080113
- Bé, A. W. H., & Gilmer, R. W. (1977). A zoogeographic and taxonomic review of euthecosomatous Pteropoda. *Oceanic micropaleontology*, 1(6), 733–808.
- Beare, D., McQuatters-Gollop, A., van der Hammen, T., Machiels, M., Teoh, S. J., & Hall-Spencer, J. M. (2013, 5). Long-Term Trends in Calcifying Plankton and pH in the North Sea. *PLOS ONE*, 8(5), e61175. Retrieved from <https://doi.org/10.1371/journal.pone.0061175>
- Beaugrand, G., Edwards, M., & Legendre, L. (2010). Marine biodiversity, ecosys-

- tem functioning, and carbon cycles. *Proceedings of the National Academy of Sciences of the United States of America*, 107(22), 10120–10124. doi: 10.1073/pnas.0913855107
- Beaugrand, G., Mcquatters-Gollop, A., Edwards, M., & Goberville, E. (2013). Long-term responses of North Atlantic calcifying plankton to climate change. *Nature Climate Change*, 3(3), 263–267. doi: 10.1038/nclimate1753
- Beckmann, W., Auras, A., & Hemleben, C. (1987). Cyclonic cold-core eddy in the eastern North Atlantic. III. Zooplankton. *Marine Ecology Progress Series*, 165–173.
- Bednaršek, N., Carter, B. R., McCabe, R. M., Feely, R. A., & Howard, E. (2022). Pelagic calcifiers face increased mortality and habitat loss with warming and ocean acidification. *Authorea*. doi: DOI:10.22541/au.164865092.29568156/v1
- Bednaršek, N., Feely, R. A., Beck, M. W., Glippa, O., Kanerva, M., & Engström-Öst, J. (2018). El Niño-related thermal stress coupled with upwelling-related ocean acidification negatively impacts cellular to population-level responses in pteropods along the California current system with implications for increased bioenergetic costs. *Frontiers in Marine Science*, 5(DEC), 1–17. doi: 10.3389/fmars.2018.00486
- Bednaršek, N., Feely, R. A., Tolimieri, N., Hermann, A. J., Siedlecki, S. A., Waldbusser, G. G., ... Pörtner, H. O. (2017). Exposure history determines pteropod vulnerability to ocean acidification along the US West Coast article. *Scientific Reports*, 7(1), 1–12. doi: 10.1038/s41598-017-03934-z
- Bednaršek, N., Harvey, C. J., Kaplan, I. C., Feely, R. A., & Možina, J. (2016). Pteropods on the edge: Cumulative effects of ocean acidification, warming, and deoxygenation. *Progress in Oceanography*, 145, 1–24. doi: 10.1016/j.pocean.2016.04.002
- Bednaršek, N., Klinger, T., Harvey, C. J., Weisberg, S., McCabe, R. M., Feely, R. A., ... Tolimieri, N. (2017). New ocean, new needs: Application of pteropod shell dissolution as a biological indicator for marine resource management. *Ecological Indicators*, 76, 240–244. doi: 10.1016/j.ecolind.2017.01.025
- Bednaršek, N., Možina, J., Vogt, M., O'Brien, C., & Tarling, G. A. (2012). The global distribution of pteropods and their contribution to carbonate and carbon biomass in the modern ocean. *Earth System Science Data*, 4(1), 167–186. doi: 10.5194/essd-4-167-2012
- Bednaršek, N., & Ohman, M. D. (2015). Changes in pteropod distributions and shell dissolution across a frontal system in the California Current System. *Marine Ecology Progress Series*, 523, 93–103. doi: 10.3354/meps11199
- Bednaršek, N., Tarling, G. A., Fielding, S., & Bakker, D. C. (2012). Population dynamics and biogeochemical significance of *Limacina helicina antarctica* in the Scotia Sea (Southern Ocean). *Deep-Sea Research Part II: Topical Studies in Oceanography*, 59-60, 105–116. doi: 10.1016/j.dsr2.2011.08.003
- Bell, D. M., & Schlaepfer, D. R. (2016). On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological Modelling*, 330, 50–59. Retrieved from <http://dx.doi.org/10.1016/j.ecolmodel.2016.03.012> doi: 10.1016/j.ecolmodel.2016.03.012
- Benedetti, F., Vogt, M., Elizondo, U. H., Righetti, D., Zimmermann, N. E., & Gruber, N. (2021). Major restructuring of marine plankton assemblages under global warming. *Nature Communications*, 12(1), 1–15. Retrieved from <http://dx.doi.org/10.1038/s41467-021-25385-x> doi: 10.1038/s41467-021-25385-x
- Berelson, W. M., Balch, W. M., Najjar, R., Feely, R. A., Sabine, C., & Lee, K. (2007). Relating estimates of CaCO₃ production, export, and dissolution in the water column to measurements of CaCO₃ rain into sediment traps and dissolution on the sea floor: A revised global carbonate budget. *Global Biogeochemical Cycles*, 21(1).

- Boehmke, B., & Greenwell, B. (2019a). Deep Learning. In *Hands-on machine learning with r* (1st ed., chap. 12). New York: Chapman and Hall/CRC. doi: <https://doi.org/10.1201/9780367816377>
- Boehmke, B., & Greenwell, B. (2019b). Gradient Boosting. In *Hands-on machine learning with r* (1st ed., chap. 11). New York: Chapman and Hall/CRC. doi: <https://doi.org/10.1201/9780367816377>
- Boehmke, B., & Greenwell, B. (2019c). Interpretable Machine Learning. In *Hands-on machine learning with r* (chap. 16). New York: Chapman and Hall/CRC. doi: <https://doi.org/10.1201/9780367816377>
- Boehmke, B., & Greenwell, B. (2019d). Modeling Process. In *Hands-on machine learning with r2* (1st ed., chap. 2). New York: Chapman and Hall/CRC.
- Boehmke, B., & Greenwell, B. (2019e). Random Forests. In *Hands-on machine learning with r* (1st ed., chap. 11). New York: Chapman and Hall/CRC. doi: <https://doi.org/10.1201/9780367816377>
- Boltovskoy, E. (1971). Patchiness in the distribution of planktonic foraminifera. In *Proceedings of the 2. planktonic conference* (pp. 107–115). Rome, Italy: Ed. Tecnoscienza.
- Boyce, D. G., Lewis, M. R., & Worm, B. (2010). Global phytoplankton decline over the past century. *Nature*, 466(7306), 591–596. doi: 10.1038/nature09268
- Boyer, T. P., García, H. E., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Reagan, J. R., ... Smolyar, I. V. (2018). *World Ocean Atlas 2018*. Retrieved from <https://www.ncei.noaa.gov/archive/accession/NCEI-WOA18>
- Bradshaw, J. S. (1959). Ecology of living planktonic foraminifera in the north and equatorial Pacific. *Ocean. Cushman Found. Foram. Res. Contr.*, 10(2), 25–64.
- Brandão, M. C., Benedetti, F., Martini, S., Soviadan, Y. D., Irisson, J. O., Romagnan, J. B., ... Lombard, F. (2021). Macroscale patterns of oceanic zooplankton composition and size structure. *Scientific Reports*, 11(1), 1–19. doi: 10.1038/s41598-021-94615-5
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, 85(7), 1771–1789.
- Brun, P., Kiørboe, T., Licandro, P., & Payne, M. R. (2016). The predictive skill of species distribution models for plankton in a changing climate. *Global change biology*, 22(9), 3170–3181. doi: 10.1111/gcb.13274
- Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R. O., Wang, Z., & Zimmermann, N. E. (2020). Model complexity affects species distribution projections under climate change. *Journal of Biogeography*, 47(1), 130–142. doi: 10.1111/jbi.13734
- Brun, P., Vogt, M., Payne, M. R., Gruber, N., O'Brien, C. J., Buitenhuis, E. T., ... Luo, Y. W. (2015). Ecological niches of open ocean phytoplankton taxa. *Limnology and Oceanography*, 60(3), 1020–1038. doi: 10.1002/lno.10074
- Buitenhuis, E. T., Hashioka, T., & Quéré, C. L. (2013). Combined constraints on global ocean primary production using observations and models. *Global Biogeochemical Cycles*, 27(3), 847–858. doi: 10.1002/gbc.20074
- Buitenhuis, E. T., Le Quéré, C., Bednaršek, N., & Schiebel, R. (2019). Large Contribution of Pteropods to Shallow CaCO₃ Export. *Global Biogeochemical Cycles*, 33(3), 458–468. doi: 10.1029/2018GB006110
- Buitenhuis, E. T., Vogt, M., Moriarty, R., Bednaršek, N., Doney, S. C., Leblanc, K., ... Swan, C. (2013). MAREDAT: Towards a world atlas of MA-Rine Ecosystem DATA. *Earth System Science Data*, 5(2), 227–239. doi: 10.5194/essd-5-227-2013
- Burba, G., & Anderson, D. (2005). *A Brief Practical Guide to Eddy Covariance Flux Measurements* (Tech. Rep.). Lincoln: LI-COR Biosciences. doi: 10.1076/ceyr.18.1.62.5393
- Burridge, A. K., Goetze, E., Wall-Palmer, D., Le Double, S. L., Huisman, J., & Peijnenburg, K. T. (2017). Diversity and abundance of pteropods and het-

