AGU ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2025GL120153

Identifying the Environmental Drivers of Spatial Variability in Zooplankton Community Grazing Dynamics

S. A. Meyjes[1], B. B. Cael[2], C. M. Petrik[3], T. Panaiotis[4], T. Rohr[5,6], and A. Mashayek[1,4]

[1]Department of Earth Sciences, University of Cambridge, Cambridge, UK, [2]Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA, [3]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA, [4]National Oceanography Centre, Southampton, UK, [5]Institute for Marine and Antarctic Science, University of Tasmania, Hobart, TAS, Australia, [6]Australian Antarctic Program Partnership, University of Tasmania, Hobart, TAS, Australia

Key Points:

- Spatial variation in zooplankton grazing dynamics can be explained by sea surface temperature, mixed layer depth and chlorophyll
- Chlorophyll is the best predictor of community-averaged zooplankton grazing dynamics, both globally and regionally
- Globally, slower grazing plankton communities are associated with higher chlorophyll concentrations and deeper mixed layer depths

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. A. Meyjes,
sam296@cam.ac.uk

Citation:

Meyjes, S. A., Cael, B. B., Petrik, C. M., Panaiotis, T., Rohr, T., & Mashayek, A. (2026). Identifying the environmental drivers of spatial variability in zooplankton community grazing dynamics. *Geophysical Research Letters*, *53*, e2025GL120153. https://doi.org/10.1029/2025GL120153

Received 27 OCT 2025
Accepted 30 APR 2026

Abstract Biogeochemical models used to estimate carbon export fluxes are highly sensitive to their parameterization of zooplankton grazing dynamics. Zooplankton grazing dynamics have been shown to vary spatially, however it is less clear how they are responding to underlying environmental drivers. In this study, we use machine learning regression models to quantify the relationship between spatial variability in observationally inferred zooplankton grazing dynamics and three environmental drivers. The majority (R^2 = 0.80) of the variance in zooplankton grazing dynamics can be explained by sea surface temperature, mixed layer depth and chlorophyll. Globally, chlorophyll is the best predictor of zooplankton grazing dynamics, with slower grazing plankton communities associated with higher chlorophyll concentrations and deeper mixed layer depths. The findings show that zooplankton community grazing and composition is well constrained by biological drivers, which could provide a pathway to parameterize more diverse grazing dynamics in models and in turn simulate more realistic carbon export estimates.

Plain Language Summary The community of marine organisms that exists in any location is influenced by environmental conditions such as sea surface temperature, ocean mixing and prey availability. The grazing activity of tiny marine animals called zooplankton has been shown to vary spatially, however, it is less clear what environmental conditions drive this variation. In this study, we use models to examine the relationship between the marine environment and zooplankton grazing activity. Prey abundance is shown to be the key determinant of spatial variation in zooplankton grazing, when compared to sea surface temperature and ocean mixing conditions. Zooplankton grazing rates can be used to model how much marine carbon is stored in the global ocean, therefore these findings have implications for modeling the ocean carbon budget within a changing climate.

1. Introduction

The biological carbon pump transports organic carbon from the ocean's upper sunlit layer to depths of over a thousand meters (Turner, 2015). Biogeochemical (BGC) models estimate the amount of particulate organic carbon exported via this pathway to be between 5 and 12 GtC per year (Siegel et al., 2022). One key determinant of the amount of carbon exported is the grazing dynamics of the zooplankton community (Rohr et al., 2022). Accurate representation of these dynamics in BGC models is vital to reduce uncertainty in carbon export estimates used for climate change mitigation (Meyjes et al., 2024; Rohr et al., 2023).

Laboratory studies show that zooplankton grazing rates are influenced by species characteristics such as size, age and behavior (Hansen et al., 1997; Hirst & Bunker, 2003). Small herbivorous microzooplankton, for example, demonstrate faster grazing rates compared to larger suspension-feeding copepods (Hansen et al., 1997; Hirst & Bunker, 2003; Rohr et al., 2022). However, due to computational restraints, grazing rates within BGC models are frequently estimated for entire communities, rather than individual species. Therefore the diversity in grazing shown at a species level, needs to be appropriately expressed as community-averaged dynamics.

Recent modeling studies have shown large spatial variability in global community grazing dynamics, when models are tuned toward satellite-derived phytoplankton biomass and forced with realistic bottom-up controls on phytoplankton growth (Meyjes et al., 2024; Rohr et al., 2024). Variability in zooplankton grazing dynamics can therefore be derived indirectly from environmental conditions, however the explicit relationships between