- eropods along a latitudinal gradient across the Atlantic Ocean. *Progress in Oceanography*, 158, 213–223. Retrieved from <http://dx.doi.org/10.1016/j.pocean.2016.10.001> doi: 10.1016/j.pocean.2016.10.001
- Caron, D. A., & Bé, A. W. H. (1984). Predicted and observed feeding rates of the spinose planktonic foraminifer *Globigerinoides sacculifer*. *Bulletin of Marine Science*, 35(1), 1–10.
- Carton, J. A., Chepurin, G. A., & Chen, L. (2018). SODA3: A new ocean climate reanalysis. *Journal of Climate*, 31(17), 6967–6983. doi: 10.1175/jcli-d-18-0149.1
- Chapperon, C., & Seuront, L. (2011). Behavioral thermoregulation in a tropical gastropod: links to climate change scenarios. *Global Change Biology*, 17(4), 1740–1749.
- Chiswell, S. M., Calil, P. H., & Boyd, P. W. (2014). Spring blooms and annual cycles of phytoplankton: A unified perspective. *Journal of Plankton Research*, 37(3), 500–508. doi: 10.1093/plankt/fbv021
- Clerc, C., Bopp, L., Benedetti, F., Vogt, M., & Aumont, O. (2022). Including filter-feeding gelatinous macrozooplankton in a global marine biogeochemical model: model-data comparison and impact on the ocean carbon cycle. *Authorea Preprints*.
- Conley, K. R., Lombard, F., & Sutherland, K. R. (2018). Mammoth grazers on the ocean’s minuteness: A review of selective feeding using mucous meshes. *Proceedings of the Royal Society B: Biological Sciences*, 285(1878). doi: 10.1098/rspb.2018.0056
- Copernicus. (2021). *Climate Data Store of the Copernicus Marine Environment*. Retrieved from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/10.24381/cds.4c328c78?tab=doc> doi: 10.24381/cds.4c328c78
- Corney, P. M., Duc, M. G., Smart, S. M., Kirby, K. J., Bunce, R. G., & Marrs, R. H. (2006). Relationships between the species composition of forest field-layer vegetation and environmental drivers, assessed using a national scale survey. *Journal of Ecology*, 94(2), 383–401. doi: 10.1111/j.1365-2745.2006.01094.x
- Dadon, J. R., & Masello, J. F. (1999). Mechanisms generating and maintaining the admixture of zooplanktonic molluscs (Euthecosomata: Opisthobranchiata: Gastropoda) in the Subtropical Front of the South Atlantic. *Marine Biology*, 135(1), 171–179. doi: 10.1007/s002270050614
- Davis, C. S., & Wiebe, P. H. (1985). Macrozooplankton biomass in a warm-core Gulf Stream ring: time series changes in size structure, taxonomic composition, and vertical distribution. *Journal of Geophysical Research*, 90(C5), 8871–8884. doi: 10.1029/JC090iC05p08871
- De Broyer, C., et al. (Eds.). (2014). *Biogeographic Atlas of the Southern Ocean*. Cambridge, United Kingdom: SCAR.
- de Garidel-Thoron, T., Chaabane, S., Giraud, X., Meilland, J., Jonkers, L., Kucera, M., . . . Schiebel, R. (2022). The Foraminiferal Response to Climate Stressors Project: Tracking the Community Response of Planktonic Foraminifera to Historical Climate Change. *Frontiers in Marine Science*, 9(May), 1–6. doi: 10.3389/fmars.2022.827962
- Dietrich, G., Kalle, K., Krauss, W., & Siedler, G. (1975). *Allgemeine Meereskunde* (3rd ed.). Berlin/Stuttgart: Gebrüder Bornträger.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., . . . Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. doi: 10.1111/j.1600-0587.2012.07348.x
- Doubek, J. P., Goldfarb, S. K., & Stockwell, J. D. (2020). Should we be sampling zooplankton at night? *Limnology and Oceanography Letters*, 5(4), 313–321. doi: 10.1002/lol2.10151

- Dufois, F., Hardman-Mountford, N. J., Greenwood, J., Richardson, A. J., Feng, M., & Matear, R. J. (2016). Anticyclonic eddies are more productive than cyclonic eddies in subtropical gyres because of winter mixing. *Science Advances*, 2(5), 1–7. doi: 10.1126/sciadv.1600282
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342. doi: 10.1111/j.2041-210x.2010.00036.x
- Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697. doi: 10.1146/annurev.ecolsys.110308.120159
- Fabry, V. J. (1989). Aragonite production by pteropod molluscs in the subarctic Pacific. *Deep Sea Research Part A, Oceanographic Research Papers*, 36(11), 1735–1751. doi: 10.1016/0198-0149(89)90069-1
- Fabry, V. J., Seibel, B. A., Feely, R. A., & Orr, J. C. (2008). Impacts of ocean acidification on marine fauna and ecosystem processes. *ICES Journal of Marine Science*, 65(3), 414–432.
- Fallet, U., Ullgren, J. E., Castañeda, I. S., van Aken, H. M., Schouten, S., Ridderinkhof, H., & Brummer, G.-J. A. (2011). Contrasting variability in foraminiferal and organic paleotemperature proxies in sedimenting particles of the Mozambique Channel (SW Indian Ocean). *Geochimica et Cosmochimica Acta*, 75(20), 5834–5848. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0016703711004650> doi: <https://doi.org/10.1016/j.gca.2011.08.009>
- Fisher, R. A., & Yates, F. (1953). *Statistical tables for biological, agricultural and medical research*. Hafner Publishing Company.
- Frenger, I., Bianchi, D., Stührenberg, C., Oschlies, A., Dunne, J., Deutsch, C., ... Schütte, F. (2018). Biogeochemical Role of Subsurface Coherent Eddies in the Ocean: Tracer Cannonballs, Hypoxic Storms, and Microbial Stewpots? *Global Biogeochemical Cycles*, 32(2), 226–249. doi: 10.1002/2017GB005743
- Frerichs, W. E., Heiman, M. E., Borgman, L. E., & Be, A. W. H. (1972). Latitudinal variations in planktonic foraminiferal test porosity; Part 1, Optical studies. *The Journal of Foraminiferal Research*, 2(1), 6–13.
- Gangstø, R., Gehlen, M., Schneider, B., Bopp, L., Aumont, O., & Joos, F. (2008). Modeling the marine aragonite cycle: Changes under rising carbon dioxide and its role in shallow water CaCO₃ dissolution. *Biogeosciences*, 5(4), 1057–1072. doi: 10.5194/bg-5-1057-2008
- Garcia, H. E., Weathers, K. W., Paver, C. R., Smolyar, I., Boyer, T. P., Locarnini, M. M., ... Seidov, D. (2019a). *World ocean atlas 2018. Vol. 4: Dissolved inorganic nutrients (phosphate, nitrate and nitrate + nitrite, silicate)* (Tech. Rep.). NOAA. Retrieved from https://data.nodc.noaa.gov/woa/WOA18/DOC/woa18_vol4.pdf
- Garcia, H. E., Weathers, K. W., Paver, C. R., Smolyar, I., Boyer, T. P., Locarnini, M. M., ... Seidov, D. (2019b). *World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Dissolved Oxygen Saturation*. (Tech. Rep.). NOAA. Retrieved from <https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=o>
- Gardner, J., Manno, C., Bakker, D. C. E., Peck, V. L., & Tarling, G. A. (2017). Southern Ocean pteropods at risk from ocean warming and acidification. *Marine Biology*, 165(1), 8. Retrieved from <https://doi.org/10.1007/s00227-017-3261-3> doi: 10.1007/s00227-017-3261-3
- Gilmer, R. W., & Harbison, G. R. (1986). Morphology and field behavior of pteropod molluscs: feeding methods in the families Cavoliniidae, Limacinae and Peraclididae (Gastropoda: Thecosomata). *Marine Biology*, 91(1), 47–57.
- Govoni, J. J., Hare, J. A., Davenport, E. D., Chen, M. H., & Marancik, K. E.

- (2010). Mesoscale, cyclonic eddies as larval fish habitat along the south-east United States shelf: A Lagrangian description of the zooplankton community. *ICES Journal of Marine Science*, 67(3), 403–411. doi: 10.1093/icesjms/fsp269
- Gregor, L., & Gruber, N. (2021). OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data*, 13(2), 777–808.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3), 147–186. doi: 10.1016/S0304-3800(00)00354-9
- H2O.ai. (2021). *R Interface for H2O*. Retrieved from <https://github.com/h2oai/h2o-3>
- Halpern, B. S., Longo, C., Hardy, D., McLeod, K. L., Samhouri, J. F., Katona, S. K., ... Zeller, D. (2012). An index to assess the health and benefits of the global ocean. *Nature*, 488(7413), 615–620. doi: 10.1038/nature11397
- Hastie, T. J., & Tibshirani, R. J. (1990). Generalized additive models. In *Generalized additive models* (1st ed., pp. 136–173). New York: Routledge. doi: <https://doi.org/10.1201/9780203753781>
- Hauri, C., Gruber, N., McDonnell, A. M., & Vogt, M. (2013). The intensity, duration, and severity of low aragonite saturation state events on the California continental shelf. *Geophysical Research Letters*, 40(13), 3424–3428. doi: 10.1002/grl.50618
- Helaouët, P., & Beaugrand, G. (2009). Physiology, ecological niches and species distribution. *Ecosystems*, 12(8), 1235–1245. doi: 10.1007/s10021-009-9261-5
- Hemleben, C., Spindler, M., & Anderson, O. R. (1989). Host and symbiont relationships. In *Modern planktonic foraminifera* (pp. 86–111). Springer.
- Hofmann Elizondo, U., Righetti, D., Benedetti, F., & Vogt, M. (2021). Biome partitioning of the global ocean based on phytoplankton biogeography. *Progress in Oceanography*, 194 (February), 102530. Retrieved from <https://doi.org/10.1016/j.pocean.2021.102530> doi: 10.1016/j.pocean.2021.102530
- Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. *Ecological Modelling*, 468, 109944. doi: 10.1016/j.ecolmodel.2022.109944
- Horton, T., Gofas, S., Kroh, A., Poore, G. C., Read, G., Rosenberg, G., ... Vranken, S. (2017). Improving nomenclatural consistency: a decade of experience in the World Register of Marine Species. *European Journal of Taxonomy*, 389, 1–24. Retrieved from <https://www.marinespecies.org/aphia.php?p=sourcedetails&id=287923> doi: 10.5852/ejt.2017.389
- Hosie, G. (2021). *Southern Ocean Continuous Plankton Recorder Zooplankton Records, Ver. 9*. Australian Antarctic Data Centre. doi: doi:10.26179/ksds-s610
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Assessing the Fit of the Model. In *Applied logistic regression* (Vol. 398, pp. 153–225). John Wiley & Sons.
- Howes, E. L., Stemmann, L., Assailly, C., Irisson, J. O., Dima, M., Bijma, J., & Gattuso, J. P. (2015). Pteropod time series from the North Western Mediterranean (1967-2003): Impacts of pH and climate variability. *Marine Ecology Progress Series*, 531, 193–206. doi: 10.3354/meps11322
- Hunt, B. P., Pakhomov, E. A., Hosie, G. W., Siegel, V., Ward, P., & Bernard, K. (2008). Pteropods in Southern Ocean ecosystems. *Progress in Oceanography*, 78(3), 193–221. doi: 10.1016/j.pocean.2008.06.001
- Iglesias-Rodriguez, M. D., Armstrong, R., Feely, R., Hood, R., Kleypas, J., Milliman, J. D., ... Sarmiento, J. (2002). Progress made in study of ocean's calcium carbonate budget. *Eos*, 83(34), 2000–2002. doi: 10.1029/2002EO000267
- IMOS. (2022). *IMOS - AusCPR: Zooplankton Abundance*. NASA/Global Change

- Master Directory (GCMD) Earth Science Keywords. Version 6.0.0.0. Retrieved from <https://catalogue-imos.aodn.org.au/geonetwork/srv/eng/catalog.search#/metadata/c1344e70-480e-0993-e044-00144f7bc0f4>
- Ivanova, E. M., Conan, S. M.-H., Peeters, F. J. C., & Troelstra, S. R. (1999). Living Neogloboquadrina pachyderma sin and its distribution in the sediments from Oman and Somalia upwelling areas. *Marine Micropaleontology*, 36(2-3), 91–107.
- Jentzen, A., Schönfeld, J., & Schiebel, R. (2018). Assessment of the effect of increasing temperature on the ecology and assemblage structure of modern planktic foraminifers in the Caribbean and surrounding seas. *Journal of Foraminiferal Research*, 48(3), 251–272. doi: 10.2113/gsjfr.48.3.251
- Jin, X., Gruber, N., Dune, J. P., Sarmiento, J. L., & Armstrong, R. A. (2006). Diagnosing the contributions of phytoplankton functional groups to the production and export of particulate organic carbon, CaCO₃, and opal from global nutrient and alkalinity distributions. *Global Biogeochemical Cycles*, 20(2), 1–17. doi: 10.1029/2005GB002532
- Johns, D. M. B. A. o. t. U. M. (2021). *Continuous Plankton Recorder data for all coccolithophores, foraminifera and thecosomata - all areas*. The Archive for Marine Species and Habitats Data (DASSH). doi: <https://doi.org/10.17031/1763>
- Joint, I., Doney, S. C., & Karl, D. M. (2011). Will ocean acidification affect marine microbes. *ISME Journal*, 5(1), 1–7. doi: 10.1038/ismej.2010.79
- Jonkers, L., Hillebrand, H., & Kucera, M. (2019). Global change drives modern plankton communities away from the pre-industrial state. *Nature*, 570(7761), 372–375. Retrieved from <http://dx.doi.org/10.1038/s41586-019-1230-3> doi: 10.1038/s41586-019-1230-3
- Kämpf, J., & Chapman, P. (2016). *Upwelling systems of the world*. Springer.
- Kirby, R. R., & Beaugrand, G. (2009). Trophic amplification of climate warming. *Proceedings of the Royal Society B: Biological Sciences*, 276(1676), 4095–4103.
- Koppelman, R., Kullmann, B., Lahajnar, N., Martin, B., & Mohrholz, V. (2013). Onshore - offshore distribution of Thecosomata (Gastropoda) in the Benguela Current upwelling region off Namibia: species diversity and trophic position. *Journal of the Marine Biological Association of the United Kingdom*, 93(6), 1625–1640. Retrieved from <https://www.cambridge.org/core/article/onshoreoffshore-distribution-of-thecosomata-gastropoda-in-the-benguela-current-upwelling-region-off-namibia-species-diversity-and-trophic-position/9231CA48082E65EFFCC11D50AFDE45F4> doi: DOI:10.1017/S0025315413000052
- Kretschmer, K., Jonkers, L., Kucera, M., & Schulz, M. (2018). Modeling seasonal and vertical habitats of planktonic foraminifera on a global scale. *Biogeosciences*, 15(14), 4405–4429. doi: 10.5194/bg-15-4405-2018
- Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260), 583–621. doi: 10.1080/01621459.1952.10483441
- Kucera, M. (2007). Chapter Six Planktonic Foraminifera as Tracers of Past Oceanic Environments. *Developments in Marine Geology*, 1(07), 213–262. doi: 10.1016/S1572-5480(07)01011-1
- Kupferman, S. L., Becker, G. A., Simmons, W. F., Schauer, U., Marietta, M. G., & Nies, H. (1986). An intense cold core eddy in the North-East Atlantic. *Nature*, 319(6053), 474–477.
- Lalli, C. M., & Gilmer, R. W. (1989). *Pelagic snails: the biology of holoplanktonic gastropod mollusks*. Stanford University Press.
- Larson, R. J. (1986). Water content, organic content, and carbon and nitrogen composition of medusae from the northeast Pacific. *Journal of Experimental Marine Biology and Ecology*, 99(2), 107–120. doi: 10.1016/0022-0981(86)90231-5

- Lee, K. (2001). Global net community production estimated from the annual cycle of surface water total dissolved inorganic carbon. *Limnology and Oceanography*, 46(6), 1287–1297.
- Leinweber, A., & Gruber, N. (2013). Variability and trends of ocean acidification in the Southern California Current System: A time series from Santa Monica Bay. *Journal of Geophysical Research: Oceans*, 118(7), 3622–3633. doi: 10.1002/jgrc.20259
- Le Quéré, C., Buitenhuis, E. T., Moriarty, R., Alvain, S., Aumont, O., Bopp, L., ... Vallina, S. M. (2016). Role of zooplankton dynamics for Southern Ocean phytoplankton biomass and global biogeochemical cycles. *Biogeosciences*, 13(14), 4111–4133. doi: 10.5194/bg-13-4111-2016
- Le Quéré, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., ... Giraud, X. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11(11), 2016–2040.
- Lischka, S., Büdenbender, J., Boxhammer, T., & Riebesell, U. (2011). Impact of ocean acidification and elevated temperatures on early juveniles of the polar shelled pteropod *Limacina helicina*: Mortality, shell degradation, and shell growth. *Biogeosciences*, 8(4), 919–932. doi: 10.5194/bg-8-919-2011
- Lischka, S., & Riebesell, U. (2012). Synergistic effects of ocean acidification and warming on overwintering pteropods in the Arctic. *Global Change Biology*, 18(12), 3517–3528. doi: 10.1111/gcb.12020
- Locarnini, M. M., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., ... Smolyar, I. (2018). *World ocean atlas 2018, Vol. 1: Temperature* (Tech. Rep.). NOAA. Retrieved from <https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=t>
- Lombard, F., Labeyrie, L., Michel, E., Bopp, L., Cortijo, E., Retailleau, S., ... Jorissen, F. (2011). Modelling planktic foraminifer growth and distribution using an ecophysiological multi-species approach. *Biogeosciences*, 8(4), 853–873. doi: 10.5194/bg-8-853-2011
- Lombard, F., Labeyrie, L., Michel, E., Spero, H. J., & Lea, D. W. (2009). Modelling the temperature dependent growth rates of planktic foraminifera. *Marine Micropaleontology*, 70(1-2), 1–7. Retrieved from <http://dx.doi.org/10.1016/j.marmicro.2008.09.004> doi: 10.1016/j.marmicro.2008.09.004
- Longhurst, A. R. (2007). Physical Control of Ecological Processes. In A. R. Longhurst (Ed.), *Ecological geography of the sea* (2nd ed., pp. 51–70). Burlington: Academic Press. doi: 10.1016/b978-012455521-1/50005-x
- Maas, A. E., Lawson, G. L., & Tarrant, A. M. (2015). Transcriptome-wide analysis of the response of the thecosome pteropod *Clio pyramidata* to short-term CO₂ exposure. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 16, 1–9. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1744117X15000398> doi: <https://doi.org/10.1016/j.cbd.2015.06.002>
- Mack, H. R., Conroy, J. D., Blocksom, K. A., Stein, R. A., & Ludsin, S. A. (2012). A comparative analysis of zooplankton field collection and sample enumeration methods. *Limnology and Oceanography: Methods*, 10(JANUARY), 41–53. doi: 10.4319/lom.2012.10.41
- MacKas, D. L., & Galbraith, M. D. (2012). Pteropod time-series from the NE Pacific. *ICES Journal of Marine Science*, 69(3), 448–459. doi: 10.1093/icesjms/fsr163
- Mackas, D. L., Tsurumi, M., Galbraith, M. D., & Yelland, D. R. (2005). Zooplankton distribution and dynamics in a North Pacific Eddy of coastal origin: II. Mechanisms of eddy colonization by and retention of offshore species. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 52(7-8), 1011–1035. doi: 10.1016/j.dsr2.2005.02.008