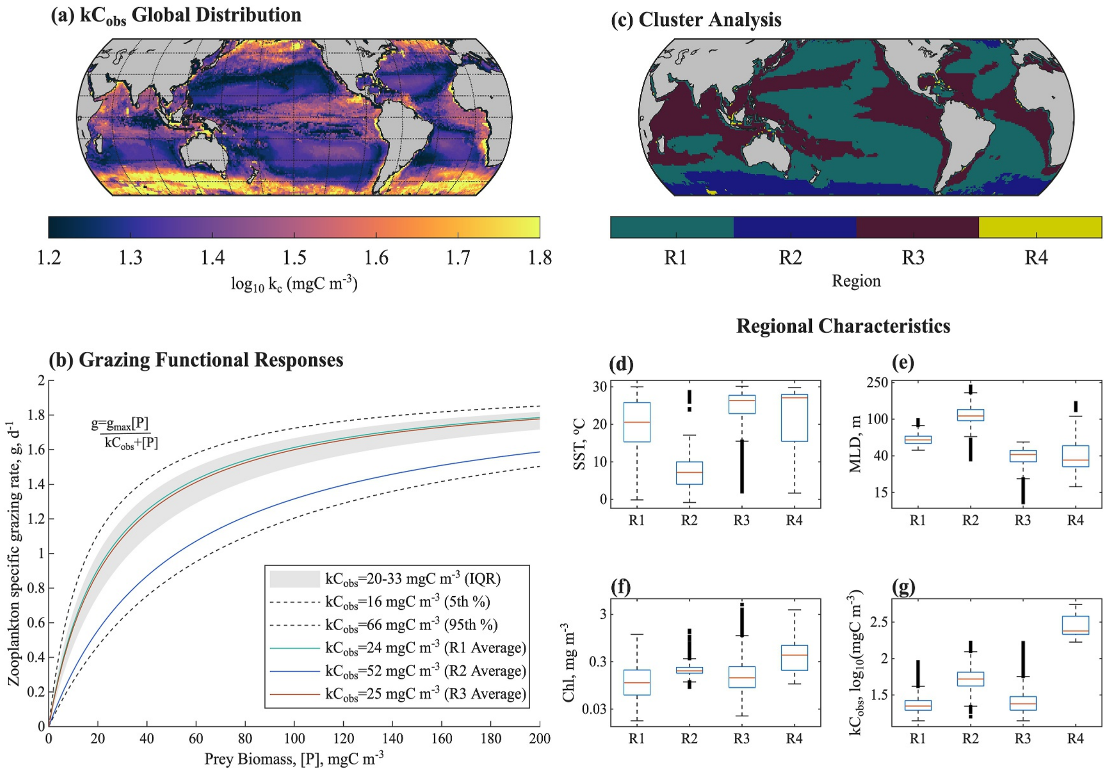

Figure 1. Estimation of zooplankton grazing dynamics and regional cluster analysis. (a) Global distribution of observationally inferred community grazing dynamics (kC_{obs}. (b) Illustrations of the grazing functional responses for: the inter-quartile range (IQR) of kC_{obs} values; the 5th and 95th percentiles of kC_{obs} values and the mean kC_{obs} values for Regions 1, 2 and 3 (R1, R2 and R3 respectively). Region 4 was not included due to size of the data set (Section 4.2). (c) Regional cluster analysis. Results of k-means cluster analysis, which divides the ocean into four distinct regions. (d–g) Box plots showing the satellite-derived data for sea surface temperature (SST), mixed layer depth (MLD) and chlorophyll (Chl), and the kC_{obs} values found in each region. Axes for mixed layer depth and chlorophyll are log scaled.

environmental drivers and zooplankton grazing dynamics are less defined. Environmental drivers of the ocean could be linked to zooplankton grazing dynamics through predator-prey interactions. Sea surface temperature and mixed layer depth, for example, impact the amount of light and nutrients available for phytoplankton communities, the prey species of zooplankton (Siegel et al., 2013; Talley, 2011). As phytoplankton species are adapted to thrive in different environments, with smaller species dominating in warmer, low nitrogen regions (Henson et al., 2021), the phytoplankton community composition in any location will therefore be determined by environmental conditions. In turn, the phytoplankton community dictates the type of zooplankton present due to differences in their grazing physiology and behavior (Hansen et al., 1997; Hirst & Bunker, 2003). However, the global spatial variability seen in zooplankton community grazing dynamics could be an emergent property of many competing environmental processes. In this study, we therefore use machine learning to disentangle their influence and quantify the dominant environmental controls of grazing. In doing so, we assess whether the underlying complexity in emergent zooplankton community grazing dynamics can be captured by environmental conditions.

2. Methodology

2.1. Overview

Grazing dynamics refer to the way in which zooplankton specific grazing rates vary with prey abundance and are often quantified by a functional response curve (Gentleman et al., 2003; Rohr et al., 2022; Figure 1). This response curve can be parameterized in global BGC models by two constants: the maximum grazing rate and the half-saturation constant or k value. Empirical measurements have shown k values vary with species characteristics such as size and age (Hansen et al., 1997; Hirst & Bunker, 2003; Rohr et al., 2022). Therefore, k values could reflect zooplankton community composition (Meyjes et al., 2024; Rohr et al., 2024).

To understand how zooplankton community grazing dynamics vary with environmental drivers, this study compared zooplankton community k values (Meyjes et al., 2024), with three climatological environmental drivers derived from remote sensing: sea surface temperature (SST), mixed layer depth (MLD) and chlorophyll. We used machine learning regression models to firstly quantify how much of the variance in community-averaged zooplankton grazing dynamics can be explained by the three environmental variables considered, and secondly to analyze the relationship between each variable and k. All methods were applied both globally and regionally, to understand if the resulting relationships varied with the spatial scale of the analysis. Below we describe in detail the data sources (Section 2.2) and the data analysis techniques (Section 2.3) employed.

2.2. Data Sources

2.2.1. Environmental Variables

The following environmental variables were considered for this study: sea surface temperature (SST), mixed layer depth (MLD) and chlorophyll (Figure S2 in Supporting Information S1). Monthly SST climatologies, averaged between 1955 and 2017, were sourced from the World Ocean Atlas 2018 (Garcia et al., 2019). Surface chlorophyll concentrations were taken from Sea Viewing Wide Field of View Sensor (SeaWiFS) measurements taken between 1997 and 2008 (http://orca.science.oregonstate.edu/1080.by.2160.monthly.hdf.chl.seawifs.php). The mixed layer depth monthly climatology was estimated from 5 million vertical profiles following de Boyer Montégut et al. (2004) and sourced from https://cerweb.ifremer.fr/deboyer/mld/Surface_Mixed_Layer_Depth.php (Siegel et al., 2014). These three variables were selected due to their roles as physical and biological drivers of ocean productivity and their widely available satellite-derived climatologies (Behrenfeld et al., 2005; Buitenhuis et al., 2013; Talley, 2011). Annual averages were calculated from each climatology as spatial, rather than temporal variation, was the focus of this study.

The resulting global distributions of the three environmental variables were then evaluated for correlation and collinearity. Correlation and collinearity between variables used in machine learning regression models may have several impacts including the detection of associations between variables and predictors, increasing model variance and the accurate estimation of variable importance (Chan et al., 2022; Dormann et al., 2013). Pairwise correlations between variables (e.g., annual mean MLD and SST) were estimated using a Kendall rank correlation matrix (Figure S3 in Supporting Information S1). Correlation coefficient values above 0.8 are generally considered to represent high correlation (Chan et al., 2022), however for this study only coefficient values below 0.5 were considered usable. The Variance Inflation Factor is a commonly used measure of collinearity, with values below 10 considered an acceptable level (Chan et al., 2022). All three environmental variables in this study meet these two criteria (Figure S3 in Supporting Information S1).

All data sources are gridded onto a 1×1 degree grid across the global ocean between roughly -50 and $+50°$ North. Polar regions are excluded due to a lack of satellite-derived products in these regions. The ocean area considered in this study covers approximately 2.93×10^8 km^2 (80% of ocean area).

2.2.2. Observationally Inferred Zooplankton Community Grazing Dynamics, kC_{obs}

Zooplankton community-averaged grazing dynamics are estimated using an inverse modeling approach detailed in Meyjes et al. (2024). Inverse modeling refers to estimating parameters from satellite-derived climatologies. A 0-D BGC box model is used to estimate the optimal grazing dynamics for both micro- and mesozooplankton (k_0 and k_1 values respectively) within the mixed layer of every grid cell of a $1 \times 1°$ global domain. The model is forced with phytoplankton growth rates derived from climatologies based on observations by SeaWiFS (McClain, 2009; Siegel et al., 2013). A suite of simulations determines the combination of k_0 and k_1 values which results in the best match between prognostic and satellite-derived phytoplankton biomass. This pair of k values is then selected as the optimal grazing dynamics for that grid cell. The biomass-weighted mean of k_0 and k_1 is then estimated, to generate a global distribution of observationally inferred community-averaged kC_{obs} values. These values are constant in time and reflect mean annual grazing dynamics.