- Manno, C., Bednaršek, N., Tarling, G. A., Peck, V. L., Comeau, S., Adhikari, D., ... Ziveri, P. (2017). Shelled pteropods in peril: Assessing vulnerability in a high CO₂ ocean. *Earth-Science Reviews*, 169(April), 132–145. doi: 10.1016/j.earscirev.2017.04.005
- Manno, C., Peck, V. L., & Tarling, G. A. (2016). Pteropod eggs released at high pCO₂ lack resilience to ocean acidification. *Scientific Reports*, 6(November 2015), 1–10. doi: 10.1038/srep25752
- Manno, C., Sandrini, S., Tositti, L., & Accornero, A. (2007). First stages of degradation of *Limacina helicina* shells observed above the aragonite chemical lysocline in Terra Nova Bay (Antarctica). *Journal of Marine Systems*, 68(1-2), 91–102.
- McGowan, J. A. (1967). Distributional atlas of pelagic molluscs in the California Current region. *CalCOFI Atlas*, 6.
- Meilland, J., Fabri-Ruiz, S., Koubbi, P., Monaco, C. L., Cotte, C., Hosie, G. W., ... Howa, H. (2016). Planktonic foraminiferal biogeography in the Indian sector of the Southern Ocean: Contribution from CPR data. *Deep-Sea Research Part I: Oceanographic Research Papers*, 110, 75–89. Retrieved from <http://dx.doi.org/10.1016/j.dsr.2015.12.014> doi: 10.1016/j.dsr.2015.12.014
- Mekkes, L., Renema, W., Bednaršek, N., Alin, S. R., Feely, R. A., Huisman, J., ... Peijenburg, K. T. C. A. (2021). Pteropods make thinner shells in the upwelling region of the California Current Ecosystem. *Scientific Reports*, 11(1), 1–11. Retrieved from <https://doi.org/10.1038/s41598-021-81131-9> doi: 10.1038/s41598-021-81131-9
- Mekkes, L., Sepúlveda-Rodríguez, G., Bielkinité, G., Wall-Palmer, D., Brummer, G. J. A., Dämmer, L. K., ... Peijenburg, K. T. (2021). Effects of Ocean Acidification on Calcification of the Sub-Antarctic Pteropod *Limacina retroversa*. *Frontiers in Marine Science*, 8(March), 1–12. doi: 10.3389/fmars.2021.581432
- Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., Normand, S., ... Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, 37(12), 1267–1281. doi: 10.1111/ecog.00845
- Michaels, A. F., Caron, D. A., Swanberg, N. R., & Howse, F. A. (1995). Primary productivity by symbiont-bearing planktonic sarcodines (Acantharia, Radiolaria, Foraminifera) in surface waters near Bermuda. *Journal of Plankton Research*, 17(1), 103–129. doi: 10.1093/plankt/17.1.103
- Milosavljić, M., Lučić, D., Gangai, B., & Onofri, I. (2014). Mesh size effects on mesozooplankton community structure in a semi-enclosed coastal area and surrounding sea (South Adriatic Sea). *Marine Ecology*, 35(4), 445–455. doi: 10.1111/maec.12101
- Morard, R., Darling, K. F., Mahé, F., Audic, S., Ujiié, Y., Weiner, A. K. M., ... Quillévéré, F. (2015). PFR2: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. *Molecular Ecology Resources*, 15(6), 1472–1485.
- Moriarty, R., Buitenhuis, E. T., & Le Quéré, C. (2013). Distribution of known macrozooplankton abundance and biomass in the global ocean. *Earth System Science Data*, 5(2), 241–257. doi: 10.5194/essd-5-241-2013
- Mucci, A. (1983). The solubility of calcite and aragonite in seawater at various salinities, temperatures, and one atmosphere total pressure. *Am. J. Sci.*, 283(7), 780–799.
- Myers, T. D. (1968). *Horizontal and vertical distribution of thecosomatous pteropods off Cape Hatteras*. Duke University.
- Naidu, P. D., & Malmgren, B. A. (1996). A high-resolution record of late Quaternary upwelling along the Oman Margin, Arabian Sea based on planktonic foraminifera. *Paleoceanography*, 11(1), 129–140.

- NASA OB.DAAC. (2018a). *Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Chlorophyll Data*. Retrieved from https://oceandata.sci.gsfc.nasa.gov/directaccess/SeaWiFS/Mapped/Monthly_Climatology/9km/chlor_a/ doi: 10.5067/ORBVIEW-2/SEAWIFS/L3M/CHL/2018
- NASA OB.DAAC. (2018b). *Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Downwelling Diffuse Attenuation Coefficient Data*. Retrieved from https://oceandata.sci.gsfc.nasa.gov/directaccess/SeaWiFS/Mapped/Monthly_Climatology/9km/Kd.490/ doi: 10.5067/ORBVIEW-2/SEAWIFS/L3M/KD/2018
- NASA OB.DAAC. (2018c). *Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Euphotic Depth Data*. Retrieved from https://oceandata.sci.gsfc.nasa.gov/directaccess/SeaWiFS/Mapped/Monthly_Climatology/9km/Zeu_lee/ doi: 10.5067/ORBVIEW-2/SEAWIFS/L3M/ZLEE/2018
- NASA OB.DAAC. (2018d). *Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Garver-Siegel-Maritorena (GSM) Model Data*. Retrieved from https://oceandata.sci.gsfc.nasa.gov/directaccess/SeaWiFS/Mapped/Monthly_Climatology/9km/bbp.443_gsm/ doi: 10.5067/ORBVIEW-2/SEAWIFS/L3M/GSM/2018
- NASA OB.DAAC. (2018e). *Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Photosynthetically Available Radiation Data*. doi: 10.5067/ORBVIEW-2/SEAWIFS/L3M/PAR/2018
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of hydrology*, 10(3), 282–290.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370–384.
- OB.DAAC, N. (2018). *Sea-viewing Wide Field-of-view Sensor (SeaWiFS)*.
- O'Brien, C. J. (2015). *Global-scale distributions of marine haptophyte phytoplankton* (Unpublished doctoral dissertation). ETH Zurich.
- O'Brien, C. J., Vogt, M., & Gruber, N. (2016). Global coccolithophore diversity: Drivers and future change. *Progress in Oceanography*, 140, 27–42. Retrieved from <http://dx.doi.org/10.1016/j.pocean.2015.10.003> doi: 10.1016/j.pocean.2015.10.003
- O'Brien, T. D. (2010). *COPEPOD, a global plankton database : a review of the 2010 database contents, processing methods, and access interface*. Retrieved from <https://repository.library.noaa.gov/view/noaa/5040>
- Ohman, M. D., Lavaniegos, B. E., & Townsend, A. W. (2009). Multi-decadal variations in calcareous holozooplankton in the California Current System: thecosome pteropods, heteropods, and foraminifera. *Geophysical Research Letters*, 36(18), 2–6. doi: 10.1029/2009GL039901
- Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., ... Yool, A. (2005). Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature*, 437(7059), 681–686. doi: 10.1038/nature04095
- Peijnenburg, K. T. C. A. (2021). *Personal communication*.
- Peijnenburg, K. T. C. A., Janssen, A. W., Wall-Palmer, D., Goetze, E., Maas, A. E., Todd, J. A., & Marlétaz, F. (2020). The origin and diversification of pteropods precede past perturbations in the Earth's carbon cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 117(41), 25609–25617. doi: 10.1073/pnas.1920918117
- Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., ... Coordinators, T. O. C. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, 2(1), 150023. Retrieved from <https://doi.org/10.1038/sdata.2015.23> doi: 10.1038/sdata.2015.23
- Pinkerton, M. H., Décima, M., Kitchener, J. A., Takahashi, K. T., Robinson, K. V.,

- Stewart, R., & Hosie, G. W. (2020). Zooplankton in the Southern Ocean from the continuous plankton recorder: Distributions and long-term change. *Deep-Sea Research Part I: Oceanographic Research Papers*, 162(May 2019). doi: 10.1016/j.dsr.2020.103303
- Pinkerton, M. H., Smith, A. N., Raymond, B., Hosie, G. W., Sharp, B., Leathwick, J. R., & Bradford-Grieve, J. M. (2010). Spatial and seasonal distribution of adult *Oithona similis* in the Southern Ocean: Predictions using boosted regression trees. *Deep-Sea Research Part I: Oceanographic Research Papers*, 57(4), 469–485. Retrieved from <http://dx.doi.org/10.1016/j.dsr.2009.12.010> doi: 10.1016/j.dsr.2009.12.010
- Qiao, H., Feng, X., Escobar, L. E., Peterson, A. T., Soberón, J., Zhu, G., & Papeş, M. (2019). An evaluation of transferability of ecological niche models. *Ecography*, 42(3), 521–534. doi: 10.1111/ecog.03986
- Rembauville, M., Meilland, J., Ziveri, P., Schiebel, R., Blain, S., & Salter, I. (2016). Planktic foraminifer and coccolith contribution to carbonate export fluxes over the central Kerguelen Plateau. *Deep-Sea Research Part I: Oceanographic Research Papers*, 111, 91–101. Retrieved from <http://dx.doi.org/10.1016/j.dsr.2016.02.017> doi: 10.1016/j.dsr.2016.02.017
- Rhumbler, L. (1911). Die Foraminiferen (Thalamophoren) der Plankton Expedition, Pt. 1, Die Allgemeinen Organisationsverhältnisse der Foraminifera. *Lipsius & Tischer, Kiel und Leipzig*, 331.
- Richardson, A. J., Walne, A. W., John, A. W., Jonas, T. D., Lindley, J. A., Sims, D. W., ... Witt, M. (2006). Using continuous plankton recorder data. *Progress in Oceanography*, 68(1), 27–74. doi: 10.1016/j.pocean.2005.09.011
- Righetti, D., Vogt, M., Gruber, N., Psomas, A., & Zimmermann, N. E. (2019). Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Science Advances*, 5(5), 1–11. doi: 10.1126/sciadv.aau6253
- Rillo, M. C., Woolley, S., & Hillebrand, H. (2022). Drivers of global pre-industrial patterns of species turnover in planktonic foraminifera. *Ecography*, 2022(1), 1–11. doi: 10.1111/ecog.05892
- Romagnan, J.-B., Legendre, L., Guidi, L., Jamet, J.-L., Jamet, D., Mousseau, L., ... Sardet, C. (2015). Comprehensive model of annual plankton succession based on the whole-plankton time series approach. *PLoS One*, 10(3), e0119219.
- Rost, B., & Riebesell, U. (2004). Coccolithophores and the biological pump: responses to environmental changes. In H. R. Thierstein & J. R. Young (Eds.), *Coccolithophores* (pp. 99–125). Springer. doi: 10.1007/978-3-662-06278-4{_}5
- Rothschild, B. J., & Osborn, T. R. (1988). Small-scale turbulence and plankton contact rates. *Journal of plankton Research*, 10(3), 465–474.
- Sallée, J. B., Pellichero, V., Akhoudas, C., Pauthenet, E., Vignes, L., Schmidtke, S., ... Kuusela, M. (2021). Summertime increases in upper-ocean stratification and mixed-layer depth. *Nature*, 591(7851), 592–598. doi: 10.1038/s41586-021-03303-x
- Sarmiento, J., & Gruber, N. (2006). Calcium Carbonate Cycle. In *Ocean biogeochemical dynamics* (p. 359 - 391). Princeton University Press.
- Sarmiento, J. L., & Gruber, N. (2006). Carbon Cycle. In *Ocean biogeochemical dynamics* (pp. 318–358). Princeton University Press.
- Schiebel, R. (2002). Planktic foraminiferal sedimentation and the marine calcite budget. *Global Biogeochemical Cycles*, 16(4), 3–1. doi: 10.1029/2001gb001459
- Schiebel, R. (2021). *Personal Communication*.
- Schiebel, R., Barker, S., Lentz, R., Thomas, H., & Bollmann, J. (2007). Planktic foraminiferal dissolution in the twilight zone. *Deep Sea Research Part II: Topical Studies in Oceanography*, 54(5-7), 676–686.
- Schiebel, R., & Hemleben, C. (2000). Interannual variability of planktic foraminiferal populations and test flux in the eastern North Atlantic Ocean (JGOFS). *Deep-Sea Research Part II: Topical Studies in Oceanography*, 47(9-11), 1809–1852.

- doi: 10.1016/S0967-0645(00)00008-4
- Schiebel, R., & Hemleben, C. (2005). Modern planktic foraminifera. *Paläontologische Zeitschrift*, 79(1), 135–148. doi: 10.1007/bf03021758
- Schiebel, R., & Hemleben, C. (2017). *Planktic foraminifera in the modern ocean* (2nd ed.). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-662-50297-6
- Schiebel, R., Hiller, B., & Hemleben, C. (1995). Impacts of storms on Recent planktic foraminiferal test production and CaCO₃ flux in the North Atlantic at 47 °N, 20 °W (JGOFS). *Marine Micropaleontology*, 26(1-4), 115–129. doi: 10.1016/0377-8398(95)00035-6
- Schiebel, R., & Movellan, A. (2012). First-order estimate of the planktic foraminifer biomass in the modern ocean. *Earth System Science Data*, 4(1), 75–89. doi: 10.5194/essd-4-75-2012
- Schiebel, R., Waniek, J., Bork, M., & Hemleben, C. (2001). Planktic foraminiferal production stimulated by chlorophyll redistribution and entrainment of nutrients. *Deep-Sea Research Part I: Oceanographic Research Papers*, 48(3), 721–740. doi: 10.1016/S0967-0637(00)00065-0
- Schiebel, R., Waniek, J., Zeltner, A., & Alves, M. (2002). Impact of the Azores Front on the distribution of planktic foraminifera, shelled gastropods, and coccolithophorids. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 49(19), 4035–4050. doi: 10.1016/S0967-0645(02)00141-8
- Schiebel, R., Zeltner, A., Treppke, U. F., Waniek, J. J., Bollmann, J., Rixen, T., & Hemleben, C. (2004). Distribution of diatoms, coccolithophores and planktic foraminifera along a trophic gradient during SW monsoon in the Arabian Sea. *Marine Micropaleontology*, 51(3-4), 345–371. doi: 10.1016/j.marmicro.2004.02.001
- Schmidt, D. N., Renaud, S., Bollmann, J., Schiebel, R., & Thierstein, H. R. (2004). Size distribution of Holocene planktic foraminifer assemblages: Biogeography, ecology and adaptation. *Marine Micropaleontology*, 50(3-4), 319–338. doi: 10.1016/S0377-8398(03)00098-7
- Schmidt, G. A., & Mulitza, S. (2002). Global calibration of ecological models for planktic foraminifera from core-top carbonate oxygen-18. *Marine Micropaleontology*, 44(3-4), 125–140. doi: 10.1016/S0377-8398(01)00041-X
- Schmidt-Nielsen, K. (1997). *Animal physiology: adaptation and environment*. Cambridge university press.
- Seuront, L., Schmitt, F., & Lagadeuc, Y. (2001). Turbulence intermittency, small-scale phytoplankton patchiness and encounter rates in plankton: Where do we go from here? *Deep-Sea Research Part I: Oceanographic Research Papers*, 48(5), 1199–1215. doi: 10.1016/S0967-0637(00)00089-3
- Shapiro, S. S., & Wilk, M. B. (1965, 12). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591–611. Retrieved from <https://doi.org/10.1093/biomet/52.3-4.591> doi: 10.1093/biomet/52.3-4.591
- Siccha, M., Schiebel, R., Schmidt, S., & Howa, H. (2012). Short-term and small-scale variability in planktic foraminifera test flux in the Bay of Biscay. *Deep Sea Research Part I: Oceanographic Research Papers*, 64, 146–156.
- Skjoldal, H. R., Wiebe, P. H., Postel, L., Knutsen, T., Kaartvedt, S., & Sameoto, D. D. (2013). Intercomparison of zooplankton (net) sampling systems: Results from the ICES/GLOBEC sea-going workshop. *Progress in Oceanography*, 108, 1–42. Retrieved from <http://dx.doi.org/10.1016/j.pocean.2012.10.006> doi: 10.1016/j.pocean.2012.10.006
- Soviadan, Y. D., Benedetti, F., Brandão, M. C., Ayata, S. D., Irisson, J. O., Jamet, J. L., ... Stemann, L. (2022). Patterns of mesozooplankton community composition and vertical fluxes in the global ocean. *Progress in Oceanography*, 200(September 2021). doi: 10.1016/j.pocean.2021.102717
- Spieß, A.-N., & Neumeyer, N. (2010, 6). An evaluation of R² as an inadequate measure for nonlinear models in pharmacological and biochemical research: a