Due to the importance of temperature to biological processes, temperature dependence terms were added to the model by Meyjes et al. (2024) before estimating the kC_{obs} values for this study (See Supplement). Overall, the explicit representation of temperature improved the model's ability to reproduce satellite derived phytoplankton biomass estimates by 28%, in comparison to Meyjes et al. (2024) (Figures S4 and S5 in Supporting Information S1).

2.3. Machine Learning Models and kC_{ML} Values

We use machine learning regression models to analyze the relationship between each variable and k. Machine Learning (ML) algorithms are considered appropriate for this study, over statistical models more regularly used in ecology (e.g., rank correlations (Figure S6 in Supporting Information S1), linear regression (Figure S7 in Supporting Information S1), GLMs), due to their greater predictive power and higher degree of flexibility, without the need to eliminate outliers (Elith et al., 2008; Pichler & Hartig, 2023). In addition, a ML approach learns natural relationships and patterns within data sets (Breiman, 2001; Elith et al., 2008) and as such, their application to ecological problems is becoming increasingly popular (Elith et al., 2008; Pichler & Hartig, 2023; Prasad et al., 2006).

Each ML model uses the environmental variable data to estimate k values which replicate the kC_{obs} as accurately as possible. The k values predicted by the ML models are referred to as kC_{ML} hereafter. The difference between the kC_{obs} and kC_{ML} values can be compared to understand the accuracy of the ML models applied. Further analysis can then be carried out using the ML models to understand the relationship between kC_{ML} values and the environmental variables (Section 2.4). Mixed layer depth, chlorophyll and the kC_{obs} values are log-transformed for the machine learning analysis.

Two popular classes of machine learning models are applied in this study: tree-based ensemble regression models and neural networks. Tree-based ensemble regression models average across multiple models, resulting in increased predictive power and reduced variance (Pichler & Hartig, 2023; Schapire, 1990). Examples include bagging (Breiman, 1996), random forest (Breiman, 2001), and boosting (Friedman, 2001). The latter differs in growing a suite of trees sequentially rather than independently (Hastie et al., 2009; Pichler & Hartig, 2023). Neural networks (McCulloch & Pitts, 1943) have strong predictive abilities and are based on artificial neurons, each layer receiving inputs from a previous layer and transmitting outputs to the next layer (Hastie et al., 2009). Both types of models allow the use of multiple predictors as inputs, can fit non-linear relationships and account for interactions between predictors—all of which are requirements for this study.

The same process of model training, optimisation and testing was used for both the tree-based ensemble and neural network models. Each model was trained using 10 fold cross-validation (Stone, 1974) on 90% of the data set. During model training hyperparameters were optimized using the Grid Search method (Figures S6 and S9–S12 in Supporting Information S1). Hyperparameter optimisation is vital to maximize predictive power of the models. Hyperparameters considered for tree-based ensemble model optimisation were: ensemble method (bagging using random forest algorithm or boosting), number of learners and minimum leaf size. Hyperparameters considered for neural network model optimisation were: number of fully connected layers, activation, regularization strength, size of each layer. In all scenarios, bagging using a random forest algorithm outperformed boosting and was selected as the ensemble method. All selected hyperparameters and search ranges can be seen in Table S1 in Supporting Information S1. The remaining 10% of the data set was then used to test the optimized models. Data was split 90:10 randomly using the MATLAB Regression Learner app. The R-squared and Root Mean Squared Error (RMSE) were calculated for each machine learning model.

2.4. Data Analysis

To understand the relationship between the three environmental variables and kC_{obs} on a regional scale, the ocean area was divided using cluster analysis (Figure 1). The aim of cluster analysis is for observations within groups to have lower dissimilarities compared to observations in different groups (Hastie et al., 2009). Here we apply a commonly used clustering method called k-means where squared Euclidean distance is used to measure dissimilarity (Hastie et al., 2009). K-means requires the selection of a cluster number. To provide assurances on the correct cluster number, two methods known as “Silhouette” (Rousseeuw, 1987) and “Davies Bouldin” (Davies & Bouldin, 1979) were used. The same machine learning approach, detailed above, was then applied to each region created from the cluster analysis. Cluster analysis was perfomed using the three environmental variables and observationally inferred kC_{obs} values.

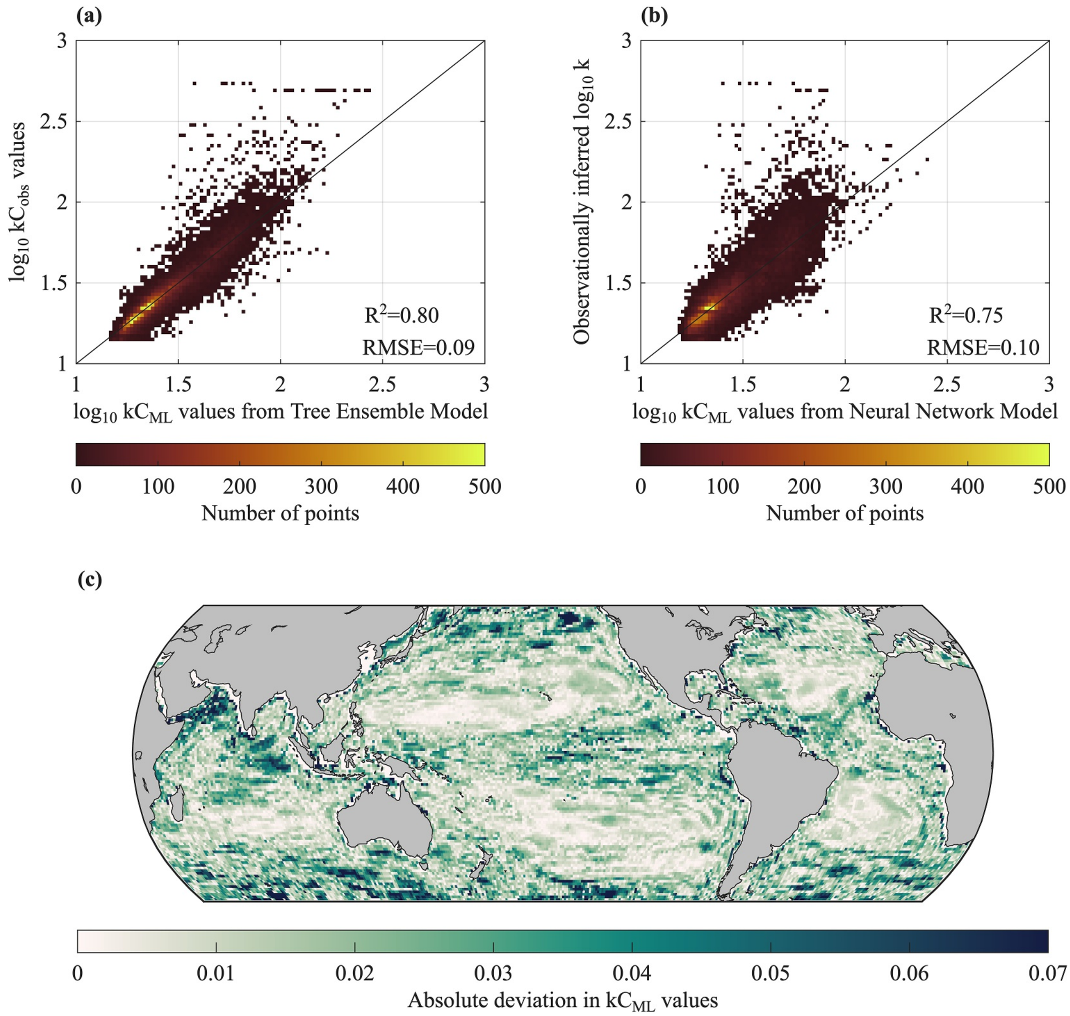

Figure 2. Global ML regression analysis. (a) Comparison between the k values predicted by the optimized tree-based ensemble model (kC_{ML} values) and those observationally inferred (kC_{obs}). (b) Comparison between the k values predicted by the optimized neural network model (kC_{ML} values) and those observationally inferred (kC_{obs}). Color bars represent the density of data in the binned scatter plot. (c) Absolute deviation of kC_{ML} values across the two ML models applied (Tree-based Ensemble Model and Neural Network Model). k values and deviation are in mgC m^{-3}.

Predictor importance analyses (Friedman, 2001; Molnar, 2022) and partial dependence plots (Friedman, 2001) were used to understand the dominance and relationship of environmental covariates to the estimation of kC_{ML} for the best-performing (highest R-squared and lowest RMSE) models within each region. These were used on both a global and regional scale. Predictor importance is based on the number of times a variable is selected at a branch node and the mean squared error from each resulting split (which indicates model improvement due to the variable; Elith et al., 2008; Friedman, 2001; Hastie et al., 2009). Partial dependence shows the average effect of the selected variable (e.g., SST) on kC_{ML} values, having taken into account other variables (Elith et al., 2008; Friedman, 2001; Friedman & Meulman, 2003).

3. Results

3.1. Predictability of Spatial Variation in Zooplankton Grazing Dynamics on a Global Scale

The majority of global spatial variation in zooplankton grazing dynamics can be explained by the three environmental variables considered in this study (SST, mixed layer depth and chlorophyll). The optimized neural network and tree-based ensemble models, respectively, have an R^2 of 0.75 and 0.80 and a root mean square error (RMSE) of 0.10 and 0.09 (Figures 2a and 2b and Table S2 in Supporting Information S1). This suggests high correlation between kC_{ML} and kC_{obs} values. In addition, there is good agreement between the two ML models used which results in a low absolute deviation in kC_{ML} values estimated by the two models (Figure 2c). However, neither ML model predicts kC_{ML} values above 273 mgC m^{-3} despite the observationally inferred kC_{obs} values reaching a maximum of 551 mgC m^{-3}.

3.2. Predictability of Spatial Variation in Zooplankton Grazing Dynamics on a Regional Scale

The cluster analysis divided the ocean into four regions, each with different environmental and zooplankton grazing characteristics (Figures 1c–1g). Both methods used for cluster analysis (Davies Boudin and Silhouette) were consistent in estimating cluster size (Figure S9 in Supporting Information S1). Region 1 has the lowest kC_{obs} values, on average for all regions (22 mgC m^{-3}), which would suggest this region is characterized by a community of smaller zooplankton species, as on average they demonstrate faster grazing rates when prey are limiting. It is also characterized by the second deepest mixed layer depths (median = 50 m) and average temperatures of around 21°C. It consists of large open ocean areas including the oligotrophic subtropical gyres. Region 2 is characterized by very low sea surface temperatures (7°C) and includes both the Southern Ocean and a small portion of the north Atlantic. This region contains the deepest mixed layer depths (median = 108 m) of the study area. Region 3 consists of generally warm waters (median = 26°C) with low kC_{obs} values, low chlorophyll and shallow mixed layer depths on average (24 mgC m^{-3}, 0.15 mg m^{-3} and 41 m respectively). The region covers large coastal areas in the subtropics and tropics, including where productive upwelling occurs. Region 4 is a very small data set of just 119 grid points (compared to 13,664, 3,602 and 10,344 for Regions 1, 2 and 3 respectively) and consists of extremely high kC_{obs} values. This size data set is too small for the machine learning methods used and therefore no further analysis for this region was made.

Within the three regions analyzed, SST, mixed layer depth and chlorophyll explain the majority of spatial variation in zooplankton grazing dynamics or kC_{obs} values (Figure S10 in Supporting Information S1). R^2 values from the optimized ML models range between 0.51 and 0.79 across regions 1, 2 and 3. Within these regions there is good correlation between k values estimated by the ML models (kC_{ML}) and those observationally inferred (kC_{obs}), with low deviation between models (Figure S10 in Supporting Information S1).

3.3. Environmental Constraints on Zooplankton Grazing Dynamics

Chlorophyll is the strongest predictor of kC_{ML} values both globally and regionally (Figure 3). The positive relationship between kC_{ML} values and chlorophyll, at concentrations above 0.1 mg m^{-3}, indicates low kC_{ML} values and fast community grazing occurs in areas of lower productivity. Higher kC_{ML} values and slower grazing is correlated with higher productivity. This relationship is seen both globally and regionally. Below chlorophyll concentrations of 0.1 mg m^{-3}, an inverse relationship occurs, which could suggest slower community grazing occurs when productivity is below a threshold productivity level.

Mixed layer depth is the second strongest predictor of kC_{ML} values. Globally, mixed layer depth is positively correlated with kC_{ML}, suggesting deeper mixed layer depths are associated with higher kC_{ML} values and a zooplankton community consisting of larger, slower grazing species. At shallower depths there is a weak relationship between kC_{ML} and mixed layer depth. Regionally, the relationship between mixed layer depth and kC_{ML} differs. In region 2, kC_{ML} values decrease with a deepening mixed layer until a depth of around 100 m is reached.