- Monte Carlo approach. *BMC pharmacology*, 10, 6. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/20529254><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2892436/> doi: 10.1186/1471-2210-10-6
- Spindler, M., Hemleben, C., Salomons, J. B., & Smit, L. P. (1984). Feeding behavior of some planktonic foraminifers in laboratory cultures. *The Journal of Foraminiferal Research*, 14(4), 237–249.
- Steinhardt, J., Cl  roux, C., Ullgren, J., de Nooijer, L., Durgadoo, J. V., Brummer, G.-J., & Reichart, G.-J. (2014). Anti-cyclonic eddy imprint on calcite geochemistry of several planktonic foraminiferal species in the Mozambique Channel. *Marine Micropaleontology*, 113, 20–33.
- Stepien, J. C. (1980). The occurrence of chaetognaths, pteropods and euphausiids in relation to deep flow reversals in the Straits of Florida. *Deep Sea Research Part A. Oceanographic Research Papers*, 27(12), 987–1011.
- Str  mberg, K. H., Smyth, T. J., Allen, J. I., Pitois, S., & O’Brien, T. D. (2009). Estimation of global zooplankton biomass from satellite ocean colour. *Journal of Marine Systems*, 78(1), 18–27. Retrieved from <http://dx.doi.org/10.1016/j.jmarsys.2009.02.004> doi: 10.1016/j.jmarsys.2009.02.004
- Student. (1908). The Probable Error of A Mean. *Biometrika*, 6(1), 1–25.
- Takahashi, K., & B  , A. W. (1984). Planktonic foraminifera: factors controlling sinking speeds. *Deep Sea Research Part A, Oceanographic Research Papers*, 31(12), 1477–1500. doi: 10.1016/0198-0149(84)90083-9
- Thibodeau, P. S., Steinberg, D. K., Stammerjohn, S. E., & Hauri, C. (2019). Environmental controls on pteropod biogeography along the Western Antarctic Peninsula. *Limnology and Oceanography*, 64, S240-S256. doi: 10.1002/lno.11041
- Thuiller, W., Gu  guen, M., Renaud, J., Karger, D. N., & Zimmermann, N. E. (2019). Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, 10(1), 1–9. Retrieved from <http://dx.doi.org/10.1038/s41467-019-09519-w> doi: 10.1038/s41467-019-09519-w
- Tittensor, D. P., Novaglio, C., Harrison, C. S., Heneghan, R. F., Barrier, N., Bianchi, D., ... Blanchard, J. L. (2021). Next-generation ensemble projections reveal higher climate risks for marine ecosystems. *Nature Climate Change*, 11(11), 973–981. doi: 10.1038/s41558-021-01173-9
- Tseng, L. C., Dahms, H. U., Hung, J. J., Chen, Q. C., & Hwang, J. S. (2011). Can different mesh sizes affect the results of copepod community studies? *Journal of Experimental Marine Biology and Ecology*, 398(1-2), 47–55. Retrieved from <http://dx.doi.org/10.1016/j.jembe.2010.12.007> doi: 10.1016/j.jembe.2010.12.007
- Tsurumi, M., Mackas, D. L., Whitney, F. A., DiBacco, C., Galbraith, M. D., & Wong, C. S. (2005). Pteropods, eddies, carbon flux, and climate variability in the Alaska Gyre. *Deep Sea Research Part II: Topical Studies in Oceanography*, 52(7-8), 1037–1053.
- Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean’s biological pump. *Progress in Oceanography*, 130, 205–248. Retrieved from <http://dx.doi.org/10.1016/j.pocean.2014.08.005> doi: 10.1016/j.pocean.2014.08.005
- Vereshchaka, A. L., Lunina, A. A., & Mikaelyan, A. S. (2022). Surface chlorophyll concentration as a mesoplankton biomass assessment tool in the Southern Ocean region. *Global Ecology and Biogeography*, 31(3), 405–424. doi: 10.1111/geb.13435
- Waldock, C., Stuart-Smith, R. D., Albouy, C., Cheung, W. W. L., Edgar, G. J., Mouillot, D., ... Pellissier, L. (2022, 1). A quantitative review of abundance-based species distribution models. *Ecography*, 2022(1). Retrieved from <https://doi.org/10.1111/ecog.05694> doi: <https://doi.org/10.1111/ecog.05694>

- Wall-Palmer, D., Smart, C. W., Kirby, R., Hart, M. B., Peijnenburg, K. T., & Janssen, A. W. (2016). A review of the ecology, palaeontology and distribution of atlantid heteropods (Caenogastropoda: Pterotracheoidea: Atlantidae). *Journal of Molluscan Studies*, 82(2), 221–234. doi: 10.1093/mollus/eyv063
- Wang, K., Hunt, B. P., Liang, C., Pauly, D., & Pakhomov, E. A. (2017). Reassessment of the life cycle of the pteropod *Limacina helicina* from a high resolution interannual time series in the temperate North Pacific. *ICES Journal of Marine Science*, 74(7), 1906–1920. doi: 10.1093/icesjms/fsx014
- Weinfurt, K. P. (1995). Multivariate analysis of variance. In G. L. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (p. 245 - 276). American Psychological Association.
- Weinkauff, M. F., Kunze, J. G., Waniek, J. J., & Kučera, M. (2016). Seasonal variation in shell calcification of planktonic foraminifera in the ne atlantic reveals species-specific response to temperature, productivity, and optimum growth conditions. *PLoS ONE*, 11(2), 1–33. doi: 10.1371/journal.pone.0148363
- Wells, F. E. (1973). Effects of mesh size on estimation of population densities of tropical euthecosomatous pteropods. *Marine Biology*, 20(4), 347–350. doi: 10.1007/BF00354276
- Wilson, R. W., Millero, F. J., Taylor, J. R., Walsh, P. J., Christensen, V., Jennings, S., & Grosell, M. (2009). Contribution of fish to the marine inorganic carbon cycle. *Science*, 323(5912), 359–362.
- Wormuth, J. H. (1981). Vertical distributions and diel migrations of Euthecosomata in the northwest Sargasso Sea. *Deep Sea Research Part A. Oceanographic Research Papers*, 28(12), 1493–1515.
- Zamelczyk, K., Fransson, A., Chierici, M., Jones, E., Meilland, J., Anglada-Ortiz, G., & Lødemel, H. H. (2021). Distribution and Abundances of Planktic Foraminifera and Shelled Pteropods During the Polar Night in the Sea-Ice Covered Northern Barents Sea. *Frontiers in Marine Science*, 8(October), 1–19. doi: 10.3389/fmars.2021.644094
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., ... Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261–1277. doi: 10.1111/ecog.04960
- Zweng, M. M., Seidov, D., Boyer, T., Locarnini, M., Garcia, H., Mishonov, A., ... Smolyar, I. (2019). *World ocean atlas 2018, Vol. 2: Salinity* (Tech. Rep.). NOAA. Retrieved from <https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=s>

Figure 1.



Plankton observation data – 2.1

Abundance counts

- COPEPOD, MAREDAT, NA-NP CPR, SO-CPR, AUS-CPR, AMT, unpublished datasets



Preprocessing and biomass calculation – 2.1.1 – 2.1.3

- Taxonomic harmonization
- Remove incomplete data
- Biomass calculation
- Exclude coastal data, NA-NP CPR and gymnosomata
- Surface data aggregation
- Outlier removal
- Match to environmental conditions



Environmental predictor data – 2.2.1

- Monthly gridded products
- World Ocean Atlas 2018
 - SeaWiFs



Environmental predictor choice – 2.2.1

- Exclude highly correlated predictors
- Choose predictors that explain the highest fraction of TC biomass variance



SDMs – 2.2.2

- Generalized Linear Model (GLM)
- Generalized Additive Model (GAM)
- Random Forest (RF)
- Boosted Regression Tree (GBM)
- Neural Network (DL)



Growth rate parametrization – 2.3.2

- Pteropoda: average overturn periods (Bednaršek, Mozina, et al., 2012)
- Foraminifera: temperature-dependent growth rates (Lombard et al., 2009)