Temperature is the weakest predictor of kC_{ML} values globally, although it is a stronger predictor than mixed layer depth in regions 1 and 3 (Figure 3). No clear relationship is seen either globally or regionally between temperature and kC_{ML} values. Around 15 degrees Celsius, the lowest kC_{ML} values can be found globally, corresponding to a faster zooplankton grazing community. This suggests an optimal temperature for grazing zooplankton at the community level. In region 2, the average temperature only surpasses 15°C in a few instances (Section 4) and the lowest kC_{ML} values are associated with the lowest temperatures.

4. Discussion

4.1. Can Environmental Variables Explain Variability in Zooplankton Grazing Dynamics?

The majority of global spatial variation in community-averaged zooplankton grazing dynamics can be explained by chlorophyll, mixed layer depth (MLD) and sea surface temperature (SST). This suggests that underlying

10.1029/2025GL120153

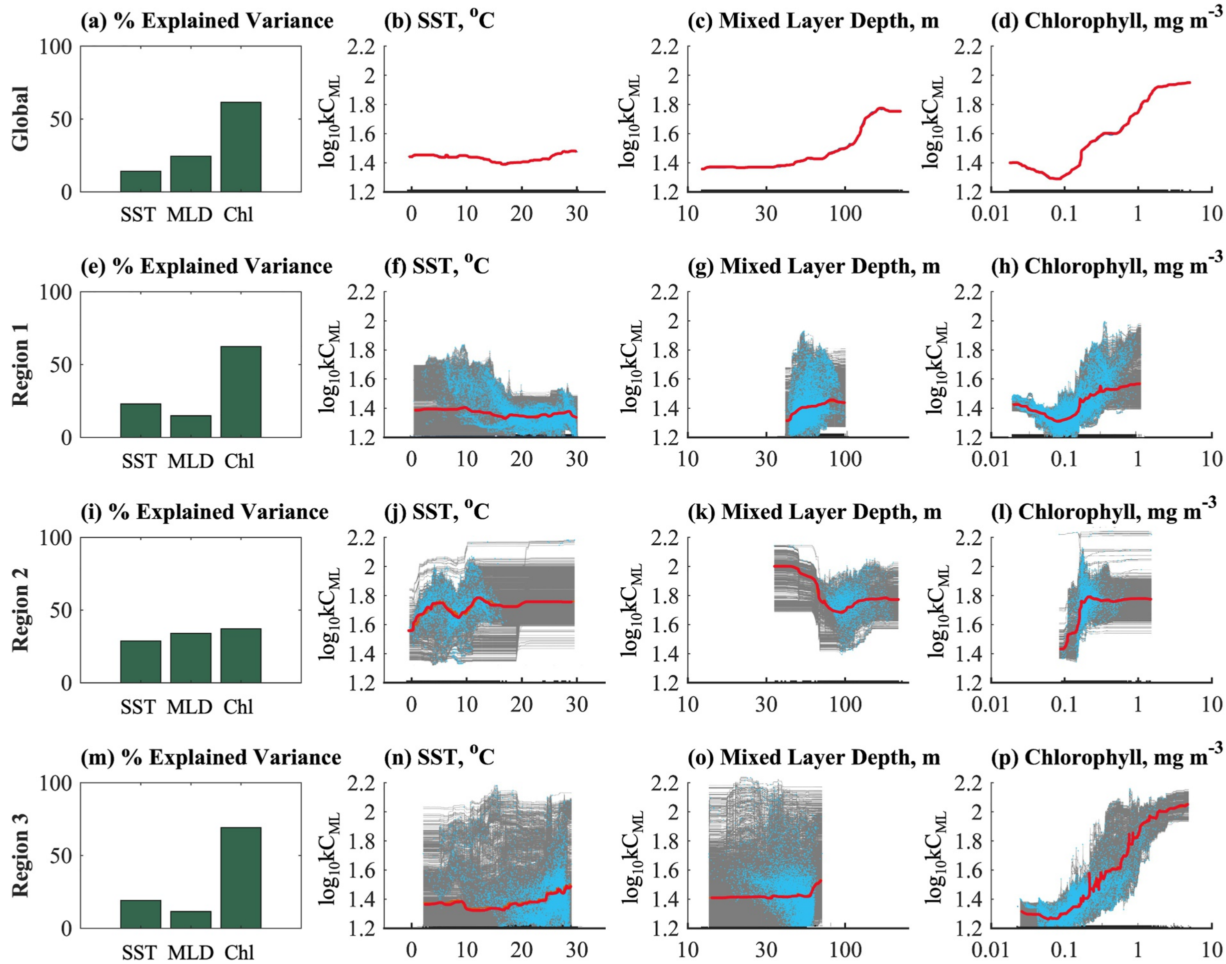

Figure 3. Predictor importance and partial dependence plots from both global and regional analyses using the tree-based ensemble ML models. Region 4 is excluded due to the very small data set. Bar charts indicate the importance of each variable to the prediction of kC_{ML} values. Importance is expressed as a percentage of explained variance in kC_{ML} values. The red lines on the line plots represent the partial dependence of the prediction of kC_{ML} values on the chosen variable. This is the average relationship across all the data for that region. The gray lines show the relationships for each data point, that is, they show the range of values that are averaged to form the red line. The scatter plots show the data points. The black bar, or rug plot, along the *x*-axis indicates available data for the analysis. Mixed layer depth and chlorophyll values are log-scaled. kC_{ML} values are in mgC m^{-3}.

complexity in environmental conditions can be captured in emergent zooplankton community grazing dynamics using the methods described by Meyjes et al. (2024). The good agreement between ML models and the observationally derived k values (kC_{obs}) also supports the use of these models in subsequent analysis to determine environmental drivers. The ML models used in this study cannot create an explicit parametric function to relate the variables to kC_{obs}, however these findings should act as a motivator for further work toward parametrizing this relationship in BGC models.

Both global ML models were limited in their ability to predict high kC_{ML} values (273 mgC m^{-3}), however these values make up less than 1% of the observationally derived data set. A review by Rohr et al. (2022) estimated an inter-quartile range of empirical measurements of k to be between 77 and 516 mgC m^{-3}. Therefore, the observationally derived high values in the training data set are thought to be within realistic ranges, and instead the discrepancy between ML model predictions and observations indicates a limitation of the modeling approach. Several previous studies have highlighted the limitations of both statistical and ML models in predicting extremes (e.g., Qi & Majda, 2020; Ribeiro & Moniz, 2020). To capture univariate extremes in a data set, alternative approaches such as those using extreme value theory could be considered (Gomes & Guillou, 2015).

The proportion of explained variation in community-averaged zooplankton grazing dynamics varies regionally. Regions were defined by cold SSTs (Region 2) and differences in mean MLDs (Regions 1 and 3) in line with

10.1029/2025GL120153

previous findings that defined biomes using mixed layer depth (Banse, 1992; Luo et al., 2020; Stock et al., 2014) and phytoplankton biogeography (Elizondo et al., 2021). The ML models were more effective in predicting kC_{ML} values in regions 1 and 3 which were substantially larger data sets, with larger dynamic ranges. Smaller data set sizes reduce the effectiveness of ML regression analysis (Pichler & Hartig, 2023).

4.2. The Relationship Between Chlorophyll and Zooplankton Grazing Dynamics

The relationship between chlorophyll and zooplankton grazing is the strongest of all variables considered in this study, both globally and regionally. Chlorophyll is an emergent biological response that integrates over physical properties. This means that SST and MLD rarely add more predictability after taking into account chlorophyll, which has integrated their effects on phytoplankton biological processes.

An increase in chlorophyll is generally associated with an increase in kC_{ML} values, which suggests a shift in the zooplankton community to slower grazing species. Zooplankton communities in high biomass areas of the ocean, for example, coastal upwelling areas, are dominated by slow grazing copepods (Steinberg & Landry, 2016), in contrast to the low biomass subtropical gyres where faster grazing pico- and nano-size flagellates dominate (Calbet & Calbet, 2008). Therefore, a shift from low biomass (low chlorophyll) areas to high biomass would be associated with a shift to slower grazing plankton communities, characterized by high k values.

Below chlorophyll concentrations of 0.1 mg m^{-3} (32% of the data set) the relationship with kC_{ML} values is reversed, with kC_{ML} values increasing as chlorophyll decreases. Chlorophyll is lowest in the center of the oligotrophic subtropical gyres, where nutrient limitation keeps biomass low and drives diversity (Signorini et al., 2015). The model used in the inverse modeling approach has a simplistic structure (2P2Z, Meyjes et al., 2024) and these diverse regions are more difficult for the mechanistic model to capture. The poor skill in modeling oligotrophic regions is quantified and made evident via the inverse modeling approach, with greater residuals between prognostic and observationally derived phytoplankton biomass in these regions (Figure S4b in Supporting Information S1). This limitation could be addressed by the inclusion of more phytoplankton types (e.g., diazotrophs) or other processes (e.g., photoacclimation) not currently represented.

The relationship between chlorophyll and kC_{ML} values is consistent with previous findings by Rohr et al. (2024) despite key differences in the approach. Rohr et al. (2024) found a sigmoidal relationship between k and observed phytoplankton concentrations. This study controls for several limitations of the methodology used by Rohr et al. (2024) by including: observed phytoplankton growth rates and temperature-limited grazing to infer k values; a more sophisticated ML approach which isolates the role of chlorophyll; and the use of chlorophyll, a more directly observed predictor variable than biomass. Despite these differences, a similar sigmoidal relationship is found. This suggests that, firstly, community-averaged grazing dynamics and therefore zooplankton community composition is well constrained by prey stocks, and secondly that biological drivers could be used to parameterize community-averaged dynamics in BGC models.

4.3. The Relationship Between Mixed Layer Depth and Zooplankton Grazing Dynamics

Globally, an increased mixed layer depth is associated with higher values of the grazing parameter kC_{ML}. Mixed layer depth is greatest in the higher latitudes where large phytoplankton species such as diatoms dominate (Murphy et al., 2021; Figure S2 in Supporting Information S1). These larger prey cannot be consumed by smaller microzooplankton. This results in microzooplankton being outcompeted by larger copepod species, despite having faster prey capture rates, which in turn could lead to the emergent property of a zooplankton community with a higher average k value (Hansen et al., 1997; Hirst & Bunker, 2003; Rohr et al., 2022).

Mixed layer depth is a weak predictor of global zooplankton grazing dynamics compared to chlorophyll, however, previous research has shown that mixed layer depth has a strong seasonal influence on plankton communities, driving spring blooms (e.g., Behrenfeld et al. (2013)). This study focused on spatial variation only due to the lack of temporal variability in the observationally derived kC_{obs} values. However, further temporal analysis that details seasonal time scales and the succession of plankton communities over it, could result in an increased importance of mixed layer depth to zooplankton community grazing dynamics. The lack of seasonal variability due to the use of annual climatological means is a limitation of this study and an area for future research.

The relationship between mixed layer depth and kC_{ML} values appears to be reversed in region 2, once the mixed layer is shallower than 100 m (Figure 3). Shallow mixed layer depths may occur from ice melt or low salinity waters that keep phytoplankton near the surface and increase productivity, leading to higher k values. However, a lack of data points at these shallower depths results in low confidence and high error estimates for the predicted relationship (Figure S16 in Supporting Information S1).

4.4. The Relationship Between Temperature and Zooplankton Grazing Dynamics

There is no clear relationship between sea surface temperature and zooplankton grazing dynamics, either globally or regionally. It is possible that temperature (and mixed layer depth) could be an indirect driver of kC_{obs}, by orchestrating changes in chlorophyll concentrations, however this covariablity was controlled for in the machine learning methods used. Additionally, in the model used to infer kC_{obs} values, the same Q_{10} values are used for grazing, growth and mortality for both micro- and mesozooplankton species. This is common practice in global models but likely does not reflect reality (Ferreira et al., 2022).

4.5. Limitations

There are several limitations to this modeling study. Firstly, kC_{obs} values were estimated using an inverse modeling approach (Meyjes et al., 2024) which uses phytoplankton division rates derived from Net Primary Productivity from the Carbon-based Productivity Model v2 (CbPMv2) as a forcing variable. For consistency the same forcing variables were used in this study, however CAFE (Silsbe et al., 2016) is thought to be a more realistic estimate (Bisson et al., 2018). For a full list of limitations in the kC_{obs} inverse modeling approach see Meyjes et al. (2024). Secondly, the environmental variables used in this analysis will also have influenced the estimation of kC_{obs} values using climatological-derived phytoplankton growth rates, following Meyjes et al. (2024). This potential circularity is necessary to enable the subsequent analysis of the environmental drivers of kC_{obs}, however further work could experiment with additional drivers not considered here. Thirdly, machine learning methods have various constraints. Increasing model flexibility can lead to overfitting and small changes in training data sets can result in largely different decision trees (Elith et al., 2008; Pichler & Hartig, 2023). Additionally, the test and training data sets were allocated randomly in a 90:10 split, however in the spatial data used, nearby points will be correlated. Spatial sampling could be used to avoid the test data inadvertently peaking due to spatially close data points, which may result in an overestimation of model performance (Roberts et al., 2017). Finally, the use of ML models for causal inference is debated, particularly neural networks, however research into this area is increasing (Hastie et al., 2009; Pichler & Hartig, 2023).