TC biomass distribution predictions – 2.3.1

- Monthly gridded global predictions



TIC surface flux estimates – 2.3.2

- Global annual TIC flux magnitude

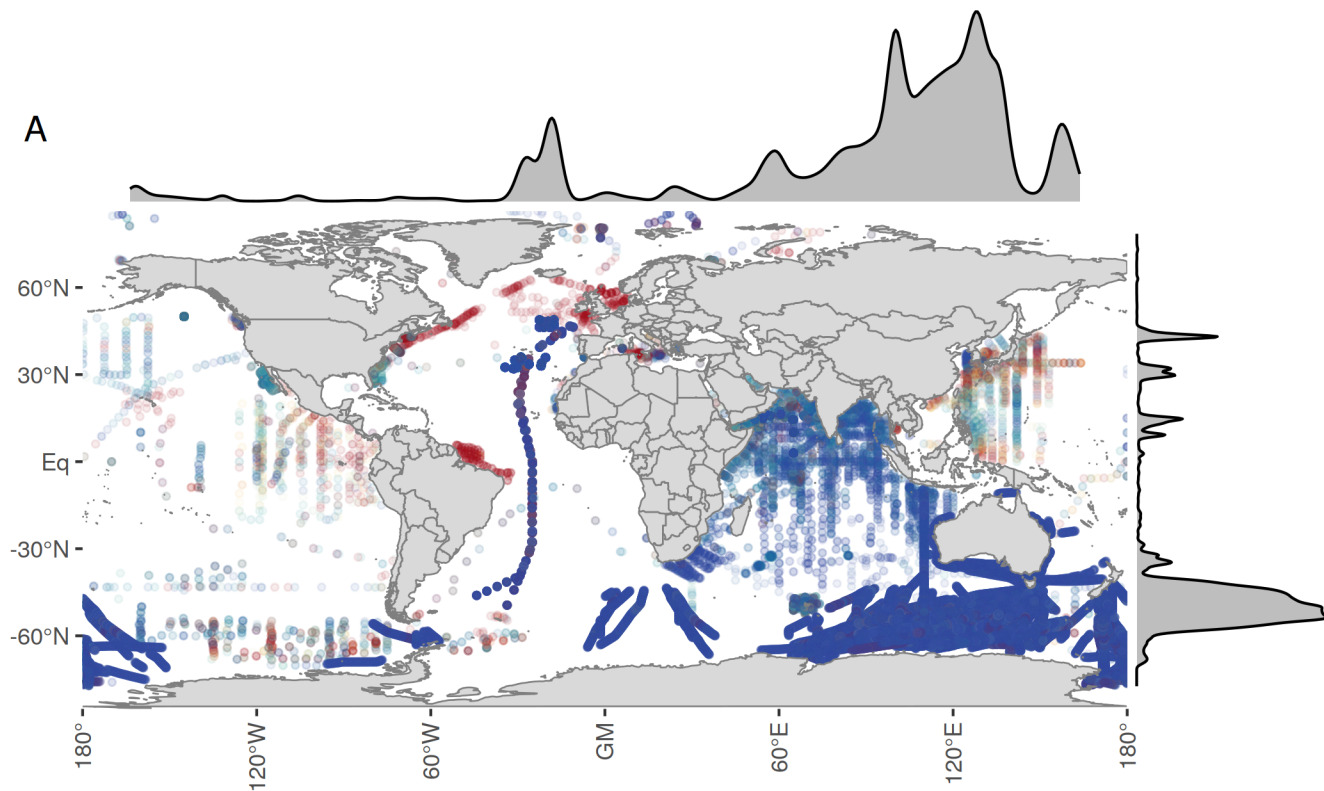


Assessment & uncertainty analysis – 2.3.3 – 2.3.4

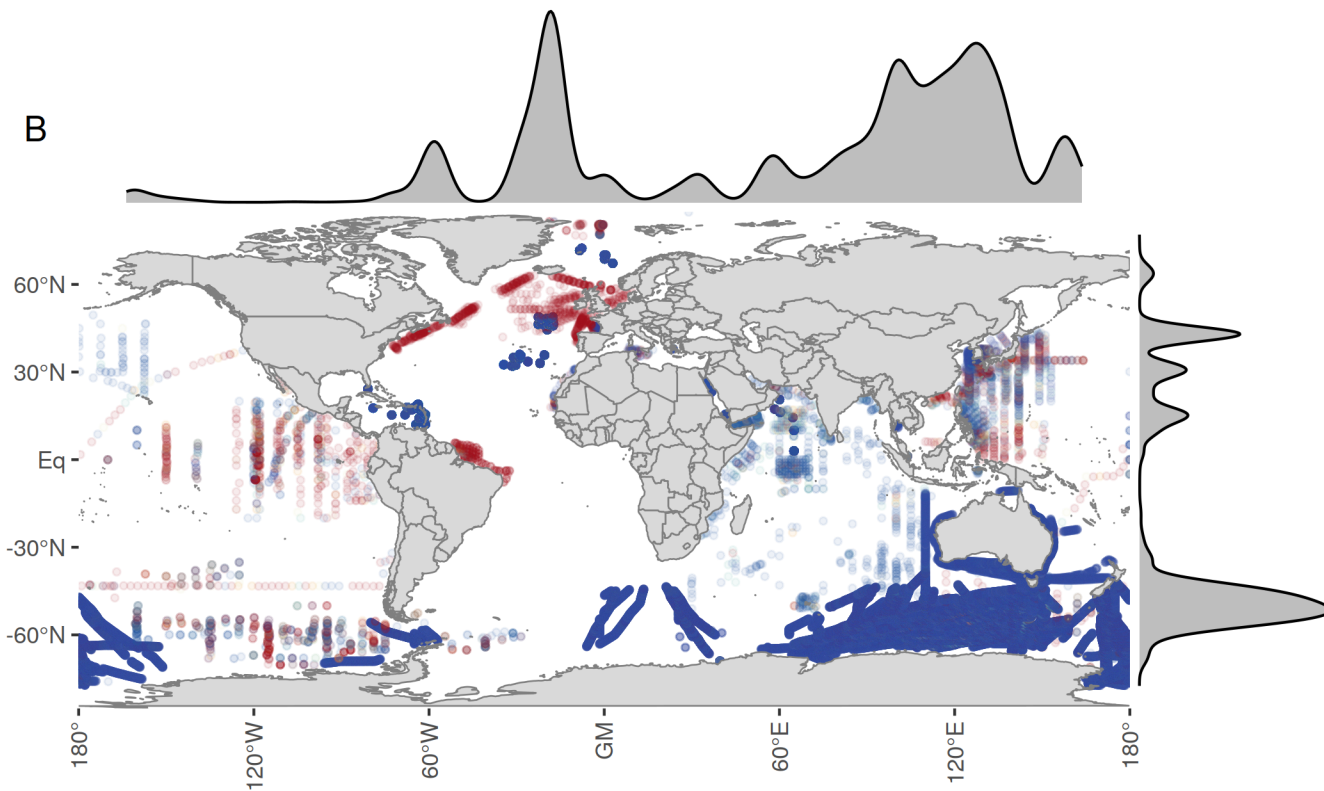
- Environmental driver analysis
 - Permutation analysis
 - Partial dependence plots
- Model performance
 - RMSE, r², residual patterns
- Uncertainty
 - Spatial inter-model uncertainty
 - Effect of model parameter choice on predictions

Figure 2.

A



B



log₁₀ Abundance (# m⁻³)

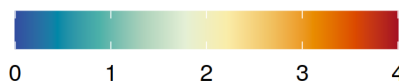
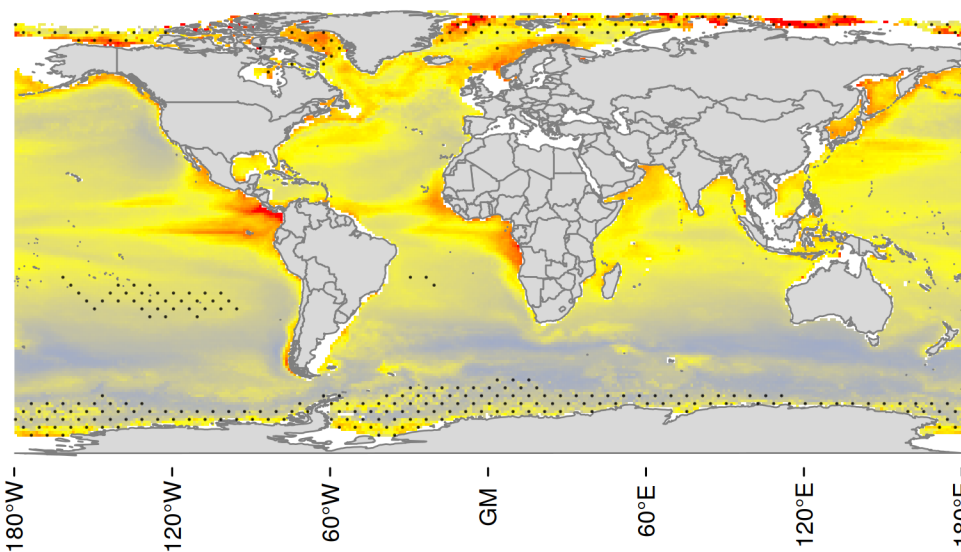
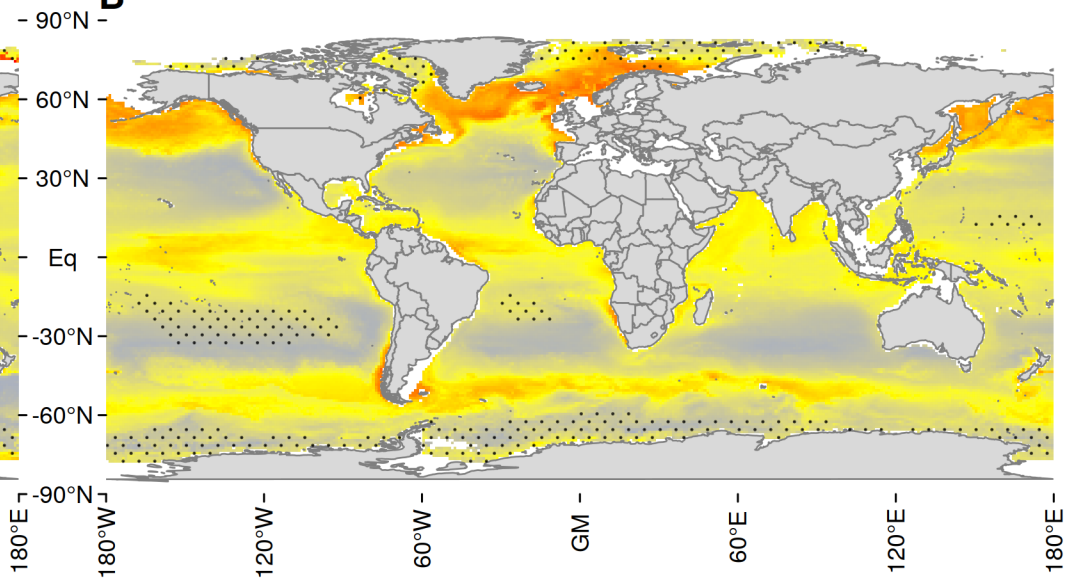


Figure 3.