4.6. Future Considerations

The three environmental variables considered in this study explain the majority of variation in the spatial distribution of zooplankton community grazing dynamics. Unexplained spatial variation could be described by the addition of further variables not considered here such as photosynthetically important nutrients. However, the findings show that biological drivers (i.e., chlorophyll) could be used to parameterize more realistic zooplankton grazing in biogeochemical (BGC) models. The sigmoid relationship found between chlorophyll and k values could be used as a diagnostic tool to derive spatially varying k values for zooplankton communities from observationally derived chlorophyll estimates for a given area. This relationship is in agreement with findings by Rohr et al. (2024), however, modellers must consider that the ML models cannot produce parametric relationships between drivers and response and additional considerations need to be made to reflect how community composition changes over time. It is also possible that spatial and temporal variability in community-integrated grazing dynamics could be achieved through explicit competition if adequate zooplankton groups are included (Liu et al., 2025). Either way, physical and biological ocean drivers are continually being influenced by climate change (Hoegh-Guldberg & Bruno, 2010), and any parametrization of grazing in BGC models would need to be responsive to climate projections. Finally, we also found improved model agreement with observations of plankton biomass when temperature-dependence of zooplankton processes is included. Therefore, this application of temperature-dependence is recommended for BGC models.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

Data is available on Zenodo via Meyjes (2025). All analysis was carried out in Matlab version R2023b. Monthly SST climatologies were sourced from the World Ocean Atlas 2018 (Garcia et al., 2019). Surface chlorophyll concentrations were taken from Sea Viewing Wide Field of View Sensor (SeaWiFS) measurements (McClain, 2009; Siegel et al., 2013). The mixed layer depth monthly climatology was sourced from de Boyer Montégut et al. (2004).

Acknowledgments

SM acknowledges support from the Natural Environment Research Council through the C-CLEAR Doctoral Training Partnership (Grant NE/S007164/1). TR is an ARC DECRA recipient (DE240100115) funded by the Australian Government. AM would like to acknowledge funding from NERC Grant NE/P018319/1 and ONR Grant N00014-22-1-2082. BBC and TP received support through Schmidt Sciences (project: CALIPSO) and by the Horizon Europe (project: 101059915, BIOcean5D). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. CMP was supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections Program Grant NA20OAR4310438, NA20OAR4310441 and NA20OAR4310442. Authors would like to thank Dr Joo-Eun Yoon for their help with initial project development.

References

Banse, K. (1992). Grazing, temporal changes of phytoplankton concentrations, and the microbial loop in the open sea. In *Primary productivity and biogeochemical cycles in the sea*. https://doi.org/10.1007/978-1-4899-0762-2_22

Behrenfeld, M. J., Boss, E., Siegel, D. A., Shea, D. M., Behrenfeld, M. J., Boss, E., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, *19*, 1–14. https://doi.org/10.1029/2004GB002299

Behrenfeld, M. J., Doney, S. C., Lima, I., Boss, E. S., & Siegel, D. A. (2013). Annual cycles of ecological disturbance and recovery underlying the subarctic Atlantic spring plankton bloom. *Global Biogeochemical Cycles*, *27*(2), 526–540. https://doi.org/10.1002/gbc.20050

Bisson, K. M., Siegel, D. A., DeVries, T., Cael, B. B., & Buesseler, K. O. (2018). How data set characteristics influence ocean carbon export models. *Global Biogeochemical Cycles*, *32*(9), 1312–1328. https://doi.org/10.1029/2018GB005934

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/bf00058655

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12343 LNCS.

Buitenhuis, E. T., Vogt, M., Moriarty, R., Bednaršek, N., Doney, S. C., Leblanc, K., et al. (2013). Earth system science data MAREDAT: Towards a world atlas of MARine ecosystem DAta. *Earth System Science Data*, *5*(2), 227–239. https://doi.org/10.5194/essd-5-227-2013

Calbet, A., & Calbet, A. (2008). The trophic roles of microzooplankton in marine systems. *ICES Journal of Marine Science*, *65*(3), 325–331. https://doi.org/10.1093/ICESJMS/FSN013

Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the multicollinearity problem and its machine learning approach. *A review*, *10*(8), 1283. https://doi.org/10.3390/math10081283

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. In *IEEE transactions on pattern analysis and machine intelligence, PAMI-1*. https://doi.org/10.1109/TPAMI.1979.4766909

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*(C12). https://doi.org/10.1029/2004JC002378

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elizondo, U. H., Righetti, D., Benedetti, F., & Vogt, M. (2021). Biome partitioning of the global ocean based on phytoplankton biogeography. *Progress in Oceanography*, *194*. https://doi.org/10.1016/j.pocean.2021.102530

Ferreira, G. D., Grigoropoulou, A., Saiz, E., & Calbet, A. (2022). The effect of short-term temperature exposure on vital physiological processes of mixoplankton and protozooplankton. *Marine Environmental Research*, *179*, 105693. https://doi.org/10.1016/j.marenvres.2022.105693

Friedman, J. H. (2001). 1999 Reitz lecture greedy function approximation: A gradient boosting machine (Technical Report) (Vol. *29*).

Friedman, J. H., & Meulman, J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, *22*(9), 1365–1381. https://doi.org/10.1002/sim.1501

Garcia, H., Boyer, T., Baranova, O., Locarnini, R., Mishonov, A., Grodsky, A., et al. (2019). World ocean atlas 2018: Product documentation. In A. Mishonov (Technical Editor). Retrieved from https://data.nodc.noaa.gov/woa/WOA18/DOC/woa18documentation.pdf

Gentleman, W., Leising, A., Frost, B., Strom, S., & Murray, J. (2003). Functional responses for zooplankton feeding on multiple resources: A review of assumptions and biological dynamics. *Deep Sea Research Part II: Topical Studies in Oceanography*, *50*(22–26), 2847–2875. https://doi.org/10.1016/J.DSR2.2003.07.001

Gomes, M. I., & Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review*, *83*(2), 263–292. https://doi.org/10.1111/insr.12058

Hansen, P. J., Bjørnsen, P. K., & Hansen, B. W. (1997). Zooplankton grazing and growth: Scaling within the 2-2-μm body size range. *Limnology and Oceanography*, *42*(4), 687–704. https://doi.org/10.4319/LO.1997.42.4.0687

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd ed.). Springer.

Henson, S. A., Cael, B. B., Allen, S. R., & Dutkiewicz, S. (2021). Future phytoplankton diversity in a changing climate. *Nature Communications*, *12*(1), 5372. https://doi.org/10.1038/S41467-021-25699-W

Hirst, A. G., & Bunker, A. J. (2003). Growth of marine planktonic copepods: Global rates and patterns in relation to chlorophyll a, temperature, and body weight. *Limnology & Oceanography*, *48*(5), 1988–2010. https://doi.org/10.4319/lo.2003.48.5.1988

Hoegh-Guldberg, O., & Bruno, J. F. (2010). The impact of climate change on the world’s marine ecosystems. *Science*, *328*(5985), 1523–1528. https://doi.org/10.1126/science.1189930

Liu, Y., Richardson, A., Strutton, P. G., & Rohr, T. W. (2025). Assessment of emergent spatial variability of community-integrated grazing dynamics in cmip6 models. https://doi.org/10.22541/essoar.175587856.62298607/v1

Luo, J. Y., Condon, R. H., Stock, C. A., Duarte, C. M., Lucas, C. H., Pitt, K. A., & Cowen, R. K. (2020). Gelatinous zooplankton-mediated carbon flows in the global oceans: A data-driven modeling study. *Global Biogeochemical Cycles*, *34*(9), e2020GB006704. https://doi.org/10.1029/2020GB006704

McClain, C. R. (2009). A decade of satellite ocean color observations. *Annual Review of Marine Science*, *1*, 19–42. https://doi.org/10.1146/ANNUREV.MARINE.010908.163650

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

Meyjes, S. A. (2025). Data for journal submission to GRL by Meyjes et al [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.17433221

Meyjes, S. A., Petrik, C. M., Rohr, T. W., Cael, B. B., & Mashayek, A. (2024). Impact of spatial variability in zooplankton grazing rates on carbon export flux. https://doi.org/10.22541/essoar.170431173.37869152/v1

Molnar, C. (2022). Interpretable machine learning.
Murphy, E. J., Johnston, N. M., Hofmann, E. E., Phillips, R. A., Jackson, J. A., Constable, A. J., et al. (2021). Global connectivity of Southern Ocean ecosystems. *Frontiers in Ecology and Evolution*, *9*, 624451. https://doi.org/10.3389/fevo.2021.624451
Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, *14*(4), 994–1016. https://doi.org/10.1111/2041-210X.14061
Prasad, A., Iverson, L., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, *9*(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1
Qi, D., & Majda, A. J. (2020). Using machine learning to predict extreme events in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(1), 52–59. https://doi.org/10.1073/pnas.1917285117
Ribeiro, R. P., & Moniz, N. (2020). Imbalanced regression and extreme value prediction. *Machine Learning*, *109*(9–10), 1803–1835. https://doi.org/10.1007/s10994-020-05900-9
Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881
Rohr, T., Anthony, J. R., Lenton, A., Chamberlain, M. A., & Shadwick, E. H. (2023). Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in cmip6 models. *Nature Communications Earth & Envrionment*, *4*.
Rohr, T., Richardson, A., Lenton, A., Chamberlain, M., & Shadwick, E. (2024). The global distribution of grazing dynamics estimated from inverse modelling. *Geophysical Research Letters*, *51*(8), e2023GL107732. https://doi.org/10.1029/2023GL107732
Rohr, T., Richardson, A. J., Lenton, A., & Shadwick, E. (2022). Recommendations for the formulation of grazing in marine biogeochemical and ecosystem models. *Progress in Oceanography*, *208*, 102878. https://doi.org/10.1016/J.POCEAN.2022.102878
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, *5*(2), 197–227. https://doi.org/10.1007/bf00116037
Siegel, D. A., Behrenfeld, M. J., Maritorena, S., McClain, C. R., Antoine, D., Bailey, S. W., et al. (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sensing of Environment*, *135*, 77–91. https://doi.org/10.1016/J.RSE.2013.03.025
Siegel, D. A., Buesseler, K. O., Doney, S. C., Sailley, S. F., Behrenfeld, M. J., & Boyd, P. W. (2014). Global assessment of ocean carbon export by combining satellite observations and food-web models. *Global Biogeochemical Cycles*, *28*(3), 181–196. https://doi.org/10.1002/2013GB004743
Siegel, D. A., DeVries, T., Cetinić, I., & Bisson, K. M. (2022). Quantifying the ocean's biological pump and its carbon cycle impacts on global scales. *Annual Review of Marine Science*, *15*(1), 329–356. https://doi.org/10.1146/ANNUREV-MARINE-040722-115226
Signorini, S. R., Franz, B. A., & McClain, C. R. (2015). Chlorophyll variability in the oligotrophic gyres: Mechanisms, seasonality and trends. *Frontiers in Marine Science*, *2*. https://doi.org/10.3389/fmars.2015.00001
Silsbe, G. M., Behrenfeld, M. J., Halsey, K. H., Milligan, A. J., & Westberry, T. K. (2016). The cafe model: A net production model for global ocean phytoplankton. *Global Biogeochemical Cycles*, *30*(12), 1756–1777. https://doi.org/10.1002/2016GB005521
Steinberg, D. K., & Landry, M. R. (2016). Zooplankton and the ocean carbon cycle. *Annual Review of Marine Science*, *9*(1), 413–444. https://doi.org/10.1146/annurev-marine-010814-015924
Stock, C. A., Dunne, J. P., & John, J. G. (2014). Global-scale carbon and energy flows through the marine planktonic food web: An analysis with a coupled physical–biological model. *Progress in Oceanography*, *120*, 1–28. https://doi.org/10.1016/J.POCEAN.2013.07.001
Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society - Series B: Statistical Methodology*, *36*(2), 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x
Talley, L. D. (2011). Descriptive physical oceanography: An introduction. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=Chb14jomm08C&oi=fnd&pg=PP1&dq=descriptive+physical+oceanography+lynne+talley&ots=fBjKTXuozg&sig=-6ryQMcqDkf3jzDIlRwcHLiKzLo&redir_esc=y#v=onepage&q=descriptive%20physical%20oceanography%20lynne%20talley&f=false
Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, *130*, 205–248. https://doi.org/10.1016/j.pocean.2014.08.005

References From the Supporting Information

Kearney, K. A., Bograd, S. J., Drenkard, E., Gomez, F. A., Haltuch, M., Hermann, A. J., et al. (2021). Using global-scale Earth system models for regional fisheries applications. *Frontiers in Marine Science*, *8*, 622206. https://doi.org/10.3389/fmars.2021.622206
Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., et al. (2020). Ocean biogeochemistry in GFDL's Earth system model 4.1 and its response to increasing atmospheric CO_2. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2019MS002043. https://doi.org/10.1029/2019MS002043
Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, *22*(2). https://doi.org/10.1029/2007GB003078