A

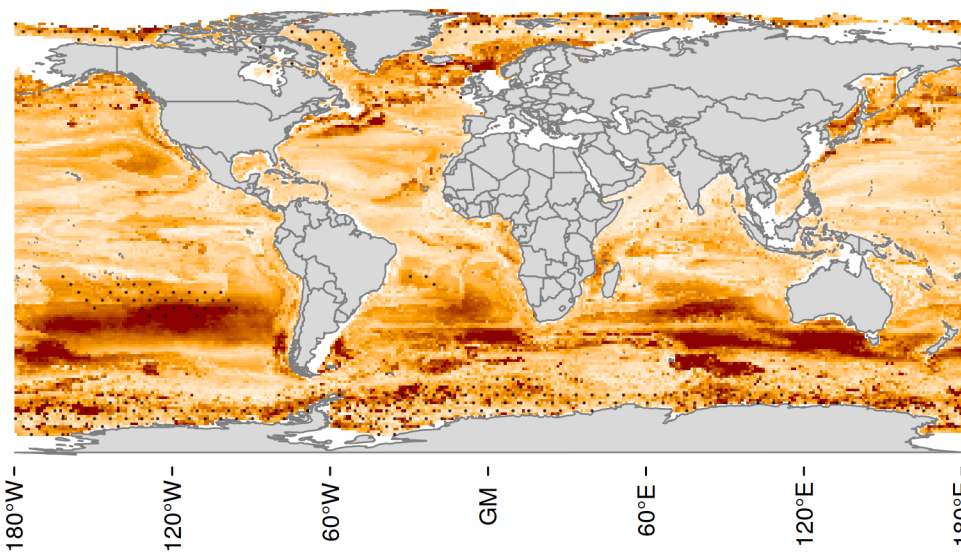
\log_{10} Total carbon (mg C m^{-3})

0.0 0.1 0.2 0.3 0.4 0.5

B

\log_{10} Total carbon ($\mu\text{g C m}^{-3}$)

0.0 0.5 1.0 1.5 2.0 2.5

C

Relative SD (%)

0 20 40 60

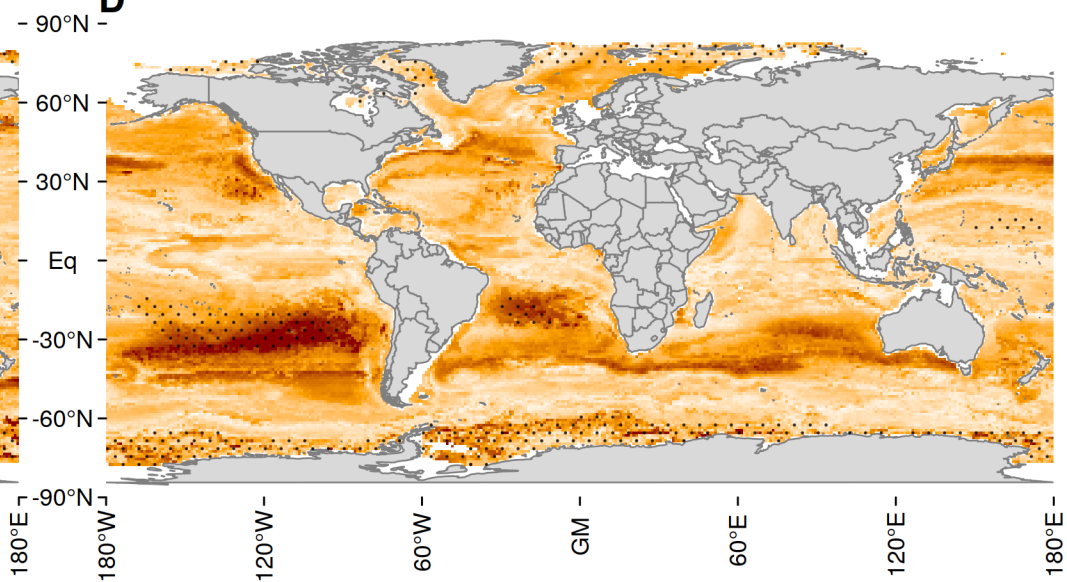
D

Figure 4.

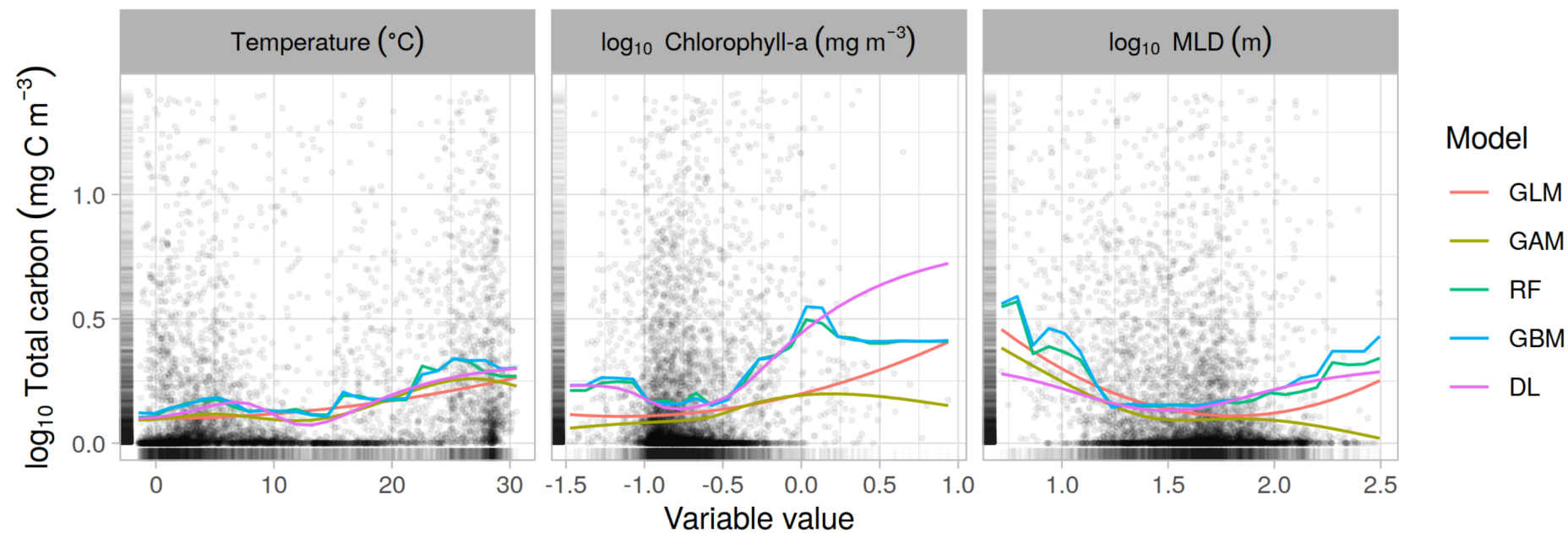
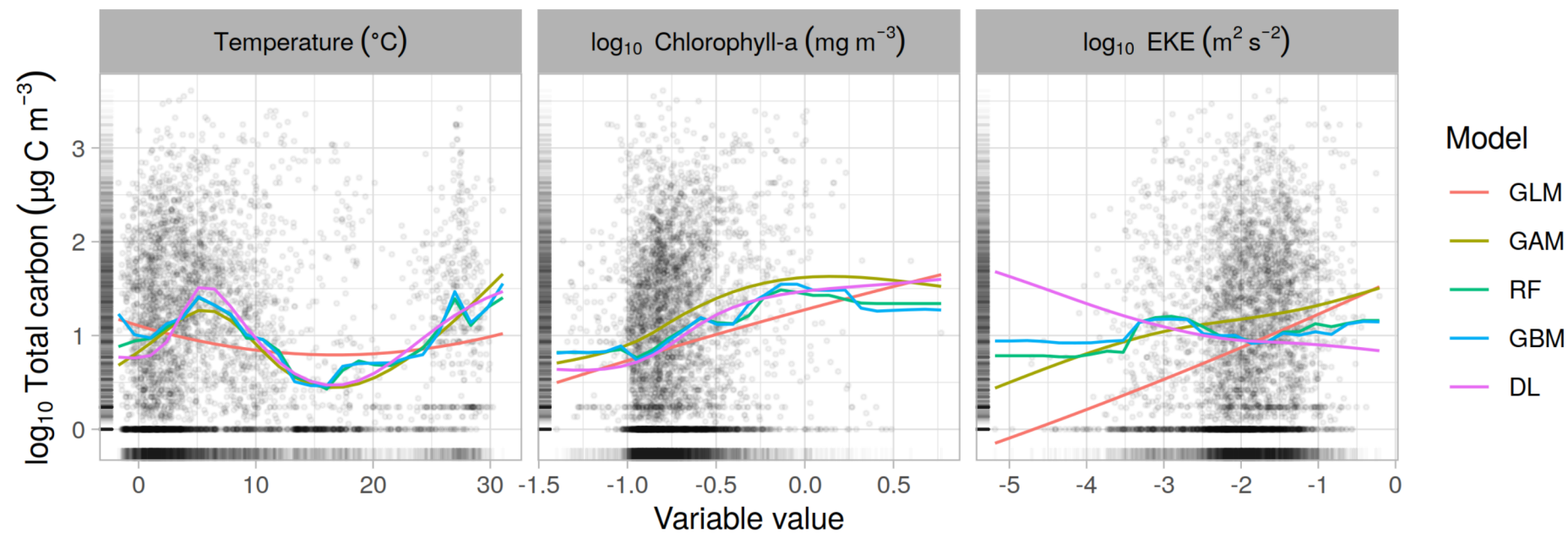
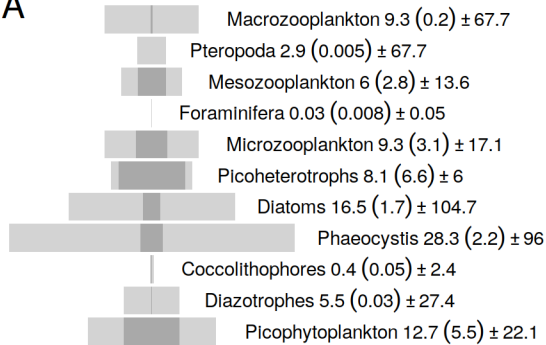
A**B**

Figure 5.

A



B

